

PAPER • OPEN ACCESS

Masked unbiased principles for parameter selection in variational image restoration under Poisson noise

To cite this article: Francesca Bevilacqua *et al* 2023 *Inverse Problems* **39** 034002

View the [article online](#) for updates and enhancements.

You may also like

- [THE HERschel INVENTORY OF THE AGENTS OF GALAXY EVOLUTION IN THE MAGELLANIC CLOUDS. A HERSCHEL OPEN TIME KEY PROGRAM](#)
M. Meixner, P. Panuzzo, J. Roman-Duval et al.
- [Masked and unmasked error-related potentials during continuous control and feedback](#)
Catarina Lopes Dias, Andreea I Sburlea and Gernot R Müller-Putz
- [INTERPOLATING MASKED WEAK-LENSING SIGNAL WITH KARHUNEN-LOËVE ANALYSIS](#)
J. T. VanderPlas, A. J. Connolly, B. Jain et al.

Masked unbiased principles for parameter selection in variational image restoration under Poisson noise

Francesca Bevilacqua¹ , Alessandro Lanza¹ ,
Monica Pragliola^{2,*}  and Fiorella Sgallari¹ 

¹ Department of Mathematics, University of Bologna, Bologna, Italy

² Department of Mathematics and Applications, University of Naples Federico II, Naples, Italy

E-mail: monica.pragliola@unina.it

Received 21 September 2022; revised 6 December 2022

Accepted for publication 6 January 2023

Published 23 January 2023



CrossMark

Abstract

In this paper we address the problem of automatically selecting the regularization parameter in variational models for the restoration of images corrupted by Poisson noise. More specifically, we first review relevant existing *unmasked* selection criteria which fully exploit the acquired data by considering all pixels in the selection procedure. Then, based on an idea originally proposed by Carlan and Blanc-Feraud to effectively deal with dark backgrounds and/or low photon-counting regimes, we introduce and discuss the *masked* versions—some of them already existing—of the considered unmasked selection principles formulated by simply discarding the pixels measuring zero photons. However, we prove that such a blind masking strategy yields a bias in the resulting principles that can be overcome by introducing a novel positive Poisson distribution correctly modeling the statistical properties of the undiscarded noisy data. Such distribution is at the core of newly proposed *masked unbiased* counterparts of the discussed strategies. All the unmasked, masked biased and masked unbiased principles are extensively compared on the restoration of different images in a wide range of photon-counting regimes. Our tests allow to conclude that the novel masked unbiased selection strategies, on average, compare favorably with unmasked and masked biased counterparts.

Keywords: inverse problems, variational methods, Poisson noise, regularization parameter, discrepancy principle, whiteness property

(Some figures may appear in colour only in the online journal)

* Author to whom any correspondence should be addressed.



1. Introduction

Poisson noise is a very pervasive noise statistics in imaging problems, as it arises whenever the acquired data is formed by counting the number of photons irradiated by a source and hitting the image domain [3]. Typical applications where the Poisson noise removal is a particularly relevant problem are astronomical and microscopy imaging; both scenarios are characterized by a low-light condition, which is intrinsically related to the acquisition set-up in the former case [22], while in the latter it is somehow preferable so as to preserve the specimen of interest [18]. However, the weaker the light intensity, the stronger the degradation in the acquired images.

The image formation model under Poisson noise corruption can be formulated in vectorized form as follows

$$\mathbf{y} = \text{poiss}(\bar{\boldsymbol{\lambda}}), \quad \bar{\boldsymbol{\lambda}} = \mathbf{H}\bar{\mathbf{x}} + \mathbf{b}, \quad (1)$$

where $\mathbf{y} \in \mathbb{N}^n$, $\bar{\mathbf{x}} \in \mathbb{R}_+^n$ and $\mathbf{b} \in \mathbb{R}_+^n$ —with \mathbb{N} and \mathbb{R}_+ denoting the sets of natural numbers including zero and of non-negative real numbers, respectively—are vectorized forms of the $n_1 \times n_2$ observed degraded image, unknown uncorrupted image and (typically) known background emission image, respectively, with $n := n_1 n_2$. The blurring matrix $\mathbf{H} \in \mathbb{R}^{n \times n}$ models the action of a known linear blurring operator, which typically arises in the context of astronomical and microscopy image processing. Finally, $\text{poiss}(\bar{\boldsymbol{\lambda}}) := (\text{poiss}(\bar{\lambda}_1), \dots, \text{poiss}(\bar{\lambda}_n))^T$, with $\text{poiss}(\bar{\lambda}_i)$ indicating the realization of a Poisson-distributed random variable with parameter (mean) $\bar{\lambda}_i$. Hence $\bar{\boldsymbol{\lambda}} \in \mathbb{R}_+^n$ —as all entries of \mathbf{H} are non-negative—is the vectorized form of the $n_1 \times n_2$ noise-free degraded image.

The problem of recovering $\bar{\mathbf{x}}$ starting from the knowledge of the observation \mathbf{y} and of the acquisition model in (1) is typically ill-posed, so that one rather seeks for an estimate $\hat{\mathbf{x}}$ of the target image $\bar{\mathbf{x}}$ which solves a well-posed problem as close as possible to the original one. In the variational framework, the estimate $\hat{\mathbf{x}}$ is obtained as a global minimizer of a cost (or energy) functional $\mathcal{J} : \mathbb{R}^n \rightarrow \mathbb{R}$ typically taking the form

$$\hat{\mathbf{x}}(\mu) \in \arg \min_{\mathbf{x} \in \mathbb{R}_+^n} \{\mathcal{J}(\mathbf{x}; \mu) = \mathcal{R}(\mathbf{x}) + \mu \mathcal{F}(\boldsymbol{\lambda}; \mathbf{y})\}, \quad \boldsymbol{\lambda} = \mathbf{H}\mathbf{x} + \mathbf{b}, \quad (2)$$

where \mathcal{R} and \mathcal{F} are the so-called *regularization* and *data fidelity* term, respectively, while $\mu > 0$ is the *regularization parameter* balancing the action of \mathcal{F} and \mathcal{R} in the overall functional.

The data fidelity term $\mathcal{F}(\boldsymbol{\lambda}; \mathbf{y})$ measures the distance between $\boldsymbol{\lambda}$ and \mathbf{y} with a metric induced by the noise statistics. In presence of Poisson noise, based on the maximum likelihood estimation approach (see, e.g. [5]), the fidelity term is typically set as the generalized Kullback–Leibler (KL) divergence between $\boldsymbol{\lambda}$ and \mathbf{y} , that is

$$\mathcal{F}(\boldsymbol{\lambda}; \mathbf{y}) = \text{KL}(\boldsymbol{\lambda}; \mathbf{y}) := \sum_{i \in \mathcal{I}} F(\lambda_i; y_i), \quad (3)$$

where

$$F(\lambda_i; y_i) = \begin{cases} \lambda_i - y_i \ln \lambda_i + y_i \ln y_i - y_i & \text{if } i \in \mathcal{I}_+ \\ \lambda_i & \text{if } i \in \mathcal{I}_0 := \mathcal{I} \setminus \mathcal{I}_+ \end{cases}, \quad (4)$$

with \mathcal{I} , \mathcal{I}_+ , \mathcal{I}_0 the sets of pixel indices defined by

$$\mathcal{I} := \{1, \dots, n\}, \quad \mathcal{I}_+ := \{i \in \mathcal{I} \mid y_i > 0\}, \quad \mathcal{I}_0 := \mathcal{I} \setminus \mathcal{I}_+ = \{i \in \mathcal{I} \mid y_i = 0\}.$$

On the other hand, the regularization term $\mathcal{R}(\mathbf{x})$ encodes information or beliefs that may be available *a priori* on the target image $\bar{\mathbf{x}}$, such as smoothness or sparsity properties. One of the most popular regularizers in image processing is the total variation (TV) semi-norm [19], which reads

$$\mathcal{R}(\mathbf{x}) = \text{TV}(\mathbf{x}) := \sum_{i \in \mathcal{I}} \|(\nabla \mathbf{x})_i\|_2, \quad (5)$$

where $(\nabla \mathbf{x})_i \in \mathbb{R}^2$ represents the discrete gradient of image \mathbf{x} computed at pixel i . The TV term, which induces sparsity of gradient magnitudes, is known to be particularly effective for the restoration of piece-wise constant images; however, it is also well-established that the TV regularizer tends to promote edges thus producing the so-called *staircasing effect* on the smooth parts of the image. As a way to overcome the classical drawbacks of TV, one can employ the TV^2 regularizer [17] defined by

$$\mathcal{R}(\mathbf{x}) = \text{TV}^2(\mathbf{x}) := \sum_{i \in \mathcal{I}} \|(\nabla^2 \mathbf{x})_i\|_F, \quad (6)$$

with $(\nabla^2 \mathbf{x})_i \in \mathbb{R}^{2 \times 2}$ indicating the discrete Hessian of image \mathbf{x} at pixel i and $\|\cdot\|_F$ denoting the Frobenius norm. The TV^2 regularizer promotes piecewise-affine structures in the image, however its ability to recover sharp edges is less than TV.

A particularly flexible regularization term that couples the benefits of the TV and of the TV^2 regularizers while mitigating their shortcomings is the popular total generalized variation (TGV) [14], in particular its second-order version which reads

$$\mathcal{R}(\mathbf{x}) = \text{TGV}^2(\mathbf{x}) := \min_{\mathbf{w} \in \mathbb{R}^{2n}} \left\{ \alpha_0 \sum_{i \in \mathcal{I}} \|(\nabla \mathbf{x})_i - \mathbf{w}_i\|_2 + \alpha_1 \sum_{i \in \mathcal{I}} \|(\mathcal{E} \mathbf{w})_i\|_F \right\}, \quad (7)$$

where $\mathbf{w} = (\mathbf{w}_1; \mathbf{w}_2)$ with $\mathbf{w}_1, \mathbf{w}_2 \in \mathbb{R}^n$, $\mathbf{w}_i := (w_{1,i}; w_{2,i}) \in \mathbb{R}^2$ and $(\mathcal{E} \mathbf{w})_i \in \mathbb{R}^{2 \times 2}$ denotes the discrete symmetric Jacobian of vector field \mathbf{w} at pixel i , with α_0, α_1 being positive parameters.

Besides the choice of suitable data and regularization terms, the selection of the regularization parameter μ can highly influence the quality of the output restoration. Generally speaking, a criterion for the selection of μ in variational models of the form (2) can be formulated as follows

$$\text{Select } \mu = \hat{\mu} \text{ such that } \mathcal{C}(\hat{\mathbf{x}}(\hat{\mu})) \text{ is satisfied,}$$

where $\hat{\mathbf{x}}(\mu) : \mathbb{R}_{++} \rightarrow \mathbb{R}^n$, with $\mathbb{R}_{++} = \mathbb{R}_+ \setminus \{0\}$, is the image restoration function introduced in (2) and $\mathcal{C}(\cdot)$ is some selection criterion or principle.

Historically, the criteria proposed for the μ -selection under Poisson noise corruption have been designed so as to extend the particularly wide literature for the parameter selection in presence of additive white Gaussian noise. In this perspective, some attempts have been made in order to modify the generalized cross validation function, originally proposed in [10], so as to be applied to the case of Poisson noise [13, 24]. Nonetheless, such strategies have never been significantly employed in imaging problems.

A particularly successful strategy, proposed by Bertero and co-authors in the seminal works [4, 26, 27], extends the popular discrepancy principle (DP) to Poisson noise scenarios. Ultimately, one requires that the KL fidelity term, evaluated at $\hat{\mathbf{x}}(\mu)$ and regarded as a function of μ , equals a given discrepancy value. In [4, 26, 27] this value is obtained by approximating the expected value of the KL term now interpreted as a function of the Poisson-distributed random vector \mathbf{Y} , of which the observed image \mathbf{y} has to be considered a realization. This instance of the DP is easy to implement and relies on pleasant theoretical guarantees—see, e.g. [21, 23, 28]. However, it is also well established that it does not yield a good selection in low photon-counting situations. This has been ascribed to the approximation of the original underlying expected value, that several strategies tried to overcome [5, 7].

Different methods rely on transforming the noise statistics. More specifically, in [2] the authors consider a Gaussian approximation of the noise and then employ the popular Morozov DP. More recently, a criterion based on the whiteness property of a standardized Poisson distribution has been proposed [6].

In [9], the authors raise the possibility that the sub-optimal results obtained via the DPs in [2, 26] can be explained in light of the fact that the pixels corresponding to zero-entries in the data should not be considered when solving the corresponding non-linear equations. Nonetheless, such modified principles still rely on modeling the corrupting noise as a Poisson distribution that does not explicitly encode the discarding of the zero-pixels, thus introducing a bias in the final formulation of the strategies.

1.1. Contribution

In this paper, we review, discuss and compare experimentally all the existing state-of-the-art principles and five new ones for the automatic selection of the regularization parameter μ in the class of so-called \mathcal{R} -KL variational models for image restoration,

$$\hat{\mathbf{x}}(\mu) \in \arg \min_{\mathbf{x} \in \mathbb{R}_+^n} \{ \mathcal{R}(\mathbf{x}) + \mu \text{KL}(\boldsymbol{\lambda}; \mathbf{y}) \}, \quad \boldsymbol{\lambda} = \mathbf{H}\mathbf{x} + \mathbf{b}, \quad (8)$$

with \mathcal{R} being a general closed, proper and convex regularization term such as, e.g. (5)–(7). More specifically, our contribution can be summarized as follows.

- (a) We provide a detailed and organic review of the most popular and effective existing *unmasked* principles (including the two very recent ones proposed in [5, 6]), which fully exploit the information in the data without discarding any pixel.
- (b) Inspired by the idea originally proposed in [9] to deal with low photon-counting scenarios, we introduce and discuss the *masked biased* versions of the previously reviewed unmasked principles, some of them already proposed in [9] and other new. These approaches come from simply discarding the zero-pixels in the acquired image when applying the principles while, at the same time, keeping the (Poisson) distribution of undiscarded data unchanged.
- (c) As the main contribution of this work, we propose a whole new class of *masked unbiased* selection criteria based on the introduction of a novel positive Poisson distribution which suitably models the data statistics after discarding the zero-photon pixels. A theoretical analysis of the biases eliminated by using the new unbiased principles is also carried out.
- (d) As a final contribution, we extensively and reliably test the three categories of principles—i.e. the unmasked, masked biased and novel masked unbiased strategies—on the restoration of different images, using different regularizers and in a wide range of different photon-counting scenarios, from very low to high counting situations. The computed

examples strongly indicate that, on average, the proposed masked unbiased criteria outperform both their unmasked and their masked biased counterparts, especially in the low-counting regime.

The paper is organized as follows. In section 2 we review and discuss the most relevant existing unmasked principles, while the associated masked biased criteria are outlined in section 3. The novel positive Poisson distribution as well as the resulting masked unbiased approaches are illustrated in detail in section 4. Numerical solution of the class of \mathcal{R} -KL variational models by the alternating direction method of multipliers (ADMM) is dealt with in section 5. Extensive numerical tests assessing the performance of the considered principles are carried out in section 6. Section 7 concludes our work with some final considerations and outlook for future research.

2. Unmasked principles

In this section, we recall and outline the four most relevant unmasked parameter selection principles for Poisson noise proposed in literature so far, namely those principles which fully exploit the information encoded in the observed image \mathbf{y} without discarding the zero-photon pixels. To this aim, first we introduce the μ -dependent image

$$\widehat{\boldsymbol{\lambda}}(\mu) := \mathbf{H}\widehat{\mathbf{x}}(\mu) + \mathbf{b} \in \mathbb{R}^n, \quad (9)$$

which represents, for each selected μ value, an estimate of the unknown true noise-free image $\overline{\boldsymbol{\lambda}}$ defined in (1), obtained by solving the \mathcal{R} -KL variational model in (8). We also introduce the true and estimated standardized images

$$\bar{\mathbf{z}} := \frac{\mathbf{y} - \overline{\boldsymbol{\lambda}}}{\sqrt{\overline{\boldsymbol{\lambda}}}} \in \mathbb{R}^n, \quad \widehat{\mathbf{z}}(\mu) := \frac{\mathbf{y} - \widehat{\boldsymbol{\lambda}}(\mu)}{\sqrt{\widehat{\boldsymbol{\lambda}}(\mu)}} \in \mathbb{R}^n. \quad (10)$$

where all operations in the above definitions have to be intended component-wise.

It can be proved—see Proposition 1 in [6]—that the original matrix (or image) form of vector $\bar{\mathbf{z}}$ above is the realization of a 2D white (i.e. uncorrelated) random field. In particular, each entry \bar{z}_i of $\bar{\mathbf{z}}$ is the realization of a scalar random variable with zero mean and unitary variance.

2.1. The approximate DP

The abstract form of the DP applied to selecting the regularization parameter μ in the class of \mathcal{R} -KL variational models defined in (8) is as follows:

$$\text{Select } \mu = \widehat{\mu} \in \mathbb{R}_{++} \text{ such that } \mathcal{D}(\widehat{\mu}; \mathbf{y}) = \Delta, \quad (11)$$

where the last equality and the scalar $\Delta \in \mathbb{R}_+$ in (11) are commonly referred to as the *discrepancy equation* and the *discrepancy value*, respectively, while the *discrepancy function* $\mathcal{D}(\cdot; \mathbf{y}) : \mathbb{R}_{++} \rightarrow \mathbb{R}_+$ is defined by

$$\mathcal{D}(\mu; \mathbf{y}) := \text{KL}(\widehat{\boldsymbol{\lambda}}(\mu); \mathbf{y}) = \sum_{i \in \mathcal{I}} \left(\mathcal{D}_i(\mu; y_i) := F(\widehat{\lambda}_i(\mu); y_i) \right), \quad (12)$$

with the function F and the estimated noise-free image $\widehat{\boldsymbol{\lambda}}(\mu)$ defined in (4) and (9), respectively.

The DP formalizes a simple idea: choose a value $\hat{\mu}$ of the regularization parameter μ in the \mathcal{R} -KL variational model in (8) such that the value of the KL data fidelity term associated with the solution $\hat{\mathbf{x}}(\hat{\mu})$ is equal to a prescribed discrepancy value Δ .

The direct extension of the Morozov DP—originally designed for additive white Gaussian noise—to the case of Poisson corruption consists in selecting Δ as the expected value of the KL fidelity term in (3) regarded as a function of the n -variate Poisson-distributed random vector \mathbf{Y} (of which the deterministic measure $\mathbf{y} \in \mathbb{N}^n$ is a realization). This version of the DP for Poisson noise, that we refer to as the exact or expected value DP (EDP), can be formalized as follows

$$\begin{aligned} &\text{Select } \mu = \hat{\mu} \in \mathbb{R}_{++} \text{ such that } \mathcal{D}(\hat{\mu}; \mathbf{y}) = \Delta^{(E)}(\hat{\mu}), \\ &\text{with } \Delta^{(E)}(\mu) := \sum_{i \in \mathcal{I}} \left(\delta^{(E)}(\hat{\lambda}_i(\mu)) := \mathbb{E} \left[F(\hat{\lambda}_i(\mu); Y_i) \right] \right), \end{aligned} \quad (13)$$

where $\mathbb{E}[F(\hat{\lambda}_i(\mu); Y_i)]$ denotes the expected value of $F(\hat{\lambda}_i(\mu); Y_i)$ regarded as a function of the Poisson-distributed scalar random variable Y_i . Nonetheless, unlike the Gaussian case, the discrepancy value is not a constant but is a function $\Delta^{(E)}(\mu)$ of the regularization parameter μ , and deriving its analytic expression is a very hard task.

A popular and widespread strategy, originally proposed in [26] for denoising purposes and extended in [4] to the image restoration task, replaces the exact expected value function $\Delta^{(E)}(\mu)$ with a constant approximation coming from truncating its Taylor series expansion. We refer to this DP version as approximate DP (ADP). It reads:

$$\begin{aligned} &\text{Select } \mu = \hat{\mu} \in \mathbb{R}_{++} \text{ such that } \mathcal{D}(\hat{\mu}; \mathbf{y}) = \Delta^{(A)}, \\ &\text{with } \Delta^{(A)} := \sum_{i \in \mathcal{I}} \left(\delta^{(A)} := \frac{1}{2} \right) = \frac{n}{2}. \end{aligned} \quad (14)$$

The popularity of the ADP mainly relies on the strong theoretical guarantees that it brings along: in fact, existence and uniqueness of the solution of the discrepancy equation in (14) can be proven under very mild conditions. However, it is well-established (see, e.g. [5]) that the ADP tends to return either over-smoothed or under-smoothed restorations in low photon-counting scenarios.

2.2. The quadratic DP

In the volume where the ADP has been originally proposed, a different selection criterion also inspired by the Morozov DP has been published. Instead of approximating the expected value of the KL fidelity term, in [2] the authors propose to directly approximate (quadratically) the KL term in such a way that the expected value of the approximate term admits a simple closed-form expression. The approximation reads

$$\text{KL}(\boldsymbol{\lambda}; \mathbf{y}) := \sum_{i \in \mathcal{I}} F(\lambda_i; y_i) \approx \text{KL}^{(Q)}(\boldsymbol{\lambda}; \mathbf{y}) := \sum_{i \in \mathcal{I}} F^{(Q)}(\lambda_i; y_i),$$

with the introduced function $F^{(Q)}$ (approximating the function F in (4)) defined by

$$F^{(Q)}(\lambda_i; y_i) = \left(\frac{y_i - \lambda_i}{\sqrt{\lambda_i}} \right)^2. \quad (15)$$

The quadratically approximated version $\mathcal{D}^{(Q)}$ of the exact discrepancy function \mathcal{D} defined in (12) and used in the ADP in (14) thus reads

$$\mathcal{D}^{(Q)}(\mu; \mathbf{y}) := \text{KL}^{(Q)}(\hat{\boldsymbol{\lambda}}(\mu); \mathbf{y}) = \sum_{i \in \mathcal{I}} \left(\mathcal{D}_i^{(Q)}(\mu; y_i) := F^{(Q)}(\hat{\lambda}_i(\mu); y_i) \right). \quad (16)$$

By regarding $F^{(Q)}$ in (15) as a function of the Poisson-distributed random variable Y_i with mean λ_i , it is immediate to prove that [12]

$$\delta^{(Q)} := \mathbb{E} \left[F^{(Q)}(\lambda_i; Y_i) \right] = 1.$$

Hence, the DP version proposed in [2], referred to as quadratic DP (QDP), reads

$$\begin{aligned} &\text{Select } \mu = \hat{\mu} \in \mathbb{R}_{++} \text{ such that } \mathcal{D}^{(Q)}(\hat{\mu}; \mathbf{y}) = \Delta^{(Q)}, \\ &\text{with } \Delta^{(Q)} := \sum_{i \in \mathcal{I}} \left(\delta^{(Q)} = 1 \right) = n. \end{aligned} \quad (17)$$

2.3. The nearly exact DP

The two aforementioned strategies perform an approximation either on the discrepancy function or on its expected value. As anticipated, the latter scenario is known to return poor quality results especially in low counting regimes when the considered approximation becomes particularly rough. This issue has been first commented in [4], where the authors state (in remark 3) that the choice of the constant value $\delta^{(A)} = 1/2$ in the ADP in (14) may not be ‘optimal’ and suggest to replace it with $1/2 + \varepsilon$, where ε is a generic small positive or negative real number. Later, in [7] the authors proposed to introduce a non-constant ε , which is function $\varepsilon(\mu)$ of μ and that is set as the sum of the second to tenth terms of the same Taylor expansion used in [26]. However, such expansion converges only for λ approaching $+\infty$ —see [5]—and cannot aspire to improve the performance of ADP in low-count regimes.

Recently, in [5] the authors proposed a novel and more accurate approximation of the discrepancy value $\Delta^{(E)}$ in the EDP in (13) based on Monte Carlo simulation and nonlinear fitting. More specifically, on a fine grid of selected λ_i , a very large number of pseudo-random realizations of random quantity $F(\lambda_i; Y_i)$ are generated. By computing the sample mean of such realizations for each λ_i one gets back estimates for $\mathbb{E}[F(\lambda_i; Y_i)]$. A weighted least square fitting is then employed to infer a nearly-exact estimate $\delta^{(\text{NE})}(\lambda)$ of the exact expected value function $\delta^{(E)}(\lambda)$ used in (13), namely

$$\delta^{(\text{NE})}(\lambda) := \delta^{(A)} + \varepsilon(\lambda) = \frac{1}{2} + \frac{\lambda^2 + 2.5792\lambda - 1.5205}{12\lambda^3 - 5.6244\lambda^2 + 17.9347\lambda + 3.0410}. \quad (18)$$

Based on (18), the novel DP version proposed in [5], referred to by the authors as the nearly exact DP (NEDP), takes the form

$$\begin{aligned} &\text{Select } \mu = \hat{\mu} \in \mathbb{R}_{++} \text{ such that } \mathcal{D}(\hat{\mu}; \mathbf{y}) = \Delta^{(\text{NE})}(\hat{\mu}), \\ &\text{with } \Delta^{(\text{NE})}(\mu) = \sum_{i \in \mathcal{I}} \left(\delta^{(\text{NE})}(\hat{\lambda}_i(\mu)) \right) = \frac{n}{2} + \sum_{i \in \mathcal{I}} \varepsilon(\hat{\lambda}_i(\mu)). \end{aligned} \quad (19)$$

2.4. The whiteness principle

The popularity of the DP when the underlying noise is Gaussian has motivated the introduction of the three unmasked selection principles described above for Poisson noise. Besides the DP, a successful selection strategy in the Gaussian noise case consists in choosing μ so as to

maximize the whiteness of the residual image or, equivalently, minimize the cross-correlation between its entries [1, 15].

In the recent paper [6], the authors extended the above principle, referred to as the whiteness principle (WP), to the case of Poisson noise. In this scenario, the WP can be applied thanks to the introduction of the standardized images \bar{z} and $\hat{z}(\mu)$ defined in (10). In fact, it can be proved [6] that \bar{z} is the realization of a white random field and, hence, it makes sense to seek for μ maximizing the whiteness of the estimate $\hat{z}(\mu)$.

The WP for Poisson noise thus reads:

$$\text{Select } \mu = \hat{\mu} \in \arg \min_{\mu \in \mathbb{R}_{++}} \{W(\mu) := \mathcal{W}(\hat{z}(\mu))\}, \text{ with } \mathcal{W}(z) := \|\mathbf{S}(z)\|_2^2, \quad (20)$$

where $\mathbf{S}(z)$ denotes the 2D normalized auto-correlation of image z (see [6] for details).

3. Masked biased principles

After noting that the ADP and QDP principles defined in (14) and (17) can yield sub-optimal results in case of many zero-photon pixels, in [9] the authors proposed masked versions of those principles based on simply discarding all pixels measuring zero photons—i.e. pixels for which $y_i = 0$. We refer to these masked principles as *biased* since they do not consider that by carrying out a selection of pixels based on the value of the noise realization should require to change the statistics of the selected pixels, as it will be illustrated in section 4.

The masked versions of the exact discrepancy function \mathcal{D} in (12) used in the ADP (14) and of the quadratically approximated discrepancy function $\mathcal{D}^{(Q)}$ in (16) used in the QDP (17)—indicated by \mathcal{D}_+ and $\mathcal{D}_+^{(Q)}$, respectively—take clearly the following forms

$$\mathcal{D}_+(\mu; \mathbf{y}) := \sum_{i \in \mathcal{I}_+} F(\hat{\lambda}_i(\mu); y_i), \quad \mathcal{D}_+^{(Q)}(\mu; \mathbf{y}) := \sum_{i \in \mathcal{I}_+} F^{(Q)}(\hat{\lambda}_i(\mu); y_i), \quad (21)$$

with functions F and $F^{(Q)}$ defined in (4) and (15), respectively. Hence, based on their unmasked versions in (14) and (17), the ADP and QDP masked biased principles proposed in [9]—that we shortly refer to as ADP-MB and QDP-MB, respectively—can be formulated as follows:

$$\begin{aligned} &\text{Select } \mu = \hat{\mu} \in \mathbb{R}_{++} \text{ such that } \mathcal{D}_+(\hat{\mu}; \mathbf{y}) = \Delta_+^{(A)}, \\ &\text{with } \Delta_+^{(A)} := \sum_{i \in \mathcal{I}_+} \left(\delta^{(A)} = \frac{1}{2} \right) = \frac{n_+}{2}, \end{aligned} \quad (22)$$

$$\begin{aligned} &\text{Select } \mu = \hat{\mu} \in \mathbb{R}_{++} \text{ such that } \mathcal{D}_+^{(Q)}(\hat{\mu}; \mathbf{y}) = \Delta_+^{(Q)}, \\ &\text{with } \Delta_+^{(Q)} := \sum_{i \in \mathcal{I}_+} \left(\delta^{(Q)} = 1 \right) = n_+, \end{aligned} \quad (23)$$

where n_+ indicates the cardinality of set \mathcal{I}_+ , namely the number of non-zero pixels.

Also the NEDP in (19), which was proposed after [9], admits a masked biased version (NEDP-MB), which clearly reads:

$$\begin{aligned} &\text{Select } \mu = \hat{\mu} \in \mathbb{R}_+ \text{ such that } \mathcal{D}_+(\hat{\mu}; \mathbf{y}) = \Delta_+^{(\text{NE})}(\mu), \\ &\text{with } \Delta_+^{(\text{NE})}(\mu) := \sum_{i \in \mathcal{I}_+} \left(\delta^{(\text{NE})}(\hat{\lambda}_i(\mu)) \right) = \frac{n_+}{2} + \sum_{i \in \mathcal{I}_+} \epsilon(\hat{\lambda}_i(\mu)). \end{aligned} \quad (24)$$

Finally, by introducing the masked versions of the standardized image $\widehat{\mathbf{z}}(\mu)$ in (10), namely

$$\widehat{\mathbf{z}}_+(\mu) = (\widehat{z}_{+,1}(\mu), \dots, \widehat{z}_{+,n}(\mu))^T \quad \text{with} \quad \widehat{z}_{+,i}(\mu) = \begin{cases} \widehat{z}_i(\mu) & \text{if } i \in \mathcal{I}_+ \\ 0 & \text{otherwise} \end{cases}, \quad (25)$$

the masked biased version of the WP (WP-MB) can be formulated as follows

$$\text{Select } \mu = \widehat{\mu} \in \arg \min_{\mu \in \mathbb{R}_{++}} \{W_+(\mu) := \mathcal{W}(\widehat{\mathbf{z}}_+(\mu))\}, \quad (26)$$

with function $\mathcal{W}(\mathbf{z})$ defined as in (20).

4. The proposed masked unbiased principles

In the next subsection we introduce and analyze a novel scalar discrete distribution, called positive Poisson distribution, which correctly models the statistics of non-zero pixels considered in the masked selection principles. Based on such distribution, in subsections 4.2–4.5 we introduce the novel masked unbiased principles.

4.1. Positive Poisson distribution

In the following definition 4.1 we recall the Poisson distribution and introduce the positive Poisson distribution, then in proposition 4.2 we outline some important properties of positive Poisson-distributed random variables.

Definition 4.1 (Poisson and positive Poisson random variables). A discrete random variable Y is said to be Poisson-distributed with parameter $\lambda \in \mathbb{R}_{++}$, denoted by $Y \sim \mathcal{P}(\lambda)$, if its probability mass function reads

$$P_Y(y) = \frac{\lambda^y \exp(-\lambda)}{y!}, \quad y \in \mathbb{N}.$$

The expected value, variance and second-order raw moment of Y are given by

$$E[Y] = \text{Var}[Y] = \lambda, \quad E[Y^2] = \lambda(\lambda + 1). \quad (27)$$

The discrete random variable Y_+ defined by

$$Y_+ := Y \text{ if } Y > 0 \quad (28)$$

is said to be positive Poisson-distributed with parameter λ and denoted by $Y_+ \sim \mathcal{P}_+(\lambda)$.

Proposition 4.2. Let $Y \sim \mathcal{P}(\lambda)$ and $Y_+ \sim \mathcal{P}_+(\lambda)$, with $\lambda \in \mathbb{R}_{++}$, and let $T, V: \mathbb{R}_{++} \rightarrow \mathbb{R}$ be the functions defined by

$$T(\lambda) = \frac{1}{1 - \exp(-\lambda)}, \quad V(\lambda) = \frac{1 - (1 + \lambda) \exp(-\lambda)}{(1 - \exp(-\lambda))^2}. \quad (29)$$

Then, the probability mass function, expected value, second-order raw moment and variance of the positive Poisson-distributed random variable Y_+ read

$$P_{Y_+}(y) = T(\lambda) P_Y(y) = \frac{1}{\exp(\lambda) - 1} \frac{\lambda^y}{y!}, \quad y \in \mathbb{N}_0 := \mathbb{N} \setminus \{0\}, \quad (30)$$

$$E[Y_+] = T(\lambda) E[Y] = \frac{\lambda}{1 - \exp(-\lambda)}, \quad (31)$$

$$E[Y_+^2] = T(\lambda) E[Y^2] = \frac{\lambda(1 + \lambda)}{1 - \exp(-\lambda)}, \quad (32)$$

$$\text{Var}[Y_+] = V(\lambda) \text{Var}[Y] = \frac{\lambda}{(1 - \exp(-\lambda))^2} \left(1 - \frac{1 + \lambda}{\exp(\lambda)}\right). \tag{33}$$

For any $\lambda \in \mathbb{R}_{++}$, P_{Y_+} , $E[Y_+]$, $E[Y_+^2]$ and $\text{Var}[Y_+]$ in (30)–(33) satisfy

$$P_{Y_+}(y) > P_Y(y) \quad \forall y \in \mathbb{N}_0, \quad E[Y_+] > E[Y], \quad E[Y_+^2] > E[Y^2], \quad \text{Var}[Y_+] < \text{Var}[Y]. \tag{34}$$

Moreover, for λ tending to 0 and λ tending to $+\infty$, we have

$$\lim_{\lambda \rightarrow 0} P_{Y_+}(y) = \begin{cases} 1 & \text{for } y = 1, \\ 0 & \text{for } y > 1 \end{cases}, \quad \lim_{\lambda \rightarrow +\infty} P_{Y_+}(y) = P_Y(y) \quad \forall y \in \mathbb{N}_0, \tag{35}$$

$$\lim_{\lambda \rightarrow 0} E[Y_+] = 1, \quad \lim_{\lambda \rightarrow +\infty} E[Y_+] = E[Y], \tag{36}$$

$$\lim_{\lambda \rightarrow 0} E[Y_+^2] = 1, \quad \lim_{\lambda \rightarrow +\infty} E[Y_+^2] = E[Y^2], \tag{37}$$

$$\lim_{\lambda \rightarrow 0} \text{Var}[Y_+] = 0, \quad \lim_{\lambda \rightarrow +\infty} \text{Var}[Y_+] = \text{Var}[Y]. \tag{38}$$

Proof. It easily follows from definition (28) that, for any $\lambda \in \mathbb{R}_{++}$, the probability mass function $\mathcal{P}_{Y_+}(y)$ of $Y_+ \sim \mathcal{P}_+(\lambda)$ is a (positively) scaled version of the probability mass function $\mathcal{P}_Y(y)$ of $Y \sim \mathcal{P}(\lambda)$ for $y \in \mathbb{N}_0$. In formula,

$$\forall \lambda \in \mathbb{R}_{++}, \quad \exists \alpha(\lambda) \in \mathbb{R}_{++} : \quad P_{Y_+}(y) = \alpha(\lambda) P_Y(y) \quad \forall y \in \mathbb{N}_0. \tag{39}$$

By imposing that the probability mass function P_{Y_+} sums to one, it is easy to prove that $\alpha(\lambda)$ in (39) coincides with function $T(\lambda)$ in (29), thus demonstrating (30):

$$\begin{aligned} \sum_{y=1}^{\infty} P_{Y_+}(y) = 1 &\iff \alpha(\lambda) \sum_{y=1}^{\infty} P_Y(y) = 1 \iff \alpha(\lambda) = \frac{1}{\sum_{z=0}^{\infty} P_Y(z) - P_Y(0)} \\ &\iff \alpha(\lambda) = \frac{1}{1 - \exp(-\lambda)} = T(\lambda). \end{aligned} \tag{40}$$

Then, it easily follows from (30) that the m th order raw moments of Y_+ are given by

$$E[Y_+^m] = \sum_{y=1}^{\infty} y^m P_{Y_+}(y) = T(\lambda) \sum_{y=1}^{\infty} y^m P_Y(y) = T(\lambda) \sum_{y=0}^{\infty} y^m P_Y(y) = T(\lambda) E[Y^m],$$

for any $m \in \mathbb{N}$. By specifying the above formula for $m = 1$ and $m = 2$, one gets (31) and (32), respectively. It follows from (31)–(32) that

$$\begin{aligned} \text{Var}[Z] &= E\left[(Y_+ - E[Y_+])^2\right] = E[Y_+^2] - (E[Y_+])^2 = T(\lambda) E[Y^2] - (T(\lambda) E[Y])^2 \\ &= T(\lambda) (\lambda(1 + \lambda) - T(\lambda) \lambda^2) = T(\lambda) (1 + \lambda - \lambda T(\lambda)) \lambda \\ &= V(\lambda) \text{Var}[Y], \end{aligned} \tag{41}$$

where the last equality in (41) comes from the definition of functions T and V in (29) and from recalling that $\text{Var}[Y] = \lambda$.

Then, the inequalities in (34) and the four asymptotic properties (for $\lambda \rightarrow +\infty$) in (35)–(38) come from (30)–(33) and the following easily provable—see the plots in figure 1—properties of functions T and V defined in (29):

$$T(\lambda) > 1 \quad \forall \lambda \in \mathbb{R}_{++}, \quad V(\lambda) < 1 \quad \forall \lambda \in \mathbb{R}_{++}, \quad \lim_{\lambda \rightarrow +\infty} T(\lambda) = \lim_{\lambda \rightarrow +\infty} V(\lambda) = 1.$$

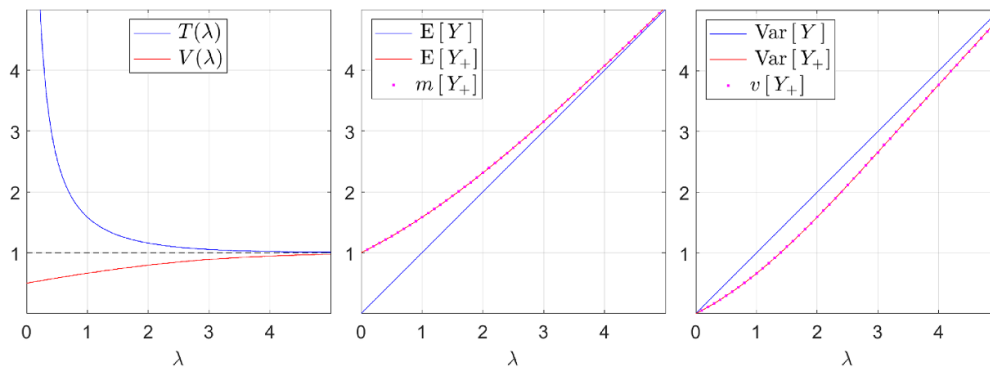


Figure 1. Plots of functions $T(\lambda)$, $V(\lambda)$ defined in (29) (left) and comparison between the expected values (center) and the variances (right) of $Y \sim \mathcal{P}(\lambda)$ and $Y_+ \sim \mathcal{P}_+(\lambda)$ given in (27) and in (31), (33), respectively, for varying λ . The magenta dots represent the sample means $m[Y_+]$ (center) and the sample variances $v[Y_+]$ (right) of very large numbers of realizations of the random variable Y_+ , for 50 different values of the parameter $\lambda \in \{0.1, 0.2, \dots, 4.9, 5\}$.

Finally, the four asymptotic properties (for $\lambda \rightarrow 0$) in (35)–(38) comes directly from taking the limits as λ approaches 0 of the four functions defined in (30)–(33). \square

In figure 1 we give a graphical representation of some of the quantities introduced in proposition 4.2. In particular, the sample means $m[Y_+]$ and sample variances $v[Y_+]$ (indicated by magenta dots and obtained by a simple Monte Carlo simulation) provide an experimental validation of the theoretically derived formulas for the expected value $E[Y_+]$ in (31) and the variance $\text{Var}[Y_+]$ in (33).

It is now clear that the Poisson and the positive Poisson random variables are characterized by significantly different statistical properties, especially for small values of the parameter λ . In this perspective, proposition 4.2 already warns on the approximations that the masked formulations of the principles given in section 3 bring along. An analysis of the introduced biases will be carried out in subsection 4.6.

In the next subsections, we are going to show how the newly introduced positive Poisson distribution can be adopted so as to formulate masked unbiased versions of the original principles reviewed here.

4.2. Masked unbiased approximate DP

As previously outlined, the approximate discrepancy value $\delta^{(A)} = 1/2$ used in the ADP relies on truncating at the first order the Taylor expansion of $E[F(\lambda; Y)]$, with Y a Poisson-distributed random variable with mean λ . It can be proved that, in the masked unbiased case (where Y is replaced by Y_+), the expected value $E[F(\lambda; Y_+)]$ admits a Taylor expansion which also coincides with $1/2$ when truncated at the first order. Hence, masked biased and masked unbiased versions of the ADP coincide; in what follows, they will be indistinctly referred to as ADP-M.

4.3. Masked unbiased quadratic DP

In light of statements (31) and (33) in proposition 4.2, we introduce the function

$$F^{(\text{QU})}(\lambda_i; y_i) = \left(\frac{y_i - \lambda_i T(\lambda_i)}{\sqrt{\lambda_i V(\lambda_i)}} \right)^2, \quad y_i \in \mathbb{N}_0.$$

Relying on the properties of the novel positive Poisson distribution, it is easy to observe that

$$\delta^{(\text{QU})} := \mathbb{E} \left[F^{(\text{QU})}(\lambda_i; Y_{+,i}) \right] = 1.$$

After introducing the unbiased version of the masked quadratically approximated discrepancy function $\mathcal{D}_+^{(\text{Q})}$ defined in (21) and used in the QDP-MB (23), namely

$$\mathcal{D}_+^{(\text{QU})}(\mu; \mathbf{y}) := \sum_{i \in \mathcal{I}_+} F^{(\text{QU})}(\hat{\lambda}_i(\mu); y_i),$$

we get the following formulation for the masked unbiased QDP, referred to as QDP-MU,

$$\begin{aligned} &\text{Select } \mu = \hat{\mu} \in \mathbb{R}_{++} \text{ such that } \mathcal{D}_+^{(\text{QU})}(\hat{\mu}; \mathbf{y}) = \Delta_+^{(\text{QU})}, \\ &\text{with } \Delta_+^{(\text{QU})} := \sum_{i \in \mathcal{I}_+} (\delta^{(\text{QU})} = 1) = n_+. \end{aligned} \quad (42)$$

4.4. Masked unbiased nearly exact DP

The masked unbiased version of NEDP (NEDP-MU) is obtained—analogously to the unmasked NEDP illustrated in [5]—by applying the weighted least squares fitting method to approximate the behavior of the sample means of large numbers of realizations of random quantity $F(\lambda_i; Y_{+,i})$, with F defined in (4). We thus get:

$$\text{Select } \mu = \hat{\mu} \in \mathbb{R}_{++} \text{ such that } \mathcal{D}_+(\hat{\mu}; \mathbf{y}) = \frac{n_+}{2} + \sum_{i \in \mathcal{I}_+} \epsilon^{(U)}(\hat{\lambda}_i(\mu)), \quad (43)$$

with

$$\epsilon^{(U)}(\lambda) = \frac{\lambda^3 + 30.7436\lambda^2 - 2.2968\lambda + 1.0827}{12\lambda^4 + 90.1921\lambda^3 - 1.8872\lambda^2 + 6.1778\lambda}. \quad (44)$$

4.5. Masked unbiased whiteness principle

We start noticing that the standardized image $\hat{\mathbf{z}}_+(\mu)$ in (25), which comes from a blind masking of the original $\hat{\mathbf{z}}(\mu)$ cannot be considered a realization of a white random process. Therefore, we introduce the novel standardized image

$$\hat{\mathbf{z}}_+^{(U)}(\mu) = \frac{\mathbf{y} - \hat{\boldsymbol{\lambda}} \mathbf{T}(\hat{\boldsymbol{\lambda}}(\mu))}{\sqrt{\hat{\boldsymbol{\lambda}}(\mu) \mathbf{V}(\hat{\boldsymbol{\lambda}}(\mu))}} \in \mathbb{R}^n,$$

where $\mathbf{T}(\boldsymbol{\lambda}) := (T(\lambda_1), \dots, T(\lambda_n))^T$ and $\mathbf{V}(\boldsymbol{\lambda}) := (V(\lambda_1), \dots, V(\lambda_n))^T$, with functions T and V defined in (29). The masked vector $\hat{\mathbf{z}}_+^{(U)} \in \mathbb{R}^n$ corresponding to the new standardization reads

$$\hat{\mathbf{z}}_+^{(U)}(\mu) = \left(\hat{z}_{+,1}^{(U)}(\mu), \dots, \hat{z}_{+,n}^{(U)}(\mu) \right)^T \quad \text{with} \quad \hat{z}_{+,i}^{(U)}(\mu) = \begin{cases} \hat{z}_i^{(U)}(\mu) & \text{if } i \in \mathcal{I}_+ \\ 0 & \text{otherwise} \end{cases}.$$

One can clearly observe that, in light of the results summarized in proposition 4.2, $\hat{\mathbf{z}}_+^{(U)}$ is a realization of a white random process, thus suggesting the following formulation for the masked unbiased version of the WP (WP-MU)

$$\text{Select } \mu = \hat{\mu} \in \mathbb{R}_{++} \left\{ \mathcal{W}_+^{(U)}(\mu) := \mathcal{W}(\hat{\mathbf{z}}_+^{(U)}(\mu)) \right\}, \quad (45)$$

with function $\mathcal{W}(\mathbf{z})$ defined as in (20).

4.6. Analysis of bias

In light of the introduced unbiased masked principles, in this section we carry out some analysis of the pixel-wise biases of the masked biased principles illustrated in section 3. To this purpose, first we give the following result.

Proposition. Let $Y \sim \mathcal{P}(\lambda)$ and $Y_+ \sim \mathcal{P}_+(\lambda)$, with $\lambda \in \mathbb{R}_{++}$, and let $Z_+^{(B)}$ and $Z_+^{(U)}$ be the biased and unbiased standardized positive Poisson random variables defined by

$$Z_+^{(B)} = \frac{Y_+ - \mathbf{E}[Y]}{\sqrt{\text{Var}[Y]}}, \quad Z_+^{(U)} = \frac{Y_+ - \mathbf{E}[Y_+]}{\sqrt{\text{Var}[Y_+]}}.$$

Then, it holds true that

$$\mathbf{E} \left[Z_+^{(B)} \right] = \sqrt{\lambda} (T(\lambda) - 1), \quad \mathbf{E} \left[\left(Z_+^{(B)} \right)^2 \right] = T(\lambda) - \lambda (T(\lambda) - 1), \quad (46)$$

$$\text{Var} \left[Z_+^{(B)} \right] = T(\lambda) (1 - \lambda (T(\lambda) - 1)), \quad (47)$$

$$\mathbf{E} \left[Z_+^{(U)} \right] = 0, \quad \mathbf{E} \left[\left(Z_+^{(U)} \right)^2 \right] = \text{Var} \left[Z_+^{(U)} \right] = 1. \quad (48)$$

Proof. First, the fact that $Z_+^{(U)}$ has zero mean and unitary second-order raw moment and variance comes in a straightforward manner from its definition.

Then, for what concern $Z_+^{(B)}$, after recalling that $\mathbf{E}[Y] = \text{Var}[Y] = \lambda$, it holds true that

$$\begin{aligned} \mathbf{E} \left[Z_+^{(B)} \right] &= \mathbf{E} \left[\frac{Y_+ - \mathbf{E}[Y]}{\sqrt{\text{Var}[Y]}} \right] = \mathbf{E} \left[\frac{Y_+ - \lambda}{\sqrt{\lambda}} \right] = \frac{1}{\sqrt{\lambda}} \mathbf{E}[Y_+ - \lambda] \\ &= \frac{1}{\sqrt{\lambda}} (\mathbf{E}[Y_+] - \lambda) = \sqrt{\lambda} (T(\lambda) - 1), \end{aligned} \quad (49)$$

where the last equality in (49) comes from replacing the expression of $\mathbf{E}[Y_+]$ given in (31). Then, by recalling also the expression of $\mathbf{E}[Y_+^2]$ in (32), we have that

$$\begin{aligned} \mathbf{E} \left[\left(Z_+^{(B)} \right)^2 \right] &= \mathbf{E} \left[\frac{(Y_+ - \mathbf{E}[Y])^2}{\text{Var}[Y]} \right] = \mathbf{E} \left[\frac{(Y_+ - \lambda)^2}{\lambda} \right] \\ &= \frac{1}{\lambda} \mathbf{E} [Y_+^2 - 2\lambda Y_+ + \lambda^2] = \frac{1}{\lambda} (\mathbf{E}[Y_+^2] - 2\lambda \mathbf{E}[Y_+] + \lambda^2) \\ &= \frac{1}{\lambda} (T(\lambda) (\lambda + \lambda^2) - 2\lambda^2 T(\lambda) + \lambda^2) \\ &= \frac{1}{\lambda} (\lambda T(\lambda) - \lambda^2 T(\lambda) + \lambda^2) = \lambda + T(\lambda) - \lambda T(\lambda) \\ &= T(\lambda) - \lambda (T(\lambda) - 1). \end{aligned} \quad (50)$$

Finally, based on (49) and (50), the variance in (47) can be computed as follows

$$\begin{aligned} \text{Var} \left[Z_+^{(B)} \right] &= \mathbf{E} \left[\left(Z_+^{(B)} \right)^2 \right] - \left(\mathbf{E} \left[Z_+^{(B)} \right] \right)^2 \\ &= T(\lambda) - \lambda (T(\lambda) - 1) - \lambda (T(\lambda) - 1)^2 \\ &= T(\lambda) - \lambda T(\lambda) + \lambda - \lambda T^2(\lambda) + 2\lambda T(\lambda) - \lambda \\ &= T(\lambda) + \lambda T(\lambda) - \lambda T^2(\lambda) = T(\lambda) (1 + \lambda - \lambda T(\lambda)) \\ &= T(\lambda) (1 - \lambda (T(\lambda) - 1)). \end{aligned} \quad (51)$$

□

In what follows, we compare the masked biased and masked unbiased versions of the QDP, NEDP and WP in terms of some pixel-based bias functions. We recall that for the ADP the bias has to be considered constantly zero as the masked biased and the masked unbiased versions coincide.

As the definition of the QDP, in its unmasked, masked biased and masked unbiased version, is related to the behavior of the sample variance of the noise realization vector approximated by the residual image, we introduce the following bias function $\beta_{\text{QDP}} : \mathbb{R}_{++} \rightarrow \mathbb{R}$ to measure the inaccuracy introduced by the QDP-MB at each pixel

$$\beta_{\text{QDP}}(\lambda) = \mathbb{E} \left[\left(Z_+^{(B)} \right)^2 \right] - \mathbb{E} \left[\left(Z_+^{(U)} \right)^2 \right] = T(\lambda) - \lambda(T(\lambda) - 1) - 1. \quad (52)$$

In the case of NEDP, the bias can be measured in terms of the difference between the building functions used to approximate the behavior of the exact expected value $\delta^{(E)}(\lambda)$ arising in the EDP (13). We thus introduce the pixel-based bias function $\beta_{\text{NEDP}} : \mathbb{R}_{++} \rightarrow \mathbb{R}$ which is defined as

$$\beta_{\text{NEDP}}(\lambda) = \epsilon(\lambda) - \epsilon_U(\lambda) \quad (53)$$

with $\epsilon(\lambda), \epsilon_U(\lambda)$ given in (18), (44), respectively.

Finally, for what concerns the WP, we point out that measuring the bias in terms of the auto-correlation of the normalized random variables $Z_+^{(B)}, Z_+^{(U)}$ —that would be the most natural choice in this scenario—is unfeasible; hence we rather measure how far is $Z_+^{(B)}$ from being a zero-mean random variable with constant (unitary) standard deviation. In other words, we introduce the two pixel-based bias functions $\beta_{\text{WP},\eta}, \beta_{\text{WP},\sigma} : \mathbb{R}_{++} \rightarrow \mathbb{R}$ defined as

$$\begin{aligned} \beta_{\text{WP},\eta}(\lambda) &= \mathbb{E} \left[Z_+^{(B)} \right] - \mathbb{E} \left[Z_+^{(U)} \right] = \sqrt{\lambda} (T(\lambda) - 1), \\ \beta_{\text{WP},\sigma}(\lambda) &= \sqrt{\text{Var} \left(Z_+^{(B)} \right)} - \sqrt{\text{Var} \left(Z_+^{(U)} \right)} = \sqrt{T(\lambda) (1 - \lambda (T(\lambda) - 1))} - 1. \end{aligned} \quad (54)$$

In figure 2, we show the behavior of the pixel-based bias functions in (52)–(54) for three different ranges of λ , namely $(0, 0.1]$, $(0.1, 5]$, $(5, 200]$. Notice that the bias is particularly relevant for very small values of λ , while it tends to 0 as λ increases. Such behavior reflects the theoretical results given in proposition 4, whence we have that the statistical properties of $Z_+^{(B)}, Z_+^{(U)}$ coincide as $\lambda \rightarrow +\infty$.

5. Numerical methods for application of the selection principles

The application of all previously reviewed and newly proposed selection criteria rely on numerically solving the \mathcal{R} -KL variational model in (8) which, we remark, is convex under the assumption that the regularization term \mathcal{R} is convex (this is the case of (5)–(7) regularizers). Efficient solvers based on the ADMM [8] for the \mathcal{R} -KL model when \mathcal{R} is the TV, TV^2 or TGV^2 regularizer—that is, for the so-called TV-KL, TV^2 -KL and TGV^2 -KL models—have been proposed, e.g. in [6, 16, 20], respectively. However, for the sake of completeness of the presentation as well as of reproducibility of the results that we will present in the next section, in this section we recall the main concepts and computational steps of the ADMM applied to the three models. In particular, we are going to show how the TV-KL, TV^2 -KL and TGV^2 -KL models can all be equivalently reformulated as two-blocks separable optimization problems with linear constraints, which can be solved by standard two-blocks ADMM schemes with guaranteed convergence.

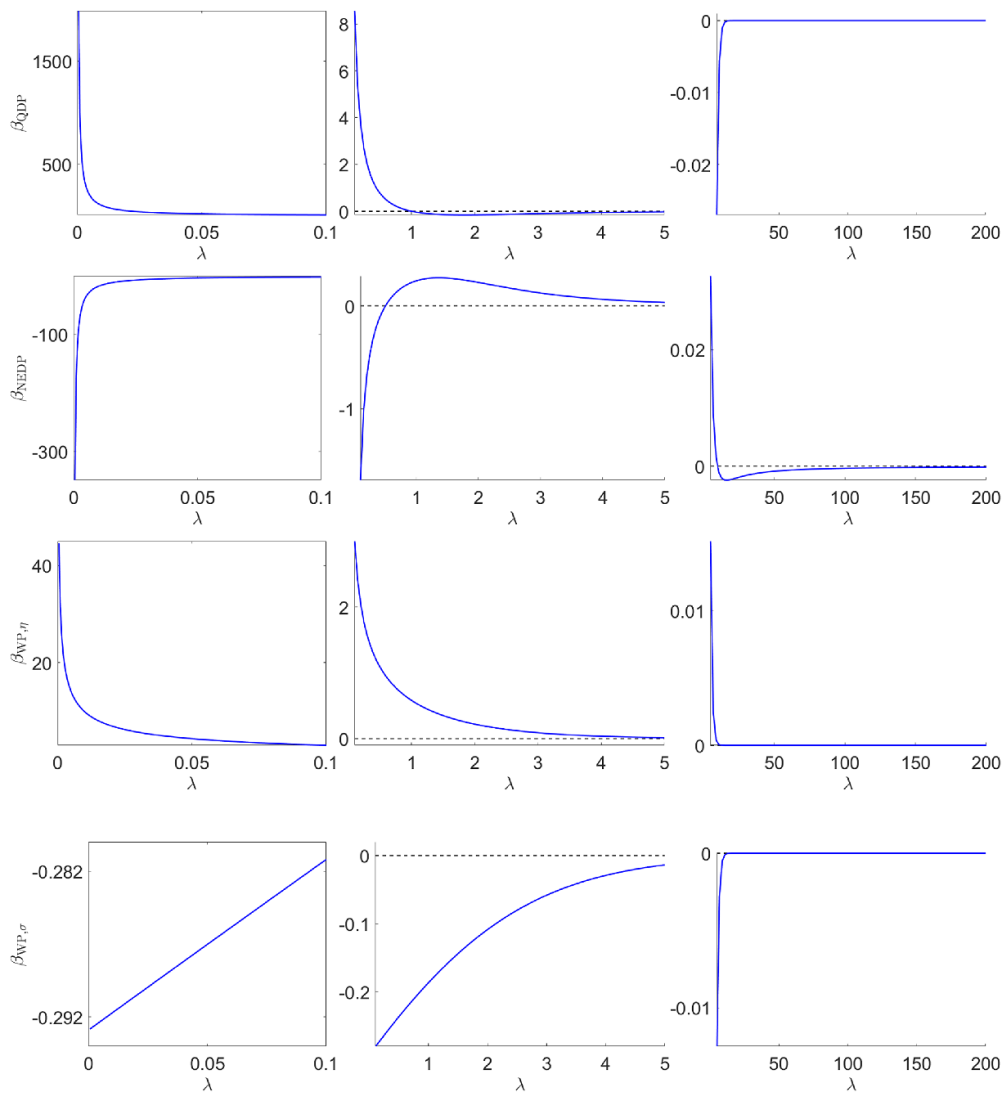


Figure 2. From top to bottom: plot of the pixel-based bias functions modeling the approximation introduced by the masked biased version of the QDP, NEDP and WP on different λ intervals in the range $(0, 200]$.

Recalling the definitions of the \mathcal{R} -KL model in (8) and of the TV, TV^2 and TGV^2 regularizers in (5)–(7), and by introducing the three matrices

$$\mathbf{D}_1 = \begin{pmatrix} \mathbf{D}_h \\ \mathbf{D}_v \end{pmatrix} \in \mathbb{R}^{2n \times n}, \quad \mathbf{D}_2 = \begin{pmatrix} \mathbf{D}_{hh} \\ \mathbf{D}_{vv} \\ \mathbf{D}_{hv} \\ \mathbf{D}_{vh} \end{pmatrix} \in \mathbb{R}^{4n \times n}, \quad \mathbf{D}_S = \begin{pmatrix} \mathbf{D}_h & 0 \\ 0 & \mathbf{D}_v \\ \frac{1}{2}\mathbf{D}_v & \frac{1}{2}\mathbf{D}_h \\ \frac{1}{2}\mathbf{D}_v & \frac{1}{2}\mathbf{D}_h \end{pmatrix} \in \mathbb{R}^{4n \times 2n},$$

with $\mathbf{D}_h, \mathbf{D}_v, \mathbf{D}_{hh}, \mathbf{D}_{vv}, \mathbf{D}_{hv}, \mathbf{D}_{vh} \in \mathbb{R}^{n \times n}$ finite difference matrices discretizing the first-order partial derivatives of the vectorized $n_1 \times n_2$ image \mathbf{x} in the horizontal and vertical direction and

the second-order partial derivatives of image \mathbf{x} in the horizontal, vertical, mixed horizontal-vertical and mixed vertical-horizontal directions (with $\mathbf{D}_{vh} = \mathbf{D}_{hv}$), respectively, the TV-KL, TV²-KL and TGV²-KL models can be equivalently written as

$$\hat{\mathbf{x}}(\mu) \in \arg \min_{\mathbf{x} \in \mathbb{R}^n} \left\{ \iota_{\mathbb{R}_+^n}(\mathbf{x}) + \mu \text{KL}(\mathbf{H}\mathbf{x} + \mathbf{b}; \mathbf{y}) + \sum_{i \in \mathcal{I}} \|(\mathbf{D}_1 \mathbf{x})_i\|_2 \right\}, \quad (55)$$

$$\hat{\mathbf{x}}(\mu) \in \arg \min_{\mathbf{x} \in \mathbb{R}^n} \left\{ \iota_{\mathbb{R}_+^n}(\mathbf{x}) + \mu \text{KL}(\mathbf{H}\mathbf{x} + \mathbf{b}; \mathbf{y}) + \sum_{i \in \mathcal{I}} \|(\mathbf{D}_2 \mathbf{x})_i\|_2 \right\}, \quad (56)$$

$$\begin{aligned} \{\hat{\mathbf{x}}(\mu), \hat{\mathbf{w}}(\mu)\} \in \arg \min_{\substack{\mathbf{x} \in \mathbb{R}^n, \\ \mathbf{w} \in \mathbb{R}^{2n}}} & \left\{ \iota_{\mathbb{R}_+^n}(\mathbf{x}) + \mu \text{KL}(\mathbf{H}\mathbf{x} + \mathbf{b}; \mathbf{y}) \right. \\ & \left. + \alpha_0 \sum_{i \in \mathcal{I}} \|(\mathbf{D}_1 \mathbf{x})_i - \mathbf{w}_i\|_2 + \alpha_1 \sum_{i \in \mathcal{I}} \|(\mathbf{D}_S \mathbf{w})_i\|_2 \right\}, \quad (57) \end{aligned}$$

respectively, where $\iota_{\mathbb{R}_+^n}(\mathbf{x})$ denotes the indicator function of the non-negative orthant \mathbb{R}_+^n and where, with a little abuse of notation, we indicate by $(\mathbf{D}_1 \mathbf{x})_i := ((\mathbf{D}_h \mathbf{x})_i; (\mathbf{D}_v \mathbf{x})_i) \in \mathbb{R}^2$ and by $(\mathbf{D}_2 \mathbf{x})_i := ((\mathbf{D}_{hh} \mathbf{x})_i; (\mathbf{D}_{vv} \mathbf{x})_i; (\mathbf{D}_{hv} \mathbf{x})_i; (\mathbf{D}_{vh} \mathbf{x})_i) \in \mathbb{R}^4$ the discrete gradient and the vectorized discrete Hessian of image \mathbf{x} at pixel i , respectively. Moreover, we denote by $(\mathbf{D}_S \mathbf{w})_i := ((\mathbf{D}_h \mathbf{w}_1)_i; (\mathbf{D}_v \mathbf{w}_2)_i; (1/2)(\mathbf{D}_v \mathbf{w}_1)_i + (1/2)(\mathbf{D}_h \mathbf{w}_2)_i; (1/2)(\mathbf{D}_v \mathbf{w}_1)_i + (1/2)(\mathbf{D}_h \mathbf{w}_2)_i) \in \mathbb{R}^4$ the vectorized discrete symmetric Jacobian of vector field $\mathbf{w} = (\mathbf{w}_1; \mathbf{w}_2)$ at pixel i .

Then, by introducing for the models in (55)–(57) the auxiliary variable \mathbf{u} defined in the three cases, respectively, by

$$\mathbf{u} = \begin{pmatrix} \mathbf{u}_1 \\ \mathbf{u}_2 \\ \mathbf{u}_3 \end{pmatrix} = \begin{pmatrix} \mathbf{x} \\ \mathbf{H}\mathbf{x} \\ \mathbf{D}_1 \mathbf{x} \end{pmatrix}, \quad \mathbf{u} = \begin{pmatrix} \mathbf{u}_1 \\ \mathbf{u}_2 \\ \mathbf{u}_3 \end{pmatrix} = \begin{pmatrix} \mathbf{x} \\ \mathbf{H}\mathbf{x} \\ \mathbf{D}_2 \mathbf{x} \end{pmatrix}, \quad \mathbf{u} = \begin{pmatrix} \mathbf{u}_1 \\ \mathbf{u}_2 \\ \mathbf{u}_3 \\ \mathbf{u}_4 \end{pmatrix} = \begin{pmatrix} \mathbf{x} \\ \mathbf{H}\mathbf{x} \\ \mathbf{D}_1 \mathbf{x} - \mathbf{w} \\ \mathbf{D}_S \mathbf{w} \end{pmatrix},$$

and setting $\mathbf{t} = \mathbf{x}$ for TV-KL and TV²-KL, $\mathbf{t} = (\mathbf{x}; \mathbf{w})$ for TGV²-KL, it is easy to verify that all the three models can be equivalently reformulated as the following standard two-blocks (additively) separable minimization problem with linear constraints:

$$\{\hat{\mathbf{t}}(\mu), \hat{\mathbf{u}}(\mu)\} \in \arg \min_{\mathbf{t}, \mathbf{u}} \{ C_1(\mathbf{t}) + C_2(\mathbf{u}; \mu) \} \quad \text{subject to: } \mathbf{M}_1 \mathbf{t} + \mathbf{M}_2 \mathbf{u} = \mathbf{0}. \quad (58)$$

In particular, for TV-KL, TV²-KL and TGV²-KL models, functions C_1, C_2 read, respectively,

$$C_1(\mathbf{t}) = 0, \quad C_2(\mathbf{u}; \mu) = \iota_{\mathbb{R}_+^n}(\mathbf{u}_1) + \mu \text{KL}(\mathbf{u}_2 + \mathbf{b}; \mathbf{y}) + \sum_{i \in \mathcal{I}} \|\mathbf{u}_{3,i}\|_2, \quad (59)$$

$$C_1(\mathbf{t}) = 0, \quad C_2(\mathbf{u}; \mu) = \iota_{\mathbb{R}_+^n}(\mathbf{u}_1) + \mu \text{KL}(\mathbf{u}_2 + \mathbf{b}; \mathbf{y}) + \sum_{i \in \mathcal{I}} \|\mathbf{u}_{3,i}\|_2, \quad (60)$$

$$C_1(\mathbf{t}) = 0, \quad C_2(\mathbf{u}; \mu) = \iota_{\mathbb{R}_+^n}(\mathbf{u}_1) + \mu \text{KL}(\mathbf{u}_2 + \mathbf{b}; \mathbf{y}) + \alpha_0 \sum_{i \in \mathcal{I}} \|\mathbf{u}_{3,i}\|_2 + \alpha_1 \sum_{i \in \mathcal{I}} \|\mathbf{u}_{4,i}\|_2, \quad (61)$$

and matrices $\mathbf{M}_1, \mathbf{M}_2$ take the form, respectively,

$$\mathbf{M}_1 = \begin{pmatrix} \mathbf{I} \\ \mathbf{H} \\ \mathbf{D}_1 \end{pmatrix}, \mathbf{M}_2 = -\mathbf{I}, \mathbf{M}_1 = \begin{pmatrix} \mathbf{I} \\ \mathbf{H} \\ \mathbf{D}_2 \end{pmatrix}, \mathbf{M}_2 = -\mathbf{I}, \mathbf{M}_1 = \begin{pmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{H} & \mathbf{0} \\ \mathbf{D}_1 & -\mathbf{I} \\ \mathbf{0} & \mathbf{D}_s \end{pmatrix}, \mathbf{M}_2 = -\mathbf{I}. \quad (62)$$

The Lagrangian function \mathcal{L} and augmented Lagrangian function \mathcal{L}_γ associated with problem (58) read

$$\mathcal{L}(\mathbf{t}, \mathbf{u}, \boldsymbol{\rho}; \mu) = C_1(\mathbf{t}) + C_2(\mathbf{u}; \mu) + \boldsymbol{\rho}^\top (\mathbf{M}_1 \mathbf{t} + \mathbf{M}_2 \mathbf{u}), \quad (63)$$

$$\mathcal{L}_\gamma(\mathbf{t}, \mathbf{u}, \boldsymbol{\rho}; \mu) = \mathcal{L}(\mathbf{t}, \mathbf{u}, \boldsymbol{\rho}; \mu) + \frac{\gamma}{2} \|\mathbf{M}_1 \mathbf{t} + \mathbf{M}_2 \mathbf{u}\|_2^2, \quad (64)$$

where $\boldsymbol{\rho}$ is the vector of Lagrange multipliers associated to the system of linear constraints in (58) and $\gamma \in \mathbb{R}_{++}$ is a penalty parameter.

Solving problem (58) amounts to seek the saddle point(s) $\{\mathbf{t}^*(\mu), \mathbf{u}^*(\mu), \boldsymbol{\rho}^*(\mu)\}$ of the augmented Lagrangian \mathcal{L}_γ in (64) which, according to the standard two-blocks ADMM [8], can be computed as the limit point of the following iterative procedure:

$$\begin{aligned} \mathbf{t}^{(k+1)} &= \arg \min_{\mathbf{t}} \mathcal{L}_\gamma(\mathbf{t}, \mathbf{u}^{(k)}, \boldsymbol{\rho}^{(k)}; \mu) \\ &= \arg \min_{\mathbf{t}} \left\{ C_1(\mathbf{t}) + \frac{\gamma}{2} \left\| \mathbf{M}_1 \mathbf{t} + \mathbf{M}_2 \mathbf{u}^{(k)} + \frac{1}{\gamma} \boldsymbol{\rho}^{(k)} \right\|_2^2 \right\}, \end{aligned} \quad (65)$$

$$\begin{aligned} \mathbf{u}^{(k+1)} &= \arg \min_{\mathbf{u}} \mathcal{L}_\gamma(\mathbf{t}^{(k+1)}, \mathbf{u}, \boldsymbol{\rho}^{(k)}; \mu) \\ &= \arg \min_{\mathbf{u}} \left\{ C_2(\mathbf{u}) + \frac{\gamma}{2} \left\| \mathbf{M}_2 \mathbf{u} + \mathbf{M}_1 \mathbf{t}^{(k+1)} + \frac{1}{\gamma} \boldsymbol{\rho}^{(k)} \right\|_2^2 \right\}, \end{aligned} \quad (66)$$

$$\boldsymbol{\rho}^{(k+1)} = \boldsymbol{\rho}^{(k)} + \gamma (\mathbf{M}_1 \mathbf{t}^{(k+1)} + \mathbf{M}_2 \mathbf{u}^{(k+1)}). \quad (67)$$

After recalling from (59)–(62) that $C_1(\mathbf{t}) = 0$ and $\mathbf{M}_2 = -\mathbf{I}$, it is immediate to prove that for the three models the \mathbf{t} -update step in (65) takes the same following form

$$\mathbf{t}^{(k+1)} = (\mathbf{M}_1^\top \mathbf{M}_1)^{-1} \mathbf{M}_1^\top \mathbf{q}^{(k)}, \quad \mathbf{q}^{(k)} = \mathbf{u}^{(k)} - \frac{1}{\gamma} \boldsymbol{\rho}^{(k)}, \quad (68)$$

that is, $\mathbf{t}^{(k+1)}$ is obtained by solving a linear system with coefficient matrix $\mathbf{M}_1^\top \mathbf{M}_1$. This matrix is symmetric and positive definite—hence, non-singular—in all the three cases and, by assuming periodic boundary conditions for all the involved finite difference matrices, the linear system can be solved very efficiently based on the 2D discrete Fourier transform, implemented by 2D fast Fourier transform (see, e.g. [6, 16, 20]). We note that $\mathbf{t}^{(k+1)} = \mathbf{x}^{(k+1)}$ for TV-KL and TV²-KL, $\mathbf{t}^{(k+1)} = (\mathbf{x}^{(k+1)}; \mathbf{w}^{(k+1)})$ for TGV²-KL.

For what regards the \mathbf{u} -update in (66), it is easy to verify that it takes the form

$$\mathbf{u}^{(k+1)} = \arg \min_{\mathbf{u}} \left\{ \sum_{j=1}^m U_j(\mathbf{u}_j) \right\} \iff \mathbf{u}_j^{(k+1)} = \arg \min_{\mathbf{u}_j} U_j(\mathbf{u}_j), \quad j = 1, \dots, m,$$

with $m = 3$ for TV-KL and TV²-KL models, $m = 4$ for TGV²-KL. This means that (66) is equivalent to 3 or 4 independent minimization subproblems each giving the updated value of one of the solution subvectors \mathbf{u}_j . In order to outline the solution of all the m subproblems for the three models, it is useful to introduce a partition of the vector of Lagrange multipliers $\boldsymbol{\rho}^{(k)}$ into m subvectors $\boldsymbol{\rho}_j^{(k)}$ having the same size of the corresponding solution subvectors \mathbf{u}_j .

Hence, for TV-KL and TV²-KL we define $\boldsymbol{\rho}^{(k)} = (\boldsymbol{\rho}_1^{(k)}; \boldsymbol{\rho}_2^{(k)}; \boldsymbol{\rho}_3^{(k)})$, whereas for TGV²-KL we set $\boldsymbol{\rho}^{(k)} = (\boldsymbol{\rho}_1^{(k)}; \boldsymbol{\rho}_2^{(k)}; \boldsymbol{\rho}_3^{(k)}; \boldsymbol{\rho}_4^{(k)})$.

For all the three models the two subproblems for variables $\mathbf{u}_1, \mathbf{u}_2 \in \mathbb{R}^n$ admit the same pixel-wise close-form solutions which, after introducing the two vectors

$$\mathbf{q}_1^{(k)} = \mathbf{x}^{(k+1)} + \frac{1}{\gamma} \boldsymbol{\rho}_1^{(k)}, \quad \mathbf{q}_2^{(k)} = \mathbf{H}\mathbf{x}^{(k+1)} + \frac{1}{\gamma} \boldsymbol{\rho}_2^{(k)},$$

and setting $\tau = \mu/\gamma$, read, respectively,

$$\mathbf{u}_{1,i}^{(k+1)} = \max \left\{ q_{1,i}^{(k)}, 0 \right\}, \quad (69)$$

$$\mathbf{u}_{2,i}^{(k+1)} = \frac{1}{2} \left(-(\tau + b_i - q_{2,i}) + \sqrt{(\tau + b_i - q_{2,i})^2 + 4(b_i q_{2,i} + \tau(y_i - b_i))} \right), \quad (70)$$

for $i = 1, \dots, n$.

Also the third subproblem for variable \mathbf{u}_3 , after introducing the vector

$$\mathbf{q}_3^{(k)} = \begin{cases} \mathbf{D}_1 \mathbf{x}^{(k+1)} + \frac{1}{\gamma} \boldsymbol{\rho}_3^{(k)} \in \mathbb{R}^{2n} & \text{for TV - KL model,} \\ \mathbf{D}_2 \mathbf{x}^{(k+1)} + \frac{1}{\gamma} \boldsymbol{\rho}_3^{(k)} \in \mathbb{R}^{4n} & \text{for TV}^2 \text{ - KL model,} \\ \mathbf{D}_1 \mathbf{x}^{(k+1)} + \frac{1}{\gamma} \boldsymbol{\rho}_3^{(k)} - \mathbf{w}^{(k+1)} \in \mathbb{R}^{2n} & \text{for TGV}^2 \text{ - KL model,} \end{cases}$$

admits a pixel-wise explicit solution (proximal map of ℓ_2 -norm function) reading

$$\mathbf{u}_{3,i}^{(k+1)} = \max \left\{ 1 - \frac{1}{\gamma \|\mathbf{q}_{3,i}^{(k)}\|_2}, 0 \right\} \mathbf{q}_{3,i}^{(k)},$$

for $i = 1, \dots, n$, where $\max\{1 - 1/0, 0\} = 0$ is assumed and where $\mathbf{u}_{3,i}^{(k+1)}, \mathbf{q}_{3,i}^{(k)} \in \mathbb{R}^2$ for TV-KL and TGV²-KL models, $\mathbf{u}_{3,i}^{(k+1)}, \mathbf{q}_{3,i}^{(k)} \in \mathbb{R}^4$ for TV²-KL.

Finally, for the TGV²-KL model, the fourth subproblem for variable $\mathbf{u}_4 \in \mathbb{R}^{4n}$ can also be solved in pixel-wise closed-form based on the ℓ_2 -norm proximal map; in formula,

$$\mathbf{u}_{4,i}^{(k+1)} = \max \left\{ 1 - \frac{1}{\gamma \|\mathbf{q}_{4,i}^{(k)}\|_2}, 0 \right\} \mathbf{q}_{4,i}^{(k)}, \quad \mathbf{q}_4^{(k)} = \mathbf{D}_S \mathbf{w}^{(k+1)} + \frac{1}{\gamma} \boldsymbol{\rho}^{(k)}, \quad (71)$$

for $i = 1, \dots, n$, with $\mathbf{u}_{4,i}^{(k+1)}, \mathbf{q}_{4,i}^{(k)} \in \mathbb{R}^4$.

To conclude, in the following proposition 5.1 we apply to the three ADMM schemes outlined above (solving the TV-KL, TV²-KL and TGV²-KL models) a general and classical convergence result for the two-blocks ADMM given in the seminal paper by Eckstein and Bertsekas [11].

Proposition 5.1. *Let $\mathcal{L}, \mathcal{L}_\gamma$ be the Lagrangian and the augmented Lagrangian functions in (63) and (64) with functions C_1, C_2 and matrices $\mathbf{M}_1, \mathbf{M}_2$ defined as in (59)–(62), depending on the regularizer (TV, TV² or TGV²) considered. Then, if a saddle point $\{\mathbf{t}^*(\mu), \mathbf{u}^*(\mu), \boldsymbol{\rho}^*(\mu)\}$ of \mathcal{L} exists, the ADMM scheme in (65)–(67) with (65) and (66) solved as in (68)–(71) converges to $\{\mathbf{t}^*(\mu), \mathbf{u}^*(\mu), \boldsymbol{\rho}^*(\mu)\}$ for any $\mu, \gamma \in \mathbb{R}_{++}$, with $\mathbf{t}^*(\mu)$ equal to a solution $\hat{\mathbf{x}}(\mu)$ of model (55) or model (56), or to a solution $\{\hat{\mathbf{x}}(\mu), \hat{\mathbf{w}}(\mu)\}$ of model (57). If a saddle point of \mathcal{L} does not exist, then at least one of the two sequences of ADMM iterates $\{\mathbf{u}^{(k)}\}$ or $\{\boldsymbol{\rho}^{(k)}\}$ is unbounded.*

Proof. For the TV-KL, TV²-KL and TGV²-KL variational models considered, for any $\mu \in \mathbb{R}_{++}$ the functions C_1, C_2 defined in (59)–(61) are all clearly proper, closed and convex and the matrices $\mathbf{M}_1, \mathbf{M}_2$ defined in (62) are such that \mathbf{M}_2 is the negative identity and \mathbf{M}_1 has full

(column) rank. Moreover, the two ADMM minimization subproblems for the primal variables \mathbf{t}, \mathbf{u} in (65) and (66) are solved exactly by formulas in (68)–(71). Hence, by applying the classical convergence result for the two-blocks ADMM given in [11] (Theorem 8), the proof of the statement follows easily. \square

6. Computed examples

In this section, we perform a reliable quantitative experimental comparison among the performance of the eleven different parameter selection criteria outlined in section 2 (unmasked), section 3 (masked biased) and section 4 (masked unbiased), namely the four unmasked principles ADP in (14), QDP in (17), NEDP in (19), WP in (20) proposed in [2, 5, 6, 26], respectively, the two masked biased criteria ADP-MB in (22), QDP-MB in (23) presented in [9], the two newly proposed masked biased principles NEDP-MB in (24), WP-MB in (26) and the three novel masked unbiased criteria QDP-MU in (42), NEDP-MU in (43) and WP-MU in (45).

In order to make the results of the comparison as solid as possible, we act in three directions. First, we consider the three test images `satellite`, `stem` and `cells` shown in figure 3—with the associated experiments reported in sections 6.3, 6.4 and 6.5, respectively—which are characterized by different properties and thus allow to test the eleven selection criteria for the three (55–57) image restoration models. Second, for each test image/model we simulate different photon-counting scenarios, ranging from very low- to high-counting ones. Third, for each image/model and each photon-counting level, we consider a number of different (independent) Poisson noise realizations and collect statistics (minimums, maximums and averages) of the quantitative accuracy results achieved by the principles. In particular, we measure the quality of the restored images $\hat{\mathbf{x}}(\mu)$ (with respect to the target uncorrupted image $\bar{\mathbf{x}}$) obtained by applying the different criteria by means of two accuracy metrics, namely the structural similarity index measure (SSIM) defined in [25] and computed by the Matlab routine `ssim` and the improved-signal-to-noise-ratio (ISNR) defined by

$$\begin{aligned} \text{ISNR}(\hat{\mathbf{x}}(\mu), \bar{\mathbf{x}}, \mathbf{y}) &= \text{SNR}(\hat{\mathbf{x}}(\mu), \bar{\mathbf{x}}) - \text{SNR}(\mathbf{y}, \bar{\mathbf{x}}), \\ \text{SNR}(\mathbf{z}, \bar{\mathbf{x}}) &= 20 \log_{10} \frac{\|\bar{\mathbf{x}} - m_{\bar{\mathbf{x}}}\|_2}{\|\bar{\mathbf{x}} - \mathbf{z}\|_2}, \end{aligned}$$

measured in decibel and with $m_{\bar{\mathbf{x}}}$ denoting the mean value of the target image $\bar{\mathbf{x}}$.

All tests have been performed in Matlab R2022b, on a Windows 10 Platform. The code is available at <https://github.com/MonicaPragliola/MU-principles>.

6.1. Data generation

For each of the three test images, in order to simulate different photon-counting scenarios, first we (affinely) scale the image in the range $[0, 1]$ and denote by $\bar{\mathbf{x}}_{\text{norm}}$ the obtained normalized image. Then, nine uncorrupted images $\bar{\mathbf{x}}^{\kappa}$ for nine different photon-counting levels are simulated by multiplying $\bar{\mathbf{x}}_{\text{norm}}$ by as many photon-level factors $\kappa \in \mathbb{R}_{++}$, in particular:

$$\bar{\mathbf{x}}^{\kappa} = \kappa \bar{\mathbf{x}}_{\text{norm}}, \quad \kappa \in \{1, 3, 5, 10, 20, 50, 100, 500, 1000\}. \quad (72)$$

The photon-scaled images $\bar{\mathbf{x}}^{\kappa}$ are then corrupted by space-invariant Gaussian blur, with blur kernel generated by the Matlab routine `fspecial`, which is characterized by two parameters: the `band` parameter, representing the side length (in pixels) of the square support of the kernel, and `sigma`, that is the standard deviation (in pixels) of the isotropic bivariate Gaussian distribution defining the kernel in the continuous setting. In all our tests, we set `band` = 5, `sigma` =

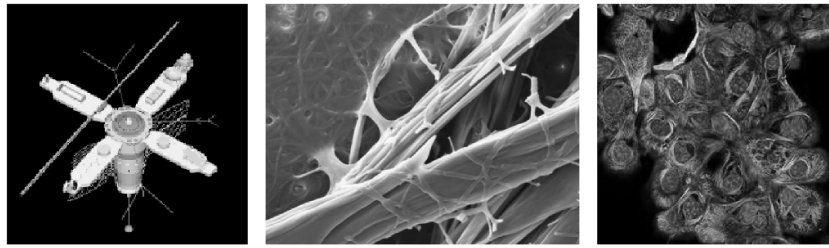


Figure 3. From left to right: original `satellite` (256×256 pixels), `stem` (453×592 pixels) and `cells` (236×236 pixels) test images considered in the numerical experiments.

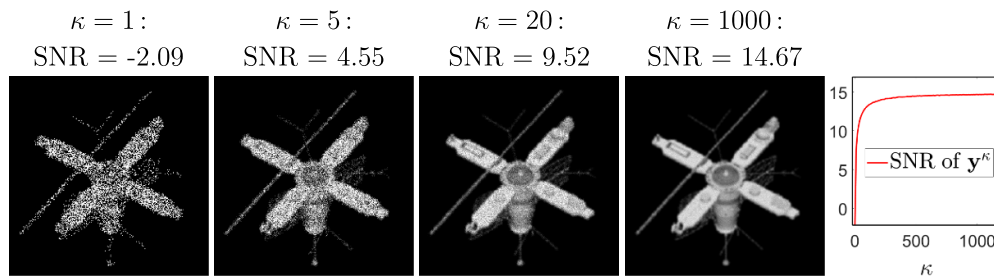


Figure 4. Test image `satellite` corrupted by blur and, from left to right, by decreasing levels of Poisson noise yielded by increasing values of the photon-level factor κ in (72). In the right-most graph, SNR value of the noisy observation \mathbf{y}^κ as a function of factor κ .

1. Then, a constant background emission image \mathbf{b} is added to the blurred images, so as to get the nine noise-free degraded images $\bar{\lambda}^\kappa = \mathbf{H}\bar{\mathbf{x}}^{(k)} + \mathbf{b}$. Finally, for each noise-free image $\bar{\lambda}^\kappa$, ten different noisy observations

$$\mathbf{y}^\kappa(j), \quad j = 1, \dots, 10, \quad (73)$$

are generated by sampling as many independent realizations from an n -variate Poisson random process with mean $\bar{\lambda}^\kappa$, using the Matlab routine `poissrnd`.

We remark that the factor κ in (72) represents the maximum number of photons that, on average, can hit any pixel of the image domain if no blur degradation ($\mathbf{H} = \mathbf{I}$) and a null emission background ($\mathbf{b} = \mathbf{0}$) are considered. In fact, in this case the noise-free image $\bar{\lambda}^\kappa$ —which, we notice, contains the mean values of the Poisson noise distributions at all pixels—is given by $\bar{\lambda}^\kappa = \mathbf{H}\bar{\mathbf{x}}^\kappa + \mathbf{b} = \bar{\mathbf{x}}^\kappa$, hence $\max\{\bar{\lambda}^\kappa\} = \max\{\bar{\mathbf{x}}^\kappa\} = \kappa$. In general, for any given blur corruption and emission background, the factor κ is positively related to the photon-range of the experiment and, recalling that for a scalar Poisson random variable with parameter $\bar{\lambda}^\kappa = \kappa\bar{\lambda}$ the ratio between its mean (true signal) and its standard deviation (noise level) is equal to $\bar{\lambda}^\kappa / \sqrt{\bar{\lambda}^\kappa} = \sqrt{\kappa} \sqrt{\bar{\lambda}}$, also to the signal-to-noise ratio of the observed degraded image \mathbf{y}^κ to restore. To highlight clearly the effect of κ on the noise-level in the observation \mathbf{y}^κ and, hence, on the difficulty of the restoration process, in figure 4 we show the test image `satellite` corrupted by Gaussian blur and by a realization of Poisson noise for four different values of κ as well as, on the right, the graph of the SNR value of the observation \mathbf{y}^κ as a function of the factor κ . This graph justifies the non-uniform grid of κ -values considered in (72) (the grid is finer for small κ -values where the SNR changes more rapidly) as well as the maximum

value $k = 1000$ considered (the SNR curve stabilizes, hence taking $\kappa > 1000$ does not change significantly the results of the criteria comparison).

6.2. Analysis

For each of the three test images and each of the nine photon-level factors κ in (72), the ten generated degraded images $\mathbf{y}^\kappa(j)$ are processed as follows. For each $\mathbf{y}^\kappa(j)$, we compute the solution of the \mathcal{R} -KL variational model for a very fine grid of different μ -values and then, based on the obtained μ -dependent restorations, we apply the eleven different criteria to get the selected regularization parameter values

$$\hat{\mu}_C^\kappa(j), \quad C \in \mathcal{C} := \{\text{ADP, QDP, NEDP, WP, ADP-M, QDP-MB, NEDP-MB, WP-MB, QDP-MU, NEDP-MU, WP-MU}\},$$

and the corresponding restored images $\hat{\mathbf{x}}(\hat{\mu}_C^\kappa(j))$. We then compute and record the associated ISNR and SSIM values denoted by $\iota_C^\kappa(j)$ and $\sigma_C^\kappa(j)$, respectively, as well as the optimal (i.e. maximum) ISNR and SSIM values achieved on the fine grid of μ -values considered, denoted by $\iota_{\text{OPT}}^\kappa(j)$ and $\sigma_{\text{OPT}}^\kappa(j)$, respectively.

After processing the ten degraded observations $\mathbf{y}^\kappa(j)$, we thus get the following sets of quantitative results:

$$\begin{aligned} I_C^\kappa &:= \{\iota_C^\kappa(1), \dots, \iota_C^\kappa(10)\}, & S_C^\kappa &:= \{\sigma_C^\kappa(1), \dots, \sigma_C^\kappa(10)\}, & C \in \mathcal{C}, \\ I_{\text{OPT}}^\kappa &:= \{\iota_{\text{OPT}}^\kappa(1), \dots, \iota_{\text{OPT}}^\kappa(10)\}, & S_{\text{OPT}}^\kappa &:= \{\sigma_{\text{OPT}}^\kappa(1), \dots, \sigma_{\text{OPT}}^\kappa(10)\}. \end{aligned}$$

Then, for each $\iota_C^\kappa(j) \in I_C^\kappa$ and each $\sigma_C^\kappa(j) \in S_C^\kappa$ we compute the percentage difference with respect to the corresponding optimal values $\iota_{\text{OPT}}^\kappa(j)$ and $\sigma_{\text{OPT}}^\kappa(j)$, respectively,

$$\epsilon_{I_C^\kappa}(j) := 100 \times \frac{\iota_{\text{OPT}}^\kappa(j) - \iota_C^\kappa(j)}{\iota_{\text{OPT}}^\kappa(j)}, \quad \epsilon_{S_C^\kappa}(j) := 100 \times \frac{\sigma_{\text{OPT}}^\kappa(j) - \sigma_C^\kappa(j)}{\sigma_{\text{OPT}}^\kappa(j)}.$$

The behavior of each selection criterion for a given photon-counting level is thus synthesized by the expected values (or, better, sample means) of the ISNR and SSIM percentage errors achieved for the ten different noise realizations,

$$\eta_{I_C^\kappa} := \frac{1}{10} \sum_{j=1}^{10} \epsilon_{I_C^\kappa}(j), \quad \eta_{S_C^\kappa} := \frac{1}{10} \sum_{j=1}^{10} \epsilon_{S_C^\kappa}(j). \quad (74)$$

Moreover, to monitor the variability of the performance of each criterion with respect to different noise realizations, we also compute

$$\bar{\epsilon}_{I_C^\kappa} = \max_j \epsilon_{I_C^\kappa}(j), \quad \bar{\epsilon}_{S_C^\kappa} = \max_j \epsilon_{S_C^\kappa}(j), \quad \underline{\epsilon}_{I_C^\kappa} = \min_j \epsilon_{I_C^\kappa}(j), \quad \underline{\epsilon}_{S_C^\kappa} = \min_j \epsilon_{S_C^\kappa}(j). \quad (75)$$

We remark that, for all the experiments, restoration is performed by means of the \mathcal{R} -KL variational model—in particular, TV-KL model for satellite image, TV²-KL for stem and TGV²-KL for cells—solved numerically by the iterative ADMM schemes outlined in section 5. In all the tests, the ADMM iterations are stopped as soon as the relative change between two subsequent \mathbf{x} -iterates satisfies

$$\frac{\|\mathbf{x}^{(k+1)} - \mathbf{x}^{(k)}\|_2}{\|\mathbf{x}^{(k)}\|_2} < 10^{-6},$$

while the ADMM penalty parameter γ is set manually so as to fasten the convergence of the alternating scheme. More specifically, the numerical tests indicated the range $\gamma \in [1, 10]$ as a good choice.

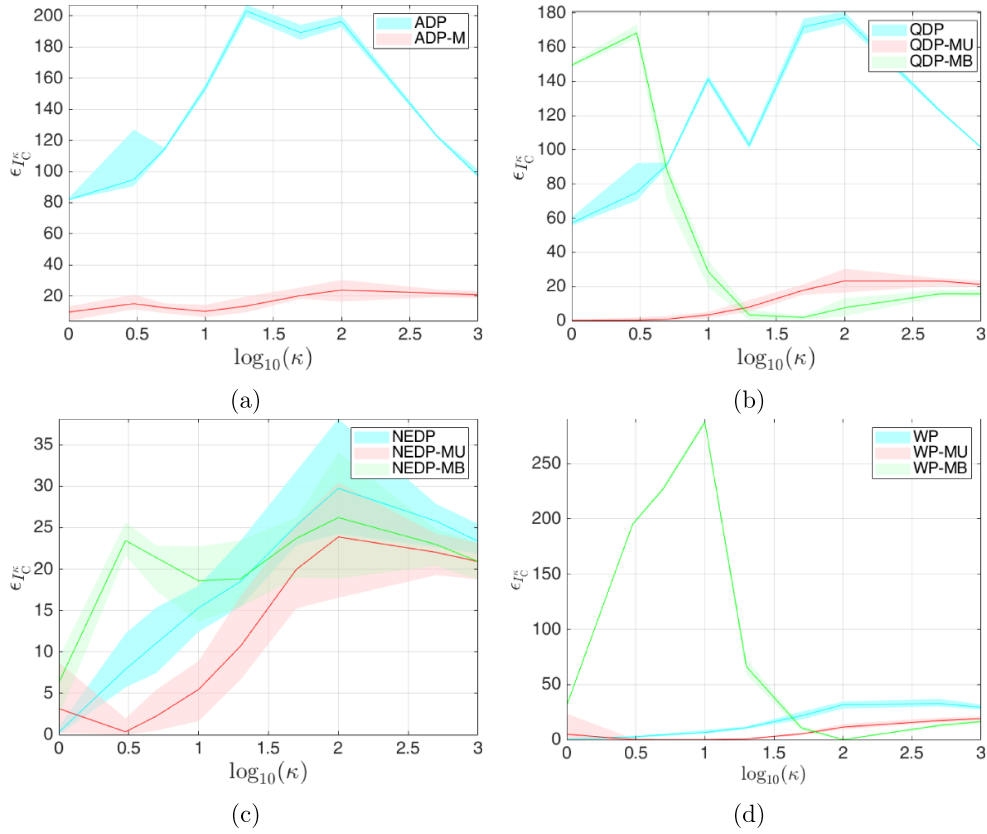


Figure 5. Test image satellite. Expected values and confidence intervals for the ISNR values achieved in different photon-counting regimes.

6.3. Test image satellite: parameter selection in the TV-KL model (55)

We consider the restoration of the test image satellite. In this first example, we set $\mathbf{b} \equiv 2 \times 10^{-3}$ and, in light of the dominant piece-wise constant features present in the image, we employ the TV regularization term in (5).

We start analyzing the behavior of the expected and limiting values defined in (74) and (75), respectively, within the four classes of ADP-, QDP-, NEDP- and WP-based criteria. In figure 5 for each class we plot the sample means $\eta_{I_C}^{\%}$ corresponding to the different photon-counting regimes κ expressed in \log_{10} -scale, and also the confidence intervals determined by the limiting values $\bar{\epsilon}_{I_C}^{\%}, \underline{\epsilon}_{I_C}^{\%}$.

For what concerns the ADP-based strategies, we recall that the MB and MU versions of the principle coincide. Notice that the masked criterion achieves significantly better results as the red band stays below the 30% regardless of the counting regime, while the unmasked method stays above the 80%

In the case of QDP-based approaches, one can immediately observe that the percentage differences achieved by the unmasked principle are particularly large for every κ . In the lower counting regimes, i.e. $\kappa \leq 5$, the MB version returns very poor results, while its performance

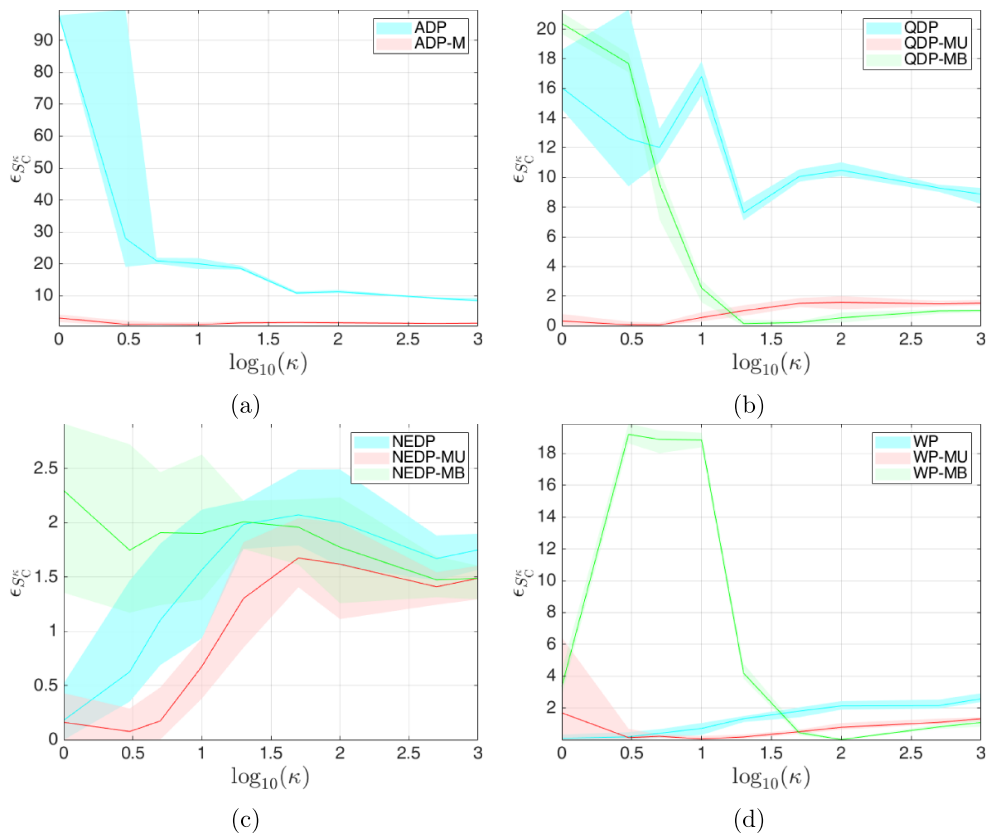


Figure 6. Test image satellite. Expected values and confidence intervals for the SSIM values achieved in different photon-counting regimes.

improves and stays below the 20% for $\kappa \geq 20$. On the other hand, we highlight that the MU principle presents a very robust behavior along the whole range of counting factors, as the corresponding percentage differences are always below the 30% and approach 0 for the smaller κ s.

When analyzing the NEDP-based approaches, one can notice that the MB method returns the poorer results in the low counting regimes while, in expectation, it outperforms the unmasked version in the mid- and high-counting regimes. On the other hand, the MU principle achieves the larger INSR values for all the considered regimes.

Finally, for the WP-based principles we notice that the MB criterion performs poorly in the low- and mid- counting regimes, while the green band stays below the 30% for $\kappa \geq 50$. One can also observe that the unmasked and the masked unbiased principles present a robust behavior with respect to the considered κ -values, with the unmasked approaching 0 for $\kappa \leq 3$ and the masked unbiased outperforming the competitors for $\kappa > 3$.

Figure 6 shows, for each class of methods, the sample means η_{SSIM}^{κ} and the confidence intervals related to the computed $\bar{\epsilon}_{SSIM}^{\kappa}, \epsilon_{SSIM}^{\kappa}$, i.e. the performances of the considered principles in terms of the SSIM. One can easily notice that the SSIM bands present the same behavior of the INSR

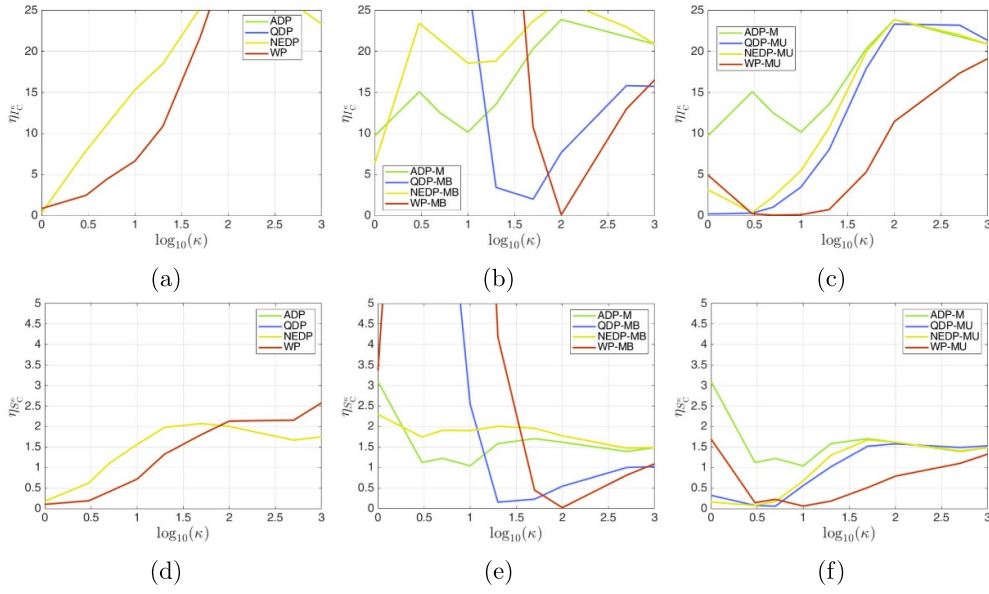


Figure 7. Test image satellite. Zoom of the expected values $\eta_{I_C}^E$ (top row) and $\eta_{S_C}^E$ (bottom row) for the ISNR and SSIM values achieved in different photon-counting regimes.

bands shown in figure 5, so that similar conclusions on the performances of the unmasked, MB and MU criteria can be drawn.

To analyze the results from a different point of view, in figures 7(a)–(c) we show the sample means $\eta_{I_C}^E$ of the unmasked, masked biased and masked unbiased principles, respectively, in the range $[0\%, 25\%]$, so as to visualize the best performances. In the unmasked category, NEDP and WP are the only principles staying below the 25% for $\kappa \geq 20$, with WP obtaining better results than the others for the low- and mid- counting regimes. Among the MB principles, QDP-MB and WP-MB stay in the interval of interest only in correspondence of the high-counting regimes, while ADP-M and NEDP-MB stay between 10% and 25% for each κ . Finally, in the MU class, all methods are in the visible range, with WP-MU being the best for $\kappa > 3$, followed by QDP-MU which reaches the highest ISNR for $\kappa \leq 3$.

Figures 7(d)–(f) show the sample means $\eta_{S_C}^E$ of the unmasked, MB and MU principles, respectively, in the range up to $[0\%, 5\%]$. The three plots confirm the considerations done for the ISNR about the best method within each group.

6.4. Test image stem: parameter selection in the TV^2 -KL model (56)

For the second example, we consider the restoration of the test image stem, with constant background emission $\mathbf{b} \equiv 2 \times 10^{-3}$, this time using the TV^2 regularization term defined in (6) due to the target image resembling a piece-wise linear function more than a piece-wise constant one.

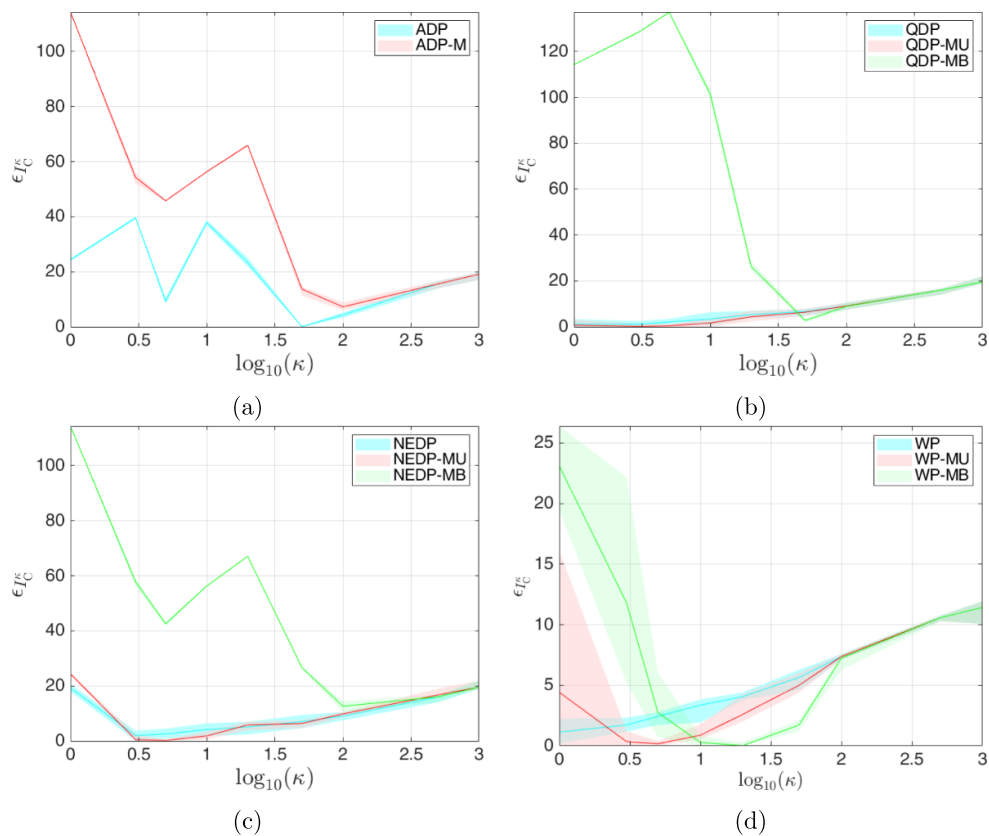


Figure 8. Test image stem. Expected values and confidence intervals for the ISNR values achieved in different photon-counting regimes.

In figure 8 we plot the sample means $\eta_{I^{\kappa}}$ and the confidence intervals corresponding to the different counting regimes κ for the four classes of ADP-, QDP-, NEDP-, and WP-based criteria.

For the ADP-based approaches, one can notice that the masked criterion reaches worse results than the unmasked one, which stays below the 40% for every counting regime.

For what concerns the QDP-based strategies, we can immediately notice that, as for the previous image, the biased masked version returns poor results for lower counting regimes, but it improves for $\kappa \geq 50$. On the other hand, both the unmasked and the unbiased masked achieve good and similar results, as they return percentage differences that are always below 20% for all factors κ .

In the case of the NEDP-based approaches one can notice a similar behavior to the QDP-based strategies, where the unmasked and unbiased masked achieve analogue results (except a little more variably across the different realization for the unmasked one) staying below the 20% for all κ and the biased masked strategy works poorly for $\kappa \leq 50$.

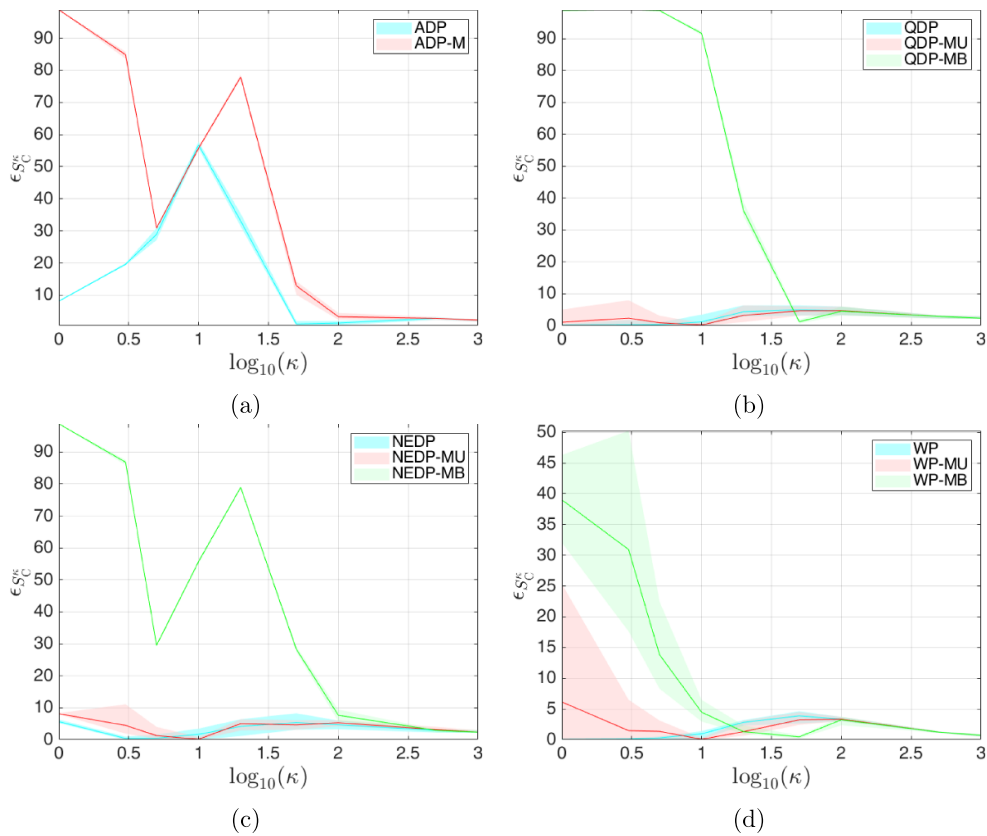


Figure 9. Test image stem. Expected values and confidence intervals for the SSIM values achieved in different photon-counting regimes.

Finally, for the WP-based principles, all of them return percentage differences less than 25%, with the biased masked working better for the mid- and high-count range, but worst for $\kappa \leq 5$. On the other hand, the unbiased masked reaches good results in the low- and mid-count range, with the exception of $\kappa = 1$ where the sample mean and the variability across the noise realization are higher than the one obtained with the unmasked strategy.

The same observations can be done by analyzing the SSIM results plotted in figure 9.

In figure 10 we show the expected values $\eta_{I_c^\kappa}$ (top row) and $\eta_{S_c^\kappa}$ (bottom row), in the range [0%, 25%] and [0%, 5%] respectively, divided by the type of principle: unmasked, biased masked and unbiased masked. Looking at both the ISNR and SSIM graphs we can note that, in the unmasked category, the best result is achieved by the WP in all the range of κ . Among the biased masked strategies, the WP-MB seems the best as it achieves the highest value of ISNR for $\kappa = 10, 20$, but the behavior for $\kappa \leq 5$ is poorer when compared to the other plots (but is the best in its category). Finally, for the unbiased masked principles, the QDP-MU obtains the best result in terms of ISNR for the low-count regime (even among all the methods), while the WP-MU works well for the mid- and high-range.

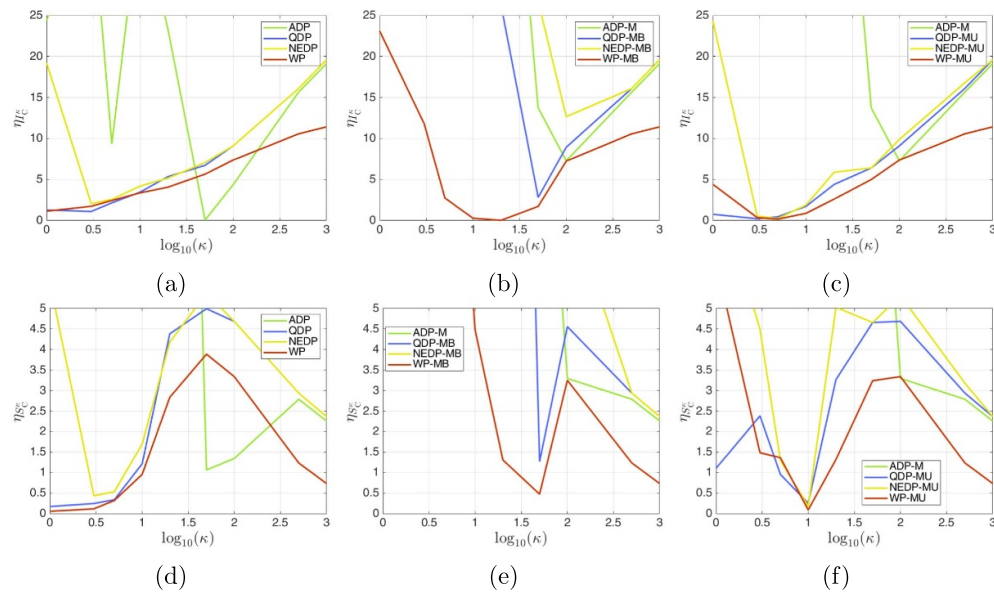


Figure 10. Test image stem. Zoom of the expected values $\eta_{I_c}^{\text{ISNR}}$ (top row) and $\eta_{S_c}^{\text{SSIM}}$ (bottom row) for the ISNR and SSIM values achieved in different photon-counting regimes.

6.5. Test image cells: parameter selection in the TGV²-KL model (57)

In this third example, we consider the restoration of the test image `cells` by employing the TGV² regularization term to effectively deal with the composite nature of the specimen. For the TGV², we set $\alpha_0 = 0.8, \alpha_1 = 0.3$ so as to maximize the ISNR for the highest counting regime considered here, i.e. $\kappa = 1000$. Moreover, we set a constant background emission $\mathbf{b} \equiv 10^{-1}$.

In figure 11 we show the ISNR bands achieved by the ADP-, QDP-, NEDP- and WP-based principles. As for the test image `stem`, the ADP-M strategy performs poorly in the lowest counting regimes, while it outperforms the unmasked version for $\kappa \geq 20$. On the other hand, the remaining three classes of methods present the same behavior: the MB versions of the principles achieve very low ISNR values in low- and mid-counting regimes, whereas the unmasked and MU principles present a very robust behavior with the latter achieving the best results.

Similar considerations can be drawn by looking at the SSIM bands reported in figure 12.

Finally, in figure 13 we show a close-up on the expected values $\eta_{I_c}^{\text{ISNR}}, \eta_{S_c}^{\text{SSIM}}$ in the range [0%, 25%] dividing the principles into unmasked, MB and MU. It is easy to conclude that the MU versions of the QDP, NEDP and WP return the best results both in terms of robustness and of quality measures achieved.

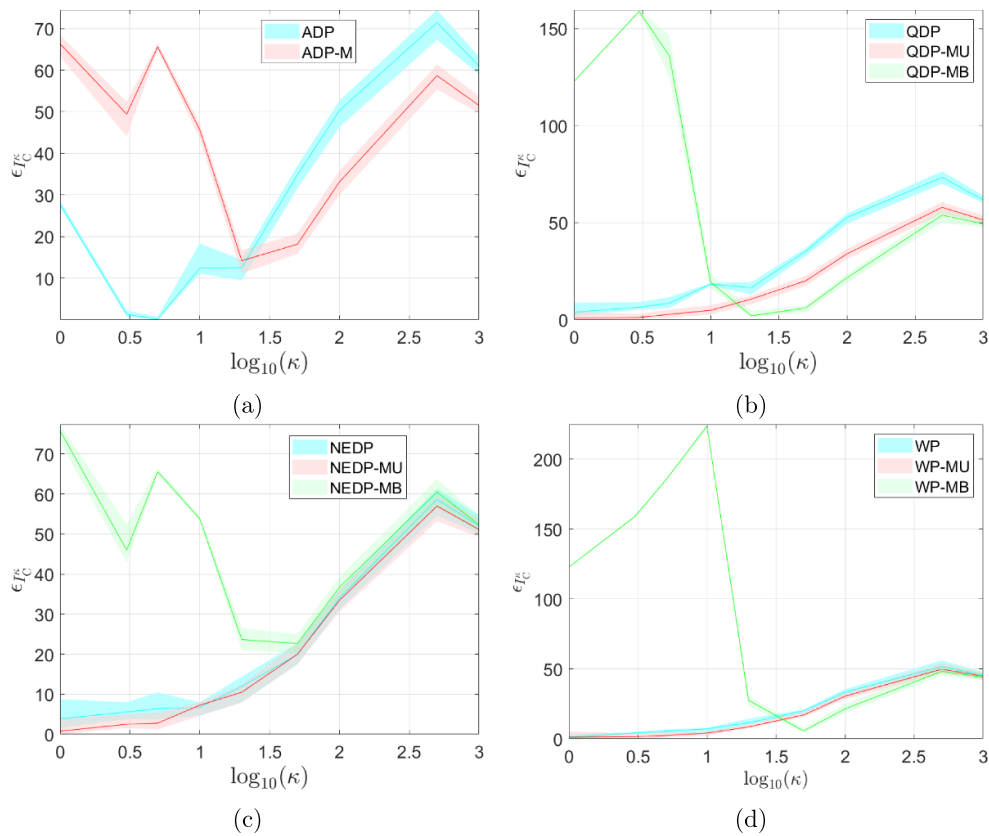


Figure 11. Test image cells. Expected values and confidence intervals for the ISNR values achieved in different photon-counting regimes.

6.6. Discussion

The detailed analysis carried out so far allows us to conclude that neglecting the zero-pixels in the acquired images as proposed in [9] can lead to particularly robust and successful parameter selection strategies provided that the proposed positive Poisson distribution is employed to model the undiscarded data. Generally speaking, in terms of quality measures the QDP-MU achieves the best results for $\kappa \leq 3$, while the WP-MU returns higher quality restorations when $\kappa > 3$. Moreover, in accordance with the theoretical results given in proposition 4.2, in the higher counting regimes the performed criteria show similar behaviors.

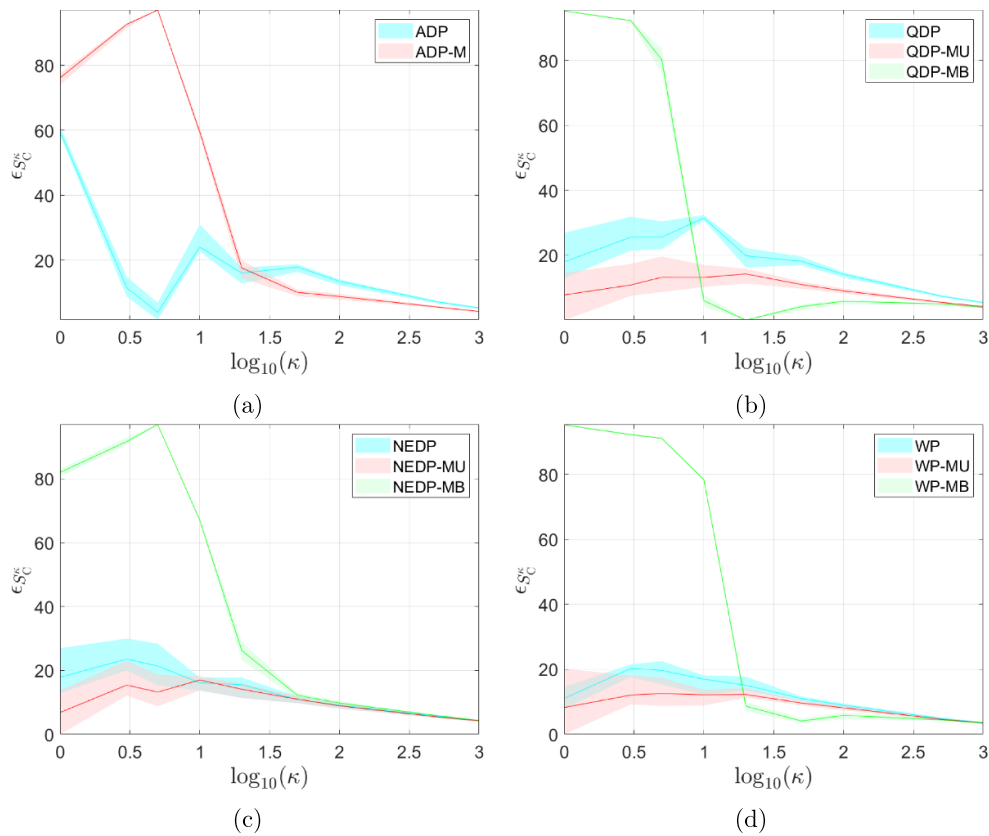


Figure 12. Test image cells. Expected values and confidence intervals for the SSIM values achieved in different photon-counting regimes.

We highlight that the improvements yielded by employing the proposed MU principles with respect to their unmasked versions are particularly relevant in the first example, that is for the test image `satellite`. To highlight a possible reason behind this phenomenon, in table 1 we report the average percentages of zero-pixels in the acquired images for the counting regimes considered. Such values are clearly influenced by the gray-level statistics of the underlying test images as well as by the selected background emissions. It is immediate to observe that for the test image `satellite` the number of zero pixels is very large for all the κ -values. As a consequence, masking the data in this scenario is particularly effective.

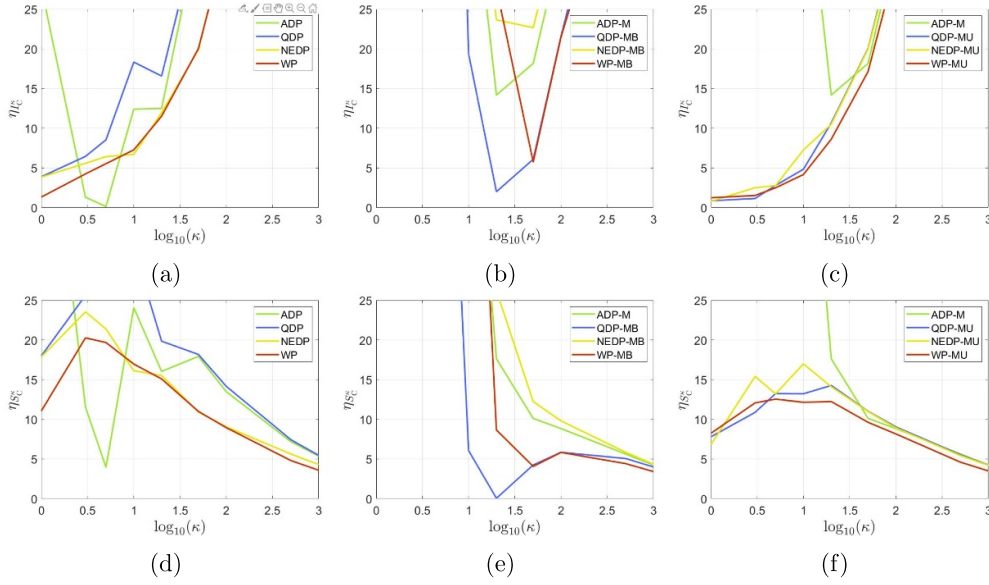


Figure 13. Test image cells. Zoom of the expected values η_C^E (top row), and $\eta_{S\Sigma}^E$ (bottom row) for the ISNR and SSIM values achieved in different photon-counting regimes.

Table 1. Average percentages of zero-pixels in the observed images for the different photon-counting regimes considered in the tests.

κ	1	3	5	10	20	50	100	500	1000
satellite	90%	83%	80%	78%	76%	75%	74%	73%	72%
stem	86%	66%	51%	28%	10%	1%	0.1%	0%	0%
cells	72%	49%	36%	22%	15%	11%	9%	6%	5%

7. Conclusions

In the present work, we addressed the parameter selection problem in the general \mathcal{R} – KL image restoration variational model under Poisson noise corruption. After recalling the classical unmasked principles typically used for setting the parameter, we discussed the idea of masking the observed data, as originally proposed in [9] to effectively deal with low photon-counting regimes. Nonetheless, we proved that neglecting the zero-pixels in the observed data without modifying the distribution of undiscarded pixels introduces a bias in the resulting principles. Hence, after defining the novel positive Poisson distribution, we proposed the masked unbiased versions of the original criteria. The unmasked, masked biased and masked unbiased strategies have been extensively tested on different images, with different regularization terms and for a wide range of counting regimes. The masked unbiased principles have been proven to outperform, on average, the corresponding unmasked and masked biased counterparts. As a future work, we plan to employ the proposed masked unbiased strategies so as to deal with different imaging problems such as computed tomography reconstruction and image super-resolution. Moreover, inspired by [28], we believe that the design of procedures faster than the grid-search employed here for the selection of μ based on the proposed principles is a matter that deserves further investigations.

Data availability statement

The data that support the findings of this study are available upon reasonable request to the authors.

Acknowledgment

Research of FB, AL, and FS was supported by the “National Group for Scientific Computation (GNCS-INDAM)” and by the ex60 project “Funds for selected research topics”, while MP acknowledges the contribution of ‘Young researchers funding’ awarded by GNCS-INDAM.

ORCID iDs

Francesca Bevilacqua  <https://orcid.org/0000-0001-8244-3355>

Alessandro Lanza  <https://orcid.org/0000-0002-4904-0682>

Monica Pragliola  <https://orcid.org/0000-0002-3074-1550>

Fiorella Sgallari  <https://orcid.org/0000-0002-9166-8879>

References

- [1] Almeida M S C and Figueiredo M A T 2013 Parameter estimation for blind and non-blind deblurring using residual whiteness measures *IEEE Trans. Image Process.* **22** 2751–63
- [2] Bardsley J M and Golde J 2009 Regularization parameter selection methods for ill-posed poisson maximum likelihood estimation *Inverse Problems* **25** 095005
- [3] Bertero M, Boccacci P, Desiderà G and Vicidomini G 2009 Image deblurring with poisson data: from cells to galaxies *Inverse Problems* **25** 123006
- [4] Bertero M, Boccacci P, Talenti G, Zanella R and Zanni L 2010 A discrepancy principle for poisson data *Inverse Problems* **26** 105004
- [5] Bevilacqua F, Lanza A, Pragliola M and Sgallari F 2022 Nearly exact discrepancy principle for low-count Poisson image restoration *J. Imaging* **8** 1
- [6] Bevilacqua F, Lanza A, Pragliola M and Sgallari F 2023 Whiteness-based parameter selection for Poisson data in variational image processing *Appl. Math. Modelling* **117** 197–218
- [7] Bonettini S and Prato M 2014 Accelerated gradient methods for the x-ray imaging of solar flares *Inverse Problems* **30** 055004
- [8] Boyd S, Parikh N, Chu E, Peleato B and Eckstein J 2011 Distributed optimization and statistical learning via the alternating direction method of multipliers *Found. Trends Mach. Learn.* **3** 1–122
- [9] Carlván M and Blanc-Feraud L 2012 Sparse Poisson noisy image deblurring *IEEE Trans. Image Process.* **21** 1834–46
- [10] Craven P and Wahba G 1978 Smoothing noisy data with spline functions *Numer. Math.* **31** 377–403
- [11] Eckstein J and Bertsekas D P 1992 On the Douglas-Rachford splitting method and the proximal point algorithm for maximal monotone operators *Math. Program.* **55** 293–318
- [12] Grinvald A and Steinberg I Z 1974 On the analysis of fluorescence decay kinetics by the method of least-squares *Anal. Biochem.* **59** 583–98
- [13] Gu C 1992 Cross-validating non-Gaussian data *J. Comput. Graph. Stat.* **1** 169–79
- [14] Kunisch K, Bredies K and Pock T 2010 Total generalized variation *J. Imaging Sci.* **3** 492–526
- [15] Lanza A, Pragliola M and Sgallari F 2020 Residual whiteness principle for parameter-free image restoration *Electron. Trans. Numer. Anal.* **53** 329–51
- [16] Lefkimmiatis S and Unser M 2013 Poisson image reconstruction with Hessian Schatten-norm regularization *IEEE Trans. Image Process.* **22** 4314–27
- [17] Lysaker M, Lundervold A and Tai X-C 2003 Noise removal using fourth-order partial differential equation with applications to medical magnetic resonance images in space and time *IEEE Trans. Image Process.* **12** 1579–90
- [18] Meinel W, Olivo-Marin J C and Angelini E D 2018 Denoising of microscopy images: a review of the state-of-the-art and a new sparsity-based method *IEEE Trans. Image Process.* **27** 3842–56

- [19] Rudin L I, Osher S and Fatemi E 1992 Nonlinear total variation based noise removal algorithms *Physica D* **60** 259–68
- [20] Serafino D, Landi G and Viola M 2021 Directional TGV-based image restoration under Poisson noise *J. Imaging* **7** 99
- [21] Sixou B, Hohweiller T and Ducros N 2018 Morozov principle for Kullback-Leibler residual term and Poisson noise *Inverse Problems Imaging* **12** 607–34
- [22] Starck J L and Murtagh F 2002 *Astronomical Image and Data Analysis* (Berlin: Springer)
- [23] Teuber T, Steidl G and Chan R H 2013 Minimization and parameter estimation for seminorm regularization models with l -divergence constraints *Inverse Problems* **29** 035007
- [24] Wahba G and Xiang D 1996 A generalized approximate cross validation for smoothing splines with non-Gaussian data *Stat. Sin.* **6** 675–92 (available at: www3.stat.sinica.edu.tw/statistica/oldpdf/A6n312.pdf)
- [25] Wang Z, Bovik A C, Sheikh H R and Simoncelli E P 2004 Image quality assessment: from error visibility to structural similarity *IEEE Trans. Image Process.* **13** 600–12
- [26] Zanella R, Boccacci P, Zanni L and Bertero M 2009 Efficient gradient projection methods for edge-preserving removal of Poisson noise *Inverse Problems* **25** 045010
- [27] Zanella R, Boccacci P, Zanni L and Bertero M 2013 Corrigendum: efficient gradient projection methods for edge-preserving removal of Poisson noise *Inverse Problems* **29** 119501
- [28] Zanni L, Benfenati A, Bertero M and Ruggiero V 2015 Numerical methods for parameter estimation in Poisson data inversion *J. Math. Imaging Vis.* **52** 397–413