Towards the numerical quantification of source conditions

Daniel Gerth

Chemnitz University of Technology

Modern Challenges in Imaging
Tufts University, Medford, MA
August 5, 2019

Overview

- Introduction
- The Kurdyka-Łojasiewicz property
- Observable lower bounds
- Estimating μ
- Numerical results

Setting

Solve

$$Ax = y$$

where:

- X, Y Hilbert spaces (unless specified otherwise)
- $A : X \to Y$ a linear, compact operator
- $\mathcal{R}(A) \neq \overline{\mathcal{R}(A)}$, $A^\dagger$ unbounded
- exact data $y = Ax^\dagger \in Y$, $x^\dagger \in X$
- available data y^δ, $||y - y^\delta|| \leq \delta$, $\delta > 0$.

practical task: from A, y^δ, find “good” approximation x^δ to $x^\dagger$
theoretical task: find φ such that $||x^\delta - x^\dagger|| \leq \varphi(\delta)$.

Classical regularization theory

- in general no such φ exists
- have to restrict possible solutions $x^\dagger$
- classical approach: **source condition**

$$x^\dagger \in \mathcal{R}((A^*A)^\mu), \quad \mu > 0, \qquad (SC)$$

- then $\varphi_\mu(\delta) \sim \delta^{\frac{2\mu}{2\mu+1}}$, namely

$$\sup\{||x - x^\dagger|| : ||Ax - y|| \leq \delta, ||x||_\mu \leq \varrho\} \leq \delta^{\frac{2\mu}{2\mu+1}} \varrho^{\frac{1}{2\mu+1}}.$$

Regularization

classical Tikhonov regularization

$$x_\alpha^\delta = \operatorname{argmin} ||Ax - y^\delta||^2 + \alpha||x||^2.$$

If (SC) with $0 < \mu \leq 1$ and

$$\alpha \sim \delta^{\frac{2}{2\mu+1}},$$

then $||x_\alpha^\delta - x^\dagger|| \leq c\varphi_\mu(\delta)$.
If (SC) with $0 < \mu \leq \frac{1}{2}$ and

$$\alpha \quad \text{s.t.} \quad ||Ax_\alpha^\delta - y^\delta|| \sim \delta,$$

then $||x_\alpha^\delta - x^\dagger|| \leq c\varphi_\mu(\delta)$.

Regularization

classical Tikhonov regularization

$$x_\alpha^\delta = \operatorname{argmin} ||Ax - y^\delta||^2 + \alpha ||x||^2.$$

If (SC) with $0 < \mu \leq 1$ and

$$\alpha \sim \delta^{\frac{2}{2\mu+1}},$$

then $||x_\alpha^\delta - x^\dagger|| \leq c\varphi_\mu(\delta)$.
If (SC) with $0 < \mu \leq \frac{1}{2}$ and

$$\alpha \quad \text{s.t.} \quad ||Ax_\alpha^\delta - y^\delta|| \sim \delta,$$

then $||x_\alpha^\delta - x^\dagger|| \leq c\varphi_\mu(\delta)$.

Parameter choice requires knowledge of μ.

Regularization, ctd.

Landweber method: pick $x_0 \in X$. Iterate

$$x_{k+1} = x_k + \beta A^*(Ax_k - y^\delta),$$

$\beta \leq \frac{2}{||A||}$, until STOP. Then if (SC) for $\mu > 0$ and

$$k_{stop} \sim \delta^{-\frac{2}{2\mu+1}}$$

or k_{stop} such that

$$||Ax_{k_{stop}} - y^\delta|| \sim \delta,$$

then $||x_\alpha^\delta - x^\dagger|| \leq c\varphi_\mu(\delta)$.

IF $x^\dagger \in \mathcal{R}((A^*A)^\mu)$.

The typical situation in practice

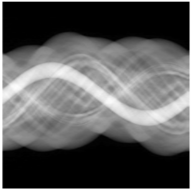
given: $A \in \mathbb{R}^{m\times n}$, ONE sample $y^\delta \in \mathbb{R}^m$, δ unknown, μ unknown

Questions/Tasks:

- How to choose α?
- How large is δ?
- How large is μ?
- Are the (infinite dimensional) source conditions relevant in a discretized setting?

Consider X-ray Tomography:

y^δ



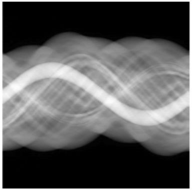
The typical situation in practice

given: $A \in \mathbb{R}^{m\times n}$, ONE sample $y^\delta \in \mathbb{R}^m$, δ unknown, μ unknown
Questions/Tasks:

- How to choose α?
- How large is δ?
- How large is μ?
- Are the (infinite dimensional) source conditions relevant in a discretized setting?

Consider X-ray Tomography:

y^δ



$x^\dagger$

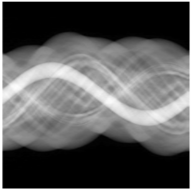
The typical situation in practice

given: $A \in \mathbb{R}^{m \times n}$, ONE sample $y^\delta \in \mathbb{R}^m$, δ unknown, μ unknown
Questions/Tasks:

- How to choose α? not the focus of this talk
- How large is δ? known algorithms / ongoing research
- How large is μ? This talk. (example below: $\mu \approx 0.22$)
- Are the (infinite dimensional) source conditions relevant in a discretized setting? Yes.

Consider X-ray Tomography:

y^δ



$x^\dagger$

Overview

- Introduction
- The Kurdyka-Łojasiewicz property
- Observable lower bounds
- Estimating μ
- Numerical results

Introduction

- S. Łojasiewicz, 1960s: Let f be a real-analytic function. Then

$$\exists\theta \in [0,1): \quad |f - f(\bar{x})|^{\theta} ||\nabla f||^{-1}$$

  remains bounded around any critical point $\bar{x}$

- K. Kurdyka, 1998: f a C^1 function on a Hilbert space, then there is $\varphi \in \mathcal{K}(0,r_0)$, $\mathcal{K}(0,r_0) :=$
  $\{\varphi : [0,r_0) \to \mathbb{R} \in C[0,r_0) \cap C^1(0,r_0),\ \varphi(0) = 0,\ \varphi'(x) > 0\}$

$$||\nabla(\varphi \circ (f - \min f))|| \geq 1$$

Introduction

- S. Łojasiewicz, 1960s: Let f be a real-analytic function. Then

$$\exists\theta \in [0,1): \quad |f - f(\bar{x})|^{\theta} ||\nabla f||^{-1}$$

  remains bounded around any critical point $\bar{x}$

- K. Kurdyka, 1998: f a C^1 function on a Hilbert space, then there is $\varphi \in \mathcal{K}(0, r_0)$, $\mathcal{K}(0, r_0) :=$ $\{\varphi : [0, r_0) \to \mathbb{R} \in C[0, r_0) \cap C^1(0, r_0),\ \varphi(0) = 0,\ \varphi'(x) > 0\}$

$$||\nabla(\varphi \circ (f - \min f))|| \geq 1$$

- this is the *Kurdyka-Łojasiewicz (KL) inequality* or -property
- applications in, e.g., asymptotic analysis of nonlinear heat equation (Łojasiewicz), PDE analysis, nonsmooth analysis, neural networks, complexity analysis,.... Inverse Problems?

KL can be connected directly to regularization theory.
Let $T_\alpha(x) := \|Ax - y\|^2 + \alpha J(x)$, $J : X \to \mathbb{R}$ a stabilizing penalty term, X a Banach space.

Theorem (D.G., S. Kindermann, 2019)

The following statements are equivalent: There are index functions $\Psi_1, \Psi_2, \Psi_3, \Psi_4, \Psi_5, \varphi$, *that are all related, such that*

(a) *(J-rate)* $J(x^\dagger) - J(x_\alpha) \le \Psi_1(\alpha) \quad \text{for all} \quad \alpha > 0.$

(b) *(T-rate)* $\frac{1}{\alpha}\left(T_\alpha(x^\dagger) - T_\alpha(x_\alpha)\right) \le \Psi_2(\alpha) \quad \text{for all} \quad \alpha > 0.$

(c) *(Variational inequality)*
$J(x^\dagger) - J(x) \le \Psi_3(\|Ax^\dagger - Ax\|) \qquad \text{for all} \quad x \in X.$

(d) *(Distance function)* $D(\frac{1}{r}) \le \Psi_4(r) \qquad \forall r > 0$
where $D(r) := \sup_{x \in X}\left(J(x^\dagger) - J(x) - r\|Ax - Ax^\dagger\|\right)$.

(e) *(Dual T-rate)*
$J^*(A^*z) - J(x^*) - (x^\dagger, A^*z - x^*)_{X,X^*} + \alpha\frac{1}{2}\|z\|^2 \le \Psi_5(\alpha).$

(f) *(KL-inequality)* $\left\|\partial\left(\varphi \circ \left(T_\alpha(x^\dagger) - T_\alpha(x_\alpha)\right)\right)\right\| \ge \frac{1}{k}$.

Theorem

Let any of the equivalent assumptions hold. Define $\Theta(\alpha) := \sqrt{\alpha\Psi_2(\alpha)}$. *Then with the choice*

$$\alpha = \Theta^{-1}\left(\frac{\delta^2}{2}\right)$$

we obtain the convergence rates

$$B_{\xi_\alpha^\delta}(x_\alpha^\delta, x^\dagger) \leq 2\Psi_2\left(\Theta^{-1}\left(\frac{\delta}{\sqrt{2}}\right)\right) \qquad .$$

in the Bregman distance

$$B_\xi(z,x) := J(x) - J(z) - \langle \xi, x-z \rangle \geq 0, \quad x \in X, \quad \xi \in \partial J(z) \subset X^*.$$

Theorem

Let any of the equivalent assumptions hold. Define $\Theta(\alpha) := \sqrt{\alpha\Psi_2(\alpha)}$. *Then with the choice*

$$\alpha = \Theta^{-1}\left(\frac{\delta^2}{2}\right) \sim \frac{1}{\partial\varphi(\delta^2)}$$

we obtain the convergence rates

$$B_{\xi_\alpha^\delta}(x_\alpha^\delta, x^\dagger) \leq 2\Psi_2\left(\Theta^{-1}\left(\frac{\delta}{\sqrt{2}}\right)\right) \sim \partial\varphi(\delta^2)\delta^2.$$

in the Bregman distance

$$B_\xi(z,x) := J(x) - J(z) - \langle \xi, x-z \rangle \geq 0, \quad x \in X, \quad \xi \in \partial J(z) \subset X^*.$$

A connection between a KL inequality and convergence rates is general and well-known in the optimization community:

Proposition, J. Bolte et. al., 2010, Cor. 7

Let X be a metric space, $f : X \to \mathbb{R} \cup \{\infty\}$ be lower semicontinuous. Then the following assumptions are equivalent:

- $\varphi \circ f$ is k-metrically regular on $[0 < f < r_0] \times (0, \varphi(r_0))$,
- for all $r_1, r_2 \in (0, r_0)$ it is

$$D([f = r_1 - \inf f], [f = r_2 - \inf f]) \leq k|\varphi(r_1 - \inf f) - \varphi(r_2 - \inf f)|,$$

- for all $x \in [0 < f < r_0]$ it holds that

$$|\nabla(\varphi \circ f)|(x) \geq \frac{1}{k}.$$

In Banach spaces X (generalizable to complete metric spaces)

Corollary

Let either A be injective or J be strictly convex. Then, for the Tikhonov functional $T_\alpha(x)$, the following are equivalent for a smooth index function $\varphi \in \mathcal{K}(0, \tilde{r})$, $x \in [T_\alpha(x_\alpha) \leq T_\alpha(x) \leq \tilde{r}]$, and $0 < k < \infty$.

- $$\|x - x_\alpha\| \leq k\varphi(T_\alpha(x) - T_\alpha(x_\alpha)),$$
- $$\varphi'(T_\alpha(x) - T_\alpha(x_\alpha)) \, \|\partial T_\alpha(x)\|_- \geq \frac{1}{k}.$$

Bringing in the source condition

Theorem (D.G. 2018)

Let $x - x^\dagger = (A^*A)^\mu w$, $w \in X$, $f(x) = ||Ax - y||^2$. *Then*

$$\varphi'(f(x) - f(\bar{x}))||\nabla f(x)|| \geq \frac{1}{k}.$$

with $\varphi(t) = t^{\frac{\mu}{2\mu+1}}$ *and* $k = ||w||^{\frac{1}{2\mu+1}}$.

Bringing in the source condition

Theorem (D.G. 2018)

Let $x - x^\dagger = (A^*A)^\mu w$, $w \in X$, $f(x) = ||Ax - y||^2$. *Then*

$$\varphi'(f(x) - f(\bar{x}))||\nabla f(x)|| \geq \frac{1}{k}.$$

with $\varphi(t) = t^{\frac{\mu}{2\mu+1}}$ *and* $k = ||w||^{\frac{1}{2\mu+1}}$.

- rewritten: $||Ax - y||^{-\frac{2\mu+2}{2\mu+1}}||A^*(Ax - y)|| \geq c$
- Let $\varrho := ||w||$ and x such that $||Ax - y|| \sim \delta$. Then $||x - x^\dagger|| \leq c_\mu \delta^{\frac{2\mu}{2\mu+1}} \varrho^{\frac{1}{2\mu+1}}$
- proof of theorem: interpolation inequality

Overview

- Introduction
- The Kurdyka-Łojasiewicz property
- Observable lower bounds
- Estimating μ
- Numerical results

Theorem (D.G. 2018)

Let $A : X \to Y$ be a linear operator between Hilbert spaces X and Y and $x^\dagger \in X$. Then, whenever $A^*(Ax - Ax^\dagger) \neq 0$, there holds for any $x \in X$

$$\frac{\|Ax - Ax^\dagger\|^2}{\|A^*(Ax - Ax^\dagger)\|} \leq \|x - x^\dagger\|.$$ [1]

If additionally $x - x^\dagger = (A^*A)^\mu w$, $\|w\| < \infty$, then

$$\|Ax - Ax^\dagger\|^{\frac{2\mu}{2\mu+1}} \|w\|^{\frac{1}{2\mu+1}} \leq \|x - x^\dagger\| \leq c_\mu \|Ax - Ax^\dagger\|^{\frac{2\mu}{2\mu+1}} \|w\|^{\frac{1}{2\mu+1}}$$

with $c_\mu = \frac{2\mu+1}{2\mu}$.

[1] compare [Brezinski, Rodriguez, Seatzu 2008]

Overview

- Introduction
- The Kurdyka-Łojasiewicz property
- Observable lower bounds
- Estimating μ
- Numerical results

estimating μ: idea

- Lojasiewicz-property for $f(x) = ||Ax - y||^2 \to \min_x$ relates $||Ax - y||$ and $||A^*(Ax - y)||$
- consider Landwebers method. Let $x_0 \in X$, iterate for $k = 0, \ldots$, $\beta < 2/||A||^2$

$$x_{k+1} = x_k - \beta A^*(Ax - y)$$

- residual and gradient are computed anyway, norms "for free":

$$R := (||Ax_1 - y||, ||Ax_2 - y||, ||Ax_3 - y||, \ldots, ||Ax_K - y||)^T$$
$$G := (||A^*(Ax_1 - y)||, ||A^*(Ax_2 - y)||, \ldots, ||A^*(Ax_K - y)||)^T.$$

Regression

- Łojasiewicz property:

$$\varphi'(R_i^2) \cdot G_i \geq c \quad \forall i = 1, \dots, k$$

- from source condition: $\varphi(t) = ct^{\frac{\mu}{2\mu+1}}$ and $\varphi'(t) = ct^{-\frac{\mu+1}{2\mu+1}}$.
- Set $\gamma := \frac{2\mu+2}{2\mu+1}$. Then

$$\varphi'(R_i^2) = c(R_i^2)^{-\frac{\mu+1}{2\mu+1}} = cR_i^{-\gamma} \quad \forall i = 1, \dots, k,$$

and therefore we obtain

$$\frac{R_i^\gamma}{c} \leq G_i \quad \forall i = 1, \dots, k.$$

- we have measured R_i and G_i, we find γ and c by linear regression. This immediately yields $\mu_k = \frac{2-\gamma}{2\gamma-2}$ and $c_k = \exp(\hat{c})$.

Overview

- Introduction
- The Kurdyka-Łojasiewicz property
- Observable lower bounds
- Estimating μ
- Numerical results

First try: diagonal operators

- $A : \ell^2 \to \ell^2$, $A : (x_1, x_2, \cdots) \mapsto (\sigma_1 x_2, \sigma_2 x_2, \dots)$ for $\sigma_i \in \mathbb{R}$, $i \in \mathbb{N}$
- let $\sigma_i = i^{-\beta}$ for some $\beta > 0$
- let x be given as $x_i = i^{-\eta}$ with $\eta > 0$
- Because for a compact linear operator A between Hilbert spaces X and Y with singular system $\{\sigma_i, u_i, v_i\}_{i=1}^{\infty}$

$$x \in \mathcal{R}((A^*A)^\mu) \quad \Leftrightarrow \quad \sum_{i=1}^{\infty} \frac{|\langle Ax, u_i \rangle|^2}{\sigma_i^{2+4\mu}} < \infty$$

we have $x \in \mathcal{R}((A^*A)^\mu)$ for $\mu \leq \frac{2\eta-1}{4\beta} - \epsilon$ and $\epsilon > 0$

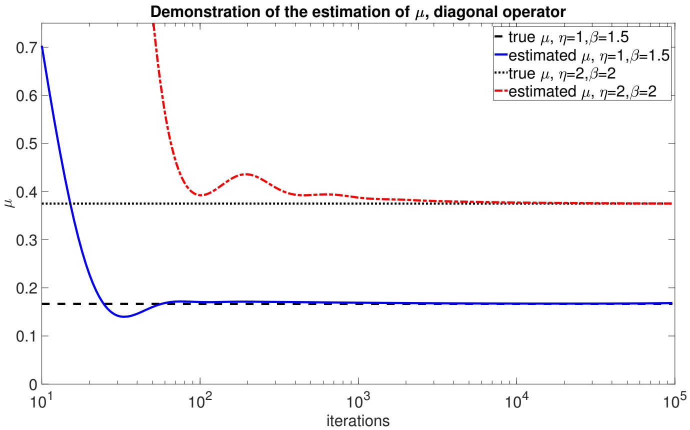


Figure: Demonstration of the method for $\eta = 1$ and $\beta = 1.5$ (red, dash-dotted), with correct $\mu = 0.375$ (black, dotted), and for $\eta = 2$ and $\beta = 2$ (blue, solid), with correct $\mu = 0.175$ (black, dashed).

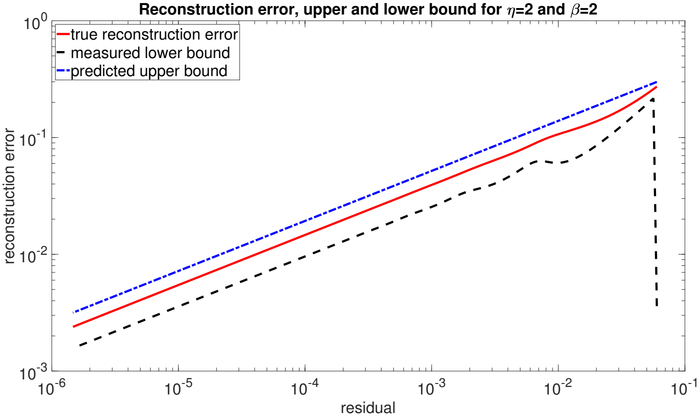


Figure: Reconstruction error (red, solid) for $\eta = 2$ and $\beta = 2$. The upper bound (dash-dotted, blue) and the observed lower bound (black, dashed) are parallel as expected.

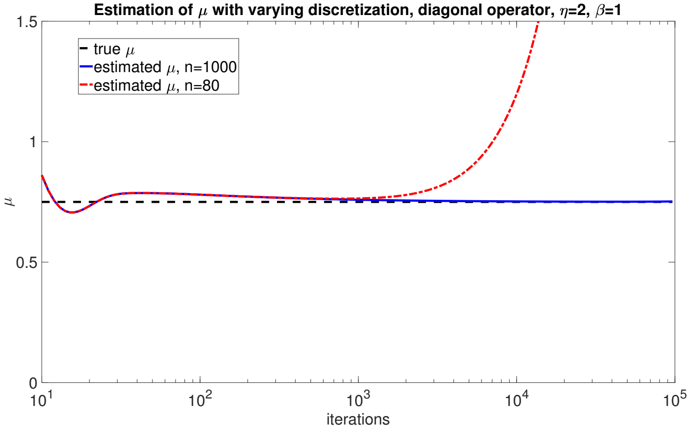


Figure: Demonstration of the method for $\eta = 2$ and $\beta = 1$ with two discretization levels.

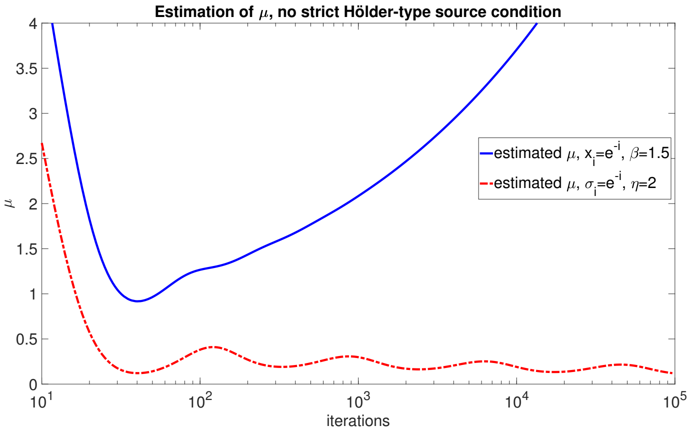


Figure: Demonstration of the method when the source condition holds for all $\mu > 0$ (solid, blue; diagonal setting with $\eta = 2$ and $\sigma_i = e^{-i}$), and when it fails for every $\mu > 0$ (dash-dotted, red; diagonal setting with $x_i = e^{-i}$ and $\beta = 1.5$).

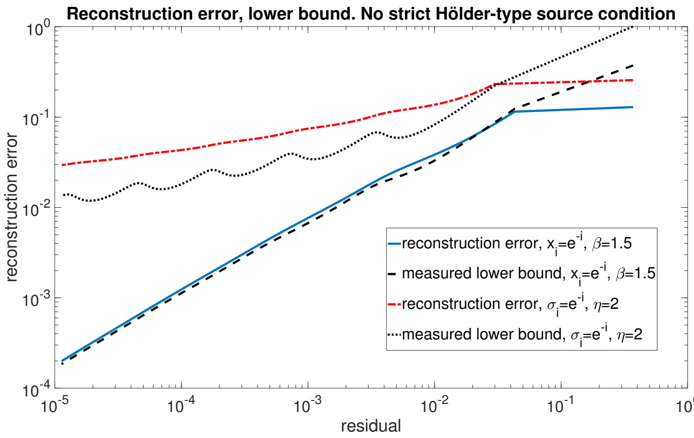


Figure: Reconstruction error and observed lower bounds when the Hölder-type source condition is inadequate. In the first case, the SC holds for all $\mu > 0$ (blue, solid: reconstruction error, black, dashed: observed lower bound). In the second case there is no $\mu > 0$ such that the SC is fulfilled.

Examples from Regularization Tools

- Inversion Toolbox by P.C. Hansen, `http://www.imm.dtu.dk/~pcha/Regutools/`
- 14 linear inverse problems as examples
- after trying our algorithm: two positive results (tomo, $\mu \approx 0.2$, deriv2, $\mu \approx 0.13$), 12 failures

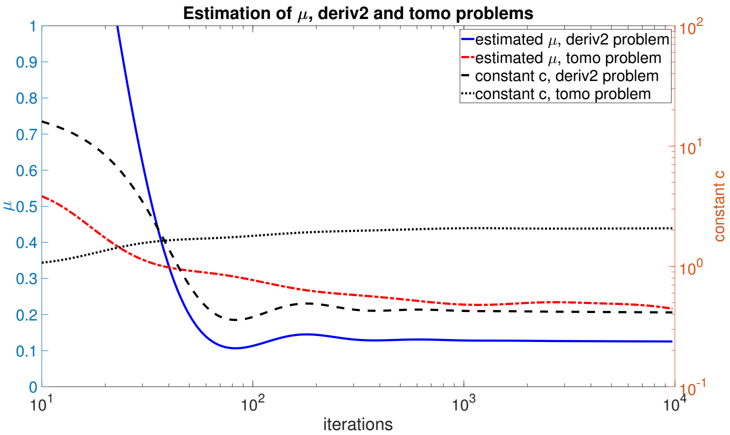


Figure: Demonstration of the method for the problems *deriv2* (solid, blue: μ; black, dashed: the constant c) and *tomo* (dash-dotted, red: μ; black, dotted: the constant c).

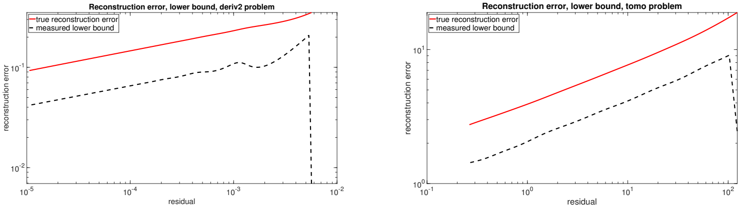


Figure: Reconstruction error (red, solid) and lower bounds (black, dashed) for the problems *deriv2* (left) and *tomo* (right).

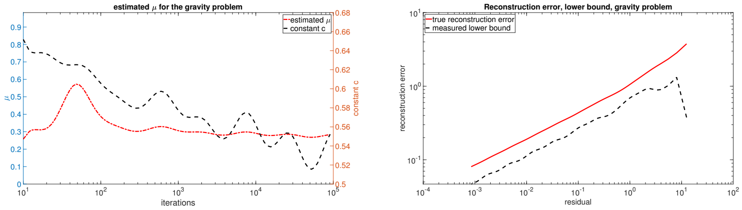


Figure: Demonstration of the method for the *gravity* problem. Left: estimated μ and constant c. Right: reconstruction error and observed lower bound.

Final test: we compute SVD of all 14 problems for several discretization levels and estimate for which μ the sum

$$\sum_{i=1}^{\infty} \frac{|\langle Ax, u_i\rangle|^2}{\sigma_i^{2+4\mu}}$$

converges. It does only for tomo ($\mu \approx 0.2$) and deriv2 ($\mu \approx 0.1$).

Noisy data

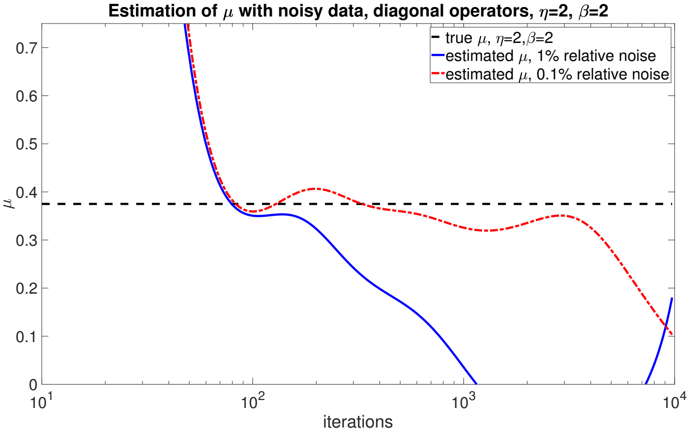
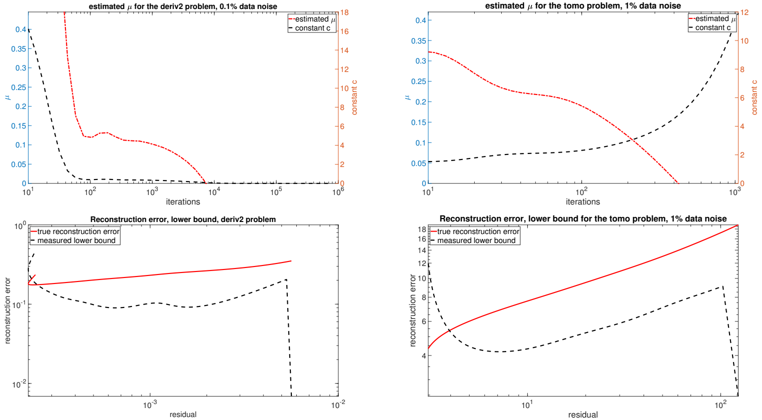
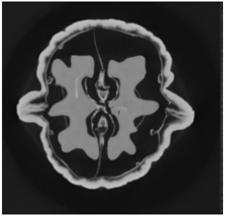


Figure: Demonstration of the method for $\eta = 2$ and $\beta = 2$ with 1% and 0.1% relative data noise.


estimated μ for the deriv2 problem, 0.1% data noise
estimated μ
constant c
iterations
constant c
estimated μ for the tomo problem, 1% data noise
estimated μ
constant c
iterations
constant c
Reconstruction error, lower bound, deriv2 problem
true reconstruction error
measured lower bound
reconstruction error
residual
Reconstruction error, lower bound for the tomo problem, 1% data noise
true reconstruction error
measured lower bound
reconstruction error
residual

Now: real tomographic X-ray data of a carved cheese, a lotus root, and of a walnut which are freely available at `http://www.fips.fi/dataset.php`.

That means: no noise information, approximate forward operator, large system and not enough discretization for SVD analysis $\Rightarrow$ our method is the only way

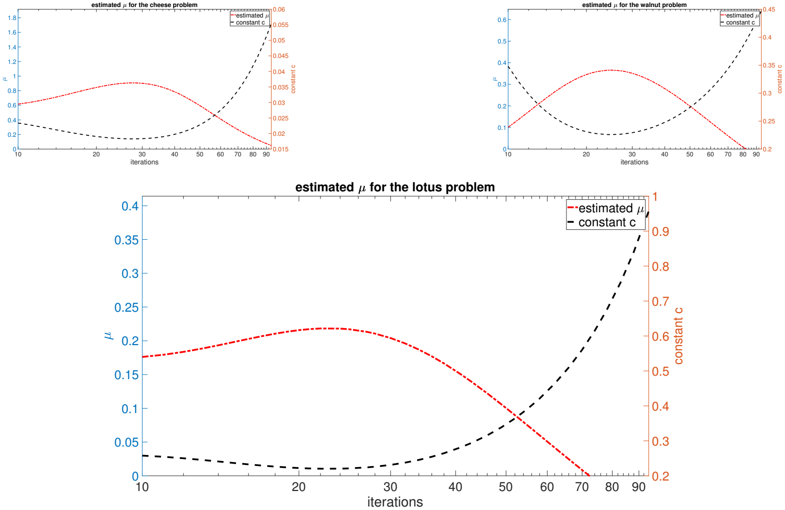


Figure: Demonstration of the method for the problems cheese (upper left), walnut (upper right), and lotus (lower row).

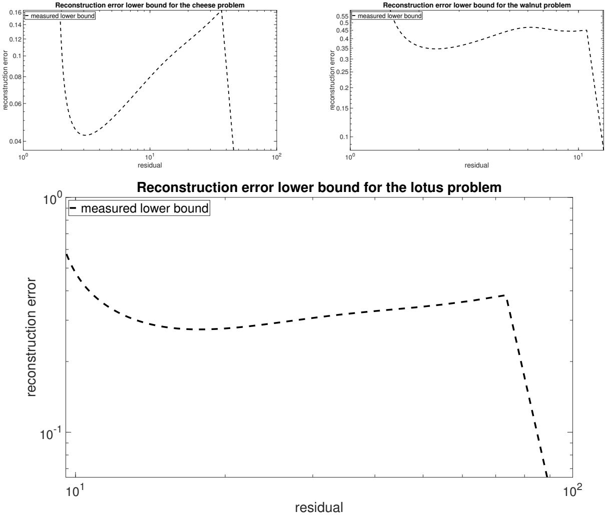


Figure: Observed lower bound for the three real data problems cheese (upper left), walnut (upper right), and lotus (lower row).

Conclusion: It is possible to extract source smoothness information given a single set of matrix A and data y. There are plenty open questions.

Implications/Future work:

- understand KL with noise
- understand relation between convergence in Bregman distance and in norm
- estimate δ simultaneously (a prototype is working)
- with μ and δ: parameter choice rules/stopping criteria
- a-posteriori error estimates may become feasible
- extension to other algorithms

- D.G. *Using Landwebers iteration to quantify source conditions - a numerical study*, J. Inverse Ill-posed Probl., 2018 (available online ahead of print)
- D.G., S. Kindermann *The Kurdyka-Łojasiewicz inequality as regularity condition*, submitted
- S. Łojasiewicz *Division d'une distribution par une fonction analytique de variables reeles* C.R. Acad. Sci, Paris 246, 683–686, 1958
- C. Brezinski, G. Rodriguez and S. Seatzu *Error estimates for linear systems with applications to regularization* Numer. Algor. 104, 2008
- J. Bolte, A. Daniilidis, O. Ley and L. Mazet. *Characterizations of Lojasiewicz inequalities: subgradient flows, talweg, convexity*. T. Am. Math. Soc. 382(6):3319–3363, 2010

Thank you for your attention!