A Stochastic Convergence Analysis for Tikhonov-Regularization with Sparsity Constraints

D. Gerth, R. Ramlau

IFIP TC7, September 2013, Klagenfurt

Doctoral Program
Computational Mathematics
Numerical Analysis and Symbolic Computation

JKU
JOHANNES KEPLER
UNIVERSITY LINZ

Overview

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?

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?

- 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.
- Basic deterministic model:

$$||\mathbf{A}\mathbf{x} - \mathbf{y}^\delta||^2 + \hat{\alpha}\Phi_{\mathbf{w},p}(\mathbf{x}) \to \min_{\mathbf{x}} \tag{1}$$

- Penalty $\Phi_{\mathbf{w},p}(\mathbf{x}) = \sum_{\lambda \in \Lambda} w_\lambda |\langle \mathbf{x}, \psi_\lambda \rangle|^p$ for an ONB $\{\psi_\lambda\}$

- computation requires discretization, done via projections

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

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

- computation requires discretization, done via projections

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

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

- computation requires discretization, done via projections

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

  where $\{\psi_i\}_{i=1}^{\infty}$ is ONB 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.

$$Ax = y^\sigma \tag{2}$$

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 $\mathcal{X}$
- our choice: Besov-Space $B^s_{p,q}(\mathbb{R}^d)$ prior (notation: $B^s_p(\mathbb{R}^d)$)
- Reasons:
  - "easy" characterization with coefficients of a wavelet expansion
  - sparsity-promoting properties
  - discretization invariance (Lassas, Saksman, Siltanen '09)

Definition of the wavelet expansion

Let $\phi \in C^{\tilde{s}}(\mathbb{R})$ be a scaling function and $\psi \in C^{\tilde{s}}(\mathbb{R})$ a compactly supported wavelet, $\tilde{s} > s$ and define

$$\phi_{j,k}(t) = 2^{\frac{j}{2}}\phi(2^j t - k),$$
$$\psi_{j,k}(t) = 2^{\frac{j}{2}}\psi(2^j t - k), \quad j,k \in \mathbb{Z}.$$

such that the collection $\{\psi_{j,k} : j,k \in \mathbb{Z}\}$ constitutes an orthonormal basis of $L_2(\mathbb{R})$ and for each $\mathbf{h} \in L_2(\mathbb{R})$ we have

$$\mathbf{h} = \sum_{k=-\infty}^{\infty} \langle \mathbf{h}, \phi_{0,k}\rangle \phi_{0,k} + \sum_{j=0}^{\infty}\sum_{k=-\infty}^{\infty} \langle \mathbf{h}, \psi_{j,k}\rangle \psi_{j,k}.$$

Definition of the wavelet expansion

Let $\phi \in C^{\tilde{s}}(\mathbb{R})$ be a scaling function and $\psi \in C^{\tilde{s}}(\mathbb{R})$ a compactly supported wavelet, $\tilde{s} > s$ and define

$$\phi_{j,k}(t) = 2^{\frac{j}{2}}\phi(2^j t - k),$$
$$\psi_{j,k}(t) = 2^{\frac{j}{2}}\psi(2^j t - k), \quad j,k \in \mathbb{Z}.$$

such that the collection $\{\psi_{j,k} : j,k \in \mathbb{Z}\}$ constitutes an orthonormal basis of $L_2(\mathbb{R})$ and for each $\mathbf{h} \in L_2(\mathbb{R})$ we have

$$\mathbf{h} = \sum_{k=-\infty}^{\infty} \langle \mathbf{h}, \phi_{0,k}\rangle \phi_{0,k} + \sum_{j=0}^{\infty}\sum_{k=-\infty}^{\infty} \langle \mathbf{h}, \psi_{j,k}\rangle \psi_{j,k}.$$

assumption:
only finitely many nonzero scalar products on each scale

- with this, a d-dimensional basis can be constructed
- let $\{\psi_\lambda : \lambda \in \Lambda\}$ denote the set of all wavelets ψ, also including the scaling functions ϕ where Λ is an appropriate index set, possibly infinite
- set $|\lambda| = j$, then
- $\mathbf{x} \in B_p^s(\mathbb{R}^d) \subset L^2(\mathbb{R}^d)$, $s < \tilde{s}$, if

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

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 U be the random function

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

where the coefficients $(X_\lambda^\alpha)_{\lambda \in \Lambda}$ are independent identically distributed real-valued random variables with probability density function

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

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

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

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

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

Let U be a B_p^s-random function defined 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) $\mathbb{E} \exp\left(||U||^p_{B_p^r(\mathbb{R}^d)}\right) < \infty$,

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

same result as [LSS 2009], but here $\mathbb{R}^d$ instead of $\mathbb{T}^d$ considered

How to avoid this phenomenon?

- “finite model” (MI)

- “infinite model” (MII)

How to avoid this phenomenon?

- "finite model" (MI)
  - consider discretization level $n < \infty$ fixed, finite index set Λ_n
  - Then

$$X_n(t) := \sum_{\lambda \in \Lambda_n} 2^{-\varsigma|\lambda|} X_\lambda^\alpha \psi_\lambda(t) \Rightarrow ||X_n||_{B_p^s}^p = \sum_{\lambda \in \Lambda_n} |X_\lambda^\alpha|^p < \infty$$

  - and $\mathbb{P}(||X_n||_{B_p^s} > \varrho) = \frac{\Gamma(\frac{n}{p}, \frac{\alpha \varrho^p}{2})}{\Gamma(\frac{n}{p})} \leq \frac{1}{\varrho} \sqrt[p]{\frac{2n}{\alpha p}}$
- "infinite model" (MII)

How to avoid this phenomenon?

- "finite model" (MI)
  - consider discretization level $n < \infty$ fixed, finite index set Λ_n
  - Then

$$X_n(t) := \sum_{\lambda \in \Lambda_n} 2^{-\varsigma|\lambda|} X_\lambda^\alpha \psi_\lambda(t) \Rightarrow ||X_n||_{B_p^s}^p = \sum_{\lambda \in \Lambda_n} |X_\lambda^\alpha|^p < \infty$$

  - and $\mathbb{P}(||X_n||_{B_p^s} > \varrho) = \frac{\Gamma(\frac{n}{p}, \frac{\alpha \varrho^p}{2})}{\Gamma(\frac{n}{p})} \leq \frac{1}{\varrho} \sqrt[p]{\frac{2n}{\alpha p}}$
- "infinite model" (MII)
  - forget about discretization level n
  - define $X(t)$ in $B_p^r(\mathbb{R}^d)$ with $s < r - \frac{d}{p}$, then
  - $\mathbb{E}(||X||_{B_p^s(\mathbb{R}^d)}) = \left(\frac{2}{\alpha p}\left(c_\lambda^1 + c_\lambda^2 \sum_{j=0}^\infty 2^{-j((r-s)p-d)}\right)\right)^{\frac{1}{p}} < \infty$
  - and $\mathbb{P}(||X||_{B_p^s} > \varrho) \leq \frac{1}{\varrho}\mathbb{E}(||X||_{B_p^s(\mathbb{R}^d)})$

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
$\Rightarrow \pi_{post}(x|y^\sigma) \propto \exp(-\frac{1}{2\sigma^2}||Ax - y^\sigma||^2) \cdot \exp(-\frac{\alpha}{2}||T_n^* x||_{B_p^s}^p)$

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
$\Rightarrow \pi_{post}(x|y^\sigma) \propto \exp(-\frac{1}{2\sigma^2}||Ax - y^\sigma||^2) \cdot \exp(-\frac{\alpha}{2}||T_n^* x||_{B_p^s}^p)$

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

$$x_\alpha^{\text{map}} = \underset{x\in\mathbb{R}^n}{\text{argmin}} \quad ||Ax - y^\sigma||^2 + \alpha\sigma^2 ||T_n^* x||_{B_p^s(\mathbb{R}^d)}^p \tag{3}$$

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
$\Rightarrow \pi_{post}(x|y^\sigma) \propto \exp(-\frac{1}{2\sigma^2}||Ax - y^\sigma||^2) \cdot \exp(-\frac{\alpha}{2}||T_n^* x||_{B_p^s}^p)$

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

$$x_\alpha^{\text{map}} = \underset{x\in\mathbb{R}^n}{\text{argmin}} \quad ||Ax - y^\sigma||^2 + \underbrace{\alpha\sigma^2}_{\hat{\alpha}}||T_n^* x||_{B_p^s(\mathbb{R}^d)}^p \qquad (3)$$

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

- 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

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

- 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

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

- allows combination of deterministic and stochastic quantities

- 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

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

- allows combination of deterministic and stochastic quantities
- metric for convergence in probability

Ky Fan error estimate

Theorem (Neubauer, Pikkarainen, 2008)

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

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

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

Ky Fan error estimate

Theorem (Neubauer, Pikkarainen, 2008)

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

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

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

in practice $\ln$-term mostly inactive, then

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

c.f. $\mathbb{E}(||y^\sigma - y||) \approx \sigma\sqrt{m}$

Overview

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

Theorem (adapted from Hofinger, '06)

Let $\alpha, \sigma > 0$, $1 \leq p \leq 2$ and $N(A) = 0$ for $p = 1$. Assume $w_\lambda \geq c > 0 \; \forall \lambda \in \Lambda$. Let x_α^{map} be the solution of (3). 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 $\alpha, \sigma > 0$, $1 \leq p \leq 2$ *and* $N(A) = 0$ *for* $p = 1$. *Assume* $w_\lambda \geq c > 0 \; \forall \lambda \in \Lambda$. *Let* x_α^{map} *be the solution of (3). 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*

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

as long as $\sigma^2 2\pi m^2 \left(\frac{e}{2}\right)^m > 1$, then $\alpha \to \infty$ is sufficient

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

Theorem (adapted from Hofinger, '06)

Let $\alpha, \sigma > 0$, $1 \leq p \leq 2$ and $N(A) = 0$ for $p = 1$. Assume $w_\lambda \geq c > 0\ \forall \lambda \in \Lambda$. Let x_α^{map} be the solution of (3). 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

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

as long as $\sigma^2 2\pi m^2 \left(\frac{e}{2}\right)^m > 1$, then $\alpha \to \infty$ is sufficient

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

Overview

Deterministic Convergence Rate, Daubechies et al, 2004

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

$$A_l^2 \sum_\lambda 2^{-2|\lambda|\beta} |\langle h, \psi_\lambda \rangle|^2 \leq ||Ah||^2 \leq A_u^2 \sum_\lambda 2^{-2|\lambda|\beta} |\langle h, \psi_\lambda \rangle|^2. \quad (4)$$

and $||\mathbf{x}^\dagger||_{B_p^s} \leq \varrho$, $\varrho > 0$. Then

$$\sup\{||\mathbf{x}_\alpha^{\text{map}} - \mathbf{x}|| : \mathbf{x} \in \mathcal{X}, \mathbf{y} \in \mathcal{Y}, ||\mathbf{Ax} - \mathbf{y}|| \leq \delta, ||\mathbf{x}||_{B_p^s} \leq \varrho\} < C \left(\frac{\delta + \delta'}{A_l} \right)^{\frac{\varsigma}{\beta+\varsigma}} (\varrho + \varrho')^{\frac{\beta}{\beta+\varsigma}}$$

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

Theorem

Let all previous assumptions hold. 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*

$$f(\alpha) := \min\left\{1, 2\left(\frac{\sqrt{2}}{A_l}\sigma\sqrt{a_m - 2\ln\sigma + \frac{\alpha\varrho^p}{2}}\right)^{\frac{\varsigma}{\beta+\varsigma}}\left(\left(\varrho^p + \frac{2}{\alpha}(a_m - 2\ln\sigma)\right)^{1/p}\right)^{\frac{\beta}{\beta+\varsigma}}\right\} - \frac{\Gamma(\frac{m}{2}, m)}{\Gamma(\frac{m}{2})} - \mathbb{P}(||x.||_{B_p^s} > \varrho) = 0$$

to the unique solution $x^\dagger$ *and*

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

where $\mathbb{P}(||x_n||_{B_p^s} > \varrho) = \frac{\Gamma(\frac{n}{p}, \frac{\alpha\varrho^p}{2})}{\Gamma(\frac{n}{p})}$ or $\mathbb{P}(||x||_{B_p^s} > \varrho) = \frac{\mathbb{E}||x||_{B_p^s}}{\varrho}$

Overview

- Introduction
- A Convergence Theorem
- Convergence Rates
- Numerical Examples

- We consider a convolution problem

$$[Ax](s) = [k * x](s) = \int_{\mathbb{R}^d} k(s-t)x(t)dt, \quad s \in \mathbb{R}^d \qquad (5)$$

- using a kernel

$$\widehat{k}(\xi) = \frac{c_{\kappa,\beta}}{(1+\kappa|\xi|^2)^{\beta/2}}, \quad \xi \in \mathbb{R}^d, \quad c_{\kappa,\beta} \text{ s.t. } ||\widehat{k}||_{L_2(\mathbb{R}^d)} < 1$$

- thus (4) is fulfilled with chosen β
- $p = 1$, $d = 1$

Iteration [Daubechies, De Mol, Defrise 2004]:

With $x_0 = 0$,

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

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 ($p = 1$) via

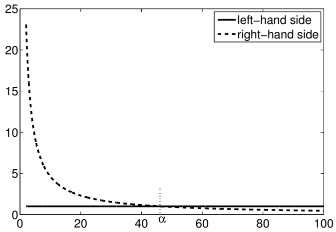
$$S_{\omega,1}(\xi) := \begin{cases} \xi - \frac{\omega}{2} & \text{if } \xi \geq \frac{\omega}{2} \\ 0 & \text{if } |\xi| < \frac{\omega}{2} \\ \xi + \frac{\omega}{2} & \text{if } \xi \leq -\frac{\omega}{2} \end{cases}.$$

converges since $||A|| < 1$

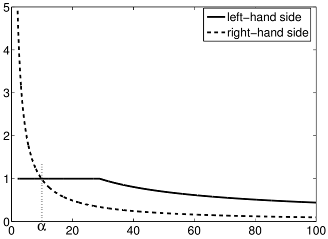
Parameter choice rule illustrated

$$\sigma = 0.01,\ m = 2500,\ \varsigma = 0.5,\ \beta = 1,\ \varrho = 2.16$$

model (MI), $s = 1$



model (MII), $s = 1$, $r = 2$



example of a solution

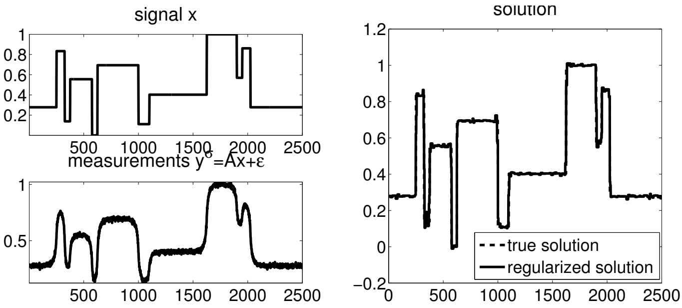


Figure : (MI), $\sigma = 0.01$, exact ϱ, $s = 1$, $\beta = 1$. $\alpha = 45.85$
$\Rightarrow \hat{\alpha} = \alpha\sigma^2 = 0.004585$

comparison of (MI) and (MII), m, n fixed, $\sigma \to 0$

all plots averaged over 20 individual simulations

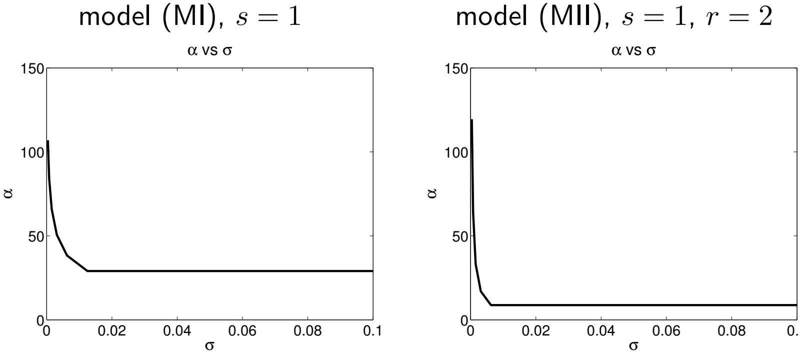


Figure : α plotted against σ, $n = m = 2500$, $\beta = 1$, exact ϱ

comparison of (MI) and (MII), m, n fixed, $\sigma \to 0$

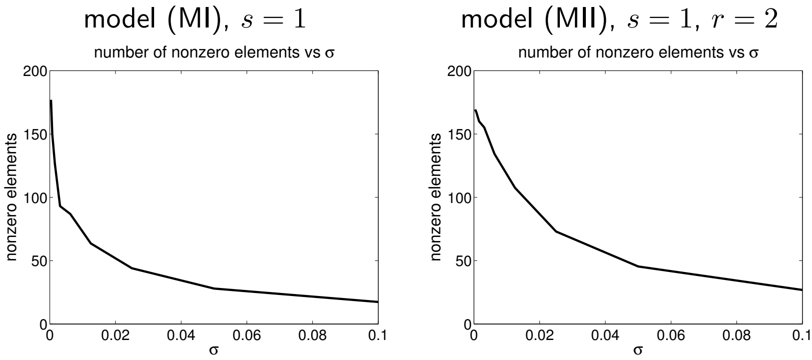


Figure : number of recovered nonzero coefficients plotted against σ, $n = m = 2500$, $\beta = 1$, exact ϱ

comparison of (MI) and (MII), m, n fixed, $\sigma \to 0$

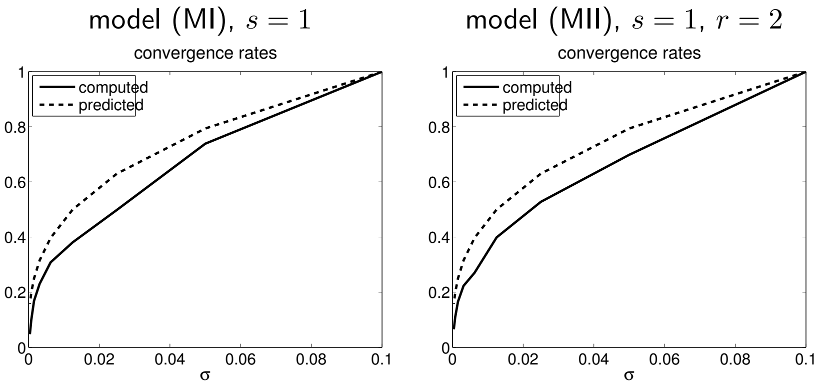


Figure : predicted and observed convergence rates plotted against σ, $n = m = 2500$, $\beta = 1$, exact ϱ

comparison of (MI) and (MII), σ fixed, m, n variable

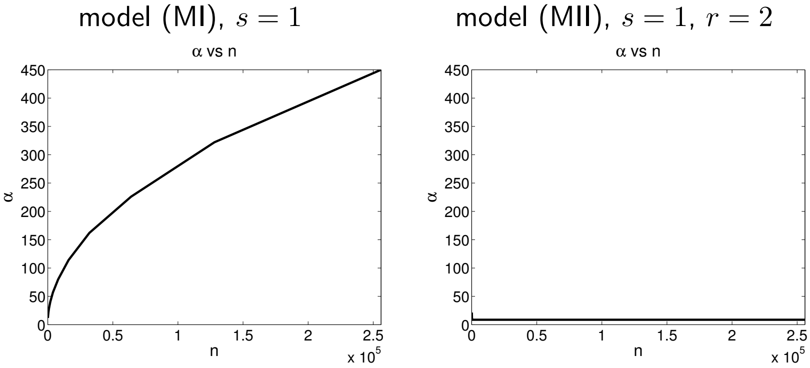
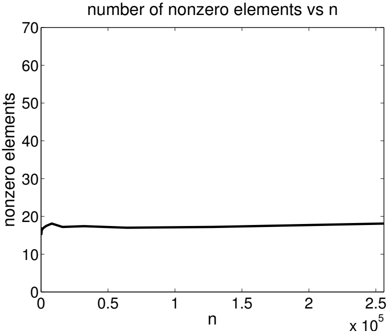


Figure : α plotted against n, $\sigma = 0.01$, $\beta = 1$, exact ϱ

comparison of (MI) and (MII), σ fixed, m, n variable

model (MI), $s = 1$



model (MII), $s = 1$, $r = 2$

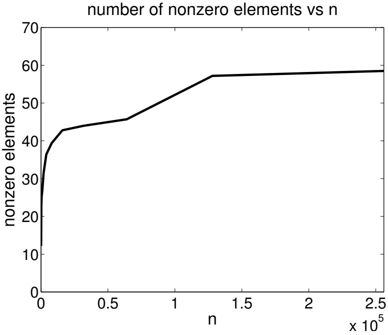


Figure : number of recovered nonzeros plotted against n, $\sigma = 0.01$, $\beta = 1$, exact ϱ

comparison of (MI) and (MII), σ fixed, m, n variable

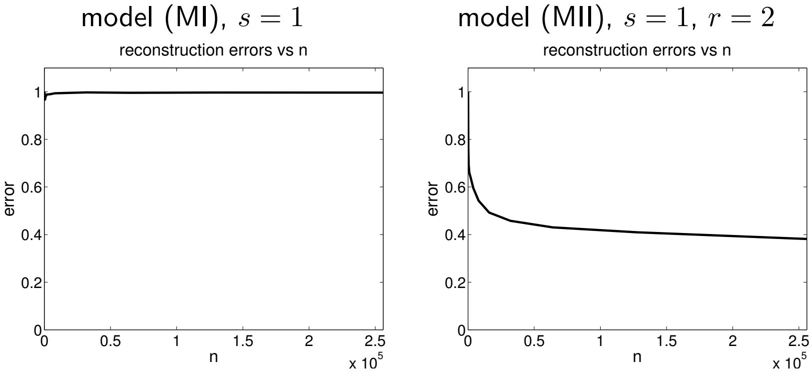


Figure : reconstruction error plotted against n, $\sigma = 0.01$, $\beta = 1$, exact ϱ

A 2D convolution example

$\sigma = 0.1$, $\beta = 1$, $\alpha = 130.5$, $\hat{\alpha} = 1.3$

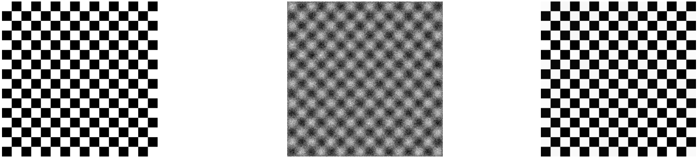


Figure : true solution - measurements - recovered solution

exactly the 68 original coefficients were reconstructed

- G., R. Ramlau, *A stochastic convergence analysis for Tikhonov-Regularization with sparsity constraints*, submitted
- 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. 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?