Sparse Representations and Bayesian Denoising

Jalal M. FADILI and Larbi BOUBCHIR

GREYC CNRS UMR 6072, Caen France

Outline

- Introduction.
  - Observation model.
  - Bayesian paradigm.
- Ingredients from modern harmonic analysis.
- Statistical sparse priors:
  - Univariate (marginal).
  - Multivariate (joint).
- Bayesian estimation.
- Combining transforms.
- Conclusion and extensions.

Observation model

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

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

where:

- $\odot$ is any composition of two arguments (e.g. '+', '·').
- $s \in S$ 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).
- Ψ 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 (or signal) is viewed as realization(s) of a RV or a random field whose degradation equation is:

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

where:

- $\odot$ is any composition of two arguments (e.g. '+', '·').
- $s \in S$ 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).
- Ψ 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

$$Y = \Psi((B\,X)) \odot \epsilon$$

- $p(x, \mathbf{z})$: *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	MF	$\mathsf{E}[X\|Y]$
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.
- MF involves solving an integration problem.
- For mutually independent iid gaussian signal and noise, MAP, MF and Wiener are the same.

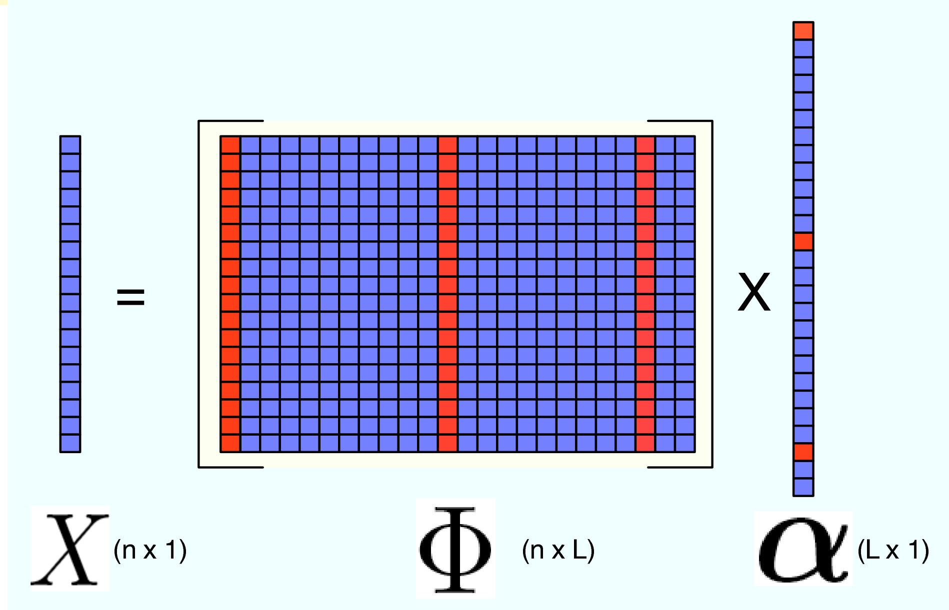
What prior ?

- 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].
- Transported Generator Models
  - Random objects randomly placed (according to some probability law) while imposing some axioms such as scale invariance (multi-scale nature of images) [Gidas et Mumford 01, Grenander et al. 99-03].
- Random Field Theory (e.g. MRF) [Besag 86, Geman and Geman 88].
- Sparse representation-based prior [Mallat 89, Simoncelli et al. 98, Wainwright et al. 00, Grenander et al. 01, Achim et al. 01, Fadili et al. 03, etc].

Elements from modern harmonic analysis

- An n-sample signal or image X can be written as the superposition of elementary functions $\phi_\gamma(s)$ (*atoms*) parameterized by γ s.t. (Γ is denumerable):

$$X_s = \sum_{\gamma \in \Gamma} \boldsymbol{\alpha}_\gamma \phi_\gamma(s), \ \phi_\gamma \in \mathcal{L}, \|\boldsymbol{\alpha}\|_0 \ll n$$



- The atoms $\{\phi_l\}_{l=1,\ldots,L}$ are normalized to a unit ℓ_2 norm.
- $\Phi = [\phi_1 \ldots \phi_L] \in \mathbb{R}^{N \times L}, \mathrm{Card}\,\Gamma = L \geq N$ (bases, tight frames or frames).
- 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).

Sparse representation-based denoising

- Stable recovery of a sparse representation in presence of (AWG) noise (with bounded variance):

$$\min_{\boldsymbol{\alpha}} \|\boldsymbol{\alpha}\|_0 \text{ s.t. } \|Y - \Phi\boldsymbol{\alpha}\|_2 \leq \delta$$

- Convexify and relax [Chen et al. 01, Donoho et al. 04, Gribonval et al. 04, Fuchs 05, etc]:

$$\min_{\boldsymbol{\alpha}} \frac{1}{2} \|Y - \Phi\boldsymbol{\alpha}\|_2 + \lambda \|\boldsymbol{\alpha}\|_1$$

- Corresponds to the MAP estimate with Laplacian prior.
- Here, the ℓ_1 norm prior will be replaced by a flexbile family of sparsity promoting priors and the MAP by the MF.

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 1

- *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 *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.*

- For sparse representations, empirical coefficient pdfs are symmetric around 0, leptokurtic and heavy tailed.
- These pdfs have their 1st and 2nd derivatives of alternating signs on $\mathbb{R}^+$.
- The SMG family satisfies all these requirements.
- 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.

Relating USMG to Besov space

- An explicit relationship between the parameters of the USMG prior on the wavelet coefficients of g and the Besov space within which g may fall (a.s.).

Theorem 1 *Let* $X_{j,k} \stackrel{d}{=} Z_j\sqrt{U}$ *iid RVs at each scale such that* $Z \sim \mathcal{N}(0, \sigma_j^2)$, $\mathsf{E}[U] = 1$ *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* $c_{0,0}$*,*

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

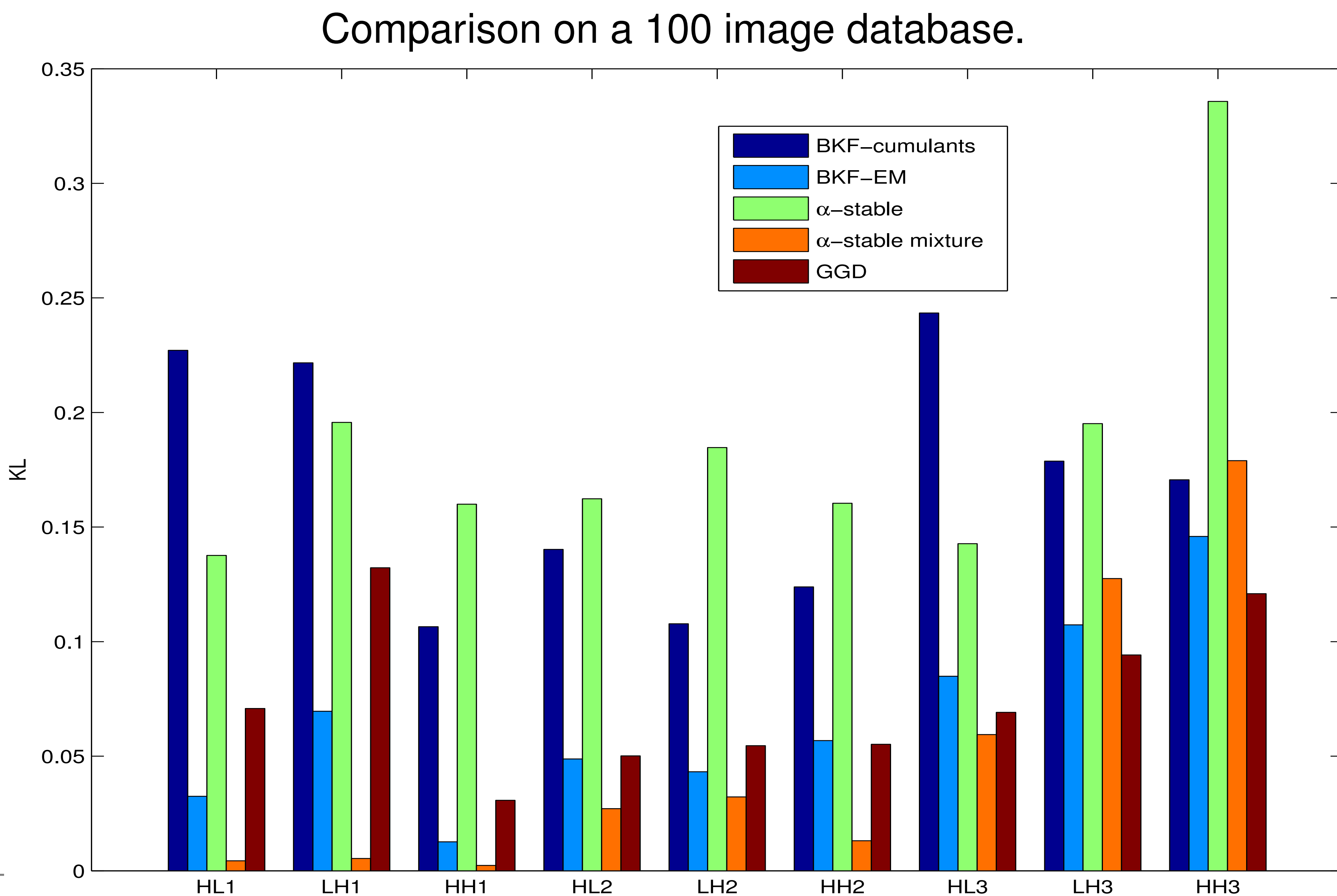
USMG: empirical Bayes

- 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.

USMG: examples

- The Bessel K form prior [Grenander et al. 03, Fadili et al. 03]: the mixing RV is Gamma distributed.
- The α-stable prior [Achim et al. 01, Boubchir and Fadili 03]: the mixing RV is also α-stable.
- The GGD prior [Mallat 89, Moulin and Liu 99]: the mixing RV is not known in a closed form.

Example on the DWT



Multivariate SMG family

Definition 2 *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:*

$$\mathbf{X} \stackrel{d}{=} \mathbf{Z}\sqrt{U} \tag{7}$$

Property 2

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

$$f_{\mathbf{X}}(x) = (2\pi)^{-d/2}|\Sigma|^{-1} \int_0^{+\infty} u^{-1/2} \exp\left[-\frac{\mathbf{x}^T\Sigma^{-1}\mathbf{x}}{2u}\right] f_U(u)du \tag{8}$$

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

$$\Phi_{\mathbf{X}}(\omega) = \mathcal{L}\left[f_U\right]\left(\frac{\boldsymbol{\omega}^T\Sigma\boldsymbol{\omega}}{2}\right) \tag{9}$$

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.*
- *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).*

- 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.

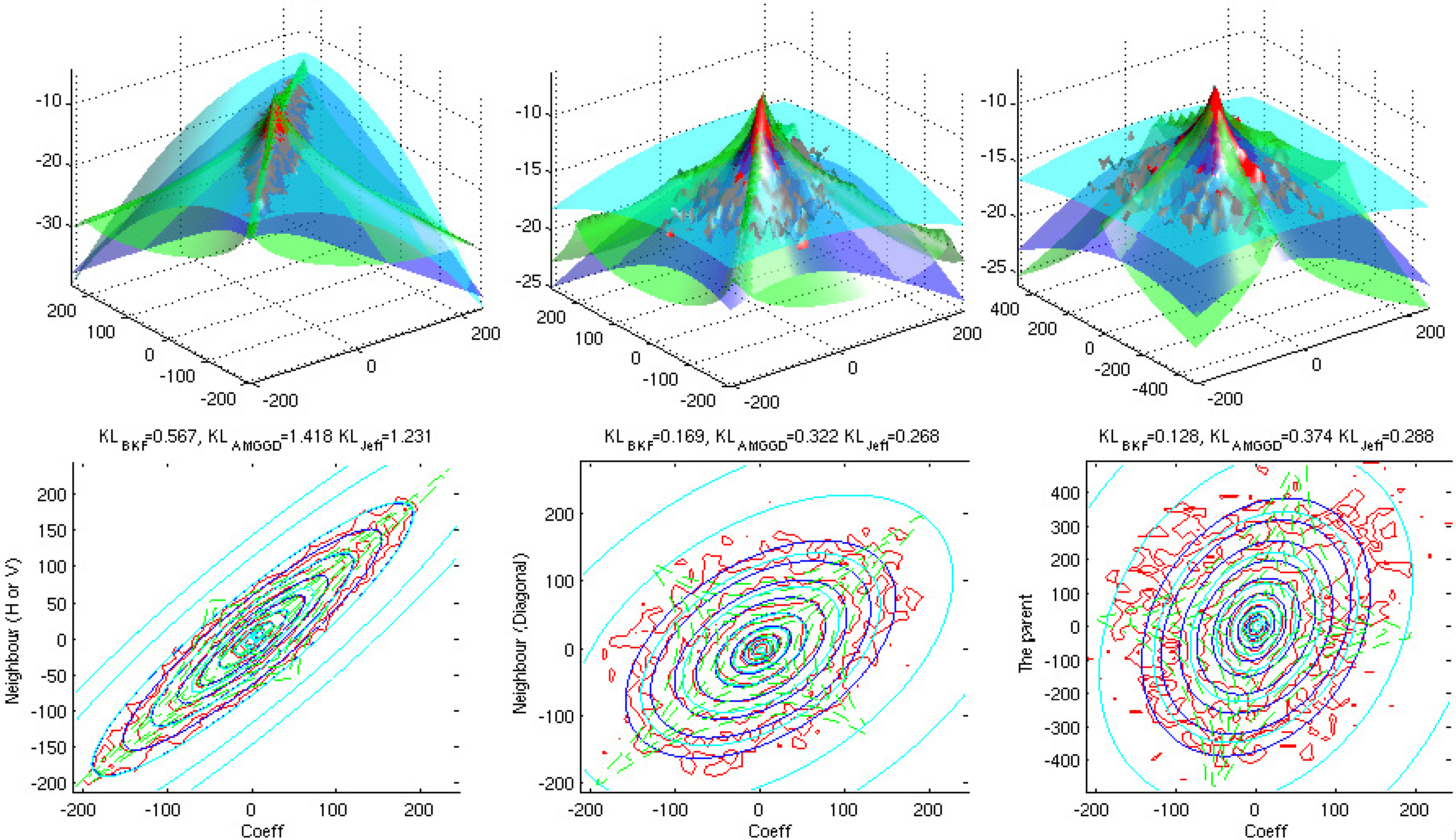
MSMG: empirical Bayes

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

MSMG: examples

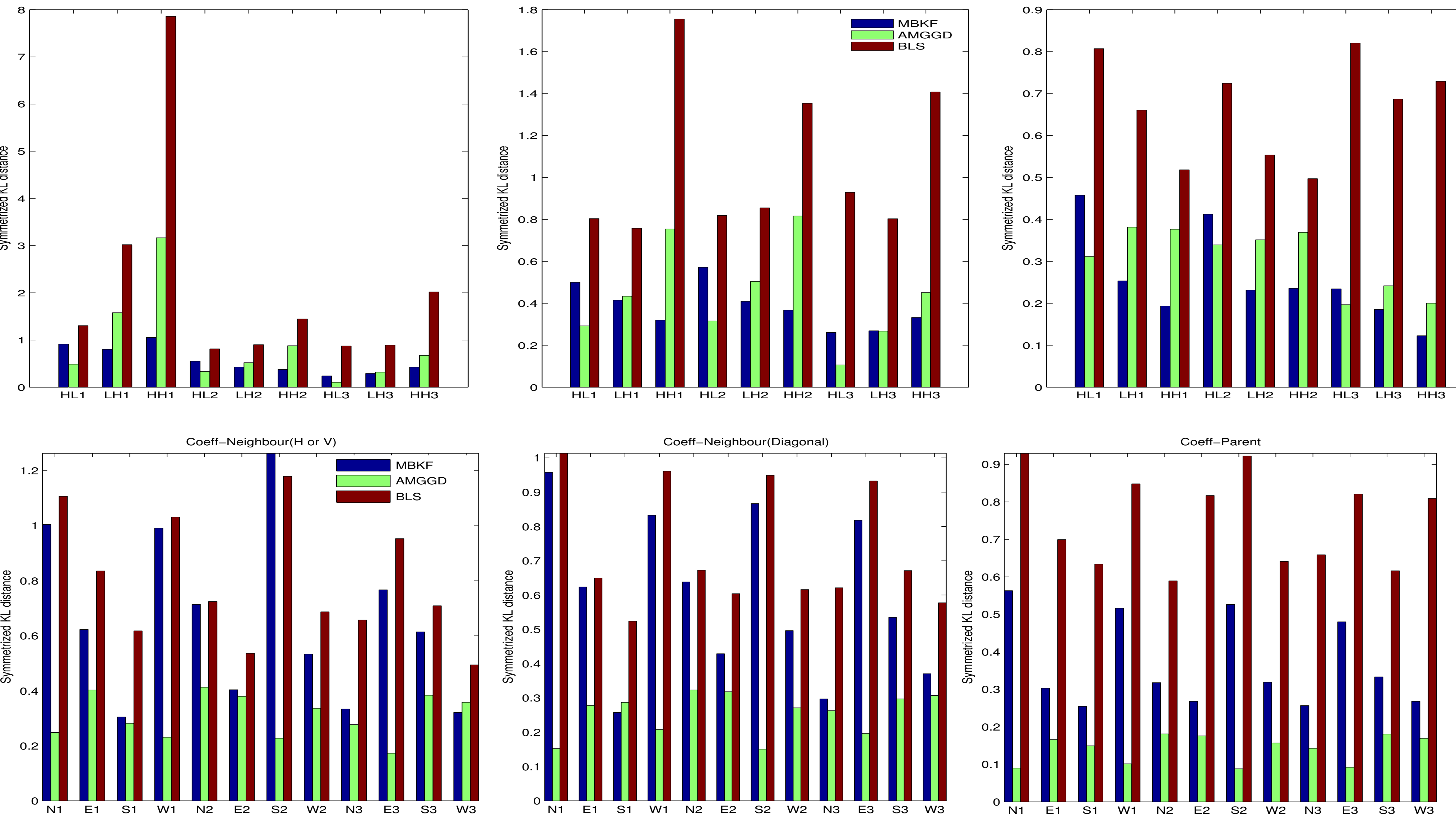
- The multivariate Bessel K form prior [Fadili et al. 06]: the mixing RV is Gamma distributed.
- The α-stable prior [Kuruoglu et al. 04]: the mixing RV is also α-stable.
- The GGD prior [Fadili et al. 05].
- Non-informative Jeffrey's prior [Portilla et al. 03].

Visual illustration ($d = 2$)



Empirical (red), MBKF (blue), AMGGD (green), MSMG with Jeffrey's multiplier (cyan).

Example on the DWT and FDCT



Application to Bayesian denoising

- The MF estimator corresponds to a scale mixture of local Wiener estimates:

$$\hat{\boldsymbol{\alpha}}_{\mathrm{MF}} = \frac{\int_0^{+\infty} \underbrace{\hat{\boldsymbol{\alpha}}|u}_{\text{Wiener estimate}} \phi\left(\mathbf{d}; u\Sigma + \Sigma_\varepsilon\right) f_U(u) du}{\int_0^{+\infty} \phi\left(\mathbf{d}; u\Sigma + \Sigma_\varepsilon\right) f_U(u) du}$$

- Closed-form expressions.
- Deploy the bayesian integration technology:
  - Analytic approximation (e.g. Laplace, Saddlepoint).
  - Quadrature numerical integration (accurate but slow).
  - Monte-Carlo Integration (fast and accurate).

Application to Bayesian denoising (cont'd)

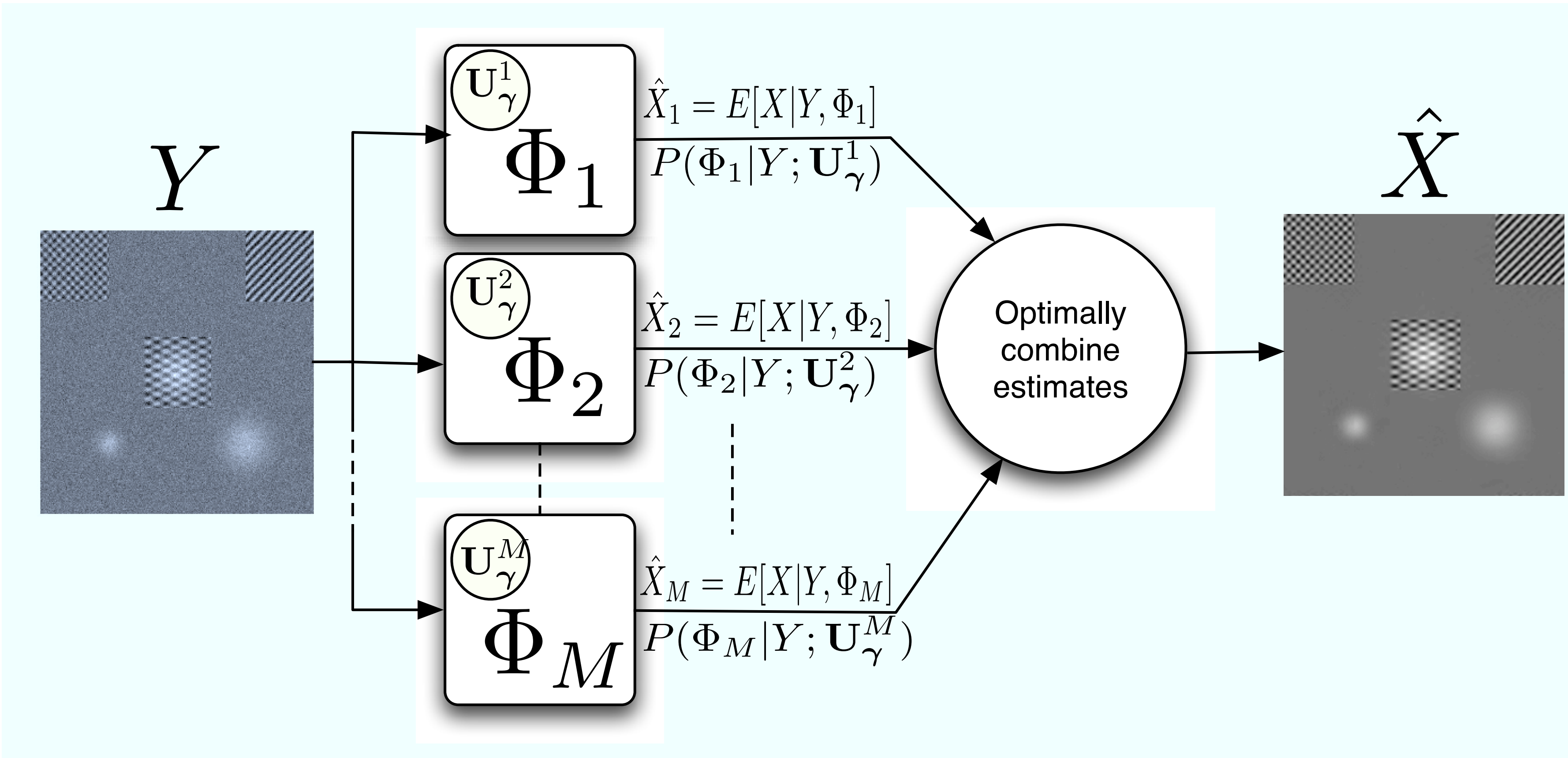
Main results

Theorem 2

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

Combining transforms: bases

- The dictionary is a union of M (sufficiently incoherent) bases $\{\Phi_m\}_{m=1,\ldots,M}$.



$$\hat{X} = \sum_{m=1}^{M} \hat{X}_m P(\Phi_m|Y;\mathbf{U}_\gamma^m)$$

Combining transforms: bases

- Suppose that:
  - The SMG prior is independent of the transform.
  - The transforms are equiprobable (or given by a learning step).
- For the USMG:

$$P(\Phi_m|Y_s;\mathbf{U}_\gamma^m) = \frac{\int\cdots\int_0^{+\infty}\mathcal{N}(Y_s;\underbrace{\mu_m}_{\text{LP component}},V_m(s)+\sigma_\varepsilon^2)\underbrace{f_{U_\gamma^m}(u_\gamma^m)}_{\text{SMG prior}}\,du_1^m\ldots du_{|\Gamma_m|}^m}{\sum_{m'=1}^M\int\cdots\int_0^{+\infty}\mathcal{N}(Y_s;\mu_{m'},V_{m'}(s)+\sigma_\varepsilon^2)f_{U_\gamma^{m'}}(u_\gamma^{m'})\,du_1^m\ldots du_{|\Gamma_m|}^m}$$

$$V_m(s) = \sum_{\gamma_m}|\phi_{\gamma_m}(s)|^2\,u_\gamma^m$$

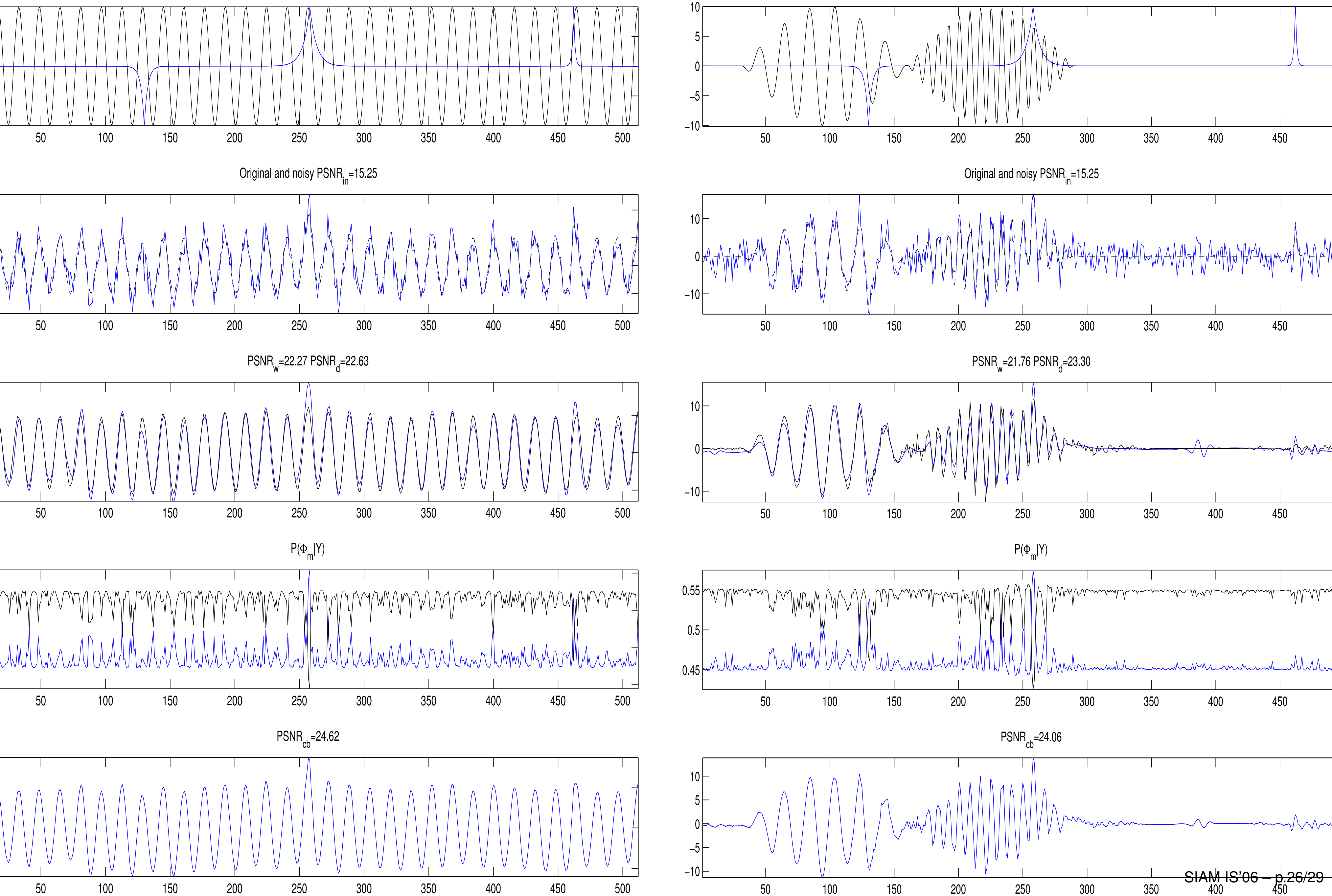
- For USMG priors with mixing RVs subband-independent and rapidly decreasing pdfs.

$$P(\Phi_m|Y_s;\mathbf{U}_\gamma^m) = \frac{\mathcal{N}(Y_s;\mu_m,\hat{V}_m+\sigma_\varepsilon^2)}{\sum_{m'=1}^M\mathcal{N}(Y_s;\mu_{m'},\hat{V}_{m'}+\sigma_\varepsilon^2)}$$

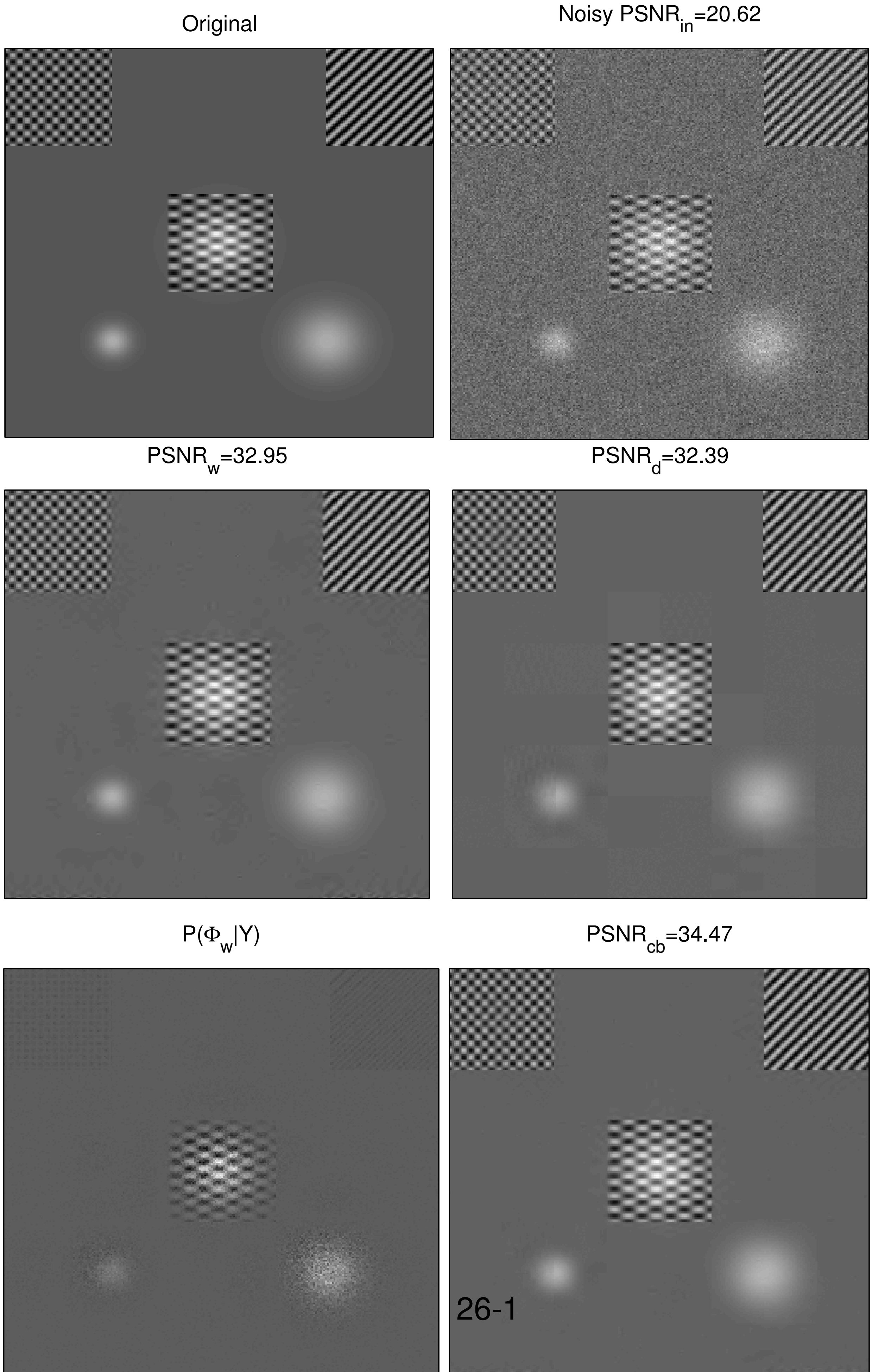
- For many usual transform bases, $\hat{V}_m$ is exactly (or to a good approximation):

$$\hat{V}_m = (\approx)\sum_{\gamma_m}\underbrace{\hat{u}_\gamma^m}_{\text{Empirical bayes estimate}}$$

1D Examples

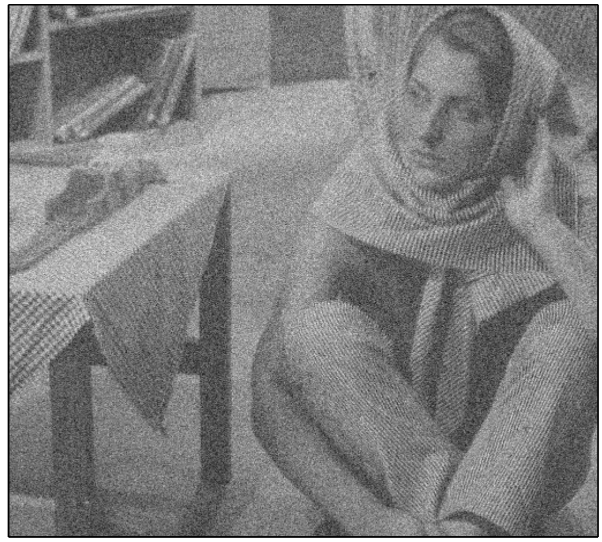


Original and noisy PSNR_in=15.25
Original and noisy PSNR_in=15.25
PSNR_w=22.27 PSNR_d=22.63
PSNR_w=21.76 PSNR_d=23.30
P(Φ_m|Y)
P(Φ_m|Y)
PSNR_cb=24.62
PSNR_cb=24.06


Original
Noisy PSNR$_{in}$=20.62
PSNR$_w$=32.95
PSNR$_d$=32.39
P(Φ_w|Y)
PSNR$_{cb}$=34.47
26-1

Original

Noisy $PSNR_{in}$=16.09



$PSNR_w$=23.95

$PSNR_d$=24.22

$P(\Phi_w|Y)$

$PSNR_{cb}$=25

Combined PSNR=25

MBKF UDWT PSNR=24.7

MBKF FDCT PSNR=26.12

Conclusion and perspectives

- A flexible statistical prior 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 MF and MAP estimators also derived.
- Combined MF denoising with bases.
- Extension to other transforms.
- Extension to handle more rigorously dependencies (e.g. global MRF).

More details at

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

Thanks
Any questions ?