Sparsity and convergence analysis of Generalized Conditional Gradient Methods

Silvio Fanzon

Based on joint works with

Kristian Bredies, Marcello Carioni, Francisco Romero, Daniel Walter

13 September 2022

Frank-Wolfe and Generalized Conditional Gradient methods

Classical Frank-Wolfe-type algorithms aim at solving

$$\min_{x \in C} F(x)$$

where C is a convex set in $\mathbb{R}^N$ and $F \colon \mathbb{R}^N \to \mathbb{R}$ is regular convex function

Algorithm: Given an iterate x^n, compute x^{n+1} in two steps:

- **Insertion step**: Solve the linearized problem around x^n as

$$\widehat{x} \in \operatorname{argmin}_{x \in C} \langle \nabla F(x^n), x \rangle$$

- **Line search step**: Obtain x^{n+1} by interpolating

$$x^{n+1} = x^n + s^*(\widehat{x} - x^n)$$

  for a suitably chosen step-size s^*

Known facts about Frank-Wolfe

- The convergence rate is typically sublinear and it can be improved to linear under strong convexity assumptions on F and other interpolation steps [1] [2].
- The algorithm has been generalized to infinite dimensional spaces (Generalized Conditional Gradient methods) [3]
- Classical algorithms in infinite dimensional optimization have been shown to be particular instances of GCG [4] [5]

---

[1] Revisiting Frank-Wolfe: Projection-Free Sparse Convex Optimization Jaggi, M. (2013)
[2] On the Global Linear Convergence of Frank-Wolfe Optimization Variants Lacoste-Julien, S. and Jaggi, M. (2015)
[3] Approximate methods in optimization problems Demyanov, V. F. and Rubinov A. M. (1970)
[4] An iterative thresholding algorithm for linear inverse problems with a sparsity constraint Daubechies, I., Defrise, M. and De Mol, C. (2004)
[5] Iterated hard shrinkage for minimization problems with sparsity constraints Bredies, K. Lorenz, D. (2006)

GCG in the space of measures

Consider the BLASSO problem

$$\min_{\mu \in \mathcal{M}(\Omega)} G(\mu) := F(K\mu) + \|\mu\|_{\mathcal{M}(\Omega)}$$

with $\Omega \subset \mathbb{R}^d$ domain, $K : \mathcal{M}(\Omega) \to Y$ linear and weak*-to-strong continuous, Y Hilbert space, $F : Y \to [0, \infty)$ strictly convex and smooth

Algorithm: Given an iterate μ^n one computes the next one μ^{n+1} in two steps:

- **Insertion step**: Solve the partially linearized problem around μ^n as

$$\widehat{\mu} \in \operatorname{argmax}_{\|\mu\|_{\mathcal{M}(\Omega)} \leq C} \langle -K_* \nabla F(K\mu^n), \mu \rangle$$

- **Line search step**: μ^{n+1} is obtained interpolating

$$\mu^{n+1} = \mu^n + s^*(\widehat{\mu} - \mu^n) \,, \quad s^* \text{ suitable step-size}$$

Inverse problems in the space of measures K. Bredies, H.K. Pikkarainen. ESAIM:COCV (2013)
The Alternating Descent Conditional Gradient Method for Sparse Inverse Problems N. Boyd, G. Schiebinger, B. Recht. *SIAM Journal on Optimization* (2017)

Key observation: iterate μ^n can be constructed as a combination of Dirac deltas

$$\mu^n = \sum_{i=1}^{k} c_i \, \delta_{x_i}$$

for suitable $c_i \in \mathbb{R}$ and $x_i \in \Omega$

Why: Define the dual variable at n-th iteration as

$$p^n := -K_* \nabla F(K\mu^n) \in C(\Omega)$$

Key Lemma

A solution to

$$\max_{\|\mu\|_{\mathcal{M}(\Omega)} \leq C} \langle \mu, p^n \rangle$$

is given by $c \, \delta_{\hat{x}}$ for some $c \in \mathbb{R}$ and

$$\hat{x} \in \arg\max_{x \in \Omega} |p^n(x)|$$

The next iterate μ^{n+1} is obtained by

$$\mu^{n+1} = \mu^n + s^*(\widehat{\mu} - \mu^n)\,, \quad s^* \text{ suitable step-size}$$

where

$$\widehat{\mu} = c\,\delta_{\hat{x}}\,, \quad \hat{x} \in \arg\max_{x \in \Omega} |p^n(x)|$$

Theorem (Bredies, Pikkarainen (2013))

μ^n converges weakly (up to subsequences) to a minimizer of G. Moreover the rate of convergence is sublinear, i.e.*

$$G(\mu^n) - \min_{\mu \in \mathcal{M}(\Omega)} G(\mu) \leq \frac{C}{n}$$

Remarks:

- The GCG method for BLASSO exploits sparsity of the problem (iterates are linear combinations of Dirac deltas)
- GCG allows to design a grid-free algorithm that does not need an a priori discretization of the domain - we only need to find the max of $|p^n|$
- The Line Search Step can be be improved by optimizing the coefficients $(c_1, \dots, c_{k+1})$ of the full linear combination

$$\sum_{i=1}^{k} c_i \, \delta_{x_i} + c_{k+1} \, \delta_{\hat{x}}$$

with respect the energy G:

$$(c_i^*)_i \in \operatorname{argmin}_{c_i \in \mathbb{R}} \, G\left(\sum_{i=1}^{k} c_i \, \delta_{x_i} + c_{k+1} \, \delta_{\hat{x}} \right) \rightsquigarrow \mu^{n+1} = \sum_{i=1}^{k} c_i^* \, \delta_{x_i} + c_{k+1}^* \, \delta_{\hat{x}}$$

Remark: Coefficient optimization $\rightsquigarrow$ **linear** rate of convergence in experiments

Linear convergence in the space of measures

Question: Is it possible prove linear convergence?

Define the dual variable of a minimizer $\bar{\mu}$ of G

$$\bar{p} = -K_* \nabla F(K\bar{\mu}) \in C(\Omega)$$

We make the following set of assumptions [6] [7]:

- Strong convexity of F, uniqueness and sparsity of the minimizer

(i) F is strongly convex
(ii) There exists $\{x_i\}_i \subset \Omega$ finite such that $\operatorname{argmax}_x |\bar{p}(x)| = \{x_i\}_i$
(iii) The set $\{K\delta_{x_i}\}_i \subset Y$ is linearly independent in Y

(i) + (ii) + (iii) imply that the minimizer $\bar{\mu}$ of G is **unique** and **sparse**, i.e.,

$$\bar{\mu} = \sum_i c_i \, \delta_{x_i}$$

[6] Linear convergence of generalized conditional gradient methods in the space of measures K. Pieper, D. Walter. ESAIM: COCV (2021)

[7] On the linear convergence rates of exchange and continuous methods for total variation minimization A. Flinth, F. De Gournay, P. Weiss. Math. Prog. (2021)

- Quadratic growth of $\nabla^2 \bar{p}$ around u_i

$$-\text{sign}(\bar{p}(x_i))\langle \xi, \nabla^2 \bar{p}(x_i)\xi \rangle \gtrapprox |\xi|^2 \qquad \text{for all } \ i$$

- Lipschitz growth of K around δ_{x_i}

$$\|K(\delta_x - \delta_{x_i})\|_Y \lesssim |x - x_i| \qquad \text{for all } \ i$$

Theorem (Pieper, Walter (2020))

Under the previous assumptions the rate of convergence of μ^n is linear, i.e., there exists $C > 0$, $\zeta \in [1/2, 1)$ such that

$$G(\mu^n) - \min_{\mu \in \mathcal{M}(\Omega)} G(\mu) \leq C\zeta^n$$

GCG methods in infinite dimensions

Goal: generalize the previous algorithm and convergence results to Banach spaces

We consider minimization problems of the form

$$\min_{u \in X} G(u) := F(Au) + R(u)$$

where we assume:

- X separable Banach space with predual X_*
- $F : Y \to \mathbb{R}$ data fidelity term, Y Hilbert space
- $K : X \to Y$ linear continuous measurement operator
- $K_* : Y \to X_*$ pre-adjoint of K
- $R : X \to [0, \infty]$ **convex**, **1-homogeneous** and **coercive**

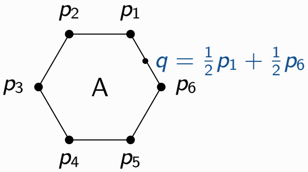
Asymptotic linear convergence of Fully-Corrective Generalized Conditional Gradient methods
K. Bredies, M. Carioni, S. Fanzon, D. Walter. Preprint (2021)

Question: How does **sparsity** enter the algorithm design?

Extremal Points: Given a set B we say that $u \in B$ is an **extremal point** if

$$u = \lambda u_1 + (1-\lambda) u_2 \quad \text{for} \quad \lambda \in (0,1) \quad \Rightarrow \quad u = u_1 = u_2$$

The set of extremal points of B is denoted by $Ext(B)$



Key Lemma

Define unit ball of the regularizer $B := \{u \in X : \ R(u) \leq 1\}$. Let $p \in X_*$. Then

$$c\bar{v} \in \arg\max_{u \in B} \langle u, p \rangle \,, \quad c \in \mathbb{R} \,, \ \bar{v} \in \text{Ext}(B)$$

In the insertion step we add an **extremal point** of B $\rightsquigarrow$ **Sparse** iterates

Accelerated Generalized Conditional Gradient algorithm

Assume the n-th iterate is sparse

$$u^n = \sum_{i=1}^{k} c_i u_i \,, \quad c_i \in \mathbb{R}, \quad u_i \in Ext(B)$$

We compute u^{n+1} in the following way:

- **Insertion step:** Set $p^n = -K_* \nabla F(Ku^n)$ and $u_{k+1} := \widehat{u}$ where

$$\widehat{u} \in \operatorname{argmax}_{u \in B} \langle p^n, u \rangle \,, \quad \widehat{u} \in Ext(B)$$

- **Coefficients optimization step:** solve the finite-dimensional problem

$$(c_1^*, \dots, c_{k+1}^*) \in \operatorname{argmin}_{c_i \in \mathbb{R}_+} \left[F\left(\sum_{i=1}^{k+1} c_i K u_i \right) + \sum_{i=1}^{k+1} c_i \right]$$

Next iterate is

$$u^{n+1} := \sum_{i=1}^{k+1} c_i^* u_i$$

Worst-case convergence rate

Theorem (Bredies, Carioni, Fanzon, Walter (2021))

The iterate u^n weakly converges (up to subsequences) to a minimizer of G at a sublinear rate, i.e.*

$$G(u^n) - \min_u G(u) \leq \frac{C}{n}$$

Remarks

- We recover the GCG for the minimization of the total variation noticing that

$$Ext(\{\mu : \|\mu\|_{\mathcal{M}(\Omega)} \leq 1\}) = \{\pm\delta_x : x \in \Omega\}$$

- Wide class of applications: dynamic inverse problems, group sparsity regularizers, PDE constrained inverse problems, learning problems
- Representer theorems show that extremal points of the regularizer are the natural atoms constituting sparse solutions [8]

[8] Sparsity of solutions for variational inverse problems with finite dimensional data
K. Bredies, M. Carioni. Calc. Var. PDE (2020)

Linear convergence

Goal: Find suitable assumptions under which linear convergence holds

Let $\bar{u}$ be a minimizer of G. Define the associated dual variable

$$\bar{p} := -K_* \nabla F(K\bar{u}) \in X_*$$

Assumptions:

- Strong convexity of F, uniqueness and sparsity of the minimizer

(i) F is strongly convex

(ii) There exists $\{u_i\}_i \subset Ext(B)$ s.t. $\operatorname{argmax}_{v \in \overline{Ext(B)}^*} \langle \bar{p}, v \rangle = \{u_i\}_i$

(iii) The set $\{Ku_i\}_i \subset Y$ is linearly independent in Y

(i) + (ii) + (iii) imply that the minimizer $\bar{u} \in X$ is unique and sparse, with

$$\bar{u} = \sum_i c_i u_i$$

There exists a function $g : Ext(B) \times Ext(B) \to [0, \infty)$ such that

- Quadratic growth of $\bar{p}$ around u_i

$$1 - \langle \bar{p}, u \rangle \gtrapprox g(u, u_i)^2 \quad \text{for every } i$$

- Lipschitz growth of K around u_i

$$\|K(u - u_i)\|_Y \lesssim g(u, u_i) \quad \text{for every } i$$

Theorem (Bredies, Carioni, Fanzon, Walter (2021))

Under the previous assumptions the rate of convergence of u^n is linear, i.e. there exists $C > 0$, $\zeta \in [1/2, 1)$ s.t.

$$G(u^n) - \min_u G(u) \leq C\zeta^n$$

Lifting to the space of measures

Proof Strategy: **lift** the problem (and algorithm) to the space of measures on extremal points and prove convergence in the lifted space

Denote $\mathcal{B} := \overline{Ext(B)}^*$ endowed with the metric that metrizes the weak* convergence. We consider positive measures on $\mathcal{B}$, i.e.

$$\mathcal{M}^+(\mathcal{B})$$

Definition

We say that a measure $\mu \in \mathcal{M}^+(\mathcal{B})$ represents $u \in X$ if

$$\langle p, u \rangle = \int_{\mathcal{B}} \langle p, v \rangle \, d\mu(v) \quad \forall p \in X_*$$

Alternatively, u is said to be the weak barycenter of μ

Example: the measure δ_u represents u for all $u \in \mathcal{B}$

As a consequence of Choquet's Theorem we have:

Proposition (Bredies, Carioni, Fanzon, Walter)

- *For every $\mu \in \mathcal{M}^+(\mathcal{B})$ we have*

$$R(u) \leq \|\mu\|_{\mathcal{M}(\mathcal{B})}$$

  where $u \in dom(R)$ is the weak barycenter of μ
- *For every $u \in dom(R)$ there exists $\mu \in \mathcal{M}^+(\mathcal{B})$ concentrated on $Ext(B)$ that represents u and such that*

$$R(u) = \|\mu\|_{\mathcal{M}(\mathcal{B})}$$

Lifted variational problem:

$$\min_{\mu \in \mathcal{M}^+(\mathcal{B})} \hat{G}(\mu) := F(\mathcal{K}\mu) + \|\mu\|_{\mathcal{M}(\mathcal{B})}$$

$\mathcal{K} : \mathcal{M}^+(\mathcal{B}) \to Y$ is the lift of $K : X \to Y$ defined by the relation

$$(\mathcal{K}\mu, y)_Y = \int_{\mathcal{B}} (Kv, y)\, d\mu(v) \quad \forall y \in Y$$

The following problems are **equivalent**

$$\min_{u \in X} G(u) := F(Ku) + R(u) \qquad \text{(original problem)} \qquad \text{(OP)}$$

$$\min_{\mu \in \mathcal{M}^+(\mathcal{B})} \hat{G}(\mu) := F(\mathcal{K}\mu) + \|\mu\|_{\mathcal{M}(\mathcal{B})} \qquad \text{(lifted problem)} \qquad \text{(LP)}$$

Theorem (Bredies, Carioni, Fanzon, Walter)

- $\bar{u} \in X$ *solves* (OP) $\implies \exists\, \bar{\mu} \in \mathcal{M}^+(\mathcal{B})$ *that represents* $\bar{u}$ *and solves* (LP)
- $\bar{\mu} \in \mathcal{M}^+(\mathcal{B})$ *solves* (LP) $\implies$ *weak barycenter of* $\bar{\mu}$ *solves* (OP)

Idea: GCG algorithm for (LP) $\implies$ GCG algorithm for (OP) taking barycenters:

$$\mu = \sum_i c_i \, \delta_{u_i} \in \mathcal{M}^+(\mathcal{B}) \qquad \rightsquigarrow \qquad u = \sum_i c_i \, u_i \in X$$

$\Rightarrow$ Prove linear convergence for the Lifted Problem

Dynamic inverse problems

Inverse Problem: Given $f_t \in H_t$, find a curve $t \mapsto \rho_t \in \mathcal{M}(\Omega)$ s.t.

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

- **Unknown:** curve of measures $t \mapsto \rho_t \in \mathcal{M}(\Omega)$, with $\Omega \subset \mathbb{R}^d$ bounded
- **Data**: curve $t \mapsto f_t \in H_t$ with $\{H_t\}_t$ family of Hilbert spaces
- **Measurements:** linear continuous operators $K_t^* \colon \mathcal{M}(\Omega) \to H_t$

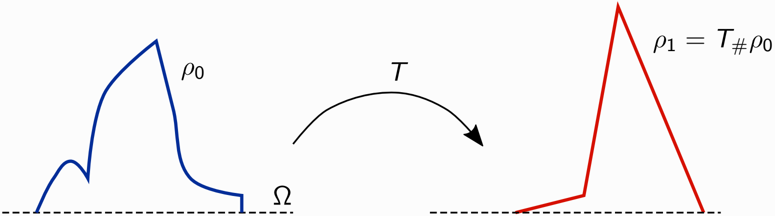
Assumptions: very weak time-regularity for $\{H_t\}_t$ and K_t^*

Goal: Regularize (P) with an Optimal Transport energy acting on the measure ρ_t
This will enforce time regularity of the reconstructions

An optimal transport approach for solving dynamic inverse problems in spaces of measures.
K. Bredies, S. Fanzon. ESAIM: M2AN (2020)

Optimal Transport - Static Formulation

$\Omega \subset \mathbb{R}^d$ bounded domain, $\rho_0, \rho_1 \in \mathcal{P}(\Omega)$, $T\colon \Omega \to \Omega$ measurable displacement



Goal: move ρ_0 to ρ_1 in the cheapest way, with cost of moving mass from x to y

$$c(x,y) := |x-y|^2$$

Optimal Transport: a transport plan $\hat{T}$ solving

$$\hat{T} \in \arg\min \left\{ \int_\Omega |T(x)-x|^2 \, d\rho_0(x) : \ T\colon \Omega \to \Omega, \ T_{\#}\rho_0 = \rho_1 \right\}$$

Optimal Transport - Dynamic Formulation

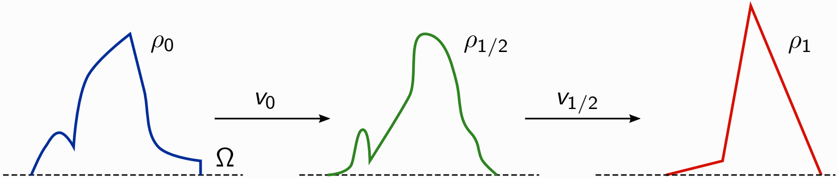
Idea: introduce a time variable $t \in [0,1]$ and consider the density evolution

- time dependent probability measures

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

- ρ_t is advected by the velocity field

$$v_t(x) \colon [0,1] \times \Omega \to \mathbb{R}^d$$



Dynamic model: (ρ_t, v_t) solves the continuity equation with initial conditions

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

Connection and Advantages

Theorem (Benamou-Brenier '00)

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

Advantages of Dynamic Formulation:

1. By introducing the momentum $m_t := \rho_t v_t$ we have

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

   which is **convex** in (ρ_t, m_t)

2. The continuity equation becomes **linear**

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

3. We know the full trajectory ρ_t and can recover the velocity field v_t from m_t

Optimal transport regularization

Time-space domain $X := (0,1) \times \Omega$, measures $\mathcal{M} := \mathcal{M}(X) \times \mathcal{M}(X; \mathbb{R}^d)$

We regularize $K_t^* \rho_t = f_t$ via minimization in $(\rho, m) \in \mathcal{M}$ of

$$G_{\alpha,\beta}(\rho, m) := \frac{1}{2} \int_0^1 \|K_t^* \rho_t - f_t\|_{H_t}^2 \, dt + J_{\alpha,\beta}(\rho, m)$$

where the optimal transport regularizer is

$$J_{\alpha,\beta}(\rho, m) := \underbrace{\frac{\alpha}{2} \int_0^1 \int_\Omega \left| \frac{dm}{d\rho} \right|^2 d\rho(t,x)}_{\text{Optimal Transport Regularizer}} + \beta \underbrace{\|\rho\|_{\mathcal{M}(X)}}_{\text{TV Regularizer}}$$

s.t. $\partial_t \rho_t + \operatorname{div} m_t = 0$ (Continuity Equation - No IC)

Existence and Sparsity

Theorem (Bredies, Fanzon '20)

(With assumptions on f_t, K_t^*, H_t) The functional $G_{\alpha,\beta}$ **admits solution**

$$\rho = dt \otimes \rho_t \,, \;\; m = v\rho \,, \;\; v \colon X \to \mathbb{R}^d$$

with v measurable **velocity field** *and $t \mapsto \rho_t \in \mathcal{M}^+(\Omega)$* **narrowly continuous**.
Moreover we have **stability**

Definition: An **atom** in $\mathcal{M}$ is a pair (ρ_γ, m_γ) with $\gamma \in H^1([0,1];\Omega)$,

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

Generalized Conditional Gradient Method

Theorem (Bredies, Carioni, Fanzon, Romero '20)

Consider the sublevel set $C_{\alpha,\beta} := \{J_{\alpha,\beta} \leq 1\}$. The **extremal points** *of $C_{\alpha,\beta}$ are*

$$\mathrm{Ext}(C_{\alpha,\beta}) = \{ \text{ atoms } \} \cup (0,0)$$

Goal: Find numerical solutions to the minimization problem for $G_{\alpha,\beta}$ by GCG

Key Step: Find a descent direction around $(\tilde{\rho}, \tilde{m})$ by solving

$$\min_{(\rho,m)\in C_{\alpha,\beta}} -\int_0^1 \langle \rho_t, w_t \rangle \, dt \,, \quad w_t := -K_t(K_t^* \tilde{\rho}_t - f_t) \in C(\Omega) \tag{D}$$

Theorem (SF, Bredies, Carioni, Romero '20)

Problem (D) admits a solution which is either an **atom** *or $(0,0)$.*

Therefore (D) can be casted in $H^1([0,1];\Omega)$, and is hence numerically feasible

Numerical Algorithm

Let $t \mapsto f_t$ be given data. Initialize $\rho^0 := 0$

1. **Insertion:** Given $\rho^n = \sum_{i=1}^N c_i \, \rho_{\gamma_i}$, set $w_t^n := -K_t(K_t^* \rho_t^n - f_t)$ and find

$$\gamma^* \in \operatorname*{arg\,min}_{\gamma \in H^1} -a_\gamma \int_0^1 w_t(\gamma(t)) \, dt \,, \qquad \rho^{n+1/2} := \rho^n + c_{N+1} \, \rho_{\gamma^*}$$

2. **Coefficients Optimization:** Solve the quadratic problem

$$(c_j^*)_j \in \operatorname*{arg\,min}_{c_j \geq 0} \;\; G_{\alpha,\beta}(\rho^{n+1/2}, m^{n+1/2}) \,, \qquad \rho^{n+1} := \sum_{i=1}^{N+1} c_i^* \, \rho_{\gamma_i}$$

Theorem (Bredies, Carioni, Fanzon, Romero '20)

The sequence (ρ^n, m^n) in Algorithm converges weak to a minimizer of $G_{\alpha,\beta}$*

The convergence rate is of order $O(1/n)$

Application: Undersampled Fourier Measurements

- $\Omega := [0,1]^2$ image frame, $t \mapsto \sigma_t \in \mathcal{M}^+(\mathbb{R}^2)$ frequencies sampling measure
- Fourier transform $\mathfrak{F} \colon \mathcal{M}(\Omega) \to C^\infty(\mathbb{R}^2; \mathbb{C})$
- $H_t := L^2_{\sigma_t}(\mathbb{R}^2; \mathbb{C})$ and $K_t^* \colon \mathcal{M}(\Omega) \to H_t$ defined by $K_t^* \rho := \mathfrak{F}\rho$

Note. K_t^* corresponds to the Fourier transform undersampled according to σ_t

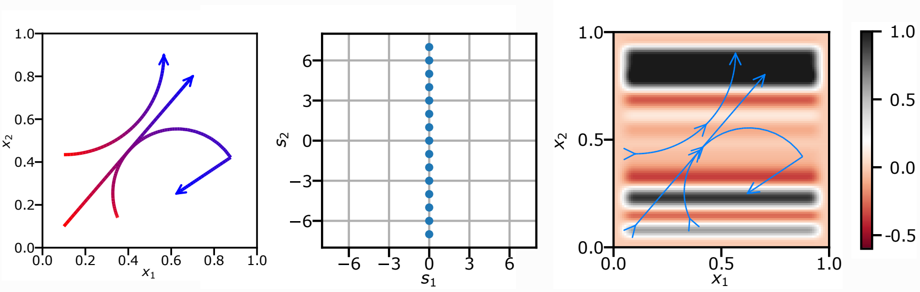
Time-discrete sampling: Fix $T+1$ times samples, $t_i := i/T$ for $i = 0, \ldots, T$

- At each time t_i sample $n_i \in \mathbb{N}$ frequencies $\{S_{i,1}, \ldots, S_{i,n_i}\} \subset \mathbb{R}^2$
- Define $t \mapsto \sigma_t$ so that $\sigma_{t_i} = \sum_{k=1}^{n_i} \delta_{S_{i,k}}$. In this case $H_{t_i} = \mathbb{C}^{n_i}$ and

$$K_{t_i}^* \rho = \left(\int_{\mathbb{R}^2} \exp(-2\pi \mathbf{i} x \cdot S_{i,k}) \, d\rho(x) \right)_{k=1}^{n_i} \in \mathbb{C}^{n_i}$$

Experiment: Dynamic Spikes Tracking

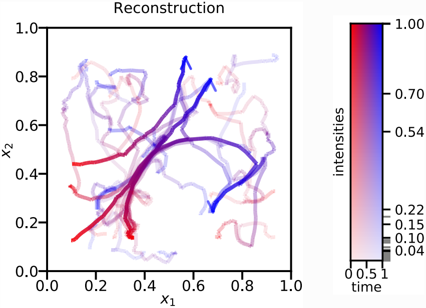
- $T = 50$, $n_i = 15$ freq. sampled on lines L_i through the origin with angle $\frac{i\pi}{4}$
- Ground Truth: $\tilde{\rho}_t = \delta_{\gamma_1(t)} + \delta_{\gamma_2(t)} + \delta_{\gamma_3(t)}$ as depicted (color=position in time)
- Synthetic Data: $f_{t_i} := K_{t_i}^* \tilde{\rho}_{t_i} + 60\%$ Gaussian Noise
- Data Visualization: By plotting the initial dual variable $w_{t_i}^0 := K_{t_i} f_{t_i} \in C(\Omega)$



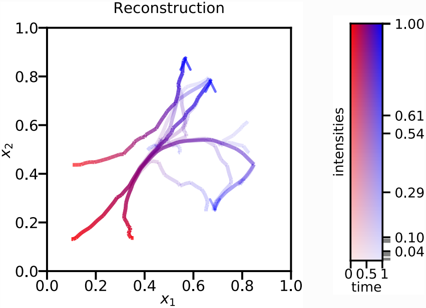
Ground Truth | Sampled Freq. $t = 1$ | $w_t^0(x)$, 60% noise

Reconstructions

$\alpha = \beta = 0.1$



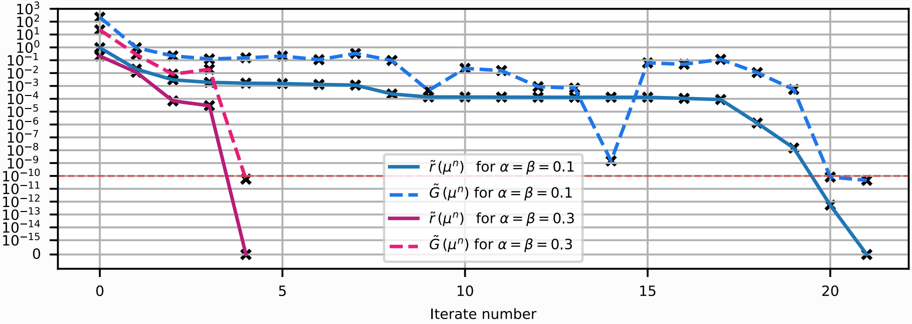
$\alpha = \beta = 0.3$



- **Low reg.** $\alpha, \beta = 0.1 \rightsquigarrow$ many low-energy artefacts around main trajectories
- **High reg.** $\alpha, \beta = 0.3 \rightsquigarrow$ improved reconstruction

Note! At each t_i the inverse problem $K_{t_i}^* \rho = f_{t_i}$ is heavily ill-posed: Indeed $K_{t_i}^* \delta_{\hat{x}} = K_{t_i}^* \delta_{\hat{x} + \lambda S_i^\perp}$ for $\lambda \in \mathbb{R}$, $S_i^\perp \perp L_i \rightsquigarrow$ Static methods cannot resolve location of $\hat{x}$

Convergence Plot



Note! Proven sublinear rate of convergence but empirical **linear rate**

As expected, higher regularization results in faster convergence

References

1. **Linear convergence of conditional gradient methods**
   - *Asymptotic linear convergence of Fully-Corrective Generalized Conditional Gradient methods*
     Preprint (2021), with K. Bredies, M. Carioni, D. Walter
2. **OT Regularization of Dynamic Inverse Problems**
   - *An optimal transport approach for solving dynamic inverse problems in spaces of measures*
     ESAIM: M2AN (2020), with K. Bredies
   - *A generalized conditional gradient method for dynamic inverse problems with optimal transport regularization*
     Found. of Comp. Math. (2021), with K. Bredies, M. Carioni, F. Romero
3. **Extremal Points of Transport Energies**
   - *On the extremal points of the ball of the Benamou-Brenier energy*
     Bull. London Math. Soc. (2021), with K. Bredies, M. Carioni, F. Romero
   - *A superposition principle for the inhomogeneous continuity equation with Hellinger-Kantorovich-regular coefficients*
     Communications in in PDE (2022), with K. Bredies, M. Carioni