Weighted Averaging for Denoising with Overcomplete Dictionaries

Onur G. Guleryuz
DoCoMo Communications Laboratories USA, Inc.
3240 Hillview Avenue,
Palo Alto, CA 94304
guleryuz@docomolabs-usa.com, 650-496-4719, 650-493-9601(fax)

Abstract

We consider the scenario where additive, independent and identically distributed (i.i.d) noise in an image is removed using an overcomplete set of linear transforms and thresholding. Rather than the standard approach where one obtains the denoised signal by ad hoc averaging of the denoised estimates provided by denoising with each of the transforms, we formulate the optimal combination as a conditional linear estimation problem and solve it for optimal estimates. Our approach is independent of the utilized transforms and the thresholding scheme, and as we illustrate using oracle based denoisers, it extends established work by exploiting a separate degree of freedom that is in general not reachable using previous techniques. Our derivation of the optimal estimates specifically relies on the assumption that the utilized transforms provide sparse decompositions. At the same time, our work is robust as it does not require any assumptions about image statistics beyond sparsity. Unlike existing work which tries to devise ever more sophisticated transforms and thresholding algorithms to deal with the myriad types of image singularities, our work uses basic tools to obtain very high performance on singularities by taking better advantage of the sparsity that surrounds them. With well-established transforms we obtain results that are competitive with state-of-the-art methods.

EDICS: RST-DNOI, FLT-LFLT
Keywords: Weighted Denoising, Sparse, Overcomplete, Translation Invariant Denoising, Cycle Spinning

The author would like to thank the three anonymous reviewers and the Associate Editor for their comments and guidance which have significantly improved this paper. Part of this work was done when the author was with Epson Palo Alto Laboratory. Part of this work was supported by NSF (CAREER award IIS-0093179).

LIST OF FIGURES

1 Noise corrupted piecewise constant signal and an overcomplete system of four block DCTs tiling it. The sample at n_e is close to the singularity and the sample at n_s is away from it. One DCT block from each transform overlaps each sample resulting in blocks $b_{n_e,1}, ..., b_{n_e,4}$ overlapping n_e and blocks $b_{n_s,1}, ..., b_{n_s,4}$ overlapping n_s. 7

2 Equivalent filters for denoising at n_s. The final denoised signal can be obtained by filtering the noisy signal y with per-sample adaptive filters. For the sample at n_s, the standard solution corresponds to obtaining a scalar product of y with the filter in (a) (shown superimposed on n_s) whereas the optimal combination corresponds to a scalar product with the filter in (b). 7

3 Equivalent filters for the derived three solutions and for the standard solution at pixel n_e using the ideal denoising rule (Table I (a)). The filters are displayed so that the tap at 0 multiplies the pixel value $y(n_e)$, the tap at -1 multiplies the pixel value $y(n_e - 1)$, etc. With the aid of Figure 1, observe that taps at $-1, 0, 1, \ldots, 3$ multiply values on the same side of the step edge as the pixel at n_e. Since it averages more, the full solution accomplishes the best filtering which results in the most noise suppression. This is followed by the identical diagonal and significant-only solutions. The standard solution accomplishes the worst filtering as it gives too much influence to $y(n_e)$ and thereby averages less over the pixels that are on the same side of the singularity. 13

4 Toy images. (a) Graphics (512×512), (b) Criss-Cross (512×512), (c) Voronoi (256×256), (d) Teapot(512×512). 13

5 PSNR v.s. threshold denoising results on the simple images Graphics ($\sigma_w = 5$), Criss-Cross ($\sigma_w = 5$), Voronoi ($\sigma_w = 5$), and Teapot ($\sigma_w = 10$). On these images with sharp singularities, the derived solutions significantly outperform the standard solution and BLS-GSM. 14

6 Portions from Graphics (a) and Criss-Cross (b). The singularities in Graphics are straightforward edges. Criss-Cross manifests more sophisticated singularities 15

7 Visual quality after denoising on the image Voronoi. Standard, full, diagonal, and significant-only solutions are at their respective peak PSNR thresholds. The standard solution has residual noise around the edges and BLS-GSM has residual noise and ringing. The proposed solutions alleviate these problems significantly. 16

8 Visual quality after denoising on the image Teapot. Standard, and significant-only solutions are at their respective peak PSNR thresholds. The standard solution and BLS-GSM have ringing. The significant-only solution alleviates this problem significantly. 16

9 Coefficient denoising rule of Equation (42). When the $(c_i(k), a_i(k))$ pair fall in the shaded region $\hat{c}_i(k)$ is set to zero. Otherwise $\hat{c}_i(k) = c_i(k)$. (Only the $c_i(k) \geq 0$ case is shown since the denoising rule is radially symmetric.) 18

10 Test images Lena, Barbara, Peppers, Cameraman, Boat, House, Photo1, Photo2. Cameraman and House are 256×256. All other images (including peppers) are 512×512. 19

11 Visual quality after denoising on the image (a) Cameraman ($\sigma_w = 10$), (b) Boat ($\sigma_w = 10$), (c) House ($\sigma_w = 15$), (d) Photo1 ($\sigma_w = 10$), (e) Photo2 ($\sigma_w = 10$). 24

LIST OF TABLES

I Denoising the example signal in Figure 1. Solutions at n_e and n_s, assuming the ideal denoising rule. 11

II Denoising the simple images using the oracle-based denoising rule. 17

III Denoising results on test images. Rows corresponding to "Signif.-only" illustrate the performance of the proposed solution. Each result is an average of eight runs with different noise realizations. BLS-GSM results are either from [28] (when available) or obtained using the authors' software. The standard deviation of the results as a function of the tested σ_w are $[0.017, 0.022, 0.022, 0.027, 0.028, 0.033]$ for the standard solution and $[0.016, 0.022, 0.023, 0.027, 0.030, 0.032]$ for the "Signif.-only" solution. . . 20

I. INTRODUCTION

The signal in additive i.i.d. noise problem continues to receive significant attention as it provides a benchmark for the accurate statistical modeling and representation of signals. In this paper we are interested in the popular scenario where noise in an image is removed with the aid of a set of linear transforms and thresholding nonlinearities[1]. Under sparse decomposition assumptions, it can be seen that additive noise leads to very low SNR on the many transform coefficients with small magnitudes. Hence a thresholding nonlinearity that simply detects and removes these coefficients results in robust and high performance denoising [9]. Such simple nonlinear shrinkage techniques can also be generalized with the aid of sophisticated regularization constructs to arrive at elaborate regression models and associated thresholding solutions [8], [14], [6]. For further details on the theory behind the shrinkage operation the reader is referred to [11].

While significant denoising improvements can be shown for classes of signals over which the utilized linear transforms provide sparse decompositions, it is well-known that typical transforms such as wavelets or block DCTs have limited sparsity over images with singularities along smooth curves [7]. Hence, after the initial set of results in [9], research in transform based denoising has so far concentrated on obtaining transforms with better mathematical properties on singularities (see for e.g., [21], [31], [27], [13], [12]), better thresholding techniques that account for spatial and scale variations over singularities (see for e.g., [25], [2], [22]), and better statistical modeling in transform domain that addresses the dependencies among coefficients over singularities (see for e.g., [5], [24], [29], [28], [23]).

The significant understanding and value provided by new transform design and better coefficient modeling notwithstanding, it is clear that this line of research must overcome significant obstacles. Newly designed transforms must surmount the practical challenges posed by real-world digital images and coefficient interdependency rules must be able to account for the many different types of singularities depicted on even the simplest images. Formulation of robust statistical models that govern groups of coefficients faces similar issues and is particularly difficult for transforms such as wavelet/complex-wavelet packets, DCTs, etc. With such transforms, statistical dependencies must be learned or modeled over large dimensional spaces since each singularity affects many transform coefficients. While these transforms do not have good approximation properties on singularities, they are nevertheless very valuable over image regions showing texture and similar high frequency variations ([15], [26]). A high performance approach that is broadly applicable to such transforms is very desirable.

An interesting alternative that can also be complementary to transform design and statistical dependency learning research is via the use of overcomplete decompositions established with the aid of an overcomplete set of transforms [4]. Denoising with an overcomplete set of transforms allows one to retain the simplicity of the

$$transform \rightarrow threshold \rightarrow inverse\ transform \tag{1}$$

denoising recipe while producing high performance results even with mathematically suboptimal transforms especially when the transform set is chosen to accomplish translation invariance [4]. In the overcomplete setting several denoised estimates (typically produced by applying (1) with shifted versions of the same transform, i.e., by "cycle

[1]There are of course many other approaches possible, for example, see [1] for a recent denoising algorithm not directly based on transforms.

spinning") are combined by *averaging* them at every pixel. It is hoped that some of the estimates will provide better performance than others, which will then be rectified in the combination via an average. While the main signal processing tools (choice of transforms, thresholds, etc.) that are used in denoising with an overcomplete set of transforms have been examined extensively, the final combination operation has not received much attention. In this paper, we examine this little explored but important aspect, namely the management of how the denoised estimates should be optimally combined at every pixel.

The method we present uses a set of denoised estimates provided by using (1) with an overcomplete set of transforms. However, the bulk of the work leaves the transform and thresholding aspects aside and concentrates solely on the final combination of the denoised estimates. In forming this combination, rather than obtaining an ad hoc average, the simple method we propose adaptively determines the better denoised estimates and prefers them on a per-pixel basis. As we show using oracle based denoisers, we build on the existing body of literature and exploit a degree of freedom that is in general not reachable by using previous approaches with better-tuned thresholds. Most importantly, our formulation does not lose the simplicity of averaging (and of the recipe in Equation (1)), but it still manages to provide significant improvements.

One of the interesting things to see in this paper is that our techniques enable even "singularity-blind" transforms like DCTs to obtain high performance over image singularities with virtually no sophisticated statistical modeling. This becomes possible because close to singularities at least some denoised estimates tend to provide good results even with transforms like DCTs. Hence achieving high performance simply becomes a matter of adaptively finding these estimates and emphasizing them. While our approach does not directly model singularities, it implicitly obtains very high performance on them by aggressively determining the regions around them where at least some of the utilized transforms provide sparse decompositions. Our combination step simply gives preference to the corresponding denoised estimates in order to obtain substantial improvements over direct averaging. As we show, with the techniques of this paper, singularity-blind DCTs readily outperform elaborate directional basis over and near singularities.

Given the denoised estimates provided by the transforms in the overcomplete set, our task in this paper is to form an overall estimate of the denoised image[2]. In this quest, we derive three estimators in order of diminishing computational complexity. Our first estimator is also the most potent and improves denoising performance throughout, even when no singularities are present in the signal. Our second estimator is very simple and solely requires the determination of the amount of expected noise energy remaining in the pixels of each denoised estimate. In the case of hard-thresholding, this per-pixel residual noise energy can be determined using the coefficients from each transform that survive thresholding with the aid of the associated basis functions. For our third and simplest estimate, we further approximate this energy by the number of non-zero coefficients remaining (associated with basis functions overlapping the given pixel) subsequent to denoising with each transform.

In related work, [16] considers hard-thresholding operations as projections to signal-subspaces and tries to find the signal that lies in the intersection of all signal-subspaces as determined by an overcomplete set of transforms and hard-thresholding. The resulting technique is very interesting but requires accurate thresholding (subspace

[2]Our early work appeared in [20], [18].

determination) since incorrect subspace determination can easily lead one to shrink to the all-zero signal. Beyond being more computationally complex and less robust to thresholding compared to our work, their approach is also difficult to generalize to other types of thresholding and denoising rules.

The paper is organized as follows. Section II introduces the basic ideas of this work. After introducing notation in Section II-A, we discuss intuitions behind the work in Section II-B and provide the main derivation in Section II-C. The three solutions and a brief solution comparison example comprise Section II-D. Section II-E expands further on this comparison by introducing the equivalent, per-pixel adaptive denoising filters. In Section III we deploy the three solutions on a set of simple images with sharp singularities in order to illustrate the solutions' properties and also look at results with oracle-based denoisers. The observed properties are used in Section IV to derive an image denoising algorithm whose performance on test images is illustrated in Section V. Section VI concludes the paper.

II. Main Ideas and Derivation

A. Definitions and Notation

Let $x\ (N \times 1)$ denote the original signal, and let the noise corrupted signal be given by

$$y = x + w, \tag{2}$$

where w is i.i.d. noise of variance σ_w^2. Let $\mathbf{H}_i\ (N \times N),\ i = 1, \ldots, M$ denote a set of invertible linear transforms. The transform coefficients of each transform are thresholded to obtain a denoised estimate $\hat{x}_i$ of x. These estimates are then combined to arrive at an overall estimate $\hat{x}$.

Denoised Estimates: The denoised estimates $\hat{x}_i$ corresponding to each transform can be obtained in three stages for $i = 1, \ldots, M$ as follows.

$$\begin{aligned} c_i &= \mathbf{H}_i y, \\ \hat{c}_i &= \mathcal{T}(c_i), \\ \hat{x}_i &= \mathbf{H}_i^{-1}(\hat{c}_i), \end{aligned} \tag{3}$$

where $c_i\ (N \times 1)$ are the transform coefficients due to $\mathbf{H}_i$, $\hat{c}_i\ (N \times 1)$ are the thresholded coefficients, $\mathcal{T}(.)$ is the thresholding function, and $\hat{x}_i$ is the i^{th} denoised estimate of x.

Thresholding: As will become clear, our work can accommodate a large variety of thresholding techniques. For convenience, the examples in this paper will utilize hard-thresholding on each coefficient via

$$\hat{c}_i(k) = \begin{cases} 0, & |c_i(k)| \leq \tau_{i,k} \\ c_i(k), & |c_i(k)| > \tau_{i,k} \end{cases} \tag{4}$$

for $k = 1, \ldots, N$, where $c_i(k)$ is the k^{th} transform coefficient of the i^{th} transform, and $\tau_{i,k}$ is a given threshold for the coefficient $c_i(k)$. Except for the simple examples in Section III, where will use spatially uniform thresholds (i.e., $\tau_{i,k} = \tau$), the manuscript will allow per coefficient variation in the thresholds.

Ideal Denoising Rule: Consider the original signal's transform coefficients given by

$$d_i \quad = \quad \mathbf{H}_i x, \quad i = 1, \ldots, M. \tag{5}$$

For convenience in discussing the toy examples and the simple formulas for the weights, we define the infeasible rule

$$\hat{c}_i(k) = \begin{cases} 0, & d_i(k) = 0 \\ c_i(k), & |d_i(k)| > 0 \end{cases} \tag{6}$$

which can be written in terms of (4) by picking $\tau_{i,k} = |c_i(k)|$ when $d_i(k) = 0$ and $\tau_{i,k} = 0$ otherwise. Of course this rule only provides denoising improvements for ideal signals that provide many zero-valued coefficients with the given transforms.

Oracle-Based Denoising Rule: In order to show that weighting improves performance under more elaborate thresholding rules and to motivate the "coefficient denoising rule" that will be used in our actual denoising algorithm in Section IV, we define the infeasible rule

$$\hat{c}_i(k) = \begin{cases} 0, & |d_i(k)| \leq |c_i(k) - d_i(k)| \\ c_i(k), & \text{otherwise,} \end{cases} \tag{7}$$

i.e., if the distortion caused by keeping the noisy coefficient unchanged is less than the distortion caused by setting it to zero, the coefficient is left as is; otherwise, the coefficient is set to zero. Note that the oracle can always find a per-coefficient adaptive threshold (as a function of $c_i(k)$ and $d_i(k)$) to carry out this rule using the notation of (4).

Index Sets and Selection Matrices: To help later notation we also define the index sets

$$\mathcal{V}_i = \{k| \ |c_i(k)| > \tau_{i,k}\} \tag{8}$$

and the related diagonal selection matrices

$$\mathbf{S}_i(k,l) = \begin{cases} 1 & k = l, |c_i(k)| > \tau_{i,k} \\ 0 & \text{otherwise.} \end{cases} \tag{9}$$

Observe that with this notation, Equation (3) becomes

$$\hat{x}_i = \mathbf{H}_i^{-1} \mathbf{S}_i \mathbf{H}_i y, \tag{10}$$

and by assuming orthonormal transforms it further simplifies to

$$\hat{x}_i = \mathbf{H}_i^T \mathbf{S}_i \mathbf{H}_i y, \tag{11}$$

where $(.)^T$ denotes transpose.

Residual Noise: By combining (2) and (11) we have $\hat{x}_i = \mathbf{H}_i^T \mathbf{S}_i \mathbf{H}_i x + \mathbf{H}_i^T \mathbf{S}_i \mathbf{H}_i w$, and the noise remaining after denoising with the i^{th} transform can be obtained as

$$\begin{aligned} w_i \quad &= \quad \hat{x}_i - \mathbf{H}_i^T \mathbf{S}_i \mathbf{H}_i x \\ &= \quad \mathbf{H}_i^T \mathbf{S}_i \mathbf{H}_i w. \end{aligned} \tag{12}$$

Combination: In the traditional simple average-based combination the denoised estimates are combined at every pixel to arrive at the overall denoised signal via

$$\hat{x}(n) = 1/M \sum_{i=1}^{M} \hat{x}_i(n),\ n = 1, \ldots, N. \tag{13}$$

In this paper, we would like to replace the above equation with

$$\hat{x}(n) = \sum_{i=1}^{M} \gamma_i(n) \hat{x}_i(n),\ n = 1, \ldots, N, \tag{14}$$

where the weights $\gamma_i(n),\ i = 1, \ldots, M,\ n = 1, \ldots, N$, are to be determined optimally in Section II-C[3]. An interesting viewpoint can be arrived at by noting that when one uses (10) to substitute for $\hat{x}_i$, Equation (14) corresponds to doing a pixel adaptive inverse transform. For denoising with naturally overcomplete transforms such as complex wavelets [21], where the inverse transform can be written in terms of a bank of real $\mathbf{H_i}$, $i = 1, \ldots, M$, the proposed work can be straightforwardly inserted in the inversion stage to again allow a pixel adaptive inverse for better denoising performance.

Sparse Regions: Throughout, the manuscript will refer to regions in the image where the utilized localized transforms provide sparse decompositions as sparse regions. These are regions where localized transform coefficients decay rapidly in magnitude sorted order. The more rapid the decay the better the expected approximation-denoising-compression performance, with the transform and the rate of decay to quantifying the families of images over which this expected performance will be valid [9], [10], [3]. It is important to note that with DCT-like transforms sparse regions can be much more general than locally smooth regions, i.e., non-smooth, textured patches in images can also be sparse with respect to the utilized transform. For example, both Figure 4 (a) and (b) have large sparse regions with respect to the utilized block DCT basis. As will become clear, the proposed work allows one to obtain high performance on both images since it exploits the more general notion of "localized sparsity" rather than the more restricted notion of localized smoothness.

B. Intuition

As an intuitive example consider a one dimensional piecewise constant signal, where an ideal step edge separates the two piecewise constant portions. Assume this signal has been corrupted with additive, i.i.d. noise to result in the signal depicted at the bottom of Figure 1. For simplicity, assume an overcomplete system of four 4×1 block DCTs ($DCT1, \ldots, DCT4$) are evaluated on this signal corresponding to four translations of the block tiling. Suppose, the coefficients of each transform are thresholded and inverse transformed to yield four denoised estimates $\hat{x}_i(n),\ i = 1, \ldots, 4$ at each sample n.

Observe that the sample identified at n_e has four block DCTs overlapping it, one block from each transform, labeled $b_{n_e,1}, \ldots, b_{n_e,4}$. Of these four blocks the first two overlap the "edge" in the figure (shown dark shaded)

[3]Reviewer 2 has pointed out potential connections to the data fusion literature [33], [34]. The reader should note though that compared to cases considered in general fusion setups, all of the "fusing" sources in (14) are simple versions of the same signal, i.e., the problem we are trying to solve is much simpler compared to the general fusion problem which will allow us to obtain high performance solutions with little apriori information.

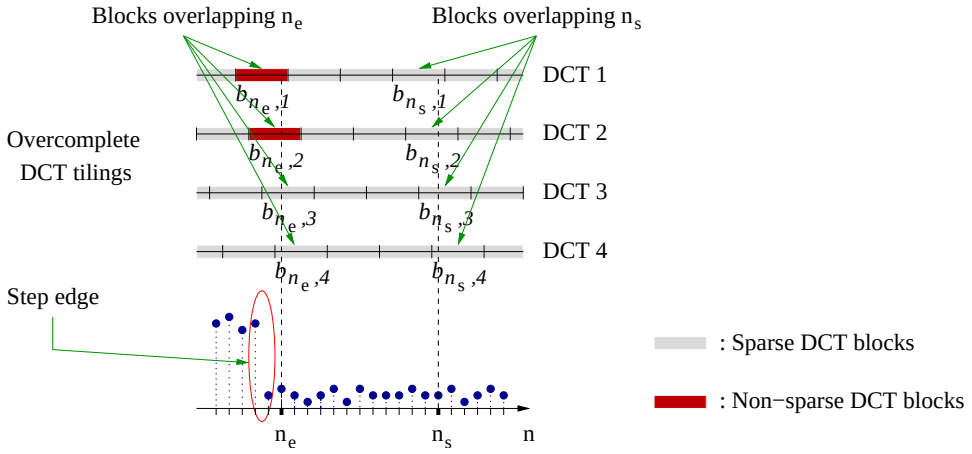
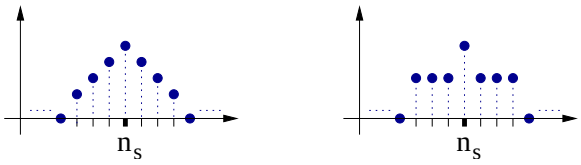


Fig. 1. Noise corrupted piecewise constant signal and an overcomplete system of four block DCTs tiling it. The sample at n_e is close to the singularity and the sample at n_s is away from it. One DCT block from each transform overlaps each sample resulting in blocks $b_{n_e,1}, ..., b_{n_e,4}$ overlapping n_e and blocks $b_{n_s,1}, ..., b_{n_s,4}$ overlapping n_s.

whereas the latter two overlap smooth portions of the signal (shown lightly shaded). Using well-known approximation arguments [10], [3], we expect the two block DCTs overlapping the smooth regions to provide sparse decompositions and yield good individual denoised estimates for the sample at n_e. The two block DCTs overlapping the edge on the other hand will not give rise to good denoised estimates at n_e. Hence, rather than setting $\gamma_i(n_e) = 1/4$ for $i = 1, \ldots, 4$ as in Equation (13), one can see that it is beneficial to give more weight to $i = 3, 4$ and less to $i = 1, 2$ when the individual estimates are combined.

For the sample at n_s in Figure 1, whether there should be a difference in the weights is not clear. However, consider the example case where the thresholding operation retains *only* the DCT-DC terms in the four overlapping blocks, i.e., except for the DC coefficient, all coefficients in blocks $b_{n_s,1}, ..., b_{n_s,4}$ are thresholded to zero. For this case, it can be seen that while Equation (13) forms the equivalent denoising filter shown in Figure 2 (a) (see Section II-E for the general derivation), the optimal filter results by choosing $\gamma_1(n_s) = \gamma_4(n_s) = 1/2,\ \gamma_2(n_s) = \gamma_3(n_s) = 0$ as shown in Figure 2 (b). Thus it is clear that the optimal assignment of the $\gamma_i(n)$ in Equation (14) has the potential to benefit samples away from singularities as well.



(a) Standard (b) Optimal

Fig. 2. Equivalent filters for denoising at n_s. The final denoised signal can be obtained by filtering the noisy signal y with per-sample adaptive filters. For the sample at n_s, the standard solution corresponds to obtaining a scalar product of y with the filter in (a) (shown superimposed on n_s) whereas the optimal combination corresponds to a scalar product with the filter in (b).

C. Main Derivation

For convenience assume zero mean quantities. We would like to derive the optimal weights in Equation (14) such that they minimize the conditional mean squared error, given the noisy signal y and the denoised estimates $\hat{x}_1, \ldots, \hat{x}_i, \ldots, \hat{x}_M$,

$$E[|x(n) - \hat{x}(n)|^2 | y, \hat{x}_1, \ldots, \hat{x}_i, \ldots, \hat{x}_M]. \tag{15}$$

Assuming the denoisers generating the individual estimates do not have side information beyond y, it is clear that the above quantity is equivalent to $E[|x(n) - \hat{x}(n)|^2 | y]$. Furthermore, if we are allowed to adapt the weights $\gamma_i(n)$ for each y in an unconstrained way it is clear that, unless all of the $\hat{x}_i(n)$ are zero, the optimal combination $\hat{x}^*(n) = \sum_{i=1}^{M} \gamma_i^*(n)\hat{x}_i(n)$ should be equal to $E[x(n)|y]$, i.e., to the global minimizer of Equation (15) [32] (simply pick an i for which $\hat{x}_i(n)$ is non-zero, set the weight $\gamma_i^*(n) = E[x(n)|y]/\hat{x}_i(n)$, and set the remaining weights to zero). Hence, our per-pixel setup neither loses nor gains anything compared to the most general formulation of the signal in noise problem. Some assumptions are needed to convert this most ambitious formulation into a tractable setting. The assumptions below are motivated by tractability concerns especially given the nonlinearities used in the individual denoisers.

i- For the results in this paper, we will assume that the overall estimate will only be conditioned on the index sets of Equation (8), i.e, on the available sparsity information as established by the denoisers. Our aim then is to minimize

$$E[|x(n) - \hat{x}(n)|^2 | \mathcal{V}_1, \ldots, \mathcal{V}_M] \tag{16}$$

rather than (15). In the rest of this section we let $\mathcal{E}[.]$ denote the conditional expectation $E[.|\mathcal{V}_1, \ldots, \mathcal{V}_M]$.

ii- We will assume that

$$\sum_{i=1}^{M} \gamma_i(n) = 1. \tag{17}$$

In conjunction with the simplifications proposed in Section II-D, this assumption will allow us to side-step issues related to the determination of the particulars of signal statistics[4].

Given these assumptions our goal becomes the minimization of

$$\mathcal{E}[|x(n) - \sum_{i=1}^{M} \gamma_i(n)\hat{x}_i(n)|^2] \tag{18}$$

in terms of $\gamma_i(n),\ i = 1, \ldots, M$, subject to $\sum_{i=1}^{M} \gamma_i(n) = 1$. Let $e_i = x - \hat{x}_i$ and define

$$\Gamma_n = [\gamma_1(n)\ \ldots\ \gamma_M(n)]^T, \tag{19}$$

$$\chi_n = [\hat{x}_1(n)\ \ldots\ \hat{x}_M(n)]^T, \tag{20}$$

$$\epsilon_n = [e_1(n)\ \ldots\ e_M(n)]^T, \tag{21}$$

$$u = [1 \ldots 1]^T, \tag{22}$$

$$\boldsymbol{\eta}_n = \mathcal{E}[\epsilon_n \epsilon_n^T]. \tag{23}$$

[4]Under the ideal denoising rule of (6), one can manipulate the relevant equations to see that this assumption becomes equivalent to a high signal to noise ratio assumption. Combining (11) with (6) leads to $x_i = x + w_i$ with x and w_i being independent under (6). The resulting optimization of (16) in terms of $\gamma_i(n)$ assuming high signal to noise ratio results in $\sum_{i=1}^{M} \gamma_i(n) \approx 1$.

The minimization of Equation (18) becomes the minimization of

$$\begin{aligned} & \mathcal{E}[||x(n) - \Gamma_n^T \chi_n||^2] - \lambda_n \Gamma_n^T u \\ = \; & \mathcal{E}[||x(n)(1 - \Gamma_n^T u) - \Gamma_n^T \epsilon_n||^2] - \lambda_n \Gamma_n^T u, \end{aligned} \tag{24}$$

where λ_n is chosen to enforce $\Gamma_n^T u = 1$. We thus arrive at

$$\boldsymbol{\eta}_n \Gamma_n = \lambda_n' u \tag{25}$$

where λ_n' ensures $\Gamma_n^T u = 1$.

D. Practical Solutions

It is clear that the exact determination of $\boldsymbol{\eta}_n = \mathcal{E}[\epsilon_n \epsilon_n^T]$ in order to enable the solution of (25) is difficult given the nonlinearities used in the individual denoisers. In order to obtain simple expressions for the weights we will assume that the errors remaining after denoising with each transform are mostly due to residual noise, i.e., $\mathcal{E}[e_i(n)e_j(n)] \approx \mathcal{E}[w_i(n)w_j(n)]$. Define the vector δ_n $(N \times 1)$ as

$$\delta_n(i) = \begin{cases} 1, & i = n \\ 0, & otherwise. \end{cases} \tag{26}$$

We have

$$\begin{aligned} \boldsymbol{\eta}_n(i,j) &= \mathcal{E}[e_i(n)e_j(n)] && (27) \\ &\approx \mathcal{E}[w_i(n)w_j(n)] && (28) \\ &= \delta_n^T \mathcal{E}[w_i w_j] \delta_n. && (29) \end{aligned}$$

Substituting (12) and approximating $\mathcal{E}[ww^T] \approx E[ww^T]$ we obtain[5]

$$\begin{aligned} \boldsymbol{\eta}_n(i,j) &\approx \sigma_w^2 \delta_n^T \mathbf{H}_i^T \mathbf{S}_i \mathbf{H}_i \mathbf{H}_j^T \mathbf{S}_j \mathbf{H}_j \delta_n && (30) \\ &= \sigma_w^2 \mathbf{G}_n(i,j), && (31) \end{aligned}$$

where σ_w^2 is the noise variance and $\mathbf{G}_n$ is a matrix that is solely dependent on the overcomplete transform set and the index sets $\mathcal{V}_i,\ i = 1, \ldots, M$.

With the aid of (31), it can be seen that Equation (25) reduces to

$$\mathbf{G}_n \Gamma_n = \lambda_n^* u, \tag{32}$$

where λ_n^* is the scaling factor that achieves $\Gamma_n^T u = 1$. In this paper we consider the following solutions to (32).

Full Solution: Since $\mathbf{G}_n$ is positive semidefinite, Equation (32) can be solved within the positive eigenspace of $\mathbf{G}_n$ to yield Γ_n whenever the rank of $\mathbf{G}_n$ is non-zero. (For the zero rank case we assume that Γ_n is set to its standard value of $[1/M \ldots 1/M]^T$.) We will refer to this solution of Equation (32) as the full solution.

[5]Note that using the ideal denoising rule of (6), both approximations become exact and (30) becomes an equality.

Diagonal Solution: The full solution is complex as it requires the formation and the inversion of $\mathbf{G}_n$ at every pixel. Looking at the form of $\mathbf{G}_n$ through Equations (30) and (31), we observe that the matrix element $\mathbf{G}_n(i,j)$ represents the cross correlation of the noise remaining at pixel n after denoising with the i^{th} and j^{th} transforms divided by σ_w^2. Assuming these cross terms are negligible in comparison to the correlations at diagonals and thereby setting the off-diagonal elements of $\mathbf{G}_n$ to zero yields a diagonal matrix $\mathbf{D}_n$ where

$$\mathbf{D}_n(i,j) = \begin{cases} \mathbf{G}_n(i,i) = \delta_n^T \mathbf{H}_i^T \mathbf{S}_i \mathbf{H}_i \delta_n, & \text{if } i = j \\ 0, & \text{otherwise.} \end{cases} \tag{33}$$

We will refer to the solution of (32) with $\mathbf{G}_n$ replaced by $\mathbf{D}_n$ via

$$\mathbf{D}_n \Gamma_n = \lambda_n^* u \tag{34}$$

as the diagonal solution.

Significant-Only Solution: While the diagonal solution is much simpler than the full solution, it still requires us to keep track of the square of the value of each transform basis function at every pixel since

$$\delta_n^T \mathbf{H}_i^T \mathbf{S}_i \mathbf{H}_i \delta_n = \sum_{k \in \mathcal{V}_i} (\mathbf{H}_i(k,n))^2. \tag{35}$$

If we assume that the transform basis functions have values of constant magnitude in their spatial support we can obtain an even simpler equation. Consider the m^{th} basis function of the i^{th} transform, i.e., the m^{th} row of $\mathbf{H}_i$ given as $\mathbf{H}_i(m,n)$, $n = 1, \ldots, N$. Let $\alpha_i(m)$ determine the number of non-zero entries in $\mathbf{H}_i(m,.)$. Fix $\varepsilon \geq 0$. For the given m and for $n = 1, \ldots, N$ define

$$\mathbf{F}_i(m,n) = \begin{cases} 1/\sqrt{\alpha_i(m)}, & \text{if } \mathbf{H}_i(m,n) > \varepsilon \\ -1/\sqrt{\alpha_i(m)}, & \text{if } \mathbf{H}_i(m,n) < \varepsilon \\ 0, & \text{otherwise.} \end{cases} \tag{36}$$

(For the results in this paper $\varepsilon = 0$.) We obtain what we refer to as the *significant-only* solution by replacing the diagonal matrix $\mathbf{D}(n)$ in Equation (34) with

$$\tilde{\mathbf{D}}_n(i,j) = \begin{cases} \delta_n^T \mathbf{F}_i^T \mathbf{S}_i \mathbf{F}_i \delta_n, & \text{if } i = j \\ 0, & \text{otherwise.} \end{cases} \tag{37}$$

Block Transform Simplification of the Significant-Only Solution: One of the benefits of the significant-only solution is that it leads to a particularly simple form when the utilized transforms are block. For block transforms, the transform basis functions have block support. Assuming all the transforms' basis functions have the same block size, the $\alpha_i(m)$ in Equation (36) can be made independent of i and m, i.e., $\alpha_i(m) = \alpha$. Since transforms are block, there is one block from each transform that overlaps any pixel. Suppose pixel n is in block $b_{n,i}$ of transform i. After denoising with transform i suppose this block has $\mathcal{N}(b_{n,i})$ non-zero transform coefficients. Then we have

$$\begin{aligned} \delta_n^T \mathbf{F}_i^T \mathbf{S}_i \mathbf{F}_i \delta_n &= \sum_{k \in \mathcal{V}_i} (\mathbf{F}_i(k,n))^2 \\ &= \sum_{k \in \{\textit{Non-zero coefficients in } b_{n,i}\}} 1/\alpha \\ &= \frac{\mathcal{N}(b_{n,i})}{\alpha}. \end{aligned} \tag{38}$$

As a result, at pixel n the weight for the denoised estimate provided by the i^{th} transform is

$$\gamma_i(n) = \frac{C_n}{\mathcal{N}(b_{n,i})}, \tag{39}$$

where C_n is a constant independent of i chosen to ensure $\sum_i \gamma_i(n) = 1$. Hence, for the block transform case, the significant-only solution results in intuitively simple weights that are only dependent on the sparsity of the transform blocks in the denoised estimates. Observe that the transform blocks with many zero coefficients subsequent to denoising receive larger weights in the combination. An example algorithm that implements the significant-only solution with block transforms and hard-thresholding is sketched in Appendix I.

The following example illustrates the derived solutions for the toy signal in Figure 1.

Solution Comparison:

Example 1: Consider the noisy piecewise constant signal and the overcomplete transform set depicted in Figure 1. In order to compare the three solutions assume that the denoised estimates are obtained through the ideal denoising rule of (6) so that the coefficients of blocks overlapping the edge have not been thresholded, and the coefficients of all the other blocks have been thresholded such that only the DCT-DC coefficient is retained. For the pixel at n_e we have

$$\mathcal{N}(b_{n_e,1}) = 4,\ \mathcal{N}(b_{n_e,2}) = 4,\ \mathcal{N}(b_{n_e,3}) = 1,\ \mathcal{N}(b_{n_e,4}) = 1,$$

and for n_s

$$\mathcal{N}(b_{n_s,1}) = 1,\ \mathcal{N}(b_{n_s,2}) = 1,\ \mathcal{N}(b_{n_s,3}) = 1,\ \mathcal{N}(b_{n_s,4}) = 1.$$

Solving for the weights at n_e and n_s we obtain the results in Table I. (The mean squared error (mse) column in the table assumes that the same denoising rule has been carried out for many realizations, i.e., that we fix the index sets and obtain the mse over many realizations.)

Solution	$\gamma_1(n_e)$	$\gamma_2(n_e)$	$\gamma_3(n_e)$	$\gamma_4(n_e)$	$mse(n_e)$
Full	-0.02	-0.02	0.52	0.52	$0.22\sigma_w^2$
Diagonal	0.10	0.10	0.40	0.40	$0.26\sigma_w^2$
Significant-only	0.10	0.10	0.40	0.40	$0.26\sigma_w^2$
Standard	0.25	0.25	0.25	0.25	$0.43\sigma_w^2$

(a)

Solution	$\gamma_1(n_s)$	$\gamma_2(n_s)$	$\gamma_3(n_s)$	$\gamma_4(n_s)$	$mse(n_s)$
Full	0.50	0.00	0.00	0.50	$0.16\sigma_w^2$
Diagonal	0.25	0.25	0.25	0.25	$0.17\sigma_w^2$
Significant-only	0.25	0.25	0.25	0.25	$0.17\sigma_w^2$
Standard	0.25	0.25	0.25	0.25	$0.17\sigma_w^2$

(b)

TABLE I

DENOISING THE EXAMPLE SIGNAL IN FIGURE 1. SOLUTIONS AT n_e AND n_s, ASSUMING THE IDEAL DENOISING RULE.

As evidenced from Table I (a), for the pixel at n_e, all solutions significantly outperform the standard solution. The diagonal and significant-only solutions generate the same weights (since $\mathbf{D}_{n_e}$ in (33) and $\tilde{\mathbf{D}}_{n_e}$ in (37) become identical) with mse performance closely following that of the full solution. On the other hand, as indicated in Table I (b), for the pixel at n_s only the full solution deviates from the standard solution. This is because the $\mathcal{N}(b_{n_s,i})$ are identical and the surviving coefficients correspond to DC, i.e., all constant, basis functions. Note that the full solution is only marginally better than the standard solution since simple averaging of the denoised estimates works well in portions of the signal where all of the utilized transforms yield sparse decompositions. Hence, based on these results, we expect the weighted solutions to obtain significant improvements primarily around singularities.□

E. Equivalent Filters

Using Equations (11) and (14) we can construct the final denoised estimate as

$$\begin{aligned} \hat{x}(n) &= \sum_{i=1}^{M} \gamma_i(n)\hat{x}_i(n) \\ &= [\sum_{i=1}^{M} \gamma_i(n)\delta_n^T \mathbf{H}_i^T \mathbf{S}_i \mathbf{H}_i]\ y \\ &= L_n^T y, \end{aligned} \tag{40}$$

where $L_n = [\delta_n^T(\sum_{i=1}^{M} \gamma_i(n)\mathbf{H}_i^T\mathbf{S}_i\mathbf{H}_i)]^T$. L_n can be thought of as an equivalent, pixel adaptive filter or projection that is applied to the noisy signal y to obtain the denoised output, i.e., the whole operation of overcomplete transforms, thresholding, and weighted combination can be thought to result in the per-pixel adaptive denoising filter L_n.

Example 1 (contd.) : Figure 3 shows the equivalent filters for the pixel at n_e using the ideal denoising rule, corresponding to the situation in Table I (a). Using the generic statistical model which dictates that pixels from different sides of a singularity are decorrelated [3], we expect the optimal equivalent filter at n_e to obtain a spatial average over the pixels that are on the same side of the singularity as n_e. It is easily seen from the figure that the derived solutions perform better spatial averaging around n_e compared to the standard solution and thereby remove more noise.

The equivalent filters for the pixel at n_s turn out to be as shown earlier in Figure 2, with the full solution accomplishing the optimal filter in Figure 2 (b) and the other solutions accomplishing the standard filter in Figure 2 (a). □

III. Simple Examples and Properties

In this section we consider four toy images that illustrate the main properties of the derived solutions. These images have sharp singularities that better delineate the performance around such features and thus form good test cases of denoising around singularities. Singularity geometry is especially important on the images Graphics and Criss-Cross. The denoising results on these images are particularly impacted by performance on singularities.

Assume that the images shown in Figure 4 are corrupted with white Gaussian noise having standard deviation σ_w. The noisy images are denoised using an overcomplete set of block DCT (8×8) transforms so that the resulting

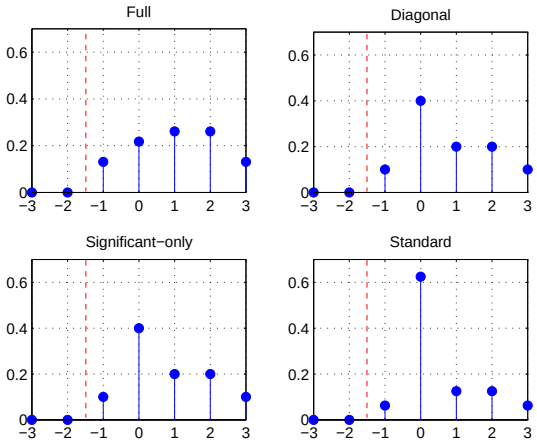
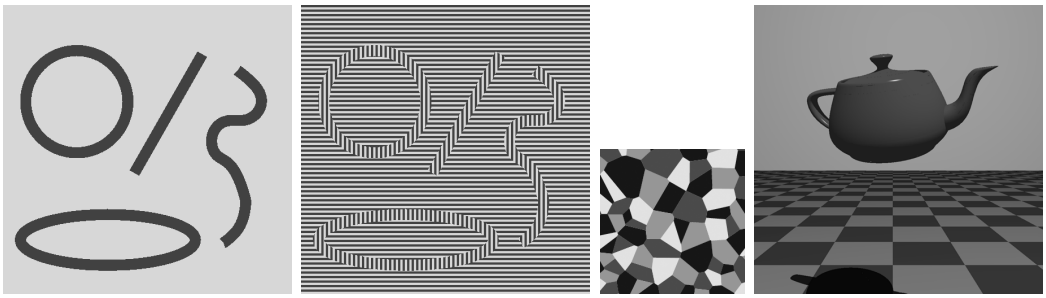


Fig. 3. Equivalent filters for the derived three solutions and for the standard solution at pixel n_e using the ideal denoising rule (Table I (a)). The filters are displayed so that the tap at 0 multiplies the pixel value $y(n_e)$, the tap at -1 multiplies the pixel value $y(n_e - 1)$, etc. With the aid of Figure 1, observe that taps at $-1, 0, 1, \ldots, 3$ multiply values on the same side of the step edge as the pixel at n_e. Since it averages more, the full solution accomplishes the best filtering which results in the most noise suppression. This is followed by the identical diagonal and significant-only solutions. The standard solution accomplishes the worst filtering as it gives too much influence to $y(n_e)$ and thereby averages less over the pixels that are on the same side of the singularity.

denoising is fully translation invariant, i.e., $\mathbf{H}_i$, $i = 1, \ldots, M$, represent all translations of a 8×8 block DCT with $M = 64$. Denoising is carried out by spatially uniform hard-thresholding with a single threshold τ using the recipe of Equation (3) (DCT-DC coefficients are not thresholded). In what follows, when comparing the three derived solutions over the standard system, we also compare them to the high performance BLS-GSM denoising method proposed in [28].



(a) (b) (c) (d)

Fig. 4. Toy images. (a) Graphics (512×512), (b) Criss-Cross (512×512), (c) Voronoi (256×256), (d) Teapot(512×512).

PSNR Performance: The PSNR v.s. threshold (τ) curves[6] obtained by denoising the four simple images are illustrated in Figure 5. It is clear that on this set of images with sharp singularities the three solutions of Section II-D (full, diagonal, and significant-only) substantially outperform the standard solution, which itself outperforms BLS-GSM. Observe that there are significant improvements in PSNR, which sometimes approach 10 dB improvements over BLS-GSM. These images are almost completely described by their singularity geometry as their uniform portions

[6]We defer issues related to threshold selection to Section V where we will present our full-fledged denoising algorithm.

can be denoised very effectively, practically with any technique, i.e., performance around singularities is the main bottleneck in improving overall PSNR. The PSNR improvements are thus a clear indication of the improvements the proposed work achieves close to singularities.

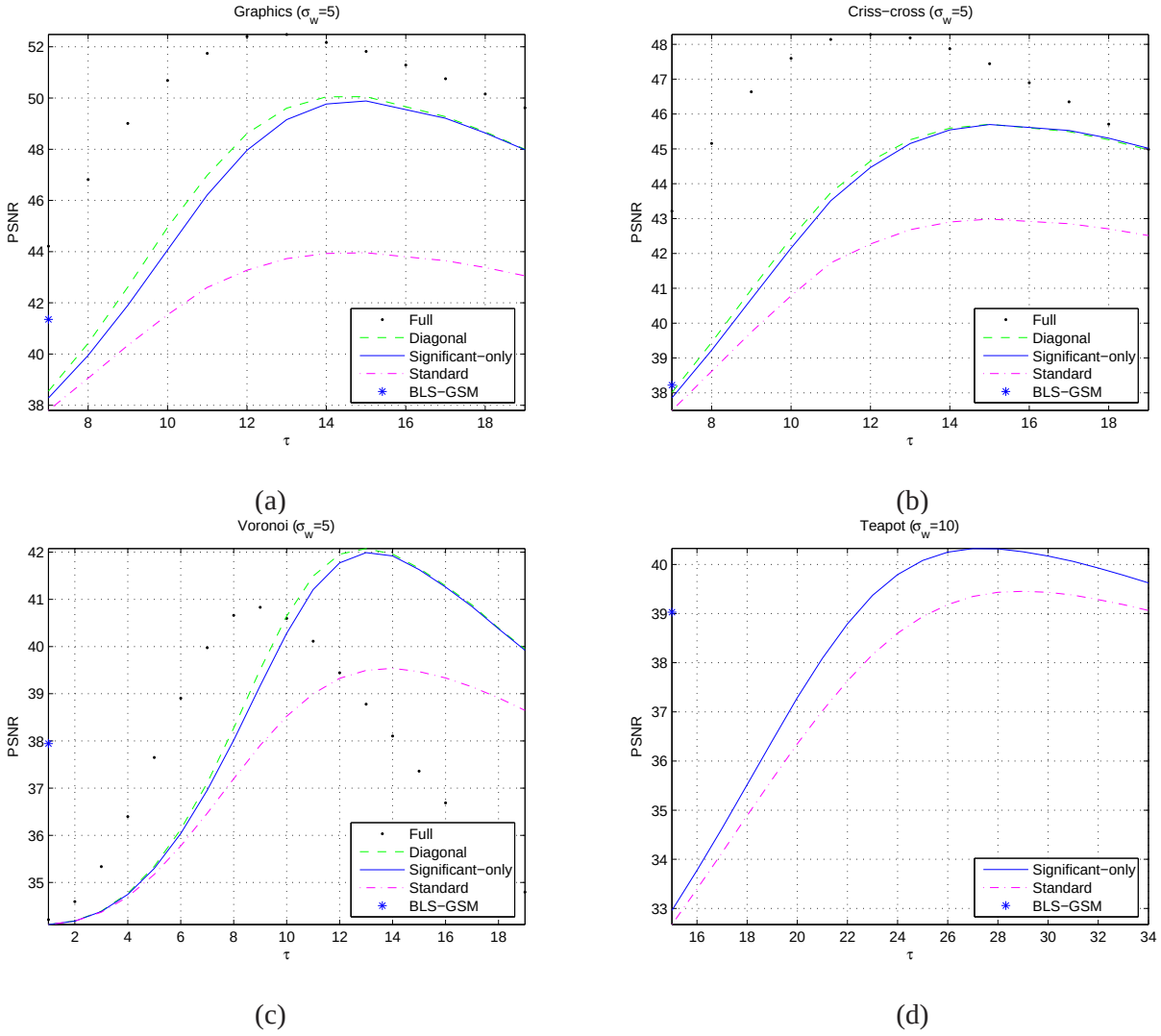
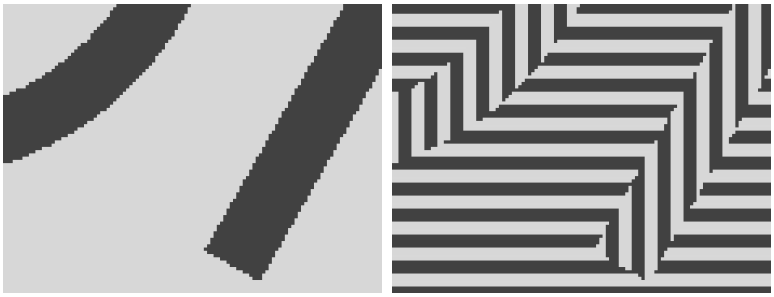


Fig. 5. PSNR v.s. threshold denoising results on the simple images Graphics ($\sigma_w = 5$), Criss-Cross ($\sigma_w = 5$), Voronoi ($\sigma_w = 5$), and Teapot ($\sigma_w = 10$). On these images with sharp singularities, the derived solutions significantly outperform the standard solution and BLS-GSM.

Complex singularities: It is important to note that the piecewise uniform images, Graphics and Criss-Cross, depicted in Figure 4 (a) and (b) have the same singularity geometry but exhibit very different behavior in their uniform portions. Due to this difference, the singularities in these two images also show remarkably different behaviour. As can be seen in the magnified portions depicted in Figure 6, while Graphics has straightforward edges for which the many singularity models devised in the literature apply, transform coefficients over the singularities of Criss-Cross are much more difficult to model. A technique dependent on straightforward edge models will likely not achieve full performance on Criss-Cross due to model failures. The derived solutions on the other hand are robust to the change in the nature of the singularities since they only depend on the sparsity around them. These observations are confirmed in Figure 5 (a) and (b), where the proposed solutions significantly outperform the standard solution and



(a) (b)

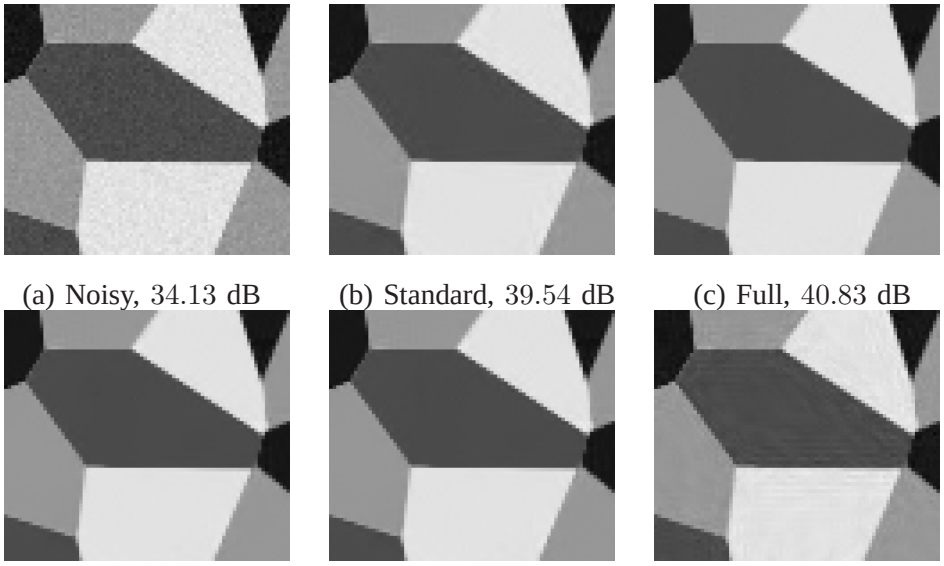
Fig. 6. Portions from Graphics (a) and Criss-Cross (b). The singularities in Graphics are straightforward edges. Criss-Cross manifests more sophisticated singularities

BLS-GSM[7]. Note also that while BLS-GSM accomplishes performance close to the standard solution in Graphics, it falls significantly behind in Criss-Cross.

Sensitivity to thresholding: As is evident from figures 5(a) through 5(c), the full solution obtains the peak performance at smaller thresholds compared to diagonal, significant-only, and standard solutions. This follows since the full solution is the one that is most tied to the assumption of Equation 30: As the threshold is raised, hard-thresholding zeros out progressively more significant portions of the signal which makes the assumptions leading to (30) progressively less valid. Deviations from the model are especially pronounced in the image Voronoi; for many pixels close to corner points where several edges intersect there are no denoised estimates for which (30) is justified. While still outperforming the standard solution, the full solution no longer enjoys performance advantages over diagonal and significant-only solutions on the image Voronoi. Since the full solution is also the most complicated solution due to the need for the formation and inversion of $\mathbf{G}_n$ of Equation (32) at every pixel, in what follows we will concentrate on the significant-only solution which obtains comparable performance to the diagonal solution while being the computationally simplest solution.

Performance around singularities: Recent developments in image processing have shunned classical transforms like DCTs for the sake of sophisticated directional transforms which have better mathematical properties around singularities along curves. Using Figures 7 and 8 observe that one can obtain very sophisticated results with the proposed techniques using block DCTs, which are well-known to be suboptimal around such singularities. Without the proposed solutions, small thresholds lead to substantial residual noise, whereas large thresholds lead to ringing-like artifacts. As illustrated in Figure 7 on the image Voronoi, the proposed solutions obtain substantially better performance by having less noise compared to the standard solution and by having less noise and artifacts compared to BLS-GSM. On the image Teapot, the proposed significant-only solution again obtains better performance due to less ringing. It is interesting to note that while BLS-GSM utilizes fully-steerable transforms which are theoretically better than the DCT, the practical nuances of even simple image singularities prevent it from obtaining better performance.

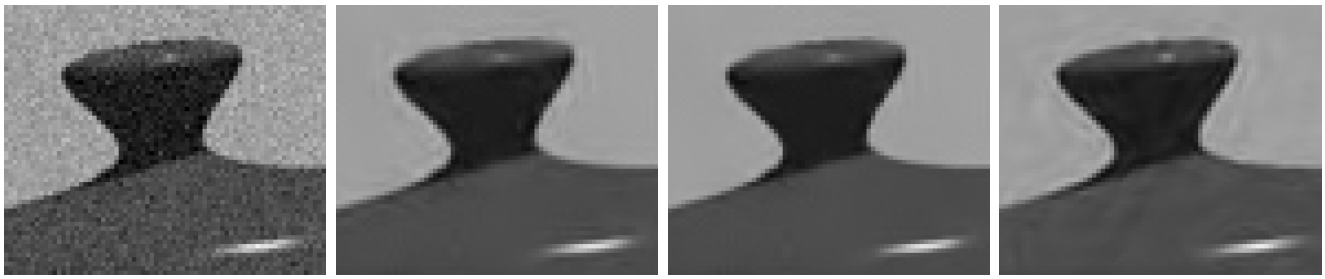
[7]Some performance is lost on Criss-Cross since the square wave pattern in the piecewise uniform portions leads to less sparsity compared to the DC pattern in Graphics.



(a) Noisy, 34.13 dB (b) Standard, 39.54 dB (c) Full, 40.83 dB

(d) Diagonal, 42.08 dB (e) Significant-only, 41.99 dB (f) BLS-GSM, 37.94 dB

Fig. 7. Visual quality after denoising on the image Voronoi. Standard, full, diagonal, and significant-only solutions are at their respective peak PSNR thresholds. The standard solution has residual noise around the edges and BLS-GSM has residual noise and ringing. The proposed solutions alleviate these problems significantly.



(a) Noisy, 28.11 dB (b) Standard, 39.45 dB (c) Signif.-only, 40.32 dB (d) BLS-GSM, 39.03 dB

Fig. 8. Visual quality after denoising on the image Teapot. Standard, and significant-only solutions are at their respective peak PSNR thresholds. The standard solution and BLS-GSM have ringing. The significant-only solution alleviates this problem significantly.

Oracle-Based Denoising Rule: The examples so far have been on spatially uniform thresholding. However, the proposed solutions do extend their effectiveness when the thresholding is more sophisticated. Suppose we utilize the infeasible Oracle-Based Denoising rule of (7) so that thresholding decisions are done by an oracle that has access to the noiseless transform coefficients. Table II shows the results on the simple images using the oracle-based denoising rule. The "Oracle (Standard)" column corresponds to using the oracle-based denoising rule to obtain each denoised estimate, which are then combined using the standard solution. The "Oracle (Significant-only)" column corresponds to using the oracle-based denoising rule to obtain each denoised estimate, which are then combined using the significant-only solution. For reference we also tabulate the results obtained using the standard and significant-only solutions with spatially uniform hard-thresholding (both of these solutions use the respective thresholds that result in peak PSNR). It is clear that switching to the oracle's infeasible thresholding strategy provides noticeable improvements over the corresponding uniformly thresholded results. Yet the derived

	σ_w	Noisy	Standard	Significant-only	Oracle (Standard)	Oracle (Significant-only)
Graphics	5	34.13dB	43.96dB	49.77dB	46.45dB	52.79dB
Criss-Cross	5	34.13dB	42.98dB	45.70dB	44.36dB	46.89dB
Voronoi	5	34.13dB	39.54dB	41.99dB	42.81dB	45.33dB
Teapot	10	28.11dB	39.45dB	40.32dB	43.43dB	44.43dB

TABLE II

DENOISING THE SIMPLE IMAGES USING THE ORACLE-BASED DENOISING RULE.

significant-only solution manages to obtain further improvements so that "Oracle (Significant-only)" substantially outperforms "Oracle (Standard)". Hence, it is clear that optimally weighting the denoised estimates is beneficial even when one is using such an infeasible thresholding strategy.

IV. IMAGE DENOISING ALGORITHM

In our image denoising algorithm we will use the derived significant-only solution but with a modified thresholding (or coefficient denoising) strategy. This strategy uses an auxiliary variable for better thresholding, and is influenced by the significantly better results oracle-based thresholding obtains over spatially uniform hard-thresholding (Table II). As in (7), let $c_i(k)$ be the k^{th} coefficient of the i^{th} transform applied to the noisy signal y, and let $d_i(k)$ denote the corresponding noise-free coefficient.

Auxiliary Variable: Suppose $c_i(k)$ is to be thresholded and assume further that during this operation we have access to an auxiliary variable $a_i(k)$, which we expect to be less noisy than $c_i(k)$ (the infeasible oracle denoiser uses $a_i(k) = d_i(k)$). We obtain this auxiliary variable as follows. We first denoise the noisy image y once using uniform hard-thresholding with threshold τ and combine the denoised estimates via the significant-only solution to result in an intermediate denoised image y'. We then obtain $a_i(k)$ as the k^{th} coefficient of the i^{th} transform applied to the intermediate denoised image y'.

Coefficient Denoising Rule: We would like to set $\hat{c}_i(k) = 0$ if on the average this action will result in less distortion compared to $\hat{c}_i(k) = c_i(k)$. Since we have $c_i(k)$ and $a_i(k)$ available, we can use conditional averages. This results in the rule,

$$\hat{c}_i(k) = \begin{cases} 0, & E[(d_i(k))^2 | a_i(k), c_i(k)] \leq E[(c_i(k) - d_i(k))^2 | a_i(k), c_i(k)] \\ c_i(k), & \text{otherwise.} \end{cases} \tag{41}$$

Observe that when $a_i(k) = d_i(k)$ we obtain the oracle-based denoising rule of (7). Simplifying (41) results in

$$\hat{c}_i(k) = \begin{cases} 0, & c_i(k) \geq 0,\ c_i(k) > 2E[d_i(k) | a_i(k), c_i(k)] \\ 0, & c_i(k) < 0,\ c_i(k) < 2E[d_i(k) | a_i(k), c_i(k)] \\ c_i(k), & \text{otherwise.} \end{cases} \tag{42}$$

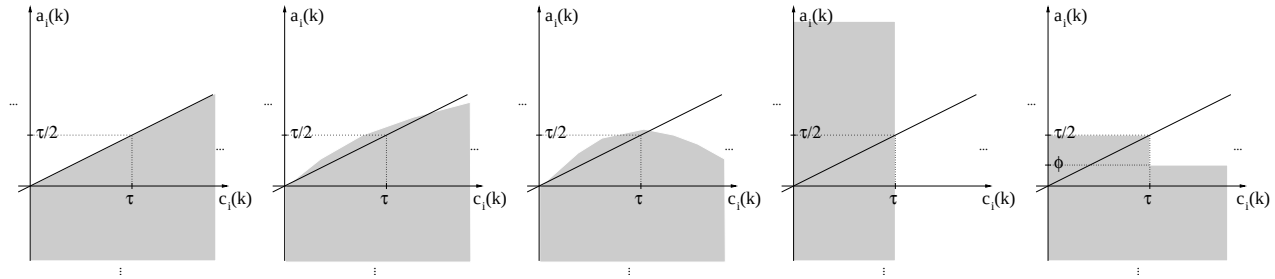
Since $c_i(k)$ is a noise corrupted version of $d_i(k)$ it can be expressed as

$$c_i(k) = d_i(k) + \nu_i(k), \tag{43}$$

where $\nu_i(k)$ is the portion of the additive noise w seen by this coefficient. Suppose the auxiliary variable $a_i(k)$ is also a version of $d_i(k)$ corrupted with noise so that

$$a_i(k) = d_i(k) + \rho_i(k), \tag{44}$$

with $\rho_i(k)$ and $\nu_i(k)$ decorrelated.



(a) $\sigma_{\rho_i(k)} = 0$ (b) $\sigma_{\rho_i(k)} = \frac{\sigma_{\nu_i(k)}}{2}$ (c) $\sigma_{\rho_i(k)} = \frac{3\sigma_{\nu_i(k)}}{2}$ (d) $\sigma_{\rho_i(k)} = \infty$ (e) Heuristic

Fig. 9. Coefficient denoising rule of Equation (42). When the $(c_i(k), a_i(k))$ pair fall in the shaded region $\hat{c}_i(k)$ is set to zero. Otherwise $\hat{c}_i(k) = c_i(k)$. (Only the $c_i(k) \geq 0$ case is shown since the denoising rule is radially symmetric.)

Consider the example case where $d_i(k)$ is a Laplacian random variable of standard deviation $\sigma_{d_i(k)}$ and the additive noise is Gaussian with the noise standard deviation $\sigma_{\nu_i(k)} = \sigma_{d_i(k)}$. The denoising rule of Equation (42) is shown shaded in Figure 9 for this example case for different values of $\sigma_{\rho_i(k)}$ as indicated in the figure. Observe that with $\sigma_{\rho_i(k)} = 0$, $a_i(k) = d_i(k)$ (Figure 9(a)) and we obtain the oracle-based denoising rule. This rule bulges upward as $\sigma_{\rho_i(k)}$ increases (Figures 9(b) and (c)) to arrive at the hard-thresholding rule when $a_i(k)$ no longer provides any information about $d_i(k)$ (Figure 9(d)). Influenced by the nature of these curves, and assuming that $a_i(k)$ is less noisy than $c_i(k)$, we have experimentally determined a heuristic denoising rule

$$\hat{c}_i(k) = \begin{cases} 0, & c_i(k) \geq 0,\ a_i(k) \leq \phi \\ 0, & c_i(k) \leq 0,\ a_i(k) \geq -\phi \\ 0, & |c_i(k)| \leq \tau,\ |a_i(k)| \leq \tau/2 \\ c_i(k), & \text{otherwise.} \end{cases} \tag{45}$$

as shown in Figure 9(e). This rule is easily parameterized using two threshold parameters τ and ϕ. For all our results in Section V, these two parameters were determined as a function of the noise standard deviation σ_w using a training set of images outside of our simulation set. The overall denoising algorithm is thus obtained via the application of the algorithm outlined in Appendix I twice. First with a spatially uniform threshold τ to obtain y' and then with the hard-thresholding replaced with the heuristic rule of (45) for the final result.

V. Simulation results

We applied the significant-only solution combined with the heuristic denoising rule of Section IV on the test images shown in Figure 10. Each image was corrupted with Gaussian noise of standard deviation σ_w. We used a fully overcomplete set of block DCT (8×8) transforms so that the resulting denoising is fully translation invariant, i.e., $\mathbf{H}_i$, $i = 1, \ldots, M$, represent all translations of a 8×8 block DCT with $M = 64$. For immediate comparison

we utilize the high performance BLS-GSM method of [28] (using the fully-steerable pyramid) and the standard solution with uniform hard-thresholding (using thresholds optimized as a function of σ_w over a training set outside of our simulation set). We refer the reader to [28], [30] for a compilation of many other results in the literature. The PSNR of denoised images can be seen tabulated in Table III. We note that our aim is only to show the competitiveness of our techniques as the utilized algorithm can be fine tuned and optimized in many ways. For example, the utilized 8×8 block DCT limits the equivalent denoising filters to be of spatial size 15×15, which tends to hinder performance on uniform portions of the image and at high noise levels. Such issues can be alleviated by using larger or multiscale transforms. Furthermore as Reviewer 1 has pointed out, the denoising rule in Section IV can be carried out with more iterations so that auxiliary variables having more fidelity can be generated and utilized in the rule.

Fig. 10. Test images Lena, Barbara, Peppers, Cameraman, Boat, House, Photo1, Photo2. Cameraman and House are 256×256. All other images (including peppers) are 512×512.

As expected, the proposed solution obtains the biggest improvements on images with high contrast, sharp singularities. Except for a few cases at small noise levels, the proposed significant-only solution consistently outperforms the standard solution. BLS-GSM tends to have comparatively higher performance on images like Lena and at higher noise levels, where performance on large smooth patches is of considerable importance. Due the multiresolutional dictionary it utilizes, BLS-GSM can take advantage of large spatial size basis functions. On the other hand, both the standard and the significant-only solutions are limited to smaller spatial size equivalent filters, which limits their performance in such cases. While the relative performance increases drop compared to the toy examples of Section III, the significant-only solution remains consistently better on images that contain high contrast singularities.

In terms of visual quality around image singularities, the proposed solution typically provides the best results as illustrated in Figure 11. Consistent with earlier examples, the standard solution and BLS-GSM have more residual noise and ringing-like artifacts in comparison.

VI. Conclusion

In this paper we concentrated on the final estimate combination stage of denoising with an overcomplete set of transforms and proposed a method that optimally combines the denoised estimates provided by denoising with each of the transforms. Our formulation is equivalent to deriving a per-pixel optimal inverse transform, and as long as $\boldsymbol{\eta}_n$ in (23) can be suitably approximated, it allows practical solutions for general transforms and thresholding rules. The derived solutions rely on sparse decompositions with the utilized transforms but do not require further assumptions on image statistics. Unlike existing work, which attempts to devise elaborate statistical models for

	σ_w	Lena	Barbara	Peppers	Cameraman	Boat	House	Photo1	Photo2
	Standard	38.47 dB	**37.98** dB	37.46 dB	37.80 dB	**37.13** dB	39.04 dB	38.58 dB	37.41 dB
$\sigma_w = 5$	Signif.-only	38.43 dB	37.84 dB	**37.47** dB	**37.95** dB	37.11 dB	**39.07** dB	**38.67** dB	**37.50** dB
(34.13 dB)	BLS-GSM	**38.49** dB	37.79 dB	37.12 dB	37.41 dB	36.97 dB	38.65 dB	38.43 dB	37.17 dB
	Standard	35.33 dB	**34.07** dB	34.61 dB	33.40 dB	33.52 dB	35.51 dB	34.53 dB	33.07 dB
$\sigma_w = 10$	Signif.-only	35.40 dB	**34.07** dB	**34.65** dB	**33.74** dB	**33.60** dB	**35.70** dB	**34.83** dB	**33.33** dB
(28.11 dB)	BLS-GSM	**35.61** dB	34.03 dB	34.57 dB	33.04 dB	33.58 dB	35.35 dB	34.53 dB	32.91 dB
	Standard	33.46 dB	31.78 dB	33.06 dB	30.99 dB	31.53 dB	33.62 dB	32.17 dB	30.61 dB
$\sigma_w = 15$	Signif.-only	33.66 dB	**31.90** dB	**33.28** dB	**31.46** dB	31.69 dB	**33.93** dB	**32.57** dB	**30.97** dB
(24.59 dB)	BLS-GSM	**33.90** dB	31.86 dB	33.13 dB	30.73 dB	**31.70** dB	33.64 dB	32.29 dB	30.55 dB
	Standard	32.09 dB	30.11 dB	31.92 dB	29.36 dB	30.14 dB	32.27 dB	30.53 dB	28.91 dB
$\sigma_w = 20$	Signif.-only	32.38 dB	**30.33** dB	**32.24** dB	**29.87** dB	30.33 dB	**32.66** dB	**30.97** dB	**29.34** dB
(22.09 dB)	BLS-GSM	**32.66** dB	30.32 dB	32.08 dB	29.29 dB	**30.38** dB	32.39 dB	30.74 dB	28.94 dB
	Standard	31.03 dB	28.84 dB	30.97 dB	28.18 dB	29.05 dB	31.17 dB	29.29 dB	27.60 dB
$\sigma_w = 25$	Signif.-only	31.36 dB	29.09 dB	**31.40** dB	**28.66** dB	29.30 dB	**31.67** dB	**29.75** dB	**28.05** dB
(20.15 dB)	BLS-GSM	**31.69** dB	**29.13** dB	31.21 dB	28.22 dB	**29.37** dB	31.40 dB	29.60 dB	27.72 dB
	Standard	30.15 dB	27.82 dB	30.18 dB	27.28 dB	28.17 dB	30.25 dB	28.33 dB	26.58 dB
$\sigma_w = 30$	Signif.-only	30.53 dB	28.06 dB	**30.64** dB	**27.75** dB	28.46 dB	**30.83** dB	**28.74** dB	**27.00** dB
(18.57 dB)	BLS-GSM	**30.85** dB	**28.15** dB	30.45 dB	27.38 dB	**28.55** dB	30.55 dB	28.68 dB	26.72 dB

TABLE III

DENOISING RESULTS ON TEST IMAGES. ROWS CORRESPONDING TO "SIGNIF.-ONLY" ILLUSTRATE THE PERFORMANCE OF THE PROPOSED SOLUTION. EACH RESULT IS AN AVERAGE OF EIGHT RUNS WITH DIFFERENT NOISE REALIZATIONS. BLS-GSM RESULTS ARE EITHER FROM [28] (WHEN AVAILABLE) OR OBTAINED USING THE AUTHORS' SOFTWARE. THE STANDARD DEVIATION OF THE RESULTS AS A FUNCTION OF THE TESTED σ_w ARE $[0.017, 0.022, 0.022, 0.027, 0.028, 0.033]$ FOR THE STANDARD SOLUTION AND $[0.016, 0.022, 0.023, 0.027, 0.030, 0.032]$ FOR THE "SIGNIF.-ONLY" SOLUTION.

transform coefficients over the myriad types of image singularities, i.e., over non-sparse regions of the signal, our work successfully solves the dual problem by simply modeling coefficients over sparse regions of the signal. Our results allow even singularity-blind transforms to obtain very high performance around image singularities as our solutions actively determine and prefer the better estimates at the combination step. As we illustrated, armed with the proposed work, well-known suboptimal transforms such as block DCTs can readily obtain substantial improvements over elaborate directional basis.

Our results clearly demonstrate the effectiveness of the proposed techniques. When compared to some of the best results in the literature that use sophisticated directional transforms and statistical modeling, a naive implementation of our work consistently provides competitive PSNR values and better quality around image singularities. Better thresholding strategies that better retain the noisy coefficients over singularities, while accomplishing high performance denoising over sparse regions, are expected to improve the performance of the proposed work. Such strategies, while providing slightly worse performance on some denoised estimates (in order to keep Equation (30) valid), will allow the final combination to have superior performance as the proposed work is designed to correct

the shortcomings of some of the denoised estimates using others.

Since the proposed method can be seen as a per-pixel adaptive inverse transform, it can also be deployed effectively with transform optimization approaches such as [12]. We hence point to joint optimization of transform, coefficient denoising-rule, and weighted combination as an interesting future research item. Another interesting issue is that with the weight normalization constraint $\sum_i \gamma_i(n) = 1$, the solutions of Equation (25) are expected to be robust to signal to noise cross-correlations in the sense that one no longer needs to assume that the noise is independent of the signal. One can hence deploy the proposed work on denoising quantization noise since the statistical signal model provided under the umbrella of sparsity remains robust in that case with "on-off" coefficient denoising rules [17]. A preliminary version of this work successfully deploying the significant-only solution in denoising quantization noise in h.264 compressed video can be found in [18], [19].

As illustrated in this paper there is a distinct advantage to deploying transforms in a translation invariant fashion. Yet one is tempted to think that fully expansive transforms, such as translation invariant DCTs, may bring an unreasonable computational complexity burden. On this account it is useful to note that one can implement DCTs using very fast multiplier-less algorithms and implement translation invariant transforms in fast ways that do not duplicate computations[8]. While the number of computations is large, these computations are highly parallelizable, predictable, and they can be executed with minimal code-branching, pipeline-stalling, etc. On modern hardware that can execute many instructions concurrently, unpredictable branches and pipeline stalls become so expensive that a sophisticated but critically decimated transform can take more execution time compared to an overcomplete DCT implementation. One can also utilize DSP chips, graphics cards, etc., since acceleration of DCT computations is ubiquitous on modern multimedia related add-on boards. In terms of memory requirements, we note that the utilized translation invariant DCTs allow one to deploy software and hardware with very small memory requirements. This is because the spatial sizes of the equivalent filters are small and one only needs to have access to limited portions of input data. In [18] we used such a fast, very low-memory implementation to denoise quantization noise in compressed video at video frame-rates.

When one considers denoising with a critically decimated basis one can resort to well-known mathematical approximation tools on statistical image models to determine basis that are optimal or near optimal in a well defined mathematical sense. Yet when one is allowed operation in an overcomplete, translation invariant fashion, one can no longer rely on such calculations to rule out "suboptimal" basis, as such basis, armed with good combination strategies, can be made to perform close to if not better than optimal transforms. The distinguishing factors that are in favor of sophisticated transforms also become fuzzy when one allows realistic variations in image singularities. Rather than trying to do sophisticated modeling of image singularities or designing new directional transforms, this paper has shown an alternative avenue that leads to high performance around singularities. This avenue is enabled by using a localized, overcomplete decomposition, judiciously sorting through the resulting denoised estimates at every pixel, preferring those estimates that offer good performance, and doing damage control on those that fail. The main contribution of this paper is thus the proposal of a simple combination strategy that enables this additional

[8]On current generation of CPUs an unoptimized implementation of the proposed algorithm runs on the order of a second for 512×512 images.

avenue which can be used to better application performance around image singularities.

APPENDIX I
EXAMPLE ALGORITHM FOR BLOCK TRANSFORMS USING HARD-THRESHOLDING

Suppose the image is $N_1 \times N_2$ and we are using a block transform where each transform basis function has spatial support $B \times B$. Let $T > 0$ be a given threshold. The following pseudo-code implements the significant-only solution using hard-thresholding.

Algorithm 1 (Significant-only solution with block transforms and hard-thresholding)

```
// assume norm(.,.), denoised(.,.) are
// initialized to zero.
  for i=1,...,N1-B+1 for j=1,...,N2-B+1 {
      for k=1,...,B for l=1,...,B
          b(k,l)=image(i+k-1,j+l-1);                  // copy data into block b
      c_b=block_transform(b);                         // transform coefficients of block b
      t_c_b=hard_threshold(cb,T);
      z=1.0/(number_of_nonzero_coefficients(t_c_b)+eps); // unnormalized weight
      de_b=inverse_block_transform(t_c_b);            // denoised estimate
      for k=1,...,B for l=1,...,B {
          denoised(i+k-1,j+l-1) += z * de_b(k,l);     // weighted combination
          norm(i+k-1,j+l-1) += z;                     // unnormalized weight accounting
      }
  }
  for i=1,...,N1 for j=1,...,N2
      denoised(i,j) /= norm(i,j);                     // normalize
```

REFERENCES

[1] A. Buades, B. Coll, and J.M Morel, "A non local algorithm for image denoising," IEEE Int. Conf. on Computer Vision and Pattern Recognition, CVPR 2005.vol. 2, pp: 60 - 65.

[2] S.G. Chang, Bin Yu, M. Vetterli, "Spatially adaptive wavelet thresholding with context modeling for image denoising," *IEEE Trans. on Image Proc.*, vol. 9, pp. 1522 -1531, Sept. 2000.

[3] A. Cohen, I. Daubechies, O. G. Guleryuz, and M. T. Orchard, "On the importance of combining wavelet-based nonlinear approximation with coding strategies," *IEEE Trans. Info. Theory*, vol. 48, no. 7, pp. 1895-1921, July 2002.

[4] R. R. Coifman and D. L. Donoho, "Translation invariant denoising," in Wavelets and Statistics, Springer Lecture Notes in Statistics 103, pp. 125-150, New York:Springer-Verlag.

[5] M. S. Crouse, R. D. Nowak, and R. G. Baraniuk, "Wavelet-based statistical signal processing using hidden markov models", IEEE Trans. On Signal Processing, vol. 46, no. 4, pp. 886-902, 1998.

[6] 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, LVII: 1413-1457, 2004.

[7] M. N. Do, P. L. Dragotti, R. Shukla, and M. Vetterli, "On the compression of two-dimensional piecewise smooth functions,' Proc. IEEE Int. Conf. on Image Proc. ICIP 01, Thessaloniki, Greece, 2001.

[8] D. L. Donoho, "Denoising by soft-thresholding", IEEE Trans. Info. Theory, 43:613–627, 1995.

[9] D. L. Donoho and I. M. Johnstone, "Ideal spatial adaptation via wavelet shrinkage," *Biometrika*, vol. 81., pp. 425-455, 1994.

[10] D. L. Donoho, M. Vetterli, R. A. DeVore, and I. Daubechies, "Data Compression and Harmonic Analysis". IEEE Transactions on Information Theory 44(6): 2435-2476, 1998.

[11] M. Elad, "Why Simple Shrinkage is Still Relevant for Redundant Representations?", to appear in the IEEE Trans. On Information Theory.

[12] M. Elad and M. Aharon, "Image Denoising Via Sparse and Redundant representations over Learned Dictionaries", Submitted to the IEEE Trans. on Image Processing.

[13] R. Eslami and H. Radha, "Image Denoising using Translation-Invariant Contourlet Transform," Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Processing (ICASSP), vol. 4, pp. 557-560, Philadelphia, PA, Mar. 2005.

[14] M. Figueiredo and R. Nowak, “An em algorithm for wavelet-based image restoration,” IEEE Transactions on Image Processing, July 2003.

[15] A. Foi, V. Katkovnik, and K. Egiazarian, ”Pointwise Shape-Adaptive DCT for High-Quality Denoising and Deblocking of Grayscale and Color Images”, (accepted) IEEE Trans. Image Process. (preprint)

[16] A. K. Fletcher, K. Ramchandran, and V. K Goyal, ”Wavelet Denoising by Recursive Cycle Spinning”, Proc. IEEE Int. Conf. on Image Proc., (ICIP), Rochester, NY, Sept. 2002.

[17] O. G. Guleryuz, “Linear, Worst-Case Estimators for Denoising Quantization Noise in Transform Coded Images,” IEEE Transactions on Image Processing, to appear.

[18] O. G. Guleryuz, “A Nonlinear Loop Filter for Quantization Noise Removal in Hybrid Video Compression,” Proc. IEEE Int’l Conf. on Image Proc. (ICIP2005), Genova, Italy, Sept. 2005.

[19] O. G. Guleryuz, “A Nonlinear Loop Filter for Quantization Noise Removal in Hybrid Video Compression,” presentation material, online: http://eeweb.poly.edu/∼onur/publish/guleryuz_hybrid_loop.ppt

[20] O. G. Guleryuz, “Weighted Overcomplete Denoising,” Proc. Asilomar Conference on Signals and Systems, Pacific Grove, CA, Nov. 2003.

[21] N. G. Kingsbury, “A Dual-Tree Complex Wavelet Transform with improved orthogonality and symmetry properties”, in *Proc. IEEE Conf. on Image Proc.*, Sept. 2000.

[22] X. Li and M. T. Orchard, “Spatially adaptive image denoising under overcomplete expansion,” in *Proc. IEEE Int. Conf. on Image Proc.*, Sept. 2000.

[23] S. Mallat, “A Wavelet Tour of Signal Processing.” New York: Academic Press, 1998.

[24] M. K. Mihcak, I. Kozintsev, K. Ramchandran, and P. Moulin “Low-complexity image denoising based on statistical modeling of wavelet coefficients,” IEEE Trans. Sig. Proc., vol. 6, no. 12, pp. 300-303, December 1999.

[25] P. Moulin and J. Liu “Analysis of multiresolution image denoising schemes using generalized Gaussian and complexity priors”, IEEE Transactions on Information Theory, Vol. 45, No. 3, pp. 909-919, April 1999.

[26] R. Oktem, L. Yaroslavsky, K. Egiazarian, and J. Astola, “Transform domain approaches for image denoising,” Journal of Electronic Imaging, Vol. 11, Issue 2, pp. 149-156, April 2002.

[27] D. D.-Y. Po and M. N. Do, “Directional multiscale modeling of images using the contourlet transform,” IEEE Transactions on Image Processing, vol. 15, no. 6, pp. 1610-1620, June 2006.

[28] J. Portilla, V. Strela, M. Wainwright, and E. P. Simoncelli, “Image Denoising using Scale Mixtures of Gaussians in the Wavelet Domain,” IEEE Transactions on Image Processing. vol 12, no. 11, pp. 1338-1351, November 2003.

[29] L. Sendur and I. Selesnick, “Bivariate shrinkage functions for wavelet-based denoising exploiting interscale dependency,” IEEE Transactions on Signal Processing. 50(11):2744 – 2756, November 2002.

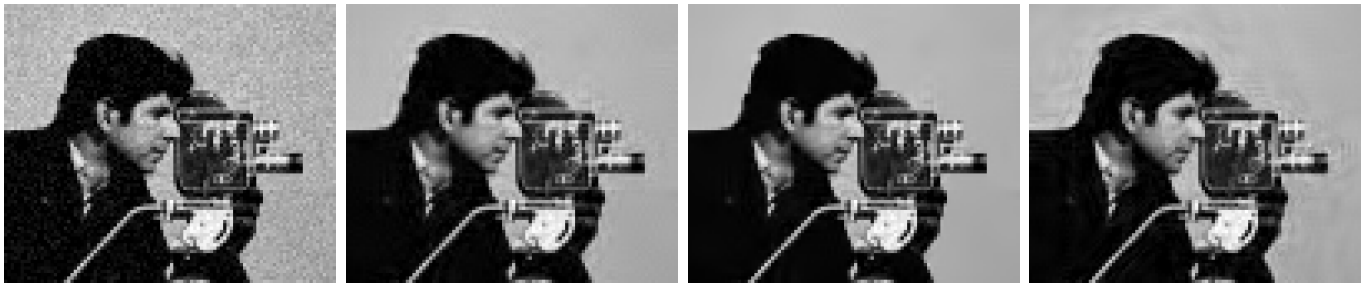
[30] L. Shen, M. Papadakis, I. A. Kakadiaris, I. Konstantinidis, D. Kouri, and D. Hoffman, “Image Denoising Using a Tight Frame,” IEEE Transactions on Image Processing, Vol. 15, no. 5, May 2006.

[31] J.L. Starck, E. Candes, and D.L. Donoho, “The Curvelet Transform for Image Denoising”, IEEE Transactions on Image Processing , 11, 6, pp 670 -684, 2002.

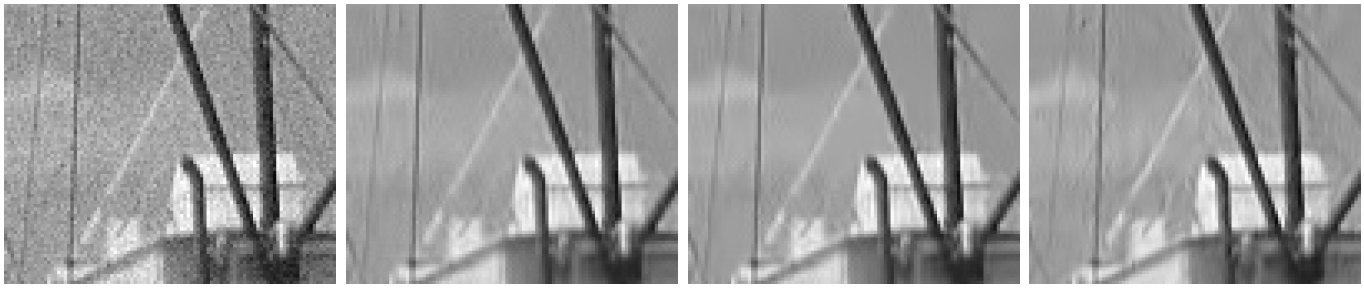
[32] H. Stark and J. W. Woods, “Probability, Random Processes, and Estimation Theory for Engineers,” Prentice Hall, Englewood Cliffs, NJ, 1986.

[33] P. K. Varshney, “Distributed Detection and Data Fusion,” Springer-Verlag New York, Inc., NY 1997.

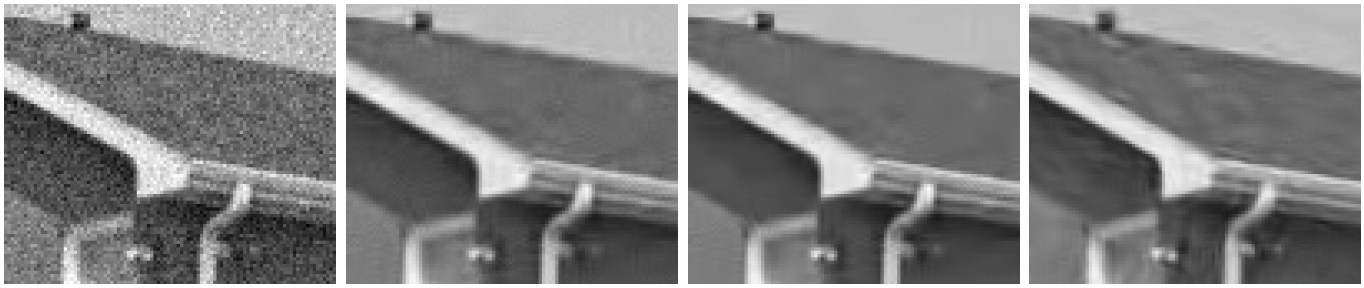
[34] Z. W. Ziou, D. Armenakis, C. Li, and D. Q. Li, “A comparative analysis of image fusion methods”, IEEE Transactions on Geoscience and Remote Sensing, 43, 6, pp 1391- 1402, 2005



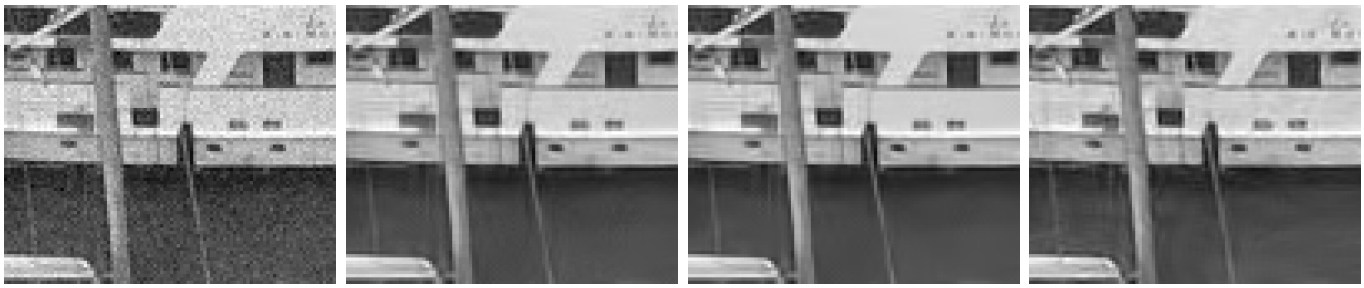
(a) Noisy, 28.11 dB | Standard, 33.39 dB | Signif.-only, 33.73 dB | BLS-GSM, 32.99 dB



(b) Noisy, 28.11 dB | Standard, 33.51 dB | Signif.-only, 33.58 dB | BLS-GSM, 33.58 dB



(c) Noisy, 24.59 dB | Standard, 33.58 dB | Signif.-only, 33.90 dB | BLS-GSM, 33.64 dB



(d) Noisy, 28.11 dB | Standard, 34.50 dB | Signif.-only, 34.81 dB | BLS-GSM, 34.50 dB

(e) Noisy, 28.11 dB | Standard, 33.02 dB | Signif.-only, 33.29 dB | BLS-GSM, 32.91 dB

Fig. 11. Visual quality after denoising on the image (a) Cameraman ($\sigma_w = 10$), (b) Boat ($\sigma_w = 10$), (c) House ($\sigma_w = 15$), (d) Photo1 ($\sigma_w = 10$), (e) Photo2 ($\sigma_w = 10$).