Regularization of statistical inverse problems

Thorsten Hohage

Dep. Numerical and Applied Math, Univ. of Göttingen

joint work with Axel Munk and Nicolai Bissantz

Dep. Stochastics, Univ. of Göttingen

“Conference on Inverse Problems: Modeling and Simulation”

Fethiye (Turkey), June 7-12, 2004

framework and notations

unknown quantity: $a \in H$, H real, separable Hilbert space
observable quantity: $u \in L^2(\Omega, \mu)$, Ω metric space, μ Borel measure

operator equation: $F(a) = u+$ "noise"

$F : D(F) \subset H \to L^2(\Omega, \mu)$ (possibly nonlinear) one-to-one operator

Unlike in a Bayesian view on inverse problems (cf. e.g. Kaipio, Päivärinta, Somersalo et al.), a is considered as an unknown parameter, not as a random variable!

exact solution: $a^\dagger \in D(F) \subset H$

exact data: $u^\dagger := F(a^\dagger)$.

statistical model

measured data are random variables

$$Y_i = u^\dagger(X_i) + v(X_i)\epsilon_i, \qquad i = 1, \ldots, n.$$

$v : \Omega \to [0, \infty)$ variance function.
ϵ_i are independent, identically distributed random variables satisfying

$$\mathbf{E}\,\epsilon_i = 0, \qquad \mathbf{Var}\,\epsilon_i = 1.$$

Both v and the distribution of the ϵ_i's are unknown!

measurement points: $X_i \in \Omega$. Either random variables, independent of ϵ_i (random design) or deterministic quantities (deterministic design).

aims

Our aim is to construct estimators

$$\begin{array}{rcl} (Y_i, X_i)_{i=1,\dots,n} & \mapsto & \hat{a}_n \\ (\mathbb{R} \times \Omega)^n & \longrightarrow & D(F) \end{array}$$

such that the expected error $\mathbf{E}\,\|\hat{a}_n - a^\dagger\|^2$ is as small as possible uniformly for all possible solutions $a^\dagger \in D(F)$.
In particular, we want the family of estimators to be consistent in the sense that

$$\lim_{n\to\infty} \mathbf{E}\,\|\hat{a}_n - a^\dagger\|^2 = 0$$

for all $a^\dagger \in D(F)$. Moreover, we would like to estimate the speed of convergence if $a^\dagger$ belongs to certain smoothness classes.

method of Ruymgaart el al. for linear problems

van Rooij & Ruymgaart 1992,1996; Mair & Ruymgaart 1996

Let $F = T$ be a linear integral operator

$$(Ta)(x) = \int k(x,t)f(t)\,\mathrm{d}t.$$

We multiply the eq. $Ta = u^\dagger$ from the left by an operator $Q : L^2(\mu) \to H$ such that QT is self-adjoint and positive definite (usually $Q = T^*$):

$$QTa = Qu$$

Obviously, for $Q = T^*$ this introduces additional ill-posedness. However, it is possible under reasonable assumptions on the kernel k to construct an estimator $\hat{q}_n$ of $q^\dagger := Qu^\dagger$ which is **unbiased, $\sqrt{n}$-consistent** ($\mathbf{E}\,\hat{q}_n = q^\dagger$ and $\mathbf{Var}\,\hat{q}_n \leq \frac{C}{n}$). In the simplest case, this estimator is given by

$$\hat{q}_n := \frac{1}{n}\sum_{i=1}^{n} k(X_i,\cdot)Y_i.$$

method of Ruymgaart el al., cont'd

An estimator of a can be obtained by applying a regularized inverse $R_\alpha \approx (QT)^{-1}$ to $\hat{q}_n$:

$$\hat{a}_{\alpha,n} := R_\alpha \hat{q}_n$$

Such regularized inverses can be constructed by any linear regularization method, e.g.

- Tikhonov regularization: $R_\alpha = (\alpha I + QT)^{-1}$
- Landweber iteration: $R_k = \sum_{j=0} k - 1(I - QT)^j$
- spectral cut-off
- ...

decomposition of the MISE

Let $f_\alpha = R_\alpha q^\dagger = \mathbf{E}\,\hat{a}_{\alpha,n}$. Then the mean integrated square error (MISE) satisfies:

$$\underbrace{\mathbf{E}\,\|\hat{a}_{\alpha,n} - a^\dagger\|^2}_{\text{MISE}} = \underbrace{\mathbf{E}\,\|\hat{a}_{\alpha,n} - a_\alpha\|^2}_{\text{variance}} + \underbrace{\|a_\alpha - a^\dagger\|^2}_{\text{bias}}$$

The bias is exactly the approximation error in deterministic theory.

The variance term corresponds to the data noise error, but should be treated differently.

Halmos version of spectral theorem

Theorem: Let $A : H \longrightarrow H$ be a, bounded, self-adjoint linear operator. Then there exists a measure space $(\mathcal{S}, \Sigma)$, a function $\rho \in L^\infty(\mathcal{S}, \Sigma)$, and a unitary operator $U : H \rightarrow L^2(\mathcal{S}, \Sigma)$ such that

$$A = U^* M_\rho U$$

where $M_\rho : L^2(\mathcal{S}, \Sigma) \rightarrow L^2(\mathcal{S}, \Sigma)$ is the multiplication operator $(M\rho\varphi)(\sigma) := \rho(\sigma)\varphi(\sigma)$.

convergence rate results

It has been shown in a number of situations that for spectral cut-off with a suitable choice of regularization parameters the MISE converges to 0 at an order-optimal rate as $n \to \infty$. (Garza, Hall, Mair, Rooij, Ruymgaart,. . .)

These cases include compact operators with singular-values behaving like $\sigma_n(T) \sim \frac{1}{n^\gamma}$, $\gamma > 0$ (mildly-ill-posed problems) or $\sigma_n(T) \sim (\ln n)^{-p}$, $p > 0$ (exponentially ill-posed problems).

For Tikhonov regularization a convergence rate analysis for mildly ill-posed problems has been provided by Nychka and Cox, 1988,89 in a special setting involving reproducing kernel Hilbert spaces.

assumptions on spectral data of operator

Assume that $\rho \in L^1(\mathcal{S}, \Sigma)$ and that $R(\alpha) := \tilde{\Sigma}(\{\rho > \alpha\})$, $\alpha \geq 0$ satisfies

1. $S(\alpha) \leq R(\alpha) \leq S(\alpha) + c$ for some $c > 0$
2. $S' \leq 0$
3. $\alpha S'(\alpha)$ is integrable
4. $\lim_{\alpha \to 0} \alpha S(\alpha) = 0$
5. There exists $C \in (0, 2)$ s.t. $S''(\alpha) \leq C \frac{-S'(\alpha)}{\alpha}$ for all $\alpha \in (0, \bar{\alpha}]$.

Motivation:

2: R is decreasing

3: $-\int_0^\infty \alpha dR(\alpha) = \int_{\mathcal{S}} \rho d\Sigma < \infty$ as $\rho \in L^1(\Sigma)$.

4. $\lim_{\alpha \to 0} \alpha R(\alpha) = \lim_{\alpha \to 0} \int_{\mathcal{S}} \alpha 1_{\rho \geq \alpha} d\Sigma = 0$ by dominated convergence

5. holds true if QT is compact with $\sigma_n(QT) \sim n^{-\gamma}$, $\gamma > 1$ or $\sigma_n(QT) \sim (\ln n)^{-p}$, $p > 0$.

convergence of spectral regularization methods

Theorem: (Bissantz, H., Munk & Ruymgaart) Under the assumptions above, the variance term of all commonly used spectral regularization methods (in particular Tikhonov regularization, Landweber iteration and ν-methods) is uniformly bounded by a constant times the variance term of spectral cut-off.

Corollary: The methods above converge at the same (optimal) rate as $n \to \infty$ if no saturation effects occur.

nonlinear statistical inverse problems

There has been a lot of work on consistency and convergence rates for linear statical inverse problems (see Evans & Stark, 2002 for a recent review), but very little on nonlinear problems.

nonlinear method of regularization (MOR):

$$\hat{a}_n = \mathrm{argmin}_{a \in D(F)} \sum_{j=1}^{n} |(Fa)(X_j) - Y_j|^2 + \alpha_n \|a - a_0\|_H^2$$

$\alpha_n > 0$ regularization parameter, a_0 a-priori guess.

O'Sullivan, 1990: consistency and convergence rate result for a special class of nonlinear problems, which are well-posed if $L^2(\Omega)$ is replaced by a Sobolev space $W^s(\Omega)$ (nonlinear extension of the approach of Nychka & Cox).

another approach

For many inverse problems in partial differential equations, the exact data function $u^\dagger \in L^2(\Omega)$ enjoys certain a-priori known smoothness properties. Therefore, we can reliably estimate $u^\dagger$ itself from the given data. However, we will have to sacrifice unbiasedness of the estimator $\hat{u}_n \in L^2(\Omega)$.

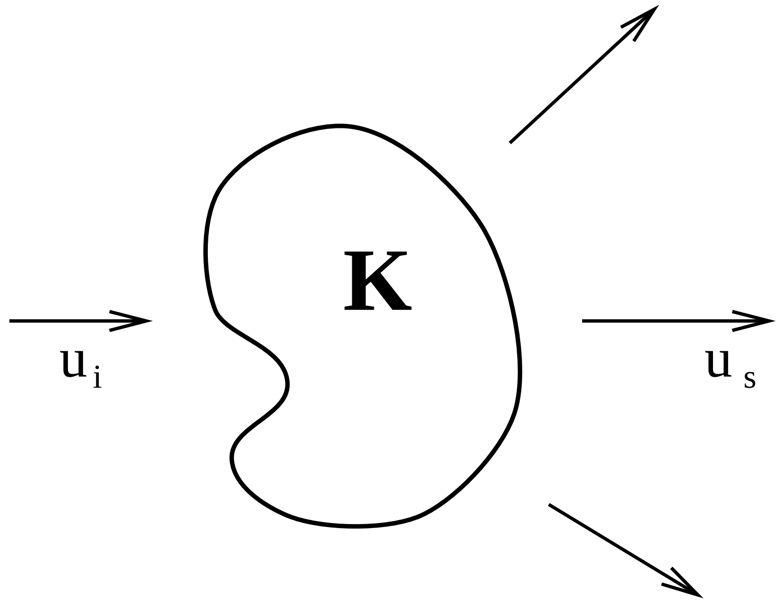
In the second step we determine an estimate $\hat{a}_n$ of $a^\dagger \in D(F)$ by computing a minimizer of the functional

$$a \mapsto \sum_{j=1}^{n} \|F(a) - \hat{u}_n\|^2 + \alpha_n \|a - a_0\|_H^2, \quad a \in D(F).$$

Again, $\alpha_n > 0$ is a regularization parameter and a_0 an a-priori guess.

example: inverse obstacle scattering problem

$K \subset \mathbb{R}^m$ star-shaped, $m = 2, 3$
$u_i(x) = e^{ikx \cdot d}$ incident wave,
$d \in \mathbb{R}^m$,$|d| = 1$ direction,
$k^{-1} \approx \operatorname{diam} K$ wave number



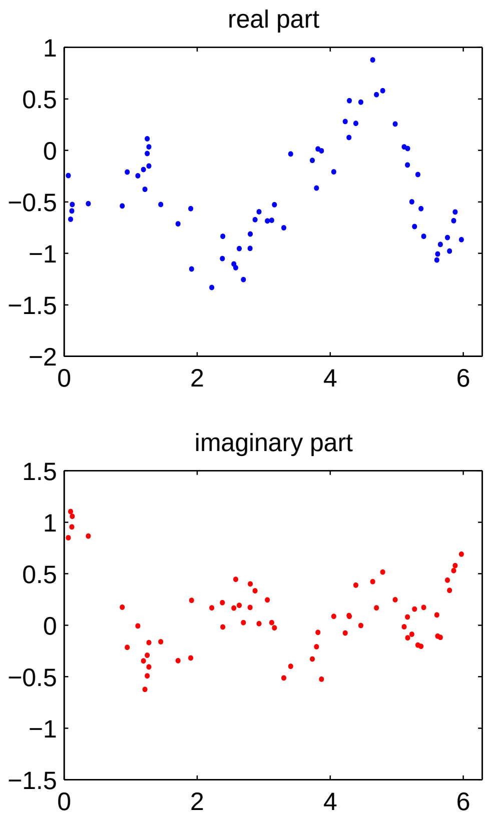
direct problem: Given K and u_i
determine a scattered field u_s such that

- $\Delta u_s + k^2 u_s = 0$ in $\mathbb{R}^m \backslash K$
- $r^{\frac{m-1}{2}} \left(\frac{\partial u_s}{\partial r} - iku_s \right) \overset{r \to \infty}{\longrightarrow} 0$
- $u = 0$ on ∂K, $u := u_i + u_s$

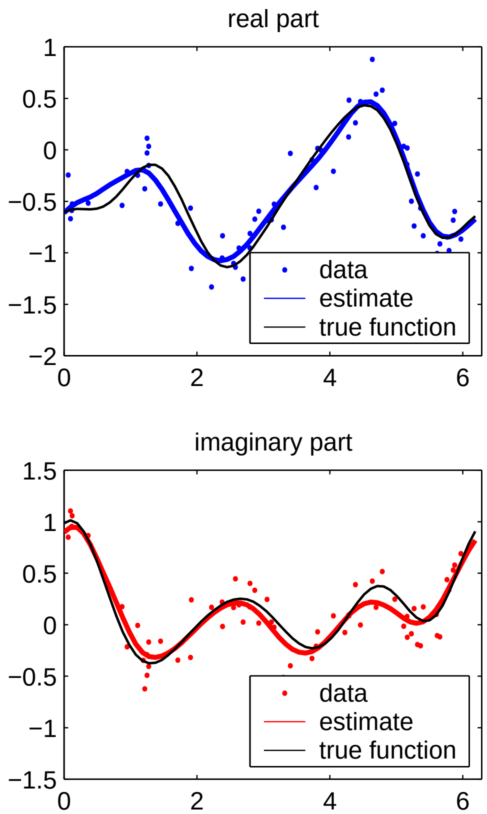
inverse problem: Given the far field pattern u_∞ of u_s and u_i, find the obstacle K!

illustation of the reconstruction algorithm

data



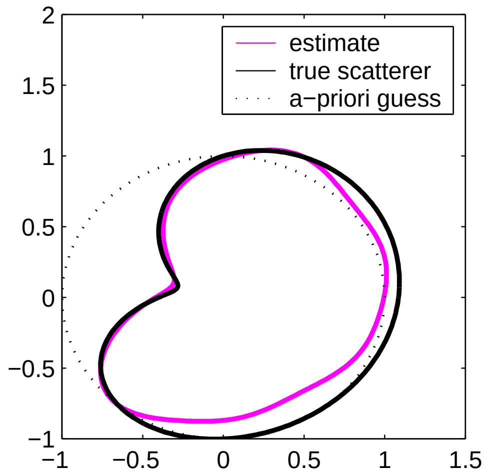
step 1: estimation of far field pattern



variance $\sqrt{v}$ estimated by 0.3005 (0.3000)

$\frac{\|\hat{u}^{\infty}_{100} - u^{\infty}\|}{\|\hat{u}^{\infty}_{100}\|}$ estimated by 0.1321 (0.1349)

step 2: estimation of shape



comparison with nonlinear MOR

The MOR has the apparent advantage that only one regularization parameter has to be chosen whereas in our method both steps involve the choice of a regularization parameter.
However, if we aim at a completely data-driven method, it is actually advantageous both for practical and theoretical reasons to tackle the simpler problem of estimating u without information on the variance of the data errors in a first step.

The choice of the regularization parameter in the second step can be done by well-known methods from deterministic regularization theory since we have obtained reliable information on the data error in the first step.

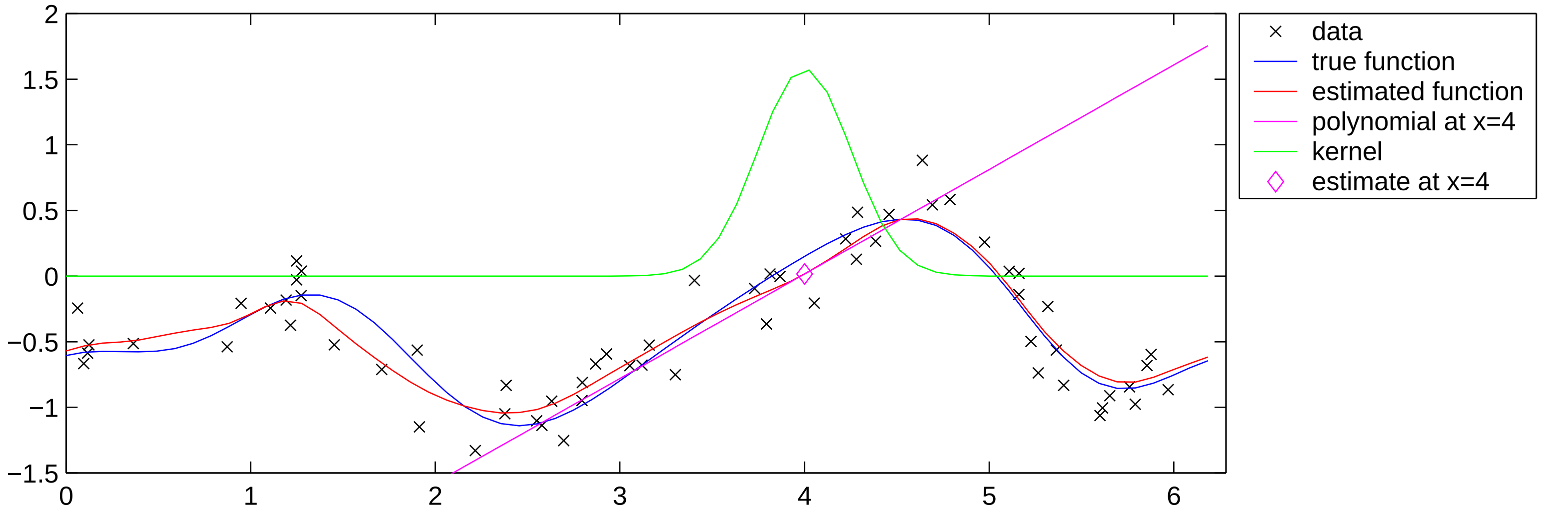
local polynomial estimators

data: $Y_i = u(X_i) + \epsilon_i \ , i = 1, \dots, n$

aim: Estimate $u \in L^2(\Omega)$, $\Omega \subset \mathbb{R}^d$ smooth, bounded

Let $K : \mathbb{R}^d \to \mathbb{R}$ be a smooth symmetric *weight function (kernel)* such that $\int_{\mathbb{R}^d} x x^T K(x)\, dx = \mu_2(K) I_d$ with $\mu_2(K) > 0$ (e.g. the Gauss bell function).
The polymial estimator of u at $x \in \Omega$ of degree $p = 0, 1, \dots$ and bandwidth $h > 0$ is defined as $\hat{u}_n(x; h, p) := \varphi(x)$ where φ is the d-variate polynomial of degree $\leq p$ which solves the minimization problem

$$\sum_{i=1}^{n} |\varphi(x - X_i) - Y_i|^2 \, h^{-d} K\left(\frac{x - X_i}{h}\right) = \min!$$



convergence of local polynomial estimators

Theorem: Assume that

- $\Omega \subset \mathbb{R}^d$ is a smooth, bounded domain and $u^\dagger \in C^2(\overline{\Omega})$.
- K is continuous, has compact support, $\int xx^T K(x)\,\mathrm{d}x = \mu_2(K)I$ for some constant $\mu_2(K) > 0$, and all odd-order moments of K vanish, i.e. $\int x_1^{\kappa_1} \cdots x_d^{\kappa_d} K(x) dx = 0$ for all combinations of integers $\kappa_1, \ldots, \kappa_d \geq 0$ with $\sum_{i=1}^d \kappa_i$ odd.
- The bandwidth $h = h(n)$ tends to 0 as $n \to \infty$ such that $nh^d \to \infty$.

Define $R(K) := \int_{\mathbb{R}^d} K(x)^2\,\mathrm{d}x$ and denote by $H_{u^\dagger}(x)$ the Hessian of $u^\dagger$ at $x \in \Omega$. Then the linear polynomial estimator is well defined with probability tending to 1 and satisfies

$$\mathbf{E}\left(\|\hat{u}_n(\cdot;h) - u^\dagger\|^2_{L^2(\Omega)} | X_1, \ldots, X_n\right) = \frac{h^4}{4}\mu_2(K)^2 \int_\Omega \left(\mathrm{Tr}\, H_{u^\dagger}(x)\right)^2\,\mathrm{d}x$$

$$+\frac{1}{nh^d}R(K)\int_\Omega v(x)\,\mathrm{d}x + o_P\left(\frac{1}{nh^d} + h^4\right).$$

estimation of the variance

Let the $n \times n$-matrix $W = (w_{ij})$ be defined through the linear mapping $\hat{u}_n(X_i) = \sum_{i=1}^n w_{ij} Y_j$, $i = 1, \ldots, n$.
The variance $v := \mathbf{Var}\, \epsilon_i$ of the measurement errors can be estimated by

$$\hat{v}_n := \frac{1}{\nu} \sum_{i=1}^{n} |Y_i - \hat{u}_n(X_i)|^2 \,.$$

The constant $\nu = n - 2\sum_{i=1}^n w_{ij} + \sum_{i=1}^n \sum_{j=1}^n w_{ij}^2$ is chosen such that $\mathbf{E}\, \hat{v}_n = v$ if u is a polynomial of degree $\leq p$.

It can be shown that the estimator $\hat{v}$ is asymptotically optimal and

$\sqrt{n}$-consistent. (Wagner, 1999)

assumptions

- uniqueness: $F(a_1) = F(a_2) \Rightarrow a_1 = a_2$
- F is weakly sequentially closed, i.e. if a sequence $(a_n)_{n\in\mathbb{N}} \subset D(F)$ convergences weakly to some $a \in D(F)$, $a_n \rightharpoonup a$ and if $F(a_n) \rightharpoonup u$ for some $u \in G$, then $a \in D(F)$ and $F(a) = u$.
- $\mathbf{E}\left\|\hat{u}_n - u^\dagger\right\|^2 = O\left(\beta_n^2\right)$ for some sequence $\beta_n \to 0$.

consistency result

Theorem: (Bissantz, H. & Munk)
Under the assumptions listed above, let $\hat{a}_n$ denote a (not necessarily unique) solution to the minimization problem

$$\|F(a) - \hat{u}_n\|^2 + \alpha_n \|a - a_0\|^2 = \min \qquad a \in D(F)$$

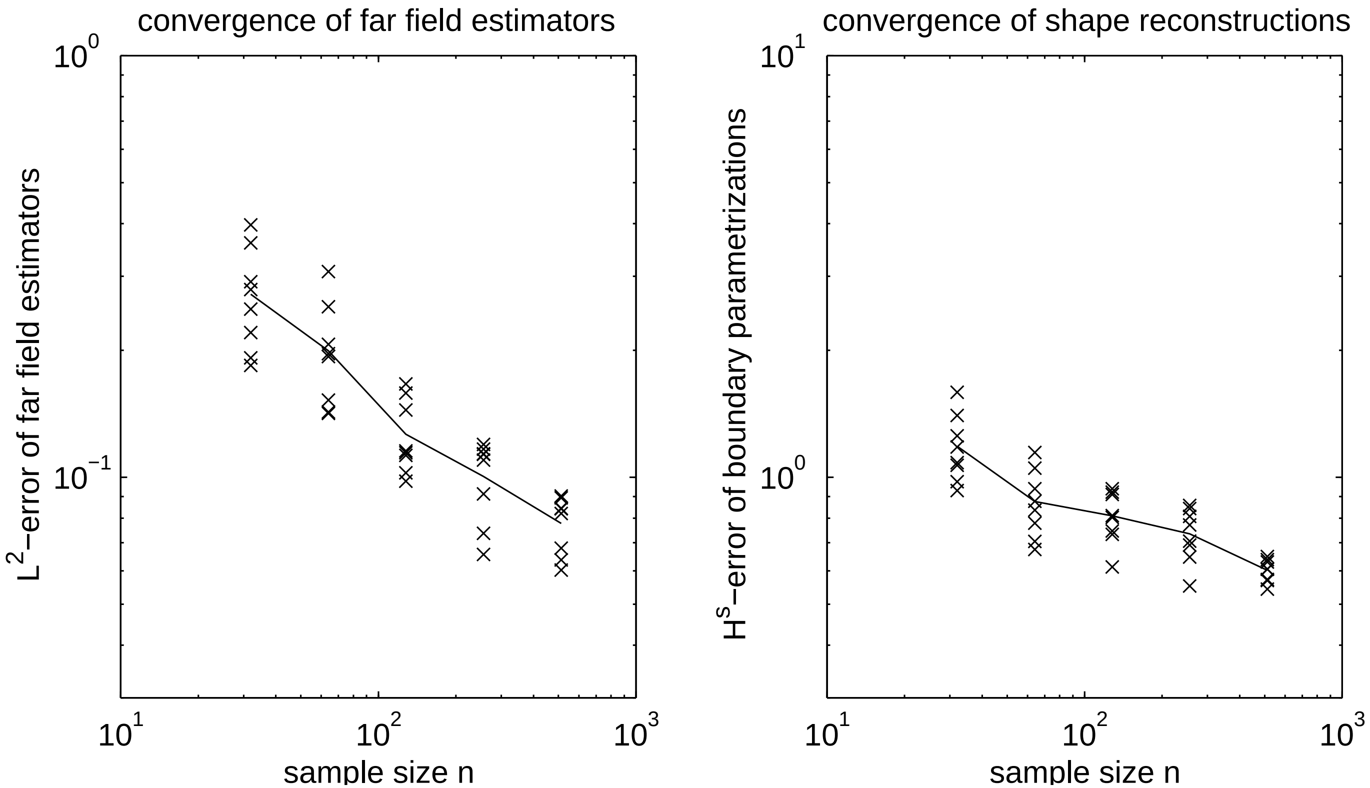
and assume that the regularization parameters $\alpha_n > 0$ are chosen such that

$$\alpha_n \to 0, \qquad \frac{\beta_n^2}{\alpha_n} \to 0 \qquad \text{as } n \to \infty.$$

Then

$$\mathbf{E}\,\|\hat{a}_n - a^\dagger\|^2 \to 0 \qquad \text{as } n \to \infty.$$

convergence plots for the scattering problem



Left: $\|\hat{u}_{n,\infty} - u_\infty^\dagger\|_{L^2}$ over n. Right: $\|\hat{a}_n - a^\dagger\|_{H^s}$ over n.
The solid lines connect the mean values for each sample size.

convergence rate result

Theorem: (Bissantz, H. & Munk) Assume that:

- F is Fréchet differentiable.
- $D(F)$ is convex, and that there exists a Lipschitz constant $L > 0$ such that
$$\|F'[a_1] - F'[a_2]\| \leq L\|a_1 - a_2\| \qquad \text{for all } a_1, a_2 \in D(F).$$
- source condition: There exists $w \in \tilde{H}$ such that
$$a^\dagger - a_0 = F'[a^\dagger]^* w, \qquad L\|w\| < 1$$
- $\mathbf{E}\,\|\hat{u}_n - u^\dagger\|^2 = O\left(\beta_n^2\right)$ for some sequence $\beta_n \to 0$.

Then, for a choice of regularization parameters satisfying $\alpha_n \sim \beta_n$, any sequence $\hat{a}_n$ of minimizers to the Tikhonov functional obeys

$$\mathbf{E}\,\|\hat{a}_n - a^\dagger\|^2 \;=\; O(\beta_n), \qquad \text{and} \qquad \mathbf{E}\,\|F(\hat{a}_n) - u^\dagger\|^2 = O(\beta_n^2).$$

(cf. result of Engl, Kunisch & Neubauer, 1989 in deterministic theory)

potential optimality of the convergence rate

Consider the linear integral operator $F : L^2([0,1]) \to L^2([0,1])$, $(Fa)(x) := \int_0^1 k(x,y)a(y)\,\mathrm{d}y$ with kernel

$$k(x,y) := \begin{cases} x(1-y), & x \le y, \\ y(1-x), & y > x. \end{cases}$$

The source condition $a^\dagger \in R(F)$ is equivalent to

$$(u^\dagger)^{(4)} \in L^2([0,1]), \qquad u^\dagger(0) = u^\dagger(1) = (u^\dagger)''(0) = (u^\dagger)''(1) = 0.$$

For this smoothness class, cubic polynomial estimators with bandwidth $h \sim n^{-1/9}$ convergence with the rate

$$\mathbf{E}\,\|\hat{u}_n - u^\dagger\|^2 = O(n^{-8/9}),$$

i.e. $\beta_n = n^{-4/9}$. Hence, by the previous theorem,

$$\mathbf{E}\,\|\hat{a}_n - a^\dagger\|^2 = O\left(n^{-4/9}\right).$$

potential optimality, cont'd

On the other hand, Mair & Ruymgaart (1996) have shown the lower bound (or *mini-max rate*)

$$\inf_{\hat{a}_n \in \mathcal{A}_n} \sup_{a^\dagger \in \{Fw : \|w\| \leq 1\}} \mathbf{E}\, \|\hat{a}_n - a^\dagger\|^2 \geq c n^{-4/9}.$$

Here $c > 0$ is a constant, and $\mathcal{A}_n$ denotes the class of all mappings $\hat{a}_n : ([0,1] \times \mathbb{R})^n \to L^2([0,1])$.

Since the upper bound is of the same order as the lower bound, the convergence of our method is optimal up to a constant for this example.

a groundwater filtration problem

a	diffusitivity of the sediment
u	piezometric head
g	sinks and sources of water
$\Omega \subset \mathbb{R}^d$	smooth domain

differential equation:

$$\nabla a \nabla u = g \quad \text{in } \Omega$$
$$u = 0 \quad \text{on } \partial\Omega.$$

We introduce the parameter-to-solution operator

$F : D(F) \subset H_0^s(\Omega) \to L^2(\Omega)$ with $D(F) := \{a \in H_0^s(\Omega) : a \geq \underline{a}\}$

where $s > d/2$ and $\underline{a} \in (0, \infty)$ by

$$F(a) := u.$$

verification of the assumptions

- uniqueness: Richter, 1981 under the assumption that the right hand side g be positive and Hölder continuous.
- weak closedness ... an argument involving compactness of embedding operators $H^s \hookrightarrow H^{s'}$ with $s > s' > d/2$
- F is Fréchet differentiable, and the differential equation can be formally differentiated w.r.t. a:

$$\nabla a \nabla (u'[a]h) = -\nabla h \nabla u, \qquad u|_{\partial\Omega} = 0$$

- F' is Lipschitz continuous since F is twice continuously differentiable.
- source condition: For $d = 1$ the source condition can be interpreted as a smoothness condition (see Engl, Hanke, Neubauer, 96). For $d \geq 2$ a precise characterization is not known.

groundwater filtration problem: results