Nonlinear Ill-Posed Problems and Noise Models for Inverse Problems and Moment Discretization

Zuhair Nashed
University of Central Florida

Inverse Problems and Applications

April 2-6, 2013

Linköping University, Sweden

I. Variational Methods

Potential operators

Quasipotential Operators

II. Monotoncity Methods

Monotone Operator Theory for

Ill-Posed Operators and Variational Inequalities

III. Noise Models

Weakly Bounded Noise

I. Variational Methods

1. Formulation of linear problems of math physics as variational problems

Symmetry or self-adjointness of the linear operator

$Au = f$ $\qquad J(u) = \langle Au, u \rangle - 2\langle u, f \rangle$

Regularization: $\quad J_\alpha(u) = J(u) + \alpha \|u\|^2, \quad \alpha > 0$

$\langle Au, u \rangle \geq 0$

Courant and Hilbert / Mikhlin

2. There is a direct analogy between

conservative force (vector) fields

and

variational methods for linear and nonlinear operator equations

$\vec{F}$ vector field on a region $\Omega \subset \mathbb{R}^3$

$\vec{F} = (F_1, F_2, F_3), \quad F_i : \Omega \subset \mathbb{R}^3 \to \mathbb{R}.$

Ω simply connected

Criteria for $\vec{F}$ to be a conservative force field:

$\text{grad } f = \vec{F}$

(i) $\displaystyle\int_{P_1}^{P_2} \vec{F} \cdot d\vec{r}$ is independent of the path

(ii) If $\vec{F} \in C'(\Omega)$, then $\vec{F} = \text{grad } f$

$$\Leftrightarrow \quad \text{curl } \vec{F} = 0 \quad \text{in } \Omega.$$

We seek analogous criteria in infinite dimensional spaces.
E. Rothe: Gradient mappings
M. Vainberg: Potential operators

$$\text{Curl } \vec{F} = \begin{bmatrix} i & j & k \\ \frac{\partial}{\partial x_1} & \frac{\partial}{\partial x_2} & \frac{\partial}{\partial x_3} \\ F_1 & F_2 & F_3 \end{bmatrix} = 0 \iff \frac{\partial F_i}{\partial x_j} = \frac{\partial F_j}{\partial x_i}, \quad i, j = 1, 2, 3$$

The vanishing of curl $\vec{F}$ is equivalent to the **symmetry** of the Jacobian matrix (= the Fréchet derivative of the operator F defining the vector field.

<u>Kerner's symmetry condition</u>

M a dense linear manifold in a Hilbert space H. $F: M \longrightarrow H$. Assume that the Gateaux differential of F exists for $x, h \in M$, and is continuous in every two dimensional manifold.
Then a necessary and sufficient conditions for F to be a gradient mapping is that $F'(x)$ be **symmetric** for every $x \in M$, i.e., for every $x, h, k \in M$,

$$\langle h, F'(x)k \rangle = \langle k, F'(x)h \rangle$$

$$f(x) = \int_0^1 \langle F(tx), x \rangle \, dt \qquad \text{Riemann-Graves integral}$$

Tikhonov Regularization

Lavrentiev Regularization (by perturbation).

II. Monotone Operator Theory in Ill-Posed Operator Equations and Variational Inequalities

$A : H \longrightarrow H$ H is a real Hilbert space

Monotone : $\langle x-y, A(x)-A(y)\rangle \geq 0$

Strictly Monotone > 0 for $y \neq x$

Strongly Monotone $\geq m\|x-y\|^2$

1. Suppose A is strongly monotone and Lipschitzian. $\|Ax - Ay\| \leq M\|x-y\|$. Consider the operator equation $Au = f$. Let $T_\alpha u := u - \alpha Au + \alpha f$.
$\hat{u}$ is a solution of $Au = f \Leftrightarrow \hat{u}$ is a fixed point of T_α.

$$\|T_\alpha x - T_\alpha y\|^2 = \|x-y\|^2 - 2\alpha\langle x-y, Ax-Ay\rangle + \alpha^2\|Ax-Ay\|^2$$
$$\leq \{1 - 2\alpha m + \alpha^2 M^2\}\|x-y\|^2$$

$g(\alpha) := 1 - 2\alpha m + \alpha^2 M^2 < 1$ for $0 < \alpha < \frac{2m}{M^2}$.

Contractive Averaging Principle

3. Theory of Integrability of Vector Fields (Pfaff's Problem) provides another extension of variational principles.

- If $\vec{F} = (F_1, F_2, F_3)$, and if F_i, $i = 1, 2, 3$ have continuous partial derivatives in a simply connected region, then a necessary and sufficient condition for the existence of a scalar function $\varphi(x_1, x_2, x_3)$ is that curl $\vec{F}$ is orthogonal to $\vec{F}$.

$$\operatorname{curl}(\varphi \vec{F}) = 0 \iff \vec{F} \cdot \operatorname{curl} \vec{F} = 0$$

$$\iff F_1\left(\frac{\partial F_3}{\partial x_2} - \frac{\partial F_2}{\partial x_3}\right) + F_2\left(\frac{\partial F_1}{\partial x_3} - \frac{\partial F_3}{\partial x_1}\right) + F_3\left(\frac{\partial F_2}{\partial x_1} - \frac{\partial F_1}{\partial x_2}\right) = 0$$

Generalization: Quasipotential Operator

$$F : X \longrightarrow X^* \qquad X \text{ Banach space}$$

Assume that the Fréchet derivative of F exists and is continuous in a neighborhood $\mathcal{U}$ of $x_0 \in X$ with $F(x_0) \neq 0$. Define the trilinear form

$$I(x; h, j, k) = \langle F(x), h\rangle\{\langle F'(x)j, k\rangle - \langle F'(x)k, j\rangle\}$$
$$+ \langle F(x), j\rangle\{\langle F'(x)k, h\rangle - \langle F'(x)h, k\rangle\}$$
$$+ \langle F(x), k\rangle\{\langle F'(x)h, j\rangle - \langle F'(x)j, h\rangle.$$

If $I(x; h, j, k) = 0$ for all $x \in \mathcal{U}$, $h, j, k \in X$, then there exists a positive integrating factor

$$\varphi(x)\, F(x) = \operatorname{grad} f(x).$$

2. Now suppose that A is strongly monotone and continuous. Then A is one-to-one, onto,
(Minty / Browder Theorem)
and its inverse is continuous.

Thus the operator equation $Au = f$ is globally well posed

Extensive literature in Banach spaces

$$A : X \longrightarrow X^*$$

and weaking of continuity and strong montonicity

3. F monotone and continuous, then $Au=f$ is ill-posed.

4. Operator Inclusions and Set-Valued Mappings

$$y \in Ax \quad \text{subject to} \quad z \in Bx$$
$$\text{or} \quad Bx = z.$$

(F. Liu and N)

Regularization: $y \in Ax + \alpha B$

Rates of convergence are still lacking except in special cases

5. Variational Inequalities and the important class of "Inverse Monotone Map

$A : H \longrightarrow 2^H$ is said to be inverse monotone $\Leftrightarrow$ $\exists$ a constant $\beta > 0$ such that

$$\langle u^* - v^*, u - v \rangle \geq \beta \, \| u^* - v^* \|^2$$

for all $u^* \in Au$, $v^* \in Av$.

For the case of a single-valued mapping, an operator $A : H \rightarrow H$ is inverse-monotone $\Leftrightarrow$ for some constant $\beta > 0$,

$$\langle Au - Av, u - v \rangle \geq \beta \, \| Au - Av \|^2$$

$A : H \rightarrow H$ is inverse-monotone $\Leftrightarrow$ $A^{-1} : H \rightarrow 2^H$ is strongly monotone

H Hilbert space

K (nonempty) closed, convex set

$f \in H$ given

$A : H \longrightarrow H$ monotone

Find $u \in K$ for which

$$\langle Au - f , v - u \rangle \geq 0 \quad \text{for all } v \in K.$$

Regularization with noisy data

Perturbed variational inequalities:

For $f_\delta \in H$, $\| f_\delta - f \| \leq \delta$, find $u_\varepsilon^{\delta,\eta}$ in K_η for which

$$\langle A u_\varepsilon^{\delta,\eta} + \varepsilon u_\varepsilon^{\delta,\eta} - f_\delta , v - u_\varepsilon^{\delta,\eta} \rangle \geq 0$$

$$\text{for all } v \in K_\eta$$

η : perturbation "parameter" for K

δ : noise

ε : regularization parameter

Strong Noise Model in Inverse/Ill-Posed Problems

$$Tu = f \qquad T : H \longrightarrow H$$

$$Tu = f_\delta \qquad f_\delta = f + \text{noise}$$

$$\| f_\delta - f \| \le \delta, \qquad \delta > 0 \text{ is "known"}$$

Approximate regularizing operators:

$$\Gamma_\alpha : H \longrightarrow H$$

$$\| \Gamma_\alpha f \| \le C_\alpha \| f \|, \quad \alpha > 0$$

$u_\alpha := \Gamma_\alpha f$ is an approximation to the solution or least squares solution $\hat{u}$.

$$C_\alpha \nearrow \infty \quad \text{as } \alpha \searrow 0$$

$$\| u_\alpha^\delta - \hat{u} \| = \| \Gamma_\alpha f_\delta - \hat{u} \|$$

$$= \| u_\alpha^\delta - u_\alpha \| + \| u_\alpha - \hat{u} \|$$

$$\le C_\alpha \| f_\delta - f \| + M(\alpha)$$

requires rate of convergence.

$$\le \underset{\nearrow \infty}{\delta C_\alpha} + \underset{\searrow 0}{M(\alpha)} \quad \text{as } \alpha \searrow 0$$

- Tikhonov regularization

  Linear problems

$$\Gamma_\alpha = \left(T^* T + \alpha L^* L\right)^{-1} T^*, \quad \alpha > 0$$

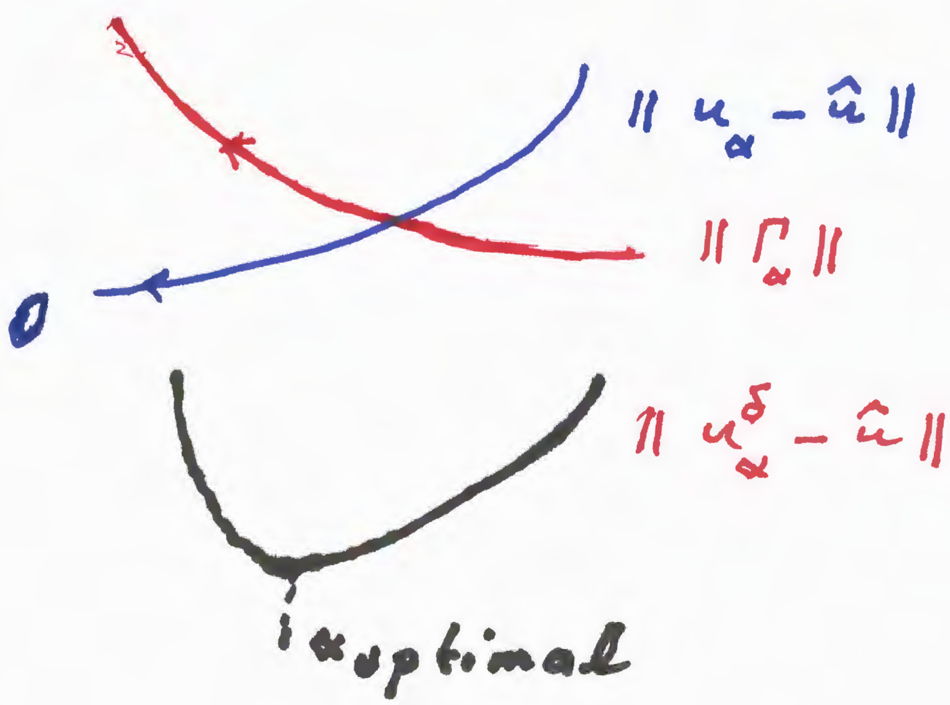
- Finite Differences $\quad \Gamma_h \quad h = \text{meshsize}$

$$\Gamma_\beta \qquad \beta = \frac{1}{h}$$

- Projection Schemes, splines, wavelet approximations

$$\Gamma_\beta \qquad \beta = \frac{1}{n}$$

  n dimension of approximation subspaces



Weakly Bounded Noise

P. P. B. Eggermont, V. N. LaRiccia and M. Z. Nashed

1. On weakly bounded noise
   Inverse Problems 25 (2009) 115018 (14 pp)

2. Noise Models for Ill-Posed Problems
   in Handbook of Geomathematics
   W. Freeden, M. Z. Nashed, T. Sonar (Eds.)
   Springer-Verlag, Berlin Heidelberg 2010
   chapter 24, pp. 742 - 762

3. Moment discretization for ill-posed problems
   with discrete weakly bounded noise
   International Journal on Geomathematics
   Published online: 09 April 2012
   July 2012

1 Noise in ill-posed problems

(a) X, Y Hilbert spaces

$K : X \longrightarrow Y$ linear, compact, 1-1
(Range of K not closed in Y)

"Data" $y \in Y$ such that

$$y = K x_0 + \delta$$

for some $x_0 \in X$, with "noise" δ.

(b) Goal: Recover x_0.

Problem is ill-posed b/c K^{-1} is not bounded.

(c) Classical view: $\|\delta\|_Y$ is "small".

Analyze what happens when $\|\delta\|_Y \to 0$.

(d) Cases not covered

- high frequency noise (not "small")
- white noise ξ ($\xi \notin Y$!)
- random discrete noise (not "small"; more data becomes available)
- signal-to-noise ratio < 10% !

(e) High-frequency noise in $L^2(0,1)$

$$y(t) = [Kx_0](t) + \delta(t), \quad 0<t<1$$

e.g., $\delta_\eta(t) = \sin(t/\eta)$ for $\eta \to 0$.

So $\|\delta_\eta\|^2_{L^2} \approx \frac{1}{2} \not\to 0$.

(f) But, e.g., Tikhonov regularization

$$x_{\alpha\delta} = (K^*K + \alpha I)^{-1} K^* y$$

Then $\|K^*\delta_\eta\|_X \to 0$ b/c $\delta_\eta \to 0$ weakly

So for $\alpha \approx \|K^*\delta\|_X^{1/2}$ it "works".

(g) One never observes $\delta(t)$, $0<t<1$.

$$y(t_i) = [Kx_0](t_i) + \delta_i, \quad i=1,2,\dots,n$$

with $t_i = \frac{i}{n}$ say; ~~xxxx~~ $\delta_i = \sin(t_i/\eta)$

$$\frac{1}{n}\sum_{i=1}^{n} |\delta_i|^2 \not\to 0.$$

(h) White noise

$$y = Kx_0 + \eta \xi$$

ξ white noise on Y, $\eta \to 0$.

$$\xi \text{ "=" } \sum_{i=1}^{\infty} Z_i u_i$$

where $u_1, u_2, \ldots$ orthonormal basis for Y,
$Z_1, Z_2, \ldots$ independent Normal$(0,1)$.

So $\xi \notin Y$: Formally

$$\|\xi\|_Y^2 = \sum_{i=1}^{\infty} |Z_i|^2 = +\infty \quad \text{a.s.}$$

But, for any $\ell \in Y$, $|\langle \ell, \xi \rangle| < \infty$ a.s.,
and then

$$|\langle \ell, \eta \xi \rangle_Y| \to 0 \text{ a.s. for } \eta \to 0.$$

So, "almost" $\eta \xi \to 0$ <u>weakly</u> in Y.

(i) Discrete random noise.

$$(K : L^2(0,1) \to C[0,1])$$

$$y_i = [Kx_0](t_{i,n}) + \delta_{i,n}, \quad i = 1, 2, \cdots, n$$

with $t_{i,n} = \frac{i}{n+1}$, $i = 1, 2, \cdots, n$ (or some such)
and

$$\delta_{1,n}, \delta_{2,n}, \cdots, \delta_{n,n} \quad \text{i.i.d. Normal}(0, \sigma^2).$$

σ^2 : variance of the noise (<u>fixed</u>)

So: the noise does not tend to 0

$$\frac{1}{n} \sum_{i=1}^{n} |\delta_{in}|^2 \approx \sigma^2$$ (Weak law of large numbers)

But: as $n \to \infty$, one can get <u>consistent</u> estimators of x_0

$$\| \hat{x}_n - x_0 \|_X \to 0 \quad \text{(in probability)}$$

(j) Note: The signal-to-noise ratio

$$\frac{\frac{1}{n} \sum_{i=1}^{n} \left| [Kx_0](t_i) - \mu_0 \right|^2}{\sigma^2}$$

can be arbitrarily small $\left(\mu_0 = \frac{1}{n} \sum_{i=1}^{n} [Kx_0](t_i) \right)$.

2 Weakly bounded noise

(a) Consider

$$y = K x_0 + \delta \qquad \text{in } Y$$

with $\delta \to 0$ weakly, $\|\delta\|_Y \asymp 1$.

We ~~want~~ need rates on "$\delta \to 0$ weakly".

(b) Precise conditions:

$$T : Y \to Y \qquad \text{linear, compact Hermitian, positive definite}$$

$$\eta^2 \triangleq \langle \delta, T\delta \rangle_Y \longrightarrow 0$$

$$K(X) \underset{\text{cont.}}{\subset} T^m(Y) \quad \text{for some } m \geq 1.$$

(c) T compact, $\delta \to 0$ weakly implies that

$$\langle \delta, T\delta \rangle \longrightarrow 0$$

but

T is "related" to K

e.g. $T = (KK^*)^{1/2m}$.

(d) Example of weakly bounded noise.

$\{v_j\}_j$, $\{w_j\}_j$ orthonormal bases for X, Y.

<u>Define</u> $K : X \to Y$ ~~by~~ , $T : Y \to Y$ by

$$Kx = \sum_j \kappa_j \langle v_j, x\rangle_X \, w_j \quad , \quad x \in X,$$

for $1 = \kappa_1 \geq \kappa_2 \geq \cdots > 0$.

$$Ty = \sum_j \tau_j \langle w_j, y\rangle_Y \quad , \quad y \in Y,$$

for $\tau_j = (\kappa_j)^{1/m}$.

noise: $$\delta_\eta = \eta \sum_j \delta_{1,j} w_j + \sum_{\tau_j \leq \eta} \delta_{\eta,j} w_j$$

for
$$\sum_j |\delta_{1,j}|^2 \cong \tfrac{1}{2} \quad , \quad \sum_{\tau_j \leq \eta} |\delta_{\eta,j}|^2 \cong \tfrac{1}{2}.$$

<u>Then</u> $T^{-m}K : X \to Y$ bdd, bdd inverse

$\langle \delta, T\delta \rangle \leq \eta$

$\langle \delta, T^q \delta \rangle = O(\eta)$ fails for $q < 1$

$\langle \delta, T^p \delta \rangle = o(\eta)$ fails for $p > 1$

3. Tikhonov regularization

For $\alpha > 0$, $x_{\alpha\delta}$ is the solution of

$$\min \|Kx - y\|_Y^2 + \alpha \|x\|_X^2 \triangleq J(x).$$

Thm If $\alpha \to 0$, $\alpha^{-1}\eta^{4m} \to 0$ then

$$\|x_{\alpha\delta} - x_0\|_X \to 0.$$

For rates, assume the source condition

(SC) $\qquad x_0 = (K^*K)^{\nu/2} z_0$ for some $z_0 \in X$

$$0 < \nu \leq 1.$$

Thm Under the assumptions 2(b) and (SC)

$$\|x_{\alpha\delta} - x_0\|_\alpha \leq c\left(\alpha^{-\frac{1}{4m}}\eta + \alpha^{\frac{\nu+1}{2}}\right)$$

Corr For $\alpha \asymp \eta^{4m/(2m(\nu+1)+1)}$

$$\|x_{\alpha\delta} - x_0\|_X \leq c\,\eta^{2m\nu/(2m(\nu+1)+1)}.$$

(Corollary holds for $0 < \nu \leq 2$, actually. The theorem does not.)

$$\boxed{\|x\|_\alpha^2 = \|Kx\|_Y^2 + \alpha\|x\|_X^2}$$

"Proof". Expand $J(x_0) - J(x_{\alpha\delta})$ around $x_{\alpha\delta}$ (Frechet derivative vanishes) and x_0.

This gives $(\cdots)$ /w $\varepsilon_{\alpha\delta} \equiv x_{\alpha\delta} - x_0$

$$\star \quad \| \varepsilon_{\alpha\delta} \|_\alpha^2 = \langle \delta, K\varepsilon \rangle_Y + \alpha \langle x_0, \varepsilon_{\alpha\delta} \rangle_X$$

The source condition (SC) gives (for $\nu \le 1$)

$$\star\star \quad \alpha \langle x_0, \varepsilon_{\alpha\delta} \rangle_X \le c \, \alpha^{\frac{\nu+1}{2}} \| \varepsilon_{\alpha\delta} \|_\alpha .$$

The assumptions 2(b) on the weakly bounded noise:

For all $\beta > 0$:

$$\langle \delta, K\varepsilon \rangle_Y = \langle (1+\beta T^{-1})^{-1} \delta, (1+\beta T^{-1}) K\varepsilon \rangle_Y$$

$$\star\star\star \quad \le \langle \delta, (1+\beta T^{-1})^{-1} \delta \rangle^{1/2} \langle K\varepsilon, (1+\beta T^{-1}) K\varepsilon \rangle_Y^{1/2}$$

(a) $(1+\beta T^{-1})^{-1} \lesssim (2\beta)^{-1} T$ (just compare eigenvalues)

so

$$\langle \delta, (1+\beta T^{-1})^{-1} \delta \rangle_Y \le (2\beta)^{-1} \langle \delta, T\delta \rangle_Y \le (2\beta)^{-1} \eta^2 .$$

(b)
$$\langle z, (1+\beta T)^{-1} z \rangle_Y = \| z \|_Y^2 + \beta \| T^{-1/2} z \|_Y^2$$
$$\le c_m \left(\| z \|_Y^2 + \beta^{2m} \| T^{-m} z \|_Y^2 \right)$$

provided $z \in \mathcal{R}(T^{-m})$. So:

$$\star\star\star\star \quad \langle K\varepsilon, (1+\beta T^{-1}) K\varepsilon \rangle \le c_m \| \varepsilon \|_{\beta^{2m}}^2 \implies \beta = \alpha^{\frac{1}{2m}}$$

substitute everything into $\star$

$$\|\varepsilon_{\alpha\delta}\|_\alpha^2 \le \tilde{c}\left\{\alpha^{-\frac{1}{4m}}\eta + \alpha^{\frac{\nu+1}{2}}\right\}\|\varepsilon_{\alpha\delta}\|_\alpha$$

and cancel $\|\varepsilon_{\alpha\delta}\|_\alpha$.

$\vdots$

Constraints The proof also works for

$$\min \|Kx-y\|_Y^2 + \alpha\|x\|^2$$
$$\text{s.t. } x \in C \subset X$$

for C closed, convex. The source condition must be replaced by

$$\text{(C-SC)} \qquad x_0 = P_C (K^*K)^{\nu/2} z_0 \quad \text{for some } z_0 \in X,$$

$$0 < \nu \le 1$$

and P_C is the orthogonal projector onto C:

$$x = P_C z \text{ solves } \min_{x\in C} \|x - z\|_X .$$

(Pereverzev (2006), Mathé and Pereverzev (2006))

Optimality lies if $T^{-m}K$ is bicontinuous.
(Follows as in Natterer (1984).)

4. Necessity of source condition.

Note that

$$x_{\alpha\delta} = (K^*K + \alpha I)^{-1} K^* y$$

equals

$$x_{\alpha\delta} = K (KK^* + \alpha I)^{-1} y$$

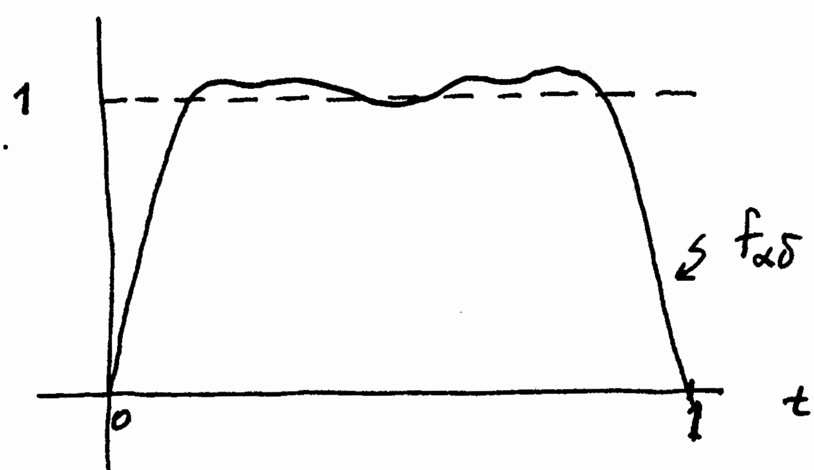
so $\quad x_{\alpha\delta} \in \mathcal{R}(K) .= \mathcal{R}\left(\left(KK^*\right)^{1/2}\right)$

If $\quad x_0 \notin \mathcal{R}(K) \quad$ then $x_{\alpha\delta} \to x_0$ is difficult.

Ex $\quad g_0(t) = \int_0^1 (t \wedge s - ts) f_0(s)\, ds \quad , 0 < t < 1$

$u = g_0$ solves $\begin{cases} -u'' = f_0 & \text{on } (0,1) \\ u(0) = u(1) = 0 \end{cases}$

e.g. $f_0 \equiv 1$



b/c $\quad f_{\alpha\delta}(0) = f_{\alpha\delta}(1) = 0 \quad$ for all $\alpha > 0$.

5. Smoothing parameter selection

Lepskiĭ principle, see Mathé (2006)

Define $\Psi(\alpha) = 2\eta\, \alpha^{-(2m+1)/4m}$

(Roughly speaking,

$$\Psi(\alpha) = 2\, \| x_{\alpha,\delta} - x_{\alpha,0} \|_X ,$$

the noise part in $x_{\alpha,\delta}$.)

Requires m (known), η (must be estimated)

but not ν (the source condition)

Select $\alpha = \alpha_L$ as

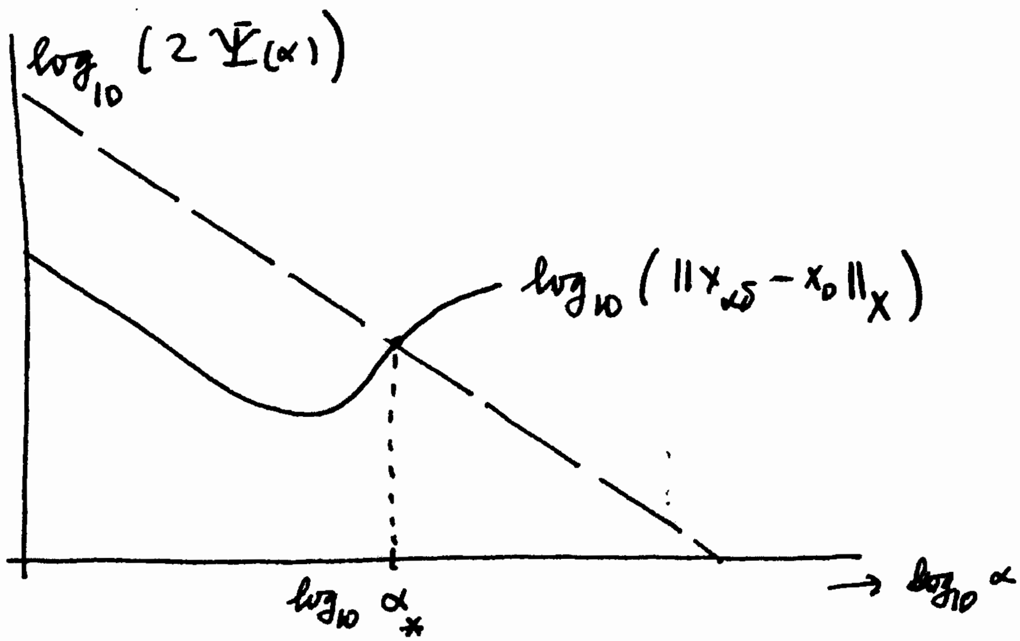
$$\alpha_L := \sup \left\{ \alpha : \forall\, 0 < \beta < \alpha \left(\| x_{\beta,\delta} - x_{\alpha,\delta} \|_X \le 2\Psi(\beta) \right) \right\}$$

Define the "optimal" $\alpha = \alpha_*$ as

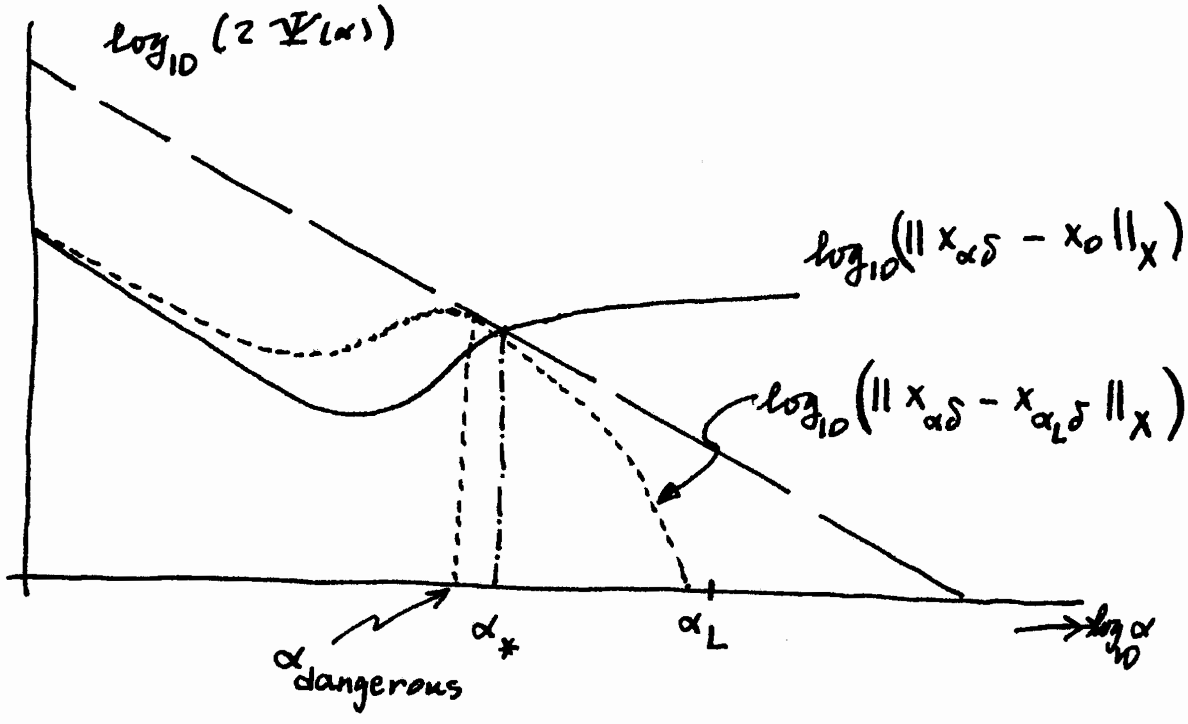
$$\alpha_* = \sup \left\{ \alpha : \forall\, 0 < \beta < \alpha \quad \| x_{\beta,\delta} - x_0 \|_X \le \Psi(\beta) \right\}$$

Then $\alpha_L > \alpha_*$

$$\| x_{\alpha_*,\delta} - x_0 \|_X = \Psi(\alpha_*)$$



Typical situation for the selection of α_* and α_L.
(For small α the graphs are parallel.)



α_d is OK if $\alpha_d \geq \alpha_*$; maybe not if $\alpha_d < \alpha_*$.

$$\alpha_L = \sup\{\alpha : \forall \beta < \alpha \ (\| x_{\beta,\delta} - x_{\alpha,\delta} \| \le 2\Psi(\beta)) \}$$

$$\alpha_* = \sup\{\alpha : \forall \beta < \alpha \ (\| x_{\beta,\delta} - x_0 \| \le \Psi(\beta) \}$$

For $0 < \gamma < \beta \le \alpha_*$

$$\| x_{\gamma\delta} - x_{\beta\delta} \| \le \| x_{\gamma\delta} - x_0 \| + \| x_{\beta\delta} - x_0 \|$$
$$\le \Psi(\gamma) + \Psi(\beta) \le 2\Psi(\gamma)$$

or $\quad \| x_{\gamma\delta} - x_{\beta\delta} \| \le 2\Psi(\gamma) \quad , \forall \gamma < \beta .$

So: $\quad \alpha_* < \alpha_L$. But then

$$\| x_{\alpha_L\delta} - x_0 \| \le \| x_{\alpha_L\delta} - x_{\alpha_*\delta} \| + \| x_{\alpha_*\delta} - x_0 \|$$
$$\le 2\Psi(\alpha_*) + \Psi(\alpha_*)$$
$$\boxed{\| x_{\alpha_L\delta} - x_0 \| \le 3\Psi(\alpha_*)} .$$

And by continuity

$$\| x_{\alpha_*\delta} - x_0 \| = \Psi(\alpha_*)$$

so

$$\Psi(\alpha_*) \le \| x_{\alpha_*\delta} - x_{\alpha_*,0} \| + \| x_{\alpha_*,0} - x_0 \|$$
$$\le \tfrac{1}{2}\Psi(\alpha_*) + c \cdot \alpha_*^{\nu/2}$$

Gives the right rate for α_*.

6. Discrete noise in $L^2(0,1)$

Model $\quad y_i = [Kx_0](t_i) + d_i \, , \quad i = 1, 2, \cdots, n \, .$

Tikhonov regularization

$$\min \frac{1}{n} \sum_{i=1}^{n} \left| [Kx](t_i) - y_i \right|^2 + \alpha \|x\|_{L^2}^2$$

(In) equality: $\quad \varepsilon_{\alpha\delta} \equiv x_{\alpha\delta} - x_0$

$$\frac{1}{n} \sum_{i=1}^{n} |[K\varepsilon](t_i)|^2 + \alpha \|\varepsilon\|^2 =$$

$$= \frac{1}{n} \sum_{i=1}^{n} d_i \, [K\varepsilon](t_i) + \alpha \langle x_0, \varepsilon \rangle_{L^2} \, .$$

Replace $\quad \frac{1}{n} \sum_{i=1}^{n} |[K\varepsilon](t_i)|^2 \quad$ by $\quad \|K\varepsilon\|_{L^2}^2 \, . \qquad (\cdots\cdot)$

Assume $\quad T^{-m} K$ bdd and

$$Ty(t) = \int_0^1 T(t,s) \, y(s) \, ds \, , \quad 0 < t < 1$$

with $T(t,s)$ continuous.

Noise Note that

$$\frac{1}{n}\sum_{i=1}^{n} [K\varepsilon](t_i)\, d_i = \int_0^1 [K\varepsilon](t)\, dF_n(t) = \langle K\varepsilon, dF_n\rangle_{L^2}$$

/w

$$F_n(t) = \frac{1}{n}\sum_{t_i \le t} d_i$$

Then (informally)

$$\langle K\varepsilon, dF_n\rangle = \langle (1+\beta T^{-1}) K\varepsilon, (1+\beta T^{-1})^{-1} dF_n\rangle$$

$$\le \langle K\varepsilon, (1+\beta T^{-1}) K\varepsilon\rangle^{1/2} \times$$

$$\langle dF_n, (1+\beta T^{-1})^{-1} dF_n\rangle^{1/2}$$

and

$$\langle dF_n, (1+\beta T^{-1})^{-1} dF_n\rangle \le (2\beta)^{-1}\langle dF_n, T\, dF_n\rangle$$

$$= \frac{1}{2\beta n^2}\sum_{i,j=1}^{n} d_i\, T(t_i, t_j)\, d_j$$

Assumption on noise

$$\frac{1}{n^2}\sum_{i,j=1}^{n} d_i\, T(t_i, t_j)\, d_j \le \eta_n \longrightarrow 0$$

Assumption on noise

$$\frac{1}{n^2} \sum_{i,j=1}^{n} d_i \, T(t_i, t_j) \, d_j \le \eta_n^2 \longrightarrow 0 .$$

Romdom noise

If $d_1, d_2, \cdots, d_n$ independent Normal$(0, \sigma^2)$

then

$$\mathbb{E}\left[\frac{1}{n^2} \sum_{i,j=1}^{n} d_i \, T(t_i, t_j) \, d_j \right]$$

$$= \frac{\sigma^2}{n} \cdot \frac{1}{n} \sum_{i=1}^{n} T(t_i, t_i)$$

$$\approx \frac{\sigma^2}{n} \, \mathrm{trace}(T).$$

Etc

7. Numerical experiments (with random noise)

$$\frac{\mathbb{E}\left[\| x_{\alpha_L,\delta} - x_0 \|\right]}{\mathbb{E}\left[\| x_{\alpha_*,\delta} - x_0 \|\right]} \approx 2.5$$

and

$$\frac{\mathbb{E}\left[\| x_{\alpha_D,\delta} - x_0 \|\right]}{\mathbb{E}\left[\| x_{\alpha_*,\delta} - x_0 \|\right]} \approx 1.5$$

but

$$\frac{\mathbb{E}\left[\| x_{\alpha_*,\delta} - x_0 \|\right]}{\mathbb{E}\left[\min_\alpha \| x_{\alpha,\delta} - x_0 \|\right]}$$ is problematic.

$$g_0(t) = \int_0^1 (t \wedge s - ts)\, f_0(s)\, ds = [K f_0](t)$$

$$g_0(t) = (t(1-t))^p \quad \text{for } p = 3, 4 .$$

Then $f_0(s) = s(1-s)\{ \; \ldots\ldots \; \}$

so $f_0 \in \mathcal{R}(K)$.

Simulation results

$$t_i = \frac{i}{n+1} \quad ; \quad y_i = g_0(t_i) + d_i \ , \ i = 1, 2, \cdots, n$$

$$(d_1, d_2, \cdots, d_n)^T \sim \text{Normal}(0, \sigma^2 I)$$

$$\sigma = \frac{1}{10} \{ \max g_0(t) - \min g_0(t) \}$$

$$\text{error} = \mathbb{E}\left[\frac{1}{\sqrt{n}} \| f_{\alpha\delta} - f_0 \|_{\mathbb{R}^n} \right] \text{ (estimated)}$$

1000 replications

n	optimal	Star	Dangerous	Lepskiĭ
50	$1.79_{10}{}^{-2}$	$2.15_{10}{}^{-2}$	$3.02_{10}{}^{-2}$	$5.69_{10}{}^{-2}$
100	$1.46_{10}{}^{-2}$	$1.77_{10}{}^{-2}$	$2.35_{10}{}^{-2}$	$4.74_{10}{}^{-2}$
200	$1.20_{10}{}^{-2}$	$1.45_{10}{}^{-2}$	$1.85_{10}{}^{-2}$	$3.89_{10}{}^{-2}$
400	$9.90_{10}{}^{-2}$	$1.19_{10}{}^{-2}$	$1.48_{10}{}^{-2}$	$3.17_{10}{}^{-2}$

$$g_0(t) = (t(1-t))^4$$

$$\Psi(\alpha) = \frac{2\sigma}{\sqrt{n}} \alpha^{-9/16}$$

Observed rate $\approx n^{-9/32}$
(in all cases)

Simulation results

$$t_i = \frac{i}{n+1} \; ; \quad y_i = g_0(t_i) + d_i \; , \quad i = 1, 2, \cdots, n$$

$$(d_1, d_2, \cdots, d_n)^T \sim \text{Normal}(0, \sigma^2 I)$$

$$\sigma = \frac{1}{10} \{ \max g_0(t) - \min g_0(t) \}$$

$$\text{error} = \mathbb{E}\left[\frac{1}{\sqrt{n}} \| f_{\alpha\delta} - f_0 \|_{\mathbb{R}^n} \right] \quad \text{(estimated)}$$

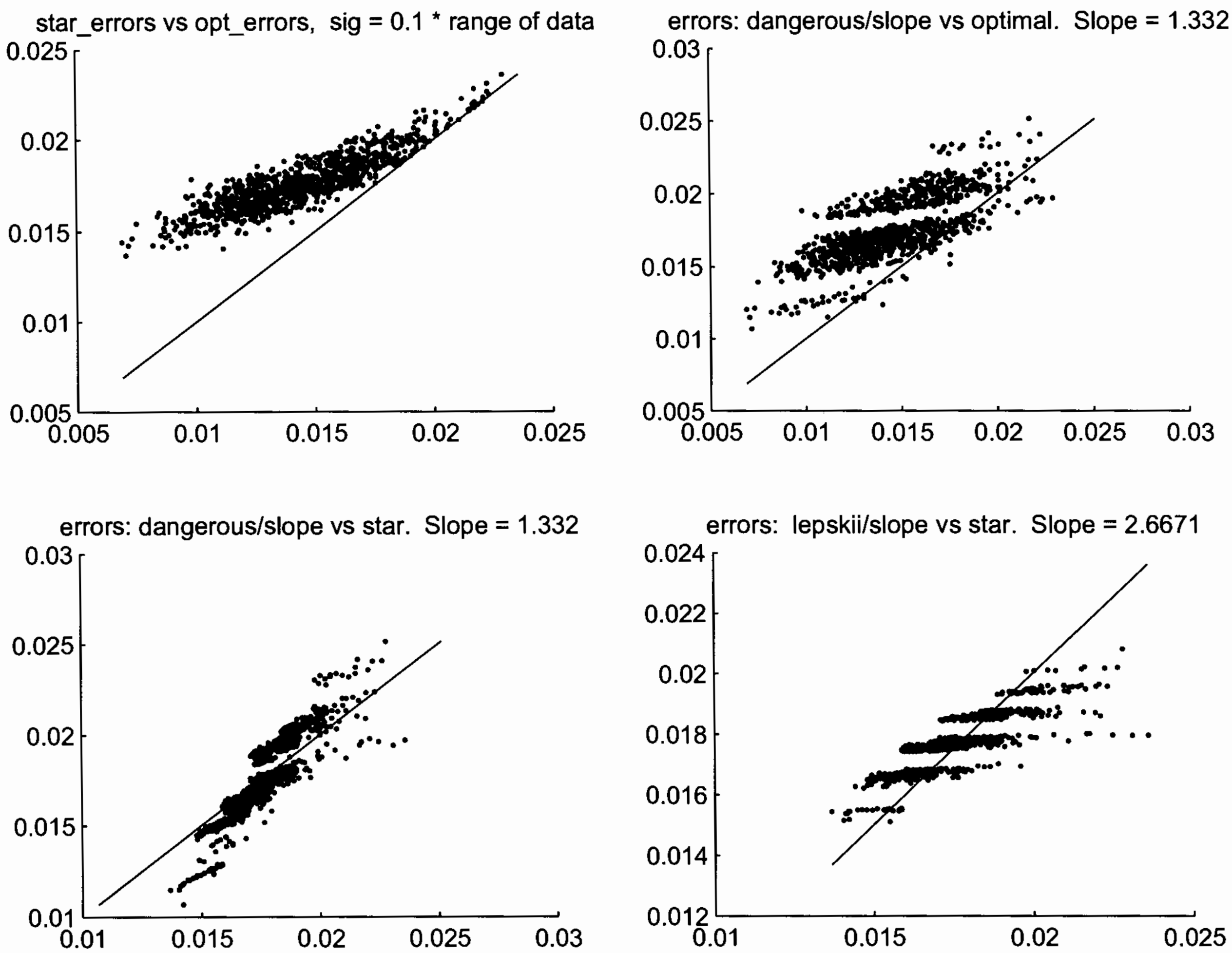
1000 replications

n	optimal	Star	Dangerous	Lepskiĭ
50	$7.13_{10^{-2}}$	$8.35_{10^{-2}}$	$1.26_{10^{-1}}$	$2.10_{10^{-1}}$
100	$5.98_{10^{-2}}$	$7.04_{10^{-2}}$	$9.70_{10^{-2}}$	$1.79_{10^{-1}}$
200	$5.00_{10^{-2}}$	$5.89_{10^{-2}}$	$7.83_{10^{-2}}$	$1.50\ 10^{-1}$
400	$4.21_{10^{-2}}$	$4.94_{10^{-2}}$	$6.47_{10^{-2}}$	$1.24\ 10^{-1}$

$$g_0(t) = (t(1-t))^3$$

$$\Psi(\alpha) = \frac{2\sigma}{\sqrt{n}} \alpha^{-9/16}$$

$$\text{Observed rate} \approx n^{-9/32}$$


star_errors vs opt_errors, sig = 0.1 * range of data
0.025
0.02
0.015
0.01
0.005
0.005
0.01
0.015
0.02
0.025
errors: dangerous/slope vs optimal. Slope = 1.332
0.03
0.025
0.02
0.015
0.01
0.005
0.005
0.01
0.015
0.02
0.025
0.03
errors: dangerous/slope vs star. Slope = 1.332
0.03
0.025
0.02
0.015
0.01
0.01
0.015
0.02
0.025
0.03
errors: lepskii/slope vs star. Slope = 2.6671
0.024
0.022
0.02
0.018
0.016
0.014
0.012
0.01
0.015
0.02
0.025