The Role of Redundant Bases and Shrinkage Functions in Image Denoising

Yacov Hel-Or and Gil Ben-Artzi

***Abstract*— Wavelet denoising is a classical and effective approach for reducing noise in images and signals. Suggested in 1994, this approach is carried out by rectifying the coefficients of a noisy image, in the transform domain, using a set of shrinkage functions (SFs). A plethora of papers deals with the optimal shape of the SFs and the transform used. For example, it is widely known that applying SFs in a redundant basis improves the results. However, it is barely known that the shape of the SFs should be changed when the transform used is redundant. In this paper, we introduce a complete picture of the interrelations between the transform used, the optimal shrinkage functions, and the domains in which they are optimized. We suggest three schemes for optimizing the SFs and provide bounds of the remaining noise, in each scheme, with respect to the other alternatives. In particular, we show that for subband optimization, where each SF is optimized independently for a particular band, optimizing the SFs in the spatial domain is always better than or equal to optimizing the SFs in the transform domain. Furthermore, for redundant bases, we provide the expected denoising gain that can be achieved, relative to the unitary basis, as a function of the redundancy rate.**

***Index Terms*— Wavelet transforms, image restoration, image denoising, shrinkage denoising, cycle spinning, noise removal, overcomplete representation.**

I. Introduction

CONSIDER a noisy image

$$\mathbf{y} = \mathbf{x} + \mathbf{n} \tag{1}$$

where $\mathbf{y}$ is the observed image, $\mathbf{x}$ the unknown original image and $\mathbf{n}$ the contaminating noise (all in vector notation). The goal is to reconstruct the original image $\mathbf{x}$ given the noisy measurement $\mathbf{y}$. This is a classical formulation of image denoising, which is a typical instance of an inverse problem. Using the maximum a posteriori (MAP) criterion, the solution aims at maximizing the a posteriori probability given the noisy image. The MAP solution must consider prior knowledge about the distribution of $\mathbf{x}$, and generally speaking, the prior distribution of natural images or any other specific class of images plays a key role in any denoising approach.

Manuscript received April 21, 2020; revised December 25, 2020 and February 7, 2021; accepted February 24, 2021. Date of publication March 17, 2021; date of current version March 23, 2021. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Zhu Li. *(Corresponding author: Yacov Hel-Or.)*

Yacov Hel-Or is with the Efi Arazi School of Computer Science, The Interdisciplinary Center, Herzliya 30900, Israel (e-mail: toky@idc.ac.il).

Gil Ben-Artzi is with the Department of Computer Science, Ariel University, Ariel 40700, Israel.

Digital Object Identifier 10.1109/TIP.2021.3065226

In the last few years, with the emergence of deep neural networks (DNN), a large body of works suggests performing image denoising by feedforward neural networks when the network aims at learning image-specific or general statistics of natural images [2]–[5]. Although DNN approaches are very effective and are the main focus these days, in this paper we remain loyal to the classical approaches where denoising is applied in the transform domain using shrinkage mappings. The reason for taking this position is that we are motivated by the theoretical bounds and the insights we gain by analyzing these type of approaches.

Transform based denoising is often implemented using some type of wavelet transform. The main motivation for this approach stems from the observation that the wavelet transform of natural images tends to reduce pixel dependencies [6]–[9]. Hence, it is possible to make a reasonable estimate about the joint distribution of the wavelet coefficients from their marginal distributions. When dealing with image denoising, this leads to a family of classical techniques known as the *wavelet shrinkage* methods, first introduced by Donoho and Johnstone in 1994 [1], [10], [11]. The shrinkage denoising approach is composed of a wavelet transform:

$$\mathbf{y}_u = U\mathbf{y} \tag{2}$$

where U is a matrix comprising the transform basis. The transform coefficients are then rectified by a correction step in which they are modified according to a set of scalar *shrinkage functions*, $\{\psi_i : \Re \to \Re\}$:

$$\tilde{\mathbf{y}}_u = \psi(\mathbf{y}_u) \tag{3}$$

where $\psi = (\psi_1, \psi_2, \cdots)$ is a vector of scalar mapping functions applied to each coefficient independently: $\tilde{\mathbf{y}}_u[i] = \psi_i(\mathbf{y}_u[i])$. The denoised image is then obtained by applying the pseudo-inverse transform to the modified coefficients:

$$\tilde{\mathbf{y}}_u^{\mathcal{S}} = U^{+}\tilde{\mathbf{y}}_u \tag{4}$$

where the superscript $\mathcal{S}$ indicates that we have transformed back to the spatial (image) domain. In cases where the transform is unitary or a tight frame, the pseudo-inverse yields the adjoint; thus, $\tilde{\mathbf{y}}_u^{\mathcal{S}} = U^T\tilde{\mathbf{y}}_u$. The resulting image $\tilde{\mathbf{y}}_u^{\mathcal{S}}$ serves as an estimate of the original image; hence, $\hat{\mathbf{x}}(\mathbf{y}) = \tilde{\mathbf{y}}_u^{\mathcal{S}}$. The denoising process is summarized in Figure 1.

The performance of shrinkage denoising is intimately dependent on two factors. The first factor is related to the choice of the shrinkage functions (SF) $\{\psi_i\}$ applied to the

1941-0042 © 2021 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission.
See https://www.ieee.org/publications/rights/index.html for more information.

Authorized licensed use limited to: Herzeliya IDC. Downloaded on May 08,2021 at 19:53:14 UTC from IEEE Xplore. Restrictions apply.

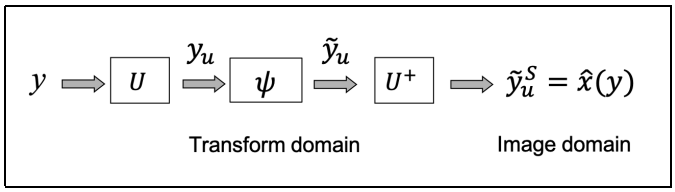

Fig. 1. The procedure for shrinkage denoising. A noisy image $\mathbf{y}$ is transformed by the wavelet matrix U into the transform coefficients $\mathbf{y}_u$. A set of shrinkage functions ψ rectify the coefficients providing $\tilde{\mathbf{y}}_u$. The inverse transform U^+ transforms the rectified coefficients back into the image domain $\tilde{\mathbf{y}}_u^S$, which is the denoised result.

transform coefficients. The justification for applying a marginal (scalar) SF to each coefficient independently emerges from the independence assumption of the wavelet coefficients when the transform is unitary. Assuming the statistical distribution of a wavelet band is stationary, and using the independence assumption of the wavelet coefficients, the SFs for all coefficients in a particular wavelet band can be shown to be identical [12]–[14]. Therefore, if the wavelet transform U is composed of K bands, only K SFs need to be estimated, when $\tilde{\mathbf{y}}_u[i] = \psi_{band(i)}(\mathbf{y}_u[i])$, where $band(i)$ indicates the band index of pixel i. Applying the SFs independently for each wavelet band, and detecting the optimal SF profile for each band was investigated previously (see e.g., [15], [16]). In principle, having a marginal prior distribution for a wavelet band, the associated SF can be derived using Bayesian estimation (e.g., [12], [14]). Alternatively, the SFs can be learnt directly from the noisy input [1], [10], [11] or from a set of example images that are given offline along with their clean counterparts [17]–[19].

The second factor that influences the denoising performance is the transform used during the process. Although the shrinkage approach using unitary wavelet transforms provides good results, significant improvement is achieved when implementing this technique with redundant transforms. Such transforms include preselected bases such as the undecimated wavelets [20], steerable wavelets [12], and other suggested transforms [21]–[26], or generated transforms that are adaptively learnt from the noisy image [27]–[30]. Note that scalar SFs can no longer be justified in redundant bases, as the transform coefficients are mutually dependent due to the transform redundancy. Nevertheless, the superior results of applying scalar SFs in the over-complete case suggest that such a scheme is still very effective in addition to its appealing efficiency.

The above-mentioned two factors influencing the denoising performance, namely, the transform used and the applied SFs, are mutually dependant and cannot be treated independently. The type of transform used directly influences the shape of the optimal SFs. Moreover, optimal SFs for a redundant transform, such as an undecimated wavelet, are shown to differ from the SFs optimally designed for the unitary basis [13], [27], [31]. To clarify, consider finding the optimal SFs for the unitary case with respect to the MMSE criterion. In other words, finding a ψ that minimizes

$$\Delta = E\{\|\hat{\mathbf{x}}(\mathbf{y}) - \mathbf{x}\|^2\}$$

when we have determined that $\hat{\mathbf{x}}(\mathbf{y}) = U^T \psi\,(U\mathbf{y})$, the norm $\|\cdot\|$ stands for the ℓ_2 norm and $E\{\cdot\}$ indicates the expectation taken over $\mathbf{x}$ and $\mathbf{y}$. Whenever U is unitary, this minimization can be formulated equivalently in the transform domain (since $UU^T = U^T U = I$) as:

$$E\{\|U^T \psi\,(U\mathbf{y}) - \mathbf{x}\|^2\} = E\{\|\psi\,(U\mathbf{y}) - U\mathbf{x}\|^2\} \tag{5}$$

i.e., ψ is optimized so that the noisy transform coefficients $U\mathbf{y}$ should be as close as possible to the transform coefficients of the clean image $U\mathbf{x}$. For an over-complete transform, however, this equality is no longer valid (since $UU^T \neq I$). This implies that the optimization for ψ should be expressed in the spatial domain, which is the relevant domain in our case. Because the inverse transform couples wavelet coefficients (inside subband and between subbands), spatial domain optimization requires a joint minimization of all SFs simultaneously, and this optimization is far more complicated to apply. In fact, this might be the reason that SFs applied in redundant bases are commonly borrowed from the unitary case or optimized in the transform domain with no real justification.

In [13], Raphan and Simoncelli showed that as long as the statistics of the image and the noise are stationary, the expected MSE of the denoised image resulting from applying the SFs, ψ in the unitary bases, is always greater or equal to the MSE of the denoised image resulting from applying the same ψ in the redundant basis (by spatially replicating the unitary basis using, e.g., cycle spinning or undecimated subbands [20], [32]). Note that this property was proven irrespective of the type of applied SFs, ψ. They also showed that when working with a redundant basis, there is an advantage in optimizing the SFs (with respect to the expected error) in the spatial domain rather than in the transform domain. This requires, however, optimizing jointly all SFs simultaneously, making the optimization process a demanding task.

In this paper, we extend the results of [13] and establish a complete picture of the interrelations between the transform used, the optimal shrinkage functions, and the domains in which they are optimized. In particular, we show that for subband optimization, where each ψ_i is optimized independently, optimizing each SF in the spatial domain is always better than or equal to optimizing the SFs in the transform domain. This option, besides being simple to implement, is proven to outperform the traditional transform domain optimization while avoiding the demanding spatial domain optimization of all SFs simultaneously.

Additionally, for redundant bases, we provide the expected denoising gain we may achieve, relative to the unitary basis, as a function of the basis redundancy. This result allows a user to make a clever decision about the redundancy used by taking into account the expected denoising gain and the computational time allocated for this process.

II. Redundant vs. Unitary Transforms

A common axiom in image denoising is that denoising applied in redundant bases (cycle-spinning or undecimated wavelets) outperforms the results obtained in unitary transforms [13], [20]. In this section we examine the relationships

Authorized licensed use limited to: Herzeliya IDC. Downloaded on May 08,2021 at 19:53:14 UTC from IEEE Xplore. Restrictions apply.

between the unitary and redundant transforms. In particular, we offer theoretical justification for applying shrinkage denoising in redundant bases. Some of the relationships in this section were already proven in [13] but we repeat them here for clarity and to provide a complete picture of the interrelations between the transformed used and the domain where the MSE is optimized. To be able to compare different transforms on a common basis, we limit our discussion to the unitary basis and their corresponding cycle-spinning transforms. That is to say, the difference between a unitary and a redundant transform is that the former is properly decimated and thus forms a complete basis, while the latter is formed by cycle-spinning the unitary basis.

A common technique for shrinkage denoising in a redundant basis is the cycle-spinning framework [20]. Cycle-spinning is performed by applying a unitary transform on a set of shifted versions of the image, denoising each version independently, then averaging the results after properly shifting them back. Since the transform of a spatially shifted image can be applied equivalently by shifting the transform basis (by the same amount but in the opposite direction), the transform can be seen as a redundant transform, composed of a set of shifted versions of the original unitary transform:

$$\mathbf{y}_{u_i} = US_i\mathbf{y} = U_i\mathbf{y} \quad i = 1 \ldots N \tag{6}$$

where S_i is a (cyclic) shift operator by the i-th displacement, and $U_i = US_i$ is a unitary transform composed of the wavelet basis after applying the respective shift. The entire transform is constructed by concatenating together all shifted transforms:

$$\mathbf{y}_w = W\mathbf{y} \tag{7}$$

where the redundant transform is defined as follows:

$$W = \frac{1}{\sqrt{N}} \begin{bmatrix} U_1 \\ U_2 \\ \vdots \\ U_N \end{bmatrix} \tag{8}$$

Note the W is over-complete and tight frame, satisfying $W^TW = I$; however, $WW^T \neq I$. Since WW^T is a projection matrix,[1] it can be shown that this restricts the eigen-values of WW^T to be 1 or 0. If W is an $m \times n$ matrix ($m > n$), then there are n eigen-values of 1, and $m - n$ eigen-values of 0 [33]. Consequently, for any $n \times 1$ vector $\mathbf{x}$ and $m \times 1$ vector $\mathbf{z}$, we have:

$$\|W\mathbf{x}\| = \|\mathbf{x}\| \tag{9}$$

and

$$\|W^T\mathbf{z}\| \leq \|\mathbf{z}\| \tag{10}$$

Similarly, since U_i is unitary, we have $U_i^TU_i = U_iU_i^T = I$ and accordingly:

$$\|U_i\mathbf{x}\| = \|U_i^T\mathbf{x}\| = \|\mathbf{x}\|, \quad \forall\ i \tag{11}$$

[1] A square matrix A is a projection matrix iff $AA = A$.

Denote by $\mathbf{n} = \mathbf{x} - \mathbf{y}$ the contaminated noise in the image domain before denoising. By applying a unitary transform, the error is transformed as well:

$$U(\mathbf{x} - \mathbf{y}) = U\mathbf{n} \doteq \mathbf{n}_u$$

and similarly, by the redundant transform:

$$W(\mathbf{x} - \mathbf{y}) = W\mathbf{n} \doteq \mathbf{n}_w$$

Since the transforms U and W are tight frames, we have (following Equations 9 and 11):

$$\|\mathbf{n}\| = \|\mathbf{n}_u\| = \|\mathbf{n}_w\| \tag{12}$$

which means that the norm of the error in the transform domain equals its norm in the image domain, and this is true for the unitary as well as the redundant case.

Now, after applying the shrinkage functions ψ to the transform coefficients, the distortion value may change. We define:

$$\tilde{\mathbf{n}}_u = U\mathbf{x} - \psi(U\mathbf{y}) \quad \text{and similarly} \quad \tilde{\mathbf{n}}_w = W\mathbf{x} - \psi(W\mathbf{y})$$

For the unitary case, the distortion is propagated to the image domain via the inverse transform:

$$U^T(Ux - \psi(U\mathbf{y})) = U^T\tilde{\mathbf{n}}_u \doteq \tilde{\mathbf{n}}_u^{\mathcal{S}}$$

and following Equation 11, we have:

$$\|\tilde{\mathbf{n}}_u\| = \|\tilde{\mathbf{n}}_u^{\mathcal{S}}\| \tag{13}$$

i.e., after applying the SFs, the MSE distortion in the transform domain is identical to its distortion in the image domain. As we will see next, in redundant transforms this property is not satisfied. In redundant transforms, the error in the image domain is:

$$W^T(Wx - \psi(W\mathbf{y})) = W^T\tilde{\mathbf{n}}_w \doteq \tilde{\mathbf{n}}_w^{\mathcal{S}}$$

Nevertheless, following Equation 10, we have:

$$\|\tilde{\mathbf{n}}_w\| \geq \|\tilde{\mathbf{n}}_w^{\mathcal{S}}\| \tag{14}$$

Note that the above relations (Equations 14 and 13) are valid for *any* shrinkage functions ψ and for any $\mathbf{x}$ and $\mathbf{y}$.

Letting $\mathbf{s}$ be a vector value depending on $\mathbf{x}$ and $\mathbf{y}$, we define the expected RMSE of $\mathbf{s}$:

$$\|\mathbf{s}\|_E \doteq \sqrt{E\,\|\mathbf{s}\|^2} \text{ where } E\,\|\mathbf{s}\|^2 = \int \|\mathbf{s}\|^2\ \ P(\mathbf{x}, \mathbf{y})d\mathbf{x}d\mathbf{y}$$

Since relations 14 and 13 are true for any $\mathbf{x}$ and $\mathbf{y}$, we can rephrase these relations using a statistical point of view:

$$\|\tilde{\mathbf{n}}_u\|_E = \|\tilde{\mathbf{n}}_u^{\mathcal{S}}\|_E \tag{15}$$

and

$$\|\tilde{\mathbf{n}}_w\|_E \geq \|\tilde{\mathbf{n}}_w^{\mathcal{S}}\|_E \tag{16}$$

These relations are illustrated along the two rows of Figure 2. We now establish the relationships between the unitary and redundant transforms that are indicated in the two columns of Figure 2.

We first show that in the transform domain, for the two transforms, the expected MSE distortion is equal. This outcome stems from the stationary property of natural images,

Authorized licensed use limited to: Herzeliya IDC. Downloaded on May 08,2021 at 19:53:14 UTC from IEEE Xplore. Restrictions apply.

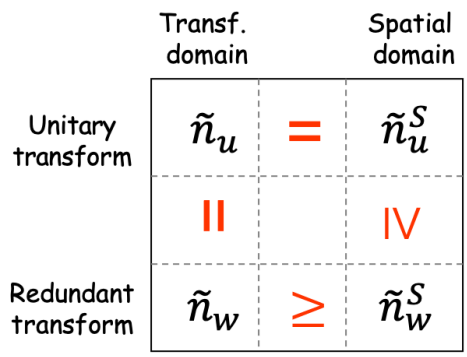

Fig. 2. The expected remaining noise for unitary vs. redundant bases, and spatial vs. transform domains. If the transform is unitary, the expected remaining noise is similar in the transform domain and in the spatial domain: $\tilde{n}_u^T = \tilde{n}_u^S$. However, for redundant transform, the remaining noise in the transform domain is expected to be no less than the remaining noise in the spatial domain: $\tilde{n}_w^T \geq \tilde{n}_w^S$.

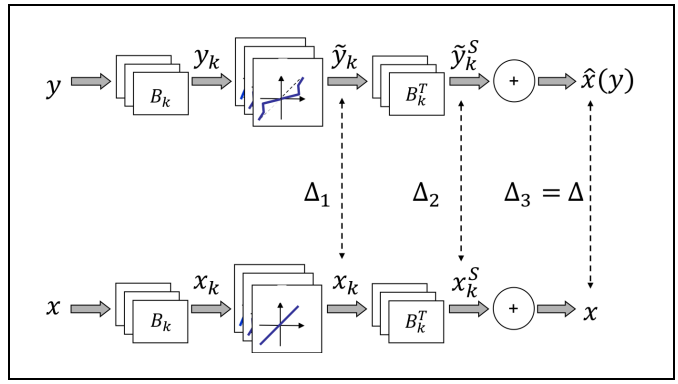

Fig. 3. The denoising process and three optimization schemes for the shrinkage functions. Δ_1, Δ_2 and Δ_3 are the three quantities used to optimize ψ^1, ψ^2 and ψ^3, respectively.

where it is assumed that the statistical properties of natural images are shift invariant.

Theorem 1: After denoising, the expected MSE distortions in the transform domain are equal for the unitary and for the redundant transforms. In other words, for any given ψ:

$$\|\tilde{\mathbf{n}}_u\|_E = \|\tilde{\mathbf{n}}_w\|_E$$

Proof 1: In Appendix A.

The last theorem leads to a theoretical justification for applying shrinkage denoising in a redundant basis. This is explicitly expressed in the next theorem:

Theorem 2: For any given ψ,

$$\|\tilde{\mathbf{n}}_u^{\mathcal{S}}\|_E \geq \|\tilde{\mathbf{n}}_w^{\mathcal{S}}\|_E$$

Proof 2: In Appendix B.

Theorem 2 completes the entire picture of Figure 2: In the transform domain, the expected remaining noise, after shrinkage, is identical for the unitary and the redundant (cycle-spinning) wavelet transforms for *any* shrinkage functions. When transforming back into the spatial domain, however, the remaining noise is expected to decrease in the redundant transform while staying the same in the unitary transform. This main conclusion suggests that it is preferable to apply shrinkage denoising in a redundant basis rather than in the unitary basis.

III. Optimizing the Shrinkage Functions

In this section we deal with the objectives to which the shrinkage functions (SFs) are optimized. As mentioned earlier, SFs play a significant role in the resulting performance, and their optimization is a longstanding topic of study (see, e.g., [27], [34]–[36], just to name a few). In principle, SFs can be derived from the joint statistics of the transform coefficients [13], [34], [35], but, unfortunately, modeling the precise joint statistics is a complicated and still intractable problem. Alternatively, one can optimize the SF of each subband independently using marginal statistics, but as mentioned earlier, this is not optimal in the redundant case. Another option is to learn the optimal SFs from an ensemble of images using a set of noisy and clean examples [13], [37], where the SFs are designed to clean the noisy examples in an optimal manner towards their clean counterparts. As shown next, there are several domains in which the SFs can be optimized in and the resulting quality depends on the selected domain.

In undecimated wavelet transforms, the number of coefficients is K times the size of the image, where K is the number of the wavelet bands. To facilitate the notation for band operations, we reorder the rows of a transform W so that transform rows corresponding to a wavelet band are co-located in a block. Naturally, we extend the same reordering to $\mathbf{y}_w$. Assuming we have K different wavelet bands and a corresponding permutation matrix P:

$$B = PW = \begin{bmatrix} B_1 \\ B_2 \\ \vdots \\ B_K \end{bmatrix} \quad \text{and accordingly} \quad \mathbf{y}_B = B\mathbf{y} = \begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \\ \vdots \\ \mathbf{y}_K \end{bmatrix}$$

where $\mathbf{y}_k = B_k\mathbf{y}$ represents the coefficients of the kth band. The new reordering does not change the tight frame property; thus, if $W^T W = I$, we have $B^T B = I$ as well. In the new reordering, a vector of SFs, $\psi = [\psi_1, \psi_2, \cdots, \psi_K]$, can be represented efficiently as follows in Equation 17. Since ψ_k is applied similarly to all coefficients in the kth band, we can rewrite Equation 3 as

$$\tilde{\mathbf{y}}_k = \psi_k(\mathbf{y}_k) \tag{17}$$

which means that the scalar mapping $\psi_k : \Re \to \Re$ is applied individually to each entry in $\mathbf{y}_k$. The clean image is then estimated using the adjoint:

$$\hat{\mathbf{x}}(\mathbf{y}) = B^T \psi(\mathbf{y}_B) = \sum_{k=1}^{K} B_k^T \tilde{\mathbf{y}}_k \doteq \sum_{k=1}^{K} \tilde{\mathbf{y}}_k^{\mathcal{S}} \tag{18}$$

where we define $\tilde{\mathbf{y}}_k^{\mathcal{S}} = B_k^T \tilde{\mathbf{y}}_k$. This process is illustrated in the upper pipeline of Figure 3. Let the SFs be a set of mapping functions taken from a given function space Ψ. The optimal set of SFs with respect to the MSE criterion is then obtained by

Authorized licensed use limited to: Herzeliya IDC. Downloaded on May 08,2021 at 19:53:14 UTC from IEEE Xplore. Restrictions apply.

finding the function set $\psi \subset \Psi$ that minimizes the following objective:

$$\hat{\psi} = \arg\min_{\psi \in \Psi} \Delta(\psi)$$

where

$$\Delta(\psi) = \sqrt{E\left\{\|\mathbf{x} - \hat{\mathbf{x}}(\mathbf{y})\|^2\right\}} \tag{19}$$

where $\hat{\mathbf{x}}(\mathbf{y})$ is as estimate defined in Equation 18 and $E\{\cdot\}$ stands for the expectation taken over $(\mathbf{x}, \mathbf{y})$.

The above minimization is complicated to accomplish as it requires modeling the entire joint statistics of natural images. Below we consider other alternatives for the objective functions. We examine three objectives expressing the optimal set of SFs. We refer to the definitions illustrated in Figure 3.

- *Method 1 (Transform Domain – Independent Bands):* A set of SFs are optimized in the transform domain. The optimization is applied by minimizing the objective function:

$$\Delta_1 = \sum_k \sqrt{E\left\{\|\mathbf{x}_k - \tilde{\mathbf{y}}_k\|^2\right\}}$$

  where $\mathbf{x}_k = B_k\mathbf{x}$ is the clean counterpart of $\tilde{\mathbf{y}}_k$. Since this objective is composed of a sum of independent terms, each of which contains a particular band, the minimization of this objective can be applied at each wavelet band independently, using only its marginal statistics (since the SFs are scalars); namely:

$$\hat{\psi}_k = \arg\min_{\psi \in \Psi} E\left\{\|\mathbf{x}_k - \tilde{\mathbf{x}}_k\|^2\right\} \quad \forall\ k \in \{1, 2, .., K\}$$

  where the expectation is over $\mathbf{y}_k, \mathbf{x}_k$.
- *Method 2 (Spatial Domain – Independent Bands):* A set of SFs is optimized in the spatial domain. The objective term for this method reads:

$$\Delta_2 = \sum_k \sqrt{E\left\{\|\mathbf{x}_k^{\mathcal{S}} - \tilde{\mathbf{y}}_k^{\mathcal{S}}\|^2\right\}}$$

  where $\mathbf{x}_k^{\mathcal{S}} = B_k^T\mathbf{x}_k$ and $\tilde{\mathbf{y}}_k^{\mathcal{S}} = B_k^T\tilde{\mathbf{y}}_k$. Note that though the objective criterion is expressed in the spatial domain, the SFs can be optimized for each band independently. Thus, although intra-band dependencies are conveyed through the adjoint transform and must be considered, the inter-band dependencies are ignored.
- *Method 3 (Spatial Domain – Joint Bands):* The objective goal is expressed in the spatial domain:

$$\Delta_3 = \sqrt{E\left\{\left\|\sum_k (\mathbf{x}_k^{\mathcal{S}} - \tilde{\mathbf{y}}_k^{\mathcal{S}})\right\|^2\right\}}$$

  It is easy to verify that this objective gives the actual expected error as defined in Equation 19; thus, $\Delta_3 = \Delta$. In this scheme, the SFs are evaluated simultaneously while inter-band as well as intra-band dependencies must be taken into account.

Denote the deviation of the approximated coefficients from the clean coefficients by $\mathbf{d}_k = \mathbf{x}_k - \tilde{\mathbf{y}}_k$ and similarly $\mathbf{d}_k^S = B_k^T(\mathbf{x}_k - \tilde{\mathbf{y}}_k) = B_k^T\mathbf{d}_k$. Using this notation, the above objectives read:

$$\Delta_1 = \sum_k \sqrt{E\left\{\|\mathbf{d}_k\|^2\right\}}, \quad \Delta_2 = \sum_k \sqrt{E\left\{\left\|\mathbf{d}_k^{\mathcal{S}}\right\|^2\right\}}, \tag{20}$$

and

$$\Delta_3 = \sqrt{E\left\{\left\|\sum_k \mathbf{d}_k^{\mathcal{S}}\right\|^2\right\}} = \Delta$$

For each method defined above, denote an associated optimal SF, $\hat{\psi}^i$, as follows:

$$\hat{\psi}^i = \arg\min_{\psi \in \Psi} \Delta_i, \quad \text{for } i = 1..3 \tag{21}$$

Additionally, the objective $\Delta(\hat{\psi}^i)$ denotes the actual expected error as defined in Equation 19 when applying the SF $\hat{\psi}^i$. In the following, we show that if the wavelet transform is unitary, then all three methods produce the same result. This is illustrated in the upper line of Figure 5.

Theorem 3: For the unitary case we have:

$$\Delta(\hat{\psi}^1) = \Delta(\hat{\psi}^2) = \Delta(\hat{\psi}^3)$$

Proof 3: To show the above relations, we prove that actually, for the unitary case, $\hat{\psi}^1 = \hat{\psi}^2 = \hat{\psi}^3$, which derives the theorem. Recall that $\hat{\psi} = [\hat{\psi}_1 \cdots \hat{\psi}_K]$ is composed of K SFs, each of which applies to a particular band. Thus, $\mathbf{d}_k$ and $\mathbf{d}_k^S$ depend only on ψ_k, and we can apply the optimization to each band independently. For the kth band, we have:

$$\hat{\psi}_k^1 = \arg\min_{\psi_k} E\left\{\|\mathbf{d}_k\|^2\right\}$$

and

$$\psi_k^2 = \arg\min_{\psi_k} E\left\{\left\|B_k^T\mathbf{d}_k\right\|^2\right\}$$

Since W is unitary, $WW^T = I$, and accordingly, $B_iB_j^T = \delta_{i,j}I$. Using this relation we get:

$$\left\|B_k^T\mathbf{d}_k\right\|^2 = \|\mathbf{d}_k\|^2$$

which gives

$$\hat{\psi}_k^1 = \hat{\psi}_k^2, \quad \text{for } k = 1..K$$

and accordingly $\hat{\psi}^1 = \hat{\psi}^2$, which implies the first relation in the theorem.

Similarly,

$$\hat{\psi}_k^3 = \arg\min_{\psi_k} E\left\{\left\|\sum_j B_j^T\mathbf{d}_j\right\|^2\right\}$$

However,

$$\left\|\sum_j B_j^T\mathbf{d}_j\right\|^2 = \left(\sum_i \mathbf{d}_i^T B_i\right)\left(\sum_j B_j^T\mathbf{d}_j\right) = \sum_j \|\mathbf{d}_j\|^2$$

Authorized licensed use limited to: Herzeliya IDC. Downloaded on May 08,2021 at 19:53:14 UTC from IEEE Xplore. Restrictions apply.

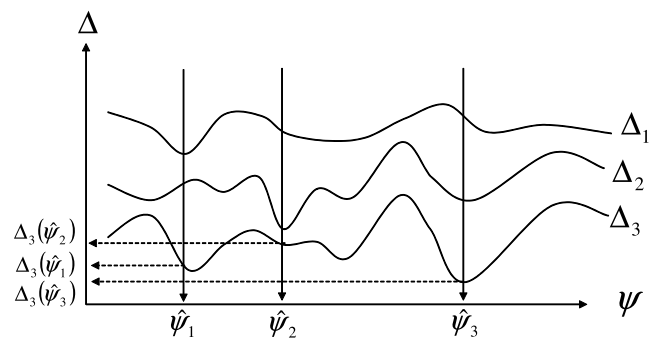

Fig. 4. Illustrated profiles of penalties following Theorem 4. The x-axis indicates various ψ values. $\Delta_1(\psi) \geq \Delta_2(\psi) \geq \Delta_3(\psi)$ and $\hat{\psi}_i = \min_\psi \Delta_i(\psi)$. Note that in this example, $\Delta(\hat{\psi}^1) < \Delta(\hat{\psi}^2)$.

Thus,

$$\hat{\psi}_k^3 = \arg\min_{\psi_k} E\left\{\sum_j \|\mathbf{d}_j\|^2\right\} = \arg\min_{\psi_k} E\left\{\|\mathbf{d}_k\|^2\right\} = \hat{\psi}_k^1$$

where $k \in \{1..K\}$. This yields $\hat{\psi}^3 = \hat{\psi}^1$, which implies the second relation of the theorem. Hence, in the unitary case, optimizing the SFs using any one of the above methods is equivalent. □

Theorem 3 establishes the justification for optimizing the SFs in the transform domain, in cases where the transform used is unitary. Using Method 1, each individual SF can be optimized independently, collecting only marginal statistics. This property makes this scheme very appealing and thus very popular (e.g., [10], [12], [20]).

The next theorem shows that in the over-complete transform, the situation is totally different, and the domain in which we apply the optimization makes a difference (see Figure 4 for an illustration).

Theorem 4: Let the transform W be over-complete and tight frame. In such a case, for each ψ,

$$\Delta_1(\psi) \geq \Delta_2(\psi) \geq \Delta_3(\psi)$$

Proof 4: Since W is tight frame, it follows that $W^T W = I$. It can easily be shown that this restricts the norm of each B_k: $\|B_k\| = \lambda_{(k)} \leq 1$, where $\lambda_{(k)}$ denotes the maximal eigen-value of $B_k B_k^T$ [33]. This yields that for any vector $\mathbf{z}$, $\|B_k^T \mathbf{z}\| \leq \|\mathbf{z}\|$; hence,

$$\Delta_1 = \sum_j \sqrt{E\left\{\|\mathbf{d}_j\|^2\right\}} \geq \sum_j \sqrt{E\left\{\|B_j^T \mathbf{d}_j\|^2\right\}}$$
$$= \sum_j \sqrt{E\left\{\|\mathbf{d}_j^S\|^2\right\}} = \Delta_2$$

and this proves the first inequality. Due to the triangular inequality of a norm,[2] we also have:

$$\Delta_2 = \sum_j \sqrt{E\left\{\|\mathbf{d}_j^S\|^2\right\}} \geq \sqrt{E\left\{\left\|\sum_j \mathbf{d}_j^S\right\|^2\right\}} = \Delta_3$$

which gives the second inequality in the theorem. □.

[2]Note that the expectation value can be inserted into the norm definition.

Theorem 4 reveals that in the over-complete case, optimizing the SFs in the transform domain is not optimal. In the following we provide justification for optimizing the SFs using Methods 1 and 2 as they provide upper bounds for the desired penalty (Δ of Method 3), that might be difficult to achieve.

Theorem 5: Let $\hat{\psi}^i = \arg\min_{\psi\in\Psi} \Delta_i(\psi)$ as defined in Equation 21. In the over-complete case,

$$\Delta_1(\hat{\psi}^1) \geq \Delta_2(\hat{\psi}^2) \geq \Delta_3(\hat{\psi}^3) \tag{22}$$

Proof 5: The SF $\hat{\psi}^3$ minimizes Δ_3; thus, $\Delta_3(\hat{\psi}^2) \geq \Delta_3(\hat{\psi}^3)$. Following Theorem 4, however, we have that $\Delta_2(\hat{\psi}^2) \geq \Delta_3(\hat{\psi}^2)$, from which it readily follows that $\Delta_2(\hat{\psi}^2) \geq \Delta_3(\hat{\psi}^3)$. The second inequality can be shown using a similar argument. Q.E.D.

Note also that according to the proof above, the actual errors (i.e., $\Delta_3 = \Delta$) using $\hat{\psi}^2$ and $\hat{\psi}^1$ are even tighter, i.e.:

$$\Delta_2(\hat{\psi}^2) \geq \Delta(\hat{\psi}^2) \geq \Delta(\hat{\psi}^3) \quad \text{for Method 2} \tag{23}$$
$$\Delta_1(\hat{\psi}^1) \geq \Delta(\hat{\psi}^1) \geq \Delta(\hat{\psi}^3) \quad \text{for Method 1} \tag{24}$$

and since $\Delta_1(\hat{\psi}^1) \geq \Delta_2(\hat{\psi}^2)$ (Theorem 5), the SF $\hat{\psi}^2$ has a better bound than $\hat{\psi}^1$. Thus, it is *expected* that $\Delta(\hat{\psi}^1) \geq \Delta(\hat{\psi}^2)$. Nevertheless, it cannot be assured that the actual error for $\hat{\psi}_2$ outperforms the actual error of $\hat{\psi}_1$, i.e., the relation:

$$\Delta(\hat{\psi}^1) \geq \Delta(\hat{\psi}^2) \tag{25}$$

is not necessarily true. To prove this, see a counter-example in Figure 4.

To conclude, in redundant transforms $\Delta_3 = \Delta$ determines the actual error and it is the optimal penalty to minimize. Nevertheless, since it requires inter- and intra-bands statistics, it is sometimes complicated to optimize. Δ_1 is the easiest term to minimize as it requires collecting only marginal statistics. Indeed, this approach is commonly used in the traditional techniques (hard/soft thresholding originated from this penalty). Δ_2 is a better penalty to minimize than Δ_1 as its bound is tighter, although it might be harder to optimize as it requires modeling intra-band statistics. Nevertheless, it is not guaranteed that $\Delta(\hat{\psi}^1) \geq \Delta(\hat{\psi}^2)$. Thus, there is an inherent trade-off between the three methods, while spatial domain optimization (Method 3) is preferable with respect to denoising quality, transform domain optimization (Method 1) is the most efficient to apply. Weak spatial domain (Method 2) is a good compromise between quality and efficiency.

In unitary transforms, all optimization objectives (Methods 1, 2, and 3) will generate similar results. Using Theorem 2 above, however, it was proven that it is expected that denoising in redundant transforms will generate better results than using unitary transforms. These relations are summarized in Figure 5.

IV. Improvement Rates for Redundant Transforms

In this section we analyze the expected improvement of the remaining MSE with respect to the redundant rate of the transform used. We assume an over-complete wavelet transform made by cycle spinning as given in Equation 7:

$$\mathbf{y}_w = W\mathbf{y} \tag{26}$$

Authorized licensed use limited to: Herzeliya IDC. Downloaded on May 08,2021 at 19:53:14 UTC from IEEE Xplore. Restrictions apply.

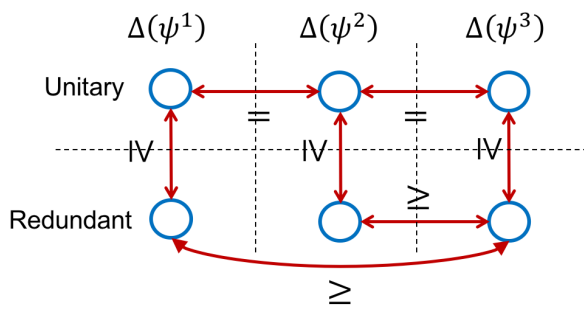

Fig. 5. The relationships of the expected error, $\Delta(\psi)$, when applying shrinkage functions that have been optimized using penalty 1..3 and in unitary vs. redundant transforms.

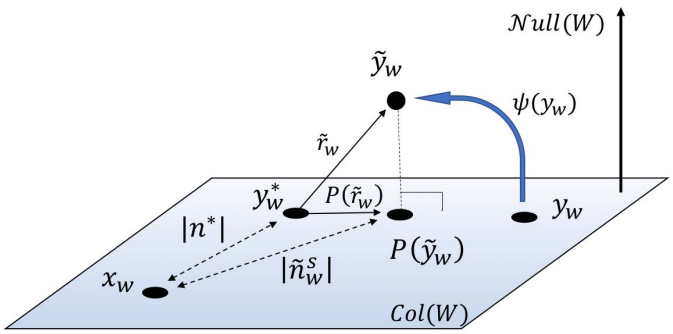

Fig. 6. Due to the triangular inequality, $\|\tilde{\mathbf{n}}_w^S\| \leq \|\mathbf{n}^*\| + \|P(\tilde{\mathbf{r}}_w)\|$. Since $\|P(\tilde{\mathbf{r}}_w)\|_E = \frac{1}{\sqrt{k}}\|\tilde{\mathbf{r}}_w\|_E$, Theorem 6 follows.

where W is composed of k shifted versions of the unitary transform U.

$$W = \frac{1}{\sqrt{k}} \begin{bmatrix} U_1 \\ U_2 \\ \vdots \\ U_k \end{bmatrix} \tag{27}$$

The redundancy rate of this transform is k, where $k \in \{1 \cdots n\}$. Namely, each subband has k shifts of the corresponding basis function. Note that if $k = 1$, the transform reduces to the unitary transform while the maximal redundancy is when $k = n$. The transform is composed of a $kn \times n$ matrix W, and is tight frame ($W^T W = I$). Denote by $Col(W)$ the column space of W (see Figure 6). The column space forms an n-dimensional subspace embedded in $\mathbb{R}^{kn}$. Since W is tight frame, it can easily be verified that the distance between two vectors in the transform domain that are in $Col(W)$ is identical to their distance in the spatial domain; i.e., let $\mathbf{z}_1 = W\mathbf{x}_1$ and $\mathbf{z}_2 = W\mathbf{x}_2$. The vectors $\mathbf{z}_1$ and $\mathbf{z}_2$ are in $Col(W)$; thus,

$$\|\mathbf{x}_1 - \mathbf{x}_2\| = \|\mathbf{z}_1 - \mathbf{z}_2\| \tag{28}$$

On the other hand, if $\mathbf{z}_1$ and $\mathbf{z}_2$ are two vectors in the transform domain that are not in $Col(W)$, their distance in the spatial domain is identical to their distance in the transform domain after projecting onto $Col(W)$. Namely, if $\mathbf{x}_1 = W^T\mathbf{z}_1$ and $\mathbf{x}_2 = W^T\mathbf{z}_2$ (i.e., $\mathbf{z}_1, \mathbf{z}_2$ are two vectors in the transform domain, and $\mathbf{x}_1, \mathbf{x}_2$ are their counterparts in the spatial domain), then

$$\|\mathbf{x}_1 - \mathbf{x}_2\| = \|P(\mathbf{z}_1) - P(\mathbf{z}_2)\| \tag{29}$$

where $P(\mathbf{z}) = WW^T\mathbf{z}$ is the projection of vector $\mathbf{z}$ onto $Col(W)$.

Denote by $\|\mathbf{n}\| = \|\mathbf{y} - \mathbf{x}\|$ and $\|\mathbf{n}_w\| = \|\mathbf{y}_w - \mathbf{x}_w\|$ the RMSE between $\mathbf{x}$ and $\mathbf{y}$ in the spatial domain and between $\mathbf{x}_w$ and $\mathbf{y}_w$ in the transform domain, respectively, as defined in Section II. Since both vectors, $\mathbf{x}_w$ and $\mathbf{y}_w$, are in $Col(W)$, the above relations give readily that $\|\mathbf{n}\| = \|\mathbf{n}_w\|$. This was also verified in Equation 12 above. After applying the shrinkage functions $\tilde{\mathbf{y}}_w = \psi(\mathbf{y}_w)$, however, the signal $\tilde{\mathbf{y}}_w$ is not necessarily in $Col(W)$. Denote by $\mathbf{y}^*$ the optimal possible reconstruction result[3] and its representation in the transform domain by $\mathbf{y}_w^* = W\mathbf{y}^*$. Clearly $\mathbf{y}_w^* \in Col(W)$ and accordingly, the optimal reconstructed RMSE is:

$$\|\mathbf{n}^*\| = \|\mathbf{y}^* - \mathbf{x}\| = \|\mathbf{y}_w^* - \mathbf{x}_w\|$$

The shrinkage functions, however, provide $\tilde{\mathbf{y}}_w$, which deviates from $\mathbf{y}_w^*$ by $\tilde{\mathbf{r}}_w$ (see Figure 6):

$$\tilde{\mathbf{r}}_w = \tilde{\mathbf{y}}_w - \mathbf{y}_w^* \tag{30}$$

Theorem 6: For an over-complete transform with redundancy k, the expected RMSE is bounded from above by:

$$\|\tilde{\mathbf{n}}_w^S\|_E \leq \|\mathbf{n}^*\|_E + \frac{1}{\sqrt{k}}\|\tilde{\mathbf{r}}_w\|_E \tag{31}$$

where $\Delta = \|\tilde{\mathbf{n}}_w^S\|_E$ is the resulting RMSE in the spatial domain. In other words, the larger the redundancy, the closer the resulting RMSE is to the optimal one and the convergence rate goes like $1/\sqrt{k}$.

Proof 6: Recall that

$$\begin{aligned} |\tilde{\mathbf{n}}_w^{\mathcal{S}}\|_E &\doteq \|\tilde{\mathbf{y}}_w^S - \mathbf{x}\|_E = \|P(\tilde{\mathbf{y}}_w) - \mathbf{x}_w\|_E \\ &= \|P(\mathbf{y}_w^* + \tilde{\mathbf{r}}_w) - \mathbf{x}_w\|_E = \|\mathbf{y}_w^* + P(\tilde{\mathbf{r}}_w) - \mathbf{x}_w\|_E \end{aligned}$$

where the second equality is due to Equation 29 and the fourth equality is due to the fact that $\mathbf{y}_w^* \in Col(W)$. Using the triangular inequality, we get:

$$\begin{aligned} &\|\mathbf{y}_w^* + P(\tilde{\mathbf{r}}_w) - \mathbf{x}_w\| \\ &\qquad \leq \|\mathbf{y}_w^* - \mathbf{x}_w\| + \|P(\tilde{\mathbf{r}}_w)\| = \|\mathbf{n}^*\| + \|P(\tilde{\mathbf{r}}_w)\| \end{aligned}$$

Thus,

$$\|\tilde{\mathbf{n}}_w^{\mathcal{S}}\| \leq \|\mathbf{n}^*\| + \|P(\tilde{\mathbf{r}}_w)\| \tag{32}$$

We follow the same argument that was used in Appendix A where we showed that due to the stationarity of natural images, for any W (any redundancy rate), we have: $\|\tilde{\mathbf{n}}_w\|_E = \|\tilde{\mathbf{n}}_u\|_E$ where $\tilde{\mathbf{n}}_w = \tilde{\mathbf{y}}_w - \mathbf{x}_w$ and $\tilde{\mathbf{n}}_u = \tilde{\mathbf{y}}_u - \mathbf{x}_u$. This argument also holds if we switch $\mathbf{x}$ with $\mathbf{y}^*$ providing:

$$\|\tilde{\mathbf{r}}_w\|_E = \|\tilde{\mathbf{r}}_u\|_E$$

where $\tilde{\mathbf{r}}_w = \tilde{\mathbf{y}}_w - \mathbf{y}_w^*$ and $\tilde{\mathbf{r}}_u = \tilde{\mathbf{y}}_u - \mathbf{y}_u^*$. This gives that the expected value of $\|\tilde{\mathbf{r}}_w\|$ is the same for any W of any redundancy rate and it equals $\|\tilde{\mathbf{r}}_u\|$. Thus, the vector $\tilde{\mathbf{r}}_w \in \mathbb{R}^{kn}$

[3]The solution is biased as we apply scalar mapping functions whereas the optimal mapping function should be a scalar field [13].

Authorized licensed use limited to: Herzeliya IDC. Downloaded on May 08,2021 at 19:53:14 UTC from IEEE Xplore. Restrictions apply.

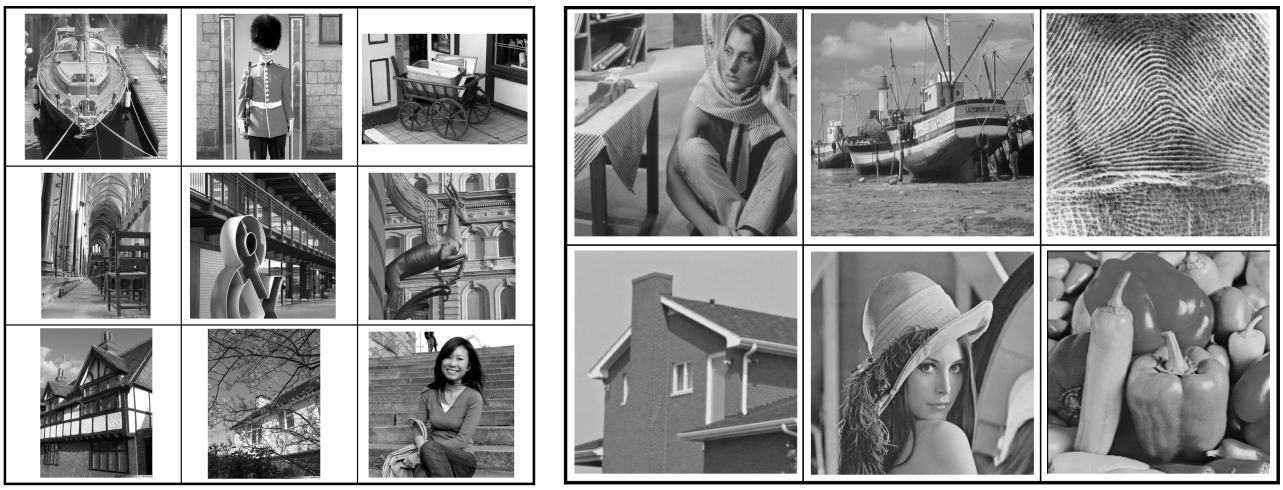

Fig. 7. Left: The images on which the SFs were trained. Right: The images on which the denoising schemes were applied.

can be seen as a random vector in $\mathbb{R}^{kn}$ whose expected length is constant for any W. Since $P(\mathbf{r}_w)$ is an orthogonal projection of a random vector from kn-dimensional space onto an n-dimensional space, the expected length of $P(\mathbf{r}_w)$ is:

$$\|P(\tilde{\mathbf{r}}_w)\|_E = \frac{1}{\sqrt{k}}\|\tilde{\mathbf{r}}_w\|_E \tag{33}$$

Moreover, the Johnson-Lindenstrauss Lemma shows that $\|P(\tilde{\mathbf{r}}_w)\|$ is fairly tight concentrated around $\|P(\tilde{\mathbf{r}}_w)\|_E$ (see [38]). Combining Equations 33 and 32, we obtain the relation given in Theorem 6. Q.E.D.

V. Results

In the previous sections we presented three different optimization schemes and how they relate to each other. In this section we test the empirical behaviour of these methods on real data. The images we used in the first part for our experiments are shown in Fig. 7.

To test the three methods, we used the optimization scheme suggested in [17] in which the shrinkage functions are modelled using piecewise linear mappings:

$$\psi_k(\mathbf{y}_k) = \mathcal{M}_k(\mathbf{y}_k; \mathbf{p}_k)$$

where $\mathbf{p}_k$ is a parameter vector controlling the piecewise function. Since the shrinkage function $\mathcal{M}_k$ is linear with respect to the parameter vector $\mathbf{p}_k$, optimizing for $\mathbf{p}_k$ can be solved in a closed form solution using a set of noisy images along with their clean counterparts. In contrast to the statistical approaches, this technique does not require estimation of a prior model for $\mathbf{y}_k$ or modelling the noise characteristics. The SFs are designed to perform "optimally" with respect to the given examples, under the assumption that they will perform equally well with similar new examples. Using Methods 1 and 2, the optimization is performed on each $\mathbf{p}_k$ independently; however, for Method 3, all $\mathbf{p}_k$ are optimized simultaneously. For more information about the optimization and the implementation, the reader is referred to [17].

In all the experiments described below, we used the undecimated windowed Discrete Cosine Transform (DCT) as the image transform. Since the DCT transform is unitary, the undecimated DCT is a tight frame. Due to the undecimated form, each wavelet band can be calculated using a single 2D convolution (with the corresponding DCT basis as the convolution kernel). Additionally, the inverse transform can be applied by convolving the rectified coefficients with the kernels forming B_k^T, which are the reflected (180 degree rotation) DCT kernels. More details are given in [17].

A. Comparing All Methods Across a Single Noise Level

In the first experiment, we tested the three shrinkage methods on six images in order to examine the performance on each one of these particular images. In the following experiments we will evaluate the methods on larger sets of images. Unless mentioned otherwise, the setting parameters were defined as follows: (1) Training images were grayscale natural images (see examples in Figure 7-left). (2) Test images were taken from Figure 7-right. The resolutions of these images are 512×512 but the "House" and the "Pepper" images whose resolution is 256×256. (3) The transform bases were the undecimated 8×8 DCT. (4) Image noise consisted of additive Gaussian noise at various STD values.

Figure 8 displays several of the SFs obtained for an 8×8 DCT basis, using the three methods described above for noise level with STD $\sigma = 20$. SFs on each row correspond to band indices (i, i) of the 8×8 DCT basis, where $i = 2..6$ (left to right). Note that a DCT band with an index (i, j) is the result of convolving the image with a DCT basis whose frequency is i along the x-axis and j along the y-axis. The top, middle and bottom rows show the SFs resulting from the first, second and the third methods, respectively. It can be seen that the SFs of the three methods differ from each other because each takes into consideration different statistical correlations as explained above.

The obtained SFs were applied to several images[4] shown in Figure 7-right. Figure 9 compares the resulting MSE across the 3 methods and for 6 different images. Each bar

[4]Taken from http://decsai.ugr.es/javier/denoise/test_images/index.htm

Authorized licensed use limited to: Herzeliya IDC. Downloaded on May 08,2021 at 19:53:14 UTC from IEEE Xplore. Restrictions apply.

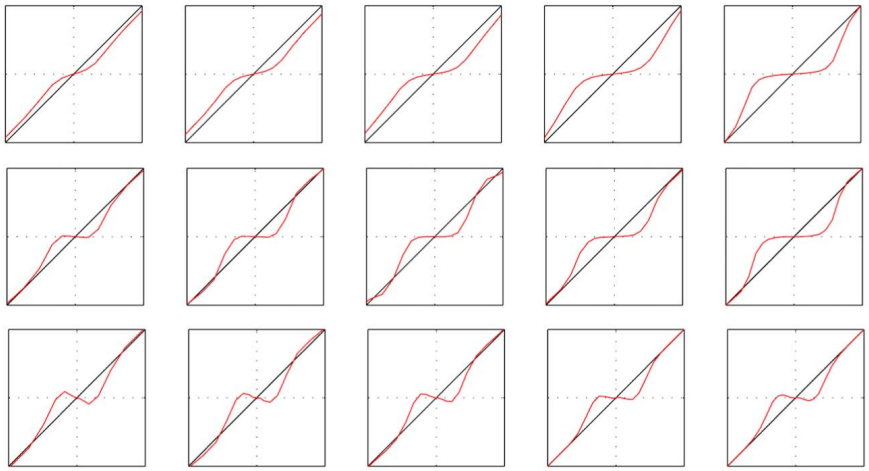

Fig. 8. Comparison of the produced SFs using Method 1 (top row), Method 2 (middle row), and Method 3 (bottom row). SFs on each row correspond to band (i, i) of the 8×8 DCT basis, where $i = 2..6$ (left to right). Graph axes are shown in the range [-120,120]. Each graph represents the coefficient mapping, where the x-axis is the input value and the y-axis is the rectified (output) value. The black line represents the identity mapping.

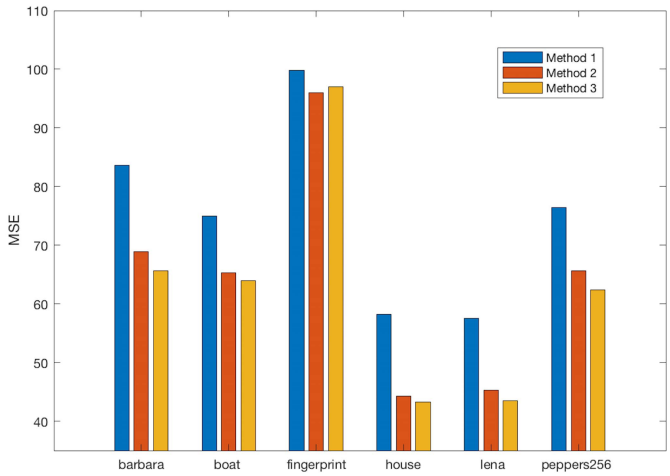

Fig. 9. MSE after applying the SFs produced by methods 1–3. Each bar is an average over 10 different noise realizations.

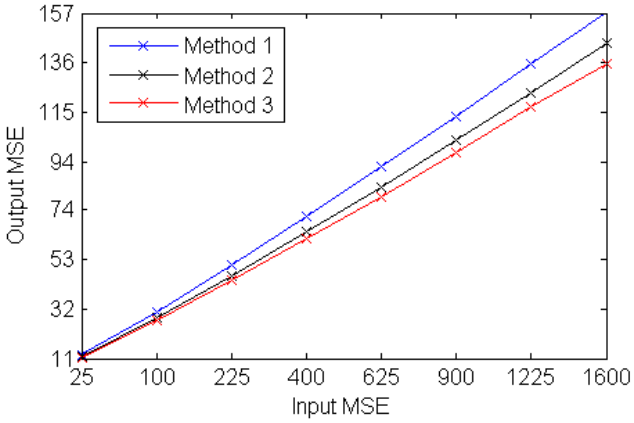

Fig. 10. The resulting MSE with respect to various noise levels. The plot shows a comparisons for the different optimization schemes. As expected, the optimal optimization scheme is Method 3. The x-axis shows the input MSE and the y-axis presents the output MSE.

in the figure presents the resulting MSE averaged over 10 realizations of noise with $\sigma = 20$. The results demonstrate the improvement of the second method over the first method, and the superiority of the third method over the other two. It can be seen that most of the improvement is achieved when applying the objective in the spatial domain (Method 2). Further improvement, although less significant, is achieved when incorporating the band dependencies (Method 3). Note, however, that Method 2 slightly improves over Method 3 for the FINGERPRINT image. This result is incontistent with Equation 23. The reason for this outcome is that the training set for this experiment does not seem to form a good representative of the statistics of the textured FINGERPRINT image. This implies that $\hat{\psi}_2$, $\hat{\psi}_3$ are not necessarily the optimal SFs that minimize Δ_2, Δ_3, respectively. Indeed, training the SFs on statistically similar images and applying them on the same noisy FINGERPRINT image, provides an MSE value of 89.55 for Method 3 (compared to 97.07 in the current plot) and MSE values of 91.43 and 98.36 for methods 2 and 1, respectively.

B. Comparing All Methods on Various Noise Levels

In the following experiments, we tested the relationships between the different methods and for various Gaussian noise levels. Performance was tested for eight different equally spaced noise levels, ranging from $\sigma = 5$ up to $\sigma = 40$. We tested the six images presented in Figure 7. We used each image in turn as a training image, from which the SFs are constructed, and the remaining images in the set served as test images. This was performed with 5 different noise realizations for each noise level. The resulting scores for 8 different noise levels can be seen in Fig. 10. Each value in this plot is the mean over all noise realizations and over all images. The x-axis shows the input MSE and the y-axis presents the output MSE. It can be seen that for each noise level $\Delta(\psi_1) \geq \Delta(\psi_2) \geq \Delta(\psi_3)$. Thus, optimizing the SFs using Method 3 gives the optimal result while Method 2 gives better bounds than Method 1 which is the traditional optimization method.

Authorized licensed use limited to: Herzeliya IDC. Downloaded on May 08,2021 at 19:53:14 UTC from IEEE Xplore. Restrictions apply.

TABLE I

IMAGE DENOISING FOR NOISE LEVELS $\sigma = 10{:}10{:}70$. PSNR AND SSIM SCORES ARE SHOWN FOR EACH NOISE LEVEL. SCORE VALUES INDICATE THE MEAN OVER 65 IMAGES FROM THE CBSD68 DATASET

STD	Score	Method 1	Method 2	Method 3
10	PSNR	31.8999	32.4799	**32.8171**
	SSIM	0.8610	0.8860	**0.9020**
20	PSNR	28.2266	28.7658	**29.0639**
	SSIM	0.7513	0.7858	**0.8078**
30	PSNR	26.4177	26.8723	**27.1420**
	SSIM	0.6766	0.7115	**0.7339**
40	PSNR	25.2718	25.6445	**25.8843**
	SSIM	0.6229	0.6560	**0.6758**
50	PSNR	24.4302	24.7742	**24.9928**
	SSIM	0.5806	0.6134	**0.6295**
60	PSNR	23.7813	24.0794	**24.2715**
	SSIM	0.5478	0.5782	**0.5904**
70	PSNR	23.2260	23.5209	**23.6623**
	SSIM	0.5207	0.5512	**0.5556**

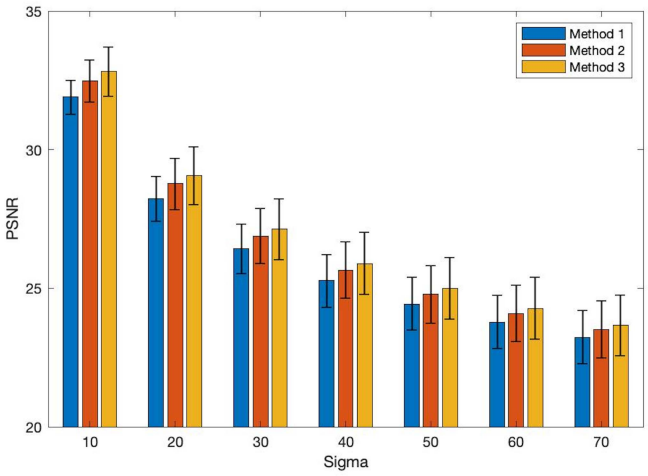

Fig. 11. The PSNR scores averaged over 65 images from the CBSD68 benchmark dataset. For each noise level the resulting PSNR and error bar are shown.

In the next experiment we tested the proposed methods on a large set of images and calculated the PSNR as well as the perceptual SSIM score [39] of the denoised results. We used 68 images from the CBSD68[5] benchmark dataset and applied methods 1-3 using 7 levels of noise whose σ =10:10:70. The first 3 images in the dataset were used for training the SFs, and the rest of the images were tested for denoising. The score values were averaged over the image set. Table I shows the denosing results. It can be seen that for all noise levels, the PSNRs of Methods 3 are greater than Method 2, and the PSNRs of Method 2 are greater than those of Method 1. Since the PSNR score is monotonic with the MSE score, this finding is expected. However, it can be seen that this order is also preserved for the SSIM scores which is not necessarily monotonic with the MSE. Figure 11 shows the PSNR scores shown as a bar plot along with error bars for each noise-level/method.

[5]https://github.com/clausmichele/CBSD68-dataset

TABLE II

JPEG COMPRESSION ARTIFACT CORRECTION. FOR EACH COMPRESSION RATE PSNR AND SSIM SCORES ARE SHOWN. SCORE VALUES INDICATE THE MEAN OVER 50 RANDOMLY SELECTED IMAGES FROM THE DIV2K800 DATASET

Quality	Score	Method 1	Method 2	Method 3
5	PSNR	26.8620	26.9937	**27.1237**
	SSIM	0.7331	0.7404	**0.7439**
10	PSNR	29.4808	29.6079	**29.7068**
	SSIM	0.8363	0.8415	**0.8425**
20	PSNR	31.8405	31.9474	**32.0541**
	SSIM	0.8936	0.8962	**0.8969**
30	PSNR	33.2380	33.3315	**33.4530**
	SSIM	0.9179	0.9199	**0.9202**
40	PSNR	34.2319	34.3148	**34.4478**
	SSIM	0.9315	0.9331	**0.9333**

TABLE III

IMAGE SCALING BY A BILINEAR INTERPOLATION THEN SHRINKAGE CORRECTION. SCORE VALUES INDICATE THE MEAN OVER 100 IMAGES FROM THE BSD100 DATASET

Scaling	Score	Method 1	Method 2	Method 3
$\times 2$	PSNR	28.9384	29.0803	**29.1733**
	SSIM	0.8602	**0.8652**	0.8639
$\times 3$	PSNR	26.2243	26.2681	**26.3768**
	SSIM	0.7427	0.7432	**0.7460**
$\times 4$	PSNR	24.8835	24.9390	**25.0221**
	SSIM	0.6615	**0.6644**	0.6631

Figure 12 presents visual examples of the denoising results for methods 1-3. It can be visually verified that Method 3 produces the cleanest and the most visually pleasing result.

C. Cost-Effective Analysis

Optimizing SFs according to Method 2 offers both tight bounds on the mean MSE distortion and fast implementation. Figure 13-left shows the fraction of deviation of Methods 1 and 2 from the optimal scheme (Method 3). It can be seen that Method 2 deviates 4% on average relative to Method 3, while Method 1 deviates about 16%.

Figure 13-right shows the computation time required to train the SFs for each method. It can be seen that training using the optimization scheme of Method 2 takes one quarter of the time needed by Method 3. Thus, Method 2 introduces a cost-efficient advantage; By allowing a deviation of 4% in quality from the optimal optimization scheme, an average gain of 70% in speed can be obtained.

D. Shrinkage Framework for Additional Corruption Models

In the following experiments we apply the shrinkage framework to two additional reconstruction cases. We used the shrinkage framework to remove JPEG artifacts and to improve image upscaling.

In the de-JPEG case, we compress and uncompress an image using the JPEG compression scheme at various compression rates. The larger the compression rate, the more artifacts introduced in the uncompressed images. In order to remove compression artifacts, the shrinkage method is applied. A set of SFs are trained on clean and uncompressed versions

Authorized licensed use limited to: Herzeliya IDC. Downloaded on May 08,2021 at 19:53:14 UTC from IEEE Xplore. Restrictions apply.

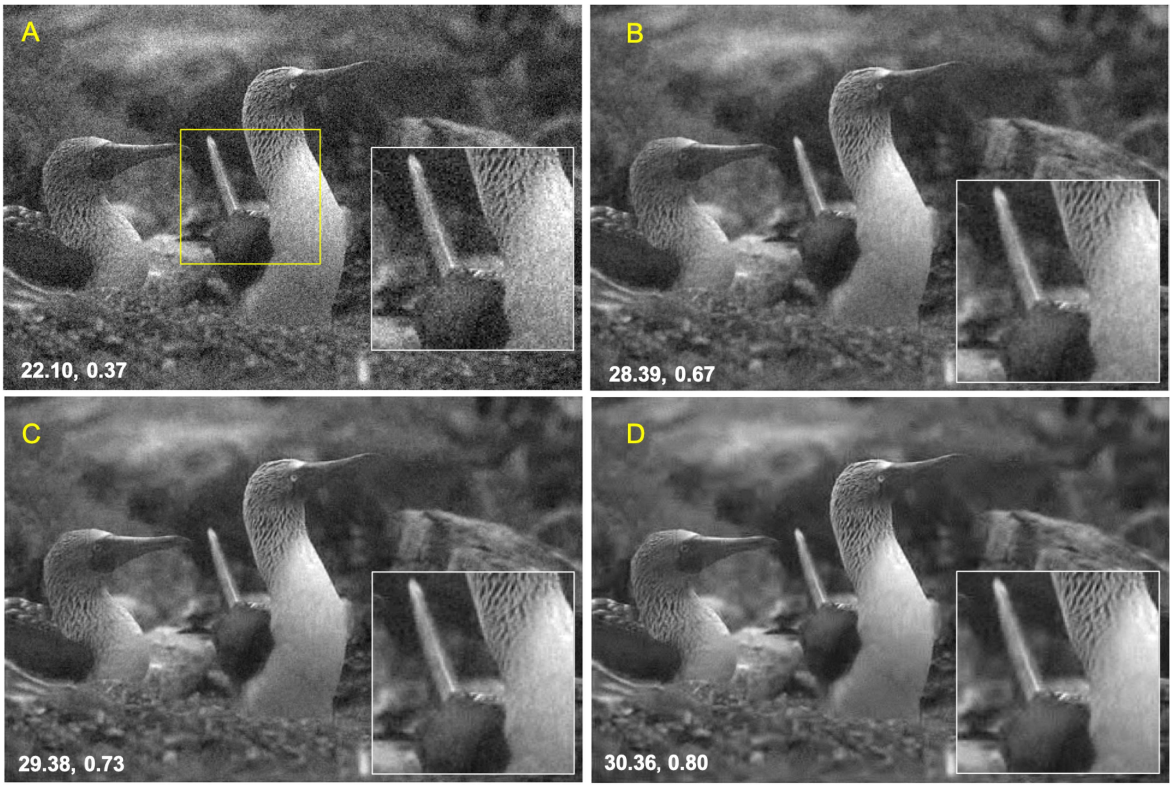

Fig. 12. Denoising results for noisy image with noise level $\sigma = 20$. A) The original noisy image. B,C,D) The resulting images for methods 1,2,3 respectively. The PSNR and the SSIM scores are written on each image.

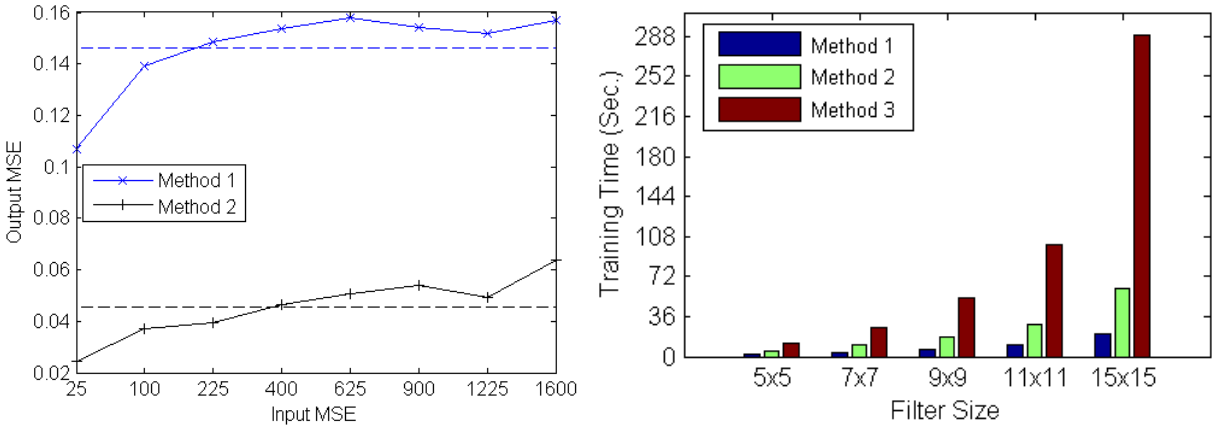

Fig. 13. Left: The plot shows the relative deviation of the output MSE in Methods 1 and 2 relative to Method 3. The x-axis shows the input MSE and the y-axis presents the relative MSE with respect to Method 3. The deviation is approx. 4% for Method 2 and approx. 16% for Method 1. Right: This plot shows the training time required for each method.

of images. The original image (before compression) and the uncompressed image provide clean and "noisy" counterparts on which the optimal SFs are trained, using methods 1-3. We tested the restoration results on 50 randomly selected images from the DIV2K800 dataset [40]. The SFs were trained on 5 images taken from the same dataset (selected randomly).

Table II presents the resulting scores in PSNR and SSIM for various compression rates (the lower the quality rate, the more compression is applied, where 100% quality rate indicates no compression). It can be seen from the table that Method 3 outperforms the other two methods, and that Method 2 provides better results than Method 1. The above is true for the PSNR and for the SSIM scores as well. Figure 14 shows examples of the de-JPEG results using the 3 methods. The improvement in image quality due to removing JPEG artifacts is clearly visible, in particular when using Method 3.

In the second case, we use the shrinkage framework for image upscaling. In this experiment low resolution images are scale up using naive bilinear interpolation. These up scaled images are of low quality, since the high frequencies are not reproduced correctly. Using a set of SFs, the shrinkage framework is applied to improve the degraded images. In order to train the SFs (using methods 1-3) we used a set of high res images along with their degraded counterparts that were

Authorized licensed use limited to: Herzeliya IDC. Downloaded on May 08,2021 at 19:53:14 UTC from IEEE Xplore. Restrictions apply.

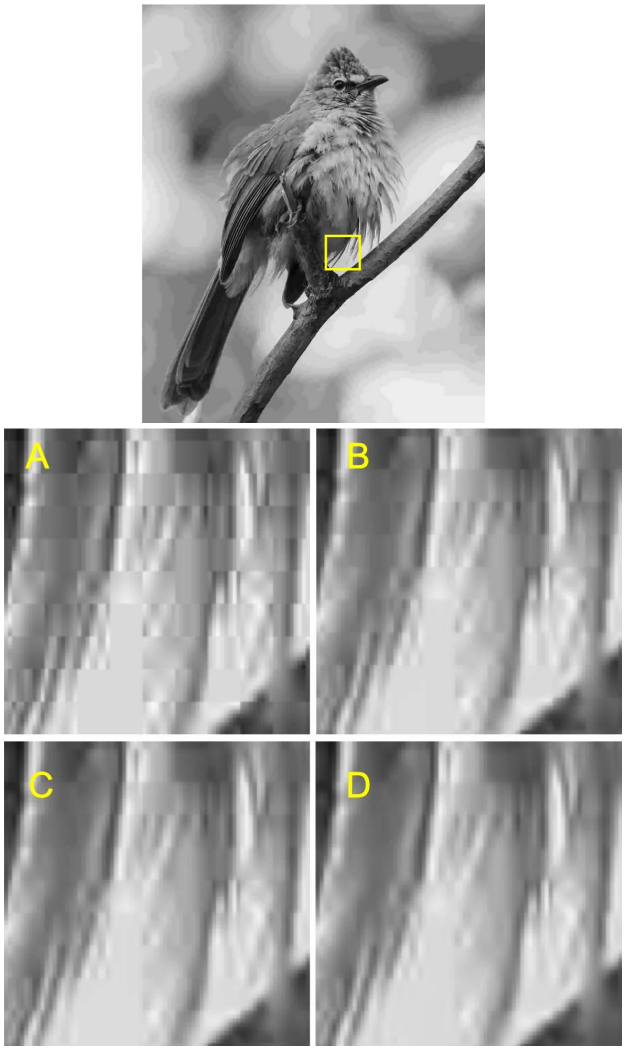

Fig. 14. Top image - The original uncompressed image. A) A scaled region of the uncompressed image. JPEG artifacts are clearly visible B,C,D) Scaled versions of the de-jpeg results using methods 1,2,3, respectively.

Fig. 15. Up scaled image using $\times 2$ bilinear interpolation (left) and a sharpened version using the shrinkage method (right).

generated by downscaling and upscaling the images using bilinear interpolation.

Table III presents the resulting scores in PSNR and SSIM for various scaling factors. The scores are the mean over 100 images from the BSD100 dataset.[6] For the PSNR score,

[6]https://deepai.org/dataset/bsd100-4x-upscaling

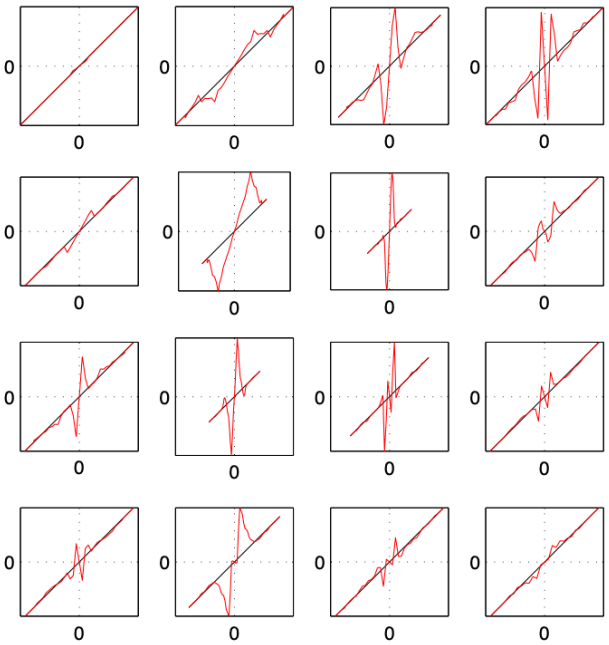

Fig. 16. The SFs for upscaling and sharpen images after $\times 2$ upscaling. The SFs shown are for the 8×8 DCT bands whose indices are: $[3..6] \times [3..6]$ (left to right $\times$ top to bottom). Each graph represents the coefficient mapping, where the x-axis is the input value and the y-axis is the rectified (output) value. The black line represents the identity mapping.

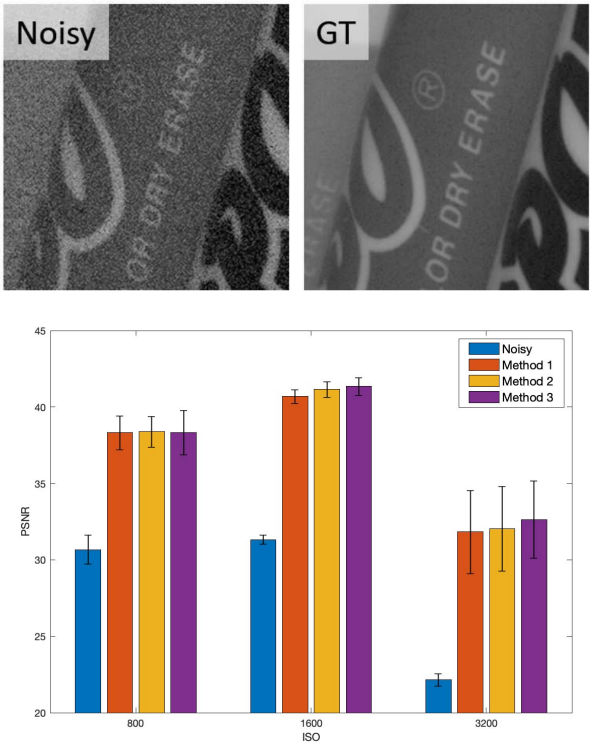

Fig. 17. Top row: An example of a noisy image and its estimated ground truth. Image is taken from [41]. Bottom row: Average PSNR values for the noisy images (blue bars) and after denoising with Methods 1-3 (orange, yellow, and purple bars). Results are shown for three ISO levels. Note that all PSNR values were calculated with respect to the estimated ground truth.

Method 3 outperforms the other two methods in all cases. For the SSIM score, Method 2 provides better scores in two cases. However this result does not contradict Equation 23 as

Authorized licensed use limited to: Herzeliya IDC. Downloaded on May 08,2021 at 19:53:14 UTC from IEEE Xplore. Restrictions apply.

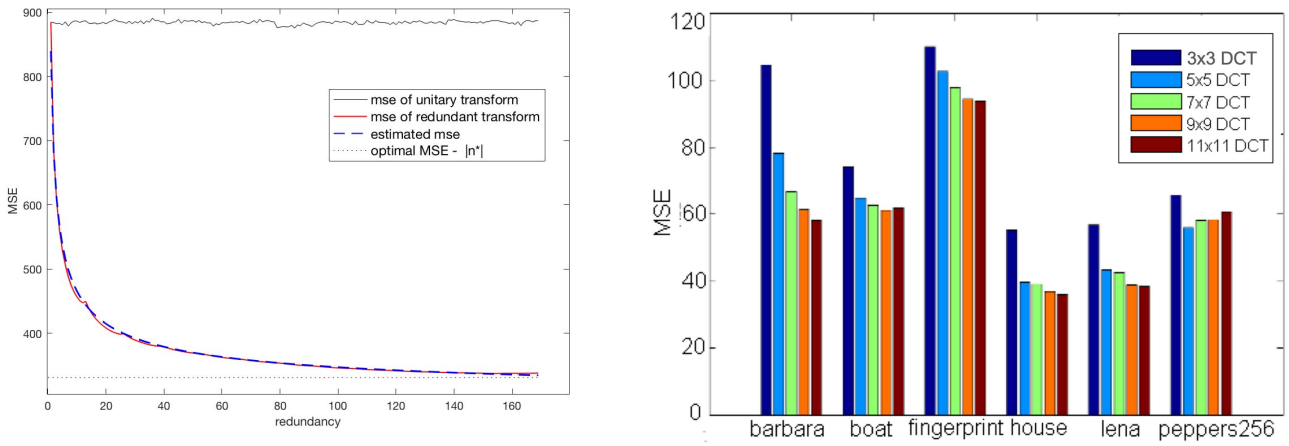

Fig. 18. The performance is improved with the redundancy; The output MSE as a function of the redundancy rate. Left: The filter size was 13×13 and the noise STD was $\sigma = 50$. The plot shows the RMSE as a function of redundancy rate. Right: The wider the filter, the more redundancy achieved and the better the denoising results.

the SSIM is not monotonic with the MSE score. Visual results of using shrinkage functions for image upscaling can be seen in Figure 15.

The SFs that were generated by Method 3 are shown in Figure 16. The SFs shown are for the 8×8 DCT bands whose indices are: $[3..6] \times [3..6]$ (left to right $\times$ top to bottom). In contrast to the typical SFs for image denoising where low values in the transform domain are attenuated, the SFs for image upscaling boost (up or down) low values in order to increase the associated high frequencies.

In the last experiment we tested the three methods on images with real world noise. We used the *Smartphone Image Denoising Dataset* (SIDD).[7] The noisy images in this dataset are given along with their estimated ground truth [41]. This dataset was captured by five representative smartphone cameras, under different light brightness levels, illumination temperatures, and different ISO levels. The ground truth of the noisy images were estimated by capturing a sequence of images and applying a noise estimation procedure [41]. Figure 17-top row, shows an example of a noisy image and its estimated ground truth. The images used in this experiment were from the *SIDD small dataset*. For each ISO level, a set of nine images was used as the training set, from which a set of shrinkage functions were estimated (using the ground truth counterparts). We tested Methods 1-3 on 120 images, randomly selected from the dataset. Figure 17-bottom row, presents the resulting average PSNR of the original noisy images (blue bars), and the resulting PSNR for Methods 1-3. Three clusters of bars are presented, one for each ISO value. It can be seen that all methods significantly improve the original PSNR, however, the PSNR of Method 3 is not significantly higher than Methods $1 + 2$, showing comparable results for ISO 800 and slightly better results for ISO 1600 and 3200. Although it is difficult to provide an educated explanation for this behaviour, we speculate that since the images were acquired using various smartphone cameras, under different light brightness levels and temperatures conditions, the statistics of the noise is not distinctive and the training regresses the shrinkage functions to the mean. Further profiling of the noise statistics for each sub-category is left for future study.

[7]https://www.eecs.yorku.ca/ kamel/sidd

E. Denoising Improvements Vs. Redundancy Rate

To validate the observations concerning the denoising improvement with respect to the redundancy rate, we measured the resulting MSEs for various redundancy rates. Figure 18-left shows the MSE resulting from denoising applied to a 13×13 windowed DCT transform, where the redundancy rates were implemented by shifting the basis functions by (i, j) along the x-axis and the y-axis, where $(i, j) \in \{0, 12\} \times \{0, 12\}$. Thus, the redundancy rate ranges between $k = 1$ and $k = 169 = 13 \times 13$. We assume the optimal MSE is given for $k = 169$ (maximum redundancy) where in this case $\hat{\mathbf{y}} = \mathbf{y}^*$ and $\|\mathbf{n}^*\| = \|\mathbf{x} - \mathbf{y}^*\| = \mathcal{E}_{opt}$. On the other hand, when $k = 1$, we have that $\tilde{\mathbf{y}}_w \in Col(W)$ (in this case $W = U$) and thus $\|P(\tilde{\mathbf{r}}_w)\| = \|\tilde{\mathbf{r}}_u\| = \|\tilde{\mathbf{y}}_u^S - \mathbf{y}^*\| = \Delta^*$. Accordingly, following Theorem 6, the RMSE should be:

$$RMSE(k) \leq \mathcal{E}_{opt} + \frac{1}{\sqrt{k}}\Delta^* \tag{34}$$

Figure 18-left shows the decrease in the measured MSE as a function of the redundancy rate. The solid red curve shows the actual measures while the dashed blue curve shows the expected RMSE following Equation 34. It can be seen that the two plots are very similar. The measures were taken for the BARBARA image where the noise level was $\sigma = 50$. Each MSE measure in this plot is an average over five noise realizations.

Figure 18-right shows the optimal achieved MSE for each DCT transform, of $5, 7, 9$ and 11 pixels wide. It can be seen that, in most cases, the wider the filter, the greater the redundancy that can be achieved and the better the denoising results.

VI. Conclusion

Transform denoising using shrinkage functions is a classical framework that is widely used in numerous applications.

Authorized licensed use limited to: Herzeliya IDC. Downloaded on May 08,2021 at 19:53:14 UTC from IEEE Xplore. Restrictions apply.

In this paper we provide a complete picture of the interrelations between the transform used, the optimal shrinkage functions, and the domains in which they are optimized. In particular, we provide a theoretical justification for applying the shrinkage functions in the transform domain and the benefit of applying them in redundant bases.

Additionally, we provided theoretical bounds for the three possible optimization schemes of the shrinkage functions. We showed that for subband optimization, where each ψ_i is optimized independently, optimizing the shrinkage function in the spatial domain is always better than or equal to optimizing them in the transform domain. This option, besides being simple to implement, is proven to outperform the traditional transform domain optimization while avoiding the demanding spatial domain optimization of all the shrinkage functions, simultaneously.

For redundant bases, we provided the expected denoising gain we may achieve, relative to the unitary basis, as a function of the basis redundancy. This result allows a user to make a clever decision about the redundancy used by taking into account the expected denoising gain and the computational time allocated for this process.

Appendix A

Theorem 1: After denoising, the expected MSE distortions in the transform domain are equal for the unitary and for the redundant transforms, i.e.:

$$\|\tilde{\mathbf{n}}_u\|_E = \|\tilde{\mathbf{n}}_w\|_E$$

Proof 1: The MSE value $\|\tilde{\mathbf{n}}_u\|_E^2$ is defined as:

$$\|\tilde{\mathbf{n}}_u\|_E^2 = E\left\{\|\tilde{\mathbf{n}}_u\|^2\right\} = \int \|U\mathbf{x} - \psi\{U\mathbf{y}\}\|^2 \;\; P(\mathbf{x},\mathbf{y})d\mathbf{x}d\mathbf{y}$$

where $P(\mathbf{x},\mathbf{y})$ denotes the probability distribution function of $(\mathbf{x},\mathbf{y})$. By changing variables, the above expression can be rewritten as:

$$\|\tilde{\mathbf{n}}_u\|_E^2 = \int \|US_i\mathbf{x} - \psi\{US_i\mathbf{y}\}\|^2 \;\; P(S_i\mathbf{x}, S_i\mathbf{y})|\det(S_i^T)| \;\; d\mathbf{x}d\mathbf{y}$$

where S_i is a shift operator by the ith displacement. Now, we exploit the stationary property of natural images. This property gives that for each S_i:

$$P(S_i\mathbf{x}, S_i\mathbf{y}) = P(\mathbf{x},\mathbf{y})$$

Additionally, we can apply the (adjoint) shift operator to the transform basis, rather than to the images. Using the notation $US_i = U_i$ (Equation 6) and having $|\det(S_i^T)| = 1$, we get:

$$\|\tilde{\mathbf{n}}_u\|_E^2 = \int \|U_i\mathbf{x} - \psi\{U_i\mathbf{y}\}\|^2 P(\mathbf{x},\mathbf{y})d\mathbf{x}d\mathbf{y} = \|\tilde{\mathbf{n}}_{u_i}\|_E^2 \quad (35)$$

Now, since $\|\tilde{\mathbf{n}}_w\|_E^2 = \frac{1}{n}\sum_{i=1}^n \|\tilde{\mathbf{n}}_{u_i}\|_E^2$, we conclude:

$$\|\tilde{\mathbf{n}}_w\|_E^2 = \frac{1}{n}\sum_{i=1}^n \left\|\tilde{\mathbf{n}}_{u_i}\right\|_E^2 = \frac{1}{n}\sum_{i=1}^n \|\tilde{\mathbf{n}}_u\|_E^2 = \|\tilde{\mathbf{n}}_u\|_E^2 \qquad \text{Q.E.D.}$$

Appendix B

Theorem 2: For any given ψ,

$$\|\tilde{\mathbf{n}}_u^{\mathcal{S}}\|_E \geq \|\tilde{\mathbf{n}}_w^{\mathcal{S}}\|_E$$

Proof 2: After Equation 13 and Theorem 1, we have:

$$\|\tilde{\mathbf{n}}_u^{\mathcal{S}}\|_E = \|\tilde{\mathbf{n}}_u\|_E = \|\tilde{\mathbf{n}}_w\|_E$$

However, from Equation 14 it follows that

$$\|\tilde{\mathbf{n}}_w\|_E \geq \|\tilde{\mathbf{n}}_w^{\mathcal{S}}\|_E$$

and therefore:

$$\|\tilde{\mathbf{n}}_u^{\mathcal{S}}\|_E \geq \|\tilde{\mathbf{n}}_w^{\mathcal{S}}\|_E \qquad \text{Q.E.D.}$$

Acknowledgment

The authors would like to thank Michael Elad for helpful discussions and useful suggestions while writing this manuscript.

References

[1] D. L. Donoho and I. M. Johnston, "Ideal denoising in an orthonormal basis chosen from a library of bases," *CR Acad. Sci.*, vol. 319, pp. 1317–1322, Dec. 1994.
[2] H. R. Shahdoosti and Z. Rahemi, "Edge-preserving image denoising using a deep convolutional neural network," *Signal Process.*, vol. 159, pp. 20–32, Jun. 2019.
[3] C. Tian, L. Fei, W. Zheng, Y. Xu, W. Zuo, and C.-W. Lin, "Deep learning on image denoising: An overview," 2019, *arXiv:1912.13171*. [Online]. Available: http://arxiv.org/abs/1912.13171
[4] H. C. Burger, C. J. Schuler, and S. Harmeling, "Image denoising: Can plain neural networks compete with BM3D?" in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2012, pp. 2392–2399.
[5] K. Isogawa, T. Ida, T. Shiodera, and T. Takeguchi, "Deep shrinkage convolutional neural network for adaptive noise reduction," *IEEE Signal Process. Lett.*, vol. 25, no. 2, pp. 224–228, Feb. 2018.
[6] B. A. Olshausen and D. J. Field, "Emergence of simple-cell receptive field properties by learning a sparse code for natural images," *Nature*, vol. 381, no. 6583, pp. 607–609, Jun. 1996.
[7] B. A. Olshausen and D. J. Field, "Natural image statistics and efficient coding," *Netw., Comput. Neural Syst.*, vol. 7, no. 2, pp. 333–339, Jan. 1996.
[8] J. Hurri, A. Hyv, R. Karhunen, and E. Oja, "Wavelets and natural image statistics," in *Proc. Scandin. Conf. Image Anal.*, Lappenranta, Finland, 1998, pp. 13–18.
[9] S. G. Mallat, "A theory for multiresolution signal decomposition: The wavelet representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 11, no. 7, pp. 674–693, Jul. 1989.
[10] D. L. Donoho, "De-noising by soft-thresholding," *IEEE Trans. Inf. Theory*, vol. 41, no. 3, pp. 613–627, May 1995.
[11] D. L. Donoho and I. M. Johnstone, "Ideal spatial adaptation by wavelet shrinkage," *Biometrika*, vol. 81, no. 3, pp. 425–455, Sep. 1994.
[12] E. P. Simoncelli and E. H. Adelson, "Noise removal via Bayesian wavelet coring," in *Proc. 3rd IEEE Int. Conf. Image Process.*, vol. 1. Lausanne, Switzerland: IEEE Sig Proc Society, Dec. 1996, pp. 379–382.
[13] M. Raphan and E. P. Simoncelli, "Optimal denoising in redundant representations," *IEEE Trans. Image Process.*, vol. 17, no. 8, pp. 1342–1352, Aug. 2008.
[14] E. P. Simoncelli, "Bayesian denoising of visual images in the wavelet domain," in *Bayesian Inference in Wavelet Based Models*. New York, NY, USA: Springer, 1999, pp. 291–308.
[15] M. Shahdloo, E. Ilicak, M. Tofighi, E. U. Saritas, A. E. Çetin, and T. Çukur, "Projection onto epigraph sets for rapid self-tuning compressed sensing mri," *IEEE Trans. Med. Imag.*, vol. 38, no. 7, pp. 1677–1689, Dec. 2018.
[16] M. Tofighi, K. Kose, and A. E. Cetin, "Denoising using projections onto the epigraph set of convex cost functions," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Oct. 2014, pp. 2709–2713.

Authorized licensed use limited to: Herzeliya IDC. Downloaded on May 08,2021 at 19:53:14 UTC from IEEE Xplore. Restrictions apply.

[17] Y. Hel-Or and D. Shaked, "A discriminative approach for wavelet denoising," *IEEE Trans. Image Process.*, vol. 17, no. 4, pp. 443–457, Apr. 2008.
[18] A. Adler, Y. Hel-Or, and M. Elad, "A weighted discriminative approach for image denoising with overcomplete representations," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, Mar. 2010, pp. 782–785.
[19] J. Sun and Z. Xu, "Color image denoising via discriminatively learned iterative shrinkage," *IEEE Trans. Image Process.*, vol. 24, no. 11, pp. 4148–4159, Nov. 2015.
[20] R. R. Coifman and D. L. Donoho, "Translation invariant de-noising," in *Wavelets and Statistics*, A. Antoniadis and G. Oppenheim, Eds. New York, NY, USA: Springer-Verlag, 1995, pp. 125–150.
[21] E. J. Candès, "Harmonic analysis of neural networks," *Appl. Comput. Harmon. Anal.*, vol. 6, no. 2, pp. 197–218, Mar. 1999.
[22] P. Carré and D. Helbert, "Ridgelet decomposition: Discrete implementation and color denoising," *Proc. SPIE Wavelet Appl. Ind. Process.*, vol. 6001, Nov. 2005, Art. no. 60010F.
[23] N. Nezamoddini-Kachouie, P. Fieguth, and E. Jernigan, "Bayesshrink ridgelets for image denoising," in *Proc. ICIAR*, Porto, Portugal, Sep. 2004, pp. 163–170.
[24] M. N. Do and M. Vetterli, "The contourlet transform: An efficient directional multiresolution image representation," *IEEE Trans. Image Process.*, vol. 14, no. 12, pp. 2091–2106, Dec. 2005.
[25] B. Matalon, M. Elad, and M. Zibulevsky, "Image denoising with the contourlet transform," in *Proc. SPARSE*, Rennes, France, Nov. 2005.
[26] J.-L. Starck, E. J. Candes, and D. L. Donoho, "The curvelet transform for image denoising," *IEEE Trans. Image Process.*, vol. 11, no. 6, pp. 670–684, Jun. 2002.
[27] M. Elad, B. Matalon, and M. Zibulevsky, "Image denoising with shrinkage and redundant representations," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, Jun. 2006, pp. 1924–1931.
[28] W. Dong, X. Li, L. Zhang, and G. Shi, "Sparsity-based image denoising via dictionary learning and structural clustering," in *Proc. CVPR*, Jun. 2011, pp. 457–464.
[29] H. Krim, D. Tucker, S. Mallat, and D. Donoho, "On denoising and best signal representation," *IEEE Trans. Inf. Theory*, vol. 45, no. 7, pp. 2225–2238, Nov. 1999.
[30] N. Ouarti and G. Peyre, "Best basis denoising with non-stationary wavelet packets," in *Proc. 16th IEEE Int. Conf. Image Process. (ICIP)*, Nov. 2009, pp. 3825–3828.
[31] M. Elad, "Why simple shrinkage is still relevant for redundant representations?" *IEEE Trans. Inf. Theory*, vol. 52, no. 12, pp. 5559–5569, Dec. 2006.
[32] J.-L. Starck, J. Fadili, and F. Murtagh, "The undecimated wavelet decomposition and its reconstruction," *IEEE Trans. Image Process.*, vol. 16, no. 2, pp. 297–309, Feb. 2007.
[33] G. H. Golub and C. F. Van Loan, *Matrix Computations*, vol. 3. Baltimore, MD, USA: JHU Press, 2012.
[34] A. Hyvärinen, P. O. Hoyer, and E. Oja, "Sparse code shrinkage: Denoising by nonlinear maximum likelihood estimation," in *Proc. Adv. Neural Inf. Process. Syst.*, 1999, pp. 473–479.
[35] P. Moulin and J. Liu, "Analysis of multiresolution image denoising schemes using generalized-Gaussian priors," in *Proc. IEEE-SP Int. Symp. Time-Frequency Time-Scale Anal.*, 1998, pp. 633–636.
[36] F. Xiao and Y. Zhang, "A comparative study on thresholding methods in wavelet-based image denoising," *Procedia Eng.*, vol. 15, pp. 3998–4003, Dec. 2011.
[37] Y. Hel-Or and D. Shaked, "Slicing the transform discriminative approach for Wavelet denoising," Hewlett-Packard Labs, Palo Alto, CA, USA, Tech. Rep. HPL-2006-103R1, 2006.
[38] S. Dasgupta and A. Gupta, "An elementary proof of a theorem of Johnson and lindenstrauss," *Random Struct. Algorithms*, vol. 22, no. 1, pp. 60–65, Jan. 2003.
[39] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, Apr. 2004.
[40] E. Agustsson and R. Timofte, "NTIRE 2017 challenge on single image super-resolution: Dataset and study," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jul. 2017, pp. 126–135.
[41] A. Abdelhamed, S. Lin, and M. S. Brown, "A high-quality denoising dataset for smartphone cameras," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 1692–1700.

Yacov Hel-Or received the Ph.D. degree in computer science from Hebrew University, Jerusalem. He is currently a Full Professor with the School of Computer Science, The Interdisciplinary Center (IDC), Israel, where he has been a Faculty Member, since 1998. He was a Postdoctorate Fellow with the Department of Applied Mathematics and Computer Science, The Weizmann Institute of Science, and with the NASA Ames Research Center, Mountain View, CA, USA. He has been a Visiting Scholar with Google and HP Labs and a Principal Scientist with Amazon. His recent interests include the fields of computer vision, image processing, and deep learning.

Gil Ben-Artzi received the B.Sc. and M.Sc. *(cum laude)* degrees in mathematics and computer science from Bar-Ilan University, Israel, and the Ph.D. degree from Hebrew University, Israel. Prior to that, he was a Postdoctoral Researcher with the Weizmann Institute (IL) and Intel Labs, Santa Clara, CA, USA. He is currently an Assistant Professor with Ariel University, Israel. His research interests include deep learning in computer vision and image processing.

Authorized licensed use limited to: Herzeliya IDC. Downloaded on May 08,2021 at 19:53:14 UTC from IEEE Xplore. Restrictions apply.