A General Framework for Kernel Similarity-based Image Denoising

Amin Kheradmand
Department of Electrical Engineering
University of California, Santa Cruz, CA, USA
aminkh@soe.ucsc.edu

Peyman Milanfar
Department of Electrical Engineering
University of California, Santa Cruz, CA, USA
Google, Inc., Mountain View, CA, USA
milanfar@ucsc.edu

Abstract—Any image can be represented as a function defined on a discrete weighted graph whose vertices are image pixels. Each pixel can be linked to other pixels via graph edges with corresponding weights derived from similarities between image pixels (graph vertices) measured in some appropriate fashion. Image structure is encoded in the Laplacian matrix derived from these similarity weights. Taking advantage of this graph-based point of view, we present a general regularization framework for image denoising. A number of well-known existing denoising methods like bilateral, NLM, and LARK, can be described within this formulation. Moreover, we present an analysis for the filtering behavior of the proposed method based on the spectral properties of Laplacian matrices. Some of the well established iterative approaches for improving kernel-based denoising like diffusion and boosting iterations are special cases of our general framework. The proposed approach provides a better understanding of enhancement mechanisms in self similarity-based methods, which can be used for their further improvement. Experimental results verify the effectiveness of this approach for the task of image denoising.

Index Terms—Image Denoising, Graph Laplacian, Kernel Similarity Matrix.

I. INTRODUCTION

Most real images contain some level of noise and/or blur distortions. The following general model describes these degradations:

$$\mathbf{y} = A\mathbf{z} + \mathbf{n}, \tag{1}$$

where $\mathbf{y}$ is the observed image pixel values lexicographically ordered in an n-element column vector (n is the total number of pixels), $\mathbf{z}$ is the corresponding latent image and $\mathbf{n}$ is additive white noise vector which is assumed to be zero mean with standard deviation σ. Moreover, A is an $n \times n$ blurring matrix resulting from a blur kernel. Generally speaking, (1) is an ill-posed problem and requires an appropriate regularization technique to avoid artifacts in the restored image. There is a very rich literature on different ways to come up with a desirable estimate of the unknown image $\mathbf{z}$ for the general deblurring problem and the special denoising case, where $A = I$. In this paper, we focus on image denoising. In general, regularization methods for denoising can be divided into two categories. The first group are those classical methods that take advantage of some prior knowledge about images. Algorithms that exploit either image smoothness priors [1] or sparsity of image spectrum coefficients in some specific

domain (e.g., Wavelet or DCT) [2] fall into this group. The second class of methods exploit the existing self similarity in images. Kernel-based denoising methods like bilateral, non-local means (NLM), and LARK [3]–[6], fall into this category. In a wide angle view, all the above mentioned algorithms perform denoising based on some type of subset selection or shrinkage operation in a fixed or adaptive basis. Note that state-of-the-art image denoising methods like BM3D try to find the optimal type of shrinkage operation by combining self similarity information with sparsity property of image spectrum coefficients in some appropriate domain [7]. There are some approaches that attempt to use the non-local similarity idea in a variational formulation [8], [9]. Some other papers connect the idea of graph signal representation and associated Laplacian matrix in graph theory with non-local similarity in image processing [6], [10], [11]. Inspired by these works, in this paper, we present a new general two-step restoration algorithm. The first step involves the computation of a kernel similarity matrix and its corresponding Laplacian operator. In the second step, an objective function is formulated and iteratively optimized in which data fidelity and smoothness terms are coupled via Laplacian and similarity matrices of the underlying image. In Section II, a kernel based denoising framework based on similarity and Laplacian matrices along with its filtering interpretation based on the eigenvectors of the corresponding Laplacian matrices are introduced. In Section III, two widely used iterative denoising methods are discussed as special cases of the framework introduced in Section II. Numerical results are presented in Section IV.

II. KERNEL FORMULATION FOR DENOISING

We propose the following unified cost function for kernel-based image denoising:

$$E(\mathbf{z}) = (\mathbf{y} - \mathbf{z})^T F(K)(\mathbf{y} - \mathbf{z}) + \eta \, \mathbf{z}^T G(L) \, \mathbf{z}, \tag{2}$$

in which K is a data-dependent kernel similarity matrix whose (i, j)th element is the kernel similarity coefficient between pairs of pixels i and j (Fig. 1). These similarity coefficients can be derived either locally using e.g., bilateral or LARK kernels or non-locally using NLM. L is the corresponding Laplacian matrix computed from K, and η is a positive regularization parameter which balances the first term (data fidelity term) and the second term (smoothness term). Also,

$F(.)$ and $G(.)$ are functions of K and L, to be specified shortly. While the core discussion is applicable to any valid choice of kernels [6], to keep focus here, we use the NLM kernel as a canonical example in the remainder of the paper. Kernel weight coefficients are computed from a pre-filtered version $\tilde{\mathbf{z}}$ of the observed image. If we denote a patch centered at pixel i in the image $\tilde{\mathbf{z}}$ as $\tilde{\mathbf{z}}_i$, the (i,j)th element of matrix K demonstrates the degree of similarity between pixels i and j and is computed as

$$K(i,j) = \exp\{-\frac{\|\tilde{\mathbf{z}}_i - \tilde{\mathbf{z}}_j\|^2}{h^2}\}, \qquad (3)$$

where h is the smoothing parameter. In the following subsections, we discuss two instances of the above energy function that describe some of the existing kernel-based methods.

A. Case 1: Un-normalized Laplacian

By defining $G(L) = L = D - K$ and $F(K) = K$, we have

$$E(\mathbf{z}) = (\mathbf{y} - \mathbf{z})^T K(\mathbf{y} - \mathbf{z}) + \eta \mathbf{z}^T(D - K)\mathbf{z}, \qquad (4)$$

where D is a diagonal matrix whose ith diagonal element is the sum of the elements of ith row of K; i.e., $D = K\mathbf{1}_n$ where $\mathbf{1}_n$ is an n-dimensional vector of ones. Note that $D - K$ is the un-normalized Laplacian matrix widely used in graph theory applications [12]. The first term is essentially a weighted data fidelity term and the second term is a non-local term that adaptively penalizes large derivatives based on the structure of data encoded in the Laplacian matrix.

The steepest descent (SD) iterations can be used to minimize (4) with respect to $\mathbf{z}$:

$$\begin{aligned}\hat{\mathbf{z}}_k &= \hat{\mathbf{z}}_{k-1} - \mu \nabla E(\mathbf{z})_{|\mathbf{z}=\hat{\mathbf{z}}_{k-1}} \\ &= \hat{\mathbf{z}}_{k-1} + \mu K(\mathbf{y} - \hat{\mathbf{z}}_{k-1}) - \mu\eta(D - K)\hat{\mathbf{z}}_{k-1}\end{aligned} \qquad (5)$$

Here, μ is the step size for SD iterations. At convergence, the corresponding estimate would be

$$\hat{\mathbf{z}}_{un-norm} = (K + \eta(D - K))^{-1}K\mathbf{y}. \qquad (6)$$

With K as e.g., the NLM kernel, and for the specific choice of $\eta = 1$ (both terms contribute equally strongly), we have precisely the NLM denoising algorithm:

$$\hat{\mathbf{z}}_{NLM} = D^{-1}K\mathbf{y}. \qquad (7)$$

Moreover, as shown in Section IV, with a fixed smoothing parameter h, tuning the regularization parameter η, yields an estimation with lower mean squared error (MSE) compared to standard NLM solution. We can use SURE-based MSE estimation approach for adjusting the regularization parameter η [13].

B. Case 2: Normalized Laplacian

The second approach is to apply Sinkhorn-Knop matrix scaling algorithm [14], [15] to the symmetric non-negative matrix K to construct the filtering matrix $W = C^{-1/2}KC^{-1/2}$, where $C^{-1/2}$ is a diagonal non-negative normalizing matrix which scales the kernel similarity matrix K. The resulting matrix W is a symmetric non-negative doubly stochastic
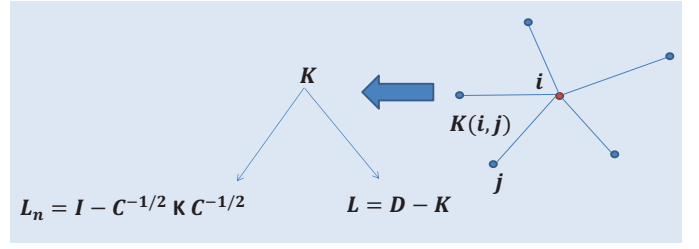

Figure 1. Graph representation of images and construction of kernel similarity matrix K, un-normalized Laplacian L and normalized Laplacian L_n.

matrix. Hence, based on Perron-Frobenius theory, W has spectral radius $0 \leq \lambda(W) \leq 1$ with largest eigenvalue $\lambda_1 = 1$ whose corresponding eigenvector is $\mathbf{v}_1 = (\frac{1}{\sqrt{n}})[1,1,...,1]^T = (\frac{1}{\sqrt{n}})\mathbf{1}_n$. Moreover, it can be decomposed as $W = VSV^T$, in which V is an orthonormal matrix whose ith column $\mathbf{v}_i$ is the ith eigenvector of W. The corresponding ith eigenvalue λ_i is the ith diagonal element of S. At this point, we are able to define the normalized Laplacian matrix $I - W$ (Fig. 1). In this case, our objective function becomes:

$$E(\mathbf{z}) = (\mathbf{y} - \mathbf{z})^T W(\mathbf{y} - \mathbf{z}) + \eta \mathbf{z}^T(I - W)\mathbf{z}. \qquad (8)$$

By computing the gradient of (8) with respect to $\mathbf{z}$, one obtains an iterative SD update equation as

$$\hat{\mathbf{z}}_k = \hat{\mathbf{z}}_{k-1} + \mu W(\mathbf{y} - \hat{\mathbf{z}}_{k-1}) - \mu\eta(I - W)\hat{\mathbf{z}}_{k-1}. \qquad (9)$$

By appropriate selection of step size parameter μ, (9) converges to

$$\hat{\mathbf{z}}_{norm} = (W + \eta(I - W))^{-1}W\mathbf{y}. \qquad (10)$$

Again, for the case $\eta = 1$, optimization of (8) leads to $\hat{\mathbf{z}}_{norm} = W\mathbf{y}$ which has been shown to outperform its non-symmetric counterpart $\hat{\mathbf{z}}_{NLM} = D^{-1}K\mathbf{y}$ [16].

C. Spectral analysis based on eigenvectors of Laplacian

Since the filtering matrix W and the normalized Laplacian matrix $I - W$ have the same eigenvectors, (10) can be interpreted as filtering the observed image in a space spanned by the eigenvectors of the Laplacian. Noting that $W = VSV^T$, (10) can be expressed as

$$\hat{\mathbf{z}}_{norm} = W'\mathbf{y} = VS'V^T\mathbf{y}, \qquad (11)$$

where S' is a diagonal matrix whose ith diagonal element λ_i' is a function of the corresponding eigenvalue λ_i of W as

$$\lambda_i' = p(\lambda_i) = \frac{\lambda_i}{(1 - \eta)\lambda_i + \eta}, \qquad (12)$$

Similarly by defining $D^{-1/2}KD^{-1/2} = U\Phi U^T$, (6) can be written as

$$D^{1/2}\hat{\mathbf{z}}_{un-norm} = U\Phi'U^T D^{1/2}\mathbf{y}. \qquad (13)$$

U is an orthonormal matrix which contains eigenvectors of $I - D^{-1/2}KD^{-1/2}$ in its columns. Φ' is a diagonal matrix
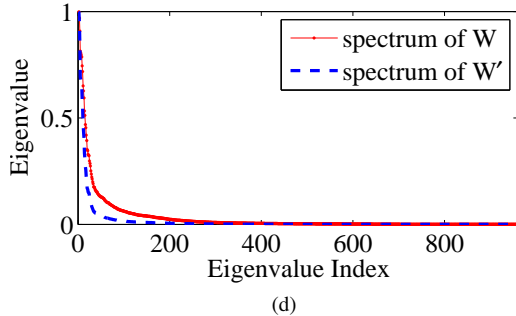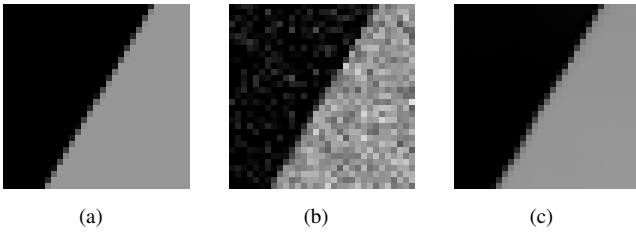
(a)   (b)   (c)

(d)

Figure 2. Denoising experiment on 31×31 piece-wise constant synthetic patch (a) clean patch, (b) noisy patch by adding white Gaussian noise with $\sigma = 25$ (PSNR = 20.31dB) , (c) output of iterative algorithm (9) (PSNR= 35.96dB, $\eta = 19$), and (d) spectrum of filter matrices W (λ_i's) and W' (λ_i''s).

whose ith diagonal element φ_i' can be derived in terms of the ith diagonal element of Φ, namely, φ_i as

$$\varphi_i' = p(\varphi_i) = \frac{\varphi_i}{(1-\eta)\varphi_i + \eta}. \tag{14}$$

Equation (13) can be regarded as a filtering interpretation of un-normalized Laplacian case in the space spanned by U applied to a scaled version of the input to obtain a scaled version of the desired estimate. Optimizing the value of η in (12) or (14) with respect to an appropriate measure (e.g., using SURE [13]) gives the desired estimate for both normalized and un-normalized cases. As an illustration, an edge patch of size 31×31 is considered in Fig. 2(a) and white Gaussian noise of standard deviation 25 is added to it (Fig. 2(b)). The resulting output of iterative algorithm (9) is shown in Fig. 2(c). Fig. 2(d) shows the spectrum of W (λ_i) and the corresponding spectrum of W' (λ_i'). As will be shown in experiments, normalized Laplacian formulation results in denoising outputs with slightly better visual quality. Moreover, filtering analysis of the normalized case is straightforward in the space spanned by V.

III. DIFFUSION AND BOOSTING AS SPECIAL CASES OF THE PROPOSED ALGORITHM

Two widely used iterative methods, namely diffusion and boosting, have been effectively used for improving the performance of kernel-based denoising algorithms [6]. These iterative methods are two extreme cases of our more general iterative approach (Eq. 9), as we illustrate below.

By setting $\eta = \infty$ in (8), the corresponding SD update equation becomes diffusion iterations as

$$\hat{\mathbf{z}}_k = W\hat{\mathbf{z}}_{k-1}. \tag{15}$$

(a) Lena   (b) Barbara   (c) Parrot   (d) House

Figure 3. Set of images used for evaluating the performance of our denoising algorithm.

Initializing (15) with $\hat{\mathbf{z}}_0 = \mathbf{y}$ leads to $\hat{\mathbf{z}}_\infty = \frac{1}{n}\mathbf{1}_n\mathbf{y}^T\mathbf{1}_n$, which corresponds to a final constant estimate (an estimation without variance).

On the other hand, by choosing η to be equal to 0, the effect of the smoothness term in (8) is omitted, for which, boosting iterations is derived as

$$\hat{\mathbf{z}}_k = \hat{\mathbf{z}}_{k-1} + W(\mathbf{y} - \hat{\mathbf{z}}_{k-1}). \tag{16}$$

In this case, initializing the boosting algorithm in (16) with $\hat{\mathbf{z}}_0 = \mathbf{0}$, results in $\hat{\mathbf{z}}_\infty = \mathbf{y}$. This corresponds to an un-biased estimate of the original image with variance equal to the noise variance.

IV. EXPERIMENTAL RESULTS

In order to evaluate the effectiveness of the proposed kernel-based restoration approach, we apply this algorithm for restoration of 256×256 benchmark images. For denoising experiments, Gaussian noise with standard deviation 20 is added to images shown in Fig. 3 and performance of iterative algorithms (5) and (9) are compared against the output of standard NLM. Peak signal to noise ratio (PSNR) in dB, and SSIM [17] are used as quantitative measures for comparison. As can be seen in Table I, in all cases we get better results in terms of PSNR and SSIM with respect to standard NLM. Also, note that SSIM values in Table I reflect slightly better visual quality of the results of normalized iterative algorithm (9) compared to un-normalized algorithm (5). Fig. 4 illustrates House image denoised using the general iterative kernel-based approaches (5) and (9) compared to standard NLM denoising. Additionally, the result of applying normalized iterative algorithm (9) to a noisy color image with the same experimental settings as for the previous examples is shown in Fig. 5.

V. CONCLUSION

We have presented a new general framework for image denoising in this paper. This graph-based approach encompasses some well-known existing denoising methods, and provides a path for further improvements. The corresponding cost function considered in this paper contains quadratic terms for both data and prior terms. Also, this work can be extended to the more general deblurring problem.
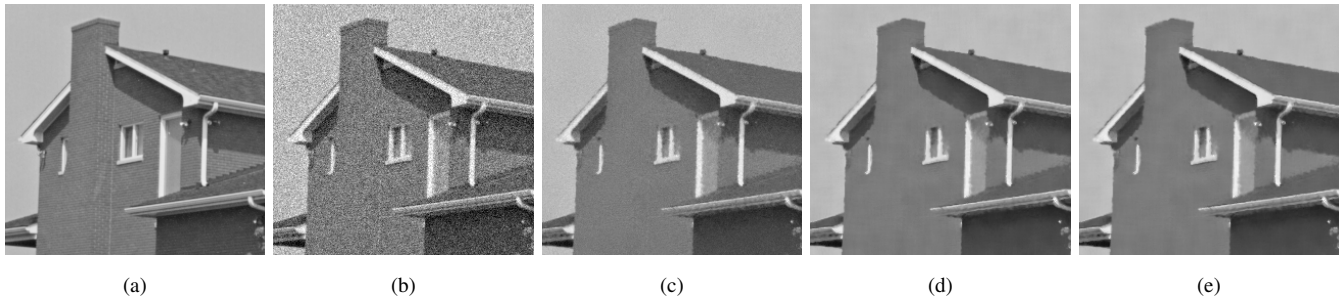
Figure 4. Denoising experiment on 256×256 House image, (a) clean image, (b) noisy image ($\sigma = 20$), (c) standard NLM output image (PSNR = 30.80dB), (d) output of iterative algorithm (5) (PSNR= 32.16dB, $\eta = 0.67$), and (e) output of iterative algorithm (9) (PSNR= 32.37dB, $\eta = 0.82$).

Table I
DENOISING PSNR AND SSIM PERFORMANCE OF THE ITERATIVE
ALGORITHMS (5) AND (9).

	NLM	Un-normalized		Normalized	
	PSNR (SSIM)	PSNR (SSIM)	η	PSNR (SSIM)	η
Lena	28.68 (0.761)	29.66 (0.807)	0.53	**29.73** **(0.814)**	0.58
Barbara	28.02 (0.794)	**29.36** **(0.855)**	0.53	29.35 **(0.860)**	0.58
Parrot	28.81 (0.786)	**29.16** **(0.841)**	0.53	29.13 **(0.848)**	0.67
House	30.80 (0.768)	32.16 (0.845)	0.67	**32.37** **(0.849)**	0.82

REFERENCES

[1] L. I. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *Physica D: Nonlinear Phenomena*, vol. 60, no. 1, pp. 259–268, 1992.

[2] D. L. Donoho, "De-noising by soft-thresholding," *Information Theory, IEEE Transactions on*, vol. 41, no. 3, pp. 613–627, 1995.

[3] C. Tomasi and R. Manduchi, "Bilateral filtering for gray and color images," in *Computer Vision, 1998. Sixth International Conference on*, Jan. 1998, pp. 839 –846.

[4] A. Buades, B. Coll, and J. M. Morel, "A review of image denoising algorithms, with a new one," *Simul*, vol. 4, pp. 490–530, 2005.

[5] H. Takeda, S. Farsiu, and P. Milanfar, "Kernel regression for image processing and reconstruction," *Image Processing, IEEE Transactions on*, vol. 16, no. 2, pp. 349 –366, Feb. 2007.

[6] P. Milanfar, "A tour of modern image filtering," *IEEE Signal Process. Mag.*, vol. 30, no. 1, pp. 106–128, 2013.

[7] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, "Image denoising by sparse 3-D transform-domain collaborative filtering," *IEEE Transactions on Image Processing*, vol. 16, no. 8, pp. 2080 –2095, Aug. 2007.

[8] G. Peyré, "Image processing with nonlocal spectral bases," *Multiscale Modeling & Simulation*, vol. 7, no. 2, pp. 703–730, 2008.

[9] L. Pizarro, P. Mrázek, S. Didas, S. Grewenig, and J. Weickert, "Generalised nonlocal image smoothing," *International Journal of Computer Vision*, vol. 90, no. 1, pp. 62–87, 2010.

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

[11] D. Shuman, S. Narang, P. Frossard, A. Ortega, and P. Vandergheynst, "The emerging field of signal processing on graphs: Extending high-dimensional data analysis to networks and other irregular domains," *Signal Processing Magazine, IEEE*, vol. 30, no. 3, pp. 83–98, 2013.

[12] U. Von Luxburg, "A tutorial on spectral clustering," *Statistics and computing*, vol. 17, no. 4, pp. 395–416, 2007.

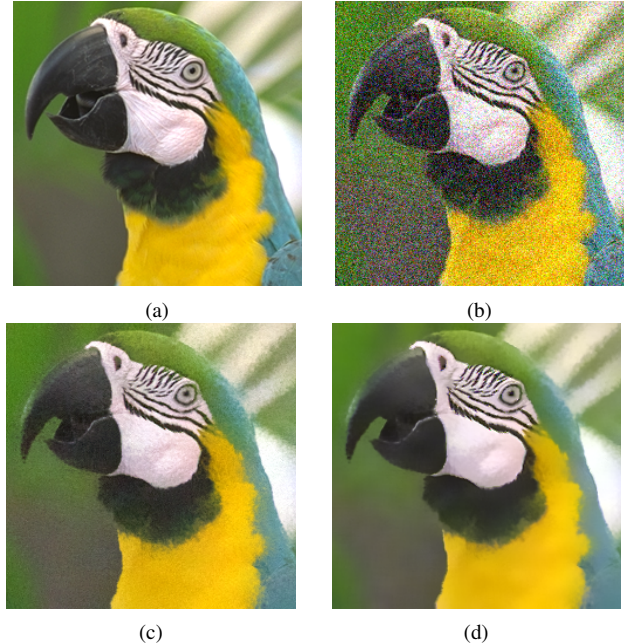[13] S. Ramani, T. Blu, and M. Unser, "Monte-Carlo SURE: A black-box optimization of regularization parameters for general denoising algorithms," *IEEE Transactions on Image Processing*, vol. 17, no. 9, pp. 1540–1554, September 2008.

[14] R. Sinkhorn and P. Knopp, "Concerning nonnegative matrices and doubly stochastic matrices," *Pacific J. Math*, vol. 21, no. 2, pp. 343–348, 1967.

[15] P. A. Knight and D. Ruiz, "A fast algorithm for matrix balancing," *IMA Journal of Numerical Analysis*, 2012.

[16] P. Milanfar, "Symmetrizing smoothing filters," *SIAM Journal on Imaging Sciences*, vol. 6, no. 1, pp. 263–284, 2013.

[17] Z. Wang, A. Bovik, H. Sheikh, and E. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *Image Processing, IEEE Transactions on*, vol. 13, no. 4, pp. 600 –612, April 2004.

Figure 5. Denoising experiment on 256×256 color parrot image, (a) original image, (b) noisy image ($\sigma = 20$, PSNR= 17.06dB), (c) standard NLM output image (PSNR= 27.68dB), and (d) output of iterative algorithm (9) (PSNR= 28.54dB).