On Bayesian Estimation of Wavelets Coefficients

Fabrizio Ruggeri

Istituto di Matematica Applicata e Tecnologie Informatiche
Consiglio Nazionale delle Ricerche
Via Alfonso Corti 12, I-20133, Milano, Italy, European Union

fabrizio@mi.imati.cnr.it

www.mi.imati.cnr.it/fabrizio

*Joint work with **Brani Vidakovic**, Gabriel Katul, Luisa Cutillo, Yoon Young Jung and Ilya Lavrik*

INTRODUCTION TO WAVELETS

$\psi \in L^2(\mathbb{R})$:

- $\int_{\mathbb{R}} \psi(x)dx = 0$
- $\int_{\mathbb{R}} \psi^2(x)dx = 1$

$\psi_{j,k}(x) = 2^{j/2}\psi(2^j x - k), j, k \in \mathbb{Z}$

ψ *wavelet* iff $\{\psi_{j,k}\}_{j,k\in\mathbb{Z}}$ orthonormal basis in $L^2(\mathbb{R})$, i.e.

$\forall f \in L^2(\mathbb{R}) \Rightarrow f(x) = \sum_{j,k\in\mathbb{Z}} d_{j,k}\psi_{j,k}(x)$

with $d_{j,k} =< f, \psi_{j,k} >= \int f(x)\psi_{j,k}(x)dx$

INTRODUCTION TO WAVELETS

$\phi(x)$ *scaling function* s.t.

- $\int_{\mathbb{R}} \phi(x)dx = 1$
- $\phi(x) = \sqrt{2}\sum_{k\in\mathbb{Z}} h_k \phi(2x-k)$

h_k filter coefficients ($\exists$ many)

$\Rightarrow \psi(x)$ *wavelet* s.t.

$\psi(x) = \sqrt{2}\sum_{k\in\mathbb{Z}}(-1)^k h_{1-k}\phi(2x-k)$

$\Rightarrow$ wavelet system

$\phi_{j,k}(x) = 2^{j/2}\phi(2^j x - k)$ and

$\psi_{j,k}(x) = 2^{j/2}\psi(2^j x - k), j, k \in \mathbb{Z}$

MULTIRESOLUTION ANALYSIS

$V_j, j \in \mathbb{Z}$, closed subspaces of $L^2(\mathbb{R})$:

- $\ldots V_{-2} \subset V_{-1} \subset V_0 \subset V_1 \subset V_2 \ldots$
- $\bigcap_j V_j = \emptyset$ and $\bigcup_j V_j = L^2(\mathbb{R})$
- $f(x) \in V_j \Leftrightarrow f(2x) \in V_{j+1}$

$V_0 = \{f \in L^2(\mathbb{R}) : f(x) = \sum_k \alpha_k \phi(x-k)\}$ from scaling function ϕ

$\Rightarrow \{\phi_{j,k}\}_{j,k}$ orthonormal basis for V_j

$\Rightarrow P_j f = \sum_k < f, \phi_{j,k} > \phi_{j,k}$

V_j-approximation (orthogonal projection) of $f \in L^2(\mathbb{R})$

$\Rightarrow W_j$ orthogonal complement of V_j in V_{j+1}, spanned by *details* $\psi_{j,k}, k \in \mathbb{Z}$

$\Rightarrow f(x) = \sum_{k\in\mathbb{Z}} c_{j_0,k}\phi_{j_0,k}(x) + \sum_{j\geq j_0}\sum_{k\in\mathbb{Z}} d_{j,k}\psi_{j,k}(x)$

Finite approximations

INTRODUCTION TO WAVELETS

[Haar (1910)]: $\forall f \in \mathcal{C}([0,1]) \Rightarrow$

- $f(x) \approx f_n(x) =< \xi_0, f > \xi_0(x) + < \xi_1, f > \xi_1(x) + \ldots + < \xi_n, f > \xi_n(x)$, with $< \xi_i, f >= \int \xi_i(x) f(x) dx, \forall i$
- f_n converges uniformly to f, as $n \to \infty$

$$\begin{aligned}
\xi_0(x) &= \mathbf{1}(0 \leq x \leq 1) \\
\xi_1(x) &= \mathbf{1}(0 \leq x \leq 1/2) - \mathbf{1}(1/2 \leq x \leq 1) \\
\xi_2(x) &= \sqrt{2}[\mathbf{1}(0 \leq x \leq 1/4) - \mathbf{1}(1/4 \leq x \leq 1/2)] \\
&\ldots \\
\xi_n(x) &= 2^{j/2}[\mathbf{1}(k2^{-j} \leq x \leq (k+1/2)2^{-j}) - \mathbf{1}((k+1/2)2^{-j} \leq x \leq (k+1)2^{-j})]
\end{aligned}$$

$n = 2^j + k, j \geq 0, 0 \leq k \leq 2^j - 1$

Note:

$\xi_n(x) = \xi_{jk}(x) = 2^{j/2}\xi_1(2^j x - k), n = 2^j + k$

$\xi_0 \to$ "average" and $\xi_n(x), n > 0 \to$ "details"

ξ_0 *scaling function* and $\xi_1(x)$ *wavelet*

INTRODUCTION TO WAVELETS

- *Wavelets generate local bases*

  (few basis functions at each point
  $\Rightarrow$ more adaptivity and parsimony)

- *Wavelets filter data*

  (split of periodic functions with different periods)

- *Wavelets ”disbalance” data*

  (time series data $y_i \rightarrow$ wavelets coefficients d_i
  $\Rightarrow$ energy ($\sum y_i^2 = \sum d_i^2$) concentrates in few coefficients)

- *Wavelets whiten data*
  (few coefficients in ACF are significant)

- *Wavelets compress data*
  (FBI: fingerprints compression 20:1)

WAVELETS IN STATISTICS

Topics (out of many ...)

- Signal processing and denoising
- Image compression
- Density estimation
- Regression

Estimation of wavelet coefficients

- Shrinkage
- Thresholding

Methods

- Classical (e.g. Universal thresholding $\lambda = \sqrt{2 \log n}\hat{\sigma}$)
- Empirical Bayes
- **Bayesian**

NONLINEAR REGRESSION

Equispaced points $x_1, \ldots, x_N, N(= 2^n)$

$y_i = f(x_i) + \eta_i, i = 1, N$

- y_i: noisy measurement
- f_i: unknown signal
- η_i: random noise, i.i.d. $\mathcal{N}(0, \sigma^2)$

Apply (discrete) wavelet transform W *(a matrix)*

$\Rightarrow \; d_i = \theta_i + \varepsilon_i, \; i = 1, N$

$[\underline{y} \to \underline{d} = W\underline{y}, \underline{f} \to \underline{\theta} = W\underline{f}, \underline{\eta} \to \underline{\varepsilon} = W\underline{\eta}]$

Estimate θ_i

$\Rightarrow \underline{\widehat{\theta}}$ and $\underline{\widehat{f}} = W^{-1}\underline{\widehat{\theta}}$

BAYESIAN MODELING

Shrinkage and thresholding (e.g. `Dirac at 0`*)*

- Small coefficients (in absolute value)
  $\Rightarrow$ function smoothing and denoising

- Few coefficients keep most of the energy

- Thresholded coefficients $\Rightarrow$ more compression

Level influence (e.g. Covariance Σ)

- Variance on both d and θ might decrease as level increases

- Correlation of coefficients across and within levels

THRESHOLDING

Donoho and Johnstone (1994), Biometrika

- $X_1, \ldots, X_n$ i.i.d. $\mathcal{N}(0, 1)$
- $\Rightarrow P(|X_{(n)}| > \sqrt{c \log n}) \sim \dfrac{\sqrt{2}}{n^{c/2-1}\sqrt{c\pi \log n}}$
- *Universal threshold* $\lambda = \sqrt{2 \log n}\widehat{\sigma}$
- Hard thresholding rule: $\delta_\lambda(d) = d \cdot \mathbf{1}(|d| > \lambda)$
- $\Rightarrow$ no noise in the data after thresholding (with high probability)

THRESHOLDING: BAYES RISK

$d_i = \theta_i + \eta_i,\ i = 1, \ldots, N$

- $d_i|\theta_i \sim f(d_i - \theta_i)$, f symmetric
- $\theta_i \sim \pi(\theta_i)$, π symmetric
- Symmetric loss $L(\theta, a) = L(\theta - a)$
- Hard thresholding rule: $\delta_\lambda(d) = d \cdot \mathbf{1}(|d| > \lambda),\ \ \lambda \geq 0$

$\exists$ *finite* $\lambda^* > 0$ *minimising the Bayes risk*

$r(\pi, \delta_\lambda(d)) = E^\theta E^{D|\theta} L(\theta, \delta_\lambda(d)) = \int_{\mathcal{R}} L(t) f(t) dt + \int_{\mathcal{R}} [\int_{\theta-\lambda}^{\theta+\lambda} [L(\theta) - L(t)] f(t) dt] \pi(\theta) d\theta$?

THRESHOLDING: BAYES RISK

- $f(t) = \pi(t), \forall t \Rightarrow r(\pi, \delta_\lambda(d)) =$ constant

- f and π in the same family (e.g. normal) but with different parameters
  $\Rightarrow \lambda^* = 0$ or ∞

$\Rightarrow$ *Choose f and π from different families*

Example: $d_i|\theta_i \sim \mathcal{N}(\theta_i, \tilde{\sigma}^2)$ and $\theta_i \sim \mathcal{DE}(0, \beta)$

THRESHOLDING: BAYES RISK

Simulation: 100 runs for different sample sizes N from "usual" noisy test signals:
`Blocks, Bumps, Doppler, Heavisine`

Signals scaled s.t.

- Noise: i.i.d. normal, variance $= 1$
- Signal-to-noise ratio (SNR) $= 7$

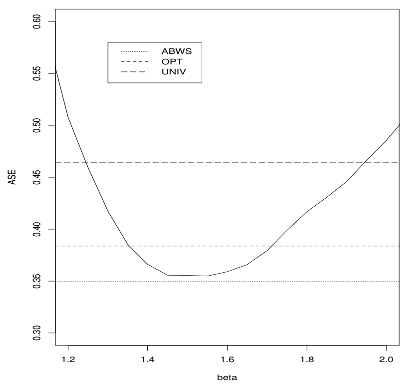
  SNR $= \frac{\text{sample std. dev signal}}{\text{std. dev. noise}}$

Comparison of MSE of estimator $\widehat{y} = (\widehat{y}_1, \ldots, \widehat{y}_n)$

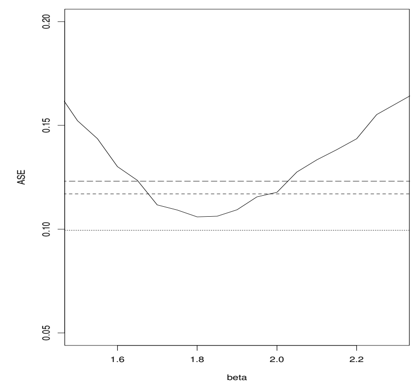
$MSE = \frac{1}{MN} \sum_{j=1}^{M} \sum_{i=1}^{N} (\widehat{y}_{ij} - y_i)^2$

THRESHOLDING: BAYES RISK

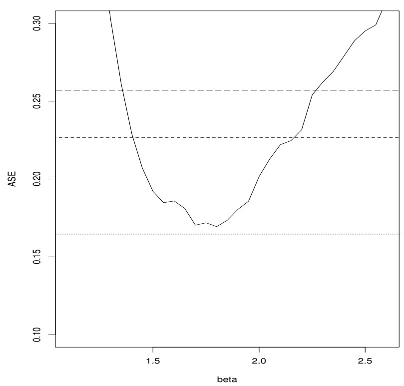
AMSE as a function of β compared with ABWS, optimal minimax, and universal AMSE



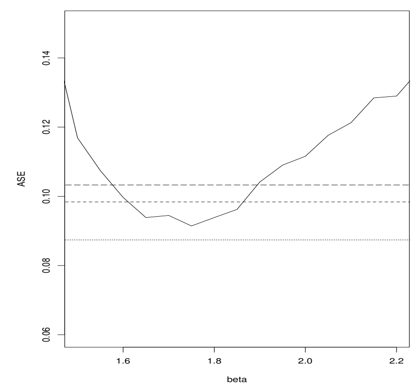
(A) `bumps`



(B) `blocks`



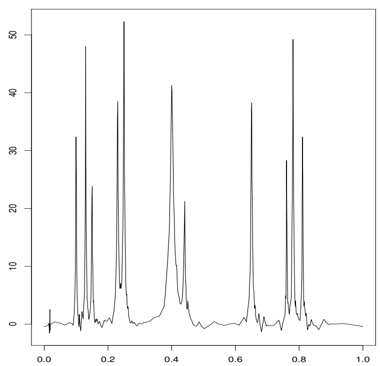
(A) `doppler`



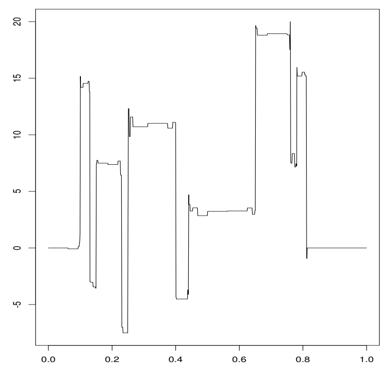
(B) `heavisine`

THRESHOLDING: BAYES RISK

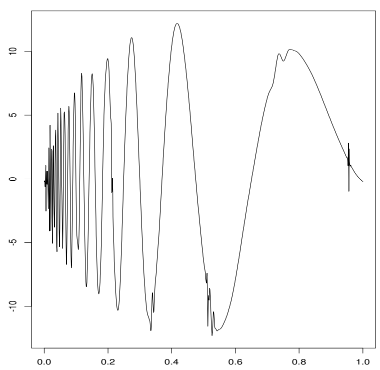
BDT estimators of the test signals



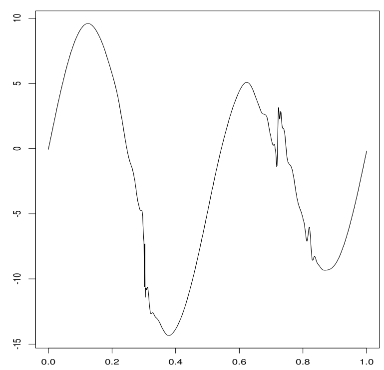
(A) `bumps`



(B) `blocks`



(A) `doppler`



(B) `heavisine`

THRESHOLDING: Bayes risk

Descriptive statistics of MSE for BDT, SNR=7, N=1024

	MSE BDT (MSE Univ)	SD MSE (SD MSE)	Ave λ^* (λ^*/λ_{univ})	SD λ^*
bumps	.378 (.481)	.032 (.041)	2.74 (.737)	.274
blocks	.104 (.122)	.018 (.024)	3.35 (.901)	.366
doppler	.210 (.252)	.025 (.029)	3.12 (.840)	.271
heavisine	.088 (.099)	.015 (.017)	3.52 (.946)	.395

THRESHOLDING: POSTERIOR EXPECTED LOSS

Decision rules $\mathcal{D} = \{\delta_\lambda, \lambda \geq 0\}$: $\forall x \in \mathcal{R}$

- $\delta_0(x) = x$
- $\lim_{\lambda \to \infty} \delta_\lambda(x) = 0$
- $\exists \lambda_x : \delta_\lambda(x) = 0, \lambda \geq \lambda_x$, and continuous, strictly decreasing o.w.

Examples of classes $\mathcal{D}$

Soft-thresholding : $\delta_\lambda^{soft}(x) = (x - \lambda \,\text{sign}(x))\mathbf{1}(|x| > \lambda)$

Rational : $\delta_\lambda^r(x) = x^{2n+1}/(\lambda^{2n} + x^{2n})$

Hyperbola : $\delta_\lambda^{hi}(x) = \text{sign}(x)\sqrt{x^2 - \lambda^2}\mathbf{1}(|x| \geq \lambda)$

Garrote : $\delta_\lambda^g(x) = (x - \lambda^2/x)\mathbf{1}(|x| > \lambda)$

Convex loss $L(\theta, a)$ *and Bayes action* $a_L(x) \Rightarrow a^* = \min\{x, a_L(x)\}$ optimal rule in $\mathcal{D}$
Hard thresholding rules and squared loss $\Rightarrow \delta(x) = x\mathbf{1}(0 \leq x \leq 2E^{\theta|X}\theta)$

SHRINKAGE

Loss $L(\theta, a) = (\theta - a)^2 \Rightarrow E^{\theta|d}\theta$ optimal

$d|\theta \sim f(d|\theta) = f(d - \theta)$, symmetric and unimodal

How to choose prior to have shrinkage,
i.e. $\Delta = |E^{\theta|d}\theta/d| < 1$*?*

- $\Gamma_S = \{\text{all symmetric}\}$
  $\Rightarrow \sup_{\pi \in \Gamma_S} \Delta > 1$ ($= \infty$ for normal model)

- $\Gamma_{Sp} = \{\text{all symmetric + mass } p \text{ at } 0\}$
  $\Rightarrow \sup_{\pi \in \Gamma_S} \Delta < \infty$ but > 1 for "small" p

- $\Gamma_{SU} = \{\text{all symmetric, unimodal}\}$
  $\Rightarrow \sup_{\pi \in \Gamma_S} \Delta \leq 1$

- $\Gamma_S = \{\text{all symmetric, unimodal+ mass } p \text{ at } 0\}$
  $\Rightarrow \sup_{\pi \in \Gamma_S} \Delta < 1$

BAYESIAN MODELING

Model for $d = \theta + \varepsilon$

$d|\theta, \sigma^2 \sim \mathcal{N}(\theta, \sigma^2)$

Many choices of priors on (θ, σ^2)

- $\sigma^2 = \hat{\sigma}^2 \Rightarrow d|\theta \sim \mathcal{N}$
- σ^2 "close" to 0 $\rightarrow \sigma^2 \sim \mathcal{E}(\mu) \Rightarrow d|\theta \sim \mathcal{DE}$
- σ^2 "far" from 0 $\rightarrow \sigma^2 \sim \mathcal{IG}(\mu) \Rightarrow d|\theta \sim t$

Vidakovic

- $\sigma^2 \sim \mathcal{E}(\mu) \Rightarrow d|\theta \sim \mathcal{DE}\left(\theta, 1/\sqrt{2\mu}\right)$
- $\theta \sim t_n(0, \tau) \Rightarrow \delta(d) = d - \dfrac{\Pi_1'(c) - \Pi_2'(c)}{\Pi_1(c) + \Pi_2(c)}$
- Π_1 and Π_2 Laplace transforms of $\pi(\theta + d)$ and $\pi(\theta - d)$, respectively, and $c = \sqrt{2\mu}$

ABWS: Adaptive Bayesian Wavelet Shrinkage

Chipman, Kolaczyk and McCulloch (1997), JASA

$$d|\theta, \sigma^2 \sim N(\theta, \widehat{\sigma}^2)$$

$$\theta|\gamma_j \sim \gamma_j \mathcal{N}(0, (c_j\tau_j)^2) + (1 - \gamma_j)\mathcal{N}(0, \tau_j^2)$$

$$\gamma_j \sim \mathcal{B}er(p_j)$$

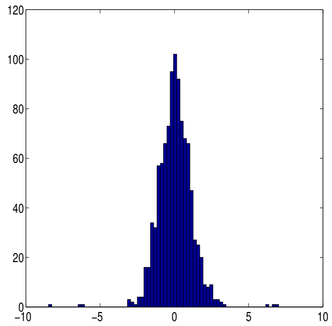
$$\widehat{\theta} = P(\gamma_j = 1|d)\frac{(c_j\tau_j)^2}{\sigma^2 + (c_j\tau_j)^2} + P(\gamma_j = 0|d)\frac{\tau_j^2}{\sigma^2 + \tau_j^2}$$

$$P(\gamma_j = 1|d) = \frac{p_j\pi(d|\gamma_j = 1)}{p_j\pi(d|\gamma_j = 1) + (1 - p_j)\pi(d|\gamma_j = 0)}$$

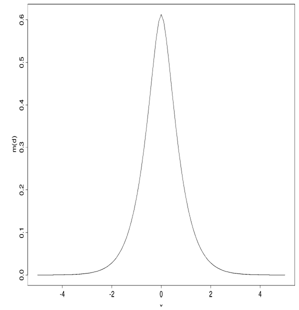
- Adaptive: hyperparameters depend on the level d belongs and could be level-wise different
- Shrinkage: $|\widehat{\theta}| < |d|$
- Hyperparameters chosen by empirical arguments

BAMS: BAYESIAN ADAPTIVE MULTIRESOLUTION SHRINKER

Three goals: reality, simplicity, adaptivity



(A) 10TH (OUT OF 15) DETAIL LEVEL OF A NOISY DOPPLER SIGNAL



(B) MARGINAL $m(d)$, FOR $\tau = .5, \mu = 5$

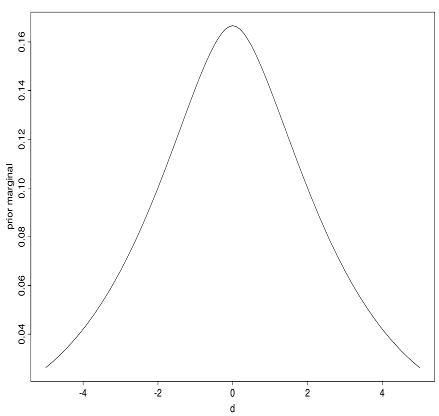
- Empirical distribution of coefficients
- $d|\theta, \sigma^2 \sim N(\theta, \sigma^2)$
- Estimator $\widehat{\sigma}^2$: $d|\theta \sim N(\theta, \widehat{\sigma}^2) \Rightarrow$ “unlikely” $m(d)$
- $\sigma^2 \sim \mathcal{E}(\mu) \Rightarrow d|\theta \sim \mathcal{DE}\left(\theta, 1/\sqrt{2\mu}\right)$
- $\theta \sim \mathcal{DE}(0, \tau) \Rightarrow m(d) = \dfrac{\tau e^{-|d|/\tau} - \frac{1}{\sqrt{2\mu}} e^{-\sqrt{2\mu}|d|}}{2\tau^2 - 1/\mu}$

BAMS: Bayesian Adaptive Multiresolution Shrinker

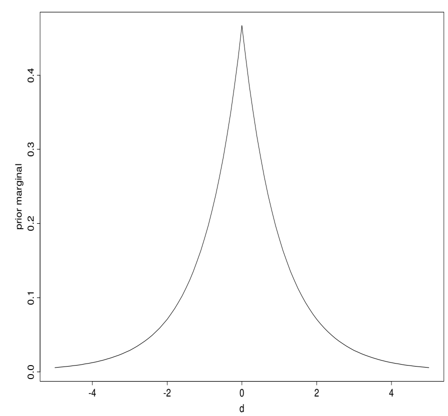
Point mass at 0 $\Rightarrow$ *more shrinkage*

$$\theta|\epsilon \sim \epsilon\delta_0 + (1-\epsilon)\mathcal{DE}(0,\tau)$$

$$\Rightarrow m^*(d) = \epsilon\mathcal{DE}\left(0, 1/\sqrt{2\mu}\right) + (1-\epsilon)m(d)$$



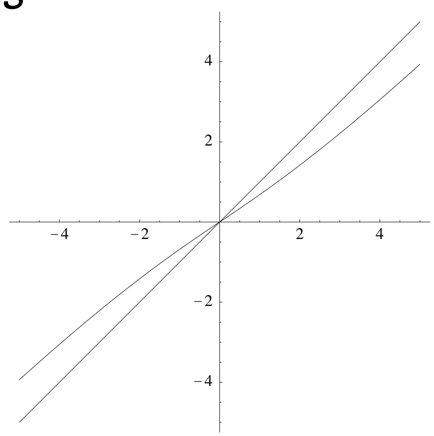
(A) $\epsilon = 0, \tau = 2,\ \mu = 1/2$



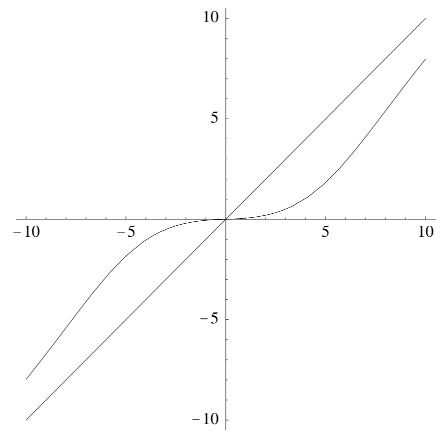
(B) $\epsilon = 0.9, \tau = 2,\ \mu = 1/2$

BAMS: Bayesian Adaptive Multiresolution Shrinker

Bayes rules



(A) $\epsilon = 0, \tau = 2,\ \mu = 1/2$

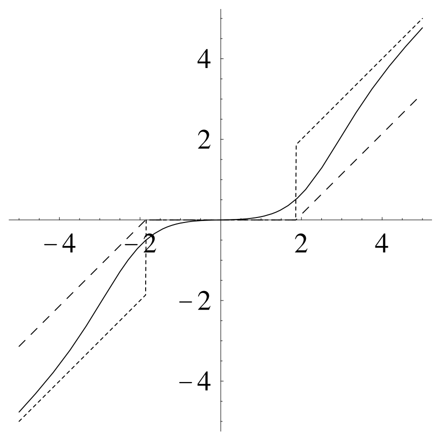


(B) $\epsilon = 0.9, \tau = 2,\ \mu = 1/2$

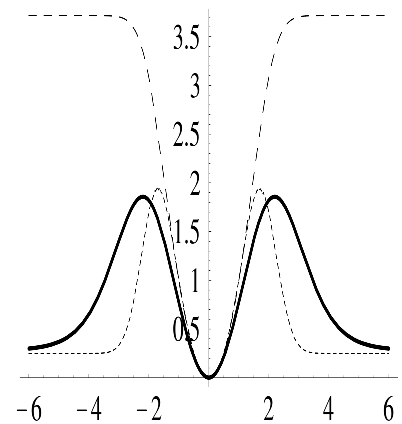
- $$\delta(d) = \frac{\tau(\tau^2 - 1/(2\mu))de^{-|d|/\tau} + \tau^2(e^{-|d|\sqrt{2\mu}} - e^{-|d|/\tau})/\mu}{(\tau^2 - 1/(2\mu))(\tau e^{-|d|/\tau} - (1/\sqrt{2\mu})e^{-|d|\sqrt{2\mu}})}$$

- $$\delta^*(d) = \frac{(1-\epsilon)m(d)\delta(d)}{\epsilon\mathcal{DE}\left(0, 1/\sqrt{2\mu}\right) + (1-\epsilon)m(d)}$$

BAMS: BAYESIAN ADAPTIVE MULTIRESOLUTION SHRINKER



(A) δ^* AND COMPARABLE HARD AND SOFT THRESHOLDING RULES



(B) RISK $R(\theta, \delta) = E^{d|\theta}(\theta - \delta(d))^2$ FOR RULES IN (A)

BAMS: Bayesian Adaptive Multiresolution Shrinker

Simulation: 1000 runs for different sample sizes N from ”usual” noisy test signals:
`Blocks, Bumps, Doppler, Heavisine`

Signals scaled s.t.

- Noise: i.i.d. normal, variance $= 1$

- Signal-to-noise ratio (SNR) $= 7$

  SNR $= \frac{\text{sample std. dev. signal}}{\text{std. dev. noise}}$

Comparison of AMSE of estimator $\widehat{y} = (\widehat{y}_1, \ldots, \widehat{y}_n)$

$AMSE = \frac{1}{MN} \sum_{j=1}^{M} \sum_{i=1}^{N} (\widehat{y}_{ij} - y_i)^2$

BAMS codes in the Antoniadis et al.’s MATLAB toolbox at

`www-mc.imag.fr/SMS/software/GaussianWaveDen/`

BAMS: BAYESIAN ADAPTIVE MULTIRESOLUTION SHRINKER

Tuning the model parameters

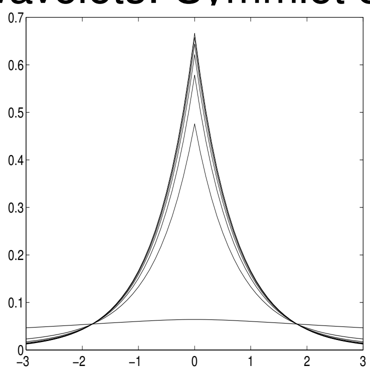
Subjective priors
Default priors
Empirical approach

- μ: $\mu = 1/\sqrt{Var\sigma} = 1/\tilde{\sigma}$
  - $\tilde{\sigma} = |Q_1 - Q_3|/C$ (Tukey (robust) estimator)
  - $1.3 \leq C \leq 1.5$, Q_1 and Q_3 empirical first and third quartiles of the finest level
- ϵ: $\epsilon(j) = 1 - \frac{1}{(j - \texttt{coarsest} + 1)^\gamma}$
  - $\texttt{coarsest} \leq j \leq \log_2 n$
  - $\texttt{coarsest} = 3$ and $\gamma = 1.5$
- τ: $\tau = \sqrt{\max\{(\sigma_d^2 - \frac{1}{\mu}), 0\}}$

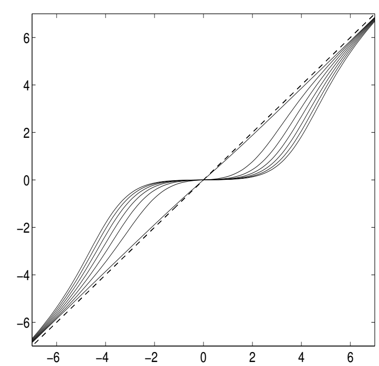
BAMS: Bayesian Adaptive MULTIRESOLUTION SHRINKER

Simulations

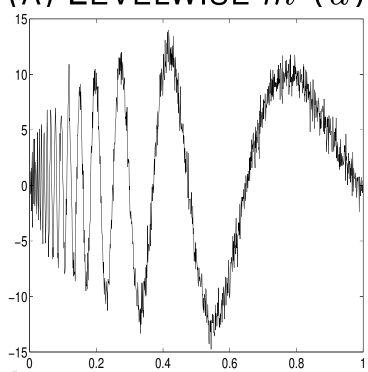
- Doppler signal with $SNR = 7$ and $N = 1024$
- Used wavelets: Symmlet 8



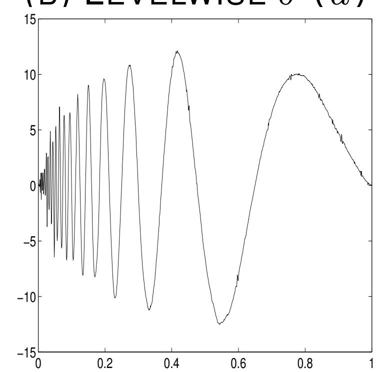
(A) Levelwise $m^*(d)$



(B) Levelwise $\delta^*(d)$



(A) A noisy `doppler` signal



(B) Reconstructed signal

BAMS: BAYESIAN ADAPTIVE MULTIRESOLUTION SHRINKER

Simulations

- $SNR = 7$, sample size $N = 1024$ and $M = 1000$ simulations
- Comparison of MSE of estimator $\widehat{y} = (\widehat{y}_1, \ldots, \widehat{y}_n)$
  $MSE = \frac{1}{MN} \sum_{j=1}^{M} \sum_{i=1}^{N} (\widehat{y}_{ij} - y_i)^2$
- MSE split into Variance + Bias^2
- Code in MATLAB and WAVELAB
- 1000 runs $\rightarrow$ approx. 1 min on AlphaStation 500

	`blocks`	`bumps`
VISUSHRINK	.6840 (.0719 + .6122)	1.5707 (.1165 + 1.4543)
SURESHRINK	.2225 (.1369 + .0856)	0.6827 (.2660 + 0.4167)
ABWS	.0995 (.0874 + .0121)	0.3495 (.2228 + 0.1267)
BAMS	.1107 (.0965 + .0142)	0.3404 (.1976 + 0.1428)
BAMS (10^6)	.1108 (.0962 + .0146)	0.3378 (.1967 + 0.1411)
	`doppler`	`heavisine`
VISUSHRINK	.4850 (.0523 + .4327)	0.1204 (.0339 + 0.0864)
SURESHRINK	.2285 (.0946 + .1340)	0.0949 (.0416 + 0.0534)
ABWS	.1646 (.1006 + .0640)	0.0874 (.0442 + 0.0433)
BAMS	.1482 (.0899 + .0584)	0.0815 (.0511 + 0.0304)
BAMS (10^6)	.1474 (.0883 + .0591)	0.0805 (0.504 + 0.0301)

BAMS: Bayesian Adaptive MULTIRESOLUTION SHRINKER

Simulations

FUNCTION	n	SNR=3	SNR=5	SNR=7	SNR=10
BLOCKS	256	0.3343	0.2835	0.2412	0.2080
	512	0.2101	0.1943	0.1763	0.1567
	1024	0.1583	0.1311	0.1107	0.0942
	2048	0.0921	0.0788	0.0665	0.0549
	4096	0.0560	0.0458	0.0390	0.0343
BUMPS	256	0.6419	0.6996	0.7554	0.8607
	512	0.4834	0.5132	0.5573	0.6093
	1024	0.2969	0.3263	0.3404	0.3508
	2048	0.1823	0.1978	0.2049	0.2144
	4096	0.1009	0.1070	0.1114	0.1176
DOPPLER	256	0.3378	0.3821	0.3887	0.4114
	512	0.1954	0.2131	0.2264	0.2391
	1024	0.1180	0.1350	0.1482	0.1590
	2048	0.0687	0.0783	0.0868	0.0939
	4096	0.0484	0.0497	0.0487	0.0460
HEAVISINE	256	0.1462	0.1754	0.1985	0.2245
	512	0.0957	0.1185	0.1374	0.1584
	1024	0.0607	0.0707	0.0815	0.0958
	2048	0.0402	0.0471	0.0531	0.0611
	4096	0.0332	0.0351	0.0363	0.0382

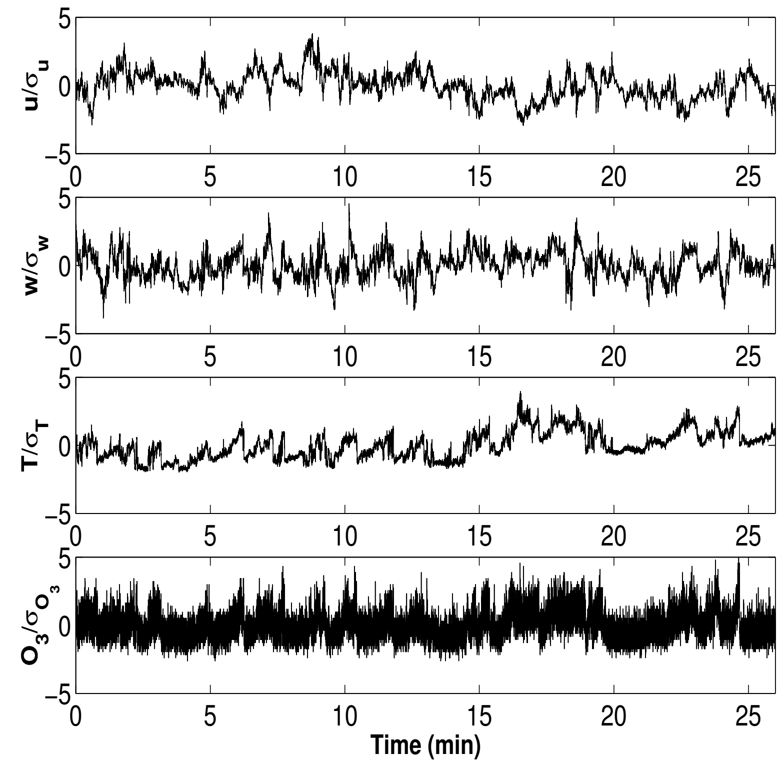
BAMS AND OZONE CONCENTRATION DATA

Data collected at Duke Forest, Durham, NC, USA, 40m from ground, over a 33m tall, 180 year old mixed hardwood forest, over a total period of 60.5 hours, under different atmospheric conditions.

Interest in

- $\mathtt{O}_3$: high frequency ozone concentration measurements
- T: air temperature
- u, v, w: (longitudinal, latitudinal, vertical) turbulent velocity components (we concentrate on first and third)

BAMS AND OZONE CONCENTRATION DATA



Time series of simultaneously measured turbulent velocity components, air temperature, and ozone concentration. For comparison purposes, we normalized all the time series measurements to zero-mean and unit variance.

BAMS AND OZONE CONCENTRATION DATA

Observations: $y_i = f_i + \eta_i$

Haar basis

$\Rightarrow d_i = \theta_i + \varepsilon_i$

$\Rightarrow$ modified BAMS for $d \mid \theta, \sigma^2$:

- self-similar signal
- autoregressive noise, due to the instrument (gas analyser)

X_t *self-similar iff* $X(st) =_d H^s X(t)$

H*: Hurst parameter*

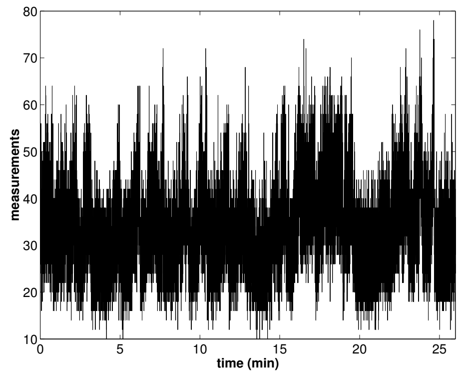
BAMS AND OZONE CONCENTRATION DATA

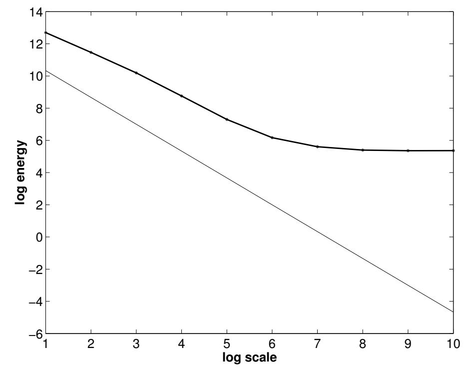
Kolmogorov **K41** scaling law applies to turbulent velocities, temperatures and gas concentration fluctuation

Spectrum of such time series exhibits a power-law decay of $-5/3$

In the wavelet domain a straight line should be obtained when considering the log of average level energies (i.e. $2^{-j} \sum_{k=0}^{2^j-1} d_{j,k}^2$) w.r.t. level j

BAMS AND OZONE CONCENTRATION DATA

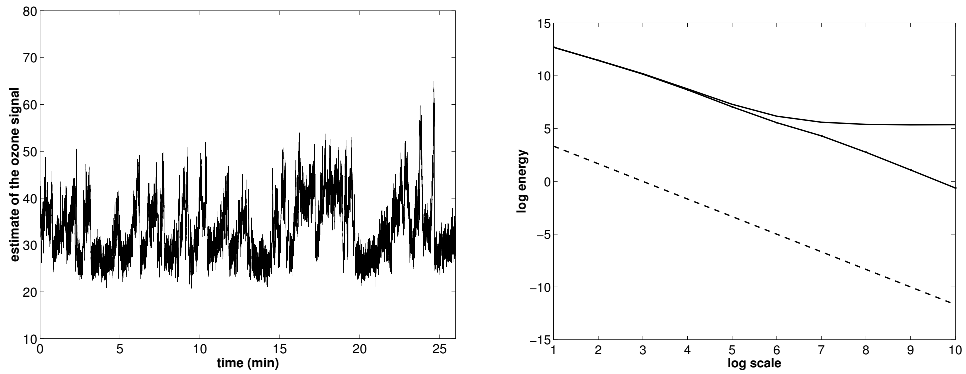




Left: The original (normalized) ozone concentration time series sampled at 21 Hz

Right: The Haar wavelet-spectra of the time series demonstrating that the observations contain noise (flattening of spectral line) at the high resolution levels (or for a large scale index). For reference, the **K41** power-law is also shown

BAMS AND OZONE CONCENTRATION DATA



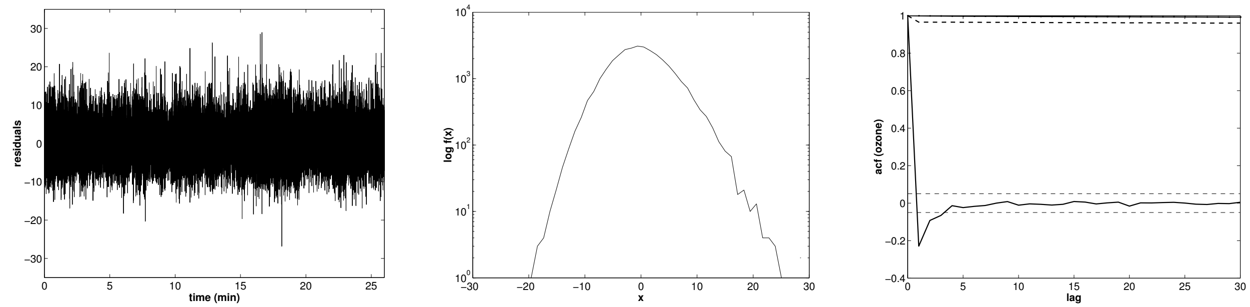
Left: The (normalized) filtered times series (s_i) derived from the noisy time series (y_i)
Right: The Haar wavelet power spectrum of the filtered time series s_i. For reference, we also repeat the Haar wavelet spectrum of y_i from previous Figure. The dotted line is the theoretical **K41** power law.

BAMS AND OZONE CONCENTRATION DATA

Tuning the model parameters

- μ: $\mu = 1/\sqrt{Var\sigma} = 1/\tilde{\sigma}$
  - $\tilde{\sigma} = |Q_1 - Q_3|/C$ (Tukey (robust) estimator)
  - $1.3 \le C \le 1.5$, Q_1 and Q_3 empirical first and third quartiles of the finest level
- ε, weight of point mass at 0 in prior on θ. Level dependent for smooth signals (close to 0 for coarsest levels and to 1 for finest ones). Here $\varepsilon = 0.5$
- τ level dependent as a (non trivial) consequence of **K41** scaling law

BAMS AND OZONE CONCENTRATION DATA



Left: Time series of the (normalized) residuals ($\epsilon_i' = y_i - s_i$); Center: The probability density function (pdf) of ϵ_i' suggesting marginal normality (near-parabola in the log-scale); Right: The autocorrelation function (acf) of ϵ_i' (solid), s_i (dotted), and y_i (dot-solid) as a function of the lag index. For ϵ_i', the 95% confidence levels demonstrate reasonable whiteness. Note that top dot-solid line for y_i hardly decayed after 30 lags (≈ 1.4 s)

BAMS AND OZONE CONCENTRATION DATA

- We split observations into signal and residual and analysed both.
- Analyses of cospectra between w (vertical velocity) and O_3, w and T, w and u (longitudinal velocity) showed that filtered data recovered the theoretical properties.
- Residuals e_k have normal marginals but they are not exactly ”white”. Previous plot shows significant 2-to-3 lag autocorrelation.
- $e_k - 0.2971 * e_{k-1} - 0.2056 * e_{k-2} - 0.1605 * e_{k-3} = Z_k$, with Z_k white noise time series.

LARGER POSTERIOR MODE

Goal: find larger (in absolute value) posterior (local) mode as an estimator of θ

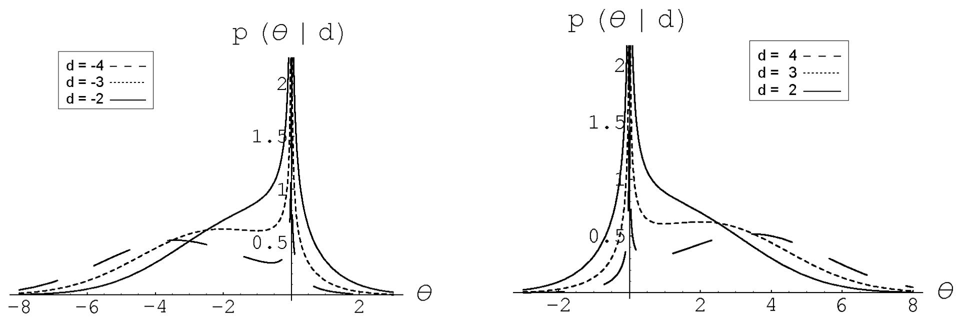
$$\begin{aligned} d|\theta &\sim \mathcal{N}(\theta, \sigma^2) \\ \theta|\tau^2 &\sim \mathcal{N}(0, \tau^2) \\ \tau^2 &\sim (\tau^2)^{-k},\ k > 1/2 \end{aligned}$$

- σ^2 known or estimated
- $p(\theta|d) \propto p(d,\theta) \propto e^{-(d-\theta)^2/(2\sigma^2)}\ |\theta|^{-2k+1}$

  $\Rightarrow \widehat{\theta} = \frac{d + \text{ sign}(d)\ \sqrt{d^2 - 4\sigma^2(2k-1)}}{2}\mathbf{1}(|d| \geq \lambda)$ with $\lambda = 2\sigma\sqrt{2k-1}$
- First order Taylor expansion of square root
  - $\Rightarrow (1-u)^\alpha \approx 1 - \alpha u$
  - $\Rightarrow \widehat{\theta} \approx \left(1 - \frac{\sigma^2(2k-1)}{d^2}\right)_+ d$

    (James-Stein estimator, widely considered in wavelet shrinkage)

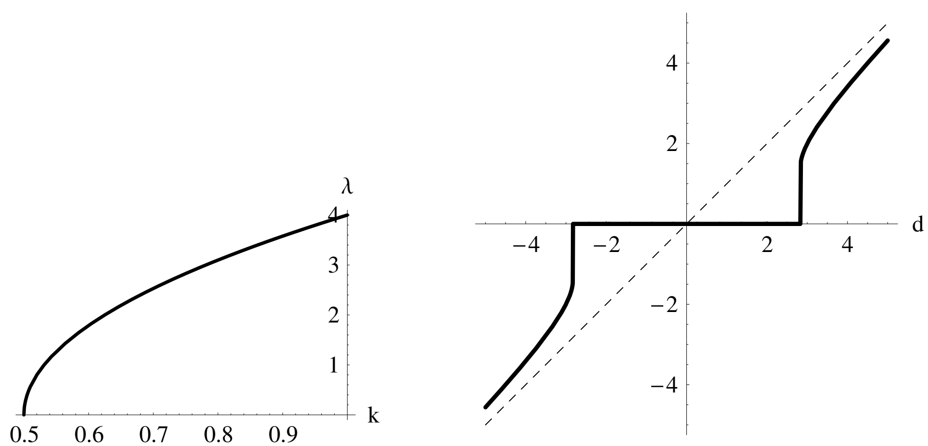
LARGER POSTERIOR MODE



Posterior distribution for $k = 3/4$ and $\sigma^2 = 2^2$; (a) $d = -4, -3, -2$; (b) $d = 2, 3, 4$.

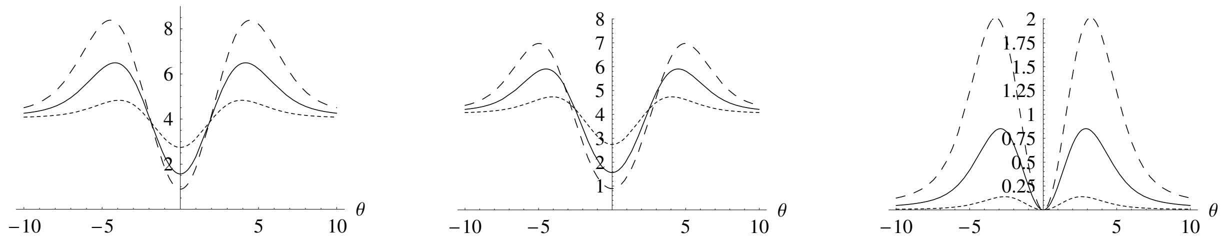
The unimodal density graphs in panels (a) and (b) correspond to $d = -2, 2$, respectively

LARGER POSTERIOR MODE



(a) Influence on the threshold λ by power parameter k; (b) LPM thresholding rule for $k = 3/4$ and $\sigma^2 = 4$.

LARGER POSTERIOR MODE

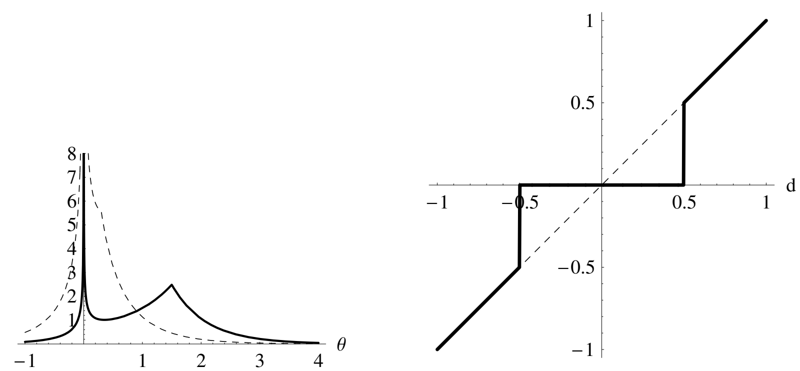


Exact risk plots for LPM rule, for $k = 0.6$ (short dash), $k = 0.75$ (solid), and $k = 0.9$ (long dash). Corresponding thresholds λ are 1.79, 2.83, and 3.58, respectively. For all three cases $\sigma^2 = 2^2$. (a) Risk; (b) Variance, and (c) Bias squared.

- Typical shape of risk for hard thresholding rule
- Risk minimal at 0 and stabilizes for $|\theta|$ large
- Largest value for θ close to threshold λ
  $\Rightarrow$ largest errors due to *keep-or-kill* policy
- Very small bias
- k *resembles* sample size

LARGER POSTERIOR MODE

$\sigma^2 \sim \mathcal{E}(\mu) \Rightarrow \widehat{\theta} = d\,\mathbf{1}(|d| \geq \frac{2k-1}{\sqrt{2\mu}})$



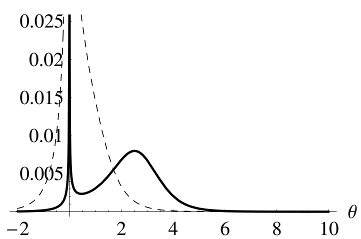
(a) Influence on the posterior in Model 1 by two different values of d. The dashed graph corresponds to $d = 0.3$ while the solid graph corresponds to $d = 1.5$; (b) LPM rule for Model 1

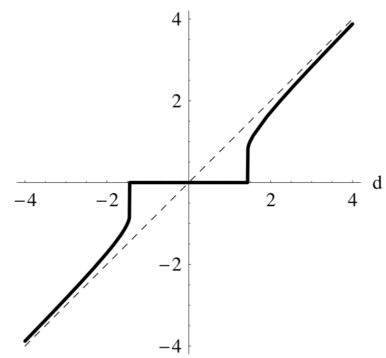
LARGER POSTERIOR MODE

$\sigma^2 \sim \mathcal{IG}(\alpha, \beta)$

$$\hat{\theta} = \frac{(2\alpha+4k-1)d + \mathrm{sign}(d)\sqrt{(2\alpha+1)^2 d^2 + 16(1-2k)(k+\alpha)\beta}}{4(k+\alpha)} \ \mathbf{1}(|d| \geq \lambda)$$

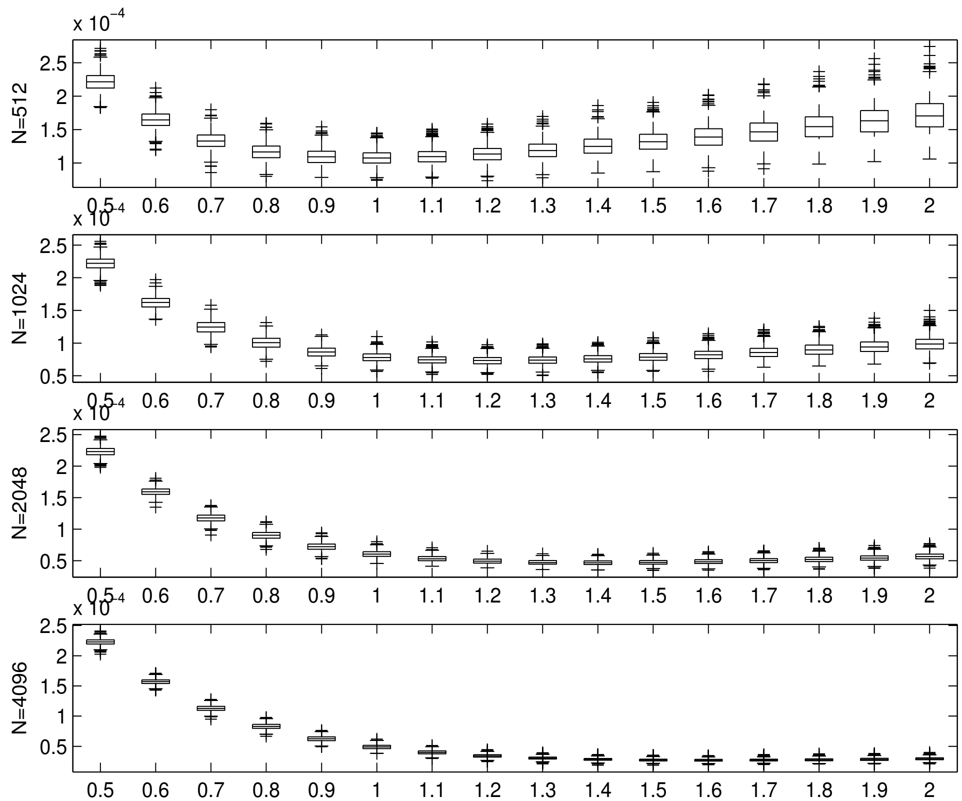
with $\lambda = \frac{2}{2\alpha-1}\sqrt{(2k-1)(k+\alpha)\beta}$





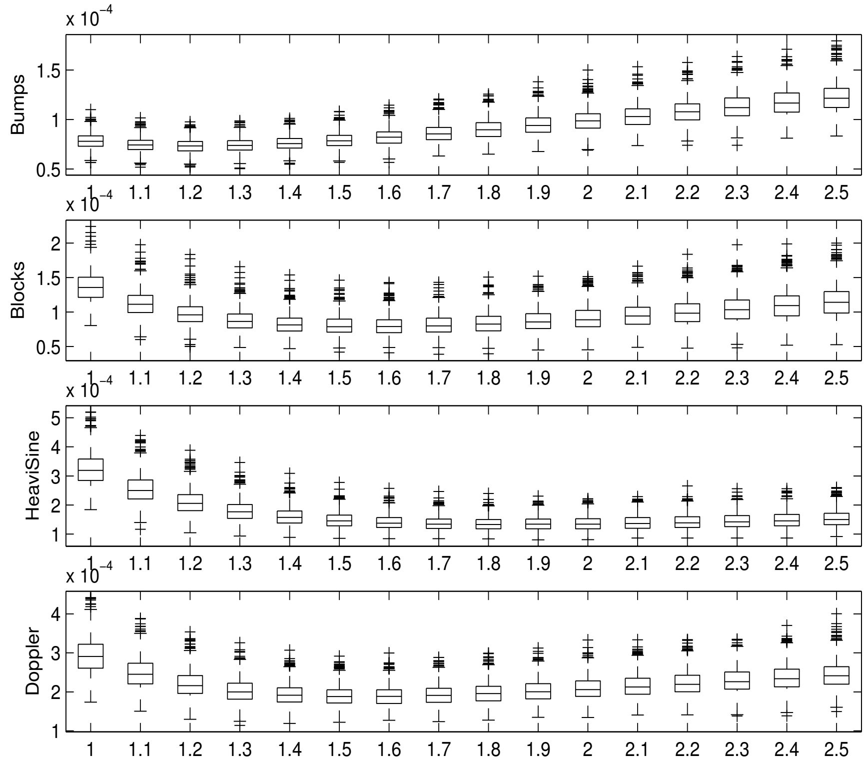
(a) Influence on the posterior in Model 2 by two different values of d. The dashed graph corresponds to $d = 0.7$ while the solid graph corresponds to $d = 2.7$; (b) LPM rule for Model 2

LARGER POSTERIOR MODE



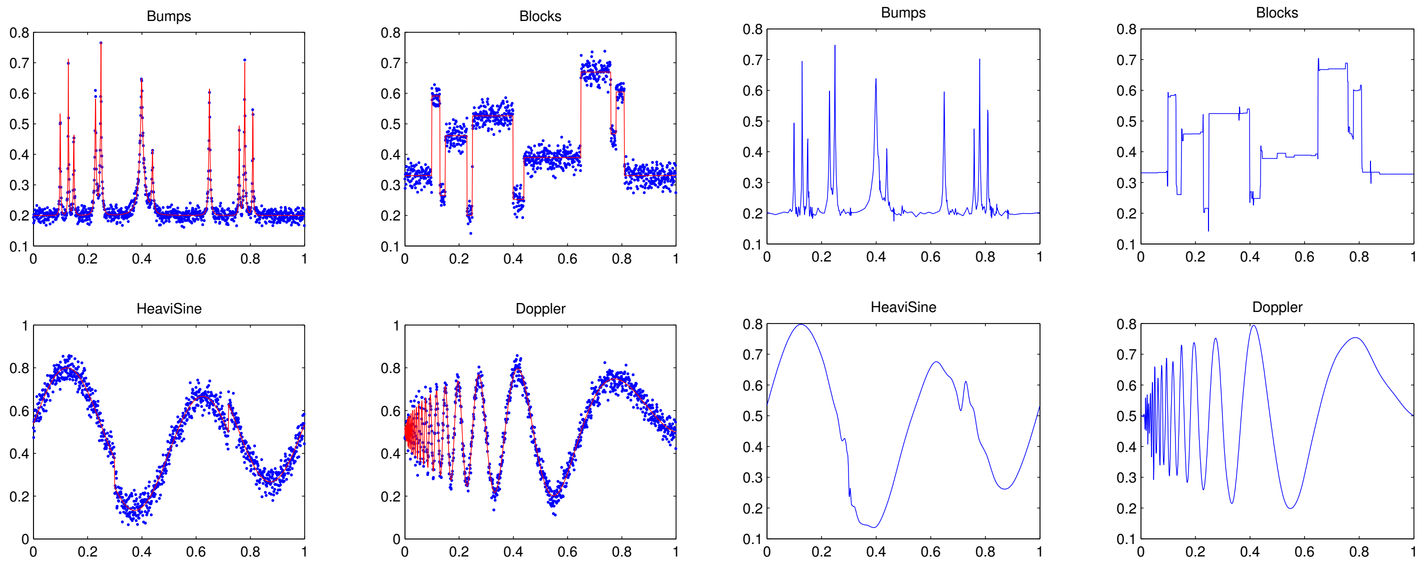
AMSE for Bumps for 4 sample sizes $n = 512$ (top), $1024, 2048, 4096$ (bottom), evaluated at different power parameters k, with SNR=5

LARGER POSTERIOR MODE



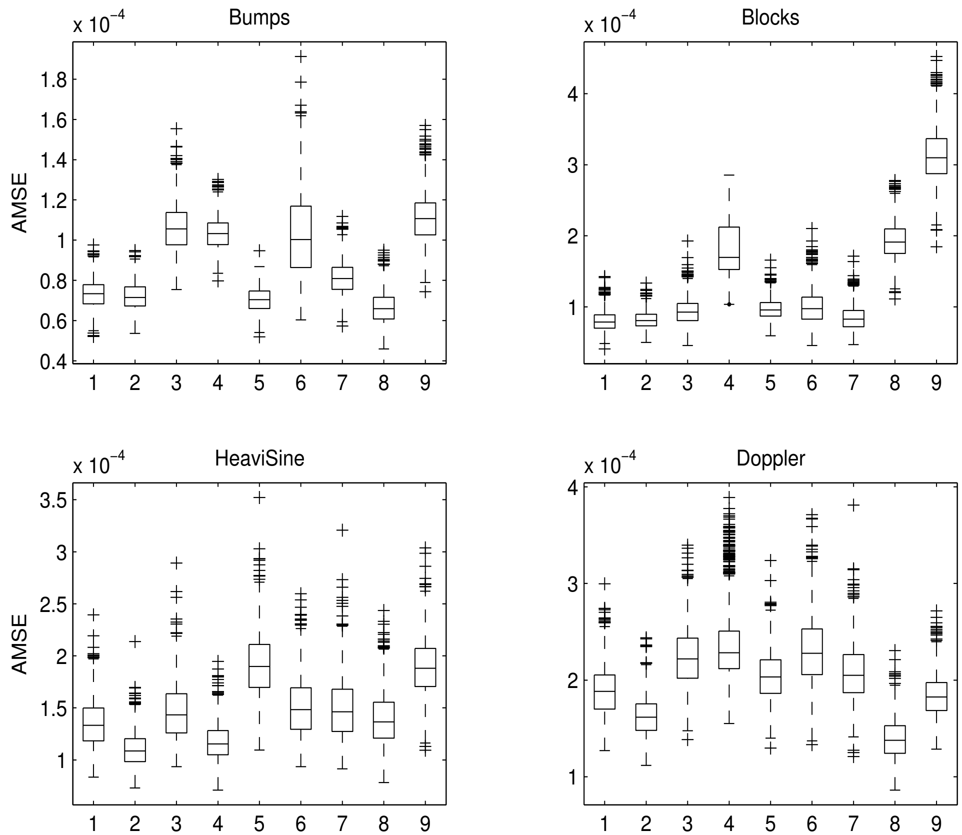
Boxplots of the AMSE for the various values of the power parameter k for four test signals: Bumps, Blocks, HeaviSine, and Doppler. Sample size was $n = 1024$ and SNR = 5

LARGER POSTERIOR MODE



(a) Test signals with superimposed noisy versions at SNR=5. The sample size is $n = 1024$. (b) Estimates obtained by using the LPM method with optimal k

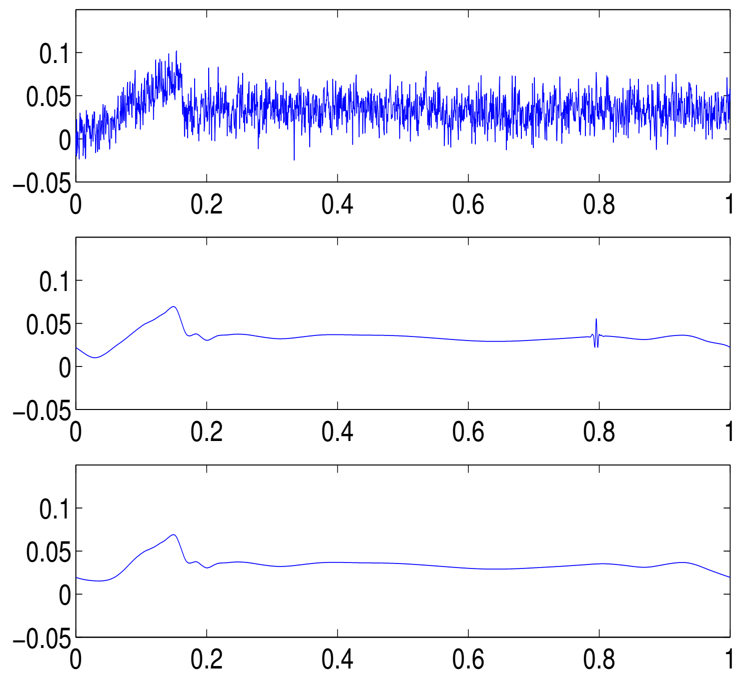
LARGER POSTERIOR MODE



Boxplots of the AMSE for the various methods (1) LPM, (2) BAMS, (3) VisuShrink, (4) Hybrid, (5) ABE, (6) CV, (7) FDR, (8) NC, (9) BJS, based on $n = 1024$ points at SNR=5

LARGER POSTERIOR MODE

AFM (atomic force microscopy) can measure the adhesion strength between two materials at the nanonewton scale. Noisy data collected at Georgia Tech.



Original AFM measurements (top), LPM estimator with the default parameter $k = 1$ (middle), LMP estimator with the parameter $k = 1.4$ (bottom)