



Projected Nesterov's Proximal-Gradient Algorithm for Sparse Signal Recovery[†]

Aleksandar Dogandžić

Electrical and Computer Engineering
IOWA STATE UNIVERSITY

[†] joint work with Renliang Gu, Ph.D. student

supported by



Terminology and Notation

- 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 λ :

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

- ε -subgradient (Rockafellar 1970, Sec. 23):

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



Introduction I

For most natural signals x ,

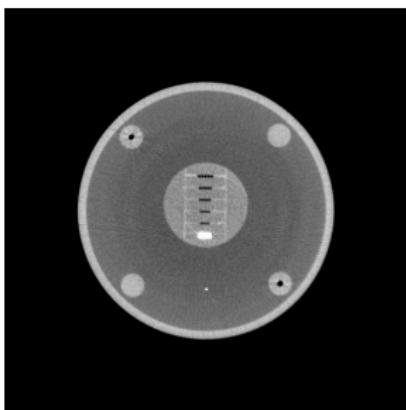
$$\# \text{ significant coefficients of } \psi(x) \ll \text{signal size } p$$

where

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

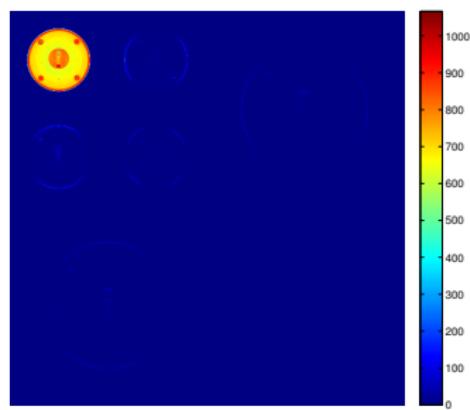
is *sparsifying transform*.

Sparsifying Transforms I



p pixels

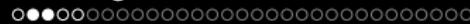
linear
transform
 \leftrightarrow
DWT



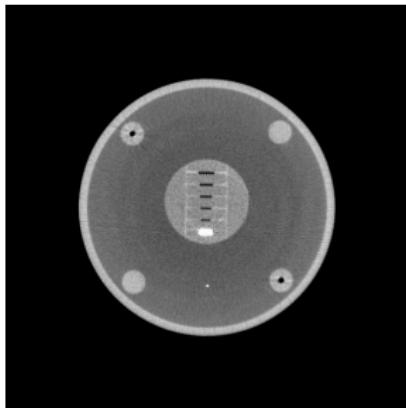
significant coeffs << *p*

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

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

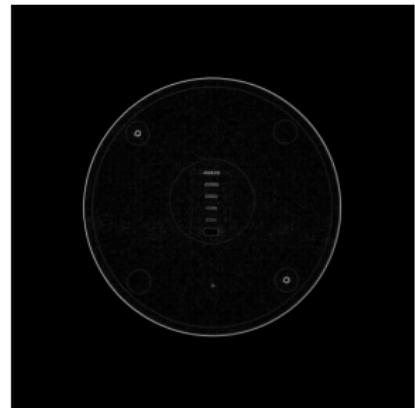


Sparsifying Transforms II



p pixels

transform
 \leftrightarrow
gradient map



significant coeffs $\ll p$

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

\mathcal{N}_i is the index set of neighbors of x_i .



Convex-Set Constraint

$$\boldsymbol{x} \in C$$

where C is a nonempty 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.



Convex-Set Constraint

$$x \in C$$

where C is a nonempty 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.

Goal

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

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

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

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

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

Example: Linear model

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

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

Penalized NLL

- objective function

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

- convex differentiable negative log-likelihood (NLL)
- convex penalty term
 $u > 0$ is a scalar tuning constant
 $C \subseteq \text{cl}(\text{dom } \mathcal{L}(x))$

Penalized NLL

- objective function

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

- convex differentiable negative log-likelihood (NLL)
- convex penalty term
 $u > 0$ is a scalar tuning constant
 $C \subseteq \text{cl}(\text{dom } \mathcal{L}(x))$

Penalized NLL

- objective function

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

- convex differentiable negative log-likelihood (NLL)
- convex penalty term
 $u > 0$ is a scalar tuning constant
 $C \subseteq \text{cl}(\text{dom } \mathcal{L}(x))$



Comment

Our objective function $f(x)$ is

- convex
 - has a convex set as minimum (unique if $\mathcal{L}(x)$ is strongly convex),
- not differentiable with respect to the signal x
 - cannot apply usual gradient- or Newton-type algorithms,
 - need proximal-gradient (PG) schemes.



Goals

Develop a fast algorithm with

- $\mathcal{O}(k^{-2})$ convergence-rate and
- iterate convergence guarantees

for minimizing $f(x)$ that

- is general (for a diverse set of NLLs),
- requires minimal tuning, and
- is matrix-free*, a must for solving large-scale problems.

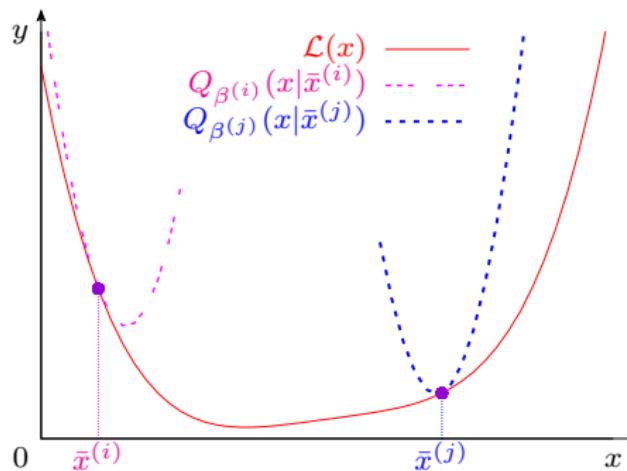
*involves only matrix-vector multiplications implementable using, e.g.,
function handle in Matlab

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



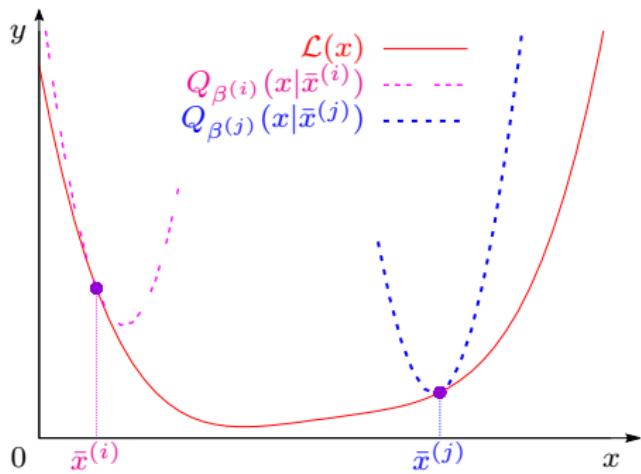


Figure 1: Majorizing function: Impact of β .

No need for strict majorization, sufficient to majorize in the neighborhood of \bar{x} where we wish to move next!

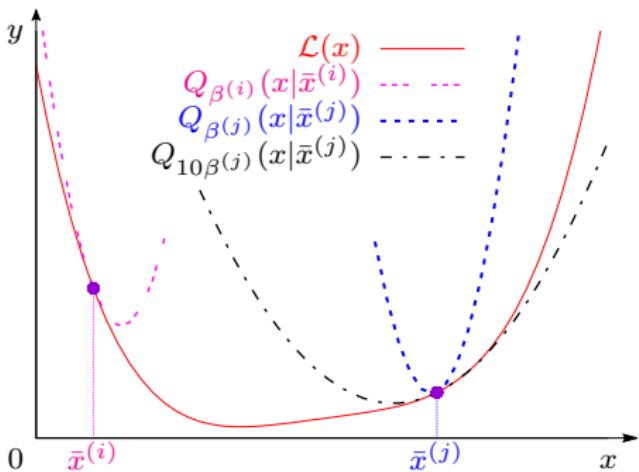


Figure 1: Majorizing function: Impact of β .

No need for strict majorization, sufficient to majorize in the neighborhood of \bar{x} where we wish to move next!

PNPG Method: Iteration i

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

$$\theta^{(i)} = \begin{cases} 1, & i \leq 1 \\ \frac{1}{\gamma} + \sqrt{b + B^{(i)}(\theta^{(i-1)})^2}, & i > 1 \end{cases}$$

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

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

where $\beta^{(i)} > 0$ is an *adaptive step size*:

- satisfies

$$\mathcal{L}(x^{(i)}) \leq Q_{\beta^{(i)}}(x^{(i)} | \bar{x}^{(i)}) \quad \text{majorization condition,}$$

- is as large as possible.

▶ more

PNPG Method: Iteration i

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

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

$$\theta^{(i)} = \begin{cases} i, & i \leq 1 \\ \frac{1}{\gamma} + \sqrt{b + B^{(i)}(\theta^{(i-1)})^2}, & i > 1 \end{cases}$$

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

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

where $\beta^{(i)} > 0$ is an *adaptive step size*:

- satisfies

$$\mathcal{L}(x^{(i)}) \leq Q_{\beta^{(i)}}(x^{(i)} | \bar{x}^{(i)}) \quad \text{majorization condition,}$$

- is as large as possible.

PNPG Method: Iteration i

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

$$\theta^{(i)} = \begin{cases} i, & i \leq 1 \\ \frac{1}{\gamma} + \sqrt{b + B^{(i)}(\theta^{(i-1)})^2}, & i > 1 \end{cases}$$

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

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

where $\beta^{(i)} > 0$ is an *adaptive step size*:

- satisfies

$$\mathcal{L}(x^{(i)}) \leq Q_{\beta^{(i)}}(x^{(i)} | \bar{x}^{(i)}) \quad \text{majorization condition,}$$

needs to hold for $x^{(i)}$, not for all x !

$\mathcal{L}(x) \not\leq Q_{\beta^{(i)}}(x | \bar{x}^{(i)})$ in general, for an arbitrary x !

* more

PNPG Method: Iteration i

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

$$\theta^{(i)} = \begin{cases} i, & i \leq 1 \\ \frac{1}{\gamma} + \sqrt{b + B^{(i)}(\theta^{(i-1)})^2}, & i > 1 \end{cases}$$

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

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

where $\beta^{(i)} > 0$ is an *adaptive step size*:

- satisfies

$$\mathcal{L}(x^{(i)}) \leq Q_{\beta^{(i)}}(x^{(i)} | \bar{x}^{(i)}) \quad \text{majorization condition,}$$

needs to hold for $x^{(i)}$, not for all x !

$\mathcal{L}(x) \not\leq Q_{\beta^{(i)}}(x | \bar{x}^{(i)})$ in general, for an arbitrary x !

▶ more

Algorithm 1: PNPG method

Input: $x^{(-1)}$, u , m , ξ , η , and threshold ϵ

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

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

repeat

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

while *true* **do**

evaluate $B^{(i)}$, $\theta^{(i)}$, and $\bar{x}^{(i)}$

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

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

solve the PG step

if majorization condition holds **then**

| break

else

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

$m \leftarrow m + m$

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

// backtracking search

// domain restart

// increase m

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

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

// function restart

if convergence condition holds **then**

| declare convergence

if $\kappa \geq m$ **then**

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

// adapt step size

else

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

until convergence declared or maximum number of iterations exceeded

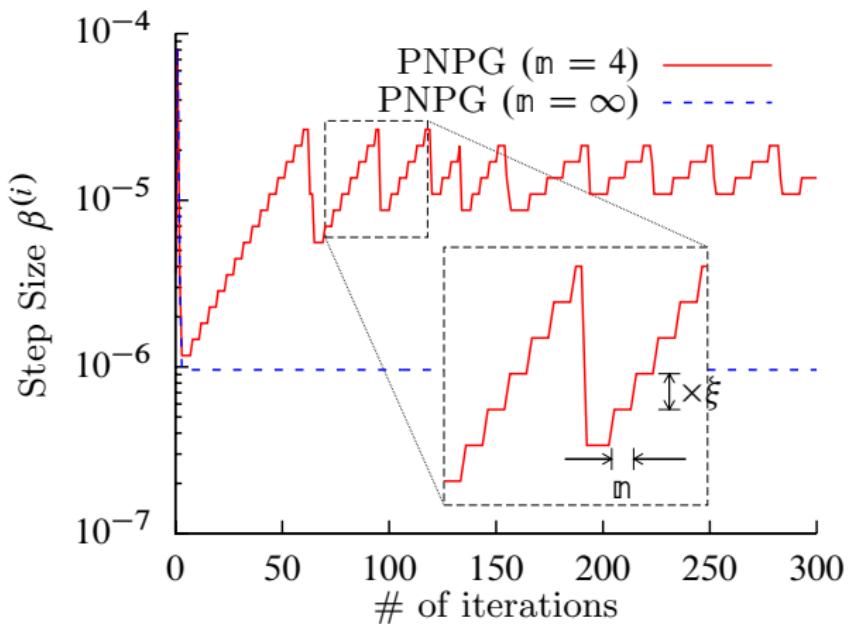
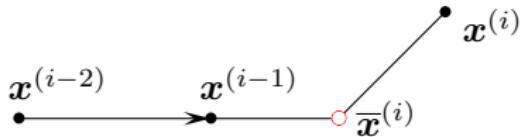


Figure 2: Illustration of step-size selection for Poisson generalized linear model (GLM) with identity link.

Momentum Illustration

??? add a better figure depicting momentum

$$\bar{x}^{(i)} = P_C \left(x^{(i-1)} + \underbrace{\frac{\theta^{(i-1)} - 1}{\theta^{(i)}} (x^{(i-1)} - x^{(i-2)})}_{\text{momentum term prevents zigzagging}} \right)$$





Comments on the extrapolation term $\theta^{(i)}$ |

$$\theta^{(i)} = \frac{1}{\gamma} + \sqrt{b + B^{(i)}(\theta^{(i-1)})^2}, \quad i \geq 2$$

where

$$\gamma \geq 2, \quad b \in [0, 1/4]$$

are momentum tuning constants.

- To establish $\mathcal{O}(k^{-2})$ convergence of PNPG, need

$$\theta^{(i)} \leq \frac{1}{2} + \sqrt{\frac{1}{4} + B^{(i)}(\theta^{(i-1)})^2}, \quad i \geq 2.$$

- γ controls the rate of increase of $\theta^{(i)}$.

Comments on the extrapolation term $\theta^{(i)}$ ||

$\theta^{(i)} \uparrow$ implies stronger momentum.

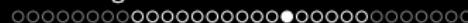
Effect of step size

In the “steady state” where $\beta^{(i-1)} = \beta^{(i)}$, $\theta^{(i)} \uparrow$ approximately linearly with i , with slope $1/\gamma$.

Changes in the step size affect $\theta^{(i)}$:

$\beta^{(i)} < \beta^{(i-1)}$ step size decrease, faster increase of $\theta^{(i)}$,

$\beta^{(i)} > \beta^{(i-1)}$ step size increase, decrease or slower increase of $\theta^{(i)}$ than in the steady state.



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 ℓ_1 -norm penalty with $\psi(\mathbf{x}) = \Psi^T \mathbf{x}$, alternating direction method of multipliers (ADMM)
 - iterative
- 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).
 - iterative

Inexact PG Steps

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

$$\theta^{(i)} = \begin{cases} i, & i \leq 1 \\ \frac{1}{\gamma} + \sqrt{b + B^{(i)}(\theta^{(i-1)})^2}, & i > 1 \end{cases}$$

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

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

Because of their iterative nature, PG steps are *inexact*: $\varepsilon^{(i)}$ quantifies the precision of the PG step in Iteration i .

▶ more



Remark (Monotonicity)

The projected Nesterov's proximal-gradient (PNPG) iteration with restart is non-increasing:

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

if the inexact PG steps are sufficiently accurate and satisfy

$$\varepsilon^{(i)} \leq \sqrt{\delta^{(i)}}$$

where

$$\delta^{(i)} \triangleq \|\mathbf{x}^{(i)} - \mathbf{x}^{(i-1)}\|_2^2$$

is the local variation of signal iterates.

Convergence Criterion

$$\sqrt{\delta^{(i)}} < \epsilon \|x^{(i)}\|_2$$

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

▶ more

Restart

The goal of *function* and *domain* restarts is to ensure that

- the PNPG iteration is *monotonic* and
- $\bar{x}^{(i)}$ and $x^{(i)}$ remain *within* $\text{dom } f$.

▶ more



Summary of PNPG Approach

Combine

- convex-set projection with
- Nesterov acceleration.

Apply

- adaptive step size,
- restart.

▶ more



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 .

Extends the application of the Nesterov's acceleration[†] to more general measurement models than those used previously.

[†]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.



Theorem (Convergence of the Objective Function)

Assume

- NLL $\mathcal{L}(x)$ is convex and differentiable and $r(x)$ is convex,
- $C \subseteq \text{dom } \mathcal{L}$: no need for domain restart.

Consider the PNPG iteration without restart.



Theorem (Convergence of the Objective Function)

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

where

$$\mathcal{E}^{(k)} \triangleq \sum_{i=1}^k (\theta^{(i)} \varepsilon^{(i)})^2 \quad \text{error term, accounts for inexact PG steps}$$

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



Comments

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

- Step sizes $\beta^{(i)} \uparrow$, convergence-rate upper bound \downarrow .
- better initialization, convergence-rate upper bound \downarrow .
- smaller prox-step approx. error, convergence-rate bound \downarrow .

Corollary

Under the condition of the Theorem,

$$f(\mathbf{x}^{(k)}) - f(\mathbf{x}^*) \leq \underbrace{\gamma^2 \frac{\|\mathbf{x}^{(0)} - \mathbf{x}^*\|_2^2 + \mathcal{E}^{(k)}}{2k^2\beta_{\min}}}_{\mathcal{O}(k^{-2}) \text{ if } \mathcal{E}^{(+\infty)} < +\infty}$$

provided that

$$\beta_{\min} \triangleq \min_{k=1}^{+\infty} \beta^{(k)} > 0.$$

The assumption that the step-size sequence is lower-bounded by a strictly positive quantity is weaker than Lipschitz continuity of $\nabla \mathcal{L}(\mathbf{x})$ because it is guaranteed to have $\beta_{\min} > \xi/L$ if $\nabla \mathcal{L}(\mathbf{x})$ has a Lipschitz constant L .

Corollary

Under the condition of the Theorem,

$$f(\mathbf{x}^{(k)}) - f(\mathbf{x}^*) \leq \underbrace{\gamma^2 \frac{\|\mathbf{x}^{(0)} - \mathbf{x}^*\|_2^2 + \mathcal{E}^{(k)}}{2k^2\beta_{\min}}}_{\mathcal{O}(k^{-2}) \text{ if } \mathcal{E}^{(+\infty)} < +\infty}$$

provided that

$$\beta_{\min} \triangleq \min_{k=1}^{+\infty} \beta^{(k)} > 0.$$

The assumption that the step-size sequence is lower-bounded by a strictly positive quantity is weaker than Lipschitz continuity of $\nabla \mathcal{L}(\mathbf{x})$ because it is guaranteed to have $\beta_{\min} > \xi/L$ if $\nabla \mathcal{L}(\mathbf{x})$ has a Lipschitz constant L .



Theorem (Convergence of Iterates)

Assume that

- ① NLL $\mathcal{L}(\mathbf{x})$ is convex and differentiable and $r(\mathbf{x})$ is convex,
- ② $C \subseteq \text{dom } \mathcal{L}$, hence no need for domain restart,
- ③ cumulative error term $\mathcal{E}^{(k)}$ converges: $\mathcal{E}^{(+\infty)} < +\infty$,
- ④ momentum tuning constants satisfy $\gamma > 2$ and $b \in [0, 1/\gamma^2]$,
- ⑤ the step-size sequence $(\beta^{(i)})_{i=1}^{+\infty}$ is bounded within the range $[\beta_{\min}, \beta_{\max}]$, ($\beta_{\min} > 0$).



The sequence of PNPG iterates $\mathbf{x}^{(i)}$ without restart converges weakly to a minimizer of $f(\mathbf{x})$. a minimizer of $f(\mathbf{x})$.



Theorem (Convergence of Iterates)

strict inequality

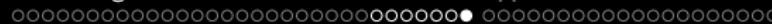
Assume that

- ① NLL $\mathcal{L}(x)$ is convex and differentiable and $r(x)$ is convex,
- ② $C \subseteq \text{dom } \mathcal{L}$, hence no need for domain restart,
- ③ cumulative error term $\mathcal{E}^{(k)}$ converges: $\mathcal{E}^{(+\infty)} < +\infty$,
- ④ momentum tuning const. satisfy $\gamma > 2$ and $b \in [0, 1/\gamma^2]$,
- ⑤ the step-size sequence $(\beta^{(i)})_{i=1}^{+\infty}$ is bounded within the range $[\beta_{\min}, \beta_{\max}]$, ($\beta_{\min} > 0$).



The sequence of PNPG iterates $x^{(i)}$ without restart converges weakly to a minimizer of $f(x)$. a minimizer of $f(x)$.

narrower than $[0, 1/4]$



Idea of Proof.

Recall the inequality

$$\theta^{(i)} = \frac{1}{\gamma} + \sqrt{b + B^{(i)}(\theta^{(i-1)})^2} \leq \frac{1}{2} + \sqrt{\frac{1}{4} + B^{(i)}(\theta^{(i-1)})^2}$$

used to establish convergence of the objective function.

??? Assumption 4:

$$\gamma > 2, \quad b \in [0, 1/\gamma^2]$$

creates a sufficient “gap” in this inequality that allows us to

- show faster convergence of the objective function than the previous theorem and
- establish the convergence of iterates.



Inspired by (Chambolle and Dossal 2015).



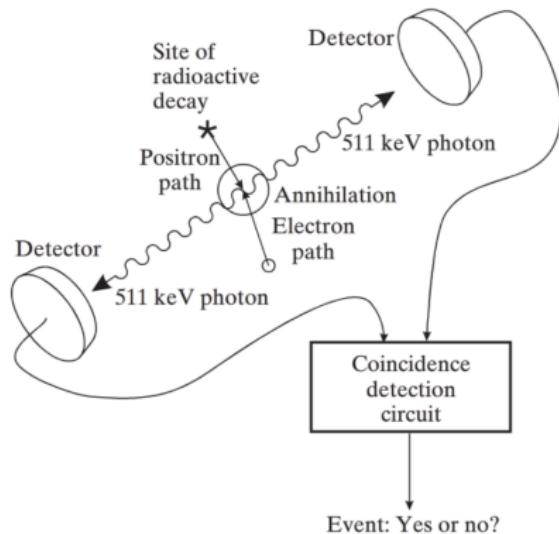


Introduction

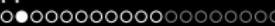
Signal reconstruction from Poisson-distributed measurements with affine model for the mean-signal intensity is important for

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

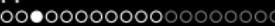
$$[\Phi \mathbf{x} + \mathbf{b}]_n$$

where

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

are the *known sensing matrix* and *intercept term*[§].

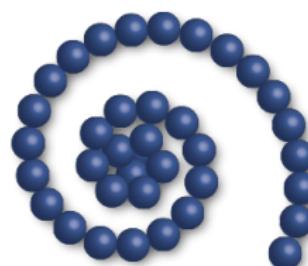
[§]the intercept \mathbf{b} models background radiation and scattering, obtained, e.g., by calibration before the measurements \mathbf{y} have been collected



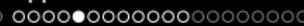
Existing Work

The sparse Poisson-intensity reconstruction algorithm (SPIRAL)[¶]

- 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.

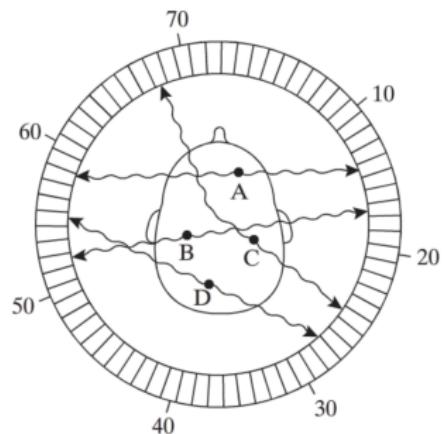


[¶]Z. T. Harmany *et al.*, “This is SPIRAL-TAP: Sparse Poisson intensity reconstruction algorithms—theory and practice,” *IEEE Trans. Image Process.*, vol. 21, no. 3, pp. 1084–1096, Mar. 2012.

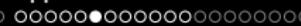


PET Image Reconstruction

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



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_{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 $f(x)$ with nonnegative x :

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

All iterative methods initialized by FBP reconstructions.

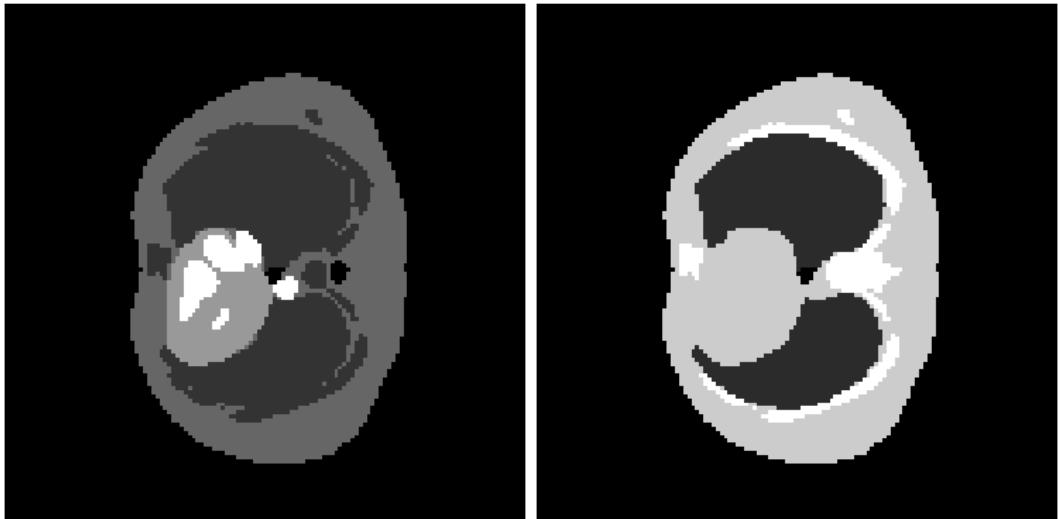
Matlab implementation available at

<http://isucsp.github.io/imgRecSrc/npg.html>.



PG Methods

- PNPG with $(\gamma, b) = (2, 1/4)$.
- 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 3: (a) True emission image and (b) density map.

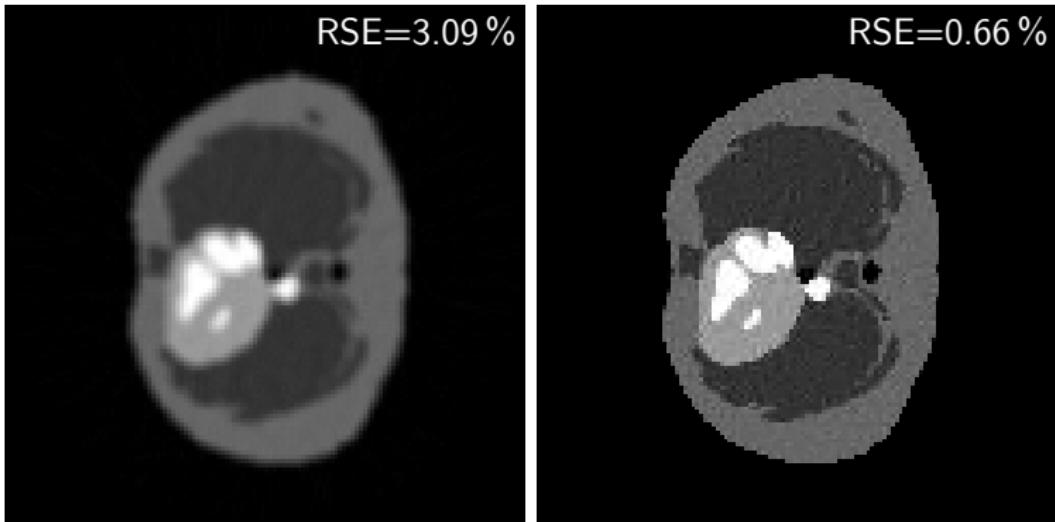


Figure 4: Reconstructions of the emission concentration map for expected total annihilation photon count (SNR) equal to 10^8 .

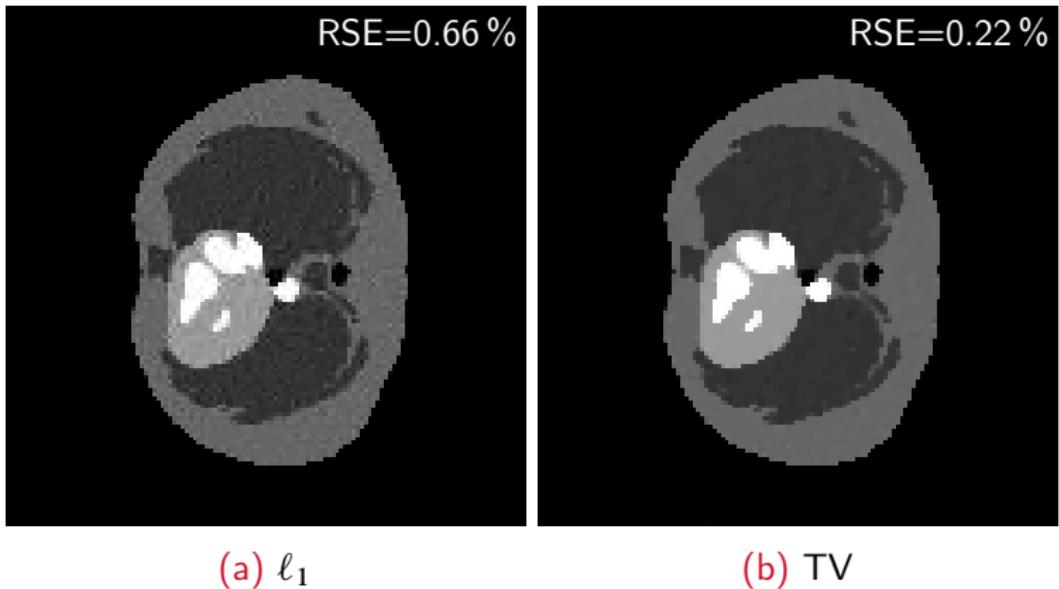


Figure 5: Comparison of the two sparsity regularizations.

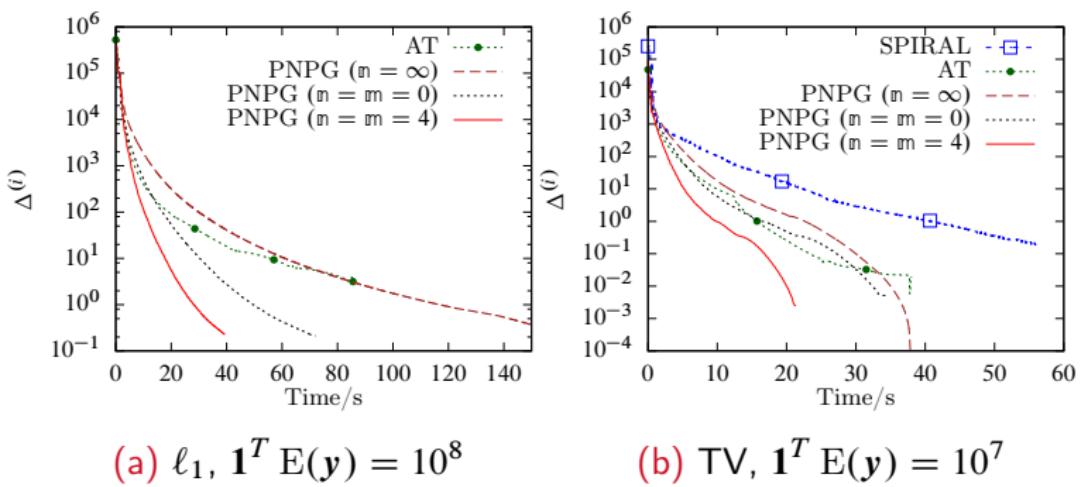


Figure 6: Centered objectives as functions of CPU time.

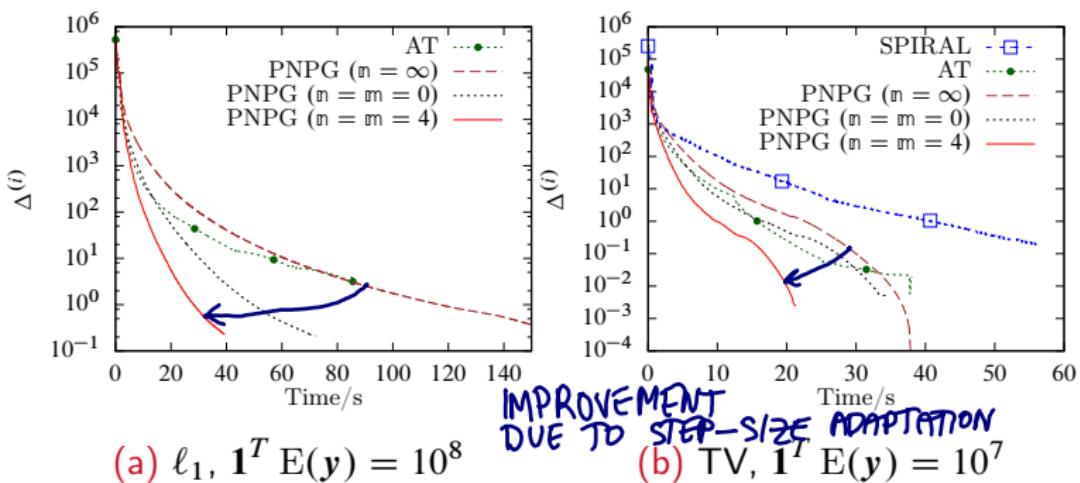
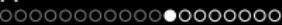


Figure 7: Centered objectives as functions of CPU time.



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 and
- the elements of the sensing matrix Φ are independent, identically distributed (i.i.d.), drawn from the standard normal distribution.

We select the ℓ_1 -norm sparsifying signal penalty with linear map:

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

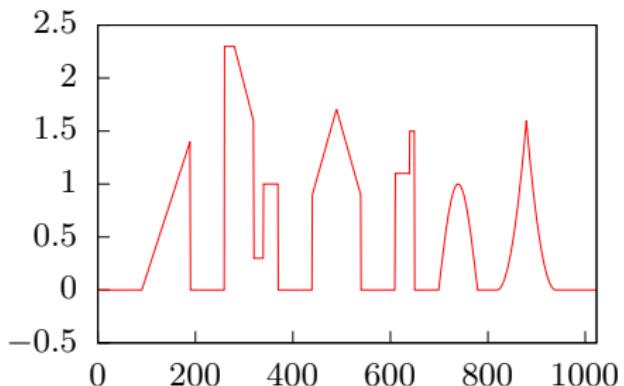


Figure 7: True signal.

Comments I

- More methods available for comparison:
 - 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).
- 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 $C = \mathbb{R}^p$, PNPG iteration reduces to the Nesterov's proximal gradient iteration with adaptive step size that imposes signal sparsity *only* in the analysis form (termed NPG_S).

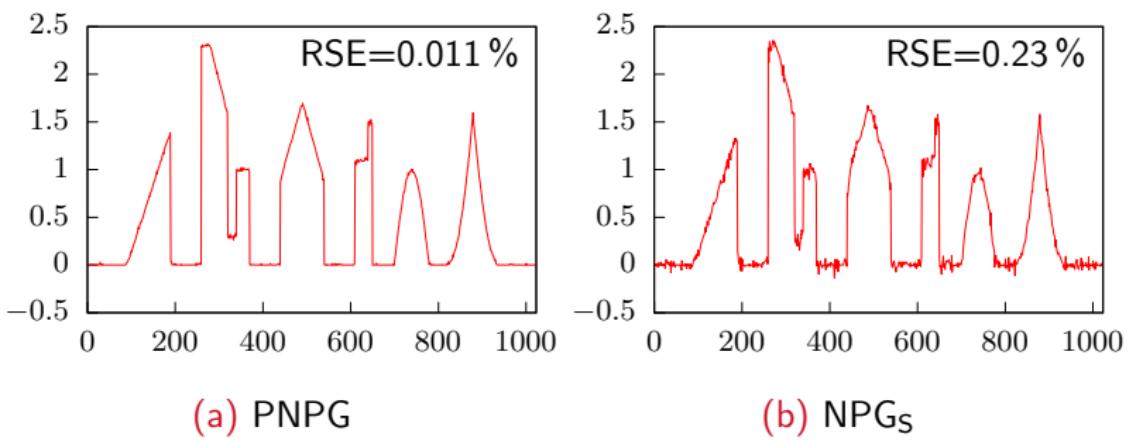


Figure 8: PNPG and NPG_S reconstructions for $N/p = 0.34$.

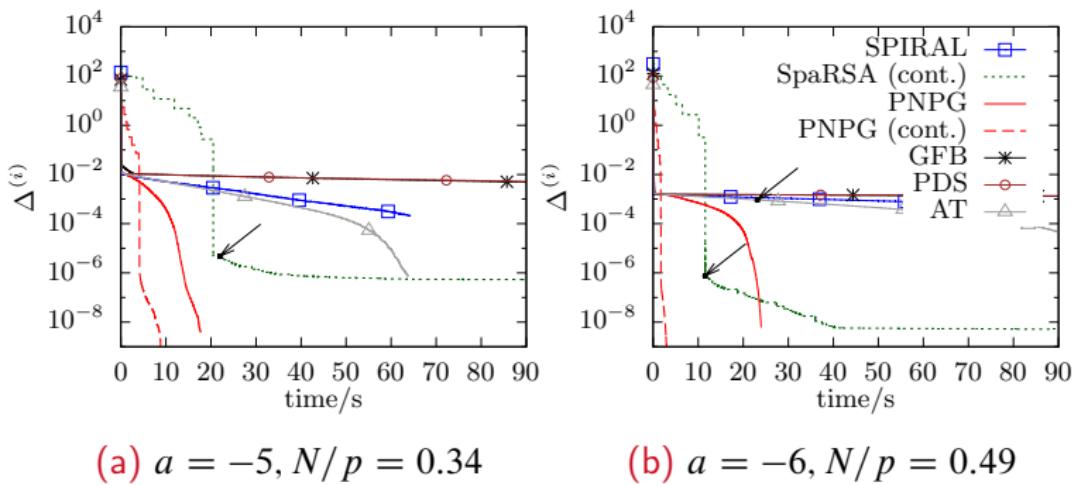


Figure 9: Centered objectives as functions of CPU time.

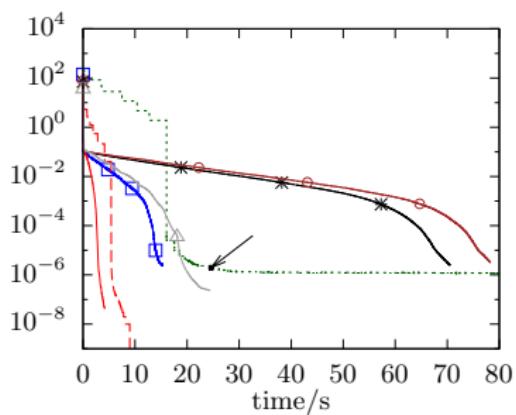
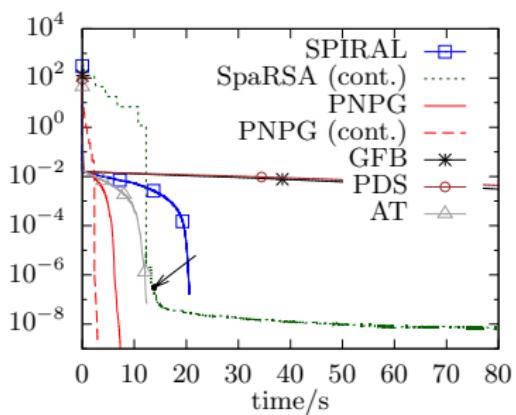
(a) $a = -4, N/p = 0.34$ (b) $a = -5, N/p = 0.49$

Figure 10: Centered objectives as functions of CPU time.

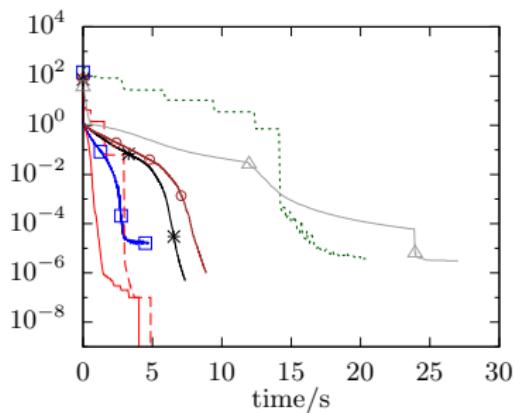
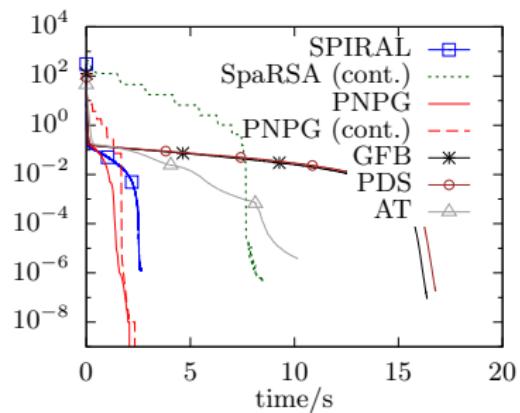
(a) $a = -3, N/p = 0.34$ (b) $a = -4, N/p = 0.49$

Figure 11: Centered objectives as functions of CPU time.

Publications

-  R. G. and A. D., "Projected Nesterov's proximal-gradient algorithms for sparse signal reconstruction with a convex constraint," , 2016, *in preparation*.

-  R. G. and A. D., "Projected Nesterov's proximal-gradient signal recovery from compressive Poisson measurements," *Proc. Asilomar Conf. Signals, Syst. Comput.*, Pacific Grove, CA, Nov. 2015, pp. 1490–1495.

More Terminology and Notation

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

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

- Laplace transform with vector argument:

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

X-ray CT

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

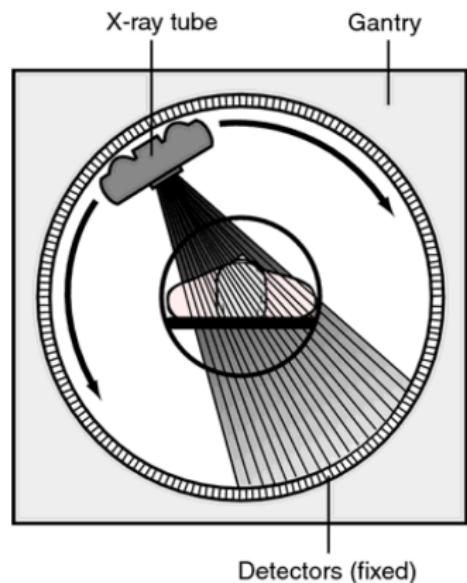


Figure 12: Fan-beam CT system.

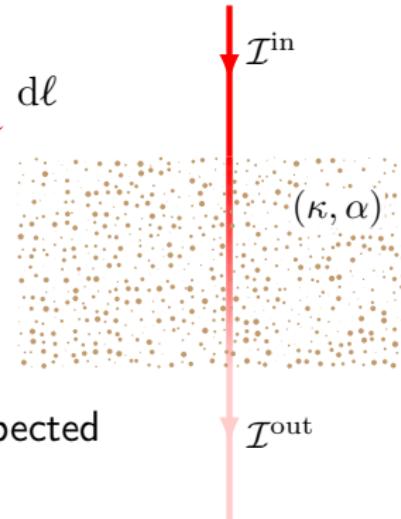
Exponential Law of Absorption

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

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

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
 - ε is **photon energy**.



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

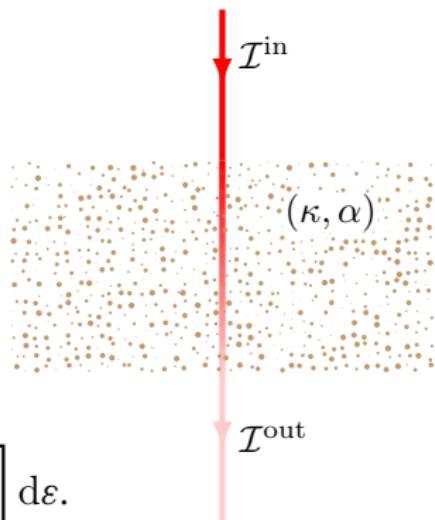
Polychromatic X-ray CT Model

- Incident energy \mathcal{I}^{in} spreads along photon energy ε 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

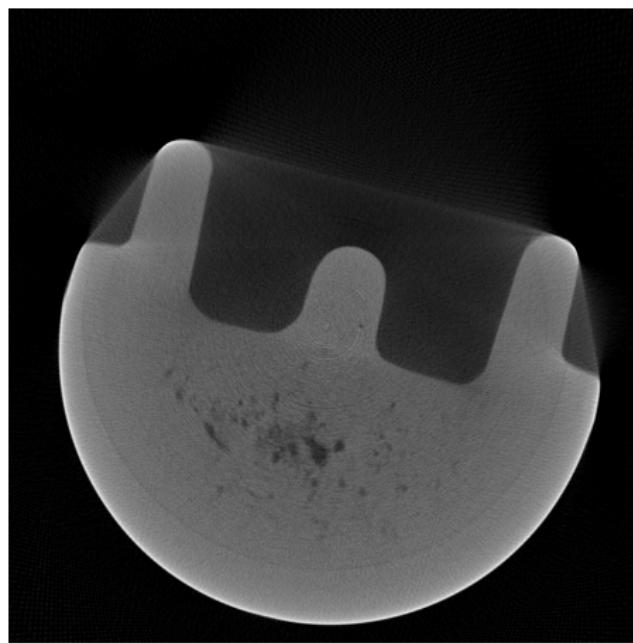
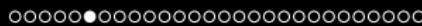


Figure 13: FBP reconstruction of an industrial object.

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



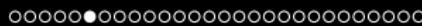
Problem Formulation and Goal

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**.

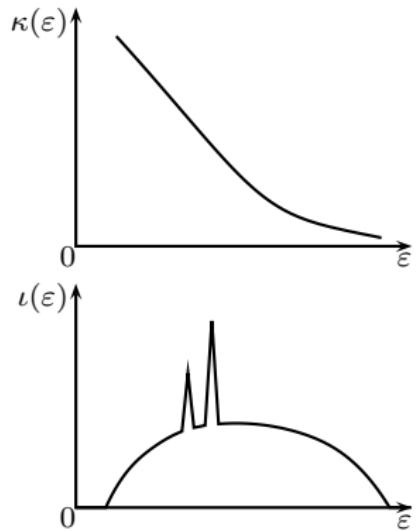
Goal: Estimate the density map $\alpha(x, y)$.

Polychromatic X-ray CT Model Using Mass-Attenuation Spectrum

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

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

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



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

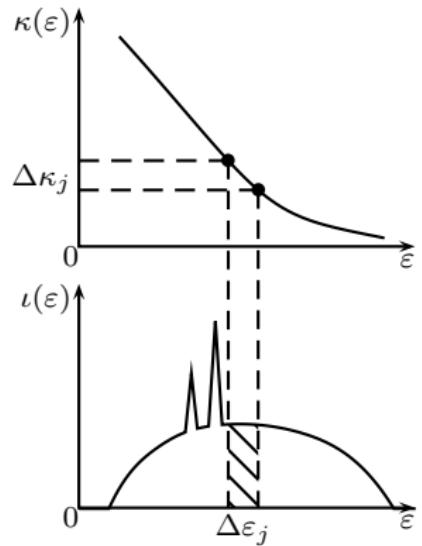


Polychromatic X-ray CT Model Using Mass-Attenuation Spectrum

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

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

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



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

Mass-Attenuation Spectrum

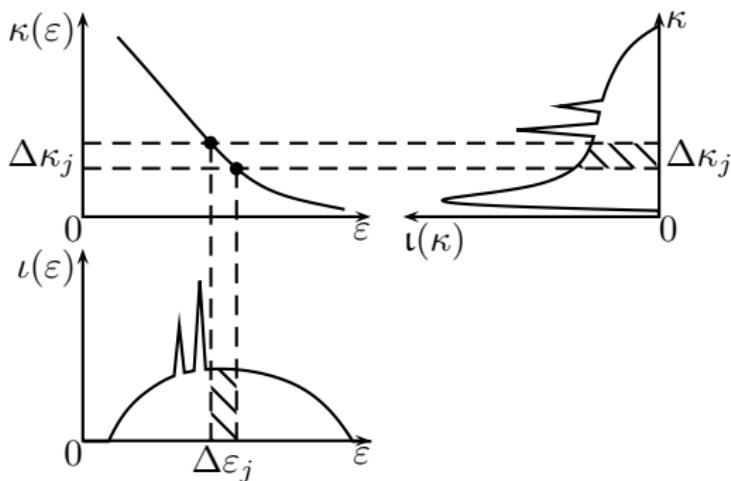


Figure 14: Relationship between mass attenuation κ , incident spectrum ι , photon energy ε , and **mass attenuation spectrum** $\iota(\kappa)$.

Basis-function expansion of mass-attenuation spectrum

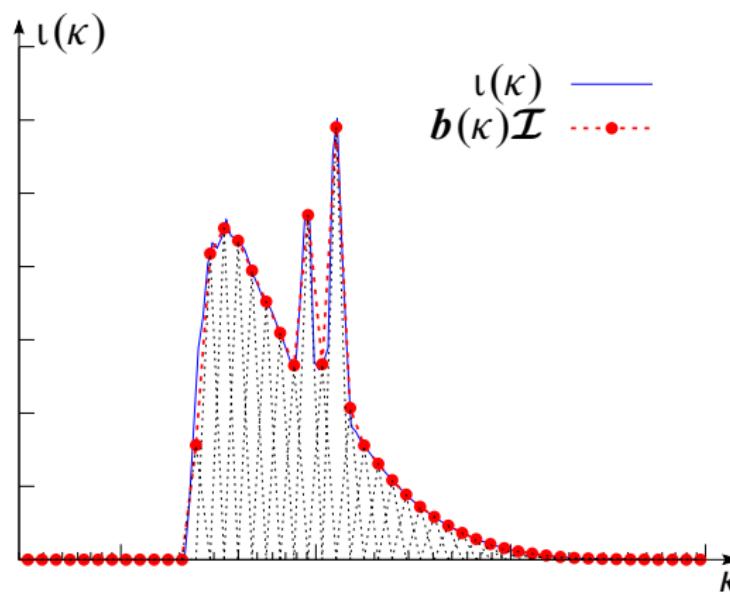
$$\iota(\kappa) = \mathbf{b}(\kappa)\mathcal{I}$$


Figure 15: 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 Φ 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,

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

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

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

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

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:

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

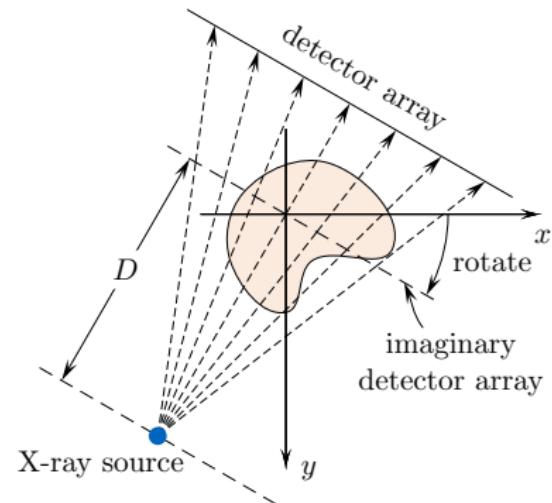
- B1-spline constants set to satisfy

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

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

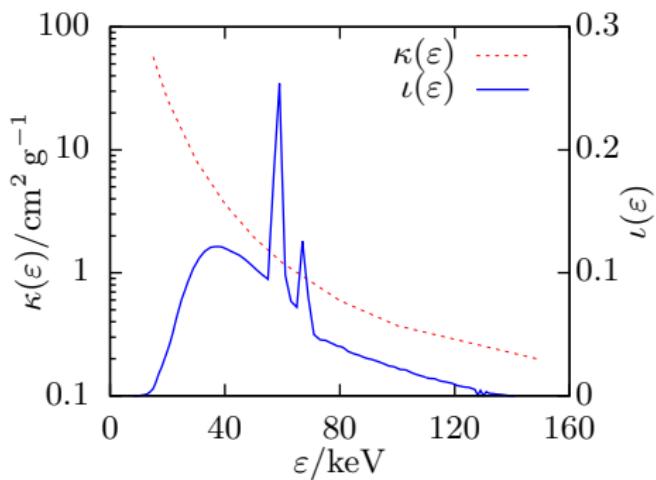
Simulated X-ray CT Example

- Equi-spaced fan-beam projections over 360° ,
- X-ray source to rotation center is $2000 \times$ detector size,
- measurement array size of 512 elements, and
- image to reconstruct has size 512×512 .
- 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.

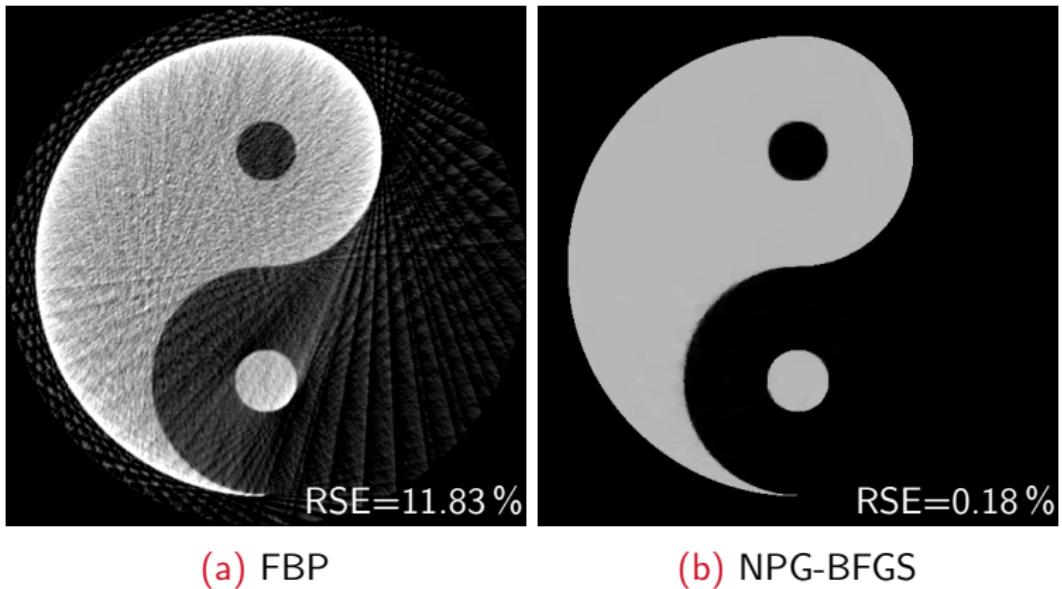


Figure 16: Reconstructions from 60 projections.

Simulated X-ray CT Example

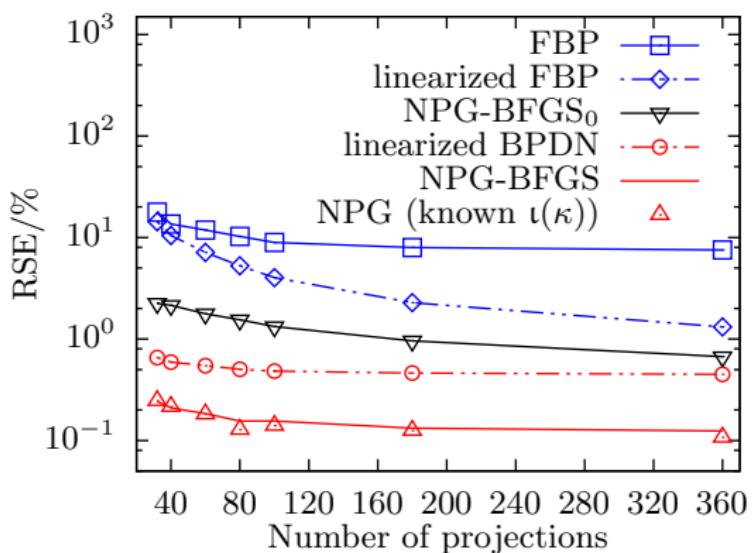


Figure 17: Average RSEs as functions of the number of projections.

Real X-ray CT Example I

- 360 equi-spaced fan-beam projections with 1° spacing,
- X-ray source to rotation center is $3492 \times$ detector size,
- measurement array size of 694 elements,
- projection matrix Φ 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 512×512 image to reconstruct.

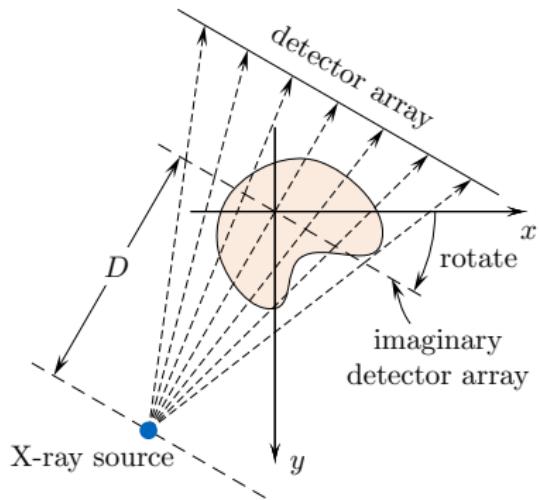
Real X-ray CT Example I

- 360 equi-spaced fan-beam projections with 1° spacing,
- X-ray source to rotation center is $3492 \times$ detector size,
- measurement array size of 694 elements,
- projection matrix Φ 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 512×512 image to reconstruct.

Implementation available at github.com/isucsp/imgRecSrc.

Real data provided by Joe Gray, CNDE. Thanks!



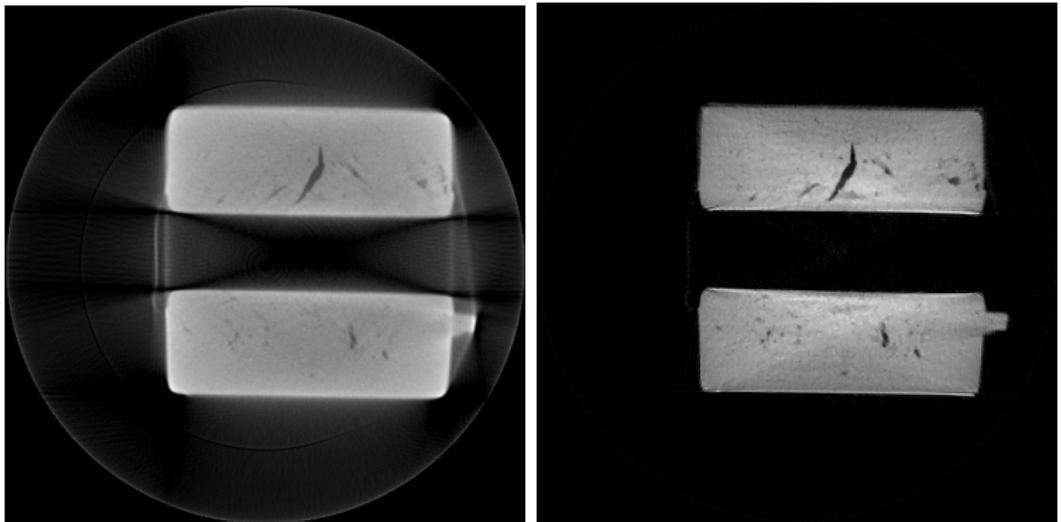


Figure 18: Real X-ray CT: Full projections.

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 μ is tuned for the best reconstruction.

Inverse Linearization Function Estimate

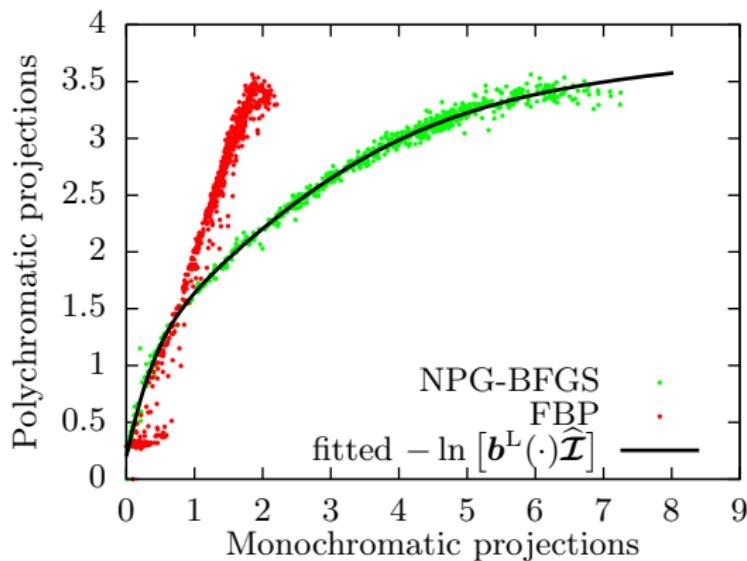
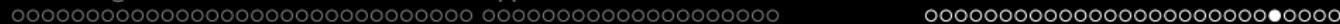


Figure 19: The polychromatic measurements as function of the monochromatic projections and its corresponding fitted curve.

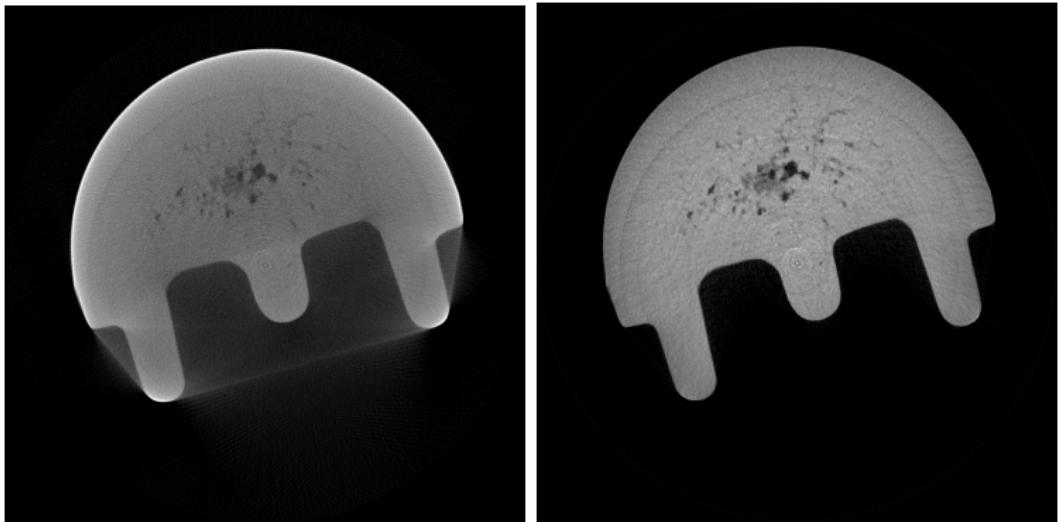
residuals: large, biased for FBP; small, unbiased for NPG-BFGS, increasing variance



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 Φ 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 1024×1024 image to reconstruct.



(a) FBP

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

Figure 20: Real X-ray CT: 360 fan-beam projections over 360°.

Figure 21: Estimated x and $-\ln(b^L(\cdot)\mathcal{I})$ from 360 fan-beam real X-ray CT projections.

Inverse Linearization Function Estimate

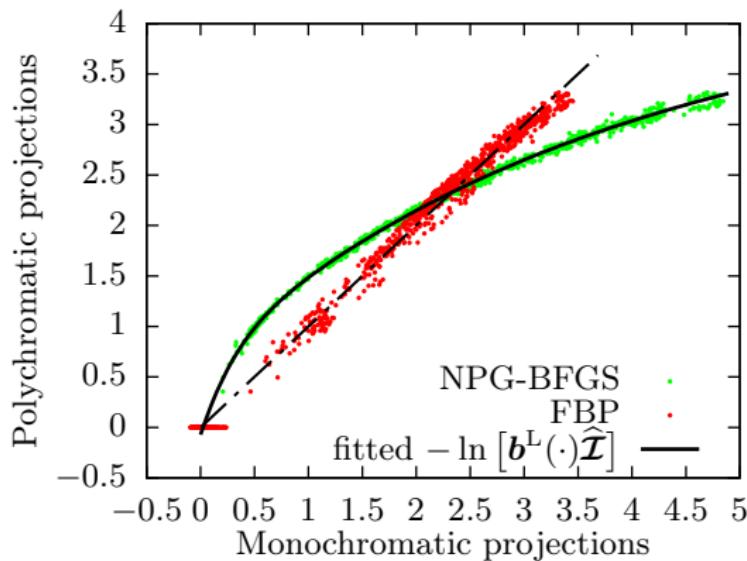
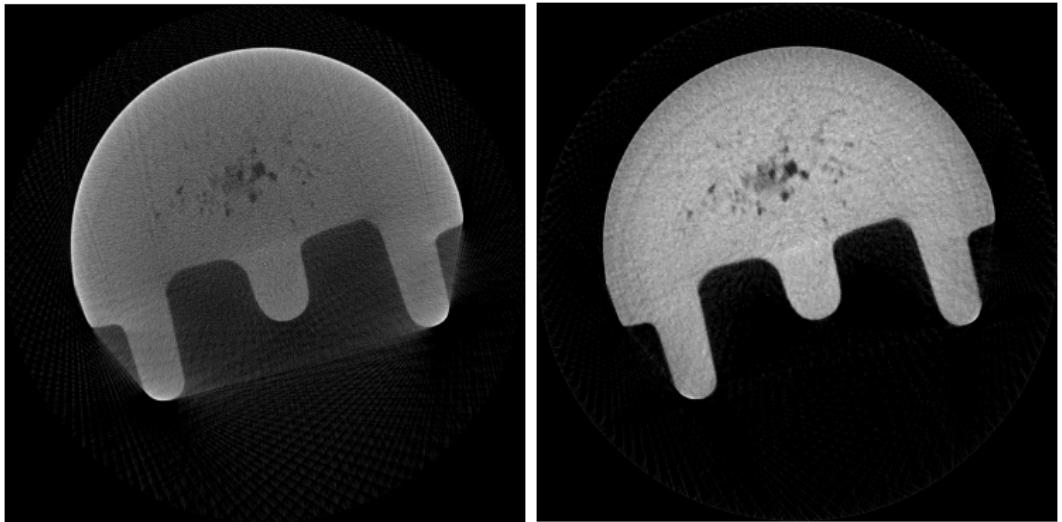


Figure 22: The polychromatic measurements as function of the monochromatic projections and its corresponding fitted curve.

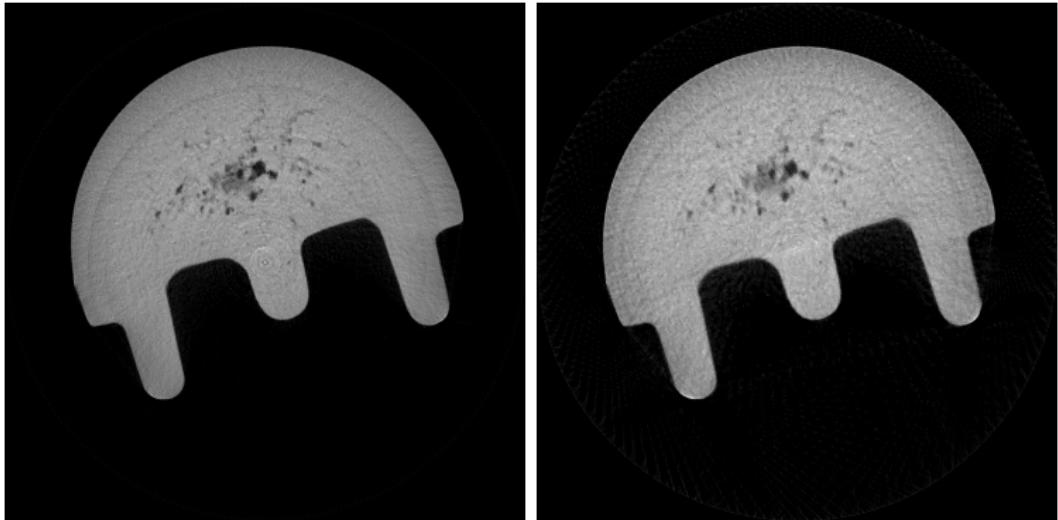


(a) FBP

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

Figure 24: Real X-ray CT: 120 fan-beam projections over 360° .

Observe the aliasing artifacts in the FBP reconstruction.



(a) 360 projections

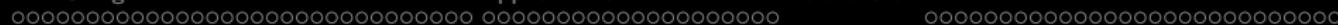
(b) 120 projections

Figure 25: NPG-BFGS ($u = 10^{-5}$) reconstructions from fan-beam projections over 360° .

Selected Publications I



- R. G. and A. D., "Blind X-ray CT image reconstruction from polychromatic Poisson measurements," *IEEE Trans. Comput. Imag.*, vol. 2, no. 2, pp. 150–165, 2016.



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, pp. 898–902.



Conclusion I

PNPG framework:

- Developed a fast framework 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.
- Proved convergence of iterates.
- Our PNPG approach is computationally efficient compared with the state-of-the-art.



Conclusion II

Polychromatic X-ray CT:

Developed a blind method for sparse density-map image reconstruction from polychromatic X-ray CT measurements in Poisson noise.

Future work: Generalize our polychromatic signal model to handle multiple materials and develop corresponding reconstruction schemes.

References I

-  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

-  A. Chambolle and C. Dossal, "On the convergence of the iterates of the "fast iterative shrinkage/thresholding algorithm"," *J. Optim. Theory Appl.*, vol. 166, no. 3, pp. 968–982, 2015.
-  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.

References III

-  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.
-  B. T. Polyak, "Some methods of speeding up the convergence of iteration methods," *USSR Comput. Math. Math. Phys.*, vol. 4, no. 5, pp. 1–17, 1964.
-  B. T. Polyak, *Introduction to Optimization*. New York: Optimization Software, 1987.
-  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.



References IV

-  **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

$$\beta^{(i)} = \xi^{t_i} \bar{\beta}^{(i)} \quad (2)$$

where $t_i \geq 0$ is the smallest integer such that (2) satisfies the majorization condition (2); *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 $f(\mathbf{x}^{(i)}) > f(\mathbf{x}^{(i-1)})$ or $\bar{\mathbf{x}}^{(i)} \in C \setminus \text{dom } \mathcal{L}$, 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

$$\text{TV: } \|\mathbf{x}^{(i,j)} - \mathbf{x}^{(i,j-1)}\|_2 \leq \eta \sqrt{\delta^{(i-1)}} \quad (3a)$$

$$\begin{aligned} \ell_1: \quad & \max \left(\|\mathbf{s}^{(i,j)} - \Psi^T \mathbf{x}^{(i,j)}\|_2, \|\mathbf{s}^{(i,j)} - \mathbf{s}^{(i,j-1)}\|_2 \right) \\ & \leq \eta \left\| \Psi^T (\mathbf{x}^{(i-1)} - \mathbf{x}^{(i-2)}) \right\|_2 \end{aligned} \quad (3b)$$

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.

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

We say that x is an approximation of $\text{prox}_{ur} a$ with ε -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

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

◀ back



Relationship with FISTA |

PNPG can be thought of as a generalized FISTA (Beck and Teboulle 2009a) that accommodates

- convex constraints,
- more general NLLs,^{||} and (increasing) adaptive step size
 - thanks to this step-size adaptation, PNPG *does not* require Lipschitz continuity of the gradient of the NLL.

◀ back

^{||}FISTA has been developed for the linear Gaussian model.

Relationship with FISTA II

- Need $B^{(i)}$ 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

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

for all i :**

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

- A simpler version of FISTA is

$$\theta^{(i)} = \frac{1}{2} + \theta^{(i-1)} = \frac{i+1}{2}$$

for $i \geq 1$.

◀ back

**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 AT

(Auslender and Teboulle 2006):

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

$$\bar{x}^{(i)} = \left(1 - \frac{1}{\theta^{(i)}} \right) x^{(i-1)} + \frac{1}{\theta^{(i)}} \tilde{x}^{(i-1)}$$

$$\tilde{x}^{(i)} = \text{prox}_{\theta^{(i)} \beta^{(i)} u r} \left(\tilde{x}^{(i-1)} - \theta^{(i)} \beta^{(i)} \nabla \mathcal{L}(\bar{x}^{(i)}) \right)$$

$$x^{(i)} = \left(1 - \frac{1}{\theta^{(i)}} \right) x^{(i-1)} + \frac{1}{\theta^{(i)}} \tilde{x}^{(i)}$$

Other variants with infinite memory are available at (Becker *et al.* 2011).

◀ back



Heavy-ball Methods

(Polyak 1964; Polyak 1987):

$$\mathbf{x}^{(i)} = \text{prox}_{\beta^{(i)} \mathcal{U} \mathcal{R}} \left(\mathbf{x}^{(i-1)} - \beta^{(i)} \nabla \mathcal{L}(\mathbf{x}^{(i-1)}) \right) + \Theta^{(i)} (\mathbf{x}^{(i-1)} - \mathbf{x}^{(i-2)}).$$

◀ back