Nonlinear tomographic reconstruction via nonsmooth optimization

Vasilis Charisopoulos — Data Science Institute, UChicago

NITMB WIP Meeting – April 8, 2025

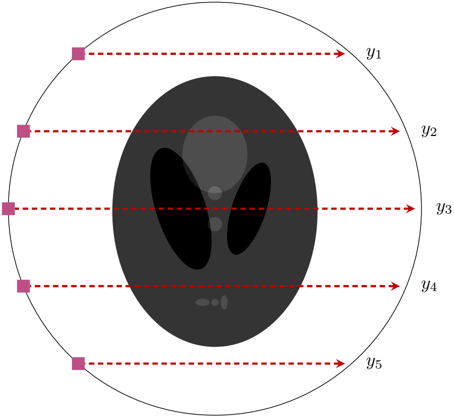
What is computed tomography?

Setup: measurements $\{y_i\}_{i=1}^m$ of object $X_\star \in \Omega^2$.

$X_\star \in \mathbb{R}^{n \times n}$

What is computed tomography?

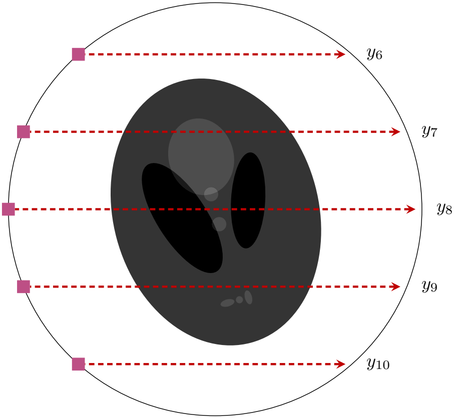
Setup: measurements $\{y_i\}_{i=1}^m$ of object $X_\star \in \Omega^2$.



$\theta = 0$

What is computed tomography?

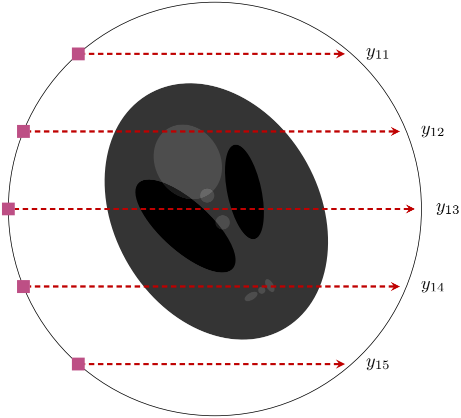
Setup: measurements $\{y_i\}_{i=1}^m$ of object $X_\star \in \Omega^2$.



$\theta = \frac{\pi}{12}$

What is computed tomography?

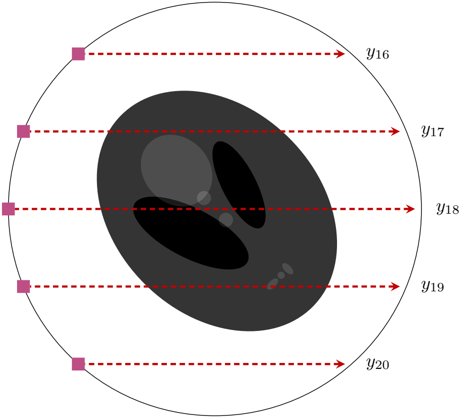
Setup: measurements $\{y_i\}_{i=1}^m$ of object $X_\star \in \Omega^2$.



$\theta = \frac{\pi}{6}$

What is computed tomography?

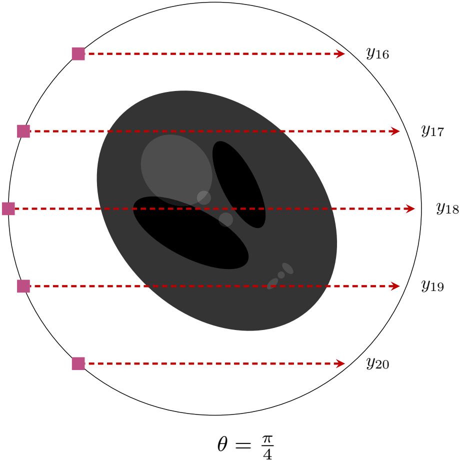
Setup: measurements $\{y_i\}_{i=1}^m$ of object $X_\star \in \Omega^2$.



$\theta = \frac{\pi}{4}$

What is computed tomography?

Setup: measurements $\{y_i\}_{i=1}^m$ of object $X_\star \in \Omega^2$.



Goal: recover $X_\star \in \mathbb{R}^{n \times n}$.

Computed tomography: measurement model

Model: attenuation of beam over X-ray spectrum.

$$y_i = \int_E S(t) \exp\Big(- \mu \int_{\ell_i} f(u) \mathrm{d}u \Big) \mathrm{d}t \tag{1}$$

Computed tomography: measurement model

Model: attenuation of beam over X-ray spectrum.

$$y_i = \int_E S(t) \exp\Big(-\mu \int_{\ell_i} f(u) \mathrm{d}u \Big) \mathrm{d}t \tag{1}$$

X-ray intensity at freq t

Computed tomography: measurement model

Model: attenuation of beam over X-ray spectrum.

$$y_i = \int_E S(t) \exp\Big(-\mu \int_{\ell_i} f(u) \mathrm{d}u\Big) \mathrm{d}t \tag{1}$$

X-ray intensity at freq t

Travel path of i^{th} beam

Computed tomography: measurement model

Model: attenuation of beam over X-ray spectrum.

$$y_i = \int_E S(t) \exp\Big(-\mu \int_{\ell_i} f(u)\mathrm{d}u \Big)\mathrm{d}t \tag{1}$$

X-ray intensity at freq t

object density at u

Travel path of i^{th} beam

Computed tomography: measurement model

Model: attenuation of beam over X-ray spectrum.

$$y_i = \int_E S(t) \exp\Big(- \mu \int_{\ell_i} f(u)\mathrm{d}u\Big)\mathrm{d}t \tag{1}$$

Radon transform: collection of all line integrals.

$$\Big\{ \mathcal{R}f(s,\theta) := \int_{\ell_{s,\theta}} f(u)\mathrm{d}u \mid s \in \mathbb{R}, \theta \in [0,\pi) \Big\}$$

Computed tomography: measurement model

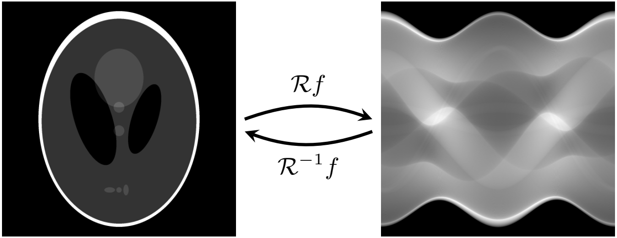
Model: attenuation of beam over X-ray spectrum.

$$y_i = \int_E S(t) \exp\Big(-\mu \int_{\ell_i} f(u) \mathrm{d}u \Big) \mathrm{d}t \tag{1}$$

Radon transform: collection of all line integrals.

$$\Big\{ \mathcal{R}f(s,\theta) := \int_{\ell_{s,\theta}} f(u)\mathrm{d}u \mid s \in \mathbb{R}, \theta \in [0,\pi) \Big\}$$

Object Sinogram



Monochromatic beams and reconstruction

Simplified model: single X-ray intensity.

$$y_i = S_{\mathsf{base}} \cdot \exp(-\langle a_i, x_\star \rangle_+), \quad i = 1, \ldots, m. \tag{2}$$

Monochromatic beams and reconstruction

Simplified model: single X-ray intensity.

$$y_i = \underbrace{S_{\mathsf{base}}}_{\text{X-ray intensity}} \cdot \exp(-\langle a_i, x_\star \rangle_+), \quad i = 1, \ldots, m. \tag{2}$$

Monochromatic beams and reconstruction

Simplified model: single X-ray intensity.

$$y_i = \underbrace{S_{\mathsf{base}}}_{\text{X-ray intensity}} \cdot \exp(-\langle \underbrace{a_i}_{\text{design vectors (known)}}, x_\star \rangle_+), \quad i = 1, \ldots, m. \tag{2}$$

Monochromatic beams and reconstruction

Simplified model: single X-ray intensity.

$$y_i = S_{\mathsf{base}} \cdot \exp(-\langle a_i, x_\star \rangle_+), \quad i = 1, \ldots, m. \tag{2}$$

X-ray intensity

$f(u)$ on $n \times n$ grid

design vectors (known)

Monochromatic beams and reconstruction

Simplified model: single X-ray intensity.

$$y_i = S_{\text{base}} \cdot \exp(-\langle a_i, x_\star \rangle_+), \quad i = 1, \ldots, m. \tag{2}$$

Assume $a_i, x_\star \in \mathbb{R}^d$, $S_{\text{base}} = 1$. How to recover $x_\star$?

Monochromatic beams and reconstruction

Simplified model: single X-ray intensity.

$$y_i = S_{\text{base}} \cdot \exp(-\langle a_i, x_\star \rangle_+), \quad i = 1, \dots, m. \tag{2}$$

Assume $a_i, x_\star \in \mathbb{R}^d$, $S_{\text{base}} = 1$. How to recover $x_\star$?

- **FBP** method: solve linear system $Ax_\star = -\log(y)$.

Monochromatic beams and reconstruction

Simplified model: single X-ray intensity.

$$y_i = S_{\mathsf{base}} \cdot \exp(-\langle a_i, x_\star \rangle_+), \quad i = 1, \ldots, m. \tag{2}$$

Assume $a_i, x_\star \in \mathbb{R}^d$, $S_{\mathsf{base}} = 1$. How to recover $x_\star$?

- **FBP** method: solve linear system $Ax_\star = -\log(y)$.
  - **Issue**: log-transform ill-conditioned; reconstruction artifacts as $\|x_\star\| \uparrow$.
  - Insufficient for polychromatic CT measurements ("beam hardening").

Monochromatic beams and reconstruction

Simplified model: single X-ray intensity.

$$y_i = S_{\text{base}} \cdot \exp(-\langle a_i, x_\star \rangle_+), \quad i = 1, \ldots, m. \tag{2}$$

Assume $a_i, x_\star \in \mathbb{R}^d$, $S_{\text{base}} = 1$. How to recover $x_\star$?

- **FBP** method: solve linear system $Ax_\star = -\log(y)$.
  - **Issue**: log-transform ill-conditioned; reconstruction artifacts as $\|x_\star\| \uparrow$.
  - Insufficient for polychromatic CT measurements ("beam hardening").
- **Alternatively**: minimize loss directly on measurement space.

$$\underset{x \in \mathcal{X} \subset \mathbb{R}^d}{\operatorname{argmin}} \mathcal{L}(x) := \frac{1}{m} \sum_{i=1}^{m} \rho(e^{-\langle a_i, x \rangle_+}; y_i), \qquad \underset{\text{penalty function}}{\rho} : \mathbb{R}_+ \times \mathbb{R}_+ \to \mathbb{R}_+.$$

Monochromatic beams and reconstruction

Simplified model: single X-ray intensity.

$$y_i = S_{\mathsf{base}} \cdot \exp(-\langle a_i, x_\star \rangle_+), \quad i = 1, \dots, m. \tag{2}$$

Assume $a_i, x_\star \in \mathbb{R}^d$, $S_{\mathsf{base}} = 1$. How to recover $x_\star$?

- **FBP** method: solve linear system $Ax_\star = -\log(y)$.
  - **Issue**: log-transform ill-conditioned; reconstruction artifacts as $\|x_\star\| \uparrow$.
  - Insufficient for polychromatic CT measurements ("beam hardening").
- **Alternatively**: minimize loss directly on measurement space.

$$\operatorname*{argmin}_{x \in \mathcal{X} \subset \mathbb{R}^d} \mathcal{L}(x) := \frac{1}{m} \sum_{i=1}^m \rho(e^{-\langle a_i, x \rangle_+}; y_i), \qquad \underbrace{\rho}_{\text{penalty function}} : \mathbb{R}_+ \times \mathbb{R}_+ \to \mathbb{R}_+.$$

- When is $\mathcal{L}$ "easy" to optimize?
  - How many measurements do we need? **(sample complexity)**
  - How fast are "reasonable" methods? **(iteration complexity)**
  - Is $x_\star$ even reachable? **(global convergence)**

Prior work: recovery with gradient descent

Natural starting point: let $\rho(\hat{y}; y) := (\hat{y} - y)^2$ (MSE loss).

$$\hat{x}_\star = \operatorname*{argmin}_{x \in \mathcal{X}} \frac{1}{m} \sum_{i=1}^{m} \left(y_i - e^{-\langle a_i, x \rangle_+} \right)^2 . \tag{3}$$

Prior work: recovery with gradient descent

Natural starting point: let $\rho(\hat{y}; y) := (\hat{y} - y)^2$ (MSE loss).

$$\hat{x}_\star = \operatorname*{argmin}_{x \in \mathcal{X}} \frac{1}{m} \sum_{i=1}^{m} \left(y_i - e^{-\langle a_i, x \rangle_+}\right)^2. \tag{3}$$

Algorithm: gradient descent with constant stepsize.

$$\begin{aligned} x_0 &:= \mathbf{0}_d, \\ x_{k+1} &:= x_k - \eta \cdot \nabla \mathcal{L}(x_k), \quad k = 0, 1, \dots \end{aligned} \tag{GD}$$

Prior work: recovery with gradient descent

Natural starting point: let $\rho(\hat{y}; y) := (\hat{y} - y)^2$ (MSE loss).

$$\hat{x}_\star = \operatorname*{argmin}_{x \in \mathcal{X}} \frac{1}{m} \sum_{i=1}^{m} \left(y_i - e^{-\langle a_i, x \rangle_+}\right)^2. \tag{3}$$

Algorithm: gradient descent with constant stepsize.

$$\boxed{\begin{aligned} x_0 &:= \mathbf{0}_d, \\ x_{k+1} &:= x_k - \eta \cdot \nabla \mathcal{L}(x_k), \quad k = 0, 1, \dots \end{aligned} \qquad \texttt{(GD)}}$$

Guarantees: exponential sample and iteration complexity.

Prior work: recovery with gradient descent

Natural starting point: let $\rho(\hat{y}; y) := (\hat{y} - y)^2$ (MSE loss).

$$\hat{x}_\star = \operatorname*{argmin}_{x \in \mathcal{X}} \frac{1}{m} \sum_{i=1}^{m} \left(y_i - e^{-\langle a_i, x \rangle_+}\right)^2. \tag{3}$$

Algorithm: gradient descent with constant stepsize.

$$\begin{aligned} x_0 &:= \mathbf{0}_d, \\ x_{k+1} &:= x_k - \eta \cdot \nabla \mathcal{L}(x_k), \quad k = 0, 1, \dots \end{aligned} \tag{GD}$$

Guarantees: exponential sample and iteration complexity.

Theorem: (Fridovich-Keil et al., 2023)

Suppose $a_i \sim_{\text{iid}} \mathcal{N}(0, I_d)$. Then, if $m/d \gtrsim e^{c_1 \|x_\star\|}$ and $\eta \lesssim e^{-c_2 \|x_\star\|}$,

$$\|x_{k+1} - x_\star\|^2 \leq \left(1 - e^{-c_3 \|x_\star\|}\right) \|x_k - x_\star\|^2, \quad \text{for } k = 0, 1, \dots$$

Prior work: recovery with gradient descent

Natural starting point: let $\rho(\hat{y}; y) := (\hat{y} - y)^2$ (MSE loss).

$$\hat{x}_\star = \operatorname*{argmin}_{x \in \mathcal{X}} \frac{1}{m} \sum_{i=1}^{m} \left(y_i - e^{-\langle a_i, x \rangle_+}\right)^2. \tag{3}$$

Algorithm: gradient descent with constant stepsize.

$$\begin{aligned} x_0 &:= \mathbf{0}_d, \\ x_{k+1} &:= x_k - \eta \cdot \nabla \mathcal{L}(x_k), \quad k = 0, 1, \dots \end{aligned} \tag{GD}$$

Guarantees: exponential sample and iteration complexity.

Question: is the exponential dependence on $\|x_\star\|$ unavoidable?

Recovery via nonsmooth optimization

Nonsmooth penalty: let $\rho(\hat{y}; y) := |\hat{y} - y|$ (ℓ_1 loss).

$$\hat{x}_\star = \operatorname*{argmin}_{x \in \mathcal{X}} \mathcal{L}(x) := \frac{1}{m} \sum_{i=1}^{m} \left| y_i - e^{-\langle a_i, x \rangle_+} \right|. \tag{4}$$

Penalty ρ is non-differentiable. How to optimize?

Recovery via nonsmooth optimization

Nonsmooth penalty: let $\rho(\hat{y}; y) := |\hat{y} - y|$ (ℓ_1 loss).

$$\hat{x}_\star = \operatorname*{argmin}_{x \in \mathcal{X}} \mathcal{L}(x) := \frac{1}{m} \sum_{i=1}^{m} \left| y_i - e^{-\langle a_i, x \rangle_+} \right|. \tag{4}$$

Penalty ρ is non-differentiable. How to optimize?

Method: `PolyakSGM`

$$x_0 := \mathbf{0}_d,$$

$$x_{k+1} := x_k - \bar{\eta} \cdot \frac{\mathcal{L}(x_k)}{\|v_k\|} \frac{v_k}{\|v_k\|}, \quad v_k \in \partial \mathcal{L}(x_k)$$

Recovery via nonsmooth optimization

Nonsmooth penalty: let $\rho(\hat{y}; y) := |\hat{y} - y|$ (ℓ_1 loss).

$$\hat{x}_\star = \operatorname*{argmin}_{x \in \mathcal{X}} \mathcal{L}(x) := \frac{1}{m} \sum_{i=1}^{m} \left| y_i - e^{-\langle a_i, x \rangle_+} \right|. \tag{4}$$

Penalty ρ is non-differentiable. How to optimize?

Method: `PolyakSGM`

$$x_0 := \mathbf{0}_d,$$

$$x_{k+1} := x_k - \bar{\eta} \cdot \overbrace{\boxed{\frac{\mathcal{L}(x_k)}{\|v_k\|}}}^{\text{Polyak step}} \frac{v_k}{\|v_k\|}, \quad v_k \in \partial \mathcal{L}(x_k)$$

Recovery via nonsmooth optimization

Nonsmooth penalty: let $\rho(\hat{y}; y) := |\hat{y} - y|$ (ℓ_1 loss).

$$\hat{x}_\star = \operatorname*{argmin}_{x \in \mathcal{X}} \mathcal{L}(x) := \frac{1}{m} \sum_{i=1}^{m} \left| y_i - e^{-\langle a_i, x \rangle_+} \right|. \tag{4}$$

Penalty ρ is non-differentiable. How to optimize?

Method: `PolyakSGM`

$$x_0 := \mathbf{0}_d,$$

$$x_{k+1} := x_k - \bar{\eta} \cdot \frac{\mathcal{L}(x_k)}{\|v_k\|} \frac{v_k}{\|v_k\|}, \quad v_k \in \partial \mathcal{L}(x_k)$$

Here, $\partial \mathcal{L}(x)$ is the *Clarke subdifferential*:

$$\partial \mathcal{L}(x) := \operatorname{conv} \left\{ \lim_{i \to \infty} \nabla \mathcal{L}(y_i) \mid y_i \to x, \ \nabla \mathcal{L} \text{ exists at } y_i \right\}$$

Convergence of `PolyakSGM`

Theorem:[1] suppose that a *convex* function h satisfies:

[1] Goffin '77

Convergence of `PolyakSGM`

Theorem:[1] suppose that a *convex* function h satisfies:

(Lipschitz) There is a constant $L > 0$ such that

$$|h(x) - h(\bar{x})| \leq L \|x - \bar{x}\|, \quad \text{for all } x, \bar{x} \in \mathbb{R}^d.$$

[1] Goffin '77

Convergence of `PolyakSGM`

Theorem:[1] suppose that a *convex* function h satisfies:

(Lipschitz) There is a constant $L > 0$ such that

$$|h(x) - h(\bar{x})| \leq L \|x - \bar{x}\|, \quad \text{for all } x, \bar{x} \in \mathbb{R}^d.$$

(Sharpness) There is a constant $\mu > 0$ such that

$$h(\bar{x}) - \min_x h(x) \geq \mu \cdot \mathrm{dist}(\bar{x}, \operatorname{argmin} h), \quad \text{for all } \bar{x} \in \mathbb{R}^d.$$

[1] Goffin '77

Convergence of `PolyakSGM`

Theorem:[1] suppose that a *convex* function h satisfies:

(Lipschitz) There is a constant $L > 0$ such that

$$|h(x) - h(\bar{x})| \leq L \|x - \bar{x}\|, \quad \text{for all } x, \bar{x} \in \mathbb{R}^d.$$

(Sharpness) There is a constant $\mu > 0$ such that

$$h(\bar{x}) - \min_x h(x) \geq \mu \cdot \operatorname{dist}(\bar{x}, \operatorname{argmin} h), \quad \text{for all } \bar{x} \in \mathbb{R}^d.$$

Then (`PolyakSGM`) generates iterates $\{x_k\}_{k \geq 1}$ satisfying

$$\|x_{k+1} - x_\star\|^2 \leq \|x_k - x_\star\|^2 \left(1 - \frac{\mu^2}{4L^2}\right), \quad \text{for all } k \in \mathbb{N}.$$

[1] Goffin '77

Convergence of PolyakSGM

Theorem:[1] suppose that a *convex* function h satisfies:

(Lipschitz) There is a constant $L > 0$ such that

$$|h(x) - h(\bar{x})| \leq L\,\|x - \bar{x}\|\,, \quad \text{for all } x, \bar{x} \in \mathbb{R}^d.$$

(Sharpness) There is a constant $\mu > 0$ such that

$$h(\bar{x}) - \min_x h(x) \geq \mu \cdot \operatorname{dist}(\bar{x}, \operatorname{argmin} h), \quad \text{for all } \bar{x} \in \mathbb{R}^d.$$

Then (PolyakSGM) generates iterates $\{x_k\}_{k \geq 1}$ satisfying

$$\|x_{k+1} - x_\star\|^2 \leq \|x_k - x_\star\|^2 \left(1 - \frac{\mu^2}{4L^2}\right), \quad \text{for all } k \in \mathbb{N}.$$

"Effective" condition number $\kappa := L/\mu$

[1] Goffin '77

Convergence of `PolyakSGM`

Theorem:[1] suppose that a *convex* function h satisfies:

(**Lipschitz**) There is a constant $L > 0$ such that

$$|h(x) - h(\bar{x})| \leq L \|x - \bar{x}\|, \quad \text{for all } x, \bar{x} \in \mathbb{R}^d.$$

(**Sharpness**) There is a constant $\mu > 0$ such that

$$h(\bar{x}) - \min_x h(x) \geq \mu \cdot \operatorname{dist}(\bar{x}, \operatorname{argmin} h), \quad \text{for all } \bar{x} \in \mathbb{R}^d.$$

Then (`PolyakSGM`) generates iterates $\{x_k\}_{k \geq 1}$ satisfying

$$\|x_{k+1} - x_\star\|^2 \leq \|x_k - x_\star\|^2 \left(1 - \frac{\mu^2}{4L^2}\right), \quad \text{for all } k \in \mathbb{N}.$$

Caveats for our loss:

- Sharpness nontrivial to verify directly for $\mathcal{L}$.
- Proof relies on convexity of the loss function.

[1] Goffin '77

Computed tomography: loss function properties

$$\mathcal{L}(x) := \frac{1}{m} \sum_{i=1}^{m} \left| y_i - e^{-\langle a_i, x_\star \rangle_+} \right| .$$

Computed tomography: loss function properties

$$\mathcal{L}(x) := \frac{1}{m}\sum_{i=1}^{m}\left|y_i - e^{-\langle a_i, x_\star\rangle_+}\right|.$$

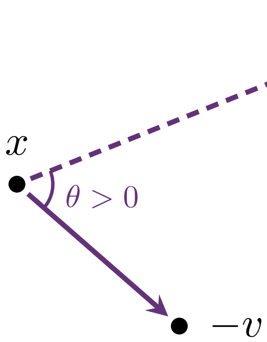
Proposition: (**C**. & Willett, 2024)

Suppose $a_i \sim_{\text{iid}} \mathcal{N}(0, I_d)$ and that $m/d \gtrsim \|x_\star\|^4$. Then $\mathcal{L}(x)$ satisfies

$$\min_{v\in\partial\mathcal{L}(x)} \frac{\langle v, x - x_\star\rangle}{\|x - x_\star\|} \gtrsim \frac{1}{\|x_\star\|^2}, \quad \forall x \in \mathcal{B}(0, 3\|x_\star\|) \setminus \{\mathbf{0}\} \quad \texttt{(Aiming)}$$

Computed tomography: loss function properties

$$\mathcal{L}(x) := \frac{1}{m}\sum_{i=1}^{m}\left|y_i - e^{-\langle a_i, x_\star\rangle_+}\right|.$$



Computed tomography: loss function properties

$$\mathcal{L}(x) := \frac{1}{m} \sum_{i=1}^{m} \left| y_i - e^{-\langle a_i, x_\star \rangle_+} \right| .$$

Proposition: (**C**. & Willett, 2024)

Suppose $a_i \sim_{\text{iid}} \mathcal{N}(0, I_d)$ and that $m/d \gtrsim \|x_\star\|^4$. Then $\mathcal{L}(x)$ satisfies

$$\min_{v \in \partial \mathcal{L}(x)} \frac{\langle v, x - x_\star \rangle}{\|x - x_\star\|} \gtrsim \frac{1}{\|x_\star\|^2}, \quad \forall x \in \mathcal{B}(0, 3\|x_\star\|) \setminus \{\mathbf{0}\} \quad \texttt{(Aiming)}$$

- **Implication**: convergence from *any* $x_0 \in \mathcal{B}(0, 3\|x_\star\|)$ (for small $\bar{\eta}$).

Loss function properties: Aiming implies decrease

$$\min_{v \in \partial \mathcal{L}(x)} \frac{\langle v, x - x_\star \rangle}{\|x - x_\star\|} \geq \mu, \quad |\mathcal{L}(x) - \mathcal{L}(\bar{x})| \leq L \, \|x - \bar{x}\|$$

Loss function properties: Aiming implies decrease

$$\min_{v \in \partial \mathcal{L}(x)} \frac{\langle v, x - x_\star \rangle}{\|x - x_\star\|} \geq \mu, \quad |\mathcal{L}(x) - \mathcal{L}(\bar{x})| \leq L \|x - \bar{x}\|$$

$$\|x_{k+1} - x_\star\|^2 = \left\| x_k - x_\star - \bar{\eta} \frac{\mathcal{L}(x_k)}{\|v_k\|^2} v_k \right\|^2$$

Loss function properties: Aiming implies decrease

$$\min_{v \in \partial \mathcal{L}(x)} \frac{\langle v, x - x_\star \rangle}{\|x - x_\star\|} \geq \mu, \quad |\mathcal{L}(x) - \mathcal{L}(\bar{x})| \leq L \|x - \bar{x}\|$$

$$\begin{aligned} \|x_{k+1} - x_\star\|^2 &= \left\| x_k - x_\star - \bar{\eta} \frac{\mathcal{L}(x_k)}{\|v_k\|^2} v_k \right\|^2 \\ &= \|x_k - x_\star\|^2 + \bar{\eta}^2 \frac{\mathcal{L}(x_k)^2}{\|v_k\|^2} - 2 \frac{\bar{\eta}\mathcal{L}(x_k)}{\|v_k\|^2} \cdot \langle v_k, x_k - x_\star \rangle \end{aligned}$$

Loss function properties: Aiming implies decrease

$$\min_{v \in \partial \mathcal{L}(x)} \frac{\langle v, x - x_\star \rangle}{\|x - x_\star\|} \geq \mu, \quad |\mathcal{L}(x) - \mathcal{L}(\bar{x})| \leq L \|x - \bar{x}\|$$

$$\begin{aligned} \|x_{k+1} - x_\star\|^2 &= \left\| x_k - x_\star - \bar{\eta} \frac{\mathcal{L}(x_k)}{\|v_k\|^2} v_k \right\|^2 \\ &= \|x_k - x_\star\|^2 + \bar{\eta}^2 \frac{\mathcal{L}(x_k)^2}{\|v_k\|^2} - 2 \frac{\bar{\eta}\mathcal{L}(x_k)}{\|v_k\|^2} \cdot \langle v_k, x_k - x_\star \rangle \\ &= \|x_k - x_\star\|^2 - \frac{\bar{\eta}\mathcal{L}(x_k)}{\|v_k\|^2} \left(2 \langle v_k, x_k - x_\star \rangle - \bar{\eta}\mathcal{L}(x_k)\right) \end{aligned}$$

Loss function properties: Aiming implies decrease

$$\min_{v\in\partial\mathcal{L}(x)} \frac{\langle v, x - x_\star\rangle}{\|x - x_\star\|} \geq \mu, \quad |\mathcal{L}(x) - \mathcal{L}(\bar{x})| \leq L\,\|x - \bar{x}\|$$

$$\begin{aligned}
\|x_{k+1} - x_\star\|^2 &= \left\|x_k - x_\star - \bar{\eta}\frac{\mathcal{L}(x_k)}{\|v_k\|^2} v_k\right\|^2 \\
&= \|x_k - x_\star\|^2 + \bar{\eta}^2 \frac{\mathcal{L}(x_k)^2}{\|v_k\|^2} - 2\frac{\bar{\eta}\mathcal{L}(x_k)}{\|v_k\|^2} \cdot \langle v_k, x_k - x_\star\rangle \\
&= \|x_k - x_\star\|^2 - \frac{\bar{\eta}\mathcal{L}(x_k)}{\|v_k\|^2}\left(2\langle v_k, x_k - x_\star\rangle - \bar{\eta}\mathcal{L}(x_k)\right) \\
&\leq \|x_k - x_\star\|^2 - \frac{\bar{\eta}\mathcal{L}(x_k)\cdot\|x_k - x_\star\|}{\|v_k\|^2}\left(2\mu - \bar{\eta}L\right)
\end{aligned}$$

Loss function properties: Aiming implies decrease

$$\min_{v \in \partial \mathcal{L}(x)} \frac{\langle v, x - x_\star \rangle}{\|x - x_\star\|} \geq \mu, \quad |\mathcal{L}(x) - \mathcal{L}(\bar{x})| \leq L \|x - \bar{x}\|$$

$$\begin{aligned}
\|x_{k+1} - x_\star\|^2 &= \left\| x_k - x_\star - \bar{\eta} \frac{\mathcal{L}(x_k)}{\|v_k\|^2} v_k \right\|^2 \\
&= \|x_k - x_\star\|^2 + \bar{\eta}^2 \frac{\mathcal{L}(x_k)^2}{\|v_k\|^2} - 2 \frac{\bar{\eta}\mathcal{L}(x_k)}{\|v_k\|^2} \cdot \langle v_k, x_k - x_\star \rangle \\
&= \|x_k - x_\star\|^2 - \frac{\bar{\eta}\mathcal{L}(x_k)}{\|v_k\|^2} \left(2 \langle v_k, x_k - x_\star \rangle - \bar{\eta}\mathcal{L}(x_k)\right) \\
&\leq \|x_k - x_\star\|^2 - \frac{\bar{\eta}\mathcal{L}(x_k) \cdot \|x_k - x_\star\|}{\|v_k\|^2} \left(2\mu - \bar{\eta} L\right)
\end{aligned}$$

Takeaway: any $\bar{\eta} < {}^{2\mu}/_{L}$ gives decrease (how much?)

Loss function properties: Aiming implies (local) sharp growth

Theorem:[a] suppose $\mathcal{L}$ satisfies (`Aiming`) on $\mathcal{B}(0; 3\|x_\star\|)$. Then

$$\mathcal{L}(x) - \mathcal{L}_\star \geq \mu \cdot \min\{\|x\|, \|x - x_\star\|\}, \quad \text{for all } x \in \mathcal{B}\Big(x_\star; \|x_\star\|\Big).$$

[a] "Solvability Lemma"; see (Clarke, 1990).

Loss function properties: Aiming implies (local) sharp growth

Theorem:[a] suppose $\mathcal{L}$ satisfies (`Aiming`) on $\mathcal{B}(0; 3\|x_\star\|)$. Then

$$\mathcal{L}(x) - \mathcal{L}_\star \geq \mu \cdot \min\{\|x\|, \|x - x_\star\|\}, \quad \text{for all } x \in \mathcal{B}\Big(x_\star; \|x_\star\|\Big).$$

[a] "Solvability Lemma"; see (Clarke, 1990).

- **Fast local convergence**: if "close enough", can readily deduce

$$\begin{aligned} \|x_{k+1} - x_\star\|^2 &\leq \|x_k - x_\star\|^2 - \frac{\bar{\eta}\mathcal{L}(x_k)\|x_k - x_\star\|}{\|v_k\|^2}(2\mu - \bar{\eta}L) \\ &\leq \|x_k - x_\star\|^2 \cdot \left(1 - \frac{\mu\bar{\eta}(2\mu - \bar{\eta}L)}{\|v_k\|^2}\right). \end{aligned}$$

Loss function properties: Aiming implies (local) sharp growth

Theorem:[a] suppose $\mathcal{L}$ satisfies (`Aiming`) on $\mathcal{B}(0; 3\|x_\star\|)$. Then

$$\mathcal{L}(x) - \mathcal{L}_\star \geq \mu \cdot \min\{\|x\|, \|x - x_\star\|\}, \quad \text{for all } x \in \mathcal{B}\Big(x_\star; \|x_\star\|\Big).$$

[a] "Solvability Lemma"; see (Clarke, 1990).

- **Fast local convergence**: if "close enough", can readily deduce

$$\begin{aligned} \|x_{k+1} - x_\star\|^2 &\leq \|x_k - x_\star\|^2 - \frac{\bar{\eta}\mathcal{L}(x_k)\|x_k - x_\star\|}{\|v_k\|^2}(2\mu - \bar{\eta}L) \\ &\leq \|x_k - x_\star\|^2 \cdot \left(1 - \frac{\mu\bar{\eta}(2\mu - \bar{\eta}L)}{\|v_k\|^2}\right). \end{aligned}$$

- Suffices to prove $\|x_k\| = \Omega(1)$ (since $x_k \to x_\star$, possibly slowly)
  - Our theory shows this for the particular case $x_0 = \mathbf{0}$.

Loss function properties: Aiming implies (local) sharp growth

Theorem:[a] suppose $\mathcal{L}$ satisfies (`Aiming`) on $\mathcal{B}(0; 3\|x_\star\|)$. Then

$$\mathcal{L}(x) - \mathcal{L}_\star \geq \mu \cdot \min\{\|x\|, \|x - x_\star\|\}, \quad \text{for all } x \in \mathcal{B}\Big(x_\star; \|x_\star\|\Big).$$

[a] "Solvability Lemma"; see (Clarke, 1990).

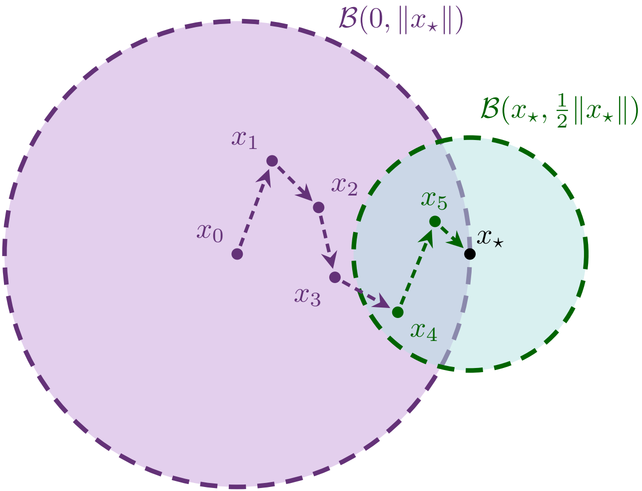
- **Fast local convergence**: if "close enough", can readily deduce

$$\begin{aligned} \|x_{k+1} - x_\star\|^2 &\leq \|x_k - x_\star\|^2 - \frac{\bar{\eta}\mathcal{L}(x_k)\|x_k - x_\star\|}{\|v_k\|^2}(2\mu - \bar{\eta}L) \\ &\leq \|x_k - x_\star\|^2 \cdot \left(1 - \frac{\mu\bar{\eta}(2\mu - \bar{\eta}L)}{\|v_k\|^2}\right). \end{aligned}$$

- Suffices to prove $\|x_k\| = \Omega(1)$ (since $x_k \to x_\star$, possibly slowly)
  - Our theory shows this for the particular case $x_0 = \mathbf{0}$.

- Final result: initial "slow" phase (until $\|x_k - x_\star\| \leq \frac{1}{2}\|x_\star\|$).

Algorithm behavior: high-level sketch



Overview of theoretical results

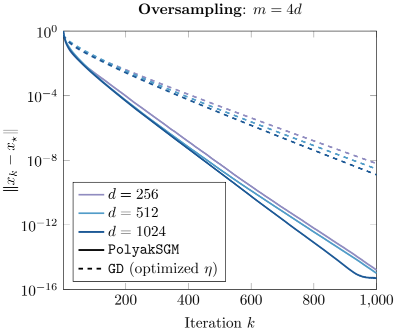
Method	Iterations	Samples	Reference
`PolyakSGM`	$O\left(\|x_\star\|^6 \log\left(\frac{\|x_\star\|}{\varepsilon}\right)\right)$	$O(\|x_\star\|^4 \cdot d)$	`arXiv:2407.12984`
`GD`	$O\left(e^{c_1\|x_\star\|} \log\left(\frac{\|x_\star\|}{\varepsilon}\right)\right)$	$O\left(\frac{e^{c_2\|x_\star\|}}{\|x_\star\|^2} \cdot d\right)$	`arXiv:2310.03956`

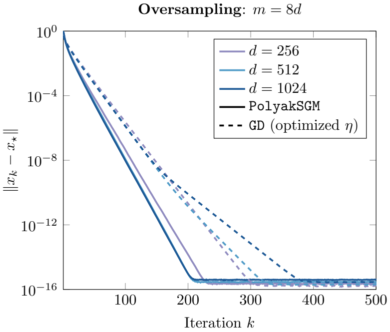
Table: Iteration and sample complexity of iterative reconstruction methods.

Takeaway: *exponential* improvements in both metrics.

Numerics

Question 1: is the algorithm fast?

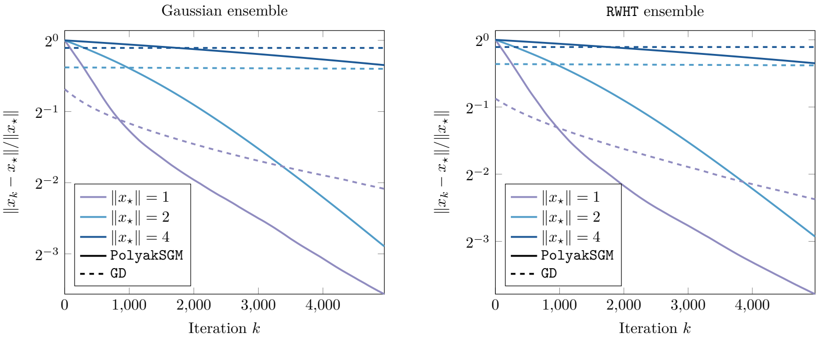




Setup: Gaussian measurements, $\|x_\star\| = 1$, $\bar{\eta} = 1$ (no stepsize scaling).

Numerics

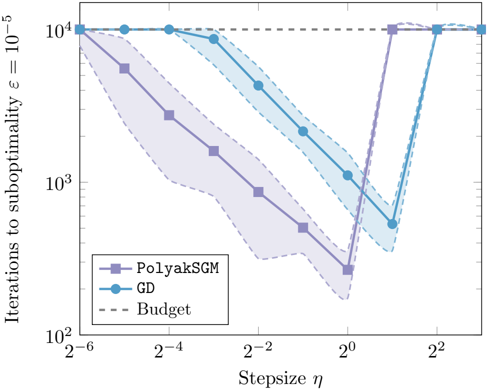
Question 2(a): how "loose" are the stepsize bounds? (η and $\bar{\eta}$)



Setup: Gaussian measurements, $\bar{\eta} \propto \frac{1}{\|x_\star\|^2}$, $\eta \propto e^{-5\|x_\star\|}$.

Numerics

Question 2(b): is the algorithm stable?



Setup: Gaussian measurements, 10 random instances per η value.

Numerics

Question 3: is the algorithm sample-efficient?

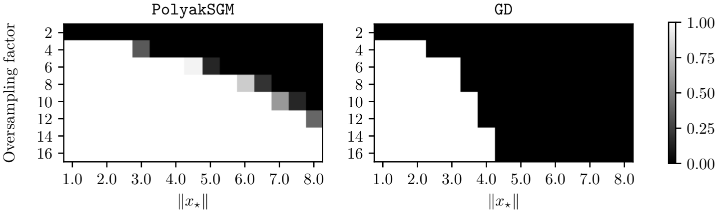


Figure: Recovery probability (random instances). Tile color indicates probability.

Setup: Gaussian measurements, 25 instances per tile, threshold $\varepsilon = 10^{-5}$.

Numerics

Question 4: does it work with "real" measurements?

(a) `GD`; `PSNR` $= 31.77$ **(b)** `PolyakSGM`; `PSNR` $= 38.12$ **(c)** Ground truth

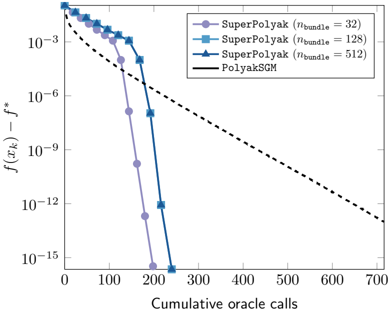
Setup: Radon-type measurements ($^m/_d = {}^1/_4$); project to TV norm[2] ball:

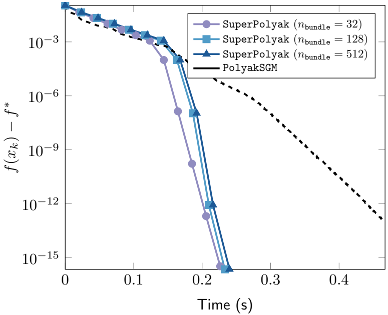
$$\mathcal{X} := \{x \mid \|x\|_{\mathsf{TV}} \leq \lambda\}.$$

[2] Rudin-Osher-Fatemi '92

Numerics

Question 5: can we accelerate it?





- **Setup**: back to Gaussian measurements $(m = 4d)$
- `SuperPolyak`: method using multiple linearizations per step.[3]

[3] `arXiv:2201.04611`

Thank you!

arXiv:2407.12984

vchariso.com