Bayesian Denoising in Oriented and Non-Oriented Multi-Scale Pyramids

Jalal M. FADILI and Larbi BOUBCHIR

Image Processing group

GREYC UMR UMR 6072, Caen France

Outline

- Introduction.
  - Observation model.
  - Bayesian paradigm.
- Ingredients from modern harmonic analysis: beyond wavelets.
- Wavelet-like domain denoising: an overview.
- Statistical priors:
  - Univariate (marginal).
  - Multivariate (joint).
- Bayesian estimation:
  - Univariate.
  - Multivariate.
- Conclusion and extensions.

Observation model

- The image is viewed as realization(s) of a RV or a random fields whose degradation equation is:

$$Y_s = \mathcal{M}\left[\Phi((\mathcal{B}X)_s) \odot \epsilon_s\right] \tag{1}$$

where:

  - $\odot$ is any composition of two arguments (e.g. '+', '$\cdot$').
  - $s \in S$ is the location index.
  - ϵ_s is the noise (random) (generally assumed AWGN but not necessarily so, e.g. speckle, Poisson, $\frac{1}{f}$).
  - $\mathcal{B}$ is a (possibly non-linear) degradation operator (e.g. convolution with a PSF).
  - Φ is a transformation not necessarily linear nor invertible (e.g. sensor-specific, etc).
  - $\mathcal{M}$ missing data mechanism.

Restoration problem: how to estimate unobserved X from observed Y

An inverse ill-posed problem

Observation model

- The image is viewed as realization(s) of a RV or a random fields whose degradation equation is:

$$Y_s = \mathcal{M}\left[\Phi((\mathcal{B}X)_s) \odot \epsilon_s\right] \tag{1}$$

where:

  - $\odot$ is any composition of two arguments (e.g. '+', '$\cdot$').
  - $s \in S$ is the location index.
  - ϵ_s is the noise (random) (generally assumed AWGN but not necessarily so, e.g. speckle, Poisson, $\frac{1}{f}$).
  - $\mathcal{B}$ is a (possibly non-linear) degradation operator (e.g. convolution with a PSF).
  - Φ is a transformation not necessarily linear nor invertible (e.g. sensor-specific, etc).
  - $\mathcal{M}$ missing data mechanism.

Restoration problem: how to estimate unobserved X from observed Y

An inverse ill-posed problem

Bayesian paradigm

$$Y = \Phi((B X)) \odot \epsilon$$

- $p(x, \mathbf{z})$: *prior* distribution. $\mathbf{z}$ some other image features (e.g. local regularity, texture, etc).
- $p(y|x, \mathbf{z})$: likelihood (given x and $\mathbf{z}$). ($p(\epsilon)$).
- $p(y)$: marginal distribution $= \int p(y|x, \mathbf{z})p(x, \mathbf{z})dxd\mathbf{z}$.
- $p(x, \mathbf{z}|y)$, *posterior* distribution: $\frac{p(y|x,\mathbf{z})p(x,\mathbf{z})}{p(y)}$

Bayesian estimation amounts to finding the operator $\mathcal{D}$ s.t.:

$$\hat{x} = \arg\inf_{\mathcal{D} \in \mathcal{O}_n} R\left(x, \hat{x} = \mathcal{D}y\right) = \mathsf{E}_{Y,X}\left[L(x, \mathcal{D}y)\right] \tag{2}$$

What estimator for what risk ?

Cost $L(x, \hat{x})$	Estimator	$\hat{x}$
0-1	MAP	$\arg\max_{x \in \mathcal{X}} p_{X\|Y}(x\|y)$
L_2	MMSE	$\mathsf{E}[X\|Y]$
L_1	MMAE	$\Pr(X > \hat{x}\|Y = y) = \frac{1}{2}$
$\mathrm{Card}\{s \in S : \hat{x}_s \neq x_s\}$	MPM	MAP at each site.

- MAP involves solving an optimization problem.
- MMSE involves solving an integration problem.
- For mutually independent iid gaussian signal and noise, MAP, MMSE and Wiener are the same.

What prior ?

Image corpus based models

- Existence of a probability space on some particular corpus of images (e.g. natural images) [Olshausen and Field 96, Zhu and Mumford 96, Gousseau 00, Gousseau-Morel-Alvarez 99, Mumford and Huang 99].
- Only general properties (e.g. scale invariance, power-law) are accessible but no explicit distribution.

Transported Generator Models

$$X(u,v) = \sum_i a_i g_i(s_i u + t_{i_u}, s_i v + t_{i_v})$$

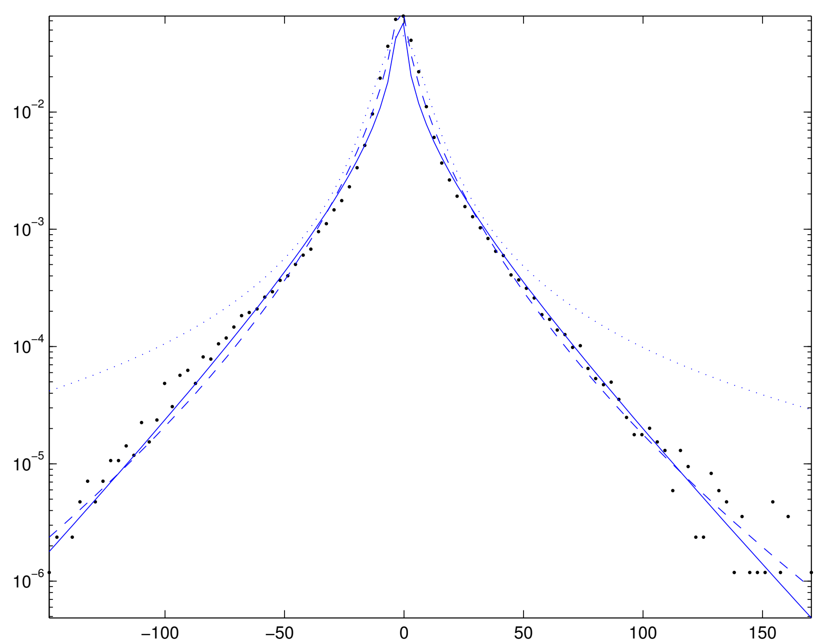
- g_i are the random object templates.
- Clearly, this states that random objects are randomly placed (according to some probability law) while imposing some axioms such as scale invariance (multi-scale nature of images !).
- [Gidas et Mumford 01] showed (invariance principle and the Lévy-Khintchine theorem) that (s_i, t_{i_u}, t_{i_v}) has a Poisson law in affine transformation group with a density $dsdt_u dt_v / s$.
- [Grenander et al. 99-03] imposed $ai \sim \mathcal{N}(0,1)$ and a Gamma prior on $\sum_i g_i^2(s_i u + t_{i_u}, s_i v + t_{i_v})$ (an instance of the scale-mixture of gaussians SMG).

Random Field Theory (e.g. MRF)

What prior ? (cont'd)

Sparse representation prior

- Choose a Hilbert space equipped with a basis or frame (Fourier, wavelets, X-lets, etc) [Mallat 89, Simoncelli et al. 98, Wainwright et al. 00, Grenander et al. 01, Fadili and Boubchir 03].
- Project the original data in that space where all components (coefficients) but a few are zero (notion of parsimony or sparsity).
- Expected statistical behavior: unimodal, centered a zero, non-gaussian sharply peaked distribution with heavy tails.
- Many images encountered in practice have a sparse gradient and then fall within this intuitive model.



Wavelet coefficients model. Dark points: observed pdf, dashed: GGD [Mallat 89], dotted: α-stable [Lévy 24, Achim et al. 01], solid: Bessel K form [Grenander et al. 01, Fadili et al. 03].

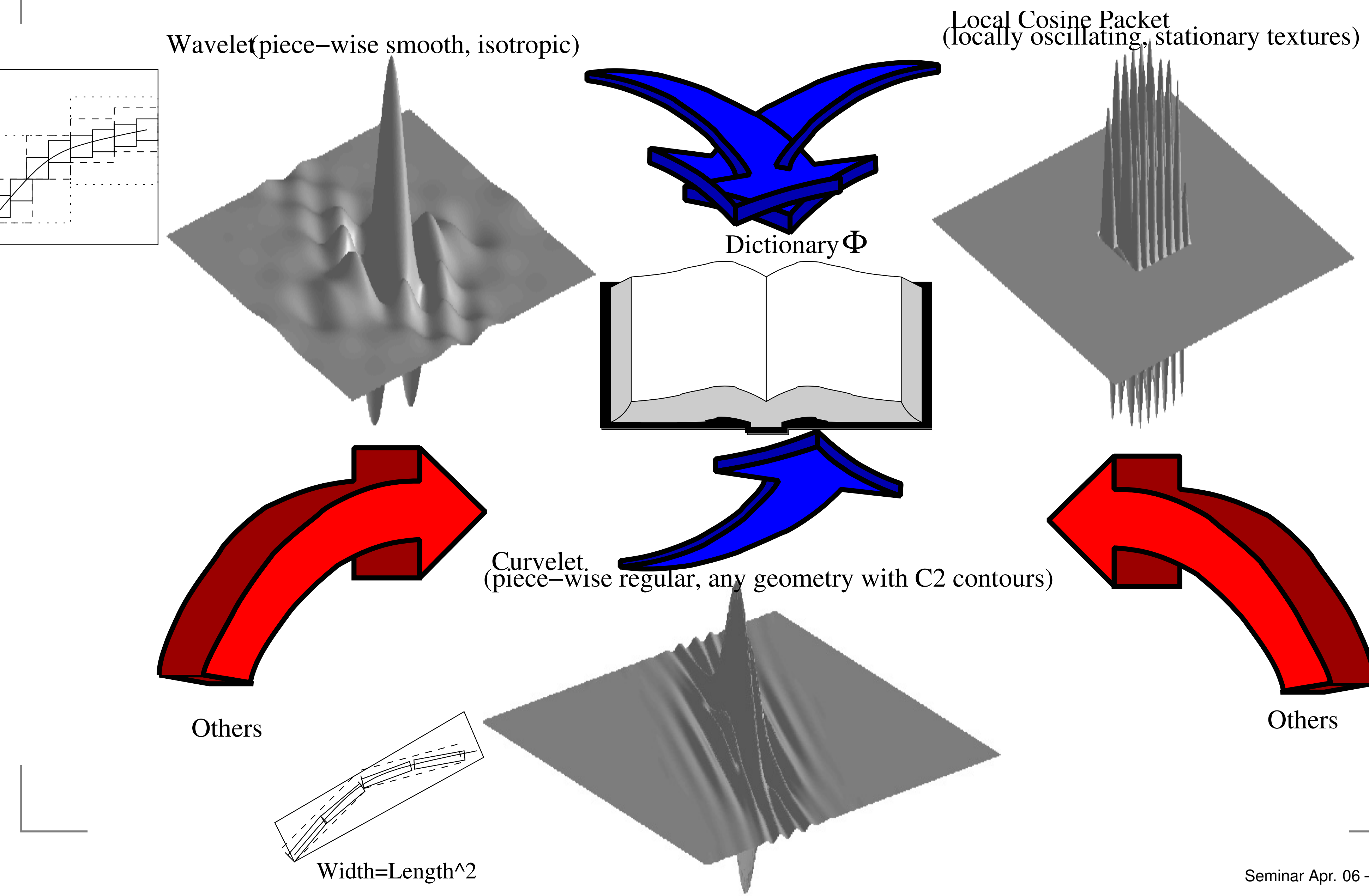
Elements from modern harmonic analysis

- Ingredients: $X \in \mathcal{H}$ a Hilbert space.
- An $\sqrt{n} \times \sqrt{n}$ image X can be written as the superposition of elementary functions $\varphi_\gamma(u, v)$ (atoms) parameterized by γ s.t. (Γ is denumerable):

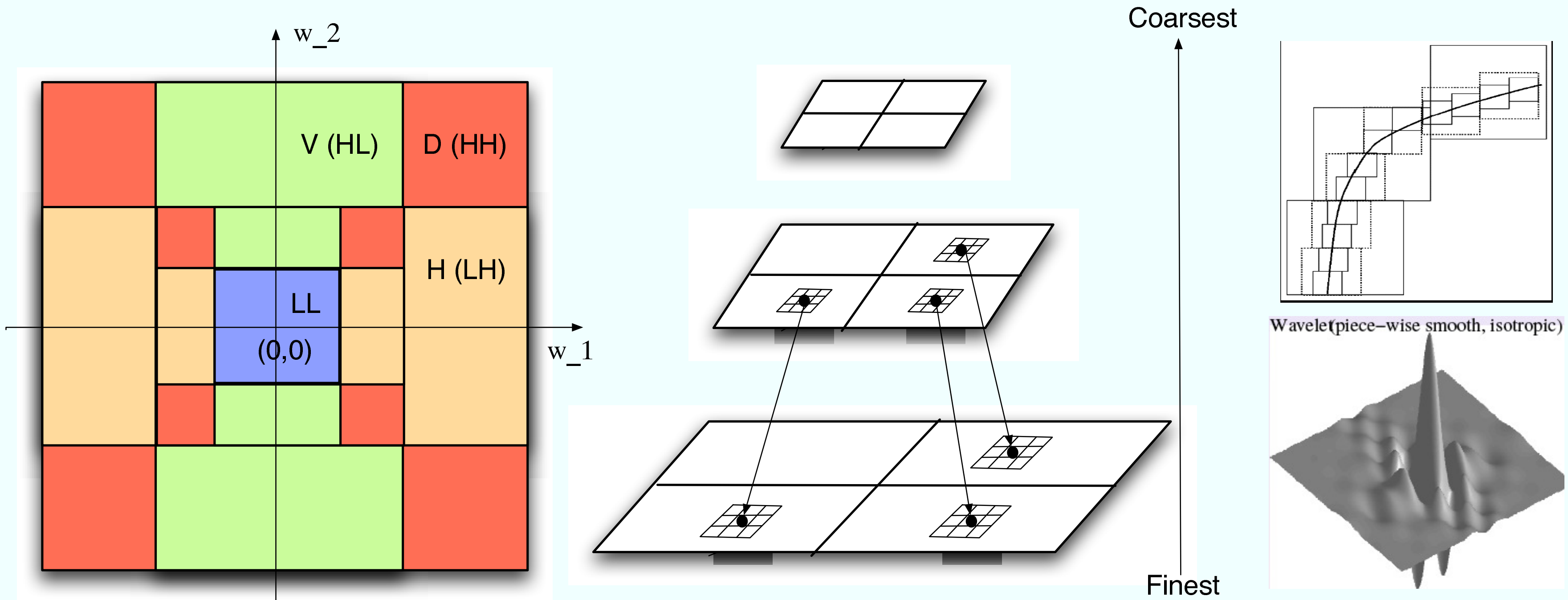
$$X(u, v) = \sum_{\gamma \in \Gamma} d_\gamma \varphi_\gamma(u, v), \ \varphi_\gamma \in \mathcal{L}$$

- The atoms $\{\varphi_l\}_{l=1,\ldots,L}$ are normalized to a unit norm.
- The forward transform is defined by $\Phi = [\varphi_1 \ldots \varphi_L] \in \mathbb{R}^{N \times L}, L \geq N$ (defines a basis, a frame or a tight frame).
- Examples of Γ: frequency (Fourier), scale-translation (wavelets), scale-translation-frequency (wavelet packets), translation-duration-frequency (cosine packets), scale-translation-angle (geometrical X-lets, curvelets, bandlets, contourlets, wedgelets, etc).

What those atoms look like: morphological diversity



Non-oriented pyramids: WT

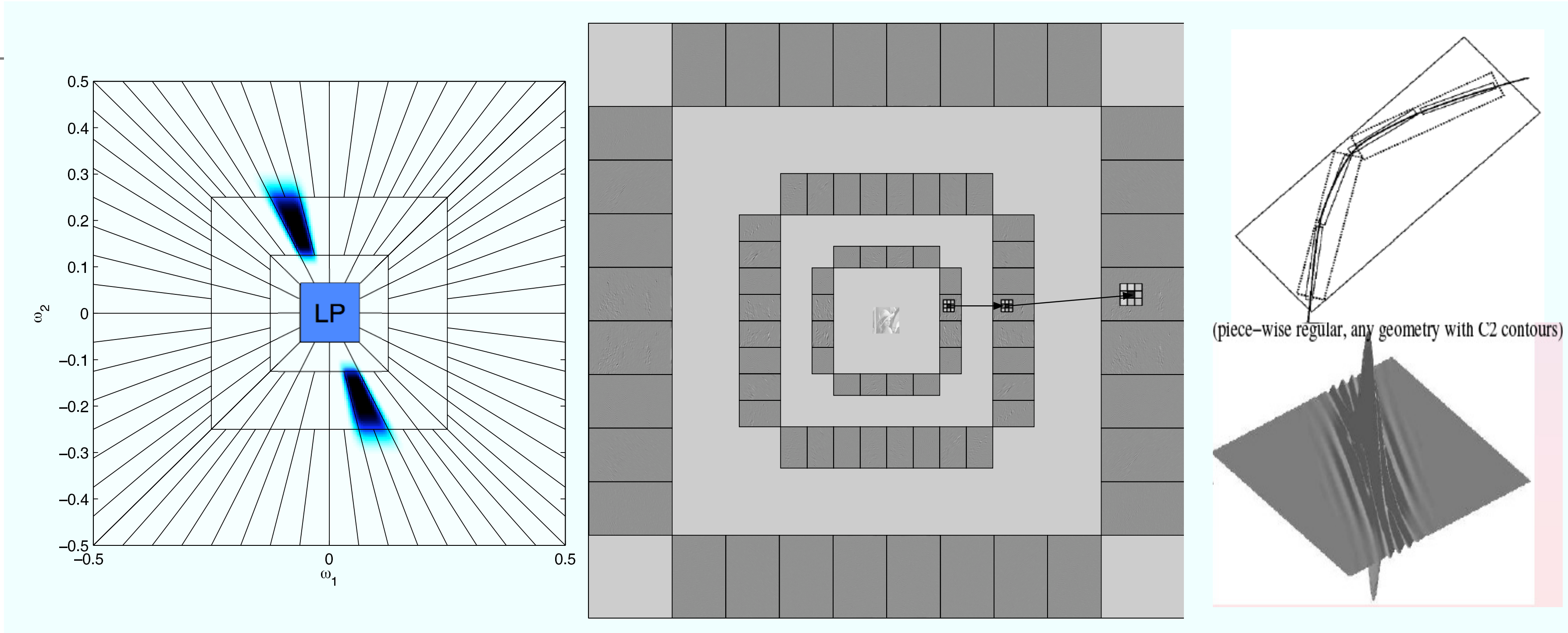


$$y(t) = \sum_{k=0}^{2^{J_c}-1} c_{J_c,k}\phi_{J_c,k}(t) + \sum_{j=J_c}^{\infty}\sum_{k=0}^{2^j-1} d_{j,k}\psi_{j,k}(t) \text{ where } J_c \geq 0, y \in L_2(\mathbb{R})$$

$$c_{J_c,k} =< y, \phi_{j_0,k} > and\ d_{j,k} =< y, \psi_{j,k} >$$

- ϕ and ψ are defined by a (FIR) filter bank (h, g).
- h and g are in most cases power-complementary.
- $\mathbf{d} = \mathcal{W}\mathbf{y}$: corresponds to a (tight) frame expansion (oversampled) or a basis (critically sampled).
- For piece-wise smooth images: $\|f - \hat{f}_M\|^2_{L_2} \leq CM^{-1}$.

Oriented pyramids: FDCT



- 2nd generation curvelets [E. Candès, D. Donoho and L. Demanet].
- Curvelet atoms are formed by translating, scaling and rotating a single mother function.
- Its Fourier transform is defined as (polar coordinates):

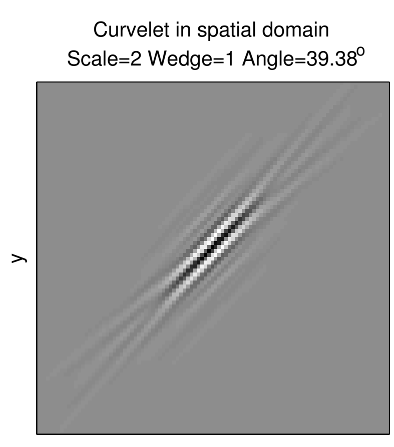
$$\hat{\varphi}_j(r,\theta) = 2^{3j/4} W(2^j r) V\left(\frac{\lfloor 2^{j/2} \rfloor \theta}{2\pi}\right)$$

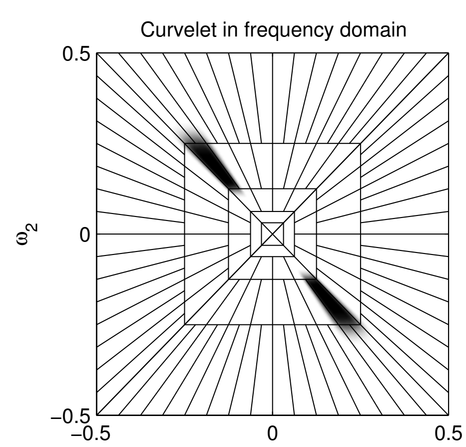
- The support of $|\hat{\varphi}_j|$ is a polar wedge defined by those of W (radial window) and V (angular window).

Oriented pyramids: FDCT (cont'd)

Property 1

- *The curvelet transform provides a multiresolution, directional representation with basis elements well localized in both space and frequency.*
- *Parabolic scaling law: atoms are highly anisotropic with* $(2^{-j}) \approx width \approx length^2 (2^{-j/2})$.
- *Oscillating behavior:* φ_j *is a little needle whose envelope is a specified "ridge" of effective length* $2^{-j/2}$ *and width* 2^{-j}*, and which displays an oscillatory behavior across the main "ridge".*
- *Tight frame.*
- *Optimally sparse representation of piecewise smooth images with* C^2 *edges:* $\|f - \hat{f}_M\|_{L_2}^2 \leq C\,(\log M)^3\, M^{-2}$.





Sparse representation-based denoising

- Seminal work of Dave Donoho and Ian Johnstone in early 90's on wavelet non-parametric regression with AWGN.
- Argument from functional approximation: owing to the sparsity of the wavelet expansion, signal is essentially concentrated on a small fraction of coefficients while most others are due to noise, which contaminates uniformly all coefficients.
- Basic idea: apply a non-expansive operator (e.g. threshold or shrink) to the representation coefficients.
- Since then, literature have been inundated by modifications or extensions of this original idea.

$$\mathbf{y} \xrightarrow{\mathcal{W}} \{c_{mn}, d_\gamma\} \xrightarrow{\text{Non-linear estimator } \mathcal{D}_\theta} \{c_{mn}, \mathcal{D}_\theta(\hat{d}_\gamma)\} \xrightarrow{\mathcal{R}} \hat{\mathbf{x}}$$

A brief overview

Classical term-by-term

Minimax estimation, SureShrink, etc [Donoho et al. 92-95]. Modifications on Donoho's shrinkage operators [Bruce and Gao, Antoniadis and Fan]. Translation invariant threshold [Coifman and Donoho 95]. Hypothesis testing [Abramovich and Benjamini 95-96, Ogden and Parzen 96]. Cross-validation [Green et Silverman 94, Eubank 99].

Classical block

Non-overlapping block thresholding [Cai 99]. Overlapping block thresholding [Cai et Silverman 00].

Bayesian term-by-term (univariate)

Bernoulli-Gaussian FM [Abramovich et al. 98, Clyde and George 99,00]. Bayesian hypothesis testing [Vidakovic et al. 98]. SMG with exponential multiplier prior [Vidakovic et al. 00]. Two Gaussians FM [Chipman et al. 97]. t-Student prior [Vidakovic 98]. GGD [Mallat99, Liu et Moulin 99]. Adaptive variance gaussian prior [Simoncelli 99]. α-stable [Achim et al. 01].

Bayesian block (multivariate)

Non-overlapping block bayesian estimation [Abramovich et Sapatinas 00]. Multivariate gaussian prior [Huang and Cressie 00]. Mixed effects models [Huang and Lu 00]. MRF [Malfait et al. 97, Crouse et al. 98, Pizurica et al. 02]. HMT model [DSP Rice (Romberg, Baraniuk et al. 00-02)]. Scale mixture of gaussians [Li and Orchard 00, Mihchak et al. 99, Portilla et al. 03]. Multi-variate α-stable [Koruglu et Achim 04].

A comprehensive comparative study in [Antoniadis, Bigot and Sapatinas 01].

Univariate Scale Mixture of Gaussians family

Definition 1 (Andrews and Mallows 74) *Let X be a RV with real-valued realizations. Under the SMG, there exist two independent RVs $U \geq 0$ and $Z \sim \mathcal{N}(0,1)$ such that:*

$$X \stackrel{d}{=} Z\sqrt{U} \tag{3}$$

Property 2

- *SMG is a subset of the elliptically symmetric distributions [Kotz et al. 89]*
- *$f_X(0)$ exists iif $\mathbb{E}[U^{-1/2}] < +\infty$.*
- *The pdf of X is:*

$$f_X(x) = \frac{1}{\sqrt{2\pi}} \int_0^{+\infty} u^{-1/2} e^{-\frac{x^2}{2u}} f_U(u) \tag{4}$$

  It is unimodal, symmetric around the mode and differentiable almost everywhere (except perhaps at 0).
- *The characteristic function (CF) of X is:*

$$\Phi_X(\omega) = \mathcal{L}[f_U]\left(\frac{\omega^2}{2}\right) \tag{5}$$

  $\mathcal{L}$ is the Laplace transform.
- *The pdf of U is closely related to the inverse Laplace transform of f_X.*

USMG family

The following proposition establishes necessary and sufficient conditions for such a representation to exist:

Proposition 1 (Andrews and Mallows 74) *The RV X has a SMG representation iff the k^{th} derivatives of $f_X(\sqrt{y})$ have alternating sign, i.e.:*

$$\left(-\frac{d}{dy}\right)^k f_X(\sqrt{y}) \geq 0 \;\forall y > 0 \tag{6}$$

Lemma 1 *If $X \stackrel{d}{=} Z\sqrt{U}$ with random $U \geq 0$ and $Z \sim \mathcal{N}(0, \sigma^2)$, then kurtosis$(X) > 0 \Longrightarrow$ the symmetric distribution of X is necessarily sharply peaked (leptokurtic) with heavy tails.*

- As stated before, in many sparse representations, empirical coefficient pdfs are symmetric around 0, leptokurtic and heavy tailed.
- Moreover, these pdfs have their 1st and 2nd derivatives of alternating signs on $\mathbb{R}^+$.
- The SMG family satisfies all these requirements (above results).
- Consequence: this family is well adapted to capture the sparsity of decompositions and is then legitimate as a prior for the coefficients.
- A key advantage of SMG is that it transfers desirable properties of the gaussian distribution through the mixing RV.

Back to wavelets: USMG and Besov space

Is the SMG family well adapted as a prior for wavelet coefficients of highly irregular functions) ?

- Besov spaces are very general tool in describing the smoothness properties of functions, e.g. piece-wise smooth or with isolated singularities.
- Instead of the modulus of continuity based definition, we use a practical characterization of Besov space norm with the wavelet coefficients.

Theorem 1 (Meyer 92) *Let $g = \sum_{j,k} d_{j,k}\psi_{j,k}$ where $d_{j,k}$ are the wavelet coefficients and ψ is a wavelet with sufficient number of vanishing moments. The Besov norm for the function $g \in B^s_{p,q}$ is related to a sequence space norm on its wavelet coefficients and is given by:*

$$\|x\|_{\mathcal{B}^s_{p,q}} = \begin{cases} |c_{0,0}| + \left[\sum_{j=0}^{\infty} 2^{js'q}\left(\sum_{k=0}^{2^j-1}|d_{j,k}|^p\right)^{\frac{q}{p}}\right]^{\frac{1}{q}} & \text{if } 1 \le q < \infty \\ |c_{0,0}| + \sup\limits_{j\ge 0}\left[2^{js'}\left(\sum_{k=0}^{2^j-1}|d_{j,k}|^p\right)^{\frac{1}{p}}\right] & \text{if } q = \infty \end{cases}$$

for $1 \le p < \infty$ where $s' = (s + \frac{1}{2} - \frac{1}{p})$. s can be viewed as a regularity parameter of the image g.

Back to wavelets: USMG and Besov space (cont'd)

- The Besov space norm can be related to the prior distribution of the wavelet coefficients at each detail scale.
- An explicit relationship between the parameters of the SMG prior model and the Besov space.

Theorem 2 *Let* $X_{j,k} \stackrel{d}{=} Z_j \sqrt{U}$ *iid RVs at each scale such that* $Z \sim \mathcal{N}(0, \sigma_j^2)$*,* $\mathrm{E}\left[U\right] = 1$ *and* $M_U(p) < +\infty \;\; 1 \leq p < +\infty$ *and* $\sigma_j = \sigma_0 2^{-j\beta}$ *(the scale invariance property of images), with* $(0 < \sigma_0 < +\infty, \beta \geq 0)$*. Then, for a fixed* $c_{0,0}$*,*

$$g \in B^s_{p,q} \text{ almost surely if and only if } \beta > (s + \frac{1}{2}), \text{ for } 1 \leq p < \infty \text{ and } 1 \leq q \leq \infty.$$

USMG: hyperparameters estimation

- Generally, distribution of U depends on some hyperparameters θ.
- They are estimated directly from the coefficients at each subband:
  - MLE.
  - Quantile methods.
  - Characteristic function methods.
  - Cumulants (easily extended in presence of AWGN).
  - EM.
- This step is crucial for the final performance of the denoiser.
- Is somewhat easy in noiseless case and becomes much more complex with corrupting noise.

Case (1): Bessel K form

- $X = \sqrt{U}Z$ where $U \sim$ Gamma(α, c)
- Raised in the TGM [Grenander et al. 01].

Definition 2 *The BKF pdf is:*

$$f(x;\alpha,c) = \frac{1}{\sqrt{\pi}\Gamma(\alpha)} \left(\frac{c}{2}\right)^{-\frac{\alpha}{2}-\frac{1}{4}} \left|\frac{x}{2}\right|^{\alpha-\frac{1}{2}} K_{\alpha-\frac{1}{2}}\left(\sqrt{\frac{2}{c}}|x|\right)$$

K_ν is the modified Bessel function. $\alpha > 0$ and $c > 0$ are the shape and scale parameters.

Property 3

- *The BKF distribution tends to a gaussian as $\alpha \longrightarrow \infty$.*
- *Let $X \sim BKF(\alpha, c)$,*

$$\kappa_{2i} = \alpha \left(\frac{c}{2}\right)^i \frac{(2i)!}{i}, i \geq 1 \qquad (7)$$

Bessel K form: estimate α and c

- Using cumulants:

$$\hat{\alpha} = \frac{3}{\mathrm{Kurt}(X) - 3}, \hat{c} = \frac{\mathrm{Var}(X)}{\hat{\alpha}}$$

- Using EM algorithm: consider the multiplier as the missing data.
  - The sufficient statistics are calculated in closed form in the E-step.
  - Proof of convergence (to a local minimum).
  - cumulant-based estimator is used as an initialization.

Case (2): α-stable

- $X = \sqrt{U}Z$ where $U \sim \mathcal{S}_{\alpha/2}(-1, \sigma = (\cos(\frac{\pi\alpha}{4}))^{\frac{2}{\alpha}}))$ and $Z \sim \mathcal{N}(0, 2\gamma)$

Definition 3 *The α-stable pdf is given (Zolotarev formulation) [Levy 24]:*

$$f(x;\alpha,\beta,\sigma) = \begin{cases} \frac{1}{\pi}\int_0^\infty \exp(-\sigma^\alpha x^\alpha)\cos[x^2 + \beta x^\alpha \tan(\frac{\alpha\pi}{2})]dx \\ \text{if } \alpha \neq 1 \\ \frac{1}{\pi}\int_0^\infty \exp(-\sigma^\alpha x^\alpha)\cos[x^2 + \beta x^\alpha \frac{2}{\pi}\log|x|]dx \\ \text{if } \alpha = 1 \end{cases}$$

- Characteristic exponent α, $0 < \alpha \leq 2$ (leptokurticity degree).
- Asymmetry parameter β (symmetric for $\beta = 0$),
- Scale parameter $\sigma = \gamma^{\frac{1}{\alpha}}$.
- Tails have an algebraic decay rate.

α-stable marginal pdf

- The α-stable exists pdf exists and is continuous, but not analytical form except very special cases.
  - Gaussian $\mathcal{S}_2(0, \sigma)$
  - Cauchy $\mathcal{S}_1(0, \sigma)$
  - Lévy $\mathcal{S}_{1/2}(1, \sigma)$
- Inverse FT necessitate intensive numerical integration with infinite bounds.
- Exact approach [Nolan 97]
  - finite integration bounds.
  - numerically stable, but ...
  - computationally intensive (e.g. for images).
  - hyperparameter estimation is an issue (especially in the noisy case).
  - no analytical form for the bayesian denoiser.
- Our approach:
  - use the SMG property to provide an finite mixture approximation with a controlled accuracy.
  - fast and numerically stable.
  - By-products: analytical form for the bayesian denoiser, good estimator of the hyperparameters.

α-stable marginal pdf

Require: Number of components N_{mixt}.

1: Estimate α and σ (e.g. using a quantile estimator).

2: Calculate the CF of the f_U with $(\frac{\alpha}{2}, \beta = -1, \sigma = (\cos(\frac{\pi\alpha}{4}))^{\frac{2}{\alpha}})$.

3: Evaluate f_U at N_{mixt} points by taking the inverse FFT of the CF.

4: Get the mixing function:

$$h(v) = 2v f_U(v^2; \alpha, -1, \sigma)$$

5: Take the finite mixture approximation for the $S\alpha S$ pdf:

$$\tilde{f}(x; \alpha, \gamma) = \frac{\sum_{i=1}^{N_{mixt}} \phi(x; 0, \sigma^2 v_i^2) h(v_i)}{\sum_{i=1}^{N_{mixt}} h(v_i)}$$

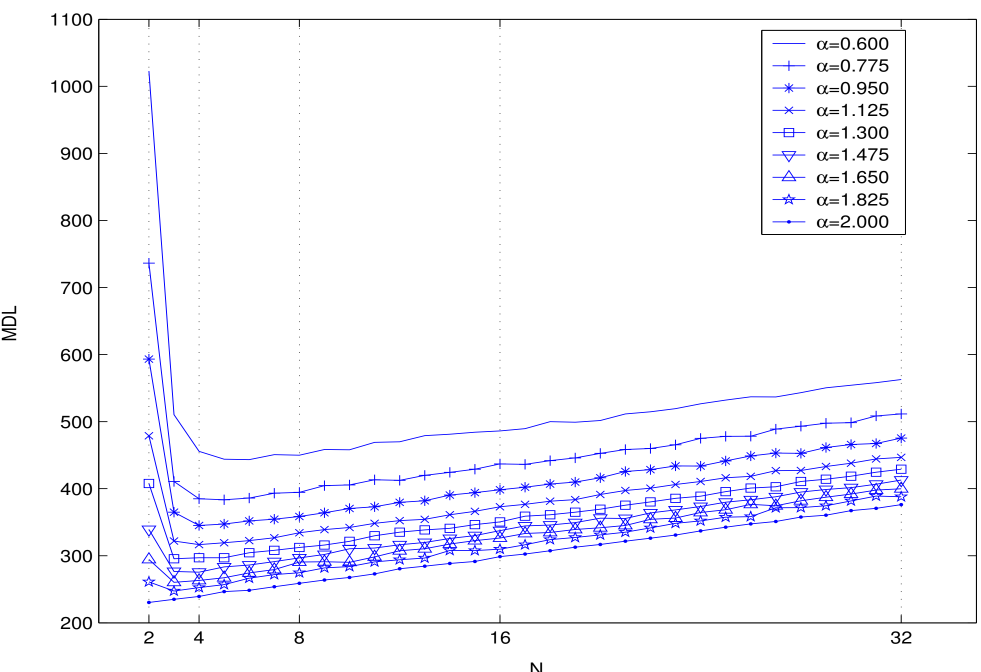
6: Fine tune using the EM algorithm.

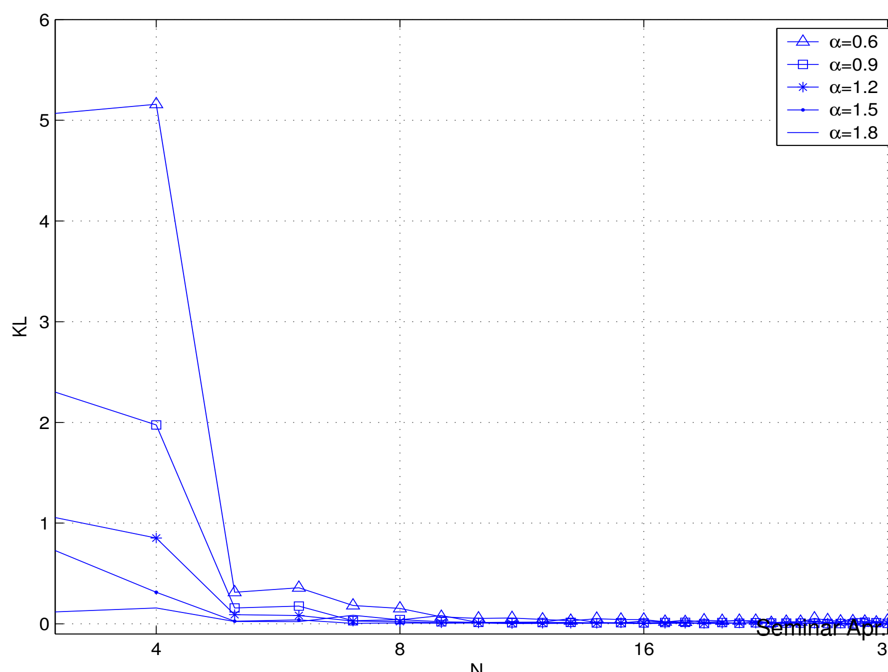
α-stable marginal pdf: N_{mixt}

- Model selection theory: BIC, AIC, MDL (balance between accuracy and complexity):

$$C_{\mathrm{MDL}}(N_{mixt}) = -\ell\ell(\mathbf{x}) + g(N_{mixt}), g \text{ positive and strictly increasing.}$$

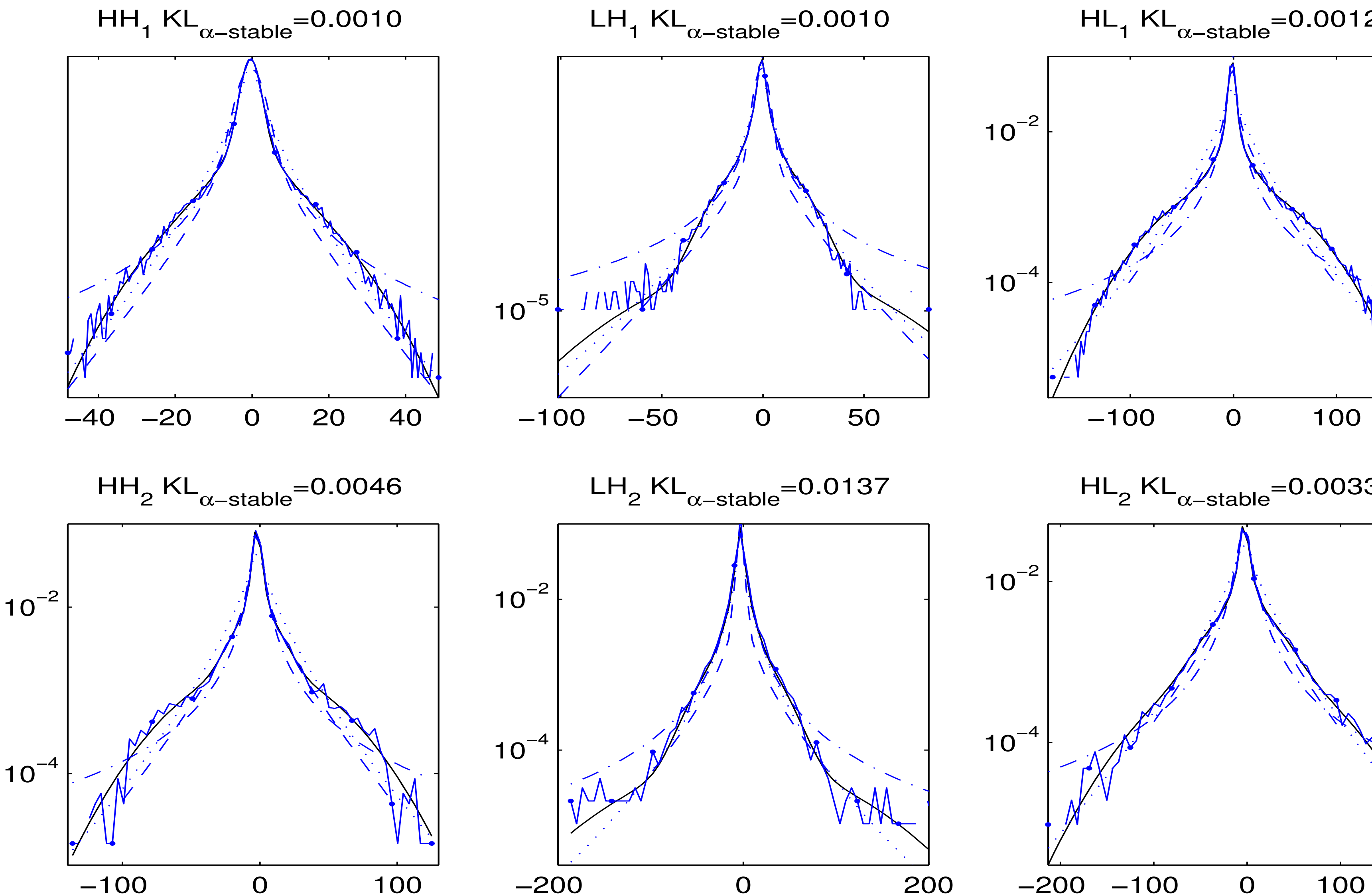
- Conclusion: $N_{mixt} \in [4, 8]$ is fair enough.
- MC simulations with the KL divergence confirmed these findings.
- For a 256×256 image and DWT, less than 1 min (on a 1.5 GHz SunW) to fit all subbands under Matlab.





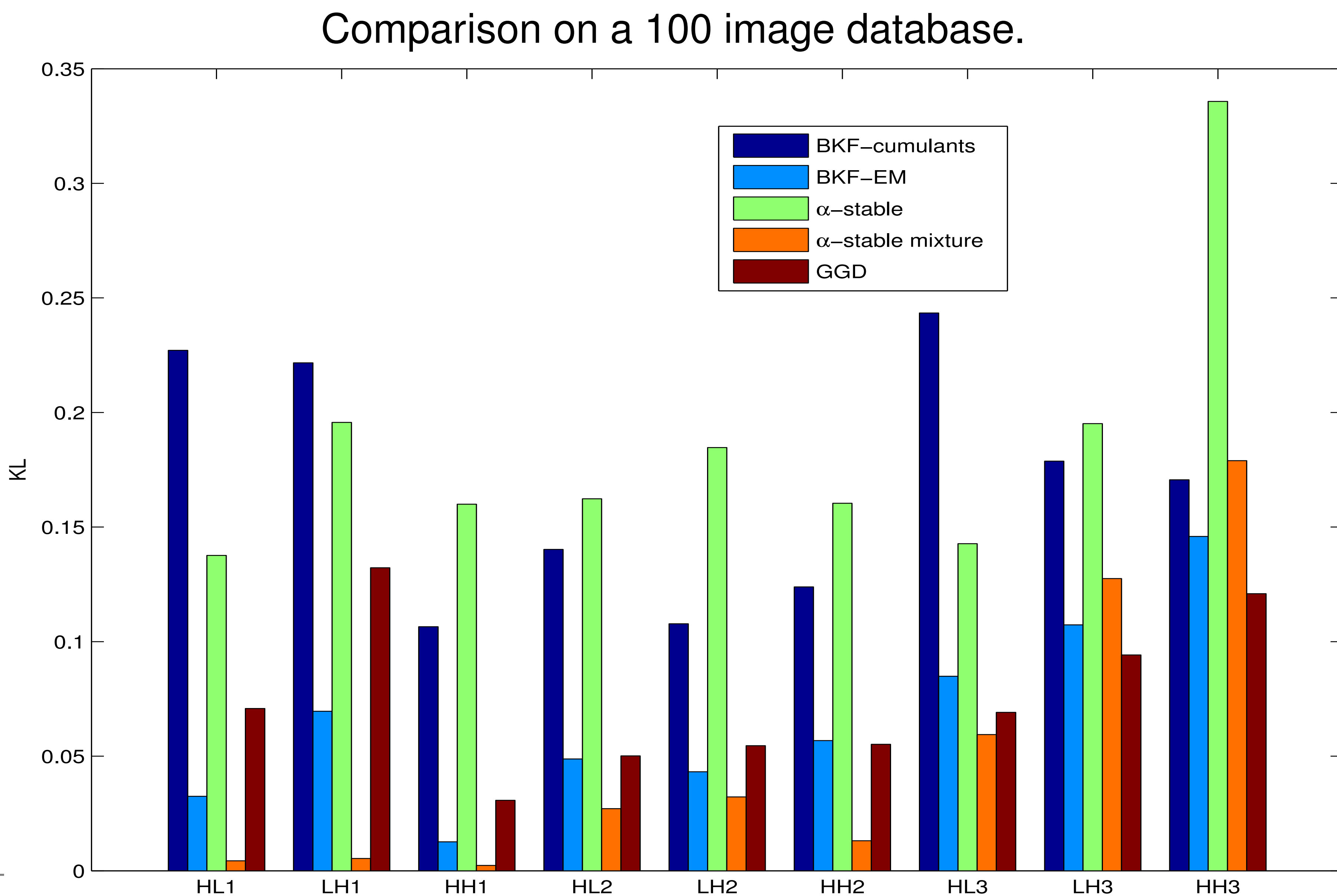
Visual illustration on DWT

Barbara DWT coeffs. modeling



Empirical (-●-), FM α-stable (solid), α-stable (dash-dotted), BKF (dashed), GGD (dotted).

Univariate priors comparison

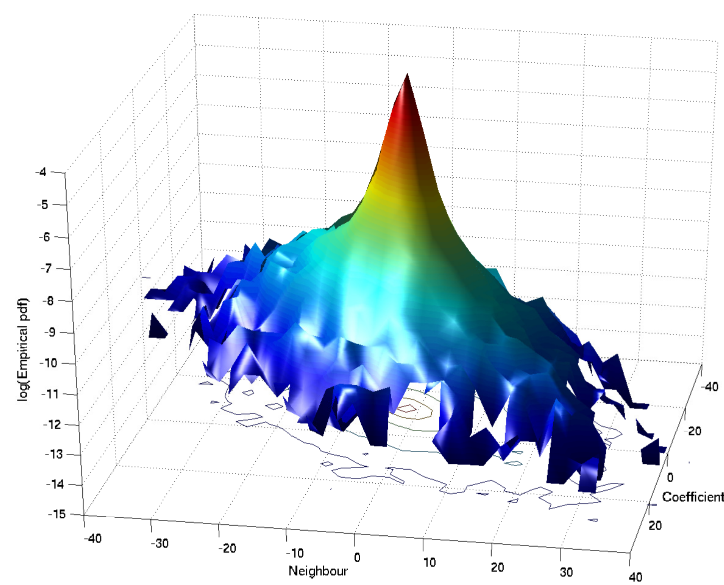


Limitations of univariate modeling

- Coefficients of transforms tend to cluster around fine details in images (exhibit geometrical structures).
- Persistence across scales although orientations are almost decoupled.
- Thus, independence hypothesis is not valid.
- Multivariate priors offer a natural way to handle this behavior.

	Scale J	Scale $J-1$
$IM(X;PX)$	0.164	0.194
$IM(X;VX)$	0.374	0.555
$IM(X;CX)$	0.142	0.151

Average mutual information between coeffs (100 images).



Multivariate SMG family

Definition 4 *Let* $\mathbf{X}$ *be a VRV taking values in* $\mathbb{R}^d$*. Under the SMG, there exists a RV* $U \geq 0$ *and a VRV* $\mathbf{Z} \sim \mathcal{N}(0, \Sigma)$*,* $\Sigma > 0$ *(*U *and* $\mathbf{Z}$ *mutually independent) such that:*

$$\mathbf{X} \stackrel{d}{=} \mathbf{Z}\sqrt{U} \tag{8}$$

Property 4

- *MSMG is a sub-family of elliptically symmetric distributions [Kotz et al. 89].*
- *The pdf of* $\mathbf{X}$ *is:*

$$f_{\mathbf{X}}(x) = (2\pi)^{-d/2}|\Sigma|^{-1}\int_0^{+\infty} u^{-1/2}\exp\left[-\frac{\mathbf{x}^T\Sigma^{-1}\mathbf{x}}{2u}\right] f_U(u)du \tag{9}$$

- *It is unimodal, elliptically symmetric with elliptically symmetric CF:*

$$\Phi_{\mathbf{X}}(\omega) = \mathcal{L}\left[f_U\right]\left(\frac{\boldsymbol{\omega}^T\Sigma\boldsymbol{\omega}}{2}\right) \tag{10}$$

- *A necessary and sufficient condition for a MSMG representation to exist is the alternation of sign of the derivatives of its functional parameter (density generator).*

Multivariate SMG family

Lemma 2 *For a RV U, the measure of multivariate kurtosis of $\mathbf{X}$ in the sense of Mardia is always strictly positive.*

- The multivariate SMG family satisfies the requirements of leptokurticity, heavy-tailness and symmetry.
- This family is again adapted to capture the sparsity and dependency structure of the representation coefficients and is then legitimate as a multivariate prior.

MSMG: hyperparameters estimation

- Again, the distribution of the multiplier U depends on some hyperparameters θ.
- Σ and these hyperparameters are estimated directly from the coefficients:
  - MLE (very time consuming).
  - Moments and Cumulants ($\mathrm{E}\left[\left(\mathbf{X}\Sigma^{-1}\mathbf{X}\right)^i\right] = 2^i \frac{\Gamma(d/2+i)}{\Gamma(d/2)} \mathrm{E}\left[U^i\right]$).
  - EM (easily adapted if univariate EM is accessible).
- Again, this step is a chief obstacle towards good performance of the denoiser.
- Somewhat easy in noiseless case but more complex with corrupting noise.

Case (1): Multivariate BKF

- $\mathbf{X} = \sqrt{U}\mathbf{Z}$ where $U \sim$ Gamma$(\alpha, 1/\alpha), \alpha > 0$

Proposition 2 *The EC MBKF pdf is given by:*

$$f(\mathbf{x}; \alpha, \Sigma) = \frac{2^{1-\alpha}}{\sqrt{(2\pi)^d|\Sigma|}} (2\alpha)^{\frac{\alpha}{2}+\frac{d}{4}} \|\mathbf{x}\|_\Sigma^{\alpha-\frac{d}{2}} K_{\alpha-\frac{d}{2}}\left(\sqrt{2\alpha}\|\mathbf{x}\|_\Sigma\right)$$

K_ν *is the modified Bessel function.*

Property 5

- *The BKF distribution tends to a gaussian as $\alpha \longrightarrow \infty$.*
- *Let $\mathbf{X} \sim MBKF(\alpha, \Sigma)$,*

$$\kappa_j(2i) = \alpha^{1-i}\frac{(2i)!}{2^i i}\sum_{l=1}^{d}\sigma^{2i}(j,l), \quad i \geq 1, j = 1, \ldots, d \qquad (11)$$

$$\mathsf{E}\left[\left(\mathbf{X}\Sigma^{-1}\mathbf{X}\right)^i\right] = 2^i\frac{\Gamma(d/2+i)\Gamma(\alpha+i)}{\alpha\Gamma(d/2)\Gamma(\alpha)} \qquad (12)$$

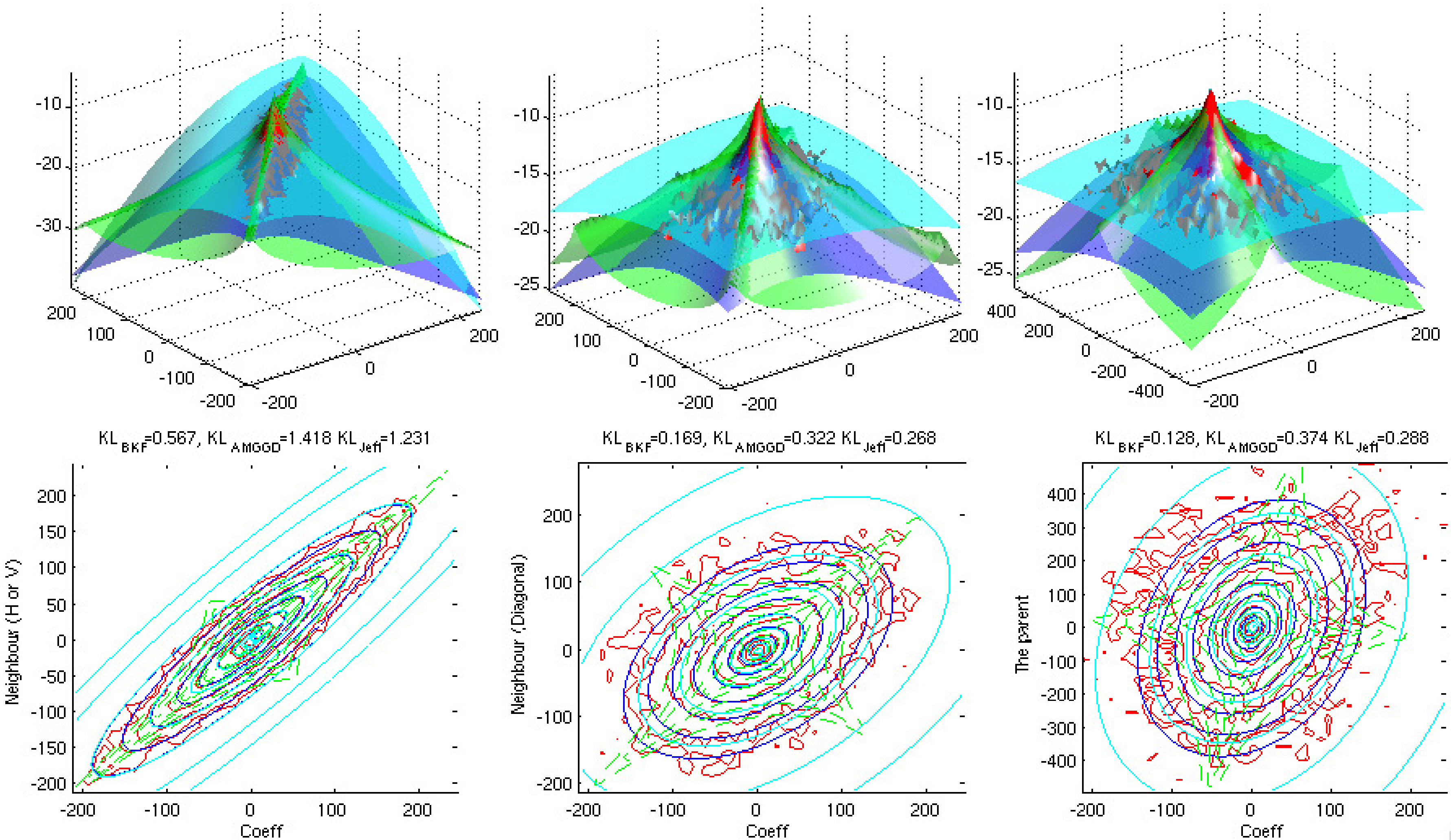
MBKF: estimate α and Σ

- Moments, e.g.:

$$\hat{\Sigma} = \widehat{\mathrm{Cov}\,[\mathbf{X}]}, \hat{\alpha} = \left(\overline{\left(\mathbf{x}\hat{\Sigma}^{-1}\mathbf{x}\right)^2} / (d(d+2)) - 1 \right)^{-1}$$

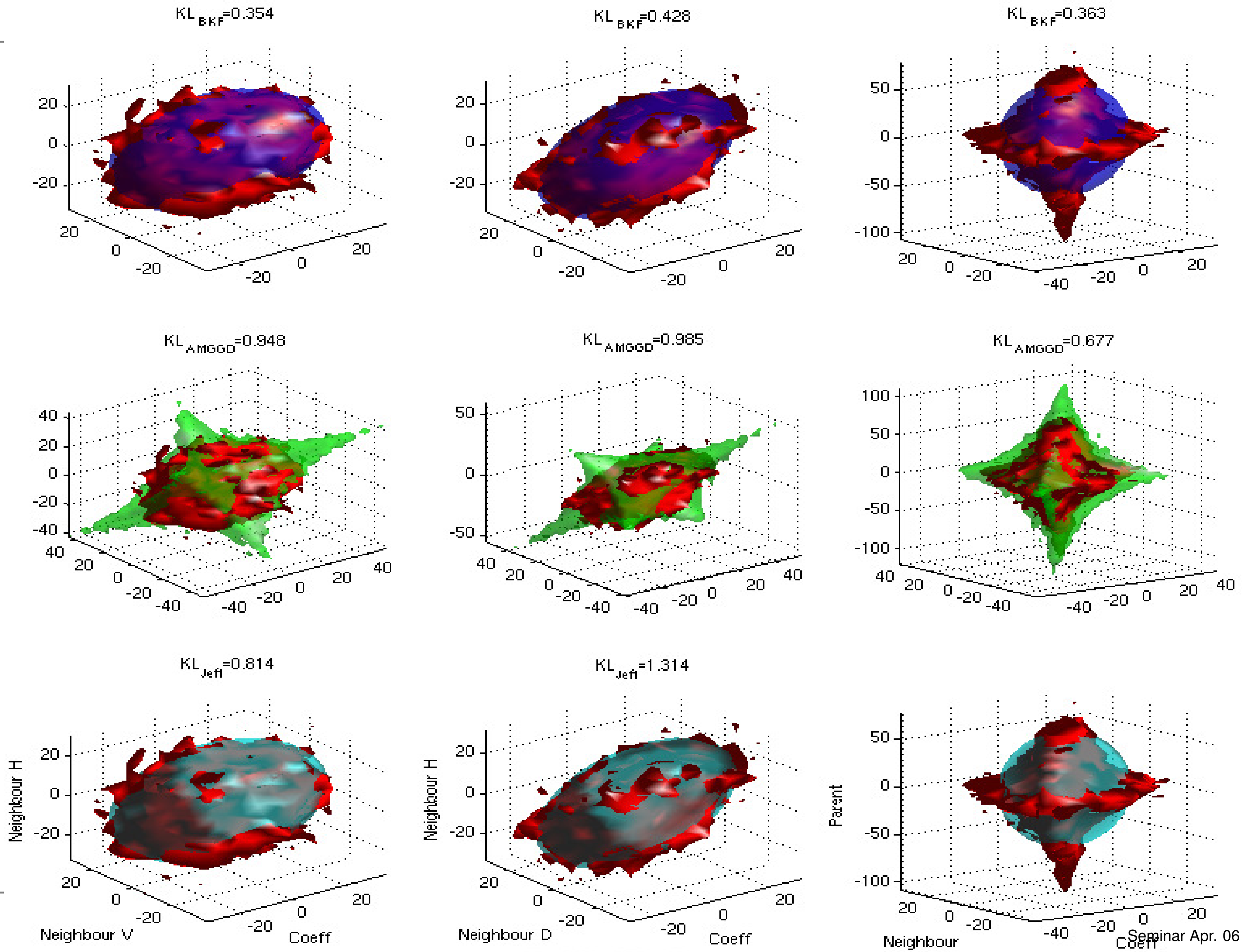
- Using cumulants.
- Using EM algorithm: consider the multiplier as the missing data.
  - The sufficient statistics are calculated in closed form in the E-step.
  - Proof of convergence (to a local minimum).
  - Moment (or cumulant) based estimator is used as an initialization.

Visual illustration on DWT ($d = 2$)

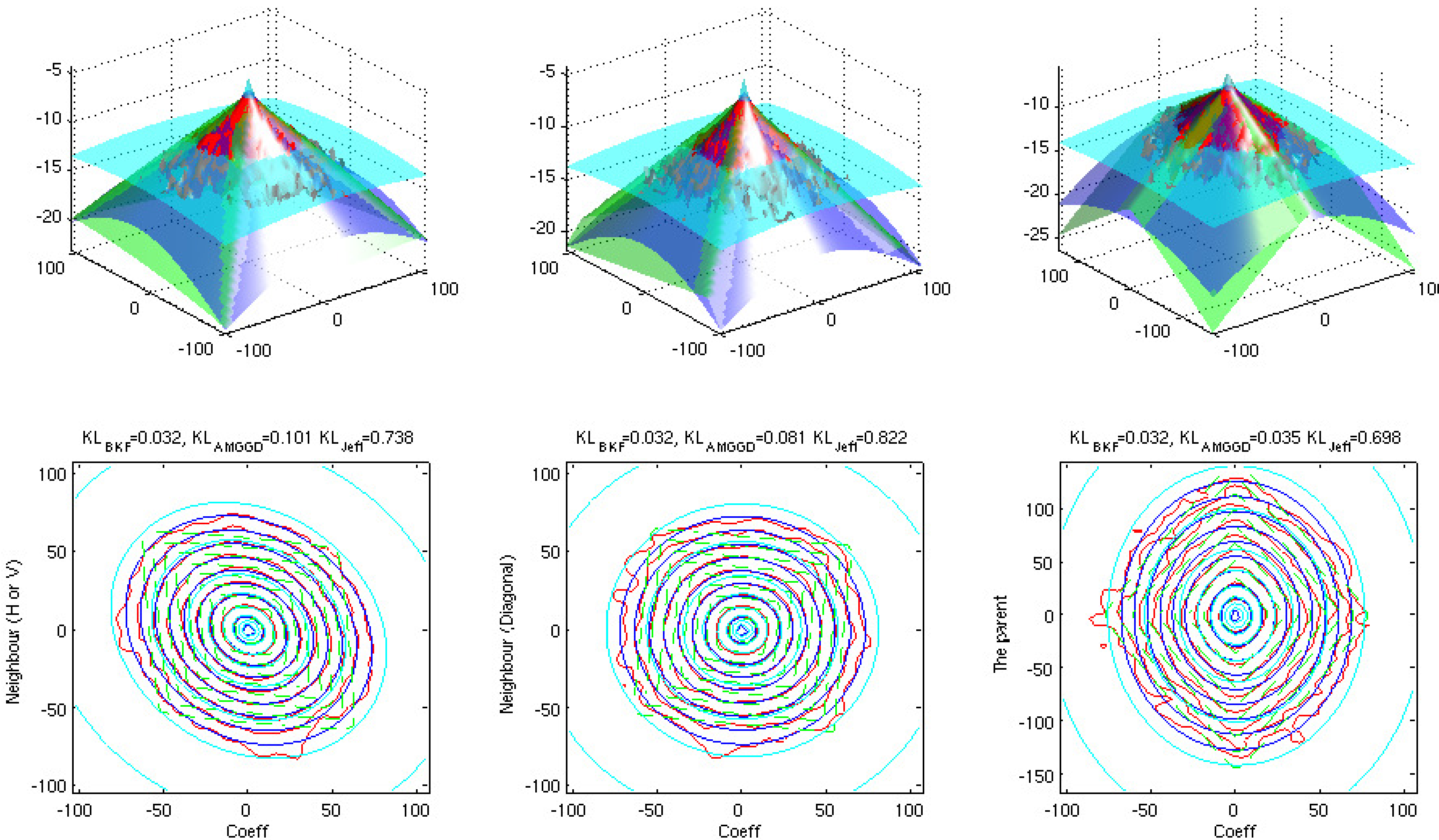


Empirical (red), MBKF (blue), AMGGD (green), MSMG with Jeffrey's multiplier [Portilla et al. 03] (cyan).

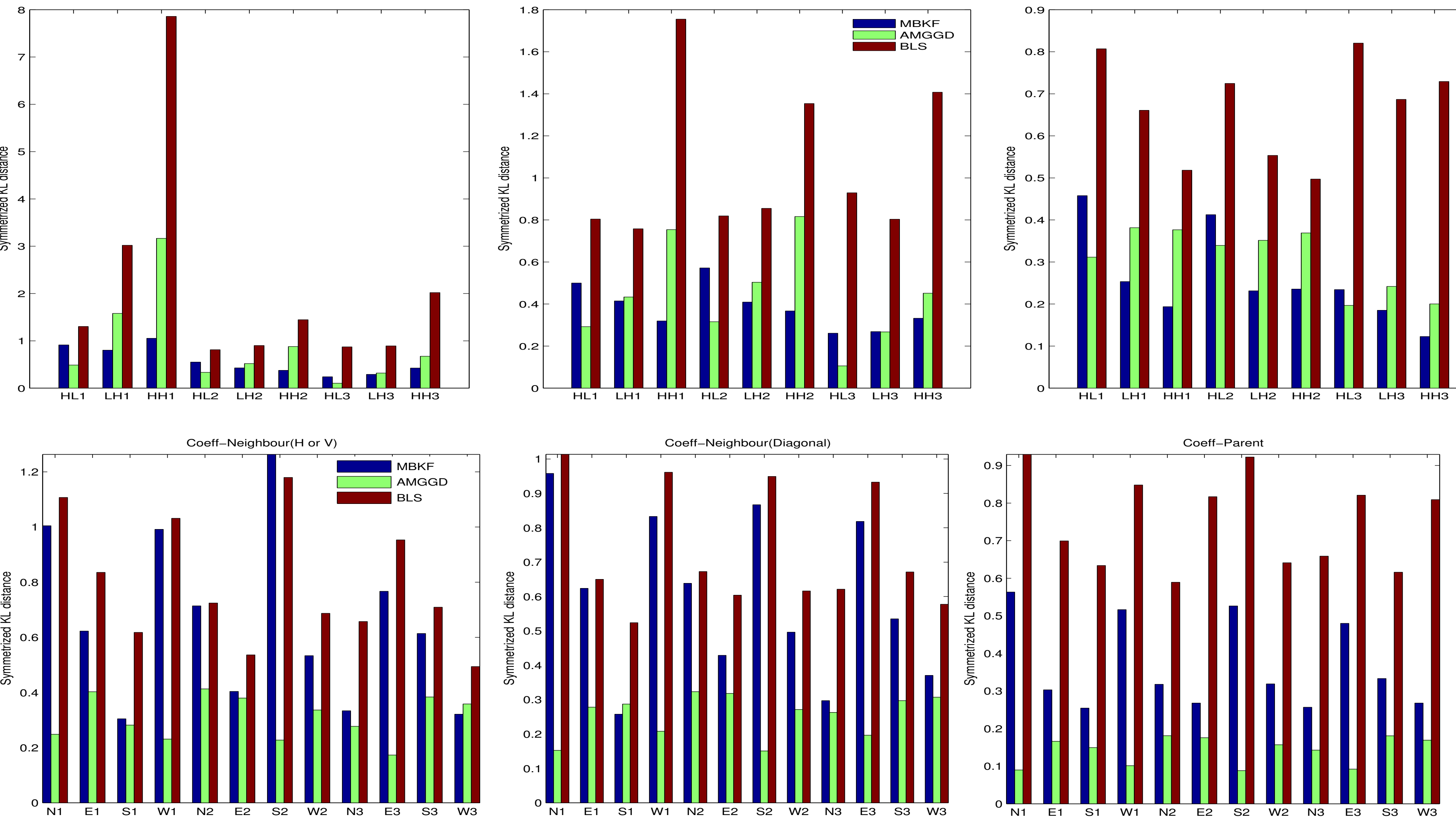
Visual illustration on DWT ($d = 3$)



Visual illustration on FDCT ($d = 2$)



Multivariate priors comparison



Bayesian denoising: univariate

Setting

- The noise in image domain is AWGN.
- The multiscale sparse representation is a basis (typically DWT).
- Observe,

$$d_{\text{sub}} = x_{\text{sub}} + \epsilon, \epsilon \sim \mathcal{N}(0, \sigma_\varepsilon^2)$$

- $x_{\text{sub}} \sim \text{SMG}_{\theta_{\text{sub}}}$ are iid in each subband (+ inter-scale independence).
- We developed bayesian estimators (MMSE and MAP) for BKF and α-stable priors.

Main results

Theorem 3

- *Both the MMSE and the MAP estimators of x_{sub} under the BKF prior (given α and c) have closed analytical forms.*
- *The BKF MAP estimator is equivalent to universal hard thresholding for $\frac{\sigma_\varepsilon^2}{c} = \log N$ as $\alpha \to 1$ (Laplacian prior) or large N.*
- *Bayesian CLT: the BKF MAP estimator is asymptotically gaussian (as $N \to +\infty$).*
- *The MMSE estimator under the approximate α-stable prior has a closed analytical form.*

Hyperparameters estimation

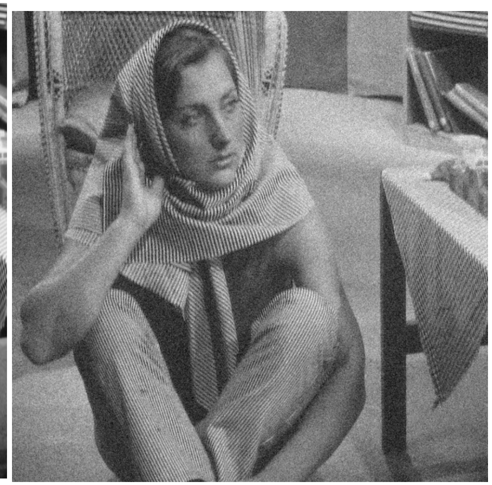
- Cumulant method extends easily for the BKF: $\hat{\sigma}_\varepsilon$ with the MAD of the finest scale, then:

$$\hat{\alpha} = 3\frac{\left(\hat{\kappa}_2 - \hat{\sigma}_\varepsilon^2\right)_+^2}{\hat{\kappa}_4}, \hat{c} = \frac{\left(\hat{\kappa}_2 - \hat{\sigma}_\varepsilon^2\right)_+}{\hat{\alpha}}$$

- EM algorithm for BKF: closed-form expressions are no longer available (not used).
- EM algorithm for approximate α-stable: OK.

Original

Noisy SNR_{in}=15 dB



α−stable mixture 20.84 dB

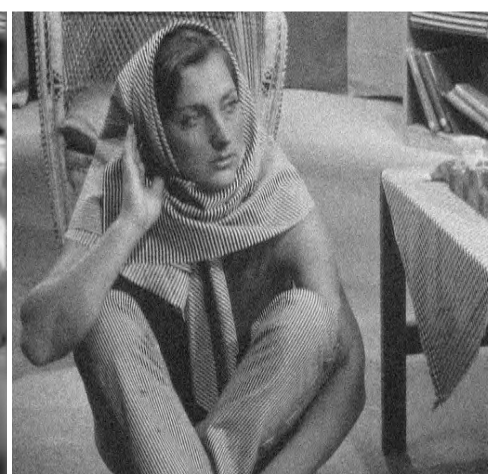
α−stable 19.04 dB

BKF 20.64 dB

Hard universal 17.00 dB

Soft universal 15.55 dB

SURE 18.14 dB



Oracle Threshold 18.87 dB

Original

Noisy SNR_{in}=15 dB

α–stable mixture 21.34 dB

α–stable 19.12 dB

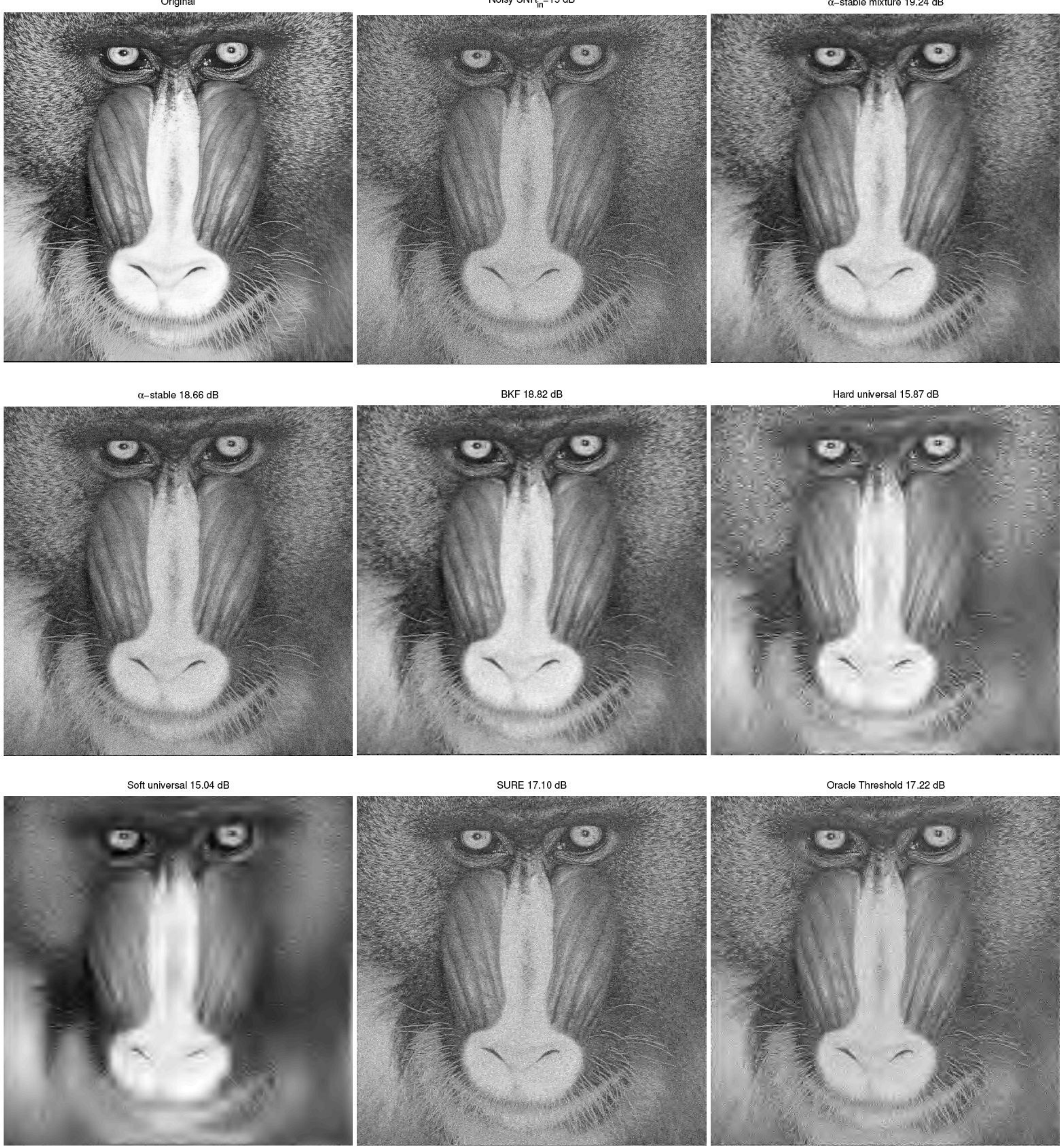
BKF 21.30 dB

Hard universal 17.92 dB

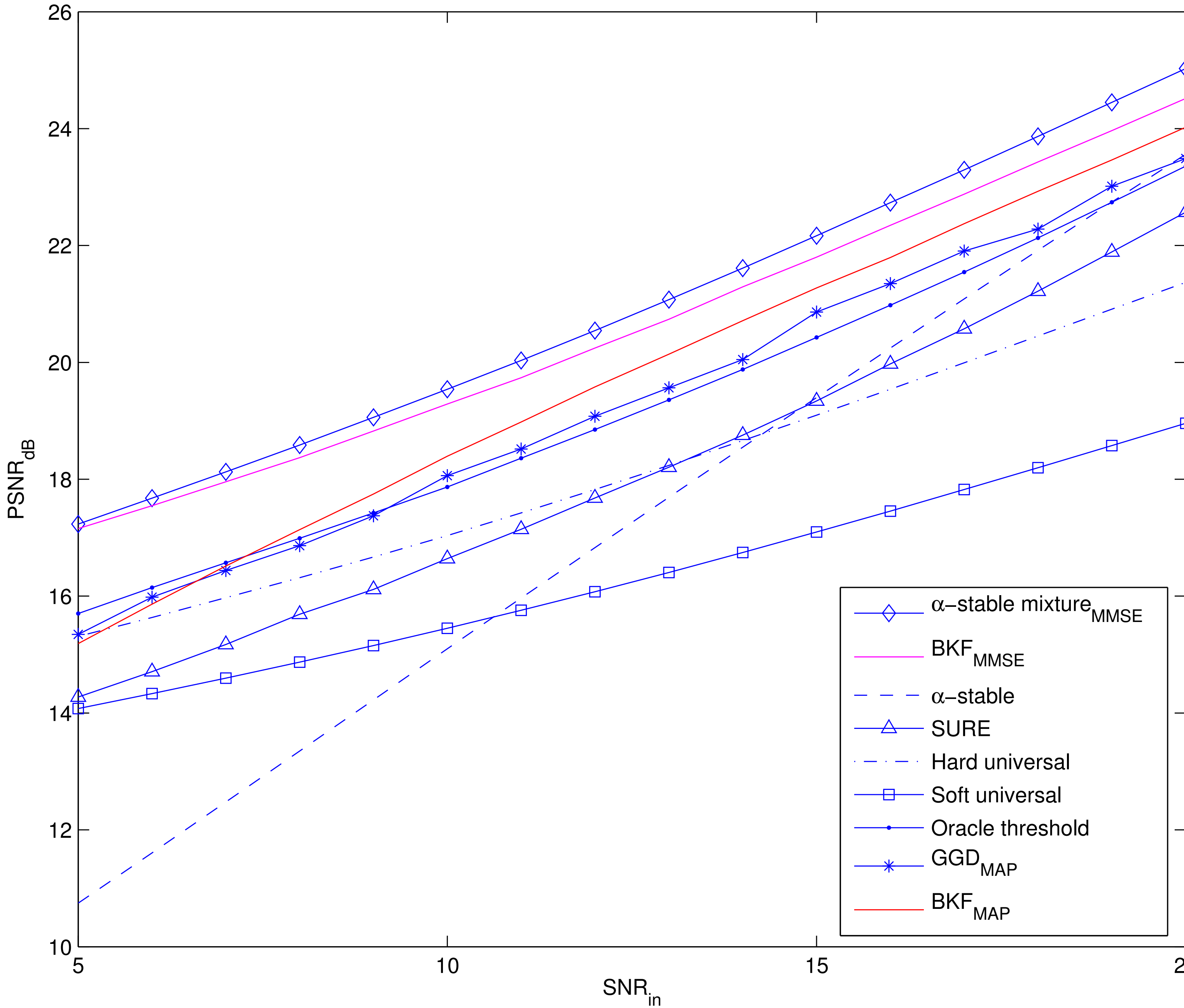
Soft universal 15.47 dB

SURE 18.23 dB

Oracle Threshold 19.33 dB


Original
Noisy SNR$_{in}$=15 dB
α–stable mixture 19.24 dB
α–stable 18.66 dB
BKF 18.82 dB
Hard universal 15.87 dB
Soft universal 15.04 dB
SURE 17.10 dB
Oracle Threshold 17.22 dB

PSNR comparison



Limitations of univariate denoising

- Lack of translation invariance with orthogonal transforms.
- Gibbs-like ringing artifacts.
- Better use directional and (almost) translation invariant representations.
- Taking into account dependency structure of neighboring coefficients is likely to improve the denoising performance.

Bayesian denoising: multivariate

Setting

- The noise in image domain is AWGN (could be generalized to colored noise).
- The representation is a frame, a tight frame or a basis (typically FDCT, UDWT).
- Observe,
$$\mathbf{d} = \mathbf{x} + \boldsymbol{\epsilon}, \boldsymbol{\epsilon} \sim \mathcal{N}(0, \Sigma_\varepsilon)$$
- Each $\mathbf{d}$ is a vector of packed coefficients (neighbors + parent).
- $\mathbf{x} \sim \mathrm{MSMG}_\theta$.

Preliminary results

- So far, we developed the MMSE bayesian estimator for the MBKF prior, but no analytical expression.

$$\hat{\mathbf{x}}_{\text{MMSE}} = \frac{\int_0^{+\infty} u\Sigma\left(u\Sigma + \Sigma_\varepsilon\right)^{-1}\mathbf{d}\;\phi\left(\mathbf{d}; u\Sigma + \Sigma_\varepsilon\right) f_U(u)du}{\int_0^{+\infty} \phi\left(\mathbf{d}; u\Sigma + \Sigma_\varepsilon\right) f_U(u)du}$$

- If the dictionary is sufficiently incoherent, then $\Sigma_\varepsilon \approx \sigma_\varepsilon^2 I_d$.
- This estimator corresponds to a scale mixture of local Wiener estimates.
- Deploy the bayesian integration technology:
  - Analytic approximation (e.g. Laplace, Saddlepoint).
  - Quadrature numerical integration (accurate but slow).
  - Monte-Carlo Integration (fast and accurate).

Hyperparameters estimation

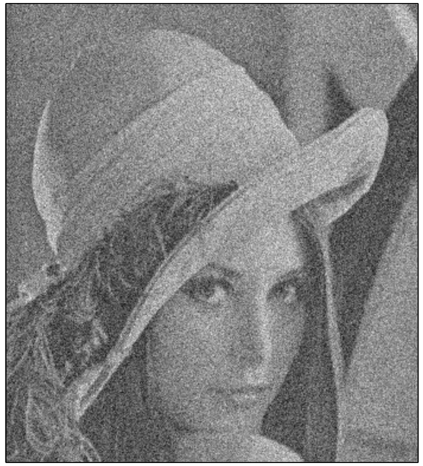
- Cumulant method extends easily for the noisy MBKF:

$$\hat{\Sigma} = \widehat{\mathrm{Cov}\,[\mathbf{d}]} - \Sigma_\varepsilon, \hat{\alpha} = \frac{3}{d} \sum_{j=1}^{d} \frac{\sum_{l=1}^{d} \hat{\sigma}^4(j,l)}{\hat{\kappa}_j(4)}$$

- EM algorithm for BKF: closed-form expressions are no longer available (not used).

Original

Noisy 18.94 dB (σ_ε=40)



UDWT H 28.00 dB

FDCT H 28.01 dB

UDWT UBKF 27.87 dB

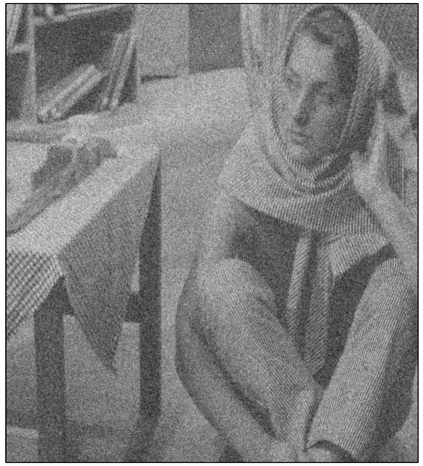
UDWT MBKF 28.60 dB

FDCT MBKF 28.12 dB

[Starck et al. 01] 28.93 dB

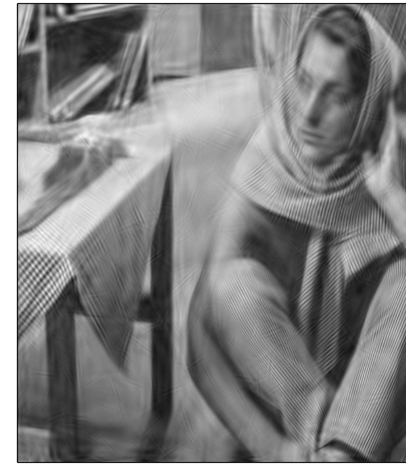
Original

Noisy 19.26 dB (σ_{ε}=40)

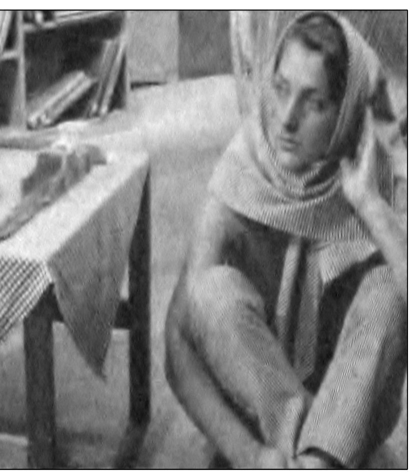


UDWT H 23.77 dB

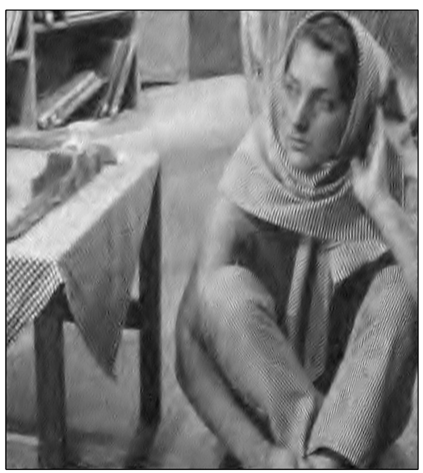
FDCT H 25.95 dB



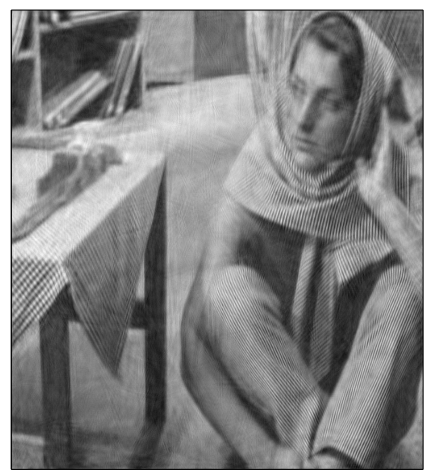
UDWT UBKF 24.21 dB



UDWT MBKF 25.79 dB



FDCT MBKF 26.72 dB



[Starck et al. 01] 26.08 dB

Take away messages

- For bases (i.e. DWT):
  - Bayesian estimators with USMG prior are clearly better than many other denoisers (including bayesian).
  - But bases suffer from many limitations, e.g. translation invariance.
- For frames and tight frames (i.e. UDWT, FDCT):
  - Saliency of univariate bayesian methods is limited (compared e.g. to simple hard thresholding).
  - Incorporation of MSMG priors clearly improves the performance of denoisers.
  - MBKF prior based denoiser compares favorably with state-of-the art denoisers [Portilla et al. 03 (Jeffrey's prior), Starck et al. 01] (courtesy of JLS for providing his code).
  - Curvelets are crucial for faint curvilinear structures.

Conclusion and perspectives

- A flexible statistical prior is proposed to model both marginal and joint statistics of sparse representation coefficients.
- Univariate and multivariate properties derived and special cases fully considered.
- Application to statistical modeling of real images.
- Bayesian term-by-term MMSE and MAP estimators were also derived.
- Preliminary results for the joint MMSE estimator.
- Experimental results are very encouraging.
- Simplification: the joint prior denoiser considered independence of overlapping blocks of coefficients.
- Rigor: this local model actually defines a global Markov model $\longrightarrow$ must be rigorously considered, but estimation will be more complicated.
- Investigate other special priors for whom bayesian estimators could be calculated analytically.
- A more comprehensive comparative study for the multivariate case.

More details

- L. Boubchir, M.J. Fadili, "A Closed-form Nonparametric Bayesian Estimator in The Wavelet-domain of Images Using an Approximate α-stable Prior", Pattern Recognition Letters, in press, 2006.
- M.J. Fadili, L. Boubchir, "Analytical form for a Bayesian wavelet estimator of images using the Bessel K form densities", IEEE Transactions on Image Processing, Vol. 14, No. 2, pp. 231-240, 2005.
- L. Boubchir, M.J. Fadili, "Multivariate Statistical Modeling of Images with The Curvelet Transform", IEEE ISSPA 2005, pp. 747-750, Sydney, Australia, August 28-31, 2005.
- L. Boubchir, M.J. Fadili, "Bayesian Denoising Based on The MAP Estimation in Wavelet-domain Using Bessel K Form Prior", IEEE ICIP'05, Vol. I, pp. 113-116, Genoa, Italy, September 11-14, 2005.
- L. Boubchir, M.J. Fadili, D. Bloyet, "Bayesian Denoising in the Wavelet-domain Using an Analytical Approximate α-stable prior", IEEE ICPR 2004, Vol. 4, pp. 889-892, Cambridge, United Kingdom, August 23-26, 2004.

`http://www.greyc.ensicaen.fr/~jfadili`

`http://www.greyc.ensicaen.fr/~boubchir`

Any questions ?