A generalized conditional gradient method for dynamic inverse problems with optimal transport regularization

Kristian Bredies Marcello Carioni Silvio Fanzon
Francisco Romero Daniel Walter

Institute of Mathematics and Scientific Computing
University of Graz

Workshop on Recent Advances in Iterative Reconstruction
University College London

May 23, 2023

Motivation

Motion-aware tomographic reconstruction

Motion on sub-acquisition time scales $\rightsquigarrow$ artefacts in reconstructed images

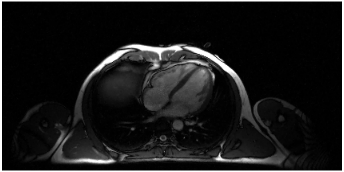
- Imaging of the lung or heart (motion cannot be suppressed)
- High-resolution imaging (sub-millimeter motion poses problems)

Motivation

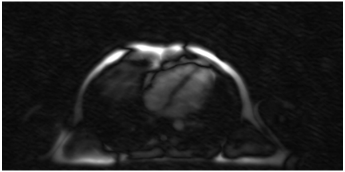
Motion-aware tomographic reconstruction

Motion on sub-acquisition time scales $\rightsquigarrow$ artefacts in reconstructed images

- Imaging of the lung or heart (motion cannot be suppressed)
- High-resolution imaging (sub-millimeter motion poses problems)



Reference data



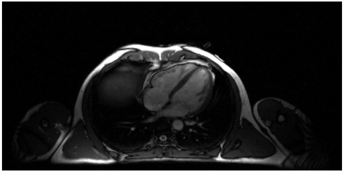
Unregularized reconstruction

Motivation

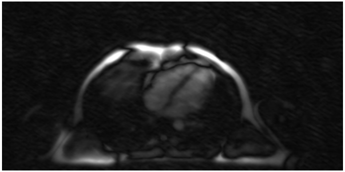
Motion-aware tomographic reconstruction

Motion on sub-acquisition time scales $\rightsquigarrow$ artefacts in reconstructed images

- Imaging of the lung or heart (motion cannot be suppressed)
- High-resolution imaging (sub-millimeter motion poses problems)



Reference data

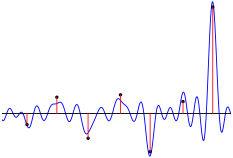


Unregularized reconstruction

$\rightsquigarrow$ **Optimal-transport regularization for for dynamic inverse problems**

Motivation: Sparse recovery

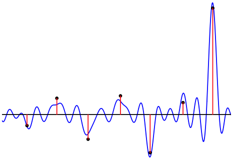
Well-studied problem: Superresolution



- Solve $\mathfrak{F}u = f$ on Σ
- $\mathfrak{F}$ Fourier transform, $\Sigma \subset \mathbb{R}^d$ finite set
- Sparsity assumption: $u = \sum_{i=1}^{N} c_i \delta_{x_i}$

Motivation: Sparse recovery

Well-studied problem: Superresolution



- Solve $\mathfrak{F}u = f \quad \text{on} \quad \Sigma$
- $\mathfrak{F}$ Fourier transform, $\Sigma \subset \mathbb{R}^d$ finite set
- Sparsity assumption: $u = \sum_{i=1}^{N} c_i \delta_{x_i}$

Radon-norm regularization

- Solve variational problem in space of Radon measures

$$\min_{u \in \mathcal{M}(\Omega)} \|u\|_{\mathcal{M}} \quad \text{subject to} \quad \mathfrak{F}u = f \quad \text{on} \quad \Sigma$$

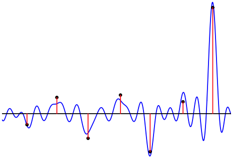
- Relaxed/regularized version (noisy data)

$$\min_{u \in \mathcal{M}(\Omega)} \tfrac{1}{2}\|\mathfrak{F}u - f\|_{\Sigma,2}^2 + \alpha\|u\|_{\mathcal{M}}$$

[Candès/Fernandez-Granda '13] and many more

Motivation: Sparse recovery

Well-studied problem: Superresolution



- Solve $\mathfrak{F}u = f$ on Σ
- $\mathfrak{F}$ Fourier transform, $\Sigma \subset \mathbb{R}^d$ finite set
- Sparsity assumption: $u = \sum_{i=1}^{N} c_i \delta_{x_i}$

Radon-norm regularization

- Solve variational problem in space of Radon measures

$$\min_{u \in \mathcal{M}(\Omega)} \|u\|_{\mathcal{M}} \quad \text{subject to} \quad \mathfrak{F}u = f \quad \text{on} \quad \Sigma$$

- Relaxed/regularized version (noisy data)

$$\min_{u \in \mathcal{M}(\Omega)} \frac{1}{2}\|\mathfrak{F}u - f\|_{\Sigma,2}^2 + \alpha\|u\|_{\mathcal{M}}$$

[Candès/Fernandez-Granda '13] and many more

$\rightsquigarrow$ **Generalization and (fast) recovery algorithms**

Outline

Outline

Static optimal transport

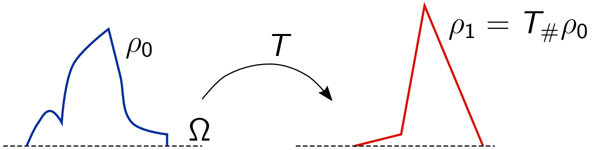
Situation

- $\Omega \subset \mathbf{R}^d$ bounded domain, $\rho_0, \rho_1 \in \mathcal{P}(\Omega)$
- $T : \Omega \to \Omega$ measurable, $\rho_1 = T_{\#}\rho_0$

Static optimal transport

Situation

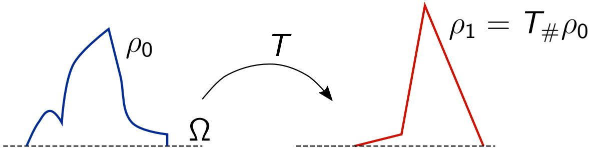
- $\Omega \subset \mathbf{R}^d$ bounded domain, $\rho_0, \rho_1 \in \mathcal{P}(\Omega)$
- $T : \Omega \to \Omega$ measurable, $\rho_1 = T_{\#}\rho_0$



Static optimal transport

Situation

- $\Omega \subset \mathbf{R}^d$ bounded domain, $\rho_0, \rho_1 \in \mathcal{P}(\Omega)$
- $T : \Omega \to \Omega$ measurable, $\rho_1 = T_\# \rho_0$



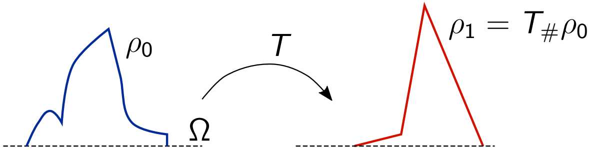
Goal

- Move ρ_0 to ρ_1 in an optimal way
- Cost of moving mass from x to y: $c(x, y) = \frac{1}{2}|x - y|^2$

Static optimal transport

Situation

- $\Omega \subset \mathbf{R}^d$ bounded domain, $\rho_0, \rho_1 \in \mathcal{P}(\Omega)$
- $T : \Omega \to \Omega$ measurable, $\rho_1 = T_\# \rho_0$



Goal

- Move ρ_0 to ρ_1 in an optimal way
- Cost of moving mass from x to y: $c(x, y) = \frac{1}{2}|x - y|^2$

Optimal transport

Solve $\displaystyle \min_{T:\Omega\to\Omega} \frac{1}{2} \int_\Omega |T(x) - x|^2 \, d\rho_0(x)$ subject to $T_\# \rho_0 = \rho_1$

Dynamic optimal transport

Idea

Introduce a time variable $t \in [0,1]$ and consider evolution of ρ_t

- Time-dependent probability measures

$$t \mapsto \rho_t \in \mathcal{P}(\Omega) \quad \text{for} \quad t \in [0,1]$$

- Velocity field advecting ρ_t

$$v_t \colon [0,1] \times \Omega \to \mathbf{R}^d$$

- (ρ_t, v_t) solves the continuity equation with initial conditions

$$\begin{cases} \partial_t \rho_t + \operatorname{div}(\rho_t v_t) = 0 \\ \text{Initial data } \rho_0, \text{ final data } \rho_1 \end{cases} \qquad \text{(CE-IC)}$$

Dynamic optimal transport

Idea

Introduce a time variable $t \in [0, 1]$ and consider evolution of ρ_t

- Time-dependent probability measures

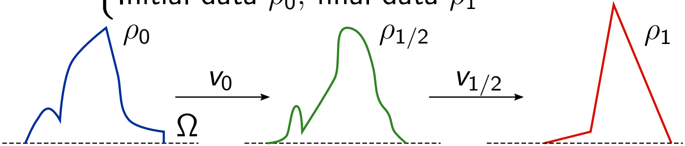
$$t \mapsto \rho_t \in \mathcal{P}(\Omega) \quad \text{for} \quad t \in [0, 1]$$

- Velocity field advecting ρ_t

$$v_t \colon [0, 1] \times \Omega \to \mathbf{R}^d$$

- (ρ_t, v_t) solves the continuity equation with initial conditions

$$\begin{cases} \partial_t \rho_t + \operatorname{div}(\rho_t v_t) = 0 \\ \text{Initial data } \rho_0, \text{ final data } \rho_1 \end{cases} \qquad \text{(CE-IC)}$$



Dynamic optimal transport

Theorem *[Benamou/Brenier '00]*

$$\min_{\substack{(\rho_t, v_t) \\ \text{solving (CE-IC)}}} \int_0^1 \int_\Omega |v_t(x)|^2 \rho_t(x) \, dx \, dt = \min_{\substack{T : \Omega \to \Omega \\ T_\# \rho_0 = \rho_1}} \int_\Omega |T(x) - x|^2 \rho_0(x) \, dx$$

Dynamic optimal transport

Theorem *[Benamou/Brenier '00]*

$$\min_{\substack{(\rho_t, v_t) \\ \text{solving (CE-IC)}}} \int_0^1 \int_\Omega |v_t(x)|^2 \rho_t(x) dx\, dt = \min_{\substack{T:\, \Omega \to \Omega \\ T_\# \rho_0 = \rho_1}} \int_\Omega |T(x) - x|^2 \rho_0(x)\, dx$$

Advantages of the dynamic formulation

- By introducing $m_t = \rho_t v_t$, we have the convex energy

$$\int_0^1 \int_\Omega |v_t(x)|^2 \rho_t(x)\, dx\, dt = \int_0^1 \int_\Omega \frac{|m_t(x)|^2}{\rho_t(x)}\, dx\, dt$$

- The continuity equation becomes linear

$$\partial_t \rho_t + \operatorname{div} m_t = 0$$

- Full trajectory ρ_t is known and v_t can be recovered from m_t

Optimal transport energy

Definition

- Let $X = (0,1) \times \overline{\Omega}$
- For $(\rho, m) \in \mathcal{M}(X) \times \mathcal{M}(X)^d$, let

$$B(\rho, m) = \int_X \Psi\left(\frac{d\rho}{d\lambda}, \frac{dm}{d\lambda}\right)\, d\lambda$$

where $\lambda \in \mathcal{M}^+(X)$ is such that $\rho, m \ll \lambda$ and

$$\Psi(t, x) = \frac{|x|^2}{2t} \quad \text{if} \quad t > 0, \quad \Psi = \infty \text{ else}$$

Optimal transport energy

Definition

- Let $X = (0,1) \times \overline{\Omega}$
- For $(\rho, m) \in \mathcal{M}(X) \times \mathcal{M}(X)^d$, let

$$B(\rho, m) = \int_X \Psi\left(\frac{d\rho}{d\lambda}, \frac{dm}{d\lambda}\right) \, d\lambda$$

where $\lambda \in \mathcal{M}^+(X)$ is such that $\rho, m \ll \lambda$ and

$$\Psi(t, x) = \frac{|x|^2}{2t} \quad \text{if} \quad t > 0, \quad \Psi = \infty \text{ else}$$

- Generalizes

$$\frac{1}{2} \int_0^1 \int_{\overline{\Omega}} \frac{|m|^2}{\rho} \, \mathrm{d}x \, \mathrm{d}t$$

for functions $\rho : X \to [0, \infty)$, $m : X \to \mathbb{R}^d$, $\mu : X \to \mathbb{R}$ to arbitrary Radon measures

Optimal transport energy

Proposition *[B./Fanzon '20]*

- The functional B is proper, convex, weak* lower semi-continuous and 1-homogeneous

Optimal transport energy

Proposition *[B./Fanzon '20]*

- The functional B is proper, convex, weak* lower semi-continuous and 1-homogeneous
- If $B(\rho, m) < \infty$ and $\partial_t \rho + \operatorname{div} m = 0$, then
  - $\rho = dt \otimes \rho_t$ for a weak*-continuous curve $t \mapsto \rho_t \in \mathcal{M}^+(\overline{\Omega})$
  - $m = \rho v_t$ for some velocity field $v_t \colon (0, 1) \times \overline{\Omega} \to \mathbf{R}^d$

Optimal transport energy

Proposition *[B./Fanzon '20]*

- The functional B is proper, convex, weak* lower semi-continuous and 1-homogeneous
- If $B(\rho, m) < \infty$ and $\partial_t \rho + \operatorname{div} m = 0$, then
  - $\rho = dt \otimes \rho_t$ for a weak*-continuous curve $t \mapsto \rho_t \in \mathcal{M}^+(\overline{\Omega})$
  - $m = \rho v_t$ for some velocity field $v_t \colon (0,1) \times \overline{\Omega} \to \mathbf{R}^d$

$$B(\rho, m) = \frac{1}{2} \int_0^1 \int_{\overline{\Omega}} |v_t(x)|^2 \, d\rho_t(x) \, dt$$

Optimal transport energy

Proposition *[B./Fanzon '20]*

- The functional B is proper, convex, weak* lower semi-continuous and 1-homogeneous
- If $B(\rho, m) < \infty$ and $\partial_t \rho + \operatorname{div} m = 0$, then
  - $\rho = dt \otimes \rho_t$ for a weak*-continuous curve $t \mapsto \rho_t \in \mathcal{M}^+(\overline{\Omega})$
  - $m = \rho v_t$ for some velocity field $v_t \colon (0,1) \times \overline{\Omega} \to \mathbf{R}^d$

$$B(\rho, m) = \frac{1}{2} \int_0^1 \int_{\overline{\Omega}} |v_t(x)|^2 \, d\rho_t(x) \, dt$$

$\rightsquigarrow$ **Use as an energy for Tikhonov regularization**

Dynamic inverse problem

General setting

- $\Omega \subset \mathbf{R}^d$ bounded open domain, $d \geq 1$
- For $t \in [0, 1]$ assume given
  - H_t Hilbert space (measurement space)
  - $K_t^* : \mathcal{M}(\overline{\Omega}) \to H_t$ linear continuous operator (forward operator)

$\rightsquigarrow$ time dependence allows for spatial undersampling

UNI GRAZ

Dynamic inverse problem

General setting

- $\Omega \subset \mathbf{R}^d$ bounded open domain, $d \geq 1$
- For $t \in [0, 1]$ assume given
  - H_t Hilbert space (measurement space)
  - $K_t^* : \mathcal{M}(\overline{\Omega}) \to H_t$ linear continuous operator (forward operator)

$\rightsquigarrow$ time dependence allows for spatial undersampling

Inverse problem

Given some data $\{f_t\}_{t\in[0,1]}$ with $f_t \in H_t$, find a curve of measures $t \mapsto \rho_t \in \mathcal{M}(\overline{\Omega})$ such that

$$K_t^* \rho_t = f_t \qquad \text{for a.e.} \quad t \in [0, 1]\,. \tag{P}$$

Tikhonov regularization

Inverse problem

Solve $K_t^* \rho_t = f_t$ in H_t for a.e. $t \in [0, 1]$

Tikhonov regularized problem

$$\min_{(\rho,m)\in\mathcal{M}(X)^{d+1}} \underbrace{\frac{1}{2}\int_0^1 \|K_t^*\rho_t - f_t\|_{H_t}^2 \, dt}_{\text{fidelty term}} + \underbrace{\alpha B(\rho, m)}_{\text{optimal-transport term}} + \underbrace{\beta\|\rho\|_{\mathcal{M}}}_{\text{total-variation term}}$$

subject to $\partial_t \rho + \operatorname{div} m = 0$ (CE)

- Regularization parameters $\alpha > 0$, $\beta > 0$

Tikhonov regularization

Inverse problem

Solve $\quad K_t^* \rho_t = f_t \quad \text{in} \quad H_t \quad \text{for a.e. } t \in [0,1]$

Tikhonov regularized problem

$$\min_{(\rho,m)\in\mathcal{M}(X)^{d+1}} \underbrace{\frac{1}{2}\int_0^1 \|K_t^*\rho_t - f_t\|_{H_t}^2 \, dt}_{\text{fidelty term}} + \underbrace{\alpha B(\rho, m)}_{\text{optimal-transport term}} + \underbrace{\beta \|\rho\|_{\mathcal{M}}}_{\text{total-variation term}}$$

$$\text{subject to} \quad \partial_t \rho + \operatorname{div} m = 0 \qquad \text{(CE)}$$

- Regularization parameters $\alpha > 0$, $\beta > 0$

- (CE) ensures $\rho = \, dt \otimes \rho_t$ and $m \ll \rho$
- $m = v_t \rho_t \rightsquigarrow$ motion field

Dynamic data spaces

Assumption (H)

The spaces H_t vary in a "measurable" way w.r.t $t \in [0, 1]$

- $\exists$ Banach space D and $i_t \colon D \to H_t$ linear continuous
- $i_t(D) \subset H_t$ dense, $\sup_t \|i_t\| \leq C$
- for each $\varphi, \psi \in D$ the map $t \mapsto \langle i_t\varphi, i_t\psi\rangle_{H_t}$ is Lebesgue measurable

Dynamic data spaces

Assumption (H)

The spaces H_t vary in a "measurable" way w.r.t $t \in [0,1]$

- $\exists$ Banach space D and $i_t \colon D \to H_t$ linear continuous
- $i_t(D) \subset H_t$ dense, $\sup_t \|i_t\| \leq C$
- for each $\varphi, \psi \in D$ the map $t \mapsto \langle i_t\varphi, i_t\psi\rangle_{H_t}$ is Lebesgue measurable

Dynamic data spaces

Assumption (H)

The spaces H_t vary in a "measurable" way w.r.t $t \in [0,1]$

- $\exists$ Banach space D and $i_t \colon D \to H_t$ linear continuous
- $i_t(D) \subset H_t$ dense, $\sup_t \|i_t\| \leq C$
- for each $\varphi, \psi \in D$ the map $t \mapsto \langle i_t\varphi, i_t\psi\rangle_{H_t}$ is Lebesgue measurable

Dynamic data spaces

Assumption (H)

The spaces H_t vary in a "measurable" way w.r.t $t \in [0,1]$

- $\exists$ Banach space D and $i_t\colon D \to H_t$ linear continuous
- $i_t(D) \subset H_t$ dense, $\sup_t \|i_t\| \leq C$
- for each $\varphi, \psi \in D$ the map $t \mapsto \langle i_t\varphi, i_t\psi\rangle_{H_t}$ is Lebesgue measurable

Definition $$L^2_H = \Big\{ f\colon [0,1] \to \cup_t H_t \;\Big|\; f_t \in H_t, \; f \text{ strongly measurable}, \int_0^1 \|f_t\|^2_{H_t} \,\mathrm{d}t < \infty \Big\}$$

Dynamic data spaces

Assumption (H)
The spaces H_t vary in a "measurable" way w.r.t $t \in [0,1]$

- $\exists$ Banach space D and $i_t \colon D \to H_t$ linear continuous
- $i_t(D) \subset H_t$ dense, $\sup_t \|i_t\| \leq C$
- for each $\varphi, \psi \in D$ the map $t \mapsto \langle i_t\varphi, i_t\psi\rangle_{H_t}$ is Lebesgue measurable

Definition $$L^2_H = \Big\{ f \colon [0,1] \to \cup_t H_t \;\Big|\; f_t \in H_t, \; f \text{ strongly measurable}, \int_0^1 \|f_t\|^2_{H_t} \,\mathrm{d}t < \infty \Big\}$$

Theorem *[B./Fanzon '20]*
The space L^2_H is a Hilbert space.

Forward operators

Assumption (K)

The operators $K_t^*\colon \mathcal{M}(\overline{\Omega}) \to H_t$ satisfy

- K_t^* linear continuous and weak*-to-weak continuous
- $\sup_t \|K_t^*\| \leq C$
- for $\rho \in \mathcal{M}(\overline{\Omega})$ the map $t \mapsto K_t^* \rho$ is strongly measurable

Forward operators

Assumption (K)

The operators $K_t^*\colon \mathcal{M}(\overline{\Omega}) \to H_t$ satisfy

- K_t^* linear continuous and weak*-to-weak continuous
- $\sup_t \|K_t^*\| \leq C$
- for $\rho \in \mathcal{M}(\overline{\Omega})$ the map $t \mapsto K_t^* \rho$ is strongly measurable

Forward operators

Assumption (K)

The operators $K_t^*\colon \mathcal{M}(\overline{\Omega}) \to H_t$ satisfy

- K_t^* linear continuous and weak*-to-weak continuous
- $\sup_t \|K_t^*\| \leq C$
- for $\rho \in \mathcal{M}(\overline{\Omega})$ the map $t \mapsto K_t^* \rho$ is strongly measurable

Forward operators

Assumption (K)

The operators $K_t^* \colon \mathcal{M}(\overline{\Omega}) \to H_t$ satisfy

- K_t^* linear continuous and weak*-to-weak continuous
- $\sup_t \|K_t^*\| \leq C$
- for $\rho \in \mathcal{M}(\overline{\Omega})$ the map $t \mapsto K_t^* \rho$ is strongly measurable

Proposition *[B./Fanzon '20]*

$t \mapsto \rho_t \in \mathcal{M}(\overline{\Omega})$ weak* continuous $\Rightarrow$ $t \mapsto K_t^* \rho_t$ is in L_H^2

Forward operators

Assumption (K)

The operators $K_t^*\colon \mathcal{M}(\overline{\Omega}) \to H_t$ satisfy

- K_t^* linear continuous and weak*-to-weak continuous
- $\sup_t \|K_t^*\| \leq C$
- for $\rho \in \mathcal{M}(\overline{\Omega})$ the map $t \mapsto K_t^* \rho$ is strongly measurable

Proposition *[B./Fanzon '20]*

$t \mapsto \rho_t \in \mathcal{M}(\overline{\Omega})$ weak* continuous $\Rightarrow$ $t \mapsto K_t^* \rho_t$ is in L_H^2

Tikhonov functional

Let $f \in L_H^2$ given data. For $(\rho, m) \in \mathcal{M}(X)^{d+1}$ set

$$T_{\alpha,\beta}(\rho, m) = \frac{1}{2} \int_0^1 \|K_t^* \rho_t - f_t\|_{H_t}^2 \,\mathrm{d}t + \alpha B(\rho, m) + \beta \|\rho\|_{\mathcal{M}}$$

if $\partial_t \rho + \operatorname{div} m = 0$, and $T_{\alpha,\beta}(\rho, m) = \infty$ else

Existence and stability

Assume **(H)**-**(K)**.

Theorem *[B./Fanzon '20]*

$$\min_{(\rho,m)\in\mathcal{M}(X)^{d+1}} T_{\alpha,\beta}(\rho, m) \qquad \text{(Tikh)}$$

admits a solution for $f \in L^2_H$.

- If K_t^* is injective for a.e. t, then the solution is unique.

Existence and stability

Assume **(H)**-**(K)**.

Theorem *[B./Fanzon '20]*

$$\min_{(\rho,m)\in\mathcal{M}(X)^{d+1}} T_{\alpha,\beta}(\rho,m) \qquad \text{(Tikh)}$$

admits a solution for $f \in L^2_H$.

- If K_t^* is injective for a.e. t, then the solution is unique.

Theorem *[B./Fanzon '20]*

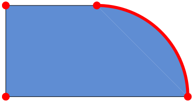
- $\{f^n\}$ noisy data such that $f^n \to f^\dagger$ strongly in L^2_H
- $K_t^* \rho_t^\dagger = f_t^\dagger$ for a.e. $t \in [0,1]$
- (ρ^n, m^n) be a solution to (Tikh) with data f^n and $\alpha_n, \beta_n \to 0$ suitably

Then: $$(\rho^n, m^n) \overset{*}{\rightharpoonup} (\rho^\dagger, m^\dagger) \ \text{ in } \ \mathcal{M}(X)^{d+1}$$

Extremal points of the regularizer

Extremal points

- A $u \in C$ is extremal for a convex set C, if



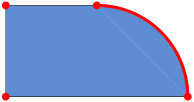
$$\begin{cases} u = \lambda u_1 + (1-\lambda) u_2 \text{ for } u_1, u_2 \in C, \\ 0 < \lambda < 1 \text{ implies } u = u_1 = u_2 \end{cases}$$

- $\mathrm{Extr}(C)$ is the set of extremal points of C

Extremal points of the regularizer

Extremal points

- A $u \in C$ is extremal for a convex set C, if



$$\begin{cases} u = \lambda u_1 + (1-\lambda) u_2 \text{ for } u_1, u_2 \in C, \\ 0 < \lambda < 1 \text{ implies } u = u_1 = u_2 \end{cases}$$

- $\mathrm{Extr}(C)$ is the set of extremal points of C

Goal

- Consider the unit ball of $J_{\alpha,\beta}(\rho, m) = \alpha B(\rho, m) + \beta \|\rho\|_{\mathcal{M}}$
- Determine the extremal points of $J_{\alpha,\beta}$-balls

$$C = \{(\rho, m) \in \mathcal{M}(X) \times \mathcal{M}(X)^d \mid J_{\alpha,\beta}(\rho, m) \leq 1\}$$

Extremal points of the regularizer

Definition

- For $\gamma \in \mathrm{AC}^2([0,1], \overline{\Omega})$ define $\rho_\gamma \in \mathcal{M}(X)$, $m_\gamma \in \mathcal{M}(X)^d$:

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

Extremal points of the regularizer

Definition

- For $\gamma \in \mathrm{AC}^2([0,1],\overline{\Omega})$ define $\rho_\gamma \in \mathcal{M}(X)$, $m_\gamma \in \mathcal{M}(X)^d$:

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

Theorem *[B./Carioni/Fanzon/Romero '21]*

The extremal points of C are characterized by

$$\mathrm{Extr}(C) = \{(0,0)\} \cup \mathcal{C}$$

where $\mathcal{C} = \big\{(\rho_\gamma, m_\gamma) \,\big|\, \gamma \in \mathrm{AC}^2([0,1];\overline{\Omega})\big\}$

Sparsity for finite-dimensional data

Fix $N \geq 1$ times $0 < t_1 < t_2 < \cdots < t_N < 1$, let

- H_i finite-dimensional Hilbert space, $\mathcal{H} = \bigtimes_{i=1}^{N} H_i$
- $K_i^* : \mathcal{M}(\overline{\Omega}) \to H_i$ linear and weak*-continuous

Sparsity for finite-dimensional data

Fix $N \geq 1$ times $0 < t_1 < t_2 < \cdots < t_N < 1$, let

- H_i finite-dimensional Hilbert space, $\mathcal{H} = \bigtimes_{i=1}^N H_i$
- $K_i^* : \mathcal{M}(\overline{\Omega}) \to H_i$ linear and weak*-continuous

Inverse problem

- For $(f_1, \ldots, f_N) \in \mathcal{H}$ find a curve $t \mapsto \rho_t \in \mathcal{M}(\overline{\Omega})$ such that

$$K_i^* \rho_{t_i} = f_i \quad \text{for} \quad i = 1, \ldots, N$$

Sparsity for finite-dimensional data

Fix $N \geq 1$ times $0 < t_1 < t_2 < \cdots < t_N < 1$, let

- H_i finite-dimensional Hilbert space, $\mathcal{H} = \bigtimes_{i=1}^N H_i$
- $K_i^* : \mathcal{M}(\overline{\Omega}) \to H_i$ linear and weak*-continuous

Inverse problem

- For $(f_1, \ldots, f_N) \in \mathcal{H}$ find a curve $t \mapsto \rho_t \in \mathcal{M}(\overline{\Omega})$ such that

$$K_i^* \rho_{t_i} = f_i \quad \text{for} \quad i = 1, \ldots, N$$

Theorem *[B./Carioni/Fanzon/Romero '21]*

$$\min_{(\rho,m) \in \mathcal{M}(X)^{d+1}} \frac{1}{2} \sum_{i=1}^N \|K_i \rho_{t_i} - f_i\|_{H_i}^2 + \alpha B(\rho, m) + \beta \|\rho\|_{\mathcal{M}(X)}$$

admits a solution of the form $\quad (\rho^*, m^*) = \sum_{i=1}^p c_i \, (\rho_{\gamma_i}, m_{\gamma_i})$

Sparsity for finite-dimensional data

Fix $N \geq 1$ times $0 < t_1 < t_2 < \cdots < t_N < 1$, let

- H_i finite-dimensional Hilbert space, $\mathcal{H} = \bigtimes_{i=1}^N H_i$
- $K_i^* : \mathcal{M}(\overline{\Omega}) \to H_i$ linear and weak*-continuous

Inverse problem

- For $(f_1, \ldots, f_N) \in \mathcal{H}$ find a curve $t \mapsto \rho_t \in \mathcal{M}(\overline{\Omega})$ such that

$$K_i^* \rho_{t_i} = f_i \quad \text{for} \quad i = 1, \ldots, N$$

Theorem *[B./Carioni/Fanzon/Romero '21]*

$$\min_{(\rho,m)\in\mathcal{M}(X)^{d+1}} \frac{1}{2} \sum_{i=1}^N \|K_i \rho_{t_i} - f_i\|_{H_i}^2 + \alpha B(\rho, m) + \beta \|\rho\|_{\mathcal{M}(X)}$$

admits a solution of the form $(\rho^*, m^*) = \sum_{i=1}^p c_i \, (\rho_{\gamma_i}, m_{\gamma_i})$

Proof Also see *[Boyer et al. '19]*, *[B./Carioni '20]*

Outline

A conditional gradient method

Consider the equivalent time-continuous problem

$$\min_{(\rho,m)\in\mathcal{M}(X)^{d+1}} \tilde{T}_{\alpha,\beta}(\rho, m)$$

for $$\tilde{T}_{\alpha,\beta}(\rho, m) = \frac{1}{2}\int_0^1 \|K_t^* \rho_t - f_t\|_{H_t}^2 \, \mathrm{d}t + \varphi(J_{\alpha,\beta}(\rho, m))$$

where, e.g., $\varphi(t) = t + \chi_{\{s \leq M_0\}}(t)$, $M_0 = \frac{1}{2}\int_0^1 \|f_t\|_{H_t}^2 \, dt$

A conditional gradient method

Consider the equivalent time-continuous problem

$$\min_{(\rho,m)\in\mathcal{M}(X)^{d+1}} \tilde{T}_{\alpha,\beta}(\rho, m)$$

for $$\tilde{T}_{\alpha,\beta}(\rho, m) = \frac{1}{2}\int_0^1 \|K_t^*\rho_t - f_t\|_{H_t}^2 \,\mathrm{d}t + \varphi(J_{\alpha,\beta}(\rho, m))$$

where, e.g., $\varphi(t) = t + \chi_{\{s \leq M_0\}}(t)$, $M_0 = \frac{1}{2}\int_0^1 \|f_t\|_{H_t}^2 \, dt$

Conditional gradient method

- Linearization of the smooth term around $(\tilde{\rho}, \tilde{m})$

$$\min_{(\rho,m)\in\mathcal{M}(X)^{d+1}} -\int_0^1 \langle \rho_t, w_t\rangle_{\mathcal{M}(\overline{\Omega})\times C(\overline{\Omega})} \, dt + \varphi(J_{\alpha,\beta}(\rho, m))$$

- $w_t = -K_t(K_t^*\tilde{\rho}_t - f_t)$

A conditional gradient method

Consider the convex unit ball of $J_{\alpha,\beta}$

$$C = \{(\rho, m) \in \mathcal{M}(X)^{d+1} : \; J_{\alpha,\beta}(\rho, m) \leq 1\}$$

and denote by $\mathrm{Extr}(C) = \{0\} \cup \mathcal{C}$ its extremal points

A conditional gradient method

Consider the convex unit ball of $J_{\alpha,\beta}$

$$C = \{(\rho, m) \in \mathcal{M}(X)^{d+1} : \ J_{\alpha,\beta}(\rho, m) \leq 1\}$$

and denote by $\mathrm{Extr}(C) = \{0\} \cup \mathcal{C}$ its extremal points

Theorem *[B./Carioni/Fanzon/Romero '21]*

- Assume **(H)-(K)**, let $f \in L^2_H$, $t \mapsto \tilde{\rho}_t \in \mathcal{M}(\overline{\Omega})$ weak* continuous, set $w_t = -K_t(K_t^* \tilde{\rho}_t - f_t)$

There exists a solution $(\rho^*, m^*) \in \mathrm{Extr}(C)$ to

$$\min_{(\rho,m)\in C} -\int_0^1 \langle \rho_t, w_t \rangle_{\mathcal{M}(\overline{\Omega})\times C(\overline{\Omega})} \, dt$$

A conditional gradient method

Consider the convex unit ball of $J_{\alpha,\beta}$

$$C = \{(\rho, m) \in \mathcal{M}(X)^{d+1} : \ J_{\alpha,\beta}(\rho, m) \leq 1\}$$

and denote by $\mathrm{Extr}(C) = \{0\} \cup \mathcal{C}$ its extremal points

Theorem *[B./Carioni/Fanzon/Romero '21]*

- Assume **(H)-(K)**, let $f \in L^2_H$, $t \mapsto \tilde{\rho}_t \in \mathcal{M}(\overline{\Omega})$ weak* continuous, set $w_t = -K_t(K_t^* \tilde{\rho}_t - f_t)$

There exists a solution $(\rho^*, m^*) \in \mathrm{Extr}(C)$ to

$$\min_{(\rho,m)\in C} -\int_0^1 \langle \rho_t, w_t \rangle_{\mathcal{M}(\overline{\Omega})\times C(\overline{\Omega})} \, dt$$

and an $M \geq 0$ such that $(M\rho^*, Mm^*)$ is a solution to

$$\min_{(\rho,m)\in \mathcal{M}(X)^{d+1}} -\int_0^1 \langle \rho_t, w_t \rangle_{\mathcal{M}(\overline{\Omega}), C(\overline{\Omega})} \, dt + \varphi(J_{\alpha,\beta}(\rho, m))$$

A conditional gradient algorithm

Let $f \in L^2_H$ be given. Initialize $(\rho^0, m^0) = (0, 0) \in \mathcal{M}(X)^{d+1}$

A conditional gradient algorithm

Let $f \in L^2_H$ be given. Initialize $(\rho^0, m^0) = (0, 0) \in \mathcal{M}(X)^{d+1}$

1 **Insertion** Assume $(\rho^n, m^n) = \sum_j c_j^n (\rho_{\gamma_j^n}, m_{\gamma_j^n})$ for $\gamma_j^n \in \mathrm{AC}^2([0,1]; \overline{\Omega})$ pairwise distinct, $c_j^n > 0$

A conditional gradient algorithm

Let $f \in L^2_H$ be given. Initialize $(\rho^0, m^0) = (0, 0) \in \mathcal{M}(X)^{d+1}$

1 **Insertion** Assume $(\rho^n, m^n) = \sum_j c_j^n (\rho_{\gamma_j^n}, m_{\gamma_j^n})$ for $\gamma_j^n \in \mathrm{AC}^2([0,1];\overline{\Omega})$ pairwise distinct, $c_j^n > 0$

- Compute the dual variable $w_t^n = -K_t(K_t^* \rho_t^n - f_t)$ and solve

$$\gamma_0^n \in \operatorname*{arg\,min}_{\gamma \in \mathrm{AC}^2([0,1];\overline{\Omega})} -\left(\frac{\alpha}{2}\int_0^1 |\dot{\gamma}(t)|^2 \, dt + \beta\right)^{-1} \int_0^1 w_t^n(\gamma(t)) \, dt$$

A conditional gradient algorithm

Let $f \in L^2_H$ be given. Initialize $(\rho^0, m^0) = (0,0) \in \mathcal{M}(X)^{d+1}$

1 **Insertion** Assume $(\rho^n, m^n) = \sum_j c_j^n (\rho_{\gamma_j^n}, m_{\gamma_j^n})$ for $\gamma_j^n \in \mathrm{AC}^2([0,1];\overline{\Omega})$ pairwise distinct, $c_j^n > 0$

- Compute the dual variable $w_t^n = -K_t(K_t^* \rho_t^n - f_t)$ and solve

$$\gamma_0^n \in \operatorname*{arg\,min}_{\gamma \in \mathrm{AC}^2([0,1];\overline{\Omega})} -\left(\frac{\alpha}{2}\int_0^1 |\dot{\gamma}(t)|^2 \, dt + \beta\right)^{-1} \int_0^1 w_t^n(\gamma(t)) \, dt$$

2 **Optimization** Solve the quadratic program

$$\bar{c}^n = (\bar{c}_j^n)_j \in \operatorname*{arg\,min}_{c_j^n \geq 0} T_{\alpha,\beta}\left(\sum_j c_j^n (\rho_{\gamma_j^n}, m_{\gamma_j^n})\right)$$

A conditional gradient algorithm

Let $f \in L^2_H$ be given. Initialize $(\rho^0, m^0) = (0, 0) \in \mathcal{M}(X)^{d+1}$

1 **Insertion** Assume $(\rho^n, m^n) = \sum_j c_j^n (\rho_{\gamma_j^n}, m_{\gamma_j^n})$ for $\gamma_j^n \in \mathrm{AC}^2([0,1]; \overline{\Omega})$ pairwise distinct, $c_j^n > 0$

- Compute the dual variable $w_t^n = -K_t(K_t^* \rho_t^n - f_t)$ and solve

$$\gamma_0^n \in \operatorname*{arg\,min}_{\gamma \in \mathrm{AC}^2([0,1];\overline{\Omega})} -\Big(\frac{\alpha}{2}\int_0^1 |\dot{\gamma}(t)|^2 \, dt + \beta\Big)^{-1} \int_0^1 w_t^n(\gamma(t)) \, dt$$

2 **Optimization** Solve the quadratic program

$$\bar{c}^n = (\bar{c}_j^n)_j \in \operatorname*{arg\,min}_{c_j^n \geq 0} T_{\alpha,\beta}\Big(\sum_j c_j^n (\rho_{\gamma_j^n}, m_{\gamma_j^n})\Big)$$

- Set $(\rho^{n+1}, m^{n+1}) = \sum_j \bar{c}_j^n (\rho_{\gamma_j^n}, m_{\gamma_j^n})$

Convergence

- Define functional distance $r(\rho, m) = T_{\alpha,\beta}(\rho, m) - \min T_{\alpha,\beta}$

Theorem *[B./Carioni/Fanzon/Romero '22]*
Let $f \in L^2_H$, $\alpha, \beta > 0$, $\{(\rho^n, m^n)\}$ in $\mathcal{M}(X)^{d+1}$ the sequence in the conditional gradient method. Then,

- $\{(\rho^n, m^n)\}$ is minimizing with $$r(\rho^n, m^n) \leq \frac{C}{n}$$ where $C > 0$ depends only on f, α, β
- Each weak* accumulation point of $\{(\rho^n, m^n)\}$ is a minimizer for $T_{\alpha,\beta}$

Convergence

- Define functional distance $r(\rho, m) = T_{\alpha,\beta}(\rho, m) - \min T_{\alpha,\beta}$

Theorem *[B./Carioni/Fanzon/Romero '22]*
Let $f \in L^2_H$, $\alpha, \beta > 0$, $\{(\rho^n, m^n)\}$ in $\mathcal{M}(X)^{d+1}$ the sequence in the conditional gradient method. Then,

- $\{(\rho^n, m^n)\}$ is minimizing with $$r(\rho^n, m^n) \leq \frac{C}{n}$$ where $C > 0$ depends only on f, α, β
- Each weak* accumulation point of $\{(\rho^n, m^n)\}$ is a minimizer for $T_{\alpha,\beta}$

Improving convergence
Simulations suggest linear convergence. Therefore one could expect that the bound can be improved.

Details and additionals tweaks

- Solve the curve insertion problem

$$\gamma_0^n \in \operatorname*{arg\,min}_{\gamma \in \mathrm{AC}^2([0,1];\overline{\Omega})} -\left(\frac{\alpha}{2}\int_0^1 |\dot{\gamma}(t)|^2 \, dt + \beta\right)^{-1} \int_0^1 w_t^n(\gamma(t)) \, dt$$

  via gradient descent with suitable stepsize rule

Details and additionals tweaks

- Solve the curve insertion problem

$$\gamma_0^n \in \operatorname*{arg\,min}_{\gamma \in \mathrm{AC}^2([0,1];\overline{\Omega})} -\left(\frac{\alpha}{2}\int_0^1 |\dot{\gamma}(t)|^2\,dt + \beta\right)^{-1} \int_0^1 w_t^n(\gamma(t))\,dt$$

via gradient descent with suitable stepsize rule

Theorem *[B./Carioni/Fanzon/Romero '22]*
Under suitable regularity assumptions, the gradient descent procedure converges subsequentially to stationary points and strongly in $\mathrm{AC}^2([0,1];\overline{\Omega})$.

Details and additionals tweaks

- Solve the curve insertion problem

$$\gamma_0^n \in \operatorname*{arg\,min}_{\gamma \in \mathrm{AC}^2([0,1];\overline{\Omega})} -\left(\frac{\alpha}{2}\int_0^1 |\dot{\gamma}(t)|^2 \, dt + \beta\right)^{-1} \int_0^1 w_t^n(\gamma(t)) \, dt$$

via gradient descent with suitable stepsize rule

Theorem *[B./Carioni/Fanzon/Romero '22]*
Under suitable regularity assumptions, the gradient descent procedure converges subsequentially to stationary points and strongly in $\mathrm{AC}^2([0,1];\overline{\Omega})$.

- Multiple starts with suitable initial guess (crossovers, random curves, etc.) to increase chance to obtain global minimizer
- Multiple insertion $\rightsquigarrow$ insert all obtained stationary points

Details and additionals tweaks

- **Alternative** Curve insertion via dynamic programming
  $\rightsquigarrow$ *[Duval/Tovey '21]*

Details and additionals tweaks

- **Alternative** Curve insertion via dynamic programming
  $\rightsquigarrow$ *[Duval/Tovey '21]*

- **Sliding step** Perform gradient descent steps for

$$\min_{c_j^n \geq 0,\ \gamma_j^n \in \mathrm{AC}^2([0,1];\overline{\Omega})} T_{\alpha,\beta}\Big(\sum_j c_j^n (\rho_{\gamma_j^n}, m_{\gamma_j^n})\Big)$$

UNI GRAZ

Details and additionals tweaks

- **Alternative** Curve insertion via dynamic programming
  $\rightsquigarrow$ *[Duval/Tovey '21]*

- **Sliding step** Perform gradient descent steps for

$$\min_{c_j^n \geq 0,\ \gamma_j^n \in AC^2([0,1];\overline{\Omega})} T_{\alpha,\beta}\Big(\sum_j c_j^n(\rho_{\gamma_j^n}, m_{\gamma_j^n})\Big)$$

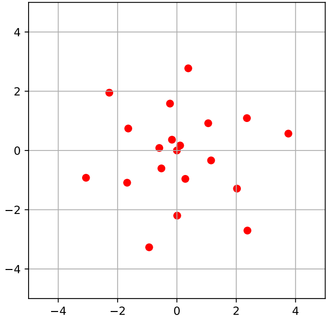
- **Stopping criterion**

$$\left(\frac{\alpha}{2}\int_0^1 |\dot{\gamma}_0^n(t)|^2\,dt + \beta\right)^{-1}\int_0^1 w_t^n(\gamma_0^n(t))\,dt \leq 1$$

or up to some tolerance

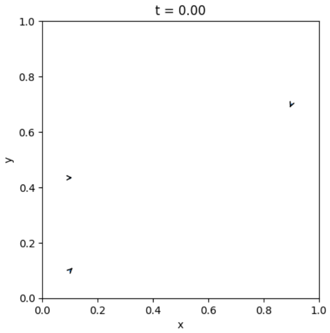
Numerical experiments

- $\Omega = (0,1)^2$, $\sigma = \mathcal{H}^0 \llcorner s$ where $s =$ spiral points in Ω
- $H_t = L^2_\sigma(\mathbf{R}^2, \mathbf{C})$ (time independent)
- $K_t^* : \mathcal{M}(\overline{\Omega}) \to H_t$ masked Fourier transform

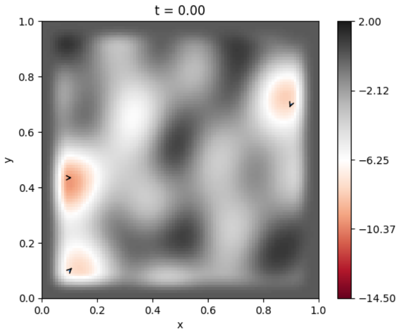


Numerical experiments

A simple example



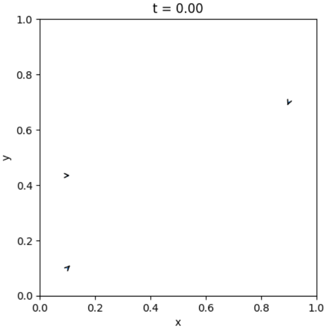
Ground truth



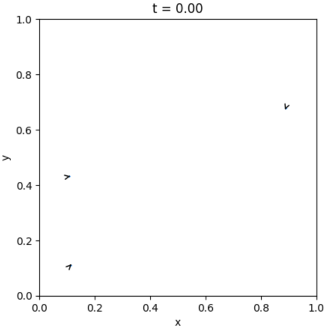
Backprojected data

Numerical experiments

A simple example



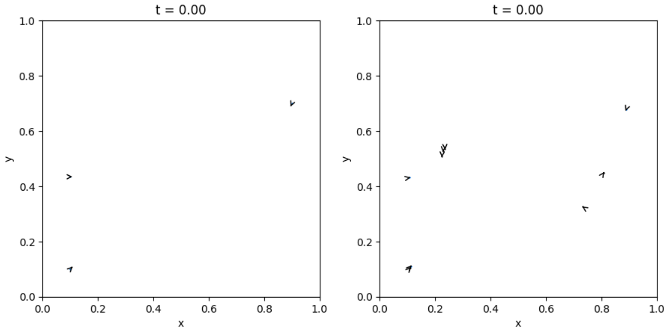
Ground truth



Reconstruction
(thresholded at 0.01)

Numerical experiments

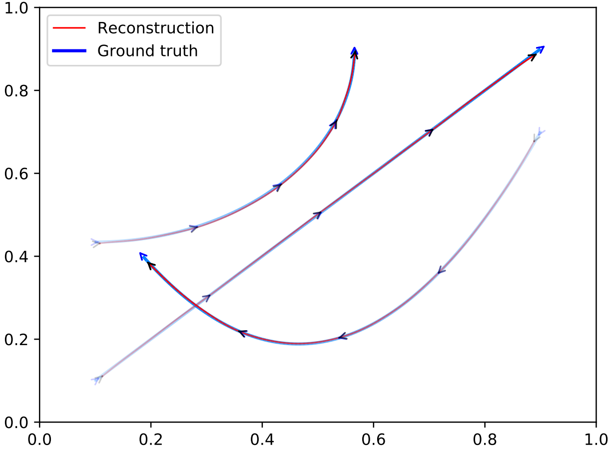
A simple example



Ground truth

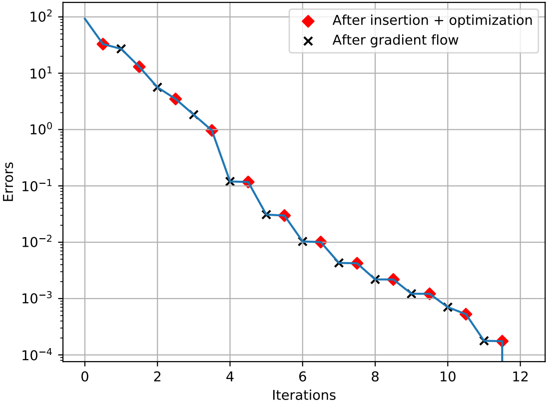
Reconstruction (no thresholding)

Numerical experiments



Reconstructed trajectories

Numerical experiments

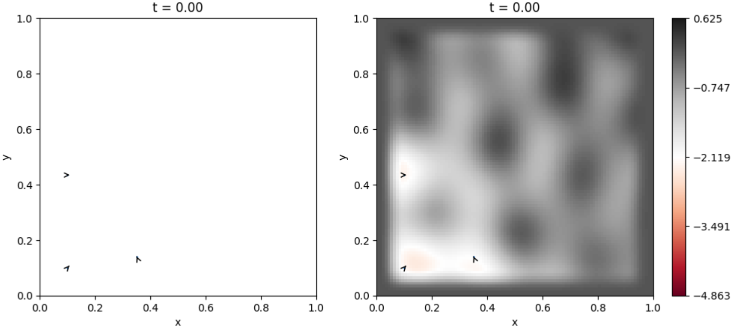


Convergence plot: exhibits linear rate

$$\text{Error} = T_{\alpha,\beta}(\rho^n, m^n) - T_{\alpha,\beta}(\rho^{n+1}, m^{n+1})$$

Numerical experiments

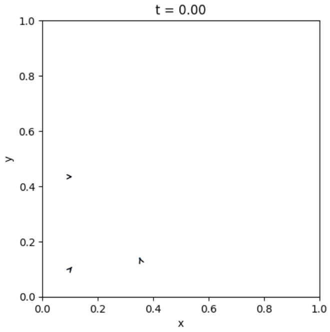
A more difficult example



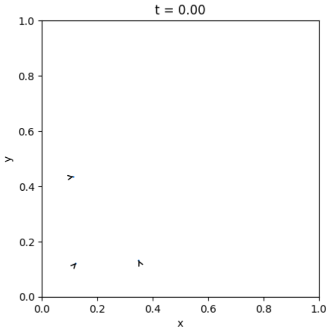
Ground truth | Backprojected data

Numerical experiments

A more difficult example



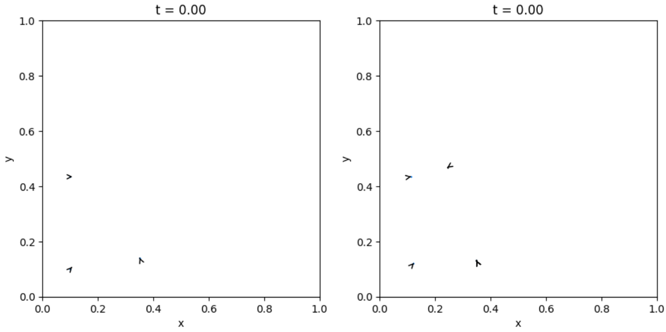
Ground truth



Reconstruction (thresholded at 0.05)

Numerical experiments

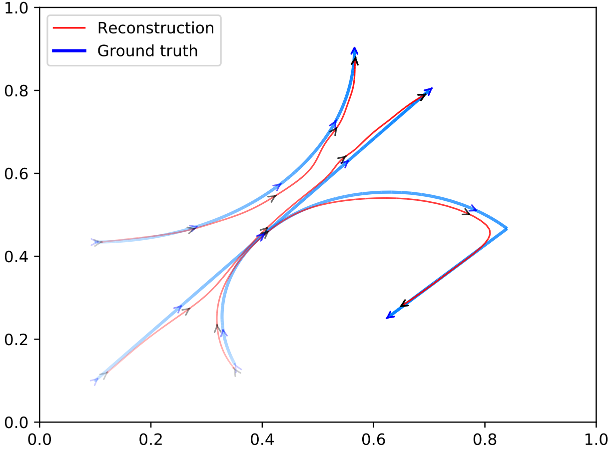
A more difficult example



Ground truth

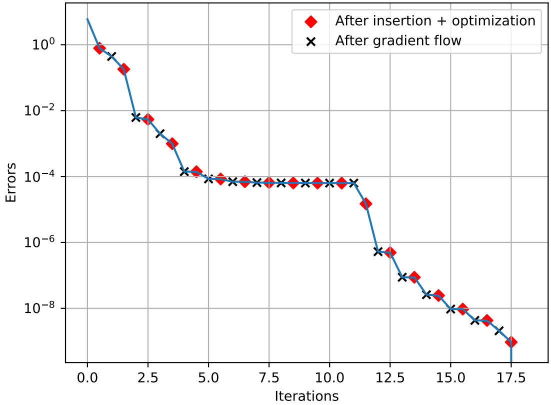
Reconstruction (no thresholding)

Numerical experiments



Reconstructed trajectories

Numerical experiments

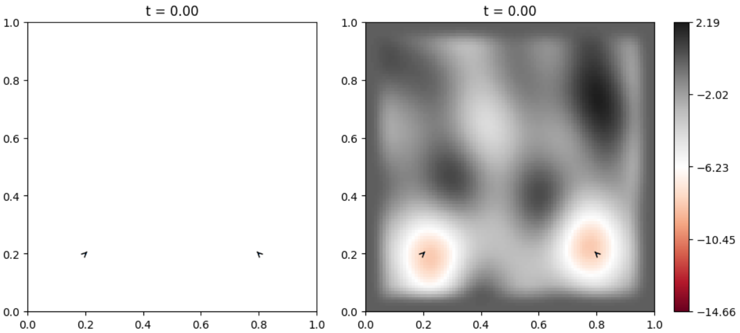


Convergence plot: exhibits linear rate

$$\text{Error} = T_{\alpha,\beta}(\rho^n, m^n) - T_{\alpha,\beta}(\rho^{n+1}, m^{n+1})$$

Numerical experiments

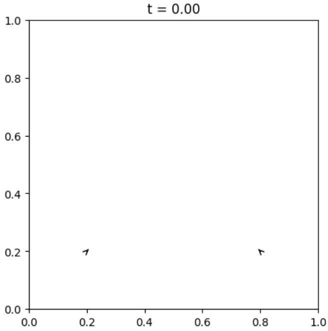
A crossing example



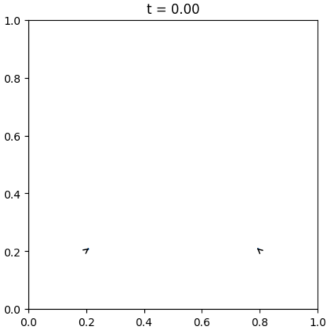
Ground truth Backprojected data

Numerical experiments

A crossing example



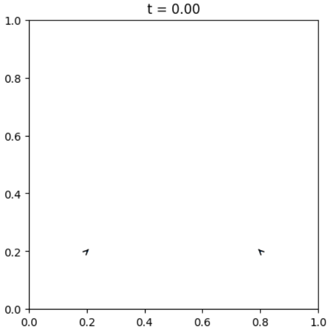
Ground truth



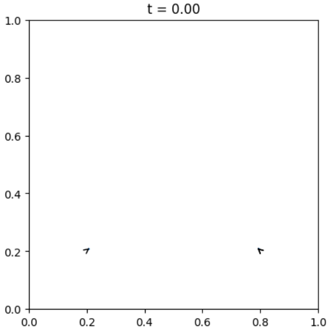
Reconstruction
(thresholded at 0.01)

Numerical experiments

A crossing example

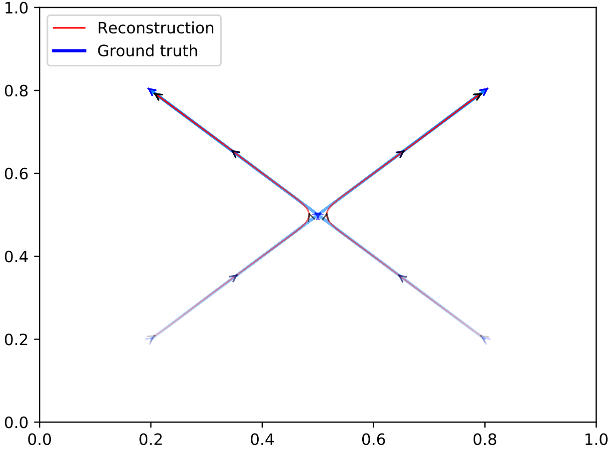


Ground truth



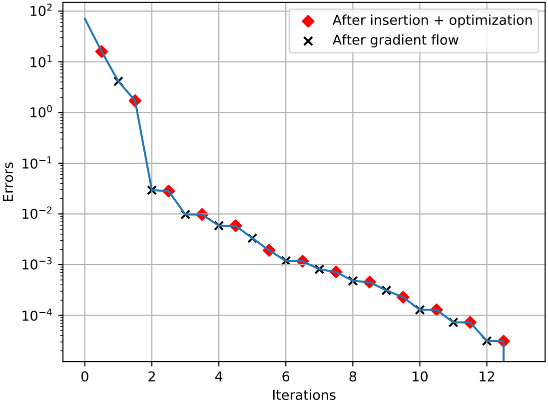
Reconstruction (no thresholding)

Numerical experiments



Reconstructed trajectories

Numerical experiments



Convergence plot: exhibits linear rate

$$\text{Error} = T_{\alpha,\beta}(\rho^n, m^n) - T_{\alpha,\beta}(\rho^{n+1}, m^{n+1})$$

Outline

General setting

Sparse inverse problems

Solve $$\min_{u \in \mathcal{M}} F(Ku) + \mathcal{G}(u)$$

- $\mathcal{M} = \mathcal{C}^*$ dual space, $\mathcal{C}$ separable, Y Hilbert space
- $K : \mathcal{M} \to Y$, weak*-to-weak continuous and weak*-to-strong continuous in $\mathrm{dom}(\mathcal{G})$
- $F : Y \to \mathbb{R}$ bounded from below, strictly convex, Fréchet differentiable, ∇F Lipschitz on compact sets
- $\mathcal{G} : \mathcal{M} \to [0, \infty]$, convex, weak* lower semi-continuous, positive one-homogeneous, coercive

General setting

Sparse inverse problems

Solve
$$\min_{u \in \mathcal{M}} F(Ku) + \mathcal{G}(u)$$

- $\mathcal{M} = \mathcal{C}^*$ dual space, $\mathcal{C}$ separable, Y Hilbert space
- $K : \mathcal{M} \to Y$, weak*-to-weak continuous and weak*-to-strong continuous in $\mathrm{dom}(\mathcal{G})$
- $F : Y \to \mathbb{R}$ bounded from below, strictly convex, Fréchet differentiable, ∇F Lipschitz on compact sets
- $\mathcal{G} : \mathcal{M} \to [0, \infty]$, convex, weak* lower semi-continuous, positive one-homogeneous, coercive

What does sparsity mean in this context?

$\rightsquigarrow$ **extremal points**

General representer theorems

Theorem *[B./Carioni '20]*

- If Y is finite-dimensional, then there are sparse solutions of $\min_{u\in\mathcal{M}} \; F(Ku) + \mathcal{G}(u)$, i.e.,

$$u^* = \sum_{i=1}^{N} v_i^* u_i^*, \qquad v_i^* > 0,\; u_i^* \in \operatorname{Extr}(\{\mathcal{G} \leq 1\})$$

- Also see *[Boyer/Chambolle/De Castro/Duval/De Gournay/Weiss '19]*

UNI GRAZ

Choquet's theorem

Theorem *[Choquet]*

- X locally convex space, $K \subset X$ non-empty, convex, metrizable, compact
- For each $v \in K$, there is a probability measure μ over X concentrated on $\mathrm{Extr}(K)$ with

$$T(v) = \int_X T \, \mathrm{d}\mu \qquad \text{for each } T \in X^*$$

Choquet's theorem

Theorem *[Choquet]*

- X locally convex space, $K \subset X$ non-empty, convex, metrizable, compact
- For each $v \in K$, there is a probability measure μ over X concentrated on $\mathrm{Extr}(K)$ with

$$T(v) = \int_X T \, \mathrm{d}\mu \qquad \text{for each } T \in X^*$$

Lifting of $\mathcal{G}$:

- With $\mathcal{B} = \overline{\mathrm{Extr}(\{\mathcal{G} \leq 1\})}^*$, there is $\mathcal{I} : \mathcal{M}_+(\mathcal{B}) \to \mathrm{dom}(\mathcal{G})$

$$\begin{cases} \mathcal{G}(\mathcal{I}\mu) \leq \|\mu\|_{\mathcal{M}} \text{ for all } \mu \in \mathcal{M}_+(\mathcal{B}) \\ \mathcal{G}(u) = \|\mu\|_{\mathcal{M}} \text{ for } u \in \mathrm{dom}(\mathcal{G}), \mu \in \mathcal{M}_+(\mathcal{B}), \mathcal{I}\mu = u \end{cases}$$

Lifted conditional gradient method

Equivalent problem: $\min_{\mu \in \mathcal{M}_+(\mathcal{B})} F(K\mathcal{I}\mu) + \|\mu\|_{\mathcal{M}}$

- Employ variant of PDAP algorithm $\rightsquigarrow$ fully corrective generalized conditional gradient method (FC-GCG)

UNI GRAZ

Lifted conditional gradient method

Equivalent problem: $\min_{\mu \in \mathcal{M}_+(\mathcal{B})} F(K\mathcal{I}\mu) + \|\mu\|_{\mathcal{M}}$

- Employ variant of PDAP algorithm $\rightsquigarrow$ fully corrective generalized conditional gradient method (FC-GCG)

Algorithm

1. Compute $p^n = -K_* \nabla F(Ku^n)$, find $v^n \in \arg\max_{v \in \mathrm{Extr}(\{\mathcal{G} \leq 1\})} \langle p^n, v \rangle$, set $\mathcal{A}_{n+1/2} = \mathcal{A}_n \cup \{v^n\}$
2. Compute solution λ^{n+1} of
$$\min_{\lambda \in [0,\infty[^{\mathcal{A}_{n+1/2}}} F\Big(\sum_{v \in \mathcal{A}_{n+1/2}} \lambda(v) v\Big) + \sum_{v \in \mathcal{A}_{n+1/2}} \lambda(v)$$
3. Set $\mathcal{A}_{n+1} = \operatorname{supp} \lambda^{n+1}$, $u^{n+1} = \sum_{v \in \mathcal{A}_{n+1}} \lambda^{n+1}(v) v$

Lifted conditional gradient method

Equivalent problem: $\min_{\mu \in \mathcal{M}_+(\mathcal{B})} F(K\mathcal{I}\mu) + \|\mu\|_{\mathcal{M}}$

- Employ variant of PDAP algorithm $\rightsquigarrow$ fully corrective generalized conditional gradient method (FC-GCG)

Algorithm

1. Compute $p^n = -K_* \nabla F(Ku^n)$, find $v^n \in \underset{v \in \mathrm{Extr}(\{\mathcal{G} \leq 1\})}{\arg\max} \langle p^n, v \rangle$, set $\mathcal{A}_{n+1/2} = \mathcal{A}_n \cup \{v^n\}$
2. Compute solution λ^{n+1} of
$$\min_{\lambda \in [0,\infty[^{\mathcal{A}_{n+1/2}}} F\Big(\sum_{v \in \mathcal{A}_{n+1/2}} \lambda(v) v\Big) + \sum_{v \in \mathcal{A}_{n+1/2}} \lambda(v)$$
3. Set $\mathcal{A}_{n+1} = \mathrm{supp}\, \lambda^{n+1}$, $u^{n+1} = \sum_{v \in \mathcal{A}_{n+1}} \lambda^{n+1}(v) v$

Lifted conditional gradient method

Equivalent problem: $\min_{\mu \in \mathcal{M}_+(\mathcal{B})} F(K\mathcal{I}\mu) + \|\mu\|_{\mathcal{M}}$

- Employ variant of PDAP algorithm $\rightsquigarrow$ fully corrective generalized conditional gradient method (FC-GCG)

Algorithm

1. Compute $p^n = -K_* \nabla F(Ku^n)$, find $v^n \in \arg\max_{v \in \mathrm{Extr}(\{\mathcal{G} \leq 1\})} \langle p^n, v \rangle$, set $\mathcal{A}_{n+1/2} = \mathcal{A}_n \cup \{v^n\}$
2. Compute solution λ^{n+1} of
$$\min_{\lambda \in [0,\infty[^{\mathcal{A}_{n+1/2}}} F\Big(\sum_{v \in \mathcal{A}_{n+1/2}} \lambda(v) v\Big) + \sum_{v \in \mathcal{A}_{n+1/2}} \lambda(v)$$
3. Set $\mathcal{A}_{n+1} = \operatorname{supp} \lambda^{n+1}$, $u^{n+1} = \sum_{v \in \mathcal{A}_{n+1}} \lambda^{n+1}(v) v$

Lifted conditional gradient method

Equivalent problem: $\min_{\mu \in \mathcal{M}_+(\mathcal{B})} F(K\mathcal{I}\mu) + \|\mu\|_{\mathcal{M}}$

- Employ variant of PDAP algorithm $\rightsquigarrow$ fully corrective generalized conditional gradient method (FC-GCG)

Algorithm

1. Compute $p^n = -K_* \nabla F(Ku^n)$, find $v^n \in \arg\max_{v \in \mathrm{Extr}(\{\mathcal{G} \leq 1\})} \langle p^n, v \rangle$, set $\mathcal{A}_{n+1/2} = \mathcal{A}_n \cup \{v^n\}$
2. Compute solution λ^{n+1} of
$$\min_{\lambda \in [0,\infty[^{\mathcal{A}_{n+1/2}}} F\Big(\sum_{v \in \mathcal{A}_{n+1/2}} \lambda(v) v\Big) + \sum_{v \in \mathcal{A}_{n+1/2}} \lambda(v)$$
3. Set $\mathcal{A}_{n+1} = \operatorname{supp} \lambda^{n+1}$, $u^{n+1} = \sum_{v \in \mathcal{A}_{n+1}} \lambda^{n+1}(v) v$

GCG method $\rightsquigarrow$ convergence rate $\mathcal{O}(n^{-1})$

Linear convergence: Assumptions

Uniqueness and sparsity assumption

- For u^* optimal solution, $y^* = Ku^*$ optimal observation, $p^* = -K_*\nabla F(y^*)$ dual certificate:
- F is strongly convex around y^*
- $\langle p^*, u\rangle = 1$ only at u_i^*, $i = 1, \ldots, N$, $p^* < 1$ otherwise on $\mathcal{B}$
- $Ku_1^*, \ldots, Ku_N^*$ are linearly independent

Linear convergence: Assumptions

Uniqueness and sparsity assumption

- For u^* optimal solution, $y^* = Ku^*$ optimal observation, $p^* = -K_* \nabla F(y^*)$ dual certificate:
- F is strongly convex around y^*
- $\langle p^*, u \rangle = 1$ only at u_i^*, $i = 1, \ldots, N$, $p^* < 1$ otherwise on $\mathcal{B}$
- $Ku_1^*, \ldots, Ku_N^*$ are linearly independent

Linear convergence: Assumptions

Uniqueness and sparsity assumption

- For u^* optimal solution, $y^* = Ku^*$ optimal observation, $p^* = -K_* \nabla F(y^*)$ dual certificate:
- F is strongly convex around y^*
- $\langle p^*, u \rangle = 1$ only at u_i^*, $i = 1, \ldots, N$, $p^* < 1$ otherwise on $\mathcal{B}$
- $Ku_1^*, \ldots, Ku_N^*$ are linearly independent

Linear convergence: Assumptions

Uniqueness and sparsity assumption

- For u^* optimal solution, $y^* = Ku^*$ optimal observation, $p^* = -K_* \nabla F(y^*)$ dual certificate:
- F is strongly convex around y^*
- $\langle p^*, u \rangle = 1$ only at u_i^*, $i = 1, \ldots, N$, $p^* < 1$ otherwise on $\mathcal{B}$
- $Ku_1^*, \ldots, Ku_N^*$ are linearly independent

UNI GRAZ

Linear convergence: Assumptions

Uniqueness and sparsity assumption

- For u^* optimal solution, $y^* = Ku^*$ optimal observation, $p^* = -K_* \nabla F(y^*)$ dual certificate:
- F is strongly convex around y^*
- $\langle p^*, u \rangle = 1$ only at u_i^*, $i = 1, \ldots, N$, $p^* < 1$ otherwise on $\mathcal{B}$
- $Ku_1^*, \ldots, Ku_N^*$ are linearly independent

Non-degeneracy assumption

- $u^* = \sum_{i=1}^N \lambda_i^* u_i^*$ with $\lambda_1^*, \ldots, \lambda_N^* > 0$
- There is $g : \mathrm{Extr}(\{\mathcal{G} \leq 1\})^2 \to [0, \infty[$ such that locally around each u_i^*:

$$\|K(v - u_i^*)\| \leq \tau g(v, u_i^*), \quad \langle p^*, v \rangle \leq 1 - \kappa g(v, u_i^*)^2$$

Linear convergence: Assumptions

Uniqueness and sparsity assumption

- For u^* optimal solution, $y^* = Ku^*$ optimal observation, $p^* = -K_*\nabla F(y^*)$ dual certificate:
- F is strongly convex around y^*
- $\langle p^*, u\rangle = 1$ only at u_i^*, $i = 1, \ldots, N$, $p^* < 1$ otherwise on $\mathcal{B}$
- $Ku_1^*, \ldots, Ku_N^*$ are linearly independent

Non-degeneracy assumption

- $u^* = \sum_{i=1}^N \lambda_i^* u_i^*$ with $\lambda_1^*, \ldots, \lambda_N^* > 0$
- There is $g : \mathrm{Extr}(\{\mathcal{G} \leq 1\})^2 \to [0, \infty[$ such that locally around each u_i^*:

$$\|K(v - u_i^*)\| \leq \tau g(v, u_i^*), \quad \langle p^*, v\rangle \leq 1 - \kappa g(v, u_i^*)^2$$

Linear convergence: Result

Theorem *[B./Carioni/Fanzon/Walter '21]*

The FC-GCG method converges linearly under the above assumptions.

Linear convergence: Result

Theorem *[B./Carioni/Fanzon/Walter '21]*

The FC-GCG method converges linearly under the above assumptions.

Ingredients for the proof:

- For n large, all $\mathcal{A}_n$ are contained in an appropriate neighborhood of $\{u_1^*, \dots, u_N^*\}$
- Letting $\mu^n = \sum_{v \in \mathcal{A}_n} \lambda^n(v)\delta_v$, relate
$$g(v^n, u_i^*) \leq c\sqrt{r(\mu^n)} \qquad \text{for appropriate } i$$
- Relate $\|K\mathcal{I}(\hat{\mu}^n - \mu^n)\| \leq c\sqrt{r(\mu^n)}$ where $\hat{\mu}^n$ concentrates the mass of μ^n around v^n to v^n
- $r(\mu^{n+1}) \leq qr(\mu^n)$ using a convex combination of μ^n and $\hat{\mu}^n$

(Nontrivial) examples

Minimum effort problems $$\min_{u \in L^\infty(\Omega)} F(Ku) + \alpha\|u\|_\infty$$

(Nontrivial) examples

Minimum effort problems $\min_{u \in L^\infty(\Omega)} F(Ku) + \alpha \|u\|_\infty$

- $\mathrm{Extr}(\{\mathcal{G} \leq 1\}) = \{u \in L^\infty(\Omega) \mid |u| = \alpha^{-1} \text{ a.e. in } \Omega\}$
- For $p \in L^1(\Omega)$, we have $\bar{v} \in \arg\max_{v \in \mathrm{Extr}(\{\mathcal{G} \leq 1\})} \langle p, v \rangle$ if
$$\bar{v} = \alpha^{-1} \operatorname{sign}(p)$$
- Non-degeneracy assumption satisfiable for
$g(v, u) = \|v - u\|_1$ and $|\{|p^*| \leq \varepsilon\}| \leq C\varepsilon$

(Nontrivial) examples

Minimum effort problems $\min_{u \in L^\infty(\Omega)} F(Ku) + \alpha \|u\|_\infty$

- $\mathrm{Extr}(\{\mathcal{G} \leq 1\}) = \{u \in L^\infty(\Omega) \mid |u| = \alpha^{-1} \text{ a.e. in } \Omega\}$
- For $p \in L^1(\Omega)$, we have $\bar{v} \in \arg\max_{v \in \mathrm{Extr}(\{\mathcal{G} \leq 1\})} \langle p, v \rangle$ if
$$\bar{v} = \alpha^{-1} \operatorname{sign}(p)$$
- Non-degeneracy assumption satisfiable for $g(v, u) = \|v - u\|_1$ and $|\{|p^*| \leq \varepsilon\}| \leq C\varepsilon$

Trace-norm regularized problems $\min_{\mathcal{U} \geq 0} F(K\mathcal{U}) + \alpha \operatorname{Tr}(\mathcal{U})$

(Nontrivial) examples

Minimum effort problems $\min_{u \in L^\infty(\Omega)} F(Ku) + \alpha \|u\|_\infty$

- $\mathrm{Extr}(\{\mathcal{G} \leq 1\}) = \{u \in L^\infty(\Omega) \mid |u| = \alpha^{-1} \text{ a.e. in } \Omega\}$
- For $p \in L^1(\Omega)$, we have $\bar{v} \in \arg\max_{v \in \mathrm{Extr}(\{\mathcal{G} \leq 1\})} \langle p, v \rangle$ if
$$\bar{v} = \alpha^{-1} \operatorname{sign}(p)$$
- Non-degeneracy assumption satisfiable for $g(v, u) = \|v - u\|_1$ and $|\{|p^*| \leq \varepsilon\}| \leq C\varepsilon$

Trace-norm regularized problems $\min_{\mathcal{U} \geq 0} F(K\mathcal{U}) + \alpha \operatorname{Tr}(\mathcal{U})$

- $\mathrm{Extr}(\{\mathcal{G} \leq 1\}) = \{\alpha^{-1} u \otimes u \mid \|u\| = 1\} \cup \{0\}$
- Solve $\bar{\mathcal{V}} \in \arg\max_{\mathcal{V} \in \mathrm{Extr}(\{\mathcal{G} \leq 1\})} \langle \mathcal{P}, \mathcal{V} \rangle$ via power iteration
- Non-degeneracy assumption satisfiable for rank-1 solutions and $g(\mathcal{V}, \mathcal{U}) = \|\mathcal{V} - \mathcal{U}\|_{\mathrm{HS}}$

Outline

Conclusions

- Introduced rigorous framework for optimal transport regularization of time dependent inverse problems

Conclusions

- Introduced rigorous framework for optimal transport regularization of time dependent inverse problems
- Characterization of the extremal points of the Benamou-Brenier regularizer

Conclusions

- Introduced rigorous framework for optimal transport regularization of time dependent inverse problems
- Characterization of the extremal points of the Benamou-Brenier regularizer
- Numerical algorithm for dynamic spike reconstruction

Conclusions

- Introduced rigorous framework for optimal transport regularization of time dependent inverse problems
- Characterization of the extremal points of the Benamou-Brenier regularizer
- Numerical algorithm for dynamic spike reconstruction
- Application to dynamic superresolution

Conclusions

- Introduced rigorous framework for optimal transport regularization of time dependent inverse problems
- Characterization of the extremal points of the Benamou-Brenier regularizer
- Numerical algorithm for dynamic spike reconstruction
- Application to dynamic superresolution
- A generalization to general sparse reconstruction is possible ($\rightsquigarrow$ FC-GCG method)

Conclusions

- Introduced rigorous framework for optimal transport regularization of time dependent inverse problems
- Characterization of the extremal points of the Benamou-Brenier regularizer
- Numerical algorithm for dynamic spike reconstruction
- Application to dynamic superresolution
- A generalization to general sparse reconstruction is possible ($\rightsquigarrow$ FC-GCG method)
- Under degeneracy assumption on the solution, the FC-GCG method converges linearly

Literature

K. Bredies and M. Carioni.
Sparsity of solutions for variational inverse problems with finite-dimensional data.
Calculus of Variations and Partial Differential Equations, 59:14, 2020.

K. Bredies and S. Fanzon.
An optimal transport approach for solving dynamic inverse problems in spaces of measures.
ESAIM: Mathematical Modelling and Numerical Analysis, 54(6):2351–2382, 2020.

K. Bredies, M. Carioni, S. Fanzon and F. Romero.
On the extremal points of the ball of the Benamou-Brenier energy.
Bulletin of the London Mathematical Society, 53(5):1436–1452, 2021.

K. Bredies, M. Carioni, S. Fanzon and F. Romero.
A generalized conditional gradient method for dynamic inverse problems with optimal transport regularization.
Foundations of Computational Mathematics, 2022.

K. Bredies, M. Carioni, S. Fanzon and D. Walter.
Asymptotic linear convergence of fully-corrective generalized conditional gradient methods.
Mathematical Programming, 2023. To appear.