Convergence Analysis of Tikhonov Regularization for Nonlinear Statistical Inverse Learning Problems

Presented by

Abhishake

Institut für Mathematik
Universität Potsdam

Colloque du Collège Doctoral Franco-Allemand
Stochastic processes and statistical machine learning II
Institut de Mathématiques de Toulouse

Mar 15, 2019

Nonlinear statistical inverse learning problem

Model

- A is a known nonlinear operator from a separable Hilbert space $\mathcal{H}_1$ to another Hilbert space $\mathcal{H}_2$.
- **The problem of interest** can be described as

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

  at a given set of observations $\mathbf{z} = \{(x_i, y_i)\}_{i=1}^m$.
- The random observations $\mathbf{z}$ are drawn independently and identically according to the unknown joint probability distribution ρ.
- $(\varepsilon_i)_{i=1}^m$ are independent centered noise variables satisfying $E_\rho[\varepsilon_i | x_i] = 0$.
- **The goal:** Provide an estimator $f_\mathbf{z}$ of f from the given set of examples $\mathbf{z} = \{(x_i, y_i)\}_{i=1}^m$.
- This is commonly called the statistical learning setting and the model is referred as nonlinear statistical inverse learning problem.

Expected risk

The goodness of the estimator f can be measured by the expected risk:

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

- The goal is to find an estimator which minimizes the above risk function $\mathcal{E}(f)$ over an admissible class of functions which is referred as **hypothesis space**.

Ideal function

Under the condition $E_\rho[\varepsilon|x] = 0$ for $y := A(f)(x) + \varepsilon$.

Assumption

The conditional expectation w.r.t. ρ of y given x exists, and it holds for all $x \in X$:

$$E_\rho[y|x] = \int_Y y d\rho(y|x) = A(f)(x) = A(f_\rho)(x), \text{ for some } f_\rho \in \mathcal{D}(A) \subset \mathcal{H}_1.$$

Ideal function

Under the condition $E_\rho[\varepsilon|x] = 0$ for $y := A(f)(x) + \varepsilon$.

Assumption

The conditional expectation w.r.t. ρ of y given x exists, and it holds for all $x \in X$:

$$E_\rho[y|x] = \int_Y y d\rho(y|x) = A(f)(x) = A(f_\rho)(x), \text{ for some } f_\rho \in \mathcal{D}(A) \subset \mathcal{H}_1.$$

Proposition (Cucker, Smale (2002))

For every $f : X \to Y$,

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

where $\sigma_\rho^2 = \int_X \int_Y ||y - A(f_\rho)(x)||_Y^2 d\rho(y|x) d\rho_X(x)$ and $\rho(\cdot|x)$, ρ_X are conditional probability, marginal probability, respectively.

Tikhonov regularization

- But, in general, the probability measure ρ is unknown.
- Given a training set $\mathbf{z}$, we define the empirical error:

$$\mathcal{E}_{\mathbf{z}}(f) = \frac{1}{m}\sum_{i=1}^{m} ||A(f)(x_i) - y_i||_Y^2.$$

Tikhonov regularization

- But, in general, the probability measure ρ is unknown.
- Given a training set $\mathbf{z}$, we define the empirical error:

$$\mathcal{E}_{\mathbf{z}}(f) = \frac{1}{m}\sum_{i=1}^{m} ||A(f)(x_i) - y_i||_Y^2.$$

- A widely used approach to the estimation problem is nonlinear **Tikhonov regularization**:

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

  where λ is the positive regularization parameter.
- Here $f^* \in \mathcal{D}(A) \subset \mathcal{H}_1$ denotes some initial guess of the ideal solution, which offers the possibility to incorporate **a-priori information**.
- The regularizer should encode some notion of smoothness/complexity of the solution
- The regularization parameter λ trade-offs the two terms.

Tikhonov regularization

- But, in general, the probability measure ρ is unknown.
- Given a training set $\mathbf{z}$, we define the empirical error:

$$\mathcal{E}_{\mathbf{z}}(f) = \frac{1}{m}\sum_{i=1}^{m} ||A(f)(x_i) - y_i||_Y^2.$$

- A widely used approach to the estimation problem is nonlinear **Tikhonov regularization**:

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

  where λ is the positive regularization parameter.
- Here $f^* \in \mathcal{D}(A) \subset \mathcal{H}_1$ denotes some initial guess of the ideal solution, which offers the possibility to incorporate **a-priori information**.
- The regularizer should encode some notion of smoothness/complexity of the solution
- The regularization parameter λ trade-offs the two terms.
- If A is **one-to-one and weakly sequentially closed**, then there **exists a global minimum** of the Tikhonov functional. But it is not necessarily unique, since A is nonlinear.

Definition (Reproducing Kernel Hilbert Space (RKHS))

Let X be an arbitrary set and $\mathcal{H}$ be a Hilbert space of real-valued functions on X. The evaluation functional over the Hilbert space of functions $\mathcal{H}$ is a linear functional that evaluates each function at a point x,

$$L_x : f \mapsto f(x) \ \ \forall f \in \mathcal{H}.$$

We say that $\mathcal{H}$ is a reproducing kernel Hilbert space if L_x is a continuous function for any x in X.

Definition (Mercer kernel)

$K : X \times X \to \mathbb{R}$ is a Mercer kernel if it is continuous, symmetric, and positive semidefinite.

Remark (N. Aronszajn, 1950)

There is one to one correspondence between the reproducing kernel Hilbert spaces and the reproducing kernels.

Mercer kernels

Construction of $\mathcal{H}_K$ from a given kernel K

1) K: $X \times X \to \mathbb{R}$ is the mercer kernel; $K_x = K(x, \cdot)$.

2) $H_K = \{f : f = \sum_{j=1}^{r} c_j K_{x_j}\}$, $K_{x_j} = K(x_j, \cdot)$

3) $\langle f, g \rangle_K = \langle \sum_{j=1}^{r} c_j K_{x_j}, \sum_{i=1}^{s} d_i K_{t_i} \rangle_K := \sum_{j=1}^{r} \sum_{i=1}^{s} c_j d_i K(x_j, t_i)$

4) $\mathcal{H}_K$ is the completion of H_K w.r.t $\|\cdot\|_K$

$$\forall f \in \mathcal{H}_K \quad f(x) = \langle K_x, f \rangle_K$$

Examples

- Gaussian RBF kernel $K(x, t) = e^{-||x-t||^2}$
- Polynomial of degree d kernel function $K(x, t) = (1 + x \cdot t)^d$
- Suppose $k \in \mathcal{L}^2(\mathbb{R}^n, \nu; \mathbb{R})$ be continuous, even function and the Fourier transform of k is nonnegative. Then the kernel $K(x, y) = k(x - y)$ is a Mercer kernel on $\mathbb{R}^n$.

Vector-valued reproducing kernel Hilbert space

Micchelli and Pontil (2005) introduced the concept of vector-valued reproducing kernel Hilbert space.

Definition (Vector-valued reproducing kernel Hilbert space (RKHSvv))

For non-empty set X and the real Hilbert space $(Y, \langle\cdot,\cdot\rangle_Y)$, the Hilbert space $(\mathcal{H}, \langle\cdot,\cdot\rangle_{\mathcal{H}})$ of functions from X to Y is called reproducing kernel Hilbert space if for any $x \in X$ and $y \in Y$ the linear functional which maps $f \in \mathcal{H}$ to $\langle y, f(x)\rangle_Y$ is continuous.

Definition (Operator-valued positive definite kernel)

Suppose $\mathcal{L}(Y)$ be the Banach space of bounded linear operators on Y. A function $K : X \times X \to \mathcal{L}(Y)$ is said to be an operator-valued positive definite kernel if for each pair $(x, z) \in X \times X$, $K(x, z)^* = K(z, x)$, and for every finite set of points $\{x_i\}_{i=1}^N \subset X$ and $\{y_i\}_{i=1}^N \subset Y$,

$$\sum_{i,j=1}^{N} \langle y_i, K(x_i, x_j)y_j\rangle_Y \geq 0.$$

- *Tikhonov regularization* for the **direct learning** scheme ($A = I$ and $\mathcal{H}_1 = \mathcal{H}_2 = \mathcal{H}$):

$$f_{\mathbf{z},\lambda} = \underset{f \in \mathcal{H}}{\operatorname{argmin}} \left\{ \frac{1}{m} \sum_{i=1}^{m} ||f(x_i) - y_i||_Y^2 + \lambda ||f||_{\mathcal{H}}^2 \right\},$$

An important result

The minimizer of the Tikhonov regularization problem over RKHS $\mathcal{H}$ can be represented by the expression:

$$f_{\mathbf{z},\lambda} = \sum_{i=1}^{m} c_i K_{x_i}, \text{ for } \mathbf{c} = (c_1, \ldots, c_m) = (\mathbb{K} + \lambda m \mathbb{I})^{-1} \mathbf{y},$$

where $\mathbb{K} = (K(x_i, x_j))_{i,j=1}^{m}$ and $\mathbb{I}$ is identity of size $m \times m$.

Hence, minimizing over the (possibly infinite dimensional) Hilbert space, boils down to minimizing over R^m.

Empirical risk minimization

Similarly we can prove that the solution of empirical risk minimization

$$\min_{f \in \mathcal{H}} \frac{1}{m} \sum_{i=1}^{m} (y_i - f(x_i))^2$$

can be written as

$$f_{\mathbf{z}}(x) = \sum_{i=1}^{m} c_i K(x, x_i)$$

where the coefficients satisfy

$$\mathbb{K}c = \mathbf{y}.$$

The Role of Regularization

Now we can observe that adding a penalty has an effect from a numerical point of view:

$$\mathbb{K}c = \mathbf{y} \Rightarrow (\mathbb{K} + m\lambda\mathbb{I})c = \mathbf{y}$$

it stabilizes a possibly ill-conditioned matrix inversion problem.

This is the point of view of regularization for (ill-posed) inverse problems.

Hadamard introduced the definition of ill-posedness. Ill-posed problems are typically inverse problems.

If $g \in G$ and $f \in F$, with G, F Hilbert spaces, a linear, continuous operator L, consider the equation

$$g = Lf$$

The direct problem is to compute g given f; the inverse problem is to compute f given the data g.

The inverse problem of finding f is well-posed when

- the solution exists,
- is unique and
- is stable, that is depends continuously on the initial data g.

Otherwise the problem is ill-posed.

In the finite dimensional case the main problem is numerical stability.

For example, in the learning setting the kernel matrix can be decomposed as $\mathbb{K} = Q\Sigma Q^T$, with $\Sigma = \mathrm{diag}(\sigma_1, \dots, \sigma_n)$, $\sigma_1 \geq \sigma_2 \geq \dots \sigma_n \geq 0$ and $q_1, \dots, q_n$ are the corresponding eigenvectors.
Then

$$c = \mathbb{K}^{-1}\mathbf{y} = Q\Sigma^{-1}Q^T\mathbf{y} = \sum_{i=1}^{m} \frac{1}{\sigma_i}\langle q_i, \mathbf{y}\rangle q_i$$

In correspondence of small eigenvalues, small perturbations of the data can cause large changes in the solution. The problem is ill-conditioned.

For Tikhonov regularization

$$\begin{aligned} c &= (\mathbb{K} + m\lambda\mathbb{I})^{-1}\mathbf{y} \\ &= Q(\Sigma + m\lambda\mathbb{I})^{-1}Q^T\mathbf{y} \\ &= \sum_{i=1}^{m} \frac{1}{\sigma_i + m\lambda}\langle q_i, \mathbf{y}\rangle q_i \end{aligned}$$

Regularization filters out the undesired components.

For $\sigma \gg \lambda m$, then $\frac{1}{\sigma_i + m\lambda} \sim \frac{1}{\sigma_i}$.
For $\sigma \gg \lambda m$, then $\frac{1}{\sigma_i + m\lambda} \sim \frac{1}{\lambda m}$.

Objective

- **Main objective:** To analyze the **theoretical properties** of the regularized estimator $f_{\mathbf{z},\lambda}$.
- In particular, **the rates of convergence** of its estimator $f_{\mathbf{z},\lambda}$ to the ideal function f_ρ in a **reproducing kernel Ansatz**.

Let the input space X be a locally compact countable Hausdorff space and the output space $(Y, \langle\cdot,\cdot\rangle_Y)$ be a real separable Hilbert space.

Assumption

- *For all* $x \in X$, $K_x : Y \to \mathcal{H}$ *is a Hilbert-Schmidt operator and* $\kappa := \sqrt{\sup_{x\in X} Tr(K_x^* K_x)} < \infty$, *where for Hilbert-Schmidt operator* $F \in \mathcal{L}(\mathcal{H}_2)$, $Tr(F) := \sum_{k=1}^{\infty} \langle Fe_k, e_k\rangle$ *for an orthonormal basis* $\{e_k\}_{k=1}^{\infty}$ *of* $\mathcal{H}_2$.
- *The real-valued function* $\phi : X \times X \to \mathbb{R}$, *defined by* $\phi(x,t) = \langle K_t v, K_x w\rangle_{\mathcal{H}_2}$, *is measurable* $\forall v, w \in Y$.

Class of probability measures $\mathcal{P}_{\phi,b}$

Covariance operator

For the canonical injection $I_K : \mathcal{H} \to \mathcal{L}^2(X, \rho_X; Y)$ the covariance operator $T : \mathcal{H} \to \mathcal{H}$ is defined as

$$T = I_K^* I_K.$$

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

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

  for $\epsilon = y - f_\rho(x)$.

- $f_\rho \in \Omega_{r,R} := \{f \in \mathcal{H} : f - f^* = \phi(T)g \text{ and } ||g||_{\mathcal{H}} \leq R\}$,
  where ϕ is a continuous increasing index function defined on the interval $[0, \kappa^2]$ with the assumption $\phi(0) = 0$. This condition is usually referred to as **general source condition**.

- The eigenvalues $(t_n)_{n \in \mathbb{N}}$ of the operator T follow the polynomial decay:

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

Remark

General source condition $f_\rho \in \Omega_{\phi,R}$ *corresponding to index function* ϕ *covers wide range of source conditions as Hölder's source condition* $\phi(t) = t^r$, *logarithm source condition* $\phi(t) = t^p \log^{-\nu}\left(\frac{1}{t}\right)$.

Effective dimension

The effective dimension $\mathcal{N}(\lambda)$, measures the complexity of RKHS, can be defined as:

$$\mathcal{N}(\lambda) := Tr\left((T + \lambda I)^{-1} T\right).$$

Convergence rates

We are interested in exponential tail inequalities such that with probability at least $1-\eta$

$$||f_{\mathbf{z}} - f_\rho|| \leq \varepsilon(m) log\left(\frac{1}{\eta}\right)$$

for some positive decreasing function $\varepsilon(m)$ and $0 < \eta \leq 1$.

$\mathbf{x} = (x_1, \ldots, x_m)$
$\mathbf{y} = (y_1, \ldots, y_m)$
$S_\mathbf{x} = (f(x_1), \ldots, f(x_m))$

- For the regularized solution
  $f_{\mathbf{z},\lambda} = (S_\mathbf{x}^* S_\mathbf{x} + \lambda I)^{-1} S_\mathbf{x}^* \mathbf{y}$
  and
  $f_\lambda = (T + \lambda I)^{-1} T f_\rho.$
- Now $f_{\mathbf{z},\lambda} - f_\rho$ can be expressed as

$$f_{\mathbf{z},\lambda} - f_\lambda = \underbrace{(S_\mathbf{x}^* S_\mathbf{x} + \lambda I)^{-1}\{S_\mathbf{x}^* \mathbf{y} - S_\mathbf{x}^* S_\mathbf{x} f_\lambda - T(f_\rho - f_\lambda)\}}_{\text{Sample error}} + \underbrace{f_\lambda - f_\rho}_{\text{Approximation error}}$$

- **First term:** Under the noise condition
  $\mathbb{P}_{\mathbf{z} \in Z^m} \left\{ ||f_{\mathbf{z},\lambda} - f_\lambda||_\mathcal{H} \leq C \left(\frac{1}{m\lambda} + \sqrt{\frac{\mathcal{N}(\lambda)}{m\lambda}} \right) \log\left(\frac{4}{\eta}\right) \right\} \geq 1 - \eta$

  Second term: Under the source condition
  $||f_\lambda - f_\rho||_\mathcal{H} \leq R\phi(\lambda)$

Convergence rates for Tikhonov regularization

Theorem

Let $\mathbf{z}$ *be i.i.d. samples drawn according to the probability measure* $\rho \in \mathcal{P}_{\phi,b}$ *where* ϕ *is the index function satisfying the conditions that* $\phi(t)$*,* $t/\phi(t)$ *are nondecreasing functions. Then for all* $0 < \eta < 1$ *and the parameter choice* $\lambda \in (0,1]$*,* $\lambda = \Psi^{-1}(m^{-1/2})$ *where* $\Psi(t) = t^{\frac{1}{2}+\frac{1}{2b}}\phi(t)$*, the convergence of the estimator* $f_{\mathbf{z},\lambda}$ *to the target function* f_ρ *can be described as*

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

Corollary

For Hölder's source condition $f_\rho \in \Omega_{\phi,R}$*,* $\phi(t) = t^r$*, for all* $0 < \eta < 1$*, with confidence* $1 - \eta$*, for the parameter choice* $\lambda = m^{-\frac{b}{2br+b+1}}$*, we have*

$$||f_{\mathbf{z},\lambda} - f_\rho||_{\mathcal{H}} \leq Cm^{-\frac{br}{2br+b+1}}\log\left(\frac{4}{\eta}\right) \text{ for } 0 \leq r \leq 1.$$

Assumptions on the nonlinear operator A

- $\mathcal{D}(A)$: convex
- $A : \mathcal{D}(A) \subset \mathcal{H}_1 \to \mathcal{H}_2 \hookrightarrow \mathcal{L}^2(X, \rho_X; Y)$ is weakly sequentially closed.
  [i.e., if a sequence $(f_m)_{m\in\mathbb{N}} \subset \mathcal{D}(A)$ converges weakly to some $f \in \mathcal{H}_1$ and if the sequence $(A(f_m))_{m\in\mathbb{N}}$ converges weakly to some $g \in \mathcal{L}^2(X, \rho_X; Y)$, then $f \in \mathcal{D}(A)$ and $A(f) = g$.]
- A: Fréchet differentiable
- $||A'(f_\rho)||_{\mathcal{H}_1 \to \mathcal{H}_2} \leq L$
- $\exists\ \gamma \geq 0\ \ni\ \forall\ f \in \mathcal{D}(A) \subset \mathcal{H}_1$ in a sufficiently large ball around f_ρ:

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

- Let I_K denote the canonical injection map $\mathcal{H}_2 \to \mathcal{L}^2(X, \rho_X; Y)$.

We define the operator:

$$\begin{aligned} B : \mathcal{H}_1 &\to \mathcal{L}^2(X, \rho_X; Y) \\ f &\mapsto Bf := [I_K \circ (A'(f_\rho))]f = I_K(A'(f_\rho)f), \end{aligned}$$

- The operator B is bounded and satisfies $||B||_{\mathcal{H}_1 \to \mathcal{L}^2(X,\rho_X;Y)} \leq \kappa L$.
- $T := B^*B$ are positive, self-adjoint operators.

Theorem

Assume that $\mathcal{D}(A)$ is weakly closed with nonempty interior and $A : \mathcal{D}(A) \subset \mathcal{H}_1 \to \mathcal{H}_2$ is Lipschitz continuous, one-to-one and that noise condition holds true and $\sigma_\rho^2 := \int_Z ||y - A(f_\rho)(x)||_Y^2 d\rho(x, y) < \infty$. Let $f_{\mathbf{z},\lambda}$ denote a (not necessarily unique) solution to the minimization problem and assume that the regularization parameter $\lambda(m) > 0$ is chosen such that

$$\lambda \to 0, \quad \frac{1}{\lambda\sqrt{m}} \to 0 \text{ as } m \to \infty.$$

Then we have that

$$\mathbb{E}_{\mathbf{z}}\left(\|f_{\mathbf{z},\lambda} - f_\rho\|_{\mathcal{H}_1}^2\right) \to 0 \text{ as } |\mathbf{z}| = m \to \infty.$$

Error bounds for nonlinear Tikhonov regularization under general source condition

Theorem

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*, for all* $0 < \eta < 1$*, with confidence* $1 - \eta$*, for the regularized estimator* $f_{\mathbf{z},\lambda}$ *the following upper bound holds:*

$$||f_{\mathbf{z},\lambda} - f_\rho||_{\mathcal{H}_1} \leq C \left\{ R\phi(\lambda) + \frac{\kappa M}{m\lambda} + \sqrt{\frac{\Sigma^2 \mathcal{N}(\lambda)}{m\lambda}} \right\} \log\left(\frac{4}{\eta}\right)$$

provided that

$$8\kappa^2 \max(1, L^2) \log(4/\eta) \leq \sqrt{m}\lambda$$

and

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

Theorem

Under the same assumptions of above theorem, for the polynomial decay condition on the eigenvalues of T and the parameter choice $\lambda \in (0,1]$, $\lambda = \Psi^{-1}(m^{-1/2})$ where $\Psi(t) = t^{\frac{1}{2}+\frac{1}{2b}}\phi(t)$, the convergence of the estimator $f_{\mathbf{z},\lambda}$ to the function f_ρ can be described as:

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

and

$$\lim_{\tau\to\infty}\limsup_{m\to\infty}\sup_{\rho\in\mathcal{P}_{\phi,b}} Prob_{\mathbf{z}}\left\{||f_{\mathbf{z},\lambda} - f_\rho||_{\mathcal{H}_1} > \tau\phi(\Psi^{-1}(m^{-1/2}))\right\} = 0.$$

Theorem

Let $\mathbf{z}$ *be i.i.d. samples drawn according to the probability measure* $\rho \in \mathcal{P}_{\phi,b}$ *under the hypothesis* $dim(Y) = d < \infty$*. Then for* $\Psi(t) = t^{\frac{1}{2}+\frac{1}{2b}}\phi(t)$*, the estimator* $f_{\mathbf{z}}$ *corresponding to any learning algorithm (*$\mathbf{z} \rightarrow f_{\mathbf{z}} \in \mathcal{H}_1$*) converges to the regression function* f_ρ *with the following lower rate:*

$$\lim_{\tau \to 0} \liminf_{m \to \infty} \inf_{l \in \mathcal{A}} \sup_{\rho \in \mathcal{P}_{\phi,b}} Prob_{\mathbf{z}} \left\{ ||f_{\mathbf{z}}^l - f_\rho||_{\mathcal{H}_1} > \tau\phi\left(\Psi^{-1}(m^{-1/2})\right)\right\} = 1.$$

Convergence rates for Tikhonov regularization under Hölder's source condition

Theorem

Under the same assumptions of above theorem, for the parameter choice $\lambda = m^{-\frac{b}{2br+b+1}}$, for all $0 < \eta < 1$, we have with confidence $1-\eta$, for the regularized estimator $f_{\mathbf{z},\lambda}$ the following convergence rate holds:

$$||f_{\mathbf{z},\lambda} - f_\rho||_{\mathcal{H}_1} \leq Cm^{-\frac{br}{2br+b+1}} \log\left(\frac{4}{\eta}\right) \text{ for } \frac{1}{2} \leq r \leq 1.$$

- **Model** $y = A(f)(x) + \varepsilon$

	$\|\|f_{\mathbf{z},\lambda} - f_\rho\|\|_{\mathcal{H}}$	Scheme	Optimal rates
Rastogi et al. (2017)	$m^{-\frac{br}{2br+b+1}}$	Direct learning	$\surd$
Blanchard et al. (2016)	$m^{-\frac{br}{2br+b+1}}$	Linear inverse learning	$\surd$
Our Results	$m^{-\frac{br}{2br+b+1}}$	Nonlinear inverse learning	$\surd$

Further questions and developments

- develope statistically and computationally effective algorithms.
- obtain confidence regions for the nonparametric model.
- 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.

N. Aronszajn, Theory of reproducing kernels. *Trans. Amer. Math. Soc.* **68** (1950) 337–404.

G. Blanchard and N. Mücke. Optimal rates for regularization of statistical inverse learning problems. *Found. Comut. Math.* **18**(4) (2018) 971–1013.

H.W. Engl, M. Hanke, and A. Neubauer. *Regularization of Inverse Problems*, volume **375**, Math. Appl., Kluwer Academic Publishers Group, Dordrecht, The Netherlands, 1996.

A. Rastogi and S. Sivananthan. Optimal rates for the regularized learning algorithms under general source condition. *Front. Appl. Math. Stat.*, **3**:3, 2017.

A. Rastogi, G. Blanchard, and P. Mathé. Convergence analysis of Tikhonov regularization for non-linear statistical inverse learning problems. *arXiv preprint arXiv:1902.05404*, 2019.

Thank you !