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

Antoine Houdard, Charles Bouveyron, Julie Delon

HIGH-DIMENSIONAL MIXTURE MODELS FOR UNSUPERVISED IMAGE DENOISING (HDMI)

ANTOINE HOUDARD & CHARLES BOUVEYRON & JULIE DELON

Abstract. This work addresses the problem of patch-based image denoising through the unsupervised learning of a probabilistic high-dimensional mixture models on the noisy patches. The model, named hereafter HDMI, proposes a full modeling of the process that is supposed to have generated the noisy patches. To overcome the potential estimation problems due to the high dimension of the patches, the HDMI model adopts a parsimonious modeling which assumes that the data live in group-specific subspaces of low dimensionalities. This parsimonious modeling allows in turn to get a numerically stable computation of the conditional expectation of the image which is applied for denoising. The use of such a model also permits to rely on model selection tools, such as BIC, to automatically determine the intrinsic dimensions of the subspaces and the variance of the noise. This yields a blind denoising algorithm that demonstrates state-of-the-art performance, both when the noise level is known and unknown.

1. Introduction

In the last decade, patch-based models have created a new paradigm in image processing, leading to significant improvements both for classical image restoration problems (denoising [8, 12, 25], *inpainting* [10, 30, 38], interpolation [41]) or for image synthesis [16, 23] and editing [4, 18, 19]. Among these problems, image denoising, which amounts to estimate an image from an observation degraded by additive noise, is probably the one that has received the most attention over the past twenty years. Inspired by the success of patch-based texture synthesis [16] and inpainting [11], the first non local denoising algorithms emerge in 2004 with the discrete universal denoiser (DUDE) for binary images [31, 37], the UINTA filters [3] and the now classical non-local means (NLMeans) [8]. Relying on the assumption that similar patches can be seen as independent realizations of the same distribution, the central idea of these approaches is to average these repeated structures to reduce noise variance. These non local methods have inspired a considerable body of works ever since, under the form of variants and improvements [21, 22], extensions to other noise models [13, 15], or to more complex inverse problems [2, 9, 32]. In order to overcome the ill-posedness of the denoising problem, most of the state-of-the-art approaches [25, 36, 41, 42] rely on a probabilistic framework, which necessitates good prior distributions on patches. Image patches can be seen as vectors in a high-dimensional space (see fig. 1) and estimating prior distributions in such spaces in practice is difficult. In this paper, we explore the use of parsimonious Gaussian mixture models designed for high-dimensional data for this task. Let us mention that a preliminary and short version of this work has appeared in french in [20].

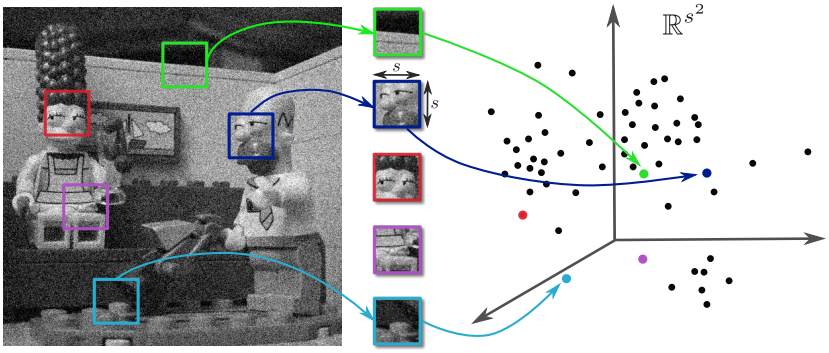


FIGURE 1. Image patches can be seen as vectors in a high-dimensional space. Assuming a Gaussian mixture model for these patches, and because the noise is also Gaussian, we can denoise the patches and hence the original image.

Patch priors for image denoising. The classical denoising model can be expressed in the patch space under the form

$$\mathrm{Y} = \mathrm{X} + \mathrm{N}, \tag{1}$$

where X is the unknown patch before degradation and Y is the observed patch, degraded by some additive noise N. The noise N is usually assumed to follow a Gaussian distribution $\mathcal{N}(0, \sigma^2 \mathrm{I}_p)$, since the Anscombe transform permits to transform the more realistic Poisson noise in a nearly Gaussian noise with fixed variance. While the question of the appropriate statistical prior for the image patches remains essentially open, the most simple and surprisingly effective models used to represent patches distributions are local Gaussian models [25] or mixtures of Gaussians [36, 41, 42]. Under the latter models, the vector X is assumed to follow a distribution

$$p(x) = \sum_{k=1}^{K} \pi_k \mathcal{N}(x; \mu_k, \Phi_k), \tag{2}$$

with μ_k and Φ_k the mean and covariance of the group k, and π_k is the probability that X has been drawn from the group k (with $\sum_{k=1}^{K} \pi_k = 1$). Assuming such a known prior on X, and because the noise is also Gaussian and independent from X, it is quite easy to derive the estimator minimizing the expected mean square error (MSE) to the patch X. This estimator, given by the conditional expectation $\mathbb{E}[\mathrm{X}|\mathrm{Y}]$, takes the form of a (non linear) combination of K linear filters:

$$\mathbb{E}[\mathrm{X}|\mathrm{Y}] = \sum_{k=1}^{K} \psi_k(\mathrm{Y}) \tau_k(\mathrm{Y}), \tag{3}$$

where $\tau_k(\mathrm{Y})$ denotes the probability that, knowing Y, X comes from the group k, and ψ_k the fixed filter

$$\psi_k(y) = \mu_k + \Phi_k(\Phi_k + \sigma^2 \mathrm{I}_p)^{-1}(y - \mu_k).$$

The mixture model being known, each image patch can be denoised by this filter.

Estimating the parameters of this Gaussian mixture model (GMM) from patches is a complex task in practice. Indeed, since the patch sizes are typically greater than 3×3, the dimensions of the corresponding patch spaces can be quite high and estimation in such high-dimensional spaces is not trivial. In the denoising literature, such Gaussian mixture models can be learned from the image itself or from a basis of natural image patches and possibly adapted to each image [36, 41, 42]. This learning stage is made more difficult when it is applied on the degraded patches. Estimating the mixture model also presents other challenges, such as the choice of the number K of mixture components, the choice of the relevant learning bases, and of the inherent dimensions of each group. While recent approaches [36, 41] of the denoising literature impose a fixed value for K and use covariance matrices with predefined ranks, we explore in this paper ways to learn automatically these different parameters. To this aim, we propose to explore recent model-based clustering approaches that have been specifically developed for high-dimensional data. These approaches have the great advantage of respecting the subspaces and the specific intrinsic dimension of each Gaussian in the mixture. In the following paragraphs, we start by briefly reviewing some key-methods in model-based clustering for high-dimensional data.

Model-based clustering for high-dimensional data. Model-based clustering [17, 29] with Gaussian mixtures is a popular approach which is renowned for its probabilistic foundations and its flexibility. One of the main advantages of this approach is the fact that the obtained partition can be interpreted from a statistical point of view. For a data set of n observations in $\mathbb{R}^p$ that one wants to cluster into K homogeneous groups, model-based clustering assumes that the overall population is a realization of a mixture of K Gaussian distributions. Unfortunately, model-based clustering methods show a disappointing behavior in high-dimensional spaces which is mainly due to the fact that they are significantly over-parametrized. Since the dimension of observed data is usually higher than their intrinsic dimension, it is theoretically possible to reduce the dimension of the original space without loosing any information. For this reason, dimension reduction methods are frequently used in practice to reduce the dimension of the data before the clustering step. Feature extraction methods, such as principal component analysis (PCA), or feature selection methods are very popular. However, dimension reduction techniques usually provide a sub-optimal data representation for the clustering step since they imply an information loss which could have been discriminative. To avoid the drawbacks of dimension reduction, several recent approaches have been proposed to allow model-based methods to efficiently cluster high-dimensional data. Subspace clustering methods are searching to model the data in subspaces of much lower dimension and, thereby, avoid numerical problems and boost clustering capability. The mixture of probabilistic principal component analyzers (MPPCA, [34]) may be considered as the earliest and the most popular subspace clustering method. In a few words, MPPCA assumes that the data live in group-specific subspaces with a common intrinsic dimensionality and that the noise has an isotropic variance. This model has become popular in the past decades due, in particular, to its links with PCA. It is worth noticing that the recent denoising approach [36] make use of this model. The authors of [36] however noticed that the fact that all groups must have the same intrinsic dimension in MPPCA is a limiting factor for image denoising.

They consequently removed this constraint and arbitrally fixed the intrinsic dimensions of the groups to be either 1, $p/2$ or $p-1$. We refer to [6] for a recent review of model-based clustering techniques for high-dimensional data.

Model-based clustering for image denoising and contributions of the paper. As explained before, model-based clustering has already been considered many times in the image denoising literature. However, since Gaussian models on patches are usually over-parameterized, their inference requires huge quantities of samples. This estimation is possible on external patch databases, as done in [42], but it becomes completely ill-posed if we just rely on the patches extracted from an image to be restored. In this latter case, regularization becomes essential. As we have seen in the previous paragraph, a first possibility consists in imposing low rank constraints on the groups. This not only makes the model easier to infer, but also reduces the overall computational complexity. One of the first papers using this approach is [36], but the authors impose fixed dimensions to the groups, which makes little sense in practice. The low rank idea is also used in the very recent [14] to drastically accelerate the computation time of [42]. Another possible regularization approach consists in imposing an hyperprior on the GMM parameters. This is the strategy investigated in [27], which first estimates a full GMM on an external patch database (as in [42]) and uses this full GMM as an hyperprior to estimate a GMM on the noisy image data. In this paper, we aim at a much simpler approach, relying only on the noisy data.

Our contribution in this paper is three-fold. First, we propose a probabilistic Gaussian mixture model for image denoising, called HDMI (High Dimensional Mixture models for Image denoising), inspired by the family of models introduced in [7]. The HDMI model proposes a full modeling of the process that is supposed to have generated the noisy patches and adopts a parsimonious modeling to overcome the potential estimation problems due to the high dimension of the data. The parsimony of the model comes from the assumption that the patches live in group-specific subspaces of low dimensionalities. Conversely to the MPPCA model, the HDMI model allows each subspace to have its own intrinsic dimensionality and, thus, proposes a finer modeling of the clusters. Second, we exhibit an expression of the conditional expectation $\mathbb{E}[\mathrm{X}|\mathrm{Y}]$ which is based on explicit inverses of the group covariance matrices. This results in a numerically stable computation of the denoising rule for a given image. Finally, the use a full probabilistic model for the image denoising problem also permits to rely on the model selection tools to determine in an automatic way the intrinsic dimensions of the subspaces and the variance of the noise. This results in a blind image denoising algorithm, that demonstrates state-of-the-art performances both in situations where the level of noise is assumed to be known or not.

It should be noted that the recent paper [40] builds on the same ideas, and proposes to incorporate low rank constraints in a GMM for compressed sensing and denoising applications. However, in [40], the low-rank assumption (including a noise term) is assumed on the actual (unknown) image X, and inferred from the observation Y. This makes the whole estimation process much more complex than in our approach, since the authors maximize the marginal likelihood with the actual signal marginalized out as a latent variable, while we maximize the classical log-likelihood for the observed signal. In addition, the inference and denoising in their model require the inversion of covariance matrices, while our model permits to infer

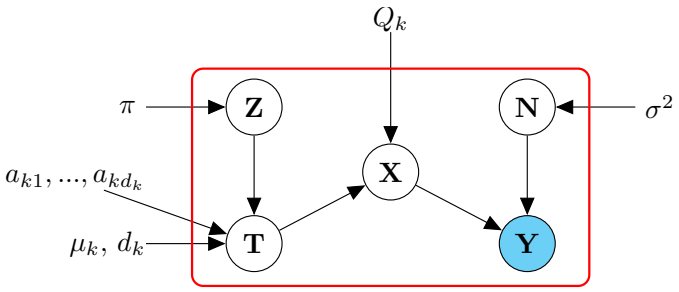


FIGURE 2. Graphical summary of the HDMI model: the circled nodes correspond to random variables whereas other nodes are model parameters; the blue node denotes the observed variable; non-filled variables are latent.

and denoise without matrix inversion. Finally, in HDMI, the intrinsic dimensions of the different groups are inferred (in relation to noise variance) from the early stages of the algorithm and these dimensions evolve during all the stages of the algorithm, whereas in [40], these dimensions are estimated after several iterations of the EM approach on a full GMM model.

Outline of the paper. The paper is organized as follows. In section 2, we present the HDMI model that we introduce to model the generation process of the noisy patches and the associated image denoising rule. Section 3 is devoted to the inference procedure and to model selection, including the estimation of group intrinsic dimensionalities and noise variance. In section 4, we provide numerical experiments that highlight the main features of our approach and demonstrate its effectiveness for image denoising, along with comparisons with the state-of-the-art. Finally, section 6 provides some concluding remarks and tracks for further work.

2. Model-based clustering for image denoising

In this section, we present a parsimonious and flexible statistical model for image denoising. The links with existing models of the literature and the associated denoising procedure are also discussed.

2.1. A parsimonious Gaussian model for image denoising.

Let us consider a data set of n observed noisy patches extracted from an image. These patches are all square sub-images of size $p = s \times s$, extracted from the noisy image and written as vectors $\{y_1, \dots y_n\} \in \mathbb{R}^p$. We assume that these patches are noisy versions of unknown patches $\{x_1, \dots x_n\} \in \mathbb{R}^p$. We consider the unknown patches $\{x_1, \dots x_n\}$ as independent realizations of a random vector $\mathrm{X} \in \mathbb{R}^p$ following a Gaussian mixture model with K groups. We model the unobserved group memberships as realizations of a random variable $\mathrm{Z} \in \{1, ..., K\}$. As pointed out in [36], it is reasonable to assume that most groups in this model should not be full rank, and that each group should have its own dimension. In order to take account of the dimensionality of each group we assume that the random vector X is, conditionally to $\mathrm{Z} = k$, linked to a low-dimensional latent random vector $\mathrm{T} \in \mathbb{R}^{d_k}$, of dimensionality d_k, through a linear transformation of the form:

$$\mathrm{X}_{|\mathrm{Z}=k} = U_k \mathrm{T} + \mu_k, \tag{4}$$

where U_k is a $p \times d_k$ orthonormal transformation matrix and $\mu_k \in \mathbb{R}^p$ is the mean vector of the kth group. The dimension d_k of the latent vector is such that $d_k < p$, $\forall k = 1, ..., K$ (the choice of the intrinsic dimensionalities d_k is discussed in section 3). Besides, the unobserved latent factor T is assumed to be, conditionally on Z, distributed according to a Gaussian density function such as:

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

where $\Lambda_k = \mathrm{diag}(\lambda_{k1}, \ldots, \lambda_{kd_k})$.

Under the degradation model (1) and assuming that the noise variable N is Gaussian with a diagonal covariance matrix $\sigma^2 \mathrm{I}_p$, not depending on the groups:

$$\mathrm{N} \sim \mathcal{N}(0, \sigma^2 \mathrm{I}_p),$$

the conditional distribution of Y is also Gaussian:

$$\mathrm{Y} \mid \mathrm{T}, \mathrm{Z} = k \sim \mathcal{N}(U_k \mathrm{T} + \mu_k, \sigma^2 \mathrm{I}_p). \tag{5}$$

The marginal distribution of Y is therefore a mixture of Gaussians:

$$p(y) = \sum_{k=1}^{K} \pi_k \mathcal{N}(y; \mu_k, \Sigma_k)$$

where π_k is the mixture proportion for the kth component and Σ_k has a specific structure:

$$\Sigma_k = U_k \Lambda_k U_k^t + \sigma^2 \mathrm{I}_p. \tag{6}$$

The specific structure of Σ_k can be exhibited by considering the projected covariance matrix $\Delta_k = Q_k^t \Sigma_k Q_k$, where $Q_k = [U_k, R_k]$ is the $p \times p$ matrix made of U_k and an orthonormal complementary R_k. With these notations, Δ_k has the following form:

$$\Delta_k = \begin{pmatrix} \boxed{\begin{matrix} a_{k1} & & 0 \\ & \ddots & \\ 0 & & a_{kd} \end{matrix}} & & 0 \\ & \boxed{\begin{matrix} \sigma^2 & & 0 \\ & \ddots & \\ 0 & & \sigma^2 \end{matrix}} & \\ 0 & & \end{pmatrix} \begin{matrix} \left.\vphantom{\begin{matrix} a \\ a \\ a \end{matrix}}\right\} d_k \\ \left.\vphantom{\begin{matrix} a \\ a \\ a \end{matrix}}\right\} (p - d_k) \end{matrix}$$

where $a_{kj} = \lambda_{kj} + \sigma^2$ and $a_{kj} > \sigma^2$, for $k = 1, \ldots, K$ and $j = 1, ..., d_k$. The model is therefore fully parametrized by the set of parameters $\theta = \{\pi_k, \mu_k, Q_k, a_{kj}, \sigma^2, d_k;\ k = 1, ..., K,\ j = 1, ..., d_k\}$ and will be referred to as the HDMI model hereafter. Figure 2 presents a graphical representation associated with this model.

2.2. **Links with existing models.** First, it is worth to notice that the model presented above is a specialization of the classical Gaussian mixture model (GMM). Indeed, if $d_k = p$ for $k = 1, ..., K$, then the HDMI model reduces to the usual GMM. Second, it is possible to obtain less or more constrained models than the one presented earlier, corresponding to weaker or stronger regularizations. In particular, it is possible to relax the constraint that the noise variance is common between groups. In this case, the model corresponds to the one presented in [7], and known as $[a_{kj} b_k Q_k d_k]$. From this general model, it is also possible to constrain the dimensions d_k to be common between the groups, which exactly corresponds to the MPPCA model proposed by [34]. Notice that the SPLE denoising approach [36] makes use of

Model	Number of parameters	Asymptotic order	Nb of prms $K = 4$, $d = 10$, $p = 100$
HDDC ($[a_{kj}b_kQ_kd_k]$)	$\rho + \bar{\tau} + 2K + D$	Kpd	4231
HDMI ($[a_{kj}\sigma^2Q_kd_k]$)	$\rho + \bar{\tau} + K + D + 1$	Kpd	4228
MPPCA ($[a_{kj}b_kQ_kd]$)	$\rho + K(\tau + d + 1) + 1$	Kpd	4228
GMM full cov.	$\rho + Kp(p+1)/2$	$Kp^2/2$	20603
GMM common cov.	$\rho + p(p+1)/2$	$p^2/2$	5453
GMM diagonal cov.	$\rho + Kp$	$2Kp$	803

TABLE 1. Properties of the HD-GMM models and some classical Gaussian models: $\rho = Kp + K - 1$ is the number of parameters required for the estimation of means and proportions, $\bar{\tau} = \sum_{k=1}^{K} d_k[p - (d_k + 1)/2]$ and $\tau = d[p - (d + 1)/2]$ are the number of parameters required for the estimation of orientation matrices Q_k, and $D = \sum_{k=1}^{K} d_k$. For asymptotic orders, the assumption that $K \ll d \ll p$ is made.

this latter model. However, the authors noticed that the use of an unique dimension for the groups in MPPCA is a limiting factor for image denoising. In this view, the model that we presented in the previous paragraph should be more appropriate for image restoration problems. Let us finally notice that a family of 28 models was proposed in [5, 7] to accommodate with different practical situations, from the most complex to simple ones. Table 1 provides orders of magnitude for the complexity (*i.e.* the number of parameters to estimate) of the HDMI model as well as some of the models discussed above, in a comparison purpose.

2.3. **Denoising with the HDMI model.** With the assumptions of the HDMI model, the best approximation of the original vector X can be estimated by computing the conditional expectation $\mathbb{E}[\mathrm{X}|\mathrm{Y}]$. Due to the Gaussian mixture distributions, this conditional expectation is a (non linear) combination of linear functions of Y, with weights $\mathbb{P}[\mathrm{Z} = k|\mathrm{Y}]$. These affine functions can be seen as Wiener filters, and require to invert the group covariance matrices. The following proposition gives both the (classical) closed form equation for this conditional expectation, and a second formula which shows how to efficiently compute these filters in the HDMI model case, avoiding numerically sensitive matrix inversions.

Proposition 1. *Assume that the random vector* X *follows the model* (4) *and that* Y *is obtained by the degradation model* (1)*. Then*

$$\mathbb{E}[\mathrm{X}|\mathrm{Y}] = \sum_{k=1}^{K} \psi_k(\mathrm{Y})\tau_k(\mathrm{Y}), \tag{7}$$

with $\tau_k(\mathrm{Y}) = \mathbb{P}[\mathrm{Z} = k|\mathrm{Y}]$ *and*

$$\psi_k(y) = \mu_k + (\Sigma_k - \sigma^2\mathrm{I}_p)\Sigma_k^{-1}(y - \mu_k),$$

where the covariance matrix Σ_k *if defined as in Equation* (6)*. Moreover,* $\psi_k(y)$ *can also be written*

$$\psi_k(y) \quad = \quad \mu_k + \widetilde{Q}_k(\mathrm{I}_p - \sigma^2\Delta_k^{-1})\widetilde{Q}_k^t(y - \mu_k), \tag{8}$$

where $\widetilde{Q}_k = [U_k, 0_{p,p-d_k}]$ *is made of the matrix* U_k *of Equation* (4)*, completed by* $p - d_k$ *zeros columns.*

Proof. If $\mathrm{Z} = k$ is known, then $(\mathrm{X}_{|\mathrm{Z}=k}, \mathrm{N})$ is a Gaussian random vector and so is $(\mathrm{X}_{|\mathrm{Z}=k}, \mathrm{Y}_{|\mathrm{Z}=k})$. The conditional expectation $\mathbb{E}[\mathrm{X} \mid \mathrm{Y}, \mathrm{Z} = k]$ can thus be written

$$\mathbb{E}[\mathrm{X}|\mathrm{Y}, \mathrm{Z} = k] \quad = \quad \mu_k + (\Sigma_k - \sigma^2 \mathrm{I}_p)\Sigma_k^{-1}(\mathrm{Y} - \mu_k) = \psi_k(Y),$$

since Σ_k is the covariance of $\mathrm{Y} \mid \mathrm{Z} = k$ and $\Sigma_k - \sigma^2 \mathrm{I}_p$ the covariance of $(\mathrm{X}_{|\mathrm{Z}=k}, \mathrm{Y}_{|\mathrm{Z}=k})$. Thus, we can write

$$\mathbb{E}[\mathrm{X} \mid \mathrm{Y}, \mathrm{Z}] = \psi_{\mathrm{Z}}(\mathrm{Y}) = \sum_{k=1}^{K} \psi_k(\mathrm{Y}) 1_{\mathrm{Z}=k}.$$

It follows that

$$\begin{aligned}
\mathbb{E}[\mathrm{X}|\mathrm{Y}] &= \mathbb{E}[\mathbb{E}[\mathrm{X} \mid \mathrm{Y}, \mathrm{Z}] \mid \mathrm{Y}] \quad \text{because} \quad \sigma(\mathrm{Z}) \subset \sigma(\mathrm{Z}, \mathrm{Y}) \\
&= \mathbb{E}[\psi_{\mathrm{Z}}(\mathrm{Y}) \mid \mathrm{Y}] = \sum_{k=1}^{K} \mathbb{E}[\psi_k(\mathrm{Y}) 1_{\mathrm{Z}=k} \mid \mathrm{Y}] \\
&= \sum_{k=1}^{K} \psi_k(\mathrm{Y}) \mathbb{E}[1_{\mathrm{Z}=k} \mid \mathrm{Y}] \quad \text{since} \quad \psi_k(\mathrm{Y}) \text{ is } \sigma(\mathrm{Y})\text{-measurable.}
\end{aligned}$$

As a consequence,

$$\mathbb{E}[\mathrm{X}|\mathrm{Y}] = \sum_{k=1}^{K} \psi_k(\mathrm{Y}) \mathbb{E}[1_{\mathrm{Z}=k} \mid \mathrm{Y}] = \sum_{k=1}^{K} \psi_k(\mathrm{Y}) \mathbb{P}[\mathrm{Z} = k|\mathrm{Y}].$$

Now, let $Q_k = [U_k, R_k]$ be the $p \times p$ matrix made of U_k and an orthonormal complementary R_k. The projected covariance matrix $\Delta_k = Q_k^t \Sigma_k Q_k$ is diagonal and can be written

$$\Delta_k = \mathrm{diag}(a_{k1}, \ldots, a_{kd_k}, \sigma^2, \ldots, \sigma^2),$$

where $a_{ki} > \sigma^2 \ \forall i \in \{1, \ldots d_k\}$.

It follows that

$$\begin{aligned}
\psi_k(\mathrm{Y}) &= \mu_k + (\Sigma_k - \sigma^2 \mathrm{I}_p)\Sigma_k^{-1}(\mathrm{Y} - \mu_k) = \mu_k + (\mathrm{I}_p - \sigma^2 \Sigma_k^{-1})(\mathrm{Y} - \mu_k) \\
&= \mathrm{Y} - \sigma^2 \Sigma_k^{-1}(\mathrm{Y} - \mu_k) = \mathrm{Y} - \sigma^2 Q_k \Delta_k^{-1} Q_k^t (\mathrm{Y} - \mu_k).
\end{aligned}$$

We could stop here and be satisfied with this formula which is quite simple and easy to compute. Nevertheless, it is possible to exploit the specific covariance structure of the HDMI model to exhibit a formulation of $\psi_k(\mathrm{Y})$ which is based on explicit inverses of the group covariance matrices. Using the decomposition $Q_k = \widetilde{Q}_k + \overline{Q}_k$, where $\widetilde{Q}_k$ is made of the d_k first columns of Q_k completed by $p - d_k$ zeros columns, we obtain

$$\begin{aligned}
Q_k \Delta_k^{-1} Q_k^t &= \widetilde{Q}_k \Delta_k^{-1} \widetilde{Q}_k^t + \overline{Q}_k \Delta_k^{-1} \overline{Q}_k^t + \widetilde{Q}_k \Delta_k^{-1} \overline{Q}_k^t + \overline{Q}_k \Delta_k^{-1} \widetilde{Q}_k^t \\
&= \widetilde{Q}_k \Delta_k^{-1} \widetilde{Q}_k^t + \overline{Q}_k \Delta_k^{-1} \overline{Q}_k^t + 0 + 0.
\end{aligned}$$

Thus

$$\psi_k(\mathrm{Y}) = \mathrm{Y} - \sigma^2 \left(\widetilde{Q}_k \Delta_k^{-1} \widetilde{Q}_k^t + \overline{Q}_k \Delta_k^{-1} \overline{Q}_k^t \right) (\mathrm{Y} - \mu_k).$$

Now, if we take into account the structure of Δ_k and the fact that the first d_k columns of $\overline{Q}_k$ are composed of zeros, it follows easily that

$$\overline{Q}_k \Delta_k^{-1} \overline{Q}_k^t = \frac{1}{\sigma^2} \overline{Q}_k \overline{Q}_k^t.$$

Since

$$\mathrm{I}_p = Q_k Q_k^t = \widetilde{Q}_k \widetilde{Q}_k^t + \overline{Q}_k \overline{Q}_k^t + \widetilde{Q}_k \overline{Q}_k^t + \overline{Q}_k \widetilde{Q}_k^t = \widetilde{Q}_k \widetilde{Q}_k^t + \overline{Q}_k \overline{Q}_k^t,$$

this yields $\overline{Q}_k \Delta_k^{-1} \overline{Q}_k^t = \frac{1}{\sigma^2}(\mathrm{I}_p - \widetilde{Q}_k \widetilde{Q}_k^t)$. It follows that

$$\begin{aligned} \psi_k(\mathrm{Y}) &= \mathrm{Y} - \sigma^2 \left(\widetilde{Q}_k \Delta_k^{-1} \widetilde{Q}_k^t + \frac{1}{\sigma^2}(\mathrm{I}_p - \widetilde{Q}_k \widetilde{Q}_k^t) \right) (\mathrm{Y} - \mu_k) \\ &= \mathrm{Y} - \left(\mathrm{I}_p + \widetilde{Q}_k (\sigma^2 \Delta_k^{-1} - \mathrm{I}_p) \widetilde{Q}_k^t \right) (\mathrm{Y} - \mu_k). \end{aligned}$$

This allows to conclude. □

At this point, it is interesting to notice that the computation of $\mathbb{E}[\mathrm{X}|\mathrm{Y}]$ usually requires the inversion of the empirical covariances matrices Σ_k. In recent denoising methods such as [25, 41], there is nothing ensuring that these empirical covariances estimate are full rank. To overcome this limitation, the authors of [41] use a standard regularization $\Sigma_k + \varepsilon \mathrm{I}_p$ to ensure invertibility. For the HDMI model, Equation (8) gives explicit and stable inverses of the covariance matrices and consequently an efficient and numerically stable way of denoising the image, without any further regularization.

3. Model inference and model selection

In this section, we discuss the inference procedure and model selection for the HDMI model, including the estimation of the group intrinsic dimensions and the noise variance.

3.1. **Model inference.**

The inference of the HDMI model cannot be done in a straightforward manner by maximizing the likelihood, which is unfortunately intractable. To overcome this problem, the expectation-maximization (EM) algorithm iteratively maximizes the conditional expectation of the complete-data log-likelihood:

$$\mathbb{E}\left[\ell_c\left(\theta; \mathrm{y}, z\right) | \theta^*\right] = \sum_{k=1}^{K} \sum_{i=1}^{n} t_{ik} \log\left(\pi_k p\left(\mathrm{y}_i; \theta_k\right)\right),$$

where $t_{ik} = \mathbb{P}\left[Z = k | \mathrm{y}_i, \theta^*\right]$ and θ^* is a given set of mixture parameters. From an initial solution $\theta^{(0)}$, the EM algorithm alternates two steps: the E-step and the M-step. First, the expectation step (E-step) computes the expectation of the complete log-likelihood $\mathbb{E}\left[\ell_c\left(\theta; \mathrm{y}, z\right) | \theta^{(q)}\right]$ conditionally to the current value of the parameter set $\theta^{(q)}$. Then, the maximization step (M-step) maximizes $\mathbb{E}\left[\ell_c\left(\theta; \mathrm{y}, z\right) | \theta^{(q)}\right]$ over θ to provide an update for the parameter set. This algorithm therefore forms a sequence $\left(\theta^{(q)}\right)_q$ which is guaranteed to converge toward a local optimum of the likelihood [39]. The reader may refer to [28] for further details on the EM algorithm. The two steps of the EM algorithm are iteratively applied until a stopping criterion is satisfied. The stopping criterion may be simply $|\ell(\theta^{(q)}; \mathrm{y}) - \ell(\theta^{(q-1)}; \mathrm{y})| < \varepsilon$ where ε is a positive value to provide. Once the EM algorithm has converged, the partition $\{\hat{z}_1, \ldots, \hat{z}_K\}$ of the data can be deduced from the posterior probabilities

$t_{ik} = \mathbb{P}(\mathrm{Z} = k|\mathrm{y}_i, \hat{\theta})$ by using the *maximum a posteriori* (MAP) rule which assigns the observation y_i to the group with the highest posterior probability.

Proposition 2. *In the particular case of the HDMI model, the update formulas for the M-step of the EM algorithm are as follows:*

- *the proportion π_k and the the mean μ_k of the kth group are respectively estimated by*

$$\hat{\pi}_k = \frac{1}{n}\sum_{i=1}^{n} t_{ik}, \;\; \hat{\mu}_k = \frac{1}{n\hat{\pi}_k}\sum_{i=1}^{n} t_{ik} y_i,$$

- *the d_k first columns of the orientation matrix Q_k are estimated by the eigenvectors associated with the d_k largest eigenvalues of the empirical covariance matrix of the kth group*

$$\mathrm{S}_k = \frac{1}{n\hat{\pi}_k}\sum_{i=1}^{n} t_{ik}(y_i - \hat{\mu}_k)(y_i - \hat{\mu}_k)^t,$$

- *the variance a_{kj} of the data along the jth axis of the subspace of the kth group is estimated by the jth largest eigenvalues $\hat{a}_{kj}$ of S_k, $j = 1, ..., d_k$.*

Proof. Proof of these results is straightforward from the proof of Proposition 4.2.1 in [7]. □

It is worth noticing that these update formulas allow to see the strong link between the HDMI model and the principal component analysis (PCA) method. Indeed, since the d_k first columns of the subspace orientation matrices Q_k are estimated by the eigenvalues of the associated empirical covariance matrices, one can say that the method performs a sort of fuzzy PCA per group, but without loosing any information.

3.2. **Model selection.** The use of the EM algorithm for parameter estimation makes the method almost automatic, except for the estimation of its hyper-parameters: the number K of groups, the group intrinsic dimensionalities d_k and, if unknown, the noise variance σ^2. Indeed, those parameters cannot be determined by maximizing the likelihood since they control the model complexity. However, since the methodology presented here has a sound statistical background, it is possible to rely on model selection tools to select for instance the most appropriate combination of the number K of groups and the dimensionalities d_k. Classical tools for model selection includes the BIC [33] criterion which asymptotically approximates the integrated likelihood. BIC penalizes the log-likelihood $\ell(\hat{\theta})$ as follows, for model $\mathcal{M}$:

$$\mathrm{BIC}(\mathcal{M}) = \ell(\hat{\theta}) - \frac{\xi(\mathcal{M})}{2}\log(n), \tag{9}$$

where $\xi(\mathcal{M})$ is the number of free parameters of the model and n is the number of observations (here the patches). The value of $\xi(\mathcal{M})$ is of course specific to the model considered (*cf.* Table 1 which provides the complexity of the HDMI model). Hence, BIC would allow the user to choose between using the HDMI model in place of the MPPCA model, or using the HDMI model with different intrinsic dimensions. To select the most appropriate configuration for the considered data, the EM algorithm is run for all possible combinations of model parameters, and the one with the highest BIC value is retained. Notice that, all configurations being independent,

Algorithm 1 Intrinsic dimension estimation for a given value of σ^2.

Require: K sets of the p eigenvalues $\lambda_{k1}, ..., \lambda_{kp}$ for each group
Ensure: the dimensions d_k for each k
  for k from 1 to K **do**
    $d_k \leftarrow \text{argmin}_d \, |\text{mean}(\lambda_{kd+1}, \ldots, \lambda_{kp}) - \sigma|$.
  end for

the model selection can be done using parallel computing. Let us finally notice that we do not expect that choosing the number K of groups with BIC, in the specific context of image denoising, would yield the best denoising performance. Indeed, BIC has a modeling objective and it is not aware of the denoising goal: it only aims at selecting the most parsimonious model which best fits the data. We discuss in section 4.1 the influence of K on the denoising performance.

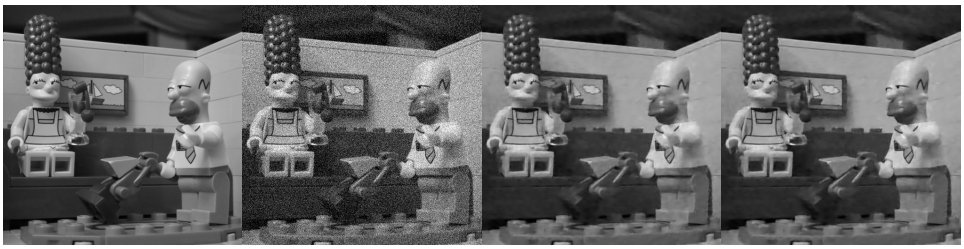
3.3. **Estimation of the intrinsic dimensions d_k.** Regarding the estimation of the intrinsic dimensions d_k, it is unfortunately impossible to test all the K-tuple of dimensions in order to keep the better one in term of BIC. To avoid this drawback, Bouveyron *et al.* proposed in [7] a strategy which avoids the exploration of all possible combinations of dimensions by relying on a unique threshold. The strategy is based on the eigenvalues scree of the covariance matrices Σ_k of the groups. The intrinsic dimension d_k, $k = 1, ..., K$ can be estimated by looking for a break in the eigenvalues scree of Σ_k. For group k the selected dimension is the one for which all subsequent eigenvalues differences are smaller than a threshold τ. The threshold τ is common to all groups and is selected using BIC. However, in the context of image restauration problems, it is expected that some groups have very low intrinsic dimensionalities (uniform zones) whereas other groups have quite large dimensionalities (highly structured zones) and this heuristic can not cover such a range of dimensionalities. To take into account this specific properties of image restauration problems, we propose hereafter two alternatives for the situations where σ^2 is known or not.

Estimation of d_k when σ^2 is known. In the specific context of image denoising, it may be of interest to denoise the image at hand at a specific level of noise. In this case, the variance of the noise is assumed to be known and we propose the heuristic of algorithm 1 to determine the intrinsic dimensions d_k from the known value of σ^2. The idea of this heuristic is, for each group $k = 1, ..., K$, to search the dimensionality d_k such that the mean of the $p - d_k$ smallest eigenvalues of the empirical covariance matrix S_k of the kth group is as close as possible to σ^2. The retained dimensionality $\hat{d}_k$ for the kth group is the solution of the following minimization problem:

$$\hat{d}_k = \text{argmin}_d \left\| \frac{1}{p-d} \sum_{j=d+1}^{p} \lambda_{kj} - \sigma^2 \right\|,$$

where λ_{kj} is the jth largest eigenvalue of the empirical covariance matrix S_k of the kth group.

Estimation of d_k when σ^2 is unknown. In the case where the variance σ^2 of the noise is unknown (unsupervised image denoising), we simply propose to run the above heuristic (algorithm 1) for a range of values for σ^2 and compute the value of



Original Noisy ($\sigma = 30$) HDMI ($K = 40$) HDMI ($K = 90$)

FIGURE 3. Influence of the number K of groups on the denoising with HDMI of the *Simpson* image for $\sigma = 30$ (see text for details).

K	3	5	10	15	20	30	40
PSNR	37.38	37.39	38.19	38.45	38.59	38.72	38.83
K	50	70	100	140	200	400	600
PSNR	38.91	38.97	39.05	**39.07**	39.06	39.01	38.96

TABLE 2. Denoising performance (evaluated through the PSNR) according to the number K of groups in HDMI on the *Simpson* image with $\sigma = 10$.

BIC criterion for the associated model. The retained noise variance $\hat{\sigma}^2$ will be the one which conduces to the highest BIC value.

3.4. **Algorithm.** Algorithm 2 summarizes the different steps of the inference procedure for the HDMI model, for given values of K and σ. Algorithm 3 describes the whole unsupervised denoising procedure using HDMI. Let us notice that the for loop on σ in algorithm 3 can be parallelized since the inferences of HDMI models with different values σ are independent. In the supervised image denoising case (noise standard deviation σ is known), algorithm 3 has to be run with $\sigma_{min} = \sigma_{max} = \sigma$.

4. NUMERICAL EXPERIMENTS

In this section, we provide several numerical experiments to illustrate the characteristics of the HDMI method and its ability to denoise images. The HDMI model is also compared with recent state of the art denoising approaches. Comparison results are provided both under the form of PSNR tables and of visual experiments. For the sake of completeness, let us recall that the PSNR is a way to measure the quality of a restored image $\hat{u}$ in comparison to the original one u. For an image with values between 0 and 255, the PSNR is given by the formula

$$\mathrm{PSNR}(u, \hat{u}) = 10 \log_{10} \frac{255^2 |\Omega|}{\sum_{x \in \Omega} (u(x) - \hat{u}(x))^2},$$

where $|\Omega|$ is the number of pixels in u. All the following experiments are run with patches of size 10×10 (the space dimension is consequently $p = 100$).

4.1. **Influence of the number K of groups.** Let us first focus on the influence of the number K of patch groups on the denoising result. We first consider the denoising of a single 512×512 image, *Simpson*, with the HDMI model for different

Algorithm 2 The HDMI inference algorithm

Require: the noisy patches $\{y_1, \dots, y_n\}$, the number K of groups, the noise variance σ^2.

Ensure: parameter estimates $\{\hat{\mu}_k, \hat{Q}_k, \hat{a}_{kj}, \hat{d}_k;\ k = 1, ..., K,\ j = 1, ..., d_k\}$ and BIC value for the HDMI model.

Initialisation Run the k-means algorithm for K groups on $\{y_1, \dots, y_n\}$.

Set $t_{ik} = 1$ if y_i is in group k and 0 otherwise.

Set $lex \leftarrow -\infty$, $dl \leftarrow \infty$.

while $dl > \epsilon$ **do**

M step Update the estimates for $\theta = \{\pi_k, \mu_k, Q_k, a_{kj}, d_k;\ k = 1, ..., K,\ j = 1, ..., d_k\}$.

$$\hat{\pi}_k = \frac{1}{n}\sum_i t_{ik},\ \hat{\mu}_k = \frac{1}{n\hat{\pi}_k}\sum_{i=1}^{n} t_{ik} y_i,\ (\hat{Q}_k, \hat{\lambda}_k) = \text{eigendec}(S_k).$$

where $S_k = \frac{1}{n\hat{\pi}_k}\sum_{i=1}^{n} t_{ik}(y_i - \hat{\mu}_k)(y_i - \hat{\mu}_k)^t$.

Compute the intrinsic dimension $\hat{d}_k$ thanks to algorithm 1.

Set $\hat{a}_{kj} = \hat{\lambda}_{kj}$ for $j = 1, \dots, d_k$. Set the $p - d_k$ last column of $\hat{Q}_k$ to 0.

E step Compute the probabilities $t_{ik} = P(Z = k | y_i, \hat{\theta})$ as follows

$$t_{ik} = \frac{\hat{\pi}_k p(y_i; \theta_k)}{\sum_{\ell=1}^{K} \hat{\pi}_\ell p(y_i; \theta_\ell)}.$$

Update the likelihood $l = \sum_{i=1}^{n} \log \sum_{k=1}^{K} \pi_k p(y_i; \theta_k)$ and compute the relative error between the two successive likelihoods $dl = |l - lex|/|l|$.

$lex \leftarrow l$.

end while

Compute the BIC $\leftarrow 2l - m\log(n)$, where m is the number of free parameters of the model.

Algorithm 3 The unsupervised HDMI image denoising algorithm.

Require: A noisy grey image u, a patch size s, a range $[\sigma_{min}, \sigma_{max}]$ and a discretization step σ_{step} for the noise standard deviation, a number of groups K.

Ensure: A denoised image $\widehat{u}$.

Patch Extraction Extract all $s \times s$ patches from u, to obtain $\{y_1, \dots, y_n\}$.

Inference and model selection

for σ from σ_{min} to σ_{max} with step σ_{step} **do**

Model inference Run algorithm 2 to obtain $\hat{\theta}_\sigma$ and the corresponding BIC value.

end for

Select the model $\hat{\theta} = \hat{\theta}_\sigma$ with the largest BIC.

Denoising

for $i = 1$ to n **do**

compute $\widehat{y}_i = \sum_{k=1}^{K} \hat{\pi}_k \left(\hat{\mu}_k + \hat{\widetilde{Q}}_k (\mathrm{I}_p - \hat{\sigma}^2 \hat{\Delta}_k^{-1}) \hat{\widetilde{Q}}_k^t (y_i - \hat{\mu}_k) \right)$,

end for

Aggregate all patches $\widehat{y}_i$ to compute $\widehat{u}$.

values of K and for a noise level of $\sigma = 10$. Figure 3 shows the original *Simpson* image, the noisy version with $\sigma = 10$ and two denoising results with HDMI at $K = 40$ and $K = 90$.

Table 2 presents the PSNR values for different values of K. First, it is worth noticing that, even when using extremely few mixture components, the denoising with HDMI is rather satisfying. Indeed, the difference in PSNR between the best result ($K = 140$) and the one with $K = 3$ is only 1.69 dB. This is an information that can be useful if one would be interested in implementing a fast version of HDMI since the computing time is almost linear in the number K of groups. Second, table 2 confirms the expected behavior that using too much patch groups in HDMI deteriorates the denoising performance. Indeed, even though a large number of groups might better represents the diversity of patches in the image, this assertion turns to be false when the number of groups become too large compared with the data size. In this case, the model overfits the data. One can see that for values of K larger than 200, the PSNR slowly decreases and goes back under 39 dB for $K = 600$ groups. Finally, one can observe on table 2 that, for a large range of K, the PSNR has a plateau. Indeed, between $K = 40$ and $K = 200$ the observed PSNR values do not vary more than $0.25dB$ (38.83 – 39.07). This allows us to conclude that the number K of mixture components for HDMI is not a sensitive parameter and that $K = 40$ may be recommended since it realizes a good compromise between efficiency and performance.

Observe that we did not used the BIC criterion to select K. Indeed, this criterion aims at selecting the most parsimonious model which best fits the data and does not take into account the denoising goal. As a summary of these experiments, we simply recommend to use a number K of groups for HDMI equal to 40 for good and fast results, and equal to 90 for optimum results.

4.2. **Role of the intrinsic dimensions d_k.** In this Section, we investigate both the relevance of the clustering provided by the mixture model and the choice of the intrinsic dimension d_k for each group.

The computed mixture model naturally provides a clustering of all image patches. Indeed, once the EM algorithm has converged, each patch y_i of the original image can be associated to the group k with the highest posterior probability t_{ik}. Figure 4 shows the resulting segmentation for several images, degraded with i.i.d. Gaussian noise with $\sigma = 20$ and restored with HDMI for $K = 40$. In this experiment, each color represents a group, and we assign this color to the central pixel of each patch of the group. The clustering is shown on the third column of the Figure, and the respective group dimensions are shown on the fourth column. In these experiments, flat regions seem to be associated with groups of smaller dimensions: the wall in the *Simpson* image, the shoulder of *Lena*, the floor of *Barbara*. Edges of similar orientations also seem to be grouped together and associated to slightly larger group dimensions (see for instance the top of the wall in *Simpson*). This is also the case for some very regular textures, as the one present on the trousers in *Barbara*. Finally, highly textured regions are usually grouped in groups of high dimensions. This is particularly visible on *Man* and *Alley*, which both contain complex textures (the feathers in *Man*, the brick walls in *Alley*).

Note that the ability of the HDMI model to infer automatically the dimension of each group is a real novelty when compared to state of the art algorithms like NL-Bayes [25] or SURE-PLE [36], which use an unrestricted Gaussian model (for

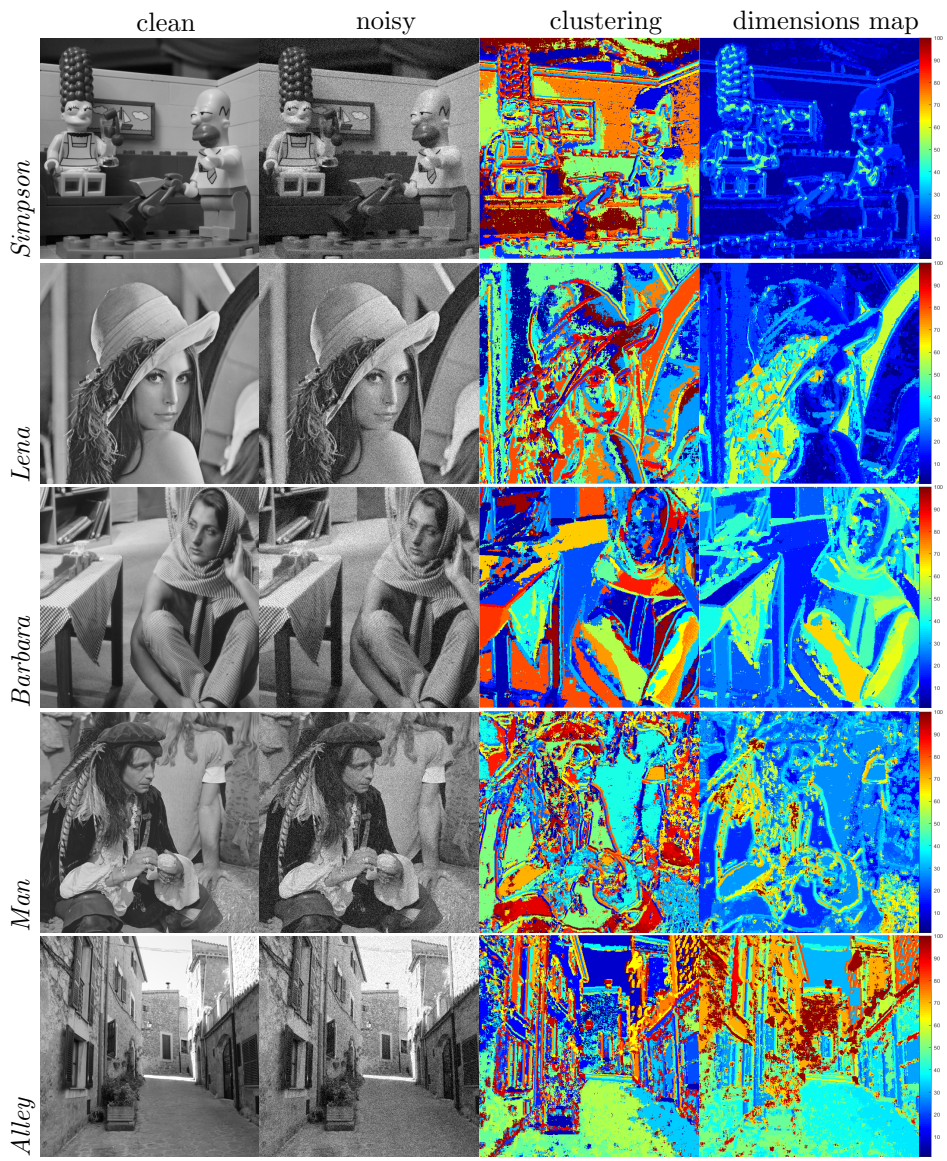


FIGURE 4. From top to bottom, on the left column, the five images *Simpson*, *Lena*, *Barbara*, *Man* and *Alley*. On the second column, the same images degraded with i.i.d. Gaussian noise with $\sigma = 20$. On the third column, the corresponding image segmentation obtained with HDMI for $K = 40$. On the last column, the corresponding maps of intrinsic dimensions for each group.

NLBayes) or a MPPCA with predefined group dimensions (for SURE-PLE), and are forced to detect and treat flat patches separately. Observe also that unlike traditional patch-based methods such as NLmeans [8], which were shown to work

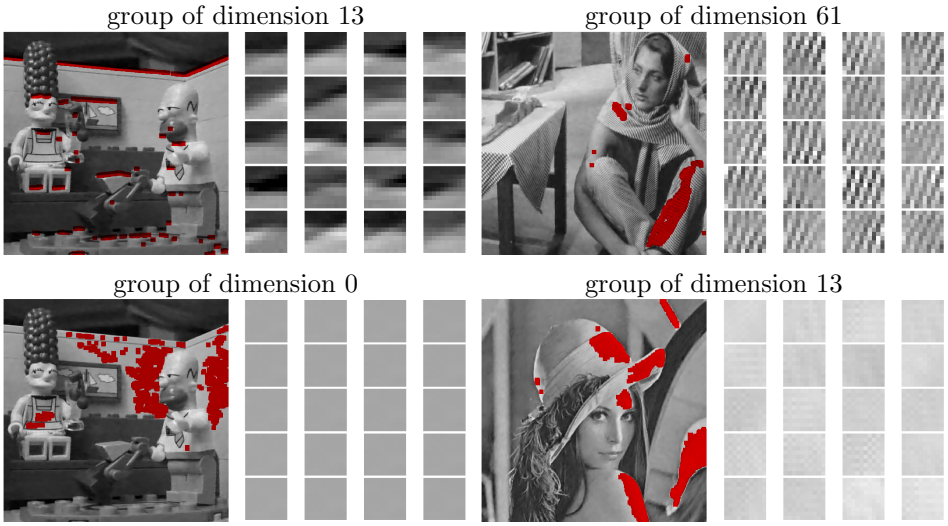


FIGURE 5. Examples of different groups: for each image, we show on the left the patches belonging to the same group k, and on the right 16 patches randomly sampled from the underlying Gaussian model.

better by limiting the search neighborhood for similar patches, each patch is able to collaborate with patches located everywhere in the image.

Figure 5 shows a selection of 4 different groups of various dimensions for the images *Barbara*, *Lena* and *Simpson*. For each group, we also show 16 patches randomly sampled from the group Gaussian model. As expected, the Gaussian model inferred from the top edge wall in *Simpson* generates patches representing more or less horizontal edges. For the group of dimension 61 in *Barbara*, the model generates textured patches which look very similar to the texture present on the trousers. The model of dimension 0 in *Simpson* produces flat patches. Finally, we show a group of dimension 13 on *Lena* which seems to group together flat patches and poorly contrasted but slightly textured ones (from Lena's hat for instance). Unfortunately, this group results in a slightly textured model which is not perfectly adapted to denoise flat regions. When this happens, small artifacts can be introduced in the denoising results. This tends to happen when the chosen number of groups K is too small.

At this point, let us stress out that the intrinsic dimensions d_k act as a regularization for the clustering. Indeed, we might wonder what happens when the EM algorithm is run without dimension reduction, with the reduction applied afterward. In the HDMI model, the dimension reduction is performed from the beginning of the EM algorithm and updated at each iteration, and thus influences the underlying clustering from the E-step. Figure 6 presents two clusterings of the same image, obtained with the same initialization. The first one is obtained by applying a standard GMM model to the patches, and the second one is obtained with the HDMI model. As one can observe on the top of Figure 6, the full GMM clustering turns out to be quite fuzzy and the associated denoising result is not convincing

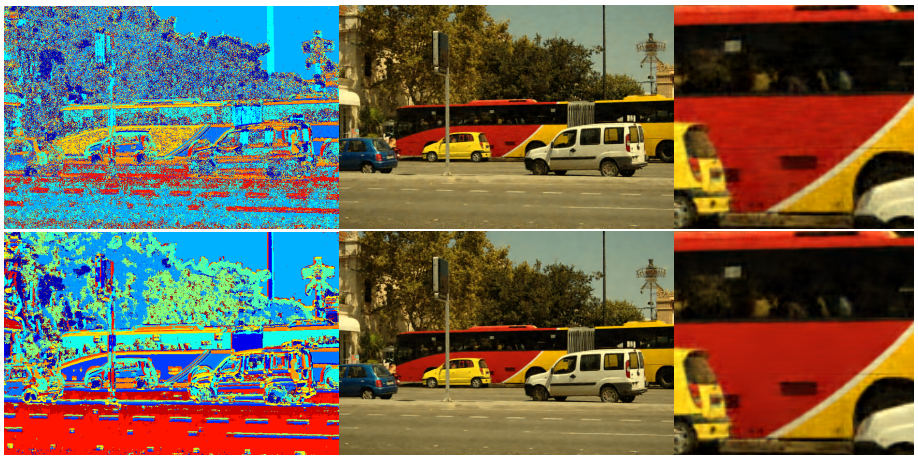


FIGURE 6. First line: Denoising with a full GMM model (50 groups) on all the patches and the HDMI dimension regularization done after the EM algorithm. The clustering (left) is quite fuzzy and the denoising result (middle) is not very good (PSNR: 28.92dB). Second line: Denoising with the HDMI model (50 groups) with intrinsic dimension regularization during the EM process. The clustering (left) is smoother and the denoising yields better results (PSNR: 29.28dB). The noise variance is $\sigma = 30$ and a zoom on the denoising results is proposed in the left column.

(PSNR: 28.92dB). Alternatively, as shown on the bottom of the figure, the HDMI clustering is smoother and the denoising yields better results (PSNR: 29.28dB).

Figure 7 shows the evolution of the intrinsic dimensions during the EM algorithm in HDMI. In this example, for the sake of simplicity, we use only $K = 10$ on the Simpson image. There is a clear stabilization of the intrinsic dimensions at some point in the algorithm. The regularization induced by these smaller dimensions plays a crucial role in the final clustering result.

4.3. **Selection of σ for unsupervised denoising.** In this section, we study how the BIC criterion can be used in order to select the unknown noise standard deviation σ. For unsupervised denoising, we run the HDMI algorithm for different σ_i within a given range of values, and we choose the model with the largest BIC criterion. Figure 8 illustrates the evolutions of the BIC and PSNR when σ_i changes, for the two images *Lena* and *Simpson*, for $\sigma = 10$ and 20. Observe that the form of the BIC curve suggests that the optimal value might be estimated very fast in practice, for instance by dichotomy. In these experiments, the PSNR obtained with the selected model is in practice very close to the best denoising performance (the difference is always smaller than 0.2 dB). Interestingly, the standard deviation estimated by the BIC is always slightly larger than the one used for the synthetic additive noise. This is also confirmed by table 3, which provides the selected σ_i for three different images and three different values of σ. This slight overestimation can be explained by the mere fact that the original images also contain a small amount of intrinsic noise, which seems to be taken into account in the model selection.

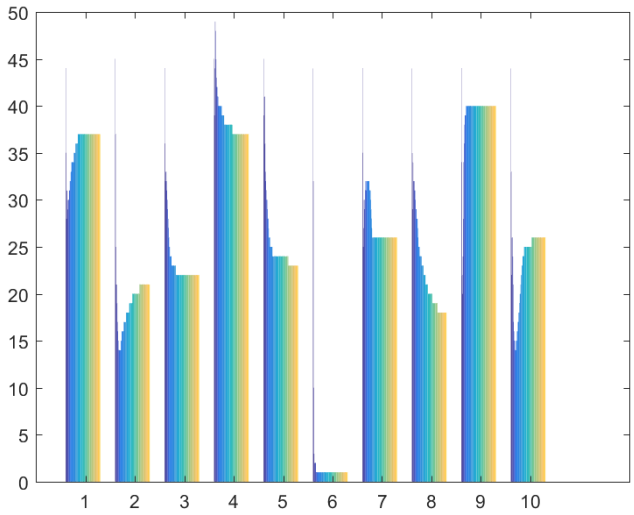


FIGURE 7. Evolution of the dimensions during the iterations of the EM algorithm in HDMI, with a small number of classes ($K = 10$) and 100 iterations. Each group of 100 colored bars represents a class and the 100 bars in each group represent the iterations. The vertical axis represents the dimension.

TABLE 3. Dimensions selection and noise estimation with BIC

Artificial noise std	Estimated noise std		
	Lena	*Simpson*	*Barbara*
10	11	10.5	11
20	21	20.5	21.5
30	31	31.5	31.5

4.4. **Effect of the subsampling on the computing time.** Even though the inference can be parallelized over σ^2 and K, the HDMI algorithm, that we propose in this paper, remains computationally intensive in its unsupervised version (algorithm 3) for large images. Nevertheless, the fact that the HDMI method relies on a sound statistical model allows us to first infer model parameters from a small proportion of the data and to classify afterward the remaining observations to the estimated groups. Indeed, the mixture model fitted by the EM algorithm can be used to compute the posterior probabilities $P(Z = k|y; \hat{\theta})$ for any new observation y.

In order to figure out the potential gain in computing time and the quality of the denoising in a subsampling scenario, we denoise the *Lena* image, degraded with a noise of standard deviation $\sigma = 10$, with the HDMI model fitted from subsamples of the image patches: 1, 2, 5, 10, 20, 30, 50 and 100% of the data. Figure 9 shows the evolution of the PSNR (left panel) and of the computation time (right panel) according to the sampling ratio for the HDMI model with $K = 20$ groups. First,

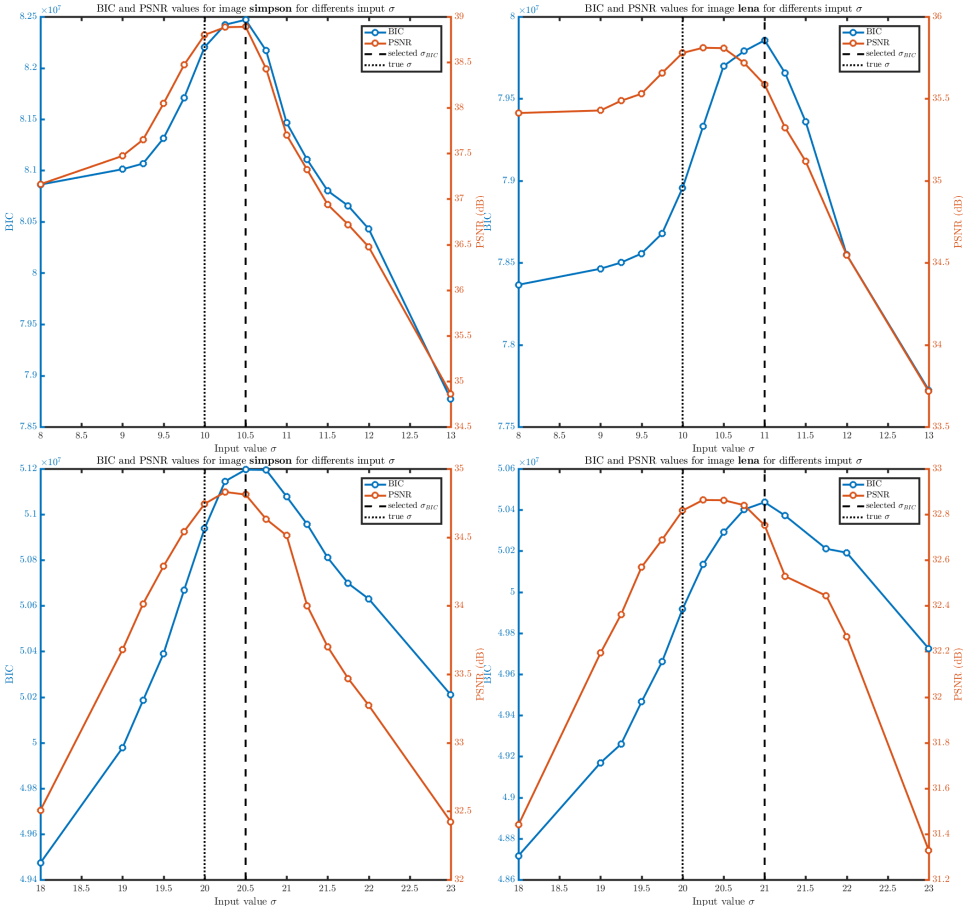


FIGURE 8. **Model selection for unsupervised denoising**. We run the HDMI algorithm for different σ_i within a given range of values. The different curves show the evolution of the BIC and of the PSNR with σ_i. **Top**: image *Simpson*. **Bottom**: image *Lena*. Left column: $\sigma = 10$. Right column: $\sigma = 20$.

the left panel shows that the computing time of the HDMI algorithm is quasi-linear in the number of observations, ranging from less than 10 seconds for 1% of the data to almost 12 minutes for the whole patches. Second, it is worth to notice that even with 1% of the patches the denoising quality is surprisingly good: PSNR of 35.1 dB with 1% whereas the denoising with all patches has a PNSR of 35.8 dB. Finally, as indicated by the vertical dashed lines on both panels of fig. 9, one can notice that there is a relative plateau of the PSNR curve after a sampling ratio of 20%. The denoising result that we obtained with 20% of the patches turns out to be a good compromise between performance and computing time: 0.04 dB less in PSNR than HDMI with 100% of the patches, obtained in 2 minutes instead of 12 minutes for all patches. As a summary, this experiment shows that we can safely run the algorithm on only 20% of the patches to obtain a scalable algorithm on large images without loosing much denoising performance.

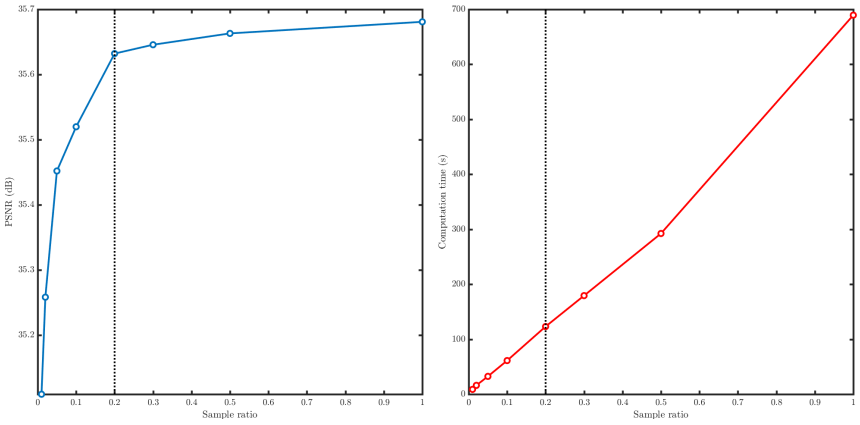


FIGURE 9. Effect of the subsampling on the computing time and the denoising performance with HDMI ($K = 20$) on *Lena* with $\sigma = 10$. *Left:* Evolution of the PSNR versus the sampling size. *Right:* evolution of the computation time versus the same sampling size. The dotted-lines correspond to a subsampling of 20% of the image patches.

4.5. **Influence of the initialization.** As mentioned earlier, the EM algorithm only converges toward a local maximum of the likelihood. This local maximum may therefore depends on the choice of the initialization. In this section, we experiment four different strategies for initializing the HDMI algorithm:

- *Random*: The patches are uniformly assigned to the K groups;
- *Local*: The patches are grouped locally in the image space;
- *K-means*: We run a K-means algorithm on the patches and use it as initialization;
- *K++*: We use the initialization of the K-means++ algorithm.

Figure 10 presents the obtained denoising results for these four initialization strategies. As we can observe, although the final grouping is different, it groups the same kind of structures and the denoising results are quite similar, both visually and in term of PSNRs (standard deviation of 0.01). As a summary, this experiment shows that the choice of the initialization procedure is not discriminant for the purpose of denoising with HDMI.

5. BENCHMARK AND COMPARISON WITH STATE-OF-THE-ART METHODS

We finally focus on the denoising performance of HDMI, and provide a comparison with state-of-the-art approaches. Section 5.1 and Section 5.2 are respectively devoted to grey-scale and color images. In Section 5.3, we propose a more precise discussion about the pros and cons of HDMI.

5.1. **Results for grey-scale images .** Table 4 presents the PSNR results for grey-scale images of HDMI with both known and unknown noise standard deviation σ, for two number of groups $K = 40$ and $K = 90$, and this for the five images (*Lena*, *Barbara*, *Simpson*, *Alley*, *Man*) which have been noised with $\sigma = 10,\ 20,\ 30$. In

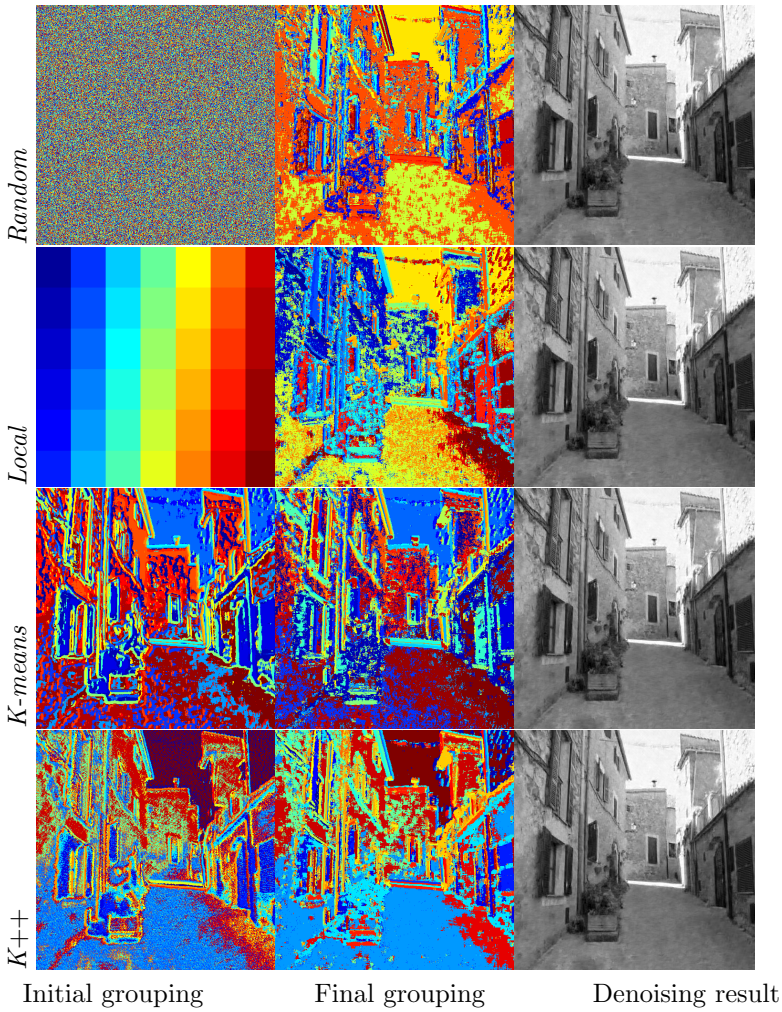


FIGURE 10. Influence of the initialization on the HDMI result. Each line corresponds to a different initialization strategy, on the same noisy image. The left column shows the clustering used to initialize the EM algorithm. The middle column shows the final clustering obtained by the HDMI model. The right column is the corresponding denoising result. *Random*: 27.36dB, *Local*: 27.37dB, *K-means*: 27.35dB, *K++*: 27.37dB.

a comparison purpose, table 4 also provides for these scenarios the results of NL-Bayes [25], with and without the "flat area trick", and the results of SURE-PLE [36], which shares some similarities with HDMI, as explained in the introduction.

As a summary of the comparison, one can first notice that HDMI$_{sup}$ (σ known) outperforms SPLE and NLBayes without the "flat area trick" in almost all the scenarios. It is interesting to notice that removing the constraints on the group intrinsic dimensions of SPLE and estimating them through our proposal allows to clearly improve the denoising. Second, HDMI$_{sup}$ turns out to also compare

TABLE 4. Comparison of HDMI, NL-Bayes [26], SURE-PLE [35] and BM3D [24] for grey-scale images.

		Supervised denoising						Unsupervised	
Image	σ	NL-Bayes		S-PLE	BM3D	HDMI_{sup}		HDMI_{unsup}	
		original	*no flat*			$K = 40$	$K = 90$	$K = 40$	$K = 90$
	10	35.85	35.57	35.34	35.91	35.78	35.83	35.59	35.23
Lena	20	32.90	32.40	32.34	33.00	32.82	32.90	32.75	32.87
	30	31.20	30.49	30.46	31.16	30.99	31.04	30.94	30.93
	10	34.93	34.77	33.89	34.79	34.77	35.01	34.71	34.67
Barbara	20	31.52	31.29	30.37	31.59	31.32	31.61	31.11	31.31
	30	29.72	29.44	28.22	29.61	29.31	29.49	29.10	28.92
	10	38.76	37.59	38.16	38.98	38.80	38.98	38.89	39.07
Simpson	20	34.74	33.72	34.08	35.05	34.74	34.91	34.81	34.79
	30	32.53	31.54	31.53	32.72	32.33	32.50	32.19	32.40
	10	32.53	32.50	32.05	32.46	32.40	32.47	31.95	31.94
Alley	20	29.10	29.07	28.67	29.15	29.03	29.07	28.89	28.96
	30	27.43	27.37	26.92	27.51	27.31	27.39	27.19	27.17
	10	34.14	34.01	33.61	33.99	33.85	33.91	33.59	33.49
Man	20	30.63	30.49	30.15	30.63	30.44	30.47	30.32	30.23
	30	28.81	28.65	28.32	28.89	28.65	28.71	28.58	28.56

equally to NL-bayes and BM3D with an advantage for the two last methods in term of PSNR. Let us finally emphasize that, even if HDMI_{unsup} is not aware of the actual noise level and has to estimate it, HDMI_{unsup} also performs results close to the NLBayes, BM3D and HDMI_{sup} ones (which are supervised methods). This emphasize the efficiency of our approach for blind image denoising.

Figure 11 finally provides a visual comparison of the different approaches on the four different images *Alley*, *Barbara*, *Lena* and *Man* when $\sigma = 30$ (images should be seen at full resolution on the electronic version of the paper). Although the PSNR values are very close, visual results are quite different in practice. While constant regions are better handled by the flat area trick of NL-Bayes and SURE-PLE, some fine geometrical structures (for instance the wall and textures in the *Alley* image) are clearly better preserved by HDMI and oversmoothed by the other methods.

5.2. **Results on color images.** Most recent denoising approaches, when applied to color images, first convert RGB images to a different color space, and then denoise each channel independently. The space conversion is applied to avoid creating color artifacts by applying the denoising independently on each channel. HDMI can easily be applied directly on RGB images, by considering color patches as points in a space of dimension $3 \times p$ ($p = s \times s$ is the patch spatial size). Figure 12 and Table 5 show color denoising results for several images and methods. The HDMI algorithm outperforms state-of-the-art denoising methods in most of these experiments. In practice, on color images, HDMI results often better preserve image details than concurrent methods. However, when the noise variance increases, some low-frequency noise or slight residual textures seem to appear in flat areas. We discuss this issue in the following section.

FIGURE 11. Comparative results on the grey-scale images *Alley*, *Barbara*, *Lena* and *Man* with $\sigma = 30$. For each column, from top to bottom: original image, noisy image, NL-Bayes [25], SURE-PLE [36], HDMI. Images should be seen at full resolution on the electronic version of the paper.

5.3. **Discussion.** In this part, we propose discuss some of the advantages and limitations of our approach. Figure 13 proposes closer views on the denoising results for the color images *Alley*, *Traffic* and *Dice*. The first column of Figure 13 is a zoom on the wires in the top of *Alley*. This really thin structure is difficult to reconstruct from a noisy image, especially when the noise is strong (in this

FIGURE 12. Compared results for the RGB images *Alley, Traffic*, *Dice* and *Flowers*. The S-PLE, NL-bayes and BM3D methods are run with default settings and the HDMI method uses $K = 50$ groups. The noise variance is set to $\sigma = 50$. Images should be seen at full resolution on the electronic version of the paper.

experiment, $\sigma = 30$). In the NL-bayes and S-PLE results, this structure has almost completely vanished, whereas HDMI is able to recover the major part of these wires. A closer view on the house shutters in *Alley* is shown on the second column of Figure 13. The shutters present texture patterns that are partially smoothed by NL-Bayes and S-PLE. In contrast, HDMI seems to restore much more precisely this textured area. Finally, the third column of Figure 13 shows a closer view of the denoising results in the tree area of *Traffic*. In this case, HDMI also appears to yield a more precise restoration than NL-bayes and S-PLE. Now, one could argue that this better structure preservation is done at the expense of a good regularization in flat regions. Indeed, the last column of Figure 13 shows a closer view on a flat part of the *Dice* image and shows that the NL-bayes and the S-PLE methods produce

TABLE 5. Comparison of HDMI_{sup}, NLBayes, SURE-PLE and BM3D for color images. The HDMI algorithm is performed with $K = 50$ and the NL-Bayes, SURE-PLE and BM3D algorithms are run from www.ipol.im with default settings. The PSNRs are averaged on five noise realization and rounded at precision 10^{-2}.

Image	σ	NL-bayes	S-PLE	BM3D	HDMI_{sup}
	10	34.83	34.36	34.82	34.85
	20	31.17	30.71	31.18	31.22
Alley	30	29.14	28.84	29.30	29.37
	40	27.75	27.61	28.04	28.16
	50	27.14	26.74	27.10	27.25
	10	43.20	42.51	43.11	43.69
	20	40.17	39.73	39.98	40.89
Dice	30	37.95	37.95	38.01	39.10
	40	36.14	36.51	36.52	37.58
	50	36.50	35.30	35.19	36.47
	10	39.57	39.19	39.49	40.33
	20	36.14	35.44	35.89	36.87
Flower	30	33.82	33.29	33.74	34.81
	40	32.16	31.78	32.13	33.40
	50	31.89	30.57	30.94	32.25
	10	35.16	34.34	34.54	35.12
	20	31.23	30.56	30.81	31.29
Traffic	30	29.02	28.53	28.83	29.28
	40	27.51	27.17	27.45	27.97
	50	26.85	26.16	26.43	27.03
	10	36.94	36.88	37.46	37.61
	20	34.24	33.98	34.59	34.72
Lena	30	32.50	32.30	32.93	33.13
	40	31.12	31.01	31.70	31.97
	50	30.85	29.97	30.72	31.02

nicer results in this region. In the same vein, observe that HDMI can sometimes create undesired artifacts in flat regions. For example, the first column of Figure 14 presents a closer view on the background of *Barbara*. In this case, the NL-bayes and the S-PLE methods perform better than HDMI which seems to add undesired structure to this flat region. We discuss further this limitation in the following paragraphs.

The usual denoising cuisine. Most really powerful image denoising methods use *tricks* or *hacks* to improve their performances. A striking example is the special treatments reserved to flats regions in NL-bayes and S-PLE. NL-bayes detects flat patches by comparing their standard deviation to the noise standard deviation (multiplied by a constant c close to 1). S-PLE defines a group of dimension 1 that will encode flat patches. In the case of HDMI, a group of flat regions is sometimes merged with a group of weakly contrasted textures, especially when the noise is strong. This result in the introduction of textured artifacts in smooth image areas. To avoid this behaviour, a *flat area trick* can be easily added to HDMI

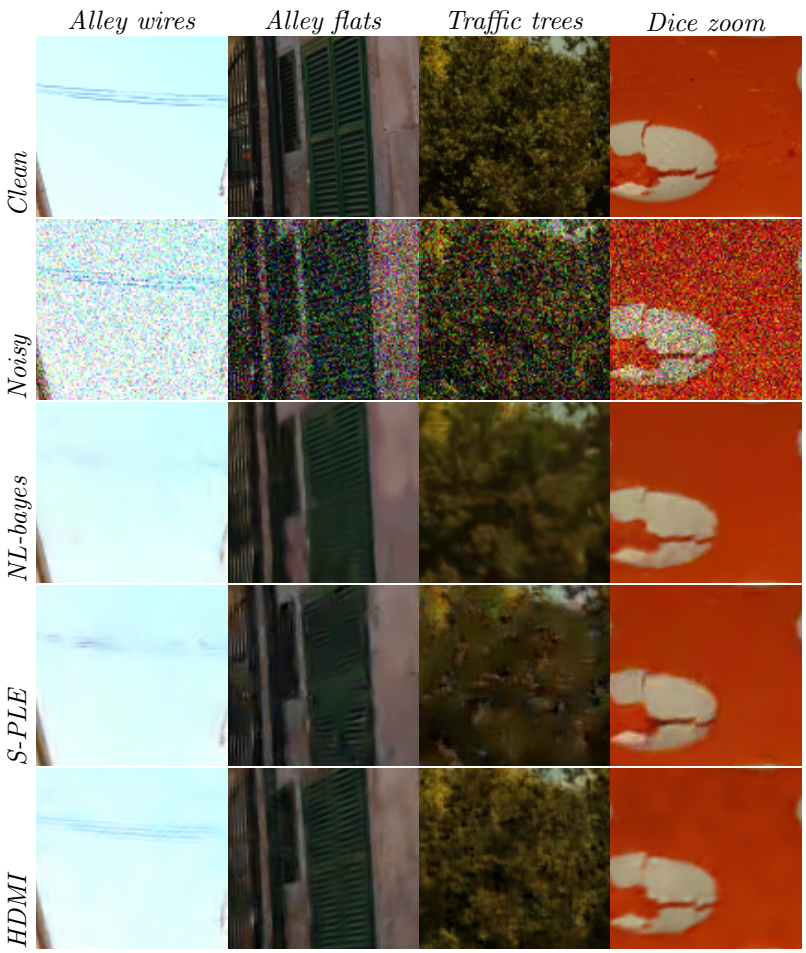


FIGURE 13. Closer views on some details from the RGB images *Alley, Traffic* and *Dice*. The S-PLE and NL-bayes methods are run with default settings and the HDMI method uses $K = 50$ groups. The noise variance is set to $\sigma = 50$.

by replacing patches detected as flat (those patches whose standard deviation is smaller than σ) by a constant patch whose value is the average value of the patch. The center of Figure 14 shows how this simple trick removes most of the annoying residual textures introduced by HDMI. Another explanation for the addition of slight textures in the flat regions is the overestimation of the intrinsic dimensions in the case of the supervised version of HDMI. Indeed, the clean images we use here do contain a small residual noise. The synthetic value σ used for the dimension estimation is thus below the real image noise level. As a consequence, residual noise is treated as structure and is matched with some existing texture in the image. To illustrate this point, the third line of Figure 14 shows the result for HDMI_{unsup}, where the noise variance, and hence the dimensions, are estimated with the BIC criterion. In this case, the slight residual noise is treated as noise and the residual texture issue tends to disappear.

FIGURE 14. Result of HDMI denoising with $K = 40$ groups for the image Barbara with noise $\sigma = 30$. Left HDMI (PSNR = 29.35dB), middle HDMI with the flat area trick (PSNR = 29.36dB), right unsupervised HDMI (PSNR = 29.11dB).

6. Conclusion

In this paper, it is shown that a probabilistic high-dimensional Gaussian mixture model can be learned efficiently on the patches of a noisy image, and used to obtain a blind patch-based denoising. The resulting model HDMI shows state of the art denoising performances, both in the supervised and unsupervised cases. Contrary to previous approaches, this model automatically detects the groups of low dimensionalities within the data. We also provide a numerically stable computation of the conditional expectation for patch denoising, overcoming the traditional limitation encountered in the denoising literature when inverting empirical covariance matrices. We show how to use model selection to automatically estimate the intrinsic dimension of the groups and the noise variance. This work opens several perspectives. The first one concerns the possibility to extend the previous approach to several patch sizes in parallel. Another possible extension is the generalization of the previous model to more general restoration problems. In this case, a nice possibility would be to include hyperpriors in order to stabilize the estimation procedure, as was recently shown in [1].

References

[1] C. Aguerrebere, A. Almansa, J. Delon, Y. Gousseau, and P. Musé, A bayesian hyperprior approach for joint image denoising and interpolation, with an application to hdr imaging, IEEE Transactions on Computational Imaging, (2017).

[2] P. Arias, V. Caselles, and G. Facciolo, Analysis of a variational framework for exemplar-based image inpainting, Multiscale Model. Simul., 10 (2012), pp. 473–514, https://doi.org/10.1137/110848281.

[3] S. Awate and R. Whitaker, Image denoising with unsupervised information-theoretic adaptive filtering, in International Conference on Computer Vision and Pattern Recognition (CVPR 2005), 2004, pp. 44–51.

[4] C. Barnes, E. Shechtman, A. Finkelstein, and D. Goldman, Patchmatch: A randomized correspondence algorithm for structural image editing, ACM Transactions on Graphics-TOG, 28 (2009), p. 24.

[5] L. Bergé, C. Bouveyron, and S. Girard, Hdclassif: An r package for model-based clustering and discriminant analysis of high-dimensional data, Journal of Statistical Software, 46 (2012), pp. 1–29.

[6] C. Bouveyron and C. Brunet-Saumard, Model-based clustering of high-dimensional data: A review, Computational Statistics & Data Analysis, 71 (2014), pp. 52–78.

[7] C. Bouveyron, S. Girard, and C. Schmid, High-dimensional data clustering, Computational Statistics & Data Analysis, 52 (2007), pp. 502–519.

[8] A. Buades, B. Coll, and J. Morel, A review of image denoising algorithms, with a new one, Multiscale Modeling and Simulation, 4 (2006), pp. 490–530.

[9] A. Buades, B. Coll, J.-M. Morel, and C. Sbert, Self-similarity driven color demosaicking, IEEE Trans. Image Process., 18 (2009), pp. 1192–202, https://doi.org/10.1109/TIP.2009.2017171.

[10] A. Criminisi, P. Pérez, and K. Toyama, Region filling and object removal by exemplar-based image inpainting, IEEE Transactions on image processing, 13 (2004), pp. 1200–1212.

[11] A. Criminisi, P. Pérez, and K. Toyama, Region filling and object removal by exemplar-based image inpainting, Image Process. IEEE Trans., 13 (2004), pp. 1200–1212.

[12] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, Image denoising by sparse 3-d transform-domain collaborative filtering, IEEE Transactions on image processing, 16 (2007), pp. 2080–2095.

[13] C.-A. Deledalle, L. Denis, and F. Tupin, Iterative weighted maximum likelihood denoising with probabilistic patch-based weights, IEEE Trans. Image Process., 18 (2009), pp. 2661–72, https://doi.org/10.1109/TIP.2009.2029593, http://www.ncbi.nlm.nih.gov/pubmed/19666338.

[14] C.-A. Deledalle, S. Parameswaran, and T. Q. Nguyen, Image restoration with generalized gaussian mixture model patch priors, arXiv preprint arXiv:1802.01458, (2018).

[15] C.-A. Deledalle, F. Tupin, and L. Denis, Poisson {NL} means: Unsupervised non local means for poisson noise, in 2010 IEEE Int. Conf. Image Process., IEEE International Conference on Image Processing ICIP, 345 E 47TH ST, New York, NY 10017 USA, 2010, IEEE; IEEE Signal Process Soc, IEEE.

[16] A. Efros and T. Leung, Texture synthesis by non-parametric sampling, in Proc. Seventh IEEE Int. Conf. Comput. Vis., vol. 2, 1999, pp. 1033–1038, https://doi.org/10.1109/ICCV.1999.790383, http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=790383.

[17] C. Fraley and A. Raftery, Model-based clustering, discriminant analysis, and density estimation, Journal of the American Statistical Association, 97 (2002), pp. 611–631.

[18] O. Frigo, N. Sabater, J. Delon, and P. Hellier, Split and match: example-based adaptive patch sampling for unsupervised style transfer, in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 553–561.

[19] A. Hertzmann, C. E. Jacobs, N. Oliver, B. Curless, and D. H. Salesin, Image analogies, in Proceedings of the 28th annual conference on Computer graphics and interactive techniques, ACM, 2001, pp. 327–340.

[20] A. Houdard, C. Bouveyron, and J. Delon, Clustering en haute dimensions pour le débruitage d'image, in XXVIième colloque GRETSI.

[21] C. Kervrann and J. Boulanger, Optimal spatial adaptation for patch-based image denoising, IEEE Trans. Image Process., 15 (2006), pp. 2866–2878, https://doi.org/10.1109/TIP.2006.877529.

[22] C. Kervrann and J. Boulanger, Local adaptivity to variable smoothness for exemplar-based image regularization and representation, Int. J. Comput. Vis., 79 (2007), pp. 45–69, https://doi.org/10.1007/s11263-007-0096-2.

[23] V. Kwatra, I. Essa, A. Bobick, and N. Kwatra, Texture optimization for example-based synthesis, ACM Transactions on Graphics (ToG), 24 (2005), pp. 795–802.

[24] M. Lebrun, An Analysis and Implementation of the BM3D Image Denoising Method, Image Processing On Line, 2 (2012), pp. 175–213, https://doi.org/10.5201/ipol.2012.l-bm3d.

[25] M. Lebrun, A. Buades, and J. M. Morel, A Nonlocal Bayesian Image Denoising Algorithm, SIAM J. Imaging Sci., 6 (2013), pp. 1665–1688, https://doi.org/10.1137/120874989.

[26] M. Lebrun, A. Buades, and J.-M. Morel, Implementation of the "non-local bayes" (nl-bayes) image denoising algorithm, Image Processing On Line, 3 (2013), pp. 1–42, https://doi.org/10.5201/ipol.2013.16.

[27] E. Luo, S. H. Chan, and T. Q. Nguyen, Adaptive image denoising by mixture adaptation, IEEE transactions on image processing, 25 (2016), pp. 4489–4503.

[28] G. McLachlan and T. Krishnan, The EM Algorithm and Extensions., Wiley, New York, 1997.

[29] G. McLachlan and D. Peel, Finite mixture models, Wiley-Interscience, 2000.

[30] A. Newson, A. Almansa, M. Fradet, Y. Gousseau, and P. Pérez, Video inpainting of complex scenes, SIAM Journal on Imaging Sciences, 7 (2014), pp. 1993–2019.

[31] E. Ordentlich, G. Seroussi, S. Verdu, M. Weinberger, and T. Weissman, A discrete universal denoiser and its application to binary images, in Image Processing, 2003. ICIP 2003. Proceedings. 2003 International Conference on, vol. 1, IEEE, 2003, pp. I–117.

[32] G. Peyré, S. Bougleux, and L. Cohen, Non-local Regularization of Inverse Problems, in ECCV 2008, Springer-Verlag, Marseille, France, 2008, pp. 57–68, https://doi.org/10.1007/978-3-540-88690-7_5.

[33] G. Schwarz, Estimating the dimension of a model, The annals of statistics, 6 (1978), pp. 461–464.

[34] M. Tipping and C. Bishop, Mixtures of probabilistic principal component analyzers, Neural computation, 11 (1999), pp. 443–482.

[35] Y.-Q. Wang, The Implementation of SURE Guided Piecewise Linear Image Denoising, Image Processing On Line, 3 (2013), pp. 43–67, https://doi.org/10.5201/ipol.2013.52.

[36] Y.-Q. Wang and J.-M. Morel, SURE Guided Gaussian Mixture Image Denoising, SIAM J. Imaging Sci., 6 (2013), pp. 999–1034, https://doi.org/10.1137/120901131, http://epubs.siam.org/doi/abs/10.1137/120901131.

[37] T. Weissman, E. Ordentlich, G. Seroussi, S. Verdú, and M. J. Weinberger, Universal discrete denoising: Known channel, IEEE Transactions on Information Theory, 51 (2005), pp. 5–28.

[38] Y. Wexler, E. Shechtman, and M. Irani, Space-time completion of video, IEEE Transactions on pattern analysis and machine intelligence, 29 (2007).

[39] C. Wu, On the convergence properties of the EM algorithm, The Annals of Statistics, 11 (1983), pp. 95–103.

[40] J. Yang, X. Liao, X. Yuan, P. Llull, D. J. Brady, G. Sapiro, and L. Carin, Compressive sensing by learning a gaussian mixture model from measurements, IEEE Transactions on Image Processing, 24 (2015), pp. 106–119.

[41] G. Yu, G. Sapiro, and S. Mallat, Solving inverse problems with piecewise linear estimators: from gaussian mixture models to structured sparsity, IEEE Trans. Image Process., 21 (2012), pp. 2481–99, https://doi.org/10.1109/TIP.2011.2176743.

[42] D. Zoran and Y. Weiss, From learning models of natural image patches to whole image restoration, in 2011 Int. Conf. Comput. Vis., IEEE, Nov. 2011, pp. 479–486, https://doi.org/10.1109/ICCV.2011.6126278.