Stochastic convergence analysis for Tikhonov-Regularization with sparsity constraints

D. Gerth, R. Ramlau

Nanjing, 2012-10-22

Doctoral Program
Computational Mathematics
Numerical Analysis and Symbolic Computation

JOHANNES KEPLER UNIVERSITY LINZ

- Introduction
- Convergence results
- A numerical example

Overview

- Introduction
- Convergence results
- A numerical example

Starting question:

Can we prove convergence (-rates) for Tikhonov-Regularization with sparsity-penalty if instead of $||y - y^\delta|| \leq \delta$ an explicit stochastic error model is used?

Setting

We study the solution of the linear ill-posed problem

$$\mathbf{A}\mathbf{x} = \mathbf{y}$$

with $\mathbf{A} \in \mathcal{L}(\mathcal{X}, \mathcal{Y})$ where $\mathcal{X}$ and $\mathcal{Y}$ are Hilbert spaces.

Setting

We study the solution of the linear ill-posed problem

$$\mathbf{A}\mathbf{x} = \mathbf{y}$$

with $\mathbf{A} \in \mathcal{L}(\mathcal{X}, \mathcal{Y})$ where $\mathcal{X}$ and $\mathcal{Y}$ are Hilbert spaces.
computation requires discretization, done via projections

$$P_m : \mathcal{Y} \to \mathbb{R}^m, \quad \mathbf{y} \mapsto y, \quad \text{e.g. point evaluation}$$
$$T_n : \mathcal{X} \to \mathbb{R}^n, \quad x = T_n\mathbf{x} = \{\langle \mathbf{x}, \psi_i \rangle\}_{i=1,\dots,n}$$

where $\{\psi_i\}_{i=1}^{\infty}$ is a basis in $\mathcal{X}$.

Setting

We study the solution of the linear ill-posed problem

$$\mathbf{A}\mathbf{x} = \mathbf{y}$$

with $\mathbf{A} \in \mathcal{L}(\mathcal{X}, \mathcal{Y})$ where $\mathcal{X}$ and $\mathcal{Y}$ are Hilbert spaces.
computation requires discretization, done via projections

$$P_m : \mathcal{Y} \to \mathbb{R}^m, \quad \mathbf{y} \mapsto y, \quad \text{e.g. point evaluation}$$
$$T_n : \mathcal{X} \to \mathbb{R}^n, \quad x = T_n\mathbf{x} = \{\langle \mathbf{x}, \psi_i \rangle\}_{i=1,\dots,n}$$

where $\{\psi_i\}_{i=1}^{\infty}$ is a basis in $\mathcal{X}$.
each component of y carries stochastic noise, $y^\sigma = y + \varepsilon$, $\varepsilon \sim \mathcal{N}(0, \sigma^2 I_m)$.

Setting

We study the solution of the linear ill-posed problem

$$\mathbf{A}\mathbf{x} = \mathbf{y}$$

with $\mathbf{A} \in \mathcal{L}(\mathcal{X}, \mathcal{Y})$ where $\mathcal{X}$ and $\mathcal{Y}$ are Hilbert spaces.
computation requires discretization, done via projections

$$P_m : \mathcal{Y} \to \mathbb{R}^m, \quad \mathbf{y} \mapsto y, \quad \text{e.g. point evaluation}$$
$$T_n : \mathcal{X} \to \mathbb{R}^n, \quad x = T_n\mathbf{x} = \{\langle \mathbf{x}, \psi_i \rangle\}_{i=1,\dots,n}$$

where $\{\psi_i\}_{i=1}^{\infty}$ is a basis in $\mathcal{X}$.
each component of y carries stochastic noise, $y^\sigma = y + \varepsilon$, $\varepsilon \sim \mathcal{N}(0, \sigma^2 I_m)$.
Define $A := P_m \mathbf{A} T_n^*$, then we want to find x s.t.

$$Ax = y^\sigma \tag{1}$$

We use Bayes' formula

to calculate the solution. In this framework, every quantity is treated as a random variable in a complete probability space $(\Omega, \mathcal{F}, \mathbb{P})$.

$$\pi_{post}(x|y^\sigma) = \frac{\pi_\varepsilon(y^\sigma|x)\pi_{pr}(x)}{\pi_{y^\sigma}(y^\sigma)}.$$

- $\pi_{post}(x|y^\sigma)$ posterior density
- $\pi_\varepsilon(y^\sigma|x)$ likelihood function
- $\pi_{pr}(x)$ prior distribution
- $\pi_{y^\sigma}(y^\sigma)$ data distribution (irrelevant)

We use Bayes' formula

to calculate the solution. In this framework, every quantity is treated as a random variable in a complete probability space $(\Omega, \mathcal{F}, \mathbb{P})$.

$$\pi_{post}(x|y^\sigma) = \frac{\pi_\varepsilon(y^\sigma|x)\pi_{pr}(x)}{\pi_{y^\sigma}(y^\sigma)}.$$

- $\pi_{post}(x|y^\sigma)$ posterior density
- $\pi_\varepsilon(y^\sigma|x)$ likelihood function
- $\pi_{pr}(x)$ prior distribution
- $\pi_{y^\sigma}(y^\sigma)$ data distribution (irrelevant)

gaussian error model:

$$\pi_\varepsilon \propto exp(-\frac{1}{2\sigma^2}||Ax - y^\sigma||^2),$$

Besov spaces

- We are looking for sparse reconstructions w.r.t. a basis in $\mathcal{X}$

Besov spaces

- We are looking for sparse reconstructions w.r.t. a basis in $\mathcal{X}$
- our choice: Besov-Space $B_p^s(\mathbb{R}^d)$ prior w.r.t. a wavelet expansion.
- Reason: such a prior is *discretization invariant* (Lassas, Saksman, Siltanen '09) and sparsity-promoting

Besov spaces

- We are looking for sparse reconstructions w.r.t. a basis in $\mathcal{X}$
- our choice: Besov-Space $B_p^s(\mathbb{R}^d)$ prior w.r.t. a wavelet expansion.
- Reason: such a prior is *discretization invariant* (Lassas, Saksman, Siltanen '09) and sparsity-promoting
- Let $\{\psi_\lambda : \lambda \in \Lambda\}$ be a wavelet system. Then $\mathbf{x} \in B_p^s(\mathbb{R}^d) \subset L^2(\mathbb{R}^d)$ if

$$||\mathbf{x}||_{s,p} := \left(\sum_{\lambda \in \Lambda} 2^{\varsigma p |\lambda|} |\langle \mathbf{x}, \psi_\lambda \rangle|^p \right)^{1/p} < \infty$$

and $\varsigma = s + d(\frac{1}{2} - \frac{1}{p}) \geq 0$. We focus on $1 \leq p \leq 2$.

Besov-space random variables

Definition (adapted from Lassas/Saksman/Siltanen, 2009)

Let $1 \leq p < \infty$ and $s \in \mathbb{R}$. Let $(X_\lambda^\alpha)_{\lambda \in \Lambda}$ be independent identically distributed real-valued random variables with probability density function

$$\pi_X(\tau) = c_p^\alpha \exp(-\frac{\alpha|\tau|^p}{2}), \quad c_p^\alpha = \left(\frac{\alpha}{2}\right)^{\frac{1}{p}} \frac{p}{2\Gamma(\frac{1}{p})}, \quad \tau \in \mathbb{R}.$$

Let U be the random function

$$U(t) = \sum_{\lambda \in \Lambda} 2^{-\varsigma|\lambda|} X_\lambda^\alpha \psi_\lambda(t), \quad t \in \mathbb{R}^d.$$

Then we say U is distributed according to a B_p^s-prior, $U \sim c \exp(-\frac{\alpha}{2}||U||_{s,p}^p)$.

"Problem": $\mathbb{P}(U \in B_p^s(\mathbb{R}^d)) = 0$ (of course $T_n^* T_n U \in B_p^s(\mathbb{R}^d)$)

"Problem": $\mathbb{P}(U \in B_p^s(\mathbb{R}^d)) = 0$ (of course $T_n^* T_n U \in B_p^s(\mathbb{R}^d)$)

Theorem (adapted from Lassas/Saksman/Siltanen, 2009)

Let U be as before, $2 < \alpha < \infty$ and take $r \in \mathbb{R}$. Then the following three conditions are equivalent:

(i) $||U||_{B_p^r(\mathbb{R}^d)} < \infty$ *almost surely,*

(ii) $\mathcal{E} \exp\left(||U||^p_{B_p^r(\mathbb{R}^d)}\right) < \infty$,

(iii) $r < s - \frac{d}{p}$.

finnish group considered $\mathbb{T}^d$, here: $\mathbb{R}^d$ but compactly supported (sparsity assumption)

Recall

$$\pi_{post}(x|y^\sigma) = \frac{\pi_{pr}(x)\pi_\varepsilon(y^\sigma|x)}{\pi_{y^\sigma}(y^\sigma)}.$$

$\pi_\varepsilon(y^\sigma|x)$ gaussian noise, $\pi_{pr}(x)$ Besov-Space prior
$\Rightarrow \pi_{post}(x|y^\sigma) \propto \exp(-\frac{1}{2\sigma^2}||Ax - y^\sigma||^2) \cdot \exp(-\frac{\alpha}{2}||T_n^* x||_{\cdot,p}^p)$

Recall

$$\pi_{post}(x|y^\sigma) = \frac{\pi_{pr}(x)\pi_\varepsilon(y^\sigma|x)}{\pi_{y^\sigma}(y^\sigma)}.$$

$\pi_\varepsilon(y^\sigma|x)$ gaussian noise, $\pi_{pr}(x)$ Besov-Space prior
$\Rightarrow \pi_{post}(x|y^\sigma) \propto \exp(-\frac{1}{2\sigma^2}||Ax - y^\sigma||^2) \cdot \exp(-\frac{\alpha}{2}||T_n^* x||_{\cdot,p}^p)$

we are interested in the *maximum a-priori* solution

$$x_{\text{map}} = \underset{x\in\mathbb{R}^n}{\text{argmax}} \quad \pi_{post}(x|y^\sigma)$$

or equivalently

$$x_\alpha^{\text{map}} = \underset{x\in\mathbb{R}^n}{\text{argmin}} \quad ||Ax - y^\sigma||^2 + \alpha\sigma^2 ||T_n^* x||_{B_p^s(\mathbb{R}^d)}^p \qquad (2)$$

Recall

$$\pi_{post}(x|y^\sigma) = \frac{\pi_{pr}(x)\pi_\varepsilon(y^\sigma|x)}{\pi_{y^\sigma}(y^\sigma)}.$$

$\pi_\varepsilon(y^\sigma|x)$ gaussian noise, $\pi_{pr}(x)$ Besov-Space prior
$\Rightarrow \pi_{post}(x|y^\sigma) \propto \exp(-\frac{1}{2\sigma^2}||Ax - y^\sigma||^2) \cdot \exp(-\frac{\alpha}{2}||T_n^* x||^p_{\cdot,p})$

we are interested in the *maximum a-priori* solution

$$x_{\text{map}} = \underset{x\in\mathbb{R}^n}{\text{argmax}} \quad \pi_{post}(x|y^\sigma)$$

or equivalently

$$x_\alpha^{\text{map}} = \underset{x\in\mathbb{R}^n}{\text{argmin}} \quad ||Ax - y^\sigma||^2 + \underbrace{\alpha\sigma^2}_{\hat{\alpha}}||T_n^* x||^p_{B^s_p(\mathbb{R}^d)} \qquad (2)$$

same functional as in deterministic case, but $||y - y^\sigma|| \leq \delta$ does not hold

Iteration [Daubechies, De Mol, Defrise 2004]:

$$x_{k+1} = \mathcal{S}_{\mathbf{w},p}\left(x_k + A^*(y^\sigma - Ax_k)\right), \qquad k = 1, 2, \ldots,$$

where $\mathcal{S}_{\mathbf{w},p}(h) := \sum_{\lambda\in\Lambda} S_{w_\lambda,p}(\langle h, \psi_\lambda\rangle)\psi_\lambda$ is defined component-wise via

$$S_{w,p}(\xi) := \left(\xi + \frac{wp}{2}\text{sign}(\xi)|\xi|^{p-1}\right)^{-1}$$

for the case $1 < p \leq 2$ and for $p = 1$

$$S_{\omega,1}(\xi) := \begin{cases} \xi - \frac{\omega}{2} & \text{if } \xi \geq \frac{\omega}{2} \\ 0 & \text{if } |\xi| < \frac{\omega}{2} \\ \xi + \frac{\omega}{2} & \text{if } \xi \leq -\frac{\omega}{2} \end{cases}.$$

Iteration [Daubechies, De Mol, Defrise 2004]:

$$x_{k+1} = \mathcal{S}_{\mathbf{w},p}\left(x_k + A^*(y^\sigma - Ax_k)\right), \qquad k = 1, 2, \dots,$$

where $\mathcal{S}_{\mathbf{w},p}(h) := \sum_{\lambda\in\Lambda} S_{w_\lambda,p}(\langle h, \psi_\lambda\rangle)\psi_\lambda$ is defined component-wise via

$$S_{w,p}(\xi) := \left(\xi + \frac{wp}{2}\text{sign}(\xi)|\xi|^{p-1}\right)^{-1}$$

for the case $1 < p \leq 2$ and for $p = 1$

$$S_{\omega,1}(\xi) := \begin{cases} \xi - \frac{\omega}{2} & \text{if } \xi \geq \frac{\omega}{2} \\ 0 & \text{if } |\xi| < \frac{\omega}{2} \\ \xi + \frac{\omega}{2} & \text{if } \xi \leq -\frac{\omega}{2} \end{cases}.$$

here: weights w_λ depend on smoothness parameters s and r

stochastic setting requires different measure for convergence

stochastic setting requires different measure for convergence
convergence in expectation may be too strict

stochastic setting requires different measure for convergence
convergence in expectation may be too strict
instead, we use the *Ky Fan metric*

Definition

Let x_1 and x_2 be random variables in a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with values in a metric space (χ, d_χ). The distance between x_1 and x_2 in the *Ky Fan metric* is defined as

$$\rho_K(x_1, x_2) := \inf\{\epsilon > 0 : \mathbb{P}(d_\chi(x_1(\omega), x_2(\omega)) > \epsilon) < \epsilon\}.$$

Theorem (Neubauer, Pikkarainen, 2008)

Let ξ be a random variable with values in $\mathbb{R}^m$. Assume that the distribution of ξ is $\mathcal{N}(y_0, \sigma^2 I)$ with $\sigma > 0$. Then it holds in $(\mathbb{R}^m, ||\cdot||)$ that

$$\rho_K(\xi, y_0) \leq \min\left\{1, \sqrt{2}\sigma\sqrt{m - \ln^-\left(\sigma^2 2\pi m^2 \left(\frac{e}{2}\right)^m\right)}\right\},$$

where $f^-(h) := \min\{0, f(h)\}$.

Theorem (Neubauer, Pikkarainen, 2008)

Let ξ be a random variable with values in $\mathbb{R}^m$. Assume that the distribution of ξ is $\mathcal{N}(y_0, \sigma^2 I)$ with $\sigma > 0$. Then it holds in $(\mathbb{R}^m, ||\cdot||)$ that

$$\rho_K(\xi, y_0) \leq \min\left\{1, \sqrt{2}\sigma\sqrt{m - \ln^-\left(\sigma^2 2\pi m^2 \left(\frac{e}{2}\right)^m\right)}\right\},$$

where $f^-(h) := \min\{0, f(h)\}$.

in practice $\ln$-term mostly inactive, for simplicity

$$\rho_K(\xi, y_0) \leq \min\left\{1, \sqrt{2}\sigma\sqrt{m}\right\},$$

c.f. $\mathcal{E}(||\xi - y_0||) = \sigma\sqrt{m}$

Overview

- Introduction
- Convergence results
- A numerical example

Let $x^\dagger$ be the unique solution of the equation $Ax = y$ with minimum value of $\Phi(\cdot)$.

Theorem (adapted from Hofinger, '06)

Let $\alpha, \sigma > 0$, $1 \leq p \leq 2$, $||A|| < 1$ *and* $N(A) = 0$ *for* $p = 1$. *Let* x_α^{map} *be the solution of (2). If* $\alpha = \alpha(\sigma)$ *is chosen such that* $\hat{\alpha} = \alpha\sigma^2 \to 0$ *and* $\frac{|\ln \sigma|}{\alpha} \to 0$ *as* $\sigma \to 0$, *then*

$$\lim_{\sigma \to 0} \rho_K(x_\alpha^{map}, x^\dagger) = 0.$$

Let $x^\dagger$ be the unique solution of the equation $Ax = y$ with minimum value of $\Phi(\cdot)$.

Theorem (adapted from Hofinger, '06)

Let $\alpha, \sigma > 0$, $1 \le p \le 2$, $||A|| < 1$ and $N(A) = 0$ for $p = 1$. Let x_α^{map} be the solution of (2). If $\alpha = \alpha(\sigma)$ is chosen such that $\hat{\alpha} = \alpha\sigma^2 \to 0$ and $\frac{|\ln \sigma|}{\alpha} \to 0$ as $\sigma \to 0$, then

$$\lim_{\sigma \to 0} \rho_K(x_\alpha^{map}, x^\dagger) = 0.$$

main idea: use Ky Fan metric and split $\Omega = \Omega_{\text{det}}(\sigma) \cup \Omega_{\text{unbound}}(\sigma)$

deterministic convergence rate, DDD '04

Assume $\mathbf{A}$ fulfils, for all $h \in L^2$

$$A_l^2 \sum_\lambda 2^{-2|\lambda|\beta} |\langle h, \psi_\lambda \rangle|^2 \leq ||Ah||^2 \leq A_u^2 \sum_\lambda 2^{-2|\lambda|\beta} |\langle h, \psi_\lambda \rangle|^2. \quad (3)$$

Set $\eta := \frac{\varsigma}{\beta+\varsigma}$ and $\eta' := \frac{\beta}{\beta+\varsigma}$. Then

$$\begin{aligned}\sup\{||\mathbf{x}_\alpha^{\text{map}} - \mathbf{x}|| : \mathbf{x} \in \mathcal{X}, \mathbf{y} \in \mathcal{Y}, ||\mathbf{A}\mathbf{x} - \mathbf{y}|| \leq \delta, ||\mathbf{x}||_{s,p} \leq \varrho\} \\ < C \left(\frac{\delta + \delta'}{A_l} \right)^\eta (\varrho + \varrho')^{\eta'}\end{aligned}$$

with $\delta' = (\delta^2 + \hat{\alpha}\varrho^p)^{\frac{1}{2}}$ and $\varrho' = (\varrho^p + \frac{\delta^2}{\hat{\alpha}})^{\frac{1}{p}}$.

especially, if

$$||Ax^\dagger - y^\sigma|| \le \delta \text{ and } ||x||_{s,p} \le \varrho, \text{ then}$$
$$||x_\alpha^{\mathsf{map}} - x^\dagger|| < \mathcal{C}\left(\delta + \delta'\right)^\eta (\varrho + \varrho')^{\eta'}, \text{ or}$$

$$\mathbb{P}(\{\omega \in \Omega : ||x_{\hat\alpha}^{\mathsf{map}}(\omega) - x^\dagger(\omega)|| > C(\delta + \delta')^\eta(\varrho + \varrho')^{\eta'}\})$$
$$\le \mathbb{P}(\{\omega : ||Ax^\dagger(\omega) - y^\sigma(\omega)|| > \delta\}) + \mathbb{P}(\{\omega : ||T^* x^\dagger(\omega)||_{s,p} \ge \varrho\})$$

especially, if

$$||Ax^\dagger - y^\sigma|| \leq \delta \text{ and } ||x||_{s,p} \leq \varrho, \text{ then}$$
$$||x_\alpha^{\text{map}} - x^\dagger|| < \mathcal{C}\left(\delta + \delta'\right)^\eta (\varrho + \varrho')^{\eta'}, \text{ or}$$

$$\mathbb{P}(\{\omega \in \Omega : ||x_{\hat{\alpha}}^{\text{map}}(\omega) - x^\dagger(\omega)|| > C(\delta + \delta')^\eta (\varrho + \varrho')^{\eta'}\})$$
$$\leq \mathbb{P}(\{\omega : ||Ax^\dagger(\omega) - y^\sigma(\omega)|| > \delta\}) + \mathbb{P}(\{\omega : ||T^* x^\dagger(\omega)||_{s,p} \geq \varrho\})$$

$$\mathbb{P}(||Ax^\dagger(\omega) - y^\sigma(\omega)|| > \delta||) = \frac{\Gamma(\frac{m}{2}, \frac{\delta^2}{2\sigma^2})}{\Gamma(\frac{m}{2})},$$

$$\mathbb{P}(||T_n^* x||_{s,p} > \varrho||) = \frac{\Gamma(\frac{n}{p}, \frac{\alpha\varrho^p}{2})}{\Gamma(\frac{n}{p})} \text{ for } \mathbf{x} \sim B_p^s(\mathbb{R}^d)$$

here: $\delta = \sqrt{2}\sigma\sqrt{m - \ln^-\left(\sigma^2 2\pi m^2 \left(\frac{e}{2}\right)^m\right)}$

Theorem

Let A fulfil (3) and assume that we have an a-priori estimate $||T_n^* x||_{s,p} \leq \varrho$ (x measured in prior norm!) for some $\varrho > 0$. Set $a_m := \ln\left(\frac{2^m}{2\pi m^2}\right)$. Then as $\sigma \to 0$, x_α^{map} converges with the parameter choice $\alpha = \alpha(\sigma, \varrho, \beta, \varsigma, p, m, n)$ fulfilling

$$\left(2\sqrt{2}\sigma\sqrt{a_m - 2\ln\sigma + \frac{\alpha\varrho^p}{2}}\right)^{\frac{\varsigma}{\beta+\varsigma}} \left(\varrho + \left(\varrho^p + \frac{2}{\alpha}(a_m - 2\ln\sigma)\right)^{1/p}\right)^{\frac{\beta}{\beta+\varsigma}} = \frac{\Gamma(\frac{m}{2}, m)}{\Gamma(\frac{m}{2})} + \frac{\Gamma(\frac{n}{p}, \frac{\alpha\varrho^p}{2})}{\Gamma(\frac{n}{p})}$$

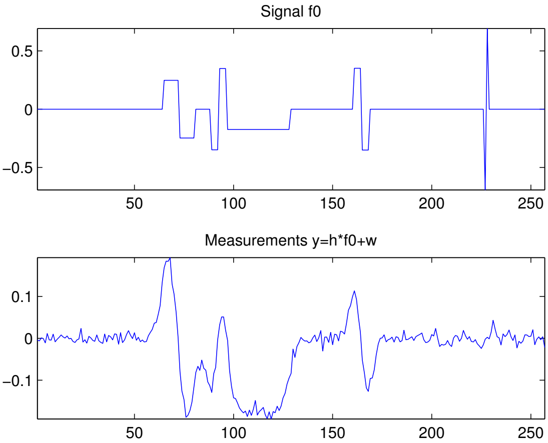
to the unique solution $x^\dagger$ and

$$\rho_K(x_\alpha^{map}, x^\dagger) = \mathcal{O}\left(\left(\sigma\sqrt{1 + |\ln\sigma| + \alpha}\right)^{\frac{\varsigma}{\beta+\varsigma}} \left(\varrho + \left(\varrho^p + \frac{1+|\ln\sigma|}{\alpha}\right)^{1/p}\right)^{\frac{\beta}{\beta+\varsigma}}\right).$$

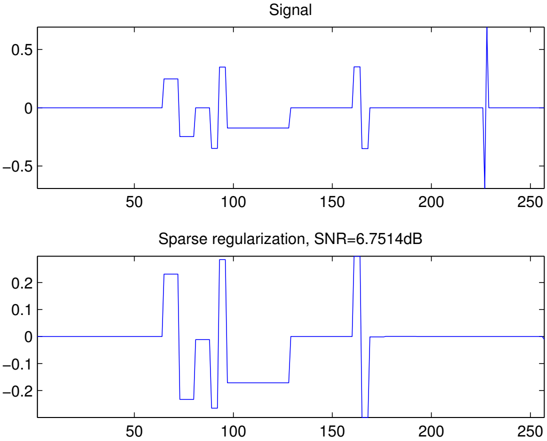
Overview

1D deconvolution

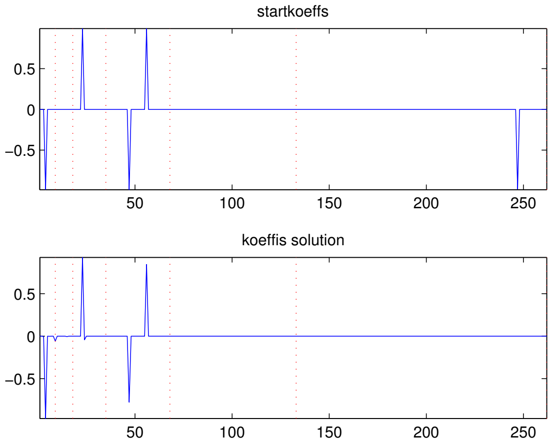
$p = 1$, $\sigma = 0.01$, $m = 128$, $n = 154$, signal measurements:



reconstruction, $\varrho = 0.1$, $\alpha = 3183.2$:



coefficients comparison:



2D deconvolution

$p = 1$, $\sigma = 0.001$, $m = 65536$, $n = 72334$, image and measurements:

reconstructions, x distributed according to $B_1^{3.01}$-prior, $\varrho = 0.1$, $\alpha = 1.4 \cdot 10^6$

measured in $B_1^{3.01}$-norm

measured in B_1^{1}-norm

- M. Lassas, E. Saksman, S. Siltanen, *Discretization-invariant Bayesian inversion and Besov space priors*, Inverse Probl. Imaging, 2009.
- I. Daubechies, M. Defrise, C. De Mol, *An iterative thresholding algorithm for linear inverse problems with a sparsity constraint*, Comm. Pure Appl. Math. 57, 2004.
- A. Hofinger, *Ill-posed problems: Extending the Deterministic Theory to a Stochastic Setup*, Thesis, JKU Linz, Austria, 2006.
- A. Neubauer, H.K. Pikkarainen, *Convergence results for the Bayesian inversion theory*, J. Inverse Ill-Posed Probl. 16, 2008.
- J. Kaipio, E. Somersalo, *Statistical and computational inverse problems*, Springer-Verlag, New York, 2005.

Thank you for attention! Are there questions?