Modern Challenges in Large-Scale and High Dimensional Data Analysis

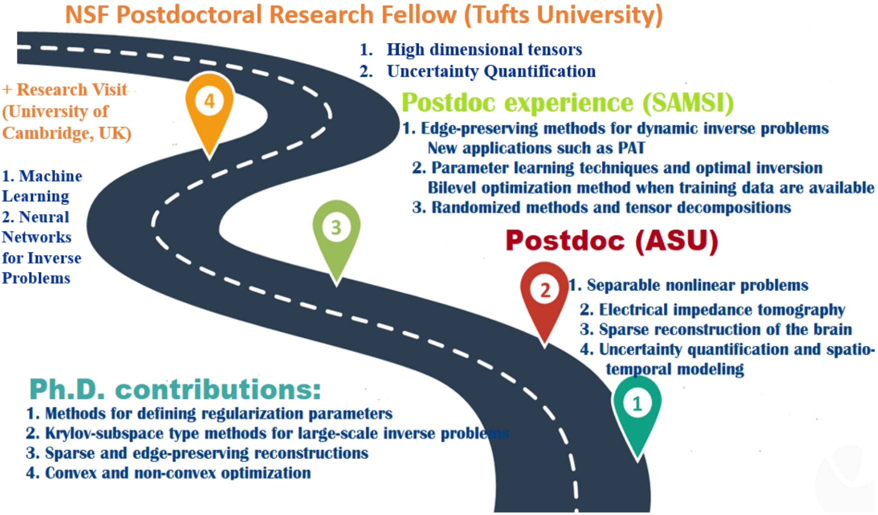
Mirjeta Pasha[1]

in collaboration with
Arvind Saibaba (NCSU), Silvia Gazzola, (UoB), Malena Espanol, (ASU), and Eric de Sturler, (VT)
Shiwei Lan (ASU) and Shuyi Li (ASU)

[1]National Science Foundation Postdoctoral Fellow and Visiting Scholar
Department of Mathematics, Tufts University, Medford, MA, USA

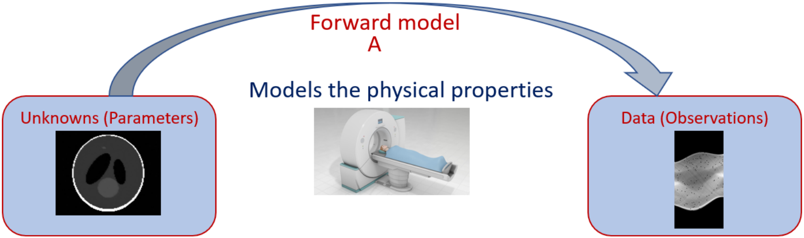
Applied Mathematics Seminar, Hunter College, City University of New York
November 10, 2022

My research journey



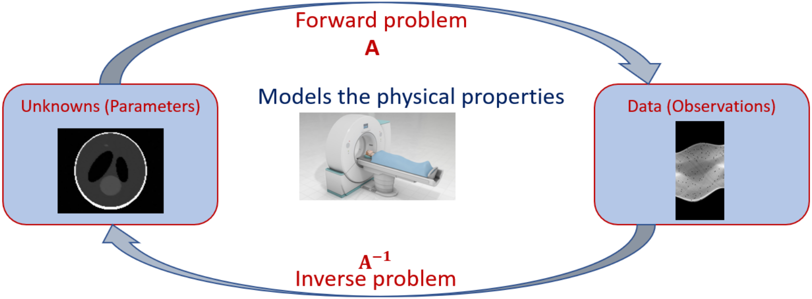
What are inverse problems?

Overview



What are inverse problems?

Overview



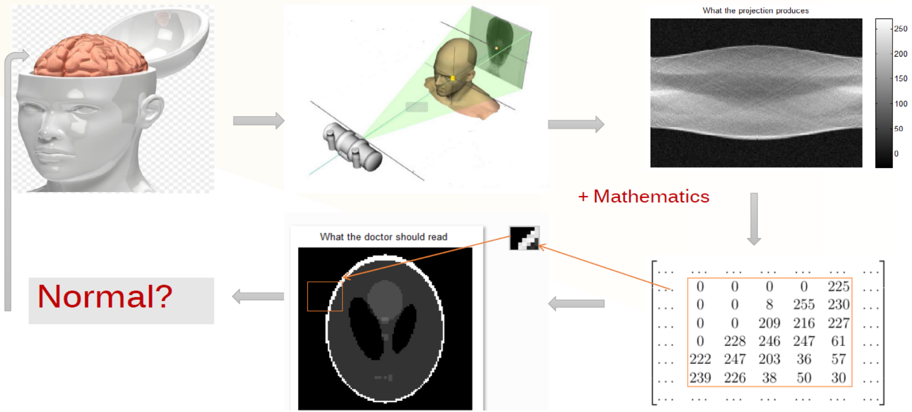
[1]

[1] Image courtesy: Google

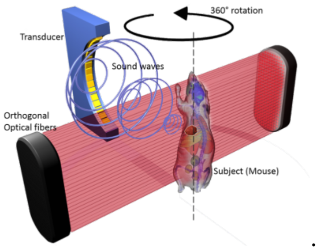
Challenges in large-scale data analysis

Computerized tomography

- Medical Imaging:



Time-elapsed Photoacoustic Tomography (PAT[2])



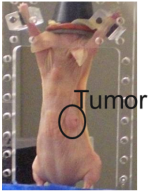
→ Given spherical projections

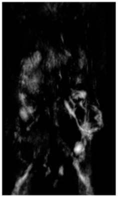
→ Find initial pressure

⚠ Computationally expensive

✓ Non-invasive, non-ionizing

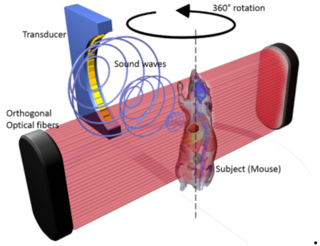
✓ PAT generates high-resolution images in optically ballistic and diffusive regimes.



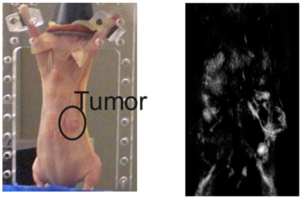


[2] tomowave.com, Wang, Anastasio (2011), Xia, Yao, Yang (2014), Chung, Nguyen (2017)

Time-elapsed Photoacoustic Tomography (PAT^2)

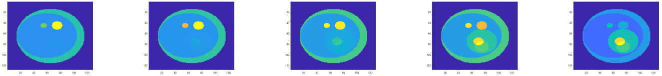


- ✓ Non-invasive, non-ionizing
- ✓ PAT generates high-resolution images in optically ballistic and diffusive regimes.



- → Given spherical projections
- → Find initial pressure
- ⚠ Computationally expensive

$i = 1$ $i = 20$ $i = 35$ $i = 55$ $i = 80$

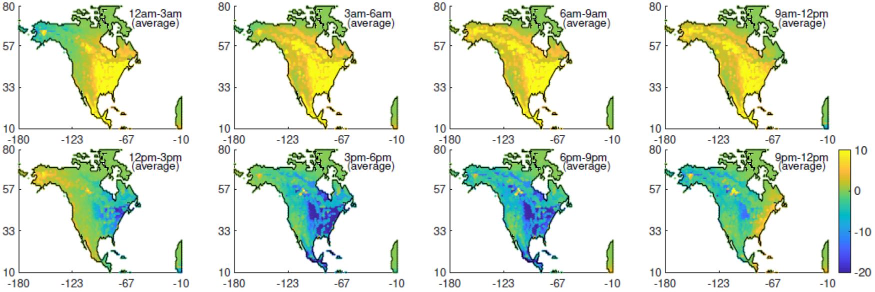


[2] tomowave.com, Wang, Anastasio (2011), Xia, Yao, Yang (2014), Chung, Nguyen (2017)

Atmospheric imaging problem

Track greenhouse gases using satellites

- Estimate spatiotemporal green house gas fluxes at the Earth's surface using observations of gases in the atmosphere.
- IP help generate detailed maps of surface emissions using atmospheric observations.



3

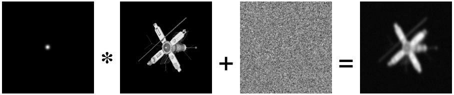
[3] Nehrkorn, Eluszkiewicz, Wofsy, Lin, Gerbig, Longo, Freitas (2010); NOAA Global Monitoring Division: CarbonTracker CT2017. (2019); Cho (2019).

Ill-posed inverse problems

Mathematical problem setup

$$\min_{\mathbf{u}\in\mathbb{R}^n} \|\mathbf{A}(\mathbf{y})\mathbf{u} - \mathbf{b}\|_2^2, \quad \text{where} \quad \mathbf{A}(\mathbf{y}_{\text{true}})\mathbf{u}_{\text{true}} + \mathbf{e} = \mathbf{b}, \quad \text{and} \quad \mathbf{b} = \mathbf{b}_{\text{true}} + \mathbf{e}$$

- $\mathbf{b} \in \mathbb{R}^m$ — available observations (measurements)
- $\mathbf{u}_{\text{true}} \in \mathbb{R}^n$ — desired solution (unknown quantity of interest)
- $\mathbf{A} \in \mathbb{R}^{m\times n}$ — parameter-to-observable map
- $\mathbf{b}_{\text{true}} \in \mathbb{R}^m$ — data without noise (not available)
- $\mathbf{e} \in \mathbb{R}^m$ — additive noise (Gaussian, Poisson, Laplace, or mixed.)
- $\mathbf{y} \in \mathbb{R}^d$ — parameters that parametrize the forward operator



A **u** **e** **b**

Introduction

Discrete ill-posed problems (continued)

- Would like to determine an approximate solution of $\mathbf{A}\mathbf{u} = \mathbf{b}$ by solving

$$\min_{\mathbf{u}\in\mathbb{R}^n} \|\mathbf{A}\mathbf{u} - \mathbf{b}\|_2^2$$

Introduction

Discrete ill-posed problems (continued)

- Would like to determine an approximate solution of $\mathbf{Au} = \mathbf{b}$ by solving

$$\min_{\mathbf{u}\in\mathbb{R}^n} \|\mathbf{Au} - \mathbf{b}\|_2^2$$

$$\|\mathbf{u}\|_2^2 = \mathbf{u}^T\mathbf{u}, \text{ so } \quad \|\mathbf{Au} - \mathbf{b}\|_2^2 = (\mathbf{Au} - \mathbf{b})^T(\mathbf{A}^T\mathbf{u} - \mathbf{b})$$

Introduction

Discrete ill-posed problems (continued)

- Would like to determine an approximate solution of $\mathbf{Au} = \mathbf{b}$ by solving

$$\min_{\mathbf{u}\in\mathbb{R}^n} \|\mathbf{Au} - \mathbf{b}\|_2^2$$

$$\|\mathbf{u}\|_2^2 = \mathbf{u}^T\mathbf{u}, \text{ so} \quad \|\mathbf{Au} - \mathbf{b}\|_2^2 = (\mathbf{Au} - \mathbf{b})^T(\mathbf{A}^T\mathbf{u} - \mathbf{b})$$

$$\mathbf{u}_{\text{naive}} = (\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T(\mathbf{b}_{\text{true}} + \mathbf{e}) = \underbrace{(\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\mathbf{b}_{\text{true}}}_{\mathbf{u}_{\text{true}}} + \underbrace{(\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\mathbf{e}}_{\text{large}}$$

- $\mathbf{b}_{\text{true}} \in \mathcal{R}(\mathbf{A})$, $(\mathbf{A}^T\mathbf{A})^{-1}\mathbf{b}_{\text{true}}$, bounded. $\mathbf{e} \notin \mathcal{R}(\mathbf{A})$, $(\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\mathbf{e}$ arbitrarily large.

Introduction

Discrete ill-posed problems (continued)

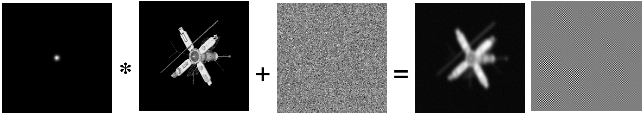
- Would like to determine an approximate solution of $\mathbf{Au} = \mathbf{b}$ by solving

$$\min_{\mathbf{u}\in\mathbb{R}^n} \|\mathbf{Au}-\mathbf{b}\|_2^2$$

$$\|\mathbf{u}\|_2^2 = \mathbf{u}^T\mathbf{u}, \text{ so } \quad \|\mathbf{Au}-\mathbf{b}\|_2^2 = (\mathbf{Au}-\mathbf{b})^T(\mathbf{A}^T\mathbf{u}-\mathbf{b})$$

$$\mathbf{u}_{\text{naive}} = (\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T(\mathbf{b}_{\text{true}}+\mathbf{e}) = \underbrace{(\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\mathbf{b}_{\text{true}}}_{\mathbf{u}_{\text{true}}} + \underbrace{(\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\mathbf{e}}_{\text{large}}$$

- $\mathbf{b}_{\text{true}} \in \mathcal{R}(\mathbf{A})$, $(\mathbf{A}^T\mathbf{A})^{-1}\mathbf{b}_{\text{true}}$, bounded. $\mathbf{e} \notin \mathcal{R}(\mathbf{A})$, $(\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\mathbf{e}$ arbitrarily large.



$\mathbf{A}$ $\quad \mathbf{u}_{\text{true}}$ $\quad \mathbf{e}$ $\quad \mathbf{b}$ $\quad \mathbf{u}_{\text{naive}}$

Regularization

Direct solvers

If $\mathbf{A}$ not too large, consider Singular Value Decomposition, (SVD) $\mathbf{A} = \mathbf{U}\mathbf{S}\mathbf{V}^T$.

Regularization

Direct solvers

If $\mathbf{A}$ not too large, consider Singular Value Decomposition, (SVD) $\mathbf{A} = \mathbf{U}\mathbf{S}\mathbf{V}^T$.

1. Truncated SVD (tSVD): $\mathbf{u}^* = \sum_{i=1}^{k} \frac{\mathbf{u}_i\mathbf{b}}{\boldsymbol{\sigma}}\mathbf{v}_i$, $n = \text{rank}(\mathbf{A})$, $k < n$.

Regularization

Direct solvers

If $\mathbf{A}$ not too large, consider Singular Value Decomposition, (SVD) $\mathbf{A} = \mathbf{U}\mathbf{S}\mathbf{V}^T$.

1. Truncated SVD (tSVD): $\mathbf{u}^* = \sum_{i=1}^{k} \frac{\mathbf{u}_i\mathbf{b}}{\boldsymbol{\sigma}}\mathbf{v}_i$, $n = \text{rank}(\mathbf{A})$, $k < n$.
2. General SVD filtering: $\mathbf{u}^* = \sum_{i=1}^{n} \phi_i(\lambda)\frac{\mathbf{u}_i\mathbf{b}}{\boldsymbol{\sigma}}\mathbf{v}_i$, $\phi_i(\lambda) = \begin{cases} 1, & \text{if } \sigma_i \quad \text{large} \\ 0, & \sigma_i \quad \text{small} \end{cases}$

Regularization

Direct solvers

If $\mathbf{A}$ not too large, consider Singular Value Decomposition, (SVD) $\mathbf{A} = \mathbf{U}\mathbf{S}\mathbf{V}^T$.

1. Truncated SVD (tSVD): $\mathbf{u}^* = \sum_{i=1}^{k} \frac{\mathbf{u}_i\mathbf{b}}{\boldsymbol{\sigma}} \mathbf{v}_i$, $n = \text{rank}(\mathbf{A})$, $k < n$.
2. General SVD filtering: $\mathbf{u}^* = \sum_{i=1}^{n} \phi_i(\lambda) \frac{\mathbf{u}_i\mathbf{b}}{\boldsymbol{\sigma}} \mathbf{v}_i$, $\phi_i(\lambda) = \begin{cases} 1, & \text{if } \sigma_i \quad \text{large} \\ 0, & \sigma_i \quad \text{small} \end{cases}$
3. Tikhonov regularization:

$$\mathbf{u}^* = \arg\min_{\mathbf{u}\in\mathbb{R}^n} \{\|\mathbf{A}\mathbf{u} - \mathbf{b}\|_2^2 + \lambda\|\mathbf{u}\|_2^2\} = \sum_{i=1}^{n} \frac{\sigma_i^2}{\sigma_i^2 + \lambda^2} \frac{\mathbf{u}_i\mathbf{b}}{\boldsymbol{\sigma}} \mathbf{v}_i$$

Regularization

Direct solvers

If $\mathbf{A}$ not too large, consider Singular Value Decomposition, (SVD) $\mathbf{A} = \mathbf{U}\mathbf{S}\mathbf{V}^T$.

1. Truncated SVD (tSVD): $\mathbf{u}^* = \sum_{i=1}^{k} \frac{\mathbf{u}_i \mathbf{b}}{\boldsymbol{\sigma}} \mathbf{v}_i$, $n = \text{rank}(\mathbf{A})$, $k < n$.
2. General SVD filtering: $\mathbf{u}^* = \sum_{i=1}^{n} \phi_i(\lambda) \frac{\mathbf{u}_i \mathbf{b}}{\boldsymbol{\sigma}} \mathbf{v}_i$, $\phi_i(\lambda) = \begin{cases} 1, & \text{if } \sigma_i \quad \text{large} \\ 0, & \sigma_i \quad \text{small} \end{cases}$
3. Tikhonov regularization:

$$\mathbf{u}^* = \arg\min_{\mathbf{u} \in \mathbb{R}^n} \{ \|\mathbf{A}\mathbf{u} - \mathbf{b}\|_2^2 + \lambda \|\mathbf{u}\|_2^2 \} = \sum_{i=1}^{n} \frac{\sigma_i^2}{\sigma_i^2 + \lambda^2} \frac{\mathbf{u}_i \mathbf{b}}{\boldsymbol{\sigma}} \mathbf{v}_i$$

Closed form solution: $\mathbf{u}_\lambda = (\mathbf{A}^T\mathbf{A} + \lambda^2 \mathbf{L}^T\mathbf{L})^{-1}(\mathbf{A}^T\mathbf{b})$

Regularization

Direct solvers

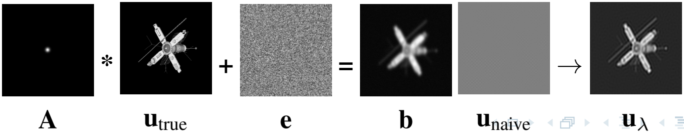
If $\mathbf{A}$ not too large, consider Singular Value Decomposition, (SVD) $\mathbf{A} = \mathbf{U}\mathbf{S}\mathbf{V}^T$.

1. Truncated SVD (tSVD): $\mathbf{u}^* = \sum_{i=1}^{k} \frac{\mathbf{u}_i \mathbf{b}}{\boldsymbol{\sigma}} \mathbf{v}_i$, $n = \text{rank}(\mathbf{A})$, $k < n$.
2. General SVD filtering: $\mathbf{u}^* = \sum_{i=1}^{n} \phi_i(\lambda) \frac{\mathbf{u}_i \mathbf{b}}{\boldsymbol{\sigma}} \mathbf{v}_i$, $\phi_i(\lambda) = \begin{cases} 1, & \text{if } \sigma_i \quad \text{large} \\ 0, & \sigma_i \quad \text{small} \end{cases}$
3. Tikhonov regularization:

$$\mathbf{u}^* = \arg\min_{\mathbf{u}\in\mathbb{R}^n} \{\|\mathbf{A}\mathbf{u} - \mathbf{b}\|_2^2 + \lambda \|\mathbf{u}\|_2^2\} = \sum_{i=1}^{n} \frac{\sigma_i^2}{\sigma_i^2 + \lambda^2} \frac{\mathbf{u}_i \mathbf{b}}{\boldsymbol{\sigma}} \mathbf{v}_i$$

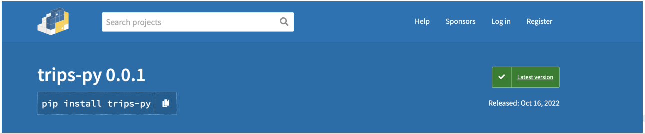
Closed form solution: $\mathbf{u}_\lambda = (\mathbf{A}^T\mathbf{A} + \lambda^2 \mathbf{L}^T\mathbf{L})^{-1}(\mathbf{A}^T\mathbf{b})$

> A good solution depends on the choice of a good regularization parameter!!!
> —- Large scale problems — difficult to solve.
> —- Regularization parameter nontrivial to be estimated.



Large scale problems
packages

1. Books
   - Hansen, Per Christian. Discrete inverse problems: insight and algorithms., SIAM, 2010.
   - Hansen, Per Christian, James G. Nagy, and Dianne P. O'leary. Deblurring images: matrices, spectra, and filtering., SIAM, 2006.
   - Hansen, Per Christian. Rank-deficient and discrete ill-posed problems: numerical aspects of linear inversion., SIAM, 1998.
2. Packages
   - Gazzola, Silvia, Per Christian Hansen, and James G. Nagy. IR Tools: a MATLAB package of iterative regularization methods and large-scale test problems. Numerical Algorithms 81.3 (2019): 773-811.
   - Hansen, Per Christian, and Maria Saxild-Hansen. AIR Tools: A MATLAB Package of Algebraic Iterative Reconstruction Techniques. DTU Informatics, 2010.
   - Van Aarle, Wim, et al. The ASTRA Toolbox: A platform for advanced algorithm development in electron tomography., Ultramicroscopy 157 (2015): 35-47.
   - Pasha, Mirjeta, and Sanderford, Connor. TRIPs-Py: Techniques for Regularization of Inverse Problems in Python (in preparation)



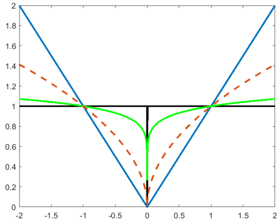
Regularization methods

Discrete ill-posed problems

Regularization methods

$$\mathbf{u}_\lambda = \arg\min_{\mathbf{u}} \frac{1}{p}\|\mathbf{A}\mathbf{u} - \mathbf{b}\|_p^p + \frac{\lambda}{q}\|\mathbf{L}(\beta)\mathbf{u}\|_q^q \quad 0 < p, q \leq 2$$

algorithm error	model error	data error	regularization error



- $p = 2$ and $q = 2$ Tikhonov reg.

4

[4] Lanza, Morigi, Reichel, Sgallari (2015), Buccini, Pasha, Reichel (2020), Buccini, Reichel (2021)

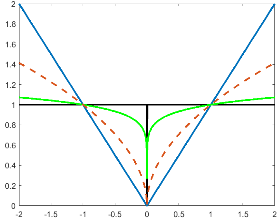
Regularization methods

Discrete ill-posed problems

Regularization methods

$$\mathbf{u}_\lambda = \arg\min_{\mathbf{u}} \frac{1}{p}\|\mathbf{A}\mathbf{u} - \mathbf{b}\|_p^p + \frac{\lambda}{q}\|\mathbf{L}(\beta)\mathbf{u}\|_q^q \quad 0 < p, q \leq 2$$

algorithm error	model error	data error	regularization error



- $p = 2$ and $q = 2$ Tikhonov reg.
- $1 \leq p, q < 2$ yields a convex minimization problem No closed form!!!!

4

[4] Lanza, Morigi, Reichel, Sgallari (2015), Buccini, Pasha, Reichel (2020), Buccini, Reichel (2021)

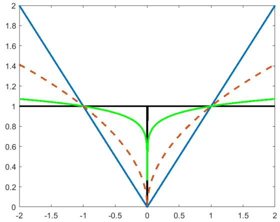
Regularization methods

Discrete ill-posed problems

Regularization methods

$$\mathbf{u}_\lambda = \arg\min_{\mathbf{u}} \frac{1}{p}\|\mathbf{A}\mathbf{u} - \mathbf{b}\|_p^p + \frac{\lambda}{q}\|\mathbf{L}(\beta)\mathbf{u}\|_q^q \quad 0 < p, q \leq 2$$

algorithm error | model error | data error | regularization error



- $p = 2$ and $q = 2$ Tikhonov reg.
- $1 \leq p, q < 2$ yields a convex minimization problem No closed form!!!!
- $0 < q < 1$ or $0 < p < 1$ yield to a non-convex minimization problem.

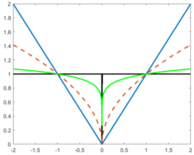
4

[4] Lanza, Morigi, Reichel, Sgallari (2015), Buccini, Pasha, Reichel (2020), Buccini, Reichel (2021)

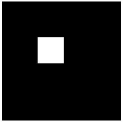
Total variation reconstruction

Edge preserving

Illustration $\|\mathbf{Lu}\|_q^q \quad 0 < q \leq 2$



Solid blue - ℓ_1, the dotted red-$\ell_{0.5}$, the solid green - $\ell_{0.1}$, the solid black - ℓ_0.

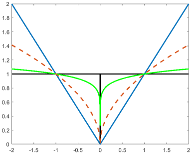


True

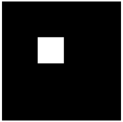
Total variation reconstruction

Edge preserving

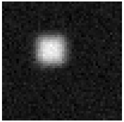
Illustration $\|\mathbf{Lu}\|_q^q \quad 0 < q \leq 2$



Solid blue - ℓ_1, the dotted red-$\ell_{0.5}$, the solid green - $\ell_{0.1}$, the solid black - ℓ_0.



True

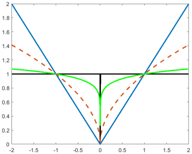


Observed

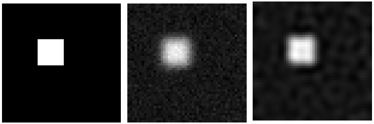
Total variation reconstruction

Edge preserving

Illustration $\|\mathbf{Lu}\|_q^q \quad 0 < q \leq 2$



Solid blue - ℓ_1, the dotted red-$\ell_{0.5}$, the solid green - $\ell_{0.1}$, the solid black - ℓ_0.

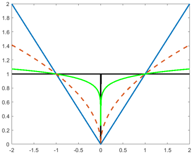


True   Observed ℓ_2 no TV

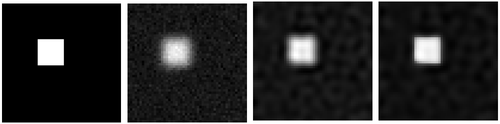
Total variation reconstruction

Edge preserving

Illustration $\|\mathbf{Lu}\|_q^q \quad 0 < q \leq 2$



Solid blue - ℓ_1, the dotted red-$\ell_{0.5}$, the solid green - $\ell_{0.1}$, the solid black - ℓ_0.

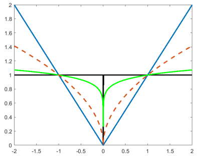


True Observed ℓ_2 no TV ℓ_1 + TV

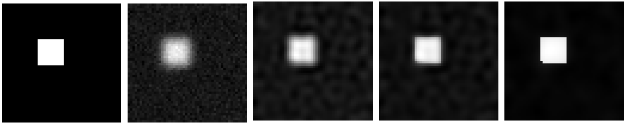
Total variation reconstruction

Edge preserving

Illustration $\|\mathbf{Lu}\|_q^q \quad 0 < q \leq 2$



Solid blue - ℓ_1, the dotted red-$\ell_{0.5}$, the solid green - $\ell_{0.1}$, the solid black - ℓ_0.

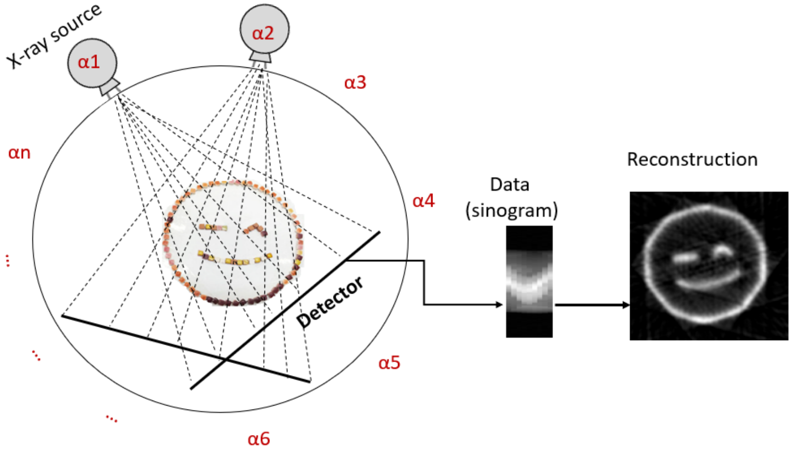


True Observed ℓ_2 no TV ℓ_1 + TV $\ell_{0.1} + TV$

Motivation for dynamic inverse problems–Computerized Tomography

Limited angles

Application: Computerized Tomography (CT)



5

[5] Meaney, Purish, Siltanen, (2018)

Dynamic inverse problems[6]

Computational challenges

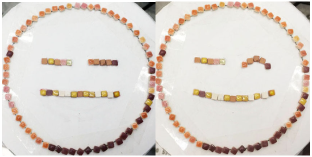
- Large-scale problems $\geq \mathcal{O}(10^6)$ measurements
  - Speed up the computational time
  - Lower the memory requirements
- Ill-posed problems
  - Efficiently determine the regularization parameters
  - Define new loss functions
  - Reconstruct solutions with specific properties
- Only a few measurements from limited angles
- Reconstruct solutions with edges

[6] Pasha, Saibaba, Gazzola, Espanol, de Sturler, https://arxiv.org/abs/2107.05727).

Dynamic inverse problems[6]

Computational challenges

- Large-scale problems $\geq \mathcal{O}(10^6)$ measurements
  - Speed up the computational time
  - Lower the memory requirements
- Ill-posed problems
  - Efficiently determine the regularization parameters
  - Define new loss functions
  - Reconstruct solutions with specific properties
- Only a few measurements from limited angles
- Reconstruct solutions with edges



[6] Pasha, Saibaba, Gazzola, Espanol, de Sturler, https://arxiv.org/abs/2107.05727).

Dynamic inverse problems[6]

Computational challenges

- Large-scale problems $\geq \mathcal{O}(10^6)$ measurements
  - Speed up the computational time
  - Lower the memory requirements
- Ill-posed problems
  - Efficiently determine the regularization parameters
  - Define new loss functions
  - Reconstruct solutions with specific properties
- Only a few measurements from limited angles
- Reconstruct solutions with edges

[6] Pasha, Saibaba, Gazzola, Espanol, de Sturler, https://arxiv.org/abs/2107.05727).

Dynamic inverse problems[6]

Computational challenges

- Large-scale problems $\geq \mathcal{O}(10^6)$ measurements
  - Speed up the computational time
  - Lower the memory requirements
- Ill-posed problems
  - Efficiently determine the regularization parameters
  - Define new loss functions
  - Reconstruct solutions with specific properties
- Only a few measurements from limited angles
- Reconstruct solutions with edges

[6] Pasha, Saibaba, Gazzola, Espanol, de Sturler, https://arxiv.org/abs/2107.05727).

Dynamic inverse problems[6]

Computational challenges

- Large-scale problems $\geq \mathcal{O}(10^6)$ measurements
  - Speed up the computational time
  - Lower the memory requirements
- Ill-posed problems
  - Efficiently determine the regularization parameters
  - Define new loss functions
  - Reconstruct solutions with specific properties
- Only a few measurements from limited angles
- Reconstruct solutions with edges

[6] Pasha, Saibaba, Gazzola, Espanol, de Sturler, https://arxiv.org/abs/2107.05727).

Dynamic inverse problems[6]

Computational challenges

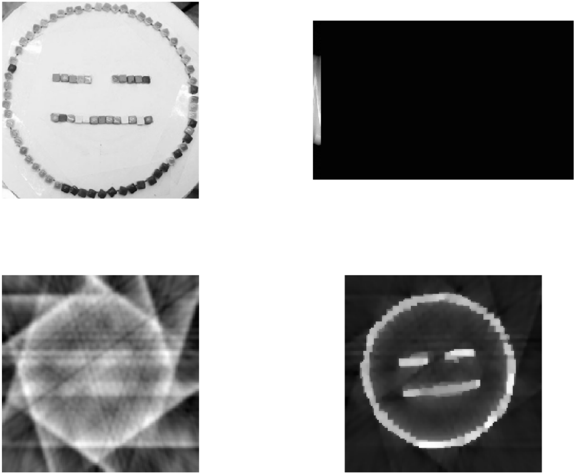
- Large-scale problems $\geq \mathcal{O}(10^6)$ measurements
  - Speed up the computational time
  - Lower the memory requirements
- Ill-posed problems
  - Efficiently determine the regularization parameters
  - Define new loss functions
  - Reconstruct solutions with specific properties
- Only a few measurements from limited angles
- Reconstruct solutions with edges

[6] Pasha, Saibaba, Gazzola, Espanol, de Sturler, https://arxiv.org/abs/2107.05727).

Dynamic inverse problems

Applications to limited angle tomography

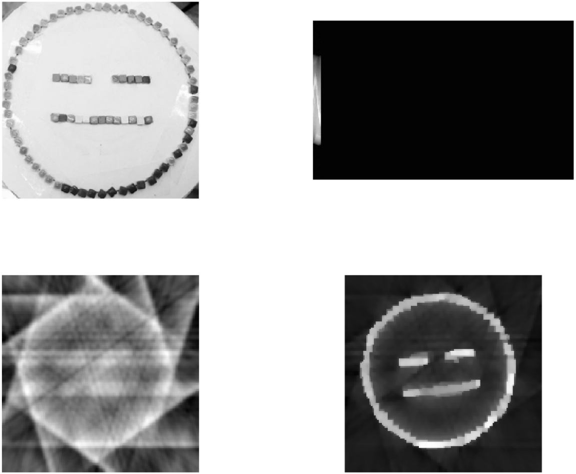
Time step = 1



Dynamic inverse problems

Applications to limited angle tomography

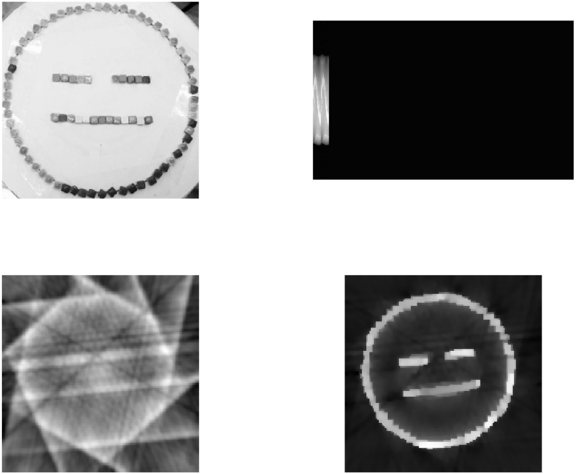
Time step = 1



Dynamic inverse problems

Applications to limited angle tomography

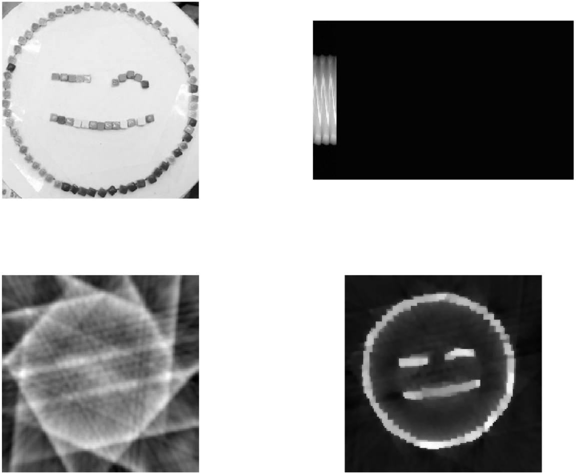
Time step = 2



Dynamic inverse problems

Applications to limited angle tomography

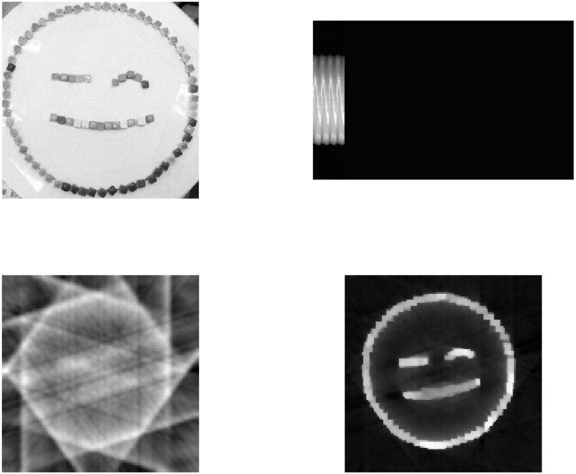
Time step = 3



Dynamic inverse problems

Applications to limited angle tomography

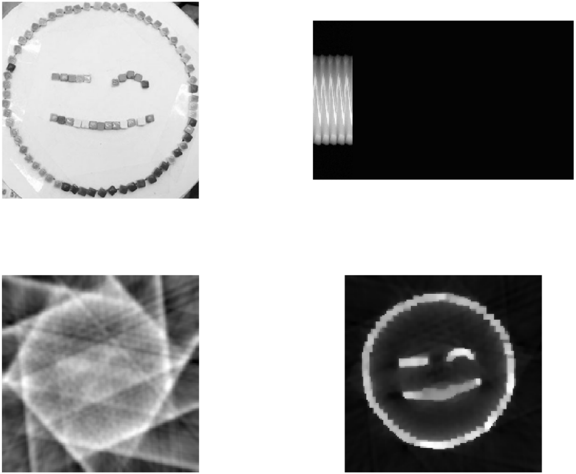
Time step = 4



Dynamic inverse problems

Applications to limited angle tomography

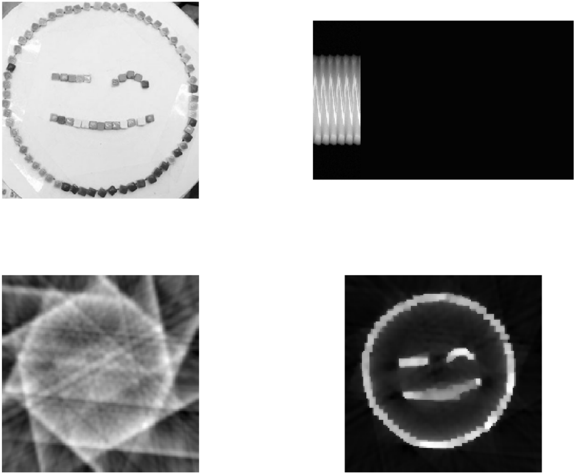
Time step = 5



Dynamic inverse problems

Applications to limited angle tomography

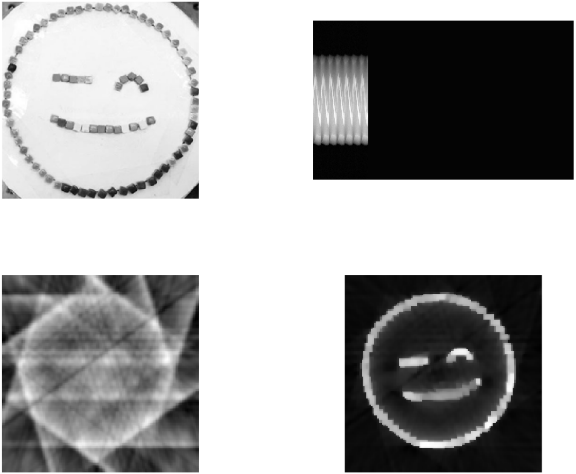
Time step = 6



Dynamic inverse problems

Applications to limited angle tomography

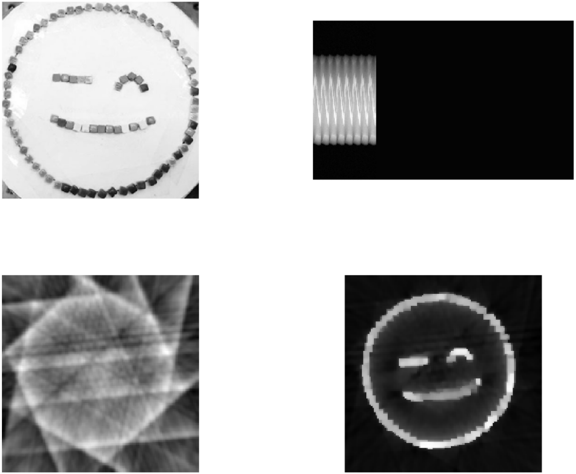
Time step = 7



Dynamic inverse problems

Applications to limited angle tomography

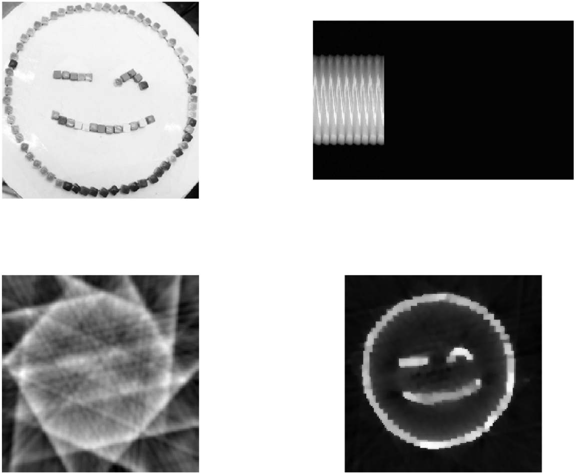
Time step = 8



Dynamic inverse problems

Applications to limited angle tomography

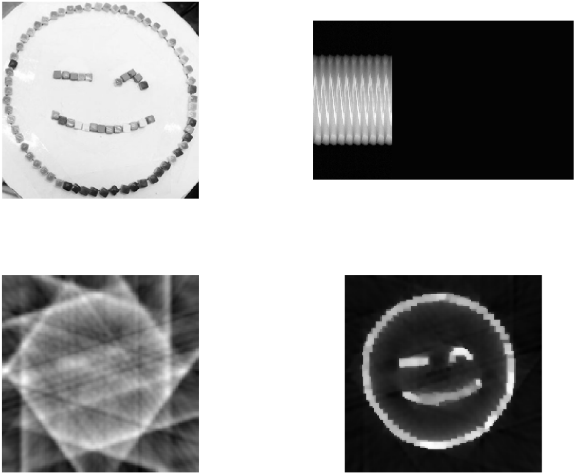
Time step = 9



Dynamic inverse problems

Applications to limited angle tomography

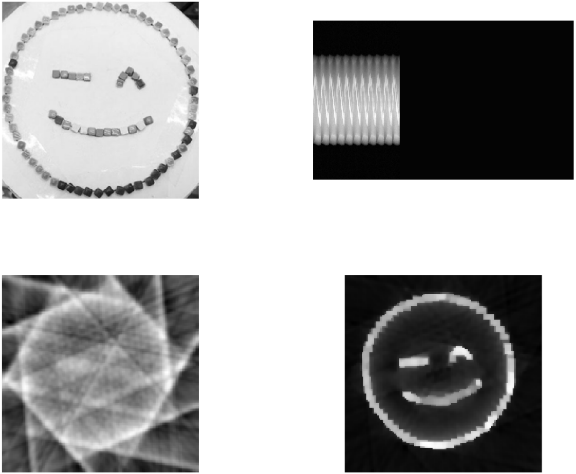
Time step = 10



Dynamic inverse problems

Applications to limited angle tomography

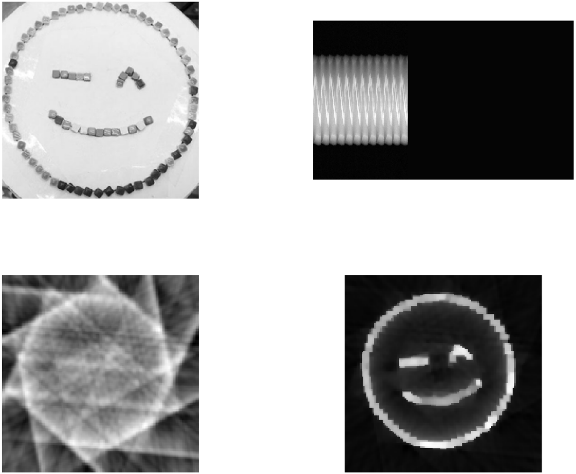
Time step = 11



Dynamic inverse problems

Applications to limited angle tomography

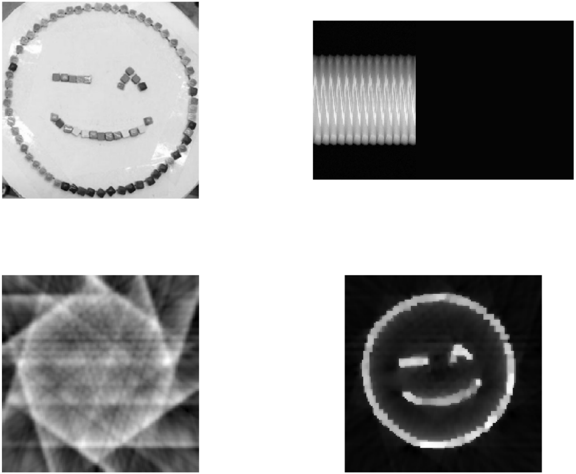
Time step = 12



Dynamic inverse problems

Applications to limited angle tomography

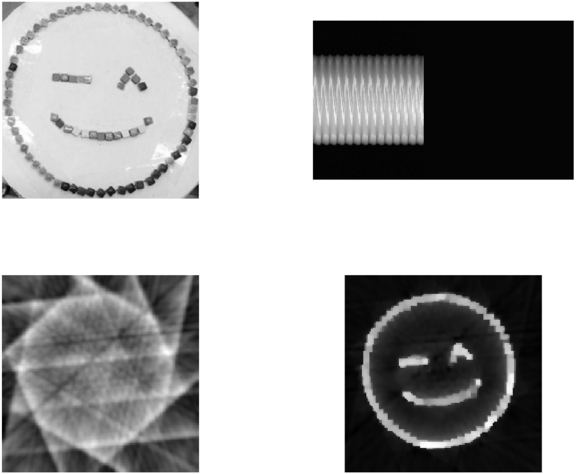
Time step = 13



Dynamic inverse problems

Applications to limited angle tomography

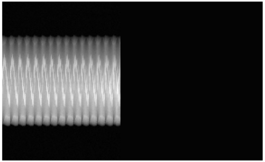
Time step = 14

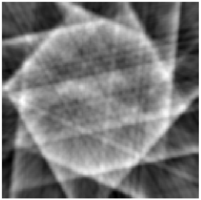


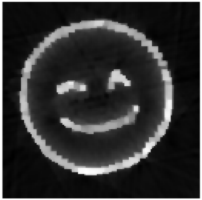
Dynamic inverse problems

Applications to limited angle tomography

Time step = 15



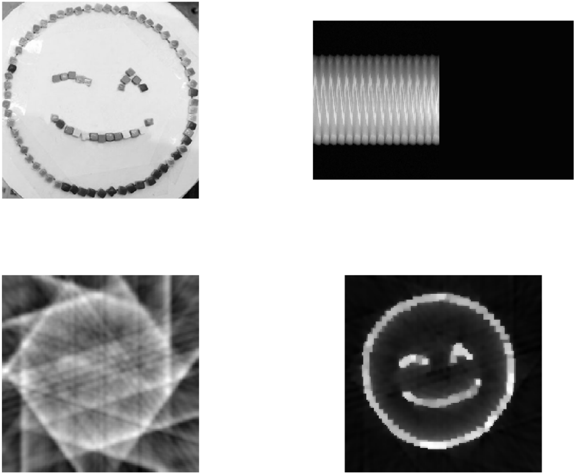




Dynamic inverse problems

Applications to limited angle tomography

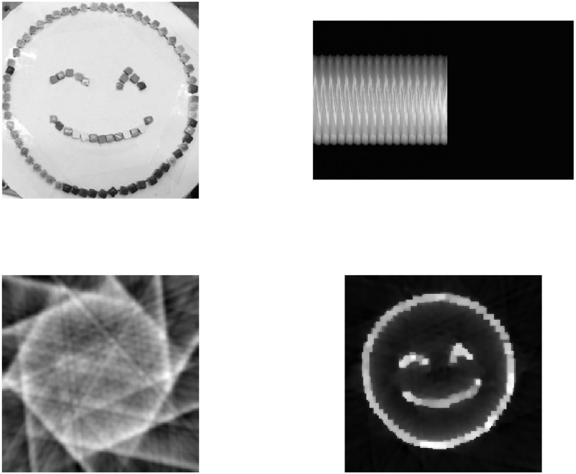
Time step = 16



Dynamic inverse problems

Applications to limited angle tomography

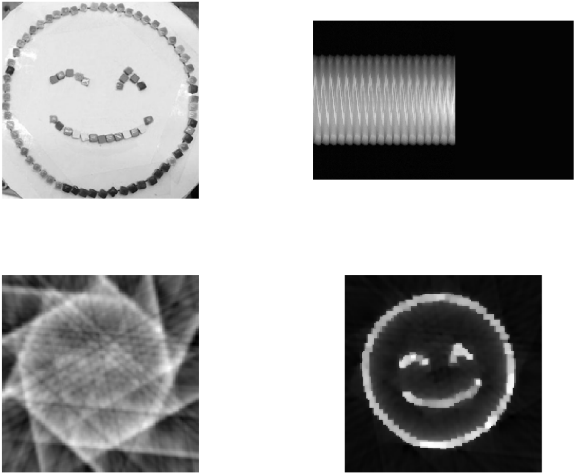
Time step = 17



Dynamic inverse problems

Applications to limited angle tomography

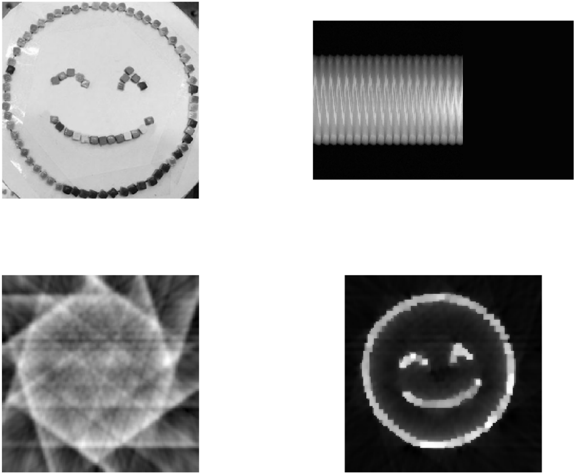
Time step = 18



Dynamic inverse problems

Applications to limited angle tomography

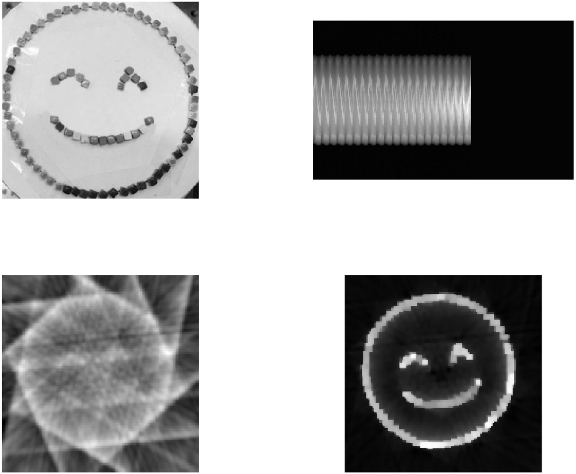
Time step = 19



Dynamic inverse problems

Applications to limited angle tomography

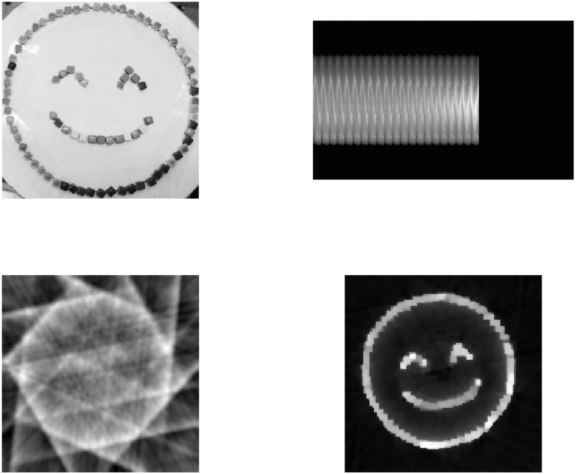
Time step = 20



Dynamic inverse problems

Applications to limited angle tomography

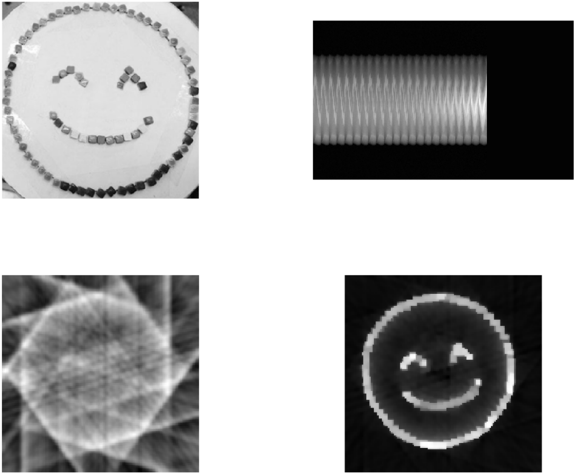
Time step = 21



Dynamic inverse problems

Applications to limited angle tomography

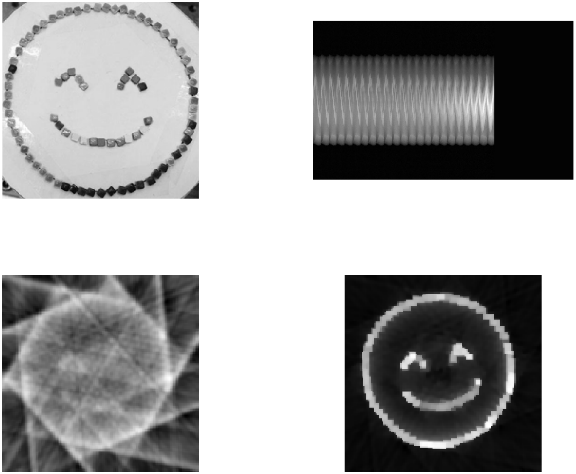
Time step = 22



Dynamic inverse problems

Applications to limited angle tomography

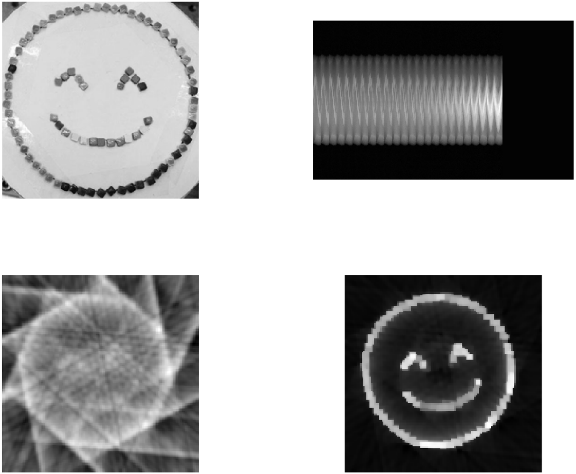
Time step = 23



Dynamic inverse problems

Applications to limited angle tomography

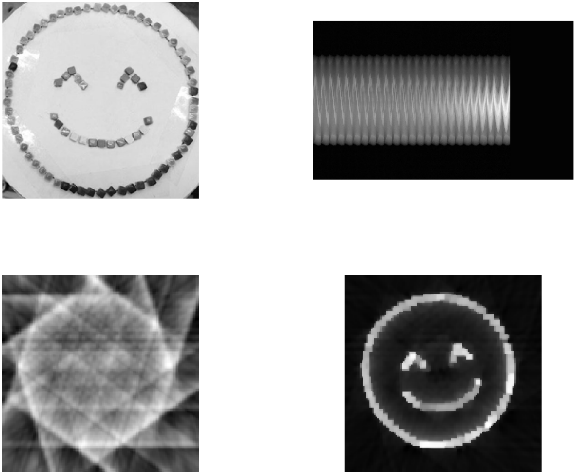
Time step = 24



Dynamic inverse problems

Applications to limited angle tomography

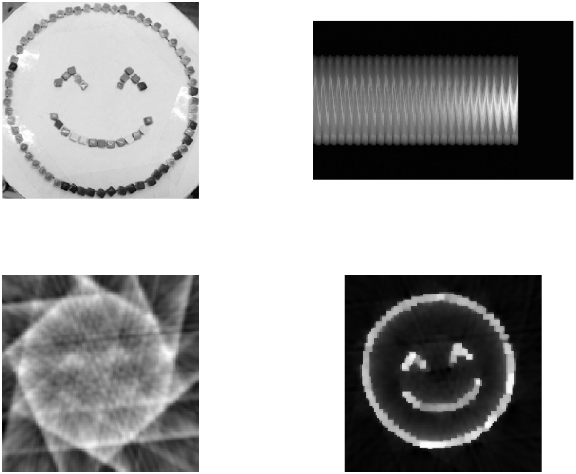
Time step = 25



Dynamic inverse problems

Applications to limited angle tomography

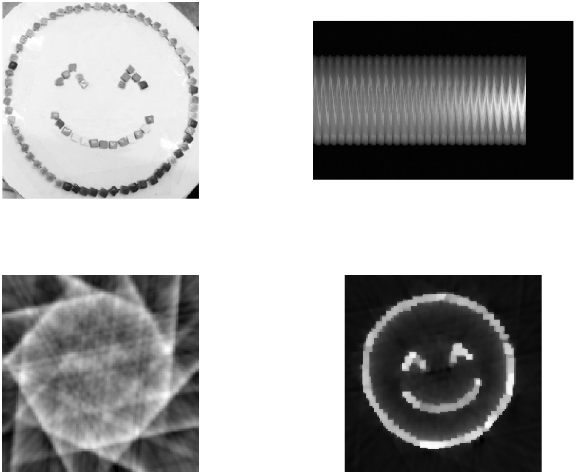
Time step = 26



Dynamic inverse problems

Applications to limited angle tomography

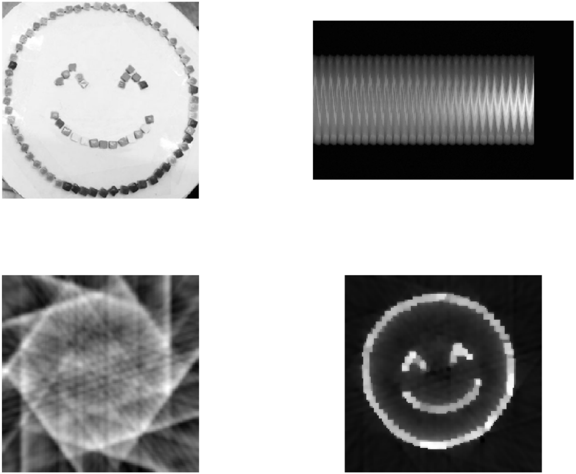
Time step = 27



Dynamic inverse problems

Applications to limited angle tomography

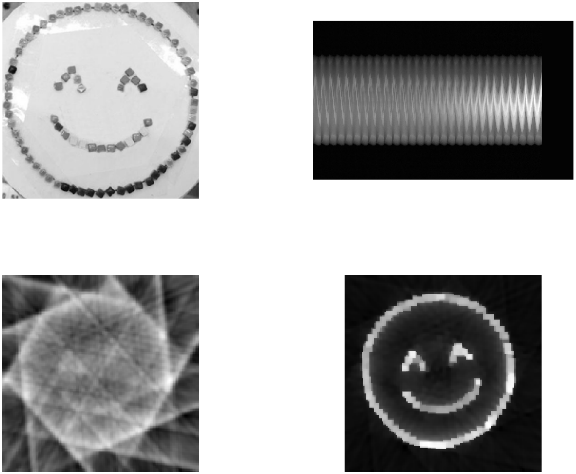
Time step = 28



Dynamic inverse problems

Applications to limited angle tomography

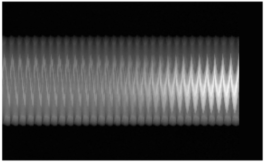
Time step = 29

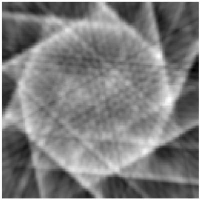


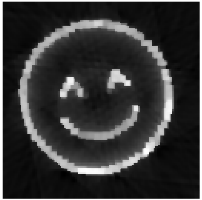
Dynamic inverse problems

Applications to limited angle tomography

Time step = 30



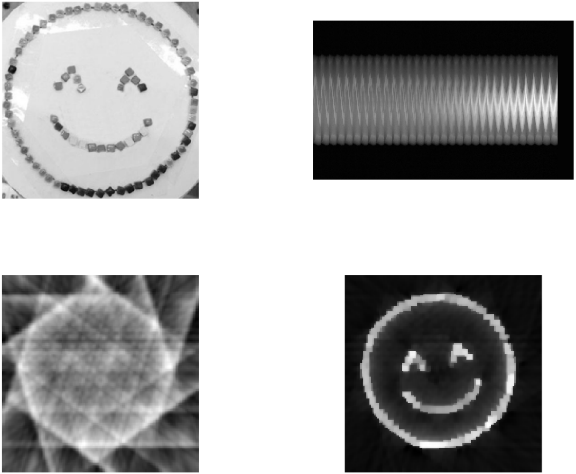




Dynamic inverse problems

Applications to limited angle tomography

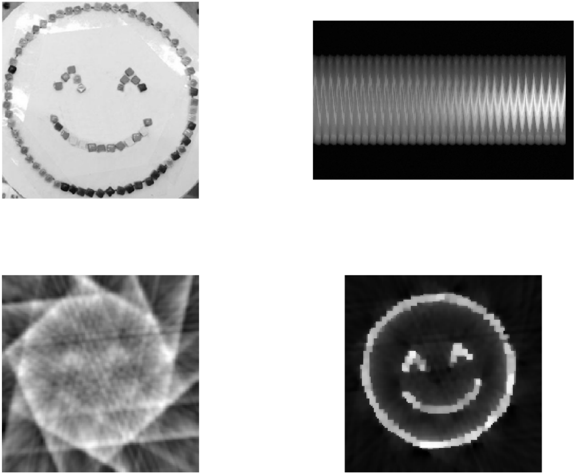
Time step = 31



Dynamic inverse problems

Applications to limited angle tomography

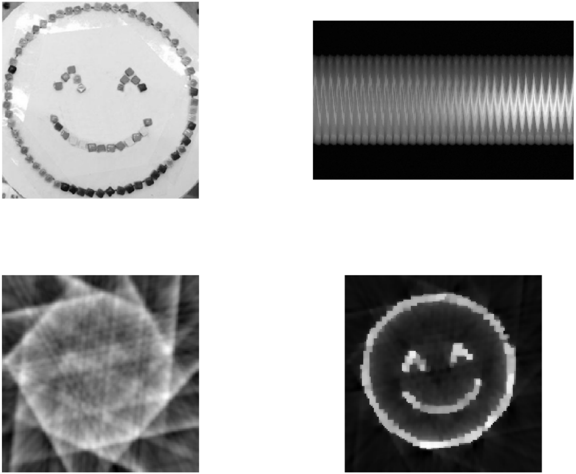
Time step = 32



Dynamic inverse problems

Applications to limited angle tomography

Time step = 33



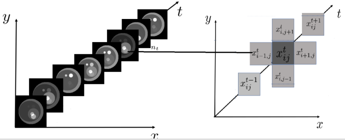
A new problem set up

Dynamic time-dependent inverse problem

$$\underbrace{\begin{bmatrix} \mathbf{A}^{(1)} & & \\ & \ddots & \\ & & \mathbf{A}^{(n_t)} \end{bmatrix}}_{\mathbf{F}} \underbrace{\begin{bmatrix} \mathbf{u}^{(1)} \\ \vdots \\ \mathbf{u}^{(n_t)} \end{bmatrix}}_{\mathbf{u}} + \underbrace{\begin{bmatrix} \mathbf{e}^{(1)} \\ \vdots \\ \mathbf{e}^{(n_t)} \end{bmatrix}}_{\mathbf{e}} = \underbrace{\begin{bmatrix} \mathbf{d}^{(1)} \\ \vdots \\ \mathbf{d}^{(n_t)} \end{bmatrix}}_{\mathbf{d}}$$

$$\mathbf{F} \in \mathbb{R}^{6553600 \times 6553600}, \quad \mathbf{u} \in \mathbb{R}^{6553600}, \quad \mathbf{d} \in \mathbb{R}^{6553600}$$

$$\mathcal{R}_1(\mathbf{u}) = \lambda_s^2 \sum_{t=1}^{n_t} \|\mathbf{L}_s \mathbf{u}^{(t)}\|_1 + \lambda_t^2 \sum_{t=1}^{n_t - 1} \|\mathbf{u}^{(t+1)} - \mathbf{u}^{(t)}\|_1$$



- ⚠ Millions of parameters
- ⚠ Solutions with edges
- ⚠ Discover dynamics of the data

Methods based on Total Variation

Space time total variation – modeling the regularization term

- Let $\mathbf{L}_s$ be a matrix that represents the discretized finite difference operator corresponding to the first derivative.

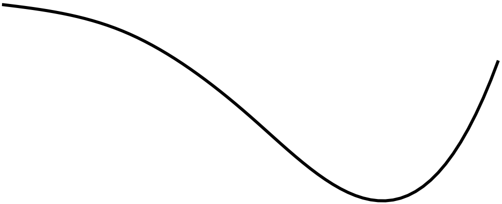
$$\mathbf{L}_s = \begin{bmatrix} \mathbf{I} \otimes \mathbf{L}_v \\ \mathbf{L}_h \otimes \mathbf{I} \end{bmatrix},$$

- $\mathbf{L}_v \in \mathbb{R}^{(n_v-1)\times n_v}$, $\mathbf{L}_h \in \mathbb{R}^{(n_h-1)\times n_h}$ represent the discretized first derivative operators in the v- and h-directions respectively.
- The discrete total variation (TV) norm $||\mathbf{L}_s\mathbf{u}^{(t)}||^1$ – sparse gradients.

$$\begin{aligned} \mathbf{u}_\lambda &= \arg\min_{\mathbf{u}} \frac{1}{2}||\mathbf{F}\mathbf{u}-\mathbf{b}||_2^2 + \mathcal{R}_1(\mathbf{u}) \\ \mathcal{R}_1(\mathbf{u}) &= \lambda_s^2 \sum_{t=1}^{n_t} ||\mathbf{L}_s\mathbf{u}^{(t)}||_1 + \lambda_t^2 \sum_{t=1}^{n_t-1} ||\mathbf{u}^{(t+1)} - \mathbf{u}^{(t)}||_1 \\ &= \lambda_s^2 ||(\mathbf{I}_{n_t} \otimes \mathbf{L}_s)\mathbf{u}||_1 + \lambda_t^2 ||(\mathbf{L}_t \otimes \mathbf{I}_{n_v n_h})\mathbf{u}||_1. \end{aligned}$$

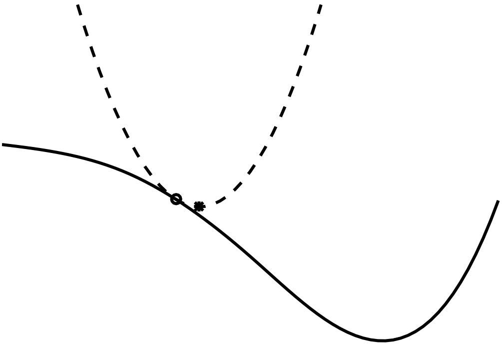
$\ell_2 - \ell_q$ minimization by MM-GKS

- Construct a sequence $\mathbf{u}^{(k)}$ that converges to a stationary point of $\mathcal{J}_{\lambda,q}(\mathbf{u})$.
- At each step the functional $\mathcal{J}_{\lambda,q}(\mathbf{u})$ is majorized by a quadratic function $\mathcal{Q}_{\mathbf{u}^{(k)}}(\mathbf{u})$ that is tangent to $\mathcal{J}_{\lambda,q}(\mathbf{u})$ at $\mathbf{u}^{(k)}$.
- The next iterate $\mathbf{u}^{(k+1)}$ is the unique minimizer of $\mathcal{Q}_{\mathbf{u}^{(k)}}(\mathbf{u})$.



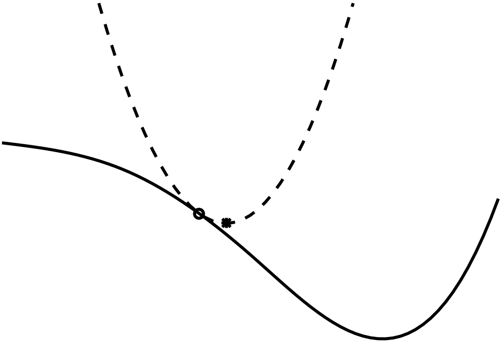
$\ell_2 - \ell_q$ minimization by MM-GKS

- Construct a sequence $\mathbf{u}^{(k)}$ that converges to a stationary point of $\mathcal{J}_{\lambda,q}(\mathbf{u})$.
- At each step the functional $\mathcal{J}_{\lambda,q}(\mathbf{u})$ is majorized by a quadratic function $\mathcal{Q}_{\mathbf{u}^{(k)}}(\mathbf{u})$ that is tangent to $\mathcal{J}_{\lambda,q}(\mathbf{u})$ at $\mathbf{u}^{(k)}$.
- The next iterate $\mathbf{u}^{(k+1)}$ is the unique minimizer of $\mathcal{Q}_{\mathbf{u}^{(k)}}(\mathbf{u})$.



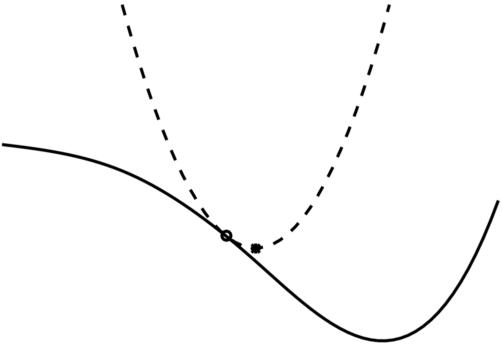
$\ell_2 - \ell_q$ minimization by MM-GKS

- Construct a sequence $\mathbf{u}^{(k)}$ that converges to a stationary point of $\mathcal{J}_{\lambda,q}(\mathbf{u})$.
- At each step the functional $\mathcal{J}_{\lambda,q}(\mathbf{u})$ is majorized by a quadratic function $\mathcal{Q}_{\mathbf{u}^{(k)}}(\mathbf{u})$ that is tangent to $\mathcal{J}_{\lambda,q}(\mathbf{u})$ at $\mathbf{u}^{(k)}$.
- The next iterate $\mathbf{u}^{(k+1)}$ is the unique minimizer of $\mathcal{Q}_{\mathbf{u}^{(k)}}(\mathbf{u})$.



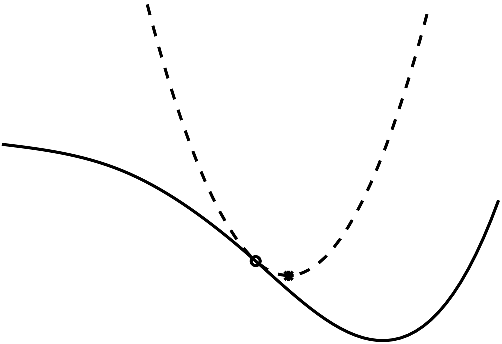
$\ell_2 - \ell_q$ minimization by MM-GKS

- Construct a sequence $\mathbf{u}^{(k)}$ that converges to a stationary point of $\mathcal{J}_{\lambda,q}(\mathbf{u})$.
- At each step the functional $\mathcal{J}_{\lambda,q}(\mathbf{u})$ is majorized by a quadratic function $\mathcal{Q}_{\mathbf{u}^{(k)}}(\mathbf{u})$ that is tangent to $\mathcal{J}_{\lambda,q}(\mathbf{u})$ at $\mathbf{u}^{(k)}$.
- The next iterate $\mathbf{u}^{(k+1)}$ is the unique minimizer of $\mathcal{Q}_{\mathbf{u}^{(k)}}(\mathbf{u})$.



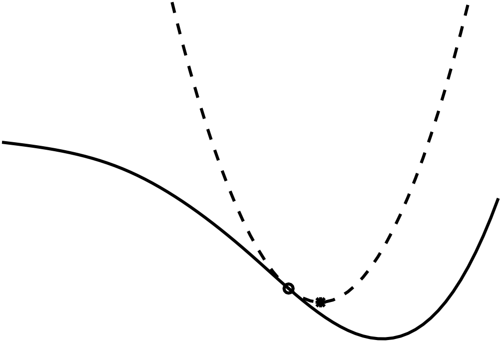
$\ell_2 - \ell_q$ minimization by MM-GKS

- Construct a sequence $\mathbf{u}^{(k)}$ that converges to a stationary point of $\mathcal{J}_{\lambda,q}(\mathbf{u})$.
- At each step the functional $\mathcal{J}_{\lambda,q}(\mathbf{u})$ is majorized by a quadratic function $\mathcal{Q}_{\mathbf{u}^{(k)}}(\mathbf{u})$ that is tangent to $\mathcal{J}_{\lambda,q}(\mathbf{u})$ at $\mathbf{u}^{(k)}$.
- The next iterate $\mathbf{u}^{(k+1)}$ is the unique minimizer of $\mathcal{Q}_{\mathbf{u}^{(k)}}(\mathbf{u})$.



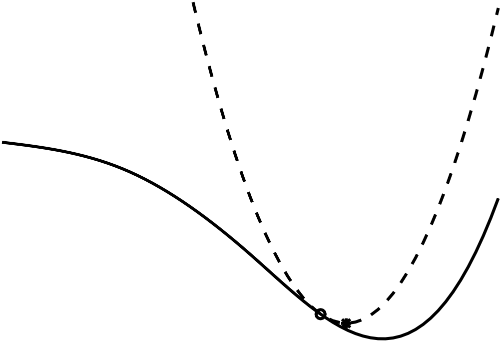
$\ell_2 - \ell_q$ minimization by MM-GKS

- Construct a sequence $\mathbf{u}^{(k)}$ that converges to a stationary point of $\mathcal{J}_{\lambda,q}(\mathbf{u})$.
- At each step the functional $\mathcal{J}_{\lambda,q}(\mathbf{u})$ is majorized by a quadratic function $\mathcal{Q}_{\mathbf{u}^{(k)}}(\mathbf{u})$ that is tangent to $\mathcal{J}_{\lambda,q}(\mathbf{u})$ at $\mathbf{u}^{(k)}$.
- The next iterate $\mathbf{u}^{(k+1)}$ is the unique minimizer of $\mathcal{Q}_{\mathbf{u}^{(k)}}(\mathbf{u})$.



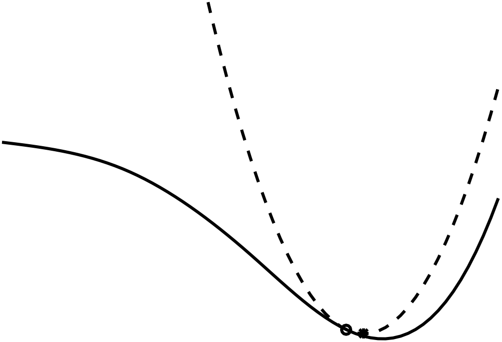
$\ell_2 - \ell_q$ minimization by MM-GKS

- Construct a sequence $\mathbf{u}^{(k)}$ that converges to a stationary point of $\mathcal{J}_{\lambda,q}(\mathbf{u})$.
- At each step the functional $\mathcal{J}_{\lambda,q}(\mathbf{u})$ is majorized by a quadratic function $\mathcal{Q}_{\mathbf{u}^{(k)}}(\mathbf{u})$ that is tangent to $\mathcal{J}_{\lambda,q}(\mathbf{u})$ at $\mathbf{u}^{(k)}$.
- The next iterate $\mathbf{u}^{(k+1)}$ is the unique minimizer of $\mathcal{Q}_{\mathbf{u}^{(k)}}(\mathbf{u})$.



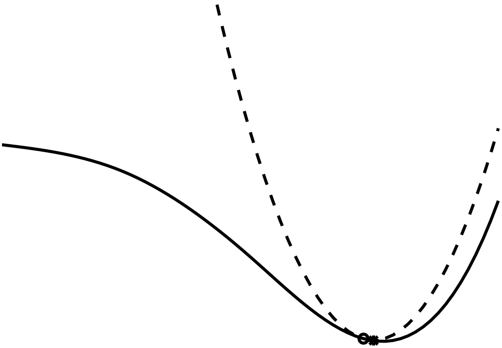
$\ell_2 - \ell_q$ minimization by MM-GKS

- Construct a sequence $\mathbf{u}^{(k)}$ that converges to a stationary point of $\mathcal{J}_{\lambda,q}(\mathbf{u})$.
- At each step the functional $\mathcal{J}_{\lambda,q}(\mathbf{u})$ is majorized by a quadratic function $\mathcal{Q}_{\mathbf{u}^{(k)}}(\mathbf{u})$ that is tangent to $\mathcal{J}_{\lambda,q}(\mathbf{u})$ at $\mathbf{u}^{(k)}$.
- The next iterate $\mathbf{u}^{(k+1)}$ is the unique minimizer of $\mathcal{Q}_{\mathbf{u}^{(k)}}(\mathbf{u})$.



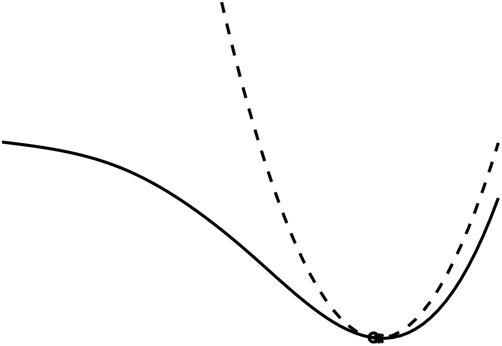
$\ell_2 - \ell_q$ minimization by MM-GKS

- Construct a sequence $\mathbf{u}^{(k)}$ that converges to a stationary point of $\mathcal{J}_{\lambda,q}(\mathbf{u})$.
- At each step the functional $\mathcal{J}_{\lambda,q}(\mathbf{u})$ is majorized by a quadratic function $\mathcal{Q}_{\mathbf{u}^{(k)}}(\mathbf{u})$ that is tangent to $\mathcal{J}_{\lambda,q}(\mathbf{u})$ at $\mathbf{u}^{(k)}$.
- The next iterate $\mathbf{u}^{(k+1)}$ is the unique minimizer of $\mathcal{Q}_{\mathbf{u}^{(k)}}(\mathbf{u})$.



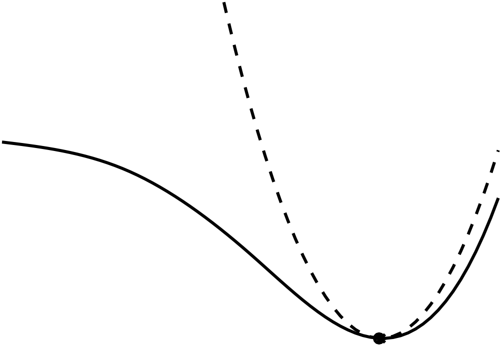
$\ell_2 - \ell_q$ minimization by MM-GKS

- Construct a sequence $\mathbf{u}^{(k)}$ that converges to a stationary point of $\mathcal{J}_{\lambda,q}(\mathbf{u})$.
- At each step the functional $\mathcal{J}_{\lambda,q}(\mathbf{u})$ is majorized by a quadratic function $\mathcal{Q}_{\mathbf{u}^{(k)}}(\mathbf{u})$ that is tangent to $\mathcal{J}_{\lambda,q}(\mathbf{u})$ at $\mathbf{u}^{(k)}$.
- The next iterate $\mathbf{u}^{(k+1)}$ is the unique minimizer of $\mathcal{Q}_{\mathbf{u}^{(k)}}(\mathbf{u})$.



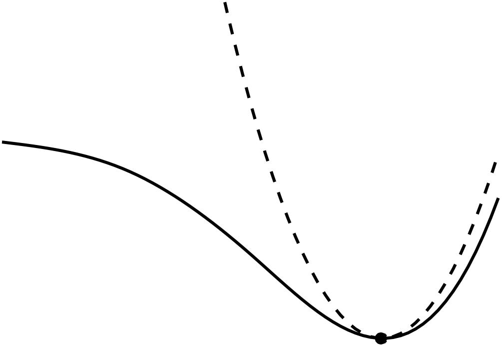
$\ell_2 - \ell_q$ minimization by MM-GKS

- Construct a sequence $\mathbf{u}^{(k)}$ that converges to a stationary point of $\mathcal{J}_{\lambda,q}(\mathbf{u})$.
- At each step the functional $\mathcal{J}_{\lambda,q}(\mathbf{u})$ is majorized by a quadratic function $\mathcal{Q}_{\mathbf{u}^{(k)}}(\mathbf{u})$ that is tangent to $\mathcal{J}_{\lambda,q}(\mathbf{u})$ at $\mathbf{u}^{(k)}$.
- The next iterate $\mathbf{u}^{(k+1)}$ is the unique minimizer of $\mathcal{Q}_{\mathbf{u}^{(k)}}(\mathbf{u})$.



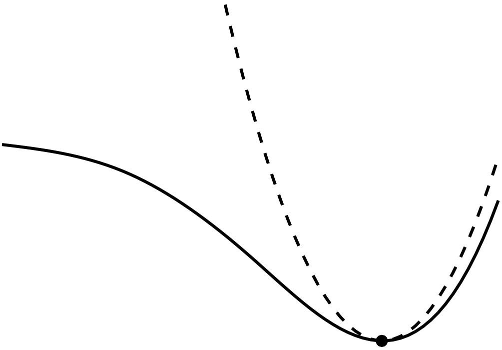
$\ell_2 - \ell_q$ minimization by MM-GKS

- Construct a sequence $\mathbf{u}^{(k)}$ that converges to a stationary point of $\mathcal{J}_{\lambda,q}(\mathbf{u})$.
- At each step the functional $\mathcal{J}_{\lambda,q}(\mathbf{u})$ is majorized by a quadratic function $\mathcal{Q}_{\mathbf{u}^{(k)}}(\mathbf{u})$ that is tangent to $\mathcal{J}_{\lambda,q}(\mathbf{u})$ at $\mathbf{u}^{(k)}$.
- The next iterate $\mathbf{u}^{(k+1)}$ is the unique minimizer of $\mathcal{Q}_{\mathbf{u}^{(k)}}(\mathbf{u})$.



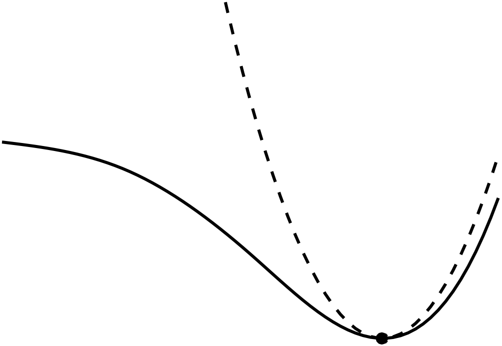
$\ell_2 - \ell_q$ minimization by MM-GKS

- Construct a sequence $\mathbf{u}^{(k)}$ that converges to a stationary point of $\mathcal{J}_{\lambda,q}(\mathbf{u})$.
- At each step the functional $\mathcal{J}_{\lambda,q}(\mathbf{u})$ is majorized by a quadratic function $\mathcal{Q}_{\mathbf{u}^{(k)}}(\mathbf{u})$ that is tangent to $\mathcal{J}_{\lambda,q}(\mathbf{u})$ at $\mathbf{u}^{(k)}$.
- The next iterate $\mathbf{u}^{(k+1)}$ is the unique minimizer of $\mathcal{Q}_{\mathbf{u}^{(k)}}(\mathbf{u})$.



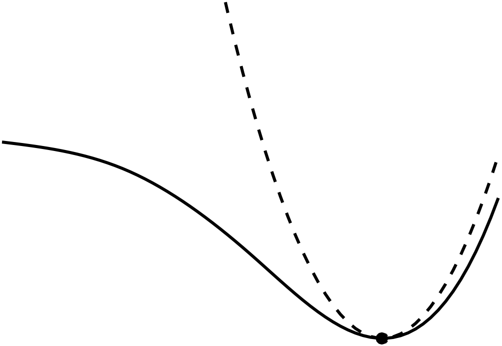
$\ell_2 - \ell_q$ minimization by MM-GKS

- Construct a sequence $\mathbf{u}^{(k)}$ that converges to a stationary point of $\mathcal{J}_{\lambda,q}(\mathbf{u})$.
- At each step the functional $\mathcal{J}_{\lambda,q}(\mathbf{u})$ is majorized by a quadratic function $\mathcal{Q}_{\mathbf{u}^{(k)}}(\mathbf{u})$ that is tangent to $\mathcal{J}_{\lambda,q}(\mathbf{u})$ at $\mathbf{u}^{(k)}$.
- The next iterate $\mathbf{u}^{(k+1)}$ is the unique minimizer of $\mathcal{Q}_{\mathbf{u}^{(k)}}(\mathbf{u})$.



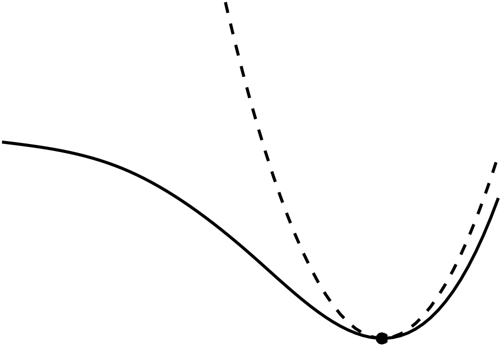
$\ell_2 - \ell_q$ minimization by MM-GKS

- Construct a sequence $\mathbf{u}^{(k)}$ that converges to a stationary point of $\mathcal{J}_{\lambda,q}(\mathbf{u})$.
- At each step the functional $\mathcal{J}_{\lambda,q}(\mathbf{u})$ is majorized by a quadratic function $\mathcal{Q}_{\mathbf{u}^{(k)}}(\mathbf{u})$ that is tangent to $\mathcal{J}_{\lambda,q}(\mathbf{u})$ at $\mathbf{u}^{(k)}$.
- The next iterate $\mathbf{u}^{(k+1)}$ is the unique minimizer of $\mathcal{Q}_{\mathbf{u}^{(k)}}(\mathbf{u})$.



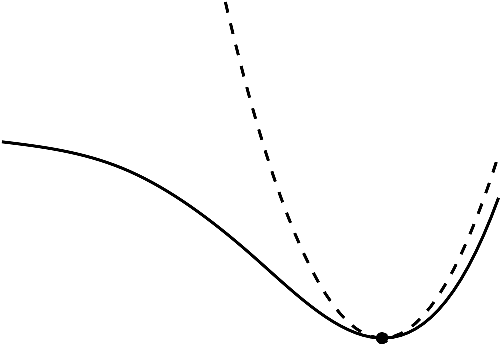
$\ell_2 - \ell_q$ minimization by MM-GKS

- Construct a sequence $\mathbf{u}^{(k)}$ that converges to a stationary point of $\mathcal{J}_{\lambda,q}(\mathbf{u})$.
- At each step the functional $\mathcal{J}_{\lambda,q}(\mathbf{u})$ is majorized by a quadratic function $\mathcal{Q}_{\mathbf{u}^{(k)}}(\mathbf{u})$ that is tangent to $\mathcal{J}_{\lambda,q}(\mathbf{u})$ at $\mathbf{u}^{(k)}$.
- The next iterate $\mathbf{u}^{(k+1)}$ is the unique minimizer of $\mathcal{Q}_{\mathbf{u}^{(k)}}(\mathbf{u})$.



$\ell_2 - \ell_q$ minimization by MM-GKS

- Construct a sequence $\mathbf{u}^{(k)}$ that converges to a stationary point of $\mathcal{J}_{\lambda,q}(\mathbf{u})$.
- At each step the functional $\mathcal{J}_{\lambda,q}(\mathbf{u})$ is majorized by a quadratic function $\mathcal{Q}_{\mathbf{u}^{(k)}}(\mathbf{u})$ that is tangent to $\mathcal{J}_{\lambda,q}(\mathbf{u})$ at $\mathbf{u}^{(k)}$.
- The next iterate $\mathbf{u}^{(k+1)}$ is the unique minimizer of $\mathcal{Q}_{\mathbf{u}^{(k)}}(\mathbf{u})$.



$\ell_2 - \ell_q$ minimization by MM-GKS [7]

◇ Smooth $\Phi_{z,\varepsilon}(t) = t$ by

$$\Phi_{z,\varepsilon}(t) = \left(t^2 + \varepsilon^2\right)^{z/2} \quad \text{with} \quad \begin{cases} \varepsilon > 0 \ \text{ for } \ 0 < z \leq 1, \\ \varepsilon = 0 \ \text{ for } \ z > 1, \end{cases}$$

◇ Consider the functional

$$\min_{\mathbf{x}} \mathcal{J}_{\lambda,q,\varepsilon}(\mathbf{u}) = \min_{\mathbf{u}} \|\mathbf{Au} - \mathbf{b}\|_2^2 + \lambda \sum_{j=1}^{n} \phi_{q,\varepsilon}((\mathbf{Lu})_j).$$

◇ Compute the quadratic tangent majorant for $\mathcal{J}_{\lambda,q,\varepsilon}$ at $\mathbf{u}^{(k)}$, (c-arbitrary constant)

$$\mathcal{Q}_{\mathbf{u}^{(k)}}(\mathbf{u}) = \frac{1}{2}\|\mathbf{Au} - \mathbf{b}\|_2^2 + \frac{\lambda}{2}\|\mathbf{P}_{q,\varepsilon}^{(k)}\mathbf{Lu}\|_2^2 + c$$

◇ Find the minimizer of $\mathcal{Q}$

$$(\mathbf{A}^T\mathbf{A} + \lambda\mathbf{L}^T\mathbf{P}_{q,\varepsilon}^{(k)}\mathbf{L})\mathbf{u}^{(k+1)} = \mathbf{A}^T\mathbf{b},$$

[7] Lanza, Morigi, Reichel, Sgallari (2015), Buccini, Pasha, Reichel (2020), Buccini, Reichel (2021)

Remedy the large-dimension issue

Krylov subspaces as dimension reduction and regularization methods

Given $\mathbf{A} \in \mathbb{R}^{m \times n}$, where m, n - large

Task: Goal: Solve $\min_{\mathbf{u}} \|\mathbf{A}\mathbf{u} - \mathbf{b}\|_2^2$

$\mathcal{K}_d(\mathbf{A}^T\mathbf{A}, \mathbf{A}^T\mathbf{b}) = span\{\mathbf{A}^T\mathbf{b}, \ldots, (\mathbf{A}^T\mathbf{A})^{d-1}\mathbf{A}^T\mathbf{b}\}$. $\mathbf{A}\mathbf{V}_d = \mathbf{U}_{d+1}\mathbf{B}_{d+1,d}$, $\mathbf{A}^T\mathbf{U}_{d+1} = \mathbf{V}_d\mathbf{B}_{d,d}^T$

New task: $\min_{\mathbf{u} \in \mathcal{K}_d(\mathbf{A}^T\mathbf{A}, \mathbf{A}^T\mathbf{b})} \|\mathbf{A}\mathbf{u} - \mathbf{b}\|_2^2 = \min_{\mathbf{y} \in \mathbb{R}^d} \|\mathbf{B}_{d+1,d}\mathbf{y} - \|\mathbf{b}\|_2 e_1\|_2^2$

$\ell_2 - \ell_q$ minimization by MM-GKS to solve the linear problem
continued

STEP 1: Generate the starting subspace.

1. The GKS method first determines an initial reduction of $\mathbf{A}$ to a small bidiagonal matrix by applying $1 \leq \ell \ll \min\{m, n\}$ steps of Golub–Kahan and get

$$\mathbf{A}\mathbf{V}_0 = \mathbf{U}_0\mathbf{B}_0.$$

   - $\mathbf{B}_0 \in \mathbb{R}^{(\ell+1)\times\ell}$ is lower bidiagonal.
   - $\mathbf{V}_0, \mathbf{U}_0$ have orthonormal columns.

2. The subspace $K_\ell(\mathbf{A}^T\mathbf{A}, \mathbf{A}^T\mathbf{b}) = \text{span}\{\mathbf{A}^T\mathbf{b}, (\mathbf{A}^T\mathbf{A})\mathbf{A}^T\mathbf{b}, \ldots, (\mathbf{A}^T\mathbf{A})^{\ell-1}\mathbf{A}^T\mathbf{b}\}$ is generated. Compute QR factorizations
   - $\mathbf{A}\mathbf{V}_0 = \mathbf{Q}_\mathbf{A}\mathbf{R}_\mathbf{A}$
   - $\mathbf{L}\mathbf{V}_0 = \mathbf{Q}_\mathbf{L}\mathbf{R}_\mathbf{L}$

$$\mathbf{z}^{(1)} = \arg\min_{\mathbf{z}\in\mathbb{R}^\ell} \left\| \begin{bmatrix} \mathbf{R}_\mathbf{A} \\ \eta^{1/2}\mathbf{R}_\mathbf{L} \end{bmatrix} \mathbf{z} - \begin{bmatrix} \mathbf{Q}_\mathbf{A}^T\mathbf{b} \\ 0 \end{bmatrix} \right\|_2^2, \quad \mathbf{x}^{(1)} = \mathbf{V}_0\mathbf{z}^{(1)}$$

$\ell_2 - \ell_q$ minimization by MM-GKS to solve the linear problem

continued

STEP 1: Generate the starting subspace.

1. The GKS method first determines an initial reduction of $\mathbf{A}$ to a small bidiagonal matrix by applying $1 \leq \ell \ll \min\{m, n\}$ steps of Golub–Kahan and get

$$\mathbf{A}\mathbf{V}_0 = \mathbf{U}_0\mathbf{B}_0.$$

   - $\mathbf{B}_0 \in \mathbb{R}^{(\ell+1)\times\ell}$ is lower bidiagonal.
   - $\mathbf{V}_0$, $\mathbf{U}_0$ have orthonormal columns.

2. The subspace $K_\ell(\mathbf{A}^T\mathbf{A}, \mathbf{A}^T\mathbf{b}) = \text{span}\{\mathbf{A}^T\mathbf{b}, (\mathbf{A}^T\mathbf{A})\mathbf{A}^T\mathbf{b}, \dots, (\mathbf{A}^T\mathbf{A})^{\ell-1}\mathbf{A}^T\mathbf{b}\}$ is generated. Compute QR factorizations
   - $\mathbf{A}\mathbf{V}_0 = \mathbf{Q}_\mathbf{A}\mathbf{R}_\mathbf{A}$
   - $\mathbf{L}\mathbf{V}_0 = \mathbf{Q}_\mathbf{L}\mathbf{R}_\mathbf{L}$

$$\mathbf{z}^{(1)} = \arg\min_{\mathbf{z}\in\mathbb{R}^\ell} \left\| \begin{bmatrix} \mathbf{R}_\mathbf{A} \\ \eta^{1/2}\mathbf{R}_\mathbf{L} \end{bmatrix} \mathbf{z} - \begin{bmatrix} \mathbf{Q}_\mathbf{A}^T\mathbf{b} \\ 0 \end{bmatrix} \right\|_2^2, \quad \mathbf{x}^{(1)} = \mathbf{V}_0\mathbf{z}^{(1)}$$

3. Enlarge the subspace $\mathbf{A}\mathbf{V}_{\ell+1} = [\mathbf{A}\mathbf{V}_\ell, \mathbf{A}\mathbf{v}_{\text{new}}]$, $\quad \mathbf{L}\mathbf{V}_{\ell+1} = [\mathbf{L}\mathbf{V}_\ell, \mathbf{L}\mathbf{v}_{\text{new}}]$
   - (By the residual)

$$\mathbf{r}^{(1)} = \mathbf{A}^T(\mathbf{A}\mathbf{V}_\ell\mathbf{z}^{(1)} - \mathbf{d}) + \eta\mathbf{L}^T\mathbf{P}_{q,\varepsilon}^{(0)}\mathbf{L}\mathbf{V}_\ell\mathbf{z}^{(1)}$$

$\ell_2 - \ell_q$ minimization by MM-GKS to solve the linear problem

Continued

GENERAL STEP k:

1. Let $\mathbf{V}_k \in \mathbb{R}^{n\times\hat{k}}$ form an orthonormal basis of the solution subspace.

$\ell_2 - \ell_q$ minimization by MM-GKS to solve the linear problem
Continued

GENERAL STEP k:

1. Let $\mathbf{V}_k \in \mathbb{R}^{n\times\hat{k}}$ form an orthonormal basis of the solution subspace.
2. Introduce the QR factorizations

$$\mathbf{A}\mathbf{V}_k = \mathbf{Q}_\mathbf{A}\mathbf{R}_\mathbf{A} \quad \text{with} \quad \mathbf{Q}_\mathbf{A} \in \mathbb{R}^{m\times\hat{k}}, \;\; \mathbf{R}_\mathbf{A} \in \mathbb{R}^{\hat{k}\times\hat{k}},$$
$$(\mathbf{P}_{q,\varepsilon}^{(k)})^{1/2}\mathbf{L}\mathbf{V}_k = \mathbf{Q}_\mathbf{L}\mathbf{R}_\mathbf{L} \quad \text{with} \quad \mathbf{Q}_\mathbf{L} \in \mathbb{R}^{\ell\times\hat{k}}, \;\; \mathbf{R}_\mathbf{L} \in \mathbb{R}^{\hat{k}\times\hat{k}}.$$

$\ell_2 - \ell_q$ minimization by MM-GKS to solve the linear problem
Continued

GENERAL STEP k:

1. Let $\mathbf{V}_k \in \mathbb{R}^{n\times \hat{k}}$ form an orthonormal basis of the solution subspace.
2. Introduce the QR factorizations

$$\mathbf{A}\mathbf{V}_k = \mathbf{Q}_\mathbf{A}\mathbf{R}_\mathbf{A} \quad \text{with} \quad \mathbf{Q}_\mathbf{A} \in \mathbb{R}^{m\times \hat{k}},\ \mathbf{R}_\mathbf{A} \in \mathbb{R}^{\hat{k}\times \hat{k}},$$
$$(\mathbf{P}_{q,\varepsilon}^{(k)})^{1/2}\mathbf{L}\mathbf{V}_k = \mathbf{Q}_\mathbf{L}\mathbf{R}_\mathbf{L} \quad \text{with} \quad \mathbf{Q}_\mathbf{L} \in \mathbb{R}^{\ell\times \hat{k}},\ \mathbf{R}_\mathbf{L} \in \mathbb{R}^{\hat{k}\times \hat{k}}.$$

3.
$$\mathbf{z}^{(k+1)} = \arg\min_{\mathbf{z}\in\mathbb{R}^\ell} \left\| \begin{bmatrix} \mathbf{R}_\mathbf{A} \\ \eta^{1/2}\mathbf{R}_\mathbf{L} \end{bmatrix} \mathbf{z} - \begin{bmatrix} \mathbf{Q}_\mathbf{A}^T\mathbf{b} \\ 0 \end{bmatrix} \right\|_2^2.$$

$\ell_2 - \ell_q$ minimization by MM-GKS to solve the linear problem

Continued

GENERAL STEP k:

1. Let $\mathbf{V}_k \in \mathbb{R}^{n\times\hat{k}}$ form an orthonormal basis of the solution subspace.
2. Introduce the QR factorizations

$$\mathbf{A}\mathbf{V}_k = \mathbf{Q}_{\mathbf{A}}\mathbf{R}_{\mathbf{A}} \quad \text{with} \quad \mathbf{Q}_{\mathbf{A}} \in \mathbb{R}^{m\times\hat{k}}, \ \ \mathbf{R}_{\mathbf{A}} \in \mathbb{R}^{\hat{k}\times\hat{k}},$$
$$(\mathbf{P}_{q,\varepsilon}^{(k)})^{1/2}\mathbf{L}\mathbf{V}_k = \mathbf{Q}_{\mathbf{L}}\mathbf{R}_{\mathbf{L}} \quad \text{with} \quad \mathbf{Q}_{\mathbf{L}} \in \mathbb{R}^{\ell\times\hat{k}}, \ \ \mathbf{R}_{\mathbf{L}} \in \mathbb{R}^{\hat{k}\times\hat{k}}.$$

3. $$\mathbf{z}^{(k+1)} = \arg\min_{\mathbf{z}\in\mathbb{R}^{\ell}} \left\| \begin{bmatrix} \mathbf{R}_{\mathbf{A}} \\ \eta^{1/2}\mathbf{R}_{\mathbf{L}} \end{bmatrix} \mathbf{z} - \begin{bmatrix} \mathbf{Q}_{\mathbf{A}}^T\mathbf{b} \\ 0 \end{bmatrix} \right\|_2^2 .$$

4. Compute the residual

$$\mathbf{r}^{(k+1)} = \mathbf{A}^T(\mathbf{A}\mathbf{V}_{\ell}\mathbf{z}^{(k+1)} - \mathbf{b}) + \eta\mathbf{L}^T\mathbf{P}_{q,\varepsilon}^{(k)}\mathbf{L}\mathbf{V}_{\ell}\mathbf{z}^{(k+1)}.$$

$\ell_2 - \ell_q$ minimization by MM-GKS to solve the linear problem

Continued

GENERAL STEP k:

1. Let $\mathbf{V}_k \in \mathbb{R}^{n\times\hat{k}}$ form an orthonormal basis of the solution subspace.
2. Introduce the QR factorizations

$$\mathbf{A}\mathbf{V}_k = \mathbf{Q}_\mathbf{A}\mathbf{R}_\mathbf{A} \quad \text{with} \quad \mathbf{Q}_\mathbf{A} \in \mathbb{R}^{m\times\hat{k}}, \;\; \mathbf{R}_\mathbf{A} \in \mathbb{R}^{\hat{k}\times\hat{k}},$$

$$(\mathbf{P}_{q,\varepsilon}^{(k)})^{1/2}\mathbf{L}\mathbf{V}_k = \mathbf{Q}_\mathbf{L}\mathbf{R}_\mathbf{L} \quad \text{with} \quad \mathbf{Q}_\mathbf{L} \in \mathbb{R}^{\ell\times\hat{k}}, \;\; \mathbf{R}_\mathbf{L} \in \mathbb{R}^{\hat{k}\times\hat{k}}.$$

3.

$$\mathbf{z}^{(k+1)} = \arg\min_{\mathbf{z}\in\mathbb{R}^\ell} \left\| \begin{bmatrix} \mathbf{R}_\mathbf{A} \\ \eta^{1/2}\mathbf{R}_\mathbf{L} \end{bmatrix} \mathbf{z} - \begin{bmatrix} \mathbf{Q}_\mathbf{A}^T\mathbf{b} \\ 0 \end{bmatrix} \right\|_2^2.$$

4. Compute the residual

$$\mathbf{r}^{(k+1)} = \mathbf{A}^T(\mathbf{A}\mathbf{V}_\ell\mathbf{z}^{(k+1)} - \mathbf{b}) + \eta\mathbf{L}^T\mathbf{P}_{q,\varepsilon}^{(k)}\mathbf{L}\mathbf{V}_\ell\mathbf{z}^{(k+1)}.$$

5. Expand the solution subspace $\mathbf{V}_{\ell+1} = [\mathbf{V}_\ell, \mathbf{v}_{\text{new}}] \in \mathbb{R}^{n\times(\hat{k}+1)}$.

$\ell_2 - \ell_q$ minimization by MM-GKS to solve the linear problem
Continued

GENERAL STEP k:

1. Let $\mathbf{V}_k \in \mathbb{R}^{n\times \hat{k}}$ form an orthonormal basis of the solution subspace.
2. Introduce the QR factorizations

$$\mathbf{A}\mathbf{V}_k = \mathbf{Q}_\mathbf{A}\mathbf{R}_\mathbf{A} \quad \text{with} \quad \mathbf{Q}_\mathbf{A} \in \mathbb{R}^{m\times \hat{k}}, \;\; \mathbf{R}_\mathbf{A} \in \mathbb{R}^{\hat{k}\times \hat{k}},$$
$$(\mathbf{P}_{q,\varepsilon}^{(k)})^{1/2}\mathbf{L}\mathbf{V}_k = \mathbf{Q}_\mathbf{L}\mathbf{R}_\mathbf{L} \quad \text{with} \quad \mathbf{Q}_\mathbf{L} \in \mathbb{R}^{\ell\times \hat{k}}, \;\; \mathbf{R}_\mathbf{L} \in \mathbb{R}^{\hat{k}\times \hat{k}}.$$

3.

$$\mathbf{z}^{(k+1)} = \arg\min_{\mathbf{z}\in\mathbb{R}^\ell} \left\| \begin{bmatrix} \mathbf{R}_\mathbf{A} \\ \eta^{1/2}\mathbf{R}_\mathbf{L} \end{bmatrix} \mathbf{z} - \begin{bmatrix} \mathbf{Q}_\mathbf{A}^T\mathbf{b} \\ 0 \end{bmatrix} \right\|_2^2.$$

4. Compute the residual

$$\mathbf{r}^{(k+1)} = \mathbf{A}^T(\mathbf{A}\mathbf{V}_\ell\mathbf{z}^{(k+1)} - \mathbf{b}) + \eta\mathbf{L}^T\mathbf{P}_{q,\varepsilon}^{(k)}\mathbf{L}\mathbf{V}_\ell\mathbf{z}^{(k+1)}.$$

5. Expand the solution subspace $\mathbf{V}_{\ell+1} = [\mathbf{V}_\ell, \mathbf{v}_{\text{new}}] \in \mathbb{R}^{n\times(\hat{k}+1)}$.

Other edge-preserving methods we propose

◇ Total variation in space and Tikhonov in time (TVplusTikhonov)

$$\mathcal{R}_2(\mathbf{u}) := \sum_{t=1}^{n_t} \|\mathbf{L}_s\mathbf{u}^{(t)}\|_1 + \sum_{t=1}^{n_t-1} \|\mathbf{u}^{(t+1)} - \mathbf{u}^{(t)}\|_2^2 = \|(\mathbf{I}_{n_t} \otimes \mathbf{L}_s)\mathbf{u}\|_1 + \|(\mathbf{L}_t \otimes \mathbf{I}_{n_s})\mathbf{u}\|_2^2$$

Other edge-preserving methods we propose

◇ Total variation in space and Tikhonov in time (TVplusTikhonov)

$$\mathcal{R}_2(\mathbf{u}) := \sum_{t=1}^{n_t} \|\mathbf{L}_s \mathbf{u}^{(t)}\|_1 + \sum_{t=1}^{n_t-1} \|\mathbf{u}^{(t+1)} - \mathbf{u}^{(t)}\|_2^2 = \|(\mathbf{I}_{n_t} \otimes \mathbf{L}_s)\mathbf{u}\|_1 + \|(\mathbf{L}_t \otimes \mathbf{I}_{n_s})\mathbf{u}\|_2^2$$

◇ Anisotropic space-time total variation (Aniso3DTV)

$$\boldsymbol{\mathcal{Y}} = \boldsymbol{\mathcal{U}} \times_1 \mathbf{L}_v \times_2 \mathbf{L}_h \times_3 \mathbf{L}_t, \quad \mathcal{R}_3(\mathbf{u}) = \|\boldsymbol{\mathcal{Y}}\|_1 = \sum_{v=1}^{n_v} \sum_{h=1}^{n_h} \sum_{t=1}^{n_t} |y_{v,h,t}|.$$

Other edge-preserving methods we propose

- Total variation in space and Tikhonov in time (TVplusTikhonov)

$$\mathcal{R}_2(\mathbf{u}) := \sum_{t=1}^{n_t} \|\mathbf{L}_s \mathbf{u}^{(t)}\|_1 + \sum_{t=1}^{n_t-1} \|\mathbf{u}^{(t+1)} - \mathbf{u}^{(t)}\|_2^2 = \|(\mathbf{I}_{n_t} \otimes \mathbf{L}_s)\mathbf{u}\|_1 + \|(\mathbf{L}_t \otimes \mathbf{I}_{n_s})\mathbf{u}\|_2^2$$

- Anisotropic space-time total variation (Aniso3DTV)

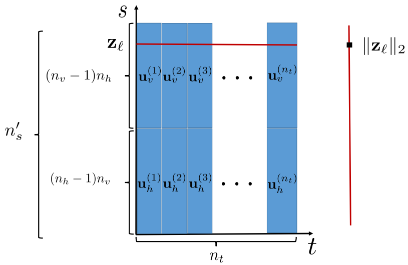
$$\mathcal{Y} = \mathcal{U} \times_1 \mathbf{L}_v \times_2 \mathbf{L}_h \times_3 \mathbf{L}_t, \quad \mathcal{R}_3(\mathbf{u}) = \|\mathcal{Y}\|_1 = \sum_{v=1}^{n_v}\sum_{h=1}^{n_h}\sum_{t=1}^{n_t} |y_{v,h,t}|.$$

- 3D space-time isotropic total variation (Iso3DTV)

$$\begin{aligned} \bar{\mathbf{z}}_v(\mathbf{u}) &:= (\mathbf{I}_{n_t} \otimes \mathbf{I}_{n_h} \otimes \bar{\mathbf{L}}_v)\mathbf{u}\,, \\ \bar{\mathbf{z}}_h(\mathbf{u}) &:= (\mathbf{I}_{n_t} \otimes \bar{\mathbf{L}}_h \otimes \mathbf{I}_{n_v})\mathbf{u}\,, \\ \bar{\mathbf{z}}_t(\mathbf{u}) &:= (\bar{\mathbf{L}}_t \otimes \mathbf{I}_{n_h} \otimes \mathbf{I}_{n_v})\mathbf{u}\,. \end{aligned}$$

$$\mathcal{R}_4(\mathbf{u}) := \sum_{\ell=1}^{n_v n_h n_t} \sqrt{(\bar{\mathbf{z}}_v(\mathbf{u}))_\ell^2 + (\bar{\mathbf{z}}_h(\mathbf{u}))_\ell^2 + (\bar{\mathbf{z}}_t(\mathbf{u}))_\ell^2} = \|\,[\bar{\mathbf{z}}_v(\mathbf{u}), \bar{\mathbf{z}}_h(\mathbf{u}), \bar{\mathbf{z}}_t(\mathbf{u})]\,\|_{2,1}.$$

Other edge-preserving methods we propose

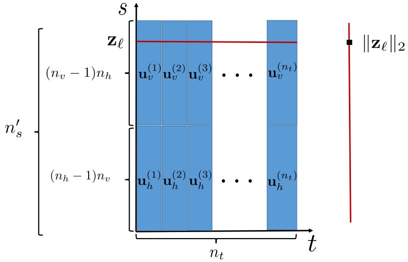


Group sparsity (GS)

$$\mathbf{z}_\ell = \left[(\mathbf{L}_s\mathbf{u}^{(1)})_\ell, \dots, (\mathbf{L}_s\mathbf{u}^{(n_t)})_\ell\right] \in \mathbb{R}^{n_t},$$

$$\ell = 1, \dots, n'_s, n'_s = (n_v-1)n_h + (n_h-1)n_v.$$

Other edge-preserving methods we propose



Group sparsity (GS)

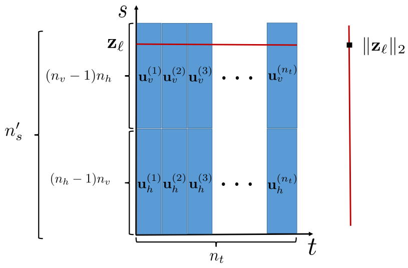
$$\mathbf{z}_\ell = \left[(\mathbf{L}_s\mathbf{u}^{(1)})_\ell, \dots, (\mathbf{L}_s\mathbf{u}^{(n_t)})_\ell\right] \in \mathbb{R}^{n_t},$$

$$\ell = 1, \dots, n_s', n_s' = (n_v-1)n_h + (n_h-1)n_v.$$

$$\mathcal{R}_5(\mathbf{u}) := \sum_{\ell=1}^{n_s'} \|\mathbf{z}_\ell\|_2$$

$$= \sum_{\ell=1}^{n_s'} \left(\sum_{t=1}^{n_t} (\mathbf{L}_s\mathbf{u}^{(t)})_\ell^2\right)^{1/2} = \|\mathbf{L}_s\mathbf{U}\|_{2,1}.$$

Other edge-preserving methods we propose



Group sparsity (GS)

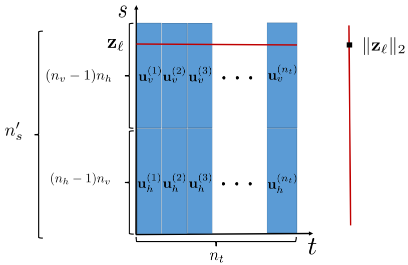
$$\mathbf{z}_\ell = \left[(\mathbf{L}_s\mathbf{u}^{(1)})_\ell, \ldots, (\mathbf{L}_s\mathbf{u}^{(n_t)})_\ell\right] \in \mathbb{R}^{n_t},$$

$$\ell = 1, \ldots, n'_s, \; n'_s = (n_v-1)n_h + (n_h-1)n_v.$$

$$\mathcal{R}_5(\mathbf{u}) := \sum_{\ell=1}^{n'_s} \|\mathbf{z}_\ell\|_2$$

$$= \sum_{\ell=1}^{n'_s} \left(\sum_{t=1}^{n_t} (\mathbf{L}_s\mathbf{u}^{(t)})_\ell^2\right)^{1/2} = \|\mathbf{L}_s\mathbf{U}\|_{2,1}.$$

Other edge-preserving methods we propose



Group sparsity (GS)

$$\mathbf{z}_\ell = \left[(\mathbf{L}_s\mathbf{u}^{(1)})_\ell, \ldots, (\mathbf{L}_s\mathbf{u}^{(n_t)})_\ell\right] \in \mathbb{R}^{n_t},$$

$$\ell = 1, \ldots, n_s', n_s' = (n_v-1)n_h + (n_h-1)n_v.$$

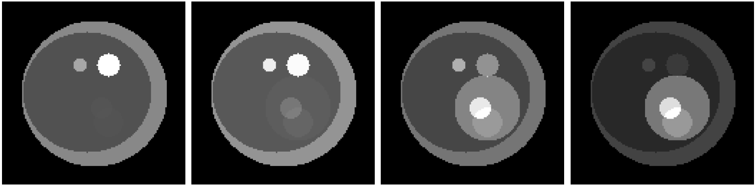
$$\mathcal{R}_5(\mathbf{u}) := \sum_{\ell=1}^{n_s'} \|\mathbf{z}_\ell\|_2$$

$$= \sum_{\ell=1}^{n_s'} \left(\sum_{t=1}^{n_t} (\mathbf{L}_s\mathbf{u}^{(t)})_\ell^2\right)^{1/2} = \|\mathbf{L}_s\mathbf{U}\|_{2,1}.$$

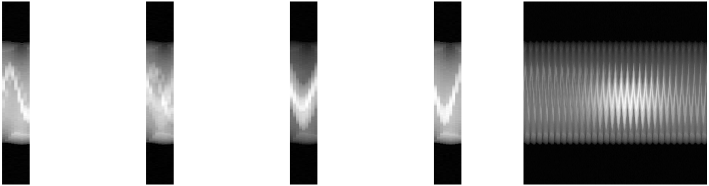
The main takeaway: All proposed models yield functionals that can be majorized by quadratic tangent majorants and minimized by GKS.

Example 1: Dynamic photoacoustic tomography (PAT)

[Sample of true images at time steps $t = 1, 10, 20, 30$ respectively from left to right.]



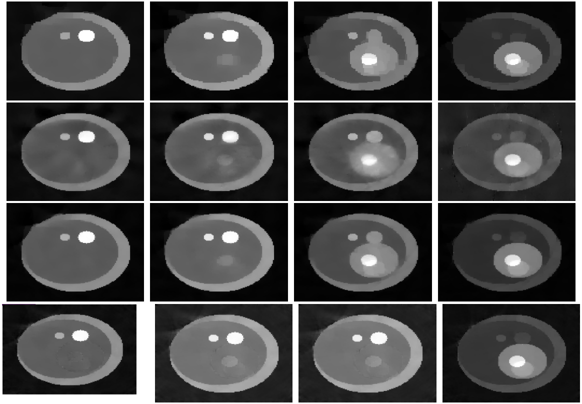
[Sample of sinograms at time steps $t = 1, 10, 20, 30$ and the full sinogram.]



PAT test problem. True images at time steps $t = 1, 10, 20, 30$.

Example 1: Dynamic photoacoustic tomography (PAT)

reconstructions



PAT test problem: First row – by solving the static problems, second row – by Iso3DTV, third row – by AnisoTV, and fourth row – by GS method.

Numerical Examples

PAT test problem

-	MM-GKS [8]	IRN-aTV (DP)[9]	IRN-aTV (L-curve)	MM-LSQR
RRE	0.096	0.081	0.071	0.299
CPU time (h)	0.31	10.1	4.16	5.54

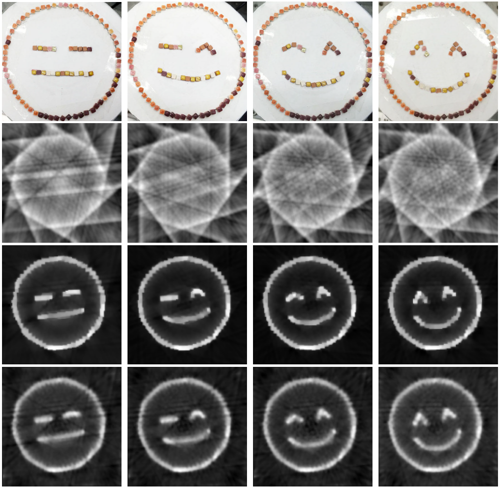
1. MM-LSQR is not competitive either in run time or in RRE. Incrementing the number of inner iterations will only increase the computational cost but will reduce the RRE.
2. IRN methods have slightly lower RRE but considerably higher run times than MM-GKS. When MM-GKS is run until 150 iterations (the maximum that we set), we get comparable RRE in about one hour.

[8] Huang, G., Lanza, A., Morigi, S., Reichel, L., Sgallari, F. (2017). Majorization–minimization generalized Krylov subspace methods for $\ell_p - \ell_q$ optimization applied to image restoration. BIT Numerical Mathematics, 57(2), 351-378.

[9] Gazzola, S., Kilmer, M. E., Nagy, J. G., Semerci, O., Miller, E. L. (2020). An inner–outer iterative method for edge preservation in image restoration and reconstruction. Inverse Problems, 36(12), 124004.

Example 2: Emoji dataset

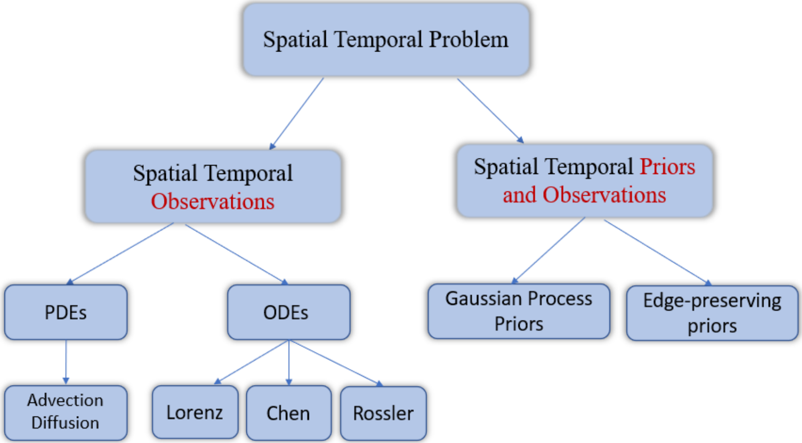
10 projection angles



10 angles. First row – the original images, second row – static problems, third row – AnisoTV, fourth row – 3DTV at time $n_t = 1, 5, 9, 15$.

Spatial temporal Bayesian Inverse Problems

Challenges



Bayesian inverse problems

Problem formulation

Let $\mathcal{G} : \mathbb{X} \mapsto \mathbb{Y}$ such that $\mathbf{b} = \mathcal{G}\mathbf{u} + \mathbf{e}$, $\quad \mathbf{e} \sim \mathcal{N}(0, \lambda^{-1}\mathbf{I}), \mathbf{u} \sim \mathcal{N}(0, \delta^{-1}\mathbf{I})$

Bayes' Law:

$$\pi_{\text{pos}}(\mathbf{u}) = \pi(\mathbf{u} \mid \mathbf{b}) = \frac{\pi_{\text{like}}(\mathbf{b} \mid \mathbf{u})\pi_{\text{pr}}(\mathbf{u})}{\mathcal{Z}} \propto \pi_{\text{like}}(\mathbf{b} \mid \mathbf{u})\pi_{\text{pr}}(\mathbf{u}), \mathcal{Z} = \int_Z \pi_{\text{like}}(\mathbf{b} \mid \mathbf{z})\pi_{\text{pr}}(\mathbf{z})d\mathbf{z}.$$

- π_{pr} prior density - encodes prior knowledge.

Bayesian inverse problems

Problem formulation

Let $\mathcal{G} : \mathbb{X} \mapsto \mathbb{Y}$ such that $\mathbf{b} = \mathcal{G}\mathbf{u} + \mathbf{e}$, $\quad \mathbf{e} \sim \mathcal{N}(0, \lambda^{-1}\mathbf{I}), \mathbf{u} \sim \mathcal{N}(0, \delta^{-1}\mathbf{I})$

Bayes' Law:

$$\pi_{\text{pos}}(\mathbf{u}) = \pi(\mathbf{u} \mid \mathbf{b}) = \frac{\pi_{\text{like}}(\mathbf{b} \mid \mathbf{u})\pi_{\text{pr}}(\mathbf{u})}{\mathcal{Z}} \propto \pi_{\text{like}}(\mathbf{b} \mid \mathbf{u})\pi_{\text{pr}}(\mathbf{u}), \mathcal{Z} = \int_{Z} \pi_{\text{like}}(\mathbf{b} \mid \mathbf{z})\pi_{\text{pr}}(\mathbf{z})d\mathbf{z}.$$

- π_{pr} prior density - encodes prior knowledge.
- $\pi(\mathbf{u}) = \pi(\mathbf{u} \mid \mathbf{b})$ is the posterior density - represents the solution.

Bayesian inverse problems

Problem formulation

Let $\mathcal{G} : \mathbb{X} \mapsto \mathbb{Y}$ such that $\mathbf{b} = \mathcal{G}\mathbf{u} + \mathbf{e}$, $\quad \mathbf{e} \sim \mathcal{N}(0, \lambda^{-1}\mathbf{I})$, $\mathbf{u} \sim \mathcal{N}(0, \delta^{-1}\mathbf{I})$

Bayes' Law:

$$\pi_{\text{pos}}(\mathbf{u}) = \pi(\mathbf{u} \mid \mathbf{b}) = \frac{\pi_{\text{like}}(\mathbf{b} \mid \mathbf{u})\pi_{\text{pr}}(\mathbf{u})}{\mathcal{Z}} \propto \pi_{\text{like}}(\mathbf{b} \mid \mathbf{u})\pi_{\text{pr}}(\mathbf{u}), \mathcal{Z} = \int_Z \pi_{\text{like}}(\mathbf{b} \mid \mathbf{z})\pi_{\text{pr}}(\mathbf{z})d\mathbf{z}.$$

- π_{pr} prior density - encodes prior knowledge.
- $\pi(\mathbf{u}) = \pi(\mathbf{u} \mid \mathbf{b})$ is the posterior density - represents the solution.
- $\pi_{\text{like}}((\mathbf{b} \mid \mathbf{u}))$ is the likelihood – encodes knowledge for the observations

Bayesian inverse problems

Problem formulation

Let $\mathcal{G} : \mathbb{X} \mapsto \mathbb{Y}$ such that $\mathbf{b} = \mathcal{G}\mathbf{u} + \mathbf{e}$, $\quad \mathbf{e} \sim \mathcal{N}(0, \lambda^{-1}\mathbf{I})$, $\mathbf{u} \sim \mathcal{N}(0, \delta^{-1}\mathbf{I})$

Bayes' Law:

$$\pi_{\text{pos}}(\mathbf{u}) = \pi(\mathbf{u} \mid \mathbf{b}) = \frac{\pi_{\text{like}}(\mathbf{b} \mid \mathbf{u})\pi_{\text{pr}}(\mathbf{u})}{\mathcal{Z}} \propto \pi_{\text{like}}(\mathbf{b} \mid \mathbf{u})\pi_{\text{pr}}(\mathbf{u}),\ \mathcal{Z} = \int_Z \pi_{\text{like}}(\mathbf{b} \mid \mathbf{z})\pi_{\text{pr}}(\mathbf{z})d\mathbf{z}.$$

- π_{pr} prior density - encodes prior knowledge.
- $\pi(\mathbf{u}) = \pi(\mathbf{u} \mid \mathbf{b})$ is the posterior density - represents the solution.
- $\pi_{\text{like}}((\mathbf{b} \mid \mathbf{u}))$ is the likelihood – encodes knowledge for the observations
- $\mathcal{Z}$ normalizing constant (model evidence)

Bayesian inverse problems

Problem formulation

Let $\mathcal{G} : \mathbb{X} \mapsto \mathbb{Y}$ such that $\mathbf{b} = \mathcal{G}\mathbf{u} + \mathbf{e}$, $\quad \mathbf{e} \sim \mathcal{N}(0, \lambda^{-1}\mathbf{I}), \mathbf{u} \sim \mathcal{N}(0, \delta^{-1}\mathbf{I})$

Bayes' Law:

$$\pi_{\text{pos}}(\mathbf{u}) = \pi(\mathbf{u} \mid \mathbf{b}) = \frac{\pi_{\text{like}}(\mathbf{b} \mid \mathbf{u})\pi_{\text{pr}}(\mathbf{u})}{\mathcal{Z}} \propto \pi_{\text{like}}(\mathbf{b} \mid \mathbf{u})\pi_{\text{pr}}(\mathbf{u}), \mathcal{Z} = \int_Z \pi_{\text{like}}(\mathbf{b} \mid \mathbf{z})\pi_{\text{pr}}(\mathbf{z})d\mathbf{z}.$$

- π_{pr} prior density - encodes prior knowledge.
- $\pi(\mathbf{u}) = \pi(\mathbf{u} \mid \mathbf{b})$ is the posterior density - represents the solution.
- $\pi_{\text{like}}((\mathbf{b} \mid \mathbf{u}))$ is the likelihood – encodes knowledge for the observations
- $\mathcal{Z}$ normalizing constant (model evidence)

Bayesian inverse problems
Problem formulation

Let $\mathcal{G} : \mathbb{X} \mapsto \mathbb{Y}$ such that $\mathbf{b} = \mathcal{G}\mathbf{u} + \mathbf{e}$, $\qquad \mathbf{e} \sim \mathcal{N}(0, \lambda^{-1}\mathbf{I}), \mathbf{u} \sim \mathcal{N}(0, \delta^{-1}\mathbf{I})$

Bayes' Law:

$$\pi_{\text{pos}}(\mathbf{u}) = \pi(\mathbf{u} \mid \mathbf{b}) = \frac{\pi_{\text{like}}(\mathbf{b} \mid \mathbf{u})\pi_{\text{pr}}(\mathbf{u})}{\mathcal{Z}} \propto \pi_{\text{like}}(\mathbf{b} \mid \mathbf{u})\pi_{\text{pr}}(\mathbf{u}), \mathcal{Z} = \int_Z \pi_{\text{like}}(\mathbf{b} \mid \mathbf{z})\pi_{\text{pr}}(\mathbf{z})d\mathbf{z}.$$

- π_{pr} prior density - encodes prior knowledge.
- $\pi(\mathbf{u}) = \pi(\mathbf{u} \mid \mathbf{b})$ is the posterior density - represents the solution.
- $\pi_{\text{like}}((\mathbf{b} \mid \mathbf{u}))$ is the likelihood – encodes knowledge for the observations
- $\mathcal{Z}$ normalizing constant (model evidence)

Maximum a posterior (MAP):

$$\pi(\mathbf{u} \mid \mathbf{b}) = \left(\lambda/2\pi\right)^{n/2} \exp\left(-\lambda/2\|\mathcal{G}\mathbf{u} - \mathbf{b}\|_2^2\right)\left(\delta/2\pi\right)^{n/2} \exp\left(-\delta/2\|\mathbf{u}\|_2^2\right)$$

Edge-preserving priors

From deterministic to Bayesian inverse problems

Edge-preserving via the prior probability distribution

- Shrinkage priors (shrinks small components to zero while maintaining true large ones)
  - Elastic net priors
  - Discrete Gaussian mixture priors
  - Horseshoe priors
  - Ridge priors

Edge-preserving priors

From deterministic to Bayesian inverse problems

Edge-preserving via the prior probability distribution

- Shrinkage priors (shrinks small components to zero while maintaining true large ones)
  - Elastic net priors
  - Discrete Gaussian mixture priors
  - Horseshoe priors
  - Ridge priors
- Heavy-tailed Markov random fields (increase the probability of large jumps by heavy tail distributions)
  - Total Variation (TV) priors
  - Laplace Markov random field priors

⚠The conditional mean estimates for the TV prior are not edge preserving through fine discretizations of the model space.

Edge-preserving priors

From deterministic to Bayesian inverse problems

Edge-preserving via the prior probability distribution

- Shrinkage priors (shrinks small components to zero while maintaining true large ones)
  - Elastic net priors
  - Discrete Gaussian mixture priors
  - Horseshoe priors
  - Ridge priors
- Heavy-tailed Markov random fields (increase the probability of large jumps by heavy tail distributions)
  - Total Variation (TV) priors
  - Laplace Markov random field priors

  ⚠The conditional mean estimates for the TV prior are not edge preserving through fine discretizations of the model space.

- Random fields with jumps of discontinuities (usage of level set functions that determine the shapes or bases)
  - Level-set priors
  - Besov priors

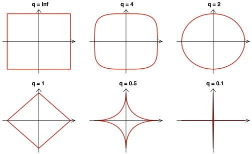
Discretization invariant

Besov priors

⚠Unwanted phenomena

- Representation of the a priori knowledge is incompatible with discretization
- The estimates diverge with adding more measurements.

✓ TV regularization is known to preserve edges by imposing the $\|\mathbf{x}\|_1$.
Equivalent to computing MAP using a TV prior and Gaussian likelihood.
⚠ Bayesian inversion with discretized TV prior is not discretization invariant.
Conditional mean looses the edge preserving property.
✓ We seek to develop priors that are discretization invariant [10]



[10] Saksman, Matti Lassas, and Samuli Siltanen.“Discretization-invariant Bayesian inversion and Besov space priors.” arXiv preprint arXiv:0901.4220 (2009), Lan, Shiwei, and Babak Shahbaba. “Sampling constrained probability distributions using spherical augmentation.” Algorithmic Advances in Riemannian Geometry and Applications. Springer, Cham, 2016. 25-71.

Bayesian inverse problems

Besov priors

- Consider a basis $\{\phi_\ell\}_{\ell=1}^{\infty}$ for $L^2(\mathbb{T}^d)$, $\mathbb{T}^d = (0,1]^d$ for $d \leq 3$ s.t. any $f \in L^2(\mathbb{T}^d)$

$$f(\mathbf{x}) = \sum_{\ell=1}^{\infty} f_\ell \phi_\ell(\mathbf{x}).$$

Bayesian inverse problems

Besov priors

- Consider a basis $\{\phi_\ell\}_{\ell=1}^{\infty}$ for $L^2(\mathbb{T}^d)$, $\mathbb{T}^d = (0,1]^d$ for $d \leq 3$ s.t. any $f \in L^2(\mathbb{T}^d)$

$$f(\mathbf{x}) = \sum_{\ell=1}^{\infty} f_\ell \phi_\ell(\mathbf{x}).$$

- Denote $\mathbb{X}^{s,q}$ as a Banach space with norm $\|\cdot\|_{s,q}$ defined as

$$\|f\|_{s,q} = \left(\sum_{\ell=1}^{\infty} \ell^{\left(\frac{sq}{d} + \frac{q}{2} - 1\right)} |f_\ell|^q \right)^{\frac{1}{q}}, \text{ with } s > 0 \text{ and } q \geq 1.$$

Bayesian inverse problems

Besov priors

- Consider a basis $\{\phi_\ell\}_{\ell=1}^{\infty}$ for $L^2(\mathbb{T}^d)$, $\mathbb{T}^d = (0,1]^d$ for $d \leq 3$ s.t. any $f \in L^2(\mathbb{T}^d)$

$$f(\mathbf{x}) = \sum_{\ell=1}^{\infty} f_\ell \phi_\ell(\mathbf{x}).$$

- Denote $\mathbb{X}^{s,q}$ as a Banach space with norm $\|\cdot\|_{s,q}$ defined as

$$\|f\|_{s,q} = \left(\sum_{\ell=1}^{\infty} \ell^{\left(\frac{sq}{d} + \frac{q}{2} - 1\right)} |f_\ell|^q \right)^{\frac{1}{q}}, \text{ with } s > 0 \text{ and } q \geq 1.$$

- Let $s > 0$, $1 \leq q < \infty$ and $\kappa > 0$ be fixed. We consider a sequence of i.i.d random variables $\{\xi_\ell\}_{\ell=1}^{\infty}$ whose probability density function is a q-exponential distribution:

$$\pi_\xi(\cdot) \propto \exp\left(-\frac{1}{2}|\xi|^q\right).$$

Bayesian inverse problems

Besov priors

- Consider a basis $\{\phi_\ell\}_{\ell=1}^{\infty}$ for $L^2(\mathbb{T}^d)$, $\mathbb{T}^d = (0,1]^d$ for $d \leq 3$ s.t. any $f \in L^2(\mathbb{T}^d)$

$$f(\mathbf{x}) = \sum_{\ell=1}^{\infty} f_\ell \phi_\ell(\mathbf{x}).$$

- Denote $\mathbb{X}^{s,q}$ as a Banach space with norm $\|\cdot\|_{s,q}$ defined as

$$\|f\|_{s,q} = \left(\sum_{\ell=1}^{\infty} \ell^{(\frac{sq}{d}+\frac{q}{2}-1)} |f_\ell|^q\right)^{\frac{1}{q}}, \text{ with } s > 0 \text{ and } q \geq 1.$$

- Let $s > 0$, $1 \leq q < \infty$ and $\kappa > 0$ be fixed. We consider a sequence of i.i.d random variables $\{\xi_\ell\}_{\ell=1}^{\infty}$ whose probability density function is a q-exponential distribution:

$$\pi_\xi(\cdot) \propto \exp\left(-\frac{1}{2}|\xi|^q\right).$$

- For an orthonormal basis $\{\phi_\ell\}_{\ell=1}^{\infty}$, we define a random function u as follows

$$u(\mathbf{x}) = \sum_{\ell=1}^{\infty} u_\ell \phi_\ell(\mathbf{x}) = \sum_{\ell=1}^{\infty} \gamma_\ell \xi_\ell \phi_\ell(\mathbf{x}), \quad \xi_\ell \overset{iid}{\sim} \pi_\xi, u_\ell := \gamma_\ell \xi_\ell, \gamma_\ell = \kappa^{-\frac{1}{q}} \ell^{-(\frac{s}{d}+\frac{1}{2}-\frac{1}{q})}.$$

We refer to the induced measure on functions u as Besov measure, denoted as μ_0.

Besov Priors

Generalization to Spatial-Temporal domain

- Let $f \in L^p(\mathcal{T})$ over temporal domain $\mathcal{T}$. We have the following series expansion for a function defined on spatial and temporal domains

$$f(\mathbf{x},t) = \sum_{\ell=1}^{\infty} f_\ell(t)\phi_\ell(\mathbf{x}), \text{ where for each } \ell \in, f_\ell \in L^p(\mathcal{T}).$$

[11] Lan, Li, Pasha, "Edge-Preserving Priors for Spatio-Temporal Inverse Problems" (in preparation)

Besov Priors

Generalization to Spatial-Temporal domain

- Let $f \in L^p(\mathcal{T})$ over temporal domain $\mathcal{T}$. We have the following series expansion for a function defined on spatial and temporal domains

$$f(\mathbf{x}, t) = \sum_{\ell=1}^{\infty} f_\ell(t)\phi_\ell(\mathbf{x}), \text{ where for each } \ell \in, f_\ell \in L^p(\mathcal{T}).$$

- For the infinite sequence $f := \{f_\ell\}_{\ell=1}^{\infty}$, we define the following (r, q, p) norm with spatial (Besov) index q and temporal index p:

$$\|f\|_{r,q,p} = \left(\sum_{\ell=1}^{\infty} \ell^{rq} \|f_\ell\|_p^q \right)^{\frac{1}{q}}, r = r_0 := \frac{s}{d} + \frac{1}{2} - \frac{1}{q}.$$

[11] Lan, Li, Pasha, "Edge-Preserving Priors for Spatio-Temporal Inverse Problems" (in preparation)

Besov Priors

Generalization to Spatial-Temporal domain

- Let $f \in L^p(\mathcal{T})$ over temporal domain $\mathcal{T}$. We have the following series expansion for a function defined on spatial and temporal domains

$$f(\mathbf{x}, t) = \sum_{\ell=1}^{\infty} f_\ell(t)\phi_\ell(\mathbf{x}), \text{ where for each } \ell \in, f_\ell \in L^p(\mathcal{T}).$$

- For the infinite sequence $f := \{f_\ell\}_{\ell=1}^{\infty}$, we define the following (r, q, p) norm with spatial (Besov) index q and temporal index p:

$$\|f\|_{r,q,p} = \left(\sum_{\ell=1}^{\infty} \ell^{rq} \|f_\ell\|_p^q \right)^{\frac{1}{q}}, r = r_0 := \frac{s}{d} + \frac{1}{2} - \frac{1}{q}.$$

- Denote such space $\ell^{r,q}(L^p(\mathcal{T})) := \{f | \|f\|_{r,q,p} < \infty\}$. We define the (r, q, p) norm for $f(\mathbf{x}, t)$ and denote the Banach space $\mathbb{X}^{r,q,p} = \{f(\mathbf{x}, t) | \|f\|_{r,q,p} < \infty\}$.

[11] Lan, Li, Pasha, "Edge-Preserving Priors for Spatio-Temporal Inverse Problems" (in preparation)

Besov Priors

Generalization to Spatial-Temporal domain

- Let $f \in L^p(\mathcal{T})$ over temporal domain $\mathcal{T}$. We have the following series expansion for a function defined on spatial and temporal domains

$$f(\mathbf{x},t) = \sum_{\ell=1}^{\infty} f_\ell(t)\phi_\ell(\mathbf{x}), \text{ where for each } \ell \in, f_\ell \in L^p(\mathcal{T}).$$

- For the infinite sequence $f := \{f_\ell\}_{\ell=1}^{\infty}$, we define the following (r,q,p) norm with spatial (Besov) index q and temporal index p:

$$\|f\|_{r,q,p} = \left(\sum_{\ell=1}^{\infty} \ell^{rq}\|f_\ell\|_p^q\right)^{\frac{1}{q}}, r = r_0 := \frac{s}{d} + \frac{1}{2} - \frac{1}{q}.$$

- Denote such space $\ell^{r,q}(L^p(\mathcal{T})) := \{f \,|\, \|f\|_{r,q,p} < \infty\}$. We define the (r,q,p) norm for $f(\mathbf{x},t)$ and denote the Banach space $\mathbb{X}^{r,q,p} = \{f(\mathbf{x},t) \,|\, \|f\|_{r,q,p} < \infty\}$.
- We generalize the Besov process $\mathcal{B}(\kappa, \mathbb{X}^{s,q})$ to be spatiotemporal by varying random coefficients $\{\xi_\ell\}$ in time according to a process[11]:

$$u(\mathbf{x},t) = \sum_{\ell=1}^{\infty} u_\ell(t)\phi_\ell(\mathbf{x}) = \sum_{\ell=1}^{\infty} \gamma_\ell \xi_\ell(t)\phi_\ell(\mathbf{x}), \quad \xi_\ell(\cdot) \overset{iid}{\sim} qEP(0,\mathcal{C})$$

[11] Lan, Li, Pasha, "Edge-Preserving Priors for Spatio-Temporal Inverse Problems" (in preparation)

Besov Priors

Generalization to Spatial-Temporal domain

- Let $f \in L^p(\mathcal{T})$ over temporal domain $\mathcal{T}$. We have the following series expansion for a function defined on spatial and temporal domains

$$f(\mathbf{x}, t) = \sum_{\ell=1}^{\infty} f_\ell(t)\phi_\ell(\mathbf{x}), \text{ where for each } \ell \in, f_\ell \in L^p(\mathcal{T}).$$

- For the infinite sequence $f := \{f_\ell\}_{\ell=1}^{\infty}$, we define the following (r, q, p) norm with spatial (Besov) index q and temporal index p:

$$\|f\|_{r,q,p} = \left(\sum_{\ell=1}^{\infty} \ell^{rq} \|f_\ell\|_p^q\right)^{\frac{1}{q}}, r = r_0 := \frac{s}{d} + \frac{1}{2} - \frac{1}{q}.$$

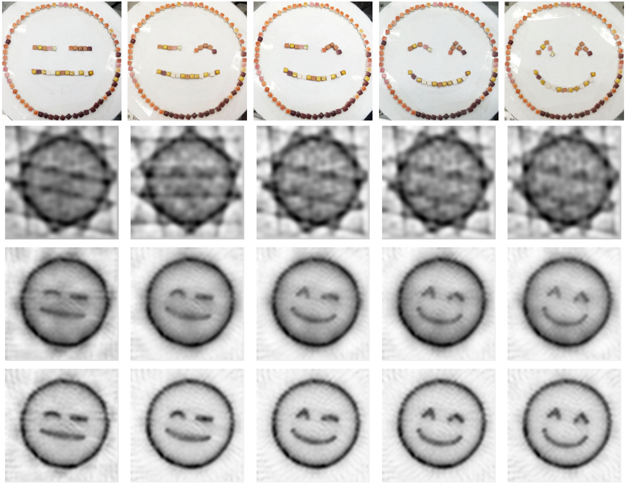
- Denote such space $\ell^{r,q}(L^p(\mathcal{T})) := \{f \| \|f\|_{r,q,p} < \infty\}$. We define the (r, q, p) norm for $f(\mathbf{x}, t)$ and denote the Banach space $\mathbb{X}^{r,q,p} = \{f(\mathbf{x}, t) \| \|f\|_{r,q,p} < \infty\}$.
- We generalize the Besov process $\mathcal{B}(\kappa, \mathbb{X}^{s,q})$ to be spatiotemporal by varying random coefficients $\{\xi_\ell\}$ in time according to a process[11]:

$$u(\mathbf{x}, t) = \sum_{\ell=1}^{\infty} u_\ell(t)\phi_\ell(\mathbf{x}) = \sum_{\ell=1}^{\infty} \gamma_\ell \xi_\ell(t)\phi_\ell(\mathbf{x}), \quad \xi_\ell(\cdot) \overset{iid}{\sim} qEP(0, \mathcal{C})$$

- This stochastic process is the *spatiotemporal Besov process* $(\kappa, \mathcal{C}, \mathbb{X}^{r,q,p})$.

[11] Lan, Li, Pasha, "Edge-Preserving Priors for Spatio-Temporal Inverse Problems" (in preparation)

Revisiting Emoji example

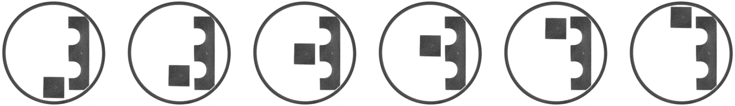


Reconstruction results for the emoji test problem with $n_a = 10$.

STEMPO test problem

Simulated data

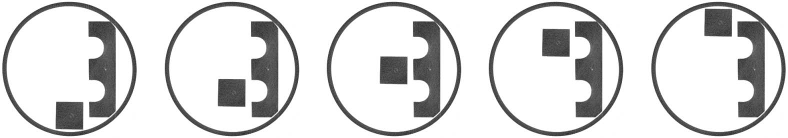
- Consider images from Spatio-TEmporal Motor-POwered (STEMPO) phantom [12].
- Select $n_t = 20$ images of size 560×560.
- Generate the forward operators $\mathbf{A}^{(t)}$, $t = 1, 2, \ldots, n_t$ by considering n_t vectors of length 11 containing projection angles.
- Each forward operator $\mathbf{A} \in \mathbb{R}^{8701 \times 313600}$ and the blockdiagonal matrix $\mathbf{A} \in \mathbb{R}^{174020 \times 6272000}$.
- Obtain n_t sinograms $\mathbf{d}^{(t)} \in \mathbb{R}^{8701}$, with $\mathbf{D}^{(t)} \in \mathbb{R}^{791 \times 11}$, for $t = 1, 2, \ldots, n_t$.



[12] Tommi Heikkilä. Stempo–dynamic x-ray tomography phantom. arXiv preprint arXiv:2209.12471, 2022

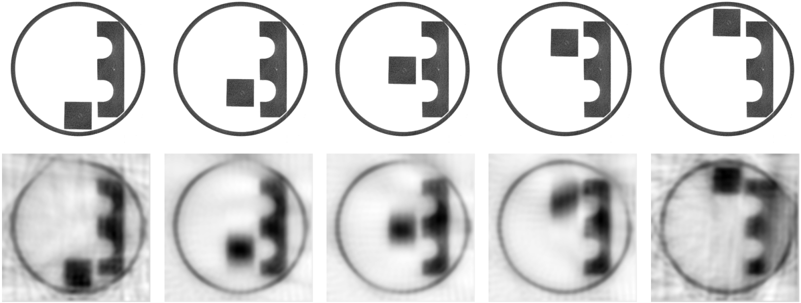
STEMPO test problem

Simulated data



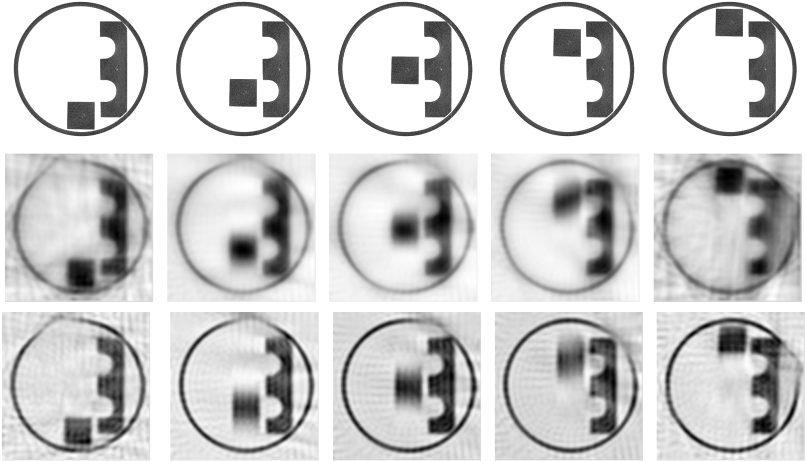
STEMPO test problem

Simulated data



STEMPO test problem

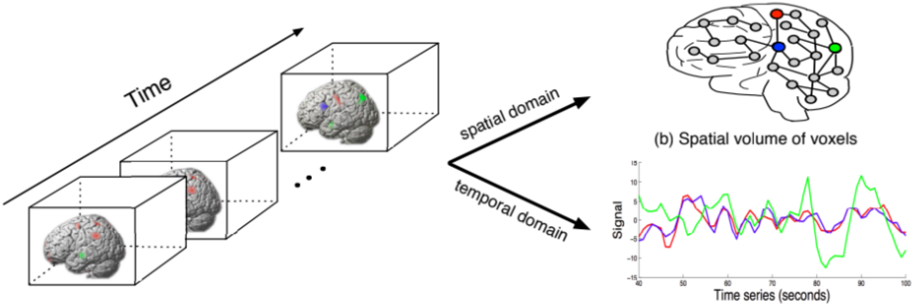
Simulated data



Dynamic STEMPO test problem: First row, from left to right: True images at time steps $t = 1, 10, 20, 30$. Second row, from left to right: Reconstructions with Besov Priors $s = 2, q = 1$ for spatial domain and $q = 1$ for time domain at time steps $t = 1, 10, 15, 20$. Third row, from left to right: Reconstructions with

High-Dimensional Data and Inverse Problems

Spatio-Temporal Brain fMRI



(b) Spatial volume of voxels

(a) Spatial-temporal structure of fMRI data

(c) Time series of voxels

1. Neurological disorders are characterized in the early stages by hidden ongoing brain injury.
2. Most of traditional kernel methods convert a tensor to a vector (or a matrix)
3. Conversion to vectors would cause the loss of structural information such as the spatial arrangement of voxel-based features.

Ma, Guixiang, et al."Spatio-temporal tensor analysis for whole-brain fMRI classification. "Proceedings of the 2016 SIAM International Conference on Data Mining. SIAM, 2016.

New exciting work

Tensor representation of high-dimensional data

- Data from many applications are natively high dimensional.

New exciting work

Tensor representation of high-dimensional data

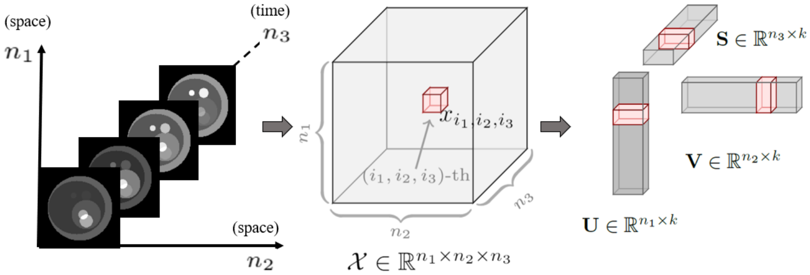
- Data from many applications are natively high dimensional.
- Standard linear algebra tools can not be used.

New exciting work

Tensor representation of high-dimensional data

- Data from many applications are natively high dimensional.
- Standard linear algebra tools can not be used.
- Emergent high need for developing and using tensor framework for a variety of applications, including image reconstruction and compression.

$$\mathcal{X}^* = \arg\min_{\mathcal{X}} \mathcal{J}(\mathcal{X};\mathcal{B}) = \frac{1}{p}\|\mathcal{B} - \mathcal{A}(\mathcal{X})\|_p^p + \mathcal{R}(\mathcal{X})$$



Concluding remarks and outlook

✓ Accomplished so far

- We propose 6 main methods for solving time-dependent inverse problems based on a generalized Krylov subspace.
- Explored UQ methods for spatial and temporal priors and observations.
- Developed non-Gaussian priors for dynamic IP

⇒ Potential future directions

- Develop efficient methods for sampling in large-scale dynamic IP.
- Develop decompositions for higher dimension representations.

Thank you for your attention!

S. Lan, S. Li, and M. Pasha.
Spatiotemporal Besov Priors for Bayesian Inverse Problems (in preparation).
S. Lan, S. Li, and M. Pasha.
Bayesian Spatiotemporal Modeling for Inverse Problems.
https://arxiv.org/abs/2204.10929
M. Pasha, A. K. Saibaba, S. Gazzola, M. I. Espanol, and E. de Sturler.
Efficient edge-preserving methods for dynamic inverse problems.
https://arxiv.org/abs/2107.05727