A Stochastic Convergence Analysis for Tikhonov-Regularization with Sparsity Constraints

Daniel Gerth, Ronny Ramlau

Sparse Tomo Days, Lyngby, Denmark
28.03.14

Doctoral Program
Computational Mathematics
Numerical Analysis and Symbolic Computation

JKU
JOHANNES KEPLER UNIVERSITY LINZ

Overview

- Introduction
- Bayesian approach
- Convergence theorem
- Convergence rates
- Numerical examples

- 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
- we seek solutions x which are sparse w.r.t to a given ONB
- the observed data is assumed to be noisy
- 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 $\{\psi_\lambda\}$

noise modelling

- two different approaches

deterministic		stochastic
worst case error		stochastic information
$\|\|\mathbf{y}^\delta - \mathbf{y}\|\| \leq \delta$		e.g. $y^\sigma \sim \mathcal{N}(y, \sigma^2)$,
		$\mathbb{E}(\|\|y^\sigma - y\|\|) = f(\sigma)$,...
$\vdots$		$\vdots$
"easy"	analysis	"hard"
"fast"	algorithms	"slow"
δ hard to get	parameters	σ easy to get
	$\Leftarrow ? \Rightarrow$	

noise modelling

- two different approaches

deterministic		stochastic
worst case error		stochastic information
$\|\|\mathbf{y}^\delta - \mathbf{y}\|\| \leq \delta$		e.g. $y^\sigma \sim \mathcal{N}(y, \sigma^2)$,
		$\mathbb{E}(\|\|y^\sigma - y\|\|) = f(\sigma)$,...
$\vdots$		$\vdots$
"easy"	analysis	"hard"
"fast"	algorithms	"slow"
δ hard to get	parameters	σ easy to get
	$\Leftarrow ? \Rightarrow$	

- We want to combine the advantages and find links between both branches. Question: Can we prove convergence (rates) for sparsity regularization, if we use an explicit stochastic noise model instead of the worst case error?

- stochastic noise model based on discretization, also 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 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.

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

Overview

We use Bayes' formula

to characterize 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 characterize 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:

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

Now we need a prior

Besov spaces

- We are looking for sparse reconstructions w.r.t. a basis in $\mathcal{X}$
- our choice: Besov-space $B^s_{p,p}(\mathbb{R}^d)$ prior

  Reasons:
  - "easy" characterization with coefficients of a wavelet expansion
  - sparsity-promoting properties known, connection to TV regularization
  - discretization invariance (Lassas, Saksman, Siltanen '09), avoiding the following phenomena:
    - solutions diverge as $m \to \infty$
    - solutions diverge as $n \to \infty$
    - Representation of a-priori knowledge is incompatible with discretization (this is the case, e.g., for a TV prior)

- we consider a wavelet basis suitable for multi resolution analysis
- let $\{\psi_\lambda : \lambda \in \Lambda\}$ denote the set of all wavelets ψ, also including the scaling functions where Λ is an appropriate index set, possibly infinite
- set $|\lambda| = j$, then
- $\mathbf{x} \in B^s_{p,p}(\mathbb{R}^d) \subset L^2(\mathbb{R}^d)$, $s < \tilde{s}$, if

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

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 X be the random function

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

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

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

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

"Problem": $\mathbb{P}(X \in B^s_{p,p}(\mathbb{R}^d)) = 0$

"Problem": $\mathbb{P}(X \in B^s_{p,p}(\mathbb{R}^d)) = 0$

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

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

(i) $||X||_{B^r_{p,p}(\mathbb{R}^d)} < \infty$ *almost surely,*

(ii) $\mathbb{E} \exp\left(||X||^p_{B^r_{p,p}(\mathbb{R}^d)} \right) < \infty$,

(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 m and n fixed, finite index set Λ_n
  - Then

$$X_n(t) := \sum_{\lambda \in \Lambda_n} 2^{-\varsigma|\lambda|} X_\lambda^\alpha \psi_\lambda(t) \Rightarrow ||X_n||^p_{B^s_{p,p}(\mathbb{R}^d)} = \sum_{\lambda \in \Lambda_n} |X_\lambda^\alpha|^p < \infty$$

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

How to avoid this phenomenon?

- "finite model" (MI)
  - consider discretization level m and n fixed, finite index set Λ_n
  - Then

$$X_n(t) := \sum_{\lambda \in \Lambda_n} 2^{-\varsigma|\lambda|} X_\lambda^\alpha \psi_\lambda(t) \Rightarrow ||X_n||^p_{B^s_{p,p}(\mathbb{R}^d)} = \sum_{\lambda \in \Lambda_n} |X_\lambda^\alpha|^p < \infty$$

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

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
$\Rightarrow \pi_{post}(x|y^\sigma) \propto \exp(-\frac{1}{2\sigma^2}||Ax - y^\sigma||^2) \cdot \exp(-\frac{\alpha}{2}||x||^p_{B^s_{p,p}(\mathbb{R}^d)})$

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

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

or equivalently

$$x_\alpha^{\mathsf{map}} = \underset{x\in\mathbb{R}^n}{\mathrm{argmin}} \quad ||Ax - y^\sigma||^2 + \alpha\sigma^2||x||^p_{B^s_p(\mathbb{R}^d)} \tag{3}$$

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
$\Rightarrow \pi_{post}(x|y^\sigma) \propto \exp(-\frac{1}{2\sigma^2}||Ax - y^\sigma||^2) \cdot \exp(-\frac{\alpha}{2}||x||^p_{B^s_{p,p}(\mathbb{R}^d)})$

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

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

or equivalently

$$x_\alpha^{\mathsf{map}} = \underset{x\in\mathbb{R}^n}{\mathrm{argmin}} \quad ||Ax - y^\sigma||^2 + \underbrace{\alpha\sigma^2}_{\hat{\alpha}}||x||^p_{B^s_p(\mathbb{R}^d)} \tag{3}$$

same functional as in deterministic case, but we only know
$\mathbb{E}(||y - y^\sigma||) = f(\sigma)$

- stochastic setting requires different measure for convergence
- 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
- 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. $\mathbb{E}(||y^\sigma - y||^2) = \sigma^2 m$

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 $\alpha, \sigma > 0$ and (3) have a unique minimizer. Let x_α^{map} be this solution. 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)

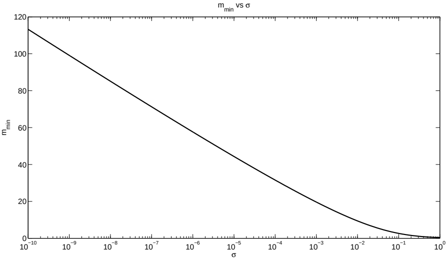
Let $\alpha, \sigma > 0$ *and* (3) *have a unique minimizer. Let* x_α^{map} *be this solution. If* $\alpha = \alpha(\sigma)$ *is chosen such that* $\hat{\alpha} = \alpha\sigma^2 \to 0$ *and* $\frac{|\ln \sigma|}{\alpha} \to 0$ *as* $\sigma \to 0$, *then*

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

Uniqueness:

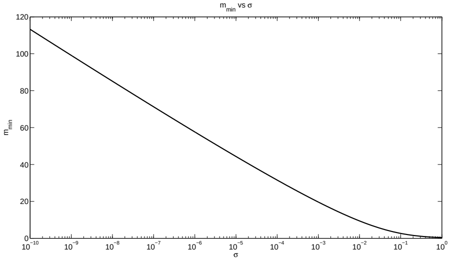
- $p > 1$
- A injective
- A injective on any finite linear subspace

Discussion



- as long as $\sigma^2 2\pi m^2 \left(\frac{e}{2}\right)^m > 1$, then $\alpha \to \infty$ is sufficient
- $\frac{1}{\alpha}$ corresponds to variance of the prior
- main idea for the proof: use Ky Fan metric and split $\Omega = \Omega_{\text{det}}(\sigma) \cup \Omega_{\text{unbound}}(\sigma)$

Discussion



- as long as $\sigma^2 2\pi m^2 \left(\frac{e}{2}\right)^m > 1$, then $\alpha \to \infty$ is sufficient
- $\frac{1}{\alpha}$ corresponds to variance of the prior
- main idea for the proof: use Ky Fan metric and split $\Omega = \Omega_{\text{det}}(\sigma) \cup \Omega_{\text{unbound}}(\sigma)$

The condition $\alpha \to \infty$ strange from a Bayesian perspective. To explain the discrepancy, it has to be interpreted relative to σ.

almost sure convergence

- convergence in probability implies convergence a.s. of subsequences
- we can identify such subsequences

Theorem (D.G.)

Let m, n *fixed and* $\{\sigma_k\}_{k=1}^{\infty}$ *be such that*

$$\sum_{k=1}^{\infty} \rho_k(y, y^{\sigma_k}) = \sum_{k=1}^{\infty} \sqrt{2}\sigma_k \sqrt{m - \ln^{-}\left(\sigma_k^2 2\pi m^2 \left(\frac{e}{2}\right)^m\right)} < \infty$$

then

$$x_{\alpha(\sigma_k)}^{\textit{map}} \overset{\text{a.s.}}{\to} x^{\dagger}$$

- convergence a.s. allows no quantitative estimates

Overview

deterministic convergence rate, DDD '04

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

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

and $||\mathbf{x}^\dagger||_{B^s_{p,p}(\mathbb{R}^d)} \leq \varrho$, $\varrho > 0$. Then

$$\sup\{||\mathbf{x}_\alpha^{\text{map}} - \mathbf{x}|| : \mathbf{x} \in \mathcal{X}, \mathbf{y} \in \mathcal{Y}, ||\mathbf{A}\mathbf{x} - \mathbf{y}|| \leq \delta, ||\mathbf{x}||_{B^s_{p,p}(\mathbb{R}^d)} \leq \varrho\} < C \left(\frac{\delta + \delta'}{A_l} \right)^{\frac{\varsigma}{\beta+\varsigma}} (\varrho + \varrho')^{\frac{\beta}{\beta+\varsigma}}$$

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

using Ky-Fan

in particular, if

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

$$\mathbb{P}(\{\omega \in \Omega : ||x_{\hat{\alpha}}^{\text{map}}(\omega) - x^{\dagger}(\omega)|| > C(\delta + \delta')^{\eta}(\varrho + \varrho')^{\eta'}\})$$
$$\leq \mathbb{P}(\{\omega : ||Ax^{\dagger}(\omega) - y^{\sigma}(\omega)|| > \delta\}) + \mathbb{P}(\{\omega : ||T^{*}x^{\dagger}(\omega)||_{B^s_{p,p}(\mathbb{R}^d)} \geq \varrho\}) \tag{5}$$

using Ky-Fan

in particular, if

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

$$\begin{aligned}&\mathbb{P}(\{\omega \in \Omega : ||x^{\mathsf{map}}_{\hat{\alpha}}(\omega) - x^\dagger(\omega)|| > C(\delta+\delta')^\eta(\varrho+\varrho')^{\eta'}\}) \\ &\le \mathbb{P}(\{\omega : ||Ax^\dagger(\omega) - y^\sigma(\omega)|| > \delta\}) + \mathbb{P}(\{\omega : ||T^* x^\dagger(\omega)||_{B^s_{p,p}(\mathbb{R}^d)} \ge \varrho\})\end{aligned} \tag{5}$$

compare with definition of the Ky-Fan-metric:

$$\rho_K(x^{\mathsf{map}}_{\hat{\alpha}}, x^\dagger) := \inf\{\epsilon > 0 : \mathbb{P}(||x^{\mathsf{map}}_{\hat{\alpha}} - x^\dagger|| > \epsilon) < \epsilon\}$$

$\Rightarrow$ balance terms in (5),use $\delta = \sqrt{2}\sigma\sqrt{m - \ln^-\left(\sigma^2 2\pi m^2 \left(\frac{e}{2}\right)^m\right)}$

Convergence rates, simplified

Theorem (D.G.)

Let all previous assumptions hold. Then there exists an explicit parameter choice rule

$$\alpha = \alpha(\sigma, \varrho, \beta, \varsigma, p, m, n)$$

depending also on the choice of model (I) or (II), such that $x_\alpha^{\mathit{map}} \to x^\dagger$ *and*

$$\rho_K(x_\alpha^{\mathit{map}}, x^\dagger) = \mathcal{O}\left(f(\alpha, \sigma, \varrho, \beta, \varsigma, p, m, n)\right)$$

both f and α are known

Theorem (D.G.)

Let A fulfil (4) and assume that we have an a-priori estimate $||x^\dagger||_{B^s_{p,p}(\mathbb{R}^d)} \leq \varrho$ 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

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

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

Overview

- We consider a convolution problem

$$[Ax](s) = [k * x](s) = \int_{\mathbb{R}^d} k(s-t)x(t)dt, \quad s \in \mathbb{R}^d \qquad (6)$$

- using a kernel

$$\widehat{k}(\xi) = \frac{c_{\kappa,\beta}}{(1+\kappa|\xi|^2)^{\beta/2}}, \quad \xi \in \mathbb{R}^d, \quad c_{\kappa,\beta} \text{ s.t. } ||k||_{L_1(\mathbb{R}^d)} < 1$$

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

Iteration [Daubechies, De Mol, Defrise 2004]:

With $x_0 = 0$,

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

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 ($p = 1$) via

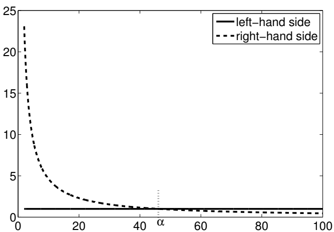
$$S_{\omega,1}(\xi) := \begin{cases} \xi - \frac{\omega}{2} & \text{if } \xi \geq \frac{\omega}{2} \\ 0 & \text{if } |\xi| < \frac{\omega}{2} \\ \xi + \frac{\omega}{2} & \text{if } \xi \leq -\frac{\omega}{2} \end{cases}.$$

converges since $||A|| < 1$

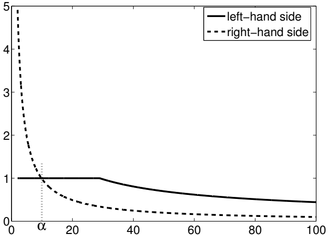
Parameter choice rule illustrated

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

model (MI), $s = 1$



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



example of a solution

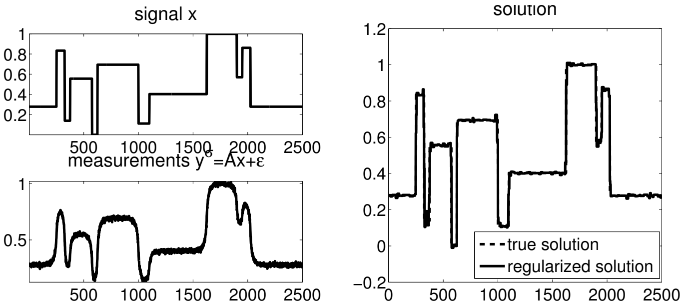
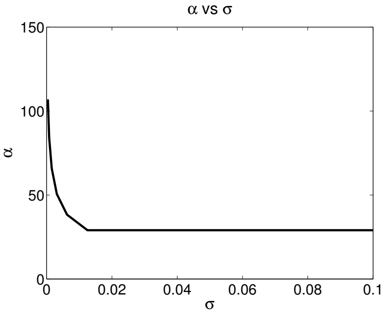


Figure : (MI), $\sigma = 0.01$, exact ϱ, $s = 1$, $\beta = 1$. $\alpha = 45.85$ according to our parameter choice rule $\Rightarrow \hat{\alpha} = \alpha\sigma^2 = 0.004585$

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

all plots averaged over 20 individual simulations

model (MI), $s = 1$



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

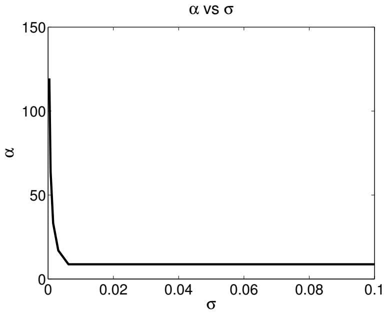


Figure : α plotted against σ, $n = m = 2500$, $\beta = 1$, exact ϱ

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

all plots averaged over 20 individual simulations

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

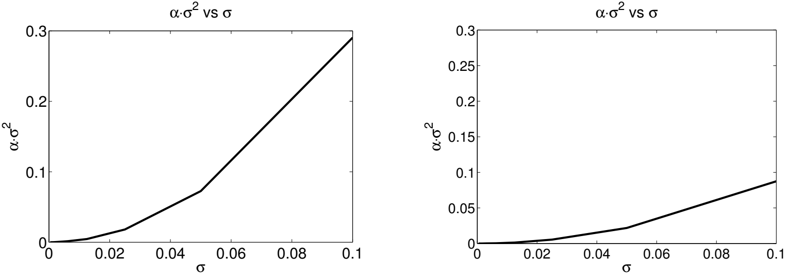
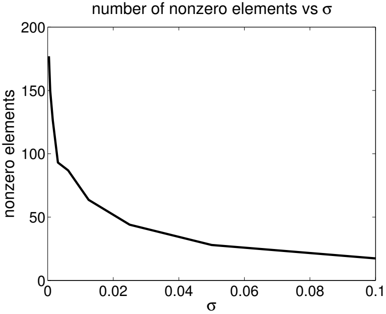


Figure : $\alpha \cdot \sigma^2$ plotted against σ, $n = m = 2500$, $\beta = 1$, exact ϱ

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

model (MI), $s = 1$



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

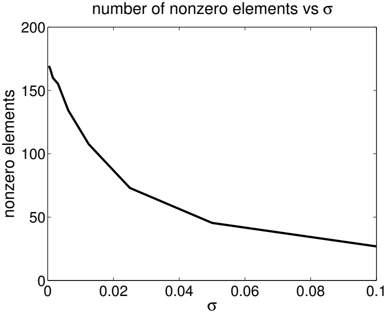


Figure : number of recovered nonzero coefficients plotted against σ, $n = m = 2500$, $\beta = 1$, exact ϱ

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

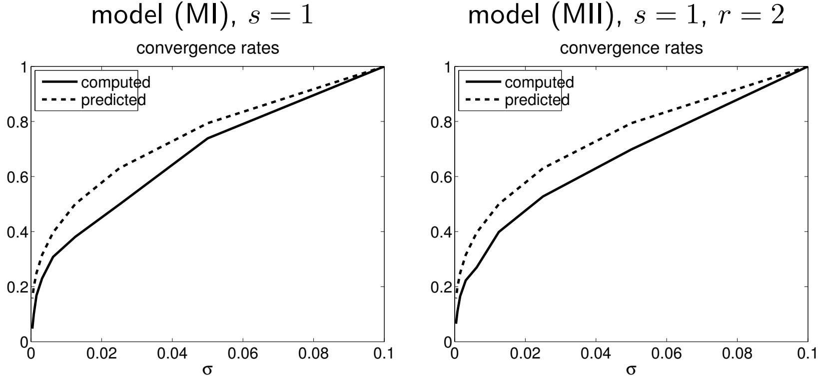


Figure : predicted and observed convergence rates plotted against σ, $n = m = 2500$, $\beta = 1$, exact ϱ

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

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

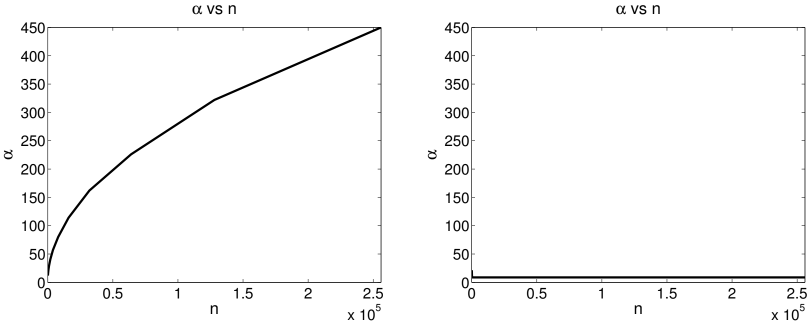


Figure : α plotted against n, $\sigma = 0.01$, $\beta = 1$, exact ϱ

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

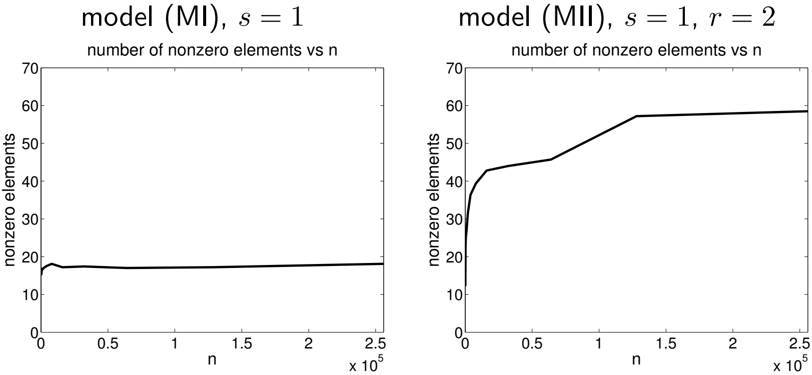


Figure : number of recovered nonzeros plotted against n, $\sigma = 0.01$, $\beta = 1$, exact ϱ

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

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

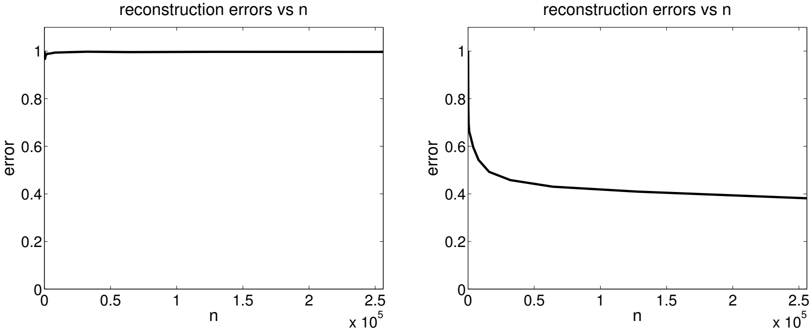


Figure : reconstruction error plotted against n, $\sigma = 0.01$, $\beta = 1$, exact ϱ

A 2D convolution example

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

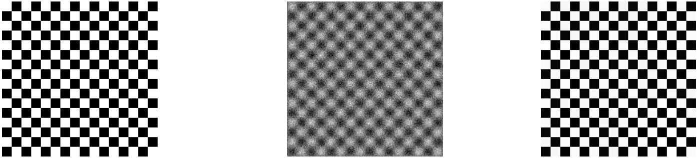


Figure : true solution - measurements - recovered solution

exactly the 68 original coefficients (out of 65536) 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. Hofinger, H.K. Pikkarainen, *Convergence Rates for Linear Inverse Problems in the Presence of Additive Normal Noise*, Stoch. Anal. and Appl. 27:2, 2009.
- J. Kaipio, E. Somersalo, *Statistical and computational inverse problems*, Springer-Verlag, New York, 2005.

Thank you for attention! Are there questions?