



ADMM-Based Residual Whiteness Principle for Automatic Parameter Selection in Single Image Super-Resolution Problems

Monica Pragliola¹ · Luca Calatroni² · Alessandro Lanza³ · Fiorella Sgallari³

Received: 3 August 2021 / Accepted: 23 May 2022 / Published online: 5 July 2022
© The Author(s) 2022

Abstract

We propose an automatic parameter selection strategy for the single image super-resolution problem for images corrupted by blur and additive white Gaussian noise with unknown standard deviation. The proposed approach exploits the structure of both the down-sampling and the blur operators in the frequency domain and computes the optimal regularisation parameter as the one optimising a suitably defined residual whiteness measure. Computationally, the proposed strategy relies on the fast solution of generalised Tikhonov ℓ_2 – ℓ_2 problems as proposed in Zhao et al. (IEEE Trans Image Process 25:3683–3697, 2016). These problems naturally appear as substeps of the Alternating Direction Method of Multipliers used to solve single image super-resolution problems with non-quadratic, non-smooth, sparsity-promoting regularisers both in convex and in non-convex regimes. After detailing the theoretical properties allowing to express the whiteness functional in a compact way, we report an exhaustive list of numerical experiments proving the effectiveness of the proposed approach for different type of problems, in comparison with well-known parameter selection strategies such as, e.g., the discrepancy principle.

Keywords Single image super-resolution · Residual whiteness principle · Total variation regularisation · Sparse regularisation · ADMM

1 Introduction

The problem of single image super-resolution (in short, SR) consists in finding a high-resolution (HR) image starting from a single low-resolution (LR), blurred and noisy image measurement. Several applications benefit from the recovery of HR information from LR ones: their non-exhaustive list range from remote sensing to medical and microscopy

imaging, where, typically, the SR problem aims to reconstruct high-quality and fine-detailed images overcoming the physical limitations imposed by the acquisition setting, such as, e.g., light diffraction phenomena [14,40].

The SR problem can be formulated in mathematical terms as follows. Let $\mathbf{X} \in \mathbb{R}^{N_r \times N_c}$ denote the original HR image, with $\mathbf{x} = \text{vec}(\mathbf{X}) \in \mathbb{R}^N$, $N = N_r N_c$, being its row-wise vectorisation. The degradation process describing the mapping from HR to LR data can be described by the linear model

$$\mathbf{b} = \mathbf{S}\mathbf{K}\mathbf{x} + \mathbf{e}, \quad \text{with } \mathbf{e} \text{ realisation of } \mathbf{E} \sim \mathcal{N}(\mathbf{0}_n, \sigma^2 \mathbf{I}_n), \quad (1)$$

where $\mathbf{b}, \mathbf{e} \in \mathbb{R}^n$ are the vectorised LR observed image and additive white Gaussian noise (AWGN) realisation, respectively, both of size $n_r \times n_c$, with $n = n_r n_c$ and $\mathbf{S} \in \mathbb{R}^{n \times N}$ is a binary selection matrix inducing a pixel decimation with factors d_r and d_c along the N_r rows and the N_c columns of the HR image \mathbf{X} , respectively—i.e. $N_r = n_r d_r$, $N_c = n_c d_c$. Then, $\mathbf{K} \in \mathbb{R}^{N \times N}$ is the matrix representing a space-invariant blurring operator and \mathbf{E} is an n -variate Gaussian-distributed random vector with zero mean and scalar covariance matrix, with $\sigma > 0$ indicating the (often unknown) noise standard deviation. In the following, we will use the notation

✉ Monica Pragliola
monica.pragliola@unina.it

Luca Calatroni
calatroni@i3s.unice.fr

Alessandro Lanza
alessandro.lanza2@unibo.it

Fiorella Sgallari
fiorella.sgallari@unibo.it

¹ Department of Mathematics and Applications, University of Naples Federico II, Naples, Italy

² CNRS, UCA, INRIA, Morpheme, I3S, Sophia-Antipolis, France

³ Department of Mathematics, University of Bologna, Bologna, Italy

$\mathbf{I}_n \in \mathbb{R}^{n \times n}$ and $\mathbf{0}_n, \mathbf{1}_n \in \mathbb{R}^n$ to denote the $n \times n$ identity matrix and the n -dimensional vectors with all zeros and ones, respectively and set $d := d_r d_c$, so that $N = nd$.

Determining $\mathbf{x} \in \mathbb{R}^N$ solving (1) given $\mathbf{b} \in \mathbb{R}^n$ and the operators \mathbf{S}, \mathbf{K} is an ill-posed inverse problem. Many approaches for solving it have been proposed, ranging from interpolation techniques, see, e.g., [35], sparse representation approaches [41], reconstruction-based methods [5,10,25,38,43] and, recently, approaches based on learning techniques, see the recent review [42].

In this paper, we consider a reconstruction-based approach based on suitable regularisation approaches. In particular, we seek an estimate \mathbf{x}^* of \mathbf{x} being the minimiser of a suitable cost function $\mathcal{J} : \mathbb{R}^N \rightarrow \mathbb{R}_+$, with \mathbb{R}_+ denoting the set of non-negative real numbers, which codifies both *a-priori* information on the solution and on the Gaussian noise statistics. A standard approach for doing so consists in considering the problem:

$$\mathbf{x}^*(\mu) \in \arg \min_{\mathbf{x} \in \mathbb{R}^N} \left\{ \mathcal{J}(\mathbf{x}; \mu) := \frac{\mu}{2} \|\mathbf{S}\mathbf{K}\mathbf{x} - \mathbf{b}\|_2^2 + \mathcal{R}(\mathbf{x}) \right\}, \quad (2)$$

where $\mathcal{R} : \mathbb{R}^N \rightarrow \mathbb{R}_+$ is a possibly non-convex and non-smooth regularisation term, and where the data fidelity term $(1/2)\|\mathbf{S}\mathbf{K}\mathbf{x} - \mathbf{b}\|_2^2$ encodes the AWGN assumption on \mathbf{e} , while the regularisation parameter $\mu > 0$ balances the action of the fidelity against regularisation.

The optimal choice of μ in (2) is in fact crucial for obtaining high-quality reconstructions. Many heuristic approaches have been proposed for its automatic selection, such as, e.g., L-curve [6] and generalised cross-validation (GCV) [13]. Alternatively, several methods exploiting available information on the noise corruption have been designed. Among them, a standard strategy is the Morozov Discrepancy Principle (DP)—see [7,16,31] for general problems and [37] for applications to SR—which can be formulated as follows:

$$\begin{aligned} \text{Select } \mu = \mu^* \text{ such that } \|\mathbf{r}^*(\mu^*)\|_2 &= \|\mathbf{S}\mathbf{K}\mathbf{x}^*(\mu^*) - \mathbf{b}\|_2 \\ &= \tau \sqrt{n} \sigma, \end{aligned} \quad (3)$$

where $\mathbf{x}^*(\mu) \in \mathbb{R}^N$ solves (2), and $\mathbf{r}^*(\mu) := \mathbf{S}\mathbf{K}\mathbf{x}^*(\mu) - \mathbf{b} \in \mathbb{R}^n$ is defined as the LR residual with $\tau > 0$ denoting the discrepancy coefficient. When σ is known, τ is set equal to 1, otherwise a value slightly greater than 1 is typically chosen so to avoid noise under-estimation. In many real world applications an accurate estimate of σ is not available, which often limits the applicability of DP-based strategies.

Alternative strategies overcoming this issue and leveraging the noise whiteness property have been used in the context of image denoising and deconvolution problems (i.e. where $\mathbf{S} = \mathbf{I}_N$). Using the whiteness property of the residual image—i.e. of the noise—is an effective idea as on the one

hand it does not require to know the noise standard deviation and on the other it exploits much more information than the one encoded in the first moments of the noise distribution. Variational models for image denoising and deblurring containing data fidelity terms or regularisers aimed to explicitly enforce the residual whiteness property have been proposed; see, e.g., [2,18,20,21,27]. The good results achieved by such models demonstrate the potentiality of using the noise whiteness property. However, these models are strongly non-convex, with all the associated numerical and computational difficulties. More classically, whiteness and other statistical and deterministic properties of the residual have been used as *a-posteriori* criteria for evaluating the quality of the restored image or selecting the regularisation parameter of convex variational models in image denoising and deblurring; see, e.g., [1,4,17,29]. In particular, in [1] the authors proposed an effective parameter selection strategy for variational image deconvolution based on minimising the residual normalised auto-correlation. This approach, that has been applied in [1] as an *a posteriori* criterion, has then been revisited in [22], where the authors design a measure of whiteness of the residual image $\mathbf{r}^*(\mu) = \mathbf{K}\mathbf{x}^*(\mu) - \mathbf{b}$ that is regarded as a function of μ . This strategy, called the *residual whiteness principle* (RWP), allows for an automatic estimation of μ and can be naturally embedded within an iterative ADMM optimisation scheme and shown to be effective for different choices of non-quadratic non-smooth regularisers \mathcal{R} . In fact, whenever $\mathbf{S} = \mathbf{I}_N$ and upon the assumption of periodic boundary conditions, models of the form (2) can be easily manipulated through an ADMM-type scheme where matrix-vector products and matrix inversions can be efficiently computed in the frequency domain by means of fast discrete Fourier transform solvers, due to the circulant structure of the operator \mathbf{K} (see [22] for details).

In SR problems, however, due to the presence of the decimation operator \mathbf{S} , the operator $\mathbf{A} := \mathbf{S}\mathbf{K}$ is, typically, unstructured and, as a consequence, the solution of (2) becomes more challenging. To motivate this better, let us consider the special choice $\mathcal{R}(\mathbf{x}) := \frac{1}{2}\|\mathbf{L}\mathbf{x} - \mathbf{v}\|_2^2$, with $\mathbf{L} \in \mathbb{R}^{M \times N}$ being a known regularisation matrix and $\mathbf{v} \in \mathbb{R}^M$ a given vector. Then, problem (2) takes the form of the generalised Tikhonov-regularised ℓ_2 - ℓ_2 problem which reads

$$\mathbf{x}^*(\mu) = \arg \min_{\mathbf{x} \in \mathbb{R}^N} \left\{ \frac{\mu}{2} \|\mathbf{S}\mathbf{K}\mathbf{x} - \mathbf{b}\|_2^2 + \frac{1}{2} \|\mathbf{L}\mathbf{x} - \mathbf{v}\|_2^2 \right\}, \quad (4)$$

which has been previously employed, e.g., in [34,41] in the context of SR problems. The solution of (4) can be computed by considering the corresponding optimality condition upon the inversion of unstructured operators, thus requiring in principle the use of iterative solvers, such as, e.g., the Conjugate Gradient algorithm.

However, for problems like (4) and upon a specific choice of \mathbf{S} , it was shown in [43] that an efficient solution strategy based on the use of the Woodbury’s formula can be derived. The resulting algorithm, therein called Fast Super-Resolution (FSR) algorithm significantly reduces the computational efforts required by iterative solvers solving (4) as it boils down the problem to the inversion of diagonal matrices in the Fourier domain. As far as the parameter selection strategy is concerned, in [43] a Generalised Cross Validation strategy [11] is used to select the optimal μ , which is known to be impractical for large-scale problems, see, e.g., [8]. Note, furthermore, that ℓ_2 - ℓ_2 problems in the form (4) naturally arise when attempting to solve the more general class of variational models (2) by means of classical iterative optimisation solvers such as the ADMM. In this context, such problems appear to enforce suitable variable splitting and are defined in terms of appropriate penalty parameters. Here, the FSR algorithm can thus still be used as an efficient solver, similarly as done, e.g., in [5,24,33,33].

Contribution We propose an automatic parameter selection strategy for the regularisation parameter μ in (2) which does not require any prior knowledge on the AWG noise level and can be applied to general non-smooth and possibly non-convex regularisers \mathcal{R} . Our approach is based on the optimisation of a suitably defined measure of whiteness of the residual image in the frequency domain. The proposed approach generalises the results presented in [22] for image deconvolution problems to the more challenging scenario of single-image super-resolution problems. At the same time, it extends the results contained in an earlier conference version of this work [26] where only problems in the form (4) were considered. By designing an ADMM-based optimisation strategy for solving the general problem (2) with an appropriate variable splitting, the residual whiteness principle can be applied iteratively along the ADMM iterations and used jointly as part of the solution of the ℓ_2 - ℓ_2 substeps in the form (4). Several numerical results confirming the effectiveness of the proposed method in comparison to the standard Discrepancy Principle for popular image regularisers such as the Tikhonov and Total Variation are reported. Moreover, to provide some insights about the extension of such strategy to non-convex setting, we propose to embed the automatic estimation strategy yielded by the adoption of the RWP within an iterative reweighted ℓ_1 scheme for tackling the non-convex continuous relaxation of the ℓ_0 norm [32].

2 Notations, Preliminaries and Assumptions

We start by setting the notations and recalling useful results that will be used in the following discussion. Then, we detail the adopted assumptions for our derivations.

2.1 Notations and Preliminaries

In the following, for $z \in \mathbb{C}$ we use $\bar{z}, |z|$ to indicate the conjugate and the modulus of z , respectively. We denote by $\mathbf{F}, \mathbf{F}^H \in \mathbb{C}^{N \times N}$ the unitary coefficient matrices of the 2D discrete Fourier transform operator and its inverse, respectively, when applied to vectorised $N_r \times N_c$ images, with $\mathbf{F}\mathbf{F}^H = \mathbf{F}^H\mathbf{F} = \mathbf{I}_N$. For any $\mathbf{v} \in \mathbb{R}^N$ and any $\mathbf{A} \in \mathbb{R}^{N \times N}$, we use the notations $\tilde{\mathbf{v}} = \mathbf{F}\mathbf{v}$ and $\tilde{\mathbf{A}} = \mathbf{F}\mathbf{A}\mathbf{F}^H$ to denote the action of the 2D Fourier transform operator \mathbf{F} on vectors and matrices, respectively. Given a permutation matrix $\mathbf{P} \in \mathbb{R}^{N \times N}$, we denote by $\hat{\mathbf{v}} = \mathbf{P}\tilde{\mathbf{v}}$ and by $\hat{\mathbf{A}} = \mathbf{P}\tilde{\mathbf{A}}\mathbf{P}^T$ the action of \mathbf{P} on the Fourier-transformed vector $\tilde{\mathbf{v}}$ and matrix $\tilde{\mathbf{A}}$, respectively. We indicate by $\check{\mathbf{A}}$ the product $\check{\mathbf{A}} = \mathbf{P}\tilde{\mathbf{A}}^H\mathbf{P}^T$, i.e. the action of \mathbf{P} on $\tilde{\mathbf{A}}^H$. Finally, we denote by \mathbf{J}_s the $s \times s$ matrix of all ones and by \mathbf{c}_i^s the i -th canonical basis vector of \mathbb{R}^s .

In the following Lemmas 1–4 we recall two well-known properties of the Kronecker product ‘ \otimes ’—see, e.g., [12]—as well as two results that will be useful in the paper.

Lemma 1 Let $\mathbf{A}_1, \dots, \mathbf{A}_m, \mathbf{B}_1, \dots, \mathbf{B}_m, m \geq 2$, be matrices such that all the products $\mathbf{A}_i\mathbf{B}_i$ exist, $i = 1, \dots, m$. Then, there holds:

$$\begin{aligned} &(\mathbf{A}_1 \otimes \dots \otimes \mathbf{A}_m)(\mathbf{B}_1 \otimes \dots \otimes \mathbf{B}_m) \\ &= (\mathbf{A}_1\mathbf{B}_1 \otimes \dots \otimes \mathbf{A}_m\mathbf{B}_m). \end{aligned} \tag{5}$$

Lemma 2 Let \mathbf{A}, \mathbf{B} be square matrices of orders r and s , respectively. Then:

$$\mathbf{B} \otimes \mathbf{A} = \mathbf{P}_s^r (\mathbf{A} \otimes \mathbf{B}) (\mathbf{P}_s^r)^T,$$

where $\mathbf{P}_s^r \in \mathbb{R}^{rs \times rs}$ is a special permutation matrix, called the perfect shuffle matrix for a set of $s \times r$ elements, defined by

$$\mathbf{P}_s^r = \sum_{i=1}^s \left(\mathbf{c}_i^s \otimes \mathbf{I}_r \otimes (\mathbf{c}_i^s)^T \right). \tag{6}$$

Lemma 3 (Woodbury formula) Let $\mathbf{A}_1, \mathbf{A}_2, \mathbf{A}_3, \mathbf{A}_4$ be conformable matrices with \mathbf{A}_1 and \mathbf{A}_3 invertible. Then, the following inversion formula holds:

$$\begin{aligned} &(\mathbf{A}_1 + \mathbf{A}_2\mathbf{A}_3\mathbf{A}_4)^{-1} \\ &= \mathbf{A}_1^{-1} + \mathbf{A}_1^{-1}\mathbf{A}_2(\mathbf{A}_3^{-1} + \mathbf{A}_4\mathbf{A}_1^{-1}\mathbf{A}_2)^{-1}\mathbf{A}_4\mathbf{A}_1^{-1}. \end{aligned} \tag{7}$$

Lemma 4 [36] Let $\mathbf{S} \in \mathbb{R}^{n \times N}$ be the binary selection matrix introduced. Then:

$$\widetilde{\mathbf{S}^H\mathbf{S}} = \frac{1}{d} (\mathbf{J}_{d_r} \otimes \mathbf{I}_{n_r}) \otimes (\mathbf{J}_{d_c} \otimes \mathbf{I}_{n_c}). \tag{8}$$

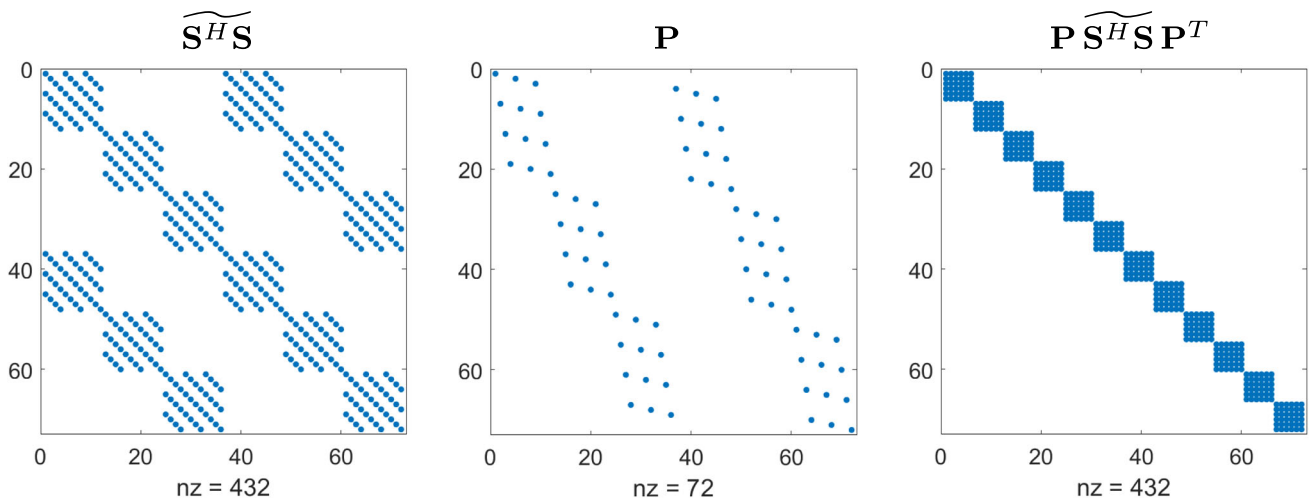


Fig. 1 Structure of matrix $\widetilde{\mathbf{S}^H \mathbf{S}}$ in (8), on the left, of permutation matrix \mathbf{P} in (9), in the centre, and of permuted matrix $\mathbf{P} \widetilde{\mathbf{S}^H \mathbf{S}} \mathbf{P}^T$ in (10), on the right, for $N_r = 6, N_c = 12, d_r = 2, d_c = 3 \implies n_r = 3, n_c = 4$. Blue dots indicate positions of non-zero entries and nz denotes the number of non-zero entries

The non-zero entries of the matrix $\widetilde{\mathbf{S}^H \mathbf{S}}$ in (8) are all equal to $1/d$ and are arranged along replicated patterns, as shown in Fig. 1(left) for the particular case $N_r = 6, N_c = 12, d_r = 2, d_c = 3 \implies n_r = 3, n_c = 4$. In the following Proposition 1 we prove that there always exists a permutation matrix \mathbf{P} , as the one represented in Fig. 1(centre), capable of shuffling the rows and columns of $\widetilde{\mathbf{S}^H \mathbf{S}}$ to get a well-structured, block-diagonal matrix, as shown in Fig. 1(right).

Proposition 1 Let $\mathbf{P} \in \mathbb{R}^{N \times N}$ be the permutation matrix defined by

$$\mathbf{P} := \mathbf{P}_n^d \left(\mathbf{I}_{d_r} \otimes \mathbf{P}_{d_c}^{n_r} \otimes \mathbf{I}_{n_c} \right), \tag{9}$$

with $\mathbf{P}_n^d, \mathbf{P}_{d_c}^{n_r}$ perfect shuffle matrices as defined in (6). Then, there holds

$$\mathbf{P} \widetilde{\mathbf{S}^H \mathbf{S}} \mathbf{P}^T = \frac{1}{d} (\mathbf{I}_n \otimes \mathbf{J}_d) = \frac{1}{d} \text{diag}(\overbrace{\mathbf{J}_d, \dots, \mathbf{J}_d}^{n \text{ blocks}}). \tag{10}$$

Proof First, by setting $\overline{\mathbf{P}} := \mathbf{I}_{d_r} \otimes \mathbf{P}_{d_c}^{n_r} \otimes \mathbf{I}_{n_c}$ and based on the definition (8), we have

$$\overline{\mathbf{P}} \widetilde{\mathbf{S}^H \mathbf{S}} = \frac{1}{d} \left(\mathbf{I}_{d_r} \otimes \mathbf{P}_{d_c}^{n_r} \otimes \mathbf{I}_{n_c} \right) \left(\mathbf{J}_{d_r} \otimes (\mathbf{I}_{n_r} \otimes \mathbf{J}_{d_c}) \otimes \mathbf{I}_{n_c} \right) \tag{11}$$

$$= \frac{1}{d} \left(\mathbf{J}_{d_r} \otimes \left(\mathbf{P}_{d_c}^{n_r} (\mathbf{I}_{n_r} \otimes \mathbf{J}_{d_c}) \right) \otimes \mathbf{I}_{n_c} \right), \tag{12}$$

where in (11) we used the associative property of Kronecker products and in (12) we applied Lemma 1. Then, starting

from (12), we have

$$\begin{aligned} \overline{\mathbf{P}} \widetilde{\mathbf{S}^H \mathbf{S}} \overline{\mathbf{P}}^T &= \frac{1}{d} \left(\mathbf{J}_{d_r} \otimes \left(\mathbf{P}_{d_c}^{n_r} (\mathbf{I}_{n_r} \otimes \mathbf{J}_{d_c}) \right) \otimes \mathbf{I}_{n_c} \right) \\ &\quad \left(\mathbf{I}_{d_r} \otimes \left(\mathbf{P}_{d_c}^{n_r} \right)^T \otimes \mathbf{I}_{n_c} \right) \end{aligned} \tag{13}$$

$$= \frac{1}{d} \left(\mathbf{J}_{d_r} \otimes \left(\mathbf{P}_{d_c}^{n_r} (\mathbf{I}_{n_r} \otimes \mathbf{J}_{d_c}) \left(\mathbf{P}_{d_c}^{n_r} \right)^T \right) \otimes \mathbf{I}_{n_c} \right) \tag{14}$$

$$= \frac{1}{d} \left(\mathbf{J}_{d_r} \otimes \mathbf{J}_{d_c} \otimes \mathbf{I}_{n_r} \otimes \mathbf{I}_{n_c} \right) = \frac{1}{d} (\mathbf{J}_d \otimes \mathbf{I}_n), \tag{15}$$

where in (13) we used the property of the transposed of a Kronecker product, in (14) we applied Lemma 1 and in (15) Lemma 2. Finally, by recalling the definition of matrix \mathbf{P} in (9), namely $\mathbf{P} = \mathbf{P}_n^d \overline{\mathbf{P}}$, we have

$$\begin{aligned} \mathbf{P} \widetilde{\mathbf{S}^H \mathbf{S}} \mathbf{P}^T &= \mathbf{P}_n^d \overline{\mathbf{P}} \widetilde{\mathbf{S}^H \mathbf{S}} \overline{\mathbf{P}}^T \left(\mathbf{P}_n^d \right)^T = \frac{1}{d} \mathbf{P}_n^d (\mathbf{J}_d \otimes \mathbf{I}_n) \left(\mathbf{P}_n^d \right)^T \\ &= \frac{1}{d} (\mathbf{I}_n \otimes \mathbf{J}_d), \end{aligned} \tag{16}$$

where in (16) we again applied Lemma 2. □

2.2 Assumptions

In this section, we detail the class of variational models of interest by listing the assumptions adopted for the regularisation term $\mathcal{R}(\mathbf{x})$ as well as for the decimation matrix \mathbf{S} and the blurring matrix \mathbf{K} .

In this paper, we focus on the automatic selection of the regularisation parameter μ in single image super-resolution

variational models of the form:

$$\mathbf{x}^*(\mu) = \arg \min_{\mathbf{x} \in \mathbb{R}^N} \left\{ \mathcal{J}(\mathbf{x}; \mu) = \frac{\mu}{2} \|\mathbf{SKx} - \mathbf{b}\|_2^2 + \mathcal{R}(\mathbf{x}) \right\},$$

$$\mathcal{R}(\mathbf{x}) := G(\mathbf{Lx}). \tag{17}$$

We refer to $\mathbf{L} \in \mathbb{R}^{M \times N}$ as the regularisation matrix, whereas the regularisation function $G : \mathbb{R}^M \rightarrow \overline{\mathbb{R}} := \mathbb{R} \cup \{+\infty\}$ is nonlinear and possibly non-smooth.

The general variational model (17) undergoes the following assumptions:

- (A1) The blur matrix $\mathbf{K} \in \mathbb{R}^{N \times N}$, decimation matrix $\mathbf{S} \in \mathbb{R}^{n \times N}$ and regularisation matrix $\mathbf{L} \in \mathbb{R}^{M \times N}$ are such that $\text{null}(\mathbf{SK}) \cap \text{null}(\mathbf{L}) = \mathbf{0}_N$.
- (A2) The regularisation function $G : \mathbb{R}^M \rightarrow \overline{\mathbb{R}}$ is proper, lower semi-continuous, convex and coercive.
- (A3) The blur matrix \mathbf{K} represents a 2D discrete convolution operator—which follows from the blur being space-invariant—and the regularisation matrix \mathbf{L} is of the form:

$$\mathbf{L} = \left(\mathbf{L}_1^T, \dots, \mathbf{L}_s^T \right)^T \in \mathbb{R}^{sN \times N}, \quad s \in \mathbb{N} \setminus \{0\},$$

with $\mathbf{L}_j \in \mathbb{R}^{N \times N}, \quad j = 1, \dots, s,$

matrices also representing 2D discrete convolution operators.

- (A4) The decimation matrix $\mathbf{S} \in \mathbb{R}^{n \times N}$ is a binary selection matrix, such that $\mathbf{SS}^H = \mathbf{I}_n$ and the operator $\mathbf{S}^H \in \mathbb{R}^{N \times n}$ interpolates the decimated image with zeros.
- (A5) The regularisation function G is *easily proximable*, that is, the proximity operator of G at any $\mathbf{t} \in \mathbb{R}^M$,

$$\text{prox}_G(\mathbf{t}) = \arg \min_{\mathbf{z} \in \mathbb{R}^M} \left\{ G(\mathbf{z}) + \frac{1}{2} \|\mathbf{t} - \mathbf{z}\|_2^2 \right\},$$

can be efficiently computed.

Assumptions (A1)–(A2) guarantee the existence—and, upon suitable assumptions, uniqueness—of solutions of the considered class of variational super-resolution models (17), as formally stated in Proposition 2 below, whose proof can be easily derived from the one of Proposition 2.1 in [22].

Proposition 2 *If the assumptions (A1)–(A2) above are fulfilled, for any fixed $\mu \in \mathbb{R}_{++}$ the function $\mathcal{J}(\cdot; \mu) : \mathbb{R}^N \rightarrow \overline{\mathbb{R}}$ in (17) is proper, lower semi-continuous, convex and coercive, hence it admits global minimisers. Furthermore, if matrix \mathbf{SK} has full rank, then the global minimiser is unique.*

Assumptions (A3)–(A4) are crucial for our proposal. In fact, as it will be detailed in the paper, they allow for the efficient automatic selection of the regularisation parameter

in the frequency domain based on the RWP. More specifically, (A3) guarantees that, under the assumption of periodic boundary conditions, the blur matrix \mathbf{K} and the regularisation matrices $\mathbf{L}_j, j = 1, \dots, s$, are all block-circulant matrices with circulant blocks, which can be diagonalised by the 2D discrete Fourier transform, i.e.:

$$\mathbf{K} = \mathbf{F}^H \boldsymbol{\lambda} \mathbf{F}, \quad \mathbf{L}_j = \mathbf{F}^H \boldsymbol{\Gamma}_j \mathbf{F}, \quad j \in \{1, \dots, s\},$$

$$\mathbf{F}^H \mathbf{F} = \mathbf{F} \mathbf{F}^H = \mathbf{I}_N, \tag{18}$$

where $\boldsymbol{\lambda}, \boldsymbol{\Gamma}_j \in \mathbb{C}^{N \times N}$ are diagonal matrices defined by

$$\boldsymbol{\lambda} = \text{diag} \left(\tilde{\lambda}_1, \dots, \tilde{\lambda}_n \right),$$

$$\boldsymbol{\Gamma}_j = \text{diag} \left(\tilde{\gamma}_{j,1}, \dots, \tilde{\gamma}_{j,N} \right), \quad j \in \{1, \dots, s\}. \tag{19}$$

Assumption (A4) allows to apply Lemma 4 and Proposition 1 which, together with Fourier-diagonalisation formulas (18)–(19) allow to solve in the frequency domain the linear systems arising in the proposed iterative solution procedure. The Fourier representation allows to obtain an explicit form of the whiteness measure function such that the regularisation parameter μ can be efficiently updated along the solver iterations based on a simple univariate minimisation problem, according to the RWP. We remark that assumptions (A3)–(A4) cannot be relaxed without making the computational burden of our proposal impractical: even in the case when the arising linear systems can be solved very efficiently (e.g. if matrix \mathbf{L} is a tall matrix such that $\mathbf{L}^H \mathbf{L}$ has small size and can be inverted cheaply), if an explicit frequency domain solution is not available, then the repeated evaluation of the whiteness measure function becomes impractical.

Assumption (A5) is required to compute efficiently the solution of all the proximity steps arising in the solution algorithm.

We now briefly discuss how stringent the above assumptions are. First, assumption (A4) on \mathbf{S} is not stringent at all. In fact, if the binary selection matrix \mathbf{S} is preceded by the integration of the HR image over the support of each LR pixel, a *pixel blur* operator—representing a 2D convolution operator - can be introduced and incorporated in the original blur matrix \mathbf{K} .

Next, as far as popular regularisation terms satisfying (A2),(A3),(A5) are concerned, we mention the Tikhonov (TIK) regulariser, which, in its general form reads $G(\mathbf{Lx}) = \|\mathbf{Lx} - \mathbf{v}\|_2^2$ for suitably chosen \mathbf{L} and $\mathbf{v} \in \mathbb{R}^N$ reflecting any prior knowledge on the solution. When $\mathbf{L} = \mathbf{D}$, the discrete image gradient, such regulariser is commonly referred to as discrete Sobolev regularisation and it is typically adopted when the image of interest is characterised by smooth regions. Note, that in the following we will consider only such gradient-choice of the Tikhonov regulariser, hence we will denote by TIK the discrete Sobolev regularisation

functional $G(\mathbf{Lx}) = \|\mathbf{Dx}\|_2^2$. Alternatively, under the same choice for $\mathbf{L} = \mathbf{D}$, another common choice for $G(\mathbf{Lx})$ is the Total Variation (TV) regulariser [30], which is employed for the reconstruction of piece-wise constant images with sharp edges, and that admits an Isotropic (TVI) and Anisotropic (TVA) formulation. They can be expressed in terms of the regularisation matrix \mathbf{L} and of the nonlinear regularisation function $G(\mathbf{t})$, $\mathbf{t} = \mathbf{Lx}$, as follows:

$$\mathbf{L} = \mathbf{D}, \quad G(\mathbf{t}) = \|\mathbf{t}\|_2^2, \quad [\text{TIK}]$$

$$\mathbf{L} = \mathbf{D}, \quad G(\mathbf{t}) = \sum_{i=1}^N \|(t_i, t_{i+N})\|_2, \quad [\text{TVI}]$$

$$\mathbf{L} = \mathbf{D}, \quad G(\mathbf{t}) = \sum_{i=1}^N \|(t_i, t_{i+N})\|_1, \quad [\text{TVA}]$$

where $\mathbf{D} := (\mathbf{D}_h^T, \mathbf{D}_v^T)^T \in \mathbb{R}^{2N \times N}$ with $\mathbf{D}_h, \mathbf{D}_v \in \mathbb{R}^{N \times N}$ finite difference operators discretising the first-order horizontal and vertical partial derivatives.

In presence of images characterised by different local features, the global nature of TIK and TV compromises their performance. As a way to promote regularisation with different strength over the image, in [9] a class of weighted TV-based regularisers has been proposed; in formula:

$$\mathbf{L} = \mathbf{D}, \quad G(\mathbf{t}) = \sum_{i=1}^N \alpha_i \|(t_i, t_{i+N})\|_2, \quad \alpha_i \in \mathbb{R}_{++}, \quad [\text{WTV}]$$

Regardless of its local or global nature, a regularisation term designed by setting $\mathbf{L} = \mathbf{D}$ is expected to promote gradient sparsity. When dealing with sparse-recovery problems, i.e. problems where the signal itself is assumed to be sparse, regularisations based, e.g., on the use of the ℓ_1 norm are often considered and possibly combined with a space-variant modelling so as to get a Weighted ℓ_1 (WL1) regulariser, which reads:

$$\mathbf{L} = \mathbf{I}_N, \quad G(\mathbf{t}) = \sum_{i=1}^N w_i |t_i|, \quad w_i \in \mathbb{R}_{++}, \quad [\text{WL1}]$$

Remark 1 (Non-convex regularisations) Despite our convexity assumption (A2), we will detail in Sect. 6 few insights on the use of the proposed parameter selection strategy to non-convex regularisers corresponding to continuous approximations of the ℓ_0 pseudo-norm, such as, e.g., the CEL0 penalty [32]. The iterative strategy we are going to discuss next does not directly apply to this scenario; nonetheless, upon the suitable definition of appropriate surrogate convex functions approximating the original non-convex problems

(by means, for instance, of reweighted ℓ_1 algorithms [23]), its use is still possible.

3 The RWP for Single Image Super-Resolution

In this section, we recall some key results originally reported in [26] and concerning the application of the RWP to Tikhonov-regularised super-resolution least squares problems of the form (4). In fact, what follows represents the building block of the iterative procedures introduced in Sect. 5.

Let us consider the noise realisation \mathbf{e} in (1) in its original $n_r \times n_c$ matrix form:

$$\mathbf{e} = \{e_{i,j}\}_{(i,j) \in \Omega}, \quad \Omega := \{0, \dots, n_r - 1\} \times \{0, \dots, n_c - 1\}.$$

The *sample auto-correlation* $a : \mathbb{R}^{n_r \times n_c} \rightarrow \mathbb{R}^{(2n_r-1) \times (2n_c-1)}$ of realisation \mathbf{e} is

$$a(\mathbf{e}) = \{a_{l,m}(\mathbf{e})\}_{(l,m) \in \Theta}, \quad \Theta := \{-(n_r - 1), \dots, n_r - 1\} \times \{-(n_c - 1), \dots, n_c - 1\},$$

with each scalar component $a_{l,m}(\mathbf{e}) : \mathbb{R}^{n_r \times n_c} \rightarrow \mathbb{R}$ given by

$$\begin{aligned} a_{l,m}(\mathbf{e}) &= \frac{1}{n} (\mathbf{e} \star \mathbf{e})_{l,m} = \frac{1}{n} (\mathbf{e} * \mathbf{e}')_{l,m} \\ &= \frac{1}{n} \sum_{(i,j) \in \Omega} e_{i,j} e_{i+l,j+m}, \quad (l,m) \in \Theta, \end{aligned} \quad (20)$$

where index pairs (l,m) are commonly called *lags*, \star and $*$ denote the 2-D discrete correlation and convolution operators, respectively, and where $\mathbf{e}'(i,j) = \mathbf{e}(-i,-j)$. Clearly, for (20) being defined for all lags $(l,m) \in \Theta$, the noise realisation \mathbf{e} must be padded with at least $n_r - 1$ samples in the vertical direction and $n_c - 1$ samples in the horizontal direction by assuming periodic boundary conditions, such that \star and $*$ in (20) denote 2-D circular correlation and convolution, respectively. This allows to consider only lags

$$(l,m) \in \bar{\Theta} := \{0, \dots, n_r - 1\} \times \{0, \dots, n_c - 1\}.$$

If the corruption \mathbf{e} in (1) is the realisation of a white Gaussian noise process—as in our case—it is well known that as $n \rightarrow +\infty$, the sample auto-correlation $a_{l,m}(\mathbf{e})$ satisfies the following asymptotic property [20]:

$$\lim_{n \rightarrow +\infty} a_{l,m}(\mathbf{e}) = \begin{cases} \sigma^2 & \text{for } (l,m) = (0,0) \\ 0 & \text{for } (l,m) \in \bar{\Theta}_0 := \bar{\Theta} \setminus \{(0,0)\}. \end{cases} \quad (21)$$

We remark that the DP exploits only the information at lag $(0, 0)$. In fact, according to (3), the standard deviation of the residual image is required to be equal to the noise standard deviation σ . Imposing whiteness of the residual image by constraining its auto-correlation at non-zero lags to be small is a stronger requirement.

In order to make such whiteness principle independent of the noise level, we consider the *normalised* sample auto-correlation of noise realisation \mathbf{e} , namely

$$\rho(\mathbf{e}) = \frac{1}{a_{0,0}(\mathbf{e})} a(\mathbf{e}) = \frac{1}{\|\mathbf{e}\|_2^2} (\mathbf{e} \star \mathbf{e}),$$

where $\|\cdot\|_2$ denotes here the Frobenius norm. It follows easily from (21) that $\rho(\mathbf{e})$ satisfies the following asymptotic properties:

$$\lim_{n \rightarrow +\infty} \rho_{l,m}(\mathbf{e}) = \begin{cases} 1 & \text{for } (l, m) = (0, 0) \\ 0 & \text{for } (l, m) \in \widetilde{\Theta}_0. \end{cases}$$

In [22], the authors introduce the following non-negative scalar measure of whiteness $\mathcal{W} : \mathbb{R}^{n_r \times n_c} \rightarrow \mathbb{R}_+$ of noise realisation \mathbf{e} :

$$\mathcal{W}(\mathbf{e}) := \|\rho(\mathbf{e})\|_2^2 = \frac{\|\mathbf{e} \star \mathbf{e}\|_2^2}{\|\mathbf{e}\|_2^4} = \widetilde{\mathcal{W}}(\tilde{\mathbf{e}}), \tag{22}$$

where the last equality comes from Proposition 3 below—the proof being reported in [22]—with $\tilde{\mathbf{e}} \in \mathbb{C}^{n_r \times n_c}$ the 2D discrete Fourier transform of \mathbf{e} and $\widetilde{\mathcal{W}} : \mathbb{C}^{n_r \times n_c} \rightarrow \mathbb{R}_+$ the function defined in (23).

Proposition 3 *Let $\mathbf{e} \in \mathbb{R}^{n_r \times n_c}$ and $\tilde{\mathbf{e}} = \mathbf{F}\mathbf{e} \in \mathbb{C}^{n_r \times n_c}$. Then, assuming periodic boundary conditions for \mathbf{e} , the function \mathcal{W} defined in (22) satisfies:*

$$\mathcal{W}(\mathbf{e}) = \widetilde{\mathcal{W}}(\tilde{\mathbf{e}}) = \left(\sum_{(l,m) \in \widetilde{\Theta}} |\tilde{e}_{l,m}|^4 \right) / \left(\sum_{(l,m) \in \widetilde{\Theta}} |\tilde{e}_{l,m}|^2 \right)^2. \tag{23}$$

By now looking at the considered class of super-resolution variational models (17), we observe that, as a general rule, the closer the attained super-resolved image $\mathbf{x}^*(\mu)$ to the target HR image \mathbf{x} , the closer the associated residual image $\mathbf{r}^*(\mu) = \mathbf{S}\mathbf{K}\mathbf{x}^*(\mu) - \mathbf{b}$ to the white noise realisation \mathbf{e} in (1) and, hence, the whiter the residual image according to the scalar measure in (22).

This motivates the application of the RWP for automatically selecting the regularisation parameter μ in variational models of the form (17), which reads:

$$\begin{aligned} &\text{select } \mu = \mu^* \text{ s.t.} \\ &\mu^* \in \arg \min_{\mu \in \mathbb{R}_{++}} \{ W(\mu) := \mathcal{W}(\mathbf{r}^*(\mu)) \}, \end{aligned} \tag{24}$$

where, according to the definition of function \mathcal{W} in (22)–(23), the scalar non-negative cost function $W : \mathbb{R}_{++} \rightarrow \mathbb{R}_+$ in (24), from now on referred to as the *residual whiteness function*, takes the following form:

$$W(\mu) = \frac{\|\mathbf{r}^*(\mu) \star \mathbf{r}^*(\mu)\|_2^2}{\|\mathbf{r}^*(\mu)\|_2^4} = \widetilde{\mathcal{W}}(\widetilde{\mathbf{r}^*(\mu)}), \tag{25}$$

with $\widetilde{\mathbf{r}^*(\mu)}$ the 2D discrete Fourier transform of the residual image and function $\widetilde{\mathcal{W}}$ defined in (23). In [26], the authors derived an explicit expression for the super-resolution residual whiteness function $W(\mu)$ in (25) in the frequency domain which generalises the one for the restoration-only case (i.e. the case where $\mathbf{S} = \mathbf{I}_N$) reported in [22]. The results of derivations in [26] are summarised in the following proposition, whose proof is reported the appendix.

Proposition 4 *Let $\mathbf{r}_H(\mu) := \mathbf{K}\mathbf{x}(\mu) - \mathbf{b}_H$, with $\mathbf{b}_H = \mathbf{S}^H\mathbf{b}$, be the high-resolution residual image and let $\mathbf{P} \in \mathbb{R}^{N \times N}$ be the permutation matrix defined in (9). Then, we have:*

$$\begin{aligned} W(\mu) &= \left(\sum_{i=1}^N w_i(\mu)^4 \right) / \left(\sum_{i=1}^N w_i(\mu)^2 \right)^2, \\ w_i(\mu) &= \left| \sum_{\ell=0}^{d-1} (\hat{\mathbf{r}}_H(\mu))_{i+\ell} \right|, \end{aligned} \tag{26}$$

where

$$\hat{\mathbf{r}}_H(\mu) = \mathbf{P}\tilde{\mathbf{r}}_H(\mu) \quad \text{and} \quad \iota := 1 + \left\lfloor \frac{i-1}{d} \right\rfloor d.$$

4 RWP for Tikhonov-Regularised SR Problems

Here, we derive the analytical expression of the whiteness function $W(\mu)$ defined in (26) when addressing Tikhonov-regularised least squares problems as the one in (4). More specifically, following the derivations reported in [43], in Proposition 5 we give an efficiently computable expression for the solution $\mathbf{x}^*(\mu)$ of the general ℓ_2 - ℓ_2 variational model.

Proposition 5 *Let*

$$\underline{\lambda} := (\mathbf{I}_n \otimes \mathbf{1}_d) \mathbf{P} \mathbf{\Lambda}, \quad \underline{\lambda}^H := \lambda^H \mathbf{P}^T (\mathbf{I}_n \otimes \mathbf{1}_d) \tag{27}$$

and

$$\Psi = \left(\sum_{j=1}^s \Gamma_j^H \Gamma_j + \epsilon \right), \quad 0 < \epsilon \ll 1,$$

with ϵ guaranteeing the inversion of $\sum_{j=1}^s \Gamma_j^H \Gamma_j$. The solution of the Tikhonov-regularised least squares problem can be expressed as:

$$\mathbf{x}^*(\mu) = \mathbf{F}^H \left[\Psi - \mu \Psi \underline{\lambda}^H \left(d\mathbf{I} + \mu \underline{\lambda} \Psi \underline{\lambda}^H \right)^{-1} \underline{\lambda} \Psi \right] \left(\mu \underline{\lambda}^H \tilde{\mathbf{b}}_H + \sum_{j=1}^s \Gamma_j^H \tilde{\mathbf{v}}_j \right), \tag{28}$$

where $\mathbf{v} = (\mathbf{v}_1^T, \dots, \mathbf{v}_s^T)^T$.

From (28), we can easily derive the Fourier transform of the high resolution residual $\mathbf{r}_H^*(\mu) = \mathbf{K}\mathbf{x}^*(\mu) - \mathbf{b}$, that takes the form

$$\tilde{\mathbf{r}}_H^*(\mu) = \underline{\lambda} \left[\Psi - \mu \Psi \underline{\lambda}^H \left(d\mathbf{I} + \mu \underline{\lambda} \Psi \underline{\lambda}^H \right)^{-1} \underline{\lambda} \Psi \right] \left(\mu \underline{\lambda}^H \tilde{\mathbf{b}}_H + \sum_{j=1}^s \Gamma_j^H \tilde{\mathbf{v}}_j \right) - \tilde{\mathbf{b}}_H, \tag{29}$$

Recalling Lemma 1 and the property (10), we prove the following proposition and corollary.

Proposition 6 Let $\Phi \in \mathbb{R}^{n \times n}$ be a diagonal matrix and consider the matrix $\underline{\lambda}$ defined in (27). Then, the following equality holds:

$$\underline{\lambda}^H \Phi \underline{\lambda} = \mathbf{P}^T (\Phi \otimes \mathbf{I}_d) \mathbf{P} \underline{\lambda}^H \underline{\lambda}.$$

Corollary 1 Let $\Phi = (d\mathbf{I} + \mu \underline{\lambda} \Psi \underline{\lambda}^H)^{-1}$. Then, the expression in (29) turns into

$$\tilde{\mathbf{r}}_H^*(\mu) = \underline{\lambda} \left[\Psi - \mu \Psi \mathbf{P}^T \left((d\mathbf{I} + \mu \underline{\lambda} \Psi \underline{\lambda}^H)^{-1} \otimes \mathbf{I}_d \right) \mathbf{P} \underline{\lambda}^H \underline{\lambda} \Psi \right] \left(\mu \underline{\lambda}^H \tilde{\mathbf{b}}_H + \sum_{j=1}^s \Gamma_j^H \tilde{\mathbf{v}}_j \right) - \tilde{\mathbf{b}}_H.$$

Recalling now the action of the permutation matrix \mathbf{P} on vectors, we have that the product $\hat{\mathbf{r}}_H^*(\mu) = \mathbf{P} \tilde{\mathbf{r}}_H^*(\mu)$ reads

$$\hat{\mathbf{r}}_H^*(\mu) = \left[\widehat{\underline{\lambda}} \Psi - \mu \widehat{\underline{\lambda}} \Psi \left((d\mathbf{I} + \mu \widehat{\underline{\lambda}} \Psi \widehat{\underline{\lambda}}^H)^{-1} \otimes \mathbf{I}_d \right) \widehat{\underline{\lambda}}^H \widehat{\underline{\lambda}} \Psi \right] \left(\mu \widehat{\underline{\lambda}}^H \tilde{\mathbf{b}}_H + \sum_{j=1}^s \tilde{\Gamma}_j^H \tilde{\mathbf{v}}_j \right) - \hat{\mathbf{b}}_H,$$

where the matrix $\widehat{\underline{\lambda}}^H \widehat{\underline{\lambda}} \Psi = \mathbf{P} \underline{\lambda}^H \mathbf{P}^T (\mathbf{I}_n \otimes \mathbf{J}_d) \mathbf{P} \underline{\lambda} \Psi \mathbf{P}^T$ acts on $\mathbf{g} \in \mathbb{R}^N$ as

$$\left(\widehat{\underline{\lambda}}^H \widehat{\underline{\lambda}} \Psi \mathbf{g} \right)_i = \bar{\lambda}_i \sum_{\ell=0}^{d-1} \frac{\hat{\lambda}_{i+\ell}}{\zeta_{i+\ell} + \epsilon} g_{i+\ell},$$

with $\zeta_{i+\ell} = \sum_{j=1}^s |\hat{\gamma}_{j,i+\ell}|^2$

Combining altogether, we finally deduce:

Proposition 7 The residual whiteness function for the generalised Tikhonov least squares problem takes the form

$$W(\mu) = \left(\sum_{i=1}^N \left| \frac{v_i - \rho_i}{1 + \eta_i \mu} \right|^4 \right) / \left(\sum_{i=1}^N \left| \frac{v_i - \rho_i}{1 + \eta_i \mu} \right|^2 \right)^2, \tag{30}$$

where the parameters $\eta_i \in \mathbb{R}_+$, $\rho_i \in \mathbb{C}$, $v_i \in \mathbb{C}$ are defined as:

$$\eta_i := \frac{1}{d} \sum_{j=0}^{d-1} \frac{|\hat{\lambda}_{i+j}|^2}{\zeta_{i+j} + \epsilon}, \quad \rho_i := \sum_{\ell=0}^{d-1} \hat{b}_{H,i+\ell},$$

$$v_i := \sum_{\ell=0}^{d-1} \frac{\hat{\lambda}_{i+\ell} \sum_{j=1}^s \tilde{\gamma}_{j,i+\ell} \tilde{v}_{j,i+\ell}}{\zeta_{i+\ell} + \epsilon}. \tag{31}$$

When $d = 1$, i.e. no decimation is considered and $\mathbf{S} = \mathbf{I}_N$, the expression in (30) reduces to the one derived in [22] for image restoration problems.

According to the RWP, the optimal μ^* is selected as the one minimising the whiteness measure function in (30). The value μ^* can be efficiently detected via grid-search or applying the Newton-Raphson algorithm to the nonlinear equation $W'(\mu) = 0$. Finally, the optimal μ^* is used for the computation of the reconstruction $\mathbf{x}^*(\mu^*)$ based on (28).

The main steps of the proposed procedure are summarised in Algorithm 1.

Algorithm 1: Exact RWP approach for image super-resolution ℓ_2 - ℓ_2 variational models of the form (17) under assumptions (A1), (A3)

inputs: observed degraded image $\mathbf{b} \in \mathbb{R}^n$,
 decimation factors d_r, d_c ,
 blur and regularisation convolution operators $\underline{\lambda}, \Gamma_j \in \mathbb{R}^{N \times N}$, $j \in \{1, \dots, s\}$,
 permutation matrix $\mathbf{P} \in \mathbb{R}^{N \times N}$,

output: estimated super-resolved image $\mathbf{x}^* \in \mathbb{R}^N$

1. **compute:** $\underline{\lambda} = (\mathbf{I}_n \otimes \mathbf{1}_d^T) \mathbf{P} \underline{\lambda}$, $\Psi = (\sum_j \Gamma_j^H \Gamma_j + \epsilon)^{-1}$
 2. **select:** $\mu^* \in \arg \min W(\mu)$
 3. **compute:** $\mathbf{x}^*(\mu^*)$ by (63)
-

5 Iterated RWP for Convex-Regularised SR Problems

In this section, we present an ADMM-based iterative algorithm for the solution of SR variational models of the general form (17) under assumptions (A1)–(A5), where the regularisation parameter μ is automatically updated along the iterations based on the RWP. The approach relies on the iterative application of the results reported in the previous section.

First, we employ variable splitting to rewrite our family of variational models (17) in the following equivalent linearly constrained form:

$$\{\mathbf{x}^*(\mu), \mathbf{t}^*(\mu)\} \in \arg \min_{\mathbf{x}, \mathbf{t}} \left\{ \frac{\mu}{2} \|\mathbf{S}\mathbf{K}\mathbf{x} - \mathbf{b}\|_2^2 + G(\mathbf{t}) \right\} \text{ s.t. : } \mathbf{t} = \mathbf{L}\mathbf{x}. \tag{32}$$

To solve problem (32), we defined the augmented Lagrangian function,

$$\mathcal{L}(\mathbf{x}, \mathbf{t}, \boldsymbol{\lambda}; \mu) = \frac{\mu}{2} \|\mathbf{S}\mathbf{K}\mathbf{x} - \mathbf{b}\|_2^2 + G(\mathbf{t}) - \langle \boldsymbol{\lambda}, \mathbf{t} - \mathbf{L}\mathbf{x} \rangle + \frac{\beta}{2} \|\mathbf{t} - \mathbf{L}\mathbf{x}\|_2^2, \tag{33}$$

where $\beta > 0$ is a scalar penalty parameter and $\boldsymbol{\lambda} \in \mathbb{R}^M$ is the dual variable, i.e. the vector of Lagrange multipliers associated with the set of M linear constraints in (32). Solving (32) corresponds to seek for the saddle point(s) of the augmented Lagrangian function in (33), i.e.:

$$\{\mathbf{x}^*(\mu), \mathbf{t}^*(\mu), \boldsymbol{\lambda}^*(\mu)\} \in \arg \min_{\mathbf{x}, \mathbf{t}} \max_{\boldsymbol{\lambda}} \mathcal{L}(\mathbf{x}, \mathbf{t}, \boldsymbol{\lambda}; \mu). \tag{34}$$

Upon suitable initialisation, and for any $k \geq 0$, the k -th iteration of the standard ADMM applied to solving the saddle-point problem (34) with \mathcal{L} defined in (33) reads as follows:

$$\mathbf{x}^{(k+1)} \in \arg \min_{\mathbf{x} \in \mathbb{R}^n} \mathcal{L}(\mathbf{x}, \mathbf{t}^{(k)}, \boldsymbol{\lambda}^{(k)}; \mu) \tag{35}$$

$$= \arg \min_{\mathbf{x} \in \mathbb{R}^n} \left\{ \frac{\mu/\beta}{2} \|\mathbf{S}\mathbf{K}\mathbf{x} - \mathbf{b}\|_2^2 + \frac{1}{2} \|\mathbf{L}\mathbf{x} - \mathbf{v}^{(k+1)}\|_2^2 \right\} \tag{36}$$

$$\text{with } \mathbf{v}^{(k+1)} = \mathbf{t}^{(k)} - \frac{\boldsymbol{\lambda}^{(k)}}{\beta},$$

$$\mathbf{t}^{(k+1)} \in \arg \min_{\mathbf{t} \in \mathbb{R}^M} \mathcal{L}(\mathbf{x}^{(k+1)}, \mathbf{t}, \boldsymbol{\lambda}^{(k)}; \mu) \tag{37}$$

$$= \arg \min_{\mathbf{t} \in \mathbb{R}^M} \left\{ G(\mathbf{t}) + \frac{\beta}{2} \|\mathbf{t} - \mathbf{q}^{(k+1)}\|_2^2 \right\} \tag{38}$$

$$\text{with } \mathbf{q}^{(k+1)} = \mathbf{L}\mathbf{x}^{(k+1)} + \frac{\boldsymbol{\lambda}^{(k)}}{\beta},$$

$$\boldsymbol{\lambda}^{(k+1)} = \boldsymbol{\lambda}^{(k)} - \beta \left(\mathbf{t}^{(k+1)} - \mathbf{L}\mathbf{x}^{(k+1)} \right). \tag{39}$$

One can notice that when introducing the additional parameter $\gamma := \mu/\beta$, the minimisation sub-problem (35) for the primal variable \mathbf{x} has the form of the Tikhonov-regularised least-squares problem (4). Hence, we adjust the regularisation parameter μ —i.e. γ —along the ADMM iterations by applying the RWP to problem (35) as illustrated in Sect. 4.

The complete \mathbf{x} -update procedure reads as follows:

$$\mathbf{v}^{(k+1)} = \mathbf{t}^{(k)} - \frac{\boldsymbol{\lambda}^{(k)}}{\beta}, \tag{40}$$

$$\gamma^{(k+1)} = \arg \min_{\gamma \in \mathbb{R}_+} W(\gamma) \text{ with } W \text{ in (30)–(31)} \tag{41}$$

$$\mathbf{x}^{(k+1)} = \mathbf{F}^H \left[\boldsymbol{\Psi} - \gamma^{(k+1)} \boldsymbol{\Psi} \boldsymbol{\Lambda}^H \left(d\mathbf{I} + \gamma^{(k+1)} \boldsymbol{\Lambda} \boldsymbol{\Psi} \boldsymbol{\Lambda}^H \right)^{-1} \boldsymbol{\Lambda} \boldsymbol{\Psi} \right] \left(\gamma^{(k+1)} \boldsymbol{\Lambda}^H \tilde{\mathbf{b}}_H + \sum_{j=1}^s \boldsymbol{\Gamma}_j^H \tilde{\mathbf{v}}_j \right). \tag{42}$$

The minimisation sub-problem (37) for the primal variable \mathbf{t} can be written in the form of a proximity operator, namely

$$\mathbf{t}^{(k+1)} \in \text{prox}_{\frac{1}{\beta}G} \left(\mathbf{q}^{(k+1)} \right), \quad \mathbf{q}^{(k+1)} = \mathbf{L}\mathbf{x}^{(k+1)} + \frac{\boldsymbol{\lambda}^{(k)}}{\beta}. \tag{43}$$

According to assumption (A5), the regularisation function G is easily proximable, which means that problem (43) can be solved efficiently.

We now report the closed-form expression of the proximity operators for the regularisation terms listed in Sect. 2.2.

For what concerns the case of the TV and WTV regularisers, the associated $2N$ -variate proximity operators associated to the splitting performed are both separable into N independent bivariate proximity operators. In particular, after introducing the N vectors $\check{\mathbf{t}}_i^{(k+1)}, \check{\mathbf{q}}_i^{(k+1)} \in \mathbb{R}^2$ defined by

$$\check{\mathbf{t}}_i^{(k+1)} := \left(t_i^{(k+1)}, t_{i+N}^{(k+1)} \right), \quad \check{\mathbf{q}}_i^{(k+1)} := \left(q_i^{(k+1)}, q_{i+N}^{(k+1)} \right), \quad i \in \{1, \dots, N\}, \tag{44}$$

the proximity operators admit the following closed-form expressions:

$$\check{\mathbf{t}}_i^{(k+1)} = \max \left(1 - \frac{1}{\beta \|\check{\mathbf{q}}_i^{(k+1)}\|_2}, 0 \right) \check{\mathbf{q}}_i^{(k+1)}, \quad i \in \{1, \dots, N\}, \quad [\text{TVI}]$$

$$\check{\mathbf{t}}_i^{(k+1)} = \max \left(1 - \frac{1}{\beta \|\check{\mathbf{q}}_i^{(k+1)}\|_1}, 0 \right) \check{\mathbf{q}}_i^{(k+1)}, \quad i \in \{1, \dots, N\}, \quad [\text{TVA}]$$

$$\check{\mathbf{t}}_i^{(k+1)} = \max \left(1 - \frac{\alpha_i^{(k)}}{\beta \|\check{\mathbf{q}}_i^{(k+1)}\|_2}, 0 \right) \check{\mathbf{q}}_i^{(k+1)},$$

$$i \in \{1, \dots, N\}, \quad [\text{WTV}] \quad (45)$$

where $\alpha_i^{(k)}$ denote the spatial weights of the WTV regulariser (45).

Finally, the proximity operator for the WL1 regularisation term reads:

$$t_i = \text{sign}(q_i) \max \left(0, |q_i| - \frac{w_i}{\beta} \right), i = 1, \dots, N. \quad [\text{WL1}] \quad (46)$$

Algorithm 2: IRWP-ADMM approach for single-image SR variational models of the form (17) under assumptions (A1)-(A5)

inputs: observed degraded image $\mathbf{b} \in \mathbb{R}^n$,
 decimation factors d_r, d_c
 blur and regularisation convolution operators λ ,
 $\Gamma_j \in \mathbb{R}^{N \times N}$, $j \in \{1, \dots, s\}$
 regularisation weights w_i , $i \in \{1, \dots, N\}$
 [only for WL1]

output: estimated super-resolved image $\mathbf{x}^* \in \mathbb{R}^N$

1. **initialise:** compute $\mathbf{x}^{(0)}$ by TIK-L₂, then $\mathbf{t}^{(0)} = \mathbf{L}\mathbf{x}^{(0)}$
 2. **for** $k = 0, 1, 2, \dots$ **until convergence** **do:**
 3. · compute $\gamma^{(k+1)} = \mu^{(k+1)}/\beta$ by (40)-(41)
 4. · compute $\mathbf{x}^{(k+1)}$ by (42)
 5. · compute $\alpha^{(k+1)}$ by (13) in [9]
 [only for WTV]
 6. · compute $\mathbf{t}^{(k+1)}$ by solving (43)
 [(44), (45), (46) for TV, WTV, WL1]
 7. · compute $\lambda^{(k+1)}$ by (39)
 8. **end for**
 9. $\mathbf{x}^* = \mathbf{x}^{(k+1)}$
-

In Algorithm 2 we outline the main computational steps of the proposed approach, which we refer to as Iterative RWP-ADMM, in short IRWP-ADMM. As initial guess of the IRWP-ADMM, we select $\mathbf{x}^{(0)}$ to be the solution of the TIK-L₂ model in (4) with $\mathbf{L} = \mathbf{D}$ and \mathbf{v} the null vector (i.e. by using Sobolev regularisation), and the regularisation parameter $\mu^{(0)}$ computed by applying the exact RWP outlined in Sect. 4, i.e. by minimising (30). In the context of image restoration [22], this choice has already been proven to facilitate and speed up the convergence of the employed iterative scheme.

Convergence of two-blocks ADMM applied to the solution of convex optimisation problems of the form (17) with fixed parameters is well-established (see, e.g., [3]). One can thus exploit such standard result for convergence of Algorithm 2 by simply setting fixed values of the parameters (regularisation parameter μ and, possibly, spatial weights of the WTV and WL1 regularisers) starting from a given iteration. In the numerical tests shown in Sect. 7, we observed that the parameters typically stabilise after 400/500 iterations of the IRWP-ADMM.

6 An Extension of IRWP for SR Problems with Nonconvex Regularisers

In this section we show an example of how IRWP can be applied to nonconvex regularisation problems that do not satisfy assumption (A2) on the convexity of the nonlinear function G .

In several sparse-recovery problems, such as, e.g., variable selection, biological image super-resolution one typically would like to consider relaxations of the ℓ_0 -norm tighter than the ℓ_1 so as to improve the sparsity-promoting behaviour. Among the many approaches proposed, we consider here the CEL0 penalty term considered in [32] which, combined with a non-negativity constraint yields the following optimisation problem

$$\mathbf{x}^*(\mu) \in \arg \min_{\mathbf{x} \in \mathbb{R}_+^N} \left\{ \mathcal{J}_{\text{CEL0}}(\mathbf{x}; \mu, \mathbf{A}) \right.$$

$$\left. := \Phi_{\text{CEL0}}(\mathbf{x}; \mu, \mathbf{A}) + \frac{\mu}{2} \|\mathbf{A}\mathbf{x} - \mathbf{b}\|_2^2 \right\}, \quad (47)$$

with

$$\mathbf{A} = \mathbf{S}\mathbf{K}, \quad \Phi_{\text{CEL0}}(\mathbf{x}; \mu, \mathbf{A}) = \sum_{i=1}^N \phi_{\text{CEL0}}(x_i; \mu, \|\mathbf{a}_i\|_2),$$

where $\mathbf{a}_i \in \mathbb{R}^N$ denotes the i -th column of matrix \mathbf{A} and the (parametric) penalty function $\Phi_{\text{CEL0}} : \mathbb{R} \rightarrow \mathbb{R}_+$ reads

$$\phi_{\text{CEL0}}(x_i; \mu, \|\mathbf{a}_i\|_2) = 1 - \frac{\mu \|\mathbf{a}_i\|_2^2}{2}$$

$$\left(|x_i| - \frac{\sqrt{2/\mu}}{\|\mathbf{a}_i\|_2} \right)^2 \chi_{[0, \sqrt{2/\mu}/\|\mathbf{a}_i\|_2]}(|x_i|). \quad (48)$$

Notice that, in agreement with our convention by which the desired parameter μ is the one multiplying the fidelity term, in (47)–(48) we divided [32] by its regularisation parameter λ and set $\mu := 1/\lambda$.

In [15], problem (47) is solved by means of the iterative reweighted ℓ_1 (IRL1) algorithm [23], which belongs to the class of so-called Majorise-Minimise (MM) optimisation

approaches (see, e.g., [19]). More precisely, at any iteration $h > 0$ of the IRL1 scheme, one minimises a convex majorant of the non-convex cost function J_{CEL0} which is tangent to J_{CEL0} at the current iterate $\mathbf{x}^{(h)}$. The majorising variational model can be defined as a convex weighted L_1 - L_2 (WL1- L_2) model so that, upon suitable initialisation $\mathbf{x}^{(0)}$ the h -th iteration of the IRL1 algorithm applied to solving (47) reads (see [15]):

$$w_i^{(h)} = \mu \left(\sqrt{\frac{2}{\mu}} \|\mathbf{a}_i\|_2 - \|\mathbf{a}_i\|_2^2 |x_i^{(h)}| \right) X_{[0, \sqrt{2/\mu} \|\mathbf{a}_i\|_2]}(|x_i^{(h)}|), \tag{49}$$

$$\mathbf{x}^{(h+1)} = \arg \min_{\mathbf{x} \in \mathbb{R}^N} \left\{ \sum_{i=1}^N w_i^{(h)}(\mu) |x_i| + \frac{\mu}{2} \|\mathbf{A}\mathbf{x} - \mathbf{b}\|_2^2 \right\}, \tag{50}$$

where the weights $w_i(\mu)$ are the generalised derivatives of ϕ_{CEL0} computed at $\mathbf{x}^{(h)}$. The minimisation problem in (50) can be addressed, e.g., by ADMM—see (35)–(39)—with $\mathbf{L} = \mathbf{I}_N$. The sub-problem with respect to \mathbf{t} reduces to compute the proximity operator of function

$$G(\mathbf{t}) = \sum_{i=1}^N \left(w_i^{(h)} |t_i| + t_{\mathbb{R}_+}(t_i) \right),$$

whose closed-form expression comes easily from (46) and reads

$$t_i = \max \left(0, |q_i| - \frac{w_i}{\beta} \right), \quad i = 1, \dots, N. \tag{51}$$

In this setting, the RWP can again be heuristically applied so as to automatically update μ along the outer iterations of the iterative scheme (49) and (50). At the ADMM iteration $k = 0$ of the general outer iteration h , $\mu^{(h)}$ is updated by applying the residual whiteness principle to problem (35). Then, the weights $w_i^{(h)}$ are computed by (49) with $\mu = \mu^{(h)}$ and $\mathbf{x}^{(h)}$ current iterate. The regularisation parameter and the weights are kept fixed along the ADMM iterations and updated at the beginning of the following outer iteration.

The proposed procedure, to which we refer as IRWP-IRL1, is outlined in Algorithm 3.

7 Computed Examples

In this section, we evaluate the performance of the proposed RWP-based procedure for the automatic selection of the regularisation parameter μ in denoising, deblurring and super-resolution variational models of the form (17). More specifically, we consider first the TV- L_2 which is particularly

Algorithm 3: IRWP-IRL1 approach for image super-resolution variational models of the form (47)

inputs: observed degraded image $\mathbf{b} \in \mathbb{R}^n$,
 decimation factors d_r, d_c
 blur and regularisation convolution operators $\lambda, \Gamma_j \in \mathbb{R}^{N \times N}$, $j \in \{1, \dots, s\}$

output: estimated restored image $\mathbf{x}^* \in \mathbb{R}^N$

1. **initialise:** compute $\mathbf{x}^{(0)}$ and $\mathbf{t}^{(0)} = \mathbf{x}^{(0)}$ by IRWP-ADMM for L_1 - L_2
 2. **for** $h = 0, 1, 2, \dots$ **until convergence** **do:**
 3. · compute $\gamma^{(h)} = \mu^{(h)} / \beta$ by (40)–(41)
 4. · compute weights $w_i^{(h)} = w_i^{(h)}(\mu^{(h)}, \mathbf{x}^{(h)})$ by (49)
 5. **for** $k = 1, 2, \dots$ **until convergence** **do:**
 6. · compute $\mathbf{x}^{(h+1, k+1)}$ by (42)
 7. · compute $\mathbf{t}^{(h+1, k+1)}$ by solving (51)
 8. · compute $\lambda^{(h+1, k+1)}$ by (39)
 9. **end for**
 10. **end for**
 11. $\mathbf{x}^* = \mathbf{x}^{(k+1)}$
-

suitable for the reconstruction of piece-wise constant images, such as, e.g., QR-codes; then, we focus our attention on the reconstruction of natural images for which the more flexible WTV- L_2 is employed. For the first two sets of numerical tests, we also perform the TIK- L_2 variational model with $\mathbf{L} = \mathbf{D}$. Finally, we compare the L_1 - L_2 and the CEL0- L_2 variational models for the problem of super-resolution microscopy of phantom images representing biological samples.

In what follows, the output restorations obtained by means of the proposed RWP-based strategy will be denoted by $\mathbf{x}_{\text{REG}}^*$, with $\text{REG} \in \{\text{TIK}, \text{TV}, \text{WTV}, L_1, \text{CEL0}\}$ denoting the regularisation term employed in the solved variational model. Note that when $\text{REG} = \text{TIK}$ the exact RWP-based approach described in Sect. 4 is used, while for $\text{REG} = \text{TV}, \text{WTV}, L_1$ and $\text{REG} = \text{CEL0}$ the IRWP-ADMM and IRWP-IRL1 iterative methods detailed in Algorithms 2 and 3 are used, respectively.

The proposed RWP-based approach is compared with the DP parameter selection strategy, defined by criterion (3) with $\tau = 1$ and known value σ of the true noise standard deviation.

With the proposed numerical tests we aim to:

- prove that the RWP is capable of selecting optimal μ^* values returning high quality results in variational image super-resolution;
- prove that the proposed IRWP-ADMM approach is capable of automatically selecting such optimal μ^* values in a robust (and efficient) manner for non-quadratic super-resolution variational models.

The latter point will be confirmed by showing that the iterative IRWP-ADMM/IRWP-IRL1 strategies and the RWP applied *a posteriori* behave similarly.

For all the variational models considered in the experiments, there is a one-to-one relationship between the value of the regularisation parameter μ and the standard deviation of the associated residual image $\mathbf{r}^*(\mu) = \mathbf{S}\mathbf{K}\mathbf{x}^*(\mu) - \mathbf{b}$. Hence, in all the reported results where μ represents the independent variable, we will substitute the μ -values with the corresponding $\tau^*(\mu)$ -values defined, according to (3), as the ratio between the residual image standard deviation and the true noise standard deviation σ , in formula

$$\tau^*(\mu) := \frac{\|\mathbf{S}\mathbf{K}\mathbf{x}^*(\mu) - \mathbf{b}\|_2}{\sqrt{n}\sigma}. \quad (52)$$

In the first two set of examples, the quality of the restorations \mathbf{x}^* computed for different values of τ^* with respect to the original uncorrupted image \mathbf{x} will be assessed by means of three scalar measures, namely the Structural Similarity Index (SSIM) [39], and the Peak-Signal-to-Noise-Ratio (PSNR) and the Improved-Signal-to-Noise Ratio (ISNR), defined by

$$\text{PSNR} = \frac{20 \log_{10}(\sqrt{N} \max(\mathbf{x}, \mathbf{x}^*))}{\|\mathbf{x} - \mathbf{x}^*\|_2},$$

$$\text{ISNR} = 20 \log_{10} \frac{\|\mathbf{x} - \mathbf{x}_{\text{BIC}}\|_2}{\|\mathbf{x} - \mathbf{x}^*\|_2}$$

respectively, with $\max(\mathbf{x}, \mathbf{x}^*)$ representing the largest component-wise value of \mathbf{x} and \mathbf{x}^* , while \mathbf{x}_{BIC} denotes the bicubic interpolation of \mathbf{b} . In the third example, we select as measure of quality the Jaccard index $J_\delta \in [0, 1]$, which is the ratio between correct detections up to some tolerance δ and the sum of correct detections, false negatives and false positives. In particular, we consider three different values of $\delta \in \{0, 2, 4\}$.

7.1 IRWP-ADMM for TV Regularisation

We start testing the IRWP-ADMM on the TV- L_2 model for the reconstruction of the test image `qrcode` (256×256 pixels) with pixel values between 0 and 1.

First, the original image has been corrupted by Gaussian blur, generated by the MATLAB routine `fspecial` with parameters `band=13` and `sigma=3`. The action of the decimation matrix \mathbf{S} has been synthesised by applying a uniform blur with a $d_r \times d_c$ kernel, with $d_r = d_c = 4$, and then selecting the rows and the columns of the blurred image according to the decimation factors d_c, d_r . Finally, the blurred and decimated image has been further corrupted by AWGN with standard deviation $\sigma = 0.1$. The original image and the acquired data are shown in Fig. 3a, b, respectively.

Due to the strongly anisotropic nature of the geometric image contents, we employ both the isotropic and the anisotropic version of TV as image regularisers.

The black solid curves in Fig. 2a, d represent the residual whiteness functions $W(\mu)$ as defined in (30), with μ replaced by $\tau^*(\mu)$ defined in (52), for the TVI- L_2 and TVA- L_2 models. They have been computed by solving the models for a fine grid of different μ -values, and then calculating for each μ -value the two associated $\tau^*(\mu)$ and $W(\mu)$ quantities. Note that independently from the selected regulariser, W has a global minimiser over the considered domain. The optimal value $\tau^* = \tau^*(\mu^*)$ according to the proposed RWP is depicted by the vertical solid magenta lines, while the vertical dashed black lines correspond to $\tau = 1$ corresponding to the standard value of classical DP.

The ISNR and SSIM curves for different values of τ are plotted in Fig. 2b, c, where the vertical lines have the same meaning as in the first column figures. Note that the RWP tends to select a value for τ that maximises the ISNR rather than the SSIM.

We are also interested in verifying that the proposed IRWP-ADMM scheme outlined in Algorithm 2 succeeds in automatically selecting such optimal τ^* in a robust and efficient way. To this purpose, the output $\bar{\tau}$ of the iterative scheme is indicated with a dashed green line in Fig. 2a–d. The blue and red markers in Fig. 2b, c represent the final ISNR and SSIM values, respectively, of the image reconstructed via IRWP-ADMM. We observe that the algorithm returns a $\bar{\tau}$ which is close to τ^* detected *a posteriori*.

The image reconstructed by bicubic interpolation, the initial guess computed by the TIK- L_2 models and the output reconstructions obtained with the TVI- L_2 and the TVA- L_2 variational models solved by the IRWP-ADMM approach in Algorithm 2 are shown in Fig. 3. We note that the bicubic interpolation in Fig. 3c performs very poorly. The TVI- L_2 reconstruction preserves image edges; however, as observable in Fig. 3e, the rounding of corners, which is a typical drawback of isotropic TV regularisation, affects the quality of the final reconstruction. The anisotropic TV returns a high quality reconstruction—shown in Fig. 3f—as it naturally drives the regularisation along the true horizontal and the vertical edge directions.

Quantitative assessment in terms of PSNR, ISNR and SSIM values are reported in Table 1 where values corresponding to the reconstructions in Fig. 3 are reported to the right part.

As a second example, we consider the test image `geometric` (320×320) corrupted by the same blur, decimation factors and AWGN as the ones considered for the test image `qrcode`. The original image and the observed data are shown in Fig. 5a, b. In this case, we only perform the isotropic version of TV, to which we are going to refer as TV.

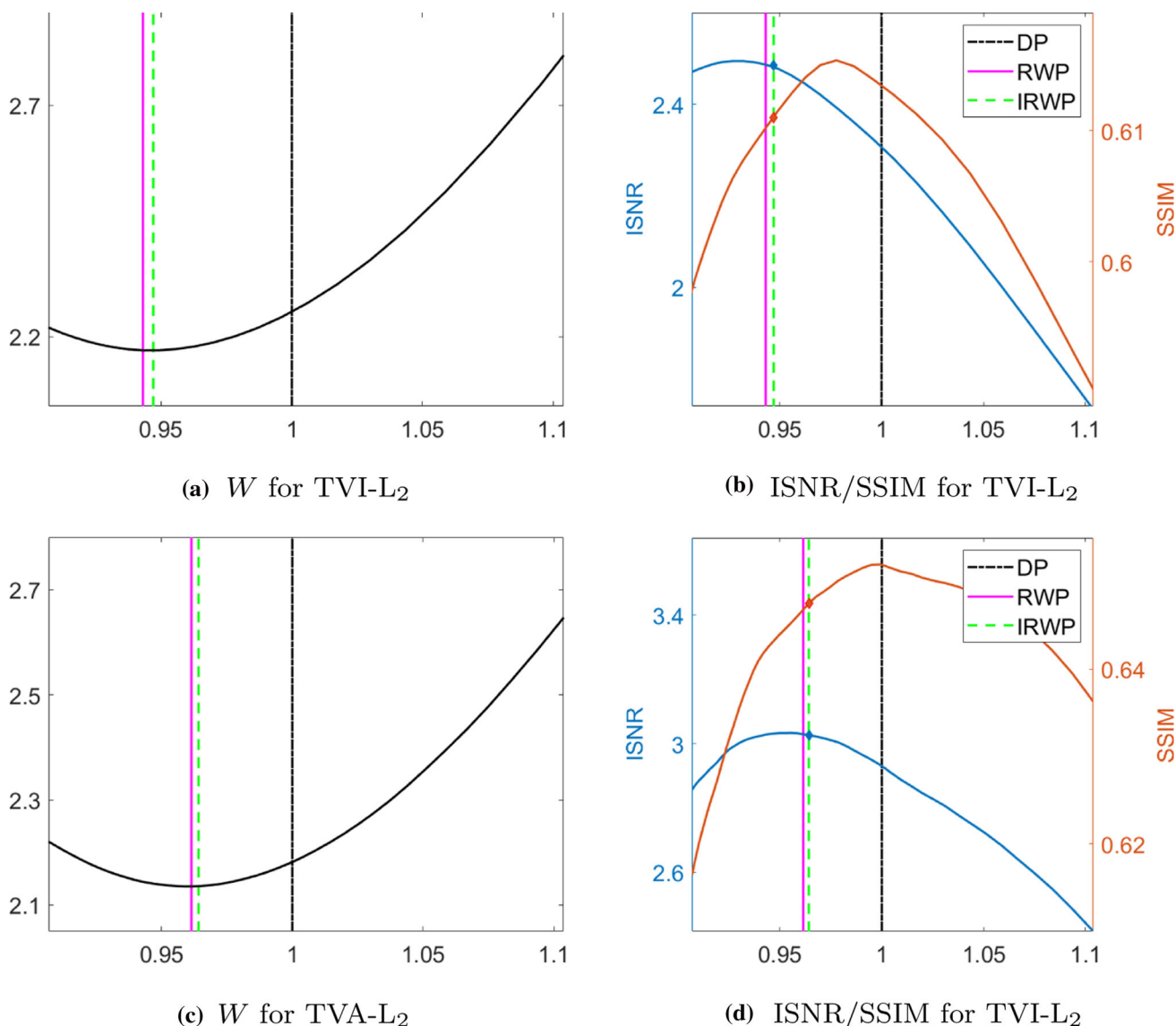


Fig. 2 Results for test image `qrcode`, Fig. 3. First column: whiteness measure functions for the TVI- L_2 and the TVA- L_2 variational models. Second column: ISNR/SSIM values for different values of τs

Table 1 PSNR, ISNR and SSIM values achieved by the bicubic interpolation and the considered variational models for which the proposed RWP-based procedure has been adopted

	band=9, sigma=2, $\sigma = 0.05$			band=13, sigma=3, $\sigma = 0.1$		
	PSNR	ISNR	SSIM	PSNR	ISNR	SSIM
<code>qrcode</code>						
x_{BIC}^*	14.1919	–	0.4334	12.4708	–	0.3129
x_{TIK}^*	14.9034	0.7115	0.4458	13.5492	1.0784	0.3720
x_{TVI}^*	18.3906	4.1987	0.7705	14.9433	2.4724	0.6123
x_{TVA}^*	19.5895	5.3976	0.8100	15.4972	3.0264	0.6475
<code>geometric</code>						
x_{BIC}^*	16.8617	–	0.3687	14.8191	–	0.2292
x_{TIK}^*	16.7718	–0.0900	0.2696	15.5181	0.6990	0.2834
x_{TV}^*	20.4136	3.5519	0.7664	17.0677	2.2486	0.5861

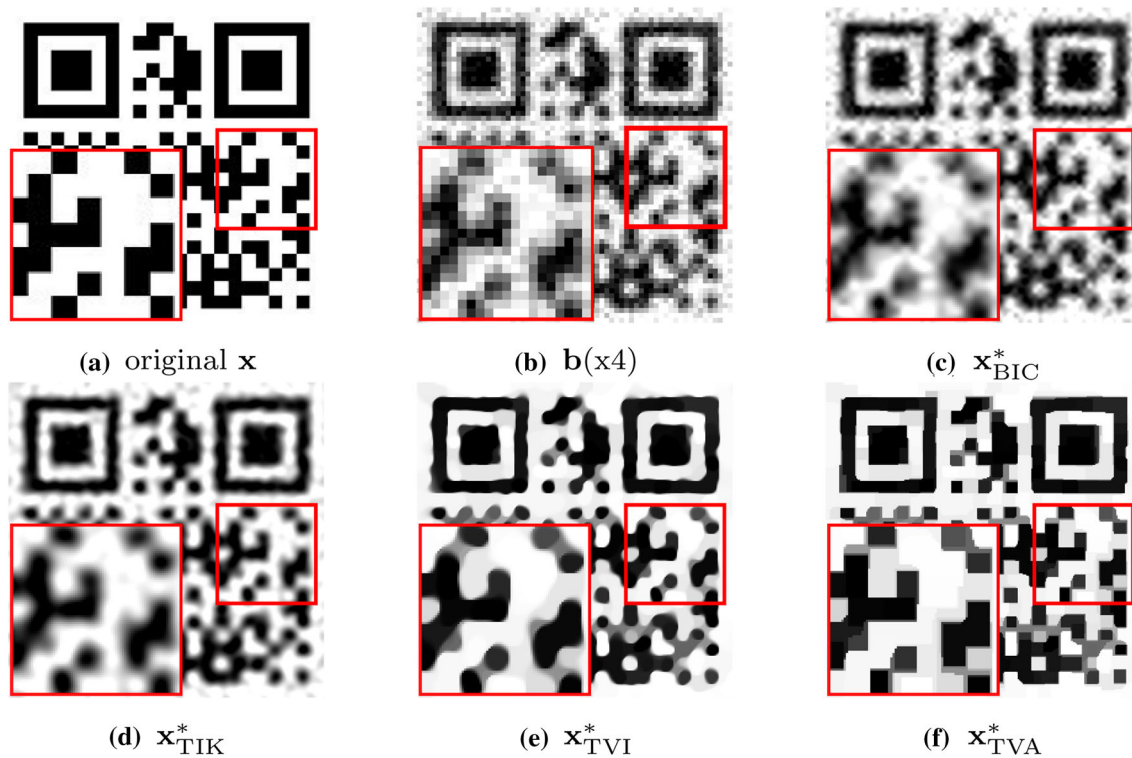


Fig. 3 Original test image *qr_code* (256×256) (a), observed image *b* (64×64) (b), reconstruction via bicubic interpolation (c), and reconstructions obtained by applying the RWP-based procedure to the TIK- L_2 (d), the TVI- L_2 (e) and the TVA- L_2 (f) variational models

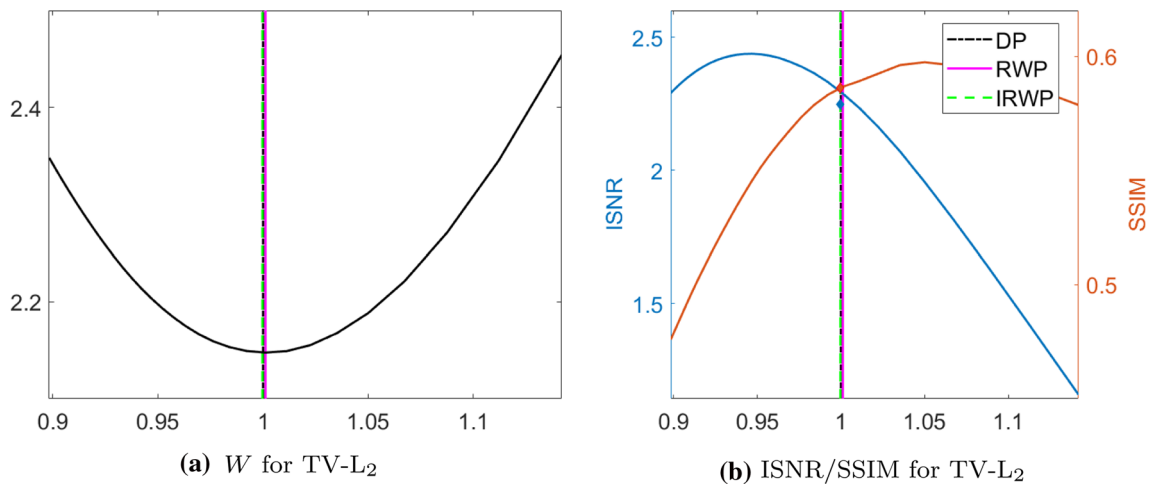


Fig. 4 Test image *geometric*. Whiteness measure function for the TV- L_2 model (first column) and ISNR/SSIM values for different τ s (second column)

The residual whiteness function and the ISNR and SSIM curves for the TV- L_2 model are shown in Fig. 4a, b, respectively. Also in this case, the W function exhibits a global minimiser over the considered domain for τ and the τ selected by RWP and IRWP are very close to each other and return high-quality reconstructions in terms of ISNR/SSIM metrics. We also notice that the τ values selected by RWP

and IRWP are in this case very close to the one selected by DP, the only difference being that RWP-based approaches do not require prior knowledge of the noise standard deviation.

The quality indices of the restorations computed via the IRWP-ADMM approach are reported in Table 1, while the corresponding reconstructions images are shown in Fig. 5.

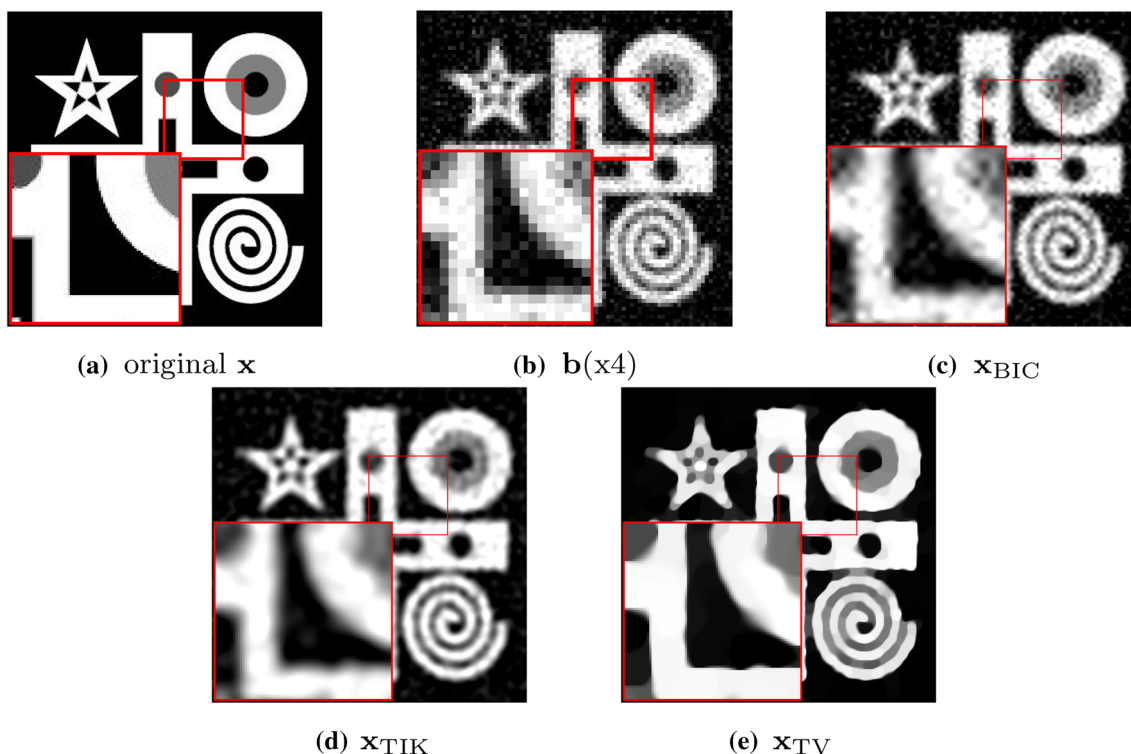


Fig. 5 Original test image *geometric* (320×320) (a), observed image **b** (80×80) (b), reconstruction via bicubic interpolation (c), and reconstructions obtained by applying the RWP-based procedure to the TIK- L_2 (d) and the TV- L_2 (e) variational models

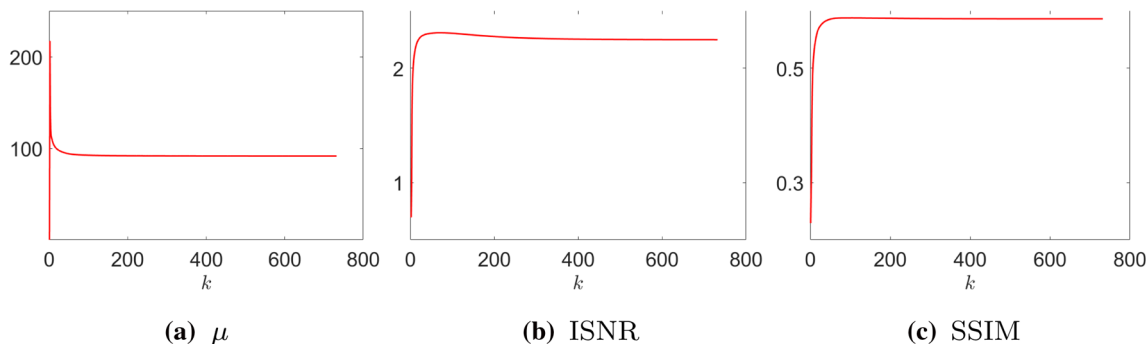


Fig. 6 Convergence plots for the IRWP-ADMM approach outlined in Algorithm 2 applied to restoring the test image *geometric* via the TV- L_2 variational model

For this second example, we also show the behaviour of the regularisation parameter μ , of the ISNR and of the SSIM along the iterations of the IRWP-ADMM for the TV- L_2 variational model - see Fig. 6a–c, respectively.

Finally, for the two test images *qr code* and *geometric*, we report in Table 1 the PSNR/ISNR/SSIM achieved by the bicubic interpolation and the considered variational models when applying a less severe noise degradation to the original images (left side of the table).

7.2 IRWP-ADMM for WTV Regularisation

We are now testing the IRWP-ADMM for the WTV- L_2 variational model employed for the reconstruction of natural images. First, we consider the test images *church* (480×320) and *monarch* (512×512) both corrupted by Gaussian blur with `band=13` and `sigma = 3`, decimated with factors $d_c = d_r = 2$ and further degraded by an AWGN with $\sigma = 0.1$. The original uncorrupted images are shown in Figs. 8a and 9a, while the acquired data are shown in Figs. 8b and 9b, respectively. We show the behaviour of the residual whiteness measure for the two test images in the first column of Fig. 7. Notice that, as already highlighted in [22],

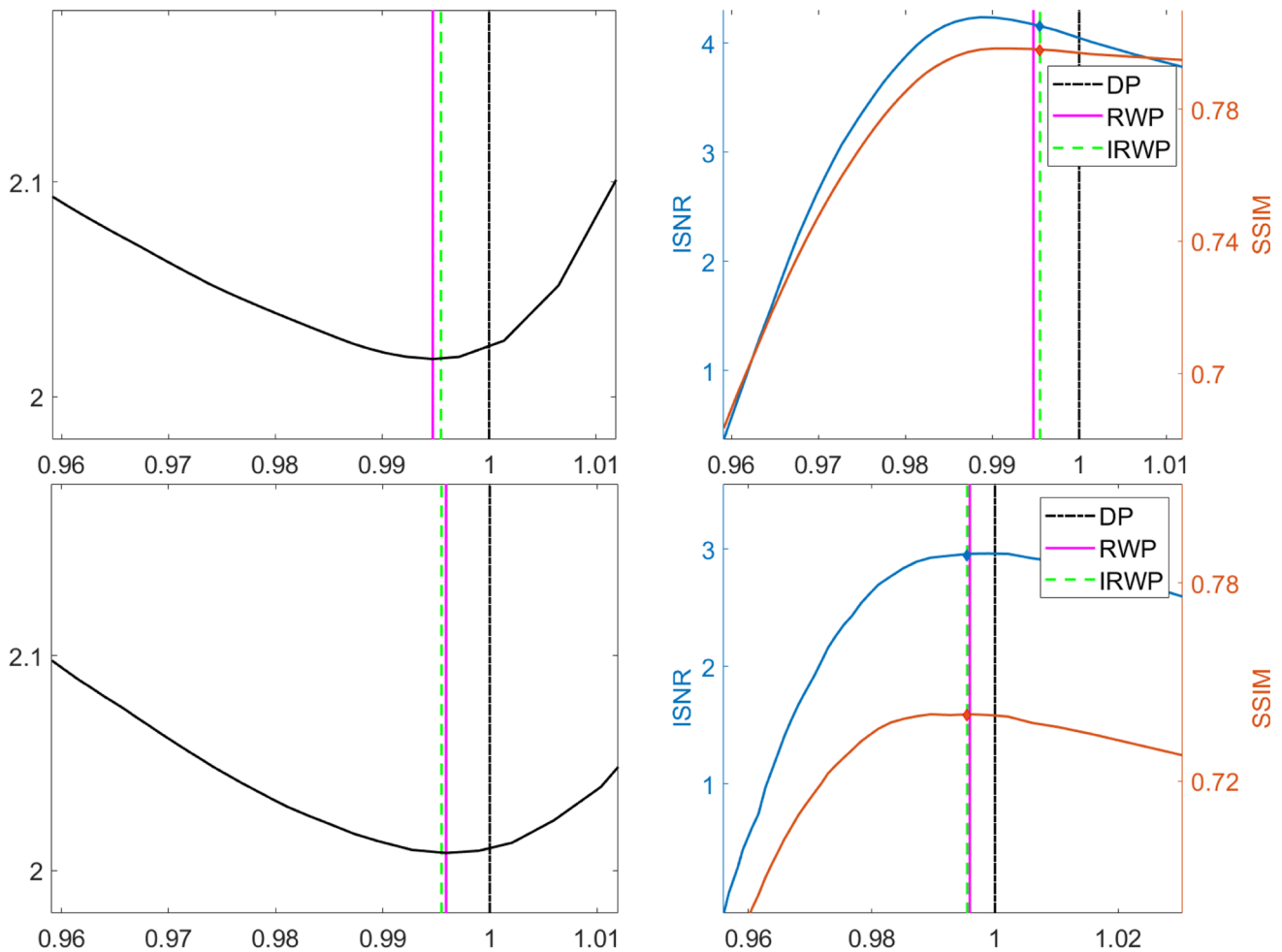


Fig. 7 Whitens measure functions for the WTV- L_2 variational model (first column) and ISNR/SSIM values for different τ 's (second column) for the test image church (top row) and the test image monarch (bottom row)

the adoption of a more flexible regularisation term yields that the ISNR and the SSIM achieve their maximum value at approximately the same τ . In both cases, the IRWP-ADMM for the solution of the WTV- L_2 returns a $\tau^*(\mu)$ very close to the one selected by the RWP; moreover, the ISNR/SSIM values corresponding to the output τ^* 's are close to the optimal ones and, in any case, larger than the one achieved by means of the DP.

The image reconstructed via bicubic interpolation, the initial guess computed by the TIK- L_2 model coupled with the RWP and the final reconstructions obtained by employing the IRWP-ADMM for the WTV- L_2 model are shown in Figs. 8 and 9 for the test image church and monarch, respectively.

Moreover, for the test image church, we also report in Fig. 10 the convergence plots of the regularisation parameter μ , the ISNR and the SSIM along the iterations of the IRWP-ADMM.

Finally, the PSNR/ISNR/SSIM values achieved for the reconstructions shown in Figs. 8 and 9 and for the ones obtained considering a less severe corruption level are reported in Table 2.

7.3 IRWP-IRL1 for CEL0 Regularisation

In this last example, we consider the problem of molecule localisation starting from a downsampled, blurred and noisy microscopy image. In the test image molecules shown in Fig. 12a, the samples are represented by sparse point sources. The original image has been corrupted by Gaussian blur with parameters `band=13` and `sigma=3`, then decimated with factors $d_c = d_r = 2$, and finally degraded by adding an AWGN realisation with σ equal to the 2% of the noiseless signal. The acquired image is shown in Fig. 12b.

In Fig. 11, we show the behaviour of the residual whiteness measure (left) and of the Jaccard index J_4 (right). Also in this example, the W function exhibits a global minimiser and

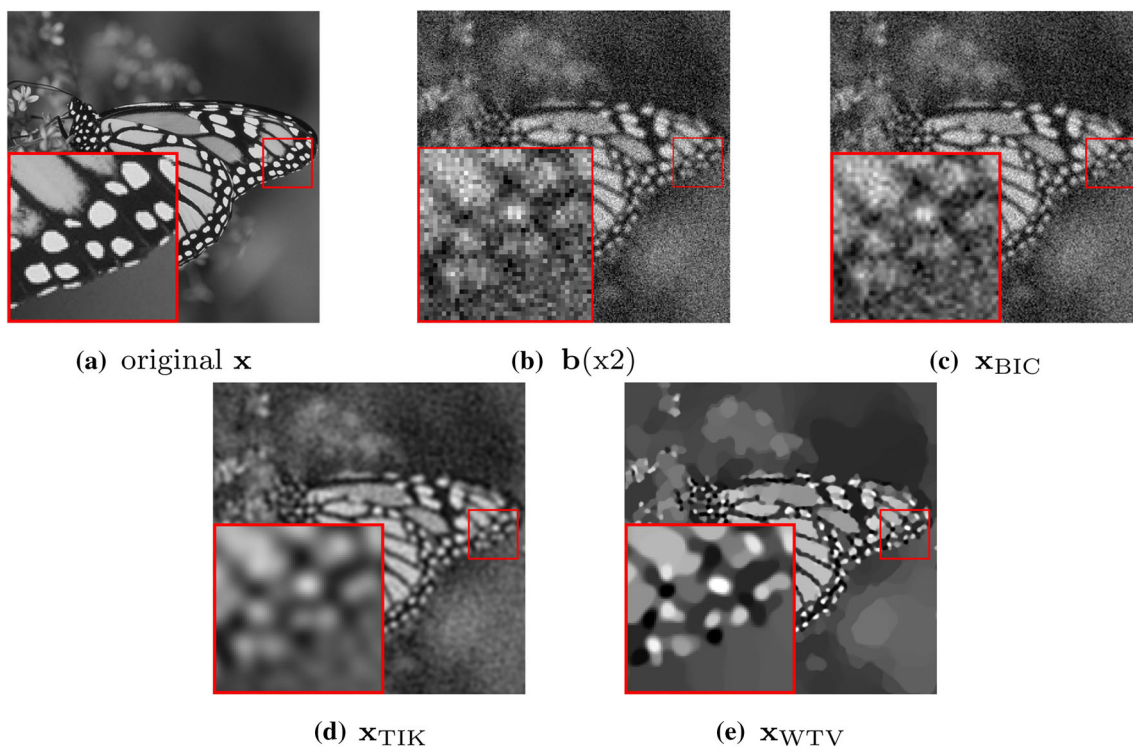


Fig. 8 Original test image monarch (520×520) (a), observed image \mathbf{b} (260×260) (b), reconstruction via bicubic interpolation (c) and reconstructions obtained by applying the RWP-based procedure to the TIK- L_2 (d), and the WTV- L_2 (e) variational models

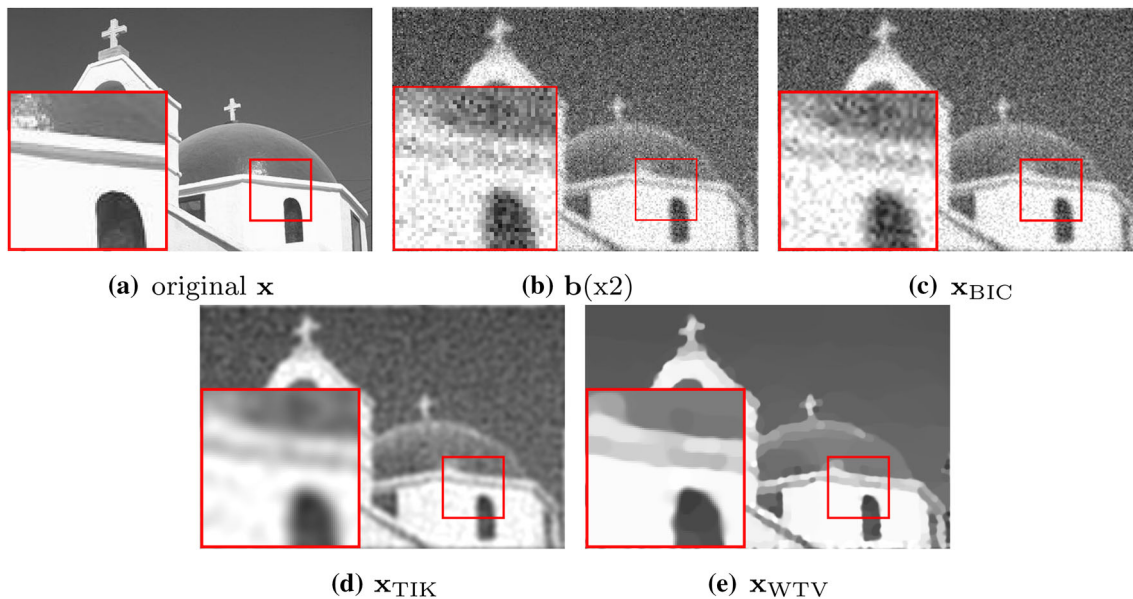


Fig. 9 Original test image church (480×320) (a), observed image \mathbf{b} (240×160) (b), reconstruction via bicubic interpolation (c) and reconstructions obtained by applying the RWP-based procedure to the TIK- L_2 (d), and the WTV- L_2 (e) variational models

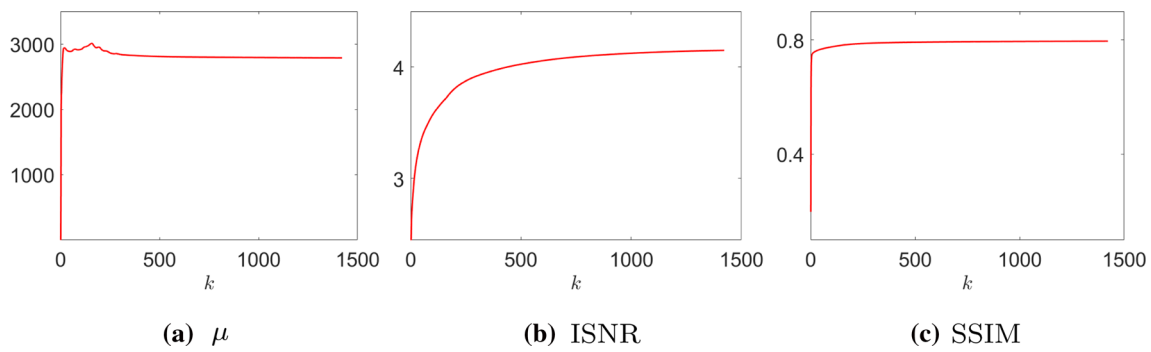


Fig. 10 Convergence plots for the proposed IRWP-ADMM approach outlined in Algorithm 2 applied to restoring the test image *church* via the WTV- L_2 variational model

Table 2 PSNR, ISNR and SSIM values achieved by the bicubic interpolation and the considered variational models for which the proposed RWP-based procedure has been adopted

	band=9, sigma=2, $\sigma=0.05$			band=13, sigma=3, $\sigma=0.1$		
	PSNR	ISNR	SSIM	PSNR	ISNR	SSIM
<i>church</i>						
\mathbf{x}_{BIC}	22.8363	–	0.4378	19.1949	–	0.1988
\mathbf{x}_{TIK}	23.7592	0.9229	0.6739	21.6448	2.4499	0.6565
\mathbf{x}_{WTV}	26.4215	3.5851	0.8592	23.3450	4.1501	0.7979
<i>monarch</i>						
\mathbf{x}_{BIC}	21.9537	–	0.5008	18.3283	–	0.2397
\mathbf{x}_{TIK}	22.8309	0.8772	0.7321	20.4384	2.1101	0.6686
\mathbf{x}_{WTV}	24.4333	2.4797	0.8231	21.2729	2.9446	0.7399

the τ^* values selected by RWP and IRWP are very close and slightly larger than $\tau^* = 1$ corresponding to DP. Unlike the previously considered quality measures, the Jaccard index does not present a *smooth* behaviour, as it measures the precision of the molecules localisation rather than some global visual properties such as the ISNR and the SSIM. However, the J_4 value selected by the IRWP-IRL1 is closer to the achieved maximum when compared to the DP.

The output reconstruction is shown in Fig. 12, together with the initialisation for the IRWP-IRL1 algorithm computed by employing the IRWP-ADMM for solving the L_1 - L_2 variational model.

A few more insights about the performance of the proposed approach are given in Table 3, where we report the Jaccard indices J_δ , $\delta \in \{0, 2, 4\}$, for the reconstruction obtained by the IRWP-ADMM for the L_1 - L_2 model and by the IRWP-IRL1 for the CEL0 model (right). In the Table, we also show the Jaccard indices achieved when applying a lower degradation level (left).

We conclude by showing, for the most severe corruption, the behaviour of the regularisation parameter μ and of the Jaccard indices along the outer iterations of Algorithm 3. One can observe that, although the monitored quantities do not exhibit a smooth nor a monotonic behaviour (as a further consequence of the definition of J_d), they stabilise thus reflecting the empirical convergence of the scheme (Fig. 13).

8 Conclusions

We proposed an automatic selection strategy for the regularisation parameter of single image super-resolution variational models based on the Residual Whiteness Principle applied along the iterations of an ADMM-based optimisation scheme. Such approach proved to be successfully applicable to several scenarios of highly degraded images by means of a large family of convex variational models, among which we performed the TIK- L_2 , the TV- L_2 and the WTV- L_2 model. Moreover, we generalised the proposed approach so as to effectively deal with a specific class of non-convex variational models, i.e. the ones admitting a convex majorant. In particular, we focused on the case of the CEL0 functional whose minimisation is carried out by the IRL1 strategy coupled with the proposed automatic estimation approach. When compared to standard parameter estimation techniques such as the Discrepancy Principle, our method has been shown to outperform and to be more applicable as it does not require any prior knowledge on the noise variance. The empirical results on the convergence of the nested iterative schemes employed reflect the robustness of our method. A more detailed theoretical study on this matter is left for future research.

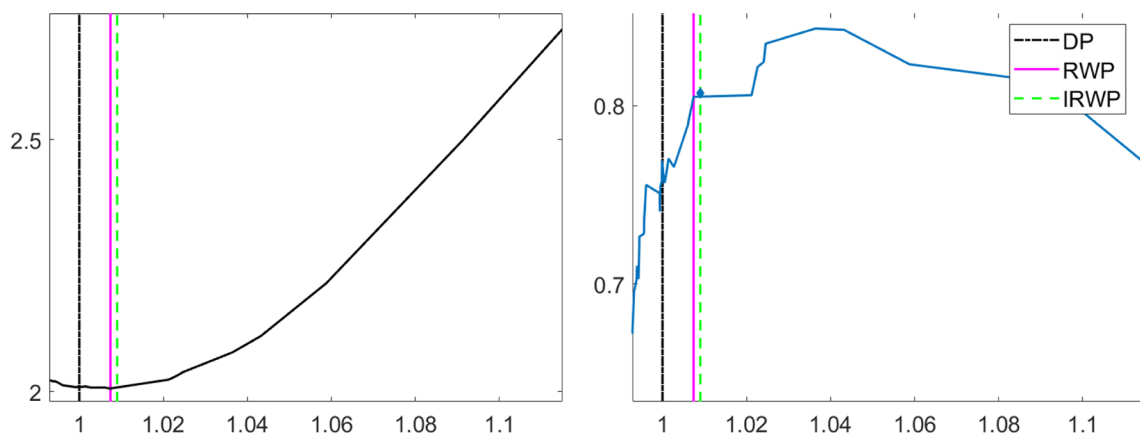


Fig. 11 Whiteness measure functions for the IRWP-IRL1 (first column) and values of J_4 for different τ s (second column) for the test image molecules

Fig. 12 Original test image molecules (256×256) (a), observed image **b** (128×128) (b), and reconstructions obtained by applying the RWP-based procedure to the L_1 - L_2 (c) and the CEL0- L_2 (d) variational models

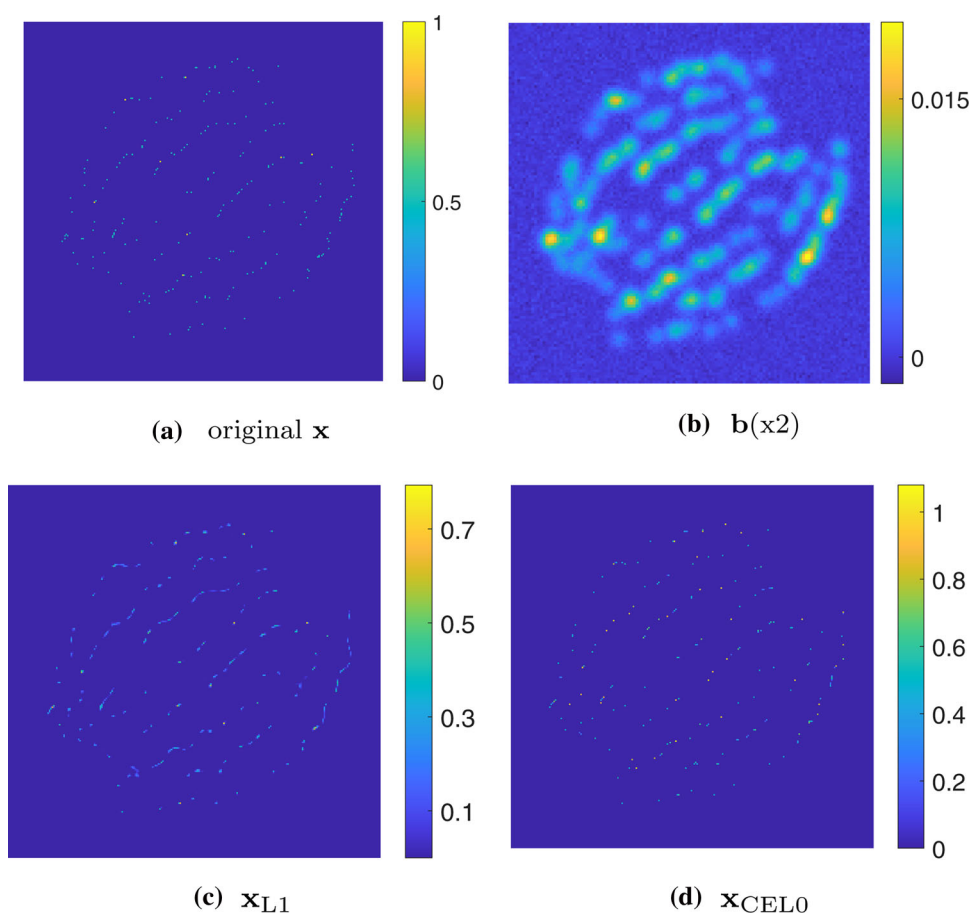
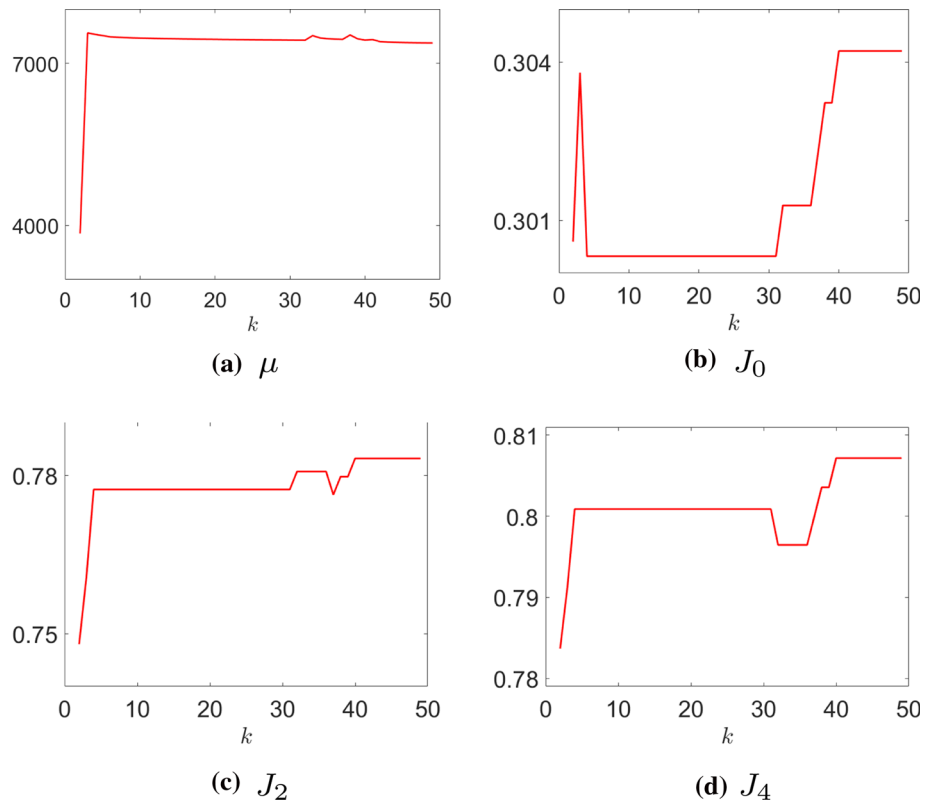


Table 3 Jaccard indices achieved by the L_1 - L_2 and the CEL0- L_2 variational models, for which the proposed RWP-based procedure has been adopted

	band=9, sigma=2, %σ = 1			band=13, sigma=3, %σ = 2		
	J_0	J_2	J_4	J_0	J_2	J_4
x_{CEL0}	0.9951	0.9951	0.9951	0.3042	0.7832	0.8072
x_{L1}	0.6126	0.6126	0.6126	0.2451	0.2957	0.2957

Fig. 13 Convergence plots for the proposed IRWP-IRL1 approach outlined in Algorithm 3 applied to restoring the test image molecules via the CEL0-L₂ variational model



Acknowledgements Research of AL, MP, FS was supported by the “National Group for Scientific Computation (GNCS-INDAM)” and by ex60 project by the University of Bologna “Funds for selected research topics”. LC acknowledges the support received by the EU H2020 RISE NoMADS, GA 777826, the UCA JEDI IDEX grant DEP “Attractivité du territoire” and the GdR ISIS project SPLIN.

Funding Open access funding provided by Università degli Studi di Napoli Federico II within the CRUI-CARE Agreement.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

A Proofs of the Results

Proof of Proposition 4 Recalling that $\mathbf{S}\mathbf{S}^H = \mathbf{I}_n$, the LR residual $\mathbf{r}^*(\mu)$ can be equivalently rewritten as

$$\mathbf{r}^*(\mu) = \mathbf{S}\mathbf{K}\mathbf{x}^*(\mu) - \mathbf{b} = \mathbf{S}\mathbf{K}\mathbf{x}^*(\mu) - \mathbf{S}\mathbf{S}^H\mathbf{b} = \mathbf{S}\mathbf{r}_H^*(\mu), \tag{53}$$

where $\mathbf{r}_H^*(\mu) = \mathbf{K}\mathbf{x}^*(\mu) - \mathbf{b}_H$ is the HR residual with $\mathbf{b}_H = \mathbf{S}^H\mathbf{b}$ the zero-interpolated, HR observation. The denominator in (25) can be thus expressed as follows

$$\begin{aligned} \|\mathbf{r}^*(\mu)\|_2^4 &= \|\mathbf{S}\mathbf{r}_H^*(\mu)\|_2^4 = \|\mathbf{S}^H\mathbf{S}\mathbf{r}_H^*(\mu)\|_2^4 \\ &= \|\mathbf{F}^H(\mathbf{F}\mathbf{S}^H\mathbf{S}\mathbf{F}^H)\mathbf{F}\mathbf{r}_H^*(\mu)\|_2^4 \end{aligned} \tag{54}$$

$$= \|\mathbf{F}^H\widetilde{\mathbf{S}}\widetilde{\mathbf{S}}^H\mathbf{F}\mathbf{r}_H^*(\mu)\|_2^4 = \|\widetilde{\mathbf{S}}\widetilde{\mathbf{S}}^H\mathbf{r}_H^*(\mu)\|_2^4, \tag{55}$$

where, in (54), the first equality comes from (53), the second from recalling that \mathbf{S}^H interpolates $\mathbf{S}\mathbf{r}_H^*(\mu)$ with zeros giving null contribution to the norm, the third from $\mathbf{F}^H\mathbf{F} = \mathbf{I}_N$, and where the last equality in (55) follows from Parseval’s theorem. Then, recalling that $\mathbf{P}^T\mathbf{P} = \mathbf{I}_N$ and $\|\mathbf{P}z\|_2 = \|z\|_2$ for any permutation matrix $\mathbf{P} \in \mathbb{R}^{N \times N}$ and any vector $z \in \mathbb{C}^N$, based on Proposition 1, the expression in (55) can be equivalently rewritten as

$$\begin{aligned} \|\mathbf{r}^*(\mu)\|_2^4 &= \left\| \left(\mathbf{P}\widetilde{\mathbf{S}}\widetilde{\mathbf{S}}^H\mathbf{P}^T \right) \mathbf{P}\mathbf{r}_H^*(\mu) \right\|_2^4 \\ &= \frac{1}{d^4} \left\| (\mathbf{I}_n \otimes \mathbf{J}_d) \hat{\mathbf{r}}_H^*(\mu) \right\|_2^4, \end{aligned}$$

where

$$\begin{aligned}
 (\mathbf{I}_n \otimes \mathbf{J}_d \hat{\mathbf{r}}_H^*(\mu))_i &= \sum_{j=0}^{d-1} (\hat{\mathbf{r}}_H^*(\mu))_{i+j}, \\
 &\text{with } \iota := 1 + \lfloor \frac{i-1}{d} \rfloor d,
 \end{aligned}$$

for every $i = 1, \dots, N$. The denominator in (25) can be thus expressed as

$$\|\mathbf{r}^*(\mu)\|_2^4 = \frac{1}{d^4} \left(\sum_{i=1}^N \left| \sum_{j=0}^{d-1} (\hat{\mathbf{r}}_H^*(\mu))_{i+j} \right|^2 \right)^2.$$

Let us now consider the numerator of the function $W(\mu)$ in (25), which, based on the definitions of auto-correlation given in (20) and of \mathbf{S}^H , reads

$$\begin{aligned}
 \|\mathbf{r}^*(\mu) \star \mathbf{r}^*(\mu)\|_2^2 &= \|\mathbf{S} \mathbf{r}_H^*(\mu) \star \mathbf{S}^H \mathbf{r}_H^*(\mu)\|_2^2 \\
 &= \|\mathbf{S}^H \mathbf{S} \mathbf{r}_H^*(\mu) \star \mathbf{S}^H \mathbf{r}_H^*(\mu)\|_2^2.
 \end{aligned}$$

By applying again the Parseval’s theorem and the convolution theorem, we get

$$\begin{aligned}
 \|\mathbf{r}^*(\mu) \star \mathbf{r}^*(\mu)\|_2^2 &= \|\mathbf{F} \left((\mathbf{S}^H \mathbf{S} \mathbf{r}_H^*(\mu)) \star (\mathbf{S}^H \mathbf{r}_H^*(\mu)) \right)\|_2^2 \\
 &= \|\mathbf{F} (\mathbf{S}^H \mathbf{S} \mathbf{r}_H^*(\mu)) \odot \overline{\mathbf{F} (\mathbf{S}^H \mathbf{r}_H^*(\mu))}\|_2^2 \\
 &= \|\mathbf{F} (\mathbf{S}^H \mathbf{S}) \mathbf{F}^H \mathbf{F} \mathbf{r}_H^*(\mu) \\
 &\quad \odot \overline{\mathbf{F} (\mathbf{S}^H \mathbf{S}) \mathbf{F}^H \mathbf{F} \mathbf{r}_H^*(\mu)}\|_2^2, \tag{56}
 \end{aligned}$$

where \odot denotes the Hadamard matrix product operator. The expression in (56) is manipulated by applying Lemma 4 and the permutation in (10), so as to give

$$\|\mathbf{r}^*(\mu) \star \mathbf{r}^*(\mu)\|_2^2 = \frac{1}{d^4} \sum_{i=1}^N \left| \sum_{j=0}^{d-1} (\hat{\mathbf{r}}_H^*(\mu))_{i+j} \right|^4. \tag{57}$$

Finally, plugging (57) and (55) into (25), we get the following form for the whiteness measure $W(\mu)$ for a super-resolution problem

$$\begin{aligned}
 W(\mu) &= \left(\sum_{i=1}^N |w_i(\mu)|^4 \right) / \left(\sum_{i=1}^N |w_i(\mu)|^2 \right)^2, \\
 w_i(\mu) &= \sum_{j=0}^{d-1} (\hat{\mathbf{r}}_H^*(\mu))_{i+j}. \tag{58}
 \end{aligned}$$

Proof of Proposition 5 We impose a first-order optimality condition on the cost function in (4) with respect to \mathbf{x} , thus getting:

$$\mathbf{x}^*(\mu) = (\mu(\mathbf{S}\mathbf{K})^H(\mathbf{S}\mathbf{K}) + \mathbf{L}^H\mathbf{L})^{-1}(\mu(\mathbf{S}\mathbf{K})^H\mathbf{b} + \mathbf{L}^H\mathbf{v}),$$

which can be manipulated in terms of \mathbf{F} and \mathbf{F}^H to deduce

$$\begin{aligned}
 \mathbf{x}^*(\mu) &= (\mu \mathbf{F}^H \mathbf{F} \mathbf{K}^H \mathbf{F}^H \mathbf{F} \mathbf{S}^H \mathbf{S} \mathbf{F}^H \mathbf{F} \mathbf{K} \mathbf{F}^H \mathbf{F} \\
 &\quad + \mathbf{F}^H \mathbf{F} \mathbf{L}^T \mathbf{L} \mathbf{F}^H \mathbf{F})^{-1} (\mu \mathbf{K}^H \mathbf{S}^H \mathbf{b} + \mathbf{L}^H \mathbf{v}) \\
 &= (\mu \mathbf{F}^H \boldsymbol{\lambda}^H (\mathbf{F} \mathbf{S}^H \mathbf{S} \mathbf{F}^H) \boldsymbol{\lambda} \mathbf{F} + \mathbf{F}^H \sum_{j=1}^s \boldsymbol{\Gamma}_j^H \boldsymbol{\Gamma}_j \mathbf{F})^{-1} \\
 &\quad (\mu \mathbf{K}^H \mathbf{S}^H \mathbf{b} + \mathbf{L}^H \mathbf{v}), \tag{59}
 \end{aligned}$$

with $\boldsymbol{\lambda}$, $\boldsymbol{\Gamma}_j$ defined in (19). From Proposition 1 we have that $\mathbf{F} \mathbf{S}^H \mathbf{S} \mathbf{F}^H = \mathbf{P}^T (\mathbf{I}_n \otimes \mathbf{J}_d) \mathbf{P}$, hence (59) becomes:

$$\begin{aligned}
 \mathbf{x}^*(\mu) &= \left(\frac{\mu}{d} \mathbf{F}^H \boldsymbol{\lambda}^H \mathbf{P}^T (\mathbf{I}_n \otimes \mathbf{J}_d) \mathbf{P} \mathbf{3} \mathbf{F} + \mathbf{F}^H \sum_{j=1}^s \boldsymbol{\Gamma}_j^H \boldsymbol{\Gamma}_j \mathbf{F} \right)^{-1} \\
 &\quad \left(\mu \mathbf{K}^H \mathbf{S}^H \mathbf{b} + \sum_{j=1}^s \mathbf{L}_j^H \mathbf{v}_j \right) \\
 &= \mathbf{F}^H \left(\frac{\mu}{d} \boldsymbol{\lambda}^H \mathbf{P}^T (\mathbf{I}_n \otimes \mathbf{J}_d) \mathbf{P} \mathbf{3} + \sum_{j=1}^s \boldsymbol{\Gamma}_j^H \boldsymbol{\Gamma}_j \right)^{-1} \\
 &\quad \mathbf{F} \left(\mu \mathbf{K}^H \mathbf{F}^H \mathbf{F} \mathbf{S}^H \mathbf{b} + \sum_{j=1}^s \mathbf{L}_j^H \mathbf{F}^H \mathbf{F} \mathbf{v}_j \right) \\
 &= \mathbf{F}^H \left(\frac{\mu}{d} \boldsymbol{\lambda}^H \mathbf{P}^T (\mathbf{I}_n \otimes \mathbf{J}_d) \mathbf{P} \mathbf{3} + \sum_{j=1}^s \boldsymbol{\Gamma}_j^H \boldsymbol{\Gamma}_j \right)^{-1} \\
 &\quad \left(\mu \boldsymbol{\lambda}^H \tilde{\mathbf{b}}_H + \sum_{j=1}^s \boldsymbol{\Gamma}_j^H \tilde{\mathbf{v}}_j \right), \tag{60}
 \end{aligned}$$

where $\mathbf{v} = (\mathbf{v}_1^T, \dots, \mathbf{v}_s^T)^T$, and $\tilde{\mathbf{b}}_H = \mathbf{F} \mathbf{b}_H = \mathbf{F} \mathbf{S}^H \mathbf{b}$ contains d replication of $\tilde{\mathbf{b}}$ —see, e.g., [28]. We now introduce the following operators

$$\square \quad \underline{\boldsymbol{\lambda}} := (\mathbf{I}_n \otimes \mathbf{1}_d) \mathbf{P} \boldsymbol{\lambda} \quad \underline{\boldsymbol{\lambda}}^H := \boldsymbol{\lambda}^H \mathbf{P}^T (\mathbf{I}_n \otimes \mathbf{1}_d) \tag{61}$$

where $\mathbf{1}_d \in \mathbb{R}^d$ is a vector of ones. In compact form, equation (60) reads:

$$\mathbf{x}^*(\mu) = \mathbf{F}^H \left(\frac{\mu}{d} \underline{\lambda}^H \underline{\lambda} + \sum_{j=1}^s \Gamma_j^H \Gamma_j \right)^{-1} \left(\mu \underline{\lambda}^H \tilde{\mathbf{b}}_H + \sum_{j=1}^s \Gamma_j^H \tilde{\mathbf{v}}_j \right). \tag{62}$$

Proceeding as in [43], we can now apply the Woodbury formula (7) and perform few manipulations, so as to obtain that the expression in (62) becomes:

$$\mathbf{x}^*(\mu) = \mathbf{F}^H \left[\underline{\Psi} - \mu \underline{\Psi} \underline{\lambda}^H \left(d\mathbf{I} + \mu \underline{\lambda} \underline{\Psi} \underline{\lambda}^H \right)^{-1} \underline{\lambda} \underline{\Psi} \right] \left(\mu \underline{\lambda}^H \tilde{\mathbf{b}}_H + \sum_{j=1}^s \Gamma_j^H \tilde{\mathbf{v}}_j \right), \tag{63}$$

where $\underline{\Psi} = \left(\sum_{j=1}^s \Gamma_j^H \Gamma_j + \epsilon \right)^{-1}$ and the parameter $0 < \epsilon \ll 1$ guarantees the inversion of $\sum_{j=1}^s \Gamma_j^H \Gamma_j$. \square

Proof of Proposition 6 Recalling property (5) in Lemma 1, we get the following chain of equalities

$$\begin{aligned} \underline{\lambda}^H \underline{\Phi} \underline{\lambda} &= \underline{\lambda}^H \mathbf{P}^T (\mathbf{I}_n \otimes \mathbf{1}_d) \underline{\Phi} (\mathbf{I}_n \otimes \mathbf{1}_d^T) \mathbf{P} \mathbf{3} \\ &= \underline{\lambda}^H \mathbf{P}^T (\mathbf{I}_n \otimes \mathbf{1}_d) (\underline{\Phi} \otimes \mathbf{1}_d^T) \mathbf{P} \mathbf{3} \\ &= \underline{\lambda}^H \mathbf{P}^T (\mathbf{I}_n \underline{\Phi} \otimes \mathbf{1}_d \mathbf{1}_d^T) \mathbf{P} \mathbf{3} = \underline{\lambda}^H \mathbf{P}^T (\underline{\Phi} \mathbf{I}_n \otimes \mathbf{J}_d) \mathbf{P} \mathbf{3} \\ &= \underline{\lambda}^H \mathbf{P}^T (\underline{\Phi} \mathbf{I}_n \otimes \mathbf{I}_d \mathbf{J}_d) \mathbf{P} \mathbf{3} = \underline{\lambda}^H \mathbf{P}^T (\underline{\Phi} \otimes \mathbf{I}_d) (\mathbf{I}_n \otimes \mathbf{J}_d) \mathbf{P} \mathbf{3} \\ &= \underline{\lambda}^H \mathbf{P}^T (\underline{\Phi} \otimes \mathbf{I}_d) \mathbf{P} \mathbf{P}^T (\mathbf{I}_n \otimes \mathbf{J}_d) \mathbf{P} \mathbf{3}, \end{aligned}$$

where the sparse block-diagonal matrix $\mathbf{P}^T (\underline{\Phi} \otimes \mathbf{I}_d) \mathbf{P} \in \mathbb{R}^{N \times N}$ commutes with $\underline{\lambda}^H$, so that $\underline{\lambda}^H \mathbf{P}^T (\underline{\Phi} \otimes \mathbf{I}_d) \mathbf{P} = \mathbf{P}^T (\underline{\Phi} \otimes \mathbf{I}_d) \mathbf{P} \underline{\lambda}^H$. Recalling (61), this yields:

$$\underline{\lambda}^H \underline{\Phi} \underline{\lambda} = \mathbf{P}^T (\underline{\Phi} \otimes \mathbf{I}_d) \mathbf{P} \underline{\lambda}^H \underline{\lambda},$$

which completes the proof. \square

Proof of Corollary 1 We first notice that

$$\underline{\lambda} \underline{\Psi} \underline{\lambda}^H = (\mathbf{I}_n \otimes \mathbf{1}_d^T) \widehat{\underline{\Lambda} \underline{\Psi} \underline{\Lambda}^H} (\mathbf{I}_n \otimes \mathbf{1}_d), \tag{64}$$

is diagonal as $\widehat{\underline{\Lambda} \underline{\Psi} \underline{\Lambda}^H} = \mathbf{P} \mathbf{3} \mathbf{3}^H \mathbf{P}^T$ is. The matrix in (64) can thus be written as

$$\underline{\lambda} \underline{\Psi} \underline{\lambda}^H = \text{diag}(\omega_1, \dots, \omega_n), \quad \omega_i = \sum_{\ell=0}^{d-1} \frac{|\hat{\lambda}_{i+\ell}|^2}{\zeta_{i+\ell} + \epsilon},$$

with $\zeta_{i+\ell} = \sum_{j=1}^s |\hat{\gamma}_{j,i+\ell}|^2$.

Hence, since $\underline{\Phi}$ is the inverse of the sum of two diagonal matrices, it is diagonal so we can apply Proposition 6 and deduce the thesis. \square

Proof of Proposition 7 We start writing the explicit expression of the action of \mathbf{P} on the Fourier transform of the high resolution residual image:

$$\begin{aligned} \hat{\mathbf{r}}_H^*(\mu) &= \mu \widehat{\underline{\Lambda} \underline{\Psi} \tilde{\mathbf{b}}_H} + \widehat{\underline{\Lambda} \underline{\Psi}} \sum_{j=1}^s \check{\Gamma}_j \tilde{\mathbf{v}}_j \\ &\quad - \mu^2 \widehat{\underline{\Lambda} \underline{\Psi}} \left[(d\mathbf{I} + \mu \underline{\lambda} \underline{\Psi} \underline{\lambda}^H)^{-1} \otimes \mathbf{I}_d \right] \widehat{\underline{\lambda}^H \underline{\lambda} \underline{\Psi} \tilde{\mathbf{b}}_H} \\ &\quad - \mu \widehat{\underline{\Lambda} \underline{\Psi}} \left[(d\mathbf{I} + \mu \underline{\lambda} \underline{\Psi} \underline{\lambda}^H)^{-1} \otimes \mathbf{I}_d \right] \widehat{\underline{\lambda}^H \underline{\lambda} \underline{\Psi}} \sum_{j=1}^s \check{\Gamma}_j \tilde{\mathbf{v}}_j - \hat{\mathbf{b}}_H, \end{aligned}$$

whence we can explicitly compute the expression for each component $i = 1, \dots, n$:

$$\begin{aligned} (\hat{\mathbf{r}}_H^*(\mu))_i &= \mu \left[\frac{|\hat{\lambda}_i|^2}{\zeta_i + \epsilon} \hat{b}_{H,i} \right] + \frac{\hat{\lambda}_i \sum_{j=1}^s \tilde{\gamma}_{j,i} \tilde{v}_{j,i}}{\zeta_i + \epsilon} \\ &\quad - \mu^2 \left[\frac{|\hat{\lambda}_i|^2}{\zeta_i + \epsilon} \sum_{\ell=0}^{d-1} \frac{|\hat{\lambda}_{i+\ell}|^2 \hat{b}_{H,i+\ell}}{\zeta_{i+\ell} + \epsilon} \right] \\ &\quad - \mu \left[\frac{|\hat{\lambda}_i|^2}{\zeta_i + \epsilon} \sum_{\ell=0}^{d-1} \frac{\hat{\lambda}_{i+\ell} \sum_{j=1}^s \tilde{\gamma}_{j,i+\ell} \tilde{v}_{j,i+\ell}}{\zeta_{i+\ell} + \epsilon} \right] - \hat{b}_{H,i}. \end{aligned}$$

By easy manipulations, we get:

$$\begin{aligned}
 (\hat{\mathbf{r}}_H^*(\mu))_i = & \mu \left[\frac{|\hat{\lambda}_i|^2}{\zeta_i + \epsilon} \hat{b}_{H,i} \right] + \frac{\hat{\lambda}_i \sum_{j=1}^s \tilde{\gamma}_{j,i} \hat{v}_{j,i}}{\zeta_i + \epsilon} \\
 & - \left[\mu^2 \sum_{\ell=0}^{d-1} \frac{|\hat{\lambda}_{i+\ell}|^2 \hat{b}_{H,i+\ell}}{\zeta_{i+\ell} + \epsilon} \right. \\
 & \left. + \mu \sum_{\ell=0}^{d-1} \frac{\hat{\lambda}_{i+\ell} \sum_{j=1}^s \tilde{\gamma}_{j,i+\ell} \hat{v}_{j,i+\ell}}{\zeta_{i+\ell} + \epsilon} \right]^{-1} \\
 & \frac{|\hat{\lambda}_i|^2}{\zeta_i + \epsilon} \left(d + \mu \sum_{j=0}^{d-1} \frac{|\hat{\lambda}_{i+j}|^2}{\zeta_{i+j} + \epsilon} \right)^{-1} - \hat{b}_{H,i}.
 \end{aligned}$$

We can thus deduce the following expression of the terms in formula (58):

$$\begin{aligned}
 \sum_{\ell=0}^{d-1} (\hat{\mathbf{r}}_H^*(\mu))_{i+\ell} = & \frac{1}{d + \mu \sum_{\ell=0}^{d-1} \frac{|\hat{\lambda}_{i+\ell}|^2}{\zeta_{i+\ell} + \epsilon}} \\
 \left[\mu \left(d \sum_{\ell=0}^{d-1} \frac{|\hat{\lambda}_{i+\ell}|^2}{\zeta_{i+\ell} + \epsilon} \hat{b}_{H,i+\ell} \right. \right. \\
 & - \sum_{\ell=0}^{d-1} \hat{b}_{H,i+\ell} \sum_{\ell=0}^{d-1} \frac{|\hat{\lambda}_{i+\ell}|^2}{\zeta_{i+\ell} + \epsilon} \Big) + d \left(\sum_{\ell=0}^{d-1} \frac{\hat{\lambda}_{i+\ell} \sum_j \tilde{\gamma}_{j,i+\ell} \hat{v}_{j,i+\ell}}{\zeta_{i+\ell} + \epsilon} \right. \\
 & \left. \left. - \sum_{\ell=0}^{d-1} \hat{b}_{H,i+\ell} \right) \right]. \tag{65}
 \end{aligned}$$

In light of its replicating structure, we observe that the action of the permutation \mathbf{P} on $\tilde{\mathbf{b}}_H$ will cluster the identical entries, so that the $\hat{b}_{H,i+j}$ can be written as the mean of the set of d values $\{\hat{b}_{H,i}, \dots, \hat{b}_{H,i+d-1}\}$. This allows to simplify formula (65) as the difference in the first bracket vanishes. By now setting

$$\begin{aligned}
 \eta_i := & \frac{1}{d} \sum_{\ell=0}^{d-1} \frac{|\hat{\lambda}_{i+\ell}|^2}{\zeta_{i+\ell} + \epsilon}, \quad \varrho_i := \sum_{j=0}^{d-1} \hat{b}_{H,i+j}, \\
 v_i := & \sum_{\ell=0}^{d-1} \frac{\hat{\lambda}_{i+\ell} \sum_{j=1}^s \tilde{\gamma}_{j,i+\ell} \hat{v}_{j,i+\ell}}{\zeta_{i+\ell} + \epsilon},
 \end{aligned}$$

which can all be computed beforehand. Plugging (65) into (58) we finally get

$$W(\mu) = \left(\sum_{i=1}^N \left| \frac{v_i - \varrho_i}{1 + \eta_i \mu} \right|^4 \right) / \left(\sum_{i=1}^N \left| \frac{v_i - \varrho_i}{1 + \eta_i \mu} \right|^2 \right)^2.$$

This proves the proposition. □

References

1. Almeida, M.S.C., Figueiredo, M.A.T.: Parameter estimation for blind and non-blind deblurring using residual whiteness measures. *IEEE Trans. Image Process.* **22**, 2751–2763 (2013)
2. Baloch, G., Ozkaramanli, H., Yu, R.: Residual correlation regularization based image denoising. *IEEE Signal Process. Lett.* **25**, 298–302 (2018)
3. Boyd, S.P., Parikh, N., Chu, E., Peleato, B., Eckstein, J.: Distributed optimization and statistical learning via the alternating direction method of multipliers. *Found. Trends Mach. Learn.* **3**, 1–122 (2011)
4. Brunet, D., Vrscaj, E.R., Wang, Z.: The use of residuals in image denoising. In: *Proceedings of 6th International Conference on Image Analysis and Recognition (ICIAR 2009)*, pp. 1–12 (2009)
5. Cascarano, P., Calatroni, L., Piccolomini, E.: Efficient ℓ_0 gradient-based super-resolution for simplified image segmentation. *IEEE Trans. Comput. Imaging* **7**, 399–408 (2021)
6. Calvetti, D., Hansen, P.C., Reichel, L.: L-curve curvature bounds via Lanczos bidiagonalization. *Electron. Trans. Numer. Anal.* **14**, 20–35 (2002)
7. Chen, A.Z., Huo, B.X., Wen, C.Y.: Adaptive regularization for color image restoration using discrepancy principle. In: *ICSPCC*, vol. 2013, pp. 1–6 (2013)
8. Clason, C.: *Regularization of Inverse Problems*. Kluwer, Dordrecht (1996)
9. Calatroni, L., Lanza, A., Pragliola, M., Sgallari, F.: Adaptive parameter selection for weighted-TV image reconstruction problems. *J. Phys.: Conf. Ser.* **1476**, 012003 (2020)
10. Chan, T.F., Ng, M.K., Yau, A.C., Yip, A.M.: Super resolution image reconstruction using fast inpainting algorithms. *Appl. Comput. Harmon. A.* **23**(1), 3–24 (2007)
11. Craven, P., Wahba, G.: Smoothing noisy data with spline functions. *Numer. Math.* **31**, 377–403 (1978)
12. D’Angeli, D., Donno, A.: Shuffling matrices, Kronecker product and Discrete Fourier Transform. *Discrete Appl. Math.* **233**, 1–18 (2017)
13. Fenu, C., Reichel, L., Rodriguez, G., Sadok, H.: GCV for Tikhonov regularization by partial SVD. *BIT* **57**, 1019–1039 (2017)
14. Galbraith, C.G., Galbraith, J.A.: Super-resolution microscopy at a glance. *J. Cell Sci.* **124**(10), 1607–1611 (2011)
15. Gazagnes, S., Soubies, Emmanuel, Blanc-Féraud, L.: High density molecule localization for super-resolution microscopy using ℓ_0 based sparse approximation. In: *2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017)*, pp. 28–31 (2017)
16. Hansen, P.: Rank-deficient and discrete ill-posed problems: numerical aspects of linear inversion (1987)
17. Hansen, P.C., Kilmer, M.E., Kjeldsen, R.H.: Exploiting residual information in the parameter choice for discrete ill-posed problems. *BIT Numer. Math.* **46**, 41–59 (2006)
18. Lanza, A., Morigi, S., Sgallari, F.: Variational image restoration with constraints on noise whiteness. *J. Math. Imaging Vis.* **53**, 61–67 (2015)

19. Lanza, A., Morigi, S., Selesnick, I., Sgallari, F.: Nonconvex nonsmooth optimization via convex-nonconvex majorization-minimization. *Numerische Mathematik* **136**, 343–381 (2017)
20. Lanza, A., Morigi, S., Sciacchitano, F., Sgallari, F.: Whiteness constraints in a unified variational framework for image restoration. *J. Math. Imaging Vis.* **60**, 1503–1526 (2018)
21. Lanza, A., Morigi, S., Sgallari, F., Yezzi, A.J.: Variational image denoising based on autocorrelation whiteness. *SIAM J. Imaging Sci.* **6**, 1931–1955 (2013)
22. Lanza, A., Pragliola, M., Sgallari, F.: Residual whiteness principle for parameter-free image restoration. *Electron. Trans. Numer. Anal.* **53**, 329–351 (2020)
23. Ochs, P., Dosovitskiy, A., Brox, T., Pock, T.: On iteratively reweighted algorithms for nonsmooth nonconvex optimization in computer vision. *SIAM J. Imaging Sci.* **8**(1), 331–72 (2015)
24. Ono, S.: l_0 gradient projection. *IEEE Trans. Image Process.* **26**(4), 1554–1564 (2017)
25. Osher, S.J., Marquina, A.: Image super-resolution by TV-regularization and bregman iteration. *J. Sci. Comput.* **37**, 367–382 (2008)
26. Pragliola, M., Calatroni, L., Lanza, A., Sgallari, F.: Residual whiteness principle for automatic parameter selection in $l_2 - l_2$ image super-resolution problems. In: Elmoataz, A., Fadili, J., Quéau, Y., Rabin, J., Simon, L. (eds.) *Scale Space and Variational Methods in Computer Vision*, pp. 476–488. Springer, Cham (2021)
27. Riot, P., Almansa, A., Gousseau, Y., Tupin, F.: Penalizing local correlations in the residual improves image denoising performance. In: 24th European Signal Processing Conference (EUSIPCO 2016), pp. 1867–1871 (2016)
28. Robinson, M.D., Farsiu, S., Lo, J.Y., Milanfar, P., Toth, C.: Efficient registration of aliased x-ray images. In: *ACSSC*, pp. 215–219 (2007)
29. Rust, B.W., O’Leary, D.P.: Residual periodograms for choosing regularization parameters for ill-posed problems. *Inverse Probl.* **24**, 034005 (2008)
30. Rudin, L.I., Osher, S., Fatemi, E.: Nonlinear total variation based noise removal algorithms. *Phys. D: Nonlinear Phenomena* **60**, 259–268 (1992)
31. Reichel, L., Rodriguez, G.: Old and new parameter choice rules for discrete ill-posed problems. *Numer. Algorithms* **63**, 65–87 (2013)
32. Soubies, E., Blanc-Féraud, L., Aubert, G.: A continuous exact l_0 penalty (CELO) for least squares regularized problem. *SIAM J. Imaging Sci.* **8**(3), 1607–1639 (2015)
33. Storath, M., Weinmann, A., Demaret, L.: Jump-sparse and sparse recovery using Potts functionals. *IEEE Trans. Signal Process.* **62**(14), 3654–3666 (2014)
34. Sun, J., Xu, Z., Shum, H.Y.: Image super-resolution using gradient profile prior. In: 2008 IEEE Conference on Computer Vision and Pattern Recognition, pp. 1–8 (2008)
35. Thevenaz, P., Blu, T., Unser, M.: Image interpolation and resampling. In: *Handbook of Medical Imaging, Processing and Analysis*. Academic Press, pp. 393–420 (2000)
36. Tuador, N.K., Pham, D., Michetti, J., Basarab, A., Kouamé, D.: A novel fast 3D single image super-resolution algorithm. In: 2021 IEEE 18th International Symposium on Biomedical Imaging (ISBI), pp. 73–76 (2021)
37. Toma, A., Sixou, B., Peyrin, F.: Iterative choice of the optimal regularization parameter in TV image restoration. *Inverse Probl. Imaging* **9**, 1171 (2015)
38. Tao, M., Yang, J., He, B.: Alternating direction algorithms for total variation deconvolution in image reconstruction. TR0918, Department of Mathematics, Nanjing University (2009)
39. Wang, Z., Bovik, A., Sheikh, H.R., Simoncelli, E.P.: Image quality assessment: from error visibility to structural similarity. *IEEE Trans. Image Process.* **13**, 600–612 (2004)
40. Willett, R.M., Jermyn, I., Nowak, R. D., Zerubia, J.: Wavelet-based superresolution in astronomy. In: *ADASS XIII*, vol. 314, p. 107 (2004)
41. Yang, J., Wright, J., Huang, T.S., Ma, Y.: Image super-resolution via sparse representation. *IEEE Trans. Image Process.* **19**(11), 2861–2873 (2010)
42. Yang, W., Zhang, X., Tian, Y., Wang, W., Xue, J.H., Liao, Q.: Deep learning for single image super-resolution: a brief review. *IEEE Trans. Multi.* **21**(12), 3106–3121 (2019)
43. Zhao, N., Wei, Q., Basarab, A., Dobigeon, N., Kouamé, D., Tourneret, J.: Fast single image super-resolution using a new analytical solution for $l_2 - l_2$ problems. *IEEE Trans. Image Process.* **25**, 3683–3697 (2016)

Publisher’s Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Monica Pragliola received her Ph.D. in Pure and Applied Mathematics from the University of Bologna in 2020. After a post-doctoral fellowship at the University of Bologna, she is now a researcher at the Department of Mathematics of University of Naples Federico II. Her research interest includes numerical methods for variational approaches in image processing and Bayesian inverse problems.



Luca Calatroni is a junior research scientist at CNRS, working at the I3S laboratory of Sophia-Antipolis, France since 2019. After completing his Ph.D. in 2015 in the Cambridge Image Analysis research group, he was Marie Skłodowska-Curie post-doctoral researcher within the Nano2fun FP7 ITN at the University of Genova and Lecteur Hadamard at the Ecole Polytechnique of Palaiseau, France, funded by the FMJH. His research interests lie in the field of mathematical image processing, applied inverse problems and non-smooth optimisation with several applications to the fields of biological imaging, digital restoration of artworks and computational neurosciences.



Alessandro Lanza received the European Ph.D. degree in Information Technology from the Advanced Research Centre on Electronic Systems for Information and Communication Technologies (ARCES), University of Bologna, and, since 2019, he is senior assistant Professor of Numerical Analysis at the Department of Mathematics of the same University. He is the author of more than 50 peer-reviewed papers and one international patent. His research

interests include numerical methods for image processing, computer vision and for the solution of large-scale discrete ill-posed inverse problems, with particular focus on variational methods of non-linear, non-differentiable and convex or non-convex type applied to inverse imaging problems.



Fiorella Sgallari is professor of Numerical Analysis at University of Bologna. She is a member of Academy of Sciences of Bologna and author of more than 160 publications. She coordinated several national, international and industrial research projects and is associate editor of international journals. From 2006–2013, she was the Chair of C.I. R. A. M.-Research Centre of Applied Mathematics, Univ.Bologna. Her research concerns PDE models and numerical methods for image processing and

solution of very large linear systems arising from discrete ill-posed problems.