Universität Potsdam
www.uni-potsdam.de

SFB 1294
Data Assimilation

Regularization Scheme for Statistical Inverse Problems

Abhishake Rastogi (UP) Peter Mathé (WIAS) Gilles Blanchard (UP)

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

Problem

The problem of interest:

$$A(f) = g, \quad y_i := g(x_i) + \varepsilon_i, \quad i = 1, \dots, m.$$

$$\underbrace{A}_{\text{Nonlinear operator}} : \mathcal{D}(A) \subset \underbrace{\mathcal{H}_1}_{\text{Hilbert space}} \to \underbrace{\mathcal{H}_2}_{\text{Reproducing kernel Hilbert space}}$$

$x_i \in X$: Polish space
$y_i \in Y$: Real separable Hilbert space
Training data:

$$\underbrace{\rho(x,y)}_{\text{Unknown}} \Rightarrow \underbrace{\{(x_1,y_1),\dots,(x_m,y_m)\}}_{\text{i.i.d. obersevations}} = \mathbf{z}$$

$$\rho(x,y) = \rho(y|x)\rho_X(x)$$

The goal: Develop an algorithm

$$\mathbf{z} \mapsto f_{\mathbf{z}} \quad \text{s.t.} \quad f_{\mathbf{z}} \approx f$$

Statistical framework:
The **expected risk** of the estimator $f \in \mathcal{D}(A) \subset \mathcal{H}_1$:

$$\mathcal{E}(f) = \int_Z ||A(f)(x) - y||_Y^2 d\rho(x,y).$$

Suppose given ρ, there exists $f_\rho \subset \mathcal{D}(A) \in \mathcal{H}_1$ such that

$$E_\rho[Y|X=x] = \int_Y y d\rho(y|x) = A(f_\rho)(x), \text{ for all } x \in X.$$

- The **minimum risk** is attained by the function f_ρ.

$$\mathcal{E}(f) = \int_X ||A(f)(x) - A(f_\rho)(x)||_Y^2 d\rho_X(x) + \mathcal{E}(f_\rho)$$

- But the probability measure ρ is **unknown**.

Tikhonov regularization:

$$f_{\mathbf{z},\lambda} = \text{arg} \min_{f \in \mathcal{D}(A) \subset \mathcal{H}_1} \left\{ \frac{1}{m} \sum_{i=1}^m ||A(f)(x_i) - y_i||_Y^2 + \lambda ||f - f^*||_{\mathcal{H}_1}^2 \right\}$$

- λ is the positive regularization parameter.
- $f^* \in \mathcal{H}_1$ denotes some initial guess of the ideal solution.
- The operator A is assumed to be one-to-one and weakly sequentially closed $\Rightarrow \exists$ **a global minimizer**, but not necessarily unique.

Assumptions

Assumptions on nonlinear operator A:

- $\mathcal{D}(A)$ is convex.
- $A : \mathcal{D}(A) \subset \mathcal{H}_1 \to \mathcal{H}_2 \hookrightarrow \mathcal{L}^2(X, \rho_X; Y)$ is one-to-one and weakly sequentially closed.
- A is Fréchet differentiable.
- Fréchet derivative of A at f_ρ is bounded, i.e., $||A'(f_\rho)||_{\mathcal{H}_1 \to \mathcal{H}_2} \le L$.
- $\exists \gamma \ge 0 \ni \forall f \in \mathcal{D}(A) \subset \mathcal{H}_1$ in a sufficiently large ball around f_ρ we have,

$$\left\| I_K \left\{ A(f) - A(f_\rho) - A'(f_\rho)(f - f_\rho) \right\} \right\|_{\mathcal{L}^2(X,\rho_X;Y)} \le \frac{\gamma}{2} ||f - f_\rho||_{\mathcal{H}_1}^2 .$$

Assumptions on kernel K:

- $\forall x \in X$, $K_x : Y \to \mathcal{H}_2$ is a Hilbert-Schmidt operator.
- $\kappa := \sqrt{\sup_{x \in X} Tr(K_x^* K_x)} < \infty.$

The class of probability measures $\mathcal{P}_\phi$:

- There exist some constants M, Σ such that for almost all $x \in X$,

$$\int_Y \left(e^{||y - A(f_\rho)(x)||_Y/M} - \frac{||y - A(f_\rho)(x)||_Y}{M} - 1 \right) d\rho(y|x) \le \frac{\Sigma^2}{2M^2}.$$

- $f_\rho \in \Omega_{\phi,R} := \left\{ f \in \mathcal{H}_1 : f - f^* = \phi(T)g \text{ and } ||g||_{\mathcal{H}_1} \le R \right\}$, where $T = I_K^* A'(f_\rho)^* A'(f_\rho) I_K$ for operator $I_K : \mathcal{H}_2 \to \mathcal{L}^2(X, \rho_X; Y)$.
- The eigenvalues $(t_n)_{n \in \mathbb{N}}$ of the operator T follow the polynomial decay:

$$\alpha n^{-b} \le t_n \le \beta n^{-b} \quad \forall n \in \mathbb{N}, \quad \alpha, \beta > 0, \quad b > 1.$$

General source condition

General source condition $f_\rho \in \Omega_{\phi,R}$, by allowing for the index functions ϕ, cover a wide range of source conditions, such as

Hölder source condition $\phi(t) = t^r$ with $r \ge 0$, and

logarithmic-type source condition $\phi(t) = t^p \log^{-\nu}\left(\frac{1}{t}\right)$ with $p \in \mathbb{N}, \ \nu \in [0,1]$.

Optimal convergence rates

Theorem 1 *Let $\mathbf{z}$ be i.i.d. samples drawn according to the probability measure $\rho \in \mathcal{P}_{\phi,b}$ where $\phi(t) = \sqrt{t}\psi(t)$ is the index function satisfying the conditions that $\phi(t)$ and $t/\phi(t)$ are nondecreasing functions. Then under Assumption on the operator A and the parameter choice $\lambda \in (0,1]$, $\lambda = \Psi^{-1}(m^{-1/2})$ where $\Psi(t) = t^{\frac{1}{2}+\frac{1}{2b}}\phi(t)$, for all $0 < \eta < 1$, with confidence $1-\eta$, for the regularized estimator $f_{\mathbf{z},\lambda}$ the following convergence rate holds:*

$$||f_{\mathbf{z},\lambda} - f_\rho||_{\mathcal{H}_1} \le C\phi(\Psi^{-1}(m^{-1/2})) \log\left(\frac{4}{\eta}\right)$$

provided that

$$8\kappa^2 \max\left(1, \frac{L(M+\Sigma)}{\kappa d}\right) \log(4/\eta) \le \sqrt{m}\lambda$$

and

$$2\gamma ||T^{-1/2}(f_\rho - f^*)||_{\mathcal{H}_1} < 1.$$

Comparison for Hölder's source condition $\phi(t) = t^r$:

	$\lVert f_{\mathbf{z},\lambda} - f \rVert_{\mathcal{H}_1}$	Smoothness	Scheme	general source condition	Optimal rates
Rastogi et al. (2017)	$m^{-\frac{br}{2br+b+1}}$	$0 \le r \le 1$	Direct learning	✓	✓
Blanchard et al. (2018)	$m^{-\frac{br}{2br+b+1}}$	$0 \le r \le 1$	Linear inverse learning		✓
Rastogi et al. (2020)	$m^{-\frac{br}{2br+b+1}}$	$\frac{1}{2} \le r \le 1$	Non-linear inverse learning	✓	✓

Further questions and developments

Statistical properties of the inverse problem and applications in covariate modeling.

- develope statistically and computationally effective algorithms.
- consider iterative schemes based on different regularization methods.
- obtain confidence regions for the nonparametric model to design a statistical goodness-of-fit test for parametric models.
- evaluate the performance of nonparametric covariate-parameter modeling against simulated data from a so-called physiologically based pharmacokinetic model and design specific kernels for the application field.
- focusing on methodological aspects of the inverse problem and on applications in pharmacology.

Numerical algorithms with statistical guarantees:

- investigate the algorithmic cost and performance of local iterative procedures, with controlled statistical performance
- construct statistically adaptive early stopping rules.
- study the performance of stochastic optimization methods.

References

[1] A. Rastogi and S. Sivananthan. Optimal rates for the regularized learning algorithms under general source condition, *Front. Appl. Math. Stat.*, **3** (2017) 3.
[2] G. Blanchard and N. Mücke, Optimal rates for regularization of statistical inverse learning problems, *Found. Comput. Math.*, **18**, no. 4 (2018) 971–1013.
[3] A. Rastogi, G. Blanchard, and P. Mathé. Convergence analysis of Tikhonov regularization for non-linear statistical inverse learning problems, *Electronic Journal of Satistics,* Accepted for publication (2020).