Sparsity and convergence analysis of Generalized Conditional Gradient Methods

Silvio Fanzon

Based on joint works with

Kristian Bredies, Marcello Carioni, Francisco Romero, Daniel Walter

18 February 2022

Summary

- Review: Frank-Wolfe algorithms and generalized conditional gradient methods (GCG) in the space of measures
- GCG for arbitrary convex regularizers: sparsity and convergence rates
- Application: Dynamic inverse problems with Optimal Transport regularization
- Numerical simulations

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 is a convex function (regular enough)

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

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

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

- **Coefficient optimization step**: Obtain x^{n+1} by interpolating

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

  for a suitably chosen s^*

- 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

GCG are adapted to solve

$$\min_{u \in \mathcal{M}(\Omega)} G(u) = \min_{u \in \mathcal{M}(\Omega)} F(Au) + \|u\|_{\mathcal{M}(\Omega)}$$

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

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

- **Insertion step**: Solve the partially linearized problem in u^n as

$$\widetilde{u}^n \in \operatorname{argmin}_{\|u\|_{\mathcal{M}(\Omega)} \leq C} \langle A_* \nabla F(Au^n), u \rangle + \|u\|_{\mathcal{M}(\Omega)}$$

- **Coefficient optimization step**: u^{n+1} is obtained interpolating

$$u^{n+1} = u^n + s^*(\widetilde{u}^n - u^n)$$

  for a suitably chosen s^*

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 u^n can be constructed as a combination of Dirac deltas

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

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

This is a consequence of the next lemma. Define the dual variable at n-th iteration as

$$P^n := -A_* \nabla F(Au^n) \in C(\Omega)$$

Key Lemma

A solution to

$$\min_{\|u\|_{\mathcal{M}(\Omega)} \leq C} -\langle u, P^n \rangle + \|u\|_{\mathcal{M}(\Omega)}$$

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 u^{n+1} is obtained by adding to u^n a new Dirac delta $\delta_{\hat{x}}$ with

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

and adjusting coefficients

Theorem (Bredies, Pikkarainen (2013))

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

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

Remarks

- The GCG method for total variation regularization of measures takes advantage of the sparsity of the problem (iterates are Dirac deltas)
- It 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 coefficient optimization step can be improved by optimizing the coefficients $(c_1, \dots, c_{k+1})$ of the full linear combination

$$\sum_{i=1}^{k} c_i \delta_{x_i^n} + c_{k+1} \delta_{x_{k+1}^n}$$

with respect the energy G:

$$(c_1^*, \dots, c_{k+1}^*) \in \operatorname{argmin}_{c_i \in \mathbb{R}} \; G\left(\sum_{i=1}^{k} c_i \delta_{x_i^n} + c_{k+1} \delta_{x_{k+1}^n} \right)$$

(**This improved coefficient optimization step is needed for proving linear convergence. From now on we always consider this variant**)

Linear convergence in the space of measures

Is it possible to improve the sublinear convergence rate to linear?

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

$$\bar{P} = -A_* \nabla F(A\bar{u}) \in C(\Omega)$$

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 $\{x_i\}_i \subset \Omega$ such that $\operatorname{argmax}_x |\bar{P}(x)| = \{x_i\}_i$

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

(i) + (ii) + (iii) imply that the minimizer $\bar{u} \in \mathcal{M}(\Omega)$ is **unique** and **sparse**, i.e.

$$\bar{u} = \sum_i c_i \delta_{x_i} \tag{0.1}$$

[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)

- **Second order condition on the dual certificate:** $\bar{P}$ is C^2 and there exists $\theta > 0$ such that

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

- **Higher regularity of A:** There exists $C > 0$ such that around x_i

$$\|A(\delta_x - \delta_{x_i})\|_Y \leq C|x - x_i| \qquad \text{for all } \ i$$

Strong convexity of fidelity + Uniqueness and sparsity of the minimizer + Second order condition on the dual variable

Theorem (Pieper, Walter (2020))

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.

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

Generalized conditional gradient methods for convex regularizers

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

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

X separable Banach space with predual X_*, $F : Y \to \mathbb{R}$ with Y Hilbert space
$R : X \to [0, \infty]$ **convex**, 1-**homogeneous** and **coercive**

The structure of the algorithm is unchanged

- **Insertion step**: Solve the partially linearized problem in u^n as

$$\widetilde{u}^n \in \operatorname{argmin}_{R(u) \leq M} \; -\langle P^n, u \rangle + R(u)$$

  with $P^n := -A_* \nabla F(Au^n) \in X_*$ dual variable

- **Coefficient optimization step**: suitable, we see it later

---

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?

Key Lemma

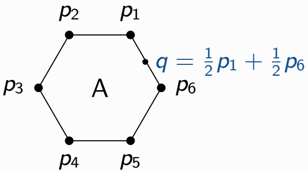
Define the unit ball of the regularizer $B := \{u \in X : \ R(u) \leq 1\}$. A solution of

$$\min_{R(u) \leq M} -\langle u, P^n \rangle + R(u)$$

is given by $c\overline{v}$ for some $c \in \mathbb{R}$ and $\overline{v} \in Ext(B)$

$Ext(A)$ is the set of extremal points of A: $u \in A$ such that

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



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

The algorithm

Define

$$B = \{u \in X : R(u) \leq 1\}$$

The iterates are then of the form

$$u^n = \sum_i c_i u_i^n$$

where $c_i \in \mathbb{R}$ and $u_i^n \in Ext(B)$. We compute u^{n+1} by solving

- **Insertion step:** Find a solution to

$$\widetilde{u}^n \in \operatorname{argmin}_{u \in Ext(B)} - \langle P^n, u \rangle$$

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

$$(c_1^*, \dots, c_{k+1}^*) \in \operatorname{argmin}_{c_i \in \mathbb{R}_+} G\left(\sum_{i=1}^k c_i u_i^n + c_{k+1} \widetilde{u}^n\right)$$

Next iterate is $u^{n+1} := \sum_{i=1}^k c_i^* u_i^n + c_{k+1}^* \widetilde{u}^n$

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

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

- We cover further coercive regularizers: dynamic optimal transport energies, group sparsity regularizers, PDE constrained inverse 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 (2020)

Linear convergence

Goal: Find suitable assumptions to prove linear convergence

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

$$\bar{P} := -A_* \nabla F(A\bar{u}) \in X_*$$

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. $\text{argmax}_{v \in \overline{Ext(B)}^*} \langle \bar{P}, v \rangle = \{u_i\}_i$

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

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

$$\bar{u} = \sum_i c_i u_i \tag{0.2}$$

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

- **Second order condition on the dual variable:** there exists a constant $\eta > 0$ such that around u_i

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

- **Higher regularity of A:** there exists $C > 0$ such that around u_i

$$\|A(u - u_i)\|_Y \leq C g(u, u_i) \quad \text{for every } i$$

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

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

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

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

Definition

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

$$\langle p, u \rangle = \int_W \langle p, v \rangle \, d\mu(v) \quad \forall p \in X_*$$

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

Example: the measure δ_u represents u for all $u \in W$

As a consequence of Choquet's theorem we have:

Proposition (Bredies, Carioni, Fanzon, Walter)

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

$$R(u) \leq \|\mu\|_{\mathcal{M}^+(W)}$$

where $u \in dom(R)$ is the weak barycenter of μ

- *For every $u \in dom(R)$ there exists $\mu \in \mathcal{M}^+(W)$ concentrated on $Ext(B)$ that represents u and such that*

$$R(u) = \|\mu\|_{\mathcal{M}^+(W)}$$

This proposition suggests to define the **lifted variational problem**

$$\min_{\mu \in \mathcal{M}^+(W)} \hat{G}(\mu) = \min_{\mu \in \mathcal{M}^+(W)} F(\hat{A}\mu) + \|\mu\|_{\mathcal{M}^+(W)}$$

where $\hat{A} : \mathcal{M}^+(W) \to Y$ is the lift of $A : X \to Y$ defined by the relation

$$(\hat{A}\mu, y)_Y = \int_W (Av, y)\, d\mu(v) \quad \forall y \in Y$$

It turns out that the original problem

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

and the lifted one

$$\min_{\mu \in \mathcal{M}^+(W)} \hat{G}(\mu) = \min_{\mu \in \mathcal{M}^+(W)} F(\hat{A}\mu) + \|\mu\|_{\mathcal{M}^+(W)} \tag{LP}$$

are equivalent

Theorem (Bredies, Carioni, Fanzon, Walter)

- *If $\overline{u} \in X$ is a solution of (P), then there exists $\overline{\mu} \in \mathcal{M}^+(W)$ that represents $\overline{u}$ and minimizes (LP)*
- *If $\overline{\mu} \in \mathcal{M}^+(W)$ is a solution of (LP), then the weak barycenter of $\overline{\mu}$ minimizes (P)*

Idea: We can construct a GCG algorithm for (LP) and obtain a GCG algorithm for (P) by taking the weak barycenters:

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

$\Rightarrow$ it is enough to prove linear convergence for the lifted problem

Dynamic variational inverse problems

Data: time dependent curve $t \mapsto f_t \in H_t$ with $\{H_t\}_t$ family of Hilbert spaces

Unknown: curve of measures $t \mapsto \rho_t \in \mathcal{M}(\Omega)$, with $\Omega \subset \mathbb{R}^d$ bounded

Forward operators: linear continuous operators $K_t^* \colon \mathcal{M}(\Omega) \to H_t$

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

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

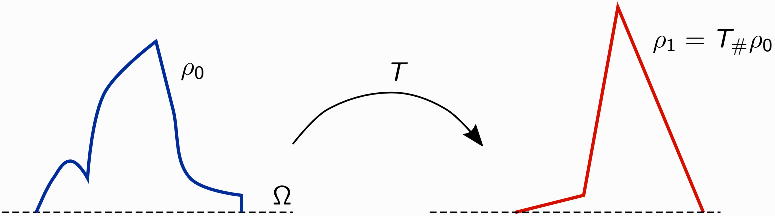
Assumptions: very weak time-regularity for $\{H_t\}_t$ and K_t^*

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 evolution of ρ_t

- time dependent probability measures

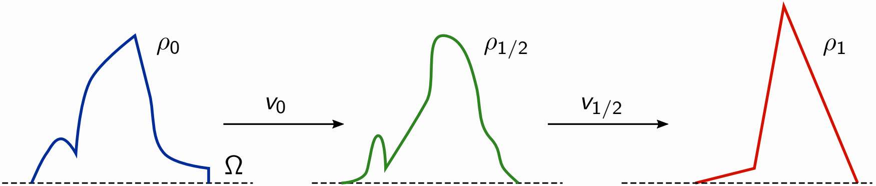
$$t \mapsto \rho_t \in \mathcal{P}(\Omega) \;\; \text{for} \;\; t \in [0,1]$$

- velocity field advecting ρ_t

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

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

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



Connection and Advantages

Theorem (Benamou-Brenier '00)

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

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). The continuity equation becomes **linear**

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

2. we know the full trajectory ρ_t and can recover the velocity field v_t from m_t

Optimal transport regularization

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

We propose to regularize (P) 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 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)

Theorem (Bredies, Fanzon '20)

(With assumptions on f_t, K_t^*, H_t) The functional $G_{\alpha,\beta}$ **admits a solution** *$\rho = dt \otimes \rho_t$, $m = v\rho$ with $v \colon X \to \mathbb{R}^d$ measurable* **velocity field** *and $t \mapsto \rho_t \in \mathcal{M}^+(\Omega)$* **narrowly continuous**. *Moreover we have* **stability**

Towards a numerical algorithm: Sparsity

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

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

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*

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

Generalized Conditional Gradient Method

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

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

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

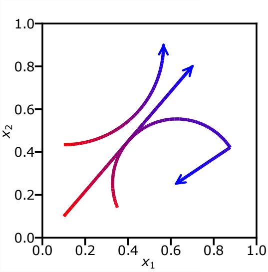
Time-discrete sampling: Fix $T + 1$ times samples, $t_i := i/T$ for $i = 0, \ldots, T$

- 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

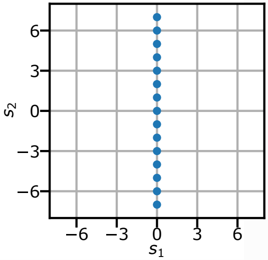
$$K_{t_i}^* \rho = \left(\int_{\mathbb{R}^2} \exp(-2\pi \mathbf{i} x \cdot S_{i,k}) \, d\rho(x) \right)_{k=1}^{n_i} \in \mathbb{C}^{n_i}$$

Experiment: Dynamic Spikes Tracking

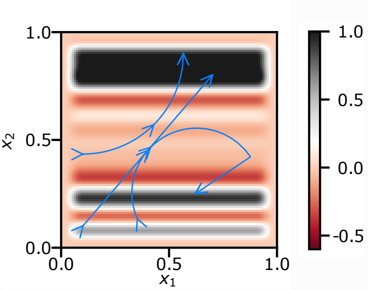
- $T = 50$, $n_i = 15$ freq. sampled on lines L_i through the origin with angle $\frac{i\pi}{4}$
- Ground Truth: $\tilde{\rho}_t = \delta_{\gamma_1(t)} + \delta_{\gamma_2(t)} + \delta_{\gamma_3(t)}$ as depicted (color=position in time)
- Synthetic Data: $f_{t_i} := K^*_{t_i}\tilde{\rho}_{t_i} + 60\%$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



Ground Truth



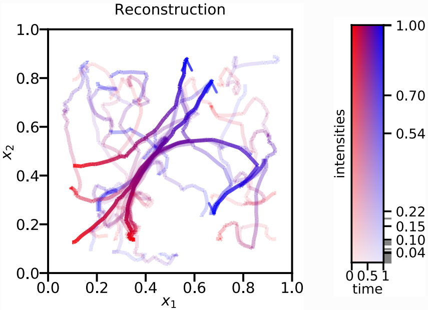
Sampled Freq. $t = 1$



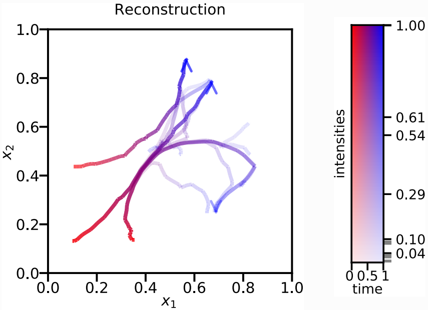
$w^0_t(x)$, 60% noise

Reconstructions

$\alpha = \beta = 0.1$



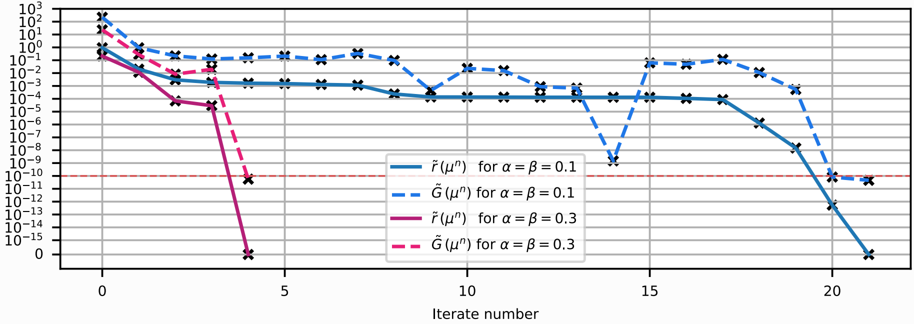
$\alpha = \beta = 0.3$



- **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*
     Comm. in PDE (2022), with K. Bredies, M. Carioni