

# GENERALISED M-LASSO FOR ROBUST, SPATIALLY REGULARISED HURST ESTIMATION

*J. D. B. Nelson*

Department of Statistical Science  
University College London  
j.nelson@ucl.ac.uk

*C. Naornita, A. Isar*

Communications Department  
Politehnica University Timisoara  
{corina.naornita@upt.ro, alexandru.isar@upt.ro}

## ABSTRACT

A generalised Lasso iteratively reweighted scheme is here introduced to perform spatially regularised Hurst estimation on semi-local, weakly self-similar processes. This is extended further to the robust, heavy-tailed case whereupon the generalised M-Lasso is proposed. The design successfully incorporates both a spatial derivative in the generalised Lasso regulariser operator and a weight matrix formulated in the wavelet domain. The result simultaneously spatially smooths the Hurst estimates and downweights outliers. Experiments using a Hampel score function confirm that the method yields superior Hurst estimates in the presence of strong outliers. Moreover, it is shown that the inferred weight matrix can be used to perform wavelet shrinkage and denoise fractional Brownian surfaces in the presence of strong, localised, band-limited noise.

## 1. INTRODUCTION

Characterised by their long-range dependency structure, weak self-similar models find myriad applications in signal and image processing and beyond including texture analysis for super-resolution [1], classification [2], segmentation [3], condition monitoring [4], decluttering [5], and denoising [6, 7, 8].

The Hurst parameter, which controls the power-law decay, and therefore the smoothness or texture, is usually of great importance and various estimation schemes have been developed. However, the majority of these either assume a globally constant value or, as in the multifractal formalism, forgo any attempt to estimate the pointwise value and instead describe how it varies stochastically over space [9, 10, 11]. Alternatively, one can consider a case somewhere in-between these two extremes whereby the Hurst parameter varies as a piecewise low-order polynomial. The piecewise constant case could, for example, form a model outline to describe multiple textures or features in an image scene. Indeed, Pižurica [12] exploited interscale wavelet ratios and deviation from a power-law of sorts to perform adaptive denoising.

Recently, Naornita et al [13] formulated a spatially regularised Hurst estimator by exploiting Tibshirani's newly developed path algorithm for the generalised Lasso [14]. Their

method delivered superior Hurst estimation and Hurst-based adaptive denoising compared to ordinary least-squares. However, like other non-robust methods which assume normal error terms, their method is vulnerable to outlier noise in the wavelet energy spectrum. If the normality assumption is broken, the Hurst estimates can become very poor. We here propose a significant generalisation of [13] which allows for a heavier-tailed distribution and which mitigates outlier noise. Although robust regression approaches have been considered for Hurst estimation [15], they have not been combined with spatial regularisation and, indeed, have not even been attempted for local or semi-local Hurst estimation (the Hurst parameter is assumed constant throughout the data). Likewise, although robust approaches have been applied to the Lasso [16, 17, 18], they did not consider the generalised Lasso or Hurst estimation.

In Section 2 we discuss wavelet-based Hurst estimation and introduce an iteratively reweighted (IRLS) scheme to solve the generalised Lasso method. In Section 3 we describe robust estimation and propose a new robust IRLS M-estimator for the generalised Lasso problem. Experiments in Section 4 illustrate the superiority of the method for Hurst estimation and denoising in the presence of band-limited noise.

## 2. BACKGROUND

The computational machinery used by Naornita et al [13] is restricted to Tibshirani's generalised Lasso [14] path algorithm. We here proposed a simple, more flexible, alternative via IRLS.

### 2.1. Regularised Hurst estimation

There exists a natural and elegant connection between weak statistical self-similarity, the Hurst parameter, and multiresolution analysis as summarised in the following result.

**Theorem 1** (e.g. Nelson & Kingsbury [5]). *Let  $\mathcal{T} \subset \mathbb{R}^2$  and suppose the stochastic field  $z: \mathcal{T} \mapsto \mathbb{R}$ , has weak statistical self-similarity namely  $\mathbb{E}z(\lambda \cdot) = \lambda^H \mathbb{E}z$  and  $\mathbb{E}z(\lambda t)z(\lambda \cdot) = \lambda^{2H} \mathbb{E}z(t)z(\cdot)$ . Then  $\mathbb{E}|(\mathcal{W}z)(\cdot; k, \theta)|^2 \propto 2^{2k(H+1)}$  where  $\mathcal{W}$  is the wavelet operator defined by  $(\mathcal{W}z)(t; k, \theta) := 2^{-k} \langle z, \psi_\theta(2^{-k} \cdot - t) \rangle$ , with wavelet  $\psi$  defined over space  $t$ , orientation  $\theta$ , and  $k$ th finest scale level.*

A lesser studied generalisation of this problem is when the Hurst parameter is allowed to vary slowly over space. Here,

---

J. D. B. Nelson was partially supported by grants from Innovate UK and the Dstl

$\mathbb{E}|\cdot|^2$  is approximated by the sample 2nd moment in some spatial region. When localisation is of particular importance, the pointwise estimate  $E_{k,\theta}(\cdot) := |(\mathcal{W}z)(\cdot; k, m)|^2$  is used [19].

The spatial domain is discretised as  $t \in \mathcal{T} = \{t_i\}_{i=1}^{n_0}$ . We drop  $\theta$  (the same analysis is performed in each direction) and put  $y_k[i] := \log E_{k,m}(t_i)$ . Then the power law in Theorem 1 motivates the linear model  $y_k[i] = k\beta_2[i] + \beta_1[i]$ , i.e.  $\mathbf{y} = \mathbf{X}\boldsymbol{\beta}$ , with  $\mathbf{y} = (\mathbf{y}_1^\top, \dots, \mathbf{y}_{n_0}^\top)^\top \in \mathbb{R}^n$ ,  $\mathbf{y}_i = (y_{k_-}[i], \dots, y_{k_+}[i])^\top$ ,  $\boldsymbol{\beta} = (\boldsymbol{\beta}_1^\top, \dots, \boldsymbol{\beta}_n^\top)^\top \in \mathbb{R}^p$ ,  $\boldsymbol{\beta}_i = (\beta_1[i], \beta_2[i])^\top$ , and  $\mathbf{X} = \mathbf{I}_{n_0} \otimes \mathbf{X}_0 \in \mathbb{R}^{n \times p}$ , with

$$\mathbf{X}_0^\top = \begin{bmatrix} 1 & \dots & 1 \\ k_- & \dots & k_+ \end{bmatrix} \in \mathbb{R}^{2 \times K},$$

where  $\mathbf{I}_{n_0}$  is the  $n_0 \times n_0$  identity matrix, and where  $n = Kn_0$  and  $p = 2n_0$ . Here the regression is performed over a subset of scales  $k_-, \dots, k_+$  because the finest levels have poor SNR and the coarsest levels have poor spatial localisation.

The spatially varying Hurst parameter can then be estimated from the least-squares estimate via  $H[t_i] = 1/2\hat{\beta}_2[i] - 1$ . If the Hurst parameter varies slowly or piecewise over space then the generalised Lasso formulation of Naornita et al [13], namely  $\operatorname{argmin}_{\boldsymbol{\beta}} \|\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\|_2^2 + \lambda \|\nabla\boldsymbol{\beta}\|_1$ , where  $\nabla \in \mathbb{R}^{m \times p}$  performs spatial differentiation on  $\boldsymbol{\beta}$ , can be used to spatially regularise the estimates for  $\boldsymbol{\beta}$  and hence  $H$ . As shown in [13] and the Experiments section below, the regulariser smooths the estimated Hurst parameter over space and leads to significantly superior estimates for piecewise varying Hurst when compared to the pointwise, ordinary least-squares approach.

## 2.2. Iteratively reweighted least-squares Lasso

The basic form of Lasso (with  $\nabla = \mathbf{I}$  above) is not directly useful for Hurst estimation, per se, but it is nonetheless instructive to consider how it is solved via iteratively reweighted least-squares (IRLS). As discussed further in Section 3, although myriad alternatives exist (proximal methods, LARS, coordinate descent, Landweber iteration, grafting, Osbourne's algorithm,  $\epsilon$ -boosting, etc) the IRLS algorithm provides a simple means to extend Lasso to the robust case where the usual square-loss term is replaced by a cost function which down-weights the erroneous affects of anomalous, or outlier, values. Unlike many other methods, IRLS can be easily extended to overlapping mixed (grouped-sparse) norms [20] and, although its exponential convergence advantage [21] is mitigated by the iterations per loop, recent fast methods based on preconditioning have emerged [20]. The basic form [16] proceeds by rewriting the Lasso objective as a weighted ridge regression

$$\|\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\|_2^2 + \lambda \|\boldsymbol{\beta}\|_1 = \|\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\|_2^2 + \lambda \|\boldsymbol{\Omega}^{1/2}\boldsymbol{\beta}\|_2^2,$$

with  $\mathbf{y} \in \mathbb{R}^n$ ,  $\mathbf{X} \in \mathbb{R}^{n \times p}$ ,  $\boldsymbol{\beta} \in \mathbb{R}^p$ , and where  $\boldsymbol{\Omega} = \operatorname{diag}(|\beta_i|^{-1})_{i=1}^p \in \mathbb{R}^{p \times p}$ . One then arrives at the following simple iterative procedure for Lasso which solves a weighted ridge problem and then updates the weights:

$$\begin{aligned} \boldsymbol{\beta}^{(\ell+1)} &\leftarrow (\mathbf{X}^\top \mathbf{X} + \lambda \boldsymbol{\Omega}^{(\ell)})^{-1} \mathbf{X}^\top \mathbf{y}, \\ \boldsymbol{\Omega}^{(\ell)} &\leftarrow \operatorname{diag}(|\beta_i^{(\ell)}| + \delta)^{-1}_{i=1}^p. \end{aligned} \quad (1)$$

where  $\delta$  can be a small, fixed constant or updated as in [21].

## 2.3. Iteratively reweighted least-squares for regularised Hurst estimation

Since  $\nabla$  does not have full rank the regularised Hurst estimator cannot collapse down to the basic Lasso. We here fuse the ideas of the preceding two subsections and propose an IRLS scheme for generalised Lasso which offers an alternative means to solve the spatially regularised Hurst estimation problem. Building on the previous section we note that the generalised penalty can be written as  $\|\nabla\boldsymbol{\beta}\|_1 = \sum_i |\nabla_i^\top \boldsymbol{\beta}| = \sum_i |\nabla_i^\top \boldsymbol{\beta}|^{-1} (\nabla_i^\top \boldsymbol{\beta})^2$ , where  $\nabla_i \in \mathbb{R}^{p \times 1}$  such that  $\nabla = (\nabla_1^\top, \dots, \nabla_m^\top)^\top$ . We then formulate the generalised Lasso objective as a weighted ridge regression:

$$\|\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\|_2^2 + \lambda \|\nabla\boldsymbol{\beta}\|_1 = \|\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\|_2^2 + \lambda \|\boldsymbol{\Omega}^{1/2}\nabla\boldsymbol{\beta}\|_2^2,$$

where, now,  $\boldsymbol{\Omega} = \operatorname{diag}(|\nabla_i^\top \boldsymbol{\beta}|^{-1})_{i=1}^m \in \mathbb{R}^{m \times m}$  and we thus propose the IRLS scheme for the generalised Lasso:

$$\begin{aligned} \boldsymbol{\beta}^{(\ell+1)} &\leftarrow (\mathbf{X}^\top \mathbf{X} + \lambda \nabla^\top \boldsymbol{\Omega}^{(\ell)} \nabla)^{-1} \mathbf{X}^\top \mathbf{y}, \\ \boldsymbol{\Omega}^{(\ell)} &\leftarrow \operatorname{diag}(|\nabla_i^\top \boldsymbol{\beta}^{(\ell)}| + \delta)^{-1}_{i=1}^m. \end{aligned} \quad (2)$$

Unlike Tibshirani's generalised Lasso [14], IRLS does not yield the full solution path (for all  $\lambda$ ). However, in practice, it is common to pick a suitable  $\lambda$  and solve over the  $\boldsymbol{\beta}$  coefficients. Either  $\lambda$  is chosen according to some theoretical criterion or is 'learnt' via cross-validation or other means.

## 3. ROBUST HURST ESTIMATION

It is well known that the least-squares based estimators are vulnerable to outliers. Just a small number of anomalous measurements or noise artefacts can render poor estimates. Even when just a minority of data points lie significantly away from the model assumptions, robust approaches become attractive. In this scenario the assumption that the residuals are Gaussian, which gives rise to the least-squares loss, must be dropped in favour of a more heavier-tailed distribution. Typically, it is assumed that the density can be written in the form  $f(\epsilon) \propto \exp -\rho(\epsilon)$ . Assuming the noise is iid, minimisation of the negative log-likelihood yields the problem  $\operatorname{argmin}_{\boldsymbol{\beta}} \sum_i \rho(y_i - \mathbf{x}_i^\top \boldsymbol{\beta})$ , where  $\mathbf{X} = (\mathbf{x}_1^\top, \dots, \mathbf{x}_n^\top)^\top$ . Differentiating with respect to  $\boldsymbol{\beta}$  and setting to zero gives  $\sum_i \psi(\epsilon_i) \mathbf{x}_i^\top = 0$ , where the so-termed score function  $\psi$  is the derivative of the cost function:  $\psi = \rho'$ . Setting  $w_i := \psi(\epsilon_i)/\epsilon_i$  then yields the solution  $\boldsymbol{\beta} = (\mathbf{X}^\top \mathbf{W} \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{W} \mathbf{y}$ , with the weight matrix  $\mathbf{W} = \operatorname{diag}(w_i)_{i=1}^n$ . Here, again, IRLS can be used; the algorithm is initialised with the ordinary least squares solution. The weight matrix is then updated before the next iteration of the solution is computed and so on:

$$\begin{aligned} \boldsymbol{\beta}^{(\ell+1)} &\leftarrow (\mathbf{X}^\top \mathbf{W}^{(\ell)} \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{W}^{(\ell)} \mathbf{y}, \\ \boldsymbol{\epsilon}^{(\ell)} &\leftarrow \mathbf{y} - \mathbf{X}\boldsymbol{\beta}^{(\ell)}, \quad \mathbf{W}^{(\ell)} \leftarrow \operatorname{diag} \left( \frac{\psi(\epsilon_i^{(\ell)})}{\epsilon_i^{(\ell)}} \right)_{i=1}^n \end{aligned} \quad (3)$$

### 3.1. Robust Lasso

The robust Lasso or M-Lasso results from combining the robust cost function  $\rho$  with the sparsifying  $\ell_1$ -norm penalty:

**Table 1.** The family of the proposed robust, generalised Lasso problems

Condition	Problem
$\lambda \neq 0, \mathbf{W}, \nabla \neq \mathbf{I}$ or $\mathbf{0}$	proposed generalised M-Lasso
$\nabla = \mathbf{I}$	M-Lasso
$\mathbf{W} = \mathbf{I}$	(IRLS) generalised Lasso
$\nabla = \mathbf{W} = \mathbf{I}$	Lasso
$\lambda = 0, (\nabla = \mathbf{0})$	robust least-squares
$\lambda = 0, (\nabla = \mathbf{0}), \mathbf{W} = \mathbf{I}$	ordinary least-squares

$$\operatorname{argmin}_{\beta} \rho(\mathbf{y} - \mathbf{X}\beta) + \lambda \|\beta\|_1.$$

Since both robust least-squares and the Lasso can be solved via IRLS it is quite natural to form a combination of the two (cf. (1) and (3)) to solve M-Lasso. See, e.g., Zhang et al [16].

$$\begin{aligned} \beta^{(\ell+1)} &\leftarrow (\mathbf{X}^\top \mathbf{W}^{(\ell)} \mathbf{X} + \lambda \Omega^{(\ell)})^{-1} \mathbf{X}^\top \mathbf{W}^{(\ell)} \mathbf{y}, \\ \epsilon^{(\ell)} &\leftarrow \mathbf{y} - \mathbf{X}\beta^{(\ell)}, \quad \mathbf{W}^{(\ell)} \leftarrow \operatorname{diag} \left( \frac{\psi(\epsilon_i^{(\ell)})}{\epsilon_i^{(\ell)}} \right)_{i=1}^n, \\ \Omega^{(\ell)} &\leftarrow \operatorname{diag} (|\beta_i^{(\ell)} + \delta|^{-1})_{i=1}^p. \end{aligned} \quad (4)$$

### 3.2. Robust generalised Lasso

We propose the combination of our IRLS scheme (2) with (3) to solve the robust, generalised Lasso or generalised M-Lasso:

$$\operatorname{argmin}_{\beta} \rho(\mathbf{y} - \mathbf{X}\beta) + \lambda \|\nabla \beta\|_1,$$

with the iterations:

$$\begin{aligned} \beta^{(\ell+1)} &\leftarrow (\mathbf{X}^\top \mathbf{W}^{(\ell)} \mathbf{X} + \lambda \nabla^\top \Omega^{(\ell)} \nabla)^{-1} \mathbf{X}^\top \mathbf{W}^{(\ell)} \mathbf{y}, \\ \epsilon^{(\ell)} &\leftarrow \mathbf{y} - \mathbf{X}\beta^{(\ell)}, \quad \mathbf{W}^{(\ell)} \leftarrow \operatorname{diag} \left( \frac{\psi(\epsilon_i^{(\ell)})}{\epsilon_i^{(\ell)}} \right)_{i=1}^n, \\ \Omega^{(\ell)} &\leftarrow \operatorname{diag} (|\nabla_i^\top \beta^{(\ell)} + \delta|^{-1})_{i=1}^m. \end{aligned} \quad (5)$$

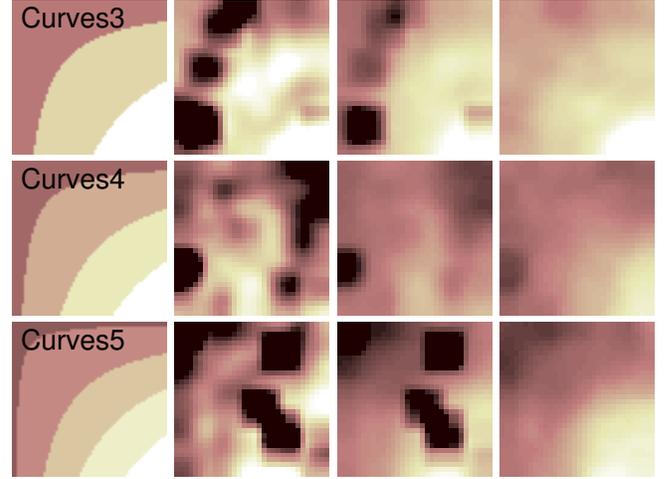
Note that this model and scheme can be thought of as a generalisation of the M-Lasso and the (IRLS) generalised Lasso. Indeed, it collapses down to the less general versions according to the Table 1. The  $\Omega$  facilitates the iterative procedure. The  $\nabla$  matrix spatially regularises the solution and exploits the assumption that the Hurst parameter varies slowly or as a piecewise constant function. The weights  $\mathbf{W}$  downweight and mitigate outliers in the wavelet coefficient magnitudes  $\mathbf{y}$  caused by band-limited noise or any other pseudo-periodic content which interferes with the power-law assumptions. Such artefacts (cf. top-centre of Fig. 2; and Section 4) are typical of textures which present edges or contours that may obscure the signal/object of interest. For instance, see the elephant in long-grass example in [22] or the mine hunting in the presence of sand ripples using sonar imagery example in [19].

## 4. EXPERIMENTS

The robust Lasso was compared to closely associated methods on a Hurst estimation, and denoising task. In both cases, the robustness to outliers was tested.

**Table 2.** Mean absolute error (and standard deviation) of the Hurst estimators: OLS, Lasso, and M-Lasso for data sets ‘CurvesQ’ ( $Q = 3, 4, 5$ ), with noise prob  $p$  (%) and var  $\sigma^2$ .

$p$	Method	$\sigma^2 = 0$	$\sigma^2 = 10$	$\sigma^2 = 1000$
$Q = 3$				
1	Ols	0.265 (0.331)	0.283 (0.338)	0.353 (0.401)
	Lasso	0.167 (0.186)	0.194 (0.193)	0.262 (0.209)
	M-Lasso	0.170 (0.195)	0.181 (0.187)	0.188 (0.193)
5	Ols	0.265 (0.331)	0.344 (0.363)	0.721 (0.601)
	Lasso	0.167 (0.186)	0.255 (0.199)	0.665 (0.317)
	M-Lasso	0.170 (0.195)	0.223 (0.207)	0.238 (0.228)
$Q = 4$				
1	Ols	0.253 (0.282)	0.266 (0.287)	0.350 (0.365)
	Lasso	0.197 (0.182)	0.217 (0.192)	0.301 (0.217)
	M-Lasso	0.184 (0.189)	0.200 (0.196)	0.216 (0.196)
5	Ols	0.253 (0.282)	0.321 (0.292)	0.778 (0.589)
	Lasso	0.197 (0.182)	0.281 (0.178)	0.754 (0.373)
	M-Lasso	0.184 (0.189)	0.230 (0.183)	0.262 (0.210)
$Q = 5$				
1	Ols	0.254 (0.270)	0.285 (0.284)	0.383 (0.380)
	Lasso	0.206 (0.182)	0.242 (0.188)	0.340 (0.227)
	M-Lasso	0.189 (0.183)	0.214 (0.187)	0.225 (0.193)
5	Ols	0.254 (0.270)	0.363 (0.305)	0.764 (0.576)
	Lasso	0.206 (0.182)	0.331 (0.196)	0.745 (0.367)
	M-Lasso	0.189 (0.183)	0.273 (0.196)	0.275 (0.217)



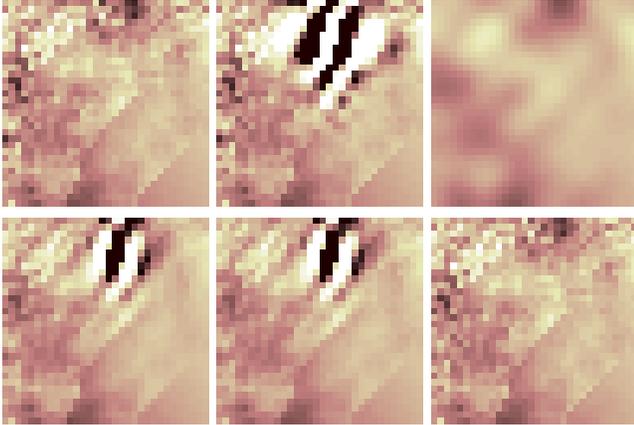
**Fig. 1.** Hurst estimates of three fractional Brownian surfaces. 1st column: true Hurst; 2nd: OLS; 3rd: Lasso, 4th: M-Lasso. Examples chosen have comparable performance to the respective means over the relevant parameters (see Table 2), namely  $p = 0.01, \sigma^2 = 1000$ ;  $p = 0.05, \sigma^2 = 10$ ;  $p = 0.01, \sigma^2 = 1000$  for curves ‘3’, ‘4’, and ‘5’ respectively.

### 4.1. Hurst estimation

Hurst estimation, used in wide-ranging applications [1, 2, 4, 5], offers a measure of long-range dependency, or texture, present in data. An adaptation [19] of the incremental Fourier synthesis method [23, 24] was used to synthesise fractional Brownian surfaces with a piecewise varying Hurst parameter. One hundred instantiations of three image types, of varying complexity, were simulated. The ground truth (of the underlying Hurst

**Table 3.** Sum of absolute reconstruction error (and standard deviation) using various denoising methods for data set ‘Curves4’

$\sigma^2$	Ols	Hard	Lasso	M-Lasso
	$Q = 3$			
10	1.124 (0.257)	2.192 (0.172)	1.130 (0.261)	0.416 (0.288)
1000	15.586 (10.291)	2.167 (0.130)	15.603 (10.124)	0.427 (0.141)
	$Q = 4$			
10	1.329 (0.287)	2.509 (0.210)	1.244 (0.327)	0.620 (0.281)
1000	18.345 (14.431)	2.515 (0.174)	18.534 (14.574)	0.632 (0.448)
	$Q = 5$			
10	1.393 (0.286)	2.592 (0.261)	1.395 (0.289)	0.625 (0.319)
1000	25.181 (40.091)	2.565 (0.186)	21.768 (23.568)	0.678 (0.385)



**Fig. 2.** Denoising example for  $\sigma^2 = 1000$ ,  $p = 0.01$ . Top row: original image, noisy image, hard thresholded reconstruction; bottom row: reconstruction using OLS, Lasso, and M-Lasso weight methods.

parameter) is shown in the first column of Fig. 1 for each type.

Three methods were used to estimate the Hurst parameter from dual-tree ([25]) wavelet coefficients: ordinary least-squares (OLS), (generalised) Lasso, and the proposed (generalised) M-Lasso. For the latter, we set the score  $\psi$  to be the Hampel’s three part redescending function and, as suggested by [16], fixed the parameters thereof to  $\{1.96, 2.24, 2.58\} \times \hat{\sigma}_\rho$ , where we compute the estimate  $\hat{\sigma}_\rho = \text{median}(|y_i - y_{i-1}|)_{i=2}^n / (0.6745\sqrt{2})$ . To emphasise the importance of locality, we used image patches of size  $32 \times 32$  and fixed  $k_- = 2, k_+ = 4$  to discard the finest and coarsest scales. The best results (over  $\lambda$ ) were reported for both the Lasso methods. In practice this can be learned via cross-validation or set according to how rapidly one expects the piecewise Hurst parameter to vary. Outlier, multiplicative noise was added to a proportion  $p$  of the 2nd finest scale level coefficients according to  $d_{2,\theta}^\sim = (1 + \mathbb{I}_{u < p} \varepsilon) d_{2,\theta}$ , with  $u \sim \mathcal{U}(0, 1)$ ,  $\varepsilon \sim \mathcal{N}(0, \sigma^2)$ , and  $d_{k,\theta} := (\mathcal{W}z)(\cdot; k, \theta)$ . Here  $p$  controls the probability of a coefficient being corrupted by an outlier and  $\sigma$  determines the probability of the outlier being large. Figure 1 illustrates example estimates on the Curves3-5 data. One can see that the outliers manifest themselves as black blobs in the Hurst estimates and that the least-squares estimates are especially sensitive. The Lasso has some limited success in smoothing these artefacts away but the M-Lasso

is notably superior. Table 2 confirms the advantage of the proposed generalised M-Lasso over the other methods grows as the outliers become more significant (as either  $p$  or  $\sigma$  is increased). At small outlier variance it does no worse.

## 4.2. Denoising

The Hurst parameter has been exploited in a variety of denoising schemes. Naornita et al [13] applied a framework based on that of Echelard and Lévy-Véhel [6]. The scheme adaptively shrinks finer-level wavelet coefficients if they lie above the power-law decay, as estimated by the log-slope (i.e. Hurst parameter) of the coarser wavelet coefficients.

However, in our example, the image patch is smaller ( $32 \times 32$ ). Furthermore, (outlier) noise is added to the 2nd finest scale level, the finest level has very low SNR, and the coarsest level is very poorly localised. In this case, the best option is to use only the second and third finest scales to estimate the slope. Unfortunately, our experiments revealed that this is highly vulnerable to noise. On the other hand, the ‘robustified’ Hurst estimator proposed here can use the 2nd-4th finest scale levels to estimate the slope. The outliers in the 2nd finest level are automatically downweighted by the weight matrix  $\mathbf{W}$ . In this way, the weight matrix which is iteratively derived during the estimation procedure, is directly used to shrink the wavelet coefficients in the 2nd finest scale. After shrinkage, the inverse wavelet transform is computed to reconstruct the image. Note that the weights  $\mathbf{W}$  (and, hence, this denoising scheme) is only available in the IRLS-type scheme.

A typical comparative example is illustrated in Fig 2. The band-limited noise can be clearly seen in the stripy patch in the centre top of the image. Performing hard thresholding (zeroing all coefficients in the 2nd finest level) results in a blurry image. The generalised Lasso method of Naornita et al [13] does little better than OLS at this noise value ( $\sigma^2 = 1000$ ) whereas using the weight matrix derived from generalised M-Lasso, we see that the reconstruction closely matches the original image. Table 3 summarises the results over relatively small and large outlier variance and for the three image types. Lasso and OLS perform respectably only for relatively small noise levels ( $\sigma^2 = 10$ ) but are extremely poor for large noise ( $\sigma^2 = 1000$ ) where even hard thresholding is better. The proposed M-Lasso weight matrix approach holds a small but significant advantage at low noise and is clearly superior for large  $\sigma$ .

## 5. CONCLUSION

A new Hurst estimator has been developed which combines the robustness of M-estimators with spatial regularisation. This results in an estimator which performs significantly better in the presence of strong outliers and no worse than least-squares and generalised Lasso for small, or no, outlier noise.

Further work could include the consideration of other M-estimator functions, a more rigorous treatment of the regression statistics including confidence intervals, results on the asymptotics of the estimators discussed, and a relaxation of the piecewise model assumptions.

## 6. REFERENCES

- [1] I. Zachevsky and Y. Y. Zeevi, "Single-image superresolution of self-similar textures," *IEEE International Conference on Image Processing*, 2013.
- [2] P. Abry, H. Wendt, and S. Jaffard, "When Van Gogh meets Mandelbrot: Multifractal classification of painting's texture," *Signal Processing*, vol. 93, no. 3, pp. 554–572, 2013.
- [3] N. Pustelnik, H. Wendt, P. Abry, and N. Dobigeon, "Local regularity, wavelet leaders and total variation based procedures for texture segmentation," arXiv:1504.05776, 2015.
- [4] J. Blackledge and D. Dubovitski, "A surface inspection machine vision system that includes fractal analysis," *International Society for Advanced Science and Technology, Journal of Electronics and Signal Processing*, vol. 3, no. 2, pp. 76–89, 2008.
- [5] J. D. B. Nelson and N. G. Kingsbury, "Fractal dimension based sand ripple suppression for mine hunting with sidescan sonar," *International Conference on Synthetic Aperture Sonar and Synthetic Aperture Radar*, 2010.
- [6] A. Echelard and J. Lévy Véhel, "Wavelet denoising based on local regularity information," *Proceedings of the European Signal Processing Program*, 2008.
- [7] P. Flandrin, P. Gonçalves, and G. Rilling, "Detrending and denoising with empirical mode decompositions," in *Proceedings of the European Signal Processing Program*, 2004, pp. 1581–1584.
- [8] B. D. Vidakovic, G. G. Katul, and J. D. Albertson, "Multiscale denoising of self-similar processes," *Journal of Geophysical Research*, vol. 105, no. D22, pp. 27049–27058, 2000.
- [9] P. Abry, P. Gonçalves, and J. Lévy Véhel, *Scaling Fractals and wavelets*, Wiley, 2009.
- [10] P. Abry, P. Flandrin, M. S. Taqqu, and D. Veitch, *Self-similarity and long-range dependence through the wavelet lens*, pp. 527–556, Birkhäuser, 2002.
- [11] S. Jaffard, B. Lashermes, and P. Abry, "Wavelet leaders in multifractal analysis," in *Wavelet Analysis and Applications*, M. I. Vai T. Qian and X. Yuesheng, Eds., Applied and Numerical Harmonic Analysis, pp. 201–246. Birkhäuser, 2007.
- [12] A. Pižurica, W. Philips, I. Lemahieu, and M. Acheroy, "A joint inter- and intrascale statistical model for bayesian wavelet based image denoising," *IEEE Trans. Image Processing*, vol. 11, no. 5, pp. 545–557, 2002.
- [13] C. Naornita, A. Isar, and J. D. B. Nelson, "Regularised, semi-local hurst estimation via generalised lasso and dual-tree complex wavelets," *IEEE International Conference on Image Processing*, pp. 2689–2693, 2014.
- [14] R. J. Tibshirani and J. Taylor, "The solution path of the generalised Lasso," *Annals of Statistics*, vol. 39, no. 3, pp. 1335–1371, 2010.
- [15] J. Park and C. Park, "Robust estimation of the hurst parameter and selection of an onset scaling," *Statistica Sinica*, vol. 19, pp. 1531–1555, 2009.
- [16] Z. G. Zhang, S. C. Chang, Y. Zhou, and Y. Hu, "Robust linear estimation using m-estimation and weighted l1 regularization: Model selection and recursive implementation," *IEEE International Symposium on Circuits and Systems*, pp. 1193–1196, 2009.
- [17] J. Fan and R. Li, "Variable selection via nonconcave penalized likelihood and its oracle properties," *Journal of the American Statistical Association*, vol. 96, no. 456, pp. 1348–1360, 2001.
- [18] V. Roth, "The generalized lasso," *IEEE Transactions on Neural Networks*, vol. 15, no. 1, pp. 16–28, 2004.
- [19] J. D. B. Nelson and N. G. Kingsbury, "Dual-tree wavelets for estimation of locally varying and anisotropic fractal dimension," *IEEE International Conference on Image Processing*, pp. 341–344, 2010.
- [20] C. Chen, J. Huang, L. He, and H. Li, "Preconditioning for accelerated iteratively reweighted least squares in structured sparsity reconstruction," *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2713–2720, 2014.
- [21] I. Daubechies, R. DeVore, M. Fornasier, and S. Gunturk, "Iteratively reweighted least squares minimization for sparse recovery," *Communications on Pure and Applied Mathematics*, vol. 63, no. 1, pp. 1–38, 2010.
- [22] G. Papari, P. Campisi, N. Petkov, and A. Neri, "A biologically motivated multiresolution approach to contour detection," *EURASIP Journal on Advances in Signal Processing*, 2007.
- [23] L. M. Kaplan and C. C. J. Kuo, "An improved method for 2-d self-similar image synthesis," *IEEE Transactions on Image Processing*, vol. 5, no. 5, pp. 754–761, 1996.
- [24] Inria Saclay/Ecole Centrale de Paris Regularity Team, "Fraclab Toolbox," <http://fraclab.saclay.inria.fr>, 2014.
- [25] I. W. Selesnick, R. G. Baraniuk, and N. G. Kingsbury, "The dual-tree complex wavelet transform," *IEEE Signal Processing Magazine*, vol. 22, no. 6, pp. 123–151, 2005.