Statistical Inverse Problems : Theory II.2
(Spectral Regularisation)

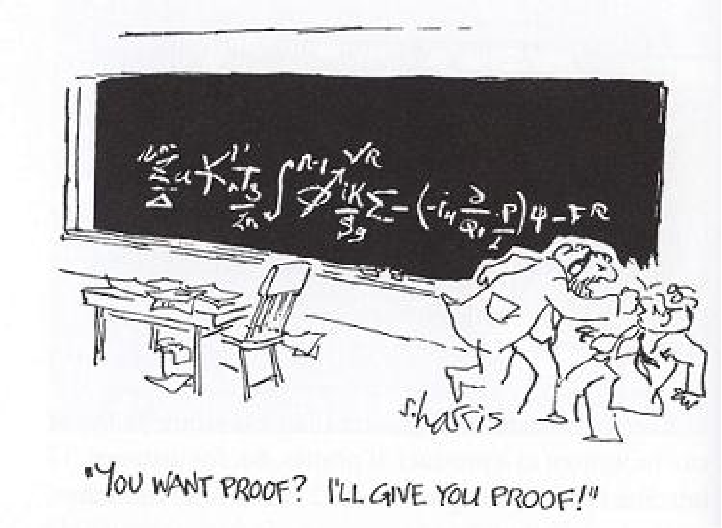
Axel Munk

Institut für Mathematische Stochastik, Göttingen

`www.stochastik.math.uni-goettingen.de/munk`
`munk@math.uni-goettingen.de`

January 2010

Apologizes II



Agenda

1. Linear Spectral Methods
    - Model and Examples
    - Reconstruction Methods
2. Risk bounds, MISE computations
    - Abstract Nonsense
    - Spectral Estimators and its Performance
    - Example: boosting
3. Summary

Continuous SIP Model

$\mathbb{H}_1, \mathbb{H}_2$ separabel, $K : \mathbb{H}_1 \to \mathbb{H}_2$, linear, injective, bounded

$$Y = Kf + \delta_{det}\,\xi_{det} + \delta_{ran}\,\xi_{ran}$$

Deterministic error: $||\xi_{det}|| = 1, \quad \delta_{det}$ noise level

Random noise: ξ_{ran} H-valued r.v. (H-space process), centered.

$||\mathbf{COV}\xi_{ran}|| = 1$, δ^2_{ran} variance.

Inverse regression, Fredholm first kind

Example

$$\begin{aligned} Y_i &= Kf(X_i) + \epsilon_i, \quad (X_i, \epsilon_i),\ i = 1, \cdots, n,\ i.i.d. \\ \mathbf{E}[Y|X] &= Kf(X), \qquad K \text{ lin. integral operator} \end{aligned}$$

Generalized empirical process (GEP)

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

unbiased for $q = (K^* K)f$, $\sqrt{n}$-consistent. ξ_{ran} ist defined as

$$\delta_{ran} K^* \xi_{ran} = \hat{q}_n - q$$

K^* is called preconditioner

Regression, density estimation

Example

Regression

$$\begin{aligned} Y_i &= f(X_i) + \epsilon_i, \quad (X_i, \epsilon_i),\ i = 1, \cdots, n,\ i.i.d. \\ \mathbf{E}\left[Y|X\right] &= f(X) \end{aligned}$$

Smoothness assumption, e.g. f in a Sobolev space

$$W_2^{\beta} = \left\{ f : f\ (\beta - 1) \text{ times cont. diffb. and} \int |f^{\beta}|^2 < \infty \right\}$$

Regression, density estimation

Example

Regression

$$\begin{aligned} Y_i &= Kf(X_i) + \epsilon_i, \quad (X_i, \epsilon_i),\ i = 1, \cdots, n,\ i.i.d. \\ \mathbf{E}[Y|X] &= Kf(X) \end{aligned}$$

Smoothness assumption, e.g. f in a Sobolev space

$$W_2^\beta = \left\{ f : f\ (\beta - 1) \text{ times cont. diffb. and } \int |f^\beta|^2 < \infty \right\}$$

Then $\mathbf{E}[Y|X] = Kf()$, the embedding operator (compact) from W_2^β in $L^2(X)$. $\mathbb{H}_1$ is a RKHS with reproducing kernel $k(x, \cdot)$.
Generalized empirical process (GEP)

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

Regression, density estimation

Example

Regression

$$\begin{aligned} Y_i &= Kf(X_i) + \epsilon_i, \quad (X_i, \epsilon_i),\ i = 1, \cdots, n,\ i.i.d. \\ \mathbf{E}[Y|X] &= Kf(X) \end{aligned}$$

Smoothness assumption, e.g. f in a Sobolev space

$$W_2^\beta = \left\{ f : f\ (\beta - 1) \text{ times cont. diffb. and } \int |f^\beta|^2 < \infty \right\}$$

Then $\mathbf{E}[Y|X] = Kf()$, the embedding operator (compact) from W_2^β in $L^2(X)$. $\mathbb{H}_1$ is a RKHS with reproducing kernel $k(x, \cdot)$.
Generalized empirical process (GEP)

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

Error in variables, deconvolution

Example

$$X_1, \cdots, X_n \sim X = F + W$$

F, W independet with densities $f, w \in L^2$, (w known).
Goal: Estimate f

$$g = Kf = w * f$$

$$\hat{q}_n(\cdot) \quad = \quad \frac{1}{n}\sum_{i=1}^{n} w(X_j - \cdot)$$

unbiased, $\sqrt{n}$-consistent for $q = K^* K f$.

$$\delta_{ran} K^* \xi_{ran} = \hat{q}_n - q$$

Regularisation scheme

$K : \mathbb{H}_1 \to \mathbb{H}_2$, linear, bounded (hence continuous), K injective.
Assumption: K compact, i.e. $cl(K(\mathbb{H}_1))$ compact.

Theorem

Assume, K compact. Then the (linear) inverse is unbounded (not continuous).

Fredholm I operators are compact.

Generalized LSE (Moore-Penrose inverse):

$$K^\dagger = (K^*K)^{-1}K^* : R(K) \oplus R(K)^\perp \to \mathbb{H}_1$$

unbounded (not continuous).

Regularised inverse: Approximate $K^\dagger$ by a sequence of bounded operators $R_\alpha, \alpha > 0$.

Ridge regression

Example

(Tikhonov regularisation / ridge regression / MOR, penalised least squares, Bayes with normal prior, ...)

$$||Kf - Y||^2 + \alpha ||f||^2 \longrightarrow \mathbf{min}$$

Tikhonov estimator (Tikhonov/Arsenin'77, Nychka/Cox '89, Ann. Stat., Mathe, Pereverzev'01, SIAM Num. Anal.,...)

$$\hat{f}_\alpha \quad = \quad (K^*K + \alpha I)^{-1} K^* Y,$$

$\alpha > 0$ regularisation parameter.

- Requires inversion of an operator

Ridge regression

Example

(Tikhonov regularisation / ridge regression / MOR, penalised least squares, Bayes with normal prior, ...)

$$||Kf - Y||^2 + \alpha||f||^2 \longrightarrow \mathbf{min}$$

Tikhonov estimator (Tikhonov/Arsenin'77, Nychka/Cox '89, Ann. Stat., Mathe, Pereverzev'01, SIAM Num. Anal.,...)

$$\hat{f}_\alpha \quad = \quad (K^*K + \alpha I)^{-1}K^*Y,$$

$\alpha > 0$ regularisation parameter.

- Requires inversion of an operator

Singular value decomposition

Theorem

$K : \mathbb{H}_1 \to \mathbb{H}_2$ compact, injectice. Then there exists an ON system $\phi_1, \ldots$ of $\mathbb{H}_1$ and $g_1, \cdots$ of $\mathbb{H}_2$ and singular values $\sigma_0 \geq \sigma_1 \geq \cdots > 0$, s.t. for any $f \in \mathbb{H}_1$

$$Kf = \sum_{n=0}^{\infty} \sigma_n < f, \phi_n > g_n.$$

Definition

Singular system: $(\sigma_n, \phi_n, g_n)_{n \in \mathbb{N}}$

Singular value decomposition

Theorem

$K : \mathbb{H}_1 \rightarrow \mathbb{H}_2$ compact, injectice. Then there exists an ON system $\phi_1, \ldots$ of $\mathbb{H}_1$ and $g_1, \cdots$ of $\mathbb{H}_2$ and singular values $\sigma_0 \geq \sigma_1 \geq \cdots > 0$, s.t. for any $f \in \mathbb{H}_1$

$$Kf = \sum_{n=0}^{\infty} \sigma_n < f, \phi_n > g_n.$$

Definition

Singular system: $(\sigma_n, \phi_n, g_n)_{n \in \mathbb{N}}$

Picard's theorem

Theorem

Let K linear, compact, $(\sigma_n, \phi_n, g_n)_{n\in\mathbb{N}}$ a singular system. Let $Kf = g$ solvable ($g \in N(T^*)^\perp$) and

$$\sum_{n=0}^{\infty} \frac{1}{\sigma_n^2} | < g, g_n > |^2 < \infty.$$

Then the solution is given by

$$f = \sum_{i=0}^{\infty} \frac{1}{\sigma_n} < g, g_n > \phi_n$$

Definition

If $\sigma_n \geq Cn^{-p}$ mildly ill posed (PS),
if $\sigma_n \leq C\exp(-n^p)$ severly ill posed (ES).

Picard's theorem

Theorem

Let K linear, compact, $(\sigma_n, \phi_n, g_n)_{n\in\mathbb{N}}$ *a singular system. Let* $Kf = g$ *solvable* ($g \in N(T^*)^\perp$) *and*

$$\sum_{n=0}^{\infty} \frac{1}{\sigma_n^2} | < g, g_n > |^2 < \infty.$$

Then the solution is given by

$$f = \sum_{i=0}^{\infty} \frac{1}{\sigma_n} < g, g_n > \phi_n$$

Definition

If $\sigma_n \geq Cn^{-p}$ mildly ill posed (PS),
if $\sigma_n \leq C \exp(-n^p)$ severly ill posed (ES).

Spectral cut off, Orthogonal series estimator

Example

$$g = Kf = \sum_{n=0}^{\infty} \sigma_n < f, \phi_n > g_n,$$

singular system $(\sigma_n, \phi_n, g_n)_{n \in \mathbb{N}}$. Under Picard condition:

$$f = \sum_{n=1}^{\infty} \sigma_n^{-1} < g, g_n > \phi_n$$

SC-estimator (Diggle/Hall'93, JRSSB, Mair/Ruymgaart'96,SIAM Appl. Math., Efromovich'97 IEEE Trans. Inf. Theory ...)

$$\hat{f}_\alpha \quad = \quad \sum_{n: \sigma_n \geq \alpha} \sigma_n^{-1} < Y, g_n > \phi_n$$

Spectral cut off, Orthogonal series estimator

SC-estimator

$$\hat{f}_\alpha = \sum_{n:\sigma_n \geq \alpha} \sigma_n^{-1} < Y, g_n > \phi_n$$

- Nice for theoretical computations (minimax)

Spectral cut off, Orthogonal series estimator

SC-estimator

$$\hat{f}_\alpha = \sum_{n:\sigma_n \geq \alpha} \sigma_n^{-1} < Y, g_n > \phi_n$$

- Nice for theoretical computations (minimax)
- Easy to compute

Spectral cut off, Orthogonal series estimator

SC-estimator

$$\hat{f}_\alpha = \sum_{n:\sigma_n \geq \alpha} \sigma_n^{-1} < Y, g_n > \phi_n$$

- Nice for theoretical computations (minimax)
- Easy to compute
- Requires full spectral information

Spectral cut off, Orthogonal series estimator

SC-estimator

$$\hat{f}_\alpha = \sum_{n:\sigma_n \geq \alpha} \sigma_n^{-1} < Y, g_n > \phi_n$$

- Nice for theoretical computations (minimax)
- Easy to compute
- Requires full spectral information

Iterative methods: Landweber'51 iteration/L^2-boosting

Example

Assume no noise. Rewrite $Kf = Y$ as

$$f = (I - \beta K^* K)f + \beta K^* Y$$

and iterate

$$f_{k+1} = (I - \beta K^* K)f_k + \beta K^* Y.$$

This is steepest decent with stepsize β applied to $f \rightarrow ||Y - Kf||^2$:

$$f_{k+1} = f_k - \beta K^*(Kf_k - Y)$$

Iterative methods: Landweber'51 iteration/L^2-boosting

Example

Assume no noise. Rewrite $Kf = Y$ as

$$f = (I - \beta K^* K)f + \beta K^* Y$$

and iterate

$$f_{k+1} = (I - \beta K^* K)f_k + \beta K^* Y.$$

This is steepest decent with stepsize β applied to $f \to ||Y - Kf||^2$:

$$f_{k+1} = f_k - \beta K^*(Kf_k - Y)$$

Iterative methods: Landweber'51 iteration/L^2-boosting

Example

Initialize by a 'weak learner' (e.g. spline):

$$\hat{1}/n \sum_{j=1}^{n} Y_i k(\cdot, X_i) =: K^* Y$$

Iterate ($\beta = 1$):

$$f_0 = 0,\ \hat{f}_{k+1} \quad = \quad (I - K^*K)\hat{f}_k + K^*Y, \quad k = 0, 1, \ldots$$

This gives (linear recursion, $Y = Kf$+noise)

$$\begin{aligned} L_m(Y) \quad &:= \quad \sum_{k=0}^{m-1} (I - K^*K)^k K^* Y \\ &= \quad (I - (I - K^*K)^m) f + \text{noise} \end{aligned}$$

Here $1/k \sim \alpha$ is a regularisation parameter!

Iterative methods: Landweber'51 iteration/L^2-boosting

Example

Initialize by a 'weak learner' (e.g. spline):

$$\hat{1}/n \sum_{j=1}^{n} Y_i k(\cdot, X_i) =: K^* Y$$

Iterate ($\beta = 1$):

$$f_0 = 0,\ \hat{f}_{k+1} \quad = \quad (I - K^* K)\hat{f}_k + K^* Y, \quad k = 0, 1, \dots$$

This gives (linear recursion, $Y = Kf$+noise)

$$\begin{aligned} L_m(Y) \quad &:= \quad \sum_{k=0}^{m-1} (I - K^* K)^k K^* Y \\ &= \quad (I - (I - K^* K)^m) f + \text{noise} \end{aligned}$$

Here $1/k \sim \alpha$ is a regularisation parameter!

Other methods

- ν-methods
- iterated Tikhonov
- tapered OSE's, ...

Goal: Analyze linear spectral methods in the general SIP model

$$Y = Kf + \delta_{det}\,\xi_{det} + \delta_{ran}\,\xi_{ran}$$

Other methods

- ν-methods
- iterated Tikhonov
- tapered OSE's, ...

Goal: Analyze linear spectral methods in the general SIP model

$$Y = Kf + \delta_{det}\,\xi_{det} + \delta_{ran}\,\xi_{ran}$$

Other approaches

Many ways of analysis and refinements:

- gaussian sequence model (Cavalier, Tsybakov'02 PTRF), ...
- oracle inequalties (Cavalier et al.'02, Ann. Stat.), ...

Other approaches

Many ways of analysis and refinements:

- gaussian sequence model (Cavalier, Tsybakov'02 PTRF), ...
- oracle inequalties (Cavalier et al.'02, Ann. Stat.), ...
- SVD Johnstone, Silverman'90 Ann. Stat.
- wavelets, vaguelettes, (Donoho'95, Appl. Comp. Harm. Anal., Abramovich, Silverman'98, Biometrika, Candes, Donoho'02, Ann. Stat., Johnstone et al.'04 JRSSB...)

Other approaches

Many ways of analysis and refinements:

- gaussian sequence model (Cavalier, Tsybakov'02 PTRF), ...
- oracle inequalties (Cavalier et al.'02, Ann. Stat.), ...
- SVD Johnstone, Silverman'90 Ann. Stat.
- wavelets, vaguelettes, (Donoho'95, Appl. Comp. Harm. Anal., Abramovich, Silverman'98, Biometrika, Candes, Donoho'02, Ann. Stat., Johnstone et al.'04 JRSSB...)

We will use Halmos'63 spectral theorem (Mair,Ruymgaart'96, SIAM Num. Anal.)

Other approaches

Many ways of analysis and refinements:

- gaussian sequence model (Cavalier, Tsybakov'02 PTRF), ...
- oracle inequalties (Cavalier et al.'02, Ann. Stat.), ...
- SVD Johnstone, Silverman'90 Ann. Stat.
- wavelets, vaguelettes, (Donoho'95, Appl. Comp. Harm. Anal., Abramovich, Silverman'98, Biometrika, Candes, Donoho'02, Ann. Stat., Johnstone et al.'04 JRSSB...)

We will use Halmos'63 spectral theorem (Mair,Ruymgaart'96, SIAM Num. Anal.)

Abstract Nonsense: Functional Calculus

- So far, K compact (SVD).
- Convolution is bounded, linear, but not compact.

Abstract Nonsense: Functional Calculus

- So far, K compact (SVD).
- Convolution is bounded, linear, but not compact.
- Halmos multiplicative spectral theorem: General representation of linear methods:

Theorem (Halmos)

$A = K^*K$ selfadjoint, bounded, $A : D(A) \to \mathbb{H}$, $D(A)$ dense in $\mathbb{H}$, separable. Then there exists a σ -finite measure space $(S, \mathcal{S}, \Sigma)$, unitary operator $U : \mathbb{H} \to L^2(\Sigma)$, $\rho : S \to \mathbb{R}$ σ-finite measure (spectral measure), a measurable function $\rho : S \to \mathbb{R}$, s.t.

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

where M_ρ is the multiplication operator.

Abstract Nonsense: Functional Calculus

- So far, K compact (SVD).
- Convolution is bounded, linear, but not compact.
- Halmos multiplicative spectral theorem: General representation of linear methods:

Theorem (Halmos)

$A = K^*K$ selfadjoint, bounded, $A : D(A) \rightarrow \mathbb{H}$, $D(A)$ dense in $\mathbb{H}$, separable. Then there exists a σ -finite measure space $(S, \mathcal{S}, \Sigma)$, unitary operator $U : \mathbb{H} \rightarrow L^2(\Sigma)$, $\rho : S \rightarrow \mathbb{R}$ σ-finite measure (spectral measure), a measurable function $\rho : S \rightarrow \mathbb{R}$, s.t.

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

where M_ρ is the multiplication operator.

functional calculus: $\Phi : \sigma(A) \to \mathbb{R}$, bounded, measurable,

$$\Phi(A) \quad = \quad U^* M_{\Phi(\rho)} U$$

Do whatever you want, but do it on the spectrum!

functional calculus: $\Phi : \sigma(A) \to \mathbb{R}$, bounded, measurable,

$$\Phi(A) \quad = \quad U^* M_{\Phi(\rho)} U$$

Do whatever you want, but do it on the spectrum!

Example

A compact:

$$Af = \sum_{i=1}^{\infty} \rho_j < \phi_j, f > \phi_j$$

In multiplicative form (Halmos):

- Σ counting measure on $S = \mathbb{N}$
- multiplicator operator: $\rho_i = \rho(i),\ i \in \mathbb{N}$
- unitary operator: $U(f)(i) =< f, \phi_i >$

Deconvolution

k convolution kernel, then the unitary operator is Fourier transform

$$Uf(\xi) = \psi_f(\xi) = (2\pi)^{-d/2} \int_{\mathbb{R}^d} f(x) \exp\left(-2i\pi < x, \xi >\right) dx$$

convolution operator: A_k, selfadjoint, $\psi^* = \psi^{-1}$.

Deconvolution

k convolution kernel, then the unitary operator is Fourier transform

$$Uf(\xi) = \psi_f(\xi) = (2\pi)^{-d/2} \int_{\mathbb{R}^d} f(x) \exp\left(-2i\pi < x, \xi >\right) dx$$

convolution operator: A_k, selfadjoint, $\psi^* = \psi^{-1}$.

Further, we know: $\psi_{k*f} = (2\pi)^d \psi_k \cdot \psi_f$.

Deconvolution

k convolution kernel, then the unitary operator is Fourier transform

$$Uf(\xi) = \psi_f(\xi) = (2\pi)^{-d/2} \int_{\mathbb{R}^d} f(x) \exp\left(-2i\pi < x, \xi >\right) dx$$

convolution operator: A_k, selfadjoint, $\psi^* = \psi^{-1}$.

Further, we know: $\psi_{k*f} = (2\pi)^d \psi_k \cdot \psi_f$.

Hence, with $\rho = (2\pi)^d \psi_k$,

$$A_k = \psi^{-1} \cdot (2\pi)^d \psi_k \cdot \psi$$

Deconvolution

k convolution kernel, then the unitary operator is Fourier transform

$$Uf(\xi) = \psi_f(\xi) = (2\pi)^{-d/2} \int_{\mathbb{R}^d} f(x) \exp\left(-2i\pi < x, \xi >\right) dx$$

convolution operator: A_k, selfadjoint, $\psi^* = \psi^{-1}$.

Further, we know: $\psi_{k*f} = (2\pi)^d \psi_k \cdot \psi_f$.

Hence, with $\rho = (2\pi)^d \psi_k$,

$$A_k = \psi^{-1} \cdot (2\pi)^d \psi_k \cdot \psi$$

Regularized approximation

Definition (regularised inverse)

Approximate $K^\dagger$ by $\Phi_\alpha(K^*K)K^*$.
Requires conditions:

$$\lim_{\alpha\to 0} \Phi_\alpha(t) = 1/t, \qquad t \in \sigma(A)$$
$$\sup_{t\in\sigma(A)} |t\Phi_\alpha(t)| \leq C \quad \text{unif. in } \alpha$$

This implies

$$||\Phi_\alpha(K^*K)K^*Kf - f|| \to 0, \qquad \alpha \searrow 0.$$

Spectral regularisation estimator (SRE)

$$\hat{f}_\alpha = \Phi_\alpha(K^*K)K^*Y$$

How good is an estimator?

smoothness classes: $\Lambda : [0,\infty) \to [0,\infty)$ continuous, $\nearrow$, $\Lambda(0) = 0$

$$\mathcal{F}_\Lambda \quad = \quad \{f : f = \Lambda(K^*K)w : w \in \mathbb{H}_1,\ ||w|| \leq 1\}$$

$\Lambda(t) = t^\nu$: Hölder-type source condition (PS), e.g.:

$$Kf(x) \quad = \quad \int_0^x f(t)dt$$

$$R(K^*K)^\mu = H^{2\mu}[0,2\pi], \quad \sum_{n\in\mathbb{Z}} (1+n^2)^{2\mu} |\hat{\varphi}(n)|^2 < \infty$$

Logarithmic source condition (ES) (Hohage, '00, Num. Funct. Anal. Opt.):

$$\Lambda(t) = (-\ln(t))^{-p}$$

How good is an estimator?

smoothness classes: $\Lambda : [0,\infty) \to [0,\infty)$ continuous, $\nearrow$, $\Lambda(0) = 0$

$$\mathcal{F}_\Lambda \quad = \quad \{f : f = \Lambda(K^*K)w : w \in \mathbb{H}_1,\ ||w|| \leq 1\}$$

$\Lambda(t) = t^\nu$: Hölder-type source condition (PS), e.g.:

$$Kf(x) \quad = \quad \int_0^x f(t)dt$$

$$R(K^*K)^\mu = H^{2\mu}[0,2\pi], \quad \sum_{n\in\mathbb{Z}} (1+n^2)^{2\mu} |\hat{\varphi}(n)|^2 < \infty$$

Logarithmic source condition (ES) (Hohage, '00, Num. Funct. Anal. Opt.):

$$\Lambda(t) = (-\ln(t))^{-p}$$

Approximation error

Approximation error for exact data $g = Kf$:

$$\begin{aligned} \|\Phi_\alpha(K^*K)K^*g - f\| &= \|(\Phi_\alpha(A)A - I)\Lambda(A)w\| \\ &\leq \sup_{t\in\sigma(A)} |(\Phi_\alpha(t)t - 1)\Lambda(t)| \end{aligned}$$

Qualification

$$\sup_{t\in\sigma(A)} |t^{\nu}(1-t\Phi_{\alpha}(t))| \le \gamma_{\nu}\alpha^{\nu}, \qquad \text{for all } \alpha \text{ and } 0 \le \nu \le \nu_0.$$

This measures the maximal degree of smoothness for which a method converges of optimal order.
More general assumptions: (Hohage'00, Mathé/Pereverzev'03, exp. ill posed problems)

$$\sup_{t\in\sigma(A)} |\Lambda(t)(1-t\Phi_{\alpha}(t))| \le \gamma_{\Lambda}\Lambda(\alpha), \qquad \alpha \searrow 0.$$

Linear regularisation methods

method of regularisation	$\Phi_\alpha(t)$	$1 - t\Phi_\alpha(t)$	ν_0
spectral cut-off	$\begin{cases} t^{-1}, & t \geq \alpha \\ 0, & t < \alpha \end{cases}$	$\begin{cases} 0, & t \geq \alpha \\ 1, & t < \alpha \end{cases}$	∞
Tikhonov regularization	$\frac{1}{t+\alpha}$	$\frac{\alpha}{t+\alpha}$	1
iterated Tikhonov, $k \geq 1$	$\frac{(t+\alpha)^k-\alpha^k}{t(t+\alpha)^k}$	$\left(\frac{\alpha}{t+\alpha}\right)^k$	k
Landweber iteration	$\sum_{j=0}^{k-1}(1-t)^j$	$(1-t)^k$	∞
ν-method ($\nu > 0$)	$\in P_{k-1}$	$\in P_k$	ν
Pinsker filter	$(1-\kappa(t))_+/t$	$1-(1-\kappa(t))_+$	β

Table: Spectral methods of regularization

MISE

Mean Integrated Square Error:

$$\mathbf{E}\,\|\hat{f}_\alpha - f\|^2 \quad = \quad \underbrace{\mathbf{B}^2(\hat{f}_{\alpha,\delta})}_{\text{bias}} + \underbrace{\mathbf{E}\,\|\hat{f}_\alpha - \mathbf{E}\,\hat{f}_\alpha\|^2}_{\text{variance}},$$

Deterministic error:

$$\mathbf{B}(\hat{f}_\alpha) \quad := \quad \|\mathbf{E}\,\hat{f}_\alpha - f\| \le \underbrace{\|\Phi_\alpha(A)Af - f\|}_{\text{approx. error}} + \delta_{\text{det}} \underbrace{\|\Phi_\alpha(A)K^*\xi_{\text{det}}\|}_{\text{propag. det. noise } \le \sqrt{C/\alpha}}$$

Approx. error $\le \gamma_\Lambda \Lambda(\alpha),\ \alpha \to 0$.

MISE: Bounding the variance

Statistical noise: variance

$$\begin{aligned} \mathbf{E}\,\|\hat{f}_\alpha - \mathbf{E}\,\hat{f}_\alpha\|^2 &= \delta_{ran}^2 \mathbf{E}\,\|\Phi_\alpha(\rho) U K^* \xi_{ran}\|^2 \\ &\le \delta_{ran}^2 C \int_{\mathbb{S}} \Phi_\alpha^2(\rho)\rho \,\mathrm{d}\Sigma. \end{aligned}$$

if

$$C_1\rho(s) \le \mathrm{Var}(UK^*\xi_{ran}(s)) \le C_2\rho(s).$$

Assumption:

$$\Sigma\{\rho \ge \alpha\} \sim S(\alpha), \quad \alpha \searrow 0$$

for S decreasing, smooth.

MISE: upper bound

This yields

$$\mathbf{E}\,\|\hat{f}_\alpha - \mathbf{E}\,\hat{f}_\alpha\|^2 \quad\le\quad \frac{C_2\delta_{ran}^2}{\alpha^2}\int_0^\alpha S(\beta)\,\mathrm{d}\beta, \qquad \alpha \searrow 0.$$

Theorem (Bissantz, Hohage, M., Ruymgaart'07, SIAM Num. Anal.)

$$\mathbf{E}\,\|\hat{f}_{\alpha,\delta} - f\|^2 \lesssim \left(\gamma_\Lambda\Lambda(\alpha) + \sqrt{\frac{C_1}{\alpha}}\delta_{det}\right)^2 + \frac{C_2\delta_{ran}^2}{\alpha^2}\int_0^\alpha S(\beta)\,\mathrm{d}\beta,\ \alpha \searrow 0$$

uniformly in $f \in \mathcal{F}_\Lambda$.

Upper bound: Spectral Methods

Theorem (BHMR'07)

Let $\delta_{det} = 0$*. Then there is a constant* $C > 0$ *s.t.*

$$\sup_{f \in \mathcal{F}_\Lambda} \mathbf{E} \|\hat{f}_\alpha - f\|^2 \leq C \sup_{f \in \mathcal{F}_\Lambda} \mathbf{E} \|\hat{f}_\alpha^{\mathrm{SC}} - f\|^2$$

for all $\delta_{ran} > 0$ *and all* $\alpha > 0$ *sufficiently small.*

Corollary

Spectral cut off achieves statistical minimax rates (Mair, Ruymgaart'96) and hence all spectral methods achieve statistical minimax rates within its qualification! α *has to be chosen s.t. the r.h.s of the upper bound is minimized (balanced).*

Moore on boosting

Weak learner (smoothing spline operator)

$$f_{0,n} := K^* Y = \operatorname*{argmin}_{f \in W_2^\nu} \frac{1}{n} \sum_{i=1}^{n} (Y_i - f(x_i))^2 + c \int |f^\nu|^2$$

The eigenvalues behave like (c fixed) $k^{-2\nu}$

- Assume: Eigenvalues of RKHS kernel k of the weak learner (e.g. spline in W_2^ν) $\kappa_0 \geq \kappa_1 \geq \cdots$ behave as

$$\#\{\kappa_j \geq \alpha\} \asymp \alpha^{-2\nu},\ \alpha \searrow 0$$

- Let $f \in W_2^\beta$, $\beta \geq \nu$
- Put $\sigma \asymp n^{-1/2}$, $\alpha \asymp 1/k$
- Then optimal stopping index is $k^*(n) \asymp n^{(2\nu)/(2\beta+1)}$ with MISE (minimax optimal in Sobolev ball $W_2^{\beta,L}$)

$$\mathbf{E}\,||\hat{f}_{k,n} - f||^2 \asymp n^{-2\beta/(2\beta+1)}$$

Moore on boosting

Weak learner (smoothing spline operator)

$$f_{0,n} := K^* Y = \operatorname*{argmin}_{f \in W_2^\nu} \frac{1}{n} \sum_{i=1}^{n} (Y_i - f(x_i))^2 + c \int |f^\nu|^2$$

The eigenvalues behave like (c fixed) $k^{-2\nu}$

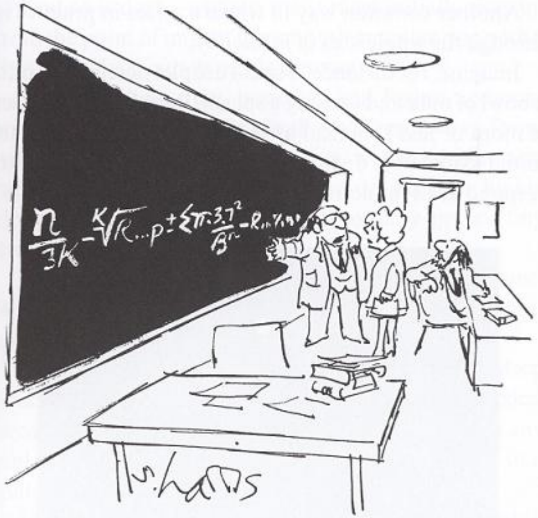
- Assume: Eigenvalues of RKHS kernel k of the weak learner (e.g. spline in W_2^ν) $\kappa_0 \geq \kappa_1 \geq \cdots$ behave as

$$\#\{\kappa_j \geq \alpha\} \asymp \alpha^{-2\nu},\ \alpha \searrow 0$$

- Let $f \in W_2^\beta$, $\beta \geq \nu$
- Put $\sigma \asymp n^{-1/2}$, $\alpha \asymp 1/k$
- Then optimal stopping index is $k^*(n) \asymp n^{(2\nu)/(2\beta+1)}$ with MISE (minimax optimal in Sobolev ball $W_2^{\beta,L}$)

$$\mathbf{E}\,||\hat{f}_{k,n} - f||^2 \asymp n^{-2\beta/(2\beta+1)}$$

Apologizes III



"BUT THIS IS THE SIMPLIFIED VERSION FOR THE GENERAL PUBLIC."

Linear statistical inverse problems and spectral regularisation

- $Y = Kf + \epsilon$ is a rich model.
- Nonparametric regression $\subset$ SIP (embedding)
- BHMR'07: Unifying error analysis of general spectral regularisation.

Linear statistical inverse problems and spectral regularisation

- $Y = Kf + \epsilon$ is a rich model.
- Nonparametric regression $\subset$ SIP (embedding)
- BHMR'07: Unifying error analysis of general spectral regularisation.
- Methods are mainly discriminated by their qualification, i.e. approximation error.

Linear statistical inverse problems and spectral regularisation

- $Y = Kf + \epsilon$ is a rich model.
- Nonparametric regression $\subset$ SIP (embedding)
- BHMR'07: Unifying error analysis of general spectral regularisation.
- Methods are mainly discriminated by their qualification, i.e. approximation error.
- Minimax rates can be achieved.

Linear statistical inverse problems and spectral regularisation

- $Y = Kf + \epsilon$ is a rich model.
- Nonparametric regression $\subset$ SIP (embedding)
- BHMR'07: Unifying error analysis of general spectral regularisation.
- Methods are mainly discriminated by their qualification, i.e. approximation error.
- Minimax rates can be achieved.
- Some caution with minimax is advisable (recall Gaussian deconvolution)

Linear statistical inverse problems and spectral regularisation

- $Y = Kf + \epsilon$ is a rich model.
- Nonparametric regression $\subset$ SIP (embedding)
- BHMR'07: Unifying error analysis of general spectral regularisation.
- Methods are mainly discriminated by their qualification, i.e. approximation error.
- Minimax rates can be achieved.
- Some caution with minimax is advisable (recall Gaussian deconvolution)
- Adaption possible (Lepski'90, Theor. Prob. Appl.)

Linear statistical inverse problems and spectral regularisation

- $Y = Kf + \epsilon$ is a rich model.
- Nonparametric regression $\subset$ SIP (embedding)
- BHMR'07: Unifying error analysis of general spectral regularisation.
- Methods are mainly discriminated by their qualification, i.e. approximation error.
- Minimax rates can be achieved.
- Some caution with minimax is advisable (recall Gaussian deconvolution)
- Adaption possible (Lepski'90, Theor. Prob. Appl.)

Spectral methods

- Spectral cut off is difficult to perform (complete spectral information required)
- SVD methods require same information
- Tikhonov regularisation (ridge regression) is computationally more expensive and does not achieve minimax rates in general

Spectral methods

- Spectral cut off is difficult to perform (complete spectral information required)
- SVD methods require same information
- Tikhonov regularisation (ridge regression) is computationally more expensive and does not achieve minimax rates in general
- In practice: Iterative methods are often preferred (next talk)

Spectral methods

- Spectral cut off is difficult to perform (complete spectral information required)
- SVD methods require same information
- Tikhonov regularisation (ridge regression) is computationally more expensive and does not achieve minimax rates in general
- In practice: Iterative methods are often preferred (next talk)
- Landweber = L^2-boosting

Spectral methods

- Spectral cut off is difficult to perform (complete spectral information required)
- SVD methods require same information
- Tikhonov regularisation (ridge regression) is computationally more expensive and does not achieve minimax rates in general
- In practice: Iterative methods are often preferred (next talk)
- Landweber $= L^2$-boosting
- Many iterative methods achieve optimal statistical minimax rates, are computationally feasible, but are quite unknown in statistics (ν - methods).

Spectral methods

- Spectral cut off is difficult to perform (complete spectral information required)
- SVD methods require same information
- Tikhonov regularisation (ridge regression) is computationally more expensive and does not achieve minimax rates in general
- In practice: Iterative methods are often preferred (next talk)
- Landweber $= L^2$-boosting
- Many iterative methods achieve optimal statistical minimax rates, are computationally feasible, but are quite unknown in statistics (ν - methods).
- The other way around: Statistical regularisation filters (Pinsker) are sharp minimax, but entirely unknown in numerical analysis/learning theory.

Spectral methods

- Spectral cut off is difficult to perform (complete spectral information required)
- SVD methods require same information
- Tikhonov regularisation (ridge regression) is computationally more expensive and does not achieve minimax rates in general
- In practice: Iterative methods are often preferred (next talk)
- Landweber $= L^2$-boosting
- Many iterative methods achieve optimal statistical minimax rates, are computationally feasible, but are quite unknown in statistics (ν - methods).
- The other way around: Statistical regularisation filters (Pinsker) are sharp minimax, but entirely unknown in numerical analysis/learning theory.

General Remarks

I have missed many important aspects:

- Bayesian regularisation (Kaipio,Somersalo'04)

General Remarks

I have missed many important aspects:

- Bayesian regularisation (Kaipio,Somersalo'04)
- choice of α (SMA)

General Remarks

I have missed many important aspects:

- Bayesian regularisation (Kaipio,Somersalo'04)
- choice of α (SMA)
- nonlinear SIPs (O'Sullivan'90, Ann. Stat., Bissantz, Hohage, M.'04, Inv. Prob.)

General Remarks

I have missed many important aspects:

- Bayesian regularisation (Kaipio,Somersalo'04)
- choice of α (SMA)
- nonlinear SIPs (O'Sullivan'90, Ann. Stat., Bissantz, Hohage, M.'04, Inv. Prob.)

Didn't we forget something?

General Remarks

I have missed many important aspects:

- Bayesian regularisation (Kaipio,Somersalo'04)
- choice of α (SMA)
- nonlinear SIPs (O'Sullivan'90, Ann. Stat., Bissantz, Hohage, M.'04, Inv. Prob.)

Didn't we forget something?

What is a SIP?

General Remarks

I have missed many important aspects:

- Bayesian regularisation (Kaipio,Somersalo'04)
- choice of α (SMA)
- nonlinear SIPs (O'Sullivan'90, Ann. Stat., Bissantz, Hohage, M.'04, Inv. Prob.)

Didn't we forget something?

What is a SIP?

Do you like my definition more now (at least slightly)?

$$\liminf \sqrt{n} \inf_{\hat{f}_n} R(\hat{f}_n) = \infty$$