Whiteness-based parameter selection for Poisson data in variational image processing

Whiteness-based parameter selection for Poisson data in variational image processing

Francesca Bevilacqua*[a], Alessandro Lanza[a], Monica Pragliola[b], Fiorella Sgallari[a]

[a] *Department of Mathematics, University of Bologna, Piazza di Porta San Donato 5, 40126 Bologna, Italy*
[b] *Department of Mathematics and Applications, University of Napoli Federico II, Via Vicinale Cupa Cintia 26, 80126 Napoli, Italy*

Abstract

We propose a novel parameter selection strategy for variational imaging problems under Poisson noise corruption. The selection of a suitable value of the regularization parameter, which is crucial for achieving high quality reconstructions, is known to be a particularly hard task in low photon-counting regimes. In this work, we extend the so-called residual whiteness principle originally designed for additive white noise to Poisson data. The proposed strategy relies on exploiting the whiteness property of a suitably standardized Poisson noise process. After deriving the theoretical properties underlying our proposal, we solve the target optimization problem by the alternating direction method of multipliers, in its standard two-blocks version or in a semi-linearized version depending on the imaging problem. Our strategy is extensively tested on image restoration and computed tomography reconstruction problems, and compared to the state-of-the-art discrepancy principle for Poisson noise proposed by Zanella at al. as well as to a nearly exact version of it recently proposed by the authors.

Keywords: Poisson noise, whiteness principle, image restoration, computed tomography, ADMM

*corresponding author

1. Introduction

In many research areas related to imaging applications, such as astronomy, microscopy and computed tomography, the acquired images are formed by counting the number of photons irradiated by a source and hitting the image domain. The number of photons measured by the sensor differs from the expected one due to random fluctuations that are faithfully modelled by a Poisson noise [1].

The general image formation (or degradation) model under Poisson noise corruption in vectorized form reads

$$\boldsymbol{y} \,=\, \mathbf{poiss}\left(\,\overline{\boldsymbol{\lambda}}\,\right)\,, \quad \overline{\boldsymbol{\lambda}} = \boldsymbol{g}\left(\mathbf{H}\bar{\boldsymbol{x}}\right) + \boldsymbol{b}\,, \tag{1}$$

where $\boldsymbol{y} \in \mathbb{N}^m$, $\overline{\boldsymbol{x}} \in \mathbb{R}_+^n$ and $\boldsymbol{b} \in \mathbb{R}_+^m$ - 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 observed degraded $m_1 \times m_2$ image, the unknown uncorrupted $n_1 \times n_2$ image and the so-called (usually known) background emission $m_1 \times m_2$ image, respectively, with $m = m_1 m_2$, $n = n_1 n_2$. Matrix $\mathbf{H} \in \mathbb{R}^{m \times n}$ contains the coefficients of a linear degradation operator, whereas the vectorial function $\boldsymbol{g} : \mathbb{R}^m \to \mathbb{R}^m$ is the identity function or a nonlinear function modelling the possible presence of (deterministic) nonlinearities in the degradation process. $\mathbf{H}$ and $\boldsymbol{g}$ are determined by the specific application at hand and here are assumed to be known. In particular, for the applications of interest in this paper, the function $\boldsymbol{g}$ can be restricted to the simplified form $\boldsymbol{g}(\boldsymbol{h}) = (g(h_1), g(h_2), \ldots, g(h_m))^T$, with $g : \mathbb{R}_+ \to \mathbb{R}_+$. Finally, $\mathbf{poiss}\left(\,\overline{\boldsymbol{\lambda}}\,\right) := \left(\mathrm{poiss}\left(\overline{\lambda}_1\right), \mathrm{poiss}\left(\overline{\lambda}_2\right), \ldots, \mathrm{poiss}\left(\overline{\lambda}_m\right)\right)^T$, with $\mathrm{poiss}(\,\overline{\lambda_i}\,)$ indicating the realization of a Poisson-distributed random variable with parameter (mean) $\overline{\lambda}_i$, hence $\overline{\boldsymbol{\lambda}} \in \mathbb{R}_+^m$ is the vectorized form of the $m_1 \times m_2$ noise-free degraded image.

The inverse problem of determining a good estimate $\boldsymbol{x}^*$ of the uncorrupted image $\overline{\boldsymbol{x}}$ given the degraded observation $\boldsymbol{y}$ is a hard task even when $\mathbf{H}$, $\boldsymbol{g}$ and $\boldsymbol{b}$ are known. In fact, such an inverse problem is typically ill-posed and, hence,

some a priori information or belief on the target image $\overline{\boldsymbol{x}}$ must necessarily be encoded, e.g. in the form of regularization, in order to obtain an acceptable estimate $\boldsymbol{x}^*$. A popular and effective approach allowing to explicitly include regularization is the so-called variational approach, according to which an estimate $\boldsymbol{x}^*$ of the original image $\bar{\boldsymbol{x}}$ is sought as the global minimizer of a suitably designed cost (or energy) function $\mathcal{J} : \mathbb{R}^n \to \mathbb{R}$; in formula

$$\boldsymbol{x}^*(\mu) \in \operatorname*{arg\,min}_{\boldsymbol{x} \in \mathbb{R}^n_+} \{\, \mathcal{J}(\boldsymbol{x}; \mu) := \mathcal{R}(\boldsymbol{x}) + \mu\, \mathcal{F}(\boldsymbol{\lambda}; \boldsymbol{y}) \,\}\,, \quad \boldsymbol{\lambda} := \boldsymbol{g}(\mathbf{H}\boldsymbol{x}) + \boldsymbol{b}\,, \tag{2}$$

where $\mathcal{R}$ and $\mathcal{F}$ are referred to as the *regularization term* and the *data fidelity term*, respectively, and where the so-called regularization parameter $\mu \in \mathbb{R}_{++}$ allows to balance the contribution of the two terms in the overall cost function.

The data fidelity term $\mathcal{F}$ measures the discrepancy between the noise-free degraded image $\boldsymbol{\lambda}$ and the noisy observation $\boldsymbol{y}$ in a way that accounts for the noise statistics. In presence of Poisson noise, according to the Maximum Likelihood (ML) estimation approach, the fidelity term is typically set as the (generalized) Kullback-Leibler (KL) divergence between $\boldsymbol{\lambda}$ and $\boldsymbol{y}$ (see, e.g., [2]), that is

$$\mathcal{F}(\boldsymbol{\lambda}; \boldsymbol{y}) = \mathrm{KL}(\boldsymbol{\lambda}; \boldsymbol{y}) := \sum_{i=1}^{m} \left(\lambda_i - y_i \ln \lambda_i + y_i \ln y_i - y_i\right)\,, \tag{3}$$

where $0 \ln 0 = 0$ is assumed. We indicate by $\mathcal{R}$-KL the class of variational models defined as in (2) with $\mathcal{F}$ equal to the KL divergence term in (3).

The regularization term $\mathcal{R}$ in (2) encodes prior information or beliefs on the target uncorrupted image $\overline{\boldsymbol{x}}$. One of the most popular and widely adopted regularizers in imaging is the Total Variation (TV) semi-norm [3], which reads

$$\mathcal{R}(\boldsymbol{x}) = \mathrm{TV}(\boldsymbol{x}) := \sum_{i=1}^{n} \|(\boldsymbol{\nabla}\boldsymbol{x})_i\|_2\,, \tag{4}$$

where $(\boldsymbol{\nabla}\boldsymbol{x})_i \in \mathbb{R}^2$ denotes the discrete gradient of image $\boldsymbol{x}$ at pixel location i. The TV term is known to be particularly effective for the regularization of piecewise constant images as it promotes sparsity of gradient magnitudes. We denote by TV-KL the $\mathcal{R}$-KL variational model with $\mathcal{R}$ equal to the TV function in (4).

The regularization parameter μ in (2) is of crucial importance for getting high quality reconstructions $\boldsymbol{x}^*(\mu)$. In fact, it is well established that, even upon

the selection of suitable fidelity and regularization terms, an incorrect value of μ can easily lead to meaningless reconstructions.

For this reason, a lot of research has been devoted to the design of effective strategies for the μ-selection task under Poisson noise corruption. In abstract form, such strategies can be formulated as follows:

$$\text{Select } \ \mu = \mu^* \ \text{ such that } \ \mathcal{C}(\boldsymbol{x}^*(\mu^*)) \ \text{ is satisfied}\,,$$

where $\boldsymbol{x}^*(\mu) : \mathbb{R}_{++} \to \mathbb{R}^n$ is the image reconstruction function introduced in (2) and where $\mathcal{C}(\cdot)$ is some selection criterion or principle.

The selection principles designed so far to deal with Poisson noise have mostly been inspired by the very wide literature related to the parameter selection under additive white Gaussian noise corruption, and they can be thus divided into two main classes according to their original derivation set-up:

1. Principles derived from imposing the value of some μ-dependent quantity;
2. Principles derived from optimizing some μ-dependent quantity.

As typical examples of first class strategies, we mention the *discrepancy principles* (DP) whose general form reads

$$\mathcal{C}(\boldsymbol{x}^*(\mu^*)) : \quad \mathcal{D}\,(\mu^*; \boldsymbol{y}) \,= \Delta \in \mathbb{R}_{++}\,, \tag{5}$$

with the so-called *discrepancy function* $\mathcal{D}\,(\mu; \boldsymbol{y})$ defined by

$$\mathcal{D}\,(\mu; \boldsymbol{y}) \,:=\, \mathrm{KL}(\boldsymbol{\lambda}^*(\mu); \boldsymbol{y})\,, \quad \boldsymbol{\lambda}^*(\mu) = \boldsymbol{g}(\mathbf{H}\boldsymbol{x}^*(\mu)) + \boldsymbol{b}\,. \tag{6}$$

The equality in (5) is commonly referred to as the *discrepancy equation* while Δ is the so-called *discrepancy value* which changes when considering different DP versions.

The Morozov discrepancy principle, which is widely adopted in presence of additive Gaussian noise, naturally induces a DP version according to which the scalar value Δ is replaced by the expected value of the KL fidelity term regarded as a function of the m-variate random vector $(Y_1, Y_2, \ldots, Y_m)$, with vector $\boldsymbol{\lambda}$ fixed. Unfortunately, recovering an exact closed-form expression for the target

expected value has been proven to be theoretically unfeasible, especially in low photon-counting regimes, i.e. when the entries of $\boldsymbol{\lambda}$ are small [2].

A popular alternative to the aforementioned *exact* but only theoretical DP, which has been proposed in [4] for the image denoising problem and extended in [5] to the restoration task, is based on truncating the Taylor series expansion of the theoretical expected value. For this reason, later in [2] this version of DP has been referred to as Approximate DP (ADP); in formula, it reads

$$\mathcal{C}(\boldsymbol{x}^*(\mu^*)): \quad \mathcal{D}(\mu^*; \boldsymbol{y}) = \frac{m}{2}\,, \tag{ADP}$$

with m indicating the total number of pixels in the observed image $\boldsymbol{y}$. The ADP is particularly robust from the theoretical viewpoint as it has been proven that the associated discrepancy equation admits a unique solution. Nonetheless, it has also been observed that, as the number of photons hitting the image domain gets smaller, the approximation considered by ADP becomes particularly rough and the principle returns low-quality reconstructions - see, e.g., [2].

In order to overcome the limitations presented by the ADP, in [2] the authors proposed the Nearly Exact DP (NEDP), which is a novel version of DP based on a more accurate approximation - both in low-counting and mid/high-counting regimes - of the expected value of the KL divergence term. In the NEDP, the discrepancy value Δ is replaced by a weighted least-square fitting of Montecarlo realizations of the target expected value and it is regarded as a function f of the regularization parameter μ; in formula:

$$\mathcal{C}(\boldsymbol{x}^*(\mu^*)): \quad \mathcal{D}\left(\mu^*; \boldsymbol{y}\right) = f\left(\mu^*\right)\,. \tag{NEDP}$$

Despite its very good experimental performances, the NEDP is characterized by theoretical limitations which are mostly related to the lack of guarantees on the uniqueness of the solution for the discrepancy equation; such limitations are also combined with the empirical evidence of multiple solutions in very extreme scenarios where the number of zero-pixels in the acquired data is relevant.

The general formulation of minimization-based principles is

$$\mathcal{C}(\boldsymbol{x}^*(\mu^*)) \;: \quad \mu^* \in \underset{\mu \in \mathbb{R}_{++}}{\arg\min}\; \mathcal{V}(\mu)\,,$$

where $\mathcal{V} : \mathbb{R}_{++} \to \mathbb{R}$ represents some demerit function to be minimized for selecting μ.

For Poisson data, this class of strategies has not been explored as much as the former. A few decades ago, some attempts have been made in order to adapt the popular Generalized Cross Validation (GCV) approach [6] to non-Gaussian data [7, 8]; nonetheless, these strategies, which ultimately rely on a weighted approximation of the KL fidelity term and on a slight reformulation of the classical GCV score, have not been diffusively employed for imaging problems.

Among the parameter selection strategies that have been developed in the context of additive white noise corruption, the class of minimization-based principles exploiting the noise *whiteness* property is one of the best performing [9, 10, 11, 12, 13, 14]. More specifically, one selects μ by minimizing the normalized correlation between the residual image components, that is by guaranteeing that the residual image resembles as much as possible the underlying additive noise in terms of whiteness. The whiteness-based approaches have been proven to outperform the Morozov discrepancy principle in different imaging tasks, such as, e.g., denoising/restoration [11] and super-resolution [12, 13] and, very importantly, do not need to know (or estimate) the noise standard deviation. Nonetheless, despite the encouraging results on Gaussian data, so far the whiteness principle has not been extended to Poisson noise corruption. In this work, we are going to address such extension.

1.1. Contribution

The main contribution of this paper is to provide the first extension of the whiteness principle proposed for additive white noise to the case of Poisson noise. In particular, we will illustrate theoretically how our proposal simply relies on applying the standard whiteness principle to a suitably standardized version of the Poisson-corrupted observation and that it can be used to select the regularization parameter μ in any variational model of the $\mathcal{R}$-KL class.

In this work, we apply the proposed selection strategy to the very popular

TV-KL model

$$\boldsymbol{x}^*(\mu) \in \underset{\boldsymbol{x}\in\mathbb{R}^n_+}{\arg\min}\{\,\mathrm{TV}(\boldsymbol{x}) + \mu\,\mathrm{KL}(\boldsymbol{\lambda};\boldsymbol{y})\,\}\,, \quad \boldsymbol{\lambda} = \boldsymbol{g}(\mathbf{H}\boldsymbol{x}) + \boldsymbol{b}\,, \qquad \text{(TV-KL)}$$

employed for the image restoration (IR) and X-rays CT image reconstruction (CTIR) tasks. In this two application scenarios, the matrix $\mathbf{H} \in \mathbb{R}^{m\times n}$ and the vectorial function $\boldsymbol{g}(\boldsymbol{h}) = (g(h_1), \ldots, g(h_m))^T$ in (TV-KL) are specified as follows:

$$\begin{aligned} \text{IR}: \quad & \mathbf{H} \in \mathbb{R}^{n\times n} \ \text{ blurring matrix}, \quad g(h_i) = h_i, \\ \text{CTIR}: \quad & \mathbf{H} \in \mathbb{R}^{m\times n} \ \text{ Radon matrix}, \quad g(h_i) = I_0\, e^{-h_i}, \ \ I_0 \in \mathbb{R}_{++}\,. \end{aligned} \tag{7}$$

From the numerical optimization viewpoint, in both scenarios the TV-KL model (TV-KL) will be solved by means of the alternating direction method of multipliers (ADMM), which for the CTIR problem will be adopted in a semi-linearized version so as to significantly decrease the per-iteration computational cost.

Experimental tests will show that in most cases the proposed selection principle outperforms the aforementioned ADP and NEDP state-of-the-art competitors and returns output images characterized by quality measures which are close to the ones achievable by manually tuning the regularization parameter μ.

The paper is organized as follows. In Section 2 we set the notations and recall some preliminary results on white random processes. The proposed selection strategy is introduced in Section 3, while in Section 4 we outline the numerical scheme employed for the solution of model (TV-KL). In Section 5 we extensively test the newly introduced strategy on IR and CTIR problems. Finally, we draw some conclusions and provide an outlook for future research in Section 6.

2. Notations and Preliminaries

In this paper, scalars, vectors and matrices are denoted, e.g., by x, $\boldsymbol{x} = \{x_i\}$ and $\mathbf{X} = \{x_{i,j}\}$, respectively, whereas scalar random variables and random matrices (also referred to as random fields) are indicated by X and $\boldsymbol{\mathcal{X}} = \{X_{i,j}\}$,

respectively. We indicate by $\mathrm{E}[X]$, $\mathrm{Var}[X]$, $\mathrm{Corr}[X,Y] = \mathrm{E}[XY]$ and P_X the expected value (or mean) of random variable X, the variance of X, the correlation between random variables X and Y and the probability mass function of (discrete) random variable X, respectively. We denote by $\mathbf{0}_d$, $\mathbf{I}_d$ and ι_{S} the d-dimensional null (column) vector, the identity matrix of order d and the indicator function of set S, respectively, with $\iota_{\mathrm{S}}(\boldsymbol{x}) = 0$ for $\boldsymbol{x} \in \mathrm{S}$ and $\iota_{\mathrm{S}}(\boldsymbol{x}) = +\infty$ for $\boldsymbol{x} \notin \mathrm{S}$. We indicate respectively by $(\,\cdot\,,\,\cdot\,)$ and $(\,\cdot\,;\,\cdot\,)$ the concatenation by rows and by columns of scalars, vectors and matrices. Finally, $\|\cdot\|_2$ denotes the vector Euclidean norm or the matrix Frobenius norm, depending on the context.

In order to introduce the theory underlying our proposal, it is useful to rewrite the vectorized image formation model (1) in its equivalent matrix form. Denoting by $\mathbf{Y}, \overline{\mathbf{\Lambda}} \in \mathbb{R}^{m_1 \times m_2}$ and $\mathbf{B}, \overline{\mathrm{X}} \in \mathbb{R}^{n_1 \times n_2}$ the matrix forms of vectors $\boldsymbol{y}, \overline{\boldsymbol{\lambda}} \in \mathbb{R}^m$ and $\boldsymbol{b}, \overline{\boldsymbol{x}} \in \mathbb{R}^n$, respectively, it reads

$$\mathbf{Y} \,=\, \mathbf{POISS}\left(\overline{\mathbf{\Lambda}}\right), \quad \overline{\mathbf{\Lambda}} = \mathbf{G}\left(\mathbf{H}\left(\overline{\boldsymbol{X}}\right)\right) + \mathbf{B}, \tag{8}$$

where, with a little abuse of notation, $\mathbf{H} : \mathbb{R}^{n_1 \times n_2} \to \mathbb{R}^{m_1 \times m_2}$ indicates here the linear operator encoded by matrix $\mathbf{H} \in \mathbb{R}^{m \times n}$ in the vectorized model (1), and where $\mathbf{POISS}\left(\overline{\mathbf{\Lambda}}\right) = \left\{\mathrm{poiss}\left(\overline{\lambda}_{i,j}\right)\right\}$ and $\mathbf{G}\left(\mathbf{H}\left(\overline{\boldsymbol{X}}\right)\right) = \left\{g\left(\left(\mathbf{H}\left(\overline{\boldsymbol{X}}\right)\right)_{i,j}\right)\right\}$, i.e. the matrix forms of vectors $\mathbf{poiss}\left(\overline{\boldsymbol{\lambda}}\right)$ and $\boldsymbol{g}(\mathbf{H}\bar{\boldsymbol{x}})$ in (1).

We now recall the definitions of weak stationary random field, ensemble normalized auto-correlation, sample normalized auto-correlation and, of particular importance for our purposes, white random field. To shorten notations in the definitions, we preliminarily define the following two sets of integer index pairs

$$\begin{aligned} \mathrm{I} \;&:=\; \left\{(i,j) \;\in \mathbb{Z}^2 \colon\; (i,j) \;\in [1,m_1] \times [1,m_2]\right\}, \\ \mathrm{L} \;&:=\; \left\{(l,m) \in \mathbb{Z}^2 \colon\; (l,m) \in [-(m_1-1),(m_1-1)] \times [-(m_2-1),(m_2-1)]\right\}. \end{aligned}$$

Definition 2.1 (weak stationary random field)**.** A $m_1 \times m_2$ random field $\boldsymbol{\mathcal{Z}} = \{Z_{i,j}\}$, $(i,j) \in \mathrm{I}$, is said to be weak stationary if

- $\mathrm{E}\left[Z_{i,j}\right] = \mu_{\boldsymbol{\mathcal{Z}}} \in \mathbb{R}$, $\mathrm{Var}\left[Z_{i,j}\right] = \sigma^2_{\boldsymbol{\mathcal{Z}}} \in \mathbb{R}_{++}$, $\forall\,(i,j) \in \mathrm{I}$;
- $\mathrm{Corr}\left[Z_{i_1,j_1}, Z_{i_1+l,j_1+m}\right] = \mathrm{Corr}\left[Z_{i_2,j_2}, Z_{i_2+l,j_2+m}\right]$, $\forall\,(i_1,j_1) \in \mathrm{I}$, $\forall\,(i_2,j_2) \in \mathrm{I}$, $\forall\,(l,m) \in \mathrm{L} \colon\; (i_2+l, j_2+m) \in \mathrm{I}$.

Definition 2.2 (ensemble normalized auto-correlation)**.** The ensemble normalized auto-correlation of a $m_1 \times m_2$ weak stationary random field $\boldsymbol{\mathcal{Z}} = \{Z_{i,j}\}$, $(i,j) \in \mathrm{I}$, is a $(2m_1 - 1) \times (2m_2 - 1)$ matrix $\mathbf{A}[\boldsymbol{\mathcal{Z}}] = \{a_{l,m}[\boldsymbol{\mathcal{Z}}]\}$, $(l,m) \in \mathrm{L}$, defined by

$$a_{l,m}[\boldsymbol{\mathcal{Z}}] = \frac{\mathrm{Corr}[Z_{i,j}, Z_{i+l,j+m}]}{\sigma^2_{\boldsymbol{\mathcal{Z}}}}, \quad (l,m) \in \mathrm{L}, \ (i,j) \in \mathrm{I} : \ (i+l, j+m) \in \mathrm{I}.$$

Definition 2.3 (white random field)**.** A $m_1 \times m_2$ random field $\boldsymbol{\mathcal{Z}} = \{Z_{i,j}\}$, $(i,j) \in \mathrm{I}$, is said to be white if

- it is weak stationary with $\mu_{\boldsymbol{\mathcal{Z}}} = 0$;
- it is uncorrelated, that is its ensemble normalized autocorrelation $\mathbf{A}[\boldsymbol{\mathcal{Z}}] = \{a_{l,m}[\boldsymbol{\mathcal{Z}}]\}$, $(l,m) \in \mathrm{L}$, satisfies:

$$a_{l,m}[\boldsymbol{\mathcal{Z}}] = \begin{cases} 0 & \forall\, (l,m) \in \mathrm{L} \setminus \{(0,0)\}\,, \\ 1 & \text{if } (l,m) = (0,0)\,. \end{cases}$$

Definition 2.4 (sample normalized auto-correlation)**.** The sample normalized auto-correlation of a $m_1 \times m_2$ non-zero matrix $\mathbf{Z} = \{z_{i,j}\}$, $(i,j) \in \mathrm{I}$, is a $(2m_1 - 1) \times (2m_2 - 1)$ matrix $\mathbf{S}(\mathbf{Z}) = \{s_{l,m}(\mathbf{Z})\}$, $(l,m) \in \mathrm{L}$, defined by

$$s_{l,m}(\mathbf{Z}) = \frac{1}{\|\mathbf{Z}\|_2^2} \sum_{(i,j) \in \mathrm{I}} z_{i,j}\, z_{i+l,j+m}\,. \tag{9}$$

It follows from Definition 2.4 that, given a non-zero matrix $\mathbf{Z} = \{z_{i,j}\}$, $(i,j) \in \mathrm{I}$, one can measure the global amount of normalized auto-correlation between the entries of $\mathbf{Z}$, that is how far is $\mathbf{Z}$ from being the realization of a white random field, via the following non-negative and scale-invariant scalar measure of whiteness (or, better, of non-whiteness) ([10, 11, 13]):

$$\mathcal{W}(\mathbf{Z}) := \|\mathbf{S}(\mathbf{Z})\|_2^2 = \sum_{(l,m) \in \mathrm{L}} \left(s_{l,m}(\mathbf{Z})\right)^2, \tag{10}$$

with scalars $s_{l,m}(\mathbf{Z})$ defined in (9).

As proved in [11], when periodic boundary conditions are assumed for the $m_1 \times m_2$ matrix $\mathbf{Z}$, the $(2m_1-1) \times (2m_2-1)$ sample normalized auto-correlation matrix $\mathbf{S}(\mathbf{Z})$ defined in (9) presents some symmetries and the whiteness measure

$\mathcal{W}(\mathbf{Z})$ in (10) can be computed very efficiently (with $\mathrm{O}(m \log m)$ computational complexity, $m = m_1 m_2$) based on the preliminary calculation of the 2D discrete Fourier transform of $\mathbf{Z}$ (implemented by 2D fast Fourier transform). For completeness, in Proposition 2.1 below we recall the result reported in [11].

Proposition 2.1. Let $\mathbf{Z} = \{z_{i,j}\} \in \mathbb{R}^{m_1 \times m_2}$ be a non-zero matrix and let $\widetilde{\mathbf{Z}} = \{\widetilde{z}_{i,j}\} \in \mathbb{C}^{m_1 \times m_2}$ be its 2D discrete Fourier transform. Then, assuming periodic boundary conditions for $\mathbf{Z}$, the function $\mathcal{W}$ in (10) can be written as:

$$\mathcal{W}(\mathbf{Z}) \;=\; \frac{\displaystyle\sum_{(i,j)\,\in\,\mathrm{I}} |\widetilde{z}_{i,j}|^4}{\left(\displaystyle\sum_{(i,j)\,\in\,\mathrm{I}} |\widetilde{z}_{i,j}|^2\right)^2} \,. \tag{11}$$

3. The proposed whiteness principle for Poisson noise

In this section, we show how the residual whiteness principle proposed for additive white noise can be quite easily extended to the case of Poisson noise based on suitable random variable standardizations.

For this purpose, first in Definition 3.1 we recall the formal definition of Poisson random variable and Poisson independent random field, then in Definition 3.2 we introduce their standard(ized) versions, whose main properties are finally highlighted in Proposition 3.1.

Definition 3.1 (Poisson random variable and independent random field)**.** 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

$$\mathrm{P}_Y(y \mid \lambda) \,=\, \frac{\lambda^y e^{-\lambda}}{y\,!}\,, \quad y \in \mathbb{N}\,.$$

The expected value and variance of random variable Y are given by

$$\mathrm{E}\,[Y] \;=\; \mathrm{Var}\,[Y] \;=\; \lambda\,.$$

A random field $\boldsymbol{\mathcal{Y}} = \{Y_{i,j}\}$ is said to be independent Poisson distributed with parameter $\boldsymbol{\Lambda} = \{\lambda_{i,j}\}$, denoted by $\boldsymbol{\mathcal{Y}} \sim \boldsymbol{\mathcal{P}}(\boldsymbol{\Lambda})$, if it satisfies:

$$Y_{i,j} \sim \mathcal{P}(\lambda_{i,j}) \;\; \forall\,(i,j) \in \mathrm{I}\,, \quad \mathrm{P}_{\boldsymbol{\mathcal{Y}}}(\mathbf{Y} \mid \boldsymbol{\Lambda}) = \prod_{(i,j)\in \mathrm{I}} \mathrm{P}_{Y_{i,j}}(y_{i,j} \mid \lambda_{i,j})\,. \tag{12}$$

Definition 3.2 (standard Poisson random variable and independent random field)**.** Let $Y \sim \mathcal{P}(\lambda)$. We call the discrete random variable Z defined by

$$Z \;=\; S_\lambda(Y) \;:=\; \frac{Y - \mathrm{E}\,[Y]}{\sqrt{\mathrm{Var}\,[Y]}} \;=\; \frac{Y-\lambda}{\sqrt{\lambda}} \;=\; \frac{1}{\sqrt{\lambda}}\,Y - \sqrt{\lambda}\,, \tag{13}$$

as standard Poisson distributed with parameter λ, denoted by $Z \sim \widetilde{\mathcal{P}}(\lambda)$.

Let $\boldsymbol{\mathcal{Y}} \sim \boldsymbol{\mathcal{P}}(\boldsymbol{\Lambda})$. We call the random field defined by

$$\boldsymbol{\mathcal{Z}} \;=\; \{Z_{i,j}\} \quad \text{with} \quad Z_{i,j} \;=\; S_{\lambda_{i,j}}(Y_{i,j}) \;\; \forall\,(i,j) \in \mathrm{I}\,, \tag{14}$$

as independent standard Poisson distributed with parameter $\boldsymbol{\Lambda}$, denoted by $\boldsymbol{\mathcal{Z}} \sim \widetilde{\boldsymbol{\mathcal{P}}}(\boldsymbol{\Lambda})$.

Proposition 3.1. Let $Z \sim \widetilde{\mathcal{P}}(\lambda)$ and let S_λ be the standardization function defined in (13). Then, the probability mass function, expected value and variance of random variable Z are given by:

$$\mathrm{P}_Z\,(z|\lambda) = \frac{\lambda^{S_\lambda^{-1}(z)}\,e^{-\lambda}}{\big(S_\lambda^{-1}(z)\big)\,!}, \;\; z \in \{S_\lambda(0), S_\lambda(1), \ldots\}\,, \;\; S_\lambda^{-1}(z) \;=\; \sqrt{\lambda}\,z + \lambda, \tag{15}$$

$$\mathrm{E}\,[\,Z\,] = 0\,, \quad \mathrm{Var}\,[\,Z\,] = 1\,. \tag{16}$$

Hence, any independent standard Poisson random field $\boldsymbol{\mathcal{Z}} \sim \widetilde{\boldsymbol{\mathcal{P}}}(\boldsymbol{\Lambda})$ is white.

Proof. The scalar affine standardization function $S_\lambda : \mathbb{N} \to \{S_\lambda(0), S_\lambda(1), \ldots\}$ in (13) is bijective (as $\lambda \in \mathbb{R}_{++}$), hence it admits the inverse S_λ^{-1} defined in (15). The expression of P_Z in (15) thus comes from specifying the general form of the probability mass function of a discrete random variable defined by a bijective function of another discrete random variable. The fact that Z has zero-mean and unit-variance - as stated in (16) - comes immediately from the definition of S_λ in (13).

It thus follows from the definition of a standard Poisson independent random field $\boldsymbol{\mathcal{Z}} = \{Z_{i,j}\}$ given in (14) and from statement (16) that:

$$Z_{i,j} \sim \widetilde{\mathcal{P}}\left(\lambda_{i,j}\right) \implies \begin{cases} \mathrm{E}\left[Z_{i,j}\right] = \mu_{\boldsymbol{\mathcal{Z}}} = 0 \\ \mathrm{Var}\left[Z_{i,j}\right] = \sigma^2_{\boldsymbol{\mathcal{Z}}} = 1 \end{cases}, \forall\,(i,j)\,. \tag{17}$$

Moreover, it clearly comes from independence of a non-standard Poisson random field $\boldsymbol{\mathcal{Y}}$ - formalized in (12) - and from the entry-wise definition of random field standardization in (14) that independence also holds true for a standard Poisson random field $\boldsymbol{\mathcal{Z}}$; in formula:

$$\mathrm{P}_{\boldsymbol{\mathcal{Z}}}(\mathbf{Z} \mid \boldsymbol{\Lambda}) = \prod_{(i,j)} \mathrm{P}_{Z_{i,j}}(z_{i,j} \mid \lambda_{i,j})\,.$$

Since independence implies uncorrelation and based on (17), we have

$$\mathrm{Corr}\left[Z_{i_1,j_1}, Z_{i_2,j_2}\right] = \begin{cases} 0 & \text{for } (i_1,j_1) \neq (i_2,j_2)\,, \\ \mathrm{Var}\left[Z_{i_1,j_1}\right] = \sigma^2_{\boldsymbol{\mathcal{Z}}} = 1 & \text{for } (i_1,j_1) = (i_2,j_2)\,. \end{cases} \tag{18}$$

It follows from (17), (18) and from Definition 2.1 that $\boldsymbol{\mathcal{Z}}$ is a weak stationary random field. Then, it comes from (18) and from Definition 2.2 that the ensemble normalized auto-correlation $\mathbf{A}\left[\boldsymbol{\mathcal{Z}}\right] = \{a_{l,m}\left[\boldsymbol{\mathcal{Z}}\right]\}$ satisfies

$$a_{l,m}\left[\boldsymbol{\mathcal{Z}}\right] = \begin{cases} 0 & \text{for } (l,m) \neq (0,0)\,, \\ 1 & \text{for } (l,m) = (0,0)\,. \end{cases}$$

Hence, based on Definition 2.3, we can conclude that any standard Poisson independent random field $\boldsymbol{\mathcal{Z}}$ is white. □

In light of Definition 3.1, the image formation model (8) can be written in probabilistic terms as follows:

$$\mathbf{Y} \text{ realization of } \boldsymbol{\mathcal{Y}} \sim \boldsymbol{\mathcal{P}}\left(\overline{\boldsymbol{\Lambda}}\right)\,, \tag{19}$$

with matrix $\overline{\boldsymbol{\Lambda}}$ defined in (8).

Then, based on Definition 3.2, after introducing the matrix

$$\mathbf{Z} = \{z_{i,j}\} \text{ with } z_{i,j} = S_{\overline{\lambda}_{i,j}}(y_{i,j}) = \frac{y_{i,j} - \overline{\lambda}_{i,j}}{\sqrt{\overline{\lambda}_{i,j}}}\,, \tag{20}$$

the probabilistic model (19) can be equivalently written in standardized form as

$$\mathbf{Z} \text{ realization of } \boldsymbol{\mathcal{Z}} \sim \widetilde{\boldsymbol{\mathcal{P}}}\left(\overline{\boldsymbol{\Lambda}}\right) .$$

That is, matrix $\mathbf{Z}$ in (20) with $\overline{\boldsymbol{\Lambda}}$ in (8) is the realization of an independent standard Poisson random field $\boldsymbol{\mathcal{Z}}$ which, according to Proposition 3.1, is white.

We note that, unlike the observed realization $\mathbf{Y}$ in (19), the realization $\mathbf{Z}$ in (20) is not available as it depends on $\overline{\boldsymbol{\Lambda}}$ which, in its turn, depends on the unknown uncorrupted image $\overline{\mathbf{X}}$. However, the whiteness property of $\mathbf{Z}$ can be exploited for stating a new principle for automatically selecting the value of the regularization parameter μ in the class of $\mathcal{R}$-KL variational models.

Denoting by $\mathbf{X}^*(\mu) = \left\{x^*_{i,j}(\mu)\right\}$ the matrix form of the solution of a $\mathcal{R}$-KL model - e.g., of the TV-KL model in (TV-KL) - we introduce the μ-dependent matrices $\boldsymbol{\Lambda}^*(\mu), \mathbf{Z}^*(\mu) \in \mathbb{R}^{m_1 \times m_2}$ given by

$$\boldsymbol{\Lambda}^*(\mu) \;=\; \left\{\lambda^*_{i,j}(\mu)\right\} \;=\; \mathbf{G}\left(\mathbf{H}\left(\mathbf{X}^*(\mu)\right)\right) + \mathbf{B}, \tag{21}$$

$$\mathbf{Z}^*(\mu) \;=\; \left\{z^*_{i,j}(\mu)\right\} \quad \text{with} \quad z^*_{i,j}(\mu) \;=\; S_{\lambda^*_{i,j}(\mu)}(y_{i,j}) \;=\; \frac{y_{i,j} - \lambda^*_{i,j}(\mu)}{\sqrt{\lambda^*_{i,j}(\mu)}}, \tag{22}$$

The ideal goal of any criterion for choosing μ in the class of $\mathcal{R}$-KL models is to select the value μ^* yielding the closest solution image $\mathbf{X}^*(\mu^*)$ to the target uncorrupted image $\overline{\mathbf{X}}$, according to some distance metric. The conjecture behind our proposal is that the closer the solution $\mathbf{X}^*(\mu)$ is to the target $\overline{\mathbf{X}}$, the closer the matrix $\mathbf{Z}^*(\mu)$ defined in (21)-(22) will be to $\mathbf{Z}$ in (20), so the more $\mathbf{Z}^*(\mu)$ will resemble the realization of a white random field. Hence, the proposed criterion, that we refer to as the Poisson Whiteness Principle (PWP), consists in choosing the value of μ leading to the less auto-correlated matrix $\mathbf{Z}^*(\mu)$. Based on the scalar normalized auto-correlation measure introduced in (10), the PWP reads:

$$\boxed{\begin{array}{c} \text{Select } \; \mu = \mu^* \in \underset{\mu \in \mathbb{R}_{++}}{\arg\min} \left\{\, W(\mu) \;:=\; \mathcal{W}\left(\mathbf{Z}^*(\mu)\right) \right\}, \\ \text{with matrix } \mathbf{Z}^*(\mu) \text{ defined in (21)-(22) and function } \mathcal{W} \text{ in (10).} \end{array}} \tag{PWP}$$

4. Numerical implementation of the Poisson whiteness principle

In this section, first we address the numerical solution of the (TV-KL) model for the IR and CTIR imaging problems, that is for matrix $\mathbf{H}$ and function $\boldsymbol{g}$ defined as in (7), then we outline the overall PWP-based reconstruction approach.

Recalling the definition of TV in (4) and introducing the discrete gradient matrix $\mathbf{D} := (\mathbf{D}_h; \mathbf{D}_v) \in \mathbb{R}^{2n \times n}$ with $\mathbf{D}_h, \mathbf{D}_v \in \mathbb{R}^{n \times n}$ two finite difference matrices discretizing the first-order partial derivatives of image $\boldsymbol{x}$ in the horizontal and vertical direction, respectively, we write the (TV-KL) model in the following form:

$$\boldsymbol{x}^* \in \operatorname*{arg\,min}_{\boldsymbol{x} \in \mathbb{R}^n} \left\{ \sum_{i=1}^{n} \|(\mathbf{D}\boldsymbol{x})_i\|_2 + \mu \, \mathrm{KL}\left(\boldsymbol{g}(\mathbf{H}\boldsymbol{x}) + \boldsymbol{b}; \boldsymbol{y}\right) + \iota_{\mathbb{R}^n_+}(\boldsymbol{x}) \right\}, \tag{23}$$

where, with a little abuse of notation, $(\mathbf{D}\boldsymbol{x})_i := ((\mathbf{D}_h\boldsymbol{x})_i \, ; \, (\mathbf{D}_v\boldsymbol{x})_i) \in \mathbb{R}^2$, the discrete gradient of image $\boldsymbol{x}$ at pixel i.

By introducing the three auxiliary variables $\boldsymbol{t}_1 = \mathbf{D}\boldsymbol{x} \in \mathbb{R}^{2n}$, $\boldsymbol{t}_2 = \mathbf{H}\boldsymbol{x} \in \mathbb{R}^m$ and $\boldsymbol{t}_3 = \boldsymbol{x} \in \mathbb{R}^n$, problem (23) can be equivalently rewritten in the following linearly constrained form:

$$\begin{aligned} \{\boldsymbol{x}^*, \boldsymbol{t}_1^*, \boldsymbol{t}_2^*, \boldsymbol{t}_3^*\} \in \operatorname*{arg\,min}_{\boldsymbol{x}, \boldsymbol{t}_1, \boldsymbol{t}_2, \boldsymbol{t}_3} & \left\{ \sum_{i=1}^{n} \|\boldsymbol{t}_{1,i}\|_2 + \mu \, \mathrm{KL}\left(\boldsymbol{g}(\boldsymbol{t}_2) + \boldsymbol{b}; \boldsymbol{y}\right) + \iota_{\mathbb{R}^n_+}(\boldsymbol{t}_3) \right\} \\ & \text{subject to: } \boldsymbol{t}_1 = \mathbf{D}\boldsymbol{x}, \;\; \boldsymbol{t}_2 = \mathbf{H}\boldsymbol{x}, \;\; \boldsymbol{t}_3 = \boldsymbol{x}, \end{aligned} \tag{24}$$

where $\boldsymbol{t}_{1,i} := (\mathbf{D}\boldsymbol{x})_i \in \mathbb{R}^2$, $i = 1. \ldots, n$.

Problem (24) can be equivalently and usefully reformulated as a standard two-blocks (additively) separable optimization problem with linear constraints. In fact, it is easy to prove - see, e.g., [15] - that, after introducing the total auxiliary variable $\boldsymbol{t} := (\boldsymbol{t}_1; \boldsymbol{t}_2; \boldsymbol{t}_3) \in \mathbb{R}^{m+3n}$, problem (24) takes the form:

$$\begin{aligned} \{\boldsymbol{x}^*, \boldsymbol{t}^*\} \;\; \in \;\; & \operatorname*{arg\,min}_{\boldsymbol{x}, \boldsymbol{t}} \{ \, C_1(\boldsymbol{x}) + C_2(\boldsymbol{t}) \, \} \\ & \text{subject to: } \mathbf{M}_1\boldsymbol{x} + \mathbf{M}_2\boldsymbol{t} = \mathbf{0}, \end{aligned} \tag{25}$$

where the two cost functions $C_1 : \mathbb{R}^n \to \mathbb{R}$ and $C_2 : \mathbb{R}^{m+3n} \to \mathbb{R}$ are defined by

$$C_1(\boldsymbol{x}) = 0, \quad C_2(\boldsymbol{t}) = \sum_{i=1}^{n} \|\boldsymbol{t}_{1,i}\|_2 + \mu \, \mathrm{KL}\left(\boldsymbol{g}(\boldsymbol{t}_2) + \boldsymbol{b}; \boldsymbol{y}\right) + \iota_{\mathbb{R}^n_+}(\boldsymbol{t}_3), \tag{26}$$

and the two matrices $\mathbf{M}_1 \in \mathbb{R}^{(m+3n)\times n}$ and $\mathbf{M}_2 \in \mathbb{R}^{(m+3n)\times(m+3n)}$ read

$$\mathbf{M}_1 = (\mathbf{D}; \mathbf{H}; \mathbf{I}_n) \,, \quad \mathbf{M}_2 = -\mathbf{I}_{m+3n} \,. \tag{27}$$

Functions C_1 and C_2 in (26) are both proper, lower semi-continuous and convex, hence the two-blocks separable problem (25)-(27) can be solved by the standard two-blocks ADMM [16], with guaranteed convergence.

In particular, the augmented Lagrangian function associated to (25) reads

$$\mathcal{L}(\boldsymbol{x}, \boldsymbol{t}, \boldsymbol{\rho}; \beta) = C_1(\boldsymbol{x}) + C_2(\boldsymbol{t}) + \langle \boldsymbol{\rho}, \mathbf{M}_1\boldsymbol{x} + \mathbf{M}_2\boldsymbol{t} \rangle + \frac{\beta}{2}\|\mathbf{M}_1\boldsymbol{x} + \mathbf{M}_2\boldsymbol{t}\|_2^2, \tag{28}$$

where $\boldsymbol{\rho} \in \mathbb{R}^{m+3n}$ is the vector of Lagrange multipliers associated to the linear constraint in (25) and $\beta \in \mathbb{R}_{++}$ is the ADMM penalty parameter.

Solving problem (25) amounts to seek the saddle point(s) of the augmented Lagrangian function which, according to the standard two-blocks ADMM, can be computed as the limit point of the following iterative procedure:

$$\boldsymbol{x}^{(k+1)} \in \underset{\boldsymbol{x} \in \mathbb{R}^n}{\arg\min}\, \mathcal{L}(\boldsymbol{x}, \boldsymbol{t}^{(k)}, \boldsymbol{\rho}^{(k)}; \beta) \tag{29}$$

$$\boldsymbol{t}^{(k+1)} \in \underset{\boldsymbol{t} \in \mathbb{R}^{m+3n}}{\arg\min}\, \mathcal{L}(\boldsymbol{x}^{(k+1)}, \boldsymbol{t}, \boldsymbol{\rho}^{(k)}; \beta) \tag{30}$$

$$\boldsymbol{\rho}^{(k+1)} = \boldsymbol{\rho}^{(k)} + \beta \left(\mathbf{M}_1\boldsymbol{x}^{(k+1)} + \mathbf{M}_2\boldsymbol{t}^{(k+1)} \right) . \tag{31}$$

In the following subsections 4.1 and 4.2 we detail how to solve the $\boldsymbol{x}$-subproblem in (29) and the $\boldsymbol{t}$-subproblem in (30), respectively, when tackling the IR and CTIR imaging problems. Then, in subsection 4.3 we outline and discuss the overall PWP-based reconstruction approach.

4.1. The $\boldsymbol{x}$-subproblem

Recalling the definition of the augmented Lagrangian function $\mathcal{L}$ in (28), with functions C_1, C_2 in (26) and matrices $\mathbf{M}_1, \mathbf{M}_2$ in (27), after dropping the constant terms the $\boldsymbol{x}$-update problem in (29) reads

$$\begin{aligned} \boldsymbol{x}^{(k+1)} &\in \underset{\boldsymbol{x} \in \mathbb{R}^n}{\arg\min} \left\{ \langle \boldsymbol{\rho}^{(k)}, \mathbf{M}_1\boldsymbol{x} - \boldsymbol{t}^{(k)} \rangle + \frac{\beta}{2}\|\mathbf{M}_1\boldsymbol{x} - \boldsymbol{t}^{(k)}\|_2^2 \right\} \\ &= \underset{\boldsymbol{x} \in \mathbb{R}^n}{\arg\min} \left\{ Q^{(k)}(\boldsymbol{x}) := \frac{1}{2}\|\mathbf{M}_1\boldsymbol{x} - \boldsymbol{v}^{(k)}\|_2^2 \right\}, \;\; \boldsymbol{v}^{(k)} = \boldsymbol{t}^{(k)} - \frac{1}{\beta}\,\boldsymbol{\rho}^{(k)}. \end{aligned} \tag{32}$$

Since the cost function $Q^{(k)}$ in (32) is quadratic and convex, it admits global minimizers which are the solutions of the linear system of normal equations:

$$\mathbf{M}_1^{\mathrm{T}}\mathbf{M}_1\,\boldsymbol{x}^{(k+1)} = \mathbf{M}_1^{\mathrm{T}}\boldsymbol{v}^{(k)} \iff \left(\mathbf{D}^{\mathrm{T}}\mathbf{D} + \mathbf{H}^{\mathrm{T}}\mathbf{H} + \mathbf{I}_n\right)\boldsymbol{x}^{(k+1)} = \mathbf{M}_1^{\mathrm{T}}\boldsymbol{v}^{(k)}. \quad (33)$$

The coefficient matrix in (33) has full rank independently of matrices $\mathbf{D}$ and $\mathbf{H}$ - i.e., of the finite difference discretization used for the gradient and of the imaging application considered - hence the solution of (32) is unique and reads

$$\boldsymbol{x}^{(k+1)} = \left(\mathbf{M}_1^{\mathrm{T}}\mathbf{M}_1\right)^{-1}\mathbf{M}_1^{\mathrm{T}}\boldsymbol{v}^{(k)}\,. \quad (34)$$

For the IR inverse problem, upon the assumption of space-invariant blur and periodic boundary conditions, the coefficient matrix in (33) is block-circulant with circulant blocks. Hence, the above linear system can be solved very efficiently by one application of the 2D Fast Fourier Transform (FFT) and one application of the inverse 2D FFT.

When addressing the CTIR problem, the structure of matrix $\mathbf{H}$ - which, we recall, in this case is a Radon matrix - does not allow for a Fourier diagonalization of matrix $\mathbf{M}_1^{\mathrm{T}}\mathbf{M}_1$, thus yielding a significative computational burden related to the solution of linear system (33). A popular strategy for avoiding such difficulty is the linearized ADMM. It relies on computing $\boldsymbol{x}^{(k+1)}$ as the global minimizer of a surrogate function $\widehat{Q}^{(k)}$ of $Q^{(k)}$ in (32), namely

$$\boldsymbol{x}^{(k+1)} = \underset{\boldsymbol{x}\,\in\,\mathbb{R}^n}{\arg\min}\ \widehat{Q}^{(k)}(\boldsymbol{x})\,, \quad (35)$$

where $\widehat{Q}^{(k)}$ is a quadratic function of the following form

$$\begin{aligned}\widehat{Q}^{(k)}(\boldsymbol{x}) \;&=\; Q^{(k)}(\boldsymbol{x}^{(k)}) + \langle \nabla Q^{(k)}(\boldsymbol{x}^{(k)}), \boldsymbol{x}-\boldsymbol{x}^{(k)}\rangle \\ &\quad + \frac{\eta}{2}\|\boldsymbol{x}-\boldsymbol{x}^{(k)}\|_2^2, \quad \eta \geq \|\mathbf{M}_1\|_2^2\,. \end{aligned} \quad (36)$$

It can be easily proved that any function $\widehat{Q}^{(k)}$ in (36) is a *quadratic tangent majorant* of the original function $Q^{(k)}$ in (32) at point $\boldsymbol{x}^{(k)}$, that is it satisfies

$$\widehat{Q}^{(k)}(\boldsymbol{x}^{(k)}) = Q^{(k)}(\boldsymbol{x}^{(k)}), \quad \nabla\widehat{Q}^{(k)}(\boldsymbol{x}^{(k)}) = \nabla Q^{(k)}(\boldsymbol{x}^{(k)}),$$
$$\widehat{Q}^{(k)}(\boldsymbol{x}) \geq Q^{(k)}(\boldsymbol{x})\ \forall\,\boldsymbol{x}\in\mathbb{R}^n\,.$$

It follows from (35)-(36) that the new iterate $\boldsymbol{x}^{(k+1)}$ computed by the linearized ADMM is given by

$$\boldsymbol{x}^{(k+1)} = \underset{\boldsymbol{x}\in\mathbb{R}^n}{\arg\min} \left\{ \langle \nabla Q^{(k)}(\boldsymbol{x}^{(k)}), \boldsymbol{x}\rangle + \frac{\eta}{2}\|\boldsymbol{x}-\boldsymbol{x}^{(k)}\|_2^2 \right\} \tag{37}$$

$$= \boldsymbol{x}^{(k)} - \frac{1}{\eta}\nabla Q^{(k)}(\boldsymbol{x}^{(k)}) \tag{38}$$

$$= \boldsymbol{x}^{(k)} - \frac{1}{\eta}\mathbf{M}_1^{\mathrm{T}}\left(\mathbf{M}_1\boldsymbol{x}^{(k)} - \boldsymbol{v}^{(k)}\right), \quad \eta \geq \|\mathbf{M}_1\|_2^2, \tag{39}$$

where in (37) we dropped the constant terms, in (38) we set $\boldsymbol{x}^{(k+1)}$ equal to the unique stationary point of the strongly convex cost function in (37) and, finally, in (39) we substituted the explicit expression of the gradient of the original cost function $Q^{(k)}$ defined in (32).

4.2. The $\boldsymbol{t}$*-subproblem*

Recalling definitions (26)-(28), the $\boldsymbol{t}$-subproblem in (30) reads

$$\begin{aligned}\boldsymbol{t}^{(k+1)} &\in \underset{\boldsymbol{t}\in\mathbb{R}^{m+3n}}{\arg\min} \left\{ C_2(\boldsymbol{t}) + \langle \boldsymbol{\rho}^{(k)}, \mathbf{M}_1\boldsymbol{x}^{(k+1)} - \boldsymbol{t}\rangle + \frac{\beta}{2}\|\mathbf{M}_1\boldsymbol{x}^{(k+1)} - \boldsymbol{t}\|_2^2 \right\} \\ &= \underset{\boldsymbol{t}\in\mathbb{R}^{m+3n}}{\arg\min} \left\{ C_2(\boldsymbol{t}) + \frac{\beta}{2}\|\boldsymbol{t}-\boldsymbol{q}^{(k)}\|_2^2 \right\}, \quad \boldsymbol{q}^{(k)} = \mathbf{M}_1\boldsymbol{x}^{(k+1)} + \frac{1}{\beta}\,\boldsymbol{\rho}^{(k)}. \end{aligned} \tag{40}$$

Then, by recalling the definition of function C_2 in (26) and introducing the vectors $\boldsymbol{\rho}_1^{(k)} \in \mathbb{R}^{2n}$, $\boldsymbol{\rho}_2^{(k)} \in \mathbb{R}^m$ and $\boldsymbol{\rho}_3^{(k)} \in \mathbb{R}^n$ such that $\boldsymbol{\rho}^{(k)} = \left(\boldsymbol{\rho}_1^{(k)}; \boldsymbol{\rho}_2^{(k)}; \boldsymbol{\rho}_3^{(k)}\right)$ and the vectors

$$\begin{gathered}\boldsymbol{q}_1^{(k)} = \mathbf{D}\boldsymbol{x}^{(k+1)} + \frac{1}{\beta}\boldsymbol{\rho}_1^{(k)} \in \mathbb{R}^{2n}, \quad \boldsymbol{q}_2^{(k)} = \mathbf{H}\boldsymbol{x}^{(k+1)} + \frac{1}{\beta}\boldsymbol{\rho}_2^{(k)} \in \mathbb{R}^m, \\ \boldsymbol{q}_3^{(k)} = \boldsymbol{x}^{(k+1)} + \frac{1}{\beta}\boldsymbol{\rho}_3^{(k)} \in \mathbb{R}^n,\end{gathered} \tag{41}$$

such that $\boldsymbol{q}^{(k)} = \left(\boldsymbol{q}_1^{(k)}; \boldsymbol{q}_2^{(k)}; \boldsymbol{q}_3^{(k)}\right)$, problem (40) can be equivalently written as

$$\boldsymbol{t}^{(k+1)} \in \underset{\boldsymbol{t}\in\mathbb{R}^{m+3n}}{\arg\min} \left\{ T_1(\boldsymbol{t}_1) + T_2(\boldsymbol{t}_2) + T_3(\boldsymbol{t}_3) \right\}, \text{ with:}$$

$$\begin{aligned} T_1(\boldsymbol{t}_1) &= \sum_{i=1}^{n} \|\boldsymbol{t}_{1,i}\|_2 + \frac{\beta}{2}\|\boldsymbol{t}_1 - \boldsymbol{q}_1^{(k)}\|_2^2, \\ T_2(\boldsymbol{t}_2) &= \mu\,\mathrm{KL}(\boldsymbol{g}(\boldsymbol{t}_2) + \boldsymbol{b}; \boldsymbol{y}) + \frac{\beta}{2}\|\boldsymbol{t}_2 - \boldsymbol{q}_2^{(k)}\|_2^2, \\ T_3(\boldsymbol{t}_3) &= \iota_{\mathbb{R}^n_+}(\boldsymbol{t}_3) + \frac{\beta}{2}\|\boldsymbol{t}_3 - \boldsymbol{q}_3^{(k)}\|_2^2. \end{aligned} \tag{42}$$

Therefore, the updates of variables $\boldsymbol{t}_1$, $\boldsymbol{t}_2$ and $\boldsymbol{t}_3$ can be addressed separately.

Update of $\boldsymbol{t}_1$. It comes from (42) that the update of $\boldsymbol{t}_1$ reads

$$\boldsymbol{t}_1^{(k+1)} = \underset{\boldsymbol{t}_1 \in \mathbb{R}^{2n}}{\arg\min} \left\{ \sum_{i=1}^{n} \left[\|\boldsymbol{t}_{1,i}\|_2 + \frac{\beta}{2} \left(\boldsymbol{t}_{1,i} - \boldsymbol{q}_{1,i}^{(k)} \right)^2 \right] \right\} . \tag{43}$$

Hence, problem (43) is separable into n independent 2-dimensional problems

$$\boldsymbol{t}_{1,i}^{(k+1)} = \underset{\boldsymbol{t}_{1,i} \in \mathbb{R}^{2n}}{\arg\min} \left\{ \|\boldsymbol{t}_{1,i}\|_2 + \frac{\beta}{2} \left(\boldsymbol{t}_{1,i} - \boldsymbol{q}_{1,i}^{(k)} \right)^2 \right\}, \quad i = 1, \ldots, n\,, \tag{44}$$

which represent the proximal map of the Euclidean norm function $\| \cdot \|_2$ in $\mathbb{R}^2$ calculated at points $\boldsymbol{q}_{1,i}^{(k)}$, $i = 1, \ldots, n$. Such a proximal map admits a well-known explicit expression which leads to the following closed-form solution of problem (44):

$$\boldsymbol{t}_{1,i}^{(k+1)} = \max \left\{ \left\| \boldsymbol{q}_{1,i}^{(k)} \right\|_2 - \frac{1}{\beta} \, , \, 0 \right\} \frac{\boldsymbol{q}_{1,i}^{(k)}}{\left\| \boldsymbol{q}_{1,i}^{(k)} \right\|_2}, \quad i = 1, \ldots, n\,. \tag{45}$$

where $0 \cdot \mathbf{0} \,/\, 0 = \mathbf{0}$ is assumed.

Update of $\boldsymbol{t}_2$. It follows from (42) that, after introducing the scalar $\tau = \mu/\beta$, the updated vector $\boldsymbol{t}_2^{(k+1)}$ is given by

$$\begin{aligned} \boldsymbol{t}_2^{(k+1)} &\in \underset{\boldsymbol{t}_2 \in \mathbb{R}^m}{\arg\min} \left\{ \tau \, \mathrm{KL}(\boldsymbol{g}(\boldsymbol{t}_2) + \boldsymbol{b}; \boldsymbol{y}) + \frac{1}{2} \|\boldsymbol{t}_2 - \boldsymbol{q}_2^{(k)}\|_2^2 \right\} \\ &= \underset{\boldsymbol{t}_2 \in \mathbb{R}^m}{\arg\min} \left\{ \sum_{i=1}^{m} \left[\tau \, g(t_i) - \tau \, y_i \ln \left(g(t_i) + b_i \right) + \frac{1}{2} \left(t_i - q_i \right)^2 \right] \right\}, \end{aligned} \tag{46}$$

where in (46) we substituted the explicit expression of the KL divergence term reported in (3), we dropped the constants and, for simplicity of notation, we set $t_i := t_{2,i} \in \mathbb{R}$ and $q_i = q_{2,i}^{(k)} \in \mathbb{R}$. Hence, similarly to the $\boldsymbol{t}_1$ update problem in (43), the m-dimensional minimization problem (46) is equivalent to the m following 1-dimensional problems

$$t_i^{(k+1)} = \underset{t_i \in \mathbb{R}}{\arg\min} \left\{ \tau \, g(t_i) - \tau \, y_i \ln \left(g(t_i) + b_i \right) + \frac{1}{2} \left(t_i - q_i \right)^2 \right\}, \tag{47}$$

$i = 1, \ldots, m$.

In the IR scenario, i.e. when $g(t_i) = t_i$, the cost function in (47) is infinitely many times differentiable, strictly convex and coercive in its domain

$t_i \in (-b_i, +\infty)$. Hence, the solution $t_i^{(k+1)}$ of (47) exists, is unique and coincides with the unique stationary point of the cost function, given by

$$t_i^{(k+1)} = \frac{1}{2}\left[-(\tau + b_i - q_i) + \sqrt{(\tau + b_i - q_i)^2 + 4\,(q_i\,b_i + \tau(y_i - b_i))}\right]. \quad (48)$$

For the CTIR problem, i.e. when $g(t_i) = I_0 e^{-t_i}$, problem (47) reads

$$t_i^{(k+1)} = \underset{t_i \in \mathbb{R}}{\arg\min} \left\{ \tau\, I_0\, e^{-t_i} - \tau\, y_i \ln\left(I_0\, e^{-t_i} + b_i\right) + \frac{1}{2}(t_i - q_i)^2 \right\}. \quad (49)$$

The cost function in (49) is infinitely many times differentiable and coercive in its domain $t_i \in \mathbb{R}$, hence it admits global minimizers. However, in the general case of a nonzero background, i.e. when $b_i \in \mathbb{R}_{++}$, problem (49) does not admit a closed-form solution and can only be addressed by employing iterative solvers.

On the other hand, when $b_i = 0$ the cost function is also strictly convex, hence $t_i^{(k+1)}$ in (49) is given by the unique solution of the first-order optimatily condition

$$-\tau\, I_0\, e^{-t_i} + \tau\, y_i + t_i - q_i = 0\,.$$

The above nonlinear equation can be manipulated so as to give

$$w_i\, e^{w_i} = \tau\, I_0\, e^{\tau\, y_i - q_i}\,, \quad \text{with } \; w_i = t_i + \tau\, y_i - q_i\,. \quad (50)$$

Equations of the form in (50) admit solutions that can be expressed in closed-form in terms of the so-called Lambert W function [17]. In particular, when the right-hand side is non-negative - which is our case as $\tau\, I_0 e^{\tau\, y_i - q_i} \in \mathbb{R}_{++}$ - then the equation admits a unique solution given by

$$w_i = W\left(\tau\, I_0\, e^{\tau\, y_i - q_i}\right).$$

It follows that problem (49) admits the unique solution

$$t_i^{(k+1)} = -(\tau\, y_i - q_i) + W\left(\tau\, I_0\, e^{\tau\, y_i - q_i}\right). \quad (51)$$

Update of $\boldsymbol{t}_3$. It comes from (42) that the $\boldsymbol{t}_3$-update problem reads

$$\boldsymbol{t}_3^{(k+1)} \in \underset{\boldsymbol{t}_3 \in \mathbb{R}^n_+}{\arg\min} \|\boldsymbol{t}_3 - \boldsymbol{q}_3^{(k)}\|_2^2\,,$$

that is $\boldsymbol{t_3}^{(k+1)}$ is given by the unique Euclidean projection of vector $\boldsymbol{q}_3^{(k)}$ onto the non-negative orthant $\mathbb{R}^n_+$, which admits the following component-wise closed-form expression:

$$t_{3,i}^{(k+1)} = \max\left\{q_{3,i}^{(k)}, 0\right\}, \quad i = 1, \ldots, n\,. \tag{52}$$

4.3. Overall PWP-based reconstruction approach

In Algorithm 1 we report the main computational steps by which the proposed (PWP) approach can be applied to select the regularization parameter μ in the (TV-KL) variational model so as to obtain high-quality reconstructions in IR and CTIR imaging problems.

First, the function `SolveTV-KL` implements the ADMM-based approach outlined in previous subsections: for any given value of the regularization parameter μ, an estimate of the μ-dependent solution $\boldsymbol{x}^*(\mu)$ of the (TV-KL) model is obtained by means of a two-blocks standard (for IR) or semi-linearized (for CTIR) iterative ADMM scheme. We notice that the standard scheme presented is equivalent to the so-called PIDSplit+ algorithm proposed in [18], which is proved to converge for any value $\beta \in \mathbb{R}_{++}$ of the ADMM penalty parameter. The semi-linearized scheme proposed for CTIR also converges for any $\beta \in \mathbb{R}_{++}$ as it belongs to the class of semi-linearized ADMM schemes whose convergence has been proved, e.g., in [19]. We remark that, as suggested in [18], in the last line of function `SolveTV-KL`, the estimated solution $\boldsymbol{x}^*(\mu)$ is not set equal to the last iterate $\boldsymbol{x}^{(k+1)}$ (as it would be natural to do) but to the last iterate $\boldsymbol{t}_3^{(k+1)}$. This in fact allows us to guarantee that the output image $\boldsymbol{x}^*(\mu)$ has non-negative intensities at all pixels (thus satisfying the non-negativity constraint on $\boldsymbol{x}$), as $\boldsymbol{t}_3^{(k+1)}$ is obtained by projecting $\boldsymbol{x}^{(k+1)}$ onto the non-negative orthant - see (52).

In the main body of Algorithm 1 we outline the main computational steps by which the proposed PWP approach can be applied in practice: for a set μ_j, $j = 1, 2, \ldots$, of different μ-values the reconstructions $\boldsymbol{x}^*(\mu_j)$ are computed by function `SolveTV-KL`, then the associated whiteness function values $W(\mu_j)$ are calculated and, finally, the value μ^* - and, hence, the corresponding reconstruction $\boldsymbol{x}^*(\mu^*)$ - is selected as the μ_j yielding the minimum $W(\mu_j)$.

In Algorithm 1, we do not formalize a specific strategy for the (a-priori fixed or online adaptive) selection of the μ_j values to test, as this is not the focus of this paper. In fact, in the experimental section we will simply test a fine uniform grid of μ_j values between a minimum $\mu_{\min}$ and a maximum $\mu_{\max}$, after making sure (empirically) that the minimum of the computed whiteness function values $W(\mu_j)$ is achieved at some μ_j different from the extremes $\mu_{\min}$ and $\mu_{\max}$. However, we think that this issue deserves some more words as a suitable selection strategy together with nice properties of the whiteness function $W(\mu)$ to minimize can make the PWP a fully-automatic approach.

After recalling the matrix notations outlined in Section 2 and introducing the function $F : \mathbb{R}^{n_1 \times n_2} \to \mathbb{R}_+$ defined by

$$F(\mathbf{X}) := \mathcal{W}\left(\frac{\mathbf{Y} - \mathbf{G}\left(\mathbf{H}\left(\mathbf{X}\right)\right) + \mathbf{B}}{\sqrt{\mathbf{G}\left(\mathbf{H}\left(\mathbf{X}\right)\right) + \mathbf{B}}}\right), \tag{53}$$

with function $\mathcal{W}$ defined in (10) and where the operators $\sqrt{\cdot}$ and $\div$ are to be intended component-wise, the (PWP) can be equivalently rewritten as follows

$$\mu^* \in \underset{\mu \in \mathbb{R}_{++}}{\arg\min} \left\{ W(\mu) \, = \, F\left(\mathbf{X}^*(\mu)\right) \right\}, \tag{54}$$

with $\mathbf{X}^*(\mu)$ the matrix form of the solution function $\boldsymbol{x}^*(\mu)$ of the (TV-KL) model, which is well-known not to admit an explicit expression. Hence, the cost function $W(\mu)$ we aim to minimize in the reformulated PWP in (54) can be evaluated for any given positive μ-value (by first solving numerically the associated TV-KL model and then directly applying the function F in (53) to the result) but does not have an explicit form and, hence, the derivatives can not be calculated. A theoretical analysis of the properties of function W is thus a very hard task and is out of the scope of this paper. However, all the numerical tests that we will present in the next section seem to indicate that W is continuous (if not differentiable) and unimodal with a unique well-shaped global minimizer. Hence, by applying some convergent zero-order minimization algorithm - i.e., relying on evaluations of the cost function but not of its derivatives - for unimodal functions of one variable, the PWP approach could be made fully automatic.

Algorithm 1: Overall (PWP)-based reconstruction approach for the IR and CTIR problems based on the (TV-KL) variational model

data: observed degraded image $\mathbf{y} \in \mathbb{R}^m$,
background emission image $\boldsymbol{b} \in \mathbb{R}^m$,
linear/nonlinear operators $\mathbf{H} \in \mathbb{R}^{m \times n}$, $\boldsymbol{g} : \mathbb{R}^m \to \mathbb{R}^m$ (see (7))

output: (PWP)-selected μ^*-value and associated reconstruction $\boldsymbol{x}^*(\mu^*)$

1. **for** *j = 1, 2, 3, ...* **do**:
2. · for the current value μ_j of regulariation parameter **do**:
3. · compute $\boldsymbol{x}^*(\mu_j) = \texttt{SolveTV-KL}(\boldsymbol{y}, \boldsymbol{b}, \mathbf{H}, \boldsymbol{g}, \mu_j)$
4. · compute $\boldsymbol{\Lambda}^*(\mu_j)$ by (21) and then $\mathbf{Z}^*(\mu_j)$ by (22)
5. · compute $W(\mu_j) = \mathcal{W}(\mathbf{Z}^*(\mu_j))$ by (11)
6. **end for**
7. compute $\mu^* = \underset{\mu=\mu_1,\mu_2,\ldots}{\arg\min}\, W(\mu)$ and the associated reconstruction $\boldsymbol{x}^*(\mu^*)$

==========

function $\boldsymbol{x}^*(\mu) = \texttt{SolveTV-KL}(\boldsymbol{y}, \boldsymbol{b}, \mathbf{H}, \boldsymbol{g}, \mu)$

1. **initialize:** set $\boldsymbol{x}^{(0)} = \boldsymbol{b}$, $\boldsymbol{\rho}^{(0)} = \mathbf{0}_{m+3n}$, $\boldsymbol{t}^{(0)} = \mathbf{M}_1 \boldsymbol{x}^{(0)}$
2. **for** *k = 0, 1, 2, ... until convergence* **do**:
3. · compute $\boldsymbol{x}^{(k+1)}$ by (32) and then (34) (for IR) or (39) (for CTIR)
4. · compute $\boldsymbol{t}^{(k+1)}$ by (41) and then:
5. · (45) to compute $\boldsymbol{t}_1^{(k+1)}$
6. · (48) (for IR) or (51) (for CTIR) to compute $\boldsymbol{t}_2^{(k+1)}$
7. · (52) to compute $\boldsymbol{t}_3^{(k+1)}$
8. · compute $\boldsymbol{\rho}^{(k+1)}$ by (31)
9. **end for**
10. set $\boldsymbol{x}^*(\mu) = \boldsymbol{t}_3^{(k+1)}$

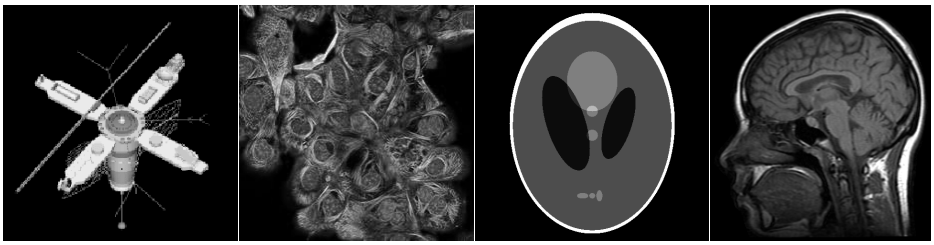
5. Computed examples

In this section, we evaluate the performance of the proposed (PWP) for the selection of the regularization parameter μ in the (TV-KL) model employed for the solution of the IR and CTIR imaging problems. For a quantitative evaluation, the accuracy of the reconstructed images $\boldsymbol{x}^*(\mu)$ with respect to the original image $\bar{\boldsymbol{x}}$ is measured by means of two scalar metrics, the Structural Similarity Index (SSIM) [20] and the Signal-to-Noise-Ratio (SNR) defined by

$$\mathrm{SNR}(\boldsymbol{x}^*, \bar{\boldsymbol{x}}) = 10 \log_{10} \frac{||\bar{\boldsymbol{x}} - \mathrm{E}[\bar{\boldsymbol{x}}]||_2^2}{||\bar{\boldsymbol{x}} - \boldsymbol{x}^*||_2^2},$$

where $\mathrm{E}[\bar{\boldsymbol{x}}]$ denotes the mean intensity of the original image $\bar{\boldsymbol{x}}$.

The proposed PWP approach is compared with the state-of-the-art ADP and NEDP selection criteria proposed in [4] and [2], respectively. The considered parameter selection rules are all applied *a posteriori*. The (TV-KL) model is solved by the ADMM approach outlined in Section 4 for a very fine grid of different μ-values and for each reconstructed image $\boldsymbol{x}^*(\mu)$ we compute both the value of the discrepancy function $\mathcal{D}(\mu; \boldsymbol{y})$ defined in (6) and involved in the (ADP) and (NEDP) and the value of the whiteness function $W(\mu)$ defined and used in (PWP), as outlined in Algorithm 1. In particular, we calculate $W(\mu)$ efficiently based on the Fourier-domain formula in (11). Then, we select the μ-values according to the (ADP) and (NEDP), denoted by $\mu^{(A)}$ and $\mu^{(NE)}$, as the ones corresponding to the intersection of $\mathcal{D}(\mu; \boldsymbol{y})$ with $\Delta^{(A)}$ and $\Delta^{(NE)}(\mu)$, respectively. The μ-value selected by the (PWP) is the one minimizing the function $W(\mu)$ and is denoted by $\mu^{(W)}$. For each μ-value in the considered grid, we also compute the SNR and SSIM of the associated reconstructed image $\boldsymbol{x}^*(\mu)$. Moreover, to evaluate in absolute terms the performance of the three compared selection criteria, we also compute the values $\mu^{(SNR)}$ and $\mu^{(SSIM)}$ of the regularization parameter which yield the reconstructed images exhibiting the highest SNR and the highest SSIM values, respectively. The two reconstructed images $\boldsymbol{x}^*(\mu^{(SNR)})$, $\boldsymbol{x}^*(\mu^{(SSIM)})$ and the associated SNR and SSIM quality metrics are then regarded as the best theoretical results achievable by the compared selection strategies.



(a) (b) (c) (d)

Figure 1: From left to right: original `satellite` (256×256), `cells` (236×236), `shepp logan` (500×500) and `brain` (238×253) test images considered for the numerical experiments.

In all the performed tests, the iterations of the ADMM approach outlined in Algorithm 1 and used for the solution of the TV-KL model are stopped (in both the standard form for IR and semi-linearized form for CTIR) as soon as

$$\delta_{\boldsymbol{x}}^{(k)} := \frac{\|\boldsymbol{x}^{(k)} - \boldsymbol{x}^{(k-1)}\|_2}{\|\boldsymbol{x}^{(k-1)}\|_2} < 10^{-6}\,, \qquad k \in \mathbb{N} \setminus \{0\}\,,$$

while the ADMM penalty parameter β is set manually so as to fasten the convergence of the alternating scheme. More specifically, we made some prior numerical tests and found that a suitable value in the IR case is $\beta = 5$, while for the CTIR we set $\beta = 10$.

All tests have been performed in Matlab R2022b, on a Windows 10 Platform.

5.1. Image restoration

We start testing our proposal on the image restoration task, and consider two test images, namely `satellite` (256×256 pixels) and `cells` (236×236 pixels), with pixel values between 0 and 1, shown in Figures 1a, 1b.

We simulate the acquisition process by first multiplying the original images by a factor $\kappa \in \mathbb{R}_{++}$ representing the maximum number of photons hitting the image domain, in expectation. Clearly, the lower the value of κ the more noisy the data, yielding a more difficult image restoration problem. Then, the resulting images are 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 our tests, we set `band` $= 5$, `sigma` $= 1$. Then, we add a constant emission background $\boldsymbol{b}$ equal to 2×10^{-3} thus obtaining the noise-free degraded image $\bar{\boldsymbol{\lambda}} = \mathbf{H}\bar{\boldsymbol{x}} + \boldsymbol{b}$. Finally, the observed image $\boldsymbol{y} = \mathbf{poiss}(\bar{\boldsymbol{\lambda}})$ is pseudo-randomly generated by sampling from a m-variate independent Poisson random process with mean vector $\bar{\boldsymbol{\lambda}}$, using the Matlab command `poissrnd`.

In Figure 2c,d we show the whiteness function $W(\mu)$, as defined in PWP, for the first image `satellite` and $\kappa = 5$ (left) and $\kappa = 10$ (right). The vertical dashed red lines correspond to the minimum of the function $W(\mu)$, i.e. to the value $\mu^{(W)}$ selected by the proposed (PWP).

The black curves in Figure 2a,b represent the discrepancy function $\mathcal{D}(\mu; \boldsymbol{y})$ as defined in (6), while the green and magenta dashed lines represent the discrepancy values $\Delta^{(A)}$ and $\Delta^{(NE)}(\mu)$ as defined in ADP and NEDP, respectively.

In Figure 2e,f, we show the SNR (in blue) and SSIM (in orange) values achieved for different μ values with $\kappa = 5, 10$. The red, green and magenta vertical lines correspond to the μ values chosen with the newly proposed method and the two considered versions of the DP. Note that, in the low-counting regime, the PWP achieves higher values of SNR and SSIM if compared to the ADP and NEDP.

In Table 1, we report, for different counting regimes κ, the selected μ-values and the SNR/SSIM metrics for the three considered strategies as well as for the cases of maximum SNR and SSIM. For each κ, the highest values of SNR and SSIM achieved among the three compared methods, i.e. the closest to the maximum achievable, are reported in bold. As already observed in Figure 2e,f, the PWP outperforms the ADP and NEDP in terms of SNR and SSIM for the low/middle counts acquisitions (up to $\kappa = 50$); in such scenarios, the obtained metrics are also particularly close to the highest SNR and SSIM achieved. For the higher counts NEDP and PWP return similar quality measures, with NEDP being slightly better.

	κ	1.5	5	10	20	50	100	1000
ADP	$\mu^{(A)}$	2×10^{-5}	0.065	0.068	0.188	0.380	0.760	8.260
	SNR	-0.001	3.408	3.508	6.580	8.688	10.574	14.747
	SSIM	0.009	0.625	0.619	0.708	0.724	0.742	0.805
NEDP	$\mu^{(NE)}$	0.841	1.205	2.348	3.848	6.800	11.380	60.760
	SNR	10.270	11.286	12.384	13.179	14.206	**15.017**	**17.540**
	SSIM	0.786	0.779	0.787	0.792	0.805	**0.823**	**0.862**
PWP	$\mu^{(W)}$	1.201	2.045	3.068	4.388	7.460	11.080	45.220
	SNR	**10.618**	**11.944**	**12.719**	**13.328**	**14.313**	14.983	17.225
	SSIM	**0.787**	**0.785**	**0.791**	**0.794**	**0.808**	0.822	0.857
SNR_{max}	$\mu^{(SNR)}$	1.561	3.845	5.948	10.841	20.563	35.562	251.685
	SNR	10.699	12.363	13.132	13.872	14.978	15.870	18.367
	SSIM	0.786	0.788	0.796	0.805	0.822	0.851	0.863
SSIM_{max}	$\mu^{(SSIM)}$	1.141	3.065	6.887	12.325	19.675	33.315	129.476
	SNR	10.589	12.302	13.105	13.851	14.977	15.860	18.130
	SSIM	0.787	0.789	0.797	0.806	0.823	0.852	0.866

Table 1: Test image `satellite`. Output μ-values and SNR/SSIM metrics for the restoration by ADP, NEDP, PWP and for the output restorations corresponding to the maximum SNR and SSIM achieved, for different κ.

For a visual comparison, in Figure 3, we show the observed images, the output restorations obtained by employing ADP, NEDP and PWP, and the restorations corresponding to the maximum SNR and SSIM for $\kappa = 5$ (top panels) and $\kappa = 10$ (bottom panels). In both cases, the NEDP and the PWP return similar results, with the latter being more capable of preserving the original contrast in the image. On the other hand, the output images obtained by selecting μ according to ADP are strongly over-regularized.

For the second test image, `cells`, we report in Figure 4 the behavior of the discrepancy function $\mathcal{D}(\mu, \boldsymbol{y})$, of the whiteness function $W(\mu)$ and of the SNR/SSIM curves obtained by applying the NEDP, the ADP and the PWP, for $\kappa = 5$ (left) and $\kappa = 10$ (right). The PWP returns larger quality measures, as it is the closest to the maximum SNR/SSIM achievable.

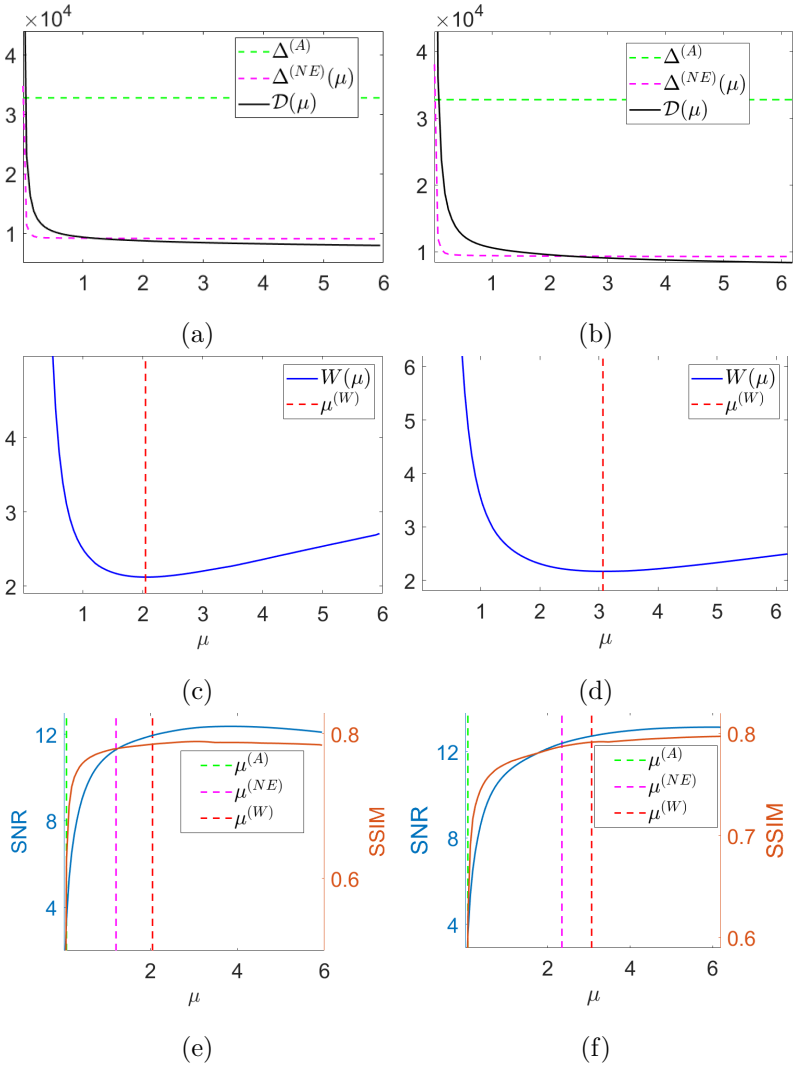


Figure 2: Test image `satellite`. From top to bottom: discrepancy curves, whiteness curves and achieved SNR/SSIM for $\kappa = 5$ (left) and $\kappa = 10$ (right).

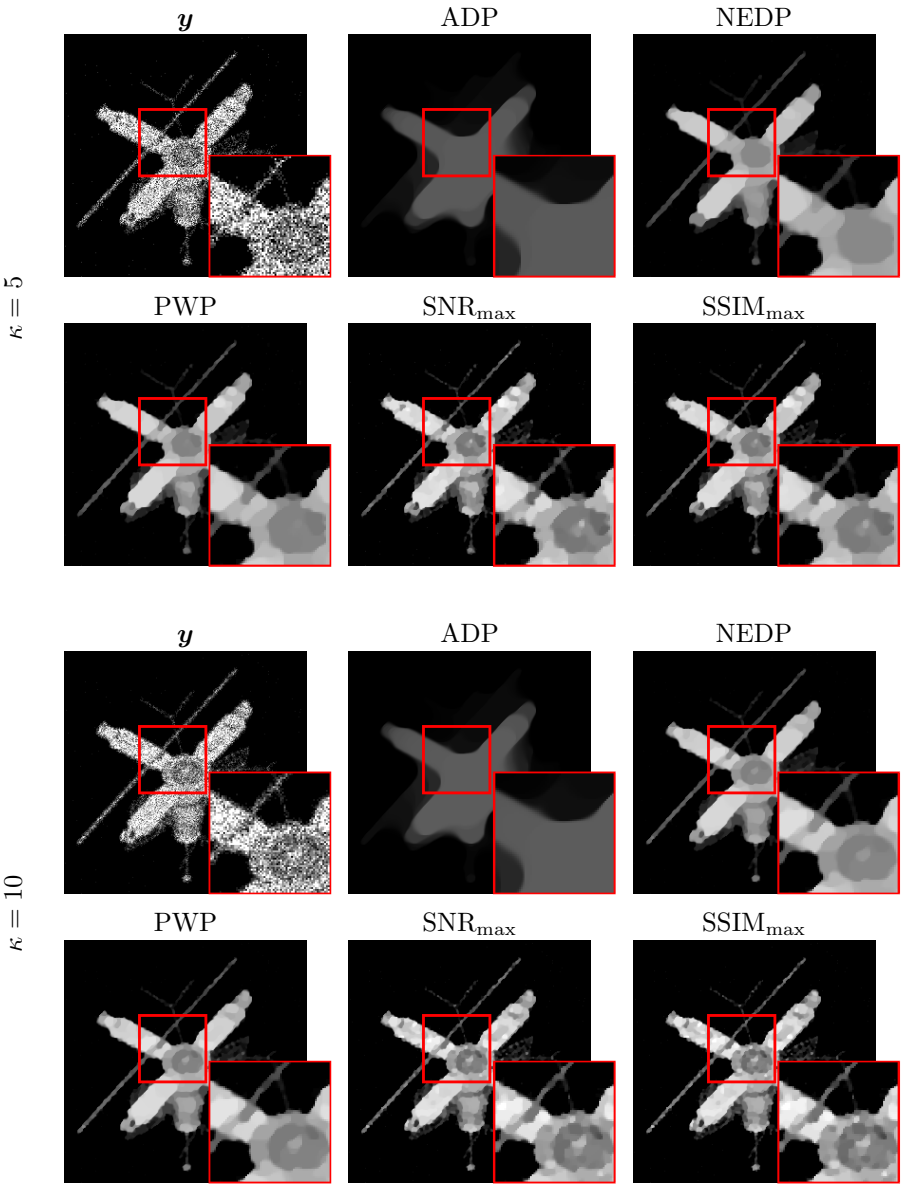


Figure 3: Test image `satellite`. Observed image $\boldsymbol{y}$, restorations using the ADP, NEDP and PWP, and restorations corresponding to the maximum SNR and SSIM for $\kappa = 5$ (top panels) and $\kappa = 10$ (bottom panels).

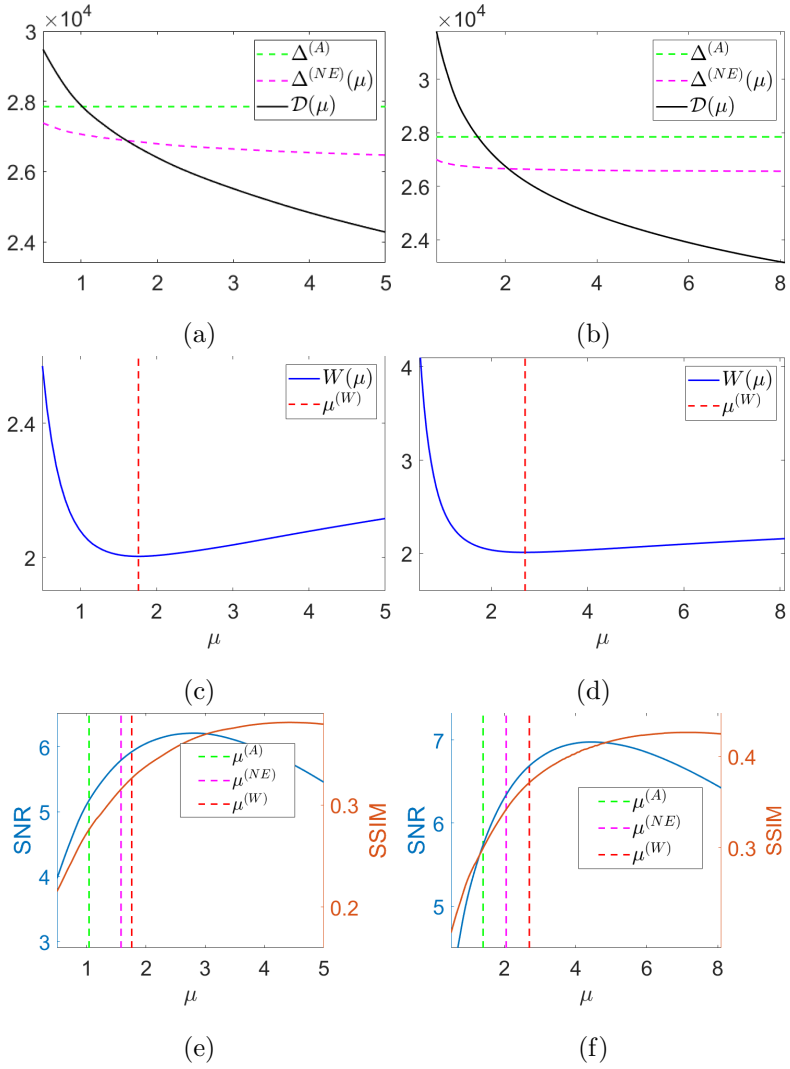


Figure 4: Test image `cells`. From top to bottom: discrepancy curves, whiteness curves and achieved SNR/SSIM for $\kappa = 5$ (left) and $\kappa = 10$ (right).

From Table 2, we note that the proposed μ-selection criterion returns restored images outperforming the ones obtained via the NEDP and ADP both in terms of SNR and SSIM, for every $\kappa \geq 5$. For $\kappa = 1.5$ the SNR and SSIM values of the PWP restoration are slightly lower, but very similar, to the one obtained with NEDP, while in all the other cases the difference between the PWP and the NEDP is more marked. In general, the percentage difference between the quality metrics achieved by the PWP and the ones corresponding to the maximum SNR and SSIM values appear to be very small, especially in low-counting regimes.

The restored images in Figure 5 reflect the values recorded in the tables: the output of the PWP preserve more details and the original contrast if compared to NEDP, while the ADP restoration seems to be less subject to over-regularization if compared to the results obtained on the test image `satellite`. This can be ascribed to the number of zeros in the image, being significantly smaller in `cells`.

5.2. CT image reconstruction

For the CT reconstruction problem we consider the test images `shepp logan` (500×500, pixel size = 0.2mm) and `brain` (238×253, pixel size=0.4mm), with pixel values between 0 and 1, shown Figures 1c, 1d, respectively. The acquisition process of the fan beam CT setup, i.e. the projection operator $\mathbf{H}$, is built using the ASTRA Toolbox [21] with the following parameters: 180 equally spaced angles of projections (from 0 to 2π), a detector with 500 pixels (detector pixel size = 1/3mm), distance between the source and the center of rotation = 300mm, distance between the center of rotation and the detector array = 200mm. Then, according to (7), we take the exponential of $-\mathbf{H}\bar{\boldsymbol{x}}$ and multiply it by a factor $I_0 \in \mathbb{N} \setminus \{0\}$ that plays the role of κ in the restoration scenario and represents the maximum emitted photon counts, i.e., the maximum number of photons that can reach each detector pixel if the X-rays are not attenuated. In the CT tests, we consider the background emission $\boldsymbol{b} = \mathbf{0}$ so that the solution of (50) can be expressed in closed-form in terms of the Lambert function. We thus

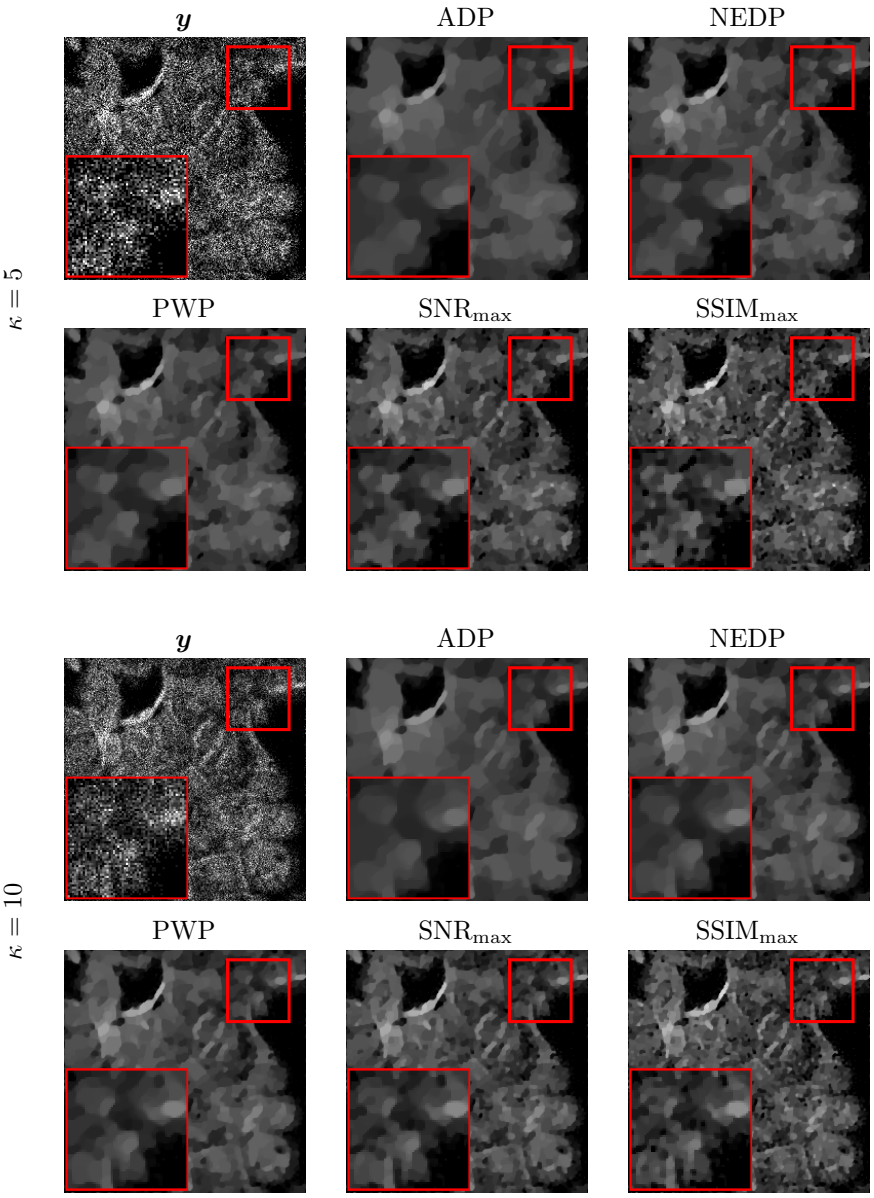


Figure 5: Test image `cells`. Observed image $\boldsymbol{y}$, restorations using the ADP, NEDP and PWP, and restorations corresponding to the maximum SNR and SSIM for $\kappa = 5$ (top panels) and $\kappa = 10$ (bottom panels).

	κ	1.5	5	10	20	50	100	1000
ADP	$\mu^{(A)}$	7×10^{-5}	1.040	1.400	2.280	4.500	7.920	45.000
	SNR	0.004	5.176	5.737	6.626	7.830	8.735	11.075
	SSIM	0.077	0.276	0.299	0.363	0.452	0.532	0.717
NEDP	$\mu^{(NE)}$	0.875	1.580	2.060	3.600	6.600	10.680	52.140
	SNR	**4.625**	5.794	6.347	7.316	8.337	9.071	11.207
	SSIM	**0.313**	0.315	0.342	0.417	0.493	0.560	0.730
PWP	$\mu^{(W)}$	0.850	1.760	2.720	4.200	7.440	12.000	54.660
	SNR	4.601	**5.924**	**6.695**	**7.503**	**8.470**	**9.186**	**11.248**
	SSIM	0.311	**0.326**	**0.372**	**0.433**	**0.505**	**0.571**	**0.733**
SNR_{max}	$\mu^{(SNR)}$	1.193	2.840	4.285	7.784	15.332	24.675	179.852
	SNR	4.685	6.212	6.973	7.826	8.904	9.590	11.841
	SSIM	0.332	0.365	0.408	0.479	0.556	0.617	0.767
SSIM_{max}	$\mu^{(SSIM)}$	1.732	4.463	6.857	10.562	22.678	33.119	182.857
	SNR	4.064	5.766	6.700	7.659	8.751	9.531	11.840
	SSIM	0.340	0.380	0.426	0.487	0.563	0.623	0.768

Table 2: Test image `cells`. Output μ-values and SNR/SSIM metrics for the restoration by ADP, NEDP, PWP and for the output restorations corresponding to the maximum SNR and SSIM achieved, for different κ.

compute the noise-free data $\bar{\boldsymbol{\lambda}} = I_0 e^{-\mathbf{H}\bar{\boldsymbol{x}}}$, while the observation $\boldsymbol{y} = \mathbf{poiss}(\bar{\boldsymbol{\lambda}})$ is obtained - like for IR - by sampling from a m-variate independent Poisson random process with mean vector $\bar{\boldsymbol{\lambda}}$.

In analogy to the restoration case, in Figure 6 we report for the test image `shepp logan` the curve of the discrepancy function $\mathcal{D}(\mu, \boldsymbol{y})$, as well as the whiteness curve $W(\mu)$ and the curves of the SNR and SSIM for the limiting values of I_0, i.e. $I_0 = 1.5$ (left) and $I_0 = 1000$ (right). In the case of $I_0 = 1.5$ the SNR/SSIM values achieved by ADP and NEDP are significanlty far from the optimal ones. On the other hand, PWP one is very close to the maximum of both the SNR and the SSIM. For $I_0 = 1000$, the NEDP and the ADP select the same μ, which allows to achieve a larger SSIM with respect to the one obatined by PWP, while our method still outperforms the other in terms of SNR.

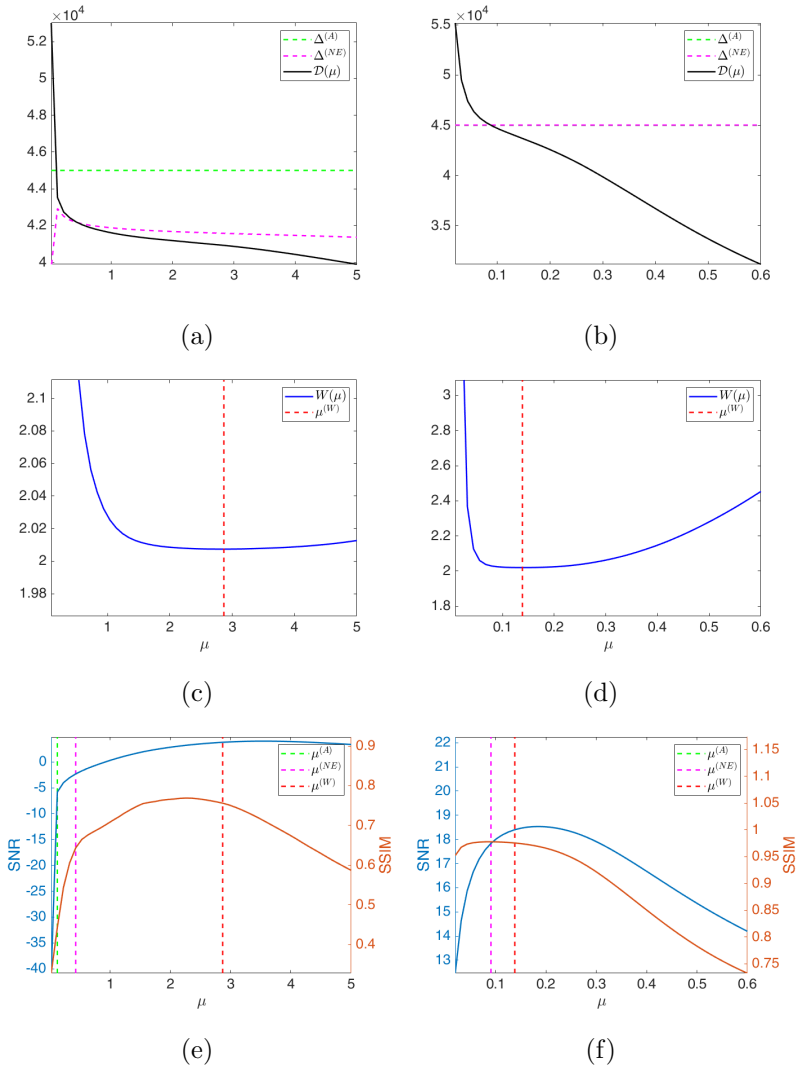


Figure 6: Test image `shepp logan`. From top to bottom: discrepancy curves, whiteness curves and achieved SNR/SSIM for $I_0 = 1.5$ (left) and $I_0 = 1000$ (right).

From Table 3, we observe that the PWP outperforms the ADP and the NEDP in terms of SNR for each I_0 value, while the NEDP returns slightly better results in terms of SSIM for high-count acquisitions. We highlight that for mid- and high-doses, the quality metrics of the reconstructions by PWP are particularly close to the highest achievable.

The reconstruction results shown in Figure 7 reflect the behavior of the plots. More specifically, for $I_0 = 1.5$ the ADP reconstruction appears to be over-regularized; NEDP allows to reconstruct only the central ellipsis, which appear to be merged; finally, in the PWP reconstruction the two ellipsis are more visible and the white edge of the phantom is sharper. In the case of $I_0 = 1000$, the three reconstructions are similar, with the PWP being more capable of separating the three fine details highlighted in the super-imposed close-up.

For the last test image, `brain`, we show in Figure 8 the behaviour of the discrepancy function $\mathcal{D}(\mu, \boldsymbol{y})$, of the whiteness function $W(\mu)$, as well as of the SNR and SSIM values for $I_0 = 1.5$ and $I_0 = 1000$. From Table 4, we note that the PWP achieves higher SNR and SSIM values compared to the ADP and NEDP for lower values of I_0. However, we observe that when considering higher values of I_0, the ADP reconstruction can outperform PWP for some of the considered doses. Moreover, notice that for some doses in the high-dose regime - e.g., $\kappa = 50, 1000$ - the PWP returns quality measures almost coinciding with the maximum SNR and SSIM achieved.

The reconstruction computed by ADP, NEDP and PWP are shown in Figure 9: we can see a higher level of details in the PWP reconstruction, both in the the low-dose and high-dose case. For $I_0 = 1.5$, only PWP is able to recover the upper part of the skull bone, while for $I_0 = 1000$ the difference mainly concerns the level of details present in the reconstruction, as shown in the close-ups.

	κ	1.5	5	10	20	50	100	1000
ADP	$\mu^{(A)}$	0.122	5.555	3.351	1.522	0.564	0.322	0.091
	SNR	-5.838	2.899	4.449	8.976	**11.626**	13.141	17.837
	SSIM	0.443	0.455	0.515	0.755	**0.974**	0.944	0.977
NEDP	$\mu^{(NE)}$	0.426	0.684	0.733	0.530	0.352	0.261	0.091
	SNR	-2.329	3.569	6.403	8.550	10.717	12.698	17.837
	SSIM	0.643	0.793	0.853	**0.889**	0.992	**0.945**	**0.977**
PWP	$\mu^{(W)}$	2.865	1.363	1.024	0.861	0.564	0.442	0.138
	SNR	**3.785**	**5.997**	**7.441**	**9.786**	**11.626**	**13.436**	**18.401**
	SSIM	**0.756**	**0.816**	**0.856**	0.883	**0.974**	0.935	0.975
SNR_{max}	$\mu^{(SNR)}$	3.470	2.042	1.461	1.001	0.654	0.443	0.186
	SNR	4.008	6.667	7.953	9.888	11.682	13.436	18.534
	SSIM	0.717	0.782	0.825	0.868	0.902	0.936	0.969
SSIM_{max}	$\mu^{(SSIM)}$	2.256	1.364	0.879	0.641	0.382	0.262	0.091
	SNR	3.202	5.997	7.009	9.165	10.945	12.698	17.838
	SSIM	0.769	0.817	0.858	0.892	0.919	0.946	0.978

Table 3: Test image `shepp logan`. Output μ-values and SNR/SSIM metrics for the restoration by ADP, NEDP, PWP and for the output restorations corresponding to the maximum SNR and SSIM achieved, for different κ.

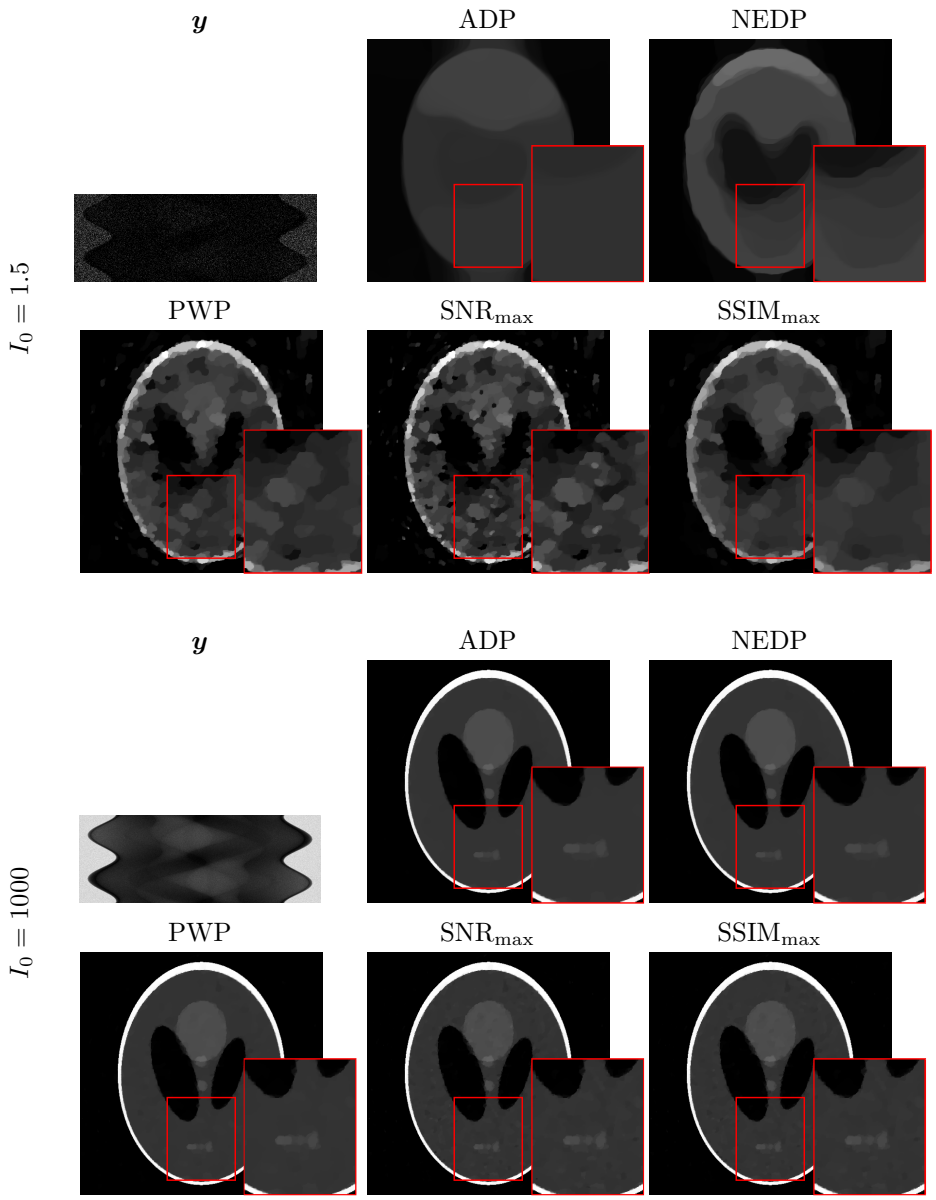


Figure 7: Test image `shepp logan`. Observed image $\boldsymbol{y}$, restorations using the ADP, NEDP and PWP, and restorations corresponding to the maximum SNR and SSIM for $I_0 = 1.5$ (top panels) and $I_0 = 1000$ (bottom panels).

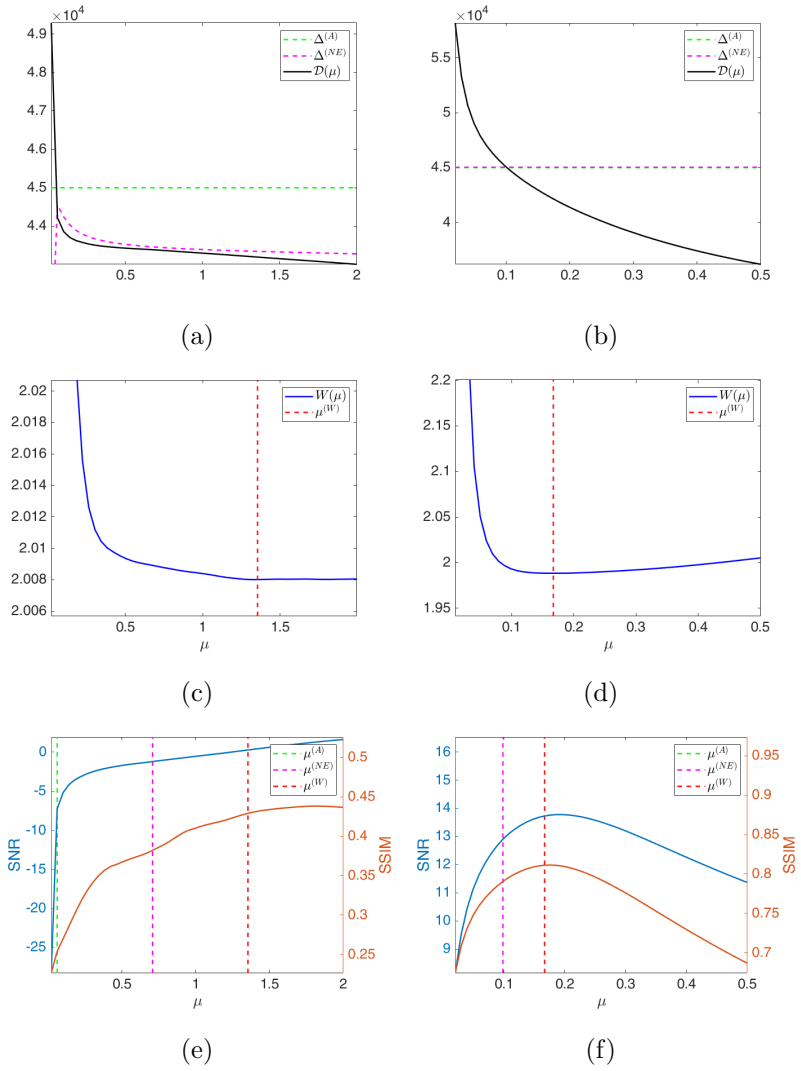


Figure 8: Test image **brain**. From top to bottom: discrepancy curves, whiteness curves and achieved SNR/SSIM for $I_0 = 1.5$ (left) and $I_0 = 1000$ (right).

	κ	1.5	5	10	20	50	100	1000
ADP	$\mu^{(A)}$	0.060	2.657	2.428	1.302	0.516	0.300	0.098
	SNR	-7.251	3.308	3.432	**6.154**	8.162	**8.970**	12.901
	SSIM	0.254	0.323	0.311	0.466	**0.622**	**0.672**	0.790
NEDP	$\mu^{(NE)}$	0.706	0.428	0.542	0.383	0.320	0.257	0.098
	SNR	-1.200	0.468	3.340	4.714	7.636	8.580	12.901
	SSIM	0.382	**0.432**	0.523	0.556	0.612	0.666	0.790
PWP	μ^{W}	1.853	2.200	0.771	0.420	0.589	0.286	0.166
	SNR	**0.278**	**3.866**	**4.472**	5.040	**8.310**	8.854	**13.728**
	SSIM	**0.429**	0.381	**0.534**	**0.562**	0.614	0.670	**0.811**
SNR_{max}	$\mu^{(SNR)}$	3.000	1.686	1.286	0.935	0.639	0.471	0.196
	SNR	2.327	4.138	5.308	6.801	8.333	9.549	13.776
	SSIM	0.374	0.444	0.499	0.547	0.606	0.663	0.810
SSIM_{max}	$\mu^{(SSIM)}$	1.798	1.057	0.829	0.641	0.467	0.343	0.177
	SNR	1.262	3.383	4.668	6.272	7.984	9.247	13.761
	SSIM	0.438	0.490	0.534	0.574	0.624	0.674	0.811

Table 4: Test image `brain`. Output μ-values and SNR/SSIM metrics for the restoration by ADP, NEDP, PWP and for the output restorations corresponding to the maximum SNR and SSIM achieved, for different κ.

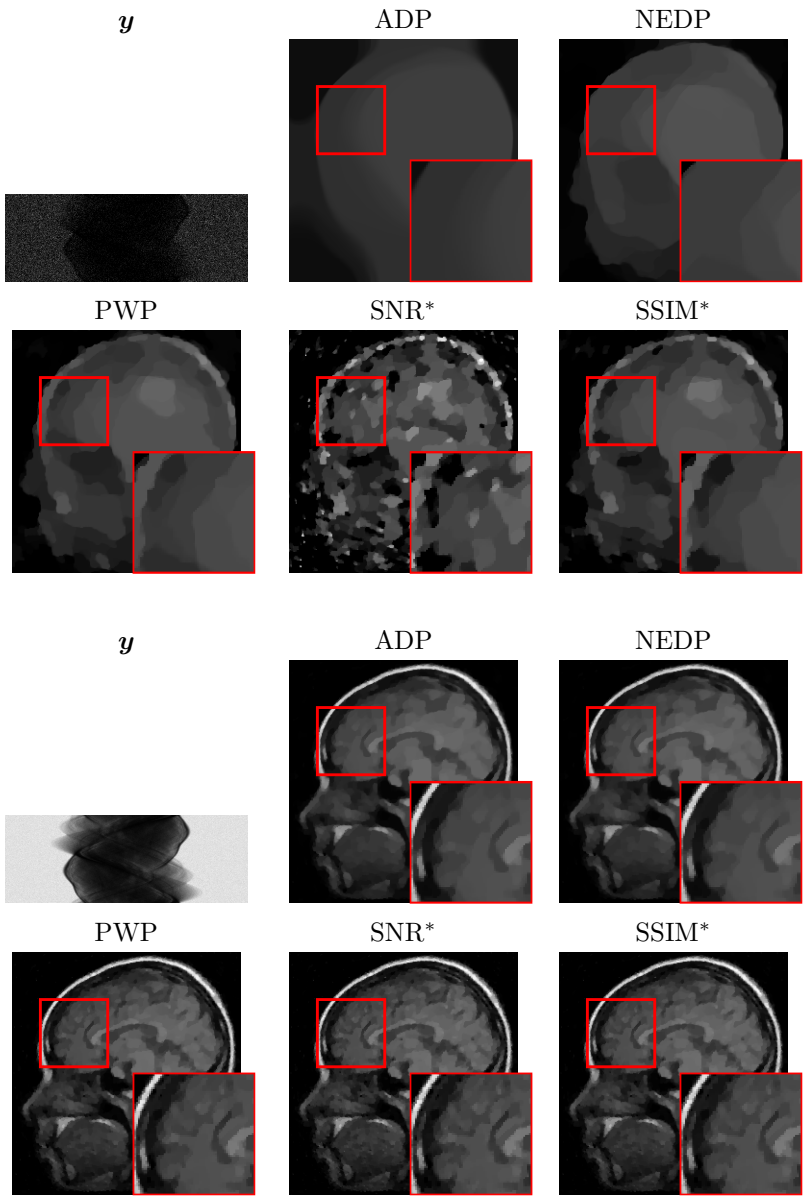


Figure 9: Test image `brain`. Observed image $\boldsymbol{y}$, restorations using the ADP, NEDP and PWP, and restorations corresponding to the maximum SNR and SSIM for $I_0 = 1.5$ (top panels) and $I_0 = 1000$ (bottom panels).

6. Conclusion

We introduced a novel parameter selection strategy in variational models under Poisson noise corruption. Our proposal relies on the extension of the whiteness principle designed for additive white noise to a suitably standardized version of the Poisson-corrupted observation. The derived Poisson whiteness principle has been extensively tested on image restoration and computed tomography reconstruction problems. In the latter case, we employed a linearized version of the ADMM which fasten the computations when the forward model operator does not present an advantageous structure. The Poisson whiteness principle has been compared with the popular approximate discrepancy principle as well as with a nearly exact version of it recently proposed by the authors. The newly introduced approach has been shown to outperform the competitors especially in the low-counting regime.

Acknowledgements All the authors are members of the "National Group for Scientific Computation (GNCS-INDAM)". The research of FB, AL, FS has been funded by the ex60 project "Funds for selected research topics", while MP acknowledges the contribution of "Young researchers funding" awarded by GNCS-INDAM.

References

[1] M. Bertero, P. Boccacci, V. Ruggiero, Inverse Imaging with Poisson Data, IOP Publishing, 2018.

[2] F. Bevilacqua, A. Lanza, M. Pragliola, F. Sgallari, Nearly exact discrepancy principle for low-count Poisson image restoration, Journal of Imaging 8 (1) (2022) 1–35.

[3] L. I. Rudin, S. Osher, E. Fatemi, Nonlinear total variation based noise removal algorithms, Physica D: Nonlinear Phenomena 60 (1–4) (1992) 259–268.

[4] R. Zanella, P. Boccacci, L. Zanni, M. Bertero, Efficient gradient projection methods for edge-preserving removal of Poisson noise, Inverse Problems 25 (2009) 045010.

[5] M. Bertero, P. Boccacci, G. Talenti, R. Zanella, L. Zanni, A discrepancy principle for Poisson data, Inverse Problems 26 (10) (2010) 105004.

[6] P. Craven, G. Wahba, Smoothing noisy data with spline functions, Numerische Mathematik 31 (1978) 377–403.

[7] C. Gu, Cross-validating non-Gaussian data, Journal of Computational and Graphical Statistics 1 (1992) 169–179.

[8] D. Xiang, G. Wahba, A generalized approximate cross validation for smoothing splines with non-Gaussian data, Statistica Sinica 6 (3) (1996) 675–692.

[9] P. Hansen, M. Kilmer, R. Kjeldsen, Exploiting residual information in the parameter choice for discrete ill-posed problems, BIT Numerical Mathematics 46 (2006) 41–59.

[10] M. Almeida, M. Figueiredo, Parameter estimation for blind and non-blind deblurring using residual whiteness measures, IEEE Transactions on Image Processing 22 (2013) 2751–2763.

[11] A. Lanza, M. Pragliola, F. Sgallari, Residual whiteness principle for parameter-free image restoration, Electronic Transactions on Numerical Analysis 53 (2020) 329–351.

[12] M. Pragliola, L. Calatroni, A. Lanza, F. Sgallari, Residual whiteness principle for automatic parameter selection in $\ell_2 - \ell_2$ image super-resolution problems, in: A. Elmoataz, J. Fadili, Y. Quéau, J. Rabin, L. Simon (Eds.), Scale Space and Variational Methods in Computer Vision, Springer International Publishing, Cham, 2021, pp. 476–488.

[13] M. Pragliola, L. Calatroni, A. Lanza, F. Sgallari, ADMM-based residual whiteness principle for automatic parameter selection in single image super-resolution problems, Journal of Mathematical Imaging and Vision (2022) 1–35.

[14] A. Lanza, M. Pragliola, F. Sgallari, Automatic fidelity and regularization terms selection in variational image restoration, BIT Numerical Mathematics 62 (3) (2022) 931–964.

[15] D. di Serafino, G. Landi, M. Viola, Directional TGV-based image restoration under Poisson noise, Journal of Imaging 7 (6) (2021) 99.
URL https://www.mdpi.com/2313-433X/7/6/99

[16] S. Boyd, N. Parikh, E. Chu, B. Peleato, J. Eckstein, Distributed optimization and statistical learning via the alternating direction method of multipliers, Foundations and Trends in Machine Learning 3 (1) (2011) 1–122. doi:10.1561/2200000016.
URL http://dx.doi.org/10.1561/2200000016

[17] R. M. Corless, G. H. Gonnet, D. E. G. Hare, D. J. Jeffrey, D. E. Knuth, On the LambertW function, Advances in Computational Mathematics 5 (1996) 329–359.

[18] S. Setzer, G. Steidl, T. Teuber, Deblurring poissonian images by split bregman techniques, Journal of Visual Communication and Image Representation 21 (3) (2010) 193–199.

[19] B. He, F. Ma, Z. Yuan, Optimally linearizing the alternating direction method of multipliers for convex programming, Computational Optimization and Applications 75 (2020) 361–388.

[20] Z. Wang, A. C. Bovik, H. R. Sheikh, E. P. Simoncelli, Image quality assessment: from error visibility to structural similarity, IEEE Transactions on Image Processing 13 (2004) 600–612.

[21] W. van Aarle, W. J. Palenstijn, J. Cant, E. Janssens, F. Bleichrodt, A. Dabravolski, J. D. Beenhouwer, K. J. Batenburg, J. Sijbers, Fast and flexible X-ray tomography using the ASTRA toolbox, Optics Express 24(22) (2016) 25129–25147.