Universität Potsdam
www.uni-potsdam.de

SFB 1294
Data Assimilation

Inverse learning in Hilbert Scales

Abhishake Rastogi (UP) Peter Mathé (WIAS)

Weierstraß-Institut für Angewandte Analysis und Stochastik
www.wias-berlin.de

Problem of Interest

Within inverse learning we are given data $\{(x_i, y_i)\}_{i=1}^m$ which follow the model

$$y_i = g(x_i) + \varepsilon_i, \quad i = 1, \dots, m, \tag{1}$$

where ε_i is the observational noise, and m denotes the sample size.

The unknown g follows the model

$$g = A(f) \quad \text{for} \quad f \in \mathcal{H}.$$

Penalized least squares: In this standard approach the estimator $f_{\mathbf{z},\lambda}$ is the minimizer of

$$\frac{1}{m}\sum_{i=1}^{m} \|A(f)(x_i) - y_i\|_Y^2 + \lambda \|f\|_{\mathcal{H}}^2, \tag{2}$$

where λ is a positive regularization parameter in form of the penalty $\|f\|_{\mathcal{H}}^2$.

Regularization in Hilbert Scales: Suppose that additional information is given by $f_\rho \in \mathcal{D}(L)$, where L is some unbounded self-adjoint operator $L\colon \mathcal{D}(L) \subset \mathcal{H} \to \mathcal{H}$.

This additional information $f_\rho \in \mathcal{D}(L)$ is taken into account by replacing the above minimization problem (2) by

$$f_{\mathbf{z},\lambda} = \arg\min_{f \in \mathcal{D}(L)} \left\{ \frac{1}{m}\sum_{i=1}^{m} \|A(f)(x_i) - y_i\|_Y^2 + \lambda \|Lf\|_{\mathcal{H}}^2 \right\}. \tag{3}$$

The minimizer $f_{\mathbf{z},\lambda}$ will be the estimator for f_ρ.

Focus of the study

- What will be the rate of reconstruction (as the sample size tends to infinity), and is this comparable to the known (minimax) rate when using (2)?
- What happens under mis-specification, i.e., when the true underlying element f_ρ does not belong to $\mathcal{D}(L)$?

Link condition

In inverse learning the following operators, in population and sampling form are relevant:

- The operator $A : \mathcal{H} \to \mathcal{H}'$
- the smoothness promoting unbounded operator L
- the canonical injection map
$$I_\nu : \mathcal{H}' \to \mathcal{L}^2(X, \rho_X; Y)$$
- Composition of operators
$$B_\nu := I_\nu A L^{-1} : \mathcal{H} \to \mathcal{L}^2(X, \rho_X; Y),$$
and its self-adjoint companion
$$T_\nu := B_\nu^* B_\nu : \mathcal{H} \to \mathcal{H},$$
$$L_\nu := A^* I_\nu^* I_\nu A : \mathcal{H} \to \mathcal{H}$$
- The sampling variants
$$B_{\mathbf{x}}, \quad T_{\mathbf{x}}, \quad L_{\mathbf{x}}$$

For the analysis it is important to adjust how the covariance operator T_ν relates to the smoothness promoting L. This is done by a link condition of the form

Assumption 1 (Link condition) *There exist a power $q > 1$ and an index function ϱ, for which the function ϱ^2 is sub-linear.*

There are constants $1 \le \beta < \infty$ such that

$$\left\|L^{-q}u\right\|_{\mathcal{H}} \le \|\varrho^q(T_\nu)u\|_{\mathcal{H}} \le \beta^q \left\|L^{-q}u\right\|_{\mathcal{H}}, \quad u \in \mathcal{H}.$$

The function $t \mapsto \varphi(t) := \varrho^{q-1}(t)$ belongs to the class $\mathcal{F} = \{\varphi = \varphi_1\varphi_2 : \varphi_1, \varphi_2 : [0, \kappa^2] \to [0, \infty),\ \varphi_1$ nondecreasing continuous sub-linear, φ_2 nondecreasing Lipschitz, $\varphi_1(0) = \varphi_2(0) = 0\}$.

Calculus from Interpolation theory is used to transfer this to other powers of T_ν.

Distance functions

Instead of assuming smoothness of f_ρ in terms of source conditions we use *distance functions* to analyze the behavior of the estimator $f_{\mathbf{z},\lambda}$.

Distance functions measure for an element f the *violation of some benchmark source condition*, they are thus called to *approximate source conditions*.

Regular case

Definition 1 *Given $q \ge 1$ we define the distance function $d_q : [0,\infty) \to [0,\infty)$ by*

$$d_q(R) = \inf\left\{\|L(f - f_\rho)\|_{\mathcal{H}} : f = L^{-q}v \text{ and } \|v\|_{\mathcal{H}} \le R\right\}, \quad R > 0. \tag{4}$$

Oversmoothing case

Of particular interest is the case when $f_\rho \notin \mathcal{D}(L)$, i.e., when the promoted smoothness is not met.

Definition 2 *We define the distance function $d : [0,\infty) \to [0,\infty)$ by*

$$d(R) = \inf\left\{\|f_\rho - f\|_{\mathcal{H}} : f = L^{-1}v \text{ and } \|v\|_{\mathcal{H}} \le R\right\}, \quad R > 0. \tag{5}$$

We denote f_ρ^R the element which realizes the above minimization problem.

Results

Under 'standard assumptions' on the noise, and the 'effective dimension', we obtain the following results.

Regular case

Theorem 1 *Let ϱ, φ from the link condition. The function ϱ^q is sublinear.*

For all $0 < \eta < 1$, and for λ small enough, the following upper bound holds for the regularized solution $f_{\mathbf{z},\lambda}$ (3) with confidence $1 - \eta$:

$$\left\|f_{\mathbf{z},\lambda} - f_\rho\right\|_{\mathcal{H}} \le C\varrho(\lambda)\left\{d_q(R) + R\left(\varphi(\lambda) + \frac{1}{\sqrt{m}}\right) + C'\sqrt{\frac{\mathcal{N}_{T_\nu}(\lambda)}{m\lambda}}\right\} \times \log^4\left(\frac{4}{\eta}\right).$$

Oversmoothing case

Theorem 2 *Let ϱ, φ from the link condition.*

Under certain concavity assumptions *for the function ϱ^{-1} we have the following.*

For all $0 < \eta < 1$, and for λ small enough, the following upper bound holds for the regularized solution $f_{\mathbf{z},\lambda}$ (3) with confidence $1 - \eta$:

$$\left\|f_{\mathbf{z},\lambda} - f_\rho\right\|_{\mathcal{H}} \le C\left\{d(R) + 2R\varrho(\lambda)\right\}\log^4\left(\frac{4}{\eta}\right), \quad R \text{ large enough.}$$

Suppose that smoothness is given explicitly in terms of a **General source condition**:

Assumption 2 *For an index function θ, the true solution f_ρ belongs to the class $\Omega(\theta, R^\dagger)$ with*

$$\Omega(\theta, R^\dagger) := \left\{f \in \mathcal{H} : f = \theta(L^{-1})v \text{ and } \|v\|_{\mathcal{H}} \le R^\dagger\right\}.$$

Then the above bounds can be expressed in terms of the composed function $\theta \circ \varrho$, and it is shown that in many cases this yields order optimal bounds.

References

[1] A. Rastogi, G. Blanchard, and P. Mathé. Convergence analysis of Tikhonov regularization for non-linear statistical inverse learning problems, To appear in *Electronic Journal of Satistics*.
[2] G. Blanchard and N. Mücke, Optimal rates for regularization of statistical inverse learning problems, *Found. Comput. Math.*, **18** (4) (2018) 971–1013.
[3] B. Hofmann and P. Mathé. Analysis of profile functions for general linear regularization methods. *SIAM J. Numer. Anal.*, **45** (3) (2007) 1122–1141.
[4] B. Hofmann and P. Mathé. Tikhonov regularization with oversmoothing penalty for non-linear ill-posed problems in Hilbert scales. *Inverse Probl.*, **34** (1) (2018) 15007.
[5] F. Natterer. Error bounds for Tikhonov regularization in Hilbert scales. *Appl. Anal.*, **18** (1-2) (1984) 29–37.