Convergence analysis for nonlinear Tikhonov regularization in Hilbert scales with adaptive choice of the regularization parameter

Mihaela Pricop-Jeckstadt

University of Bonn

Bonn, 27 May 2014

Outline

1. Statistical inverse problems
2. Balancing principle in Hilbert scales
3. Applications

universität bonn

Abstract noise model

$$Y = u + \delta\xi + \sigma\epsilon, \; u = F(a^\dagger)$$

Mathé & Pereverzev, 2003
Bissantz & Hohage & Munk & Ruymgaart, 2007

- $\xi \in \mathcal{Y}$, $\|\xi\|_{\mathcal{Y}} = 1$ deterministic error
- ϵ is a Hilbert-space process with $\mathbf{E}\,\langle \epsilon, \phi \rangle_{\mathcal{Y}} = 0$, $\|\mathbf{cov}_\epsilon\| \leq 1$.

Remark

- White noise models occur as limits of discrete noise models as the sample size n tends to infinity.
- We have $\sigma \sim \frac{1}{\sqrt{n}}$.

Definitions

Definition

A continuous linear operator $\epsilon : \mathcal{Y} \to L^2(\Omega, \mathcal{K}, \mathbf{P})$ is called a Hilbert-space process.

The covariance $\mathbf{cov}_\epsilon : \mathcal{Y} \to \mathcal{Y}$ of ϵ is the bounded operator defined by $\langle \mathbf{cov}_\epsilon \phi_1, \phi_2 \rangle = \mathbf{Cov}(\langle \epsilon, \phi_1 \rangle, \langle \epsilon, \phi_2 \rangle)$, $\phi_1, \phi_2 \in \mathcal{Y}$.

Definition

ϵ is a white noise process if $\mathbf{cov}_\epsilon = I$.

A Gaussian white noise process in an infinite-dimensional Hilbert space can not be identified with a Hilbert-space valued random variable with finite second moment.

Aims in statistical inverse problems

1. Approximate the discontinuous operator F^{-1} by a family of continuous operators $\{R_\alpha : \alpha > 0\}$.
2. Choose a parameter choice rule $\alpha = \alpha(Y, \sigma)$ to obtain an estimate $\widehat{a} = R_{\alpha(Y,\sigma)}(Y)$.
3. Prove consistency for $\widehat{a}$ i.e.

$$\mathbf{E}\|\widehat{a} - a^\dagger\|_{\mathcal{X}}^2 \stackrel{\sigma\to 0}{\longrightarrow} 0$$

4. Compute rates of convergence under further a-priori information on the solution, e.g. that $a^\dagger$ belongs to a smoothness class $\mathcal{X}_q$.

2-step method for nonlinear inverse problems

- $F : \mathcal{X} \to \mathcal{Y}$ is a nonlinear, injective operator.
- An estimator $\widehat{u}$ of $u \in \mathcal{Y}$ is chosen, $\mathcal{Y}$ a Hilbert space, such that $\sqrt{\mathbf{E}\|\widehat{u} - u\|_{\mathcal{Y}}^2} \leq \tau$ with known τ.
- $\widehat{a} \in D(F)$ is the Tikhonov estimator of a:

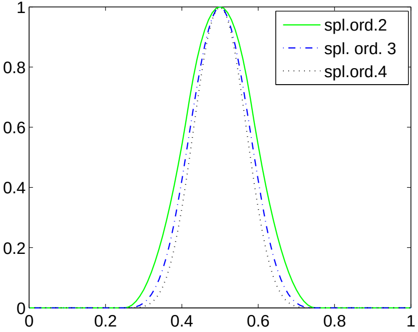
$$\widehat{a} := \operatorname{argmin}_{a \in D(F)} \{\|F(a) - \widehat{u}\|_{\mathcal{Y}}^2 + \alpha \|a - a_0\|_{\mathcal{X}}^2\}$$

- Tikhonov regularization corresponds to ridge regression for linear models in statistics.
- Bissantz & Hohage & Munk 2004

Convergence rates for nonlinear statistical inverse problems

- O'Sullivan 1990: first convergence rate result (suboptimal rates with restrictive assumptions)
- Bissantz & Hohage & Munk 2004: consistency and optimal rates for one smoothness class
- Hohage & Pricop 2008: optimal rates in a range of smoothness classes

Hilbert scales



$L : D(L) \to \mathcal{X}$ unbounded, selfadjoint, strictly positive

$D(L) \subset \mathcal{X}$ dense

$\mathcal{X}_s := D(L^s)$, $s \geq 0$

$\langle x, y \rangle_s := \langle L^s x, L^s y \rangle_{\mathcal{X}}, x, y \in \mathcal{X}_s$

Natterer 1984: Rates of convergence for deterministic linear inverse problems

$\widehat{a}$ is the solution of

$$\|F(a) - \widehat{u}\|_{\mathcal{Y}}^2 + \alpha\|a - a_0\|_s^2 \rightarrow \min, a \in D(F) \cap (a_0 + \mathcal{X}_s)$$

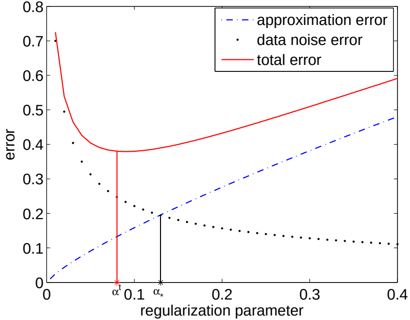
with $\widehat{u} = F(a^\dagger) + \delta\xi$, δ deterministic noise level und $\tau \in \mathcal{Y}$.

Assumptions

1. $D(F)$ is convex, F is continuous, injective, Fréchet-differentiable on $\mathcal{X}$ and weakly closed on $\mathcal{X}_s$ for some $s \geq 0$.
2. $\|F'(a^\dagger)h\|_{\mathcal{Y}} \sim \|h\|_{-p}, \forall h \in \mathcal{X}$, for some known $p > 0$.
3. There exists $L > 0$ such that $a \in D(F) \cap (a_0 + \mathcal{X}_s)$

$$\|F'(a^\dagger) - F'(a)\|_{\mathcal{Y} \leftarrow \mathcal{X}_{-p}} \leq L\|a^\dagger - a_0\|_0 \leq \frac{\lambda}{2\Lambda}.$$

Lepskiĭ choice of the regularization parameter



- Lepskiĭ 1990: adaptive choice of the regularization parameter for regression problems
- Mathé, Pereverzev 2003, 2006: the Lepskiĭ principle for linear inverse problems

Convergence for exact data

We use the error splitting $\|a^\dagger - \widehat{a}\| \leq \|a^\dagger - a_\alpha\| + \|a_\alpha - \widehat{a}\|$ where

$$a_\alpha := \operatorname{argmin}_{a \in D(F) \cap (a_0 + \mathcal{X}_s)} \left(\|F(a) - F(a^\dagger)\|^2_{\mathcal{Y}} + \alpha \|a - a_0\|^2_s \right).$$

Theorem

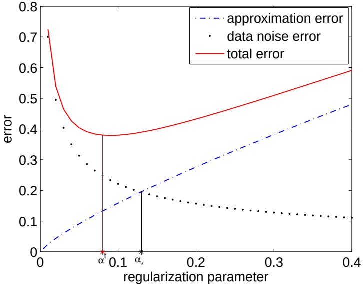
Let Assumptions $1-3$, $a^\dagger - a_0 \in \mathcal{X}_q$, $q \in [s, p+2s]$, $s \geq p$ *and a deterministic noise model hold. Then it holds*

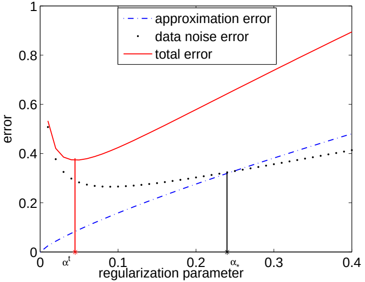
$$\|a_\alpha - a^\dagger\|_{\mathcal{X}} \leq C\alpha^{\frac{q}{2(p+s)}}$$

$$\|\widehat{a} - a_\alpha\|_{\mathcal{X}} \leq c(\delta\alpha^{\frac{-p}{2(p+s)}} + \alpha^{\frac{q}{2(p+s)}})$$

with the constants C *and* c *depending on* $a^\dagger, p, q, s$.

Balancing principle for deterministic nonlinear inverse problems





We choose $\alpha_j = \delta^2 (q^2)^{j-1}$, $q > 1, j = 1, \ldots, m$, denote $a_i = a_{\alpha_i}$ and determine $\alpha_+ = \alpha_{i_+}$ such that

$$i_+ = \max \left\{ i : \|a_i - a_j\| \le 4 C^* \delta \alpha_j^{\frac{-p}{2(s+p)}}, j = 1, 2, \ldots, i \right\}.$$

Balancing principle for deterministic nonlinear inverse problems

Theorem

Under the Assumptions $1-3$*, for deterministic noise model and for the choice of the regularization parameter* $\alpha = \alpha_+$*, the order-optimal error bound*

$$\|a_+ - a^\dagger\|_{\mathcal{X}} \leq 6C^* \delta^{\frac{q}{p+q}}$$

holds true, where $a_+ = a_{\alpha_+}$*.*

Shuai, Pereverzev, Ramlau 2007: the balancing principle for nonlinear inverse problems

Balancing principle for statistical nonlinear inverse problems

Let us assume a stochastic setting and choose

$$i_+ = \max\left\{ i : \|a_i - a_j\|_{\mathcal{X}} \leq 4C^*\tau \ln\frac{1}{\tau}\alpha_j^{\frac{-p}{2(s+p)}}, j = 1, 2, \ldots, i \right\}.$$

Theorem

If, besides the Assumptions $1-3$ *for stochastic setting, the probability distribution for the estimator* $\widehat{u}$ *fulfills the exponential inequality*

$$\mathbf{P}\left\{ \|\widehat{u} - \mathbf{E}\widehat{u}\|^2 \geq (t-1)\mathbf{E}\left(\|\widehat{u} - \mathbf{E}\widehat{u}\|^2\right)\right\} \leq c_1 \exp(-c_2 t)$$

for any $t > 1$ *and for a constant* $k > 0$*, then it holds*

$$\mathbf{E}(\|a_+ - a^\dagger\|_{\mathcal{X}}^2) \leq \frac{2qK}{p+q}\tau^{\frac{2q}{p+q}} \ln\frac{1}{\tau}.$$

universitätbonn

Parameter estimation as inverse problem

Direct problem

find u given a and f

$$\begin{cases} -u''(x) + a(x)u(x) = f(x) \\ u(0) = g_0, u(1) = g_1 \end{cases}$$

- $x \in (0, 1)$
- u is the population density of a biological species
- Malthus model the rate of change f linearly dependent on a population density u

Inverse Problem

estimate a from u given f and g

$$F : D(F) \to L^2(0, 1), F(a^\dagger) := u^\dagger$$

$$D(F) = \{a \in L^2(0, 1) : 0 \leq a \leq \gamma\}$$

- The inverse problem is the not so well understood model.
- For any $u^\dagger \in L^2(\Omega)$ there exists an unique $a^\dagger \in D(F)$.
- F^{-1} is a discontinuous operator $\to$ ill-posed problem

universitätbonn

Hilbert scale

$$\mathcal{X}_{-1} =:= \left\{ v \in L^2 : \int_0^1 v\,dx = 0 \right\},$$

$$\mathcal{X}_0 = H^1 \cap \left\{ v \in L^2 : \int_0^1 v\,dx = 0 \right\},$$

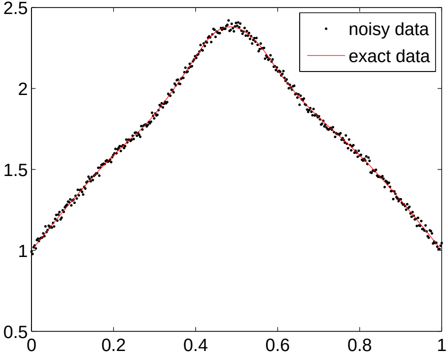
$$\mathcal{X}_1 = \left\{ u \in H^2 : u'(0) = u'(1) = 0, \int_0^1 u\,dx = 0 \right\},$$

$$\mathcal{X}_2 = H^3 \cap \mathcal{X}_1,$$

$$\mathcal{X}_3 = \{\phi \in H^4 \cap \mathcal{X}_1 : \phi'''(0) = \phi'''(1) = 0\}.$$

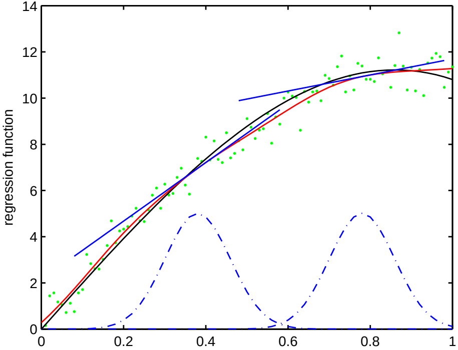
For fast rates of convergence the mean values of $a^\dagger$ and its odd derivatives at boundaries must be known a-priori. This a-priori knowledge must be incorporated in the initial guess a_0.
Verification of assumptions for F: Hohage & Pricop 2008

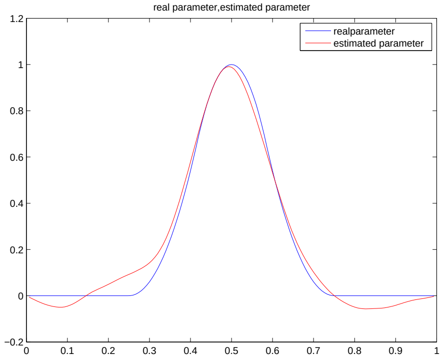
Noise model



Data: $(X_1, Y_1) \cdots (X_n, Y_n)$
$\{X_1, \ldots, X_n\}$ fixed design in $[0, 1]$
Regression model
$Y_i = u(X_i) + \varepsilon_i, \; i = 1, .., n$
with errors ε_i i.i.d. with
$\mathbf{E}(\varepsilon_i) = 0$ and
$\mathbf{var}(\varepsilon_i) = 0.01^2$, $n = 398$

Parameter reconstruction





$C^3(0,1)$, grid $m = 100$
Gauss. kernel, bandwith by CV

is spline of order 2
$s = 2$, $p = 2$, $q = 2.5$

universitätbonn

Some references

- Mair B. A., Ruymgaart F. H.:
  Statistical Inverse Estimation in Hilbert Scales
  SIAM J. Appl. Math. 56 **56**, *1424- 1444 (1996)*
- Mathé, Peter and Pereverzev, Sergei V.:
  Geometry of linear ill-posed problems in variable Hilbert scales
  Inverse Problems **19**, *789- 803 (2003)*
- Bissantz N., Hohage T., Munk A.:
  Consistency and rates of convergence of nonlinear Tikhonov regularization with random noise
  Inverse Problems **20**, *1753- 1771 (2004)*
- Hohage T., Pricop M.
  Nonlinear Tikhonov regularization in Hilbert scales for inverse boundary value problems with random noise
  Inverse Problems and Imaging **2**, *271-290 (2008)*

Thank you for your attention!