Original articles

Denoising strategies for general finite frames

D. De Canditiis*, M. Pensky, P.J. Wolfe

CNR, Rome, Italy
UCF, Orlando, USA
UCL, London, UK

Abstract

Overcomplete representations such as wavelets and windowed Fourier expansions have become mainstays of modern statistical data analysis. In the present work, in the context of general finite frames, we derive an oracle expression for the mean quadratic risk of a linear diagonal de-noising procedure which immediately yields the optimal linear diagonal estimator. Moreover, we obtain an expression for an unbiased estimator of the risk of any smooth shrinkage rule. This last result motivates a set of practical estimation procedures for general finite frames that can be viewed as the generalization of the classical procedures for orthonormal bases. A simulation study verifies the effectiveness of the proposed procedures with respect to the classical ones and confirms that the correlations induced by frame structure should be explicitly treated to yield an improvement in estimation precision.
ⓒ 2017 International Association for Mathematics and Computers in Simulation (IMACS). Published by Elsevier B.V. All rights reserved.

Keywords: Finite frames; Block thresholding; Shrinkage; Signal de-noising; SURE

1. Introduction

Regression using more predictors than observations has received a great deal of attention in recent years, from viewpoints as diverse and fundamental as high-dimensional inference and regularization, approximation theory, and sparse coding. While an orthogonal basis yields fast algorithms and classical asymptotic theory, it can often fail to represent a particular function of interest efficiently. As a result, overcomplete representations such as wavelets and windowed Fourier expansions have become mainstays of modern statistics and signal processing.

Such representations are formalized through the theory of frames. Frames can be generated by the action of operators on a template function (mother wavelet or Gabor atom), or be unstructured and random (as in compressive sensing). Results for frame-based regression come in two flavors: those that derive from statistics which usually aim for universality, and those that derive from signal processing which usually exploit structure of specific frame families (usually wavelet or Fourier frames).

In the context of thresholding estimators for de-noising purposes, [6] gave some of the first related results, for multiwavelets. Some of the latest results are by [12] for a particular windowed Fourier frame and [7], where an

---

* Corresponding author at: CNR, Rome, Italy.
  E-mail address: d.decanditiis@iac.cnr.it (D. De Canditiis).

universal threshold is derived for frames satisfying rather stringent conditions which ensure that the threshold depends on the number of the frame functions but not on the frame structure. In general, frame based de-noising procedures are often derived using adaptation of de-noising methodologies developed for the case of orthonormal bases by exploring specific characteristic of an unknown signal (as in Bayesian framework, see, e.g., [13]), and/or by exploring specific characteristic of the frame structure (see, e.g., [10,8,14]).

The objective of this paper is to bridge the existing gap between mathematical and statistical theories on one hand and engineering practice on the other. In particular, the purpose of this paper is to give a comprehensive study of de-noising properties of frames in a way that allows practitioners to trade off between computational convenience and statistical efficacy. We do this by providing several new and practically implementable techniques that generalize standard ones taking the correlation structure of the frame coefficients into account.

We start with deriving an oracle expression for the quadratic risk of a linear diagonal estimator which, in the case of an orthonormal basis, immediately yields the popular Wiener estimator. We also show that it is naturally related to block thresholding, which is proved to be very effective (see, e.g., [2]). We then derive the Stein's (1981) unbiased risk estimator (SURE) for all smooth shrinkage functions and general frames, recovering the results of [1,9] from the image processing literature. Using SURE, we then obtain explicit expressions for the linear shrinkage and the soft and the hard thresholding de-noising procedures for general frames that are standard in signal processing literature for the case of orthonormal bases. A finite-sample simulation study verifies the effectiveness of the proposed methods with respect to classical techniques.

2. Oracle risk expression for frame-based regression

Consider the canonical problem of recovering some $f \in \mathcal{R}^n$ from its corrupted observation x:

$$x = f + \delta, \quad \delta \sim \mathcal{N}(0, \sigma^2 I_n). \tag{2.1}$$

Any matrix $W \in \mathcal{C}^{N \times n}$ of full column rank is a frame for $\mathcal{R}^n$ when n and N are finite; columns of its conjugate transpose W^* are the corresponding frame vectors. Applying W to (2.1), we obtain

$$Wx = Wf + W\delta \implies y = \theta + \varepsilon, \quad \varepsilon \sim \mathcal{N}(0, \sigma^2 U) \tag{2.2}$$

where covariance $U = WW^* \in \mathcal{C}^{N \times N}$ is of rank n, and we let $y = Wx$, $\theta = Wf$, and $\varepsilon = W\delta$.

The goal of our analysis is first to reduce noise in the vector y of frame coefficients by shrinking or thresholding its components – thus yielding vector $\hat{\theta}$ – and subsequently to estimate f by

$$\hat{f} = (W^*W)^{-1}W^*\hat{\theta} = W^+\hat{\theta}, \tag{2.3}$$

with $W^+ = (W^*W)^{-1}W^*$ the Moore–Penrose inverse of W, termed its canonical dual frame. When, in particular, $W^*W = \alpha I_n$, then the frame is said to be tight, in which case $U^- = (W^+)^*W^+$ reduces to $U^- = \alpha^{-2}U$ and $\|Wf\|_2^2 = \frac{1}{\alpha^2}\|f\|_2^2$.

Let Γ be a diagonal matrix with vector γ on the diagonal, so that $\Gamma = diag(\gamma)$. If we consider any estimator of type $\hat{\theta} = \Gamma y$ with Γ fixed (not depending on y), then an expression for its oracle risk is given by the following statement.

Theorem 1. *Let $U = WW^*$ and $U^- = (W^+)^*W^+$. If $\hat{\theta} = \Gamma y$ for fixed, diagonal Γ, then*

$$\mathbb{E}\|\hat{f} - f\|^2 = Tr\left\{U^-(I_N - \Gamma)\theta\theta^*(I_N - \Gamma) + \sigma^2 \Gamma U \Gamma U^-\right\}. \tag{2.4}$$

In particular, if the frame is tight, then $U^- = \alpha^{-2}U$.

Proof. Note that for all $x \in \mathcal{C}^{n \times 1}$ it holds $\|x\|^2 = \sum_{i=1}^n x_i^2 = Tr(xx^*)$. Hence, we obtain

$$\begin{aligned}
\mathbb{E}\|\hat{f} - f\|^2 &= \mathbb{E}\|W^+(\Gamma y - \theta)\|^2 \\
&= \mathbb{E}Tr\left\{W^+(\Gamma y - \theta)(\Gamma y - \theta)^*(W^+)^*\right\} \equiv \Delta_1 + \Delta_2.
\end{aligned}$$

Then, by using $\mathbb{E}(\epsilon) = 0$ and the cyclic permutation property of the trace operator, we derive

$$
\begin{aligned}
\Delta_1 &= Tr\left\{W^+(I_N - \Gamma)\theta\theta^*(I_N - \Gamma)(W^+)^*\right\} \\
&= Tr[U^-(I_N - \Gamma)\theta\theta^*(I_N - \Gamma)]; \\
\Delta_2 &= Tr\left\{W^+\Gamma\mathbb{E}\left(\epsilon\epsilon^*\right)\Gamma^*(W^+)^*\right\} = \sigma^2 Tr\left(\Gamma U\Gamma U^-\right). \quad \square
\end{aligned}
$$

It is worth to observe that Theorem 1 implies that in the case of hard thresholding, one needs to minimize the risk of (2.4) over the set of binary diagonal matrices. In the case of an orthonormal basis, the oracle expression in (2.4) takes the familiar form $\sum_{i=1}^{n}\left\{\theta_i^2\,\mathbb{I}(\gamma_i = 0) + \sigma^2\,\mathbb{I}(\gamma_i = 1)\right\}$. This motivates the choice of a threshold based on the magnitude of a coefficient for any kind of an orthonormal basis. This situation changes, however, whenever matrix W of frame elements ceases to be unitary—because the specific frame structure should now be taken into account. Indeed, Theorem 1 implies that the choice of coefficients to "keep" ($\gamma_i = 1$) or "kill" ($\gamma_i = 0$) depends not only on their values, but also on the entries of matrix U.

Denote by $A \circ B$ the Hadamard (element-wise) matrix product of A and B. Then, Theorem 1 yields the following expression for the optimal diagonal shrinkage rule.

Corollary 1. *The optimal diagonal linear shrinkage rule $\hat{\theta} = \gamma \circ y$ in the setting of Theorem 1 is given by*

$$
\gamma = \left\{(\theta\theta^*) \circ U^- + \sigma^2(U \circ U^-)\right\}^{-1}\left\{(\theta\theta^*) \circ U^-\right\}1_N, \tag{2.5}
$$

where, again, $U^- = \alpha^{-2}U$ for the special case of a tight frame.

Proof. Note that minimization of (2.4) with respect to $\Gamma = diag(\gamma)$ leads to minimization, with respect to γ, of the quadratic form

$$
\arg\min_{\gamma}(\gamma^*A\gamma - 2\gamma^*b), \tag{2.6}
$$

where $A = (\theta\theta^*) \circ U^- + \sigma^2(U \circ U^-)$ and $b = ((\theta\theta^*) \circ U^-)1_N$, with 1_N the column vector of all ones. The existence of the inverse in (2.5) is guaranteed to exist, since the Hadamard product of any two positive-definite matrices is positive-definite. $\quad\square$

Corollary 1 generalizes the linear diagonal oracle estimator for an orthonormal basis. In a special case of a tight frame, when $N = n$ and $\alpha = 1$ and $U = U^- = I_n$, we recover, directly from (2.5), the well-known Wiener linear shrinkage rule $\gamma_i = \theta_i^2/(\theta_i^2 + \sigma^2)$.

Corollary 1 also implies that the optimal weights are functions not only of θ_i but also of other coefficients in their respective neighborhoods. Thus, (2.5) represents an overlapping block shrinkage procedure where the lengths of the blocks are automatically determined by the correlations induced by the frame operator. The latter directly motivates the block shrinkage procedures that are known to have good risk properties (see, e.g., [2,4,14] among others).

3. Unbiased risk estimation for the frame-based regression

The oracle expression of Theorem 1 enables the construction of unbiased estimators of the risk $\mathbb{E}\|\hat{f} - f\|^2$. Indeed, if we re-write the matrix $\Theta = \theta\theta^*$ appearing in (2.4) as $\Theta = \mathbb{E}(yy^*) - \sigma^2 U$ and estimate it via $\hat{\Theta} = yy^* - \sigma^2 U$, then, we obtain the following unbiased estimator of the risk.

Corollary 2. *Let $\hat{\theta} = \Gamma y$ for a fixed, diagonal matrix Γ. Then, an unbiased estimator for the risk $R = \mathbb{E}\|\hat{f} - f\|^2$ is*

$$
\widehat{R} = \sigma^2 n + \Delta, \quad \Delta = y^*(I_N - \Gamma)U^-(I_N - \Gamma)y - 2\sigma^2 Tr\left\{U^-U\,(I_N - \Gamma)\right\}. \tag{3.1}
$$

In particular, if Γ induces a hard thresholding rule, so that $\hat{\theta} = \gamma \circ y$ with $\gamma \in \{0, 1\}^N$, then

$$
\Delta = \sum_{i,j=1}^{N}\left\{y_i y_j U_{ij}^- - 2\sigma^2(U^-U)_{ii}\,\mathbb{I}(i = j)\right\}\mathbb{I}(\gamma_i = 0)\,\mathbb{I}(\gamma_j = 0). \tag{3.2}
$$

Proof. Note that

$$Tr\left(\Gamma U \Gamma U^-\right) = Tr\left\{UU^- + (I_N - \Gamma)U(I_N - \Gamma)U^- - 2(I_N - \Gamma)UU^-\right\}.$$

Now replace $\theta\theta^*$ in (2.4) by $yy^* - \sigma^2 U$ and observe that, since $W^+W = I_n$,

$$Tr\left(UU^-\right) = Tr\left\{WW^*(W^+)^*W^+\right\} = Tr\left\{(W^+W)^* W^+W\right\} = n, \tag{3.3}$$

which proves expression (3.1). $\quad\square$

We shall later use Corollary 2 and the specific form of (3.2) to implement hard thresholding. First, however, note that since Θ is positive semi-definite, its diagonal elements must be nonnegative, implying $\hat{\Theta}_{ii} = y_i^2 - \sigma^2 U_{ii} \geq 0$ for each i. These inequalities themselves enforce thresholds $\sigma\sqrt{U_{ii}}$ on the values of y_i.

The oracle risk (2.4) and its unbiased estimator (3.1) are very useful when working with linear estimators, i.e., estimators for which Γ does not depend on y. However, very often linear estimator is poor and more sophisticated formulae are required for Γ depending on y. To take into account this dependence, in the following theorem we propose a modification of SURE for a general estimator $\hat{\theta} = y + g(y)$.

Theorem 2. *Assume the settings of Section 2 and let $\hat{\theta} = y + g(y)$, where $g(y) : \mathcal{R}^N \to \mathcal{R}^N$ is a continuous and piecewise differentiable column vector function. Let $Z = \nabla_y g^*(y)$ be the $N \times N$ matrix with components $Z_{ij} = \partial g_j(y)/\partial y_i$. Then SURE for $\mathbb{E}\|\hat{f} - f\|^2$ is given by (3.1) with*

$$\Delta = g^*(y)U^- g(y) + 2\sigma^2 Tr\left\{U^- U Z\right\}. \tag{3.4}$$

Here, again $U^- = \alpha^{-2} U$ and $U^- U = \alpha^{-1} U$ for the special case of a tight frame.

Proof. First, let us show that under conditions of Theorem 2, one has

$$\mathbb{E}\left\{(\hat{\theta} - \theta)(\hat{\theta} - \theta)^*\right\} = \sigma^2 U + \mathbb{E}\left\{g(y)g^*(y)\right\} + 2\sigma^2 U\mathbb{E}(Z). \tag{3.5}$$

To see this, write $\mathbb{E}\left\{(\hat{\theta}-\theta)(\hat{\theta}-\theta)^*\right\} = \mathbb{E}\left\{(y - \theta)(y - \theta)^* + g(y)g^*(y)\right\} + 2\mathbb{E}\left\{(y - \theta)g^*(y)\right\} \equiv \Omega_1 + 2\Omega_2$. Observe $\Omega_1 = \sigma^2 U + \mathbb{E}\left\{g(y)g^*(y)\right\}$ and $\Omega_2 = W\mathbb{E}\left\{(x - f)g^*(Wx)\right\}$, as $y = Wx$ and $\theta = Wf$.

Let $C_\sigma = (2\pi\sigma^2)^{-n/2}$ and observe that $Q = \mathbb{E}[(x - f)g^*(Wx)]$ is the $n \times N$ matrix with components

$$
\begin{aligned}
Q_{lj} &= \mathbb{E}[(x_l - f_l)g_j(Wx)] = C_\sigma \int (x_l - f_l) \, g_j(Wx) \, \exp\left(\frac{-\|x - f\|^2}{2\sigma^2}\right) dx \\
&= -C_\sigma\sigma^2 \int g_j(Wx) \, d_l \left\{\exp\left(\frac{-\|x - f\|^2}{2\sigma^2}\right)\right\} dx_1 \cdots dx_{l-1}dx_{l+1} \cdots dx_n \\
&= C_\sigma\sigma^2 \int \frac{\partial}{\partial x_l} \left\{g_j(Wx)\right\} \exp\left(\frac{-\|x - f\|^2}{2\sigma^2}\right) dx \\
&= \sigma^2\mathbb{E}\left\{\frac{\partial}{\partial x_l} g_j(Wx)\right\},
\end{aligned}
$$

obtained using integration by parts, and denoting the differential with respect to x_l by d_l. Applying the chain rule, we derive that

$$\frac{\partial}{\partial x_l}\left\{g_j(Wx)\right\} = \sum_{i=1}^{N} \frac{\partial}{\partial y_i}\left\{g_j(y)\right\} W_{il} = \sum_{i=1}^{N} Z_{ij}W_{il} = (W^*Z)_{lj}.$$

Therefore, we obtain that $\Omega_2 = \sigma^2\,\mathbb{E}(WW^*Z) = \sigma^2\,U\,\mathbb{E}(Z)$, which yield expression (3.5). Now, to complete the proof of the theorem, recall that $Tr(UU^-) = n$ by (3.3), and observe that

$$
\begin{aligned}
\mathbb{E}\|\hat{f} - f\|^2 &= \mathbb{E}Tr\left\{(\hat{\theta} - \theta)(\hat{\theta} - \theta)^*(W^+)^*W^+\right\} \\
&= \sigma^2 Tr(UU^-) + \mathbb{E}Tr\left\{g(y)g^*(y)U^- + 2\sigma^2 U Z U^-\right\}. \quad\square
\end{aligned}
$$

We emphasize that Theorem 2 generalizes the classical Stein's results presented in [11] to the case of general frames, which includes any full-rank linear expansion of a data vector in a finite-dimensional setting. It also recovers results of [1,9] in the image processing literature. Moreover, Theorem 2 applies to any arbitrary de-noising strategy which satisfies assumptions of the theorem. In particular, it applies to various types of thresholding or shrinkage procedures for construction of unbiased estimators of the risk, as it is shown by the following four examples.

Example 1. Consider the case of linear shrinkage procedure $\hat{\theta} = \Gamma y = I_N y + (\Gamma - I_N)y$ where Γ is diagonal and independent of y. Then, Theorem 2 with $g(y) = (\Gamma - I_N)y$ and $Z = \Gamma - I_N$ recovers expression (3.1) for Δ.

Example 2. Consider the case of "keep-or-kill" hard thresholding procedure $\hat{\theta} = \Gamma y = I_N y + (\Gamma - I_N)y$ where $\Gamma = diag(\gamma)$ and $\gamma \in \{0, 1\}^N$ independent on data. Then, Theorem 2 with $g(y) = (\Gamma - I_N)y$ and $Z = \Gamma - I_N$ recovers expression (3.2) for Δ.

Example 3. In the case of a hard thresholding procedure with variable threshold t (where the decision to "keep" or "kill" a coefficient depends upon the coefficient itself), $\hat{\theta} = y - y\mathbb{I}(|y_i| < t)$. Hence, in Theorem 2, $g(y) = -y\mathbb{I}(|y_i| < t)$ and Z is diagonal with elements $Z_{ii} = -\mathbb{I}(|y_i| < t)$. Thus, Theorem 2 yields the SURE for $\mathbb{E}\|\hat{f} - f\|^2$ of the form (3.1) with

$$\Delta = \sum_{i,j=1}^{N} \left\{ y_i y_j U_{ij}^- \mathbb{I}(|y_i| < t)\mathbb{I}(|y_j| < t) - 2\sigma^2 (U^- U)_{ii}\, \mathbb{I}(i = j)\mathbb{I}(|y_i| < t) \right\}. \tag{3.6}$$

Note that (3.6) generalizes the hard thresholding rule from orthonormal bases to frames. Indeed, in the former case, (3.6) recovers the familiar expression

$$\Delta = \left\{ \sum_{i=1}^{n} y_i^2 \mathbb{I}(|y_i| < t) - 2\sigma^2 \sum_{i=1}^{n} \mathbb{I}(|y_i| < t) \right\}. \tag{3.7}$$

Example 4. In the case of the soft thresholding with a variable threshold t, one has $\hat{\theta}_i = \{y_i - sign(y_i)\, t\}\, \mathbb{I}(|y_i| > t)$, so that $g_i(y)$ is of the form

$$g_i(y) = -sign(y_i)\min(|y_i|, t) \quad i = 1, \ldots, N. \tag{3.8}$$

Hence, Z in Theorem 2 is diagonal, with elements $Z_{ii} = -\mathbb{I}(|y_i| < t)$, so that Theorem 2 yields the SURE for $\mathbb{E}\|\hat{f} - f\|^2$ of the form (3.1) with

$$\Delta = \sum_{i,j=1}^{N} \left\{ sign(y_i y_j)\min(|y_i|, t)\min(|y_j|, t)U_{ij}^- - 2\sigma^2 (U^- U)_{ii}\, \mathbb{I}(i = j)\mathbb{I}(|y_i| < t) \right\}. \tag{3.9}$$

Note that (3.9) generalizes the soft thresholding rule from orthonormal bases to frames. Indeed, in the former case, (3.9) recovers the familiar risk reported in [5]

$$\Delta = \left\{ \sum_{i=1}^{n} \min(y_i^2, t^2) - 2\sigma^2 \sum_{i=1}^{n} \mathbb{I}(|y_i| < t) \right\}. \tag{3.10}$$

4. Optimal strategies for finite frames

In this section, we use results obtained in the previous sections to obtain optimal frame based estimators of some specific forms. In particular, we construct an optimal linear shrinkage estimator, an optimal hard thresholding and an optimal soft thresholding estimator in the case of a general frame. These choices are motivated, on one hand, by the actual possibility of minimizing the SURE obtained in the previous section, on the other hand, by our intention to compare them with their classical counterparts adapted from the setting with an orthonormal basis. Indeed, the numerical simulations below show the advantage of taking into account the frame correlation structure while minimizing the error estimator. However, we would like to point out that Theorem 2, in principle, offers a valid instrument for obtaining an optimal estimator for any $\hat{\theta} = y + g(y)$ which satisfies the assumptions.

4.1. The linear shrinkage

We begin with the SURE strategy for linear shrinkage in the frame setting. By Corollary 2, one has $\mathbb{E}\|\hat{f} - f\|^2 = \sigma^2 n + \mathbb{E}\Delta$ with $\Delta = Tr\{U^-(I_N - \Gamma)yy^*(I_N - \Gamma) + 2\sigma^2 U^- U\Gamma - 2\sigma^2 U^- U\}$. Writing $\hat{\theta} = \gamma \circ y$, we recognize that, similarly to Corollary 1, the problem of finding Γ reduces to the minimization of a quadratic form (2.6) where, now, $A = (yy^*) \circ U^-$ and $b = \{A - \sigma^2(U \circ U^-)\}1_N$, so that

$$\gamma = A^{-1}b = \left\{(yy^*) \circ U^-\right\}^{-1}\left\{(yy^*) \circ U^- - \sigma^2(U \circ U^-)\right\}1_N. \tag{4.1}$$

Since matrices U, U^- and yy^* are nonnegative definite and Hermitian, matrix A is also nonnegative definite and Hermitian, and, thus, the minimum exists and it is unique. Furthermore, note that matrix A and vector b are often sparse. For example, in the case of a sparse tight frame, expressions for A and b take the forms $A = \alpha^{-2}(yy^*) \circ U$ and $b = (A - \sigma^2 U \circ U)1_N$. Since the majority of entries of matrix U are equal to zero, respective entries of matrix A also vanish.

Note that the objective function in (2.6) can be modified and improved in a variety of ways. For example, adding a quadratic penalty term $\gamma^* P \gamma$ with any positive-definite matrix P leads to the Tikhonov regularization while adding a penalty of the form $\beta \|\gamma\|_{\ell_p}$, where $\|\cdot\|_{\ell_p}$ is a vector norm in ℓ_p space, induces sparsity whenever $0 \le p \le 1$. We refer to [3] for the discussion on how the estimating procedure above can be improved and specialized.

4.2. The hard thresholding

In the case of a hard thresholding rule, our SURE strategy again follows Corollary 2 and Example 3 with Δ given by (3.6). To minimize the resulting expression, we introduce a matrix H with components

$$H_{ij} = \begin{cases} y_i y_j U_{ij}^- & \text{if } i \ne j, \\ y_i^2 U_{ii}^- - 2\sigma^2 (U^- U)_{ii} & \text{if } i = j. \end{cases}$$

Consider a set of indices J such that $j \in J$ if $\gamma_j = 0$ and $j \notin J$ otherwise. Then Δ can be re-written as

$$\Delta = \sum_{i,j \in J} H_{ij}$$

and the goal is to find a set of indices J such that the sum of respective row and column elements of matrix H is minimal. This minimizations can be accomplished by a kind of a greedy algorithm which can be carried out as follows.

The greedy algorithm

1. Since the diagonal values of matrix H are counted once while all other elements are counted twice, introduce modified matrix $\tilde{H}$ with elements

$$\tilde{H}_{ij} = \begin{cases} H_{ij}, & \text{if } i \ne j, \\ H_{ij}/2, & \text{if } i = j. \end{cases}$$

   Set $J = \{1, \dots, N\}$.
2. Find a column l of $\tilde{H}$ with the maximum sum of elements.
3. If the sum of elements of column l is positive, then eliminate column l and row l from $\tilde{H}$ and index l from set J, and RETURN TO STEP 2. If the sum of elements of column l is zero or negative, then FINISH.
4. Set $\gamma_j = 0$ if $j \in J$ and $\gamma_j = 1$ if $j \notin J$.

4.3. The soft thresholding

In the case of the soft thresholding, our SURE strategy follows Example 4 with Δ given by (3.9) and threshold chosen as the minimum of

$$\sum_{i,j=1}^{N} \left\{ sign(y_i y_j) \min(|y_i|, t) \min(|y_j|, t) U_{ij}^- - 2\sigma^2 (U^- U)_{ii} \mathbb{I}(i = j)\mathbb{I}(|y_i| < t) \right\}. \tag{4.2}$$
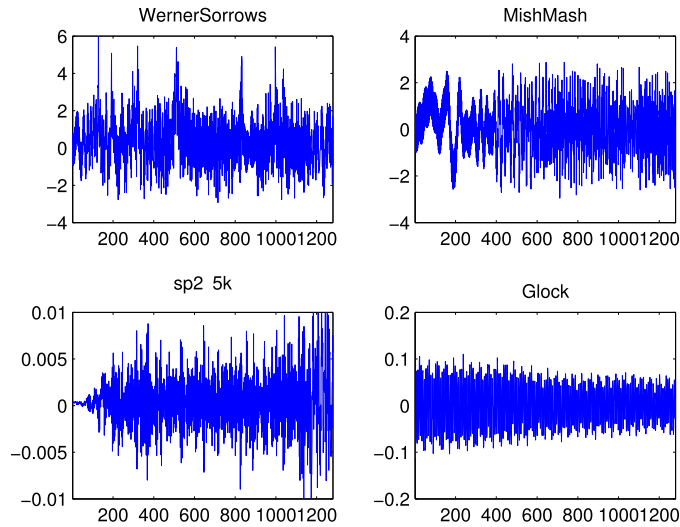
Fig. 1. Normalized test signals of length 1280.

This optimization problem can be solved in $\mathcal{O}(N \log N)$ steps. Indeed, if $|y_i|$ are arranged in an increasing order, then the rule (4.2) implies that for the values of t which lie between two consecutive values of $|y_i|$ SURE is strictly increasing, see [5]. Therefore, t should coincide with one of the values $|y_i|$.

5. Simulation study

In this section, we carry out some numerical experiments to study the finite sample performances of the proposed estimators. In our simulation study, we use the classical Gabor frame with the Hamming window. This is a tight frame which is particularly suitable for representation of fast oscillating signals such as audio signals. For that reason, we consider two fast oscillating standard test signals, *WernerSorrows* and *Mishmash*, reproducible by MakeSignal of the toolbox Wavelab, and two pieces of real audio signals *sp2-5k.wav* and *Glock.wav*. The test signals listed above are displayed in Fig. 1.

The objective of this simulation study is to illustrate the gain in de-noising precision obtained by taking into the account the frame structure rather than to be an exhaustive study of signal de-noising by frames which is usually addressed by application-specific schemes. Results of all comparisons are represented in terms of the means and the standard deviations of $\|\hat{f} - f\|_2$. In order to show the advantage attained by accounting for the frame structure, we compare the ideal best diagonal estimator obtained minimizing the true risk in (2.5) with the ideal best diagonal estimator obtained by minimizing the true risk without taking into account the frame structure, i.e., considering $U = I$. We denote these two estimators $WIENER_U$ and $WIENER_I$, respectively. Note that estimators $WIENER_U$ and $WIENER_I$ are not available in practice, but their comparison can give an idea of the **best** possible gain obtained by taking into account the frame structure. The empirical versions of these estimators, to which we refer as EMP_U and EMP_I, respectively, are derived by substituting $\theta\theta^*$ with its unbiased estimator $yy^* - \sigma^2 U$ for EMP_U, and $yy^* - \sigma^2 I_N$ for EMP_I. Note that both estimators are of the form $\gamma \circ y$ where γ is given by expression (4.1) for EMP_U, while γ reduces to the well known empirical Wiener estimator $\gamma_i = (y_i^2 - \sigma^2)/y_i^2$ for EMP_I. Since matrices $(yy^T \circ U)$ and $(yy^T \circ I_N)$ have high condition numbers, in order to stabilize their inversion, in our simulation study, we add a quadratic penalty term $\gamma^t P \gamma$ with matrix $P = \zeta I_N$ to the objective function (2.6).

Results for $\zeta = 10^{-4.5}$ are reported in Table 1 and are based on 100 simulation runs with the signal-to-noise ratios (SNR) 1, 3 and 5, that represent, respectively, the high, the moderate and the low noise levels. As it is standard in the statistical literature, the signal-to-noise ratio (SNR) is defined as the ratio of the standard deviations of the signal and the noise. The empirical estimators EMP_U and EMP_I approximate the corresponding ideal estimators $WIENER_U$ and $WIENER_I$ when the noise level is low (SNR = 5) and can be quite far from them when the noise level is high (SNR = 1). However, for all the test signals, the ideal gain (the difference between the first and the second columns) and the empirical gain (the difference between the third and the fourth columns) obtained by accounting for the frame structure is quite significant, especially, in the case of high noise.

Table 1
Results obtained over 100 runs and with parameter choices, $n = 1280$ and 64-sampled Hamming window.

	$WIENER_U$	$WIENER_I$	EMP_U	EMP_I
WERNERSORROWS				
SNR = 1	0.1327 (0.0096)	0.2274 (0.0116)	0.4964 (0.0240)	5.7420 (0.1847)
SNR = 3	0.0284 (0.0019)	0.0404 (0.0022)	0.0777 (0.0034)	0.1343 (0.0049)
SNR = 5	0.0126 (0.0006)	0.0167 (0.0007)	0.0321 (0.0011)	0.0412 (0.0015)
MISHMASH				
SNR = 1	0.1026 (0.0106)	0.1837 (0.0139)	0.4881 (0.0254)	6.2411 (0.2290)
SNR = 3	0.0211 (0.0017)	0.0284 (0.0021)	0.0752 (0.0036)	0.1113 (0.0044)
SNR = 5	0.0094 (0.0007)	0.0122 (0.0008)	0.0324 (0.0013)	0.0286 (0.0015)
SP2-5K				
SNR = 1	0.1533 (0.0112)	0.2474 (0.0127)	0.5201 (0.0254)	6.2648 (0.1745)
SNR = 3	0.0363 (0.0022)	0.0548 (0.0024)	0.0849 (0.0039)	0.1771 (0.0058)
SNR = 5	0.0168 (0.0009)	0.0244 (0.0011)	0.0349 (0.0014)	0.0614 (0.0022)
GLOCK				
SNR = 1	0.0845 (0.0075)	0.1305 (0.0093)	0.4529 (0.0245)	6.4889 (0.2079)
SNR = 3	0.0192 (0.0014)	0.0278 (0.0016)	0.0737 (0.0037)	0.1232 (0.0043)
SNR = 5	0.0089 (0.0006)	0.0123 (0.0007)	0.0322 (0.0013)	0.0326 (0.0015)

Table 2
Results obtained over 100 runs and with parameter choices, $n = 1280$ and 64-sampled Hamming window.

	$SOFT_U$	$SOFT_I$	$HARD_U$	$HARD_I$
WERNERSORROWS				
SNR = 1	0.3748 (0.0188)	0.8511 (0.0414)	0.4939 (0.0277)	0.9024 (0.0200)
SNR = 3	0.0763 (0.0041)	0.1342 (0.0114)	0.0737 (0.0037)	0.3965 (0.3965)
SNR = 5	0.0327 (0.0016)	0.0481 (0.0039)	0.0298 (0.0014)	0.1275 (0.0041)
MISHMASH				
SNR = 1	0.3519 (0.0216)	0.8970 (0.0695)	0.4884 (0.0312)	0.9756 (0.0158)
SNR = 3	0.0602 (0.0040)	0.1063 (0.0095)	0.0659 (0.0034)	0.2573 (0.0160)
SNR = 5	0.0251 (0.0013)	0.0414 (0.0035)	0.0266 (0.0015)	0.0786 (0.0039)
SP2-5K				
SNR = 1	0.3893 (0.0197)	0.8555 (0.0693)	0.5204 (0.0281)	0.9952 (0.0105)
SNR = 3	0.0917 (0.0047)	0.1689 (0.0164)	0.0842 (0.0036)	0.4023 (0.0142)
SNR = 5	0.0457 (0.0026)	0.0626 (0.0059)	0.0353 (0.0015)	0.1799 (0.0050)
GLOCK				
SNR = 1	0.2853 (0.0186)	0.5064 (0.0462)	0.4523 (0.0285)	0.9350 (0.0462)
SNR = 3	0.0516 (0.0029)	0.0981 (0.0083)	0.0656 (0.0037)	0.1628 (0.0078)
SNR = 5	0.0228 (0.0012)	0.0406 (0.0034)	0.0273 (0.0015)	0.0919 (0.0026)

In a similar manner, we carry out comparisons between the soft thresholding procedures obtained with and without consideration of the specific frame structure. In particular, we construct estimators $SOFT_U$ and $SOFT_I$ which are provided by the formula $\hat{\theta}_i = (y_i - sign(y_i)t)\mathbb{I}(|y_i| > t)$ with the global threshold t obtained by minimizing, respectively, expression (3.9) derived above and the classical one (3.10) as in [5]. Moreover, we compare estimators $HARD_U$ and $HARD_I$ obtained using the hard thresholding scheme presented in Section 4.2 and the universal hard thresholding estimator defined as $\hat{\theta}_i = y_i\mathbb{I}(|y_i| > \sigma\sqrt{2\log N})$. We emphasize that this last estimator represents both the classical universal thresholding (when $U = I$) as well as an asymptotically optimal estimator when U is a stable frame (as the one adopted here) according to results presented in [7].

Results of comparisons are reported in Table 2 and are based on 100 simulation runs. As in the previous study, results show a very significant gain when taking frame structure into account. Note that, performance of the estimator $HARD_I$ is very poor, but this is not surprising since the universal threshold $\sigma\sqrt{2\log N}$ is known to be too large for de-noising applications when the true signal is not sufficiently sparse (see, e.g., [5]).
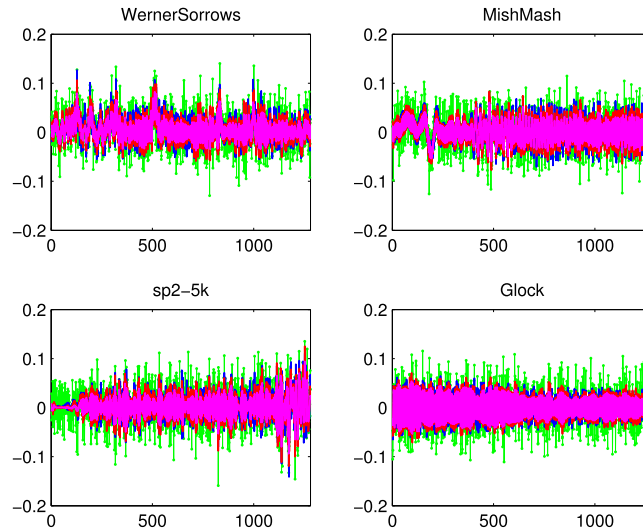
Fig. 2. Data (green dotted line); true signals (blue); *WIENER_U* estimators (red); *WIENER_I* estimators (magenta). SNR = 1. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)
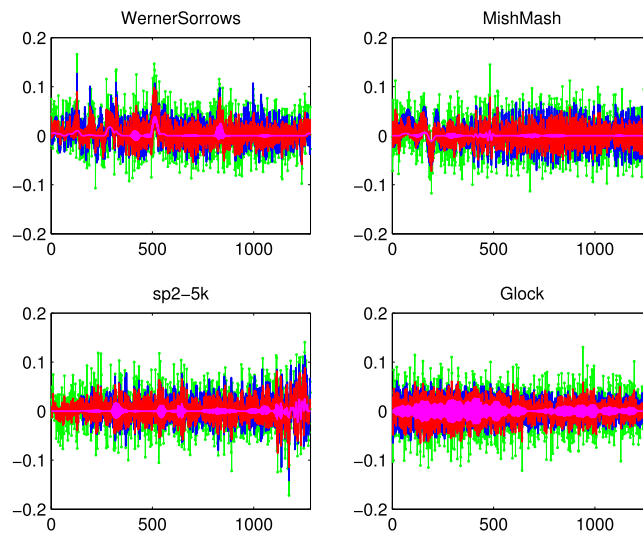

Fig. 3. Data (green dotted line); true signals (blue); *SOFT_U* estimators (red); *SOFT_I* estimators (magenta). SNR = 1. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

For completeness, one of the 100 estimators obtained by *WIENER_U* and *WIENER_I* are shown in Fig. 2 and one obtained by *SOFT_U* and *SOFT_I* are shown in Fig. 3, both in the case of SNR = 1.

Acknowledgments

References

[1] T. Blu, F. Luisier, The sure-let approach to image denoising, IEEE Trans. Image Process. 16 (3) (2007) 2778–2786.
[2] T. Cai, Adaptive wavelet estimation: A block thresholding and oracle inequality approach, Ann. Statist. 27 (1999) 898–924.
[3] D. De Canditiis, A frame based shrinkage procedure for fast oscillating functions, Comput. Statist. Data Anal. 75 (2014) 142–150.

[4] D. De Canditiis, B. Vidakovic, Wavelet Bayesian block shrinkage via mixtures of normal-inverse gamma priors, J. Comput. Graph. Statist. 13 (2004) 383–398.

[5] D. Donoho, I. Johnstone, Adapting to unknown smoothness via wavelet shrinkage, J. Amer. Statist. Assoc. 90 (1995) 1200–1224.

[6] T. Downie, B. Silverman, The discrete multiple wavelet transform and thresholding methods, IEEE Trans. Signal Process. 46 (1998) 2558–2561.

[7] M. Haltmeier, A. Munk, Extreme value analysis of frame coefficients and implications for image denoising, Appl. Comput. Harmon. Anal. 36 (3) (2014) 434–460.

[8] N. Nezamoddini-Kachouie, P. Fieguth, A gabor based technique for image denoising, IEEE Electr. Comput. Eng. 36 (3) (2005) 980–983.

[9] M. Raphan, E.P. Simoncelli, Optimal denoising in redundant representations, IEEE Trans. Image Process. 17 (2008) 1342–1352.

[10] D. Rudoy, P. Basu, P.J. Wolfe, Superposition frames for adaptive time-frequency analysis and fast reconstruction, IEEE Trans. Signal Process. 58 (2010) 2581–2596.

[11] C.M. Stein, Estimation of the mean of a multivariate normal distribution, Ann. Statist. 9 (6) (1981) 1135–1151.

[12] J.S. Walker, Y.J. Chen, Denoising gabor transforms, Preprint http://people.uwec.edu/walkerjs/media/dgt.pdf.

[13] P.J. Wolfe, S. Godsill, W.J. Ng, Bayesian variable selection and regularization for time-frequency surface estimation, J. R. Stat. Soc. Ser. B Stat. Methodol. 66 (6) (2004) 575–589.

[14] G. Yu, S. Mallat, E. Bacry, Audio denoising by time-frequency block thresholding, IEEE Trans. Signal Process. 56 (2008) 1830–1839.