Whiteness Constraints in a Unified Variational Framework for Image Restoration

(Article begins on next page)

28 April 2024

Whiteness constraints in a unified variational framework for image restoration

**Alessandro Lanza · Serena Morigi ·
Federica Sciacchitano · Fiorella Sgallari**

Abstract We propose a robust variational model for the restoration of images corrupted by blur and the general class of additive white noises. The key idea behind our proposal relies on a novel hard constraint imposed on the residual of the restoration, namely we characterize a residual whiteness set to which the restored image must belong. As the feasible set is unbounded, solution existence results for the proposed variational model are given. Moreover, based on theoretical derivations as well as on Monte Carlo simulations, we provide well-founded guidelines for setting the whiteness constraint limits. The solution of the non-trivial optimization problem, due to the non-smooth non-convex proposed model, is efficiently obtained by an Alternating Directions Method of Multipliers (ADMM), which in particular reduces the solution to a sequence of convex optimization sub-problems. Numerical results show the potentiality of the proposed model for restoring blurred images corrupted by several kinds of additive white noises.

A. Lanza
Department of Mathematics, University of Bologna, Bologna, IT
E-mail: alessandro.lanza2@unibo.it

S. Morigi
Department of Mathematics, University of Bologna, Bologna, IT
E-mail: serena.morigi@unibo.it

F. Sciacchitano
Department of Mathematics, University of Genova, Genova, IT
E-mail: sciacchitano@dima.unige.it

F. Sgallari
Department of Mathematics, University of Bologna, Bologna, IT
E-mail: fiorella.sgallari@unibo.it

1 Introduction

During the image acquisition and transmission processes, degradation effects such as those due to blur and noise always occur. One of the main goals of image processing is to eliminate these unwanted effects and to recover clean images from the acquired blurred and noisy images. Over the years, one of the most studied class of noises is that of additive, independent identically distributed (in short i.i.d.) noises, which affect all the pixels by independent random corruptions coming from the same distribution [9]. This class includes important noises such as those characterized by Gaussian [39], uniform [17], Laplacian [35] and Cauchy [42] distributions, which can be found in many applications, such as e.g. medical and astronomic imaging [9,28]. For any of these noise distributions, ad hoc variational models have been devised in the past for image restoration; see [9]. However, in many practical applications it is difficult to know a priori the noise distribution and, in any case, the noise might be the outcome of several sources thus giving raise to mixed noise models [24] with very specific/complex distributions.

To overcome these inherent difficulties, in this paper we propose a robust variational model aimed at restoring gray level images corrupted by blur and by the generic wide class of additive white - or uncorrelated - noises [9], which include i.i.d noises as a particular case and whose precise definition will be given later. Since an extension to color images would require some different definitions of our variational model, see [16], we postpone the analysis to color images to a future work. Without loss of generality, we consider images with a square $d \times d$ domain. The discrete model of the image degradation process we consider in this paper reads as

$$g \, = \, Ku \, + \, n \, , \tag{1}$$

where $u, n, g \in \mathbb{R}^{d^2}$ represent column-major vectorized forms of the unknown uncorrupted image, unknown noise realization and observed corrupted image, respectively, and where $K \in \mathbb{R}^{d^2 \times d^2}$ is a known linear blurring operator. Given K and g, and under the assumption that the additive noise process is white, our goal is to solve the ill-conditioned - or even numerically singular, depending on K - inverse problem of recovering an as accurate as possible estimate u^* of the unknown clean image u.

The proposed variational model, referred to as TV-W, is as follows:

$$u^* \leftarrow \arg \min_{u \in \mathcal{W}_\alpha} \mathrm{TV}(u) \, , \qquad \mathrm{TV}(u) := \sum_{i=1}^{d^2} \|(\nabla u)_i\|_2 \, , \tag{2}$$

where $\mathrm{TV}(u)$ represents the discrete Total Variation semi-norm of image u and $(\nabla u)_i \in \mathbb{R}^2$ denotes the discrete gradient of image u at pixel i. We chose the TV regularizer for its popularity and effectiveness but potentially any other regularizer could be used as well, such as nonlocal TV [18], TV_p [23] or fractional TV [13]. The key novelty behind our proposal is the feasible set $\mathcal{W}_\alpha$,

referred to as the *residual whiteness set* – or, in short, simply the *whiteness set* – that will be formally defined and widely analyzed in Section 4.

Coarsely speaking, $\mathcal{W}_\alpha$ contains restored images u^* for which the associated residue images $g - Ku^*$ resemble the realization of a white noise process. This is motivated by the fact that, according to the considered degradation model (1), the residue image $g - Ku$ is exactly equal to n – the realization of a white noise process – when u is the target uncorrupted image we aim to recover. Hence, the idea is that, by explicitly constraining the restoration residual $g - Ku^*$ to resemble a white noise realization, the restored image u^* is effectively pushed towards the target uncorrupted image u, independently of the actual noise distribution.

As pointed out in the paper, the potentiality of the proposed model is twofold:

a) it can be used effectively for the restoration of images corrupted by blur and the large class of additive white noises, under mild assumptions on the noise distribution;
b) it holds the potential for obtaining restorations of quality similar if not superior to that achievable by variational models containing data fidelity terms suitably constructed for a specific noise distribution.

With the aim of highlighting potentiality b) above, in Figure 1 we show some restoration results obtained on a photographic image containing neat edges and also some texture (a). The image has been synthetically corrupted by space-invariant Gaussian blur and strong additive white Gaussian noise (b) and then restored by using both a classical TV-ℓ_2 - or ROF - variational model (c) and the proposed TV-W model (d). From visual inspection of Figures 1(c)-(d), one can observe the quality improvement provided by our model, especially in correspondence of textured regions (see, e.g., the central skyscraper).

The proposed constrained variational model (2) is non-smooth due to the TV objective function and non-convex due to the feasible whiteness set $\mathcal{W}_\alpha$ (non-convexity of $\mathcal{W}_\alpha$ will be demonstrated in Section 5). Non-convex problems raise difficulties from an optimization point of view, due to the fact that in general there is no guarantee that the numerical optimization algorithm converges to the global optimum. For what concerns the latter difficulty, however, in literature it has been shown that many approaches specifically devised for convex optimization can also be effectively used in the non-convex setting under appropriate conditions/assumptions, e.g. the augmented Lagrangian approach, the gradient projection method, the forward-backward splitting algorithm, the majorization-minimization approach, the graph cut method, the iPiano algorithm and the ADMM; for an overview, we refer the reader to [1, 5, 2, 36, 41, 37] and the references therein. In this paper, we propose an ADMM-based iterative algorithm [46] for solving our non-convex non-smooth minimization problem (2). This choice allows us to reduce the solution of (2) to the solution of a sequence of easier, convex and efficiently solvable minimization sub-problems.

The paper is organized as follows. In Section 2 we introduce basic definitions on white noise processes, and in Section 3 we discuss the distributions
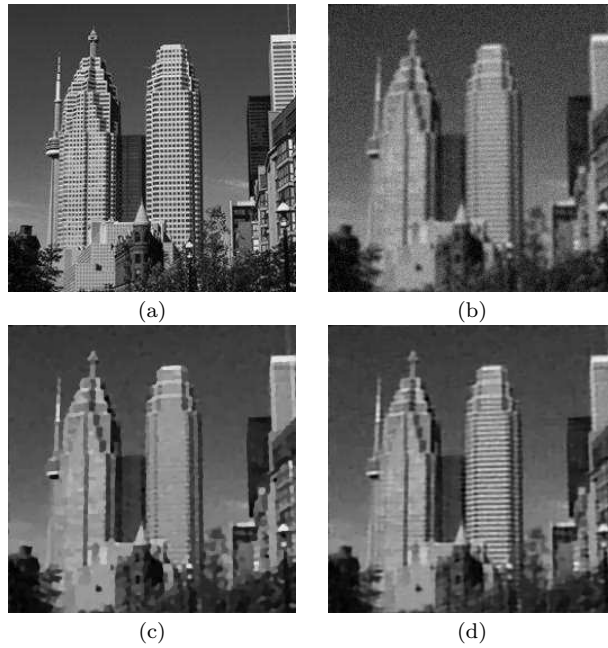
Fig. 1 Restoration results obtained by using the TV-ℓ_2 (c) and the proposed TV-W (d) variational models on an approximately piecewise constant and slightly textured photographic image (a) corrupted by space-invariant Gaussian blur and additive white Gaussian noise (b).

related to some important cases of additive white noises considered in this paper. In Section 4 we formally define and then analyze the residual whiteness set $\mathcal{W}_\alpha$, which constitutes the key ingredient in our proposal. Section 5 is devoted to the proof of existence of solutions for the proposed TV-W variational model. The ADMM-based iterative algorithm used to numerically solve the model is illustrated in Section 6 and numerical results are presented in Section 7. Conclusions are drawn in Section 8.

1.1 Related Work

Various image restoration strategies have been proposed in literature, but invariably they make strong assumptions about the properties of the image and, in particular, of the degradation process including the noise characteristics. Hence, some of these methods lack the generality to be easily applied in practical cases where, e.g., the noise distribution is a priori unknown or too complex to be dealt with explicitly. In this section, we briefly review some early approaches which exploit, more or less explicitly, the noise statistics together with the properties of the restoration residual in order to improve the restoration results.

In [6], where the popular Nonlocal means (NLM) algorithm for image denoising has been proposed, the properties of the denoising residual, referred to as the "method noise", have been used to design the NLM regularizer and, then, to visually evaluate a posteriori the accuracy of the obtained denoised images. In [12], the authors demonstrated how many full-reference image quality measures, e.g. the mean-squared-error and the structural similarity index, can be estimated from the residual image without the reference image. This suggests how the information contained in the residual image is strongly correlated with the quality of the restoration.

During the years, the statistical properties of the residual - including its distribution and its whiteness - have been also used for the selection of the regularization parameter of unconstrained variational models for image restoration, see e.g. [52], [53] and [19]. For instance, a bilevel programming approach has been proposed in [19] to optimize the choice of the regularization parameters for denoising images corrupted by additive Gaussian noise. In particular, the method relies on testing Gaussianity of the residuals of the denoising process.

In [4], the authors introduce an unsupervised, information-theoretic, adaptive filter (UINTA) for image denoising which adapts automatically to the statistics of the input image. In practice, the filter implicitly implements a general-purpose regularizer which does not assume any particular smoothness constraint on the target uncorrupted image, similarly to what we do in this paper by the proposed general-purpose fidelity term based on only the noise whiteness property. For what concerns noise, in [4] only additive i.i.d. Gaussian noise is considered. Similarly to [4], other works have been proposed [14, 45] where the Stein's unbiased risk estimator (SURE) has been exploited for robust denoising of images corrupted by additive Gaussian i.i.d. noise. The method relies on a patch-based learning of the unknown uncorrupted image prior - that is, of a patch-based regularizer - which allows a robust estimation of the noise-free image.

In [25] the authors were the first to propose a variational model containing a penalty term devoted to promote whiteness of the residue image. The model in [25] was specifically aimed to perform denoising of images corrupted by additive white Gaussian noise. Then, an extensions of [25] to the image deblurring problem has been proposed in [26], where whiteness of the residue image is enforced through a hard instead of a soft constraint. Compared to the work in [26], the proposal in this work exhibits several important novelties.

First, in [26] the whiteness constraint was defined in the frequency domain (based on the normalized cumulative periodogram), whereas here we impose whiteness in the space domain (based on the sample auto-correlation). This choice yields a constrained variational model which is more easily tractable than the one in [26] from a numerical point of view.

Second, the whiteness bounds in the frequency domain could only be chosen empirically based on a computationally heavy Monte Carlo approach. Here, instead, we are able to theoretically characterize the statistics of the sample auto-correlation values, thus obtaining explicit expressions for the whiteness bounds in the space domain. Monte Carlo simulations are in this case carried

out only for an off-line empirical validation of the derived expressions. The existence of closed-form formulas for the whiteness bounds in the space domain not only allows for a much computationally lighter approach but also yields a parameter-free variational model, which only requires knowledge of the noise standard deviation.

Third, a theoretical proof of the existence of solution(s) to the new model is provided - see Proposition 2.

Last but not the least, the new model is suitable to entail a large class of additive white noises, under mild assumptions on the noise distribution, while the model in [26] was only proposed to restore images corrupted by additive white Gaussian noise.

We finally remark that this work is an extended version of the preliminary proceeding paper in [27].

2 Basics on white noise processes

The statistical version of the degradation model in (1) with images in matrix-form is

$$G[i,j] \;=\; v\,[i,j] \;+\; N[i,j]\,, \qquad [i,j] \in \Omega \,:=\, \{1,\ldots,d\}^2\,, \qquad (3)$$

with capital letters indicating random quantities and where $v[i,j] := (K\,u)[i,j]$ denotes the value of the blurred image at pixel $[i,j]$. In particular, the additive noise is modeled as a $d \times d$ discrete random process $\mathcal{N} := \{\, N[i,j] : [i,j] \in \Omega \,\}$ with $N[i,j]$ denoting the scalar random variable modeling noise at pixel $[i,j]$. This means that different images of the same subject under the same blurring operator will differ due to the inherently random nature of the noise.

We introduce in the following some basic definitions, which will be useful in the construction of the whiteness set and in the derivation of related theoretical results.

Definition 1 A $d \times d$ discrete random process $\mathcal{N}$ is said to be *weak stationary* if the following three conditions are satisfied:

· $\mathbb{E}(N[i,j]^2)$ is finite, $\forall[i,j] \in \Omega$,
· $\mathbb{E}(N[i,j]) = \mathbb{E}(N[i+l,j+m])$, $\forall[l,m]$,
· $\mathrm{Cov}(N[i_1,j_1],N[i_1+l,j_1+m]) = \mathrm{Cov}(N[i_2,j_2],N[i_2+l,j_2+m]), \forall[l,m]$,

where $\mathbb{E}$ and Cov denote the expectation and covariance operators, respectively.

In other words, Definition 1 states that stationary processes must have three features: finite variance, constant mean and the covariance that depends only on the two-dimensional *lag* denoted by $[l,m]$.

Definition 2 The *ensemble* auto-correlation of a weak stationary process $\mathcal{N}$ is a function $\rho_{\mathcal{N}}$ which maps any lag $[\ell,m]$ into a scalar value given by

$$\rho_{\mathcal{N}}[l,m] \;=\; \mathbb{E}\left(\, N[i,j]\,N[i+l,j+m]\,\right). \qquad (4)$$

Definition 3 A $d \times d$ discrete random process $\mathcal{N}$ is said to be *white* if it is

- weak stationary;
- zero-mean;
- uncorrelated, that is

$$\rho_{\mathcal{N}}[l, m] \;=\; \begin{cases} \sigma_n^2 & \text{if } [l, m] = (0, 0) \\ 0 & \text{otherwise} \end{cases}, \; [l, m] \in \Theta := \{-(d-1), \, \dots, d-1\}^2, \quad (5)$$

where σ_n^2 represents the variance of the 2D random process.

Equation (5) says that a white noise process is characterized by zero values of the auto-correlation function at all non-vanishing lags.

Given a single realization $n := \{\, n[i, j] \in \mathbb{R} : \; [i, j] \in \Omega \,\} \in \mathbb{R}^{d \times d}$ of the noise process $\mathcal{N}$, that is the series of noise values corrupting the particular observed image according to the deterministic degradation model in (1), the *sample auto-correlation* of n is a function r_n mapping all the possible lags $[l, m] \in \Theta$ into a scalar value given by

$$r_n[l, m] \;:=\; \frac{1}{d^2} \left(n \star n \right)[l, m] \;=\; \frac{1}{d^2} \left(n' * n \right)[l, m]$$

$$=\; \frac{1}{d^2} \sum_{i,j=1}^{d} n[i, j] \, n[i + l, j + m] \,, \quad [l, m] \in \Theta \,, \quad (6)$$

where $\star$ and $*$ denote the 2-D discrete correlation and convolution operators, respectively, and where $n'(i, j) = n(-i, -j)$.

Clearly, for (6) being defined for all lags $[l, m] \in \Theta$, the noise realization n must be padded with at least $d - 1$ samples in all directions, that is boundary conditions for the noise realization have to be chosen.

In this paper, we assume that the noise process n is periodic, such that $\star$ and $*$ in the definition (6) of the sample auto-correlation function indicate circular correlation and circular convolution, respectively. This assumption yields symmetry properties of the sample auto-correlation function, such that the only informative lags belong to the following subset $\overline{\Theta}$ of Θ defined in (5):

$$\overline{\Theta} := \{0, \, \dots, d-1\}^2. \quad (7)$$

Therefore, in the rest of the paper we consider the restriction of the sample auto-correlation function to the domain $\overline{\Theta}$.

With the further assumption that the noise process is ergodic, i.e. any ensemble statistical properties of the process can be estimated from a single, sufficiently long, random sample of it, it can be demonstrated that the sample auto-correlation r_n is a good estimate of the ensemble auto-correlation $\rho_{\mathcal{N}}$, see [16]. In particular, we have:

$$\lim_{d \to \infty} r_n[l, m] = \rho_{\mathcal{N}}[l, m] \;=\; \begin{cases} \sigma_n^2 & \text{for } [l, m] = [0, 0] \\ 0 & \text{for } [l, m] \in \overline{\Theta}_0 := \overline{\Theta} \setminus \{0, 0\} \,, \end{cases} \quad (8)$$

A set of white noise realizations can thus be defined in the spatial domain by constraining the values of the sample auto-correlation function r_n in (6) to lie within a band around the theoretical limit which, in accordance with (8), is given by the ensemble auto-correlation $\rho_\mathcal{N}$ in (4).

3 Variational models for popular additive white noises

In this section, we characterize the distributions of the popular additive white noises considered in this paper and discuss how these distributions lead to variational models containing different data fidelity terms.

A commonly used paradigm for image restoration is the probabilistic Maximum A Posteriori (MAP) approach [9]: the restored image is obtained as the maximizer of the posterior probability of the unknown clean image u given the observed image g and the blurring operator K, considered as a deterministic parameter. In formulas:

$$u^* \leftarrow \arg\max_{u \in \mathbb{R}^{d^2}} \Pr(u|g;K) = \arg\min_{u \in \mathbb{R}^{d^2}} \left\{ -\log \Pr(u) - \log \Pr(g|u;K) \right\}, \quad (9)$$

where (9) follows by applying the Bayes' rule, by dropping the *evidence* term $\Pr(g)$ since it does not depend on u, and by reformulating maximization as a minimization of the negative logarithm. The two terms $\Pr(u)$ and $\Pr(g|u;K)$ represent the *prior* and the *likelihood*, respectively [22]. The prior term embodies prior knowledge on the unknown clean image u, typically in the form of smoothness constraints. The likelihood term encodes information on the observation model and forces closeness of the estimate u^* to the observed data g according to such model. For what concerns the prior, a common choice is to model the unknown image u as a Markov Random Field (MRF) [50], such that the image can be characterized by its Gibbs prior distribution, whose general form is:

$$\Pr(u) = \frac{1}{Z} \prod_{i=1}^{d^2} \exp\left(-\alpha V_{c_i}(u) \right) = \frac{1}{Z} \exp\left(-\alpha \sum_{i=1}^{d^2} V_{c_i}(u) \right), \quad (10)$$

where $\alpha > 0$ is the MRF parameter, $\{c_i\}_{i=1}^{d^2}$ is the set of all cliques (a clique is a set of neighboring pixels) for the MRF, V_{c_i} is the potential function defined on the clique c_i and Z is the partition function, that is a function not depending on u which allows for the normalization of the prior. Choosing as potential function at the generic i-th pixel the magnitude of the discrete gradient at the same pixel, i.e. $V_{c_i}(u) := \|(\nabla u)_i\|_2$, the Gibbs prior in (10) reduces to the popular TV prior:

$$\Pr(u) = \frac{1}{Z} \exp\left(-\alpha \sum_{i=1}^{d^2} \|(\nabla u)_i\|_2 \right) = \frac{1}{Z} \exp(-\alpha \, \mathrm{TV}(u)) . \quad (11)$$

Using the TV prior in (11) can be regarded as implicitly assuming that the gradient magnitude at each pixel of the unknown clean image, $\|(\nabla u)_i\|_2$, follows a half-Laplacian (or exponential) distribution, whose probability density function (pdf) is given by:

$$\Pr(x; \alpha) = \begin{cases} \alpha \exp(-\alpha x) & \text{for } x \geq 0 \\ 0 & \text{for } x < 0 \end{cases}, \tag{12}$$

where $\alpha > 0$ is called the *scale* parameter of the distribution.

For what regards the likelihood term in (9), by assuming that noise is additive independent identically distributed - which is a common choice and, as previously highlighted, is a stronger assumption than that of noise being white - then we have:

$$\Pr(g|u; K) = \prod_{i=1}^{d^2} \Pr(g_i|u; K). \tag{13}$$

Replacing the the TV prior (11) and the likelihood (13) into the MAP inference formula (9), one gets:

$$u^* \leftarrow \arg\min_{u \in \mathbb{R}^{d^2}} \left\{ \ln(Z) + \alpha \, \mathrm{TV}(u) - \sum_{i=1}^{d^2} \ln\left[\Pr(g_i|u; K)\right] \right\}$$

$$\leftarrow \arg\min_{u \in \mathbb{R}^{d^2}} \left\{ \mathrm{TV}(u) + \frac{1}{\alpha} \sum_{i=1}^{d^2} \ln\left[\frac{1}{\Pr(g_i|u; K)}\right] \right\}. \tag{14}$$

where in (14) we dropped the constant term $\ln(Z)$ and divided the objective function by the positive constant α. Substituting any specific noise distribution, the MAP inference formula in (14) turns into an unconstrained variational model of the form

$$u^* \leftarrow \arg\min_{u \in \mathbb{R}^{d^2}} \left\{ \mathrm{TV}(u) + \mu \, \mathcal{F}(u; K, g) \right\}, \tag{15}$$

where the functional $\mathcal{F}(u; K, g)$ and the positive constant μ are commonly referred to as the data fidelity term and the regularization parameter.

The choice of the regularization parameter μ is crucial for obtaining a good reconstruction of the image; in fact, a too small value of μ would lead to an over-regularized (over-smoothed) image while if the value is too large, a poor reduction of noise would be observed. As we will see later, the proposed variational model somehow circumvents the use of this regularization parameter by introducing a feasible set which depends only on the whiteness bounds.

Some important examples of distributions of additive white noises are the uniform, the Gaussian, the Laplacian and the Cauchy. In Table 1 we report, for any of these distributions, the expressions of the probability density function (pdf), of the variance and of the associated fidelity term $\mathcal{F}(u; K, g)$ and regularization parameter μ one obtains by applying the MAP derivation above.

Here, we can see that each noise model has its own MAP-based fidelity term, therefore the a-priori knowledge of the noise distribution is required in order to apply one of the classical variational models. The Gaussian noise has been extensively studied partially since it produces simple and very tractable noise models and partially due to the celebrated central limit theorem in probability theory. The uniform noise [17] mainly appears as a statistical model of quantization errors, thus it is common during the digital acquisition [44]. The Laplacian and Cauchy noises model impulsive corruptions. Laplacian noise is somehow related with the impulse noise [35], where the main difference between these two noise models is that in the Laplacian case all the pixels are corrupted while in the impulse noise some pixels are noisy free. Cauchy noise is a degradation more impulsive than the Laplacian noise, which appears in atmospheric and underwater acoustic noises, radar and sonar applications, air turbulence, wireless communication system, biomedical images, synthetic aperture radar (SAR) images, see [42] and references therein. Cauchy and Gaussian noises belong to the class of alpha-stable distributions, characterized by a bell-shaped pdf.

Table 1 Popular additive (zero-mean) white noise distributions and some properties.

Distribution	pdf $p(x)$	Variance	Fidelity $\mathcal{F}(u; K, g)$	μ
Uniform	$\dfrac{1}{2a}\chi_{[-a,a]}(x)$	$a^2/3$	$\imath_{[0,a]}\left(\|Ku - g\|_\infty\right)$	$\dfrac{1}{\alpha}$
Gaussian	$\dfrac{1}{\sqrt{2\pi\sigma^2}}\exp\left(-\dfrac{x^2}{2\sigma^2}\right)$	σ^2	$\frac{1}{2}\|Ku - g\|_2^2$	$\dfrac{1}{\alpha\,\sigma^2}$
Laplace	$\dfrac{1}{2b}\exp\left(-\dfrac{|x|}{b}\right)$	$2b^2$	$\|Ku - g\|_1$	$\dfrac{1}{\alpha\,b}$
Cauchy	$\dfrac{1}{\pi}\dfrac{\gamma}{\gamma^2 + x^2}$	not finite	$\displaystyle\sum_i \log\left(\gamma^2 + (Ku - g)_i^2\right)$	$\dfrac{1}{\alpha}$

For an accurate description of these noise models we refer the reader to [8, 9, 17, 35, 39, 42, 44] and the references therein.

To highlight the differences between the considered distributions, in Figure 2 we show the pdfs of the Uniform, Gaussian, Laplace and Cauchy distributions with zero-mean and unit-variance. Comparing the pdfs, we can see that Cauchy pdf has the highest tails, which means that for large values the density approaches zero slower than the other distributions, i.e. rare events have the probability to occur more often. Therefore, Cauchy is the most impulsive among the considered noises, Laplace is more impulsive than Gaussian and the uniform pdf has no tails.

It is worth highlighting that Cauchy distribution does not have finite variance. Since in our assumptions (see Section 4) the variance has to be finite, from now on we consider a modified version of the Cauchy distribution: the truncated Cauchy distribution with scale parameter $\gamma > 0$, [7], whose pdf
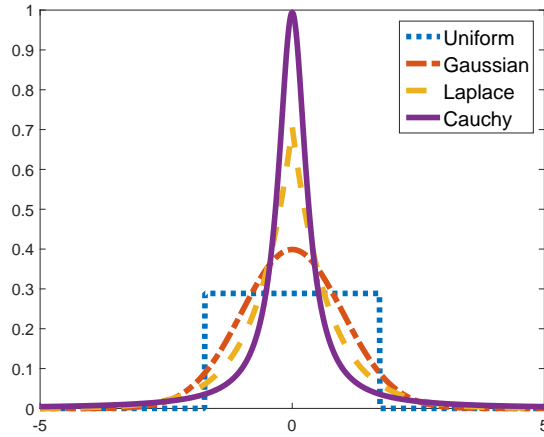
Fig. 2 Plots of zero-mean, unit-variance pdfs for the distributions in Table 1.

reads as

$$p(x) = \left(2\gamma \arctan(L/\gamma)\left(1 + (x/\gamma)^2\right)\right)^{-1} \text{ if } |x| \le L, \quad p(x) = 0 \text{ if } |x| > L,$$

where $L > 0$ defines the finite support $[-L, L]$ of the pdf. Note that it is reasonable to consider the truncated version of the Cauchy pdf since in practice the range of images is always a finite interval, in our case $[0, 1]$. However, for this distribution the fidelity term is unknown and it will be matter of future studies.

In Figure 3 we show some examples of noisy images corrupted by additive white uniform, Gaussian, Laplacian, and truncated Cauchy noise with the same standard deviation. The noisy images are computed by using the equation (1) with K the identity, i.e. they are given by a sum of the original image and a realization of a random variable which is Gaussian, Laplacian, uniform and Cauchy distributed, respectively. We can observe that Cauchy noise is the most impulsive one, where some pixels are totally corrupted, while some others are not so corrupted, Laplace is less impulsive than the Cauchy one, but it still has some peaks which are bigger than the Gaussian distribution.

Furthermore, one could also consider the Poisson and the multiplicative noises. Poisson noise is a signal dependent noise, which can be found in many applications, in particular astronomic and medical imaging. As pointed out in [32], the noisy image obtained by applying the Anscombe root transformation to an image corrupted by Poisson noise, can be treated as additive white Gaussian noise. The multiplicative noise generally appears in synthetic aperture radar (SAR), ultrasound imaging, laser images, and so on. As suggested in [3], and the references therein, one can apply the log-transformation in order to obtain a noise model that corresponds to the white additive noise. In this paper, we do not focus on the Poisson and multiplicative noises, but we point out that by using the above preliminary transformations, the proposed model works also for these two noise models.

Fig. 3 Noisy images corrupted by additive white noise: Uniform, Gaussian, Laplacian and truncated Cauchy, with the same variance.

4 Whiteness constraints

In this section, we formally define the residual whiteness set $\mathcal{W}_\alpha$, which represents the key novelty behind our proposed variational model (2). First, based on the asymptotic property (8) of the sample auto-correlation of a white noise realization, we define $\mathcal{W}_\alpha$. Then, in order to provide well-grounded basis to this definition and to the choice of the whiteness bounds - i.e., the limits of the whiteness set - in subsections 4.1-4.2 we statistically characterize the sample auto-correlation values of a white noise realization and then validate empirically the given theoretical results by means of some Monte Carlo simulations. Finally, in subsection 4.3 we give an explicit expression for the whiteness bounds.

First, based on (2), we define the *noise whiteness set* as follows:

$$W_\alpha := \left\{\, n \in \mathbb{R}^{d \times d} : \ -w_\alpha \leq r_n[l,m] \leq w_\alpha \quad \forall\, [l,m] \in \overline{\Theta}_0 \,\right\}. \qquad (16)$$

We notice that only the auto-correlation values at non-zero lags are constrained. In fact, the auto-correlation at lag $[0,0]$ represents the variance of the noise realization, which is not related to the whiteness of the process.

In our restoration model, we impose that the residue image of the restoration $Ku - g \in \mathbb{R}^{d^2}$ resembles the realization of a white noise process, that is the residual belongs to the whiteness set W_α defined in (16). Hence, the *whiteness set* in our model (2) takes the form

$$\mathcal{W}_\alpha := \left\{ u \in \mathbb{R}^{d^2} : \ -w_\alpha \leq \ r_{\mathrm{mat}(Ku-g)}[l,m] \ \leq w_\alpha \ \forall \, [l,m] \in \overline{\Theta}_0 \right\}, \qquad (17)$$

with

$$r_{\mathrm{mat}(Ku-g)}[l,m] = \frac{1}{d^2} \left(\mathrm{mat}(Ku - g) \star \mathrm{mat}(Ku - g) \right)[l,m] \qquad (18)$$

denoting the auto-correlation of the residue image and $\mathrm{mat}(v) \in \mathbb{R}^{d \times d}$ indicating the matrix formed column-wise by taking the elements of the vector $v \in \mathbb{R}^{d^2}$. To simplify notations, in the rest of the paper we drop the $\mathrm{mat}(\cdot)$ symbol when applied to the operands of the 2D correlation binary operator $\star$ or to the subscript of the sample auto-correlation, such that $v \star w$ stands for $\mathrm{mat}(v) \star \mathrm{mat}(w)$ and r_v stands for $r_{\mathrm{mat}(v)}$ when $v, w \in \mathbb{R}^{d^2}$.

The non-negative scalar w_α in (17), referred to as the *whiteness bound*, represents a degree of freedom which allows to set the actual size of the whiteness set or, equivalently, the probability that the sample auto-correlation of a white noise realization belongs to the whiteness set. In order to justify the fact that a unique symmetric bound is used in (17) for any lag pair $[l,m]$ and to provide a founded insight on how such bound must be set, in the next two subsections we characterize the sample auto-correlation of a white noise realization from a statistical point of view. In particular, in Subsection 4.1 we give theoretical results concerning the distribution of the sample auto-correlation values, whereas in Subsection 4.2 we provide empirical evidence of such results by means of suitable Monte Carlo simulations. Based on these results, in Subsection 4.3 we finally suggest a simple yet sound strategy for choosing the value of the whiteness bound w_α in (17).

4.1 Statistical characterization of the sample-autocorrelation

The sample auto-correlation defined in (6) is related to a given realization n of the random process $\mathcal{N}$, hence it is clearly a deterministic quantity. However, since here we want to analyze the statistics of the sample auto-correlation, we need to consider it as a random quantity depending on the random process $\mathcal{N}$ instead of on the deterministic realization n. We thus introduce the random variables $R_\mathcal{N}[l,m]$ which, according to definition (6) of the sample auto-correlation, are defined as

$$R_\mathcal{N}[l,m] = \frac{1}{d^2} \sum_{i,j=1}^{d} N[i,j] \, N[i+l, j+m], \quad [l,m] \in \overline{\Theta}_0, \qquad (19)$$

where, with respect to (6), we clearly replaced the realized, deterministic values $n[\cdot, \cdot]$ with the associated random variables $N[\cdot, \cdot]$.

In Proposition 1 below we give results on the distribution of the sample auto-correlation values of a white noise process.

Proposition 1 *Let $\mathcal{N}$ be a $d \times d$ white random process with distribution having (finite) variance σ_n^2 and let $\mathcal{N}$ be stationary to the fourth order and with finite fourth order moments. Then, as the dimension d tends to $+\infty$, the random variables $R_{\mathcal{N}}[l,m]$ representing the sample auto-correlation at any non-zero lag $[l,m]$ are asymptotically uncorrelated and their limiting distribution is a Gaussian distribution defined as follows:*

$$R_{\mathcal{N}}[l,m] \sim \mathcal{G}(0,\sigma_r^2), \quad \sigma_r = \frac{\sigma_n^2}{d}, \quad [l,m] \in \overline{\Theta}_0 \,. \tag{20}$$

Proof First, by using the central limit theorem for dependent stationary processes, see [15, Thm 2A], we have that, as the dimension d tends to $+\infty$, the distribution of any random variable $R_{\mathcal{N}}[l,m]$ defined in (19) approaches a Gaussian distribution.

As far as the expected value of the random variables $R_{\mathcal{N}}[l,m]$ is concerned, based on definition (19) and on linearity of the expectation operator, we have

$$\mathbb{E}\left(R_{\mathcal{N}}[l,m]\right) = \mathbb{E}\left(\frac{1}{d^2}\sum_{i,j=1}^{d} N[i,j]\,N[i+l,j+m]\right)$$

$$= \frac{1}{d^2}\sum_{i,j=1}^{d}\mathbb{E}\left(N[i,j]\,N[i+l,j+m]\right) = 0 \quad \forall [l,m] \in \overline{\Theta}_0 \,, \tag{21}$$

where the last equality in (21) comes from whiteness of the random process $\mathcal{N}$ - see previous Definitions 2 and 3.

For what finally concerns the variances and covariances of the random variables $R_{\mathcal{N}}[l,m]$, the asymptotic results given in the proposition statement can be obtained based on the well known Bartlett's formula - see [10, Chapters 2 and 7].

4.2 Monte Carlo simulations

In this section we provide empirical evidence of the theoretical asymptotic results stated in Proposition 1 by means of suitable Monte Carlo simulations. The simulations have been carried out in Matlab, and the experimental setting is as follows. For a given dimension d, a given distribution among the zero-mean Uniform, Gaussian, Laplace, (truncated) Cauchy and balanced mixed Uniform-Gaussian distributions, and a given standard deviation σ_n of the distribution, we generated pseudo-randomly a large number – namely, 20000 – of realizations $n \in \mathbb{R}^{d \times d}$ from the $d \times d$ white noise random process having the selected distribution with variance σ_n^2. Then, for each realization n we computed the sample auto-correlation r_n according to definition (6) and, finally, we constructed, for each lag $[l,m]$, the normalized histogram of the obtained

20000 sample auto-correlation values. We remark that, as the number of generated samples/realizations tends to $+\infty$, the constructed histograms approach the theoretical distributions of the sample auto-correlation values, that is the distributions of the random variables $R_{\mathcal{N}}[l, m]$ defined in (19). Hence, if the number of considered realizations is sufficiently large (we noticed that increasing the number of realizations beyond 20000 does not change significantly the histograms), the obtained histograms can be utilized to check the validity of theoretical results given in Proposition 1.
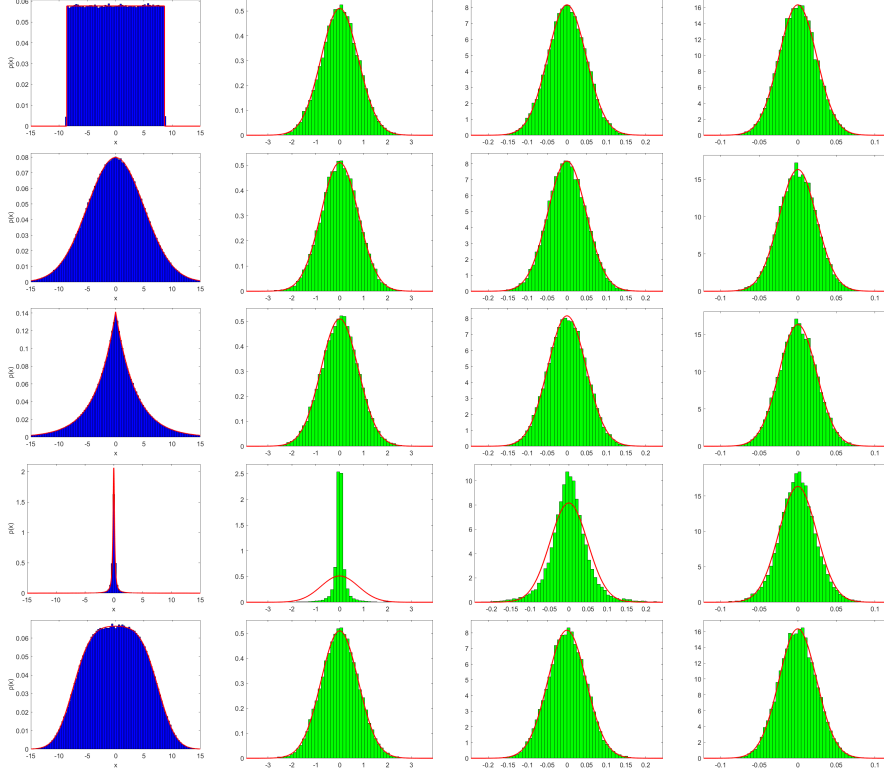
Fig. 4 First column: plots of the theoretical pdfs (solid red lines) and of the measured normalized histograms for the five considered white noise distributions, namely Uniform, Gaussian, Laplacian, Cauchy and mixed Uniform-Gaussian, from top to bottom. Columns 2-4: theoretical asymptotic pdfs (solid red lines) - see Proposition 1 - and measured normalized histograms of the autocorrelation values for lag $[l, m] = [3, 4]$ and image dimensions $d = 32$, $d = 512$ and $d = 1024$, respectively.

In Figure 4 we show the results of the Monte Carlo simulations for one auto-correlation lag, namely $[l, m] = [3, 4]$ (for other lags, results are similar), and for a noise standard deviation $\sigma_n = 5$. In particular, in the first column we show the five theoretical pdfs (solid red lines) from which we extracted the white noise samples and also the normalized histograms (in blue) of the noise

samples we actually generated. The good accordance between the pdfs and the histograms allows to validate the procedures used to extract the samples. In columns 2-4 we report the normalized histograms (in green) of the generated sample auto-correlation values $r_n[3,4]$ together with the theoretical, asymptotic pdfs (solid red lines) - see results in Proposition 1 - for increasing dimension $d \in \{32, 512, 1024\}$ of the generated $d \times d$ noise realizations. One can observe how, as the dimension d increases, the empirical histograms approach (more or less quickly, depending on the noise distribution) the asymptotic theoretical zero-mean Gaussian pdf with standard deviation σ_r, as stated in Proposition 1. In case of Cauchy noise, the empirical validation of this important theoretical result requires that the dimension of the sample realizations is increased to $d = 1024$ to better appreciate the Gaussian-shaped distribution, as illustrated in Figure 4 (fourth row).

We remark that the more impulsive is the white noise process - that is, the highest are the tails of the noise distribution - the slower is the convergence of the empirical pdf to the theoretical, asymptotic one.

4.3 Setting the whiteness bound w_α

Since according to Proposition 1 and to Monte Carlo simulations the distribution of the sample auto-correlation of a white noise realization at any non-zero lag can be well approximated - for sufficiently and reasonably large images - by a zero-mean Gaussian with standard deviation σ_r given in (20), it results natural the choice of using a unique symmetric bound for any lag in the definition (17) of our whiteness set. Moreover, the whiteness bound w_α in (17) can be simply set according to the formula

$$w_\alpha = \alpha \, \sigma_r, \tag{22}$$

such that the *whiteness coefficient* $\alpha \geq 0$ allows to directly set the probability that the sample auto-correlation of a white noise realization at any given non-zero lag falls inside the whiteness set. In fact, due to the fact that the limiting distribution is Gaussian, there is a well known one-to-one relationship between α and the probability that the sample auto-correlation values belong to the interval $[-w_\alpha, +w_\alpha]$, we refer the reader to [34] for more details. For instance, setting $\alpha = 2$ yields a whiteness set $\mathcal{W}_2$ which represents the smallest size set containing, for any lag, about 95% of all the possible realizations of white noise processes. It is worth remarking that the choice of the whiteness coefficient α should not depend on the value of σ_r and, hence, on the dimension d of the considered images or on the value σ_n of the noise standard deviation.

5 Existence of solutions

In this section we investigate the existence of solutions to the proposed variational model (2), which is not straightforwardly derivable due to properties of

the feasible whiteness set $\mathcal{W}_\alpha$ defined in (17)–(18). In the following, $\mathbb{1}$ denotes the (column) vector of all ones in $\mathbb{R}^{d^2}$, e_m the m-th element of the canonical basis of $\mathbb{R}^{d^2}$, $\{x^{(k)}\}$ the one-sided infinite sequence $\{x^{(k)}\}_{k\in\mathbb{N}}$ with $\mathbb{N}$ the set of natural numbers.

First, we give some useful definitions.

Definition 4 (unbounded sequence) A sequence $\{x^{(k)}\} \subset \mathbb{R}^n$ is unbounded if and only if $\lim_{k\to\infty}\left\|x^{(k)}\right\|_2 = +\infty$.

Definition 5 (unbounded set) A set $S \subseteq \mathbb{R}^n$ is unbounded if and only if it contains (at least) one unbounded sequence.

Definition 6 (coercive function) A function $f : \mathbb{R}^n \to \mathbb{R}$ is coercive over an unbounded set $S \subseteq \mathbb{R}^n$ if and only if $\lim_{k\to\infty} f(x^{(k)}) = +\infty$ for any unbounded sequence $\{x^{(k)}\} \subset S$.

In Lemma 1 and Lemma 2 below, whose proofs are postponed to the appendix, we analyze the feasible set and the cost function of the constrained minimization problem (2) associated to our proposal, respectively, then solution existence results for the problem are provided in Proposition 2.

Lemma 1 *For any real $\alpha \geq 0$, the whiteness set $\mathcal{W}_\alpha \subset \mathbb{R}^{d^2}$ defined in (17)–(18) is closed, unbounded and non-convex.*

Lemma 2 *The Total Variation objective function* $\mathrm{TV} : \mathbb{R}^{d^2} \to \mathbb{R}$ *defined in (2) is proper, continuous, convex, bounded from below by zero and coercive over the (unbounded) whiteness set $\mathcal{W}_\alpha \subset \mathbb{R}^{d^2}$ defined in (17)–(18), for any real $\alpha \geq 0$.*

The result in the following final proposition, which is the one of our main interest, follows directly from Lemma 1 and Lemma 2.

Proposition 2 *The constrained minimization problem defined in (2) admits at least one solution.*

6 Applying ADMM to the proposed model

In this section, we illustrate the ADMM-based [46] iterative algorithm used to numerically solve the proposed constrained minimization problem (2) with the feasible set $\mathcal{W}_\alpha$ defined in (17).

Towards this aim, first we resort to the variable splitting technique and introduce three auxiliary variables $r \in V$, $s \in V$ and $t \in Q$, with $V := \mathbb{R}^{d^2}$ and $Q := \mathbb{R}^{2d^2}$, such that model (2) is rewritten in the following equivalent form:

$$
\begin{aligned}
\{u^*, t^*, r^*, s^*\} \quad &\leftarrow \arg\min_{u,t,r,s} \left\{ \sum_{i=1}^{d^2} \|t_i\|_2 + \imath_{\mathcal{B}}(r,s) \right\} \\
\text{s.t. :} \quad &t = Du,\, r = Ku - g,\, s = Ku - g\,,
\end{aligned} \tag{23}
$$

where $D := (D_h^T, D_v^T)^T \in \mathbb{R}^{2d^2 \times d^2}$, $t_i := \left((D_h u)_i, (D_v u)_i \right)^T \in \mathbb{R}^2$ represents the discrete gradient of image u at pixel i, and $\imath_{\mathcal{B}}$ is the indicator function of the feasible set $\mathcal{B}$ for the variables r and s defined by

$$\mathcal{B} := \left\{ (r,s) \in V \times V \colon -w_\alpha d^2 \le r \star s \le w_\alpha d^2 \right\} , \qquad (24)$$

where $\imath_{\mathcal{B}}(r,s)$ takes the value 0 for $(r,s) \in \mathcal{B}$ and $+\infty$ otherwise.

The auxiliary variable t is introduced to transfer the discrete gradient operator $(\nabla u)_i$ out of the non-differentiable term $\| \cdot \|_2$. The variables r and s play the role of the restoration residue $Ku - g$ within the whiteness constraint (17) so that constraint (24) is now imposed on r and s. The rationale of introducing these two auxiliary variables which are defined to be equal, is to guarantee convexity of all the related subproblems.

To solve (23), we define the augmented Lagrangian functional

$$
\begin{aligned}
\mathcal{L}(u,t,r,s;\lambda_t,\lambda_r,\lambda_s) \ = \ & \sum_{i=1}^{d^2} \| t_i \|_2 + \imath_{\mathcal{B}}(r,s) \\
& - \langle \lambda_t, t - Du \rangle + \frac{\beta_t}{2} \| t - Du \|_2^2 \\
& - \langle \lambda_r, r - (Ku - g) \rangle + \frac{\beta_{rs}}{2} \| r - (Ku - g) \|_2^2 \\
& - \langle \lambda_s, s - (Ku - g) \rangle + \frac{\beta_{rs}}{2} \| s - (Ku - g) \|_2^2 (25)
\end{aligned}
$$

where $\beta_t, \beta_{rs} > 0$ are scalar penalty parameters and $\lambda_t \in Q$, $\lambda_r, \lambda_s \in V$ are the vectors of Lagrange multipliers associated with the linear constraints in (23). We used a common penalty parameter β_{rs} for the variables r and s since they represent the same quantity, namely the restoration residue.

Given the previously computed (or initialized for $k = 0$) vectors u^k, s^k, λ_t^k, λ_r^k and λ_s^k, the k-th iteration of the proposed ADMM-based iterative scheme [46] applied to the solution of the saddle-point problem related to the functional (25) reads as follows:

$$t^{k+1} \quad \leftarrow \quad \arg \min_{t \in Q} \mathcal{L}(u^k, t, r^k, s^k; \lambda_t^k, \lambda_r^k, \lambda_s^k) \qquad (26)$$

$$r^{k+1} \quad \leftarrow \quad \arg \min_{r \in V} \mathcal{L}(u^k, t^{k+1}, r, s^k; \lambda_t^k, \lambda_r^k, \lambda_s^k) \qquad (27)$$

$$s^{k+1} \quad \leftarrow \quad \arg \min_{s \in V} \mathcal{L}(u^k, t^{k+1}, r^{k+1}, s; \lambda_t^k, \lambda_r^k, \lambda_s^k) \qquad (28)$$

$$u^{k+1} \quad \leftarrow \quad \arg \min_{u \in V} \mathcal{L}(u, t^{k+1}, r^{k+1}, s^{k+1}; \lambda_t^k, \lambda_r^k, \lambda_s^k) \qquad (29)$$

$$
\begin{pmatrix} \lambda_t^{k+1} \\ \lambda_r^{k+1} \\ \lambda_s^{k+1} \end{pmatrix} \leftarrow \begin{pmatrix} \lambda_t^k - \beta_t \left(t^{k+1} - Du^{k+1} \right) \\ \lambda_r^k - \beta_{rs} \left(r^{k+1} - (Ku^{k+1} - g) \right) \\ \lambda_s^k - \beta_{rs} \left(s^{k+1} - (Ku^{k+1} - g) \right) \end{pmatrix} . \qquad (30)
$$

In the following subsections we show how to solve the four minimization sub-problems for the primal variables t, r, s and u, respectively. Although the minimization sub-problems are all strictly convex and admit a unique solution,

convergence of the overall ADMM algorithm is clearly not guaranteed. We postpone the analysis of convergence of the proposed ADMM scheme to a future extended version of this work.

6.1 Solving the sub-problem for t

The minimization sub-problem for t in (26) can be written as follows:

$$t^{k+1} \;\leftarrow\; \arg\min_{t \in Q} \left\{ \sum_{i=1}^{d^2} \|t_i\|_2 \;-\; \langle \lambda_t^k, t - Du^k \rangle \;+\; \frac{\beta_t}{2} \left\| t - Du^k \right\|_2^2 \right\} \, , \quad (31)$$

and the solution is obtained by the following closed-form shrinkage operators:

$$t_i^{k+1} \;=\; \max\left\{ \|q_i^k\|_2 - \frac{1}{\beta_t} \, , \, 0 \right\} \frac{q_i^k}{\|q_i^k\|_2} \, , \quad i = 1, \ldots, d^2 \, , \quad (32)$$

where $0 \cdot (0/0) = 0$ is assumed, and the constant vectors $q_i^k \in \mathbb{R}^2$ are defined as

$$q_i^k \;:=\; \left(Du^k\right)_i + \frac{1}{\beta_t} \left(\lambda_t^k\right)_i \, , \quad i = 1, \ldots, d^2 \, . \quad (33)$$

The overall cost of this subproblem is linear in the number of pixels d^2.

6.2 Solving the sub-problems for r

The minimization sub-problem for r in (27) is as follows:

$$\begin{aligned} r^{k+1} \;\leftarrow\;& \arg\min_{r \in V} \left\{ \iota_{\mathcal{B}}(r, s^k) - \langle \lambda_r^k, r - \left(Ku^k - g\right) \rangle + \frac{\beta_{rs}}{2} \left\| r - \left(Ku^k - g\right) \right\|_2^2 \right\} \\ \;\leftarrow\;& \arg\min_{r \in \mathcal{B}_r^k} \left\{ \frac{1}{2} \left\| r - v_r^k \right\|_2^2 \right\} \, , \end{aligned} \quad (34)$$

that is, the solution r^{k+1} of (34) is given by the Euclidean projection of the constant (with respect to the optimization variable r) vector

$$v_r^k \;:=\; Ku^k - g + \frac{1}{\beta_{rs}} \lambda_r^k \quad (35)$$

onto the convex feasible set

$$\mathcal{B}_r^k \;:=\; \left\{ r \in V \colon \; -w_\alpha d^2 \le r \star s^k \le w_\alpha d^2 \right\} \, . \quad (36)$$

An efficient solver for the quadratic problem (34)–(36) is given in Section 6.5.

6.3 Solving the sub-problems for s

The minimization sub-problem for s in (28) is as follows:

$$
\begin{aligned}
s^{k+1} \;\leftarrow\; & \arg\min_{s \in V} \left\{ \imath_{\mathcal{B}}(r^{k+1}, s) - \left\langle \lambda_s^k, s - \left(Ku^k - g\right) \right\rangle + \frac{\beta_{rs}}{2} \left\| s - \left(Ku^k - g\right) \right\|_2^2 \right\} \\
\leftarrow\; & \arg\min_{s \in \mathcal{B}_s^k} \left\{ \frac{1}{2} \left\| s - v_s^k \right\|_2^2 \right\} ,
\end{aligned}
\tag{37}
$$

that is, the solution s^{k+1} of (37) is given by the Euclidean projection of the constant (with respect to the optimization variable s) vector

$$
v_s^k \;:=\; Ku^k - g + \frac{1}{\beta_{rs}} \lambda_s^k
\tag{38}
$$

onto the convex feasible set

$$
\mathcal{B}_s^k \;:=\; \left\{ s \in V \colon \; -w_\alpha d^2 \leq r^{k+1} \star s \leq w_\alpha d^2 \right\} .
\tag{39}
$$

An efficient solver for the quadratic problem (37)–(39) is given in Section 6.5.

6.4 Solving the sub-problem for u

The minimization sub-problem for u in (29) can be re-written as follows:

$$
\begin{aligned}
u^{k+1} \;\leftarrow\; \arg\min_{u \in V} \Big\{ & -\left\langle \lambda_t^k, t^{k+1} - Du \right\rangle + \frac{\beta_t}{2} \left\| t^{k+1} - Du \right\|_2^2 \\
& -\left\langle \lambda_r^k, r^{k+1} - (Ku - g) \right\rangle + \frac{\beta_{rs}}{2} \left\| r^{k+1} - (Ku - g) \right\|_2^2 \\
& -\left\langle \lambda_s^k, s^{k+1} - (Ku - g) \right\rangle + \frac{\beta_{rs}}{2} \left\| s^{k+1} - (Ku - g) \right\|_2^2 \Big\} .
\end{aligned}
$$

The above quadratic minimization problem can be solved for u^{k+1} by computing the solution of the following $d^2 \times d^2$ linear system:

$$
\begin{aligned}
\left(D^T D + 2\frac{\beta_{rs}}{\beta_t} K^T K \right) u \;=\; & D^T \left(t^{k+1} - \frac{1}{\beta_t} \lambda_t^k \right) \\
& + \frac{\beta_{rs}}{\beta_t} K^T \left(r^{k+1} + s^{k+1} - \frac{1}{\beta_{rs}} \lambda_r^k - \frac{1}{\beta_{rs}} \lambda_s^k + 2g \right)
\end{aligned}
\tag{40}
$$

The coefficient matrix of the linear system in (40) is symmetric positive definite, highly sparse and, under the assumption of periodic boundary conditions for u, block-circulant with circulant blocks. Hence, at each ADMM iteration the linear system in (40) can be solved by one forward and one inverse FFT, each at a cost of $O(d^2 \log d^2)$. We remark that symmetric or anti-symmetric boundary conditions could be assumed as well, thus simply replacing FFT with fast cosine or sine transforms and retaining the same computational complexity.

6.5 Computing projection

In this subsection, we are interested in solving efficiently quadratic programs of the form of the sub-problems for r and s , see equations (34) and (37). Let $v \in V$ and $m \in V$ be given (vectorized) images and $b \in \mathbb{R}_+$ a given scalar positive bound. We want to solve:

$$x^* = \arg \min_{x \in \mathcal{S}_x} \left\{ \frac{1}{2} \|x - v\|_2^2 \right\} , \tag{41}$$

where $\mathcal{S}_x \subseteq V$ is the convex feasible set

$$\mathcal{S}_x = \left\{ x \in V : -b \leq m \star x \leq b \right\} , \tag{42}$$

with $\star$ denoting the 2-D circular correlation operator.

In order to solve (41), we use the ADMM procedure. By introducing the auxiliary variable $y \in V$, (41) can be rewritten in the following equivalent form:

$$\{x^*, y^*\} = \arg \min_{x,y \in V} \left\{ \imath_{\mathcal{S}_y}(y) + \frac{1}{2} \|x - v\|_2^2 \right\} \text{ s.t. :} \quad y = Mx , \tag{43}$$

where $\mathcal{S}_y \subseteq V$ is the convex feasible set

$$\mathcal{S}_y = \left\{ y \in V : -b \leq y \leq b \right\} , \tag{44}$$

and $M \in \mathbb{R}^{d^2 \times d^2}$ is the matrix associated with the linear operator corresponding to the circular correlation with the image m. To solve (43)–(44), we define the augmented Lagrangian functional

$$\mathcal{L}_p(x, y; \lambda_p) = \imath_{\mathcal{S}_y}(y) + \frac{1}{2} \|x - v\|_2^2 - \langle \lambda_p , y - Mx \rangle + \frac{\beta_p}{2} \| y - Mx \|_2^2 , \tag{45}$$

where $\beta_p > 0$ is a scalar penalty parameter and $\lambda_p \in V$ is the vector of Lagrange multipliers associated with the system of linear constraints in (43).

In the following we outline a standard ADMM iterative procedure applied to solve (45).

Given the previously computed (or initialized for $j = 0$) vectors x^j, and λ_p^j, the j-th iteration of the proposed ADMM-based iterative scheme [46] applied to the solution of problem (43)–(44) reads as follows:

$$y^{j+1} = \arg \min_{y \in V} \mathcal{L}_p(x^j, y; \lambda_p^j) \tag{46}$$

$$x^{j+1} = \arg \min_{x \in V} \mathcal{L}_p(x, y^{j+1}; \lambda_p^j) \tag{47}$$

$$\lambda_p^{j+1} = \lambda_p^j - \beta_p \left(y^{j+1} - Mx^{j+1} \right) \tag{48}$$

In particular, the minimization sub-problem for y in (46) is as follows:

$$
\begin{aligned}
y^{j+1} &= \arg\min_{y \in V} \left\{ \imath_{\mathcal{S}_y}(y) - \langle \lambda_p^j, y - Mx^j \rangle + \frac{\beta_p}{2} \left\| y - Mx^j \right\|_2^2 \right\} \\
&= \arg\min_{y \in V} \left\{ \imath_{\mathcal{S}_y}(y) + \frac{\beta_p}{2} \left\| y - \left(Mx^j + \frac{1}{\beta_p} \lambda_p^j \right) \right\|_2^2 \right\} \\
&= \arg\min_{y \in \mathcal{S}_y} \left\{ \frac{1}{2} \left\| y - z^j \right\|_2^2 \right\} ,
\end{aligned}
\tag{49}
$$

that is, the solution y^{j+1} of (49) is given by the Euclidean projection of the constant (with respect to the optimization variable y) vector

$$
z^j \; := \; Mx^j + \frac{1}{\beta_p} \lambda_p^j
\tag{50}
$$

onto the convex set $\mathcal{S}_y$ defined in (44), which represents a box-constraint for y, such that the projection admits the closed-form solution

$$
y^{j+1} = \; \min \left\{ \max \left\{ z^j \, , \, -b \right\} , \, b \right\} .
\tag{51}
$$

Given the augmented Lagrangian functional in (45), the minimization sub-problem for x in (47) is as follows:

$$
x^{j+1} = \arg\min_{x \in V} \left\{ \frac{1}{2} \| x - v \|_2^2 + \langle \lambda_p^j, Mx \rangle + \frac{\beta_p}{2} \left\| y^{j+1} - Mx \right\|_2^2 \right\} , \quad
\tag{52}
$$

where constant terms have been omitted. We notice that (52) is a quadratic minimization problem whose first-order optimality conditions lead to the following linear system:

$$
\left(\frac{1}{\beta_p} I + M^T M \right) x = \frac{1}{\beta_p} v + M^T \left(y^{j+1} - \frac{1}{\beta_p} \lambda_p^j \right) ,
\tag{53}
$$

such that the new iterate is obtained as follows:

$$
x^{j+1} = \underbrace{\left(\frac{1}{\beta_p} I + M^T M \right)^{-1} \frac{1}{\beta_p} v}_{\tilde{v}} + \underbrace{\left(\frac{1}{\beta_p} I + M^T M \right)^{-1} M^T}_{\widetilde{C}} \left(y^{j+1} - \frac{1}{\beta_p} \lambda_p^j \right) .
$$

$$
\tag{54}
$$

The vector $\tilde{v}$ and the matrix $\widetilde{C}$ in (54) are constant during iterations, hence they are computed once and for all at the beginning. In particular, (54) is solved by FFT.

6.6 ADMM-based minimization algorithm

To summarize previous results, in Algorithm 1 we report the main steps of the proposed ADMM-based iterative scheme used to solve the minimization problem (23)–(24) associated to our restoration model (2) with the whiteness set defined in (17)–(18).

Recently, many works have been proposed to study the convergence properties for the ADMM-based algorithms for non-convex optimization problems, see [21,30,43,47]. Nevertheless the convergence of our proposed ADMM scheme, which presents a non-convex non-smooth objective with multiple variable splitting and multiple penalty parameters, is not easy to derive relying on the results presented so far. We thus postpone the investigation on the theoretical convergence of our ADMM scheme to a future dedicated work. However, we will provide some evidence of the numerical convergence in the experimental section.

Algorithm 1 ADMM-based iterative scheme for the solution of problem (23)–(24)

---

inputs: observed corrupted image $g \in \mathbb{R}^{d^2}$, noise standard deviation σ_n

output: approximate solution $u^* \in \mathbb{R}^{d^2}$ of (2)–(17)

parameters: MODEL: whiteness coefficient $\alpha \geq 0 \to$ whiteness bound w_α by (22)

                  ADMM: penalty parameters $\beta_t, \beta_{rs}, \beta_p > 0$

1. initialization: $u^{(0)} = b$, $\rho^{(0)} = 0$
2. **for** $k = 0, 1, 2, \ldots$ *until convergence*   **do**:
3.   · **update primal variables:**
4.   · compute t^{k+1} by (33)–(32)
5.   · compute r^{k+1} by computing the projection (34)–(36)
6.   · compute s^{k+1} by computing the projection (37)–(39)
7.   · compute u^{k+1} by solving the linear system (40)
8.   · **update dual variables:**

$$\lambda_t^{k+1} = \lambda_t^k - \beta_t \left(t^{k+1} - Du^{k+1} \right)$$

9.   · compute $\lambda_r^{k+1} = \lambda_r^k - \beta_{rs} \left(r^{k+1} - (Ku^{k+1} - g) \right)$

$$\lambda_s^{k+1} = \lambda_s^k - \beta_{rs} \left(s^{k+1} - (Ku^{k+1} - g) \right)$$

10. **end for**
11. $u^* = u^{k+1}$

---

7 Numerical Experiments

This section is devoted to evaluate the effectiveness of the proposed TV-W restoration model (2) when applied to images synthetically corrupted by space-invariant Gaussian blur and additive white noises of different types among uniform (AWUN), Gaussian (AWGN), Laplacian (AWLN) and truncated Cauchy (AWCN). We consider three test images - `geometry` (256×256 pixels), `skyscrapers` (256×256 pixels) and `checkboard` (200×200 pixels) - which contain flat regions, neat edges and textures. The experiments have been carried out on test problems where, in accordance with the considered degradation model (1), the synthetically corrupted images were obtained by blurring the original image (the blur kernel is created through the MATLAB command `fspecial('Gaussian',band,sigma)`) and then adding a realization of white noise with standard deviation σ_n. In particular, we compare the performance of our TV-W model with that of the unconstrained variational model

$$u^* \leftarrow \arg \min_{u \in \mathbb{R}^{d^2}} \left\{ \mathrm{TV}(u) + \lambda \, \mathcal{F}(u; K, g) \right\} , \tag{55}$$

that is the model having the same TV regularizer and a suitable fidelity term according to the noise type - see Table 1 - referred to as TV-F model. More precisely, for AWGN-corrupted images we used the TV-ℓ_2 model implemented by the ADMM-based algorithm contained in the software package freely available at [49]. For AWLN-corrupted images, we solved the TV-ℓ_1 model by implementing the ADMM-based algorithm presented in [48], for AWCN-corrupted images we used the code provided by the authors of [42], while for AWUN-corrupted images we implemented an ad hoc ADMM-based algorithm for solving the TV-ℓ_∞ model. For all the examples, the regularization parameter λ of the TV-F model (55) has been hand-tuned so that the solution satisfies the discrepancy principle, that is the variance of the residue image $g - Ku^*$ must be (approximately) equal to the variance σ_n^2 of the synthetically added noise. In the interest of a fair comparison, for all the examples in our TV-W model we set the standard deviation σ_r of the residue image auto-correlation - see the definition of σ_r in Section 4.1 - by taking the true value σ_n^2 of the noise variance.

For what concerns the computation of the whiteness bound by (22), numerical tests indicated that the performance of TV-W are quite robust to choices of the whiteness coefficient α. Hence, we fixed $\alpha = 2.5$ for all the experiments. We used the following parameters for the ADMM algorithm: $\beta_t = 200$, $\beta_{rs} = 150$, $\beta_p = 1$. As approximate solution u^* of all the image restoration problems we consider u^k obtained as soon as two successive iterates satisfy the condition

$$\delta^k := \frac{\|u^k - u^{k-1}\|_2}{\|u^{k-1}\|_2} < 10^{-4}, \tag{56}$$

or after a maximum number of 500 iterations.

Finally, since the TV-W model is non-convex, it is also worth mentioning the initial iterate of the ADMM algorithm which can affect the final results. In

general, good performance has been obtained using the constant image with gray level equal to the average of the observed corrupted image g as the initial iterate. For what concerns the tests on images corrupted by Cauchy noise, a detailed discussion can be found in [33]. In this case, we set as initial iterate of the ADMM algorithm the result of a simple median filter applied to the observed corrupted image.

The quality of the corrupted and restored images is measured by the Blurred Signal-to-Noise Ratio (BSNR) and the Improved Signal-to-Noise Ratio (ISNR):

$$\text{BSNR}(u, g) = 10 \log_{10} \frac{\|Ku - E[Ku]\|_2^2}{\|Ku - g\|_2^2}, \quad \text{ISNR}(u, g, u^*) = 10 \log_{10} \frac{\|g - u\|_2^2}{\|u^* - u\|_2^2},$$

where $u, g, u^* \in \mathbb{R}^{d^2}$ are the uncorrupted, corrupted and restored images, respectively, and where $E[Ku]$ represents the mean gray level of image Ku. The ISNR provides a measure of the quality of the restored image: a high ISNR value indicates that u^* is an accurate approximation of u.

For a quantitative assessment, in Table 2 we report the results obtained with one noise realization for each of the discussed white noise distributions. In particular, we report in Table 2 the restoration results obtained for the test images `geometry`, `skyscrapers` and `checkboard` corrupted by different Gaussian blurs, namely with kernels defined by parameters `band = 7`, `sigma = 1.5` for `geometry` and `checkboard` and `band = 5`, `sigma = 1.5` for `skyscrapers`, and by AWUN, AWGN, AWLN, AWCN with standard deviations $\sigma_n \in \{5, 10\}$. In the columns labeled TV-F and TV-W we report the ISNR values achieved by the compared models. For TV-W we also display, in parenthesis, the standard deviation (or, better, the l_2-norm) of the residue image, in formulas $\|g - Ku^*\|_2$. From the results in Table 2, we can state that the proposed general-purpose TV-W model significantly outperforms the specific TV-F models for all the considered examples. It is worth noting that the standard deviations of the residue images of the TV-W model are always approximately equal to the standard deviation of the noise corruption, even if no explicit constraint on this quantity is contained in the whiteness set $\mathcal{W}_\alpha$ in (17).

The numerical experiments carried out clearly indicate that the proposed TV-W model holds the potential not only for dealing effectively with additive white noises of unknown distribution but also for yielding restorations of higher quality than those obtainable by noise-specific models, especially for images containing textures.

To provide further evidence of that, in Figures 5-8 we show qualitative, visual restoration results obtained on the `geometry` test image corrupted by AWUN, AWGN, AWLN and AWCN of standard deviation $\sigma_n = 10$, respectively. Together with the original and the corrupted images (first column), we report in the second column of Figures 5-8 the restored images obtained by the TV-F (top) and TV-W (bottom) models, and in the third column the associated residue images. From a visual inspection of the restored and residue images, one can notice how the TV-W model well recovers the textured re-

		$\sigma_n = 5$			$\sigma_n = 10$		
IMAGE	NOISE	BSNR	TV-F	TV-W	BSNR	TV-F	TV-W
geometry	AWUN	19.55	3.10	4.98 (5.04)	13.54	2.26	4.95 (10.03)
	AWGN	19.54	2.42	5.02 (5.04)	13.54	1.08	5.04 (10.03)
	AWLN	19.57	3.14	5.01 (5.03)	13.55	1.44	4.96 (10.00)
	AWCN	19.60	4.26	6.09 (5.01)	13.57	3.78	7.60 (9.99)
skyscrapers	AWUN	18.39	1.40	1.92 (5.06)	12.37	1.14	1.44 (10.09)
	AWGN	18.37	1.14	1.88 (5.07)	12.35	0.63	1.47 (10.09)
	AWLN	18.40	1.46	1.88 (5.07)	12.38	0.84	1.45 (10.09)
	AWCN	18.26	2.04	2.48 (5.19)	12.26	1.54	2.36 (10.20)
checkboard	AWUN	23.48	8.52	10.28 (5.06)	17.46	6.43	8.78 (10.09)
	AWGN	23.47	8.16	10.06 (5.07)	17.47	5.68	8.43 (10.09)
	AWLN	23.46	8.23	10.05 (5.00)	17.44	6.34	8.29 (10.00)
	AWCN	23.49	10.08	10.44 (5.00)	17.48	8.19	9.37 (10.00)

Table 2 ISNR values obtained by restoring the test images `geometry`, `skyscrapers`, and `checkboard`.

gions while the residuals produced by the TV-F models contain parts of the textures which, hence, are not present in the restored images.
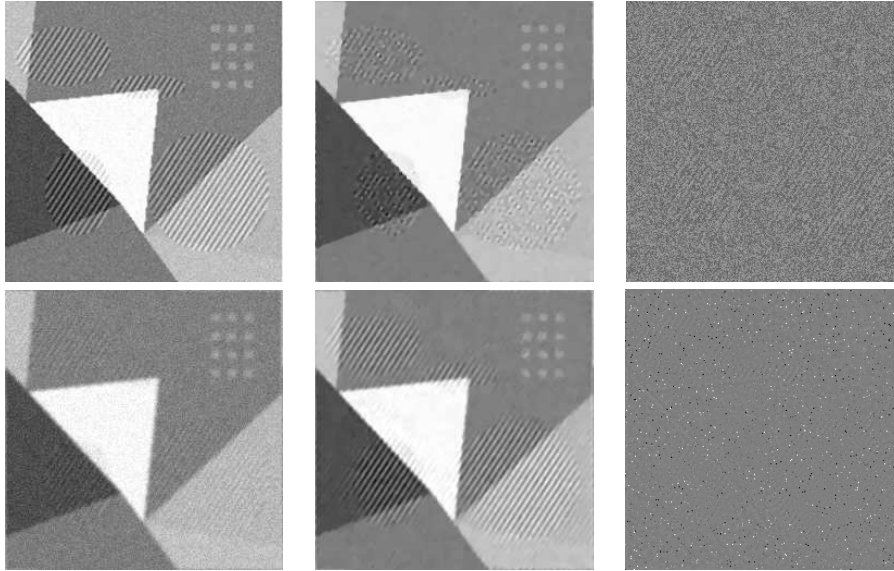

Fig. 5 Visual restoration results for the `geometry` test image corrupted by blur and AWUN of standard deviation $\sigma_n = 10$. First column: original image (top) and corrupted image (bottom). Second and third column: restored and residue images with TV-F (top) and TV-W (bottom) models, respectively.
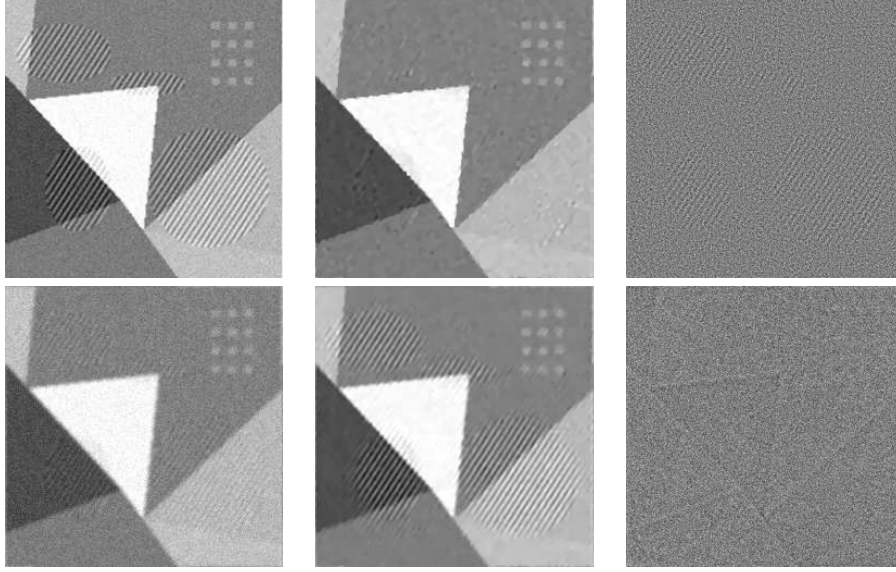
Fig. 6 Visual restoration results for the `geometry` test image corrupted by blur and AWGN of standard deviation $\sigma_n = 10$. First column: original image (top) and corrupted image (bottom). Second and third column: restored and residue images with TV-F (top) and TV-W (bottom) models, respectively.

In order to highlight more clearly the ability of the proposed TV-W model in dealing effectively with additive white noises characterized by unknown, possibly complex distributions, we present a further example consisting in the restoration of the `checkboard` test image corrupted by space-invariant Gaussian blur of parameters `band` $= 7$, `sigma` $= 1.5$, and by a balanced mixture of AWUN, AWGN and AWLN of total standard deviation $\sigma_n = 20$ - that is, the additive mixed noise corruption is the sum of the realizations of AWUN, AWGN and AWLN having the same standard deviation equal to $20\sqrt{3}/3$. Since the fidelity term associated - according to the MAP principle - with this kind of mixed noise is potentially very complicated and, indeed, has not been derived so far, we compare the performance of our TV-W model with those of the three models TV-ℓ_∞, TV-ℓ_2 and TV-ℓ_1 which represent the best MAP-based choices associated with the separate AWUN, AWGN and AWLN noise components, respectively. In Figure 9 we show the original image, the corrupted image and the restored images computed by the four compared models together with the associated ISNR values. It appears evident also from these results how the proposed TV-W model can potentially provide high quality restorations for images corrupted by additive white noises characterized by a wide range of distributions.

As previously stated, the key idea of our proposal consists of a new general-purpose fidelity term (in the form of a residual whiteness constraint) that can be coupled with whichever regularizer. In this paper we have considered the
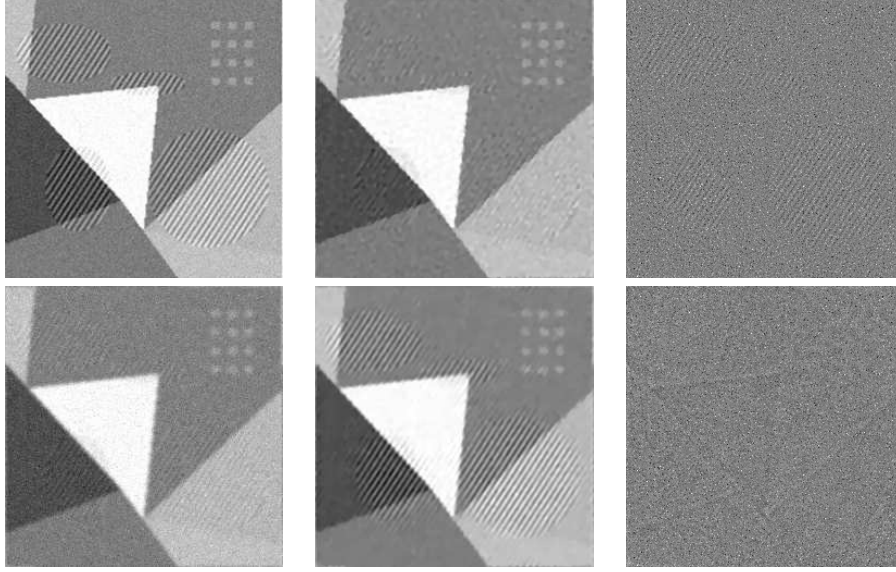
Fig. 7 Visual restoration results for the `geometry` test image corrupted by blur and AWLN of standard deviation $\sigma_n = 10$. First column: original image (top) and corrupted image (bottom). Second and third column: restored and residue images with TV-F (top) and TV-W (bottom) models, respectively.

		$\sigma_n = 10$		
IMAGE	BSNR	TIK-F	TIK-W	TV-W
geometry	13.54	0.80	3.92	5.04
skyscrapers	12.35	0.53	1.21	1.47
checkboard	17.47	1.49	2.65	8.43

Table 3 ISNR values obtained by restoring the test images `geometry`, `skyscrapers`, and `checkboard` using the Tikhonov instead of the TV regularizer.

popular TV regularizer, hence in the previous experiments we compared our TV-W model with models containing the same TV regularizer and suitable fidelity terms according to the specific noise corruption. We are confident that by replacing the TV regularizer in TV-W with more powerful regularizers such as, e.g., TV_p [23], nonlocal TV [18], nonlocal means [6], IBM3D [29], and RED [51], higher quality results can be achieved. However, in order to emphasize the capabilities of the proposed fidelity term, it is clarifying to test the performance of the variational model obtained by coupling it with a basic well-known Tikhonov regularizer, that is

$$u^* \leftarrow \arg \min_{u \in \mathcal{W}_\alpha} \|\mathrm{D}u\|_2^2 \,, \tag{57}$$

where D is the same first-order finite difference matrix as in (23). Table 3 reports the ISNR results obtained by model (57), referred to as TIK-W, the
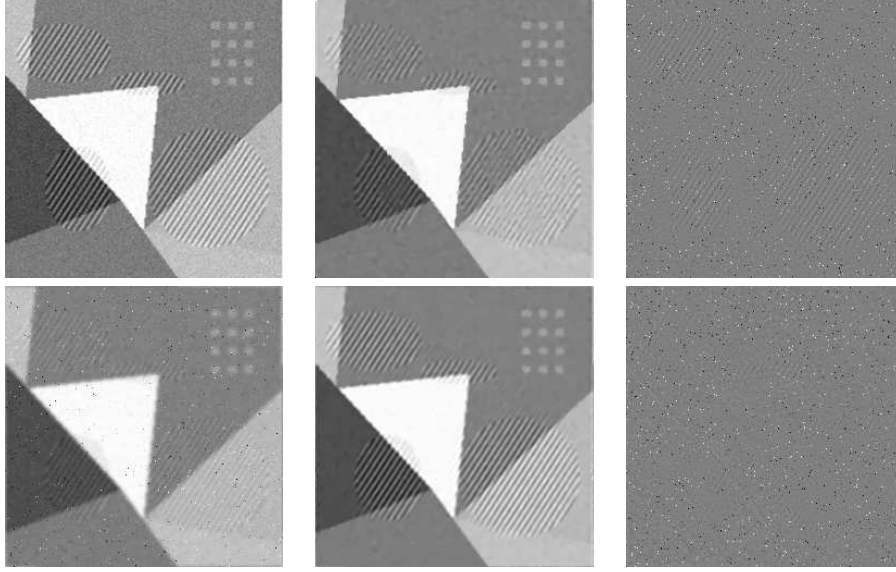
Fig. 8 Visual restoration results for the `geometry` test image corrupted by blur and AWCN of standard deviation $\sigma_n = 10$. First column: original image (top) and corrupted image (bottom). Second and third column: restored and residue images with TV-F (top) and TV-W (bottom) models, respectively.

unconstrained model obtained by coupling the Tikhonov regularizer with the ℓ_2 fidelity term (TIK-F), and the proposed TV-W model (reported also in Table 2), when applied to the restoration of the three test images `geometry`, `skyscrapers` and `checkboard` corrupted by space-invariant Gaussian blur (same parameters `band` and `sigma` as in the previous examples) and by AWGN of standard deviation $\sigma_n = 10$. The results confirm that replacing the ideal MAP-based fidelity term with the proposed whiteness constraint yields to a remarkable improvement of the restoration quality. In Figure 10 we show the restored and residue images obtained by applying the TIK-F and TIK-W models to the `geometry` test image. By visual inspection we notice how in spite of the very basic regularizer used our whiteness constraint allows to restore effectively the textured regions such that the residue image resembles a white noise image.

Finally, in order to assess the good performance of the proposed TV-W model with respect to the state-of-the-art for image deblurring, we compared it with the IBM3D method [29] on some of the previous examples. By the way of illustration, in Figure 11 we show the restored images obtained by our TV-W model (left) and by the IBM3D method (right). We notice that even if the ISNR values are slightly worse, from a visual inspection we can appreciate the better quality of our restored image.

We conclude this section by presenting an empirical investigation on the numerical convergence of the proposed ADMM-based minimization scheme.
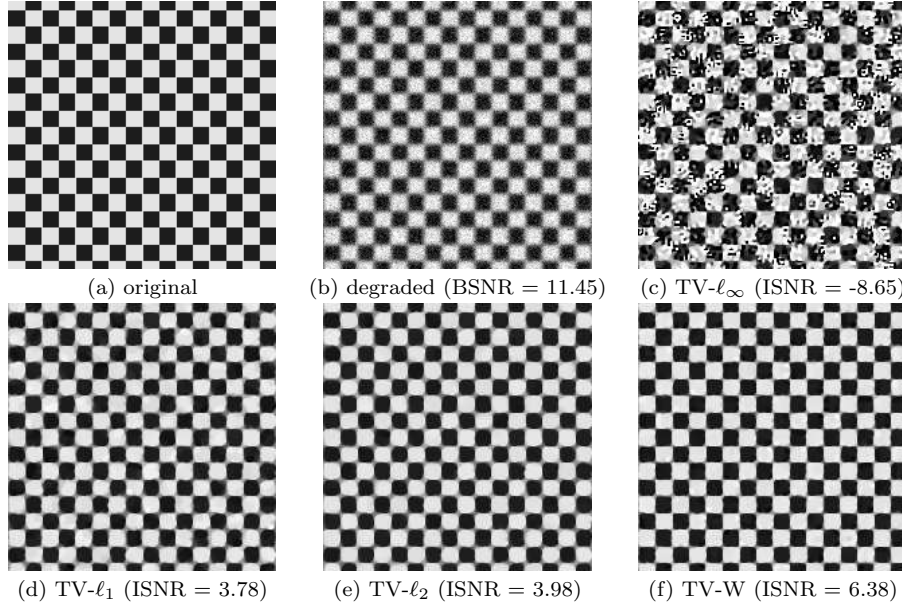
(a) original   (b) degraded (BSNR = 11.45)   (c) TV-ℓ_∞ (ISNR = -8.65)

(d) TV-ℓ_1 (ISNR = 3.78)   (e) TV-ℓ_2 (ISNR = 3.98)   (f) TV-W (ISNR = 6.38)

Fig. 9 Restoration results for the `checkboard` test image corrupted by space-invariant Gaussian blur of parameters `band` = 7, `sigma` = 1.5, and by a balanced mixture of AWUN, AWGN and AWLN of total standard deviation $\sigma_n = 20$.

Towards this aim, in Figure 12 we report some convergence plots concerning the test image `geometry` corrupted by space-invariant Gaussian blur of parameters `band` = 7, `sigma` = 1.5, and by AWGN of standard deviation $\sigma_n = 5$ (left column) and by AWLN of standard deviation $\sigma_n = 5$ (right column); results associated to the second and third rows of Table 2. In particular, the plots in the first and the second row of Figure 12 represent the logarithm of the relative change δ^k of the iterates defined in (56), and the values of the objective function defined in (2), versus the iteration index k for the first 2000 iterations of Algorithm 1. The plots in Figure 12 indicate the good convergence behavior of the proposed ADMM-based Algorithm 1. In particular, it is worth observing from the plots in the first row that less than 100 iterations are required to drop below the stopping criterion threshold defined in (56). To finally investigate the robustness of the ADMM-based Algorithm to the choice of the initial guess - which is important due to non-convexity of the proposed model - we run the algorithm on the same examples considered in Table 2 by using the observed image g instead of its average as the ADMM initial guess. We observed that at numerical convergence (after 4000 iterations) the achieved ISNR values associated to the restored images for the two different initial guesses differ by less than 10^{-3}. This could indicate that, at least in the considered cases, the same local minimizer has been achieved.
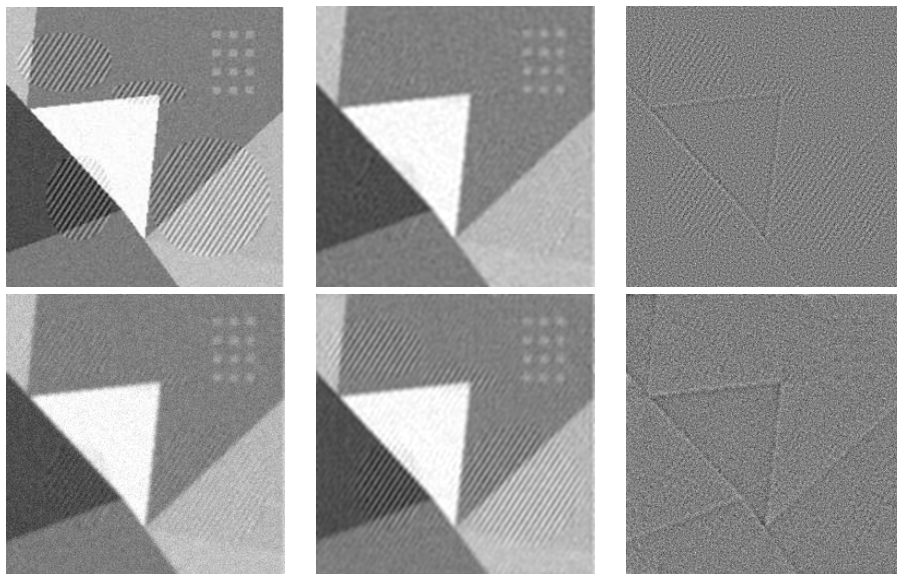
Fig. 10 Visual restoration results for the `geometry` test image corrupted by blur and AWGN of standard deviation $\sigma_n = 10$. First column: original image (top) and corrupted image (bottom). Second and third column: restored and residue images with TIK-F (top) and TIK-W (bottom) models, respectively.
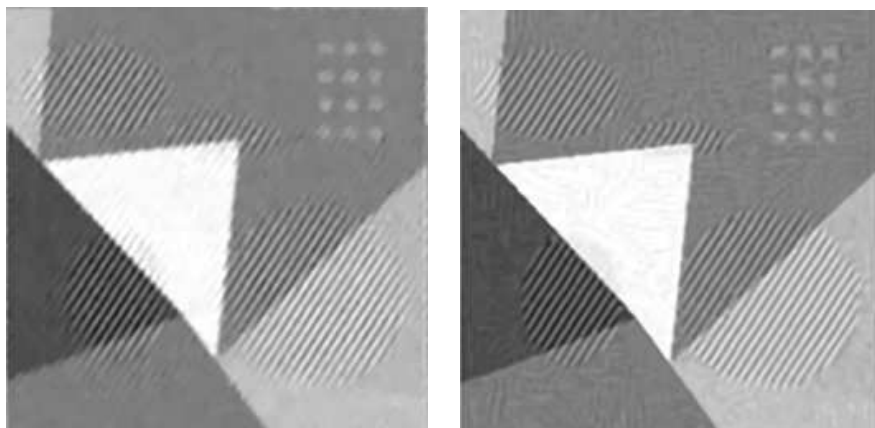

Fig. 11 Visual restoration results for the `geometry` test image corrupted by space-invariant Gaussian blur of parameters `band` $= 7$, `sigma` $= 1.5$ and by AWGN of standard deviation $\sigma_n = 10$. (left) restored image with TV-W, ISNR $= 5.82$, and (right) restored image with IBM3D, ISNR $= 6.01$.

8 Conclusions

We presented a new robust variational model for the restoration of gray level images corrupted by blur and by the general class of additive white noises.
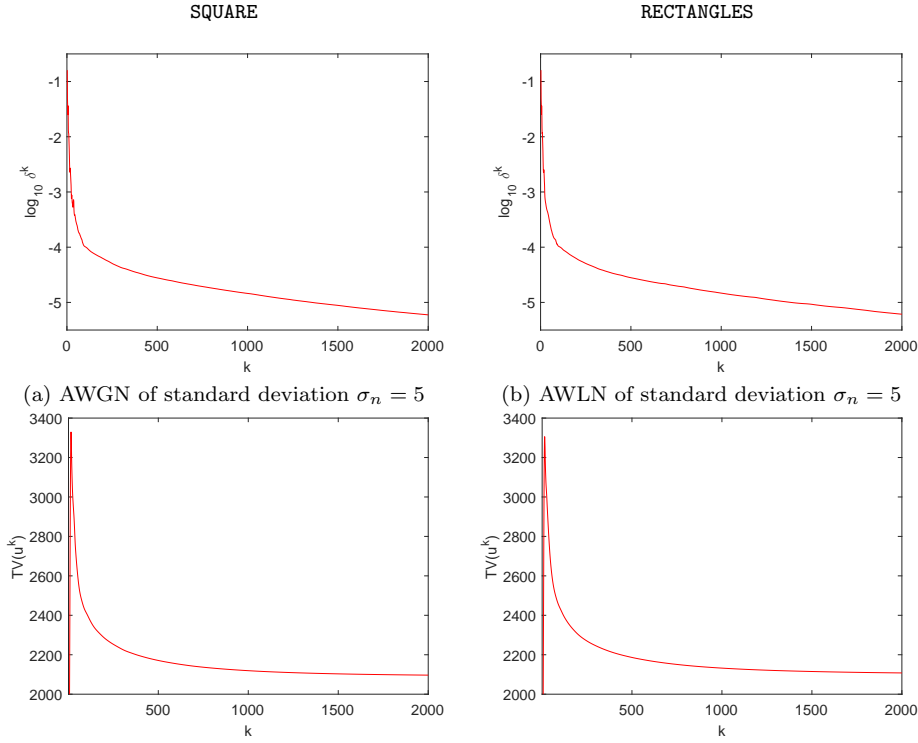
Fig. 12 Empirical convergence related to the test image `geometry` corrupted by space-invariant Gaussian blur of parameters `band = 7`, `sigma = 1.5` and by AWGN of standard deviation $\sigma_n = 5$ (left column) and by AWLN of standard deviation $\sigma_n = 5$ (right column).

This class includes important noises such as those characterized by Gaussian, uniform, Laplacian and Cauchy distributions, which can be found in many applications, such as e.g. medical and astronomic imaging.

The proposed constrained model is non-smooth and non-convex and the feasible set - referred to as the whiteness set - contains restored images u such that the associated residue images $Ku - g$ resemble the realization of a white noise process. We provided well-founded guidelines for setting the whiteness constraint limits, in particular we derived theoretical and empirical results on the distribution of the sample auto-correlation values. The variational model is solved by an efficient ADMM-based algorithm which reduces the solution to a sequence of convex optimization sub-problems. Experimental comparisons with the variational models having suitable fidelity terms demonstrate the effectiveness of the proposed model, especially for images with textured parts.

The main advantage of our proposal is that it provides a unified framework for deblurring images corrupted by white noises characterized by any finite-variance distribution. This is of particular importance since in most real applications the physics of image acquisition yields noise distributions which

are difficult to be known a priori and, in any case, can be very complex as they are the outcome of several noise sources.

Future work will investigate theoretical convergence of the proposed ADMM numerical scheme. Furthermore, the possibility to extend the model to deal with Poisson and multiplicative noises and to color images will also be considered. The extension to Poisson and multiplicative noises should be quite straightforward by using the Anscombe transformation and log-transformation, respectively; while the generalization to color images is not obvious, mainly due to the per-pixel inter-channel correlation.

Acknowledgements

References

1. M. Artina, M. Fornasier, and F. Solombrino, Linearly constrained non-smooth and non-convex minimization, SIAM Journal on Optimization 23(3), 1904-1937, 2013.
2. H. Attouch, J. Bolte, and B. F. Svaiter, Convergence of descent methods for semi-algebraic and tame problems: proximal algorithms, forward-backward splitting, and regularized Gauss-Seidel methods, Mathematical Programming 137(1-2), 91-129, 2013.
3. G. Aubert and J. Aujol, A variational approach to removing multiplicative noise, SIAM J. Appl. Math., 68, 925-946 (2008).
4. S. P. Awate, R. T. Whitaker, Unsupervised, information-theoretic, adaptive image filtering for image restoration. IEEE Transactions on pattern analysis and machine intelligence, 28(3), 364-376, 2006.
5. K. Bredies, A. L. Dirk, and S. Reiterer, Minimization of non-smooth, non-convex functionals by iterative thresholding, Journal of Optimization Theory and Applications 165(1), 78-112, 2015.
6. A. Buades, B. Coll, and J.-M. Morel, A review of image denoising algorithms, with a new one, SIAM Journal Multiscale Model. Simul., 4(2), 490-530, 2005.
7. A. Ben-Israel, A Concentrated Cauchy Distribution with Finite Moments, Annals of Operations Research, 208(1), 147–153, 2013.
8. C. Bouman, K. Sauer, A Generalized Gaussian Image Model for Edge-preserving MAP Estimation, IEEE Trans. on Image Processing, 2(3), 296–310, 1993.
9. A.C. Bovik, Handbook of Image and Video Processing, Academic press, 2010.
10. P.J. Brockwell, and R. A. Davis. Introduction to time series and forecasting. Springer, 2002.
11. S. D. Babacan, R. Molina, A. K. Katsaggelos, Parameter estimation in TV image restoration using variational distribution approximation, IEEE transactions on image processing 17.3, 326-339, 2008.
12. Brunet, Dominique, Edward R. Vrscay, and Zhou Wang, The use of residuals in image denoising, International conference image analysis and recognition. Springer, Berlin, Heidelberg, 2009.
13. R. H. Chan, A. Lanza, S. Morigi, F. Sgallari, An Adaptive Strategy for the Restoration of Textured Images Using Fractional Order Regularization, Numerical Mathematics: Theory, Methods and Applications, 6(1), 276–296, 2013.
14. C. Chaux, L. Duval, A. Benazza-Benyahia and J. C. Pesquet, A Nonlinear Stein-Based Estimator for Multichannel Image Denoising, in IEEE Transactions on Signal Processing, 56(8), 3855-3870, 2008.

15. Diananda, P. H., and M. S. Bartlett, Some probability limit theorems with statistical applications, Mathematical Proceedings of the Cambridge Philosophical Society, 49(2), Cambridge University Press, 1953.
16. T.F. Chan, J.J. Shen, Image Processing and Analysis: Variational, PDE, Wavelet, and Stochastic Methods. Siam, 2005.
17. C. Clason, L^∞ Fitting for Inverse Problems with Uniform Noise, Inverse Problems, vol. 28(10), 104007, 2012.
18. G. Gilboa, S. Osher, Nonlocal Operators with Applications to Image Processing, Multiscale Modeling & Simulation, 7(3), 1005–1028, 2009.
19. J. Fehrenbach, M. Nikolova, G. Steidel and P. Weiss, Bilevel image denoising using gaussianity tests, Proc. SSVM 2015.
20. Per Christian Hansen, *Discrete inverse problems: insight and algorithms*, 7, SIAM, 2010.
21. M. Hong, Z. Luo, and M. Razaviyayn, Convergence analysis of alternating direction method of multipliers for a family of non-convex problems, SIAM Journal on Optimization 26(1), 337-364, 2016.
22. D. Keren, M. Werman, Probabilistic Analysis of Regularization, IEEE Transactions on Pattern Analysis and Machine Intelligence, 15(10), 982–995, 1993.
23. A. Lanza, S. Morigi, F. Sgallari, Constrained TV_p-ℓ_2 Model for Image Restoration, Journal of Scientific Computing, 68(1), 64–91, 2016.
24. A. Lanza, S. Morigi, F. Sgallari, Y.W. Wen, Image Restoration with Poisson-Gaussian Mixed Noise, Computer Methods in Biomechanics and Biomedical Engineering: Imaging and Visualization, 2(1), 12–24, 2013.
25. A. Lanza, S. Morigi, F. Sgallari, A.J. Yezzi, Variational Image Denoising Based on Autocorrelation Whiteness, SIAM Journal on Imaging Sciences, 6(4), 1931–1955, 2013.
26. A. Lanza, S. Morigi, F. Sgallari, Variational Image Restoration with Constraints on Noise Whiteness, Journal of Mathematical Imaging and Vision, 53(1), 61–77, 2015.
27. A. Lanza, F. Sciacchitano, S. Morigi, and F. Sgallari, A Unified Framework for the Restoration of Images Corrupted by Additive White Noise, International Conference on Scale Space and Variational Methods in Computer Vision, 498-510. Springer, Cham, 2017.
28. D. Lazzaro, S. Morigi, P. Melpignano, E. Loli Piccolomini, L. Benini, Image enhancement variational methods for Enabling Strong Cost Reduction in OLED-based Point-of-Care Immunofluorescent Diagnostic Systems, International Journal for Numerical Methods in Biomedical Engineering, DOI: 10.1002/cnm.2932, 34(3), e2932, 2018.
29. A. Danielyan, V. Katkovnik, K. Egiazarian, BM3D frames and variational image deblurring. IEEE Transactions on image processing, 21(4), 1715-1728, 2012.
30. G. Li, and T. K. Pong, Global convergence of splitting methods for nonconvex composite optimization, SIAM Journal on Optimization 25(4), 2434-2460, 2015.
31. F. Luisier, T. Blu, and M. Unser, Image denoising in mixed Poisson-Gaussian noise, IEEE Transactions on Image Proc., 20(3), 696-708, 2011.
32. M. Makitalo and A. Foi, Optimal inversion of the Anscombe transformation in low-count Poisson image denoising, Image Processing, IEEE Trans. on 20(1), 99-109, 2011.
33. JJ Mei, Y Dong, TZ Huang, W Yin, Cauchy Noise Removal by Nonconvex ADMM with Convergence Guarantees, Journal of Scientific Computing, DOI 10.1007/s10915-017-0460-5, 2017.
34. J. Neyman, Outline of a theory of statistical estimation based on the classical theory of probability, Philosophical Trans. of the Royal Society of London. Series A, Mathematical and Physical Sciences, 236(767), 333-380, 1937.
35. M. Nikolova, A Variational Approach to Remove Outliers and Impulse Noise, Journal of Mathematical Imaging and Vision, 20(1), 99–120, 2004.
36. P. Ochs, A. Dosovitskiy, T. Brox, and T. Pock, On iteratively reweighted algorithms for nonsmooth nonconvex optimization in computer vision, SIAM Journal on Imaging Sciences, 8(1), 331-372, 2015.
37. P. Ochs, Y. Chen, T. Brox, and T. Pock. iPiano: Inertial proximal algorithm for nonconvex optimization. SIAM Journal on Imaging Sciences, 7(2), 1388-1419, 2014.
38. J. Portilla, V. Strela, M. J. Wainwright, and E. P. Simoncelli, Image denoising using scale mixtures of Gaussians in the wavelet domain, IEEE Trans. Image Process., 12 , 1338-1351, 2003.

39. L. Rudin, S. Osher, E. Fatemi, Nonlinear Total Variation Based Noise Removal Algorithms, Phys. D, 60(1-4), 259–268, 1992.
40. Riot, P., Almansa, A., Gousseau, Y., and Tupin, F. (2017, June). A Correlation-based Dissimilarity Measure for Noisy Patches. In International Conference on Scale Space and Variational Methods in Computer Vision (pp.184-195). Springer, Cham.
41. O. Scherzer, ed. Handbook of mathematical methods in imaging, Springer Science & Business Media, 2010.
42. F. Sciacchitano, Y. Dong, T. Zeng, Variational Approach for Restoring Blurred Images with Cauchy Noise, SIAM Journal on Imaging Sciences, 8(3), 1894–1922, 2015.
43. Y. Wang, Y. Wotao, and J. Zeng, Global convergence of ADMM in nonconvex nonsmooth optimization, arXiv preprint arXiv:1511.06324 (2015).
44. B. Widrow, I. Kollár, Quantization Noise: Roundoff Error in Digital Computation, Signal Processing, Control, and Communications, 485–528, 2008.
45. Y.Q. Wang, J. M. Morel, SURE guided Gaussian mixture image denoising, SIAM Journal on Imaging Sciences, 6(2), 999-1034, 2013.
46. 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, vol. 3(1), pp. 1-122, 2011.
47. L. Yang, T. K. Pong, and X. Chen, Alternating direction method of multipliers for a class of nonconvex and nonsmooth problems with applications to background/foreground extraction, SIAM Journal on Imaging Sciences 10(1), 74-110, 2017.
48. C. Wu, J. Zhang, X.C. Tai, Augmented Lagrangian Method for Total Variation Restoration with Non-Quadratic Fidelity, Inverse Problems and Imaging, 5(1), 237–261, 2011.
49. http://www.caam.rice.edu/optimization/L1/ftvd/v4.0/
50. S. Geman and D. Geman, Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 6, pp. 721–741, 1984.
51. Y. Romano, M. Elad, P. MilanFar, The little engine that could: regularization by denoising (RED), SIAM Journal on Imaging Sciences, 10(4), 1808–1844, 2017.
52. Hansen, P. C., Kilmer, M. E., and Kjeldsen, R. H. (2006). Exploiting residual information in the parameter choice for discrete ill-posed problems. BIT Numerical Mathematics, 46(1), 41-59.
53. Rust, B. W., and O'Leary, D. P. (2008). Residual periodograms for choosing regularization parameters for ill-posed problems. Inverse Problems, 24(3), 034005.

APPENDIX

Proof of Lemma 1.

Proof First, we notice that the whiteness set $\mathcal{W}_\alpha \subset \mathbb{R}^{d^2}$ defined in (17)–(18) is given by the intersection of $\left(d^2 - 1\right)$ different sets, namely

$$\mathcal{W}_\alpha = \bigcap_{[l,m]\in\overline{\Theta}_0} \mathcal{W}_\alpha^{(l,m)}, \tag{58}$$

where $\overline{\Theta}_0 \subset \mathbb{Z}^2$ is defined in (8) and

$$\mathcal{W}_\alpha^{(l,m)} = \left\{ u \in \mathbb{R}^{d^2}: \ -w_\alpha d^2 \le (Ku - g)^T \ \overline{S}^{(l,m)} \ (Ku - g) \ \le \ +w_\alpha d^2 \right\} \tag{59}$$

Any matrix $\overline{S}^{(l,m)} \in \mathbb{R}^{d^2 \times d^2}$ in (59) is given by the symmetric part of the matrix $S^{(l,m)} \in \mathbb{R}^{d^2 \times d^2}$ representing the 2D circular (l,m)-shift operator, that is the operator which circularly shifts the elements of a $d \times d$ matrix of l rows and m columns. It is easy to verify that matrices $\overline{S}^{(l,m)}$ are all indefinite, hence the associated sets $\mathcal{W}_\alpha^{(l,m)}$ in (59) are closed and can be non-convex. It clearly follows that the whiteness set $\mathcal{W}_\alpha$ in (58) is closed (since intersection of closed sets) and can be non-convex.

According to Definition 5, in order to prove that $\mathcal{W}_\alpha$ is unbounded it is sufficient to demonstrate that there exists an unbounded sequence $\{u^{(k)}\} \subset \mathcal{W}_\alpha$. After recalling that the blur matrix K can be strongly ill-conditioned or even numerically singular but in purely mathematical sense it is invertible, we consider the following sequences:

$$\left\{u^{(k)}\right\} := \left\{K^{-1}\left(g + s^{(k)}\right)\right\}, \quad \left\{s^{(k)}\right\} := \left\{\nu^{(k)} e_m\right\}, \quad \left\{\nu^{(k)}\right\} \subset \mathbb{R} \text{ unbounded.} \tag{60}$$

Such sequences are unbounded, in fact

$$
\begin{aligned}
\lim_{k \to +\infty} \left\|u^{(k)}\right\|_2 &= \lim_{k \to +\infty} \left\|K^{-1}\left(g + s^{(k)}\right)\right\|_2 = \lim_{k \to +\infty} \left\|K^{-1}g + \nu^{(k)} K^{-1} e_m\right\}\right\|_2 \\
&= \lim_{k \to +\infty} \left\|\nu^{(k)}\left(\frac{1}{\nu^{(k)}} K^{-1}g + K^{-1} e_m\right)\right\|_2 \\
&= \lim_{k \to +\infty} \left(\left|\nu^{(k)}\right| \left\|K^{-1} e_m\right\|_2\right) = \left\|K^{-1} e_m\right\|_2 \lim_{k \to +\infty} \left|\nu^{(k)}\right| = +\infty
\end{aligned}
\tag{61}
$$

To prove that the unbounded sequences in (60) belong to the whiteness set in (17)–(18), we derive the expression of $r_{Ku^{(k)} - g}$, that is the sample auto-correlation of the residue image associated to the generic term of the sequences:

$$
\begin{aligned}
r_{Ku^{(k)} - g} &= \frac{1}{d^2}\left(Ku^{(k)} - g\right) \star \left(Ku^{(k)} - g\right) = \frac{1}{d^2} s^{(k)} \star s^{(k)} \\
&= \frac{1}{d^2}\left(\nu^{(k)} e_m\right) \star \left(\nu^{(k)} e_m\right) = \frac{\left(\nu^{(k)}\right)^2}{d^2} e_m \star e_m \\
&= \frac{\left(\nu^{(k)}\right)^2}{d^2} \operatorname{mat}(e_1),
\end{aligned}
$$

that is

$$
r_{Ku^{(k)} - g}[l,m] = \begin{cases} \left(\nu^{(k)}\right)^2 / d^2 & \text{for } [l,m] = [0,0] \\ 0 & \text{for } [l,m] \in \overline{\Theta}_0 \end{cases} \quad \forall\, k \in \mathbb{N}. \tag{62}
$$

It follows from (61), (62) and (17)–(18) that any sequence $\{u^{(k)}\}$ defined as in (60) is unbounded and belongs to the whiteness set $\mathcal{W}_\alpha$ for any real $\alpha \geq 0$. This implies that $\mathcal{W}_\alpha$ is unbounded.

Finally, in order to demonstrate that $\mathcal{W}_\alpha$ in (17)–(18) is non-convex, it is sufficient to prove that, for any given real $\alpha \geq 0$, there always exist two images

$u_\alpha, v_\alpha \in \mathbb{R}^{d^2}$ belonging to $\mathcal{W}_\alpha$ and a scalar $\gamma \in (0, 1)$ such that the image $z_\alpha := \gamma\, u_\alpha + (1 - \gamma)\, v_\alpha$ do not belong to $\mathcal{W}_\alpha$. By taking

$$
\begin{cases}
u_\alpha := K^{-1}\big(g + 2\, d\, \sqrt{\rho_\alpha}\, e_p \big) \\
v_\alpha := K^{-1}\big(g + 2\, d\, \sqrt{\rho_\alpha}\, e_q \big)
\end{cases}
\quad \text{with} \quad p, q \in \{1, \dots, d^2\},\ p \neq q, \quad \rho_\alpha > w_\alpha,
\tag{63}
$$

and, according to the choice $\gamma = 1/2$,

$$
z_\alpha := \frac{1}{2}\, u_\alpha + \frac{1}{2}\, v_\alpha = K^{-1}\left(g + d\, \sqrt{\rho_\alpha}\, (e_p + e_q) \right),
\tag{64}
$$

we have

$$
r_{Ku_\alpha - g}[l, m] = r_{Kv_\alpha - g}[l, m] = 0 \quad \forall\, (l, m) \in \overline{\Theta}_0
\tag{65}
$$

and

$$
\begin{aligned}
r_{Kz_\alpha - g}[l, m] &= \frac{1}{d^2}\left((Kz_\alpha - g) \star (Kz_\alpha - g) \right)[l, m] \\
&= \frac{1}{d^2}\left((d\, \sqrt{\rho_\alpha}\, (e_p + e_q)) \star (d\, \sqrt{\rho_\alpha}\, (e_p + e_q)) \right)[l, m] \\
&= \rho_\alpha\, \left((e_p + e_q) \star (e_p + e_q) \right)[l, m] \\
&= \rho_\alpha\, \left(e_p \star e_p + e_q \star e_q + e_p \star e_q + e_q \star e_p \right)[l, m] \\
&= \rho_\alpha\, \left(\mathrm{mat}(e_1) + \mathrm{mat}(e_1) + e_p \star e_q + e_q \star e_p \right)[l, m] \\
&= \begin{cases}
2\, \rho_\alpha & \text{for } [l, m] = [0, 0] \\
\rho_\alpha & \text{for } [l, m] \in \{[\bar{l}, \bar{m}], [-\bar{l}, -\bar{m}]\} \text{ for some } [\bar{l}, \bar{m}] \neq [0, 0] \\
0 & \text{otherwise}
\end{cases}
\end{aligned}
\tag{66}
$$

Since $\rho_\alpha > w_\alpha$ by assumption, then z_α does not belong to $\mathcal{W}_\alpha$, and the proof is completed.

Proof of Lemma 2.

Proof The fact that the discrete TV semi-norm function is proper, continuous, convex and bounded from below by zero is well known and immediate to verify. It is also well known that the TV function is not coercive over its entire domain $\mathbb{R}^{d^2}$. To prove that the TV function is coercive over the unbounded whiteness set $\mathcal{W}_\alpha \subset \mathbb{R}^{d^2}$ defined in (17)–(18), first we outline the set of all the unbounded sequences $\{u^{(k)}\} \subset \mathbb{R}^{d^2}$ for which the TV function is not coercive, that is for which $\lim\limits_{k \to \infty} \mathrm{TV}\big(u^{(k)}\big) < +\infty$, then we demonstrate that such sequences are not contained into $\mathcal{W}_\alpha$.

Let us define the matrix $D := (D_h^T, D_v^T)^T \in \mathbb{R}^{2d^2 \times d^2}$ with $D_h, D_v \in \mathbb{R}^{d^2}$ the coefficient matrices of linear finite difference operators approximating the horizontal and vertical partial derivatives of image u, respectively. Then, the TV semi-norm of u defined in (2) can be regarded as the composition of a

linear map $\mathcal{D} : \mathbb{R}^{d^2} \to \mathbb{R}^{2d^2}$ with coefficient matrix D and a suitable nonlinear function $\mathcal{G} : \mathbb{R}^{2d^2} \to \mathbb{R}$, that is

$$\mathrm{TV}(u) = \mathcal{G}(\mathcal{D}(u)), \qquad \mathcal{D}(u) := Du, \qquad \mathcal{G}(v) := \sum_{i=1}^{d^2} \|(v_i, v_{i+d^2})\|_2 . \quad (67)$$

We now study coerciveness of the function $\mathcal{G}$ above. As it is immediate to verify that

$$\mathcal{G}(v) = \sum_{i=1}^{d^2} \|(v_i, v_{i+d^2})\|_2 \geq \big\| (v_1, v_2, \ldots, v_{2d^2}) \big\|_2 = \|v\|_2 , \qquad (68)$$

for any unbounded sequence $\{v^{(k)}\} \subseteq \mathbb{R}^{2d^2}$ - that is, according to Definition 4, any sequence satisfying $\|v^{(k)}\|_2 \xrightarrow{k \to \infty} +\infty$ - we have that

$$\lim_{\|v^{(k)}\|_2 \xrightarrow{k \to \infty} +\infty} \mathcal{G}\big(v^{(k)}\big) \geq \lim_{\|v^{(k)}\|_2 \xrightarrow{k \to \infty} +\infty} \big\|v^{(k)}\big\|_2 = +\infty , \qquad (69)$$

that is the function $\mathcal{G}$ is coercive over its entire domain $\mathbb{R}^{2d^2}$.

For what concerns the linear operator $\mathcal{D}$, clearly the kernel of the coefficient matrix D has dimension 1 and is given by

$$\ker(D) = \left\{ u \in \mathbb{R}^{d^2} : \ u = \nu \mathbb{1}, \ \nu \in \mathbb{R} \right\} . \qquad (70)$$

Since for any vector $u \in \mathbb{R}^{d^2}$ there always exists only one pair of vectors $u_1 \in \ker(D)$, $u_2 \in (\ker(D))^\perp$ such that $u = u_1 + u_2$, then any unbounded sequence $\{u^{(k)}\} \subseteq \mathbb{R}^{d^2}$ can be additively split as follows

$$\{u^{(k)}\} = \{u_1^{(k)}\} + \{u_2^{(k)}\}, \quad \{u_1^{(k)}\} \subseteq \ker(D), \ \{u_2^{(k)}\} \subseteq (\ker(D))^\perp , \quad (71)$$

where either $\{u_1^{(k)}\}$ or $\{u_2^{(k)}\}$ is unbounded. In case that $\{u_2^{(k)}\}$ is unbounded, then clearly $\{Du_2^{(k)}\}$ is also unbounded and, due to coerciveness of the function $\mathcal{G}$, $\mathrm{TV}\big(u_2^{(k)}\big) = \mathcal{G}\big(Du_2^{(k)}\big)$ tends to ∞. On the other hand, any unbounded $\{u_1^{(k)}\}$ is mapped by D into a null (bounded) sequence. It follows that all the unbounded sequences for which the TV function is not coercive are of the form

$$\{u^{(k)}\} = \{u_2^{(k)} + \nu^{(k)} \mathbb{1}\}, \qquad (72)$$

with $\{u_2^{(k)}\} \subset \mathbb{R}^{d^2}$ any bounded sequence and $\{\nu^{(k)}\} \subset \mathbb{R}$ any (scalar) unbounded sequence.

We now prove that no unbounded sequence of the form (72) belongs to the whiteness set $\mathcal{W}_\alpha$ in (17)–(18), for any real $\alpha \geq 0$. According to definition (18) of the residual sample auto-correlation, we have:

$$
\begin{aligned}
r_{Ku^{(k)}-g} &= \frac{1}{d^2}\left(Ku^{(k)}-g\right)\star\left(Ku^{(k)}-g\right)\\
&= \frac{1}{d^2}\left(K\left(u_1^{(k)}+\nu^{(k)}\mathbb{1}\right)-g\right)\star\left(K\left(u_1^{(k)}+\nu^{(k)}\mathbb{1}\right)-g\right)\\
&= \frac{1}{d^2}\left(\nu^{(k)}K\mathbb{1}+Ku_1^{(k)}-g\right)\star\left(\nu^{(k)}K\mathbb{1}+Ku_1^{(k)}-g\right). \quad (73)
\end{aligned}
$$

Since the blur PSF is typically energy-preserving, which implies that the sum of the elements of each row of the blur matrix K is equal to one, then $K\mathbb{1}=\mathbb{1}$. Moreover, since K is bounded, the sequence $c^{(k)}:=Ku_1^{(k)}-g$ is bounded. From (73) we have

$$
\begin{aligned}
r_{Ku^{(k)}-g}[l,m] &= \frac{1}{d^2}\left(\left(\nu^{(k)}\mathbb{1}+c^{(k)}\right)\star\left(\nu^{(k)}\mathbb{1}+c^{(k)}\right)\right)[l,m]\\
&= \frac{1}{d^2}\left(\left(\nu^{(k)}\right)^2\mathbb{1}\star\mathbb{1}+\nu^{(k)}\,\mathbb{1}\star c^{(k)}+\nu^{(k)}\,c^{(k)}\star\mathbb{1}+c^{(k)}\star c^{(k)}\right)[l,m]\\
&= \frac{1}{d^2}\left(\left(\nu^{(k)}\right)^2 d^2\,\mathrm{mat}(\mathbb{1})+2\,\nu^{(k)}\left(\sum_{i=1}^{d^2}c_i^{(k)}\right)\mathrm{mat}(\mathbb{1})+c^{(k)}\star c^{(k)}\right)[l,m]\\
&= \frac{1}{d^2}\left(\left(\nu^{(k)}\right)^2 d^2+2\,\nu^{(k)}\left(\sum_{i=1}^{d^2}c_i^{(k)}\right)+\left(c^{(k)}\star c^{(k)}\right)[l,m]\right)\\
&= \left(\nu^{(k)}\right)^2\left(1+\frac{2}{d^2\nu^{(k)}}\left(\sum_{i=1}^{d^2}c_i^{(k)}\right)+\frac{\left(c^{(k)}\star c^{(k)}\right)[l,m]}{d^2\left(\nu^{(k)}\right)^2}\right). \quad (74)
\end{aligned}
$$

As the sequence $c^{(k)}\subset\mathbb{R}^{d^2}$ is bounded and the sequence $\nu^{(k)}\subset\mathbb{R}$ is unbounded, the second and third terms within parenthesis in (74) both represent bounded sequences in $\mathbb{R}$ (more precisely, they both tend to zero as k approaches $+\infty$). Hence, we have that

$$
\lim_{k\to+\infty}r_{Ku^{(k)}-g}[l,m]=\lim_{k\to+\infty}\left(\nu^{(k)}\right)^2=+\infty\quad\forall\,[l,m]\in\overline{\Theta}_0. \quad (75)
$$

It thus follows from (75) that unbounded sequences of the form (72) do not belong to the whiteness set $\mathcal{W}_\alpha$ in (17) for any real $\alpha\geq 0$, at least for k greater than a certain value. This implies coercivity of the TV function over $\mathcal{W}_\alpha$ and concludes the proof.