Convergence rates for stochastic inverse problems using variational methods

Benjamin Sprung,
joint work with Thorsten Hohage and Frederic Weidling

Institute for numerical and applied mathematics
Georg-August Universität Göttingen

London, December 16, 2017

SFB 755

inv opt RTG 2088 stat

Outline

Inverse problems

... consists of finding causes for observed effects.

Notation

- *$f^\dagger$: the unknown cause*
- *T: the forward operator. Takes a possible cause as input and gives the corresponding effect as output:*

$$T : f^\dagger \mapsto g^\dagger$$

- *T describes the experimental setup*

Ill-posed: T^{-1} is not continuous and instead of the true data $g^\dagger$ we are given experimentally observed noisy data g^{obs} with a certain bound on the noiselevel, e.g.

$$\|g^{\text{obs}} - g^\dagger\| \leq \delta.$$

Convergence rates

Let R be some regularization method.

- we are not concerned with convergence of an iterative algorithm
- $R(g^{\text{obs}}) = f^{\dagger}$ is not possible for fixed noisy data g^{obs}
- however possible to show: $\|R(g^{\text{obs}}) - f^{\dagger}\| \leq \phi(\delta)$

What is the use of convergence rates w.r.t. the noiselevel?

Convergence rates

Let R be some regularization method.

- we are not concerned with convergence of an iterative algorithm
- $R(g^{\text{obs}}) = f^\dagger$ is not possible for fixed noisy data g^{obs}
- however possible to show: $\|R(g^{\text{obs}}) - f^\dagger\| \leq \phi(\delta)$

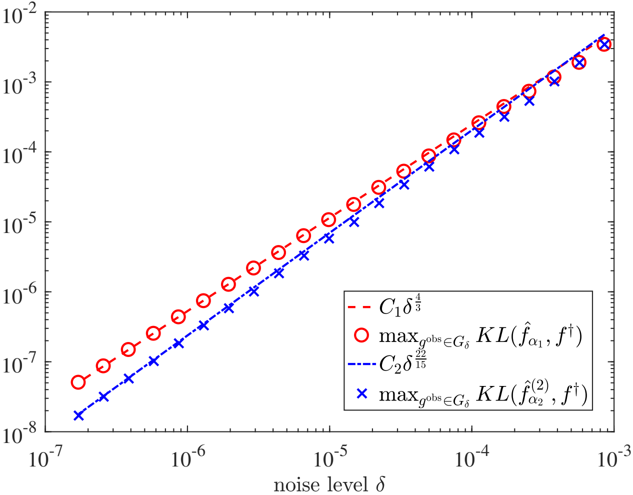
What is the use of convergence rates w.r.t. the noiselevel?
They allow to compare different regularization methods:
Assume you have two methods R_1, R_2, with

$$\|R_1(g^{\text{obs}}) - f^\dagger\| \leq \phi_1(\delta), \quad \|R_2(g^{\text{obs}}) - f^\dagger\| \leq \phi_2(\delta)$$

with

$$\phi_1(\delta) \leq \phi_2(\delta)$$

Example



T. Hohage, S. *Higher order convergence rates for Bregman iterated variational regularization of inverse problems*. **arXiv:1710.09244** 2017.

Outline

Noise

Forward operator: $T\colon X \to Y$, $f^\dagger \mapsto g^\dagger$

We are given observed noisy data G^{obs}, which are not necessarily an element of Y.

Example

- $G^{obs} = g^{obs} \in Y$, $\|g^\dagger - g^{obs}\| \leq \delta$
- $G^{obs} = g^\dagger + \varepsilon W$, *W is a Gaussian white noise*
- $G^{obs} = G_t$, *G_t is a Poisson process*

Regularization method

We apply generalized Tikhonov regularization

$$R(G^{\text{obs}}) = \hat{f}_\alpha \in \text{argmin}_{f \in X} \left[\mathcal{S}_{G^{\text{obs}}}(Tf) + \alpha \mathcal{R}(f) \right],$$

where $\mathcal{S}_{G^{\text{obs}}}$ and $\mathcal{R}$ are convex, lower-semicontinuous functionals.

Example

- $\mathcal{S}_{g^{obs}}(Tf) = \frac{1}{2}\|Tf - g^{obs}\|^2$
- $\mathcal{S}_{g^\dagger + \varepsilon W}(Tf) = \frac{1}{2}\|Tf\|^2 + \langle g^\dagger + \varepsilon W, Tf\rangle$
- $\mathcal{S}_{G_t}(Tf) = -\ln \mathbb{P}_{Tf}(G_t)$

Effective noise level

deterministic data: $\|g^\dagger - g^{\text{obs}}\| \leq \delta$.
For stochastic data we need a more general definition:

Definition (Hohage, Werner)

Let $\mathcal{T}_{g^\dagger}: Y \to [0, \infty)$, $C_{\text{err}} \geq 1$, for $g \in Y$ define

$$\mathbf{err}(g) := \mathcal{S}_{G^{\text{obs}}}(g^\dagger) - \mathcal{S}_{G^{\text{obs}}}(g) + \frac{1}{C_{\text{err}}} \mathcal{T}_{g^\dagger}(g).$$

Example

- *deterministic:* $\mathcal{T}_{g^\dagger}(g) = \frac{1}{2}\left\|g - g^\dagger\right\|^2$, $\mathbf{err}(g) \leq \delta^2 \, \forall g \in Y$
- *white noise:* $\mathcal{T}_{g^\dagger}(g) = \frac{1}{2}\left\|g - g^\dagger\right\|^2$, $\mathbf{err}(g) = \langle \varepsilon W, g - g^\dagger \rangle$
- *Poisson:*
  $\mathcal{T}_{g^\dagger}(g) = \mathrm{KL}(g^\dagger, g)$, $\mathbf{err}(g) = \int \ln\left(\frac{g}{g^\dagger}\right)\left(\mathrm{d}G_t - g^\dagger \, \mathrm{d}x\right)$

Convergence rates

Theorem (Hohage, Werner)

Let $f^\dagger$ satisfy a variational source condition (VSC), i.e.

$$\langle f^\dagger, f - f^\dagger \rangle \leq \frac{1}{4}\|f - f^\dagger\|^2 + \psi\left(\mathcal{T}_{g^\dagger}(Tf)\right),$$

then we have

$$\|f - f^\dagger\|^2 \leq \frac{2\,\mathbf{err}(T\hat{f}_\alpha)}{\alpha} + (-\psi)^*\left(-\frac{1}{C_{\mathrm{err}}\alpha}\right).$$

For $X = L^2(\mathbb{T}^d)$ the above VSC is equivalent to the smoothness condition of $f^\dagger$ belonging to a certain Besov space (depending on the Forward operator and the function ψ).

T. Hohage, F. Werner *Convergence rates in expectation for Tikhonov-type regularization of inverse problems with Poisson data*. **Inverse Problems** 28, 2012.

Approach for white noise

From now on we assume $Y = L^2(\mathbb{T}^d)$. It was shown by Veraar, that $W \in B^{-d/2}_{2,\infty}(\mathbb{T}^d)$ a.s., so we have

$$\mathbf{err}\left(T\hat{f}_\alpha\right) \le \langle \varepsilon W, T\hat{f}_\alpha - g^\dagger \rangle \le \varepsilon \left\| W \right\|_{B^{-d/2}_{2,\infty}} \left\| T\hat{f}_\alpha - g^\dagger \right\|_{B^{d/2}_{2,1}}$$

and $\left\| W \right\|_{B^{-d/2}_{2,\infty}}$ can be controlled by the following concentration inequality.

Theorem (Veraar)

Let M be the median of $\left\| W \right\|_{B^{-d/2}_{2,\infty}}$, then we have for all $r > 0$ that

$$\mathbb{P}\left(\left| \left\| W \right\|_{B^{-d/2}_{2,\infty}} - M \right| > r\right) \le \exp\left(-r^2/4\right)$$

M. Veraar *Regularity of Gaussian white noise on the d-dimensional torus.* **Banach Center Publications** 95, 2011.

Outline

Models for the forward operator

We treat two different models:

- an a-times smoothing forward operator T_a in the sense that for all $s \in \mathbb{R}$
$$T_a : H^s(\mathbb{T}^d) \to H^{s+a}(\mathbb{T}^d)$$
is well-defined, bounded and has a bounded inverse (for example $T_a = (I - \Delta)^{-a/2}$).
- the backwards heat equation on $\mathbb{T}^d$, that is $(T_{\mathrm{BH}} f)(x) = u(x, \bar{t})$ for some fixed $\bar{t} > 0$ where u solves
$$\begin{aligned} \partial_t u &= \Delta u && \text{in } \mathbb{T}^d \times (0, \bar{t}) \\ u(\cdot, 0) &= f && \text{on } \mathbb{T}^d. \end{aligned}$$
Note that T_{BH} can be conveniently expressed via the Fourier transform $\mathcal{F}$ as
$$T_{\mathrm{BH}} f = \mathcal{F}^* \exp\left(-\bar{t}|\cdot|^2\right) \mathcal{F} f.$$

Optimal rates in Hilbert spaces

Let $X = Y = L_2(\mathbb{T}^d)$, $\mathcal{R} = \|\cdot\|_{L_2}^2$, let $\|f^\dagger\|_{B^s_{2,\infty}} \leq \varrho$.

Theorem (Hohage, Weidling, S)

Let the forward operator be given by T_a. For an optimal choice of α and for $s \in (0, a)$ we find the following convergence rate

$$\mathbb{P}\left[\|f^\dagger - \hat{f}_\alpha\|_{L^2} > (c + r)\varrho^{\frac{a+d/2}{s+a+d/2}}\varepsilon^{\frac{s}{s+a+d/2}}\right] \leq \exp(-Cr^2),$$

which implies $\mathbb{E}\left(\|f^\dagger - \hat{f}_\alpha\|_{L^2}\right) \leq C\varrho^{\frac{a+d/2}{s+a+d/2}}\varepsilon^{\frac{s}{s+a+d/2}}$.

Optimal rates in Hilbert spaces (2)

Let $X = L_2(\mathbb{T}^d)$, $\mathcal{R} = \|\cdot\|_{L_2}^2$, let $\|f^\dagger\|_{B^s_{2,\infty}} \leq \varrho$.

Theorem (Hohage, Weidling, S)

Let the forward operator be given by T_{BH}. For an optimal choice of α we find the following convergence rate

$$\mathbb{P}\left[\|f^\dagger - \hat{f}_\alpha\|_{L^2} > c\left(\varrho\left(\ln\left(\frac{\varrho}{\varepsilon}\right)\right)^{-s/2} + r\right)\right] \leq \exp\left(-C\varepsilon^{-\frac{1}{2}} r^2\right)$$

which implies $\mathbb{E}\left(\|f^\dagger - \hat{f}_\alpha\|_{L^2}\right) \leq C\varrho\left(\ln\left(\frac{\varrho}{\varepsilon}\right)\right)^{-s/2}$.

These rates for T_a and T_{BH} for Tikhonov regularization on Hilbert spaces were found and proved to be optimal in

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), 2007.

Poisson data

For Poisson data the approach is similar. One additionally needs:

Theorem (Hohage, Simayi)

There exists a constant c such that for r sufficiently large

$$\mathbb{P}\left(\left\| \mathrm{d}G_t - g^\dagger \,\mathrm{d}x \right\|_{B^{-d/2}_{2,\infty}} > \frac{r + C}{\sqrt{t}}\right) \leq 2 \exp\left(-cr\right).$$

With this they find for $T = T_a$

Theorem (Hohage, Simayi)

$$\mathbb{E}\left(\left\| \hat{f}_\alpha - f^\dagger \right\|\right) = \mathcal{O}\left(\sqrt{t}^{\frac{s}{s+a+d/2}}\right).$$

Besov penalty term

An interesting choice for the penalty term is $\mathcal{R} = \|\cdot\|_{B^0_{p,q}}$.

- can be seen as wavelet regularization
- promotes sparsity

We show for the white noise model the following

Theorem (Hohage, Weidling, S)

Let $p \in (1,2]$, $q \in (1,\infty)$. Let T_a be as above and $f^\dagger \in B^s_{p,\infty}$ for $s \in (0, \frac{a}{q-1})$. For $q \leq 2$ there exists c s.t.

$$\mathbb{P}\left[\|f^\dagger - \hat{f}_\alpha\|_{B^0_{p,q}} > (c+r)\varrho^{\frac{a+d/2}{s(q-1)+a+d/2}}\varepsilon^{\frac{s(q-1)}{s(q-1)+a+d/2}}\right] \leq \exp(-Cr^2),$$

for $q \geq 2$ there exists c s.t.

$$\mathbb{P}\left[\|f^\dagger - \hat{f}_\alpha\|_{B^0_{p,q}} > (c+r)\varrho^{\frac{a+d/2}{s+a+d/2}}\varepsilon^{\frac{s}{s+a+d/2}}\right] \leq \exp\left(-Cr^{\left(q+\frac{(q-2)d}{2a}\right)}\right).$$

Thank you for your attention!