Residual whiteness principle for parameter-free image restoration

(Article begins on next page)

04 December 2024

RESIDUAL WHITENESS PRINCIPLE FOR PARAMETER-FREE IMAGE RESTORATION

ALESSANDRO LANZA[†], MONICA PRAGLIOLA[‡], AND FIORELLA SGALLARI[§]

Abstract. Selecting the regularization parameter in the image restoration variational framework is of crucial importance, since it can highly influence the quality of the final restoration. In this paper, we propose a parameter-free approach for automatically selecting the regularization parameter when the blur is space-invariant and known and the noise is additive white Gaussian with unknown standard deviation, based on the so-called *residual whiteness principle*. More precisely, the regularization parameter is required to minimize the *residual whiteness function*, namely the normalized auto-correlation of the residual image of the restoration. The proposed method can be applied to a wide class of variational models, such as those including in their formulation regularizers of Tikhonov and Total Variation type. For non-quadratic regularizers, the residual whiteness principle is nested in an iterative optimization scheme based on the alternating direction method of multipliers. The effectiveness of the proposed approach is verified by solving some test examples and performing a comparison with other parameter estimation state-of-the-art strategies, such as the discrepancy principle.

Key words. image restoration, variational methods, regularization parameter, additive white Gaussian noise, alternating direction method of multipliers.

AMS subject classifications. AMS subject classifications: 68U10, 94A08, 65K10.

1. Introduction. Image restoration is the task of recovering an unknown target image x starting from a degraded - typically, blurred and noisy - observation b of it, and exploiting all the available information on the degradation process. When the blurring operator is known, restoration is referred to as *non-blind*. Image noise corruption has probabilistic nature, hence possible a priori information is related to its joint (over all pixels) probability distribution. One of the most ubiquitous noises in practice is the so-called additive white Gaussian noise: intensity at any image pixel is independently corrupted by additive random noise coming from the same univariate zero-mean Gaussian distribution of known or unknown standard deviation.

In this paper we consider the problem of non-blind restoration of images corrupted by known linear space-invariant blur and AWG noise of unknown standard deviation.

In the continuous setting, the inverse problem of image restoration can be formalized by a Fredholm integral equations of the first kind,

$$(1.1) \quad b(\xi) = (h * x)(\xi) + e(\xi) = \int_{\varphi \in \Omega} h(\xi - \varphi)x(\varphi)d\varphi, \quad \xi = (\xi_1, \xi_2) \in \Omega \subset \mathbb{R}^2,$$

where, given the observed image b and the convolution blur kernel h, one seeks for a solution x, with $*$ the 2-D convolution operator and Ω a compact rectangular domain. After suitable discretization and with a little abuse of notation, problem (1.1) takes the following finite-dimensional form

$$(1.2) \quad b = \mathrm{H}x + e, \quad e \text{ realization of } E \sim \mathcal{N}(0, \sigma^2 \mathrm{I}_n),$$

where $x, b, e \in \mathbb{R}^n$, $n = n_1 n_2$, are column major forms of the uncorrupted, corrupted and noise images, respectively, all made up of $n_1 \times n_2$ pixels, the matrix $\mathrm{H} \in \mathbb{R}^{n \times n}$ represents the discrete linear space-invariant blur operator, $\mathrm{I}_n \in \mathbb{R}^{n \times n}$ denotes the identity matrix and E is an n-variate Gaussian-distributed random vector with zero mean and scalar covariance matrix, with σ indicating the noise standard deviation. The image resoration problem (1.2) is a discrete

---

[†]alessandro.lanza2@unibo.it.

[‡]monica.pragliola2@unibo.it

[§]fiorella.sgallari@unibo.it

linear inverse problem whose main difficulty comes typically from its large size coupled with ill-conditioning of the blur operator H. A possible strategy to overcome ill-conditioning of problem (1.2) is *regularization*, which amounts to seek for an estimate x^* of the target image x by solving a nearby well-posed problem.

In the *variational* approach, one seeks for x^* minimizing a cost functional $\mathcal{J} : \mathbb{R}^n \to \mathbb{R}$,

$$(1.3) \qquad x^*(\mu) \in \arg\min_{x \in \mathbb{R}^n} \mathcal{J}(x;\mu), \quad \mathcal{J}(x;\mu) := \mu\mathcal{F}(x;\mathrm{H},b) + \mathcal{R}(x)\,,$$

with the functionals $\mathcal{F}$ and $\mathcal{R}$ commonly referred to as the *data fidelity term* and the *regularization term*, respectively. The fidelity term $\mathcal{F}$ measures the 'distance' between the observed data b and the target image x after the action of the operator H with respect to some norm related to statistics of the noise corrupting data, cf., e.g., [21]. In the considered case of AWG noise corruption, it is well-known that a suitable choice for $\mathcal{F}$ is the so-called $\mathcal{L}_2$ fidelity term which is related to the standard finite-dimensional Euclidean norm of the residual image; in formula,

$$\mathcal{F}(x;\mathrm{H},b) \,=\, \mathcal{L}_2(x;\mathrm{H},b) := \frac{1}{2}\|\mathrm{H}x - b\|_2^2\,.$$

The regularization term $\mathcal{R}$ in (1.3) encodes prior information on the unknown target image x, typically in the form of smoothness properties. The best choice for $\mathcal{R}$ can thus strongly depend on the particular application at hand.

Finally, the so-called regularization parameter $\mu \in \mathbb{R}_*^+$ in (1.3) - in the paper, we denote by $\mathbb{R}_*^+$ and $\mathbb{R}^+$ the sets of positive and non-negative real numbers, respectively - balances the action of the fidelity and regularization terms and is of crucial importance for achieving high quality restorations. In fact, even in case that $\mathcal{F}$ and $\mathcal{R}$ are suitably chosen, a bad value of μ leads to meaningless restorations. For this reason, the problem of selecting a good value of μ has attracted a lot of attention in the image restoration research field, and is the focus of this paper. Solutions to such a problem can be abstractly formalized as follows

$$\text{select } \mu = \mu^* \text{ such that } \mathcal{P}\left(x^*(\mu^*)\right) \text{ is satisfied,}$$

with $x^*(\mu) : \mathbb{R}_*^+ \to \mathbb{R}^n$ the image restoration solution function in (1.3) and $\mathcal{P}(\cdot)$ some selection criterion, or *principle*.

Two of the most popular among the classical parameter selection strategies are the L-curve [5, 10] and the generalized cross-validation (GCV) [6], providing an estimate of the optimal μ based on the behavior of the norm of the residual image. Both methods bring along some downsides, such as, just to name a few, the existence of a corner in the L-curve and the feasibility of the computations in the GCV, especially when dealing with large-scale problems - see [7] for a more extensive discussion - and non-quadratic regularizers. A plethora of literature has been focused on developing strategies to overcome the downsides of these classical methods [8, 17, 12, 18, 2].

Then, there is a class of selection strategies which try to exploit the available information on noise corruption. The simple idea is that, the nearer the restored image $x^*(\mu)$ is to the desired target image, the closer the associated residual image $r^*(\mu) := \mathrm{H}x^*(\mu) - b$ is to the noise realization e in (1.2). The general principle reads:

$$\mathcal{P}\left(x^*(\mu^*)\right): \quad r^*(\mu^*) = \mathrm{H}x^*(\mu^*) - b \approx e\,.$$

However, as previously noted, no deterministic information is typically available on the noise realization e, but only information on the noise probability distribution. When noise is AWG

with known standard deviation σ, then one can use the popular *discrepancy principle* (DP) [16, 9], which reads:

$$(1.4) \qquad \mathcal{P}\left(x^*(\mu^*)\right): \quad \|r^*(\mu^*)\|_2 \, = \, \|\mathrm{H}x^*(\mu^*) - b\|_2 \, = \, \varrho_\tau := \tau\sqrt{n}\,\sigma\,, \qquad \left[\mathrm{DP}\right]$$

where τ is the so-called *discrepancy coefficient*. When adopting DP, the noise standard deviation σ is assumed to be known and, as a consequence, $\tau = 1$. Nonetheless, numerical tests have shown that better restoration results can be obtained by letting τ be less than 1 - see, e.g., [4]. As a further downside of DP, we remark that typically, in real world applications, σ is unknown and has to be estimated.

To overcome the limitations of DP, other principles not using information on the noise standard deviation but on its whiteness property have been introduced as well. Among these strategies, it is worth mentioning [11, 20] and [1], where the authors propose two statistically-motivated parameter selection procedures based on the maximization of the residual whiteness by the normalized cumulative periodogram and the normalized auto-correlation, respectively. Using the whiteness property of the residual image - i.e. of the noise - is a very effective idea, since, on one hand it does not need any *a priori* knowledge or possibly unreliable estimates of σ. On the other hand, it exploits much more information than the one encoded in some moments of the noise distribution. Recently, variational models containing fidelity terms or constraints aimed at explicitly enforcing the residual whiteness property have been proposed; see, e.g., [15, 14, 13]. The good results obtained by such models demonstrate the potentiality of using the noise whiteness property. However, the proposed models are strongly nonconvex, with all the associated difficulties (existence of local minimizers, convergence of iterative solvers).

Here, we aim at selecting automatically the regularization parameter of a large class of convex variational models containing the quadratic $\mathcal{L}_2$ fidelity term, based on a suitable measure of the residual image whiteness. In fact, all previous strategies for the selection of the regularization parameter μ based on residual whiteness can not be considered *automatic* in a strict sense, as the selection of μ is performed *a posteriori*: the variational model is solved for a grid of different μ-values and the optimal μ is chosen as the one maximizing the residual whiteness. Such a procedure is quite costly, even if a warm-starting strategy like the one proposed in [1] is adopted.

In this paper, we propose a fully automatic and parameter-free selection strategy for the regularization parameter μ, which - like [11, 20, 1] - is based on a criterion, referred to as the *residual whiteness principle* (RWP), according to which we select the μ value maximizing the whiteness of the residual image $r^*(\mu) = \mathrm{H}x^*(\mu) - b$. More precisely, our principle is based on minimizing the squared Euclidean norm of the normalized auto-correlation of the residual image. No knowledge about the noise standard deviation is required. For the simple case of a Tikhonov-regularized least-squares problem, our proposal allows to get the optimal μ^* by simply solving a scalar optimization problem. Based on this result, and more interestingly, we propose an iterative framework based on the *alternating direction method of multipliers* (ADMM) which allows to apply the RWP to the automatic selection of μ for a large class of non-quadratically regularized variational models.

The paper is organized as follows. In Sect. 2 we introduce the class of variational models for which the proposed automatic selection strategy can be applied. The considered RWP is motivated and formally defined in Sect. 3 and its application to simple Tikhonov-regularized least-squares problems is illustrated in Sect. 4. The proposed ADMM-based iterative approach aimed to automatically select the regularization parameter for the class of considered variational models is described in detail in Sect. 5. Some numerical results are presented in Sect. 6 and conclusions are drawn in Sect. 7.

.

2. The class of considered variational models. The class of variational image restoration models for which the proposed automatic selection strategy can be applied take the form

$$(2.1) \qquad x^*(\mu) = \arg \min_{x \in \mathbb{R}^n} \left\{ \mathcal{J}(x; \mu) = \frac{\mu}{2} \|\mathrm{H}x - b\|_2^2 + \mathcal{R}(x) \right\}, \quad \mathcal{R}(x) = G(\mathrm{L}x),$$

under the following assumptions on the blur matrix $\mathrm{H} \in \mathbb{R}^{n \times n}$, the regularization matrix $\mathrm{L} \in \mathbb{R}^{m \times n}$ and the possibly nonlinear function $G : \mathbb{R}^m \to \mathbb{R} \cup \{+\infty\}$:

($\mathcal{A}$1) the blur matrix H and the regularization matrix L are such that:

$$(2.2) \qquad \qquad \mathrm{null}(\mathrm{H}) \cap \mathrm{null}(\mathrm{L}) = \{0_n\},$$

with 0_n the n-dimensional null vector.

($\mathcal{A}$2) the function G is proper, lower semi-continuous, convex and coercive.

($\mathcal{A}$3) the blur matrix H represents a 2D discrete convolution operator - i.e., the blur is space-invariant - and the regularization matrix L is of the form:

$$\mathrm{L} = \left(\mathrm{L}_1^T, \ldots, \mathrm{L}_d^T\right)^T \in \mathbb{R}^{dn \times n}, \ d \in \mathbb{N} \setminus \{0\}, \quad \text{with } \mathrm{L}_j \in \mathbb{R}^{n \times n}, \ j = 1, \ldots, d,$$

matrices also representing 2D discrete convolution operators.

($\mathcal{A}$4) the function $G : \mathbb{R}^{dn} \to \mathbb{R}$ is *easily proximable*, that is, the proximity operator of G at any $t \in \mathbb{R}^{dn}$,

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

can be easily - i.e. efficiently - computed.

The first two assumptions ($\mathcal{A}$1)-($\mathcal{A}$2) together constitute a sufficient condition for the existence - and, eventually, uniqueness - of solutions of the considered class of variational models defined in (2.1), as formally stated in Proposition 2.1 below. Assumptions ($\mathcal{A}$3)-($\mathcal{A}$4) allow for efficient automatic selection of the regularization parameter, as it will be illustrated in Sects. 4-5. In particular, ($\mathcal{A}$3) guarantees that, under suitable boundary conditions assumptions, the blur matrix H and the regularization matrices L_j are all diagonalizable by fast discrete 2D transforms. More specifically, in this paper we assume periodic boundary conditions for images, such that H and L_j are all block-circulant matrices with circulant blocks, which can be diagonalized by the 2D discrete Fourier transform. Formally, we can thus write:

$$(2.3) \qquad \mathrm{H} = \mathrm{F}^* \widetilde{\mathrm{H}} \mathrm{F}, \quad \mathrm{L}_j = \mathrm{F}^* \widetilde{\mathrm{L}}_j \mathrm{F}, \ j \in \{1, \ldots, d\}, \quad \mathrm{F}^* \mathrm{F} = \mathrm{F} \mathrm{F}^* = \mathrm{I}_n,$$

where $\mathrm{F}, \mathrm{F}^* \in \mathbb{C}^{n \times n}$ are unitary matrices representing the 2D discrete Fourier transform operator and its conjugate transpose (i.e., its inverse), respectively, and $\widetilde{\mathrm{H}}, \widetilde{\mathrm{L}}_j \in \mathbb{C}^{n \times n}$ are diagonal matrices defined by

$$(2.4) \qquad \widetilde{\mathrm{H}} = \mathrm{diag}\left(\tilde{h}_1, \ldots, \tilde{h}_n\right), \quad \widetilde{\mathrm{L}}_j = \mathrm{diag}\left(\tilde{l}_{j,1}, \ldots, \tilde{l}_{j,n}\right), \ j \in \{1, \ldots, d\}.$$

PROPOSITION 2.1. *In case that assumptions ($\mathcal{A}$1) and ($\mathcal{A}$2) above are both fulfilled, for any fixed $\mu \in \mathbb{R}_*^+$ the function $\mathcal{J}(\,\cdot\,; \mu) : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ in (2.1) is proper, lower semi-continuous, convex and coercive, hence it admits global minimizers. Furthermore, if matrix H has full rank, then the global minimizer is unique.*

Proof. After noticing that, for any fixed $\mu \in \mathbb{R}_*^+$, the quadratic fidelity term in (2.1) is clearly a proper, continuous and convex function, and that by assumption ($\mathcal{A}$2) on function

G the regularization term $\mathcal{R}(x) = G(\mathrm{L}x)$ is a proper, lower semi-continuous and convex function, it follows that the total cost function $\mathcal{J}(x; \mu)$ in (2.1) is also proper, lower semi-continuous and convex in x. In order to prove that $\mathcal{J}$ is also coercive, such that existence of global minimizers for $\mathcal{J}$ follows, we introduce the function $Z : \mathbb{R}^n \to \mathbb{R}$ defined by

$$(2.5) \qquad Z(y) := \frac{\mu}{2} \|y - b\|_2^2, \quad y \in \mathbb{R}^n,$$

and then write the cost function $\mathcal{J}$ in (2.1) in the equivalent form

$$(2.6) \qquad \mathcal{J}(x; \mu) = Z(\mathrm{H}x) + G(\mathrm{L}x).$$

Under assumptions $(\mathcal{A}1)$-$(\mathcal{A}2)$, and noticing that the function Z in (2.5) is clearly proper, lower semi-continuous and coercive in y, we can apply Lemma 5.1 in [3] and conclude that the function $\mathcal{J}$ in (2.6) is coercive in x.

Finally, in the special case that matrix H has full rank, the quadratic fidelity term in (2.1) is a strongly convex function of x, hence the total cost function $\mathcal{J}$ is also strongly convex in x. This implies that $\mathcal{J}$ admits a unique global minimizer.

⬜

As far as the class of considered regularizers is concerned, first we notice that two popular choices for the regularization matrix L satisfying assumptions $(\mathcal{A}1)$, $(\mathcal{A}3)$ are the following:

$$(2.7) \ \ \mathrm{L} = \mathrm{D}^{(1)} := \left(\mathrm{D}_h^T, \mathrm{D}_v^T \right)^T \in \mathbb{R}^{2n \times n}, \qquad \mathrm{L} = \mathrm{D}^{(2)} := \left(\mathrm{D}_{hh}^T, \mathrm{D}_{vv}^T, \mathrm{D}_{hv}^T \right)^T \in \mathbb{R}^{3n \times n},$$

with $\mathrm{D}_h, \mathrm{D}_v, \mathrm{D}_{hh}, \mathrm{D}_{vv}, \mathrm{D}_{hv} \in \mathbb{R}^{n \times n}$ finite difference operators discretizing the first-order horizontal and vertical partial derivatives and the second-order horizontal, vertical and mixed horizontal-vertical partial derivatives, respectively. In this paper, we will consider regularizers based on the first-order finite difference matrix $\mathrm{D}^{(1)}$.

Within this class, Tikhonov (TIK) regularization is usually preferred when the image to restore is characterized by smooth features, while the Total Variation (TV) regularizer [19] is known to be particularly suited to preserving edges. The TIK and TV regularizers are defined as in (2.1) by taking the matrix L and the function G as follows:

$$(2.8) \qquad \mathrm{L} = \mathrm{D}^{(1)}, \qquad G(t) = \|t\|_2^2, \qquad\qquad \left[\mathrm{TIK}\right]$$

$$(2.9) \qquad \mathrm{L} = \mathrm{D}^{(1)}, \qquad G(t) = \sum_{i=1}^n \|(t_i, t_{i+n})\|_2. \qquad \left[\mathrm{TV}\right]$$

The TIK and TV regularizers are both *global* or *space-invariant*, and, as a consequence, the regularization strength at each pixel is the same regardless of the local image structures. In the light of this observation, recently in [4] the authors proposed a class of spatially Weighted Total Variation (WTV) regularization terms, defined as in (2.1) by taking the matrix L and the function G as follows:

$$(2.10) \qquad \mathrm{L} = \mathrm{D}^{(1)}, \qquad G(t) = \sum_{i=1}^n \alpha_i \|(t_i, t_{i+n})\|_2, \qquad \left[\mathrm{WTV}\right]$$

where the regularization weights $\alpha_i > 0$ are automatically estimated according to the local image properties, in order to enforce the regularization on constant and smoothly varying regions, while keeping its contribution limited in correspondence of edges and texture.

We notice that the three functions G defined in (2.8)-(2.10) all satisfy assumptions $(\mathcal{A}2)$ and $(\mathcal{A}4)$.

3. Residual Whiteness Principle. Let us consider the noise realization $e \in \mathbb{R}^n$ in (1.2) in its original $n_1 \times n_2$ matrix form, with $n = n_1 n_2$, that is

$$e = \{e_{i,j}\}_{(i,j) \in \Omega}, \quad \Omega := \{0, \dots, n_1 - 1\} \times \{0, \dots, n_2 - 1\}.$$

The *sample auto-correlation* of e is a function $a : \mathbb{R}^{n_1 \times n_2} \to \mathbb{R}^{(2n_1-1) \times (2n_2-1)}$ defined by

$$a(e) = \{a_{l,m}(e)\}_{(l,m) \in \Theta}, \quad \Theta := \{-(n_1-1), \dots, n_1-1\} \times \{-(n_2-1), \dots, n_2-1\},$$

with each scalar component $a_{l,m}(e) : \mathbb{R}^{n_1 \times n_2} \to \mathbb{R}$ given by

$$
\begin{aligned}
a_{l,m}(e) &= \frac{1}{n} \left(e \star e \right)_{l,m} = \frac{1}{n} \left(e * e' \right)_{l,m} \\
&= \frac{1}{n} \sum_{(i,j) \in \Omega} e_{i,j}\, e_{i+l, j+m}, \quad (l,m) \in \Theta,
\end{aligned}
\tag{3.1}
$$

where index pairs (l, m) are commonly called *lags*, $\star$ and $*$ denote the 2-D discrete correlation and convolution operators, respectively, and where $e'(i, j) = e(-i, -j)$. Clearly, for (3.1) being defined for all lags $(l, m) \in \Theta$, the noise realization e must be padded with at least $n_1 - 1$ samples in the vertical direction and $n_2 - 1$ samples in the horizontal direction. We assume here periodic boundary conditions for e, such that $\star$ and $*$ in (3.1) denote 2-D circular correlation and convolution, respectively. Under this assumption, the auto-correlation presents some symmetries that allow to consider only lags

$$(l, m) \in \overline{\Theta} := \{0, \dots, n_1 - 1\} \times \{0, \dots, n_2 - 1\}. \tag{3.2}$$

If the corruption e in (1.2) is the realization of a white Gaussian noise process - which is our case - then it is well known that the sample auto-correlation $a(e)$ satisfies the following asymptotic property:

$$
\lim_{n \to +\infty} a_{l,m}(e) =
\begin{cases}
\sigma^2 & \text{for } (l,m) = (0,0) \\
0 & \text{for } (l,m) \in \overline{\Theta}_0 := \overline{\Theta} \setminus \{(0,0)\}.
\end{cases}
\tag{3.3}
$$

For additive white Gaussian noise corruptions affecting images of realistic (finite) size - namely, n between 10^4 and 10^7 - we can roughly say that the auto-correlation values at non-zero lags $(l, m) \in \overline{\Theta}_0$ of the noise realization e have all *small* magnitude, depending on n. More precisely, it can be demonstrated - see Proposition 1 in [13] - that all quantities $a_{l,m}(e), (l,m) \in \overline{\Theta}_0$ can be regarded as random variables having (approximately) Gaussian distribution with zero mean and standard deviation $\sigma_a = \sigma^2/n$, with σ^2 the noise variance. It clearly follows that, as a general rule, the larger is the image size, the smaller will be (on average) the auto-correlation values of the white noise process. Also in case of realistically small images (e.g., $n = 100 \times 100 = 10^4$), the auto-correlation values are given by a very small fraction of the noise variance.

It is worth noting that the popular discrepancy principle relies on exploiting only one - namely, the first one for lag $(0,0)$ - among the n asymptotic properties of the noise auto-correlation given in (3.3). In fact, according to the discrepancy principle, the restoration residual has only to present a standard deviation approximately equal to the standard deviation of noise σ. The other $n - 1$ properties, related to noise whiteness, are not used at all, such that the restoration residual can contain auto-correlated structures belonging to the image to be restored.

Imposing whiteness of the restoration residual by constraining the residual auto-correlation at non-zero lags to be small is a much stronger requirement.

In order to render such whiteness principle completely independent of the noise level, we consider the *normalized* sample auto-correlation of noise realization e, namely

$$(3.4) \qquad \rho(e) \;=\; \frac{1}{a_{0,0}(e)}\, a(e) \;=\; \frac{1}{\|e\|_2^2}\, \big(e \star e \big),$$

where $\|M\|_2$ denotes the Frobenius norm of matrix M. It follows easily from (3.3) that $\rho(e)$ satisfies the following asymptotic properties:

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

We can thus introduce the following σ-independent non-negative scalar measure of whiteness $\mathcal{W} : \mathbb{R}^{n_1 \times n_2} \to \mathbb{R}^+$ of noise realization e:

$$(3.5) \qquad \mathcal{W}(e) \;:=\; \|\rho(e)\|_2^2 \;=\; \frac{\| e \star e \|_2^2}{\|e\|_2^4} \;=\; \widetilde{\mathcal{W}}(\tilde{e})\,,$$

where the second equality follows from the definition of ρ in (3.4) and the last equality comes from Proposition 3.1 below, with $\tilde{e}$ denoting the 2D discrete Fourier transform of e and $\widetilde{\mathcal{W}} : \mathbb{C}^{n_1 \times n_2} \to \mathbb{R}^+$ the function defined in (3.6).

PROPOSITION 3.1. *Let $e \in \mathbb{R}^{n_1 \times n_2}$ and let $\tilde{e} \in \mathbb{C}^{n_1 \times n_2}$ be its 2D discrete Fourier transform. Then, under the assumption of periodic boundary conditions for e, the function $\mathcal{W}$ defined in (3.5) satisfies:*

$$(3.6) \qquad \mathcal{W}(e) = \widetilde{\mathcal{W}}(\tilde{e}) := \frac{\displaystyle\sum_{(l,m)\in\overline{\Theta}} |\tilde{e}_{l,m}|^4}{\left(\displaystyle\sum_{(l,m)\in\overline{\Theta}} |\tilde{e}_{l,m}|^2 \right)^{\!2}}\,, \quad \text{with } \overline{\Theta} \text{ defined in } (3.2)\,.$$

Proof. First, the convolution - or, better, correlation - theorem allows to write

$$(3.7) \qquad \widetilde{(e \star e)} \;=\; \big(\tilde{e} \odot \overline{\tilde{e}} \big) \;=\; \Big\{ |\tilde{e}_{l,m}|^2 \Big\}_{(l,m)\in\overline{\Theta}}\,,$$

where $\odot$ denotes the Hadamard matrix product operator and $\bar{c}$, $|c|$ indicate the conjugate and the modulus of complex number c, respectively. Then, by applying the Parseval's theorem we have

$$(3.8) \qquad \| e \star e \|_2^2 \;=\; \big\| \widetilde{e \star e} \big\|_2^2 \;=\; \sum_{(l,m)\in\overline{\Theta}} |\tilde{e}_{l,m}|^4\,,$$

$$(3.9) \qquad \|e\|_2^4 \;=\; \Big(\|e\|_2^2 \Big)^2 = \Big(\|\tilde{e}\|_2^2 \Big)^2 \;=\; \left(\sum_{(l,m)\in\overline{\Theta}} |\tilde{e}_{l,m}|^2 \right)^{\!2}\,,$$

where the last equality in (3.8) comes from (3.7). Statement (3.6) follows straightforwardly by replacing (3.8) and (3.9) into the fractional expression of $\mathcal{W}$ in (3.5). ☐

Clearly, the nearer the restored image $x^*(\mu)$ in (2.1) is to the target uncorrupted image, the closer the associated residual image $r^*(\mu) = \mathrm{H}x^*(\mu) - b$ is to the white noise realization e in (1.2) and, hence, the whiter is the residual image according to the scalar measure in (3.5).

This motivates the proposed residual whiteness principle for automatically selecting the regularization parameter μ of variational models of the general form (2.1), which reads:

$$(3.10) \quad \mu^* \in \arg \min_{\mu \in \mathbb{R}^+_*} W(\mu), \quad W(\mu) := \mathcal{W}\left(r^*(\mu)\right) = \widetilde{\mathcal{W}}\left(\widetilde{r^*(\mu)}\right), \quad r^*(\mu) = \mathrm{H}x^*(\mu) - b.$$

Recalling (3.5)-(3.6), the scalar non-negative cost function $W : \mathbb{R}^+_* \to \mathbb{R}^+$ in (3.10), from now on referred to as the *residual whiteness function*, takes the following form:

$$(3.11) \quad W(\mu) \;=\; \|\rho(r^*(\mu))\|_2^2 \;=\; \frac{\| r^*(\mu) \star r^*(\mu) \|_2^2}{\|r^*(\mu)\|_2^4} \;=\; \frac{\displaystyle\sum_{i=1}^n w_i^4(\mu)}{\left(\displaystyle\sum_{i=1}^n w_i^2(\mu)\right)^2},$$

with the n scalar non-negative functions $w_i : \mathbb{R}^+_* \to \mathbb{R}^+$, $i \in \{1, \ldots, n\}$, defined by

$$(3.12) \quad w_i(\mu) \;=\; \left|\widetilde{r}^*_{l,m}(\mu)\right|, \quad (l, m) \in \overline{\Theta}, \quad i = l + n_1 m + 1.$$

4. Exact Residual Whiteness Principle for generalized Tikhonov regularization. In this section, we derive the analytical expression of the whiteness function $W(\mu)$ defined in (3.11)–(3.12) for a particular choice of the regularization function G in (2.1), namely $G(t) = \|t - v\|_2^2$, $v \in \mathbb{R}^m$. The corresponding variational model, referred to as generalized Tikhonov-$\mathcal{L}_2$ model, reads

$$(4.1) \quad x^*(\mu) \;=\; \arg \min_{x \in \mathbb{R}^n} \left\{ \mathcal{J}(x; \mu) = \frac{\mu}{2}\|\mathrm{H}x - b\|_2^2 + \frac{1}{2}\|\mathrm{L}x - v\|_2^2 \right\},$$

where vector $v \in \mathbb{R}^m$ reduces to the null vector for classical Tikhonov regularization. It is worth remarking that here we consider a generic vector $v \in \mathbb{R}^m$ - which not only can be non-null but also does not necessarily belong to the range of matrix L - since this choice facilitates the extension of the principle described below to more general regularizers, as it will be illustrated in Sect. 5.

Thanks to assumption ($\mathcal{A}1$), for any given $\mu \in \mathbb{R}^+_*$ the quadratic cost function $\mathcal{J}(x; \mu)$ in (4.1) is strongly convex in x, hence the minimization problem admits the unique solution

$$x^*(\mu) \;=\; \left(\mu \mathrm{H}^T\mathrm{H} + \mathrm{L}^T\mathrm{L}\right)^{-1} \left(\mu \mathrm{H}^T b + \mathrm{L}^T v\right).$$

The associated residual image thus reads

$$r^*(\mu) = \mathrm{H}x^*(\mu) - b = \mathrm{H}\left(\mu \mathrm{H}^T\mathrm{H} + \mathrm{L}^T\mathrm{L}\right)^{-1}\left(\mu \mathrm{H}^T b + \mathrm{L}^T v\right) - b$$

$$(4.2) \quad = \mathrm{H}\left(\mu \mathrm{H}^T\mathrm{H} + \sum_{j=1}^d \mathrm{L}_j^T\mathrm{L}_j\right)^{-1}\left(\mu \mathrm{H}^T b + \sum_{j=1}^d \mathrm{L}_j^T v_j\right) - b$$

$$(4.3) \quad = \mathrm{F}^*\widetilde{\mathrm{H}}\mathrm{F}\left(\mu \mathrm{F}^*|\widetilde{\mathrm{H}}|^2\mathrm{F} + \mathrm{F}^*\sum_{j=1}^d |\widetilde{\mathrm{L}}_j|^2\mathrm{F}\right)^{-1}\left(\mu \mathrm{F}^*\overline{\widetilde{\mathrm{H}}}\mathrm{F}b + \mathrm{F}^*\sum_{j=1}^d \overline{\widetilde{\mathrm{L}}}_j\mathrm{F}v_j\right) - b$$

$$(4.4) \quad = \mathrm{F}^*\widetilde{\mathrm{H}}\left(\mu|\widetilde{\mathrm{H}}|^2 + \sum_{j=1}^d |\widetilde{\mathrm{L}}_j|^2\right)^{-1}\left(\mu\overline{\widetilde{\mathrm{H}}}\widetilde{b} + \sum_{j=1}^d \overline{\widetilde{\mathrm{L}}}_j\widetilde{v}_j\right) - \mathrm{F}^*\widetilde{b},$$

where (4.2) comes from assumption ($\mathcal{A}3$) on matrix L, with $v = \left(v_1^T, \ldots, v_d^T\right)^T \in \mathbb{R}^{dn}$, $v_j \in \mathbb{R}^n \, \forall j \in \{1, \ldots, d\}$, and (4.3) comes from (2.3)-(2.4), with

$$|\widetilde{\mathrm{H}}|^2 = \mathrm{diag}\left(|\tilde{h}_1|^2, \ldots, |\tilde{h}_n|^2\right), \quad |\widetilde{\mathrm{L}}_j|^2 = \mathrm{diag}\left(|\tilde{l}_{j,1}|^2, \ldots, |\tilde{l}_{j,n}|^2\right), \;\; j \in \{1, \ldots, n\}.$$

The Fourier-transformed residual image takes the form

$$\widetilde{r}^*(\mu) = \mathrm{F}\, r^*(\mu) = \left(\mu|\widetilde{\mathrm{H}}|^2 + \sum_{j=1}^{d}|\widetilde{\mathrm{L}}_j|^2\right)^{-1}\left(\mu|\widetilde{\mathrm{H}}|^2\tilde{b} + \widetilde{\mathrm{H}}\sum_{j=1}^{d}\overline{\widetilde{\mathrm{L}}}_j\tilde{v}_j\right) - \tilde{b},$$

$$(4.5) \qquad = \left(\mu|\widetilde{\mathrm{H}}|^2 + \sum_{j=1}^{d}|\widetilde{\mathrm{L}}_j|^2\right)^{-1}\left(\widetilde{\mathrm{H}}\sum_{j=1}^{d}\overline{\widetilde{\mathrm{L}}}_j\tilde{v}_j - \tilde{b}\sum_{j=1}^{d}|\widetilde{\mathrm{L}}_j|^2\right).$$

We thus introduce the following notations

$$(4.6) \qquad \tilde{z} = \widetilde{\mathrm{H}}\sum_{j=1}^{d}\overline{\widetilde{\mathrm{L}}}_j\tilde{v}_j = \mathrm{F}z \in \mathbb{C}^n, \quad z = \mathrm{HL}^T v,$$

and

$$(4.7) \qquad \zeta_i = \left|\tilde{h}_i\right|^2, \quad \eta_i = \sum_{j=1}^{d}\left|\tilde{l}_{j,i}\right|^2, \quad \epsilon_i = \left|\tilde{z}_i - \eta_i\,\tilde{b}_i\right|, \quad i \in \{1, \ldots, n\}.$$

with $\zeta_i, \eta_i, \epsilon_i$ non-negative real constants - i.e. independent of μ. Hence, the modulus of the i-th entry of vector $\tilde{r}^*(\mu) \in \mathbb{C}^n$, which corresponds to function $w_i(\mu)$ in the Fourier domain definition (3.11)–(3.12) of the residual whiteness function, is given by

$$(4.8) \qquad w_i(\mu) = \left|\tilde{r}_i^*(\mu)\right| = \frac{\epsilon_i}{\mu\,\zeta_i + \eta_i},$$

Substituting the derived expressions (4.8) of functions $w_i(\mu)$ into definition (3.11), we obtain the following analytic form for the residual whiteness function $W(\mu)$:

$$(4.9) \qquad W(\mu) = \frac{\displaystyle\sum_{i=1}^{n}w_i^4(\mu)}{\left(\displaystyle\sum_{i=1}^{n}w_i^2(\mu)\right)^2} = \frac{\displaystyle\sum_{i=1}^{n}\left(\frac{\epsilon_i}{\mu\,\zeta_i + \eta_i}\right)^4}{\left(\displaystyle\sum_{i=1}^{n}\left(\frac{\epsilon_i}{\mu\,\zeta_i + \eta_i}\right)^2\right)^2}.$$

PROPOSITION 4.1. *Let W be the residual whiteness function and I, S, T be the following set of indices:*

$$I = \{1\ldots, n\}, \quad S = \{i \in I : \zeta_i = 0, \, \epsilon_i \neq 0\}, \quad T = \{i \in I : \zeta_i \neq 0, \, \epsilon_i \neq 0\}.$$

Moreover, let ℓ_1, ℓ_2 denote the following limits:

$$\ell_1 := \lim_{\mu\to 0^+}W(\mu), \quad \ell_2 := \lim_{\mu\to +\infty}W(\mu).$$

Then, the following results hold:

  (i) $W(\mu) \in C^\infty(\mathbb{R}^+)$, with $\mathbb{R}^+$ denoting the set of non-negative real numbers;
  (ii) $\ell_1, \ell_2 \in \mathbb{R}_*^+$;
  (iii) If $\ell_1 < \ell_2$, then $W(\mu)$ admits a minimum in $\mathbb{R}^+$;
  (iv) If $\ell_1 > \ell_2$ and, in addition,

$$\left| \frac{\tilde{z}_i}{\eta_i} - \tilde{b}_i \right|^2 \le \frac{\sum_{j \in S} |\tilde{b}_j|^4}{\sum_{j \in S} |\tilde{b}_j|^2}\,, \quad \forall i \in T\,,$$

  then $W(\mu)$ admits a minimum in $\mathbb{R}^+$.

The proof of Proposition 4.1 is reported in the Appendix. Condition (iv) can be efficiently verified, however future investigation will be aimed at studying other conditions guaranteeing the existence of a minimum for W.

REMARK 4.2. The discrepancy principle in (1.4) can be equivalently rewritten as

$$(4.10) \qquad \|r^*(\mu^*)\|_2^2 - \varrho_\tau^2 = \|\widetilde{r}^*(\mu^*)\|_2^2 - \varrho_\tau^2 = 0\,,$$

where the first equality comes from applying the Parseval's theorem. Hence, based on (4.10) and on the previously derived expression (4.5)-(4.8) for the Fourier-transformed residual image $\widetilde{r}^*(\mu)$, the discrepancy principle applied to the generalized Tikhonov-regularized least-squares problem (4.1) reduces to selecting $\mu = \mu^*$ with μ^* a root of the following non-linear equation:

$$(4.11) \qquad D(\mu) := \sum_{i=1}^n w_i^2(\mu) - \varrho_\tau^2 = 0\,, \quad \text{with } w_i \text{ defined in } (4.6)-(4.8).$$

It is easy to prove that the *discrepancy function* $D : \mathbb{R}_*^+ \to \mathbb{R}$ in (4.11) is monotonically decreasing in μ, such that for reasonable values of the constant ϱ_τ - that is, of the discrepancy coefficient τ in (1.4) - the root μ^* of the non-linear equation in (4.11) is unique.

5. Iterated Residual Whiteness Principle for easily-proximable regularizers. In the previous section we illustrated how the proposed residual whiteness principle allows to automatically select the regularization parameter μ of quadratically-regularized least-squares problems based only on minimizing a scalar function of one variable.

For more general regularization terms, such as those belonging to the considered class of regularizers defined in (2.1), this is not the case as it is very hard - if not impossible - to get an explicit analytical expression for the model solution function $x^*(\mu)$. Indeed, for any given μ value, the solution of most of the variational models in (2.1) can only be obtained by resorting to iterative minimization algorithms.

Our idea is to propose an iterative solution framework which, on one hand, can be applied to the efficient solution of any variational model of the form (2.1) under assumptions $(\mathcal{A}1)$-$(\mathcal{A}4)$ and, on the other hand, allows to iteratively adjust the regularization parameter μ such that the residual whiteness principle is exactly satisfied at each iteration for some surrogate Tikhonov-regularized least-squares problem. In the following, we illustrate in detail our proposal, referred to in short as IRWP-ADMM since it is iterative and based on the ADMM minimization approach.

First, we resort to the variable splitting strategy and rewrite our family of variational models (2.1) in the following equivalent linearly constrained form:

$$(5.1) \qquad \{x^*(\mu), t^*(\mu)\} \in \arg\min_{x,t} \left\{ \frac{\mu}{2} \|\mathrm{H}x - b\|_2^2 + G(t) \right\} \quad \text{s.t.} : \quad t = \mathrm{L}x,$$

where $t \in \mathbb{R}^m$ is the newly introduced variable.

To solve problem (5.1), we introduce the augmented lagrangian (AL) function,

$$(5.2) \qquad \mathrm{AL}(x, t, \lambda; \mu) = \frac{\mu}{2} \|\mathrm{H}x - b\|_2^2 + G(t) - \langle \lambda, t - \mathrm{L}x \rangle + \frac{\beta}{2} \| t - \mathrm{L}x \|_2^2,$$

where $\beta > 0$ is a scalar penalty parameter and $\lambda \in \mathbb{R}^m$ is the dual variable, i.e. the vector of Lagrange multipliers associated with the set of m linear constraints in (5.1).

We observe that solving (5.1) is tantamount to seek for the saddle point(s) of the Augmented Lagrangian function in (5.2). The saddle-point problem reads as follows:

$$(5.3) \qquad \{x^*(\mu), t^*(\mu), \lambda^*(\mu)\} \in \arg \min_{x,t} \max_{\lambda} \mathrm{AL}(x, t, \lambda; \mu).$$

Upon suitable initialization, and for any $k \geq 0$, the k-th iteration of the standard ADMM applied to solving the saddle-point problem (5.3) with AL defined in (5.2) reads as follows:

$$(5.4) \quad x^{(k+1)} \in \arg \min_{x \in \mathbb{R}^n} \mathrm{AL}(x, t^{(k)}, \lambda^{(k)}; \mu)$$

$$= \arg \min_{x \in \mathbb{R}^n} \left\{ \frac{\mu/\beta}{2} \|\mathrm{H}x - b\|_2^2 + \frac{1}{2} \left\| \mathrm{L}x - v^{(k+1)} \right\|_2^2 \right\}, \quad v^{(k+1)} = t^{(k)} - \frac{\lambda^{(k)}}{\beta},$$

$$(5.5) \quad t^{(k+1)} \in \arg \min_{t \in \mathbb{R}^m} \mathrm{AL}(x^{(k+1)}, t, \lambda^{(k)}; \mu)$$

$$= \arg \min_{t \in \mathbb{R}^m} \left\{ G(t) + \frac{\beta}{2} \left\| t - q^{(k+1)} \right\|_2^2 \right\}, \quad q^{(k+1)} = \mathrm{L}x^{(k+1)} + \frac{\lambda^{(k)}}{\beta},$$

$$(5.6) \quad \lambda^{(k+1)} = \lambda^{(k)} - \beta \left(t^{(k+1)} - \mathrm{L}x^{(k+1)} \right).$$

First, we notice that the regularization parameter μ only appears in the minimization sub-problem (5.4) for the primal variable x and is there coupled with the ADMM penalty parameter β to form the sub-problem regularization parameter $\gamma := \mu/\beta$. Moreover, sub-problem (5.4) has the form of the Tikhonov-regularized problem in (4.1), with γ in place of μ. We thus propose to adjust μ - that is, γ - at each ADMM iteration by applying the exact residual whiteness principle outlined in Sect. 4 to the Tikhonov-regularized quadratic problem (5.4).

The complete x-update procedure reads as follows:

$$(5.7) \qquad v^{(k+1)} = t^{(k)} - \frac{\lambda^{(k)}}{\beta},$$

$$(5.8) \qquad \gamma^{(k+1)} = \arg \min_{\gamma \in \mathbb{R}^+} W(\gamma) \quad \text{with } W \text{ in } (4.6)-(4.9) \text{ and } v = v^{(k+1)},$$

$$(5.9) \qquad x^{(k+1)} = \left(\gamma^{(k+1)} \mathrm{H}^T \mathrm{H} + \mathrm{L}^T \mathrm{L} \right)^{-1} \left(\gamma^{(k+1)} \mathrm{H}^T b + \mathrm{L}^T v^{(k+1)} \right).$$

The minimization sub-problem (5.5) for the primal variable t can be written in the form of a proximity operator, namely

$$(5.10) \qquad t^{(k+1)} \in \mathrm{prox}_{\frac{1}{\beta} G} \left(q^{(k+1)} \right), \qquad q^{(k+1)} = \mathrm{L}x^{(k+1)} + \frac{\lambda^{(k)}}{\beta}.$$

According to assumption $(\mathcal{A}4)$, the regularization function G is easily proximable, which means that problem (5.10) admits a closed-form solution or, at least, it can be solved very efficiently. Actually, this is the case for many popular and effective regularizers, among

which the TV and WTV regularizers considered in this paper. Recalling expressions (2.9) and (2.10) for functions G defining the TV and WTV regularizers, one notices that the associated $2n$-variate proximity operators are both separable into n independent bivariate proximity operators. In particular, after introducing the n vectors $\breve{t}_i^{(k+1)}$, $\breve{q}_i^{(k+1)} \in \mathbb{R}^2$ defined by

$$(5.11) \quad \breve{t}_i^{(k+1)} := \left(t_i^{(k+1)}, t_{i+n}^{(k+1)} \right), \quad \breve{q}_i^{(k+1)} := \left(q_i^{(k+1)}, q_{i+n}^{(k+1)} \right), \quad i \in \{1, \dots, n\},$$

the two proximity operators for TV and WTV admit the following closed-form expressions:

$$(5.12) \quad \breve{t}_i^{(k+1)} = \max \left(1 - \frac{1}{\beta \left\| \breve{q}_i^{(k+1)} \right\|_2}, 0 \right) \breve{q}_i^{(k+1)}, \quad i \in \{1, \dots, n\}, \qquad [\,\text{TV}\,]$$

$$(5.13) \quad \breve{t}_i^{(k+1)} = \max \left(1 - \frac{\alpha_i^{(k)}}{\beta \left\| \breve{q}_i^{(k+1)} \right\|_2}, 0 \right) \breve{q}_i^{(k+1)}, \quad i \in \{1, \dots, n\}, \qquad [\,\text{WTV}\,]$$

where the spatial weights of the WTV regularizer are denoted in (5.13) by $\alpha_i^{(k)}$, so as to recall that in the original WTV proposal [4] such weights can be updated along iterations.

Coupling two simultaneous iterated automatic selections, namely of the WTV weights α_i (acting on the local regularization strengths) as proposed in [4], and of the regularization parameter μ (acting on the global residual whiteness) as proposed in this paper, holds the potential for a very effective parameter-free image restoration approach. On the other hand, convergence of the overall iterative approach can potentially become an issue, and a theoretical analysis of such convergence is very hard to come by.

Also in case of a regularizer which does not change during iterations - such as, e.g. the standard TV or the WTV with fixed weights - proving convergence of the proposed iterated residual whiteness approach is not trivial at all and will be matter of future investigation. In the experimental section, we will provide some evidence for the convergence behaviour of the method.

With the twofold aim of facilitating and speeding up convergence, we propose to use as initial iterate $x^{(0)}$ for our approach, not the degraded observation b, but a higher quality image (i.e., less blurred and noisy) which can be obtained very efficiently by applying the exact residual whiteness principle outlined in Sect. 4. In particular, given the linear operators H and L, we compute $x^{(0)}$ as the solution of the Tikhonov-regularized model (4.1), with v the null vector and the regularization parameter $\mu^{(0)}$ selected by minimizing the residual whiteness function, that is, in formula,

$$(5.14) \qquad \mu^{(0)} = \arg \min_{\mu \in \mathbb{R}^+} W(\mu) \quad \text{with } W \text{ in } (4.6)\text{--}(4.9) \text{ and } v = 0_m,$$

$$(5.15) \qquad x^{(0)} = \left(\mu^{(0)} \mathrm{H}^T \mathrm{H} + \mathrm{L}^T \mathrm{L} \right)^{-1} \left(\mu^{(0)} \mathrm{H}^T b \right).$$

The main steps of the overall proposed IRWP-ADMM approach are summarized in Alg. 1.

Finally, we notice that the residual whiteness principle problem in (5.8) is addressed by applying the Newton-Raphson method to $W'(\gamma) = 0$, using as initial guess the γ value selected at previous ADMM iteration. Notice that properties (iii) and (iv) in Proposition 4.1 guarantee that, upon suitable conditions on the data b and on vector $v^{(k+1)}$ in (5.7), W admits a minimum in $\mathbb{R}^+$. If at iteration k of Alg. 1 neither condition (iii) nor (iv) is satisfied, one can simply keep γ unchanged, i.e. $\gamma^{(k+1)} = \gamma^{(k)}$.

REMARK 5.1. The iterative approach outlined in Alg. 1 can be easily modified such that the discrepancy principle is used in place of the proposed residual whiteness principle. In fact,

---

Algorithm 1: Proposed parameter-free IRWP-ADMM approach for image restoration based on variational models of the form (2.1) under assumptions ($\mathcal{A}$1)-($\mathcal{A}$4)

---

    inputs: observed degraded image $b \in \mathbb{R}^n$,

                 blur and regularization convolution operators $H, L_j \in \mathbb{R}^{n \times n}$, $j \in \{1, \ldots, d\}$

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

   1. **initialize:** compute $x^{(0)}$ by (5.14)-(5.15), then $t^{(0)} = Lx^{(0)}$

   2. **for** $k = 0, 1, 2, \ldots$ *until convergence* **do**:

   3.   · compute $\gamma^{(k+1)} = \mu^{(k+1)}/\beta$ by (5.7)-(5.8)

   4.   · compute $x^{(k+1)}$ by (5.9)

   5.   · compute $\alpha^{(k+1)}$ by (13) in [4] $\left[\,\text{only for WTV}\,\right]$

   6.   · compute $t^{(k+1)}$ by solving (5.10) $\left[\,\text{(5.11)-(5.13) for TV and WTV}\,\right]$

   7.   · compute $\lambda^{(k+1)}$ by (5.6)

   8. **end for**

   9. $x^* = x^{(k+1)}$

---

only the γ-update step at line 3 of Alg. 1 - more precisely, only formula (5.8) - is affected by the adopted parameter selection criterion. In particular, based on Remark 4.2, in order to use the discrepancy principle it suffices to replace (5.8) with the following γ-update rule:

$$(5.16) \quad \gamma^{(k+1)} = \text{root of equation}: \ D(\gamma) = 0, \quad \text{with } D \text{ in (4.11) and } v = v^{(k+1)}.$$

Analogously to (5.8) where $\gamma^{(k+1)}$ is computed by solving the non-linear equation $W'(\gamma) = 0$ through the Newton-Raphson algorithm, $\gamma^{(k+1)}$ in (5.16) can only be computed through an iterative root-finding algorithm. It follows that the proposed iterative framework outlined in Alg. 1 presents approximately the same computational cost when coupled with the residual whiteness principle or with the discrepancy principle - i.e., with (5.8) or (5.16).

6. Numerical results. In this section, we evaluate experimentally the performance of the proposed RWP-based automatic procedure for selecting the regularization parameter μ in image restoration variational models of the form (2.1) under assumption ($\mathcal{A}$1)-($\mathcal{A}$4). More specifically, we consider the TIK-$\mathcal{L}_2$, TV-$\mathcal{L}_2$ and WTV-$\mathcal{L}_2$ variational models, defined as in (2.1) with the TV, TIK and WTV regularizers given in (2.8), (2.9) and (2.10), respectively, all based on the first-order finite difference regularization matrix $D^{(1)}$ in (2.7). For the TIK-$\mathcal{L}_2$ model, we apply the exact RWP-based approach described in Sect. 4, whereas for the non-quadratic TV-$\mathcal{L}_2$ and WTV-$\mathcal{L}_2$ models we use the IRWP-ADMM iterative method outlined in Alg. 1. The proposed RWP-based approach is compared with the DP-based parameter selection strategy, defined by criterion (1.4) with $\tau = 1$ and σ the true noise standard deviation and implemented in accordance with Remarks 4.2 and 5.1.

The goal of this numerical section is twofold. First, we aim at highlighting that the proposed RWP defined in (3.10)-(3.12) is capable of selecting a regularization parameter value μ^* - which, we recall, is the one minimizing the whiteness function $W(\mu)$ - yielding high quality restorations. Then, for non-quadratic models it is equally important to prove that the proposed IRWP-ADMM approach outlined in Alg. 1 allows to automatically select such μ^* in a very robust and efficient way.

For the three considered variational models there is a one-to-one relationship between the μ-value and the norm of the associated residual image. Hence, in all the presented results we will substitute the μ-values with the corresponding τ-values, with τ defined according to (1.4) by

$$(6.1) \qquad \tau^*(\mu) := \frac{\|\mathrm{H}x^*(\mu) - b\|_2}{\sqrt{n}\sigma} .$$

The quality of the restorations x^*, for different values of τ^*, with respect to the original uncorrupted image x, will be assessed by means of two scalar measures, namely the Improved Signal-to-Noise Ratio (ISNR),

$$\mathrm{ISNR}(b, x, x^*) := 10 \log_{10} \frac{\|b - x\|_2^2}{\|x^* - x\|_2^2},$$

and the Structural Similarity Index (SSIM) [22]. The larger the ISNR and SSIM values, the higher the quality of the restoration.

For all tests, iterations of the IRWP-ADMM approach in Alg. 1 are stopped as soon as

$$\delta_x^{(k)} := \frac{\left\|x^{(k)} - x^{(k-1)}\right\|_2}{\left\|x^{(k-1)}\right\|_2} < 10^{-5}, \quad k \in \mathbb{N} \setminus \{0\},$$

and the ADMM penalty parameter β has been set manually. It is also worth remarking that, as far as the minimization of function W in (5.8) is concerned, we observed that at each iteration of Alg. 1 a global minimum exists and it is always attained, i.e. Newton-Raphson algorithm for finding a zero of W' converges in few iterations (typically around 10).

As a first example, we consider the `skyscraper` (256×256) test image, with pixel values normalized in the range [0,1], corrupted by AWG noise of standard deviation $\sigma = 0.05$ and space-invariant Gaussian blur defined by a convolution kernel generated using the Matlab routine `fspecial` with parameters `band`$= 5$ and `sigma`$= 1$. The `band` parameter represents the side length (in pixels) of the square support of the kernel, whereas `sigma` is the standard deviation (in pixels) of the isotropic bivariate Gaussian distribution defining the kernel in the continuous setting. The original and corrupted images are shown in Fig. 6.1a and Fig. 6.1b, respectively.

The black curves shown in Figs. 6.2a, 6.2c, 6.2e represent the residual whiteness functions $W(\mu)$ as defined in (3.11)-(3.12) - with μ replaced by $\tau^*(\mu)$ defined in (6.1) - for the three considered restoration models. They have been experimentally 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. The optimal τ according to the proposed RWP - which, we recall, corresponds to the global minimizer of W - is depicted by the vertical solid magenta lines, while the vertical dashed black lines correspond to $\tau = 1$, that is to the classical DP. It is worth observing that, independently from the selected regularizer, W has a global minimizer over the considered domain. The ISNR and SSIM curves for different values of τ are plotted in Figs. 6.2b, 6.2d, 6.2f, where the vertical lines have the same meaning as in the first column figures. One can notice that the RWP tends to select a value for τ that maximizes the ISNR rather than the SSIM. However, the choice of a more flexible and effective regularizer, such as the WTV, not only increases the values of the attained quality indices with respect to TIK and TV, but at the same time reduces the mismatching between the ISNR and SSIM curves. More importantly, in this case the proposed RWP selects a τ-value yielding nearly optimal restoration quality, higher than the quality achieved by using the DP .

It is now crucial to verify that the proposed IRWP-ADMM scheme outlined in Alg. 1 succeeds in automatically selecting such optimal τ in a robust and efficient way. To this

(a) Original `skyscraper`.

(b) Corrupted `skyscraper`.

(c) Original `geometric`.
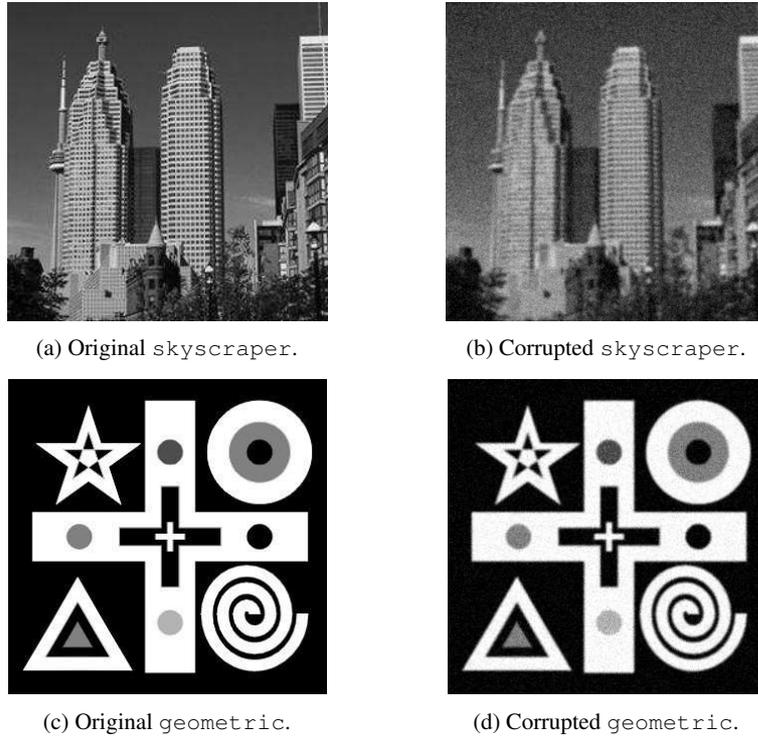

(d) Corrupted `geometric`.

Fig. 6.1: Original (first column) test images `skyscraper` and `geometric` and corresponding observed data (second column) corrupted by space-invariant Gaussian blur with parameters `band`= 5, `sigma`= 1, and AWG noise of standard deviation $\sigma = 0.05$.

purpose, the output τ of the iterative scheme is indicated with a dashed green line in Figs. 6.2c, 6.2e, 6.2d, 6.2f . The blue and red dots in Figs. 6.2d, 6.2f represent the final ISNR and SSIM values, respectively, of the image restored via IRWP-ADMM. We remark that the comparison can be carried out only for TV-$\mathcal{L}_2$ and WTV-$\mathcal{L}_2$ models, since the solution of TIK-$\mathcal{L}_2$ is in closed-form - see Sect. 4. It is easy to observe that, although in the IRWP-ADMM framework the RWP is actually applied to one of the sub-problems of the ADMM and not to the original problem, the algorithm still returns a τ which is very close, if not coincident, with the optimal τ detected *a posteriori* via the RWP.

The restored images for the three considered models solved by the IRWP-ADMM approach in Alg. 1 are shown in Fig. 6.3. The very low optimal τ selected for the TIK-$\mathcal{L}_2$ model is such that the noise is not completely removed - see Figs. 6.3a, 6.3e. When more flexible regularizers are used, the optimal τ increases and the noise is better removed, while preserving texture - see Figs. 6.3c, 6.3g.

The ISNR and SSIM values corresponding to the restored images in Fig. 6.3 are reported in Table 6.1.

As a second example, we consider the piece-wise constant test image `geometric` (320×320) in Fig. 6.1c, corrupted by the same blur and noise degrading the `skyscraper` test image in the first example. The residual whiteness functions for the three considered models are shown in Figs. 6.4a, 6.4c, 6.4e. Notice that, also in this case, the W functions
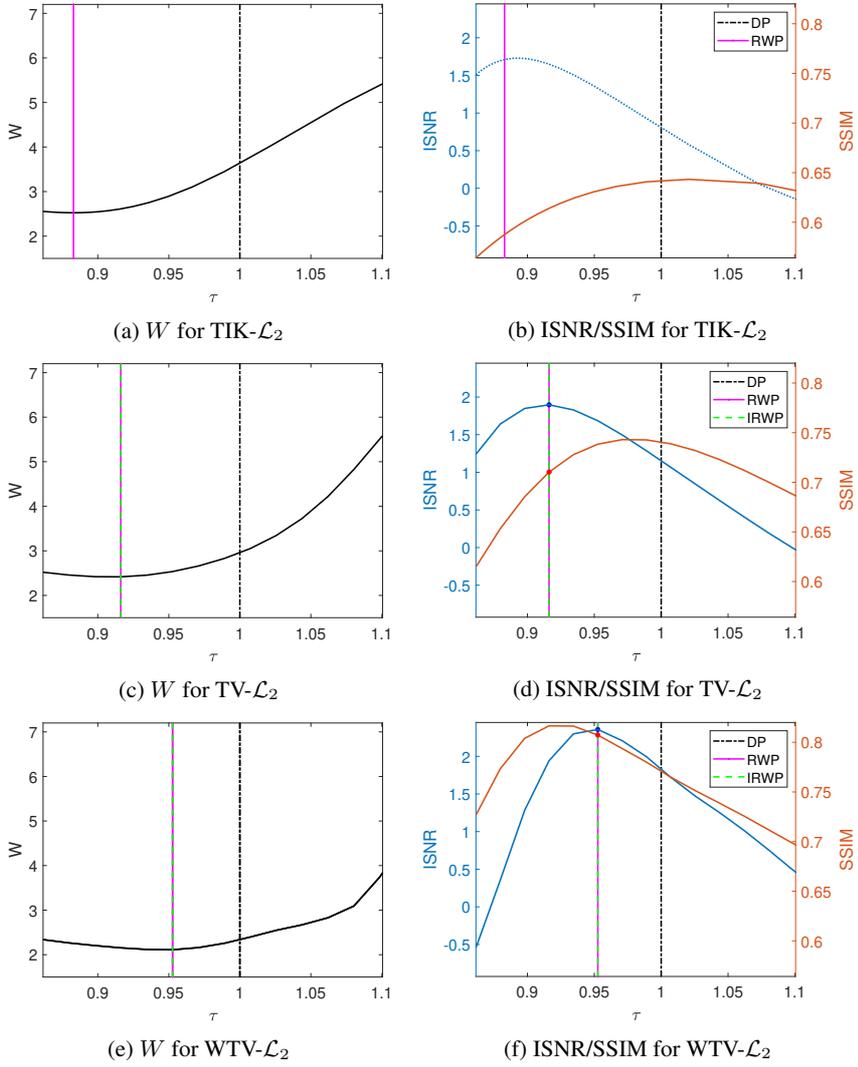
(a) W for TIK-$\mathcal{L}_2$

(b) ISNR/SSIM for TIK-$\mathcal{L}_2$

(c) W for TV-$\mathcal{L}_2$

(d) ISNR/SSIM for TV-$\mathcal{L}_2$

(e) W for WTV-$\mathcal{L}_2$

(f) ISNR/SSIM for WTV-$\mathcal{L}_2$

Fig. 6.2: *Test image* `skyscraper`. Whiteness measure functions for the three variational models performed (first column) and ISNR/SSIM values for different τs (second column).

	skyscraper			geometric		
	TIK-$\mathcal{L}_2$	TV-$\mathcal{L}_2$	WTV-$\mathcal{L}_2$	TIK-$\mathcal{L}_2$	TV-$\mathcal{L}_2$	WTV-$\mathcal{L}_2$
ISNR	1.7144	1.8967	2.3567	1.8287	8.1858	9.5665
SSIM	0.5886	0.7103	0.8074	0.5584	0.9422	0.9567

Table 6.1: ISNR and SSIM values achieved at convergence by the three considered variational models coupled with the proposed RWP and solved by the IRWP-ADMM approach in Alg. 1.
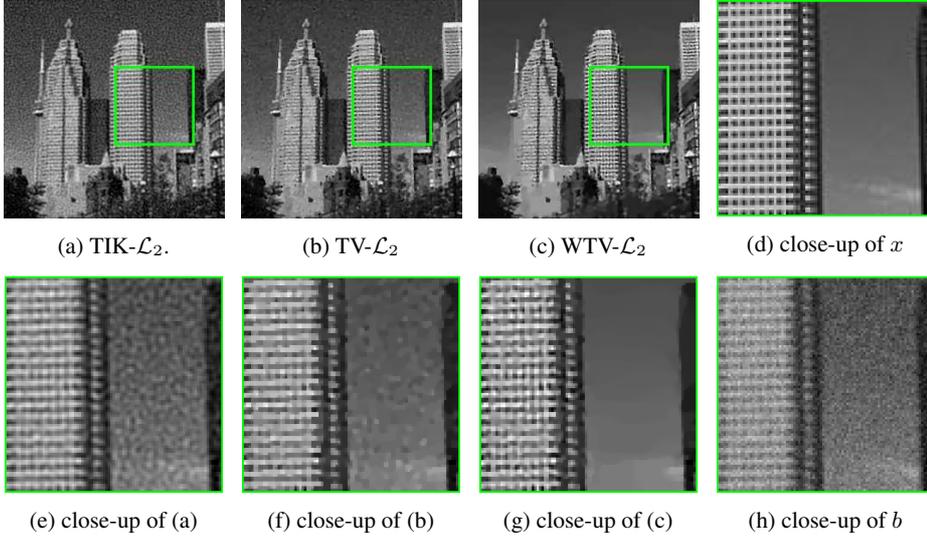
| (a) TIK-$\mathcal{L}_2$. | (b) TV-$\mathcal{L}_2$ | (c) WTV-$\mathcal{L}_2$ | (d) close-up of x |

| (e) close-up of (a) | (f) close-up of (b) | (g) close-up of (c) | (h) close-up of b |

Fig. 6.3: Restorations of test image `skyscraper` via TIK-$\mathcal{L}_2$ (a), TV-$\mathcal{L}_2$ (b) and WTV-$\mathcal{L}_2$ (c) coupled with RWP, and corresponding close-ups.

exhibit global minimizers over the considered domain for τ. In the second column of Fig. 6.4, the ISNR and SSIM curves for the models of interest are plotted. In particular, one can observe that the very low values reported in Fig. 6.4b reflect the weakness of Tikhonov regularizer when dealing with piece-wise constant images. As expected, TV-type regularizers return higher quality indices. Also in this case, when comparing the RWP with the IRWP, the resulting optimal τs are very close to each other and, in addition, they simultaneously maximize, at least approximately, both the ISNR and SSIM indices. Also in this case, at each iteration of the IRWP-ADMM approach we observed the existence of a global minimizer for the residual whiteness function of the quadratic ADMM sub-problem. The quality indices of the restorations computed via the IRWP-ADMM approach are reported in Table 6.1, while the corresponding restored images are shown in Fig. 6.5. Comparing the restorations with the original image - which, we notice, is affected by lossy compression artifacts - one can observe that the WTV-$\mathcal{L}_2$ model (see Figs. 6.5c, 6.5g) outperforms the TV-$\mathcal{L}_2$ model also in this second example.

As a further analysis of the proposed IRWP-ADMM approach for the WTV-$\mathcal{L}_2$ model applied to the restoration of the `geometric` test image, in Fig. 6.6 we plot the behavior of some quantities of interest along the ADMM iterations, namely the regularization parameter μ (Fig. 6.6a), the residual whiteness function W (Fig. 6.6d), the relative change δ_x (Fig. 6.6b), the discrepancy coefficient τ (Fig. 6.6e), the ISNR (Fig. 6.6c) and the SSIM (Fig. 6.6f) values. Notice that, at each iteration of the IRWP-ADMM algorithm applied to the solution of the WTV-$\mathcal{L}_2$ model, the updates of the regularization parameter μ, of the primal variables u, t, of the dual variable λ and, finally, of the weights α_i in the WTV regularizer are performed. Despite this chain of nested iterative updates, all the monitored quantities present a convergence behavior characterized by a smooth and monotonic (after the very first iterations) trend.

7. Conclusions. We proposed an automatic strategy for the selection of the regularization parameter μ based on the Residual Whiteness Principle. The RWP has been applied for the
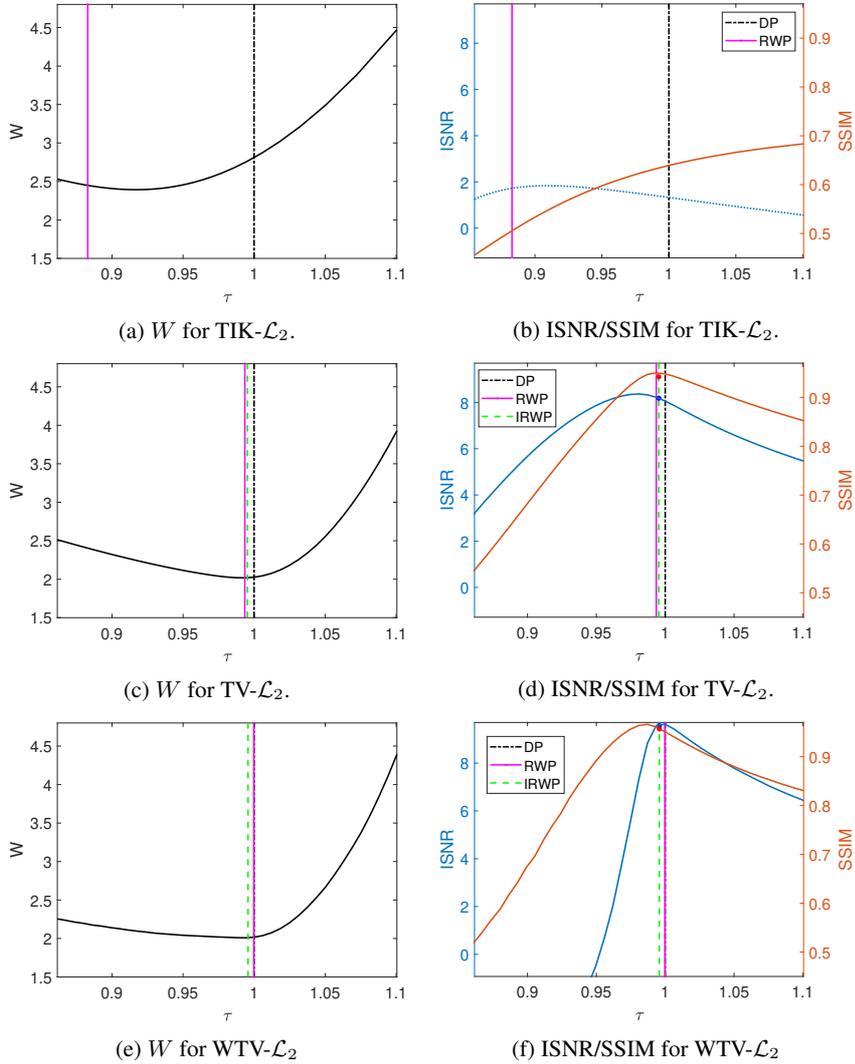
Fig. 6.4: *Test image* `geometric`. Whiteness measure functions for the three variational models performed (first column) and ISNR/SSIM values for different τs (second column).

solution of TIK-$\mathcal{L}_2$, TV-$\mathcal{L}_2$ and WTV-$\mathcal{L}_2$ when the only information available on the corrupting additive Gaussian noise is its whiteness. Nonetheless, it has been highlighted that the RWP can be easily combined with a wider class of variational models, and it can be also efficiently implemented in the case of easy-proximable regularization terms. Numerically, the TV-$\mathcal{L}_2$ and the WTV-$\mathcal{L}_2$ have been solved by means of an ADMM-based iterative scheme. Even when compared to standard parameter estimation techniques based on the knowledge of the noise level, such as the DP, the proposed approach outperforms the competitor both in terms of ISNR and SSIM. The convergence plot of μ, together with the behavior of other monitored quantities along the ADMM iterations, confirms the robustness of our method.
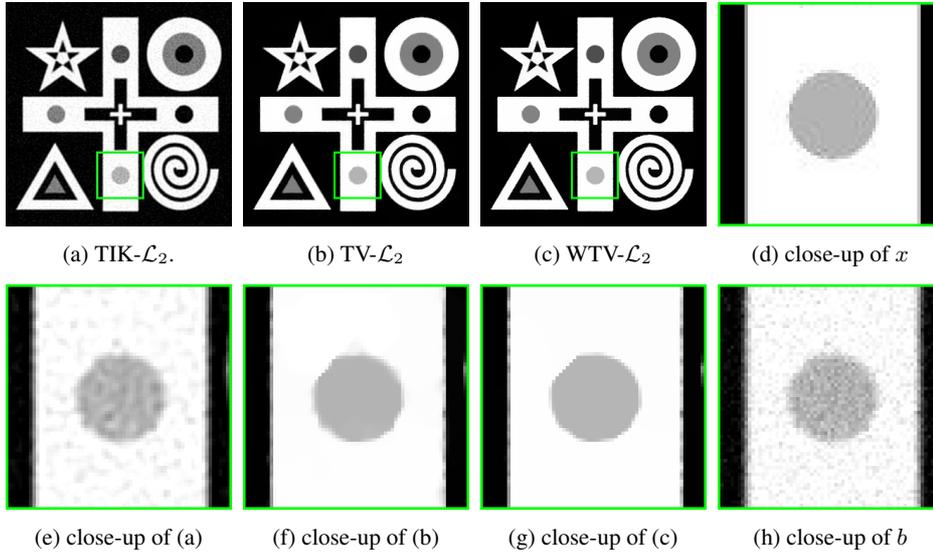
(a) TIK-$\mathcal{L}_2$.   (b) TV-$\mathcal{L}_2$ (c) WTV-$\mathcal{L}_2$ (d) close-up of x

(e) close-up of (a)   (f) close-up of (b)   (g) close-up of (c)   (h) close-up of b

Fig. 6.5: Restorations of test image `geometric` via TIK-$\mathcal{L}_2$ (a), TV-$\mathcal{L}_2$ (b) and WTV-$\mathcal{L}_2$ (c) coupled with RWP, and corresponding close-ups.

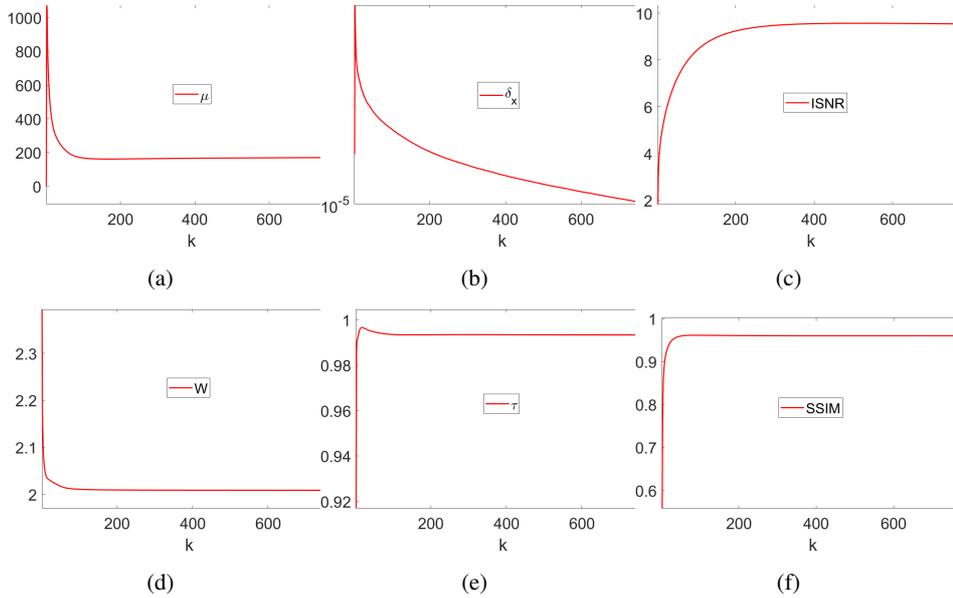(a)   (b)   (c)

(d)   (e)   (f)

Fig. 6.6: Some convergence plots for the proposed IRWP-ADMM approach outlined in Alg. 1 applied to restoring the test image `geometric` via the WTV-$\mathcal{L}_2$ variational model.

Appendix.

Proof of Proposition 4.1. .

(i) We start recalling the analytic expression of $W(\mu)$,

$$W(\mu) = \frac{\displaystyle\sum_{i \in I} \left(\frac{\epsilon_i}{\mu\,\zeta_i + \eta_i} \right)^4}{\left(\displaystyle\sum_{i \in I} \left(\frac{\epsilon_i}{\mu\,\zeta_i + \eta_i} \right)^2 \right)^2} \cdot$$

Observe that, due to property (2.2), we also have,

$$\{i \in I : \eta_i = 0\} \cap \{i \in I : \zeta_i = 0\} = \emptyset.$$

Also notice that, after recalling the definition of ϵ_i in (4.7) and introducing $Q = \{i \in I : \epsilon_i = 0\}$, it holds,

$$\{i \in I : \eta_i = 0\} \subset Q.$$

Hence, the index set I can be partitioned as

$$I = S \cup Q \setminus S \cup T,$$

and function W takes the form

(7.1) $$W(\mu) = \frac{\displaystyle\sum_{i \in S} \left(\frac{\epsilon_i}{\eta_i} \right)^4 + \sum_{i \in T} \left(\frac{\epsilon_i}{\mu\,\zeta_i + \eta_i} \right)^4}{\left(\displaystyle\sum_{i \in S} \left(\frac{\epsilon_i}{\eta_i} \right)^2 + \left(\sum_{i \in T} \frac{\epsilon_i}{\mu\,\zeta_i + \eta_i} \right)^2 \right)^2} \cdot$$

Hence, W is a ratio of rational polynomial functions which is always defined except for the case $I = Q$, that is, according to (4.6) and (4.7),

(7.2) $$\mathrm{H L}^T v - \mathrm{L}^T \mathrm{L} b = 0.$$

Notice that it is unlikely that condition (7.2) is verified. For instance, it may happen when b and v are both constant images, which is highly improbable and, in any case, easily detectable. We can thus conclude that $W(\mu) \in C^\infty(\mathbb{R}^+)$, since the definition of W can be extended by continuity in $\mu = 0$.

(ii) Based on the expression in (7.1), the limits of W as μ goes to 0 or to $+\infty$ can be easily computed a follows

$$\ell_1 := \lim_{\mu \to 0^+} W(\mu) = \frac{\alpha + \displaystyle\sum_{i \in T} \left(\frac{\epsilon_i}{\eta_i} \right)^4}{\left(\beta + \displaystyle\sum_{i \in T} \left(\frac{\epsilon_i}{\eta_i} \right)^2 \right)^2}$$

$$\ell_2 := \lim_{\mu \to +\infty} W(\mu) = \frac{\alpha}{\beta^2},$$

with

$$\alpha = \sum_{i \in S} \left(\frac{\epsilon_i}{\eta_i} \right)^4 \quad \text{and} \quad \beta = \left(\sum_{i \in S} \left(\frac{\epsilon_i}{\eta_i} \right)^2 \right)^2.$$

In light of the observations in (i), ℓ_1 and ℓ_2 are strictly positive.

(iii) The only configuration in which W does not admit a minimum in $\mathbb{R}^+$ is when W monotonically decreases from ℓ_1 to the asymptotic value ℓ_2. This is clearly not allowed if $\ell_1 < \ell_2$.

(iv) Consider now the case $\ell_1 > \ell_2$. After recalling that

$$w_i = \frac{\epsilon_i}{\zeta_i \mu + \eta_i}\,, \quad i \in T\,,$$

the derivative of W reads

$$(7.3)\quad W'(\mu) = \frac{1}{\left(\beta + \sum_{i \in T} w_i(\mu)^2\right)^4} \Bigg[4\left(\sum_{i \in T} w_i^3(\mu)w_i'(\mu)\right)\left(\beta + \sum_{i \in T} w_i^2(\mu)\right)^2$$
$$- 4\left(\beta + \sum_{i \in T} w_i^2(\mu)\right)\left(\sum_{i \in T} w_i(\mu)w_i'(\mu)\right)\left(\alpha + \sum_{i \in T} w_i^4(\mu).\right)\Bigg]$$

In order to exclude the case in which W monotonically decreases over the domain, i.e. $W'(\mu) < 0\,,\ \forall \mu \in \mathbb{R}^+$, we can require $W'(\mu)$ to satisfy the following condition:

$$(7.4)\qquad\qquad \lim_{\mu \to 0^+} W'(\mu) > 0\,.$$

Imposing condition (7.4) in (7.3) is equivalent to require

$$(7.5)\qquad \left(\sum_{i \in T} w_i^3(0)w_i'(0)\right)\left(\beta + \sum_{i \in T} w_i^2(0)\right)$$
$$- \left(\sum_{i \in T} w_i(0)w_i'(0)\right)\left(\alpha + \sum_{i \in T} w_i^4(0)\right) > 0\,,$$

where

$$w_i(0) = \frac{\epsilon_i}{\eta_i}\,, \quad w_i'(0) = -\frac{\zeta_i}{\epsilon_i}w_i^2(0) = -\frac{\zeta_i}{\eta_i}\frac{\epsilon_i}{\eta_i}\,.$$

Hence, condition (7.5) turns into,

$$\left(-\sum_{i \in T} \frac{\zeta_i}{\eta_i}\left(\frac{\epsilon_i}{\eta_i}\right)^4\right)\left(\beta + \sum_{i \in T}\left(\frac{\epsilon_i}{\eta_i}\right)^2\right) + \left(\sum_{i \in T} \frac{\zeta_i}{\eta_i}\left(\frac{\epsilon_i}{\eta_i}\right)^2\right)\left(\alpha + \sum_{i \in T}\left(\frac{\epsilon_i}{\eta_i}\right)^4\right) > 0\,,$$

or equivalently,

$$(7.6)\quad \sum_{i \in T} \frac{\zeta_i}{\eta_i}\left(\frac{\epsilon_i}{\eta_i}\right)^2 \underbrace{\left[\frac{\alpha + \sum_{i \in T}\left(\frac{\epsilon_i}{\eta_i}\right)^4}{\left(\beta + \sum_{i \in T}\left(\frac{\epsilon_i}{\eta_i}\right)^2\right)^2}\right]}_{= \ell_1} - \left[\frac{\sum_{i \in T} \frac{\zeta_i}{\eta_i}\left(\frac{\epsilon_i}{\eta_i}\right)^4}{\left(\beta + \sum_{i \in T}\left(\frac{\epsilon_i}{\eta_i}\right)^2\right)}\right] > 0\,.$$

Recalling that we are considering the case $\ell_1 > \ell_2$, (7.6) is satisfied whenever

$$\sum_{i \in T} \frac{\zeta_i}{\eta_i} \left(\frac{\epsilon_i}{\eta_i} \right)^2 \frac{\alpha}{\beta^2} - \left[\frac{\sum_{i \in T} \frac{\zeta_i}{\eta_i} \left(\frac{\epsilon_i}{\eta_i} \right)^4}{\left(\beta + \sum_{i \in T} \left(\frac{\epsilon_i}{\eta_i} \right)^2 \right)} \right] > 0,$$

or equivalently, after simple algebraic manipulations,

$$(7.7) \qquad \sum_{i \in T} \frac{\zeta_i}{\eta_i} \left(\frac{\alpha}{\beta} \left(\frac{\epsilon_i}{\eta_i} \right)^2 - \left(\frac{\epsilon_i}{\eta_i} \right)^4 \right)$$
$$+ \underbrace{\frac{\alpha}{\beta^2} \left(\sum_{i \in T} \left(\frac{\epsilon_i}{\eta_i} \right)^2 \right) \left(\sum_{i \in T} \frac{\zeta_i}{\eta_i} \left(\frac{\epsilon_i}{\eta_i} \right)^2 \right)}_{> 0} > 0$$

A sufficient condition for (7.7) to be satisfied is

$$\left(\frac{\epsilon_i}{\eta_i} \right)^2 < \frac{\alpha}{\beta}, \quad \text{or equivalently} \quad \left| \frac{\tilde{z}_i}{\eta_i} - \tilde{b}_i \right|^2 < \frac{\sum_{i \in S} |\tilde{b}_i|^4}{\sum_{i \in S} |\tilde{b}_i|^2}, \quad \forall i \in T.$$

$\square$

REFERENCES

[1] M.S.C. Almeida and M.A.T. Figueiredo. Parameter estimation for blind and non-blind deblurring using residual whiteness measures. *IEEE Transactions on Image Processing*, 22:2751–2763, 2013.

[2] A. Buccini, Y. Park, and L. Reichel. Comparison of a-posteriori parameter choice rules for linear discrete ill-posed problems. 2019.

[3] L. Calatroni, A. Lanza, M. Pragliola, and F. Sgallari. A flexible space-variant anisotropic regularization for image restoration with automated parameter selection. *SIAM Journal on Imaging Sciences*, 12:1001–1037, 2019.

[4] L. Calatroni, A. Lanza, M. Pragliola, and F. Sgallari. Adaptive parameter selection for weighted-tv image reconstruction problems. To appear on Journal of Physics: Conference Series.

[5] D. Calvetti, B. Lewis, and L. Reichel. L-curve for the minres method. *Advanced Signal Processing Algorithms, Architectures, and Implementations X*, 4116:385–395, 2000.

[6] P.G. Craven and G. Wahba. Smoothing noisy data with spline functions: Estimating the correct degree of smoothing by the method of generalized cross-validation. *Numerische mathematik*, 31:377–403, 1978.

[7] H.W. Engl, M. Hanke, and A. Neubauer. Regularization of inverse problems. 1996.

[8] C. Fenu, L. Reichel, and G. Rodriguez. Gcv for Tikhonov regularization via global golub-kahan decomposition. *Numerical Lin. Alg. with Applic.*, 23:467–484, 2016.

[9] P.C. Hansen. Rank-deficient and discrete ill-posed problems. *SIAM*, 4, 1998.

[10] P.C. Hansen, T.K. Jensen, and Giuseppe Rodriguez. An adaptive pruning algorithm for the discrete L-curve criterion. *Journal of computational and applied mathematics*, 198:483–492, 2007.

[11] P.C. Hansen, M.E. Kilmer, and R.H. Kjeldsen. Exploiting residual information in the parameter choice for discrete ill-posed problems. *BIT Numerical Mathematics*, 46:41–59, 2006.

[12] M.E. Hochstenbach, L. Reichel, and G. Rodriguez. Regularization parameter determination for discrete ill-posed problems. *J. Computational Applied Mathematics*, 273:132–149, 2014.

[13] A. Lanza, S. Morigi, F. Sciacchitano, and F. Sgallari. Whiteness constraints in a unified variational framework for image restoration. *Journal of Mathematical Imaging and Vision*, 60:1503–1526, 2018.

[14] A. Lanza, S. Morigi, and F. Sgallari. Variational image restoration with constraints on noise whiteness. *Journal of Mathematical Imaging and Vision*, 53:61–67, 2015.

[15] A. Lanza, S. Morigi, F. Sgallari, and A.J. Yezzi. Variational image denoising based on autocorrelation whiteness. *SIAM Journal on Imaging Sciences*, 6:1931–1955, 2013.

[16] V.A. Morozov. On the solution of functional equations by the method of regularization. *Dokl. Akad. Nauk SSSR*, 167:510–512, 1966.

[17] Y. Park, L. Reichel, G. Rodriguez, and X. Yu. Parameter determination for Tikhonov regularization problems in general form. *J. Computational Applied Mathematics*, 343:12–25, 2018.

[18] L. Reichel and G. Rodriguez. Old and new parameter choice rules for discrete ill-posed problems. *Numerical Algorithms*, 63:65–87, In press.

[19] L.I. Rudin, S. Osher, and E. Fatemi. Nonlinear total variation based noise removal algorithms. *Physica D: nonlinear phenomena*, 60:259–268, 1992.

[20] B. W. Rust and D. P. O'Leary. Residual periodograms for choosing regularization parameters for ill-posed problems. *Inverse Problems*, 24:034005, 2008.

[21] A.M. Stuart. Inverse problems: A bayesian perspective. *Acta Numer.*, 19:451–559, 2010.

[22] Z. Wang, A.C. Bovik, H.R. Sheikh, and E.P. Simoncelli. Image quality assessment: From error visibility to structural similarity. *IEEE Transactions on Image Processing*, 4:600–612, 2004.