Estimation of the noise level function based on a non-parametric detection of homogeneous image regions*

Camille Sutour[†‡], Charles-Alban Deledalle[†], and Jean-François Aujol[†]

Abstract. We propose a two-step algorithm that automatically estimates the noise level function of stationary noise from a single image, i.e., the noise variance as a function of the image intensity. First, the image is divided into small square regions and a non-parametric test is applied to decide weather each region is homogeneous or not. Based on Kendall's τ coefficient (a rank-based measure of correlation), this detector has a non-detection rate independent on the unknown distribution of the noise, provided that it is at least spatially uncorrelated. Moreover, we prove on a toy example, that its overall detection error vanishes with respect to the region size as soon as the signal to noise ratio level is non-zero. Once homogeneous regions are detected, the noise level function is estimated as a second order polynomial minimizing the ℓ^1 error on the statistics of these regions. Numerical experiments show the efficiency of the proposed approach in estimating the noise level function, with a relative error under 10% obtained on a large data set. We illustrate the interest of the approach for an image denoising application.

Key words. Noise level estimation, non-parametric detection, signal-dependent noise, least absolute deviation

AMS subject classifications.

1. Introduction.

Most image processing applications, including denoising [6, 40], super-resolution [17], or feature extraction [31], require the knowledge of the noise level. Noise estimation from multiple images is an over-constrained problem that was solved in [19], but its estimation from a single image remains a challenging problem. We review in the following some of the main approaches that have been proposed for the estimation of the noise level when the noise is spatially uncorrelated. Methods dedicated to correlated noise can be found, e.g, in [13, 35, 27].

A first approach used for the estimation of the noise level consists in applying a linear transform that maps the image into a suitable domain where the signal and the noise are relatively well separated. The noise level can then be estimated by performing some statistics in its associated subspace. For instance, when the noise is additive white Gaussian, a popular estimator of the noise standard deviation is given by the mean absolute deviation of the finest wavelets coefficients, see e.g. [15]. Nevertheless, this approach tends to over-estimate the noise level as it assumes there is enough separation between the signal and the noise coefficients. Indeed, singularities such as edges usually produce high responses in all sub-bands of the transformed domain. To avoid this issue, some authors [33, 41] have proposed to pre-filter the image before estimating the noise in order to remove such high frequency components, e.g., using a Laplacian-based filter together with a Sobel edge detector [38]. Unfortunately, such approaches still lead to a subsequent over-estimation of the noise level. Rather than attempting to remove the high frequency components, some authors have proposed to take them into account by assuming some *prior* knowledge on the distribution of the signal coefficients. In [14], the authors use a method of "trained"

*This study has been carried out with financial support from the French State, managed by the French National Research Agency (ANR) in the frame of the "Investments for the future" Programme IdEx Bordeaux - CPU (ANR-10-IDEX-03-02). C. Sutour would like to thank the DGA and the Aquitaine region for funding her PhD. J.-F. Aujol acknowledges the support of the Institut Universitaire de France.

[†]IMB, CNRS, UMR 5251, Université de Bordeaux, Talence, France (jaujol,cdeledal@math.u-bordeaux.fr.).

[‡]LaBRI, Université de Bordeaux, Talence, France (camille.sutour@math.u-bordeaux.fr.).

moments from an assumption of Laplacian distribution of the signal coefficients, while the authors of [43] exploit the invariance of the kurtosis when assuming a generalized Gaussian distribution. Patch-based approaches have also been proposed for such a task, e.g., in [39, 30, 37]. While such techniques can be very powerful, they are nevertheless limited to constant noise levels. In real-world applications, the noise is signal-dependent and its level varies spatially according to the underlying intensity. Variance stabilizing transforms [1] can reduce the dependency between the signal intensity and the noise. The authors of [36] iteratively perform the estimation of Poisson-Gaussian noise by applying a variance stabilizing transform, then they estimate the noise variance of the stabilized noise using a principal component analysis. In [5], Poisson-Gaussian noise is stabilized using the generalized Anscombe transform. However, this transform requires a prior knowledge of the noise parameters, obtained by linear regression on areas with homogeneous variance [18]. The generalized Anscombe transform is also optimized for the stabilization of Poisson-Gaussian noise in [32].

Separation techniques have been extended to specific signal-dependent models, e.g., using a wavelet transform for a Poissonian-Gaussian model [16, 2] or using a Gaussian mixture model of patches for additive noise with affine variance [3]. Unfortunately, in the general case, it appears pointless to design a linear transform able to separate the noise from the signal. The noise can also be distinguished from the signal components by principal component analysis [12] or by selecting blocks with lowest variance [11]. An alternative consists in pre-filtering the image to analyze the residue as if it was the noise component. For instance, [20, 9] have proposed different pre-filtering techniques for some specific types of signal-dependent noise. However, for our general problem, as such pre-filtering techniques require a certain knowledge of the noise, this leads to a chicken and egg problem.

Another popular approach encountered in [8, 4] and that we follow here relies on the fact that natural images contain homogeneous areas, where the signal to noise ratio is very weak, so only the statistics of the noise intervene. The statistics of the noise can then be deduced by first detecting homogeneous areas and next estimating the noise parameters in those regions. However, classic detectors also require assumptions on the noise statistics (e.g, Gaussian), preventing their use when the nature of the noise is completely unknown.

Our main contribution is to propose a non-parametric detection of homogeneous areas that only requires the noise to be spatially uncorrelated. This means that the detection performance of our detector, based on Kendall's τ coefficient [22], is independent of the noise statistical distribution. In Section 3, we introduce our homogeneous detector based on the rank correlation. We show that the performance of the detector relies on the underlying signal to noise ratio, with an error rate that decreases with the number of samples available. Once homogeneous areas are detected, we estimate the noise level function (NLF) [29], i.e., the function of the noise variance with respect to the image intensities. The NLF estimation is detailed in Section 4. At this step we assume that the noise is signal-dependent with finite first and second order moments linked together by a second order polynomial. This can encompass a wide range of noise statistics encountered in image processing such as additive, multiplicative or Poissonian noise, as well as hybrid models. In the last step we perform a robust estimation of this polynomial NLF as the best ℓ^1 approximation of the mean/variance relationship of homogeneous regions. Section 5 details how the estimation is performed in practice, then Section 6 shows the efficiency of the detection of homogeneous areas and of the parameters estimation for different noise models, by comparing them to the state of the art. The interest of performing a reliable noise estimation is also illustrated with an application to image denoising: the non-local

means algorithm [6] is adapted to deal with hybrid noises, thanks to the estimated NLF parameters. Finally, Section 7 provides discussions and conclusions.

2. Context and notations. We denote by $g : \Omega \to \mathbb{R}$ an image defined on a 2D discrete domain $\Omega \subset \mathbb{Z} \times \mathbb{Z}$ such that g can be identified to an element of $\mathbb{R}^{N_0}$ with $N_0 = \#\Omega$ the number of pixels of the image. Hence, for an image $g \in \mathbb{R}^{N_0}$, $i \in [1, \dots, N_0]$ is the index of its i-th pixel and $g_i \in \mathbb{R}$ the value of g at pixel i.

We assume that g is an observation of a clean unknown image $g^0 \in \mathbb{R}^{N_0}$ corrupted by a spatially uncorrelated signal dependent noise. More precisely, we assume that g can be modeled as the realization of a random vector G that verifies

$$\mathbb{E}[G] = g^0 \quad \text{and} \quad \mathrm{Cov}[G] = \begin{pmatrix} \mathrm{NLF}(g_1^0) & & & 0 \\ & \mathrm{NLF}(g_2^0) & & \\ & & \ddots & \\ 0 & & & \mathrm{NLF}(g_{N_0}^0) \end{pmatrix} . \tag{2.1}$$

where $\mathrm{NLF} : \mathbb{R} \to \mathbb{R}^+$ is coined the noise level function. As $\mathrm{Cov}[G]$ is assumed to be diagonal, this model restricts to spatially uncorrelated noise. But since the i-th diagonal element depends on g_i^0, the noise is said to be signal dependent. Remark, that the same noise level function applies for each diagonal elements, so the noise is assumed to be stationary. Noise models with spatially varying noise level functions are out of the scope of this study.

Example 2.1 (Additive white Gaussian noise). *A classic example is to consider an additive white Gaussian noise such that $G = g^0 + \varepsilon$, with $\varepsilon \sim \mathcal{N}(0, \sigma^2 \mathrm{Id})$. In such a case, G satisfies eq. (2.1) with $\mathrm{NLF}(\cdot) = \sigma^2$ which reveals that the noise is signal independent.*

Example 2.2 (Poisson noise). *Another example is to consider a Poisson noise such that $g^0 \in \mathbb{N}$, $G_i^0 \sim \mathcal{P}(g_i^0)$ and, for all $i \neq j$, G_i and G_j are independent. In such a case, G satisfies eq. (2.1) with $\mathrm{NLF}(g_i^0) = g_i^0$ which reveals that the noise is signal dependent.*

Example 2.3 (Multiplicative noise). *Another example is to consider a multiplicative noise such that $G_i = g_i^0 \times S_i$, with $\mathbb{E}[S] = 1$ and $\mathbb{E}[SS^\top] = \gamma \mathrm{Id}$, $\gamma > 0$. In such a case, G satisfies eq. (2.1) with $\mathrm{NLF}(g_i^0) = \gamma (g_i^0)^2$ which reveals again that the noise is signal dependent. A typical example is the gamma noise of parameter $L \in \mathbb{N}^*$, for which $\gamma = \frac{1}{L}$, whose probability density function f is defined for $g_i \in \mathbb{R}^+$ as*

$$f(g_i \mid g_i^0, L) = \frac{L^L g_i^{L-1}}{\Gamma(L)(g_i^0)^L} \exp\left(-\frac{L g_i}{g_i^0}\right) . \tag{2.2}$$

The goal of our work is to estimate the unknown noise level function NLF. To do so, we consider K small disjoint blocks $(\omega_k)_{k \in [1,\dots,K]} \subset \Omega$ of size N such that $KN \leq N_0$. We denote by ω any of such block ω_k and by $b = g_\omega \in \mathbb{R}^N$ (resp. $B = G_\omega$ and $b^0 = g_\omega^0$) the restriction of g in ω (resp. the restriction of G and g^0 in ω). Our motivation relies on the fact that most natural images, provided that their resolution is high enough, contain small blocks ω for which b^0 is constant. In this case we say that B is homogeneous, and otherwise when b^0 varies, B is said to be inhomogeneous. Hence, according to eq. (2.1), the empirical expectation and variance inside such a block b should provide a punctual estimation of the noise level function.

In practice b^0 is unknown as well as the homogeneous blocks. We hence propose in Section 3, a statistical test that can decide weather a block is likely to be homogeneous or not without more assumptions regarding the noise model introduced in (2.1). Once homogeneous regions are detected, several punctual estimations of the noise level function are available. Section 4 describes how these punctual estimates can be used to robustly estimate a noise level function with few additional assumptions, by fitting a second order polynomial function to the scatterplot.

3. Detection of homogeneous areas.

The noise statistics can be estimated from several small homogeneous areas, or in fact areas that are close to be homogeneous, i.e. where the signal to noise ratio is weak. Indeed, in those regions the signal fluctuations can be assumed to be negligible compared to the noise fluctuations, so only the statistics of the noise intervene. Hence, provided that a large number of such regions has been detected, the statistics of each regions can be used to infer the underlying noise level function. The goal of this section is then to develop a method that automatically selects homogeneous areas in the image, based on the fact that natural images contain such regions.

3.1. Homogeneity and uncorrelation.

In the light of detection theory, we aim at designing a detector of inhomogeneous blocks that answers the following statistical hypothesis test from the observation b only

$$\begin{array}{ll} H_0 : B \text{ is homogeneous, i.e., } b^0 \text{ constant} & \text{(null hypothesis)}, \\ H_1 : B \text{ is inhomogeneous, i.e., } b^0 \text{ varies} & \text{(alternative hypothesis)}. \end{array} \tag{3.1}$$

As b^0 is unknown, we cannot directly provide an answer to this test. To alleviate this difficulty, we suggest making use of this first proposition.

Proposition 3.1. *Let $B \in \mathbb{R}^N$ be homogeneous. We denote by B_I the sub random vector of B whose entries are indexed by $I \subset \{1, \dots, N\}$. Then, for any disjoint subsets I and J of n elements of $\{1, \dots, N\}$, $2n \leq N$, $X = B_I$ and $Y = B_J$ are uncorrelated, i.e.:*

$$\mathrm{Corr}(X, Y) = 0 \ .$$

This proposition tells us that if two random disjoint sequences X and Y of a block are correlated, then the block is inhomogeneous. In practice, the random variables are unknown and only their single realizations, the noisy sequences $x = b_I$ and $y = b_J$, are available. As a consequence, the correlation between X and Y is also unknown and the arising question is "Are the realizations x and y significantly correlated?".

Hence, the goal is to design a score $s : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ that can be used to answer correctly to our hypothesis problem. More precisely, the score s should ensure that, for any constant $P_{\mathrm{FA}} \in [0, 1[$, there exists a threshold $\alpha > 0$ such that, whatever the distribution of X and Y, the followings are satisfied

$$\text{(C1)} \qquad \lim_{n \to \infty} \mathbb{P}(\underbrace{|s(X, Y)| \geq \alpha}_{\text{false alarms}} \mid H_0) = P_{\mathrm{FA}},$$

$$\text{(C2)} \qquad \lim_{n \to \infty} \mathbb{P}(\underbrace{|s(X, Y)| < \alpha}_{\text{miss detections}} \mid H_1) = 0.$$

The first condition is very important and states that as the size n gets large enough, the same threshold α can maintain the same probability of false alarms P_{FA} as soon as X and Y are uncorrelated (regardless of the distribution of X and Y). The test is said to be an asymptotically constant false-alarm rate (CFAR) detector. The second condition

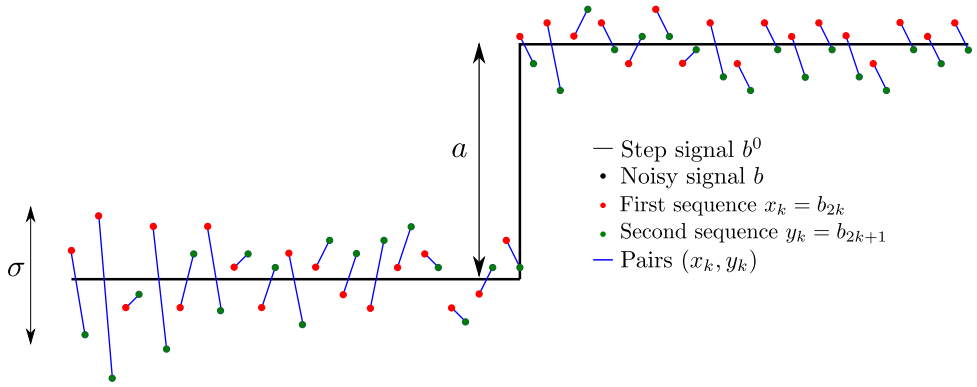


Figure 1. *Step signal associated to the alternative hypothesis H_1 (see Def. 3.9) and selection of the neighboring sequences.*

is natural and states that with the same threshold α the probability of miss detections vanishes as the size n of the sequences increases.

In the following, we propose to build such a statistical test between pairs of disjoint sequences of a block. As in practice, we cannot test every disjoint sequence, we focus on the correlation between neighboring sequences that we denote by $x = (b_{2k})$ and $y = (b_{2k+1})$, where $2k$ and $2k+1$ represent neighbor pixel indexes for a given scan path of the underlying block ω. An example of the construction of the two neighboring sequences $x = (b_{2k})$ and $y = (b_{2k+1})$ is displayed on Figure 1. If these two variables are found to be significantly correlated, this means that there is some dependencies between the pixels of the blocks and their neighbors, so we can assume that there is some structure and all fluctuations are not only due to noise.

A naive approach for answering the above hypothesis test is to use the empirical correlation between the two sequences x and y, i.e., $s(x, y) = \mathrm{Corr}(x, y)$. Unfortunately, the significance of such a test depends on the parameters of the random process that generates the observation, i.e., it cannot lead to a CFAR detection without extra assumptions on the distribution of X and Y. As our goal is to estimate the nature of the noise, it is important for our test of homogeneity not to rely on such parameters. We therefore consider a non-parametric approach whose statistical answer is independent of the noise model.

3.2. Non parametric test of independence.

As mentioned above, the empirical correlation test does not provide a CFAR detector as the threshold α would depend on the unknown parametrization of the random process that generates the observation. The key idea of non-parametric tests is to focus on the rank (i.e., on the relative order) of the values rather than on the values themselves. In this vein, a block is considered homogeneous if the ranking of the pixel values is uniformly distributed, regardless of the spatial organization of the pixels. Such correlation between the sequences x and y can be assessed based on the ranking of their values using Kendall's τ coefficient.

3.2.1. Kendall's τ coefficient.

The Kendall's τ coefficient is a rank correlation measure [22] that provides a non-parametric hypothesis test for statistical dependence. Before turning to the definition of the Kendall's τ coefficient, let us first introduce the definitions of concordant, discordant and tied pairs.

Definition 3.2. *Two pairs of observations (x_i, y_i) and (x_j, y_j) are said to be **concordant** if the ranks of both elements agree, i.e., if ($x_i < x_j$ and $y_i < y_j$) or ($x_i > x_j$ and $y_i > y_j$). They are said to be **discordant** if ($x_i < x_j$ and $y_i > y_j$) or if ($x_i > x_j$ and $y_i < y_j$). If $x_i = x_j$ or $y_i = y_j$, the pair is neither concordant nor discordant; it is said to be **tied**.*

We can now turn to the definition of the Kendall's τ coefficient as introduced in [22].

Definition 3.3. *Let $x \in \mathbb{R}^n$ and $y \in \mathbb{R}^n$ be two sequences without tied pairs. Kendall's $\tau : \mathbb{R}^n \times \mathbb{R}^n \to [-1, 1]$ coefficient is the quantity given by*

$$\tau(x, y) = \frac{n_c - n_d}{\frac{n(n-1)}{2}} = \frac{1}{n(n-1)} \sum_{1 \leq i,j \leq n} \operatorname{sign}(x_i - x_j) \operatorname{sign}(y_i - y_j), \tag{3.2}$$

where n_c is the number of concordant pairs and n_d the number of discordant pairs.

Remark that Kendall's $\tau(x, y)$ coefficient depends only on the relative order of the values of x and y. It lies in the range $-1 \leq \tau(x, y) \leq 1$. If the agreement between the two rankings is perfect, all pairs are concordant, then $\tau(x, y) = 1$. If the disagreement between the two rankings is perfect, all pairs are discordant, then $\tau(x, y) = -1$. A value $\tau(x, y) = 0$ indicates the absence of significant correlation between x and y.

Definition 3.3 is however limited to sequences without tied pairs. This restrict the measure of correlation to the case where the unknown random process is continuous, e.g., Gaussian, and the apparition of tied pairs is hence of zero measure. In our imaging context, for instance with optical systems, the image formation processes typically include a photon counting step that is by nature discrete. Then ties in the sequence might appear with non-zero probability so it is important to take them into account properly.

When the data suffers from tied pairs, the definition of Kendall's τ coefficient might be extended according to [23] as follows.

Definition 3.4. *Let $x \in \mathbb{R}^n$ and $y \in \mathbb{R}^n$ be two arbitrary sequences. Define from x and y the quantities*

$n_0 = n(n-1)/2$	*the total number of pairs,*
$n_1 = \sum_i t_i(t_i - 1)/2$	*the total number of tied values for the first quantity,*
$n_2 = \sum_j u_j(u_j - 1)/2$	*the total number of tied values for the second quantity,*
t_i	*the number of tied values for the i^{th} group for the first quantity,*
u_j	*the number of tied values for the j^{th} group for the second quantity.*

Kendall's $\tau : \mathbb{R}^n \times \mathbb{R}^n \in [-1, 1]$ coefficient is the quantity given by

$$\tau(x, y) = \frac{n_c - n_d}{\sqrt{(n_0 - n_1)(n_0 - n_2)}} \,. \tag{3.3}$$

Remark that in the absence of ties, Definition 3.4 matches with Definition 3.3. Equipped with such a measure of correlation, we are now ready to study its ability to answer our hypothesis test under the requirements (C1) and (C2).

3.2.2. Distribution of τ in homogeneous blocks (under the null hypothesis H_0). In this section, we study the distribution of Kendall's τ coefficient under the null hypothesis H_0 and show accordingly that it can lead to a statistical test satisfying (C1). In particular as it is based only on the relative order of the values of x and y, it provides a non-parametric measure of correlation leading to a CFAR detector.

We first give the distribution of τ in the absence of tied pairs.

Proposition 3.5. *Let X and Y be two arbitrary sequences under H_0 without tied pairs. The random variable $\tau(X, Y)$ has an expected value of 0 and a variance of $\frac{2(2n+5)}{9n(n-1)}$. Moreover, in case of large samples n, its distribution is approximated by the normal distribution. More precisely, for any sequences x and y, let $z : \mathbb{R}^n \times \mathbb{R}^n \in \mathbb{R}$ be the z-score defined as*

$$z(x, y) = \frac{3(n_c - n_d)}{\sqrt{n(n-1)(2n+5)/2}}. \tag{3.4}$$

The z-score is asymptotically distributed according to the standard normal distribution

$$z(X,Y) \underset{n\to\infty}{\sim} \mathcal{N}(0,1) \ .$$

The proof can be found in [24].

We now turn to the extension in the case of tied pairs.

Proposition 3.6. *For any sequences $x \in \mathbb{R}^n$ and $y \in \mathbb{R}^n$, let $z : \mathbb{R}^n \times \mathbb{R}^n \in \mathbb{R}$ be the z-score defined as*

$$z(x,y) = \frac{n_c - n_d}{\sqrt{v}} \tag{3.5}$$

with

$$\begin{array}{rcl} v & = & (v_0 - v_t - v_u)/18 + v_1 + v_2 \\ v_0 & = & n(n-1)(2n+5) \\ v_t & = & \sum_i t_i(t_i-1)(2t_i+5) \\ v_u & = & \sum_j u_j(u_j-1)(2u_j+5) \\ v_1 & = & \sum_i t_i(t_i-1)\sum_j u_j(u_j-1)/(2n(n-1)) \\ v_2 & = & \sum_i t_i(t_i-1)(t_i-2)\sum_j \frac{u_j(u_j-1)(u_j-2)}{9n(n-1)(n-2)}. \end{array} \ .$$

Then, for any sequences X and Y under H_0, the z-score is asymptotically distributed according to the standard normal distribution

$$z(X,Y) \underset{n\to\infty}{\sim} \mathcal{N}(0,1) \ .$$

The proof can be found in [23]. A direct consequence of Proposition 3.6 is the following.

Corollary 3.7. *Let X and Y be two arbitrary sequences under H_0, then*

$$\lim_{n\to\infty} \mathbb{P}(z(X,Y) > \alpha \mid H_0) = \int_\alpha^{+\infty} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{t^2}{2}\right) \mathrm{d}t = 1 - \phi(\alpha)\,, \tag{3.6}$$

where ϕ is the cumulative distributive function (cdf) of the Gaussian distribution

$$\phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} \mathrm{e}^{-\frac{t^2}{2}} \mathrm{d}t. \tag{3.7}$$

The proof is immediate and comes from the definition of the probability of false-alarm given in (C1), combined with the normal distribution of $z(X,Y)$ under H_0.

We now turn to the first theorem that shows that the z-score satisfies the condition (C1) as its distribution does not rely on any assumption regarding the distribution of X and Y.

Theorem 3.8. *Let X and Y be two arbitrary sequences under H_0. Let $P_{FA} \in [0,1[$ and choose*

$$\alpha = \phi^{-1}(1 - P_{FA}/2) = \sqrt{2}\,\mathrm{erf}^{-1}(1 - P_{FA}) \ , \tag{3.8}$$

where erf *is the Gauss error function,* $\mathrm{erf}(x) = \frac{1}{\sqrt{\pi}} \int_{-x}^{x} \mathrm{e}^{-t^2} \mathrm{d}t$*, then*

$$\lim_{n\to\infty} \mathbb{P}(|z(X,Y)| > \alpha \mid H_0) = P_{FA} \ . \tag{3.9}$$

The proof is given in appendix A. Besides, apart from the standard distribution approximation, the z-score still satisfies the CFAR condition (C1) for low numbers of samples n, as the distribution remains the same whatever the noise distribution.

In practice, the detection is performed by computing the associated p-value

$$p(x,y) = 2 - 2\phi(|z(x,y)|), \tag{3.10}$$

i.e., the probability under the null hypothesis of having $|z(X,Y)|$ greater than $|z(x,y)|$.

We then reject the null hypothesis if the p-value is smaller than the predetermined significance level of false alarm P_{FA}. In general, the smaller the p-value, the larger the significance of the rejection of the null hypothesis H_0.

3.2.3. Distribution of τ subject to a step-signal (an alternative hypothesis H_1). The computation of the distribution of τ under the null hypothesis reflects the non-parametric properties of Kendall's test, since no assumption on the distribution of X and Y (except the absence of spatial correlation) has been made.

When X and Y are somehow correlated, a model of dependence has to be assumed in order to derive some statistical properties for τ. Accordingly, we propose to study the ability of Kendall's τ coefficient to accept or to reject the alternative hypothesis H_1 when X and Y correspond to sequences extracted from a 1D step signal corrupted by additive white Gaussian noise. See Fig. 1 for the illustration of such a signal.

Definition 3.9. *Let n be an even size of sequence and H_1 be the hypothesis of dependence between X and Y parametrized by $a > 0$ and σ^2 as follows*

$$\forall\ 1 \le k \le n, \quad X_k = \begin{cases} \varepsilon_k & \text{if} \quad k \le n/2 \\ a + \varepsilon_k & \text{if} \quad k > n/2 \end{cases} \quad \text{and} \quad Y_k = \begin{cases} \eta_k & \text{if} \quad k \le n/2 \\ a + \eta_k & \text{if} \quad k > n/2 \end{cases},$$

where ε and η are independent random variables distributed along $\mathcal{N}(0, \sigma^2)$.

Proposition 3.10. *Let X and Y be two sequences satisfying H_1. Then*

$$\mathbb{E}[\tau(X,Y) \mid H_1] = \frac{n}{2(n-1)} \left(1 - 2\phi\left(\frac{a}{\sqrt{2}\sigma}\right)\right)^2, \tag{3.11}$$

and in particular

$$\lim_{\frac{a}{\sigma} \to +\infty, n \to +\infty} \mathbb{E}[\tau(X,Y) \mid H_1] = \frac{1}{2}. \tag{3.12}$$

The proof is detailed in appendix B. This proposition shows that, asymptotically to the dimension n, the expected value of τ in the case of a step tends towards $1/2$ when the signal to noise ratio (a versus σ) is large enough. In other words, since the expected value of τ under the null hypothesis is zero, the two distributions are expected to be well separated as soon as the sequences are large enough and the signal to noise ratio is sufficient. It remains to show that the variance of the sampling distribution of τ decreases with the length of the sequence.

Proposition 3.11. *Let X and Y be two sequences satisfying H_1. Then*

$$\operatorname{Var}[\tau(X,Y) \mid H_1] = O\left(\frac{1}{n}\right).$$

The proof is detailed in appendix C. This proposition guarantees that as the number of samples n increases, the variance of the sample distribution of τ under H_1 decreases towards zero. Since the expected value of τ under H_1 is strictly positive for $a > 0$, this statement provides that for a sufficient number of samples, the distributions of τ under both cases can be distinguished.

The next corollary extends this result to the case of the z-score defined in (3.4),(3.5).

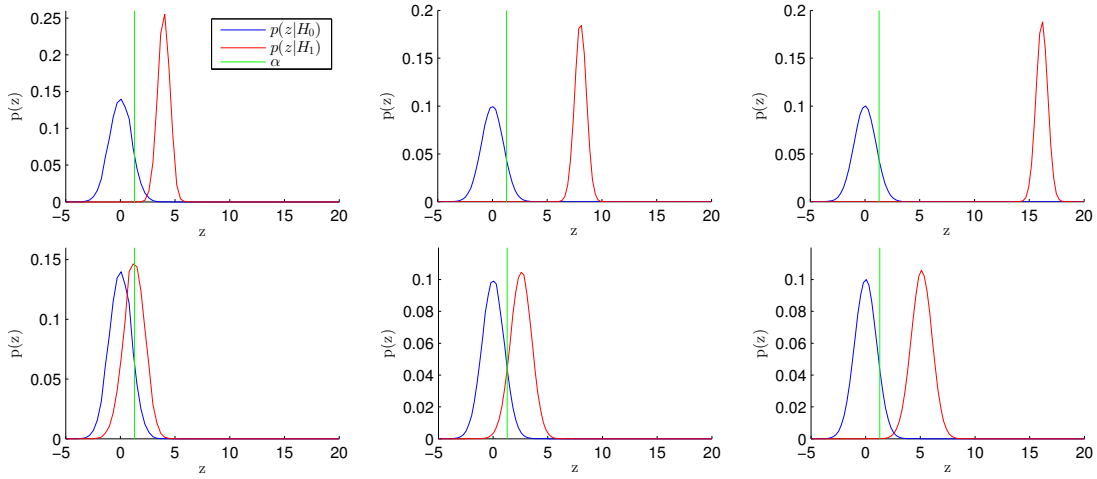


Figure 2. *Empirical distributions of Kendall's z-score under the null hypothesis H_0 (blue) and under the H_1 hypothesis of statistical dependence (red) for a growing number of samples ($n = 32, 128, 512$). The threshold α associated to the probability of false alarm $P_{FA} = 0.15$ is displayed as a green line. On the top line, the noise power is small ($\sigma = 40$) compared to the step size ($a = 128$), resulting in a signal to noise ratio $a/\sigma = 3.2$, whereas on the bottom line, the noise power is stronger ($\sigma = 60$) compared to the step size ($a = 64$), resulting in a lower signal to noise ratio $a/\sigma = 1.07$, which makes the detection trickier.*

Corollary 3.12. *Let X and Y be two sequences satisfying H_1. Then*

$$\mathbb{E}[z(X,Y) \mid H_1] = O(\sqrt{n}) \quad \textit{and} \quad \mathrm{Var}[z(X,Y) \mid H_1] = O(1). \tag{3.13}$$

The proof is detailed in appendix D.

As the distribution of z under H_0 is zero mean with a standard deviation of 1, this corollary guarantees that as the number of samples n increases, the z-score can be used to discriminates between H_0 and H_1. Indeed, as a consequence of Chebyshev's inequality, the next theorem shows that the z-score satisfies condition (C2).

Theorem 3.13. *Let X and Y be two sequences satisfying H_1. For any $\alpha > 0$, we have*

$$\mathbb{P}(z(X,Y) < \alpha \mid H_1) \leq \frac{\mathrm{Var}(z(X,Y)|H_1)}{(\mathbb{E}[z(X,Y)|H_1] - \alpha)^2} = O\left(\frac{1}{n}\right), \tag{3.14}$$

and in particular, as $\mathbb{P}(|z(X,Y)| < \alpha \mid H_1) \leq \mathbb{P}(z(X,Y) < \alpha \mid H_1)$, we have

$$\mathbb{P}(|z(X,Y)| < \alpha \mid H_1) = O\left(\frac{1}{n}\right). \tag{3.15}$$

The proof is detailed in appendix E.

Illustrations. Figure 2 displays the empirical distributions of Kendall's z-score under the null hypothesis H_0 (blue) and the alternative hypothesis H_1 of statistical dependence in case of a step (red) for a growing number of samples ($n = 32, 128, 512$). The threshold α associated to the probability of non detection $P_{\mathrm{ND}} = 0.15$ is displayed as a green line. On the top line, the noise power is small ($\sigma = 40$) compared to the step size ($a = 128$) so the signal to noise ratio $a/\sigma = 3.2$ is important, whereas on the bottom line, the noise power is stronger ($\sigma = 60$) compared to the step size ($a = 64$), resulting in a lower signal to noise ratio $a/\sigma = 1.07$, which makes the detection trickier. These empirical distributions show that the error rate is related to the signal to noise ratio, i.e. the relationship between the noise power σ and the step size a, due to the proximity of both distributions. Indeed, on the bottom line where both values are close, according to Proposition 3.10 the expected

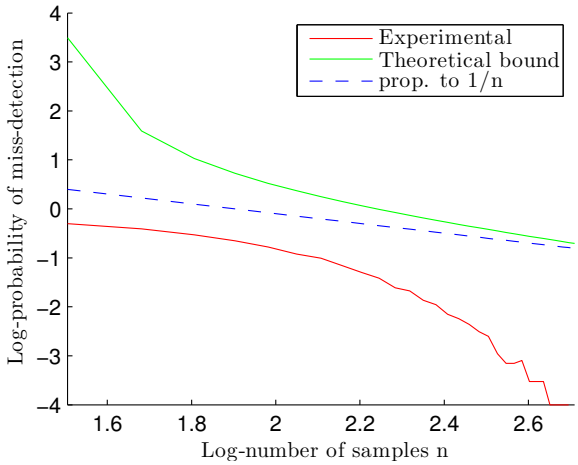


Figure 3. *Evolution of the empirical probability of miss-detections as a function of the number of samples n, with probability of false alarm $P_{FA} = 0.15$. Displayed in a log-log scale, the bound in $1/n$ is verified.*

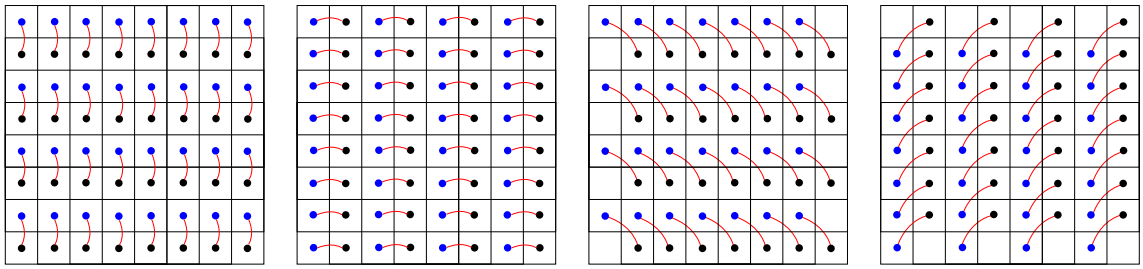


Figure 4. *Neighborhood selection for the independence tests.*

value of τ, hence z, under H_1 is close to zero, hence to the expected value under H_0, which makes the two distributions harder to distinguish. However, when the number of samples n increases, Corollary 3.12 guarantees that the expected value of z under H_1 grows towards the infinity, enabling to tell them apart more precisely.

Figure 3 illustrates condition (C2). The empirical probability of miss-detection is displayed as a function of the number of samples, in a log-log scale. The theoretical bound we derived is shown to decrease asymptotically as $1/n$, while the empirical probability of miss-detection seems to offer an even better decrease rate.

3.3. Extension to homogeneous areas in images. The above study shows that Kendall's coefficient is able to distinguish uniform regions from areas that contain an edge. Based on the fact that natural images can be approximated by a piece-wise constant model, this guarantees that homogeneous regions can be detected, and the error is controlled by the number of samples and the signal to noise ratio.

In practice, we run $K = 4$ Kendall's τ test for four pair of sequences $(x^{(1)}, y^{(1)})$, $(x^{(2)}, y^{(2)})$, $(x^{(3)}, y^{(3)})$, $(x^{(4)}, y^{(4)})$ corresponding respectively to horizontal, vertical and the two diagonal neighbors, as shown on Figure 4. We could also investigate other relationships, for example not directly related neighbors but k-neighbors (located at a distance of k pixels), for every k, or look at these correlations in another domain, for example in the Fourier domain, to expose frequency dependencies. The area is considered to be homogeneous if each of the K obtained p-values $p(x^{(k)}, y^{(k)})$ reaches a given level of significance α. By doing so, the overall level of detection α_{eq} after aggregation is no longer α but smaller

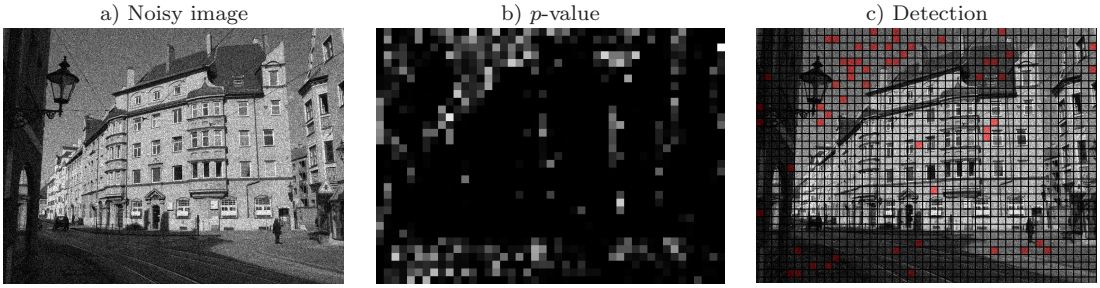


Figure 5. *Detection of homogeneous areas in an image corrupted with a hybrid noise as the sum of Gaussian, Poissonian and multiplicative gamma noise whose NLF parameters are* $(a, b, c) = (0.0312, 0.75, 400)$*, resulting in an initial PSNR of 17.93dB. a) Noisy image (range [0, 255]), b) p-value (range* $[black = 0, white = 1]$*) of the associated Kendall's* τ *coefficient computed within blocks of size* $N = 16 \times 16$*, and c) selected homogeneous blocks (red) by thresholding the p-value to reach a probability of detection of* $P_D = 1 - P_{FA} = 0.7$*.*

and given by

$$\alpha_{eq} = \mathbb{P}\left(\bigcap_{k=1}^{K} \left\{ p(X^{(k)}, Y^{(k)}) > \alpha \right\}\right). \tag{3.16}$$

In order to control the overall level of detection α_{eq}, we empirically estimated offline the relation between α_{eq} and α.

Figure 5 illustrates the selection of homogeneous areas. We synthesize a hybrid noise as the sum of Gaussian, Poissonian and multiplicative gamma noise, resulting in an initial PSNR of 17.93dB. We compute Kendall's τ coefficient within non-overlapping blocks of size $N = 16 \times 16$, by comparing horizontal, vertical, and diagonal neighbors. Figure 5-b displays the p-value associated to the rank correlation computed between horizontal neighbors. Then we perform the selection by thresholding the p-value, in order to obtain an overall level of detection given by the probability of detection $P_{\mathrm{D}} = 1 - P_{\mathrm{FA}} = 0.7$. The blocks that have been considered homogeneous are displayed in red on Figure 5-c.

4. Model estimation. Once the homogeneous blocks $B = G_\omega$ are detected, the noise statistics can be estimated on those blocks. More precisely, for an homogeneous block B of size N, Eq. (2.1) implies that there exists $\mu \in \mathbb{R}$ and $\sigma^2 > 0$ such that $\mathbb{E}[B_i] = \mathbb{E}[B_j] = \mu$ and $\mathrm{Var}[B_i] = \mathrm{Var}[B_j] = \sigma^2$ for all $i, j \in [1, \dots, N]$. Again as B is unknown, we can only relies on empirical statistics based on the observation b that unbiasedly estimate μ and σ^2 as follows

$$\widehat{\mu}(b) = \frac{1}{N}\sum_{i=1}^{N} b_i \quad \text{and} \quad \widehat{\sigma}^2(b) = \frac{1}{N-1}\sum_{i=1}^{N} (b_i - \widehat{\mu})^2, \tag{4.1}$$

meaning that $\mathbb{E}[\widehat{\mu}(B)] = \mu$ and $\mathbb{E}[\widehat{\sigma}^2(B)] = \sigma^2$. Accordingly, each homogeneous block gives a punctual estimation of the the noise level function defined in (2.1) as $\widehat{\sigma}^2(b) \approx \mathrm{NLF}(\widehat{\mu}(b))$. In this section, we explain how to robustly estimate the noise level function given P pairs of statistics $(\widehat{\mu}(b^p), \widehat{\sigma}^2(b^p)), p \in [1, \dots, P]$ from P detected homogeneous blocks $b^p = g_{\omega_p}$.

Before turning to the estimation of the noise level function, it is important to understand the behavior of the above variance estimator inside a single homogeneous region.

4.1. Punctual estimation from a single homogeneous area. The unbiased estimator of the variance σ^2 computed inside a block B of N pixels might suffer from statistical errors. This section aims at studying the performance of the variance estimator in the best

possible case, i.e., when all fluctuations in b are only due to the noise (no false alarms). More precisely, we are interested in quantifying the relative error $\frac{|\widehat{\sigma}^2(B)-\sigma^2|}{\sigma^2}$ under H_0. To do so, we first introduce the following proposition.

Proposition 4.1. *Let B be a homogeneous block satisfying H_0. We have*

$$\mathrm{Var}(\widehat{\sigma}^2(B) \mid H_0) \underset{N\to\infty}{\sim} \frac{\sigma^4}{N}\left(\beta_2 - \frac{N-3}{N-1}\right), \tag{4.2}$$

where β_2 is the kurtosis given, for any $i \in [1,\dots,N]$, by the following formula

$$\beta_2 \underset{N\to\infty}{\sim} \mathbb{E}\left[\left(\frac{B_i-\mu}{\sigma}\right)^4 \,\middle|\, H_0\right]. \tag{4.3}$$

The proof can be found in [10]. Not surprisingly, this proposition shows that as N increases, the variance vanishes, and as $\widehat{\sigma}^2$ is an unbiased estimator, $\widehat{\sigma}^2(B)$ tends towards σ^2. However, it shows also that the variance and hence the efficiency of the unbiased estimator of σ^2 depend on the variance σ^2 itself and the kurtosis β_2 of the unknown noise process that generates b.

We are now ready to introduce our third theorem.

Theorem 4.2. *Let B be a homogeneous block satisfying H_0. The relative error in estimating the variance σ^2 has an expected value and a variance given by*

$$\mathbb{E}\left[\frac{|\widehat{\sigma}^2(B)-\sigma^2|}{\sigma^2} \,\middle|\, H_0\right] \underset{N\to\infty}{\sim} \sqrt{\frac{1}{N}\left(\beta_2 - \frac{N-3}{N-1}\right) \times \frac{2}{\pi}}, \tag{4.4}$$

$$\mathrm{Var}\left(\frac{|\widehat{\sigma}^2(B)-\sigma^2|}{\sigma^2} \,\middle|\, H_0\right) \underset{N\to\infty}{\sim} \frac{1}{N}\left(\beta_2 - \frac{N-3}{N-1}\right) \times \left(1 - \frac{2}{\pi}\right). \tag{4.5}$$

The proof is detailed in appendix F. This formula gives the expected value and the variance of the estimation error when computing the variance on a single homogeneous block.

Interestingly, like for the variance of $\widehat{\sigma}^2(B)$, this theorem shows that the relative error vanishes with high probability as the block size N increases. But, unlike the variance of $\widehat{\sigma}^2(B)$, the relative error under H_0 solely depends on the kurtosis and the size of the block irrespectively of the variance σ^2 to be estimated. This shows that, for a given size of block N, the estimation error can then be predicted as a function of the kurtosis, as displayed on Figure 6-a where $N = 16 \times 16$. Hence, without any assumptions on the underlying variance, one can bound the expected relative error made by our approach under H_0 provided some mild assumptions on the kurtosis. For instance, assuming the kurtosis is lower than 40 (that corresponds to a very high peakedness of the noise random process and so such high kurtosis is very unlikely to appear in practical imaging problem), the expected relative error will be lower than 30%.

The following examples support our claim in the case of noise distributions classically encountered in imaging problems.

Example 4.3 (Additive white Gaussian noise). *Consider the example of an additive white Gaussian noise of parameter σ^2 as given in Example 2.1. In this case, the kurtosis β_2 is constant with respect to σ^2 and equal to 3. Provided the block is of size $N = 16 \times 16$, Theorem 4.2 shows that the expected relative error in estimating σ^2 is of about 7%.*

Figure 6-b displays the behavior of the variance estimation error as a function of the block size N in the case of Gaussian noise. The red line displays the mean relative error, that decreases as $1/\sqrt{N}$ as N increases, the blue curves show the standard deviation around

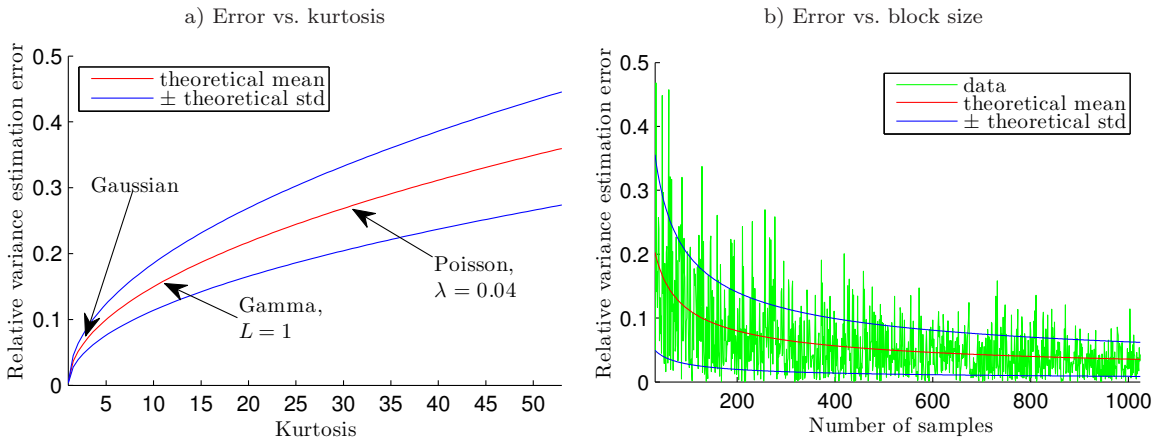


Figure 6. *a) Evolution of the expected value and standard deviation of the relative error of the variance estimator as a function of the kurtosis β_2 for a size of block $N = 16 \times 16$, and b) Evolution of the expected value and standard deviation of the relative variance estimation error as a function of the block size for an additive white Gaussian noise.*

the expected value that tends towards zero at the same rate, and the green line shows the variance estimation error at each bloc size for different noise realizations.

Example 4.4 (Multiplicative gamma case). *Consider the example of the multiplicative gamma noise of parameter L as given in Example 2.3. In this case, the kurtosis is given by the formula $\beta_2 = \frac{6}{L} + 3$. Here, contrary to the Gaussian case, the estimation error depends on the parameter of the distribution L. In the worst case when $L = 1$, Theorem 4.2 shows that for a block of size $N = 16 \times 16$ the expected relative error in estimating the variance is of about 14%.*

Example 4.5 (Poisson case). *Consider the example of the Poisson noise as given in Example 2.3. In this case, the kurtosis varies spatially with the image intensity, and is given by $\beta_2 = \frac{1}{g_i^0} + 3$, where g_i^0 is the underlying intensity. for example, in a pessimistic case, where $\min_{i \in \omega} g_i^0 = 0.04$ for a block of size $N = 16 \times 16$, which corresponds in an imaging problem of a pixel with a very low expected number of photons, Theorem 4.2 shows that the expected relative error in estimating the variance is of about 28%.*

Besides, the estimation error computed here gives a bound on the error that affects the estimation of the variance on a single block, i.e., for one point of the NLF. The next step consists in fitting a NLF to the set of estimated points, so the overall error based on all the selected blocks should be reduced by the number of selected blocks, which depends on the image size and the desired level of false alarms.

However, when some false detections arise, i.e., if inhomogeneous blocks are selected, the variance will be overestimated, resulting in a higher estimation error.

The bounds computed above hence give a safety check for the error one might expect: if the overall error ends up being higher than the predicted bound, this means either that too many outliers were taken into account or that the NLF estimator did not fit properly.

4.2. Noise level function estimation from all homogeneous areas. Provided that uniform regions have been detected, the noise model can be estimated from the mean/variance couples extracted from these areas. We now explain how to robustly estimate a noise level function (NLF) as defined in Eq. 2.1 given P pairs of statistics $(\hat{\mu}_p, \hat{\sigma}_p^2) = (\widehat{\mu}(b^p), \widehat{\sigma}^2(b^p)), p \in [1, \dots, P]$ from P detected homogeneous blocks $b^p = g_{\omega_p}$. Within the p-th homogeneous block, the intensity variations should be only ascribed to noise fluctuations, hence $\widehat{\sigma}_p^2 \approx \mathrm{NLF}(\widehat{\mu}_p)$. As each block provides only a punctual infor-

mation about the NLF, we are confronted to an under constraint approximation problem from P points of $\mathbb{R}\times\mathbb{R}^+$ to a function of $\mathbb{R}\to\mathbb{R}^+$. It is then necessary to inject some extra regularity assumptions. We suggest that the NLF is a positively increasing second order polynomial of the image intensity, parametrized by $\theta=(a,b,c)^t\in(\mathbb{R}^+)^3$ and given by

$$\sigma_p^2=\mathrm{NLF}_\theta(\mu_p)=a\mu_p^2+b\mu_p+c. \tag{4.6}$$

According to examples 2.1, 2.2 and 2.3, this model encompasses Gaussian noise, with $(a,b)=(0,0)$, in which case the NLF is constant, Poisson noise, with $(a,c)=(0,0)$, in which case the NLF is linear, and multiplicative noise, with $(b,c)=(0,0)$, in which case the NLF is parabolic, as well as hybrid models, e.g., Poissonian-Gaussian. Note that the positivity constraints can be dropped to embrace more general models to allow non increasing noise level function.

The goal is to find the polynomial coefficients (a,b,c) such that the estimated variances $\hat\sigma_p^2$ can be well approached as $\mathrm{NLF}_\theta(\hat\mu_p)=a\hat\mu_p^2+b\hat\mu_p+c$. Denoting by $\hat\mu\in\mathbb{R}^P$, $\hat\sigma^2\in(\mathbb{R}^+)^P$ and $\mathrm{NLF}_\theta(\hat\mu)$, the vectors obtained by stacking each of these estimations, our approximation problem can be expressed in a vectorial form as

$$\hat\sigma^2\approx\mathrm{NLF}_\theta(\hat\mu)=A\theta\quad\text{where}\quad A=\begin{pmatrix}\hat\mu_1^2&\hat\mu_1&1\\ \hat\mu_2^2&\hat\mu_2&1\\ \vdots&\vdots&\vdots\\ \hat\mu_P^2&\hat\mu_P&1\end{pmatrix} \tag{4.7}$$

Of course, if the noise is known to be Gaussian, Poissonian or multiplicative respectively, the operator A should be adapted, keeping only a single column of ones, of the means or of the means to the square respectively. In case of the Poisson-Gaussian noise, only the two columns of ones and of the means should be kept.

We focus in the following on two different estimators in order to robustly estimate θ.

4.2.1. Least squares estimation. The least squares (LS) method consists in minimizing a ℓ^2-norm of the residue $\mathrm{NLF}_\theta(\hat\mu)-\hat\sigma^2$. The estimation is formalized as the solution of the follow constraint optimization problem

$$\hat\theta^{\mathrm{LS}}=\operatorname*{argmin}_{\theta\in(\mathbb{R}^+)^3}\|\mathrm{NLF}_\theta(\hat\mu)-\hat\sigma^2\|_2=\operatorname*{argmin}_{(a,b,c)\in(\mathbb{R}^+)^3}\|a\hat\mu^2+b\hat\mu+c-\hat\sigma^2\|_2 \tag{4.8}$$

whose solutions can be obtained using quadratic programming tools. Remark that without the positivity constraint on (a,b,c) the least square solution is explicit and the following holds

$$\widehat\theta^{\mathrm{LS}}=(A^tA)^{-1}A^t\hat\sigma^2\quad\text{and}\quad\mathrm{NLF}_\theta^{\mathrm{LS}}(\hat\mu)=\Pi_A\hat\sigma^2 \tag{4.9}$$

where $\Pi_A=A(A^tA)^{-1}A^t$ is the so-called hat function that is the projector on the space of second order polynomial.

Assuming $\hat\sigma^2$ would have approximately a normal distribution with covariance proportional to the identity, the above least square estimator would correspond to the maximum likelihood estimator that is guaranteed to converge in probability towards the underlying second order polynomial NLF function. However, even though $\hat\sigma^2$ has a diagonal covariance matrix (as the blocks are independent), it is not proportional to the identity. Worst, the distribution of $\hat\sigma^2$ deviates from normality as the data might suffer from outliers. Indeed, large errors can arise when some non-homogeneous samples happen to be selected. This leads us to studying other estimation methods more robust to outliers.

Algorithm 1 A preconditioned primal-dual algorithm of [7] for least absolute deviation

Set $\gamma \in \mathbb{R}^{P,1}$ and $\tau \in \mathbb{R}^{1,P}$ such that $\gamma_i = 1/\sum_{j=1}^{3} |A_{ij}|$, and $\tau_j = 1/\sum_{i=1}^{P} |A_{ij}|$.
$\Pi_{|z|\leq 1}$ the projector of $\mathbb{R}^P$ such that $|z_i| \leq 1 \; \forall 1 \leq i \leq p$,
$\Pi_{\theta \geq 0}$ the projector of $\mathbb{R}^3$ such that $\theta_j \geq 0 \; \forall 1 \leq j \leq 3$,
Initialize $z_0 = 0 \in \mathbb{R}^P$, $\bar{\theta}_0 = 0 \in \mathbb{R}^3$ and $\theta_0 = 0 \in \mathbb{R}^3$,
Iterate for $k \geq 0$

$$\left\{ \begin{array}{rcl} z_{k+1} & = & \Pi_{|z|\leq 1}\left(z_k + \gamma(A\bar{\theta}_k - \hat{\sigma})\right), \\ \theta_{k+1} & = & \Pi_{\theta \geq 0}\left(\theta_k - \tau A^* z_{k+1}\right), \\ \bar{\theta}_{k+1} & = & 2\theta_{k+1} - \theta_k. \end{array} \right.$$

4.2.2. Least absolute deviation. If miss-detections happen as inhomogeneous samples are selected, the presence of structures in the signal might result in a highly over estimated variance resulting from the summation of the noise variations with the signal variations. In order to reduce the impact of such outliers, we suggest using the least absolute deviation (LAD) method. Unlike the LS estimation, the LAD estimation is shown to be more robust to outliers [34].

The LAD estimation is based on the minimization of the ℓ^1-norm of the residue (instead of the ℓ^2-norm) as the solution of the following constraint problem

$$\hat{\theta}^{\mathrm{LAD}} = \operatorname*{argmin}_{\theta \in (\mathbb{R}^+)^3} \| \mathrm{NLF}_\theta(\hat{\mu}) - \hat{\sigma}^2 \|_1 = \operatorname*{argmin}_{(a,b,c)\in(\mathbb{R}^+)^3} \| a\hat{\mu}^2 + b\hat{\mu} + c - \hat{\sigma}^2 \|_1. \tag{4.10}$$

There is no close form solution for such a problem even in the un-constrained case. Fortunately, we can derive a solution based on a fast iterative algorithm, using the preconditioned primal-dual algorithm of Chambolle and Pock [7], as displayed in Algorithm 1 and where Π is the projection operator on a convex set.

While the LS estimator corresponds to the maximum likelihood estimator in the case of estimated variances following a specific normal distribution, the LAD estimator is the maximum likelihood estimator assuming a specific Laplacian distribution. Unlike the Gaussian distribution, the Laplacian distribution has heavy tails meaning that it is suitable to describe data subject to extreme values or outliers. According to the maximum likelihood interpretation of the LAD estimator, errors are assumed to follow the same Laplacian distribution, which is not verified in practice, see Theorem 4.2. And even if it were Laplacian, the scale of the Laplacian should be rescaled with respect to the unknown underlying variance as suggested by Eq. 4.4. However, due to the performance of the homogeneous detector the scatterplot is reliable (the amount of false detections is reasonable and the false detections are not too aberrant). The LAD estimator hence offers a fast and reliable estimation that is robust to outliers.

A more realistic model would be to use distributions with heavier tails, such as the Cauchy distribution, as proposed in [2]. This leads to optimization problems that are typically non convex, and hence harder to minimize. The Laplacian distribution is in this sense a good trade-off between robustness and computational efficiency.

Examples of other robust estimators include Theil-Sen's estimator [42] (also referred to as Kendall robust line-fit method) used to fit affine functions based on median slopes, its generalization as the quantile regression [26] for finding curves corresponding to the conditional median or other quantiles of the data and Hubert's M-estimators [21] designed to be robust under partial knowledge of the data distribution.

Figure 7. *Images extracted from the database of 150 high quality grayscale images of resolution 540×720, taken from a 7 Mpix Canon Powershot S70 or a 8 Mpix Canon IXY Digital 910IS, where the noise can be considered to be negligible.*

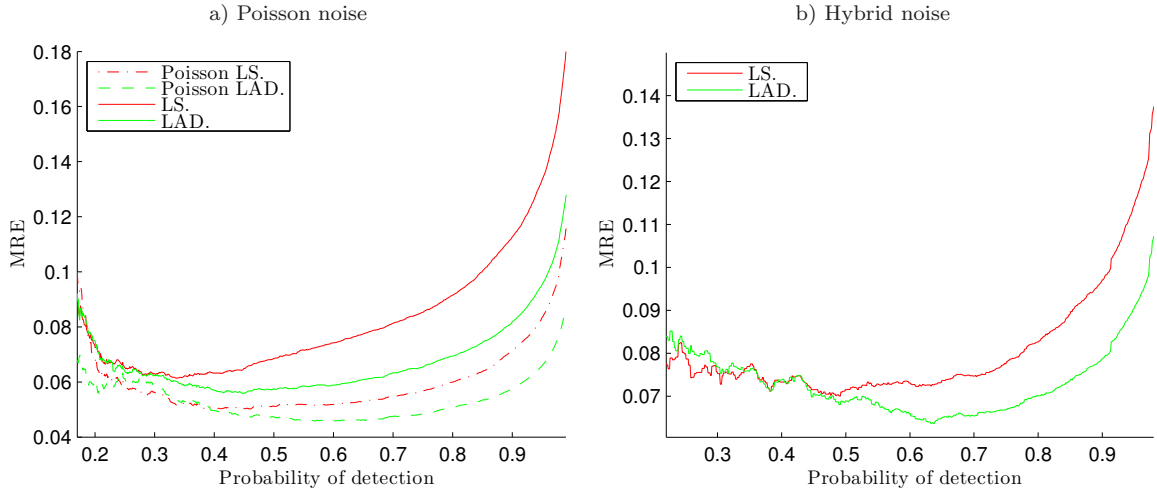


Figure 8. *Mean relative error as a function of the desired probability of detection $P_D = 1 - P_{FA}$. On the left, the noise is Poissonian, with different levels. It can be estimated with the prior knowledge of the noise model, using the least squares or the least absolute deviation adapted to the noise model (Poisson LS or Poisson LAD), or it can be described by the general second order model with the hybrid estimators, LS or LAD. On the right, the noise is hybrid with different NLF parameters. Hybrid noise are estimated with the general second order model using the least square and the least absolute deviation estimators. Whereas the noise model is known or not, the LAD estimator is shown to be more accurate that the LS estimator.*

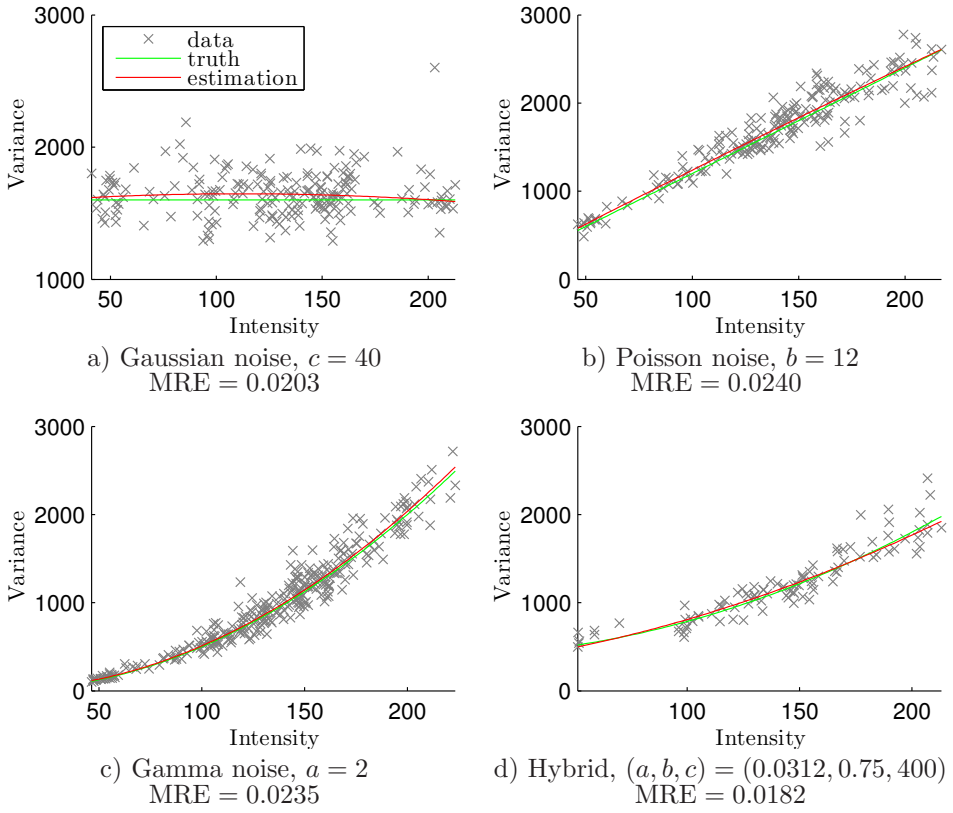
4.3. Comparison of estimators. To compare the least absolute deviations estimator with the least squares estimator, and for all the following experiments, we have generated a set of noisy images with different noise levels, using a database of 150 clean grayscale high quality images of resolution 540×720, taken from a 7 Mpix Canon Powershot S70 or a 8 Mpix Canon IXY Digital 910IS[1]. The noise on these images can be considered to negligible, as shown on the sample of this database displayed on Figure 7. Based on the knowledge of the real noise parameters $\theta = (a, b, c)$, we can compute the mean relative error

$$\mathrm{MRE}(\widehat{\theta}) = \frac{1}{|I|} \sum_{f_i \in I} \frac{\left|\mathrm{NLF}_{\theta}(f_i) - \mathrm{NLF}_{\widehat{\theta}}(f_i)\right|}{\mathrm{NLF}_{\theta}(f_i)}, \tag{4.11}$$

where I is a discretization of the interval of image intensities.

Figure 8 displays the evolution of the mean relative error as a function of the desired level of detection $P_{\mathrm{D}} = 1 - P_{\mathrm{FA}}$. On Figure 8-a, the images have been corrupted by Poissonian noise with a varying noise level. We can use the prior knowledge of a Poissonian model, and estimate solely the parameter b with the least square estimator (*Poisson LS*) and the least absolute deviation method (*Poisson LAD*) adapted to the Poisson model,

[1] http://www.gipsa-lab.grenoble-inp.fr/~laurent.condat/imagebase.html



a) Gaussian noise, $c = 40$
MRE $= 0.0203$

b) Poisson noise, $b = 12$
MRE $= 0.0240$

c) Gamma noise, $a = 2$
MRE $= 0.0235$

d) Hybrid, $(a, b, c) = (0.0312, 0.75, 400)$
MRE $= 0.0182$

Figure 9. *Parameter estimation for different types of noise with the LAD estimator. The noise model is supposed to be unknown so the NLF is estimated as a general second order polynomial function.*

while the other two parameters a and c are set to zero. We then compare such estimators to the least square (LS) and least absolute deviation (LAD) estimators that estimate the three parameters (a, b, c) of a general second order model without the knowledge of a Poissonian model. On Figure 8-b, the images have been corrupted by hybrid noise with varying NLF parameters (a, b, c). In this case, a general second order model is directly estimated.

4.4. Parameter estimation results. Figure 9 displays the estimation of the NLF with the least absolute deviations estimator for different noise models (Gaussian, Poisson, multiplicative gamma and hybrid). The mean/variance couples extracted from the homogeneous patches are shown in gray crosses, the real NLF is displayed in green and the estimated one using the least absolute deviation estimator appears in red. There is no prior on the noise model, so the estimation is performed by seeking a second order polynomial NLF. The results fit closely, and the shape of each estimated curve shows that even with no prior knowledge of the noise level function, the model is accurately guessed. The knowledge of the noise model can make the parameters estimation even more accurate.

5. Practical settings.

5.1. Automatic adaptation of the block size N. The optimal block size has to satisfy a compromise between the reliability of the detection and the performance of the estimation. Indeed, both Kendall's test (see Section 3.2.3) and the computation of the statistics on each block (see Section 4.1) need a sufficient number of samples, dictated by the block size. But if the block size is too big then the number of homogeneous zones might be too small or even null, hence limiting the performance of the NLF estimation.

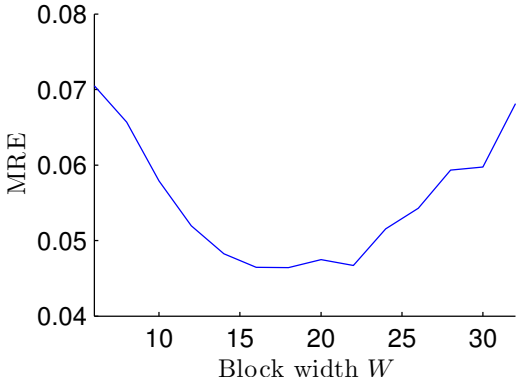


Figure 10. *Evolution of the mean relative error as a function of the block width W.*

To make sure that the estimation remains efficient or even possible, we impose a minimum number of detected areas. We divide the intensities of the input image into three bins of equal size, and we force the detection step to find at least three regions in each bin. If this condition is not satisfied, then the block width is reduced by two pixels in each direction, until the number of detected regions is sufficient. This guarantees that most images can be processed, and that the initial block size can be set to be large enough to limit the error in Kendall's test and in the computation of the statistics.

5.2. Selection of the maximum block size. The maximal block size can then be set as the size that balances the compromise between the reliability of the detection and the performance of the estimation. We have studied on a dataset of images corrupted by Gaussian noise the evolution of the mean relative error as a function of the block width W, where the block size $N = W \times W$, displayed on Figure 10. Since we cannot in this case automatically reduce the block size in order to guarantee a minimum number of regions, we impose as before a minimum of three detected regions in each bin by selecting the ones that have the best Kendall score, even if the associated p-value is below the threshold. Again, this was only done to determine offline on our dataset the maximum size window. This guarantees a sufficient number of data for the estimation. However, if the block size is too big the data might be irrelevant, meaning that the statistics would arise from non-homogeneous areas. The optimum appears to be reached around $W = 16$ pixels.

5.3. Selection of the probability of detection. To set the optimum probability of detection, rules by the threshold α, we have used the results obtained from Figure 8. Indeed, the convex shape of the error curves shows two opposite influences. At low levels of detection, the number of selected areas is small. False detections do not arise very likely, but the NLF estimation is not finely estimated as the number of data is insufficient. When the level of detection increases, the number of samples for the estimation increases, which reduces the estimation error, but the number of false detections increases as well, leading to the presence of outliers. The outliers are responsible for the increasing error at higher levels of detection, and they also explain the superiority of the least absolute deviation estimator over the least square estimator, at high levels of detection in particular. This leads to a trade-off between detection and outliers, indicating that for a level of detection $\approx 60\%$, a relative error lower than 10% can be reached without prior knowledge of the noise model. Accordingly, in all the following experiments, we chose the parameter $P_{\mathrm{D}} = 0.6$.

6. Experiments. In this section, we discuss and compare the efficiency of our proposed approach in terms of NLF estimation errors and image denoising applications. For the sake of replicability, a Matlab implementation for the automatic

Table 1

Estimation error under low Gaussian noise, with increasing variance. The estimators are MAD [15], the Gaussian-Cauchy mixture model [2], the PCA method [12], the percentile method [11], Noise Clinic [13], and our algorithm. In our method, we automatically adapt the block size to provide enough homogeneous zones, so the average block size for each level is displayed.

	$\sigma = 1$	$\sigma = 2$	$\sigma = 5$	$\sigma = 10$
MAD [15]	3.263	1.546	0.621	0.314
Gaussian-Cauchy [2]	55.767	11.281	1.619	0.4669
PCA [12]	**0.376**	**0.163**	0.136	0.235
Percentile [11]	0.421	0.181	**0.077**	**0.0378**
Noise Clinic [13]	0.856	0.3324	0.113	0.112
Proposed	2.028	0.640	0.134	0.047
(Average block size W)	(12)	(14)	(15)	(16)

noise estimation and its application to image denoising is available for download at https://github.com/csutour/RNLF.

We validate the proposed approach with respect to the state of the art algorithms that perform noise estimation, in terms of efficiency, accuracy, and validation for other processing tasks. Based on the database of 150 natural images, we generate a set of noisy images, for different noise statistics (Gaussian, Poisson, Poisson-Gaussian, hybrid...) and levels. To our knowledge, few state of the art methods can deal with such a general noise model as the one we propose here, so we estimate the noise parameters with the different estimators, according to the noise model involved and the properties of each estimator: the mean absolute deviation (MAD) estimator [15], which is only suitable for the estimation of Gaussian noise, the Gaussian-Cauchy mixture model [2] which is the most general model, the PCA method [12], the percentile method [11], and Noise Clinic [13, 27], that estimate frequency-dependent noise but that we use here for the estimation of affine noise, the estimation based on the variance stabilization transform (VST) [36] that applies for Poisson-Gaussian noise, and our algorithm that can estimate either a given model or a general second order one.

6.1. Estimation of low Gaussian noise. A fine estimation of very low Gaussian noise is hard to achieve. In our case, this is due to the difficulty of finding homogeneous areas. Indeed, according to the characterization of homogeneous areas that we define in the introduction, a region is homogeneous if its signal to noise ratio is weak. However, with such low noise levels, the signal fluctuations cannot be assumed to be negligible compared to the noise variations, so the detection is harder to achieve.

Table 1 gives the mean relative error when estimating very low Gaussian noise ($\sigma = 1, 2, 5, 10$) with the different estimators adapted for Gaussian noise: MAD [15], the Gaussian-Cauchy mixture model [2], the PCA method [12], the percentile method [11], Noise Clinic [13], and our algorithm. It does indeed increase when the noise level gets too small. In the proposed method, we automatically adapt the block size to provide enough homogeneous zones, which is more difficult at smaller noise levels. Table 1 also displays in the bottom line the average optimal block size, which has to be reduced under lower noise levels.

6.2. Estimation of strong Poisson noise. Estimating pure Poisson noise can also be challenging, due to the behavior of the noise, especially in the darkest areas when it can appear as shot noise. Figure 11 illustrates the estimation of strong Poisson noise, with noise parameter $b = 1000$ and initial PSNR$\approx$2dB. The original and noisy images are displayed

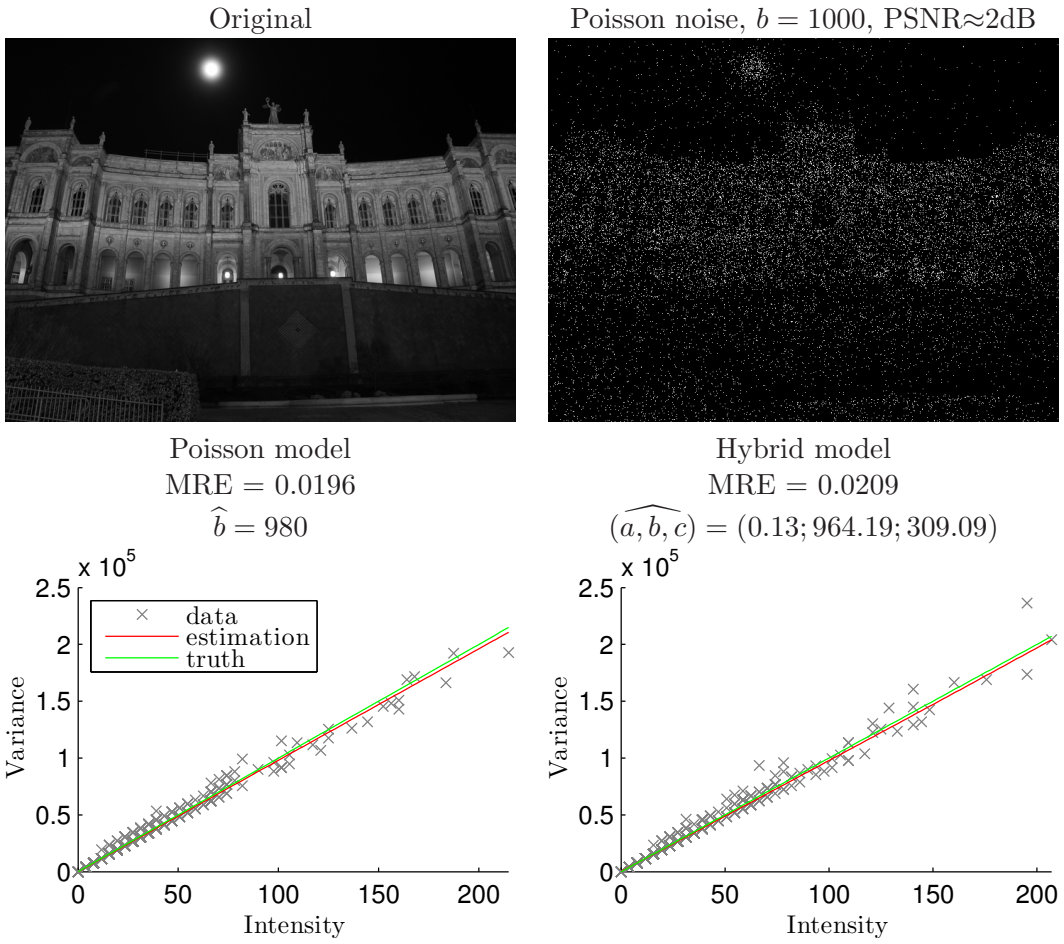


Figure 11. *Estimation of strong Poisson noise, with parameter $b = 1000$ and initial PSNR≈2dB. The original and noisy images are displayed on the top line. On the bottom line, the noise parameter are estimated first assuming a Poisson model (on the left), then with the general second order model (on the right). In both cases, the relative error is below 5%, showing the reliability of the estimation.*

Table 2

Estimation error under strong Poisson noise, with different powers. The estimators are MAD [15], the Gaussian-Cauchy mixture model [2], the PCA method [12], the percentile method [11], Noise Clinic [13], the VST based method [36], and our algorithm.

	$b = 60$	$b = 100$	$b = 200$	$b = 1000$
Gaussian-Cauchy [2]	0.088	0.116	0.183	0.571
PCA [12]	0.681	0.788	0.933	1.000
Percentile [11]	0.279	0.371	0.477	0.834
Noise Clinic [13]	0.430	0.457	0.411	0.998
VST [36]	0.281	\	\	\
Proposed (Poisson)	**0.053**	**0.053**	**0.072**	**0.238**
Proposed (hybrid)	0.121	0.131	0.106	0.433

on the top line. On the bottom line, the noise parameter are estimated first assuming a Poisson model (on the left), then with the general second order model (on the right). In both cases, the relative error is below 5%, showing the reliability of the estimation.

Table 2 gives the mean relative error when estimating strong Poisson noise ($b = 60, 100, 200, 1000$) with the different estimators adapted for Poisson noise: the Gaussian-

Cauchy mixture model [2], the PCA method [12], the percentile method [11], Noise Clinic [13], the VST based method [36], which does not work for very high levels, and our algorithm, which offers the most reliable estimation results.

6.3. Comparison on general noise models. Tables 3 to 6 illustrate the estimation performance of the suitable estimators for commonly encountered noise models: Gaussian 3, Poisson-Gaussian 4, multiplicative gamma noise 5 and hybrid noise 6. The different estimators that are used are the mean absolute deviation (MAD) estimator [15], which is only suitable for the estimation of Gaussian noise, the Gaussian-Cauchy mixture model [2] which is the most general model, the PCA method [12], the percentile method [11], and Noise Clinic [13, 27], that can estimate frequency-dependent noise but are used here for the estimation of affine noise, and the estimation based on the variance stabilizing transform (VST) [36] that applies for Poisson-Gaussian noise. We compare these methods to our propose algorithm that can estimate either a given model or a general second order one. In each case, we can either estimate the suitable model with the involved parameters, e.g. constant for Gaussian noise, affine for Poisson-Gaussian, etc, or estimate a second order polynomial NLF, which corresponds to the 'hybrid' case. Results show that the proposed method offers an accurate noise level estimation. It is quite competitive for affine noise, and it offers state of the art estimation results for second order noise models, in particular multiplicative gamma noise.

6.4. Application to image denoising. Automatic noise estimation can be useful for many applications that require the knowledge of the noise level, in particular image denoising. In order to illustrate the use of the proposed method, the estimated NLF is re-injected for blind image denoising.

The non-local means algorithm (NL-means) [6] has been adapted in order to use only the knowledge of the noise parameters (a, b, c) provided by the estimated NLF. The NL-means algorithm is based on the natural redundancy of the image structures, not just locally but in the whole image. While local filters average pixels that are spatially close to each other, the NL-means algorithm compares patches, i.e., small windows extracted around each pixel, in order to average pixels whose surroundings are similar. For each pixel $i \in \Omega$, the solution of the NL-means is

$$u_i^{\mathrm{NL}} = \frac{\sum_{j\in\Omega} w_{i,j} g_j}{\sum_{j\in\Omega} w_{i,j}}, \tag{6.1}$$

where the weights $w_{i,j} \in [0, 1]$ are computed in order to select the pixels j whose surrounding $\rho_j \subset \Omega$ is similar to the surrounding $\rho_i \subset \Omega$ of the central pixel i, for instance, as proposed in [40] by

$$w_{i,j} = \exp\left(-\frac{|d(g^{\rho_i}, g^{\rho_j}) - m_d^{\rho}|}{s_d^{\rho}}\right) \tag{6.2}$$

where $m_d^{\rho} = \mathbb{E}[d(G^{\rho_i}, G^{\rho_j})]$ and $s_d^{\rho} = \sqrt{\mathrm{Var}[d(G^{\rho_i}, G^{\rho_j})]}$ are respectively the expectation and the standard deviation of the dissimilarity d computed between two noisy patches of size $|\rho|$ following the same distribution.

In order to take into account the signal-dependence of the noise, we approximate the noise to additive Gaussian noise with polynomial variance, and we adapt the dissimilarity measure d as follows

$$d(g^{\rho_i}, g^{\rho_j}) = \frac{1}{|\rho|} \sum_{k=1}^{|\rho|} \frac{\left(g_k^{\rho_i} - g_k^{\rho_j}\right)^2}{\mathrm{NLF}_\theta(g_k^{\rho_i}) + \mathrm{NLF}_\theta(g_k^{\rho_j})}. \tag{6.3}$$

Table 3

Mean relative error (MRE) when estimating Gaussian noise with MAD [15], the Gaussian-Cauchy mixture model [2], the PCA method [12], the percentile method [11], Noise Clinic [13, 27], and our algorithm when estimating either a Gaussian model or a second order one, and PSNR achieved after denoising with the NLF-means algorithm, using the estimated NLF parameters.

Estimator	$\lvert\widehat{a}\rvert$	$\widehat{b}$	$\frac{\lvert c-\widehat{c}\rvert}{c}$	MRE	PSNR
MAD [15]	\	\	0.08	0.081	26.33
Gaussian-Cauchy [2]	\	0.55	0.05	0.040	26.41
PCA [12]	\	\	0.51	0.513	24.12
Percentile [11]	\	\	0.11	0.105	25.69
Noise Clinic [13]	\	\	0.39	0.389	23.69
[27]					(24.29)
Proposed, Gaussian	\	\	**0.03**	**0.030**	**26.72**
Proposed, hybrid	0.00	1.56	0.12	0.056	26.64

Table 4

Mean relative error (MRE) when estimating Poisson-Gaussian noise with the Gaussian-Cauchy mixture model [2], the PCA method [12], the percentile method [11], Noise Clinic [13, 27], the VST based method [36] and our algorithm when estimating either an affine model or a second order one, and PSNR achieved after denoising with the NLF-means algorithm, using the estimated NLF parameters.

Estimator	$\lvert\widehat{a}\rvert$	$\frac{\lvert b-\widehat{b}\rvert}{b}$	$\frac{\lvert c-\widehat{c}\rvert}{c}$	MRE	PSNR
Gaussian-Cauchy [2]	\	**0.06**	**0.11**	0.040	26.10
PCA [12]	\	0.54	2.71	0.88	23.67
Percentile [11]	\	0.28	0.44	0.12	26.36
Noise Clinic [13]	\	0.45	0.34	0.390	**26.44**
[28]					(25.18)
VST [36]	\	0.07	0.15	**0.034**	26.23
Proposed, affine	\	0.11	0.25	0.055	26.07
Proposed, hybrid	0.01	0.13	0.26	0.054	26.02

Table 5

Mean relative error (MRE) when estimating multiplicative gamma noise with the Gaussian-Cauchy mixture model [2] or and our algorithm, and PSNR achieved after denoising with the NLF-means algorithm, using the estimated NLF parameters.

Estimator	$\frac{\lvert a-\widehat{a}\rvert}{a}$	$\widehat{b}$	$\lvert\widehat{c}\rvert$	MRE	PSNR
Gaussian-Cauchy [2]	0.80	2.01	120.62	10.521	17.70
Proposed, gamma	**0.05**	\	\	**0.046**	**25.69**
Proposed, hybrid	0.06	1.14	8.02	0.086	25.14

Table 6

Mean relative error (MRE) when estimating hybrid second order noise with the Gaussian-Cauchy mixture model [2] or and our algorithm, and PSNR achieved after denoising with the NLF-means algorithm, using the estimated NLF parameters.

Estimator	$\frac{\lvert a-\widehat{a}\rvert}{a}$	$\frac{\lvert b-\widehat{b}\rvert}{b}$	$\frac{\lvert c-\widehat{c}\rvert}{c}$	MRE	PSNR
Gaussian-Cauchy [2]	0.27	2.53	0.60	**0.070**	**27.81**
Proposed	**0.20**	**1.75**	**0.36**	0.078	27.74

Table 7

Mean PSNRs achieved after denoising a set of noisy images corrupted with hybrid noise, with the standard NL-means assuming a Gaussian model, then with the NLF-means with the estimated NLF and the true NLF.

Initial PSNR	NL-means (Gaussian)	NLF-means (estimated NLF)	NLF-means (true NLF)
19.52	26.26	28.85	28.96
18.87	25.88	28.48	28.58
18.04	26.24	27.93	28.09
17.51	25.86	27.73	27.87

Figure 12. *Denoising of a hybrid noise with true parameters* $(a, b, c) = (0.0312, 0.625, 100)$*, initial PSNR = 20.34dB. The noisy image is displayed on Fig. 5-a. a) Standard NL-means assuming and estimating Gaussian noise, b) NLF-means with the estimated NLF and c) NLF-means with the true NLF.*

for which $m_d^\rho = 1$ and $s_d^\rho = \sqrt{\frac{2}{|\rho|}}$. These two quantities reflect the noise dependence to the signal, and they reduce the sensitivity of the dissimilarity to the noise level and patch size.

The application of the automatic noise estimation to image denoising shows both the performance of the estimation and the necessity of the knowledge of the noise statistics to perform image processing tasks. Table 7 displays the PSNRs achieved after blind denoising using the NLF-means algorithm proposed above. The images have been corrupted with hybrid noise with different levels. In the first column, the denoising results are obtained using the standard NL-means algorithm (that assumes Gaussian noise). Only the noise variance (i.e. the σ parameter) has been estimated, using the proposed estimation method for a specific Gaussian model. Both the low PSNRs and the artifacts on Figure 9.a) show that the denoising is quite inappropriate; hence the necessity to properly estimate the noise parameters and/or model. The other two columns use the NLF-means algorithm. In the second column, the NLF parameters have been estimated with our proposed method while on the third column we have used to true parameters used to simulate the noise. The superiority of these methods over the standard Gaussian NL-means shows the interest of assuming a relevant noise model, while the feeble difference between the last two columns

Table 8

Computational time (in seconds) for the estimation of the noise parameters on an image of size 540×720 corrupted with Gaussian noise with MAD [15], the Gaussian-Cauchy mixture model [2], the PCA method [12], the percentile method [11], Noise Clinic [13, 27], the VST based method [36], and our algorithm.

Estimator	Time (s)
MAD [15]	0.58
Gaussian-Cauchy [2]	32
PCA [12]	1
Percentile [11]	0.5
Noise Clinic [13]	20
VST [36]	530
Proposed	1.4

illustrates the reliability of our estimation procedure for denoising purposes.

Figure 12 displays the result of such denoising on a natural image, corrupted with a hybrid noise of parameters $(a, b, c) = (0.0312, 0.625, 100)$, with an initial PSNR of about 20dB. On Figure 12-a, the standard NL-means have been used, assuming a Gaussian distribution. The single noise parameter σ^2 has been estimated and injected in the NL-means algorithm. Since the noise model is inappropriate, the denoised image suffers from artifacts on uniform areas and loss of details, on the hat for example. On Figure 12-b, the NLF has been estimated using the proposed method then the estimated parameters $\widehat{(a, b, c)}$ are used for the NLF-means algorithm, while on Figure 12-c the true NLF parameters have been used. Both denoising results offer a satisfying result on which details are preserved, showing that the NLF parameters offer a good model and that the estimation is accurate enough to perform efficient denoising.

Besides, the estimations performed by each estimator in Tables 3 to 6 are re-injected in the NLF-means algorithm to evaluate the noise estimation through the denoising performance, using the PSNR computation. Results show once again that the proposed method offers efficient denoising thanks to the accurate noise level estimation.

6.5. Computational time. We use a fast implementation [25] that computes Kendall's coefficient in $O(N \log N)$ for a block of size N, resulting in a global complexity in $O(N_0 \log N)$, for an image of size N_0 divided into non-overlapping blocks of size N. For an image of size $N_0 = 512 \times 512$ and blocks of size $N = 16 \times 16 = 256$ pixels, the fast implementation allows to compute Kendall's coefficient for one direction in 0.12 seconds, and the detection on the whole image as well as the computation of the noise statistics on the selected homogeneous areas are performed in about 0.6 seconds.

Table 8 displays the computational time for the estimation of the noise parameters on an image of size 540×720 corrupted with Gaussian noise with MAD [15], the Gaussian-Cauchy mixture model [2], the PCA method [12], the percentile method [11], Noise Clinic [13, 27], the VST based method [36], and our algorithm, which is quite competitive.

7. Discussion and conclusion. We have developed a fully automatic noise estimation method that relies on the non-parametric detection of homogeneous blocks. These blocks are selected by measuring Kendall's τ coefficient between neighboring pixels, a non-parametric test that requires few hypothesis on the distribution of the samples. In other words, it means that our detector of homogeneous blocks has the exact same performance without extra tuning for any kind of uncorrelated signal-dependent noise including weak/strong Poisson noise, weak/strong Gaussian noise, multiplicative, Rician, hybrid,

impulsive, or salt and pepper noise. More precisely, we have shown that the detection errors are controlled with respect to the signal to noise ratio and the number of samples. Next, the noise level function is assumed to follow a second order polynomial model, and is estimated thanks to an estimation method of least absolute deviation. Results on synthetic images show that the proposed method gives an accurate noise level estimation with few assumptions. This NLF estimation can then be used for further image processing applications, for example image denoising, as demonstrated with the use of the NLF estimation in the NL-means algorithm to perform adapted denoising.

Future work might lead to the study of a more general noise model, beyond second order polynomial. First, it would be interesting to encompass spatially varying noise level functions. Indeed, in many image devices the noise level varies with the distance to the sensor for example. Another issue lies in spatially correlated noise. Spatial correlation is not handled by our Kendall-based detection, yet it occurs in many natural image corruptions. Dealing with spatial correlation would extend the noise estimation to artifacts detection, for example compression artifacts. Beyond noise estimation, it might in fact be interesting to extend the proposed approach to other corruptions such as blur detection.

Appendix A. Proof of Theorem 3.8.

Proof. Theorem 3.8 gives the relationship between the level of detection α and the probability of false alarm. In fact, the detection threshold α is actually chosen in order to achieve the desired probability of false alarm associated to condition (C1). Indeed, using the definition of the probability of false alarm and Corollary 3.7, we have:

$$\begin{aligned}\lim_{n\to\infty} \mathbb{P}(|z(X,Y)| > \alpha \mid H_0) &= \lim_{n\to\infty} \mathbb{P}(z(X,Y) > \alpha \mid H_0) + \mathbb{P}(-z(X,Y) < -\alpha \mid H_0) \\ &= (1-\phi(\alpha)) + \phi(-\alpha) = 2\,(1-\phi(\alpha))\,. \end{aligned} \tag{A.1}$$

Now using the choice of α given by $\alpha = \phi^{-1}(1 - P_{\mathrm{FA}}/2)$, we get:

$$2\,(1-\phi(\alpha)) = 2\left(1-\phi\left(\phi^{-1}(1-P_{\mathrm{FA}}/2)\right)\right) = P_{\mathrm{FA}}. \tag{A.2}$$

■

Appendix B. Proof of Proposition 3.10.

We compute in the following proofs the expectation and variance of Kendall's coefficient under the alternative hypothesis H_1.

Under H_1, due to the nature of Gaussian noise, ties are not an issue. We can use the definition given in Def. 3.3 that does not take ties into account. Besides, to simplify the notations, we drop the dependency on (X,Y) in the notation of τ or z. We also drop the conditional hypothesis in the expressions $\mathbb{E}[\tau|H_1]$ and $\mathrm{Var}[\tau|H_1]$, recalling that the results we demonstrate hereafter apply for the H_1 hypothesis.

Preliminary results. This section shows some basic properties that are useful for the computation of the mean and variance of τ. Let ε be a random variable, with density φ. Unless mentioned otherwise, no distribution is supposed for ε. We note $\varepsilon_i, \varepsilon_j$ realizations of the random variable ε, and ϕ the cumulative distributive function (cdf) of ε:

$$\phi(x) = \int_{-\infty}^{x} \varphi(\varepsilon)\,\mathrm{d}\varepsilon \tag{B.1}$$

Proposition B.1. *Let ε_i and ε_j be two independent realizations of ε. Then*

$$\mathbb{E}[\mathrm{sign}(\varepsilon_i - \varepsilon_j)] = 0. \tag{B.2}$$

Proof. Using the definition of the expectation and the independence between the two variables, we have:

$$\begin{aligned}
\mathbb{E}[\mathrm{sign}(\varepsilon_i - \varepsilon_j)] &= \iint_{\mathbb{R}^2} \mathrm{sign}(\varepsilon_i - \varepsilon_j)\varphi(\varepsilon_i)\varphi(\varepsilon_j)\,\mathrm{d}\varepsilon_i\,\mathrm{d}\varepsilon_j \\
&= \int_{\varepsilon_i\in\mathbb{R}} \left(\int_{\varepsilon_j=-\infty}^{\varepsilon_i} \varphi(\varepsilon_j)\,\mathrm{d}\varepsilon_j - \int_{\varepsilon_j=\varepsilon_i}^{+\infty} \varphi(\varepsilon_j)\,\mathrm{d}\varepsilon_j \right) \varphi(\varepsilon_i)\,\mathrm{d}\varepsilon_i \\
&= \int_{\varepsilon_i\in\mathbb{R}} \left(\int_{\varepsilon_j=-\infty}^{\varepsilon_i} \varphi(\varepsilon_j)\,\mathrm{d}\varepsilon_j - \int_{\varepsilon_j=\varepsilon_i}^{+\infty} \varphi(\varepsilon_j)\,\mathrm{d}\varepsilon_j \right) \varphi(\varepsilon_i)\,\mathrm{d}\varepsilon_i \\
&= \int_{\varepsilon_i\in\mathbb{R}} \left(\phi(\varepsilon_i) - (1-\phi(\varepsilon_i))\right) \varphi(\varepsilon_i)\,\mathrm{d}\varepsilon_i \\
&= \int_{\varepsilon_i\in\mathbb{R}} \left(2\phi(\varepsilon_i) - 1\right) \varphi(\varepsilon_i)\,\mathrm{d}\varepsilon_i = \left[\frac{1}{4}\left(2\phi(\varepsilon_i)-1\right)^2\right]_{-\infty}^{+\infty} = \frac{1}{4} - \frac{1}{4} = 0.
\end{aligned}$$

■

Proposition B.2. *Let ε_i, ε_j and ε_k be independent realizations of ε.*

$$\mathbb{E}[\mathrm{sign}(\varepsilon_i - \varepsilon_j)\,\mathrm{sign}(\varepsilon_i - \varepsilon_k)] = \frac{1}{3}. \tag{B.3}$$

Proof. Using once again the independence between the three variables and following the same idea as for Prop. B.1, we have:

$$\begin{aligned}
&\mathbb{E}[\mathrm{sign}(\varepsilon_i - \varepsilon_j)\mathrm{sign}(\varepsilon_i - \varepsilon_k)] \\
&\quad= \iiint_{\mathbb{R}^3} \mathrm{sign}(\varepsilon_i - \varepsilon_j)\,\mathrm{sign}(\varepsilon_i - \varepsilon_k)\varphi(\varepsilon_i)\varphi(\varepsilon_j)\varphi(\varepsilon_k)\,\mathrm{d}\varepsilon_i\,\mathrm{d}\varepsilon_j\,\mathrm{d}\varepsilon_k \\
&\quad= \int_{\varepsilon_i\in\mathbb{R}} \left(\int_{\varepsilon_j\in\mathbb{R}} \mathrm{sign}(\varepsilon_i - \varepsilon_j)\varphi(\varepsilon_j)\,\mathrm{d}\varepsilon_j \int_{\varepsilon_k\in\mathbb{R}} \mathrm{sign}(\varepsilon_i - \varepsilon_k)\varphi(\varepsilon_k)\,\mathrm{d}\varepsilon_k \right) \varphi(\varepsilon_i)\,\mathrm{d}\varepsilon_i \\
&\quad= \int_{\varepsilon_i\in\mathbb{R}} \left(2\phi(\varepsilon_i) - 1\right)^2 \varphi(\varepsilon_i)\,\mathrm{d}\varepsilon_i = \left[\frac{1}{6}\left(2\phi(\varepsilon_i)-1\right)^3\right]_{-\infty}^{+\infty} = \frac{1}{6} + \frac{1}{6} = \frac{1}{3}.
\end{aligned}$$

■

Proposition B.3. *Let ε_i and ε_j be two independent realizations of ε and $a \in \mathbb{R}$.*

$$\mathbb{E}[\mathrm{sign}(\varepsilon_i - \varepsilon_j - a)] = 2\int_{\mathbb{R}} \phi(\varepsilon_i - a)\varphi(\varepsilon_i)\,\mathrm{d}\varepsilon_i - 1. \tag{B.4}$$

Besides, if ε follows a standard distribution $\mathcal{N}(0,\sigma^2)$, then:

$$\mathbb{E}[\mathrm{sign}(\varepsilon_i - \varepsilon_j - a)] = 2\phi\left(\frac{-a}{\sqrt{2}\sigma}\right) - 1 = 1 - 2\phi\left(\frac{a}{\sqrt{2}\sigma}\right). \tag{B.5}$$

Proof. In the general case where no particular distribution is assumed for ε, we use the same idea as for the previous propositions.

$$\mathbb{E}[\mathrm{sign}(\varepsilon_i - \varepsilon_j - a)] = \iint_{\mathbb{R}^2} \mathrm{sign}(\varepsilon_i - \varepsilon_j - a)\varphi(\varepsilon_i)\varphi(\varepsilon_j)\,\mathrm{d}\varepsilon_i\,\mathrm{d}\varepsilon_j$$

$$= \int_{\varepsilon_i \in \mathbb{R}} \left(\int_{\varepsilon_j = -\infty}^{\varepsilon_i - a} \varphi(\varepsilon_j)\, \mathrm{d}\varepsilon_j - \int_{\varepsilon_j = \varepsilon_i - a}^{+\infty} \varphi(\varepsilon_j)\, \mathrm{d}\varepsilon_j \right) \varphi(\varepsilon_i)\, \mathrm{d}\varepsilon_i$$
$$= \int_{\mathbb{R}} \left(2\phi(\varepsilon_i - a) - 1\right) \varphi(\varepsilon_i)\, \mathrm{d}\varepsilon_i = 2 \int_{\mathbb{R}} \phi(\varepsilon_i - a)\varphi(\varepsilon_i)\, \mathrm{d}\varepsilon_i - 1.$$

If $\varepsilon \sim \mathcal{N}(0, \sigma^2)$, then $\varepsilon_i - \varepsilon_j = \mu_{ij}$ follows a Gaussian distribution $\mathcal{N}(0, 2\sigma^2)$, with density $\varphi_{2\sigma^2}$, and

$$\begin{aligned}
\mathbb{E}[\mathrm{sign}(\varepsilon_i - \varepsilon_j - a)] = \mathbb{E}[\mathrm{sign}(\mu_{ij} - a)] &= \int_{\mathbb{R}} \mathrm{sign}(\mu - a)\varphi_{2\sigma^2}(\mu)\mathrm{d}\mu \\
&= \int_{\mathbb{R}} \mathrm{sign}(\sqrt{2\sigma^2}t - a)\varphi(t)\mathrm{d}t, \text{ with } t = \frac{\mu}{\sqrt{2\sigma^2}} \\
&= -\int_{t=-\infty}^{a/\sqrt{2\sigma^2}} \varphi(t)\mathrm{d}t + \int_{t=a\sqrt{2\sigma^2}}^{+\infty} \varphi(t)\mathrm{d}t \\
&= 1 - 2\phi\left(\frac{a}{\sqrt{2}\sigma}\right) = 2\phi\left(\frac{-a}{\sqrt{2}\sigma}\right) - 1.
\end{aligned}$$

■

This last result is confirmed by the following proposition:

Proposition B.4. *Let $\varepsilon \sim \mathcal{N}(0, \sigma^2)$ be a random Gaussian variable. For all $a \in \mathbb{R}$, we have:*

$$\int_{\varepsilon} \phi(\varepsilon - a)\varphi(\varepsilon)\, \mathrm{d}\varepsilon = \phi\left(\frac{-a}{\sqrt{2}\sigma}\right). \tag{B.6}$$

Proof. This proposition can be shown using the property of the error function (defined in Theorem 3.8): for all $a, b, c, d \in \mathbb{R}$,

$$\mathrm{erf}\left(\frac{b - ac}{\sqrt{1 + 2a^2d^2}}\right) = \int_{\mathbb{R}} \frac{\mathrm{erf}(ax + b)}{\sqrt{2\pi d^2}} \exp\left(-\frac{(x + c)^2}{2d^2}\right) \mathrm{d}x. \tag{B.7}$$

■

Proof. [Back to the proof of Prop. 3.10]

Based on Def. 3.3, and using the expression of X and Y defined in (3.9), we separate the cases according to the location of the indexes in the step signal, and we have:

$$\begin{aligned}
\tau &= \frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{j=1}^{n} \mathrm{sign}(X_i - X_j)\, \mathrm{sign}(Y_i - Y_j) \\
&= \frac{1}{n(n-1)} \Bigg[\sum_{\substack{i,j \le n/2 \\ \text{or } i,j > n/2, \\ i \neq j}} \underbrace{\mathrm{sign}(\varepsilon_i - \varepsilon_j)\, \mathrm{sign}(\eta_i - \eta_j)}_{(*)} \\
&\quad + \sum_{\substack{i \le n/2, \\ j > n/2}} \underbrace{\mathrm{sign}(\varepsilon_i - \varepsilon_j - a)\, \mathrm{sign}(\eta_i - \eta_j - a)}_{(**)} + \sum_{\substack{i > n/2, \\ j \le n/2}} \underbrace{\mathrm{sign}(\varepsilon_i - \varepsilon_j + a)\, \mathrm{sign}(\eta_i - \eta_j + a)}_{(***)} \Bigg].
\end{aligned}$$

Each of the three sums can be evaluated separately using the propositions shown in the first part.

$$\mathbb{E}[(*)] = 0 \text{ using prop. B.1 and the independence of } \varepsilon_i \text{ and } \varepsilon_j. \tag{B.8}$$

$$\text{(B.9)}\qquad \mathbb{E}[(**)] = \mathbb{E}[\operatorname{sign}(\varepsilon_i - \varepsilon_j - a)] \times \mathbb{E}[\operatorname{sign}(\eta_i - \eta_j - a)] = \left(1 - 2\phi\left(\frac{a}{\sqrt{2}\sigma}\right)(a)\right)^2,$$

using the independence between the two variables ε,η and Proposition B.3. The sum involving $(**)$ occurs when i and j are in the two different halves of the set of samples, so it contains $\frac{n}{2} \times \frac{n}{2}$ terms.

$$\text{(B.10)}\quad \mathbb{E}[(***)] = \mathbb{E}[\operatorname{sign}(\varepsilon_i - \varepsilon_j + a)] \times \mathbb{E}[\operatorname{sign}(\eta_i - \eta_j + a)] = \left(2\phi\left(\frac{a}{\sqrt{2}\sigma}\right)(a) - 1\right)^2,$$

using Prop. B.3 and the independence. As for $(**)$, $(***)$ occurs when i and j are in different subsets so the sum contains $\frac{n}{2} \times \frac{n}{2}$ terms.

Hence, combining the three terms,

(B.11)

$$\mathbb{E}[\tau|H_1] = \frac{1}{n(n-1)}\left[2 \times \left(\frac{n}{2}\right)^2 \left(1 - 2\phi\left(\frac{a}{\sqrt{2}\sigma}\right)\right)^2\right] = \frac{n}{2(n-1)}\left(1 - 2\phi\left(\frac{a}{\sqrt{2}\sigma}\right)\right)^2.$$

■

Appendix C. Proof of Proposition 3.11.

Proof. In order to assess the variance of τ under the hypothesis H_1, we need to evaluate $\mathbb{E}[\tau^2]$. Using the definition of τ in Def. 3.3, we compute τ^2 then distinguish the cases where X and Y are independent or not. This consists in considering the cases when the indexes can coincide or not.

$$\begin{aligned}\tau^2 = \frac{1}{n^2(n-1)^2}\Bigg[&\sum_{\substack{i,j=1\\ i\neq j}}^{n} \sum_{\substack{k,l=1\\ k\neq l\neq i\neq j}}^{n} \underbrace{\operatorname{sign}(X_i - X_j)\operatorname{sign}(Y_i - Y_j)\operatorname{sign}(X_k - X_l)\operatorname{sign}(Y_k - Y_l)}_{(*)} \\ &+4 \sum_{\substack{i,j=1\\ i\neq j}}^{n} \sum_{\substack{k=1\\ k\neq i,j}}^{n} \underbrace{\operatorname{sign}(X_i - X_j)\operatorname{sign}(Y_i - Y_j)\operatorname{sign}(X_i - X_k)\operatorname{sign}(Y_i - Y_k)}_{(**)} \\ &+2 \underbrace{\sum_{\substack{i,j=1\\ i\neq j}}^{n} (\operatorname{sign}(X_i - X_j)\operatorname{sign}(Y_i - Y_j))^2}_{=n(n-1)}\Bigg].\end{aligned}$$

Each of the three sums is then evaluated separately, taking into account the location of the indexes in the step.

$(*)$: $k \neq l \neq i \neq j$.

$$\begin{aligned}(*) &= \sum_{\substack{i,j=1\\ i\neq j}}^{n} \sum_{\substack{k,l=1\\ k\neq l\neq i\neq j}}^{n} \operatorname{sign}(X_i - X_j)\operatorname{sign}(Y_i - Y_j)\operatorname{sign}(X_k - X_l)\operatorname{sign}(Y_k - Y_l) \\ &= \sum_{\substack{i,j\leq n/2\\ \text{or } i,j>n/2}}^{n} \Bigg[\sum_{\substack{k,l\leq n/2\\ \text{or } k,l>n/2}}^{n} \underbrace{\operatorname{sign}(\varepsilon_i - \varepsilon_j)\operatorname{sign}(\eta_i - \eta_j)\operatorname{sign}(\varepsilon_k - \varepsilon_l)\operatorname{sign}(\eta_k - \eta_l)}_{\mathbb{E}=0}\end{aligned}$$

$$
\begin{aligned}
&+ \sum_{\substack{k \le n/2 \\ l > n/2}}^{n} \underbrace{\operatorname{sign}(\varepsilon_i - \varepsilon_j)\operatorname{sign}(\eta_i - \eta_j)\operatorname{sign}(\varepsilon_k - \varepsilon_l - a)\operatorname{sign}(\eta_k - \eta_l - a)}_{\mathbb{E}=0} \\
&+ \sum_{\substack{k > n/2 \\ l \le n/2}}^{n} \underbrace{\operatorname{sign}(\varepsilon_i - \varepsilon_j)\operatorname{sign}(\eta_i - \eta_j)\operatorname{sign}(\varepsilon_k - \varepsilon_l + a)\operatorname{sign}(\eta_k - \eta_l + a)}_{\mathbb{E}=0}\Big] \\
&+ \sum_{\substack{i \le n/2 \\ j > n/2}}^{n} \Big[\sum_{\substack{k,l \le n/2 \\ \text{or } k,l > n/2}}^{n} \underbrace{\operatorname{sign}(\varepsilon_i - \varepsilon_j - a)\operatorname{sign}(\eta_i - \eta_j - a)\operatorname{sign}(\varepsilon_k - \varepsilon_l)\operatorname{sign}(\eta_k - \eta_l)}_{\mathbb{E}=0} \\
&+ \sum_{\substack{k \le n/2 \\ l > n/2}}^{n} \underbrace{\operatorname{sign}(\varepsilon_i - \varepsilon_j - a)\operatorname{sign}(\eta_i - \eta_j - a)\operatorname{sign}(\varepsilon_k - \varepsilon_l - a)\operatorname{sign}(\eta_k - \eta_l - a)}_{(*1)} \\
&+ \sum_{\substack{k > n/2 \\ l \le n/2}}^{n} \underbrace{\operatorname{sign}(\varepsilon_i - \varepsilon_j - a)\operatorname{sign}(\eta_i - \eta_j - a)\operatorname{sign}(\varepsilon_k - \varepsilon_l + a)\operatorname{sign}(\eta_k - \eta_l + a)}_{(*2)}\Big] \\
&+ \sum_{\substack{i > n/2 \\ j \le n/2}}^{n} \Big[\sum_{\substack{k,l \le n/2 \\ \text{or } k,l > n/2}}^{n} \underbrace{\operatorname{sign}(\varepsilon_i - \varepsilon_j + a)\operatorname{sign}(\eta_i - \eta_j + a)\operatorname{sign}(\varepsilon_k - \varepsilon_l)\operatorname{sign}(\eta_k - \eta_l)}_{\mathbb{E}=0} \\
&+ \sum_{\substack{k \le n/2 \\ l > n/2}}^{n} \underbrace{\operatorname{sign}(\varepsilon_i - \varepsilon_j + a)\operatorname{sign}(\eta_i - \eta_j + a)\operatorname{sign}(\varepsilon_k - \varepsilon_l - a)\operatorname{sign}(\eta_k - \eta_l - a)}_{(*2)} \\
&+ \sum_{\substack{k > n/2 \\ l \le n/2}}^{n} \underbrace{\operatorname{sign}(\varepsilon_i - \varepsilon_j + a)\operatorname{sign}(\eta_i - \eta_j + a)\operatorname{sign}(\varepsilon_k - \varepsilon_l + a)\operatorname{sign}(\eta_k - \eta_l + a)}_{(*3)}\Big]
\end{aligned}
$$

Since $k \neq l \neq i \neq j$ and using the independence, some of these expectations are null. The other terms are evaluated using Prop. B.3 and the independence, and we have:

$$\mathbb{E}[(*1)] = \left(1 - 2\phi\left(\frac{a}{\sqrt{2}\sigma}\right)\right)^4, \tag{C.1}$$

$$\mathbb{E}[(*2)] = \left(1 - 2\phi\left(\frac{a}{\sqrt{2}\sigma}\right)\right)^4, \tag{C.2}$$

$$\mathbb{E}[(*3)] = \left(1 - 2\phi\left(\frac{a}{\sqrt{2}\sigma}\right)\right)^4. \tag{C.3}$$

$(*1)$ and $(*3)$ come from sums where i, j and k, l are in the two distinct subsets, so each term occurs $\left(\frac{n}{2}\left(\frac{n}{2} - 1\right)\right)^2$ times. $(*2)$ appears twice and it comes from sums where i, j and k, l are in the two distinct subsets, so this term occurs $2 \times \left(\frac{n}{2}\left(\frac{n}{2} - 1\right)\right)^2$ times. Adding up the expectations,

$$\mathbb{E}[(*)] = \frac{n^2(n-2)^2}{4}\left(1 - 2\phi\left(\frac{a}{\sqrt{2}\sigma}\right)\right)^4. \tag{C.4}$$

$(**)$: $i \neq j \neq k$. In a similar fashion, we distinguish different cases based on the location of the indexes in the step signal, resulting in several sums that we evaluate separately.

$$
\begin{aligned}
(**) &= \sum_{\substack{i,j=1\\ i\neq j}}^{n} \sum_{\substack{k=1\\ k\neq i,j}}^{n} \operatorname{sign}(X_i - X_j)\operatorname{sign}(Y_i - Y_j)\operatorname{sign}(X_i - X_k)\operatorname{sign}(Y_i - Y_k) \\
&= \sum_{i,j\le n/2}\sum_{k\le n/2} \underbrace{\operatorname{sign}(\varepsilon_i - \varepsilon_j)\operatorname{sign}(\varepsilon_i - \varepsilon_k)\operatorname{sign}(\eta_i - \eta_j)\operatorname{sign}(\eta_i - \eta_k)}_{(*1)} \\
&\quad + \sum_{i,j\le n/2}\sum_{k>n/2} \underbrace{\operatorname{sign}(\varepsilon_i - \varepsilon_j)\operatorname{sign}(\varepsilon_i - \varepsilon_k - a)\operatorname{sign}(\eta_i - \eta_j)\operatorname{sign}(\eta_i - \eta_k - a)}_{(*2)} \\
&\quad + \sum_{i\le n/2}\sum_{j>n/2}\sum_{k\le n/2} \underbrace{\operatorname{sign}(\varepsilon_i - \varepsilon_j - a)\operatorname{sign}(\varepsilon_i - \varepsilon_k)\operatorname{sign}(\eta_i - \eta_j - a)\operatorname{sign}(\eta_i - \eta_k)}_{(*2)} \\
&\quad + \sum_{i\le n/2}\sum_{j>n/2}\sum_{k>n/2} \underbrace{\operatorname{sign}(\varepsilon_i - \varepsilon_j - a)\operatorname{sign}(\varepsilon_i - \varepsilon_k - a)\operatorname{sign}(\eta_i - \eta_j - a)\operatorname{sign}(\eta_i - \eta_k - a)}_{(*4)} \\
&\quad + \sum_{i>n/2}\sum_{j\le n/2}\sum_{k\le n/2} \underbrace{\operatorname{sign}(\varepsilon_i - \varepsilon_j + a)\operatorname{sign}(\varepsilon_i - \varepsilon_k + a)\operatorname{sign}(\eta_i - \eta_j + a)\operatorname{sign}(\eta_i - \eta_k + a)}_{(*5)} \\
&\quad + \sum_{i>n/2}\sum_{j\le n/2}\sum_{k>n/2} \underbrace{\operatorname{sign}(\varepsilon_i - \varepsilon_j + a)\operatorname{sign}(\varepsilon_i - \varepsilon_k)\operatorname{sign}(\eta_i - \eta_j + a)\operatorname{sign}(\eta_i - \eta_k)}_{(*3)} \\
&\quad + \sum_{i,j>n/2}\sum_{k\le n/2} \underbrace{\operatorname{sign}(\varepsilon_i - \varepsilon_j)\operatorname{sign}(\varepsilon_i - \varepsilon_k + a)\operatorname{sign}(\eta_i - \eta_j)\operatorname{sign}(\eta_i - \eta_k + a)}_{(*3)} \\
&\quad + \sum_{i,j>n/2}\sum_{k>n/2} \underbrace{\operatorname{sign}(\varepsilon_i - \varepsilon_j)\operatorname{sign}(\varepsilon_i - \varepsilon_k)\operatorname{sign}(\eta_i - \eta_j)\operatorname{sign}(\eta_i - \eta_k)}_{(*1)}.
\end{aligned}
$$

$\mathbb{E}[(*1)] = \frac{1}{9}$ using Prop. B.3, and the expression appears in two sums where i, j, k are in the same subset, resulting in $2 \times \frac{n}{2}\left(\frac{n}{2}-1\right)\left(\frac{n}{2}-2\right) = \frac{n(n-2)(n-4)}{4}$ terms.

$(*2)$ is the product of two independent terms, so we first evaluate

$$
\begin{aligned}
&\mathbb{E}[\operatorname{sign}(\varepsilon_i - \varepsilon_j) \cdot \operatorname{sign}(\varepsilon_i - \varepsilon_k - a)] \\
&\qquad = \iiint_{\mathbb{R}^3} \operatorname{sign}(\varepsilon_i - \varepsilon_j)\operatorname{sign}(\varepsilon_i - \varepsilon_k - a)\varphi(\varepsilon_i)\varphi(\varepsilon_j)\varphi(\varepsilon_k)\,\mathrm{d}\varepsilon_i\,\mathrm{d}\varepsilon_j\,\mathrm{d}\varepsilon_k \\
&\qquad = \int_{\varepsilon_i\in\mathbb{R}} \left(\int_{\varepsilon_j\in\mathbb{R}} \operatorname{sign}(\varepsilon_i - \varepsilon_j)\varphi(\varepsilon_j\,\mathrm{d}\varepsilon_j \int_{\varepsilon_k\in\mathbb{R}} \operatorname{sign}(\varepsilon_i - \varepsilon_k - a)\varphi(\varepsilon_k)\,\mathrm{d}\varepsilon_k \right) \varphi(\varepsilon_i)\,\mathrm{d}\varepsilon_i \\
&\qquad = \int_{\mathbb{R}} \left(2\phi(\varepsilon_i) - 1\right)\left(2\phi(\varepsilon_i - a) - 1\right)\varphi(\varepsilon_i)\,\mathrm{d}\varepsilon_i \quad \text{with Prop. B.3,} \\
&\qquad = \int_{\mathbb{R}} 2\phi(\varepsilon_i - a)\left(2\phi(\varepsilon_i) - 1\right)\varphi(\varepsilon_i)\,\mathrm{d}\varepsilon_i \\
&\qquad\quad \text{since } \int_{\mathbb{R}} \left(2\phi(\varepsilon_i) - 1\right)\varphi(\varepsilon_i)\,\mathrm{d}\varepsilon_i = 0 \text{ with Prop. B.1,} \\
&\qquad = 4\int_{\mathbb{R}} \phi(\varepsilon_i - a)\phi(\varepsilon_i)\varphi(\varepsilon_i)\,\mathrm{d}\varepsilon_i - 2\underbrace{\int_{\mathbb{R}} \phi(\varepsilon_i - a)\varphi(\varepsilon_i)\,\mathrm{d}\varepsilon_i}_{=\phi\left(\frac{-a}{\sqrt{2}\sigma}\right) \text{ with Prop. B.4}}.
\end{aligned}
$$

In order to evaluate the left integral, we use the fact that $a > 0$ and ϕ is a non-decreasing function (begin a CDF), so $\phi(\varepsilon_i - a) \leq \phi(\varepsilon_i)$. This leads to

$$\int_{\varepsilon_i} \phi(\varepsilon_i - a)\phi(\varepsilon_i)\varphi(\varepsilon_i)\,\mathrm{d}\varepsilon_i \leq \int_{\varepsilon_i} \phi(\varepsilon_i)^2\varphi(\varepsilon_i)\,\mathrm{d}\varepsilon_i = \frac{1}{3}, \text{ so}$$

$$\begin{aligned}\mathbb{E}[\operatorname{sign}(\varepsilon_i - \varepsilon_j)\operatorname{sign}(\varepsilon_i - \varepsilon_k - a)] &\leq \frac{4}{3} - 2\phi\left(\frac{-a}{\sqrt{2}\sigma}\right) \\ &\leq \frac{1}{3} + (1 - 2\phi\left(\frac{-a}{\sqrt{2}\sigma}\right)) = \frac{1}{3} + (2\phi\left(\frac{a}{\sqrt{2}\sigma}\right) - 1).\end{aligned}$$

With the product of the two terms,

$$\mathbb{E}[(*2)] \leq \left(\frac{1}{3} + (2\phi\left(\frac{a}{\sqrt{2}\sigma}\right) - 1)\right)^2, \tag{C.5}$$

and the term $(*2)$ appears $2 \times \frac{n}{2}\frac{n}{2}\left(\frac{n}{2} - 1\right) = \frac{n^2(n-2)}{4}$ times.

$(*3)$ is also a product of two independent terms, so in a similar way we first evaluate

$$\begin{aligned}&\mathbb{E}[\operatorname{sign}(\varepsilon_i - \varepsilon_j) \cdot \operatorname{sign}(\varepsilon_i - \varepsilon_k + a)] \\ &\quad= \iiint_{\mathbb{R}^3} \operatorname{sign}(\varepsilon_i - \varepsilon_j)\operatorname{sign}(\varepsilon_i - \varepsilon_k + a)\varphi(\varepsilon_i)\varphi(\varepsilon_j)\varphi(\varepsilon_k)\,\mathrm{d}\varepsilon_i\,\mathrm{d}\varepsilon_j\,\mathrm{d}\varepsilon_k \\ &\quad= \int_{\varepsilon_i\in\mathbb{R}} \left(\int_{\varepsilon_j\in\mathbb{R}} \operatorname{sign}(\varepsilon_i - \varepsilon_j)\varphi(\varepsilon_j\,\mathrm{d}\varepsilon_j \int_{\varepsilon_k\in\mathbb{R}} \operatorname{sign}(\varepsilon_i - \varepsilon_k + a)\varphi(\varepsilon_k)\,\mathrm{d}\varepsilon_k\right)\varphi(\varepsilon_i)\,\mathrm{d}\varepsilon_i \\ &\quad= \int_{\mathbb{R}} (2\phi(\varepsilon_i) - 1)\,(2\phi(\varepsilon_i + a) - 1)\,\varphi(\varepsilon_i)\,\mathrm{d}\varepsilon_i \text{ using Prop. B.3,} \\ &\quad= \int_{\mathbb{R}} 2\phi(\varepsilon_i + a)\,(2\phi(\varepsilon_i) - 1)\,\varphi(\varepsilon_i)\,\mathrm{d}\varepsilon_i \\ &\qquad\text{since } \int_{\mathbb{R}} (2\phi(\varepsilon_i) - 1)\,\varphi(\varepsilon_i)\,\mathrm{d}\varepsilon_i = 0 \text{ with Prop. B.1,} \\ &\quad= 4\int_{\mathbb{R}} \phi(\varepsilon_i + a)\phi(\varepsilon_i)\varphi(\varepsilon_i)\,\mathrm{d}\varepsilon_i - 2\int_{\mathbb{R}} \phi(\varepsilon_i + a)\varphi(\varepsilon_i)\,\mathrm{d}\varepsilon_i\,.\end{aligned}$$

To evaluate this integral, we use that $\phi(\varepsilon) \leq 1$ for all $\varepsilon \in \mathbb{R}$:

$$\mathbb{E}[\operatorname{sign}(\varepsilon_i - \varepsilon_j)\operatorname{sign}(\varepsilon_i - \varepsilon_k + a)] \leq 4\int_{\mathbb{R}} \phi(\varepsilon_i + a)\varphi(\varepsilon_i)\,\mathrm{d}\varepsilon_i - 2\int_{\mathbb{R}} \phi(\varepsilon_i + a)\varphi(\varepsilon_i)\,\mathrm{d}\varepsilon_i = 2\phi\left(\frac{a}{\sqrt{2}\sigma}\right).$$

With the products of the two terms, this leads to

$$\mathbb{E}[(*3)] \leq 4\phi\left(\frac{a}{\sqrt{2}\sigma}\right)^2, \tag{C.6}$$

and the expression $(*3)$ arises $2 \times \frac{n}{2}\frac{n}{2}\left(\frac{n}{2} - 1\right) = \frac{n^2(n-2)}{4}$ times.

$(*4)$ is handled in the same manner:

$$\mathbb{E}[\operatorname{sign}(\varepsilon_i - \varepsilon_j - a) \cdot \operatorname{sign}(\varepsilon_i - \varepsilon_k - a)]$$

$$
\begin{aligned}
&= \iiint_{\mathbb{R}^3} \operatorname{sign}(\varepsilon_i - \varepsilon_j - a)\operatorname{sign}(\varepsilon_i - \varepsilon_k - a)\varphi(\varepsilon_i)\varphi(\varepsilon_j)\varphi(\varepsilon_k)\,\mathrm{d}\varepsilon_i\,\mathrm{d}\varepsilon_j\,\mathrm{d}\varepsilon_k \\
&= \int_{\varepsilon_i\in\mathbb{R}} \left(\int_{\varepsilon_j\in\mathbb{R}} \operatorname{sign}(\varepsilon_i - \varepsilon_j - a)\varphi(\varepsilon_j\,\mathrm{d}\varepsilon_j \int_{\varepsilon_k\in\mathbb{R}} \operatorname{sign}(\varepsilon_i - \varepsilon_k - a)\varphi(\varepsilon_k)\,\mathrm{d}\varepsilon_k \right) \varphi(\varepsilon_i)\,\mathrm{d}\varepsilon_i \\
&= \int_{\mathbb{R}} \left(2\phi(\varepsilon_i - a) - 1\right)^2 \varphi(\varepsilon_i)\,\mathrm{d}\varepsilon_i \quad \text{using Prop. B.3,} \\
&= 4\int_{\mathbb{R}} \phi(\varepsilon_i - a)^2\varphi(\varepsilon_i)\,\mathrm{d}\varepsilon_i - 4\underbrace{\int_{\mathbb{R}} \phi(\varepsilon_i - a)\varphi(\varepsilon_i)\,\mathrm{d}\varepsilon_i}_{=\phi\left(\frac{-a}{\sqrt{2}\sigma}\right)} + \underbrace{\int_{\mathbb{R}} \varphi(\varepsilon_i)\,\mathrm{d}\varepsilon_i}_{=1}.
\end{aligned}
$$

Using $\phi(\varepsilon_i - a) \leq \phi(\varepsilon_i)$, $\int_{\mathbb{R}} \phi(\varepsilon_i - a)^2\varphi(\varepsilon_i)\,\mathrm{d}\varepsilon_i \leq \int_{\mathbb{R}} \phi(\varepsilon_i)^2\varphi(\varepsilon_i)\,\mathrm{d}\varepsilon_i = \frac{1}{3}$, so

$$
\begin{aligned}
\mathbb{E}[\operatorname{sign}(\varepsilon_i - \varepsilon_j - a) \cdot \operatorname{sign}(\varepsilon_i - \varepsilon_k - a)] &\leq \frac{4}{3} - 4\phi\left(\frac{-a}{\sqrt{2}\sigma}\right) + 1 \\
&\leq \frac{7}{3} - 4\phi\left(\frac{-a}{\sqrt{2}\sigma}\right) = \frac{1}{3} + 2(1 - 2\phi\left(\frac{-a}{\sqrt{2}\sigma}\right)) = \frac{1}{3} + 2(2\phi\left(\frac{a}{\sqrt{2}\sigma}\right) - 1).
\end{aligned}
$$

Taking into account the two multiplied terms, we have

$$
\mathbb{E}[(*4)] \leq \left(\frac{1}{3} + 2(2\phi\left(\frac{a}{\sqrt{2}\sigma}\right) - 1)\right)^2. \tag{C.7}
$$

Expression $(*4)$ refers to a sum where j and k are in the same subset, so the number of terms is $\frac{n}{2}\frac{n}{2}\left(\frac{n}{2} - 1\right) = \frac{n^2(n-2)}{8}$.

Finally, $(*5)$ is evaluated as follows:

$$
\begin{aligned}
\mathbb{E}[\operatorname{sign}(\varepsilon_i - \varepsilon_j + a) \cdot \operatorname{sign}(\varepsilon_i - \varepsilon_k + a)] &= \int_{\mathbb{R}} \left(2\phi(\varepsilon_i + a) - 1\right)^2 \varphi(\varepsilon_i)\,\mathrm{d}\varepsilon_i \\
&= 4\int_{\mathbb{R}} \phi(\varepsilon_i + a)^2\varphi(\varepsilon_i)\,\mathrm{d}\varepsilon_i - 4\int_{\mathbb{R}} \phi(\varepsilon_i + a)\varphi(\varepsilon_i)\,\mathrm{d}\varepsilon_i + \int_{\mathbb{R}} \varphi(\varepsilon_i)\,\mathrm{d}\varepsilon_i.
\end{aligned}
$$

Since $\phi(\varepsilon) \leq 1$ for all $\varepsilon \in \mathbb{R}$, $\phi(\varepsilon)^2 \leq \phi(\varepsilon)$, and

$$
\begin{aligned}
&\mathbb{E}[\operatorname{sign}(\varepsilon_i - \varepsilon_j + a) \cdot \operatorname{sign}(\varepsilon_i - \varepsilon_k + a)] \\
&\leq 4\int_{\mathbb{R}} \phi(\varepsilon_i + a)\varphi(\varepsilon_i)\,\mathrm{d}\varepsilon_i - 4\int_{\mathbb{R}} \phi(\varepsilon_i + a)\varphi(\varepsilon_i)\,\mathrm{d}\varepsilon_i + \underbrace{\int_{\mathbb{R}} \varphi(\varepsilon_i)\,\mathrm{d}\varepsilon_i}_{=1} \leq 1.
\end{aligned}
$$

This leads to

$$
\mathbb{E}[(*5)] \leq 1. \tag{C.8}
$$

As for $(*4)$, the number of terms here is $\frac{n}{2}\frac{n}{2}\left(\frac{n}{2} - 1\right) = \frac{n^2(n-2)}{8}$.

Using the results computed for $(*1) \to (*5)$, the expectation of $(**)$ is bounded by:

$$
\begin{aligned}
\mathbb{E}[(**)] \leq \frac{n(n-2)(n-4)}{4 \times 9} + \frac{n^2(n-2)}{4} \Bigg[&\left(\frac{1}{3} + (2\phi\left(\frac{a}{\sqrt{2}\sigma}\right) - 1)\right)^2 \\
&+ 4\phi\left(\frac{a}{\sqrt{2}\sigma}\right)^2 + \frac{1}{2}\left(\frac{1}{3} + 2(2\phi\left(\frac{a}{\sqrt{2}\sigma}\right) - 1)\right)^2 + \frac{1}{2}\Bigg]
\end{aligned}
$$

$$\leq \frac{n(n-2)(n-4)}{4\times 9} + \frac{n^2(n-2)}{4}\left[\frac{2}{3}+\frac{4}{3}(2\phi\left(\frac{a}{\sqrt{2}\sigma}\right)-1)+3\left(2\phi\left(\frac{a}{\sqrt{2}\sigma}\right)-1\right)^2+4\phi\left(\frac{a}{\sqrt{2}\sigma}\right)^2\right].$$

Recalling the first result on $(*)$ and adding it to $(**)$, we have:

$$\begin{aligned}\mathbb{E}[\tau^2|H_1] &\leq \frac{1}{n^2(n-1)^2}\left[\frac{n^2(n-2)^2}{4}\left(1-2\phi\left(\frac{a}{\sqrt{2}\sigma}\right)\right)^4+\frac{n(n-2)(n-4)}{4\times 9}\right.\\ &\left.+\frac{n^2(n-2)}{4}\left(\frac{2}{3}+\frac{4}{3}(2\phi\left(\frac{a}{\sqrt{2}\sigma}\right)-1)+3\left(2\phi\left(\frac{a}{\sqrt{2}\sigma}\right)-1\right)^2+4\phi\left(\frac{a}{\sqrt{2}\sigma}\right)^2\right)\right]\\ &\leq \frac{1}{(n-1)^2}\left[\frac{(n-2)^2}{4}\left(1-2\phi\left(\frac{a}{\sqrt{2}\sigma}\right)\right)^4+\frac{(n-2)(n-4)}{4\times 9n}\right.\\ &\left.+\frac{(n-2)}{4}\left(\frac{2}{3}+\frac{4}{3}(2\phi\left(\frac{a}{\sqrt{2}\sigma}\right)-1)+3\left(2\phi\left(\frac{a}{\sqrt{2}\sigma}\right)-1\right)^2+4\phi\left(\frac{a}{\sqrt{2}\sigma}\right)^2\right)\right].\end{aligned}$$

In the end, we can derive a bound for the variance of τ under the hypothesis H_1:

$$\begin{aligned}\mathrm{Var}[\tau|H_1] &= \mathbb{E}[\tau^2|H_1]-\mathbb{E}[\tau|H_1]^2\\ &\leq \frac{(n-2)^2}{4(n-1)^2}\left(1-2\phi\left(\frac{a}{\sqrt{2}\sigma}\right)\right)^4+\frac{1}{(n-1)^2}\left[\frac{(n-2)(n-4)}{4\times 9n}\right.\\ &\left.+\frac{(n-2)}{4}\left(\frac{2}{3}+\frac{4}{3}(2\phi\left(\frac{a}{\sqrt{2}\sigma}\right)-1)+3\left(2\phi\left(\frac{a}{\sqrt{2}\sigma}\right)-1\right)^2+4\phi\left(\frac{a}{\sqrt{2}\sigma}\right)^2\right)\right]\\ &-\frac{n^2}{4(n-1)^2}\left(1-2\phi\left(\frac{a}{\sqrt{2}\sigma}\right)\right)^4\\ &\leq \frac{1}{(1-n)}\left(1-2\phi\left(\frac{a}{\sqrt{2}\sigma}\right)\right)^4+\frac{1}{(n-1)^2}\left[\frac{(n-2)(n-4)}{4\times 9n}\right.\\ &\left.+\frac{(n-2)}{4}\left(\frac{2}{3}+\frac{4}{3}(2\phi\left(\frac{a}{\sqrt{2}\sigma}\right)(a)-1)+3\left(2\phi\left(\frac{a}{\sqrt{2}\sigma}\right)-1\right)^2+4\phi\left(\frac{a}{\sqrt{2}\sigma}\right)^2\right)\right].\end{aligned}$$

This bound effectively shows the decrease rate of τ as the number of samples n increases, hence concluding the proof. ■

Appendix D. Proof of Corollary 3.12.

Corollary 3.12 uses the results on the expectation and variance of τ under H_1 to describe the behavior of the associated z-score.

Proof. Under the hypothesis H_1, due to the nature of Gaussian noise, ties are not an issue. Hence, z is defined as $z = \frac{\tau}{\sqrt{\frac{2(2n+5)}{9n(n-1)}}}$. According to Prop. 3.10 and 3.11 that rule the expectation and variance of τ under H_1, the corollary is then immediate. ■

Appendix E. Proof of Theorem 3.13.

Theorem 3.13 controls the probability of miss-detection, as defined in condition (C2).

Proof. Chebyshev's inequality states that if X is a random variable with finite expected value μ and non-zero variance σ^2, then

$$\forall k>0, \Pr(|X-\mu|\geq k) \leq \frac{\sigma^2}{k^2}. \tag{E.1}$$

Using this inequality for Kendall's z-score, we have

$$\Pr(z(X,Y) \leq \mathbb{E}[z(X,Y)|H_1] - k) \leq \Pr(|z(X,Y) - \mu| \geq k) \leq \frac{\mathrm{Var}[z(X,Y)|H_1]}{k^2}. \tag{E.2}$$

Then with $\alpha = \mathbb{E}[z(X,Y)|H_1] - k$,

$$\mathbb{P}(z(X,Y) < \alpha \mid H_1) \leq \frac{\mathrm{Var}(z(X,Y)|H_1)}{(\mathbb{E}[z(X,Y)|H_1] - \alpha)^2}. \tag{E.3}$$

The bound in $O\left(\frac{1}{n}\right)$ is then ensured by Corr. 3.12. ■

Appendix F. Proof of Proposition 4.2.

Proposition 4.2 rules the estimation error between the estimator $\widehat{\sigma}^2$ and the true value σ^2.

Proof.

Following Proposition 4.1, the estimator variance is

$$\mathrm{Var}(\widehat{\sigma}^2) = \frac{\sigma^4}{N}\left(\beta_2 - \frac{N-3}{N-1}\right), \tag{F.1}$$

where $\beta_2 = \mathbb{E}\left[\left(\frac{X-\mu}{\sigma}\right)^4\right] = \frac{\mu_4}{\sigma^4}$ is the kurtosis of the noise distribution and N is the number of pixels in an image block.

If the block size N is large enough, we can use the Law of large numbers to assume that $\widehat{\sigma}^2$ follows a standard distribution, i.e.:

$$\widehat{\sigma}^2 \sim \mathcal{N}(\sigma^2, \mathrm{Var}(\widehat{\sigma}^2)) \Leftrightarrow \widehat{\sigma}^2 - \sigma^2 \sim \mathcal{N}(0, \mathrm{Var}(\widehat{\sigma}^2)) \tag{F.2}$$

Based on the properties that if X follows a normal distribution $\mathcal{N}(0,\sigma^2)$, then the central absolute moments are given by the following formula:

$$\mathbb{E}\left[|X|^p\right] = \sigma^p \frac{2^{\frac{p}{2}}\Gamma\left(\frac{p+1}{2}\right)}{\sqrt{\pi}}. \tag{F.3}$$

We can then deduce the expected value and the variance of the estimation error.

$$\mathbb{E}\left[\frac{\left|\widehat{\sigma}^2 - \sigma^2\right|}{\sigma^2}\right] = \frac{\sqrt{\mathrm{Var}(\widehat{\sigma}^2)}}{\sigma^2} \times \frac{\sqrt{2}}{\sqrt{\pi}}, \quad \text{and} \tag{F.4}$$

$$\begin{aligned}\mathrm{Var}\left(\frac{\left|\widehat{\sigma}^2 - \sigma^2\right|}{\sigma^2}\right) &= \mathbb{E}\left[\frac{\left|\widehat{\sigma}^2 - \sigma^2\right|^2}{\sigma^4}\right] - \mathbb{E}\left[\frac{\left|\widehat{\sigma}^2 - \sigma^2\right|}{\sigma^2}\right]^2 \\ &= \frac{\mathrm{Var}(\widehat{\sigma}^2)}{\sigma^4} - \frac{\mathrm{Var}(\widehat{\sigma}^2)}{\sigma^4} \times \frac{2}{\pi} = \frac{\mathrm{Var}(\widehat{\sigma}^2)}{\sigma^4} \times \left(1 - \frac{2}{\pi}\right).\end{aligned} \tag{F.5}$$

Injecting the variance of the estimator $\widehat{\sigma}^2$ of Proposition 4.1, this concludes the proof. ■

REFERENCES

[1] F. Anscombe, *The transformation of poisson, binomial and negative-binomial data*, Biometrika, (1948), pp. 246–254.

[2] L. Azzari and A. Foi, *Gaussian-cauchy mixture modeling for robust signal-dependent noise estimation*, in IEEE Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP), 2014, pp. 5357–5361.

[3] ———, *Indirect estimation of signal-dependent noise with nonadaptive heterogeneous samples.*, IEEE Trans. Image Process., 23(8):34–59 (2014).
[4] L. Beaurepaire, K. Chehdi, and B. Vozel, *Identification of the nature of noise and estimation of its statistical parameters by analysis of local histograms*, in IEEE Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP), vol. 4:2805–2808, 1997.
[5] J. Boulanger, C. Kervrann, P. Bouthemy, P. Elbau, J.-B. Sibarita, and J. Salamero, *Patch-based nonlocal functional for denoising fluorescence microscopy image sequences*, IEEE Trans. Med. Imag., 29(2):442–454 (2010).
[6] A. Buades, B. Coll, and J.-M. Morel, *A review of image denoising algorithms, with a new one*, SIAM Journal Multiscale Model. Simul., 4(2):490–53 (2005).
[7] A. Chambolle and T. Pock, *A first-order primal-dual algorithm for convex problems with applications to imaging*, Journal of Mathematical Imaging and Vision, 40(1):120–145 (2011).
[8] K. Chehdi and M. Sabri, *A new approach to identify the nature of the noise affecting an image*, in IEEE Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP), vol. 3:285–288, 1992.
[9] Y. Chen and M. Das, *An automated technique for image noise identification using a simple pattern classification approach*, in IEEE Int. Midwest Symp. Circuits and Systems (MWSCAS), 2007, pp. 819–822.
[10] E. Cho and M. Cho, *Variance of sample variance*, in Joint Statistical Meetings, Section on Survey Research Methods, 2008, pp. 1291–1293.
[11] M. Colom and A. Buades, *Analysis and extension of the percentile method, estimating a noise curve from a single image*, Image Process. On Line, 3:332–359 (2013).
[12] ———, *Analysis and extension of the pca method, estimating a noise curve from a single image*, Image Process. On Line, (2014).
[13] M. Colom, M. Lebrun, A. Buades, and J.-M. Morel, *A non-parametric approach for the estimation of intensity-frequency dependent noise*, in IEEE Int. Conf. Image Process. (ICIP), IEEE, 2014, pp. 4261–4265.
[14] A. De Stefano, P. White, and W. Collis, *Training methods for image noise level estimation on wavelet components*, EURASIP Journal on Applied Signal Processing, (2004), pp. 2400–2407.
[15] D. Donoho, *De-noising by soft-thresholding*, IEEE Trans. Inf. Theory, 41(3):613–627 (1995).
[16] A. Foi, M. Trimeche, V. Katkovnik, and K. Egiazarian, *Practical Poissonian-Gaussian noise modeling and fitting for single-image raw-data*, IEEE Trans. Image Process., 17(10):1737–1754 (2008).
[17] W. Freeman, T. Jones, and E. Pasztor, *Example-based super-resolution*, IEEE Comput. Graph. Appl., 22(2):56–65 (2002).
[18] T. Gasser, L. Sroka, and C. Jennen-Steinmetz, *Residual variance and residual pattern in nonlinear regression*, Biometrika, 73(3):625–633 (1986).
[19] G. Healey and R. Kondepudy, *Radiometric ccd camera calibration and noise estimation*, IEEE Trans. Pattern Anal. Mach. Intell., 16(3):267–276 (1994).
[20] M. Hensel, T. Pralow, and R. Grigat, *Modeling and real-time estimation of signal-dependent noise in quantum-limited imaging*, in WSEAS Int. Conf. Signal Processing, Robotics and Automation (ISPRA), 2007, pp. 183–191.
[21] P. Huber et al., *Robust estimation of a location parameter*, The Annals of Statistics, 35(1):73–101 (1964).
[22] M. Kendall, *A new measure of rank correlation*, Biometrika, 30(1–2):81–93 (1938).
[23] ———, *The treatment of ties in ranking problems*, Biometrika, (1945), pp. 239–251.
[24] ———, *Rank correlation methods.*, Griffin, 1948.
[25] W. Knight, *A computer method for calculating kendall's tau with ungrouped data*, Journal of the American Statistical Association, 61(314):436–439 (1966).
[26] R. Koenker, *Quantile regression*, Cambridge university press, 2005.
[27] M. Lebrun, M. Colom, and J.-M. Morel, *The noise-clinic: a universal blind denoising algorithm*, in IEEE Int. Conf. Image Process. (ICIP), 2014, pp. 2674–2678.
[28] ———, *The noise-clinic: a blind image denoising algorithm*, Image Process. On Line, 5:1–54 (2015).
[29] C. Liu, W. Freeman, R. Szeliski, and S. Kang, *Noise estimation from a single image*, in IEEE Comput. Vis. and Pattern Recognition (CVPR), vol. 1:901–908, 2006.
[30] X. Liu, M. Tanaka, and M. Okutomi, *Single-image noise level estimation for blind denoising*, IEEE Trans. Image Process., 22(12):5226–5237 (2013).
[31] D. Lowe, *Distinctive image features from scale-invariant keypoints*, International Journal of Computer Vision, 60(2):91–110 (2004).
[32] M. Makitalo and A. Foi, *Noise parameter mismatch in variance stabilization, with an application to poisson–gaussian noise estimation*, IEEE Trans. Image Process., 23(12):5348–5359 (2014).
[33] S. Olsen, *Estimation of noise in images: an evaluation*, Graphical Models and Image Processing

(CVGIP), 55(4):319–323 (1993).
[34] D. Pollard, *Asymptotics for least absolute deviation regression estimators*, Econometric Theory, 7(2):186–199 (1991).
[35] J. Portilla, *Blind non-white noise removal in images using gaussian scale mixtures in the wavelet domain*, in Benelux Signal Processing Symp., 2004.
[36] S. Pyatykh and J. Hesser, *Image sensor noise parameter estimation by variance stabilization and normality assessment*, IEEE Trans. Image Process., 23(9):3990–3998 (2014).
[37] S. Pyatykh, J. Hesser, and L. Zheng, *Image noise level estimation by principal component analysis*, IEEE Trans. Image Process., 22(2):687–699 (2013).
[38] K. Rank, M. Lendl, and R. Unbehauen, *Estimation of image noise variance*, Proc. Vision, Image, and Signal Processing, 146(2):80–84 (1999).
[39] D. Shin, R. Park, S. Yang, and J. Jung, *Block-based noise estimation using adaptive gaussian filtering*, IEEE Trans. Consum. Electron., 51(1):218–226 (2005).
[40] C. Sutour, C.-A. Deledalle, and J.-F. Aujol, *Adaptive regularization of the NL-means: Application to image and video denoising*, IEEE Trans. Image Process., 23(8):3506–3521 (2014).
[41] S. Tai and S. Yang, *A fast method for image noise estimation using laplacian operator and adaptive edge detection*, in Int. Symp. Communications, Control and Signal Processing (ISCCSP), 2008, pp. 1077–1081.
[42] H. Theil, *A rank-invariant method of linear and polynomial regression analysis*, in Henri Theil Contributions to Economics and Econometrics, Springer, 1992, pp. 345–381.
[43] D. Zoran and Y. Weiss, *Scale invariance and noise in natural images*, in Int. Conf. Computer Vision (ICCV), 2009, pp. 2209–2216.