Sparse optimization methods for infinite-dimensional variational problems

Marcello Carioni

University of Twente
Mathematics of Imaging and AI (MIA)

NUCLEI 2022, GSSI

FUNDED BY
THE ROYAL SOCIETY

Summary

- Sparsity in finite dimensions
- Sparsity for infinite dimensional problems with finite dimensional constraints – Representer theorems.
- Generalized conditional gradient methods – Infinite-dimensional Frank-Wolfe-type algorithms
- Applications and numerics – Dynamic inverse problems with Optimal transport regularizers.

Summary

- Sparsity in finite dimensions
- Sparsity for infinite dimensional problems with finite dimensional constraints – Representer theorems.
- Generalized conditional gradient methods – Infinite-dimensional Frank-Wolfe-type algorithms
- Applications and numerics – Dynamic inverse problems with Optimal transport regularizers.

A particular thanks to my collaborators: Kristian Bredies (KFU Graz), Silvio Fanzon (KFU Graz), Yury Korolev (University of Cambrige), Martin Holler (KFU Graz) , Francisco Romero (KFU Graz), Carola Schönlieb (University of Cambrige), Daniel Walter (RICAM Linz)

Sparsity in finite dimensions (compressed sensing)

Consider a linear operator

$$A : \mathbb{R}^n \to \mathbb{R}^k \qquad \text{for} \qquad k << n.$$

We aim to reconstruct the data $u \in \mathbb{R}^n$ given the observation $y \in \mathbb{R}^k$ solving

$$Au = y \qquad \text{for given} \qquad y \in \mathbb{R}^k \qquad \text{ill posed}$$

Need to impose extra constraints to select a specific solution.

[1] Stable Signal Recovery from Incomplete and Inaccurate Measurements E. Candes, J. Romberg, T. Tao (2005)

Sparsity in finite dimensions (compressed sensing)

Consider a linear operator

$$A : \mathbb{R}^n \to \mathbb{R}^k \qquad \text{for} \qquad k << n.$$

We aim to reconstruct the data $u \in \mathbb{R}^n$ given the observation $y \in \mathbb{R}^k$ solving

$$Au = y \qquad \text{for given} \qquad y \in \mathbb{R}^k \qquad \text{ill posed}$$

Need to impose extra constraints to select a specific solution.

Theory of compressed sensing [1] aims to select sparse solutions solving the inverse problem.

Definition (Sparse vectors)

$u \in \mathbb{R}^n$ is sparse if it has minimal $\|\cdot\|_0$ norm ($\|u\|_0 = \#\{u_i \neq 0\}$).

One could solve the variational problem

$$\min_{u \in \mathbb{R}^n} \|u\|_0 \quad \text{subjected to} \quad Au = y$$

[1] Stable Signal Recovery from Incomplete and Inaccurate Measurements E. Candes, J. Romberg, T. Tao (2005)

Unfortunately

The $\|\cdot\|_0$ norm is not suitable for optimization.

Unfortunately

The $\|\cdot\|_0$ norm is not suitable for optimization.

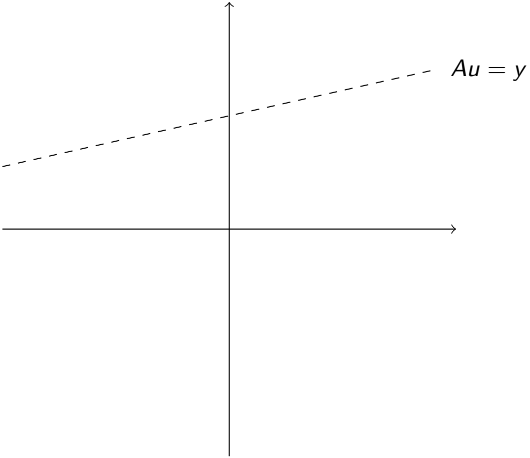
However, It is well-known (Candes-Tao-Romberg) that the $\|\cdot\|_0$ norm can be replaced by the 1-norm $\|\cdot\|_1$ and still leading to a sparse solution. One then solves

$$\min_{u \in \mathbb{R}^n} \|u\|_1 \quad \text{subjected to} \quad Au = y \qquad \textbf{(LASSO)}$$

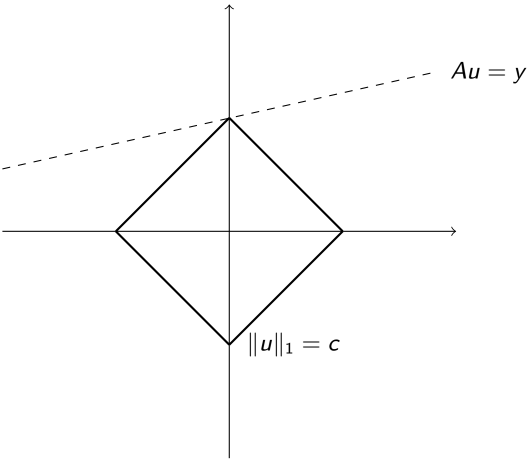
Under some assumptions on A the solution is still sparse.

$\|\cdot\|_1$ promotes sparsity

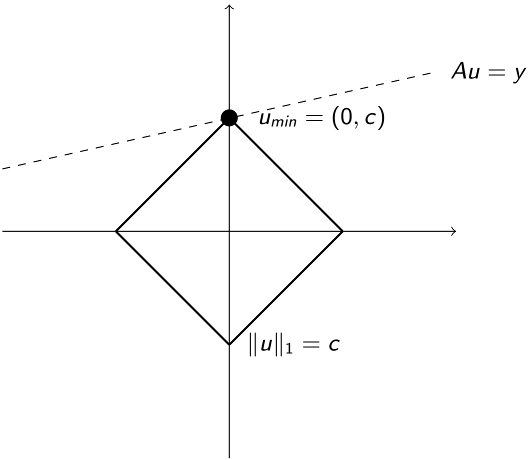
Why is $\|\cdot\|_1$ promoting sparsity? For example for a linear $A : \mathbb{R}^2 \to \mathbb{R}$:



Why is $\|\cdot\|_1$ promoting sparsity? For example for a linear $A : \mathbb{R}^2 \to \mathbb{R}$:



Why is $\|\cdot\|_1$ promoting sparsity? For example for a linear $A : \mathbb{R}^2 \to \mathbb{R}$:



$$\boxed{u_{min} = (0, c) \text{ is sparse}}$$

Sparsity for infinite dimensional inverse problems

How this theory transfers to infinite dimensional problems?

Heuristically: "A sparse solution is any solution that can be represented as finite combination of simpler objects (atoms)"

Sparsity for infinite dimensional inverse problems

How this theory transfers to infinite dimensional problems?

Heuristically: "A sparse solution is any solution that can be represented as finite combination of simpler objects (atoms)"

We consider an infinite dimensional variational problem with the crucial assumption that **the constraint is finite dimensional**.

$$\min_{u \in X} R(u) \quad \text{subjected to} \quad Au = y$$

where X is an infinite dimensional space and $A : X \to \mathbb{R}^k$ is linear and continuous and R is convex.

Sparsity for infinite dimensional inverse problems

How this theory transfers to infinite dimensional problems?

Heuristically: "A sparse solution is any solution that can be represented as finite combination of simpler objects (atoms)"

We consider an infinite dimensional variational problem with the crucial assumption that **the constraint is finite dimensional**.

$$\min_{u \in X} R(u) \quad \text{subjected to} \quad Au = y$$

where X is an infinite dimensional space and $A : X \to \mathbb{R}^k$ is linear and continuous and R is convex.

How can we make the heuristic definition of sparse solutions rigorous?

How R is promoting sparsity? Which is the role of the operator A and the data $y \in \mathbb{R}^k$?

Setting and general representer theorem

We consider the following variational problem

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

where X is a locally convex space, $A : X \to \mathbb{R}^k$ is linear and continuous, R is a seminorm and F is convex. (+ necessary hypotheses to ensure existence of minimizers).

Setting and general representer theorem

We consider the following variational problem

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

where X is a locally convex space, $A : X \to \mathbb{R}^k$ is linear and continuous, R is a seminorm and F is convex. (+ necessary hypotheses to ensure existence of minimizers).

- We recover the strong constraint $Au = y$ for

$$F(z) = \begin{cases} 0 & z = y \\ +\infty & otherwise. \end{cases}$$

- We treat general penalization terms like L^2-penalization.

On Representer Theorems and Convex Regularization C. Boyer, A. Chambolle, Y. De Castro, V. Duval, F. De Gournay, P. Weiss
Splines Are Universal Solutions of Linear Inverse Problems with Generalized TV Regularization M. Unser, J. Fageout, J.P. Ward

Theorem (Bredies, C., 2020)

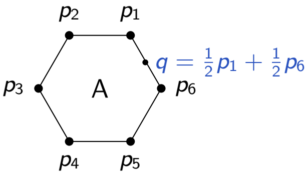
There exists $\overline{u} \in X$, *a solution of the variational problem such that:*

$$\overline{u} = \overline{\psi} + \sum_{i=1}^{p} \gamma_i u_i \,,$$

$\overline{\psi} \in Null(R)$, $p \leq k$, $\gamma_i > 0$ *and* $u_i \in Ext(\{u : R(u) \leq 1\} + Null(R))$.

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

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



Theorem (Bredies, C., 2020)

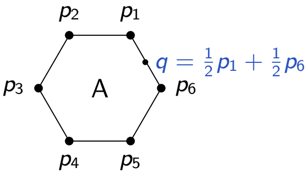
There exists $\overline{u} \in X$, a solution of the variational problem such that:

$$\overline{u} = \overline{\psi} + \sum_{i=1}^{p} \gamma_i u_i \, ,$$

$\overline{\psi} \in Null(R)$, $p \leq k$, $\gamma_i > 0$ *and* $u_i \in Ext\left(\{u : R(u) \leq 1\} + Null(R)\right)$.

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

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



The atoms are the extremal points of the unit ball of R.

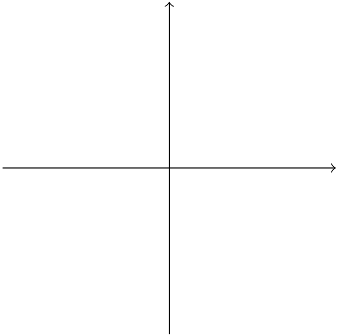
Sparsity of solutions for variational inverse problems with finite-dimensional data K. Bredies, M. Carioni. Calc. Var. and PDE, (2020)

Applications and revisitation of known results

To apply our abstract result is essential to

Characterize the extremal points of the ball of the regularizer

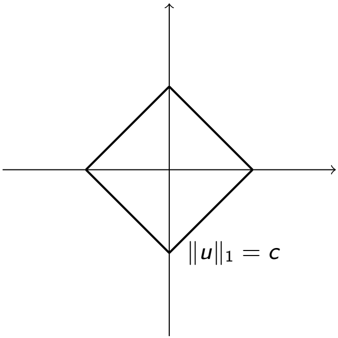
- **Finite dimensional LASSO**



To apply our abstract result is essential to

Characterize the extremal points of the ball of the regularizer

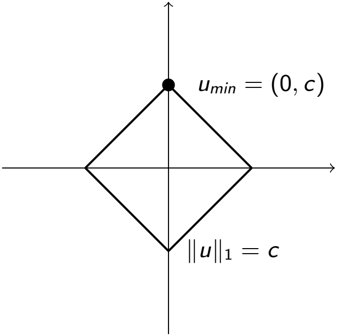
- **Finite dimensional LASSO**



To apply our abstract result is essential to

Characterize the extremal points of the ball of the regularizer

- **Finite dimensional LASSO**



- **Radon norm of measures**

With choices $X = \mathcal{M}(\Omega)$, $R(u) = \|u\|_{\mathcal{M}}$, $A : X \to \mathbb{R}^k$

$$\min_{u \in \mathcal{M}(\Omega)} \|u\|_{\mathcal{M}} \quad \text{subjected to} \quad Au = y$$

Then there exists a solution of the variational problem of the form

$$\overline{u} = \sum_{i=1}^{p} c_i \delta_{x_i}$$

where $p \leq k$ and $c_i \in \mathbb{R}$.

Lemma (Characterization of the extremal points for $\|\cdot\|_{\mathcal{M}}$)

The extremal points of the set

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

are $\{\pm\delta_x : x \in \Omega\}$

TV regularization for denoising

We choose $X = BV(\Omega)$ and $R(u) = TV(u)$ where

$$TV(u) := \sup\left\{ \int_\Omega u \,\mathrm{div}\varphi \, dx : \varphi \in C_c^1(\Omega),\ \|\varphi\|_\infty \leq 1 \right\} .$$

TV regularization for denoising

We choose $X = BV(\Omega)$ and $R(u) = TV(u)$ where

$$TV(u) := \sup\left\{ \int_\Omega u \,\mathrm{div}\varphi \, dx : \varphi \in C_c^1(\Omega),\ \|\varphi\|_\infty \leq 1 \right\}.$$

We solve the ROF model (Rudin-Osher-Fatemi, '92)

$$\min_{u \in BV(\Omega)} F(Au) + TV(u)$$

TV regularization for denoising

We choose $X = BV(\Omega)$ and $R(u) = TV(u)$ where

$$TV(u) := \sup \left\{ \int_\Omega u \, \mathrm{div}\varphi \, dx : \varphi \in C_c^1(\Omega), \ \|\varphi\|_\infty \leq 1 \right\} .$$

We solve the ROF model (Rudin-Osher-Fatemi, '92)

$$\min_{u \in BV(\Omega)} F(Au) + TV(u)$$

Theorem (Bredies, Carioni)

The extremal points of $\{u \in BV(\Omega) : TV(u) \leq 1\}$ *are*

$$\left\{ \pm \frac{\chi_E}{P(E,\Omega)} : E \subset \Omega \ \textit{simple} \right\} .$$

Here χ_E is the characteristic function of E and $P(E,\Omega)$ is the perimeter of E in Ω. (Ambrosio-Caselles-Morel-Masnou '99 for $\Omega = \mathbb{R}^n$).

Sparsity of solutions for variational inverse problems with finite-dimensional data K. Bredies, M. Carioni. Calc. Var. and PDE, (2020)
Connected components of sets of finite perimeter and applications to image processing L. Ambrosio, V. Caselles; S. Masnou, J.-M. Morel. JEMS, (2001)

Theorem (Bredies, C.)

There exists $\overline{u} \in BV(\Omega)$ *a solution of the variational problem such that*

$$\overline{u} = c + \sum_{i=1}^{p} \frac{\gamma_i}{P(E_i, \Omega)} \chi_{E_i} \,,$$

where $c \in \mathbb{R}$, $p \leq k$, $\gamma_i \in \mathbb{R}$ *and* $E_i \subset \Omega$ *are simple sets in* Ω.

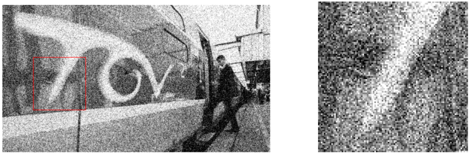
Theorem (Bredies, C.)

There exists $\overline{u} \in BV(\Omega)$ a solution of the variational problem such that

$$\overline{u} = c + \sum_{i=1}^{p} \frac{\gamma_i}{P(E_i,\Omega)} \chi_{E_i}\,,$$

where $c \in \mathbb{R}$, $p \leq k$, $\gamma_i \in \mathbb{R}$ and $E_i \subset \Omega$ are simple sets in Ω.

This is a theoretical justification of the staircase effect for TV-denoising



(a) Noisy Image

(b) ROF

Classical Frank-Wolfe-type algorithms aim at solving

$$\min_{x \in \mathbb{R}^N} F(x) + R(x)$$

where R is convex, non-smooth F is a convex, smooth function.

Classical Frank-Wolfe-type algorithms aim at solving

$$\min_{x \in \mathbb{R}^N} F(x) + R(x)$$

where R is convex, non-smooth F is a convex, smooth function. Given an iterate x^n one computes the next one x^{n+1} in two steps:

- Solve the partially linearized problem in x^n as

  $$\widetilde{x}^n \in \operatorname{argmin}_{x \in \mathbb{R}^N} \langle \nabla F(x^n), x \rangle + R(x)$$

  (**Insertion step**)
- Obtain x^{n+1} by interpolating

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

  for a suitably chosen s^*.
  (**Coefficients optimization step**)

- The convergence rate is typically sublinear and it can be improved to linear under strong convexity assumptions on F and better coefficient optimization steps [2] [3].
- Generalization to infinite dimensional spaces are called Generalized Conditional Gradient methods (GCG) [4]
- Classical algorithms in infinite dimensional optimization have been shown to be particular instances of GCG [5] [6]

[2] Revisiting Frank-Wolfe: Projection-Free Sparse Convex Optimization Jaggi, M. (2013)

[3] On the Global Linear Convergence of Frank-Wolfe Optimization Variants Lacoste-Julien, S. and Jaggi, M. (2015)

[4] Approximate methods in optimization problems Demyanov, V. F. and Rubinov A. M. (1970)

[5] An iterative thresholding algorithm for linear inverse problems with a sparsity constraint Daubechies, I., Defrise, M. and De Mol, C. (2004)

[6] Iterated hard shrinkage for minimization problems with sparsity constraints Bredies, K. Lorenz, D. (2006)

Generalized conditional gradient methods for convex regularizers

Goal: Generalize Frank-Wolfe algorithms in Banach spaces taking advantage of the sparsity of the problem to prove convergence results.

Generalized conditional gradient methods for convex regularizers

Goal: Generalize Frank-Wolfe algorithms in Banach spaces taking advantage of the sparsity of the problem to prove convergence results.

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

where X is a Banach space with predual X_*, $R : X \to [0, \infty]$ is convex, 1-homogeneous and coercive, F is convex and smooth and $A : X \to Y$ is linear and weak*-to-weak continuous.

Generalized conditional gradient methods for convex regularizers

Goal: Generalize Frank-Wolfe algorithms in Banach spaces taking advantage of the sparsity of the problem to prove convergence results.

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

where X is a Banach space with predual X_*, $R : X \to [0, \infty]$ is convex, 1-homogeneous and coercive, F is convex and smooth and $A : X \to Y$ is linear and weak*-to-weak continuous.

The structure of the algorithm is unchanged

- Solve the partially linearized problem in u^n as

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

  with $P^n = -A_* \nabla F(Au^n) \in X_*$ dual variable. (**Insertion step**)
- Obtaining u^{n+1} by interpolating $u^{n+1} = u^n + s^*(\widetilde{u}^n - u^n)$ for suitable s^* (**Coefficients optimization step**).

Linear convergence of accelerated generalized conditional gradient methods K. Bredies, M. Carioni, S. Fanzon, D. Walter (2021)

How sparsity is entering in the design of the conditional gradient method?

How sparsity is entering in the design of the conditional gradient method?

Lemma (Key lemma)

A solution of

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

is given by $c\overline{v}$ *for* $c \in \mathbb{R}$ *and* $\overline{v} \in Ext(\{u \in X : R(u) \le 1\})$.

A solution of the insertion step is an extremal point $\Rightarrow$ At each insertion step we can add a new extremal point of the ball of R.

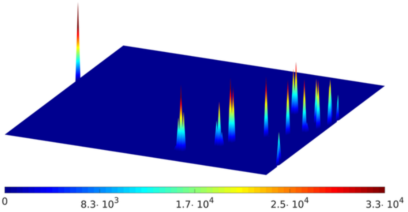


Figure from Controllability of the one-dimensional fractional heat equation under positivity constraints. U. Biccari, M. Warma and E. Zuazua (2019).

The atoms used to construct the iterates are the extremal points of the ball of R. In this sense the iterates are sparse.

The atoms used to construct the iterates are the extremal points of the ball of R. In this sense the iterates are sparse.

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)**

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

- **(Improved coefficients optimization step)**

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

So that $u^{n+1} = \sum_{i=1}^{k} c_i^* u_i^n + c_{k+1}^* \widetilde{u}^n$.

We obtain the worst-case convergence rate.

Theorem (Bredies, C., Fanzon, Walter (2021))

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

$$G(u^n) - \min_u G(u) \le Cn^{-1}$$

[7] Inverse problems in spaces of measures K. Bredies and H.-K. Pikkaranien. (2013)

[8] The alternating descent conditional gradient method for sparse inverse problems N. Boyd, G. Schiebinger, B. Recht (2015)

We obtain the worst-case convergence rate.

Theorem (Bredies, C., Fanzon, Walter (2021))

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

$$G(u^n) - \min_u G(u) \leq Cn^{-1}$$

- The algorithm is grid-free. No discretizations of X are needed.
- The notion of sparsity for infinite-dimensional problems is central
- If $R(u) = \|u\|_{\mathcal{M}}$ the algorithm is a classical (by now) grid-free sparse optimization algorithm [7][8] using

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

- The effectiveness of such algorithms relies on
  - The characterization of the extremal points of $\{R(u) \leq 1\}$.
  - The computational feasibility of the insertion step

[7] Inverse problems in spaces of measures K. Bredies and H.-K. Pikkaranien. (2013)

[8] The alternating descent conditional gradient method for sparse inverse problems N. Boyd, G. Schiebinger, B. Recht (2015)

Linear convergence

We now want to make suitable assumptions to prove linear convergence

Linear convergence

We now want to make suitable assumptions to prove linear convergence

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

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

We make the following **set of 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 $\{Au_i\}_i \subset Y$ is linearly independent in Y

Linear convergence

We now want to make suitable assumptions to prove linear convergence

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

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

We make the following **set of 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 $\{Au_i\}_i \subset Y$ is linearly independent in Y

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

$$\bar{u} = \sum_i c_i u_i \tag{1}$$

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

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

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

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

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

$$\|A(u_i - u)\|_Y \leq Cg(u, \bar{u}_i) \quad \text{for every } i$$

Theorem (Bredies, C., Fanzon, Walter (2021))

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

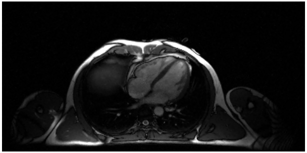
$$G(u^n) - \min_u G(u) \leq C\zeta^n$$

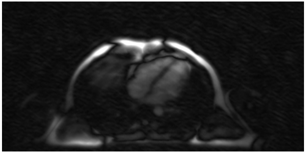
Dynamic inverse problems

Motivation: Reconstruct motion of objects (organs, blood flow, cells tracking) from sub-acquisition time samples.

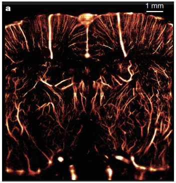
Several difficulties:

- Motion on sub-acquisition time scales $\rightsquigarrow$ artifacts in reconstructions.
- Drawbacks: need for unrealistic assumptions (e.g. periodicity in heart beating). Still limited to low-resolution





Infimal convolution of Total Generalized Variation functionals for dynamic. M. Schloegl, M. Holler, A. Schwarzl, K. Bredies and R. Stollberger, 2017



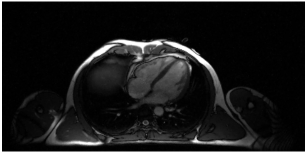
Ultrafast ultrasound localization microscopy for deep super-resolution vascular imaging. Errico et al., 2015

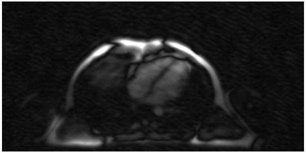
Dynamic inverse problems

Motivation: Reconstruct motion of objects (organs, blood flow, cells tracking) from sub-acquisition time samples.

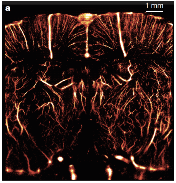
Several difficulties:

- Motion on sub-acquisition time scales $\rightsquigarrow$ artifacts in reconstructions.
- Drawbacks: need for unrealistic assumptions (e.g. periodicity in heart beating). Still limited to low-resolution





Infimal convolution of Total Generalized Variation functionals for dynamic. M. Schloegl, M. Holler, A. Schwarzl, K. Bredies and R. Stollberger, 2017



Ultrafast ultrasound localization microscopy for deep super-resolution vascular imaging. Errico et al., 2015

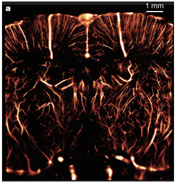
Dynamic inverse problems

Motivation: Reconstruct motion of objects (organs, blood flow, cells tracking) from sub-acquisition time samples.

Several difficulties:

- Motion on sub-acquisition time scales $\rightsquigarrow$ artifacts in reconstructions.
- Drawbacks: need for unrealistic assumptions (e.g. periodicity in heart beating). Still limited to low-resolution

Infimal convolution of Total Generalized Variation functionals for dynamic. M. Schloegl, M. Holler, A. Schwarzl, K. Bredies and R. Stollberger, 2017



Ultrafast ultrasound localization microscopy for deep super-resolution vascular imaging. Errico et al., 2015

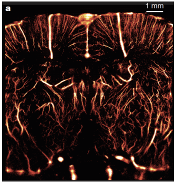
Dynamic inverse problems

Motivation: Reconstruct motion of objects (organs, blood flow, cells tracking) from sub-acquisition time samples.

Several difficulties:

- Motion on sub-acquisition time scales $\rightsquigarrow$ artifacts in reconstructions.
- Drawbacks: need for unrealistic assumptions (e.g. periodicity in heart beating). Still limited to low-resolution

Infimal convolution of Total Generalized Variation functionals for dynamic. M. Schloegl, M. Holler, A. Schwarzl, K. Bredies and R. Stollberger, 2017



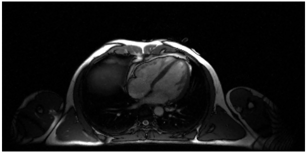
Ultrafast ultrasound localization microscopy for deep super-resolution vascular imaging. Errico et al., 2015

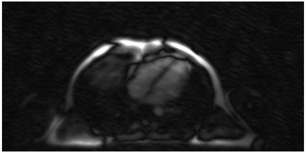
Dynamic inverse problems

Motivation: Reconstruct motion of objects (organs, blood flow, cells tracking) from sub-acquisition time samples.

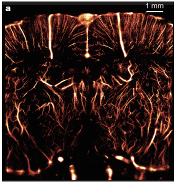
Several difficulties:

- Motion on sub-acquisition time scales $\rightsquigarrow$ artifacts in reconstructions.
- Drawbacks: need for unrealistic assumptions (e.g. periodicity in heart beating). Still limited to low-resolution





Infimal convolution of Total Generalized Variation functionals for dynamic. M. Schloegl, M. Holler, A. Schwarzl, K. Bredies and R. Stollberger, 2017



Ultrafast ultrasound localization microscopy for deep super-resolution vascular imaging. Errico et al., 2015

We want to reconstruct a time dependent measure $\rho_t \in \mathcal{M}(\overline{\Omega})$ for $t \in (0,1)$. At each time t we have a linear observation operator

$$K_t : \mathcal{M}(\overline{\Omega}) \to H_t$$

where H_t are time dependent Hilbert spaces.

We want to reconstruct a time dependent measure $\rho_t \in \mathcal{M}(\overline{\Omega})$ for $t \in (0,1)$. At each time t we have a linear observation operator

$$K_t : \mathcal{M}(\overline{\Omega}) \to H_t$$

where H_t are time dependent Hilbert spaces.
Then we solve the following variational inverse problem

$$\frac{1}{2}\int_0^1 \|K_t\rho_t - f_t\|_{H_t}^2 \, dt + \text{OT Regularizer}$$

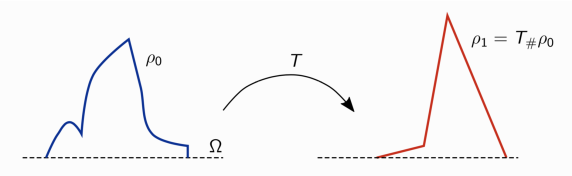
where $f_t \in H_t$ is the data.

The goal is to choose a regularizer that penalizes an Optimal Transport energy acting on the measure ρ_t at different time instants. In this way the regularized motion will be as *natural* as possible.

An optimal transport approach for solving dynamic inverse problems in spaces of measures. K. Bredies, S. Fanzon. M2AN (2020)
Dynamic Cell Imaging in PET with Optimal Transport Regularization. Schmitzer, B., Schäfers, K.P., Wirth, B., 2019

Optimal Transport - Static formulation

Let $\Omega \subset \mathbb{R}^d$ be a 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 distance between ρ_0 and ρ_1 can be defined as

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

(Monge formulation of Optimal Transport)

Kinetic Formulation of Optimal Transport - Benamou Brenier Energy

Kinetic Formulation of Optimal Transport - Benamou Brenier Energy

Given a vector field

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

we say that the pair (ρ_t, v_t) solves the continuity equation with initial conditions if

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

Kinetic Formulation of Optimal Transport - Benamou Brenier Energy

Given a vector field

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

we say that the pair (ρ_t, v_t) solves the continuity equation with initial conditions if

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

Theorem (Benamou-Brenier '00)

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

On the left side we minimize the kinetic energy of a pair (ρ_t, v_t) solving the continuity equation with initial and final data ρ and ρ_1.

The right hand side is the classical Monge formulation of OT.

Consider the following variational inverse problem

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

where the regularizer is

$$J_{\alpha,\beta}(\rho_t, v_t) = \underbrace{\frac{\alpha}{2}\int_0^1 \int_{\overline{\Omega}} |v_t(x)|^2 \, d\rho_t(x) dt}_{\text{Optimal Transport Regularizer}} + \beta \underbrace{\int_0^1 \|\rho_t\|_{\mathcal{M}(\overline{\Omega})} \, dt}_{\text{TV Regularizer}}$$

$$\text{s.t.} \quad \partial_t \rho_t + \operatorname{div}(\rho_t v_t) = 0 \qquad \text{(Continuity Equation)}$$

Consider the following variational inverse problem

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

where the regularizer is

$$J_{\alpha,\beta}(\rho_t, v_t) = \underbrace{\frac{\alpha}{2}\int_0^1 \int_{\overline{\Omega}} |v_t(x)|^2 \, d\rho_t(x) dt}_{\text{Optimal Transport Regularizer}} + \beta \underbrace{\int_0^1 \|\rho_t\|_{\mathcal{M}(\overline{\Omega})} \, dt}_{\text{TV Regularizer}}$$

$$\text{s.t.} \quad \partial_t \rho_t + \operatorname{div}(\rho_t v_t) = 0 \qquad \text{(Continuity Equation)}$$

Now consider the substitution $\rho_t v_t = m_t \in \mathcal{M}(\overline{\Omega}, \mathbb{R}^d)$ for every t. Then one can rewrite the previous functional in a convex way as:

$$J_{\alpha,\beta}(\rho_t, m_t) = \underbrace{\frac{\alpha}{2}\int_0^1 \int_{\overline{\Omega}} \left(\frac{|m_t(x)|}{\rho_t}\right)^2 \, d\rho_t(x) dt}_{\text{Optimal Transport Regularizer}} + \beta \underbrace{\int_0^1 \|\rho_t\|_{\mathcal{M}(\overline{\Omega})} \, dt}_{\text{TV Regularizer}}$$

$$\text{s.t.} \quad \partial_t \rho_t + \operatorname{div}(m_t) = 0 \quad \text{(Continuity Equation)}$$

We design a **generalized conditional gradient algorithm** (DGCG) following the general GCG algorithm described in the first part of the talk.

As sparse iterates we consider linear combinations of extremal points of the ball of $J_{\alpha,\beta}(\rho_t, m_t)$.

We design a **generalized conditional gradient algorithm** (DGCG) following the general GCG algorithm described in the first part of the talk.

As sparse iterates we consider linear combinations of extremal points of the ball of $J_{\alpha,\beta}(\rho_t, m_t)$.

Theorem (Bredies, C., Fanzon, Romero)

Let $B := \{(\rho_t, m_t) : \ J_{\alpha,\beta}(\rho_t, m_t) \leq 1\}$. *Then*

$$\mathrm{Ext}(B) = \left\{c(\gamma,\alpha,\beta)\left(dt \otimes \delta_{\gamma(t)}, \dot{\gamma}(t)dt \otimes \delta_{\gamma(t)}\right) : \ \gamma \in \mathrm{AC}([0,1];\overline{\Omega})\right\}$$

where $c(\gamma,\alpha,\beta) = \left(\alpha + \beta \int_0^1 |\dot{\gamma}(t)|^2 \, dt\right)^{-1}$.

Extremal points of the Benamou-Brenier energy are pairs of measures concentrated on absolutely continuous curves in Ω.

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

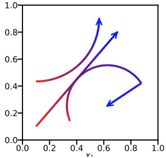
Description of DGCG algorithm

Definition (Iterates)

The iterates of the the algorithm are sparse measures $\mu = (\rho, m)$ that are linear combination of extremal points, i.e. such that

$$\mu = \sum_{j=1}^{N} c_j \mu_{\gamma_j} = \sum_{j=1}^{N} c_j (dt \otimes \delta_{\gamma_j(t)}, \dot{\gamma}_j(t) dt \otimes \delta_{\gamma_j(t)}) \tag{2}$$

for some $N \in \mathbb{N}$, $c_j > 0$ and $\gamma_j \in \mathrm{AC}([0,1]; \overline{\Omega})$.



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

Insertion step:

We add an extremal point of the ball of the Benamou-Brenier energy to the iterate μ^n. The extremal point is obtained by computing a minimizer to the partially linearized problem. Defining the dual variable $p_t^n := K_t(K_t^* \rho_t^n - f_t)$, we compute

$$\begin{aligned} &\min_{(\rho_t,m_t)} \int_0^1 \langle \rho_t, p_t^n \rangle_{\mathcal{M}(\overline{\Omega}),C(\overline{\Omega})}\, dt + J_{\alpha,\beta}(\rho_t,m_t) \\ &= \min_{(\rho_t,m_t)\in Ext(\{J_{\alpha,\beta}\leq 1\})} \int_0^1 \langle \rho_t, p_t^n \rangle_{\mathcal{M}(\overline{\Omega}),C(\overline{\Omega})}\, dt \\ &= \min_{\gamma\in \mathrm{AC}([0,1];\overline{\Omega})} \left(\alpha + \beta \int_0^1 |\dot\gamma(t)|^2\, dt\right)^{-1} \int_0^1 p_t^n(\gamma(t))\, dt\,, \end{aligned} \tag{3}$$

Insertion step:

We add an extremal point of the ball of the Benamou-Brenier energy to the iterate μ^n. The extremal point is obtained by computing a minimizer to the partially linearized problem. Defining the dual variable $p_t^n := K_t(K_t^* \rho_t^n - f_t)$, we compute

$$\begin{aligned}
&\min_{(\rho_t, m_t)} \int_0^1 \langle \rho_t, p_t^n \rangle_{\mathcal{M}(\overline{\Omega}), C(\overline{\Omega})} \, dt + J_{\alpha,\beta}(\rho_t, m_t) \\
&\quad = \min_{(\rho_t, m_t) \in Ext(\{J_{\alpha,\beta} \leq 1\})} \int_0^1 \langle \rho_t, p_t^n \rangle_{\mathcal{M}(\overline{\Omega}), C(\overline{\Omega})} \, dt \\
&\quad = \min_{\gamma \in \mathrm{AC}([0,1];\overline{\Omega})} \left(\alpha + \beta \int_0^1 |\dot{\gamma}(t)|^2 \, dt \right)^{-1} \int_0^1 p_t^n(\gamma(t)) \, dt \,, \qquad (3)
\end{aligned}$$

We reduce the insertion step to a variational problem in $\mathrm{AC}([0,1];\overline{\Omega})$.

The next iterate is constructed first by adding to μ^n the atom (extremal point) μ_{γ^*} associated to the minimizer of (3) denoted by γ^*.

Coefficient optimization step

We optimize the coefficients of the linear combination

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

by solving

$$\min_{(c_1, c_2, \ldots, c_{N_n+1}) \in \mathbb{R}_+^{N_n+1}} G_{\alpha,\beta}\left(\sum_{i=1}^{N_n} c_i \mu_{\gamma_i^n} + c_{N_n+1}\mu_{\gamma^*}\right). \tag{4}$$

Coefficient optimization step

We optimize the coefficients of the linear combination

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

by solving

$$\min_{(c_1,c_2,\dots,c_{N_n+1})\in\mathbb{R}_+^{N_n+1}} G_{\alpha,\beta}\left(\sum_{i=1}^{N_n} c_i \mu_{\gamma_i^n} + c_{N_n+1}\mu_{\gamma^*}\right). \tag{4}$$

Calling $(c_1^*, c_2^*, \dots, c_{N_n+1}^*)$ a solution of (4) we obtain the next iterate as

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

Theorem (Bredies, C., Fanzon, Romero)

Let $\mu^n = (\rho_t^n, m_t^n)$ an iterate of the DGCG algorithm. Then μ^n converges weakly to a minimizer of $G_{\alpha,\beta}$ with sublinear rate.*

Numerical simulations

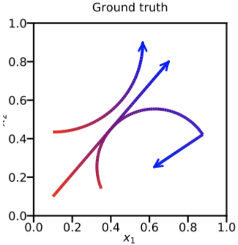
We now want to test our DGCG algorithm to reconstruct observations for the inverse problem

$$\min_{(\rho_t, m_t)} \frac{1}{2} \int_0^1 \|K_t \rho_t - f_t\|_{H_t}^2 \, dt + J_{\alpha,\beta}(\rho_t, m_t)$$

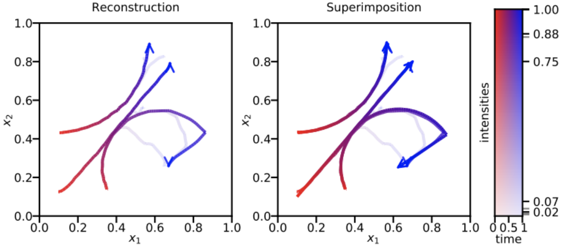
where $J_{\alpha,\beta}$ is the Benamou-Brenier energy.

- The domain is $\overline{\Omega} = [0,1]^2$
- The observation operators $K_t : \mathcal{M}(\overline{\Omega}) \to H_t$ are time dependent Fourier measurements that at each time detect different sets of Fourier frequencies of ρ_t (severely ill-posed).
- The data f_t is the image by K_t of sparse dynamic measures (with 20% and 60% of noise)

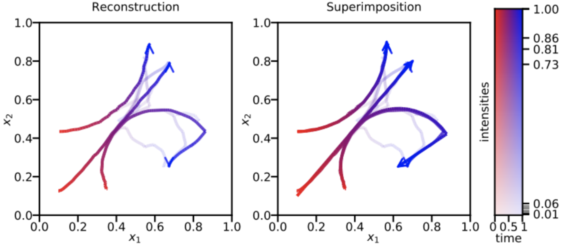
Noiseless reconstructions and with 20% of noise:



(A) Considered ground-truth.

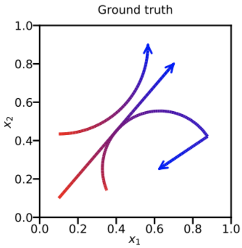


(B) Reconstruction with noiseless data.

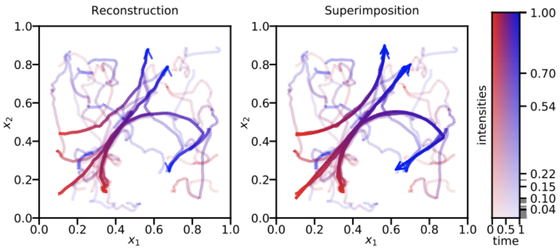


(C) Reconstruction with 20% of relative noise.

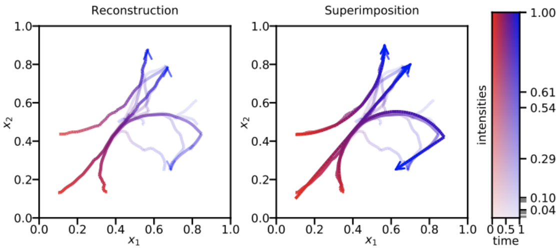
Reconstructions with 60% of noise:



(A) Considered ground-truth.



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



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

References

- **Sparsity of Solutions for variational inverse problem with finite dimensional constraints** K. Bredies, M. Carioni. Calc. Var. and PDE (2020)
- **Linear convergence of accelerated generalized conditional gradient methods** K. Bredies, M. Carioni, S. Fanzon, D. Walter (2021)
- **A generalized conditional gradient method for dynamic inverse problems with optimal transport regularization.** K. Bredies, M. Carioni, S. Fanzon, F. Romero. Foundations of Computational Mathematics (2022)
- **On the extremal points of the ball of the Benamou-Brenier energy** K. Bredies, M. Carioni, S. Fanzon, F. Romero. Bull. Lond. Math. Soc (2021)
- **Inverse problems in the space of measures** K. Bredies, H. K. Pikkaranien. ESAIM: COCV, (2013)
- **An optimal transport approach for solving dynamic inverse problems in spaces of measures** K. Bredies, S. Fanzon. M2AN (2020)