Convergence Analysis of (Statistical) Inverse Problems under Conditional Stability Estimates

Frank Werner

joint work with Bernd Hofmann

Institut für Mathematische Stochastik

Universität Göttingen

Outline

Model

In this talk we consider **nonlinear** inverse problems

$$F(f) = g$$

with $F : D(F) \subset \mathcal{X} \to \mathcal{Y}$, $\mathcal{X}, \mathcal{Y}$ Hilbert spaces in the data model

$$g^{\mathrm{obs}} = g^\dagger + \sigma Z + \delta \xi$$

with

- exact data $g^\dagger = F\left(f^\dagger\right)$
- random noise given by a (centered) **noise process** $Z : \mathcal{Y} \to \mathbf{L}^2(\Omega, \mathcal{A}, \mathbb{P})$
- deterministic noise $\xi \in \mathcal{Y}$, $\|\xi\|_{\mathcal{Y}} \leq 1$
- and corresponding noise levels $\sigma, \delta \geq 0$.

Model (cont')

$$g^{\mathrm{obs}} = g^\dagger + \sigma Z + \delta\xi$$

As $Z \notin \mathcal{Y}$, whole model has to be understood in the weak sense:
For each $g \in \mathcal{Y}$ we can access the corresponding *coefficient*

$$\left\langle g^{\mathrm{obs}}, g \right\rangle = \left\langle g^\dagger, g \right\rangle + \sigma \langle Z, g\rangle + \delta \langle \xi, g\rangle$$

under additive

deterministic noise $\delta \langle \xi, g\rangle \leq \delta \|g\|_{\mathcal{Y}}$ and

random noise $\sigma \langle Z, g\rangle$ - a centered rv with finite variance

Note: For $g_1, g_2 \in \mathcal{Y}$ it holds

$$\mathrm{Cov}\left[\langle Z, g_1\rangle, \langle Z, g_2\rangle\right] = \langle g_1, \mathrm{Cov}\left[Z\right] g_2\rangle$$

with the covariance operator $\mathrm{Cov}\left[Z\right] \in \mathcal{L}\left(\mathcal{Y}\right)$.

Examples

$$g^{\mathrm{obs}} = g^{\dagger} + \sigma Z + \delta\xi$$

allows for

- completely deterministic noise models: $\sigma = 0$
- continuous Gaussian white noise models: $\delta = 0$, $\mathrm{Cov}\,[Z] = \mathrm{id}$, Z Gaussian process
- mixtures of both, e.h. discretized Gaussian white noise models (with ξ being the normalized discretization error)

N. Bissantz, T. Hohage, A. Munk, and F. Ruymgaart.
Convergence rates of general regularization methods for statistical inverse problems and applications.
SIAM Journal on Numerical Analysis 45(6): 2610-2636, 2007

Hilbert scales

We assume that there is a Hilbert scale $\{\mathcal{X}_\nu\}_{\nu\in\mathbb{R}}$ such that

- $\mathcal{X}_\nu := D\left(L^\nu\right)$ with a densely defined linear self-adjoint $L : D\left(L\right) \subset \mathcal{X} \to \mathcal{X}$
- $\mathcal{X}_0 := \mathcal{X}$
- $\left\|f\right\|_\nu := \left\|L^\nu f\right\|_{\mathcal{X}}$
- Interpolation for $-a < t \leq s$:

$$\left\|f\right\|_t \leq \left\|f\right\|_{-a}^{\frac{s-t}{s+a}} \left\|f\right\|_s^{\frac{t+a}{s+a}}$$

Regularization

We consider variational regularization of the form

$$\hat{f}_\alpha \in \operatorname*{argmin}_{f \in D(F)} \left[\frac{1}{2} \|F(f)\|_{\mathcal{Y}}^2 - \left\langle F(f), g^{\mathrm{obs}} \right\rangle + \alpha \|f\|_s^2 \right]$$
$$\hat{=} \operatorname*{argmin}_{f \in D(F)} \left[\frac{1}{2} \left\| F(f) - g^{\mathrm{obs}} \right\|_{\mathcal{Y}}^2 + \alpha \|f\|_s^2 \right]$$

Well-known properties of minimizers (if $D(F)$ is closed and convex, and F is weak-to-weak sequentially continuous):

- existence (with probability 1 in the random noise case)
- stability if $s \geq 0$ (w.r.t. ξ clear, w.r.t. Z more complicated)
- convergence in the deterministic noise case (if α is chosen appropriately)

Here we will focus on **rates of convergence**!

Outline

Main assumption

Conditional stability estimate (CSE)

There exists a concave **index function** φ, a set Q and a constant $R > 0$ such that

$$\left\| f - f^{\dagger} \right\|_{-a} \leq R \varphi \left(\left\| F(f) - g^{\dagger} \right\|_{\mathcal{Y}} \right)$$

for all $f \in Q$.

- If F is linear, $Q = \mathcal{X}$ and $\varphi = \mathrm{id}$, then this corresponds to ill-posedness of degree a
- Usually Q is a subset of some unit ball in $\mathcal{X}_\theta$ with some $\theta > 0$.

Sufficient conditions for CSE

$$\left\| f - f^\dagger \right\|_{-a} \leq R\varphi\left(\left\| F\left(f\right) - g^\dagger \right\|_{\mathcal{Y}} \right)$$

For differentiable F, this condition is satisfied if

(A) F' is ill-posed of degree a

$$\|h\|_{-a} \leq \bar{K}\, \|F'(f^\dagger)\, h\|_{\mathcal{Y}}$$

and a tangential-cone-type condition holds true

$$\|F'(f^\dagger)(f - f^\dagger)\|_{\mathcal{Y}} \leq \tilde{K}\, \varphi(\|F(f) - F(f^\dagger)\|_{\mathcal{Y}})$$

(B) F' is ill-posed of degree a and a Hölder continuity condition holds true

$$\|F'(f) - F'(f^\dagger)\|_{\mathcal{L}(\mathcal{X},\mathcal{Y})} \leq \check{K}\, \|f - f^\dagger\|^\eta$$

CSE vs. variational source conditions

$$\left\|f - f^\dagger\right\|_{-a} \leq R\varphi\left(\left\|F(f) - g^\dagger\right\|_{\mathcal{Y}}\right) \quad \text{for all} \quad f \in Q \qquad (1)$$

CSE looks similar to variational source conditions:

$$\left\|f - f^\dagger\right\|_s^2 \leq \|f\|_s^2 - \left\|f^\dagger\right\|_s^2 + R\varphi\left(\left\|F(f) - g^\dagger\right\|_{\mathcal{Y}}\right) \quad \text{for all} \quad f \in M \quad (2)$$

But depending on M, the roles of f and $f^\dagger$ might not be interchangeable (hence no relation in general)

If M allows for interchanging f and $f^\dagger$, then (2) implies (1)

Both can be verified using global estimates in specific problems:

T. Hohage and F. Weidling.
Verification of a variational source condition for acoustic inverse medium scattering problems.
Inverse Problems 31:075006, 2015.

A prominent example

Consider the identification of f in

$$\begin{aligned} \frac{\partial}{\partial t} u(x,t) - \Delta u(x,t) + f(t)\, u(x,t) &= 0 && \text{in } \Omega \times (0,T] \\ \frac{\partial}{\partial n} u(x,t) &= 0 && \text{on } \partial\Omega \times (0,T] \\ u(x,0) &= u_0(x) && \text{on } \Omega \end{aligned}$$

from integral data $g(t) = \int_\Omega u(x,t)\, \mathrm{d}x$.
Then for each $\frac{1}{2} \leq \theta \leq 1$ and each $\rho > 0$ there exists $C(\rho) > 0$ such that

$$\left\| f - f^\dagger \right\|_{L^2} \leq R(\rho) \left\| F(f) - F\left(f^\dagger\right) \right\|_{L^2}^{\frac{\theta}{\theta+1}}$$

for all $f, f^\dagger \in Q := \left\{ u \in H^\theta : \|u\|_\theta \leq \rho \right\}$.

B. Hofmann and M. Yamamoto.
On the interplay of source conditions and variational inequalities for nonlinear ill-posed problems.
Appl. Anal., 89:1705–1727, 2010.

Smoothness conditions

With

- a from the CSE $\left\|f - f^\dagger\right\|_{-a} \leq R\varphi\left(\left\|F(f) - g^\dagger\right\|_{\mathcal{Y}}\right)$
- and s as in our regularization $\hat{f}_\alpha \in \operatorname{argmin}_{f\in D(F)}\left[\frac{1}{2}\|F(f)\|_{\mathcal{Y}}^2 - \left\langle F(f), g^{\text{obs}}\right\rangle + \alpha\|f\|_s^2\right]$

we assume:

Smoothing properties

Furthermore $f^\dagger \in \mathcal{X}_u$ is the unique solution to $F\left(f^\dagger\right) = g^\dagger$ and the indices satisfy

- $a \geq 0$ (smoothing property of F)
- $-a < s < u$ (smoothing of the regularization, undersmoothing case)
- $u \leq 2s + a$ (smoothness of $f^\dagger$, saturation)

Smoothing properties of F

In the random noise case ($\sigma > 0$) we assume additionally:

Interpolation in $\mathcal{Y}$

- There is a Gelfand triple $(\mathcal{V}, \mathcal{Y}, \mathcal{V}')$, $\iota : \mathcal{V} \hookrightarrow \mathcal{Y}$ s.th. $\iota^* \mathrm{Cov}\,[Z]\,\iota$ is trace-class.
- F satisfies the interpolation inequality

$$\left\| F(f) - g^\dagger \right\|_{\mathcal{V}} \leq C \left\| F(f) - g^\dagger \right\|_{\mathcal{Y}}^{\theta} \left\| f - f^\dagger \right\|_{s}^{1-\theta}$$

for all $f \in Q$ with some constant $C > 0$ and $\theta \in (0,1)$.

Note: Assumption implies

$$\mathbb{E}\left[\|Z\|_{\mathcal{V}'}^2 \right] = \mathrm{trace}\left(\iota^* \mathrm{Cov}\,[Z]\,\iota\right) < \infty,$$

i.e. especially $\|Z\|_{\mathcal{V}'} < \infty$ a.s.

Examples

$$\left\| F(f) - g^\dagger \right\|_{\mathcal{V}} \leq C \left\| F(f) - g^\dagger \right\|_{\mathcal{Y}}^{\theta} \left\| f - f^\dagger \right\|_{s}^{1-\theta}$$

This condition holds true whenever

- $(\mathcal{V}, \mathcal{Y}, \mathcal{V}')$ is part of a Hilbert scale $\{\mathcal{Y}_\mu\}_{\mu \in \mathbb{R}}$ (i.e. $\mathcal{V} = \mathcal{Y}_t$, $\mathcal{V}' = \mathcal{Y}_{-t}$, $\mathcal{Y} = \mathcal{Y}_0$) and
- $F : \mathcal{X}_s \to \mathcal{Y}_r$ is Lipschitz continuous for some $r > t$

Proof:

$$\begin{aligned} \left\| F(f) - g^\dagger \right\|_{\mathcal{V}} &\leq \left\| F(f) - g^\dagger \right\|_{\mathcal{Y}}^{\theta} \left\| F(f) - g^\dagger \right\|_{\mathcal{Y}_r}^{1-\theta} \\ &\leq L^{1-\theta} \left\| F(f) - g^\dagger \right\|_{\mathcal{Y}}^{\theta} \left\| f - f^\dagger \right\|_{s}^{1-\theta} \end{aligned}$$

Classical example: Sobolev spaces $\mathcal{V} = H^s(\Omega)$, $\mathcal{Y} = \mathbf{L}^2(\Omega)$, $s > d/2$.

Outline

Notation

- For an index function h denote by

$$h^*(y) := \sup_{x \geq 0} [xy - h(x)]$$

the **Fenchel** conjugate of h.

- Introduce

$$\psi_{u,s,a}(t) := \left(\varphi\left(\sqrt{t}\right)\right)^{\frac{2(u-s)}{a+u}}, \qquad t \geq 0$$

and

$$\varphi_{\mathrm{app}}(\alpha) := (-\psi_{u,s,a})^* \left(-\frac{1}{\alpha}\right), \qquad \alpha \geq 0$$

Deterministic noise case ($\sigma = 0$)

Error estimates

Whenever $\hat{f}_\alpha \in Q$ (validity set of CSE), then it holds

$$\left\| \hat{f}_\alpha - f^\dagger \right\|_s^2 \leq \frac{\delta^2}{\alpha} + C\varphi_{\mathrm{app}}\left(8C\alpha\right)$$

with a constant $C = C\left(R, \left\|f^\dagger\right\|_u, u, s, a\right)$.

The function φ_{app} (approximation error) is

- non-negative, monotonically increasing, $\varphi_{\mathrm{app}}\left(\alpha\right) \to 0$ as $\alpha \to 0$
- satisfies $\varphi_{\mathrm{app}}\left(C\alpha\right) \leq \max\left\{1, C^{\frac{u-s}{a+s}}\right\} \varphi_{\mathrm{app}}\left(\alpha\right)$ for all $\alpha, C > 0$.

Note: Assumption $\hat{f}_\alpha \in Q$ can be verified using convergence statements!

Deterministic noise case ($\sigma = 0$)

Convergence rates

Assume additionally that $\psi_{u,s,a}$ is concave, and choose $\alpha = \alpha_*$ such that

$$-\frac{1}{\alpha^*} \in \partial \left(-\psi_{u,s,a}\right) \left(\delta^2\right).$$

Then we obtain the convergence rate

$$\left\| \hat{f}_{\alpha_*} - f^\dagger \right\|_s = \mathcal{O}\left(\sqrt{\psi_{u,s,a}\left(\delta^2\right)}\right) = \mathcal{O}\left(\left(\varphi\left(\delta\right)\right)^{\frac{u-s}{a+u}}\right) \qquad \text{as} \qquad \delta \to 0.$$

- If F is linear and φ a power function (Hölder case), then this rate is order optimal.
- A posteriori choice (Lepskiĭ) is also possible.

Random noise case ($\sigma > 0$)

Error estimates

Whenever $\hat{f}_\alpha \in Q$ (validity set of CSE), then it holds (surely) that

$$\left\| \hat{f}_\alpha - f^\dagger \right\|_s^2 \leq C \left[\sigma^2 \left\| Z \right\|_{\mathcal{V}'}^2 \alpha^{\theta - 2} + \frac{\delta^2}{\alpha} + \varphi_{\mathrm{app}} \left(8 C \alpha \right) \right]$$

with a constant $C > 0$.

- $\|Z\|_{\mathcal{V}'} < \infty$ a.s.. If Z is Gaussian noise and $\mathcal{V}$ 'nice', then $\mathbb{P}\left[\left| \|Z\|_{\mathcal{V}'} - \mathbf{E}\left[\|Z\|_{\mathcal{V}'} \right] \right| \geq t \right] \leq 2 \exp\left(-\frac{2t^2}{\pi^2 \sigma^2} \right)$, $t \geq 0$

E. Giné and R. Nickl.
Mathematical Foundations of Infinite-Dimensional Statistical Models.
Cambridge University Press, 2015.

- Assumption $\hat{f}_\alpha \in Q$ might be problematic, as consistency is unclear (can however be verified if local Lipschitz constants of F do not increase too fast).

Random noise case ($\sigma > 0$)

Convergence rates

Let

$$\Sigma(\alpha) = \sqrt{\alpha}\sqrt{\varphi_{\mathrm{app}}(\alpha)} \qquad \text{and} \qquad \widetilde{\Sigma}(\alpha) = \alpha^{1-\frac{\theta}{2}}\sqrt{\varphi_{\mathrm{app}}(\alpha)}, \quad \alpha > 0$$

and choose α such that

$$\alpha \sim \left(\Sigma^{-1}(\delta) + \widetilde{\Sigma}^{-1}(\sigma)\right) \qquad \text{as} \qquad \max\{\delta,\sigma\} \to 0.$$

Then we obtain the a.s. convergence rate

$$\left\|\hat{f}_\alpha - f^\dagger\right\|_s = \mathcal{O}\left(\sqrt{\varphi_{\mathrm{app}}\left(\Sigma^{-1}(\delta) + \widetilde{\Sigma}^{-1}(\sigma)\right)}\right)$$

as $\max\{\delta,\sigma\} \to 0$.

Outline

Conclusion and outlook

- Statistical Inverse Problems differ from deterministic ones ...
  - ... by the fact, that the data is not an element of the space $\mathcal{Y}$.
  - ... and hence regularization has to be treated differently.
- Convergence analysis under conditional stability estimates ...
  - ... avoids nonlinearity assumptions on the operator.
  - ... can be carried out both in the deterministic and random noise case.
- Future research should address ...
  - ... validity of conditional stability estimates ($\rightsquigarrow Q$ and φ).
  - ... what can be done to ensure $\hat{f}_\alpha \in Q$ under random noise?

F. Werner and B. Hofmann.
Convergence Analysis of (Statistical) Inverse Problems under Conditional Stability Estimates.
arXiv preprint 1905.09765, 2019.

Thank you for your attention!