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 $A : X \to Y$, $X$, $Y$ Hilbert spaces, be an operator and $x \in X$, $y \in Y$ be such that $y = A(x) + \delta$, $\delta$ 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 $\lambda$ 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 $\delta$ 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:

$$\tilde{\lambda}_n \in \arg \min_{\lambda \in \Lambda} \frac{1}{n} \sum_{i=1}^{n} l(X_\lambda, Y_i),$$

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 $L(X_\lambda)$, where

$$\lambda_n \in \arg \min_{\lambda \in \Lambda} L(X_\lambda) \simeq \mathbb{E} \left[ l(X, X_\lambda) \right].$$

Assumptions and main result

**Assumption 1** We assume that $\mathbb{E}[\delta | X] = 0$ and, moreover, that there exists $\tau > 0$ such that $\mathbb{E}[\|\delta\|^2 | X] \leq \tau^2$.

**Assumption 2** We assume, for every $\lambda > 0$, that $L(X_\lambda) \leq \epsilon(\lambda)$.

where $\epsilon$ attains its minimum in $\lambda_n$, and there exists $C = C(q)$ such that, for all $q > 1$,

$$\epsilon(q, \lambda) \leq C(q, \epsilon(\lambda)).$$

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

$$\lambda_1 \leq \lambda_n \leq \lambda_N.$$

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

$$L(X_\lambda) \leq C(Q) \epsilon(\lambda) + \frac{M}{n} \log \frac{2N}{\eta},$$

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

**Idea:** Combine

$$L(X_\lambda) \leq 2L(X_{\lambda_n}) + \frac{M}{n} \log \frac{2N}{\eta}$$

and

$$L(X_\lambda) \leq L(X_{\lambda_n}) \leq \epsilon(q, \lambda_n) \leq C(Q) \epsilon(\lambda_n).$$

Spectral Regularization Methods

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

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

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

**“Nonlinear” Tikhonov Regularization**

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

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

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

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

$$\epsilon(\lambda) \leq c_1 (\tau^2 + c_2 \lambda)(1 + \lambda).$$

Sparsity Inducing Regularizers

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

$$X_\lambda = \arg \min_{x \in \mathbb{R}^n} \|A(x) - Y\|^2 + \|Gx\|_1,$$

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

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

$$\epsilon(\lambda) \leq c \frac{\tau^2 \lambda}{\lambda} + c^2 \frac{\lambda}{\lambda^2}.$$

Expected risk behaviour

![Expected risk behaviour graph](image)

We plot here the expected risk w.r.t. $\hat{\lambda}(n)$ 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