2016 ICSEE International Conference on the Science of Electrical Engineering

Gaussian Mixture Diffusion

Jeremias Sulam
Department of Computer Science
Technion, Haifa 32000, Israel
jsulam@cs.technion.ac.il

Yaniv Romano
Department of Electrical Engineering
Technion, Haifa 32000, Israel
yromano@tx.technion.ac.il

Michael Elad
Department of Computer Science
Technion, Haifa 32000, Israel
elad@cs.technion.ac.il

***Abstract*—Most state-of-the-art denoising algorithms employ a patch-based approach by enforcing a local model or prior, such as self similarity, sparse representation, or Gaussian Mixture Model (GMM). While applying these models, these algorithms implicitly build a notion of similarity between the image pixels. This can be formulated as an image-adaptive linear-filter which is then used to denoise or restore the degraded image. In this work we focus on such a filter emerging from the GMM and study its properties.**

Focusing then on the denoising formulation, we incorporate a graph-based regularization term by leveraging the corresponding GMM graph. From a variational interpretation, the resulting algorithm extends and improves the non-local diffusion algorithm by replacing the Non-Local Means kernel with a GMM one. Our results indicate that this approach, termed Gaussian Mixtures Diffusion (GMD), consistently improves over both the original GMM scheme and the non-local diffusion algorithm. Furthermore, GMD is competitive or even better than the state-of-the-art method of EPLL.

I. Introduction

Image restoration is the problem concerned with inverting a degradation model and recovering or estimating the underlying image. In the particular case of image denoising, which will be the main concern of our work, the degradation of the image $\mathbf{u} \in \mathbb{R}^N$ is modeled as

$$\mathbf{f} = \mathbf{u} + \boldsymbol{\eta}, \tag{1}$$

where $\mathbf{f} \in \mathbb{R}^N$ is the noisy image, and $\boldsymbol{\eta} \in \mathbb{R}^N$ is a zero-mean Gaussian noise realization, with standard deviation σ. Most algorithms assume that patches from natural images can be well represented by a specific model, such as sparse representations [1], [2] or Gaussian Mixtures Model (GMM) [3], [4]. Given the noisy measurements $\mathbf{f}$, the image restoration task can be expressed in terms of an optimization problem, minimizing a cost function over the unknown image $\mathbf{u}$:

$$\min_{\mathbf{u}} \frac{1}{2}\|\mathbf{f} - \mathbf{u}\|_2^2 + \lambda \mathcal{R}(\mathbf{u}). \tag{2}$$

From a maximum a posteriori perspective, the first term corresponds to the log-likelihood function while $\mathcal{R}(\mathbf{u})$ enforces the specific model on the unknown image, with parameter λ. This last term acts as the regularizer, promoting *smoothness* or other qualities that – we believe – characterize natural images.

Broadly speaking, often times the denoising process can be decomposed into two stages. The first one involves highly non-linear decisions which enforce quite sophisticated *local* priors on small patches extracted from the image, whereas the second stage accounts for projections and averaging in order to obtain the final *global* image. Interestingly, as pointed out in [5], [6], once the non-linear part is fixed, these algorithms can be formulated as an image-adaptive linear filter, $\mathbf{W}$. This way, the entire denoising process in this framework can be expressed as

$$\hat{\mathbf{u}} = \mathbf{W}\mathbf{f}, \tag{3}$$

where $\hat{\mathbf{u}} \in \mathbb{R}^N$ is the deniosed image, and $\mathbf{W} \in \mathbb{R}^{N \times N}$ is the matrix form of the denoiser (we will describe this operator in more detail in Section II).

On the one hand, Equation (3) states that each denoised pixel in $\hat{\mathbf{u}}$ is the outcome of a weighted average over the noisy image pixels, where the weights are determined by the specific denoising algorithm. On the other hand, this shows that most denoising algorithms implicitly build a notion of similarity between the i-th and j-th image pixels, given by the entry $\mathbf{W}(i,j)$. Based on this observation, the denoiser can be formulated as a weighted graph, where the vertices refer to the image pixels, and the (weighted) edges represent the pixels' similarity. Previous works have addressed the graph formulation (and its properties) of the K-SVD [6], the Non-Local Means, Bilateral and LARK kernels [5]. Yet, despite the popularity of the GMM prior in the image processing community, this analysis has not been addressed for its resulting operator – this will be the first concern of our work.

Recently, the graph formulation has been employed to regularize the denoising process [5]–[7], designing an image-adaptive term $\mathcal{R}(\mathbf{u})$ in Equation (2). This term is usually expressed in terms of the Laplacian operator, defined by $\boldsymbol{\mathcal{L}} = \mathbf{I} - \mathbf{W}$, where $\mathbf{I} \in \mathbb{R}^{N \times N}$ is the identity matrix and $\mathbf{W}$ is induced by different denoisers (from Equation (3)). Broadly speaking, the eigenvectors that correspond to the small eigenvalues of $\boldsymbol{\mathcal{L}}$ encapsulate most of the structure of the underlying signal [8]. As such, one may propose a graph-based regularization term that penalizes those components in $\mathbf{u}$ corresponding to the large eigenvalues of $\boldsymbol{\mathcal{L}}$; e.g., as done in [7], [9],

$$\min_{\mathbf{u}} \frac{1}{2}\|\mathbf{f} - \mathbf{u}\|_2^2 + \frac{\rho}{2}\mathbf{u}^T \boldsymbol{\mathcal{L}} \mathbf{u}. \tag{4}$$

Notice that when moving from Equation (2) to Equation (4), we are constraining ourselves to priors that have a graph-Laplacian interpretation. In these cases, the performance of the

resulting algorithm depends on the choice of $\boldsymbol{\mathcal{L}}$, which is in turn determined by the similarity measure we use to construct the corresponding graph.

The problem in Equation (4) is certainly not the only way to enforce a graph-based regularization when dealing with inverse problems. Recent works have also considered replacing the data-fidelity term by a weighted norm induced by the matrix $\mathbf{W}$, as in [11]. Another alternative, presented in [6], is to enforce the reconstructed image $\mathbf{u}$ to be close to the filtered image $\mathbf{Wf}$. Formally,

$$\mathcal{J}(\mathbf{u}) = \min_{\mathbf{u}} \frac{1}{2}\|\mathbf{Wf} - \mathbf{u}\|_2^2 + \frac{\rho}{2}\mathbf{u}^T\boldsymbol{\mathcal{L}}\mathbf{u}. \tag{5}$$

This formulation generally provides better results than the problem in Equation (4), as it is related to boosting methods. In particular, this is the cost function minimized by the SOS boosting [6].

Interestingly, the minimization of this kind of problems can be interpreted from a variational perspective. The Non-Local Diffusion (NLD) algorithm [7] suggests a non-local generalization of the diffusion framework by employing a functional defined over a set of pixels which are not necessarily near each other. Unlike the conventional diffusion methods, the minimization of the corresponding functional amounts to a diffusion process between pixels that can now be *far apart* by leveraging some notion of similarity or affinity. In particular, the NLD employed the affinity measure induced by the Non Local Means (NLM) kernel [10], defining the distance between pixels as a function of the Euclidean distance between their corresponding patches. The resulting algorithm effectively minimizes the cost function in Equation (4), where the Laplacian is the one corresponding to the NLM operator.

In this work, we explore the algorithm resulting from the problem in Equation (5) in the case of a Laplacian operator induced by the GMM prior. Unlike the work of [6], we provide a detailed analysis of the denoiser resulting from GMM, and employ the formulation in terms of a non-local diffusion process. This way, our work extends and improves the non-local diffusion algorithm of [7] by (1) employing a similarity measure induced by the GMM operator, and (2) considering the cost function in Equation (5) instead of the original problem in (4). As we will show in the experimental section, our proposed Gaussian Mixture Diffusion (GMD) approach outperforms both the initial formulation of the NLD with the NLM kernel [7], and the original GMM algorithm. Interestingly, the GMD is also competitive or even better than the state-of-the-art method of EPLL [4], which builds upon GMM as well.

This paper is organized as follows: In Section II we provide a brief description of the GMM denoising scheme along with the derivation of its equivalent linear filter. Then, we present the properties of the GMM filter. Implementation details of the proposed GMD method are described in Section III, before moving to the experimental results in Section IV. Conclusions are lastly drawn in Section V.

II. Gaussian Mixture Model

GMM is a popular prior for natural image patches, which has been shown to be very effective in several image restoration tasks [3], [4]. This prior models the distribution of patches as the sum of multivariate Gaussians learned from real data. Applying this prior for image denoising accounts to formulating a MAP estimator for each independent patch from the corrupted image. This can be approximated by choosing the Gaussian with the highest conditional weight for each patch, and then applying a plain Wiener filter with the corresponding covariance matrix [4]. Finally, a patch averaging step is applied in order to obtain the final denoised image.

Given K (learned) Gaussian distributions, characterized by their covariance matrices Σ_k, with zero mean[1], denoising each patch $\mathbf{z}_i \in \mathbb{R}^n$ can be formally expressed by the following minimization problem

$$\hat{\mathbf{p}}_i = \arg\min_{\mathbf{p}} \|\mathbf{p} - \mathbf{z}_i\|_2^2 + \sigma^2\mathbf{p}^T\boldsymbol{\Sigma}_{k(i)}^{-1}\mathbf{p}, \tag{6}$$

where $k(i)$ is the index of the chosen Gaussian with highest conditional weight [4] for the i^{th} patch, and $\hat{\mathbf{p}}_i$ is its estimated clean version. This problem has a closed form solution in terms of the Wiener filter, given by

$$\begin{aligned}\hat{\mathbf{p}}_i &= \left(\mathbf{I} + \sigma^2\boldsymbol{\Sigma}_{k(i)}^{-1}\right)^{-1}\mathbf{z}_i \\ &= \mathbf{F}_i\ \mathbf{z}_i.\end{aligned} \tag{7}$$

Next, the denoised patches $\hat{\mathbf{p}}_i$ are merged together by averaging. This is done by minimizing the following cost function

$$\hat{\mathbf{u}} = \arg\min_{\mathbf{u}}\ \mu\|\mathbf{f} - \mathbf{u}\|_2^2 + \sum_{i=1}^{N}\|\hat{\mathbf{p}}_i - \mathbf{R}_i\mathbf{u}\|_2^2, \tag{8}$$

where $\hat{\mathbf{u}} \in \mathbb{R}^N$ is the estimated (denoised) global image, and $\mathbf{R}_i \in \mathbb{R}^{n\times N}$ is a matrix that extracts the i^{th} patch from the image. Following the procedure that described in [6], the closed form solution of Equation (8) is given by

$$\begin{aligned}\hat{\mathbf{u}} &= \left(\mu\mathbf{I} + \sum_i \mathbf{R}_i^T\mathbf{R}_i\right)^{-1}\left(\mu\mathbf{I} + \sum_i \mathbf{R}_i^T\mathbf{F}_i\mathbf{R}_i\right)\mathbf{f} \\ &= \mathbf{W}_{\text{GMM}}\mathbf{f},\end{aligned} \tag{9}$$

where the filters $\mathbf{F}_i$ are the Wiener filters in Equation (7). In the above derivation we have used the fact that $\mathbf{z}_i = \mathbf{R}_i\mathbf{f}$. Notice that this linear operator can be written as a matrix $\mathbf{W}_{\text{GMM}}$, and thus the denoised image is simply expressed as $\hat{\mathbf{u}} = \mathbf{W}_{\text{GMM}}\ \mathbf{f}$.

While the GMM model is a popular choice for image denoising, a formal analysis of $\mathbf{W}_{\text{GMM}}$ has not yet been addressed. In this section, we present the properties of this filter and provide their corresponding proofs in an accompanying appendix. We should note that all these properties are also

[1]For simplicity, we make the common assumption that the image patches have zero-mean.

shared by the KSVD filter [6].

Theorem 1. *Under the assumption of periodic boundary conditions*[2]*, the matrix* $\mathbf{W}_{GMM}$*, defined in Equation* (9)*, has the following properties:*

1) $\mathbf{W}_{GMM} = \mathbf{W}_{GMM}^T$*: it is symmetric.*
2) $\mathbf{W}_{GMM} \succ 0$*: it is positive definite, and has minimal eigenvalue equal to* $\frac{\mu}{\mu+n}$.
3) $\|\mathbf{W}_{GMM}\|_2 \leq 1$*: its spectral radius* ≤ 1.

III. The Proposed Approach

Based on the properties provided in Theorem 1, we can draw interesting conclusions. As it was done for the K-SVD operator matrix in [6], the matrix $\mathbf{W}_{\text{GMM}}$ can be decomposed into a similarity matrix $\mathbf{K}_{\text{GMM}}$ and a normalization matrix $\mathcal{D}$. Formally,

$$\begin{aligned}\mathbf{W}_{\text{GMM}} =& \left(\mu\mathbf{I} + \sum_i \mathbf{R}_i^T\mathbf{R}_i\right)^{-1}\left(\mu\mathbf{I} + \sum_i \mathbf{R}_i^T\mathbf{F}_i\mathbf{R}_i\right) \\ =& \mathcal{D}^{-1}\mathbf{K}_{\text{GMM}}. \end{aligned} \tag{10}$$

As a consequence, a graph-Laplacian can be constructed from this operator by

$$\boldsymbol{\mathcal{L}}_{\text{GMM}} = \mathbf{I} - \mathbf{W}_{\text{GMM}}, \tag{11}$$

where the eigenvalues of $\boldsymbol{\mathcal{L}}_{\text{GMM}}$ are in $(0, 1]$.

The denoising algorithm is obtained by minimizing the function $\mathcal{J}(\mathbf{u})$, defined in Equation (5), which can be done using a gradient descend strategy. As such, the estimated image is found by iterating:

$$\mathbf{u}^{k+1} = \mathbf{u}^k - \gamma\nabla\mathcal{J}(\mathbf{u}) \tag{12}$$

where

$$\nabla\mathcal{J}(\mathbf{u}) = \mathbf{u} - \mathbf{W}\mathbf{f} + \rho\boldsymbol{\mathcal{L}}\mathbf{u}. \tag{13}$$

For a given step-size γ and regularizer-strength ρ (fixed for each noise level), we run the gradient descent process a fixed number of iterations. In practice, we do not build $\boldsymbol{\mathcal{L}}$ nor $\mathbf{W}$ explicitly, but rather apply it by using the local filters $\mathbf{F}_i$. In the case of the GMM, this corresponds to knowing the Gaussians chosen for each patch. Before moving to the experimental section, we should note that the non-local diffusion method of [7] employs a minimization driven by an update very similar to that in Equation (13), where $\mathbf{f}$ is used instead of $\mathbf{W}\mathbf{f}$ – thus, the name. This shows more clearly the connection to the NLD algorithm, as well as providing a variational interpretation to the SOS boosting algorithm [6].

IV. Experimental Results

In this section we present image denoising results corresponding to the minimization of the cost function in Equation (4) for the Laplacian matrix $\boldsymbol{\mathcal{L}}$ induced by the GMM prior, for various standard test images and noise levels. As for the parameters the proposed GMD, for all noise levels we set $\gamma = 0.1$, $\rho = 4$, and 2 diffusion steps are applied.

We compare the proposed approach to the Non Local Diffusion work [7], which corresponds to a diffusion process guided by the graph built with the NLM kernel (minimizing Equation (4)). In addition, we also compare the GMD to the plain GMM denoiser as a baseline. We include for completion the results obtained by the EPLL [4]. This algorithm essentially minimizes a cost function similar to the one in Equation (2), where the prior is enforced *on the reconstructed patches*. This idea boils down to applying a GMM-based denoiser iteratively, with a set of parameters which need to be tuned. Note that the EPLL algorithm is still a patch-based method which also updates its operator (choosing the Gaussian Mixtures) at every iteration, whereas in our approach these remain constant.

Table I provides a comparison between NLD, GMM, EPLL and the proposed GMD approach in terms of Peak Signal to Noise Ratio (PSNR). As can be seen, for $\sigma = 20$, GMD achieves the best reconstruction performance. For $\sigma = 30$, EPLL and GMD obtain comparable results, whereas for $\sigma = 50$, the EPLL slightly outperforms GMD. We remind the reader that the EPLL, unlike our approach, updates its operator at each iteration. We believe that by updating the matrix $\boldsymbol{\mathcal{L}}$ (i.e., re-run the Gaussian selection step) the GMD results can be further improved, too. However, we choose not to include this step in our algorithm in order to focus the attention on the minimization of the problem in Equation (5).

Figure 1 provides a visual comparison between the NLD approach of [7], GMM and ours GMD method. As can be seen the GMD reconstruction has less artifacts than the baseline methods, complying with the quantitative PSNR measure.

V. Conclusion

In this paper, we introduced a graph interpretation of the GMM denoiser, followed by an analysis of the resulting operator. The denoising effect is obtained by minimizing a cost function with a graph-Laplacian regularization, as suggested by the SOS formulation. We have shown that the proposed approach can be understood from a variational perspective, resulting in close variant of the NLD algorithm. Following the experimental results, it was evidenced that our approach is more effective than the traditional NLD counterpart. Our results not only outperform those by the regular GMM denoising algorithm, but they are also competitive with those of the state-of-the-art EPLL method. We believe that updating the graph (and the corresponding operator) along the iterations (as done by the EPLL and the SOS boosting) will result in increased performance, and this is a promising direction of future work.

Appendix

In this appendix we provide the proves for Theorem 1.

Proof: We will start by showing that property 1 holds. Under the periodic boundary conditions, we have that

[2]By assuming cyclic boundary conditions, the following holds: $\sum_i \mathbf{R}_i^T\mathbf{R}_i = n\mathbf{I}$, where n is the dimension of the image patch.

TABLE I
DENOISING RESULTS FOR VARIOUS NOISE LEVELS AND IMAGES, GIVEN IN TERMS OF PSNR. THE BEST RESULT IS HIGHLIGHTED.

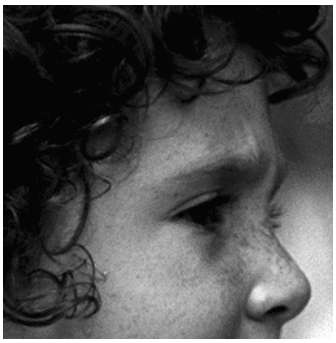
$\sigma \backslash$**Image**	House	Saturn	Foreman	Lena	Peppers	Girl	Woman	**Averages**
Non Local Diffusion (NLM kernel)								
20	32.27	35.21	32.65	31.35	31.51	30.03	30.99	32.00
30	30.00	32.57	30.42	29.36	26.65	28.60	28.86	29.49
50	27.14	29.64	27.76	26.86	27.09	26.72	26.06	27.40
GMM								
20	32.59	35.54	33.07	32.23	32.11	30.56	31.81	32.56
30	30.84	33.38	31.24	30.48	30.56	29.38	29.86	30.82
50	28.13	30.33	28.66	28.03	28.23	27.82	27.26	28.35
EPLL-GMM								
20	32.98	36.68	33.63	32.60	32.51	30.71	32.08	33.03
30	**31.22**	34.23	31.66	**30.78**	**30.90**	**29.54**	**30.04**	**31.20**
50	**28.76**	**31.15**	**29.16**	**28.41**	**28.68**	28.00	**27.57**	**28.82**
GMD (Proposed)								
20	**33.07**	**36.78**	**33.68**	**32.61**	**33.53**	**30.75**	**32.14**	**33.22**
30	31.20	**34.35**	**31.70**	30.70	30.86	29.53	29.97	31.19
50	28.42	31.12	29.08	28.23	28.47	**28.07**	27.27	28.66

(a) Input: Foreman

(b) NLD: PSNR = 32.65dB

(c) GMM: PSNR = 33.07dB

(d) GMD: PSNR = 33.68dB



(e) Input: Girl

(f) NLD: PSNR = 30.03dB

(g) GMM: PSNR = 30.56dB

(h) GMD: PSNR = 30.75dB

Fig. 1. Denoising of the images Foreman (a-d) and Girl (e-h), when $\sigma = 20$.

$$\mathbf{W}_{\text{GMM}} = \left(\mu\mathbf{I} + \sum_i \mathbf{R}_i^T\mathbf{R}_i\right)^{-1}\left(\mu\mathbf{I} + \sum_i \mathbf{R}_i^T\mathbf{F}_i\mathbf{R}_i\right) = \frac{1}{\mu+n}\left(\mu\mathbf{I} + \sum_i \mathbf{R}_i^T\mathbf{F}_i\mathbf{R}_i\right). \quad (14)$$

Recall that the filters $\mathbf{F}_i$, expressed in Equation (7) are symmetric (and more so, positive definite) as they are the inverse of symmetric matrices. Therefore, $\mathbf{W}_{\text{GMM}}$ is the sum of symmetric matrices, and it is then also symmetric.

The second property can be deduced using the same rationale. The matrices given by $\mathbf{R}_i^T\mathbf{F}_i\mathbf{R}_i$ are symmetric and positive semidefinite. Thus, their sum $\sum_i \mathbf{R}_i^T\mathbf{F}_i\mathbf{R}_i$ is also positive semidefinite. Moreover, we can express

$$\mathbf{W}_{\text{GMM}} = \frac{\mu}{\mu+n}\mathbf{I} + \frac{1}{\mu+n}\left(\sum_i \mathbf{R}_i^T\mathbf{F}_i\mathbf{R}_i\right). \quad (15)$$

The first term in this sum is obviously positive definite, while the second is positive semidefinite. From this, the minimal eigenvalue of the GMM matrix $\lambda_{min}(\mathbf{W}_{\text{GMM}}) = \frac{\mu}{\mu+n} > 0$, and $\mathbf{W}_{\text{GMM}} \succ 0$.

To prove the last property, consider the operator norm of $\mathbf{W}_{\text{GMM}}$ given the (square of the) decomposition presented above:

$$\begin{aligned}\|\mathbf{W}_{\text{GMM}}\|_2 &= \left\| \frac{\mu}{\mu+n}\mathbf{I} + \frac{1}{\mu+n}\left(\sum_i \mathbf{R}_i^T\mathbf{F}_i\mathbf{R}_i\right)\right\|_2 \\ &\leq \left\|\frac{\mu}{\mu+n}\mathbf{I}\right\|_2 + \left\|\frac{1}{\mu+n}\left(\sum_i \mathbf{R}_i^T\mathbf{F}_i\mathbf{R}_i\right)\right\|_2 .\end{aligned} \tag{16}$$

Note first that the operator norm of the first term is given by $\frac{\mu}{\mu+n}$. Focusing on the second term, we consider a similar decomposition to that presented in [6] (Appendix B). The sum in the last Equation, over all N patches, considers overlapping structures. We can decompose this term by considering the sum over $\{\Omega_j\}_{j=1}^n$ groups of *non overlapping* patches only. With this, we have that

$$\begin{aligned}\left\|\sum_i \mathbf{R}_i^T\mathbf{F}_i\mathbf{R}_i\right\|_2 &= \left\|\sum_j^n \sum_{k\in\Omega_j} \mathbf{R}_k^T\mathbf{F}_k\mathbf{R}_k\right\|_2 \\ &\leq \sum_j^n \|\mathbf{M}_j\|_2 ,\end{aligned} \tag{17}$$

where we have denoted $\mathbf{M}_j = \sum_{k\in\Omega_j} \mathbf{R}_k^T\mathbf{F}_k\mathbf{R}_k$. Notice that $\mathbf{M}$ is a block diagonal matrix, having the filters $\mathbf{F}_k$ as leading minors.

To show that $\|\mathbf{M}_j\|_2 \leq 1$ we will rely on the definition of the induced norm. Consider thus any vector $\mathbf{X} \in \mathbb{R}^N$ such that $\|\mathbf{X}\|_2 = 1$. Moreover, denote $\mathbf{R}_k\mathbf{X} = x_k \in \mathbb{R}^n$. Then,

$$\begin{aligned}\|\mathbf{M}_j\mathbf{X}\|_2^2 &= \left\|\sum_{k\in\Omega_j} \mathbf{M}_j^k\mathbf{X}\right\|_2^2 \\ &= \left\|\sum_{k\in\Omega_j} \mathbf{R}_k^T\mathbf{F}_k x_k\right\|_2^2 .\end{aligned} \tag{18}$$

Due to the fact that $\mathbf{R}_k^T\mathbf{R}_j = \mathbf{0}$, $\forall i \neq k$ (because the corresponding patches are non-overlapping), we have that

$$\begin{aligned}\|\mathbf{M}_j\mathbf{X}\|_2^2 &\leq \sum_{k\in\Omega_j} \left\|\mathbf{R}_k^T\mathbf{F}_k x_k\right\|_2^2 , \\ &\leq \sum_{k\in\Omega_j} \left\|\mathbf{R}_k^T\right\|_2^2 \|\mathbf{F}_k x_k\|_2^2 , \\ &\leq \sum_{k\in\Omega_j} \|\mathbf{F}_k x_k\|_2^2 ,\end{aligned} \tag{19}$$

where we have used the multiplicative property of the operator norm and the fact that $\|\mathbf{R}_k^T\|_2 = 1$.

Looking now at the square of the operator norm of $\mathbf{F}_k x_k$:

$$\begin{aligned}\|\mathbf{F}_k x_k\|_2^2 &= \left\|(\mathbf{I} + \sigma^2\mathbf{\Sigma}_k^{-1})^{-1} x_k\right\|_2^2 \\ &\leq \|x_k\|_2^2 ,\end{aligned} \tag{20}$$

where the inequality holds since $\lambda_{max}(\mathbf{I} + \sigma^2\mathbf{\Sigma}_k^{-1}) \geq 1$, as $\mathbf{\Sigma}_k^{-1} \succ 0$.

By using Equations (18), (19), (20), and the fact that $\|\mathbf{X}\|_2^2 = \sum_{k\in\Omega_j} \|x_k\|_2^2$, we have that

$$\begin{aligned}\frac{\|\mathbf{M}_j\mathbf{X}\|_2^2}{\|\mathbf{X}\|_2^2} &= \frac{\left\|\sum_{k\in\Omega_j} \mathbf{R}_k^T\mathbf{F}_k x_k\right\|_2^2}{\|\sum_{k\in\Omega_j} x_k\|_2^2} \\ &\leq \frac{\sum_{k\in\Omega_j} \|\mathbf{F}_k x_k\|_2^2}{\sum_{k\in\Omega_j} \|x_k\|_2^2} \\ &\leq \frac{\sum_{k\in\Omega_j} \|x_k\|_2^2}{\sum_{k\in\Omega_j} \|x_k\|_2^2} = 1.\end{aligned} \tag{21}$$

Therefore, the squared maximal singular value (its operator norm) of $\mathbf{M}_j$ is ≤ 1. Incorporating this into Equation (16), we have that

$$\begin{aligned}\|\mathbf{W}_{\text{GMM}}\|_2 &\leq \left\|\frac{\mu}{\mu+n}\mathbf{I}\right\|_2 + \left\|\frac{1}{\mu+n}\left(\sum_i \mathbf{R}_i^T\mathbf{F}_i\mathbf{R}_i\right)\right\|_2 \\ &\leq \frac{\mu}{\mu+n} + \frac{1}{\mu+n}\sum_{j=1}^n \|\mathbf{M}_j\|_2 \\ &\leq \frac{\mu}{\mu+n} + \frac{n}{\mu+n} \leq 1.\end{aligned} \tag{22}$$

■

REFERENCES

[1] M. Elad and M. Aharon, "Image denoising via sparse and redundant representations over learned dictionaries.," *IEEE Trans. Image Process.*, vol. 15, pp. 3736–3745, Dec. 2006. 1

[2] J. Mairal, G. Sapiro, and M. Elad, "Multiscale Sparse Image Representation with Learned Dictionaries," *Int. Conf. Image Process.*, vol. 3, pp. 105–108, 2007. 1

[3] J. Portilla, V. Strela, M. J. Wainwright, and E. P. Simoncelli, "Image denoising using scale mixtures of Gaussians in the wavelet domain.," *IEEE Transactions on Image Processing.*, vol. 12, pp. 1338–51, Jan. 2003. 1, 2

[4] D. Zoran and Y. Weiss, "From learning models of natural image patches to whole image restoration," *2011 International Conference on Computer Vision, ICCV.*, pp. 479–486, Nov. 2011. 1, 2, 3

[5] P. Milanfar, "A tour of modern image filtering: New insights and methods, both practical and theoretical," *Signal Processing Magazine, IEEE*, vol. 30, no. 1, pp. 106–128, 2013. 1

[6] Y. Romano and M. Elad, "Boosting of Image Denoising Algorithms," *SIAM Journal on Imaging Sciences*, vol. 8, no. 2, pp. 1187–1219, 2015. 1, 2, 3, 5

[7] G. Gilboa and S. Osher, "Nonlocal linear image regularization and supervised segmentation," *Multiscale Modeling & Simulation*, vol. 6, no. 2, pp. 595–630, 2007. 1, 2, 3

[8] F. G. Meyer and X. Shen, "Perturbation of the eigenvectors of the graph laplacian: Application to image denoising," *Applied and Computational Harmonic Analysis*, vol. 36, no. 2, pp. 326–334, 2014. 1

[9] A. Elmoataz, O. Lezoray, and S. Bougleux, "Nonlocal discrete regularization on weighted graphs: a framework for image and manifold processing," *IEEE Transactions on Image Processing*, vol. 17, no. 7, pp. 1047–1060, 2008. 1

[10] A. Buades, B. Coll, and J. M. Morel, "A non-local algorithm for image denoising," *Conference on Computer Vision and Pattern Recognition, CVPR IEEE.*, pp. 60–65, 2005. 2

[11] A. Kheradmand and P. Milanfar, "A general framework for regularized, similarity-based image restoration," *IEEE Transactions on Image Processing*, vol. 23, no. 12, pp. 5136–5151, 2014. 2