Convergence analysis of balancing principle for nonlinear Tikhonov regularization in Hilbert scales for statistical inverse problems

Mihaela Pricop-Jeckstadt

University of Bonn

Cambridge, 26 March 2014
1 Statistical inverse problems

2 Balancing principle in Hilbert scales

3 Applications
Inverse problems and metagenomics

Definition
"Solution of an inverse problem entails determining unknown causes based on observation of their effects."
A. Tikhonov (~ 1963)

- Inverse problems are often encountered in the mechanistical approach in biology.
- The interactions that occur during the ripening of smear cheeses
  Jérôme Mounier, Christophe Monnet et al, Appl. Env. Microbiology, 2008
- The effect of an antibiotic on the intestinal microbiome from time-series data
  Richard R. Stein, Vanni Bucci, et.al, Comp.Biology, 2013
Inverse problem: basic example

Numerical Differentiation

- Differentiation and Integration are inverse operations.
- **Direct Problem = Integration:** For any \( a \in C([0,1]) \) we define

\[
F(a)(x) = \int_0^x a(s) \, ds, \quad x \in [0,1].
\]

- **Inverse Problem = Differentiation:** Compute \( a^\dagger = u^\dagger' \) from known \( u^\dagger \in C^1([0,1]) \) with \( u^\dagger(0) = 0 \).

- The existence and stability of the solution \( a^\dagger \in C([0,1]) \) hold for \( u^\dagger \in C^1([0,1]) \) with the norm \( C^1([0,1]) \) but not in \( \| \cdot \|_{\infty} \).

- The inverse problem is usually ill-posed i.e. the inverse of the operator \( F \) is not continuous.
We illustrate the instability in $\| \cdot \|_\infty$ by the noisy data

$$u_\delta^n(x) := u^\dagger(x) + \delta \sin\left(\frac{nx}{\delta}\right), \ x \in [0, 1].$$

Error in the data:

$$\|u_\delta^n - u^\dagger\|_\infty \leq \delta \to 0 \text{ mit } \delta \to 0.$$

Error in the solution:

$$\|a_\delta^n - a^\dagger\|_\infty = \|(u_\delta^n)' - u^\dagger'\|_\infty = n \to \infty \text{ mit } n \to \infty.$$
Symmetric difference quotient as a continuous approximation of the $F^{-1}$:

$$u^\dagger'(x) \approx \frac{u^\dagger(x+\alpha)-u^\dagger(x-\alpha)}{2\alpha} =: (R_\alpha u^\dagger)(x)$$

Apply $R_\alpha$ on noisy data $u_\delta^\dagger$

$$u^\dagger'(x) \approx \frac{u_\delta^\dagger(x+\alpha)-u_\delta^\dagger(x-\alpha)}{2\alpha} =: (R_\alpha u_\delta^\dagger)(x)$$

Given: noisy values $u_i^\delta = u_n^\delta(x_i)$ on a grid $x_i \in [0, 1]$ with $|u_i^\delta - u^\dagger(x_i)| \leq \delta, i = 0, 1, 2, \ldots, n$
Exact and noisy data

Exact data and $\alpha = 0.1$

Exact data and $\alpha = 0.01$

1% noise level and $\alpha = 0.1$

1% noise level and $\alpha = 0.01$
Reconstruction error:

\[ |(R_\alpha u_n^\delta)(x) - u^\dagger'(x)| \leq |(R_\alpha u^\dagger)(x) - u^\dagger'(x)| + |(R_\alpha u_n^\delta)(x) - (R_\alpha u^\dagger)(x)| \]

Approximation error: \[ |(R_\alpha u^\dagger)(x) - u^\dagger'(x)| \xrightarrow{\alpha \to 0} 0 \]

Data error: \[ |(R_\alpha u_n^\delta)(x) - (R_\alpha u^\dagger)(x)| \leq \alpha^{-1} \delta \xrightarrow{\alpha \to 0} 0 \]

\( R_\alpha \) with a choice of the regularization parameter \( \alpha = \alpha(\delta) \) such that \( \alpha(\delta) \xrightarrow{\delta \to 0} 0 \) and \( \frac{\delta}{\alpha(\delta)} \xrightarrow{\delta \to 0} 0 \) is a regularization method.

Convergence rates are computable only when a-priori information (smoothness of \( a^\dagger \)) is available

\[ |(R_\alpha u^\dagger)(x) - u^\dagger'(x)| \leq \frac{\alpha^2}{6} \max_{x \in [0,1]} |u^{'''}(x)| \text{ if } a^\dagger \in C^2([0, 1]) \]

\( \alpha_{opt}(\delta) \sim \delta^{\frac{1}{3}} \) and \( \| R_{\alpha_{opt}(\delta)} u_n^\delta - u^\dagger' \|_{\infty} = O(\delta^{\frac{2}{3}}) \).
Statistical inverse problems

Direct Regression

\[ Y_i = a^\dagger(X_i) + \sigma \epsilon_i \]

kernel, projection or local polynomial estimators

Design: \((X_1, \ldots, X_n)\) i.i.d., uniformly distributed on \([0, 1]\)
(stochastic) or \(X_i = i/n, i = 1 \ldots, n\) (deterministic)

Error: \((\epsilon_i)\) random variables, \(\mathbb{E}(\epsilon_i) = 0, \mathbb{E}(\epsilon_i^2) < \infty\)

Inverse Regression

\[ Y_i = (Fa^\dagger)(X_i) + \sigma \epsilon_i \]

\(F\) is a possibly nonlinear, injective operator with discontinuous inverse

Aim: estimate \(a^\dagger\) in a nonparametric class of functions from
\((X_1, Y_1), \ldots, (X_n, Y_n)\) by \(\hat{a}: [0, 1] \rightarrow \mathcal{R}\) and study the quality of
the estimator.
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 \( \hat{a} = R_\alpha(Y, \sigma)(Y) \).

3. Prove consistency for \( \hat{a} \) i.e.

\[
\mathbb{E} \| \hat{a} - a^\dagger \|_{\mathcal{X}}^2 \xrightarrow{\sigma \to 0} 0
\]

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

- \( F : \mathcal{X} \rightarrow \mathcal{Y} \) is a nonlinear, injective operator.

- An estimator \( \hat{u} \) of \( u \in \mathcal{Y} \) is chosen, \( \mathcal{Y} \) a Hilbert space, such that \( \sqrt{\mathbb{E}\|\hat{u} - u\|^2_\mathcal{Y}} \leq \tau \) with known \( \tau \).

- \( \hat{a} \in D(F) \) is the Tikhonov estimator of \( a \):
  \[
  \hat{a} := \arg\min_{a \in D(F)} \{ \|F(a) - \hat{u}\|^2_\mathcal{Y} + \alpha \|a - a_0\|^2_\mathcal{X} \}
  \]

- 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
Tikhonov regularization in Hilbert scales

Nonlinear Inverse Problems

\( \hat{a} \) is the solution of

\[
\| F(a) - \hat{u} \|_Y^2 + \alpha \| a - a_0 \|_s^2 \rightarrow \min, \quad a \in D(F) \cap (a_0 + \mathcal{X}_s)
\]

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. \( \| F'(a^\dagger) h \|_Y \sim \| h \|_\mathcal{X}_p, \forall h \in \mathcal{X}, \) for some known \( p > 0 \).

3. There exists \( L > 0 \) such that \( a \in D(F) \cap (a_0 + \mathcal{X}_s) \)

\[
\| F'(a^\dagger) - F'(a) \|_{\mathcal{Y} \leftarrow \mathcal{X}_p} \leq L \| a^\dagger - a_0 \|_0 \leq \frac{\lambda}{2\Lambda}.
\]
Lepskiï choice of the regularization parameter

• Lepskiï 1990: adaptive choice of the regularization parameter for regression problems
• Mathé, Pereverzev 2003, 2006: the Lepskiï principle for linear inverse problems
Convergence for exact data

We use the error splitting \( \|a^\dagger - \hat{a}\| \leq \|a^\dagger - a_\alpha\| + \|a_\alpha - \hat{a}\| \) where

\[
a_\alpha := \arg\min_{a \in D(F) \cap (a_0 + \mathcal{X}_s)} \left( \|F(a) - F(a^\dagger)\|^2_Y + \alpha \|a - a_0\|^2_s \right).
\]

**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

\[
\|a_\alpha - a^\dagger\|_X \leq C \alpha^{\frac{q}{2(p+s)}}
\]

\[
\|
\hat{a} - a_\alpha\|_X \leq c \left( \delta \alpha^{\frac{-p}{2(p+s)}} + \alpha^{\frac{q}{2(p+s)}} \right)
\]

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

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

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

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

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

**Theorem**

*If, besides the Assumptions 1 – 3 for stochastic setting, the probability distribution for the estimator \( \hat{u} \) fulfills the exponential inequality*

\[
P \left\{ \| \hat{u} - E\hat{u} \|^2 \geq (t - 1)E \left( \| \hat{u} - E\hat{u} \|^2 \right) \right\} \leq c_1 \exp(-c_2 t)
\]

*for any \( t > 1 \) and for a constant \( k > 0 \), then it holds*

\[
E(\| a_+ - a^\dagger \|^2 ) \leq \frac{2qK}{p+q} \tau^{\frac{2q}{p+q}} \ln \frac{1}{\tau}.
\]
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
Hilbert scale

\[ X_{-1} := \left\{ v \in L^2 : \int_0^1 v \, dx = 0 \right\}, \]

\[ X_0 = H^1 \cap \left\{ v \in L^2 : \int_0^1 v \, dx = 0 \right\}, \]

\[ X_1 = \left\{ u \in H^2 : u'(0) = u'(1) = 0, \int_0^1 u \, dx = 0 \right\}, \]

\[ X_2 = H^3 \cap X_1, \]

\[ X_3 = \{ \phi \in H^4 \cap X_1 : \phi'''(0) = \phi'''(1) = 0 \}. \]

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
\(Y_i = u(X_i) + \varepsilon_i, \ i = 1, \ldots, n\)
with errors \(\varepsilon_i\) i.i.d. with \(E(\varepsilon_i) = 0\) and \(\text{var}(\varepsilon_i) = 0.01^2, n = 398\)

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

\[ u \in C^3(0,1), \text{grid } m = 100, \]
Gauss. kernel, bandwith by CV

is spline of order 2
\[ s = 2, p = 2, q = 2.5 \]
Let the Hammerstein operator be

\[ F : H^1 \rightarrow L^2 \]

\[ a \rightarrow \int_0^\bullet \Phi(a(t)) \, dt \]

where \( \Phi \in C^{2,1}(\mathbb{R}) \) such that

\[ \| \Phi''(t) \|_{C^{0,1}} \leq K, \forall t \in \mathcal{R}. \]

**Inverse problem** To determine \( a \) from the knowledge of \( u = F(a) \).
Hilbert scale for Hammerstein operator

\[ D(L) = \left\{ w \in H^5 : w'(0) = w'(1) = w^{(3)}(0) = 0, w(1) = w''(1) \right\} \]

\[ L : D(L) \to H^1, \quad Lw := w - 2w'' + w^{(4)}, \quad D(L^{\frac{1}{2}}) = R(F'(a^\dagger)^*) \]

The first elements of Hilbert scale with integer index are

\[ x_0 = H^1 \]

\[ x_2 = \left\{ w \in H^3 : w'(0) = w'(1) = 0, w(1) = w''(1) \right\} \]

\[ \langle w, v \rangle_{x_2} = \int_0^1 w^{(3)}v^{(3)} + 3w''v'' + 3w'v' + wv \, dx \]
Some references

Mair B. A., Ruymgaart F. H.:
Statistical Inverse Estimation in Hilbert Scales

Mathé, Peter and Pereverzev, Sergei V.:
Geometry of linear ill-posed problems in variable Hilbert scales

Bissantz N., Hohage T., Munk A.:
Consistency and rates of convergence of nonlinear Tikhonov regularization with random noise

Hohage T., Pricop M.
Nonlinear Tikhonov regularization in Hilbert scales for inverse boundary value problems with random noise
Thank you for your attention!

Acknowledgement due to Isaac Newton Institute for Mathematical Sciences and the University of Bonn