

REGULARISED ESTIMATION OF 2D-LOCALLY STATIONARY WAVELET PROCESSES

Alex J. Gibberd, James D. B. Nelson

Department of Statistical Science, University College London, Gower Street, London WC1E 6BT

ABSTRACT

Locally Stationary Wavelet processes provide a flexible way of describing the time/space evolution of autocovariance structure over an ordered field such as an image/time-series. Classically, estimation of such models assume continuous smoothness of the underlying spectra and are estimated via local kernel smoothers. We propose a new model which permits spectral jumps, and suggest a regularised estimator and algorithm which can recover such structure from images. We demonstrate the effectiveness of our method in a synthetic experiment where it shows desirable estimation properties. We conclude with an application to real images which illustrate the qualitative difference between the proposed and previous methods.

1. INTRODUCTION

Locally Stationary Wavelet (LSW) processes, as originally introduced by Nason et al. [11] provide a statistically well-principled method to capture the localised covariance of a signal. Additionally, given their wavelet underpinnings, they also allow the de-coupling of this covariance structure across multiple length/time scales.

The extension of the LSW framework and theory to 2-dimensional fields by Eckley et al. [3, 4] has led to new applications in the analysis of textured images. Alternative wavelet based approaches for describing random-fields either study the distribution of the wavelet coefficients directly [9], or learn the wavelet coefficients through machine-learning based multiple-kernel learning algorithms [16]. The well-principled design of the LSW process means that statistical properties such as asymptotic bias [11] and variance [10] of estimators are relatively well understood. Knowledge of such properties enables the construction of statistical tests for non-stationarity in both time-series [10] and random fields [15, 12].

There are however, still some large open research challenges relating to LSW models, both in terms of how relevant they are to modeling real phenomena (in our case real world images/textures), and how we can effectively we can estimate model structure from data. In this paper we attempt to tackle some of these challenges. In Section 3, we examine how one might relax the continuous smoothness require-

ments of the traditional LSW process to allow for piecewise-variation of spectra. We then show in Section 4 how an Alternating Directed Method of Multipliers (ADMM) algorithm can be utilised to efficiently find optimal estimates from an image. Experiments (Section 5) demonstrate that even when the sample/image size is very small, our proposed estimator selects structure consistent with an LSW model. In the next section we formally introduce the 2D-LSW process, and further discuss the motivation for extending the LSW modeling framework.

2. THE 2D LSW PROCESS

Let the 2D-discrete wavelet filters $\{\psi_j^l\}$ be defined as a square matrix of size $L_j \times L_j$, where L_j gives the support of the wavelet at the j th scale level, i.e. $L_j = (2^j - 1)(N_f - 1) + 1$ and N_f is the number of non-zero elements in the associated low-pass mirror filter. The individual elements ψ_{j,k_1,k_2}^l are defined through the tensor products of the corresponding one-dimensional wavelets. In the horizontal direction h we have $\psi_{j,k}^h = \phi_{j,k_1} \psi_{j,k_2}$, in the vertical direction $\psi_{j,k}^v = \psi_{j,k_1} \phi_{j,k_2}$ and in the diagonal direction $\psi_{j,k}^d = \psi_{j,k_1} \psi_{j,k_2}$, where $\phi_{j,k}$ are the associated father/scaling wavelets.

In order to construct a LSW process we utilise the non-decimated wavelet transform such that there is an equivalent number of wavelet coefficients at each scale level $j = 1, \dots, J$. To achieve this we simply translate the discrete wavelets $\{\psi_j^l\}$ over the space \mathbb{Z}^2 , the non-decimated set of discrete wavelets are defined as $\psi_{j,\mathbf{u}}^l(\mathbf{r}) := \psi_{j,\mathbf{u}-\mathbf{r}}^l$ for all j, l , and $\mathbf{u}, \mathbf{r} \in \mathbb{Z}^2$. The *2d-Locally Stationary Wavelet process* can now be defined as a doubly indexed stochastic process over the field \mathbb{R}^2 , it is indexed by position $\mathbf{r} = (r_1, r_2)$ and the size of the image $\mathbf{R} = (R_1, R_2)$ as:

$$X_{\mathbf{r};\mathbf{R}} = \sum_l \sum_{j=1}^{\infty} \sum_{\mathbf{u}} w_{j,\mathbf{u}}^l \psi_{j,\mathbf{u}}^l(\mathbf{r}) \epsilon_{j,\mathbf{u}}^l,$$

where $l = h, v, d$ refer to horizontal, vertical or diagonal directions. The process is broken down into three components: 1) A stochastic term $\{\epsilon_{j,\mathbf{u}}^l\}$, this encodes no independence structure in itself and is an i.i.d zero-mean random variable, they are orthogonal such that $\epsilon_{j,\mathbf{u}}^l \perp \epsilon_{j',\mathbf{u}'}^{l'}$ for all $j \neq j'$, $\mathbf{u} \neq \mathbf{u}'$ and $l \neq l'$. 2) The non-decimated wavelets $\{\psi_{j,\mathbf{u}-\mathbf{r}}^l\}$, allowing us to separate localise signal in both space and frequency, and 3) A collection of real valued amplitudes $\{w_{j,\mathbf{u}}^l\}$ which can be interpreted as discrete transfer functions.

Thanks to the Defence Science Technology Laboratory (Dstl) for funding this project through their National PhD Scheme

2.1. Smoothness Assumptions

In previous works authors enforce smoothness on the process via a Lipschitz continuous transfer function $W_j^l(\mathbf{u}/\mathbf{R})$, where the ratio \mathbf{u}/\mathbf{R} is understood to be component-wise. This function relates to the discrete amplitudes $w_{j,\mathbf{u}}^l$ such that the maximum deviation between the two functions decays asymptotically, i.e. $\sup_{\mathbf{u}} |w_{j,\mathbf{u}}^l - W_j^l(\mathbf{u}/\mathbf{R})| \leq \mathcal{O}(R^{-1})$. In the original LSW formulation [11, 3], the above constraints are used to impose asymptotic smoothness on the transfer function, as sample size increases the authors then show asymptotic properties relating to convergence of the (auto)covariance (of the process) to the Local Wavelet Spectrum (LWS), defined as: $S_j^l(\mathbf{u}/\mathbf{R}) = |W_j^l(\mathbf{u}/\mathbf{R})|^2$, for $j = 1, \dots, J(\mathbf{R})$.

Remark 1. *Continuously Smooth Spectra – In real images we do not realistically expect the spectral structure of images to vary in a continuous manner, rather the spectral properties of the image may change abruptly – for example where there exist two or more neighbouring regions of different materials or textures. In Section 3 we propose an estimator which relaxes such smoothness conditions and can permit spectral jumps.*

2.2. Kernel Estimators for the LWS

Let $x_{\mathbf{r};\mathbf{R}}$ be a sample pixel (at position \mathbf{r}) drawn from the corresponding LSW-process $X_{\mathbf{r};\mathbf{R}}$. A natural estimator for the spectrum can be constructed around the *empirical periodogram* which is defined in as $\hat{I}_{j,\mathbf{r}}^l = |d_{j,\mathbf{r}}^l|^2$ where $d_{j,\mathbf{r}}^l = \sum_{\mathbf{k}} \psi_{j,\mathbf{r}-\mathbf{k}}^l x_{\mathbf{r};\mathbf{R}}$ are the 2D-wavelet coefficients. It has been shown by Nason et al [11] in the 1D-case, and Eckley et al. [3] in the 2D case, that the empirical periodogram by itself is both biased and inconsistent, more specifically $E[\hat{I}_{j,\mathbf{r}}^l] = \sum_{j,l} A_{(j,l),(j',l')} S_{j'}^{l'}(\mathbf{r}/\mathbf{R}) + \mathcal{O}(\min(\mathbf{R})^{-1})$. Mixing between scales/directions is encoded by the matrix $\mathbf{A} = \sum_{\mathbf{k}} \Psi_j^l(\mathbf{k}) \Psi_{j'}^{l'}(\mathbf{k}) \in \mathbb{R}^{JL \times JL}$, where $\Psi_j^l(\mathbf{r}) = \sum_{\mathbf{r}'} \psi_{j,\mathbf{r}}^l \psi_{j',\mathbf{r}'}^l$ is referred to as the auto-correlation wavelet. Such results are extensions of work in the 1D-case, however, rather than just mixing over scale, in the 2D-case we observe a dispersion of power over direction l .

Inverting \mathbf{A} now allows one to construct an unbiased estimator as $\hat{S}_{j,\mathbf{r}}^l = \sum_{j',l'} A_{(j,l),(j',l')}^{-1} \hat{I}_{j',\mathbf{r}}^{l'}$, although unfortunately this estimator is still not consistent, i.e. $\text{Var}(\hat{S}) \not\rightarrow 0$ as $R \rightarrow \infty$. To encourage consistency one must perform some kind of smoothing over the image/samples. Typically it is suggested to smooth $\hat{I}_{j,\mathbf{r}}^l$ and then perform de-biasing, where smoothing is performed by either a second stage of wavelet transform and then thresholding, or adopting a moving average/kernel smoother [11, 4, 10]. In particular, in this paper we consider comparison to the de-biased kernel estimator given as:

$$\hat{K}_{j,\mathbf{r}}^l = \sum_{j',l'} A_{(j,l),(j',l')}^{-1} \left(\sum_{\mathbf{k}} \mathcal{K}(\mathbf{r} - \mathbf{k}) \hat{I}_{j',\mathbf{r}}^{l'} \right), \quad (1)$$

where $\mathcal{K}(\mathbf{r} - \mathbf{k})$ is a 2-dimensional kernel function defined over \mathbb{Z}^2 with bounded support. For simplicity, in this work we use a box-car kernel with width h .

Remark 2. *Finite Sample Performance – Whilst asymptotically, one may show that kernel smoothers akin to (1) can recover the LWS spectrum in a consistent manner [10, 14], the finite sample performance of such estimators is often lacking. For example, in the multivariate case we demonstrated [5] such an estimator often produced negative estimates for the variance, a quantity that through the model construction is required to be positive.*

3. REGULARISED ESTIMATION OF THE LWS SPECTRUM

An alternative to kernel estimation is to construct a regularised estimator that can incorporate additional prior information about the LWS estimate. We define a quadratic loss function over the whole set of scales $j = 1, \dots, J(\mathbf{R})$ and directions $l = \{h, v, d\}$ as:

$$L(\bar{\mathbf{I}}; \mathbf{B}) := \sum_{j,l} \|\bar{I}_{j,\mathbf{r}}^l - \sum_{j',l'} A_{(j,l),(j',l')} B_{j',\mathbf{r}}^{l'}\|_F^2, \quad (2)$$

where $\bar{I}_{j,\mathbf{r}}^l$ is an estimator of the biased spectrum. We note, that the above can be written in matrix form as $L(\bar{\mathbf{I}}; \mathbf{B}) := \|\bar{\mathbf{I}} - \mathbf{A}\mathbf{B}\|_F^2$, where $\bar{\mathbf{I}} \in \mathbb{R}^{LJ \times R_1 R_2}$ is a reshaped estimate of the periodogram and $\mathbf{B} \in \mathbb{R}^{R_1 R_2 \times LJ}$ is our estimator relating to the set of B . A natural choice for the link function $\bar{I}_{j,\mathbf{r}}^l$ would be to use the raw empirical periodogram $\bar{I}_{j,\mathbf{r}}^l = \hat{I}_{j,\mathbf{r}}^l$, however, we instead opt to use the kernel estimator $\bar{I}_{j,\mathbf{r}}^l = \bar{K}_{j,\mathbf{r}}^l := \sum_{\mathbf{k}} \mathcal{K}(\mathbf{r} - \mathbf{k}) \hat{I}_{j',\mathbf{r}}^{l'}$, we note that when $\mathbf{B} = \hat{K}_{j,\mathbf{r}}^l$ as defined in (1), then $L(\bar{\mathbf{I}}; \mathbf{B}) = 0$ and is minimised.

Given we have defined a loss function for estimating the LWS, it is natural that we should consider whether we are required to incorporate additional constraints into the estimation procedure. The beauty of the formulation in Eq. (2) is that it allows us to consider estimation of the LWS spectrum as a convex optimisation problem. For example, since the LWS estimate should be positive, we can restrict solutions such that $\hat{\mathbf{B}} \geq 0$ through the simple addition of an indicator function, i.e. $l_{\mathbb{R}^+}(\mathbf{B}) = 0$ if $B_{i,j} \geq 0 \forall i, j$ and $+\infty$ otherwise. Further to this positivity requirement, we can introduce additional constraints that relate to how we want or expect the estimates to behave when applied to real images. Inspired by work in change-point detection, and image segmentation, we here introduce a total-variation inspired penalty which actively constrains the estimator variation across the whole image according to;

$$R_j^l(\mathbf{B}) = \lambda \sum_{m=2}^{R_1} \sum_{n=2}^{R_2} |B_{j,(r_1^{(m)}, r_2^{(n)})}^l - B_{j,(r_1^{(m-1)}, r_2^{(n-1)})}^l|,$$

or in matrix form as; $R(\mathbf{B}) = \lambda \|[\mathbf{D}_H \mathbf{B}; \mathbf{D}_V \mathbf{B}]\|_1$, where $\mathbf{D}_H, \mathbf{D}_V$ are differencing matrices operating respectively in

the horizontal and vertical dimensions. The table below describes the estimators we consider in this paper.

Name - Abbreviation	Loss Fn.	Penalty
De-Biased Kernel - K	$\hat{K}_{j,r}^l$	n/a
Positive Kernel - K(P)	$L(K_{j,r}^l; \mathbf{B})$	$l_{\mathbb{R}^+}$
TV-Positive - TV(P)	$L(\hat{\mathbf{I}}_{j,r}; \mathbf{B})$	$l_{\mathbb{R}^+} + R$
TV-Kernel Positive - TVK(P)	$L(K_{j,r}^l; \mathbf{B})$	$l_{\mathbb{R}^+} + R$

Remark 3. *Effect of Regularisation – Much like the kernel estimator in (1) the TV-estimators attempt to smooth the raw periodogram over space. However, unlike simply using $\hat{K}_{j,r}^l$ alone, such a TV-constraint fuses estimates across the whole image resulting in a global estimator. Due to the ℓ_1 form of the norm, such estimators should promote piecewise structure in the spectra, for example see Harchaoui et al. [6] for an application in the univariate change-point detection setting.*

4. ALTERNATING DIRECTED METHOD OF MOMENTS ALGORITHM

Due to the biased nature of the raw periodogram we need to consider jointly all the scale levels within one optimisation problem. Our strategy here is to tackle this optimisation task with an *Alternating Directed Method of Multipliers (ADMM)* approach. Such methods are becoming increasingly popular for solving convex optimization problems that are formed as the sum of convex objectives. The ADMM method allows one to split up the optimisation problem across linearly separable portions of the objective.

Taking the K(P)-LWS objective for example, we reformulate the optimisation problem in an explicitly constrained form; $\arg \min_{\mathbf{B}; \mathbf{U}=\mathbf{B}} \|\bar{\mathbf{K}} - \mathbf{AB}\|_F^2 + l_{\mathbb{R}^+}(\mathbf{U})$, where \mathbf{U} is referred to as an auxiliary variable. In practice, and to ensure sufficient curvature, an augmentation term $\mu/2 \|\mathbf{U} - \mathbf{B}\|_F^2$ is added to the Lagrangian, in the general case, we introduce auxiliary variables for each term $\mathbf{Q}_B = \mathbf{Q}_{\mathbb{R}^+} = \mathbf{B}, \mathbf{Q}_{AB} = \mathbf{AB}, \mathbf{Q}_D = \mathbf{DQ}_B$, i.e. in matrix form $\mathbf{FB} = \mathbf{GQ}$, where

$$\mathbf{F} = \begin{bmatrix} \mathbf{A} \\ \mathbf{I} \\ \mathbf{0} \\ \mathbf{I} \end{bmatrix}, \quad \mathbf{G} = \begin{bmatrix} \mathbf{I} & & & \\ & \mathbf{I} & & \\ & -\mathbf{D} & \mathbf{I} & \\ & & & \mathbf{I} \end{bmatrix}.$$

We now formulate the augmented Lagrangian $\mathcal{L}(\mathbf{B}, \mathbf{Q}, \mathbf{V}) :=$

$$\|\bar{\mathbf{K}} - \mathbf{Q}_{AB}\|_F^2 + l_{\mathbb{R}^+}(\mathbf{Q}_{\mathbb{R}^+}) + \lambda \|\mathbf{Q}_D\|_1 + \frac{\mu}{2} \|\mathbf{FB} - \mathbf{GQ} - \mathbf{V}\|_F^2,$$

where \mathbf{V}/μ are Lagrange multipliers. The Lagrangian problem can now be solved through a series of updates, minimising sequentially, $\arg \min_{\mathbf{U}} \mathcal{L}(\mathbf{B}, \mathbf{Q}, \mathbf{V})$, $\arg \min_{\mathbf{Q}} \mathcal{L}(\mathbf{B}, \mathbf{Q}, \mathbf{V})$, and then updating \mathbf{V} to keep track of the cumulative errors (see [13, 2] for a review)

One benefit of ADMM, is that the updates required for $\arg \min_{\mathbf{U}} \mathcal{L}(\mathbf{B}, \mathbf{Q}, \mathbf{V})$ are simply proximity operators.

These can be calculated extremely quickly, projecting onto the positive real line for $l_{\mathbb{R}^+}$, and updating via the soft-thresholding operator for the TV-denoising methods. Algorithm 1 provides more details, in practice we utilised a version of Iordache et al.'s SunSAL-TV algorithm [7]. Whilst the ADMM scheme is guaranteed to converge if a solution exists for any $\mu > 0$, we use $\mu = 100$, which converged relatively quickly ~ 1 minute for an $R = 256$ size image, the algorithm appears to scale computationally as $\mathcal{O}(R^2)$.

Algorithm 1 ADMM algorithm for LWS smoothing.

Require: $\mu > 0, \bar{\mathbf{I}}, \mathbf{A}$

- 1: **while** not converged **do**
- 2: $\mathbf{B}^{(k+1)} \leftarrow \arg \min_{\mathbf{B}} \mathcal{L}(\mathbf{B}, \mathbf{Q}^{(k)}, \mathbf{V}^{(k)})$
- 3: $\mathbf{Q}^{(k+1)} \leftarrow \arg \min_{\mathbf{Q}} \mathcal{L}(\mathbf{B}^{(k+1)}, \mathbf{Q}, \mathbf{V}^{(k)})$
- 4: Solve via proximity operators
- 5: i.e. $\mathbf{Q}_{\mathbb{R}^+}^{(k+1)} \leftarrow \max(\mathbf{B}^{(k+1)} - \mathbf{V}_{\mathbb{R}^+}^{(k)}, 0)$
- 6: $\mathbf{Q}_H^{(k+1)} \leftarrow \text{soft}(\mathbf{D}_H \mathbf{B}^{(k+1)} - \mathbf{V}_H, \lambda/\mu)$
- 7: $\mathbf{V}^{(k+1)} \leftarrow \mathbf{V}^{(k)} + (\mathbf{FB}^{(k+1)} - \mathbf{GQ}^{(k+1)})$
- 8: **end while**

5. EXPERIMENTS

To test the recovery ability of our estimators and examine how they perform as a function of problem size we generate synthetic data-sets according to a simple piecewise constant texture model. We split variance structure as encoded through $S_{j,r}^l$ into a set of blocks with alternating values and denote the in block structure as $\{w_{j,b}^l = w_{j,r}^l \forall r \in \mathcal{R}_b \subset \mathbb{Z}^2\}$ for blocks $b = 1, \dots, B$. In this experiment we consider recovery of the true structure $\{S_{j,b}^l\}$ from synthetic data $\{x_{r;R}\} \sim LSW(\{w_{j,b}^l\})$ where $\epsilon_{j,r}^l \sim \mathcal{N}(0, 1)$ over images of varying size. The simulated structure looks broadly like that estimated in Fig. (1), where dark regions have $S_{j,b}^l = 0$ and light regions have $S_{j,b}^l = 1$, in these experiments true signal is restricted to scale $j = 1$, and is identically distributed across directions $l = h, v, d$. In order to exam-

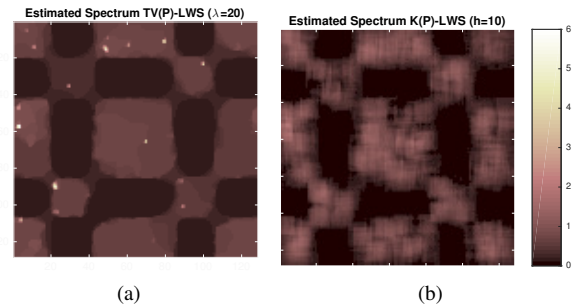


Fig. 1: Examples of recovered spectrum $\hat{S}_{1,b}^1$ (at $R = 128$), with a) TV(P)-LWS $\lambda = 20$, and b) K(P)-LWS $h = 10$.

ine statistical properties of the proposed estimators across the generative distribution, we consider a cross-validation setup. We first train on $N_{\text{train}} = 20$ images and select an optimal set of tuning parameters $(\hat{h}, \hat{\lambda})$ which minimise the error

$\epsilon_{\text{test}}(h, \lambda) = \sum_{j,r,l} (|w_{j,r}^l|^2 - \hat{B}_{j,r}^l) / JR^2 D$. The error surfaces (Fig. 2) can be quite insightful in terms of understanding how performance trades off when introducing different levels of prior smoothness knowledge through λ and h .

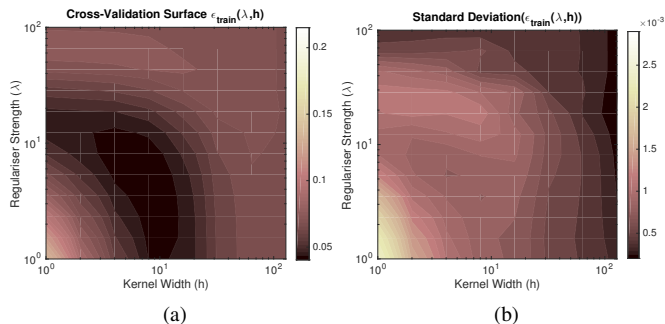


Fig. 2: a) The mean cross-validation error surface ϵ_{train} , and b) the standard-deviation of the surface, taken over $N_{\text{train}} = 20$ synthetic images of size $R = 128$.

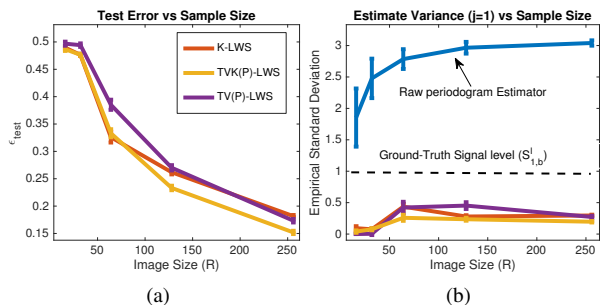


Fig. 3: Test performance of the various LWS estimators. a) The error at scale level $j = 1$ as a function of image size, b) standard-deviation of estimate taken over $N_{\text{test}} = 100$ simulations.

Clearly, estimation performance is enhanced by performing smoothing using either the kernel and/or regularisation. The distinctive kink in the cross-validation surface is typical for estimation across image sizes. It is interesting to note, that whilst optimal performance requires some kernel *and* some regularisation, the regularised estimate with $h = 1$ performs relatively well. In practice this relates in TV-LWS resolving more clearly defined edge detail as seen in Fig. (1).

Figure 3 summarises performance of the estimators on a test set (of size $N_{\text{test}} = 100$) with specifically chosen tuning parameters $(\hat{h}, \hat{\lambda}) = \arg \min \epsilon_{\text{train}}(h, \lambda)$. These results gives us some measure of the sample efficiency of the different estimators. We can see that the error for all smoothed estimators converges, both at scale $j = 1$ and when averaged across all scales. This is in contrast to the raw-periodogram, which does not converge for $j = 1$ (where all the true spectral structure is simulated). We see that the hybrid method TVK(P)-LWS appears to perform best, converging faster at all scales. Additionally, it appears to have favourable rates of consistency, as can be established from Fig (3b), which also serves to highlight the inconsistency of the bias-corrected raw periodogram.

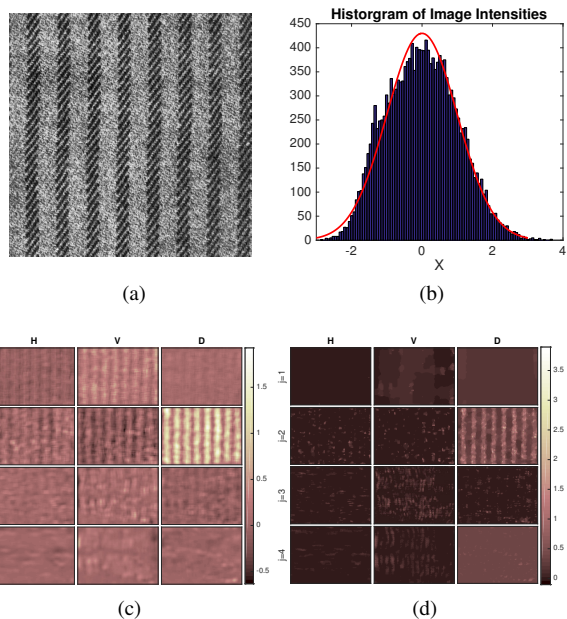


Fig. 4: Comparison of LWS estimation with regularised and kernel approach. a) Original image. b) Histogram of z-scored image (normalised across both rows/columns). c) Estimated LWS with KLWS $h = 16$, and d) Estimated LWS with TV(P)-LWS $\lambda = 10$.

5.1. Real Images

We here compare the qualitative effect of regularised LWS estimation applied to a carpet like texture taken from the Brodatz data set ($R = 640$). Given that we wish to model the image as a Gaussian process, we first pre-process the image down-sampling to a size of $R_{\text{eff}} = 128 \times 128$ pixels. The image is then standardized by *z-scoring* the data with respect to all pixels in the rescaled image. Figure 4 demonstrates the output of the analysis and how the pre-processing effectively produces a Gaussian intensity distribution. One clearly observes how the regularised solution is more parsimonious, and is positive valued across the spectra, whilst still preserving the key spectral features, namely, the diagonal banding in $j = 2, l = d$.

Conclusion

Typically, wavelet decompositions may be used to break down an image into a set of features which can then be used for general image understanding tasks, for example image classification [8] or clustering [1]. Enhanced estimation of wavelet spectra may aid in many of these applications.

We have demonstrated in this paper how one can construct a regularised estimator for the LWS and that this has beneficial estimation properties, such as the ability to model piecewise spectra. Future work may attempt to extend the theoretical estimation framework of Nason [11] and Eckley [3] to cope with such piecewise-smoothness.

6. REFERENCES

- [1] P. Bhattacharya, A. Biswas, and S. P. Maity. *Wavelets-Based Clustering Techniques for Efficient Color Image Segmentation*, pages 237–244. 2014.
- [2] S. Boyd, N. Parikh, and E. Chu. Distributed optimization and statistical learning via the alternating direction method of multipliers. *Foundations and Trends in Machine Learning*, 3(1):1–122, 2011.
- [3] I. A. Eckley and G. P. Nason. Efficient computation of the discrete autocorrelation wavelet inner product matrix. *Statistics and Computing*, 15(2):83–92, 2005.
- [4] I. A. Eckley, G. P. Nason, and R. L. Treloar. Locally stationary wavelet fields with application to the modelling and analysis of image texture. *Journal of the Royal Statistical Society Series C Applied Statistics*, 59(4):595–616, 2010.
- [5] A. J. Gibberd and J. D. B. Nelson. Estimating Multiresolution Dependency Graphs within the Stationary Wavelet Framework. *IEEE Global Conference on Signal and Information Processing*, 2015.
- [6] Z. Harchaoui and C. Lévy-Leduc. Multiple Change-Point Estimation With a Total Variation Penalty. *Journal of the American Statistical Association*, 105(492):1480–1493, 2010.
- [7] M. D. Iordache, J. M. Bioucas-Dias, and A. Plaza. Total variation spatial regularization for sparse hyperspectral unmixing. *IEEE Transactions on Geoscience and Remote Sensing*, 50(11 PART1):4484–4502, 2012.
- [8] S. K. Meher, B. U. Shankar, and A. Ghosh. Wavelet-feature-based classifiers for multispectral remote-sensing images. *IEEE Transactions on Geoscience and Remote Sensing*, 45(6):1881–1886, 2007.
- [9] D. Mondal and D. B. Percival. Wavelet variance analysis for random fields on a regular lattice. *IEEE Transactions on Image Processing*, 21(2):537–549, 2012.
- [10] G. Nason. A test for second-order stationarity and approximate confidence intervals for localized autocovariances for locally stationary time series. *Journal of the Royal Statistical Society. Series B: Statistical Methodology*, 75(5):879–904, 2013.
- [11] G. P. Nason, R. von Sachs, and G. Kroisandt. Wavelet processes and adaptive estimation of the evolutionary wavelet spectrum. *Journal of the Royal Statistical Society. Series B: Statistical Methodology*, pages 1–28, 2000.
- [12] M. A. Nunes, S. L. Taylor, and I. A. Eckley. A Multi-scale Test of Spatial Stationarity for Textured Images in R. *The R Journal*, 2014.
- [13] N. Parikh and S. Boyd. Proximal algorithms. *Foundations and Trends in optimization*, 1(3):123–231, 2013.
- [14] T. Park, I. A. Eckley, and H. C. Ombao. Estimating Time-Evolving Partial Coherence Between Signals via Multivariate Locally Stationary Wavelet Processes. *IEEE Transactions on Signal Processing*, 62(20):5240–5250, 2014.
- [15] S. L. Taylor, I. A. Eckley, and M. A. Nunes. A Test of Stationarity for Textured Images. *Technometrics*, 56(3):291–301, 2014.
- [16] F. Yger and A. Rakotomamonjy. Wavelet Kernel Learning. *Pattern Recognition*, 44, 2011.