

Symmetric Smoothing Filters from Global Consistency Constraints

Sk. Mohammadul Haque[†], Gautam Pai and Venu Madhav Govindu^{*}

Abstract

Many patch-based image denoising methods can be viewed as data-dependent smoothing filters that carry out a weighted averaging of similar pixels. It has recently been argued that these averaging filters can be improved using their doubly stochastic approximation which are symmetric and stable smoothing operators. In this paper we introduce a simple principle of consistency that argues that the relative similarities between pixels as imputed by the averaging matrix should be preserved in the filtered output. The resultant consistency filter has the theoretically desirable properties of being symmetric and stable, and is a generalized doubly stochastic matrix. In addition, we can also interpret our consistency filter as a specific form of Laplacian regularization. Thus, our approach unifies two strands of image denoising methods, i.e. symmetric smoothing filters and spectral graph theory. Our consistency filter provides high quality image denoising and significantly outperforms the doubly stochastic version. We present a thorough analysis of the properties of our proposed consistency filter and compare its performance with that of other significant methods for image denoising in the literature.

Index Terms

Image denoising, restoration, symmetric smoothers, graph Laplacian, manifold regularization.

I. INTRODUCTION

IN recent years denoising of natural images has attracted an enormous amount of interest. Earlier attempts at image denoising involved building global models or parametric smooth approximations to the underlying function and using them as a prior in a Bayesian setting. Examples of such approaches include total variation norm minimization [1], wavelet [2] and curvelet methods [3]. While such techniques were largely based on an analysis of the spectral content of an image, a major breakthrough was achieved when image denoising was motivated as a non-parametric, data-dependent filtering operation that leads to pixel or patch-based regression techniques. These include the bilateral filter [4] and the non-local means [5] filter which estimated intensity at a pixel as a weighted combination of itself and all other pixels where the weights were based on a choice of kernel that captured *similarity* between a pair of pixels. [6] provides an excellent tutorial on image denoising methods based on the above idea. More recently, in [7] it is suggested that symmetrization of filters using their nearest doubly stochastic versions leads to improved denoising performance. Regularization methods are also used in image denoising using pixel similarity, e.g. [8] and [9]. Table I summarizes the categorization of some image denoising methods. Almost all of the currently best performing methods are patch-based approaches [5], [10], [11]. Indeed, the success of this approach has even lead to the question of whether image denoising has reached the optimal levels of performance [12], [13], [14].

While there are a variety of methods for image denoising in the literature, in this paper we will confine ourselves to the class of data-dependent linear filters that define the desired denoised intensity of a pixel as a linear weighted combination of the intensities of all pixels. Specifically, for the i -th pixel in an $M \times N$ image, our observation model for an individual pixel intensity is

$$y_i = z_i + n_i, \quad i = 1, 2, \dots, N_p \quad (1)$$

Copyright (c) 2014 IEEE. Personal use of this material is permitted. However, permission to use this material for any other purposes must be obtained from the IEEE by sending a request to pubs-permissions@ieee.org.

The authors are with the Department of Electrical Engineering, Indian Institute of Science, Bengaluru, India. (e-mail: smhaque@ee.iisc.ernet.in[†], venu@ee.iisc.ernet.in^{*}). The first two authors contributed equally to this work.

TABLE I: Schematic classification of image denoising methods.

| | Methods | Examples |
|----|------------------|---|
| 1. | Global | PCA-based, TV-norm-minimization [1], wavelet-based [2], curvelet-based [3]. |
| 2. | Parametric | Spline-based [15], [16]. |
| 3. | Non-parametric | |
| | Data-independent | Gaussian filter, Fourier/DCT domain filter. |
| | Data-dependent | Median filter, mode filter, bilateral [4], steering kernel [17], non-local means [5]. |

where z_i represents the true (but unknown) intensity of the i -th pixel, y_i is the observed intensity and n_i represents the observation noise that is assumed to have a zero mean *i.i.d.* Gaussian distribution throughout the image. Here $N_p = M \times N$ is the total number of pixels in the image. The observations of all pixel intensities of the form (1) can be collected into a vector form,

$$\mathbf{y} = \mathbf{z} + \mathbf{n} \quad (2)$$

where \mathbf{y} and \mathbf{z} are N_p -dimensional vectors and $\mathbf{n} \sim \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{I})$. The problem of image denoising is to recover \mathbf{z} from the observed noisy image \mathbf{y} . It is easy to see that the problem is ill-posed and hence a denoising algorithm needs more information on the structure of the image \mathbf{z} . In our case, this additional information comes from two sources. Firstly, like many patch-based approaches including the non-local means of [5], we use a linear weighting function to model pixel intensities as linear combinations of other similar pixels. Secondly, we introduce a novel modeling criterion of consistency which is used to estimate the desired denoised image. In Sections II and III we present background material on non-parametric regression and graph representations respectively that we use in our approach. In Section V we introduce and develop our notion of consistency which leads to a novel consistency filter. The properties of our proposed filter are examined in Sec. VI. Our proposed solution for image denoising can be characterized in terms of two distinct filtering approaches, i.e. symmetric, stable operators and regularized estimates on graphs. In Sec. VII and VIII we compare the behavior and performance of our filter with respect to representative methods from the two above-mentioned filtering approaches. Specifically, in Sec. VII we explain the differences between our approach and that of [7]. In Sec. VIII we explain the differences between our approach and the regularized filters of [8] and demonstrate that our approach arises from a very different perspective leads to a novel framework that gives better performance. In Sec. IX we present results of comparing the performance of our denoising approach to other methods in the literature, including some state-of-the-art image denoisers.

II. NON PARAMETRIC REGRESSION IN IMAGE RESTORATION

Many modern image denoising methods can be written in the form

$$\hat{z}_i = \sum_{j=1}^{N_p} W_{ij} y_j \quad (3)$$

where W_{ij} is a measure of the similarity between pixels i and j . Thus, \mathbf{W} is an $N_p \times N_p$ weight matrix that encapsulates the similarities between all image pixels. Typically, the weights W_{ij} are defined by the choice of a kernel function that represents the notion of similarity we wish to capture. The intensity \hat{z}_i is estimated as a weighted average of all pixels j with the weights given by W_{ij} . In practice, due to computational considerations, only the largest weights W_{ij} are considered resulting in a sparse representation for matrix \mathbf{W} .¹

The image filtering approach of (3) can be considered as a solution to the weighted least squares problem [6]

$$\hat{z}_i = \arg \min_z \sum_{j=1}^{N_p} (y_j - z)^2 K(i, j) \quad (4)$$

$$= \arg \min_z (\mathbf{y} - z\mathbf{1})^T \mathbf{K}_i (\mathbf{y} - z\mathbf{1}) \quad (5)$$

¹Recent work in [18] estimates a full matrix \mathbf{W} by using the Nystrom extension.

where $\mathbf{y} = [y_1, y_2, \dots, y_{N_p}]^T$, $K(i, j)$ is a kernel indicating the similarity between pixels i and j , \mathbf{K}_i is diagonal with entries $[K(i, 1), K(i, 2) \dots K(i, N_p)]^T$ and $\mathbf{1}$ is an N_p dimensional vector of all ones. The solution to the above weighted least squares problem is given by (3) where $W_{ij} = \frac{K(i, j)}{\sum_j K(i, j)}$. Stacking all estimates we obtain

$$\hat{\mathbf{z}} = \mathbf{W}\mathbf{y} = \mathbf{D}^{-1}\mathbf{K}\mathbf{y} \quad (6)$$

i.e. $\mathbf{W} = \mathbf{D}^{-1}\mathbf{K}$ where \mathbf{D} is diagonal with i -th entry of $\sum_j K(i, j)$. It will be observed that the matrix \mathbf{W} of (6) is data dependent and satisfies the following properties :

Proposition 1 (Properties of \mathbf{W}). a) *The matrix \mathbf{W} is row-stochastic.*

b) *The matrix \mathbf{W} has all eigenvalues real lying in $[0, 1]$ with the largest eigenvalue $\lambda_1 = 1$.*

Proof :

- a) From the definition of \mathbf{W} , we have $W(i, j) \geq 0$. Also, $(\mathbf{W}\mathbf{1})_i = \sum_j W(i, j) = \frac{1}{\sum_j K(i, j)} \sum_j K(i, j) = 1$ implying that \mathbf{W} is row-stochastic.
- b) From (6), $\mathbf{W} = \mathbf{D}^{-1}\mathbf{K} = \mathbf{D}^{-\frac{1}{2}} \left(\mathbf{D}^{-\frac{1}{2}}\mathbf{K}\mathbf{D}^{-\frac{1}{2}} \right) \mathbf{D}^{\frac{1}{2}}$. Since \mathbf{K} is a kernel function and \mathbf{D} is a sum of such kernel functions, it can be observed that $\mathbf{D}^{-\frac{1}{2}}\mathbf{K}\mathbf{D}^{-\frac{1}{2}}$ has real non-negative eigenvalues. $\mathbf{D}^{-\frac{1}{2}}\mathbf{K}\mathbf{D}^{-\frac{1}{2}}$ is related to \mathbf{W} through a diagonal similarity transformation [6]. Hence $\mathbf{D}^{-\frac{1}{2}}\mathbf{K}\mathbf{D}^{-\frac{1}{2}}$ and \mathbf{W} have the same eigenvalues. Hence, all eigenvalues of \mathbf{W} are non-negative real.

We note that \mathbf{W} is row-stochastic i.e. $\mathbf{W}\mathbf{1} = \mathbf{1}$ implying $\|\mathbf{W}\|_\infty = 1$, where $\|\cdot\|_\infty$ is the max-absolute row sum matrix norm. Now, for any eigenvalue and eigenvector pair (λ, \mathbf{x}) of \mathbf{W} we have $\mathbf{W}\mathbf{x} = \lambda\mathbf{x}$, which implies $\|\mathbf{W}\mathbf{x}\|_\infty = \lambda\|\mathbf{x}\|_\infty$. Also since this norm is self-consistent, we have $\lambda\|\mathbf{x}\|_\infty = \|\mathbf{W}\mathbf{x}\|_\infty \leq \|\mathbf{W}\|_\infty \|\mathbf{x}\|_\infty = \|\mathbf{x}\|_\infty$ or $\lambda \leq 1$.

Importantly, since the eigenvalues of such matrices are real and lie between $[0, 1]$, filters of the form (6) belong to the class of *shrinking smoothers* [19], [7], i.e. $\|\hat{\mathbf{z}}\| = \|\mathbf{W}\mathbf{y}\| \leq \|\mathbf{y}\|$. A filter that is a shrinking smoother guarantees that the estimated output is stable even when the filter is repeatedly applied to the signal.

A. Symmetric Stable Smoothers

In [7], the author presents an extensive argument for converting the filter $\mathbf{W} = \mathbf{D}^{-1}\mathbf{K}$ of (6) into a doubly stochastic matrix. He argues that such a conversion : (a) leads to improved performance; (b) provides stable filters that can be repeatedly applied; and (c) allows analysis and interpretation in terms of the eigendecomposition. This argument for deriving symmetric operators is based on the work of [20] wherein it is proved that a linear estimate of the form $\mathbf{W}\mathbf{y}$ is admissible only if \mathbf{W} is symmetric. For any non-negative \mathbf{W} it is easy to find a doubly stochastic approximation $\mathbf{S} \in \mathcal{D}_{N_p}$ where \mathcal{D}_{N_p} is the space of all $N_p \times N_p$ doubly stochastic matrices. Specifically, there exist diagonal matrices \mathbf{R} and \mathbf{C} such that $\mathbf{S} = \mathbf{R}\mathbf{W}\mathbf{C}$ [21], [22]. In practice, the solution \mathbf{S} is obtained by using the Sinkhorn iteration of alternately scaling the rows and columns of \mathbf{W} to sum to one, till convergence. Moreover, it can be shown (see Theorem 2.3 of [7]) that the change in the eigenvalues due to the Sinkhorn normalization is upper bounded by the Frobenius norm of $\|\mathbf{W} - \mathbf{S}\|$. Thus, [7] argues for converting a given filter \mathbf{W} into its symmetric doubly stochastic version \mathbf{S} via the Sinkhorn iteration.

III. THE GRAPH PERSPECTIVE OF AN IMAGE

The linear algebraic view of image denoising can be fruitfully complemented by a graph-theoretic perspective. Recalling that the kernel $K(i, j)$ represents the similarity between pixels i and j , we can recognize that \mathbf{K} also provides an interpretation of the structure of the image as a graph which has several advantages. Consider the graph $\mathcal{G} = (V, E)$ where V is the set of all pixels in the image and $E \subseteq V \times V$ represents the edges which encapsulate the relationships between pairs of individual image pixels. In our case, choosing the edge weights to be given by

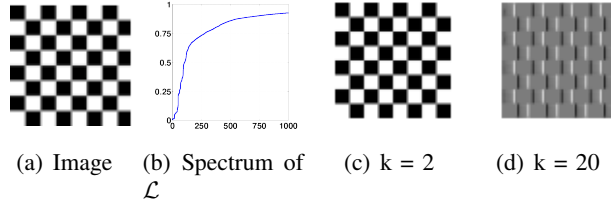


Fig. 1: Binary Checkerboard image, normalized Laplacian spectrum of the smallest 1000 singular values.

kernel $K(i, j)$, we can encode all similarity relationships into a single graph \mathcal{G} . Many familiar data-adaptive image denoising methods can be represented in a kernel form, i.e.

$$[\mathbf{K}]_{ij} = \exp\left(-\frac{d^2(i, j)}{h^2}\right) \quad (7)$$

where $d(i, j)$ is a measure of the distance between pixels i and j and h is a tuning parameter representing the ‘scale’ of the distance values $d(\cdot, \cdot)$. These include the isotropic Gaussian kernel \mathbf{K}_G , the bilateral filter \mathbf{K}_{bl} [4], the LARK filter \mathbf{K}_{LARK} [17] and the non-local filter \mathbf{K}_{nl} [5].

IV. THE GRAPH LAPLACIAN

The Laplacian of a graph plays an important role for representing functions on the graph [23]. Some examples of image denoising in the graph spectral domain are [8], [24], [9]. The Laplacian can be defined as follows : For an undirected graph, the degree of vertex i is denoted as d_i and is equal to the sum of the weights on the edges connected to this vertex. Evidently, for a graph with an adjacency matrix \mathbf{K} , we have $d_i = \sum_j K(i, j)$. The corresponding degree matrix \mathbf{D} is a diagonal matrix where the entries on the diagonal are given by $D(i, i) = d_i$. While the unnormalized graph Laplacian is given by $\mathbf{L} = \mathbf{D} - \mathbf{K}$, there are multiple definitions for a normalized graph Laplacian. In this paper, we use the asymmetric form that has a random walk interpretation and is given by

$$\mathcal{L} = \mathbf{D}^{-1}\mathbf{L} = \mathbf{I} - \mathbf{D}^{-1}\mathbf{K} = \mathbf{I} - \mathbf{W}. \quad (8)$$

The nature of the graph is encoded by the eigen-structure of \mathbf{W} or equivalently that of \mathcal{L} which has a number of interesting properties. We state some of the relevant properties of \mathcal{L} here and refer the reader to [23] for proofs and more extensive discussions:

- The multiplicity k of the value 0 of \mathcal{L} equals the number of connected components A_1, A_2, \dots, A_k in the graph.
- The nullspace of \mathcal{L} is spanned by the connected component indicator vectors $\mathbf{1}_{A_1}, \mathbf{1}_{A_2}, \mathbf{1}_{A_3}, \dots, \mathbf{1}_{A_k}$ for each of these components $A_1, A_2, A_3, \dots, A_k$ respectively.

For instance, for the checkerboard image in Fig. 1(a), if we used a kernel $K(i, j) = \delta(y_j - y_i)$ that assigns a similarity measure of 1 for all pixels of the same intensity (black or white), then the corresponding nullspace of \mathcal{L} would be a matrix of rank 2 with corresponding eigenvectors that would span the space of all possible checkerboard images where the black or white pixels could independently take on any intensity.

A. Spectrum of the normalized Laplacian

In an ideal sense, to denoise an image we would like to represent it in a basis that separates out the signal and noise components of the input as best as possible. Many denoising methods are based on computing the appropriate basis where such separation is best achieved. These methods project the input image onto this basis and estimate the underlying image by shrinking the magnitude of those projected components that are deemed to be noise according to a specified image modeling criterion. Amongst such shrinkage approaches, threshold methods are very popular

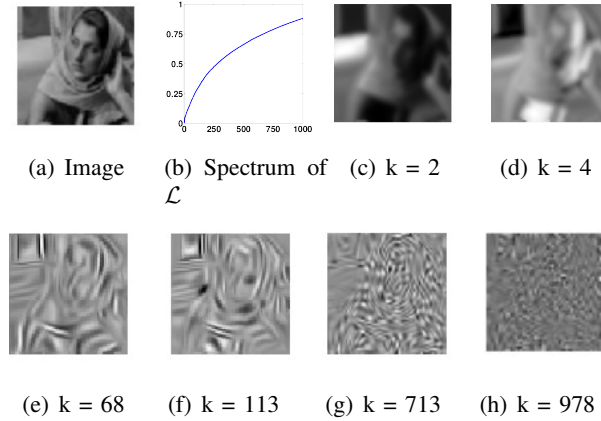


Fig. 2: Face from the Barbara image, normalized Laplacian spectrum of the smallest 1000 singular values.

[25], [26], [27]. In this sense, the basis defined by the eigenvectors of the normalized graph Laplacian \mathcal{L} plays a special role. Specifically, the smallest eigenvectors of \mathcal{L} can be interpreted as capturing different structures present in the image whereas the eigenvectors with the largest eigenvalues represent the noise present in the image. We illustrate this property on a binary and a natural image. In Fig. 1, we present a checkerboard binary image of size 80×80 pixels, i.e. $N_p = 6400$. We construct a kernel $\mathbf{K} = \mathbf{K}_{bl} + \mathbf{K}_{nl}$ that combines both local (bilateral) and non-local information. Fig. 1(b) illustrates the 1000 smallest eigenvalues of \mathcal{L} . As can be observed, the structure of the image is captured in the smallest eigenvectors of \mathcal{L} and denoising can be performed by projecting the observed signal into the subspace spanned by these vectors. Such an interpretation is also obtained when we examine a natural image in a similar manner in Fig. 2. As can be observed from the illustrative eigenvectors shown here, the eigenvectors corresponding to the smallest eigenvalues encode the macroscopic structure of the image. With increasing index, the spectrum encodes finer spatial structures in the image resulting in a spectral decomposition of \mathcal{L} that encodes significant structures such as strong edges, smooth regions, fine-grained details etc. Thus, the smallest eigenvectors of \mathcal{L} provide a suitable *basis* for representing the denoised image. In [28], the authors studied the stability of the eigenvectors of the Laplacian with respect to perturbations. While [28] used the symmetric form of the normalized Laplacian, i.e. $\mathbf{I} - \mathbf{D}^{-\frac{1}{2}} \mathbf{W} \mathbf{D}^{-\frac{1}{2}}$, their general inferences also hold in our case. They noted that the smallest eigenvectors of \mathcal{L} are stable even when a large amount of noise is added to the image. This suggests that these eigenvectors should play a dominant role in the estimation of the denoised image.

V. CONSISTENCY IN MULTI-DIMENSIONAL FILTERING

With these preliminaries we now introduce our notion of consistency in the context of image filtering. From (3), we observe that for a given weighting matrix \mathbf{W} , in the estimate \hat{z}_i the contributions of the j -th and k -th pixels are weighted by W_{ij} and W_{ik} respectively. Therefore, in \hat{z}_i the relative contributions of the j -th and k -th pixels are in the ratio $K(i, j) : K(i, k)$. However, as can be seen from (6), in the linear estimate $\hat{\mathbf{z}} = \mathbf{W}\mathbf{y}$, each pixel intensity in $\hat{\mathbf{z}}$ is estimated independently as a dot-product between \mathbf{y} and the corresponding row of \mathbf{W} . In other words, while we carry out the filtering operation using \mathbf{W} that encodes a set of similarity relationships between pairs of pixels, we do not ensure that the estimated image $\hat{\mathbf{z}}$ maintains the same set of similarity relationships as given by \mathbf{W} . Thus, the linear filtering method of (6) is *inconsistent* with the relationships encoded by \mathbf{W} . Consider the case where two pixels i and j (and their corresponding neighborhoods) in the observed image \mathbf{y} are identical to each other. The inconsistent filtering of (6) does not guarantee that \hat{z}_i and \hat{z}_j (and their corresponding neighborhoods) will remain identical to each other. We may remedy this problem by requiring our estimation method to be consistent in the sense stated above.

Consistent filtering of an image requires that all geometric relationships encoded in the weight matrix \mathbf{W} be preserved in the filtered output $\hat{\mathbf{z}}$. To derive such a consistent filter, we note that in practice, most filtering operations

are data-dependent and \mathbf{W} is itself derived from the image being filtered. Consistency of image filtering demands that when we filter the estimate $\hat{\mathbf{z}}$, the resulting image should preserve all the pairwise pixel intensity relationships present in $\hat{\mathbf{z}}$ and encoded by \mathbf{W} which is data-dependent and equal to $\mathbf{W}(\hat{\mathbf{z}})$. Therefore, our consistency constraint is given by

$$\hat{\mathbf{z}} = \mathbf{W}(\hat{\mathbf{z}})\hat{\mathbf{z}} = \mathbf{W}\hat{\mathbf{z}}. \quad (9)$$

Here, we have simplified our representation for notational simplicity, while recognizing that \mathbf{W} is itself derived from the image. The consistency constraint tells us that the filtered image has the same intensity relationships as encoded by \mathbf{W} , and no amount of further filtering will change this condition. If \mathbf{W} requires that a certain group of pixels are similar, then the filtered image must observe this relationship irrespective of the effect of filtering. The consistency constraint also motivates a transformation of the observed signal whereby the noise and signal are completely separated, since any amount of further filtering does not change the image, i.e. $\hat{\mathbf{z}} = \mathbf{W}\hat{\mathbf{z}} = \mathbf{W}^2\hat{\mathbf{z}} = \mathbf{W}^3\hat{\mathbf{z}} = \dots$. Our notion of consistency implies that we parameterize the individual pixel intensities by a linear model given as

$$z_i = \sum_{j \in \mathcal{N}(i)} W(i, j) z_j \quad (10)$$

where $\mathcal{N}(i)$ is the set of local and non-local neighbors of the i -th pixel as defined by \mathbf{W} . We note that this approach of local linear fitting of the data is similar in spirit to dimensionality reduction methods which seek to preserve the local linear structure of the empirical manifold representation of observed data [29]. In our context, we fit the data locally through kernel regression to learn a linear representation of a manifold of which the estimated image $\hat{\mathbf{z}}$ is a member.

A. Strict Consistency

Strictly enforcing our notion of consistency implies that

$$\hat{\mathbf{z}} = \mathbf{W}\hat{\mathbf{z}} \Rightarrow (\mathbf{I} - \mathbf{W})\hat{\mathbf{z}} = \mathcal{L}\hat{\mathbf{z}} = \mathbf{0}. \quad (11)$$

This consistency criterion in combination with the noise model in (2) yields a maximum likelihood estimate for the denoised image $\hat{\mathbf{z}}$. Using Bayes rule, we can write the conditional likelihood of the estimate $\hat{\mathbf{z}}$ in terms of the observed image \mathbf{y} as

$$p(\mathbf{z}|\mathbf{y}) \propto p(\mathbf{y}|\mathbf{z})p(\mathbf{z}). \quad (12)$$

While $p(\mathbf{y}|\mathbf{z})$ is defined by the *i.i.d.* Gaussian noise assumption in (1), the prior $p(\mathbf{z})$ is defined by our assumption of the nature of the true image \mathbf{z} . Since strictly enforcing the consistency constraint means that the denoised image estimate $\hat{\mathbf{z}}$ lies in the nullspace of \mathcal{L} , maximizing likelihood is equivalent to the constrained minimization problem

$$\arg \min_{\mathbf{z}} \|\mathbf{z} - \mathbf{y}\|^2 \text{ such that } \mathcal{L}\mathbf{z} = \mathbf{0}. \quad (13)$$

Therefore, here we are implicitly assuming that $p(\mathbf{z}) \propto \delta(\|\mathcal{L}\mathbf{z}\|)$ and we can see that any estimate $\hat{\mathbf{z}}$ has to be a linear combination of the indicator vectors of each connected component, i.e.

$$\hat{\mathbf{z}} = \alpha_1 \mathbf{1}_{\mathbf{A}_1} + \alpha_2 \mathbf{1}_{\mathbf{A}_2} + \alpha_3 \mathbf{1}_{\mathbf{A}_3} + \dots + \alpha_k \mathbf{1}_{\mathbf{A}_k} \quad (14)$$

where the coefficients α_i , $i = 1, 2, \dots, k$ are the intensity values of each of the individual connected components. While, any solution of the form in (14) is equally likely in terms of $p(\mathbf{z})$, maximizing the likelihood in (12) specifies the choice of the individual intensity values for each connected component, i.e. α_i . It will be immediately recognized that the optimal coefficient for the c -th connected component is given as $\alpha_c = \frac{\mathbf{y}^T \mathbf{1}_{\mathbf{A}_c}}{\|\mathbf{1}_{\mathbf{A}_c}\|}$. Such a simple piecewise constant solution has an important role in image segmentation problems and can also be seen to be a specific simple form of the piecewise-smooth Mumford-Shah functional used in variational approaches to image segmentation [30].

B. Relaxed Consistency

Although (14) is the optimal solution for (13), the solution of piecewise constant images is only useful for image segmentation. However, this is excessively restrictive for our problem of denoising natural images which are far from being piecewise constant. In fact, for natural images, the graph Laplacian often has a single connected component of the whole image, implying that the only solution given by (14) is a constant image with intensity equal to the mean intensity of the pixels in \mathbf{y} .

A remedy to this restrictive scenario is to replace our strict consistency criterion of $\mathbf{z} = \mathbf{W}\mathbf{z}$ by a more general consistency model

$$\mathbf{z} = \mathbf{W}\mathbf{z} + \epsilon(\mathbf{z}) \quad (15)$$

where $\epsilon(\mathbf{z})$ signifies the degree to which any estimated image deviates from the consistency requirement specified by \mathbf{W} and the argument \mathbf{z} emphasizes that this measure of inconsistency is dependent on \mathbf{z} . For relaxing the strict consistency criterion, we need to specify an image prior $p(\mathbf{z})$ that applies to natural images. The literature on the statistics of natural images is large and the likelihood models in terms of image derivatives have been extensively studied and utilized in a variety of image processing problems [31], [32]. However, here we emphasize that $p(\mathbf{z})$ needs to capture our model assumptions specified by the consistency criterion of (15). In other words, in our problem we do not seek to model the distribution of *all* natural images but instead restrict ourselves to modeling those images that satisfy the linear inter-pixel similarity relationships specified by \mathbf{W} . While this still leaves us with a variety of possible choices, with the aim of designing a linear filter in mind, we specify the form of $p(\mathbf{z})$ with an empirical Bayes prior of

$$p(\mathbf{z}) \propto e^{-\frac{1}{2\gamma^2} \|\epsilon(\mathbf{z})\|^2} \quad (16)$$

where γ is a constant. Noting that $\epsilon(\mathbf{z}) = \mathbf{z} - \mathbf{W}\mathbf{z} = \mathcal{L}\mathbf{z}$, we have the maximum likelihood estimate of \mathbf{z} as the solution of the unconstrained minimization problem

$$\arg \min_{\mathbf{z}} \|\hat{\mathbf{z}} - \mathbf{y}\|^2 + \lambda \|\mathcal{L}\hat{\mathbf{z}}\|^2 \quad (17)$$

where $\lambda = \frac{\sigma^2}{\gamma^2}$. The cost function in (17) can be seen to be a regularized cost function and in Sec. VIII we will compare it with the conventional form of Laplacian regularization.

In (17), we notice that in the regularization term $\|\mathcal{L}\hat{\mathbf{z}}\|$, all deviations for individual pixels are penalized uniformly. However, with our purpose of image denoising in mind it would be useful to penalize the regularization term for each pixel according to an appropriate weighting criterion. This can be achieved by a pixel-dependent weight factor in the penalty term of $\|\mathbf{\Lambda}\mathcal{L}\hat{\mathbf{z}}\|$ where $\mathbf{\Lambda}$ is a diagonal matrix of weights. Thus, the cost function to be minimized for consistent image filtering is given as

$$\begin{aligned} E(\mathbf{z}) &= \|\mathbf{z} - \mathbf{y}\|^2 + \lambda \|\mathbf{\Lambda}(\mathcal{L}\mathbf{z})\|^2 \\ &= \|\mathbf{z} - \mathbf{y}\|^2 + \lambda \mathbf{z}^T \mathcal{L}^T \mathbf{\Lambda}^2 \mathcal{L} \mathbf{z} \end{aligned} \quad (18)$$

which when differentiated with respect to \mathbf{z} and set to zero, leads to

$$\begin{aligned} (\mathbf{I} + \lambda \mathcal{L}^T \mathbf{\Lambda}^2 \mathcal{L}) \hat{\mathbf{z}} &= \mathbf{y} \\ \Rightarrow \hat{\mathbf{z}} &= (\mathbf{I} + \lambda \mathcal{L}^T \mathbf{\Lambda}^2 \mathcal{L})^{-1} \mathbf{y}. \end{aligned} \quad (19)$$

While in Eqn. 19, we can choose any appropriate weighting $\mathbf{\Lambda}$, for simplicity, in this paper we set it to the degree of each pixel, i.e. $\mathbf{\Lambda} = \mathbf{D}$. Since $\mathbf{\Lambda}\mathcal{L} = \mathbf{D}(\mathbf{I} - \mathbf{D}^{-1}\mathbf{K}) = (\mathbf{D} - \mathbf{K}) = \mathbf{L}$, where \mathbf{L} is the un-normalized Laplacian of \mathcal{G} , we can immediately see that

$$\hat{\mathbf{z}} = (\mathbf{I} + \lambda \mathbf{L}^T \mathbf{L})^{-1} \mathbf{y} = \mathbf{C} \mathbf{y} \quad (20)$$

where $\mathbf{C} = (\mathbf{I} + \lambda \mathbf{L}^T \mathbf{L})^{-1}$ is our *consistency* filter for image denoising. We note here that the weighting used ($\mathbf{\Lambda} = \mathbf{D}$) in deriving Eqn. (20) means that the diffusion caused by applying the filter \mathbf{C} depends on the degree of

the pixel in the graph \mathcal{G} . If a pixel has a high degree, i.e. it is deemed to be similar to many pixels, then we allow a higher rate of diffusion. It will be noted that in practice we do not need to explicitly compute the filter \mathbf{C} , rather we estimate the denoised image $\hat{\mathbf{z}}$ by solving the sparse linear system of equations given by $(\mathbf{I} + \lambda \mathbf{L}^T \mathbf{L}) \hat{\mathbf{z}} = \mathbf{y}$. In our approach, we efficiently solve for $\hat{\mathbf{z}}$ using the bi-conjugate gradient stabilized method for optimization. Our filtering approach can be schematically represented as given in algorithm 1.

Algorithm 1 Denoising using consistency filter \mathbf{C}

Require: Noisy image \mathbf{y} , noise variance σ^2 .

 Compute non-local kernel matrix \mathbf{K}_{nl} .

 Compute bilateral kernel matrix \mathbf{K}_{bl} .

 Set $\mathbf{K} \leftarrow \mathbf{K}_{nl} + \mathbf{K}_{bl}$.

 Compute degree matrix $\mathbf{D} \leftarrow \text{diag}(\mathbf{K})$.

 Compute Laplacian matrix $\mathbf{L} \leftarrow \mathbf{D} - \mathbf{K}$.

 Solve $(\mathbf{I} + \lambda \mathbf{L}^T \mathbf{L}) \hat{\mathbf{z}} = \mathbf{y}$ for $\hat{\mathbf{z}}$.

return Denoised image $\hat{\mathbf{z}}$.

VI. PROPERTIES OF FILTER \mathbf{C}

We can now specify and examine the properties of our image denoising filter \mathbf{C} that is based on a relaxed form of consistency. For a graph Laplacian \mathbf{L} specified by a weight matrix \mathbf{W} , we note that given the singular value decomposition $\mathbf{L} = \mathbf{V} \mathbf{M} \mathbf{U}^T$, we have $\mathbf{L}^T \mathbf{L} = \mathbf{U} \mathbf{M}^2 \mathbf{U}^T$ where \mathbf{M} is a diagonal matrix with $M_{ii} = \mu_i$. Here, we remark that although in this paper we derive \mathbf{W} from a kernel \mathbf{K} , in its most general form \mathbf{K} need not be a symmetric matrix, e.g. the LARK filter [17].

Theorem 1 (Properties). *The consistency filter matrix $\mathbf{C} = (\mathbf{I} + \lambda \mathbf{L}^T \mathbf{L})^{-1}$ has the following properties*

- \mathbf{C} is symmetric.
- \mathbf{C} is a generalized doubly stochastic matrix.
- For μ_i as the eigenvalues of \mathbf{L} , the eigenvalues of \mathbf{C} are given as $\lambda_i = \frac{1}{1 + \lambda \mu_i^2}$.
- The eigenvalues λ_i are bounded as $0 < \lambda_i \leq 1$.
- For a given image \mathbf{y} and \mathbf{C} , the denoised image $\hat{\mathbf{z}}$ is given as

$$\hat{\mathbf{z}} = \sum_{i=1}^{N_p} \frac{1}{1 + \lambda \mu_i^2} (\mathbf{u}_i^T \mathbf{y}) \mathbf{u}_i \quad (21)$$

where \mathbf{u}_i are columns of \mathbf{U} in order corresponding to λ_i .

Proof :

- Since \mathbf{C} is the inverse of a symmetric matrix $(\mathbf{I} + \lambda \mathbf{L}^T \mathbf{L})$, \mathbf{C} is symmetric.
- From the definition of \mathbf{C} , we have

$$\begin{aligned} \mathbf{C}^{-1} \mathbf{1} &= (\mathbf{I} + \lambda \mathbf{L}^T \mathbf{L}) \mathbf{1} \\ &= \mathbf{1} + \lambda \mathbf{L}^T \mathbf{L} \mathbf{1} = \mathbf{1} \end{aligned} \quad (22)$$

as the graph Laplacian \mathbf{L} always has an eigen-value of 0 for all indicator vectors $\mathbf{1}_{\mathbf{A}_k}$ (Proposition 2, [23]). Therefore, $\mathbf{1} = \sum_k \mathbf{1}_{\mathbf{A}_k}$ is an eigenvector of \mathbf{L} with eigen-value of 0 implying that $\mathbf{L} \mathbf{1} = \mathbf{0}$. From (22), we have $\mathbf{C}^{-1} \mathbf{1} = \mathbf{1}$ and $\mathbf{C} \mathbf{C}^{-1} \mathbf{1} = \mathbf{1}$, we conclude that $\mathbf{C} \mathbf{1} = \mathbf{1}$, i.e. \mathbf{C} has unit row-sum. Since \mathbf{C} is symmetric, it also has unit column-sum. However, since \mathbf{C} is not guaranteed to be a non-negative matrix, it is seen to be a generalized doubly stochastic matrix.

- We have

$$\begin{aligned} \mathbf{C} &= (\mathbf{I} + \lambda \mathbf{L}^T \mathbf{L})^{-1} \\ &= (\mathbf{U} (\mathbf{I} + \lambda \mathbf{M}^2) \mathbf{U}^T)^{-1} = \mathbf{U} (\mathbf{I} + \lambda \mathbf{M}^2)^{-1} \mathbf{U}^T. \end{aligned} \quad (23)$$

Thus, the eigenvalues of \mathbf{C} are given as $\lambda_i = \frac{1}{1 + \lambda\mu_i^2}$.

d) Also, for $\lambda > 0$, we have $0 < \frac{1}{1 + \lambda\mu_i^2} \leq 1$ implying $0 < \lambda_i \leq 1$.

e) From (23), we obtain

$$\hat{\mathbf{z}} = \mathbf{C}\mathbf{y} = \sum_{i=1}^{N_p} \frac{1}{1 + \lambda\mu_i^2} (\mathbf{u}_i^T \mathbf{y}) \mathbf{u}_i. \quad (24)$$

While \mathbf{W} may possibly be asymmetric, our consistency filter \mathbf{C} is always symmetric and is also a shrinking smoother as its eigen-values are bounded between $(0, 1]$. It is a known result that for an unweighted graph \mathcal{G} , the operator $(\mathbf{I} + \mathbf{L})^{-1}$ is doubly stochastic [33]. However, in our more general case, \mathbf{C} is not a non-negative matrix, i.e. \mathbf{C} is a generalized doubly stochastic matrix. In other words, our consistency filter \mathbf{C} satisfies the desiderata laid out in [7] for converting a given averaging filter \mathbf{W} into a symmetric, smoothing form that is stable. We also remark that the lack of non-negativity of \mathbf{C} is advantageous in that it generally allows for reducing the bias of the estimate.

Most importantly, we remark that the eigenvalues $\lambda_i = (1 + \frac{\sigma^2}{s^2})^{-1}$ have the form of a Wiener filter which minimizes the mean squared error for stationary signals. From (24) we recognize that the consistency filter \mathbf{C} projects the input signal \mathbf{y} onto the basis \mathbf{U} and applies a one-dimensional Wiener filter to the projected signal along the eigenvectors \mathbf{u}_i . Using (24) we can see that for the component projected onto the i -th basis vector \mathbf{u}_i , $SNR_i = \frac{s^2}{\sigma^2} = \frac{1}{\lambda\mu_i^2}$. Now, since $\lambda \propto \sigma^2$, we have $s^2 = \frac{K}{\mu_i^2}$ for some $K > 0$. In summary, the consistency filter \mathbf{C} uses the graph Laplacian to project the input image onto a data-dependent basis and then carries out Wiener filtering on the individual projected components. Recalling that the eigen decomposition of \mathbf{L} assigns smaller eigenvalues μ_i to the signal terms and larger μ_i to the noise terms, we can see that (24) has the desired property that the noise terms are shrunk to a larger degree than the signal terms resulting in good denoising performance. Finally we note that if \mathbf{L} has multiple connected components, their corresponding eigenvalues μ_i are zero, implying that (24) preserves the corresponding connected components in the image without any shrinkage and reduces the magnitude of the noise terms.

VII. BIAS-VARIANCE TRADEOFF IN \mathbf{C}

Since our consistency filter \mathbf{C} satisfies the properties of symmetry and stability, it would be useful to compare its performance relative to the doubly stochastic filter \mathbf{S} proposed in [7]. In this Section, we carry out such a comparison by looking at the mean squared error (MSE) of \mathbf{S} and our filter \mathbf{C} . Similar to the filter analysis done in [7], under small noise assumption, the following theorem leads to the explanation of the bias-variance trade-off of our filter \mathbf{C} for different amounts of regularization λ .

Theorem 2 (Bias, variance and mse). *Under small noise assumption such that the filter matrix \mathbf{C} is not perturbed significantly and that the true image \mathbf{z} be represented as $\mathbf{z} = \mathbf{U}\mathbf{b}$ for some \mathbf{b} , the bias and variance of the filter $\hat{\mathbf{z}} = \mathbf{C}\mathbf{y}$ are given respectively as*

- a) $\|bias(\hat{\mathbf{z}})\|_2^2 = \sum_{i=1}^{N_p} \left(\frac{\lambda\mu_i^2}{1 + \lambda\mu_i^2} \right)^2 b_i^2$.
- b) $var(\hat{\mathbf{z}}) = \sigma^2 \sum_{i=1}^{N_p} \left(\frac{1}{1 + \lambda\mu_i^2} \right)^2$ where σ^2 is the per-pixel noise variance.
- c) Hence the mean squared error is given as

$$mse(\hat{\mathbf{z}}) = \sum_{i=1}^{N_p} \left(\frac{\lambda\mu_i^2}{1 + \lambda\mu_i^2} \right)^2 b_i^2 + \sigma^2 \sum_{i=1}^{N_p} \left(\frac{1}{1 + \lambda\mu_i^2} \right)^2. \quad (25)$$

Proof: We consider the filter $\hat{\mathbf{z}} = \mathbf{C}\mathbf{y}$ where $\mathbf{y} = \mathbf{z} + \mathbf{n}$ where $\mathbf{n} \sim \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{I})$ and σ is small.

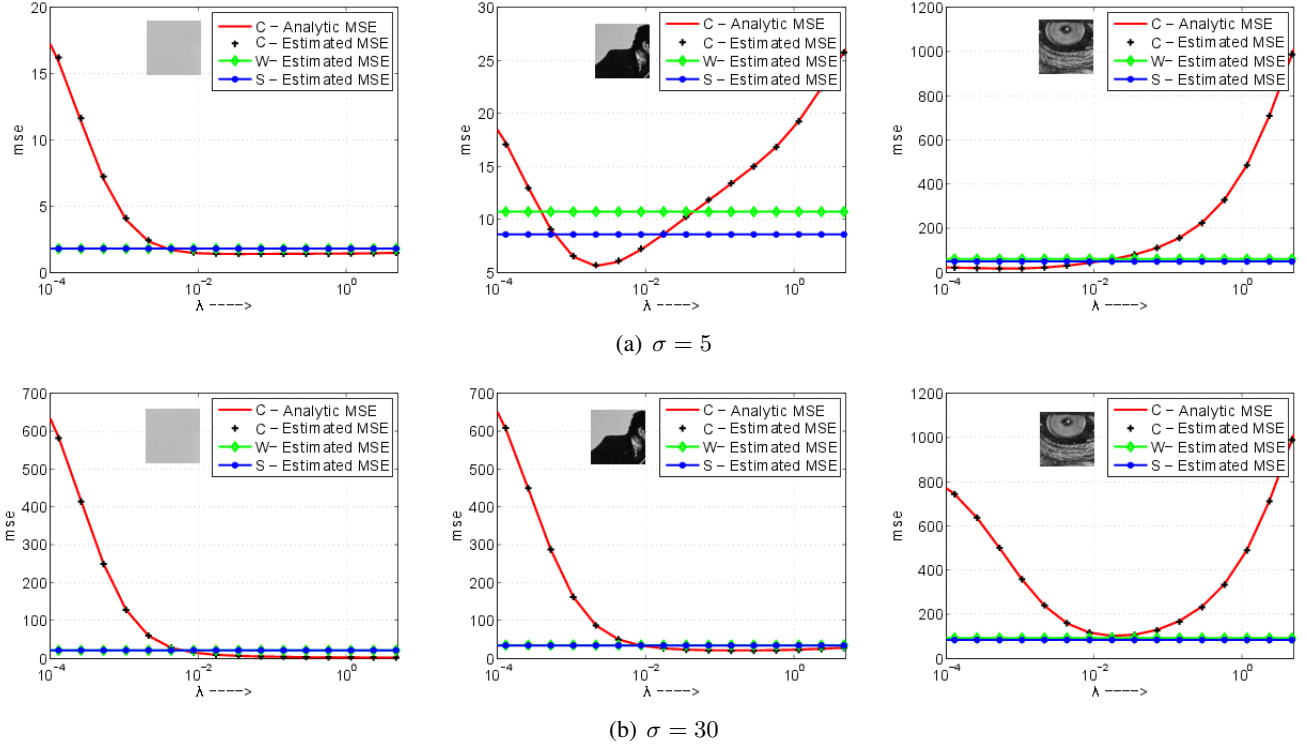


Fig. 3: Comparison of the mean squared error (MSE) of our consistency filter \mathbf{C} and those of the filter \mathbf{W} and its symmetric doubly stochastic version \mathbf{S} . We compare performance on three image patches as indicated and for two different noise levels. See text for details.

a)

$$\begin{aligned} \text{bias}(\hat{\mathbf{z}}) &= \mathbb{E}(\hat{\mathbf{z}}) - \mathbf{z} \\ &= \mathbb{E}(\mathbf{C}\mathbf{y}) - \mathbf{z} = \mathbb{E}(\mathbf{C}\mathbf{z} + \mathbf{C}\mathbf{n}) - \mathbf{z} \approx \mathbf{C}\mathbf{z} - \mathbf{z}. \end{aligned}$$

$$\begin{aligned} \text{Therefore, } \|\text{bias}(\hat{\mathbf{z}})\|_2^2 &= \|(\mathbf{C} - \mathbf{I})\mathbf{z}\|_2^2 \\ &= \left\| \left(\mathbf{U}(\mathbf{I} + \lambda\mathbf{M}^2)^{-1}\mathbf{U}^T - \mathbf{I} \right) \mathbf{U}\mathbf{b} \right\|_2^2 \\ &= \left\| \text{diag} \left(\frac{1}{1 + \lambda\mu_i^2} - 1 \right) \mathbf{b} \right\|_2^2 = \sum_{i=1}^{N_p} \left(\frac{\lambda\mu_i^2}{1 + \lambda\mu_i^2} \right)^2 b_i^2. \end{aligned}$$

b) We have the covariance as

$$\begin{aligned} \text{cov}(\hat{\mathbf{z}}) &= \text{cov}(\mathbf{C}\mathbf{y}) = \text{cov}(\mathbf{C}(\mathbf{z} + \mathbf{n})) \\ &\approx \sigma^2 \mathbf{C}\mathbf{C}^T \\ &= \sigma^2 \mathbf{U}(\mathbf{I} + \lambda\mathbf{M}^2)^{-2} \mathbf{U}^T. \end{aligned}$$

$$\text{Thus, } \text{var}(\hat{\mathbf{z}}) = \sigma^2 \sum_{i=1}^{N_p} \left(\frac{1}{1 + \lambda\mu_i^2} \right)^2.$$

c) From above, the mean squared error is given as

$$\begin{aligned} \text{mse}(\hat{\mathbf{z}}) &= \|\text{bias}(\hat{\mathbf{z}})\|_2^2 + \text{var}(\hat{\mathbf{z}}) \\ &= \sum_{i=1}^{N_p} \left(\frac{\lambda\mu_i^2}{1 + \lambda\mu_i^2} \right)^2 b_i^2 + \sigma^2 \sum_{i=1}^{N_p} \left(\frac{1}{1 + \lambda\mu_i^2} \right)^2. \end{aligned}$$

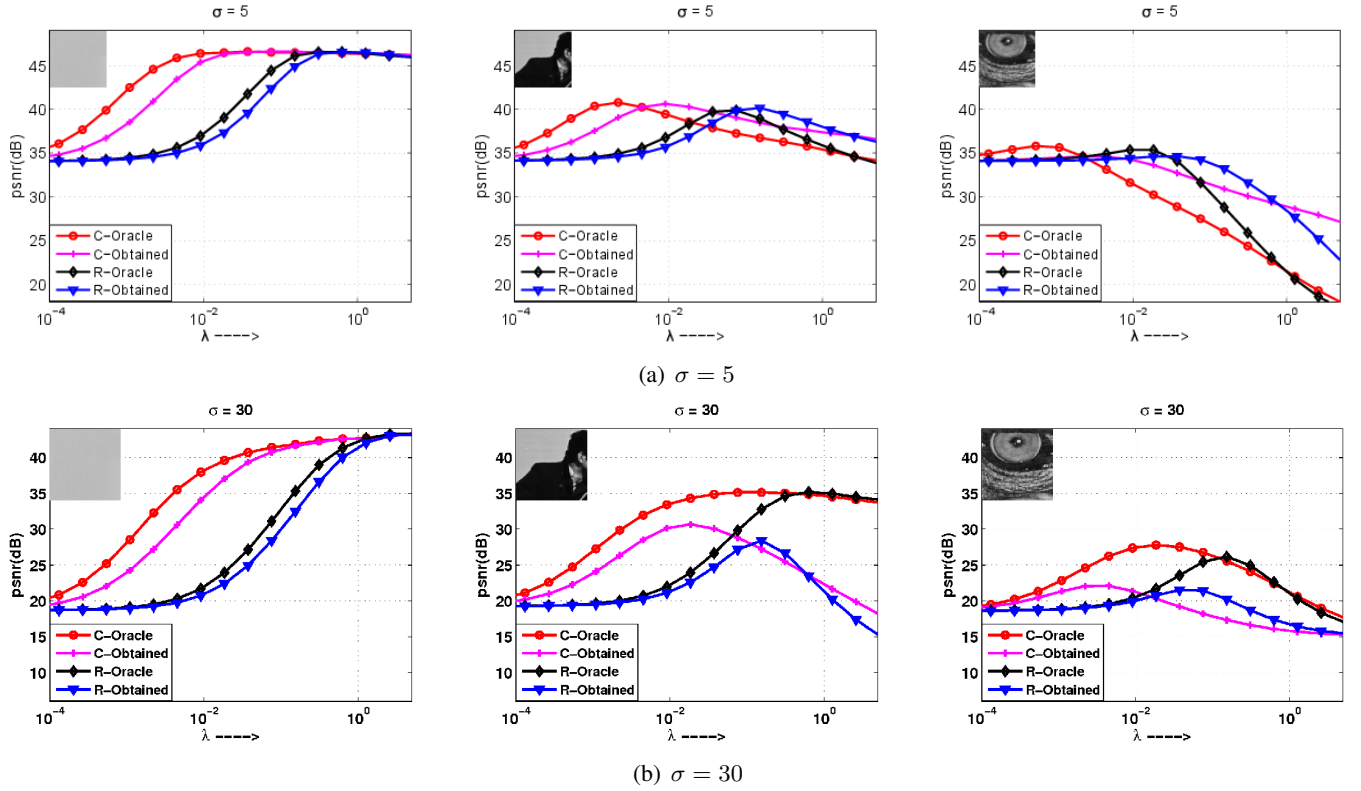


Fig. 4: Comparison of performance our consistency filter \mathbf{C} and the regularized Laplacian filter \mathbf{R} for different types of patches.

In Fig. 3, we present the results of comparing the mean squared error (MSE) performance of our consistency filter \mathbf{C} , the usual averaging filter \mathbf{W} and its symmetric doubly stochastic version \mathbf{S} . It is useful to recall here that both \mathbf{S} and \mathbf{C} are derived from \mathbf{W} . For our experiments here, we used the local bilateral kernel \mathbf{K}_{bl} for the true image, i.e. \mathbf{W} is an oracle filter. As the performance of our filter \mathbf{C} depends on λ , we plot the MSE as a function of λ . Since, both \mathbf{W} and \mathbf{S} are independent of λ , their MSE values are plotted as constant values in Fig. 3. For the constant patch image, as expected the performance of our filter \mathbf{C} improves with increasing λ since higher amounts of diffusion effectively smoothens the noisy patch. Here we note that our filter achieves MSE values that are lower than that of both \mathbf{W} and \mathbf{S} . For the edge-like image patch (part of the familiar cameraman image) shown in the middle column of Fig. 3, we note that the performance of consistency filter \mathbf{C} is superior to both \mathbf{W} and \mathbf{S} over a certain range of values for λ . It is noteworthy that this superior performance of \mathbf{C} is valid for an interval that spans one order of magnitude for the regularization term λ , i.e. our method is fairly robust to the choice of λ . Finally, in the third column of Fig. 3 we consider the denoising performance for a textured image patch which has very different properties compared to the flat and edge patches. For the low noise case of $\sigma = 5$, we note that for moderate values of λ , our filter outperforms both \mathbf{W} and \mathbf{S} . Expectedly, increasing λ causes increased diffusion that blurs the fine-scale details of the textured patch, thereby increasing the MSE of our filter. In the high noise case of $\sigma = 30$, we note that we need to apply a significant amount of diffusion to achieve the best possible performance of our filter. However, in this case, we also note that the lowest MSE for our consistency filter \mathbf{C} filter is marginally poorer than the performance of \mathbf{W} and \mathbf{S} .

We may draw some general inferences from the observations based on Fig. 3. While converting \mathbf{W} into its doubly stochastic version \mathbf{S} as proposed in [7] leads to better theoretical properties, it does not necessarily lead to improved empirical performance. This is also implicit in the fact that applying the Sinkhorn iteration to \mathbf{W} does not change its eigenvalues to any significant degree (see Theorem 2.4 of [7]). Intuitively, one may argue that while this assures us that \mathbf{S} is not very far from \mathbf{W} in a Frobenius norm sense, it also implies that the MSE cannot change significantly. However, if we wish to significantly improve on denoising performance when compared to \mathbf{W} ,

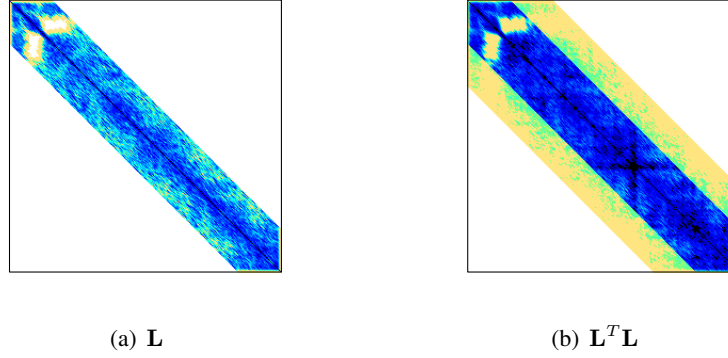


Fig. 5: Matrix visualizations of \mathbf{L} and $\mathbf{L}^T \mathbf{L}$ for Lena image. Note the increased number of neighbors in $\mathbf{L}^T \mathbf{L}$.

we would like to significantly modify the spectrum of our filter while ensuring the desirable theoretical properties of symmetry and stability. As we will show in subsequent sections, this is achieved by our consistency filter \mathbf{C} . Here, we note that in general, our consistency filter \mathbf{C} achieves better performance compared to both \mathbf{W} and \mathbf{S} . Moreover, this improved performance is valid for a fairly large range of λ (indicated on a log scale in Fig. 3), implying that our improved performance is not sensitive to the degree of regularization. We conclude our discussion in this section by remarking that all of the above analysis and interpretation holds for the oracle \mathbf{W} . In practice, we estimate \mathbf{W} from the observed noisy image or a pilot version, hence the effective performance will not only depend on the shrinkage properties of our filter but is also dependent on the robustness of our filter to errors in estimating \mathbf{W} . This is particularly true for a non-local kernel \mathbf{K}_{nl} since a practical implementation often uses a fixed number of nearest neighbors to determine the neighborhood connectivity in \mathcal{G} which can be significantly affected by the presence of noise.

VIII. COMPARISON WITH LAPLACIAN REGULARIZATION








While \mathbf{C} belongs to the class of symmetric, stable filters, as noted earlier, \mathbf{C} also belong to the class of regularized estimators. In this section, we consider the properties and performance of our filter \mathbf{C} as a form of graph Laplacian regularization. While our consistency filter in Eqn. (20) can be seen to be a regularizer on the graph Laplacian, this cost function is quite different from the conventional method of regularizing data using a graph Laplacian [34], [8], [24]. Specifically, the conventional graph Laplacian regularizer minimizes a cost function that penalizes the divergence of the gradient of the estimate $\hat{\mathbf{z}}$ with respect to the graph Laplacian, i.e.

$$\arg \min_{\hat{\mathbf{z}}} \|\hat{\mathbf{z}} - \mathbf{y}\|^2 + \lambda \hat{\mathbf{z}}^T \mathbf{L} \hat{\mathbf{z}} \quad (26)$$

which yields a graph Laplacian regularizer $\mathbf{R} = (\mathbf{I} + \lambda \mathbf{L})^{-1}$ [34], [8]. We note that the corresponding shrinkage coefficients for \mathbf{R} are given as $\frac{1}{1 + \lambda \mu_i}$. In [35], the authors apply Laplacian regularization to a bilateral filter graph. While the eventual form of the filter in [35] is similar to ours, we note here that their filter uses the normalized Laplacian form \mathcal{L} as opposed to our choice of the unnormalized version \mathbf{L} . In general, the performance of using \mathcal{L} will be significantly poorer than that of \mathbf{L} owing to the lack of appropriate weighting of the contribution of individual pixels to the overall error. Moreover, the filter used in [35] is presented as an approach to graph regularization and does not consider the principle of consistency developed in this paper.

The superiority of our consistency filter \mathbf{C} over conventional graph Laplacian regularization is also empirically evident from the following experiment. In Fig. 4 we compare the performance of our consistency filter \mathbf{C} and the regularized Laplacian filter \mathbf{R} on the three image patches used in Sec. VII. For two different noise levels of $\sigma = 5$ and 30, we plot the peak SNR values for estimates obtained using the filters \mathbf{C} and \mathbf{R} as a function of the regularization penalty factor λ . We illustrate the performance of the two filters for both the oracle similarity matrix \mathbf{W} as well the estimate obtained from the noisy observed image.

TABLE II: Comparison of denoising performance of the standard regression $\mathbf{W} = \mathbf{D}^{-1}\mathbf{K}$ with its Sinkhorn version $\mathbf{S}(\mathbf{W})$ and our filter \mathbf{C} (PSNR in dB).

| Image | Noise Level | $\mathbf{W} = \mathbf{D}^{-1}\mathbf{K}$ | $\mathbf{S}(\mathbf{W})$ | Our \mathbf{C} filter | $\mathbf{S}(\mathbf{W})$ Diff. | Our Diff. |
|---|---------------|--|--------------------------|-------------------------|--------------------------------|-----------|
|  | $\sigma = 20$ | 28.69 | 28.48 | 30.44 | -0.21 | 1.75 |
| | $\sigma = 40$ | 23.24 | 22.74 | 26.82 | -0.50 | 3.58 |
|  | $\sigma = 20$ | 27.27 | 27.11 | 28.61 | -0.16 | 1.34 |
| | $\sigma = 40$ | 22.42 | 21.97 | 25.73 | -0.45 | 3.31 |
|  | $\sigma = 20$ | 24.78 | 24.89 | 26.18 | 0.11 | 1.40 |
| | $\sigma = 40$ | 22.16 | 21.96 | 23.08 | -0.20 | 0.92 |
|  | $\sigma = 20$ | 24.89 | 25.29 | 27.35 | 0.40 | 2.46 |
| | $\sigma = 40$ | 21.83 | 21.73 | 22.39 | -0.10 | 0.56 |
|  | $\sigma = 20$ | 28.38 | 27.91 | 33.12 | -0.47 | 4.74 |
| | $\sigma = 40$ | 22.68 | 22.13 | 29.55 | -0.55 | 6.87 |
|  | $\sigma = 20$ | 28.26 | 27.98 | 30.67 | -0.28 | 2.41 |
| | $\sigma = 40$ | 22.93 | 22.41 | 27.33 | -0.52 | 4.40 |
|  | $\sigma = 20$ | 28.20 | 27.93 | 30.53 | -0.27 | 2.33 |
| | $\sigma = 40$ | 22.95 | 22.43 | 27.23 | -0.52 | 4.28 |

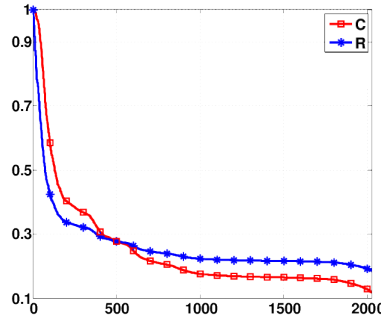









Fig. 6: Spectra of \mathbf{C} and \mathbf{R} for the cameraman edge image used in Fig. 3. The faster decay of the spectrum of \mathbf{C} indicates superior denoising performance.

The first important observation to be made from Fig. 4 is that the optimal performance of our consistency filter \mathbf{C} is in general superior to that of \mathbf{R} (for both the oracle and estimated filters \mathbf{W}). This is in conformity with the argument in Sec. VI that our filter carries out an optimal Wiener filtering operation in the eigen-space of the graph Laplacian for the image. This is in contrast with the non-optimal shrinkage carried out by \mathbf{R} . Not only is this observation true for the oracle filters, but it is also the case when the filters are estimated from the noisy image patches themselves. The gap in performance is large for both the constant and edge patches since here the eigen-decomposition results in a clear separation of piecewise-constant regions present in the image patches. While the difference in performance is reduced for the textured patch for high noise levels ($\sigma = 30$), we note that even here, the performance of our consistency filter \mathbf{C} is many dB superior to the optimal performance of the regularized Laplacian filter \mathbf{R} .

We reiterate here that the superior performance of our filter over the Laplacian regularization approach is due to the comparative role played by the form $\mathbf{L}^T\mathbf{L}$ and the \mathbf{L} in \mathbf{C} and \mathbf{R} respectively. Consider a graph adjacency

TABLE III: Comparison of denoising performance of the optimally diffused $\mathbf{W}^\#$ with its Sinkhorn version $\mathbf{S}(\mathbf{W})^\#$ and our filter \mathbf{C} (PSNR in dB).

| Image | Noise Level | $\mathbf{W}^\#$ | $\mathbf{S}(\mathbf{W})^\#$ | Our \mathbf{C} filter | $\mathbf{S}(\mathbf{W})^\#$ Diff. | Our Diff. |
|---|---------------|-----------------|-----------------------------|-------------------------|-----------------------------------|-----------|
|  | $\sigma = 20$ | 29.98 | 29.96 | 30.44 | -0.02 | 0.46 |
| | $\sigma = 40$ | 26.73 | 26.62 | 26.82 | -0.11 | 0.09 |
|  | $\sigma = 20$ | 27.94 | 28.02 | 28.61 | 0.08 | 0.67 |
| | $\sigma = 40$ | 25.36 | 25.29 | 25.73 | -0.07 | 0.37 |
|  | $\sigma = 20$ | 24.78 | 24.91 | 26.18 | 0.13 | 1.40 |
| | $\sigma = 40$ | 23.08 | 23.08 | 23.08 | 0.00 | 0.00 |
|  | $\sigma = 20$ | 24.89 | 25.29 | 27.35 | 0.40 | 2.46 |
| | $\sigma = 40$ | 21.96 | 22.03 | 22.39 | 0.07 | 0.43 |
|  | $\sigma = 20$ | 32.30 | 32.28 | 33.12 | -0.02 | 0.82 |
| | $\sigma = 40$ | 29.71 | 29.51 | 29.55 | -0.20 | -0.16 |
|  | $\sigma = 20$ | 29.72 | 29.72 | 30.67 | 0.00 | 0.95 |
| | $\sigma = 40$ | 27.09 | 27.00 | 27.33 | -0.09 | 0.24 |
|  | $\sigma = 20$ | 29.63 | 29.57 | 30.53 | -0.06 | 0.90 |
| | $\sigma = 40$ | 27.25 | 27.12 | 27.23 | -0.13 | -0.02 |

matrix where both pixels i and j are not neighbors of each other but are both neighbors of a third pixel k . In such a case, while \mathbf{L} will reflect this fact, the equivalent Laplacian matrix in our method $\mathbf{L}_{eqv} = \mathbf{L}^T \mathbf{L}$ will induce a relationship between pixels i and j by virtue of the fact that they are neighbors of pixel k . Thus, the effective set of good neighbors is enriched due to our symmetric Laplacian form resulting in improved denoising. Such an approach of symmetrization plays an important role in clustering methods based on directed graphs and is known as *co-citation* in the bibliometric literature [36], [37]. In Fig. 5 we show the matrices for \mathbf{L} and $\mathbf{L}_{eqv} = \mathbf{L}^T \mathbf{L}$ for the Lena image. While \mathbf{L} has a fixed number of entries in each row, we note that the effective number of good neighbors is significantly increased in our case of $\mathbf{L}_{eqv} = \mathbf{L}^T \mathbf{L}$, thereby leading to improved denoising performance. Finally, we present additional evidence of the impact of our symmetrization by comparing the spectra of the filters \mathbf{C} and \mathbf{R} . For the cameraman edge image used in earlier experiments, we explicitly compute the filters \mathbf{C} and \mathbf{R} and their corresponding spectra are shown in Fig. 6. As can be seen, our filter \mathbf{C} has a sharper fall-off resulting in more effective denoising performance as the higher modes that correspond to noise are better suppressed.

IX. EXPERIMENTAL RESULTS

In this Section we describe a series of experiments that test the performance of our consistency filter \mathbf{C} on a set of natural images and compare it with other existing methods. We note that while [7] studies the effect of symmetrization for denoising patches, we believe that it is useful to compare performance on full images which contain a variety of scales and texture patterns. In these tests we use the set of 7 natural images that were used in [38]². All images used in our experiments were resized to 256×256 pixels. Since we are considering the problem of denoising full images, we wish to use a kernel that combines both local and non-local similarity measurements. The kernel used in these experiments is a combination of the pixelwise bilateral kernel, \mathbf{K}_{bl} and the patchwise non-local kernel \mathbf{K}_{nl} , i.e. $\mathbf{K} = \mathbf{K}_{bl} + \mathbf{K}_{nl}$. By using such a combination, we can utilize the power of both patch based non-local methods as well as exploit the correlations between neighboring pixels that is captured by the bilateral

²Available at http://decsai.ugr.es/~javier/denoise/test_images/index.htm

method. For all our experiments we use a normalized bilateral filter \mathbf{K}_{bl} of size 9×9 pixels with $\sigma_{spatial} = 2\sigma$ and $\sigma_{photometric} = 1.2\sigma$ where σ is the noise level. For the non-local filter \mathbf{K}_{nl} we use 7 nearest neighbor patches of size 7×7 pixels. We note that a 7×7 patch size is both sufficiently large to be robust to noise and also small enough to retain fine details in the denoised image [5]. We validated this in terms of the PSNR's of the denoised images through extensive experimentation. In addition, we set $\lambda = 3 \times 10^{-3}$ and average results over 20 trials. Finally, for all experiments in this Section, for each filter, we run that filter once to obtain a pilot image and then re-estimate the filter.

It will be recalled that given a kernel measure of similarity \mathbf{K} , the standard linear average or regression is given by using the filter $\mathbf{W} = \mathbf{D}^{-1}\mathbf{K}$. While [7] proposed the use of \mathbf{S} as a way to modify \mathbf{W} , our proposed modification is the consistency filter \mathbf{C} . Hence, a meaningful measure of the value of the symmetrization approaches of \mathbf{S} and \mathbf{C} is to compare the improvement in performance they offer over the original filter \mathbf{W} . This comparison is presented in Table II where the peak-SNR of the different methods are presented for two different noise levels on different images. As can be seen here, although \mathbf{S} makes \mathbf{W} into a symmetric and stable filter, it does not lead to improved denoising performance. In fact, in most cases \mathbf{S} is poorer in quality than the original filter \mathbf{W} . This suggests that while the theoretical properties of symmetry and stability are desirable, we need other approaches to achieve these characteristics. We argue here that our consistency filter \mathbf{C} satisfies these requirements and, as is evident from Table II, we gain significantly in terms of performance (many dB over the performance of \mathbf{W}). We note that the significant improvement in performance of \mathbf{C} compared to \mathbf{W} is entirely due to the consistency criteria that we have introduced since \mathbf{C} is entirely derived from \mathbf{W} .

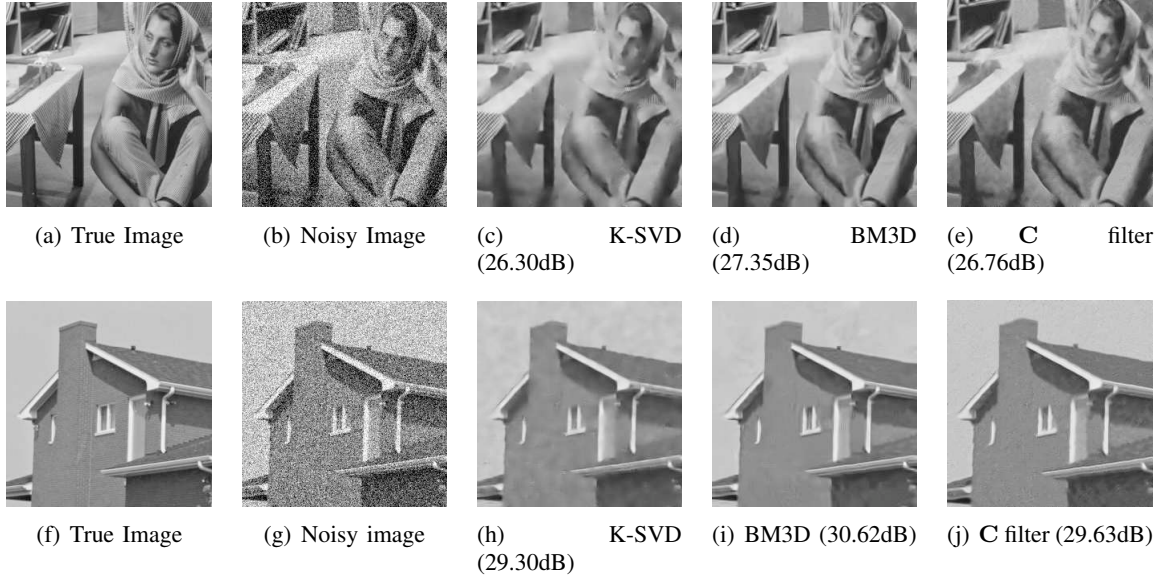


Fig. 7: Comparative visual results with output PSNRs on the Barbara and House images at a high noise level of $\sigma = 40$.

While \mathbf{W} shows relatively poor performance compared to \mathbf{C} in Table II, we also note that many papers propose the repeated application of \mathbf{W} for optimal denoising. The argument used for advocating such proposals is that there is an ‘optimal’ level of diffusion that a filter should carry out and a single application of \mathbf{W} is insufficient to achieve this desired diffusion level. To test this proposal, we carried out an additional experiment where we applied the filters \mathbf{W} and \mathbf{S} upto 6 times on the noisy image. Thereupon, since we know the ground truth image, we select that estimate with the optimal PSNR and report it. These measurements are reported in Table III. In other words, the performance reported in Table III is for $\mathbf{W}^\# = \mathbf{W}^k$ and $\mathbf{S}^\# = \mathbf{S}^k$, where $k \in \{1, \dots, 6\}$. However, the results reported for our consistency filter \mathbf{C} is for a single application as in Table II. In Table III we notice that the performance of $\mathbf{W}^\#$ and $\mathbf{S}^\#$ is significantly superior from that of a single application of \mathbf{W} and \mathbf{S} respectively. Indeed, the gap in performance between our filter \mathbf{C} and $\mathbf{W}^\#$ is significantly reduced compared to that of \mathbf{W} in

Table II. This suggests that optimal diffusion does play a significant role in improving performance. However, we strongly emphasize that the experiments in Table III are only to test the value of diffusion and do not have any significance as a real filter proposal. We note that we have selected the optimally diffused filters $\mathbf{W}^\#$ and $\mathbf{S}^\#$ by selecting the number of filter iterations (between 1 and 6) by using the ground truth image for comparison. This is unrealistic and we will need to select a fixed number of iterations for any real filter implementation. Such a filter with a fixed number of iterations will perform poorly as the amount of diffusion is fixed independent of the amount of texture and noise present in a given image. In any event, even when we favor $\mathbf{W}^\#$ and $\mathbf{S}^\#$ by using the ground truth to determine the appropriate stopping criterion, our filter \mathbf{C} that uses a single iteration exhibits superior denoising performance. We believe that this is strong evidence for the value of carrying out the form of regularization used in our filter \mathbf{C} as opposed to the straightforward application of either \mathbf{W} or its doubly stochastic counterpart \mathbf{S} .

We also compared the performance of our method with a variety of other methods to establish its value as an image denoising filter. In addition to the above mentioned regression filters $\mathbf{W} = \mathbf{D}^{-1}\mathbf{K}$ and \mathbf{S} , we also compared our performance with the regularized graph Laplacian filter \mathbf{R} described in Sec. VIII. For both filters \mathbf{R} and \mathbf{C} we selected the regularization parameters λ to obtain optimal performance in terms of PSNR for the filters \mathbf{R} and \mathbf{C} respectively. We note that this allows us to present a fair comparison of the performance of our filter \mathbf{C} against that of \mathbf{R} . While all of these methods, including our filter \mathbf{C} , are linear filters derived from the kernel form \mathbf{K} , we also compared our results with those obtained by two well-known nonlinear methods that are amongst the state-of-the-art techniques, i.e., the K-SVD algorithm [10] and BM3D [11]. For both these methods we used the online code presented by the respective authors. The comparison of our filter \mathbf{C} with all of these methods is given in terms of PSNR in Table IV. As can be seen, our filter consistently outperforms the other linear filters presented, i.e. \mathbf{W} , \mathbf{S} and \mathbf{R} . We also note the important point that our method provides an improvement over that of the K-SVD algorithm for most of the cases. In general, we note that the performance of our consistency filter is between that of K-SVD and BM3D in terms of the quality of the denoised images. We emphasize here that this quality of performance is achieved by our method which is a linear filter as opposed to the nonlinear techniques used in the K-SVD algorithm [10] and BM3D [11]. We also compared our performance in terms of SSIM which measures the structural similarity of images [39]. Since PSNR and SSIM are metrics that capture different properties of the denoised estimates, for this comparison, we marginally changed our parameter settings ($\lambda_{20} = 0.0028$ and $\lambda_{40} = 0.0035$) which improves the SSIM at a cost of marginal decrease in PSNR. From Table V, it can be seen that the SSIM indices are comparable to those of K-SVD and BM3D and outperforms K-SVD in some of the cases. Finally, in Fig. 7, we present some denoised images for various methods including our consistency filter for the purposes of visual comparison.








X. CONCLUSION

In this paper we have presented a consistency framework for image denoising. Applying the consistency criteria leads to a linear filter that has the desirable properties of being symmetric and is also a generalized doubly stochastic matrix. Our proposed consistency filter has an interpretation as performing optimal one-dimensional Wiener filtering along the basis vectors of the Laplacian of the weighting graph of the image. It significantly outperforms the linear regression filter as well as the standard regularized Laplacian filter. Its performance is also competitive with some well-known nonlinear image denoising methods.

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] J. L. Starck, E. J. Candès, and D. L. Donoho, "The curvelet transform for image denoising," *Image Processing, IEEE Transactions on*, vol. 11, no. 6, pp. 670–684, 2002.
- [4] C. Tomasi and R. Manduchi, "Bilateral filtering for gray and color images," in *Computer Vision, 1998. Sixth International Conference on*, 1998, pp. 839–846.
- [5] A. Buades, B. Coll, and J.-M. Morel, "A non-local algorithm for image denoising," in *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, vol. 2. IEEE, 2005, pp. 60–65.

TABLE IV: Denoising performance comparison (PSNR in dB) of our method with various state-of-the-art techniques.

| Image | Noise Level | $\mathbf{W} = \mathbf{D}^{-1}\mathbf{K}$ | $\mathbf{S}(\mathbf{W})$ [7] | \mathbf{R} filter = $(\mathbf{I} + \lambda\mathbf{L})^{-1}$ | K-SVD [10] | BM3D [11] | Our \mathbf{C} filter = $(\mathbf{I} + \lambda\mathbf{L}^T\mathbf{L})^{-1}$ |
|--|---------------|--|------------------------------|---|------------|-----------|---|
|  | $\sigma = 20$ | 28.69 | 28.48 | 28.84 | 30.33 | 30.92 | 30.44 |
| | $\sigma = 40$ | 23.24 | 22.74 | 26.00 | 26.37 | 27.43 | 26.82 |
|  | $\sigma = 20$ | 27.27 | 27.11 | 28.26 | 29.41 | 29.66 | 28.61 |
| | $\sigma = 40$ | 22.42 | 21.97 | 24.61 | 25.56 | 26.25 | 25.73 |
|  | $\sigma = 20$ | 24.78 | 24.89 | 25.54 | 26.53 | 26.79 | 26.18 |
| | $\sigma = 40$ | 22.16 | 21.96 | 21.96 | 22.77 | 23.23 | 23.08 |
|  | $\sigma = 20$ | 24.89 | 25.29 | 26.24 | 26.97 | 27.56 | 27.35 |
| | $\sigma = 40$ | 21.83 | 21.73 | 21.37 | 21.96 | 22.70 | 22.39 |
|  | $\sigma = 20$ | 28.38 | 27.91 | 31.39 | 33.03 | 33.64 | 33.12 |
| | $\sigma = 40$ | 22.68 | 22.13 | 27.88 | 29.20 | 30.66 | 29.55 |
|  | $\sigma = 20$ | 28.26 | 27.98 | 29.09 | 30.94 | 31.43 | 30.67 |
| | $\sigma = 40$ | 22.93 | 22.41 | 26.10 | 27.01 | 27.93 | 27.33 |
|  | $\sigma = 20$ | 28.20 | 27.93 | 29.84 | 30.96 | 31.44 | 30.53 |
| | $\sigma = 40$ | 22.95 | 22.43 | 26.17 | 27.34 | 28.18 | 27.23 |

- [6] 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.
- [7] P. Milanfar, "Symmetrizing smoothing filters," *SIAM Journal on Imaging Sciences*, vol. 6, no. 1, pp. 263–284, 2013.
- [8] A. Elmoataz, O. Lezoray, and S. Boughleux, "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.
- [9] X. Liu, D. Zhai, D. Zhao, G. Zhai, and W. Gao, "Progressive image denoising through hybrid graph laplacian regularization: A unified framework," *Image Processing, IEEE Transactions on*, vol. 23, no. 4, pp. 1491–1503, April 2014.
- [10] M. Elad and M. Aharon, "Image denoising via sparse and redundant representations over learned dictionaries," *Image Processing, IEEE Transactions on*, vol. 15, no. 12, pp. 3736–3745, 2006.
- [11] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, "Image denoising by sparse 3-d transform-domain collaborative filtering," *Image Processing, IEEE Transactions on*, vol. 16, no. 8, pp. 2080–2095, 2007.
- [12] P. Chatterjee and P. Milanfar, "Is denoising dead?" *Image Processing, IEEE Transactions on*, vol. 19, no. 4, pp. 895–911, 2010.
- [13] A. Levin and B. Nadler, "Natural image denoising: Optimality and inherent bounds," in *CVPR*, 2011, pp. 2833–2840.
- [14] A. Levin, B. Nadler, F. Durand, and W. T. Freeman, "Patch complexity, finite pixel correlations and optimal denoising," in *ECCV* (5), 2012, pp. 73–86.
- [15] M. Berman, "Automated smoothing of image and other regularly spaced data," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 16, no. 5, pp. 460–468, May 1994.
- [16] R. Kašpar and B. Zitová, "Weighted thin-plate spline image denoising," *Pattern recognition*, vol. 36, no. 12, pp. 3027–3030, 2003.
- [17] 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, 2007.
- [18] H. Talebi and P. Milanfar, "Global image denoising," *Image Processing, IEEE Transactions on*, vol. 23, no. 2, pp. 755–768, 2014.
- [19] A. Buja, T. Hastie, and R. Tibshirani, "Linear smoothers and additive models," *The Annals of Statistics*, 1989.
- [20] A. Cohen, "All admissible linear estimates of the mean vector," *The Annals of Mathematical Statistics*, vol. 37, no. 2, pp. 458–463, 1966.
- [21] R. Sinkhorn, "A relationship between arbitrary positive matrices and doubly stochastic matrices," *The Annals of Mathematical Statistics*, vol. 35, no. 2, pp. 876–879, 6 1964.
- [22] P. A. Knight, "The sinkhorn-knopp algorithm: Convergence and applications," *SIAM J. Matrix Anal. Appl.*, vol. 30, no. 1, pp. 261–275, 2008.

TABLE V: Denoising performance comparison (PSNR in dB and SSIM) of our method with some state-of-the-art techniques.

| Image | Noise Level | Measure | K-SVD [10] | BM3D [11] | Our C filter = $(\mathbf{I} + \lambda \mathbf{L}^T \mathbf{L})^{-1}$ |
|---|---------------|---------|------------|-----------|--|
|  | $\sigma = 20$ | PSNR | 30.34 | 30.93 | 30.14 |
| | | SSIM | 0.86 | 0.88 | 0.86 |
| | $\sigma = 40$ | PSNR | 26.39 | 27.46 | 26.77 |
| | | SSIM | 0.74 | 0.79 | 0.75 |
|  | $\sigma = 20$ | PSNR | 29.39 | 29.64 | 28.43 |
| | | SSIM | 0.83 | 0.85 | 0.81 |
| | $\sigma = 40$ | PSNR | 25.55 | 26.24 | 24.82 |
| | | SSIM | 0.70 | 0.74 | 0.69 |
|  | $\sigma = 20$ | PSNR | 26.54 | 26.79 | 26.21 |
| | | SSIM | 0.93 | 0.93 | 0.93 |
| | $\sigma = 40$ | PSNR | 22.78 | 23.23 | 22.97 |
| | | SSIM | 0.80 | 0.84 | 0.83 |
|  | $\sigma = 20$ | PSNR | 26.98 | 27.56 | 27.41 |
| | | SSIM | 0.91 | 0.92 | 0.90 |
| | $\sigma = 40$ | PSNR | 21.93 | 22.69 | 22.35 |
| | | SSIM | 0.77 | 0.81 | 0.74 |
|  | $\sigma = 20$ | PSNR | 33.04 | 33.64 | 32.98 |
| | | SSIM | 0.86 | 0.87 | 0.85 |
| | $\sigma = 40$ | PSNR | 29.20 | 30.69 | 29.76 |
| | | SSIM | 0.79 | 0.82 | 0.77 |
|  | $\sigma = 20$ | PSNR | 30.95 | 31.45 | 30.37 |
| | | SSIM | 0.88 | 0.89 | 0.87 |
| | $\sigma = 40$ | PSNR | 27.02 | 27.91 | 26.97 |
| | | SSIM | 0.78 | 0.81 | 0.76 |
|  | $\sigma = 20$ | PSNR | 30.95 | 31.44 | 30.49 |
| | | SSIM | 0.88 | 0.89 | 0.87 |
| | $\sigma = 40$ | PSNR | 27.15 | 27.92 | 26.78 |
| | | SSIM | 0.80 | 0.82 | 0.78 |

- [23] U. Von Luxburg, "A tutorial on spectral clustering," *Statistics and Computing*, vol. 17, no. 4, pp. 395–416, 2007.
- [24] D. I. Shuman, S. K. 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.
- [25] D. L. Donoho and I. M. Johnstone, "Ideal spatial adaptation by wavelet shrinkage," *Biometrika*, vol. 81, no. 3, pp. 425–455, 1994.
- [26] I. Daubechies, M. Defrise, and C. De Mol, "An iterative thresholding algorithm for linear inverse problems with a sparsity constraint," *Communications on Pure and Applied Mathematics*, vol. 57, no. 11, pp. 1413–1457, 2004.
- [27] A. Beck and M. Teboulle, "A fast iterative shrinkage-thresholding algorithm for linear inverse problems," *SIAM Journal on Imaging Sciences*, vol. 2, no. 1, pp. 183–202, 2009.
- [28] F. G. Meyer and X. Shen, "Perturbation of the eigenvectors of the graph laplacian: Application to image denoising," *CoRR*, vol. abs/1202.6666, 2012.
- [29] S. T. Roweis and L. K. Saul, "Nonlinear dimensionality reduction by locally linear embedding," *Science*, vol. 290, no. 5500, pp. 2323–2326, 2000.
- [30] D. Mumford and J. Shah, "Optimal approximations by piecewise smooth functions and associated variational problems," *Communications on Pure and Applied Mathematics*, vol. 42, no. 5, pp. 577–685, 1989.
- [31] J. Huang and D. Mumford, "Statistics of natural images and models," in *IEEE Conference on Computer Vision and Pattern Recognition*, 1999, pp. 1541–1547.
- [32] M. Zontak and M. Irani, "Internal statistics of a single natural image," in *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on*, 2011.
- [33] R. Merris, "Doubly stochastic graph matrices," *Publikacije-Elektrotehnickog Fakulteta Univerziteta U Beogradu Serija Matematika*, vol. 8, pp. 64–71, 1997.
- [34] A. J. Smola and R. Kondor, "Kernels and regularization on graphs," in *Learning theory and kernel machines*. Springer, 2003, pp. 144–158.
- [35] A. Gadde, S. K. Narang, and A. Ortega, "Bilateral filter: Graph spectral interpretation and extensions," in *International Conference on Image Processing*, 2013, pp. 1222–1226.
- [36] H. Small, "Co-citation in the scientific literature: A new measure of the relationship between two documents," *Journal of the American Society for Information Science*, vol. 24, no. 4, pp. 265–269, 1973.

- [37] V. Satuluri and S. Parthasarathy, "Symmetrizations for clustering directed graphs," in International Conference on Extending Database Technology, 2011, pp. 343–354.
- [38] J. Portilla, V. Strela, M. J. Wainwright, and E. P. Simoncelli, "Image denoising using scale mixtures of gaussians in the wavelet domain," Image Processing, IEEE Transactions on, vol. 12, no. 11, pp. 1338–1351, 2003.
- [39] 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.



Sk. Mohammadul Haque received the M.E. degree in Signal Processing from the Department of Electrical Engineering, Indian Institute of Science, Bengaluru, India in 2012 where he is currently pursuing the Ph. D. degree. His main areas of research interest are inverse problems in computer vision and image processing.



Gautam Pai received the M.E. degree with a major in Signal Processing from the Department Of Electrical Engineering, Indian Institute Of Science, Bengaluru, India. He is currently an Engineer with Philips Research Bangalore, India. His areas of research interest are geometric estimation problems in computer vision and image processing.



Venu Madhav Govindu received the Ph.D. degree in electrical engineering from the University of Maryland, College Park, MD, USA. He is currently an Assistant Professor with the Department of Electrical Engineering, Indian Institute of Science, Bengaluru, India. His current research interests include geometric and statistical inference problems in computer vision and image processing.