

# Projected Nesterov's Proximal-Gradient Algorithm for Constrained Sparse Signal Reconstruction<sup>†</sup>

Aleksandar Dogandžić

joint work with Renliang Gu, Ph.D. student

Electrical and Computer Engineering  
IOWA STATE UNIVERSITY

---

<sup>†</sup>Supported by 

# Terminology and Notation I

- ▶ soft-thresholding operator for  $\mathbf{a} = (a_i)_{i=1}^N \in \mathbb{R}^N$ :

$$[\mathcal{T}_\lambda(\mathbf{a})]_i = \text{sign}(a_i) \max(|a_i| - \lambda, 0);$$

- ▶ “ $\succeq$ ” is the elementwise version of “ $\geq$ ”;
- ▶ proximal operator for function  $r(\mathbf{x})$  scaled by  $\lambda$ :

$$\text{prox}_{\lambda r} \mathbf{a} = \arg \min_{\mathbf{x}} \frac{1}{2} \|\mathbf{x} - \mathbf{a}\|_2^2 + \lambda r(\mathbf{x}).$$

- ▶  $\varepsilon$ -subgradient (Rockafellar 1970, Sec. 23):

$$\partial_\varepsilon r(\mathbf{x}) \triangleq \{\mathbf{g} \in \mathbb{R}^p \mid r(\mathbf{z}) \geq r(\mathbf{x}) + (\mathbf{z} - \mathbf{x})^T \mathbf{g} - \varepsilon, \forall \mathbf{z} \in \mathbb{R}^p\}.$$

## Terminology and Notation II

- ▶  $\iota^L(s)$  is the *Laplace transform* of  $\iota(\kappa)$ :

$$\iota^L(s) \triangleq \int \iota(\kappa) e^{-s\kappa} d\kappa,$$

- ▶ Laplace transform with vector argument:

$$\mathbf{b}_o^L(s) = \mathbf{b}_o^L \left( \begin{bmatrix} s_1 \\ s_2 \\ \vdots \\ s_N \end{bmatrix} \right) = \begin{bmatrix} \mathbf{b}^L(s_1) \\ \mathbf{b}^L(s_2) \\ \vdots \\ \mathbf{b}^L(s_N) \end{bmatrix}.$$

# Outline

## Projected Nesterov's Proximal-Gradient Algorithm

Background

Optimization problem

Reconstruction algorithm

Convergence analysis

## Applications

Poisson compressed sensing

Linear model with Gaussian noise

X-Ray CT Reconstruction from Polychromatic Measurements

Polychromatic X-ray CT model

Mass-attenuation spectrum

Minimization algorithm

Numerical Examples

## Conclusion

# Introduction I

Natural signals are well described by a few significant coefficients in an appropriate transform domain, where

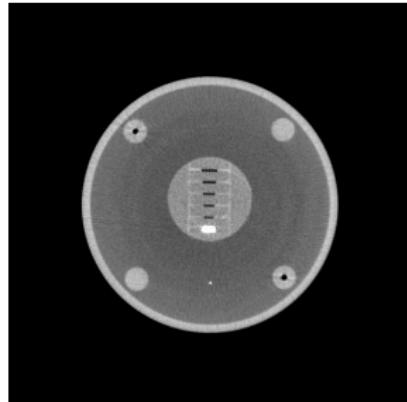
$$\#\text{ significant coefficients} \ll \text{signal size } p.$$

*Sparsifying transform:* Choose

$$\psi(\cdot) : \mathbb{R}^p \mapsto \mathbb{R}^{p'},$$

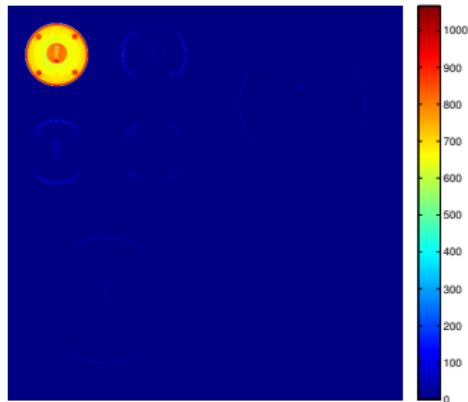
so that most elements of  $\psi(x)$  have negligible magnitudes.

# Sparsifying Transforms I



$p$  pixels

linear  
transform  
 $\Leftrightarrow$   
DWT

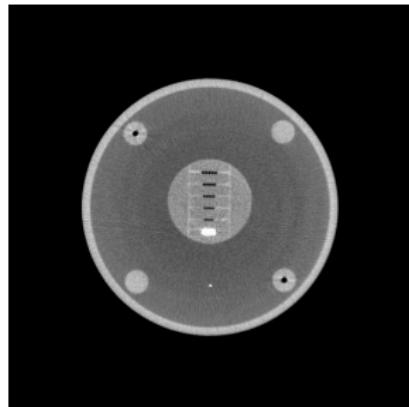


# significant coeffs  $\ll p$

$$\psi(x) = \Psi^T x$$

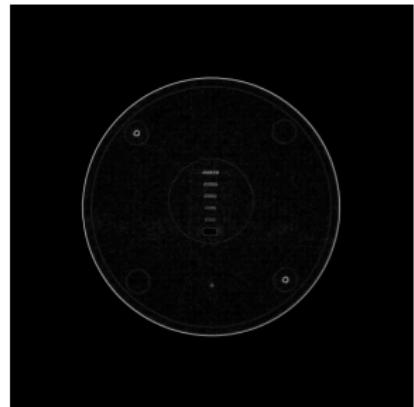
where  $\Psi \in \mathbb{R}^{p \times p'}$  is a known *sparsifying dictionary matrix*.

## Sparsifying Transforms II



$p$  pixels

transform  
 $\leftrightarrow$   
gradient map



# significant coeffs  $\ll p$

$$[\psi(x)]_i \triangleq \sqrt{\sum_{j \in \mathcal{N}_i} (x_i - x_j)^2}, \text{ for } i = 1, 2, \dots, p'$$

---

$\mathcal{N}_i$  is the index set of neighbors of  $x_i$  in an appropriate (e.g., 2D) arrangement.

## Goal

Sense the significant components of  $\psi(x)$  using a small number of measurements.

Define the noiseless measurement vector  $\phi(x)$ , where

$$\phi(\cdot) : \mathbb{R}^p \mapsto \mathbb{R}^N$$

and  $N \leq p$ .

Example: Linear model

$$\phi(x) = \Phi x$$

where  $\Phi \in \mathbb{R}^{N \times p}$  is a known *sensing* matrix.

## Goal

Sense the significant components of  $\psi(x)$  using a small number of measurements.

Define the noiseless measurement vector  $\phi(x)$ , where

$$\phi(\cdot) : \mathbb{R}^p \mapsto \mathbb{R}^N$$

and  $N \leq p$ .

Example: Linear model

$$\phi(x) = \Phi x$$

where  $\Phi \in \mathbb{R}^{N \times p}$  is a known *sensing* matrix.

## Convex Signal Constraint

$$\mathbf{x} \in C$$

where  $C$  is a closed convex set.

Example: the nonnegative signal set

$$C = \mathbb{R}_+^p$$

is of significant practical interest and applicable to X-ray computed tomography (CT), single photon emission computed tomography (SPECT), positron emission tomography (PET), and magnetic resonance imaging (MRI).

## Convex Signal Constraint

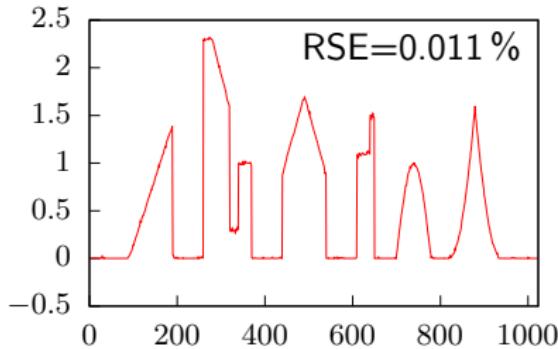
$$x \in C$$

where  $C$  is a closed convex set.

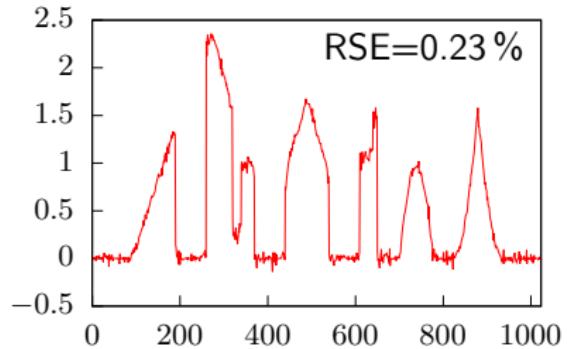
Example: the nonnegative signal set

$$C = \mathbb{R}_+^p$$

is of significant practical interest and applicable to X-ray CT, SPECT, PET, and MRI.



(a) PNPG



(b) NPGs

Figure 1: Benefit of nonnegative convex-set signal constraints.

# Outline

## Projected Nesterov's Proximal-Gradient Algorithm

Background

Optimization problem

Reconstruction algorithm

Convergence analysis

## Applications

Poisson compressed sensing

Linear model with Gaussian noise

X-Ray CT Reconstruction from Polychromatic Measurements

Polychromatic X-ray CT model

Mass-attenuation spectrum

Minimization algorithm

Numerical Examples

## Conclusion

## Penalized NLL

- ▶ objective function

$$f(\mathbf{x}) = \mathcal{L}(\mathbf{x}) + u \underbrace{[\|\psi(\mathbf{x})\|_1 + \mathbb{I}_C(\mathbf{x})]}_{r(\mathbf{x})}$$

- ▶ negative log-likelihood (NLL)
- ▶ penalty term
- $u > 0$  is a scalar tuning constant

## Penalized NLL

- ▶ objective function

$$f(x) = \mathcal{L}(x) +$$

$$u \underbrace{[\|\psi(x)\|_1 + \mathbb{I}_C(x)]}_{r(x)}$$

- ▶ NLL

- ▶ penalty term

$u > 0$  is a scalar tuning constant

## Penalized NLL

- ▶ objective function

$$f(x) = \mathcal{L}(x) + u \underbrace{\left[ \|\psi(x)\|_1 + \mathbb{I}_C(x) \right]}_{r(x)}$$

- ▶ NLL

- ▶ penalty term  
 $u > 0$  is a scalar tuning constant

## NLL Assumptions

- ▶ The negative log-likelihood (NLL)  $\mathcal{L}(\mathbf{x})$  is a convex differentiable function of the signal  $\mathbf{x}$ .
- ▶

$$C \subseteq \text{cl}(\text{dom } \mathcal{L}(\mathbf{x}))$$

ensures that  $\mathcal{L}(\mathbf{x})$  is computable for all  $\mathbf{x} \in \text{int } C$ .

- ▶ For numerical stability, normalize the likelihood function so that the corresponding NLL  $\mathcal{L}(\mathbf{x})$  is lower-bounded by zero.

## Comment

Our objective function  $f(x)$  is

- ▶ convex
  - ▶ has unique optimum,
- ▶ not differentiable with respect to the signal  $x$ 
  - ▶ cannot apply usual gradient- or Newton-type algorithms,
  - ▶ need proximal-gradient (PG) schemes.

# Outline

## Projected Nesterov's Proximal-Gradient Algorithm

Background

Optimization problem

### Reconstruction algorithm

Convergence analysis

## Applications

Poisson compressed sensing

Linear model with Gaussian noise

X-Ray CT Reconstruction from Polychromatic Measurements

Polychromatic X-ray CT model

Mass-attenuation spectrum

Minimization algorithm

Numerical Examples

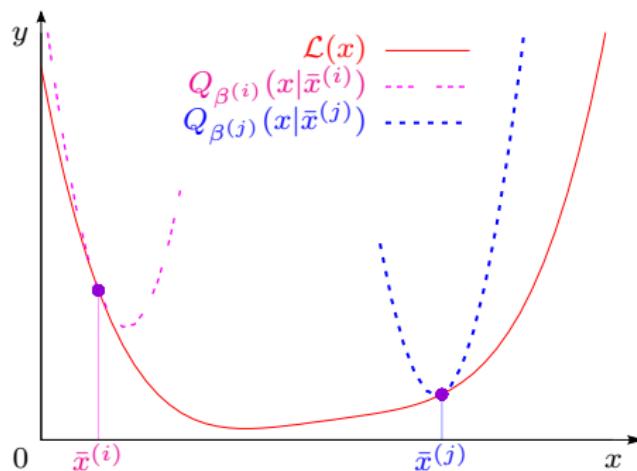
## Conclusion

## Majorization Function

Define the quadratic approximation of the NLL  $\mathcal{L}(x)$ :

$$Q_\beta(x | \bar{x}) = \mathcal{L}(\bar{x}) + (x - \bar{x})^T \nabla \mathcal{L}(\bar{x}) + \frac{1}{2\beta} \|x - \bar{x}\|_2^2$$

with  $\beta$  chosen so that  $Q_\beta(x | \bar{x})$  majorizes  $\mathcal{L}(x)$  in the neighborhood of  $x = \bar{x}$ .



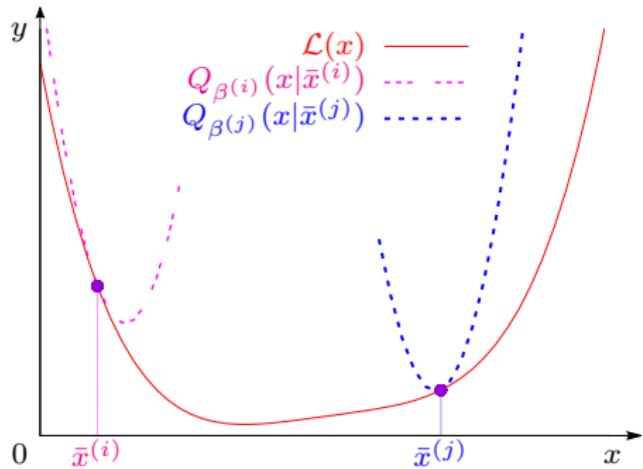


Figure 2: Majorizing function: Impact of  $\beta$ .

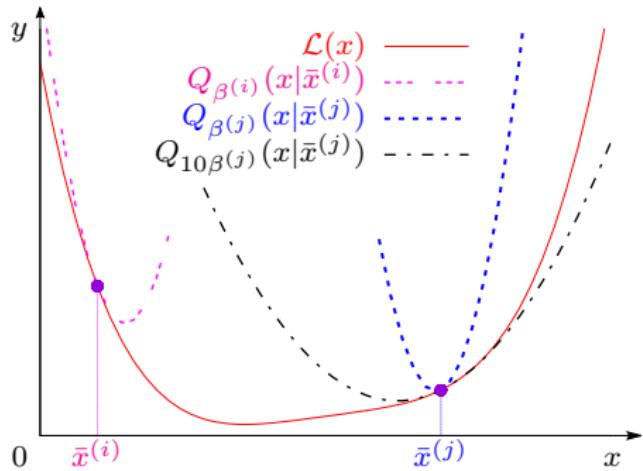


Figure 2: Majorizing function: Impact of  $\beta$ .

## PNPG Method: Iteration $i$

$$B^{(i)} = \beta^{(i-1)} / \beta^{(i)} \quad (2a)$$

$$\theta^{(i)} = \frac{1}{2} \left[ 1 + \sqrt{1 + 4 B^{(i)} (\theta^{(i-1)})^2} \right] \quad (2b)$$

$$\bar{\mathbf{x}}^{(i)} = P_C \left( \mathbf{x}^{(i-1)} + \frac{\theta^{(i-1)} - 1}{\theta^{(i)}} (\mathbf{x}^{(i-1)} - \mathbf{x}^{(i-2)}) \right) \quad \text{accel. step} \quad (2c)$$

$$\mathbf{x}^{(i)} = \text{prox}_{\beta^{(i)} \mathbf{u} \mathbf{r}} \left( \bar{\mathbf{x}}^{(i)} - \beta^{(i)} \nabla \mathcal{L}(\bar{\mathbf{x}}^{(i)}) \right) \quad \text{PG step} \quad (2d)$$

where  $\beta^{(i)} > 0$  is an *adaptive step size* chosen to satisfy the *majorization condition*

$$\mathcal{L}(\mathbf{x}^{(i)}) \leq Q_{\beta^{(i)}}(\mathbf{x}^{(i)} \mid \bar{\mathbf{x}}^{(i)}) \quad (3)$$

using a simple adaptation scheme that aims at keeping  $\beta^{(i)}$  as large as possible.

## PNPG Method: Iteration $i$

allows conv. rate guarantees for adaptive step size

$$B^{(i)} = \beta^{(i-1)} / \beta^{(i)} \quad (2a)$$

$$\theta^{(i)} = \frac{1}{2} \left[ 1 + \sqrt{1 + 4 B^{(i)} (\theta^{(i-1)})^2} \right] \quad (2b)$$

$$\bar{x}^{(i)} = P_C \left( x^{(i-1)} + \frac{\theta^{(i-1)} - 1}{\theta^{(i)}} (x^{(i-1)} - x^{(i-2)}) \right) \quad \text{accel. step} \quad (2c)$$

$$x^{(i)} = \text{prox}_{\beta^{(i)} u r} \left( \bar{x}^{(i)} - \beta^{(i)} \nabla \mathcal{L}(\bar{x}^{(i)}) \right) \quad \text{PG step} \quad (2d)$$

where  $\beta^{(i)} > 0$  is an *adaptive step size* chosen to satisfy the *majorization condition*

$$\mathcal{L}(x^{(i)}) \leq Q_{\beta^{(i)}}(x^{(i)} | \bar{x}^{(i)}) \quad (3)$$

using a simple adaptation scheme that aims at keeping  $\beta^{(i)}$  as large as possible.

## PNPG Method: Iteration $i$

allows general  $\text{dom } \mathcal{L}$

$$B^{(i)} = \beta^{(i-1)} / \beta^{(i)} \quad (2a)$$

$$\theta^{(i)} = \frac{1}{2} \left[ 1 + \sqrt{1 + 4B^{(i)} (\theta^{(i-1)})^2} \right] \quad (2b)$$

$$\bar{\mathbf{x}}^{(i)} = P_C \left( \mathbf{x}^{(i-1)} + \frac{\theta^{(i-1)} - 1}{\theta^{(i)}} (\mathbf{x}^{(i-1)} - \mathbf{x}^{(i-2)}) \right) \text{ accel. step} \quad (2c)$$

$$\mathbf{x}^{(i)} = \text{prox}_{\beta^{(i)} \mathbf{u} r} \left( \bar{\mathbf{x}}^{(i)} - \beta^{(i)} \nabla \mathcal{L}(\bar{\mathbf{x}}^{(i)}) \right) \text{ PG step} \quad (2d)$$

where  $\beta^{(i)} > 0$  is an *adaptive step size* chosen to satisfy the *majorization condition*

$$\mathcal{L}(\mathbf{x}^{(i)}) \leq Q_{\beta^{(i)}}(\mathbf{x}^{(i)} \mid \bar{\mathbf{x}}^{(i)}) \quad (3)$$

using a simple adaptation scheme that aims at keeping  $\beta^{(i)}$  as large as possible.

▶ more

---

**Algorithm 1:** PNPG method

---

**Input:**  $x^{(-1)}$ ,  $u$ ,  $\mathfrak{m}$ ,  $\mathfrak{M}$ ,  $\xi$ ,  $\eta$ , and threshold  $\epsilon$

**Output:**  $\arg \min_x f(x)$

Initialization:  $\theta^{(0)} \leftarrow 0$ ,  $x^{(0)} \leftarrow \mathbf{0}$ ,  $i \leftarrow 0$ ,  $\kappa \leftarrow 0$ ,  $\beta^{(0)} \leftarrow 0$  and  $\beta^{(1)}$  by the BB method

**repeat**

$i \leftarrow i + 1$  and  $\kappa \leftarrow \kappa + 1$

**while** true **do**

        evaluate (2a) to (2c)

**if**  $\bar{x}^{(i)} \notin \text{dom } \mathcal{L}$  **then**

$\theta^{(i-1)} \leftarrow 1$  and continue

// backtracking search

        solve the proximal-mapping step (2d)

**if** majorization condition (3) holds **then**

            break

**else**

**if**  $\beta^{(i)} > \beta^{(i-1)}$  **then**

$\mathfrak{m} \leftarrow \mathfrak{m} + \mathfrak{m}$

$\beta^{(i)} \leftarrow \xi \beta^{(i)}$  and  $\kappa \leftarrow 0$

// increase  $\mathfrak{m}$

**if**  $i > 1$  and  $f(x^{(i)}) > f(x^{(i-1)})$  **then**

// function restart

$\theta^{(i-1)} \leftarrow 1$ ,  $i \leftarrow i - 1$ , and continue

**if** convergence condition holds **then**

            declare convergence

**if**  $\kappa \geq \mathfrak{m}$  **then**

// adapt step size

$\kappa \leftarrow 0$  and  $\beta^{(i+1)} \leftarrow \beta^{(i)} / \xi$

**else**

$\beta^{(i+1)} \leftarrow \beta^{(i)}$

**until** convergence declared or maximum number of iterations exceeded

---

## Restart

The *function* and *domain* restarts guarantee that

- ▶ the projected Nesterov's proximal-gradient (PNPG) iteration is *monotonic* and
- ▶  $\bar{x}^{(i)}$  and  $x^{(i)}$  will remain *within*  $\text{dom } f$  as long as the projected initial value is within  $\text{dom } f$ :  $f(P_C(x^{(0)})) < +\infty$ .

### Remark (Monotonicity)

*The PNPG iteration with restart is non-increasing:*

$$f(x^{(i)}) \leq f(x^{(i-1)})$$

*for all  $i$ .*

▶ more

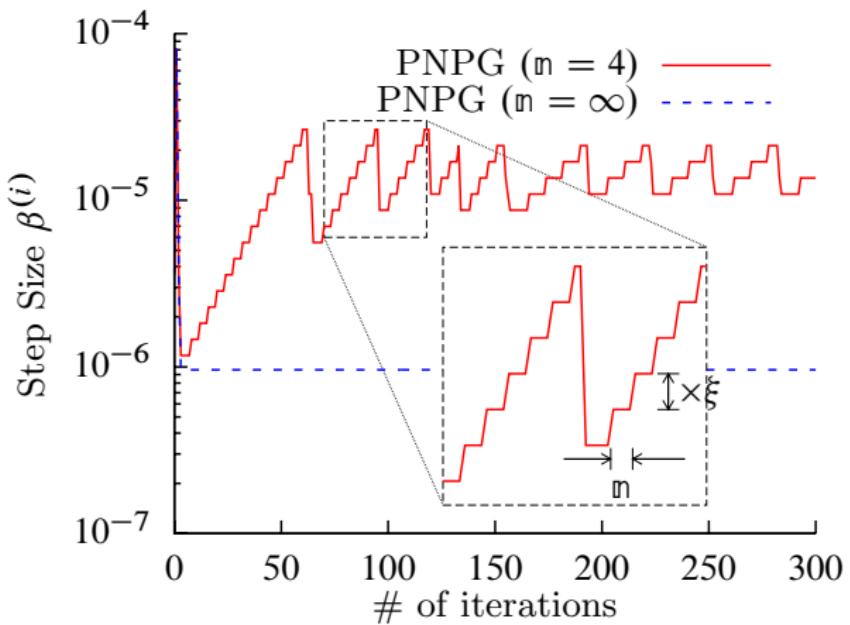


Figure 3: Illustration of step-size selection.

# Proximal Mapping

To compute

$$\text{prox}_{\lambda r} \mathbf{a} = \arg \min_{\mathbf{x}} \frac{1}{2} \|\mathbf{x} - \mathbf{a}\|_2^2 + \lambda r(\mathbf{x})$$

use

- ▶ for  $\ell_1$ -norm penalty  $\psi(\mathbf{x}) = \Psi^T \mathbf{x}$ , alternating direction method of multipliers (ADMM)
  - ▶ easy in the practically important scenario where
$$\Psi \Psi^T = I,$$
  - ▶ otherwise needs more attention.
- ▶ for total-variation (TV)-norm penalty with gradient map  $\psi(\mathbf{x})$ , an inner iteration with the TV-based denoising method in (Beck and Teboulle 2009b).

## Convergence Criterion

$$\delta^{(i)} \triangleq \|x^{(i)} - x^{(i-1)}\|_2 < \epsilon \|x^{(i)}\|_2$$

where  $\epsilon > 0$  is the convergence threshold.

▶ more

# Summary of PNPG Approach

Combine

- ▶ convex-set projection with
- ▶ Nesterov acceleration.

Apply

- ▶ adaptive step size,
- ▶ restart.

## Why?

- ▶ Thanks to step-size adaptation, no need for Lipschitz continuity of the gradient of the NLL.
- ▶  $\text{dom } \mathcal{L}$  does not have to be  $\mathbb{R}^P$ .

Consequently, extends the application of the Nesterov's acceleration<sup>†</sup> to more general measurement models than those used previously:

- ▶ we present the first application of this acceleration to Poisson compressed sensing.

---

<sup>†</sup>Y. Nesterov, "A method of solving a convex programming problem with convergence rate  $O(1/k^2)$ ," *Sov. Math. Dokl.*, vol. 27, 1983, pp. 372–376.

## Relationship with FISTA I

PNPG can be thought of as a generalized fast iterative shrinkage-thresholding algorithm (FISTA) (Beck and Teboulle 2009a) that accommodates

- ▶ convex constraints,
- ▶ more general NLLs,<sup>‡</sup> and (increasing) adaptive step size
  - ▶ thanks to this step-size adaptation, PNPG *does not* require Lipschitz continuity of the gradient of the NLL.

---

<sup>‡</sup>FISTA has been developed for the linear Gaussian model.

## Relationship with FISTA II

- ▶ Need  $B^{(i)}$  in (2a) to derive theoretical guarantee for convergence speed of the PNPG iteration.
- ▶ In contrast with PNPG, FISTA has a non-increasing step size  $\beta^{(i)}$ , which allows for setting

$$B^{(i)} = 1$$

in (2b) for all  $i$ :<sup>§</sup>

$$\theta^{(i)} = \frac{1}{2} \left[ 1 + \sqrt{1 + 4(\theta^{(i-1)})^2} \right].$$

---

<sup>§</sup>Y. Nesterov, “A method of solving a convex programming problem with convergence rate  $O(1/k^2)$ ,” *Sov. Math. Dokl.*, vol. 27, 1983, pp. 372–376.

# Outline

## Projected Nesterov's Proximal-Gradient Algorithm

Background

Optimization problem

Reconstruction algorithm

Convergence analysis

## Applications

Poisson compressed sensing

Linear model with Gaussian noise

X-Ray CT Reconstruction from Polychromatic Measurements

Polychromatic X-ray CT model

Mass-attenuation spectrum

Minimization algorithm

Numerical Examples

## Conclusion

## Theorem (Convergence of PNPG)

- ▶ Assume convex and differentiable NLL  $\mathcal{L}(\mathbf{x})$  and that the boundary set  $C \setminus \text{dom } \mathcal{L}$  is empty.
- ▶ Consider the PNPG iteration without restart.

The convergence of PNPG iterates  $\mathbf{x}^{(k)}$  to the minimum point

$$\mathbf{x}^* = \arg \min_{\mathbf{x}} f(\mathbf{x})$$

is bounded as follows:

$$f(\mathbf{x}^{(k)}) - f(\mathbf{x}^*) \leq 2 \frac{\|\mathbf{x}^{(0)} - \mathbf{x}^*\|_2^2}{\left(\sum_{i=1}^k \sqrt{\beta^{(i)}}\right)^2}$$

for all  $k \geq 1$ .

## Definition (Inexact Proximal Operator (Villa *et al.* 2013))

We say that  $x$  is an approximation of  $\text{prox}_{ur} a$  with  $\varepsilon$ -precision, denoted by

$$x \underset{\varepsilon}{\approx} \text{prox}_{ur} a$$

if

$$\frac{a - x}{u} \in \partial_{\frac{\varepsilon^2}{2u}} r(x).$$

**Note:** This definition implies

$$\|x - \text{prox}_{ur} a\|_2^2 \leq \varepsilon^2.$$

## Theorem (Convergence of PNPG for Inexact PG Steps)

$$f(\mathbf{x}^{(k)}) - f(\mathbf{x}^*) \leq 2 \frac{\|\mathbf{x}^{(0)} - \mathbf{x}^*\|_2^2 + \mathcal{E}^{(k)}}{\left(\sum_{i=1}^k \sqrt{\beta^{(i)}}\right)^2}$$

where

$$\mathcal{E}^{(k)} \triangleq \sum_{i=1}^k (\theta^{(i)} \varepsilon^{(i)})^2$$

*is the cumulative error term that accounts for the inexact PG steps.*

## Corollary

Under the condition of the Theorem, the convergence of PNPG iterates  $\mathbf{x}^{(k)}$  is bounded as follows:

$$f(\mathbf{x}^{(k)}) - f(\mathbf{x}^*) \leq 2 \frac{\|\mathbf{x}^{(0)} - \mathbf{x}^*\|_2^2 + \mathcal{E}^{(k)}}{k^2 \beta_{\min}^{(k)}}$$

  
 $\mathcal{O}(k^{-2})$

where

$$\beta_{\min}^{(k)} = \min_{i \leq k} \beta^{(i)}$$

is the minimum step size up until iteration  $k$ .

# Outline

## Projected Nesterov's Proximal-Gradient Algorithm

Background

Optimization problem

Reconstruction algorithm

Convergence analysis

## Applications

Poisson compressed sensing

Linear model with Gaussian noise

## X-Ray CT Reconstruction from Polychromatic Measurements

Polychromatic X-ray CT model

Mass-attenuation spectrum

Minimization algorithm

Numerical Examples

## Conclusion

# Outline

## Projected Nesterov's Proximal-Gradient Algorithm

Background

Optimization problem

Reconstruction algorithm

Convergence analysis

## Applications

Poisson compressed sensing

Linear model with Gaussian noise

X-Ray CT Reconstruction from Polychromatic Measurements

Polychromatic X-ray CT model

Mass-attenuation spectrum

Minimization algorithm

Numerical Examples

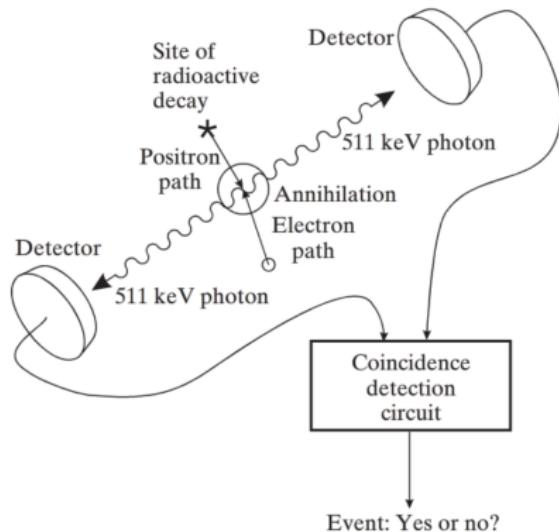
## Conclusion

# Introduction

Signal reconstruction from Poisson-distributed measurements with affine model for the mean-signal intensity is important for

- ▶ tomographic (Ollinger and Fessler 1997),
- ▶ astronomic, optical, microscopic (Bertero *et al.* 2009),
- ▶ hyperspectral (Willett *et al.* 2014)

imaging.



PET: Coincidence detection due to positron decay and annihilation  
(Prince and Links 2015).

## Measurement Model

$N$  independent measurements  $\mathbf{y} = (y_n)_{n=1}^N$  following the Poisson distribution with means

$$[\phi(\mathbf{x})]_n = [\Phi \mathbf{x} + \mathbf{b}]_n$$

where

$$\mathbf{x} = (x_i)_{i=1}^p \in C$$

is the *unknown*  $p \times 1$  signal vector that we wish to reconstruct, and

$$\Phi \in \mathbb{R}_+^{N \times p}, \quad \mathbf{b}, \quad C$$

are the *known sensing matrix*, *intercept term* ¶, and *nonempty closed convex set* that  $\mathbf{x}$  belongs to.

---

¶the intercept  $\mathbf{b}$  models background radiation and scattering, obtained, e.g., by calibration before the measurements  $\mathbf{y}$  have been collected

## Existing Work

### The sparse Poisson-intensity reconstruction algorithm (SPIRAL)<sup>||</sup>

- ▶ approximates the logarithm function in the underlying NLL by adding a small positive term to it and then
- ▶ descends a regularized NLL objective function with proximal steps that employ Barzilai-Borwein (BB) step size in each iteration, followed by backtracking.

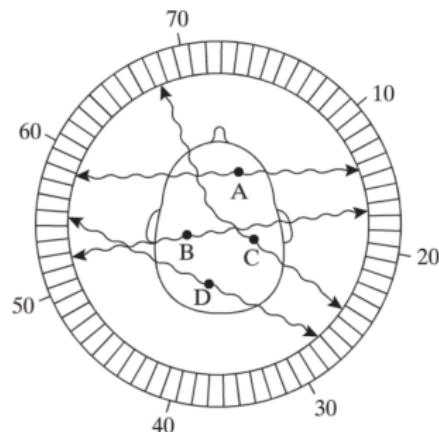


---

<sup>||</sup>Z. T. Harmany *et al.*, "This is SPIRAL-TAP: Sparse Poisson intensity reconstruction algorithms—theory and practice," *IEEE Trans. Image Process.*, vol. 21, no. 3, pp. 1084–1096, Mar. 2012.

# PET Image Reconstruction

- ▶  $128 \times 128$  concentration map  $x$ .
- ▶ Collect the photons from 90 equally spaced directions over  $180^\circ$ , with 128 radial samples at each direction,
- ▶ Background radiation, scattering effect, and accidental coincidence combined together lead to a known intercept term  $b$ .
- ▶ The elements of the intercept term are set to a constant equal to 10 % of the sample mean of  $\Phi x$ :  $b = \frac{\mathbf{1}^T \Phi x}{10N} \mathbf{1}$ .



The model, choices of parameters in the PET system setup, and concentration map have been adopted from Image Reconstruction Toolbox (IRT) (Fessler n.d.).

## Numerical Example

- ▶ Main metric for assessing the performance of the compared algorithms is relative square error (RSE)

$$\text{RSE} = \frac{\|\hat{x} - x_{\text{true}}\|_2^2}{\|x_{\text{true}}\|_2^2}$$

where  $x_{\text{true}}$  and  $\hat{x}$  are the true and reconstructed signal, respectively.

- ▶ All iterative methods use the convergence threshold

$$\epsilon = 10^{-6}$$

and have the maximum number of iterations limited to  $10^4$ .

- ▶ Regularization constant  $u$  has the form

$$u = 10^a.$$

We vary  $a$  in the range  $[-6, 3]$  with a grid size of 0.5 and search for the reconstructions with the best RSE performance.

## Compared Methods

- ▶ Filtered backprojection (FBP) (Ollinger and Fessler 1997) and
- ▶ PG methods that aim at minimizing  $f(\mathbf{x})$  with nonnegative  $\mathbf{x}$ :

$$C = \mathbb{R}_+^p.$$

All iterative methods initialized by FBP reconstructions.

## PG Methods

- ▶ PNPG with

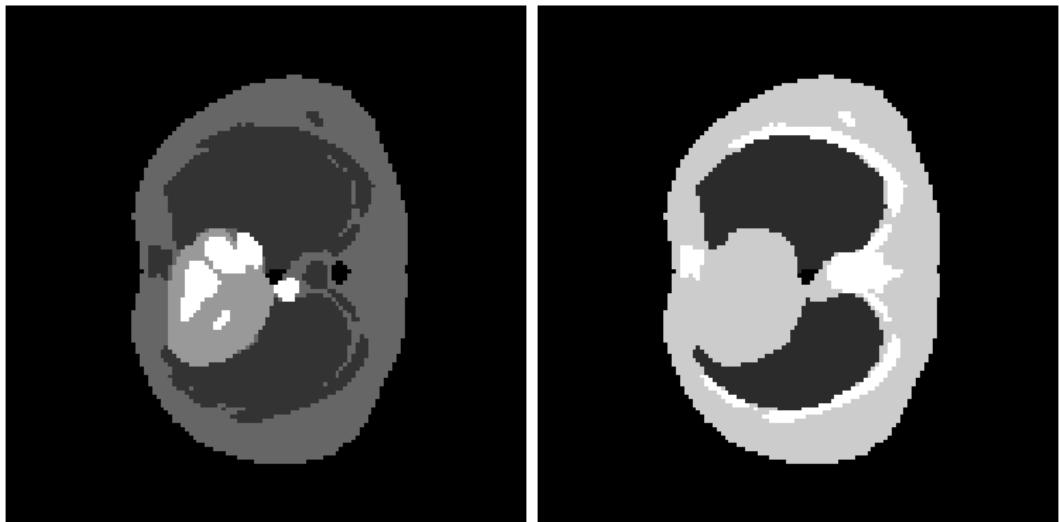
- ▶ adaptive step-size parameters

$$(\mathfrak{n}, \mathfrak{m}) = (4, 4), \quad \xi = 0.8$$

(unless specified otherwise) and

- ▶ the (initial) inner-iteration convergence constant  $\eta = 10^{-2}$ .

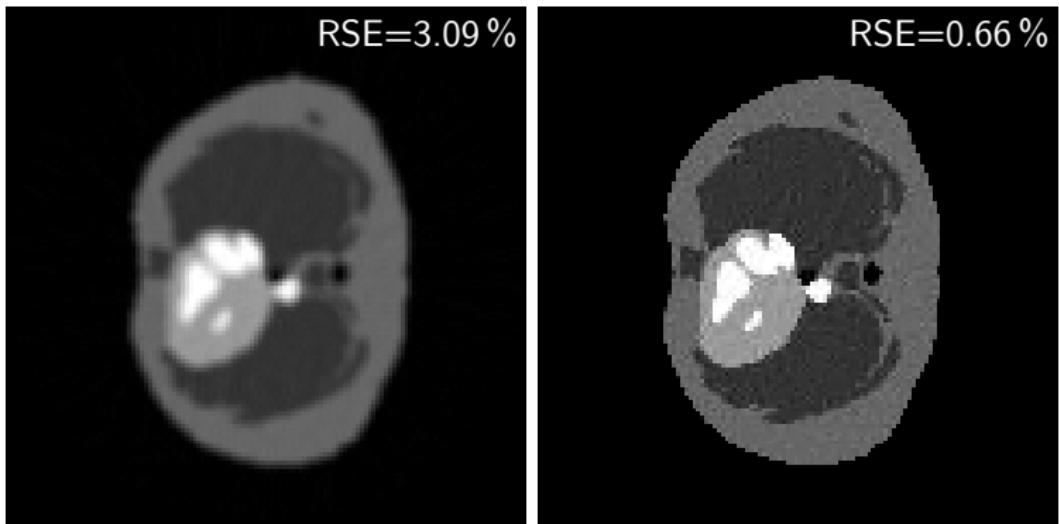
- ▶ AT (Auslender and Teboulle 2006) implemented in the templates for first-order conic solvers (TFOCS) package (Becker *et al.* 2011) with a periodic restart every 200 iterations (tuned for its best performance) and our proximal mapping.
- ▶ SPIRAL, when possible.



(a) radio-isotope concentration

(b) attenuation map

**Figure 4:** (a) True emission image and (b) density map.



**Figure 5:** Reconstructions of the emission concentration map.

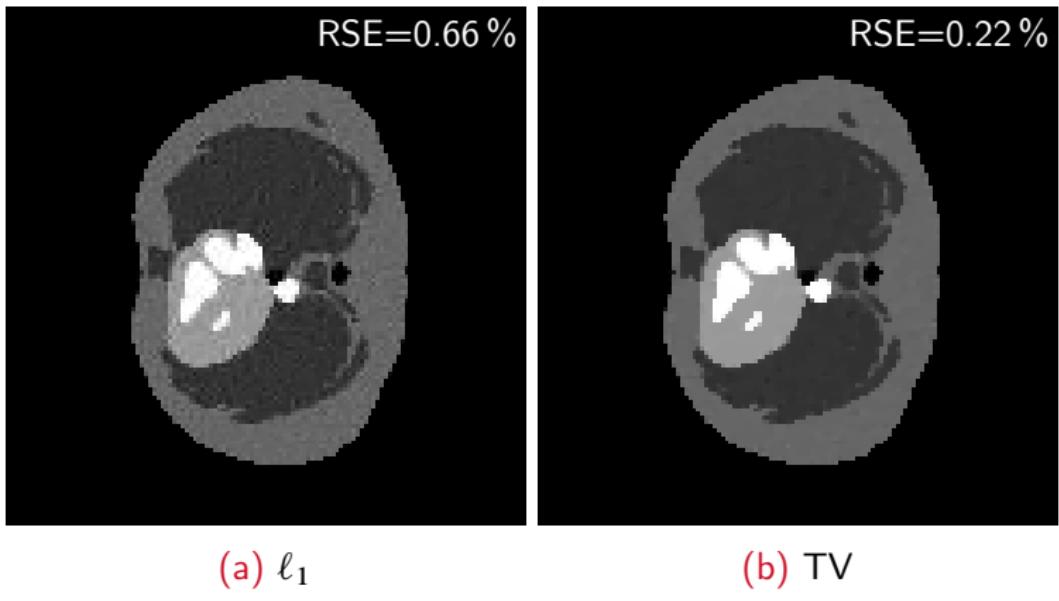


Figure 6: Comparison of the two sparsity regularizations.

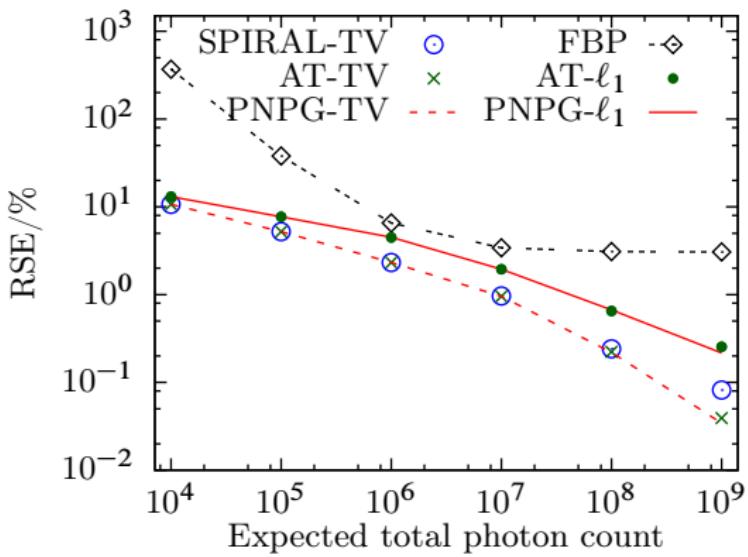
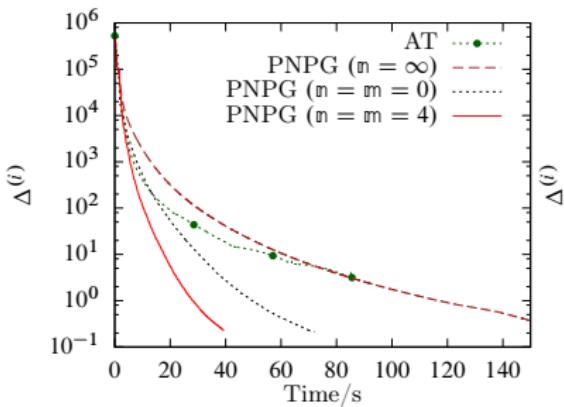


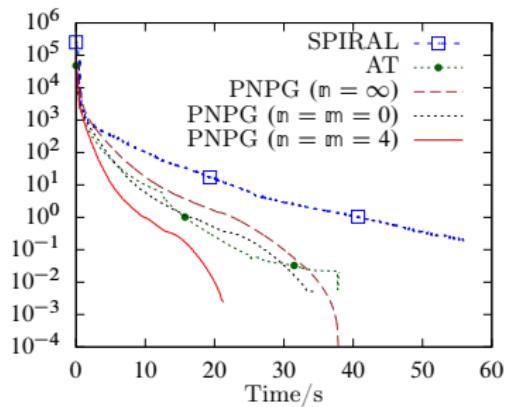
Figure 7: Minimum average RSEs as functions of SNR  $\mathbf{1}^T \mathbf{E}(\mathbf{y})$ .

## Comments

- ▶ The RSEs achieved by the methods that employ TV regularization are 1.2 to 6.3 times smaller than those by  $\ell_1$ -norm regularization.
- ▶ As the SNR increases, the convergence points of SPIRAL-TV and PNPG-TV diverge, which explains the difference between the RSEs of the two methods at large SNRs.
  - ▶ This trend is observed already when  $\mathbf{1}^T \mathbf{E}(\mathbf{y}) = 10^7$  in Fig. 7.



(a)  $\ell_1$ ,  $\mathbf{1}^T E(y) = 10^8$



(b) TV,  $\mathbf{1}^T E(y) = 10^7$

Figure 8: Centered objectives as functions of CPU time.

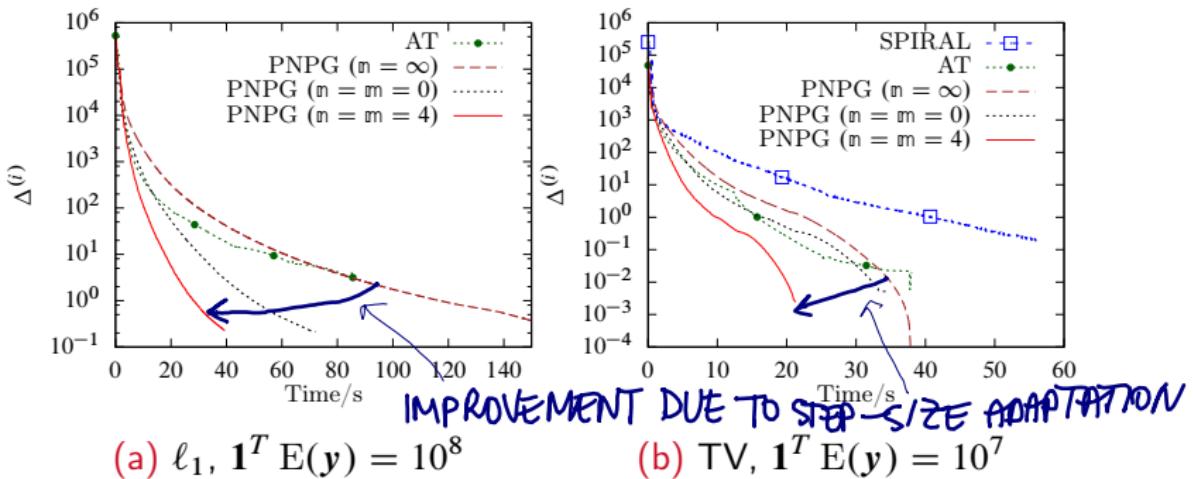


Figure 8: Centered objectives as functions of CPU time.

# Outline

## Projected Nesterov's Proximal-Gradient Algorithm

Background

Optimization problem

Reconstruction algorithm

Convergence analysis

## Applications

Poisson compressed sensing

### Linear model with Gaussian noise

X-Ray CT Reconstruction from Polychromatic Measurements

Polychromatic X-ray CT model

Mass-attenuation spectrum

Minimization algorithm

Numerical Examples

## Conclusion

## Linear Model with Gaussian Noise

$$\mathcal{L}(\mathbf{x}) = \frac{1}{2} \|\mathbf{y} - \Phi \mathbf{x}\|_2^2$$

where  $\mathbf{y} \in \mathbb{R}^N$  is the measurement vector.

Minimization of the corresponding objective function  $f(\mathbf{x})$  can be thought of as a *generalized analysis basis pursuit denoising (BPDN) problem with a convex signal constraint*.

We select the  $\ell_1$ -norm sparsifying signal penalty with linear map:

$$\psi(\mathbf{x}) = \Psi^T \mathbf{x}.$$

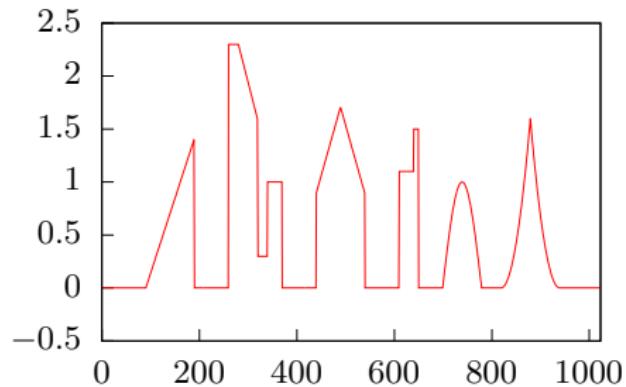


Figure 9: True signal.

## Comments I

- ▶ Here, we compare with more methods:
  - ▶ sparse reconstruction by separable approximation (SpaRSA) (Wright *et al.* 2009),
  - ▶ generalized forward-backward splitting (GFB) (Raguet *et al.* 2013),
  - ▶ primal-dual splitting (PDS) (Condat 2013).
- ▶ We select the regularization parameter  $u$  as

$$u = 10^a U, \quad U \triangleq \|\Psi^T \nabla \mathcal{L}(\mathbf{0})\|_\infty$$

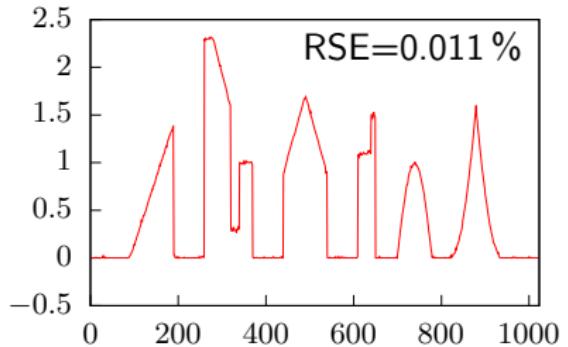
where  $a$  is an integer selected from the interval  $[-9, -1]$  and  $U$  is an upper bound on  $u$  of interest.

- ▶ Choose the nonnegativity convex set:

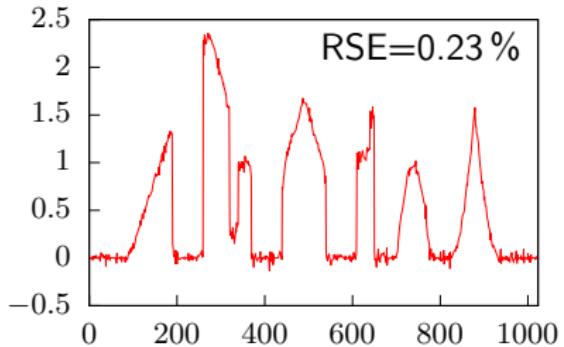
$$C = \mathbb{R}_+^p.$$

## Comments II

- ▶ If we remove the convex-set constraint by setting  $C = \mathbb{R}^p$ , iteration (2a)–(2d) reduces to the Nesterov's proximal gradient iteration with adaptive step size that imposes signal sparsity *only* in the analysis form (termed NPG<sub>S</sub>).

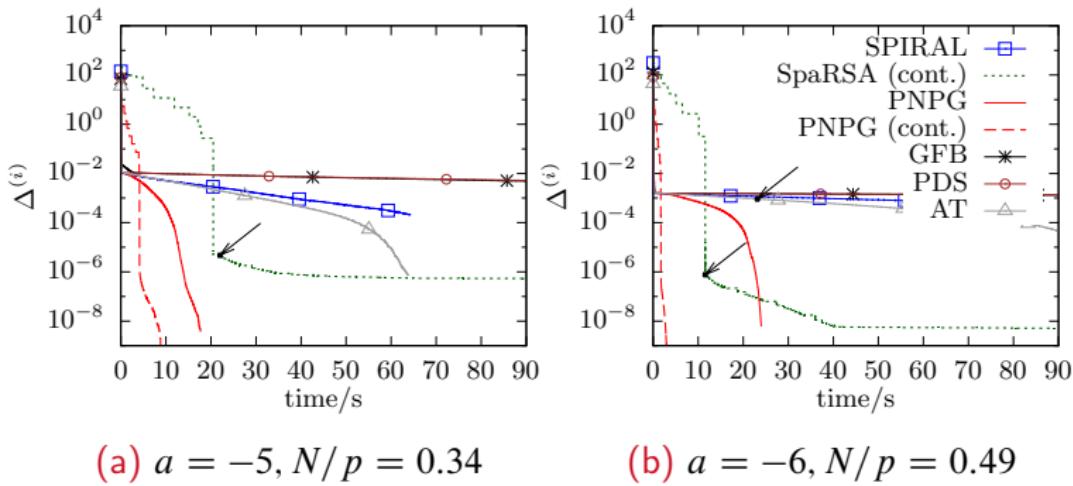


(a) PNPG

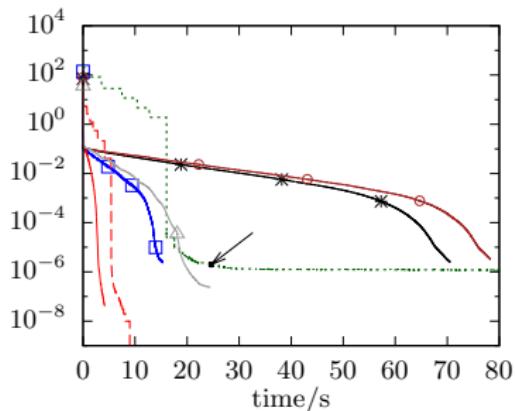


(b) NPG<sub>S</sub>

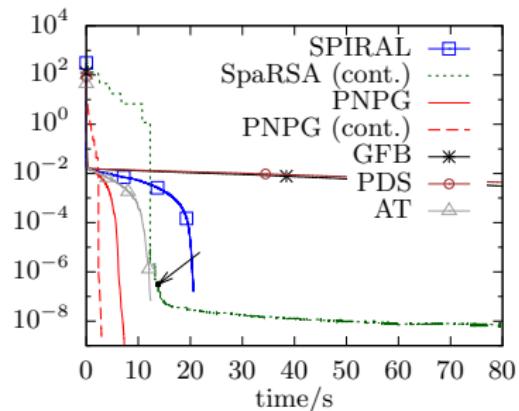
Figure 10: PNPG and NPG<sub>S</sub> reconstructions for  $N/p = 0.34$ .



**Figure 11:** Centered objectives as functions of CPU time.

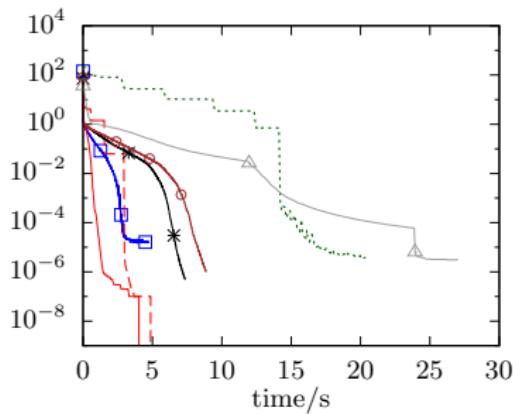


(a)  $a = -4, N/p = 0.34$

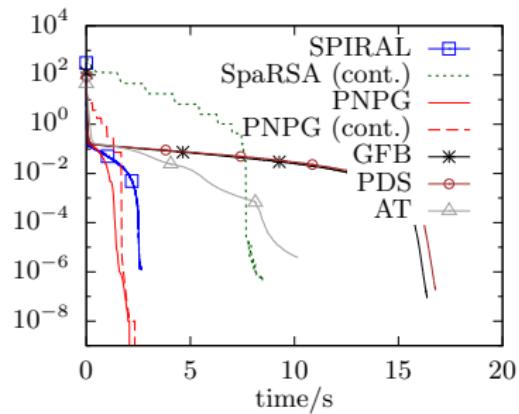


(b)  $a = -5, N/p = 0.49$

Figure 12: Centered objectives as functions of CPU time.



(a)  $a = -3, N/p = 0.34$



(b)  $a = -4, N/p = 0.49$

**Figure 13:** Centered objectives as functions of CPU time.

# Outline

## Projected Nesterov's Proximal-Gradient Algorithm

Background

Optimization problem

Reconstruction algorithm

Convergence analysis

## Applications

Poisson compressed sensing

Linear model with Gaussian noise

## X-Ray CT Reconstruction from Polychromatic Measurements

Polychromatic X-ray CT model

Mass-attenuation spectrum

Minimization algorithm

Numerical Examples

## Conclusion

# X-ray CT

An X-ray CT scan consists of multiple projections with the beam intensity measured by multiple detectors.

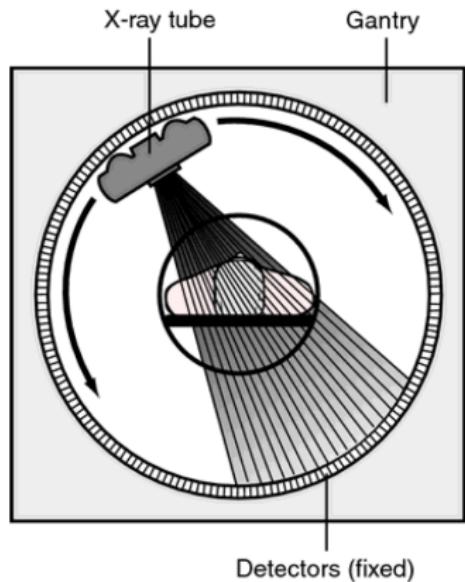


Figure 14: Fan-beam CT system.

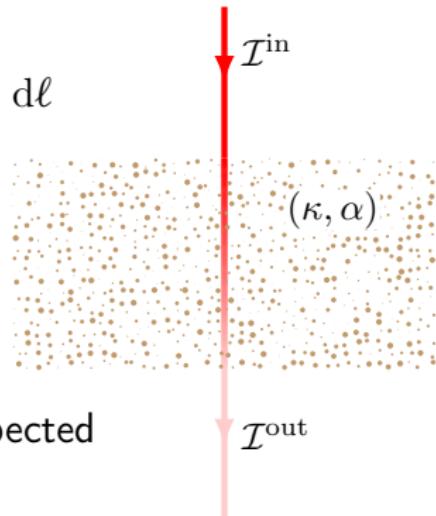
## Exponential Law of Absorption

The fraction  $d\mathcal{I}/\mathcal{I}$  of plane-wave intensity lost in traversing an infinitesimal thickness  $d\ell$  at Cartesian coordinates  $(x, y)$  is proportional to  $d\ell$ :

$$\frac{d\mathcal{I}}{\mathcal{I}} = - \underbrace{\mu(x, y, \varepsilon)}_{\text{attenuation}} d\ell = - \underbrace{\kappa(\varepsilon)\alpha(x, y)}_{\text{separable}} d\ell$$

where

- ▶  $\kappa(\varepsilon) \geq 0$  is the **mass attenuation function** of the material,
- ▶  $\alpha(x, y) \geq 0$  is the **density map** of the inspected object, and
- ▶  $\varepsilon$  is **photon energy**.



To obtain the intensity decrease along a straight-line path  $\ell = \ell(x, y)$ , integrate along  $\ell$  and over  $\varepsilon$ . The underlying measurement model is **nonlinear**.

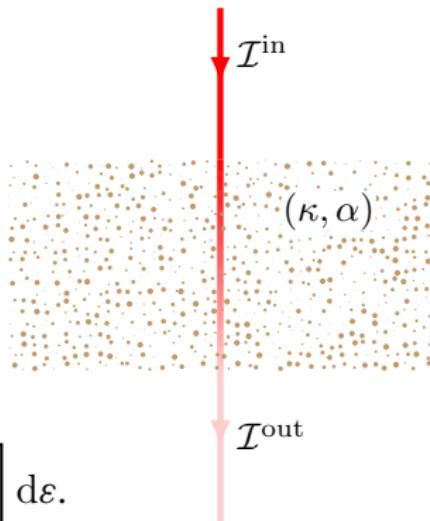
# Polychromatic X-ray CT Model

- ▶ Incident energy  $\mathcal{I}^{\text{in}}$  spreads along photon energy  $\varepsilon$  with density  $\iota(\varepsilon)$ :

$$\int \iota(\varepsilon) d\varepsilon = \mathcal{I}^{\text{in}}.$$

- ▶ Noiseless energy measurement obtained upon traversing a straight line  $\ell = \ell(x, y)$  through an object composed of a single material:

$$\mathcal{I}^{\text{out}} = \int \iota(\varepsilon) \exp \left[ -\kappa(\varepsilon) \int_{\ell} \alpha(x, y) d\ell \right] d\varepsilon.$$



## Linear Reconstruction Artifacts

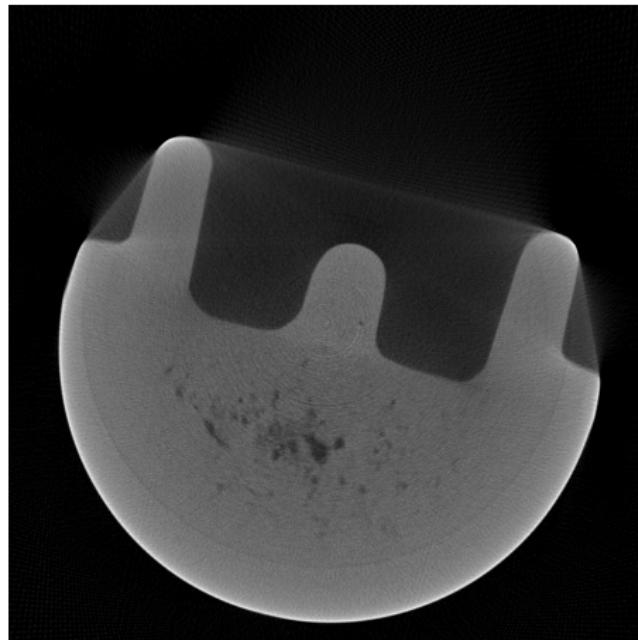


Figure 15: FBP reconstruction of an industrial object.

Note the cupping and streaking artifacts of the linear FBP reconstruction, applied to  $\ln \mathcal{I}^{\text{out}}$ .

## Problem Formulation and Goal

Assume that both

- o the incident spectrum  $\iota(\varepsilon)$  of X-ray source and
- o mass attenuation function  $\kappa(\varepsilon)$  of the object

are **unknown**.

Goal: Estimate the density map  $\alpha(x, y)$ .

## Problem Formulation and Goal

Assume that both

- o the incident spectrum  $\iota(\varepsilon)$  of X-ray source and
- o mass attenuation function  $\kappa(\varepsilon)$  of the object

are **unknown**.

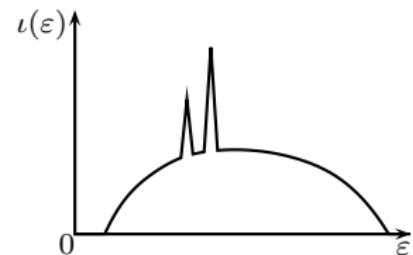
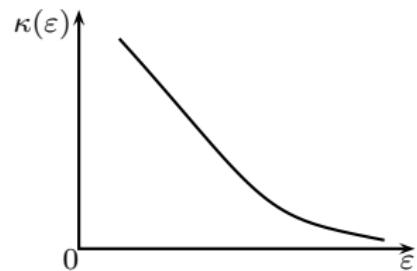
**Goal:** Estimate the density map  $\alpha(x, y)$ .

# Polychromatic X-ray CT Model Using Mass-Attenuation Spectrum

- ▶ Mass attenuation  $\kappa(\varepsilon)$  and incident spectrum density  $\iota(\varepsilon)$  are both functions of  $\varepsilon$ .
- ▶ Idea. Write the model as integrals of  $\kappa$  rather than  $\varepsilon$ :

$$\mathcal{I}^{\text{in}} = \int \iota(\kappa) d\kappa = \iota^L(0)$$

$$\begin{aligned}\mathcal{I}^{\text{out}} &= \int \iota(\kappa) \exp\left[-\kappa \int_{\ell} \alpha(x, y) d\ell\right] d\kappa \\ &= \iota^L\left(\int_{\ell} \alpha(x, y) d\ell\right)\end{aligned}$$



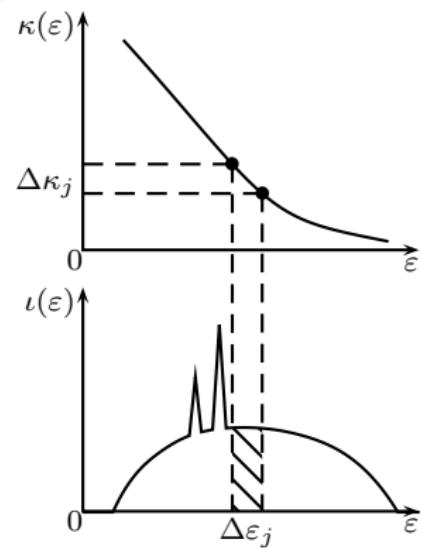
☞ Need to estimate **one** function,  $\iota(\kappa)$ , rather than **two**,  $\iota(\varepsilon)$  and  $\kappa(\varepsilon)$ !

# Polychromatic X-ray CT Model Using Mass-Attenuation Spectrum

- ▶ Mass attenuation  $\kappa(\varepsilon)$  and incident spectrum density  $\iota(\varepsilon)$  are both functions of  $\varepsilon$ .
- ▶ **Idea.** Write the model as integrals of  $\kappa$  rather than  $\varepsilon$ :

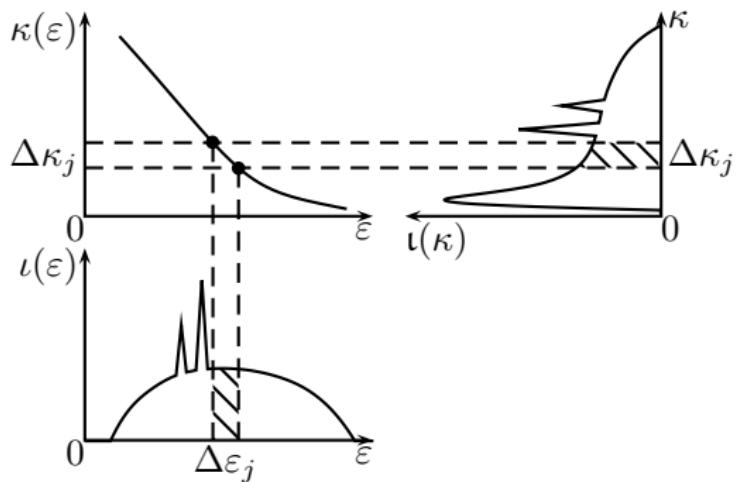
$$\mathcal{I}^{\text{in}} = \int \iota(\kappa) d\kappa = \iota^L(0)$$

$$\begin{aligned}\mathcal{I}^{\text{out}} &= \int \iota(\kappa) \exp\left[-\kappa \int_{\ell} \alpha(x, y) d\ell\right] d\kappa \\ &= \iota^L\left(\int_{\ell} \alpha(x, y) d\ell\right)\end{aligned}$$



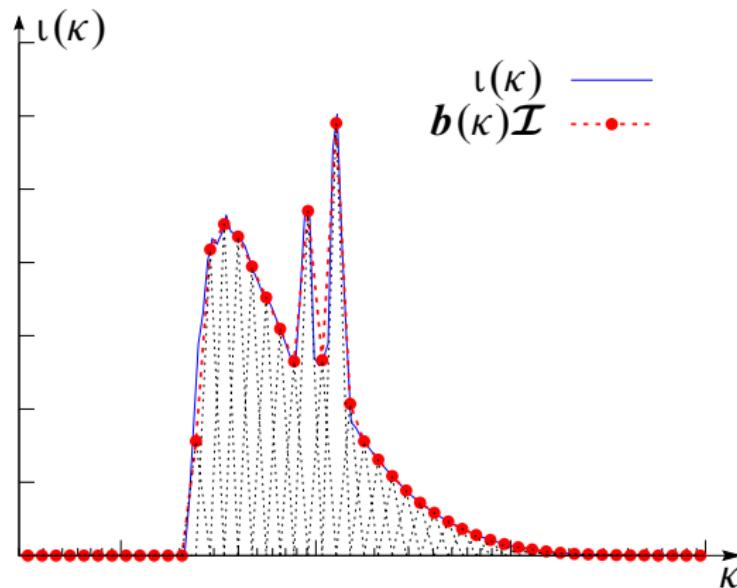
☞ Need to estimate **one** function,  $\iota(\kappa)$ , rather than **two**,  $\iota(\varepsilon)$  and  $\kappa(\varepsilon)$ !

# Mass-Attenuation Spectrum



**Figure 16:** Relationship between mass attenuation  $\kappa$ , incident spectrum  $\iota$ , photon energy  $\varepsilon$ , and **mass attenuation spectrum**  $\iota(\kappa)$ .

# Basis-function expansion of mass-attenuation spectrum

$$\iota(\kappa) = \mathbf{b}(\kappa)\mathcal{I}$$


**Figure 17:** B1-spline expansion  $\iota(\kappa) = \mathbf{b}(\kappa)\mathcal{I}$ , where the B1-spline basis is  $\underbrace{\mathbf{b}(\kappa)}_{1 \times J} = [b_1(\kappa), b_2(\kappa), \dots, b_J(\kappa)]$ .  $\iota(\kappa) \geq 0$  implies  $\mathcal{I} \succeq \mathbf{0}$ .

## Noiseless Measurement Model

$N \times 1$  vector of noiseless energy measurements:

$$\mathcal{I}^{\text{out}}(x, \mathcal{I}) = b_o^L(\Phi x) \mathcal{I}$$

where  $\Phi$  is the known projection matrix,

- ▶  $x = (x_i)_{i=1}^p \succeq \mathbf{0}$  is an *unknown*  $p \times 1$  density-map vector representing the 2D image we wish to reconstruct, and
- ▶

$$\mathcal{I} = (\mathcal{I}_j)_{j=1}^J \succeq \mathbf{0}$$

is an *unknown*  $J \times 1$  vector of corresponding mass-attenuation basis-function coefficients.

## Poisson Noise Model

For independent Poisson measurements  $\mathcal{E} = (\mathcal{E}_n)_{n=1}^N$ , the NLL is

$$\mathcal{L}(x, \mathcal{I}) = \mathbf{1}^T [\mathcal{I}^{\text{out}}(x, \mathcal{I}) - \mathcal{E}] - \sum_{n, \mathcal{E}_n \neq 0} \mathcal{E}_n \ln \frac{\mathcal{I}_n^{\text{out}}(x, \mathcal{I})}{\mathcal{E}_n}.$$

## Penalized NLL

- ▶ objective function

$$f(\mathbf{x}, \mathcal{I}) = \mathcal{L}(\mathbf{x}, \mathcal{I}) + u \underbrace{[\|\psi(\mathbf{x})\|_1 + \mathbb{I}_C(\mathbf{x})]}_{r(\mathbf{x})} + \mathbb{I}_{\mathbb{R}_+^J}(\mathcal{I})$$

- ▶ NLL
- ▶ penalty term  
 $u > 0$  is a scalar tuning constant  
we select  $\psi(\mathbf{x})$  = gradient map,  
 $C = \mathbb{R}_+^J$

## Penalized NLL

- ▶ objective function

$$f(\mathbf{x}, \mathcal{I}) = \mathcal{L}(\mathbf{x}, \mathcal{I}) + u \underbrace{[\|\psi(\mathbf{x})\|_1 + \mathbb{I}_C(\mathbf{x})]}_{r(\mathbf{x})} + \mathbb{I}_{\mathbb{R}_+^J}(\mathcal{I})$$

- ▶ NLL

- ▶ penalty term

$u > 0$  is a scalar tuning constant

we select  $\psi(\mathbf{x})$  = gradient map,

$$\mathcal{C} = \mathbb{R}_+^J$$

## Penalized NLL

- ▶ objective function

$$f(\mathbf{x}, \mathcal{I}) = \mathcal{L}(\mathbf{x}, \mathcal{I}) + u \underbrace{[\|\psi(\mathbf{x})\|_1 + \mathbb{I}_C(\mathbf{x})]}_{r(\mathbf{x})} + \mathbb{I}_{\mathbb{R}_+^J}(\mathcal{I})$$

- ▶ NLL

- ▶ penalty term

$u > 0$  is a scalar tuning constant

we select  $\psi(\mathbf{x}) = \text{gradient map}$ ,

$$C = \mathbb{R}_+^J$$

## Goal and Minimization Approach

**Goal:** Estimate the density-map and mass-attenuation spectrum parameters

$$(\mathbf{x}, \mathcal{I})$$

by minimizing the penalized NLL  $f(\mathbf{x}, \mathcal{I})$ .

**Approach:** A block coordinate-descent that uses

- ▶ Nesterov's proximal-gradient (NPG) (Nesterov 1983) and
- ▶ limited-memory Broyden-Fletcher-Goldfarb-Shanno with box constraints (L-BFGS-B) (Byrd *et al.* 1995; Zhu *et al.* 1997)

methods to update estimates of the **density map** and **mass-attenuation spectrum** parameters.

We refer to this iteration as NPG-BFGS algorithm.

## Numerical Examples

- ▶ convergence threshold:

$$\epsilon = 10^{-6}$$

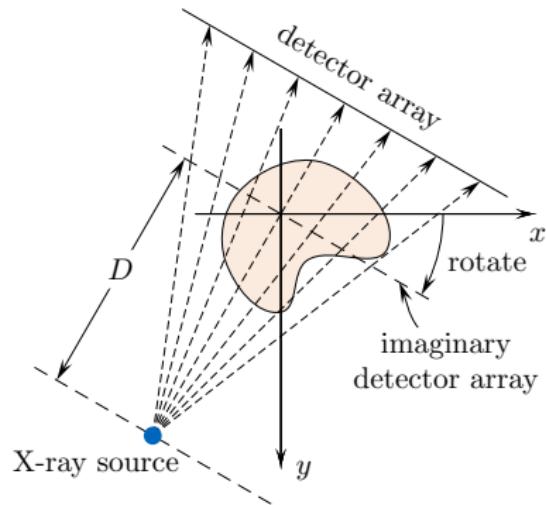
- ▶ B1-spline constants set to satisfy

$$\begin{aligned} J &= 20, & \text{\# basis functions} \\ q^J &= 10^3, & \text{span} \\ \kappa_0 q^{\lceil 0.5(J+1) \rceil} &= 1, & \text{centering} \end{aligned}$$

Implementation available at [github.com/isucsp/imgRecSrc.](https://github.com/isucsp/imgRecSrc)

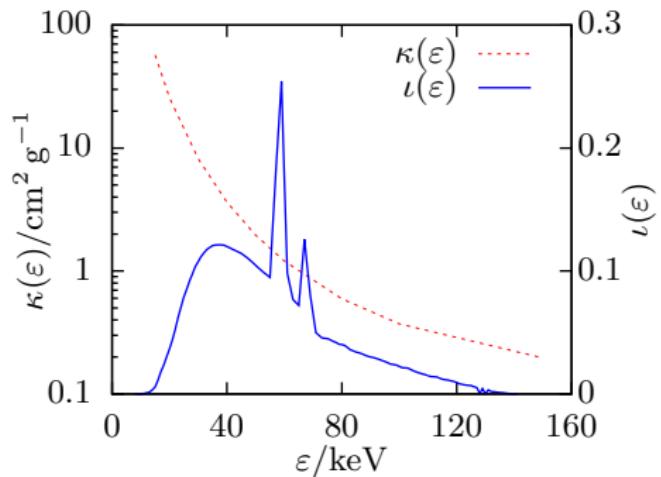
## Simulated X-ray CT Example

- ▶ Equi-spaced fan-beam projections over  $360^\circ$ ,
- ▶ X-ray source to rotation center is  $2000 \times$  detector size,
- ▶ measurement array size of 512 elements,
- ▶ image to reconstruct is of  $512 \times 512$ , and
- ▶ performance metric is the RSE of an estimate  $\hat{\mathbf{x}}$  of the signal coefficient vector:

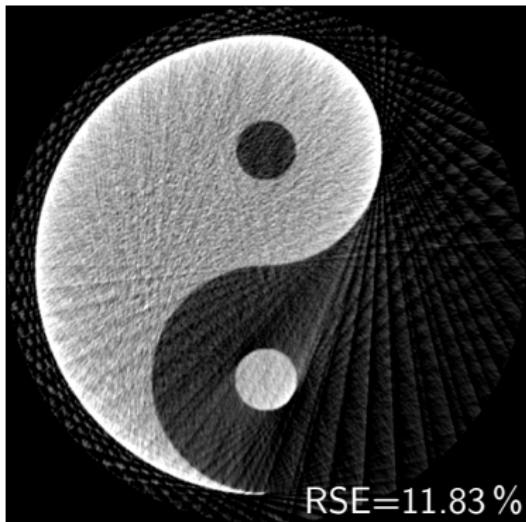


$$\text{RSE}\{\hat{\mathbf{x}}\} = 1 - \left( \frac{\hat{\mathbf{x}}^T \mathbf{x}_{\text{true}}}{\|\hat{\mathbf{x}}\|_2 \|\mathbf{x}_{\text{true}}\|_2} \right)^2$$

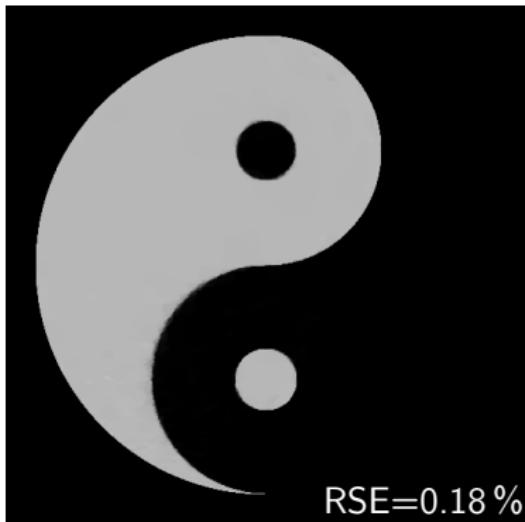
## Simulated X-ray CT Example



- ▶ Incident X-ray spectrum from tungsten anode X-ray tubes at 140 keV with 5 % relative voltage ripple, and
- ▶ using photon-energy discretization with 130 equi-spaced discretization points over the range 20 keV to 140 keV.



(a) FBP



(b) NPG-BFGS

Figure 18: Reconstructions from 60 projections.

## Simulated X-ray CT Example

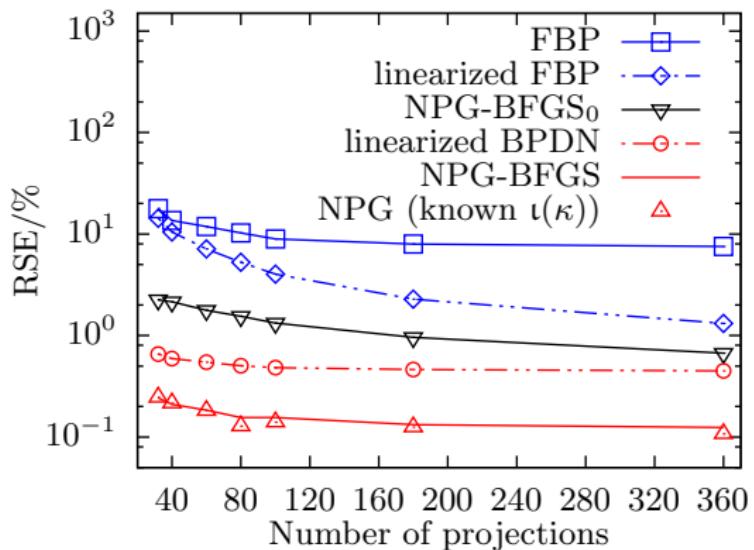


Figure 19: Average RSEs as functions of the number of projections.

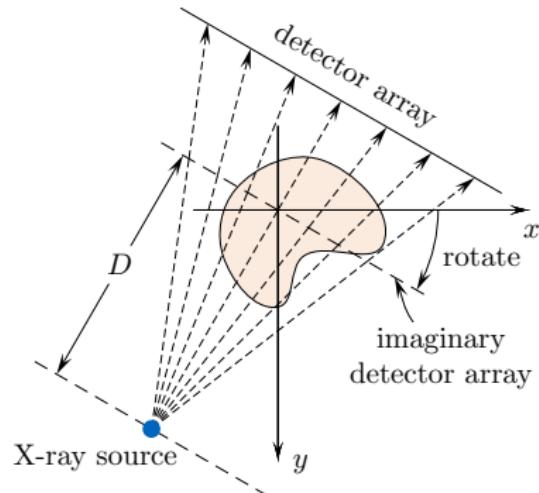
## Real X-ray CT Example I

- ▶ 360 equi-spaced fan-beam projections with  $1^\circ$  spacing,
- ▶ X-ray source to rotation center is  $3492 \times$  detector size,
- ▶ measurement array size of 694 elements,
- ▶ projection matrix  $\Phi$  constructed directly on **GPU** (multi-thread version on CPU is also available) with full circular mask (D. *et al.* 2011),

yielding a nonlinear estimation problem with  $N = 694 \times 360$  measurements and an  $512 \times 512$  image to reconstruct.

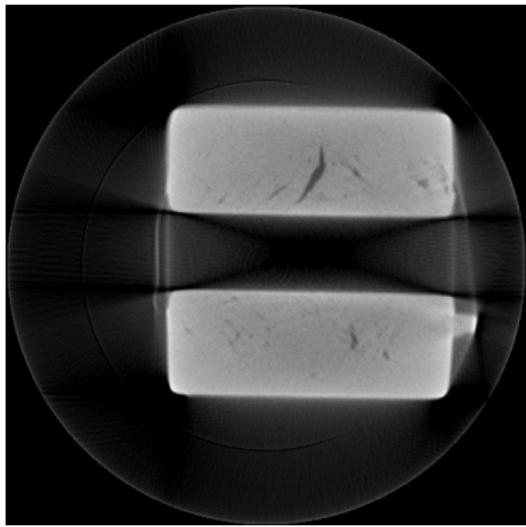
# Real X-ray CT Example I

- ▶ 360 equi-spaced fan-beam projections with  $1^\circ$  spacing,
- ▶ X-ray source to rotation center is  $3492 \times$  detector size,
- ▶ measurement array size of 694 elements,
- ▶ projection matrix  $\Phi$  constructed directly on GPU (multi-thread version on CPU is also available) with full circular mask (D. et al. 2011).

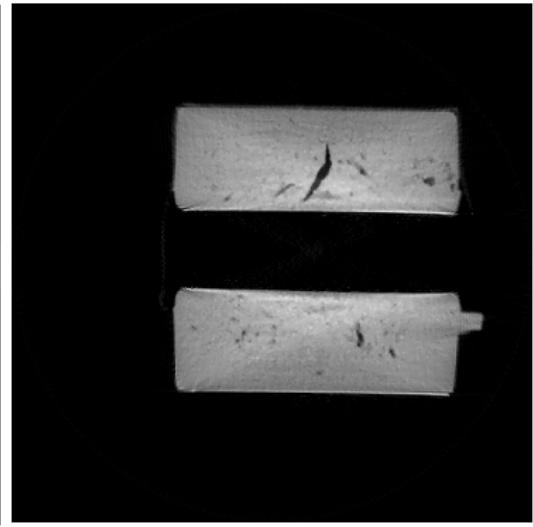


yielding a nonlinear estimation problem with  $N = 694 \times 360$  measurements and an  $512 \times 512$  image to reconstruct.

Implementation available at [github.com/isucsp/imgRecSrc](https://github.com/isucsp/imgRecSrc).



(a) FBP



(b) NPG-BFGS ( $u = 10^{-5}$ )

Figure 20: Real X-ray CT: Full projections.

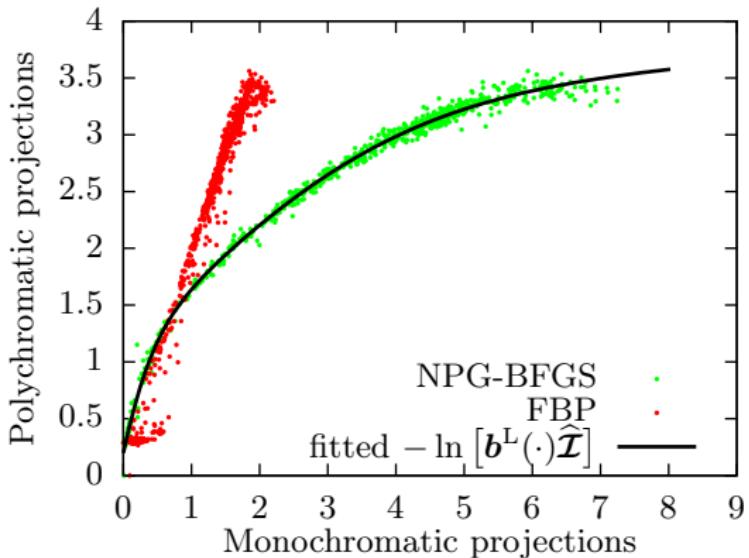
## Comments

Our reconstruction eliminates

- ▶ the streaking artifacts across the air around the object,
- ▶ the cupping artifacts with high intensity along the border.

Note that the regularization constant  $u$  is tuned for the best reconstruction.

## Inverse Linearization Function Estimate



**Figure 21:** The polychromatic measurements as function of the monochromatic projections and its corresponding fitted curve.

Observe the biased residual for FBP, the unbiased residual for NPG-BFGS and its increasing variance.

## Real X-ray CT Example II

- ▶ X-ray source to rotation center is 8696 times of a single detector size,
- ▶ measurement array size of 1380 elements,
- ▶ projection matrix  $\Phi$  constructed directly on **GPU** (multi-thread version on CPU is also available) with full circular mask.

yielding a nonlinear estimation problem with  $N = 1380 \times 360$  measurements and an  $1024 \times 1024$  image to reconstruct.

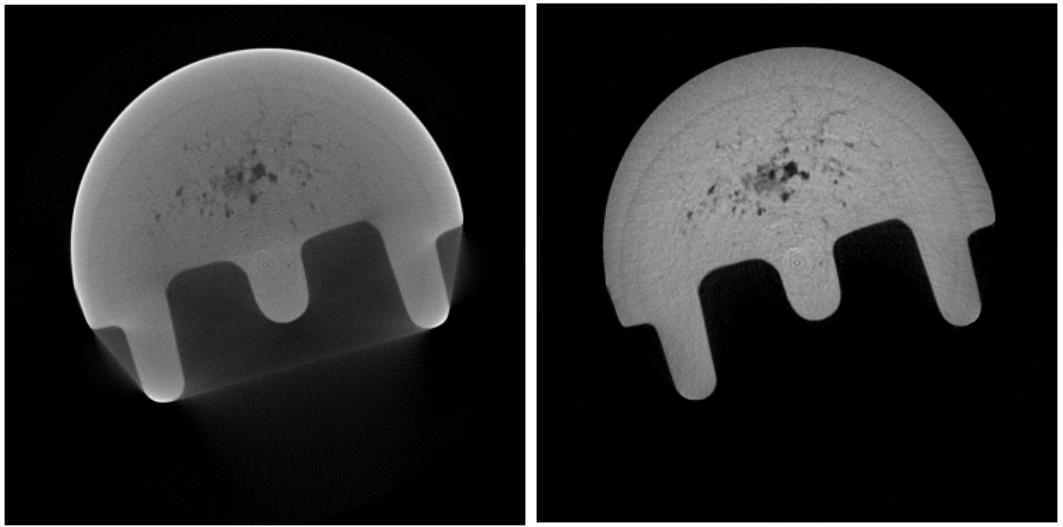
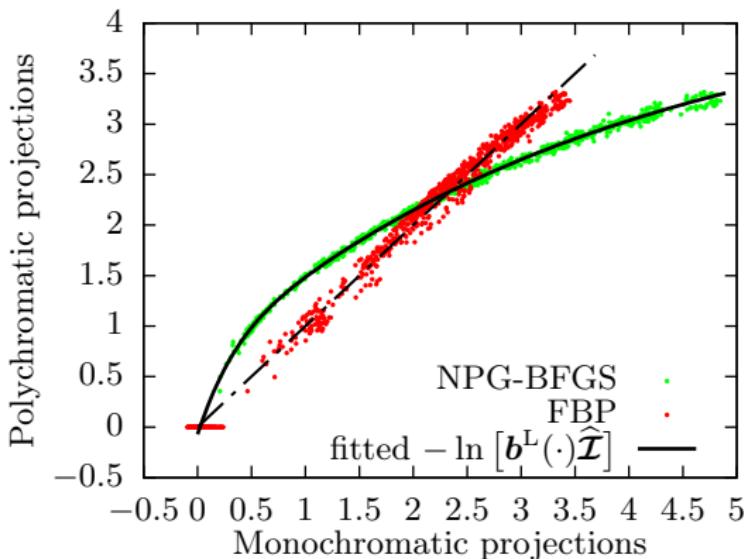


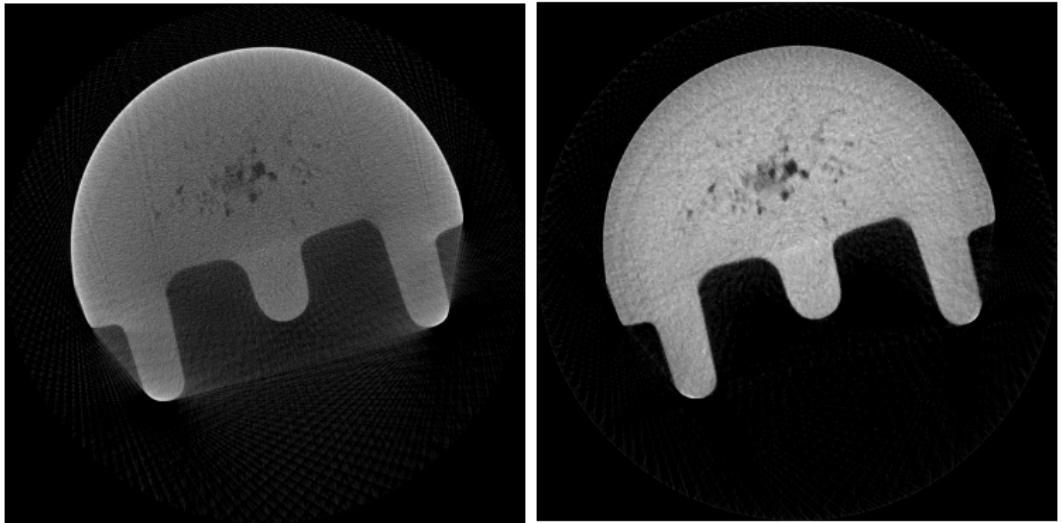
Figure 22: Real X-ray CT: 360 fan-beam projections over  $360^\circ$ .

**Figure 23:** Estimated  $x$  and  $-\ln(b^L(\cdot)\mathcal{I})$  from 360 fan-beam real X-ray CT projections.

## Inverse Linearization Function Estimate



**Figure 24:** The polychromatic measurements as function of the monochromatic projections and its corresponding fitted curve.

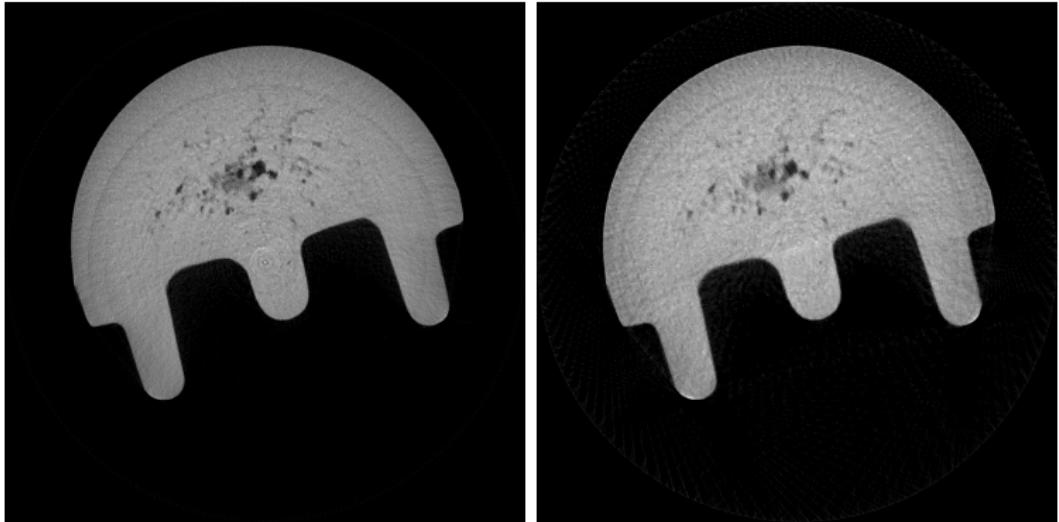


(a) FBP

(b) NPG-BFGS ( $u = 10^{-5}$ )

Figure 26: Real X-ray CT: 120 fan-beam projections over  $360^\circ$ .

Observe the aliasing artifacts in the FBP reconstruction.



(a) 360 projections

(b) 120 projections

**Figure 27:** NPG-BFGS ( $u = 10^{-5}$ ) reconstructions from fan-beam projections over  $360^\circ$ .

## Selected Publications I

-  R. G. and A. D., "Blind X-ray CT image reconstruction from polychromatic Poisson measurements," *IEEE Trans. Comput. Imag.*, 2016, to appear.
-  R. G. and A. D., "Projected Nesterov's proximal-gradient algorithms for sparse signal reconstruction with a convex constraint," , 2016, submitted.

## Selected Publications II

-  R. G. and A. D., "Beam hardening correction via mass attenuation discretization," *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, Vancouver, Canada, May 2013, pp. 1085–1089.
-  R. G. and A. D., "Polychromatic sparse image reconstruction and mass attenuation spectrum estimation via B-spline basis function expansion," *Rev. Prog. Quant. Nondestr. Eval.*, D. E. Chimenti and L. J. Bond, Eds., ser. AIP Conf. Proc. Vol. 34 1650, Melville, NY, 2015, pp. 1707–1716.
-  R. G. and A. D., "Blind polychromatic X-ray CT reconstruction from Poisson measurements," *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, Shanghai, China, 2016, to appear.

# Outline

## Projected Nesterov's Proximal-Gradient Algorithm

Background

Optimization problem

Reconstruction algorithm

Convergence analysis

## Applications

Poisson compressed sensing

Linear model with Gaussian noise

X-Ray CT Reconstruction from Polychromatic Measurements

Polychromatic X-ray CT model

Mass-attenuation spectrum

Minimization algorithm

Numerical Examples

## Conclusion

## Conclusion I

### PNPG algorithm:

- ▶ Developed a fast algorithm for reconstructing signals that are sparse in a transform domain and belong to a closed convex set by employing a projected proximal-gradient scheme with Nesterov's acceleration, restart and *adaptive* step size.
- ▶ Applied the proposed framework to construct the first Nesterov-accelerated Poisson compressed-sensing reconstruction algorithm.
- ▶ Derived convergence-rate upper-bound that accounts for inexactness of the proximal operator.
- ▶ Our PNPG approach is computationally efficient compared with the state-of-the-art.

## Conclusion II

We developed a fast algorithm for reconstructing signals that are sparse in a transform domain and belong to a closed convex set by employing a projected proximal-gradient scheme with Nesterov's acceleration, restart and *adaptive* step size.

Future work will include automating the selection of the regularization parameter  $u$ , extending the adaptive step size and acceleration to the splitting methods.

- ▶ PNPG is the first Nesterov-accelerated proximal algorithm for Poisson compressed sensing
  - ▶ i.e., for Poisson generalized linear model (GLM) with identity link;
- ▶ it is computationally efficient compared with the state-of-the-art SPIRAL method.

## References |

-  A. Auslender and M. Teboulle, "Interior gradient and proximal methods for convex and conic optimization," *SIAM J. Optim.*, vol. 16, no. 3, pp. 697–725, 2006.
-  A. Beck and M. Teboulle, "A fast iterative shrinkage-thresholding algorithm for linear inverse problems," *SIAM J. Imag. Sci.*, vol. 2, no. 1, pp. 183–202, 2009.
-  A. Beck and M. Teboulle, "Fast gradient-based algorithms for constrained total variation image denoising and deblurring problems," *IEEE Trans. Image Process.*, vol. 18, no. 11, pp. 2419–2434, 2009.
-  S. R. Becker, E. J. Candès, and M. C. Grant, "Templates for convex cone problems with applications to sparse signal recovery," *Math. Program. Comp.*, vol. 3, no. 3, pp. 165–218, 2011. [Online]. Available: <http://cvxr.com/tfocs>.
-  M. Bertero, P. Boccacci, G. Desiderà, and G. Vicidomini, "Image deblurring with Poisson data: From cells to galaxies," *Inverse Prob.*, vol. 25, no. 12, pp. 123006-1–123006-26, 2009.
-  R. H. Byrd, P. Lu, J. Nocedal, and C. Zhu, "A limited memory algorithm for bound constrained optimization," *SIAM J. Sci. Comput.*, vol. 16, no. 5, pp. 1190–1208, 1995.

## References II

-  L. Condat, "A primal-dual splitting method for convex optimization involving Lipschitzian, proximable and linear composite terms," *J. Optim. Theory Appl.*, vol. 158, no. 2, pp. 460–479, 2013.
-  A. D., R. G., and K. Qiu, "Mask iterative hard thresholding algorithms for sparse image reconstruction of objects with known contour," *Proc. Asilomar Conf. Signals, Syst. Comput.*, Pacific Grove, CA, Nov. 2011, pp. 2111–2116.
-  J. A. Fessler, *Image reconstruction toolbox*, accessed 30-October-2014. [Online]. Available: <http://www.eecs.umich.edu/~fessler/code>.
-  Z. T. Harmany, R. F. Marcia, and R. M. Willett, "This is SPIRAL-TAP: Sparse Poisson intensity reconstruction algorithms—theory and practice," *IEEE Trans. Image Process.*, vol. 21, no. 3, pp. 1084–1096, Mar. 2012.
-  Y. Nesterov, "A method of solving a convex programming problem with convergence rate  $O(1/k^2)$ ," *Sov. Math. Dokl.*, vol. 27, 1983, pp. 372–376.
-  B. O'Donoghue and E. Candès, "Adaptive restart for accelerated gradient schemes," *Found. Comput. Math.*, vol. 15, no. 3, pp. 715–732, 2013.
-  J. M. Ollinger and J. A. Fessler, "Positron-emission tomography," *IEEE Signal Process. Mag.*, vol. 14, no. 1, pp. 43–55, 1997.

## References III

-  J. L. Prince and J. M. Links, *Medical Imaging Signals and Systems*, 2nd ed. Upper Saddle River, NJ: Pearson, 2015.
-  H. Raguet, J. Fadili, and G. Peyré, "A generalized forward-backward splitting," *SIAM J. Imag. Sci.*, vol. 6, no. 3, pp. 1199–1226, 2013.
-  R. T. Rockafellar, *Convex Analysis*. Princeton, NJ: Princeton Univ. Press, 1970.
-  S. Villa, S. Salzo, L. Baldassarre, and A. Verri, "Accelerated and inexact forward-backward algorithms," *SIAM J. Optim.*, vol. 23, no. 3, pp. 1607–1633, 2013.
-  R. M. Willett, M. F. Duarte, M. A. Davenport, and R. G. Baraniuk, "Sparsity and structure in hyperspectral imaging: Sensing, reconstruction, and target detection," *IEEE Signal Process. Mag.*, vol. 31, no. 1, pp. 116–126, Jan. 2014.
-  S. J. Wright, R. D. Nowak, and M. A. T. Figueiredo, "Sparse reconstruction by separable approximation," *IEEE Trans. Signal Process.*, vol. 57, no. 7, pp. 2479–2493, 2009.
-  C. Zhu, R. H. Byrd, P. Lu, and J. Nocedal, "Algorithm 778: L-BFGS-B: Fortran subroutines for large-scale bound-constrained optimization," *ACM Trans. Math. Softw.*, vol. 23, no. 4, pp. 550–560, Dec. 1997.

## Adaptive Step Size

- ① ► if no step-size backtracking events or increase attempts for  $m$  consecutive iterations, start with a larger step size

$$\bar{\beta}^{(i)} = \frac{\beta^{(i-1)}}{\xi} \quad (\text{increase attempt})$$

where  $\xi \in (0, 1)$  is a *step-size adaptation parameter*;

- otherwise start with

$$\bar{\beta}^{(i)} = \beta^{(i-1)};$$

- ② (backtracking search) select

$$\beta^{(i)} = \xi^{t_i} \bar{\beta}^{(i)} \quad (5)$$

where  $t_i \geq 0$  is the smallest integer such that (5) satisfies the majorization condition (3); *backtracking event* corresponds to  $t_i > 0$ .

- ③ if  $\max(\beta^{(i)}, \beta^{(i-1)}) < \bar{\beta}^{(i)}$ , increase  $m$  by a nonnegative integer  $m$ :

$$m \leftarrow m + m.$$

◀ back

## Restart

Whenever  $f(x^{(i)}) > f(x^{(i-1)})$  or  $\bar{x}^{(i)} \in C \setminus \text{dom } \mathcal{L}$ , we set

$$\theta^{(i-1)} = 1 \quad (\text{restart})$$

and refer to this action as *function restart* (O'Donoghue and Candès 2013) or *domain restart* respectively.

◀ back

## Inner Convergence Criteria

$$\text{TV: } \|\mathbf{x}^{(i,j)} - \mathbf{x}^{(i,j)}\|_2 \leq \eta \delta^{(i-1)} \quad (6a)$$

$$\begin{aligned} \ell_1: \quad & \max \left( \|\mathbf{s}^{(i,j)} - \Psi^T \mathbf{x}^{(i,j)}\|_2, \|\mathbf{s}^{(i,j)} - \mathbf{s}^{(i,j-1)}\|_2 \right) \\ & \leq \eta \left\| \Psi^T (\mathbf{x}^{(i-1)} - \mathbf{x}^{(i-2)}) \right\|_2 \end{aligned} \quad (6b)$$

where  $j$  is the inner-iteration index,

- ▶  $\mathbf{x}^{(i,j)}$  is the iterate of  $\mathbf{x}$  in the  $j$ th inner iteration step within the  $i$ th step of the (outer) PNPG iteration, and
- ▶

$$\eta \in (0, 1)$$

is the convergence tuning constant chosen to trade off the accuracy and speed of the inner iterations and provide sufficiently accurate solutions to the proximal mapping.

◀ back