A Supervised Learning Approach to Regularization of Inverse Problems

Chirinos Rodríguez, J., Molinari, C., De Vito, E., Rosasco, L. and Villa, S.

jonathan.chirinosrodriguez@edu.unige.it

Abstract

- Selecting the best regularization parameter in inverse problems is a classical and yet challenging problem.
- We propose and study a statistical machine learning approach based on empirical risk minimization.
- Our main contribution is a theoretical analysis, showing that, provided with enough data, this approach can reach sharp rates.

Model and Approach

Let $A\colon \mathcal{X} \to \mathcal{Y}$, $\mathcal{X}$, $\mathcal{Y}$ Hilbert spaces, be an operator and $x^* \in \mathcal{X}$, $y \in \mathcal{Y}$ be such that

$$y = A(x^*) + \delta, \quad \delta \text{ noise.}$$

To overcome ill-posedness, we consider, for every $\lambda > 0$, the family of regularization methods $X_\lambda = X_\lambda(y)$. A good choice for λ should lead to a good reconstruction:

$$X_\lambda(y) \sim x^*.$$

Question: How to formalize this?

Data-driven approach

1. We turn from deterministic to stochastic inverse problems:

$$Y = A(X) + \delta,$$

where Y, X and δ are random variables.

2. We assume to have pairs $(y_i, x_i) \sim (Y, X)$, $i = 1, ..., n$, and we find a minimizer of the **empirical risk**:

$$\widehat{\lambda}_\Lambda \in \arg\min_{\lambda\in\Lambda} \widehat{L}(X_\lambda) := \frac{1}{n}\sum_{i=1}^{n} \ell(x_i, X_\lambda(y_i))),$$

for a certain bounded discrepancy ℓ, bounded by $M > 0$, and a set of candidates $\Lambda = \{\lambda_1, ..., \lambda_N\}$, $N \in \mathbb{N}$.

3. Compute an upper bound for $L(X_{\widehat{\lambda}_\Lambda})$, where

$$\lambda_\Lambda \in \arg\min_{\lambda\in\Lambda} L(X_\lambda) := \mathbb{E}\left[\ell(X, X_\lambda)\right].$$

Assumptions and main result

Assumption 1 *We assume that* $\mathbb{E}[\delta|X] = 0$ *and, moreover, that there exists* $\tau > 0$ *such that*

$$\mathbb{E}[\|\delta\|^2_{\mathcal{Y}}|X] \leq \tau^2.$$

Assumption 2 *We assume, for every* $\lambda > 0$, *that*

$$L(X_\lambda) \leq \varepsilon(\lambda).$$

where ε *attains its minimum in* λ_*, *and there exists* $C = C(q)$ *such that, for all* $q > 1$,

$$\varepsilon(q\lambda) \leq C(q)\varepsilon(\lambda).$$

Assumption 3 *Assume* $\Lambda = \{\lambda_j = \lambda_1 Q^{j-1}, j = 1, \dots, N\}$, $Q > 1$, *with*

$$\lambda_1 \leq \lambda_* \leq \lambda_N.$$

Theorem 1 *Let Assumptions 1, 2 and 3 be satisfied and let* $\eta \in (0, 1]$. *Then, with probability at least* $1 - \eta$,

$$L(X_{\widehat{\lambda}_\Lambda}) \lesssim C(Q)\varepsilon(\lambda_*) + \frac{M}{n}\log\frac{2N}{\eta},$$

where the r.h.s goes to 0 as $n \to \infty$ (up to a constant).

Idea: Combine

$$L(X_{\widehat{\lambda}_\Lambda}) \lesssim 2L(X_{\lambda_\Lambda}) + \frac{M}{n}\log\frac{2N}{\eta}$$

and

$$L(X_{\lambda_\Lambda}) \leq L(X_{q\lambda_*}) \leq \varepsilon(q\lambda_*) \leq C(Q)\varepsilon(\lambda_*).$$

Spectral Regularization Methods

Consider

$$X_\lambda = g_\lambda(A^*A)A^*Y$$

where A is linear and g_λ, $\lambda > 0$, are spectral functions (e.g. Tikhonov, Landweber, TSVD). Choose $\ell(X, X_\lambda) = \|X_\lambda - X\|^2$. Then,

Theorem 2 *There exist constants* α, C_1, $C_2 > 0$ *such that*

$$\varepsilon(\lambda) = \frac{C_1^2\tau^2}{\lambda} + C_2^2\lambda^{2\alpha}.$$

"Nonlinear" Tikhonov Regularization

We consider, for every $\lambda > 0$ and a certain initialization variable X_0,

$$X_\lambda = \arg\min_{x\in\mathcal{X}} \|A(x) - Y\|^2 + \lambda\|x - X_0\|^2,$$

where A is non-linear. Choose $\ell(X, X_\lambda) = \|X_\lambda - X\|^2$. Then,

Theorem 3 *There exist constants* c_1, $c_2 > 0$ *such that*

$$\varepsilon(\lambda) = c_1\frac{(\tau^2 + c_2\lambda)(1+\lambda)}{\lambda}.$$

Sparsity Inducing Regularizers

We finally consider, for every $\lambda > 0$,

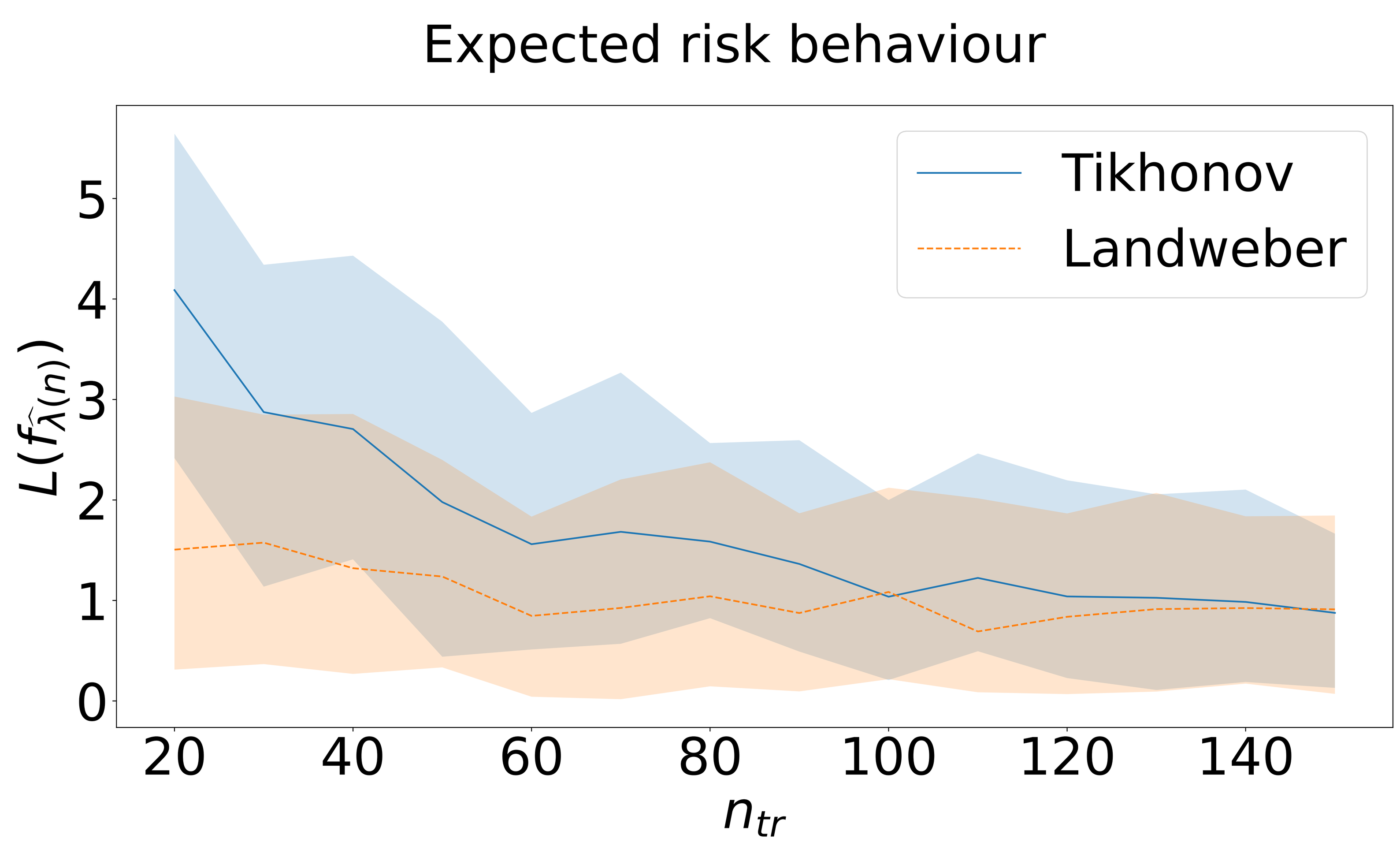
$$X_\lambda = \arg\min_{x\in\mathbb{R}^d} \|Ax - Y\|^2 + \|Gx\|_1$$

where A is linear and $G\colon \mathbb{R}^d \to \mathbb{R}^n$ is linear and bounded. Choose $\ell(X, X_\lambda) = D_J(X, X_\lambda)$ the Bregman divergence. Then,

Theorem 4 *There exists a constant* $c > 0$ *such that*

$$\varepsilon(\lambda) = \frac{\tau^2\lambda}{2} + \frac{c^2}{2\lambda}.$$

Expected risk behaviour



We plot here the expected risk w.r.t. $\widehat{\lambda}(n)$ for both Tikhonov regularization and the Landweber iteration with $\tau = 0.1$. The solid lines represent the mean value, while the shaded areas are the 1-percentiles and 99-percentiles over 30 trials.

References

[1] Cucker, F. and Smale, S.. *On the Mathematical Foundations of Learning*. 2001.

[2] Engl, H. W., Hanke, M. and Neubauer, A.. *Regularization of inverse problems*. 1996.

[3] Arridge, S., Maass, P., Öktem, O., and Schönlieb, C-B.. *Solving inverse problems using data-driven models*. 2019.