High-Dimensional Mixture Models for Unsupervised Image Denoising (HDMI)

SIAM Annual Meeting 2017, Pittsburgh, Pennsylvania

Antoine Houdard

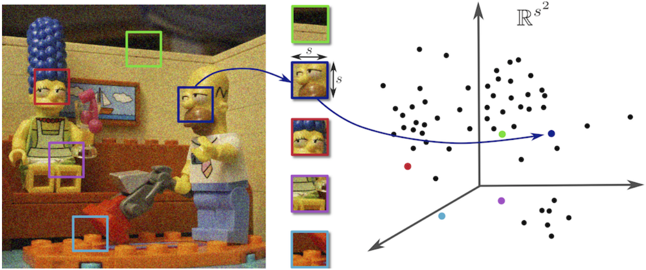
LTCI, Télécom ParisTech
MAP5, Université Paris Descartes

antoine.houdard@telecom-paristech.fr
houdard.wp.imt.fr

Joint work with C. Bouveyron & J. Delon

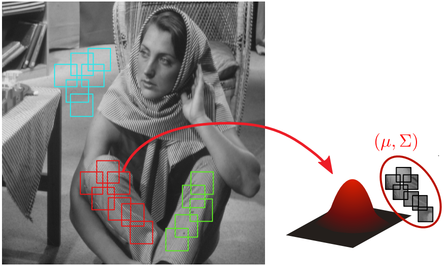
Patch-based image denoising

- most of the denoising methods rely on the description of the image by patches (NL-means, NL-Bayes, S-PLE, LDMM, PLE, BM3D, DA3D)



Patch-based image denoising

- most of the denoising methods rely on the description of the image by patches (NL-means, NL-Bayes, S-PLE, LDMM, PLE, BM3D, DA3D)
- part of them rely on a clustering in the patch space and a statistical model (NL-Bayes, S-PLE, PLE)



The curse of dimensionality

Parameters estimation for Gaussian models or GMMs suffers from the curse of dimensionality

$$\mathbb{R}^{s^2}$$

In the litterature, this issue is worked around by

- the use of small patches in NL-Bayes (5×5)
- a model of mixture with fixed lower dimensions covariances in S-PLE

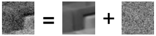
We propose a fully statistical model, that estimates a lower dimension for each group.

Noise model and notations

We denote

- $\{y_1, \ldots, y_n\} \in \mathbf{R}^p$ the (observed) noisy patches of the image;
- $\{x_1, \ldots, x_n\} \in \mathbf{R}^p$ the corresponding (unobserved) clean patches.

We suppose they are realizations of random variables Y and X that follow the classical degradation model:



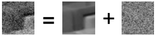
$$Y = X + N \sim \mathcal{N}(0, \sigma^2 \mathbf{I})$$

Noise model and notations

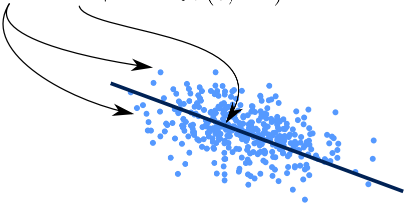
We denote

- $\{y_1, \ldots, y_n\} \in \mathbf{R}^p$ the (observed) noisy patches of the image;
- $\{x_1, \ldots, x_n\} \in \mathbf{R}^p$ the corresponding (unobserved) clean patches.

We suppose they are realizations of random variables Y and X that follow the classical degradation model:



$$Y = X + N \sim \mathcal{N}(0, \sigma^2 \mathbf{I})$$



We design for X the High-Dimensional Mixture Model for Image Denoising (HDMI)

The HDMI model

- Model on the actual patches X. Let Z be the latent random variable indicating the group from which the patch X has been generated. We assume that X lives in a low-dimensional subspace which is specific to its latent group:

$$X_{|Z=k} = U_k T + \mu_k,$$

where U_k is a $p \times d_k$ orthonormal transformation matrix and $T \in \mathbb{R}^{d_k}$ such that

$$T \mid Z = k \sim \mathcal{N}(0, \Lambda_k),$$

with $\Lambda_k = \mathrm{diag}(\lambda_1^k, \ldots, \lambda_{d_k}^k)$.

- Model on the noisy patches. This implies that Y follow

$$p(y) = \sum_{k=1}^{K} \pi_k g\left(y; \mu_k, \Sigma_k\right)$$

where π_k is the mixture proportion for the kth component and $\Sigma_k = U_k \Lambda_k U_k^T + \sigma^2 \mathrm{I}_p$.

The HDMI model

Let $Q_k = [U_k, R_k]$ be a $p \times p$ matrix made from U_k and an orthogonal complementary R_k, then the projection of the covariance matrix $\Delta_k = Q_k \Sigma_k Q_k^t$ has the specific structure:

$$\Delta_k = \left(\begin{array}{c|c} \begin{array}{ccc} a_{k1} & & 0 \\ & \ddots & \\ 0 & & a_{kd} \end{array} & \mathbf{0} \\ \hline \mathbf{0} & \begin{array}{ccc} \sigma^2 & & 0 \\ & \ddots & \\ 0 & & \sigma^2 \end{array} \end{array} \right) \begin{array}{l} \left.\vphantom{\begin{array}{c} a \\ \ddots \\ a \end{array}}\right\} \quad d_k \\ \left.\vphantom{\begin{array}{c} a \\ \ddots \\ a \end{array}}\right\} \quad (p - d_k) \end{array}$$

where $a_{kj} = \lambda_j^k + \sigma^2$ and $a_{kj} > \sigma^2$, for $j = 1, \ldots, d_k$. This model, called hereafter HDMI is fully parametrized with

$$\theta = \{\pi_k, \mu_k, Q_k, a_{kj}, d_k, \sigma; k = 1 \ldots K, j = 1 \ldots d_k\}.$$

The HDMI model

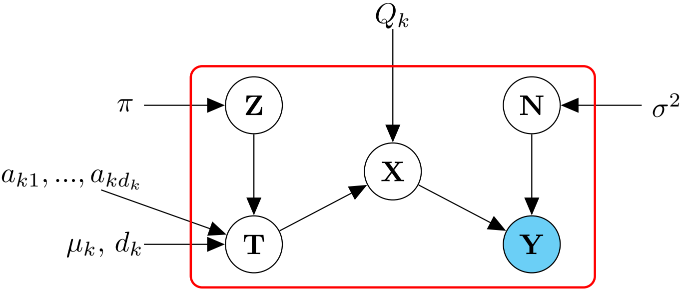


Figure: Graphical representation of the HDMI model.

Denoising with the HDMI model

The HDMI model being known, each patch is denoised with the conditional-expectation

$$\widehat{x_i} = \mathbf{E}[X|Y = y_i],$$

which can be computed as follow:

Proposition.

$$\mathbf{E}[X|Y = y_i] = \sum_{k=1}^{K} \psi_k(y_i) t_{ik},$$

with t_{ik} the posterior probability for the patch y_i to belong in the kth group and

$$\psi_k(y_i) = \mu_k + \widetilde{Q}_k(I_p - \sigma^2 \Delta_k^{-1})\widetilde{Q}_k^T(y_i - \mu_k),$$

with $\widetilde{Q}_k = [U_k, 0_{p,p-d_k}]$.

Model inference

EM algorithm: maximize *w.r.t.* θ the conditional expectation of the complete log-likelihood:

$$\Psi(\theta, \theta^*) \stackrel{\text{def}}{=} \sum_{k=1}^{K} \sum_{i=1}^{n} t_{ik} \log\left(\pi_k g\left(y_i; \theta_k\right)\right),$$

where $t_{ik} = E\left[z = k | y_i, \theta^*\right]$ and θ^* a given set of parameters.

- E-step estimation of t_{ik} knowing the current parameters
- M-step compute maximum likelihood estimators (MLE) for parameters:

$$\widehat{\pi}_k = \frac{n_k}{n}, \quad \widehat{\mu}_k = \frac{1}{n_k} \sum_i t_{ik} y_i, \quad \widehat{S}_k = \frac{1}{n_k} \sum_i t_{ik}(y_i - \mu_k)(y_i - \mu_k)^T,$$

with $n_k = \sum_i t_{ik}$. Then $\widehat{Q}_k$ is formed by the d_k first eigenvectors of $\widehat{S}_k$ and $\widehat{a}_{kj}$ is the jth eigenvalue of $\widehat{S}_k$.

Model inference

The hyper-parameters

The hyper-parameters K and $d_1, \dots, d_K$ cannot be determined by maximizing the log-likelihood since they control the model complexity.

We propose to set K at a given value (in the experiments we use $K = 40$ and $K = 90$) and to choose the intrinsic dimensions d_k:

- using an heuristic that links the d_k with the noise variance σ when known;
- using a model selection tool in order to select the best σ when unknown.

Estimation of intrinsic dimensions

when σ is known

With d_k begin fixed, the MLE for the noise variance in the kth group is

$$\widehat{\sigma}^2_{|k} = \frac{1}{p - d_k} \sum_{j=d_k+1}^{p} \widehat{a}_{kj}.$$

When the noise variance σ is known, this gives us the following heuristic:

Heuristic. Given a value of σ^2 and for $k = 1, ..., K$, we estimate the dimension d_k by

$$\widehat{d_k} = \mathrm{argmin}_d \left| \frac{1}{p - d} \sum_{j=d+1}^{p} \widehat{a}_{kj} - \sigma^2 \right|.$$

Estimation of intrinsic dimensions

when σ is unknown

Each value of σ yields a different model, we propose to select the one with the better BIC (Bayesian Information Criterion)

$$\mathrm{BIC}(\mathcal{M}) = \ell(\hat{\theta}) - \frac{\xi(\mathcal{M})}{2}\log(n),$$

where $\xi(\mathcal{M})$ is the complexity of the model.

why BIC is well-adapted for the selection of σ?

- if σ is too small, the likelihood is good but the complexity explodes;
- if σ is too high, the complexity is low but the likelihood is bad.

Estimation of intrinsic dimensions

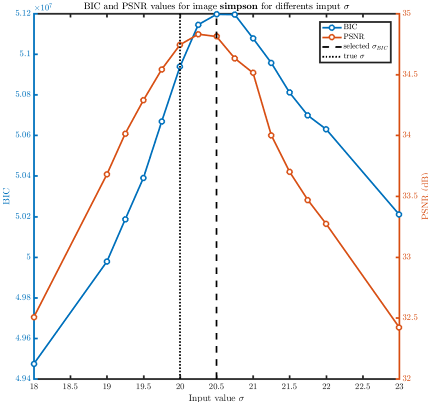
when σ is unknown

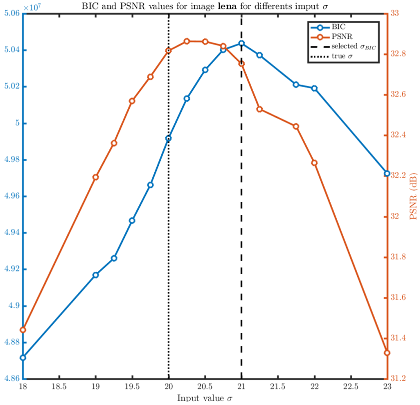
$$\Delta_k = \left(\begin{array}{c|c} \boxed{\begin{matrix} a_{k1} & & 0 \\ & \ddots & \\ 0 & & a_{kd} \end{matrix}} & \mathbf{0} \\ \hline \mathbf{0} & \boxed{\begin{matrix} \sigma^2 & & 0 \\ & \ddots & \\ 0 & & \sigma^2 \end{matrix}} \end{array}\right) \begin{array}{l} \left.\vphantom{\begin{matrix} a \\ \ddots \\ a \end{matrix}}\right\} \quad d_k \\ \left.\vphantom{\begin{matrix} a \\ \ddots \\ a \end{matrix}}\right\} \quad (p - d_k) \end{array}$$

why BIC is well-adapted for the selection of σ?

- if σ is too small, the likelihood is good but the complexity explodes;
- if σ is too high, the complexity is low but the likelihood is bad.

Experiment: selection of σ with BIC

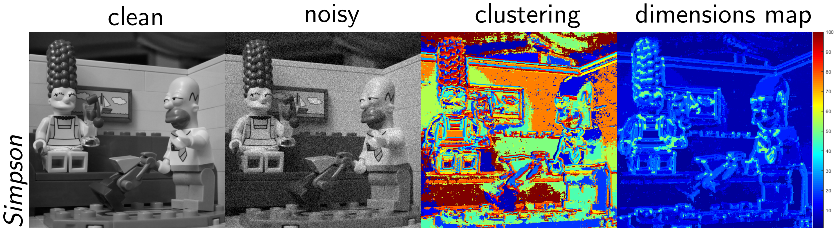




Numerical experiments

Visualization of the intrinsic dimensions

We display for each pixel the dimension of the most probable group of the patch around it.



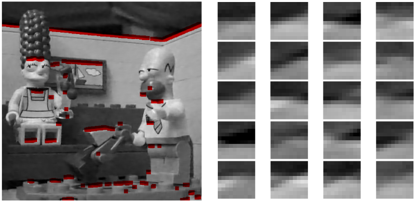
Numerical experiments

Visualization of the intrinsic dimensions

Numerical experiments

What kind of structure is encoded by the model?

group of dimension 13



group of dimension 61

group of dimension 0

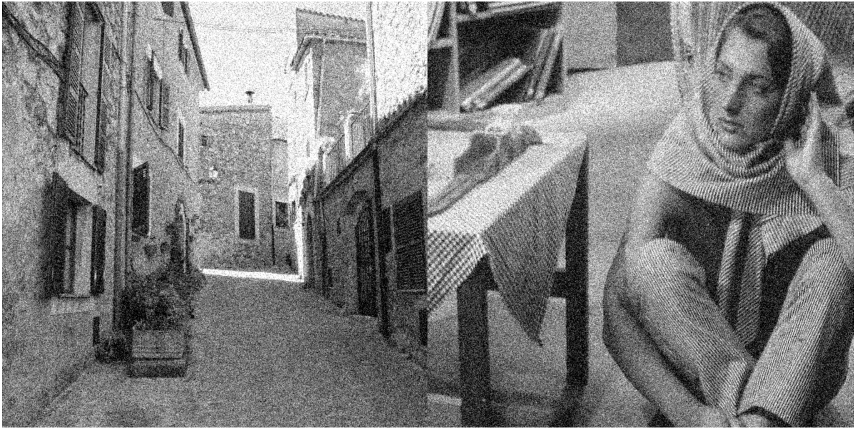
group of dimension 13

Numerical experiments

Results - clean images

Numerical experiments

Results - noisy images $\sigma = 30$



Numerical experiments

Results - denoised with NL-bayes

Numerical experiments

Results - denoised with HDMI$_{sup}$ $K = 90$

Numerical experiments

Results - denoised with S-PLE

Benchmark results

		Supervised denoising					Unsupervised denoising	
Image	σ	NLBayes		S-PLE	HDMI_{sup}		HDMI_{unsup}	
		original	*no flat*		$K = 40$	$K = 90$	$K = 40$	$K = 90$
	10	35.79	35.51	35.50	35.78	**35.83**	35.59	35.23
Lena	20	32.86	32.33	32.58	32.82	**32.90**	32.75	**32.87**
	30	**31.19**	30.42	30.75	30.99	31.04	30.94	30.93
	10	34.91	34.77	34.21	34.77	**35.01**	34.71	34.67
Barbara	20	31.51	31.25	30.67	31.32	**31.61**	31.11	31.31
	30	**29.62**	29.15	28.47	29.31	29.49	29.10	28.92
	10	38.67	37.49	38.37	**38.80**	**38.98**	**38.89**	**39.07**
Simpson	20	34.65	33.42	34.21	**34.74**	**34.91**	**34.81**	**34.79**
	30	32.21	30.59	31.44	**32.33**	**32.50**	32.19	**32.40**
	10	32.45	32.37	32.14	32.40	**32.47**	31.95	**31.94**
Alley	20	28.90	28.73	28.57	**29.03**	**29.07**	**28.89**	**28.96**
	30	26.89	26.65	26.61	**27.31**	**27.39**	**27.19**	**27.17**
	10	**34.07**	33.97	33.76	33.85	33.91	33.59	33.49
Man	20	**30.63**	30.44	30.31	30.44	30.47	30.32	30.23
	30	**28.81**	28.56	28.47	28.65	28.71	28.58	28.56

Conclusion and further work

We presented the HDMI model for image denoising

- which models the full process of the generation of the noisy patches;
- can be used in a "blind" way with the *unsup* version
- reaches state-of-the-art performances in both cases (*sup* and *unsup*)

Some issues and further work

- high computation time: about 12min on a 512×512 image $\rightarrow$ learn the model on a subsample of the patches
- in the case of high σ some miss-classification can yield artifacts $\rightarrow$ explore other initializations?
- slight low-frequency noise in flat areas $\rightarrow$ explore aggregation methods (weighted, EPLL)?

Preprint available at: up5.fr/HDMI

Thank you for your attention!

Any question?

Preprint available at: up5.fr/HDMI