Sketched learning for image denoising

Hui Shi, Yann Traonmilin, Jean-François Aujol

▶ **To cite this version:**

Hui Shi, Yann Traonmilin, Jean-François Aujol. Sketched learning for image denoising. The Eighth International Conference on Scale Space and Variational Methods in Computer Vision (SSVM), May 2021, Cabourg, France. pp.281-293, 10.1007/978-3-030-75549-2_23 . hal-03123805v2

HAL Id: hal-03123805
https://hal.science/hal-03123805v2

Submitted on 18 Mar 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Sketched learning for image denoising ⋆

Hui Shi[1]*, Yann Traonmilin[1], and Jean-François Aujol[1]

[1]Univ. Bordeaux, Bordeaux INP, CNRS, IMB, UMR 5251,F-33400 Talence, France.
*hui.shi@u-bordeaux.fr

Abstract. The Expected Patch Log-Likelihood algorithm (EPLL) and its extensions have shown good performances for image denoising. It estimates a Gaussian mixture model (GMM) from a training database of image patches and it uses the GMM as a prior for denoising. In this work, we adapt the *sketching* framework to carry out the compressive estimation of Gaussian mixture models with low rank covariances for image patches. With this method, we estimate models from a compressive representation of the training data with a learning cost that does not depend on the number of items in the database. Our method adds another dimension reduction technique (low-rank modeling of covariances) to the existing sketching methods in order to reduce the dimension of model parameters and to add flexibility to the modeling. We test our model on synthetic data and real large-scale data for patch-based image denoising. We show that we can produce denoising performance close to the models estimated from the original training database, opening the way for the study of denoising strategies using huge patch databases.

Keywords: Image denoising · Sketching · Optimisation · Machine learning.

1 Introduction

In image processing, non-local patch-based models have been producing state-of-the art results for classic image denoising problems [2, 18, 24]. Patch-based methods are also beneficial to other image inverse problems such as superresolution [8, 12], inpainting [7] and deblurring [16]. Among these various non-local methods, the Expected Patch Log-Likelihood algorithm (EPLL) [26] shows very good restoration performances.

The EPLL method uses Gaussian mixture models (GMMs) as a prior model for natural images. In order to maximize the redundancy of structural information to estimate the best possible model parameter, we would want to use a very large training database. However, estimating parameters from a large database can be impractical for classic parameter estimation techniques such as Expectation-Maximization (EM), as their memory consumption and computation time depend on the database size.

⋆ This work was partly funded by ANR project EFFIREG - ANR-20-CE40-0001

Recent works [4, 13, 17] propose a scalable technique to learn model parameters from a compressive representation: a *sketch* of the training data collection. It leverages ideas from compressive sensing [11] and streaming algorithms [6] to compress a large database into a size-fixed representation. Thus, space and time complexity of the algorithm for the estimation of the model no longer depends on the original database size, but only on the size of compressed data and on the dimensionality of the model. Sketching has been used successfully for clustering [4] and GMM estimation with diagonal covariances [17] using the greedy Continuous Orthogonal Matching Pursuit (COMP) algorithm. Sketching produces accurate estimates while requiring fewer memory space and calculations. Sketching also has the advantage to be suitable for distributed computing.

Estimating GMMs on image patches is a complex large-scale learning task. The objective of this paper is to explore the sketching method in this context. In this work, we estimate GMMs priors with non-diagonal covariances which is an extension of previous works. Moreover, for a denoising task it has been shown that the rank of covariance matrices can be reduced [20]. We implement a model using low-rank modeling for GMMs covariances in order to manage the modeling of the image patches in the most possible flexible way.

Contributions: The main contributions of this work are the following.

- We describe how we can learn a GMM prior from a compressed database of patches in the context of image denoising.
- We extend the Continuous Orthogonal Matching Pursuit algorithm to be able to estimate GMM models with non-diagonal and possibly low rank covariances.
- We demonstrate the potential of the approach on real large-scale data (over 4 millions training samples) for the task of patch-based image denoising. We show that we can obtain denoising performances with models trained with the compressed database close to the performance of the denoising with the model obtained with the classical EM algorithm. To the best of our knowledge, this is also the first time that the sketching framework has been applied for such high dimensional GMM (GMM in dimension 25).

Outline: The article is organized as follows. In Section 2, we recall the EPLL framework for image denoising. In Section 3, we describe how sketching can be implemented within the specific setting of patch based denoising and we give an implementation with a low-rank technique for GMM estimation. In Section 4, we provide experimental results both on synthetic data and real images showing that our method has denoising performances close to the EM framework. Finally, we discuss future works in the conclusion.

2 Model estimation and denoising with EPLL

2.1 Denoising with EPLL

Expected Patch Log-Likelihood (EPLL) algorithm is a patch-based image restoration algorithm introduced by Zoran and Weiss [26]. It uses priors learned on

patches extracted from a database of clean images. We consider the problem of recovering an image $u \in \mathbb{R}^N$ with N the number of pixels from a noisy version $v = u + w$, where $w \sim \mathcal{N}(0, \sigma^2 I_N)$ is a white Gaussian noise component. The EPLL framework restores an image u by using the following maximum a posteriori (MAP) estimation:

$$u^* = \arg\min_{u \in \mathbb{R}^N} \frac{P}{2\sigma^2}\|u - v\|^2 - \sum_{i=1}^{N} log(p(\mathcal{P}_i u)) \tag{1}$$

where $\mathcal{P}_i : \mathbb{R}^N \to \mathbb{R}^P$ is the linear operator that extracts a patch with P pixels centered at the position i and $p(.)$ is the density of the prior probability distribution of the patches.

Problem (1) is a large non-convex optimization problem as $p(.)$ is chosen as the density of a GMM prior. It can be extended to generalized Gaussian mixture model (GGMM) [10] for a better performance. In the following we keep the GMM model to simplify the description of the model and we leave the extension to GGMM to future work. In the case of GMM, the denoising can be performed with a simple patch by patch Wiener filter with the denoising parameter β.

$$\hat{u} = (I + \frac{\beta\sigma^2}{P}\sum_{i=1}^{N} \mathcal{P}_i^T \mathcal{P}_i)^{-1}(v + \frac{\beta\sigma^2}{P}\sum_{i=1}^{N} \mathcal{P}_i^T \hat{x}_i) \tag{2}$$

where the

$$\hat{x}_i = (\Sigma_{k_i^*} + \frac{1}{\beta} I_P)^{-1} \Sigma_{k_i^*} \tilde{x}_i \tag{3}$$

are denoised patches estimated from noisy patches $\tilde{x}_i$ which are attributed to a Gaussian prior k_i^*, where k_i^* is the component of the GMM that maximizes the likelihood for the given patch $\tilde{x}_i$, i.e. $k_i^* = \arg\max_{1 \leq k_i \leq K} p(k_i|\tilde{x}_i)$ (see e.g.[10]). Note that these operations are applied a few times with increasing β for best denoising performance.

2.2 EM

A classical technique to estimate the GMM is the Expectation-Maximization (EM) algorithm. It is an iterative algorithm to find estimates of GMM parameters, which carries out at each iteration two steps: the expectation step (E-Step), which creates a function for the expectation of the log-likelihood evaluated using the current estimate for the parameters; and the maximization step (M-Step), which computes parameters maximizing the expected log-likelihood found on the E-Step. These estimated parameters are then used to determine the distribution of the latent variables in the next E-Step.

The EM algorithm's average time complexity is $\mathcal{O}(K^2 n)$ when estimating a K-components model on a database of n elements. Learning parameters using EM technique face computational issues linked to the size of the data and the number of parameters to estimate, which would make the use of (very) large

image patches databases impractical. Moreover, the EM algorithm is not guaranteed to lead us to the global optimum, it typically converges to a local one [1, 25]. It may be arbitrarily poor in high dimensions [9].

3 Compressive GMM learning from large image patches database with sketches

We begin by recalling the sketching method, then we show how to extend previous works to manage the case of GMM prior on image patches.

3.1 Compressive mixture estimation

In the sketching framework [14, 15], a measure $f \in \mathcal{D}$ ($\mathcal{D}$ is the set of probability measures over $\mathbb{R}^d$) is encoded with a linear sketching operator $\mathcal{S} : \mathcal{D} \to \mathbb{C}^m$ into a compressed representation $z \in \mathbb{R}^m$:

$$z = \mathcal{S}f \tag{4}$$

We call z a sketch of f. In practice we only have access to the empirical probability distribution $y = \frac{1}{n}\sum_{i=1}^{n} \delta_{x_i}$ where $\chi = \{x_1, ..., x_n\} \subset \mathbb{R}^d$ is the training database (δ_{x_i} is a unit mass at x_i), which we compress into a sketched database $\tilde{y} = \frac{1}{n}\mathcal{S}\sum_{i=1}^{n} \delta_{x_i}$. The goal of the sketching framework is to recover f from $\tilde{y}$.

For some finite $K \in \mathbb{N}^*$, we define a K-sparse model in $\mathcal{D}$ with the parameters $\Theta = \{\theta_1, ..., \theta_K\}$ and the weights $\alpha = \{\alpha_1, ..., \alpha_K\}$:

$$f_{\Theta,\alpha} = \sum_{k=1}^{K} \alpha_k f_{\theta_k} \tag{5}$$

where $f_{\theta_k} \in \mathcal{D}$ are measures parametrized by θ_k, $\alpha_k \in \mathbb{R}^+$ for all components and $\sum_{k=1}^{K} \alpha_k = 1$. The vector z can then be expressed as

$$z = \mathcal{S}f_{\Theta,\alpha} = \sum_{k=1}^{K} \alpha_k \mathcal{S}f_{\theta_k} \tag{6}$$

The objective of sketched learning algorithms is to minimize the energy between the compressed database and the sketch of the estimation. It corresponds to the traditional parametric optimization Generalized Method of Moments. We estimate the parameters with the following minimization

$$(\hat{\Theta}, \hat{\alpha}) = \underset{\substack{\Theta \in \mathbb{R}^K \\ \alpha \in \mathbb{R}^K, \alpha_k > 0, \sum_{k=1}^{K} \alpha_k = 1}}{\arg\min} \|\mathcal{S}f_{\Theta,\alpha} - \tilde{y}\|_2^2, \tag{7}$$

i.e. our aim is to find the probability distribution (the parameters α, Θ) whose sketch is closest to the empirical sketch $\tilde{y}$. It was shown in [15] that we can theoretically guarantee the success of this estimation with a condition on the

sketch size. In particular, sketching uses the "lower restricted isometry property" (LRIP) for the recovery guarantee. This property, is verified, for GMM with sufficiently separated means and random Fourier sketching with high probability as long as $m \geq O(k^2 d\text{polylog}(k,d))$, i.e. when the size of the sketch essentially depends on the number of parameters k, d (empirical results seem to indicate that for Γ the number of parameters, a database size of the order of Γ is sufficient). The excess risk of the GMM learning task is then controlled by the sum of an empirical error term and a modeling error term. This guarantees that the estimated GMM approximates well the distribution of the data.

In our case, the sketched GMM learning problem reduces to the estimation of the sum of k zero-mean Gaussians with covariances $\Theta = (\Sigma_k)_{k=1}^K$, i.e $f_{\Theta,\alpha} = \sum_{k=1}^K \alpha_k g_{\Sigma_k}$ where g_Σ is the zero mean Gaussian measure with covariance Σ. The mean is not needed in the denoising process and it is removed from the patches before sketching and denoising. In this context, the notion of separation used to prove guarantees in [15] does not hold. We still show empirically that the sketching process is successful without this separation assumption.

Examples on synthetic data illustrate that a different notion of separation might be more suitable, which opens interesting new theoretical questions.

3.2 Design of sketching operator: randomly sampling the characteristic function

In [17], the sketch is a sampling of the characteristic function (*i.e* the Fourier transform of the probability distribution f). The characteristic function ψ_f of a distribution f is defined as:

$$\psi_f(\omega) = \int_{\mathbb{R}^d} e^{-i\omega^T x} df(x) \quad \forall \omega \in \mathbb{R}^d \tag{8}$$

The sketching operator is therefore expressed as:

$$\mathcal{S}f = \frac{1}{\sqrt{m}}[\psi(\omega_1), ..., \psi(\omega_m)]^T \tag{9}$$

where $\Omega = (\omega_1, ..., \omega_m)$ is a set of well chosen frequencies.

In the context of images, given a training set of n centered patches $\chi = \{x_1, ..., x_n\} \subset \mathbb{R}^P$, we define the empirical characteristic function with $\tilde{\psi}(w) = \frac{1}{n}\sum_{i=1}^n e^{-i\omega^T x_i}$. Thus the empirical sketch is:

$$\tilde{y} = \frac{1}{\sqrt{m}}[\tilde{\psi}(\omega_1), ..., \tilde{\psi}(\omega_m)]^T \tag{10}$$

In other words, a sample of the sketched database is a P-dimensional frequency component calculated by averaging over patches (not to be mixed with usual 2D Fourier components of images)

$$\tilde{\psi}(\omega_l) = \frac{1}{n}\sum_{i=1}^n e^{-i\omega_l^T x_i} \tag{11}$$

Thanks to the properties of the Fourier transform of Gaussians, the sketch of a single zero-mean Gaussian component g_Σ is

$$(\mathcal{S}(g_\Sigma))_l = e^{-\frac{1}{2}\omega_l^T \Sigma \omega_l}. \tag{12}$$

The choice of frequencies is essential to the success of sketching. Theoretical estimation results are given with random Gaussian frequencies. In practice we generate a Gaussian profile of the amplitude of the frequency using a small sample of the database and we generate randomly the angle of the frequency [17].

3.3 Extension to low rank covariances

Bayesian MAP theory permits to use degenerate covariances as a denoising prior. As we perform Wiener filtering, this is useful as we can reduce the number of parameters by just truncating the component of noisy patches supported on the lowest eigenvalues of Σ. A Gaussian covariance Σ_k is low-rank if there exists a rank r such that we can write $\Sigma_k = X_k X_k^T$ with X_k a $P \times r$ matrix. Our goal is to estimate covariances Θ^* close to the optimal $\hat{\Theta}$. Remark that:

$$\begin{aligned} \|\mathcal{S}f_{\Theta^*} - \mathcal{S}f_{\hat{\Theta}}\|^2 &= \|\sum_{k=1}^{K} \alpha_k \mathcal{S}(f_{\Sigma_k^*} - f_{\hat{\Sigma}_k})\|^2 \\ &= \sum_{l=1}^{m} e_l^2 \end{aligned} \tag{13}$$

where

$$e_l := \left| \sum_{k=1}^{K} \alpha_k (e^{-\frac{1}{2}\omega_l^T \Sigma_k^* \omega_l} - e^{-\frac{1}{2}\omega_l^T \hat{\Sigma}_k \omega_l}) \right| \tag{14}$$

We have, using the Taylor expansion of the exponential,

$$\begin{aligned} \left| e^{-\frac{1}{2}\omega_l^T \Sigma_k^* \omega_l} - e^{-\frac{1}{2}\omega_l^T \hat{\Sigma}_k \omega_l} \right| &= \left| e^{-\frac{1}{2}\omega_l^T \Sigma_k^* \omega_l} \left(1 - e^{-\frac{1}{2}\omega_l^T (\hat{\Sigma}_k - \Sigma_k^*)\omega_l}\right) \right| \\ &= e^{-\frac{1}{2}\omega_l^T \Sigma_k^* \omega_l} \mathcal{O}(\|\hat{\Sigma}_k - \Sigma_k^*\|_F) \\ &\leq C_{\Theta,\Omega} \|\Sigma_k^* - \hat{\Sigma}_k\|_F. \end{aligned} \tag{15}$$

Close to the minimizer, the energy (7) is close to the weighted sum of the Frobenius distance between covariance matrices.

Following classical ideas in low-rank matrix estimation we parametrize Σ_k by its factors X_k: $\Sigma_k = X_k X_k^T$. This is often referred as the Burer-Monteiro method [3, 5]. Assume that

$$X_k^* \in \underset{X \in \mathbb{R}^{P \times r}}{\arg\min} \|X_k X_k^T - \Sigma_k\|_F^2. \tag{16}$$

A classical result is that $X_k^* X_k^{*,T} = U_k \Lambda_k U_k^T$ with $\Lambda_k = \mathrm{diag}(\lambda_1, ..., \lambda_P)$ and $\lambda_1 \geq \lambda_2 ... \geq \lambda_P$ are the ordered eigenvalues of Σ_k (Eckart and Young theorem). Hence, minimizing the Frobenius distance with a reduced rank recovers

the largest components of Σ_k. Using this qualitative argument, we approximate minimization (7) by

$$(\hat{X},\hat{\alpha}) = \underset{\substack{X_k\in\mathbb{R}^{P\times r},\forall k\\ \alpha\in\mathbb{R}^K,\alpha_k>0,\sum_{k=1}^K\alpha_k=1}}{\arg\min} \frac{1}{\sqrt{m}}\sum_{l=1}^{m}\left|\frac{1}{n}\sum_{i=1}^{n}e^{-i\omega_l^T x_i}-\sum_{k=1}^{K}\alpha_k e^{-\frac{1}{2}\omega_l^T X_k X_k^T\omega_l}\right|^2 \tag{17}$$

where $\hat{X}=(\hat{X}_1,...,\hat{X}_K)$ is the collection of factorized rank reduced covariances.

3.4 An algorithm for patch prior learning from sketch : LR-COMP (Low Rank Continuous Orthogonal Matching Pursuit)

Problem (17) can be solved approximately using the greedy Continuous Orthogonal Matching Pursuit (COMP) algorithm (also called CL-OMP)[17]. We adapt this algorithm in the GMMs context with our low-rank approximation (Alg. 1).

Algorithm 1: LR-COMP: Compressive GMM estimation with low-rank covariances.

Data: Empirical sketch $\tilde{y}$, sketching operator $\mathcal{S}$, sparsity K, number of iterations $T \geq K$
Result: Support Θ, weights α
$\hat{r} \leftarrow \tilde{y}$; $\Theta \leftarrow \emptyset$;
for *t = 1 to T* **do**
  Step 1: Find a X such that: $X \leftarrow \arg\max_X Re\left\langle \frac{\mathcal{S}f_X}{\|\mathcal{S}f_X\|_2}, \hat{r}\right\rangle_2$, init = rand;
  Step 2: $\Theta \leftarrow \Theta \cup \{X\}$;
  Step 3: Enforce sparsity by Hard Thresholding if needed;
  if $|\Theta| > K$ **then**
    $\eta \leftarrow \arg\min_{\eta\geq 0}\left\|\tilde{y}-\sum_{k=1}^{|\Theta|}\eta_k\frac{\mathcal{S}f_{X_k}}{\|\mathcal{S}f_{X_k}\|_2}\right\|_2^2$;
    Select K largest entries $\eta_{i_1},...,\eta_{i_K}$;
    Reduce the support $\Theta \leftarrow \{X_{i_1},...,X_{i_K}\}$;
  Step 4: Project to find weights;
  $\alpha \leftarrow \arg\min_{\alpha\geq 0}\left\|\tilde{y}-\sum_{k=1}^{|\Theta|}\alpha_k\mathcal{S}f_{X_k}\right\|_2^2$;
  Step 5: Perform a gradient descent initialized with current parameters;
  $\Theta,\alpha \leftarrow \arg\min_{\Theta,\alpha\geq 0}\left\|\tilde{y}-\sum_{k=1}^{|\Theta|}\alpha_k\mathcal{S}f_{X_k}\right\|_2^2$, init = (Θ,α);
  Step 6: Update residual: $\hat{r} \leftarrow \tilde{y}-\sum_{k=1}^{|\Theta|}\alpha_k\mathcal{S}f_{X_k}$;
Normalize α such that $\sum_{k=1}^K\alpha_k = 1$.

The main tool for the implementation of Alg. 1 is to compute the gradients necessary to perform the gradients in Steps 1, 4 and 5. We define the vector $v(X)=[Re(\mathcal{S}f_X);Im(\mathcal{S}f_X)]\in\mathbb{R}^{2m}$ with

$$v(X)=\begin{bmatrix}[Re(\frac{1}{\sqrt{m}}\psi_X(\omega_l))]_{l=1,...,m}\\ [Im(\frac{1}{\sqrt{m}}\psi_X(\omega_l))]_{l=1,...,m}\end{bmatrix}=\begin{bmatrix}[\frac{1}{\sqrt{m}}e^{-\frac{1}{2}(\omega_l^T XX^T\omega_l)}]_{l=1,...,m}\\ 0\end{bmatrix} \tag{18}$$

To calculate the gradient, we only need to be able to calculate, for a given vector $y \in \mathbb{R}^{2m}$, the scalar products

$$\langle \nabla_X v(X), y \rangle = -B(v(X)_{1:m} \dot{*} y_{1:m}) \tag{19}$$

where $B \in M_{J,m}(\mathbb{R}), J = P \times r$ is a block matrix with

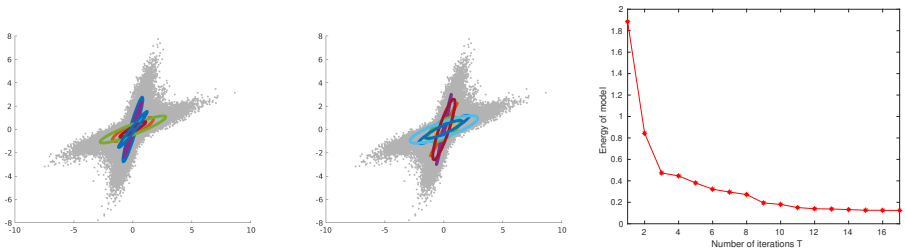
$$B(j,:) = X(:,q)^T W \dot{*} W(s,:), \quad \forall j = (q-1)P + s \tag{20}$$

where $W = [\omega_1, ..., \omega_m] \in M_{P,m}(\mathbb{R})$ the frequency matrix and $\dot{*}$ the multiplication element by element.

4 Results and analysis

4.1 Experiments with synthetic data

We generate data with the following settings: $n = 10^5$ items, dimension $d = 4$, the sparsity level of GMM $K = 8$. The parameters of sketching are: the size of sketch $m = 500$, the rank $r = 2$. We compare the estimation from sketch with the estimation of the GMM with EM algorithm. Fig. 1 shows the reconstruction performance (projected on the first 2 dimensions). We see that we are able to estimate an accurate GMM model from the sketch of the data and the energy (7) of the 2 models are closed. This figure also illustrates that, although Gaussians have zero mean, they have an angular separation instead of a separation of the means (used to give estimation guarantees in [14]). This opens the question of establishing recovery guarantees for zero mean Gaussians using a different notion of separation.



(a) (b) (c)

Fig. 1. Modeling on synthetic data : (a) with EM, (b) with sketching, (c) the model (b) is obtained by minimising energy in (17).

4.2 Results with real images

We extract randomly $n = 4 \times 10^6$ patches of size $P = 5 \times 5$ from the training images of Berkeley Segmentation Database (BSDS) [19]. We show the result of denoising with a prior model estimated with EM (with covariances truncated to have rank r) and with LR-COMP. We use $K = 20$ to demonstrate the capability of our algorithm. Our experiments show that we cannot reduce the rank too

much to keep good denoising performance. Setting a rank $r = 20$ shows no loss of performance (for both EM and LR-COMP). We set $m = 40K(P \times r + 1) \approx 4 \times 10^5$, i.e the compressed database is 250 times smaller than the original patch database. The gains in terms of memory is approximately $\frac{n}{m}$ times compared to the EM approach (most of memory is used to store the frequency matrix). We show results for noise levels $\sigma^2 = 15$ (Fig. 2) and $\sigma^2 = 20$ (Fig. 3). We observe that we obtain similar denoising performances for most images, the worst case being with the "barbara" image which has high contrast and high frequency content. Better results are obtained for the satellite image.

5 Conclusion

In this work, we provide an implementation of the sketching method to estimate a prior model from a compressed database for image denoising. It is shown that a high-dimensional Gaussian mixture model can be learned from a compressed database of patches, and then used for patch-based denoising. We achieve performance close to state-of-the art model based methods.

This work opens several perspectives. We saw that performance is degraded for a particular type of image. One possible explanation is that the sketching (i.e. the choice of frequencies) "missed" these particular images as we used frequencies from previous sketching literature. Adapting this choice to the case of zero mean GMM is still an open question (theoretically and practically). We demonstrated the feasibility of image denoising with sketches. Even if its complexity does not depend on the size of the original database, the LR-COMP algorithm still has computational issues. A possible direction is to extend algorithms proposed in [21, 22] for the estimation of sums of Diracs to the case of GMM with potential performance guarantees [23].

References

1. Balakrishnan, S., Wainwright, M.J., Yu, B., et al.: Statistical guarantees for the EM algorithm: From population to sample-based analysis. The Annals of Statistics **45**(1), 77–120 (2017)
2. Buades, A., Coll, B., Morel, J.M.: A review of image denoising algorithms, with a new one. Multiscale Modeling & Simulation **4**(2), 490–530 (2005)
3. Burer, S., Monteiro, R.D.: Local minima and convergence in low-rank semidefinite programming. Mathematical Programming **103**(3), 427–444 (2005)
4. Chatalic, A., Gribonval, R., Keriven, N.: Large-scale high-dimensional clustering with fast sketching. In: 2018 International Conference on Acoustics, Speech and Signal Processing (ICASSP). pp. 4714–4718. IEEE (2018)
5. Chi, Y., Lu, Y.M., Chen, Y.: Nonconvex optimization meets low-rank matrix factorization: An overview. IEEE Transactions on Signal Processing **67**(20), 5239–5269 (2019)
6. Cormode, G., Muthukrishnan, S.: An improved data stream summary: the count-min sketch and its applications. Journal of Algorithms **55**(1), 58–75 (2005)
7. Criminisi, A., Pérez, P., Toyama, K.: Region filling and object removal by exemplar-based image inpainting. IEEE Transactions on image processing **13**(9), 1200–1212 (2004)

8. Danielyan, A., Foi, A., Katkovnik, V., Egiazarian, K.: Image upsampling via spatially adaptive block-matching filtering. In: 2008 16th European Signal Processing Conference. pp. 1–5. IEEE (2008)
9. Dasgupta, S., Schulman, L.J.: A probabilistic analysis of EM for mixtures of separated, spherical gaussians. Journal of Machine Learning Research **8**, 203–226 (2007)
10. Deledalle, C.A., Parameswaran, S., Nguyen, T.Q.: Image denoising with generalized gaussian mixture model patch priors. SIAM Journal on Imaging Sciences **11**(4), 2568–2609 (2018)
11. Foucart, S., Rauhut, H.: A mathematical introduction to compressive sensing. Springer (2013)
12. Glasner, D., Bagon, S., Irani, M.: Super-resolution from a single image. In: 2009 12th International conference on computer vision. pp. 349–356. IEEE (2009)
13. Gribonval, R., Chatalic, A., Keriven, N., Schellekens, V., Jacques, L., Schniter, P.: Sketching datasets for large-scale learning (long version). arXiv preprint arXiv:2008.01839 (2020)
14. Gribonval, R., Blanchard, G., Keriven, N., Traonmilin, Y.: Compressive statistical learning with random feature moments (2020)
15. Gribonval, R., Blanchard, G., Keriven, N., Traonmilin, Y.: Statistical learning guarantees for compressive clustering and compressive mixture modeling (2020)
16. Katkovnik, V., Egiazarian, K.: Nonlocal image deblurring: Variational formulation with nonlocal collaborative l 0-norm prior. In: 2009 International Workshop on Local and Non-Local Approximation in Image Processing. pp. 46–53. IEEE (2009)
17. Keriven, N., Bourrier, A., Gribonval, R., Pérez, P.: Sketching for large-scale learning of mixture models. Information and Inference: A Journal of the IMA **7**(3), 447–508 (2018)
18. Lebrun, M., Buades, A., Morel, J.M.: A nonlocal bayesian image denoising algorithm. SIAM Journal on Imaging Sciences **6**(3), 1665–1688 (2013)
19. Martin, D., Fowlkes, C., Tal, D., Malik, J.: A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics. In: Proceedings 8th International Conference on Computer Vision. ICCV 2001. vol. 2, pp. 416–423. IEEE (2001)
20. Parameswaran, S., Deledalle, C.A., Denis, L., Nguyen, T.Q.: Accelerating gmm-based patch priors for image restoration: Three ingredients for a 100× speed-up. IEEE Transactions on Image Processing **28**(2), 687–698 (2018)
21. Traonmilin, Y., Aujol, J.F.: The basins of attraction of the global minimizers of the non-convex sparse spike estimation problem. Inverse Problems **36**(4), 045003 (2020)
22. Traonmilin, Y., Aujol, J.F., Leclaire, A.: Projected gradient descent for non-convex sparse spike estimation. IEEE Signal Processing Letters **27**, 1110–1114 (2020)
23. Traonmilin, Y., Aujol, J.F., Leclaire, A.: The basins of attraction of the global minimizers of non-convex inverse problems with low-dimensional models in infinite dimension (2020)
24. Wang, Y.Q., Morel, J.M.: Sure guided gaussian mixture image denoising. SIAM Journal on Imaging Sciences **6**(2), 999–1034 (2013)
25. Wu, C.J.: On the convergence properties of the EM algorithm. The Annals of statistics pp. 95–103 (1983)
26. Zoran, D., Weiss, Y.: From learning models of natural image patches to whole image restoration. In: 2011 International Conference on Computer Vision. pp. 479–486. IEEE (2011)

Fig. 2. Denoising results: (a) original, (b) noisy images with $\sigma^2 = 15$, (c) results with truncated EM model, (d) results LR-COMP model with PSNR/SSIM. Similar denoising performances are obtained with LR-COMP with a compressed database 250 times smaller.

Fig. 3. Denoising results: (a) original, (b) noisy images with $\sigma^2 = 20$, (c) results with truncated EM model, (d) results LR-COMP model with PSNR/SSIM. Similar denoising performances are obtained with LR-COMP with a compressed database 250 times smaller.