Sparsity and convergence analysis of Generalized Conditional Gradient Methods

Silvio Fanzon

University of Graz

Based on joint works with

Kristian Bredies, Marcello Carioni, Francisco Romero, Daniel Walter

3 November 2022

Frank-Wolfe Algorithm

Consider the minimization problem

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

- $F \colon \mathbb{R}^N \to \mathbb{R}$ is regular convex function
- $C \subset \mathbb{R}^N$ convex set

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

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

$$\widehat{x} \in \arg\min_{x \in C} \; \langle \nabla F(x^n), x \rangle$$

- **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^*

Frank-Wolfe bibliographical notes

- Convergence rate of Frank-Wolfe is typically **sublinear**
- It can be improved to **linear** under strong convexity assumptions on F and different interpolation steps [1] [2].
- The algorithm has been generalized to infinite dimensional spaces called **Generalized Conditional Gradient methods** [3]
- Classical algorithms in infinite dimensional optimization are 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

BLASSO Problem:

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

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

GCG Algorithm

Given u^n compute next iterate u^{n+1} in two steps:

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

$$\widehat{u} \in \arg\max_{\|u\| \leq C} \langle p^n, u \rangle \,, \quad p^n := -K_* \nabla F(Ku^n) \in C(\Omega)$$

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

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

Sparse Iterations

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

$$u^n = \sum_{i=1}^{k_n} c_i \, \delta_{x_i} \,, \quad c_i \in \mathbb{R} \,, \quad x_i \in \Omega$$

Why? Key Lemma: Let $p \in C(\Omega)$. Then $\exists \, \hat{x} \in \Omega, c \in \mathbb{R}$ s.t.

$$\hat{x} \in \arg\max_{x \in \Omega} \; |p(x)| \quad \text{and} \quad c \, \delta_{\hat{x}} \in \arg\max_{\|u\| \leq C} \langle u, p \rangle$$

Therefore

$$\text{Insertion Step} \quad \Longleftrightarrow \quad \text{Maximizing } \; |p^n|$$

For later: Dirac deltas $\pm\delta_x$ are extremal points of the set

$$\{ u \in \mathcal{M}(\Omega) : \; \|u\| \leq 1 \}$$

Convergence result

Remark: Next iterate u^{n+1} can be obtained by

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

where

$$\widehat{u} = c\,\delta_{\hat{x}}\,, \quad \hat{x} \in \operatorname*{arg\,max}_{x\in\Omega} |p^n(x)|$$

Theorem (Bredies, Pikkarainen (2013))

GCG Algorithm generates a **sparse** *sequence*

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

such that

- $u^n \overset{*}{\rightharpoonup} u$ *with* u *minimizer of* G,
- *The rate of convergence is* **sublinear**, *i.e.*

$$G(u^n) - \min_{u\in\mathcal{M}(\Omega)} G(u) \lesssim \frac{1}{n}$$

Remarks on BLASSO

- GCG for BLASSO exploits **sparsity** of the problem (iterates are linear combinations of Dirac deltas)
- GCG allows to design a **discretization-free algorithm**

Faster convergence

Line Search Step can be replaced by a **Coefficients Optimization Step**: solve

$$(c_1^*, \dots, c_{k_n+1}^*) \in \operatorname*{arg\,min}_{c_i \geq 0} \; G\left(\sum_{i=1}^{k_n} c_i \, \delta_{x_i} + c_{k+1} \, \delta_{\hat{x}} \right)$$

and set the next iterate to

$$u^{n+1} := \sum_{i=1}^{k_n} c_i^* \, \delta_{x_i} + c_{k+1}^* \, \delta_{\hat{x}}$$

Important: Coefficient optimization $\rightsquigarrow$ **Linear** rate of convergence in experiments

GCG methods in Banach spaces

Question: Is it possible prove linear convergence in some cases?

Abstract Problem:

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

- **Unknown:** X separable Banach space with predual X_*
- **Data:** Y Hilbert space
- **Measurement:** $K \colon X \to Y$ linear, weak*-to-strong continuous
- **Data Fidelity:** $F : Y \to \mathbb{R}$ smooth, strictly convex
- **Regularizer:** $R \colon X \to [0, \infty]$ **convex**, **1-homogeneous** and **coercive**

Remark: Above assumptions guarantee existence of minimizer

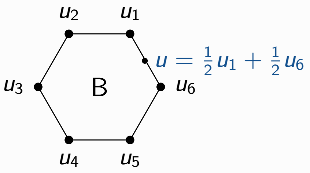
Asymptotic linear convergence of Fully-Corrective Generalized Conditional Gradient Methods
K. Bredies, M. Carioni, S. Fanzon, D. Walter. Mathematical Programming (2023)

Sparsity: Extremal points

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

$$\arg\max_{u \in B} \langle u, p \rangle = \arg\max_{u \in \text{Ext}(B)} \langle u, p \rangle$$

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

Fully-Corrective GCG algorithm

Assume the n-th iterate is sparse

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

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

- **Insertion step:** Compute dual variable $p^n := -K_* \nabla F(Ku^n) \in X_*$. Find

$$\widehat{u} \in \arg\max_{u \in B} \, \langle p^n, u \rangle \quad \text{s.t.} \quad \widehat{u} \in \mathrm{Ext}(B)$$

- **Coefficients Optimization:** Set $u_{k+1} := \widehat{u}$. Solve finite-dimensional problem

$$(c_1^*, \dots, c_{k+1}^*) \in \arg\min_{c_i \geq 0} \left[F\left(\sum_{i=1}^{k+1} c_i K u_i \right) + \sum_{i=1}^{k+1} c_i \right]$$

Next iterate is

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

Why sparse solutions?

Motivations:

- The **complexity** of the Insertion Step problem

$$\max_{u \in B} \langle p^n, u \rangle$$

  can be substantially reduced by looking at solutions in $\mathrm{Ext}(B)$

- We obtain a **discretization-free** algorithm in Banach space
- **Sparse solutions** are preferred in certain applications
- Recent **Representer Theorems** show that, under suitable assumptions, the minimization problem considered

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

  admits **sparse solutions**, i.e., finite linear combinations of points in $\mathrm{Ext}(B)$ [6]

---

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

Applications

Some classes of examples:

- The regularizer R could be a **norm** or **semi-norm**
- For $X = \mathcal{M}(\Omega)$ and $R(\mu) = \|\mu\|$ we recover GCG for **BLASSO**, since

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

- $R =$ **gauge function** of $\mathcal{M} \subset X$ weak* compact and convex:

$$R(u) := \inf\{\rho \geq 0 : \ u \in \rho\mathcal{M}\}$$

(dictionary learning, matrix completion)

Convergence rates

Global convergence: we first establish a basic convergence result for the FC-GCG

Theorem (Global convergence)

Consider the iterate generated by FC-GCG Algorithm:

$$u^n = \sum_{i=1}^{k_n} c_i \, u_i \,, \quad c_i \geq 0, \quad u_i \in \mathrm{Ext(B)}$$

We have

- $u^n \overset{*}{\rightharpoonup} u$ *with* u *minimizer of* G,
- *The convergence rate is* **sublinear**, *i.e.,*

$$G(u^n) - \min_{u \in X} G(u) \leq \frac{C}{n}$$

Linear convergence

Goal: Find additional assumptions under which linear convergence holds

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

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

Sparsity Assumptions:

(F1) F is strongly convex

(F2) There exists $\mathcal{A} = \{\bar{u}_i\}_{i=1}^N \subset \mathrm{Ext}(B)$ such that

$$\arg\max_{v \in B} \langle \bar{p}, v \rangle = \mathcal{A} = \{\bar{u}_i\}_{i=1}^N$$

(F3) The set $\{K\bar{u}_i\}_i \subset Y$ is linearly independent in Y

Theorem. (F1) + (F2) + (F3) $\implies$ the minimizer $\bar{u} \in X$ is **unique** and **sparse**:

$$\bar{u} = \sum_{i=1}^N \bar{c}_i \bar{u}_i \,, \quad \bar{c}_i > 0 \,, \quad \bar{u}_i \in \mathcal{A}$$

Linear convergence

Growth Assumptions: There exists a "distance" function

$$g : \mathrm{Ext}(B) \times \mathrm{Ext}(B) \to [0, \infty)$$

such that

(F4) g-Quadratic growth of $\bar{p}$ around $\bar{u}_i$

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

(F5) g-Lipschitz growth of K around $\bar{u}_i$

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

Theorem (Linear convergence)

Assume (F1)–(F5). Let u^n be generated by FC-GCG. Then u^n converges **linearly**

$$G(u^n) - \min_{u \in X} G(u) \lesssim C\zeta^n\,, \quad \exists\, \zeta \in [1/2, 1)$$

Proof Idea: Lifting to the space of measures

Proof Strategy: Lift the problem and algorithm to the space

$$\mathcal{M}(\mathrm{Ext}(B))$$

Definition

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

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

When μ represents u we denote

$$u = \mathcal{I}(\mu)$$

Example: For all $u \in \mathrm{Ext}(B)$ we have

$$\mathcal{I}(\delta_u) = u$$

Question: How do we lift to $\mathcal{M}(\mathrm{Ext}(B))$?

Proposition

If $u \in dom(R)$ there exists $\mu \in \mathcal{M}^+(\mathrm{Ext}(B))$ s.t.

$$R(u) = \|\mu\|_{\mathcal{M}}\,, \quad u = \mathcal{I}(\mu)$$

Proposition

There exists $\mathcal{K} : \mathcal{M}(\mathrm{Ext}(B)) \to Y$ linear and weak-to-strong continuous s.t.*

$$\mathcal{K}\mu = Ku \quad \textit{whenever} \;\; \mathcal{I}(\mu) = u$$

Lifted variational problem:

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

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

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

Theorem

(OP) *and* (LP) *are* **equivalent**. *In particular*

$$\bar{\mu} \in \mathcal{M}^+(\mathrm{Ext}(B)) \quad \textit{solves} \quad \text{(LP)} \quad \Longrightarrow \quad \bar{u} := \mathcal{I}(\bar{\mu}) \quad \textit{solves} \quad \text{(OP)}$$

Proving convergence:

1. Formulate FC-GCG Algorithm for (LP)
2. Prove linear convergence rate for such Algorithm
3. Obtain linear convergence rate for original Algorithm, since

$$\mu^n = \sum_i c_i \, \delta_{u_i} \quad \rightsquigarrow \quad u^n := \mathcal{I}(\mu^n) = \sum_i c_i \, u_i \quad \text{and} \quad \hat{G}(\mu^n) = G(u^n)$$

Example 1: Linear convergence for BLASSO

BLASSO Problem: $X = \mathcal{M}(\Omega)$ with $\Omega \subset \mathbb{R}^d$ bounded

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

Extremal points: For the unit ball $B = \{\|u\|_{\mathcal{M}(\Omega)} \leq 1\}$ we have

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

Dual variable: Let $\bar{u}$ be a minimizer of G. The dual variable is

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

Sparsity Assumptions:

(B1) F is strongly convex = **(F1)**

(B2) There exist $\mathcal{A} := \{x_i\}_{i=1}^N \subset \Omega$ such that = **(F2)**

$$\arg\max_{x \in \Omega} |\bar{p}(x)| = \{x_i\}_{i=1}^N = \mathcal{A}$$

(B3) The set $\{K\delta_{x_i}\}_i \subset Y$ is linearly independent in Y = **(F3)**

Theorem: (B1) + (B2) + (B3) $\implies$ minimizer $\bar{u}$ of G is **unique** and **sparse**:

$$\bar{u} = \sum_{i=1}^{N} \bar{c}_i \, \delta_{x_i} \,, \quad \bar{c}_i > 0 \,, \quad x_i \in \mathcal{A}$$

Growth Assumptions: Suppose that $\bar{p} \in C^2(\Omega)$

(B4) **Non-degenerate curvature** of $\nabla^2 \bar{p}$ at x_i $\implies$ **(F4)**

$$-\mathrm{sign}(\bar{p}(x_i)) \, \langle \xi, \nabla^2 \bar{p}(x_i) \xi \rangle \gtrsim |\xi|^2 \qquad \text{for all } \; i \,, \quad \text{and } \; \xi \in \mathbb{R}^d$$

(B4) **Lipschitz growth** of K around δ_{x_i} $\implies$ **(F5)**

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

Theorem: Define $g \colon \mathrm{Ext}(B) \times \mathrm{Ext}(B) \to [0, \infty)$ by

$$g(s_1 \delta_{x_1}, s_2 \delta_{x_2}) := |s_1 - s_2| + |x_1 - x_2|$$

Then **(B1)-(B5)** imply **(F1)-(F5)** wrt g, and FC-GCG converges **linearly**

Numerical Experiment: Sparse source identification

Bibliographical Note: Similar assumptions for BLASSO were made in

- K. Pieper, D. Walter. ESAIM: COCV (2021)
- A. Flinth, F. De Gournay, P. Weiss. Mathematical Programming (2021)

Setting of Experiment: $\Omega = (0,1)^2$ spatial domain, $(0, T)$ time domain

Unknown: $X = \mathcal{M}(\Omega)$ space of **initial temperature** distributions

Data: $Y = L^2(\Omega)$ space of **final temperature** measurements at time T

Operator: $K\colon \mathcal{M}(\Omega) \to L^2(\Omega)$ maps u to $y(T)$ where $y\colon [0,1] \times \Omega \to \mathbb{R}$ solves

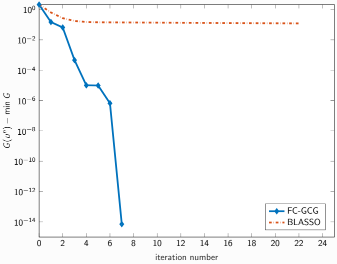
$$\begin{cases} \partial_t y - \Delta y = 0 & \text{in } (0,T) \times \Omega \\ y = 0 & \text{in } (0,T) \times \partial\Omega \\ y(0) = u & \text{in } \Omega \end{cases}$$

Problem: Given a final temperature distribution $y_d \in L^2(\Omega)$ find initial source

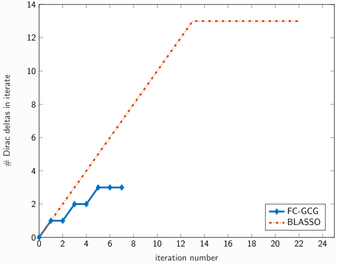
$$\hat{u} \in \arg\min_{u \in \mathcal{M}(\Omega)} \|y(T) - y_d\|^2_{L^2(\Omega)} + \|u\|_{\mathcal{M}(\Omega)}$$

Numerical Data: y_d corresponding to sparse source $u^\dagger$+ noise

$$u^\dagger := 25\delta_{x_1} - 10\delta_{x_2}\,, \quad x_1 = (0.75, 0.75)\,, \quad x_2 = (0.25, 0.25)$$



Residual at iteration k



Number of Diracs at iteration k

Remarks:

- FC-GCG is substantially faster then BLASSO
- FC-GCG correctly identifies 2 sources

  (Actually 3, but two of them are on adjacent grid points)

Example 2: Undersampled dynamic MRI

Task: Imaging via Magnetic Resonance

Mathematical Model: For the forward process

- $u\colon [0,1]^2 \to \mathbb{R}$ gray-scale image,
- $\mathfrak{F}$ denotes Fourier transform
- y Fourier data acquired by machine

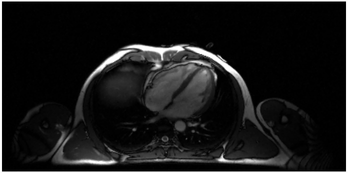
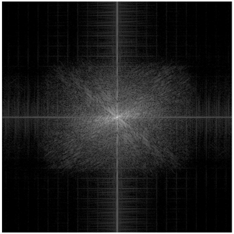


Image u

$\xrightarrow{\mathfrak{F}}$



$y = \mathfrak{F}u$

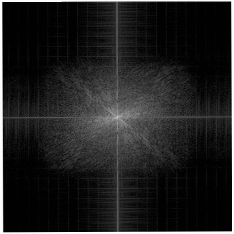
Example 2: Undersampled dynamic MRI

Inverse Problem: Given MRI data y, reconstruct u s.t.

$$\mathfrak{F}u = y$$

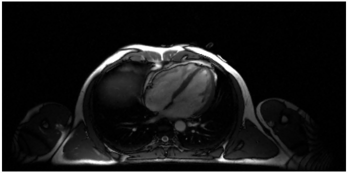
Ideal World: Easy! Just take

$$u = \mathfrak{F}^{-1}y$$



Data y

$$\xrightarrow{\mathfrak{F}^{-1}}$$



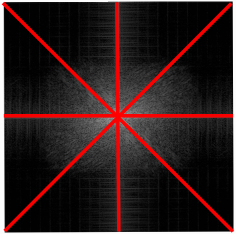
Reconstruction u

Example 2: Undersampled dynamic MRI

Reality: Things are not that straightforward:

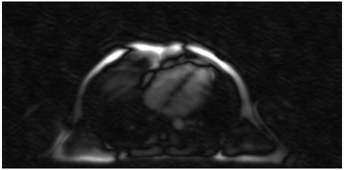
- Measurement process is inherently **noisy**
- **Limited sampling** in k-space, to limit scan time

Issue: Plain inversion leads to poor reconstructions



Undersampled noisy data y

$$\xrightarrow{\mathfrak{F}^{-1}}$$



Reconstruction u

Example 2: Undersampled dynamic MRI

Motion: Represents even bigger challenge to accurate reconstructions

- High resolution imaging
- Imaging moving organs

Dynamic inverse prob: Reconstruct movie u_t from undersampled data series y_t

$$\mathfrak{F}(u_t) = y_t \quad \text{for all} \quad t \in [0, 1]$$

Original movie

Bad Reconstruction

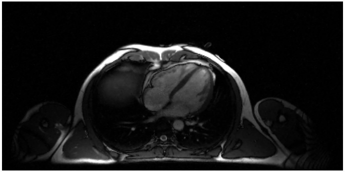
Example 2: Undersampled dynamic MRI

Motion: Represents even bigger challenge to accurate reconstructions

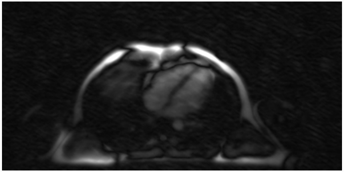
- High resolution imaging
- Imaging moving organs

Dynamic inverse prob: Reconstruct movie u_t from undersampled data series y_t

$$\mathfrak{F}(u_t) = y_t \quad \text{for all} \quad t \in [0,1]$$



Original movie



Bad Reconstruction

Solution: We need regularization for **dynamic inverse problems**

Dynamic inverse problems

Our setting:

- **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 $K_t^* \colon \mathcal{M}(\Omega) \to H_t$

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

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

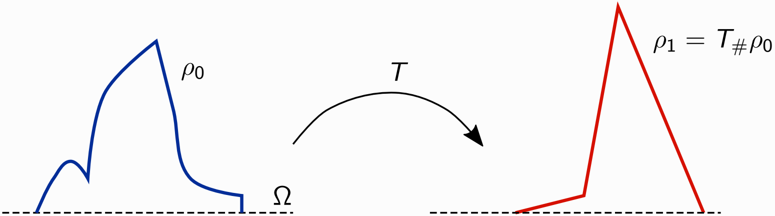
Assumptions: very weak time-regularity for $\{H_t\}_t$ and K_t^*

Proposal: Regularize (P) with an **Optimal Transport Energy** acting on ρ_t

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, $\rho_0, \rho_1 \in \mathcal{P}(\Omega)$, $T : \Omega \to \Omega$ 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

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

Optimal Transport - Dynamic Formulation

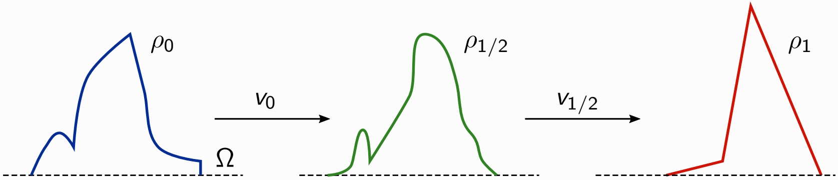
Idea: introduce a time variable $t \in [0, 1]$ and consider the density evolution

- 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

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

Benamou-Brenier Formula

Theorem (Benamou-Brenier '00)

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

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

Recall: We want to regularize the inverse problem

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

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

Regularization: Minimize in $(\rho, m) \in \mathcal{M}$ the functional

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

Optimal Transport Regularizer:

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

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

Existence and Sparsity

Theorem: (Assumptions on f_t, K_t^*, H_t) The functional $G_{\alpha,\beta}$ **admits minimizer**

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

with v measurable **velocity field** and $t \mapsto \rho_t \in \mathcal{M}^+(\Omega)$ **narrowly continuous**. Moreover we have **stability** for vanishing noise level and $\alpha, \beta \to 0$

Sparsity: In order to apply FC-GCG we need Extremal Points of $J_{\alpha,\beta}$

Atoms: pairs $(\rho_\gamma, m_\gamma) \in \mathcal{M}$ with $\gamma \in H^1([0,1];\Omega)$,

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

Theorem: Define unit ball $C_{\alpha,\beta} := \{ J_{\alpha,\beta} \le 1 \}$. The **extremal points** are

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

On the extremal points of the ball of the Benamou-Brenier energy.
K. Bredies, M. Carioni, S. Fanzon, F. Romero. Bull. London Math. Soc. (2021)

FC-GCG Method

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

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

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

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

K. Bredies, M. Carioni, S. Fanzon, F. Romero. Found. of Comp. Math. (2021)

FC-GCG Algorithm

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

$$\rho^n = \sum_{i=1}^{k} c_i \, \rho_{\gamma_i}$$

- **Insertion Step:** Set $w_t^n := -K_t(K_t^* \rho_t^n - f_t) \in C(\Omega)$ and find

$$\widehat{\gamma} \in \operatorname*{arg\,min}_{\gamma \in H^1} \; -a_\gamma \int_0^1 w_t^n(\gamma(t)) \, dt$$

- **Coefficients Optimization:** Set $\gamma_{k+1} := \widehat{\gamma}$. Solve the quadratic problem

$$(c_1^*, \dots, c_{k+1}^*) \in \operatorname*{arg\,min}_{c_j \geq 0} \; G_{\alpha,\beta}\left(\sum_{i=1}^{k+1} c_i \, \delta_{\gamma_i} \right)$$

The next iterate is

$$\rho^{n+1} := \sum_{i=1}^{k+1} c_i^* \, \rho_{\gamma_i}$$

Experiment: Tracking particles from Undersampled MRI

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

$$K_t^* := M_t \circ \mathfrak{F}$$

Note: At fixed t the inverse problem $K_t^*\rho = f_t$ is heavily ill-posed:

$$\sigma_t = \mathcal{H}^1 \llcorner L\,, \quad L \text{ line} \quad \Longrightarrow \quad K_t^*\delta_{\hat{x}} = K_t^*\delta_{\hat{x}+\lambda L^\perp} \quad \text{for} \quad \lambda \in \mathbb{R},\ L^\perp \perp L$$

$\rightsquigarrow$ Static methods cannot resolve location of $\hat{x}$

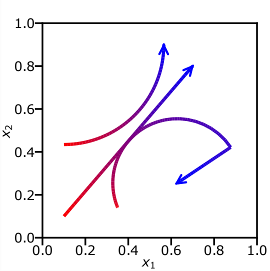
Time-discrete sampling: $T+1$ times samples, $t_i := i/T$ for $i = 0,\dots,T$

- At t_i sample $n_i \in \mathbb{N}$ frequencies $\{S_{i,1},\dots,S_{i,n_i}\} \subset \mathbb{R}^2$
- Sampling measure $\sigma_{t_i} = \sum_{k=1}^{n_i} \delta_{S_{i,k}}$. Then $H_{t_i} = \mathbb{C}^{n_i}$ and

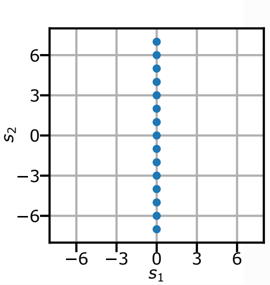
$$K_{t_i}^*\rho = \left(\int_{\mathbb{R}^2} \exp(-2\pi \mathbf{i} x \cdot S_{i,k})\, d\rho(x) \right)_{k=1}^{n_i} \in \mathbb{C}^{n_i}$$

Experiment: Dynamic Spikes Tracking

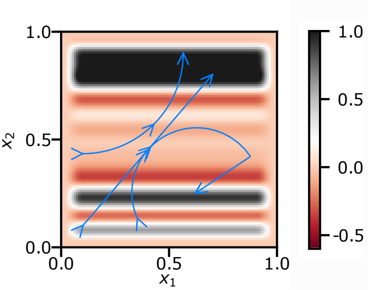
- L_i line through the origin with angle $\frac{i\pi}{4}$, $i \in \mathbb{N}$
- $T = 50$ time sample, $n_i = 15$ frequencies sampled on L_i
- Ground Truth: $\tilde{\rho}_t = \delta_{\gamma_1(t)} + \delta_{\gamma_2(t)} + \delta_{\gamma_3(t)}$ as depicted (color=position in time)
- Synthetic Data: $f_{t_i} := K_{t_i}^* \tilde{\rho}_{t_i} + 60\%$ 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



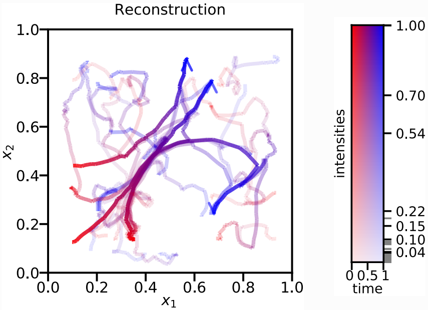
Sampled Freq. $t = 1$



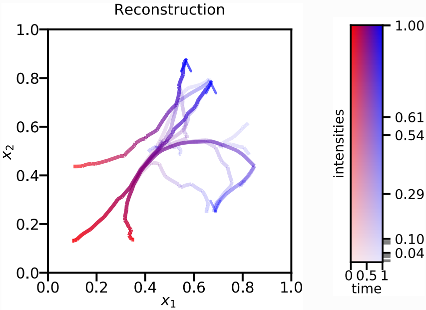
$w_t^0(x)$, 60% noise

Reconstructions

$\alpha = \beta = 0.1$

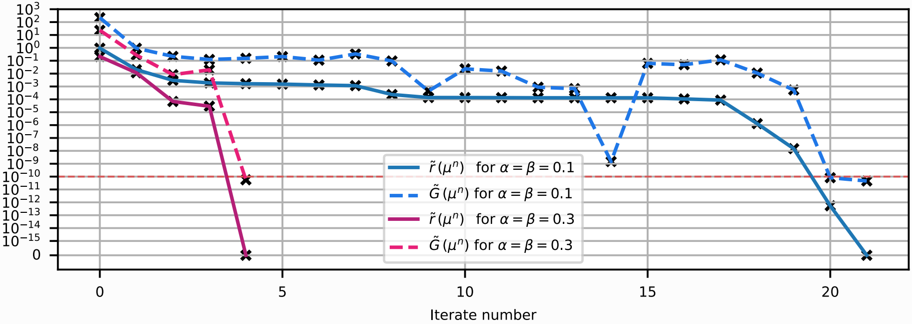


$\alpha = \beta = 0.3$



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

Convergence Plot



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

Linear convergence: proof is work in progress

References

1. **Linear convergence of conditional gradient methods**
   - *Asymptotic linear convergence of Fully-Corrective Generalized Conditional Gradient Methods*
     Mathematical Programming (2023), 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