SIAM Review
Vol. 65, No. 3, pp. 601–685

On and Beyond Total Variation Regularization in Imaging: The Role of Space Variance*

Monica Pragliola†
Luca Calatroni‡
Alessandro Lanza§
Fiorella Sgallari§

Abstract. Over the last 30 years a plethora of variational regularization models for image reconstruction have been proposed and thoroughly inspected by the applied mathematics community. Among them, the pioneering prototype often taught and learned in basic courses in mathematical image processing is the celebrated Rudin–Osher–Fatemi (ROF) model [L. I. Rudin, S. Osher, and E. Fatemi, *Phys. D*, 60 (1992), pp. 259–268], which relies on the minimization of the edge-preserving total variation (TV) seminorm as a regularization term. Despite its (often limiting) simplicity, this model is still very much employed in many applications and used as a benchmark for assessing the performance of modern learning-based image reconstruction approaches, thanks to its thorough analytical and numerical understanding. Among the many extensions to TV proposed over the years, a large class is based on the concept of *space variance*. Space-variant models can indeed overcome the intrinsic inability of TV to describe *local* features (strength, sharpness, directionality) by means of an adaptive mathematical modeling which accommodates local regularization weighting, variable smoothness, and anisotropy. Those ideas can further be cast in the flexible Bayesian framework of generalized Gaussian distributions and combined with maximum likelihood and hierarchical optimization approaches for efficient hyperparameter estimation. In this work, we review and connect the major contributions in the field of space-variant TV-type image reconstruction models, focusing, in particular, on their Bayesian interpretation which paves the way to new exciting and unexplored research directions.

Key words. mathematical image reconstruction, total variation regularization, space-variant regularization, generalized Gaussian Bayesian modeling, hierarchical modeling, ADMM

MSC codes. 68U10, 94A08, 65K10

DOI. 10.1137/21M1410683

Contents

*Received by the editors April 7, 2021; accepted for publication (in revised form) April 25, 2022; published electronically August 8, 2023.
https://doi.org/10.1137/21M1410683
Funding: The work of the first, third, and fourth authors was supported by the "National Group for Scientific Computation (GNCS-INDAM)" and by ex60 project by the University of Bologna. The work of the second author was supported by the EU H2020 RISE NoMADS, GA 777826, and the UCA JEDI IDEX grant DEP.

†Department of Mathematics and Applications, University of Naples Federico II, Naples, Italy (monica.pragliola@unina.it).
‡CNRS, UCA, INRIA, Laboratoire I3S, 06903, Sophia-Antipolis, France (calatroni@i3s.unice.fr).
§Department of Mathematics, University of Bologna, Bologna, Italy (alessandro.lanza2@unibo.it, fiorella.sgallari@unibo.it).

1. Introduction. The technological developments which favored the storage, the exploitation, and the management of large and unstructured data over the last few decades have been responsible for significant and fundamental advances in the field of applied mathematics. In particular, the field of mathematical image processing has undergone significant changes in its paradigm, which has been shifted several times over the years. Historically, imaging problems have been formulated as specific instances of (linear) ill-posed inverse problems and studied by means of classical tools such as functional calculus, partial differential equations (PDEs), and Fourier analysis. Later in the 1990s, advances in the field of nonsmooth variational calculus have drifted the attention towards the use of sparsity-promoting image regularization models as well as the development of efficient optimization algorithms tailored to compute the desired output as efficiently as possible. Over the last decade, a new class of models has attracted the attention of the applied mathematics community. Differently from the traditional formulation of imaging problems described above, whose ingredients are chosen a priori following analytical, model-driven strategies, by capitalizing on the improved technological advances, these new data-driven approaches exploit the large availability of imaging data and design a posteriori image reconstruction models tailored to fit the specific application at hand. Understanding the (deep) reasons behind the outstanding performance of these models is nowadays among the most (if not the most) prominent challenging tasks, with implications in fields such as artificial intelligence, human-to-robot interactions, and bio-inspired computer designs. In addition, the clever and efficient exploitation of training data has favored the development of new and theoretically grounded branches of applied mathematics lying at the interface between analysis, variational calculus, and statistics. Interestingly, in many imaging applications, the fusion of classical and modern approaches has been in fact capable of overcoming the intrinsic difficulties and rigidities of fully model-driven methods by incorporating appropriately data-driven information.

In this spirit, we present in the following a scientific journey across disciplines taking as example a standard problem in the context of mathematical image reconstruction in the attempt to highlight for the popular and well-studied total variation (TV) regularization model some of the many extensions proposed over the years featuring as least common denominator the description of local image features at a pixel scale. In the attempt to combine classical tools of variational calculus, optimization, and numerical analysis with more data-driven large-scale statistical approaches, we introduce a new flexible Bayesian interpretation of the imaging quantities into play and report on how such a happy marriage can be efficiently exploited as a powerful tool for the exploration of new research directions in imaging.

With the intent of providing an as-exhaustive-as-possible review on the topic, we will start our discussion by recalling in the following introductory sections the main characters in our play, providing appropriate referencing and illustrations which, we hope, will help familiarize the inexpert reader with the main notions introduced.

1.1. Imaging Inverse Problems. We start our discussion by setting up the scene. To do so, we consider the formulation of a general image reconstruction problem defined on an image domain $\boldsymbol{\Omega} := \{(h,l) : h = 1,\dots,n_1,\ l = 1,\dots,n_2\} \subset \mathbb{R}^2$ with $|\boldsymbol{\Omega}| = n_1 n_2 =: N$ given by

$$\text{find} \quad \boldsymbol{u} \quad \text{s.t.} \quad \boldsymbol{b} = \mathcal{N}(\mathbf{A}\boldsymbol{u})\,, \tag{1.1}$$

where $\boldsymbol{u} \in \mathbb{R}^N$ and $\boldsymbol{b} \in \mathbb{R}^M$ are the vectorized unknown image and observed data, respectively, $\mathbf{A} \in \mathbb{R}^{M\times N}$ is the (known) linear forward operator and $\mathcal{N} : \mathbb{R}^M \to$

$\mathbb{R}^M$ stands for the degradation operator modeling the presence of noise in $\boldsymbol{b}$. Some classical examples for $\mathbf{A}$ are, for instance, convolution (blurring), Radon transform, and undersampling operators. As is known from standard books in inverse problems (see, e.g., [68, 144]), it is in general not possible to solve (1.1) directly due to the lack of stability and/or uniqueness properties of the operators involved. As a remedy, problem (1.1) can be reformulated as the problem of finding an estimate $\boldsymbol{u}^*$ of $\boldsymbol{u}$ as accurate as possible by solving a new, well-posed problem where some a priori information on $\boldsymbol{u}^*$ is encoded in the form of a regularization term. In this work, we will focus our attention on the family of variational regularization methods where the reconstructed image $\boldsymbol{u}^* \in \mathbb{R}^N$ is computed as a minimizer of a suitable cost functional $\mathcal{J} : \mathbb{R}^N \to \mathbb{R}$ such that the problem can be formulated as

$$\text{find } \boldsymbol{u}^* \in \underset{\boldsymbol{u}\in\mathbb{R}^N}{\arg\min} \Big(\mathcal{J}(\boldsymbol{u};\mu) := \mathcal{R}(\boldsymbol{u}) + \mu\mathcal{F}(\mathbf{A}\boldsymbol{u};\boldsymbol{b})\Big). \tag{1.2}$$

The functionals $\mathcal{R}$ and $\mathcal{F}$ are commonly referred to as the *regularization* and the *data fidelity* term, respectively. While $\mathcal{R}$ encodes prior information on the desired image $\boldsymbol{u}$ (such as, e.g., its regularity and/or its sparsity patterns), the term $\mathcal{F}$ measures the "distance" between the given image $\boldsymbol{b}$ and $\boldsymbol{u}$ after the action of the operator $\mathbf{A}$ with respect to some functional describing noise statistics in the data. Finally, the regularization parameter $\mu > 0$ controls the trade-off between the two terms.

1.2. A Leading Actor: TV Regularization. Features, Drawbacks, and Limitations. Probably (if not certainly) the most popular choice in the context of imaging for $\mathcal{R}$ in (1.2) is the TV seminorm which is suited for describing meaningful image contents such as image discontinuities (edges). In its simplest discrete form, it is defined as the following nonsmooth and convex regularizer:

$$\mathcal{R}(\boldsymbol{u}) \;=\; \mathrm{TV}(\boldsymbol{u}) \;:=\; \sum_{i=1}^{N} \|(\mathbf{D}\boldsymbol{u})_i\|_2\,, \tag{TV}$$

where, for each pixel $i = 1,\dots,N$, $(\mathbf{D}\boldsymbol{u})_i = (\mathbf{D}_h\boldsymbol{u}, \mathbf{D}_v\boldsymbol{u})_i \in \mathbb{R}^2$ stands for the discrete gradient of image $\boldsymbol{u}$ at pixel i, with $\mathbf{D}_h, \mathbf{D}_v \in \mathbb{R}^{N\times N}$ suitable finite difference operators discretizing the first-order partial derivatives of image $\boldsymbol{u}$ along the horizontal and vertical directions, respectively.

The use of TV regularization in imaging was first proposed by Rudin, Osher, and Fatemi in [134], which is nowadays probably among the most cited papers in mathematical imaging.[1] In the 1990s, the use of TV in the context of imaging paved the way for the development of mathematical approaches based on the use of nonlinear, edge-preserving, and sparse gradient-based regularization models and for their application in a variety of image reconstruction problems. Analytically, the fine properties of TV in the context of image reconstruction have been thoroughly studied and understood over the years (see, e.g., [40, 38, 23] for a review), and efficient algorithmic approaches have been developed for the efficient numerical solution of TV-based problems (see, e.g., the recent review [45]).

Despite generating much interest in and greater understanding of the analytical and regularization properties of TV regularization (we will list the main contributions in both directions in due course), such a regularizer also presents significant limitations leading to undesirable artifacts even in the case of uncorrupted observed images. A

[1] 18,319 citations according to Google Scholar.

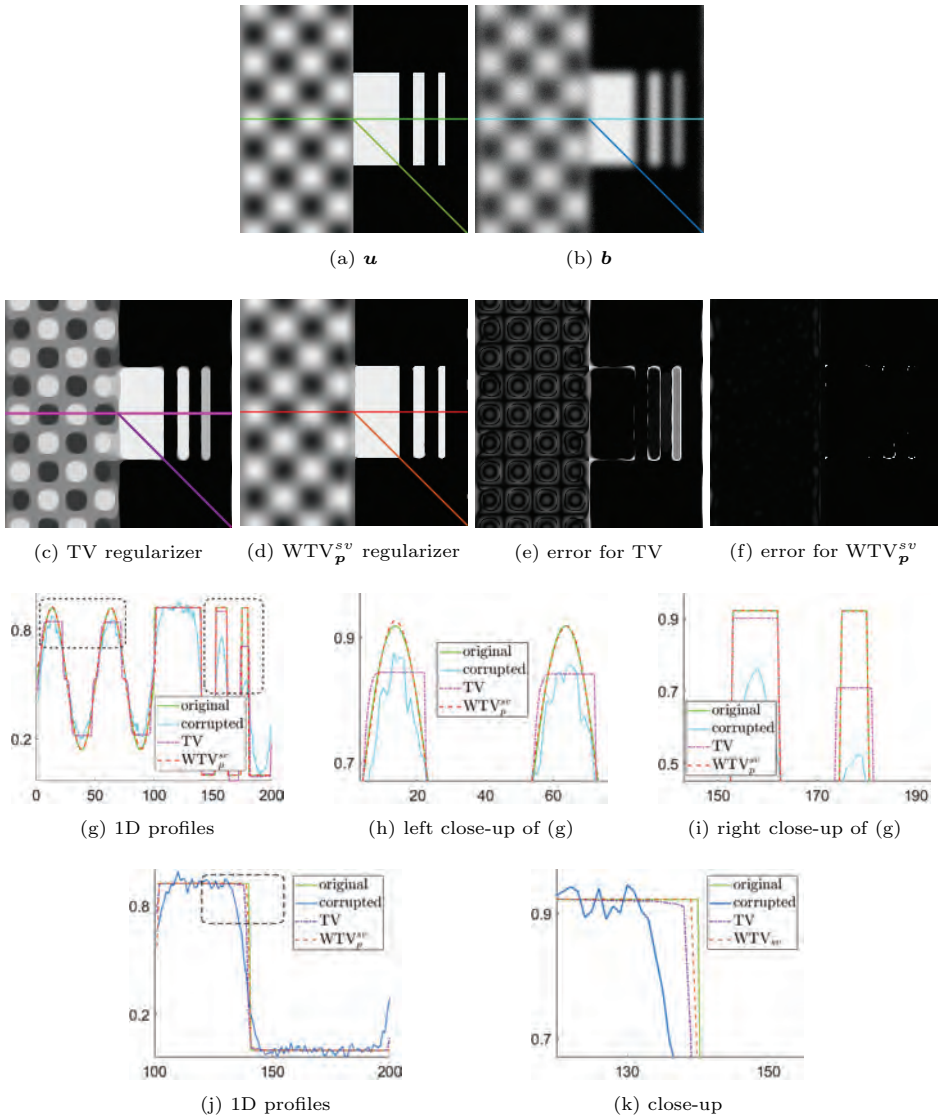
major one is the so-called *staircasing* effect, which consists of a tendency to promote edges at the expense of smooth structures; see, e.g., [39, 119, 88] for some analytical studies. Moreover, as observed, e.g., in [143, 114], TV reconstructions suffer from *loss of contrast* artifacts, which can also result in visually undesirable *rounding of corners* effects. Then, due to its dependence on the Euclidean norms of the gradients, the TV regularizer is not able to take advantage of situations where clear *directionality* appears, either at frame level [10, 93] or at neighborhood level [157]. Maybe the major limitation of TV is its *global* or *space-invariant* behavior, that is, the fact that the contribution at each pixel $i = 1, \ldots, N$ in (TV) to the whole regularization takes exactly the same functional form. Due to this "rigidity," the TV regularizer is not suited to describe possibly very heterogeneous local image structures encountered, for instance, in natural images.

In Figure 1 we provide clear visual evidence for some of the aforementioned TV limitations. We employ the TV regularizer for the restoration of the synthetic test image shown in Figure 1(a), which is composed of smooth features, in the left half, and blocky objects, in the right half, after corruption by blur and additive noise; see Figure 1(b). The restored image obtained by employing the TV regularizer shown in Figure 1(c) and the associated absolute error image reported in Figure 1(e) reveal that TV is not capable of recovering the smooth structure. Moreover, it does not succeed in restoring the original contrast of the piecewise-constant part and fails in sharpening the corners. The one-dimensional profiles plotted in Figures 1(g) and 1(j) allow us to give a closer view of the mentioned effects. In particular, from the horizontal profile in Figure 1(g), one can clearly notice the staircasing, arising in correspondence with the smooth left part—see also the close-up in Figure 1(h)—and the contrast loss, which is mostly visible in the last two rectangles on the right side of the signal; see the close-up in Figure 1(i). The rounding of corners appearing in the error image is also visible in the diagonal profile plotted in Figure 1(j) and in the corresponding close-up in Figure 1(k).

We will now review the main contributions proposed in the literature over the last few decades to improve upon the regularization capabilities of TV and, in particular, to reduce the aforementioned drawbacks within the class of space-invariant regularizers. Next, we will discuss the advantages that space-variant approaches bring along, pointing out how the notion of space variance has been used under different names in different mathematical fields.

1.3. A Partial Remedy: Space-Invariant TV Generalizations. In order to reduce some of the TV reconstruction drawbacks highlighted above, several space-invariant generalizations have been proposed over the years; see, e.g., [49, 47, 124, 139, 154] and the references therein. Within this class, we mention in particular two nowadays very popular extensions: the infimal convolution total variation (ICTV) proposed by Chambolle and Lions in [43] and the total generalized variation (TGV) regularizer introduced by Bredies, Kunisch, and Pock in [21]. This latter regularizer shares some favorable properties with TV, such as rotational invariance, lower semicontinuity, and convexity. However, differently from TV, TGV involves and balances higher-order derivatives of the desired image, which reduces staircasing, while preserving sharp edges at the same time.

Generally speaking, while the use of higher-order extensions has been shown to be very effective in practice and thoroughly analyzed from both an analytical and numerical viewpoint, the question of how to overcome the intrinsic "rigidity" of first-order TV-type regularization models is still very much open. With the intent of



(a) $\boldsymbol{u}$ (b) $\boldsymbol{b}$

(c) TV regularizer (d) $\mathrm{WTV}_{\boldsymbol{p}}^{sv}$ regularizer (e) error for TV (f) error for $\mathrm{WTV}_{\boldsymbol{p}}^{sv}$

(g) 1D profiles (h) left close-up of (g) (i) right close-up of (g)

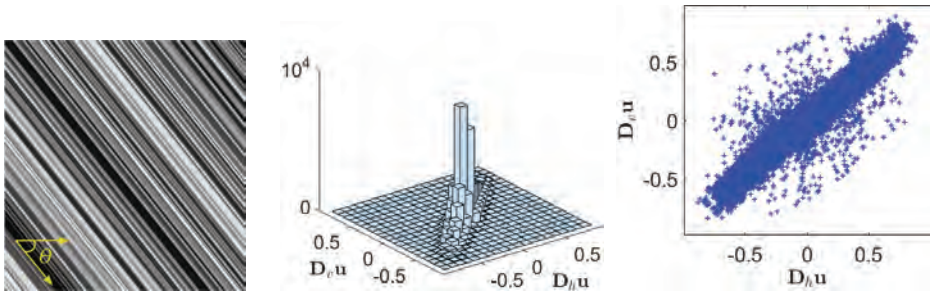
(j) 1D profiles (k) close-up

Fig. 1 Space-invariant TV regularizer vs. space-variant $\mathrm{WTV}_{\boldsymbol{p}}^{sv}$ regularizer. *Original test image* 1(a), *observed image corrupted by blur and noise* 1(b), *restoration under the adoption of the space-invariant* TV *regularizer* 1(c), *and the space-variant* $\mathrm{WTV}_{\boldsymbol{p}}^{sv}$ *regularizer* 1(d), *with the corresponding absolute error images* (×2) 1(e),1(f); *plot of one-dimensional* (1D) *profiles from the original, observed, and restored images* 1(g),1(j), *with the corresponding close-ups* 1(h),1(i),1(k).

adapting the TV-type regularization to structural image information, in [10] Bayram and Kamasak proposed a directional TV (DTV) regularizer for image denoising, whose analytic form reads

$$\text{(DTV)} \qquad \mathrm{DTV}(\boldsymbol{u}) := \sum_{i=1}^{N} \|\boldsymbol{\Lambda}_a \mathbf{R}_{-\theta} (\mathbf{D}\boldsymbol{u})_i\|_2, \qquad a \in (0,1], \qquad \theta \in [-\pi/2, \pi/2).$$

Here, $\theta \in [-\pi/2, \pi/2)$ denotes the dominant orientation in the target image $\mathbf{u}$, $\mathbf{R}_{-\theta}$



(a) (b) (c)

Fig. 2 *Test image with a clear dominant direction defined by angle* $\theta \in [\pi/2, \pi/2)$ (a)*,* $2D$ *histogram of image gradient components* (b)*, scatter plot of image gradient components* (c)*.*

denotes the rotation matrix of angle $-\theta$, while $\mathbf{\Lambda}_a = \mathrm{diag}(1, a)$ is a diagonal matrix which encodes the strength of the regularization along the direction $(\cos\theta, \sin\theta)$ and its orthogonal depending on a parameter $a \in (0, 1]$ whose value ranges from enforcing a full isotropic modeling (no directional preference) for $a = 1$ to a strong anisotropy one for $a \approx 0$. The very same modeling can be analogously used to define a directional variant of the TGV regularizer, dubbed DTGV, as done in [94] by Kongskov, Dong, and Knudsen. The use of DTV and DTGV is indeed beneficial in applications related to fibers, such as the study of glass fibers in wind-turbine blades [136] and the scanning of optical fibers from computed tomography (CT) scans [89].

To illustrate the benefits of incorporating directional information defined in terms of a dominant orientation θ in the regularizer, we consider the test image in Figure 2(a), in which the existing piecewise-constant regions as well as the smooth straight lines align along the edge direction $\mathbf{v} := (\cos\theta, \sin\theta)$. The 2D histogram of the gradients reported in Figure 2(b) and the scatter plot in Figure 2(c) show, as expected, global alignment along the perpendicular direction $\mathbf{v}^\perp = (-\sin\theta, \cos\theta)$. It appears natural for this type of image to design a regularization whose functional form could promote smoothing along the direction $\mathbf{v}$ only (i.e., anisotropically) and not indistinctly along all directions (i.e., isotropically) to better adapt to the underlying geometrical image structures, thus avoiding staircasing.

A different extension to TV regularization has been explored in [7, 95, 101]. By adopting a statistical viewpoint, in these works the authors show how the use of TV implicitly corresponds to assuming a space-invariant one-parameter half-Laplacian distribution (hLd) for the gradient magnitudes of the target image $\boldsymbol{u}$, which appears to be in general too restrictive to model the distribution of gradient magnitudes in natural images for which a more general half-generalized Gaussian distribution (hGGd) prior should be considered instead. This choice corresponds to employing the following TV_p regularization model:

$$(\mathrm{TV}_p) \qquad \mathrm{TV}_p(\boldsymbol{u}) := \sum_{i=1}^{N} \|(\mathbf{D}\boldsymbol{u})_i\|_2^p, \quad p \in (0, 2],$$

where the exponent $p \in (0, 2]$ is a free parameter providing the TV_p regularizer with higher flexibility than the TV regularizer. Its setting is indeed related to the properties of the image of interest: it can be fixed either empirically to enhance gradient-sparsity ($p < 1$, see [95]) or smoothness ($p = 2$, corresponding to Tikhonov regularization) of

solution, or regarded as information tailored to the image itself which should thus be estimated appropriately. The TV_p regularizer has proved to be effective for the solution of several imaging problems ranging from labeling and segmentation [109] to blind deblurring [95, 110] and synthetic aperture radar (SAR) image despeckling [77] and many more. Its performance strongly depends on the selected/estimated value of p, whose setting may be hard in the case of very heterogeneous images composed, for instance, by both smooth and piecewise-constant regions.

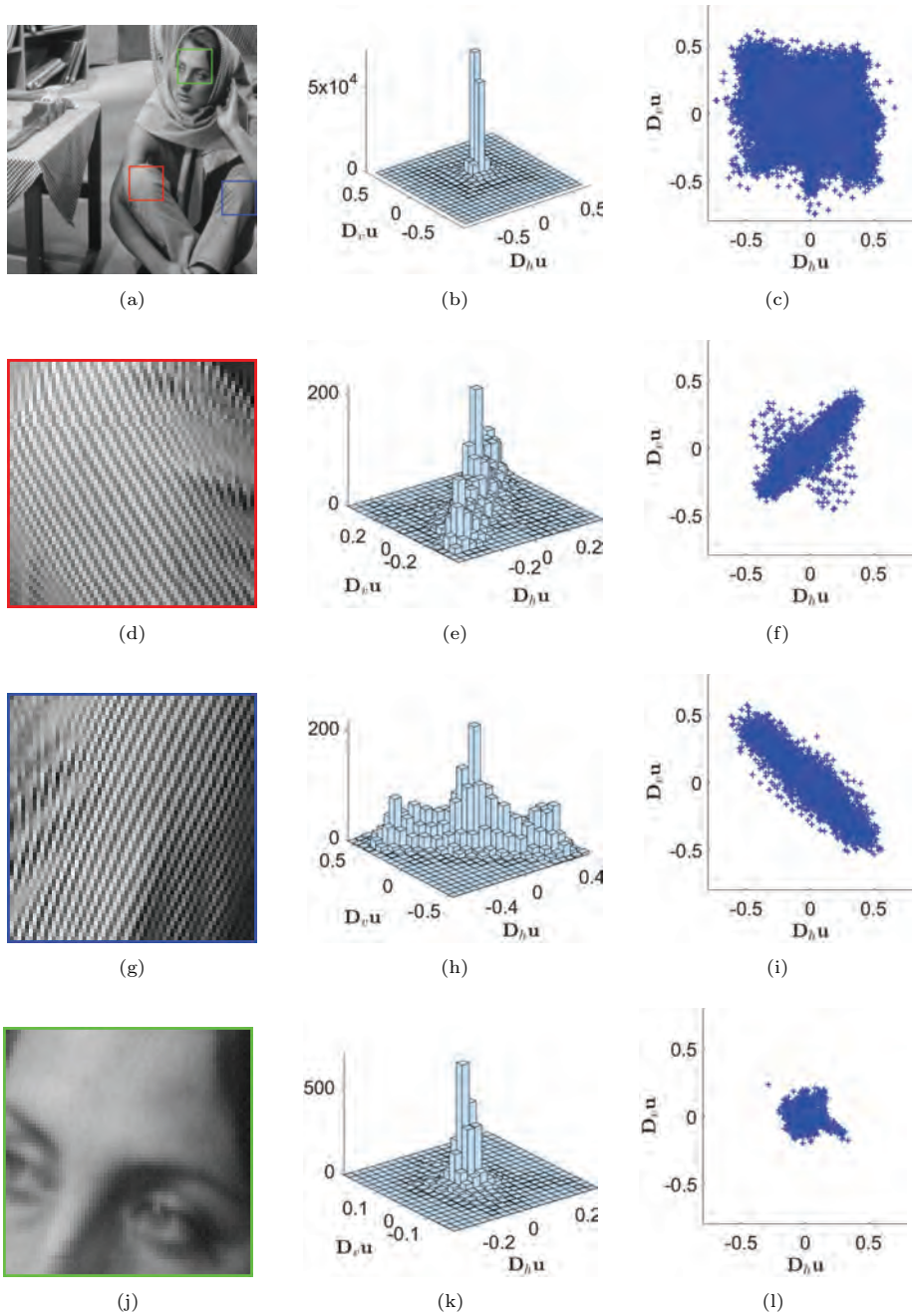
Finally, we recall that a further albeit classical limitation of global TV-type regularization consists in the choice of the optimal regularization parameter μ in (1.2). As a matter of fact, this is a common challenge for all regularized inverse problems in the form (1.2), not limited to the TV context. Among classical methods for parameter selection, we recall here those based on the discrepancy principle [78, 79], generalized cross validation [74, 71], L-curve analysis [30], and Stein's unbiased risk estimators (SURE) [142, 105]. More recent approaches based, for instance, on bilevel learning [96, 25, 52, 84, 85] or on statistical whiteness principles [102, 100] can be used as well.

1.4. Incorporating Space Variance. While the aforementioned modeling appears very restrictive from a *global* viewpoint, it is a natural question to wonder whether by considering *local* image information, that is, looking at image patches of suitably small size, image information can be "glued" together so as to define a better, more suitable image regularizer.

We motivate this idea by showing in Figure 3(a) an enlightening example concerned with local description of directional features for the popular test image image `barbara`. We start by selecting three subregions of interest: two of them are characterized by geometric textures—see Figures 3(d) and 3(g)—while the other presents smooth and homogeneous details—see Figure 3(j). The global image histogram and plot in Figures 3(b) and 3(c) clearly show that image gradients are not oriented along a single dominant direction; rather, they appear to be nonuniformly spread over a box-shaped neighborhood of the origin. Different scenarios arise when considering the two textured regions, as the gradients therein show a clear directional bias, as displayed in the histograms in Figures 3(e) and 3(h) and in the scatter plots in Figures 3(f) and 3(i). Finally, for the gradients computed within the smooth region, a different configuration occurs, where they are not aligned along any preferred direction—see Figure 3(k)—but rather concentrated in a neighborhood of the origin, due to the large homogeneous image content. As observable from the scatter plot in Figure 3(l), dispersion around the origin is, however, smaller with respect to the global scatter plot. Besides the dispersion and the directionality of the gradients, also the shape of the local histograms can significantly vary when considering different image patches (homogeneous vs. textured), thus revealing the limitations of the space-variant TV_p regularizer (TV_p).

In the following sections, we will provide examples that will motivate and justify the idea of formulating a new space-variant framework allowing for the description of image contents at a *local* scale.

However, to give an idea of the results achievable when adopting a more flexible setup where the local amount and sharpness of regularization are adjusted at each pixel depending on the local image content, we employ one of the space-variant regularization terms detailed in what follows, namely the $\mathrm{WTV}^{sv}_{\boldsymbol{p}}$ regularizer, for the restoration of the synthetic test image shown in Figure 1(a). The restored image computed by means of the $\mathrm{WTV}^{sv}_{\boldsymbol{p}}$ is shown in Figure 1(d), together with the corresponding absolute error image in Figure 1(f). From a visual viewpoint, one can clearly



(a) (b) (c)

(d) (e) (f)

(g) (h) (i)

(j) (k) (l)

Fig. 3 *Test image* barbara *and selected subregions (left column), 2D histograms of image gradient components (middle column), scatter plots of image gradient components (right column).*

observe that the $\mathrm{WTV}_{\boldsymbol{p}}^{sv}$ can successfully deal with smooth features while returning sharp blocky objects, thanks to its capability of adapting the regularization sharpness to local, possibly heterogeneous, contents. The one-dimensional horizontal profiles plotted in Figure 1(g) reveal that, under the adoption of the $\mathrm{WTV}_{\boldsymbol{p}}^{sv}$, the staircasing

effect is circumvented, and the image contrast is perfectly recovered; see the close-ups in Figures 1(h) and 1(i), respectively. Note that the edges in the right half of the image appear to be sharper than the ones recovered by the TV regularizer, thanks to the further local adjustment of regularization strength to the underlying geometry. The diagonal profiles plotted in Figure 1(j), and the corresponding close-up in Figure 1(k), reflect this tendency, as the rounding of the corner observed when employing the TV regularizer is mitigated.

The idea of incorporating space-variant information is not new in the context of mathematical methods for image reconstruction. Early approaches can already be traced back looking at the contributions in the field of diffusion-type PDEs for imaging [129, 150, 132, 138, 151] and statistical approaches [60, 131, 36, 33, 32, 128, 158]. In the last couple of years and under a different perspective, a few contributions have also been made in the context of (deep) learning approaches for imaging [75, 112, 113, 115, 97]. In the following, we summarize the ideas contained in these works by structuring our description by topical subsections, to favor readability. As the reader will notice, however, a detailed description of the (large) class of space-variant variational image regularization methods extending those described in section 1.3 will be postponed to the later sections of this work, as those models will appear naturally as soon as our combined statistical/analytical modeling is introduced.

1.4.1. PDE Approaches. In the context of PDE approaches for imaging, the idea of space variance has enriched standard linear and nonlinear diffusion-type models starting from the work of Perona and Malik in the 1990s [129]. There, the authors proposed a nonlinear space-adaptive diffusion method for avoiding the blurring and localization problems of linear diffusion filtering. To do so, an inhomogeneous and space-variant process reducing diffusivity at locations with high probability of being edges was considered. As a natural choice, for any point $\boldsymbol{x} \in \boldsymbol{O} \subset \mathbb{R}^2$, where by $\boldsymbol{O}$ we denote a regular image domain, an edge-stopping diffusion function depending locally on the quantity $|\boldsymbol{\nabla} u(\boldsymbol{x})|$ was used as a likelihood measure. Note that the Perona–Malik filter is a particular instance of the nonlinear diffusion model

$$\begin{cases} u_t = \operatorname{div}\big(W(|\boldsymbol{\nabla} u|)\boldsymbol{\nabla} u\big) & \text{on } \boldsymbol{O} \times (0,\infty], \\ u(\boldsymbol{x},0) = b(\boldsymbol{x}) & \text{on } \boldsymbol{O}, \\ \langle \boldsymbol{\nabla} u, \boldsymbol{n} \rangle = 0 & \text{on } \partial\boldsymbol{O} \times (0,T], \end{cases} \tag{1.3}$$

where $\boldsymbol{n}$ stands for the outward normal vector on $\partial\boldsymbol{O}$ and b stands for the observed image. Problem (1.3) is a standard reference model for anisotropic image restoration PDE approaches. In the case $W \equiv 1$, it acts as a convolution model of the given function $b(\cdot)$ with a Gaussian kernel with standard deviation parameter $\sigma = \sqrt{2t}$. Such an operation corresponds to the well-known low-pass spectral filtering, and it is commonly used for smoothing pictures by averaging values within a certain neighborhood. In the general case, model (1.3) produces a family of images parametrized by $t > 0$, each resulting in a combination between the original image and a filter that depends on the local content of the given image $b(\cdot)$. For $\boldsymbol{x} \in \boldsymbol{O}$, the function W in the Perona–Malik model is set as [129]

$$W(|\nabla u(\boldsymbol{x})|) = \frac{1}{1 + \left(\dfrac{|\boldsymbol{\nabla} u(\boldsymbol{x})|}{K}\right)^2}, \tag{1.4}$$

where the parameter $K > 0$ controls the sensitivity to edges. On uniform regions, where the magnitude of the gradients is weaker, the diffusion coefficient W is close

to 1, so that (1.3) turns into a heat equation which smooths out the noise. Close to edges and boundaries, the gradient magnitudes get larger instead, and thereby the diffusion coefficient in (1.4) vanishes; as a result, in correspondence with these pixels diffusion is not performed and meaningful structures are preserved.

TV regularization can also be framed within the PDE framework above; in fact, by selecting function W as follows,

$$W(|\nabla u(\boldsymbol{x})|) \;=\; \frac{1}{|\boldsymbol{\nabla} u(\boldsymbol{x})|}\,,$$

model (1.3) can be interpreted as the TV *gradient flow*, which consists of the subgradient descent flow of the isotropic TV defined in (TV) [2, 61].

However, despite their edge-adaptive behavior, scalar diffusivity functions $W(\cdot)$ are intrinsically incapable of adjusting the diffusion along the orientation of salient image structures. To do so, a diffusion tensor leading to anisotropic diffusion filters has to be introduced. The most popular choice consists in replacing the scalar W by the structure tensor $\mathbf{J}(\boldsymbol{x}) = \boldsymbol{\nabla} u(\boldsymbol{x}) \otimes \boldsymbol{\nabla} u(\boldsymbol{x}) \in \mathbb{R}^{2\times 2}$. This matrix can be written in terms of its eigenvalues and eigenvectors, the latter encoding the dominant local orientation, the former representing the strength of the diffusion along the preferred direction and its orthogonal. As a result, the action of $\mathbf{J}(\boldsymbol{x})$ can be somehow synthesized by the elongation of the associated elliptical level curves; see, e.g., [150, 132, 138, 151, 76, 69].

1.4.2. Statistical Approaches. Statistical approaches for image processing have become very popular in the last decades due to their ability to incorporate nondeterministic information in the forward model; see [35]. Here, the core idea is to model the unknown image $\boldsymbol{u}$ as a random variable $\boldsymbol{U}$ to highlight the intrinsic uncertainty about its value, which is also related to possible approximations of the model operator $\mathbf{A}$, and of the noise degradation model $\mathcal{N}$. The information or the beliefs available a priori on the random variable $\boldsymbol{U}$ are encoded in the *prior* probability density function (pdf) $\pi(\boldsymbol{u})$. Analogously, the observed image $\boldsymbol{b}$ is regarded as a realization of a random variable $\boldsymbol{B}$, whose behavior for a fixed $\boldsymbol{u}$ is encoded in the *likelihood* pdf $\pi(\boldsymbol{b} \mid \boldsymbol{u})$. In this framework, the goal is to recover the distribution of $\boldsymbol{U}$ given the observation $\boldsymbol{b}$ and the forward linear operator $\mathbf{A}$; in terms of distributions, this translates into seeking the *posterior* pdf $\pi(\boldsymbol{u} \mid \boldsymbol{b})$, which is related to prior and likelihood pdfs via the Bayes' formula:

$$\pi(\boldsymbol{u} \mid \boldsymbol{b}) = \frac{\pi(\boldsymbol{b} \mid \boldsymbol{u})\,\pi(\boldsymbol{u})}{\pi(\boldsymbol{b})}\,, \tag{1.5}$$

where $\pi(\boldsymbol{b})$ is often referred to as the evidence term and plays the role of a normalization constant.

The prior pdf $\pi(\boldsymbol{u})$ can model different characteristics of the image, ranging from the presence of smooth, piecewise-constant, piecewise-linear, or textured regions to boundary configurations. Popular priors for image restoration problems encode information on the distribution of the gray levels within an image and the transition of grayscale intensities between different areas of the image [108].

In [72], the authors interpreted the pixel-gray levels as states of atoms in a lattice-like physical system, so that the unknown image is modeled as a Markov random field (MRF). This translates into the requirement that a selected feature at the generic pixel i of $\boldsymbol{u}$ only depends on the behavior of $\boldsymbol{u}$ at pixels belonging to a set $\mathcal{C}_i$ of neighbors of i, called a *clique*. When the selected feature is the gray level, the 2D

Markovian property at pixel i reads

$$\pi\big(U_i = u_i \mid U_j = u_j\,,\ j \neq i\big) \;=\; \pi\big(U_i = u_i \mid U_j = u_j\,,\ j \in \mathcal{C}_i\big), \tag{1.6}$$

where by $\pi(\cdot)$ we denote the pdf. The prior distribution for an MRF is the so-called *Gibbs prior*,

$$\pi(\boldsymbol{u}) = \frac{1}{Z}\exp\bigg(-\sum_{i=1}^{n} V_{\mathcal{C}_i}(\boldsymbol{u})\bigg), \tag{1.7}$$

where $Z > 0$ is a normalization constant and $V_{\mathcal{C}_i}$ is referred to as the *Gibbs potential function* defined on a clique of pixels centred at pixel i; see also [16] for a more extensive discussion. The potential functions typically depend on a number of parameters that can be considered fixed in the case where our prior beliefs are informative enough to allow for their manual setting, but which, in general, may vary from pixel to pixel.

As will be explored in detail in section 3.2, there exists a strict relation between Gibbs priors and many notable regularizers, such as Tikhonov and TV, which are designed based on the properties of the discrete image gradients $\mathbf{D}\boldsymbol{u}$. Note that nonstationary (i.e., space-variant) MRF approaches have also been considered over the years (see, e.g., [60, 131, 108] for some applications) in a purely statistical framework, for which the design of efficient sampling strategies and maximum likelihood (ML) approaches is required.

By (1.5), the sought unknown image $\boldsymbol{u}$ can be recovered as a single-point representative of $\pi(\boldsymbol{u} \mid \boldsymbol{b})$ moving from a purely statistical to a more optimization-based framework. A very popular strategy for that goes under the name of maximum a posteriori (MAP) estimation [90]: it consists in summarizing the posterior pdf with its mode, i.e.,

$$\begin{aligned} \boldsymbol{u}_{\mathrm{MAP}} \;&\in\; \underset{\boldsymbol{u}\in\mathbb{R}^N}{\arg\max}\ \{\,\pi(\boldsymbol{b} \mid \boldsymbol{u})\,\pi(\boldsymbol{u})\,\} \\ &=\; \underset{\boldsymbol{u}\in\mathbb{R}^N}{\arg\min}\ \{\,-\ln\pi(\boldsymbol{b} \mid \boldsymbol{u}) - \ln\pi(\boldsymbol{u})\,\}\,, \end{aligned} \tag{1.8}$$

where the evidence term $\pi(\boldsymbol{b})$ has been neglected since it does not depend on $\boldsymbol{u}$. We remark that connections between the statistical interpretation of the inverse problem (1.1) via MAP formulation (1.8) and its analytical counterpart (1.2) are nowadays drawn for a large variety of *space-invariant* regularization models, while they are generally hard to understand in the context of *space-variant* models. We will try to fill this gap in the following sections, but for the moment we only warn the reader that moving from a global or space-invariant to a more informative local or space-variant framework requires significant modeling, and computational differences have to be taken into account as the explicit dependence on the distribution hyperparameters in (1.7) cannot be neglected any longer.

In space-variant settings, a very natural alternative to a completely supervised strategy, i.e., the parameters of the nonstationary MRFs are specified a priori, is to model the parameters themselves as random variables following a hierarchical Bayesian approach. In this sense, the MAP paradigm represents a very versatile tool in terms of imaging applications and algorithmic optimization. Among the many contributions in this field, we mention [36, 33, 32], where the authors propose an iterative alternating scheme for the solution of the hierarchical MAP formulation for sparse recovery problems. A parameter marginalization, followed by a small variance analysis, is employed instead in [128] for image inpainting applications.

We further mention that the classical literature on *compressed sensing* algorithms has been revisited and interpreted in probabilistic terms; in this perspective, the popular ℓ_1-regularization terms can be thought of as deriving from a support-informed or spatially adaptive prior, where the local weights are typically estimated starting from the observable data following an empirical Bayesian approach; see [158] and references therein.

1.4.3. Generative Learning Models. Many shortcomings are classically associated to the use of model-driven approaches. Among them, the dependence on the (supposedly known) forward model operator $\mathbf{A}$, the high complexity encountered when solving in practice large-dimensional PDE systems or computing high-dimensional integrals and, mostly, the conceptual and intrinsic ideas of representing the unknown solution in terms of a priori fixed models and distributions have been shown to represent major limitations which have been overcome in recent years by replacing knowledge-based by data-driven designs. It is outside the scope of this review to provide an extensive state-of-the-art description of these from-shallow-to-deep-learning approaches as well as their connection with the world of inverse problems. For that, we refer the reader to the recent review paper [5] and [135] where general data-driven models and generative ones are reviewed in great detail, respectively.

However, for the following description, we will limit ourselves to considering an extremely popular class of data-driven approaches which, in some sense, are in close connection with the statistically inspired space-variant modeling discussed in this work. The former class, introduced in [92] under the name of variational autoencoders (VAEs), exploit training examples to estimate, rather than the desired solution itself, the statistical distribution it is sampled from (ideally (1.7)) and use it to generate new data with analogous characteristics to the elements in the training set. VAEs perform such a process by means of the interplay between two subsequent neural networks: the first one (the *encoder*) is used to compress the data into a latent probability space which typically has a smaller dimension than the original one, while the second *decoder* network maps back the features sampled from the latent space so as to generate data as close as possible to the given data, with respect to some given distance function. First introduced in [75], generative adversarial networks (GANs) are a different class of generative models popularly used nowadays for learning the unknown distribution of given training data. Analogously to VAEs, two networks are used here as well to detect distinctive features in the data. However, differently from VAEs, the two networks are used here in an adversarial fashion with one network used to generate data and the other used to discriminate between the "true" data distribution and its "opponent" (adversarial) attacks. Convergence of the training process is achieved when the discriminator network becomes unable to make such a distinction, that is, when the learning procedure has been so effective that generated data cannot be discriminated from true data. GANs are nowadays used in many different contexts ranging from supervised to unsupervised learning. Differently from the fully model-driven statistical approaches summarized in section 1.4.2, the solution computed by GANs potentially offers a more precise way to describe local image features in natural images. However, given their fully data-driven modeling, the interpretation of GANs within both a statistical and analytical framework remains somehow unclear. The outstanding performance of GANs in the field of imaging, however, has favored their use in a large variety of imaging problems (see, e.g., [112, 113, 115, 97]) and is still a growing research area in the field which could inspire and complement the classical analytical/statistical knowledge-based modeling in future

work (see section 11 for research outlooks). In analogy to what we are going to describe in the following, generative approaches share the common feature of having as their objective the estimation of an unknown underlying probability distribution of given (imaging) data. Differently from the more model-driven approaches, however, they do not stick with any explicit (albeit general as highly parametric) expression of such distribution. We could argue that while generative approaches make the learning procedure flexible, thanks to their adaptivity to the possibly nonstandard data distribution, the statistically inspired models we are going to consider in the following take advantage of the hyperparametrization imposed a priori.

1.4.4. Exploiting Space Variance in Applications. Space-variant models have been shown to be effective not only on synthetic and/or "didactic" examples, but also on a wide class of real-world applications, such as, for instance, medical imaging.

Regarding CT applications, for instance, several image reconstruction approaches relying intrinsically on a space-variant estimation of model hyperparameters have been considered. In [146], for instance, directional TV regularization on the sinogram data is proposed to inpaint the missing range of angles and improve the inversion process, while in [62] an automatic selection strategy for a weighted reconstruction model is considered and validations on both synthetic and real data are reported.

As far as (multicontrast) magnetic resonance imaging (MRI) applications are concerned, we refer the reader to [4, 65], where local regularization weighting as well as directional (therein often called *structural*) strategies extending those used in [91] for electrical impedance tomography (EIT) are used. Analogous approaches have further been considered in [67, 66] for improving the quality of positron emission tomography (PET) imaging data by an appropriate fusion driven by structural MRI data. Similar approaches have further been considered in [22] for blind hyperspectral imaging and in [9] for magnetic particle imaging.

Within the class of medical imaging applications, we further mention photoacoustic tomography (PAT) imaging, for which in [17] a space-variant modeling well-adapted to the composite and heterogeneous nature of the target is proposed.

Among the many other real-world applications which significantly benefit from the use of space-variant approaches in terms of, in particular, local directional dependence, we mention here the work carried out in [73, 111, 26, 125], where noninvasive digital reconstruction models based on anisotropic diffusion and transport PDEs have been effectively used in the context of digital reconstruction of ancient frescoes, illuminated manuscripts, surface colorization, and inpainting to unveil missing or occluded contents via the use of inpainting, image fusion, and/or image enhancement techniques.

1.5. Motivation and Contribution of This Work. From the aforementioned sections we have seen that, though following different paths, different communities focused on the mathematical modeling of local image features. Depending on the scientific community considered and when looking at all these works, though, it is not very clear how to connect and compare these different findings, since very similar properties interpreted in different fields may be called by very different names (e.g., *nonstationary* models in a Bayesian framework, *structural* or *adaptive* approaches in an analytical context). The objective of this work is to provide a unified view of many of these different models in terms of a new, generalized Bayesian modeling which allows also for some original extensions which have not been explicitly studied before. The Bayesian framework described in this work indeed paves the way for the design of new, unexplored strategies helpful to design flexible and adaptive image regular-

ization functionals whose hyperparameters can be estimated by taking advantage of the form of the underlying gradient distributions through statistical approaches. In order to present the framework in its full generality and in view of its application to a larger class of image reconstruction models, we will not neglect to provide details on the use of a (generalized) discrepancy principle strategy needed to compute the hyperparameters associated to the likelihood functionals. We further stress that TV regularization (TV) is taken as a reference regularization model in this work due to the incredible amount of contributions developed over the last 30 years, as we have discussed and will discuss thoroughly in the following. However, the reader should be reassured that analogous considerations could (and should!) still be exploited for different types of regularization functionals, as we will shortly comment on in section 11 of this work.

1.6. Structure of the Paper. The paper is organized as follows. In section 2, we set the notations and recall the main notions and definitions which will be useful in the rest of the article. Then, in section 3, we set the Bayesian probabilistic scene by introducing the main actors, namely the space-variant (nonstationary) priors together with the likelihood pdf corresponding to the class of noise models considered in this review. The properties of the regularizers induced by the space-variant priors will be analyzed from a modeling and optimization viewpoint in section 4, while in section 5 we provide some useful insights on their geometric interpretation. Next, in section 6, we formulate the final joint image and hyperparameter estimation models, where the different space-variant regularizers proposed are combined with general data fidelity terms and the suitable prior distributions on the model hyperparameters (i.e., *hyper-priors*). In section 7, we address the hyperparameter estimation problem by designing robust ML-type strategies that will be tested on synthetic and natural examples. Then in section 8 the numerical solution of the general variational model in the form (1.2) upon the selected choices of regularizers and fidelity terms is addressed by means of an alternating direction method of multipliers (ADMM). In section 9, the effectiveness of the space-variant approach is finally assessed by applying the designed framework to the restoration of different synthetic and natural images. A short user's guide to space-variant regularized models is given in section 10. To summarize, in section 11 we discuss some open questions and challenges representing natural extensions of this work. Finally, we report in section 12 some final considerations and remarks.

2. Notation and Preliminaries. We will use the notation $\mathbb{R}_+$ and $\mathbb{R}_{++} = \mathbb{R}_+ \setminus \{0\}$ for the set of nonnegative and positive real numbers, respectively, and denote by $\mathbf{0}_d$, $\mathbf{1}_d$, $\mathbf{I}_d$ the d-dimensional vectors of all zeros and ones and the identity matrix of size $d \times d$, respectively. In the case of a matrix $\mathbf{M} \in \mathbb{R}^{d\times d}, d > 1$, we will denote by $|\mathbf{M}|$ the determinant of $\mathbf{M}$.

To indicate multivariate random variables and their realizations we will use bold capital/lowercase letters, e.g., $\boldsymbol{X}$ and $\boldsymbol{x}$, and we denote by $\pi_{\boldsymbol{X}}$, $\boldsymbol{\eta}_{\boldsymbol{X}} = \mathbb{E}(\boldsymbol{X})$, $\boldsymbol{\Sigma}_{\boldsymbol{X}}$ the pdf, mean (or expected value), and covariance matrix of the random variable $\boldsymbol{X}$, respectively. We will omit the subscript $\boldsymbol{X}$ if not necessary. The characteristic and the indicator function of a set $\boldsymbol{S}$ are defined as

$$\chi_{\boldsymbol{S}}(\boldsymbol{x}) := \begin{cases} 1 & \text{if } \boldsymbol{x} \in \boldsymbol{S}, \\ 0 & \text{otherwise}, \end{cases} \qquad \iota_{\boldsymbol{S}}(\boldsymbol{x}) := -\ln \chi_{\boldsymbol{S}}(\boldsymbol{x}) = \begin{cases} 0 & \text{if } \boldsymbol{x} \in \boldsymbol{S}, \\ +\infty & \text{otherwise}, \end{cases}$$

respectively. Moreover, we denote by Γ the Gamma function, which is defined as follows.

DEFINITION 2.1 (Gamma and incomplete Gamma functions). *The lower and upper incomplete Gamma functions, $\underline{\Gamma}$ and $\overline{\Gamma}$, respectively, are defined by*

$$\underline{\Gamma}(x,y) = \int_0^y t^{x-1}e^{-t}dt\,, \quad \overline{\Gamma}(x,y) = \int_y^{+\infty} t^{x-1}e^{-t}dt\,, \quad (x,y) \in \mathbb{R}_{++} \times \mathbb{R}_+\,. \tag{2.1}$$

The (complete) Gamma function Γ is

$$\Gamma(x) = \lim_{y\to+\infty} \underline{\Gamma}(x,y) = \overline{\Gamma}(x,0) = \int_0^{+\infty} t^{x-1}e^{-t}dt\,, \quad x \in \mathbb{R}_{++}\,. \tag{2.2}$$

Now, we recall the definitions of a few well-known distributions to which we often refer throughout the discussion and that will be mainly employed in the modeling of the regularization terms reviewed here.

DEFINITION 2.2 (univariate Laplace distribution). *A scalar random variable X is Laplacian-distributed with mean $\eta \in \mathbb{R}$ and scale parameter $\gamma \in \mathbb{R}_{++}$, denoted by $X \sim \mathrm{L}(\eta,\gamma)$, if its pdf has the form*

$$\pi(x|\eta,\gamma) = \frac{\gamma}{2}\exp\left(-\gamma\,|x-\eta|\,\right)\,, \quad x \in \mathbb{R}\,. \tag{Ld}$$

DEFINITION 2.3 (univariate generalized Gaussian distribution). *A scalar random variable X is generalized Gaussian-distributed with mean $\eta \in \mathbb{R}$, scale parameter $\gamma \in \mathbb{R}_{++}$, and shape parameter $s \in \mathbb{R}_{++}$, denoted by $X \sim \mathrm{GG}(\eta,\gamma,s)$, if its pdf has the form*

$$\pi\left(x|\eta,\gamma,s\right) = \frac{\gamma}{2}\frac{s}{\Gamma(1/s)}\,\exp\left(-\gamma^s\,|x-\eta|^s\right), \quad x \in \mathbb{R}, \tag{GGd}$$

with $\Gamma(\cdot)$ denoting the Gamma function defined in (2.2)*. In particular, for any fixed $\eta \in \mathbb{R}$, $\gamma \in \mathbb{R}_{++}$, the pdf in* (GGd) *converges pointwise to a uniform distribution as $s \to +\infty$, namely,*

$$\lim_{s\to+\infty} \pi(x|\eta,\gamma,s) = \frac{\gamma}{2}\,\chi_{\,[0,1/\gamma]}\left(|x-\eta|\right).$$

Finally, the standard deviation $\sigma \in \mathbb{R}_+$ of the GG pdf in (GGd) *can be written in terms of the scale parameter γ and the shape parameter s as follows:*

$$\sigma = (1/\gamma)\sqrt{\Gamma(3/s)\,/\,\Gamma(1/s)}\,. \tag{2.3}$$

The following definition extends the GG distribution to the bivariate case.

DEFINITION 2.4 (bivariate generalized Gaussian distribution). *A bivariate random variable $\boldsymbol{X}$ is generalized Gaussian-distributed with mean $\boldsymbol{\eta} \in \mathbb{R}^2$, symmetric positive definite covariance matrix $\boldsymbol{\Sigma} \in \mathbb{R}^{2\times2}$, and shape parameter $s \in \mathbb{R}_{++}$, denoted by $\boldsymbol{X} \sim \mathrm{BGG}(\boldsymbol{\eta},\boldsymbol{\Sigma},s)$, if its pdf has the form*

$$\pi(\boldsymbol{x}|\boldsymbol{\eta},\boldsymbol{\Sigma},s) = \frac{1}{2\pi|\boldsymbol{\Sigma}|^{1/2}}\,\frac{s}{\Gamma(2/s)2^{2/s}}\,\exp\left(-\frac{1}{2}\left((\boldsymbol{x}-\boldsymbol{\eta})^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\eta})\right)^{\frac{s}{2}}\right). \tag{BGGd}$$

We now provide definitions for the Laplace distribution (Ld) and the generalized Gaussian one (GGd) when the scalar random variable X is known to be nonnegative.

DEFINITION 2.5 (univariate exponential or half Laplacian distribution). *A scalar random variable X is exponential- or half Laplacian-distributed with scale parameter $\gamma \in \mathbb{R}_{++}$, denoted by $X \sim \mathrm{hL}(\gamma)$, if $X = |Y|$, with $Y \sim \mathrm{L}(0, \gamma)$. The pdf of X takes the form*

$$\pi(x|\gamma) = \begin{cases} \gamma \exp(-\gamma x) & \text{if } x \in \mathbb{R}_+, \\ 0 & \text{otherwise.} \end{cases} \tag{hLd}$$

DEFINITION 2.6 (univariate half generalized Gaussian distribution). *A scalar random variable X is half generalized Gaussian-distributed with scale parameter $\gamma \in \mathbb{R}_{++}$ and shape parameter $s \in \mathbb{R}_{++}$, denoted by $X \sim \mathrm{hGG}(\gamma, s)$, if $X = |Y|$, with $Y \sim \mathrm{GG}(0, \gamma, s)$. The pdf of X takes the form*

$$\pi(x|\gamma, s) = \begin{cases} \gamma \dfrac{s}{\Gamma(1/s)} \exp(-\gamma^s x^s) & \text{if } x \in \mathbb{R}_+, \\ 0 & \text{otherwise.} \end{cases} \tag{hGGd}$$

We also give the definition of Gamma distribution which will be used in the next subsection and in the appendix.

DEFINITION 2.7 (univariate Gamma distribution). *A scalar random variable X is Gamma-distributed with scale parameter $\nu \in \mathbb{R}_{++}$ and shape parameter $z \in \mathbb{R}_{++}$, denoted by $X \sim \mathrm{Gamma}(\nu, z)$, if its pdf has the form*

$$\pi(x|\nu, z) = \begin{cases} \dfrac{1}{\nu^z \Gamma(z)} x^{z-1} \exp\left(-\dfrac{x}{\nu}\right) & \text{if } x \in \mathbb{R}_{++}, \\ 0 & \text{otherwise.} \end{cases} \tag{2.4}$$

The mean $\mu \in \mathbb{R}_{++}$ and standard deviation $\sigma \in \mathbb{R}_{++}$ of the Gamma pdf in (2.4) *can be written in terms of the scale parameter ν and the shape parameter z as follows:*

$$\mu = \mu z, \qquad \sigma = \mu \sqrt{z}. \tag{2.5}$$

Finally, we recall the definition of proximal operator to which we will extensively refer in sections 4–8.

DEFINITION 2.8 (proximal operator). *Let $f : \mathbb{R}^N \to \mathbb{R}$ be a proper, lower semi-continuous and possibly nonconvex function and let $\beta \in \mathbb{R}_{++}$. The proximal operator of f with proximity parameter β is the set-valued function $\mathrm{prox}_f^\beta : \mathbb{R}^N \rightrightarrows \mathbb{R}^N$ defined for any $\boldsymbol{w} \in \mathbb{R}^N$ by*

$$\mathrm{prox}_f^\beta(\boldsymbol{w}) := \underset{\boldsymbol{x} \in \mathbb{R}^N}{\arg\min} \left\{ f(\boldsymbol{x}) + \frac{\beta}{2} \|\boldsymbol{x} - \boldsymbol{w}\|_2^2 \right\} \subset \mathbb{R}^N. \tag{2.6}$$

Note that if f in the definition above is convex, then the cost function in (2.6) is strongly convex, hence it admits a unique global minimizer. In this case, $\mathrm{prox}_f^\beta(\cdot)$ is a well-defined function from $\mathbb{R}^N$ to itself which coincides with the well-studied proximal operator frequently encountered in convex optimization contexts (see, e.g., [54]).

2.1. Generalized Discrepancy Principle. Assuming that the noise degradation operator $\mathcal{N}$ in (1.1) models the action of an additive, zero-mean, independent and identically distributed (i.i.d.) generalized Gaussian (in short, AIGG) noise, we have that (1.1) can be rewritten as

$$\boldsymbol{b} = \mathbf{A}\boldsymbol{u} + \boldsymbol{e}, \text{ with } e_j \text{ realization of } E_j \sim \mathrm{GG}(0, \omega, q), \ j = 1, \dots, M, \tag{2.7}$$

where, based on Definition 2.3, $\boldsymbol{e} \in \mathbb{R}^M$ is the vector of realizations of the M-variate random variable $\boldsymbol{E}$ whose components E_j are i.i.d. zero-mean GG random variables with scale parameter $\omega \in \mathbb{R}_{++}$ and shape parameter $q \in \mathbb{R}_{++}$, both encoding information on the noise standard deviation according to (2.3).

Before detailing a generalized discrepancy principle (in short, GDP) useful to define an automatic selection strategy for the regularization parameter μ in (1.2) under the modeling assumption (2.7), we report the following result, whose proof is based on classical probability arguments and given for completeness in the appendix.

PROPOSITION 2.9. *If* $X_i \sim \mathrm{GG}(0, \omega, q)$, $i = 1, \dots, M$, *with* $\omega, q \in \mathbb{R}_{++}$, *are independent random variables, then we have*

$$Y \;=\; \|(X_1, \dots, X_M)\|_q^q \;=\; \sum_{i=1}^{M} |X_i|^q \;\sim\; \mathrm{Gamma}(\nu, z), \quad \nu = \frac{1}{\omega^q}\,,\; z = \frac{M}{q}\,. \tag{2.8}$$

It thus follows from (2.5) *that the random variable* Y *has mean* $\eta_Y \in \mathbb{R}_{++}$ *and standard deviation* $\sigma_Y \in \mathbb{R}_{++}$, *whose expressions are given by*

$$\eta_Y \;=\; \nu\, z \;=\; \frac{1}{\omega^q}\,\frac{M}{q}\,, \qquad \sigma_Y \;=\; \nu\,\sqrt{z} \;=\; \frac{1}{\omega^q}\,\sqrt{\frac{M}{q}}\,. \tag{2.9}$$

Thanks to Proposition 2.9, we can now provide the following GDP.

DEFINITION 2.10 (generalized discrepancy principle). *Let* $\boldsymbol{u}^*(\mu)$ *indicate the parameter-dependent solution of variational problem* (1.2) *under modeling assumptions* (2.7) *and let* $\boldsymbol{y}^*(\mu) := \mathbf{A}\boldsymbol{u}^*(\mu) - \boldsymbol{b}$ *be the associated residual image. Then the generalized discrepancy principle applied to selecting* μ *can be formulated as follows:*

$$\textit{Select}\;\; \mu \;=\; \mu^* \;\; \textit{such that} \;\; \|\boldsymbol{y}^*(\mu^*)\|_q \;=\; \delta_q\,,$$

where $\delta_q \in \mathbb{R}_{++}$, *referred to as the discrepancy radius, is given by*

$$\delta_q \;:=\; \tau\,\sqrt[q]{\mathbb{E}(\|\boldsymbol{E}\|_q^q)} \;=\; \begin{cases} \tau\,(1/\omega)\,(M/q)^{1/q} & \textit{if} \quad q < +\infty\,, \\ \tau\,(1/\omega) & \textit{if} \quad q = +\infty\,, \end{cases} \tag{2.10}$$

with discrepancy coefficient $\tau \approx 1$.

Remark 2.11 (Morozov discrepancy principle from GDP with $q = 2$). We note that for $q = 2$, corresponding to the case of additive, zero-mean i.i.d. Gaussian noise, the random variables X_i in Proposition 2.9 are all generalized Gaussian-distributed with zero-mean, shape parameter $q = 2$, and scale parameter ω or, equivalently, Gaussian-distributed with zero-mean and standard deviation $\sigma = 1/(\sqrt{2}\,\omega)$, according to (2.3). Hence, the random variable Y in (2.8) is Gamma-distributed with shape parameter $z = M/q = M/2$ and scale parameter $\nu = 1/\omega^q = 2\,\sigma^2$. It follows from (2.9) that the mean (or expected value) of random variable Y is given by $\mu_Y = \mathbb{E}(\|\boldsymbol{E}\|_2^2) = M\,\sigma^2$ and, hence, the GDP in Definition 2.10 reduces to the classical Morozov discrepancy principle,

$$\text{select}\;\; \mu \;=\; \mu^* \;\text{such that}\;\; \|\boldsymbol{y}^*(\mu^*)\|_2 \;=\; \delta_2\,, \quad \text{with}\;\; \delta_2 \;=\; \tau\,\sqrt{M}\,\sigma\,,$$

detailed, e.g., in [117, 68].

3. A Flexible Bayesian-Inspired Framework. In this section, we recall the general Bayesian framework outlined in section 1.4.2 and adapt it to our purposes and considerations. As a general comment, we believe it is worth remarking here that the derivations carried out in section 3 are only meant to be *Bayesian-inspired*. In other words, the probabilistic setup is employed for providing a rational basis for the introduction of the recalled space-variant variational models; at the same time, it gives significant insights about the expected behavior of the parameters arising in the variational models and allows us to derive, in a very natural way, an algorithmic strategy for their automatic estimation. The choice of not rigorously following the Bayesian approach is motivated by the fact that we consider the variational framework as the setup of primary interest for our analysis. In this perspective, the MAP strategy represents the key which allows us to bridge the two approaches. The employment of different probabilistic-based methods for exploring the posterior pdf is thus beyond the goal of this work.

In what follows, we start specifying the different noise degradation models $\mathcal{N}$ considered in (1.1) and define suitable likelihood pdfs accordingly. Next, we specify the flexible space-variant priors focus of this work by defining a class of increasingly general distributions. Likelihoods and priors are then combined by means of a suitable MAP estimate suited to describe the case where the prior hyperparameters are unknown.

3.1. Likelihoods. In the following, the likelihood pdf will be indicated by $\pi(\boldsymbol{b} \mid \boldsymbol{u}, \boldsymbol{\Phi})$, where, in addition to $\boldsymbol{u} \in \mathbb{R}^N$, the dependence on the likelihood hyperparameter vector $\boldsymbol{\Phi} \in \mathbb{R}^k$ involved in the analytic expression of the pdf is here explicitly taken into account.

In order to benefit from the automatic parameter selection strategy provided by the GDP detailed in Definition 2.10, we will focus our attention on the class of AIGG noises corresponding to model (2.7).

Note that although not exhaustive, this class is very general as it contains some commonly used noise models, such as, e.g., the additive i.i.d. Laplacian (AIL) noise ($q = 1$), the additive i.i.d. Gaussian (AIG) noise ($q = 2$) and the additive i.i.d. uniform (AIU) noise ($q = +\infty$). AIG noise is very frequently considered in applications to model the presence of electronic noise. On the other hand, AIL noise is typically encountered in stack-filtering and to model the residuals in object-tracking applications (see, e.g., [1, 156]), although its associated functional form is also typically and frequently used to model nonadditive impulsive noise (see the seminal works [120, 48]). Finally, AIU noise is typically used to describe quantization errors arising in data acquisition and processing; see, e.g., [152, 53]. More general nonadditive noise models can also be employed to express signal-dependent or multiplicative noise degradation. From a statistical point of view, such models can be described by means of Poisson [104, 137, 153, 13] or Gamma [133, 6, 140, 87, 141] distributions, respectively. The design of GDP strategies for these models is intrinsically more challenging due to the dependence of the noise on the (unknown) local image intensity values. Nonetheless, some works in this direction exist; see, e.g., [14, 145].

Due to the assumption that the univariate GG random variables E_j in (2.7) are zero-mean i.i.d., the M-variate likelihood pdf can be written as the product of M univariate GG pdfs (see Definition 2.3) all characterized by the same pair of scale and shape parameters $\boldsymbol{\Phi} := (\omega, q)$ and with means equal to $(\mathbf{A}\boldsymbol{u})_j$, $j = 1, \ldots, M$; in

formula,

$$\pi\left(\boldsymbol{b} \mid \boldsymbol{u}, \boldsymbol{\Phi}\right) \;=\; \prod_{j=1}^{M} \pi(b_j \mid (\mathbf{A}\boldsymbol{u})_j, \omega, q)\,.$$

According to Definition 2.3, when $q < +\infty$ the likelihood pdf thus takes the form

$$\begin{aligned}\pi(\boldsymbol{b} \mid \boldsymbol{u}, \boldsymbol{\Phi}) \;&=\; \prod_{j=1}^{M} \left(\frac{\omega}{2}\frac{q}{\Gamma(1/q)} \exp\left(-\omega^q |(\mathbf{A}\boldsymbol{u})_j - b_j|^q\right)\right) \\ &=\; \left(\frac{\omega}{2}\frac{q}{\Gamma(1/q)}\right)^M \exp\left(-\omega^q \left\|\mathbf{A}\boldsymbol{u} - \boldsymbol{b}\right\|_q^q\right), \text{ with } \; \boldsymbol{\Phi}{=}(\omega, q) \in \mathbb{R}^2_{++}\,, \end{aligned} \tag{3.1}$$

while for $q = +\infty$ it reads

$$\begin{aligned}\pi(\boldsymbol{b} \mid \boldsymbol{u}, \boldsymbol{\Phi}) \;&=\; \prod_{j=1}^{M} \left(\frac{\omega}{2}\, \chi_{[0,1/\omega]}\left(|(\mathbf{A}\boldsymbol{u})_j - b_j|\right)\right) \\ &=\; \left(\frac{\omega}{2}\right)^M \chi_{[0,1/\omega]}\left(\|\mathbf{A}\boldsymbol{u} - \boldsymbol{b}\|_\infty\right), \text{ with } \; \boldsymbol{\Phi}{=}(\omega, q) \in \mathbb{R}_{++} \times \{+\infty\}\,. \end{aligned} \tag{3.2}$$

In our settings, we will assume that both parameters $\boldsymbol{\Phi} = (\omega, q)$ are known. As a consequence, from now on, the dependence on $\boldsymbol{\Phi}$ in the expression of the likelihood pdf will be omitted.

3.2. Priors. Recalling the statistical modeling introduced in section 1.4.2 and in particular the MRF structure in (1.7), we proceed similarly as in the previous section and make explicit the dependence of the prior pdf on the vector of prior hyperparameters $\boldsymbol{\Theta}$ involved in its expression, which here are assumed to be unknown. The Gibbs prior in (1.7) thus reads

$$\pi(\boldsymbol{u} \mid \boldsymbol{\Theta}) = \frac{1}{Z(\boldsymbol{\Theta})} \exp\left(-\sum_{i=1}^{N} V_{\mathcal{C}_i}(\boldsymbol{u}; \boldsymbol{\Theta})\right),$$

where $Z(\boldsymbol{\Theta}) \in \mathbb{R}_{++}$ is a normalization constant depending on the unknown parameters $\boldsymbol{\Theta}$ while $V_{\mathcal{C}_i}$ is the Gibbs potential on the ith clique $\mathcal{C}_i$, also depending on $\boldsymbol{\Theta}$.

Recalling the general Markovian property (1.6) and thinking of the description of the image in terms of its local gradients $(\mathbf{D}\boldsymbol{u})_i$ discretized by standard first-order forward finite differences, we have that (1.6) turns into

$$\pi(\,U_i = u_i \mid U_j = u_j\,,\; j \neq i\,) = \pi(\,U_i = u_i \mid U_{i,\text{right}} = u_{i,\text{right}}\,,\, U_{i,\text{down}} = u_{i,\text{down}}). \tag{3.3}$$

For better illustration, we show the corresponding configuration of the generic clique in Figure 4. Condition (3.3) states that the potential function $V_{\mathcal{C}_i}$ is defined over a discrete set of cardinality 3, namely $\{u_i, u_{i,\text{right}}, u_{i,\text{down}}\}$, which are indeed the values involved in the computation of the discrete gradient at pixel i.

We now outline a prior-derivation procedure aimed at connecting the Bayesian setup with the classical variational framework. In fact, by considering suitable simplifications, in the end the following approach will produce variational models characterized by two penalization terms, namely one for the image and one for the unknown parameters.

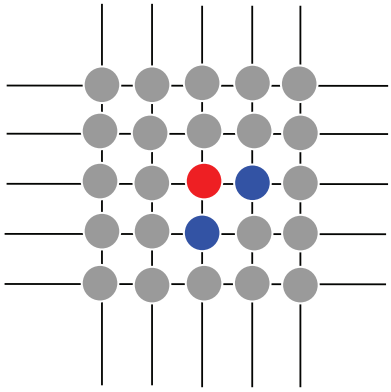


Fig. 4 *Pixels represented as atoms in a lattice. The colored ones belong to the clique related to the red atom. In particular, the blue atoms are involved in the computation of the finite difference gradient evaluated at the red atom.*

Introducing the function $\boldsymbol{z} : \mathbb{R}^N \to \mathbb{R}_+^N$ defined by

$$\boldsymbol{z}(\boldsymbol{u}) = [z_1(\boldsymbol{u}), \dots, z_N(\boldsymbol{u})]^T, \quad z_i(\boldsymbol{u}) = \|(\mathbf{D}\boldsymbol{u})_i\|_2\,, \tag{3.4}$$

and assuming that each image gradient magnitude $x_i := z_i(\boldsymbol{u}) = \|(\mathbf{D}\boldsymbol{u})_i\|_2 \in \mathbb{R}_+$ is the realization of the same univariate half-Laplacian (or exponential) distribution (hLd) with scale parameter $\alpha \in \mathbb{R}_{++}$ and that the magnitudes at different pixels are independent, we have that the N-variate pdf $\pi(\boldsymbol{z}(\boldsymbol{u}) \mid \alpha)$ takes the form

$$\begin{aligned} \pi(\boldsymbol{z}(\boldsymbol{u}) \mid \alpha) &= \prod_{i=1}^{N} \pi(z_i(\boldsymbol{u}) \mid \alpha) = \prod_{i=1}^{N} \left(\alpha \exp\left(-\alpha \|(\mathbf{D}\boldsymbol{u})_i\|_2\right)\right) \\ &= \alpha^N \exp\left(-\alpha \sum_{i=1}^{N} \|(\mathbf{D}\boldsymbol{u})_i\|_2\right). \end{aligned} \tag{3.5}$$

Starting from (3.5), one can obtain the popular Gibbs TV prior by normalization. In fact, by taking

$$\pi(\boldsymbol{u} \mid \boldsymbol{\Theta}) = c(\boldsymbol{\Theta})\, \pi(\boldsymbol{z}(\boldsymbol{u}) \mid \boldsymbol{\Theta})\,, \tag{3.6}$$

with the normalization term $c(\boldsymbol{\Theta})$ reading

$$c(\boldsymbol{\Theta}) = \frac{1}{\displaystyle\int_{\boldsymbol{u} \in \mathbb{R}^N} \pi(z(\boldsymbol{u}) \mid \boldsymbol{\Theta}) d\boldsymbol{u}}\,, \tag{3.7}$$

we have

$$\begin{aligned} \pi(\boldsymbol{u} \mid \boldsymbol{\Theta}) &= c(\boldsymbol{\Theta}) \prod_{i=1}^{N} \left(\alpha \exp\left(-\alpha \|(\mathbf{D}\boldsymbol{u})_i\|_2\right)\right) \\ &= c(\boldsymbol{\Theta})\, \alpha^N \exp\left(-\alpha \sum_{i=1}^{N} \|(\mathbf{D}\boldsymbol{u})_i\|_2\right), \text{ with } \boldsymbol{\Theta} = \alpha \in \mathbb{R}_{++}\,. \end{aligned} \tag{3.8}$$

Notice that the presence of $c(\boldsymbol{\Theta})$ guarantees that the prior pdf in (3.8) sums up to 1 when considering the space of all possible configurations.

A way to improve upon the intrinsic rigidity of (3.8), due to the dependence on the single scale parameter $\alpha \in \mathbb{R}_{++}$, consists in letting it vary at any pixel, so as to maintain the same prior hypothesis on the image gradient magnitudes x_i, while enriching it with further flexibility depending on the local scale α_i. The corresponding space-variant hLd probability density thus reads in this case

$$\text{(hLd-sv)} \qquad \pi(x_i \mid \alpha_i) = \begin{cases} \alpha_i \exp(-\alpha_i x_i), & x_i \geq 0, \\ 0, & x_i < 0, \end{cases} \quad i = 1, \dots, N, \quad \alpha_i \in \mathbb{R}_{++}\,,$$

and, according to the general expression in (3.6) and upon the selection of function $\boldsymbol{z}(\boldsymbol{u})$ as in (3.4), yields the following nonstationary prior pdf on $\boldsymbol{u}$:

$$\begin{aligned} \pi(\boldsymbol{u} \mid \boldsymbol{\Theta}) &= c(\boldsymbol{\Theta}) \prod_{i=1}^{N} \left(\alpha_i \exp\left(-\alpha_i \left\|(\mathbf{D}\boldsymbol{u})_i\right\|_2\right)\right) \\ &= c(\boldsymbol{\Theta}) \left(\prod_{i=1}^{N} \alpha_i\right) \exp\left(-\sum_{i=1}^{N} \alpha_i \left\|(\mathbf{D}\boldsymbol{u})_i\right\|_2\right), \quad \text{with} \quad \boldsymbol{\Theta} = \boldsymbol{\alpha} \in \mathbb{R}_{++}^{N}\,, \end{aligned} \tag{3.9}$$

with z defined as in (3.4) and the normalization function $c(\boldsymbol{\Theta})$ obtained by plugging the novel $\pi(\boldsymbol{z}(\boldsymbol{u}) \mid \boldsymbol{\Theta})$ into (3.7).

As shown in Figure 5(a), this choice allows for more flexibility in the description of local gradient contents; nonetheless, it has the major drawback of still limiting to the family of half-Laplacian distributions the choice of the local pdf considered.

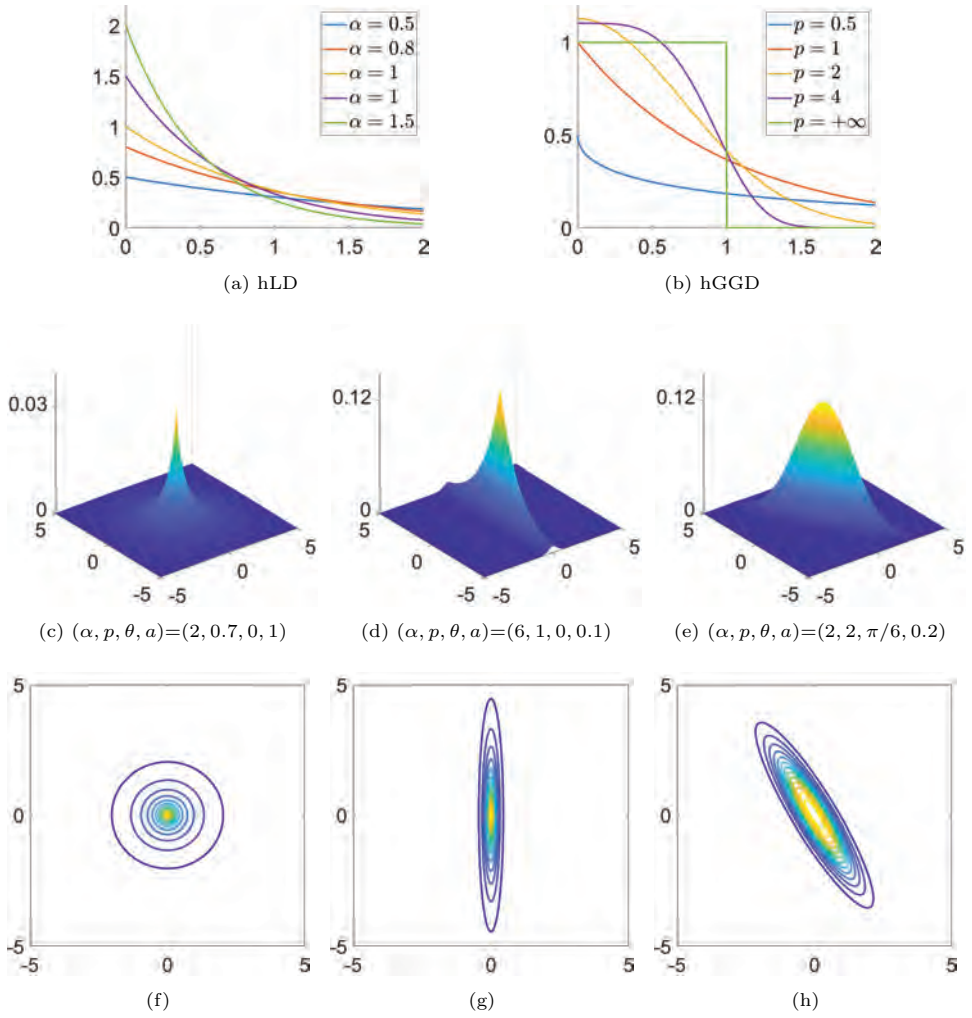
To overcome this, one can leave further freedom to the heavy- vs. light-tailed behavior of the exponential distribution considered. This can be done in practice by allowing, along with a space-variant pdf scale α_i, a different exponential behavior depending on a "sharpness" (shape) parameter p_i, still possibly varying at any $i = 1, \dots, N$. This choice corresponds to considering a space-variant half generalized Gaussian distribution (hGGD-sv) (see (GGd)), whose expression for $i = 1, \dots, N$ reads

$$\text{(hGGd-sv)} \qquad \pi(x_i; \alpha_i, p_i) = \begin{cases} \frac{\alpha_i p_i}{\Gamma(1/p_i)} \exp(-(\alpha_i x_i)^{p_i}), & x_i \geq 0, \\ 0, & x_i < 0, \end{cases} \quad \alpha_i, p_i \in \mathbb{R}_{++}\,.$$

In Figure 5(b) we show the plot of the hGGd pdf for different values of the shape parameter p while leaving the scale parameter $\alpha = 1$ fixed. One can easily notice that the family of hGG distributions is particularly rich, ranging from hyper-Laplacian distributions for $p < 1$ to uniform distributions for $p = +\infty$. According to (3.6), the prior on $\boldsymbol{u}$ corresponding to the pdf in (hGGd-sv) reads

$$\begin{aligned} \pi(\boldsymbol{u} \mid \boldsymbol{\Theta}) &= c(\boldsymbol{\Theta}) \prod_{i=1}^{N} \left(\frac{\alpha_i p_i}{\Gamma(1/p_i)} \exp\left(-\left(\alpha_i \left\|(\mathbf{D}\boldsymbol{u})_i\right\|_2\right)^{p_i}\right)\right) \\ &= c(\boldsymbol{\Theta}) \left(\prod_{i=1}^{N} \frac{\alpha_i p_i}{\Gamma(1/p_i)}\right) \exp\left(-\sum_{i=1}^{N} \alpha_i^{p_i} \left\|(\mathbf{D}\boldsymbol{u})_i\right\|_2^{p_i}\right), \quad \text{with } \boldsymbol{\Theta} = (\boldsymbol{\alpha}, \boldsymbol{p}) \in \mathbb{R}_{++}^{N\times 2}\,, \end{aligned} \tag{3.10}$$

where z is defined in (3.4), while $c(\boldsymbol{\Theta})$ is given by (3.7) when considering the current selection of $\pi(\boldsymbol{z}(\boldsymbol{u}) \mid \boldsymbol{\Theta})$.



(a) hLD (b) hGGD

(c) (α, p, θ, a)=$(2, 0.7, 0, 1)$ (d) (α, p, θ, a)=$(6, 1, 0, 0.1)$ (e) (α, p, θ, a)=$(2, 2, \pi/6, 0.2)$

(f) (g) (h)

Fig. 5 First row*: half Laplacian distribution for different values of* $\alpha \in \mathbb{R}_{++}$ (a)*, half generalized Gaussian distribution for different values of* $p \in \mathbb{R}_{++}$ *and* $\alpha = 1$ (b). Second row*: bivariate generalized Gaussian distribution for different values of* $\alpha \in \mathbb{R}_{++}$*,* $p \in \mathbb{R}_{++}$*,* $\theta \in [-\pi/2, \pi/2)$*,* $a \in (0, 1]$. Third row*: contour plots of the bivariate generalized Gaussian pdfs displayed in the second row.*

We stress that, despite their differences, the choices (hLd), (hLd-sv), and (hGGd-sv) correspond to the formulation of prior assumptions on the quantities $\|(\mathbf{D}\boldsymbol{u})_i\|_2$ for $i = 1, \ldots, N$. Clearly, this is somehow a "rigid" choice since it does not exploit the two-dimensionality of local image gradients $(\mathbf{D}\boldsymbol{u})_i = ((\mathbf{D}_{\boldsymbol{h}}\boldsymbol{u})_i, (\mathbf{D}_{\boldsymbol{v}}\boldsymbol{u})_i)$ and, possibly, the correlation between their horizontal and vertical components $(\mathbf{D}_{\boldsymbol{h}}\boldsymbol{u})_i$ and $(\mathbf{D}_{\boldsymbol{v}}\boldsymbol{u})_i$. To do so, a different, possibly space-variant, prior assumption imposing a priori information on the local image gradient can be made. Namely, after introducing a novel definition for function $\boldsymbol{z}(\boldsymbol{u})$ that reads

$$\boldsymbol{z}(\boldsymbol{u}) = [z_1(\boldsymbol{u}), \ldots, z_N(\boldsymbol{u})]^T, \quad z_i(\boldsymbol{u}) = (\mathbf{D}\boldsymbol{u})_i,$$

one can assume that $z_i(\boldsymbol{u})$ follows a bivariate generalized Gaussian distribution (BGGd; see (BGGd)) which is space-variant with respect to shape and scale and locally takes the form

(BGGd-sv)
$$\pi(z_i(\boldsymbol{u}); p_i, \boldsymbol{\Sigma}_i) = \frac{1}{2\pi|\boldsymbol{\Sigma}_i|^{1/2}} \frac{p_i}{\Gamma(2/p_i)\, 2^{2/p_i}} \exp\left(-\frac{1}{2}(z_i(\boldsymbol{u})^T \boldsymbol{\Sigma}_i^{-1} z_i(\boldsymbol{u}))^{p_i/2}\right),$$

where, for every i, the covariance matrix $\boldsymbol{\Sigma}_i \in \mathbb{R}^{2\times 2}$ is symmetric positive definite with determinant $|\boldsymbol{\Sigma}_i| > 0$. The general prior in (3.6) thus takes the form of the following nonstationary prior pdf on $\boldsymbol{u}$:

$$\begin{aligned}\pi(\boldsymbol{u} \mid \boldsymbol{\Theta}) &= c(\boldsymbol{\Theta}) \prod_{i=1}^{N} \left(\frac{1}{2\pi|\boldsymbol{\Sigma}_i|^{1/2}} \frac{p_i}{\Gamma(2/p_i)\, 2^{2/p_i}} \exp\left(-\frac{1}{2}((\mathbf{D}\boldsymbol{u})_i^T \boldsymbol{\Sigma}_i^{-1} (\mathbf{D}\boldsymbol{u})_i)^{\frac{p_i}{2}}\right)\right) \\ &= c(\boldsymbol{\Theta}) \prod_{i=1}^{N} \left(\frac{1}{2\pi|\boldsymbol{\Sigma}_i|^{1/2}} \frac{p_i}{\Gamma(2/p_i)\, 2^{2/p_i}}\right) \exp\left(-\frac{1}{2}\sum_{i=1}^{N} \left((\mathbf{D}\boldsymbol{u})_i^T \boldsymbol{\Sigma}_i^{-1} (\mathbf{D}\boldsymbol{u})_i\right)^{\frac{p_i}{2}}\right),\end{aligned} \tag{3.11}$$

where the normalization function c is given by (3.7) when, we recall, $\boldsymbol{z}(\boldsymbol{u}) = \mathbf{D}\boldsymbol{u}$, and when the pdf of $\boldsymbol{z}(\boldsymbol{u})$ given $\boldsymbol{\Theta}$ is selected as a BGGd.

For a better interpretation of such a choice, we now perform some simple manipulations to the generic ith term of the sum appearing in (BGGd-sv) so as to highlight how the information related to the local image scale and orientations is all encoded in the local covariance matrices $\boldsymbol{\Sigma_i}$. To this purpose, we consider the following eigenvalue decomposition:

$$\boldsymbol{\Sigma_i} = \mathbf{V_i}^T \mathbf{E_i} \mathbf{V_i}, \quad \mathbf{E_i} = \begin{pmatrix} e_i^{(1)} & 0 \\ 0 & e_i^{(2)} \end{pmatrix}, \quad e_i^{(2)} \geq e_i^{(1)} > 0, \quad \mathbf{V_i}^T \mathbf{V_i} = \mathbf{V_i} \mathbf{V_i}^T = \mathbf{I}_2, \tag{3.12}$$

where for every $i = 1, \ldots, N$, $e_i^{(1)}, e_i^{(2)}$ are the (positive) eigenvalues of $\boldsymbol{\Sigma_i}$ and $\mathbf{V_i}$ is an orthonormal (rotation) matrix to be made precise. We can thus rewrite the ith term of the sum in (3.11) as

$$\left((\mathbf{D}\boldsymbol{u})_i^T \boldsymbol{\Sigma_i}^{-1} (\mathbf{D}\boldsymbol{u})_i\right)^{\frac{p_i}{2}} = \left((\mathbf{D}\boldsymbol{u})_i^T \mathbf{V_i}^T \mathbf{E_i}^{-1} \mathbf{V_i} (\mathbf{D}\boldsymbol{u})_i\right)^{\frac{p_i}{2}} = \left\| \widetilde{\boldsymbol{\Lambda}}_i \mathbf{R}_{\text{-}\theta_i} (\mathbf{D}\boldsymbol{u})_i \right\|_2^{p_i}, \tag{3.13}$$

where

$$\widetilde{\boldsymbol{\Lambda}}_i = \begin{pmatrix} \widetilde{\lambda}_i^{(1)} & 0 \\ 0 & \widetilde{\lambda}_i^{(2)} \end{pmatrix} := \mathbf{E_i}^{-1/2} = \begin{pmatrix} 1/\sqrt{e_i^{(1)}} & 0 \\ 0 & 1/\sqrt{e_i^{(2)}} \end{pmatrix}, \tag{3.14}$$

$$\mathbf{R}_{\text{-}\theta_i} = \begin{pmatrix} \cos\theta_i & \sin\theta_i \\ -\sin\theta_i & \cos\theta_i \end{pmatrix} = \mathbf{V_i}, \quad \theta_i \in [-\pi/2, \pi/2), \tag{3.15}$$

and $\theta_i \in [-\pi/2, \pi/2)$ denotes the angle drawn locally with respect to the horizontal axis, as simple geometrical considerations show. By now introducing the two parameter vectors $\boldsymbol{\alpha} \in \mathbb{R}_{++}^N$ and $\boldsymbol{a} \in (0,1]^N$ with components

$$\alpha_i := \widetilde{\lambda}_i^{(1)} \in \mathbb{R}_{++}, \quad a_i := \frac{\widetilde{\lambda}_i^{(2)}}{\widetilde{\lambda}_i^{(1)}} \in (0,1], \quad i = 1, \ldots, N,$$

we have that the matrix $\widetilde{\mathbf{\Lambda}}_i$ in (3.15) can be equivalently rewritten as

$$\widetilde{\mathbf{\Lambda}}_i = \widetilde{\lambda}_i^{(1)} \begin{pmatrix} 1 & 0 \\ 0 & \widetilde{\lambda}_i^{(2)}/\widetilde{\lambda}_i^{(1)} \end{pmatrix} = \alpha_i \mathbf{\Lambda}_{a_i}, \quad \text{with } \mathbf{\Lambda}_{a_i} = \begin{pmatrix} 1 & 0 \\ 0 & a_i \end{pmatrix}. \tag{3.16}$$

Combining altogether, we have that (3.15)–(3.16) entail that the term in (3.13) can be indeed written as

$$\left((\mathbf{D}\boldsymbol{u})_i^T \mathbf{\Sigma}_{\boldsymbol{i}}^{-1} (\mathbf{D}\boldsymbol{u})_i \right)^{\frac{p_i}{2}} = \alpha_i^{p_i} \, \| \, \mathbf{\Lambda}_{a_i} \mathbf{R}_{\text{-}\theta_i} (\mathbf{D}\boldsymbol{u})_i \|_2^{p_i} \, . \tag{3.17}$$

Furthermore, based on (3.12) and (3.15)–(3.16), we observe that

$$\left| \mathbf{\Sigma}_i \right|^{-1/2} = \left| \mathbf{V}_i^T \mathbf{E}_i \mathbf{V}_i \right|^{-1/2} = \left| \mathbf{E}_i \right|^{-1/2} = \left| \widetilde{\mathbf{\Lambda}}_i \right| = \left| \alpha_i^2 \mathbf{\Lambda}_{a_i} \right| = \alpha_i^2 a_i > 0. \tag{3.18}$$

Plugging now (3.17) and (3.18) into the expression (3.11), we obtain the equivalent form

$$\pi(\boldsymbol{u} \mid \mathbf{\Theta}) = \frac{c(\mathbf{\Theta})}{(2\pi)^N} \left(\prod_{i=1}^{N} \frac{\alpha_i^2 \, p_i \, a_i}{\Gamma(2/p_i) \, 2^{2/p_i}} \right) \exp\left(- \sum_{i=1}^{N} \alpha_i^{p_i} \, \| \, \mathbf{\Lambda}_{a_i} \mathbf{R}_{\text{-}\theta_i} \, (\mathbf{D}\boldsymbol{u})_i \|_2^{p_i} \right), \tag{3.19}$$

where the vector of hyperparameters is here:

$$\mathbf{\Theta} = (\boldsymbol{\alpha}, \boldsymbol{p}, \boldsymbol{\theta}, \boldsymbol{a}) \in \mathbb{R}_{++}^{N \times 2} \times [-\pi/2, \pi/2)^N \times (0, 1]^N \, . \tag{3.20}$$

Compared to the univariate prior (3.10), prior (3.19)–(3.20) is characterized by two additional vectors of (space-variant) parameters $\theta_i \in [-\pi/2, \pi/2)$ and $a_i \in (0, 1]$, $i = 1. \ldots, N$. These parameters relate in fact to the bivariate nature of the BGGd in (BGGd-sv). In particular, the parameter θ_i represents the direction of the major axis of elliptical contour lines of the local BGGd, while a_i describes locally the eccentricity of the contour lines. More precisely, $a_i = 1$ corresponds to circular contour lines, i.e., to a maximally isotropic pdf, whereas for $a_i \approx 0$ the contour lines approach lines drawing the angle θ_i with respect to the horizontal axis, hence they are maximally anisotropic. The great flexibility of distribution in (BGGd-sv) is highlighted in Figures 5(c)–5(e), where the pdfs corresponding to the choice of different scalar parameters α_i, p_i, a_i, and θ_i are shown, while the corresponding contour plots are displayed in Figures 5(f)–5(h).

Remark 3.1. Note that the nonstationary prior in (3.9) reduces to the stationary TV prior in (3.8) for constant choices of the scale parameters $\alpha_i = \alpha \; \forall i$. Analogously, by setting $\alpha_i = \alpha$ and $p_i = p \; \forall i$, in (3.10), we recover the space-invariant prior corresponding to the TV_p regularizer in (TV_p). The same consideration holds for the DTV regularization term in (DTV), whose statistical counterpart is obtained starting from (3.19) and setting $\alpha_i = \alpha$, $p_i = 1$, $\theta_i = \theta$, and $a_i = a \; \forall i$. From a probabilistic point of view, note that moving from a stationary to a nonstationary framework corresponds to not assume identity in distribution for the functions $z_i(\boldsymbol{u})$ due to the space-variant nature of the distribution parameters, while keeping the independence assumption.

In section 4, we are going to bridge the Bayesian and the variational framework via the MAP approach, so that the relation between the prior pdfs recalled so far and the regularization terms that have been already introduced, or that will be discussed

Table 1 *Regularization terms introduced in section 1 (top) and defined in section 4 (bottom) corresponding to the stationary (i.i.d.) and nonstationary (for which only independence holds) priors recalled. For each of them, we report the function $z_i(\boldsymbol{u})$, together with the expression of the conditioned prior pdf and the parameters domain.*

	$\mathcal{R}(\boldsymbol{u},\boldsymbol{\Theta})$	$z_i(\boldsymbol{u})$	$\pi(z_i(\boldsymbol{u}) \mid \boldsymbol{\Theta}_i)$	$\boldsymbol{\Theta}_i$
stationary (i.i.d.)	TV	$\|(\mathbf{D}\boldsymbol{u})_i\|_2$	$\alpha \exp(-\alpha\, z_i(\boldsymbol{u}))$	$\alpha \in \mathbb{R}_{++}$
	TV_p	$\|(\mathbf{D}\boldsymbol{u})_i\|_2$	$\frac{\alpha\, p}{\Gamma(1/p)} \exp(-(\alpha z_i(\boldsymbol{u}))^p)$	$(\alpha, p) \in \mathbb{R}^2_{++}$
	DTV	$(\mathbf{D}\boldsymbol{u})_i$	$\frac{a}{(2\pi)} \frac{\alpha^2}{4} \exp\left(-\alpha \left\| \boldsymbol{\Lambda}_a \mathbf{R}_{-\theta}\, z_i(\boldsymbol{u}) \right\|_2\right)$	$(\alpha, \theta, a) \in \mathbb{R}_{++} \times [-\pi/2, \pi/2) \times (0, 1]$
nonstationary	WTV	$\|(\mathbf{D}\boldsymbol{u})_i\|_2$	$\alpha_i \exp(-\alpha_i\, z_i(\boldsymbol{u}))$	$\alpha_i \in \mathbb{R}_{++}$
	$\mathrm{WTV}^{sv}_{\boldsymbol{p}}$	$\|(\mathbf{D}\boldsymbol{u})_i\|_2$	$\frac{\alpha_i\, p_i}{\Gamma(1/p_i)} \exp(-(\alpha_i z_i(\boldsymbol{u}))^{p_i})$	$(\alpha_i, p_i) \in \mathbb{R}^2_{++}$
	$\mathrm{WDTV}^{sv}_{\boldsymbol{p}}$	$(\mathbf{D}\boldsymbol{u})_i$	$\frac{a_i}{(2\pi)} \frac{\alpha_i^2\, p_i}{\Gamma(2/p_i)\, 2^{2/p_i}} \exp\left(-\alpha_i^{p_i} \left\| \boldsymbol{\Lambda}_{a_i} \mathbf{R}_{-\theta_i}\, z_i(\boldsymbol{u}) \right\|_2^{p_i}\right)$	$(\alpha_i, p_i, \theta_i, a_i) \in \mathbb{R}^2_{++} \times [-\pi/2, \pi/2) \times (0, 1]$

later, will appear clear. However, we anticipate the mentioned connection and report in Table 1 the space-invariant and the space-variant regularization terms which are derived starting from the stationary and nonstationary prior pdfs listed above. For each pdf, we also report the expression of function $z_i(\boldsymbol{u})$ as well as the parameter domain. Notice that we have not explicitly considered the pdfs corresponding to the TV_p and to the DTV regularizer; however, the former corresponds to selecting a global hGGd for the gradient magnitudes, while the latter can be easily derived starting from the BGGd.

3.3. Hierarchical Modeling. The effort made in deriving the highly parametric prior distributions in the previous section would be vain if not coupled with an automatic and robust procedure for the estimation of the unknown parameters $\boldsymbol{\Theta}$. The choice of recasting the original problem in probabilistic terms makes it very natural to model the unknown vector $\boldsymbol{\Theta}$ as well as the unknown $\boldsymbol{u}$ as random variables. To do so, we thus need to introduce a further pdf encoding the a priori beliefs on $\boldsymbol{\Theta}$, which, in the following, will be denoted by $\pi(\boldsymbol{\Theta})$ and which will be referred to as *hyperprior*.

By proceeding as in (1.5), we seek the analytic expression of the joint posterior pdf, which, by leaving the dependence on $\boldsymbol{\Theta}$ explicit, is related to the prior and likelihood pdf through

$$\pi(\boldsymbol{u}, \boldsymbol{\Theta} \mid \boldsymbol{b}) = \frac{\pi(\boldsymbol{u}, \boldsymbol{\Theta})\pi(\boldsymbol{b} \mid \boldsymbol{u}, \boldsymbol{\Theta})}{\pi(\boldsymbol{b})} = \frac{\pi(\boldsymbol{u} \mid \boldsymbol{\Theta})\pi(\boldsymbol{\Theta})\pi(\boldsymbol{b} \mid \boldsymbol{u})}{\pi(\boldsymbol{b})},$$

where, in the last equality, we have used $\pi(\boldsymbol{u}, \boldsymbol{\Theta}) = \pi(\boldsymbol{u} \mid \boldsymbol{\Theta})\pi(\boldsymbol{\Theta})$ and $\pi(\boldsymbol{b} \mid \boldsymbol{u}, \boldsymbol{\Theta}) = \pi(\boldsymbol{b} \mid \boldsymbol{u})$. Proceeding by standard MAP estimation, we thus have that the sought solution pair $\{\boldsymbol{u}^*, \boldsymbol{\Theta}^*\}$ is the one maximizing the joint posterior pdf $\pi(\boldsymbol{u}, \boldsymbol{\Theta} \mid \boldsymbol{b})$, i.e.,

$$\{\boldsymbol{u}^*, \boldsymbol{\Theta}^*\} \in \arg\max_{\boldsymbol{u}, \boldsymbol{\Theta}} \{\pi(\boldsymbol{u} \mid \boldsymbol{\Theta})\pi(\boldsymbol{\Theta})\pi(\boldsymbol{b} \mid \boldsymbol{u})\},$$

or, equivalently,

$$\begin{aligned}\{\boldsymbol{u}^*, \boldsymbol{\Theta}^*\} &\in \arg\min_{\boldsymbol{u},\boldsymbol{\Theta}} \{-\ln \pi(\boldsymbol{u} \mid \boldsymbol{\Theta}) - \ln \pi(\boldsymbol{\Theta}) - \ln \pi(\boldsymbol{b} \mid \boldsymbol{u})\} \\ &= \arg\min_{\boldsymbol{u},\boldsymbol{\Theta}} \{-\ln c(\boldsymbol{\Theta}) - \ln \pi(\boldsymbol{z}(\boldsymbol{u}) \mid \boldsymbol{\Theta}) - \ln \pi(\boldsymbol{\Theta}) - \ln \pi(\boldsymbol{b} \mid \boldsymbol{u})\} ,\end{aligned} \tag{3.21}$$

where the evidence term $\pi(\boldsymbol{b})$ has been dropped as it does not depend on either $\boldsymbol{u}$ or $\boldsymbol{\Theta}$.

When tackling the joint model (3.21), two major difficulties arise, namely the computation of the highly dimensional constant $c(\boldsymbol{\Theta})$ and the choice of an efficient algorithmic scheme for the numerical solution of the minimization problem (3.21). Different strategies have been designed to overcome the former issue: most of them are based on a modification of the conditional prior $\pi(\boldsymbol{u} \mid \boldsymbol{\Theta})$ which comes from either approximating $c(\boldsymbol{\Theta})$ (see [122, 8, 7]) or neglecting it (see [155]). Here, we adopt this latter approach so that the joint hypermodel (3.21) takes the form

$$\{\boldsymbol{u}^*, \boldsymbol{\Theta}^*\} \in \arg\min_{\boldsymbol{u},\boldsymbol{\Theta}} \{-\ln \pi(\boldsymbol{z}(\boldsymbol{u}) \mid \boldsymbol{\Theta}) - \ln \pi(\boldsymbol{\Theta}) - \ln \pi(\boldsymbol{b} \mid \boldsymbol{u})\} . \tag{3.22}$$

Neglecting $c(\boldsymbol{\Theta})$ provides a significant simplification of the problem of interest. Nonetheless, as we will show in sections 7 and 9, such simplification will result in an efficient ML-type parameter estimation strategy which will be shown to produce meaningful results. Clearly, a more accurate study of (3.21) will require dealing explicitly with the computation of such constant by means, for instance, of analogous approaches such as those described in [147, 58].

From a numerical perspective, the solution of problem (3.22) can be addressed in different manners. A standard strategy illustrated in [34] is based on the design of an iterated sequential algorithm (IAS) which, for $k \geq 0$ and upon a suitable initialization for $\boldsymbol{u}^{(0)}$, reads

$$\boldsymbol{\Theta}^{(k+1)} \in \arg\min_{\boldsymbol{\Theta}} \left\{-\ln \pi(\boldsymbol{z}(\boldsymbol{u}^{(k)}) \mid \boldsymbol{\Theta}) - \ln \pi(\boldsymbol{\Theta})\right\} , \tag{3.23}$$

$$\boldsymbol{u}^{(k+1)} \in \arg\min_{\boldsymbol{u}} \left\{-\ln \pi(\boldsymbol{z}(\boldsymbol{u}) \mid \boldsymbol{\Theta}^{(k+1)}) - \ln \pi(\boldsymbol{b} \mid \boldsymbol{u})\right\} , \tag{3.24}$$

where the function $\boldsymbol{z}(\cdot)$ has been defined in section 3.2 depending on the specific form of the prior distribution at hand.

4. The Anatomy of Space-Variant Regularization Models. In this section, we derive the explicit expressions of the negative log-prior term $-\ln \pi(\boldsymbol{z}(\boldsymbol{u}) \mid \boldsymbol{\Theta})$, appearing in the cost function of (3.22), depending on the particular choice of the prior pdf among the ones described in section 3.2. For each considered prior, we will write explicitly the analytical form of the corresponding image regularizer, dissecting its properties in terms of regularization features and providing some intuitions on their sparsity promoting behavior.

4.1. From Nonstationary Priors to Space-Variant Regularizers. Recalling (TV), we start computing the negative logarithm of the stationary Gibbs TV prior in (3.8). We have

$$-\ln \pi(\boldsymbol{z}(\boldsymbol{u}) \mid \boldsymbol{\Theta}) = \alpha \mathrm{TV}(\boldsymbol{u}) - N \ln \alpha. \tag{4.1}$$

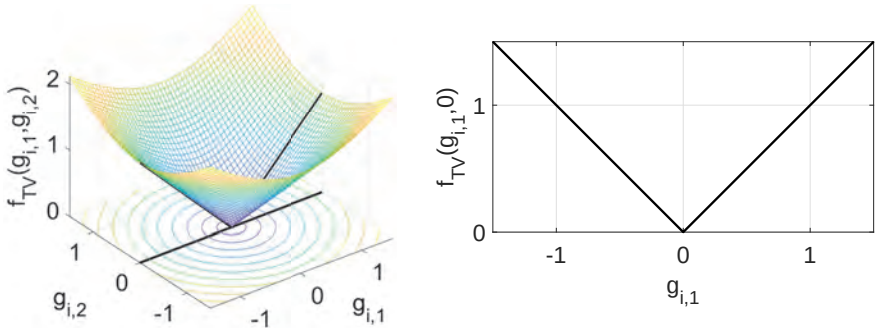
We now rewrite TV as

$$\mathrm{TV}(\boldsymbol{u}) \;=\; \sum_{i=1}^{N} f_{\mathrm{TV}}\left((\mathbf{D}\boldsymbol{u})_i\right), \tag{4.2}$$

where the space-invariant and nonparametric function $f_{\mathrm{TV}} : \mathbb{R}^2 \to \mathbb{R}_+$ is defined by

$$f_{\mathrm{TV}}(\boldsymbol{g}_i) \;:=\; \|\boldsymbol{g}_i\|_2\,, \quad \boldsymbol{g}_i = (g_{i,1}, g_{i,2}) \in \mathbb{R}^2 \tag{4.3}$$

and is referred to in the following as the TV *gradient penalty function*. As is well known, TV is bounded from below by zero, (nonstrictly) convex, noncoercive due to $\mathrm{null}(\mathbf{D}) \neq \{\mathbf{0}_N\}$, and nonsmooth. This last property is indeed responsible for the good gradient sparsity-promoting effect of TV, which favors piecewise-constant solutions. The TV gradient penalty function f_{TV} in (4.3) is shown in Figure 6(a).



(a) f_{TV} (b) 1D section of f_{TV} along the $g_{i,1}$-axis

Fig. 6 *Space-invariant gradient penalty function f_{TV} defined in* (4.3) *for the TV regularizer* (4.2).

To analyze in detail the properties of the TV regularizer, it is useful to consider the 1D sections of the gradient penalty function f_{TV} along straight lines passing through the origin of the penalty domain and having direction defined by the angle $\varphi \in [-\pi, +\pi)$. Using a standard (arc-length) parametrization for straight lines, namely, $\{\, g_{i,1}(t;\varphi) = t\,\cos(\varphi),\;\; g_{i,2}(t;\varphi) = t\,\sin(\varphi),\;\; t \in \mathbb{R}\,\}$, the sections of f_{TV} in (4.3) read

$$s_i(t;\varphi) = \,|t|\,, \quad t \in \mathbb{R}, \quad i = 1, \ldots, N\,. \tag{4.4}$$

In Figure 6(b) we show one section of f_{TV} along the direction defined by the angle $\varphi = 0$, i.e., the $g_{i,1}$-axis. However, as the expression (4.4) does not depend on φ, we deduce that the same figure could be obtained by representing the section corresponding to *any* φ, for *any* pixel location i. The TV penalty f_{TV} in (4.3)—hence, the (TV) regularizer—is in fact space and rotationally invariant (i.e., *isotropic*).

Being isotropic, TV does not take explicitly into account directionality properties in the image. Moreover, the presence of a fixed, global exponent $p = 1$ for the norms in the penalty (4.3) and of a global scale parameter $\alpha > 0$ in (3.8) and, hence, in the negative log-prior (4.1) makes the TV regularizer unable to adapt the strength (associated to α in (4.1)) or the nature (associated to the exponent of the norm in (4.3)) of the gradient sparsity-promoting effect to the local contents of the image to be recovered.

In the following, we inspect how the nonstationary priors introduced in section 3.2 can favor local regularization features, namely strength, sharpness, and directionality.

4.1.1. Local Regularization Strength. Recalling section 3.2, the first and probably the easiest way to make the (TV) regularizer spatially flexible consists in allowing for a different amount of regularization at every pixel in the image. From a Bayesian perspective, this corresponds to assuming a nonstationary hL prior distribution for the gradient magnitudes of $\boldsymbol{u}$. By computing the negative logarithm in (3.9), we have

$$-\ln \pi(\boldsymbol{z}(\boldsymbol{u}) \mid \boldsymbol{\Theta}) = \mathrm{WTV}(\boldsymbol{u}; \boldsymbol{\Theta}) - \sum_{i=1}^{N} \ln \alpha_i \,,$$

where the space-variant WTV regularizer is defined in terms of hyperparameters $\boldsymbol{\Theta} = \boldsymbol{\alpha}$ and reads

$$\mathrm{WTV}(\boldsymbol{u}; \boldsymbol{\alpha}) := \sum_{i=1}^{N} \alpha_i \|(\mathbf{D}\boldsymbol{u})_i\|_2 \,, \quad \boldsymbol{\alpha} \in \mathbb{R}^N_{++} \,. \tag{WTV}$$

Analogously to (TV), the (WTV) regularizer can be equivalently rewritten as

$$\mathrm{WTV}(\boldsymbol{u}; \boldsymbol{\alpha}) = \sum_{i=1}^{N} f_{\mathrm{WTV}}\left((\mathbf{D}\boldsymbol{u})_i; \alpha_i\right), \quad \alpha_i \in \mathbb{R}_{++},$$

where the gradient penalty function $f_{\mathrm{WTV}} : \mathbb{R}^2 \to \mathbb{R}_+$ now depends locally on the parameter α_i and reads

$$f_{\mathrm{WTV}}(\boldsymbol{g}_i; \alpha_i) := \alpha_i \|\boldsymbol{g}_i\|_2 \,, \quad \boldsymbol{g}_i = (g_{i,1}, g_{i,2}) \in \mathbb{R}^2 \,. \tag{4.5}$$

Similarly as for the (TV) regularizer, (WTV) is still convex and nondifferentiable. However, the sparsity-promoting effect can now be locally modulated thanks to the presence of the local weights α_i. To highlight this feature, we report in the first column of Figure 7 (i.e., Figures 7(a),(d),(g)) the graphs of the WTV gradient penalty function f_{WTV} defined in (4.5) for three different values $\alpha_j = 1$, $\alpha_k = 1.3$, $\alpha_l = 0.7$ of the scale parameter, respectively, assuming that they represent the local weights of the WTV regularizer at different pixel positions $i \in \{j, k, l\}$. These three graphs share the same inverted right-circular conical shape with vertex at the origin as the TV penalty drawn in Figure 6, with the one in Figure 7(a) coinciding with the TV penalty. Different values of the weight yield different slopes of the conical lateral surface; note that $\|\boldsymbol{\nabla} f_{\mathrm{WTV}}(g_{i,1}, g_{i,2})\|_2 = \alpha_i$ for any $(g_{i,1}, g_{i,2}) \in \mathbb{R}^2 \setminus \{(0,0)\}$, and, hence, different local regularization strengths. The larger (smaller) is the local weight α_i, the more (less) strongly the WTV regularizer will force $\|(\mathbf{D}\boldsymbol{u})_i\|_2$ to be small.

Similarly as before, we show the 1D sections of the three WTV penalty functions along the two directions defined by angles $\varphi = 0$ (solid lines) and $\varphi = \pi/2$ (dashed lines), corresponding to the $g_{i,1}$- and $g_{i,2}$-axis in the 3D plots in Figures 7(a),(d),(g) and, for better readability, in Figure 7(j), left. Like TV, the WTV regularizer is isotropic, hence the two sections—actually, any section along straight lines passing through the origin—of each of the three penalties coincide. Despite their space-variant feature, these sections are in fact still rotationally invariant as they take the form

$$s_i(t; \varphi) = \alpha_i |t| \,, \quad t \in \mathbb{R}, \quad i = 1, \ldots, N \,. \tag{4.6}$$

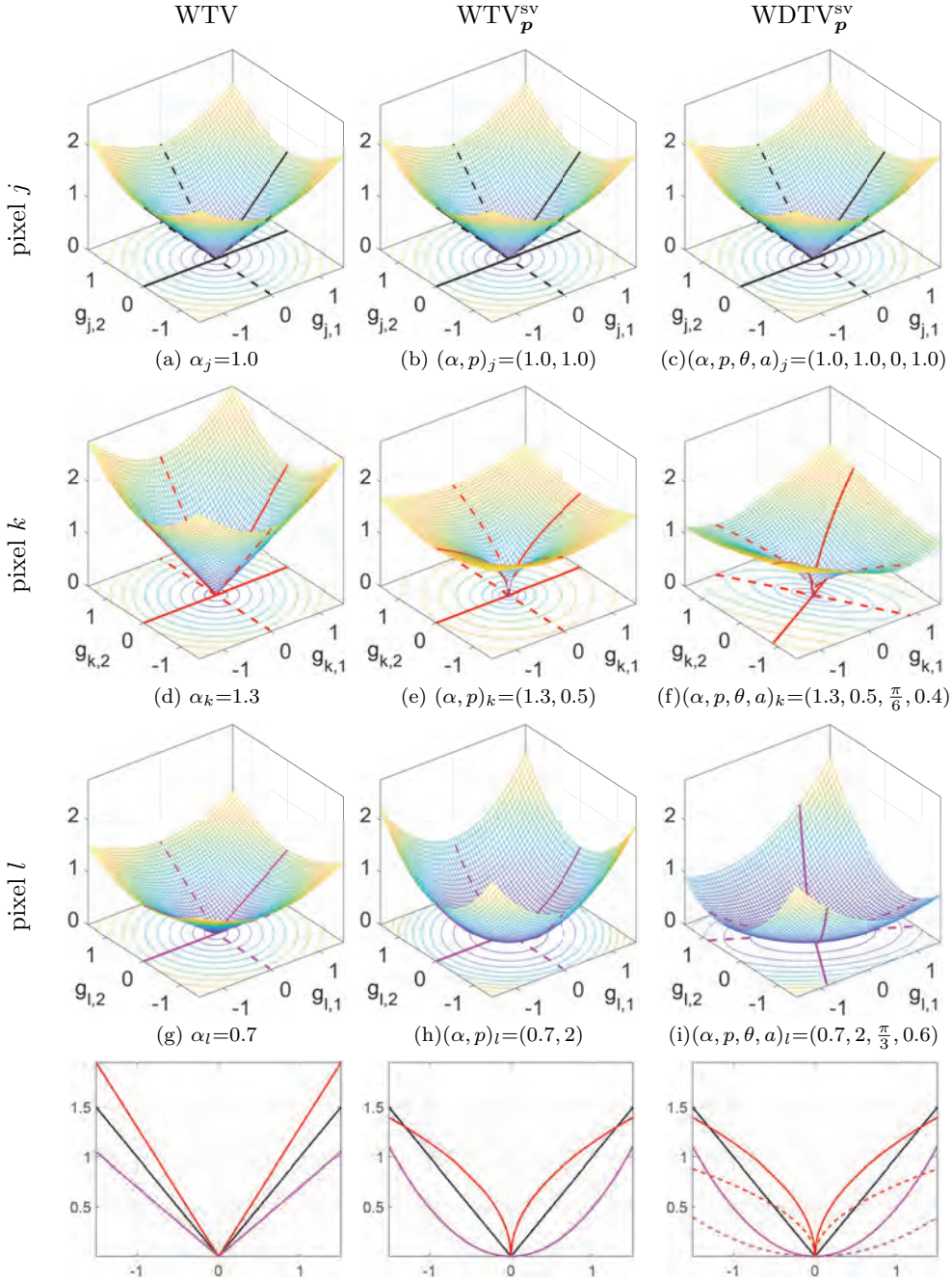


Fig. 7 *Graphs of the gradient penalty functions defined in* (4.5), (4.7), *and* (4.10) *for the space-variant* (WTV), (WTV$^{sv}_{\boldsymbol{p}}$), *and* (WDTV$^{sv}_{\boldsymbol{p}}$) *regularizers, respectively.*

Finally, one can notice from (4.6) and from Figure 7(j), left, that all sections are nothing but positively scaled versions of the absolute value function of scale parameter α_i, i.e., of the TV sections in (4.4).

Due to its ability to promote local TV smoothing, we remark that the WTV regularizer has been proposed and studied in several papers (e.g., [83, 84, 82] and many more) from an analytical point of view and motivated by means of analogous probabilistic arguments in [29].

4.1.2. Local Regularization Sharpness. As previously mentioned, the weights α_i in the (WTV) regularizer allow us to locally tune the strength of the gradient-sparsity promotion effect of the regularization which, by construction, is of fixed TV type. In fact, the presence of a global exponent 1 for the gradient norms in the definition of WTV does not allow us to change, either globally or locally, the *sharpness* of the associated gradient penalty functions, hence the nature of the involved sparsity promotion.

This motivates the introduction of a second set of space-variant parameters $p_i > 0$, $i = 1, \dots, N$, those being the exponents of the gradient norms in the (WTV) definition and corresponding to the local shape parameters of the associated hGG pdf; see Definition 2.6.

We proceed as above and compute the negative logarithm of the nonstationary hGG prior (3.10), thus getting

$$-\ln \pi(\boldsymbol{u} \mid \boldsymbol{\Theta}) = \mathrm{WTV}^{\mathrm{sv}}_{\boldsymbol{p}}(\boldsymbol{u}; \boldsymbol{\Theta}) - \sum_{i=1}^{N} \ln \frac{\alpha_i p_i}{\Gamma(1/p_i)}\,,$$

where the space-variant $\mathrm{WTV}^{\mathrm{sv}}_{p}$ regularizer, depending on the hyperparameters $\boldsymbol{\Theta} = (\boldsymbol{\alpha}, \boldsymbol{p})$, is defined by

$$(\mathrm{WTV}^{sv}_{\boldsymbol{p}}) \qquad \mathrm{WTV}^{\mathrm{sv}}_{\boldsymbol{p}}(\boldsymbol{u}; \boldsymbol{\alpha}, \boldsymbol{p}) := \sum_{i=1}^{N} \alpha_i^{p_i} \|(\mathbf{D}\boldsymbol{u})_i\|_2^{p_i}, \quad (\boldsymbol{\alpha}, \boldsymbol{p}) \in \mathbb{R}^{N\times 2}_{++}.$$

Like (WTV), the $(\mathrm{WTV}^{sv}_{\boldsymbol{p}})$ regularizer can be rewritten in terms of a parametric, space-variant gradient penalty function, namely,

$$\mathrm{WTV}^{\mathrm{sv}}_{\boldsymbol{p}}(\boldsymbol{u}; \boldsymbol{\alpha}, \boldsymbol{p}) = \sum_{i=1}^{N} f_{\mathrm{WTV}^{\mathrm{sv}}_{\boldsymbol{p}}}\left((\mathbf{D}\boldsymbol{u})_i; \alpha_i, p_i\right), \quad (\alpha_i, p_i) \in \mathbb{R}^2_{++},$$

where the function $f_{\mathrm{WTV}^{\mathrm{sv}}_{\boldsymbol{p}}}(\cdot)$ now reads

$$f_{\mathrm{WTV}^{\mathrm{sv}}_{\boldsymbol{p}}}(\boldsymbol{g}_i; \alpha_i, p_i) = \alpha_i^{p_i} \|\boldsymbol{g}_i\|_2^{p_i}\,, \quad \boldsymbol{g}_i = (g_{i,1}, g_{i,2}) \in \mathbb{R}^2\,. \tag{4.7}$$

Like (TV) and (WTV), the $(\mathrm{WTV}^{sv}_{\boldsymbol{p}})$ regularizer is bounded below by zero, continuous and noncoercive. However, its other regularity properties depend on the actual values of the parameters p_i. If $p_i \geq 1$ for any $i = 1, \dots, N$, then $\mathrm{WTV}^{\mathrm{sv}}_{p}$ is convex, whereas it is nonconvex if there exists at least one i such that $p_i < 1$. Then it is differentiable whenever $p_i > 1$ for any i; otherwise it is nonsmooth.

In the second column of Figure 7 we show the graph of the $\mathrm{WTV}^{\mathrm{sv}}_{p}$ gradient penalty function $f_{\mathrm{WTV}^{\mathrm{sv}}_{\boldsymbol{p}}}$ defined in (4.7) for three different parameter configurations $(\alpha_j, p_j) = (1, 1)$, $(\alpha_k, p_k) = (1.3, 0.5)$, and $(\alpha_l, p_l) = (0.7, 2)$, where, we remark,

the scale parameter values α_j, α_k, α_l are the same as for the WTV penalties reported in the first column. In case of unitary scale and shape parameters (see Figure 7(b)) the $\mathrm{WTV}_p^{\mathrm{sv}}$ penalty coincides with the TV penalty, and, more generally, for $p_j = p_k = p_l = 1$, the $\mathrm{WTV}_p^{\mathrm{sv}}$ penalties coincide with the WTV penalties. For nonunitary shape parameters, the $\mathrm{WTV}_p^{\mathrm{sv}}$ penalty function can assume different shapes, ranging from nonconvex and nondifferentiable ones ($p_i < 1$, Figure 7(e)) to strongly convex and differentiable ones ($p_i > 1$, Figure 7(h)). The degrees of freedom encoded by the shape parameters p_i thus provide the $\mathrm{WTV}_p^{\mathrm{sv}}$ regularizer with the ability to adapt its gradient sparsity-promoting effect to the local image content. In particular, $p_i > 1$—typically, $p_i \geq 2$—should be used to avoid TV staircasing in correspondence with smooth image regions, whereas $p_i < 1$—typically, $p_i \leq 0.5$—should be used in piecewise-constant regions to mitigate the undesirable contrast loss effect of TV.

Similarly as for the previous (TV) and (WTV) regularizers, let us now take a look at the 1D sections of the $\mathrm{WTV}_p^{\mathrm{sv}}$ penalty in (4.7) for $\varphi \in [-\pi, \pi)$, which read

$$s_i(t;\varphi) = \alpha_i^{p_i} \, |t|^{p_i} \,, \quad t \in \mathbb{R}, \quad i = 1, \ldots, N \,. \tag{4.8}$$

By looking at the plot of such sections shown in Figure 7(j), center, it is clear how the value of parameter p_i can substantially change the regularization effect at each pixel. In particular, by comparing the red, black, and magenta sections in Figure 7(j), one can notice that for $p_i < 1$ small gradients are induced to be zero in a stronger way than for $p_i = 1$, but large gradients are less penalized (weaker contrast loss effect). On the other hand, for $p_i > 1$ the sparsity-promoting effect is no longer present as the gradient penalty function is differentiable in $t = 0$. More generally, for $p_i > 1$ small gradients are less penalized than for $p_i = 1$, whereas large gradients are more penalized. We finally remark that, like for TV and WTV, the $\mathrm{WTV}_p^{\mathrm{sv}}$ sections in (4.8) do not depend on the direction angle φ, hence the ($\mathrm{WTV}_{\boldsymbol{p}}^{sv}$) regularizer still falls in the class of isotropic regularizers. This is visually confirmed by the $\mathrm{WTV}_p^{\mathrm{sv}}$ penalties shown in Figures 7(b),(e),(h), which are rotationally invariant (i.e., have circular level curves), and by the penalty sections along the $g_{i,1}$-axis and the $g_{i,2}$-axis, which coincide as is evident from Figure 7(j), center.

The $\mathrm{WTV}_{\boldsymbol{p}}^{sv}$ regularizer was first introduced in a simplified version, i.e., with $\alpha_i = \alpha \; \forall i$, and interpreted in a probabilistic framework in [99]. The general case with space-variant weights was discussed in [98].

4.1.3. Local Anisotropy. As shown above, the (TV), (WTV), and ($\mathrm{WTV}_{\boldsymbol{p}}^{sv}$) regularizers are isotropic. For this reason, such regularizers are not able to exploit any information on the directionality of local image structures and, hence, to drive their local nonlinear diffusion effect along specific directions only. As motivated in section 1.3, this can be a limitation, especially for images presenting local structures characterized by well-defined orientations. As illustrated in section 3.2, to circumvent this limitation, a nonstationary BGG prior can be assumed for modeling the local distribution of gradients of $\boldsymbol{u}$.

By computing the negative logarithm of the nonstationary BGG prior in (3.19), we have

$$-\ln \pi(\boldsymbol{z}(\boldsymbol{u}) \mid \boldsymbol{\Theta}) = \mathrm{WDTV}_{\boldsymbol{p}}^{\mathrm{sv}}(\boldsymbol{u};\boldsymbol{\Theta}) - \sum_{i=1}^{N} \ln\left(\frac{\alpha_i^2 \, p_i \, a_i}{\Gamma(2/p_i)\, 2^{2/p_i}}\right) + N \ln(2\pi) \,,$$

where the space-variant $\mathrm{WDTV}_p^{\mathrm{sv}}$ regularizer is defined in terms of the hyperparam-

eters $\boldsymbol{\Theta} = (\boldsymbol{\alpha}, \boldsymbol{p}, \boldsymbol{\theta}, \boldsymbol{a})$ and reads

$$(\mathrm{WDTV}^{sv}_{\boldsymbol{p}}) \qquad \mathrm{WDTV}^{\mathrm{sv}}_{\boldsymbol{p}}(\boldsymbol{u}; \boldsymbol{\alpha}, \boldsymbol{p}, \boldsymbol{\theta}, \boldsymbol{a}) \;=\; \sum_{i=1}^{N} \alpha_i^{p_i} \, \|\boldsymbol{\Lambda}_{a_i} \mathbf{R}_{-\theta_i} (\mathbf{D}\boldsymbol{u})_i\|_2^{p_i} \,,$$
$$(\boldsymbol{\alpha}, \boldsymbol{p}, \boldsymbol{\theta}, \boldsymbol{a}) \;\in\; \mathbb{R}^{N\times 2}_{++} \times [-\pi/2, \pi/2)^N \times (0,1]^N \,,$$

where the orthogonal (rotation) matrices $\mathbf{R}_{-\theta_i}$ and the diagonal matrices $\boldsymbol{\Lambda}_{a_i}$ have been defined in (3.15) and (3.16), respectively.

Note that the $\mathrm{WDTV}^{sv}_{\boldsymbol{p}}$ regularizer can also be written in terms of its parametric, space-variant gradient penalty functions as

$$\mathrm{WDTV}^{\mathrm{sv}}_{\boldsymbol{p}}(\boldsymbol{u}; \boldsymbol{\alpha}, \boldsymbol{p}, \boldsymbol{\theta}, \boldsymbol{a}) \;=\; \sum_{i=1}^{N} f_{\mathrm{WDTV}^{\mathrm{sv}}_{\boldsymbol{p}}} \left((\mathbf{D}\boldsymbol{u})_i; \alpha_i, p_i, \theta_i, a_i \right), \tag{4.9}$$
$$(\alpha_i, p_i, \theta_i, a_i) \;\in\; \mathbb{R}^2_{++} \times [-\pi/2, \pi/2) \times (0,1] \,,$$

with

$$f_{\mathrm{WDTV}^{\mathrm{sv}}_{\boldsymbol{p}}}(\boldsymbol{g}_i; \alpha_i, p_i, \theta_i, a_i) \;=\; \alpha_i^{p_i} \, \|\boldsymbol{\Lambda}_{a_i} \mathbf{R}_{-\theta_i} \boldsymbol{g}_i\|_2^{p_i} \;, \quad \boldsymbol{g}_i = (g_{i,1}, g_{i,2}) \in \mathbb{R}^2 \,. \tag{4.10}$$

Since $a_i \in (0,1]$ for any i, matrices $\mathbf{M}_i := \boldsymbol{\Lambda}_{a_i} \mathbf{R}_{-\theta_i} \in \mathbb{R}^{2\times 2}$ are all nonsingular. As a consequence, the $(\mathrm{WDTV}^{sv}_{\boldsymbol{p}})$ regularizer shares the same analytical properties as the $(\mathrm{WTV}^{sv}_{\boldsymbol{p}})$ regularizer. In particular, it is worth noting that the $(\mathrm{WDTV}^{sv}_{\boldsymbol{p}})$ regularizer reduces to the rotationally invariant $(\mathrm{WTV}^{sv}_{\boldsymbol{p}})$ regularizer in the special case $a_i = 1$ for any i, independently of the directionality parameters θ_i.

In the last column of Figure 7 we show the graph of the $\mathrm{WDTV}^{\mathrm{sv}}_{\boldsymbol{p}}$ gradient penalty function $f_{\mathrm{WDTV}^{\mathrm{sv}}_{\boldsymbol{p}}}$ defined in (4.10) for three different parameter configurations $(\alpha_j, p_j, \theta_j, a_j) = (1, 1, 0, 1)$, $(\alpha_k, p_k, \theta_k, a_k) = (1.3, 0.5, \pi/6, 0.4)$, and $(\alpha_l, p_l, \theta_l, a_l) = (0.7, 2, \pi/3, 0.6)$, where the scale and shape parameter values (α_j, p_j), (α_k, p_k), and (α_l, p_l) are the same as for the $\mathrm{WTV}^{\mathrm{sv}}_{\boldsymbol{p}}$ penalties (second column). It is clear from these figures that the degrees of freedom represented by parameters a_i allow us to make the $\mathrm{WDTV}^{\mathrm{sv}}_{\boldsymbol{p}}$ regularizer locally anisotropic, in the sense that it can locally penalize the gradient $\boldsymbol{g}_i = (\mathbf{D}\boldsymbol{u})_i$ with different strength according to its direction. The level curves of the penalties in Figures 7(f),(i) corresponding to $a_i < 1$ are elliptical and not circular as for the case $a_i = 1$ in Figure 7(c). Furthermore, the smaller a_i, the more eccentric the ellipses and, hence, the more anisotropic the regularizer. The local directional parameters θ_i represent local image directions along which a stronger regularization effect is typically desired (typically, edge direction). We observe that the elliptical level curves of the penalties in Figures 7(f),(i) are rotated of angle θ_i counterclockwise, with the minor and major axes aligned along the directions defined by θ_i and $\theta_i + \pi/2$, respectively, and that the 1D sections of the $\mathrm{WDTV}^{\mathrm{sv}}_{\boldsymbol{p}}$ penalty functions in (4.10) along directions defined by angle φ take the form

$$s_i(t; \varphi) = \left(\cos^2(\varphi - \theta_i) + a_i^2 \sin^2(\varphi - \theta_i) \right)^{p_i/2} \alpha_i^{p_i} \, |t|^{p_i} \,, \quad t \in \mathbb{R}, \quad i = 1, \ldots, N \,.$$

It is a simple calculation verifying that for any fixed θ_i, a_i, the positive real coefficient in brackets takes its maximum (equal to 1) and minimum (equal to $a_i^{p_i}$) values for $\varphi = \theta_i$ and $\varphi = \theta_i + \pi/2$, respectively. This entails that the 1D sections of the $\mathrm{WDTV}^{\mathrm{sv}}_{\boldsymbol{p}}$ penalty along the dominant direction θ_i and its orthogonal $\theta_i + \pi/2$ are those characterized by the strongest and the weakest regularization effects, respectively.

Note that the sections exhibit the same sharpness, i.e., the same shape, but they are differently scaled; see the pairs of solid/dashed red and magenta curves in Figure 7(j), right.

The $\mathrm{WDTV}^{sv}_{\boldsymbol{p}}$ regularizer was first introduced and analyzed in probabilistic settings in [27].

4.1.4. Comparing Regularizers: Proximal Operators. In order to gain more insights on the regularization effects yielded by the different gradient penalty functions introduced in the previous sections and, consequently, on the different space-variant regularizers considered, we compare in this section the proximal operators $\mathrm{prox}^{\beta}_{f} : \mathbb{R}^2 \rightrightarrows \mathbb{R}^2,\ f \in \left\{f_{\mathrm{TV}}, f_{\mathrm{WTV}}, f_{\mathrm{WTV}^{\mathrm{sv}}_{p}}, f_{\mathrm{WDTV}^{\mathrm{sv}}_{p}}\right\}$ (see Definition 2.8) associated to the TV, WTV, $\mathrm{WTV}^{sv}_{\boldsymbol{p}}$, and $\mathrm{WDTV}^{sv}_{\boldsymbol{p}}$ penalty functions defined in (4.3), (4.5), (4.7), and (4.10), respectively.

Having fixed the value of parameter $\beta \in \mathbb{R}_{++}$, and regarding $\boldsymbol{w}$ as an input image gradient vector to be regularized, the 2D vector field $\boldsymbol{e}^{\beta}_{f} : \mathbb{R}^2 \rightrightarrows \mathbb{R}^2$ defined by

$$\boldsymbol{e}^{\beta}_{f}(\boldsymbol{w}) := \mathrm{prox}^{\beta}_{f}(\boldsymbol{w}) - \boldsymbol{w}, \quad \boldsymbol{w} \in \mathbb{R}^2\,, \tag{4.11}$$

can be studied to represent the regularization effect of the gradient penalty function considered on $\boldsymbol{w}$.

Analytical expressions of the proximal operators of the gradient penalty functions in (4.3), (4.5), (4.7), and (4.10) associated to the (TV), (WTV), ($\mathrm{WTV}^{sv}_{\boldsymbol{p}}$), and ($\mathrm{WDTV}^{sv}_{\boldsymbol{p}}$) regularizers have been previously studied in [159], [101], and [27], respectively, and are discussed (for completeness) in section 8.3 of this review. Based on those expressions, we thus compute the vector field $\boldsymbol{e}^{\beta}_{f}$ in (4.11) for each of the nine penalty functions considered in Figure 7 and report the results in Figure 8. In order to allow for a meaningful comparison between penalties, the same proximal parameter value $\beta = 3$ has been used.

We first remark that for all the considered gradient penalty functions f and parameters $\beta > 0$ we have $\mathrm{prox}^{\beta}_{f}(\boldsymbol{0}) = \boldsymbol{0} \implies \boldsymbol{e}^{\beta}_{f}(\boldsymbol{0}) = \boldsymbol{0}$, and that for all penalties with shape parameter $p_i \leq 1$—namely, the penalties in Figures 7(a)–(g)—there exists a region in the $\boldsymbol{w}$ domain (with center in the origin, size depending on β, and shape depending on the penalty itself) for which we have $\mathrm{prox}^{\beta}_{f}(\boldsymbol{w}) = \boldsymbol{0} \implies \boldsymbol{e}^{\beta}_{f}(\boldsymbol{w}) = -\boldsymbol{w}$. This means that any input gradient vector $\boldsymbol{w}$ belonging to this region is "completely" regularized by the gradient penalty function, in the sense that it is proximal-mapped to the null gradient vector. For visualization purposes, such sparsity-promoting regions are depicted in green (without showing the arrows pointing towards the origin) in the vector field representations of Figures 8(a)–(g).

As expected, for isotropic penalties—namely, the WTV and $\mathrm{WTV}^{\mathrm{sv}}_{p}$ penalties shown in the first two columns of Figure 7 and the $\mathrm{WDTV}^{\mathrm{sv}}_{p}$ penalty with unitary anisotropy parameter depicted in Figure 7(c)—the associated vector fields $\boldsymbol{e}^{\beta}_{f}$ are radial with vectors pointing towards the origin and the sparsity-promoting regions are circularly shaped; see Figures 8(a)–(e),(g). This means that the regularization effect yielded by the isotropic penalties depicted in Figures 7(a)–(e),(g),(h) on $\boldsymbol{w}$ is only a shrinkage of its norm $\|\boldsymbol{w}\|_2$, namely,

$$\mathrm{prox}^{\beta}_{f}(\boldsymbol{w}) = \xi\,\boldsymbol{w} \Longrightarrow \boldsymbol{e}^{\beta}_{f}(\boldsymbol{w}) = -(1-\xi)\,\boldsymbol{w},$$

with a shrinkage coefficient $\xi \in [0, 1)$ only depending on the norm itself. This result has been proved, e.g., in [101, Proposition 1], where analytical expressions for ξ as a

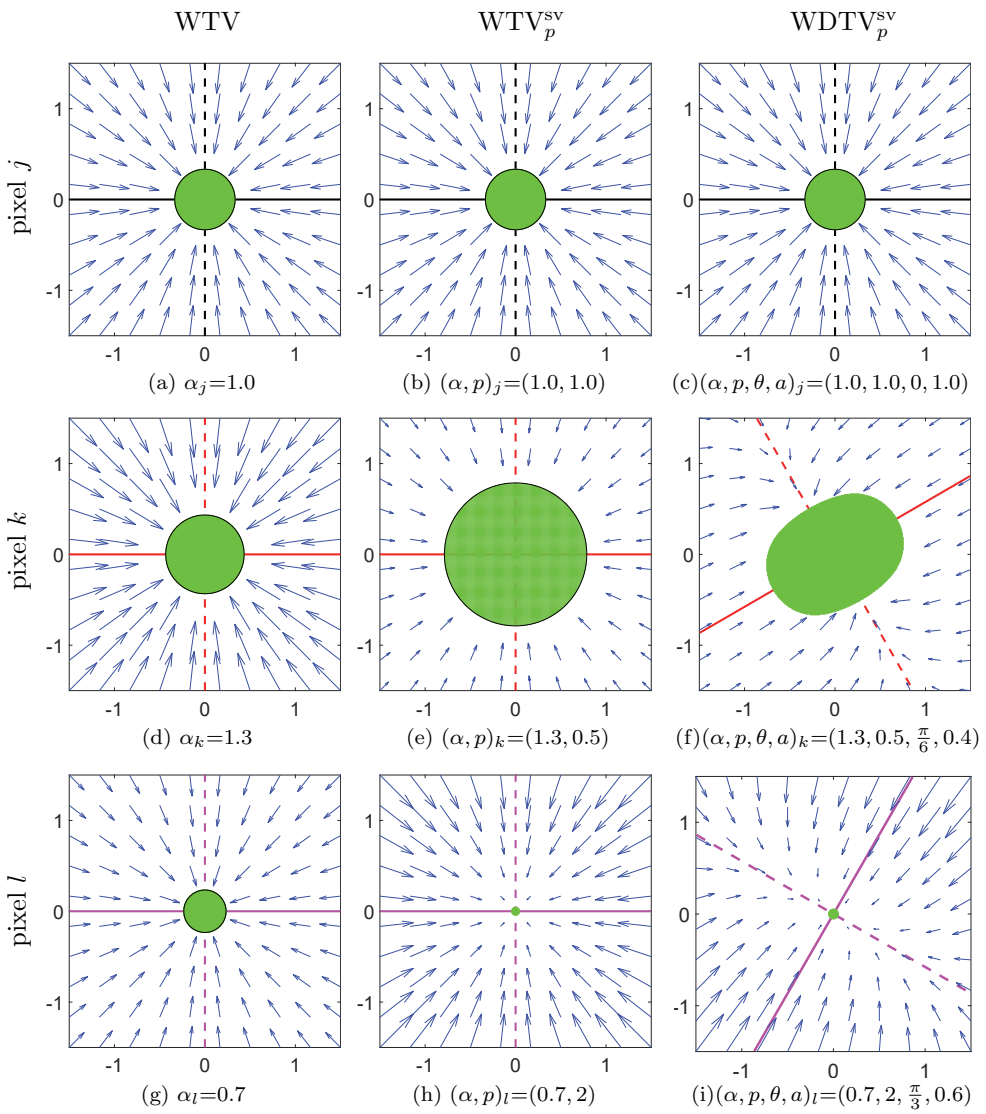


Fig. 8 *2D vector fields* $\boldsymbol{e}_f^\beta : \mathbb{R}^2 \rightrightarrows \mathbb{R}^2$ *in* (4.11)*, representing the effect of the proximal operator* prox_f^β *on a (gradient) vector* $\boldsymbol{w} = (w_1, w_2)$*, for the same gradient penalty functions* f *shown in Figures* 7(a)–(i) *and for a fixed proximity parameter* $\beta = 3$*.*

function of $\|\boldsymbol{w}\|_2$ as well as of the shape and proximal parameters have been given for the proximal operator of a TV$_p$ penalty of the form $f(\boldsymbol{g}; p) = \|\boldsymbol{g}\|_2^p$. Since it follows immediately from Definition 2.8 that

$$\mathrm{prox}_f^\beta(\boldsymbol{w}) = \mathrm{prox}_{\widetilde{f}}^{\widetilde{\beta}}(\boldsymbol{w}) \quad \forall \boldsymbol{w} \in \mathbb{R}^2$$

with $\widetilde{f}(\boldsymbol{g}; \alpha, p) = \alpha^p \|\boldsymbol{w}\|_2^p$ and $\widetilde{\beta} = \alpha^p \beta$, then the results in [101] can be straightforwardly extended to cover the more general case of a WTV$_p$ penalty. These results provide an analytical interpretation of the visual results reported in Figures 7(a)–(e),(g),(h). In particular, by observing the vector fields depicted in these figures, it

is clear how larger scale parameter values α_i in the WTV penalty yield stronger gradient shrinkage effects as well as sparsity-promoting regions of larger radii. Then, by comparing the vector fields in Figures 7(d),(e), one can notice that for a fixed scale parameter α_i, decreasing the shape parameter p_i (starting from $p_i = 1$) in the $\mathrm{WTV}_p^{\mathrm{sv}}$ penalty yields weaker shrinkage effects on gradients outside the sparsity-promoting regions but larger radii of these regions. Finally, Figures 7(g),(h) show that increasing p_i (for a fixed α_i and starting from $p_i = 1$) in the $\mathrm{WTV}_p^{\mathrm{sv}}$ penalty yields stronger gradient shrinkage effects and, for any $p_i > 1$, the sparsity-promoting regions reduce to the point $\boldsymbol{w} = \mathbf{0}_2$.

Clearly, the vector fields in Figures 8(f),(i), associated to the $\mathrm{WDTV}_p^{\mathrm{sv}}$ anisotropic penalties, are not radial. To be more precise, they are radial only when restricted to input vectors $\boldsymbol{w}$ lying on the two straight lines having direction defined by angles θ_i and $\theta_i + \pi/2$ (solid/dashed red and magenta lines in Figures 8(f),(i)). In general, the regularization effect of the $\mathrm{WDTV}_p^{\mathrm{sv}}$ penalties on input vectors $\boldsymbol{w}$ is stronger along the direction θ_i. Finally, the sparsity-promoting regions are elongated in the direction defined by θ_i and their elongation is negatively correlated with the value of the local anisotropy parameter a_i.

5. Geometrical Interpretation. In this section, we enrich the statistical and analytical study of the space-adaptive regularizers introduced in the previous sections by providing some insights helpful to understand their local behavior from a geometrical point of view. To do so, we unify and expand some considerations from [10, 94, 126, 59] and start recalling the dual definition of TV:

$$\mathrm{TV}(\boldsymbol{u}) = \sum_{i=1}^{N} \max_{\boldsymbol{w}_i \in \mathcal{B}_1(\mathbf{0})} \langle (\mathbf{D}\boldsymbol{u})_i, \boldsymbol{w}_i \rangle, \tag{5.1}$$

where $\mathcal{B}_1(\mathbf{0})$ denotes the two-dimensional Euclidean unit ball centered in the origin. Such constraint can be equivalently expressed by requiring $\|\boldsymbol{w}_i\|_2 \leq 1 \ \forall i = 1, \dots, N$.

Following [10], we can now replace the set $\mathcal{B}_1(\mathbf{0})$ in (5.1) with a *fixed* two-dimensional elliptical region $\mathcal{E}_{a,\theta}(\mathbf{0})$ centered in the origin and defined in terms of its orientation $\theta \in [-\pi/2, \pi/2)$ with respect to the horizontal x-axis and eccentricity $a \in (0, 1]$, that is,

$$\mathcal{E}_{a,\theta}(\mathbf{0}) := \left\{ (x_1, x_2) \in \mathbb{R}^2 : |x_1 \cos\theta + x_2 \sin\theta|^2 + \left| \frac{-x_1 \sin\theta + x_2 \cos\theta}{a} \right|^2 \leq 1 \right\}.$$

Note that as $a \to 0$, the set $\mathcal{E}_{a,\theta}(\mathbf{0})$ degenerates to the line $x_2 = \tan\theta \ x_1$.

Recalling definitions (3.15) and (3.16) of the matrices $\boldsymbol{\Lambda}_a$ and $\mathbf{R}_{-\theta}$ and denoting (formally, given the purely discrete setting we are working on) by $D_\theta u_i = (\mathbf{D}\boldsymbol{u})_i \cdot \boldsymbol{v}$ and $D_{\theta^\perp} u_i = (\mathbf{D}\boldsymbol{u})_i \cdot \boldsymbol{v}^\perp$ the directional derivatives along the direction $\boldsymbol{v} = (\cos\theta, \sin\theta)$ and its orthogonal $\boldsymbol{v}^\perp = (-\sin\theta, \cos\theta)$, we define elementwise the *directional gradient* $\widetilde{\mathbf{D}}_{a,\theta}\boldsymbol{u} \in (\mathbb{R}^2)^N$ of $\boldsymbol{u}$ as

$$\widetilde{\mathbf{D}}_{a,\theta}\boldsymbol{u} := (\boldsymbol{\Lambda}_a \mathbf{R}_{-\theta} (\mathbf{D}\boldsymbol{u})_i)_i = \begin{pmatrix} D_\theta u_i \\ a D_{\theta^\perp} u_i \end{pmatrix}_i.$$

By this definition we can thus write the *directional* formulation of TV first used in [10] and later applied in several other works (see, e.g., [157, 94, 59]) for promoting TV smoothness along $\boldsymbol{v}$. Note that the $(\mathrm{WDTV}_{\boldsymbol{p}}^{sv})$ reduces to this definition by choosing

$\alpha_i = p_i = 1$, $\theta_i = \theta \in [-\pi/2, \pi/2)$, and $a_i = a \in (0,1]$ $\forall i$. It reads

$$\mathrm{DTV}(\boldsymbol{u}) = \sum_{i=1}^{N} \|(\widetilde{\mathbf{D}}_{a,\theta}\boldsymbol{u})_i\|_2 = \sum_{i=1}^{N} \max_{\boldsymbol{w}_i \in \mathcal{E}_{a,\theta}(\mathbf{0})} \langle (\mathbf{D}\boldsymbol{u})_i, \boldsymbol{w}_i \rangle, \tag{5.2}$$

where here and in what follows we omit the explicit dependence of the regularizers on the hyperparameters to facilitate the overall readability. Note that, differently from TV, DTV is computed as the sum of maximum values of scalar products in which the dual functions $\boldsymbol{w}_i$ are forced to belong to $\mathcal{E}_{a,\theta}(\mathbf{0})$ at any point. Following [94], we can now observe that for $\boldsymbol{w}_i \in \mathcal{E}_{a,\theta}(\mathbf{0})$, we have

$$\langle (\mathbf{D}\boldsymbol{u})_i, \boldsymbol{w}_i \rangle = \langle (\mathbf{D}\boldsymbol{u})_i, \mathbf{R}_\theta \boldsymbol{\Lambda}_a \widetilde{\boldsymbol{w}}_i \rangle = \langle \boldsymbol{\Lambda}_a \mathbf{R}_{-\theta} (\mathbf{D}\boldsymbol{u})_i, \widetilde{\boldsymbol{w}}_i \rangle = \langle (\widetilde{\mathbf{D}}_{a,\theta}\boldsymbol{u})_i, \widetilde{\boldsymbol{w}}_i \rangle,$$

where $\widetilde{\boldsymbol{w}}_i \in \mathcal{B}_1(\mathbf{0})$ for all $i = 1, \dots, N$. Thus, we deduce

$$\max_{\boldsymbol{w}_i \in \mathcal{E}_{a,\theta}(\mathbf{0})} \langle (\mathbf{D}\boldsymbol{u})_i, \boldsymbol{w}_i \rangle = \max_{\widetilde{\boldsymbol{w}}_i \in \mathcal{B}_1(\mathbf{0})} \langle (\widetilde{\mathbf{D}}_{a,\theta}\boldsymbol{u})_i, \widetilde{\boldsymbol{w}}_i \rangle,$$

which can be used to show via the standard Cauchy–Schwarz inequality that at any point $i = 1, \dots, N$ the maximum is achieved by the normalized directional gradient vector, i.e., by the vector $\widetilde{\boldsymbol{w}}_i = (\widetilde{\mathbf{D}}_{a,\theta}\boldsymbol{u})_i / \|(\widetilde{\mathbf{D}}_{a,\theta}\boldsymbol{u})_i\|_2$, so that

$$\begin{aligned} \langle (\widetilde{\mathbf{D}}_{a,\theta}\boldsymbol{u})_i, \widetilde{\boldsymbol{w}}_i \rangle &= \left\langle (\widetilde{\mathbf{D}}_{a,\theta}\boldsymbol{u})_i, \frac{(\widetilde{\mathbf{D}}_{a,\theta}\boldsymbol{u})_i}{\|(\widetilde{\mathbf{D}}_{a,\theta}\boldsymbol{u})_i\|_2} \right\rangle \\ &= \|(\widetilde{\mathbf{D}}_{a,\theta}\boldsymbol{u})_i\|_2 = \|\boldsymbol{\Lambda}_a \mathbf{R}_{-\theta} (\mathbf{D}\boldsymbol{u})_i\|_2, \end{aligned}$$

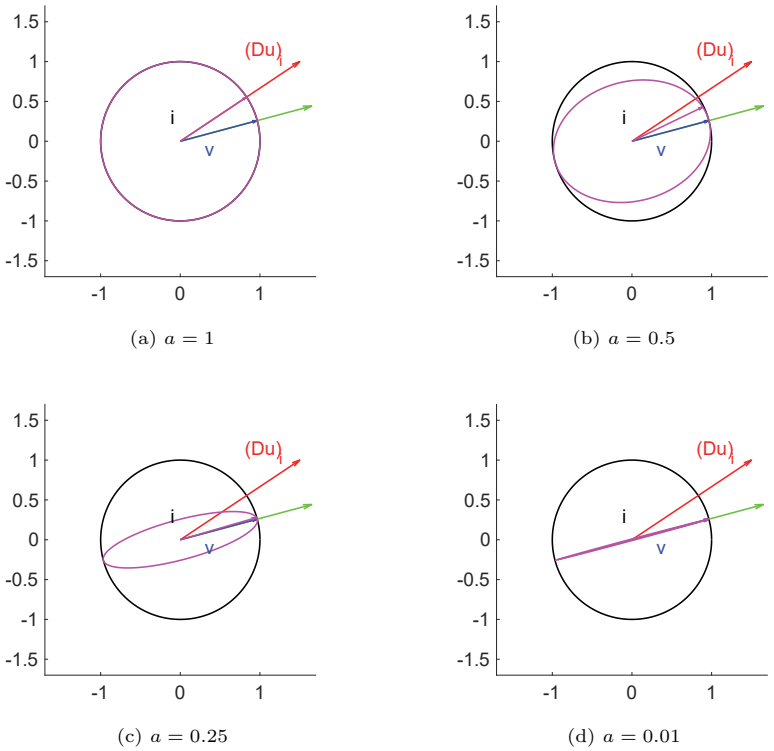
which justifies (5.2).

Inspired by [127], we report in Figure 9 a graphical representation of the considerations above. There, we denote in red a fixed nonzero gradient vector $(\mathbf{D}\boldsymbol{u})_i \in \mathbb{R}^2$ evaluated at a certain point $i = 1, \dots, N$, in blue the direction $\boldsymbol{v} = (\cos\theta, \sin\theta) \in \mathbb{R}^2$ drawing an angle θ with the x-axis, and in green the projection of $(\mathbf{D}\boldsymbol{u})_i$ along $\boldsymbol{v}$, i.e., the directional gradient $(\widetilde{\mathbf{D}}_{a,\theta}\boldsymbol{u})_i$. The unitary ball $\mathcal{B}_1(\mathbf{0})$ is colored black while the ellipses $\mathcal{E}_{a,\theta}(\mathbf{0})$ for different values of $a \in (0,1]$ are colored magenta. For each plot, the unitary vector $\widetilde{\boldsymbol{w}}_i$ realizing the maximum in (5.2) is drawn (magenta). Note that for $a = 1$ we retrieve that the vector $\widetilde{\boldsymbol{w}}_i$ maximizing the scalar product is the one parallel to $(\mathbf{D}\boldsymbol{u})_i$ (note that in such case the directionality does not affect the value computed, as $\mathbf{R}_{-\theta}$ is unitary). However, as $a \to 0$ we observe that $\widetilde{\boldsymbol{w}}_i$ progressively aligns with $\boldsymbol{v} = (\cos\theta, \sin\theta)$, thus promoting directional regularization.

As previously remarked, the unit vector $\boldsymbol{v}$ defining the orientation of the ellipse $\mathcal{E}_{a,\theta}(\mathbf{0})$ is defined in terms of the angle θ, which makes the use of the DTV regularizer useful in practice only when θ can be easily estimated. This is the case, for instance, of geometric textured images or of images of very specific scenes (see Figure 2), which limits significantly the application of DTV in practice.

Such a limitation can be overcome by considering the following natural space-variant extension of the DTV regularizer (5.2), which comes from the $(\mathrm{WDTV}^{sv}_{\boldsymbol{p}})$ regularizer with $\alpha_i = p_i = 1$ $\forall i$, and space-variant $\theta_i \in [-\pi/2, \pi/2)$ and $a_i \in (0,1]$:

$$\mathrm{DTV}^{sv}(\boldsymbol{u}) = \sum_{i=1}^{N} \|(\widetilde{\mathbf{D}}_{\boldsymbol{a},\boldsymbol{\theta}}\boldsymbol{u})_i\|_2 = \sum_{i=1}^{N} \|\boldsymbol{\Lambda}_{a_i} \mathbf{R}_{-\theta_i} (\mathbf{D}\boldsymbol{u})_i\|_2 = \sum_{i=1}^{N} \max_{\boldsymbol{w}_i \in \mathcal{E}_{a_i,\theta_i}(\mathbf{0})} \langle (\mathbf{D}\boldsymbol{u})_i, \boldsymbol{w}_i \rangle, \tag{5.3}$$



(a) $a = 1$ (b) $a = 0.5$

(c) $a = 0.25$ (d) $a = 0.01$

Fig. 9 *Directional behavior of* DTV *regularization* (5.2).

where now $\boldsymbol{a} = (a_i)_i \in (0,1]^N, \boldsymbol{\theta} = (\theta_i)_i \in [0,\pi)^N$, and where we have used the simplified notation $(\widetilde{\mathbf{D}}_{\boldsymbol{a},\boldsymbol{\theta}}\boldsymbol{u})_i = \boldsymbol{\Lambda}_{a_i}\mathbf{R}_{-\theta_i}(\mathbf{D}\boldsymbol{u})_i \in \mathbb{R}^2$. In this case, a space-variant adjustment of the directional smoothing (from strongly anisotropic along the direction $\boldsymbol{v}_i = (\cos\theta_i, \sin\theta_i)$ with $a_i = 0$ to fully isotropic with $a_i = 1$) is allowed at any point. For such a regularizer, the same geometrical considerations as before hold, the difference being that the orientations θ_j may change from one point to another. Few choices can be made here. Following the edge adaptive total variation (EATV) approach proposed in [157], one possibility consists in estimating the local directions $\boldsymbol{v}_i$ by imposing that $\boldsymbol{v}_i \perp (\mathbf{D}\boldsymbol{b}_\sigma)_i$, where for $\sigma > 0$, $\boldsymbol{b}_\sigma$ denotes a smoothed version of the given image $\boldsymbol{b}$. This choice, however, is very sensitive to noise oscillations, and it may misguide the local directional behavior if these are too large. Alternatively, as considered in [76, 69, 65, 106] and more recently in [123, 59], the dependence on the image to retrieve can be encoded explicitly in the definition of the regularization by allowing θ_i to be a function of the target image $\boldsymbol{u}$ (i.e., $\theta_i = \theta_i(\boldsymbol{u})$) using, for instance, information coming from the structure tensor. This procedure is much more robust, but the nonlinear dependence on $\boldsymbol{u}$ in the definition of θ_i may significantly complicate the problem from an optimization viewpoint. For further estimation strategies based on ML approaches, we refer the reader to [159, 98, 27] and to the following discussion in section 7.1. Whatever the approach considered, it is worth remarking that an accurate and robust estimation of the space-variant parameters $\boldsymbol{a}$ and $\boldsymbol{\theta}$ is a very challenging problem.

Remark 5.1. The values $a_i \in (0,1]$ for all $i = 1,\dots,N$ have to be interpreted as "confidence" parameters enforcing a strong anisotropic TV smoothing ($a_i \approx 0$) whenever a good local estimation of θ_i is available, while leaving the behavior to be close-to-isotropic ($a_i \approx 1$) whenever the estimation of θ_i is unreliable.

As noted in sections 3 and 4, we can further incorporate in (5.3) an additional shape/sharpness parameter vector $\boldsymbol{p} = (p_i)_i \in \mathbb{R}^N_{++}$, thus considering the regularizer

$$\text{DTV}^{sv}_{\boldsymbol{p}}(\boldsymbol{u}) = \sum_{i=1}^{N} \|\boldsymbol{\Lambda}_{a_i}\mathbf{R}_{-\theta_i}(\mathbf{D}\boldsymbol{u})_i\|_2^{p_i} = \sum_{i=1}^{N} \left(\max_{\boldsymbol{w}_i \in \mathcal{E}_{a_i,\theta_i}(\mathbf{0})} \langle (\mathbf{D}\boldsymbol{u})_i, \boldsymbol{w}_i \rangle \right)^{p_i}, \tag{5.4}$$

which is a particular instance of ($\text{WDTV}^{sv}_{\boldsymbol{p}}$) taking $\alpha_i = 1\ \forall i$. The presence of the parameters p_i in (5.4) does not alter the directional behavior of such a regularizer in comparison with the one observed for DTV^{sv}. However, as thoroughly discussed in sections 4.1.2 and 4.1.4, such behavior is made sharper for $p_i < 1$ and smoother for $p_i > 1$. Note, in particular, that when $p_i = 2$ the DTV^{sv}_2 regularizer acts locally as a Tikhonov-type squared ℓ_2-norm of the directional gradient $\widetilde{\mathbf{D}}_{a_i,\theta_i}\boldsymbol{u}$.

We conclude this section with some considerations regarding weighted models. Recalling (5.1), we notice that introducing a space-variant parameter vector $(\alpha_i)_i \in \mathbb{R}^N_{++}$ corresponds simply to inflating/deflating the Euclidean ball $\mathcal{B}_1(\mathbf{0})$ and to looking for maxima therein, which corresponds to the choice

$$\text{WTV}(\boldsymbol{u}) = \sum_{i=1}^{N} \alpha_i \|(\mathbf{D}\boldsymbol{u})_i\|_2 = \sum_{i=1}^{N} \max_{\boldsymbol{w}_i \in \mathcal{B}_{\alpha_i}(\mathbf{0})} \langle (\mathbf{D}\boldsymbol{u})_i, \boldsymbol{w}_i \rangle,$$

where for $\alpha_i > 0$, $\mathcal{B}_{\alpha_i}(\mathbf{0}) := \{\boldsymbol{z} \in \mathbb{R}^2 : \|\boldsymbol{z}\|_2 \le \alpha_i\}$ and where the vector $\widetilde{\boldsymbol{w}}_i = \alpha_i \frac{(\mathbf{D}\boldsymbol{u})_i}{\|(\mathbf{D}\boldsymbol{u})_i\|_2}$ maximizes the scalar products at any point. By analogous considerations as above, we can finally draw a connection with the regularizer defined in (4.9), which, recalling the discussion above, can be written as

$$\text{WDTV}^{sv}_{\boldsymbol{p}}(\boldsymbol{u}) = \sum_{i=1}^{N} \alpha_i^{p_i} \|(\widetilde{\mathbf{D}}_{\boldsymbol{a},\boldsymbol{\theta}}\boldsymbol{u})_i\|_2^{p_i} = \sum_{i=1}^{N} \alpha_i^{p_i} \left(\max_{\boldsymbol{w}_i \in \mathcal{E}_{a_i,\theta_i}(\mathbf{0})} \langle (\mathbf{D}\boldsymbol{u})_i, \boldsymbol{w}_i \rangle \right)^{p_i}. \tag{5.5}$$

Remark 5.2. If $p_i = 1$ for all $i = 1,\dots,N$, the regularizer (5.5) takes the form

$$\text{WDTV}^{sv}(\boldsymbol{u}) = \sum_{i=1}^{N} \max_{\boldsymbol{w}_i \in \mathcal{E}_{a_i,\theta_i,\alpha_i}(\mathbf{0})} \langle (\mathbf{D}\boldsymbol{u})_i, \boldsymbol{w}_i \rangle,$$

where $\mathcal{E}_{a_i,\theta_i,\alpha_i}(\mathbf{0})$ denotes the 2D ellipse centered in the origin with eccentricity $a_i \in (0,1]$, orientation $\theta_i \in [-\pi/2, \pi/2)$ with respect to the x-axis, and width/height equal to $2\alpha_i$ and $2\alpha_i a_i$, respectively.

The authors believe that an interesting generalization of the discussion above shall address situations where the constraint set is nonconvex and, in particular, it is defined in terms of *Lamé curves* centered in $\mathbf{0}$, i.e., defined by

$$\mathcal{L}_{\beta,a,\theta}(\mathbf{0}) := \left\{ (x_1,x_2) \in \mathbb{R}^2 : \ |x_1\cos\theta + x_2\sin\theta|^\beta + \left| \frac{-x_1\sin\theta + x_2\cos\theta}{a} \right|^\beta \le 1 \right\}. \tag{5.6}$$

Such a set is nonconvex as soon as $\beta < 1$; see Figure 10. The use of such general shapes may lead one to consider new gradient-based regularizations where the underlying geometry constraining the dual functions favors smoothing in different ways.

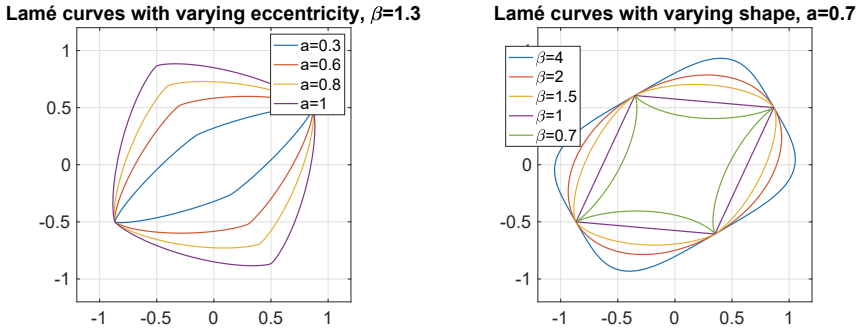


Fig. 10 *Lamé curves* (5.6) *with varying eccentricity and shape parameters,* $\theta = \pi/6$.

Finally, we remark on an analogous study on the geometrical interpretation of the so-called *anisotropic* formulation of TV regularization, where ℓ_1-norms of local gradients evaluated in place of the Euclidean ones in (TV) could also be of interest. For this choice, constraint sets there are expected to assume rhomboidal shape; see, e.g., [28].

6. Joint Hypermodeling. In this section, we provide explicit expressions of the negative log-hyperprior $-\ln \pi(\boldsymbol{\Theta})$ and of the negative log-likelihood $-\ln \pi(\boldsymbol{b} \mid \mathbf{A}\boldsymbol{u})$ in (3.22), which allows us to derive the final joint hypermodel.

6.1. Noninformative Hyperprior. When no a priori knowledge or intuition about the value of the unknown prior hyperparameters is available, a uniform distribution for the random vector $\boldsymbol{\Theta}$ can be set, thus considering a (possibly improper) noninformative hyperprior. In formulas this corresponds to setting

$$\pi(\boldsymbol{\Theta}) = \varrho\, \chi_{\mathcal{D}_{\boldsymbol{\Theta}}}(\boldsymbol{\Theta})\,, \text{ with } \varrho \in \mathbb{R}_{++}\,, \tag{6.1}$$

from which it follows that

$$-\ln \pi(\boldsymbol{\Theta}) = -\ln \varrho + \iota_{\mathcal{D}_{\boldsymbol{\Theta}}}(\boldsymbol{\Theta})\,.$$

6.2. GG Likelihood Leads to $\mathbf{L}_q$ Fidelity Term. First, based on the expression of the considered GG likelihoods in (3.1)–(3.2), the negative log-likelihood term $-\ln \pi(\boldsymbol{b} \mid \mathbf{A}\boldsymbol{u})$ in (3.21) takes the form

$$q < +\infty : \begin{cases} -\ln \pi(\boldsymbol{b} \mid \mathbf{A}\boldsymbol{u}) &= \omega^q \, \|\mathbf{A}\boldsymbol{u} - \boldsymbol{b}\|_q^q \, - M \ln\left(\dfrac{\omega}{2}\dfrac{q}{\Gamma(1/q)}\right), \\ &= \omega^q \, \|\mathbf{A}\boldsymbol{u} - \boldsymbol{b}\|_q^q \, + C_q\,, \end{cases} \tag{6.2}$$

$$q = +\infty : \begin{cases} -\ln \pi(\boldsymbol{b} \mid \mathbf{A}\boldsymbol{u}) &= \iota_{[0,1/\omega]}\left(\|\mathbf{A}\boldsymbol{u} - \boldsymbol{b}\|_\infty\right) - M \ln \dfrac{\omega}{2}\,, \\ &= \iota_{[0,1/\omega]}\left(\|\mathbf{A}\boldsymbol{u} - \boldsymbol{b}\|_\infty\right) \, + C_\infty\,, \end{cases} \tag{6.3}$$

where the quantities $C_q, C_\infty > 0$ appearing in (6.2)–(6.3) do not depend on the optimization variable $\boldsymbol{u}$, so they can be dropped in (3.21).

We now introduce the functional $F_q(\mathbf{A}\boldsymbol{u};\boldsymbol{b}) : \mathbb{R}^M \to \mathbb{R}_+$, which is defined as

$$q < +\infty : \qquad F_q(\mathbf{A}\boldsymbol{u};\boldsymbol{b}) := q\,\omega^q \mathrm{L}_q(\mathbf{A}\boldsymbol{u};\boldsymbol{b}), \tag{6.4}$$

$$q = +\infty : \qquad F_\infty(\mathbf{A}\boldsymbol{u};\boldsymbol{b}) := \iota_{[0,1/\omega]}(\|\mathbf{A}\boldsymbol{u}-\boldsymbol{b}\|_\infty), \tag{6.5}$$

with

$$\mathrm{L}_q(\mathbf{A}\boldsymbol{u};\boldsymbol{b}) \;=\; \frac{1}{q}\|\mathbf{A}\boldsymbol{u}-\boldsymbol{b}\|_q^q\,.$$

6.3. Joint Variational Bayesian Hypermodels. We are now ready to derive the explicit instances of the hypermodel (3.22) in terms of the selected priors, hyperpriors, and fidelity functionals discussed above. To improve readability, we will consider in the following data terms F_q with $q < +\infty$. However, as will be remarked upon at the end of the section, analogous derivations can be easily extended to the case $q = +\infty$.

Problem (3.22) can be reformulated as the following optimization problem:

$$\{\boldsymbol{u}^*, \boldsymbol{\Theta}^*\} \;\in\; \underset{\boldsymbol{u}\in\mathbb{R}^N, \boldsymbol{\Theta}\in\mathcal{D}_{\boldsymbol{\Theta}}}{\arg\min} \; \{\, \mathcal{R}(\boldsymbol{u},\boldsymbol{\Theta}) + \mathcal{H}(\boldsymbol{\Theta}) + \mu\,\mathrm{L}_q(\mathbf{A}\boldsymbol{u};\boldsymbol{b}) \,\}\,, \tag{6.6}$$

where $\mathcal{R}(\boldsymbol{u},\boldsymbol{\Theta})$ denotes one of the regularization terms previously discussed in section 4.1, while $\mathcal{H}(\boldsymbol{\Theta})$ accounts for possibly multiple terms depending only on the hyperparameter vector $\boldsymbol{\Theta} \in \mathcal{D}_{\boldsymbol{\Theta}}$. The uniform hyperprior (6.1) acts here simply by enforcing optimization on $\mathcal{D}_{\boldsymbol{\Theta}}$ only. The parameter $\mu > 0$ is a regularization parameter whose choice will be specified for each hypermodel in the following. Note that due to presence of the coupling term $\mathcal{R}(\boldsymbol{u},\boldsymbol{\Theta})$, there is no guarantee of joint convexity for problem (6.6) even in the case of convexity with respect to each minimization variable. This poses several questions related to the theoretical analysis as well as practical optimization of this problem (such as, e.g., choice of the initialization, parameter sensitivity, etc.), which, as we will see in the following, introduce further challenges to the development of a sound and reliable model.

Similarly to what was discussed in section 4.1, we have that $\mathcal{H}(\boldsymbol{\Theta})$ can be expressed in general form as

$$\mathcal{H}(\boldsymbol{\Theta}) = \sum_{i=1}^{N} h(\boldsymbol{\Theta}_i)\,, \quad h : \mathcal{D}_{\boldsymbol{\Theta}_i} \to \mathbb{R}\,,$$

where the function h is a *parameter penalty function* whose form will be specified for each regularizer.

We start our considerations from the TV prior (3.8). By plugging (6.1) and (3.1) into (3.22), we get

$$\begin{aligned}\{\boldsymbol{u}^*, \alpha^*\} \;&\in\; \underset{\boldsymbol{u}\in\mathbb{R}^N,\,\alpha\in\mathbb{R}_{++}}{\arg\min} \; \left\{ \alpha \sum_{i=1}^{N} \|(\mathbf{D}\boldsymbol{u})_i\|_2 - N\ln\alpha + q\,\omega^q\,\mathrm{L}_q(\mathbf{A}\boldsymbol{u};\boldsymbol{b}) \right\} \\ &= \underset{\boldsymbol{u}\in\mathbb{R}^N,\,\alpha\in\mathbb{R}_{++}}{\arg\min} \; \left\{ \mathrm{TV}(\boldsymbol{u}) - \frac{N}{\alpha}\ln\alpha + \mu\,\mathrm{L}_q(\mathbf{A}\boldsymbol{u};\boldsymbol{b}) \right\}, \quad \text{with} \quad \mu := \frac{q\,\omega^q}{\alpha}\,, \end{aligned} \tag{6.7}$$

where we recall that in this case $\boldsymbol{\Theta} = \alpha \in \mathbb{R}_{++}$, hence the parameter penalty function h_{TV} reads

$$h_{\mathrm{TV}}(\alpha) = -\frac{1}{\alpha}\ln\alpha\,, \quad i = 1,\dots,N\,. \tag{6.8}$$

For the TV_p and the DTV regularizers in (TV_p) and (DTV), respectively, hypermodels with similar form as in (6.7) can be derived. In particular, as far as the TV_p regularization term is concerned, we have that $\boldsymbol{\Theta} = (\alpha, p) \in \mathbb{R}^2_{++}$ and that the parameter penalty function takes the form

$$h_{\mathrm{TV}_p}(\alpha, p) = -\frac{1}{\alpha^p} \ln \frac{\alpha\, p}{\Gamma(1/p)}\,, \quad i = 1, \dots, N\,,$$

while for the DTV_p regularizer we have $\boldsymbol{\Theta} = (\alpha, p, \theta, a) \in \mathbb{R}^2_{++} \times [-\pi/2, \pi/2) \times (0, 1]$ and the parameter penalty function reads

$$h_{\mathrm{DTV}_p}(\alpha, \theta, a) = -\frac{1}{\alpha} \ln \left(\frac{a}{2\pi} \frac{\alpha^2}{4} \right), \quad i = 1, \dots, N\,.$$

As far as space-variant hypermodels are concerned, we start considering the WTV regularizer for which $\boldsymbol{\Theta} = \boldsymbol{\alpha} \in \mathbb{R}^N_+$. Model (6.6) thus turns into

$$\{\boldsymbol{u}^*, \boldsymbol{\alpha}^*\} \in \underset{\boldsymbol{u} \in \mathbb{R}^N,\, \boldsymbol{\alpha} \in \mathbb{R}^N_+}{\arg\min} \left\{ \sum_{i=1}^N \alpha_i \, \|(\mathbf{D}\boldsymbol{u})_i\|_2 - \sum_{i=1}^N \ln \alpha_i + \mu \, \mathrm{L}_q(\mathbf{A}\boldsymbol{u}; \boldsymbol{b}) \right\} \quad \text{with} \quad \mu := q\, \omega^q\,,$$

where the function h_{WTV} is defined by

$$h_{\mathrm{WTV}}(\alpha_i) = -\ln \alpha_i\,, \quad i = 1, \dots, N\,.$$

For the $\mathrm{WTV}^{sv}_{\boldsymbol{p}}$, we have that $\boldsymbol{\Theta} = (\boldsymbol{\alpha}, \boldsymbol{p}) \in \mathbb{R}^N_{++} \times \mathbb{R}^N_{++}$. The hypermodel (6.6) here specifies into

$$\{\boldsymbol{u}^*, \boldsymbol{\alpha}^*, \boldsymbol{p}^*\} \in \underset{\boldsymbol{u} \in \mathbb{R}^N,\, \boldsymbol{\alpha} \in \mathbb{R}^N_{++},\, \boldsymbol{p} \in \mathbb{R}^N_{++}}{\arg\min} \left\{ \sum_{i=1}^N \alpha_i^{p_i} \, \|(\mathbf{D}\boldsymbol{u})_i\|_2^{p_i} - \sum_{i=1}^N \ln \frac{\alpha_i\, p_i}{\Gamma(1\,/\,p_i)} + \mu \, \mathrm{L}_q(\mathbf{A}\boldsymbol{u}; \boldsymbol{b}) \right\} \quad \text{with} \quad \mu := q\, \omega^q\,,$$

with penalty function $h_{\mathrm{WTV}^{sv}_{\boldsymbol{p}}}$ defined by

$$h_{\mathrm{WTV}^{sv}_{\boldsymbol{p}}}(\alpha_i, p_i) = -\ln \frac{\alpha_i\, p_i}{\Gamma(1/p_i)}\,, \quad i = 1, \dots, N\,. \tag{6.9}$$

Finally, for $\mathrm{WDTV}^{sv}_{\boldsymbol{p}}$, we have $\boldsymbol{\Theta} = (\boldsymbol{\alpha}, \boldsymbol{p}, \boldsymbol{\theta}, \boldsymbol{a}) \in \mathbb{R}^N_{++} \times \mathbb{R}^N_{++} \times [-\pi/2, \pi/2)^N \times (0, 1]^N$, and hence the final hypermodel reads

$$\begin{aligned} &\{\boldsymbol{u}^*, \boldsymbol{\alpha}^*, \boldsymbol{p}^*, \boldsymbol{\theta}^*, \boldsymbol{a}^*\} \\ &\in \underset{\boldsymbol{u} \in \mathbb{R}^N, \boldsymbol{\alpha} \in \mathbb{R}^N_+,\, \boldsymbol{p} \in \mathbb{R}^N_{++},\, \boldsymbol{\theta} \in [-\pi/2, \pi/2)^N,\, \boldsymbol{a} \in (0,1]^N}{\arg\min} \left\{ \sum_{i=1}^N \alpha_i^{p_i} \, \|\boldsymbol{\Lambda}_{a_i} \mathbb{R}_{-\theta_i} (\mathbf{D}\boldsymbol{u})_i\|_2^{p_i} \right. \\ &\left. - \sum_{i=1}^N \ln \left(\frac{a_i}{2\pi} \frac{p_i \alpha_i^2}{\Gamma(2\,/\,p_i) 2^{2/p_i}} \right) + \mu \, \mathrm{L}_q(\mathbf{A}\boldsymbol{u}; \boldsymbol{b}) \right\}, \quad \text{with} \quad \mu := q\, \omega^q\,, \end{aligned}$$

with parameter penalty function taking the form

$$h_{\mathrm{WDTV}^{sv}_{\boldsymbol{p}}}(\alpha_i, p_i, \theta_i, a_i) = -\ln \left(\frac{a_i}{2\pi} \frac{p_i\, \alpha_i^2}{\Gamma(2/p_i) 2^{2/p_i}} \right), \quad i = 1, \dots, N\,. \tag{6.10}$$

Table 2 *Gradient penalty function f, parameter penalty function h, parameters domain $\mathcal{D}_{\boldsymbol{\Theta}_i}$, and regularization parameter μ for the space-invariant and space-variant regularization hypermodels considered with L_q data fidelity, $q < +\infty$.*

	$\mathcal{R}(\boldsymbol{u},\boldsymbol{\Theta})$	$f(\boldsymbol{g}_i;\boldsymbol{\Theta}_i)$	$h(\boldsymbol{\Theta}_i)$	$\mathcal{D}_{\boldsymbol{\Theta}_i}$	μ	Ref.
space-invariant	TV	$\|\boldsymbol{g}_i\|_2$	$-\frac{1}{\alpha}\ln\alpha$	$\alpha \in \mathbb{R}_{++}$	$\frac{q\,\omega^q}{\alpha}$	[134]
	TV_p	$\|\boldsymbol{g}_i\|_2^p$	$-\frac{1}{\alpha^p}\ln\frac{\alpha\,p}{\Gamma(1/p)}$	$(\alpha,p) \in \mathbb{R}^2_{++}$	$\frac{q\,\omega^q}{\alpha^p}$	[101]
	DTV	$\|\boldsymbol{\Lambda}_a\mathbf{R}_{-\theta}\boldsymbol{g}_i\|_2$	$-\frac{1}{\alpha}\ln\left(\frac{a}{2\pi}\frac{\alpha^2}{4}\right)$	$(\alpha,\theta,a) \in \mathbb{R}_{++}\times[-\pi/2,\pi/2)\times(0,1]$	$\frac{q\,\omega^q}{\alpha}$	[10]
space-variant	WTV	$\alpha_i\|\boldsymbol{g}_i\|_2$	$-\ln\alpha_i$	$\alpha_i \in \mathbb{R}_{++}$	$q\,\omega^q$	[29]
	$\mathrm{WTV}^{sv}_{\boldsymbol{p}}$	$\alpha_i^{p_i}\|\boldsymbol{g}_i\|_2^{p_i}$	$-\ln\frac{\alpha_i\,p_i}{\Gamma(1/p_i)}$	$(\alpha_i,p_i) \in \mathbb{R}^2_{++}$	$q\,\omega^q$	[98]
	$\mathrm{WDTV}^{sv}_{\boldsymbol{p}}$	$\alpha_i^{p_i}\|\boldsymbol{\Lambda}_{a_i}\mathbf{R}_{-\theta_i}\boldsymbol{g}_i\|_2^{p_i}$	$-\ln\left(\frac{a_i}{2\pi}\frac{p_i\,\alpha_i^2}{\Gamma(2/p_i)\,2^{2/p_i}}\right)$	$(\alpha_i,p_i,\theta_i,a_i) \in \mathbb{R}^2_{++}\times[-\pi/2,\pi/2)\times(0,1]$	$q\,\omega^q$	[27]

For all considered space-invariant and space-variant hypermodels, we summarize in Table 2 the gradient and parameter penalty functions f and h, respectively, as well as the parameters domains $\mathcal{D}_{\boldsymbol{\Theta}_i}$ and the regularization parameters μ. In the table, we also report the reference papers in which the aforementioned regularizers were first introduced and/or analyzed in probabilistic terms.

Remark 6.1 (additive i.i.d. uniform noise). When the corrupting noise is AIU, i.e., $q = +\infty$ and the data term is written as in (6.5), the regularization parameter μ does not appear explicitly in the final hypermodels. However, one can clearly observe that functions f, h, and the parameter domain $\mathcal{D}_{\boldsymbol{\Theta}_i}$ have the same expressions as the ones listed in Table 2.

Remark 6.2. As far as the value of the regularization parameter μ is concerned, we remark that when both the scale parameter ω and the shape parameter q of the AIGG noise distribution are assumed to be known, the parameter $\mu > 0$ in all models above is also known. However, it is quite well known that setting a priori the μ value based on the true noise parameters does not guarantee that the empirical noise level calculated starting from the output residual image $\boldsymbol{y}^*(\mu) = \mathbf{A}\boldsymbol{u}^*(\mu) - \boldsymbol{b}$ coincides with the true underlying noise level. Hence, μ will be regarded in the following as a further unknown parameter to be estimated based on the GDP strategy presented in section 2.1 and further detailed in section 8.

Recalling now the alternated IAS scheme written in (3.23)–(3.24) in a statistical form, we have that the general joint minimization problem (6.6) can now be similarly solved by means of an alternated minimization scheme whose update for $k \geq 0$ and $\boldsymbol{u}^{(0)}$ reads

$$\boldsymbol{\Theta}^{(k+1)} \in \underset{\boldsymbol{\Theta}\in\mathcal{D}_{\boldsymbol{\Theta}}}{\arg\min} \left\{\mathcal{R}(\boldsymbol{u}^{(k)},\boldsymbol{\Theta}) + \mathcal{H}(\boldsymbol{\Theta})\right\}, \tag{6.11}$$

$$\boldsymbol{u}^{(k+1)} \in \underset{\boldsymbol{u}\in\mathbb{R}^N}{\arg\min} \left\{\mathcal{R}(\boldsymbol{u},\boldsymbol{\Theta}^{(k+1)}) + \mu\,\mathrm{L}_q(\mathbf{A}\boldsymbol{u};\boldsymbol{b})\right\}. \tag{6.12}$$

Note that due to the possible joint and blockwise nonconvexity of the problems involved, providing sufficient conditions for guaranteeing the convergence of the iterative scheme (6.11)–(6.12) is a challenging problem. For what follows, we interpret the scheme above as an alternate parameter-estimation/optimization scheme where the estimation of model hyperparameters and their use in the underlying image restoration model considered are combined in an alternate fashion.

7. Coupling Image Statistics with Variational Modeling: Parameter Selection. In this section, we address the estimation of the parameters $\boldsymbol{\Theta}$ for the different instances of space-variant models introduced below. As pointed out in section 4, a key step considered in the following for tackling the $\boldsymbol{\Theta}$-update step in the alternating scheme (6.11)–(6.12) consists in neglecting the normalization constant $c(\boldsymbol{\Theta})$. Although this approximation causes of course a lack of consistency with the original model, the estimation results reported in this section will support the rationale of our choice. An extensive analysis of the good statistical properties of the estimator considered in what follows has been provided in [27, section 7]. There, the authors provided empirical evidence of the fact that the considered estimator is unbiased, by showing that the empirical variance and root mean square error decay to zero.

7.1. Inspecting Space Variance. According to (6.11), the general form of the $\boldsymbol{\Theta}$-update, under the adoption of noninformative hyperprior, reads

$$\begin{aligned}\boldsymbol{\Theta}^{(k+1)} &\in \underset{\boldsymbol{\Theta}\in\mathcal{D}_{\boldsymbol{\Theta}}}{\arg\min} \left\{ -\ln \pi(\boldsymbol{z}(\boldsymbol{u}^{(k)}) \mid \boldsymbol{\Theta}) \right\} \\ &= \underset{\boldsymbol{\Theta}\in\mathcal{D}_{\boldsymbol{\Theta}}}{\arg\min} \left\{ \sum_{i=1}^{N} \left(f((\mathbf{D}\boldsymbol{u}^{(\boldsymbol{k})})_i; \boldsymbol{\Theta}_i) + h(\boldsymbol{\Theta}_i) \right) \right\}, \end{aligned} \tag{7.1}$$

where f and h denote the general gradient and parameter penalty functions, respectively, summarized in Table 2 for the hypermodels of interest. In light of the separability induced by the summation, problem (7.1) can be addressed by solving N minimization problems of the form

$$\boldsymbol{\Theta}_{\boldsymbol{i}}^{(k+1)} \in \underset{\boldsymbol{\Theta}_{\boldsymbol{i}}\in\mathcal{D}_{\boldsymbol{\Theta}_i}}{\arg\min} \left\{ f((\mathbf{D}\boldsymbol{u}^{(k)})_i; \boldsymbol{\Theta}_{\boldsymbol{i}}) + h(\boldsymbol{\Theta}_i) \right\}, \quad i = 1 \ldots, N, \tag{7.2}$$

where, notice, the information on $\boldsymbol{u}^{(k)}$ required to perform the update of $\boldsymbol{\Theta}_i$ is synthesized in the sole value $(\mathbf{D}\boldsymbol{u}^{(k)})_i$. However, if $(\mathbf{D}\boldsymbol{u}^{(k)})_i$ is a highly blur- and noise-damaged version of the true underlying value $(\mathbf{D}\boldsymbol{u})_i$ (which is the case along the ADMM iterations), the local estimate is expected to be particularly poor and unreliable.

As a way to overcome this limitation, the estimation problem (7.2) can be recast so as to take into account the information encoded in a set of pixels close to pixel i. More specifically, for any $i = 1, \ldots, N$, we consider the square neighborhood $\mathcal{J}_i^r$ centered at pixel i with side $2r + 1$ pixels and dimension $card(\mathcal{J}_i^r) = (2r + 1)^2 =: m$ and compute the discrete gradients points in $\mathcal{J}_i^r$. These quantities will be then used for the estimation of the ith unknown parameter $\boldsymbol{\Theta}_i$. We will refer to r as the *radius* of the neighborhoods; as will be observed in section 9, the choice of r can affect the quality of the final restorations. This issue will be addressed explicitly in section 10.

Statistically, the selected strategy relies on the assumption that in each of the considered neighborhoods the gradients (or their magnitudes) are independently sampled from the same distribution.

We thus introduce the following sets of samples drawn around i for $i = 1, \dots, N$:

$$\mathcal{S}_i := \begin{cases} \left\{ \|(\mathbf{D}\boldsymbol{u}^{(k)})_j\|_2 \,:\, j \in \mathcal{J}_i^r \right\} & \text{for the WTV and WTV}_{\boldsymbol{p}}^{sv} \text{ regularizers,} \\ \left\{ (\mathbf{D}\boldsymbol{u}^{(k)})_j \,:\, j \in \mathcal{J}_i^r \right\} & \text{for the WDTV}_{\boldsymbol{p}}^{sv} \text{ regularizer.} \end{cases}$$

For each $i = 1, \dots, N$, by exploiting the mutual independence of the gradients, problem (7.2) can thus be formulated as follows:

$$\begin{aligned} \boldsymbol{\Theta}_i^{(k+1)} &\in \underset{\boldsymbol{\Theta}_i \in \mathcal{D}_{\boldsymbol{\Theta}_i}}{\arg\min} \left\{ -\ln \prod_{j \in \mathcal{J}_i^r} \pi(\mathcal{S}_i \mid \boldsymbol{\Theta}_i) = -\ln \pi((\mathbf{D}\boldsymbol{u}^{(k)})_j; \boldsymbol{\Theta}_i) \right\} \\ &= \underset{\boldsymbol{\Theta}_i \in \mathcal{D}_{\boldsymbol{\Theta}_i}}{\arg\min} \left\{ \sum_{j \in \mathcal{J}_i^r} \Big(f((\mathbf{D}\boldsymbol{u}^{(k)})_j; \boldsymbol{\Theta}_i) + h(\boldsymbol{\Theta}_i) \Big) \right\}. \end{aligned} \tag{7.3}$$

We now specify the formulation of the minimization problem (7.3) in correspondence with the regularization terms considered in this review. For the sake of better readability, in this section the outer iteration superscript k will be neglected and the discrete gradient $(\mathbf{D}\boldsymbol{u})_j$ at pixel j will be simply denoted by $\boldsymbol{g}_j$.

7.2. Parameter Estimation for the WTV Regularizer. We start considering WTV regularization. Recalling the definition of the gradient and parameter penalty functions f_{TV}, h_{TV} in (4.3), (6.8) and the hyperparameter domain $\mathcal{D}_{\boldsymbol{\Theta}_i}$ specified in Table 2, the problem of interest turns into

$$\alpha_i^* \in \underset{\alpha_i \in \mathbb{R}_{++}}{\arg\min} \left\{ \mathcal{G}(\alpha_i) := -\ln \pi(\mathcal{S}_i \mid \alpha_i) = -m \ln \alpha_i + \sum_{j \in \mathcal{J}_i^r} \alpha_i \|\boldsymbol{g}_j\|_2 \right\}. \tag{7.4}$$

The following result holds true.

PROPOSITION 7.1. *The function $\mathcal{G} : \mathbb{R}_{++} \to \mathbb{R}$ in (7.4) is smooth and convex, hence it admits a unique global minimizer.*

In particular, since $\mathcal{G}$ is differentiable on $\mathbb{R}_{++}$, the solution of the ith minimization problem (7.4) can be simply found by imposing a first-order optimality condition:

$$\mathcal{G}'(\alpha_i) = -\frac{m}{\alpha_i} + \sum_{j \in \mathcal{J}_i^r} \|\boldsymbol{g}_j\|_2 = 0\,, \text{ hence } \alpha_i^* = \left(\frac{1}{m} \sum_{j \in \mathcal{J}_i^r} \|\boldsymbol{g}_j\|_2 \right)^{-1}. \tag{7.5}$$

Notice that in order to avoid degenerate configurations arising when considering neighborhoods with null gradients, a small regularization parameter $0 < \varepsilon \ll 1$ can be added to the local mean in (7.5). The selection of pixels involved in (7.5) can be efficiently carried out based on fast 2D convolution operators (realized by a fast 2D discrete transform) of the map of gradient norms with a square $(2r+1) \times (2r+1)$ averaging kernel.

For validating the proposed parameter estimation strategy, in Figure 11 we analyze the performance of the parameter estimation strategy outlined above on selected subregions of the clean image in Figure 11(a) and on the whole image itself. We remark that in the following tests, and in the other experiments reported in this section, we perform the estimate of the unknown parameters starting from the target image for illustrative purposes. Clearly, in section 9, where we show the results of the outlined

strategy for the restoration of different test images, the parameters will be updated starting from the observed data or from the current iterate $\boldsymbol{u}^{(k)}$, as prescribed from the alternating scheme (6.11)–(6.12) whose $\boldsymbol{\Theta}$-substep relies, at each $k \geq 0$, on the knowledge of the current estimate $\boldsymbol{u}^{(k)}$. The local neighborhoods shown here consist of an almost constant red-bordered region and two textured regions—see Figure 11(d) and Figures 11(g),(j), respectively—the last two differing in terms of directional features; in fact, the magenta-bordered region presents horizontally oriented features, while the texture in the cyan-bordered neighborhood does not present a dominant directionality.

We compute the hL pdfs returning the best fitting both of the global and of the local histograms of the gradient magnitudes. More specifically, we first calculate (7.5) for the whole image, i.e., when the summation index j goes from 1 to N, that will return the global scale parameter. Then the same formula is applied when the set of samples is restricted to the gradient magnitudes of the three subregions, so as to obtain local scale parameters. The estimated parameters are reported in the caption. In Figure 11(b), and in the close-up in Figure 11(c), we show the histogram of the gradient magnitudes of the whole image. The superimposed solid green line represents the global hL distribution. The histogram of the gradient magnitudes in the selected neighborhoods together with the corresponding estimated pdfs are shown in Figures 11(e),(f) for the constant region, and in Figures 11(h),(i) and 11(k),(l) for the textured regions. In the local histograms, we also report the global pdf. The comparison immediately reveals how the space-variant approach guarantees a more accurate modeling of local features; this is also reflected in the values of the estimated global and local scale parameters, which appear to be very different from each other, except for the case of the two textured regions. In fact, as discussed before, directional dissimilarities cannot be detected when adopting a (univariate) hL prior.

In order to analyze in more detail the connection between the estimated scale parameters and the local regularization strength, in Figure 12 we show the $\boldsymbol{\alpha}$-map corresponding to different test images. We observe that the scale parameters assume higher values on smooth or piecewise-constant regions, whereas lower values are obtained in correspondence with edges and texture. In those areas, a weaker regularization is indeed preferable in order to preserve details. Note also that the $\boldsymbol{\alpha}$-maps are sensitive to the choice of radius r. When considering small values of r—see, for instance, the map on the `barbara` image with $r = 2$—possibly small artifacts due to image compression or resolution may appear. A similar effect is expected in the presence of noise. On the other hand, setting a large radius r could make some details or finer structures in the image less detectable, as in the case of the map for the `geometric` image with $r = 7$, where inner edges are not visible in the final map.

7.3. Parameter Estimation for the $\mathrm{WTV}_{\boldsymbol{p}}^{sv}$ Regularizer. We now consider the $\mathrm{WTV}_{\boldsymbol{p}}^{sv}$ regularizer. Recalling the definitions for $f_{\mathrm{WTV}_{\boldsymbol{p}}^{sv}}$ and $h_{\mathrm{WTV}_{\boldsymbol{p}}^{sv}}$ given in (4.7) and (6.9), respectively, and the hyperparameter space $\mathcal{D}_{\boldsymbol{\Theta}_i}$ in Table 2, we have that the general problem (7.3) reduces to

$$\{\alpha_i^*, p_i^*\} \in \underset{(\alpha_i,p_i)\in\mathbb{R}^2_{++}}{\arg\min} \left\{ \mathcal{G}(\alpha_i, p_i) := -\ln \pi(\mathcal{S}_i \mid \alpha_i, p_i) = -m \ln \alpha_i + m \ln \Gamma\left(1 + \frac{1}{p_i}\right) + \sum_{j\in\mathcal{J}_i^r} \alpha_i^{p_i} \|\boldsymbol{g}_j\|_2^{p_i} \right\}.$$

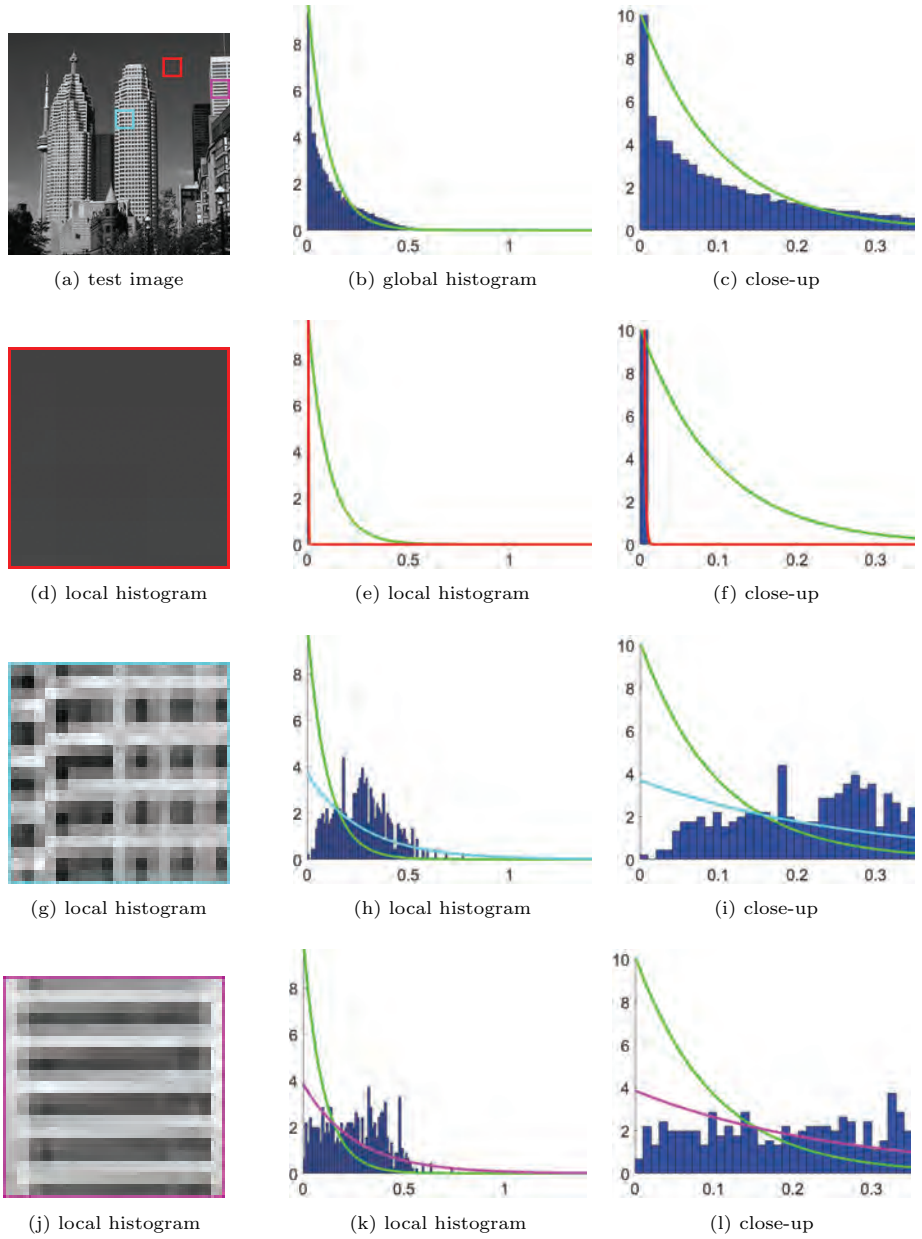


Fig. 11 Parameter estimate for WTV. *Histogram of the gradient magnitudes* (b) *on the whole test image* (a) *with the best-fitting hL pdf,* $\alpha^* = 10.17$ *(green), and corresponding close-up* (c)*; histogram of the gradient magnitudes* (e),(h),(k)*, and corresponding close-up(s)* (f),(i),(l)*, on regions* (d),(g),(j) *with the local best-fitting hL pdf, with* $\alpha^* = 200.87$ *(red),* $\alpha^* = 3.70$ *(cyan), and* $\alpha^* = 3.86$ *(magenta), respectively.*

Proceeding analogously as before, we have that by imposing a first-order optimality condition on $\mathcal{G}(\alpha_i, p_i)$ with respect to α_i, we get

$$\frac{\partial}{\partial \alpha_i}\mathcal{G}(\alpha_i, p_i) = -\frac{m}{\alpha_i} + p_i \alpha_i^{p_i - 1} \sum_{j \in \mathcal{J}_i^r} \|\boldsymbol{g}_j\|_2^{p_i} = 0\,,$$

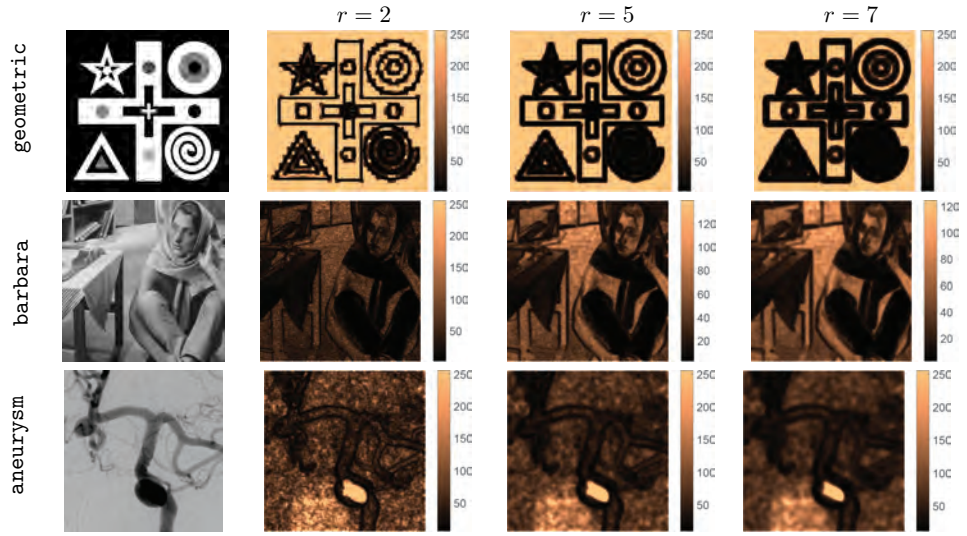


Fig. 12 *Test images with the corresponding $\boldsymbol{\alpha}$-map for different values of radius r.*

which yields the following closed-form formula for the estimation of α_i:

$$\alpha_i^*(p_i) = \left(\frac{p_i}{m} \sum_{j \in \mathcal{J}_i^r} \|\boldsymbol{g}_j\|_2^{p_i} \right)^{-\frac{1}{p_i}}. \tag{7.6}$$

It is easy to verify that the second derivative of $\mathcal{G}$ with respect to α_i computed at $\alpha_i^*(p_i)$ is strictly positive, hence the stationary point in (7.6) is a minimum. Similarly as for (7.5), also in this case a parameter $0 < \varepsilon \ll 1$ shall be added to the summation (7.6) so as to avoid degenerate configurations of gradient magnitudes. Plugging in (7.6), we have

$$p_i^* \in \underset{p_i \in \mathbb{R}_{++}}{\arg\min} \left\{ G(p_i) := \mathcal{G}(\alpha_i(p_i), p_i) = \frac{m}{p_i} \log \left(\frac{p_i}{m} \sum_{j \in \mathcal{J}_i^r} \|\boldsymbol{g}_j\|_2^{p_i} \right) + m \ln \Gamma \left(1 + \frac{1}{p_i} \right) + \frac{m}{p_i} \right\}.$$

When addressing the study of G on $\mathbb{R}_{++}$, one can immediately notice that its behavior is related to the local configurations of gradient magnitudes. As a result, drawing any conclusion on the existence of minima is in general not trivial. However, looking at the problem from a computational viewpoint, it appears reasonable to restrict the p_i feasibility set to a bounded interval $[\epsilon, R]$, with $0 < \epsilon < R$ and $R > 1$. In this case, the following result holds.

PROPOSITION 7.2. *The function $G : [\epsilon, R] \to \mathbb{R}$ defined in* (7.3) *is continuous, hence it admits a minimum in its compact domain.*

In Figure 13, the estimation of the global and local shape parameters for the $\mathrm{WTV}_{\boldsymbol{p}}^{sv}$ regularizer is performed by setting $\epsilon = 0.1$ and $R = 10$. The dashed green

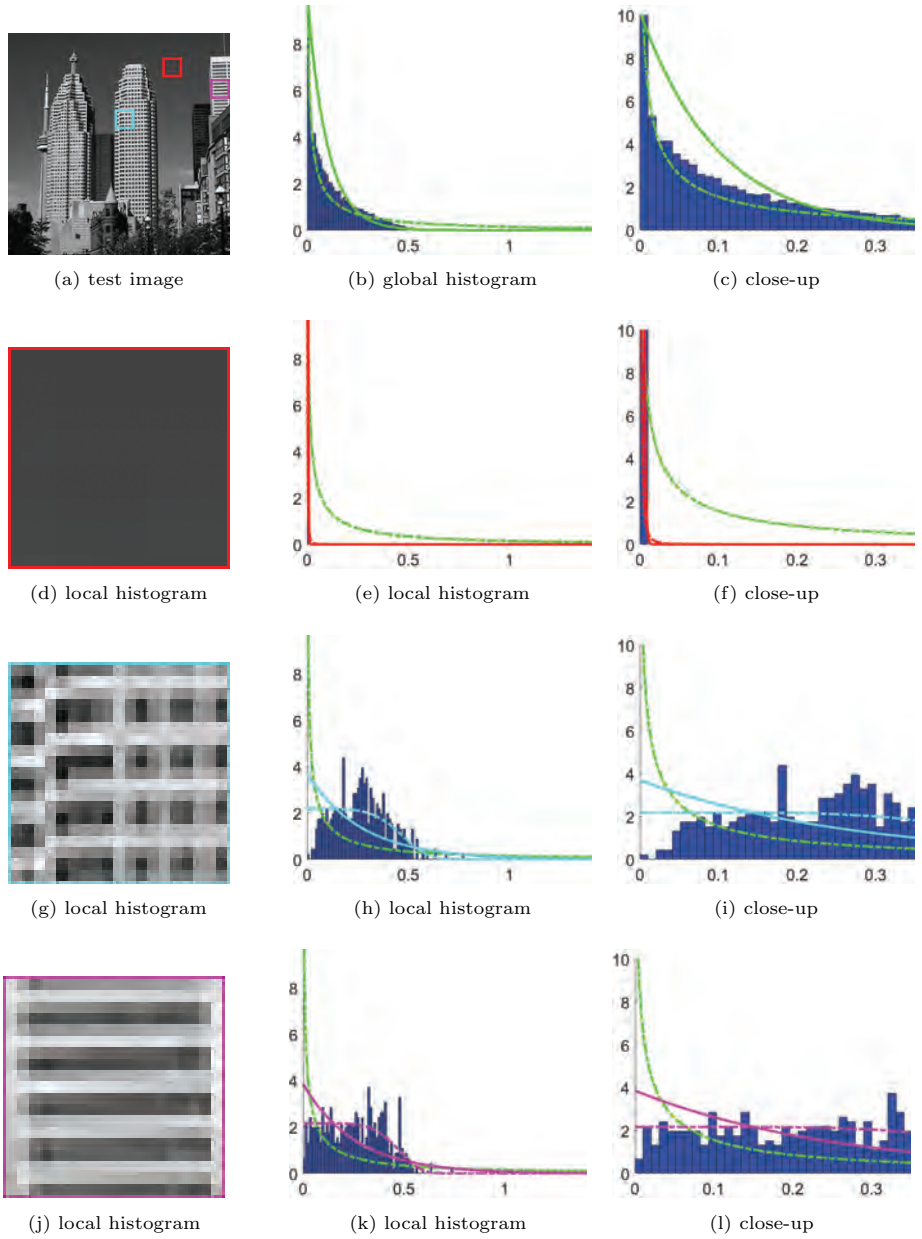


Fig. 13 Parameter estimate for $\mathrm{WTV}_{\boldsymbol{p}}^{sv}$. *Histogram of the gradient magnitudes* (b) *on the whole test image* (a) *with the best-fitting hL pdf (solid green) and hGG pdf (dashed green), with* $p^* = 0.2, \alpha^* = 465.67$*, and corresponding close-up* (c)*; histogram of the gradient magnitudes* (e),(h),(k)*, and corresponding close-up(s), on regions* (d),(g),(j) *with the local best-fitting hL pdf (solid) and hGG pdf (dashed), with* $p^* = 0.1, \alpha^* = 687.34$ *(red),* $p^* = 4.85, \alpha^* = 2.56$ *(cyan), and* $p^* = 5.65, \alpha^* = 2.78$ *(magenta).*

line in Figures 13(b),(c) represents the hGG pdf that best fits the global histogram of the gradient magnitudes where parameters have been estimated as above. One can already observe how the introduction of a further global parameter allows for a better

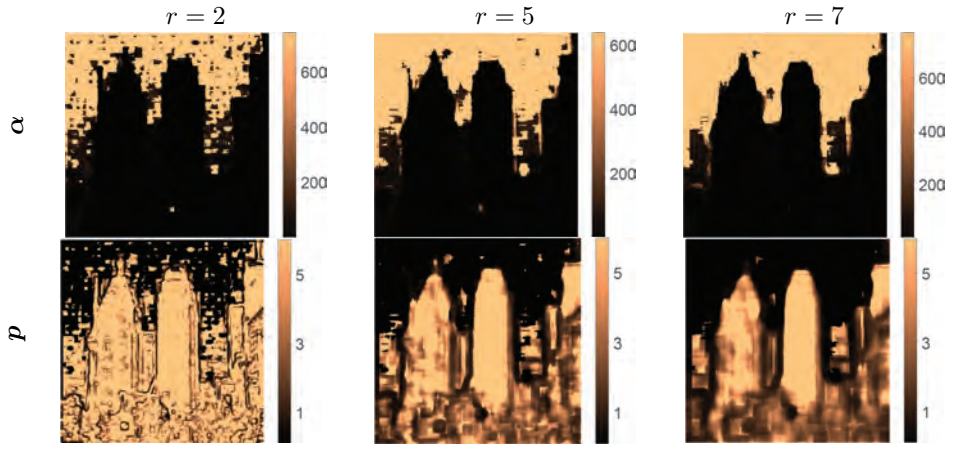


Fig. 14 *The $\boldsymbol{\alpha}$ and $\boldsymbol{p}$ maps for different values of radius r on the image* `skyscraper` *in Figure* 11(a).

modeling of the global histogram when compared to the solid green line, representing the hL pdf shown in Figure 11. In Figures 13(e),(h),(k), we report colored dashed lines corresponding to the estimated local hGG pdfs; in addition, we superimpose the global hGG pdf together with the local hL pdfs plotted in Figure 11 as solid lines. To facilitate the inspection, we also show close-up(s) of the local histograms in Figures 13(f),(g),(l). The benefits associated with the use of a second space-variant parameter are here even more significant. The differences between the selected patches, and between the patches and the global image, are accurately highlighted by the estimated global and local parameters reported in the caption. Note, however, that also in this case the selected hGG prior is not capable of detecting directional differences between the two textured subregions, due once again to its univariate behavior.

Finally, in Figure 14 we show the $\boldsymbol{\alpha}$- and $\boldsymbol{p}$-maps obtained by considering neighborhoods of different sizes (r) for the image in Figure 11(a). In all three cases, the method associates very low $\boldsymbol{p}$ values with flat regions (thus promoting enforced sparsity) and higher values with texture (where gradients show oscillations). Similarly to what was observed for WTV, the scale parameters $\boldsymbol{\alpha}$ are again smaller on regions characterized by finer details, as expected.

7.4. Parameter Estimation for the WDTV$_{\boldsymbol{p}}^{sv}$ Regularizer. For the WDTV$_{\boldsymbol{p}}^{sv}$ regularization term, after selecting the functions $f_{\text{WDTV}_{\boldsymbol{p}}^{sv}}, h_{\text{WDTV}_{\boldsymbol{p}}^{sv}}$ as in (4.10), (6.10) and the domain $\mathcal{D}_{\boldsymbol{\Theta}_i}$ as specified in Table 2, we get that the problem of interest takes the form

$$\{\alpha_i^*, p_i^*, \theta_i^*, a_i^*\} \in \underset{(\alpha_i,p_i,\theta_i,a_i)\in\mathbb{R}^2_{++}\times[-\pi/2,\pi/2)\times(0,1]}{\arg\min} \mathcal{G}(\alpha_i, p_i, \theta_i, a_i), \tag{7.7}$$

where

$$\begin{aligned}\mathcal{G}(\alpha_i, p_i, \theta_i, a_i) := & -m \ln a_i + m \ln \Gamma\left(\frac{2}{p_i} + 1\right) - 2m \ln \alpha_i \\ & + m\left(\frac{2}{p_i} - 1\right) \ln 2 + \alpha_i^{p_i} \sum_{j\in\mathcal{J}_i^r} (\, \boldsymbol{g}_j^T \mathbf{R}_{\theta_i} \boldsymbol{\Lambda}_{a_i}^2 \mathbf{R}_{-\theta_i} \boldsymbol{g}_j \,)^{p_i/2}.\end{aligned} \tag{7.8}$$

Note that $\mathcal{G}$ is differentiable on $\mathbb{R}^2_{++} \times [-\pi/2, \pi/2) \times (0, 1]$. By simply imposing a first-order optimality condition on α_i, we get the following closed formula:

$$\frac{\partial \mathcal{G}}{\partial \alpha_i} = -2m \frac{1}{\alpha_i} + p_i \alpha_i^{p_i - 1} \sum_{j \in \mathcal{J}_i^r} (\boldsymbol{g}_j^T \mathbf{R}_{\theta_i} \boldsymbol{\Lambda}_{a_i}^2 \mathbf{R}_{-\theta_i} \boldsymbol{g}_j)^{p_i/2},$$

which yields

$$\alpha_i^*(p_i, \theta_i, a_i) = \left(\frac{p_i}{2m} \sum_{j=1}^{m} (\boldsymbol{g}_j^T \mathbf{R}_{\theta_i} \boldsymbol{\Lambda}_{a_i}^2 \mathbf{R}_{-\theta_i} \boldsymbol{g}_j)^{p_i/2} \right)^{-\frac{1}{p_i}}, \tag{7.9}$$

and which can be regularized depending on $0 < \varepsilon \ll 1$ as above. The stationary point in (7.9) can be proved to be a minimum as the second derivative of $\mathcal{G}$ with respect to α_i at α_i^* is strictly positive. Plugging (7.9) into (7.8), we thus get

$$\begin{aligned} G(p_i, \theta_i, a_i) &:= \mathcal{G}(\alpha_i^*(p_i, \theta_i, a_i), p_i, \theta_i, a_i) \\ &= m \ln \left[\Gamma \left(\frac{2}{p_i} + 1 \right) \frac{1}{2\, a_i} \right] + \frac{2\, m}{p_i} \left(\ln \frac{p_i}{m} + 1 \right) \\ &\quad + \frac{2\, m}{p_i} \ln \left(\sum_{j=1}^{m} (\boldsymbol{g}_j^T \mathbf{R}_{\theta_i} \boldsymbol{\Lambda}_{a_i}^2 \mathbf{R}_{-\theta_i} \boldsymbol{g}_j)^{p_i/2} \right). \end{aligned} \tag{7.10}$$

By now making explicit the dependence of G on the entries of $(\mathbf{R}_{\theta_i} \boldsymbol{\Lambda}_{a_i}^2 \mathbf{R}_{-\theta_i})$, we have that (7.10) turns into

$$\begin{aligned} G(p_i, \theta_i, a_i) &= m \ln \left[\Gamma \left(\frac{2}{p_i} + 1 \right) \frac{1}{2\, a_i} \right] + \frac{2\, m}{p_i} \left(\ln \frac{p_i}{m} + 1 \right) \\ &\quad + \frac{2m}{p_i} \ln \left(\sum_{j \in \mathcal{J}_i^r} ((\cos^2 \theta_i + a_i^2 \sin^2 \theta_i) g_{j,1}^2 + (\sin^2 \theta_i + a_i^2 \cos^2 \theta_i) g_{j,2}^2 \right. \\ &\quad \left. + 2(1 - a_i^2) \cos \theta_i \sin \theta_i g_{j,1} g_{j,2})^{p_i/2} \right). \end{aligned} \tag{7.11}$$

Problem (7.7)–(7.8) thus takes the form

$$\{p_i^*, \theta_i^*, a_i^*\} \in \underset{(p_i, \theta_i, a_i) \in \mathbb{R}_{++} \times [-\pi/2, \pi/2) \times (0,1]}{\arg\min} G(p_i, \theta_i, a_i). \tag{7.12}$$

We now study the behavior of G as the triplet (p_i, θ_i, a_i) approaches the boundary of the set $\widehat{\mathcal{D}}_{\boldsymbol{\Theta}_i} := \mathbb{R}_{++} \times [-\pi/2, \pi/2) \times (0, 1]$. Note that, since problem (7.12) is formulated over a noncompact set of $\mathbb{R}^3$, the existence of a solution is in general not guaranteed. One possible way to overcome the problem of noncompactness consists in characterizing explicitly the configurations of the samples $\mathcal{S}_i$ for which the functional G in (7.11) does not attain its minimum in $\widehat{\mathcal{D}}_{\boldsymbol{\Theta}_i}$. To do so, let us first set

$$\begin{aligned} A(\theta_i, a_i) := \frac{2m}{p_i} \log \left[\sum_{j \in \mathcal{J}_i^r} ((\cos^2 \theta_i + a_i^2 \sin^2 \theta_i) g_{j,1}^2 + (\sin^2 \theta_i + a_i^2 \cos^2 \theta_i) g_{j,2}^2 \right. \\ \left. + 2(1 - a_i^2) \cos \theta_i \sin \theta_i g_{j,1} g_{j,2})^{p_i/2} \right]. \end{aligned}$$

For any $p_i > 0$, if $A(\theta_i, a_i)$ is bounded as $a_i \to 0^+$, then the functional G in (7.11) tends to $+\infty$ and the minimum is necessarily attained in the interior of $\widehat{\mathcal{D}}_{\boldsymbol{\Theta}_i}$. However, if $A(\theta_i, a_i)$ is unbounded as $a_i \to 0^+$, nothing can be said about the behavior of G at the boundary and, as a consequence, nothing can be said about its minima. In particular, in this situation there may exist one or multiple configurations of the samples $\boldsymbol{g}_1, \ldots, \boldsymbol{g}_m \in \mathcal{S}_i$ for which G tends to $-\infty$ at the boundary. In order to characterize such configurations, note that as $a_i \to 0^+$ we have that by continuity

$$A(\theta_i, a_i) \to \frac{2m}{p_i} \log \left[\sum_{j=1}^{m} (\cos\theta_i g_{j,1} + \sin\theta_i g_{j,2})^{p_i} \right],$$

which tends to $-\infty$ if and only if

$$g_{j,2} = -\frac{\cos\theta_i}{\sin\theta_i}\, g_{j,1} \qquad \forall j = 1, \ldots, m.$$

This situation corresponds to the very particular case when the samples $\boldsymbol{g}_j$ lie all on the line passing through the origin with slope $-\cos\theta_i / \sin\theta_i$, and they can be thus considered as realizations of a degenerate BGG pdf characterized by a positive semidefinite covariance matrix. This sort of configuration can be avoided by requiring that a_i does not get smaller than a fixed value $0 < \delta \ll 1$.

A possible way to guarantee the existence of solutions of the problem (7.12) is to reformulate the problem over a compact subset of $\mathbb{R}^3$, in analogy with what has been done in section 7.3. As noted above on the admissible values for p_i, we point out that the more we enforce sparsity (i.e., the closer p_i is to zero), the more the BGGD will tend to a Dirac delta distribution, making the estimation of local anisotropy in a neighborhood of the point considered almost impossible. Hence, the exponent p_i is considered to be confined in the closed interval $[\epsilon, R]$, with $0 < \epsilon < R$.

We can thus reformulate problem (7.12) as

$$\{p_i^*, \theta_i^*, a_i^*\} \in \min_{p_i, \theta_i, a_i} G(p_i, \theta_i, a_i) \tag{7.13}$$

$$\text{s.t.} \quad p_i \in [\epsilon, R], \quad -\pi/2 \le \theta_i \le \pi/2\,, \quad \delta \le a_i \le 1.$$

The following result holds true.

PROPOSITION 7.3. *The function $G : [\epsilon, R] \times [-\pi/2, \pi/2] \times [\delta, 1] \to \mathbb{R}$ in (7.11) is continuous and admits a minimum in its compact domain.*

In Figure 15, we analyze the performance of the outlined parameter estimation strategy for the $\text{WDTV}_{\boldsymbol{p}}^{sv}$ regularizer, where again the search interval for the local parameter p_i has been set as $[\epsilon, R] = [0.1, 10]$. More specifically, in the left column we display selected neighborhoods from a synthetic image—i.e., a vertical edge in Figure 15(a), a horizontal edge in Figure 15(d), and a circular profile in Figure 15(g)—and the two textured regions already considered in Figures 11 and 13. In the middle column of Figure 15, we report the samples extracted from each neighborhood, together with the level curves of the estimated local BGG pdfs, while in the last column we show the scatter plot of the samples by drawing once again the level curves of the underlying distribution to facilitate the analysis.

The estimated pdfs for the three geometrical profiles lie along the horizontal axis, the vertical axis, and the first quadrant bisector of the scatter diagram $\mathbf{D}_h\boldsymbol{u}$-$\mathbf{D}_v\boldsymbol{u}$,

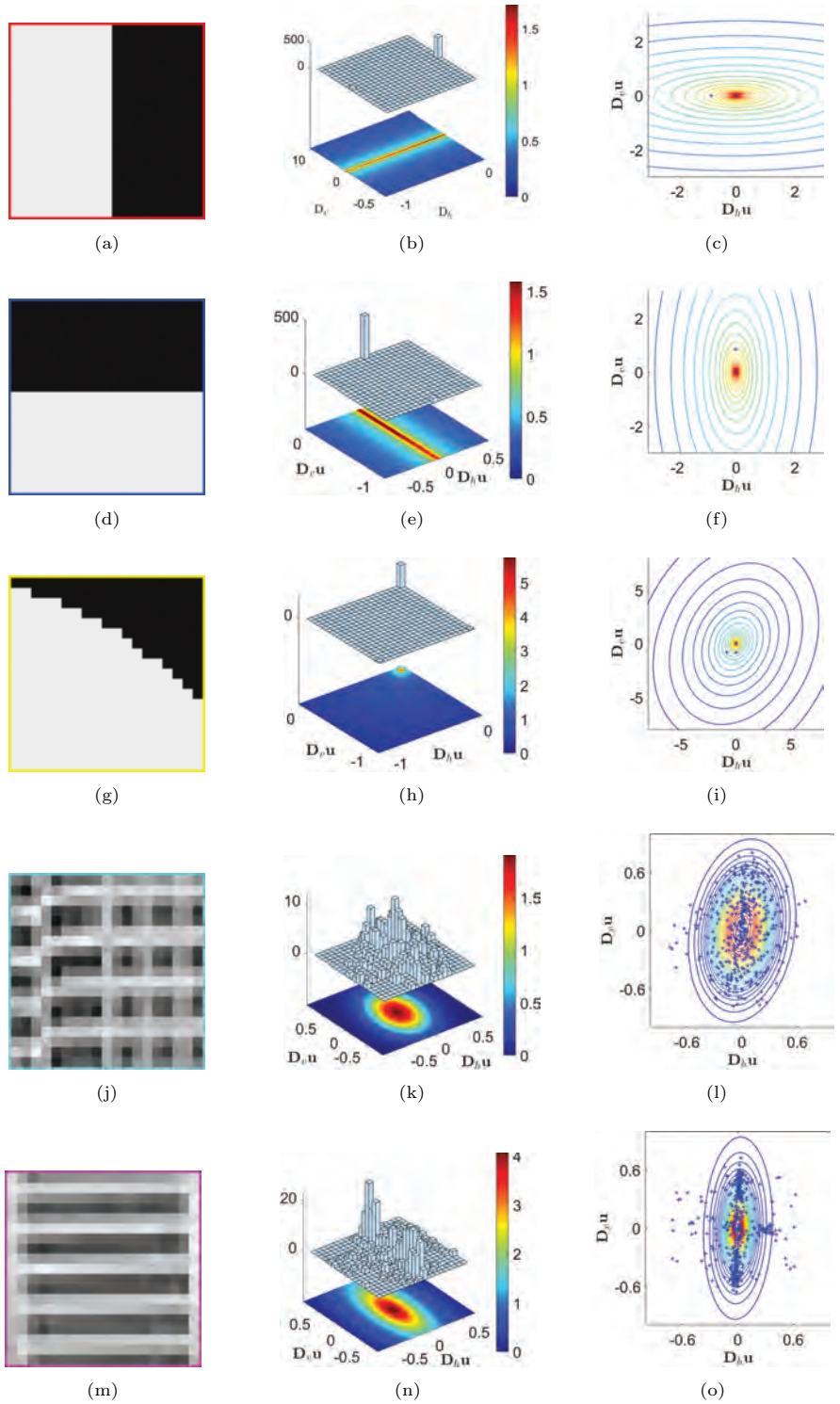


Fig. 15 Parameter estimate for $\mathrm{WDTV}_{\boldsymbol{p}}^{sv}$. *From left to right: neighborhoods, histograms, and scatter plots of gradients with level curves of the estimated BGG pdf. From top to bottom, the estimates are* $(\alpha^*, p^*, \theta^*, a^*) = (2.23, 0.1, -89.90, 0.33)$, $(\alpha^*, p^*, \theta^*, a^*) = (2.10, 0.1, 0, 0.37)$, $(\alpha^*, p^*, \theta^*, a^*) = (1.14, 0.1, 47.50, 0.78)$, $(\alpha^*, p^*, \theta^*, a^*) = (1.30, 2.97, -12.50, 0.64)$, $(\alpha^*, p^*, \theta^*, a^*) = (1.94, 1.24, -2.18, 0.39)$.

respectively. This behavior, as expected, corresponds to the dominant orientation of the gradients in the neighborhoods. Finally, the textured regions in Figures 15(g),(j), statistically indistinguishable from the hLd and hGGd viewpoint, turn out to be significantly different now; in fact, the samples in the former are spread more homogeneously in the scatter diagram, while the gradients in the latter present a dominant edge orientation which is almost aligned with the horizontal axis of the diagram. Such a difference is now reflected in the estimated BGG pdfs.

8. Algorithmic Optimization. From an optimization point of view, it is not trivial to design a unified optimization solver for the general $\boldsymbol{u}$-estimation problem in the alternating scheme (6.11)–(6.12), as it may be either extremely easy (smooth and convex) or extremely difficult (nonsmooth and nonconvex). We can surely think of specific optimization algorithms that could be effectively used for solving (6.11)–(6.12) in specific scenarios such as (l-)BFGS [24] for the smooth and convex case, Nesterov-type proximal schemes [118, 11, 55, 130] and dual/primal-dual methods [81, 50, 86, 41, 44] for the nonsmooth convex case, and, e.g., [121] for the nonsmooth nonconvex case. However, in the following we will stick with one single optimization algorithm for better clarity and consider the alternating direction method of multipliers (ADMM) [20] whose different subproblems can be solved by means of classical tools in the field of proximal calculus [54], numerical linear algebra, and adaptive discrepancy principle [79]. Note that albeit proposed and widely applied in convex scenarios, nonconvex variants of ADMM have been recently proposed and endowed with global convergence guarantees [148, 18], although not always applicable to the problem at hand due to the (often limiting) assumptions on the operators involved. However, as we will comment on in the following, empirical convergence is often observed for general nonconvex ADMM algorithms, which makes their use often amenable in practice. For further details on the recent developments of convex and nonconvex optimization algorithms for variational imaging models, we refer the reader to [45] and the references therein.

8.1. ADMM Optimization. By dropping out the terms in (6.6) which do not depend on the unknown image $\boldsymbol{u}$, the $\boldsymbol{u}$-update step (6.12) reads

$$\boldsymbol{u}^{(k+1)} \in \underset{\mathbf{u}\in\mathbb{R}^N}{\arg\min} \left\{ \sum_{i=1}^{N} f((\mathbf{D}\boldsymbol{u})_i; \boldsymbol{\Theta}_i^{(k+1)}) + F_q(\mathbf{A}\boldsymbol{u}; \boldsymbol{b}) \right\}, \tag{8.1}$$

where $F_q(\mathbf{A}\boldsymbol{u}; \boldsymbol{b})$ is defined in (6.4) for $q < +\infty$ and in (6.5) for $q = +\infty$, and the gradient penalty functions f are summarized in Table 2 for the different regularizers considered.

By introducing the auxiliary variables $\boldsymbol{g} \in \mathbb{R}^{2N}$ and $\boldsymbol{y} \in \mathbb{R}^M$, and dropping out the iteration superscript, problem (8.1) can be reformulated as

$$\{\boldsymbol{u}^*, \boldsymbol{g}^*, \boldsymbol{y}^*\} \in \underset{\boldsymbol{u},\boldsymbol{g},\boldsymbol{y}}{\arg\min} \left\{ \sum_{i=1}^{N} f(\boldsymbol{g}_i; \boldsymbol{\Theta}_i) + F_q(\boldsymbol{y}; \boldsymbol{b}) \right\} \text{ s.t. } \begin{cases} \boldsymbol{g} = \mathbf{D}\boldsymbol{u}, \\ \boldsymbol{y} = \mathbf{A}\boldsymbol{u}. \end{cases} \tag{8.2}$$

For every $i = 1, \ldots, N$, the quantity $\boldsymbol{g}_i = ((\mathbf{D}_h\boldsymbol{u})_i, (\mathbf{D}_v\boldsymbol{u})_i) \in \mathbb{R}^2$ stands for the local image gradient at pixel i. By means of this change of variable, we can avoid considering the dependence on the linear operator $\mathbf{D}$ of the (in general) nondifferentiable and possibly nonconvex function f, while the use of $\boldsymbol{y}$ is helpful for the GDP strategy introduced in section 2.1.

We define the augmented Lagrangian functional of problem (8.2) as follows:

$$\begin{aligned}\mathcal{L}(\boldsymbol{u},\boldsymbol{g},\boldsymbol{y},\boldsymbol{\rho}_g,\boldsymbol{\rho}_y;\boldsymbol{\Theta}) := &\sum_{i=1}^{N} f(\boldsymbol{g}_i;\boldsymbol{\Theta}_i) + F_q(\boldsymbol{y};\boldsymbol{b}) - \langle \boldsymbol{\rho}_g, \boldsymbol{g} - \mathbf{D}\boldsymbol{u}\rangle \\ &+ \frac{\beta_g}{2}\|\boldsymbol{g} - \mathbf{D}\boldsymbol{u}\|_2^2 - \langle \boldsymbol{\rho}_y, \boldsymbol{y} - \mathbf{A}\boldsymbol{u}\rangle + \frac{\beta_y}{2}\|\boldsymbol{y} - \mathbf{A}\boldsymbol{u}\|_2^2,\end{aligned} \tag{8.3}$$

where $\beta_g, \beta_y \in \mathbb{R}_{++}$ are the ADMM penalty parameters, while $\boldsymbol{\rho}_g \in \mathbb{R}^{2N}$, $\boldsymbol{\rho}_y \in \mathbb{R}^M$ are the vectors of Lagrange multipliers associated with the linear constraints $\boldsymbol{g} = \mathbf{D}\boldsymbol{u}$ and $\boldsymbol{y} = \mathbf{A}\boldsymbol{u}$ in (8.2), respectively.

Solving (8.2) amounts to seeking solutions of the following saddle-point problem:

$$\begin{aligned}&\text{Find } (\boldsymbol{u}^*,\boldsymbol{g}^*,\boldsymbol{y}^*) \in \mathbb{R}^N \times \mathbb{R}^{2N} \times \mathbb{R}^M \text{ and } (\boldsymbol{\rho}_g^*, \boldsymbol{\rho}_y^*) \in \mathbb{R}^{2N} \times \mathbb{R}^M \text{ such that} \\ &\mathcal{L}(\boldsymbol{u}^*,\boldsymbol{g}^*,\boldsymbol{y}^*,\boldsymbol{\rho}_g,\boldsymbol{\rho}_y;\boldsymbol{\Theta}) \le \mathcal{L}(\boldsymbol{u}^*,\boldsymbol{g}^*,\boldsymbol{y}^*,\boldsymbol{\rho}_g^*,\boldsymbol{\rho}_y^*;\boldsymbol{\Theta}) \le \mathcal{L}(\boldsymbol{u},\boldsymbol{g},\boldsymbol{y},\boldsymbol{\rho}_g^*,\boldsymbol{\rho}_y^*,\boldsymbol{\Theta}) \\ &\forall\, (\boldsymbol{u},\boldsymbol{g},\boldsymbol{y}) \in \mathbb{R}^N \times \mathbb{R}^{2N} \times \mathbb{R}^M,\ \ \forall\, (\boldsymbol{\rho}_g,\boldsymbol{\rho}_y) \in \mathbb{R}^{2N} \times \mathbb{R}^M.\end{aligned} \tag{8.4}$$

Upon suitable initialization, and for any $j \ge 0$, the jth iteration of the ADMM algorithm applied to solve the saddle-point problem (8.4) thus reads

$$\boldsymbol{u}^{(j+1)} \in \underset{\boldsymbol{u}\in\mathbb{R}^N}{\arg\min}\ \mathcal{L}\big(\boldsymbol{u}, \boldsymbol{g}^{(j)}, \boldsymbol{y}^{(j)}, \boldsymbol{\rho}_g^{(j)}, \boldsymbol{\rho}_y^{(j)}; \boldsymbol{\Theta}\big), \tag{8.5}$$

$$\boldsymbol{g}^{(j+1)} \in \underset{\boldsymbol{g}\in\mathbb{R}^{2N}}{\arg\min}\ \mathcal{L}\big(\boldsymbol{u}^{(j+1)}, \boldsymbol{g}, \boldsymbol{y}^{(j)}, \boldsymbol{\rho}_g^{(j)}, \boldsymbol{\rho}_y^{(j)}; \boldsymbol{\Theta}\big), \tag{8.6}$$

$$\boldsymbol{y}^{(j+1)} \in \underset{\boldsymbol{y}\in\mathbb{R}^M}{\arg\min}\ \mathcal{L}\big(\boldsymbol{u}^{(j+1)}, \boldsymbol{g}^{(j+1)}, \boldsymbol{y}, \boldsymbol{\rho}_g^{(j)}, \boldsymbol{\rho}_y^{(j)}; \boldsymbol{\Theta}\big), \tag{8.7}$$

$$\boldsymbol{\rho}_g^{(j+1)} = \boldsymbol{\rho}_g^{(j)} - \beta_g\big(\boldsymbol{g}^{(j+1)} - \mathbf{D}\boldsymbol{u}^{(j+1)}\big), \tag{8.8}$$

$$\boldsymbol{\rho}_y^{(j+1)} = \boldsymbol{\rho}_y^{(j)} - \beta_r\big(\boldsymbol{y}^{(j+1)} - \mathbf{A}\boldsymbol{u}^{(j+1)}\big). \tag{8.9}$$

In the following, we make precise the solution of the three subproblems for the primal variables $\boldsymbol{u}$, $\boldsymbol{g}$, and $\boldsymbol{y}$ in (8.5)–(8.7). The automatic estimation of the regularization parameter μ will be addressed in section 8.4 concerned with the $\boldsymbol{y}$-update.

8.2. Subproblem for the Primal Variable $\boldsymbol{u}$. Subproblem (8.5) reads

$$\boldsymbol{u}^{(j+1)} \in \underset{\boldsymbol{u}\in\mathbb{R}^N}{\arg\min}\ \left\{\langle \boldsymbol{\rho}_g^{(j)}, \mathbf{D}\boldsymbol{u}\rangle + \langle \boldsymbol{\rho}_y^{(j)}, \mathbf{A}\boldsymbol{u}\rangle + \frac{\beta_g}{2}\|\boldsymbol{g}^{(j)} - \mathbf{D}\boldsymbol{u}\|_2^2 + \frac{\beta_r}{2}\|\boldsymbol{y}^{(j)} - \mathbf{A}\boldsymbol{u}\|_2^2\right\},$$

which is quadratic with first-order optimality condition given by

$$\left(\beta_g\mathbf{D}^T\mathbf{D} + \beta_r\mathbf{A}^T\mathbf{A}\right)\boldsymbol{u} = \beta_g\mathbf{D}^T\left(\boldsymbol{g}^{(j)} - \frac{1}{\beta_g}\boldsymbol{\rho}_g^{(j)}\right) + \beta_r\mathbf{A}^T\left(\boldsymbol{y}^{(j)} - \frac{1}{\beta_r}\boldsymbol{\rho}_y^{(j)}\right). \tag{8.10}$$

The coefficient matrix of the linear system above is symmetric positive semidefinite and, under the assumption

$$\operatorname{null}(\mathbf{A}) \cap \operatorname{null}(\mathbf{D}) = \{\mathbf{0}_N\},$$

then it is positive definite so that $\boldsymbol{u}^{(j+1)}$ is the unique solution of linear system (8.10). Matrix $\mathbf{A}$ is typically sparse, hence (8.10) can be solved efficiently by means of (preconditioned) conjugate gradient methods. When $\mathbf{A}$ is a convolution matrix, like in image restoration with space-invariant blur, the linear system can be solved more efficiently by means of fast 2D discrete transforms.

8.3. Subproblem for the Primal Variable $\boldsymbol{g}$. After dropping all terms not depending on $\boldsymbol{g}$ in (8.3), subproblem (8.6) reads

$$\boldsymbol{g}^{(j+1)} \in \underset{\boldsymbol{g}\in\mathbb{R}^{2N}}{\arg\min} \left\{ \sum_{i=1}^{N} f(\boldsymbol{g}_i;\boldsymbol{\Theta}_i) - \langle \boldsymbol{\rho}_g^{(j)}, \boldsymbol{g} - \mathbf{D}\boldsymbol{u}^{(j+1)} \rangle + \frac{\beta_g}{2} \left\| \boldsymbol{g} - \mathbf{D}\boldsymbol{u}^{(j+1)} \right\|_2^2 \right\}$$

$$= \underset{\boldsymbol{g}\in\mathbb{R}^{2N}}{\arg\min} \left\{ \sum_{i=1}^{N} f(\boldsymbol{g}_i;\boldsymbol{\Theta}_i) + \frac{\beta_g}{2} \|\boldsymbol{g} - \boldsymbol{w}^{(j)}\|_2^2 \right\}, \tag{8.11}$$

with vector $\boldsymbol{w}^{(j)} \in \mathbb{R}^{2N}$ defined by

$$\boldsymbol{w}^{(j)} := \mathbf{D}\boldsymbol{u}^{(j+1)} + \frac{1}{\beta_g} \boldsymbol{\rho}_g^{(j)} . \tag{8.12}$$

Solving the $2N$-dimensional minimization problem above is thus equivalent to solving the following N independent two-dimensional problems:

$$\begin{aligned} \boldsymbol{g}_i^{(j+1)} &\in \underset{\boldsymbol{g}_i\in\mathbb{R}^2}{\arg\min} \left\{ f(\boldsymbol{g}_i;\boldsymbol{\Theta}_i) + \frac{\beta_g}{2} \left\| \boldsymbol{g}_i - \boldsymbol{w}_i^{(j)} \right\|_2^2 \right\} \\ &= \operatorname{prox}_{f(\,\cdot\,;\boldsymbol{\Theta}_i)}^{\beta_g} \left(\boldsymbol{w}_i^{(j)} \right), \quad i = 1, \dots, N , \end{aligned} \tag{8.13}$$

where $\operatorname{prox}_{f(\,\cdot\,;\boldsymbol{\Theta}_i)}^{\beta_g} : \mathbb{R}^2 \rightrightarrows \mathbb{R}^2$ denotes the proximal operator of the gradient penalty function $f(\,\cdot\,;\boldsymbol{\Theta}_i)$ with proximity parameter β_g (see Definition 2.8 and section 4.1.4) and where the vectors $\boldsymbol{w}_i^{(j)} \in \mathbb{R}^2$ at any iteration read

$$\boldsymbol{w}_i^{(j)} = \left(\mathbf{D}\boldsymbol{u}^{(j+1)} \right)_i + \frac{1}{\beta_g} \left(\boldsymbol{\rho}_t^{(j)} \right)_i , \quad i = 1, \dots, N .$$

We start detailing the solving procedure for problem (8.13) under the adoption of a $\mathrm{WDTV}_{\boldsymbol{p}}^{sv}$ regularization term, which corresponds to considering the gradient penalty function $f_{\mathrm{WDTV}_{\boldsymbol{p}}^{sv}}$ defined in (4.10); the proximal maps arising for the WTV and $\mathrm{WTV}_{\boldsymbol{p}}^{sv}$ regularizers will be discussed afterwards as special cases. In [27, Proposition 6.3], the authors proved a result on the existence of solutions for problem (8.13). Before reporting the statement, we recall that in the following, for $\boldsymbol{v}, \boldsymbol{w} \in \mathbb{R}^n$ we denote by $\boldsymbol{v} \circ \boldsymbol{w}$, $|\boldsymbol{v}|$, and $\operatorname{sign}(\boldsymbol{v})$ the componentwise (or Hadamard) product between $\boldsymbol{v}$ and $\boldsymbol{w}$ and the componentwise absolute value and sign of $\boldsymbol{v}$, respectively.

LEMMA 8.1. *Let $f : \mathbb{R}^2 \to \mathbb{R}_+$ be the (parametric and not necessarily convex) function defined by*

$$f(\boldsymbol{g}) := \alpha^p \, \|\boldsymbol{\Lambda}_a \mathbf{R}_{-\theta} \, \boldsymbol{g}\|_2^p , \quad \boldsymbol{g} \in \mathbb{R}^2 ,$$

with parameters $\alpha, p \in \mathbb{R}_{++}$, $a \in (0, 1]$, $\theta \in [-\pi/2, \pi/2)$, $\boldsymbol{\Lambda}_a = \operatorname{diag}(1, a)$, and $\mathbf{R}_{-\theta}$ the 2×2 rotation matrix of angle $-\theta$, and let $\operatorname{prox}_f^{\beta} : \mathbb{R}^2 \rightrightarrows \mathbb{R}^2$ be the proximal operator of f with proximity parameter $\beta \in \mathbb{R}_{++}$ defined by

$$\boldsymbol{g}^* \in \operatorname{prox}_f^{\beta}(\boldsymbol{w}) := \underset{\boldsymbol{g}\in\mathbb{R}^2}{\arg\min} \left\{ F(\boldsymbol{g}) := f(\boldsymbol{g}) + \frac{\beta}{2} \|\boldsymbol{g} - \boldsymbol{w}\|_2^2 \right\}, \quad \boldsymbol{w} \in \mathbb{R}^2 . \tag{8.14}$$

Then problem (8.14) *admits at least one solution, which is unique when $p \geq 1$. Moreover, after defining*

$$\widetilde{\boldsymbol{w}} := \mathbf{R}_{-\theta} \, \boldsymbol{w}, \quad \boldsymbol{s} := \operatorname{sign}(\widetilde{\boldsymbol{w}}), \quad \overline{\boldsymbol{w}} := |\widetilde{\boldsymbol{w}}|, \quad \bar{\beta} := \frac{\beta}{\alpha^p},$$

we have that any solution $\boldsymbol{g}^*$ *of* (8.14) *can be expressed as*

$$\boldsymbol{g}^* = \mathbf{R}_\theta\,(\boldsymbol{s} \circ \boldsymbol{z}^*)\,, \qquad \boldsymbol{z}^* \in \underset{\boldsymbol{z}\in\mathcal{H}_1\subset\mathbb{R}^2}{\arg\min}\; H(\boldsymbol{z})\,, \tag{8.15}$$

where $H : \mathbb{R}^2 \to \mathbb{R}_+$ *and* $\mathcal{H}_1 \subset \mathbb{R}^2$ *are defined by*

$$H(\boldsymbol{z}) := \|\boldsymbol{\Lambda}_a\,\boldsymbol{z}\|_2^p + \frac{\bar{\beta}}{2}\,\|\,\boldsymbol{z} - \overline{\boldsymbol{w}}\,\|_2^2\,, \qquad \mathcal{H}_1 := \mathcal{H} \cap \left(\left[0, \overline{w}_1\right] \times \left[0, \overline{w}_2\right]\right),$$

with $\mathcal{H}$ *being*

1. *the rectangular hyperbola defined by*

$$\mathcal{H} := \left\{ \boldsymbol{z} \in \mathbb{R}^2 :\; (z_1 - c_1)\,(z_2 - c_2) = c_1 c_2,\; c_1 = -\,\frac{a^2\,\overline{w}_1}{1-a^2}, c_2 = \frac{\overline{w}_2}{1-a^2} \right\} \tag{8.16}$$

   for $a \in (0,1)$ *and* $\overline{w}_1\overline{w}_2 \neq 0$*;*

2. *the line defined by*

$$\mathcal{H} := \left\{\, \boldsymbol{z} \in \mathbb{R}^2 :\;\; \overline{w}_2 z_1 - \overline{w}_1 z_2 = 0 \,\right\}$$

   for $a \in (0,1)$ *and* $\overline{w}_1\overline{w}_2 = 0$*, or for* $a = 1$ *and any* $\overline{w}_1, \overline{w}_2 \in \mathbb{R}_+$*.*

COROLLARY 8.2. *The minimizers* $\boldsymbol{z}^* \in \mathbb{R}^2$ *in* (8.15) *can be obtained as follows:*

$$\boldsymbol{z}^* = \left(z_1^*,\, \frac{c_2\, z_1^*}{z_1^* - c_1} \right),$$

where $c_1,\ c_2 \in \mathbb{R}$ *are defined in* (8.16) *and* $z_1^* \in \mathbb{R}$ *is the solution(s) of the following one-dimensional constrained minimization problem:*

$$z_1^* \;\in\; \underset{\xi\,\in\,[0,\overline{w}_1]}{\arg\min}\; \left\{\, h(\xi) := \left(h_1(\xi)\right)^{p/2} \;+\; \frac{\bar{\beta}}{2}\, h_2(\xi) \right\},$$

$$h_1(\xi) = \xi^2 \left(1 + \frac{a^2\, c_2^2}{(\xi - c_1)^2}\right), \quad h_2(\xi) = (\xi - \overline{w}_1)^2 + \left(\frac{c_2\,\xi}{\xi - c_1} - \overline{w}_2\right)^2.$$

We will omit the proof of this lemma, and provide only a brief graphical sketch of the key steps leading to (8.15). First, in order to get some clues about the approximate position of the minimizer $\boldsymbol{z}^*$ in the plane z_1-z_2, we restrict the study of the function H to the one-parameter family of ellipses:

$$\mathcal{E}_R(\mathbf{0}) = \left\{ (z_1, z_2) \in \mathbb{R}^2 \mid z_1^2 + a^2 z_2^2 = R \right\}. \tag{8.17}$$

One can prove that $\boldsymbol{z}^*$ needs to belong to the hyperbola $\mathcal{H}$ defined in (8.16). More specifically, the sought $\boldsymbol{z}^*$ has to coincide with one of the two points in $\mathcal{E}_R \cap \mathcal{H}$ belonging to the first quadrant of the plane z_1-z_2. In Figure 16(a), we show one ellipse $\mathcal{E}_R$, which is depicted by a blue dashed line, and the hyperbola $\mathcal{H}$, plotted with a solid magenta line. We conclude that $\boldsymbol{z}^*$ lies on the arc of hyperbola $\mathcal{H}_1$ which is delimited by the origin $\boldsymbol{O}$ and $\overline{\boldsymbol{w}}$; $\mathcal{H}_1$ is also illustrated in Figure 16(a) with a solid red line.

Remark 8.3. Upon the adoption of the $\mathrm{WTV}_{\boldsymbol{p}}^{sv}$ regularizer, a similar result can be proven; see [101, Proposition 1]. More specifically, in isotropic settings there holds

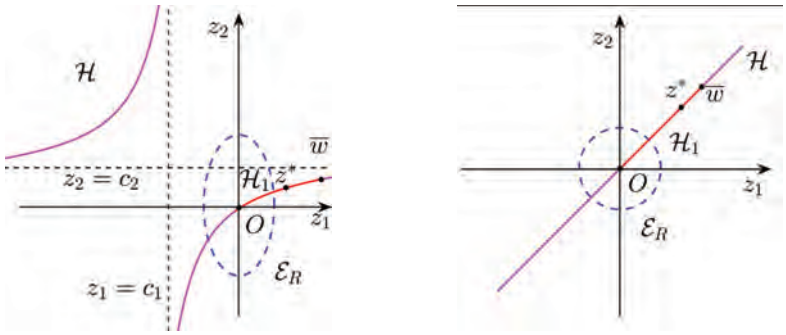


Fig. 16 *Graphical representation of the bivariate minimization problems arising for the* $\mathrm{WDTV}^{sv}_{\boldsymbol{p}}$ (a) *and the* $\mathrm{WTV}^{sv}_{\boldsymbol{p}}$ *regularizers* (b).

$a = 1$, which yields that the parametric family of ellipses in (8.17) reduces to a parametric family of circles. Moreover, the hyperbola $\mathcal{H}$ and the arc $\mathcal{H}_1$ turn into a line and a segment, respectively. The simplified configuration is reported in Figure 16(b); notice that also in this case $\mathcal{H}_1$ lies between the origin $\boldsymbol{O}$ and $\overline{\boldsymbol{w}} = |\boldsymbol{w}|$.

For the WTV regularizer, the solutions of the separable $\boldsymbol{g}_i$-subproblems can be written in closed form by means of a soft-thresholding operator; see, e.g., [159].

8.4. Subproblem for the Primal Variable $\boldsymbol{y}$. For $1 \leq q < +\infty$ and $F_q(\boldsymbol{y}; \boldsymbol{b}) = q\omega^q \mathrm{L}_q(\boldsymbol{y}; \boldsymbol{b})$ as in (6.4), recalling the notation introduced in Table 2 and Remark 6.2, after dropping all terms not depending on $\boldsymbol{y}$ in (8.3), subproblem (8.7) reads

$$\begin{aligned} \boldsymbol{y}^{(j+1)} &\in \underset{\boldsymbol{y}\in\mathbb{R}^M}{\arg\min} \left\{ \mu^{(j)}\, \mathrm{L}_q(\boldsymbol{y}; \boldsymbol{b}) - \langle \boldsymbol{\rho}_y^{(j)}, \boldsymbol{y} - \mathbf{A}\boldsymbol{u}^{(j+1)} \rangle + \frac{\beta_r}{2} \left\| \boldsymbol{y} - \mathbf{A}\boldsymbol{u}^{(j+1)} \right\|_2^2 \right\} \\ &= \underset{\boldsymbol{y}\in\mathbb{R}^M}{\arg\min} \left\{ \lambda^{(j)}\, \mathrm{L}_q(\boldsymbol{y}; \boldsymbol{b}) + \frac{1}{2} \|\boldsymbol{y} - \boldsymbol{c}^{(j)}\|_2^2 \right\}, \end{aligned} \tag{8.18}$$

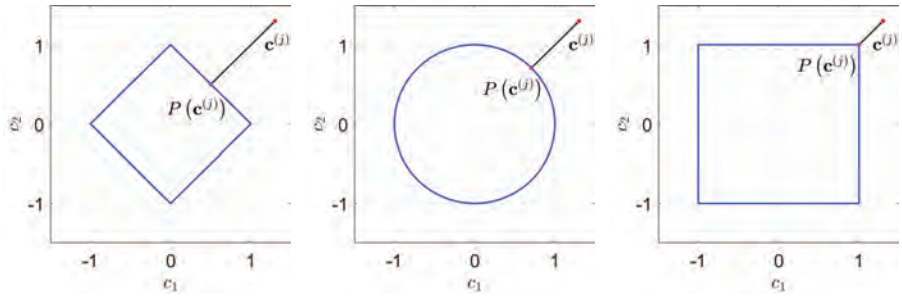
where the variables $\lambda^{(j)} \in \mathbb{R}_{++}$ and $\boldsymbol{c}^{(j)} \in \mathbb{R}^M$ are defined by

$$\lambda^{(j)} := \frac{\mu^{(j)}}{\beta_r}, \qquad \boldsymbol{c}^{(j)} := \mathbf{A}\boldsymbol{u}^{(j)} + \frac{1}{\beta_r} \boldsymbol{\rho}_y^{(j)}.$$

Note that, as already observed in section 6.3, the regularization parameter μ is not assumed to be fixed, but it is rather estimated along the ADMM iterations (hence the $^{(j)}$ superscript) based on the GDP strategy detailed in section 2.1. In order to update $\mu^{(j)}$, i.e., $\lambda^{(j)}$, so that the GDP is automatically satisfied, we can regard $\lambda^{(j)}$ as a Lagrange multiplier, and then exploit the well-known duality property (see, e.g., [15, Chapter 1]), which allows us to replace the unconstrained problem in (8.18) with its constrained formulation

$$\boldsymbol{y}^{(j+1)} \in \underset{\boldsymbol{y}\in\mathcal{B}_\delta^q}{\arg\min} \left\{ \|\boldsymbol{y} - \boldsymbol{c}^{(j)}\|_2^2 \right\} = P_{\mathcal{B}_\delta^q}\left(\boldsymbol{c}^{(j)}\right), \tag{8.19}$$

where $P_{\mathcal{B}_\delta^q}$ denotes the Euclidean projection onto the ℓ_q-ball $\mathcal{B}_\delta^q = \left\{ \boldsymbol{y} \in \mathbb{R}^M : \|\boldsymbol{y}\|_q \leq \delta_q \right\}$, with δ_q given in (2.10).



(a) $q = 1$ (b) $q = 2$ (c) $q = +\infty$

Fig. 17 *Projections on the unitary ℓ_q-balls for $q = 1$, $q = 2$, and $q = +\infty$.*

We remark that although the presence of the regularization parameter μ is not explicit in problem (8.19), it is actually embedded in the radius δ_q.

When the underlying noise is AIU, i.e., F_q is set as in (6.5), the $\boldsymbol{y}$-update can be expressed as the constrained minimization problem in (8.19), where the constraint set is the ℓ_∞-ball with radius δ_∞ defined in (2.10).

Note that the projections onto the ℓ_2- and ℓ_∞-balls can be efficiently computed by

$$q = 2: \qquad P_{\mathcal{B}^2_\delta}\left(\boldsymbol{c}^{(j)}\right) = \begin{cases} \boldsymbol{c}^{(j)} & \text{if } \|\boldsymbol{c}^{(j)}\|_2 \le \delta_2, \\ \min\left(\delta_2, \|\boldsymbol{c}^{(j)}\|_2\right) \dfrac{\boldsymbol{c}^{(j)}}{\|\boldsymbol{c}^{(j)}\|_2} & \text{otherwise,} \end{cases}$$

$$q = +\infty: \qquad P_{\mathcal{B}^\infty_\delta}\left(\boldsymbol{c}^{(j)}\right) = \begin{cases} \boldsymbol{c}^{(j)} & \text{if } \|\boldsymbol{c}^{(j)}\|_\infty \le \delta_\infty, \\ \min\left(\max(\boldsymbol{c}^{(j)}, -\delta_\infty), \delta_\infty\right) & \text{otherwise,} \end{cases}$$

where all the operations have to be intended componentwise.

For $q = 1$, the projection can be computed as follows:

$$q = 1: \qquad P_{\mathcal{B}^1_\delta}\left(\boldsymbol{c}^{(j)}\right) = \begin{cases} \boldsymbol{c}^{(j)} & \text{if } \|\boldsymbol{c}^{(j)}\|_1 \le \delta_1, \\ \operatorname{sign}(\boldsymbol{c}^{(j)})\boldsymbol{x} & \text{otherwise,} \end{cases}$$

where

$$\boldsymbol{x} = \max(\boldsymbol{c}^{(j)} - \tau, 0) \;\text{ and }\; \tau \in \mathbb{R} \;:\; \|\boldsymbol{x}\|_1 = \delta_1.$$

Setting a suitable τ for the problem at hand is possibly a very expensive task from the computational viewpoint. Nonetheless, in [56] a complexity linear $O(M)$ projection algorithm has been proposed, which improves previous $O(M^2)$ and $O(M \log M)$ strategies considered, e.g., in [80, 12, 64].

In Figure 17, we show the ℓ_q-balls $\mathcal{B}^q_\delta(\mathbf{0})$ for the considered choices of q in two-dimensional settings. In the three plots, we also report the vector $\boldsymbol{c}^{(j)}$ in the case it does not belong to $\mathcal{B}^q_\delta(\mathbf{0})$, and the projection $P\left(\boldsymbol{c}^{(j)}\right)$ onto the ball.

The alternating scheme outlined in section 3.3 requires us to solve the $\boldsymbol{u}$-update via the ADMM until a fixed tolerance has been reached after having performed the $\boldsymbol{\Theta}$-update. As a result, the adoption of a *pure* alternating scheme yields a computational burden that can be partially remedied by nesting the parameter estimation in the

ADMM scheme, as formalized in Algorithm 8.1. As clearly detailed in section 7, the estimation of the parameters involved in the expression of the $\mathrm{WTV}^{sv}_{\boldsymbol{p}}$ and $\mathrm{WDTV}^{sv}_{\boldsymbol{p}}$ regularizers represents a further computational bottleneck. Therefore, one can decide to further lighten the algorithmic scheme in Algorithm 8.1 by not performing the parameter update at each iteration j, but at every few iterations.

Algorithm 8.1 Joint ADMM-scheme for hyperparameter estimation and image reconstruction.

inputs: observed image $\boldsymbol{b}$, forward model operator $\mathbf{A}$
parameters: radius $r > 0$, discrepancy parameter $\tau = 1$,
ADMM penalty parameters $\beta_g, \beta_r > 0$
outputs: estimated image $\boldsymbol{u}^*$ and parameter vector $\boldsymbol{\Theta}^*$

- **Initialization:** $\boldsymbol{u}^{(0)}$, $\boldsymbol{\rho}_g^{(0)} = \mathbf{0}_{2N}$, $\boldsymbol{\rho}_y^{(0)} = \mathbf{0}_M$
- **Nested alternating scheme:**

for *j = 0, 1, 2, ... until convergence* **do**:

· *parameters update*
update $\boldsymbol{\Theta}^{(j+1)}$ as detailed in section 7.2, 7.3, or 7.4
· *primal variables update*
update $\boldsymbol{u}^{(j+1)}$ by solving (8.10)
update $\boldsymbol{g}^{(j+1)}$ as detailed in section 8.3
update $\boldsymbol{y}^{(j+1)}$ as detailed in section 8.4
· *dual variables update*
update $\boldsymbol{\rho}_g^{(j+1)}$ by (8.8)
update $\boldsymbol{\rho}_y^{(j+1)}$ by (8.9)
end for

8.5. The Computational Cost of Space Variance. In order to delineate how space-variant regularization is paid in terms of computational efficiency, we now give a closer look to all the updating steps in the ADMM-based scheme summarized in Algorithm 8.1.

First of all, we note that the updates of the primal variables $\boldsymbol{u}$, $\boldsymbol{y}$ and the dual variables $\boldsymbol{\rho}_g$, $\boldsymbol{\rho}_y$ are independent of the regularizer considered. In terms of efficiency, Lagrange multipliers $\boldsymbol{\rho}_g$ and $\boldsymbol{\rho}_y$ are updated by (8.8) and (8.9) with linear complexities $O(N)$ in the number of pixels of the sought image $\boldsymbol{u}$ and $O(M)$ in the number of pixels of the observation $\boldsymbol{b}$, respectively (in most imaging problems, $M \leq N$). Variable $\boldsymbol{y}$ is also updated with $O(M)$ operations by the Euclidean projection in (8.19) in the three most interesting cases of additive Gaussian, Laplace, and uniform noises. Variable $\boldsymbol{u}$ is updated by solving linear system (8.10) with cubic complexity $O(N^3)$ for general coefficient matrix, approximately quadratic complexity $O(N^2)$ for sparse matrix and solution via the conjugate gradient method, and superlinear complexity $O(N \log N)$ for convolution matrix and solution by fast 2D discrete transforms. The cost for updating $\boldsymbol{u}$ thus dominates the previous ones already for medium-size images.

Then the cost for updating the variable $\boldsymbol{g}$ through the solution of the N bivariate proximal maps in (8.13) strongly depends on the functional form of the regularizer but not on its spatial variance or invariance. In fact, by plugging in (8.13) the gradient penalty functions f reported in Table 2 (actually, instead of DTV we consider

the space-invariant analogue of $\mathrm{WDTV}_p^{\mathrm{sv}}$, which we refer to as DTV_p), one obtains formally identical solution formulas for the pairs of corresponding space-invariant and space-variant regularizers, namely,

$$\boldsymbol{g}_i^{(j+1)} \in \operatorname{prox}^{\beta_i}_{\|\cdot\|_2}\Big(\boldsymbol{w}_i^{(j)}\Big) \qquad \text{for TV, WTV,} \tag{8.20}$$

$$\boldsymbol{g}_i^{(j+1)} \in \operatorname{prox}^{\beta_i}_{\|\cdot\|_2^{p_i}}\Big(\boldsymbol{w}_i^{(j)}\Big) \qquad \text{for } \mathrm{TV}_p,\ \mathrm{WTV}_{\boldsymbol{p}}^{sv}, \qquad i = 1, \dots, N, \tag{8.21}$$

$$\boldsymbol{g}_i^{(j+1)} \in \operatorname{prox}^{\beta_i}_{\|\boldsymbol{\Lambda}_{a_i}\mathbf{R}_{-\theta_i}\cdot\|_2^{p_i}}\Big(\boldsymbol{w}_i^{(j)}\Big) \qquad \text{for } \mathrm{DTV}_p,\ \mathrm{WDTV}_{\boldsymbol{p}}^{sv}, \tag{8.22}$$

with $\beta_i = \beta_g$, $p_i = p$, $a_i = a$, $\theta_i = \theta$ for space-invariant regularizers and $\beta_i = \beta_g/\alpha_i^{(j)}$, $p_i = p_i^{(j+1)}$, $a_i = a_i^{(j+1)}$, $\theta_i = \theta_i^{(j+1)}$ for space-variant regularizers, and with vector $\boldsymbol{w}^{(j)}$ defined in (8.12).

The $\boldsymbol{g}$-update computational cost thus depends on the functional form of the regularizer. In particular, the prox in (8.20) for TV and WTV regularizers admits a well-known closed-form expression (bivariate soft-thresholding), such that updating $\boldsymbol{g}$ in this case requires $O(N)$ operations. The prox in (8.21) for TV_p and $\mathrm{WTV}_p^{\mathrm{sv}}$ has been analyzed thoroughly in [101] and, very recently, in [103]. Depending on p_i and $w_i^{(j)}$ values, each of the N proximal maps in (8.21) either admits a closed-form solution or can be solved by finding the unique root of a nonlinear equation (nle) in one unknown via iterative approaches such as, e.g., the Newton–Raphson algorithm. Complexity of the $\boldsymbol{u}$-update step for TV_p and $\mathrm{WTV}_p^{\mathrm{sv}}$ can thus vary between a best $O(N)$ case (closed-form solution for any $i = 1, \dots, N$) and a worst $O(N\, n_{its})$ case (nle solution for any pixel $i = 1, \dots, N$), where n_{its} denotes the average number of iterations required by the algorithm used for solving the N nles. Finally, according to Lemma 8.1 and Corollary 8.2, the prox in (8.22) for DTV_p and $\mathrm{WDTV}_p^{\mathrm{sv}}$ always reduces to a one-dimensional constrained optimization problem to be solved iteratively. Updating $\boldsymbol{u}$ thus costs $O(N\, n_{its})$ operations in this last case.

The actual computational overhead brought about by adopting space-variant regularization lies in updating the vector $\boldsymbol{\Theta}$ of space-variant hyperparameters. Space-invariant regularizers, even if of complicated functional form, are typically characterized by a few global—we might say $O(1)$—free parameters which, when not set by hand or inferred once and for all from the entire observed degraded image $\boldsymbol{b}$, can be updated along iterations without significantly affecting the overall efficiency of the algorithm. Instead, even the simplest space-variant regularizers exhibit $O(N)$ free parameters which can hardly be set by hand, and also their one-shot neighborhood-based estimate from $\boldsymbol{b}$ is very likely to be unreliable due to the degradations in $\boldsymbol{b}$ and the necessarily small sample sizes. Hence, the $O(N)$ space-variant parameters must be estimated repeatedly along the iterations of the overall optimization algorithm, so that the cost per iteration of their update, which is at least in the $O(N)$ order, can significantly affect the efficiency of the overall algorithm.

In the following, we analyze in detail the computational efficiency of the $\boldsymbol{\Theta}$-update step for each pair of corresponding space-invariant and space-variant regularizers considered in this review work.

TV and WTV Regularizers. When standard TV is selected as regularizer, the $\boldsymbol{\Theta}$-update step does not arise. In fact, as already remarked on in section 6.3 and Table 2, the global scale parameter $\alpha \in \mathbb{R}_{++}$ in the TV Gibbs prior (3.8) can be merged to the regularization parameter μ, whose estimate is carried out in the $\boldsymbol{y}$-update step.

When the space-variant WTV counterpart of TV is used, in accordance with the

estimation procedure outlined in section 7.2, the update of the scale parameter vector $\boldsymbol{\alpha} \in \mathbb{R}^N_{++}$ admits the following componentwise closed-form expression:

$$\alpha_i^{(j+1)} = \frac{m}{\sum_{l \in \mathcal{J}_i^r} \|\boldsymbol{g}_l^{(j)}\|_2}, \quad i = 1, \dots, N. \tag{8.23}$$

As specified in section 7.2, such an update is realized based on fast 2D discrete transforms with complexity $O(N \log N)$ comparable to that of the $\boldsymbol{u}$-update step in its most favorable case (convolution matrix).

TV$_p$ and WTV$^{\text{sv}}_{\boldsymbol{p}}$ Regularizers. When selecting the TV$_p$ regularizer, similarly as for TV, the global scale parameter $\alpha \in \mathbb{R}_{++}$ can be merged to the regularization parameter μ (see Table 2), hence it does not need to be explicitly determined. The expression for the update of the global shape parameter $p \in \mathbb{R}_{++}$ can be easily obtained based on the derivations in section 7.3 by taking as a sample set $\mathcal{S}$ the one composed by the gradient magnitudes $\|\boldsymbol{g}_i^{(j)}\|_2$ of all the pixels in the image. We thus have

$$p^{(j+1)} \in \underset{p \in [\epsilon, R]}{\arg\min} \left\{ \frac{N}{p} \log \left(\frac{p}{n} \sum_{i=1}^{N} \|\boldsymbol{g}_i^{(j)}\|_2^p \right) + N \ln \Gamma \left(1 + \frac{1}{p} \right) + \frac{N}{p} \right\}. \tag{8.24}$$

At each iteration j, the one-dimensional minimization problem (8.24) must be solved iteratively, with complexity $O(N\, n_{its})$.

When using the space-variant WTV$^{sv}_{\boldsymbol{p}}$ regularizer, both the vectors $\boldsymbol{\alpha}, \boldsymbol{p} \in \mathbb{R}^N_{++}$ of scale and shape parameters must be updated, according to the estimation procedures outlined in section 7.3. In particular, we have

$$p_i^{(j+1)} \in \underset{p_i \in [\epsilon, R]}{\arg\min} \left\{ \frac{m}{p_i} \log \left(\frac{p_i}{m} \sum_{l \in \mathcal{J}_i^r} \|\boldsymbol{g}_l\|_2^{p_i} \right) + m \ln \Gamma \left(1 + \frac{1}{p_i} \right) + \frac{m}{p_i} \right\}, \tag{8.25}$$

$$\alpha_i^{(j+1)} = \left(\frac{m}{p_i^{(j+1)} \sum_{l \in \mathcal{J}_i^r} \|\boldsymbol{g}_l^{(j)}\|_2^{p_i^{(j+1)}}} \right)^{1/p_i^{(j+1)}}, \quad i = 1, \dots, N. \tag{8.26}$$

The $\boldsymbol{p}$-update in (8.25) has complexity $O(N\, m\, n_{its})$, with an $O(m)$ overhead with respect to its space-invariant counterpart (8.24). The $\boldsymbol{\alpha}$-update in (8.26), like its analogue in (8.23) for the WTV regularizer, has complexity $O(N \log N)$.

DTV$_p$ *and* WDTV$^{\text{sv}}_{\boldsymbol{p}}$ Regularizers. Like for TV and TV$_p$, the global scale parameter $\alpha \in \mathbb{R}_{++}$ of the DTV$_p$ regularizer does not need to be estimated, whereas the scale parameter vector $\boldsymbol{\alpha} \in \mathbb{R}_{++}$ of its space-variant counterpart WDTV$^{sv}_{\boldsymbol{p}}$, in accordance with derivations in section 7.4, can be updated in closed form as follows:

$$\alpha_i^{(j+1)} = \left(\frac{2m}{p_i^{(j+1)} \sum_{l \in \mathcal{J}_i^r} \|\boldsymbol{\Lambda}_{a_i^{(j+1)}} \mathbf{R}_{-\theta_i^{(j+1)}} \boldsymbol{g}_l^{(j)}\|_2^{p_i^{(j+1)}}} \right)^{1/p_i^{(j+1)}}, \quad i = 1, \dots, N. \tag{8.27}$$

The $\boldsymbol{\alpha}$-update step above has the same $O(N \log N)$ complexity as those in (8.23) for WTV and (8.26) for WTV_p^{sv}.

For both the space-invariant DTV_p and space-variant $\text{WDTV}_p^{\text{sv}}$ regularizers, the triplets of shape, orientation, and anisotropicity parameters are updated jointly by solving (iteratively) constrained three-dimensional minimization problems of the form in (7.13), with function G defined in (7.11). The computational complexity is the same as the one for updating the shape parameters of the TV_p and WTV_p^{sv} regularizers by (8.24) and (8.26), respectively, namely, $O(N\, n_{its})$ for DTV_p and $O(N\, m\, n_{its})$ for $\text{WDTV}_p^{\text{sv}}$. However, as will be shown next, the computation time required to update the parameter triplet for DTV_p and $\text{WDTV}_p^{\text{sv}}$ is much higher than that to update the shape parameter for TV_p and WTV_p^{sv}. This is clearly due to the higher time-per-iteration spent by any iterative method to solve a three-dimensional compared to a one-dimensional problem.

We conclude this section by summarizing in Table 3 the numerical steps hidden behind the $\boldsymbol{g}$- and $\boldsymbol{\Theta}$-updates for the different space-invariant and space-variant regularization terms considered so far and that will be tested in the following experimental section. In the table, we also report the CPU times (in seconds) of a single iteration of Algorithm 8.1 employed for the restoration of a 256×256 image corrupted by Gaussian blur and AIG noise. The experiments have been performed under Windows 10 in MATLAB R2019 on an ASUS PC with an Intel Core i7 CPU 6500U @2.50GHz processor and 8GB of RAM. For the estimate of the parameters in the space-variant scenario, we used neighborhoods of radius $r = 6$.

One can immediately observe that depending on the regularizer under consideration, adding space variance affects computational efficiency differently. For instance, when considering the TV_p regularizer, denoting by M the number of $\boldsymbol{g}_i$ subproblems that can be solved via a closed-form expression, the solution of the $N - M$ nonlinear equations required to update the remaining entries of $\boldsymbol{g}$ significantly slows down the computations, while the performance of the space-variant WTV is comparable with the one of the TV, given the closed-form expression for the update of both $\boldsymbol{\alpha}$ and $\boldsymbol{g}$. The estimation of the space-variant parameters $\boldsymbol{\alpha}$ and the $\boldsymbol{p}$ for the $\text{WTV}_{\boldsymbol{p}}^{sv}$ regularizer does not drastically increase the computational time when compared to its global counterpart. However, the computational time significantly increases for the highly parametric $\text{WDTV}_{\boldsymbol{p}}^{sv}$ regularizer, as it lacks efficient procedures for the update of the triplet (p_i, θ_i, a_i). In our simulations, such a problem has been performed by means of standard constrained programming algorithms (implemented in the `fmincon` MATLAB routine). As anticipated at the end of section 8.4, in order to mitigate the computational burden provided by such an estimation step, we will thus perform the update of the space-variant parameters for the $\text{WDTV}_{\boldsymbol{p}}^{sv}$ every few (typically 30) iterations of Algorithm 8.1.

9. Applications to Image Restoration. In this section, we evaluate the performances of the space-variant regularizers discussed so far, namely the WTV, the $\text{WTV}_{\boldsymbol{p}}^{sv}$, and $\text{WDTV}_{\boldsymbol{p}}^{sv}$ regularizers in comparison with the space-invariant TV [134] and TV_p [101] regularizers. As an example, we will consider the problem of image deblurring, for which the forward linear operator $\mathbf{A} \in \mathbb{R}^{N \times N}$ in (2.7) models the action of a space-invariant blur kernel.

Test Images, Quality Measures, and Parameters. In order to highlight the flexibility of the space-variant approach described in this work, the regularizers of interest will be tested on the restoration of images characterized by different global and local properties. More specifically, we will consider the `geometric` image in Figure

Table 3 *Computational solvers used for the $\boldsymbol{\Theta}$- and $\boldsymbol{g}$-updates in Algorithm 8.1 for the space-invariant (top) and space-variant (bottom) regularization terms considered so far, together with the corresponding CPU times (in seconds) per iteration.*

	$\mathcal{R}(\boldsymbol{u},\boldsymbol{\Theta})$	$\boldsymbol{\alpha}$-update	$\boldsymbol{p}$-update	$\boldsymbol{\theta}$-update	$\boldsymbol{a}$-update	$\boldsymbol{g}$-update	Time (s)
space-invariant	TV	-	-	-	-	closed-form (8.20)	0.025
	TV_p	-	1 1D min pb (grid-search) (8.24)	-	-	M closed-form + $(N-M)$ nle (Newton–Raphson) (8.21)	1.245
space-variant	WTV	closed-form (8.23)	-	-	-	closed-form (8.20)	0.034
	$\mathrm{WTV}_{\boldsymbol{p}}^{sv}$	closed-form (8.26)	N 1D min pbs (grid-search) (8.25)	-	-	M closed-form + $(N-M)$ nle (Newton–Raphson) (8.21)	2.759
	$\mathrm{WDTV}_{\boldsymbol{p}}^{sv}$	closed-form (8.27)	N 3D min pbs (MATLAB routine fmincon) (7.13)			N 1D min pbs (grid-search) (8.22)	41.742

18(a), which is purely piecewise constant, the `skyscraper` image in Figure 18(b), which presents a mixture of piecewise constant, piecewise linear, and textured features, the `stairs` image in Figure 18(c), which is highly textured with fine oriented details, and the `heart` medical image in Figure 18(d). The test images `geometric`, `skyscraper`, and `stairs` have been corrupted by 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 continuous settings. The test image `heart` has been corrupted by a space-invariant motion blur defined by a convolution kernel generated, as before, using the MATLAB routine `fspecial` with parameters `length`=10 and `theta`=45, the former representing the length of the motion, the latter its direction.

Then, the blurred images have been degraded by AIGG noise realizations from different distributions with standard deviation $\sigma = 0.1$. More specifically, we considered $q = 1$ (Laplace noise) for the `geometric` test image, $q = 2$ (Gaussian noise) for the `skyscraper` and the `heart` test images, and $q = +\infty$ (uniform noise) for the `stairs` test image. The blur- and noise-corrupted images are displayed on the bottom row of Figure 18.

The quality of the obtained restorations $\boldsymbol{u}^*$ vs. the associated ground-truth image $\boldsymbol{u}$ is assessed by means of two scalar measures: the improved signal-to-noise ratio (ISNR)

$$\mathrm{ISNR}(\boldsymbol{b};\boldsymbol{u};\boldsymbol{u}^*) := 10\log_{10}\left(\frac{\|\boldsymbol{b}-\boldsymbol{u}\|_2^2}{\|\boldsymbol{u}^*-\boldsymbol{u}\|_2^2}\right),$$

and the structural similarity index (SSIM) [149]. The larger the ISNR and SSIM values, the higher the restoration quality. For all tests, the ADMM iterations are stopped as soon as

$$\eta^{(j+1)} := \frac{\|\boldsymbol{u}^{(j+1)}-\boldsymbol{u}^{(j)}\|_2}{\|\boldsymbol{u}^{(j)}\|_2} < 10^{-5}, \quad j \in \mathbb{N}.$$

The penalty parameters β_g, β_r are manually set so as to facilitate the convergence of the overall alternating scheme. Typically, suitable values are $\beta_g, \beta_r \approx 10^4, 10^5$.

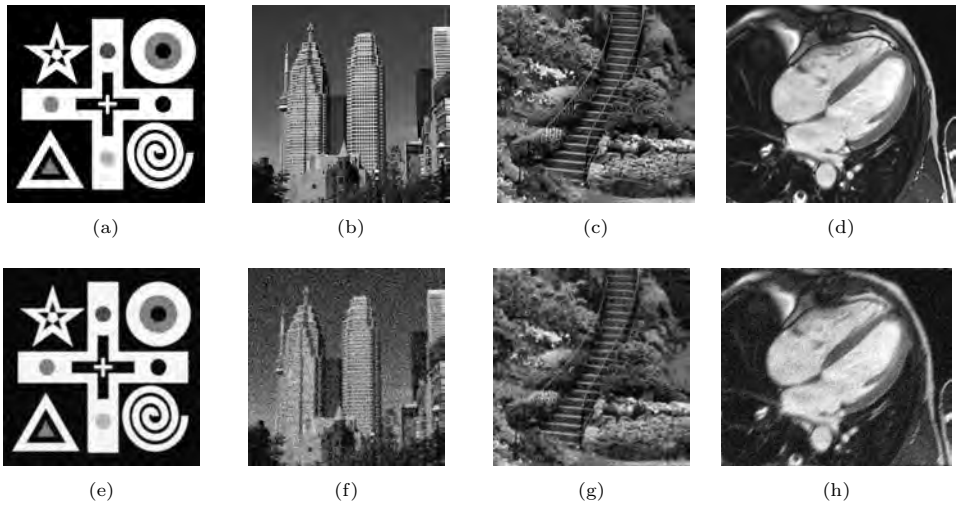


Fig. 18 *Original test images* `geometric` *(*320×320*),* `skyscraper` *(*256×256*),* `stairs` *(*320×320*), and* `heart` *(*378×442*) (top), and observed data corrupted by blur and different AIGG (bottom).*

The estimation of the hyperparameters in the space-variant regularizers WTV, $\text{WTV}_{\boldsymbol{p}}^{sv}$, and $\text{WDTV}_{\boldsymbol{p}}^{sv}$ is performed by manually setting the radius r, so as to attain the highest ISNR and SSIM values. Moreover, for the WTV regularizer, the existence of a very efficient procedure for the computation of the $\{\alpha_i\}_i$ weights allows us to update the $\boldsymbol{\alpha}$-map at each iteration of the ADMM-based scheme; in order to hold back the computational effort coming along with the estimation of the unknown $\{\alpha_i, p_i\}_i$ in the $\text{WTV}_{\boldsymbol{p}}^{sv}$ and $\{\alpha_i, p_i, \theta_i, a_i\}_i$ in the $\text{WDTV}_{\boldsymbol{p}}^{sv}$ regularizer, we update the maps of parameters every 30 iterations.

For what concerns the estimation of the local p_i for the $\text{WTV}_{\boldsymbol{p}}^{sv}$ and the $\text{WDTV}_{\boldsymbol{p}}^{sv}$, as well as the estimation of the global p in the TV_p regularizer, we fix the compact set $[\epsilon, R]$ of Propositions 7.2 and 7.3 equal to $[0.5, 2]$. Notice that the choice of the lower bound allows the $\boldsymbol{u}$-estimation problem (8.1) to result in nonconvex regularizers. This implies that particular attention has to be given to the design of a suitable initial guess, which can prevent the performed hypermodels from getting trapped in bad local minima.

We initialize Algorithm 8.1 using a suitable initialization minimizing noise whiteness for a standard Tikhonov-L_2 problem, as proposed recently in [102].

Restoration of `geometric`. First, we discuss the performance of the considered regularizers for the restoration of the `geometric` test image. The restored images are shown in Figure 19, while the achieved ISNR and SSIM values are reported in Table 4. Notice that in general, the TV regularizer is well-suited for the restoration of piecewise-constant images; however, as discussed in section 1.2, it also suffers from several drawbacks. Our results confirm that using instead a TV_p regularizer ($p = 0.5$) reduces such artifacts. Overall, the three considered space-variant regularizers appear to be more effective than plain TV.

In Figure 20, we show the output maps of parameters for the WTV, $\text{WTV}_{\boldsymbol{p}}^{sv}$, and $\text{WDTV}_{\boldsymbol{p}}^{sv}$ regularizers, obtained with $r = 1$, $r = 3$, and $r = 1$, respectively. For all three regularizers, the $\boldsymbol{\alpha}$-maps present higher weights in the background, while showing that weaker regularization is performed along the profiles of the geometrical

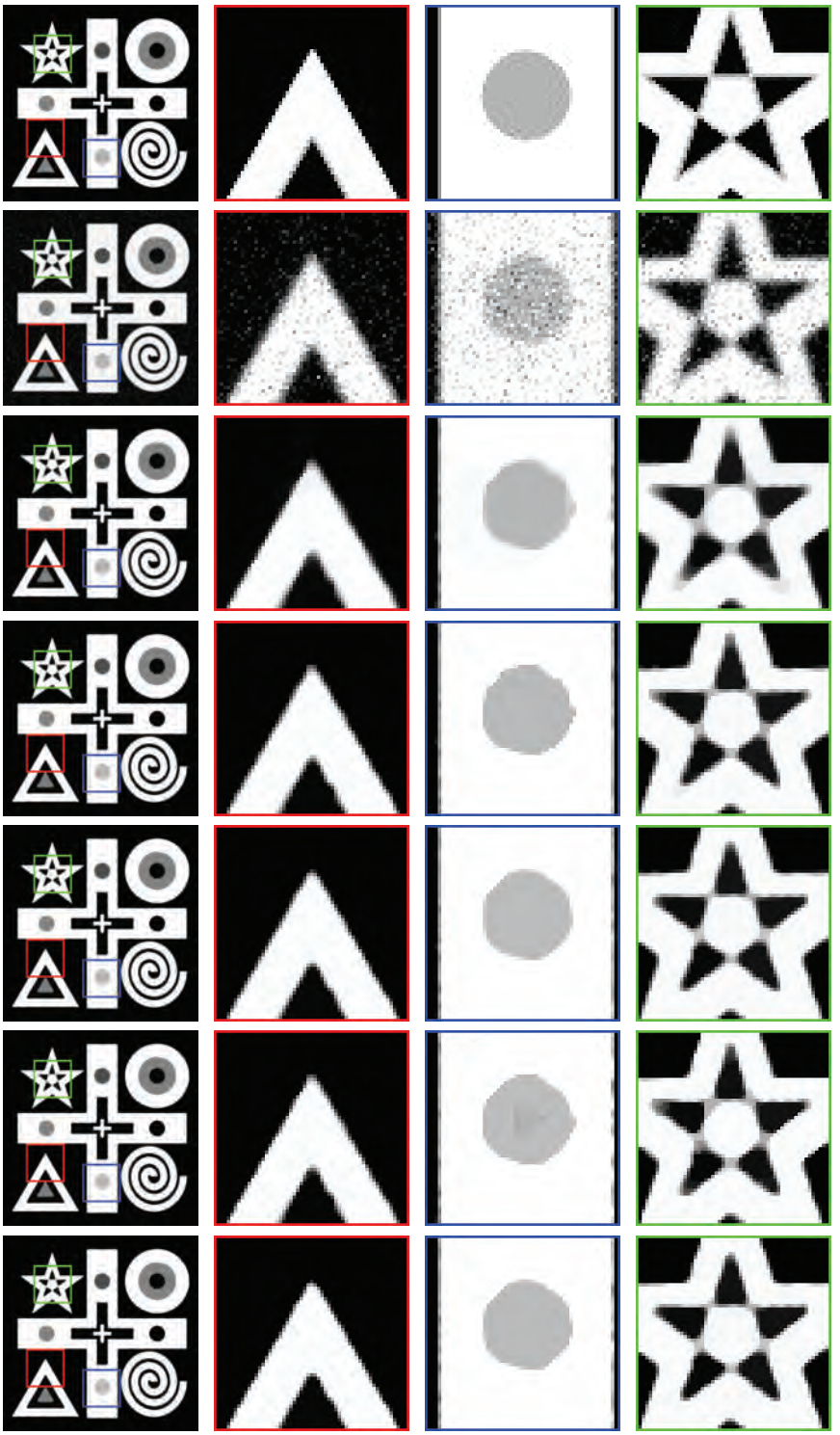


Fig. 19 *From top to bottom: for the test image* `geometric`*, original image, observed image b, performance of the TV, the TV_p (with output $p = 0.5$), the WTV, the $WTV_{\boldsymbol{p}}^{sv}$, and the $WDTV_{\boldsymbol{p}}^{sv}$ regularizers with the respective close-up(s).*

Table 4 *ISNR and SSIM values achieved by the considered regularizers for the four test images corrupted by blur and different AIGG noises.*

	TV	TV_p	WTV	$\mathrm{WTV}_{\boldsymbol{p}}^{sv}$	$\mathrm{WDTV}_{\boldsymbol{p}}^{sv}$
`geometric`					
ISNR	8.8499	9.0568	9.5567	9.6041	10.2188
SSIM	0.9227	0.9225	0.9343	0.9346	0.9388
`skyscraper`					
ISNR	2.3239	2.5775	2.7906	2.9894	3.2083
SSIM	0.6255	0.6432	0.6711	0.6789	0.7166
`stairs`					
ISNR	3.9417	4.5251	4.6836	5.0718	5.2031
SSIM	0.6515	0.6912	0.6879	0.7149	0.7307
`heart`					
ISNR	6.8080	7.0698	7.5251	8.0261	8.6283
SSIM	0.7494	0.7427	0.7529	0.7762	0.7841

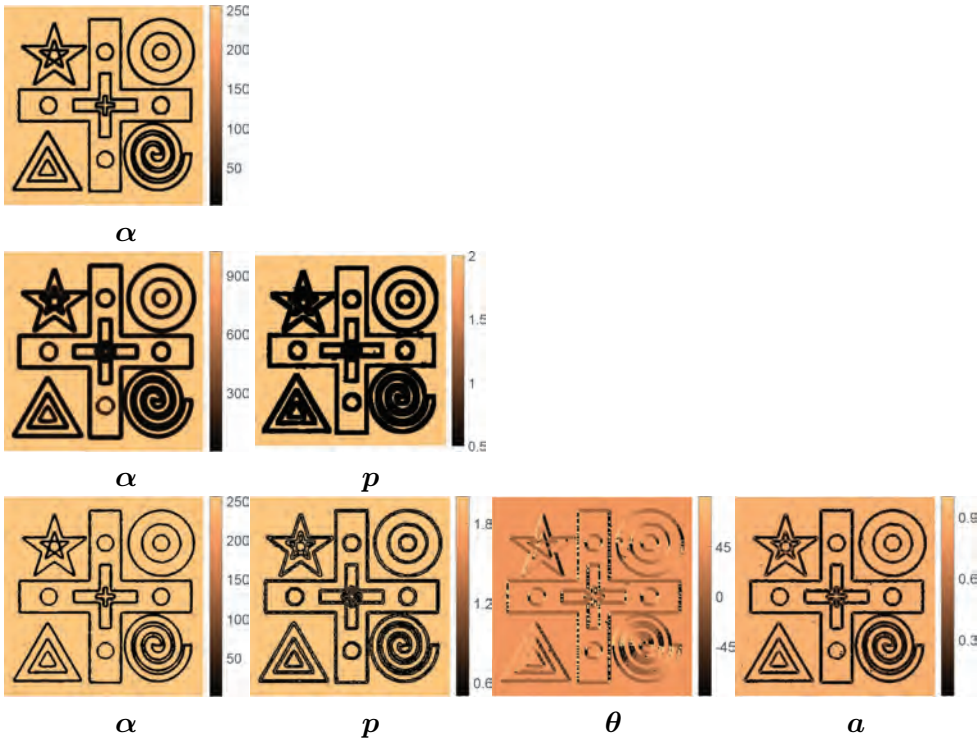


Fig. 20 *From top to bottom: for the test image* `geometric`*, output maps of the parameters for the* WTV *($r = 1$), the* $\mathrm{WTV}_{\boldsymbol{p}}^{sv}$ *($r = 3$), and the* $\mathrm{WDTV}_{\boldsymbol{p}}^{sv}$ *($r = 1$) regularizers.*

figures. Notice that the $\boldsymbol{p}$-values in the $\mathrm{WTV}_{\boldsymbol{p}}^{sv}$ and in the $\mathrm{WDTV}_{\boldsymbol{p}}^{sv}$ approach 2 in the background, which combined with the high regularization weights allows for an effective smoothing and noise removal therein. Finally, the $\boldsymbol{\theta}$ and $\boldsymbol{a}$ maps in the bottom row of Figure 20 show that the estimator detects a clear directionality in correspondence with the figure profiles, where the angles $\boldsymbol{\theta}$ have been accurately estimated and $\boldsymbol{a}$ assume small values.

Restoration of `skyscraper`. We now consider the restoration of the test image `skyscraper`, which, due to its composite nature, is expected to largely benefit from a space-variant approach. From the restored images and the selected details in Figure 21, one can clearly notice how each additional space-variant parameter effectively contributes to gradually improving the output result, as also reflected in the ISNR and SSIM values reported in Table 4. In Figure 22, we show the output map of parameters for the WTV, $\mathrm{WTV}^{sv}_{\boldsymbol{p}}$, and $\mathrm{WDTV}^{sv}_{\boldsymbol{p}}$ regularizers, computed for $r = 15$, $r = 15$, and $r = 3$, respectively.

As a general comment, we highlight that the weights α_i assume larger values on the background so that a strong regularization is performed regardless of the corresponding p_i; in fact, the $\boldsymbol{p}$-maps for the $\mathrm{WTV}^{sv}_{\boldsymbol{p}}$ and the $\mathrm{WDTV}^{sv}_{\boldsymbol{p}}$ regularizers appear to be different in this region. From the $\boldsymbol{\theta}$-map reported in the bottom row of Figure 22, we observe that also in this case the estimator is capable of detecting the direction of the buildings' profiles as well as the horizontal oriented texture. Finally, the $\boldsymbol{a}$ values indicate a stronger dominance in terms of directionality along the edges of the buildings.

Restoration of `stairs`. In this third test, we consider the highly textured image `stairs`. From the ISNR and SSIM values reported in Table 4 and from the restored images displayed in Figure 23, we notice that the WTV regularizer is outperformed by the space-invariant TV_p regularizer, with output estimated at $p = 1.56$, in terms of SSIM. In fact, the TV_p performs a type of regularization which, although global, appears to be more suitable for describing the image of interest. Further improvement is achieved by the $\mathrm{WTV}^{sv}_{\boldsymbol{p}}$ regularizer, which preserves the textured regions in the image while smoothing out the limited piecewise-constant parts. Finally, the $\mathrm{WDTV}^{sv}_{\boldsymbol{p}}$ regularization term slightly refines the output by driving the regularization along the local directionalities.

The maps of the parameters for the space-variant WTV with radius $r = 2$, and for $\mathrm{WTV}^{sv}_{\boldsymbol{p}}$ and $\mathrm{WDTV}^{sv}_{\boldsymbol{p}}$ with radius $r = 1$ are shown in Figure 24. From the $\boldsymbol{p}$-maps for $\mathrm{WTV}^{sv}_{\boldsymbol{p}}$ and $\mathrm{WDTV}^{sv}_{\boldsymbol{p}}$, one can observe that values of p_i equal or close to 2 are spread out all over the image, thus indicating that a Tikhonov-type of regularization, combined with the suitable local weights, is more effective in dealing with this sort of image, due to the presence of large textured regions where the distribution of gradients is thus very spread out. The directions in the central part of the image are precisely detected, as shown in the $\boldsymbol{\theta}$-map, as well as the confidence in the estimation, represented by $\boldsymbol{a}$, which appear to be particularly relevant along the steps.

Restoration of `heart`. In this last test, we address the restoration of the medical image `heart`. The ISNR and SSIM values achieved by the employed space-variant and space-invariant regularization terms are reported in Table 4, while the output restorations are shown in Figure 25. Notice that in this case, the TV and the TV_p, with output p-value $p = 0.65$, return similar quality measures, with the TV outperforming the TV_p in terms of SSIM. Although performing a TV-type of regularization, the locally adapting regularization weights in the WTV lead to a mitigated staircasing effect on the smooth features of the image. As in the previous example, the flexibility allowed by the local shape parameters p in the $\mathrm{WTV}^{sv}_{\boldsymbol{p}}$ yields a very high-quality restoration, from both visual and metric viewpoints. Finally, the local directional information encoded in the $\mathrm{WDTV}^{sv}_{\boldsymbol{p}}$ contributes to recovering finer and oriented details.

In Figure 26, we show the output maps for the WTV with $r = 8$, for the $\mathrm{WTV}_{\boldsymbol{p}^{sv}}$ with $r = 7$, and for the $\mathrm{WDTV}^{sv}_{\boldsymbol{p}}$ with $r = 3$. The $\boldsymbol{p}$-maps reveal that a Tikhonov-

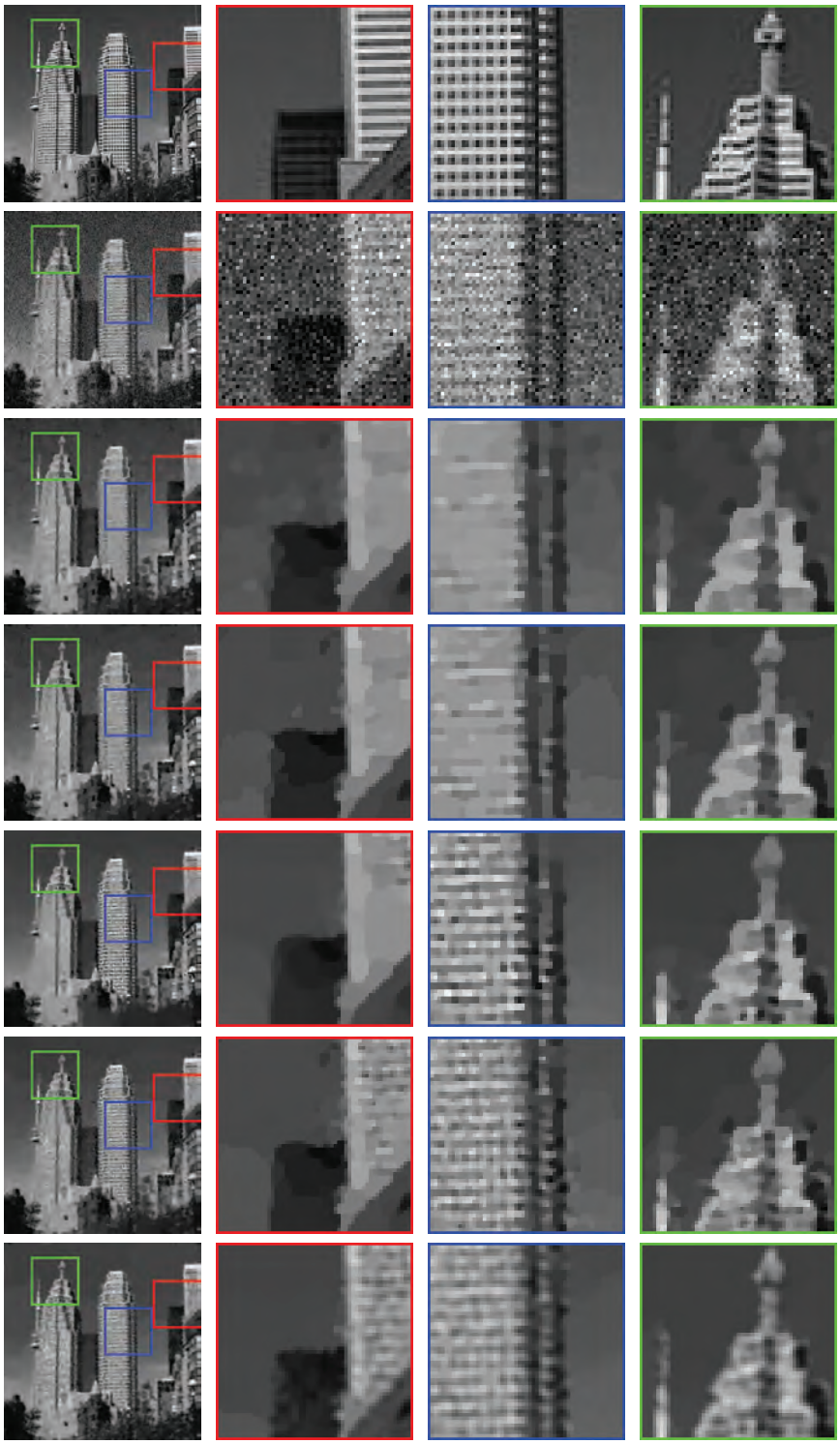


Fig. 21 *From top to bottom: for the test image* `skyscraper`*, original image, observed image b, performance of the* TV*, the* TV_p *(with output* $p = 0.5$*), the* WTV*, the* $\mathrm{WTV}^{sv}_{\boldsymbol{p}}$*, and the* $\mathrm{WDTV}^{sv}_{\boldsymbol{p}}$ *regularizers with the respective close-up(s).*

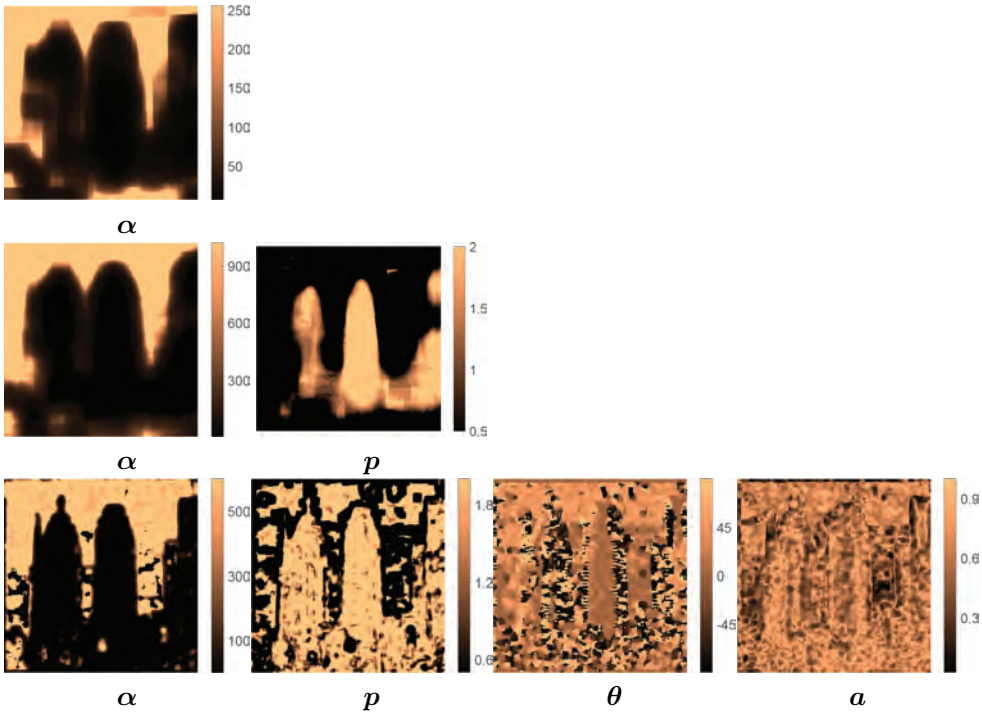


Fig. 22 *From top to bottom: for the test image* `skyscraper`*, output maps of the parameters for the* WTV *($r = 15$),* $\text{WTV}_{\boldsymbol{p}}^{sv}$ *($r = 15$), and* $\text{WDTV}_{\boldsymbol{p}}^{sv}$ *($r = 3$) regularizers.*

type of regularization is preferred to recover the predominant smooth features. Also, notice that the larger p-values on the flat background are accompanied by larger regularization weights, so that the noise in the restorations computed by $\text{WTV}\boldsymbol{p}^{sv}$ and $\text{WDTV}\boldsymbol{p}^{sv}$ is removed.

10. User Guide to Space Variance. We conclude this work by addressing several issues pointed out so far, with the purpose of guaranteeing a more conscious use of the derived hypermodels.

1. *The best space-adaptive regularizer.* As one could expect, the question about which regularizer performs *best* in absolute terms does not have a unique answer. The adoption of more and more general regularization terms does not always pay, as the overall performance has to be evaluated with respect to the trade-off between quality of the restorations and computational effort. From this perspective, on the restoration of the `geometric` and `skyscraper` images, the WTV has returned remarkably good results while keeping the computational times low, in light of the closed-form expressions existing for the $\boldsymbol{\alpha}$-update and for the $\boldsymbol{g}$-subproblem in the ADMM-based scheme. However, for the `stairs` image, the typical shortage of a TV-type regularization, even if weighted, has emerged.
   The selection of the regularizer to employ should thus be motivated by the application of interest and, ultimately, by the processed data.
2. *The optimal radius r.* In the previous tests, the radius r involved in the estimate of the parameters has been selected so as to maximize the ISNR and

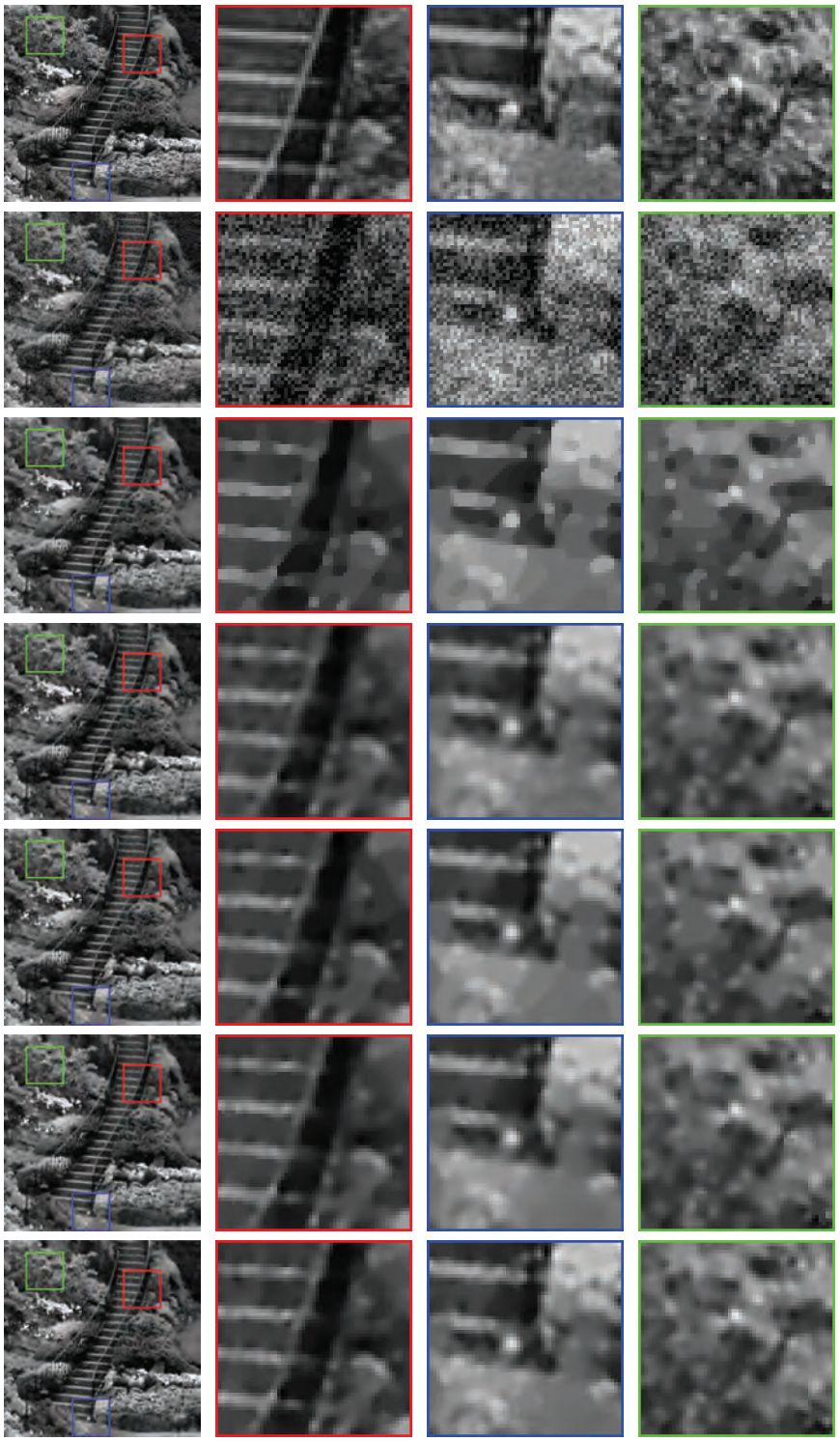


Fig. 23 *From top to bottom: for the test image* `stairs`*, original image, observed image b, performance of the* TV*, the* TV_p *(with output* $p = 1.56$*), the* WTV*, the* $\mathrm{WTV}^{sv}_{\boldsymbol{p}}$*, and the* $\mathrm{WDTV}^{sv}_{\boldsymbol{p}}$ *regularizers with the respective close-up(s).*

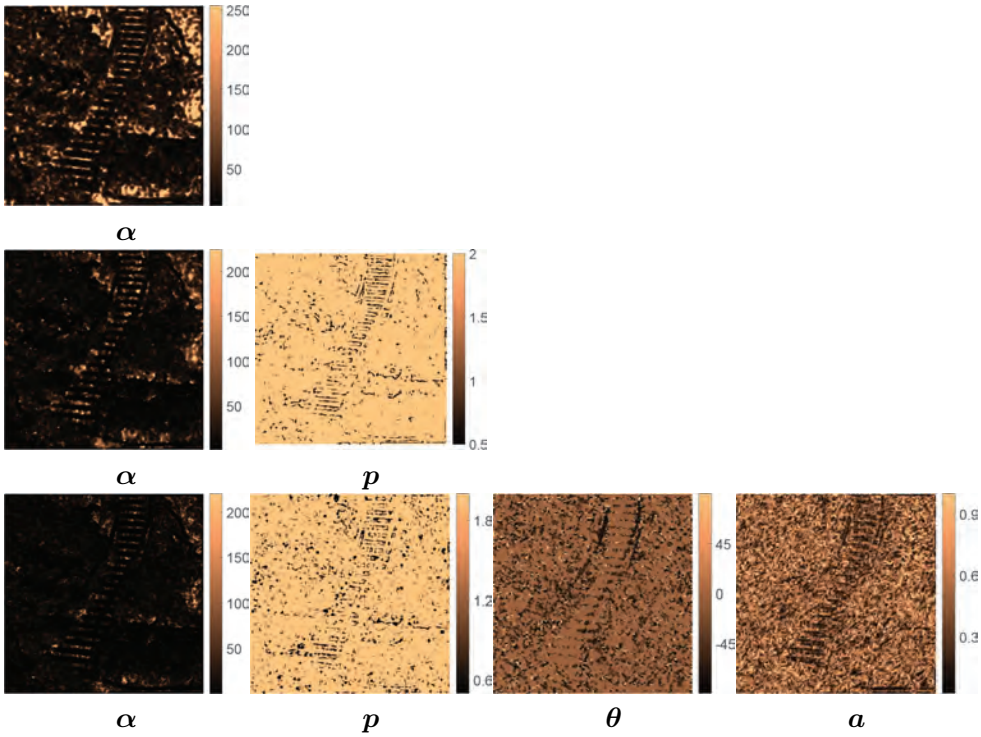


Fig. 24 *From top to bottom: for the test image* `stairs`*, output maps of the parameters for the* WTV *($r = 2$),* $\mathrm{WTV}_{\boldsymbol{p}}^{sv}$ *($r = 1$), and* $\mathrm{WDTV}_{\boldsymbol{p}}^{sv}$ *($r = 1$) regularizers.*

the SSIM of the final restorations. One can notice that its choice somehow reflects the scale of the structures to preserve, and that it can be set differently for different space-variant regularizers. In this sense, a useful example is given by the test image `skyscraper` which presents textured objects in the foreground with a smooth and constant background. For the WTV and the $\mathrm{WTV}_{\boldsymbol{p}}^{sv}$ regularizers, a large value of r can easily catch the "dual" nature of the image. However, the texture on the foreground buildings is fine-scale so that to detect the local directionalities with the $\mathrm{WDTV}_{\boldsymbol{p}}^{sv}$ regularizer, a smaller radius has to be selected.

3. *The curse of nonconvexity.* The convergence of the outlined numerical scheme aimed at solving a possibly nonconvex problem interlaced with a parameter estimation step is a very delicate issue that has not found a theoretical response yet. However, when the ADMM penalty parameters β_g, β_r are set in a suitable manner—typically, $\beta_g, \beta_r \approx 10^4, 10^5$—empirical convergence is observed.

11. Challenges. We conclude this review by listing in the following some challenging future research directions which could enrich this work from both theoretical and applied points of view. Each of the following items has to be intended not as a straightforward extension of the framework presented here, but rather as an intersection with some related mathematical fields (analysis, optimization, numerical and linear algebra, medical imaging, etc.) favoring the development of new and unexplored research.

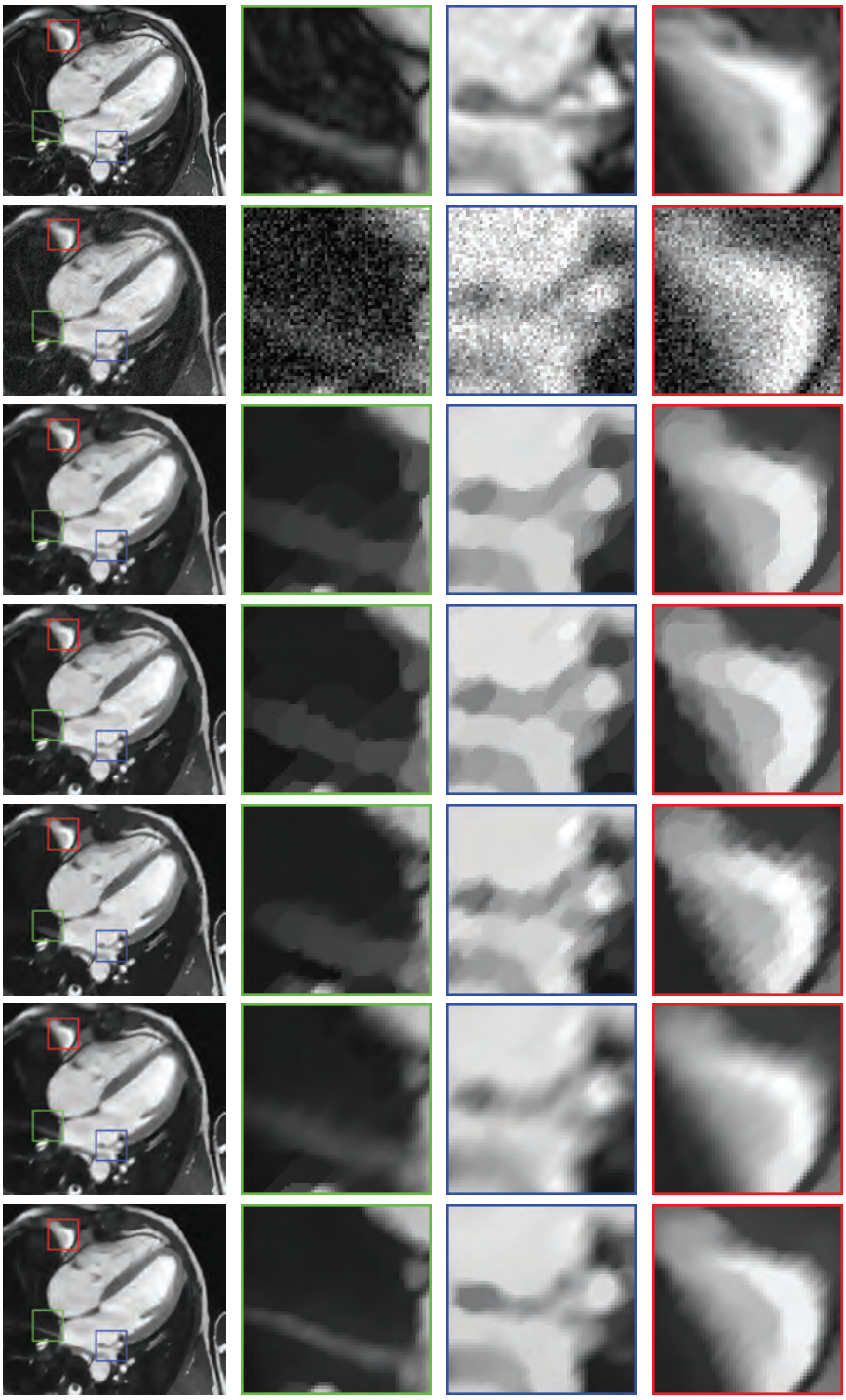


Fig. 25 *From top to bottom: for the test image* `heart`*, original image, observed image b, performance of the* TV*, the* TV_p *(with output* $p = 0.65$*), the* WTV*, the* $\mathrm{WTV}_{\boldsymbol{p}}^{sv}$*, and the* $\mathrm{WDTV}_{\boldsymbol{p}}^{sv}$ *regularizers with the respective close-up(s).*

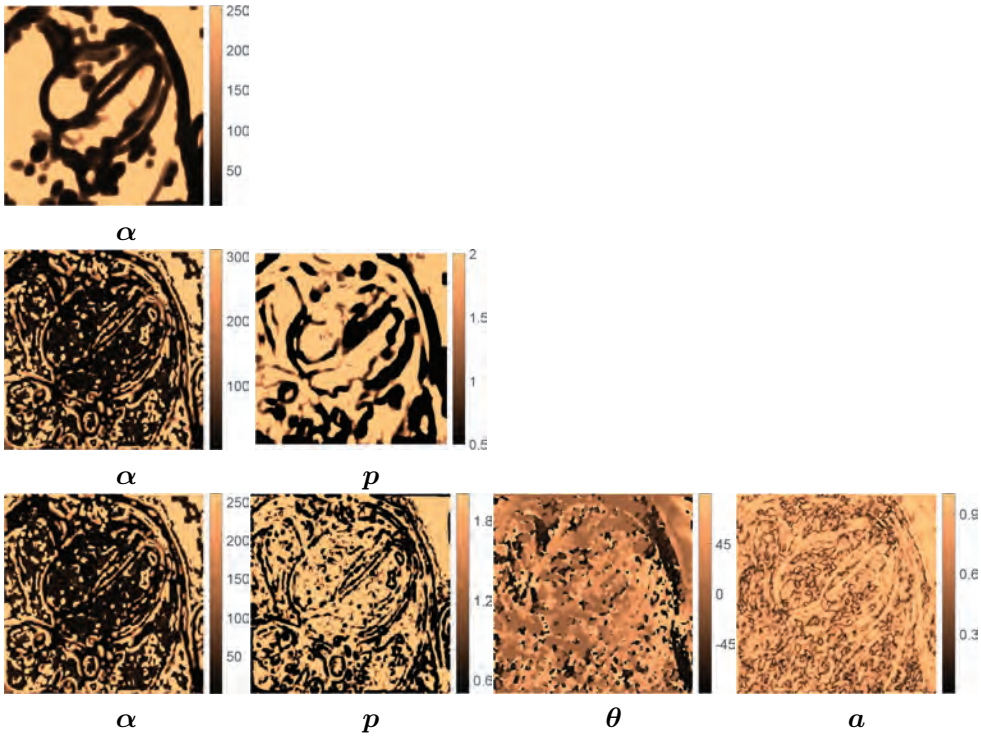


Fig. 26 *From top to bottom: for the test image* `heart`*, output maps of the parameters for the* WTV *($r = 8$),* $\mathrm{WTV}^{sv}_{\boldsymbol{p}}$ *($r = 7$), and* $\mathrm{WDTV}^{sv}_{\boldsymbol{p}}$ *($r = 3$) regularizers.*

1. The detailed analytical study in an infinite-dimensional framework of the nonsmooth, nonconvex, and space-variant regularization models discussed in this work is expected to provide more insights into the structure of the expected solutions by means of duality tools, functional calculus, and nonstandard Lebesgue/Sobolev calculus in spaces with variable exponents; see, e.g., [51] for the convex case.
2. The development of a rigorous theoretical framework guaranteeing convergence to (at least) stationary points for the nonconvex ADMM Algorithm 8.1 and of the alternating scheme (6.11)–(6.12) is highly nontrivial and practically made challenging due to the parameter estimation performed jointly along the iterations. Note that even for the IAS algorithm [31] where parameters and iterates are updated sequentially, a convergence proof in general nonconvex scenarios is still missing, as only partial results in convex (quadratic) cases are available.
3. The use of a similar space-variant modeling for more general regularizers defined, for instance, in terms of wavelet expansions [37] and higher-order differential operators (see, e.g., [107]).
4. Similarly as for the case of anisotropic diffusion, we expect that the use of suitable adaptive discretization stencils [42, 70, 57, 46] built on the estimated local directional information and/or relying on the suitable definition of appropriate transfer operators and staggered grids [126] could improve upon the quality of the numerical reconstructions by describing anisotropy on the

image grid in a more precise way.

5. Following [33], we believe that the use of a hierarchical modeling with *informative* hyperpriors encoding, for instance, local smoothness/sparsity beliefs on the solution could refine the hyperparameter selection strategy and thus, overall, the quality of the reconstruction.
6. The comparison of the proposed ML-type parameter estimation procedure described in section 7 with the recent approach based on empirical Bayes estimation proposed in [147, 58] could lead to new hybrid hyperparameter selection strategies relying on the sole observation of the given corrupted image $\boldsymbol{b}$. These ideas are expected indeed to speed up the performance of Algorithm 8.1.
7. As pointed out in section 7 and highlighted in section 9, a crucial choice strongly affecting both the parameter estimation and image restoration results is the radius r of the patches used to estimate locally the hyperparameters in play. Following previous work in the context of image completion [63, 160], we also believe that such a choice could be made *adaptive* in order to adjust itself to the local image content (cartoon vs. texture).
8. In light of the analysis carried out in section 8.5, we remark that the study of tailored optimization methods for the parameter estimate can facilitate the employment of the $\mathrm{WTV}^{sv}_{\boldsymbol{p}}$ and of the $\mathrm{WDTV}^{sv}_{\boldsymbol{p}}$ regularization terms, both penalized by the lack of closed-form expressions for the update of the space-variant parameters involved in their expression.
9. As an obvious field of application due to the recent use of analogous models in medical imaging problems such as MRI, PET, and CT applications [67, 66, 65, 146], we expect that the use of structural and adaptive modeling could significantly improve the quality of the reconstructions and favor, at the same time, the exploitation of structural information in multimodal image analysis. Moreover, as preliminarily shown in [3], we expect that a space-variant modeling based, e.g., on a local estimation of the cartoon vs. texture components could improve significantly also the performance of image analysis and, in particular, image segmentation algorithms.
10. We wonder how the flexible statistical modeling proposed in this work compares with the results obtained by VAEs [92] and GANs [75]. It would be interesting to compare qualitatively (and/or quantitatively, provided that a good quality measure is used; see, e.g., [19]) the highly parametric model-driven BGGD-type distribution estimated by means of the proposed approaches with the one estimated by means of generative approaches. In order to combine model- and data-driven approaches in an interpretable way, an alternative way of estimating the hyperparameters in (6.11) by means, e.g., of unrolling approaches [116] shall be considered.

12. Conclusions. In this work, we described a journey across time and various fields of applied mathematics with the intent of reviewing the many features of the exemplar and probably *the* most popular image regularization model over the last 30 years, the TV functional. After recalling its genesis, its main features, and shortcomings in section 1 and having fixed some notation in section 2, we described in section 3 how the rigidity of existing TV-type image regularization models can be overcome within the setting of nonstationary Markov random fields whose capability of describing local image features (i.e., scale, shape, and directionality) endows the corresponding prior distributions with more flexibility and degrees of freedom. An-

alytically, we show in section 4 that these nonstationary priors can be put in close correspondence to a large class of space-variant image regularization models which have been thoroughly studied over the last decades with the intent of improving upon well-known TV drawbacks. In their analytical form, the dependence of these models on local information (i.e., amount of regularization, sharpness, and anisotropy) is then shown in section 5 to correspond geometrically to changing at each pixel the definition of the constraint set of dual functions and changing their alignment accordingly. Having provided an expression of the corresponding data models in section 6, we then describe in section 7 a maximum-likelihood type automatic parameter estimation strategy, motivated by the underlying Bayesian formulation resulting in the definition of appropriate variational Bayesian hypermodels. Finally, the joint statistical-analytical procedure is embedded into a general alternating minimization scheme in section 8 and validated in section 9 on some exemplar image restoration models. The flexibility of the proposed approach and the accuracy and robustness of the estimator considered for the automatic selection of hyperparameters show good adaptation to both geometrical and texture image information and pave the way to challenging new research directions as finally described in section 11.

The incredible potential offered nowadays by the possibility of combining different fields of applied mathematics with the intent of improving and making more data-adaptive the performance of TV makes the use of such often taught, out-of-date image regularization models still interesting for the whole applied mathematics community. The swan song of TV is still far off. Due to both the profound understanding of this powerful, yet simple, image regularization model carried out over years and the recent advances in large-scale data exploitation and numerical optimization favoring the development of its many extensions, we expect that the descendant models stemming from TV still have much to say and could at the same time enrich and be enriched by the increasingly popular interest towards data-driven approaches shown by the analytical, signal processing, statistical, and optimization communities.

Appendix A. Generalized Gaussian Distributions.

DEFINITION A.1 (GG cumulative distribution function). *The cumulative distribution function (cdf) of a scalar random variable* $X \sim \mathrm{GG}(\eta, \gamma, s)$ *reads*

$$F_X(x) \;=\; \frac{1}{2} + \frac{\operatorname{sign}(x-\eta)}{2\,\Gamma\left(1/s\right)}\;\underline{\Gamma}\left(\frac{1}{s}\,,\,\gamma^s\,|x-\eta|^s\right)\,, \tag{A.1}$$

with $\underline{\Gamma}$ *the lower incomplete Gamma function defined in* (2.1) *and* Γ *the Gamma function defined in* (2.2).

LEMMA A.2. *If* $X_i \sim \mathrm{Gamma}(\nu, z_i)$, $i = 1, \ldots, M$, *are independent random variables, then it holds true that*

$$Y \;=\; \sum_{i=1}^{M} X_i \;\sim\; \mathrm{Gamma}\left(\nu, \sum_{i=1}^{n} z_i\right).$$

LEMMA A.3. *If* $X \sim \mathrm{GG}(0, \gamma, s)$, *then it holds true that*

$$Y \;=\; g(X) \;=\; |X|^s \;\sim\; \mathrm{Gamma}(\nu, z), \quad \nu = \frac{1}{\gamma^s},\; z = \frac{1}{s}\,. \tag{A.2}$$

Proof. We have

$$F_Y(y) = P\left(Y \in]-\infty, y]\right) = P\left(Y \in [0, y]\right), \tag{A.3}$$

$$= P\left(X \in g^{-1}\left([0, y]\right)\right) = P\left(X \in \left[-y^{1/s}, +y^{1/s}\right]\right), \tag{A.4}$$

$$= 1 - 2\,F_X\left(-y^{1/s}\right) = \frac{1}{2\,\Gamma\left(1/s\right)}\,\underline{\Gamma}\left(\frac{1}{s}\,,\,\gamma^s\,y\right), \tag{A.5}$$

where the first and second equalities in (A.3) come from the definition of cdf and from noticing that Y cannot assume negative values, respectively, $g^{-1}\left([0, y]\right)$ in (A.4) denotes the preimage of interval $[0, y]$ under the function g defined in (A.2), and (A.5) follows from the pdf of X being an even function and, then, from replacing the expression of the GG cdf given in (A.1) for F_X.

The pdf π_Y can be obtained by differentiating the cdf F_Y in (A.5). To this aim, first we rewrite F_Y in the following equivalent composite form:

$$F_Y(y) = F_2\left(F_1(y)\right), \quad F_1(y) = \gamma^s\,y, \quad F_2(w) = \frac{1}{2\,\Gamma\left(1/s\right)}\int_0^w t^{\frac{1}{s}-1}e^{-t}dt\,,$$

where we also replaced the explicit expression of the lower incomplete Gamma function $\underline{\Gamma}$ given in (2.1). By applying the chain rule of differentiation, we have

$$\pi_Y(y) = \frac{d}{dy}F_Y = \frac{d}{dy}F_1(y) \times \frac{d}{dw}F_2\left(\gamma^s\,y\right) \tag{A.6}$$

$$= \gamma^s \times \frac{1}{2\,\Gamma\left(1/s\right)}\left(\gamma^s\,y\right)^{\frac{1}{s}-1}\exp\left(-\gamma^s\,y\right) \tag{A.7}$$

$$= \frac{\gamma}{2\,\Gamma\left(1/s\right)}\;y^{\frac{1}{s}-1}\exp\left(-\gamma^s\,y\right). \tag{A.8}$$

A simple one-to-one reparametrization of (A.8), namely $s = 1/z$, $\gamma = (1/\nu)^s$, together with the recall of definition (2.4), leads to (A.2) and thus completes the proof. ∎

Proof of Proposition 2.9. Statement (2.8) follows straightforwardly from Lemmas A.2 and A.3, whereas (2.9) comes from (2.8) and the expressions given in (2.5) for the mean and standard deviation of a Gamma-distributed random variable. ∎

Acknowledgments. The authors are extremely grateful to Laure Blanc-Féraud and Xavier Descombes for their invaluable suggestions and comments on the statistical derivation of our models and to Simone Parisotto for his advice on their geometrical interpretation.

REFERENCES

[1] S. Alliney, *A property of the minimum vectors of a regularizing functional defined by means of the absolute norm*, IEEE Trans. Signal Process., 45 (1997), pp. 913–917, https://doi.org/10.1109/78.564179. (Cited on p. 619)

[2] F. Andreu, C. Ballester, V. Caselles, and J. M. Mazon, *Minimizing total variation flow*, Differential Integral Equations, 14 (2001), pp. 321–360. (Cited on p. 611)

[3] L. Antonelli, V. De Simone, and D. di Serafino, *Spatially adaptive regularization in image segmentation*, Algorithms, 13 (2020), art. 226, https://doi.org/10.3390/a13090226. (Cited on p. 675)

[4] S. Arridge, V. Kolehmainen, and M. J. Schweiger, *Reconstruction and regularisation in optical tomography*, in Mathematical Methods in Biomedical Imaging and Intensity-Modulated Radiation Therapy (IMRT), Y. A. Censor, M. Jiang, and Louis, eds., Scuola Normale Superiore, 2008, pp. 1–18. (Cited on p. 614)

[5] S. Arridge, P. Maass, O. Öktem, and C.-B. Schönlieb, *Solving inverse problems using data-driven models*, Acta Numer., 28 (2019), pp. 1–174, https://doi.org/10.1017/S0962492919000059. (Cited on p. 613)

[6] G. Aubert and J.-F. Aujol, *A variational approach to removing multiplicative noise*, SIAM J. Appl. Math., 68 (2008), pp. 925–946, https://doi.org/10.1137/060671814. (Cited on p. 619)

[7] S. D. Babacan, R. Molina, and A. Katsaggelos, *Generalized Gaussian Markov random field image restoration using variational distribution approximation*, in 2008 IEEE Internat. Conf. Acoustics, Speech and Signal Proc., 2008, pp. 1265–1268, https://doi.org/10.1109/ICASSP.2008.4517847. (Cited on pp. 607, 627)

[8] S. D. Babacan, R. Molina, and A. Katsaggelos, *Parameter estimation in TV image restoration using variational distribution approximation*, IEEE Trans. Image Process., 17 (2008), pp. 326–339, https://doi.org/10.1109/TIP.2007.916051. (Cited on p. 627)

[9] C. Bathke, T. Kluth, C. Brandt, and P. Maaß, *Improved image reconstruction in magnetic particle imaging using structural a priori information*, Internat. J. Magnetic Particle Imaging, 3 (2017), art. 703015, https://doi.org/10.1088/1361-6420/aaaf63. (Cited on p. 614)

[10] I. Bayram and M. E. Kamasak, *Directional total variation*, IEEE Signal Process. Lett., 19 (2012), pp. 781–784, https://doi.org/10.1109/LSP.2012.2220349. (Cited on pp. 605, 606, 636, 643)

[11] A. Beck and M. Teboulle, *A fast iterative shrinkage-thresholding algorithm for linear inverse problems*, SIAM J. Imaging Sci., 2 (2009), pp. 183–202, https://doi.org/10.1137/080716542. (Cited on p. 654)

[12] J. Bentley and M. McIlroy, *Engineering a sort function*, Software Practice Exper., 23 (1993), pp. 1249–1265, https://doi.org/10.1002/spe.4380231105. (Cited on p. 659)

[13] M. Bertero, P. Boccacci, and V. Ruggiero, *Inverse Imaging with Poisson Data*, IOP Publishing, 2018, https://doi.org/10.1088/2053-2563/aae109. (Cited on p. 619)

[14] M. Bertero, P. Boccacci, G. Talenti, R. Zanella, and L. Zanni, *A discrepancy principle for Poisson data*, Inverse Problems, 26 (2010), art. 105004, https://doi.org/10.1088/0266-5611/26/10/105004. (Cited on p. 619)

[15] D. P. Bertsekas, *Constrained Optimization and Lagrange Multiplier Methods*, Academic Press, 1982. (Cited on p. 658)

[16] A. Blake, P. Kohli, and C. Rother, *Markov Random Fields for Vision and Image Processing*, The MIT Press, 2011, https://doi.org/10.7551/mitpress/8579.001.0001. (Cited on p. 612)

[17] Y. G. Boink, M. J. Lagerwerf, W. Steenbergen, S. A. van Gils, S. Manohar, and C. Brune, *A framework for directional and higher-order reconstruction in photoacoustic tomography*, Phys. Med. Biol., 63 (2018), art. 045018, https://doi.org/10.1088/1361-6560/aaaa4a. (Cited on p. 614)

[18] J. Bolte, S. Sabach, and M. Teboulle, *Nonconvex Lagrangian-based optimization: Monitoring schemes and global convergence*, Math. Oper. Res., 43 (2018), pp. 1210–1232, https://doi.org/10.1287/moor.2017.0900. (Cited on p. 654)

[19] A. Borji, *Pros and cons of GAN evaluation measures*, Comput. Vis. Image Und., 179 (2019), pp. 41–65, https://doi.org/10.1016/j.cviu.2018.10.009. (Cited on p. 675)

[20] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, *Distributed optimization and statistical learning via the alternating direction method of multipliers*, Found. Trends Mach. Learn., 3 (2011), pp. 1–122, https://doi.org/10.1561/2200000016. (Cited on p. 654)

[21] K. Bredies, K. Kunisch, and T. Pock, *Total generalized variation*, SIAM J. Imaging Sci., 3 (2010), pp. 492–526, https://doi.org/10.1137/090769521. (Cited on p. 605)

[22] L. Bungert, D. A. Coomes, M. J. Ehrhardt, J. Rasch, R. Reisenhofer, and C.-B. Schönlieb, *Blind image fusion for hyperspectral imaging with the directional total variation*, Inverse Problems, 34 (2018), art. 044003, https://doi.org/10.1088/1361-6420/aaaf63. (Cited on p. 614)

[23] M. Burger and S. Osher, *A guide to the TV zoo*, in Level Set and PDE Based Reconstruction Methods in Imaging, Lecture Notes in Math. 2090, Springer, Cham, 2013, pp. 1–70, https://doi.org/10.1007/978-3-319-01712-9_1. (Cited on p. 604)

[24] R. H. Byrd, P. Lu, J. Nocedal, and C. Zhu, *A limited memory algorithm for bound constrained optimization*, SIAM J. Sci. Comput., 16 (1995), pp. 1190–1208, https://doi.org/10.1137/0916069. (Cited on p. 654)

[25] L. Calatroni, C. Cao, J. C. D. los Reyes, C.-B. Schönlieb, and T. Valkonen, *Bilevel approaches for learning of variational imaging models*, in Variational Methods, De Gruyter, Berlin, Boston, 2017, pp. 252–290, https://doi.org/10.1515/9783110430394-008. (Cited

on p. 608)

[26] L. Calatroni, M. d'Autume, R. Hocking, S. Panayotova, S. Parisotto, P. Ricciardi, and C.-B. Schönlieb, *Unveiling the invisible: Mathematical methods for restoring and interpreting illuminated manuscripts*, Heritage Sci., 6 (2018), art. 56, https://doi.org/10.1186/s40494-018-0216-z. (Cited on p. 614)

[27] L. Calatroni, A. Lanza, M. Pragliola, and F. Sgallari, *A flexible space-variant anisotropic regularization for image restoration with automated parameter selection*, SIAM J. Imaging Sci., 12 (2019), pp. 1001–1037, https://doi.org/10.1137/18M1227937. (Cited on pp. 634, 638, 643, 644, 656)

[28] L. Calatroni, A. Lanza, M. Pragliola, and F. Sgallari, *Space-adaptive anisotropic bivariate Laplacian regularization for image restoration*, in VipIMAGE 2019, J. M. R. S. Tavares and R. M. Natal Jorge, eds., Springer, Cham, 2019, pp. 67–76. (Cited on p. 640)

[29] L. Calatroni, A. Lanza, M. Pragliola, and F. Sgallari, *Adaptive parameter selection for weighted-TV image reconstruction problems*, J. Phys. Conf. Ser., 1476 (2020), pp. 541–547, https://doi.org/10.1088/1742-6596/1476/1/012003. (Cited on pp. 631, 643)

[30] D. Calvetti, S. Morigi, L. Reichel, and F. Sgallari, *Tikhonov regularization and the L-curve for large discrete ill-posed problems*, J. Comput. Appl. Math., 123 (2000), pp. 423–446, https://doi.org/10.1016/S0377-0427(00)00414-3. (Cited on p. 608)

[31] D. Calvetti, A. Pascarella, F. Pitolli, E. Somersalo, and B. Vantaggi, *A hierarchical Krylov–Bayes iterative inverse solver for MEG with physiological preconditioning*, Inverse Problems, 31 (2015), art. 125005, https://doi.org/10.1088/0266-5611/31/12/125005. (Cited on p. 674)

[32] D. Calvetti, M. Pragliola, and E. Somersalo, *Sparsity promoting hybrid solvers for hierarchical Bayesian inverse problems*, SIAM J. Sci. Comput., 42 (2020), pp. A3761–A3784, https://doi.org/10.1137/20M1326246. (Cited on pp. 610, 612)

[33] D. Calvetti, M. Pragliola, E. Somersalo, and A. Strang, *Sparse reconstructions from few noisy data: Analysis of hierarchical Bayesian models with generalized gamma hyperpriors*, Inverse Problems, 36 (2020), art. 025010, https://doi.org/10.1088/1361-6420/ab4d92. (Cited on pp. 610, 612, 675)

[34] D. Calvetti and E. Somersalo, *Hypermodels in the Bayesian imaging framework*, Inverse Problems, 24 (2008), art. 034013, https://doi.org/10.1088/0266-5611/24/3/034013. (Cited on p. 627)

[35] D. Calvetti and E. Somersalo, *Statistical methods in imaging*, in Handbook of Mathematical Methods in Imaging, Springer, New York, 2015, pp. 1343–1392, https://doi.org/10.1007/978-1-4939-0790-8_53. (Cited on p. 611)

[36] D. Calvetti, E. Somersalo, and A. Strang, *Hierarchical Bayesian models and sparsity: ℓ_2-magic*, Inverse Problems, 35 (2019), art. 035003, https://doi.org/10.1088/1361-6420/ab4d92. (Cited on pp. 610, 612)

[37] E. J. Candès, Y. C. Eldar, D. Needell, and P. Randall, *Compressed sensing with coherent and redundant dictionaries*, Appl. Comput. Harmon. Anal., 31 (2011), pp. 59–73, https://doi.org/10.1016/j.acha.2010.10.002. (Cited on p. 674)

[38] V. Caselles, A. Chambolle, D. Cremers, M. Novaga, and T. Pock, *An introduction to total variation for image analysis*, in Theoretical Foundations and Numerical Methods for Sparse Recovery, De Gruyter, Berlin, New York, 2010, pp. 263–340, https://doi.org/10.1515/9783110226157.263. (Cited on p. 604)

[39] V. Caselles, A. Chambolle, and M. Novaga, *The discontinuity set of solutions of the TV denoising problem and some extensions*, Multiscale Model. Simul., 6 (2007), pp. 879–894, https://doi.org/10.1137/070683003. (Cited on p. 605)

[40] V. Caselles, A. Chambolle, and M. Novaga, *Total variation in imaging*, in Handbook of Mathematical Methods in Imaging, Springer, New York, 2015, pp. 1455–1499, https://doi.org/10.1007/978-1-4939-0790-8_23. (Cited on p. 604)

[41] A. Chambolle, *An algorithm for total variation minimization and applications*, J. Math. Imaging Vis., 20 (2004), pp. 89–97, https://doi.org/10.1023/B:JMIV.0000011325.36760.1e/. (Cited on p. 654)

[42] A. Chambolle, S. E. Levine, and B. J. Lucier, *An upwind finite-difference method for total variation–based image smoothing*, SIAM J. Imaging Sci., 4 (2011), pp. 277–299, https://doi.org/10.1137/090752754. (Cited on p. 674)

[43] A. Chambolle and P.-L. Lions, *Image recovery via total variation minimization and related problems*, Numer. Math., 76 (1997), pp. 167–188, https://doi.org/10.1007/s002110050258. (Cited on p. 605)

[44] A. Chambolle and T. Pock, *A first-order primal-dual algorithm for convex problems with applications to imaging*, J. Math. Imaging Vis., 40 (2011), pp. 120–145, https://doi.org/

10.1007/s10851-010-0251-1. (Cited on p. 654)

[45] A. Chambolle and T. Pock, *An introduction to continuous optimization for imaging*, Acta Numer., 25 (2016), pp. 161–319, https://doi.org/10.1017/S096249291600009X. (Cited on pp. 604, 654)

[46] A. Chambolle and T. Pock, *Approximating the total variation with finite differences or finite elements*, in Handbook of Numerical Analysis: Geometric Partial Differential Equations–Part II, Elsevier, 2021, pp. 383–417, https://doi.org/10.1016/bs.hna.2020.10.005. (Cited on p. 674)

[47] T. Chan, A. Marquina, and P. Mulet, *High-order total variation-based image restoration*, SIAM J. Sci. Comput., 22 (2000), pp. 503–516, https://doi.org/10.1137/S1064827598344169. (Cited on p. 605)

[48] T. F. Chan and S. Esedoḡlu, *Aspects of total variation regularized L^1 function approximation*, SIAM J. Appl. Math., 65 (2005), pp. 1817–1837, https://doi.org/10.1137/040604297. (Cited on p. 619)

[49] T. F. Chan, S. Esedoglu, and F. Park, *A fourth order dual method for staircase reduction in texture extraction and image restoration problems*, in 2010 IEEE International Conference on Image Processing, 2010, pp. 4137–4140, https://doi.org/10.1109/ICIP.2010.5653199. (Cited on p. 605)

[50] T. F. Chan, G. H. Golub, and P. Mulet, *A nonlinear primal-dual method for total variation-based image restoration*, SIAM J. Sci. Comput., 20 (1999), pp. 1964–1977, https://doi.org/10.1137/S1064827596299767. (Cited on p. 654)

[51] Y. Chen, S. Levine, and M. Rao, *Variable exponent, linear growth functionals in image restoration*, SIAM J. Appl. Math., 66 (2006), pp. 1383–1406, https://doi.org/10.1137/050624522. (Cited on p. 674)

[52] C. V. Chung, J. C. D. los Reyes, and C. B. Schönlieb, *Learning optimal spatially-dependent regularization parameters in total variation image denoising*, Inverse Problems, 33 (2017), art. 074005, https://doi.org/10.1088/1361-6420/33/7/074005. (Cited on p. 608)

[53] C. Clason, *L_∞ fitting for inverse problems with uniform noise*, Inverse Problems, 28 (2012), art. 104007. (Cited on p. 619)

[54] P. Combettes and J. Pesquet, *Proximal splitting methods in signal processing*, in Fixed-Point Algorithms for Inverse Problems in Science and Engineering, Springer, New York, 2011, pp. 185–212, https://doi.org/10.1007/978-1-4419-9569-8_10. (Cited on pp. 617, 654)

[55] L. Condat, *Primal–dual splitting method for convex optimization involving Lipschitzian, proximable and linear composite terms*, J. Optim. Theory Appl., 158 (2013), pp. 460–479, https://doi.org/10.1007/s10957-012-0245-9. (Cited on p. 654)

[56] L. Condat, *Fast projection onto the simplex and the ℓ_1 ball*, Math. Program., 158 (2016), pp. 575–585, https://doi.org/10.1007/s10107-015-0946-6. (Cited on p. 659)

[57] L. Condat, *Discrete total variation: New definition and minimization*, SIAM J. Imaging Sci., 10 (2017), pp. 1258–1290, https://doi.org/10.1137/16M1075247. (Cited on p. 674)

[58] V. De Bortoli, A. Durmus, M. Pereyra, and A. F. Vidal, *Maximum likelihood estimation of regularization parameters in high-dimensional inverse problems: An empirical Bayesian approach. Part* II*: Theoretical analysis*, SIAM J. Imaging Sci., 13 (2020), pp. 1990–2028, https://doi.org/10.1137/20M1339842. (Cited on pp. 627, 675)

[59] E. Demircan-Tureyen and M. E. Kamasak, *Adaptive Direction-Guided Structure Tensor Total Variation*, preprint, https://arxiv.org/abs/2001.05717, 2020. (Cited on pp. 636, 638)

[60] X. Descombes, M. Sigelle, and F. Preteux, *Estimating Gaussian Markov random field parameters in a nonstationary framework: Application to remote sensing imaging*, IEEE Trans. Image Process., 8 (1999), pp. 490–503, https://doi.org/10.1109/83.753737. (Cited on pp. 610, 612)

[61] F. Dibos and G. Koepfler, *Global total variation minimization*, SIAM J. Numer. Anal., 37 (2000), pp. 646–664, https://doi.org/10.1137/S0036142998334838. (Cited on p. 611)

[62] Y. Dong and C.-B. Schönlieb, *Tomographic reconstruction with spatially varying parameter selection*, Inverse Problems, 36 (2020), art. 054002, https://doi.org/10.1088/1361-6420/ab72d4. (Cited on p. 614)

[63] I. Drori, D. Cohen-Or, and H. Yeshurun, *Fragment-based image completion*, ACM Trans. Graph., 22 (2003), pp. 303–312, https://doi.org/10.1145/882262.882267. (Cited on p. 675)

[64] J. C. Duchi, S. Shalev-Shwartz, Y. Singer, and T. Chandra, *Efficient projections onto the ℓ_1-ball for learning in high dimensions*, in ICML '08: Proceedings of the 25th International Conference on Machine Learning, Vol. 3, 2008, pp. 272–279, https:

//doi.org/10.1145/1390156.1390191. (Cited on p. 659)
[65] M. J. Ehrhardt and M. M. Betcke, *Multicontrast MRI reconstruction with structure-guided total variation*, SIAM J. Imaging Sci., 9 (2016), pp. 1084–1106, https://doi.org/10.1137/15M1047325. (Cited on pp. 614, 638, 675)
[66] M. J. Ehrhardt, P. Markiewicz, M. Liljeroth, A. Barnes, V. Kolehmainen, J. S. Duncan, L. Pizarro, D. Atkinson, B. F. Hutton, S. Ourselin, K. Thielemans, and S. R. Arridge, *PET reconstruction with an anatomical MRI prior using parallel level sets*, IEEE Trans. Med. Imaging, 35 (2016), pp. 2189–2199, https://doi.org/10.1109/TMI.2016.2549601. (Cited on pp. 614, 675)
[67] M. J. Ehrhardt, K. Thielemans, L. Pizarro, D. Atkinson, S. Ourselin, B. F. Hutton, and S. R. Arridge, *Joint reconstruction of PET-MRI by exploiting structural similarity*, Inverse Problems, 31 (2014), art. 015001, https://doi.org/10.1088/0266-5611/31/1/015001. (Cited on pp. 614, 675)
[68] H. Engl, M. Hanke, and A. Neubauer, *Regularization of Inverse Problems*, Math. Appl. 375, Springer, Dordrecht, 2000. (Cited on pp. 604, 618)
[69] V. Estellers, S. Soatto, and X. Bresson, *Adaptive regularization with the structure tensor*, IEEE Trans. Image Process., 24 (2015), pp. 1777–1790, https://doi.org/10.1109/tip.2015.2409562. (Cited on pp. 611, 638)
[70] J. Fehrenbach and J.-M. Mirebeau, *Sparse non-negative stencils for anisotropic diffusion*, J. Math. Imaging Vis., 49 (2014), pp. 123–147, https://doi.org/10.1007/s10851-013-0446-3. (Cited on p. 674)
[71] C. Fenu, L. Reichel, and G. Rodriguez, *GCV for Tikhonov regularization via global Golub–Kahan decomposition*, Numer. Linear Algebra Appl., 23 (2016), https://doi.org/10.1002/nla.2034. (Cited on p. 608)
[72] S. Geman and D. Geman, *Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images*, IEEE Trans. Pattern Anal. Mach. Intell., PAMI-6 (1984), pp. 721–741, https://doi.org/10.1109/TPAMI.1984.4767596. (Cited on p. 611)
[73] I. Giakoumis and I. Pitas, *Digital restoration of painting cracks*, in Proceedings of the 1998 IEEE International Symposium on Circuits and Systems, Vol. 4, 1998, pp. 269–272, https://doi.org/10.1109/ISCAS.1998.698812. (Cited on p. 614)
[74] G. H. Golub, M. Heath, and G. Wahba, *Generalized cross-validation as a method for choosing a good ridge parameter*, Technometrics, 21 (1979), pp. 215–223, https://doi.org/10.1080/00401706.1979.10489751. (Cited on p. 608)
[75] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, *Generative adversarial nets*, in Proceedings of the 27th International Conference on Neural Information Processing Systems, Vol. 2, NIPS'14, MIT Press, Cambridge, MA, 2014, pp. 2672–2680. (Cited on pp. 610, 613, 675)
[76] M. Grasmair and F. Lenzen, *Anisotropic total variation filtering*, Appl. Math. Opt., 62 (2010), pp. 323–339, https://doi.org/10.1007/s00245-010-9105-x. (Cited on pp. 611, 638)
[77] M. Guo, C. Han, W. Wang, S. Zhong, R. Lv, and Z. Liu, *A novel truncated nonconvex nonsmooth variational method for SAR image despeckling*, Remote Sens. Lett., 12 (2021), pp. 174–183, https://doi.org/10.1080/2150704X.2020.1846820. (Cited on p. 608)
[78] M. Hanke, *Chapter 8: The Discrepancy Principle*, in A Taste of Inverse Problems, SIAM, Philadelphia, 2017, pp. 55–60, https://doi.org/10.1137/1.9781611974942.ch8. (Cited on p. 608)
[79] C. He, C. Hu, W. Zhang, and B. Shi, *A fast adaptive parameter estimation for total variation image restoration*, IEEE Trans. Image Process., 23 (2014), pp. 4954–4967, https://doi.org/10.1109/TIP.2014.2360133. (Cited on pp. 608, 654)
[80] M. Held, P. Wolfe, and H. Crowder, *Validation of subgradient optimization*, Math. Program., 6 (1974), pp. 62–88, https://doi.org/10.1007/BF01580223. (Cited on p. 659)
[81] M. Hintermüller and K. Kunisch, *Total bounded variation regularization as a bilaterally constrained optimization problem*, SIAM J. Appl. Math., 64 (2004), pp. 1311–1333, https://doi.org/10.1137/S0036139903422784. (Cited on p. 654)
[82] M. Hintermüller and K. Papafitsoros, *Generating structured nonsmooth priors and associated primal-dual methods*, in Processing, Analyzing and Learning of Images, Shapes, and Forms: Part 2, R. Kimmel and X.-C. Tai, eds., Handb. Numer. Anal. 20, Elsevier, 2019, pp. 437–502, https://doi.org/10.1016/bs.hna.2019.08.001. (Cited on p. 631)
[83] M. Hintermüller, K. Papafitsoros, and C. N. Rautenberg, *Analytical aspects of spatially adapted total variation regularisation*, J. Math. Anal. Appl., 454 (2017), pp. 891–935, https://doi.org/10.1016/j.jmaa.2017.05.025. (Cited on p. 631)
[84] M. Hintermüller and C. N. Rautenberg, *Optimal selection of the regularization function in a weighted total variation model. Part I: Modelling and theory*, J. Math. Imaging Vis.,

59 (2017), pp. 498–514, https://doi.org/10.1007/s10851-017-0744-2. (Cited on pp. 608, 631)

[85] M. Hintermüller, C. N. Rautenberg, T. Wu, and A. Langer, *Optimal selection of the regularization function in a weighted total variation model. Part* II*: Algorithm, its analysis and numerical tests*, J. Math. Imaging Vis., 59 (2017), pp. 515–533, https://doi.org/10.1007/s10851-017-0736-2. (Cited on p. 608)

[86] M. Hintermüller and G. Stadler, *An infeasible primal-dual algorithm for total bounded variation–based inf-convolution-type image restoration*, SIAM J. Sci. Comput., 28 (2006), pp. 1–23, https://doi.org/10.1137/040613263. (Cited on p. 654)

[87] Y.-M. Huang, M. K. Ng, and Y.-W. Wen, *A new total variation method for multiplicative noise removal*, SIAM J. Imaging Sci., 2 (2009), pp. 20–40, https://doi.org/10.1137/080712593. (Cited on p. 619)

[88] K. Jalalzai, *Some remarks on the staircasing phenomenon in total variation-based image denoising*, J. Math. Imaging Vis., 54 (2016), pp. 256–268, https://doi.org/10.1007/s10851-015-0600-1. (Cited on p. 605)

[89] K. M. Jespersen, J. Zangenberg, T. Lowe, P. J. Withers, and L. P. Mikkelsen, *Fatigue damage assessment of uni-directional non-crimp fabric reinforced polyester composite using X-ray computed tomography*, Comp. Science Tech., 136 (2016), pp. 94–103, https://doi.org/10.1016/j.compscitech.2016.10.006. (Cited on p. 607)

[90] J. Kaipio and E. Somersalo, *Statistical and Computational Inverse Problems*, Springer-Verlag, New York, 2004, https://doi.org/10.1007/b138659. (Cited on p. 612)

[91] J. P. Kaipio, V. Kolehmainen, M. Vauhkonen, and E. Somersalo, *Inverse problems with structural prior information*, Inverse Problems, 15 (1999), pp. 713–729, https://doi.org/10.1088/0266-5611/15/3/306. (Cited on p. 614)

[92] D. P. Kingma and M. Welling, *Auto-Encoding Variational Bayes*, preprint, https://arxiv.org/abs/1312.6114, 2014. (Cited on pp. 613, 675)

[93] R. Kongskov and Y. Dong, *Directional total generalized variation regularization for impulse noise removal*, in Scale Space and Variational Methods in Computer Vision, F. Lauze, Y. Dong, and A. B. Dahl, eds., Springer, Cham, 2017, pp. 221–231. (Cited on p. 605)

[94] R. Kongskov, Y. Dong, and K. Knudsen, *Directional total generalized variation regularization*, BIT, 59 (2019), pp. 903–928, https://doi.org/10.1007/s10543-019-00755-6. (Cited on pp. 607, 636, 637)

[95] D. Krishnan and R. Fergus, *Fast image deconvolution using hyper-Laplacian priors*, in Advances in Neural Information Processing Systems, Y. Bengio, D. Schuurmans, J. Lafferty, C. Williams, and A. Culotta, eds., Curran Associates, 2009, pp. 1033–1041, https://proceedings.neurips.cc/paper/2009/file/3dd48ab31d016ffcbf3314df2b3cb9ce-Paper.pdf. (Cited on pp. 607, 608)

[96] K. Kunisch and T. Pock, *A bilevel optimization approach for parameter learning in variational models*, SIAM J. Imaging Sci., 6 (2013), pp. 938–983, https://doi.org/10.1137/120882706. (Cited on p. 608)

[97] K. Kurach, M. Lucic, X. Zhai, M. Michalski, and S. Gelly, *The GAN Landscape: Losses, Architectures, Regularization, and Normalization*, 2019, https://openreview.net/forum?id=rkGG6s0qKQ. (Cited on pp. 610, 613)

[98] A. Lanza, S. Morigi, M. Pragliola, and F. Sgallari, *Space-variant generalised Gaussian regularisation for image restoration*, Comput. Meth. Biomech. Biomed. Eng. Imaging Vis., 13 (2018), pp. 490–503, https://doi.org/10.1080/21681163.2018.1471620. (Cited on pp. 632, 638, 643)

[99] A. Lanza, S. Morigi, M. Pragliola, and F. Sgallari, *Space-variant TV regularization for image restoration*, in VipIMAGE 2017, J. M. R. Tavares and R. Natal Jorge, eds., Springer, Cham, 2018, pp. 160–169. (Cited on p. 632)

[100] A. Lanza, S. Morigi, F. Sciacchitano, and F. Sgallari, *Whiteness constraints in a unified variational framework for image restoration*, J. Math. Imaging Vis., 60 (2018), pp. 1573–7683, https://doi.org/10.1007/s10851-018-0845-6. (Cited on p. 608)

[101] A. Lanza, S. Morigi, and F. Sgallari, *Constrained TV_p-ℓ_2 model for image restoration*, J. Sci. Comput., 68 (2016), pp. 64–91, https://doi.org/10.1007/s10915-015-0129-x. (Cited on pp. 607, 634, 635, 643, 657, 661, 663)

[102] A. Lanza, M. Pragliola, and F. Sgallari, *Residual whiteness principle for parameter-free image restoration*, Electron. Trans. Numer. Anal., 53 (2020), pp. 329–351, https://doi.org/10.1553/etna_vol53s329. (Cited on pp. 608, 665)

[103] A. Lanza, M. Pragliola, and F. Sgallari, *Automatic fidelity and regularization terms selection in variational image restoration*, BIT, 62 (2022), pp. 931–964, https://doi.org/10.1007/s10543-021-00901-z. (Cited on p. 661)

[104] T. Le, R. Chartrand, and T. J. Asaki, *A variational approach to reconstructing images corrupted by Poisson noise*, J. Math. Imaging Vis., 27 (2007), pp. 257–263, https://doi.org/10.1007/s10851-007-0652-y. (Cited on p. 619)

[105] Y. Le Montagner, E. D. Angelini, and J. Olivo-Marin, *An unbiased risk estimator for image denoising in the presence of mixed Poisson–Gaussian noise*, IEEE Trans. Image Process., 23 (2014), pp. 1255–1268, https://doi.org/10.1109/TIP.2014.2300821. (Cited on p. 608)

[106] S. Lefkimmiatis, A. Roussos, P. Maragos, and M. Unser, *Structure tensor total variation*, SIAM J. Imaging Sci., 8 (2015), pp. 1090–1122, https://doi.org/10.1137/14098154X. (Cited on p. 638)

[107] S. Lefkimmiatis, J. P. Ward, and M. Unser, *Hessian Schatten-norm regularization for linear inverse problems*, IEEE Trans. Image Process., 22 (2013), pp. 1873–1888, https://doi.org/10.1109/TIP.2013.2237919. (Cited on p. 674)

[108] S. Z. Li, *Markov Random Field Modeling in Image Analysis*, 3rd ed., Springer-Verlag, Tokyo, Japan, 2009. (Cited on pp. 611, 612)

[109] Y. Li, C. Wu, and Y. Duan, *The TV_p regularized Mumford-Shah model for image labeling and segmentation*, IEEE Trans. Image Process., 29 (2020), pp. 7061–7075, https://doi.org/10.1109/TIP.2020.2997524. (Cited on p. 608)

[110] R. W. Liu, Y. Li, Y. Liu, J. Duan, T. Xu, and J. Liu, *Single-image blind deblurring with hybrid sparsity regularization*, in 2017 20th International Conference on Information Fusion (Fusion), 2017, pp. 1–8, https://doi.org/10.23919/ICIF.2017.8009659. (Cited on p. 608)

[111] F. Lozes, A. Elmoataz, and O. Lezoray, *PDE-based graph signal processing for 3-D color point clouds: Opportunities for cultural heritage*, IEEE Signal Process. Mag., 32 (2015), pp. 103–111, https://doi.org/10.1109/MSP.2015.2408631. (Cited on p. 614)

[112] M. Lucic, K. Kurach, M. Michalski, O. Bousquet, and S. Gelly, *Are GANs created equal? A large-scale study*, in NIPS'18: Proceedings of the 32nd International Conference on Neural Information Processing Systems, Curran Associates, Red Hook, NY, 2018, pp. 698–707. (Cited on pp. 610, 613)

[113] S. Lunz, O. Öktem, and C.-B. Schönlieb, *Adversarial regularizers in inverse problems*, in NIPS'18: Proceedings of the 32nd International Conference on Neural Information Processing Systems, S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi, and R. Garnett, eds., Vol. 31, Curran Associates, Red Hook, NY, 2018, https://proceedings.neurips.cc/paper/2018/file/d903e9608cfbf08910611e4346a0ba44-Paper.pdf. (Cited on pp. 610, 613)

[114] Y. Meyer, *Oscillating Patterns in Image Processing and Nonlinear Evolution Equations: The Fifteenth Dean Jacqueline B. Lewis Memorial Lectures*, American Mathematical Society, 2001. (Cited on p. 605)

[115] T. Miyato, T. Kataoka, M. Koyama, and Y. Yoshida, *Spectral normalization for generative adversarial networks*, in Internet. Conf. Learning Repres., 2018, https://openreview.net/forum?id=B1QRgziT-. (Cited on pp. 610, 613)

[116] V. Monga, Y. Li, and Y. C. Eldar, *Algorithm unrolling: Interpretable, efficient deep learning for signal and image processing*, IEEE Signal Process. Mag., 38 (2021), pp. 18–44, https://doi.org/10.1109/MSP.2020.3016905. (Cited on p. 675)

[117] V. A. Morozov, *On the solution of functional equations by the method of regularization*, Dokl. Akad. Nauk SSSR, 167 (1966), pp. 510–512. (Cited on p. 618)

[118] Y. Nesterov, *Introductory Lectures on Convex Optimization: A Basic Course*, Appl. Optim. 87, Kluwer Academic, Boston, 2004. (Cited on p. 654)

[119] M. Nikolova, *Local strong homogeneity of a regularized estimator*, SIAM J. Appl. Math., 61 (2000), pp. 633–658, https://doi.org/10.1137/S0036139997327794. (Cited on p. 605)

[120] M. Nikolova, *A variational approach to remove outliers and impulse noise*, J. Math. Imaging Vis., 20 (2004), pp. 99–120, https://doi.org/10.1023/B:JMIV.0000011326.88682.e5. (Cited on p. 619)

[121] P. Ochs, Y. Chen, T. Brox, and T. Pock, *iPiano: Inertial proximal algorithm for nonconvex optimization*, SIAM J. Imaging Sci., 7 (2014), pp. 1388–1419, https://doi.org/10.1137/130942954. (Cited on p. 654)

[122] J. Oliveira, J. Bioucas-Dias, and M. A. T. Figueiredo, *Adaptive total variation image deblurring: A majorization-minimization approach*, Signal Process., 89 (2009), pp. 1683–1693, https://doi.org/10.1016/j.sigpro.2009.03.018. (Cited on p. 627)

[123] Z.-F. Pang, H.-L. Zhang, S. Luo, and T. Zeng, *Image denoising based on the adaptive weighted TV^P regularization*, Signal Process., 167 (2020), art. 107325, https://doi.org/10.1016/j.sigpro.2019.107325. (Cited on p. 638)

[124] K. Papafitsoros and C.-B. Schönlieb, *A combined first and second order variational approach for image reconstruction*, J. Math. Imaging Vis., 48 (2014), pp. 308–338, https://doi.org/10.1007/s10851-013-0445-4. (Cited on p. 605)

[125] S. Parisotto, L. Calatroni, A. Bugeau, N. Papadakis, and C. B. Schönlieb, *Variational osmosis for non-linear image fusion*, IEEE Trans. Image Process., 29 (2020), pp. 5507–5516, https://doi.org/10.1109/TIP.2020.2983537. (Cited on p. 614)

[126] S. Parisotto, J. Lellmann, S. Masnou, and C. B. Schönlieb, *Higher order total directional variation: Imaging applications*, SIAM J. Imaging Sci., 13 (2020), pp. 2063–2104, https://doi.org/10.1137/19M1239209. (Cited on pp. 636, 674)

[127] S. Parisotto, S. Masnou, and C. B. Schönlieb, *Higher order total directional variation: Analysis*, SIAM J. Imaging Sci., 13 (2020), pp. 474–496, https://doi.org/10.1137/19M1239210. (Cited on p. 637)

[128] M. Pereyra and S. McLaughlin, *Fast unsupervised Bayesian image segmentation with adaptive spatial regularisation*, IEEE Trans. Image Process., 26 (2017), pp. 2577–2587. (Cited on pp. 610, 612)

[129] P. Perona and J. Malik, *Scale-space and edge detection using anisotropic diffusion*, IEEE Trans. Pattern Anal. Mach. Intell., 12 (1990), pp. 629–639, https://doi.org/10.1109/34.56205. (Cited on p. 610)

[130] H. Raguet, J. Fadili, and G. Peyré, *A generalized forward-backward splitting*, SIAM J. Imaging Sci., 6 (2013), pp. 1199–1226, https://doi.org/10.1137/120872802. (Cited on p. 654)

[131] K. N. Ramamurthy, J. J. Thiagarajan, and A. Spanias, *Fast image registration with non-stationary Gauss-Markov random field templates*, in 2009 16th IEEE International Conference on Image Processing (ICIP), 2009, pp. 185–188, https://doi.org/10.1109/ICIP.2009.5414177. (Cited on pp. 610, 612)

[132] A. Roussos and P. Maragos, *Tensor-based image diffusions derived from generalizations of the total variation and Beltrami functionals*, in 2010 IEEE International Conference on Image Processing, 2010, pp. 4141–4144, https://doi.org/10.1109/ICIP.2010.5653241. (Cited on pp. 610, 611)

[133] L. Rudin, P.-L. Lions, and S. Osher, *Multiplicative denoising and deblurring: Theory and algorithms*, in Geometric Level Set Methods in Imaging, Vision, and Graphics, Springer, New York, 2003, pp. 103–119, https://doi.org/10.1007/0-387-21810-6_6. (Cited on p. 619)

[134] L. I. Rudin, S. Osher, and E. Fatemi, *Nonlinear total variation based noise removal algorithms*, Phys. D, 60 (1992), pp. 259–268, https://doi.org/10.1016/0167-2789(92)90242-F. (Cited on pp. 604, 643, 663)

[135] L. Ruthotto and E. Haber, *An introduction to deep generative modeling*, GAMM-Mitt., 44 (2021), art. e202100008, https://doi.org/10.1002/gamm.202100008. (Cited on p. 613)

[136] S. R. Sandoghchi, G. T. Jasion, N. V. Wheeler, S. Jain, Z. Lian, J. P. Wooler, R. P. Boardman, N. Baddela, Y. Chen, J. Hayes, E. N. Fokoua, T. Bradley, D. R. Gray, S. M. Mousavi, M. Petrovich, F. Poletti, and D. J. Richardson, *X-ray tomography for structural analysis of microstructured and multimaterial optical fibers and preforms*, Opt. Express, 22 (2014), pp. 26181–26192, https://doi.org/10.1364/OE.22.026181. (Cited on p. 607)

[137] A. Sawatzky, C. Brune, J. Müller, and M. Burger, *Total variation processing of images with Poisson statistics*, in Computer Analysis of Images and Patterns, X. Jiang and N. Petkov, eds., Springer, Berlin, Heidelberg, 2009, pp. 533–540. (Cited on p. 619)

[138] H. Scharr, M. J. Black, and H. W. Haussecker, *Image statistics and anisotropic diffusion*, in Proceedings of the 9th IEEE International Conference on Computer Vision, Vol. 2, 2003, pp. 840–847, https://doi.org/10.1109/ICCV.2003.1238435. (Cited on pp. 610, 611)

[139] S. Setzer and G. Steidl, *Variational methods with higher order derivatives in image processing*, in Approximation Theory XII: San Antonio 2007, Nashboro Press, Brentwood, TN, 2008, pp. 360–385. (Cited on p. 605)

[140] J. Shi and S. Osher, *A nonlinear inverse scale space method for a convex multiplicative noise model*, SIAM J. Imaging Sci., 1 (2008), pp. 294–321, https://doi.org/10.1137/070689954. (Cited on p. 619)

[141] G. Steidl and T. Teuber, *Removing multiplicative noise by Douglas-Rachford splitting methods*, J. Math. Imaging Vis., 36 (2009), pp. 168–184. (Cited on p. 619)

[142] C. M. Stein, *Estimation of the mean of a multivariate normal distribution*, Ann. Statist., 9 (1981), pp. 1135–1151, https://doi.org/10.1214/aos/1176345632. (Cited on p. 608)

[143] D. Strong and T. Chan, *Edge-preserving and scale-dependent properties of total variation regularization*, Inverse Problems, 19 (2003), pp. S165–S187, https://doi.org/10.1088/0266-5611/19/6/059. (Cited on p. 605)

[144] A. M. Stuart, *Inverse problems: A Bayesian perspective*, Acta Numer., 19 (2010), pp. 451–559, https://doi.org/10.1017/S0962492910000061. (Cited on p. 604)

[145] T. Teuber, G. Steidl, and R. H. Chan, *Minimization and parameter estimation for seminorm regularization models with I-divergence constraints*, Inverse Problems, 29 (2013), art. 035007, https://doi.org/10.1088/0266-5611/29/3/035007. (Cited on p. 619)

[146] R. Tovey, M. Benning, C. Brune, M. J. Lagerwerf, S. M. Collins, R. K. Leary, P. A. Midgley, and C. B. Schönlieb, *Directional sinogram inpainting for limited angle tomography*, Inverse Problems, 35 (2019), art. 024004, https://doi.org/10.1088/1361-6420/aaf2fe. (Cited on pp. 614, 675)

[147] A. F. Vidal, V. De Bortoli, M. Pereyra, and A. Durmus, *Maximum likelihood estimation of regularization parameters in high-dimensional inverse problems: An empirical Bayesian approach part* I*: Methodology and experiments*, SIAM J. Imaging Sci., 13 (2020), pp. 1945–1989, https://doi.org/10.1137/20M1339829. (Cited on pp. 627, 675)

[148] Y. Wang, W. Yin, and J. Zeng, *Global convergence of ADMM in nonconvex nonsmooth optimization*, J. Sci. Comput., 78 (2019), pp. 29–63, https://doi.org/10.1007/s10915-018-0757-z. (Cited on p. 654)

[149] Z. Wang, A. Bovik, H. R. Sheikh, and E. P. Simoncelli, *Image quality assessment: From error visibility to structural similarity*, IEEE Trans. Image Process., 13 (2004), pp. 600–612, https://doi.org/10.1109/TIP.2003.819861. (Cited on p. 664)

[150] J. Weickert, *Anisotropic Diffusion in Image Processing*, B.G. Teubner, Stuttgart, 1998. (Cited on pp. 610, 611)

[151] J. Weickert and T. Brox, *Diffusion and regularization of vector- and matrix-valued images*, in Inverse Problems, Image Analysis, and Medical Imaging, AMS, 2002, pp. 251–268, http://lmb.informatik.uni-freiburg.de/Publications/2002/Bro02a. (Cited on pp. 610, 611)

[152] B. Widrow and I. Kollár, *Quantization Noise: Roundoff Error in Digital Computation, Signal Processing, Control, and Communications*, Cambridge University Press, 2008, https://doi.org/10.1017/CBO9780511754661. (Cited on p. 619)

[153] R. M. Willett, Z. T. Harmany, and R. F. Marcia, *Poisson image reconstruction with total variation regularization*, in 2010 IEEE International Conference on Image Processing, 2010, pp. 4177–4180, https://doi.org/10.1109/ICIP.2010.5649600. (Cited on p. 619)

[154] C. Wu and X.-C. Tai, *Augmented Lagrangian method, dual methods, and split Bregman iteration for ROF, vectorial TV, and high order models*, SIAM J. Imaging Sci., 3 (2010), pp. 300–339, https://doi.org/10.1137/090767558. (Cited on p. 605)

[155] Q. Xie, D. Zeng, Q. Zhao, D. Meng, Z. Xu, Z. Liang, and J. Ma, *Robust low-dose CT sinogram preprocessing via exploiting noise-generating mechanism*, IEEE Trans. Med. Imaging, 36 (2017), pp. 2487–2498, https://doi.org/10.1109/TMI.2017.2767290. (Cited on p. 627)

[156] J. Yan, F. Wang, X. Cao, and J. Zhang, *Robust object tracking using least absolute deviation*, Image Vision Comput., 32 (2014), p. 930–939, https://doi.org/10.1016/j.imavis.2014.08.008. (Cited on p. 619)

[157] H. Zhang and Y. Wang, *Edge adaptive directional total variation*, J. Engrg., 2013 (2013), pp. 61–62, https://doi.org/10.1049/joe.2013.0116. (Cited on pp. 605, 636, 638)

[158] J. Zhang, A. Gelb, and T. Scarnati, *Empirical Bayesian Inference Using Joint Sparsity*, preprint, https://arxiv.org/abs/2103.15618, 2021. (Cited on pp. 610, 613)

[159] H. Zhao, Y. Liu, C. Huang, and T. Wang, *Hybrid-weighted total variation and nonlocal low-rank-based image compressed sensing reconstruction*, IEEE Access, 8 (2020), pp. 23002–23010, https://doi.org/10.1109/ACCESS.2020.2970158. (Cited on pp. 634, 638, 658)

[160] H. Zhou and J. Zheng, *Adaptive patch size determination for patch-based image completion*, in 2010 IEEE International Conference on Image Processing, 2010, pp. 421–424, https://doi.org/10.1109/ICIP.2010.5654120. (Cited on p. 675)