A Generalized Conditional Gradient Method for Dynamic Inverse Problems with Optimal Transport Regularization

Kristian Bredies[1] · Marcello Carioni[2] · Silvio Fanzon[1] ·
Francisco Romero[1]

Abstract

We develop a dynamic generalized conditional gradient method (DGCG) for dynamic inverse problems with optimal transport regularization. We consider the framework introduced in Bredies and Fanzon (ESAIM: M2AN 54:2351–2382, 2020), where the objective functional is comprised of a fidelity term, penalizing the pointwise in time discrepancy between the observation and the unknown in time-varying Hilbert spaces, and a regularizer keeping track of the dynamics, given by the Benamou–Brenier energy constrained via the homogeneous continuity equation. Employing the characterization of the extremal points of the Benamou–Brenier energy (Bredies et al. in Bull Lond Math Soc 53(5):1436–1452, 2021), we define the *atoms* of the problem as measures concentrated on absolutely continuous curves in the domain. We propose a dynamic generalization of a conditional gradient method that consists of iteratively adding suitably chosen *atoms* to the current sparse iterate, and subsequently optimizing the coefficients in the resulting linear combination. We prove that the method converges with a sublinear rate to a minimizer of the objective functional. Additionally, we propose heuristic strategies and acceleration steps that allow to implement the algorithm

---

Communicated by Hans Munthe Kaas.

✉ Kristian Bredies
kristian.bredies@uni-graz.at

Marcello Carioni
mc2250@maths.cam.ac.uk

Silvio Fanzon
silvio.fanzon@uni-graz.at

Francisco Romero
francisco.romero-hinrichsen@uni-graz.at

1 Institute of Mathematics and Scientific Computing, University of Graz, Heinrichstraße 36, 8010
   Graz, Austria

2 Department of Applied Mathematics and Theoretical Physics, University of Cambridge, Wilberforce
   Road, Cambridge CB3 0WA, United Kingdom

🖄 Springer

efficiently. Finally, we provide numerical examples that demonstrate the effectiveness of our algorithm and model in reconstructing heavily undersampled dynamic data, together with the presence of noise.

1 Introduction

The aim of this paper is to develop a dynamic generalized condition gradient method (DGCG) to numerically compute solutions of ill-posed dynamic inverse problems regularized with optimal transport energies. The code is openly available on GitHub.[1]

   Lately, several approaches have been proposed to tackle dynamic inverse problems [42, 57, 58, 61, 67], all of which take advantage of redundancies in the data, allowing to stabilize reconstructions, both in the presence of noise or undersampling. A common challenge faced in such time-dependent approaches is understanding how to properly connect, or relate, the time-neighboring datapoints, in a way that the reconstructed object follows a presumed dynamic. In this paper, we address such issue by means of dynamic optimal transport. A wide range of applications can benefit from motion-aware approaches. In particular, the employment of dynamic reconstruction methods represents one of the latest key mathematical advances in medical imaging. For instance, magnetic resonance imaging (MRI) [48, 50, 56] and computed tomography (CT) [9, 21, 32] methods allow dynamic modalities in which the time-dependent data are further undersampled to reach high temporal sampling rates; these are required to resolve organ motion, such as the beating heart or the breathing lung. A more accurate reconstruction of the image, and of the underlying dynamics, would yield valuable diagnostic information.

1.1 Setting and Existing Approaches

Recently, it has been proposed to regularize dynamic inverse problems using dynamic optimal transport energies both in a balanced and unbalanced context [15, 59, 60], with the goal of efficiently reconstructing time-dependent Radon measures. Such regularization choice is natural: optimal transport energies incorporate information about time correlations present in the data, and are thus favoring a more stable reconstruction. Optimal transport theory was originally developed to find the most efficient way to move mass from a probability measure ρ_0 to another one ρ_1, with respect to a given cost [44, 54]. More recently, Benamou and Brenier [8] showed that the optimal transport map can be computed by solving

---

[1] https://github.com/panchoop/DGCG_algorithm/.

$$\min_{(\rho_t, v_t)} \frac{1}{2} \int_0^1 \int_{\overline{\Omega}} |v_t(x)|^2 \, d\rho_t(x) \, dt \quad \text{s.t.} \quad \partial_t \rho_t + \operatorname{div}(v_t \rho_t) = 0, \tag{1}$$

where $t \mapsto \rho_t$ is a curve of probability measures on the closure of a bounded domain $\Omega \subset \mathbb{R}^d$, v_t is a time-dependent vector field advecting the mass, and the continuity equation is intended in the sense of distributions with initial data ρ_0 and final data ρ_1. Notably, the quantity at (1), named Benamou–Brenier energy, admits an equivalent convex reformulation. Specifically, consider the space of bounded Borel measures $\mathcal{M} := \mathcal{M}(X) \times \mathcal{M}(X; \mathbb{R}^d)$, $X := (0, 1) \times \overline{\Omega}$, and define the convex energy $B :$ $\mathcal{M} \to [0, +\infty]$ by setting

$$B(\rho, m) := \frac{1}{2} \int_0^1 \int_{\overline{\Omega}} \left| \frac{dm}{d\rho}(t, x) \right|^2 \, d\rho(t, x),$$

if $\rho \geq 0$ and $m \ll \rho$, and $B := +\infty$ otherwise. Then, (1) is equivalent to minimizing B under the linear constraint $\partial_t \rho + \operatorname{div} m = 0$. Such convex reformulation can be employed as a regularizer for dynamic inverse problems, where instead of fixing initial and final data, a fidelity term is added to measure the discrepancy between the unknown and the observation at each time instant, as proposed in [15]. There the authors consider the dynamic inverse problem of finding a curve of measures $t \mapsto \rho_t$, with $\rho_t \in \mathcal{M}(\overline{\Omega})$, such that

$$K_t^* \rho_t = f_t \quad \text{for a.e. } t \in (0, 1), \tag{2}$$

where $f_t \in H_t$ is some given data, $\{H_t\}$ is a family of Hilbert spaces and $K_t^* :$ $\mathcal{M}(\overline{\Omega}) \to H_t$ are linear continuous observation operators. The problem at (2) is then regularized via the minimization problem

$$\min_{(\rho, m) \in \mathcal{M}} \frac{1}{2} \int_0^1 \left\| K_t^* \rho_t - f_t \right\|_{H_t}^2 \, dt + \beta B(\rho, m) + \alpha \left\| \rho \right\|_{\mathcal{M}(X)} \quad \text{s.t.} \quad \partial_t \rho + \operatorname{div} m = 0, \tag{3}$$

where $\|\rho\|_{\mathcal{M}(X)}$ denotes the total variation norm of the measure $\rho := dt \otimes \rho_t$, and $\alpha, \beta > 0$ are regularization parameters. Notice that any curve $t \mapsto \rho_t$ having finite Benamou–Brenier energy and satisfying the continuity equation constraint must have constant mass (see Lemma 9). As a consequence, the regularization (3) is especially suited to reconstruct motions where preservation of mass is expected. We point our that such formulation is remarkably flexible, as the measurements spaces and measurement operators are allowed to be very general. In this way, one could model, for example, undersampled acquisition strategies in medical imaging, particularly MRI [15].

The aim of this paper is to design a numerical algorithm to solve (3). The main difficulties arise due to the non-reflexivity of measure spaces. Even in the static case, solving the classical LASSO problem [62] in the space of bounded Borel measures (known as BLASSO [29]), i.e.,

$$\inf_{\rho \in \mathcal{M}(\Omega)} \frac{1}{2} \|K\rho - f\|_Y^2 + \alpha \|\rho\|_{\mathcal{M}(\Omega)}, \tag{4}$$

for a Hilbert space Y and a linear continuous operator $K : \mathcal{M}(\Omega) \to Y$ has proven to be challenging. Usual strategies to tackle (4) numerically often rely on the discretization of the domain [26, 28, 63]; however, grid-based methods are known to be affected by theoretical and practical flaws such as high computational costs and the presence of mesh-dependent artifacts in the reconstruction. The mentioned drawbacks have motivated algorithms that do not rely on domain discretization, but optimize directly on the space of measures. One class of such algorithms, first introduced in [18] and subsequently developed in different directions [10, 30, 40, 52], are named generalized conditional gradient methods (GCG) or Frank–Wolfe-type algorithms. They can be regarded as the infinite-dimensional generalization of the classical Frank–Wolfe optimization algorithm [41] and of GCG in Banach spaces [7, 17, 19, 25, 33, 43]. The basic idea behind such algorithms consists in exploiting the structure of sparse solutions to (4), which are given by finite linear combinations of Dirac deltas supported on Ω. In this case, Dirac deltas represent the extremal points of the unit ball of the Radon norm regularizer. With this knowledge at hand, the GCG method iteratively minimizes a linearized version of (4); such minimum can be found in the set of extremal points. The iterate is then constructed by adding delta peaks at each iteration, and by subsequently optimizing the coefficients of the linear combination. GCG methods have proven to be successful at solving (4), and have been adapted to related problems in the context of, e.g., super-resolution [1, 49, 52, 59].

1.2 Outline of the Main Contributions

Inspired by GCG methods, the goal of this paper is to develop a dynamic generalized conditional gradient method (DGCG) aimed at solving the dynamic minimization problem at (3). Similarly to the classical GCG approaches, our DGCG algorithm is based on the structure of sparse solutions to (3), and it is Lagrangian in essence, since it does not require a discretization of the space domain. Lagrangian approaches have been proven useful for many different dynamic applications [55], often outperforming Eulerian approaches, where the discretization in space is necessary. Indeed, since Eulerian approaches are based on the optimization of challenging discrete assignment problems in space, Lagrangian approaches allow to lower the computational costs and are more suitable to reconstruct coalescence phenomena in the dynamics. Motivated by similar considerations, our approach aims to reduce the reconstruction artifacts and lower the computational cost when compared to grid-based methods designed to solve similar inverse problems to (3) (see [60]).

    The fundamentals of our approach rest on recent results concerning sparsity for variational inverse problems: it has been empirically observed that the presence of a regularizer promotes the existence of sparse solutions, that is, minimizers that can be represented as a finite linear combination of simpler atoms. This effect is evident in reconstruction problems [34, 39, 64–66], as well as in variational problems in other applications, such as materials science [37, 38, 46, 53]. Existence of sparse solutions has been recently proven for a class of general functionals comprised of a fidelity term, mapping to a finite dimensional space, and a regularizer: in this case, atoms correspond to the extremal points of the unit ball of the regularizer [11, 14].

In the context of (3), the extremal points of the Benamou–Brenier energy have been recently characterized in [20]; this provides an operative notion of atoms that will be used throughout the paper. More precisely, for every absolutely continuous curve $\gamma : [0, 1] \to \overline{\Omega}$, we name as *atom* of the Benamou–Brenier energy the respective pair of measures $\mu_\gamma := (\rho_\gamma, m_\gamma) \in \mathcal{M}$ defined by

$$\rho_\gamma := a_\gamma \, dt \otimes \delta_{\gamma(t)}, \quad m_\gamma := \dot{\gamma}(t)\rho_\gamma, \quad a_\gamma := \left(\frac{\beta}{2} \int_0^1 |\dot{\gamma}(t)|^2 \, dt + \alpha \right)^{-1}. \tag{5}$$

The notion of atom described above can be regarded as the dynamic counterpart of the Dirac deltas for the Radon norm regularizer. Curves of measures of the form (5) constitute the building blocks used in our DGCG method to generate at iteration step n the sparse iterate $\mu^n = (\rho^n, m^n)$

$$\mu^n = \sum_{j=1}^{N_n} c_j \mu_{\gamma_j} \tag{6}$$

converging to a solution of (3), where $c_j > 0$.

The basic DGCG method proposed is comprised of two steps. The first one, called *insertion step*, operates as follows. Given a sparse iterate μ^n of the form (6), we obtain a descent direction for the energy at (3), by minimizing a version of (3) around μ^n, in which the fidelity term is linearized. We show that in order to find such descent direction, it is sufficient to solve

$$\min_{(\rho,m)\in\text{Ext}(C_{\alpha,\beta})} - \int_0^1 \langle \rho_t, w_t^n \rangle_{H_t} \, dt, \tag{7}$$

where $w_t^n := -K_t(K_t^* \rho_t^n - f_t) \in C(\overline{\Omega})$ is the dual variable of the problem at the iterate μ^n, and $\text{Ext}(C_{\alpha,\beta})$ denotes the set of extremal points of the unit ball of the regularizer in (3), namely the set $C_{\alpha,\beta} := \{(\rho,m) \in \mathcal{M} : \partial_t \rho + \text{div}\, m = 0, \ \beta B(\rho,m) + \alpha \|\rho\|_{\mathcal{M}(X)} \leq 1\}$. Formula (7) clarifies the connection between atoms and extremal points of $C_{\alpha,\beta}$, showing the fundamental role that the latter play in sparse optimization and GCG methods. In view of the characterization Theorem 1, proven in [20], the minimization problem (7) can be equivalently written in terms of atoms

$$\min_{\gamma\in\text{AC}^2} \left\{ 0, -a_\gamma \int_0^1 w_t^n(\gamma(t)) \, dt \right\}, \tag{8}$$

where AC^2 denotes the set of absolutely continuous curves with values in $\overline{\Omega}$ and weak derivative in $L^2(\overline{\Omega}; \mathbb{R}^d)$. The insertion step then consists in finding a curve γ^* solving (8), and considering the respective atom μ_{γ^*}. Afterward, naming $\gamma_{N_n+1} := \gamma^*$, the *coefficients optimization step* proceeds at optimizing the conic combination $\mu^n + c_{N_n+1}\mu_{\gamma_{N_n+1}}$ with respect to (3), among all nonnegative coefficients c_j. Denoting by $c_1^*, \ldots, c_{N_n+1}^*$ a solution to such problem, the new iterate is defined by $\mu^{n+1} := \sum_j c_j^* \mu_{\gamma_j}$. The two steps of inserting a new atom in the linear combination and optimizing the coefficients are the building blocks of our *core algorithm*,

summarized in Algorithm 1. In Theorem 7, we prove that such algorithm has a sublinear convergence rate, similarly to the GCG method for the BLASSO problem [18], and the produced iterates μ^n converge in the weak* sense of measures to a solution of (3). The core algorithm and its analysis are the subject of Sect. 4.

From the computational point of view, we observe that the coefficients optimization step can be solved efficiently, as it is equivalent to a finite-dimensional quadratic program. Concerning the insertion step, however, even if the complexity of searching for a descent direction for (3) is reduced by only minimizing in the set of atoms, (8) remains a challenging nonlinear and non-local problem. For this reason, we shift our attention to computing stationary points for (8), relying on gradient descent strategies. Specifically, we prove that under additional assumptions on H_t and K_t^*, problem (8) can be cast in the Hilbert space $H^1([0, 1]; \mathbb{R}^d)$, and that the gradient descent algorithm, with appropriate stepsize, outputs stationary points to (8) (see Theorem 14). With this theoretical result at hand, in Sect. 5.1, we formulate a solution strategy for (8) based on multistart gradient descent methods, whose initializations are chosen according to heuristic principles. More precisely, the initial curves are chosen randomly in the regions where the dual variable w_t^n has larger value, and new starting curves are produced combining pieces of stationary curves for (8) by means of a procedure named *crossover*.

We complement the core algorithm with acceleration strategies. First, we add multiple atoms in the insertion step (*multiple insertion step*). Such new atoms can be easily obtained as a byproduct of the multistart gradient descent in the insertion step. Moreover, after optimizing the coefficients in the linear combination, we perform an additional gradient descent step with respect to (3), varying the curves in the iterate μ^{n+1}, while keeping the weights fixed. Such procedure, named *sliding step*, will then be alternated with the coefficients optimization step for a fixed number of iterations, before searching for a new atom in the insertion step. These additional steps are described in Sect. 5.1. We mention that similar strategies were already employed for the BLASSO problem [18, 51]. They are then included in the basic core algorithm to obtain the complete DGCG method in Algorithm 2.

In Section 6, we provide numerical examples. As observation operators K_t^*, we use time-dependent undersampled Fourier measurements, popular in imaging and medical imaging [16, 35], as well as in compressed sensing and super-resolution [1, 22, 23]. Such examples show the effectiveness of our DGCG method in reconstructing spatially sparse data, in the presence of simultaneously strong noise and severe temporal undersampling. Indeed, satisfactory results are obtained for ground-truths with 20% and 60% of added Gaussian noise, and heavy temporal undersampling in the sense that at each time t, the observation operator K_t^* is not able to distinguish sources along lines. With such ill-posed measurements, static reconstruction methods would not be able to accurately recover any ground-truth. In contrast, the time regularization chosen in (3) allows to resolve the dynamics by correlating the information of neighboring data points. As shown by the experiments presented, our DGCG algorithm produces accurate reconstructions of the ground-truth. In case of 20% and 60% of added noise, we note a surge of low intensity artifacts; nonetheless, the obtained reconstruction is close, in the sense of the measures, to the original ground-truth. Moreover, in all of the tried out examples, a linear convergence rate has been observed; this shows that

the algorithm is, in practice, faster than the theoretical guarantees (Theorem 7), and a linear rate has to be expected in most of the cases.

We conclude the paper with Sect. 7, in which we discuss future perspectives and open questions, such as the possibility of improving the theoretical convergence rate for Algorithm 1, and alternative modeling choices.

1.3 Organization of the Paper

The paper is organized as follows. In Sect. 2, we summarize all the relevant notations and preliminary results regarding the Benamou–Brenier energy that are needed in the paper. In particular, we recall the characterization of the extremal points of the unit ball of the Benamou–Brenier energy obtained in [20]. In Sect. 3, we introduce the dynamic inverse problem under consideration and its regularization (3), following the approach of [15]. We further establish basic theory needed to setup the DGCG method. In Sect. 4, we provide the definition of atoms, and we give a high-level description of the DGCG method we propose in this paper, see Algorithm 1, proving its sublinear convergence. In Sect. 5, we describe the strategy employed to solve the insertion step problem (8), based on a multistart gradient descent method. Moreover, we outline the mentioned acceleration steps. Incorporating these procedures in the core algorithm, we obtain the complete DGCG method, see Algorithm 2. In Sect. 6, we show numerical results supporting the effectiveness of Algorithm 2. Finally, we present the reader some open questions in Sect. 7.

2 Preliminaries and Notation

In this section, we introduce the mathematical concepts and results we need to formulate our minimization problem and consequent algorithms. Throughout the paper, $\Omega \subset \mathbb{R}^d$ denotes an open-bounded domain with $d \in \mathbb{N}$, $d \geq 1$. We define the time-space cylinder $X := (0, 1) \times \overline{\Omega}$. Following [3], given a metric space Y, we denote by $\mathcal{M}(Y)$, $\mathcal{M}(Y; \mathbb{R}^d)$, $\mathcal{M}^+(Y)$, the spaces of bounded Borel measures, bounded vector Borel measures, and positive measures, respectively. For a scalar measure ρ, we denote by $\|\rho\|_{\mathcal{M}(Y)}$ its total variation. In addition, we employ the notations $\mathbb{R}_+$ and $\mathbb{R}_{++}$ to refer to the nonnegative and positive real numbers, respectively.

2.1 Time-Dependent Measures

We say that $\{\rho_t\}_{t \in [0,1]}$ is a Borel family of measures in $\mathcal{M}(\overline{\Omega})$ if $\rho_t \in \mathcal{M}(\overline{\Omega})$ for every $t \in [0, 1]$ and the map $t \mapsto \int_{\overline{\Omega}} \varphi(x) \, d\rho_t(x)$ is Borel measurable for every function $\varphi \in C(\overline{\Omega})$. Given a measure $\rho \in \mathcal{M}(X)$, we say that ρ disintegrates with respect to time if there exists a Borel family of measures $\{\rho_t\}_{t \in [0,1]}$ in $\mathcal{M}(\overline{\Omega})$ such that

$$\int_X \varphi(t, x) \, d\rho(t, x) = \int_0^1 \int_{\overline{\Omega}} \varphi(t, x) \, d\rho_t(x) \, dt \quad \text{for all} \quad \varphi \in L^1_\rho(X).$$

We denote such disintegration with the symbol $\rho = dt \otimes \rho_t$. Further, we say that a curve of measures $t \in [0, 1] \mapsto \rho_t \in \mathcal{M}(\overline{\Omega})$ is narrowly continuous if, for all $\varphi \in C(\overline{\Omega})$, the map $t \mapsto \int_{\overline{\Omega}} \varphi(x) \, d\rho_t(x)$ is continuous. The family of narrowly continuous curves will be denoted by C_{w}. We denote by C_{w}^+ the family of narrowly continuous curves with values in $\mathcal{M}^+(\overline{\Omega})$.

2.2 Optimal Transport Regularizer

Introduce the space

$$\mathcal{M} := \mathcal{M}(X) \times \mathcal{M}(X; \mathbb{R}^d).$$

We denote elements of $\mathcal{M}$ by $\mu = (\rho, m)$ with $\rho \in \mathcal{M}(X)$, $m \in \mathcal{M}(X; \mathbb{R}^d)$, and by 0 the pair $(0, 0)$. Define the set of pairs in $\mathcal{M}$ satisfying the continuity equation as

$$\mathcal{D} := \{\mu \in \mathcal{M} \; : \; \partial_t \rho + \operatorname{div} m = 0 \;\text{ in }\; X\},$$

where the solutions of the continuity equation are intended in a distributional sense, that is,

$$\int_X \partial_t \varphi \, \mathrm{d}\rho + \int_X \nabla\varphi \cdot \mathrm{d}m = 0 \quad \text{for all} \quad \varphi \in C_c^\infty(X). \tag{9}$$

The above weak formulation includes no flux boundary conditions for the momentum m on $\partial\Omega$, and no initial and final data for ρ. Notice that by standard approximation arguments, it is equivalent to test (9) against maps in $C_c^1(X)$ (see [4, Remark 8.1.1]).

We now introduce the Benamou–Brenier energy, as originally done in [8]. To this end, define the convex, one-homogeneous and lower semicontinuous map $\Psi \colon \mathbb{R} \times \mathbb{R}^d \to [0, \infty]$ as

$$\Psi(t, x) := \begin{cases} \frac{|x|^2}{2t} & \text{if } t > 0, \\ 0 & \text{if } t = |x| = 0, \\ +\infty & \text{otherwise}. \end{cases} \tag{10}$$

The Benamou–Brenier energy $B \colon \mathcal{M} \to [0, \infty]$ is defined by

$$B(\mu) := \int_X \Psi\left(\frac{\mathrm{d}\rho}{\mathrm{d}\lambda}, \frac{\mathrm{d}m}{\mathrm{d}\lambda}\right) \mathrm{d}\lambda, \tag{11}$$

where $\lambda \in \mathcal{M}^+(X)$ is any measure satisfying $\rho, m \ll \lambda$. Note that (11) does not depend on the choice of λ, as Ψ is one-homogeneous. Following [15], we introduce a coercive version of B: for fixed parameters $\alpha, \beta > 0$ define the functional $J_{\alpha,\beta} \colon \mathcal{M} \to [0, \infty]$ as

$$J_{\alpha,\beta}(\mu) := \begin{cases} \beta B(\mu) + \alpha \, \|\rho\|_{\mathcal{M}(X)} & \text{if } (\rho, m) \in \mathcal{D}, \\ +\infty & \text{otherwise}. \end{cases} \tag{12}$$

As recently shown [15], $J_{\alpha,\beta}$ can be employed as a regularizer for dynamic inverse problems in spaces of measures.

2.3 Extremal Points of $J_{\alpha,\beta}$

Define the convex unit ball

$$C_{\alpha,\beta} := \left\{ \mu \in \mathcal{M} : \ J_{\alpha,\beta}(\mu) \leq 1 \right\},$$

and the set of measures concentrated on AC^2 curves in $\overline{\Omega}$

$$\mathcal{C}_{\alpha,\beta} := \left\{ \mu_\gamma \in \mathcal{M} : \ \gamma \in AC^2([0,1]; \mathbb{R}^d), \ \gamma([0,1)) \subset \overline{\Omega} \right\}, \tag{13}$$

where we denote by μ_γ the pair (ρ_γ, m_γ) with

$$\rho_\gamma := a_\gamma \, dt \otimes \delta_{\gamma(t)}, \quad m_\gamma := \dot{\gamma}(t)\rho_\gamma, \quad a_\gamma := \left(\frac{\beta}{2} \int_0^1 |\dot{\gamma}(t)|^2 \, dt + \alpha \right)^{-1}. \tag{14}$$

Here, $AC^2([0,1]; \mathbb{R}^d)$ denotes the space of curves having metric derivative in $L^2((0,1); \mathbb{R}^d)$. We can identify $AC^2([0,1]; \mathbb{R}^d)$ with the Sobolev space $H^1((0,1); \mathbb{R}^d)$ (see [4, Remark 1.1.3]). For brevity, we will denote by $AC^2 := AC^2([0,1]; \overline{\Omega})$ the set of curves γ belonging to $AC^2([0,1]; \mathbb{R}^d)$ such that $\gamma([0,1]) \subset \overline{\Omega}$. For the extremal points of $C_{\alpha,\beta}$, we have the following characterization result, originally proven in [20, Theorem 6].

Theorem 1 *Let* $\alpha, \beta > 0$ *be fixed. Then, it holds* $\text{Ext}(C_{\alpha,\beta}) = \{0\} \cup \mathcal{C}_{\alpha,\beta}$.

We now show that $J_{\alpha,\beta}$ is linear on nonnegative combinations of points in $\mathcal{C}_{\alpha,\beta}$. Such property will be crucial for several computations in this paper, and the proof is postponed to Sect. 1

Lemma 2 *Let* $N \in \mathbb{N}$, $N \geq 1$, $c_j \in \mathbb{R}$ *with* $c_j > 0$, *and* $\gamma_j \in AC^2$ *for* $j = 1, \ldots, N$. *Let* $\mu_{\gamma_j} = (\rho_{\gamma_j}, m_{\gamma_j}) \in \mathcal{C}_{\alpha,\beta}$ *be defined according to* (14). *Then,* $J_{\alpha,\beta}(\mu_j) = 1$ *and*

$$J_{\alpha,\beta}\left(\sum_{j=1}^N c_j \mu_{\gamma_j} \right) = \sum_{j=1}^N c_j.$$

3 The Dynamic Inverse Problem and Conditional Gradient Method

In this section, we introduce the dynamic inverse problem we aim at solving, following the approach of [15]. Moreover, we set up the functional analytic framework necessary to state the numerical algorithm presented in Sect. 4. Recall that $\Omega \subset \mathbb{R}^d$ is an open-bounded domain, $d \in \mathbb{N}$, $d \geq 1$. Let $\{H_t\}_{t \in [0,1]}$ be a family of real Hilbert spaces, $K_t^* : \mathcal{M}(\overline{\Omega}) \to H_t$ a family of linear continuous forward operators parametrized by $t \in [0,1]$. Given some data $f_t \in H_t$ for a.e. $t \in (0,1)$, consider the dynamic inverse problem of finding a curve $t \mapsto \rho_t \in \mathcal{M}(\overline{\Omega})$ such that

$$K_t^* \rho_t = f_t \quad \text{for a.e. } t \in (0,1). \tag{15}$$

It has been recently proposed [15] to regularize the above problem with the optimal transport energy $J_{\alpha,\beta}$ defined in (12), where $\alpha, \beta > 0$ are fixed parameters. This leads to consider the Tikhonov functional $T_{\alpha,\beta} \colon \mathcal{M} \to [0, +\infty]$ with associated minimization problem

$$\min_{\mu \in \mathcal{M}} T_{\alpha,\beta}(\mu), \quad T_{\alpha,\beta}(\mu) := \mathcal{F}(\mu) + J_{\alpha,\beta}(\mu), \tag{$\mathcal{P}$}$$

where the fidelity term $\mathcal{F} \colon \mathcal{M} \to [0, +\infty]$ is defined by

$$\mathcal{F}(\mu) := \begin{cases} \dfrac{1}{2} \displaystyle\int_0^1 \left\| K_t^* \rho_t - f_t \right\|_{H_t}^2 \, dt & \text{if } \rho = dt \otimes \rho_t, \ (t \mapsto \rho_t) \in C_{\mathrm{w}}, \\ +\infty & \text{otherwise}. \end{cases} \tag{16}$$

In the following, we will denote by f, $K^*\rho$, Kf the maps $t \mapsto f_t$, $t \mapsto K_t^*\rho_t$, $t \mapsto K_t f_t$, respectively, where $f_t \in H_t$ for a.e. $t \in (0, 1)$ and ρ_t is the disintegration of ρ with respect to time. The fidelity term $\mathcal{F}$ serves to track the discrepancy in (15) continuously in time. Following [15], this is achieved by introducing the Hilbert space of square integrable maps $f \colon [0, 1] \to H := \cup_t H_t$, denoted by L_H^2. The data f are then assumed to belong to L_H^2. The assumptions under which this procedure can be made rigorous are briefly summarized in Sect. 3.1, see (H1)–(H3), (K1)–(K3). Under these assumptions, we have that $\mathcal{F}$ is well-defined, see Remark 1. Such framework allows to model a variety of time-dependent acquisition strategies in dynamic imaging, as shown in Sect. 1. We are now ready to recall an existence result for ($\mathcal{P}$) (see [15, Theorem 4.4]).

Theorem 3 *Assume (H1)–(H3), (K1)–(K3) as in Sect. 3.1. Let $f \in L_H^2$ and $\alpha, \beta > 0$. Then, $T_{\alpha,\beta}$ is weak* lower semicontinuous on $\mathcal{M}$, and there exists $\mu^* \in \mathcal{D}$ that solves the minimization problem ($\mathcal{P}$). Moreover, ρ^* disintegrates in $\rho^* = dt \otimes \rho_t^*$ with $(t \mapsto \rho_t^*) \in C_{\mathrm{w}}^+$. If in addition K_t^* is injective for a.e. $t \in (0, 1)$, then μ^* is unique.*

The proposed numerical approach for ($\mathcal{P}$) is based on the conditional gradient method, which consists in seeking minimizers of local linear approximations of the target functional. As standard practice [18, 19, 51], we first replace ($\mathcal{P}$) with a surrogate minimization problem, by defining the functional $\tilde{T}_{\alpha,\beta}$ as in ($\tilde{\mathcal{P}}$) below. The key step in a conditional gradient method is then to find the steepest descent direction for a linearized version of $\tilde{T}_{\alpha,\beta}$. In Sect. 3.3, we show that in order to find such direction, it is sufficient to solve the minimization problem

$$\min_{\mu \in \mathrm{Ext}(C_{\alpha,\beta})} -\langle \rho, w \rangle, \tag{17}$$

where $C_{\alpha,\beta} := \{J_{\alpha,\beta} \leq 1\}$, $w_t := -K_t(K_t^* \tilde{\rho}_t - f_t) \in C(\overline{\Omega})$ is the dual variable associated with the current iterate $(\tilde{\rho}, \tilde{m})$, and the linear term $\mu \mapsto \langle \rho, w \rangle$ is defined in (23). Finally, in Sect. 3.4, we define the primal-dual gap G associated with ($\mathcal{P}$), and prove optimality conditions.

3.1 Functional Analytic Setting for Time-Continuous Fidelity Term

In order to define the continuous sampling fidelity term $\mathcal{F}$ at (16), the authors of [15] introduce suitable assumptions on the measurement spaces H_t and on the forward operators $K_t^* : \mathcal{M}(\overline{\Omega}) \to H_t$.

Assumption 1 For a.e. $t \in (0, 1)$, let H_t be a real Hilbert space with norm $\|\cdot\|_{H_t}$ and scalar product $\langle \cdot, \cdot \rangle_{H_t}$. Let D be a real Banach space with norm denoted by $\|\cdot\|_D$. Assume that for a.e. $t \in (0, 1)$, there exists a linear continuous operator $i_t : D \to H_t$ with the following properties:

- (H1) $\|i_t\| \leq C$ for some constant $C > 0$ not depending on t,
- (H2) $i_t(D)$ is dense in H_t,
- (H3) the map $t \in [0, 1] \mapsto \langle i_t \varphi, i_t \psi \rangle_{H_t} \in \mathbb{R}$ is Lebesgue measurable for every fixed $\varphi, \psi \in D$.

Setting $H := \bigcup_{t \in [0,1]} H_t$, it is possible to define the space of square integrable maps $f : [0, 1] \to H$ such that $f_t \in H_t$ for a.e. $t \in (0, 1)$, that is,

$$
L_H^2 = L^2([0, 1]; H) := \left\{ f : [0, 1] \to H : f \text{ strongly measurable,} \right.
$$

$$
\left. \int_0^1 \|f_t\|_{H_t}^2 \, \mathrm{d}t < \infty \right\} . \tag{18}
$$

The strong measurability mentioned in (18) is an extension to time-dependent spaces of the classical notion of strong measurability for Bochner integrals. The common subset D is employed to construct step functions in a suitable way. An important property of strong measurability is that $t \mapsto \langle f_t, g_t \rangle_{H_t}$ is Lebesgue measurable whenever f, g are strongly measurable [15, Remark 3.4]. Moreover, L_H^2 is a Hilbert space with inner product and norm given by

$$
\langle f, g \rangle_{L_H^2} := \int_0^1 \langle f_t, g_t \rangle_{H_t} \, \mathrm{d}t , \quad \|f\|_{L_H^2} := \left(\int_0^1 \|f_t\|_{H_t}^2 \, \mathrm{d}t \right)^{1/2} , \tag{19}
$$

respectively [15, Theorem 3.13]. We refer the interested reader to [15, Section 3] for more details on the construction of such spaces and their properties. We will now state the assumptions required for the measurement operators K_t^*.

Assumption 2 For a.e. $t \in (0, 1)$, the linear continuous operators $K_t^* : \mathcal{M}(\overline{\Omega}) \to H_t$ satisfy:

- (K1) K_t^* is weak*-to-weak continuous, with pre-adjoint denoted by $K_t : H_t \to C(\overline{\Omega})$,
- (K2) $\|K_t^*\| \leq C$ for some constant $C > 0$ not depending on t,
- (K3) the map $t \in [0, 1] \mapsto K_t^* \rho \in H_t$ is strongly measurable for every fixed $\rho \in \mathcal{M}(\overline{\Omega})$.

Remark 1 After assuming (H1)–(H3), (K1)–(K3), the fidelity term $\mathcal{F}$ introduced at (16) is well-defined. Indeed, the conditions $\rho = \mathrm{d}t \otimes \rho_t$ and $(t \mapsto \rho_t) \in C_w$ imply

that $t \mapsto K_t^* \rho_t$ belongs to L_H^2 by Lemma 12. We further remark that $\mathcal{F}(\mu)$ is finite whenever $J_{\alpha,\beta}(\mu) < +\infty$, as in this case we have $\rho = \mathrm{d}t \otimes \rho_t$ with $(t \mapsto \rho_t) \in C_\mathrm{w}^+$, by Lemmas 9, 10.

3.2 Surrogate Minimization Problem

Let $f \in L_H^2$ and define the map $\varphi \colon \mathbb{R}_+ \to \mathbb{R}_+$

$$\varphi(t) := \begin{cases} t & \text{if } t \leq M_0, \\ \frac{t^2 + M_0^2}{2M_0} & \text{if } t > M_0, \end{cases} \tag{20}$$

where we set $M_0 := T_{\alpha,\beta}(0)$. Notice that by (A1), we have $J_{\alpha,\beta}(0) = 0$, so that

$$M_0 = \frac{1}{2} \int_0^1 \|f_t\|_{H_t}^2 \, \mathrm{d}t \,, \tag{21}$$

highlighting the dependence of φ on f. Recalling the definition of $\mathcal{F}$ at (16), define the surrogate minimization problem

$$\min_{\mu \in \mathcal{M}} \tilde{T}_{\alpha,\beta}(\mu) \,, \quad \tilde{T}_{\alpha,\beta}(\mu) := \mathcal{F}(\mu) + \varphi(J_{\alpha,\beta}(\mu)) \,. \tag{$\tilde{\mathcal{P}}$}$$

Notice that ($\mathcal{P}$) and ($\tilde{\mathcal{P}}$) share the same set of minimizers, and they are thus equivalent. This is readily seen after noting that solutions to ($\mathcal{P}$) and ($\tilde{\mathcal{P}}$) belong to the set $\{\mu \in \mathcal{M} : J_{\alpha,\beta}(\mu) \leq M_0\}$, thanks to the estimate $t \leq \varphi(t)$, and that $T_{\alpha,\beta}$ and $\tilde{T}_{\alpha,\beta}$ coincide on the said set.

We remark that the surrogate minimization problem ($\tilde{\mathcal{P}}$) is a technical modification of ($\mathcal{P}$) introduced just to ensure that the partially linearized problem defined in the following section is coercive.

3.3 Linearized Problem

Fix some data $f \in L_H^2$ and a curve $(t \mapsto \tilde{\rho}_t) \in C_\mathrm{w}$. We define the associated dual variable $t \mapsto w_t$ by

$$w_t := -K_t(K_t^* \tilde{\rho}_t - f_t) \in C(\overline{\Omega}) \,, \tag{22}$$

and the map $\mu \in \mathcal{M} \mapsto \langle \rho, w \rangle \in \mathbb{R} \cup \{\pm\infty\}$ as

$$\langle \rho, w \rangle := \begin{cases} \int_0^1 \langle \rho_t, w_t \rangle_{\mathcal{M}(\overline{\Omega}), C(\overline{\Omega})} \, \mathrm{d}t & \text{if } \rho = \mathrm{d}t \otimes \rho_t \,, \ (t \mapsto \rho_t) \in C_\mathrm{w} \,, \\ -\infty & \text{otherwise.} \end{cases} \tag{23}$$

Remark 2 The above map is well-defined; indeed, assuming that $(t \mapsto \rho_t) \in C_\mathrm{w}$, we have $(t \mapsto K_t^* \rho_t) \in L_H^2$ by Lemma 12. Similarly, also $(t \mapsto K_t^* \tilde{\rho}_t) \in L_H^2$. Thus,

recalling (K1), we infer

$$\langle \rho, w \rangle = -\langle K^* \rho, K^* \tilde{\rho} - f \rangle_{L_H^2}, \tag{24}$$

which is well-defined and finite, being a scalar product in the Hilbert space L_H^2 (see (19)). Moreover if $J_{\alpha,\beta}(\mu) < +\infty$, then $\langle \rho, w \rangle$ is finite, since $\rho = \mathrm{d}t \otimes \rho_t$ for $(t \mapsto \rho_t) \in C_w^+$, by Lemmas 9, 10.

Let φ and M_0 be as in (20)-(21). We consider the following linearized version of $(\tilde{\mathcal{P}})$

$$\min_{\mu \in \mathcal{M}} -\langle \rho, w \rangle + \varphi(J_{\alpha,\beta}(\mu)), \tag{25}$$

which is well-posed by Theorem 13.

The objective of this section is to prove the existence of a solution to (25) belonging, up to a multiplicative constant, to the extremal points of the sublevel set $C_{\alpha,\beta} := \{J_{\alpha,\beta} \leq 1\}$.

To this end, consider the problem

$$\min_{\mu \in C_{\alpha,\beta}} -\langle \rho, w \rangle. \tag{26}$$

In the following proposition, we prove that (26) admits a minimizer $\mu^* \in \mathrm{Ext}(C_{\alpha,\beta})$. Moreover, we show that a suitably rescaled version of μ^* solves (25).

Proposition 4 *Assume (H1)–(H3), (K1)–(K3) as in Sect. 3.1. Let $f \in L_H^2$, $\alpha, \beta > 0$. Then, there exists a solution $\mu^* \in \mathrm{Ext}(C_{\alpha,\beta})$ to (26). Moreover, $M\mu^*$ is a minimizer for (25), where*

$$M := \begin{cases} 0 & \text{if } \langle \rho^*, w \rangle \leq 1, \\ M_0 \langle \rho^*, w \rangle & \text{if } \langle \rho^*, w \rangle > 1. \end{cases} \tag{27}$$

The above statement is reminiscent of the classical Bauer Maximum Principle [2, Theorem 7.69]. In our case, however, there is no clear topology that makes the set $C_{\alpha,\beta}$ compact and the linearized map defined in (23) continuous (or upper semicontinuous). Therefore, an ad hoc proof is required.

Proof Let $\hat{\mu} \in C_{\alpha,\beta}$ be a solution to (26), which exists thanks to Theorem 13 with the choice $\varphi(t) := \chi_{(-\infty,1]}(t)$. Consider the set $S := \{\mu \in C_{\alpha,\beta} : \langle \rho, w \rangle = \langle \hat{\rho}, w \rangle\}$ of all solutions to (26). Note that S is bounded with respect to the total variation on $\mathcal{M}$, due to (A2) and definition of $C_{\alpha,\beta}$. In particular, the weak* topology of $\mathcal{M}$ is metrizable in S. We claim that S is compact in the same topology. Indeed, given a sequence $\{\mu^n\}$ in S we have by definition that

$$\sup_n J_{\alpha,\beta}(\mu^n) \leq 1, \qquad \langle \rho^n, w \rangle = \langle \hat{\rho}, w \rangle \quad \text{for every } n \in \mathbb{N}. \tag{28}$$

Therefore, (28) and Lemma 11 imply that, up to subsequences, μ^n converges to some $\mu \in \mathcal{M}$ in the sense of (A3). By (28) and by the weak* sequential lower semicontinuity

of $J_{\alpha,\beta}$ (Lemma 11), we infer $\mu \in C_{\alpha,\beta}$. Moreover by (28), (24) and Lemma 12, we also conclude that $\mu \in S$, hence proving compactness. Also notice that S is convex due to the convexity of $J_{\alpha,\beta}$ (Lemma 11) and linearity of the constraint. Since $S \neq \emptyset$, by Krein–Milman's theorem, we have that $\text{Ext}(S) \neq \emptyset$. Let $\mu^* \in \text{Ext}(S)$. If we show that $\mu^* \in \text{Ext}(C_{\alpha,\beta})$, the thesis is achieved by definition of S. Hence, assume that μ^* can be decomposed as

$$\mu^* = \lambda \mu^1 + (1 - \lambda)\mu^2 \tag{29}$$

with $\mu^j = (\rho^j, m^j) \in C_{\alpha,\beta}$ and $\lambda \in (0, 1)$. Assume that μ^1 belongs to $C_{\alpha,\beta} \smallsetminus S$. By (29) and the minimality of the points in S for (26), we infer $-\langle \hat{\rho}, w \rangle < -\langle \rho^*, w \rangle$, which is a contradiction since $\mu^* \in S$. Therefore, $\mu^1 \in S$. Similarly also $\mu^2 \in S$. Since $\mu^* \in \text{Ext}(S)$, from (29), we infer $\mu^* = \mu^1 = \mu^2$, showing that $\mu^* \in \text{Ext}(C_{\alpha,\beta})$.

Assume now that $\mu^* \in \text{Ext}(C_{\alpha,\beta})$ minimizes in (26). If $\mu^* = 0$, it is straightforward to check that 0 minimizes in (25). Hence, assume $\mu^* \neq 0$, so that $J_{\alpha,\beta}(\mu^*) > 0$ by (A2). Since the functional at (26) is linear and μ^* is a minimizer, we can scale by $J_{\alpha,\beta}(\mu^*)$ and exploit the one-homogeneity of $J_{\alpha,\beta}$ to obtain $J_{\alpha,\beta}(\mu^*) = 1$. For every $\mu = (\rho, m) \in \mathcal{M}$ such that $J_{\alpha,\beta}(\mu) < +\infty$, one has

$$- \langle \rho, w \rangle + \varphi(J_{\alpha,\beta}(\mu)) \geq -J_{\alpha,\beta}(\mu)\langle \rho^*, w \rangle + \varphi(J_{\alpha,\beta}(\mu)), \tag{30}$$

since μ^* is a minimizer, $J_{\alpha,\beta}$ is nonnegative and one-homogeneous, and since $J_{\alpha,\beta}(\mu) = 0$ if and only if $\mu = 0$ by Lemma 10. Again one-homogeneity implies

$$\inf_{\mu \in \mathcal{M}} -J_{\alpha,\beta}(\mu)\langle \rho^*, w \rangle + \varphi(J_{\alpha,\beta}(\mu)) = \inf_{\lambda \geq 0} -\lambda\langle \rho^*, w \rangle + \varphi(\lambda). \tag{31}$$

It is immediate to check that M defined in (27) is a minimizer for the right-hand side problem in (31).

Hence from (30)–(31), one-homogeneity of $J_{\alpha,\beta}$ and the fact that $J_{\alpha,\beta}(\mu^*) = 1$, we conclude that $M\mu^*$ is a minimizer for (25). □

3.4 The Primal-Dual Gap

In this section, we introduce the primal-dual gap associated with ($\mathcal{P}$).

Definition 1 The primal-dual gap is defined as the map $G \colon \mathcal{M} \to [0, +\infty]$ such that

$$G(\mu) := \begin{cases} J_{\alpha,\beta}(\mu) - \varphi(J_{\alpha,\beta}(\hat{\mu})) - \langle \rho - \hat{\rho}, w \rangle & \text{if } J_{\alpha,\beta}(\mu) < +\infty, \\ +\infty & \text{otherwise,} \end{cases} \tag{32}$$

for $\mu \in \mathcal{M}$. Here $w_t := -K_t(K_t^* \rho_t - f_t)$, the product $\langle \cdot, \cdot \rangle$ is defined in (23), the map φ at (20), and $\hat{\mu} \in \mathcal{M}$ is a solution to (25).

Notice that G is well-defined. Indeed, assume that $\mu \in \mathcal{M}$ is such that $J_{\alpha,\beta}(\mu) < +\infty$ and let $\hat{\mu} \in \mathcal{M}$ be a solution to (25). In particular, $J_{\alpha,\beta}(\hat{\mu}) < +\infty$ by Theorem 13. Therefore, the scalar product in (32) is finite, see Remark 2. The purpose of G

becomes clear in its relationship with the *functional distance* associated with $T_{\alpha,\beta}$, which is defined by

$$r(\mu) := T_{\alpha,\beta}(\mu) - \min T_{\alpha,\beta} , \tag{33}$$

for all $\mu \in \mathcal{M}$. Such relationship is described in the following lemma. We remark that a similar result is standard in the context of Frank–Wolfe-type algorithms and generalized conditional gradient methods (see e.g., [18, Lemma 5.5] and [43, Section 2]). Due to the specificity of our dynamic problem and for sake of completeness, we present it in our setting as well.

Lemma 5 *Let $\mu \in \mathcal{M}$ be such that $J_{\alpha,\beta}(\mu) < +\infty$. Then,*

$$r(\mu) \leq G(\mu) . \tag{34}$$

Moreover, μ^* solves ($\mathcal{P}$) if and only if $G(\mu^*) = 0$.

Proof Let w and $\hat{\rho}$ be as in (32). By using the Hilbert structure of L_H^2, for any $\tilde{\mu}$ such that $J_{\alpha,\beta}(\tilde{\mu}) < +\infty$, we have, by the polarization identity,

$$- \langle \rho - \tilde{\rho}, w \rangle = \frac{\|K^*\rho - f\|_{L_H^2}^2}{2} - \frac{\|K^*\tilde{\rho} - f\|_{L_H^2}^2}{2} + \frac{\|K^*(\tilde{\rho} - \rho)\|_{L_H^2}^2}{2} . \tag{35}$$

Let μ^* be a minimizer for $T_{\alpha,\beta}$, which exists by Theorem 3. Since $J_{\alpha,\beta}(\mu^*) \leq T_{\alpha,\beta}(\mu^*) \leq T_{\alpha,\beta}(0) = M_0$, by definition of φ, we have $\varphi(J_{\alpha,\beta}(\mu^*)) = J_{\alpha,\beta}(\mu^*)$. Using (35) and the definition of G in (32), where $\hat{\mu}$ is chosen to be a solution to (25), we obtain

$$
\begin{aligned}
G(\mu) &\geq - \langle \rho - \rho^*, w \rangle + J_{\alpha,\beta}(\mu) - J_{\alpha,\beta}(\mu^*) \\
&\geq \frac{\|K^*\rho - f\|_{L_H^2}^2}{2} - \frac{\|K^*\rho^* - f\|_{L_H^2}^2}{2} + J_{\alpha,\beta}(\mu) - J_{\alpha,\beta}(\mu^*) \\
&= T_{\alpha,\beta}(\mu) - T_{\alpha,\beta}(\mu^*) = r(\mu) ,
\end{aligned}
$$

proving (34). If $G(\mu^*) = 0$, then μ^* minimizes in ($\mathcal{P}$) by (34). Conversely, assume that μ^* is a solution of ($\mathcal{P}$) and denote by $w_t^* := -K_t(K_t^* \rho_t^* - f_t)$ the associated dual variable. Let μ be arbitrary and such that $J_{\alpha,\beta}(\mu) < +\infty$. Let $s \in [0, 1]$ and set $\mu^s := \mu^* + s(\mu - \mu^*)$. By convexity of $J_{\alpha,\beta}$ (see Lemma 11), we have that $J_{\alpha,\beta}(\mu^s) < +\infty$. As μ^* is optimal in ($\mathcal{P}$), we infer

$$
\begin{aligned}
0 &\leq T_{\alpha,\beta}(\mu^s) - T_{\alpha,\beta}(\mu^*) \\
&\leq \frac{\|K^*\rho^s - f\|_{L_H^2}^2}{2} - \frac{\|K^*\rho^* - f\|_{L_H^2}^2}{2} + s(J_{\alpha,\beta}(\mu) - J_{\alpha,\beta}(\mu^*)) \\
&= -s \langle \rho - \rho^*, w^* \rangle + s^2 \frac{\|K^*(\rho^* - \rho)\|_{L_H^2}^2}{2} + s(J_{\alpha,\beta}(\mu) - J_{\alpha,\beta}(\mu^*)) ,
\end{aligned}
$$

where we used convexity of $J_{\alpha,\beta}$, and the identity at (35) with respect to w^*, ρ^* and ρ^s. Dividing the above inequality by s and letting $s \to 0$ yields

$$0 \leq -\langle \rho - \rho^*, w^* \rangle + J_{\alpha,\beta}(\mu) - J_{\alpha,\beta}(\mu^*), \tag{36}$$

which holds for all μ with $J_{\alpha,\beta}(\mu) < +\infty$. Now notice that $J_{\alpha,\beta}(\mu^*) \leq T_{\alpha,\beta}(\mu^*) \leq M_0$, since μ^* solves ($\mathcal{P}$). Therefore, $\varphi(J_{\alpha,\beta}(\mu^*)) = J_{\alpha,\beta}(\mu^*)$. Moreover, $t \leq \varphi(t)$ for all $t \geq 0$. As a consequence of (36), we then infer

$$-\langle \rho^*, w^* \rangle + \varphi(J_{\alpha,\beta}(\mu^*)) = -\langle \rho^*, w^* \rangle + J_{\alpha,\beta}(\mu^*)$$
$$\leq -\langle \rho, w^* \rangle + J_{\alpha,\beta}(\mu) \leq -\langle \rho, w^* \rangle + \varphi(J_{\alpha,\beta}(\mu)),$$

proving that μ^* minimizes in (25) with respect to w^*. Therefore, by definition, $G(\mu^*) = 0$. □

4 The Algorithm: Theoretical Analysis

In this section, we give a theoretical description of the dynamic generalized conditional gradient algorithm anticipated in the introduction, which we call *core algorithm*. The proposed algorithm aims at finding minimizers to $T_{\alpha,\beta}$ as defined in ($\mathcal{P}$), for some fixed data $f \in L_H^2$ and parameters $\alpha, \beta > 0$. It is comprised of an *insertion step*, where one seeks a minimizer to (26) among the extremal points of the set $C_{\alpha,\beta} := \{J_{\alpha,\beta} \leq 1\}$, and of a *coefficients optimization step*, which will yield a finite dimensional quadratic program. As a result, each iterate will be a finite linear combination, with nonnegative coefficients, of points in $\text{Ext}(C_{\alpha,\beta})$. We remind the reader that $\text{Ext}(C_{\alpha,\beta}) = \{0\} \cup \mathcal{C}_{\alpha,\beta}$ in view of Theorem 1, where $\mathcal{C}_{\alpha,\beta}$ is defined at (13). From the definition of $\mathcal{C}_{\alpha,\beta}$, we see that except for the zero element, the extremal points are in 1-on-1 correspondence with the space of curves $\text{AC}^2 := \text{AC}^2([0, 1]; \overline{\Omega})$. This observation motivates us to define the atoms of our problem.

Definition 2 *(Atoms)* We denote by AC_∞^2 the one-point extension of the set AC^2, where we include a point denoted by γ_∞. For any $\gamma \in \text{AC}^2$, we name as *atom* the respective extremal point $\mu_\gamma = (\rho_\gamma, m_\gamma) \in \mathcal{M}$ defined according to (14).

For γ_∞, the corresponding atom is defined by $\mu_{\gamma_\infty} := (0, 0)$. We call *sparse* any measure $\mu \in \mathcal{M}$ such that

$$\mu = \sum_{j=1}^{N} c_j \mu_{\gamma_j} \tag{37}$$

for some $N \in \mathbb{N}$, $c_j > 0$ and $\gamma_j \in \text{AC}^2$, with $\gamma_i \neq \gamma_j$ for $i \neq j$.

Note that γ_∞ can be regarded as the infinite length curve: indeed if $\{\gamma^n\}$ in AC^2 is a sequence of curves with diverging length, that is, $\int_0^1 |\dot{\gamma}^n| \, dt \to \infty$ as $n \to \infty$, then $\rho_{\gamma^n} \overset{*}{\rightharpoonup} \rho_{\gamma_\infty}$, since $a_{\gamma^n} \to 0$ by Hölder's inequality.

Additionally, it is convenient to introduce the following map associating vectors of curves and coefficients with the corresponding sparse measure:

$$(\boldsymbol{\gamma}, \boldsymbol{c}) \mapsto \mu := \sum_{j=1}^{N} c_j \mu_{\gamma_j}, \tag{38}$$

where $N \in \mathbb{N}$ is fixed and $\boldsymbol{c} := (c_1, \ldots, c_N)$ with $c_j > 0$, $\boldsymbol{\gamma} := (\gamma_1, \ldots, \gamma_N)$, with $\gamma_j \in AC^2$.

Remark 3 The decomposition in extremal points of a given sparse measure might not be unique, that is, the map at (38) is not injective. For example, let $N := 2, \Omega := (0, 1)^2$ and

$$\gamma_1(t) := (t, t), \qquad \tilde{\gamma}_1(t) := (t, t)\chi_{[0,1/2)}(t) + (t, 1-t)\chi_{[1/2,1]}(t),$$
$$\gamma_2(t) := (t, 1-t), \quad \tilde{\gamma}_2(t) := (t, 1-t)\chi_{[0,1/2)}(t) + (t, t)\chi_{[1/2,1]}(t).$$

Note that injectivity fails for $((\gamma_1, \gamma_2), (1, 1))$ and $((\tilde{\gamma}_1, \tilde{\gamma}_2), (1, 1))$, given that they map to the same measure μ^*, but γ_1 and γ_2 cross at $t = 1/2$, while $\tilde{\gamma}_1$ and $\tilde{\gamma}_2$ rebound. This observation is relevant for the algorithms presented, seeing that they operate in terms of extremal points: if for example, μ^* was the unique solution to $(\mathcal{P})$ for some data f, due to the lack of unique sparse representation for μ^*, the numerical reconstruction could favor the representation having the least energy in terms of the regularizer $J_{\alpha,\beta}$. This is not surprising, since our method aims at reconstructing sparse measures, rather than their extremal points. A numerical example displaying the behavior of our algorithm on crossings, such as the case of μ^*, is given in Sect. 6.2.3.

The rest of the section is organized as follows. In Sect. 4.1, we present the core algorithm, describing its basic steps and summarizing it in Algorithm 1. In Sect. 4.2, we discuss the equivalence of the coefficients optimization step to a quadratic program, while in Sect. 4.3, we show sublinear convergence of Algorithm 1 in terms of the residual defined at (33). In Sect. 4.4, we detail on a theoretical stopping criterion for our algorithm. To conclude, in Sect. 4.5, we give a description of how to alter Algorithm 1 in case the fidelity term $\mathcal{F}$ at $(\mathcal{P})$ is replaced by a time-discrete version. All the results presented in this section and in the above will hold also for this particular case, with minor modifications.

4.1 Core Algorithm

The core algorithm consists of two steps. In the first one, named the *insertion step*, an atom is added to the current iterate, this atom being the minimizer of the linearized problem defined at (26). In the second step, named the *coefficients optimization step*, the atoms are fixed and their associated weights are optimized to minimize the target functional $T_{\alpha,\beta}$ defined in $(\mathcal{P})$. In what follows, $f \in L^2_H$ is a given datum and $\alpha, \beta > 0$ are fixed parameters.

4.1.1 Iterates

We initialize the algorithm to the zero atom $\mu^0 := 0$. The n-th iteration μ^n is a *sparse* element of $\mathcal{M}$ according to (37), that is,

$$\mu^n = \sum_{j=1}^{N_n} c_j^n \mu_{\gamma_j^n}, \tag{39}$$

where $N_n \in \mathbb{N} \cup \{0\}$, $\gamma_j^n \in \mathrm{AC}^2$, $c_j^n > 0$ and $\gamma_i \neq \gamma_j$ if $i \neq j$. Notice that N_n is counting the number of atoms present at the n-th iteration, and is not necessarily equal to n, since the optimization step could discard atoms by setting their associated weights to zero. In practice, Algorithm 1 operates in terms of curves and weights. That is, the n-th iteration outputs pairs (γ_j, c_j) with $\gamma_j \in \mathrm{AC}^2$, $c_j > 0$: the iterate at (39) can be then constructed via the map (38).

4.1.2 Insertion Step

Assume μ^n is the current iterate. Define the dual variable associated with μ^n as in (22), that is,

$$w_t^n := -K_t(K_t^* \rho_t^n - f_t) \in C(\overline{\Omega}) \quad \text{for a.e. } t \in (0, 1). \tag{40}$$

With it, consider the minimization problem of the form (17), that is,

$$\min_{\mu \in \mathrm{Ext}(C_{\alpha,\beta})} -\langle \rho, w^n \rangle, \tag{41}$$

where the term $\langle \cdot, \cdot \rangle$ is defined in (23). We recall that (41) admits solution by Proposition 4. Thanks to the characterization $\mathrm{Ext}(C_{\alpha,\beta}) = \{0\} \cup \mathcal{C}_{\alpha,\beta}$ provided by Theorem 1, problem (41) can be cast on the space AC_∞^2. Indeed, given $\mu \in \mathcal{C}_{\alpha,\beta}$, following the notations at (13)–(14), we have that $\mu = \mu_\gamma$ for some $\gamma \in \mathrm{AC}^2$. The curve $t \mapsto \delta_{\gamma(t)}$ belongs to C_w^+, and hence $\langle \rho_\gamma, w \rangle$ is finite (see Remark 2). Thus, by definition, we have

$$\langle \rho_\gamma, w^n \rangle = a_\gamma \int_0^1 \langle \delta_{\gamma(t)}, w_t^n \rangle_{\mathcal{M}(\overline{\Omega}), C(\overline{\Omega})} \, \mathrm{d}t = a_\gamma \int_0^1 w_t^n(\gamma(t)) \, \mathrm{d}t, \tag{42}$$

showing that (41) is equivalent to

$$\min_{\gamma \in \mathrm{AC}_\infty^2} -\langle \rho_\gamma, w^n \rangle = \min_{\gamma \in \mathrm{AC}^2} \left\{ 0, -a_\gamma \int_0^1 w_t^n(\gamma(t)) \, \mathrm{d}t \right\}. \tag{43}$$

The insertion step consists in finding a curve $\gamma^* \in \mathrm{AC}_\infty^2$ that solves (43). To such curve, we associate a new atom μ_{γ^*} via Definition 2. Note that γ^* depends on the current iterate μ^n, as well as on the datum f and parameters α, β: however, in order to simplify notations, we omit such dependencies. After, we have a stopping condition:

- if $\langle \rho_{\gamma^*}, w^n \rangle \leq 1$, then μ^n is solution to ($\mathcal{P}$). The algorithm outputs μ^n and stops,
- if $\langle \rho_{\gamma^*}, w^n \rangle > 1$, then μ^n is not a solution to ($\mathcal{P}$) and $\gamma^* \in \mathrm{AC}^2$. The found atom μ_{γ^*} is inserted in the n-th iterate μ^n and the algorithm continues.

The optimality statements in the above stopping condition correspond to positivity conditions on the subgradient of $T_{\alpha,\beta}$ and are rigorously proven in Sect. 4.4. Moreover, the mentioned stopping condition can be made quantitative as discussed in Remark 7.

Remark 4 In this section, we will always assume the availability of an exact solution γ^* to (43). In particular, this allows to obtain a sublinear convergence rate for the core algorithm (Theorem 7), and make the stopping condition rigorous. In practice, however, obtaining γ^* is not always possible, due to the nonlinearity and non-locality of the functional at (43). For this reason, in Sect. 5.1, we propose a strategy aimed at obtaining stationary points of (43). Based on such strategy, a relaxed version of the insertion step is proposed for Algorithm 2, which we employ for the numerical simulations of Sect. 6.

4.1.3 Coefficients Optimization Step

This step is realized after the stopping condition is checked, with the condition $\langle \rho_{\gamma^*}, w^n \rangle > 1$ being satisfied. In particular, as observed above, in this case $\gamma^* \in \mathrm{AC}^2$. We then set $\gamma^n_{N_n+1} := \gamma^*$ and consider the coefficients optimization problem

$$\min_{(c_1, c_2, \ldots, c_{N_n+1}) \in \mathbb{R}_+^{N_n+1}} T_{\alpha,\beta} \left(\sum_{j=1}^{N_n+1} c_j \mu_{\gamma_j^n} \right), \tag{44}$$

where $\mu_{\gamma_j^n}$ for $j = 1, \ldots, N_n$ are the atoms present in the n-th iterate μ^n. If $c \in \mathbb{R}_+^{N_n+1}$ is a solution to the above problem, the next iterate is defined by

$$\mu^{n+1} := \sum_{c_j > 0} c_j \mu_{\gamma_j^n}, \tag{45}$$

thus discarding the curves that do not contribute to (44).

Remark 5 Problem (44) is equivalent to a quadratic program of the form

$$\min_{c=(c_1, \ldots, c_{N_n+1}) \in \mathbb{R}_+^{N_n+1}} \frac{1}{2} c^T \Gamma c + b^T c, \tag{46}$$

where $\Gamma \in \mathbb{R}^{(N_n+1) \times (N_n+1)}$ is a positive-semidefinite and symmetric matrix and $b \in \mathbb{R}^{N_n+1}$, as proven in Proposition 6.

Therefore, throughout the paper, we will always assume the availability of an exact solution to (44). In practice, we solved (44) by means of the free Python software package *CVXOPT* [5, 6].

Algorithm 1: Core algorithm

Input: Data $f \in L_H^2$, parameters $\alpha, \beta > 0$, forward operators $K_t^* : \mathcal{M}(\overline{\Omega}) \to H_t$

1   $c \leftarrow (), \ \gamma \leftarrow ()$
2   **for** $n = 0, 1, \ldots$ **do**
3     $N_n \leftarrow |\gamma|, \ \mu^n \leftarrow \sum_{j=1}^{N_n} c_j \mu_{\gamma_j}, \ w_t^n \leftarrow -K_t(K_t^* \rho_t^n - f_t)$
     /* Insertion step   */
4     $\gamma^* \leftarrow \arg\min_{\gamma \in AC_\infty^2} \ -\langle \rho_\gamma, w^n \rangle$
     /* Stopping condition   */
5   **if** $\langle \rho_{\gamma^*}, w^n \rangle \leq 1$ **then**
6   **return** μ^n
7   **else**
8       $\gamma \leftarrow (\gamma, \gamma^*)$
      /* Coefficients optimization step   */
9       $c \leftarrow \arg\min_{(c_1, \ldots, c_{N_n+1}) \in \mathbb{R}_+^{N_n+1}} T_{\alpha, \beta} \left(\sum_{j=1}^{N_n} c_j \mu_{\gamma_j} + c_{N_n+1} \mu_{\gamma^*} \right)$
10      $N_{n+1} \leftarrow \#\{j : c_j > 0\}$
11      $(\gamma, c) \leftarrow$ delete_zero_weighted(γ, c)
12      $\mu^{n+1} \leftarrow \sum_{j=1}^{N_{n+1}} c_j \mu_{\gamma_j}$

4.1.4 Algorithm Summary

As discussed in Sect. 4.1.1, the iterates of Algorithm 1 are pairs of curves and weights (γ_j, c_j) for $\gamma \in AC^2$, $c_j > 0$ and $j = 1, \ldots, N_n$. In the pseudo-code, we denote such iterates with the tuples $\gamma = (\gamma_1, \ldots, \gamma_{N_n})$ and $c = (c_1, \ldots, c_{N_n})$. Note that such tuples vary in size at each iteration, and the initial iterate of the algorithm, that is, the zero atom, corresponds to the empty tuples $\gamma = ()$ and $c = ()$. We denote the number of elements contained in a tuple γ with the symbol $|\gamma|$. Via the map (38), the iterates (γ, c) define a sequence of sparse measures $\{\mu^n\}$ of the form (39). The generated sequence $\{\mu^n\}$ weakly* converges (up to subsequences) to a minimizer of $T_{\alpha, \beta}$ for the datum $f \in L_H^2$, as shown in Theorem 7. Notice that the assignments at lines 4 and 9 are meant to choose one element in the respective argmin set. The function delete_zero_weighted(γ, c) at line 11 in Algorithm 1 is designed to input a tuple (γ, c) and output another tuple where the curves γ_j and corresponding weights c_j are deleted if $c_j = 0$.

4.2 Quadratic Optimization

We prove the statement in Remark 5. To be more precise, assume (H1)–(H3), (K1)–(K3) from Sect. 3.1 and let $f \in L_H^2, \alpha, \beta > 0$ be given. Fix $N \in \mathbb{N}, \gamma_1, \ldots, \gamma_N \in AC^2$ with $\gamma_i \neq \gamma_j$ for all $i \neq j$, and consider the coefficients optimization problem

$$\min_{c=(c_1, \ldots, c_N) \in \mathbb{R}_+^N} T_{\alpha, \beta} \left(\sum_{j=1}^N c_j \mu_{\gamma_j} \right), \tag{47}$$

where μ_{γ_j} is defined according to (14). For (47), the following holds.

Proposition 6 *Problem* (47) *is equivalent to*

$$\min_{c=(c_1,\dots,c_N)\in\mathbb{R}^N_+} \frac{1}{2}c^T\Gamma c + b^T c \,, \tag{48}$$

where $\Gamma = (\Gamma_{i,j}) \in \mathbb{R}^{N\times N}$ *is a positive semi-definite symmetric matrix and* $b = (b_i) \in \mathbb{R}^N$, *with*

$$\Gamma_{i,j} := a_{\gamma_i}a_{\gamma_j}\int_0^1 \langle K_t^*\delta_{\gamma_i(t)}, K_t^*\delta_{\gamma_j(t)}\rangle_{H_t}\,\mathrm{d}t \,, \quad b_i := 1 - a_{\gamma_i}\int_0^1 \langle K_t^*\delta_{\gamma_i(t)}, f_t\rangle_{H_t}\,\mathrm{d}t \,. \tag{49}$$

Proof As γ_j is continuous, the curve $t \mapsto \delta_{\gamma_j(t)}$ belongs to C_w^+. Hence, the map $t \mapsto K_t^*\delta_{\gamma_j(t)}$ belongs to L_H^2 by Lemma 12 and the quantities at (49) are well-defined. Thanks to definition of $T_{\alpha,\beta}$ and Lemma 2, we immediately see that

$$T_{\alpha,\beta}\left(\sum_{j=1}^N c_j\mu_{\gamma_j}\right) = \frac{1}{2}\sum_{i,j=1}^N c_ic_j\langle K^*\rho_{\gamma_i}, K^*\rho_{\gamma_j}\rangle_{L_H^2}$$

$$- \sum_{j=1}^N c_j\langle K^*\rho_{\gamma_j}, f\rangle_{L_H^2} + \frac{1}{2}\|f\|_{L_H^2}^2 + \sum_{j=1}^N c_j J_{\alpha,\beta}(\mu_{\gamma_j})$$

$$= \frac{1}{2}c^T\Gamma c + b^T c + M_0 \,,$$

where $M_0 \geq 0$ is defined at (21). This shows that (47) and (48) are equivalent. The rest of the statement follows since Γ is the Gramian with respect to the vectors $K^*\rho_{\gamma_1}, \dots, K^*\rho_{\gamma_N}$ in L_H^2. $\qquad\square$

4.3 Convergence Analysis

We prove sublinear convergence for Algorithm 1. The convergence rate is given in terms of the functional distance (33) associated with $T_{\alpha,\beta}$. Throughout the section, we assume that $f \in L_H^2$ is a given datum, $\alpha, \beta > 0$ are fixed regularization parameters and (H1)–(H3), (K1)–(K3) as in Sect. 3.1 hold. The convergence result is stated as follows.

Theorem 7 *Let* $\{\mu^n\}$ *be a sequence generated by Algorithm 1. Then,* $\{T_{\alpha,\beta}(\mu^n)\}$ *is non-increasing and the residual at* (33) *satisfies*

$$r(\mu^n) \leq \frac{C}{n} \,, \quad \text{for all } n \in \mathbb{N} \,, \tag{50}$$

where $C > 0$ *is a constant depending only on* α, β, f *and* K_t^*. *Moreover, each accumulation point* μ^* *of* μ^n *with respect to the weak* topology of* $\mathcal{M}$ *is a minimizer*

for $T_{\alpha,\beta}$. If $T_{\alpha,\beta}$ admits a unique minimizer μ^*, then $\mu^n \overset{*}{\rightharpoonup} \mu^*$ along the whole sequence.

Remark 6 The proof of Theorem 7 follows similar steps to [18, Theorem 5.8] and [51, Theorem 5.4]. We highlight that the proof of monotonicity of $\{T_{\alpha,\beta}(\mu^n)\}$ and of the decay estimate (50) does not make full use of the coefficients optimization step (Sect. 4.1.3) of Algorithm 1. Rather, the proof relies on energy estimates for the surrogate iterate $\mu^s := \mu^n + s(M\mu_{\gamma*} - \mu^n)$, where $\gamma^* \in AC^2$ is a solution of the insertion step (43), $M \geq 0$ a suitable constant, and the step-size s is chosen according to the Armijo–Goldstein condition (see [18, Section 5]). Since μ^s is a candidate for problem (44), the added coefficients optimization step in Algorithm 1 does not worsen the convergence rate.

Proof Fix $n \in \mathbb{N}$, and let $\{\mu^n\}$ be a sequence generated by Algorithm 1, which by construction is of the form (39). Recall that $w_t^n := -K_t(K_t^* \rho_t^n - f_t) \in C(\overline{\Omega})$ is the associated dual variable. Let $\mu^* = (\rho^*, m^*)$ in $\text{Ext}(C_{\alpha,\beta})$ be a solution to the insertion step, that is, μ^* solves (41). The existence of such μ^* is guaranteed by Proposition 4. Without loss of generality, we can assume that the algorithm does not stop at iteration n, that is, $\langle \rho^*, w^n \rangle > 1$ according to the stopping condition in Sect. 4.1.2. In particular, $\rho^* \neq 0$. Recalling that $\text{Ext}(C_{\alpha,\beta}) = \{0\} \cup C_{\alpha,\beta}$ (Theorem 1), we then have $\mu^* = \mu_{\gamma*}$ for some $\gamma^* \in AC^2$. By the coefficients optimization step, the next iterate is of the form

$$\mu^{n+1} = \sum_{j=1}^{N_n} c_j^{n+1} \mu_{\gamma_j^n} + c^* \mu_{\gamma^*},$$

and the coefficients $(c_1^{n+1}, \ldots, c_{N_n}^{n+1}, c^*)$ solve the quadratic problem

$$\min_{c=(c_1,\ldots,c_{N_n+1})\in\mathbb{R}_+^{N_n+1}} T_{\alpha,\beta}\left(\sum_{j=1}^{N_n} c_j \mu_{\gamma_j^n} + c_{N_n+1} \mu_{\gamma^*}\right). \tag{51}$$

Claim. There exists a constant $\tilde{C} > 0$ independent of n such that, if $T_{\alpha,\beta}(\mu^n) \leq M_0$, then

$$r(\mu^{n+1}) - r(\mu^n) \leq -\tilde{C}\, r(\mu^n)^2. \tag{52}$$

To prove the above claim, assume that $T_{\alpha,\beta}(\mu^n) \leq M_0$. Define $\hat{\mu} := M\mu_{\gamma*}$, with $M := M_0\langle \rho_{\gamma*}, w^n \rangle$. Since $\mu_{\gamma*}$ solves (41) and $\langle \rho_{\gamma*}, w^n \rangle > 1$, by Proposition 4, we have that $\hat{\mu}$ minimizes in (25) with respect to w^n. Therefore, according to (32),

$$G(\mu^n) = J_{\alpha,\beta}(\mu^n) - \varphi(J_{\alpha,\beta}(\hat{\mu})) - \langle \rho^n - \hat{\rho}, w^n \rangle, \tag{53}$$

given that $J_{\alpha,\beta}(\mu^n) \leq T_{\alpha,\beta}(\mu^n) \leq M_0 < +\infty$. For $s \in [0, 1]$, set $\mu^s := \mu^n + s(\hat{\mu} - \mu^n)$. By convexity of $J_{\alpha,\beta}$ (see Lemma 11), we have $J_{\alpha,\beta}(\mu^s) < +\infty$. Hence, we can

apply the polarization identity at (35) with respect to w^n, ρ^n, ρ^s to obtain

$$
T_{\alpha,\beta}(\mu^s) - T_{\alpha,\beta}(\mu^n)
$$
$$
= -s\langle \hat{\rho} - \rho^n, w^n \rangle + \frac{s^2}{2} \left\| K^*(\rho^n - \hat{\rho}) \right\|_{L_H^2}^2 + J_{\alpha,\beta}(\mu^s) - J_{\alpha,\beta}(\mu^n)
$$
$$
\leq -s\langle \hat{\rho} - \rho^n, w^n \rangle + \frac{s^2}{2} \left\| K^*(\rho^n - \hat{\rho}) \right\|_{L_H^2}^2 + s(\varphi(J_{\alpha,\beta}(\hat{\mu})) - J_{\alpha,\beta}(\mu^n)),
$$

where in the second line, we used convexity of $J_{\alpha,\beta}$ and the inequality $t \leq \varphi(t)$ for all $t \geq 0$. Note that μ^s is a competitor for (51), so that $T_{\alpha,\beta}(\mu^{n+1}) \leq T_{\alpha,\beta}(\mu^s)$. Recalling (53), we then obtain

$$
r(\mu^{n+1}) - r(\mu^n) = T_{\alpha,\beta}(\mu^{n+1}) - T_{\alpha,\beta}(\mu^n) \leq -sG(\mu^n) + \frac{s^2}{2} \left\| K^*(\hat{\rho} - \rho^n) \right\|_{L_H^2}^2,
$$
$$
\tag{54}
$$

which holds for all $s \in [0, 1]$. Choose the stepsize $\hat{s}$ according to the Armijo–Goldstein condition (see e.g., [51, Definition 4.1]) as

$$
\hat{s} := \min \left\{ 1, \frac{G(\mu^n)}{\left\| K^*(\hat{\rho} - \rho^n) \right\|_{L_H^2}^2} \right\},
\tag{55}
$$

with the convention that $C/0 = +\infty$ for $C > 0$. If $\hat{s} < 1$, by (54) and (34), we obtain

$$
r(\mu^{n+1}) - r(\mu^n) \leq -\frac{1}{2} \frac{G(\mu^n)^2}{\left\| K^*(\hat{\rho} - \rho^n) \right\|_{L_H^2}^2} \leq -\frac{1}{2} \frac{r(\mu^n)^2}{\left\| K^*(\hat{\rho} - \rho^n) \right\|_{L_H^2}^2}.
\tag{56}
$$

Since we are assuming $T_{\alpha,\beta}(\mu^n) \leq M_0$, by definition of $T_{\alpha,\beta}$, we deduce that $\|\rho^n\|_{\mathcal{M}(X)} \leq M_0/\alpha$. Moreover, since $\hat{\mu} = M\mu_{\gamma^*}$, by (K2) in Sect. 3.1, the estimate $a_{\gamma^*} \leq 1/\alpha$, and the Cauchy–Schwarz inequality yield

$$
\|\hat{\rho}\|_{\mathcal{M}(X)} = a_{\gamma^*}M \leq a_{\gamma^*}^2 M_0 \int_0^1 \left| \langle \delta_{\gamma^*(t)}, w_t^n \rangle \right| \, \mathrm{d}t \leq \frac{M_0}{\alpha^2} \int_0^1 \|w_t^n\|_{C(\overline{\Omega})} \, \mathrm{d}t
$$
$$
\leq \frac{M_0}{\alpha^2} \left(\int_0^1 \|w_t^n\|_{C(\overline{\Omega})}^2 \, \mathrm{d}t \right)^2 \leq \frac{M_0}{\alpha^2} C\sqrt{2} \left(\frac{1}{2} \|K^*\rho^n - f\|_{L_H^2}^2 \right)^{1/2}
$$
$$
\leq \frac{M_0}{\alpha^2} C\sqrt{2}\, T_{\alpha,\beta}(\mu^n)^{1/2} \leq \frac{M_0^{3/2}}{\alpha^2} C\sqrt{2},
$$

where $C > 0$ is the constant in (K2). Thus, we can estimate

$$
\|K^*(\hat{\rho} - \rho^n)\|_{L_H^2}^2 \leq 2C^2(\|\hat{\rho}\|_{\mathcal{M}(X)}^2 + \|\rho^n\|_{\mathcal{M}(X)}^2) \leq \tilde{C},
$$

with $\tilde{C} > 0$ not depending on n. Inserting the above estimate in (56) yields (52) and the claim follows. Assume now $\hat{s} = 1$, so that $\|K^*(\hat{\rho} - \rho^n)\|_{L_H^2}^2 \leq G(\mu^n)$. From (54)

and (34), we obtain

$$r(\mu^{n+1}) - r(\mu^n) \leq -\frac{1}{2} G(\mu^n) \leq -\frac{1}{2} r(\mu^n) \,. \tag{57}$$

As $T_{\alpha,\beta}(\mu^n) \leq M_0$, we also have $r(\mu^n) \leq T_{\alpha,\beta}(\mu^n) \leq M_0$. Therefore, we can find a constant $\tilde{C} > 0$ not depending on n such that $\tilde{C} r(\mu^n)^2 \leq r(\mu^n)$. Substituting the latter in (57) yields (52), and the proof of the claim is concluded.

Finally, we are in position to prove (50). Since $\mu^0 = 0$, we have that $T_{\alpha,\beta}(\mu^0) \leq M_0$. Thus, we can inductively apply (52) and obtain that $T_{\alpha,\beta}(\mu^n) \leq M_0$ for all $n \in \mathbb{N}$. In particular, (52) holds for every $n \in \mathbb{N}$ and, as a consequence, the sequence $\{r(\mu^n)\}$ is non-increasing. Setting $r_n := r(\mu^n)$, we then get

$$\frac{1}{r_n} - \frac{1}{r_0} = \sum_{j=0}^{n-1} \frac{1}{r_{j+1}} - \frac{1}{r_j} = \sum_{j=0}^{n-1} \frac{r_j - r_{j+1}}{r_j r_{j+1}} \geq \sum_{j=0}^{n-1} \frac{\tilde{C} r_j^2}{r_j r_{j+1}} \geq \tilde{C} n \,,$$

and (50) follows. The remaining claims follow from the weak* lower semicontinuity of $T_{\alpha,\beta}$ (Theorem 3) and estimate (A2), given that $\{\mu^n\}$ is a minimizing sequence for ($\mathcal{P}$). $\qquad\square$

4.4 Stopping Condition

We prove the optimality statements in the stopping criterion for the core algorithm anticipated in Sect. 4.1.2. In the following, G denotes the primal-dual gap introduced in (32). We denote by μ^n the n-th iterate of Algorithm 1, which is of the form (39), and by w^n the corresponding dual variable (40). Moreover, let μ_{γ^*} be the atom associated with the curve $\gamma^* \in \mathrm{AC}_\infty^2$ solving the insertion step (43).

Lemma 8 *For all $n \in \mathbb{N}$, we have*

$$G(\mu^n) = \Lambda(\langle \rho_{\gamma^*}, w^n \rangle), \qquad \Lambda(t) := \begin{cases} 0 & \text{if } t \leq 1, \\ \frac{M_0}{2}(t^2 - 1) & \text{if } t > 1, \end{cases} \tag{58}$$

where $M_0 \geq 0$ is defined at (21). In particular, μ^n is a solution of ($\mathcal{P}$) if and only if $\langle \rho_{\gamma^*}, w^n \rangle \leq 1$.

Before proving Lemma 8, we give a quantitative version of the stopping condition of Sect. 4.1.2.

Remark 7 With the same notations as above, consider the condition

$$\Lambda(\langle \rho_{\gamma^*}, w^n \rangle) < \mathrm{TOL} \tag{59}$$

where $\mathrm{TOL} > 0$ is a fixed tolerance. Notice that $G(\mu^n) = \Lambda(\langle \rho_{\gamma^*}, w^n \rangle)$ by Lemma 8. Thus, assuming (59), and using (34), we see that the functional residual defined at (33) satisfies $r(\mu^n) < \mathrm{TOL}$, i.e., μ^n almost minimizes ($\mathcal{P}$), up to the tolerance.

Therefore, the condition at (59) can be employed as a quantitative stopping criterion for Algorithm 1.

Proof of Lemma 8 We start by computing $G(\mu^n)$ for a fixed $n \in \mathbb{N}$. Set $\hat{\mu} := M \mu_{\gamma^*}$, where M is defined as in (27) with $w = w^n$. Since $\mu_{\gamma^*} \in \text{Ext}(C_{\alpha,\beta})$ solves (41), we have that $\hat{\mu}$ solves (25) with respect to w^n (see Proposition 4). By Theorem 7, the sequence $\{T_{\alpha,\beta}(\mu^n)\}$ is non-increasing. Thus, $J_{\alpha,\beta}(\mu^n) \le T_{\alpha,\beta}(\mu^n) \le T_{\alpha,\beta}(\mu^0) = M_0 < +\infty$, since $\mu^0 = 0$. Then, (32) reads

$$G(\mu^n) = J_{\alpha,\beta}(\mu^n) - \langle \rho^n, w^n \rangle - \varphi(J_{\alpha,\beta}(\hat{\mu})) + \langle \hat{\rho}, w^n \rangle. \tag{60}$$

Notice that by one-homogeneity of $J_{\alpha,\beta}$ (see Lemma 11) and the fact that $\mu_{\gamma^*} \in \text{Ext}(C_{\alpha,\beta})$, we have $J_{\alpha,\beta}(\hat{\mu}) = M$. Recalling the definition of φ at (20), by direct calculation, we obtain

$$- \varphi(J_{\alpha,\beta}(\hat{\mu})) + \langle \hat{\rho}, w^n \rangle = \Lambda(\langle \rho_{\gamma^*}, w^n \rangle). \tag{61}$$

We now compute the remaining terms in (60). By (44) and (45) at the step $n - 1$, we know that the coefficients $c^n = (c_1^n, \ldots, c_{N_n}^n) \in \mathbb{R}_{++}^{N_n}$ of μ^n solve the minimization problem

$$\min_{c=(c_1,\ldots,c_{N_n}) \in \mathbb{R}_{++}^{N_n}} T_{\alpha,\beta}\left(\sum_{j=1}^{N_n} c_j \mu_{\gamma_j^n} \right). \tag{62}$$

Since γ_j^n in (39) is continuous, we have that $(t \mapsto \delta_{\gamma_j^n(t)}) \in C_{\text{w}}^+$ and thus $t \mapsto K_t^* \delta_{\gamma_j^n(t)}$ belongs to L_H^2, by Lemma 12. In view of (39), definition of $T_{\alpha,\beta}$ and Lemma 2, we can expand the expression at (62), differentiate with respect to each component of c, and recall that c^n is optimal in (62), to obtain

$$0 = \sum_{j=1}^{N_n} c_j^n a_{\gamma_i^n} a_{\gamma_j^n} \int_0^1 \langle K_t^* \delta_{\gamma_i^n(t)}, K_t^* \delta_{\gamma_j^n(t)} \rangle_{H_t} \, \mathrm{d}t - a_{\gamma_i^n} \int_0^1 \langle K_t^* \delta_{\gamma_i^n(t)}, f_t \rangle_{H_t} \, \mathrm{d}t + 1$$
$$= \langle K^* \rho_{\gamma_i^n}, K^* \rho^n \rangle_{L_H^2} - \langle K^* \rho_{\gamma_i^n}, f \rangle_{L_H^2} + 1 = 1 - \langle \rho_{\gamma_i^n}, w^n \rangle, \tag{63}$$

which holds for all $i = 1, \ldots, N_n$. By Lemma 2 and linearity of $\langle \cdot, \cdot \rangle$, we obtain the identity $J_{\alpha,\beta}(\mu^n) = \langle \rho^n, w^n \rangle$. The latter, together with (60) and (61), yields (58). For the remaining part of the statement, notice that by (58), we have $G(\mu^n) = 0$ if and only if $\langle \rho_{\gamma^*}, w^n \rangle \le 1$. Therefore, the thesis follows by Lemma 5. $\qquad \square$

4.5 Time-Discrete Version

The minimization problem ($\mathcal{P}$) presented in Sect. 3 is posed for time-continuous measurements $f \in L_H^2$, whose discrepancy to the reconstructed curve $t \mapsto \rho_t$ is modeled by the fidelity term $\mathcal{F}$ at (16). In real-world applications, however, the measured data are time-discrete, i.e., we can assume that measurements are taken at times

$0 = t_0 < t_1 < \ldots < t_T = 1$, with $T \in \mathbb{N}$ fixed. Hence, the data are of the form $f = (f_{t_0}, f_{t_1}, \ldots, f_{t_T})$, where $f_{t_i} \in H_{t_i}$.

For this reason, and with the additional goal of lowering the computational cost, we decided to present numerical experiments (Sect. 6) where the time-continuous fidelity term $\mathcal{F}$ is replaced by a discrete counterpart. In Sect. 4.5.1, we show how to modify the mathematical framework discussed so far, in order to deal with the time-discrete case. Consequently, it is immediate to adapt Algorithm 1 to the resulting time-discrete functional, as discussed in Sect. 4.5.2. We remark that all the results up to this point will hold, in a slightly modified version, also for the discrete setting discussed below.

4.5.1 Time-Discrete Framework

We replace problem ($\mathcal{P}$) with

$$\min_{\mu \in \mathcal{M}} T_{\alpha,\beta}^{\mathcal{D}}(\mu), \quad T_{\alpha,\beta}^{\mathcal{D}}(\mu) := \mathcal{F}^{\mathcal{D}}(\mu) + J_{\alpha,\beta}(\mu), \qquad (\mathcal{P}_{\text{discr}})$$

where the time-discrete fidelity term $\mathcal{F}^{\mathcal{D}} \colon \mathcal{M} \to [0, +\infty]$ is defined by

$$\mathcal{F}^{\mathcal{D}}(\mu) := \begin{cases} \dfrac{1}{2(T+1)} \displaystyle\sum_{i=0}^{T} \left\| K_{t_i}^* \rho_{t_i} - f_{t_i} \right\|_{H_{t_i}}^2 & \text{if } \rho = \mathrm{d}t \otimes \rho_t, \ (t \mapsto \rho_t) \in C_{\mathrm{w}}, \\ +\infty & \text{otherwise.} \end{cases}$$

Here, H_{t_i} are real Hilbert spaces, and the given data vector $f = (f_{t_0}, \ldots, f_{t_T})$ satisfies $f_{t_i} \in H_{t_i}$ for all $i = 0, \ldots, T$. The forward operators $K_{t_i}^* \colon \mathcal{M}(\overline{\Omega}) \to H_i$ are assumed to be linear continuous and weak*-to-weak continuous, for each $i = 0, \ldots, T$. The minimization problem ($\mathcal{P}_{\text{discr}}$) is well-posed by the direct method of calculus of variations: indeed $T_{\alpha,\beta}^{\mathcal{D}}$ is proper, $J_{\alpha,\beta}$ is weak* lower semicontinuous and coercive in $\mathcal{M}$ (Lemma 11), and $\mathcal{F}^{\mathcal{D}}$ is lower semicontinuous with respect to the convergence in (A3). In particular, a solution $\mu^* = (\rho^*, m^*)$ to ($\mathcal{P}_{\text{discr}}$) will satisfy $\rho^* = \mathrm{d}t \otimes \rho_t^*$ with $(t \mapsto \rho_t^*) \in C_{\mathrm{w}}^+$. We now define the other quantities which are needed to formulate the discrete counterpart of the theory developed so far. For a given curve of measures $(t \to \tilde{\rho}_t) \in C_{\mathrm{w}}$, the corresponding dual variable (22) is redefined to be

$$w_{t_i} := -K_{t_i}(K_{t_i}^* \tilde{\rho}_{t_i} - f_{t_i}) \in C(\overline{\Omega}), \qquad (64)$$

for each $i = 0, \ldots, T$. Consequently, we redefine the associated scalar product (23) to

$$\langle \rho, w \rangle_{\mathcal{D}} := \begin{cases} \dfrac{1}{T+1} \displaystyle\sum_{i=0}^{T} \langle \rho_{t_i}, w_{t_i} \rangle_{\mathcal{M}(\overline{\Omega}), C(\overline{\Omega})} & \text{if } \rho = \mathrm{d}t \otimes \rho_t, \ (t \mapsto \rho_t) \in C_{\mathrm{w}}, \\ -\infty & \text{otherwise.} \end{cases}$$

It is straightforward to check that all the results in Sect. 3 hold with $T_{\alpha,\beta}$ and $\langle \cdot, \cdot \rangle$ replaced by $T^{\mathcal{D}}_{\alpha,\beta}$ and $\langle \cdot, \cdot \rangle_{\mathcal{D}}$, respectively, with the obvious modifications. In particular, the problem

$$\min_{\mu \in \mathrm{Ext}(C_{\alpha,\beta})} -\langle \rho, w \rangle_{\mathcal{D}} \tag{65}$$

admits a solution μ^*, where $C_{\alpha,\beta} := \{J_{\alpha,\beta}(\mu) \le 1\}$. One can perform a similar computation to the one at (42) to obtain equivalence between (65) and

$$\min_{\gamma \in \mathrm{AC}^2_\infty} -\langle \rho_\gamma, w \rangle_{\mathcal{D}} = \min_{\gamma \in \mathrm{AC}^2} \left\{ 0, -\frac{a_\gamma}{T+1} \sum_{i=0}^{T} w_{t_i}(\gamma(t_i)) \right\}. \tag{66}$$

Remark 8 Assume additionally that $\overline{\Omega}$ is convex. Then, a solution $\gamma^* \in \mathrm{AC}^2_\infty$ to problem (66) is either $\gamma^* = \gamma_\infty$, or $\gamma^* \in \mathrm{AC}^2$ with γ^* linear in each interval $[t_i, t_{i+1}]$. Indeed, given any curve $\gamma \in \mathrm{AC}^2$, denote by $\tilde{\gamma}$ the piecewise linear version of γ sampled at t_i for $i = 0, \ldots, T$. Then $\langle \rho_\gamma, w \rangle_{\mathcal{D}} \le \langle \rho_{\tilde{\gamma}}, w \rangle_{\mathcal{D}}$, due to the inequality $\int_0^1 \left| \dot{\tilde{\gamma}}(t) \right|^2 \, \mathrm{d}t \le \int_0^1 |\dot{\gamma}(t)|^2 \, \mathrm{d}t$.

4.5.2 Adaption of Algorithm 1 to the Time-Discrete Setting

The core algorithm in Sect. 4.1 is readily adaptable to the task of minimizing ($\mathcal{P}_{\mathrm{discr}}$). To this end, let $f_{t_i} \in H_{t_i}$ be a given datum and $\alpha, \beta > 0$ be fixed parameters for ($\mathcal{P}_{\mathrm{discr}}$). Assume that μ^n is the current sparse iterate, of the form (39). The dual variable associated with μ^n is defined, according to (64), by $w^n_{t_i} := -K_{t_i}(K^*_{t_i}\rho^n_{t_i} - f_{t_i})$. Similarly to Sect. 4.1.2, the insertion step in the time-discrete version consists in finding a curve $\gamma^* \in \mathrm{AC}^2_\infty$ which solves (66) with respect to w^n_i. Such curve defines a new atom μ_{γ^*} according to Definition 2. Adapting the proofs of Lemmas 5, 8 to the discrete setting, one deduces the following stopping condition:

- if $\langle \rho_{\gamma^*}, w^n \rangle_{\mathcal{D}} \le 1$, then μ^n is solution to ($\mathcal{P}_{\mathrm{discr}}$). The algorithm outputs μ^n and stops,
- if $\langle \rho_{\gamma^*}, w^n \rangle_{\mathcal{D}} > 1$, then μ^n is not a solution to ($\mathcal{P}_{\mathrm{discr}}$) and $\gamma^* \in \mathrm{AC}^2$. The found atom μ_{γ^*} is inserted in the n-th iterate μ^n and the algorithm continues.

Set $\gamma^n_{N_n+1} := \gamma^*$. The time-discrete version of the coefficients optimization step discussed in Sect. 4.1.3 consists in solving

$$\min_{(c_1, c_2, \ldots, c_{N_n+1}) \in \mathbb{R}^{N_n+1}_+} T^{\mathcal{D}}_{\alpha,\beta} \left(\sum_{j=1}^{N_n+1} c_j \mu_{\gamma^n_j} \right). \tag{67}$$

By proceeding as in Sect. 4.2, one can check that (67) is equivalent to a quadratic program of the form (46), where the matrix $\Gamma \in \mathbb{R}^{(N_n+1) \times (N_n+1)}$ and the vector

$b \in \mathbb{R}^{N_n+1}$ are given by

$$\Gamma_{j,k} := \frac{a_{\gamma_j} a_{\gamma_k}}{T+1} \sum_{i=0}^{T} \langle K_{t_i}^* \delta_{\gamma_j(t_i)}, K_{t_i}^* \delta_{\gamma_k(t_i)} \rangle_{H_{t_i}},$$

$$b_j := 1 - \frac{a_{\gamma_j}}{T+1} \sum_{i=0}^{T} \langle K_{t_i}^* \delta_{\gamma_j}(t_i), f_{t_i} \rangle_{H_{t_i}}.$$

In view of Remark 8 and of the above construction, we note that the iterates of the discrete algorithms are of the form (39) with $\gamma_j^n \in AC^2$ piecewise linear. Finally, we remark that the time-discrete algorithm obtained with the above modifications has the same sublinear rate of convergence stated in Theorem 7.

5 The Algorithm: Numerical Implementation

This aim of this section is twofold. First, in Sect. 5.1, we describe how to approach the minimization of the insertion step problem (43) by means of gradient descent strategies. This analysis is performed under additional assumptions on the operators $K_t \colon H_t \to C(\overline{\Omega})$, which, loosely speaking, require that K_t map into the space $C^{1,1}(\overline{\Omega})$ of differentiable functions with bounded Lipschitz gradient. The strategies proposed will result in the *multistart gradient descent* Subroutine 1 (Sect. 5.1.5), which, given a dual variable w, outputs a set of stationary points for (43). Then, in Sect. 5.2, we present two acceleration steps that can be added to Algorithm 1 to, in principle, enhance its performance. The first acceleration strategy, called the *multiple insertion step*, proceeds by adding all the outputs of Subroutine 1 to the current iterate. The second strategy, termed *sliding step*, consists in locally descending the target functional $T_{\alpha,\beta}$ at ($\mathcal{P}$) in a neighborhood of the curves composing the current iterate, while keeping the coefficients fixed. These strategies are finally added to Algorithm 1. The outcome is Algorithm 2, presented in Sect. 5.3, which we name *dynamic generalized conditional gradient* (DGCG).

5.1 Insertion Step Implementation

We aim at minimizing the linearized problem (43) in the insertion step, which is of the form

$$\min_{\gamma \in H^1([0,1];\overline{\Omega})} \{0, F(\gamma)\}, \quad F(\gamma) := -a_\gamma \int_0^1 w_t(\gamma(t)) \, dt, \tag{68}$$

where the dual variable $t \mapsto w_t$ is defined for a.e. $t \in (0, 1)$ by $w_t := -K_t(K_t^* \tilde{\rho}_t - f_t) \in C(\overline{\Omega})$, for some curve $(t \mapsto \tilde{\rho}_t) \in C_w$ and data $f \in L_H^2$ fixed. In (68), we also identified AC^2 with $H^1([0, 1]; \overline{\Omega})$, and employed the notations at (14). We remind the reader that although (68) admits solutions (see Sect. 4.1.1), in practice, they may be difficult to compute numerically (Remark 4). Therefore, we turn our attention at finding stationary points for the functional F at (68), relying on gradient descent methods.

To make this approach feasible, we require additional assumptions on the operators K_t^* (see Assumption 3), which allow to extend the functional F to the Hilbert space $H^1 := H^1([0, 1]; \mathbb{R}^d)$ and make it Fréchet differentiable, without altering the value of the minimum at (68). In particular, this allows for a gradient descent procedure to be well-defined (Sect. 5.1.1). With this at hand, in Sect. 5.1.2, we define a descent operator $\mathcal{G}_w$ associated with F, which, for a starting curve $\gamma \in H^1$, outputs either a stationary point of F or the infinite length curve γ_∞. The starting curves for $\mathcal{G}_w$ are of two types:

- random starts which are guided by the values of the dual variable w (Sect. 5.1.3),
- crossovers between known stationary points (Sect. 5.1.4).

We then propose a minimization strategy for (68), which is implemented in the *multistart gradient descent algorithm* contained in Subroutine 1 (Sect. 5.1.5). Such algorithm inputs a set of curves and a dual variable w, and returns a set $\mathcal{S} \subset \mathrm{AC}^2$, which is either empty or contains stationary curves for F at (68).

5.1.1 Minimization Strategy

The additional assumptions required on K_t^* are as follows.

Assumption 3 For a.e. $t \in (0, 1)$, the linear continuous operator $K_t^*: C^{1,1}(\overline{\Omega})^* \to H_t$ satisfies

- (F1) K_t^* is weak*-to-weak continuous, with pre-adjoint denoted by $K_t: H_t \to C^{1,1}(\overline{\Omega})$,
- (F2) $\left\| K_t^* \right\| \leq C$ for some constant $C > 0$ not depending on t,
- (F3) the map $t \mapsto K_t^* \rho$ is strongly measurable for every fixed $\rho \in \mathcal{M}(\overline{\Omega})$,
- (F4) there exists a closed convex set $E \Subset \Omega$ such that $\mathrm{supp}(K_t f), \mathrm{supp}(\nabla(K_t f)) \subset E$ for all $f \in H_t$ and a.e. $t \in (0, 1)$, where ∇ denotes the spatial gradient.

Notice that (F1)–(F3) imply (K1)–(K3) of Sect. 3.1, due to the embedding $\mathcal{M}(\overline{\Omega}) \hookrightarrow C^{1,1}(\overline{\Omega})^*$. Also note that (F4) has no counterpart in the assumptions of Sect. 3.1, and is only assumed for computational convenience, as discussed below. The problem of solving (68) under (F1)–(F4) is addressed in Appendix 1; here, we summarize the main results obtained. First, we extend F to the Hilbert space H^1, by setting $w_t(x) := 0$ for all $x \in \mathbb{R}^d \setminus \overline{\Omega}$ and a.e. $t \in (0, 1)$. Due to (F4), we have that $w_t \in C^{1,1}(\mathbb{R}^d)$. We then show that F is continuously Fréchet differentiable on H^1, with locally Lipschitz derivative (Proposition 16).

Denote by $D_\gamma F \in (H^1)^*$, the Fréchet derivative of F at γ (see (A13) for the explicit computation) and introduce the set of stationary points of F with nonzero energy

$$\mathfrak{S}_F := \{\gamma \in H^1 : F(\gamma) \neq 0, \ D_\gamma F = 0\}.$$

In Proposition 17, we prove that all the points in $\mathfrak{S}_F$ satisfy $\gamma([0, 1]) \subset \Omega$. As a consequence, $\mathfrak{S}_F$ contains all the solutions to the insertion problem (68) whenever

$$\min_{\gamma \in H^1([0,1];\overline{\Omega})} \{0, F(\gamma)\} < 0. \tag{69}$$

Remark 9 The case (69) is the only one of interest: indeed Algorithm 1 stops if (69) is not satisfied, with the current iterate being a solution to the target problem ($\mathcal{P}$) (see Sect. 4.1.2). In Proposition 19, we prove that (69) is equivalent to

$$P(w) := \int_0^1 \max_{x \in \overline{\Omega}} w_t(x) \, \mathrm{d}t > 0 \,. \tag{70}$$

As $P(w)$ is easily computable, condition (70) provides an implementable test for (69). Moreover, note that (69) is satisfied when F is computed from the dual variable w^n associated with the n-th iterate $\mu^n = \sum_{j=1}^{N_n} c_j^n \delta_{\gamma_j^n}$ of Algorithm 1, and $n \geq 1$. This is because $F(\gamma_j^n) = -1$ for all $j = 1, \ldots, N_n$, as shown in (63).

If (69) is satisfied, we aim at computing points in $\mathfrak{S}_F$ by gradient descent. To this end, we say that $\{\gamma^n\}$ in H^1 is a *descent sequence* if

$$F(\gamma^0) < 0 \,, \quad \gamma^{n+1} = \gamma^n - \delta_n D_{\gamma^n} F \,, \quad \text{for all} \ \ n \in \mathbb{N} \cup \{0\} \,, \tag{71}$$

where $D_{\gamma^n} F \in (H^1)^*$ is identified with its Riesz representative in H^1, and $\{\delta_n\}$ is a stepsize chosen according to the Armijo–Goldstein or Backtracking-Armijo rules. In Theorem 14, we prove that if $\{\gamma^n\}$ is a descent sequence, there exists at least a subsequence such that $\gamma^{n_k} \to \gamma^*$ strongly in H^1; moreover, any such accumulation point γ^* belongs to $\mathfrak{S}_F$. To summarize, descent sequences in the sense of (71) enable us to compute points in $\mathfrak{S}_F$, which are candidate solutions to (68) whenever (69) holds.

5.1.2 Descent Operator

Fix a dual variable w and consider the functional F defined as in (68). The descent operator $\mathcal{G}_w \colon H^1([0, 1]; \overline{\Omega}) \to H^1 \cup \{\gamma_\infty\}$ associated with w is defined by

$$\mathcal{G}_w(\gamma) := \begin{cases} \gamma^* & \text{if} \ \ F(\gamma) < 0 \,, \\ \gamma_\infty & \text{otherwise} \,, \end{cases} \tag{72}$$

where γ^* is an accumulation point for the descent sequence $\{\gamma^n\}$ defined according to (71) with starting point $\gamma^0 := \gamma$. In view of the discussion in the previous section, we know that the image of $\mathcal{G}_w$ is contained in $\mathfrak{S}_F \cup \{\gamma_\infty\}$. Note that if $P(w) > 0$, in principle, the image of $\mathcal{G}_w$ will contain at least one stationary point $\gamma^* \in \mathfrak{S}_F$, as in this case (69) holds (Remark 9). However, in simulations, we can only compute $\mathcal{G}_w$ on some finite family of curves $\{\gamma_i\}_{i \in I}$, which we name the *starts*. Thus, in general, we have no guarantee of finding points in $\mathfrak{S}_F$, even if (69) holds. The situation improves if $w = w^n$ is the dual variable associated with the n-th iterate $\mu^n = \sum_{j=1}^{N_n} c_j^n \delta_{\gamma_j^n}$ of Algorithm 1 and $n \geq 1$. In this case, setting $\mathcal{A} := \{\gamma_1^n, \ldots, \gamma_{N_n}^n\}$, we have that $\mathcal{G}_{w^n}(\mathcal{A}) \subset \mathfrak{S}_F$ by definition (72) and Remark 9. Therefore, by including the curves in $\mathcal{A}$ in the set of considered starts, we are guaranteed of obtaining at least one point in $\mathfrak{S}_F$.

5.1.3 Random Starts

We now describe how we randomly generate starting points γ in $H^1([0, 1]; \overline{\Omega})$ for the descent operator $\mathcal{G}$ at (72). We start by selecting time nodes $0 = t_0 < t_1 < \ldots < t_T = 1$ drawn uniformly in $[0, 1]$. (If operating in the time-discrete setting, we sample instead with a uniform probability on the finite set of sampling times on which the fidelity term of ($\mathcal{P}_{\text{discr}}$) is defined.) We choose the value of a random start γ at time t_i seeking to maximize the dual variable w_{t_i}. To achieve this, let $Q : \mathbb{R} \to \mathbb{R}_+$ be non-decreasing and monotonous, and define the probability measure on E

$$\mathbb{P}_{w_{t_i}}(A) := \frac{\int_A Q(w_{t_i}(x)) \, \mathrm{d}x}{\int_\Omega Q(w_{t_i}(x)) \, \mathrm{d}x} \, ,$$

for $A \subset E$ Borel measurable and $E \Subset \Omega$ introduced in Assumption 3. We then draw samples from $\mathbb{P}_{w_{t_i}}$ with the rejection-sampling algorithm, and assign those samples to $\gamma(t_i)$. Using that E is a convex set, the random curve $\gamma \in \text{AC}^2$ is obtained by interpolating linearly the values $\gamma(t_i) \in E$. This procedure is executed by the routine sample, which inputs a dual variable w and outputs a randomly generated curve $\gamma \in \text{AC}^2$.

5.1.4 Crossovers Between Stationary Points

It is heuristically observed that stationary curves have a tendency to share common "routes," as for example seen in the reconstructions presented in Figs. 4 and 5. It is then a reasonable ansatz to combine curves which are sharing routes, in order to increase the likelihood for the newly obtained crossovers to share common routes with the sought global minimizers of F. Such crossovers will then be employed as starts for the descent operator G at (72). Formally, the crossover is achieved as follows. We fix small parameters $\varepsilon > 0$ and $0 < \delta < 1$. For $\gamma_1, \gamma_2 \in \text{AC}^2$ define the set

$$R_\varepsilon(\gamma_1, \gamma_2) := \{t \in [0, 1] : \ |\gamma_1(t) - \gamma_2(t)| < \varepsilon\} \, .$$

We say that γ_1 and γ_2 share routes if $R_\varepsilon(\gamma_1, \gamma_2) \neq \emptyset$. If $R_\varepsilon(\gamma_1, \gamma_2) = \emptyset$, we perform no operations on γ_1 and γ_2. If instead $R_\varepsilon(\gamma_1, \gamma_2) \neq \emptyset$, first notice that $R_\varepsilon(\gamma_1, \gamma_2)$ is relatively open in $[0, 1]$. Denote by I any of its connected components. Then, I is an interval with endpoints t^- and t^+, satisfying $0 \leq t^- < t^+ \leq 1$. The crossovers of γ_1 and γ_2 in I are the two curves $\gamma_3, \gamma_4 \in \text{AC}^2$ defined by

$$\gamma_3(t) := \begin{cases} \gamma_1(t), & t \in [0, \hat{t} - \delta\tilde{t}] \\ \gamma_2(t), & t \in [\hat{t} + \delta\tilde{t}, 1] \end{cases} , \qquad \gamma_4(t) := \begin{cases} \gamma_2(t), & t \in [0, \hat{t} - \delta\tilde{t}] \\ \gamma_1(t), & t \in [\hat{t} + \delta\tilde{t}, 1] \end{cases}$$

and linearly interpolated in $(\hat{t} - \delta\tilde{t}, \hat{t} + \delta\tilde{t})$, where $\hat{t} := (t^+ + t^-)/2$ and $\tilde{t} := (t^+ - t^-)/2$, i.e.,

$$\frac{t - (\hat{t} - \delta\tilde{t})}{2\delta\tilde{t}} \gamma_2(\hat{t} + \delta\tilde{t}) - \frac{t - (\hat{t} + \delta\tilde{t})}{2\delta\tilde{t}} \gamma_1(\hat{t} - \delta\tilde{t}), \quad t \in (\hat{t} - \delta\tilde{t}, \hat{t} + \delta\tilde{t})$$

is, for instance, the result of the linear interpolation of γ_3 in $(\hat{t}-\delta\tilde{t}, \hat{t}+\delta\tilde{t})$. We construct the crossovers of γ_1 and γ_2 in each connected component of $R_\varepsilon(\gamma_1, \gamma_2)$ obtaining $2M$ new curves, with M being the number of connected components of $R_\varepsilon(\gamma_1, \gamma_2)$. The described procedure is executed by the routine `crossover`, which inputs two curves $\gamma_1, \gamma_2 \in AC^2$ and outputs a set of curves in AC^2, possibly empty.

5.1.5 Multistart Gradient Descent Algorithm

In Subroutine 1, we sketch the proposed method to search for a minimizer of (68): this is implemented in the function `MultistartGD`, which inputs a set of curves $\mathcal{A}$ and a dual variable w, and outputs a (possibly empty) set $\mathcal{S}$ of stationary points for F defined at (68). We now describe how to interpret such subroutine in the context of Algorithm 1, and how it can be used to replace the insertion step operation at line 4.

Given the n-th iteration $\mu^n = \sum_{j=1}^{N_n} c_j^n \delta_{\gamma_j^n}$ of Algorithm 1, we define $\mathcal{A} :=$ $\{\gamma_1^n, \ldots, \gamma_{N_n}^n\}$ if $n \geq 1$ and $\mathcal{A} := \emptyset$ if $n = 0$. The dual variable w^n is as in (40). We initialize to empty the sets $\mathcal{S}$ and $\mathcal{O}$ of known stationary and crossover points, respectively. The condition at line 2 of Subroutine 1 checks if we are at the 0-th iteration and $P(w^0) \leq 0$. In case this is satisfied, then 0 is the minimum of (68), and no stationary point is returned; indeed in this situation, Algorithm 1 stops, with $\mu^0 = 0$ being the minimum of ($\mathcal{P}$) (see Sect. 5.1.1). Otherwise, the set $\mathcal{A}$ (possibly empty) is inserted in $\mathcal{O}$. Then, $N_{\max}$ initializations of the multistart gradient descent are performed, where a starting point γ is either chosen from the crossover set $\mathcal{O}$, if the latter is nonempty, or sampled at random by the function `sample` described in Sect. 5.1.3. We then descend γ, obtaining the new curve $\gamma^* := \mathcal{G}_{w^n}(\gamma)$, where $\mathcal{G}_{w^n}$ is defined at (72). If the outputted point γ^* does not belong to the set of known stationary points $\mathcal{S}$, and $\gamma^* \neq \gamma_\infty$, then we first compute the crossovers of γ^* with all the elements of $\mathcal{S}$, and afterward insert it in $\mathcal{S}$. After $N_{\max}$ iterations, the set $\mathcal{S}$ is returned. Notice that $\mathcal{S}$ could be empty only if $\mathcal{A} = \emptyset$, i.e., if `MultistartGD` is called at the first iteration of Algorithm 1 (see Sect. 5.1.2).

The modified insertion step, that is, line 4 of Algorithm 1, reads as follows. First, we set $\mathcal{A} = \boldsymbol{\gamma}$ and compute $\mathcal{S} := $ `MultistartGD`$(\mathcal{A}, w^n)$. If $\mathcal{S} = \emptyset$, the algorithm stops and returns the current iterate μ^n. Otherwise, the element in $\mathcal{S}$ with minimal energy with respect to F is chosen as candidate minimizer, and inserted as γ^* in line 4.

5.2 Acceleration Strategies

In this section, we describe two acceleration strategies that can be incorporated in Algorithm 1.

5.2.1 Multiple Insertion Step

This is an extension of the insertion step for Algorithm 1 described in Sect. 4.1.2. Precisely, the multiple insertion step consists in inserting into the current iterate μ^n all the atoms associated with the stationary points in the set $\mathcal{S}$ produced by Subroutine

---

Subroutine 1: Multistart gradient descent for the insertion step

Function: `MultistartGD`
Input: Set of curves $\mathcal{A}$, dual variable $w_t \in C^{1,1}(\overline{\Omega})$ with supp $w_t \subset \Omega$

```
1  S := ∅,  O := ∅ // Sets of known stationary and crossover points
2  if A == ∅ and P(w) ≤ 0 then
3  │  return S
4  O ← A
5  for k = 1, ..., Nmax do
      /* Restart from a random curve or a crossover one            */
6  │  if O == ∅ then
7  │  │  γ ← sample(w)
8  │  else
9  │  │  γ ← get from O, O ← O \ {γ}
      /* Descend, crossover, and incorporate to the stationary points
         set                                                       */
10 │  γ* ← Gw(γ)
11 │  if γ* ∉ S and γ* ≠ γ∞ then
12 │  │  for η ∈ S do
          /* crossover with all known stationary points           */
13 │  │  │  O ← O ∪ crossover(γ*, η)
14 │  │  S ← S ∪ {γ*}
15 return S
```

---

1 with respect to the dual variable w^n. This procedure is motivated by the following observations. First, computationally speaking, the coefficients optimization step described in Sect. 4.1.3 is cheap and fast. Second, it is observed that stationary points are good candidates for the insertion step in the GCG method presented in [51] to solve (4). This observation can be extended similarly to our framework, noticing that the addition of multiple stationary points is encouraging the iterates to concentrate around every atom of the ground-truth measure $\mu^\dagger$, and consequently, the algorithm could need fewer iterations to efficiently locate the support of $\mu^\dagger$.

5.2.2 Sliding Step

The sliding step proposed in this paper is a natural extension of the one introduced for BLASSO in [18] and further analyzed in [30]. Precisely, given a current iterate $\mu = \sum_{j=1}^{N} c_j \mu_{\gamma_j}$ of Algorithm 1, we fix the weights $\boldsymbol{c} = (c_1, \ldots, c_N)$ and define the functional $T_{\alpha,\beta,\boldsymbol{c}} : (H^1)^N \to \mathbb{R}$ as

$$T_{\alpha,\beta,\boldsymbol{c}}(\eta_1, \ldots, \eta_N) := T_{\alpha,\beta}\left(\sum_{j=1}^{N} c_j \mu_{\eta_j}\right) \quad \text{for all} \quad (\eta_1, \ldots, \eta_N) \in \left(H^1\right)^N. \quad (73)$$

We then perform additional gradient descent steps in the space $(H^1)^N$ for the functional $T_{\alpha,\beta,\boldsymbol{c}}$, starting from the tuple of curves $(\gamma_1, \ldots, \gamma_N)$ contained in the current iterate μ. Formally, this procedure is possible; in Proposition 20, we prove that $T_{\alpha,\beta,\boldsymbol{c}}$ is

continuously Fréchet differentiable in $(H^1)^N$ under Assumption 3, with derivative given by (A42). This step can be intertwined with the coefficients optimization one, by alternating between modifying the position of the current curves, and optimizing their associated weights.

5.3 Full Algorithm

By including Subroutine 1 and the proposed acceleration steps of Sect. 5.2 into Algorithm 1, we obtain Algorithm 2, which we name the *dynamic generalized conditional gradient* (DGCG).

5.3.1 Algorithm Summary

We employ the notations of Sect. 4.1.4. In particular, Algorithm 2 generates, as iterates, tuples of coefficients $c = (c_1, \ldots, c_N)$ with $c_j > 0$, and curves $\gamma = (\gamma_1, \ldots, \gamma_N)$ with $\gamma_j \in \mathrm{AC}^2$. We now summarize the main steps of Algorithm 2. The first 3 lines are unaltered from Algorithm 1, and they deal with tuples initializations and assembly of the measure iterate μ^n. The multiple insertion step is carried out by Subroutine 1, via the function MultistartGD at line 4, which is called with arguments γ and w^n. The output is a set of curves $\mathcal{S}$ which contains stationary points for the functional F at (68) with respect to the dual variable w^n. If $\mathcal{S} = \emptyset$, the algorithm stops and outputs the current iterate μ^n. As observed in Sect. 5.1.2, this can only happen at the first iteration of the algorithm. Otherwise, in line 7, the function order is employed to input $\mathcal{S}$ and output a tuple of curves, obtained by ordering the elements of $\mathcal{S}$ increasingly with respect to their value of F. As anticipated in Sect. 5.1, the first element of order($\mathcal{S}$), named $\gamma^*_{N_n+1}$, is considered to be the best available candidate solution to the insertion step problem (68), and is used in the stopping condition at lines 8, 9. Such stopping condition has been used similarly in Algorithm 1, with the difference that in Algorithm 2 the curve $\gamma^*_{N_n+1}$ is not necessarily the global minimum of F, but, in general, just a stationary point. We further discuss such stopping criterion in Sect. 5.3.2. After, the found stationary points are inserted in the current iterate, and the algorithm alternates between the coefficients optimization step (Sect. 4.1.3) and the sliding step (Sect. 5.2.2). Such operations are executed from line 11 to 16 in Algorithm 2, for $K_{\max}$ times.

5.3.2 Stopping Condition

The stopping condition for Algorithm 2 is implemented in lines 8, 9: when $\gamma^*_{N_n+1}$ satisfies $\langle \rho_{\gamma^*_{N_n+1}}, w^n \rangle \leq 1$, the algorithm stops and outputs the current iterate μ^n. Due to the definition of F, such condition is equivalent to say that the Subroutine 1 has not been able to find any curve $\gamma^* \in \mathrm{AC}^2$ satisfying $\langle \rho_{\gamma^*}, w^n \rangle > 1$. The main difference when compared to the stopping condition for Algorithm 1 (Sect. 4.1.2) is that in Algorithm 2, the curve $\gamma^*_{N_n+1}$ is generally not a global minimum of F. As a consequence, Lemma 8 does not hold, and the condition $\langle \rho_{\gamma^*_{N_n+1}}, w^n \rangle \leq 1$ is not equivalent to the minimality of the current iterate μ^n for ($\mathcal{P}$). The evident drawback is

that Algorithm 2 could stop even if the current iterate does not solve ($\mathcal{P}$). However, it is at least possible to say that if Algorithm 2 continues after line 9, then the current iterate does not solve ($\mathcal{P}$). Thus, in this situation, the correct decision is taken. We remark that even if the stopping condition for Algorithm 2 does not ensure the minimality of the output, from a practical standpoint, if Subroutine 1 is employed with a high number of restarts, the reconstruction is satisfactory. We also point out that the condition at line 8 can be replaced by the quantitative condition defined at (59) in Remark 7, with some predefined tolerance.

5.3.3 Convergence and Numerical Residual

In the following, we will say that Algorithm 2 converges if `MultistartGD`($\boldsymbol{\gamma}, w^0$) = $\emptyset$, or if some iterate satisfies the stopping condition at line 8. In case of convergence, we will denote by μ^N the output value of Algorithm 2. We remind the reader that due to the discussion in Sect. 5.3.2, μ^N is considered to be an approximate solution for the minimization problem ($\mathcal{P}$). In order to analyze the convergence rate for Algorithm 2 numerically, we define the numerical residual as

$$\tilde{r}(\mu^n) := T_{\alpha,\beta}(\mu^n) - T_{\alpha,\beta}(\mu^N), \qquad n = 0, \ldots, N-1, \tag{74}$$

with μ^n each of the computed intermediate iterates. We further define the numerical primal-dual gap $\tilde{G}(\mu^n)$, which we compute by employing (58) with $\gamma^* = \gamma^*_{N_n+1}$.

6 Numerical Implementation and Experiments

In this section, we present the produced numerical experiments. In order to lower the computational cost, we chose to implement Algorithm 2 for the minimization of the time-discrete functional ($\mathcal{P}_{\text{discr}}$) discussed in Sect. 4.5. The adaptation of Algorithm 2 to such setting is easily obtainable as a corollary of the discussion in Sect. 4.5.2. The simulations were produced by a Python code that is openly available at https://github.com/panchoop/DGCG_algorithm/. For all the simulations, we employ the following:

- the considered domain is $\Omega := (0, 1) \times (0, 1) \subset \mathbb{R}^2$,
- the number of time samples is fixed to $T = 50$, with $t_i := i/T$ for $i = 0, \ldots, T$,
- the data spaces H_{t_i} and forward operators $K^*_{t_i} \colon \mathcal{M}(\overline{\Omega}) \to H_{t_i}$ are as in Sect. 6.1.1, and model a Fourier transform with time-dependent spatial undersampling. A specific choice of sampling pattern will be made in each experiment,
- in each experiment problem, ($\mathcal{P}_{\text{discr}}$) is considered for specific choices of regularization parameters $\alpha, \beta > 0$ and data $f = (f_{t_0}, \ldots, f_{t_T})$ with $f_{t_i} \in H_{t_i}$,
- convergence for Algorithm 2 is intended as in Sect. 5.3.3. The number of restarts $N_{\max}$ for Subroutine 1 is stated in each experiment. Also, we employ the quantitative stopping condition for Algorithm 2 described in Remark 7, with tolerance TOL $:= 10^{-10}$.
- the crossover parameters described in Sect. 5.1.4 are chosen as $\varepsilon = 0.05$ and $\delta = 1/T$. The random starts presented in Sect. 5.1.3 are implemented using the

---

Algorithm 2: Dynamic generalized conditional gradient (DGCG)

Input: Data $f \in L_H^2$, parameters $\alpha, \beta > 0$, forward operators $K_t^* : \mathcal{M}(\overline{\Omega}) \mapsto H_t$

---

1 $\boldsymbol{c} \leftarrow ()$, $\boldsymbol{\gamma} \leftarrow ()$
2 **for** $n = 0, 1, \ldots$ **do**
3 $\quad$ $N_n \leftarrow |\boldsymbol{\gamma}|$, $\mu^n \leftarrow \sum_{j=1}^{N_n} c_j \mu_{\gamma_j}$, $w_t^n \leftarrow -K_t(K_t^* \rho_t^n - f_t)$
$\quad\quad$ /* Multiple insertion step employing Subroutine 1 $\quad\quad\quad$ */
4 $\quad$ $\mathcal{S} \leftarrow$ MultistartGD$(\boldsymbol{\gamma}, w^n)$
$\quad\quad$ /* Stopping conditions $\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad$ */
5 $\quad$ **if** $\mathcal{S} == \emptyset$ **then**
6 $\quad\quad$ **return** μ^n
7 $\quad$ $(\gamma_{N_n+1}^*, \ldots, \gamma_{N_n+|\mathcal{S}|}^*) \leftarrow$ order$(\mathcal{S})$
8 $\quad$ **if** $\langle \rho_{\gamma_{N_n+1}^*}, w^n \rangle \leq 1$ **then**
9 $\quad\quad$ **return** μ^n
10 $\quad$ $\tilde{N}_n \leftarrow N_n + |\mathcal{S}|$, $\boldsymbol{\gamma} = (\gamma_1, \ldots, \gamma_{\tilde{N}_n}) \leftarrow (\boldsymbol{\gamma}, \gamma_{N_n+1}^*, \ldots, \gamma_{N_n+|\mathcal{S}|}^*)$
11 $\quad$ **for** $k = 1, \ldots, K_{\max}$ **do**
$\quad\quad$ /* Coefficients optimization step $\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad$ */
12 $\quad\quad$ $\boldsymbol{c} \leftarrow \arg\min_{(c_1, \ldots, c_{\tilde{N}_n}) \in \mathbb{R}_+^{\tilde{N}_n}} T_{\alpha, \beta}\left(\sum_{j=1}^{\tilde{N}_n} c_j \mu_{\gamma_j} \right)$
13 $\quad\quad$ $\tilde{N}_n \leftarrow \#\{j : c_j > 0\}$, $(\boldsymbol{\gamma}, \boldsymbol{c}) \leftarrow$ delete_zero_weighted$(\boldsymbol{\gamma}, \boldsymbol{c})$
$\quad\quad$ /* Sliding step $\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad$ */
14 $\quad\quad$ $T_{\alpha, \beta, \boldsymbol{c}}(\eta_1, \ldots, \eta_{\tilde{N}_n}) := T_{\alpha, \beta}\left(\sum_{j=1}^{\tilde{N}_n} c_j \mu_{\eta_j} \right)$
15 $\quad\quad$ $(\eta_1^*, \ldots, \eta_{\tilde{N}_n}^*) \leftarrow$ descend $T_{\alpha, \beta, \boldsymbol{c}}$ from $(\gamma_1, \ldots, \gamma_{\tilde{N}_n})$
16 $\quad\quad$ $\boldsymbol{\gamma} \leftarrow (\eta_1^*, \ldots, \eta_{\tilde{N}_n}^*)$
17 $\quad$ $N_{n+1} \leftarrow \tilde{N}_n$, $\mu^{n+1} \leftarrow \sum_{j=1}^{N_{n+1}} c_j \mu_{\gamma_j}$

---

function $Q(x) = \exp(\max(x + 0.05, 0)) - 1$. These parameter choices are purely heuristical, and there is no reason to believe that they are optimal.

The remainder of the section is organized as follows. In Sect. 6.1, we first introduce the measurement spaces and Fourier-type forward operators employed in the experiments. After, we explain how the synthetic data are generated in the noiseless case, and then detail on the noise model we consider. Subsequently, we show how the data and the obtained reconstructions can be visualized, by means of the so-called backprojections and intensities. We then pass to the actual experiments in Sect. 6.2, detailing three of them. The first experiment (Sect. 6.2.1), which is basic in nature, serves the purpose of illustrating how the measurements are constructed and how the data can be visualized. We then showcase the reach of the proposed regularization and algorithm in the second example (Sect. 6.2.2). There, we consider more complex data with various levels of added noise. In particular, we show that the proposed dynamic setting is capable of reconstructing severely spatially undersampled data. The final experiment (Sect. 6.2.3) illustrates a particular behavior of our model when reconstructing sparse measures whose underling curves cross, as discussed in Remark 3. We point out that in all the experiments presented, the algorithm converged faster, indeed linearly, than the

sublinear rate predicted by Theorem 7. We conclude the section with a few general observations on the model and algorithm proposed.

6.1 Measurements, Data, Noise Model and Visualization

6.1.1 Measurements

The measurements employed in the experiments are given by the time-discrete version, in the sense of Sect. 4.5, of the spatially undersampled Fourier measurements introduced in Sect. 1. Such measurements will be suitably cut-off, in order to avoid boundary conditions when dealing with the insertion step (Sects. 4.1.2, 5.2.1) and the sliding step (Sect. 5.2.2). Precisely, at each time instant t_i, we sample $n_i \in \mathbb{N}$ frequencies, encoded in the given vector $S_i = (S_{i,1}, \ldots, S_{i,n_i}) \in (\mathbb{R}^2)^{n_i}$, for all $i \in \{0, \ldots, T\}$. The sampling spaces H_{t_i} are defined as the realification of $\mathbb{C}^{n_i}$, equipped with the inner product $\langle u, v \rangle_{H_{t_i}} := \mathrm{Re}\langle u, v \rangle_{\mathbb{C}^{n_i}}/n_i$, where Re denotes the real part of a complex number. According to (A50), we define the cut-off Fourier kernels $\psi_{t_i} : \mathbb{R}^2 \mapsto \mathbb{C}^{n_i}$ by

$$\psi_{t_i}(x) := \left(\exp(-2\pi i x \cdot S_{i,k})\chi(x_1)\chi(x_2)\right)_{k=1}^{n_i},$$

where the map $\chi : (0, 1) \mapsto (0, 1)$ is defined by

$$\chi(z) := \begin{cases} 10(z/0.1)^3 - 15(z/0.1)^4 + 6(z/0.1)^5 & \text{for } z \in [0, 0.1), \\ 1 & \text{for } z \in [0.1, 0.9], \\ 10((1-z)/0.1)^3 - 15((1-z)/0.1)^4 + 6((1-z)/0.1)^5 & \text{for } z \in (0.9, 1]. \end{cases}$$

Notice that χ is twice differentiable, strictly increasing in $[0, 0.1]$, strictly decreasing in $[0.9, 1]$, and it satisfies $\chi(0) = \chi(1) = 0$, $\chi(z) = 1$ for all $z \in [0.1, 0.9]$. Following (A49), the cut-off undersampled Fourier transform and its pre-adjoint are given by the linear continuous operators $K_{t_i}^* : \mathcal{M}(\overline{\Omega}) \to H_{t_i}$ and $K_{t_i} : H_{t_i} \to C(\overline{\Omega})$ defined by

$$K_{t_i}^*(\rho) := \int_{\mathbb{R}^2} \psi_{t_i}(x)\, d\rho(x), \qquad K_{t_i}(h) := \left(x \mapsto \langle \psi_{t_i}(x), h \rangle_{H_{t_i}} \right), \tag{75}$$

for all $\rho \in \mathcal{M}(\overline{\Omega})$, $h \in H_t$, where ρ is extended to zero outside of $\overline{\Omega}$, and the first integral is intended component-wise.

6.1.2 Data

For all the experiments, the ground-truth consists of a sparse measure of the form

$$\mu^\dagger = (\rho^\dagger, m^\dagger) := \sum_{j=1}^{N} c_j^\dagger \mu_{\gamma_j^\dagger}, \tag{76}$$

for some $N \in \mathbb{N}$, $c_j^\dagger > 0$, $\gamma_j^\dagger \in AC^2$, where we follow the notations at (14). Given a ground-truth $\mu^\dagger$, the respective noiseless data is constructed by $f_{t_i} := K_{t_i}^* \rho_{t_i}^\dagger \in H_{t_i}$, for $i = 0, \dots, T$.

6.1.3 Noise Model

Let $U_{i,k}$, $V_{i,k}$, for $i = 0, \dots, T$ and $k = 1, \dots, n_i$, be the realization of two jointly independent families of standard one-dimensional Gaussian random variables, with which we define the noise vector ν by

$$\nu_{i,k} := U_{i,k} + iV_{i,k} \in \mathbb{C}, \quad \nu_{t_i} := (\nu_{i,k})_{k=1}^{n_i} \in H_{t_i}, \quad \nu := (\nu_0, \nu_1, \dots, \nu_T). \quad (77)$$

Given some data $f = (f_{t_0}, \dots, f_{t_T})$, with $f_{t_i} \in H_{t_i}$, when $\nu \neq 0$, the corresponding noisy data f^ε with noise level $\varepsilon \geq 0$ is taken as

$$f^\varepsilon := f + \varepsilon \sqrt{\frac{\sum_{i=0}^{T} \left\| f_{t_i} \right\|_{H_{t_i}}^2}{\sum_{i=0}^{T} \left\| \nu_{t_i} \right\|_{H_{t_i}}^2}} \, \nu.$$

6.1.4 Visualization via Backprojection

In general, it is not illustrative to directly visualize the data. The proposed way to gain some insight on the data structure is by means of backprojections: given $f = (f_{t_0}, \dots, f_{t_T})$, with $f_{t_i} \in H_{t_i}$, we call backprojection the map $w_{t_i}^0 := K_{t_i} f_{t_i} \in C(\overline{\Omega})$. Note that $w_{t_i}^0$ corresponds to the dual variable at the first iteration of Algorithm 2. As $\Omega = (0, 1)^2$, such functions can be plotted at each time sample, allowing us to display the data.
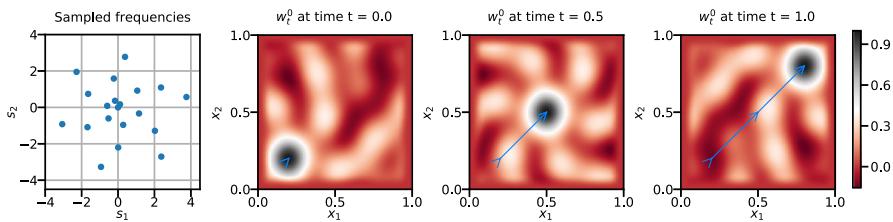
6.1.5 Reconstruction's Intensities

Given a sparse measure $\mu := \sum_{j=1}^{N} c_j \mu_{\gamma_j}$, with $N \in \mathbb{N}$, $c_j > 0$, $\gamma_j \in AC^2$, the intensity associated to the atom μ_{γ_j} is defined by $I_j := c_j a_{\gamma_j}$. The quantity I_j measures the intensity at time t_i of the signal for a single source, as $K_{t_i}^*(c_j \rho_{\gamma_j(t_i)}) = I_j K_{t_i}^*(\delta_{\gamma_j(t_i)})$. Therefore, when presenting reconstructions and comparing them to the given ground-truth, we will use the intensity I_j of each atom instead of its associated weight c_j.

6.2 Numerical Experiments

6.2.1 Experiment 1: Single Atom with Constant Speed

We start with a basic example that serves at illustrating how to observe the data, the obtained reconstructions, and their respective distortions due to the employed regularization. We use constant-in-time forward Fourier-type measurements, with frequencies sampled from an Archimedean spiral: for each sampling time $i \in \{0, \dots, 50\}$,
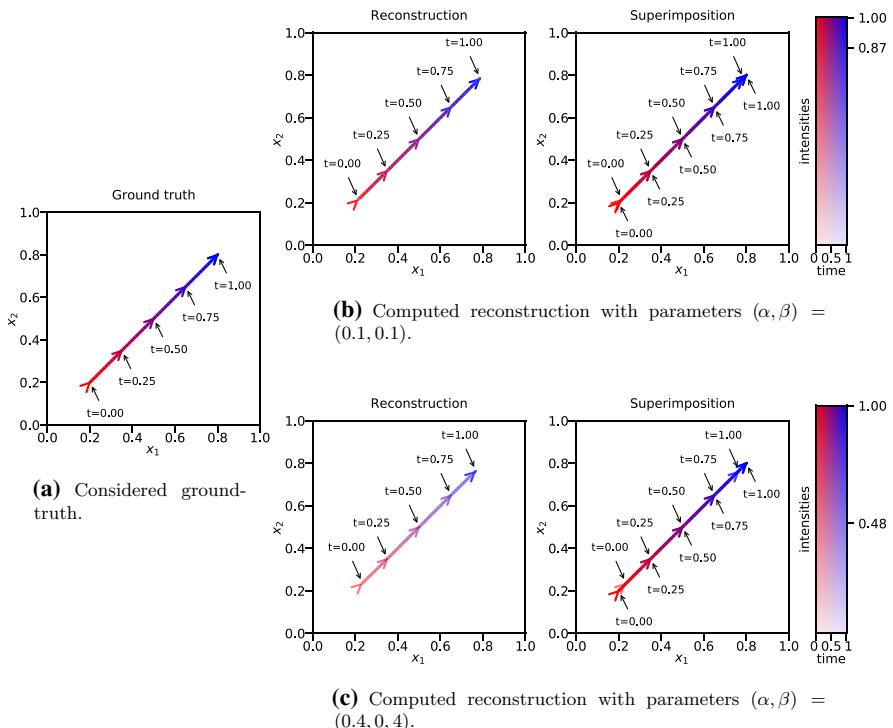
Fig. 1 Time-constant frequency samples $\{S_{i,k}\}_k$ and backprojections $w_{t_i}^0$ for data $f_{t_i} := K^*\rho_{t_i}^\dagger$ at times $t_0 = 0$, $t_{25} = 0.5$, $t_{50} = 1$, with the ground-truth curve $\gamma^\dagger$ superimposed in blue color (Color figure online)

we consider the same frequencies vector $S_i \in (\mathbb{R}^2)^{n_i}$ with $n_i := 20$ and $S_{i,k}$ lying on a spiral for $k = 1, \ldots, 20$ (see Fig. 1). Thus, the corresponding forward operators defined by (75) are constant in time, i.e., $K_{t_i}^* = K^*$. The employed ground-truth $\mu^\dagger$ is composed of a single atom with intensity $I^\dagger = 1$, and respective curve $\gamma^\dagger(t) := (0.2, 0.2) + t(0.6, 0.6)$. Accordingly, we have $\mu^\dagger = c^\dagger \mu_{\gamma^\dagger}$ with $c^\dagger = 1/a_{\gamma^\dagger}$. We consider the case of noiseless data $f_{t_i} := K^*\rho_{t_i}^\dagger$. The corresponding backprojection $w_{t_i}^0 := Kf_{t_i}$ is shown in Fig. 1 for some selected time samples. For the proposed data f, we solve the minimization problem ($\mathcal{P}_{\text{discr}}$) employing Algorithm 2. The obtained reconstructions are shown in Fig. 2, where we display the results for two different parameter choices α, β. Given the simplicity of the considered example, we employed Subroutine 1 with only 5 restarts, that is, $N_{\max} = 5$. In the respective cases of parameters $\alpha = \beta = 0.1$ and $\alpha = \beta = 0.4$, Algorithm 2 converged in 2 and 1 iterations, and had an execution time of 2 and 1 minutes (the employed CPU was an Apple M1 8 Core 3.2 GHz, running native arm64 Python 3.9.7).

As common for Tikhonov regularization methods, the obtained reconstructions differ from the ground-truth due to the effect of regularization. Specifically, in Fig. 2c, we notice a stronger effect of the regularization with parameters $\alpha = \beta = 0.4$ at the endpoints of the reconstructed curve. Such phenomenon is expected: as argued in Sect. 5.1, each curve found by Subroutine 1 belongs to the set of stationary points $\mathfrak{S}_F$; due to the optimality conditions proven in Proposition 17, any of such curves has zero initial and final speed, i.e., $\dot{\gamma}(0) = \dot{\gamma}(1) = 0$. The mentioned constraint is achieved with a slower transition for larger speed penalizations β. We can quantify the discrepancy between the ground-truth curve $\gamma^\dagger$ and a reconstructed curve $\overline{\gamma}$ with respect to the L^2 norm by computing $D(\gamma^\dagger, \overline{\gamma}) := \|\gamma^\dagger - \overline{\gamma}\|_{L^2} / \|\gamma^\dagger\|_{L^2}$. We obtain that $D(\gamma^\dagger, \overline{\gamma}) = 0.00515$ for $\alpha = \beta = 0.1$ and $D(\gamma^\dagger, \overline{\gamma}) = 0.017$ for $\alpha = \beta = 0.4$. The reconstructed intensities for the parameter choices $\alpha = \beta = 0.1$ and $\alpha = \beta = 0.4$ are 87% and 48% of the ground-truth's intensity, respectively, as shown in Fig. 2.

6.2.2 Experiment 2: Complex Example with Time-Varying Measurements

The following example is given to showcase the full strength of the proposed regularization and algorithm. The frequencies are sampled over lines through the origin of $\mathbb{R}^2$, which are rotating in time. Specifically, let $\Theta \in \mathbb{N}$ be a bound on the number of lines, $h > 0$ a fixed spacing between measured frequencies on a given line, and

(b) Computed reconstruction with parameters $(\alpha, \beta) = (0.1, 0.1)$.

(a) Considered ground-truth.

(c) Computed reconstruction with parameters $(\alpha, \beta) = (0.4, 0.4)$.
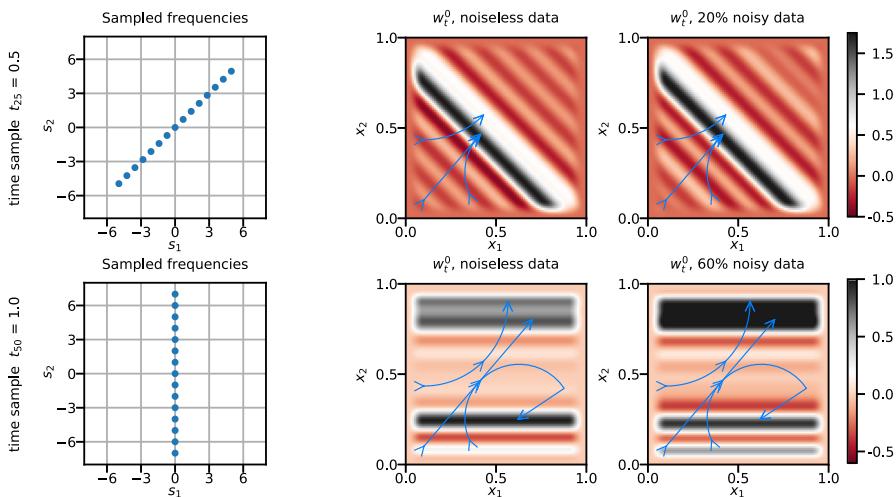
Fig. 2 Reconstruction results for Experiment 1. We use color to indicate position in time and transparency to indicate intensity of the respective atom. From left to right, we plot the employed ground-truth, the obtained reconstruction with the specified parameters, and then the superimposition of ground-truth and reconstruction

consider frequencies $S_i \in (\mathbb{R}^2)^{n_i}$ defined by

$$
S_{i,k} := \begin{pmatrix} \cos(\theta_i) & \sin(\theta_i) \\ -\sin(\theta_i) & \cos(\theta_i) \end{pmatrix} \begin{pmatrix} h(k - (n_i + 1)/2) \\ 0 \end{pmatrix},
$$
$$
i \in \{0, \ldots, 50\}, \ k \in \{1, \ldots, n_i\}. \tag{78}
$$

Here, the matrix represents a rotation of angle θ_i, with $\theta_i := \frac{i}{\Theta}\pi$. For such frequencies, we consider the associated forward operators $K_{t_i}^*$ as in (75).

In the presented experiment, the parameters are chosen as $\Theta = 4$, $h = 1$ and $n_i = 15$ for all $i = 0, \ldots, 50$: in other words, we sample along four different lines which rotate at each time-sample; see Fig. 3 for two examples of them. The ground-truth is a measure $\mu^\dagger$ composed of three atoms, whose associated curves are as in Fig. 4a. Note that $\mu^\dagger$ displays different non-constant speeds, a contact point between two underlying curves, a strong kink, and intensities equal to 1. We point out that equal intensities are considered for the sole purpose of easing graph visualization; the reconstruction quality is not affected by different intensity choices. The noiseless data are defined by $f_{t_i} := K_{t_i}^* \rho_{t_i}^\dagger$. Following the noisy model described in Sect. 6.1.3,

Springer
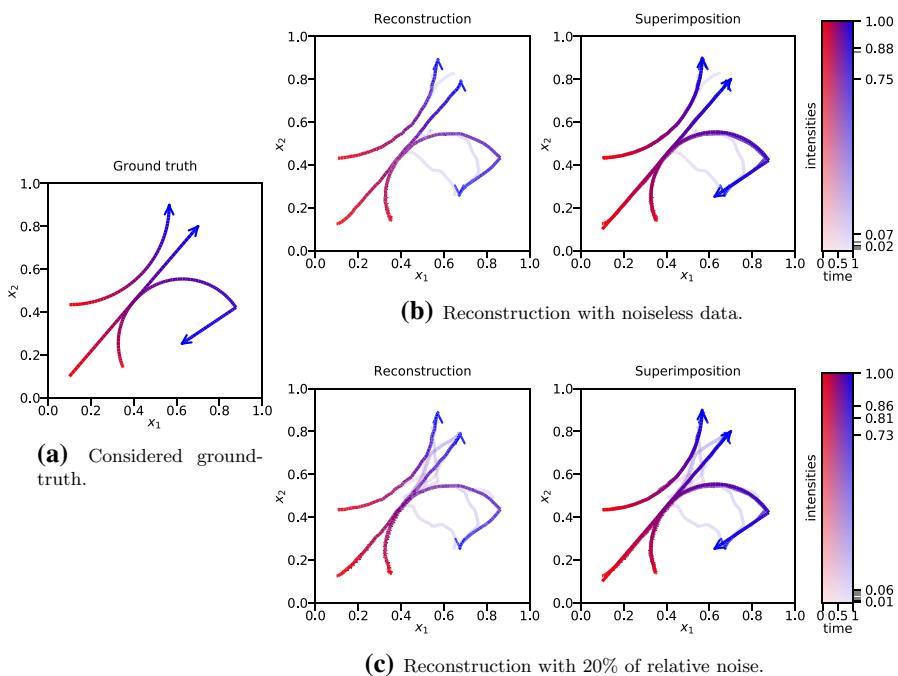
Fig. 3 Sampled frequencies and backprojected data for Experiment 2 at times $t_{25} = 0.5$, $t_{50} = 1$. The backprojected data are displayed for noiseless data and for 20% and 60% of relative noise, i.e., $f^{0.2}$ and $f^{0.6}$, respectively, with superimposed ground-truth's curves in blue color

we also consider data $f^{0.2}$ and $f^{0.6}$ with added 20% and 60% of relative noise, respectively. For the noisy data, we employed the same realization of the randomly generated noise vector ν defined in (77). In Fig. 3, we present the backprojections for the noiseless and noisy data at two different times. We can observe that at each time sample, the backprojected data exhibits a line-constant behavior, as explained in Remark 10.

We apply Algorithm 2 to obtain reconstructions for the cases of noiseless and 20% of added noise data for parameters $\alpha = \beta = 0.1$ (see Fig. 4); in Fig. 5, we present the obtained reconstructions for the case of added 60% noise, where we further show the regularization effects of employing larger α, β values, namely $\alpha = \beta = 0.1$ and $\alpha = \beta = 0.3$.

In the noiseless case, presented in Fig. 4b, we can observe an accurate reconstruction, with some low intensity artifacts that for most of the time share paths with the higher intensity atoms. In Fig. 4c, where we add 20% of noise to the data, we notice a surge of low-intensity artifacts, but nonetheless, we see that the obtained solution is close, in the sense of measures, to the original ground-truth. In Fig. 5, for the case of 60% added noise, we can notice that by increasing the regularization parameters, the quality of the obtained reconstruction increases, displaying small regularization-induced distortions. The examples in Figs. 4, 5 demonstrate the power of the proposed regularization, given its reconstruction accuracy when simultaneously employing highly ill-posed forward measurements, as pointed out in Remark 10, together with strong noise.
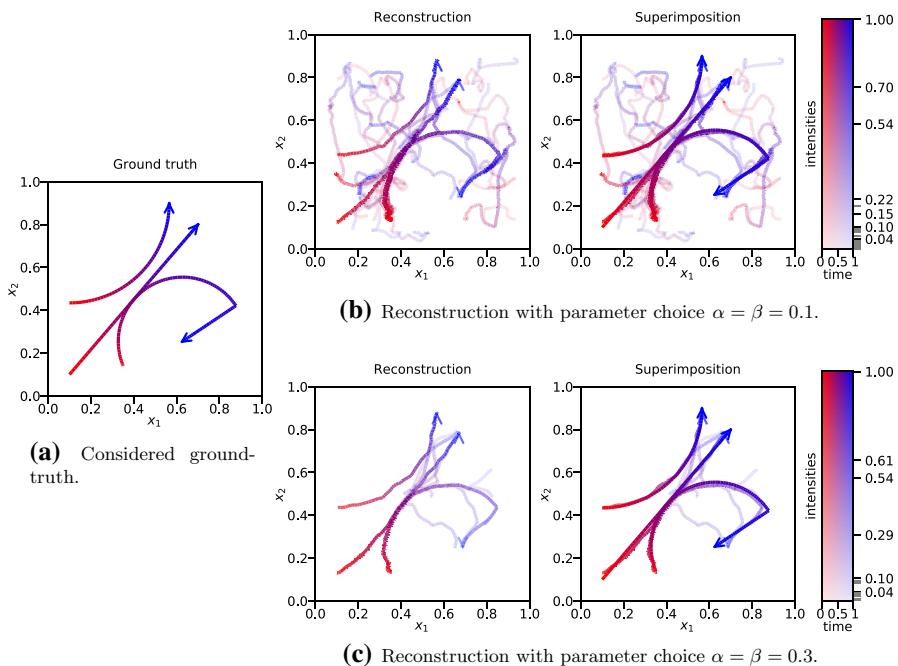
We finalize this example by presenting the convergence plots for the case of 60% added noise, this being the most complex experiment (see Fig. 6). We plot the numerical residual $\tilde{r}(\mu^n)$ and the numerical primal-dual gap $\tilde{G}(\mu^n)$ for the iterates μ^n, where $\tilde{r}$ and $\tilde{G}$ are defined in Sect. 5.3.3. We observe that the algorithm exhibits a linear

(b) Reconstruction with noiseless data.

(a) Considered ground-truth.


(c) Reconstruction with 20% of relative noise.
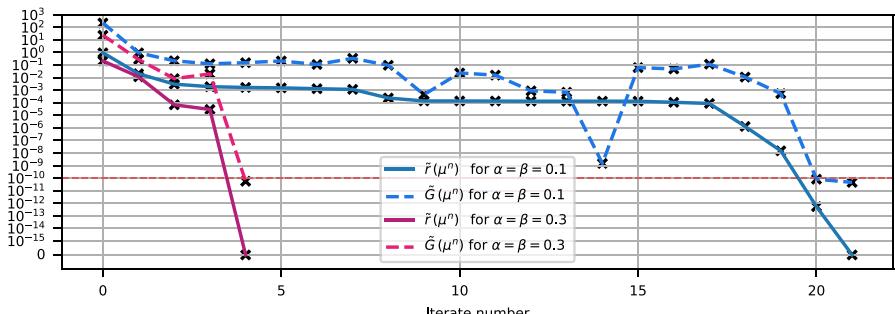
Fig. 4 Reconstruction results for Experiment 2, with noiseless data, 20% of relative noise data and parameters $\alpha = \beta = 0.1$. The intensities of the reconstructed atoms are represented by the rightmost black and grey tick-lines in the colorbars

rate of convergence, instead of the proven sublinear one. Such linear convergence has been numerically observed in all the tested examples. Additionally, we see that the algorithm is greatly accelerated when one considers strong regularization parameters α, β. Finally, the plot confirms the efficacy of the proposed descent strategy for the insertion step (68); indeed, we note that the inequality $\tilde{r}(\mu^n) \leq \tilde{G}(\mu^n)$ holds for most of the iterations in the performed experiments. As such inequality is proven in Lemma 5 for the actual residual and primal dual gap, we have confirmation that $\gamma^*_{N_n+1}$ in Algorithm 2 is a good approximated solution for (68). Regarding execution times, iterations until convergence and number of restarts, they are summarized in Table 1.

Remark 10 Consider the Fourier-type forward measurements K_t^* defined by (75) with frequencies $S_i \in (\mathbb{R}^2)^{n_i}$ sampled along rotating lines, as in (78). In this case, at each fixed time-sample t_i, the operator $K_{t_i}^*$ does not encode sufficient information to accurately resolve the location of the unknown ground-truth at time t_i. As a consequence, any static reconstruction technique, that is, one that does not jointly employ information from different time samples in order to perform a reconstruction, would not be able to accurately recover any ground-truth under these measurements. To justify this claim, notice that for all time-samples t_i, the family $\{S_{i,k}\}_{k=1}^{n_i}$ defined at (78) is collinear, and as such, there exists a corresponding vector $S_i^\perp \in \mathbb{R}^2$ such that $S_{i,k} \cdot S_i^\perp = 0$ for all $k = 1, \dots, n_i$. Therefore, for any given time-static source $\rho^\dagger = \delta_{x^\dagger}$

(b) Reconstruction with parameter choice $\alpha = \beta = 0.1$.

(c) Reconstruction with parameter choice $\alpha = \beta = 0.3$.

Fig. 5 Reconstruction results for Experiment 2, with 60% of relative noise. Algorithm 2 is applied to the same data, with regularization parameter choices $\alpha = \beta = 0.1$ and $\alpha = \beta = 0.3$

Fig. 6 Convergence plot for Experiment 2 with 60% of added noise and parameter choices $\alpha = \beta = 0.1$ and $\alpha = \beta = 0.3$. For each iterate, we plot the numerical residual $\tilde{r}(\mu^n)$, together with the corresponding numerical dual gap $\tilde{G}(\mu^n)$

with $x^\dagger \in (0.1, 0.9)^2 \subset \mathbb{R}^2$, the measured forward data are invariant along $S_i^\perp$, that is,

$$K_{t_i}^* \delta_{x^\dagger} = K_{t_i}^* \delta_{x^\dagger + \lambda S_i^\perp}, \quad \text{for all } \lambda \in \mathbb{R} \text{ such that } x^\dagger + \lambda S_i^\perp \in (0.1, 0.9)^2.$$

Hence, solely with the information of a single time sample, it is not possible to distinguish a source along a line, and therefore, it is not possible to accurately resolve it.

Table 1 Convergence information and execution times for Experiment 2

Relative Noise	(α, β)	Restarts	Iterations	Execution time
0%	(0.1, 0.1)	200	4	1.5 hours
20%	(0.1, 0, 1)	1000	7	5.8 hours
60%	(0.1, 0.1)	10000	21	10.5 days
60%	(0.3, 0.3)	5000	4	16.8 hours

We display the considered relative noise level of the data, the employed regularization parameters α, β, the number of restarts $N_{\max}$ in Subroutine 1, the number of iterations until reaching convergence, and the total execution time of Algorithm 2. The employed CPU was an Apple M1 8 Core 3.2 GHz, running native arm64 Python 3.9.7. For comments on execution times, see Sects. 6.3, 6.4

This is in contrast with the dynamic model presented in this paper, which is able to perform an efficient reconstruction, as demonstrated in Experiment 2.
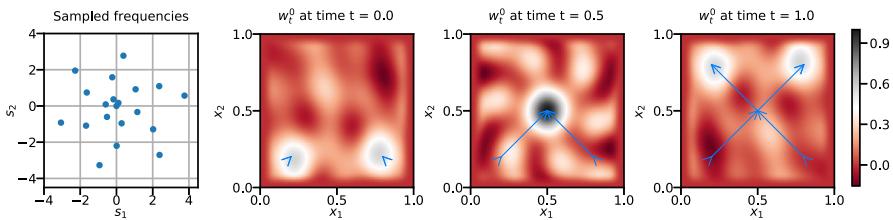
6.2.3 Experiment 3: Crossing Example

The following is an example in which the considered model is not able to track dynamic sources; although the reconstruction is close to the ground-truth in the sense of measures, its underlying curves do not resemble those of the ground-truth. This effect is due to the non-injectivity of the map at (38); even if the sparse measure we wish to recover is unique, its decomposition into atoms might not be. A simple example in which injectivity fails is given by the crossing of two curves (see Remark 3); this is the subject of the numerical experiment performed in this section. Specifically, the ground-truth $\mu^\dagger$ considered is of the form
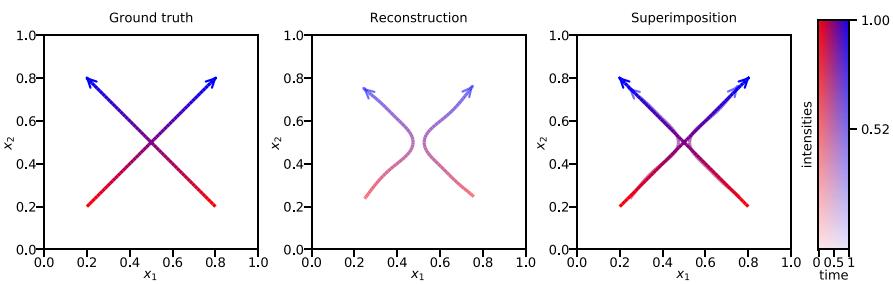
$$\mu^\dagger := a_{\gamma_1^\dagger}^{-1} \rho_{\gamma_1^\dagger} + a_{\gamma_2^\dagger}^{-1} \rho_{\gamma_2^\dagger},$$
$$\gamma_1^\dagger(t) := (0.2, 0.2) + t(0.6, 0.6), \qquad \gamma_2^\dagger(t) := (0.8, 0.2) + t(-0.6, 0.6). \tag{79}$$

Notice that $\gamma_1^\dagger$ and $\gamma_2^\dagger$ cross at time $t = 0.5$, and the respective atoms have both intensity 1. For the forward measurements, we employ the time-constant Archimedean spiral family of frequencies defined in the first numerical experiment in Sect. 6.2.1, resulting in the constant in time operator $K_{t_i}^* = K^*$. The reconstruction is performed for noiseless data $f_{t_i} := K^* \rho_{t_i}^\dagger$. In Fig. 7, we present the considered frequency samples, together with the backprojected data at selected time samples.

In Fig. 8, we display the obtained reconstruction for high regularization parameters $\alpha = 0.5$ and $\beta = 0.5$. It is observed that the reconstructed atoms are not close, as curves, to the ones in (79); rather than a crossing at time $t = 0.5$, the two curves rebound. As already mentioned, this phenomenon is a consequence of the lack of uniqueness for the sparse representation of $\mu^\dagger$, which in this case is both represented by crossing curves and rebounding curves. The fact that Algorithm 2 outputs rebounding curves is due to employed regularization; indeed, the considered Benamou–Brenier-type penalization selects a solution whose squared velocity is minimized, which discourages the reconstruction to follow the crossing path. However, we

Fig. 7 Time-constant frequency samples $\{S_{i,k}\}_k$ and corresponding backprojected data w_t^0 for Experiment 3. The backprojected data are taken in the noiseless case, and presented at times $t_0 = 0$, $t_{25} = 0.5$, $t_{50} = 1$, with associated ground-truth superimposed in blue color

(a) Considered ground-truth.

(b) Computed reconstruction with parameters $(\alpha, \beta) = (0.5, 0.5)$.

Fig. 8 Reconstruction results for Experiment 3. In this case, the method fails to reconstruct the crossing, but still approximates the ground-truth in terms of measures

remark that the algorithm proposed yields a good solution in terms of the model, given that in the sense of measures, the obtained reconstruction is very close to the ground-truth $\mu^\dagger$. More sophisticated models are needed in order to resolve the crossing, as briefly discussed in Sect. 7.

6.3 Remark on Execution Times

It is observed that the execution times of our algorithm are quite high for some of the presented examples in Table 1. This is mainly due to the computational cost of the insertion step, since the algorithm is set to run several gradient descents to find a global minimum of the linearized problem (26) at each iteration, and these descents are executed in a non-parallel fashion on a single CPU core. The other components of the algorithm, namely the routines `sample` and `crossover` included in Subroutine 1, the coefficient optimization step and the sliding step have, in comparison, negligible computational cost. In particular, the sliding step grows in execution time with the number of active atoms, but this effect appears toward the last iterations of the algorithm, and it is shadowed by the insertion step, whose gradient descents become longer as the iterate is getting closer to the optimal value. As a confirmation of the role of the insertion step in the overall computational cost, one can see that the execution times of the algorithm linearly depend on the total number of gradient descents that are run in

each example. Indeed, since the total number of gradient descents is given by the number of restarts multiplied by the number of iterations (see Table 1), the ratio between the execution times and the total number of gradient descents is of the same order for all the presented examples (between 0.0008 and 0.0019 hour/gradient descent). It is worth pointing out that the multistart gradient descent is a highly parallelizable method. Some early tests in this direction indicate that much lower computational times are achievable by simultaneously computing gradient descents on several GPUs for the presented examples.

6.4 Conclusions and Discussion

The presented numerical experiments confirm the effectiveness of Algorithm 2, and that the proposed Benamou–Brenier-type energy is an excellent candidate to regularize dynamic inverse problems. This is in particular evidenced by Experiment 2 in Sect. 6.2.2, where we consider a dynamic inverse problem that is impossible to tackle with a purely static approach, as discussed in Remark 10. Even in the extreme case of 60% added noise, our method recovered satisfactory reconstructions.

We can further observe the distortions induced by the considered regularization. Precisely, the attenuation of the reconstruction's velocities and intensities is a direct consequence of the minimization of the Benamou–Brenier energy and the total variation norm in the objective functional; for this reason, the choice of the regularization parameters α and β affects the reconstruction and the magnitude of such distortions. Additionally, a further effect of the regularization is the phenomenon presented in Experiment 3 (Sect. 6.2.3). As the dynamic inverse problem is formulated in the space of measures, if the sparse ground truth possesses many different decompositions into extremal points, the preferred reconstruction may be the one favoring the regularizer.

Finally, concerning the execution times, we emphasize that the presented simulations are a proof of concept and not the result of a carefully optimized algorithm. There are many improvement directions, where the most promising one is a GPU implementation to parallelize the multiple insertion step. To increase the likelihood of finding a global minimizer for the insertion step, Subroutine 1 is employed with a high number of restarts N_{max}, which was tuned manually, prioritizing high reconstruction accuracy over execution time. To improve on this aspect, one could include early stopping conditions in Subroutine 1, for example by exiting the routine when a sufficiently high ratio of starts descend toward the same stationary curve. Last, the code was written having in mind readability, as well as adaptability to a broad class of inverse problems. Therefore, the experienced execution times are not an accurate estimation of what would be possible in specific applications.

7 Future Perspectives

In this section, we propose several research directions to expand on the research presented in this paper. A first relevant question concerns the proposed DGCG algorithm, and, in particular, the possibility of proving a theoretical linear convergence rate under

suitable structural assumptions on the minimization problem ($\mathcal{P}$). Linear convergence has been recently proven for the GCG method applied to the BLASSO problem [40, 52]. It seems feasible to extend such an analysis to the DGCG algorithm presented in this paper, especially seeing the linear convergence observed in the experiments provided in Sect. 6.2, and the fact that our proof of sublinear convergence (see Theorem 7) does not fully exploit the coefficients optimization step, as commented in Remark 6. This line of research is currently under investigation by the authors [13].

Another interesting research direction is the extension of the DGCG method introduced in this paper to the case of unbalanced optimal transport. Precisely, one can regularize the inverse problem (2) by replacing the Benamou–Brenier energy B in (3) with the so-called Wasserstein–Fischer–Rao energy, as proposed in [15]. Such energy, first introduced in [24, 45, 47] as a model for unbalanced optimal transport, accounts for more general displacements $t \mapsto \rho_t$, in particular allowing the total mass of ρ_t to vary during the evolution. The possibility to numerically treat such a problem with conditional gradient methods would rest on the characterization of the extremal points for the Wasserstein–Fischer–Rao energy recently achieved by the authors in [12].

In addition, it is a challenging open problem to design alternative dynamic regularizers that allow to reconstruct accurately a ground-truth composed of crossing atoms, such as the ones considered in the experiment in Sect. 6.2.3. Due to the fact that the considered Benamou–Brenier-type regularizer penalizes the square of the velocity field associated with the measure, the reconstruction obtained by our DGCG algorithm does not follow the crossing route (Fig. 8b). A possible solution is to consider additional high-order regularizers in ($\mathcal{P}$), such as curvature-type penalizations. The challenging part is devising a penalization that can be enforced at the level of Borel measures, and whose extremal points are measures concentrated on sufficiently regular curves.

Finally, keeping into account the possible improvements discussed in Sect. 6.4, the implementation of an accelerated and parallelized version of Algorithm 2 will be the subject of future work.

Appendix A

A.1 Lemmas on Optimal Transport Regularization

In this section, we recall several results concerning the continuity equation (9), the functionals B and $J_{\alpha,\beta}$ introduced at (11) and (12), respectively, and the data spaces L_H^2 at (18). For proofs of such results, we refer the reader to Propositions 2.2, 2.4 and Lemmas 4.2, 4.5, 4.6 in [15], and to Proposition 5.18 in [54].

Lemma 9 (Properties of the continuity equation) *Assume that $\mu = (\rho, m) \in \mathcal{M}$ satisfies* (9) *and that $\rho \in \mathcal{M}^+(X)$. Then, ρ disintegrates with respect to time into $\rho = \mathrm{d}t \otimes \rho_t$, where $\rho_t \in \mathcal{M}^+(\overline{\Omega})$ for a.e. t, and $t \mapsto \rho_t(\overline{\Omega})$ is constant, with $\rho_t(\overline{\Omega}) = \rho(X)$ for a.e. $t \in (0,1)$. Moreover, $t \mapsto \rho_t$ belongs to C_w^+ if, in addition, $m = v\rho$ for some measurable $v\colon X \to \mathbb{R}^d$ such that*

$$\int_0^1 \int_{\overline{\Omega}} |v(t,x)| \, \mathrm{d}\rho_t(x) \, \mathrm{d}t < +\infty.$$

Lemma 10 *(Properties of B) The functional B defined in* (11) *is nonnegative, convex, one-homogeneous and sequentially lower semicontinuous with respect to the weak* topology on $\mathcal{M}$. Moreover, the following properties hold:*

  i) *if $B(\rho, m) < +\infty$, then $\rho \geq 0$ and $m \ll \rho$, that is, there exists a measurable map $v\colon X \to \mathbb{R}^d$ such that $m = v\rho$,*

  ii) *let Ψ be the map at* (10). *If $\rho \geq 0$ and $m = v\rho$ for some $v\colon X \to \mathbb{R}^d$ measurable, then*

$$B(\rho, m) = \int_X \Psi(1, v) \, \mathrm{d}\rho = \frac{1}{2} \int_X |v|^2 \, \mathrm{d}\rho. \tag{A1}$$

Lemma 11 *(Properties of $J_{\alpha,\beta}$) Let $\alpha, \beta > 0$. The functional $J_{\alpha,\beta}$ at* (12) *is nonnegative, convex, one-homogeneous and sequentially lower semicontinuous with respect to weak* convergence on $\mathcal{M}$. For $\mu = (\rho, m) \in \mathcal{M}$ such that $J_{\alpha,\beta}(\mu) < +\infty$, we have that*

$$\max\{\alpha \|\rho\|_{\mathcal{M}(X)}, C \|m\|_{\mathcal{M}(X;\mathbb{R}^d)}\} \leq J_{\alpha,\beta}(\mu), \tag{A2}$$

where $C := \min\{2\alpha, \beta\}$. Moreover, if $\{\mu^n\}$ sequence in $\mathcal{M}$ is such that $\{J_{\alpha,\beta}(\mu^n)\}$ is uniformly bounded, then $\rho^n = \mathrm{d}t \otimes \rho_t^n$ for some $(t \mapsto \rho_t^n) \in C_w^+$, and there exists $\mu = (\rho, m) \in \mathcal{D}$ with $\rho = \mathrm{d}t \otimes \rho_t$ and $(t \mapsto \rho_t) \in C_w^+$, such that, up to subsequences,

$$\begin{cases} (\rho^n, m^n) \overset{*}{\rightharpoonup} (\rho, m) \ \text{ weakly* in } \ \mathcal{M}, \\ \rho_t^n \overset{*}{\rightharpoonup} \rho_t \ \text{ weakly* in } \ \mathcal{M}(\overline{\Omega}), \ \text{ for every } \ t \in [0,1]. \end{cases} \tag{A3}$$

Lemma 12 *(Properties of K_t^*) Assume (H1)–(H3), (K1)–(K3) as in Sect. 3.1. If $t \mapsto \rho_t$ is in C_w, then the map $t \mapsto K_t^* \rho_t$ belongs to L_H^2. Let $\{(\rho^n, m^n)\}$ in $\mathcal{D}$ be such that $\rho^n = \mathrm{d}t \otimes \rho_t^n$ and $(t \mapsto \rho_t^n) \in C_w^+$. If (ρ^n, m^n) converges to (ρ, m) in the sense of* (A3), *where $\rho = \mathrm{d}t \otimes \rho_t$, $(t \mapsto \rho_t) \in C_w^+$, then we have $K^* \rho^n \rightharpoonup K^* \rho$ weakly in L_H^2.*

A.2 Proof of Lemma 2

The fact that $J_{\alpha,\beta}(\mu_j) = 1$ follows by (A1). Define the vector field

$$v(t, x) := \begin{cases} \dot{\gamma}_j(t) & \text{if } (t, x) \in \text{graph } \gamma_j, \\ 0 & \text{otherwise}, \end{cases}$$

where graph $\gamma_j := \{(t, \gamma_j(t)) : t \in (0, 1)\} \subset X$. Notice that v is well-defined up to negligibly many $t \in (0, 1)$; indeed, we have $\dot{\gamma}_i = \dot{\gamma}_j$ a.e. in $\{t : \gamma_i(t) = \gamma_j(t)\}$ for every i, j (see [36, Theorem 4.4]). Hence, $\dot{\gamma}_i(t) = \dot{\gamma}_j(t)$ for a.e. $t \in (0, 1)$ and every x such that $(t, x) \in \text{graph } \gamma_j \cap \text{graph } \gamma_i$. Set now $\mu := \sum_{j=1}^N c_j \mu_{\gamma_j}$. It is immediate to see that $v = dm/d\rho$ and that v satisfies $v(t, \gamma_j(t)) = \dot{\gamma}_j(t)$ for all $j = 1, \ldots, N$. Moreover, by linearity, μ satisfies the continuity equation (9). Employing (A1) and the definition of a_{γ_j}, we conclude noting that

$$J_{\alpha,\beta}(\mu) = \frac{\beta}{2} \int_X |v(t, x)|^2 \, d\rho + \alpha \|\rho\|_{\mathcal{M}(X)}$$

$$= \sum_{j=1}^N c_j a_{\gamma_j} \left(\frac{\beta}{2} \int_0^1 |v(t, \gamma_j(t))|^2 \, dt + \alpha \right)$$

$$= \sum_{j=1}^N c_j a_{\gamma_j} \left(\frac{\beta}{2} \int_0^1 |\dot{\gamma}_j(t)|^2 \, dt + \alpha \right) = \sum_{j=1}^N c_j.$$

A.3 Existence of Minimizers for Linearized Problems

In this section, we show existence of minimizers for the problems at (25) and (26). To this end, we prove existence for a slightly more general functional (see (A4)), which coincides with (25) and (26) for φ as in (20) and $\varphi(t) := \chi(-\infty, 1](t)$, respectively.

Theorem 13 *Assume (H1)–(H3), (K1)–(K3) and let $f \in L^2_H$, $\alpha, \beta > 0$. Given $(t \mapsto \tilde{\rho}_t) \in C_w$ define $w_t := -K_t(K_t^* \tilde{\rho}_t - f_t) \in C(\overline{\Omega})$ for a.e. $t \in (0, 1)$. Let $\varphi \colon \mathbb{R} \to [0, +\infty]$ with $\varphi(0) = 0$ be monotonically increasing, lower semicontinuous and superlinear at infinity, i.e., $\varphi(t)/t \to +\infty$ as $t \to +\infty$. Then, there exists $\mu^* = (\rho^*, m^*) \in \mathcal{D}$ that solves the minimization problem*

$$\min_{\mu \in \mathcal{M}} -\langle \rho, w \rangle + \varphi(J_{\alpha,\beta}(\mu)), \tag{A4}$$

where the product $\langle \cdot, \cdot \rangle$ is defined at (23). Moreover, $\rho^* = dt \otimes \rho_t^*$ with $(t \mapsto \rho_t^*) \in C_w^+$.

Proof First notice that the functional at (A4) is proper since $J_{\alpha,\beta}(0) = 0$ (by Lemma 10) and $\varphi(0) = 0$. Let $\{\mu^n\}$ be a minimizing sequence, so that, in particular,

$$\sup_n -\langle \rho^n, w \rangle + \varphi(J_{\alpha,\beta}(\mu^n)) < +\infty. \tag{A5}$$

We claim that $\sup_n J_{\alpha,\beta}(\mu^n) < +\infty$. Indeed, assume by contradiction that $J_{\alpha,\beta}(\mu^n) \to +\infty$ as $n \to +\infty$ (subsequentially). Fix $C' > 0$. Since φ is superlinear, there exists $n_0 \in \mathbb{N}$ such that

$$\varphi(J_{\alpha,\beta}(\mu^n)) \geq C' J_{\alpha,\beta}(\mu^n) \quad \text{for all} \ n \geq n_0. \tag{A6}$$

Moreover, notice that for n fixed, we have $J_{\alpha,\beta}(\mu^n) < +\infty$, as (A5) holds. In particular, we obtain that $\mu^n \in \mathcal{D}$ and $\rho^n = \mathrm{d}t \otimes \rho_t^n$ with $(t \mapsto \rho_t^n) \in C_{\mathrm{w}}^+$, thanks to Lemmas 9, 10. By definition of $\langle \cdot, \cdot \rangle$ at (23), assumptions (K1)–(K2) and Cauchy–Schwarz, we obtain, for all fixed $n \in \mathbb{N}$,

$$\langle \rho^n, w \rangle = -\int_0^1 \langle K_t^* \rho_t^n, K_t^* \tilde{\rho}_t - f_t \rangle_{H_t} \, \mathrm{d}t \leq \int_0^1 \left\| K_t^* \rho_t^n \right\|_{H_t} \left\| K_t^* \tilde{\rho}_t - f_t \right\|_{H_t} \, \mathrm{d}t$$

$$\leq C \sup_t \left\| \rho_t^n \right\|_{\mathcal{M}(\overline{\Omega})} \int_0^1 \left\| K_t^* \tilde{\rho}_t - f_t \right\|_{H_t} \, \mathrm{d}t \leq C \left\| \rho^n \right\|_{\mathcal{M}(X)} \left\| K^* \tilde{\rho} - f \right\|_{L_H^2}, \tag{A7}$$

where $C > 0$ is the constant from (K2), and where we used that $\left\| \rho_t^n \right\|_{\mathcal{M}(\overline{\Omega})} = \left\| \rho^n \right\|_{\mathcal{M}(X)}$ for each $t \in [0, 1]$ (see Lemma 9). From (A2), (A7) and (A6), we get

$$-\langle \rho^n, w \rangle + \varphi(J_{\alpha,\beta}(\mu^n)) \geq -C \left\| \rho^n \right\|_{\mathcal{M}(X)} \left\| K^* \tilde{\rho} - f \right\|_{L_H^2} + C' J_{\alpha,\beta}(\mu^n)$$

$$\geq J_{\alpha,\beta}(\mu^n)[C' - C\alpha^{-1} \left\| K^* \tilde{\rho} - f \right\|_{L_H^2}],$$

for all $n \geq n_0$. By choosing $C' > 0$ sufficiently large, the above estimate contradicts (A5), showing that $\sup_n J_{\alpha,\beta}(\mu^n) < +\infty$.

In this case, Lemma 11 ensures that (ρ^n, μ^n) converges to $\mu^* = (\rho^*, m^*)$ in the sense of (A3), up to subsequences, and $\mu^* \in \mathcal{D}$, $\rho^* = \mathrm{d}t \otimes \rho_t^*$ with $(t \mapsto \rho_t^*) \in C_{\mathrm{w}}^+$.

In particular, $K^* \rho^n \rightharpoonup K^* \rho^*$ weakly in L_H^2 by Lemma 12. As $w_t = -K_t(K_t^* \tilde{\rho}_t - f_t)$ with $(K^* \tilde{\rho} - f) \in L_H^2$ (Lemma 12), from (K1), we deduce that $\langle \rho^n, w \rangle \to \langle \rho^*, w \rangle$ for $n \to +\infty$.

Recall that $J_{\alpha,\beta}$ is weak* lower semicontinuous (Lemma 11). As φ is lower-semicontinuous and monotonically increasing, we deduce that $\varphi \circ J_{\alpha,\beta}$ is weak* lower-semicontinuous. As $\{\mu^n\}$ is a minimizing sequence, by (A3) and Lemma 12, we conclude that μ^* solves (A4). □

A.4 Analysis for the Insertion Step

In this section, we show that under the assumptions (F1)–(F4) of Sect. 5.1 on the forward operators K_t^*, and assumptions (H1)–(H3) of Sect. 3.1 on the sampling spaces H_t, it is possible to tackle the insertion step problem (68) numerically, by means of gradient descent methods.

To this end, it is convenient to introduce the functionals $F, W, L \colon H^1([0, 1]; \overline{\Omega}) \to \mathbb{R}$ as

$$F(\gamma) := \frac{W(\gamma)}{L(\gamma)}, \quad W(\gamma) := -\int_0^1 w_t(\gamma(t)) \, \mathrm{d}t, \quad L(\gamma) := \frac{\beta}{2} \int_0^1 |\dot{\gamma}(t)|^2 \, \mathrm{d}t + \alpha. \tag{A8}$$

As observed in Remark 9, the only case of interest is when the minimum value of the problem at (68) is strictly negative. Thus, we assume to be in such situation, and consider

$$\min_{\gamma \in H^1([0,1]; \overline{\Omega})} F(\gamma). \tag{A9}$$

As discussed in Sect. 5.1, we are interested in computing stationary points of F by gradient descent. In order to make this possible, we first extend F to the Hilbert space $H^1 := H^1([0, 1]; \mathbb{R}^d)$, in a way that the set of stationary points of F is not altered. To be more precise, by assumptions (F1)–(F3), we have that the dual variable $w_t := -K_t(K_t^* \tilde{\rho}_t - f_t)$ belongs to $C^{1,1}(\overline{\Omega})$ for a.e. $t \in (0, 1)$. Additionally, (F4) implies that supp w_t, supp $\nabla w_t \subset E$ for a.e. $t \in (0, 1)$, where $E \Subset \Omega$ is closed and convex. We can then extend w_t to the whole $\mathbb{R}^d$ by setting $w_t(x) := 0$ for all $x \in \mathbb{R}^d \smallsetminus \overline{\Omega}$ and a.e. $t \in (0, 1)$. Consequently, the functional F is well-defined via (A8) over the space H^1.

In the above setting, we are able to prove that F is continuously Fréchet differentiable over H^1 (Proposition 16). Denote by $D_\gamma F \in (H^1)^*$ the Fréchet derivative of F at γ. We also show that stationary points of F, i.e., curves $\gamma^* \in H^1$ such that $D_{\gamma^*} F = 0$, satisfy $\gamma^*([0, 1]) \subset E \subset \Omega$ whenever $F(\gamma^*) \neq 0$ (Proposition 17). Therefore, problem (A9) is equivalent to

$$\min_{\gamma \in H^1} F(\gamma). \tag{A10}$$

We can now apply the gradient descent algorithm to compute stationary points of F, in the attempt of approximating solutions to (A10), and hence to (A9). The main result of this section states that descent sequences for F, in the sense of (71), converge (subsequentially) to stationary points.

Theorem 14 *Assume (F1)–(F4) as in Sect. 5.1 and (H1)–(H3) as in Sect. 3.1. Assume given $(t \mapsto \tilde{\rho}_t) \in C_w$, $f \in L_H^2$ and $\alpha, \beta > 0$. For a.e. $t \in (0, 1)$ set $w_t := -K_t(K_t^* \tilde{\rho}_t - f_t)$ and $w_t(x) := 0$ for all $x \in \mathbb{R}^d \smallsetminus \overline{\Omega}$. Then, the corresponding functional $F \colon H^1 \to \mathbb{R}$ defined via (A8) is continuously Fréchet differentiable. If $\{\gamma^n\}$ in H^1 is a descent sequence for F in the sense of (71) then, up to subsequences, $\gamma^n \to \gamma^*$ strongly in H^1. Any such accumulation point γ^* satisfies $F(\gamma^*) < 0$ and is stationary for F, that is, $D_{\gamma^*} F = 0$. Moreover, $\gamma^*([0, 1]) \subset E$, where $E \Subset \Omega$ is the closed convex set in (F4).*

The proof of Theorem 14, postponed to Sect. 1, relies on differentiability results for F and on properties of its stationary points, as discussed in Sect. 1. Finally, in Sect. 1, we provide an implementable criterion to determine whether the minimum of (68) is strictly negative.

A.4.1 Differentiability of *F* and Stationary Points

In this section, we discuss Fréchet differentiability and stationary points properties for the (extended) operator $F: H^1 \to \mathbb{R}$ defined at (A8). Before proceeding with the discussion, we establish a few notations and make some remarks on assumptions (F1)–(F4).

In the following, for any closed $S \subseteq \mathbb{R}^d$, we denote by $C^{1,1}(S)$ the space of differentiable maps $\varphi: S \to \mathbb{R}$ such that the norm

$$\|\varphi\|_{C^{1,1}(\overline{\Omega})} := \|\varphi\|_\infty + \|\nabla\varphi\|_\infty + \mathrm{Lip}(\nabla\varphi), \quad \mathrm{Lip}(\nabla\varphi) := \sup_{x \neq y} \frac{|\nabla\varphi(x) - \nabla\varphi(y)|}{|x - y|},$$

is finite, where $\nabla: C^{1,1}(S) \to C(S; \mathbb{R}^d)$ is the gradient operator. Notice that in this case, ∇ is linear and continuous. We will also consider the Bochner space $L^2([0, 1]; C^{1,1}(S))$ equipped with the norm $\|w\|_{1,1}^2 := \int_0^1 \|w_t\|_{C^{1,1}(S)}^2 \, dt$. Finally, for two Banach spaces X, Y and a continuously Fréchet differentiable map $G: X \to Y$, we denote the differential of G by $DG: X \to Y^*$, with evaluation at $u \in X$ given by the linear continuous functional $v \mapsto D_u G(v)$ belonging to Y^*.

Remark 11 Notice that (F3) also holds for $\rho \in C^{1,1}(\overline{\Omega})^*$: indeed, let $\{\rho^n\}$ sequence in $\mathcal{M}(\overline{\Omega})$ be such that $\|\rho^n - \rho\|_{C^{1,1}(\overline{\Omega})^*} \to 0$ as $n \to +\infty$. Then, the maps $t \mapsto K_t^* \rho^n$ are strongly measurable at each fixed n by (F3). From (F2) we have $\|K_t^* \rho^n - K_t^* \rho\|_{H_t} \le C \|\rho^n - \rho\|_{C^{1,1}(\overline{\Omega})^*} \to 0$ for a.e. $t \in (0, 1)$, implying that $t \mapsto K_t^* \rho$ is strongly measurable by [15, Remark 3.4].

A way to ensure that (F1)–(F4) hold is as follows. First notice that if K_t^* satisfies (F1)–(F3) and $D: C^{1,1}(\overline{\Omega}) \to C^{1,1}(\overline{\Omega})$ is a linear bounded operator, then one can check that also $\tilde{K}_t^* := K_t^* D^*$ satisfies (F1)–(F3). Now let $E' \Subset E$ be closed and let ξ_E be a cut-off function such that

$$\xi_E \in C^{1,1}(\overline{\Omega}), \ \ 0 \le \xi_E \le 1, \ \ \xi_E = 1 \text{ in } E', \ \ \xi_E = 0 \text{ in } \overline{\Omega} \smallsetminus E. \tag{A11}$$

Defining D by $D\varphi := \xi_E \varphi$ for all $\varphi \in C^{1,1}(\overline{\Omega})$ yields that $\tilde{K}_t = K_t D$ satisfies (F4).

In order to show that F is differentiable, we first investigate the regularity for dual variables $t \mapsto w_t$ of the form considered in (68). The differentiability properties for F are considered afterward.

Lemma 15 *Assume (F1)–(F4), (H1)–(H3) and let $g \in L_H^2$. For a.e. $t \in (0, 1)$ set $w_t := K_t g_t$ and $w_t(x) := 0$ for all $x \in \mathbb{R}^d \smallsetminus \overline{\Omega}$. Then, w belongs to $L^2([0, 1]; C^{1,1}(\mathbb{R}^d))$. Moreover $w_t, \nabla w_t$ are Carathéodory functions in $[0, 1] \times \mathbb{R}^d$, that is, $x \mapsto w_t(x)$, $x \mapsto \nabla w_t(x)$ are continuous for a.e. $t \in (0, 1)$ fixed and $t \mapsto w_t(x)$, $t \mapsto \nabla w_t(x)$ are measurable for all $x \in \mathbb{R}^d$ fixed.*

Proof We first show that $w \in L^2([0, 1]; C^{1,1}(\overline{\Omega}))$ and that $w, \nabla w$ are Carathéodory in $[0, 1] \times \overline{\Omega}$. In order to do so, let us check that the map $t \mapsto K_t g_t$ is strongly measurable in the classic sense [31, Ch II]. Since $C^{1,1}(\overline{\Omega})$ is separable, by the Pettis

measurability theorem [31, Ch II.1, Thm 2], strong measurability is equivalent to weak measurability, that is, we need to prove that

$$t \mapsto \langle \rho, K_t g_t \rangle_{C^{1,1}(\overline{\Omega})^*, C^{1,1}(\overline{\Omega})} \tag{A12}$$

is measurable for each $\rho \in C^{1,1}(\overline{\Omega})^*$. Note that $\langle \rho, K_t g_t \rangle_{C^{1,1}(\overline{\Omega})^*, C^{1,1}(\overline{\Omega})} = \langle K_t^* \rho, g_t \rangle_{H_t}$ by (F1). Moreover, $t \mapsto K_t^* \rho$ is strongly measurable by (F3) and Remark 11. Since g is strongly measurable (as $g \in L^2_H$), by [15, Remark 3.4], we conclude that $t \mapsto \langle K_t^* \rho, g_t \rangle_{H_t}$ is measurable. Therefore, the measurability of the map at (A12) follows.

From (F1)–(F2), we infer $\int_0^1 \|K_t g_t\|^2_{C^{1,1}(\overline{\Omega})} \, dt < +\infty$, since $g \in L^2_H$. By [31, Ch II.2, Thm 2], we then conclude $w \in L^2([0, 1]; C^{1,1}(\overline{\Omega}))$. In particular the maps $x \mapsto w_t(x)$, $x \mapsto \nabla w_t(x)$ are continuous for a.e. $t \in (0, 1)$ fixed and x varying in $\overline{\Omega}$. Let now $x \in \overline{\Omega}$ be fixed. By (F1) and Remark 11, we have $w_t(x) = \langle \delta_x, K_t g_t \rangle_{\mathcal{M}(\overline{\Omega}), C(\overline{\Omega})} = \langle K_t^* \delta_x, g_t \rangle_{H_t}$. As the map $t \mapsto K_t^* \delta_x$ is strongly measurable by (F3), and g is strongly measurable since it belongs to L^2_H, from [15, Remark 3.4], we conclude that $t \mapsto \langle K_t^* \delta_x, g_t \rangle_{H_t}$ is measurable. Thus, w is Carathéodory in $[0, 1] \times \overline{\Omega}$. Similarly, we have $\partial_{x_i} w_t(x) = \langle \nabla^* e_i \delta_x, K_t g_t \rangle_{C^{1,1}(\overline{\Omega})^*, C^{1,1}(\overline{\Omega})}$ for all $i = 1, \ldots, d$, where e_i is the i-th coordinate vector in $\mathbb{R}^d$. Notice that $\nabla^* e_i \delta_x \in C^{1,1}(\overline{\Omega})^*$. Hence, the measurability of $t \mapsto \nabla w_t(x)$ is implied by setting $\rho = \nabla^* e_i \delta_x$ in (A12), showing that ∇w is Carathéodory in $[0, 1] \times \overline{\Omega}$. Finally, the facts that $w \in L^2([0, 1]; \mathbb{R}^d)$ and that $w, \nabla w$ are Carathéodory in $[0, 1] \times \mathbb{R}^d$, follow since w_t is extended to zero in $\mathbb{R}^d \setminus \overline{\Omega}$ and (F4) holds.   □

Proposition 16 *Assume (F1)–(F4), (H1)–(H3). Let $(t \mapsto \tilde{\rho}_t) \in C_w$, $f \in L^2_H$ and $\alpha, \beta > 0$ be given. For a.e. $t \in (0, 1)$ set $w_t := -K_t(K_t^* \tilde{\rho}_t - f_t)$ and $w_t(x) := 0$ for all $x \in \mathbb{R}^d \setminus \overline{\Omega}$. Then, the corresponding functionals F, W, L defined at (A8) are continuously Fréchet differentiable in $H^1 := H^1([0, 1]; \mathbb{R}^d)$. The derivatives of F, W, L at $\gamma \in H^1$ are given by*

$$D_\gamma F(\eta) = \frac{D_\gamma W(\eta)}{L(\gamma)} - F(\gamma) \frac{D_\gamma L(\eta)}{L(\gamma)}, \tag{A13}$$

$$D_\gamma W(\eta) = -\int_0^1 \nabla w_t(\gamma(t)) \cdot \eta(t) dt, \quad D_\gamma L(\eta) = \beta \int_0^1 \dot{\gamma}(t) \cdot \dot{\eta}(t) \, dt, \tag{A14}$$

for each $\eta \in H^1$. In addition, we have

$$\sup_{\gamma \in H^1} |F(\gamma)| \leq \frac{\|w\|_{1,1}}{\alpha}, \quad \sup_{\gamma \in H^1} \|D_\gamma F\|_{(H^1)^*} \leq \frac{\|w\|_{1,1}}{\alpha} \left(1 + \sqrt{\frac{\beta}{2\alpha}} \right), \tag{A15}$$

where $\|w\|^2_{1,1} := \int_0^1 \|w_t\|^2_{C^{1,1}(\overline{\Omega})} \, dt$. Last, the map $\gamma \mapsto D_\gamma F$ is locally Lipschitz, that is, for all $R > 0$ fixed it holds

$$\left\| D_{\gamma^1} F - D_{\gamma^2} F \right\|_{(H^1)^*} \leq (C_1 R + C_2) \left\| \gamma^1 - \gamma^2 \right\|_{H^1}, \tag{A16}$$

for all $\gamma^i \in H^1$ such that $\left\| \gamma^i \right\|_{H^1} \leq R$, $i = 1, 2$, where $C_1, C_2 > 0$ are constants depending only on α, β and w.

Proof The continuous Fréchet differentiability of L is standard, and the proof is omitted. Moreover, continuous differentiability of F and formula (A13) follow from continuous differentiability of W and L, and from the quotient rule, given that $L \geq \alpha > 0$. Therefore, let us show that W is continuously differentiable with derivative as in (A14). Since $(t \mapsto \tilde{\rho}_t) \in C_w$, by Lemma 12, we have that $t \mapsto K_t^* \tilde{\rho}_t$ belongs to L_H^2, so that also $g := -(K^*\tilde{\rho} - f)$ belongs to L_H^2. Set $w_t := K_t^* g_t$ and $w_t(x) := 0$ for all $x \in \mathbb{R}^d \setminus \overline{\Omega}$. By Lemma 15, we know that $w \in L^2([0, 1]; C^{1,1}(\mathbb{R}^d))$ and w, ∇w are Carathéodory maps in $[0, 1] \times \mathbb{R}^d$. In particular, for a fixed $\gamma \in H^1$, the maps $t \mapsto w_t(\gamma(t))$, $t \mapsto \nabla w_t(\gamma(t))$ are measurable [27, Proposition 3.7].

Since $w \in L^2([0, 1]; C^{1,1}(\mathbb{R}^d))$, we can proceed as in the proof of Theorem 3.37 in [27] and show that the Gâteaux derivative of W at γ, along the direction η, is given by the first formula in (A14).

We are left to prove that $\gamma \mapsto D_\gamma W$ is continuous from H^1 into $(H^1)^*$. To this end, fix $\gamma^1, \gamma^2 \in H^1$ and notice that

$$\left\| D_{\gamma^1} W - D_{\gamma^2} W \right\|_{(H^1)^*} \leq \sup_{\substack{\eta \in H^1, \\ \|\eta\|_{H^1} \leq 1}} \|\eta\|_\infty \int_0^1 \mathrm{Lip}(\nabla w_t) \left| \gamma^1(t) - \gamma^2(t) \right| \, dt \tag{A17}$$

$$\leq \sqrt{2} \, \|w\|_{1,1} \left\| \gamma^1 - \gamma^2 \right\|_{H^1},$$

where in the last inequality, we employed Cauchy–Schwarz and the estimate $\|\eta\|_\infty \leq \sqrt{2} \, \|\eta\|_{H^1}$. Notice that (A17) shows that the map $\gamma \mapsto D_\gamma W$ is Lipschitz from H^1 into $(H^1)^*$. Thus, in particular, W is continuously Fréchet differentiable. We will now prove the estimates at (A15)-(A16). The first bound in (A15) follows immediately from the definition of F, the fact that $w \in L^2([0, 1]; C^{1,1}(\mathbb{R}^d))$, and the estimate $L \geq \alpha > 0$. As for the second estimate in (A15), by (A13) and the triangle inequality, we have

$$\left\| D_\gamma F \right\|_{(H^1)^*} \leq \frac{\left\| D_\gamma W \right\|_{(H^1)^*}}{L(\gamma)} + \left\| \frac{D_\gamma L}{L(\gamma)} \right\|_{(H^1)^*} |F(\gamma)| . \tag{A18}$$

Notice that $\left\| D_\gamma W \right\|_{(H^1)^*} \leq \|w\|_{1,1}$, thanks to (A14) and Hölder's inequality. Moreover, by (A14) and Hölder's inequality,

$$\left\| \frac{D_\gamma L}{L(\gamma)} \right\|_{(H^1)^*} \leq \frac{\beta \left(\int_0^1 |\dot{\gamma}|^2 \, dt \right)^{1/2}}{\frac{\beta}{2} \int_0^1 |\dot{\gamma}|^2 \, dt + \alpha} \leq \sqrt{\frac{\beta}{2\alpha}}, \tag{A19}$$

where the second estimate is obtained by noting that the real map $s \mapsto \beta s / (\beta s^2 / 2 + \alpha)$ is differentiable, with maximum value given by $\sqrt{\beta/(2\alpha)}$. By the first estimate in (A15) and the fact that $L \geq \alpha$, from (A18)–(A19), we conclude (A15). Finally, we prove (A16). To this end, fix $R > 0$ and $\gamma^1, \gamma^2 \in H^1$ such that $\left\| \gamma^1 \right\|_{H^1}, \left\| \gamma^2 \right\|_{H^1} \leq R$.

Note that as a consequence of (A13), we get

$$\left\| D_{\gamma^1} F - D_{\gamma^2} F \right\|_{(H^1)^*} \leq \left\| \frac{D_{\gamma^1} W}{L(\gamma^1)} - \frac{D_{\gamma^2} W}{L(\gamma^2)} \right\|_{(H^1)^*}$$
$$+ \left\| F(\gamma^1) \frac{D_{\gamma^1} L}{L(\gamma^1)} - F(\gamma^2) \frac{D_{\gamma^2} L}{L(\gamma^2)} \right\|_{(H^1)^*}. \quad \text{(A20)}$$

Concerning the first term in (A20), observe that by the estimate $L \geq \alpha$,

$$\left| \frac{1}{L(\gamma^1)} - \frac{1}{L(\gamma^2)} \right| = \left| \frac{L(\gamma^1) - L(\gamma^2)}{L(\gamma^1) L(\gamma^2)} \right| \leq \frac{\beta}{2\alpha^2} \left| \int_0^1 (\left| \dot\gamma^1 \right| + \left| \dot\gamma^2 \right|)(\left| \dot\gamma^1 \right| - \left| \dot\gamma^2 \right|) \, dt \right|$$
$$\leq \frac{\beta}{2\alpha^2} \left(\left\| \gamma^1 \right\|_{H^1} + \left\| \gamma^2 \right\|_{H^1} \right) \left(\int_0^1 \left| \dot\gamma^1 - \dot\gamma^2 \right|^2 \, dt \right)^{1/2}$$
$$\leq R \frac{\beta}{\alpha^2} \left\| \gamma^1 - \gamma^2 \right\|_{H^1}. \quad \text{(A21)}$$

Recall that $\gamma \mapsto D_\gamma W$ is bounded, with $\left\| D_\gamma W \right\|_{(H^1)^*} \leq \|w\|_{1,1}$. Also the map $\gamma \mapsto 1/L(\gamma)$ is bounded by $1/\alpha$. Therefore by the Lipschitz estimates (A17) and (A21), we obtain

$$\left\| \frac{D_{\gamma^1} W}{L(\gamma^1)} - \frac{D_{\gamma^2} W}{L(\gamma^2)} \right\|_{(H^1)^*} \leq \left\| D_{\gamma^1} W \right\|_{(H^1)^*} \left| \frac{1}{L(\gamma^1)} - \frac{1}{L(\gamma^2)} \right|$$
$$+ \frac{1}{L(\gamma^2)} \left\| D_{\gamma^1} W - D_{\gamma^2} W \right\|_{(H^1)^*} \quad \text{(A22)}$$
$$\leq \|w\|_{1,1} \left(\frac{R\beta}{\alpha^2} + \frac{\sqrt{2}}{\alpha} \right) \left\| \gamma^1 - \gamma^2 \right\|_{H^1}.$$

We now estimate the second term in (A20). First note that as a consequence of (A15) and of the mean value theorem, the map $\gamma \mapsto F(\gamma)$ is bounded by $\|w\|_{1,1}/\alpha$ and has (global) Lipschitz constant bounded by $\|w\|_{1,1}(1 + \sqrt{\beta/(2\alpha)})/\alpha$. Moreover, the map $\gamma \mapsto G(\gamma) := D_\gamma L/L(\gamma)$ is bounded by $\sqrt{\beta/(2\alpha)}$ (see (A19)). It is easy to check that G is continuously Fréchet differentiable. Employing the estimates $L \geq \alpha$ and (A19), we also check that $\gamma \mapsto D_\gamma G$ is bounded uniformly by $3\beta/(2\alpha)$. By the mean value theorem, we then conclude that G is globally Lipschitz with constant controlled by $3\beta/(2\alpha)$. Arguing as in (A22), we compute

$$\left\| F(\gamma^1) \frac{D_{\gamma^1} L}{L(\gamma^1)} - F(\gamma^2) \frac{D_{\gamma^2} L}{L(\gamma^2)} \right\|_{(H^1)^*} \leq \frac{\|w\|_{1,1}}{\alpha} \left(\sqrt{\frac{\beta}{2\alpha}} + \frac{2\beta}{\alpha} \right) \left\| \gamma^1 - \gamma^2 \right\|_{H^1}. \quad \text{(A23)}$$

The inequality at (A16) follows from (A20), (A22), (A23), and the proof is concluded. $\square$

Finally, we show that stationary points of F with nonzero energy are curves contained in $E \subset \Omega$.

Proposition 17 *Assume (F1)–(F4), (H1)–(H3). Let $(t \mapsto \tilde{\rho}_t) \in C_w$, $f \in L^2_H$ and $\alpha, \beta > 0$. For a.e. $t \in (0, 1)$ define $w_t := -K_t(K_t^* \tilde{\rho}_t - f_t) \in C^{1,1}(\overline{\Omega})$ and $w_t(x) := 0$ for all $x \in \mathbb{R}^d \setminus \overline{\Omega}$. Consider the corresponding functional F defined at (A8). If $\gamma^* \in H^1$ is a stationary point for F, that is, $D_{\gamma^*} F = 0$, then γ^* satisfies the following system is in the weak sense*

$$\beta F(\gamma) \ddot{\gamma}(t) = \nabla w_t(\gamma(t)) \quad \text{for all } t \in (0, 1), \qquad \dot{\gamma}(0) = \dot{\gamma}(1) = 0. \quad (A24)$$

If in addition $F(\gamma^*) \neq 0$, we have $\gamma^*([0, 1]) \subset E$, where $E \Subset \Omega$ is the closed convex set in (F4).

Proof By Lemma 15, we have that $w \in L^2([0, 1]; C^{1,1}(\mathbb{R}^d))$. Moreover, Proposition 16 ensures that F is continuously Fréchet differentiable over H^1. If γ^* is such that $D_{\gamma^*} F = 0$, from (A13)–(A14) and the inequality $L \geq \alpha > 0$, we deduce the weak formulation of (A24), i.e.,

$$-\int_0^1 \nabla w_t(\gamma^*(t)) \cdot \eta(t) \, dt = \beta F(\gamma^*) \int_0^1 \dot{\gamma}^*(t) \cdot \dot{\eta}(t) \, dt, \quad \text{for all } \eta \in H^1. \quad (A25)$$

Suppose that $F(\gamma^*) \neq 0$ and set $A := \{t \in [0, 1] : \gamma^*(t) \notin E\}$. Assume by contradiction that $A \neq \emptyset$. Note that $A \neq [0, 1]$, since $F(\gamma^*) \neq 0$ and (F4) holds. Since E is closed and γ^* is continuous, then A is relatively open in $[0, 1]$. Therefore, $A = \cup_{i \in \mathbb{N}} I_i$, with I_i pairwise disjoint, which are either of the form (a_i, b_i) with $0 < a_i < b_i < 1$, or $[0, a_i)$, or $(a_i, 1]$, with $0 < a_i < 1$, or empty. Assume that there exists $i \in \mathbb{N}$ such that $I_i = (a_i, b_i)$ with $0 < a_i < b_i < 1$. Let $\varphi \in C^1_c(a_i, b_i)$ and extend it to zero to the whole $[0, 1]$. Set $\eta := e_j \varphi$, with e_j the j-th coordinate vector in $\mathbb{R}^d$. Since supp $\nabla w_t \subset E$ for a.e. $t \in (0, 1)$ (see (F4)), $\gamma^*(t) \notin E$ for $t \in (a_i, b_i)$, and $F(\gamma^*) \neq 0$, testing (A25) against η yields $\int_{a_i}^{b_i} \dot{\gamma}_j^* \dot{\varphi} \, dt = 0$, where γ_j^* is the j-th component of γ^*. Therefore γ^* is linear in $[a_i, b_i]$. Since by construction $\gamma^*(a_i), \gamma^*(b_i) \in E$, by convexity of E, we obtain $\gamma^*(t) \in E$ for all $t \in [a_i, b_i]$, which is a contradiction. Assume now that there exists $i \in \mathbb{N}$ such that $I_i = [0, a_i)$ for some $0 < a_i < 1$. Let $\varphi \in L^2(0, a_i)$, extend it to zero in $[a_i, 1]$, and set $\eta(t) := -e_j \int_{a_i}^t \varphi(s) \, ds$. Testing (A25) against η, allows to conclude that γ^* is constant in $[0, a_i]$, which is a contradiction since by construction $\gamma^*(a_i) \in E$. Similarly, the remaining case $I_i = (a_i, 1]$ for some $0 < a_i < 1$ leads to a contradiction. Thus, we conclude that $A = \emptyset$, finishing the proof. $\square$

A.4.2 Gradient Descent

In this section, we prove Theorem 14 on descent sequences for the functional F at (A8). The proof relies on the following lemma.

Lemma 18 *Assume (H1)–(H3), (F1)–(F4). Let $(t \mapsto \tilde{\rho}_t) \in C_w$, $f \in L^2_H$, $\alpha, \beta > 0$. For a.e. $t \in (0, 1)$ define $w_t := -K_t(K_t^* \tilde{\rho}_t - f_t) \in C^{1,1}(\overline{\Omega})$ and $w_t(x) := 0$ for all $x \in \mathbb{R}^d \setminus \overline{\Omega}$. Consider the corresponding functional F defined at (A8). Then, for all*

M < 0, there exists R > 0 depending only on M, Ω, w, α, β, such that

$$\{\gamma \in H^1 : F(\gamma) \le M\} \subset \{\gamma \in H^1 : \|\gamma\|_{H^1} \le R\}. \tag{A26}$$

Moreover, let $\{\gamma^n\}$ in H^1 be a sequence such that

$$F(\gamma^n) \to c, \quad \|D_{\gamma^n} F\|_{(H^1)^*} \to 0, \quad as \ n \to +\infty, \tag{A27}$$

for some c < 0. Then, up to subsequences, $\gamma^n \to \gamma^*$ strongly in H^1. Any such accumulation point γ^* satisfies $F(\gamma^*) = c$ and is stationary for F, namely $D_{\gamma^*} F = 0$.

Proof Assume that $F(\gamma) \le M$ for some $M < 0$. Since $w \in L^2([0, 1]; C^{1,1}(\mathbb{R}^d))$ by Lemma 15,

$$\int_0^1 |\dot{\gamma}(t)|^2 \, \mathrm{d}t \le -\frac{2}{\beta} \left(\frac{\|w\|_{1,1}}{M} + \alpha \right), \tag{A28}$$

where we also used that $|W| \le \|w\|_{1,1}$ and $L > 0$. As supp $w_t \subset E$ for a.e. $t \in (0, 1)$ by (F4), and $L(\gamma) > 0$, the condition $F(\gamma) < 0$, together with the continuity of γ, implies the existence of some $\hat{t} \in [0, 1]$ such that $\gamma(\hat{t}) \in E$. (Otherwise, we would have $F(\gamma) = 0$.) Hence, we can estimate

$$\sup_{t \in [0,1]} |\gamma(t)| \le \sup_{t \in [0,1]} \left| \gamma(t) - \gamma(\hat{t}) \right| + \left| \gamma(\hat{t}) \right|$$
$$\le \int_0^1 |\dot{\gamma}(s)| \, \mathrm{d}s + \max_{x \in E} |x| \le \left(\int_0^1 |\dot{\gamma}(s)|^2 \, \mathrm{d}s \right)^{1/2} + \max_{x \in \overline{\Omega}} |x|. \tag{A29}$$

From (A28)–(A29), we immediately deduce (A26) for some $R > 0$. Assume now that $\{\gamma^n\}$ in H^1 satisfies (A27) for some $c < 0$. We will prove that $\{\gamma^n\}$ has at least one accumulation point with respect to the strong convergence of H^1. As $F(\gamma^n) \to c$ with $c < 0$, from (A26) we deduce that $\{\gamma^n\}$ is uniformly bounded in H^1. Hence, there exists $\gamma \in H^1$ such that $\gamma^n \rightharpoonup \gamma$ weakly in H^1 and $\gamma^n \to \gamma$ uniformly in $[0, 1]$, up to subsequences (not relabelled). We will now prove that $\gamma^n \to \gamma$ strongly in H^1. By the uniform convergence $\gamma^n \to \gamma$ and regularity of w, dominated convergence yields

$$W(\gamma^n) \to W(\gamma) \text{ as } n \to +\infty. \tag{A30}$$

Assume that $\dot{\gamma} \not\equiv 0$ and define, for n sufficiently large,

$$\eta^n := \frac{1}{2} \gamma^n + \frac{\alpha}{\beta \int_0^1 \dot{\gamma}^n \cdot \dot{\gamma} \, \mathrm{d}t} \gamma, \quad \eta := \frac{1}{2} \gamma + \frac{\alpha}{\beta \int_0^1 |\dot{\gamma}|^2 \, \mathrm{d}t} \gamma.$$

Notice that $\eta^n \rightharpoonup \eta$ weakly in H^1. In particular, $\{\eta^n\}$ is bounded in H^1, so that

$$\left| D_{\gamma^n} F(\eta^n) \right| \le \|D_{\gamma^n} F\|_{(H^1)^*} \|\eta^n\|_{H^1} \to 0 \text{ as } n \to +\infty, \tag{A31}$$

where we employed continuous differentiability of F (Proposition 16) and (A27). Notice now that $\eta^n \to \eta$ strongly in $L^2([0, 1]; \mathbb{R}^d)$, by Sobolev embeddings. Recalling

(A14) and using the uniform convergence $\gamma^n \to \gamma$, together with the regularity of w, by dominated convergence we get

$$D_{\gamma^n} W(\eta^n) \to D_\gamma W(\eta) \quad \text{as} \quad n \to +\infty. \tag{A32}$$

Moreover, by definition of η^n and (A14) one can check that $D_{\gamma^n} L(\eta^n) = L(\gamma^n)$ for all $n \in \mathbb{N}$. Taking the latter into account and substituting γ^n and η^n into (A13) yields

$$L(\gamma^n) D_{\gamma^n} F(\eta^n) = D_{\gamma^n} W(\eta^n) - W(\gamma^n) \tag{A33}$$

for all $n \in \mathbb{N}$. Recalling that $\{\gamma^n\}$ is bounded in H^1, we also infer that $\{L(\gamma^n)\}$ is bounded. Therefore, we can invoke (A30), (A31), (A32) to pass to the limit in (A33) and infer

$$D_\gamma W(\eta) = W(\gamma). \tag{A34}$$

Substituting the definition of η into (A14) yields $D_\gamma W(\eta) = L(\gamma) D_\gamma W(\gamma)/D_\gamma L(\gamma)$. By definition of F, the previous identity, and (A34), we get that $F(\gamma) = D_\gamma W(\gamma)/D_\gamma L(\gamma)$. On the other hand, substituting γ^n and γ into (A13), and recalling that $D_{\gamma^n} F(\gamma) \to 0$ by (A27), and that $L(\gamma^n) \geq \alpha > 0$, results in

$$\left[D_{\gamma^n} W(\gamma) - F(\gamma^n) D_{\gamma^n} L(\gamma) \right] \to 0 \quad \text{as} \quad n \to +\infty. \tag{A35}$$

Concerning (A35), first note that $F(\gamma^n) \to c$ by assumption. Moreover, since $\gamma^n \rightharpoonup \gamma$ weakly in H^1 and $w \in L^2([0, 1]; C^1(\mathbb{R}^d))$, by dominated convergence we see that $D_{\gamma^n} W(\gamma) \to D_\gamma W(\gamma)$ and $D_{\gamma^n} L(\gamma) \to D_\gamma L(\gamma)$. Thus from (A35), we deduce that $D_\gamma W(\gamma) = c \, D_\gamma L(\gamma)$. Recalling that $F(\gamma) = D_\gamma W(\gamma)/D_\gamma L(\gamma)$, we conclude $F(\gamma) = c$, so that $F(\gamma^n) \to F(\gamma)$ (recalling (A27)). By the convergence $F(\gamma^n) \to F(\gamma)$, definition of F and (A30), we conclude that $L(\gamma^n) \to L(\gamma)$. By definition of L, the latter is equivalent to $\int_0^1 |\dot\gamma^n|^2 \, \mathrm{d}t \to \int_0^1 |\dot\gamma|^2 \, \mathrm{d}t$ as $n \to +\infty$. Since $\gamma^n \rightharpoonup \gamma$ weakly in H^1, we infer $\gamma^n \to \gamma$ strongly in H^1. Setting $\gamma^* := \gamma$ concludes the convergence statement. Assume now that $\dot\gamma \equiv 0$. As $\dot\gamma \equiv 0$, by (A14) we obtain $D_\gamma L = 0$. As $\gamma^n \rightharpoonup \gamma$ weakly in H^1, by dominated convergence we get $D_{\gamma^n} W(\eta) \to D_\gamma W(\eta)$ and $D_{\gamma^n} L(\eta) \to D_\gamma L(\eta) = 0$. Hence, taking the limit as $n \to +\infty$ in (A13) evaluated on γ^n and $\eta \in H^1$, and recalling that $\{F(\gamma^n)\}$ is bounded, yields $L(\gamma^n) D_{\gamma^n} F(\eta) \to D_\gamma W(\eta)$. As $\{L(\gamma^n)\}$ is bounded, by (A27) we get $D_\gamma W = 0$. We now claim that

$$\int_0^1 |\dot\gamma^n|^2 \, \mathrm{d}t \to 0 \quad \text{as} \quad n \to +\infty. \tag{A36}$$

Assume by contradiction that (A36) does not hold. Then, there exists a subsequence (not relabeled) such that $\int_0^1 |\dot\gamma^n|^2 \, \mathrm{d}t \geq q > 0$ for all $n \in \mathbb{N}$. Given that $\{\gamma^n\}$ is bounded in H^1, without loss of generality we can assume that $\int_0^1 |\dot\gamma^n|^2 \, \mathrm{d}t \to q_0$ as $n \to +\infty$, for some $q_0 > 0$. Define

$$\eta^n := \frac{1}{2} \gamma^n + \frac{\alpha}{\beta \int_0^1 |\dot\gamma^n|^2 \, \mathrm{d}t} \gamma^n, \qquad \eta := \frac{1}{2} \gamma + \frac{\alpha}{\beta q_0} \gamma.$$

Clearly, $\eta^n \rightharpoonup \eta$ weakly in H^1. Arguing as in the proof of (A34), we conclude that $D_\gamma W(\eta) = W(\gamma)$. Recalling that $D_\gamma W = 0$, we infer $W(\gamma) = 0$. Now notice that $L(\gamma^n) \to \beta q_0/2 + \alpha > 0$, because $\int_0^1 |\dot\gamma^n|^2 \, dt \to q_0$. By (A30) and the fact that $W(\gamma) = 0$, we then conclude that $F(\gamma^n) \to 0$, which contradicts (A27). Thus, (A36) holds. As $\gamma^n \rightharpoonup \gamma$ weakly in H^1 and $\dot\gamma \equiv 0$, from (A36), we infer that $\gamma^n \to \gamma$ strongly in H^1. Setting $\gamma^* := \gamma$ concludes the convergence statement. Finally, suppose that $\gamma^n \to \gamma^*$ strongly in H^1 (subsequentially). As F is continuously Fréchet differentiable (Proposition 16), thanks to (A27), we obtain that $F(\gamma^*) = c$ and $D_{\gamma^*} F = 0$. $\square$

Proof of Theorem 14 The functional F is continuously Fréchet differentiable as a consequence of Proposition 16. Moreover, recall that DF is locally Lipschitz (Proposition 16), with local Lipschitz constant in a ball $\{\gamma \in H^1 : \|\gamma\|_{H^1} \le R\}$ estimated by $C_1 R + C_2$, for some constants $C_1, C_2 > 0$ depending only on w, α, β. Assume now that $\{\gamma^n\}$ in H^1 is a descent sequence in the sense of (71) and set $M := F(\gamma^0) < 0$. By (A26) in Lemma 18, we can find some $R > 0$, depending only on $M, \Omega, w, \alpha, \beta$, such that

$$\{\gamma \in H^1 : F(\gamma) \le F(\gamma^0)\} \subset \{\gamma \in H^1 : \|\gamma\|_{H^1} \le R\}. \tag{A37}$$

For such R, consider the corresponding local Lipschitz constant $C_1 R + C_2$ for DF. It well-known that the Armijo–Goldstein or Backtracking-Armijo rules for the stepsize $\{\delta_n\}$ guarantee that

$$0 < A < \delta_n < B < \frac{2}{C_1 R + C_2}, \tag{A38}$$

for some $A, B > 0$ and all $n \in \mathbb{N}$. It is also standard that (A37)–(A38) and regularity of F imply $\|D_{\gamma^n} F\|_{(H^1)^*} \to 0$ and $F(\gamma^{n+1}) \le F(\gamma^n)$ for all $n \in \mathbb{N}$. Since $F(\gamma^0) < 0$ and $|F| \le \|w\|_{1,1}/\alpha$ by (A15), from the monotonicity of $\{F(\gamma^n)\}$, we infer that $F(\gamma^n) \to c$ for some $c < 0$. Therefore $\{\gamma^n\}$ satisfies (A27), so that we can apply Lemma 18 and infer that $\{\gamma^n\}$ is strongly precompact in H^1, and that any strong accumulation point γ^* satisfies $F(\gamma^*) = c$ and $D_{\gamma^*} F = 0$. Since $c < 0$, by Proposition 17, we also obtain that $\gamma^*([0, 1]) \subset E$, concluding. $\square$

A.4.3 Test for Zero Minimum

Proposition 19 *Assume (H1)–(H3), (F1)–(F4). Let $(t \mapsto \tilde\rho_t) \in C_w$, $f \in L_H^2$, $\alpha, \beta > 0$. For a.e. $t \in (0, 1)$, define $w_t := -K_t(K_t^* \tilde\rho_t - f_t) \in C^{1,1}(\overline\Omega)$ and $w_t(x) := 0$ for all $x \in \mathbb{R}^d \setminus \overline\Omega$. Consider the corresponding functional F defined at (A8). Then, 0 is the minimum of (68) if and only if*

$$\int_0^1 \max_{x \in \overline\Omega} w_t(x) \, dt \le 0. \tag{A39}$$

Proof First note that w is a Carathéodory map in $[0, 1] \times \overline\Omega$ by the proof of Lemma 15, since $g_t := -K_t^* \tilde\rho_t + f_t$ belongs to L_H^2 by Lemma 12. Therefore, w is also Carathéodory in $[0, 1] \times E$, because $\text{supp}\, w_t \subset E$ with $E \Subset \Omega$ closed and convex

(see (F4)). Seeing that E is compact, we can apply Theorem 18.19 in [2] to obtain that the scalar map $t \mapsto \max_{x \in E} w_t(x)$ is measurable, and that there exists a measurable curve $\hat{\gamma} \colon [0, 1] \to E$ such that $\hat{\gamma}(t) \in \arg\max_{x \in E} w_t(x)$ for all $t \in [0, 1]$. By the condition $\mathrm{supp}\, w_t \subset E$, we infer $\max_{x \in E} w_t(x) = \max_{x \in \overline{\Omega}} w_t(x)$ for a.e. $t \in (0, 1)$, showing that $t \mapsto \max_{x \in \overline{\Omega}} w_t(x)$ is measurable. Thus, the integral in (A39) is well defined. Moreover, by construction, $w_t(\hat{\gamma}(t)) = \max_{x \in \overline{\Omega}} w_t(x)$ for a.e. $t \in (0, 1)$. Assume that 0 is the minimum of (68). By the inequality $L(\gamma) \geq \alpha > 0$, we infer

$$\int_0^1 w_t(\gamma(t))\, \mathrm{d}t \leq 0 \qquad\qquad (\text{A40})$$

for all $\gamma \in H^1([0, 1]; \overline{\Omega})$. As $E \Subset \Omega$, we can find a sequence $\{\gamma_n\}$ in $H^1([0, 1]; \overline{\Omega})$ such that $\gamma_n \to \hat{\gamma}$ a.e. in $(0, 1)$ as $n \to +\infty$. Since $w_t \in C^{1,1}(\overline{\Omega})$ for a.e. t fixed, we have $w_t(\gamma_n(t)) \to w_t(\hat{\gamma}(t))$ a.e. in $(0, 1)$. Moreover, $|w_t(\gamma_n(t))| \leq \|w_t\|_{C^{1,1}(\overline{\Omega})}$. We can then substitute γ^n in (A40) and apply dominated convergence to infer that $\hat{\gamma}$ satisfies (A40) as well. By maximality of $\hat{\gamma}$, we conclude (A39). Conversely, assume that (A39) holds. For all $\gamma \in H^1([0, 1]; \overline{\Omega})$, we get

$$F(\gamma) = -\frac{\int_0^1 w_t(\gamma(t))\, \mathrm{d}t}{L(\gamma)} \geq -\frac{\int_0^1 \max_{x \in \overline{\Omega}} w_t(x)\, \mathrm{d}t}{L(\gamma)}.$$

Since $L(\gamma) > 0$, we infer that 0 is the minimum of (68). $\square$

A.5 Analysis for the Sliding Step

In this section, we rigorously justify the sliding step discussed in Sect. 5.2.2, showing that under Assumption 3, the target functional at (73) is differentiable. To fix notations, Let $N \in \mathbb{N}$, $N \geq 1$ and $c_j \in \mathbb{R}$, $c_j > 0$ be fixed. We denote by $(H^1_\Omega)^N$ the space of points $\Gamma := (\gamma_1, \ldots, \gamma_N)$ with $\gamma_j \in H^1_\Omega := H^1([0, 1]; \overline{\Omega})$. For $\Gamma \in (H^1_\Omega)^N$, we define the measure

$$\mu(\Gamma) := \sum_{j=1}^N c_j \mu_{\gamma_j} \in \mathcal{M},$$

where $\mu_{\gamma_j} := (\rho_{\gamma_j}, m_{\gamma_j}) \in \mathcal{C}_{\alpha,\beta}$, according to (14). Define the functional $\Phi \colon (H^1_\Omega)^N \to \mathbb{R}$ by

$$\Phi(\Gamma) = \Phi(\gamma_1, \ldots, \gamma_N) := T_{\alpha,\beta,\boldsymbol{c}}(\mu(\Gamma)), \qquad\qquad (\text{A41})$$

where $T_{\alpha,\beta,\boldsymbol{c}}$ is defined in (73), for some $f \in L^2_H$ and $\alpha, \beta > 0$ fixed. We also recall the notation $H^1 := H^1([0, 1]; \mathbb{R}^d)$.

Proposition 20 *Assume (F1)–(F3), (H1)–(H3). The functional Φ at (A41) is continuously Fréchet differentiable at each $\Gamma \in (H^1_\Omega)^N$ such that $\gamma_j([0, 1]) \subset \Omega$ for*

$j = 1, \ldots, N$, *with*

$$D_\Gamma \Phi(\Theta) = \sum_{j=1}^{N} c_j \, D_{\gamma_j} F(\eta_j) \,, \tag{A42}$$

for all $\Theta = (\eta_1, \ldots, \eta_N)$, $\eta_j \in H^1$. *Here, F is as in* (A8), *with respect to the dual variable*

$$w_t := -K_t \left(\sum_{j=1}^{N} c_j a_{\gamma_j} K_t^* \delta_{\gamma_j(t)} - f_t \right). \tag{A43}$$

Proof Let $\Gamma \in (H_\Omega^1)^N$ with $\gamma_j([0, 1]) \subset \Omega$. Let $\Theta = (\eta_1, \ldots, \eta_N)$ with $\eta_j \in H^1$ and $\varepsilon_0 > 0$ sufficiently small, so that $(\gamma_j + \varepsilon\eta_j)([0, 1]) \subset \Omega$ for each $0 < \varepsilon < \varepsilon_0$, $j = 1, \ldots, N$.

By Lemma 2

$$J_{\alpha,\beta}(\mu(\Gamma)) = J_{\alpha,\beta}(\mu(\Gamma + \varepsilon\Theta)) = \sum_{j=1}^{N} c_j \,, \tag{A44}$$

for all $0 < \varepsilon < \varepsilon_0$. Define w as in (A43) and notice that $w \in L^2([0, 1]; C^{1,1}(\overline{\Omega}))$ by (the proof of) Lemma 15, since $f \in L_H^2$ by assumption and $t \mapsto K_t^* \delta_{\gamma_j(t)}$ belongs to L_H^2 by Lemma 12, as $(t \mapsto \delta_{\gamma_j(t)}) \in C_w^+$. By (A44), linearity of K_t^* and the identity (35) with ρ and $\hat{\rho}$ replaced by $\sum_{j=1}^{N} c_j \rho_{\gamma_j}$ and $\sum_{j=1}^{N} c_j \rho_{\gamma_j + \varepsilon\eta_j}$, respectively, one can compute that

$$\frac{\Phi(\Gamma + \varepsilon\Theta) - \Phi(\Gamma)}{\varepsilon}$$

$$= -\frac{1}{\varepsilon} \sum_{j=1}^{N} c_j \langle \rho_{\gamma_j + \varepsilon\eta_j} - \rho_{\gamma_j}, w \rangle + \frac{1}{2\varepsilon} \left\| \sum_{j=1}^{N} c_j K^* \left(\rho_{\gamma_j + \varepsilon\eta_j} - \rho_{\gamma_j} \right) \right\|_{L_H^2}^2 \,, \tag{A45}$$

for all $0 < \varepsilon < \varepsilon_0$. By proceeding in the same way as in (42), we have

$$\lim_{\varepsilon \to 0} -\frac{1}{\varepsilon} \sum_{j=1}^{N} c_j \langle \rho_{\gamma_j + \varepsilon\eta_j} - \rho_{\gamma_j}, w \rangle$$

$$= \lim_{\varepsilon \to 0} \sum_{j=1}^{N} c_j \frac{F(\gamma_j + \varepsilon\eta_j) - F(\gamma_j)}{\varepsilon} = \sum_{j=1}^{N} c_j \, D_{\gamma_j} F(\eta_j) \,,$$

where we also used the definition of F at (A8) and Proposition 16. We claim that the second term in (A45) is infinitesimal as $\varepsilon \to 0$. By (F2) and Cauchy–Schwarz's inequality, one has

$$\left\| \sum_{j=1}^{N} c_j K^* \left(\rho_{\gamma_j + \varepsilon\eta_j} - \rho_{\gamma_j} \right) \right\|_{L_H^2}^2$$

$$\leq NC^2 \sum_{j=1}^{N} c_j^2 \int_0^1 \left\| a_{\gamma_j + \varepsilon\eta_j} \delta_{\gamma_j(t) + \varepsilon\eta_j(t)} - a_{\gamma_j} \delta_{\gamma_j(t)} \right\|_{C^1(\overline{\Omega})^*}^2 \, \mathrm{d}t$$

where $C > 0$ is the constant in (F2). Fix $t \in [0, 1]$. Since $a_\gamma = 1/L(\gamma)$ (see (A8), (14)),

$$
\begin{aligned}
&\left\| a_{\gamma_j + \varepsilon\eta_j} \delta_{\gamma_j(t) + \varepsilon\eta_j(t)} - a_{\gamma_j} \delta_{\gamma_j(t)} \right\|_{C^1(\overline{\Omega})^*} \\
&= \sup_{\|\varphi\|_{C^1(\overline{\Omega})} \leq 1} \left| \frac{\varphi(\gamma_j(t) + \varepsilon\eta_j(t))}{L(\gamma_j + \varepsilon\eta_j)} - \frac{\varphi(\gamma_j(t))}{L(\gamma_j)} \right|.
\end{aligned}
\tag{A46}
$$

For a fixed $\varphi \in C^1(\overline{\Omega})$, define the map $\Psi_t \colon H_\Omega^1 \to \mathbb{R}$ as $\Psi_t(\gamma) := \varphi(\gamma(t))/L(\gamma)$. Since $\overline{\Omega}$ is bounded and $\varphi \in C^1(\overline{\Omega})$, one can check that $\gamma \in H_\Omega^1 \mapsto \varphi(\gamma(t)) \in \mathbb{R}$ is continuously Fréchet differentiable at each $\gamma \in H_\Omega^1$ with $\gamma([0, 1]) \subset \Omega$, with derivative given by $\eta \mapsto \nabla\varphi(\gamma(t)) \cdot \eta(t)$. Moreover, L is continuously differentiable by Proposition 16. Therefore, Ψ_t is continuously differentiable, given that $L \geq \alpha > 0$. By differentiation rules and triangle inequality, we also obtain the estimate

$$
\begin{aligned}
\left\| D_\gamma \Psi_t \right\|_{(H^1)^*} &= \sup_{\|\eta\|_{H^1} \leq 1} \left| \frac{\nabla\varphi(\gamma(t)) \cdot \eta(t)}{L(\gamma)} - \frac{D_\gamma L(\eta)}{L(\gamma)} \frac{\varphi(\gamma(t))}{L(\gamma)} \right| \\
&\leq \frac{\tilde{C} \|\varphi\|_{C^1(\overline{\Omega})}}{L(\gamma)} + \left\| \frac{D_\gamma L}{L(\gamma)} \right\|_{(H^1)^*} \frac{\|\varphi\|_{C^1(\overline{\Omega})}}{L(\gamma)} \leq \frac{\|\varphi\|_{C^1}}{\alpha} \left(\tilde{C} + \sqrt{\frac{\beta}{2\alpha}} \right)
\end{aligned}
$$

where $\tilde{C} > 0$ is the Sobolev embedding constant for $H^1((0, 1); \mathbb{R}^d) \hookrightarrow C([0, 1]; \mathbb{R}^d)$, and where in the last inequality we used that $L \geq \alpha$ and (A19). By the mean value theorem and (A46)

$$\left\| a_{\gamma_j + \varepsilon\eta_j} \delta_{\gamma_j(t) + \varepsilon\eta_j(t)} - a_{\gamma_j} \delta_{\gamma_j(t)} \right\|_{C^1(\overline{\Omega})^*} \leq \varepsilon\, C\, \|\eta_j\|_{H^1},$$

where C depends only on α, β and on $\overline{\Omega}$.

Putting together the above estimates shows that the second term in (A45) is infinitesimal as $\varepsilon \to 0$. This proves that the Gâteaux derivative of Φ at Γ in the direction Θ is given by (A42). From (A42) and Proposition 16, we also conclude that $\Gamma \mapsto D_\Gamma \Phi$ is continuous from $(H^1)^N$ into its dual, completing the proof. $\qquad\square$

A.6 Dynamic Undersampled Fourier Measurements

In this section, we detail a specific example of operators K_t^* and measurement spaces H_t satisfying the assumptions (H1)–(H3), (K1)–(K3) in Sect. 3.1. Such example is contained in [15, Section 5], and realizes, within our framework, a spatially undersampled Fourier transform with time-dependent mask. Let $\Omega \subset \mathbb{R}^2$ be a bounded open domain, and $\sigma_t \in \mathcal{M}^+(\mathbb{R}^2)$ be a family of measures such that

(M1) $\|\sigma_t\|_{\mathcal{M}(\mathbb{R}^2)} \leq C$ for a.e. $t \in (0, 1)$,

(M2) the map $t \mapsto \int_{\mathbb{R}^2} \varphi(s) \, d\sigma_t(s)$ is measurable for all $\varphi \in C_0(\mathbb{R}^2)$.

The measurement spaces are defined as the real Hilbert space $H_t := L^2_{\sigma_t}(\mathbb{R}^2; \mathbb{C})$, with scalar product given by $\langle f, g \rangle_{H_t} := \mathrm{Re}\left(\int_{\mathbb{R}^2} f(s)\overline{g(s)} \, d\sigma_t(s)\right)$, where Re denotes the real part in $\mathbb{C}$. For a measure $\rho \in \mathcal{M}(\overline{\Omega})$, we denote its Fourier transform by

$$\mathfrak{F}\rho(s) := \int_{\mathbb{R}^2} \exp\left(-2\pi i x \cdot s\right) d\rho(x), \tag{A47}$$

for all $s \in \mathbb{R}^2$, where ρ is extended to zero outside of $\overline{\Omega}$. Note that $\mathfrak{F}\rho \in C^\infty(\mathbb{R}^2; \mathbb{C})$. We then define $K^*_t \colon \mathcal{M}(\overline{\Omega}) \to H_t$ by setting $K^*_t \rho := \mathfrak{F}\rho$. In this way, K^*_t corresponds to the Fourier transform sampled according to the measure σ_t. As a consequence of [15, Lemma 5.4], we have that (H1)–(H3), (K1)–(K3) hold whenever (M1)–(M2) are satisfied. It an easy check that in this case also (F1)–(F3) from Sect. 5.1 are satisfied. Moreover, define the operators $\tilde{K}^*_t \colon C^{1,1}(\overline{\Omega})^* \to H_t$ as the dense extension of $\tilde{K}^*_t \rho := \mathfrak{F}_E \rho$, where for $\rho \in \mathcal{M}(\overline{\Omega})$ we set

$$\mathfrak{F}_E \rho(s) := \int_{\mathbb{R}^2} \exp\left(-2\pi i x \cdot s\right) \xi_E(x) \, d\rho(x), \tag{A48}$$

for all $s \in \mathbb{R}^2$, and $\xi_E \colon \overline{\Omega} \to [0, 1]$ is a cut-off with respect to a closed convex set $E \Subset \Omega$ satisfying (A11). Then, arguing as in Remark 11, we can show that $\tilde{K}^*_t$ satisfies (F1)–(F4).

A.6.1 Discrete Sampling

As a particular case of the above setting, we sample the Fourier transform on a finite collection of time-dependent frequencies. Specifically, fix $T \in \mathbb{N}$ and consider a time-grid $0 \le t_0 < t_1 < \ldots < t_T \le 1$. For each time t_i, we sample a given collection of frequencies $S_{i,1}, \ldots, S_{i,n_i} \in \mathbb{R}^2$, for some $n_i \in \mathbb{N}$. In order to incorporate this in the above setting, define a partition $A_0, \ldots, A_T$ of $[0, 1]$, in a way that $t_i \in A_i$ for all $i = 0, \ldots, T$. Then, define the scalar measure

$$\sigma_t := \sum_{i=0}^{T} \sum_{k=1}^{n_i} \frac{1}{n_i} \delta_{S_{i,k}} \chi_{A_i}(t).$$

It is immediate to check that σ_t satisfies (M1)–(M2), so that the prescribed sampling falls within the above framework. In this case, the sampling space $H_t := L^2_{\sigma_t}(\mathbb{R}^2; \mathbb{C})$ is isomorphic to $\mathbb{C}^{n_i}$, whenever $t \in A_i$. By defining the Fourier-type kernels $\psi_t \colon \mathbb{R}^2 \to \mathbb{C}^{n_i}$,

$$\psi_t(x) := \left(\exp\left(-2\pi i x \cdot S_{i,k}\right)\right)_{k=1}^{n_i} \in \mathbb{C}^{n_i},$$

for any $t \in A_i$, we see that the operator $K_t^* \colon \mathcal{M}(\overline{\Omega}) \to H_t$ defined by $K_t^* \rho := \mathfrak{F}\rho$ in (A47) and its pre-adjoint $K_t \colon H_t \to C(\overline{\Omega})$ can be represented as

$$K_t^*(\rho) = \int_{\mathbb{R}^2} \psi_t(x) \, \mathrm{d}\rho(x) \,, \qquad K_t(h) = \left(x \mapsto \langle \psi_t(x), h \rangle_{H_t} \right) , \qquad (A49)$$

for all $\rho \in \mathcal{M}(\overline{\Omega})$, $h \in H_t$, where the first integral is computed component-wise. Similarly, the operator $K_t^* \rho := \mathfrak{F}_E \rho$ in (A48) and its pre-adjoint K_t are represented by (A49) with ψ_t replaced by the cut-off kernel

$$\psi_t(x) := \left(\exp\left(-2\pi i x \cdot S_{i,k}\right) \xi_E(x) \right)_{k=1}^{n_i} \in \mathbb{C}^{n_i} \,, \qquad (A50)$$

for every $t \in A_i$.

References

1. Alberti, G.S., Ammari, H., Romero, F., Wintz, T.: Dynamic spike superresolution and applications to ultrafast ultrasound imaging. SIAM Journal on Imaging Sciences **12**(3), 1501–1527 (2019)
2. Aliprantis, C.D., Border, K.: Infinite Dimensional Analysis. Springer, Berlin Heidelberg (2006)
3. Ambrosio, L., Fusco, N., Pallara, D.: Functions of Bounded Variation and Free Discontinuity Problems. Oxford University Press, Oxford (2000)
4. Ambrosio, L., Gigli, N., Savaré, G.: Gradient Flows: In Metric Spaces and in the Space of Probability Measures. Birkhäuser, Basel (2005)
5. Andersen, M.S., Dahl, J., Vandenberghe, L.: CVXOPT: A Python package for convex optimization, version 1.1.5. Available at: https://cvxopt.org/
6. Andersen, M.S., Dahl, J., Liu, Z., Vandenberghe, L.: Interior-point methods for large-scale cone programming. In: Sra, S., Nowozin, S., Wright, S.J. (eds.) Optimization for Machine Learning, pp. 55–83. MIT Press, Cambridge, Massachusetts (2012)
7. Bach, F.: Duality between subgradient and conditional gradient methods. SIAM Journal on Optimization **25**(1), 115–129 (2015)
8. Benamou, J.-D., Brenier, Y.: A computational fluid mechanics solution to the Monge-Kantorovich mass transfer problem. Numerische Mathematik **84**(3), 375–393 (2000)
9. Bonnet, S., Koenig, A., Roux, S., Hugonnard, P., Guillemaud, R., Grangeat, P.: Dynamic X-ray computed tomography. Proceedings of the IEEE **91**(10), 1574–1587 (2003)
10. Boyd, N., Schiebinger, G., Recht, B.: The alternating descent conditional gradient method for sparse inverse problems. SIAM Journal on Optimization **27**(2), 616–639 (2017)
11. Boyer, C., Chambolle, A., Castro, Y.D., Duval, V., de Gournay, F., Weiss, P.: On representer theorems and convex regularization. SIAM Journal on Optimization **29**(2), 1260–1281 (2019)
12. Bredies, K., Carioni, M., Fanzon, S.: A superposition principle for the inhomogeneous continuity equation with Hellinger-Kantorovich-regular coefficients. arXiv e-prints (2020) arXiv:2007.06964
13. Bredies, K., Carioni, M., Fanzon, S., Walter, D.: Linear convergence of accelerated generalized conditional gradient methods. arXiv e-prints (2021) arXiv:2110.06756
14. Bredies, K., Carioni, M.: Sparsity of solutions for variational inverse problems with finite-dimensional data. Calculus of Variations and Partial Differential Equations **59**(1), 14 (2020)
15. Bredies, K., Fanzon, S.: An optimal transport approach for solving dynamic inverse problems in spaces of measures. ESAIM: Mathematical Modelling and Numerical Analysis **54**(6), 2351–2382 (2020)
16. Bredies, K., Lorenz, D.: Mathematical Image Processing. Birkhäuser, Basel (2018)
17. Bredies, K., Lorenz, D.A.: Iterated hard shrinkage for minimization problems with sparsity constraints. SIAM Journal on Scientific Computing **30**(2), 657–683 (2008)
18. Bredies, K., Pikkarainen, H.K.: Inverse problems in spaces of measures. ESAIM: Control, Optimisation and Calculus of Variations **19**(1), 190–218 (2013)
19. Bredies, K., Lorenz, D.A., Maass, P.: A generalized conditional gradient method and its connection to an iterative shrinkage method. Computational Optimization and Applications **42**(2), 173–193 (2009)

20. Bredies, K., Carioni, M., Fanzon, S., Romero, F.: On the extremal points of the ball of the Benamou–Brenier energy. Bulletin of the London Mathematical Society **53**(5), 1436–1452 (2021)
21. Burger, M., Dirks, H., Frerking, L., Hauptmann, A., Helin, T., Siltanen, S.: A variational reconstruction method for undersampled dynamic X-ray tomography based on physical motion models. Inverse Problems **33**(12), 124008 (2017)
22. Candès, E.J., Romberg, J.K., Tao, T.: Stable signal recovery from incomplete and inaccurate measurements. Communications on Pure and Applied Mathematics **59**(8), 1207–1223 (2006)
23. Candès, E.J., Fernandez-Granda, C.: Towards a mathematical theory of super-resolution. Communications on Pure and Applied Mathematics **67**(6), 906–956 (2014)
24. Chizat, L., Peyré, G., Schmitzer, B., Vialard, F.-X.: An interpolating distance between optimal transport and Fisher–Rao metrics. Foundations of Computational Mathematics **18**(1), 1–44 (2018)
25. Clarkson, K.L.: Coresets, sparse greedy approximation, and the Frank-Wolfe algorithm. ACM Transactions on Algorithms **6**(4), 63 (2010)
26. Combettes, P.L., Wajs, V.R.: Signal recovery by proximal forward-backward splitting. Multiscale Modeling & Simulation **4**(4), 1168–1200 (2005)
27. Dacorogna, B.: Direct Methods in the Calculus of Variations, 2nd edn. Applied Mathematical Sciences, vol. 78. Springer, New York (2008)
28. Daubechies, I., Defrise, M., De Mol, C.: An iterative thresholding algorithm for linear inverse problems with a sparsity constraint. Communications on Pure and Applied Mathematics **57**(11), 1413–1457 (2004)
29. de Castro, Y., Gamboa, F.: Exact reconstruction using Beurling minimal extrapolation. Journal of Mathematical Analysis and Applications **395**(1), 336–354 (2012)
30. Denoyelle, Q., Duval, V., Peyré, G., Soubies, E.: The sliding Frank–Wolfe algorithm and its application to super-resolution microscopy. Inverse Problems **36**(1), 014001 (2019)
31. Diestel, J., Uhl, J.J.: Vector Measures. American Mathematical Society, Providence (1977)
32. Ding, Q., Burger, M., Zhang, X.: Dynamic SPECT reconstruction with temporal edge correlation. Inverse Problems **34**(1), 014005 (2017)
33. Dunn, J.C.: Rates of convergence for conditional gradient algorithms near singular and nonsingular extremals. SIAM Journal on Control and Optimization **17**(2), 187–211 (1979)
34. Duval, V.: An epigraphical approach to the representer theorem. Journal of Convex Analysis **28**(3), 819–836 (2021)
35. Epstein, C.L.: Introduction to the Mathematics of Medical Imaging, 2nd edn. Society for Industrial and Applied Mathematics, Philadelphia (2007)
36. Evans, L.C., Gariepy, R.F.: Measure Theory and Fine Properties of Functions. CRC Press, Boca Raton, Florida (2015)
37. Fanzon, S., Palombaro, M., Ponsiglione, M.: Derivation of linearised polycrystals from a two-dimensional system of edge dislocations. SIAM Journal on Mathematical Analysis **51**(5), 3956–3981 (2019)
38. Fanzon, S., Ponsiglione, M., Scala, R.: Uniform distribution of dislocations in Peierls–Nabarro models for semi-coherent interfaces. Calculus of Variations and Partial Differential Equations **59**(4), 141 (2020)
39. Flinth, A., Weiss, P.: Exact solutions of infinite dimensional total-variation regularized problems. Information and Inference: A Journal of the IMA **8**(3), 407–443 (2018)
40. Flinth, A., de Gournay, F., Weiss, P.: On the linear convergence rates of exchange and continuous methods for total variation minimization. Mathematical Programming **190**(1), 221–257 (2021)
41. Frank, M., Wolfe, P.: An algorithm for quadratic programming. Naval Research Logistics Quarterly **3**(1-2), 95–110 (1956)
42. Holler, M., Kunisch, K.: On infimal convolution of TV-type functionals and applications to video and image reconstruction. SIAM Journal on Imaging Sciences **7**(4), 2258–2300 (2014)
43. Jaggi, M.: Revisiting Frank-Wolfe: Projection-free sparse convex optimization. In: Dasgupta, S., McAllester, D. (eds.) Proceedings of the 30th International Conference on Machine Learning, Atlanta, Georgia, USA, pp. 427–435 (2013)
44. Kantorovitch, L.: On the translocation of masses. Comptes Rendus (Doklady) de l'Académie des Sciences de l'URSS **37**, 199–201 (1942)
45. Kondratyev, S., Monsaingeon, L., Vorotnikov, D.: A new optimal transport distance on the space of finite Radon measures. Advances in Differential Equations **21**(11/12), 1117–1164 (2016)
46. Lauteri, G., Luckhaus, S.: An energy estimate for dislocation configurations and the emergence of Cosserat-type structures in metal plasticity. arXiv e-prints (2016) arXiv:1608.06155

47. Liero, M., Mielke, A., Savaré, G.: Optimal Entropy-Transport problems and a new Hellinger-Kantorovich distance between positive measures. Inventiones mathematicae **211**, 969–1117 (2018)
48. Lingala, S.G., Hu, Y., DiBella, E., Jacob, M.: Accelerated dynamic MRI exploiting sparsity and low-rank structure: *k-t* SLR. IEEE Transactions on Medical Imaging **30**(5), 1042–1054 (2011)
49. Neitzel, I., Pieper, K., Vexler, B., Walter, D.: A sparse control approach to optimal sensor placement in PDE-constrained parameter estimation problems. Numerische Mathematik **143**(4), 943–984 (2019)
50. Otazo, R., Candès, E., Sodickson, D.K.: Low-rank plus sparse matrix decomposition for accelerated dynamic MRI with separation of background and dynamic components. Magnetic Resonance in Medicine **73**(3), 1125–1136 (2015)
51. Pieper, K., Walter, D.: Linear convergence of accelerated conditional gradient algorithms in spaces of measures. ESAIM: Control, Optimisation and Calculus of Variations **27**, 38 (2021)
52. Pieper, K., Tang, B.Q., Trautmann, P., Walter, D.: Inverse point source location with the Helmholtz equation on a bounded domain. Computational Optimization and Applications **77**(1), 213–249 (2020)
53. Read, W.T., Shockley, W.: Dislocation models of crystal grain boundaries. Physical Review **78**(3), 275–289 (1950)
54. Santambrogio, F.: Optimal Transport for Applied Mathematicians. Birkhäuser, Basel (2015)
55. Saumier, L.-P., Khouider, B., Agueh, M.: Optimal transport for particle image velocimetry: real data and postprocessing algorithms. SIAM Journal on Applied Mathematics **75**(6), 2495–2514 (2015)
56. Schloegl, M., Holler, M., Schwarzl, A., Bredies, K., Stollberger, R.: Infimal convolution of total generalized variation functionals for dynamic MRI. Magnetic Resonance in Medicine **78**(1), 142–155 (2017)
57. Schmitt, U., Louis, A.K.: Efficient algorithms for the regularization of dynamic inverse problems: I. Theory. Inverse Problems **18**(3), 645 (2002)
58. Schmitt, U., Louis, A.K., Wolters, C., Vauhkonen, M.: Efficient algorithms for the regularization of dynamic inverse problems: II. Applications. Inverse Problems **18**(3), 659–676 (2002)
59. Schmitzer, B., Wirth, B.: Dynamic models of Wasserstein-1-type unbalanced transport. ESAIM: Control, Optimisation and Calculus of Variations **25**, 23 (2019)
60. Schmitzer, B., Schafers, K.P., Wirth, B.: Dynamic cell imaging in PET with optimal transport regularization. IEEE Transactions on Medical Imaging **39**(5), 1626–1635 (2020)
61. Schuster, T., Hahn, B., Burger, M.: Dynamic inverse problems: modelling, regularization, numerics. Inverse Problems **34**(4), 040301 (2018)
62. Tibshirani, R.: Regression shrinkage and selection via the Lasso. Journal of the Royal Statistical Society. Series B (Methodological), 267–288 (1996)
63. Tseng, P.: Convergence of a block coordinate descent method for nondifferentiable minimization. Journal of Optimization Theory and Applications **109**(3), 475–494 (2001)
64. Unser, M.: A unifying representer theorem for inverse problems and machine learning. Foundations of Computational Mathematics **21**(4), 941–960 (2021)
65. Unser, M., Fageot, J.: Native Banach spaces for splines and variational inverse problems. arXiv e-prints (2019) arXiv:1904.10818
66. Unser, M., Fageot, J., Ward, J.P.: Splines are universal solutions of linear inverse problems with generalized TV regularization. SIAM Review **59**(4), 769–793 (2017)
67. Weickert, J., Schnörr, C.: Variational optic flow computation with a spatio-temporal smoothness constraint. Journal of Mathematical Imaging and Vision **14**(3), 245–255 (2001)

Springer