An Unbiased Risk Estimator for Image Denoising in the Presence of Mixed Poisson–Gaussian Noise

Yoann Le Montagner, *Student Member, IEEE*, Elsa D. Angelini, *Senior Member, IEEE*,
and Jean-Christophe Olivo-Marin, *Fellow, IEEE*

***Abstract*— The behavior and performance of denoising algorithms are governed by one or several parameters, whose optimal settings depend on the content of the processed image and the characteristics of the noise, and are generally designed to minimize the mean squared error (MSE) between the denoised image returned by the algorithm and a virtual ground truth. In this paper, we introduce a new Poisson–Gaussian unbiased risk estimator (PG-URE) of the MSE applicable to a mixed Poisson–Gaussian noise model that unifies the widely used Gaussian and Poisson noise models in fluorescence bioimaging applications. We propose a stochastic methodology to evaluate this estimator in the case when little is known about the internal machinery of the considered denoising algorithm, and we analyze both theoretically and empirically the characteristics of the PG-URE estimator. Finally, we evaluate the PG-URE-driven parametrization for three standard denoising algorithms, with and without variance stabilizing transforms, and different characteristics of the Poisson–Gaussian noise mixture.**

***Index Terms*— Denoising, Stein's unbiased risk estimate (SURE), MSE estimation, mixed Poisson–Gaussian noise, Monte Carlo methods, fluorescence microscopy, bioimaging, PG-URE.**

I. INTRODUCTION

A. Denoising Background

Image denoising is one of the most studied problem in image processing. Many algorithms have been developed to tackle this issue, with various characteristics in terms of denoising efficiency, applicability to different types of images and noise models, and running time. Among this large collection of available methods, we can single out the very classical ones: wavelet soft-thresholding [1], which has a low algorithmic complexity and can be applied quickly even on large 2D or 3D signals; total-variation (TV) based methods [2], which are very efficient in removing noise while preserving sharp edges in cartoon-like images; non-local means (NLM) [3], which exploits patch similarities that exist in natural images and is very powerful in presence of textured content. Most of the state-of-the-art denoising algorithms [4], [5] consist in refinements of and crossings between these classical ones: for instance, BM3D [5] consists in looking for image patches that present similarities (as in [3]), and then applying a thresholding operation on group of similar patches (in the manner of [1]). One can refer to [6] for a more comprehensive overview of filtering methods applied to the denoising problem.

Notation	Definition
N	Number of pixels of the considered images.
x_k, f_k	k^{th} component of a rasterized image $\boldsymbol{x}$ or a function $\boldsymbol{f}$ returning an image.
$\mathbf{1}$	Flat image with all pixels equal to 1.
$\boldsymbol{e}_k$	k^{th} image of the canonical basis: all pixels are 0, except the k^{th} one which is set to 1.
$\boldsymbol{u} \times \boldsymbol{v}$	Component-wise product between two images $\boldsymbol{u}$ and $\boldsymbol{v}$.
$\dfrac{\partial f_k}{\partial y_l}(\boldsymbol{y})$	Partial derivative of the k^{th} component of an image-valued function $\boldsymbol{f}$ along the l^{th} component of the input.
$\boldsymbol{\partial}^p \boldsymbol{f}(\boldsymbol{y})$	Image-valued object for which the value of each pixel k is equal to $\frac{\partial^p f_k}{\partial {y_k}^p}(\boldsymbol{y})$. For $p = 1$, the superscript is implicit.
$\boldsymbol{f}^{[-\zeta]}(\boldsymbol{y})$	Image-valued object for which the value of each pixel k is equal to $f_k(\boldsymbol{y} - \zeta \boldsymbol{e}_k)$ (with ζ a scalar parameter).
$\mathbb{E}\{\cdot\}$ $\mathbb{V}\mathrm{ar}\{\cdot\}$	Expected value and variance of a random variable. In this paper, the ground truth signals are considered to be deterministic, therefore the only sources of randomness come from the noise and the numerically generated random vectors $\boldsymbol{\delta}$ (in Sec. III).

Fig. 1. Common notations used in the paper. Bold font is used to denote images and image-valued functions.

All these algorithms have one or several parameters, whose optimal values are almost always dependent on the data being processed. More precisely, if $\boldsymbol{y}$ is the noisy image being observed, $\boldsymbol{f}_\theta$ a denoising algorithm depending on a set of parameters θ, and $\hat{\boldsymbol{x}} = \boldsymbol{f}_\theta(\boldsymbol{y})$ the denoised image returned by the algorithm, it is often desirable to select θ such that it optimizes a similarity criteria between $\hat{\boldsymbol{x}}$ and a *ground truth* noise-free image $\boldsymbol{x}$. In this paper, we will focus on the mean

---

Manuscript received July 21, 2013; revised November 26, 2013; accepted January 2, 2014. Date of publication January 16, 2014; date of current version February 4, 2014. This work was supported in part by a Fellowship from EDX, and in part by ANR under Grant ANR-10-INBS-04-06-FranceBioImaging and Grant ANR-10-LABX-62-IBEID. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Scott T. Acton.

Y. Le Montagner is with the Institut Mines-Télécom, Télécom ParisTech, Paris 75013, France, and also with the Institut Pasteur, Quantitative Image Analysis Unit, Paris 75724, France (e-mail: yoann.le-montagner_ieee@m4x.org).

E. D. Angelini is with the Institut Mines-Télécom, Télécom ParisTech, Paris 75013, France (e-mail: elsa.angelini@telecom-paristech.fr).

J.-C. Olivo-Marin is with the Institut Pasteur, Quantitative Image Analysis Unit, Paris 75724, France (e-mail: jcolivo@pasteur.fr).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TIP.2014.2300821

squared error (MSE) criteria:[1]

$$\mathrm{MSE} = \frac{1}{N} \left\| \boldsymbol{f}_\theta(\boldsymbol{y}) - \boldsymbol{x} \right\|_2^2 \tag{1}$$

where $\|\cdot\|_2$ is the l_2-norm applied on vectorized images (see Fig. 1 for details on notations).

However, except in special contexts such as simulation when the ground truth $\boldsymbol{x}$ is known, the MSE (1) is impossible to evaluate directly and cannot be used as an objective criteria to optimize the parameters θ of the $\boldsymbol{f}_\theta$ algorithm. The unbiased risk estimator tools, among which SURE [8], [9] is a well-known representative, aim at tackling this issue.

B. SURE and Parameter Estimation

Stein's unbiased risk estimator (SURE) [8] is a well-known tool in the statistics field, that has recently received a growing interest from the image processing community (see for instance [9], [10], and [11]).

The SURE estimator is built upon the hypothesis that the ground truth $\boldsymbol{x}$ is corrupted by a white additive Gaussian noise $\boldsymbol{b}$:

$$\boldsymbol{y} = \boldsymbol{x} + \boldsymbol{b} \quad \text{with} \quad \boldsymbol{b} \sim \mathcal{N}\left(0, \sigma^2\,\mathrm{Id}\right) \tag{2}$$

where $\boldsymbol{y}$ is the observed noisy image, and where the standard deviation parameter σ is assumed to be known. From this noise model, and given a denoising function[2] $\boldsymbol{f}$, a similarity criteria SURE is defined as:

$$\mathrm{SURE} = \frac{1}{N} \left\| \boldsymbol{f}(\boldsymbol{y}) - \boldsymbol{y} \right\|_2^2 - \sigma^2 + \frac{2\sigma^2}{N} \operatorname{Div} \boldsymbol{f}(\boldsymbol{y}) \tag{3}$$

where $\operatorname{Div} \boldsymbol{f}(\boldsymbol{y}) = \sum_k \frac{\partial f_k}{\partial y_k}(\boldsymbol{y})$ stands for the divergence of the function $\boldsymbol{f}$.

In [8], the author showed that, up to some technical points,[3] MSE and SURE have equal expected values over all the realizations of the random variable $\boldsymbol{b}$: $\mathbb{E}\{\mathrm{MSE}\} = \mathbb{E}\{\mathrm{SURE}\}$. This means that, in practice, SURE is an estimator of the MSE similarity criteria, and can be taken as a surrogate. The empirical equality SURE $\approx$ MSE has been confirmed in various particular situations: see for instance [12] and [11].

A significant difference between MSE and SURE is that the latter does not depend on the ground truth $\boldsymbol{x}$. As $\boldsymbol{x}$ is generally not available in real-life problems, this property dramatically increases the interest of SURE over MSE in practical applications. For instance, if $\theta_1, \ldots, \theta_K$ are K admissible parameter values for a denoising algorithm $\boldsymbol{f}_\theta$, it is possible to select a "best-performing" value $\theta_{k^\star}$ as the one that minimizes $\mathrm{SURE}(\theta_k)$. Such selection is data-adaptive (it depends on $\boldsymbol{y}$), and objective (it does not rely on human expert evaluation), opening the way to automated parameter estimation.

C. Paper Outline

The paper is built around the resolution of two issues that restrict in practice the use of SURE for automatic parameter tuning. First, SURE relies on the hypothesis of additive white Gaussian noise (2), which may not account for situations encountered in bio-imaging applications: for example, in this case, noise intensity may not be uniform in the whole image as assumed in (2), but rather depend on the presence of biological objects, and more generally on the value on the underlying signal (see [13] and [14]). The extension of SURE to a more realistic mixed Poisson-Gaussian noise model is thus proposed in Sec. II, extending the work in [15].

The second limitation comes from the divergence term that appears in the expression of the SURE estimator (3). More precisely, the evaluation of the partial derivatives $\frac{\partial f_k}{\partial y_k}(\boldsymbol{y})$ is not a trivial task when the denoising algorithm $\boldsymbol{f}$ is not defined by a closed-form expression: such situations include variational-based algorithms (e.g. total variation minimization [2]) and diffusion methods (e.g. anisotropic diffusion [16]). To tackle this issue, a methodology based on the introduction of small stochastic perturbations to $\boldsymbol{y}$ (similar to the one introduced by [12]) is proposed in Sec. III.

A numerical validation of the proposed framework is presented in Sec. IV, along with several practical examples of parameter estimation.

II. MIXED POISSON–GAUSSIAN NOISE MODEL

A. Generalized Unbiased Risk Estimators

The original SURE estimator [8] (3) was designed around the Gaussian noise hypothesis (2). Other types of unbiased risk estimators have been derived since to handle different noise models. It is worth noting that unbiased risk estimators can be refined to account for several phenomena that affect the image formation, beyond simple noise: see for instance [17]–[20] and references therein for applications of SURE-like estimators to deconvolution problems. An exhaustive review of the existing unbiased risk estimators applied to image restoration problems is beyond the scope of this paper, and we focus in this work just on pure denoising problems involving noise models encountered in microscopy imaging applications.

B. Poisson Noise and Associated PURE Estimator

A usual noise model in bioimaging is the Poisson model, which is quite common in low-light fluorescence microscopy imaging, and more generally in imaging modalities that operate in low-signal conditions (see for instance [13] and [14]). In this model, each observed pixel value y_k is assumed to be the result of a Poisson random process of intensity x_k, independent of the other pixels y_l. Formally:

$$\boldsymbol{y} \sim \mathcal{P}(\boldsymbol{x}) \tag{4}$$

A qualitative property of Poisson images is that the noise variance is signal dependent, and increases with the underlying

---

[1] Several image similarity criteria exist (see for instance [7]), but an exhaustive comparison and discussion of their respective qualities is beyond the scope of this paper. The MSE is not the best one in terms of correlation to the human perception system, but its mathematical tractability makes it a valuable tool in image processing.

[2] From now on, we will drop the subscript θ from $\boldsymbol{f}_\theta$ for the sake of readability, when no ambiguity is possible.

[3] For the following result to hold, $\boldsymbol{f}$ must be weakly differentiable, and its partial derivatives must fulfil $\mathbb{E}\left\{\sum_k \left|\frac{\partial f_k}{\partial y_k}(\boldsymbol{y})\right|\right\} < +\infty$. These technical conditions will always be assumed to be true, as well as all other requirements on the regularity of $\boldsymbol{f}$ that could be encountered in the paper. Please note however that some realistic denoising functions $\boldsymbol{f}$ may not be even weak-differentiable: for instance, wavelet hard-thresholding is not.

intensity of the signal. This behavior is fundamentally different from what is modeled with the additive white Gaussian noise hypothesis (2), for which the noise intensity is uniform and independent of the value of the ground truth signal.

A *Poisson unbiased risk estimator* (PURE) of the MSE similarity criteria has been derived in [21] for the Poisson noise model (4):

$$\mathrm{PURE}=\frac{1}{N}\left(\|\boldsymbol{f}(\boldsymbol{y})\|_2^2+\|\boldsymbol{y}\|_2^2-2\left\langle\boldsymbol{y}\middle|\boldsymbol{f}^{[-1]}(\boldsymbol{y})\right\rangle-\langle\mathbf{1}|\boldsymbol{y}\rangle\right)\quad(5)$$

where $\langle\cdot|\cdot\rangle$ stands for the usual inner product between vectorized images, and where the image-valued function $\boldsymbol{f}^{[-1]}(\boldsymbol{y})$ is defined in Fig. 1. For smooth functions $\boldsymbol{f}$, a first-order Taylor approximation $\boldsymbol{f}^{[-1]}(\boldsymbol{y})\approx\boldsymbol{f}(\boldsymbol{y})-\boldsymbol{\partial f}(\boldsymbol{y})$ can be considered to simplify (5) into:

$$\mathrm{PURE}=\frac{1}{N}\left(\|\boldsymbol{f}(\boldsymbol{y})-\boldsymbol{y}\|_2^2+2\,\langle\boldsymbol{y}|\boldsymbol{\partial f}(\boldsymbol{y})\rangle-\langle\mathbf{1}|\boldsymbol{y}\rangle\right)\quad(6)$$

The terms $\left\langle\boldsymbol{y}\middle|\boldsymbol{f}^{[-1]}(\boldsymbol{y})\right\rangle$ in (5) and $\langle\boldsymbol{y}|\boldsymbol{\partial f}(\boldsymbol{y})\rangle$ in (6) play roles similar to the divergence term in SURE (3), in that they probe how small modifications of the observed image $\boldsymbol{y}$ impact the output of the denoising algorithm $\boldsymbol{f}$. Their evaluation are subject to technical difficulties similar to those mentioned in Sec. I-C for SURE.

C. Mixed Poisson–Gaussian Noise

The Gaussian and Poisson noise models (2) and (4) do not individually account for the various phenomena involved with real image acquisition processes in fluorescence microscopy. Therefore, in the following, we consider a *mixed Poisson-Gaussian* (MPG) noise model, similar to the ones proposed in [13] and [22], [23], [24]:

$$\boldsymbol{y}=\zeta\boldsymbol{z}+\boldsymbol{b}\quad\text{with}\quad\begin{cases}\boldsymbol{z}\sim\mathcal{P}\left(\dfrac{\boldsymbol{x}}{\zeta}\right)\\ \boldsymbol{b}\sim\mathcal{N}\left(0,\sigma^2\,\mathrm{Id}\right)\end{cases}\quad(7)$$

where $\boldsymbol{z}$ and $\boldsymbol{b}$ are two independent random variables, following respectively a Poisson and a Gaussian distribution. This noise model introduces two numerical parameters:

- $\sigma\geq 0$ is the standard deviation of $\boldsymbol{b}$; the higher this parameter, the more the model (7) behaves like a pure Gaussian noise model.
- $\zeta\geq 0$ is the *gain* of the acquisition process;[4] the higher this parameter, the more Poisson-like is the behavior of the noise in (7).

It can be noted that the proposed MPG noise model (7) encompasses the classical Gaussian and Poisson noise models: setting $\zeta=0$ and $\sigma>0$ corresponds to the Gaussian noise model (2), while $\zeta=1$ and $\sigma=0$ leads to the Poisson noise model (4).

In what follows, we will always assume that the values of the noise parameters σ and ζ are known. However, it is worth noting that estimating these parameters from a given noisy observation $\boldsymbol{y}$ is not trivial. In particular, as noticed in [24] and [25], the cumulant based approach advised in [26] leads to unreliable estimates of the gain parameter ζ. This is due to the fact that this approach makes use of high-order empirical moments (order ≥ 3) evaluated on the noisy signal, which leads to numerical instability. As an alternative, [24] proposes an expectation-maximization approach to address this parameter estimation issue, which provides more stable and reliable estimates.

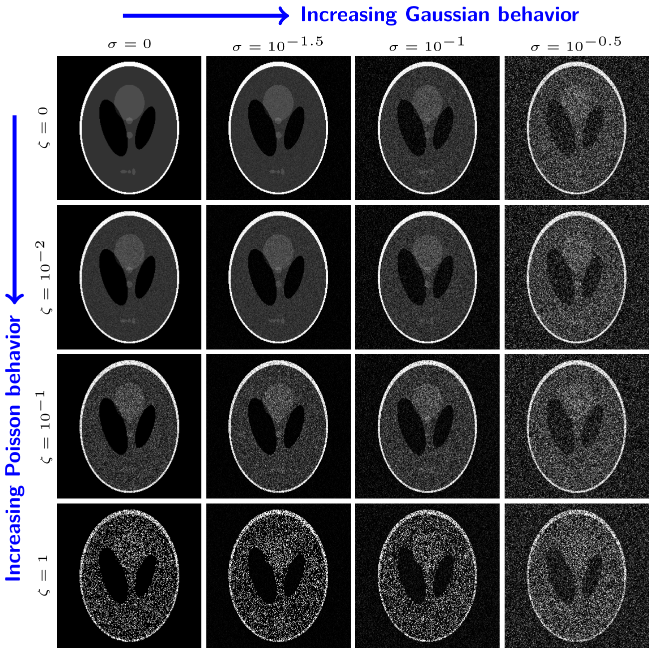

Fig. 2. Example of Shepp-Logan images $\boldsymbol{y}$ corrupted with the mixed Poisson-Gaussian noise model (7), for different values of the two parameters σ and ζ. For $\sigma=0$ and $\zeta=0$ (upper left), the image is identical to the the ground truth $\boldsymbol{x}$ (i.e. the original Shepp-Logan image).

D. Unbiased Risk Estimator for the MPG Model

Extending the pioneer work in [15], we derive the *Poisson-Gaussian unbiased risk estimator* (PG-URE) of the MSE for the MPG model (7):

$$\begin{aligned}\text{PG-URE}=\frac{1}{N}\Big(&\|\boldsymbol{f}(\boldsymbol{y})\|_2^2+\|\boldsymbol{y}\|_2^2-2\left\langle\boldsymbol{y}\middle|\boldsymbol{f}^{[-\zeta]}(\boldsymbol{y})\right\rangle\\&+2\sigma^2\,\mathrm{Div}\,\boldsymbol{f}^{[-\zeta]}(\boldsymbol{y})-\zeta\,\langle\mathbf{1}|\boldsymbol{y}\rangle\Big)-\sigma^2\end{aligned}\quad(8)$$

where the notation $\boldsymbol{f}^{[-\zeta]}(\boldsymbol{y})$ is defined in Fig. 1. The derivation of (8) and the proof that $\mathbb{E}\{\text{PG-URE}\}=\mathbb{E}\{\text{MSE}\}$ are given in appendix A, along with the technical conditions required on $\boldsymbol{f}$ for this result to hold. As for the Poisson model, if $\boldsymbol{f}^{[-\zeta]}(\boldsymbol{y})$ is replaced by its first-order Taylor expansion $\boldsymbol{f}^{[-\zeta]}(\boldsymbol{y})\approx\boldsymbol{f}(\boldsymbol{y})-\zeta\boldsymbol{\partial f}(\boldsymbol{y})$, this leads to the following simplified expression of the PG-URE estimator:

$$\begin{aligned}\text{PG-URE}=\frac{1}{N}\Big(&\|\boldsymbol{f}(\boldsymbol{y})-\boldsymbol{y}\|_2^2+2\left\langle\zeta\boldsymbol{y}+\sigma^2\mathbf{1}\middle|\boldsymbol{\partial f}(\boldsymbol{y})\right\rangle\\&-2\sigma^2\zeta\left\langle\mathbf{1}\middle|\boldsymbol{\partial}^2\boldsymbol{f}(\boldsymbol{y})\right\rangle-\zeta\,\langle\mathbf{1}|\boldsymbol{y}\rangle\Big)-\sigma^2\end{aligned}\quad(9)$$

It should be noted that this simplified expression (9) of PG-URE may significantly deviate from (8) in the case of

---

[4] By convention, when $\zeta=0$, the MPG model must be understood as $\boldsymbol{y}=\boldsymbol{x}+\boldsymbol{b}$ (i.e. pure Gaussian noise (2)). This extension is motivated by the fact that the random variable $\zeta\boldsymbol{z}$ with $\boldsymbol{z}\sim\mathcal{P}\left(\frac{\boldsymbol{x}}{\zeta}\right)$ converges in law to $\boldsymbol{x}$ (deterministic value) when $\zeta\to 0$.

large values of the gain parameter ζ, due to the Taylor approximation $\boldsymbol{f}^{[-\zeta]}(\boldsymbol{y}) \approx \boldsymbol{f}(\boldsymbol{y}) - \zeta \boldsymbol{\partial f}(\boldsymbol{y})$. However, the numerical results presented in Sec. IV show that this deviation has no consequence in the range of gain values encountered in practice.

It can be verified that the expressions (8)-(9) of the PG-URE estimator are consistent with SURE (3) and PURE (5)-(6) for the special values of the parameters σ and ζ mentioned in Sec. II-C. They are also consistent with the unbiased risk estimator derived in [15] for a simpler mixed Poisson-Gaussian noise model that does not integrate a gain parameter ζ.

Again, evaluation of the terms involving $\boldsymbol{f}^{[-\zeta]}$, $\boldsymbol{\partial f}$ or $\boldsymbol{\partial^2 f}$ in (8)-(9) raises some technical difficulties: in the next section, we propose a non-deterministic method to handle them.

III. STOCHASTIC EVALUATION OF THE POISSON–GAUSSIAN UNBIASED RISK ESTIMATOR

A. Why is a Deterministic Evaluation of PG-URE Impossible?

The expressions (8) and (9) define unbiased risk estimator of the MSE (1) under a mixed Poisson-Gaussian noise model hypothesis (7). These expressions do not involve non-accessible entities such as the ground truth $\boldsymbol{x}$, making their numerical evaluation conceivable in practical settings. However, the terms involving $\boldsymbol{f}^{[-\zeta]}$, $\boldsymbol{\partial f}$ or $\boldsymbol{\partial^2 f}$ may not be directly computable, as explained below.

For instance, let us assume that the denoising algorithm $\boldsymbol{f}$ is modeled as a *black-box* process, meaning that we do not make any assumption on how $\boldsymbol{f}$ works internally, and therefore that the only available “action” with $\boldsymbol{f}$ is to submit an input $\boldsymbol{y}$ and to retrieve an output $\boldsymbol{f}(\boldsymbol{y})$. Then, due to its definition, a direct evaluation of $\boldsymbol{f}^{[-\zeta]}(\boldsymbol{y})$ would require to run $\boldsymbol{f}$ on N perturbed versions of the input $\boldsymbol{y}$: $(\boldsymbol{y} - \zeta \boldsymbol{e}_k)$ for $k = 1$ to N. As N represents the number of pixels in the input image, such direct evaluation would be computationally irrealistic even with images of reasonable size. The same argument holds for the terms $\boldsymbol{\partial f}$ and $\boldsymbol{\partial^2 f}$, that could be approximated through finite differences: for instance, the first order difference $\frac{1}{\epsilon}\big(f_k(\boldsymbol{y} + \epsilon \boldsymbol{e}_k) - f_k(\boldsymbol{y})\big)$ for some small scalar parameter ϵ would provide a good approximation of the k^{th} component of $\boldsymbol{\partial f}(\boldsymbol{y})$, but computing all the components of this term through this scheme would require to evaluate $\boldsymbol{f}(\boldsymbol{y} + \epsilon \boldsymbol{e}_k)$ for $k = 1$ to N, which is again irrealistic.

The method developed in the following sections bypasses these problems, thanks to a stochastic scheme to evaluate the Taylor-expanded PG-URE estimator (9) in the context of the black-box denoising process mentioned above. One key advantage of this method is that the required number of evaluations of $\boldsymbol{f}$ – i.e. the significant factor in terms of computation time – is small and does not depend on N.

B. Evaluation of the First-Order Derivative Term

We first focus on the term involving the first-order partial derivatives of $\boldsymbol{f}$ in (9), namely $\big\langle \zeta \boldsymbol{y} + \sigma^2 \mathbf{1} \big| \boldsymbol{\partial f}(\boldsymbol{y}) \big\rangle$. The idea of the proposed method, which is a direct extension of the *Monte-Carlo SURE* approach proposed in [12], is to probe the behavior of $\boldsymbol{f}$ when applied on slightly modified versions of $\boldsymbol{y}$, which are obtained by adding some well-chosen random perturbations to $\boldsymbol{y}$.

Let us introduce a few notations: in what follows, $\epsilon > 0$ is a scalar parameter whose value is ideally as small as possible, $\boldsymbol{\delta}$ is a random perturbation vector generated according to a probability distribution to be specified, and $\langle \boldsymbol{u} | \boldsymbol{\partial f}(\boldsymbol{y}) \rangle$ is the quantity to evaluate.[5] Then, assuming that $\boldsymbol{f}$ is continuously differentiable, we have:

$$\boldsymbol{f}(\boldsymbol{y} + \epsilon \boldsymbol{\delta}) = \boldsymbol{f}(\boldsymbol{y}) + \epsilon \sum_l \delta_l \frac{\partial \boldsymbol{f}}{\partial y_l}(\boldsymbol{y}) + \epsilon \boldsymbol{r}(\epsilon) \tag{10}$$

where $\boldsymbol{r}(\epsilon)$ is some remainder that tends to 0 when $\epsilon \to 0$. From this Taylor expansion, it results that:

$$\lim_{\epsilon \to 0} \left\langle \boldsymbol{\delta} \times \boldsymbol{u} \left| \frac{\boldsymbol{f}(\boldsymbol{y} + \epsilon \boldsymbol{\delta}) - \boldsymbol{f}(\boldsymbol{y})}{\epsilon} \right. \right\rangle = \sum_{k,l} u_k \delta_k \delta_l \frac{\partial f_k}{\partial y_l}(\boldsymbol{y}) \tag{11}$$

where each summation index k and l visits every components of the involved vectors.

Now, let us assume the following properties on the probability distribution of the random perturbation $\boldsymbol{\delta}$:

- the components δ_k of $\boldsymbol{\delta}$ are independent,
- each δ_k has an expected value of 0 and a variance equal to 1.

Then, by considering the expected value[6] over the random variable $\boldsymbol{\delta}$ on both sides of the equality (11), we obtain:

$$\mathbb{E}_{\boldsymbol{\delta}} \left\{ \lim_{\epsilon \to 0} \left\langle \boldsymbol{\delta} \times \boldsymbol{u} \left| \frac{\boldsymbol{f}(\boldsymbol{y}+\epsilon \boldsymbol{\delta}) - \boldsymbol{f}(\boldsymbol{y})}{\epsilon} \right. \right\rangle \right\} = \sum_k u_k \frac{\partial f_k}{\partial y_k}(\boldsymbol{y}) \tag{12}$$

Finally, up to some technical hypothesis (see [12] for more details) which are also important to derive the empirical formula (14), the expectation and the limit in (12) can be switched, leading to the final expression:

$$\lim_{\epsilon \to 0} \mathbb{E}_{\boldsymbol{\delta}} \left\{ \left\langle \boldsymbol{\delta} \times \boldsymbol{u} \left| \frac{\boldsymbol{f}(\boldsymbol{y} + \epsilon \boldsymbol{\delta}) - \boldsymbol{f}(\boldsymbol{y})}{\epsilon} \right. \right\rangle \right\} = \langle \boldsymbol{u} | \boldsymbol{\partial f}(\boldsymbol{y}) \rangle \tag{13}$$

Equation (13) shows that, by taking a parameter ϵ sufficiently small, the value $\langle \boldsymbol{u} | \boldsymbol{\partial f}(\boldsymbol{y}) \rangle$ can be approximated by the expected value of the random variable $\frac{1}{\epsilon} \langle \boldsymbol{\delta} \times \boldsymbol{u} | \boldsymbol{f}(\boldsymbol{y} + \epsilon \boldsymbol{\delta}) - \boldsymbol{f}(\boldsymbol{y}) \rangle$. Moreover, as observed in [12], one realization of this random variable is likely to be sufficient to reach a reliable estimate of the expected value in the case of image processing applications (this point will be detailed in Sec. III-E). Therefore, we obtain the following empirical estimation formula for $\langle \boldsymbol{u} | \boldsymbol{\partial f}(\boldsymbol{y}) \rangle$:

$$\langle \boldsymbol{u} | \boldsymbol{\partial f}(\boldsymbol{y}) \rangle = \frac{1}{\epsilon} \big\langle \boldsymbol{\delta} \times \boldsymbol{u} \big| \boldsymbol{f}(\boldsymbol{y} + \epsilon \boldsymbol{\delta}) - \boldsymbol{f}(\boldsymbol{y}) \big\rangle \tag{14}$$

---

[5] Here, $\boldsymbol{u} = \zeta \boldsymbol{y} + \sigma^2 \mathbf{1}$, but the method does not depend on the actual definition of the image $\boldsymbol{u}$. In [12], the method is presented with $\boldsymbol{u} = \mathbf{1}$, which corresponds to $\langle \boldsymbol{u} | \boldsymbol{\partial f}(\boldsymbol{y}) \rangle = \operatorname{Div} \boldsymbol{f}(\boldsymbol{y})$.

[6] In this section, we temporarily assume that $\boldsymbol{y}$ is deterministic. However, to be fully rigorous, what is considered here is not the expectation, but rather the conditional expectation given $\boldsymbol{y}$. To avoid confusion, the latter is denoted with an additional subscript ($\mathbb{E}_{\boldsymbol{\delta}}$), indicating the remaining source of randomness.

C. Evaluation of the Second-Order Derivative Term

A similar method can be proposed to evaluate the term involving the second-order partial derivatives of $\boldsymbol{f}$ in (9), namely $\left\langle \boldsymbol{v} \middle| \boldsymbol{\partial}^2 \boldsymbol{f}(\boldsymbol{y}) \right\rangle$ with[7] $\boldsymbol{v} = \mathbf{1}$.

We use here notations similar to those introduced in Sec. III-B. Then, assuming that $\boldsymbol{f}$ is continuously twice differentiable, a second-order Taylor expansion can be written as:

$$\begin{aligned}\boldsymbol{f}(\boldsymbol{y} + \epsilon\boldsymbol{\delta}) = \boldsymbol{f}(\boldsymbol{y}) &+ \epsilon \sum_l \delta_l \frac{\partial \boldsymbol{f}}{\partial y_l}(\boldsymbol{y}) \\ &+ \frac{\epsilon^2}{2} \sum_{l,m} \delta_l \delta_m \frac{\partial^2 \boldsymbol{f}}{\partial y_l \partial y_m}(\boldsymbol{y}) + \epsilon^2 \boldsymbol{r}(\epsilon) \end{aligned} \quad (15)$$

and similarly for $\boldsymbol{f}(\boldsymbol{y} - \epsilon\boldsymbol{\delta})$. By summing these two expansions, we obtain:

$$\begin{aligned}\lim_{\epsilon \to 0} \left\langle \boldsymbol{\delta} \times \boldsymbol{v} \middle| \frac{\boldsymbol{f}(\boldsymbol{y} + \epsilon\boldsymbol{\delta}) - 2\boldsymbol{f}(\boldsymbol{y}) + \boldsymbol{f}(\boldsymbol{y} - \epsilon\boldsymbol{\delta})}{\epsilon^2} \right\rangle \\ = \sum_{k,l,m} v_k \delta_k \delta_l \delta_m \frac{\partial^2 f_k}{\partial y_l \partial y_m}(\boldsymbol{y}) \end{aligned} \quad (16)$$

In addition to the hypothesis made in Sec. III-B for $\boldsymbol{\delta}$, we impose here the additional requirement that the third moment κ of the random variables δ_k is non-zero. Then, the independence of the δ_k and their zero mean ensure that $\mathbb{E}_{\boldsymbol{\delta}}\{\delta_k \delta_l \delta_m\}$ is always zero except when $k = l = m$, while $\mathbb{E}_{\boldsymbol{\delta}}\left\{\delta_k^3\right\} = \kappa \neq 0$. Therefore, taking the expected value in (16) and switching it with the limit in the left-hand side leads to the following result:

$$\begin{aligned}\lim_{\epsilon \to 0} \mathbb{E}_{\boldsymbol{\delta}} \left\{ \left\langle \boldsymbol{\delta} \times \boldsymbol{v} \middle| \frac{\boldsymbol{f}(\boldsymbol{y} + \epsilon\boldsymbol{\delta}) - 2\boldsymbol{f}(\boldsymbol{y}) + \boldsymbol{f}(\boldsymbol{y} - \epsilon\boldsymbol{\delta})}{\epsilon^2} \right\rangle \right\} \\ = \kappa \left\langle \boldsymbol{v} \middle| \boldsymbol{\partial}^2 \boldsymbol{f}(\boldsymbol{y}) \right\rangle \end{aligned} \quad (17)$$

Finally, assuming that one realization of the random variable $\boldsymbol{\delta}$ is sufficient to estimate the expected value in (17) (see Sec. III-E), we obtain the following empirical estimation formula for $\left\langle \boldsymbol{v} \middle| \boldsymbol{\partial}^2 \boldsymbol{f}(\boldsymbol{y}) \right\rangle$:

$$\begin{aligned}\left\langle \boldsymbol{v} \middle| \boldsymbol{\partial}^2 \boldsymbol{f}(\boldsymbol{y}) \right\rangle = \frac{1}{\epsilon^2 \kappa} \langle \boldsymbol{\delta} \times \boldsymbol{v} | \boldsymbol{f}(\boldsymbol{y} + \epsilon\boldsymbol{\delta}) - 2\boldsymbol{f}(\boldsymbol{y}) \\ + \boldsymbol{f}(\boldsymbol{y} - \epsilon\boldsymbol{\delta}) \rangle \end{aligned} \quad (18)$$

D. Empirical PG-URE Estimator

Using the results obtained in Sec. III-B and III-C, we are now able to re-write the PG-URE estimator (9) without partial derivatives of $\boldsymbol{f}$:

$$\begin{aligned}\text{PG-URE} = &\frac{1}{N} \|\boldsymbol{f}(\boldsymbol{y}) - \boldsymbol{y}\|_2^2 - \frac{\zeta}{N} \langle \mathbf{1} | \boldsymbol{y} \rangle - \sigma^2 \\ &+ \frac{2}{N\dot{\epsilon}} \left\langle \dot{\boldsymbol{\delta}} \times \left(\zeta \boldsymbol{y} + \sigma^2 \mathbf{1} \right) \middle| \boldsymbol{f}\left(\boldsymbol{y} + \dot{\epsilon}\dot{\boldsymbol{\delta}} \right) - \boldsymbol{f}(\boldsymbol{y}) \right\rangle \\ &- \frac{2\sigma^2 \zeta}{N\ddot{\epsilon}^2 \kappa} \left\langle \ddot{\boldsymbol{\delta}} \middle| \boldsymbol{f}\left(\boldsymbol{y} + \ddot{\epsilon}\ddot{\boldsymbol{\delta}} \right) - 2\boldsymbol{f}(\boldsymbol{y}) + \boldsymbol{f}\left(\boldsymbol{y} - \ddot{\epsilon}\ddot{\boldsymbol{\delta}} \right) \right\rangle \end{aligned} \quad (19)$$

This expression uses four parameters that are not related to the noise model, but that are introduced for computational purposes:

- $\dot{\boldsymbol{\delta}}$ is the random perturbation vector used to evaluate the term involving the first-order partial derivatives of $\boldsymbol{f}$ in (9). To fulfil the assumptions made in Sec. III-B, its components $\dot{\delta}_k$ must be independent and identically distributed (i.i.d.) random variables with expected value 0 and variance 1. Several probability distributions with these properties can be used to generate the $\dot{\delta}_k$, and we demonstrate that a binary distribution taking values -1 and 1 with probability $\frac{1}{2}$ each is optimal, in the sense that it minimizes the variance of the PG-URE estimator with respect to the random variable $\dot{\boldsymbol{\delta}}$ (see Sec. III-E).
- $\ddot{\boldsymbol{\delta}}$ is the random perturbation vector used to evaluate the second-order derivative term. $\ddot{\boldsymbol{\delta}}$ is a random vector of i.i.d. components such that[8] $\mathbb{E}\left\{\ddot{\delta}_k\right\} = 0$, $\mathbb{E}\left\{\ddot{\delta}_k^2\right\} = 1$ and $\mathbb{E}\left\{\ddot{\delta}_k^3\right\} = \kappa \neq 0$. Again, an optimum with respect to the variance of PG-URE (see Sec. III-E for details) is reached if the $\ddot{\boldsymbol{\delta}}$ are generated according to a binary distribution π, defined as:

  $$\begin{aligned} &\pi\left(\ddot{\delta}_k = -\sqrt{\frac{q}{p}} \right) = p \qquad \pi\left(\ddot{\delta}_k = \sqrt{\frac{p}{q}} \right) = q \\ &\text{with } p = \frac{1}{2} + \frac{\kappa}{2}\left(\kappa^2 + 4 \right)^{-\frac{1}{2}} \text{ and } q = 1 - p \end{aligned} \quad (20)$$

  where κ is the third moment of the distribution π. The optimal value of κ may not be available in practical settings, and we set it to 1 in our experiments (we motivate this choice in appendix B).
- $\dot{\epsilon}$ and $\ddot{\epsilon}$ are the amplitudes of the perturbations introduced to probe the partial derivatives of $\boldsymbol{f}$. The values of these scalar parameters result from a compromise between 1) the fact that $\dot{\epsilon}$ and $\ddot{\epsilon}$ must be chosen as small as possible to limit the approximation errors in the initial Taylor expansions (10) and (15), and 2) the finite precision of floating point calculators, which causes significant rounding errors when these parameters are too small. How these values should actually be set is discussed in Sec. IV.

Finally, the computational complexity of evaluating the PG-URE estimator through the empirical formula (19) is $4\mathcal{C}_f + \mathcal{O}(N)$, where $\mathcal{C}_f$ is the computational complexity of applying the denoising algorithm $\boldsymbol{f}$.

E. Variance of the Empirical PG-URE Estimator With Respect to the Random Perturbations

In the expression (19) of the PG-URE estimator, the equality is mathematically proved in terms of expected value over the probability distribution of the two random vectors $\dot{\boldsymbol{\delta}}$ and $\ddot{\boldsymbol{\delta}}$. In practice and similarly to what is proposed in [12], we evaluate the right-hand side of this expression with a single realization of each of these random variables, as we can assume that such evaluation is close to the expected

---

[7] Again, the method does not take advantage of the fact that $\boldsymbol{v} = \mathbf{1}$, which motivates the use of a generic notation $\boldsymbol{v}$.

[8] The constraint on the second moment of $\ddot{\delta}_k$ is not compulsory with respect to the methodology developed in Sec. III-C, but is rather a normalization convention.

value. Formally, the underlying assumption is that the standard deviation $\mathbb{V}\mathrm{ar}_{\dot{\boldsymbol{\delta}},\ddot{\boldsymbol{\delta}}}\{\text{PG-URE}\}^{\frac{1}{2}}$ of the estimator (19) over the probability distribution of $\dot{\boldsymbol{\delta}}$ and $\ddot{\boldsymbol{\delta}}$ is small with respect to its expected value.

Thanks to the independence of $\dot{\boldsymbol{\delta}}$ and $\ddot{\boldsymbol{\delta}}$, the variance of PG-URE can be decomposed as:

$$\mathbb{V}\mathrm{ar}_{\dot{\boldsymbol{\delta}},\ddot{\boldsymbol{\delta}}}\{\text{PG-URE}\} = \frac{1}{N^2}\mathbb{V}\mathrm{ar}_{\dot{\boldsymbol{\delta}}}\left\{\sum_{k,l} a_{k,l}\dot{\delta}_k\dot{\delta}_l\right\} + \frac{1}{N^2\kappa^2}\mathbb{V}\mathrm{ar}_{\ddot{\boldsymbol{\delta}}}\left\{\sum_{k,l,m} b_{k,l,m}\ddot{\delta}_k\ddot{\delta}_l\ddot{\delta}_m\right\} \quad (21)$$

where the notations $a_{k,l}$ and $b_{k,l,m}$ stand for:

$$a_{k,l}=2\left(\zeta y_k+\sigma^2\right)\frac{\partial f_k}{\partial y_l}(\mathbf{y}) \quad b_{k,l,m}=2\sigma^2\zeta\frac{\partial^2 f_k}{\partial y_l \partial y_m}(\mathbf{y}) \quad (22)$$

Let us focus on the term $V_{\dot{\boldsymbol{\delta}}} = \frac{1}{N^2}\mathbb{V}\mathrm{ar}_{\dot{\boldsymbol{\delta}}}\{\cdots\}$ in (21), which corresponds to the contribution of the perturbation $\dot{\boldsymbol{\delta}}$ to the overall variance of the estimator. In what follows, the p^{th} moment of the probability distribution associated to $\dot{\boldsymbol{\delta}}$ will be denoted as $\dot{\mathfrak{m}}_p = \mathbb{E}\left\{\dot{\delta}_k^p\right\}$. Then, using the independence of the $\dot{\delta}_k$ and the properties $\dot{\mathfrak{m}}_1 = 0$ and $\dot{\mathfrak{m}}_2 = 1$ introduced in Sec. III-B, $V_{\dot{\boldsymbol{\delta}}}$ can be rewritten as:

$$V_{\dot{\boldsymbol{\delta}}} = \frac{\dot{\mathfrak{m}}_4 - 1}{N^2}\sum_k a_{k,k}^2 + \frac{1}{2N^2}\sum_{k \neq l}\left(a_{k,l} + a_{l,k}\right)^2 \quad (23)$$

This expression (23) calls for two remarks:

1) As $V_{\dot{\boldsymbol{\delta}}}$ should be made as small as possible to limit the variance of the PG-URE estimator, the probability distribution used to generate the $\dot{\delta}_k$ should be chosen so that $\dot{\mathfrak{m}}_4$ is as small as possible. Yet, with the requirements $\dot{\mathfrak{m}}_1 = 0$ and $\dot{\mathfrak{m}}_2 = 1$, it can be shown that $\dot{\mathfrak{m}}_4 \geq 1$ (see for instance [27]); the optimal value $\dot{\mathfrak{m}}_4 = 1$ is obtained with a symmetric binary distribution taking values -1 and 1 with probability $\frac{1}{2}$ each. This justifies our proposition to use this probability distribution in Sec. III-D.
2) The second summation group (the one with two summation indices k and l) involves $N(N-1)$ terms (all the pairs $k, l = 1$ to N, except those with $k = l$), but most of the $\left(a_{k,l} + a_{l,k}\right)^2$ terms are likely to be 0. Indeed, $a_{k,l}$ is proportional to $\frac{\partial f_k}{\partial y_l}(\mathbf{y})$, and the value of this partial derivative is likely to be insignificant when the indices k and l refer to pixels that are distant from each others: in particular, this is certainly true if $\boldsymbol{f}$ is a local denoising method. Furthermore, if we assume that the number of input pixels y_l that have a significant influence on the k^{th} output pixel $f_k(\mathbf{y})$ is constant – or equivalently that the number of indices k such that $a_{k,l} \neq 0$ is bounded independently of N, then the number of non-zero $\left(a_{k,l} + a_{l,k}\right)^2$ terms is proportional to N, making $V_{\dot{\boldsymbol{\delta}}}$ proportional to $\frac{1}{N}$. As N is quite large in the case of images, $V_{\dot{\boldsymbol{\delta}}}$, which represents the variance of the PG-URE estimator with respect to the perturbation $\dot{\boldsymbol{\delta}}$, is likely to be very small: this justifies the assumption made in Sec. III-B that only one realization of this perturbation is sufficient to estimate the first-order partial derivatives of $\boldsymbol{f}$ involved in the computation of PG-URE.


Shepp-Logan Disks Lena MCTS

Fig. 3. Test images used for the simulations (intensity range normalized to the interval [0, 1]).

The term corresponding to the contribution of the perturbation $\ddot{\boldsymbol{\delta}}$ in (21), namely $V_{\ddot{\boldsymbol{\delta}}} = \frac{1}{N^2\kappa^2}\mathbb{V}\mathrm{ar}_{\ddot{\boldsymbol{\delta}}}\{\cdots\}$, can also be expressed as a function of the coefficients $b_{k,l,m}$ and the moments $\ddot{\mathfrak{m}}_p = \mathbb{E}\left\{\ddot{\delta}_k^p\right\}$, similarly to (23) (see appendix B). The obtained expression leads to conclusions similar to those drawn for $V_{\dot{\boldsymbol{\delta}}}$, namely that $V_{\ddot{\boldsymbol{\delta}}}$ is proportional to $\frac{1}{N}$ for reasonable denoising functions $\boldsymbol{f}$, and that $V_{\ddot{\boldsymbol{\delta}}}$ is minimal when $\ddot{\boldsymbol{\delta}}$ is generated according to the binary probability distribution (20), for a particular value $\kappa^\star$ of the parameter κ. Unfortunately, the optimal value $\kappa^\star$ depends on the coefficients $b_{k,l,m}$ and consequently on the partial derivatives of $\boldsymbol{f}$, whose values are by definition not available. Still, we noticed that the arbitrary setting $\kappa = 1$ leads to stable results (see Sec. IV and appendix B).

IV. NUMERICAL VALIDATION AND APPLICATION

A. Simulation Goals and Process

The expression (19) defines an unbiased risk estimator of the MSE under the mixed Poisson-Gaussian noise hypothesis (7). Sections III-D and III-E describe how the random perturbation $\dot{\boldsymbol{\delta}}$ and $\ddot{\boldsymbol{\delta}}$ involved in this PG-URE estimator are generated. However, we have not discussed yet on the values that should be attributed to the scalar parameters $\dot{\epsilon}$ and $\ddot{\epsilon}$. We propose to determine how these values should be set through numerical simulations; we will also make the most of these simulations to verify the empirical equality PG-URE = MSE.

For the numerical simulations, we selected four test images (see Fig. 3):

1) the well-known *Shepp-Logan* phantom, sized 256×256 pixels;
2) a synthetic test image (referred as *Disks*), also sized 256×256 pixels, representing several disks with random gray levels, sizes and boundary sharpness, over a non-uniform dark background,
3) the gray-scale version of the standard *Lena* test image, sized 512×512 pixels,
4) a microscopy image of a multicellular tumor spheroid culture (referred as *MCTS*), sized 512×512 pixels.

All these images were normalized so that they are valued between 0 and 1, and assumed to be noise-free: in particular, the *MCTS* microscopy image was acquired such that it presents a low level of "natural" noise compared to what we added numerically afterward. Indeed, from each of these four noise-free images $\boldsymbol{x}$, we generated four noisy images $\boldsymbol{y}$ following the MPG model (7), with the following noise parameters:

- $\sigma = 10^{-1.5}$, $\zeta = 10^{-2}$ (this case is denoted as "low noise" in the following results);
- $\sigma = 10^{-1}$, $\zeta = 10^{-2}$ (denoted as "mostly Gaussian");
- $\sigma = 10^{-1.5}$, $\zeta = 10^{-1}$ (denoted as "mostly Poisson");
- $\sigma = 10^{-1}$, $\zeta = 10^{-1}$ (denoted as "high noise").

We selected six classical denoising algorithms,[9] all dependent of a scalar parameter θ:

- Wavelet soft-thresholding [1]:

$$\boldsymbol{f}_\theta^{\text{WSo}}(\boldsymbol{y}) = W^{-1} \cdot \boldsymbol{T}_\theta (W \cdot \boldsymbol{y}) \tag{24}$$

  where W is a 2D un-decimated wavelet transform (we used the Daubechies-4 orthogonal wavelet with 4 levels of decomposition), and $\boldsymbol{T}_\theta$ is a component-wise soft-thresholding function, mapping each input wavelet coefficient w to $\text{sign}(w) \max(0, |w| - \theta)$.
- TV minimization [2]:

$$\boldsymbol{f}_\theta^{\text{TV}}(\boldsymbol{y}) = \arg\min_{\boldsymbol{x}} \|\boldsymbol{x}\|_{\text{TV}} \;\text{ s.t. }\; \|\boldsymbol{y} - \boldsymbol{x}\|_2 \leq \theta \tag{25}$$

  where $\|\cdot\|_{\text{TV}}$ is the usual discrete 2D total variation semi-norm.
- Non-local means [3]: $\boldsymbol{f}_\theta^{\text{NLM}}(\boldsymbol{y})$ is defined component-wise as:

$$f_{k,\theta}^{\text{NLM}}(\boldsymbol{y}) = \frac{1}{Z_k(\boldsymbol{y}, \theta)} \sum_l K_{k,l}(\boldsymbol{y}, \theta)\, y_l \tag{26}$$

  where the positive coefficient $K_{k,l}(\boldsymbol{y}, \theta)$ measures the similarity between the neighborhoods of pixels k and l in the image $\boldsymbol{y}$, and $Z_k(\boldsymbol{y}, \theta) = \sum_l K_{k,l}(\boldsymbol{y}, \theta)$ is a normalization factor. We used the similarity measure originally proposed in [3]:

$$K_{k,l}(\boldsymbol{y}, \theta) = \exp\left(-\frac{\|\Pi_k(\boldsymbol{y}) - \Pi_l(\boldsymbol{y})\|_2^2}{\theta^2} \right) \tag{27}$$

  where $\Pi_k(\boldsymbol{y})$ represents a restriction of the observation $\boldsymbol{y}$ to a square patch centered on the pixel k; the patch width was set to 5 pixels.
- We derived three "stabilized" versions of these three denoising algorithms, for which we first applied a variance stabilization transform on the input image, to make the variance of the noisy pixel y_k independent of the ground truth value x_k, and therefore uniform over the whole image (see [13] and [14]). Formally:

$$\boldsymbol{f}_\theta^{\text{S-WSo}}(\boldsymbol{y}) = \boldsymbol{S}^{-1} \circ \boldsymbol{f}_\theta^{\text{WSo}} \circ \boldsymbol{S}(\boldsymbol{y}) \tag{28}$$

  and similarly for $\boldsymbol{f}_\theta^{\text{S-TV}}$ and $\boldsymbol{f}_\theta^{\text{S-NLM}}$. The variance stabilization transform $\boldsymbol{S}(\boldsymbol{y})$ is defined as:

$$S_k(\boldsymbol{y}) = \frac{2}{\zeta} \,\text{sign}(t) \sqrt{|t|} \quad \text{with} \quad t = \zeta y_k + \frac{3}{8}\zeta^2 + \sigma^2 \tag{29}$$

  In [13], it is shown that, under the MPG hypothesis (7), $S_k(\boldsymbol{y})$ has a variance approximately equal to 1 (except for very low values of x_k, which correspond to an extremely low-light regime).

Finally, for each pair of tested noisy image and algorithm, and for several values of the corresponding parameter θ, we computed the denoised estimate $\boldsymbol{f}_\theta(\boldsymbol{y})$ and the MSE (as we are using phantom test images, the ground truth is available), and we evaluated the PG-URE estimator (19) with different values of the amplitude parameters $\dot{\epsilon}$ and $\ddot{\epsilon}$. All simulations were performed with Matlab®, using double precision floating point arithmetic. The influence of $\dot{\epsilon}$ and $\ddot{\epsilon}$ on the PG-URE estimator is studied in the next sections.

B. Influence of the Amplitude Parameters $\dot{\epsilon}$ and $\ddot{\epsilon}$

To study how the parameters $\dot{\epsilon}$ and $\ddot{\epsilon}$ affect the PG-URE estimator (19), we decompose the latter into three terms, as $\text{PG-URE} = T_0 + T_1(\dot{\epsilon}) + T_2(\ddot{\epsilon})$, where:

$$\begin{aligned} T_0 &= \frac{1}{N} \|\boldsymbol{f}(\boldsymbol{y}) - \boldsymbol{y}\|_2^2 - \frac{\zeta}{N} \langle \mathbf{1} | \boldsymbol{y} \rangle - \sigma^2 \\ T_1(\dot{\epsilon}) &= \frac{2}{N\dot{\epsilon}} \left\langle \dot{\boldsymbol{\delta}} \times \left(\zeta \boldsymbol{y} + \sigma^2 \mathbf{1} \right) \middle| \boldsymbol{f}\left(\boldsymbol{y} + \dot{\epsilon} \dot{\boldsymbol{\delta}} \right) - \boldsymbol{f}(\boldsymbol{y}) \right\rangle \\ T_2(\ddot{\epsilon}) &= -\frac{2\sigma^2 \zeta}{N \ddot{\epsilon}^2 \kappa} \left\langle \ddot{\boldsymbol{\delta}} \middle| \boldsymbol{f}\left(\boldsymbol{y} + \ddot{\epsilon} \ddot{\boldsymbol{\delta}} \right) - 2\boldsymbol{f}(\boldsymbol{y}) + \boldsymbol{f}\left(\boldsymbol{y} - \ddot{\epsilon} \ddot{\boldsymbol{\delta}} \right) \right\rangle \end{aligned} \tag{30}$$

In this decomposition, T_0 includes the contributions to PG-URE that do not depend on $\dot{\epsilon}$ and $\ddot{\epsilon}$, while $T_1(\dot{\epsilon})$ and $T_2(\ddot{\epsilon})$ represent respectively the contributions due to the first and second order partial derivatives of $\boldsymbol{f}$. Figs. 4 and 5 present two examples of the evolution of T_0, $T_1(\dot{\epsilon})$ and $T_2(\ddot{\epsilon})$ as functions of the denoising parameter θ, for different values of $\dot{\epsilon}$ and $\ddot{\epsilon}$.

1) Parameter $\dot{\epsilon}$: Both graphs presented in Figs. 4 and 5 show that, although $T_1(\dot{\epsilon} = 0.1)$ and $T_1(\dot{\epsilon} = 1)$ have singular behaviors (the latter curve does not fall in displayed range of the graph in Fig. 5), $T_1(\dot{\epsilon})$ seems to converge to an asymptotic curve for smaller values of $\dot{\epsilon}$: indeed, for $\dot{\epsilon} \leq 10^{-3}$, we can assume that $T_1(\dot{\epsilon})$ becomes almost independent of $\dot{\epsilon}$, with a value close to the ideal one that would be obtained for $\dot{\epsilon} \to 0$.

To confirm this assumption, we measured the term $T_1(\dot{\epsilon})$ for $\dot{\epsilon} = 10^{-7}, 10^{-6.98}, \ldots, 10^{-0.04}, 10^{-0.02}, 1$, and for all the combinations of denoising algorithms, test images and noise parameter mentioned in Sec. IV-A, with the denoising parameter θ set such that the MSE is minimal; the corresponding minimal value of the MSE is denoted as $\text{MSE}^\star$. We then measured the variability among the $T_1(\dot{\epsilon})$ values through the indicator ΔT_1, defined as:

$$\Delta T_1 = \frac{1}{\text{MSE}^\star} \operatorname*{StdDev}_{\dot{\epsilon}} T_1(\dot{\epsilon}) \tag{31}$$

where $\operatorname*{StdDev}_{\dot{\epsilon}} T_1(\dot{\epsilon})$ measures the empirical standard deviation of $T_1(\dot{\epsilon})$ for $\dot{\epsilon}$ varying within a sub-range $[\dot{\epsilon}_{\min}, \dot{\epsilon}_{\max}]$ of the probed interval $\left[10^{-7}, 1\right]$. The values obtained for ΔT_1 with $[\dot{\epsilon}_{\min}, \dot{\epsilon}_{\max}] = \left[10^{-6}, 10^{-3}\right]$ are presented in Fig. 6. Theses results show that the variability of $T_1(\dot{\epsilon})$ induced by the choice of $\dot{\epsilon}$ is very small compared to the MSE (the quantity to

---

[9] We intentionally do not select the state-of-the-art denoising algorithms such as BM3D [5], as the goal here is not to compare the performances of the existing denoising methods, but rather verify that PG-URE can be used as an empirical estimator of the MSE with various families of denoising algorithms. We therefore focused on the simplest and most classical methods.

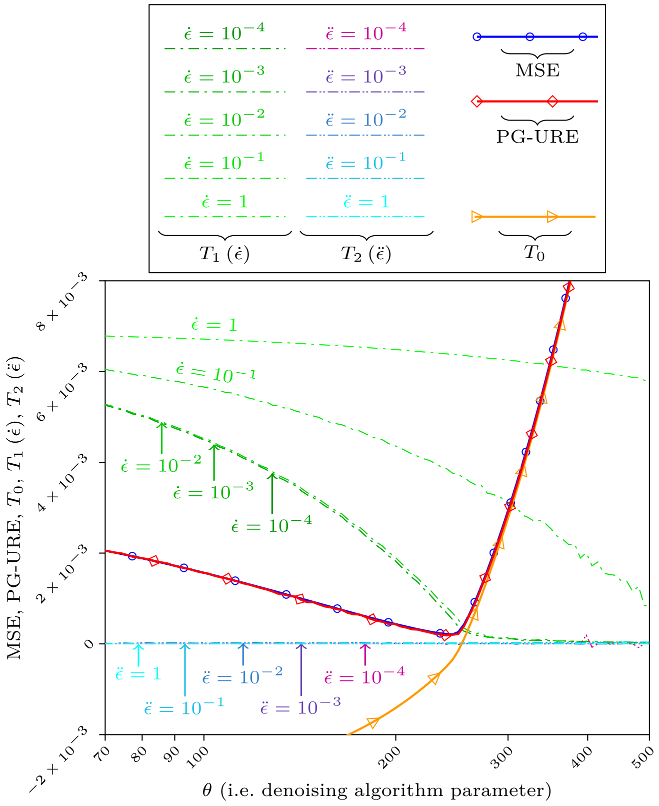

Fig. 4. Denoising of *Disks* + "low noise", using the $\boldsymbol{f}_\theta^{\text{S-TV}}$ algorithm (TV minimization together with variance stabilization transform). MSE and PG-URE values are plotted as functions of the denoising parameter θ, together with the individual PG-URE terms T_0, $T_1(\dot{\epsilon})$ and $T_2(\ddot{\epsilon})$ for several values of the parameters $\dot{\epsilon}$ and $\ddot{\epsilon}$. Only the PG-URE curve corresponding to $\dot{\epsilon} = 10^{-4}$ and $\ddot{\epsilon} = 10^{-2}$ is plotted.

estimate): indeed, whatever the value chosen for $\dot{\epsilon}$ in the range $\left[10^{-6}, 10^{-3}\right]$, the value obtained for $T_1(\dot{\epsilon})$ (and therefore for PG-URE) is constant. We therefore used in practice $\dot{\epsilon} = 10^{-4}$ in what follows.

It is important to note that this value depends on the normalization used for the intensity of the processed images: here, our images are valued between 0 and 1, but different normalizations would lead to different compromises. For instance, in case of intensity normalized between 0 and 255, a correct setting is $\dot{\epsilon} = 255 \times 10^{-4}$. The floating point precision used for the computations may also have an influence, although this aspect is less critical for $T_1(\dot{\epsilon})$ than for the second order term, as discussed in the next paragraph.

2) Parameter $\ddot{\epsilon}$: We proceeded similarly to determine a satisfactory value for $\ddot{\epsilon}$: we measured the term $T_2(\ddot{\epsilon})$ for $\ddot{\epsilon} = 10^{-4}, 10^{-3.99}, \dots, 10^{-0.02}, 10^{-0.01}, 1$, and for all the combinations of denoising algorithms, test images and noise parameters, with the denoising parameter θ set such that the MSE is minimal. The values obtained for $T_2(\ddot{\epsilon})$ as functions of $\ddot{\epsilon}$ in six of these configurations are presented in Fig. 7.

Contrary to what happens with the first order term, we did not observe a clear convergence of $T_2(\ddot{\epsilon})$ to an asymptotic value when $\ddot{\epsilon} \to 0$: the curves $T_2(\ddot{\epsilon})$ showed chaotic behaviors, with large and unpredicable oscillations when $\ddot{\epsilon} \leq 10^{-3}$. We interpret these behaviors as the consequence of rounding errors introduced by floating point operations involved when

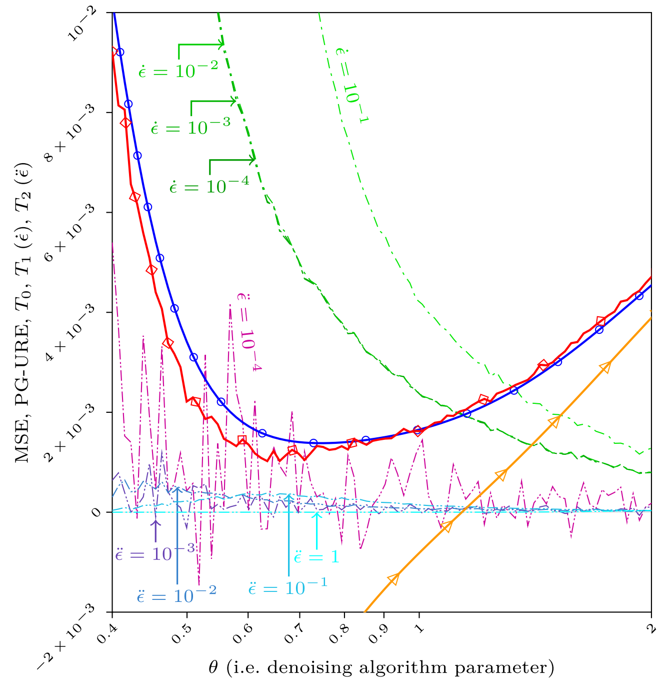

Fig. 5. Denoising of *Disks* + "mostly Poisson" noise, using the $\boldsymbol{f}_\theta^{\text{S-NLM}}$ algorithm (non-local means together with variance stabilization transform). Same representation and legend as in Fig. 4.

computing the term $T_2(\ddot{\epsilon})$. More precisely, the latter involves a second-order finite difference $\boldsymbol{f}\left(\boldsymbol{y} + \ddot{\epsilon}\ddot{\boldsymbol{\delta}}\right) - 2\boldsymbol{f}(\boldsymbol{y}) + \boldsymbol{f}\left(\boldsymbol{y} - \ddot{\epsilon}\ddot{\boldsymbol{\delta}}\right)$ whose order of magnitude might be significantly smaller than the ones of the individual terms $\boldsymbol{f}\left(\boldsymbol{y} \pm \ddot{\epsilon}\ddot{\boldsymbol{\delta}}\right)$ and $\boldsymbol{f}(\boldsymbol{y})$: then, due to cancellation events (see [28]), the error made when performing this operation is likely to be significant. A solution to avoid this problem could have been to increase the parameter $\ddot{\epsilon}$, but in this case the assumption that $T_2(\ddot{\epsilon})$ is close to its theoretical limit obtained for $\ddot{\epsilon} \to 0$ becomes erroneous: it appears that the trade-off between the need for $\ddot{\epsilon}$ to be small enough for the mathematical analysis derived in Sec. III to be valid, and the need for $\ddot{\epsilon}$ to be large enough to avoid numerical rounding errors is much more tight for $\ddot{\epsilon}$ than for $\dot{\epsilon}$.

However, the curves on Fig. 7 show that there seems to exist a narrow window around $\ddot{\epsilon} = 10^{-2}$ where both requirements hold, leading to functions $T_2(\ddot{\epsilon})$ approximately constant. To validate this hypothesis, we introduce an indicator ΔT_2 as follows:

$$\Delta T_2 = \frac{1}{\mathrm{MSE}^{\star}} \operatorname*{StdDev}_{\ddot{\epsilon}} T_2(\ddot{\epsilon}) \tag{32}$$

where the empirical standard deviation is computed for $\ddot{\epsilon}$ varying in a sub-range $[\ddot{\epsilon}_{\min}, \ddot{\epsilon}_{\max}]$ of the probed interval. The values obtained for ΔT_2 with $[\ddot{\epsilon}_{\min}, \ddot{\epsilon}_{\max}] = \left[5 \times 10^{-3}, 2 \times 10^{-2}\right]$ are presented in Fig. 8. These values show that the variability of $T_2(\ddot{\epsilon})$ (and therefore the variability of PG-URE) induced by the choice of $\ddot{\epsilon}$ represents less than 1% of the MSE to be estimated in more than half of the tested situations. This variability seems to be mainly determined by the denoising algorithm: indeed, the value of $T_2(\ddot{\epsilon})$ is very

	Shepp-Logan				*Disks*				*Lena*				*MCTS*			
	Low noise	Mostly Gaussian	Mostly Poisson	High noise	Low noise	Mostly Gaussian	Mostly Poisson	High noise	Low noise	Mostly Gaussian	Mostly Poisson	High noise	Low noise	Mostly Gaussian	Mostly Poisson	High noise
$\boldsymbol{f}_\theta^{\mathrm{WSo}}$	0.06%	0.07%	0.07%	0.06%	0.07%	0.09%	0.11%	0.11%	0.02%	0.05%	0.06%	0.05%	0.03%	0.04%	0.18%	0.07%
$\boldsymbol{f}_\theta^{\mathrm{TV}}$	0.11%	0.04%	0.03%	0.11%	0.97%	0.34%	0.64%	0.53%	0.03%	0.05%	0.02%	0.04%	0.33%	0.02%	0.25%	0.15%
$\boldsymbol{f}_\theta^{\mathrm{NLM}}$	0.01%	0.01%	0.01%	0.00%	0.02%	0.01%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.01%	0.01%	0.00%	0.00%
$\boldsymbol{f}_\theta^{\mathrm{S\text{-}WSo}}$	0.07%	0.10%	0.59%	0.30%	0.23%	0.26%	1.40%	0.52%	0.05%	0.25%	0.65%	0.36%	0.05%	0.17%	2.53%	0.52%
$\boldsymbol{f}_\theta^{\mathrm{S\text{-}TV}}$	0.06%	0.10%	0.83%	0.46%	0.13%	0.09%	0.46%	0.47%	0.08%	0.01%	0.15%	0.12%	0.13%	0.10%	0.97%	0.33%
$\boldsymbol{f}_\theta^{\mathrm{S\text{-}NLM}}$	0.05%	0.01%	0.66%	0.17%	0.02%	0.02%	0.59%	0.46%	0.01%	0.00%	0.19%	0.26%	0.03%	0.01%	0.93%	0.19%

Fig. 6. ΔT_1 obtained for $\dot{\epsilon} = 10^{-6}, 10^{-5.98}, \ldots, 10^{-3.02}, 10^{-3}$ (151 samples), given as percentages. The two values greater than 1% are highlighted in yellow.

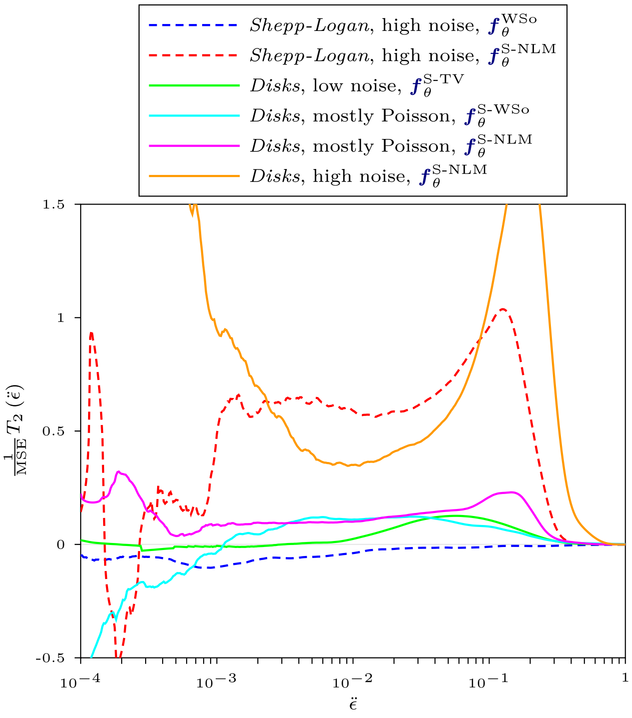

Fig. 7. Term $T_2(\ddot{\epsilon})$ as a function of $\ddot{\epsilon}$ for six of the tested combinations of test image, noise level and denoising algorithm (with in each case the parameter θ set such that the MSE is minimal). Each curve $T_2(\ddot{\epsilon})$ was normalized by the actual MSE measured for the corresponding tested combination.

stable in the case of $\boldsymbol{f}_\theta^{\mathrm{NLM}}$, and on the contrary extremely dependent on $\ddot{\epsilon}$ in the case of $\boldsymbol{f}_\theta^{\mathrm{S\text{-}TV}}$. However, as other choices of intervals $[\ddot{\epsilon}_{\min}, \ddot{\epsilon}_{\max}]$ lead to poorer results for ΔT_2, we propose $\ddot{\epsilon} = 10^{-2}$ as a reasonable compromise value for this parameter. Results presented in the next section show that this choice leads to an estimator PG-URE that can be successfully used to adaptively set the value of the parameter θ for each denoising algorithm.

Similarly to the case of the first order term, the setting for $\ddot{\epsilon}$ depends on the normalization used for the intensity of the processed images, and also on the floating point precision used for the computations.

C. Optimization of the Denoising Parameters θ Driven by the PG-URE Estimator

Finally, to evaluate the performance of the PG-URE estimator when used to optimize the parameter θ of the denoising algorithms, we performed the following simulations: for each combination of tested image, set of noise parameters, and denoising algorithm $\boldsymbol{f}_\theta$, we ran the denoising algorithm for several values of θ, and computed the resulting MSE and PG-URE values;[10] we then retained in each case the parameters $\theta^\star_{\mathrm{MSE}}$ and $\theta^\star_{\text{PG-URE}}$ that minimize respectively the MSE and the PG-URE. The corresponding image $\hat{\boldsymbol{x}}_{\text{PG-URE}} = \boldsymbol{f}_{\theta^\star_{\text{PG-URE}}}(\boldsymbol{y})$ represents the denoising result obtained by tuning the denoising parameter such that the PG-URE estimator is minimal – hence without using the ground truth – while $\hat{\boldsymbol{x}}_{\mathrm{MSE}} = \boldsymbol{f}_{\theta^\star_{\mathrm{MSE}}}(\boldsymbol{y})$ corresponds to the denoised image obtained by selecting the best denoising parameter according to the MSE, following an oracle-based approach (hence not applicable for real denoising problems). We finally compared the differences between the two denoised images by measuring the following indicator:

$$\Delta\mathrm{Estim} = \frac{\left\| \hat{\boldsymbol{x}}_{\text{PG-URE}} - \hat{\boldsymbol{x}}_{\mathrm{MSE}} \right\|_2^2}{\left\| \boldsymbol{x} - \hat{\boldsymbol{x}}_{\mathrm{MSE}} \right\|_2^2} \tag{33}$$

Here, ΔEstim relates the l_2 distance between the two denoised images to the l_2 distance between the ground truth $\boldsymbol{x}$ and the "best" denoised image, i.e. the one obtained by following the oracle based approach. The values measured for ΔEstim are presented in Fig. 9, along with the peak signal-to-noise measure reached with $\hat{\boldsymbol{x}}_{\mathrm{MSE}}$ – defined as $\mathrm{PSNR} = -10\log_{10}\left(\frac{1}{N}\left\|\boldsymbol{x} - \hat{\boldsymbol{x}}_{\mathrm{MSE}}\right\|_2^2\right)$ – which assesses the "best" denoising quality achievable following the oracle-based parameter estimation approach. Four examples of pairs of denoised images $\hat{\boldsymbol{x}}_{\mathrm{MSE}}$ and $\hat{\boldsymbol{x}}_{\text{PG-URE}}$ are also presented in Fig. 10.

Although the best performing denoising parameters $\theta^\star_{\mathrm{MSE}}$ and $\theta^\star_{\text{PG-URE}}$ selected by the MSE and the PG-URE are not always the same, Fig. 9 shows that the distance between

---

[10] We selected $\dot{\epsilon} = 10^{-4}$ and $\ddot{\epsilon} = 10^{-2}$ to evaluate PG-URE, as advised in Sec. IV-B.

	Shepp-Logan				*Disks*				*Lena*				*MCTS*			
	Low noise	Mostly Gaussian	Mostly Poisson	High noise	Low noise	Mostly Gaussian	Mostly Poisson	High noise	Low noise	Mostly Gaussian	Mostly Poisson	High noise	Low noise	Mostly Gaussian	Mostly Poisson	High noise
$\boldsymbol{f}_\theta^{\text{WSo}}$	0.2%	0.6%	0.1%	3.2%	0.6%	1.3%	0.3%	3.2%	0.0%	0.1%	0.0%	0.3%	0.2%	0.6%	0.2%	1.9%
$\boldsymbol{f}_\theta^{\text{TV}}$	0.4%	0.4%	0.0%	0.8%	1.3%	0.9%	0.4%	3.3%	0.0%	0.1%	0.0%	0.1%	0.4%	0.3%	0.2%	2.2%
$\boldsymbol{f}_\theta^{\text{NLM}}$	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
$\boldsymbol{f}_\theta^{\text{S-WSo}}$	0.8%	0.6%	4.2%	4.6%	0.3%	1.5%	0.9%	4.5%	0.1%	0.1%	0.2%	0.4%	2.2%	0.4%	2.6%	2.1%
$\boldsymbol{f}_\theta^{\text{S-TV}}$	3.2%	1.1%	12.2%	12.5%	5.5%	11.1%	8.6%	12.1%	0.1%	0.2%	0.4%	1.0%	5.3%	2.0%	25.5%	7.7%
$\boldsymbol{f}_\theta^{\text{S-NLM}}$	2.6%	0.1%	2.1%	6.1%	0.5%	0.1%	2.3%	4.4%	0.0%	0.0%	0.0%	0.5%	2.8%	0.1%	2.8%	1.5%

Fig. 8. ΔT_2 given as percentages obtained for $5 \times 10^{-3} \leq \ddot{\epsilon} \leq 2 \times 10^{-2}$ with geometric increments of $10^{0.01}$ (61 samples). Yellow cells contain values greater than 1%, while orange cells contain values greater than 10%.

	Shepp-Logan				*Disks*				*Lena*				*MCTS*			
	Low noise	Mostly Gaussian	Mostly Poisson	High noise	Low noise	Mostly Gaussian	Mostly Poisson	High noise	Low noise	Mostly Gaussian	Mostly Poisson	High noise	Low noise	Mostly Gaussian	Mostly Poisson	High noise
$\boldsymbol{f}_\theta^{\text{WSo}}$	0.4% 31 dB	1.4% 27 dB	0.9% 25 dB	0.1% 24 dB	0.3% 33 dB	0.1% 30 dB	3.5% 27 dB	0.0% 27 dB	0.1% 30 dB	0.4% 28 dB	0.8% 25 dB	0.1% 25 dB	0.1% 39 dB	0.6% 35 dB	5.4% 34 dB	2.5% 34 dB
$\boldsymbol{f}_\theta^{\text{TV}}$	1.6% 36 dB	12.3% 32 dB	1.3% 28 dB	2.8% 27 dB	0.0% 38 dB	15.4% 35 dB	0.0% 31 dB	0.0% 30 dB	0.0% 31 dB	0.0% 29 dB	0.0% 26 dB	0.0% 26 dB	3.4% 39 dB	0.0% 36 dB	0.0% 35 dB	0.0% 34 dB
$\boldsymbol{f}_\theta^{\text{NLM}}$	0.0% 35 dB	0.2% 32 dB	0.3% 27 dB	0.3% 26 dB	0.7% 39 dB	0.2% 36 dB	0.1% 29 dB	1.5% 29 dB	0.0% 31 dB	0.0% 29 dB	0.0% 25 dB	0.1% 25 dB	0.0% 38 dB	0.1% 35 dB	0.9% 32 dB	0.9% 32 dB
$\boldsymbol{f}_\theta^{\text{S-WSo}}$	1.2% 30 dB	0.8% 27 dB	2.0% 23 dB	0.5% 22 dB	0.1% 33 dB	0.5% 30 dB	4.9% 26 dB	2.2% 24 dB	0.1% 30 dB	0.1% 28 dB	2.7% 24 dB	0.1% 24 dB	0.1% 38 dB	0.5% 35 dB	23.3% 31 dB	2.5% 28 dB
$\boldsymbol{f}_\theta^{\text{S-TV}}$	0.0% 35 dB	2.8% 31 dB	0.7% 27 dB	6.5% 24 dB	4.1% 37 dB	11.3% 34 dB	15.5% 28 dB	56.1% 27 dB	0.0% 31 dB	0.0% 29 dB	2.1% 25 dB	2.7% 25 dB	0.0% 39 dB	8.8% 36 dB	25.6% 32 dB	1.7% 28 dB
$\boldsymbol{f}_\theta^{\text{S-NLM}}$	0.4% 36 dB	1.4% 32 dB	10.7% 27 dB	0.9% 25 dB	0.5% 39 dB	0.5% 36 dB	8.4% 29 dB	6.3% 28 dB	0.5% 32 dB	0.0% 29 dB	2.7% 25 dB	1.8% 25 dB	0.1% 38 dB	0.7% 35 dB	1.6% 31 dB	1.7% 29 dB

Fig. 9. ΔEstim (33) given as percentages. Yellow cells contain values greater than than 5%, while orange cells contain values greater that 20%. PSNR values (in dB) obtained for $\hat{\boldsymbol{x}}_{\text{MSE}}$ are reported, as a measure of the "best" denoising quality achievable using an oracle-based parametrization.

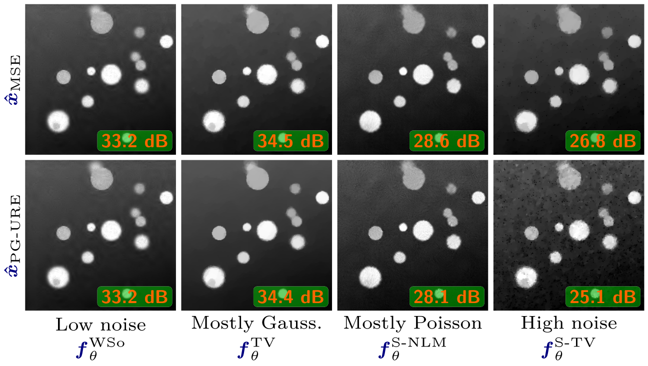

Fig. 10. Comparison between the denoised images $\hat{\boldsymbol{x}}_{\text{MSE}}$ and $\hat{\boldsymbol{x}}_{\text{PG-URE}}$ obtained for the original image *Disks*, with four different noise levels and denoising methods. PSNR values are also reported in the bottom right corner of each image.

the corresponding denoised images is, in most cases, very small compared to the distance between the oracle-denoised image and the ground truth: the indicator ΔEstim is indeed smaller than 5% in 83 of the 96 tested configurations, which corresponds to differences between the denoised images that are visually unnoticeable. The visual similarity between the denoised images $\hat{\boldsymbol{x}}_{\text{MSE}}$ and $\hat{\boldsymbol{x}}_{\text{PG-URE}}$ obtained with these parameters is illustrated on four examples in Fig. 10: in each of the three left-most columns – which correspond to situations with $\Delta\text{Estim} \leq 20\%$ (either white or yellow cells in Fig. 9)– the images $\hat{\boldsymbol{x}}_{\text{MSE}}$ and $\hat{\boldsymbol{x}}_{\text{PG-URE}}$ are indeed very similar. For all these cases, the PG-URE estimator therefore performed very well as a surrogate for the MSE value, while still being actually computable in real denoising problems, for which a ground truth is not available.

However, for the *Disks* image in the "high-noise" configuration processed by the $\boldsymbol{f}_\theta^{\text{S-TV}}$ algorithm, and for the *MCTS* image in the "Poisson noise" configuration processed by either $\boldsymbol{f}_\theta^{\text{S-WSo}}$ or $\boldsymbol{f}_\theta^{\text{S-TV}}$ (orange cells in Fig. 9 and right-most column in Fig. 10), we can clearly observe that the denoising task failed and did not return a reliable image, due to an inappropriate selection of the parameter θ value, itself derived from an erroneous estimation of the MSE with the empirical PG-URE estimate. Two scenarios can explain these erroneous estimations: drawing of a "bad" sample of the parameter $\ddot{\epsilon}$ (Fig. 8 shows that two of these configurations correspond to the least favorable cases with respect to the indicator ΔT_2), and/or a realization of one of the random variables $\dot{\boldsymbol{\delta}}$ or $\ddot{\boldsymbol{\delta}}$ that makes the PG-URE estimator significantly deviate from its expected value. These scenarios correspond to the inherent risk taken with any stochastic Monte-Carlo type of method. One way to reduce this risk would be to draw several realizations of $\dot{\boldsymbol{\delta}}$ or $\ddot{\boldsymbol{\delta}}$ and average the corresponding values of $T_1(\dot{\epsilon})$ and $T_2(\ddot{\epsilon})$, at the cost however of a higher computation time. Post-processing could also be proposed to detect failure of the denoising, or multiple runs could be performed to gauge the range of values obtained for the parameter being optimized, with detection of outliers.

V. CONCLUSION

In this paper, we presented a new unbiased risk estimator (PG-URE) for general image denoising applications, in a context where the processed images are degraded following a mixed Poisson-Gaussian noise model. This model unifies the widely used Gaussian and Poisson noise models and is relevant to describe the degradations observed in bioimaging applications, in particular low-light fluorescence microscopy. We showed that the PG-URE estimator can be used as a surrogate for the usual mean squared error measure, although its evaluation does not require any knowledge about the noise-free version (i.e. the ground truth) of the image to denoise. We also proposed a practical formula (19) to evaluate the PG-URE estimator when no specific knowledge on the partial derivatives of the denoising function $\boldsymbol{f}$ is available, making this framework usable "out of the box" with almost any available denoising algorithm.

Finally, we validated our approach through numerical simulations involving standard denoising algorithms and phantom test images. Relying on theses simulations, we discussed how to set the numerical parameters involved in PG-URE. We compared the results obtained when tuning the parameters θ of these standard denoising algorithms by minimizing the PG-URE estimator and the mean squared error, and showed that these two approaches lead to similar denoised images in most of the tested scenarios. This demonstrates the interest of the PG-URE estimator for practical applications, as MSE driven optimization is not applicable for real denoising problems.

APPENDIX A

DERIVATION OF THE PG-URE ESTIMATOR

This appendix describes how the first definition (8) of the PG-URE estimator is obtained, and proves the equality $\mathbb{E}\{\text{PG-URE}\} = \mathbb{E}\{\text{MSE}\}$. This result could be derived quite directly from the work in [15], but we propose here a more intrinsic proof, relying on the two basic properties of the Gaussian and Poisson distributions that are mentioned below. Proofs of these lemmas can be found respectively in [8], [29], and [30].

Lemma 1.1(Stein's lemma): Let $\boldsymbol{y} = \boldsymbol{x} + \boldsymbol{b}$ where $\boldsymbol{x} \in \mathbb{R}^N$ is deterministic and $\boldsymbol{b} \sim \mathcal{N}\left(0, \sigma^2\,\mathrm{Id}\right)$. Let $\boldsymbol{\phi} : \mathbb{R}^N \to \mathbb{R}^N$ be a weakly differentiable function such that $\mathbb{E}\left\{\left|\frac{\partial \phi_k}{\partial y_k}(\boldsymbol{y})\right|\right\} < +\infty$ for all k. Then:

$$\mathbb{E}\{\langle \boldsymbol{b} | \boldsymbol{\phi}(\boldsymbol{y})\rangle\} = \sigma^2 \mathbb{E}\{\mathrm{Div}\,\boldsymbol{\phi}(\boldsymbol{y})\}$$

Lemma 1.2: Let $\boldsymbol{z} \in \mathbb{R}^N$ such that $\boldsymbol{z} \sim \mathcal{P}(\boldsymbol{x})$ (i.e. the components z_k are independent random variables following Poisson distributions of parameters x_k). Let $\boldsymbol{\psi} : \mathbb{R}^N \to \mathbb{R}^N$ such that $\mathbb{E}\{|\psi_k(\boldsymbol{z})|\} < +\infty$ for all k. Then:

$$\mathbb{E}\{\langle \boldsymbol{x} | \boldsymbol{\psi}(\boldsymbol{z})\rangle\} = \mathbb{E}\left\{\left\langle \boldsymbol{z} \middle| \boldsymbol{\psi}^{[-1]}(\boldsymbol{z})\right\rangle\right\}$$

Thanks to these results, we can state the following theorem:

Theorem 1.1: Let $\boldsymbol{y} = \zeta \boldsymbol{z} + \boldsymbol{b}$ where $\boldsymbol{b} \sim \mathcal{N}\left(0, \sigma^2\,\mathrm{Id}\right)$ and $\boldsymbol{z} \sim \mathcal{P}\left(\frac{\boldsymbol{x}}{\zeta}\right)$ ($\boldsymbol{b}$ and $\boldsymbol{z}$ independent). Let $\boldsymbol{\phi} : \mathbb{R}^N \to \mathbb{R}^N$ a weakly differentiable function such that $\mathbb{E}\{|\phi_k(\boldsymbol{y})|\} < +\infty$ and $\mathbb{E}\left\{\left|\frac{\partial \phi_k}{\partial y_k}(\boldsymbol{y} - \zeta \boldsymbol{e}_k)\right|\right\} < +\infty$ for all k. Then:

$$\mathbb{E}\{\langle \boldsymbol{x} | \boldsymbol{\phi}(\boldsymbol{y})\rangle\} = \mathbb{E}\left\{\left\langle \boldsymbol{y} \middle| \boldsymbol{\phi}^{[-\zeta]}(\boldsymbol{y})\right\rangle - \sigma^2\,\mathrm{Div}\,\boldsymbol{\phi}^{[-\zeta]}(\boldsymbol{y})\right\}$$

Proof: We introduce the family of functions $\boldsymbol{\psi}_{\boldsymbol{b}} : \mathbb{R}^N \to \mathbb{R}^N$, defined by $\boldsymbol{\psi}_{\boldsymbol{b}}(\boldsymbol{z}) = \boldsymbol{\phi}(\zeta \boldsymbol{z} + \boldsymbol{b})$. Then:

$$\begin{aligned}
\mathbb{E}\{\langle \boldsymbol{x} | \boldsymbol{\phi}(\boldsymbol{y})\rangle\} &= \ldots \\
&= \mathbb{E}\left\{\zeta \mathbb{E}_{\boldsymbol{z}}\left\{\left\langle \frac{\boldsymbol{x}}{\zeta} \middle| \boldsymbol{\psi}_{\boldsymbol{b}}(\boldsymbol{z})\right\rangle\right\}\right\} \\
&\overset{(a)}{=} \mathbb{E}\left\{\zeta \mathbb{E}_{\boldsymbol{z}}\left\{\left\langle \boldsymbol{z} \middle| \boldsymbol{\psi}_{\boldsymbol{b}}{}^{[-1]}(\boldsymbol{z})\right\rangle\right\}\right\} \\
&= \mathbb{E}\left\{\left\langle \boldsymbol{y} - \boldsymbol{b} \middle| \boldsymbol{\phi}^{[-\zeta]}(\zeta \boldsymbol{z} + \boldsymbol{b})\right\rangle\right\} \\
&= \mathbb{E}\left\{\left\langle \boldsymbol{y} \middle| \boldsymbol{\phi}^{[-\zeta]}(\boldsymbol{y})\right\rangle\right\} - \mathbb{E}\left\{\mathbb{E}_{\boldsymbol{b}}\left\{\left\langle \boldsymbol{b} \middle| \boldsymbol{\phi}^{[-\zeta]}(\boldsymbol{y})\right\rangle\right\}\right\} \\
&\overset{(b)}{=} \mathbb{E}\left\{\left\langle \boldsymbol{y} \middle| \boldsymbol{\phi}^{[-\zeta]}(\boldsymbol{y})\right\rangle\right\} - \mathbb{E}\left\{\sigma^2 \mathbb{E}_{\boldsymbol{b}}\left\{\mathrm{Div}\,\boldsymbol{\phi}^{[-\zeta]}(\boldsymbol{y})\right\}\right\} \\
&= \mathbb{E}\left\{\left\langle \boldsymbol{y} \middle| \boldsymbol{\phi}^{[-\zeta]}(\boldsymbol{y})\right\rangle - \sigma^2\,\mathrm{Div}\,\boldsymbol{\phi}^{[-\zeta]}(\boldsymbol{y})\right\}
\end{aligned}$$

Steps (a) and (b) make use respectively of lemmas 1.2 and 1.1. ■

Finally, from the definition of the MSE (1), it can be noticed that:

$$\mathbb{E}\{\mathrm{MSE}\} = \frac{1}{N}\mathbb{E}\left\{\|\boldsymbol{f}(\boldsymbol{y})\|_2^2 - 2\langle \boldsymbol{x} | \boldsymbol{f}(\boldsymbol{y})\rangle + \langle \boldsymbol{x} | \boldsymbol{y}\rangle\right\} \qquad (34)$$

Theorem 1.3 applied twice on this expression with $\boldsymbol{\phi} = \boldsymbol{f}$ and $\boldsymbol{\phi} = \mathbf{Id}$ (the identity function) leads to the expected expression of PG-URE. As previously mentioned, we assume that the regularity and expectation conditions of theorem 1.3 hold for $\boldsymbol{f}$.

APPENDIX B

CONTRIBUTION OF THE PERTURBATION $\ddot{\boldsymbol{\delta}}$ TO THE VARIANCE OF THE PG-URE ESTIMATOR

In this appendix, we derive an algebraic expression for the contribution $V_{\ddot{\boldsymbol{\delta}}}$ of the perturbation $\ddot{\boldsymbol{\delta}}$ to the variance (21) of the PG-URE estimator. This expression uses only the coefficients $b_{k,l,m}$ defined by (22), and the moments $\ddot{\mathfrak{m}}_p = \mathbb{E}\left\{\ddot{\delta}_k^p\right\}$ of the probability distribution used to generate the components of $\ddot{\boldsymbol{\delta}}$. We finally derive the optimal conditions on these moments $\ddot{\mathfrak{m}}_p$ for $V_{\ddot{\boldsymbol{\delta}}}$ to be minimal.

A. Expression of $V_{\ddot{\delta}}$

First, we introduce a few notations:

- $c_k = b_{k,k,k}$ for all pixel index k,
- $d_{k,l} = b_{k,k,l} + b_{k,l,k} + b_{l,k,k}$, for all $k \neq l$,
- $D_l = \sum_{k,k\neq l} d_{k,l}$ for all l,
- $e_{k,l,m} = b_{k,l,m} + b_{k,m,l} + b_{l,k,m} + b_{m,k,l} + b_{l,m,k} + b_{m,l,k}$ for all 3-tuple (k,l,m) with $k \neq l$, $k \neq m$ and $l \neq m$.

We also recall that $\ddot{\mathfrak{m}}_1 = 0$, $\ddot{\mathfrak{m}}_2 = 1$, $\ddot{\mathfrak{m}}_3 = \kappa$. Then, starting from the definition of $V_{\ddot{\delta}}$, we have:

$$\begin{aligned} N^2\kappa^2 V_{\ddot{\delta}} &= \mathbb{V}\mathrm{ar}_{\ddot{\delta}}\left\{\sum_{k,l,m} b_{k,l,m}\ddot{\delta}_k\ddot{\delta}_l\ddot{\delta}_m\right\} \\ &= \sum_{i,j,k,l,m,n} b_{i,j,k}b_{l,m,n}\mathbb{E}_{\ddot{\delta}}\left\{\ddot{\delta}_i\ddot{\delta}_j\ddot{\delta}_k\ddot{\delta}_l\ddot{\delta}_m\ddot{\delta}_n\right\} \\ &\quad - \left(\sum_{k,l,m} b_{k,l,m}\mathbb{E}_{\ddot{\delta}}\left\{\ddot{\delta}_k\ddot{\delta}_l\ddot{\delta}_m\right\}\right)^2 \end{aligned}$$

As explained in Sec. III-C, $\mathbb{E}_{\ddot{\delta}}\left\{\ddot{\delta}_k\ddot{\delta}_l\ddot{\delta}_m\right\} = 0$ except when $k = l = m$: this is due to the independence of the components of $\ddot{\delta}$ and to the property $\ddot{\mathfrak{m}}_1 = 0$. This leads to the immediate simplification of the expression above:

$$N^2\kappa^2 V_{\ddot{\delta}} = \underbrace{\sum_{\substack{i,j,k\\ l,m,n}} b_{i,j,k}b_{l,m,n}\mathbb{E}_{\ddot{\delta}}\left\{\ddot{\delta}_i\ddot{\delta}_j\ddot{\delta}_k\ddot{\delta}_l\ddot{\delta}_m\ddot{\delta}_n\right\}}_{S_6} - \kappa^2\sum_{k,l} c_k c_l \tag{35}$$

The same arguments can be used to simplify the sixfold sum S_6, as $\mathbb{E}_{\ddot{\delta}}\left\{\ddot{\delta}_i\ddot{\delta}_j\ddot{\delta}_k\ddot{\delta}_l\ddot{\delta}_m\ddot{\delta}_n\right\} = 0$ as soon as one of the six indices is different from the others. Then, S_6 can be divided according to the four situations where $\mathbb{E}_{\ddot{\delta}}\left\{\ddot{\delta}_i\ddot{\delta}_j\ddot{\delta}_k\ddot{\delta}_l\ddot{\delta}_m\ddot{\delta}_n\right\}$ is non-zero:

$$S_6 = \ddot{\mathfrak{m}}_6 T_6 + \ddot{\mathfrak{m}}_4 T_{4,2} + \kappa^2 T_{3,3} + T_{2,2,2} \tag{36}$$

- T_6 includes the terms involved in S_6 for which all the six summation indexes are equal: obviously, we have $T_6 = \sum_k c_k^2$;
- $T_{4,2}$ groups together all the terms such that, among the six summation indices, there is one group of four equal indices on the one hand, and another group of two equal indices on the other hand (for instance: $i = j = l = n \neq k = m$);
- in the same way, $T_{3,3}$ includes all the terms such that the indices form two groups of three.
- finally, $T_{2,2,2}$ covers the situation where there are three pairs of equal indices.

A careful enumeration of the terms involved in these situations leads to the following expressions:

$$\begin{aligned} T_{4,2} &= \sum_{k\neq l} d_{k,l}^2 + 2\sum_k c_k D_k \\ T_{3,3} &= \sum_{k\neq l} c_k c_l + \sum_{k\neq l} d_{k,l}d_{l,k} \\ T_{2,2,2} &= \sum_k D_k^2 - \sum_{k\neq l} d_{k,l}^2 + \frac{1}{6}\sum_{k\neq l\neq m} e_{k,l,m}^2 \end{aligned} \tag{37}$$

By putting all things together, we finally obtain:

$$\begin{aligned} N^2 V_{\ddot{\delta}} = &\frac{\ddot{\mathfrak{m}}_6 - \ddot{\mathfrak{m}}_4^2 - \kappa^2}{\kappa^2}\sum_k c_k^2 + \frac{\ddot{\mathfrak{m}}_4 - \kappa^2 - 1}{\kappa^2}\sum_{k\neq l} d_{k,l}^2 \\ &+ \frac{1}{\kappa^2}\sum_k \left(\ddot{\mathfrak{m}}_4 c_k + D_k\right)^2 + \frac{1}{2}\sum_{k\neq l}\left(d_{k,l} + d_{l,k}\right)^2 \\ &+ \frac{1}{6\kappa^2}\sum_{k\neq l\neq m} e_{k,l,m}^2 \end{aligned} \tag{38}$$

It can be verified that this expression (38) is indeed positive, as for any probability distribution with moments $\mathfrak{m}_p$, the following Hankel matrix H_p is positive (see [27]):

$$H_p = \begin{bmatrix} 1 & \mathfrak{m}_1 & \mathfrak{m}_2 & \cdots & \mathfrak{m}_p \\ \mathfrak{m}_1 & \mathfrak{m}_2 & & ⋰ & \vdots \\ \mathfrak{m}_2 & & ⋰ & & \vdots \\ \vdots & ⋰ & & & \vdots \\ \mathfrak{m}_p & \cdots & \cdots & \cdots & \mathfrak{m}_{2p} \end{bmatrix} \tag{39}$$

In our case, this implies:

$$\ddot{\mathfrak{m}}_6 - \ddot{\mathfrak{m}}_4^2 - \kappa^2 \geq 0 \quad \text{and} \quad \ddot{\mathfrak{m}}_4 - \kappa^2 - 1 \geq 0 \tag{40}$$

As in the case of $V_{\dot{\delta}}$, we can analyze the order of magnitude of the contribution $V_{\ddot{\delta}}$ to the variance of the PG-URE estimator. As explained in Sec. III-E, for reasonable denoising operators $\boldsymbol{f}$, the second order derivative $\frac{\partial^2 f_k}{\partial y_l \partial y_m}(\boldsymbol{y})$ is likely to be zero, except when the pixels corresponding to the indexes k, l and m share some spatial proximity. As the $b_{k,l,m}$ are proportional to these second order derivatives, and due to their definition, we deduce that the number of non-zero coefficients $d_{k,l}$ and $e_{k,l,m}$ is proportional to N; the order of magnitude of $V_{\ddot{\delta}}$ is therefore proportional to $\frac{1}{N}$, as claimed in Sec. III-E.

B. Optimal Probability Distribution

Selection of the probability distribution of $\ddot{\delta}$ can be formulated as an optimization problem consisting in minimizing the right-hand side of (38) seen as a function of $\ddot{\mathfrak{m}}_6$, $\ddot{\mathfrak{m}}_4$ and κ, subject to the feasibility constraints (40).

As a first remark, it can be observed that for fixed values of the variables $\ddot{\mathfrak{m}}_4$ and κ, the minimal value of $V_{\ddot{\delta}}$ is reached with $\ddot{\mathfrak{m}}_6 = \ddot{\mathfrak{m}}_4^2 + \kappa^2$. Then, by re-injecting this optimal condition in (40), and by removing the constant terms, the problem can be restated as minimizing the following function Φ:

$$\begin{aligned} \Phi\left(\ddot{\mathfrak{m}}_4, \kappa\right) &= \frac{\alpha\ddot{\mathfrak{m}}_4^2 + \beta\ddot{\mathfrak{m}}_4 + \gamma}{\kappa^2} \quad \text{subject to} \quad \ddot{\mathfrak{m}}_4 \geq \kappa^2 + 1 \\ \text{with} \quad \alpha &= \sum_k c_k^2 \quad \beta = 2\sum_k c_k D_k + \sum_{k\neq l} d_{k,l}^2 \\ \gamma &= \sum_k D_k^2 - \sum_{k\neq l} d_{k,l}^2 + \frac{1}{6}\sum_{k\neq l\neq m} e_{k,l,m}^2 \end{aligned} \tag{41}$$

From the definition of the coefficients α, β and γ, it can be checked that $\alpha\ddot{\mathfrak{m}}_4^2 + \beta\ddot{\mathfrak{m}}_4 + \gamma \geq 0$ when $\ddot{\mathfrak{m}}_4 \geq 1$ (which is the case in the constraint domain). This implies that, for a fixed value of $\ddot{\mathfrak{m}}_4$, the function $\Phi\left(\ddot{\mathfrak{m}}_4, \kappa\right)$ decreases when κ^2 increases: the minimal value is therefore obtained on the boundary of the feasibility domain, i.e. $\ddot{\mathfrak{m}}_4 = \kappa^2 + 1$. Finally, a basic function analysis shows that $\Phi\left(\kappa^2+1,\kappa\right)$ reaches a minimum value when $\kappa^4 = (\alpha+\beta+\gamma)/\alpha$.

To summarize, the contribution $V_{\ddot{\delta}}$ to the variance of the PG-URE estimator is minimal under the following conditions:

$$\ddot{\mathrm{m}}_6 = \kappa^4 + 3\kappa^2 + 1 \qquad \ddot{\mathrm{m}}_4 = \kappa^2 + 1 \qquad \kappa = \pm\kappa^\star$$
$$\text{with } \kappa^\star = \left(\frac{\sum_k \left(c_k + D_k\right)^2 + \frac{1}{6}\sum_{k\neq l\neq m} e_{k,l,m}^2}{\sum_k c_k^2} \right)^{\frac{1}{4}} \tag{42}$$

The probability distribution π (20) defined in Sec. III-D for $\ddot{\delta}$ do verify the conditions on $\ddot{\mathrm{m}}_6$ and $\ddot{\mathrm{m}}_4$. However, in practice, the optimal value $\kappa^\star$ of the third moment cannot be evaluated, as we do not know the values of the partial derivatives involved in the definition of the coefficients c_k, D_k and $e_{k,l,m}$.

In Sec. III-D, we propose to use $\kappa = 1$: although this choice is certainly not optimal in all cases, we can propose a sketch of proof from the expression of $\kappa^\star$ (42). Indeed, the coefficients c_k are proportional to $\frac{\partial^2 f_k}{\partial y_k^2}(\boldsymbol{y})$, while the D_k and $e_{k,l,m}$ depend only on second-order partial derivatives $\frac{\partial^2 f_k}{\partial y_l \partial y_m}(\boldsymbol{y})$ for which at least $k \neq l$ or $k \neq m$: then, under the hypothesis that, for a reasonable denoising operator $\boldsymbol{f}$, the k^{th} output pixel depends mostly on the k^{th} input pixel, we can assume that $\frac{\partial^2 f_k}{\partial y_k^2}(\boldsymbol{y})$ have higher order of magnitude than $\frac{\partial^2 f_k}{\partial y_l \partial y_m}(\boldsymbol{y})$. We deduce that $|c_k| \gg |D_k|$ and $|c_k| \gg \left|e_{k,l,m}\right|$, and therefore that $\kappa^\star \approx 1$. However, several approximation and hypothesis are made here: a quantitative analysis of the statistical distribution of the second-order partial derivative values would certainly be desirable to achieve better approximations.

ACKNOWLEDGMENT

The authors would like to thank Corinne Lorenzo, from Institut des Technologies Avancées en Sciences du Vivant (Toulouse, France), and Jordi Andilla and Pablo Loza-Alvarez, from the Institute of Photonic Sciences (Barcelona, Spain), for providing the *MCTS* test image. We are also grateful to the anonymous reviewers, for their valuable remarks.

REFERENCES

[1] D. L. Donoho, "De-noising by soft-thresholding," *IEEE Trans. Inf. Theory*, vol. 41, no. 3, pp. 613–627, May 1995.

[2] L. I. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *Phys. D, Nonlinear Phenomena*, vol. 60, no. 1, pp. 259–268, 1992.

[3] A. Buades, B. Coll, and J.-M. Morel, "A review of image denoising algorithms, with a new one," *Multiscale Model. Simul.*, vol. 4, no. 2, pp. 490–530, 2005.

[4] M. Elad and M. Aharon, "Image denoising via sparse and redundant representations over learned dictionaries," *IEEE Trans. Image Process.*, vol. 15, no. 12, pp. 3736–3745, Dec. 2006.

[5] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, "Image denoising by sparse 3D transform-domain collaborative filtering," *IEEE Trans. Image Process.*, vol. 16, no. 8, pp. 2080–2095, Aug. 2007.

[6] P. Milanfar, "A tour of modern image filtering," *Signal Process. Mag.*, vol. 30, no. 1, pp. 106–128, 2013.

[7] L. Zhang, L. Zhang, X. Mou, and D. Zhang, "A comprehensive evaluation of full reference image quality assessment algorithms," in *Proc. Int. Conf. Image Process.*, 2012, pp. 1477–1480.

[8] C. M. Stein, "Estimation of the mean of a multivariate normal distribution," *Ann. Statist.*, vol. 9, no. 6, pp. 1135–1151, 1981.

[9] D. L. Donoho and I. M. Johnstone, "Adapting to unknown smoothness via wavelet shrinkage," *J. Amer. Statist. Assoc.*, vol. 90, no. 432, pp. 1200–1224, 1995.

[10] A. Benazza-Benyahia and J.-C. Pesquet, "Building robust wavelet estimators for multicomponent images using Stein's principle," *IEEE Trans. Image Process.*, vol. 14, no. 11, pp. 1814–1830, Nov. 2005.

[11] D. Van De Ville and M. Kocher, "SURE-based non-local means," *IEEE Signal Process. Lett.*, vol. 16, no. 11, pp. 973–976, Nov. 2009.

[12] S. Ramani, T. Blu, and M. Unser, "Monte-Carlo SURE: A black-box optimization of regularization parameters for general denoising algorithms," *IEEE Trans. Image Process.*, vol. 17, no. 9, pp. 1540–1554, Sep. 2008.

[13] J.-L. Starck, F. Murtagh, and A. Bijaoui, *Image Processing and Data Analysis*. Cambridge, U.K.: Cambridge Univ. Press, 1998.

[14] B. Zhang, J. M. Fadili, and J.-L. Starck, "Wavelets, ridgelets, and curvelets for Poisson noise removal," *IEEE Trans. Image Process.*, vol. 17, no. 7, pp. 1093–1108, Jul. 2008.

[15] F. Luisier, T. Blu, and M. Unser, "Image denoising in mixed Poisson-Gaussian noise," *IEEE Trans. Image Process.*, vol. 20, no. 3, pp. 696–708, Mar. 2011.

[16] P. Perona and J. Malik, "Scale-space and edge detection using anisotropic diffusion," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 12, no. 7, pp. 629–639, Jul. 1990.

[17] C. Vonesch, S. Ramani, and M. Unser, "Recursive risk estimation for non-linear image deconvolution with a wavelet-domain sparsity constraint," in *Proc. 15th Int. Conf. Image Process.*, 2008, pp. 665–668.

[18] J.-C. Pesquet, A. Benazza-Benyahia, and C. Chaux, "A SURE approach for digital signal/image deconvolution problems," *IEEE Trans. Signal Process.*, vol. 57, no. 12, pp. 4616–4632, Dec. 2009.

[19] Y. C. Eldar, "Generalized SURE for exponential families: Applications to regularization," *IEEE Trans. Signal Process.*, vol. 57, no. 2, pp. 471–481, Feb. 2009.

[20] R. Giryes, M. Elad, and Y. C. Eldar, "The projected GSURE for automatic parameter tuning in iterative shrinkage methods," *Appl. Comput. Harmon. Anal.*, vol. 30, no. 3, pp. 407–422, 2011.

[21] F. Luisier, C. Vonesch, T. Blu, and M. Unser, "Fast interscale wavelet denoising of Poisson-corrupted images," *Signal Process.*, vol. 90, no. 2, pp. 415–427, Feb. 2010.

[22] B. Zhang, J. M. Fadili, J.-L. Starck, and J.-C. Olivo-Marin, "Multiscale variance-stabilizing transform for mixed Poisson-Gaussian processes and its applications in bioimaging," in *Proc. Int. Conf. Image Process.*, 2007, pp. 233–236.

[23] 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.*, vol. 17, no. 10, pp. 1737–1754, Oct. 2008.

[24] A. Jezierska, C. Chaux, J.-C. Pesquet, and H. Talbot, "An EM approach for Poisson-Gaussian noise modeling," in *Proc. Eur. Signal Process. Conf.*, 2011, pp. 2244–2248.

[25] A. Jezierska, H. Talbot, C. Chaux, J.-C. Pesquet, and G. Engler, "Poisson-Gaussian noise parameter estimation in fluorescence microscopy," in *Proc. 9th Int. Symp. Biomed. Imag.*, 2012, pp. 1663–1666.

[26] B. Zhang, "Contributions à la microscopie à fluorescence en imagerie biologique: Modélisation de la PSF, restauration d'images et détection super-résolutive," Ph.D. dissertation, Dépt. traitement du Signal et des Images, Telecom ParisTech, Paris, France, 2007.

[27] N. I. Akhiezer, *The Classical Moment Problem and Some Related Questions in Analysis*. London, U.K.: Oliver & Boyd, 1965.

[28] D. Goldberg, "What every computer scientist should know about floating-point arithmetic," *ACM Comput. Surveys*, vol. 23, no. 1, pp. 5–48, 1991.

[29] J. C.-M. Peng. (1975). *Simultaneous Estimation of the Parameters of Independent Poisson Distribution*. Dept. Statist., Stanford University, Stanford, CA, USA. Tech. Rep. 78 [Online]. Available: https://statistics.stanford.edu/sites/default/files/EFS%2520NSF%252078.pdf

[30] K.-W. Tsui and S. J. Press, "Simultaneous estimation of several Poisson parameters under K-normalized squared error loss," *Ann. Statist.*, vol. 10, no. 1, pp. 93–100, 1982.

Yoann Le Montagner received the M.Sc. degree in applied mathematics from École Polytechnique, Palaiseau, France, in 2010, and the Ph.D. degree in signal and image processing from Télécom ParisTech, Paris, France, in 2013. From 2010 to 2013, he was with the Signal and Image Processing Department, Télécom ParisTech, Institut Mines-Télécom, Paris, and with the Quantitative Image Analysis Unit, Institut Pasteur, France. His research interests include compressed sensing, phase retrieval, statistical signal processing, denoising, and inverse problems.

Elsa D. Angelini (M'98–SM'12) received the B.Sc. degree from Ecole Centrale de Nantes, Nantes, France, in 1996, and the M.Sc. and Ph.D. degrees in biomedical engineering from Columbia University, New York, NY, USA, in 1998 and 2003, respectively. She currently holds a dual position as an Associate Professor of computer science with Institut Mines-Télécom, Télécom ParisTech, Paris, France, and a Senior Research Scientist with the Department of Biomedical Engineering, Columbia University, USA. Her research focuses on image processing for multidimensional biomedical imaging, including applications such as denoising, enhancement, segmentation, and modeling. She was an Associate Editor for the IEEE TRANSACTIONS ON BIOMEDICAL ENGINEERING from 2008 to 2012, and she is currently the Chair of the EMBS TC on Biomedical Imaging and Image Processing, and a member of the IEEE EMBS Administrative Committee.

Jean-Christophe Olivo-Marin received the Ph.D. and H.D.R. degrees in optics and signal processing from Institut d'Optique Théorique et Appliquée, University of Paris-Orsay, France. He is the Head of the Quantitative Image Analysis Unit, Institut Pasteur, and the Chair of the Cell Biology and Infection Department. He was a cofounder of the Institut Pasteur Korea, Seoul, where he held a joint appointment as a Chief Technology Officer from 2004 to 2005. He was a Staff Scientist with the European Molecular Biology Laboratory, Heidelberg, from 1990 to 1998. His research interests are in image analysis of multidimensional microscopy images, computer vision and motion analysis for cellular dynamics, and in multidisciplinary approaches for biological imaging. He is a Past Chair of the IEEE SPS Bio Imaging and Signal Processing Technical Committee, Senior Area Editor of the IEEE SIGNAL PROCESSING LETTERS, and member of the editorial board of the journals *Medical Image Analysis* and *BMC Bioinformatics*. He was the General Chair of the IEEE International Symposium on Biomedical Imaging in 2008.