arXiv:2505.07314v1 [math.OC] 12 May 2025

Sparsity for dynamic inverse problems on Wasserstein curves with bounded variation

Marcello Carioni* Julius Lohmann†

May 13, 2025

Abstract

We investigate a dynamic inverse problem using a regularization which implements the so-called Wasserstein-1 distance. It naturally extends well-known static problems such as lasso or total variation regularized problems to a (temporally) dynamic setting. Further, the decision variables, realized as BV curves, are allowed to exhibit discontinuities, in contrast to the design variables in classical optimal transport based regularization techniques. We prove the existence and a characterization of a sparse solution. Further, we use an adaption of the fully-corrective generalized conditional gradient method to experimentally justify that the determination of BV curves in the Wasserstein-1 space is numerically implementable.

Keywords: BV curves, conditional gradient methods, dynamic inverse problems, optimal transport, sparse optimization, Wasserstein distance

Contents

*Faculty of Electrical Engineering, Mathematics and Computer Science, University of Twente, m.c.carioni@utwente.nl
†International Research Fellow of Japan Society for the Promotion of Science, Department of Mathematics, Institute of Science Tokyo, lohmann.j.aa@m.titech.ac.jp

1 Introduction

Consider an *inverse problem* composed of a *fidelity term* and *regularization term*,

$$\inf_{v\in V} F(Av)+\alpha R(v),$$

where $\alpha \in \mathbb{R}_+^*$ is a *regularization parameter*, $A : V \to W$ is a *forward operator* from a *model space* V to a *data space* W, $F : W \to \mathbb{R}\cup\{\infty\}$ is a *fidelity*, and R is a *regularizer*.

In this article, we are particularly interested in so-called *sparse optimization* where the regularizer R enforces a 'sparsity property'. We briefly recall some well-known instances in this context. The first is termed ℓ^1*-regularization* or *lasso* [52], for example in the formulation

$$\inf_{v\in\mathbb{R}^N} \|Av-f\|_2^2+\alpha\|v\|_1,$$

where $A \in \mathbb{R}^{N\times N}$, $f \in \mathbb{R}^N$, and $\|\cdot\|_p$ denotes the ℓ^p-norm, that is $\|v\|_p^p = |v_1|^p+\ldots+|v_N|^p$. If A is orthonormal, then it is straightforward to check that the solution is unique and given by $S_{\alpha/2}(A^\intercal f)$, where thresholding operator $S_{\alpha/2}$ is applied component-wise and given by

$$S_{\alpha/2}(x) = \begin{cases} 0 & \text{if } x \le \alpha/2, \\ x-\operatorname{sign}(x)\alpha/2 & \text{else.} \end{cases}$$

Hence, the larger α, the more zero entries in the solution, which is called *sparsity* in this context. A variant of this problem is given by (see for instance [39, Ch. 6])

$$\inf_{v\in\mathbb{R}^N} \|Av-f\|_2^2+\alpha\|Mv\|_1,$$

where $M \in \mathbb{R}^{M\times N}$. For example, one can take matrix $M = \binom{M_1}{M_2} \in \mathbb{R}^{2N\times N}$, with $M_1, M_2 \in \mathbb{R}^{N\times N}$, as a discretization of the two-dimensional gradient operator applied to some grayscale image with $N = N_1N_2$ pixels stored in the vector v. Then the problem can be seen as a discretization of so-called *total variation regularization* (cf. [42]). The regularizer $R(v) = \|Mv\|_1$ then enforces sparsity of the (discrete) gradient which particularly yields sharp edges in the image (which is to be reconstructed from possibly noisy data f). A solution can be obtained by reformulating the problem as a quadratic program [27]. In a spatially continuous setting, one may replace $\mathbb{R}^N$ by some non-empty bounded Lipschitz domain $U \subset \mathbb{R}^n$ and regularizer $v \mapsto \|Mv\|_1$ by the total variation of a BV function [14],

$$\inf_{v\in BV(U)} F(Av)+|\mathrm{D}v|(U),$$

where we assume that $A : BV(U) \to \mathbb{R}^m$ is continuous and linear with $A(BV(U)) = \mathbb{R}^m$ and $F : \mathbb{R}^m \to \mathbb{R} \cup \{\infty\}$ is coercive, convex, lower semicontinuous, and proper. In this setting, there exists a sparse solution [14, Thm. 4.8]

$$v^{opt} = c + \sum_{i=1}^{L} \frac{c^i}{|\mathrm{D}1_{E_i}|(U)} 1_{E^i},$$

where $c \in \mathbb{R}, c_i \in \mathbb{R}_+^*, L \leq \dim(\mathbb{R}^m / A(\mathbb{R}))$, and the $E_i \subset U$ are *simple* (see [14, Def. 4.4 & 4.5]), in particular each E^i cannot be decomposed into two sets with positive volume such that the sum of their perimeters equals the perimeter of E^i, that is $|\mathrm{D}1_{E^i}|(U)$ (1_{E^i} denoting the characteristic function of E^i). Hence, sparsity in this configuration can be interpreted as being piecewise constant, where 'piecewise' is relative to the simple sets E^i (respectively their intersections). Recall that, by the Riesz representation theorem and the definition of $|\mathrm{D}v|(U)$, we can interpret the distributional gradient of $v \in BV(U)$ as a vector-valued Radon measure $\mathrm{D}v \in \mathscr{M}(U;\mathbb{R}^n)$. For general vector-valued Radon measures, one may consider the problem in [20],

$$\inf_{v \in \mathscr{M}(X;\mathbb{R}^k)} \|Av - g\|_H^2 + \alpha \|v\|_{\mathscr{M}},$$

where X is a locally compact and separable metric space, H is a Hilbert space ($g \in H$), $A : \mathscr{M}(X;\mathbb{R}^k) \to H$ is the adjoint of some continuous and linear $A_* : H \to C_0(X;\mathbb{R}^k)$, and $\|\cdot\|_{\mathscr{M}}$ denotes the total variation norm. For $X \subset \mathbb{R}^n$ non-empty, bounded, and open, $k = 1$, finite-dimensional H, and $A(\mathscr{M}(X;\mathbb{R})) = H$, there exists a minimizer which can be written [14, Thm. 4.2]

$$v^{opt} = \sum_{i=1}^{N} \lambda^i \delta_{x^i},$$

where $N \leq \dim(H), \lambda^i \in \mathbb{R}$, and $x^i \in X$ (δ_{x^i} denotes the Dirac measure centered at x^i).

In this paper, we investigate sparsity for a *dynamic inverse problem* whose regularizer is related to the above problems. As a motivation, recall that the above regularizers enforce the decision variables to be concentrated on few simple geometric objects, for example time points of a discrete signal $v \in \mathbb{R}^N$ if $R(v) = \|v\|_1$, spikes of $v \in \mathscr{M}(X;\mathbb{R})$ if $R(v) = \|v\|_{\mathscr{M}}$, adjoint pixels of $v \in \mathbb{R}^N$ if $R(v) = \|M_1 v\|_1 + \|M_2 v\|_1$, or simple sets in the support of $v \in BV(U)$ if $R(v) = |\mathrm{D}v|(U)$. In the dynamic, time-dependent case, this *simpleness* should be reflected in objects which evolve over time. Another property of the above regularizers is that they separate the contributions of the different geometric objects: ℓ^1-norm $R = \|\cdot\|_{\ell^1}$ is a sum over the entries, total variation $R = \|\cdot\|_{\mathscr{M}}$ is a supremum over partitions (this also applies to $R(v) = |\mathrm{D}v|(U)$ if $\mathrm{D}v$ is interpreted as a vector-valued measure). Therefore, this *separating* property should be observed in the dynamics as well.

A natural setup which implements the above mentioned simpleness and separating can be obtained by considering a well-known distance from optimal transport theory: We have seen that $R(v) = \|v\|_{\mathscr{M}}(X)$ is used in the case of vector-valued Radon measures $v \in \mathscr{M}(X;\mathbb{R}^k)$. This norm actually appears in the Beckmann formulation of the *Wasserstein*-1 *distance* [44, Thm. 4.6] between (compactly supported) probability measures $\rho^+, \rho^- \in \mathscr{P}(\mathbb{R}^n)$,

$$W_1(\rho^+, \rho^-) = \min_{v} \|v\|_{\mathscr{M}},$$

where the minimum is over vector-valued Radon measures $v \in \mathscr{M}(\mathbb{R}^n;\mathbb{R}^n)$ with distributional divergence equal to $\rho^+ - \rho^-$. One can restrict this minimization to compactly supported and 'loop-free' v and each candidate minimizer can be interpreted as a continuous curve $t \mapsto \mu_t \in (\mathscr{P}(\mathbb{R}^n), W_1)$ (see section 3.3), where t represents a time variable. Such μ and their evolution with respect to W_1 are a reasonable choice for a dynamic inverse problem implementing the above characteristics which we will briefly clarify. First, the Wasserstein-1 distance favors trajectories on which there is no direct interaction between the different mass particles which reflects the desired simpleness. More specifically, the Wasserstein-1 distance minimizes $\int |x-y| \mathrm{d}\pi(x,y)$ over π, where $|x-y|\mathrm{d}\pi(x,y)$ is the (infinitesimal) transportation cost (in terms of Euclidean distance) of moving (infinitesimal) amount of mass $\mathrm{d}\pi(x,y)$ from x to y — there is no efficiency gain when particles bundle because the transportation cost is linear in the mass. Regularization with variation $\mathrm{var}(\mu)$ would ensure that the total accumulated travel distance of all mass particles is kept small. In particular, if the curve μ is only concentrated in one (Euclidean) curve in $\mathbb{R}^n$, then it will tend to be close to a traverse in the corresponding dynamic inverse problem (in a discrete setting with data measured at finitely many time points). Second, the separating property is given by Smirnov's decomposition [50, Thm. C]: one can write the total variation $\|v\|_{\mathscr{M}}$ (where v can be interpreted as a normal 1-current in $\mathbb{R}^n$ whose boundary is equal to $\rho^- - \rho^+$) as an integral over the contributions of the different particle trajectories. We do not only allow for continuous curves (which may show diffusive behaviour), but also curves with jumps. The possibility to jump allows for the adjustment to heavily scattered data even on a small time scale — however, such jumps are penalized through the regularization. This approach (and the resulting weak assumption on the regularity of a curve) extends the optimal transport based regularization techniques used so far, as we will highlight in more detail below. Finally, as in the above examples, the total mass should also appear in the regularizer. Hence, we consider $\mathbb{R}_+\mathscr{P}(\mathbb{R}^n)$ instead of $\mathscr{P}(\mathbb{R}^n)$ (the Wasserstein-1 distance apparently just rescales). Our proposed regularizer of $\mu \in \mathbb{R}_+BV([0,T];(\mathscr{P}(\Omega),W_1))$ ($\Omega \subset \mathbb{R}^n$ compact and convex, $[0,T]$ some time interval) is given by

$$T\omega(\mu) + \operatorname{ess\,var}(\mu),$$

where $T\omega(\mu) \in \mathbb{R}_+$ represents the time integral of the temporally constant mass $\omega(\mu)$ of μ and $\operatorname{ess\,var}(\mu)$ is its essential variation (for convenience, we normalize the time interval and add regularization parameters to both terms). We show that this regularizer enforces a sparsity property similar to those listed above. In particular, we prove the existence and representation of a sparse minimizer for the corresponding inverse problem under mild assumptions. We also provide numerical experiments based on an adaption of the algorithm proposed in [18] (employing our sparse characterization). We use a discretization procedure (via temporal deblurring) which suits to our class of admissible paths. Our numerical experiments demonstrate that sparse or diffuse ground truths with jumps can be accurately reconstructed.

We mention that in [28] the authors study a regularizer defined on normal 1-currents in some nonempty, bounded, and open subset of $\mathbb{R}^n$. The regularizer is given as the sum of mass (which is equal to $\|\cdot\|_{\mathscr{M}}$) and boundary mass. In comparison, this regularizer acts on (static) 1-currents and the penalty on the boundary may be seen as a quantization of the number of jumps.

In general, the main goal in the modeling of dynamic inverse problems is to incorporate knowledge about the behaviour of a dynamic source which is to be reconstructed from possibly noisy measurements. A crucial challenge is to *correlate* the observations given at different time instances. The choice of an appropriate regularizer is often a non-trivial task considering, inter alia, the available data, measurement procedure, or physical laws. For example, in [46, 47] the authors require 'temporal smoothness' to model observed displacements in X-ray computed tomography or current density reconstruction.

In recent years, temporal regularization using optimal transport theory has gained a lot of interest. The underlying assumption is that the dynamics under consideration follow certain physical principles. Examples of related inverse problems can be found in positron emission tomography (PET) [48, 22], magnetic resonance imaging (MRI) [19], single-particle tracking [16, 23], and particle image velocimetry (PIV) [45]. These works use the so-called Benamou–Brenier formulation of the Wasserstein-2 distance $W_2(\rho^+,\rho^-)$ between $\rho^+,\rho^- \in \mathscr{P}(\Omega)$ [6],

$$W_2(\rho^+,\rho^-)^2 = \inf_{\rho,v} T\int_{[0,T]}\int_\Omega |v_t|^2 \mathrm{d}\rho_t \mathrm{d}t,$$

where the infimum is over sufficiently regular time-dependent mass distributions $\rho = \rho_t$ and velocity fields $v = v_t$ on Ω which satisfy the continuity equation $\partial_t\rho + \mathrm{div}(\rho v) = 0$ and initial respectively final condition $\rho_0 = \rho^+$ and $\rho_T = \rho^-$ (otherwise, the squared Wasserstein-2 distance can as well be written as an infimum over transport plans π as above, with $|x-y|$ replaced by $|x-y|^2$). Note that $\frac{1}{2T}W_2(\rho^+,\rho^-)^2$ can be interpreted as an infimization of the time integral over the (total) kinetic energy at time t. Inspired by this result, a dynamic regularizer has been devised, for example in the form (with $T = 1$)

$$R(p,\mu) = \|\mu\|_{\mathscr{M}} + \int_{(0,1)\times\Omega}\left|\frac{\mathrm{d}p}{\mathrm{d}\mu}(t,x)\right|^2 \mathrm{d}\mu(t,x),$$

where μ is a nonnegative and p (representing the physical momentum) a vector-valued Radon measure on $(0,1)\times\Omega$ with $p \ll \mu$, subject to the continuity equation (which, by definition of momentum, becomes $\partial_t\mu + \mathrm{div}(p) = 0$), cf. [19, 16]. Note that one may formally replace the second term in R by the time integral of the metric derivative of an absolutely continuous curve [44, Thm. 5.27]

$$t \mapsto \mu_t \in (\omega(\mu)\mathscr{P}(\mathbb{R}^n), W_2).$$

From the point of view of sparse optimization, it has also been shown that such dynamic regularization enforces sparsity: under reasonable assumptions (in particular, finite dimensional data space), there always exists a reconstruction of the form (after disintegration)

$$t \mapsto \mu_t^{opt} = \sum_{i=1}^N \lambda^i \delta_{\gamma_t^i}$$

with $\lambda^i \in \mathbb{R}_+^*$ and absolutely continuous curves $\gamma^i : [0,1] \to \Omega$ [17]. This characterization, which can also be extended to the unbalanced case [15], is a consequence of [50], see [4, Thm. 8.2.1]. Moreover, the sparse structure has been successfully used to design sparse optimization algorithms that directly optimize the curve μ by inserting or deleting curves of type γ and optimizing the weights λ [16, 23]. A feature of the above framework is that the reconstructed curve $t \mapsto \mu_t$ respectively the underlying paths $t \mapsto \gamma_t^i$ are *absolutely continuous*. As a consequence of this regularization bias, these models are only suitable for the tracking and reconstruction of dynamic sources that do *not* show spatial discontinuities across time. Although these regularizations are useful in many applications (as pointed out above), this prior limits the reconstruction of a ground truth that exhibits a dynamic with jumps, which appear, for example, in (stochastic) processes (in particular, càdlàg processes) or object tracking in computer vision, see [5, 29, 25, 10] for some related problems. Jumps may also be produced by defects or not adequately calibrated measurement devices. Conversely, their detection can support the maintenance or recalibration. In this paper, we address this issue by building a mathematical framework that allows for the tracking of temporally discontinuous sources using the regularizer $\mu \mapsto T\omega(\mu) + \mathrm{ess\,var}(\mu)$ applied to a *BV curve* $t \mapsto \mu_t \in (\omega(\mu)\mathscr{P}(\Omega), W_1)$. Note that there also exists a dynamic formulation of the Wasserstein-1 distance [2].

Beside our choice of a regularizer, we mention that there are generalizations of the Wasserstein-1 distance whose use in the regularization of inverse problems might also be interesting to further investigate. In [49] the authors study a Wasserstein-1-type model in which the mass can vary over time, in [37, 36] the authors introduce a multi-material transport problem (it also admits a temporally dynamic formulation [30]) which, in the single-material case, can be interpreted as Wasserstein-1 transport, and in [12, 21, 33, 31] the authors investigate the Wasserstein-1 distance with respect to the so-called (generalized) urban metric (for which the Euclidean distance is a special case).

The paper is organized as follows. In section 1.1 we set up our dynamic inverse problem with the regularizer from above. Our main results are summarized in section 1.2. Section 2 is devoted to some preliminaries for the proofs and numerical part. In section 3 we show our main result, the existence of a sparse minimizer (sections 3.1 and 3.2). We close it with a discussion on a generalization (section 3.3). In section 4 we describe the algorithm (section 4.1), which we use in our numerical experiments, and specify its discretization and implementation details (sections 4.2 and 4.3). Finally, in section 4.4 we display numerical results.

1.1 Inverse tracking problem for BV curves

Let $\Omega \subset \mathbb{R}^n$ be non-empty, compact, and convex.

Definition 1.1 (Admissible BV curves, weight functional)**.** *We define the set of* ***admissible BV curves*** *(not necessarily normalized) as*

$$\mathscr{A} = \mathbb{R}_+ BV([0,1];\mathscr{W}_1(\Omega)),$$

where $\mathscr{W}_1(\Omega) = (\mathscr{P}(\Omega), W_1)$ denotes the Wasserstein-1 space of probability measures equipped with the Wasserstein-1 distance. If $\mu \in \mathscr{A}$, then we write $\omega(\mu)$ for the ***weight*** *of μ, i.e. we have $\mu = \omega(\mu)\rho$ for some $\rho \in BV([0,1];\mathscr{W}_1(\Omega))$.*

An example of a shortest càdlàg representative of $\rho \in BV([0,1];\mathscr{W}_1(\Omega))$ matching some given data at different time instances is given in figure 1.

Remark 1.2 (Structure of $\mathscr{A}$)**.** *The set of admissible BV curves $\mathscr{A}$ is a convex cone in the vector space $L^2_w([0,1];\mathscr{M}(\Omega))$ (see section 2.1).*

Definition 1.3 (Regularizer for BV curves)**.** *We define ($\alpha, \beta \in \mathbb{R}^*_+$ fixed)*

$$\mathscr{R}_{\alpha,\beta}(\mu) = \alpha\omega(\mu) + \beta \operatorname{ess\,var}(\mu),$$

where $\operatorname{ess\,var}(\mu)$ denotes the essential variation of $\mu \in \mathscr{A}$ (see section 2.2).

Problem 1.4 (BV curve tracking problem)**.** *A* ***BV curve tracking problem*** *is any inverse problem of the form*

$$\inf_{\mu \in \mathscr{A}} \mathscr{F}(K(\mu)) + \mathscr{R}_{\alpha,\beta}(\mu),$$

where ***forward operator*** *K maps each admissible BV curve to an element of some* ***data space*** *Y and $\mathscr{F}$ is a* ***fidelity****. In particular, we make the following assumptions (which ensure existence, cf. proposition 1.5):*

- *Y is a Hilbert space,*
- *$\mathscr{F} : Y \to \mathbb{R} \cup \{\infty\}$ is bounded from below, proper, and weakly lower semi-continuous,*

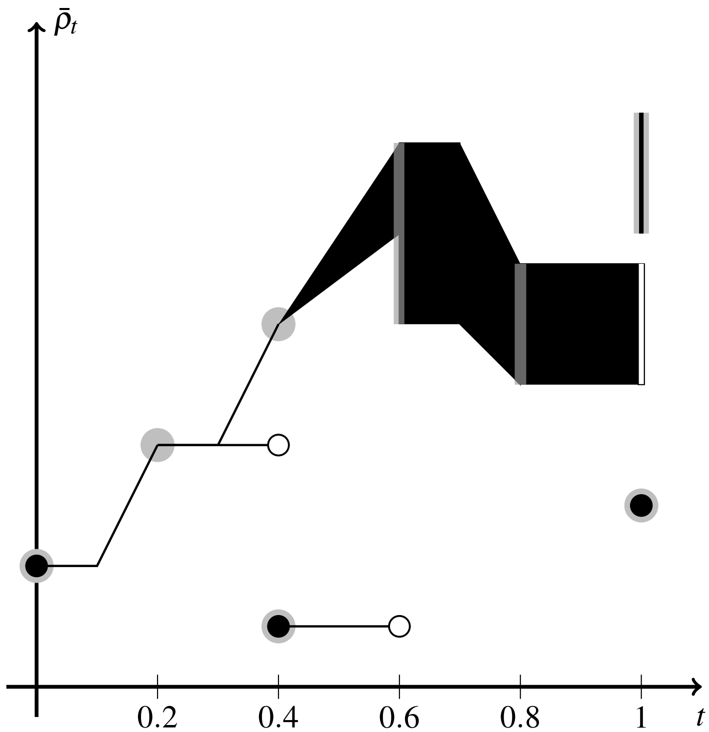


Figure 1: Example of a shortest càdlàg curve $\bar\rho : [0,1] \to \mathscr{W}_1(\Omega)$ matching given data (displayed in gray) at time points $t = 0, 0.2, \ldots, 1$, where $\Omega \subset \mathbb{R}$. In the one-dimensional case, Wasserstein-1 transport is well-understood (by the linearity of the transportation cost combined with the fact that each particle can only move in two directions), see [44, Ch. 2]. For example, we have $\bar\rho_t = \delta_{x_t}$ for $t \in [0, 0.3]$, $\bar\rho_t = \frac{1}{2}\delta_{x_t} + \frac{1}{2}\delta_y$ for $t \in [0.3, 0.4)$, and $\bar\rho_t = (2\mathscr{L}(\ell_t))^{-1}\mathscr{L}\llcorner\ell_t + \frac{1}{2}\delta_z$ for $t \in (0.4, 0.6)$. Here $\mathscr{L}\llcorner\ell_t$ denotes the restriction of one-dimensional Lebesgue measure $\mathscr{L} = \mathscr{L}^1$ to line segment ℓ_t (which would be replaced by the one-dimensional Hausdorff measure if $n > 1$). Note that the jumps at $t = 0.4$ and $t = 0.6$ are càdlàg (in particular, the right-hand limits correspond to the given data). At $t = 1$ the curve $\bar\rho$ jumps to data consisting of a line segment and a Dirac mass. In this example, each particle (with infinitesimal small mass) moves linearly with respect to t (which does not need to be satisfied), waits, or jumps.

- *$K : \mathscr{A} \to Y$ is continuous in the following sense: if $(\mu^j) \subset \mathscr{A}$ and $\mu \in \mathscr{A}$ with $\mu_t^j \overset{*}{\rightharpoonup} \mu_t$ for a.e. $t \in [0,1]$ (in short, we have $\mu^j \overset{*}{\rightharpoonup} \mu$ a.e.), then $K(\mu^j) \rightharpoonup K(\mu)$ (weakly) in Y.*

The existence of a minimizer follows readily from the direct method in the calculus of variations and [3, Thm. 2.4 (i)] (see beginning of section 3).

Proposition 1.5 (Existence)**.** *Problem 1.4 admits a solution.*

Remark 1.6 (Stability)**.** *A natural choice (which for $Y = \mathbb{R}^m$ corresponds to least square fitting) for $\mathscr{F}$ is $\mathscr{F}_f(y) = \frac{1}{2}\|y - f\|_Y^2$, where $f \in Y$ is some given reference data. In this setting, which we will use in the numerical part (section 4), the following stability property is satisfied: Let $(f^j) \subset Y$ with $f^j \to \hat f$ (strongly) in Y. Further, assume that each μ^j is a solution of problem 1.4 with $\mathscr{F} = \mathscr{F}_{f^j}$. Then, up to a subsequence, we have $\mu^j \overset{*}{\rightharpoonup} \hat\mu$ a.e. for some $\hat\mu \in \mathscr{A}$ and $\hat\mu$ is a solution of problem 1.4 with $\mathscr{F} = \mathscr{F}_{\hat f}$.*

Example 1.7 (*K* with values in infinite-dimensional Hilbert space)**.** *Let $\Phi \in C([0,1]\times\Omega^2)$. For any $\mu \in L^2_w([0,1],\mathscr{M}(\Omega))$ define $\widehat{K}\mu : \Omega \to \mathbb{R}$ by*

$$(\widehat{K}\mu)(x) = \int_{[0,1]}\int_\Omega \Phi_t(x,y)\mathrm{d}\mu_t(y)\mathrm{d}\mathscr{L}(t).$$

Let $(x_i) \subset \Omega$ with $x_i \to x \in \Omega$. For each i we have $\int_\Omega \Phi_t(x_i,y)\mathrm{d}\mu_t(y) \le \max\Phi\|\mu_t\|_{\mathscr{M}}$ for a.e. $t\in[0,1]$. Further, it holds $t \mapsto \|\mu_t\|_{\mathscr{M}} \in L^1([0,1])$ by definition of $L^2_w([0,1],\mathscr{M}(\Omega))$. Thus, twofold application of Lebesgue's dominated convergence theorem implies $(\widehat{K}\mu)(x_i) \to (\widehat{K}\mu)(x)$, hence $\widehat{K}\mu \in C(\Omega)$. In particular, we can interpret $\widehat{K}$ as a linear map $L^2_w([0,1],\mathscr{M}(\Omega)) \to L^2(\Omega)$. Now define $K = \widehat{K}|_{\mathscr{A}}$. Take a sequence $(\mu^j) \subset \mathscr{A}$ such that $\mu^j \overset{*}{\rightharpoonup} \mu$ a.e. Write $\mu^j = \omega(\mu^j)\rho^j$ and $\mu = \omega(\mu)\rho$ as in definition 1.1. Choosing the test function $\phi \equiv 1$ we obtain $\omega(\mu^j) \to \omega(\mu)$, in particular $\omega(\mu^j)$ is bounded. Using this, $\mu^j \overset{*}{\rightharpoonup} \mu$ a.e., and dominated convergence again, we get

$$\int_\Omega |K\mu^j - K\mu|^2 \mathrm{d}\mathscr{L}^n = \int_\Omega \left| \int_{[0,1]}\int_\Omega \Phi\mathrm{d}(\mu^j-\mu)\mathrm{d}\mathscr{L}\right|^2 \mathrm{d}\mathscr{L}^n \to 0,$$

that is $K\mu^j \to K\mu$ (strongly) in $L^2(\Omega)$ so that K satisfies the assumption in problem 1.4. Thanks to the regularity of Φ, we can simply apply the Arzelà–Ascoli theorem to show that $\widehat{K}$ is a compact operator implying that the inverse problem $\widehat{K}\mu = f$ with data $f \in C(\Omega)$ is ill-posed in the sense of Hadamard [39, Thm. 3.3.2]. Therefore, problem 1.4 can be seen as a sparse regularization of it. One may interpret $\widehat{K}\mu = f$ as a variant of the Fredholm integral equation of first kind (a classical ill-posed linear inverse problem) — the regularity of Φ is needed to ensure that $\widehat{K}$ is well-defined (typically, one requires lower L^2-regularity of the kernel). Indeed, the (constant) curve $\mu \equiv g\mathscr{L}^n$ with $g \in L^2(\Omega)$ is an element of $L^2_w([0,1],\mathscr{M}(\Omega))$. Taking Φ independent of t we have $(\widehat{K}\mu)(x) = \int_\Omega \Phi(x,y)g(y)\mathscr{L}^n(y)$.

Note that forward operator K cannot implement pointwise evaluation of $\mu \in L^2_w([0,1],\mathscr{M}(\Omega))$. This aspect is also addressed in the discretization in example 1.11.

1.2 Main results

First, we use the superposition principle in [1] to characterize the extremal points of sublevel sets of the regularizer $\mathscr{R}_{\alpha,\beta}$. Define

$$L^-_c(\mathscr{R}_{\alpha,\beta}) = \left\{\mu \in \mathscr{A} \,\middle|\, \mathscr{R}_{\alpha,\beta}(\mu) \le c\right\}$$

for any $c \in \mathbb{R}_+$.

Proposition 1.8 (Extremal points of $L^-_c(\mathscr{R}_{\alpha,\beta})$)**.** *For every $c \in \mathbb{R}_+$ we have*

$$\operatorname{ext}(L^-_c(\mathscr{R}_{\alpha,\beta})) = \{\mu \in \mathscr{A} \mid \mathscr{R}_{\alpha,\beta}(\mu) \in \{0,c\},\, \bar{\mu}_t = \omega(\mu)\delta_{x_t} \text{ for all } t \in [0,1)\},$$

where $\bar{\mu}$ is any càdlàg representative of μ.

We will then use proposition 1.8 and the representer theorem in [9] to prove the existence of a sparse minimizer and its characterization. To this end, we require the following.

Assumption 1.9 (Convex setup and finiteness of data)**.** *The fidelity $\mathscr{F}$ is convex. Further, the forward operator K is a linear map from $\mathscr{A}$ to a finite-dimensional data space $Y = \mathbb{R}^m$ which can be linearly extended to $L^2_w([0,1];\mathscr{M}(\Omega))$.*

Theorem 1.10 (Existence of sparse solution)**.** *Let assumption 1.9 be satisfied. Then there exists a solution $\mu^{opt} \in \mathscr{A}$ of problem 1.4 with*

$$\mu_t^{opt} = \sum_{i=1}^{N} \lambda^i \delta_{x_t^i}$$

for a.e. $t \in [0,1]$, where $N \leq m = \dim(Y)$, $\lambda^i \in \mathbb{R}_+$, and $t \mapsto x_t^i \in BV([0,1];\Omega)$ for $i = 1,\ldots,N$.

As a variant, one may fix N and minimize over weights $\lambda^1,\ldots,\lambda^N \in \mathbb{R}_+^*$ and curves $x^1,\ldots,x^N \in BV([0,1];\Omega)$. This optimization may be seen as regression by BV curves with small length and importance λ^i. The data may be represented by a (static) point cloud in Ω or recorded over time, for example, it may be given by finitely many maps $X^1,\ldots,X^{N_S} : [0,1] \to \Omega$, where each X^i represents a sample path of a càdlàg process $\{X_t \,|\, t \in [0,1]\}$, that is almost surely every path $X : [0,1] \to \Omega$ is càdlàg [40, p. 4]. Note that the sparse solution in theorem 1.10 does not contain any diffusive behaviour which, in principle, is admissible in Wasserstein-1 transport, for example when a line segment is irrigated by a point source (see figure 1).

Example 1.11 (K with values in finite-dimensional Hilbert space)**.** *Recall the definition of forward operator $\widehat{K} : L_w^2([0,1];\mathscr{M}(\Omega)) \to L^2(\Omega)$ from example 1.7. Let $x_1,\ldots,x_L \in \Omega$. Instead of considering $\Phi \in C([0,1]\times\Omega^2)$ one may locate $\Phi^i \in C([0,1]\times\Omega)$ around x_i for all $i = 1,\ldots,L$. For example, each Φ^i may represent the sensitivity (which can be different for different time points t) of a sensor placed at x_i convolved with some point spread function, that is $\Phi_t^i = \theta_t^i * \varphi_t^i$, where θ_t^i is the sensitivity (not necessarily continuous) and φ_t^i models a physical spreading process (for example, spreading of light with φ independent of t and i or spreading through diffusive material properties with φ^i depending on i), see [53, Section 3.1]. Then $\int_{[0,1]}\int_\Omega \Phi_t^i(y)\mathrm{d}\mu_t(y)\mathrm{d}\mathscr{L}(t)$ represents the (total) information collected by the sensor placed at x_i. In applications, this information is obtained by temporal sampling and one is typically interested in the information collected at specific time instances instead. We may include a temporal blurring which naturally circumvents the fact that $\widehat{K}\mu$ cannot realize pointwise evaluation of μ: Let $t_0 < t_1 < \ldots < t_M$ be given time points in $[0,1]$. For each time t_j we may consider*

$$\int_{[0,1]} \phi_j(t) \int_\Omega \Phi_t^i(y)\mathrm{d}\mu_t(y)\mathrm{d}\mathscr{L}(t),$$

where $\phi_j \in C([0,1])$ models the temporal blur at t_j. In summary, we have $\Psi_j^i = \phi_j\Phi^i \in C([0,1]\times\Omega)$. Now define linear $\widetilde{K} : L_w^2([0,1];\mathscr{M}(\Omega)) \to \mathbb{R}^{L\times(M+1)}$ by

$$(\widetilde{K}\mu)_j^i = \int_{[0,1]}\int_\Omega \Psi_j^i \mathrm{d}\mu\mathrm{d}\mathscr{L}$$

for each $i = 1,\ldots,L$ and $j = 0,1,\ldots,M$. As in example 1.7, it is not difficult to see that $\widetilde{K}$ satisfies the continuity property in problem 1.4 so that we can define K as the restriction of $\widetilde{K}$ to $\mathscr{A}$. The non-locality in time may be regulated by ϕ_j given as the density of a truncated normal distribution (for a prioritization of time points close to t_j regulated by the variance) or an indicator function $1_{[0,1]\cap[t_j-\delta_j,t_j+\delta_j]}$ (for a uniform weighting of information close to t_j, which still yields a well-defined operator K). The physical spreading described by φ_t^i may as well be given by a Gaussian filter (which is typical for the spreading of light).

2 Preliminaries

In this section, we introduce our notation and recapitulate the notions of BV curves and their càdlàg representatives.

2.1 Basic notation

We use the following standard notation.

- We write $\mathbb{R}_+ = [0,\infty)$ and $\mathbb{R}_+^* = (0,\infty)$. Further, we set $I = [0,1]$ for the *unit interval*. We predominantly use this notation if $t \in I$ represents the time variable.
- $\Omega \subset \mathbb{R}^n$ denotes some non-empty, compact, and convex (spatial) *domain*.
- If $F : X \to \mathbb{R} \cup \{\infty\}$ and $c \in \mathbb{R}$, then we write $L_c^-(F)$ for the *sublevel set* $\{x \in X \mid F(x) \leq c\}$.
- We write $e_t(\gamma) \in \Omega$ for the *evaluation* of a curve $\gamma : I \to \Omega$ at $t \in I$.
- If X is a topological space, then we write $\mathscr{B}(X)$ for its *Borel σ-algebra*. The *restriction* $\mu \llcorner B$ of a map $\mu : \mathscr{B}(X) \to \mathbb{R}$ (e.g. Radon measure) to $B \in \mathscr{B}(X)$ is defined by

$$(\mu \llcorner B)(\tilde{B}) = \mu(B \cap \tilde{B})$$

  for all $\tilde{B} \in \mathscr{B}(X)$.
- The *pushforward* $f_\# \mu$ of a measure μ on X_1 under a measurable map $f : X_1 \to X_2$ is the measure defined by $f_\# \mu(M) = \mu(f^{-1}(M))$ for all measurable subsets $M \subset X_2$.
- We denote the one-dimensional *Lebesgue measure* by $\mathscr{L}$.
- Let $A \subset \mathbb{R}$ be Lebesgue measurable. For $1 \leq p \leq \infty$ we define $L^p(A)$ as the *Lebesgue space* of equivalence classes of Lebesgue measurable functions $f : A \to \mathbb{R}$ with $\int_A |f|^p \mathrm{d}\mathscr{L} < \infty$ if $p < \infty$, $\operatorname{ess\,sup}_A |f| < \infty$ if $p = \infty$. Further, if (X,d) is a metric space, then we let $L^p(A;(X,d))$ denote the set of equivalence classes of Lebesgue measurable maps $f : A \to (X,d)$ for which $d(f,x) \in L^p(A)$ for some (and thus all if A is bounded) $x \in X$. In each of the two definitions (of $L^p(A)$ and $L^p(A;(X,d))$) two maps belong to the same class if they coincide a.e. in A.
- We write $\mathscr{P}(\Omega)$ for the space of *probability measures* on Ω,

$$\mathscr{P}(\Omega) = \{\rho : \mathscr{B}(\Omega) \to [0,1] \mid \rho \ \sigma\text{-additive with } \rho(\Omega) = 1\}.$$

- The *Wasserstein-1 space* is defined by $\mathscr{W}_1(\Omega) = (\mathscr{P}(\Omega), W_1)$. It is equipped with the *Wasserstein-1 distance* W_1 given by

$$W_1(\rho^+,\rho^-) = \inf_\pi \int_{\Omega\times\Omega} |x-y| \mathrm{d}\pi(x,y),$$

  where the infimum is over *transport plans* between ρ^+ and ρ^-, that is $\pi \in \mathscr{P}(\Omega \times \Omega)$ with $\pi(B \times \Omega) = \rho^+(B)$ and $\pi(\Omega \times B) = \rho^-(B)$ for all $B \in \mathscr{B}(\Omega)$. The space $\mathscr{W}_1(\Omega)$ is complete and separable since Ω is complete and separable [54, Thm. 6.18].

- The space of *Radon measures* on Ω (which by the Riesz representation theorem can be identified with the topological dual of the continuous functions $C(\Omega)$ with uniform norm $\|\cdot\|_\infty$) is given by

$$\mathscr{M}(\Omega) = \{\mu : \mathscr{B}(\Omega) \to \mathbb{R} \,|\, \mu \text{ } \sigma\text{-additive}\}.$$

The corresponding operator norm is called *total variation* and we denote it by

$$\|\mu\|_{\mathscr{M}} = \sup_{\|\phi\|_\infty \le 1} \int_\Omega \phi \,\mathrm{d}\mu.$$

We indicate the weak-$*$ convergence in $\mathscr{M}(\Omega)$ by $\overset{*}{\rightharpoonup}$, that is $\mu^i \overset{*}{\rightharpoonup} \mu$ in $\mathscr{M}(\Omega)$ if and only if $\langle \phi, \mu^i \rangle = \int \phi \,\mathrm{d}\mu^i \to \int \phi \,\mathrm{d}\mu = \langle \phi, \mu \rangle$ for all $\phi \in C(\Omega)$.

- Let $A \subset \mathbb{R}$ be closed. Along with the above duality $\mathscr{M}(\Omega) = C(\Omega)^*$, we define $L^2_w(A; C(\Omega))$ as the space of equivalence classes of Lebesgue measurable maps $f : A \to C(\Omega)$ with $\int_A \|f\|^2_\infty \mathrm{d}\mathscr{L} < \infty$. The space $L^2_w(A; C(\Omega))$ with norm

$$\|\phi\|_{L^2_w} = \sqrt{\int_A \|\phi\|^2_\infty \mathrm{d}\mathscr{L}}$$

is a separable Banach space (see [24, Section 8.18.1] or [55, Thm. I.5.18]). Its (topological) dual is given by $L^2_w(A; \mathscr{M}(\Omega))$ [24, Thm. 8.20.3], the space of equivalence classes of weakly measurable maps $f : A \to \mathscr{M}(\Omega)$[1] with $\int_A \|f\|^2_{\mathscr{M}} \mathrm{d}\mathscr{L} < \infty$. The dual norm becomes

$$\|\mu\|_{L^2_w} = \sqrt{\int_A \|\mu\|^2_{\mathscr{M}} \mathrm{d}\mathscr{L}}$$

and the dual pairing is given by $\langle \phi, \mu \rangle_{L^2_w} = \int_A \langle \phi, \mu \rangle \mathrm{d}\mathscr{L}$. In each of the two definitions (of $L^2_w(A; C(\Omega))$ and $L^2_w(A; \mathscr{M}(\Omega))$) two maps belong to the same class if they coincide a.e. in A. (The subscript in $L^2_w(A; C(\Omega))$ is justified by the fact that (Lebesgue) measurability and weak measurability coincide for $C(\Omega)$-valued maps, which follows from the Pettis measurability theorem and the separability of $C(\Omega)$.)

- Finally, we recall some terminology from convex analysis (see [26, p. 1] and [41, Section 8]). Let V be a real vector space and $C \subset V$ convex. An *extremal point* (or extreme point) of C is a point $v \in C$ such that $C \setminus \{v\}$ is convex (this notion goes back to Minkowski). We write $\mathrm{ext}(C)$ for the set of extremal points of C. We say that C is *linearly closed* if the intersection of C with every one-dimensional affine space of V is closed. A *ray* of V is any set of the form $\{v + tw \,|\, t > 0\}$ for $v, w \in V$ and $w \neq 0$. The *lineality space* of C is defined as $\mathrm{lin}(C) := \mathrm{rec}(C) \cap (-\mathrm{rec}(C))$, where $\mathrm{rec}(C) = \{v \in V \,|\, C + \mathbb{R}^*_+ v \subset C\}$ denotes the *recession cone* of C.

2.2 BV curves and càdlàg representatives

In this section, we recall the notions of BV curves and their càdlàg representatives in a metric space (X, d) (following [1]). For a more comprehensive theory of BV curves in metric spaces the reader may consult [3]. In this article, we are particularly interested in $(X, d) = \mathscr{W}_1(\Omega)$.

[1] By [24, Sections 1.11.1 & 1.11.3] a map $f : I \to \mathscr{M}(\Omega)$ is weakly measurable if it is Lebesgue measurable with respect to Borel σ-algebra on $\mathscr{M}(\Omega)$ induced by the subspace topology of $\mathscr{M}(\Omega) = C(\Omega)^* \subset C(\Omega)'$. The algebraic dual $C(\Omega)'$ is equipped with the weak topology, i.e. the weakest topology for which each functional $f \mapsto \ell(f)$ is continuous ($\ell \in C(\Omega)'$).

Definition 2.1 (Variation)**.** *The **variation** of a map $\mu : I \to (X,d)$ on $J \subset I$ is defined by*

$$\mathrm{var}(\mu, J) = \sup\left\{ \sum_{i=1}^{N} d(\mu_{t_i}, \mu_{t_{i-1}}) \,\middle|\, t_0 < t_1 < \ldots < t_N,\, t_i \in J \right\}.$$

We abbreviate $\mathrm{var}(\mu) = \mathrm{var}(\mu, I)$.

Definition 2.2 (Essential variation, BV curves)**.** *The **essential variation** of $\mu \in L^1(I;(X,d))$ on $J \subset I$ is given by*

$$\mathrm{ess\,var}(\mu, J) = \inf\{\mathrm{var}(\tilde{\mu}, J) \mid \mu = \tilde{\mu} \text{ a.e.}\}.$$

We write $\mathrm{ess\,var}(\mu) = \mathrm{ess\,var}(\mu, I)$*. If* $\mathrm{ess\,var}(\mu) < \infty$*, then we call μ a **BV curve** in (X,d). We write $BV(I;(X,d))$ for the space of BV curves in (X,d).*

Remark 2.3 (BV Wasserstein curves)**.** *We have $\mathbb{R}_+ BV(I;\mathscr{W}_p(\Omega)) \subset \mathbb{R}_+ BV(I;\mathscr{W}_1(\Omega)) = \mathscr{A}$ for all $p \in [1,\infty)$, where $\mathscr{W}_p(\Omega) = (\mathscr{P}(\Omega), W_p)$ denotes the Wasserstein-p space with $W_p(\rho^+, \rho^-)^p = \inf_\pi \int |x - y|^p \mathrm{d}\pi(x,y)$ (infimum over transport plans π between ρ^+ and ρ^-). This follows easily from the estimate (invoking Hölder's inequality)*

$$W_1(\rho^+, \rho^-) \le W_p(\rho^+, \rho^-) \le \mathrm{diam}(\Omega)^{\frac{p-1}{p}} W_1(\rho^+, \rho^-)^{\frac{1}{p}}.$$

Hence, our set $\mathscr{A}$ of decision variables contains all the (equivalence classes of) classical Wasserstein curves with arbitrary weight and bounded (essential) variation.

The following lemma is a version of a standard result. We briefly recall the argument since we did not find it for our setting.

Lemma 2.4 (L^p regularity of BV curves)**.** *We have $BV(I;(X,d)) \subset L^p(I;(X,d))$ for every $1 \le p \le \infty$.*

Proof. Let $\mu \in BV(I;(X,d))$ and $x \in X$ with $d(\mu, x) \in L^1(I)$. Pick any representative $\tilde{\mu}$ of μ with $\mathrm{var}(\tilde{\mu}) < \infty$. Then $d(\mu, x) \le d(\tilde{\mu}, \tilde{\mu}_0) + d(\tilde{\mu}_0, x) \le \mathrm{var}(\tilde{\mu}) + d(\tilde{\mu}_0, x)$ a.e. Thus, for $p < \infty$ we have (using the convexity of $t \to t^p$ and the boundedness of $I = [0,1]$) $\int_I d(\mu,x)^p \mathrm{d}\mathscr{L} \le 2^{p-1}(\mathrm{var}(\tilde{\mu})^p + d(\tilde{\mu}_0, x)^p) < \infty$, and for $p = \infty$ we get $\mathrm{ess\,sup}_I\, d(\mu, x) \le \sup_I d(\tilde{\mu}, x) \le \mathrm{var}(\tilde{\mu}) + d(\tilde{\mu}_0, x) < \infty$. □

As in [1], we use [51, Ch. 6] to generalize the definition of variation measure to the metric space setting.

Definition 2.5 (Variation measure)**.** *Let $\mu \in BV(I;(X,d))$. Define a non-decreasing function $V : I \to \mathbb{R}$ by $V(t) = \mathrm{ess\,var}(\mu, (0,t))$ for all $t \in I$. The **variation measure** $|D\mu| : \mathscr{B}(I) \to \mathbb{R}_+$ of μ is defined as the Lebesgue–Stieltjes measure [51, Section 6.3.3, Thm. 3.5] which satisfies*

$$|D\mu|((a,b]) = V(b) - V(a)$$

for all $(a,b] \subset I$ with $a < b$.

We recall the definition of càdlàg curves. Every $\mu \in BV(I;(X,d))$ can be represented by such a curve $\bar{\mu} : I \to (X,d)$. Further, a càdlàg representative $\bar{\mu}$ is uniquely determined on $[0,1)$.

Definition 2.6 (Càdlàg curves in (X,d), spaces $\mathscr{D}_E(\Omega), \mathscr{D}_W(\Omega)$)**.** *A map $\mu : I \to (X,d)$ is called a* ***càdlàg curve*** *if μ is right-continuous and the left limit $\lim_{\tilde{t} \nearrow t} \mu_{\tilde{t}}$ exists for every $t \in I$ (continue à droite, limite à gauche). We write $\mathscr{D}_E(\Omega)$ for the set of càdlàg curves $\gamma : I \to \Omega$ (with d the Euclidean distance) and $\mathscr{D}_W(\Omega)$ for the càdlàg curves $\mu : I \to \mathscr{W}_1(\Omega)$ (with $d = W_1$ the Wasserstein-1 distance).*

Remark 2.7 (Skorohod metric)**.** *If (X,d) is complete and separable, then the space of càdlàg curves $\mathscr{D} = \{\gamma : I \to (X,d) \mid \gamma \text{ càdlàg}\}$ equipped with the* ***Skorohod*** *J_1* ***metric*** *$d_{\mathscr{D}}$ is complete and separable [8, Thm. 12.2 & p. 131] (cf. text passage below [1, Thm. 2.10]). Thus, this property is statisfied for both $\mathscr{D}_E(\Omega)$ and $\mathscr{D}_W(\Omega)$. We recall the definition of $d_{\mathscr{D}}$ (see construction in [8, p. 123-126]). Let Λ be the set of strictly increasing homeomorphisms $\lambda : I \to I$. A complete metric is obtained by requiring that λ is close to the identity in the sense that*

$$f(\lambda) = \sup_{\substack{s,t \in I \\ s<t}} \left| \log \frac{\lambda(t) - \lambda(s)}{t - s} \right|$$

is small. The Skorohod J_1 metric is then given by

$$d_{\mathscr{D}}(\gamma_1, \gamma_2) = \inf_{\lambda \in \Lambda} \max\{f(\lambda), \|\gamma_1 - \gamma_2 \circ \lambda\|_\infty\}.$$

Next, we apply the compactness statement [3, Thm. 2.4 (i)] to our setting. Recall the definitions of admissible set $\mathscr{A}$ and regularizer $\mathscr{R}_{\alpha,\beta}$ (definitions 1.1 and 1.3). It is readily established that we can assume $\alpha = \beta = 1$ in the proofs (the arguments for general $\alpha, \beta \in \mathbb{R}_+^*$ are the same).

Notation 2.8 (Regularizer $\mathscr{R}$)**.** *We write $\mathscr{R} = \mathscr{R}_{1,1}$, that is $\mathscr{R}(\mu) = \omega(\mu) + \operatorname{ess\,var}(\mu)$ for every $\mu \in \mathscr{A}$.*

Lemma 2.9 (Compactness in $(\mathscr{A}, \mathscr{R}_{\alpha,\beta})$)**.** *Let $(\mu^j) \subset \mathscr{A}$ with $\sup_j \mathscr{R}_{\alpha,\beta}(\mu^j) < \infty$. Then, up to a subsequence, we have $\mu^j \overset{*}{\rightharpoonup} \mu$ a.e. and $\mathscr{R}_{\alpha,\beta}(\mu) \leq \liminf_j \mathscr{R}_{\alpha,\beta}(\mu^j)$ for some $\mu \in \mathscr{A}$.*

Proof. Since $\mathscr{R}(\mu^j) = \omega(\mu^j) + \operatorname{ess\,var}(\mu^j)$ is uniformly bounded, we get $\omega(\mu^j) \to \omega \in \mathbb{R}_+$ up to a subsequence. Recall that $\mu^j = \omega(\mu^j)\rho^j$ with $\rho^j \in BV(I; \mathscr{W}_1(\Omega))$.

Case $\omega > 0$: If $\omega > 0$, then $\sup_j \operatorname{ess\,var}(\rho^j) = \sup_j \omega(\mu^j)^{-1} \operatorname{ess\,var}(\mu^j)$ must be finite (neglect the finite number of indices for which $\omega(\mu^j) = 0$). Moreover, we clearly have $\sup_j \int_I W_1(\rho^j, \delta_0) \mathrm{d}\mathscr{L} < \infty$ because Ω is bounded. Further, metric space $\mathscr{W}_1(\Omega)$ is separable and every bounded and closed set in $\mathscr{W}_1(\Omega)$ is compact by the Banach–Alaoglu theorem (W_1 metrizes weak-$*$ convergence). Hence, by [3, Thm. 2.4 (i)] there is some $\rho \in BV(I; \mathscr{W}_1(\Omega))$ such that (up to a subsequence) $\lim_j W_1(\rho^j, \rho) = 0$ a.e. and $\operatorname{ess\,var}(\rho) \leq \liminf_j \operatorname{ess\,var}(\rho^j)$. Therefore, we obtain $\rho^j \overset{*}{\rightharpoonup} \rho$ a.e., thus $\mu^j \overset{*}{\rightharpoonup} \mu = \omega\rho$ a.e. In summary, we have

$$\mathscr{R}(\mu) = \omega(\mu)\mathscr{R}(\rho) \leq \omega(\mu) \liminf_j \mathscr{R}(\rho^j) = \liminf_j \omega(\mu^j)\mathscr{R}(\rho^j) = \liminf_j \mathscr{R}(\mu^j).$$

Case $\omega = 0$: If $\omega = 0$, then we directly obtain $\mu^j \overset{*}{\rightharpoonup} 0$ a.e. Noting $\mathscr{R}(0) = 0$ yields the inequality. □

3 Existence of sparse solution for BV curve tracking

In this section, we always take $\alpha = \beta = 1$ to make the proofs more readable (recall notation 2.8). The arguments are the same for general $\alpha, \beta \in \mathbb{R}_+^*$ (as we already noticed in lemma 2.9). First, we prove the existence of a minimizer for problem 1.4 (proposition 1.5) and the stability property in remark 1.6. The first statement is a direct consequence of lemma 2.9.

Proof of proposition 1.5. Recall that $\mathscr{F}$ is bounded from below and proper. Therefore, we can pick a minimizing sequence $(\mu^j) \subset \mathscr{A}$. Since $\mathscr{R}(\mu^j)$ is uniformly bounded, we can apply lemma 2.9: there exists $\mu \in \mathscr{A}$ such that $\mu^j \overset{*}{\rightharpoonup} \mu$ a.e. and $\mathscr{R}(\mu) \leq \liminf_j \mathscr{R}(\mu^j)$ (up to a subsequence). By the regularity of K (it satisfies $K(\mu^j) \rightharpoonup K(\mu)$) and $\mathscr{F}$ (it is weakly lower semicontinuous) we obtain $\mathscr{F}(K(\mu)) \leq \liminf_j \mathscr{F}(K(\mu^j))$. Hence, the BV curve μ is a minimizer of problem 1.4. □

Proof of the stability statement in remark 1.6. The argument is similar to [19, Thm. 4.7]. By the triangle inequality, the optimality of the μ^j, $\mathscr{R}(0) = 0$, and $f^j \to \hat{f}$ we have

$$\begin{aligned}\mathscr{F}_{\hat{f}}(K(\mu^j)) + \mathscr{R}(\mu^j) &= \frac{1}{2}\|K(\mu^j) - \hat{f}\|_Y^2 + \mathscr{R}(\mu^j) \leq \|K(\mu^j) - f^j\|_Y^2 + \mathscr{R}(\mu^j) + \|f^j - \hat{f}\|_Y^2 \\ &\leq 2\left(\frac{1}{2}\|K(\mu^j) - f^j\|_Y^2 + \mathscr{R}(\mu^j)\right) + \|f^j - \hat{f}\|_Y^2 \leq \|K(0) - f^j\|_Y^2 + \|f^j - \hat{f}\|_Y^2 \leq C < \infty\end{aligned}$$

for some constant $C > 0$. Hence, by the proof of proposition 1.5 there exists $\hat{\mu} \in \mathscr{A}$ with $\mu^j \overset{*}{\rightharpoonup} \hat{\mu}$ a.e. (up to a subsequence) and $\mathscr{F}_{\hat{f}}(K(\hat{\mu})) + \mathscr{R}(\hat{\mu}) \leq \liminf_j \mathscr{F}_{\hat{f}}(K(\mu^j)) + \mathscr{R}(\mu^j)$. Further, note that (by taking the lim inf on both sides of the first inequality in the above estimate and using $f^j \to \hat{f}$ again) $\liminf_j \mathscr{F}_{\hat{f}}(K(\mu^j)) + \mathscr{R}(\mu^j) \leq \liminf_j \mathscr{F}_{f^j}(K(\mu^j)) + \mathscr{R}(\mu^j)$. Using this, the optimality of the μ^j, and the definition of $\mathscr{F}_f$ (together with $f^j \to \hat{f}$) we get

$$\mathscr{F}_{\hat{f}}(K(\hat{\mu})) + \mathscr{R}(\hat{\mu}) \leq \liminf_j \mathscr{F}_{f^j}(K(\mu^j)) + \mathscr{R}(\mu^j) \leq \liminf_j \mathscr{F}_{f^j}(K(\mu)) + \mathscr{R}(\mu) = \mathscr{F}_{\hat{f}}(K(\mu)) + \mathscr{R}(\mu)$$

for all $\mu \in \mathscr{A}$, which shows the optimality of $\hat{\mu}$. □

3.1 Characterization of extremal points in $L_c^-(\mathscr{R})$

In this section, we characterize the extremal points of the sublevel sets (proposition 1.8)

$$L_c^-(\mathscr{R}) = \{\mu \in \mathscr{A} \mid \mathscr{R}(\mu) \leq c\}.$$

We will need the following statement.

Lemma 3.1 (Bound for essential variation of pushforward under evaluation map)**.** *Let $\Gamma \in \mathscr{B}(\mathscr{D}_E(\Omega))$ and $\eta \in \mathbb{R}_+\mathscr{P}(\mathscr{D}_E(\Omega) \cap \{\mathrm{var} < \infty\})$ with $\int_{\mathscr{D}_E(\Omega)} |\mathrm{D}\gamma|(I)\mathrm{d}\eta(\gamma) < \infty$. Assume that $\eta(\Gamma) \in \mathbb{R}_+^*$. Then the map $t \mapsto \mu_t = (e_t)_\#(\eta \llcorner \Gamma)$ satisfies*

$$\operatorname{ess\,var}(\mu) \leq \int_\Gamma \mathrm{var}(\gamma)\,\mathrm{d}\eta(\gamma).$$

Proof. First, we remark that $\int_\Gamma \mathrm{var}(\gamma)\,\mathrm{d}\eta(\gamma)$ is well-defined because $\gamma \mapsto \mathrm{var}(\gamma)$ is lower semicontinuous [1, Lem. 2.13][2] (which also implies that $\mathscr{D}_E(\Omega) \cap \{\mathrm{var} < \infty\}$ is Borel measurable). By [1, Thm. 3.1] the integral $\int_{\mathscr{D}_E(\Omega)} |\mathrm{D}\gamma|(I)\mathrm{d}\eta(\gamma)$ is well-defined, $\mu \in \eta(\Gamma)\mathscr{D}_W(\Omega)$, and

$$|\mathrm{D}\mu| \leq \int_\Gamma |\mathrm{D}\gamma| \mathrm{d}\eta(\gamma). \tag{1}$$

[2] It is also finite. Indeed, we have $\mathrm{var}(\gamma) = \mathrm{var}(\gamma;(0,1)) + \gamma_1^- = |\mathrm{D}\gamma|(0,1) + \gamma_1^-$ [1, Lem. 2.5 (1)] and $\gamma_1^- = \lim_{t \nearrow 1} |\gamma(t) - \gamma(1)| \leq \mathrm{diam}(\Omega)$ (γ_1^- exists because γ is càdlàg). Thus $\int_\Gamma \mathrm{var}(\gamma)\mathrm{d}\eta(\gamma) \leq \int_{\mathscr{D}_E(\Omega)} |\mathrm{D}\gamma|(I)\mathrm{d}\eta(\gamma) + \mathrm{diam}(\Omega) < \infty$ by assumption.

Using the definition of ess var and the characterizing properties of càdlàg curves we get

$$\operatorname{ess\,var}(\mu) \leq \operatorname{var}(\mu) = \operatorname{var}(\mu;(0,1)) + \mu_1^-,$$

where we use the abbreviation (for the jump at $t = 1$)

$$\mu_1^- = \lim_{t \nearrow 1} W_1(\mu_t, \mu_1).$$

For each $t \in I$ we define a transport plan (as in [11, Def. 3.4.9]) $\pi_t \in \eta(\Gamma)\mathscr{P}(\Omega \times \Omega)$ by

$$\langle \varphi, \pi_t \rangle = \int_\Gamma \varphi(\gamma(t), \gamma(1)) \mathrm{d}\eta(\gamma)$$

for all $\varphi \in C(\Omega \times \Omega)$ (the integrand $\gamma \mapsto \varphi(\gamma(t), \gamma(1))$ is Borel measurable by [1, Prop. 2.15] for all $\varphi \in C(\Omega \times \Omega)$). Moreover, the transport plans π_t are admissible for $W_1(\mu_t, \mu_1)$ since

$$\pi_t(B \times \Omega) = (\eta \llcorner \Gamma)(e_t^{-1}(B)) = \mu_t(B) \qquad \text{and} \qquad \pi_t(\Omega \times B) = \mu_1(B)$$

for all $B \in \mathscr{B}(\Omega)$. Hence, we obtain

$$\mu_1^- \leq \limsup_{t \nearrow 1} \int_{\Omega \times \Omega} |x - y| \mathrm{d}\pi_t(x, y) = \limsup_{t \nearrow 1} \int_\Gamma |\gamma(t) - \gamma(1)| \mathrm{d}\eta(\gamma),$$

where we used $\varphi(x, y) = |x - y|$ in the equality. Note that (as in footnote 2) $\gamma \mapsto |\gamma(t) - \gamma(1)|$ is bounded (uniformly in $t \in I$) by $\gamma \mapsto \operatorname{diam}(\Omega)$. Thus, using dominated convergence,

$$\limsup_{t \nearrow 1} \int_\Gamma |\gamma(t) - \gamma(1)| \mathrm{d}\eta(\gamma) = \int_\Gamma \gamma_1^- \mathrm{d}\eta(\gamma).$$

In summary, we obtain (invoking [1, Lem. 2.5 (1)] in the first equality and inequality (1) plus the above estimate in the second inequality)

$$\begin{aligned} \operatorname{ess\,var}(\mu) \leq \operatorname{var}(\mu;(0,1)) + \mu_1^- &= |\mathrm{D}\mu|((0,1)) + \mu_1^- \\ &\leq \int_\Gamma |\mathrm{D}\gamma|((0,1)) \mathrm{d}\eta(\gamma) + \int_\Gamma \gamma_1^- \mathrm{d}\eta(\gamma) = \int_\Gamma \operatorname{var}(\gamma) \mathrm{d}\eta(\gamma). \quad \square \end{aligned}$$

Proof of proposition 1.8. Fix $c \in \mathbb{R}_+$. First, we note that zero is always an extremal point of $L_c^-(\mathscr{R})$ because $\mathscr{R}(0) = \mathscr{R}(\lambda\mu + (1-\lambda)\nu)$ with $\lambda \in (0,1), \mu, \nu \in L_c^-(\mathscr{R})$ implies $\mu = \nu = 0$ by $\mu, \nu \geq 0$ and definition of $\mathscr{R}$. We abbreviate

$$\mathscr{E} = \{\mu \in \mathscr{A} \mid \mathscr{R}(\mu) \in \{0, c\}, \bar{\mu}_t = \omega(\mu)\delta_{x_t} \text{ for all } t \in [0,1)\}.$$

<u>$\operatorname{ext}(L_c^-(\mathscr{R})) \subset \mathscr{E}$</u>: Let $\mu \in \operatorname{ext}(L_c^-(\mathscr{R}))$. By the above we can assume $\mu \neq 0$, thus $\mathscr{R}(\mu) = c$. Indeed, if we had $\mathscr{R}(\mu) \in (0, c)$, then $\mu = \lambda \cdot 0 + (1-\lambda)(1-\lambda)^{-1}\mu$ and $\mathscr{R}((1-\lambda)^{-1}\mu) = (1-\lambda)^{-1}\mathscr{R}(\mu) \leq c$ for $\lambda \in (0,1)$ sufficiently small, thus $0, (1-\lambda)^{-1}\mu \in L_c^-(\mathscr{R})$ which would imply that μ is not extremal. Let $\bar{\mu}$ be a càdlàg representative of μ with $\operatorname{var}(\bar{\mu}) = \operatorname{ess\,var}(\mu)$ (which is finite by $\mathscr{R}(\mu) < \infty$). By [1, Thm. 3.3] there exists a measure $\eta \in \omega(\mu)\mathscr{P}(\mathscr{D}_E(\Omega))$ (concentrated on $\mathscr{D}_E(\Omega) \cap \{\operatorname{var} < \infty\}$) with $(e_t)_\# \eta = \bar{\mu}_t$ for all $t \in I$ such that

$$|\mathrm{D}\bar{\mu}| = \int_{\mathscr{D}_E(\Omega)} |\mathrm{D}\gamma| \mathrm{d}\eta(\gamma)$$

as well as (see footnote on [1, p. 18])

$$\operatorname{var}(\bar{\mu}) = \int_{\mathscr{D}_E(\Omega)} \operatorname{var}(\gamma) \mathrm{d}\eta(\gamma). \tag{2}$$

We want to show that $\bar{\mu}_t = \omega(\mu)\delta_{x_t}$ for every $t \in [0,1)$. To this end, we use a similar approach as in the proof of the claim in [17, Thm. 6]. By contradiction, assume that there exist $t_0 \in [0,1)$ and $B_0 \in \mathscr{B}(\Omega)$ with $\bar{\mu}_{t_0}(B_0), \bar{\mu}_{t_0}(\Omega \setminus B_0) \in \mathbb{R}_+^*$. Now set

$$\Gamma_0 = \mathscr{D}_E(\Omega) \cap e_{t_0}^{-1}(B_0).$$

Set Γ_0 is Borel measurable by [1, Prop. 2.15]. Define

$$\lambda_1 = \frac{1}{c}\left(\eta(\Gamma_0) + \int_{\Gamma_0} \operatorname{var}(\gamma)\mathrm{d}\eta(\gamma)\right) \quad \text{and} \quad \lambda_2 = \frac{1}{c}\left(\eta(\mathscr{D}_E(\Omega)\setminus\Gamma_0) + \int_{\mathscr{D}_E(\Omega)\setminus\Gamma_0} \operatorname{var}(\gamma)\mathrm{d}\eta(\gamma)\right).$$

We have $\eta(\Gamma_0) > 0$ since

$$\eta(\Gamma_0) = \eta(e_{t_0}^{-1}(B_0)) = \bar{\mu}_{t_0}(B_0) > 0.$$

Therefore, we get $\lambda_1 > 0$. Similarly, we obtain $\lambda_2 > 0$. Moreover,

$$c(\lambda_1 + \lambda_2) = \eta(\mathscr{D}_E(\Omega)) + \int_{\mathscr{D}_E(\Omega)} \operatorname{var}(\gamma)\mathrm{d}\eta(\gamma) = \omega(\mu) + \operatorname{ess\,var}(\mu) = \mathscr{R}(\mu) = c$$

by equation (2) and the choice of $\bar{\mu}$. Therefore, we get $\lambda_1, \lambda_2 \in (0,1)$ with $\lambda_1 + \lambda_2 = 1$. Now define $\bar{\mu}_t^1, \bar{\mu}_t^2 \in \mathbb{R}_+^* \mathscr{P}(\Omega)$ by

$$\bar{\mu}_t^1 = \lambda_1^{-1}(e_t)_\#(\eta \llcorner \Gamma_0) \qquad \text{and} \qquad \bar{\mu}_t^2 = \lambda_2^{-1}(e_t)_\#(\eta \llcorner (\mathscr{D}_E(\Omega)\setminus\Gamma_0))$$

for all $t \in I$. Note that we directly get

$$\bar{\mu} = \lambda_1 \bar{\mu}^1 + \lambda_2 \bar{\mu}^2.$$

Now we show that $\bar{\mu}^1, \bar{\mu}^2$ are representatives of $\mu^1, \mu^2 \in L_c^-(\mathscr{R})$ with $\mu^1 \neq \mu^2$ (which yields the desired contradiction). First, note that $\mu^1 \in L_c^-(\mathscr{R})$ by (the argument for μ^2 is similar)

$$\mathscr{R}(\mu^1) = \omega(\mu^1) + \operatorname{ess\,var}(\mu^1) = \lambda_1^{-1}(\eta(\Gamma_0) + \operatorname{ess\,var}(t \mapsto (e_t)_\#(\eta \llcorner \Gamma_0))) \leq c,$$

where the inequality follows from the definition of λ_1 and lemma 3.1. Finally, we show $\mu^1 \neq \mu^2$. We have

$$\lambda_1 \bar{\mu}_{t_0}^1(B_0) = (e_{t_0})_\#(\eta \llcorner \Gamma_0)(B_0) = \eta(\Gamma_0) > 0 \quad \text{and} \quad \lambda_2 \bar{\mu}_{t_0}^2(B_0) = (e_{t_0})_\#(\eta \llcorner (\mathscr{D}_E(\Omega)\setminus\Gamma_0))(B_0) = 0,$$

thus $\bar{\mu}_{t_0}^1(B_0) \neq \bar{\mu}_{t_0}^2(B_0)$. We prove that there exists $t_1 \in (t_0, 1)$ with $\bar{\mu}_t^1 \neq \bar{\mu}_t^2$ for all $t \in (t_0, t_1)$. For a contradiction, assume that there is a sequence $(s_i) \subset (t_0, 1)$ with $s_i \searrow t_0$ and $\bar{\mu}_{s_i}^1 = \bar{\mu}_{s_i}^2$ for all i. Since $\bar{\mu}^1, \bar{\mu}^2$ are càdlàg by [1, Thm. 3.1], we obtain

$$0 = \lim_i W_1(\bar{\mu}_{t_0}^1, \bar{\mu}_{s_i}^1) = \lim_i W_1(\bar{\mu}_{t_0}^1, \bar{\mu}_{s_i}^2),$$

thus $\bar{\mu}_{t_0}^1 = \bar{\mu}_{t_0}^2$ (which contradicts $\bar{\mu}_{t_0}^1(B_0) \neq \bar{\mu}_{t_0}^2(B_0)$).

$\underline{\mathrm{ext}(L_c^-(\mathscr{R})) \supset \mathscr{E}}$: Let $\mu \in \mathscr{E}$. The case $\mathscr{R}(\mu) = 0$ is trivial. Hence, we can assume $\mathscr{R}(\mu) = c$. Write $\mu = \lambda\mu^1 + (1-\lambda)\mu^2$ with $\lambda \in (0,1)$ and $\mu^1, \mu^2 \in L_c^-(\mathscr{R})$. We show $\mu^1 = \mu^2 = \mu$. We have

$$c = \mathscr{R}(\mu) \leq \lambda\mathscr{R}(\mu^1) + (1-\lambda)\mathscr{R}(\mu^2) \leq \lambda c + (1-\lambda)c = c.$$

Therefore, we get $\mathscr{R}(\mu^1) = \mathscr{R}(\mu^2) = c$. Write $\mu^1 = \omega(\mu^1)\rho^1$ and $\mu^2 = \omega(\mu^2)\rho^2$ with $\rho^1, \rho^2 \in BV(I, \mathscr{W}_1(\Omega))$. Let $\bar{\mu}, \bar{\rho}^1, \bar{\rho}^2$ be càdlàg representatives of μ, ρ^1, ρ^2. For every $t \in [0,1)$ and every $B \in \mathscr{B}(\Omega)$ we have (using $\bar{\mu}_t = \omega(\mu)\delta_{x_t}$)

$$\lambda\omega(\mu^1)\bar{\rho}_t^1(B) + (1-\lambda)\omega(\mu^2)\bar{\rho}_t^2(B) = \bar{\mu}_t(B) = \begin{cases} \omega(\mu) & \text{if } x_t \in B, \\ 0 & \text{else.} \end{cases}$$

In particular, we obtain

$$\bar{\rho}_t^1(\Omega \setminus \{x_t\}) = \bar{\rho}_t^2(\Omega \setminus \{x_t\}) = 0,$$

thus $\bar{\rho}_t^1 = \bar{\rho}_t^2 = \delta_{x_t}$ for every $t \in [0,1)$. Using $\mathscr{R}(\mu^1) = \mathscr{R}(\mu^2) = c$, we finally conclude $\omega(\mu^1) = \omega(\mu^2) = \omega(\mu)$ as we wanted to prove. □

3.2 Application of a representer theorem

In this section, we apply proposition 1.8 to prove theorem 1.10. We will use the following abbreviations (recall the definitions in section 2.1).

Notation 3.2 (Solution set $\mathscr{S}$, weak-$*$ topology $\mathscr{T}_*$)**.** *We define $\mathscr{S}$ as the set of solutions of problem 1.4 (recall proposition 1.5). Further, we write $\mathscr{T}_*$ for the weak-$*$ topology on $L_w^2(I;\mathscr{M}(\Omega))$.*

We will show (proposition 3.5) that $\mathrm{ext}(\mathscr{S})$ is non-empty. This will allow us to pick and characterize an arbitrary element of $\mathrm{ext}(\mathscr{S})$ in the proof of theorem 1.10. In order to show $\mathrm{ext}(\mathscr{S}) \neq \emptyset$, we use lemma 3.3 (see below) and the properties of $\mathscr{A}$ stated in remark 1.2.

Proof of remark 1.2. First, we argue that $\mathscr{A}$ is a subset of $L_w^2(I;\mathscr{M}(\Omega))$. This actually follows directly from the fact that W_1 induces the weak-$*$ topology on $\mathscr{P}(\Omega)$ [54, Thm. 6.9]. Now it is straightforward to check that the weak-$*$ topology coincides with the weak (subspace) topology on $\mathscr{P}(\Omega)$ stemming from $C(\Omega)'$. Hence, any map $f : I \to \mathscr{P}(\Omega)$ is Lebesgue measurable (relative to the weak-$*$ topology on $\mathscr{P}(\Omega)$) if and only if it is weakly measurable. Now let $\mu = \omega(\mu)\rho \in \mathscr{A}$. Then, by definition, any representative of ρ is Lebesgue measurable. Further, we clearly have $\|\rho\|_{L_w^2} = 1$. Thus, using also the above equivalence of the measurability notions, we obtain $\rho \in L_w^2(I;\mathscr{M}(\Omega))$. Further, since $L_w^2(I;\mathscr{M}(\Omega))$ is a vector space, we also have $\mu = \omega(\mu)\rho \in L_w^2(I;\mathscr{M}(\Omega))$, which proves $\mathscr{A} \subset L_w^2(I;\mathscr{M}(\Omega))$. Obviously, set $\mathscr{A}$ is a cone. Its convexity follows from

$$\begin{aligned} &\lambda\mu^1 + (1-\lambda)\mu^2 \\ &\quad = (\lambda\omega(\mu^1) + (1-\lambda)\omega(\mu^2)) \left[\frac{\lambda\omega(\mu^1)}{\lambda\omega(\mu^1) + (1-\lambda)\omega(\mu^2)}\rho^1 + \frac{(1-\lambda)\omega(\mu^2)}{\lambda\omega(\mu^1) + (1-\lambda)\omega(\mu^2)}\rho^2 \right] \in \mathscr{A} \end{aligned}$$

for all $\mu^1 = \omega(\mu^1)\rho^1, \mu^2 = \omega(\mu^2)\rho^2 \in \mathscr{A} \setminus \{0\}$ and $\lambda \in (0,1)$. We note that the convex combination in the angular brackets satisfies the required measurability (by the above) and its distance to δ_0 (with respect to W_1) is finite because Ω is bounded. □

The next lemma will be used in the proof of proposition 3.5 and corollary 4.8.

Lemma 3.3 (Properties of $\mathscr{R}$ with respect to $\mathscr{T}_*$)**.** *The regularizer $\mathscr{R}$ is lower semicontinuous with respect to $\mathscr{A}\cap\mathscr{T}_*$. Further, for every $c\in\mathbb{R}_+$ the sublevel set $L_c^-(\mathscr{R})\subset L_w^2(I;\mathscr{M}(\Omega))$ is compact relative to $\mathscr{T}_*$. In particular, if $(\mu^j)\subset L_c^-(\mathscr{R})$, then, up to a subsequence, we have $\mu^j \overset{*}{\rightharpoonup} \mu$ (in $\mathscr{T}_*$) and $\mu^j \overset{*}{\rightharpoonup} \mu$ a.e.*

Proof. Let $(\mu^j)\subset\mathscr{A}$ be an arbitrary sequence with $\mu^j \overset{*}{\rightharpoonup} \mu\in\mathscr{A}$. We can assume that $\liminf_j\mathscr{R}(\mu^j)<\infty$ (otherwise there is nothing to show) and restrict to a subsequence with $\liminf_j\mathscr{R}(\mu^j)=\lim_j\mathscr{R}(\mu^j)$. By lemma 2.9 we have (if necessary, restricting to a further subsequence) $\mu^j \overset{*}{\rightharpoonup} \tilde\mu$ a.e. for some $\tilde\mu\in\mathscr{A}$ with $\mathscr{R}(\tilde\mu)\le\liminf_j\mathscr{R}(\mu^j)$. Therefore, we only need to show that $\mu=\tilde\mu$. Let $\phi\in L_w^2(I;C(\Omega))$ be arbitrary. We have $f^j(t)=|\langle\phi_t,\tilde\mu_t-\mu_t^j\rangle|\to 0$ for a.e. $t\in I$. Further, we get $0\le f^j(t)\le f(t)=\|\phi_t\|_\infty(\omega(\tilde\mu)+\omega(\mu^j))$ for a.e. $t\in I$. Note that $f\in L^1(I)$ by the boundedness of $\int_I\|\phi\|_\infty\mathrm{d}\mathscr{L}$ (since I is bounded and $\|\phi\|_{L_w^2}<\infty$) and $\omega(\mu^j)$ (because $\omega(\mu^j)\le\mathscr{R}(\mu^j)$). Thus, by Lebesgue's dominated convergence theorem we obtain

$$\lim_j\langle\phi,\tilde\mu-\mu^j\rangle_{L_w^2}=\lim_j\int_I\langle\phi,\tilde\mu-\mu^j\rangle\mathrm{d}\mathscr{L}=\lim_j\int_I f^j(t)\mathrm{d}\mathscr{L}=\int_I\lim_j f^j(t)\mathrm{d}\mathscr{L}=0,$$

which (by the uniqueness of weak-$*$ limits) implies $\tilde\mu=\mu$. This proves the lower semicontinuity of $\mathscr{R}$ with respect to $\mathscr{A}\cap\mathscr{T}_*$. Now let $(\nu^j)\subset L_c^-(\mathscr{R})$. Since then $\mathscr{R}(\nu^j)$ is bounded by c, the weights $\omega(\nu^j)$ are also bounded. Using $\|\nu^j\|_{L_w^2}=\omega(\nu^j)$ and the Banach–Alaoglu theorem, we obtain $\nu^j \overset{*}{\rightharpoonup} \nu\in L_w^2(I;\mathscr{M}(\Omega))$ up to a subsequence. By the same argument as above we actually have (if necessary, restricting to a further subsequence) $\nu^j \overset{*}{\rightharpoonup} \tilde\nu$ a.e. for some $\tilde\nu\in\mathscr{A}$ with $\mathscr{R}(\tilde\nu)\le\liminf_j\mathscr{R}(\nu^j)\le c$ and $\nu=\tilde\nu\in\mathscr{A}$, which finalizes the proof. □

Example 3.4 (Convergence in $L_c^-(\mathscr{R})$)**.** *In the proof of lemma 3.3, we have seen that $(\mu^j)\subset L_c^-(\mathscr{R})$ with $\mu^j \overset{*}{\rightharpoonup} \mu\in L_c^-(\mathscr{R})$ (with respect to $\mathscr{T}_*$) implies $\mu^j \overset{*}{\rightharpoonup} \mu$ a.e. up to a subsequence. This is, in general, not true (in $L_c^-(\mathscr{R})$) for the whole sequence as the following example illustrates: For i and $j\in\{0,1,\dots,2^{i-1}-1\}$ define indices and (spatial and time) intervals by*

$$k=k(i,j)=2^{i-1}+j \qquad\text{and}\qquad \Omega_k=I_k=\begin{cases}2^{1-i}[j,j+1) & \text{if } j\neq 2^{i-1}-1,\\ 2^{1-i}[j,j+1] & \text{if } j=2^{i-1}-1.\end{cases}$$

For each i the intervals Ω_k partition $\Omega=[0,1]$ and $\mathscr{L}(\Omega_k)=2^{1-i}$. Now define (see figure 2)

$$\mu_t^k=\begin{cases}2^{i-1}\mathscr{L}\llcorner\Omega_k & \text{if } t\in I_k,\\ \delta_0 & \text{if } t\in I\setminus I_k.\end{cases}$$

We get $\omega(\mu^k)=1$ and $\operatorname{ess\,var}(\mu^k)\le 2W_1(\delta_0,2^{i-1}\mathscr{L}\llcorner\Omega_k)\le 2$, thus $\mathscr{R}(\mu^k)\le 3$. In particular, we obtain $\mu^k\in L_3^-(\mathscr{R})$. Next, we show that $\mu^k \overset{*}{\rightharpoonup} \mu$ in $L_w^2(I;\mathscr{M}(\Omega))$. Let $\phi: I\to(C(\Omega),\|\cdot\|_\infty)$ be continuous. Then

$$\langle\phi,\mu^k\rangle_{L_w^2}=2^{i-1}\int_{I_k}\int_{\Omega_k}\phi_t(x)\mathrm{d}\mathscr{L}(x)\mathrm{d}\mathscr{L}(t)+\int_{I\setminus I_k}\phi_t(0)\mathrm{d}\mathscr{L}(t). \tag{3}$$

We have

$$2^{i-1}\int_{I_k}\int_{\Omega_k}\phi_t(x)\mathrm{d}\mathscr{L}(x)\mathrm{d}\mathscr{L}(t)\le 2^{i-1}\max_{t\in I}\|\phi_t\|_\infty\int_{I_k}\int_{\Omega_k}1\mathrm{d}\mathscr{L}(x)\mathrm{d}\mathscr{L}(t)=2^{1-i}\max_{t\in I}\|\phi_t\|_\infty\to 0$$

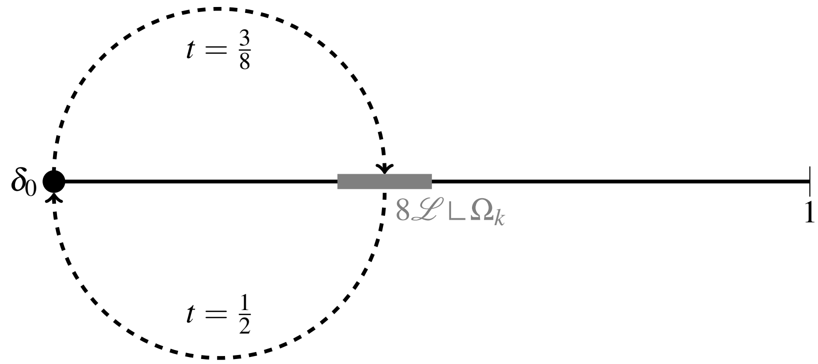


Figure 2: Sketch of the (piecewise constant) BV curve μ^k with $k=k(4,3)$ (example 3.4).

for $i\to\infty$. Further, we can estimate

$$\left|\int_{I\setminus I_k}\phi_t(0)\mathrm{d}\mathscr{L}(t)-\int_I\phi_t(0)\mathrm{d}\mathscr{L}(t)\right|\le\int_{I_k}|\phi_t(0)|\mathrm{d}\mathscr{L}(t)\le 2^{1-i}\max_{t\in I}\|\phi_t\|_\infty\to 0.$$

This yields that the right-hand side in equation (3) *converges to*

$$\langle\phi,\mu\equiv\delta_0\rangle_{L^2_w}=\int_I\phi_t(0)\mathrm{d}\mathscr{L}(t).$$

By density of continuous function in the Bochner space $L^2_w(I;C(\Omega))$, we have $\mu^k\overset{*}{\rightharpoonup}\mu$ in $L^2_w(I;\mathscr{M}(\Omega))$[3]*. On the other hand, for all $t\in I$ and $\psi\in C(\Omega)$ we get*

$$\langle\psi,\mu^k_t\rangle=\begin{cases}2^{i-1}\int_{\Omega_k}\psi\mathrm{d}\mathscr{L} & \text{if } t\in I_k,\\ \psi(0) & \text{if } t\in I\setminus I_k.\end{cases}$$

The right-hand side does, in general, not converge to $\langle\psi,\mu_t\rangle=\psi(0)$ for a.e. $t\in I$. For example, we can choose $\psi(x)=\min(2x,1)$ which yields $2^{i-1}\int_{\Omega_k}\psi\mathrm{d}\mathscr{L}=1$ for countably many k and $\psi(0)=0$. Hence, the convergence $\mu^k\to\mu$ a.e. is only valid up to a subsequence (e.g. subsequence μ^{k_ℓ} with $k_\ell=k(\ell,0)$).

Proposition 3.5 (ext($\mathscr{S}$) $\neq\emptyset$)**.** *Let assumption 1.9 be satisfied (the finiteness of Y is not needed). Then the set* ext($\mathscr{S}$) *is nonempty.*

Proof. We apply the Krein–Milman theorem. By remark 1.2 and $\mathscr{S}\subset\mathscr{A}$ we have $\mathscr{S}\subset L^2_w(I;\mathscr{M}(\Omega))$. The convexity of $\mathscr{S}$ follows directly from the convexity of $\mathscr{A}$ and assumption 1.9. The compactness of $\mathscr{S}$ (with respect to $\mathscr{T}_*$) follows from lemma 3.3: If $(\mu^i)\subset\mathscr{S}$, then $(\mu^i)\subset L^-_c(\mathscr{R})$ for some $c\in\mathbb{R}_+$ by optimality and $\mathscr{R}\ge 0$. Thus, using lemma 3.3 and the properties of $\mathscr{F},K$, we obtain $\mu^i\overset{*}{\rightharpoonup}\mu\in\mathscr{A}$ and $\mu^i\overset{*}{\rightharpoonup}\mu$ a.e. up to a subsequence and (by the proof of proposition 1.5) μ is also optimal, that is $\mu\in\mathscr{S}$. The Krein–Milman[4] theorem yields that the nonempty, convex, and compact set $\mathscr{S}\subset L^2_w(I;\mathscr{M}(\Omega))$ is equal to the convex hull of ext($\mathscr{S}$) which, in particular, implies ext($\mathscr{S}$) $\neq\emptyset$. □

[3] If $\varepsilon>0$, $\phi\in L^2_w(I;C(\Omega))$, and $\phi_\varepsilon\in L^2_w(I;C(\Omega))$ continuous with $\|\phi-\phi_\varepsilon\|_{L^2_w}<\varepsilon$, then $\langle\phi,\mu^k\rangle_{L^2_w}=\langle\phi-\phi_\varepsilon,\mu^k\rangle_{L^2_w}+\langle\phi_\varepsilon,\mu^k\rangle_{L^2_w}$ and $|\langle\phi-\phi_\varepsilon,\mu^k\rangle_{L^2_w}|\le\varepsilon\|\mu^k\|_{L^2_w}=\varepsilon$. Thus, we get $\langle\phi,\mu^k\rangle_{L^2_w}\to\langle\phi,\mu\rangle_{L^2_w}$ for $k\to\infty$.

[4] The weak-$*$ topology $\mathscr{T}_*$ is a Hausdorff locally convex vector topology [43, p. 68]. Note that the definition of 'vector topology' in [43, Section 1.6] includes the Hausdorff property (cf. [43, Thm. 1.12]).

Next, we show some further properties of the sublevel sets $L_c^-(\mathscr{R})$ which we will need in the proof of theorem 1.10.

Lemma 3.6 (Properties of $L_c^-(\mathscr{R})$)**.** *Let $c \in \mathbb{R}_+$. Then $L_c^-(\mathscr{R})$ is linearly closed, does not contain any rays, and its lineality space is trivial, that is* $\operatorname{lin}(L_c^-(\mathscr{R})) = \{0\}$.

Proof. Write $S = L_c^-(\mathscr{R})$.

S linearly closed: Consider a sequence $\mu^i = \mu^0 + \lambda^i \nu \in S$ where $\lambda^i \in \mathbb{R}$ with $\lambda^i \to \lambda \in \mathbb{R}$ and $\mu^0, \nu \in L_w^2(I;\mathscr{M}(\Omega)), \nu \neq 0$. We prove $\mu = \mu^0 + \lambda \nu \in S$. Since $\sup_i \mathscr{R}(\mu^i) \leq c$, lemma 2.9 implies the existence of $\hat{\mu} \in \mathscr{A}$ with $\mu^i \overset{*}{\rightharpoonup} \hat{\mu}$ a.e. (up to a subsequence) and $\mathscr{R}(\hat{\mu}) \leq \liminf_i \mathscr{R}(\mu^i)$, thus $\hat{\mu} \in S$. Clearly, by definition of μ^i, we also have $\mu^i \overset{*}{\rightharpoonup} \mu$ a.e. Therefore, we obtain $\mu = \hat{\mu} \in S$.

S does not contain any rays: By contradiction, assume that S contains a ray $R = \mu + \mathbb{R}_+^* \nu$, where $\mu, \nu \in L_w^2(I;\mathscr{M}(\Omega)), \nu \neq 0$. Define $\mu^k = \mu + k\nu \in R$ for $k \in \mathbb{N}$. Since $\mu^k \in S$, we have $c \geq \mathscr{R}(\mu^k) = \omega(\mu^k) + \operatorname{ess\,var}(\mu^k) \geq 0$. Therefore, we get

$$
\begin{aligned}
c \geq \omega(\mu^k) = \|\mu^k\|_{L_w^2} = \sup_\phi \langle \phi, \mu^k \rangle_{L_w^2} = \sup_\phi \langle \phi, \mu + k\nu \rangle_{L_w^2} = \sup_\phi \langle \phi, \mu \rangle_{L_w^2} + k \langle \phi, \nu \rangle_{L_w^2} \\
\geq \sup_\phi - \|\mu\|_{L_w^2} + k\langle \phi, \nu \rangle_{L_w^2} = -\|\mu\|_{L_w^2} + k\|\nu\|_{L_w^2},
\end{aligned}
$$

where the supremum is over $\phi \in L_w^2(I;C(\Omega))$ with $\|\phi\|_{L_w^2} \leq 1$. Letting $k \to \infty$ yields the contradiction.

$\operatorname{lin}(S) = \{0\}$: Note that always $0 \in \operatorname{lin}(S)$ by definition. Clearly, the lineality space $\operatorname{lin}(S)$ cannot be nontrivial, since then $\operatorname{rec}(S)$ would be nontrivial and thus S would contain a ray. This was already excluded. □

Proof of theorem 1.10. We apply [9, Thm. 1 & Cor. 2]. Write problem 1.4 as

$$
\min_{\mu \in \mathscr{A}} \mathscr{F}(K(\mu)) + \mathscr{R}(\mu) = \min_{\mu \in L_w^2(I;\mathscr{M}(\Omega))} \mathscr{F}(\widehat{K}(\mu)) + \widehat{\mathscr{R}}(\mu),
$$

where $\widehat{K}$ denotes any linear extension of K to $L_w^2(I;\mathscr{M}(\Omega))$ (assumption 1.9) and $\widehat{\mathscr{R}} : L_w^2(I;\mathscr{M}(\Omega)) \to \mathbb{R} \cup \{\infty\}$ is defined by

$$
\widehat{\mathscr{R}}(\mu) = \begin{cases} \mathscr{R}(\mu) & \text{if } \mu \in \mathscr{A}, \\ \infty & \text{else.} \end{cases}
$$

This problem is convex (assumption 1.9 and remark 1.2). Now let $\mu^{opt} \in \operatorname{ext}(\mathscr{S})$ (proposition 3.5). The assumptions in [9, Cor. 2] are satisfied: $\mathscr{F}$ is convex by assumption 1.9, the solution set $\mathscr{S}$ is nonempty by proposition 1.5, and the sublevel set $S = L_c^-(\widehat{\mathscr{R}}) = L_c^-(\mathscr{R})$ with $c = \mathscr{R}(\mu^{opt})$ is linearly closed by lemma 3.6. Further, forward operator $\widehat{K}$ is linear on $L_w^2(I;\mathscr{M}(\Omega))$ (cf. beginning of [9, Section 3.1]), regularizer $\widehat{\mathscr{R}}$ is convex (see setting in [9, Thm. 1]) because $\mathscr{R}$ is ($\mathscr{A}$ is convex by remark 1.2), and the dimension of the smallest face of $\mathscr{S}$ at μ^{opt} is equal to zero since $\mu^{opt} \in \operatorname{ext}(\mathscr{S})$. If $\mathscr{R}(\mu^{opt}) > 0$, then [9, Thm. 1] can be applied by [9, Cor. 2] if $\operatorname{lin}(S) = \{0\}$, which is true by lemma 3.6. Hence, in this case we get that μ^{opt} is a convex combination of at most m points in $\operatorname{ext}(S)$ or $m-1$ points of which each is in $\operatorname{ext}(S)$ or in an extremal ray of S (which is a ray R whose intersection with open line segments (μ^a, μ^b) equals (μ^a, μ^b)). Since S does not contain any rays (lemma 3.6), we remain in the previous case (because the convex combination is not necessarily strict). If $\mathscr{R}(\mu^{opt}) = 0$, then $\mu^{opt} = 0$ which is a convex combination of itself. Now we can apply proposition 1.8,

$$
\operatorname{ext}(S) = \operatorname{ext}\left(L^-_{\mathscr{R}(\mu^{opt})}(\mathscr{R})\right) = \{\mu \in \mathscr{A} \mid \mathscr{R}(\mu) \in \{0, \mathscr{R}(\mu^{opt})\}, \bar{\mu}_t = \omega(\mu)\delta_{x_t} \text{ for all } t \in [0,1)\}.
$$

Hence, if $\mu^{opt} = \sum_{i=1}^N \tilde{\lambda}^i \mu^i$ with $N \le m$, $\tilde{\lambda}^i \in [0,1]$, $\sum_{i=1}^N \tilde{\lambda}^i = 1$, and $\mu^i \in \mathrm{ext}(S)$, then $\mu_t^{opt} = \sum_{i=1}^N \lambda^i \delta_{x_t^i}$ for a.e. $t \in I$ with $\lambda^i = \tilde{\lambda}^i \omega(\mu^i) \in \mathbb{R}_+$. □

3.3 Discussion: extension to $\mathscr{P}(\Omega) \times \ldots \times \mathscr{P}(\Omega)$

One may consider a natural generalization of problem 1.4 in which $\mathscr{A} = \mathbb{R}_+ BV(I;\mathscr{W}_1(\Omega))$ is replaced by $\mathbb{R}_+ BV(I;\mathscr{W}_H(\Omega))$, where $\mathscr{W}_H(\Omega) = (\mathscr{P}(\Omega)^M, W_H)$ for some metric W_H, defined via a norm $H : \mathbb{R}^{M\times n} \to \mathbb{R}_+$, which generalizes W_1 (with a slight abuse of notation, we write $W_H = W_1$ if $M = 1$ and $H = |\cdot|$); metric W_H will be specified below. In this context, one may interpret M as the number of different labels of moving particles. However, a *natural* superposition principle in the spirit of [1, Thm. 3.3] (which was used to characterize the extremal points of $L_c^-(\mathscr{R})$ in proposition 1.8) does not exist for this setting. In this section, we aim to illustrate this aspect. As a motivation, let us recall the Beckmann formulation of the Wasserstein 1-distance [44, Thm. 4.6] between $\rho^+, \rho^- \in \mathscr{P}(\Omega)$,

$$W_1(\rho^+,\rho^-) = \min_T \|T\|_{\mathscr{M}},$$

where the minimum is over vector-valued Radon measures $T : \mathscr{B}(\Omega) \to \mathbb{R}^n$ with distributional divergence equal to $\rho^+ - \rho^-$ (equivalently, normal 1-currents in Ω whose boundary is equal to $\rho^- - \rho^+$). Now let $M > 1$ and $\vec{\rho}^+ = (\rho_1^+,\ldots,\rho_M^+)^\intercal, \vec{\rho}^- = (\rho_1^-,\ldots,\rho_M^-)^\intercal \in \mathscr{P}(\Omega)^M$. Fix a norm $H : \mathbb{R}^{M\times n} \to \mathbb{R}_+$. Define

$$W_H(\vec{\rho}^+,\vec{\rho}^-) = \min_{\vec{T}} \|\vec{T}\|_H,$$

where the minimum is over matrix-valued Radon measures $\vec{T} : \mathscr{B}(\Omega) \to \mathbb{R}^{M\times n}$ with row-wise distributional divergence equal to $\vec{\rho}^+ - \vec{\rho}^-$ (equivalently, normal 1-currents in Ω with coefficients in $\mathbb{R}^M$ whose component-wise boundary is equal to $\vec{\rho}^- - \vec{\rho}^+$), and $\|\cdot\|_H$ denotes the total variation with respect to H, that is $\|\vec{T}\|_H = \sup\{H(\vec{T}(B_1)) + H(\vec{T}(B_2)) + \ldots \mid B_1, B_2, \ldots \in \mathscr{B}(\Omega) \text{ partition of } \Omega\}$. If $H(\theta \otimes \vec{e}) = h(\theta)$ for all $\theta \in \mathbb{R}^M$, $\vec{e} \in \mathbb{R}^n$ with $|\vec{e}| = 1$ for some norm $h : \mathbb{R}^M \to \mathbb{R}_+$, then by [32, Thm. 1.10] $W_H(\vec{\rho}^+,\vec{\rho}^-)$ can be interpreted as a multi-material transport problem [37, 36] ($h(\theta)$ denotes the cost to transport material vector θ per unit distance), cf. [38, Ch. 4.2]. Next, we identify an admissible $\vec{T}$ in $W_H(\vec{\rho}^+,\vec{\rho}^-)$ with a càdlàg curve $\mu : I \to \mathscr{W}_H(\Omega)$. For simplicity, we assume that the divergence-free part of $\vec{T}$ is equal to zero. We follow the proof of [13, Prop. 4.1]. By [50, Thm C] there exists an M-tuple $\vec{\eta} = (\eta_1,\ldots,\eta_M)^\intercal$ of measures $\eta_i : (\Theta, \mathscr{B}(\Theta)) \to [0,\infty)$ (Θ denoting a space of Lipschitz curves which are identified modulo parameterization with topology induced by a metric d_Θ, cf. [21, Def. 2.5]) such that

$$\int_\Omega \Phi : \mathrm{d}\vec{T} = \int_\Theta \int_I \Phi(\gamma)\dot{\gamma}\mathrm{d}\mathscr{L} \cdot \mathrm{d}\vec{\eta}(\gamma)$$

for all $\Phi \in C(\Omega;\mathbb{R}^{M\times n})$, where : denotes the Frobenius inner product. By Skorohod's theorem [8, Thm. 6.7] $\vec{\eta}$ is induced by an M-tuple $\vec{\chi} = (\chi_1,\ldots,\chi_M)^\intercal$ of so-called irrigation patterns $\chi_i : [0,1] \times I \to \Omega$ between ρ_i^+ and ρ_i^- [34, 7, 35], that is χ_i is Borel measurable, $\chi_i(p,\cdot)$ is absolutely continuous for $\mathscr{L}$-a.e. $p \in [0,1]$ ($\chi_i(p,t)$ represents the position of particle p in the reference space $([0,1], \mathscr{B}([0,1]), \mathscr{L} \llcorner [0,1])$ at time $t \in I$), and $\chi_i(\cdot,0)_\#(\mathscr{L} \llcorner [0,1]) = \rho_i^+, \chi_i(\cdot,1)_\#(\mathscr{L} \llcorner [0,1]) = \rho_i^-$ for all $i = 1,\ldots,M$. Now define $\mu : I \to \mathscr{W}_H(\Omega)$ (note that $\vec{\chi}(\cdot,t) : [0,1] \to \Omega^M$ is Borel measurable) by

$$\mu_t = \vec{\chi}(\cdot,t)_\#(\mathscr{L} \llcorner [0,1])$$

for every $t \in I$. Then μ is a càdlàg curve. Indeed, for every $t \in I$, $\varepsilon \in \mathbb{R} \setminus \{0\}$ with $t+\varepsilon \in I$, and $\phi \in C(\Omega;\mathbb{R}^M)$ we have

$$\left| \int_\Omega \phi \cdot \mathrm{d}(\mu_t - \mu_{t+\varepsilon}) \right| \le \sum_{i=1}^{M} \int_{[0,1]} |\phi_i(\chi(p,t)) - \phi_i(\chi(p,t+\varepsilon))| \mathrm{d}\mathscr{L}(p) \to 0$$

for $\varepsilon \to 0$ by the regularity of ϕ and continuity of $t \mapsto \chi(p,t)$ for $\mathscr{L}$-a.e. $p \in [0,1]$ (Lebesgue's dominated convergence theorem). Thus, we have $\mu_{t+\varepsilon} \overset{*}{\rightharpoonup} \mu_t$ for $\varepsilon \to 0$, hence $W_H(\mu_{t+\varepsilon}, \mu_t) \to 0$ for $\varepsilon \to 0$ since W_H metrizes weak-$*$ convergence by norm equivalence. This shows that $\mu : I \to \mathscr{W}_H(\Omega)$ is càdlàg. Nevertheless, a natural superposition principle according to [1, Thm. 3.3] does not exist which is demonstrated in the following example.

Example 3.7 (Decompositions of $|\mathrm{D}\mu|$ for $\mu : I \to \mathscr{W}_H$)**.** *Let $M = n = 2$, $\alpha = \frac{3\pi}{4}$, and $x = \binom{\cos\alpha}{\sin\alpha}, y = \binom{\cos\alpha}{-\sin\alpha}, z = \binom{1}{0}$. Further assume that $H : \mathbb{R}^{2\times 2} \to \mathbb{R}_+$ satisfies $H(\theta \otimes \vec{e}) = h(\theta)$ for all $\theta \in \mathbb{R}^2$, $\vec{e} \in \mathbb{R}^2$ with $|\vec{e}| = 1$ for some norm $h : \mathbb{R}^2 \to \mathbb{R}_+$. Define Lipschitz curves $\gamma_1, \gamma_2 : I \to \Omega = [-1,1]^2$ (cf. figure 3a) by*

$$\gamma_1(t) = \begin{cases} (1-2t)x & \text{if } t \in [0,\frac{1}{2}), \\ (2t-1)z & \text{else,} \end{cases} \qquad \text{and} \qquad \gamma_2(t) = \begin{cases} (1-2t)y & \text{if } t \in [0,\frac{1}{2}), \\ (2t-1)z & \text{else.} \end{cases}$$

Further, let $\mu : I \to \mathscr{W}_H$ be the continuous curve given by

$$\mu_t = \begin{pmatrix} \delta_{\gamma_1(t)} \\ \delta_{\gamma_2(t)} \end{pmatrix} \qquad \text{associated with} \qquad \vec{T} = \frac{1}{2} \begin{pmatrix} \dot{\gamma}_1^\top \mathscr{H}^1 \llcorner (\gamma_1(I)) \\ \dot{\gamma}_2^\top \mathscr{H}^1 \llcorner (\gamma_2(I)) \end{pmatrix}$$

for $t \in I$, where $\mathscr{H}^1$ denotes the one-dimensional Hausdorff measure. Now we may define

$$\zeta = h(1,0)\delta_{\gamma_1} + h(0,1)\delta_{\gamma_2} \in \mathbb{R}_+\mathscr{P}(\mathscr{D}_E(\Omega)).$$

Then we have

$$\begin{aligned} |\mathrm{D}\mu|((0,1/2)) = \mathrm{var}(\mu,(0,1/2)) = h(1,0)\,\mathrm{var}(\gamma_1,(0,1/2)) + h(0,1)\,\mathrm{var}(\gamma_2,(0,1/2)) \\ = \int_{\mathscr{D}_E(\Omega)} |\mathrm{D}\gamma|((0,1/2))\mathrm{d}\zeta(\gamma). \end{aligned}$$

However, we get

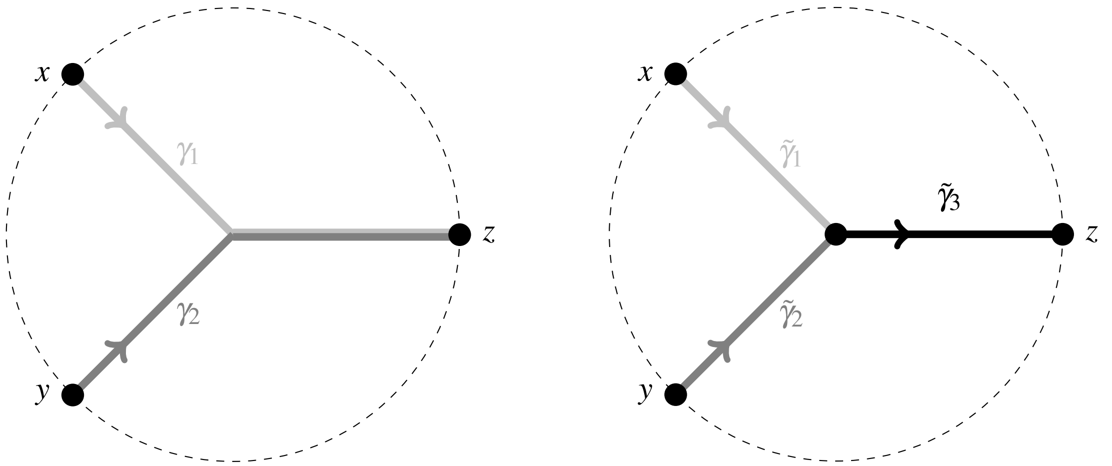
$$|\mathrm{D}\mu|((1/2,1)) = h(1,1)\,\mathrm{var}(\gamma_1,(1/2,1)) \neq \int_{\mathscr{D}_E(\Omega)} |\mathrm{D}\gamma|((1/2,1))\mathrm{d}\zeta(\gamma) \quad \text{if and only if} \quad h(1,1) < h(1,0) + h(0,1).$$

This issue may be solved by considering (see figure 3b)

$$\tilde{\gamma}_1(t) = \begin{cases} \gamma_1(t) & \text{if } t \in [0,\frac{1}{2}), \\ \gamma_1(\frac{1}{2}) & \text{else,} \end{cases} \quad \tilde{\gamma}_2(t) = \begin{cases} \gamma_2(t) & \text{if } t \in [0,\frac{1}{2}), \\ \gamma_2(\frac{1}{2}) & \text{else,} \end{cases} \quad \text{and} \quad \tilde{\gamma}_3(t) = \begin{cases} \gamma_1(\frac{1}{2}) & \text{if } t \in [0,\frac{1}{2}), \\ \gamma_1(t) & \text{else.} \end{cases}$$

Define $\tilde{\zeta} = h(1,0)\delta_{\tilde{\gamma}_1} + h(0,1)\delta_{\tilde{\gamma}_2} + h(1,1)\delta_{\tilde{\gamma}_3}$. Then $|\mathrm{D}\mu| = \int_{\mathscr{D}_E(\Omega)} |\mathrm{D}\tilde{\gamma}|\mathrm{d}\tilde{\zeta}(\tilde{\gamma})$ but $\tilde{\zeta}$ is not normalized: $\tilde{\zeta} \notin \mathscr{P}(\mathscr{D}_E(\Omega))$. Further, measure $\tilde{\zeta}$ is not in a natural sense defined via μ. Another approach would be to consider $\vec{\zeta} \in \mathscr{P}(\mathscr{D}_E(\Omega))^2$ given by

$$\vec{\zeta} = \begin{pmatrix} \delta_{\gamma_1} \\ \delta_{\gamma_2} \end{pmatrix}.$$



(a) $z = \gamma_1(1) = \gamma_2(1)$.

(b) $\{0\} = \tilde{\gamma}_1([\frac{1}{2},1]) = \tilde{\gamma}_2([\frac{1}{2},1]) = \tilde{\gamma}_3([0,\frac{1}{2}])$.

Figure 3: Lipschitz curves γ_i and $\tilde{\gamma}_i$ from example 3.7.

Then we have

$$\int_{\mathscr{D}_E(\Omega)} h\left(\frac{\mathrm{d}\vec{\zeta}}{\mathrm{d}|\vec{\zeta}|}(\gamma)\right)|\mathrm{D}\gamma|\mathrm{d}|\vec{\zeta}|(\gamma) = h(1,0)|\mathrm{D}\gamma_1| + h(0,1)|\mathrm{D}\gamma_2| \neq |\mathrm{D}\mu| \quad \text{if and only if} \quad h(1,1) < h(1,0) + h(0,1).$$

Finally, we can resolve this by defining $\vec{\xi} \in \mathscr{P}(\mathscr{D}_E(\Omega))^2$ *by*

$$\vec{\xi} = \frac{1}{2}\begin{pmatrix} \delta_{\tilde{\gamma}_1} + \delta_{\tilde{\gamma}_3} \\ \delta_{\tilde{\gamma}_2} + \delta_{\tilde{\gamma}_3} \end{pmatrix}.$$

This yields

$$\int_{\mathscr{D}_E(\Omega)} h\left(\frac{\mathrm{d}\vec{\xi}}{\mathrm{d}|\vec{\xi}|}(\gamma)\right)|\mathrm{D}\gamma|\mathrm{d}|\vec{\xi}|(\gamma) = h(1,0)|\mathrm{D}\tilde{\gamma}_1| + h(0,1)|\mathrm{D}\tilde{\gamma}_2| + h(1,1)|\mathrm{D}\tilde{\gamma}_2| = |\mathrm{D}\mu|$$

as desired. Measure $\vec{\xi}$ *is still not naturally defined via* μ. *Nevertheless, it is normalized in the sense that* $\vec{\xi} \in \mathscr{P}(\Omega)^2$. *We believe that the formula (with* $\vec{\xi}$ *defined via* μ*)*

$$|\mathrm{D}\mu| = \int_{\mathscr{D}_E(\Omega)} h\left(\frac{\mathrm{d}\vec{\xi}}{\mathrm{d}|\vec{\xi}|}(\gamma)\right)|\mathrm{D}\gamma|\mathrm{d}|\vec{\xi}|(\gamma)$$

holds in a general setting.

4 Fully-corrective generalized conditional gradient method for BV curve tracking

In this section, we design an algorithm which computes (approximate) solutions of problem 1.4 by exploiting the sparse structure given by theorem 1.10 and, in particular, using the characterization of extremal points (proposition 1.8). The algorithm will be an instance of the (grid-free) fully-corrective generalized conditional gradient (FC-GCG) algorithm introduced in [18] applied to our BV curve tracking problem. Similar algorithms have been used to approximate solutions of dynamic inverse problems regularized with Wasserstein-2 transport energies [16, 23]. In order to apply [18, Alg. 1], we require the following (in addition to the assumptions in problem 1.4).

Assumption 4.1 (Numerical setup)**.** *The fidelity $\mathscr{F}$ is strictly convex, $\mathbb{R}$-valued, and Fréchet differentiable such that the restriction of $D\mathscr{F} : Y \to Y$ to any compact subset of Y is Lipschitz. Further, forward operator K is a linear map which can be linearly extended to a weak-$*$-to-strong continuous map on $L^2_w(I;\mathscr{M}(\Omega))$.*

Note that, compared to assumption 1.9, data space Y is not necessarily finite-dimensional. In our numerical experiments, we will use $Y = \mathbb{R}^{L\times(M+1)}$ (implying the validity of theorem 1.10 under assumption 4.1) and a discretization of an instance of the class of forward operators specified in example 1.11. Further, we will use a quadratic fidelity term as in remark 1.6 (whose Fréchet derivative at y_0 applied to y is just the inner product $(y_0 - f, y)_Y$ and therefore satisfies assumption 4.1). Moreover, we re-introduce the regularization parameters $\alpha, \beta \in \mathbb{R}^*_+$ to calibrate the (discrete) regularizer in our computations.

Problem 4.2 (Auxiliary problem formulation on $L^2_w(I;\mathscr{M}(\Omega))$)**.** *Let assumption 4.1 be satisfied and $\widehat{K}$ any linear extension of K to $L^2_w(I;\mathscr{M}(\Omega))$ which is weak-$*$-to-strong continuous. Further, define $\widehat{\mathscr{R}}_{\alpha,\beta} : L^2_w(I;\mathscr{M}(\Omega)) \to \mathbb{R}\cup\{\infty\}$ as in the proof of theorem 1.10,*

$$\widehat{\mathscr{R}}_{\alpha,\beta}(\mu) = \begin{cases} \mathscr{R}_{\alpha,\beta}(\mu) & \text{if } \mu \in \mathscr{A}, \\ \infty & \text{else.} \end{cases}$$

Then problem 1.4 becomes

$$\min_{\mu \in L^2_w(I;\mathscr{M}(\Omega))} \mathscr{J}_{\alpha,\beta}(\mu), \qquad \text{where} \qquad \mathscr{J}_{\alpha,\beta}(\mu) = \mathscr{F}(\widehat{K}\mu) + \widehat{\mathscr{R}}_{\alpha,\beta}(\mu).$$

Remark 4.3 (Existence of pre-adjoint of $\widehat{K}$)**.** *Clearly, forward operator $\widehat{K} : L^2_w(I;\mathscr{M}(\Omega)) \to Y$ is weak-$*$-to-weak continuous. Thus, it admits a continuous and linear pre-adjoint [20, Rem. 3.2]. A simple application of the Hahn–Banach separation theorem shows that it is unique.*

Notation 4.4 (Pre-adjoint of $\widehat{K}$)**.** *We write $\widehat{K}_* : Y \to L^2_w(I;C(\Omega))$ for the pre-adjoint of $\widehat{K}$ in problem 4.2, that is*

$$\langle \widehat{K}_* f, \mu \rangle_{L^2_w} = (f, \widehat{K}\mu)_Y$$

for all $f \in Y, \mu \in L^2_w(I;\mathscr{M}(\Omega))$.

Proposition 4.5 (Embedding into the setting of [18, Section 2])**.** *Consider problem 4.2. Then the assumptions in [18, Section 2] (including [18, Section 2, (A1)-(A3)]) are satisfied.*

Proof. Regularizer $\widehat{\mathscr{R}}$ is convex by the convexity of $\mathscr{A}$ (remark 1.2), lower semicontinuous with compact sublevel sets (with respect to $\mathscr{T}_*$) by lemma 3.3, and nonnegative homogeneous by definition: $\widehat{\mathscr{R}}(\lambda\mu) = \lambda\widehat{\mathscr{R}}(\mu)$ for all $\lambda \in \mathbb{R}_+, \mu \in L^2_w(I;\mathscr{M}(\Omega))$ (with the convention $0\cdot\infty = 0$). The other requirements in [18, Section 2] readily follow from the setup in problem 4.2. $\square$

4.1 Description of the algorithm

In this section, we explain the FC-GCG algorithm which we will use to compute (approximate) minimizers of problem 4.2 and relate it to the steps of the general FC-GCG algorithm in [18].

In the k-th iteration, the algorithm updates two *active sets* (term 'active' indicates that they are modified at each iteration), namely a set of BV curves in Ω denoted by $\Gamma^k = \{\gamma_i^k\}_{i=1}^{N_k} \subset BV(I;\Omega)$ and a set of weights $\Lambda^k = \{\lambda_i^k\}_{i=1}^{N_k} \subset \mathbb{R}_+^*$ such that the BV curves $\mu^k \in \mathscr{A}$ given by

$$\mu^k = \sum_{i=1}^{N_k} \lambda_i^k \delta_{\gamma_i^k}$$

converge (up to a subsequence, with respect to $\mathscr{T}_*$) to a solution of problem 4.2 (see corollary 4.8). Note that each iterate μ^k is sparse but, a priori, the limit (or ground truth) may be non-sparse. At each iteration of the algorithm, the sets Γ^k and Λ^k are updated to Γ^{k+1} and Λ^{k+1} following several steps which we will describe below. We use the following abbreviation.

Notation 4.6 (Dual variable p^k). *In the remainder, we write (often referred to as 'dual variable')*

$$p^k = -\widehat{K}_* D\mathscr{F}(\widehat{K}\mu^k) \in L_w^2(I;C(\Omega)),$$

see, e.g., [18, Proposition 2.3].

The update of the active sets Γ^k and Λ^k consists of the following three steps:

1. Insertion step: A new BV curve $\gamma_{N_k+1}^k \in BV(I;\Omega)$ is determined by solving the following variational problem:

$$\gamma_{N_k+1}^k \in \underset{\gamma \in BV(I;\Omega)}{\arg\max} \max \left\{ a(\gamma) \int_I p_t^k(\gamma(t)) \mathrm{d}\mathscr{L}(t), 0 \right\}, \tag{4}$$

where

$$a(\gamma) = (\alpha + \beta \operatorname{ess\,var}(\gamma))^{-1}. \tag{5}$$

Problem (4) corresponds to the insertion step of FC-GCG algorithms defined in [18, (3.1)],

$$\hat{u} \in \underset{u \in \operatorname{ext}(L_1^-(\mathscr{R}_{\alpha,\beta}))}{\arg\max} \langle p^k, u \rangle_{L_w^2}. \tag{6}$$

Indeed, due to our characterization of extremal points in proposition 1.8 (compare with [16, (8)]) one can readily check that problem (6) is equivalent to problem (4). As a consequence, the arg max in problem (4) is non-empty by [18, Lem. A.1]. Note that, in general, functions p_t^k may be non-convex for all t in a set of positive $\mathscr{L}$-measure which makes problem (4) non-convex. Therefore, its accurate solution is challenging. For details related to the practical implementation of the optimization in problem (4) we refer to section 4.3. After problem (4) is solved, the BV curve $\gamma_{N_k+1}^k$ is added to the active set of curves, $\Gamma^{k,+} = \Gamma^k \cup \{\gamma_{N_k+1}^k\}$. This justifies the name insertion step.

2. Coefficients optimization step: The active set Λ^k is updated to $\Lambda^{k,+} = \{\hat{\lambda}_i^k\}_{i=1}^{N_k+1}$ where $\hat{\lambda}_i^k$ are obtained by solving the finite-dimensional problem

$$\hat{\lambda}^k = \underset{\lambda \in \mathbb{R}_+^{N_k+1}}{\arg\min} \mathscr{F}\left(\sum_{i=1}^{N_k+1} \lambda_i a(\gamma_i^k) \widehat{K} \delta_{\gamma_i^k} \right) + \sum_{i=1}^{N_k+1} \lambda_i, \tag{7}$$

where $\gamma_1^k, \dots, \gamma_{N_k}^k \in \Gamma^k$ are transferred from the $(k-1)$-th iteration. This problem corresponds to the coefficients optimization step of FC-GCG algorithms defined in [18, (3.4)]. Problem (7) is easily solvable by proximal methods or Newton methods.

3. Pruning: We define N_{k+1} as the number of non-zero coefficients in $\hat{\lambda}^k$ and construct Γ^{k+1} and Λ^{k+1} by removing from $\Gamma^{k,+}$ and $\Lambda^{k,+}$ the elements whose indices correspond to the zero coefficients.

Stopping criterion: We add a stopping criterion which is derived from the optimality conditions associated with problem 4.2. In particular, we terminate the algorithm in iteration $k \in \mathbb{N}$ with output μ^k if

$$a(\gamma_{N_k+1}^k) \int_I p_t^k(\gamma_{N_k+1}^k(t)) \mathrm{d}\mathscr{L}(t) \leq 1. \tag{8}$$

This criterion matches the one proposed in [18, Prop. 3.1] where it is formulated as (cf. [18, (3.5)])

$$\max_{u \in \mathrm{ext}(L_1^-(\mathscr{R}_{\alpha,\beta}))} \langle p^k, u \rangle_{L_w^2} \leq 1.$$

Again, this follows from our characterization of extremal points in proposition 1.8. In our implementation, we will add a tolerance $\varepsilon > 0$ on the right-hand side of inequality (8).

Remark 4.7 (Insertion of $0 \in \mathscr{A}$)**.** *Note that, if*

$$\max_{\gamma \in BV(I;\Omega)} \max \left\{ a(\gamma) \int_I p_t^k(\gamma(t)) \mathrm{d}\mathscr{L}(t), 0 \right\} = 0,$$

which corresponds to the insertion of the extremal point $\hat{u} = 0$ in problem (6), then it necessarily holds that $\max_{\gamma \in BV(I;\Omega)} a(\gamma) \int_I p_t^k(\gamma(t)) \mathrm{d}\mathscr{L}(t) \leq 1$, implying that the stopping criterion in inequality (8) is satisfied. Therefore, we can replace problem (4) with the simplified insertion step

$$\gamma_{N_k+1}^k \in \operatorname*{arg\,max}_{\gamma \in BV(I;\Omega)} a(\gamma) \int_I p_t^k(\gamma(t)) \mathrm{d}\mathscr{L}(t). \tag{9}$$

Algorithm 1 summarizes the steps described above. [18, Thm. 3.3] and proposition 4.5 imply the convergence properties of the iterates produced by algorithm 1. We state the rate of convergence in terms of the residuals

$$r(\mu^k) = \mathscr{J}_{\alpha,\beta}(\mu^k) - \min_{\mu \in \mathscr{A}} \mathscr{J}_{\alpha,\beta}(\mu). \tag{10}$$

Corollary 4.8 (Sublinear convergence of algorithm 1, cf. [18, Thm. 3.3])**.** *The iterates $\mu^k \in \mathscr{A}$ produced by algorithm 1 (extended with $\mu^k = \mu^{k_0}$ for $k \geq k_0$ if the algorithm terminates in iteration k_0) converge, up to subsequences, in $\mathscr{T}_*$ to a minimizer of problem 4.2. Moreover, there exists $C \in \mathbb{R}_+^*$ such that*

$$r(\mu^k) \leq \frac{C}{k+1}$$

for all $k \in \mathbb{N}$.

Algorithm 1 Fully-corrective generalized conditional gradient method for BV curve tracking

Require: $\Lambda^0 = \emptyset$, $\Gamma^0 = \emptyset$, $N_0 = 0$, $\mu^0 = 0$
  for $k = 0, 1, 2, \dots$ **do**
    $p^k \leftarrow -\widehat{K}_* D\mathscr{F}(\widehat{K}\mu^k)$
    $\gamma^k_{N_k+1} \in \arg\max_{\gamma \in BV(I;\Omega)} a(\gamma) \int_I p^k_t(\gamma(t)) \mathrm{d}\mathscr{L}(t)$
    if $k \in \mathbb{N}$ and $a(\gamma^k_{N_k+1}) \int_I p^k_t(\gamma^k_{N_k+1}(t)) \mathrm{d}\mathscr{L}(t) \leq 1$ **then**
      return μ^k
    end if
    $(\hat{\lambda}^k_1, \dots, \hat{\lambda}^k_{N_k+1})^\intercal \in \arg\min_{\lambda \in \mathbb{R}_+^{N_k+1}} \mathscr{F}\left(\sum_{i=1}^{N_k+1} \lambda_i a(\gamma^k_i) \widehat{K} \delta_{\gamma^k_i}\right) + \sum_{i=1}^{N_k+1} \lambda_i$
    $\Gamma^{k+1} \leftarrow \left\{\gamma^k_i \in \Gamma^k \cup \{\gamma^k_{N_k+1}\} \,\middle|\, \hat{\lambda}^k_i > 0\right\}$, $\Lambda^{k+1} \leftarrow \{\hat{\lambda}^k_i \,|\, \hat{\lambda}^k_i > 0\}$
    $\mu^{k+1} \leftarrow \sum_{i=1}^{N_k+1} \hat{\lambda}^k_i \delta_{\gamma^k_i}$
    $N_{k+1} \leftarrow \#\Gamma^{k+1}$
  end for

4.2 Discretization approach

In our implementation of algorithm 1, we will use discretizations which are obtained by letting the temporal blur at discrete time points go to zero. In this section, we will make this conception rigorous. We will need the following definition.

Definition 4.9 (Nascent (θ,t)-delta functions)**.** *Let $\theta \in [0,1]$ and $t \in I$. A family of functions $\phi^\delta \in C(I;\mathbb{R}_+)$ (indexed by $\delta \in \mathbb{R}^*_+$) is called a family of* ***nascent*** (θ,t)***-delta functions*** *if*

- $\phi^\delta(s) = 0$ *for all* $s \in I \setminus [t-\delta, t+\delta]$,
- *the numbers* $\theta^\delta = \int_{[0,t]} \phi^\delta \mathrm{d}\mathscr{L}$ *satisfy*

$$\theta^\delta \in [0,1], \qquad \int_{[t,1]} \phi^\delta \mathrm{d}\mathscr{L} = 1 - \theta^\delta, \qquad \text{and} \qquad \theta^\delta \to \theta \text{ for } \delta \to 0.$$

Note that necessarily $\theta^\delta \equiv 0$ if $t = 0$ and $\theta^\delta \equiv 1$ if $t = 1$. Thus, for $t = 0$ (respectively $t = 1$) such a family only exists if $\theta = 0$ (respectively $\theta = 1$).

Lemma 4.10 (‘(θ,t)-delta’)**.** *Let $\psi \in BV(I;\mathbb{R})$ and $\phi^\delta \in C(I;\mathbb{R}_+)$ a family of nascent (θ,t)-delta functions. Then we have*

$$\lim_{\delta\to 0} \int_{[0,t]} \phi^\delta \psi \mathrm{d}\mathscr{L} = \theta \bar{\psi}^-_t \qquad \text{and} \qquad \lim_{\delta\to 0} \int_{[t,1]} \phi^\delta \psi \mathrm{d}\mathscr{L} = (1-\theta)\bar{\psi}_t,$$

where $\bar{\psi}$ is any càdlàg representative of ψ.

Proof. We only prove the first equality (the proof of the second is similar). Let $\varepsilon \in \mathbb{R}^*_+$ and $\delta = \delta(\varepsilon) \in \mathbb{R}^*_+$ with $\bar{\psi}_s \in [\bar{\psi}^-_t - \varepsilon, \bar{\psi}^-_t + \varepsilon]$ for all $s \in [t-\delta, t] \cap I$ and $\delta \to 0$ for $\varepsilon \to 0$ (possible because $\bar{\psi}^-_t$ exists). Then, using $\int_{[0,t]} \phi^\delta \mathrm{d}\mathscr{L} = \theta^\delta \to \theta$, $\phi^\delta \in \mathbb{R}_+$, and $\phi = 0$ on $[0, t-\delta) \cap I$, we obtain

$$\theta \bar{\psi}^-_t = \lim_{\varepsilon\to 0} \int_{[0,t]} \phi^\delta \mathrm{d}\mathscr{L} \cdot (\bar{\psi}^-_t - \varepsilon) \leq \lim_{\varepsilon\to 0} \int_{[0,t]} \phi^\delta \psi \mathrm{d}\mathscr{L} \leq \lim_{\varepsilon\to 0} \int_{[0,t]} \phi^\delta \mathrm{d}\mathscr{L} \cdot (\bar{\psi}^-_t + \varepsilon) = \theta \bar{\psi}^-_t. \qquad \square$$

The following statement will be applied to our forward operator from example 1.11.

Proposition 4.11 (Vanishing temporal blur)**.** *Let $\mu \in \mathscr{A}$, $\phi^\delta \in C(I;\mathbb{R}_+)$ a family of nascent (θ,t)-delta functions, and $\Phi : \Omega \to \mathbb{R}$ Lipschitz. Then we get*

$$\lim_{\delta\to 0}\int_{[0,t]}\phi^\delta(s)\int_\Omega \Phi(x)\mathrm{d}\mu_s(x)\mathrm{d}\mathscr{L}(s) = \theta\int_\Omega \Phi(x)\mathrm{d}\bar{\mu}_t^-(x) \qquad \text{and}$$
$$\lim_{\delta\to 0}\int_{[t,1]}\phi^\delta(s)\int_\Omega \Phi(x)\mathrm{d}\mu_s(x)\mathrm{d}\mathscr{L}(s) = (1-\theta)\int_\Omega \Phi(x)\mathrm{d}\bar{\mu}_t(x),$$

where $\bar{\mu}$ is any càdlàg representative of μ.

Proof. As in lemma 4.10, we only prove the first equality (the proof of the second is similar). Define $\psi_s = \int_\Omega \Phi \mathrm{d}\mu_s$ for a.e. $s \in I$ (recall that $\psi \in L^1(I;\mathbb{R})$ by $\mu \in L^2_w(I;\mathscr{M}(\Omega))$). We prove $\psi \in BV(I;\mathbb{R})$ and apply lemma 4.10. We have ($\mathrm{Lip}(\Phi)$ denoting the Lipschitz constant of Φ, we can assume $\mathrm{Lip}(\Phi) \in \mathbb{R}_+^*$ because the other case is trivial)

$$\begin{aligned}
\operatorname{ess\,var}(\psi) \leq \operatorname{var}\left(s \mapsto \int_\Omega \Phi \mathrm{d}\bar{\mu}_s\right) &= \sup\left\{\sum_{i=1}^N \left|\int_\Omega \Phi \mathrm{d}(\bar{\mu}_{s_i} - \bar{\mu}_{s_{i-1}})\right| \,\middle|\, s_0 < s_1 < \ldots < s_N,\, s_i \in I\right\} \\
&= \mathrm{Lip}(\Phi)\sup\left\{\sum_{i=1}^N \left|\int_\Omega \frac{1}{\mathrm{Lip}(\Phi)}\Phi \mathrm{d}(\bar{\mu}_{s_i} - \bar{\mu}_{s_{i-1}})\right| \,\middle|\, s_0 < s_1 < \ldots < s_N,\, s_i \in I\right\} \\
&\leq \mathrm{Lip}(\Phi)\sup\left\{\sum_{i=1}^N W_1(\bar{\mu}_{s_i}, \bar{\mu}_{s_{i-1}}) \,\middle|\, s_0 < s_1 < \ldots < s_N,\, s_i \in I\right\} = \mathrm{Lip}(\Phi)\operatorname{var}(\bar{\mu}) < \infty
\end{aligned}$$

since $\mu \in \mathscr{A}$, where we used the Kantorovich–Rubinstein formula in the second inequality. Hence, we obtain $\psi \in BV(I;\mathbb{R})$. Application of lemma 4.10 yields

$$\lim_{\delta\to 0}\int_{[0,t]}\phi^\delta(s)\int_\Omega \Phi \mathrm{d}\mu_s \mathrm{d}\mathscr{L}(s) = \theta\bar{\psi}_t^- = \theta\int_\Omega \Phi \mathrm{d}\bar{\mu}_t^-,$$

where the last equation follows from the fact that W_1 metrizes weak-$*$ convergence and $\Phi \in C(\Omega)$. □

Next, let us define a reasonable discretization of a forward operator as in example 1.11. To this end, fix a discretization $0 = t_0 < \ldots < t_M = 1$ of the time interval $I = [0,1]$ and, for each $j = 0,1,\ldots,M$, let $\phi_j^\delta \in C(I;\mathbb{R}_+)$ be a family of nascent (θ_j,t_j)-delta functions (in particular, we have $\theta_0 = 0$ and $\theta_M = 1$). Define $\widetilde{K} = \widetilde{K}(\delta) : L^2_w(I;\mathscr{M}(\Omega)) \to \mathbb{R}^{L\times(M+1)}$ by

$$(\widetilde{K}\mu)^i_j = \int_I \phi_j^\delta(s)\int_\Omega \Phi^i(x)\mathrm{d}\mu_s(x)\mathrm{d}\mathscr{L}(s), \tag{11}$$

where $\Phi^1,\ldots,\Phi^L : \Omega \to \mathbb{R}$ are Lipschitz and may correspond to some spatial points $x_1,\ldots,x_L \in \Omega$, see example 1.11. Recall that ϕ_j^δ can be interpreted as the temporal blur at t_j (of order δ). Given $\mu \in \mathscr{A}$, we may define matrix $K_0\mu \in \mathbb{R}^{L\times(M+1)}$ by (proposition 4.11)

$$(K_0\mu)^i_j = \lim_{\delta\to 0}(\widetilde{K}\mu)^i_j = \theta_j\int_\Omega \Phi^i \mathrm{d}\bar{\mu}_{t_j}^- + (1-\theta_j)\int_\Omega \Phi^i \mathrm{d}\bar{\mu}_{t_j}.$$

Note that only the evaluation of any càdlàg representative $\bar{\mu}$ of μ is needed and the expression is well-defined in the sense that it does not depend on the value of $\bar{\mu}$ at $t=1$ (because $\theta_M = 1$). Hence, this definition yields a natural discrete forward operator. More precisely, only the vectors $\bar{\mu}^+ = (\bar{\mu}^{0,+},\dots,\bar{\mu}^{M,+}) = (\bar{\mu}_{t_0},\dots,\bar{\mu}_{t_M})$ and $\bar{\mu}^- = (\bar{\mu}^{0,-},\dots,\bar{\mu}^{M,-}) = (\bar{\mu}^-_{t_0},\dots,\bar{\mu}^-_{t_M})$, which may be seen as time samples of a càdlàg curve in $\mathbb{R}_+\mathscr{D}_W(\Omega)$ representing the right and left 'traces', are considered. For any pair of vectors $\bar{\nu} = (\bar{\nu}^+,\bar{\nu}^-) = ((\bar{\nu}^{0,+},\dots,\bar{\nu}^{M,+}),(\bar{\nu}^{0,-},\dots,\bar{\nu}^{M,-})) \in \mathbb{R}_+(\mathscr{P}(\Omega)^{M+1}\times\mathscr{P}(\Omega)^{M+1})$, we set (with a slight abuse of notation)

$$(K_0\bar{\nu})^i_j = \theta_j \int_\Omega \Phi^i \mathrm{d}\bar{\nu}^{j,-} + (1-\theta_j)\int_\Omega \Phi^i \mathrm{d}\bar{\nu}^{j,+}.$$

Similarly, given samples $(\bar{\nu}^+,\bar{\nu}^-)$ of a càdlàg representative of $\nu = \delta_\gamma$ with $\gamma \in BV(I;\Omega)$, we write $\bar{\gamma} = (\bar{\gamma}^+,\bar{\gamma}^-)$ with $\bar{\gamma}^+ = (\bar{\gamma}^{0,+},\dots,\bar{\gamma}^{M,+}) = (\bar{\gamma}_{t_0},\dots,\bar{\gamma}_{t_M}), \bar{\gamma}^- = (\bar{\gamma}^{0,-},\dots,\bar{\gamma}^{M,-}) = (\bar{\gamma}^-_{t_0},\dots,\bar{\gamma}^-_{t_M})$ and set

$$(K_0\bar{\gamma})^i_j = \theta_j\Phi^i(\bar{\gamma}^{j,-}) + (1-\theta_j)\Phi^i(\bar{\gamma}^{j,+}).$$

Note that, in principle, each Φ^i may also be evaluated at values of other reasonable representatives of γ, e.g. representatives whose evaluations at the t_j lie in between $\bar{\gamma}^-_{t_j}$ and $\bar{\gamma}^+_{t_j}$ (recall that Ω is convex). Since we are particularly interested in càdlàg curves, we use the construction from above.

Next, let us apply the above procedure to discretize the functional $\gamma \mapsto a(\gamma)\int_I p^k(\gamma)\mathrm{d}\mathscr{L}$ in the insertion step of algorithm 1, see problem (9). First, we define

$$p(\mu) = -\widetilde{K}_*\mathrm{D}\mathscr{F}(\widetilde{K}\mu) \qquad \text{and discretize} \qquad \langle p(\mu),\nu\rangle_{L^2_w} = -\mathrm{D}\mathscr{F}(\widetilde{K}\mu) : \widetilde{K}\nu$$

for arbitrary $\mu,\nu \in L^2_w(I;\mathscr{M}(\Omega))$, where : denotes the Frobenius inner product on $Y = \mathbb{R}^{L\times(M+1)}$. We see directly that the latter expression depends continuously on the measurements $\widetilde{K}\mu$ and $\widetilde{K}\nu$ (provided that $\mathrm{D}\mathscr{F}$ is continuous, which is satisfied by assumption 4.1). Therefore, in the spirit of our discretization approach via temporal deblurring in the data space, we discretize $\langle p(\mu),\nu\rangle_{L^2_w}$ as

$$\lim_{\delta\to 0}\langle p(\mu),\nu\rangle_{L^2_w}.$$

In particular, if $\mu = \mu^k$ is an iterate of algorithm 1, then we discretize the functional in the insertion step (problem (9)) by (recall that $p^k = p(\mu^k)$)

$$D^k_0(\bar{\gamma}^+,\bar{\gamma}^-) = a_0(\alpha,\beta,\bar{\gamma}^+,\bar{\gamma}^-)\lim_{\delta\to 0}\langle p^k,\delta_\gamma\rangle_{L^2_w}, \tag{12}$$

where $\gamma \in BV(I;\Omega)$ is arbitrary with $\bar{\gamma}^\pm_{t_j} = \bar{\gamma}^{j,\pm}$ and

$$a_0(\alpha,\beta,\bar{\gamma}^+,\bar{\gamma}^-) = \left(\alpha + \beta\sum_{j=0}^{M-1}\left(|\bar{\gamma}^{j,+}-\bar{\gamma}^{j+1,-}| + |\bar{\gamma}^{j,+}-\bar{\gamma}^{j,-}|\right)\right)^{-1} \tag{13}$$

is a discretization of $a(\gamma)$. Indeed, the right-hand side in equation (12) does not depend on the choice of γ (proposition 4.11):

$$D^k_0(\bar{\gamma}^+,\bar{\gamma}^-) = -a_0(\alpha,\beta,\bar{\gamma}^+,\bar{\gamma}^-)\sum_{i=1}^{L}\sum_{j=0}^{M}(\mathrm{D}\mathscr{F}(K_0\mu^k))_{i,j}(\theta_j\Phi^i(\bar{\gamma}^{j,-}) + (1-\theta_j)\Phi^i(\bar{\gamma}^{j,+})).$$

4.3 Implementation details

This section is devoted to the implementation of algorithm 1. We limit ourselves to the one-dimensional case. It is well-known that in FC-GCG algorithms, the numerical bottleneck lies in the efficient solution of the (in general non-convex) insertion step (problem (4)), see, for example, [16, Section 5.1] and [18, Section 1.1]. In particular, this optimization is extremely challenging in higher-dimensional domains. However, we believe that a numerical realization of algorithm 1 in the case $n = 2$ is possible by transferring the approaches in [16, 23]. Further, it is likely that [23] can be applied to speed up the computations in the insertion step.

First, let us specify our (discrete) forward operator. Assume that $n = 1$ and fix points $x_1, \dots, x_L$ in some compact interval $\Omega \subset \mathbb{R}$. Further, pick a discretization $0 = t_0 < \dots < t_M = 1$ of the time interval $I = [0,1]$. Given $C_i, \sigma_i > 0$, we define (truncated) Gaussian kernels $\Phi^i \in C(\Omega)$ (independent of t) by

$$\Phi^i(x) = \frac{C_i}{\sigma_i} e^{\frac{-|x-x_i|^2}{2\sigma_i^2}} \qquad \text{for} \qquad i = 1, \dots, L, \tag{14}$$

where σ_i^2 indicates variance and C_i can be chosen arbitrarily (e.g. such that $\int_\Omega \Phi^i \mathrm{d}\mathscr{L} = 1$). We use these functions in the definition of $\widetilde{K} = \widetilde{K}(\delta)$ (see equation (11)) and define K_0 as in section 4.2. Recall that

$$(K_0 \bar{\gamma})^i_j = \theta_j \Phi^i(\bar{\gamma}^{j,-}) + (1-\theta_j)\Phi^i(\bar{\gamma}^{j,+})$$

whenever $\bar{\gamma} = (\bar{\gamma}^+, \bar{\gamma}^-)$ with $\bar{\gamma}^+ = (\bar{\gamma}^{0,+}, \dots, \bar{\gamma}^{M,+})$ and $\bar{\gamma}^- = (\bar{\gamma}^{0,-}, \dots, \bar{\gamma}^{M,-})$ corresponds to time samples of a càdlàg representative of $\gamma \in BV(I;\Omega)$ respectively $\delta_\gamma \in BV(I;\mathscr{W}_1(\Omega))$. As mentioned in section 4.2, these vectors can be interpreted as the left and right 'traces' of this representative at the discrete time points t_j. In particular, we enforce any reconstructed jump to take place at these points. In the remainder, we take $\theta_j = 0$ for all $j \neq M$ (recall that $\theta_M = 1$), which yields evaluation of Φ^i at the right-hand limits for $t = t_0, \dots, t_{M-1}$ and evaluation at the left-hand limit for $t = t_M = 1$ (which is reasonable by the non-uniqueness of a càdlàg representative of γ at $t = 1$).

Remark 4.12 (Space and time discretization)**.** *In order to reduce the computational cost, we consider the time discretization $t_0, \dots, t_M$. This is standard for such particular algorithms [16, 23]. We stress that the discretization $x_1, \dots, x_L$ corresponds to the definition of $\widetilde{K}$ (respectively K_0). In particular, a spatial numerical discretization is not needed because algorithm 1 is grid-free (thus it allows for super-resolution).*

We consider the fidelity $\mathscr{F} : \mathbb{R}^{L\times(M+1)} \to \mathbb{R}_+$ defined by

$$\mathscr{F}_f(y) = \frac{1}{2(M+1)} \|y - f\|_F^2,$$

where $f \in \mathbb{R}^{L\times(M+1)}$ represents given reference data and $\|\cdot\|_F$ denotes the Frobenius norm.

Example 4.13 (Formula for $D_0^k(\bar{\gamma}^+, \bar{\gamma}^-)$)**.** *In the above setting, we have (proposition 4.11)*

$$\begin{aligned} D_0^k(\bar{\gamma}^+, \bar{\gamma}^-) = &-\frac{a_0(\alpha,\beta,\bar{\gamma}^+,\bar{\gamma}^-)}{M+1} \sum_{j=0}^{M-1} \sum_{i_1=1}^{L} \Phi^{i_1}(\bar{\gamma}^{j,+}) \left(\sum_{\ell_1=1}^{N} \lambda_{\ell_1}^k \Phi^{i_1}((\bar{\gamma}_\ell^k)^{j,+}) - f_j^i \right) \\ &- \frac{a_0(\alpha,\beta,\bar{\gamma}^+,\bar{\gamma}^-)}{M+1} \sum_{i_2=1}^{L} \Phi^{i_2}(\bar{\gamma}^{M,-}) \left(\sum_{\ell_2=1}^{N} \lambda_{\ell_2}^k \Phi^{i_2}((\bar{\gamma}_\ell^k)^{M,-}) - f_M^i \right), \end{aligned}$$

where $\mu^k = \sum_{\ell=1}^N \lambda_\ell^k \delta_{\gamma_\ell^k}$ is an iterate of algorithm 1.

In order to maximize $(\bar{\gamma}^+,\bar{\gamma}^-)\mapsto D_0^k(\bar{\gamma}^+,\bar{\gamma}^-)$, we adopt the multi-start gradient descent approach in [16, Section 5.1]. First, we sample from the uniform distribution in Ω two $(M+1)$-tuples $\bar{\gamma}_0^+ = (\bar{\gamma}_0^{0,+},\dots,\bar{\gamma}_0^{M,+})$ and $\bar{\gamma}_0^- = (\bar{\gamma}_0^{0,-},\dots,\bar{\gamma}_0^{M,-})$. Then, using $\bar{\gamma}_0^+$ and $\bar{\gamma}_0^-$ as initializations, we run a gradient ascent algorithm[5] to maximize functional D_0^k obtaining as output vectors $\bar{\gamma}_{0,*}^+$, $\bar{\gamma}_{0,*}^-$. We repeat this procedure for a fixed amount of random initializations $q=0,...,Q$. Then we store the vectors $\bar{\gamma}_{q_{max},*}^+$, $\bar{\gamma}_{q_{max},*}^-$ which yield the maximum value of $q\mapsto D_0^k(\bar{\gamma}_{q,*}^+,\bar{\gamma}_{q,*}^-)$. These vectors will be the output of the insertion step. In order to balance between accuracy and efficiency, it is crucial to carefully choose the number of initializations Q of the gradient ascent algorithm. In all the experiments in section 4.4, we take $Q=150$.

4.4 Numerical experiments

In the following experiments, we use $\Omega=[0,5]$ and equidistant time discretization $0=t_0<\ldots<t_M=1$ with $M=30$. We employ the forward operator $\widetilde{K}$ given by equations (11) and (14) with equidistant points $0=x_1<\ldots<x_{100}=5$ in Ω and $C_i\equiv\frac{1}{\sqrt{2\pi}}, \sigma_i^2\equiv 0.02$.

Three curves, no noise: In the first numerical experiment, we aim to reconstruct the ground truth $\bar{\mu}^\dagger\in\mathscr{D}_E(\Omega)$ given by

$$\bar{\mu}^\dagger=\delta_{\bar{\gamma}_1^\dagger}+\delta_{\bar{\gamma}_2^\dagger}+\delta_{\bar{\gamma}_3^\dagger}, \tag{15}$$

where

$$\bar{\gamma}_1^\dagger(t)=t+3.5,\qquad \bar{\gamma}_2^\dagger(t)=\sqrt{t}+2.5,\qquad\text{and}\qquad \bar{\gamma}_3^\dagger(t)=\begin{cases}1+t^2 & \text{if } t<0.5,\\ 2+t^2 & \text{if } t\geq 0.5.\end{cases} \tag{16}$$

A discretization of the ground truth $\bar{\mu}^\dagger$ is given in figure 4.

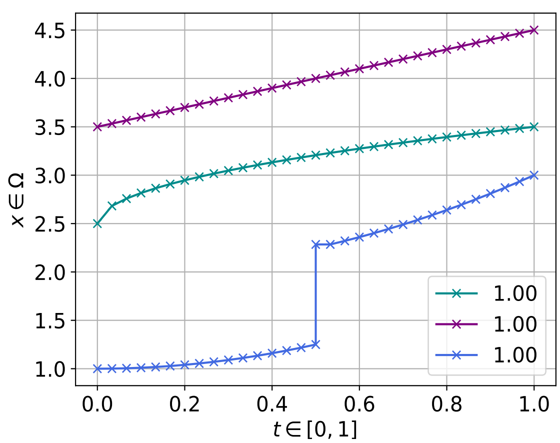
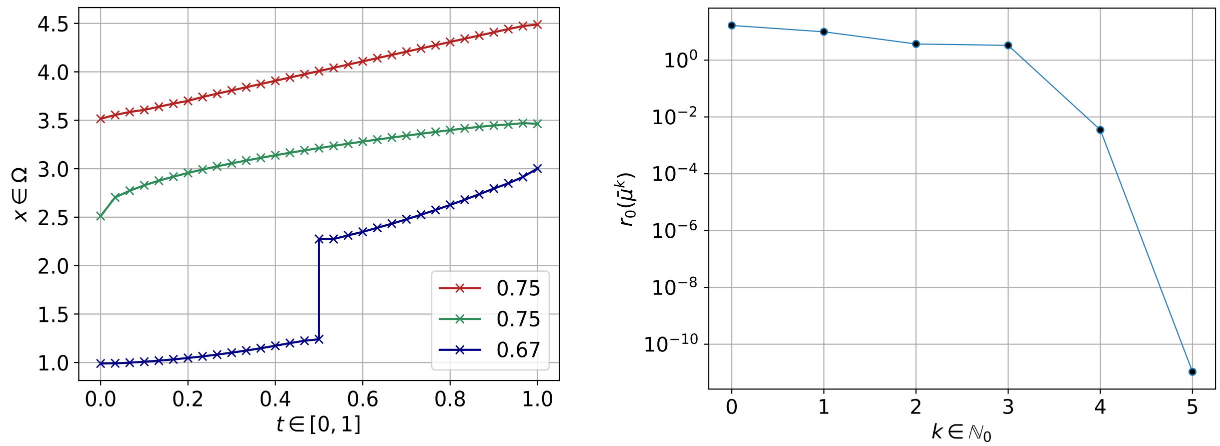


Figure 4: Discretization of the ground truth $\bar{\mu}^\dagger$ defined by equations (15) and (16). The legend shows the weights associated with the càdlàg curves $\delta_{\bar{\gamma}_i^\dagger}$.

We consider the reference data $f=K_0\bar{\mu}^\dagger\in\mathbb{R}^{L\times(M+1)}$ (recall that K_0 equals the pointwise limit of $\widetilde{K}=\widetilde{K}(\delta)$) and choose regularization parameters $\alpha=5$ and $\beta=2$. Our reconstruction is depicted in figure 5a. Figure 5b shows the convergence of the residuals (equation (10)). Note that, since min $\mathscr{J}_{\alpha,\beta}$

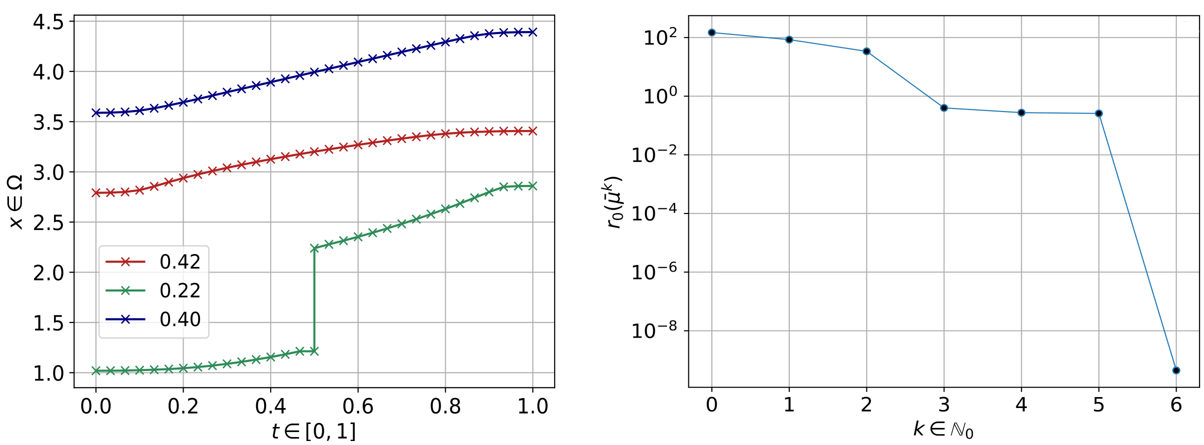
[5]The Euclidean norm in equation (13) is approximated by $\eta_\varepsilon(z)=\sqrt{|z|^2+\varepsilon}$.



(a) $\alpha = 5$ and $\beta = 2$. (b) Convergence of residuals (log scale).

Figure 5: Reconstruction of the discretized ground truth from figure 4. Note that it is faithful to the ground truth in the sense that it recovers the jump of $\bar{\gamma}_3^{\dagger}$. The effect of regularization is observable in the attenuated weights and decreased variation of each curve (in particular, close to $t = 0$ and $t = 1$).

is unknown, we approximate the residuals by evaluating our discretization[6] of $\mathscr{J}_{\alpha,\beta}$ at the last iterate returned by algorithm 1. The same idea was used in [16]. The corresponding residuals are denoted by $r_0(\bar{\mu}^k)$. To highlight the effect of our regularizer $\mathscr{R}_{\alpha,\beta}$, we perform another reconstruction using $\alpha = 12$ and $\beta = 5$. It is shown in figure 6a (next to the residuals $k \mapsto r_0(\bar{\mu}^k)$ in figure 6b).

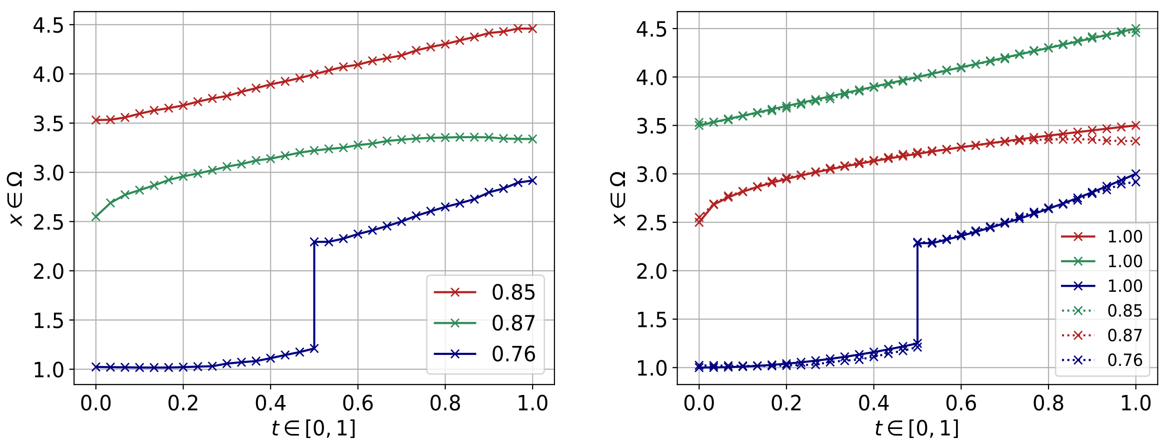


(a) $\alpha = 12$ and $\beta = 5$. (b) Convergence of residuals (log scale).

Figure 6: The larger value of β reduces the variation of each of the curves — close to $t = 0$ and $t = 1$ the Diracs move with small velocity. The weights of the reconstructed curves are further scaled down because of the larger value of α.

[6]It is naturally defined in the spirit of section 4.2 using K_0 and an approximation of the essential variation as in a_0, see equation (13).

Three curves with noise: In the second experiment, we consider the same ground truth as in the previous example (defined by equations (15) and (16)). Moreover, we perturb the measurements with Gaussian noise. In particular, we add a matrix $\mathscr{N} \in \mathbb{R}^{L\times(M+1)}$ whose entries are realizations of normally distributed (with zero mean and standard deviation equal to 0.2) random variables. Hence, the measurement $K_0\bar{\mu}^\dagger + \mathscr{N}$ is used. Further, we take $\alpha = 5$ and $\beta = 3$. In general, we expect that our regularization allows for a reconstruction which is stable with respect to (sufficiently small) additive noise (since our regularizer is a natural extension of 'static' regularizers with similar properties). This is also due to remark 1.6 and the choice of our fidelity term. The reconstruction is illustrated in figure 7.



(a) $\alpha = 5$ and $\beta = 3$. (b) Superimposition ground truth & reconstruction.

Figure 7: Reconstruction of $\bar{\mu}^\dagger$ from $K_0\bar{\mu}^\dagger + \mathscr{N}$. The superimposition of ground truth (depicted with a solid line) and reconstruction highlights the regularization bias. As expected, the reconstruction is not strongly affected by the noise. Similar regularization effects as in the previous experiments can be observed.

Crossing curves: In this experiment, we consider the problem of reconstructing two curves which cross each other at time $t = 0.5$. More specifically, we use the ground truth given by (see figure 8a)

$$\bar{\mu}^\dagger = \delta_{\bar{\gamma}_1^\dagger} + \delta_{\bar{\gamma}_2^\dagger}$$

where

$$\bar{\gamma}_1^\dagger(t) = 1 + 3t \qquad \text{and} \qquad \bar{\gamma}_2^\dagger(t) = 4 - 3t.$$

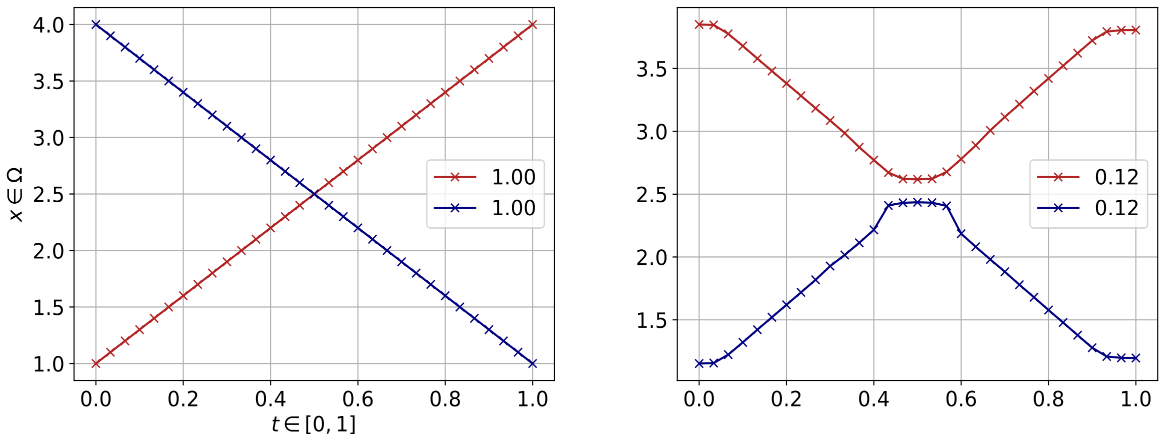
Again, we compute the reference data as $f = K_0\bar{\mu}^\dagger \in \mathbb{R}^{L\times(M+1)}$. The reconstruction is shown in figure 8b. A similar effect to the one highlighted in [16, Section 6.2.3] (where a Wasserstein-2-type regularization is considered) can be observed.

Non-sparse ground truths: Finally, we reconstruct diffuse ground truths which cannot be written as linear combinations of Diracs in elements of $\mathscr{D}_E(\Omega)$. More precisely, we take $\bar{\mu}^\dagger, \bar{\nu}^\dagger \in \mathscr{D}_W(\Omega)$ defined by

$$\bar{\mu}_t^\dagger = \mathscr{L} \llcorner [1+t, 4-t] \qquad \text{and} \qquad \bar{\nu}_t^\dagger = \mathscr{L} \llcorner [\zeta_1(t), \zeta_2(t)]$$

with

$$\zeta_1(t) = \begin{cases} 1+t & \text{if } t < 0.5, \\ 2+t & \text{if } t \geq 0.5 \end{cases} \qquad \text{and} \qquad \zeta_2(t) = \begin{cases} 2+t & \text{if } t < 0.5, \\ 3+t & \text{if } t \geq 0.5. \end{cases}$$

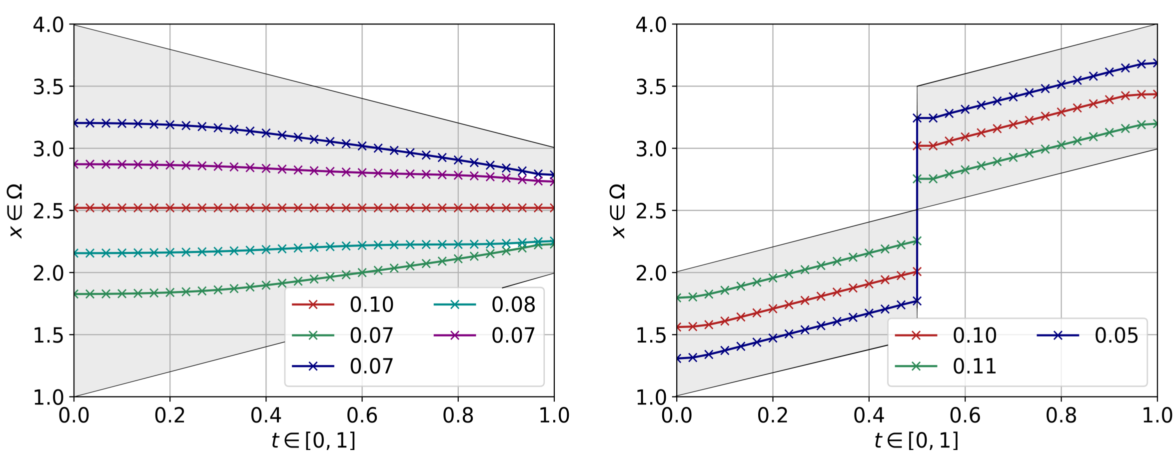


(a) Ground truth.

(b) Reconstruction for $\alpha = 13$ and $\beta = 5$.

Figure 8: Reconstruction of two crossing curves. Our algorithm is unable to identify the correct trajectories due to the regularization (which encourages small variation).

We define $f = K_0\bar{\mu}^\dagger, g = K_0\bar{\nu}^\dagger \in \mathbb{R}^{L\times(M+1)}$ as the corresponding reference data. Note that $\bar{\mu}^\dagger$ is absolutely continuous while $\bar{\nu}^\dagger$ jumps at $t = 0.5$. We use regularization parameters $\alpha = 3$ and $\beta = 2$ for the reconstruction of $\bar{\mu}^\dagger$ and $\alpha = 5$ and $\beta = 2$ for the reconstruction of $\bar{\nu}^\dagger$. The results are shown in figure 9. Recall that algorithm 1 yields sparse reconstructions despite our choice of the ground truths.



(a) Reconstruction of $\bar{\mu}^\dagger$ ($\alpha = 3, \beta = 2$).

(b) Reconstruction of $\bar{\nu}^\dagger$ ($\alpha = 5$, $\beta = 2$).

Figure 9: Reconstruction of diffuse ground truths. As expected, curve $\bar{\mu}^\dagger$ is approximated by only using absolutely continuous curves while $\bar{\nu}^\dagger$ is approximated by curves with jumps.

5 Acknowledgements

M.C.'s research was supported by the NWO-M1 grant Curve Ensemble Gradient Descents for Sparse Dynamic Problems (Grant Number OCENW.M.22.302). J.L.'s research was supported by the JSPS Postdoctoral Fellowship for Research in Japan and the KAKENHI Grant-in-Aid for Scientific Research (Grant Number JP24KF0215) awarded by the Japan Society for the Promotion of Science. J.L. would like to thank Carola-Bibiane Schönlieb for the hosting of a NoMADS secondment at the University of Cambridge, funded by the European Union's Horizon 2020 research and innovation programme (Grant Number 777826), during which this research collaboration began. J.L.'s research was also supported by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under the priority program SPP 1962, grant WI 4654/1-1, and under Germany's Excellence Strategy EXC 2044-390685587, Mathematics Münster: Dynamics-Geometry-Structure.

References

[1] Ehsan Abedi, Zhenhao Li, and Timo Schultz. "Absolutely continuous and BV-curves in 1-Wasserstein spaces". In: *Calc. Var. Partial Differential Equations* 63.1 (2024), Paper No. 16, 34. ISSN: 0944-2669,1432-0835. DOI: 10.1007/s00526-023-02616-1.

[2] Luigi Ambrosio. "Lecture notes on optimal transport problems". In: *Mathematical aspects of evolving interfaces (Funchal, 2000)*. Vol. 1812. Lecture Notes in Math. Springer, Berlin, 2003, pp. 1–52. ISBN: 3-540-14033-6. DOI: 10.1007/978-3-540-39189-0_1.

[3] Luigi Ambrosio. "Metric space valued functions of bounded variation". In: *Ann. Scuola Norm. Sup. Pisa Cl. Sci. (4)* 17.3 (1990), pp. 439–478. ISSN: 0391-173X,2036-2145. URL: http://www.numdam.org/item?id=ASNSP_1990_4_17_3_439_0.

[4] Luigi Ambrosio, Nicola Gigli, and Giuseppe Savaré. *Gradient flows in metric spaces and in the space of probability measures*. Lectures in Mathematics ETH Zürich. Birkhäuser Verlag, Basel, 2005, pp. viii+333. ISBN: 978-3-7643-2428-5; 3-7643-2428-7.

[5] Baran Tan Bacinoglu, Yin Sun, and Elif Uysal. "On the Trackability of Stochastic Processes Based on Causal Information". In: *2020 IEEE International Symposium on Information Theory (ISIT)*. Los Angeles, CA, USA: IEEE Press, 2020, pp. 2228–2233. DOI: 10.1109/ISIT44484.2020.9174285.

[6] Jean-David Benamou and Yann Brenier. "A computational fluid mechanics solution to the Monge-Kantorovich mass transfer problem". In: *Numer. Math.* 84.3 (2000), pp. 375–393. ISSN: 0029-599X,0945-3245. DOI: 10.1007/s002110050002.

[7] Marc Bernot, Vicent Caselles, and Jean-Michel Morel. "Traffic plans". In: *Publ. Mat.* 49.2 (2005), pp. 417–451. ISSN: 0214-1493,2014-4350. URL: https://eudml.org/doc/41575.

[8] Patrick Billingsley. *Convergence of probability measures*. Second. Wiley Series in Probability and Statistics: Probability and Statistics. A Wiley-Interscience Publication. John Wiley & Sons, Inc., New York, 1999, pp. x+277. ISBN: 0-471-19745-9. DOI: 10.1002/9780470316962.

[9] Claire Boyer et al. "On representer theorems and convex regularization". In: *SIAM J. Optim.* 29.2 (2019), pp. 1260–1281. ISSN: 1052-6234,1095-7189. DOI: 10.1137/18M1200750.

[10] Leif Boysen, Sophie Bruns, and Axel Munk. “Jump estimation in inverse regression”. In: *Electron. J. Stat.* 3 (2009), pp. 1322–1359. ISSN: 1935-7524. DOI: 10.1214/08-EJS204.

[11] Alessio Brancolini. *Optimization problems for transportation networks*. Dissertation. Scuola Normale Superiore, 2005.

[12] Alessio Brancolini and Giuseppe Buttazzo. “Optimal networks for mass transportation problems”. In: *ESAIM Control Optim. Calc. Var.* 11.1 (2005), pp. 88–101. ISSN: 1292-8119,1262-3377. DOI: 10.1051/cocv:2004032.

[13] Alessio Brancolini and Benedikt Wirth. “General transport problems with branched minimizers as functionals of 1-currents with prescribed boundary”. In: *Calc. Var. Partial Differential Equations* 57.3 (2018), Paper No. 82, 39. ISSN: 0944-2669,1432-0835. DOI: 10.1007/s00526-018-1364-4.

[14] Kristian Bredies and Marcello Carioni. “Sparsity of solutions for variational inverse problems with finite-dimensional data”. In: *Calc. Var. Partial Differential Equations* 59.1 (2020), Paper No. 14, 26. ISSN: 0944-2669,1432-0835. DOI: 10.1007/s00526-019-1658-1.

[15] Kristian Bredies, Marcello Carioni, and Silvio Fanzon. “A superposition principle for the inhomogeneous continuity equation with Hellinger-Kantorovich-regular coefficients”. In: *Comm. Partial Differential Equations* 47.10 (2022), pp. 2023–2069. ISSN: 0360-5302,1532-4133. DOI: 10.1080/03605302.2022.2109172.

[16] Kristian Bredies, Marcello Carioni, Silvio Fanzon, and Francisco Romero. “A generalized conditional gradient method for dynamic inverse problems with optimal transport regularization”. In: *Found. Comput. Math.* 23.3 (2023), pp. 833–898. ISSN: 1615-3375,1615-3383. DOI: 10.1007/s10208-022-09561-z.

[17] Kristian Bredies, Marcello Carioni, Silvio Fanzon, and Francisco Romero. “On the extremal points of the ball of the Benamou-Brenier energy”. In: *Bull. Lond. Math. Soc.* 53.5 (2021), pp. 1436–1452. ISSN: 0024-6093,1469-2120. DOI: 10.1112/blms.12509.

[18] Kristian Bredies, Marcello Carioni, Silvio Fanzon, and Daniel Walter. “Asymptotic linear convergence of fully-corrective generalized conditional gradient methods”. In: *Math. Program.* 205.1-2 (2024), pp. 135–202. ISSN: 0025-5610,1436-4646. DOI: 10.1007/s10107-023-01975-z.

[19] Kristian Bredies and Silvio Fanzon. “An optimal transport approach for solving dynamic inverse problems in spaces of measures”. In: *ESAIM Math. Model. Numer. Anal.* 54.6 (2020), pp. 2351–2382. ISSN: 2822-7840,2804-7214. DOI: 10.1051/m2an/2020056.

[20] Kristian Bredies and Hanna Katriina Pikkarainen. “Inverse problems in spaces of measures”. In: *ESAIM Control Optim. Calc. Var.* 19.1 (2013), pp. 190–218. ISSN: 1292-8119,1262-3377. DOI: 10.1051/cocv/2011205.

[21] Giuseppe Buttazzo, Aldo Pratelli, Sergio Solimini, and Eugene Stepanov. *Optimal urban networks via mass transportation*. Vol. 1961. Lecture Notes in Mathematics. Springer-Verlag, Berlin, 2009, pp. x+150. ISBN: 978-3-540-85798-3. DOI: 10.1007/978-3-540-85799-0.

[22] Mohammad Dawood et al. “A Continuity Equation Based Optical Flow Method for Cardiac Motion Correction in 3D PET Data”. In: *Medical Imaging and Augmented Reality*. Ed. by Hongen Liao, P. J. Edwards, Xiaochuan Pan, Yong Fan, and Guang-Zhong Yang. Berlin, Heidelberg: Springer Berlin Heidelberg, 2010, pp. 88–97. ISBN: 978-3-642-15699-1.

[23] Vincent Duval and Robert Tovey. "Dynamical programming for off-the-grid dynamic inverse problems". In: *ESAIM Control Optim. Calc. Var.* 30 (2024), Paper No. 7, 40. ISSN: 1292-8119,1262-3377. DOI: 10.1051/cocv/2023085.

[24] R. E. Edwards. *Functional analysis. Theory and applications*. Holt, Rinehart and Winston, New York-Toronto-London, 1965, pp. xiii+781.

[25] Yicheng Kang, Xiaodong Gong, Jiti Gao, and Peihua Qiu. "Errors-in-variables jump regression using local clustering". In: *Stat. Med.* 38.19 (2019), pp. 3642–3655. ISSN: 0277-6715,1097-0258. DOI: 10.1002/sim.8205.

[26] V. L. Klee Jr. "Extremal structure of convex sets". In: *Arch. Math. (Basel)* 8 (1957), pp. 234–240. ISSN: 0003-889X,1420-8938. DOI: 10.1007/BF01899998.

[27] Matti Lassas and Samuli Siltanen. "Can one use total variation prior for edge-preserving Bayesian inversion?" In: *Inverse Problems* 20.5 (2004), pp. 1537–1563. ISSN: 0266-5611,1361-6420. DOI: 10.1088/0266-5611/20/5/013.

[28] Bastien Laville, Laure Blanc-Féraud, and Gilles Aubert. "Off-the-grid curve reconstruction through divergence regularization: an extreme point result". In: *SIAM J. Imaging Sci.* 16.2 (2023), pp. 867–885. ISSN: 1936-4954. DOI: 10.1137/22M1494373.

[29] Xiaobai Liu, Donovan Lo, and Chau Thuan. "Unsupervised Learning based Jump-Diffusion Process for Object Tracking in Video Surveillance". In: *Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence, IJCAI-18*. International Joint Conferences on Artificial Intelligence Organization, July 2018, pp. 5060–5066. DOI: 10.24963/ijcai.2018/702.

[30] Julius Lohmann. *On the branched transport problem. Reformulation as geometry optimization and calibration using convex duality*. Dissertation. University of Münster, 2023.

[31] Julius Lohmann, Bernhard Schmitzer, and Benedikt Wirth. "Duality in branched transport and urban planning". In: *Appl. Math. Optim.* 86.3 (2022), Paper No. 45, 32. ISSN: 0095-4616,1432-0606. DOI: 10.1007/s00245-022-09927-3.

[32] Julius Lohmann, Bernhard Schmitzer, and Benedikt Wirth. *Formulas for the h-mass on 1-currents with coefficients in* $\mathbb{R}^m$. 2024. arXiv: 2407.10158 [math.OC].

[33] Julius Lohmann, Bernhard Schmitzer, and Benedikt Wirth. "Formulation of branched transport as geometry optimization". In: *J. Math. Pures Appl. (9)* 163 (2022), pp. 739–779. ISSN: 0021-7824,1776-3371. DOI: 10.1016/j.matpur.2022.05.021.

[34] F. Maddalena, S. Solimini, and J.-M. Morel. "A variational model of irrigation patterns". In: *Interfaces Free Bound.* 5.4 (2003), pp. 391–415. ISSN: 1463-9963,1463-9971. DOI: 10.4171/IFB/85.

[35] Francesco Maddalena and Sergio Solimini. "Synchronic and asynchronic descriptions of irrigation problems". In: *Adv. Nonlinear Stud.* 13.3 (2013), pp. 583–623. ISSN: 1536-1365,2169-0375. DOI: 10.1515/ans-2013-0303.

[36] A. Marchese, A. Massaccesi, S. Stuvard, and R. Tione. "A multi-material transport problem with arbitrary marginals". In: *Calc. Var. Partial Differential Equations* 60.3 (2021), Paper No. 88, 49. ISSN: 0944-2669,1432-0835. DOI: 10.1007/s00526-021-01967-x.

[37] Andrea Marchese, Annalisa Massaccesi, and Riccardo Tione. “A multimaterial transport problem and its convex relaxation via rectifiable G-currents”. In: *SIAM J. Math. Anal.* 51.3 (2019), pp. 1965–1998. ISSN: 0036-1410,1095-7154. DOI: 10.1137/17M1162858.

[38] Olga Minevich. *Adaptive numerical methods for optimal and branched transport problems*. Dissertation. Georg August University of Göttingen, 2023.

[39] Jennifer L. Mueller and Samuli Siltanen. *Linear and nonlinear inverse problems with practical applications*. Vol. 10. Computational Science & Engineering. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 2012, pp. xiv+351. ISBN: 978-1-611972-33-7. DOI: 10.1137/1.9781611972344.

[40] Philip E. Protter. *Stochastic integration and differential equations*. Second. Vol. 21. Stochastic Modelling and Applied Probability. Corrected third printing. Springer-Verlag, Berlin, 2005, pp. xiv+419. ISBN: 3-540-00313-4. DOI: 10.1007/978-3-662-10061-5.

[41] R. Tyrrell Rockafellar. *Convex analysis*. Vol. No. 28. Princeton Mathematical Series. Princeton University Press, Princeton, NJ, 1970, pp. xviii+451.

[42] Leonid I. Rudin, Stanley Osher, and Emad Fatemi. “Nonlinear total variation based noise removal algorithms”. In: vol. 60. 1-4. Experimental mathematics: computational issues in nonlinear science (Los Alamos, NM, 1991). 1992, pp. 259–268. DOI: 10.1016/0167-2789(92)90242-F.

[43] Walter Rudin. *Functional analysis*. Second. International Series in Pure and Applied Mathematics. McGraw-Hill, Inc., New York, 1991, pp. xviii+424. ISBN: 0-07-054236-8.

[44] Filippo Santambrogio. *Optimal transport for applied mathematicians*. Vol. 87. Progress in Nonlinear Differential Equations and their Applications. Calculus of variations, PDEs, and modeling. Birkhäuser/Springer, Cham, 2015, pp. xxvii+353. ISBN: 978-3-319-20827-5; 978-3-319-20828-2. DOI: 10.1007/978-3-319-20828-2.

[45] Louis-Philippe Saumier, Boualem Khouider, and Martial Agueh. “Optimal transport for particle image velocimetry: real data and postprocessing algorithms”. In: *SIAM J. Appl. Math.* 75.6 (2015), pp. 2495–2514. ISSN: 0036-1399,1095-712X. DOI: 10.1137/140988814.

[46] U. Schmitt and A. K. Louis. “Efficient algorithms for the regularization of dynamic inverse problems. I. Theory”. In: *Inverse Problems* 18.3 (2002), pp. 645–658. ISSN: 0266-5611,1361-6420. DOI: 10.1088/0266-5611/18/3/308.

[47] U. Schmitt, A. K. Louis, C. Wolters, and M. Vauhkonen. “Efficient algorithms for the regularization of dynamic inverse problems. II. Applications”. In: *Inverse Problems* 18.3 (2002), pp. 659–676. ISSN: 0266-5611,1361-6420. DOI: 10.1088/0266-5611/18/3/309.

[48] Bernhard Schmitzer, Klaus P. Schäfers, and Benedikt Wirth. “Dynamic Cell Imaging in PET With Optimal Transport Regularization”. In: *IEEE Trans Med Imaging* 39.5 (2020), pp. 1626–1635. DOI: 10.1109/TMI.2019.2953773.

[49] Bernhard Schmitzer and Benedikt Wirth. “Dynamic models of Wasserstein-1-type unbalanced transport”. In: *ESAIM Control Optim. Calc. Var.* 25 (2019), Paper No. 23, 54. ISSN: 1292-8119,1262-3377. DOI: 10.1051/cocv/2018017.

[50] S. K. Smirnov. “Decomposition of solenoidal vector charges into elementary solenoids, and the structure of normal one-dimensional flows”. In: *Algebra i Analiz* 5.4 (1993), pp. 206–238. ISSN: 0234-0852.

[51] Elias M. Stein and Rami Shakarchi. *Real analysis*. Vol. 3. Princeton Lectures in Analysis. Measure theory, integration, and Hilbert spaces. Princeton University Press, Princeton, NJ, 2005, pp. xx+402. ISBN: 0-691-11386-6.

[52] Robert Tibshirani. "Regression shrinkage and selection via the lasso". In: *J. Roy. Statist. Soc. Ser. B* 58.1 (1996), pp. 267–288. ISSN: 0035-9246. DOI: 10.1111/j.2517-6161.1996.tb02080.x.

[53] Tuomo Valkonen. "Proximal methods for point source localisation". In: *J. Nonsmooth Anal. Optim.* 4 (2023), Paper No. 10433, 36. ISSN: 2700-7448.

[54] Cédric Villani. *Optimal transport*. Vol. 338. Grundlehren der mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]. Old and new. Springer-Verlag, Berlin, 2009, pp. xxii+973. ISBN: 978-3-540-71049-3. DOI: 10.1007/978-3-540-71050-9.

[55] J. Warga. *Optimal control of differential and functional equations*. Academic Press, New York-London, 1972, pp. xiii+531.