Iterative Regularization Methods for Inverse Problems: Lecture 2

Thorsten Hohage

Institut für Numerische und Angewandte Mathematik
Georg-August Universität Göttingen

Madrid, April 4, 2011

Outline

1. regularizing property of spectral regularization methods
2. source conditions
3. optimality
4. linear statistical inverse problems
5. Lepskiĭ's balancing principle

aim

Show that the "regularization methods" we derived in the last lecture (spectral cut-off, Tikhonov regularization, iterated Tikhonov regularization, Landweber iteration) combined with some appropriate parameter choice rules are in fact regularization methods in the sense of our definition, i.e. the worst case error tends to 0 with the noise level.

spectral theorem (Halmos' version)

Theorem

Let $A \in L(\mathcal{X})$ be self-adjoint. Then there exist

- *a locally compact space Ω with a positive Borel measure μ*
- *a unitary map $W : \mathcal{X} \to L^2(\Omega, d\mu)$,*
- *a real-valued, bounded function $\lambda \in C(\Omega)$, with corresponding multiplication operator $M_\lambda \in L(L^2(\Omega, d\mu))$,* $(M_\lambda f)(\omega) := \lambda(\omega) f(\omega)$

such that the following diagram commutes:

$$\begin{array}{ccc} \mathcal{X} & \xrightarrow{A} & \mathcal{X} \\ W \downarrow & & W \downarrow \\ L^2(\Omega, d\mu) & \xrightarrow{M_\lambda} & L^2(\Omega, d\mu) \end{array}$$

functional calculus: $\varphi(A) := W^* M_{\varphi \circ \lambda} W$

spectral regularization methods

- $T : \mathcal{X} \to \mathcal{Y}$ bounded, $N(T) = \{0\}$, T^{-1} unbounded
- We approximate T^{-1} by a family of bounded operators of the form

$$R_\alpha := q_\alpha(T^*T)T^*, \qquad \alpha > 0.$$

- $q_\alpha : \sigma(T^*T) \to \mathbb{R}$ and $r_\alpha(t) := 1 - tq_\alpha(t)$ satisfy

$$|q_\alpha(t)| \leq \frac{C_e}{\alpha} \qquad \text{for all } \alpha > 0, t \in \sigma(T^*T),$$

$$\lim_{\alpha \to 0} q_\alpha(t) = \frac{1}{t} \qquad \text{for all } t \in \sigma(T^*T) \setminus \{0\},$$

$$|r_\alpha(t)| \leq C_r \qquad \text{for all } \alpha > 0, t \in \sigma(T^*T).$$

regularizing property of spectral regularization methods

Under the previous assumptions $(R_\alpha, \overline{\alpha})$ is a regularization method for any parameter choice rule $\overline{\alpha}$ satisfying

$$\overline{\alpha}(\delta, g^\delta) \to 0 \quad \text{and} \quad \frac{\delta}{\sqrt{\overline{\alpha}(\delta, g^\delta)}} \to 0 \quad \text{as } \delta \to 0.$$

Outline

1. regularizing property of spectral regularization methods
2. source conditions
3. optimality
4. linear statistical inverse problems
5. Lepskiĭ's balancing principle

aims

problem:

- $\|I - R_\alpha T\| \not\to 0$ if T^{-1} unbounded
- Convergence of $f^\dagger - f_\alpha = (I - R_\alpha T)f^\dagger$ arbitrarily slow
- Recall that for any regularization method (not only spectral regularization methods) the convergence of the worst case error as $\delta \to 0$ is arbitrarily slow in general.

aim:

- Formulate additional smoothness assumptions on $f^\dagger$ and prove convergence rate estimates as $\delta \to 0$!

strategy:

- Formulate smoothness of $f^\dagger$ relative to the smoothing properties of $T^*T \rightsquigarrow$ source conditions

approximation error under sources conditions

- Hölder source condition with parameter $\nu > 0$:

$$f^\dagger = (T^*T)^\nu w \qquad \text{for some } w \in \mathcal{X}.$$

- Then $f^\dagger - f_\alpha = r_\alpha(T^*T)f^\dagger = (r_\alpha \cdot t^\mu)(T^*T)w$.
- Assumption: Let $\tau := \|T^*T\|$.

$$\sup_{t\in[0,\tau]} |r_\alpha(t)t^\nu| \leq C_\nu \alpha^\nu \qquad \text{for all } \alpha > 0$$

- Definition: The qualification ν_0 of the method is the largest number $\nu > 0$ for which the previous inequality is satisfied.
- Since $\|\varphi(A)\| \leq \|\varphi\|_\infty$ we obtain

$$\|f^\dagger - f_\alpha\| \leq C_\nu \alpha^\nu \|w\|$$

convergence rates for worst case error

Corollary

Assume that

$$f^\dagger = (T^*T)^\nu w, \qquad \|w\| \le \rho$$

for some $\nu \in (0, \nu_0]$. *Then the a-priori parameter choice rule* $\alpha \sim (\delta/\rho)^{\frac{2}{2\nu+1}}$ *yields the error estimate*

$$\|f^\dagger - f_\alpha^\delta\| \le c_\nu \rho^{\frac{1}{2\nu+1}} \delta^{\frac{2\nu}{2\nu+1}}$$

with c_ν *independent of* δ, ρ, *and* $f^\dagger$.

Disadvantage: Smoothness index ν must be known a-priori!

qualification of standard regularization methods

regularization method	$r_\alpha(t)$	qualification
spectral cut-off	$\chi_{[0,\alpha)}(t)$	∞
Tikhonov regularization	$\frac{\alpha}{\alpha+t}$	1
iterated Tikhonov	$\left(\frac{\alpha}{\alpha+t}\right)^m$	m
Landweber iteration $n = 1/\alpha$	$(1-\mu t)^n$	∞

discrepancy principle

Theorem

Assume that

$$f^\dagger = (T^*T)^\nu w, \qquad \|w\| \le \rho$$

for some $\nu \in \left(0, \nu_0 - \frac{1}{2}\right]$. *Then the discrepancy principle*

$$\overline{\alpha}(\delta, g^\delta) = \sup\left\{\alpha : \|Tf_\alpha^\delta - g^\delta\| \le \tau\delta\right\}$$

with $\tau > C_r$ *yields the error estimate*

$$\|f^\dagger - f_\alpha^\delta\| \le c_\nu \rho^{\frac{1}{2\nu+1}} \delta^{\frac{2\nu}{2\nu+1}}$$

with c_ν *independent of* δ, ρ, *and* $f^\dagger$.

interpretation of source conditions: numerical differentiation

- Consider the periodic Sobolev space $H^s_{\mathrm{per}}([0,1])$, the closed subspace $H^s_\diamond([0,1]) := \{f \in H^s_{\mathrm{per}}([0,1]) : \int_0^1 f(t)\,\mathrm{d}t = 0\}$, and $L^2_\diamond([0,1]) := H^0_\diamond([0,1])$.
- Numerical differentiation with periodic boundary conditions: Define $T_{\mathrm{D}} : L^2_\diamond([0,1]) \to L^2_\diamond([0,1])$ by

$$(T_{\mathrm{D}} f)(x) := \int_0^x f(t)\,dt + c(f) \quad \text{with} \quad c(f) := -\int_0^1 \int_0^x f(t)\,\mathrm{d}t\,\mathrm{d}x\,.$$

- The following operator is a norm isomorphism for all $\nu > 0$:

$$(T_{\mathrm{D}}^* T_{\mathrm{D}})^\nu : L^2_\diamond([0,1]) \to H^{2\nu}_\diamond([0,1])$$

backwards heat equation

- Consider the heat equation on a ring:

$$\begin{aligned} \frac{\partial}{\partial t}u(x,t) &= \frac{\partial^2}{\partial x^2}u(x,t), && x \in (0,1), t \in (0,T] \\ u(0,t) &= u(1,t), && t \in (0,T) \\ u(x,0) &= u_0(x), && x \in [0,1]. \end{aligned}$$

- reconstruction of the temperature at an intermediate time $t_0 \in (0,T)$: Let $f := u(\cdot, t_0)$ and $g := u(\cdot, T)$ and define $T_{t_0} : L^2([0,1]) \to L^2([0,1])$ by $T_{t_0} f := g$.
- Then f satisfies the Hölder source condition

$$f = (T_{t_0}^* T_{t_0})^{\nu} u_0 \qquad \text{with} \qquad \nu = \frac{t_0}{2(T - t_0)}.$$

logarithmic source conditions

- A Hölder source condition for the backwards heat equation implies that the unknown f is an analytic function.
- For $t_0 = 0$ this is a very restrictive assumption. We would prefer a smoothness condition in terms of Sobolev spaces.
- Therefore, we define for $p > 0$ the functions

$$\varphi_p(t) := \begin{cases} (-\ln t)^{-p}, & \text{if } t > 0 \\ 0, & \text{else} \end{cases}$$

and formulate the logarithmic source condition

$$f^\dagger = \varphi_p(T^*T)w\,.$$

interpretation of logarithmic source conditions

Proposition

Let $T_{\mathrm{BH}} : L^2([0,1]) \to L^2([0,1])$ *be the forward operator for the periodic heat equation (with* $t_0 = 0$*). Then the following operator is a norm isomorphism for all* $p > 0$*:*

$$\varphi_p(T_{\mathrm{BH}}^* T_{\mathrm{BH}}) : L^2([0,1]) \to H^{2p}_{\mathrm{per}}([0,1])$$

For similar interpretations of logarithmic source conditions for other inverse problems, see:

- T. Hohage. *Regularization of exponentially ill-posed Problems.* **Numer. Funct. Anal. Optim.**, 21:439–464, 2000.
- T. Hohage. *Logarithmic convergence rates of the iteratively regularized Gauss-Newton method for an inverse potential and an inverse scattering problem.* **Inverse Problems**, 13:1279–1299, 1997.

general source conditions

Definition

A function $\varphi : [0, \tau] \to \mathbb{R}$ is called index function, if it is continuous, monotonically increasing, and $\varphi(0) = 0$.

general source conditions with an index function φ:

$$f^\dagger = \varphi(T^*T)w$$

Definition

We say that an index function ψ covers an index function φ with constant $c > 0$ if

$$c\frac{\psi(\alpha)}{\varphi(\alpha)} \leq \inf_{\alpha \leq t \leq \tau} \frac{\psi(t)}{\varphi(t)} \qquad \text{for all } 0 < \alpha \leq \tau.$$

estimation of the approximation error under general source conditions

examples

- ψ covers φ with constant $c = 1$ if $\frac{\psi}{\varphi}$ is increasing.
- t^{ν_1} covers t^{ν_2} if and only if $\nu_1 \geq \nu_2$.
- t^{ν} covers the logarithmic functions φ_p for all $p > 0$.

Proposition

Consider a spectral regularizaton method described by the family of functions $r_\alpha(t) = 1 - tq_\alpha(t)$ with qualification $\nu_0 > 0$. If t^{ν_0} covers φ, then

$$\sup_{t\in[0,\tau]} |r_\alpha(t)\varphi(t)| \leq \frac{C_{\nu_0}}{c}\varphi(\alpha) \qquad \textit{for all } \alpha > 0\,.$$

T. Hohage. *Regularization of exponentially ill-posed Problems.* **Numer. Funct. Anal. Optim.**, 21:439–464, 2000.

P. Mathé, S. Pereverzev. *Geometry of linear ill-posed problems in variable Hilbert scales.* **Inverse Problems,** 19:789–803, 2003.

convergence rates under general source conditions

Theorem

Assume that $f^\dagger = \varphi(T^*T)w$ with $\|w\| \le \rho$, the regularization method satisfies the assumptions above with qualification $\nu_0 > 0$, and t^{ν_0} covers the index function φ. Moreover, consider the a-priori parameter choice $\overline{\alpha} = \overline{\alpha}(\delta)$ defined implicitely by

$$\Theta(\sqrt{\overline{\alpha}}) = \frac{\delta}{\rho} \quad \text{with} \quad \Theta(t) := \sqrt{t}\varphi(t).$$

Then the error is bounded by

$$\|f^\dagger - f_\alpha^\delta\| \le C\varphi\left(\Theta^{-1}\left(\frac{\delta}{\rho}\right)\right) .$$

Corollary

For $\varphi = \varphi_p$ we get $\|f^\dagger - f_\alpha^\delta\| \le C\varphi_p\left(\frac{\delta}{\rho}\right)$.

Outline

aims

- Can we design regularization methods which yield more accurate results than the methods considered so far?
- Are the estimates we have obtained so far sharp?

definitions

- source sets:

$$M_{\varphi,\rho} := \{\varphi(T^*T)w : \|w\|_{\mathcal{X}} \leq \rho\}$$

- worst case error of a reconstruction method $R : \mathcal{Y} \to \mathcal{X}$ under the a-priori information $f^\dagger \in M_{\varphi,\rho}$ for noise level δ:

$$\Delta_R(\delta, M_{\varphi,\rho}, T) := \sup\left\{\|R(g^\delta) - f^\dagger\| : \begin{array}{l} f^\dagger \in M_{\varphi,\rho},\ g^\delta \in \mathcal{Y} \\ \|Tf^\dagger - g^\delta\| \leq \delta \end{array}\right\}$$

- best possible error bound under the a-priori information $f^\dagger \in M_{\varphi,\rho}$ for noise level δ:

$$\Delta(\delta, M_{\varphi,\rho}, T) := \inf_R \Delta_R(\delta, M_{\varphi,\rho}, T)$$

Here the infimum is over all mappings $R : \mathcal{Y} \to \mathcal{X}$.

modulus of continuity

Definition

The modulus of continuity of $(T|_{M_{f,\rho}})^{-1}$ is defined by

$$\omega(\delta, M_{\varphi,\rho}, T) := \sup\{\|f^\dagger\| : f^\dagger \in M_{\varphi,\rho}, \|Tf^\dagger\| \leq \delta\}.$$

Theorem

$$\Delta(\delta, M_{\varphi,\rho}, T) = \omega(\delta, M_{\varphi,\rho}, T)$$

A.A. Melkman and C.A Micchelli. *Optimal estimation of linear operators in Hilbert spaces from inaccurate data.* **SIAM J. Numer. Anal.**, 16:87–105, 1979.

Jensen's inequality

Lemma

Assume that $\phi : [\alpha, \beta] \to \mathbb{R}$ *with* $\alpha, \beta \in \mathbb{R} \cup \{\pm\infty\}$ *is convex, and let* μ *be a finite measure on some measure space* Ω. *Then*

$$\phi\left(\frac{\int \chi \, d\mu}{\int d\mu}\right) \leq \frac{\int \phi \circ \chi \, d\mu}{\int d\mu}$$

holds for all $\chi \in L^1(\Omega, d\mu)$ *satisfying* $\alpha \leq \chi \leq \beta$ *a.e.* $d\mu$.
The right hand side may be infinite if $\alpha = -\infty$ *or* $\beta = \infty$.

an abstract stability estimate

Theorem

Let $\tau := \|T\|^2$, let $\varphi \in C([0,\tau])$ be an index function, and set $\Theta(t) := \sqrt{t}\varphi(t)$. Assume that

- $f^\dagger = \varphi(T^*T)w$ *with* $\|w\| \le \rho$,
- $\varphi^2((\Theta^2)^{-1})$ *is concave.*

Then the stability estimate

$$\|f^\dagger\| \le \rho\varphi\left(\Theta^{-1}\left(\frac{\|Tf^\dagger\|}{\rho}\right)\right)$$

holds true. Consequently, for $\delta \le \rho\sqrt{\tau}\varphi(\tau)$,

$$\omega(\delta, M_{\varphi,\rho}, T) \le \rho\varphi\left(\Theta^{-1}\left(\frac{\delta}{\rho}\right)\right),$$

and this inequality is sharp for $(\delta/\rho)^2 \in \sigma(T^*T\varphi(T^*T))$.

Corollary

If t^{ν_0} covers φ, the error estimates of the previous section can at most be improved by a constant factor.

- M. Hegland. *An optimal order regularization method which does not use additional smoothness assumptions.* **SIAM J. Numer. Anal.** 29:1446–1461, 1992.
- M. Hegland. *Variable Hilbert scales and their interpolation inequalities with applications to Tikhonov regularization.* **Appl. Anal.** 59:207–223, 1995.
- B.A. Mair. *Tikhonov regularization for finitely and infinitely smoothing operators.* **SIAM J. Math. Anal.** 25:135–147, 1994.
- U. Tautenhahn. *Optimality for ill-posed problems under general source conditions.* **Numer. Funct. Anal. Optim.** 19:377–398, 1998.

Outline

aim

What changes if the error in the data is random rather than deterministic?

Hilbert-space processes

Let $\mathcal{Y}$ be a Hilbert space.

- A Hilbert-space process is a continuous linear operator

$$\xi : \mathcal{Y} \to L^2(\Omega, \mathcal{P}, P).$$

  Notation: $\langle \xi, \varphi \rangle := \xi\varphi$, $\varphi \in \mathcal{Y}$.
- The covariance $\mathbf{Cov}_\xi \in L(\mathcal{Y})$ of ξ is defined implicitly by $\langle \mathbf{Cov}_\xi \varphi_1, \varphi_2 \rangle = \mathbf{Cov}\left(\langle \xi, \varphi_1 \rangle, \langle \xi, \varphi_2 \rangle \right)$ for all $\varphi_1, \varphi_2 \in \mathcal{Y}$.
- ξ is called white noise if $\mathbf{Cov}_\xi = I$.
- Every Hilbert-space valued random variable Ξ satisfying $\mathbb{E}\|\Xi\|^2 < \infty$ can be identified with a Hilbert-space process $\varphi \mapsto \langle \Xi, \varphi \rangle$, $\varphi \in \mathcal{Y}$, but not vice versa. Counter-example: white noise

statistical regularization methods

setting:

$$Y = F(f^\dagger) + \sigma\xi + \delta\zeta$$

$F : D(F) \subset \mathcal{X} \to \mathcal{Y}$ Fréchet differentiable and one-to-one, $\mathcal{X}, \mathcal{Y}$ separable Hilbert spaces. F^{-1} is not continuous.

ξ	Hilbert space process in $\mathcal{Y}$ with $\|\mathbf{Cov}_\xi\| = 1$, $\mathbb{E}\xi = 0$
$\sigma \geq 0$	stochastic noise level
$\zeta \in \mathcal{Y}$	normalized deterministic noise, $\|\zeta\| = 1$
$\delta \geq 0$	deterministic noise level

Definition

A family of mapping $R_\alpha : Y \mapsto \widehat{f}_\alpha$ together with a parameter choice rule $\overline{\alpha} = \overline{\alpha}(Y, \sigma, \delta)$ is called a regularization method for the model above if $R_{\overline{\alpha}(Y,\sigma,\delta)} Y \in \mathcal{X}$ a.s. for all $f^\dagger \in D(F)$ and all $\delta, \sigma > 0$ and if for all $f^\dagger \in D(F)$

$$\lim_{\sigma,\delta \to 0} \sup_{\|\zeta\| \leq 1\|} \mathbb{E}_{f^\dagger} \|R_{\overline{\alpha}(Y,\sigma,\delta)}(Y) - f^\dagger\|^2 = 0\,.$$

bias-variance decomposition

$T : \mathcal{X} \to \mathcal{Y}$ linear, bounded

$$Y = Tf^\dagger + \sigma\xi$$

Consider a linear regularization method $R_\alpha : Y \mapsto \widehat{f}_\alpha$.

$$\mathbb{E}\|\widehat{f}_\alpha - f^\dagger\|^2 = \underbrace{\|R_\alpha Tf^\dagger - f^\dagger\|^2}_{\text{bias}} + \underbrace{\sigma^2\mathbb{E}\|R_\alpha\xi\|^2}_{\text{variance term}} + 2\mathbb{E}\underbrace{\left\langle R_\alpha^*(R_\alpha Tf^\dagger - f^\dagger), \xi\right\rangle}_{=0}$$

- Bias coincides with deterministic approximation error.
- Typical deterministic estimate $\mathbb{E}\|R_\alpha\xi\|^2 \leq \|R_\alpha\|^2\mathbb{E}\|\xi\|^2$ fails or yields suboptimal results.

analysis of the variance term

- Assume that T is compact and let $\{(\sigma_n, u_n, v_n) : n \in \mathbb{N}\}$ be a singular system of T.
- Then R_α is compact with singular system $\{(\tau_n, v_n, u_n) : n \in \mathbb{N}\}$ where $\tau_n := \sigma_n q_\alpha(\sigma_n^2)$.
- Assume that ξ is white noise.

$$\mathbb{E}\|R_\alpha \xi\|^2 = \mathbb{E}\sum_{n=0}^{\infty} \tau_n^2 |\langle \xi, v_n\rangle|^2 = \sum_{n=0}^{\infty} \tau_n^2 \mathbb{E}|\langle \xi, v_n\rangle|^2$$
$$= \sum_{n=0}^{\infty} \sigma_n^2 q_\alpha(\sigma_n^2)^2 = \mathrm{tr}\,(R_\alpha^* R_\alpha)\,.$$

Consequences:

- Unlike in the deterministic case the variance term depends not only on α and σ, but also on the distribution of singular values of T.
- Convergence rate estimates depend on additional parameter(s) characterizing T.

lower bounds (minimax rates)

Define the best possible accuracy of an estimator for the operator T, on a source set $M_{\varphi,\rho} = \{\varphi(T^*T)w : \|w\| \leq \rho\}$ and noise level $\sigma > 0$ by

$$\Delta(\sigma, M_{\varphi,\rho}, T) := \inf_R \sup_{f^\dagger \in M_{\varphi,\rho}} \mathbb{E}\|R(Tf^\dagger + \sigma\xi) - f^\dagger\|^2$$

M.S. Pinsker, *Optimal filtering of square-integrable signals in Gaussian white noise*, **Probl. Inf. Transm.** 16:52–68, 1980

Treats regression (i.e. $T = I$), but general noise and shows that a linear method, which reconstructs only a finite number of Fourier coefficients, is optimal among all linear methods and almost optimal and almost optimal among all methods. Easily generalizes to $T \neq I$.

A.B. Tsybakov Introduction à l'estimation non-paramétrique. Springer, 2004.

linear regularization methods: upper bounds

Tikhonov regularization (suboptimal rates):

D. Cox *Approximation of method of regularization estimators*, **Ann. Stat.** 16:694–712, 1988.

D.W. Nychka, D. Cox *Convergence rates for regularized solutions of integral equations from discrete noisy data*, **Ann. Stat.** 17:556–572, 1989.

spectral cut-off (truncated SVD) for general (not necessarily compact) operators:

B Mair and F Ruymgaart *Statistical inverse estimation in Hilbert scales*, **SIAM J. Appl. Math.** 56:1424–1444, 1996.

order optimality of general regularization methods:

N Bissantz, T Hohage, A Munk, F Ruymgaart *Convergence rates of general regularization methods for statistical inverse problems and applications.* **SIAM J. Numer. Anal.** 45:2610-2636, 2007.

Outline

aims

problem:

- a priori parameter choice rules for (deterministic and statistical) inverse problem require knowledge of the smoothness of the unknown solution. This is usually unrealistic.
- Discrepancy principle is not order optimal if the smoothness is close to the qualification.
- Discrepancy principle does not work for random noise.

aim:

- Describe a parameter choice rule which yields (almost) optimal results without prior knowledge of the smoothness of the solution.
- Treat both deterministic and random errors.

adaptivity for linear statistical inverse problems

Generalized cross validation:

- G.Wahba *Practical approximate solutions to linear operator equations when data are noisy*, **SIAM J. Numer. Anal.** 14:651–667, 1977.
- M.A.Lukas *Asymptotic optimality of generalized cross-validation for choosing the regularization parameter*, **Numer. Math.** 66:41–66, 1993.

negative result: Sometimes it is not possible to get the optimal rates adaptively, one looses at least a factor $|\log \sigma|$.

- A.B. Tsybakov. *On the best rate of adaptive estimation in some inverse problems*, **C.R. Acad. Sci. Paris** 330:835–840, 2000.

In other cases, one can even asymptotically get the optimal constants:

- L.Cavalier, Y.Golubev, D.Picard, A.Tsybokov *Oracle inequalities for inverse problems*, **Ann. Stat.** 30:843–874, 2002.
- L.Cavalier, A.Tsybakov *Sharp adaption for inverse problems with random noise*, **Prob. Theor. Rel. Fields** 123:323–354, 2002.

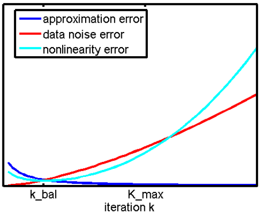
Lepskiĭ for nonlinear inverse problems

- Let $\widehat{f}_0, \widehat{f}_1, \dots, \widehat{f}_{K_{\max}}$ be estimators of $f^\dagger$ in a metric space (X, d) such that

$$d(\widehat{f}_k, f^\dagger) \le \Phi_{\mathrm{noi}}(k) + \Phi_{\mathrm{app}}(k) + \Phi_{\mathrm{nl}}(k), \qquad k \le K_{\max}.$$

- Φ_{noi} is known and non-decreasing.
- Φ_{app} is unknown and non-increasing.
- Φ_{nl} is unknown and satisfies for some $\gamma_{\mathrm{nl}} > 0$

$$\Phi_{\mathrm{nl}}(k) \le \gamma_{\mathrm{nl}} \left(\Phi_{\mathrm{noi}}(k) + \Phi_{\mathrm{app}}(k) \right), \quad k = 0, \dots, K_{\max}.$$



oracle inequality

Lepskiĭ balancing principle:

$$k_{\mathrm{bal}} := \min \left\{ k \le K_{\max} : \begin{array}{l} d(\widehat{f}_k, \widehat{f}_m) \le 4(1+\gamma_{\mathrm{nl}})\Phi_{\mathrm{noi}}(m), \\ m = k+1, \dots, K_{\max} \end{array} \right\}$$

Theorem

Assume that $\Phi_{\mathrm{noi}}(k+1) \le \overline{\gamma}_{\mathrm{noi}} \Phi_{\mathrm{noi}}(k)$ *for some constant* $\overline{\gamma}_{\mathrm{noi}} < \infty$*. Then*

$$d(\widehat{f}_{k_{\mathrm{bal}}}, f^{\dagger}) \le 6(1+\gamma_{\mathrm{nl}})\overline{\gamma}_{\mathrm{noi}} \min_{k=1,\dots,K_{\max}} \left(\Phi_{\mathrm{app}}(k) + \Phi_{\mathrm{noi}}(k) \right).$$

F. Bauer, T. Hohage, A. Munk, *Regularized Newton Methods for Nonlinear Inverse Problems with Random Noise*, **SIAM J. Numer. Anal.** 47:1827-1846, 2009.

discussion

- Lepskiĭ principle for iterative regularization method requires to compute more iterations than the optimal stopping index.
- Oracle inequalities sharper than minimax rates:
  - Every element in $\mathcal{X}$ fulfills a source condition, also the source elements w.
  - The minimax rate for $M_{\varphi,\rho}$ is not optimal for *any* of the elements of $M_{\varphi,\rho}$.
  - Oracle inequality assures optimality for each indiviual solution, not after a sup over some class.
- Strategy for random noise:
  - Distinguish a ”good case” that the variance term is not much larger than its expectation and a ”bad case” that it is much larger.
  - The ”good case” can be treated as in the deterministic case.
  - Show that the probility of the ”bad case” is small.
  - This costs a logarithmic factor in the noise level as price for adaptation (which cannot be avoided in some cases)

literature

Lepskiĭ principle for linear inverse problems:

- O.V. Lepskiĭ. *On a problem of adaptive estimation in Gaussian white noise*, **Theory Probab. Appl.** 35:454–466, 1990.
- A. Goldenshluger, S. Pereverzev. *Adaptive estimation of linear functionals in Hilbert scales from indirect white noise observations*, **Prob. Theor. Rel. Fields** 118:169–186, 2000.
- P. Mathé, S. Pereverzev. *Geometry of ill-posed problems in variable Hilbert scales*, **Inverse Problems** 19:789–803, 2003.
- P. Mathé. *The Lepskiĭ principle revisited.* **Inverse Problems** 22:L11–L15, 2006.