Data-driven Thresholding in Denoising with Spectral Graph Wavelet Transform

Basile de Loynes, Fabien Navarro, Baptiste Olivier

Data-driven Thresholding in Denoising with Spectral Graph Wavelet Transform

Basile de Loynes*, Fabien Navarro†, Baptiste Oliver‡

June 18, 2019

Abstract

This paper is devoted to adaptive signal denoising in the context of Graph Signal Processing (GSP) using Spectral Graph Wavelet Transform (SGWT). This issue is addressed *via* a data-driven thresholding process in the transformed domain by optimizing the parameters in the sense of the Mean Square Error (MSE) using the Stein's Unbiased Risk Estimator (SURE). The SGWT considered is built upon a partition of unity making the transform semi-orthogonal so that the optimization can be performed in the transformed domain. However, since the SGWT is over-complete, the SURE needs to be adapted to the context of correlated noise. Two thresholding strategies called coordinatewise and block thresholding process are investigated. For each of them, the SURE is derived for a whole family of elementary thresholding functions among which the soft threshold and the James-Stein threshold. To provide a fully data-driven method, a noise variance estimator derived from the Von Neumann estimator in the Gaussian model is adapted to the graph setting.

1 Introduction

The emerging field of Graph Signal Processing (GSP) aims to bridge the gap between signal processing and spectral graph theory (see for instance Chung and Graham (1997); Belkin and Niyogi (2008) and references therein). One objective is to generalize fundamental analysis operations from regular grid signals to irregular structures as graphs. An extensive literature on GSP exists, in particular we refer the reader to Shuman et al. (2013) for an introduction to this field and Ortega et al. (2018) for an overview of recent developments, challenges and applications. In this context, the authors of Coifman and Lafon (2006); Gavish et al. (2010); Hammond et al. (2011) have developed wavelet transforms on graphs. More specifically, in Hammond et al. (2011) a fairly general construction of a frame enjoying the usual properties of standard wavelets is developed: each vector of the frame is localized both in the graph domain and the spectral domain. The transform associated with this frame is named Spectral Graph Wavelet Transform (SGWT). Many studies based on SGWT (or some variants) explore the denoising performance of this approach using different strategies Onuki et al. (2016); Deutsch et al. (2016); Irion and Saito (2017); Dong et al. (2016); Göbel et al. (2018); Leonardi and Van De Ville (2013); Onuki et al. (2016) from signal adapted tight frames to regularization method.

The denoising approach chosen in this paper involves several thresholding processes in the transformed domain of the wavelet coefficients. Actually, this can be seen as an extension to

*Basile de Loynes
ENSAI, France, E-mail: basile.deloynes@ensai.fr
†Fabien Navarro
CREST, ENSAI, France, E-mail: fabien.navarro@ensai.fr
‡Baptiste Oliver
Orange Labs, France. E-mail: baptiste.olivier@orange.fr

SGWT of the methodology of Donoho and Johnstone (1995); Cai (1999). With this approach, the main challenge is the efficient calibration of the parameters minimizing the MSE risk in a complete data-driven way. Recently, in the setting of discrete wavelets transform on a regular grid — the so-called regular case — the Stein's unbiased risk estimate (SURE) has proven to be a powerful tool for signal/image restoration Luisier et al. (2007); Pesquet et al. (2009); Vaiter et al. (2013). Based on the Stein's lemma, this estimator acts as a proxy for the MSE which cannot be computed in practice since the original signal is unknown.

In this paper, the SURE is explicitly computed for a wide family of thresholding processes in Theorem 1. Let us point out that contrary to the regular wavelet transform, the SGWT is no longer orthogonal so that a white Gaussian noise in the graph domain is transformed in a correlated noise. Consequently, the divergence term of the resulting SURE involves the covariance of the transformed noise making the numerical evaluation less simple than in the regular case. Afterward, the SURE is specified to the case of coordinatewise and block thresholding. The latter is inspired by image denoising problems for which a Stein risk estimator has been proposed in Peyré et al. (2011) to tune both the block-sparsity structure and the threshold. A similar selection strategy has been developed by Navarro et al. (2013) in the context of deconvolution.

A key parameter in the calculation of the SURE and more generally in standard denoising methods is the noise level which is unknown for real data. In the regular case of wavelet-based denoising methods, the Median Absolute Deviation (MAD) estimator Donoho and Johnstone (1995) (calculated on the finest decomposition scale) is often the preferred method. However, the MAD is not directly applicable for SGWT since the transformed noise is correlated. We propose here to extend the estimator developed in von Neumann (1941) in the graph setting.

Note that in Onuki et al. (2016) a regularization method is proposed for which the regularization parameter is also selected optimizing an MSE proxy based on a similar argument. Nonetheless, beyond the fact that the philosophy is different (regularization *versus* thresholding), one stress that the empirical risk bias is explicitly determined while the MSE estimation in Onuki et al. (2016) is only validated numerically. Furthermore, our approach makes full use of the multiscale analysis property of the SGWT that can be interpreted as a multiple filter denoising. Note also that the authors of Onuki et al. (2016) do not address the problem of noise variance estimation.

The paper is organized as follows. Section 2 introduces the notation and briefly reviews the notions of tight frame and SGWT of (Hammond et al., 2011). Section 3 is devoted to denoising and the SURE estimator for generic thresholding process in the context of correlated work is derived. Then, the SURE is specified in the cases of coordinatewise and block thresholding processes. Finally, an estimator of the noise variance along with its performance is presented. Section 4 discusses numerical results.

2 Spectral Graph Wavelet Transform

2.1 Graphs, Frames and Tight Frames

Let G be an undirected weighted graph, with set of vertices V, and weights $(w_{ij})_{i,j\in V}$ satisfying $w_{ij} = w_{ji}$ for $i, j \in V$. The size of the graph is the number of nodes $n = |V|$. The (unnormalized) graph Laplacian matrix $\mathcal{L} \in \mathbb{R}^{V\times V}$ associated with G is the symmetric matrix defined as $\mathcal{L} = D - W$, where W is the matrix of weights with coefficients $(w_{ij})_{i,j\in V}$, and D the diagonal matrix with diagonal coefficients $D_{ii} = \sum_{j\in V} w_{ij}$. A signal f on the graph G is a function $f : V \to \mathbb{R}$.

Let $\mathfrak{F} = \{r_i\}_{i\in I}$ be a frame of vectors of $\mathbb{R}^V$, that is a family of vectors in $\mathbb{R}^V$ such that there exist $A, B > 0$ satisfying for all $f \in \mathbb{R}^V$

$$A\|f\|_2^2 \leq \sum_{i\in I} |\langle f, r_i\rangle|^2 \leq B\|f\|_2^2. \tag{1}$$

The linear map $T_{\mathfrak{F}} : \mathbb{R}^V \to \mathbb{R}^I$ defined for $f \in \mathbb{R}^V$ by $T_{\mathfrak{F}} f = (\langle f, r_i \rangle)_{i \in I}$ is called the *analysis* operator. The *synthesis* operator is the adjoint of $T_{\mathfrak{F}}$: namely, it is the linear map $T^*_{\mathfrak{F}} : \mathbb{R}^I \to \mathbb{R}^V$ defined for a vector of coefficients $(c_i)_{i \in I}$ by $T^*_{\mathfrak{F}}(c_i)_{i \in I} = \sum_{i \in I} c_i r_i$. As a frame is in particular a generating family of $\mathbb{R}^V$, a signal $f \in \mathbb{R}^V$ can be recovered from its coefficients $T_{\mathfrak{F}} f$ with the help of the synthesis operator.

2.2 Construction of Tight Frames

A frame $\mathfrak{F}$ is said to be tight if $A = B = 1$ in Equation (1) — the latter is then termed the Parseval identity. From now on, the frames considered are supposed to be tight. Let us recall the generic construction of such a frame (*c.f.* Kereta et al. (2019) for instance).

Since $\mathcal{L}$ is self-adjoint, it admits the spectral decomposition

$$\mathcal{L} = \sum_{\ell} \lambda_\ell \langle \chi_\ell, \cdot \rangle \chi_\ell,$$

where $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n = 0$ denote the (ordered) eigenvalues of the matrix $\mathcal{L}$, and $(\chi_\ell)_{1 \leq \ell \leq n}$ are the associated normalized and pairwise orthogonal eigenvectors. Then, for any function $\rho : \mathrm{sp}(\mathcal{L}) \to \mathbb{R}$ defined on the spectrum $\mathrm{sp}(\mathcal{L})$ of matrix $\mathcal{L}$, the functional calculus formula reads

$$\rho(\mathcal{L}) = \sum_{\ell} \rho(\lambda_\ell) \langle \chi_\ell, \cdot \rangle \chi_\ell.$$

A finite collection $(\psi_j)_{j=0,\dots,J}$ is a finite partition of unity on the compact $[0, \lambda_1]$ if

$$\psi_j : [0, \lambda_1] \to [0, 1] \text{ for all } j \in \mathcal{J} \text{ and } \forall \lambda \in [0, \lambda_1], \ \sum_{j=0}^{J} \psi_j(\lambda) = 1.$$

Given a finite partition of unity $(\psi_j)_{j=0,\dots,J}$, the Parseval identity implies that the following set of vectors is a tight frame:

$$\mathfrak{F} = \left\{ \sqrt{\psi_j}(\mathcal{L}) \delta_i, j = 0, \dots, J, i \in V \right\}.$$

Also, following Leonardi and Van De Ville (2013); Göbel et al. (2018), a partition of unity can be easily defined as follows: let $\omega : \mathbb{R}^+ \to [0, 1]$ be some function with support in $[0, 1]$, satisfying $\omega \equiv 1$ on $[0, b^{-1}]$ and set

$$\psi_0(x) = \omega(x) \ \text{ and } \ \psi_j(x) = \omega(b^{-j} x) - \omega(b^{-j+1} x) \ \text{ for } \ j = 1, \dots, J, \ \text{ where } \ J = \left\lfloor \frac{\log \lambda_1}{\log b} \right\rfloor + 2.$$

2.3 Discrete SGWT Associated with a Partition of Unity

Let $(\psi_j)_{j=0,\dots,J}$ be a partition of unity of $[0, \lambda_1]$. The discrete SGWT of a signal $f \in \mathbb{R}^V$ is defined as follows:

$$\mathcal{W} f = \left(\sqrt{\psi_0}(\mathcal{L}) f^T, \sqrt{\psi_1}(\mathcal{L}) f^T, \dots, \sqrt{\psi_J}(\mathcal{L}) f^T \right)^T \in \mathbb{R}^{n(J+1)}.$$

The adjoint linear transformation $\mathcal{W}^*$ of $\mathcal{W}$ is:

$$\mathcal{W}^* \left(\eta_0^T, \eta_1^T, \dots, \eta_J^T \right)^T = \sum_{j \geq 0} \sqrt{\psi_j}(\mathcal{L}) \eta_j.$$

The tightness of the underlying frame implies that $\mathcal{W}^* \mathcal{W} = \mathrm{Id}_{\mathbb{R}^V}$ so that a signal $f \in \mathbb{R}^V$ can be recovered by applying $\mathcal{W}^*$ to its wavelet coefficients $((\mathcal{W} f)_i)_{i=1,\dots,n(J+1)} \in \mathbb{R}^{n(J+1)}$ (see Hammond et al. (2011) for the details).

3 Adaptive Denoising with SGWT

Let $f \in \mathbb{R}^V$ be some signal on a graph G and ξ be an n-dimensional Gaussian vector distributed as $\mathcal{N}(0, \sigma^2 \mathrm{Id})$. The aim of denoising is to recover the unknown signal f from the observed noisy version $\tilde{f} = f + \xi$.

Basically, denoising consists of three steps: (1) compute the SWGT transform $\mathcal{W}\tilde{f} \in \mathbb{R}^{n(J+1)}$; (2) apply a given thresholding operator $h(\cdot)$ to the coefficients $\mathcal{W}\tilde{f}$; (3) apply the inverse SGWT transform to obtain an estimation $\hat{f}$ of the original signal.

The main challenge in denoising consists in choosing a suitable thresholding operator with respect to the noisy signal $\tilde{f}$ and the underlying graph. The performance measures in the sequel will be the MSE between the original signal f and the denoised signal $\hat{f}$: $\|f - \hat{f}\|_2^2$. First, it is worth noting that the Parseval identity allows direct optimization in the transformed domain of wavelet coefficients. Secondly, in practice, obviously the original signal remains unknown. To overcome this difficulty, the MSE is generally substituted with the Stein's Unbiased Risk Estimator which no longer depends on the original signals (see Donoho and Johnstone (1995) for instance). Nonetheless, contrary to the usual wavelet transform, the white noise ξ is mapped onto a correlated Gaussian noise. In the next section, the SURE is derived taking into account these correlations.

3.1 The SURE Estimator for Correlated Noise

By linearity, the denoising problem $\widetilde{f} = f + \xi$ is transferred to the denoising problem $\widetilde{F} = F + \Xi$ with $\Xi \sim \mathcal{N}(0, \sigma^2 \mathcal{W}\mathcal{W}^*)$, $\widetilde{F} = \mathcal{W}\widetilde{f}$ and $F = \mathcal{W}f$. The spectral decomposition of $\mathcal{W}\mathcal{W}^*$ reads $\mathcal{W}\mathcal{W}^* = U\Sigma U^*$ with U a unitary matrix of $\mathbb{R}^{n(J+1)}$ and $\Sigma = \left(\begin{smallmatrix} \mathrm{Id}_{\mathbb{R}^n} & 0 \\ 0 & 0 \end{smallmatrix}\right)$.

A thresholding process is a map $h : \mathbb{R}^{n(J+1)} \to \mathbb{R}^{n(J+1)}$. Typically, the map h is a coordinate-wise or a block shrinkage in applications. The following result extending the SURE's expression to correlated noise is based on the Stein's lemma in Stein (1981) in which h is assumed to be weakly differentiable. One refer the reader to Stein (1981) for the precise definition.

Theorem 1 (h-SURE)**.** *Let h be a weakly differentiable thresholding process for the denoising problem $\widetilde{F} = F + \Xi$. Then the theoretical MSE is given by*

$$\mathbf{E}\|h(\widetilde{F}) - F\|^2 = \mathbf{E}\left[-n\sigma^2 + \|h(\widetilde{F}) - \widetilde{F}\|^2 + 2\sum_{i,j=1}^{n(J+1)} \mathbf{Cov}(\Xi_i, \Xi_j)\partial_j h_i(\widetilde{F})\right],$$

where h_i is the i-th component of h.

It is worth noting that $\mathbf{Cov}(\Xi_i, \Xi_j) = \sigma^2(\mathcal{W}\mathcal{W}^*)_{i,j}$ so that, as soon as the thresholding process h is specified and the noise variance σ^2 estimated, the SURE of h defined below can be completely computed from the noisy observations as in the regular case:

$$\mathbf{SURE}(h) = -n\sigma^2 + \|h(\widetilde{F}) - \widetilde{F}\|^2 + 2\sum_{i,j=1}^{n(J+1)} \mathbf{Cov}(\Xi_i, \Xi_j)\partial_j h_i(\widetilde{F}).$$

Proof. The theoretical MSE can be rewritten as follows

$$\mathbf{E}\|h(\widetilde{F}) - F\|^2 = \mathbf{E}\|h(\widetilde{F}) - \widetilde{F}\|^2 + \mathbf{E}\|\Xi\|^2 + 2\mathbf{E}\langle h(\widetilde{F}) - \widetilde{F}, \Xi\rangle.$$

The second term is equal to $n\sigma^2$ since almost surely $\|\Xi\|^2 = \|U^*\Xi\|^2 = \|P_K U^*\Xi\|^2$ where $K = \ker(\mathcal{W}\mathcal{W}^*)^\perp$ and P_K the orthogonal projection onto K. Finally, setting $g(x) = h(x) - x$, $x \in \mathbb{R}^{n(J+1)}$, it remains to compute the last term $\mathbf{E}\langle g(\widetilde{F}), \Xi\rangle = \mathbf{E}\langle g(F + \Xi), \Xi\rangle$ where F is deterministic and $\Xi \sim \mathcal{N}(0, \sigma^2 \mathcal{W}\mathcal{W}^*)$. A simple computation gives

$$\mathbf{E}\langle g(F + \Xi), \Xi\rangle = \sum_{i=1}^{n(J+1)} \mathbf{E}[g_i(F + \Xi)\Xi_i] = \sum_{i=1}^{n(J+1)} \mathbf{Cov}(g_i(F + \Xi), \Xi_i).$$

Then, following Liu (1994), each term in the sum above are given by

$$\mathbf{Cov}(g_i(F+\Xi),\Xi_i) = \sum_{j=1}^{n(J+1)} \mathbf{Cov}(\Xi_i,\Xi_j)\mathbf{E}[(\partial_j h_i)(F+\Xi)] - n\sigma^2,$$

since $\mathrm{Tr}(\sigma^2\mathcal{W}\mathcal{W}^*) = n\sigma^2$. This ends the proof of Theorem 1. □

3.2 Coordinatewise Thresholding Process

For a coordinatewise thresholding process, the map h is of the form $h(x) = (\tau(x_i,t_i))_{i=1,\dots,n(J+1)}$ where $(t_i)_{i=1,\dots,n(J+1)}$ are the thresholds. In practice, we may choose $\tau(x,t) = x\max\{1 - t^\beta|x|^{-\beta}, 0\}$ with $\beta \geq 1$. The most popular choices are the soft thresholding ($\beta = 1$), the James-Stein thresholding ($\beta = 2$) and the hard thresholding ($\beta = \infty$). The latter will not be considered here since it does not lead to a sufficiently regular thresholding process for Theorem 1 to be applied.

For any $\beta \in [1,\infty)$, the derivative $\partial_j h_i$ vanishes whereas

$$\partial_i h_i(F+\Xi) = \mathbf{1}_{[t_i,\infty)}(|\widetilde{F}_i|)\left[1 + (\beta-1)\frac{t^\beta}{|\widetilde{F}_i|^\beta}\right].$$

Consequently, the SURE associated with h is given by

$$\mathbf{SURE}(h) = -n\sigma^2 + \sum_{i=1}^{n(J+1)} \widetilde{F}_i^2\left(1 \wedge \frac{t_i^\beta}{|\widetilde{F}_i|^\beta}\right)^2 + 2\sum_{i=1}^{n(J+1)} \mathbf{V}(\Xi_i)\mathbf{1}_{[t_i,\infty)}(|\widetilde{F}_i|)\left[1 + (\beta-1)\frac{t_i^\beta}{|\widetilde{F}_i|^\beta}\right]. \quad (2)$$

The usual expression of the SURE is recovered from the identity above remarking that $\mathbf{V}[\Xi_i]$ are identically equal to σ^2 when the transformed noise is uncorrelated.

3.3 Optimization: Donoho and Johnstone's Trick

The SURE can be optimized in the same way as in the standard case using the Donoho and Johnstone's trick of Donoho and Johnstone (1995) whose the justification is recalled below.

For the sake of simplicity, we first consider the case of the coordinatewise thresholding process with a uniform threshold: $t_i = t$ for all $i = 1,\dots,n(J+1)$. Denote by $a_1,\dots,a_{n(J+1)}$ the absolute values of the noisy wavelet coefficients $|\widetilde{F}_i|$ in the increasing order. The trick comes from the observation that, on each interval (a_k,a_{k+1}), the last term of Equation (2) is non-decreasing whereas the second term is an increasing function of t. Consequently, the SURE hits its minimum at some value a_{k^*}, $k^* = 1,\dots,n(J+1)$.

If the thresholds t_i are no longer uniform but merely tied inside blocks with values $t_1,\dots,t_L$, the same trick is still valid: group the terms in the sums along the different parameters $t_1,\dots,t_L$ and optimize each partial sum with respect to t_k, $k = 1,\dots,L$.

3.4 Block Thresholding Process

In order to take advantage of the localization properties of SGWT and the regularity of the original signal, we may introduce block thesholding processes similar to Cai (1999).

Consider a partition $(B_\ell)_{\ell\in L}$ of $\{1,\dots,n(J+1)\}$ and set $\|x\|_{B_\ell}^2 = \sum_{i\in B_\ell}(x_i)^2$. In this case, the thesholding process $h = (h_i)_{i=1,\dots,n(J+1)}$ reads

$$h_i(x) = x_i\max\left\{1 - \frac{t_\ell^\beta}{\|x\|_{B_\ell}^\beta}, 0\right\}, \quad x\in\mathbb{R}^{n(J+1)}, \quad \text{and} \quad \ell\in L : i\in B_\ell.$$

If i, j are in different blocks, then $\partial_j h_i$ vanishes. Additionally, if i, j are in B_ℓ but $i \neq j$ then

$$\partial_j h_i(\widetilde{F}) = \widetilde{F}_i \mathbf{1}_{[t_\ell,\infty)}(\|\widetilde{F}\|_{B_\ell}) \beta t_\ell^\beta \widetilde{F}_j \|\widetilde{F}\|_{B_\ell}^{-\beta-2},$$

whereas

$$\partial_i h_i(\widetilde{F}) = \mathbf{1}_{[t_\ell,\infty)}(\|\widetilde{F}\|_{B_\ell}) \left(1 - t_\ell^\beta \|\widetilde{F}\|_{B_\ell}^{-\beta} + \beta t_\ell^\beta \widetilde{F}_i^2 \|\widetilde{F}\|_{B_\ell}^{-\beta-2}\right).$$

Consequently, a straightforward computation leads to

$$\begin{aligned}\mathbf{SURE}(h) = -n\sigma^2 &+ \sum_{\ell \in L} \left(1 \wedge \frac{t_\ell^\beta}{\|\widetilde{F}\|_{B_\ell}^\beta}\right)^2 \|\widetilde{F}\|_{B_\ell}^2 \\ &+ 2 \sum_{\ell \in L} \mathbf{1}_{[t_\ell,\infty)}(\|\widetilde{F}\|_{B_\ell}) \left[\left(1 - \frac{t_\ell^\beta}{\|\widetilde{F}\|_{B_\ell}^\beta}\right) \sum_{i \in B_\ell} \mathbf{V}(\Xi_i) + \frac{\beta t_\ell^\beta}{\|\widetilde{F}\|_{B_\ell}^{\beta+2}} \sum_{i,j \in B_\ell} \mathbf{Cov}(\Xi_i, \Xi_j) \widetilde{F}_i \widetilde{F}_j\right]\end{aligned}$$

Once again, for uncorrelated transformed noise, the usual expression easily follows from the identity above. Note also that the optimization of the SURE in this case requires more sophisticated technics as the divergence term is no longer monotone.

3.5 Estimation of σ^2 Von Neumann's Estimator

Even though the SURE no longer depends on the original signal, it is needed to have an estimate of σ^2. A straightforward computation gives

$$\mathbf{E}[\widetilde{f}^T \mathcal{L} \widetilde{f}] = f^T \mathcal{L} f + \mathbf{E}[\xi^T \mathcal{L} \xi] = f^T \mathcal{L} f + \sigma^2 \mathrm{Tr}\ \mathcal{L}.$$

A biased estimator of σ^2 is given by

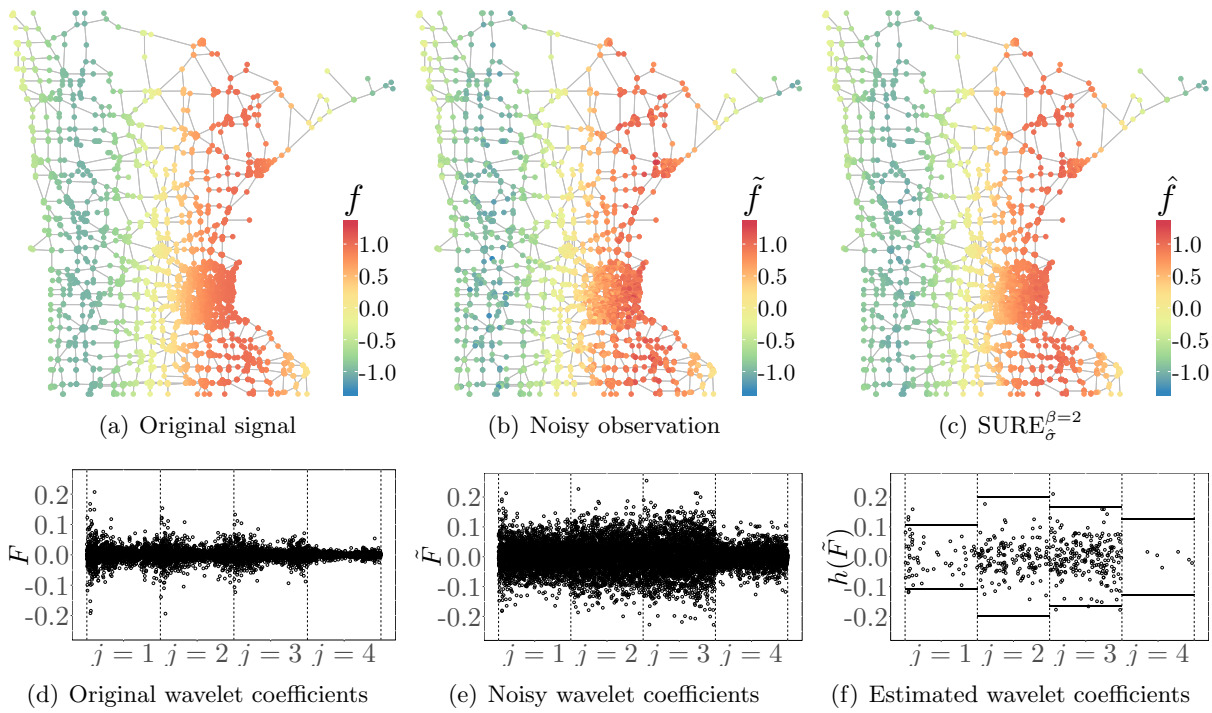
$$\hat{\sigma}^2 = \frac{\widetilde{f}^T \mathcal{L} \widetilde{f}}{\mathrm{Tr}\ \mathcal{L}} = \frac{\sum_{i,j \in V} w_{ij} |\widetilde{f}(i) - \widetilde{f}(j)|^2}{2\ \mathrm{Tr}\ \mathcal{L}}.$$

and is nothing but the graph analogue of the Von Neumann estimator of von Neumann (1941). As soon as the original signal is reasonably smooth so that $f^T \mathcal{L} f$ is negligible compared to $\sigma^2 \mathrm{Tr}\ \mathcal{L}$, then $\hat{\sigma}^2$ is an accurate enough estimation of σ^2.

4 Numerical Results

This section presents the empirical performance of the proposed automatic threshold selection for signals defined on the Minnesota roads graph. The latter is considered as a reference in many recent studies (see Behjat et al. (2016) and references therein). This graph is planar and consists of 2642 vertices and 6606 edges. Each vertex is described by its (x, y)-coordinates. The function ω chosen in the experiments is a piecewise linear function with support in $[0, 1]$ and constant equal to 1 on $[0, b^{-1}]$ with $b = 2$. From $\lambda_1 \approx 6.89$, we deduce the number of scales is $J + 1 = 5$ — see Section 2. For all the results presented below, the sine function is considered: namely, for a vertex $v \in V$ described by its coordinates $(v_x, v_y) \in \mathbb{R}^2$, $f(v) = \sin(v_x)$. The smoothness modulus $f^T \mathcal{L} f$ of f is approximately equal to 12.63 whereas the trace $\mathrm{Tr}\ \mathcal{L} = 6614$.

First, the restoration quality and the efficiency of the proposed method can be assessed visually in Figure 1 which shows the original, the noisy (with $\sigma = 0.1$) and the restored signals at the optimally chosen level-dependent thresholds in the coordinatewise level-dependent setting. The resulting input and output Signal to Noise Ratio (SNR) are respectively $\mathrm{SNR}_{\mathrm{in}} \approx 17.59$ dB and the output $\mathrm{SNR}_{\mathrm{out}} \approx 23.60$ dB whereas $\hat{\sigma} \approx 0.111$. The bottom panel also provides a representation of the coefficients at each scale excluding the coarcest scale for a convenient



(a) Original signal (b) Noisy observation (c) $\text{SURE}_{\hat{\sigma}}^{\beta=2}$

(d) Original wavelet coefficients (e) Noisy wavelet coefficients (f) Estimated wavelet coefficients

Figure 1: Typical reconstruction from a single simulation with $\sigma = 0.1$.

rendering. Each of them is separated by a vertical dashed line. Also, the value of the optimal level-dependent thresholds are represented in Figure 1(f). We observe that the value of the optimal threshold varies from one scale to another, therefore minimizing the SURE scale by scale improves noise reduction performance. We also note that the number of coefficients after thresholding provides a sparse representation (for the four finest scales, only 428 among 10568 are kept).

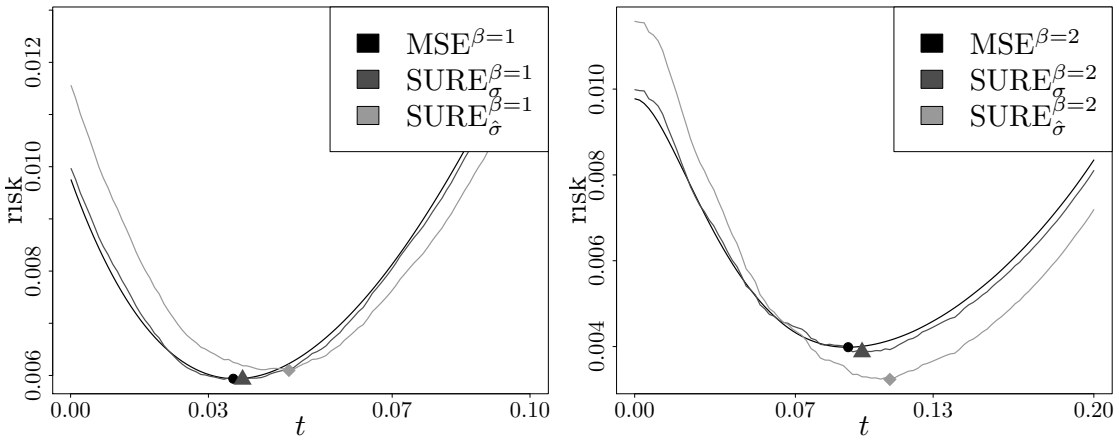
Secondly, we compare the performance in terms of SNR (computed on the functions after reconstruction) for different denoising strategies and different noise levels for the same sine function. For each noise levels $\sigma = 0.025, 0.1, 0.4$, a sample of $N = 25$ white Gaussian noise is simulated and a global (G) *versus* level-dependent (LD) coordinatewise thresholding are performed with respect the soft ($\beta = 1$) and James-Stein ($\beta = 2$) thresholding rules. For each of the four combinations of methodologies, we compare the average behavior of the SNR for parameter selected with the oracle ($\text{MSE}^{\beta=1,2}$ obtained by minimizing the MSE using the original signal f), the SURE with known σ ($\text{SURE}_{\sigma}^{\beta=1,2}$) and the SURE with estimated $\hat{\sigma}$ ($\text{SURE}_{\hat{\sigma}}^{\beta=1,2}$). Also, the standard deviation on the sample is provided. All the measurements are synthetized in Table 1.

Generally speaking, we observe that the level-depend method performs better than a uniform threshold. Similarly to the regular case, numerical experiments shows that the James-Stein threshold ($\beta = 2$) is slightly more efficient than the soft threshold in particular in the global thresholding process. Also, while the SURE with known σ^2 provides a reliable estimate of the MSE for arbitrary noise levels (see Figure 2 for an illustration with $\sigma = 0.1$), the quality of the SURE with unknown noise level depends highly on the estimation of $\hat{\sigma}^2$. This comes from the bias of the estimation. Typically it is required for the smoothness modulus $f^T \mathcal{L} f$ to be negligible compared to $\sigma^2 \text{Tr}\ \mathcal{L}$. This phenomenon is clearly illustrated in the experiments for $\sigma = 0.025$: the performance of the SURE with estimated $\hat{\sigma}$ falls quite sharply.

Finally, we report some experimental results in the context of block thresholding. For each scale $j = 0, 1, \ldots, 4$, the n wavelet coefficients are split into L blocks of uniform length (except for the last block that can be shorter). The same sample of noisy signal as above is used for

Table 1: Comparison of mean SNR performance for each coordinatewise procedure over $N = 25$ realizations of the low to high noise levels settings with corresponding empirical standard deviation.

	$\sigma = 0.025$		$\sigma = 0.1$		$\sigma = 0.4$	
	$\hat{\sigma} = 0.050 \pm 0.0004$		$\hat{\sigma} = 0.109 \pm 0.0015$		$\hat{\sigma} = 0.401 \pm 0.0074$	
	$\mathrm{SNR_{in}} = 29.71 \pm 0.11$		$\mathrm{SNR_{in}} = 17.66 \pm 0.11$		$\mathrm{SNR_{in}} = 5.64 \pm 0.14$	
	G	LD	G	LD	G	LD
$\mathrm{MSE}^{\beta=1}$	31.12 ±0.14	32.07 ±0.13	20.20 ±0.17	23.31 ±0.25	8.69 ±0.16	12.71 ±0.26
$\mathrm{SURE}^{\beta=1}_{\sigma}$	31.09 ±0.13	32.05 ±0.13	20.19 ±0.18	23.28 ±0.23	8.67 ±0.15	12.66 ±0.24
$\mathrm{SURE}^{\beta=1}_{\hat{\sigma}}$	28.26 ±0.26	29.52 ±0.17	20.13 ±0.21	23.25 ±0.26	8.67 ±0.15	12.66 ±0.25
$\mathrm{MSE}^{\beta=2}$	32.06 ±0.13	32.12 ±0.13	22.30 ±0.24	23.31 ±0.25	9.48 ±0.2	12.68 ±0.26
$\mathrm{SURE}^{\beta=2}_{\sigma}$	32.03 ±0.16	32.06 ±0.15	22.29 ±0.24	23.27 ±0.24	9.45 ±0.2	12.63 ±0.26
$\mathrm{SURE}^{\beta=2}_{\hat{\sigma}}$	29.77 ±0.29	29.67 ±0.16	22.15 ±0.27	23.26 ±0.25	9.46 ±0.2	12.63 ±0.26



(a) $\beta = 1$ (b) $\beta = 2$

Figure 2: MSE risk and its SURE estimates as a function of the threshold parameter ($\sigma = 0.1$).

an easier comparison. The chosen cardinalities $|L|$ take values in $\{25, 50, 75, 100\}$. In these experiments, a uniform threshold is optimized using a 2000 points grid search. The average SNRs along with their standard deviation are summarized in Table 2.

Table 2: Comparison of mean SNR performance for each blockwise procedure over $N = 25$ realizations with $\sigma = 0.1$ and their corresponding empirical standard deviation.

	$\lvert L\rvert = 25$	$\lvert L\rvert = 50$	$\lvert L\rvert = 75$	$\lvert L\rvert = 100$
$\mathrm{MSE}^{\beta=1}$	21.38±0.20	21.36±0.20	21.34±0.20	21.32±0.20
$\mathrm{SURE}^{\beta=1}_{\sigma}$	21.38±0.20	21.35±0.20	21.33±0.20	21.30±0.20
$\mathrm{SURE}^{\beta=1}_{\hat{\sigma}}$	21.24±0.21	21.22±0.21	21.21±0.22	21.18±0.20
$\mathrm{MSE}^{\beta=2}$	23.29±0.19	23.28±0.19	23.27±0.19	23.27±0.19
$\mathrm{SURE}^{\beta=2}_{\sigma}$	23.27±0.19	23.27±0.19	23.25±0.20	23.25±0.20
$\mathrm{SURE}^{\beta=2}_{\hat{\sigma}}$	23.21±0.18	23.21±0.16	23.20±0.18	23.20±0.18

These experiments show better performance compared to the coordinatewise with a global threshold whereas results are comparable with the level-dependent James-Stein thresholding. It is worth noting that the number of blocks in each scale does not seem to have a significant influence on the results. It is expected that the level-dependent method to give better perfor-

mance but it would require a more sophisticated optimization algorithm than grid seach to be computationally acceptable.

5 Conclusion and Perspectives

We have introduced a version of the SURE designed for SGWT, allowing automatic parameter selection in denoising tasks of signals on large graphs. Closed-form expressions for coordinatewise and block SUREs have been provided for a wide range of threshold rules and a graph analogue of Von Neumann estimator has been proposed for the parameter selection to be fully data-driven. Many experiments on the large baseline Minnesota graph has been conducted. For smooth functions on the graph, performances are comparable with those of an oracle knowing both the signal and the noise level. This shows that our blind parameter selection provides the missing piece to apply SGWT based denoising in practice.

In this work, the numerical complexity of the methodology has not been discussed. In the same vein as Hammond et al. (2011), many of the involved steps might be numerically optimized using Chebyshev polynomials. Actually, the only problematic step in the method is the computation of the weights $(\mathcal{W}\mathcal{W}^*)_{i,j}$ appearing in the SURE. The space-time complexity might be reduced taking advantage of the low-rank property of $\mathcal{W}\mathcal{W}^*$ implying several linear constraints on the weights (precisely nJ).

As shown in experiments, the quality of the noise estimation is crucial. For denoising with standard wavelets, the noise σ^2 is generally estimated by the Median Absolute Deviation (MAD). Basically, the idea is that the wavelet coefficients at the finest scale mainly encode the noise if the original signal is sufficiently regular. Consequently, the singular points are considered as outliers and swept out whereas the noise is estimated on the remaining coefficients. In our case, the noise is not only correlated but also degenerated since the rank of $\psi_J(\mathcal{L})$ is nothing but the number of eigenvalues of $\mathcal{L}$ in the support of ψ_J. Still, the MAD should be adaptable to our framework and might perform better in the case of small noise level.

References

Behjat, H., Richter, U., Van De Ville, D., and Sörnmo, L. (2016). Signal-adapted tight frames on graphs. *IEEE Trans. Signal Process.*, 64(22):6017–6029.

Belkin, M. and Niyogi, P. (2008). Towards a theoretical foundation for laplacian-based manifold methods. *Journal of Computer and System Sciences*, 74(8):1289–1308.

Cai, T. T. (1999). Adaptive wavelet estimation: a block thresholding and oracle inequality approach. *Ann. Statist.*, 27(3):898–924.

Chung, F. R. and Graham, F. C. (1997). *Spectral graph theory*. Number 92. American Mathematical Soc.

Coifman, R. R. and Lafon, S. (2006). Diffusion maps. *Applied and computational harmonic analysis*, 21(1):5–30.

Deutsch, S., Ortega, A., and Medioni, G. (2016). Manifold denoising based on spectral graph wavelets. In *Acoustics, Speech and Signal Processing (ICASSP), 2016 IEEE International Conference on*, pages 4673–4677. IEEE.

Dong, B., Jiang, Q., Liu, C., and Shen, Z. (2016). Multiscale representation of surfaces by tight wavelet frames with applications to denoising. *Applied and Computational Harmonic Analysis*, 41(2):561 – 589. Sparse Representations with Applications in Imaging Science, Data Analysis, and Beyond, Part II.

Donoho, D. L. and Johnstone, I. M. (1995). Adapting to unknown smoothness via wavelet shrinkage. *Journal of the american statistical association*, 90(432):1200–1224.

Gavish, M., Nadler, B., and Coifman, R. R. (2010). Multiscale wavelets on trees, graphs and high dimensional data: Theory and applications to semi supervised learning. In *ICML*, pages 367–374.

Göbel, F., Blanchard, G., and von Luxburg, U. (2018). *Construction of tight frames on graphs and application to denoising*, pages 503–522. Springer.

Hammond, D. K., Vandergheynst, P., and Gribonval, R. (2011). Wavelets on graphs via spectral graph theory. *Applied and Computational Harmonic Analysis*, 30(2):129–150.

Irion, J. and Saito, N. (2017). Efficient approximation and denoising of graph signals using the multiscale basis dictionaries. *IEEE Transactions on Signal and Information Processing over Networks*, 3(3):607–616.

Kereta, Z., Vigogna, S., Naumova, V., Rosasco, L., and De Vito, E. (2019). Monte carlo wavelets: a randomized approach to frame discretization. *arXiv preprint arXiv:1903.06594*.

Leonardi, N. and Van De Ville, D. (2013). Tight wavelet frames on multislice graphs. *IEEE Transactions on Signal Processing*, 61(13):3357–3367.

Liu, J. S. (1994). Siegel's formula via Stein's identities. *Statist. Probab. Lett.*, 21(3):247–251.

Luisier, F., Blu, T., and Unser, M. (2007). A new sure approach to image denoising: Interscale orthonormal wavelet thresholding. *IEEE Transactions on image processing*, 16(3):593–606.

Navarro, F., Fadili, M., and Chesneau, C. (2013). Adaptive parameter selection for block wavelet-thresholding deconvolution. *IFAC Proceedings Volumes*, 46(11):495–499.

Onuki, M., Ono, S., Yamagishi, M., and Tanaka, Y. (2016). Graph signal denoising via trilateral filter on graph spectral domain. *IEEE Transactions on Signal and Information Processing over Networks*, 2(2):137–148.

Onuki, M., Ono, S., Yamagishi, M., and Tanaka, Y. (2016). Graph signal denoising via trilateral filter on graph spectral domain. *IEEE Trans. Signal Inform. Process. Netw.*, 2(2):137–148.

Ortega, A., Frossard, P., Kovačević, J., Moura, J. M., and Vandergheynst, P. (2018). Graph signal processing: Overview, challenges, and applications. *Proceedings of the IEEE*, 106(5):808–828.

Pesquet, J.-C., Benazza-Benyahia, A., and Chaux, C. (2009). A sure approach for digital signal/image deconvolution problems. *IEEE Transactions on Signal Processing*, 57(12):4616–4632.

Peyré, G., Fadili, J., and Chesneau, C. (2011). Adaptive structured block sparsity via dyadic partitioning. In *2011 19th European Signal Processing Conference*, pages 1455–1459. IEEE.

Shuman, D. I., Narang, S. K., Frossard, P., Ortega, A., and Vandergheynst, P. (2013). The emerging field of signal processing on graphs: Extending high-dimensional data analysis to networks and other irregular domains. *IEEE Signal Processing Magazine*, 30(3):83–98.

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

Vaiter, S., Deledalle, C.-A., Peyré, G., Dossal, C., and Fadili, J. (2013). Local behavior of sparse analysis regularization: Applications to risk estimation. *Applied and Computational Harmonic Analysis*, 35(3):433–451.

von Neumann, J. (1941). Distribution of the ratio of the mean square successive difference to the variance. *Ann. Math. Statistics*, 12:367–395.