On and beyond Total Variation regularisation in imaging: the role of space variance

Monica Pragliola∗1, Luca Calatroni†2, Alessandro Lanza‡1, and Fiorella Sgallari§1

1Department of Mathematics, University of Bologna, Italy
2CNRS, UCA, INRIA, Morpheme, I3S, Sophia-Antipolis, France

Abstract

Over the last 30 years a plethora of variational regularisation models for image reconstruction has 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 [118] which relies on the minimisation of the edge-preserving Total Variation (TV) semi-norm as regularisation 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 modelling which accommodates local regularisation weighting, variable smoothness and anisotropy. Those ideas can further be cast in the flexible Bayesian framework of generalised Gaussian distributions and combined with maximum likelihood and hierarchical optimisation approaches for efficient hyper-parameter 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.

1 Introduction

The technological developments which favoured the storage, the exploitation and the management of large and unstructured data over the last decades has been responsible of 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 90’s, advances in the field of non-smooth variational calculus have drifted the attention towards the use of sparsity-promoting image regularisation models as well as the development of efficient optimisation 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 capitalising 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 (if not the) most prominent challenging tasks, with implications in fields such as

∗monica.pragliola2@unibo.it
†calatroni@i3s.unice.fr
‡alessandro.lanza2@unibo.it
§fiorella.sgallari@unibo.it
artificial intelligence, human-to-robot interactions and bio-inspired computer designs. Alongside, the clever and efficient exploitation of training data has favoured 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 travel across disciplines taking as example a standard problem in the context of mathematical image reconstruction in the attempt of highlighting for the popular and well-studied Total Variation (TV) regularisation 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 of combining classical tools of variational calculus, optimisation, 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 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 the inexpert reader to familiarise 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 \( Ω := \{(h,l) : h = 1, \ldots, n_1, \ l = 1, \ldots, n_2\} \subset \mathbb{R}^2 \) with \( |Ω| = n_1n_2 := N \) given by

\[
\text{find } u \quad \text{s.t. } \ b = N(Au),
\]

where \( u \in \mathbb{R}^N \) and \( b \in \mathbb{R}^M \) are the vectorised unknown image and observed data, respectively, \( A \in \mathbb{R}^{M \times N} \) is the (known) linear forward operator and \( N : \mathbb{R}^M \to \mathbb{R}^M \) stands for the degradation operator modelling the presence of noise in \( b \). Some classical examples for \( A \) are, for instance, convolution (blurring), Radon transform and under-sampling operators. As it is known from standard books in inverse problems (see, e.g., [57, 124]), 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 \( u^* \) of \( u \) as accurate as possible by solving a new, well-posed problem where some a priori information on \( u^* \) is encoded in the form of a regularisation term. In this work, we will focus our attention on the family of variational regularisation methods where the reconstructed image \( u^* \in \mathbb{R}^N \) is computed as a minimiser of a suitable cost functional \( J : \mathbb{R}^N \to \mathbb{R} \) such that the problem can be formulated as

\[
\text{find } u^* \in \text{arg min}_{u \in \mathbb{R}^N} \left( J(u;\mu) := \mathcal{R}(u) + \mu \mathcal{F}(Au,b) \right).
\]

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

1.2 A leading actor: TV regularisation. 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 semi-norm which is suited for describing meaningful image contents such as image discontinuities (edges). In its simplest discrete form, it is defined as the following non-smooth and convex regulariser

\[
\mathcal{R}(u) = \text{TV}(u) := \sum_{i=1}^{N} \|D_i u_i\|_2,
\]

(TV)
where, for each pixel \( i = 1, \ldots, N \), \((Du)_i = (D_h u, D_v u)_i \in \mathbb{R}^2\) stands for the discrete gradient of image \( u \) at pixel \( i \), with \( D_h, D_v \in \mathbb{R}^{N \times N} \) suitable finite difference operators discretising the partial derivatives of image \( u \) along the horizontal and vertical directions, respectively.

The use of TV regularisation in imaging was firstly proposed by Rudin, Osher and Fatemi in [118] which is nowadays probably among the most cited papers in mathematical imaging\(^1\). In the 90’s, 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, sparse gradient-based regularisation 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 years (see, e.g., [33, 31, 17] 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 [38]).

Despite the large interest and thorough understanding towards the analytical and regularisation properties of TV regularisation (we will list the main contributions in both directions in due course), such regulariser also presents significant limitations. A 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., [32, 106, 77] for some analytical studies. Moreover, as observed, e.g., in [123, 101], TV reconstructions also suffer from loss of contrast artefacts, even in the case of noise-free observed images. Another major limitation of TV as is its *global* or *space-invariant* behaviour, that is the fact that the contribution at each pixel \( i = 1, \ldots, N \) in (TV) to the whole regularisation takes exactly the same functional form. Due to this ‘rigidity’, the TV regulariser is not suited to describe possibly very heterogeneous local image structures encountered, for instance, in natural images. Furthermore, such form is not adapted to situations where clear *directionality* (either global [6, 82] or local [133]) appears.

We will now review the main contributions proposed in the literature over the last decades to improve upon the regularisation capabilities of TV and, in particular, to reduce the aforementioned drawbacks within the class of global and space-invariant regularisers. Next, we will discuss on 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 generalisations

In order to reduce some of the TV reconstruction drawbacks highlighted above, several space-invariant generalisations have been proposed over the years, see, e.g., [41, 40, 110, 121, 131] 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 [36] and the Total Generalised Variation (TGV) regulariser introduced by Bredies, Kunisch and Pock in [15]. This latter regulariser shares some favourable properties with TV, such as rotational invariance, lower semi-continuity 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 shown to be very effective in practice and thoroughly analysed from both an analytical and numerical viewpoint, the question of how to overcome the intrinsic ‘rigidity’ of first-order TV-type regularisation models is still very much open. With the intent of adapting the TV-type regularisation to structural image information, in [6] Bayram and Kamasak proposed a Directional TV (DTV) regulariser for image denoising, whose analytic form reads

\[
\text{DTV}(u) := \sum_{i=1}^{N} \| \Lambda_a R_{-\theta} (Du)_i \|_2, \quad a \in (0, 1], \quad \theta \in [-\pi/2, \pi/2). \quad (1.3)
\]

Here, \( \theta \in [-\pi/2, \pi/2) \) denotes the dominant orientation in the target image \( u \), \( R_{-\theta} \) denotes the rotation matrix of angle \(-\theta\), while \( \Lambda_a = \text{diag}(1, a) \) is a diagonal matrix which encodes the strength of the regularisation along the direction (\( \cos \theta, \sin \theta \)) and its orthogonal depending on a parameter \( a \in (0, 1] \) whose value range from enforcing a full isotropic modelling (no directional preference) for

---

\(^1\)15749 citations according to Google Scholar.
Figure 1: Test image from [83] 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).

\( a = 1 \) to a strong anisotropy one for \( a \approx 0 \). The very same modelling can be analogously used to define a Directional variant of the TGV regulariser, dubbed DTGV, as done in [83] by Kongskov, Dong and Knudsen. The use DTG and DTGV is indeed beneficial in applications related to fibres, such as the study of glass fibres in wind-turbine blades [119] and the scanning of optical fibres from computed tomography (CT) scans [78].

To illustrate the benefits of incorporating directional information defined in terms of a dominant orientation \( \theta \) in the regulariser, we consider the test image in Figure 1a, previously used in [83], in which the existing piece-wise 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 1b and the scatter plot in Figure 1c show, as expected, global alignment along the perpendicular direction \( \mathbf{v}^\perp = (-\sin \theta, \cos \theta) \). It appears natural for this type of images to design a regularisation 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 regularisation has been explored in [3, 84, 90]. By adopting a statistical viewpoint, in these works the authors show how the use of TV implicitly corresponds to assume a space-invariant one-parameter half-Laplacian distribution (hLd) for the gradient magnitudes of the target image \( u \), that appears to be in general too restrictive to model the distribution of gradient magnitudes in natural images for which a more general half-Generalised Gaussian distribution (hGGd) prior should be considered instead. This choice corresponds to employ the following \( TV_p \) regularisation model

\[
TV_p(u) := \sum_{i=1}^{N} \| (D \mathbf{v})_i \|_2^p, \quad p \in (0, 2],
\]

where the exponent \( p \in (0, 2] \) is a free parameter providing the \( TV_p \) regulariser with higher flexibility than the TV regulariser. Its setting is indeed related to the properties of the image of interest: it can be fixed either empirically to enhance sparsity (\( p < 1 \), see [84]) or regarded as an information tailored to the image itself which should thus be estimated appropriately. The \( TV_p \) regulariser has proved to be effective for the solution of several imaging problems ranging from labelling and segmentation [96] to blind deblurring [84, 97] and synthetic aperture radar (SAR) image despeckling [66] and many more. Its performance strongly depends on the selected/estimated value of \( p \), whose setting may be hard in case of very heterogeneous images composed, for instance, by both smooth and piece-wise constant regions.

Finally, we recall that a further albeit classical limitation of global TV-type regularisation consists in the choice of the optimal regularisation parameter \( \mu \) in (1.2). As a matter of facts, this is a common challenge for all regularised 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 discrepancy principle [68, 69], generalised cross validation [63, 60], L-curve analysis [23] and unbiased risk estimators (SURE) [122, 92]. Note that while being effective in practice, these approaches often require the prior knowledge of the noise level in the data, which in many practical
applications is a difficult information to obtain. To avoid this issue, different techniques based, for instance, on bilevel learning [85, 19, 44, 71, 72] or on statistical whiteness principles can be used [91, 89].

1.4 Incorporating space variance

While the aforementioned modelling appears very restrictive from a global viewpoint, it is a natural question wondering 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 regulariser.

We motivate this idea by showing in Figure 2a an enlightening example concerned with local description of directional features for the popular test image image 

\textit{barbara}. We start selecting three sub-regions of interest: two of them are characterised by geometric textures - see Figures 2d,2g - while the other presents smooth and homogeneous details - see Figure 2j. The global image histogram and plot in Figures 2b, 2c clearly show that image gradients are not oriented along a single dominant direction; rather, they appear to be non-uniformly spread over a box-shaped neighbourhood 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 2e,2h and in the scatter plots in Figures 2f,2i. Finally, for the gradients computed within the smooth region, a different configuration occurs, being them not aligned along any preferred direction - see Figure 2k, but rather concentrated in a neighbourhood of the origin, due to the large homogeneous image content. As observable from the scatter plot in Figure 2l, dispersion around the origin is however smaller with respect to the global scatter plot.

Similarly, examples illustrating the different local behaviour of the $\text{TV}_p$ regulariser (1.4) depending on the particular (homogeneous VS. textured) image patch considered could be given, as we will extensively report in the following sections. Those examples motivate and justify the idea of formulating a new space-variant framework allowing for the description of image contents at a local scale.

The idea of incorporating space-variant information is in fact 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 [115, 129, 117, 120, 130] and statistical approaches [51, 116, 29, 26, 25, 114, 134]. In the last couple of years and under a different perspective, few contributions have also been made in the context of (deep) learning approaches for imaging [64, 99, 100, 102, 86]. In the following, we summarise the ideas contained in these works by structuring our description by topical subsections, to favour readability. As the reader will notice, however, a detailed description of the (large) class of space-variant variational image regularisation 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 modelling will be introduced.

1.4.1 PDE approaches

In the context of PDE approaches for imaging, the idea of space variance has enriched standard linear and non-linear diffusion-type models starting from the work of Perona and Malik in the 90’s [115]. There, the authors proposed a nonlinear space-adaptive diffusion method for avoiding the blurring and localisation 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 $x \in O \subset \mathbb{R}^2$, where by $O$ we denote a regular image domain, an edge-stopping diffusion function depending locally on the quantity $|\nabla u(x)|$ was used as a likelihood measure. Note that the Perona-Malik filter is a particular instance of the diffusion model

\[
\begin{cases}
    u_t = \text{div}\{W(|\nabla u|)\nabla u\} & \text{on } O \times (0, \infty),
    \\
    u(x, 0) = b(x) & \text{on } O,
    \\
    (\nabla u, n) = 0 & \text{on } \partial O \times (0, T),
\end{cases}
\] (1.5)

where $n$ stands for the outward normal vector on $\partial O$ and $b$ stands for the observed image. Problem (1.5) 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
Figure 2: Test image *barbara* and selected sub-regions (left column), 2D histograms of image gradient components (middle column), scatter plots of image gradient components (right column).

standard deviation parameter $\sigma = \sqrt{2t}$. Such operation corresponds to the well-known low-pass spectral filtering and it is commonly used for smoothing pictures by averaging values within a certain neighbourhood. In the general case, model (1.5) produces a family of images parametrised 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 $x \in O$, the function $W$ can be, for instance, chosen as

$$W(|\nabla u(x)|) = \frac{1}{1 + \left(\frac{|\nabla u(x)|}{K}\right)^2},$$  \hspace{1cm} (1.6)$$

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 (1.5) turns into a heat equation which smooths out the noise. Close to edges and boundaries, the gradient magnitudes get larger instead, thereby the diffusion coefficient in (1.6) vanishes; as a result, in
correspondence of these pixels diffusion is not performed and meaningful structures are preserved.

However, despite their edge-adaptive behaviour, 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 $J(x) = \nabla u(x) \otimes \nabla u(x) \in \mathbb{R}^{2 \times 2}$. 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 $J(x)$ can be somehow synthesised by the elongation of the associated elliptical level curves, see, e.g., [129, 117, 120, 130, 65, 58].

### 1.4.2 Statistical approaches

Statistical approaches for image processing have become very popular in the last decades due to their ability to incorporate non-deterministic information in the forward model - see [28]. Here, the core idea is to model the unknown image $u$ as a random variable $U$ to highlight the intrinsic uncertainty about its value, which is also related to possible approximations of the model operator $A$, and of the noise degradation model $N$. The information or the beliefs available a priori on the random variable $U$ are encoded in the prior probability density function (pdf) $P(u)$. Analogously, the observed image $b$ is regarded as a realisation of a random variable $B$, whose behaviour for a fixed $u$ is encoded in the likelihood pdf $P(b \mid u)$. In this framework, the goal is to recover the distribution of $U$ according to the given observation $b$ and the underlying degradation model $A$; in terms of distributions, this translates in seeking the posterior pdf $P(u \mid b, A)$ which is related to prior and likelihood pdfs via the Bayes’ formula:

$$P(u \mid b, A) = \frac{P(b \mid Au)P(u)}{P(b)}, \quad (1.7)$$

where $P(b)$ is often referred to as the evidence term and plays the role of a normalisation constant.

The poorer is the information on the degradation model, the more relevant is the design of a suitable prior for the unknown image. The prior pdf $P(u)$ can model different characteristics of the image, ranging from the presence of textures to boundary configurations. Popular priors for image restoration problems encode information on the distribution of the grey levels within an image and the transition of grey-scale intensities between different areas of the image [95].

In [61], the authors interpreted the pixel-grey levels as states of atoms in a lattice-like physical system, so that the unknown image is modelled as a Markov Random Field (MRF). This translates into the requirement that a selected feature at the generic pixel $i$ of $u$ only depends on the behaviour of $u$ at pixels belonging to a set $C_i$ of neighbours of $i$, called clique. When the selected feature is the grey level, the 2D Markovian property at pixel $i$ reads

$$P(U_i = u_i \mid U_j = u_j, \ j \neq i) = P(U_i = u_i \mid U_j = u_j, \ j \in C_i), \quad (1.8)$$

where by $P(\cdot)$ we denote the probability density function (pdf). The prior distribution for a MRF is the so-called *Gibbs prior*

$$P(u) = \frac{1}{Z} \exp \left( - \sum_{i=1}^{n} V_{C_i}(u) \right), \quad (1.9)$$

where $Z > 0$ is a normalisation constant and $V_{C_i}$ is referred to as the *Gibbs potential function* defined on a clique of pixels centred at pixel $i$ - see also [10] for a more extensive discussion. The potential functions typically depend on a number of parameters that can be considered fixed in the case our prior beliefs are informative enough to allow for their manual setting, but which, in general, may vary from pixel to pixel.

As it will be explored in details in Section 3.2 there exists a strict relation between Gibbs’ priors and many notable regularisers, such as Tikhonov and TV, which are designed based on the properties of the discrete image gradients $Du$. Note that non-stationary (i.e. space-variant) MRF approaches have been also considered over the years (see, e.g., [51, 116, 95] for some applications) in a purely statistical framework, for which the design of efficient sampling strategies and Maximum Likelihood (ML) approaches is required.

By (1.7), the sought unknown image $u$ can be recovered as a single-point representative of $P(u \mid b, A)$ moving from a purely statistical to a more optimisation-based framework. A very
popular strategy for that goes under the name of Maximum A Posteriori (MAP) estimation [79]: it consists in summarising the posterior pdf with its mode, i.e.

\[
\mathbf{u}_{\text{MAP}} \in \arg \max_{\mathbf{u} \in \mathbb{R}^N} \{ \mathbb{P}(\mathbf{b} \mid \mathbf{u}, A) \mathbb{P}((\mathbf{u})) \}
\]

\[
= \arg \min_{\mathbf{u} \in \mathbb{R}^N} \{ -\ln \mathbb{P}(\mathbf{b} \mid \mathbf{u}, A) - \ln \mathbb{P}((\mathbf{u})) \}, \quad (1.10)
\]

where the evidence term \(\mathbb{P}(\mathbf{b})\) has been neglected since it does not depend on \(\mathbf{u}\). We remark that connections between the statistical interpretation of the inverse problem (1.1) via MAP formulation (1.10) and its analytical counterpart (1.2) have been nowadays drawn for a large variety of \textit{space-invariant} regularisation models, while they are generally hard to be understood in the context of \textit{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 modelling as well as computational differences have to be taken into account as the explicit dependence on the distribution hyperparameters in (1.9) cannot be neglected any longer.

In space-variant settings, a very natural alternative to a completely supervised strategy, i.e. the parameters of the non-stationary 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 optimisation. Among the many contributions in this field, we mention [29, 26, 25], where the authors propose an iterative alternating scheme for the solution of the hierarchical MAP formulation for sparse recovery problems. A parameter marginalisation, followed by a small variance analysis is employed instead in [114] for image inpainting applications.

We further mention that the classical literature on \textit{compressed sensing} algorithms has been revisited and interpreted in probabilistic terms; in this perspective, the popular \(\ell_1\)-regularisation 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 [134] and references therein.

1.4.3 Generative and unfolded learning-based approaches

Many shortcomings are classically associated to the use of model-driven approaches. Among them, the dependence on the (supposedly known) forward model operator \(A\), 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 \textit{a-priori} fixed models and distributions have been shown to represent major limitations which have been overcome over the recent years by replacing knowledge-based by data-driven designs. It is outside the scopes of this review providing 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 problem. For that, we refer the reader to the recent review paper [2] where these questions are addressed in a thorough way and where an extensive literature review is given.

For the following description, we will limit ourselves to consider two nowadays extremely popular classes of data-driven approaches which, in some sense, are in close connection with the space-variant modelling discussed in this work. The former class, introduced firstly in [64] under the name of Generative Adversarial Networks (GAN), exploit training examples to estimate, rather than the desired solution itself, the distribution it is sampled from (ideally (1.9)) by means of the interplay between two adversarial networks which force the joint machinery to discriminate between ‘true’ data distribution and its ‘opponent’ (adversarial) attacks. Differently from the fully model-driven statistical approaches summarised 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 modelling, 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 favoured their use in a large variety of imaging problems, see, e.g., [99, 100, 102, 86] and is still a growing research area in the field which could inspire and complement the classical analytical/statistical knowledge-based modelling in future work (see Section 11 for research outlooks).
We further recall a different class of learning-based methods which have become very popular due to their easy interpretation from an optimisation viewpoint. Heuristically, such approaches are indeed inspired and thus made understandable by the operative expression of the iterative solution of problem (1.2) whose algorithmic solver can be unrolled/unfolded by means of deep-learning architectures (see [103] for a review). As showed recently in some works, see, e.g., [43, 9, 76], this approach provides indeed an interpretable framework for the data-driven estimation for several model hyperparameters, such as diffusion filters, algorithmic step-sizes and many more, thus combining well-known notions of optimisation (such as gradient/proximal-gradient updates, algorithmic parameters) with more learning-based concepts (activation functions, learning rates...).

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 models hyperparameters have been considered. In [125], for instance, directional TV regularisation on the sinogram data is proposed to inpaint the missing range of angles and improve the inversion process, while in [52] 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 here to [1, 54] where local regularisation weighting as well as directional (therein often called structural) strategies extending those used in [80] for Electrical Impedence Tomography (EIT) are used. Analogous approaches have further been considered in [56, 55] 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 [16] for blind hyperspectral imaging and in [5] for magnetic particle imaging.

Within the class of medical imaging applications, we further mention Photo-Acoustic Tomography (PAT) imaging, for which in [11] a space-variant modelling 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, in particular, of local directional dependence, we mention here the work carried out in [62, 98, 20, 111] where non-invasive 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 colorisation, 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 modelling 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 with very different names (e.g. non-stationary 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 many different models in terms of a new, generalised Bayesian modelling which allows also for some original extensions which have not explicitly studied before. The Bayesian framework described in this work paves indeed the way for the design of new, unexplored strategies helpful to design flexible and adaptive image regularisation 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 omit to provide details on the use of a (generalised) discrepancy principle strategy needed to compute the hyperparameters associated to the likelihood functionals. We further stress that TV regularisation (TV) is here taken as a reference regularisation model in this work due to the incredible amount of contributions developed over the last thirty 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 type of regularisation functionals, as we will shortly comment in the final Section 11 of this work.

1.6 Structure of the paper

The paper is organised 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 (non-stationary) priors together with the likelihood pdf corresponding to the class of noise models considered in this review. The properties of the regularisers induced by the space-variant priors will be analysed from a modelling and optimisation 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 regularisers proposed are combined with general data fidelity terms and the suitable prior distributions on the model hyperparameters (i.e. hyperpriors). In Section 7, we address the hyperparameter estimation problem by designing robust maximum likelihood-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 regularisers 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. To conclude, 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 Notations and preliminaries

We will use the notation \( \mathbb{R}_+ \) and \( \mathbb{R}_{++} = \mathbb{R}_+ \setminus \{0\} \) for the set of non-negative 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 \( M \in \mathbb{R}^{d \times d} \), \( d > 1 \), we will denote by \( |M| \) the determinant of \( M \).

To indicate multi-variate random variables and their realisations we will use bold capital/lower-case letters, e.g. \( X \) and \( x \), and we denote by \( P_X, P_x, \eta_X = \mathbb{E}(X), \Sigma_X \) the probability mass function, pdf, mean and covariance matrix of the random variable \( X \), respectively. We will omit the subscript \( X \) if not necessary. The characteristic and the indicator function of a set \( S \) are defined as

\[
\chi_S(x) := \begin{cases} 
1 & \text{if } x \in S \\
0 & \text{otherwise}
\end{cases}, \quad \iota_S(x) := -\ln \chi_S(x) = \begin{cases} 
0 & \text{if } x \in S \\
+\infty & \text{otherwise}
\end{cases}, \quad (2.1)
\]

respectively. Moreover, we denote by \( \Gamma \) the Gamma function, which is defined as follows:

**Definition 1** (Gamma and incomplete Gamma functions). The lower and upper incomplete Gamma functions, \( \Gamma \) and \( \Gamma \) respectively, are defined by

\[
\Gamma(x,y) = \int_0^y t^{x-1}e^{-t}dt, \quad \Gamma(x,y) = \int_y^{+\infty} t^{x-1}e^{-t}dt, \quad (x,y) \in \mathbb{R}_{++} \times \mathbb{R}_+ . \quad (2.2)
\]

The (complete) Gamma function \( \Gamma \) is

\[
\Gamma(x) = \lim_{y \to +\infty} \Gamma(x,y) = \Gamma(x,0) = \int_0^{+\infty} t^{x-1}e^{-t}dt, \quad x \in \mathbb{R}_{++} . \quad (2.3)
\]

Now, we recall the definitions of few well-known distributions to which we are often referring throughout the discussion and that will be mainly employed in the modelling of the regularisation terms reviewed here.

**Definition 2** (Univariate Laplacian distribution). A scalar random variable \( X \) is Laplacian-distributed with mean \( \eta \in \mathbb{R} \) and scale parameter \( \gamma \in \mathbb{R}_{++} \), denoted by \( X \sim \text{L}(\eta, \gamma) \), if its pdf has the form

\[
P(x|\eta, \gamma) = \frac{\gamma}{2} \exp(-\gamma |x-\eta|), \quad x \in \mathbb{R} . \quad (Ld)
\]
Finally, the standard deviation $\sigma$ scale parameter strongly convex hence it admits a unique minimiser. In this case, proximal operator in convex optimisation contexts (see, e.g., [45]). The proximal operator of $f$ with proximity parameter $\beta$ is the set-valued function $\text{prox}_\beta^f: \mathbb{R}^N \rightrightarrows \mathbb{R}^N$ defined for any $w \in \mathbb{R}^N$ by

$$\text{prox}_\beta^f(w) := \arg \min_{x \in \mathbb{R}^N} \left\{ f(x) + \frac{\beta}{2} \|x - w\|_2^2 \right\} \subset \mathbb{R}^N. \quad (2.7)$$

Note that if $f$ in the Definition above is convex, then the minimisation problem in (2.7) is strongly convex hence it admits a unique minimiser. In this case, $\text{prox}_\beta^f(\cdot)$ is a well-defined function from $\mathbb{R}^N$ to itself which coincides with the well-studied proximal operator frequently encountered in convex optimisation contexts (see, e.g., [45]).

**Definition 3** (Univariate Generalised Gaussian distribution). A scalar random variable $X$ is generalised Gaussian-distributed with mean $\eta \in \mathbb{R}$, scale parameter $\gamma \in \mathbb{R}_{++}$ and shape parameter $s \in \mathbb{R}_{++}$, denoted by $X \sim \text{GG}(\eta, \gamma, s)$, if its pdf has the form

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

with $\Gamma(\cdot)$ denoting the Gamma function defined in (2.3). 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

$$\lim_{s \to +\infty} p(x|\eta, \gamma, s) = \frac{1}{2} \chi_{[0,1]}(|x-\eta|). \quad (2.4)$$

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

$$\sigma = (1/\gamma) \sqrt{\Gamma(3/s)/\Gamma(1/s)}. \quad (2.5)$$

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

**Definition 4** (Bivariate Generalised Gaussian distribution). A bivariate random variable $X$ is generalised Gaussian-distributed with mean $\eta \in \mathbb{R}^2$, symmetric positive definite covariance matrix $\Sigma \in \mathbb{R}^{2 \times 2}$ and shape parameter $s \in \mathbb{R}_{++}$, denoted by $X \sim \text{BGG}(\eta, \Sigma, s)$, if its pdf has the form

$$p(x|\eta, \Sigma, s) = \frac{1}{2\pi |\Sigma|^{1/2} \Gamma(2/s)^{2/s}} \exp\left(\frac{-1}{2} (x-\mu)^T \Sigma^{-1} (x-\mu) \right)^{2/s}, \quad (\text{BGGd})$$

We now provide definitions for the Laplace distribution (Ld) and the Generalised Gaussian one (GGd) when the scalar random variable $X$ is known to be non-negative.

**Definition 5** (Univariate Half Laplacian distribution). A scalar random variable $X$ is Half Laplacian-distributed with scale parameter $\gamma \in \mathbb{R}_{++}$, denoted by $X \sim \text{hL}(\gamma)$, if $X = |Y|$, with $Y \sim \text{L}(0, \gamma)$. The pdf of $X$ takes the form

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

**Definition 6** (Univariate Half Generalised Gaussian distribution). A scalar random variable $X$ is Half Generalised Gaussian-distributed with scale parameter $\gamma \in \mathbb{R}_{++}$ and shape parameter $s \in \mathbb{R}_{++}$, denoted by $X \sim \text{hGG}(\gamma, s)$, if $X = |Y|$, with $Y \sim \text{GG}(0, \gamma, s)$. The pdf of $X$ takes the form

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

We also give the definition of Gamma distribution which will be used in the next sub-section and in the Appendix.

**Definition 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 $X \sim \text{Gam}(\nu, z)$, if its pdf has the form

$$p(x|\nu, z) = \begin{cases} \frac{1}{\Gamma(z)} x^{z-1} \exp(-\frac{x}{z}) & \text{for } x \in \mathbb{R}_+, \\ 0 & \text{otherwise} \end{cases}. \quad (2.6)$$

Finally, we recall the definition of proximal operator to which we will extensively refer in Sections 4-8:

**Definition 8** (proximal operator). Let $f: \mathbb{R}^N \to \mathbb{R}$ be a proper, lower semi-continuous and possibly non-convex function and let $\beta \in \mathbb{R}_{++}$. The proximal operator of $f$ with proximity parameter $\beta$ is the set-valued function $\text{prox}_\beta^f: \mathbb{R}^N \rightrightarrows \mathbb{R}^N$ defined for any $w \in \mathbb{R}^N$ by

$$\text{prox}_\beta^f(w) := \arg \min_{x \in \mathbb{R}^N} \left\{ f(x) + \frac{\beta}{2} \|x - w\|_2^2 \right\} \subset \mathbb{R}^N. \quad (2.7)$$

Note that if $f$ in the Definition above is convex, then the minimisation problem in (2.7) is strongly convex hence it admits a unique minimiser. In this case, $\text{prox}_\beta^f(\cdot)$ is a well-defined function from $\mathbb{R}^N$ to itself which coincides with the well-studied proximal operator frequently encountered in convex optimisation contexts (see, e.g., [45]).
2.1 Generalised Discrepancy Principle

Assuming that the noise degradation operator $N$ in (1.1) models the action of an additive, zero-mean, independent and identically distributed (i.i.d.) generalised Gaussian (in short, AIGG) noise, we have that (1.1) can be rewritten as

$$b = Au + e,$$

where, based on Definition 3, $e \in \mathbb{R}^M$ is the vector of realisations of the $M$-variate random variable $E$ whose components are i.i.d. GG random variables with shape parameter $q \in \mathbb{R}^+$ and scale parameter $\omega \in \mathbb{R}^+$, the latter encoding information on the noise standard deviation according to (2.5).

Before detailing a Generalised Discrepancy Principle (in short, GDP) useful to define an automatic selection strategy for the regularisation parameter $\mu$ in (1.2) under the modelling assumption (2.8), we report the following result, whose proof is based on classical probability arguments and given for completeness in the Appendix.

**Proposition 1.** If $X_i \sim \text{GG}(0,\omega,q)$, $i = 1,\ldots,M$, with $\omega,q \in \mathbb{R}^+$, are independent random variables, then we have

$$Y = \| (X_1,\ldots,X_M) \|_q^q = \sum_{i=1}^{M} |X_i|^q \sim \text{Gamma}(\nu,z), \quad \nu = \frac{1}{\omega q}, \quad z = \frac{M}{q}. \quad (2.9)$$

The random variable $Y$ has mean $\eta_Y$ and variance $\sigma_Y^2$ whose expressions are given by

$$\eta_Y = \frac{M}{q} \frac{1}{\omega^q}, \quad \sigma_Y^2 = \frac{M}{q} \frac{1}{\omega^{2q}}. \quad (2.10)$$

Thanks to Proposition 1, we can now provide the following GDP.

**Definition 9** (Generalised Discrepancy Principle). Let $u^*(\mu)$ be the parameter-dependent solution of (1.2) for model (2.8); denoting by $r^*(\mu) = Au^*(\mu) - b$ the associated residual image, we have that the Generalised Discrepancy Principle can be formulated as:

$$\text{Select } \mu = \mu^* \text{ such that } \| r^*(\mu^*) \|_q = \delta_q, \quad (2.11)$$

where

$$\delta_q := \tau E(\| E \|_q) = \left\{ \begin{array}{ll} \tau (M/q)^{1/q} (1/\omega) & \text{if } q < +\infty \\ \tau (1/\omega) & \text{if } q = +\infty \end{array} \right., \quad (2.12)$$

with $\tau \approx 1$.

Note that in the case $q = 2$, the strategy reduces to the classical discrepancy principle strategy detailed, e.g., in [104, 57] for i.i.d. Gaussian noise.

3 A flexible Bayesian framework

In this section, we recall the general Bayesian framework outlined in Section 1.4.2 and adapt it to our purposes and considerations. We start specifying the different noise degradation models $N$ 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 $P(b \mid Au, \Phi)$, where, in addition to $Au \in \mathbb{R}^M$ the dependence on the (generally unknown) likelihood hyperparameter vector $\Phi \in \mathbb{R}^k$ 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 9, we will focus our attention to the class of AIGGGN corresponding to model (2.8).

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 \((q = +\infty)\).

Due to the independence assumption for the univariate random variables \(E_j\) in (2.8), the \(M\)-variate likelihood pdf can be written as the product of \(M\) identical univariate GG pdfs (see Definition 3). When \(q < +\infty\), it thus takes the form:

\[
\mathbb{P}(b | Au, \Phi) = \prod_{j=1}^{M} \left( \frac{\omega}{2} \right)^{q/2} \Gamma(1/q) \exp\left(-\frac{\omega^q}{q} |(Au)_j - b_j|^q\right)
\]

while for \(q = +\infty\) it reads

\[
\mathbb{P}(b | Au, \Phi) = \prod_{j=1}^{M} \left( \frac{\omega}{2} \right)^{q/2} \Gamma(1/q) \chi_{[0,1]}(|(Au)_j - b_j|)
\]

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

### 3.2 Priors

Recalling the statistical modelling introduced in Section 1.4.2 and in particular the MRF structure in (1.9), we proceed similarly as in the previous section and make explicit the dependence of the prior pdf on the vector of prior hyperparameters \(\Theta\) involved in its expression, which here are assumed to be unknown. The corresponding Gibbs’ prior reads

\[
\mathbb{P}(u | \Theta) = \frac{1}{Z(\Theta)} \exp\left(-\sum_{i=1}^{n} \Phi(i \cdot u; \Theta)\right),
\]

where \(Z(\Theta) \in \mathbb{R}_{++}\) is a normalisation constant depending on the unknown parameters \(\Theta\) while \(\Phi(i \cdot u; \Theta)\) is the Gibbs’ potential on the \(i\)-th clique \(C_i\).

Recalling the general Markovian property (1.8) and thinking of the description of the image in terms of its local gradients \((Du)\), discretised by standard first-order forward finite differences, we have that (1.8) turns into

\[
\mathbb{P}(U_i = u_i \mid U_j = u_j, j \neq i) = \mathbb{P}(U_i = u_i \mid U_{i, \text{right}} = u_{i, \text{right}}, U_{i, \text{down}} = u_{i, \text{down}}).
\]

For better illustration, we show the corresponding configuration of the generic clique in Figure 3. Condition (3.4) states that the potential function \(\Phi_C\) 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 \(i\).

Introducing the function \(z : \mathbb{R}^N \rightarrow \mathbb{R}_{++}^N\) defined by

\[
z(u) = [z_1(u), \ldots, z_N(u)]^T, \quad z_i(u) = ||(Du)_i||_2,
\]

and assuming that each image gradient magnitude \(x_i := z_i(u) = ||(Du)_i||_2 \in \mathbb{R}_{+}\) is the realisation 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 \(\mathbb{P}(u | \Theta)\) takes
the form of the Gibbs’ TV prior

$$P(u | \Theta) = c(\Theta) P(z(u) | \Theta) = c(\Theta) \prod_{i=1}^N P(z_i(u) | \Theta)$$

(3.6)

$$= c(\Theta) \prod_{i=1}^N (\alpha \exp(-\alpha \|D_i u\|_2))$$

(3.7)

$$= c(\Theta) \alpha^N \exp \left(-\alpha \sum_{i=1}^N \|D_i u\|_2\right), \text{ with } \Theta = \alpha \in \mathbb{R}_{++},$$

(3.8)

where the scalar normalisation function $c : \mathbb{R}_{++} \to \mathbb{R}_{++}$, depending only on $\Theta$, reads

$$c(\Theta) = \frac{1}{\int_{u \in \mathbb{R}^N} \alpha^N \exp \left(-\alpha \sum_{i=1}^N \|D_i u\|_2\right) du}.$$  

(3.9)

Notice that the presence of $c(\Theta)$ guarantees that the prior pdf in (3.8) sums up to one 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 $\alpha_i$. The corresponding space-variant hLd probability density thus reads in this case

$$P(x_i | \alpha_i) = \begin{cases} \alpha_i \exp(-\alpha_i x_i), & x_i \geq 0 \\ 0, & x_i < 0 \end{cases}, \quad i = 1, \ldots, N, \quad \alpha_i \in \mathbb{R}_{++},$$

(hLd-sv)

and yields the following non-stationary prior pdf on $u$

$$P(u | \Theta) = c(\Theta) P(z(u) | \Theta)$$

(3.10)

$$= c(\Theta) \prod_{i=1}^N (\alpha_i \exp(-\alpha_i \|D_i u\|_2))$$

$$= c(\Theta) \left( \prod_{i=1}^N \alpha_i \right) \exp \left(-\sum_{i=1}^N \alpha_i \|D_i u\|_2\right), \quad \text{with } \Theta = \alpha \in \mathbb{R}_{++}^N,$$

(3.11)

with $z$ defined as in (3.5) and the normalisation function $c(\Theta)$ defined by

$$c(\Theta) = \frac{1}{\int_{u \in \mathbb{R}^N} \left( \prod_{i=1}^N \alpha_i \right) \exp \left(-\sum_{i=1}^N \alpha_i \|D_i u\|_2\right) du}.$$  

(3.12)
As shown in Figure 4a, 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 probability density function considered.

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

$$
P(x_i; \alpha_i, p_i) = \begin{cases} 
\alpha_i p_i \exp(-\alpha_i x_i^p) & x_i \geq 0 \\
0 & x_i < 0
\end{cases}, \quad \alpha_i, p_i \in \mathbb{R}_{++}.
$$

(hGGD-sv)

In Figure 4b 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 $p = +\infty$. 

---

Figure 4: First row: half Laplacian distribution for different values of $\alpha \in \mathbb{R}_{++}$ (a), half Generalised Gaussian distribution for different values of $p \in \mathbb{R}_{++}$ and $\alpha = 1$ (b). Second row: bivariate Generalised 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 Generalised Gaussian pdfs displayed in the second row.
for \( p = +\infty \). The prior on \( \mathbf{u} \) corresponding to the pdf in (hGGd-sv) reads

\[
\mathbb{P}(\mathbf{u} \mid \Theta) = c(\Theta) \mathbb{P}(\mathbf{z}(\mathbf{u}) \mid \Theta) = c(\Theta) \prod_{i=1}^{N} \left( \frac{\alpha_i p_i}{\Gamma(1/p_i)} \exp\left(-\alpha_i \|(\mathbf{Du})_i\|_2^{p_i}\right) \right)
\]

\[
= c(\Theta) \prod_{i=1}^{N} \left( \frac{\alpha_i p_i}{\Gamma(1/p_i)} \right) \exp\left(-\sum_{i=1}^{N} \alpha_i p_i \|(\mathbf{Du})_i\|_2^{p_i}\right), \quad \text{with } \Theta = (\alpha, p) \in \mathbb{R}^{N \times 2}_{++},
\]

where \( c(\Theta) \) now takes the form

\[
c(\Theta) = \frac{1}{\int_{\mathbf{u} \in \mathbb{R}^N} \left( \prod_{i=1}^{N} \frac{\alpha_i p_i}{\Gamma(1/p_i)} \right) \exp\left(-\sum_{i=1}^{N} \alpha_i p_i \|(\mathbf{Du})_i\|_2^{p_i}\right) d\mathbf{u}}.
\]

We stress that, despite their differences, the choices (hLd), (hLd-sv) and (hGGd-sv) correspond to the formulation of prior assumptions on the quantities \( \|(\mathbf{Du})_i\|_2 \), for \( i = 1, \ldots, N \). Clearly, this is somehow a 'rigid' choice since it does not exploit the two-dimensionality of local image gradients \( (\mathbf{Du})_i = ((\mathbf{D}_x u)_i,(\mathbf{D}_y u)_i) \) and, possibly, the correlation between their horizontal and vertical components \( (\mathbf{D}_x u)_i \) and \( (\mathbf{D}_y u)_i \). To do so, a different, possibly space-variant, prior assumption imposing \textit{a-priori} information on the local image gradient can be made. Namely, for all \( \mathbf{y}_i := ((\mathbf{D}_x u)_i,(\mathbf{D}_y u)_i) \) and for \( i = 1, \ldots, N \) one can assume that \( \mathbf{y}_i \) follows a Bivariate Generalised Gaussian distribution (BGgd, see (BGGd)) which is space-variant with respect to shape and scale and takes locally the form:

\[
\mathbb{P}(\mathbf{y}_i; p_i, \Sigma_i) = \frac{1}{2\pi |\Sigma_i|^{1/2} \Gamma(2/p_i)^{1/2}} \exp\left(-\frac{1}{2} \left(y_i^T \Sigma_i^{-1} y_i\right)^{p_i/2}\right).
\]

(BGGd-sv)

where, for every \( i \), the covariance matrix \( \Sigma_i \in \mathbb{R}^{2 \times 2} \) is symmetric positive definite with determinant \( |\Sigma_i| > 0 \). The associated non-stationary prior pdf on \( \mathbf{u} \) can thus be written as

\[
\mathbb{P}(\mathbf{u} \mid \Theta) = c(\Theta) \mathbb{P}(\mathbf{z}(\mathbf{u}) \mid \Theta) = c(\Theta) \prod_{i=1}^{N} \left( \frac{1}{2\pi |\Sigma_i|^{1/2} \Gamma(2/p_i)^{1/2}} \exp\left(-\frac{1}{2} \left((\mathbf{Du})_i^T \Sigma_i^{-1} (\mathbf{Du})_i\right)^{p_i/2}\right) \right)
\]

\[
= c(\Theta) \prod_{i=1}^{N} \left( \frac{1}{2\pi |\Sigma_i|^{1/2} \Gamma(2/p_i)^{1/2}} \exp\left(-\frac{1}{2} \sum_{i=1}^{N} \left((\mathbf{Du})_i^T \Sigma_i^{-1} (\mathbf{Du})_i\right)^{p_i/2}\right) \right),
\]

where now \( \mathbf{z}(\mathbf{u}) = \mathbf{Du} \), and the normalisation function \( c(\Theta) \) is defined by

\[
c(\Theta) = \frac{1}{\int_{\mathbf{u} \in \mathbb{R}^N} \left( \prod_{i=1}^{N} \frac{1}{2\pi |\Sigma_i|^{1/2} \Gamma(2/p_i)^{1/2}} \right) \exp\left(-\frac{1}{2} \sum_{i=1}^{N} \left((\mathbf{Du})_i^T \Sigma_i^{-1} (\mathbf{Du})_i\right)^{p_i/2}\right) d\mathbf{u}}.
\]

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

\[
\Sigma_i = V_i^T E_i V_i, \quad E_i = \begin{pmatrix} e_i^{(1)} & 0 \\ 0 & e_i^{(2)} \end{pmatrix}, \quad e_i^{(1)} \geq e_i^{(2)} > 0, \quad V_i^T V_i = V_i V_i^T = I_2,
\]

\[(\mathbf{Du})_i^T \Sigma_i^{-1} (\mathbf{Du})_i\]
where
\[
\widetilde{\Lambda}_i = \begin{pmatrix}
\tilde{\lambda}_i^{(1)} & 0 \\
0 & \tilde{\lambda}_i^{(2)}
\end{pmatrix} := E_i^{-1/2} = \begin{pmatrix}
1/\sqrt{e_i^{(1)}} & 0 \\
0 & 1/\sqrt{e_i^{(2)}}
\end{pmatrix},
\] (3.21)
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 \( \alpha \in \mathbb{R}_+^N \) and \( a \in (0, 1]^N \) with components
\[
\alpha_i := \tilde{\lambda}_i^{(1)} \in \mathbb{R}_+, \quad a_i := \frac{\tilde{\lambda}_i^{(2)}}{\tilde{\lambda}_i^{(1)}} \in (0, 1], \quad i = 1, \ldots, N,
\] (3.23)
we have that the matrix \( \widetilde{\Lambda}_i \) in (3.22) can be equivalently rewritten as
\[
\widetilde{\Lambda}_i = \begin{pmatrix}
1 & 0 \\
0 & \tilde{\lambda}_i^{(1)}/\tilde{\lambda}_i^{(2)}
\end{pmatrix} = \alpha_i \lambda_{a_i}, \quad \text{with} \quad \lambda_{a_i} = \begin{pmatrix}
1 & 0 \\
0 & a_i
\end{pmatrix}.
\] (3.24)
Combining altogether, we have that (3.22)-(3.24) entail that the term in (3.20) can be indeed written as
\[
\left((D)u^T \Sigma_{i}^{-1}(D)u\right)^{p_i/2} = \alpha_i^{p_i} \| \lambda_{a_i} R_{\theta_i} (D)u \|_{2}^{p_i}.
\] (3.25)
Furthermore, based on (3.19) and (3.22)-(3.24), we observe that:
\[
|\Sigma_i|^{-1/2} = |\mathbf{V}_i^T \mathbf{E}_i \mathbf{V}_i|^{-1/2} = |\mathbf{E}_i|^{-1/2} = |\widetilde{\Lambda}_i| = |\alpha_i^{2} \lambda_{a_i}| = \alpha_i^{2} a_i > 0.
\] (3.26)
Plugging now (3.25) and (3.26) into the expression (3.17), we obtain the following equivalent form
\[
\mathbb{P}(u | \Theta) = \frac{c(\Theta)}{(2\pi)^{N}} \left( \prod_{i=1}^{N} \frac{\alpha_i^{2} p_i a_i}{\Gamma(2/p_i) 2^{2/p_i}} \right) \exp \left( -\sum_{i=1}^{N} \alpha_i^{p_i} \| \lambda_{a_i} R_{\theta_i} (D)u \|_{2}^{p_i} \right)
\] (3.27)
where the vector of hyperparameters is here:
\[
\Theta = (\alpha, p, \theta, a) \in \mathbb{R}_+^{N+2} \times [-\pi/2, \pi/2)^{N} \times (0, 1]^N.
\] (3.28)

Compared to the univariate prior (3.14), prior (3.27)-(3.28) is characterised by two additional vectors of (space-variant) parameters \( \theta_i \in [-\pi/2, \pi/2) \) and \( a_i \in (0, 1], \quad i = 1, \ldots, N \). These parameters relate in fact to the bivariate nature of the BGGd in (BGGd-sv). In particular, The parameter \( \theta_i \) 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 locally maximal isotropic pdf, whereas for \( a_i \approx 0 \) the contour lines approach lines drawing the angle \( \theta_i \) w.r.t. to the horizontal axis, hence they are maximally anisotropic. The great flexibility of distribution in (BGGd-sv) is highlighted in Figures 4c-4e, where the pdfs corresponding to the choice of different scalar parameters \( a_i, p_i, a_i, \) and \( \theta_i \) are shown, while the corresponding contour plots are displayed in Figures 4f-4h.

**Remark 1.** Note that the non-stationary prior in (3.11) 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 (3.14), we recover the space-invariant prior corresponding to the TV \( p \), regulariser in (1.4). The same consideration holds for the DTV regularisation term in (1.3), whose statistical counterpart is obtained starting from (3.27) and setting \( a_i = \alpha, p_i = 1, \theta_i = \theta \) and \( a_i = \alpha, \forall i \).

### 3.3 Hierarchical modelling

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 $\Theta$. The choice of recasting the original problem in probabilistic terms makes very natural to model the unknown vector $\Theta$ as well as the unknown $u$, as random variables. To do so, we thus need to introduce a further pdf encoding the \textit{a priori} beliefs on $\Theta$, which, in the following, will be denoted by $P(\Theta)$ and which will be referred to as \textit{hyperprior}.

By proceeding as in (1.7), we seek for the analytic expression of the joint posterior pdf, which, by leaving the dependence on $\Theta$ explicit, is related to the prior and likelihood pdf through

$$P(u, \Theta \mid b) = \frac{P(u, \Theta)P(b \mid Au)}{P(b)} = \frac{P(u \mid \Theta)P(\Theta)P(b \mid Au)}{P(b)},$$

(3.29)

where we have used $P(u, \Theta) = P(u \mid \Theta)P(\Theta)$. Proceeding by standard MAP estimation, we thus have that the sought solution pair $\{u^*, \Theta^*\}$ is the one maximising $P(u, \Theta \mid b)$, i.e.

$$\{u^*, \Theta^*\} \in \mathop{\arg\max}_{u, \Theta} \{P(u \mid \Theta)P(\Theta)P(b \mid Au)\},$$

(3.30)

or, equivalently,

$$\{u^*, \Theta^*\} \in \mathop{\arg\min}_{u, \Theta} \{-\ln P(u \mid \Theta) - \ln P(\Theta) - \ln P(b \mid Au)\},$$

(3.31)

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

When tackling the joint model (3.31), two major difficulties arise, namely the computation of the highly-dimensional constant $c(\Theta)$ and the choice of an efficient algorithmic scheme for the numerical solution of the minimisation problem (3.31). Different strategies have been designed to overcome the former issue: most of them are based on a modification of the conditional prior $P(\Theta \mid u)$ which comes from either approximating $c(\Theta)$ (see [108, 4, 3]) or neglecting it (see [132]). Here, we adopt this latter approach so that the joint hypermodel (3.31) takes the form:

$$\{u^*, \Theta^*\} \in \mathop{\arg\min}_{u, \Theta} \{-\ln P(z(u) \mid \Theta) - \ln P(\Theta) - \ln P(b \mid Au)\}.$$  

(3.32)

Neglecting $c(\Theta)$ 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 (3.31) will require to deal explicitly with the computation of such constant by means, for instance, of analogous approaches as those described in [126, 49].

From a numerical perspective, the solution of problem (3.32) can be addressed in different manners. A standard strategy illustrated in [27] is based on the design of an Iterated Sequential Algorithm (IAS) which, for $k \geq 0$ and upon a suitable initialisation for $u^{(0)}$, reads:

$$\Theta^{(k+1)} \in \mathop{\arg\min}_{\Theta} \{-\ln P(z(u^{(k)}) \mid \Theta) - \ln P(\Theta)\}$$

(3.33)

$$u^{(k+1)} \in \mathop{\arg\min}_{u} \{-\ln P(z(u) \mid \Theta^{(k+1)}) - \ln P(b \mid Au)\},$$

(3.34)

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

4 The anatomy of space-variant regularisation models

In this section, we derive the explicit expressions of the negative log-prior term $-\ln P(z(u) \mid \Theta)$, appearing in the cost function of (3.32), 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 regulariser, dissecting its properties in terms of regularisation features and providing some intuitions on their sparsity promoting behaviour.
4.1 From non-stationary priors to space-variant regularisers

Recalling (TV), we start computing the negative logarithm of the stationary Gibbs’ TV prior in (3.8). We have:

\[-\ln P(z(u) | \Theta) = \alpha \text{TV}(u) - N \ln \alpha. \tag{4.1}\]

We now rewrite TV as

\[\text{TV}(u) = \sum_{i=1}^{N} f_{\text{TV}}((Du)_i), \tag{4.2}\]

where the space-invariant and non-parametric function \( f_{\text{TV}} : \mathbb{R}^2 \to \mathbb{R}^+ \) is defined by

\[f_{\text{TV}}(g_i) := \|g_i\|_2, \quad g_i = (g_{i,1}, g_{i,2}) \in \mathbb{R}^2, \tag{4.3}\]

and is referred to in the following as the TV gradient penalty function. As it is well-known, TV is bounded from below by zero, (non-strictly) convex, non-coercive due to \( \text{null}(D) \neq \{0\}_2 \) and non-smooth. This last property is indeed responsible of the good gradient sparsity-promoting effect of TV, which favours piece-wise constant solutions. The TV gradient penalty function \( f_{\text{TV}} \) in (4.3) is shown in Figure 5(a).

![Figure 5: Space-invariant gradient penalty function](a) \( f_{\text{TV}} \)

To analyse in detail the properties of the TV regulariser, it is useful to consider the 1D sections of the gradient penalty function \( f_{\text{TV}} \) 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) parametrisation 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 \( f_{\text{TV}} \) in (4.3) read

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

In Figure 5(b) we show one section of \( f_{\text{TV}} \) along the direction defined by the angle \( \varphi = 0 \), i.e., the \( x \)-axis. However, as the expression (4.4) does not depend on \( \varphi \), we deduce that the same Figure could be obtained by representing the section corresponding to any \( \varphi \), for any pixel location \( i \). The TV penalty \( f_{\text{TV}} \) in (4.3) - whence, the (TV) regulariser - 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 regulariser not capable to adapt the strength (associated to \( \alpha \) in (4.1)) nor the nature (associated to the exponent of the norm in (4.3)) of the gradient sparsity-promotion effect to the local contents of the image to be recovered.

In the following, we inspect how the non-stationary priors introduced in Section 3.2 can favour local regularisation features, namely strength, sharpness and directionality.
Figure 6: Graphs of the gradient penalty functions defined in (4.7), (4.11) and (4.15) for the space-variant (WTV), (WTV\textsubscript{sv}p) and (WDTV\textsubscript{sv}p) regularisers, respectively.

4.1.1 Local regularisation strength

Recalling Section 3.2, the first and probably the easiest way to make the (TV) regulariser spatially flexible consists in allowing for a different amount of regularisation at every pixel in the image. From a Bayesian perspective, this corresponds to assuming a non-stationary hL prior distribution
for the gradient magnitudes of \( u \). By computing the negative logarithm in (3.11), we have

\[
-\ln P(z(u) \mid \Theta) = WTV(u; \Theta) - \sum_{i=1}^{N} \ln \alpha_i,
\]

where the space-variant WTV regulariser is defined in terms of hyperparameters \( \Theta = \alpha \) and reads

\[
WTV(u; \alpha) := \sum_{i=1}^{N} \alpha_i \| (Du)_i \|_2, \quad \alpha \in \mathbb{R}_+^N.
\]

(WTV)

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

\[
WTV(u; \alpha) = \sum_{i=1}^{N} f_{WTV} ((Du)_i; \alpha_i), \quad \alpha_i \in \mathbb{R}_+,
\]

where the gradient penalty function \( f_{WTV} : \mathbb{R}^2 \rightarrow \mathbb{R}_+ \) depends now locally on the parameter \( \alpha_i \) and reads

\[
f_{WTV}(g_i; \alpha_i) := \alpha_i \| g_i \|_2, \quad g_i = (g_{i,1}, g_{i,2}) \in \mathbb{R}^2.
\]

Similarly as for the (TV) regulariser, (WTV) is still convex and non-differentiable. However, the sparsity-promoting effect can now be locally modulated thanks to the presence of the local weights \( \alpha_i \). To highlight this feature, we report in the first column of Figure 6 (i.e. Figures 6(a),(d),(g)) the graphs of the WTV gradient penalty function \( f_{WTV} \) defined in (4.7) 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 regulariser 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 5, with the one in Figure 6(a) coinciding with the TV penalty. Different values of the weight yield different slopes of the conical lateral surface - note that \( \| \nabla f_{WTV}(g_{1,1}, g_{1,2}) \|_2 = \alpha_i \) for any \( (g_{1,1}, g_{1,2}) \in \mathbb{R}^2 \setminus \{(0,0)\} \) - and, hence, different local regularisation strengths. The larger (smaller) is the local weight \( \alpha_i \), the more (less) strongly the WTV regulariser will force \( \| (Du)_i \|_2 \) 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 x- and y-axis in the 3D plots in Figures 6(a),(d),(g) and, for better readability, in Figure 6(j), left. Like TV, the WTV regulariser 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

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

Finally, one can notice from (4.8) and from Figure 6(j), left, that all sections are nothing but positively-scaled versions of the absolute value function of scale parameter \( \alpha_i \), i.e. of the TV sections in (4.4).

Due to its ability of promoting local TV smoothing, we remark that the WTV regulariser has been proposed and studied in several papers (e.g. [75, 71, 74] and many more) from an analytical point of view and motivated by means of analogous probabilistic arguments in [22].

### 4.1.2 Local regularisation sharpness

As previously mentioned, the weights \( \alpha_i \) in the (WTV) regulariser allow to locally tune the strength of the gradient-sparsity promotion effect of the regularisation which, by construction, is of fixed TV type. In fact, the presence of a global exponent 1 for the gradient norms in definition of WTV does not allow to change, neither globally nor 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 \( \rho_i > 0, \ i = 1, \ldots, N \), being them the exponents of the gradient norms in the (WTV) definition and corresponding to the local shape parameters of the associated hGG pdf - see Definition 6.
We proceed as above and compute the negative logarithm of the non-stationary hGG prior (3.14), thus getting
\[
- \ln P(u \mid \Theta) = \text{WTV}^\text{sv}_p (u; \Theta) - \sum_{i=1}^{N} \ln \frac{\alpha_i p_i}{\Gamma(1/p_i)} ,
\]
where the space-variant WTV\textsuperscript{sv}_p regulariser, depending on the hyperparameters \( \Theta = (\alpha, p) \), is defined by
\[
\text{WTV}^\text{sv}_p (u; \alpha, p) := \sum_{i=1}^{N} \alpha_i^p \| (Du)_i \|_{p_i}^p , \quad (\alpha, p) \in \mathbb{R}^{N \times 2}.
\]
Like (WTV), the (WTV\textsuperscript{sv}_p) regulariser can be rewritten in terms of a parametric, space-variant gradient penalty function, namely
\[
\text{WTV}_p (u; \alpha, p) = \sum_{i=1}^{N} f_{\text{WTV}^\text{sv}_p} ((Du)_i ; \alpha_i, p_i) , \quad (\alpha_i, p_i) \in \mathbb{R}_+^2 ,
\]
where the function \( f_{\text{WTV}^\text{sv}_p} (\cdot) \) now reads:
\[
f_{\text{WTV}^\text{sv}_p} (g_i ; \alpha_i, p_i) = \alpha_i^p \| g_i \|_{p_i}^p , \quad g_i = (g_i,1, g_i,2) \in \mathbb{R}^2 .
\]
Like (TV) and (WTV), the (WTV\textsuperscript{sv}_p) regulariser is bounded below by zero, continuous and non-coercive. However, its other regularity properties depend on the actual values of the parameters \( p_i \). If \( p_i \geq 1 \) for any \( i = 1, \ldots, N \), then WTV\textsuperscript{sv}_p is convex, whereas it is non-convex 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 non-smooth.

In the second column of Figure 6 we show the graph of the WTV\textsuperscript{sv}_p gradient penalty function \( f_{\text{WTV}^\text{sv}_p} \) defined in (4.11) 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 \( \alpha_j, \alpha_k, \alpha_l \) are the same as for the WTV penalties reported in the first column. In case of unitary scale and shape parameters - see Figure 6(b) - the WTV\textsuperscript{sv}_p penalty coincides with the TV penalty and, more in general, for \( p_j = p_k = p_l = 1 \), the WTV\textsuperscript{sv}_p penalties coincide with the WTV penalties. For non-unitary shape parameters, the WTV\textsuperscript{sv}_p penalty function can assume different shapes, ranging from non-convex and non-differentiable ones \((p_i < 1\), Figure 6(e)) to strongly convex and differentiable ones \((p_i > 1\), Figure 6(h)). The degree of freedoms encoded by the shape parameters \( p_i \) thus provide the WTV\textsuperscript{sv}_p regulariser 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 of smooth image regions, whereas \( p_i < 1 \) - typically, \( p_i \leq 0.5 \) - should be used in piece-wise constant regions to mitigate the undesirable contrast loss effect of TV.

Similarly as for the previous (TV) and (WTV) regularisers, let us now take a look at the 1D sections of the WTV\textsuperscript{sv}_p penalty in (4.11) for \( \varphi \in [-\pi, \pi) \), which read
\[
s_i(t ; \varphi) = \alpha_i^p |t|^{p_i} , \quad t \in \mathbb{R} , \quad i = 1, \ldots, N .
\]
By looking at the plot of such sections shown in Figure 6(j), centre, it is clear how the value of parameter \( p_i \) can substantially change the regularisation effect at each pixel. In particular, by comparing the red, black and magenta sections in Figure 6(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 penalised (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 penalised than for \( p_i = 1 \), whereas large gradients are more penalised. We finally remark that, like for TV and WTV, the WTV\textsuperscript{sv}_p sections in (4.12) do not depend on the direction angle \( \varphi \), hence the (WTV\textsuperscript{sv}_p) regulariser still falls in the class of isotropic regularisers. This is visually confirmed by the WTV\textsuperscript{sv}_p penalties shown in Figures 6(b),(e),(h), which are rotationally invariant (i.e. have circular level curves), and by the penalty sections along the \( x \)-axis and the \( y \)-axis, which coincide as it is evident from Figure 6(j), centre.

The WTV\textsuperscript{sv}_p regulariser has been first introduced in a simplified version, i.e., with \( \alpha_i = \alpha, \forall i \), and interpreted in a probabilistic framework in [88]. The general case with space-variant weights has been discussed in [87].
4.1.3 Local anisotropy

As shown above, the (TV), (WTV) and (WTVP) regularisers are isotropic. For this reason, such regularisers 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 characterised by well-defined orientations. As illustrated in Section 3.2, to circumvent this limitation, a non-stationary BGG prior can be assumed for modelling the local distribution of gradients characterised by well-defined orientations. As illustrated in Section 3.2, to circumvent this limitation, a non-stationary BGG prior can be assumed for modelling the local distribution of gradients of $u$.

By computing the negative logarithm of the non-stationary BGG prior in (3.27), we have

$$- \ln P(z(u) | \Theta) = \text{WDTV}^\Theta_p(u; \Theta) = - \sum_{i=1}^{N} \ln \left( \frac{\alpha_i^p}{\Gamma(2/p_i)} |2^{2/p_i}| \right) + N \ln(2\pi),$$

where the space-variant WDTV$^\Theta_p$ regulariser is defined in terms of the hyperparameters $\Theta = (\alpha, p, \theta, a)$ and reads

$$\text{WDTV}^\Theta_p(u; \alpha, p, \theta, a) = \sum_{i=1}^{N} \alpha_i^p \| A_{a_i} R_{-\theta_i} (Du_i) \|_p^p,$$

where the orthogonal (rotation) matrices $R_{-\theta_i}$ and the diagonal matrices $A_{a_i}$ have been defined in (3.22) and (3.24), respectively.

Note that the (WDTV$^\Theta_p$) regulariser can also be written in terms of its parametric, space-variant gradient penalty functions as

$$\text{WDTV}^\Theta_p(u; \alpha, p, \theta, a) = \sum_{i=1}^{N} f_{\text{WDTV}^\Theta_p}((Du_i); \alpha_i, p_i, \theta_i, a_i),$$

with

$$f_{\text{WDTV}^\Theta_p}(g_i; \alpha_i, p_i, \theta_i, a_i) = \alpha_i^p \| A_{a_i} R_{-\theta_i} g_i \|_2^p,$$

where the scale and shape parameter values $(\alpha_i, p_i, \theta_i, a_i)$ are the same as for the WTV$^\Theta_p$ regulariser. In particular, it is worth noting that the (WDTV$^\Theta_p$) regulariser reduces to the rotationally-invariant regulariser in the special case $a_i = 1$ for any $i$, independently of the directionality parameters $\theta_i$.

In the last column of Figure 6 we show the graph of the WDTV$^\Theta_p$ gradient penalty function $f_{\text{WDTV}^\Theta_p}$ defined in (4.15) 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 $(\alpha_i, p_i, \theta_i, a_i)$ are the same as for the WTV$^\Theta_p$ regulariser. In this case, it is clear from these figures that the degrees of freedom represented by parameters $a_i$ allow to make the WDTV$^\Theta_p$ regulariser locally anisotropic, in the sense that it can locally penalise the gradient $g_i = (Du_i)$, with different strength according to its direction. The level curves of the penalties in Figures 6(f),(i) corresponding to $a_i < 1$ are elliptical and not circular as for the case $a_i = 1$ in Figure 6(c). Furthermore, the smaller $a_i$, the more eccentric the ellipses and, hence, the more anisotropic the regulariser. The local directional parameters $\theta_i$ represent local image directions along which a stronger regularisation effect is typically desired (typically, edge direction). We observe that the elliptical level curves of the penalties in Figures 6(f),(i) are rotated of angle $\theta_i$ counterclockwise, with the minor and major axes aligned along the directions defined by $\theta_i$ and $\theta_i + \pi/2$, respectively, and that the 1D sections of the WDTV$^\Theta_p$ penalty functions in (4.15) along directions defined by angle $\varphi$ take the form

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

It is a simple calculation verifying that, for any fixed $\theta_i$, $a_i$, the positive real coefficient in brackets takes its maximum (equal to 1) and minimum (equal to $a_i^p$) values for $\varphi = \theta_i$ and $\varphi = \theta_i + \pi/2$, respectively.
respectively. This entails that the 1D sections of the $\text{WDTV}^w_p$ penalty along the dominant direction $\theta$, and its orthogonal $\theta + \pi/2$ are those characterised by the strongest and the weakest regularisation effect, 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 6(j), right.

The $\text{WDTV}^w_p$ regulariser has been first introduced and analysed in probabilistic settings in [21].

4.1.4 Comparing regularisers: proximal operators

In order to gain more insights on the regularisation effects yielded by the different gradient penalty functions introduced in the previous sections and, consequently, on the different space-variant regularisers considered, we compare in this section the proximal operators $\text{prox}^\beta_f : \mathbb{R}^2 \mapsto \mathbb{R}^2$, $f \in \{ \text{fTV}, \text{fWTV}, \text{fWTV}^w_p, \text{fWDTV}^w_p \}$ (see Definition 8) associated to the TV, WTV, $\text{WTV}^w_p$ and $\text{WDTV}^w_p$ penalty functions defined in (4.3), (4.7), (4.11) and (4.15), respectively.

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

$$e^\beta_f (w) := \text{prox}^\beta_f (w) - w, \quad w \in \mathbb{R}^2,$$

(4.17)
can be studied to represent the regularisation effect of the gradient penalty function considered on $w$.

Analytical expressions of the proximal operators of the gradient penalty functions in (4.3), (4.7), (4.11) and (4.15) associated to the (TV), (WTV), (WTV$^w_p$) and (WDTV$^w_p$) regularisers have been previously studied in [135], [90] and [21], respectively, and are discussed (for completeness) in Section 8.3 of this review. Based on those expressions, we thus compute the vector field $e^\beta_f$ in (4.17) for each of the nine penalty functions considered in Figure 6 and report the results in Figure 7. 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 $\text{prox}^\beta_f (0) = 0 \implies e^\beta_f (0) = 0$ and that for all penalties with shape parameter $p_0 \leq 1$ - namely, the penalties in Figures 6(a)-(g) - there exists a region in the $w$ domain (with centre the origin, size depending on $\beta$ and shape depending on the penalty itself) for which we have $\text{prox}^\beta_f (w) = 0 \implies e^\beta_f (w) = -w$. This means that any input gradient vector $w$ belonging to this region is "completely" regularised by the gradient penalty function, in the sense that it is proximal-mapped to the null gradient vector. For visualisation purposes, such sparsity-promoting regions are depicted in green (without showing the arrows pointing towards the origin) in the vector field representations of Figures 7(a)-(g).

As expected, for isotropic penalties - namely, the WTV and $\text{WTV}^w_p$ penalties shown in the first two columns of Figure 6 and the $\text{WDTV}^w_p$ penalty with unitary anisotropy parameter depicted in Figure 6(c) - the associated vector fields $e^\beta_f$ are radial with vectors pointing towards the origin and the sparsity-promotion regions are circularly shaped - see Figures 7(a)-(e),(g). This means that the regularisation effect yielded by the isotropic penalties depicted in Figures 6(a)-(e),(g),(h) on $w$ is only a shrinkage of its norm $\|w\|_2$, namely

$$\text{prox}^\beta_f (w) = \xi w \implies e^\beta_f (w) = -(1 - \xi) w,$$

with a shrinkage coefficient $\xi \in [0, 1)$ only depending on the norm itself. This result has been proved, e.g., in [90](Proposition 1), where analytical expressions for $\xi$ as a function of $\|w\|_2$ as well as of the shape and proximal parameters have been given for the proximal operator of a TV$^p$ penalty of the form $f(g; p) = \|g\|_p^p$. Since it follows immediately from Definition 8 that

$$\text{prox}^\beta_f (w) = \text{prox}^\tilde{\beta}_f (w), \quad \forall w \in \mathbb{R}^2$$

with $\tilde{f}(g; \alpha, p) = \alpha^p \|w\|_p^p$ and $\tilde{\beta} = \alpha^p \beta$, then the results in [90] 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 6(a)-(e),(g),(h). In particular, by observing the
vector fields depicted in these Figures, it is clear how larger scale parameter values $\alpha_i$ 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 6(d),(e), one can notice that, for a fixed scale parameter $\alpha_i$, decreasing the shape parameter $p_i$ (starting from $p_i = 1$) in the WTV\textsuperscript{sp} penalty yields weaker shrinkage effects on gradients outside the sparsity-promoting regions but larger radii of these regions. Finally, Figures 6(g),(h) show that increasing $p_i$ (for a fixed $\alpha_i$ and starting from $p_i = 1$) in the WTV\textsuperscript{sp} penalty yields stronger gradient shrinkage effects and, for any $p_i > 1$, the sparsity-promotion regions reduce to the point $w = 0_2$.

Clearly, the vector fields in Figures 7(f),(i), associated to the WDTV\textsuperscript{sp} anisotropic penalties are not radial. To be more precise, they are radial only when restricted to input vectors $w$ lying on the two straight lines having direction defined by angles $\theta_i$ and $\theta_i + \pi/2$ (solid/dashed red and magenta lines in Figures 7(f),(i)). In general, the regularisation effect of the WDTV\textsuperscript{sp} penalties on input vectors $w$ is stronger along the direction $\theta_i$. Finally, the sparsity-promotion regions are elongated in the direction defined by $\theta_i$ and their elongation is negatively correlated with the value of the local anisotropy parameter $a_i$. 

Figure 7: 2D vector fields $e_{\beta}^f: \mathbb{R}^2 \to \mathbb{R}^2$ in (4.17), representing the effect of the proximal operator $\text{prox}_f^\beta$ on a (gradient) vector $w = (w_1, w_2)$, for the same gradient penalty functions $f$ shown in Figures 6(a)-(i) and for a fixed proximity parameter $\beta = 3$. 

25
5 Geometrical interpretation

In this section, we enrich the statistical and analytical study of the space-adaptive regularisers introduced in the previous sections by providing some insights helpful to understand their local behaviour from a geometrical point of view. To do so, we unify and expand some considerations from [6, 83, 112, 50] and start recalling the dual definition of TV:

$$\text{TV}(u) = \sum_{i=1}^{N} \max_{w_i \in B_1(0)} \langle (Du)_i, w_i \rangle,$$

(5.1)

where $B_1(0)$ denotes the two-dimensional Euclidean unit ball centred in the origin. Such constraint can be equivalently expressed by requiring $\|w_i\|_2 \leq 1$ for all $i = 1, \ldots, N$.

Following [6], we can now replace the set $B_1(0)$ in (5.1) with a fixed two-dimensional region $E_{a,\theta}(0)$ centred 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:

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

(5.2)

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

Recalling definitions (3.22) and (3.24) of the matrices $A_\theta$ and $R_\theta$ and denoting (formally, given the purely discrete setting we are working on) by $D_{\theta}u_i = (Du)_i \cdot v$ and $D_{\theta}^\perp u_i = (Du)_i \cdot v^\perp$ the directional derivatives along the direction $\mathbf{v} = (\cos \theta, \sin \theta)$ and its orthogonal $\mathbf{v}^\perp = (-\sin \theta, \cos \theta)$, we define element-wise the directional gradient $\tilde{D}_{a,\theta}u \in (\mathbb{R}^2)^N$ of $u$ as

$$\tilde{D}_{a,\theta}u := (A_\theta R_{\theta}(Du)_i)_i = \left( \frac{D_{\theta}u_i}{\partial D_{\theta}^\perp u_i} \right)_i. 

(5.3)

By this definition we can thus write the directional formulation of TV firstly used in [6] and later applied in several other works (see, e.g., [133, 83, 50]) for promoting TV smoothness along $\mathbf{v}$. Note that the (WDTV) 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

$$\text{DTV}(u) = \sum_{i=1}^{N} \| (\tilde{D}_{a,\theta}u)_i \|_2 = \sum_{i=1}^{N} \max_{w_i \in E_{a,\theta}(0)} \langle (Du)_i, w_i \rangle, 

(5.4)

where here and in what follows we omit the explicit dependence of the regularisers 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 $w_i$ are forced to belong to $E_{a,\theta}(0)$ at any point. Following [83], we can now observe that for $w_i \in E_{a,\theta}(0)$, we have

$$\langle (Du)_i, w_i \rangle = \langle (Du)_i, \mathbf{R}_\theta A_\theta \tilde{w}_i \rangle = \langle A_\theta R_{\theta}(Du)_i, \tilde{w}_i \rangle = \langle (\tilde{D}_{a,\theta}u)_i, \tilde{w}_i \rangle 

$$

where $\tilde{w}_i \in E_{1}(0)$ for all $i = 1, \ldots, N$. Thus, we deduce:

$$\max_{w_i \in E_{a,\theta}(0)} \langle (Du)_i, w_i \rangle = \max_{\tilde{w}_i \in E_{1}(0)} \langle (\tilde{D}_{a,\theta}u)_i, \tilde{w}_i \rangle, 

(5.5)

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

$$\langle (\tilde{D}_{a,\theta}u)_i, \tilde{w}_i \rangle = \frac{\langle (\tilde{D}_{a,\theta}u)_i, (\tilde{D}_{a,\theta}u)_i \rangle}{\| (\tilde{D}_{a,\theta}u)_i \|_2} = \| (\tilde{D}_{a,\theta}u)_i \|_2 = \| A_\theta R_{\theta}(Du)_i \|_2, 

(5.6)

which justifies (5.4).
Inspired by [113], we report in Figure 8 a graphical representation of the considerations above. There, we denote in red a fixed non-zero gradient vector $(\mathbf{D}u)_i \in \mathbb{R}^2$ evaluated at a certain point $i = 1, \ldots, N$, in blue the direction $\mathbf{v} = (\cos \theta, \sin \theta) \in \mathbb{R}^2$ drawing an angle $\theta$ with the $x$-axis and in green the projection of $(\mathbf{D}u)_i$, i.e., the directional gradient $(\mathbf{D}_{a,\theta}u)_i$. The unitary ball $B_1(0)$ is coloured black while the ellipses $E_{a,\theta}(0)$ for different values of $a \in (0, 1]$ are coloured magenta. For each plot, the unitary vector $\tilde{\mathbf{w}}_i$ realising the maximum in (5.4) is drawn (magenta). Note that for $a = 1$ we retrieve that the vector $\tilde{\mathbf{w}}_i$ maximising the scalar product is the one parallel to $(\mathbf{D}u)_i$ (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 $\tilde{\mathbf{w}}_i$ progressively aligns with $\mathbf{v} = (\cos \theta, \sin \theta)$, thus promoting directional regularisation.

As previously remarked, the unit vector $\mathbf{v}$ defining the orientation of the ellipse $E_{a,\theta}(0)$ is defined in terms of the angle $\theta$, which makes the use of the DTV regulariser useful in practice only when $\theta$ can be easily estimated. This is the case, for instance, of geometric textured images or of images of very specific scenes (see Figure 1), which limits significantly the application of DTV in practice.

Such limitation can be overcome by considering the following natural space-variant extension of the DTV regulariser (5.4), which comes from $\text{WDTV}^{sv}_{p}$ regulariser with $\alpha_i = p_i = 1$, $\forall i$, and space-variant $\theta_i \in [-\pi/2, \pi/2]$ and $a_i \in (0, 1]$:

$$
\text{DTV}^{sv}(\mathbf{u}) = \sum_{i=1}^{N} \| (\mathbf{D}_{a_i,\theta_i}u)_i \|_2 = \sum_{i=1}^{N} \| \Lambda_{a_i} \mathbf{R}_{-\theta_i}(\mathbf{D}u)_i \|_2 = \sum_{i=1}^{N} \max_{\mathbf{w}_i \in E_{a_i,\theta_i}(0)} \langle (\mathbf{D}u)_i, \mathbf{w}_i \rangle.
$$

(5.7)

where now $\alpha = (a_i)_i \in (0, 1]^N$, $\theta = (\theta_i)_i \in [0, \pi)^N$, and where we have used the simplified notation $(\mathbf{D}_{a,\theta}u)_i = \Lambda_{a_i} \mathbf{R}_{-\theta_i}(\mathbf{D}u)_i \in \mathbb{R}^2$. In this case, a space-variant adjustment of the directional smoothing (from strongly anisotropic along the direction $\mathbf{v}_i = (\cos \theta_i, \sin \theta_i)$ with $a_i = 0$ to fully isotropic with $a_i = 1$) is allowed at any point. For such regulariser, the same geometrical considerations as before hold, the difference being that the orientations $\theta_j$ may change from one point...
to another. Few choices can be made here. Following the Edge Adaptive Total Variation (EATV) approach proposed in [133], one possibility consists in estimating the local directions \( v_i \) by imposing that \( v_i \perp (Db)_i \), where, for \( \sigma > 0 \), \( b_\sigma \) denotes a smoothed version of the given image \( b \). This choice, however, is very sensitive to noise oscillations and it may misguide the local directional behaviour if these are too large. Alternatively, as considered in [65, 58, 54, 93] and more recently in [109, 50], the dependence on the image to retrieve can be encoded explicitly in the definition of the regularisation by allowing \( \theta_i \) to be a function of the target image \( u \) (i.e. \( \theta_i = \theta_i(u) \)) using, for instance, information coming from the structure tensor. This procedure is much more robust, but the nonlinear dependence on \( u \) in the definition of \( \theta_i \) may significantly complicate the problem from an optimisation viewpoint. For further estimation strategies based on maximum likelihood approaches, we refer the reader to [135, 87, 21] 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 \( a \) and \( \theta \) is a very challenging problem.

Remark 2. The values \( a_i \in (0,1] \) for all \( i = 1, \ldots, N \) have to be interpreted as ‘confidence’ parameters enforcing a strong anisotropic TV smoothing \( (a_i \approx 0) \) whenever a good local estimation of \( \theta_i \) is available, while leaving the behaviour to be close-to-isotropic \( (a_i \approx 1) \) whenever the estimation of \( \theta_i \) is unreliable.

As noted in Sections 3 and 4, we can further incorporate in (5.7) an additional shape/sharpness parameter vector \( p = (p_i) \in \mathbb{R}^N_+ \), thus considering the regulariser

\[
\text{DTV}_p^sv(u) = \sum_{i=1}^N \left( \max_{w_i \in E_{a_i,\theta_i}(0)} \langle (Du)_i, w_i \rangle \right)^{p_i},
\]

which is a particular instance of \( \text{WDTV}_p^sv \) taking \( \alpha_i = 1 \), \( \forall i \). The presence of the parameters \( p_i \) in (5.8) does not alter the directional behaviour of such regulariser in comparison with the one observed for \( \text{DTV}^sv \). However, as thoroughly discussed in Sections 4.1.2 and 4.1.4, such behaviour is made sharper for \( p_i < 1 \) and smoother for \( p_i > 1 \). Note, in particular, that when \( p_i = 2 \) the \( \text{DTV}_2^sv \) regulariser acts locally as a Tikhonov-type squared \( \ell_2 \)-norm of the directional gradient \( \tilde{D}_{a_i,\theta_i} u \).

We conclude this section with some considerations regarding weighted models. Recalling (5.1), we notice that introducing a space-variant parameter vector \( (\alpha_i) \in \mathbb{R}^N_+ \) corresponds simply to inflate/dilate the Euclidean ball \( B_i(0) \) and to look for maxima therein, which corresponds to the choice

\[
\text{WTV}(u) = \sum_{i=1}^N \alpha_i \| (Du)_i \|_2 = \sum_{i=1}^N \max_{w_i \in B_{\alpha_i}(0)} \langle (Du)_i, w_i \rangle
\]

where for \( \alpha_i > 0 \), \( B_{\alpha_i}(0) := \{ z \in \mathbb{R}^2 : \| z \|_2 \leq \alpha_i \} \) and where the vector \( \tilde{w}_i = \alpha_i \frac{(Du)_i}{\| (Du)_i \|_2} \) maximises the scalar products at any point. By analogous considerations as above, we can finally draw a connection with the regulariser defined in (4.14), which, recalling the discussion above, can be written as:

\[
\text{WDTV}_p^sv(u) = \sum_{i=1}^N \alpha_i^{p_i} \| (\tilde{D}_{a_i,\theta_i} u)_i \|_2 = \sum_{i=1}^N \alpha_i^{p_i} \left( \max_{w_i \in E_{a_i,\theta_i}(0)} \langle (Du)_i, w_i \rangle \right)^{p_i}.
\]

Remark 3. If \( p_i = 1 \) for all \( i = 1, \ldots, N \), the regulariser (5.10) takes the form

\[
\text{WDTV}^sv(u) = \sum_{i=1}^N \max_{w_i \in E_{a_i,\theta_i}(0)} \langle (Du)_i, w_i \rangle
\]

where \( E_{a_i,\theta_i}(0) \) denotes the 2D ellipse centred in the origin with eccentricity \( a_i \in (0,1] \), orientation \( \theta_i \in [-\pi/2, \pi/2) \) w.r.t. to the \( x \)-axis and width/height equal to \( 2a_i \) and \( 2a_ia_i \), respectively.

The authors believe that an interesting generalisation of the discussion above shall address situations where the constraint set is non-convex and, in particular, it is defined in terms of Lamé
Lamé curves with varying eccentricity, $\beta=1.3$

Lamé curves with varying shape, $a=0.7$

Figure 9: Lamé curves (5.12) with varying eccentricity and shape parameters, $\theta = \pi/6$.

curves centred in 0, i.e. defined by

$$L_{\beta,a,\theta}(\Theta) = \left\{ (x_1, x_2) \in \mathbb{R}^2 : |x_1 \cos \theta + x_2 \sin \theta|^\beta + \frac{|-x_1 \sin \theta + x_2 \cos \theta|^\beta}{a} \leq 1 \right\}. \quad (5.12)$$

Such set is non-convex as soon as $\beta < 1$, see Figure 9. The use of such general shapes may lead to consider new gradient-based regularisations where the underlying geometry constraining the dual functions favours smoothing in different ways.

6 Joint hypermodelling

In this section, we provide explicit expressions of the negative log-hyperprior $-\ln P(\Theta)$ and of the negative log-likelihood $-\ln P(b | A u)$ in (3.32), which allows to derive the final joint hypermodel.

6.1 Non-informative 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 $\Theta$ can be set, thus considering a (possibly improper) non-informative hyperprior. In formulas this corresponds to set

$$P(\Theta) = \varrho \chi_{D}(\Theta), \text{ with } \varrho \in \mathbb{R}^+, \quad (6.1)$$

from which it follows

$$-\ln P(\Theta) = -\ln \varrho + \iota P_{\Theta}(\Theta). \quad (6.2)$$

6.2 GG likelihood leads to $L_q$ fidelity term

First, based on the expression of the considered GG likelihoods in (3.1)-(3.2), the negative log-likelihood term $-\ln P(b | A u)$ in (3.31) takes the form

$$q < +\infty : \left\{ \begin{array}{l} -\ln P(b | A u) = \omega^q \|A u - b\|_q^q - M \ln \left( \frac{\omega}{2 \Gamma(1/q)} \right), \\ = \omega^q \|A u - b\|_q^q + C_q, \end{array} \right. \quad (6.3)$$

$$q = +\infty : \left\{ \begin{array}{l} -\ln P(b | A u) = \iota_{[0,1/\omega]}(\|A u - b\|_\infty) - M \ln \frac{\omega}{2}, \\ = \iota_{[0,1/\omega]}(\|A u - b\|_\infty) + C_\infty, \end{array} \right. \quad (6.5)$$

where the quantities $C_q, C_\infty > 0$ appearing in (6.3)-(6.5) do not depend on the optimisation variable $u$, so they can be dropped - in (3.31).

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

$$q < +\infty : F_q(A u; b) := q \omega^q L_q(A u; b), \quad (6.6)$$

$$q = +\infty : F_\infty(A u; b) := \iota_{[0,1/\omega]}(\|A u - b\|_\infty) \quad (6.7)$$
with
\[ L_q(Au; b) = \frac{1}{q} \|Au - b\|_q^q. \]  

\[ (6.8) \]

### 6.3 Joint variational Bayesian hypermodels

We are now ready to derive the explicit instances of the hypermodel (3.32) in terms of the selected priors, hyperpriors and fidelity functionals discussed above. To improve readability, we will consider the following data terms \( F_q \) with \( q < +\infty \). However, as it will be remarked at the end of the section, analogous derivations can be easily extended to the case \( q = +\infty \).

Problem (3.32) can be reformulated as
\[ \{u^*, \Theta^*\} \in \arg \min_{u \in \mathbb{R}^N, \Theta \in \mathcal{D}_\Theta} \left\{ \mathcal{R}(u, \Theta) + \mathcal{H}(\Theta) + \mu L_q(Au; b) \right\}, \]  

\[ (6.9) \]

where \( \mathcal{R}(u, \Theta) \) denotes one of the regularisation terms previously discussed in Section 4.1, while \( \mathcal{H}(\Theta) \) accounts for possibly multiple terms depending only on the hyperparameter vector \( \Theta \in \mathcal{D}_\Theta \). The uniform hyperprior (6.1) acts here simply by enforcing optimisation on \( \mathcal{D}_\Theta \) only. The parameter \( \mu > 0 \) is a regularisation parameter whose choice will be specified for each hypermodel in the following.

Similarly as what discussed in Section 4.1, we have that \( \mathcal{H}(\Theta) \) can be expressed in general form as
\[ \mathcal{H}(\Theta) = \sum_{i=1}^N h(\Theta_i), \quad h: \mathcal{D}_\Theta_i \to \mathbb{R}, \]  

\[ (6.10) \]

where the function \( h \) is a parameter penalty function whose form will be specified for each regulariser.

We start our considerations from the TV prior (3.8). By plugging (6.1) and (3.1) into (3.32), we get
\[ \{u^*, \alpha^*\} \in \arg \min_{u \in \mathbb{R}^N, \alpha \in \mathbb{R}_{++}} \left\{ \alpha \sum_{i=1}^N \| (Du)_i \|_2 - N \ln \alpha + q \omega^q L_q(Au; b) \right\} \]
\[ = \arg \min_{u \in \mathbb{R}^N, \alpha \in \mathbb{R}_{++}} \left\{ TV(u) - \frac{N}{\alpha} \ln \alpha + \mu L_q(Au; b) \right\}, \quad \text{with} \quad \mu := \frac{q \omega^q}{\alpha}, \]  

\[ (6.11) \]

\[ (6.12) \]

where, we recall that in this case \( \Theta = \alpha \in \mathbb{R}_{++} \), whence parameter penalty function \( h_{TV} \) reads
\[ h_{TV}(\alpha) = -\frac{1}{\alpha} \ln \alpha, \quad i = 1, \ldots, N. \]  

\[ (6.13) \]

For the TV\(_p\) and the DTV regularisers in (1.4) and (1.3), respectively, hypermodels with similar form as in (6.12) can be derived. In particular, as far as the TV\(_p\) regularisation term is concerned, we have that \( \Theta = (\alpha, p) \in \mathbb{R}_{++}^2 \) and that the parameter penalty function takes the form
\[ h_{TV_p}(\alpha, p) = -\frac{1}{\alpha^p} \ln \frac{\alpha p}{\Gamma(1/p)}, \quad i = 1, \ldots, N, \]  

\[ (6.14) \]

while for the DTV\(_p\) regulariser we have \( \Theta = (\alpha, p, \theta, a) \in \mathbb{R}_{++}^2 \times [-\pi/2, \pi/2] \times (0, 1] \) and the parameter penalty function reads
\[ h_{DTV_p}(\alpha, \theta, a) = -\frac{1}{\alpha} \ln \left( \frac{a + \frac{\alpha^2}{2\pi}}{4} \right), \quad i = 1, \ldots, N. \]  

\[ (6.15) \]

As far as space-variant hypermodels are concerned, we start considering the WTV regulariser for which \( \Theta = \alpha \in \mathbb{R}_{++}^N \). Model (6.9) thus turns into
\[ \{u^*, \alpha^*\} \in \arg \min_{u \in \mathbb{R}^N, \alpha \in \mathbb{R}_{++}^N} \left\{ \sum_{i=1}^N \alpha_i \| (Du)_i \|_2 - \sum_{i=1}^N \ln \alpha_i + \mu L_q(Au; b) \right\} \]
with \( \mu := q \omega^q \),
where the function $h_{WTV}$ is defined by
\[
h_{WTV}(\alpha_i) = -\ln \alpha_i, \quad i = 1, \ldots, N.
\] (6.16)

For the WTV$^p_{\mu}$, we have that $\Theta = (\alpha, p) \in \mathbb{R}_+^N \times \mathbb{R}_+^N$. The hypermodel (6.9) here specifies into
\[
\left\{ u^*, \alpha^*, p^* \right\} \in \arg \min_{u \in \mathcal{U}, \alpha \in \mathbb{R}_+^N, p \in \mathbb{R}_+^N} \left\{ \sum_{i=1}^N \alpha_i^p \| (Du)_i \|_2^p - \sum_{i=1}^N \ln \frac{\alpha_i p_i}{\Gamma(1/p_i)} + \mu \Lambda_q(Au; b) \right\}
\]
with $\mu := q \omega^q$,

With penalty function $h_{WTV}^p$, defined by
\[
h_{WTV}^p(\alpha_i, p_i) = -\ln \frac{\alpha_i p_i}{\Gamma(1/p_i)}, \quad i = 1, \ldots, N.
\] (6.17)

Finally, for WDTV$^p_{\mu}$, we have $\Theta = (\alpha, p, \theta, a) \in \mathbb{R}_+^N \times \mathbb{R}_+^N \times [-\pi/2, \pi/2]^N \times (0, 1]^N$, whence the final hypermodel reads
\[
\left\{ u^*, \alpha^*, p^*, \theta^*, a^* \right\} \in \arg \min_{u \in \mathcal{U}, \alpha \in \mathbb{R}_+^N, p \in \mathbb{R}_+^N, \theta \in [-\pi/2, \pi/2]^N, a \in (0, 1]^N} \left\{ \sum_{i=1}^N \alpha_i^p \| A_{\alpha_i, \theta} (Du)_i \|_2^{p_i} - \sum_{i=1}^N \ln \frac{\alpha_i p_i \alpha_i^2}{2 \pi \Gamma(2/p_i) 2^{2/p_i}} + \mu \Lambda_q(Au; b) \right\}, \quad \text{with} \quad \mu := q \omega^q,
\]
(6.18)

With parameter penalty function takes the form
\[
h_{WDTV}^p(\alpha_i, p_i, \theta_i, a_i) = -\ln \frac{\alpha_i p_i \alpha_i^2}{2 \pi \Gamma(2/p_i) 2^{2/p_i}}, \quad i = 1, \ldots, N.
\] (6.19)

For all considered space-invariant and space-variant hypermodels, we summarise in Table 1 the gradient and parameter penalty functions $f$ and $h$, respectively, as well as the parameter domains $\mathcal{D}_\Theta$, and the regularisation parameters $\mu$. In the table, we also report the reference papers in which the aforementioned regularisers have been first introduced and/or analysed in probabilistic terms.

| $\mathcal{R}(u, \Theta)$ | $f(g; \Theta)$ | $h(\Theta)$ | $\mathcal{D}_\Theta$, | $\mu$ | Ref |
|--------------------------|-------------|-------------|-----------------|-------|-----|
| TV                       | $\|g\|_2$    | $-\frac{1}{\alpha} \ln \alpha$ | $\alpha \in \mathbb{R}_+$ | $\frac{q \omega^q}{\alpha}$ | [118] |
| TV$^p$                   | $\|g\|_2^p$ | $-\frac{1}{\alpha} p \ln \frac{\alpha p}{\Gamma(1/p)}$ | $(\alpha, p) \in \mathbb{R}_+^N$ | $\frac{q \omega^q}{\alpha}$ | [90] |
| DTV                      | $\|A_{\alpha} R_{\alpha} g\|_2$ | $-\frac{1}{\alpha} p \ln \frac{\alpha p}{\Gamma(1/p)}$ | $(\alpha, \theta) \in \mathbb{R}_+ \times [-\pi/2, \pi/2] \times (0, 1]$ | $\frac{q \omega^q}{\alpha}$ | [6] |
| WTV$^p$                  | $\alpha_i g_i$ | $-\ln \alpha_i$ | $\alpha_i \in \mathbb{R}_+$ | $q \omega^q$ | [22] |
| WTV$^p_{\mu}$            | $\alpha_i^p \| g_i \|_2^p$ | $-\ln \alpha_i p_i \frac{\alpha_i^2}{\Gamma(1/p_i)}$ | $(\alpha_i, p_i) \in \mathbb{R}_+^N$ | $q \omega^q$ | [87] |
| WDTV$^p$                 | $\alpha_i^p \| A_{\alpha_i, \theta} g_i \|_2^p$ | $-\ln \alpha_i p_i \frac{\alpha_i^2}{\Gamma(1/p_i)}$ | $(\alpha_i, p_i, \theta_i, a_i) \in \mathbb{R}_+ \times [-\pi/2, \pi/2] \times (0, 1]$ | $q \omega^q$ | [21] |

Table 1: Gradient penalty function $f$, parameter penalty function $h$, parameters domain $\mathcal{D}_\Theta$, and regularisation parameter $\mu$ for the space-invariant and space-variant regularisation hypermodels considered with $L_q$ data fidelity, $q < +\infty$. 
Remark 4 (Additive i.i.d. uniform noise). When the corrupting noise is AIU, i.e. $q = +\infty$ and the data term is written as in (6.7), the regularisation parameter $\mu$ does not appear explicitly in the final hypermodels. However, one can clearly observe that functions $f, h$ and the parameter domain $\mathcal{D}_\Theta$, have the same expressions as the ones listed in Table 1.

As far as the value of the regularisation parameter $\mu$ is concerned, we remark that when both the scale parameter $\omega$ 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 $\mu$ value based on the true noise parameters does not guarantee that the empirical noise level calculated starting from the output residual image $r^*(\mu) = Au^*(\mu) - b$ coincides with the true underlying noise level.

Hence, $\mu$ 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.

7 Coupling image statistics with variational modelling: parameter selection

In this section, we address the estimation of the parameters $\Theta$ arising in the final joint hypermodel (6.9). As pointed out in Section 4, a key step considered in the following for tackling the $\Theta$-update step in the alternating scheme (3.33)-(3.34) consists in neglecting the normalisation constant $c(\Theta)$. 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 the sequel has been provided in [21, Section 7]. There, the authors showed that the considered estimator is unbiased, with empirical variance and root mean square error decaying to zero.

7.1 Inspecting space-variance

According to (6.9), the general form of the $\Theta$-update (3.33) reads

$$
\Theta^{(k+1)} \in \arg\min_{\Theta \in \mathcal{D}_\Theta} \left\{ -\ln P(z(u^{(k)}) \mid \Theta) \right\}
$$

$$
= \arg\min_{\Theta \in \mathcal{D}_\Theta} \left\{ \sum_{i=1}^{N} \left( f((D\textbf{u}^{(k)})_{i}; \Theta_{i}) + h(\Theta_{i}) \right) \right\},
$$

(7.1)

where $f$ and $h$ denote the general gradient and parameter penalty functions, respectively summarised in Table 1 for the hypermodels of interest. In light of the separability induced by the summation, problem (7.1) can be addressed by solving $N$ minimisation problems of the form

$$
\Theta_{i}^{(k+1)} \in \arg\min_{\Theta_{i} \in \mathcal{D}_{\Theta_{i}}} \left\{ f((D\textbf{u}^{(k)})_{i}; \Theta_{i}) + h(\Theta_{i}) \right\}, \quad i = 1, \ldots, N,
$$

(7.2)

where, notice, the information on $\textbf{u}^{(k)}$ required to perform the update of $\Theta_{i}$, is synthesised in the sole value $(D\textbf{u}^{(k)}).$. However, if the true underlying value $(D\textbf{u})$ is highly damaged by noise and blur, 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 neighbourhood $\mathcal{J}_i^r$ centred at pixel $i$ with side $2r + 1$ and dimension $\text{card}(\mathcal{J}_i^r) = (2r + 1)^2 =: m$ and compute the discrete gradients points in $\mathcal{J}_i^r$. These quantities will be then used for the estimation of the $i$-th unknown parameter $\Theta_{i}$. Statistically, the selected strategy relies on the assumption that in each of the considered neighbourhoods the gradients (or their magnitudes) are independently sampled from the same distribution.

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

$$
\mathcal{S}_i := \left\{ \| (D\textbf{u}^{(k)})_{j} \|_2 : j \in \mathcal{J}_i^r \right\} \quad \text{for the WTV and WTV}_{p}^{sv} \text{ regularisers}
$$

$$
\left\{ (D\textbf{u}^{(k)})_{j} : j \in \mathcal{J}_i^r \right\} \quad \text{for the WDTV}_{p}^{sv} \text{ regulariser}.
$$

(7.3)

32
For each \( i = 1, \ldots, N \), by exploiting the mutual independence of the gradients, problem (7.2) can thus be formulated as follows:

\[
\Theta_i^{(k+1)} \in \arg \min_{\Theta_i \in \mathcal{D}_{\Theta_i}} \left\{ -\ln \prod_{j \in \mathcal{J}_i} \mathbb{P}(S_i | \Theta_i) = -\ln \mathbb{P}(\{D_u^{(k)}\}_j; \Theta_j) \right\}
\]

\[
= \arg \min_{\Theta_i \in \mathcal{D}_{\Theta_i}} \left\{ \sum_{j \in \mathcal{J}_i} \left( f((D_u^{(k)})_j; \Theta_j) + h(\Theta_j) \right) \right\}.
\]

(7.4)

We now specify the formulation of the minimisation problem (7.4) in correspondence with the regularisation 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 \( (D_u)_j \) at pixel \( j \) will be simply denoted by \( g_j \).

### 7.2 Parameter estimation for the WTV regulariser

We start considering WTV regularisation. Recalling the definition of the gradient and parameter penalty functions \( f_{TV}, h_{TV} \) in (4.3),(6.13) and the hyperparameter domain \( \mathcal{D}_{\Theta} \), specified in Table 1, the problem of interest turns into

\[
\alpha_i^* \in \arg \min_{\alpha_i \in \mathbb{R}_{++}} \left\{ \mathcal{G}(\alpha_i) := -\ln \mathbb{P}(S_i | \alpha_i) = -m \ln \alpha_i + \sum_{j \in \mathcal{J}_i} \alpha_i \|g_j\|_2 \right\}.
\]

(7.5)

The following result holds true.

**Proposition 2.** The function \( \mathcal{G} : \mathbb{R}_{++} \rightarrow \mathbb{R} \) in (7.5) is smooth and convex, hence it admits a unique global minimiser.

In particular, since \( \mathcal{G} \) is differentiable on \( \mathbb{R}_{++} \), the solution of the \( i \)-th minimisation problem (7.5) can be simply found by imposing a first-order optimality condition:

\[
\mathcal{G}'(\alpha_i) = -\frac{m}{\alpha_i} + \sum_{j \in \mathcal{J}_i} \|g_j\|_2 = 0, \text{ whence } \alpha_i^* = \left( \frac{1}{m} \sum_{j \in \mathcal{J}_i} \|g_j\|_2 \right)^{-1}.
\]

(7.6)

Notice that in order to avoid degenerate configurations arising when considering neighbourhoods with null gradients, a small regularisation parameter \( 0 < \epsilon \ll 1 \) can be added to the local mean in (7.6). The selection of pixels involved in (7.6) can be efficiently carried out based on fast 2D convolution operators (realised by a fast 2D discrete transform) of the map of gradient norms with a square \((2r+1) \times (2r+1)\) averaging kernel.

In Figure 10, we analyse the performance of the parameter estimation strategy outlined above on selected sub-regions of the image in Figure 10a as well as on the image itself. The local neighbourhoods shown here consist of an almost constant red-bordered region and two textured regions - see Figure 10d and Figures 10g,10j, 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 neighbourhood 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 (7.6) 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 sub-regions, so as to obtain local scale parameters. The estimated parameters are reported in the caption. In Figure 10b, and in the close-up in Figure 10c, 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 neighbourhoods together with the corresponding estimated pdfs are shown in Figures 10e-10f, for the constant region, and in Figures 10h-10i and 10k-10l 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 modelling of local features; this is also reflected into 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 can not
be detected when adopting a (univariate) $h_L$ prior.

In order to analyse in more detail the connection between the estimated scale parameters and
the local regularisation strength, in Figure 11 we show the $\alpha$-map corresponding to different test
images. We observe that the scale parameters assume higher values on smooth or piece-wise
constant regions, whereas lower values are obtained in correspondence of edges and texture. In
those areas, a weaker regularisation is indeed preferable in order to preserve details. Note also
that the $\alpha$-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 artefacts 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 WTV$^p_{sv}$ regulariser

We now consider the WTV$^p_{sv}$ regulariser. Recalling the definitions for $f_{WTV^p_{sv}}$ and $h_{WTV^p_{sv}}$ given
in (4.11) and (6.17), respectively, and the hyperparameter space $D_{\Theta}$, in Table 1, we have that the
general problem (7.4) reduces to

$$\{ \alpha_i^*, p_i^* \} \in \arg \min_{(\alpha_i, p_i) \in \mathbb{R}_+^2} \left\{ \mathcal{G}(\alpha_i, p_i) := - \ln p(\mathcal{S} | \alpha_i, p_i) = - m \ln \alpha_i \right. \quad (7.7)$$

$$\left. + m \ln \Gamma \left( 1 + \frac{1}{p_i} \right) + \sum_{j \in J_i^*} \alpha_i \| g_j \|_{l_2}^p \right\}.$$  

(7.8)

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

$$\frac{\partial}{\partial \alpha_i} \mathcal{G}(\alpha_i, p_i) = - \frac{m}{\alpha_i} + p_i \alpha_i^{p_i-1} \sum_{j \in J_i^*} \| g_j \|_{l_2}^p = 0,$$  

(7.9)

which yields the following closed-form formula for the estimation of $\alpha_i$:

$$\alpha_i^*(p_i) = \left( \frac{p_i}{m \sum_{j \in J_i^*} \| g_j \|_{l_2}^p} \right)^{- \frac{1}{p_i}}.$$  

(7.10)

It is easy to verify that the second derivative of $\mathcal{G}$ with respect to $\alpha_i$ computed at $\alpha_i^*(p_i)$ is strictly
positive, hence the stationary point in (7.10) is a minimum. Similarly as for (7.6), also in this
case a parameter $0 < \varepsilon << 1$ shall be added to the summation (7.10) so as to avoid degenerate
configurations of gradient magnitudes. Plugging (7.10), we have

$$p_i^* \in \arg \min_{p_i \in \mathbb{R}_+^+} \left\{ \mathcal{G}(p_i) := \mathcal{G}(\alpha_i(p_i), p_i) = \frac{m}{p_i} \log \left( \frac{p_i}{m \sum_{j \in J_i^*} \| g_j \|_{l_2}^p} \right) \right. \quad (7.11)$$

$$\left. + m \ln \Gamma \left( 1 + \frac{1}{p_i} \right) + \frac{m}{p_i} \right\}.$$  

When addressing the study of $\mathcal{G}$ on $\mathbb{R}_+^+$, one can immediately notice that its behaviour 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 $[\varepsilon, R]$, with
$0 < \varepsilon < R$ and $R > 1$. In this case, the following result holds.
Figure 10: Parameters estimate for the WTV regulariser. From top to bottom: histogram of the gradient magnitudes on the whole test image ($\alpha^* = 10.17$), on a constant region ($\alpha^* = 200.87$) and on two different texture regions ($\alpha^* = 3.70$ and $\alpha^* = 3.86$), with the corresponding close-up(s).
Figure 11: Test images with the corresponding $\alpha$-map for different values of radius $r$. 

**Proposition 3.** The function $G : [\epsilon, R] \to \mathbb{R}$ defined in (7.11) is continuous, hence it admits a minimum in its compact domain.

In Figure 12, the estimation of the global and local shape parameters for the WTV$^\text{sv}$ regulariser is performed by setting $\epsilon = 0.1$ and $R = 10$. The dashed green line in Figures 12b-12c 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 modelling of the global histogram when compared to the solid green line, representing the hL pdf shown in Figure 10. In Figures 12e,12h,12k, we report coloured 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 10 as solid lines. To facilitate the inspection, we also show close-up(s) of the local histograms in Figures 12f,12g,12l. The benefits associated to 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, is 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 sub-regions, due once again to its univariate behaviour.

Finally, in Figure 13 we show the $\alpha$- and $p$-maps, obtained by considering neighbourhoods of different sizes ($r$) for the image in Figure 10a. In all three cases, the method associates very low $p$ values with flat regions (thus promoting enforced sparsity) and higher values with texture (where gradients show oscillations). Similarly as what observed for WTV, the scale parameters $\alpha$ are again smaller on regions characterised by finer details, as expected.

### 7.4 Parameter estimation for the WDTV$^\text{sv}$ regulariser

For the WDTV$^\text{sv}$ regularisation term, after selecting the functions $f_{\text{WDTV}^\text{sv}}$, $h_{\text{WDTV}^\text{sv}}$ as in (4.15),(6.19) and the domain $\mathcal{D}_\Theta$, as specified in Table 1, we get that the problem of interest takes the form

$$
\{\alpha^*_i, p^*_i, \theta^*_i, a^*_i\} \in \arg \min_{(\alpha_i, p_i, \theta_i, a_i) \in \mathbb{R}^*_+ \times [-\pi/2, \pi/2] \times (0,1]} G(\alpha_i, p_i, \theta_i, a_i)
$$

(7.12)
Figure 12: Parameters estimate for the \( WTV^p \) regulariser. From top to bottom: histogram of the gradient magnitudes on the whole test image \((p^* = 0.2, \alpha^* = 465.67)\), on a constant region \((p^* = 0.1, \alpha^* = 687.34)\) and on two different texture regions \((p^* = 4.85, \alpha^* = 2.56\) and \(p^* = 5.65, \alpha^* = 2.78)\), with the corresponding close-up(s).

where

\[
G(\alpha_i, p_i, \theta, a_i) := -m \ln a_i + m \ln \Gamma \left( \frac{2}{p_i} + 1 \right) - 2m \ln \alpha_i
+ m \left( \frac{2}{p_i} - 1 \right) \ln 2 + \alpha_i p_i \sum_{j \in J_i} (g_j^T A_{\theta_i} R \theta_i \Lambda_2 R_{\theta_i} g_j)^{p_i/2}.
\]

(7.13)

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

\[
\frac{\partial G}{\partial \alpha_i} = -2m \frac{1}{\alpha_i} + p_i \alpha_i^{p_i-1} \sum_{j \in J_i} (g_j^T A_{\theta_i} R \theta_i \Lambda_2 R_{\theta_i} g_j)^{p_i/2}
\]

(7.14)
which yields
\[ \alpha^*_i(p_i, \theta_i, a_i) = \left( \frac{p_i}{2m} \sum_{j=1}^{m} (g^T R_{\theta_i} A_{a_i}^2 R_{-\theta_i} g_j)^{p_i/2} \right)^{-\frac{1}{p_i}}, \] (7.15)
and which can be regularised depending on \( 0 < \varepsilon \ll 1 \) as above. The stationary point in (7.15) can be proved to be a minimum as the second derivative of \( G \) with respect to \( \alpha_i \) at \( \alpha^*_i \) is strictly positive. Plugging (7.15) into (7.13), we thus get:
\[ G(p_i, \theta_i, a_i) := G(\alpha^*_i(p_i, \theta_i, a_i), \theta_i, a_i) \]
\[ = m \ln \left[ \Gamma \left( \frac{2}{p_i} + 1 \right) \frac{1}{2 a_i} \right] + \frac{2m}{p_i} \left( \ln \frac{p_i}{m} + 1 \right) \]
\[ + \frac{2m}{p_i} \ln \left( \sum_{j=1}^{m} (g^T R_{\theta_i} A_{a_i}^2 R_{-\theta_i} g_j)^{p_i/2} \right). \] (7.16)

By making now explicit the dependence of \( G \) on the entries of \((R_{\theta_i} A_{a_i}^2 R_{-\theta_i})\), we have that (7.16) turns into:
\[ G(p_i, \theta_i, a_i) = m \ln \left[ \Gamma \left( \frac{2}{p_i} + 1 \right) \frac{1}{2 a_i} \right] + \frac{2m}{p_i} \left( \ln \frac{p_i}{m} + 1 \right) \]
\[ + \frac{2m}{p_i} \ln \left( \sum_{j=1}^{m} (\cos^2 \theta_i + a_i^2 \sin^2 \theta_i) g^2_{j,1} + (\sin^2 \theta_i + a_i^2 \cos^2 \theta_i) g^2_{j,2} \right) \]
\[ + 2(1 - a_i^2) \cos \theta_i \sin \theta_i g_{j,1} g_{j,2}^{p_i/2}. \] (7.17)

Problem (7.12)-(7.13) thus takes the form:
\[ \{p^*_i, \theta^*_i, a^*_i\} \in \arg \min_{(p_i, \theta_i, a_i) \in \mathbb{R}_+ \times [-\pi/2, \pi/2] \times (0, 1]} G(p_i, \theta_i, a_i) \] (7.18)

We now study the behaviour of \( G \) as the triplet \((p_i, \theta_i, a_i)\) approaches the boundary of the set \( \hat{D}_{\Theta_i} := \mathbb{R}_+ \times [-\pi/2, \pi/2] \times (0, 1] \). Note that, since problem (7.18) is formulated over a non-compact set of \( \mathbb{R}^3 \), the existence of a solution is in general not guaranteed. One possible way to overcome the problem of non-compactness consists in characterising explicitly the configurations of the samples \( S_i \) for which the functional \( G \) in (7.17) does not attain its minimum in \( \hat{D}_{\Theta_i} \). To do
so, let us first set:

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

(7.19)

For any $p_i > 0$, if $A(\theta_i, a_i)$ is bounded as $a_i \to 0^+$, then the functional $G$ in (7.17) tends to $+\infty$ and the minimum is necessarily attained in the interior of $\hat{D}_{\Theta_i}$. However, if $A(\theta_i, a_i)$ is unbounded as $a_i \to 0^+$, nothing can be said about the behaviour 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 $g_1, \ldots, g_m \in S_i$ for which $G$ tends to $-\infty$ at the boundary. In order to characterise such configurations, note that as $a_i \to 0^+$ we have that by continuity:

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

(7.21)

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

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

(7.22)

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

A possible way to guarantee the existence of solutions of the problem (7.18) is to re-formulate 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 neighbourhood of the point considered almost impossible. Hence, the exponent $p_i$ is thought as confined in the closed interval $[\epsilon, R]$, with $0 < \epsilon < R$.

We can thus reformulate problem (7.18) as

$$\{p^*_i, \theta^*_i, a^*_i\} \in \min_{p_i, \theta_i, a_i} G(p_i, \theta_i, a_i)$$

(7.23)

s.t. $p_i \in [\epsilon, R]$, $-\pi/2 \leq \theta_i \leq \pi/2$, $\delta \leq a_i \leq 1$.

The following result holds true:

**Proposition 4.** The function $G : [\epsilon, R] \times [-\pi/2, \pi/2] \times [\delta, 1] \to \mathbb{R}$ in (7.17) is continuous and admits a minimum in its compact domain.

In Figure 14, we analyse the performance of the outlined parameter estimation strategy for the WDTV regulariser, 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 neighbourhoods from a synthetic image - i.e. a vertical edge in Figure 14a, a horizontal edge in Figure 14d, and a circular profile in Figure 14g - and the two textured regions already considered in Figures 10 and 12. In the middle column of Figure 14, we report the samples extracted from each neighbourhood, 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 $D_{\theta_i, u-D_i, u}$, respectively. This behaviour, as expected, corresponds to the dominant orientation of the gradients in the neighbourhoods. Finally, the textured regions in Figures 14g,14j, statistically indistinguishable from the hLd and hGGd viewpoint, result 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 difference is now reflected into the estimated BGG pdfs.
Figure 14: Parameters estimate for the WDTV\(^p\) regulariser. From left to right: selected neighbourhoods, histogram and scatter plot of gradient samples with level curves of the estimated BGGd. From top to bottom, the estimated parameters are: \((\alpha, p, \theta, a) = (2.23, 0.1, -89.90, 0.33), (\alpha, p, \theta, a) = (2.10, 0.1, 0.0, 0.37), (\alpha, p, \theta, a) = (1.14, 0.1, 47.50, 0.78), (\alpha, p, \theta, a) = (1.30, 2.97, -12.50, 0.64), (\alpha, p, \theta, a) = (1.94, 1.24, -2.18, 0.39).\)
8 Algorithmic optimisation

From an optimisation point of view, it is not trivial to design a unified optimisation solver for the general $\mathbf{u}$-estimation problem in the alternating scheme (3.33)-(3.34), as it may be either extremely easy (smooth and convex) or extremely difficult (non-convex and non-smooth). We can surely think of specific optimisation algorithms that could be effectively used for solving (3.33)-(3.34) in specific scenarios such as [l-BFGS [18] for the smooth and convex case, Nesterov-type proximal schemes [105, 7, 46, 67] and dual/primal-dual methods [81, 42, 73, 34, 37] for the non-smooth convex case and, e.g., [107] for the non-smooth non-convex case). However, in the following we will stick with one single optimisation algorithm for better clarity and consider the Alternating Directional Method of Multipliers (ADMM) [14] whose different subproblems can be solved by means of classical tools in the field of proximal calculus [45], numerical linear algebra and adaptive discrepancy principle [69]. Note that albeit proposed and widely applied in convex scenarios, non-convex variants of ADMM have been recently proposed and endowed with global convergence guarantees [127, 12], although not always applicable to the problem at hand due to the (often limiting) assumptions on the operators involved. However, as we will comment in the following, empirical convergence is often observed for general non-convex ADMM algorithms, which makes their use often amenable in practice. For further details on the recent developments of convex and non-convex optimisation algorithms for variational imaging models, we refer the reader to [38] and the references therein.

8.1 ADMM optimisation

By dropping out the terms in (6.9) which do not depend on the unknown image $\mathbf{u}$, the $\mathbf{u}$-update step in the alternating scheme (3.33)-(3.34) reads

$$
\mathbf{u}^{(k+1)} \in \arg \min_{\mathbf{u} \in \mathbb{R}^N} \left\{ \sum_{i=1}^{N} f((\mathbf{D}_i \mathbf{u}); \Theta_i^{(k+1)}) + F_q(\mathbf{A} \mathbf{u}; \mathbf{b}) \right\},
$$

(8.1)

where $F_q(\mathbf{A} \mathbf{u}; \mathbf{b})$ is defined in (6.6)-(6.7) and the gradient penalty functions $f$ are summarised in Table 1 for the different regularisers considered.

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

$$
\{ \mathbf{u}^*, \mathbf{g}^*, \mathbf{r}^* \} \in \arg \min_{\mathbf{u}, \mathbf{g}, \mathbf{r}} \left\{ \sum_{i=1}^{N} f(\mathbf{g}_i; \Theta_i) + F_q(\mathbf{r}; \mathbf{b}) \right\} \text{ s.t. : } \begin{cases} \mathbf{g} = \mathbf{D} \mathbf{u} \\ \mathbf{r} = \mathbf{A} \mathbf{u} \end{cases}
$$

(8.2)

For every $i = 1, \ldots, N$, the quantity $\mathbf{g}_i = ((\mathbf{D}_{i1} \mathbf{u}), (\mathbf{D}_{i2} \mathbf{u})) \in \mathbb{R}^2$ 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) non-differentiable and possibly non-convex function $f$, while the use of $\mathbf{r}$ is helpful for the GDP strategy introduced in Section 2.1.

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

$$
\mathcal{L}(\mathbf{u}, \mathbf{g}, \mathbf{r}, \mathbf{\rho}_u, \mathbf{\rho}_r; \Theta) := \sum_{i=1}^{N} f(\mathbf{g}_i; \Theta_i) + F_q(\mathbf{r}; \mathbf{b}) - \langle \mathbf{\rho}_u, \mathbf{g} - \mathbf{D} \mathbf{u} \rangle + \frac{\beta_u}{2} \| \mathbf{g} - \mathbf{D} \mathbf{u} \|^2_2 + \frac{\beta_r}{2} \| \mathbf{r} - \mathbf{A} \mathbf{u} \|^2_2,
$$

(8.3)

where $\beta_u, \beta_r \in \mathbb{R}_{++}$ are the ADMM penalty parameters, while $\mathbf{\rho}_u \in \mathbb{R}^{2N}$, $\mathbf{\rho}_r \in \mathbb{R}^M$ are the vectors of Lagrange multipliers associated with the linear constraints $\mathbf{g} = \mathbf{D} \mathbf{u}$ and $\mathbf{r} = \mathbf{A} \mathbf{u}$ in (8.2), respectively.

Solving (8.2) amounts to seek for solutions of the saddle point problem:

$$
\text{Find } (\mathbf{u}^*, \mathbf{g}^*, \mathbf{r}^*) \in \mathbb{R}^{N} \times \mathbb{R}^{2N} \times \mathbb{R}^M \text{ and } (\mathbf{\rho}_u^*, \mathbf{\rho}_r^*) \in \mathbb{R}^{2N} \times \mathbb{R}^M \text{ such that: }
\mathcal{L}(\mathbf{u}, \mathbf{g}, \mathbf{r}, \mathbf{\rho}_u, \mathbf{\rho}_r; \Theta) \leq \mathcal{L}(\mathbf{u}^*, \mathbf{g}^*, \mathbf{r}^*, \mathbf{\rho}_u^*, \mathbf{\rho}_r^*; \Theta) \leq \mathcal{L}(\mathbf{u}^*, \mathbf{g}^*, \mathbf{r}^*, \mathbf{\rho}_u^*, \mathbf{\rho}_r^*; \Theta)
$$

(8.4)

forall $(\mathbf{u}, \mathbf{g}, \mathbf{r}) \in \mathbb{R}^{N} \times \mathbb{R}^{2N} \times \mathbb{R}^M$, $\forall (\mathbf{\rho}_u, \mathbf{\rho}_r) \in \mathbb{R}^{2N} \times \mathbb{R}^M$. 

41
Upon suitable initialisation, and for any \( j \geq 0 \), the \( j \)-th iteration of the ADMM algorithm applied to solve the saddle-point problem (8.4) thus reads:

\[
\begin{align*}
\mathbf{u}^{(j+1)} & \in \arg \min_{\mathbf{u} \in \mathbb{R}^N} \mathcal{L}(\mathbf{u}, g^{(j)}, \mathbf{r}^{(j)}, \rho_g^{(j)}, \rho_r^{(j)}; \Theta), \\
g^{(j+1)} & \in \arg \min_{g \in \mathbb{R}^N} \mathcal{L}(\mathbf{u}^{(j+1)}, g, \mathbf{r}^{(j)}, \rho_g^{(j)}, \rho_r^{(j)}; \Theta), \\
\mathbf{r}^{(j+1)} & \in \arg \min_{\mathbf{r} \in \mathbb{R}^M} \mathcal{L}(\mathbf{u}^{(j+1)}, \mathbf{g}^{(j+1)}, \mathbf{r}, \rho_g^{(j)}, \rho_r^{(j)}; \Theta), \\
\rho_g^{(j+1)} & = \rho_g^{(j)} - \beta_g (g^{(j+1)} - \mathbf{D} \mathbf{u}^{(j+1)}), \\
\rho_r^{(j+1)} & = \rho_r^{(j)} - \beta_r (\mathbf{r}^{(j+1)} - \mathbf{A} \mathbf{u}^{(j+1)}). 
\end{align*}
\] (8.5-8.7)

In the following, we make precise the solution of the three sub-problems for the primal variables \( \mathbf{u}, g, \) and \( \mathbf{r} \) in (8.5)-(8.7). The automatic estimation of the regularisation parameter \( \mu \) will be addressed in Section 8.4 concerned with the \( \mathbf{r} \)-update.

### 8.2 Subproblem for the primal variable \( \mathbf{u} \)

Subproblem (8.5) reads

\[
\mathbf{u}^{(j+1)} \in \arg \min_{\mathbf{u} \in \mathbb{R}^N} \left\{ (\rho_g^{(j)}, \mathbf{D} \mathbf{u}) + (\rho_r^{(j)}, \mathbf{A} \mathbf{u}) + \frac{\beta_g}{2} \|g^{(j)} - \mathbf{D} \mathbf{u}\|_2^2 + \frac{\beta_r}{2} \|\mathbf{r}^{(j)} - \mathbf{A} \mathbf{u}\|_2^2 \right\},
\] (8.9)

which is quadratic with first-order optimality condition given by

\[
\left( \beta_g \mathbf{D}^T \mathbf{D} + \beta_r \mathbf{A}^T \mathbf{A} \right) \mathbf{u} = \beta_g \mathbf{D}^T \left( g^{(j)} - \frac{1}{\beta_g} \rho_g^{(j)} \right) + \beta_r \mathbf{A}^T \left( \mathbf{r}^{(j)} - \frac{1}{\beta_r} \rho_r^{(j)} \right). 
\] (8.10)

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

\[
\text{null}(\mathbf{A}) \cap \text{null}(\mathbf{D}) = \{0_N\},
\] (8.11)

then it is positive definite so that \( \mathbf{u}^{(j+1)} \) 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 \( g \)

After dropping all terms not depending on \( g \) in (8.3), subproblem (8.6) reads

\[
\begin{align*}
g^{(j+1)} & \in \arg \min_{g \in \mathbb{R}^N} \left\{ \sum_{i=1}^{N} f(g_i; \Theta_i) - \langle \rho_g^{(j)}, g - \mathbf{D} \mathbf{u}^{(j+1)} \rangle + \frac{\beta_g}{2} \|g - \mathbf{D} \mathbf{u}^{(j+1)}\|_2^2 \right\} \\
& = \arg \min_{g \in \mathbb{R}^N} \left\{ \sum_{i=1}^{N} f(g_i; \Theta_i) + \frac{\beta_g}{2} \|g - \mathbf{w}^{(j)}\|_2^2 \right\}, 
\end{align*}
\] (8.12)

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

\[
\mathbf{w}^{(j)} := \mathbf{D} \mathbf{u}^{(j+1)} + \frac{1}{\beta_g} \rho_g^{(j)}. 
\] (8.13)

Solving the 2N-dimensional minimisation problem above is thus equivalent to solve the following N independent 2-dimensional problems:

\[
\begin{align*}
g_i^{(j+1)} & \in \arg \min_{g_i \in \mathbb{R}^2} \left\{ f(g_i; \Theta_i) + \frac{\beta_g}{2} \|g_i - \mathbf{w}_i^{(j+1)}\|_2^2 \right\} \\
& = \text{prox}_{\frac{\beta_g}{2}}(\mathbf{w}_i^{(j+1)}), \quad i = 1, \ldots, N, 
\end{align*}
\] (8.14)
where prox$^β_f(\cdot; \mathbf{e}_i) : \mathbb{R}^2 \mapsto \mathbb{R}^2$ denotes the proximal operator of the gradient penalty function $f(\cdot; \Theta_i)$ with proximity parameter $\beta$ - see Definition 8 and Section 4.1.4 - and where the vectors $\mathbf{w}_i^{(j+1)} \in \mathbb{R}^2$ at any iteration read

$$\mathbf{w}_i^{(j+1)} = \left( \mathbf{D} \mathbf{w}_i^{(j+1)} \right)_i + \frac{1}{\beta y} \left( \mathbf{p}_i^{(j)} \right)_i, \quad i = 1, \ldots, N. \quad (8.15)$$

We start detailing the solving procedure for problem (8.14) under the adoption of a WDTV$^\theta_p$ regularisation term, which corresponds to consider the gradient penalty function $f_{\text{WDTV}^\theta_p}$ defined in (4.15); the proximal maps arising for the WTV and WDTV$^\theta_p$ regularisers will be discussed afterwards as special cases. In [21], the authors proved a result on the existence of solutions for problem (8.14). Before reporting the statement, we recall that in the following, for $\mathbf{v}, \mathbf{w} \in \mathbb{R}^n$ we denote by $\mathbf{v} \circ \mathbf{w}$, $|\mathbf{v}|$ and sign$(\mathbf{v})$ the component-wise (or Hadamard) product between $\mathbf{v}$ and $\mathbf{w}$ and the component-wise absolute value and sign of $\mathbf{v}$, respectively.

**Lemma 1.** Let $f : \mathbb{R}^2 \to \mathbb{R}_+$ be the (parametric and not necessarily convex) function defined by

$$f(\mathbf{g}) := \alpha^p \| \mathbf{A}_a \mathbf{R}_- \mathbf{g} \|^2_2, \quad \mathbf{g} \in \mathbb{R}^2, \quad (8.16)$$

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

$$\mathbf{g}^* \in \text{prox}_f^β(\mathbf{w}) := \arg \min_{\mathbf{g} \in \mathbb{R}^2} \left\{ F(\mathbf{g}) := f(\mathbf{g}) + \frac{\beta}{2} \| \mathbf{g} - \mathbf{w} \|^2 \right\}, \quad \mathbf{w} \in \mathbb{R}^2. \quad (8.17)$$

Then, problem (8.17) admits at least one solution, which is unique when $p \geq 1$. Moreover, after defining

$$\tilde{\mathbf{w}} := \mathbf{R}_- \mathbf{w}, \quad s := \text{sign}(\tilde{\mathbf{w}}), \quad \overline{\mathbf{w}} := |\tilde{\mathbf{w}}|, \quad \bar{\beta} := \frac{\beta}{\alpha^p}, \quad (8.18)$$

we have that any solution $\mathbf{g}^*$ of (8.17) can be expressed as

$$\mathbf{g}^* = \mathbf{R}_a (s \circ \mathbf{z}^*), \quad \mathbf{z}^* \in \arg \min_{\mathbf{z} \in \mathcal{H}_1 \subset \mathbb{R}^2} H(\mathbf{z}), \quad (8.19)$$

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

$$H(\mathbf{z}) := \| \mathbf{A}_a \mathbf{z} \|^2_2 + \frac{\bar{\beta}}{2} \| \mathbf{z} - \overline{\mathbf{w}} \|^2, \quad \mathcal{H}_1 := \mathcal{H} \cap \{ (\overline{\mathbf{w}}, \overline{\mathbf{w}}) \times \mathbb{R}_+ \}, \quad (8.20)$$

with $\mathcal{H}$ being

1. the rectangular hyperbola defined by

$$\mathcal{H}_1 := \left\{ \mathbf{z} \in \mathbb{R}^2 : (z_1 - c_1) (z_2 - c_2) = c_1 c_2, c_1 = -a^2 \overline{\mathbf{w}}, c_2 = -\overline{\mathbf{w}} \right\}, \quad (8.21)$$

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

2. the line defined by

$$\mathcal{H}_1 := \left\{ \mathbf{z} \in \mathbb{R}^2 : \overline{\mathbf{w}}_2 z_1 - \overline{\mathbf{w}}_1 z_2 = 0 \right\} \quad (8.22)$$

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

**Corollary 1.** The minimisers $\mathbf{z}^* \in \mathbb{R}^2$ in (8.19) can be obtained as follows:

$$\mathbf{z}^* = \left( \frac{z_1^*}{z_1^* - c_1}, \frac{c_2 z_1^*}{z_1^* - c_1} \right), \quad (8.23)$$

where $c_1, c_2 \in \mathbb{R}$ are defined in (8.21) and $z_1^* \in \mathbb{R}$ is the solution(s) of the following 1-dimensional constrained minimisation problem:

$$z_1^* \in \arg \min_{\xi \in [0, \overline{\mathbf{w}}_1]} \left\{ h(\xi) := (h_1(\xi))^{p/2} + \frac{\bar{\beta}}{2} h_2(\xi) \right\}, \quad (8.24)$$

$$h_1(\xi) = \xi^2 \left( 1 + \frac{a^2 c_2}{(\xi - c_1)^2} \right), \quad h_2(\xi) = (\xi - \overline{\mathbf{w}}_1)^2 + \left( \frac{c_2 \xi}{\xi - c_1} - \overline{\mathbf{w}}_2 \right)^2. \quad (8.25)$$
Figure 15: Graphical representation of the bivariate minimisation problems arising for the WDTVsvp (a) and the WTVsvp regularisers (b).

We will omit the proof of this lemma, and provide only a brief graphical sketch of the key steps leading to (8.19). First, in order to get some clues about the approximate position of the minimiser z∗ in the plane z1-z2, we restrict the study of the function H to the one-parameter family of ellipses

\[
E_R(0) = \{ (z_1, z_2) \in \mathbb{R}^2 \mid z_1^2 + a^2 z_2^2 = R \}. \tag{8.26}
\]

One can prove that z∗ needs to belong to the hyperbola H defined in (8.21). More specifically, the sought z∗ has to coincide with one of the two points in \( E_R \cap H \) belonging to the first quadrant of the plane z1-z2. In Figure 15a, we show one ellipse \( E_R \), which is depicted with a blue dashed line, and the hyperbola H, plotted with a solid magenta line. We conclude that z∗ lies on the arc of hyperbola \( H_1 \) which is delimited by the origin O and \( w \); \( H_1 \) is also illustrated in Figure 15a with a solid red line.

Remark 5. Upon the adoption of the WTVsvp regulariser, a similar result can be proven - see [90, Proposition 1]. More specifically, in isotropic settings there holds a = 1, which yields that the parametric family of ellipses in (8.26) reduces to a parametric family of circles. Moreover, the hyperbola H and the arc \( H_1 \) turn into a line and a segment, respectively. The simplified configuration is reported in Figure 15b; notice that also in this case \( H_1 \) lies between the origin O and \( |w| \).

For the WTV regulariser, the solutions of the separable g,-subproblems can be written in closed form by means of a soft-thresholding operator, see, e.g. [135].

8.4 Subproblem for the primal variable r

For 1 ≤ q < +∞ and \( F_q \) as in (6.6), after dropping all terms not depending on r in (8.3), subproblem (8.7) reads

\[
r^{(j+1)} = \arg\min_{r \in \mathbb{R}^M} \left\{ \mu^{(j)} L_q(r; b) - \langle \rho_r^{(j)}, r - A u^{(j+1)} \rangle + \frac{\beta}{2} \| r - A u^{(j+1)} \|_2^2 \right\}
\]

\[
= \arg\min_{r \in \mathbb{R}^M} \left\{ \lambda^{(j)} L_q(r; b) + \frac{1}{2} \| r - y^{(j)} \|_2^2 \right\}. \tag{8.27}
\]

where the variables \( \lambda^{(j)} \in \mathbb{R}^M \) and \( y^{(j)} \in \mathbb{R}^M \) are defined by

\[
\lambda^{(j)} := \frac{\mu^{(j)}}{\beta_r}, \quad y^{(j)} := A u^{(j)} + \frac{1}{\beta_r} \rho_r^{(j)}. \tag{8.28}
\]

Note that, as already observed in Section 6.3, the regularisation parameter \( \mu \) is not assumed to be fixed but it is rather estimated along the ADMM iterations (whence 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
The constrained minimisation problem in (8.29), where the constraint set is the projection \( \pi_{B^q} \), has been proposed, which improves previous algorithms by not performing the computational viewpoint. Nonetheless, in [47] a complexity linear complexity algorithm has been proposed, which improves previous \( O(M^2) \) and \( O(M \log M) \) strategies considered, e.g., in [70, 8, 53].

We remark that, although the presence of the regularisation parameter \( \mu \) is not explicit in problem (8.29), it is actually embedded in the radius \( \delta_q \).

When the underlying noise is AIU, i.e. \( F_q \) is set as in (6.7), the \( \tau \)-update can be expressed as the constrained minimisation problem in (8.29), where the constraint set is the \( \ell_\infty \)-ball with radius \( \delta_\infty \) defined in (2.12).

Note that the projections onto the \( \ell_\infty \)-balls can be efficiently computed by:

\[
q = 2: \quad \pi_{B_2^q}(\mathbf{y}^{(j)}) = \begin{cases} \mathbf{y}^{(j)} & \text{if } \|\mathbf{y}^{(j)}\|_2 \leq \delta_2 \\ \text{min}(\delta_2, \|\mathbf{y}^{(j)}\|_2) \mathbf{y}^{(j)} \|\mathbf{y}^{(j)}\|_2 & \text{otherwise} \end{cases} \\
q = +\infty: \quad \pi_{B_\infty^q}(\mathbf{y}^{(j)}) = \begin{cases} \mathbf{y}^{(j)} & \text{if } \|\mathbf{y}^{(j)}\|_\infty \leq \delta_\infty \\ \text{min}(\max(\mathbf{y}^{(j)}, -\delta_\infty), \delta_\infty) & \text{otherwise} \end{cases}
\]

where the projections have to be intended componentwise.

For \( q = 1 \), the projection can be computed as follows:

\[
q = 1: \quad \pi_{B_1^q}(\mathbf{y}^{(j)}) = \begin{cases} \mathbf{y}^{(j)} & \text{if } \|\mathbf{y}^{(j)}\|_1 \leq \delta_1 \\ \text{sign}(\mathbf{y}^{(j)}) \mathbf{x} & \text{otherwise} \end{cases}
\]

where \( \mathbf{x} = \text{max}(\mathbf{y}^{(j)} - \tau, 0) \) and \( \tau \in \mathbb{R} : \|\mathbf{x}\|_1 = \delta_1 \).

Setting a suitable \( \tau \) for the problem at hand is possibly a very expensive task from the computational viewpoint. Nonetheless, in [47] 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 [70, 8, 53].

In Figure 16, we show the \( \ell_q \)-balls \( B^q_2(\mathbf{0}) \) for the considered choices of \( q \) in 2-dimensional settings. In the three plots, we also report the vector \( \mathbf{y}^{(j)} \) in the case it does not belong to \( B^q_2(\mathbf{0}) \), and the projection \( \pi(\mathbf{y}^{(j)}) \) onto the ball.

The alternating scheme outlined in Section 3.3 requires to solve the \( u \)-update via the ADMM until a fixed tolerance has been reached after having performed the \( \Theta \)-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 formalised in the following Algorithm 1. As clearly detailed in Section 7, the estimation of the parameters involved in the expression of the WTV and WDTV regularisers represents a further computational bottleneck. Therefore, one can decide to further lighten the algorithmic scheme in Algorithm 1 by not performing the parameters update at each iteration \( j \), but at every few iterations.
Algorithm 1: Joint ADMM-scheme for hyperparameter estimation and image reconstruction

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

\* Initialisation: ; ; \( u^{(0)}, \rho_t^{(0)} = 0_{2N}, \rho_r^{(0)} = 0_M \)

\* Nested alternating scheme:

for \( j = 0, 1, 2, \ldots \) until convergence do:

- parameters update
  update \( \Theta^{(j+1)} \) as detailed in Section 7.2, 7.3 or 7.4

- primal variables update
  update \( u^{(j+1)} \) by solving (8.10)
  update \( g^{(j+1)} \) as detailed in Section 8.3
  update \( r^{(j+1)} \) as detailed in Section 8.4

- dual variables update
  update \( \rho_t^{(j+1)} \) by (8.8)
  update \( \rho_r^{(j+1)} \) by (8.9)

end for

9 Applications to image restoration

In this section, we evaluate the performances of the space-variant regularisers discussed so far, namely the WTV, the WTV\( ^{sv} \), and WDTV\( ^{sv} \) regularisers in comparison with the space-invariant TV [118] and TV\( ^p \) [90] regularisers. As an example, we will consider the problem of image deblurring, for which the forward linear operator \( A \in \mathbb{R}^{N \times N} \) in (2.8) 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 regularisers of interest will be tested on the restoration of images characterised by different global and local properties. More specifically, we will consider the geometric image in Figure 17a, which is purely piece-wise constant, the skyscraper image in Figure 17b, which presents a mixture of piece-wise constant, piece-wise linear and textured features, and the stairs image in Figure 17c, which is highly textured with fine oriented details. The three test images have all been corrupted by space-invariant Gaussian blur defined by a convolution kernel generated using the Matlab routine \texttt{fspecial} with parameters \texttt{band} = 5 and \texttt{sigma} = 1. The \texttt{band} parameter represents the side length (in pixels) of the square support of the kernel, whereas \texttt{sigma} is the standard deviation (in pixels) of the isotropic bivariate Gaussian distribution defining the kernel in continuous settings. Then, the blurred images have been degraded by AIGG noise realisations 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 test image and \( q = +\infty \) (uniform noise) for the stairs test image. The blur- and noise-corrupted images are displayed on the bottom row of Figure 17.

The quality of the obtained restorations \( u^* \) versus the associated ground-truth image \( u \) is assessed by means of two scalar measures, the Improved Signal-to-Noise Ratio (ISNR)

\[
\text{ISNR}(b; u; u^*) := 10 \log_{10} \left( \frac{\|b - u\|^2_2}{\|u^* - u\|^2_2} \right),
\]

(9.1)

and the Structural Similarity Index (SSIM) [128]. The larger the ISNR and SSIM values, the higher the restoration quality. For all tests, the ADMM iterations are stopped as soon as

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

46
The penalty parameters $\beta_g$, $\beta_r$ are manually set.

The estimation of the hyperparameters in the space-variant regularisers WTV, WTV$^{sv}_p$ and WDTV$^{sv}_p$ is performed by manually setting the radius $r$, so as to attain the highest ISNR and SSIM values. Moreover, for the WTV regulariser, the existence of a very efficient procedure for the computation of the $\{\alpha_i\}_i$ weights allows to update the $\alpha$-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 WTV$^p$ and $\{\alpha_i, p_i, \theta_i, a_i\}_i$ in the WDTV$^p$ regulariser, we update the maps of parameters every 30 iterations.

For what concerns the estimation of the local $p_i$ for the WTV$^{sv}_p$ and the WDTV$^{sv}_p$, as well as of the global $p$ in the TV$^p$ regulariser, we fix the compact set $[\epsilon, R]$ of Propositions 3 and 4, equal to $[0.5, 2]$. Notice the choice of the lower bound allows the $u$-estimation problem (8.1) to result in non-convex regularisers. This implies that a particular attention has to be put in the design of a suitable initial guess, which can prevent the performed hypermodels to get trapped in bad local minima.

We initialise Algorithm 1 using a suitable initialisation minimising noise whiteness for a standard Tikhonov-L2 problem as proposed recently in [91].

**Restoration of geometric** First, we discuss the performance of the considered regularisers for the restoration of the geometric test image. The restored images are shown in Figure 18, while the achieved ISNR and SSIM values are reported in Table 2. Notice that, in general, the TV regulariser is well-suited for the restoration of piece-wise constant images; however, as discussed in 1.2, it also suffers from several drawbacks. Our results confirm that using instead a TV$^p$ regulariser ($p = 0.5$) reduces such artefacts. Overall, the three considered space-variant regularisers appear to be more effective than plain TV.

In Figure 19, we show the output maps of parameters for the WTV, WTV$^{sv}_p$ and WDTV$^{sv}_p$ regularisers, obtained with $r = 1$, $r = 3$ and $r = 1$, respectively. For all three regularisers, the $\alpha$-maps present higher weights in the background, while showing that weaker regularisation is performed along the profiles of the geometrical figures. Notice that the $p$-values in the WTV$^p$ and in the WDTV$^p$ approach 2 in the background, which combined with the high regularisation weights allow for an effective smoothing and noise removal therein. Finally, the $\theta$ and $a$ maps in the bottom row of Figure 19 show that the estimator detects a clear directionality in correspondence of the figure profiles, where the angles $\theta$ have been accurately estimated and $a$ assume small values.
Figure 18: From top to bottom: for the test image geometric, observed image $b$, performance of the TV, the TV$_p$ (with output $p = 0.5$), the WTV, the WTV$^{sv}_p$ and the WDTV$^{sv}_p$ regularisers with the respective close-up(s).
Figure 19: From top to bottom: for the test image geometric, output maps of the parameters for the WTV \((r = 1)\), the WTV\(_p\) \((r = 3)\) and the WDTV\(_p\) \((r = 1)\) regularisers.
Table 2: ISNR and SSIM values achieved by the considered regularisers for the three test images corrupted by blur and different AIGG noises.

|          | TV   | TV<sub>p</sub> | WTV | WTV<sub>sv</sub><sup>p</sup> | WDTV<sub>sv</sub><sup>p</sup> |
|----------|------|----------------|------|--------------------------|--------------------------|
| 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                   |

**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 20, one can clearly notice how each additional space-variant parameter effectively contributes in gradually improving the output result, as also reflected in the ISNR and SSIM values reported in Table 2. In Figure 21, we show the output map of parameters for the WTV, WTV<sub>sv</sub><sup>p</sup> and WDTV<sub>sv</sub><sup>p</sup> regularisers, computed for \( r = 15 \), \( r = 15 \) and \( r = 3 \), respectively.

As a general comment, we highlight that the weights \( \alpha_i \) assume larger values on the background so that a strong regularisation is performed regardless of the corresponding \( p_i \); in fact, the \( p \)-maps for the WTV<sub>sv</sub><sup>p</sup> and the WDTV<sub>sv</sub><sup>p</sup> regularisers appear to be different in this region. From the \( \theta \)-map reported in the bottom row of Figure 21, we observe that also in this case the estimator is capable of detecting the direction of the buildings profile as well as the horizontal oriented texture. Finally, the \( a \) values indicate a stronger dominance in terms of directionality along the edges of the buildings.

**Restoration of stairs** In this final test, we consider the highly textured image stairs. From the ISNR and SSIM values reported in Table 2 and from the restored images displayed in Figure 22, we notice that the WTV regulariser is outperformed by the space-invariant TV<sub>p</sub> regulariser, with output estimated \( p = 1.56 \), in terms of SSIM. In fact, the TV<sub>p</sub> performs a type of regularisation which, although global, appears to be more suitable for describing the image of interest. A further improvement is achieved by the WTV<sub>sv</sub><sup>p</sup> regulariser which preserves the textured regions in the image while smoothing out the limited piece-wise constant parts. Finally, the WDTV<sub>sv</sub><sup>p</sup> regularisation term slightly refines the output by driving the regularisation along the local directionalities.

The maps of the parameters for the space-variant WTV with radius \( r = 2 \), and for WTV<sub>sv</sub><sup>p</sup> and WDTV<sub>sv</sub><sup>p</sup> with radius \( r = 1 \) are shown in Figures 23. From the \( p \)-maps for WTV<sub>sv</sub><sup>p</sup> and WDTV<sub>sv</sub><sup>p</sup>, 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 regularisation, combined with the suitable local weights, is more effective in dealing with this sort of images, due to the presence of large textured regions where the distribution of gradients is thus very spread. The directions in the central part of the image are precisely detected, as shown in the \( \theta \)-map, as well as the confidence in the estimation, represented by \( a \), which appear to be particularly relevant along the steps.

10 User’s 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 regulariser. As one could expect, the question about which regulariser performs best in absolute terms does not have a unique answer. The adoption of more and more general regularisation terms does not always pay back as the overall performance has to be evaluated with respect to the trade-off between quality of the restorations.
Figure 20: From top to bottom: for the test image skyscraper, observed image $b$, performance of the TV, the $TV_p$ (with output $p = 0.5$), the WTV, the $WTV^p_s$ and the $WDTV^s_p$ regularisers with the respective close-up(s).
and computational effort. In 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 $\alpha$-update and for the $g$-subproblem in the ADMM-based scheme. However, for the stairs image, the typical shortage of a TV-type regularisation, even if weighted, has emerged. The selection of the regulariser 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 maximise the ISNR and 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 regularisers. 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 WTV$^sv_p$ regularisers, 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 WDTV$^sv_p$ regulariser, a smaller radius has to be selected.

3. The curse of non-convexity. The convergence of the outlined numerical scheme aimed at solving a possibly non-convex 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 parameter $\beta_g, \beta_r$ are set in a suitable manner - typically, $\beta_g, \beta_r \approx 10^4, 10^5$ - empirical convergence is observed.
Figure 22: From top to bottom: for the test image stairs, observed image b, performance of the TV, the TV\(_p\) (with output \(p = 1.56\)), the WTV, the WTV\(_{sv}\) and the WDTV\(_{sv}\) regularisers with the respective close-up(s).
Figure 23: From top to bottom: for the test image *stairs*, output maps of the parameters for the WTV \((r = 2)\), WTV\(_p\) \((r = 1)\) and WDTV\(_p\) \((r = 1)\) regularisers.
11 Challenges

We conclude this review by listing in the following some challenging future research directions which could enrich this work from both the theoretical and the applied point 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, optimisation, numerical and linear algebra, medical imaging...) favouring the development of new and unexplored research.

1. The detailed analytical study in an infinite-dimensional framework of the non-smooth, non-convex and space-variant regularisation models discussed in this work is expected to provide more insights on the structure of the expected solutions by means of duality tools, functional calculus and non-standard Lebesgue/Sobolev calculus in spaces with variable exponents.

2. The development of a rigorous theoretical framework guaranteeing convergence to (at least) stationary points for the non-convex ADMM Algorithm 1 is highly non-trivial and practically made challenging due to the parameter estimation performed jointly along the iterations. Note that even for the IAS algorithm [24] where parameters and iterates are updated sequentially, a convergence proof in general non-convex scenarios is still missing, as only partial results in convex (quadratic) cases are available.

3. The use of a similar space-variant modelling for more general regularisers defined, for instance, in terms of wavelet expansions [30] and higher-order differential operators (see, e.g., [94]).

4. Similarly as for the case of anisotropic diffusion, we expect that the use of suitable adaptive discretisation stencils [35, 59, 48, 39] built on the estimated local directional information and/or relying on the suitable definition of appropriate transfer operators and staggered grids [112] could improve upon the quality of the numerical reconstructions by describing anisotropy on the image grid in a more precise way.

5. Following [26], we believe that the use of a hierarchical modelling 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 [126, 49] could lead to new hybrid hyperparameter selection strategies relying on the sole observation of the given corrupted image \( b \). These ideas are expected indeed to speed up the performance of Algorithm 1.

7. As an obvious field of applications due to the recent use of analogous models in medical imaging problems such as MRI, PET and CT applications [56, 55, 54, 125], we expect that the use of a structural and adaptive modelling could significantly improve the quality of the reconstructions and favour, at the same time, the exploitation of structural information in multimodal image analysis.

8. We wonder whether the flexible underlying statistical modelling proposed in this work could be used in a GAN framework [64] from two different perspectives: firstly, to compare qualitatively (and/or quantitatively, provided that a good quality measure is used [13]) the highly parametric model-driven BGGD-type distribution of the target image with the target distribution estimated in the generative step; secondly, to provide an alternative way of estimating the hyperparameters, thus replacing the parameter estimation step (3.33) by a fully data-driven approach.

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 regularisation model over the last thirty years, the TV functional. After recalling its genesis,
its main features and shortcomings in Section 1 and having fixed some notations in Section 2, we described in Section 3 how the rigidity of existing TV-type image regularisation models can be overcome within the setting of non-stationary 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. Analytically, we show in Section 4 that these non-stationary priors can be put in close correspondence to a large class of space-variant image regularisation 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 regularisation, sharpness and anisotropy) is then showed in Section 5 to correspond geometrically to change at each pixel the definition of the constrain 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 minimisation 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 new challenging research directions as finally described in Section 11.

The incredible potential offered nowadays by the possibility of combining together 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 regularisation model still interesting for the whole applied mathematics community. The swan song of TV is still far away in time. Due to both the profound understanding of this powerful, yet simple, image regularisation model carried out over years and the recent advances in large-scale data exploitation and numerical optimisation favouring the development of its many extensions, we expect that the descendent models stemmed from TV have still much to say and could at the same time enrich and be enriched by the increasingly popular interest towards data-driven approaches showed by the analytical, signal-processing, statistical and optimisation communities.

Acknowledgements

The authors are warmly grateful to Laure Blanc-Féraud and Xavier Descombes for their precious suggestions and comments on the statistical derivation of our models and to Simone Parisotto for his advice on their geometrical interpretation. Research of AL, MP, FS was supported by the “National Group for Scientific Computation (GNCS-INDAM)” and by ex60 project by the University of Bologna “Funds for selected research topics”. LC acknowledges the support received by the EU H2020 RISE NoMADS, GA 777826 and the UCA JEDI IDEX grant DEP “Attractivité du territoire”.

A Generalised Gaussian distributions

**Definition 10** (GG cumulative distribution function). The cumulative distribution function (cdf) of a scalar random variable $X \sim \text{GG}(\eta, \gamma, s)$ reads

\[ F_X(x) = \frac{1}{2} + \frac{\text{sign}(x - \eta)}{2 \Gamma(1/s)} \Gamma\left(\frac{1}{s}, \gamma \frac{|x - \eta|}{s}\right), \quad (A.1) \]

with $\Gamma$ the lower incomplete Gamma function defined in (2.2) and $\Gamma$ the Gamma function defined in (2.3).

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

\[ Y = \sum_{i=1}^{M} X_i \sim \text{Gamma}\left(\nu, \sum_{i=1}^{n} z_i\right). \quad (A.2) \]
Lemma 3. If $X \sim \text{GG}(0, \gamma, s)$, then it holds true that

$$Y = g(X) = |X|^s \sim \text{Gamma}(\nu, z), \quad \nu = \frac{1}{\gamma s}, \quad z = \frac{1}{s}. \quad (A.3)$$

Proof. We have:

$$F_Y(y) = P(Y \in ]-\infty, y[) = P(Y \in [0, y]) \quad (A.4)$$

$$= P(X \in g^{-1}([0, y])) = P\left(X \in \left[-y^{1/s}, +y^{1/s}\right]\right) \quad (A.5)$$

$$= 1 - 2 F_X\left(-y^{1/s}\right) = \frac{1}{2 \Gamma(1/s)} \Gamma\left(\frac{1}{s}, \gamma^s y\right) \quad (A.6)$$

where the first and second equality in (A.4) come from the definition of cdf and from noticing that $Y$ cannot assume negative values, respectively, $g^{-1}([0, y])$ in (A.5) denotes the preimage of interval $[0, y]$ under the function $g$ defined in (A.3). (A.6) 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 $p_Y$ can be obtained by differentiating the cdf $F_Y$ in (A.6). To this aim, first we rewrite $F_Y$ in the following equivalent composite form:

$$F_Y(y) = F_2(F_1(y)), \quad F_1(y) = \gamma^s y, \quad F_2(w) = \frac{1}{2 \Gamma(1/s)} \int_0^w t^{\frac{1}{s}-1} e^{-t} dt, \quad (A.7)$$

where we also replaced the explicit expression of the lower incomplete Gamma function $\Gamma$ given in (2.2). By applying the chain rule of differentiation, we have

$$P_Y(y) = \frac{d}{dy} F_Y = \frac{d}{dy} F_1(y) \times \frac{d}{dw} F_2(\gamma^s y) \quad (A.8)$$

$$= \gamma^s \times \frac{1}{2 \Gamma(1/s)} (\gamma^s y)^{\frac{1}{s}-1} \exp(-\gamma^s y) \quad (A.9)$$

$$= \frac{\gamma}{2 \Gamma(1/s)} y^{\frac{1}{s}-1} \exp(-\gamma^s y). \quad (A.10)$$

A simple one-to-one reparameterisation of (A.10), namely $s = 1/z$, $\gamma = (1/\nu)^s$, together with the recall of definition (2.6), leads to (A.3) and thus completes the proof.

Proof of Proposition 1. Statement (2.9) follows straightforwardly from Lemma 2 and Lemma 3, whereas (2.10) comes from (2.9) and the well-known expressions for the mean and variance of a Gamma-distributed random variable.

References

[1] Arridge, S., Kolehmainen, V., and Schweiger, M. J. Reconstruction and regularization in optical tomography. In *Mathematical Methods in Biomedical Imaging and Intensity-Modulated Radiation Therapy (IMRT)* (Scuola Normale Superiore, 2008), Y. A. Censor, M. Jiang, and Louis, Eds.

[2] Arridge, S., Maass, P., Öktem, O., and Schönlieb, C. B. Solving inverse problems using data-driven models. *Acta Numer.* 28 (2019), 1–174.

[3] Babacan, S. D., Molina, R., and Katsaggelos, A. Generalized Gaussian Markov random field image restoration using variational distribution approximation. *2008 IEEE Internat. Conf. Acoustics, Speech and Signal Proc.* (2008), 1265–1268.

[4] Babacan, S. D., Molina, R., and Katsaggelos, A. Parameter estimation in TV image restoration using variational distribution approximation. *IEEE Trans. Image Proc.* 17 (2008), 326–339.

[5] Bathke, C., Kluth, T., Brandt, C., and Maass, P. Improved image reconstruction in magnetic particle imaging using structural a priori information. *Internat. J. Magnetic Particle Imaging* 3, 1 (2017).
[6] Bayram, I., and Kamasak, M. E. Directional total variation. **IEEE Signal Proc. Letters** 19, 12 (Dec 2012), 781–784.

[7] Beck, A., and Teboulle, M. A fast iterative shrinkage-thresholding algorithm for linear inverse problems. **SIAM J. Imaging Sci.** 2 (2009), 183–202.

[8] Bentley, J., and McIlroy, M. Engineering a sort function. **Software: Practice and Experience** 23 (1993), 1249–1265.

[9] Bertocchi, C., Chouzenoux, E., Corbíneau, M.-C., Pesquet, J.-C., and Prato, M. Deep unfolding of a proximal interior point method for image restoration. **Inverse Problems** 36, 3 (2020).

[10] Blake, A., Kohli, P., and Rother, C. *Markov Random Fields for Vision and Image Processing*. The MIT Press, 2011.

[11] Boink, Y. G., Lagerwerf, M. J., Steenbergen, W., van Gils, S. A., Manohar, S., and Brune, C. A framework for directional and higher-order reconstruction in photoacoustic tomography. *Physics Med. Biol.* 63 (2018).

[12] Bolte, J., Sabach, S., and Teboulle, M. Nonconvex Lagrangian-based optimization: Monitoring schemes and global convergence. *Math. Oper. Res.* 43, 4 (2018), 1210–1232.

[13] Borji, A. Pros and cons of GAN evaluation measures. *Comput. Vis. Image Und.* 179 (2019), 41–65.

[14] Boyd, S., Parikh, N., Chu, E., Peleato, B., and Eckstein, J. Distributed optimization and statistical learning via the alternating direction method of multipliers. *Found. and Trends Mach. Learning* 3 (Jan. 2011), 1–122.

[15] Bredies, K., Kunisch, K., and Pock, T. Total generalized variation. **SIAM J. Imaging Sci.** 3, 3 (2010), 492–526.

[16] Bungert, L., Coomes, D. A., Ehrhardt, J. M., Rasch, J., Reisenhofer, R., and Schönnlieb, C.-B. Blind image fusion for hyperspectral imaging with the directional total variation. **Inverse Problems** 34, 4 (Mar 2018).

[17] Burger, M., and Osher, S. *A Guide to the TV Zoo*, vol. 2090. Springer, Cham, 2013.

[18] Byrd, R., Lu, P., Nocedal, J., and Zhu, C. A limited memory algorithm for bound constrained optimization. *SIAM J. Sci. Comput.* 16 (Sept. 1995), 1190–1208.

[19] Calatroni, L., Cao, C., Los Reyes, J. C. D., Schönnlieb, C.-B., and Valkonen, T. Bilevel approaches for learning of variational imaging models. Berlin, Boston: De Gruyter, 2017, pp. 252–290.

[20] Calatroni, L., d’Autume, M., Hocking, R., Panayotova, S., Parisotto, S., Ricciardi, P., and Schönnlieb, C.-B. Unveiling the invisible: mathematical methods for restoring and interpreting illuminated manuscripts. *Heritage Science* 6, 1 (2018), 56.

[21] Calatroni, L., Lanza, A., Pragliola, M., and Sgallari, F. A flexible space-variant anisotropic regularization for image restoration with automated parameter selection. **SIAM J. Imaging Sci.** 12 (2019), 1001–1037.

[22] Calatroni, L., Lanza, A., Pragliola, M., and Sgallari, F. Adaptive parameter selection for weighted-TV image reconstruction problems. In *J.Phys.: Conf. Series, NCMIP 2019* (2020), vol. 1476, pp. 541–547.

[23] Calvetti, D., Morini, S., Reichel, L., and Sgallari, F. Tikhonov regularization and the L-curve for large discrete ill-posed problems. *J. Comput. Appl. Math.* 123, 1 (2000), 423–446. Numerical Analysis 2000. Vol. III: Linear Algebra.
Calvetti, D., Pascarella, A., Pitoli, F., Somersalo, E., and Vantaggi, B. A hierarchical krylov–bayes iterative inverse solver for MEG with physiological preconditioning. *Inverse Problems* 31, 12 (Nov 2015).

Calvetti, D., Pragliola, M., and Somersalo, E. Sparsity promoting hybrid solvers for hierarchical bayesian inverse problems. *SIAM J. Sci. Comput.* 42 (2020), A3761–A3784.

Calvetti, D., Pragliola, M., Somersalo, E., and Strang, A. Sparse reconstructions from few noisy data: analysis of hierarchical Bayesian models with generalized gamma hyperpriors. *Inverse Problems* 36, 2 (Jan 2020).

Calvetti, D., and Somersalo, E. Hypermodels in the Bayesian imaging framework. *Inverse Problems* 24, 3 (2008).

Calvetti, D., and Somersalo, E. *Statistical Methods in Imaging*. Springer, New York, 2015.

Calvetti, D., Somersalo, E., and Strang, A. Hierarchical Bayesian models and sparsity: ℓ2-magic. *Inverse Problems* 35 (2019).

Candès, E. J., Eldar, Y. C., Needell, D., and Randall, P. Compressed sensing with coherent and redundant dictionaries. *Appl. Computat. Harmon. Anal.* 31, 1 (2011), 59–73.

Caselles, V., Chambolle, A., Cremers, D., Novaga, M., and Pock, T. An introduction to total variation for image analysis. *Theoretical Foundations and Numerical Methods for Sparse Recovery, De Gruyter, Radon Series Comp. Appl. Math.* 9 (2010), 263–340.

Caselles, V., Chambolle, A., and Novaga, M. The discontinuity set of solutions of the TV denoising problem and some extensions. *Multiscale Model. Sim.* 6 (2007), 879–894.

Caselles, V., Chambolle, A., and Novaga, M. *Total Variation in Imaging*. Springer, New York, NY, 2015, pp. 1455–1499.

Chambolle, A. An algorithm for total variation minimization and applications. *J. Math. Imaging Vis.* 20 (2004), 89–97.

Chambolle, A., Levine, S. E., and Lucier, B. J. An upwind finite-difference method for total variation–based image smoothing. *SIAM J. Imaging Sci.* 4, 1 (2011), 277–299.

Chambolle, A., and Lions, P.-L. Image recovery via total variation minimization and related problems. *Numer. Math.* 76, 2 (1997), 167–188.

Chambolle, A., and Pock, T. A first-order primal-dual algorithm for convex problems with applications to imaging. *J. Math. Imaging Vis.* 40 (2011), 120–145.

Chambolle, A., and Pock, T. An introduction to continuous optimization for imaging, *Acta Numer.* 25 (2016), 161–319.

Chambolle, A., and Pock, T. Approximating the Total Variation with Finite Differences or Finite Elements. In *Handbook Numer.Anal.: Geometric Part. Diff. Eqs II*. 2021.

Chan, T., Marquina, A., and Mulet, P. High-order total variation-based image restoration. *SIAM J. Sci. Comput.* 22 (2000), 503–516.

Chan, T. F., Esedooglu, S., and Park, F. A fourth order dual method for staircase reduction in texture extraction and image restoration problems. In *2010 IEEE Internat.Conf. Image Proc.* (2010), pp. 4137–4140.

Chan, T. F., Golub, G. H., and Mulet, P. A nonlinear primal-dual method for total variation based image restoration. *SIAM J. Sci. Comput.* 20, 6 (1999), 1964–1977.

Chen, Y., and Pock, T. Trainable nonlinear reaction diffusion: A flexible framework for fast and effective image restoration. *IEEE Trans. Pattern Anal. Mach. Intell.* 39, 6 (2017), 1256–1272.
[44] Chung, C. V., Los Reyes, J. C. D., and Schönlieb, C. B. Learning optimal spatially-dependent regularization parameters in total variation image denoising. *Inverse Problems* 33, 7 (Jun 2017).

[45] Combettes, P., and Pesquet, J. *Proximal Splitting Methods in Signal Processing*. Springer New York, 2011, pp. 185–212.

[46] Condat, L. Primal–dual splitting method for convex optimization involving lipschitzian, proximable and linear composite terms. *J. Optimiz. Theory Appl.* 158 (2013), 460–479.

[47] Condat, L. Fast projection onto the simplex and the ℓ₁ ball. *Math. Program.* 158 (2016), 575–585.

[48] Condat, L. Discrete total variation: New definition and minimization. *SIAM J. Imaging Sci.* 10, 3 (2017), 1258–1290.

[49] De Bortoli, V., Durmus, A., Pereyra, M., and Vidal, A. F. Maximum likelihood estimation of regularization parameters in high-dimensional inverse problems: An empirical Bayesian approach. Part II: Theoretical analysis. *SIAM J. Imaging Sci.* 13, 4 (2020), 1990–2028.

[50] Demircan-Tureyen, E., and Kamasak, M. E. Adaptive direction-guided structure tensor total variation, 2020.

[51] Descombes, X., Sigelle, M., and Preteux, F. Estimating Gaussian Markov random field parameters in a nonstationary framework: Application to remote sensing imaging. *IEEE Trans. Image Proc.* 8, 4 (1999), 490–503.

[52] Dong, Y., and Schönlieb, C.-B. Tomographic reconstruction with spatially varying parameter selection. *Inverse Problems* 36 (2020).

[53] Duchi, J. C., Shalev-Shwartz, S., Singer, Y., and Chandra, T. Efficient projections onto the ℓ₁,1-ball for learning in high dimensions. In *ICML ’08: Proc. 25th Intern. Conf. Machine Learning* (2008), vol. 3, pp. 272–279.

[54] Ehrhardt, M., and Betcke, M. Multicontrast MRI reconstruction with structure-guided total variation. *SIAM J. Imaging Sci.* 9, 3 (2016), 1084–1106.

[55] Ehrhardt, M. J., Markiewicz, P., Liljeroth, M., Barnes, A., Kolehmainen, V., Duncan, J. S., Pizarro, L., Atkinson, D., Hutton, B. F., Ourselin, S., Thielemans, K., and Arridge, S. R. PET reconstruction with an anatomical MRI prior using parallel level sets. *IEEE Trans. Med. Imaging* 35, 9 (2016), 2189–2199.

[56] Ehrhardt, M. J., Thielemans, K., Pizarro, L., Atkinson, D., Ourselin, S., Hutton, B. F., and Arridge, S. R. Joint reconstruction of PET-MRI by exploiting structural similarity. *Inverse Problems* 31, 1 (Dec 2014).

[57] Engl, H., Hanke, M., and Neubauer, A. *Regularization of Inverse Problems*. Mathematics and Its Applications. Springer Netherlands, 2000.

[58] Estellers, V., Soatto, S., and Bresson, X. Adaptive regularization with the structure tensor. *IEEE Trans. Image Proc.* 24 (2015), 1777–1790.

[59] Fehrenbach, J., and Mirebeau, J.-M. Sparse non-negative stencils for anisotropic diffusion. *J. Math. Imaging Vis.* 49, 1 (May 2014), 123–147.

[60] Fenu, C., Reichel, L., and Rodriguez, G. GCV for tikhonov regularization via global golub-kahan decomposition. *Numer. Linear Algebr. Appl.* 23 (02 2016).

[61] Geman, S., and Geman, D. Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Trans. Pattern Anal. Mach. Intell.* PAMI-6 (1984), 721–741.

[62] Giakoumis, I., and Pitas, I. Digital restoration of painting cracks. In *ISCAS ’98. Proc. 1998 IEEE Intern. Symp. Circuits Systems (Cat. No.98CH36187)* (1998), vol. 4, pp. 269–272.
[63] Golub, G. H., Heath, M., and Wahba, G. Generalized cross-validation as a method for choosing a good ridge parameter. *Technometrics* 21, 2 (1979), 215–223.

[64] Goodfellow, I. J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., and Bengio, Y. Generative adversarial nets. In Proc. 27th Intern. Conf. Neural Information Proc. Syst.- Volume 2 (Cambridge, MA, USA, 2014), NIPS’14, MIT Press, p. 2672–2680.

[65] Grasmair, M., and Lenzen, F. Anisotropic total variation filtering. *Appl. Math. Opt.* 62, 3 (2010), 323–339.

[66] Guo, M., Han, C., Wang, W., Zhong, S., Lv, R., and Liu, Z. A novel truncated nonconvex nonsmooth variational method for SAR image despeckling. *Remote Sens. Lett.* 12 (2021), 174–183.

[67] H. Raguet, J. F., and Peyré, G. A generalized forward-backward splitting. *SIAM J. Imaging Sci.* 6, 3 (2013), 1199–1126.

[68] Hanke, M. *Chapter 8: The Discrepancy Principle*. SIAM Philadelphia, 2017, pp. 55–60.

[69] He, C., Hu, C., Zhang, W., and Shi, B. A fast adaptive parameter estimation for total variation image restoration. *IEEE Trans. Image Proc.* 23, 12 (Dec 2014), 4954–4967.

[70] Held, M., Wolfe, P., and Crowder, H. Validation of subgradient optimization. *Math. Program.* 6 (1974), 62–88.

[71] Hintermüller, M., and Rautenberg, C. N. Optimal selection of the regularization function in a weighted total variation model. part i: Modelling and theory. *J. Math. Imaging Vis.* 59, 3 (Nov 2017), 498–514.

[72] Hintermüller, M., Rautenberg, C. N., Wu, T., and Langer, A. 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, 3 (2017), 515–533.

[73] Hintermüller, M., and Stadler, G. An infeasible primal-dual algorithm for total bounded variation–based inf-convolution-type image restoration. *SIAM J. Sci. Comput.* 28, 1 (2006), 1–23.

[74] Hintermüller, M., and Papafitsoros, K. Chapter 11 - 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., vol. 20 of *Handbook of Numerical Analysis*. Elsevier, 2019, pp. 437–502.

[75] Hintermüller, M., Papafitsoros, K., and Rautenberg, C. N. Analytical aspects of spatially adapted total variation regularisation. *J.Math. Anal. App.* 454, 2 (2017), 891 – 935.

[76] Hinz, T., Navarro-Guerrero, N., Magg, S., and Wermter, S. Speeding up the hyperparameter optimization of deep convolutional neural networks. *Intern. J. Comput. Intell. Appl.* 17, 02 (2018), 1850008.

[77] Jalalzai, K. Some remarks on the staircasing phenomenon in total variation-based image denoising. *J. Math. Imaging Vis.* 54, 2 (2016), 256–268.

[78] Jespersen, K. M., Zangenberg, J., Lowe, T., Withers, P. J., and Mikkelsen, L. P. Fatigue damage assessment of uni-directional non-crimp fabric reinforced polyester composite using x-ray computed tomography. *Comp. Science Tech.* 136 (2016), 94–103.

[79] Kaipio, J., and Somersalo, E. *Statistical and computational inverse problems*. Springer-Verlag, New York, 2004.

[80] Kaipio, J. P., Kolehmainen, V., Vauhkonen, M., and Somersalo, E. Inverse problems with structural prior information. *Inverse Problems* 15, 3 (Jan 1999), 713–729.
[81] K. Kunisch, and Hintermüller, M. Total bounded variation regularization as a bilaterally constrained optimization problem. *SIAM J. Appl. Math.* 64, 4 (2004), 1311–1333.

[82] Kongskov, R., and Dong, Y. Directional total generalized variation regularization for impulse noise removal. In *Scale Space and Variat. Meth. Comp. Vision* (Cham, 2017), F. Lauze, Y. Dong, and A. B. Dahl, Eds., Springer Intern. Publ., pp. 221–231.

[83] Kongskov, R., Dong, Y., and Knudsen. Directional total generalized variation regularization. *Bit Numer Math* 59 (2019), 903–928.

[84] Krishnan, D., and Fergus, R. Fast image deconvolution using hyper-Laplacian priors. In *Advances in Neural Information Processing Systems* (2009), Y. Bengio, D. Schuurmans, J. Lafferty, C. Williams, and A. Culotta, Eds., vol. 22, Curran Associates, Inc.

[85] Kunisch, K., and Pock, T. A bilevel optimization approach for parameter learning in variational models. *SIAM J. Imaging Sci.* 6, 2 (2013), 938–983.

[86] Kurach, K., Lucic, M., Zhai, X., Michalski, M., and Gelly, S. The GAN landscape: Losses, architectures, regularization, and normalization, 2019.

[87] Lanza, A., Morici, S., Pragliola, M., and Sgallari, F. Space-variant generalised gaussian regularisation for image restoration. *Comput. Meth. Biomech.Biomed. Eng.: Imaging Vis.* 13 (2018), 490–503.

[88] Lanza, A., Morici, S., Pragliola, M., and Sgallari, F. Space-variant TV regularization for image restoration. In *VipIMAGE 2017* (Cham, 2018), J. M. R. Tavares and R. Natal Jorge, Eds., Springer Intern. Publ., pp. 160–169.

[89] Lanza, A., Morici, S., Sciaccitano, F., and Sgallari, F. Whiteness constraints in a unified variational framework for image restoration. *J. Math.Imaging Vis.* 60, 9 (2018), 1573–7683.

[90] Lanza, A., Morici, S., and Sgallari, F. Constrained TVp-ℓ2 model for image restoration. *J. Sci. Comput.* 68, 1 (Jul 2016), 64–91.

[91] Lanza, A., Pragliola, M., and Sgallari, F. Residual whiteness principle for parameter-free image restoration. *Electro. Trans. Numer.Anal.* 53 (2020), 329–351.

[92] Le Montagner, Y., Angelini, E. D., and Olivo-Marin, J. An unbiased risk estimator for image denoising in the presence of mixed poisson–gaussian noise. *IEEE Trans. Image Proc.* 23, 3 (2014), 1255–1268.

[93] Lefkimmiatis, S., Roussos, A., Maragos, P., and Unser, M. Structure tensor total variation. *SIAM J. Imaging Sci.* 8, 2 (2015), 1090–1122.

[94] Lefkimmiatis, S., Ward, J. P., and Unser, M. Hessian Schatten-norm regularization for linear inverse problems. *IEEE Trans. Image Proc.* 22, 5 (2013), 1873–1888.

[95] Li, S. Z. *Markov Random Field Modeling in Image Analysis*, 3rd ed. Springer-Verlag, Tokyo, Japan, 2009.

[96] Li, Y., Wu, C., and Duan, Y. The TVp regularized Mumford-Shah model for image labeling and segmentation. *IEEE Trans. Image Proc.* 29 (2020), 7061–7075.

[97] Liu, R. W., Li, Y., Liu, Y., Duan, J., Xu, T., and Liu, J. Single-image blind deblurring with hybrid sparsity regularization. In *2017 20th Intern. Conf. Inf. Fusion* (2017), pp. 1–8.

[98] Lozes, F., Elmoataz, A., and Lezoray, O. PDE-based graph signal processing for 3-D color point clouds: Opportunities for cultural heritage. *IEEE Signal Proc. Mag.* 32, 4 (2015), 103–111.

[99] Lucic, M., Kurach, K., Michalski, M., Bousquet, O., and Gelly, S. Are GANs created equal? a large-scale study. In *Proc. 32nd Intern.Conf. Neural Inform. Proc.Systems* (Red Hook, NY, USA, 2018), NIPS’18, Curran Associates Inc., p. 698–707.
[100] Lunz, S., Öktem, O., and Schönlieb, C.-B. Adversarial regularizers in inverse problems. In Advances in Neural Information Processing Systems (2018), S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi, and R. Garnett, Eds., vol. 31, Curran Associates, Inc.

[101] Meyer, Y. Oscillating Patterns in Image Processing and Nonlinear Evolution Equations: The Fifteenth Dean Jacqueline B. Lewis Memorial Lectures. American Mathem. Soc., USA, 2001.

[102] Miyato, T., Kataoka, T., Koyama, M., and Yoshida, Y. Spectral normalization for generative adversarial networks. In Intern. Conf. Learning Repres. (2018).

[103] Monga, V., Li, Y., and Eldar, Y. C. Algorithm unrolling: Interpretable, efficient deep learning for signal and image processing. IEEE Signal Proc. Mag. 38, 2 (2021), 18–44.

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

[105] Nesterov, Y. Introductory lectures on convex optimization: a basic course. Applied optimization. Kluwer Academic Publ., Boston, Dordrecht, London, 2004.

[106] Nikolova, M. Local strong homogeneity of a regularized estimator. SIAM J. Appl. Math. 61 (2000), 633–658.

[107] Ochs, P., Chen, Y., Brox, T., and Pock, T. iPiano: Inertial proximal algorithm for non-convex optimization. SIAM J. Imaging Sci. 7, 2 (2014), 1388–1419.

[108] Oliveira, J., Bioucas-Dias, J., and Figueiredo, M. A. T. Adaptive total variation image deblurring: A majorization-minimization approach. Signal Proc. 89 (2009), 1683–1693.

[109] Pang, Z.-F., Zhang, H.-L., Luo, S., and Zeng, T. Image denoising based on the adaptive weighted tvp regularization. Signal Proc. 167 (2020), 107325.

[110] Papafitsoros, K., and Schönlieb, C.-B. A combined first and second order variational approach for image reconstruction. J. Math. Imaging Vis. 48 (2014), 308–338.

[111] Parisotto, S., Calatroni, L., Bugeau, A., Papadakis, N., and Schönlieb, C. B. Variational osmosis for non-linear image fusion. IEEE Trans. Image Proc. 29 (2020), 5507–5516.

[112] Parisotto, S., Masnou, S., and Schönlieb, C. B. Higher order total directional variation. Part I: Imaging applications. SIAM J. Imaging Sci. 13 (2020), 2063–2104.

[113] Parisotto, S., Masnou, S., and Schönlieb, C. B. Higher order total directional variation. Part II: Analysis. SIAM J. Imaging Sci. 13 (2020), 474–496.

[114] Pereyra, M., and McLaughlin, S. Fast unsupervised bayesian image segmentation with adaptive spatial regularisation. IEEE Trans. Image Proc. 26 (2017), 2577–2587.

[115] Perona, P., and Malik, J. Scale-space and edge detection using anisotropic diffusion. IEEE Trans. Pattern Anal. Mach. Intell. 12, 7 (Jul 1990), 629–639.

[116] Ramamurthy, K. N., Thiagarajan, J. J., and Spanias, A. Fast image registration with non-stationary Gauss-Markov random field templates. In 2009 16th IEEE Internat. Conf. Image Proc. (ICIP) (2009), pp. 185–188.

[117] Roussos, A., and Maragos, P. Tensor-based image diffusions derived from generalizations of the total variation and Beltrami functionals. In 2010 IEEE Intern. Conf. Image Proc. (Sep. 2010), pp. 4141–4144.

[118] Rudin, L. I., Osher, S., and Fatemi, E. Nonlinear total variation based noise removal algorithms. Physica D: Nonlinear Phen. 60, 1 (1992), 259 – 268.
Sandoghchi, S. R., Jaslon, G. T., Wheeler, N. V., Jain, S., Lian, Z., Wooler, J. P., Boardman, R. P., Baddela, N., Chen, Y., Hayes, J., Fokoua, E. N., Bradley, T., Gray, D. R., Mousavi, S. M., Petrovich, M., Poletti, F., and Richardson, D. J. X-ray tomography for structural analysis of microstructured and multimaterial optical fibers and preforms. *Opt. Express* 22, 21 (Oct 2014), 26181–26192.

Scharr, H., Black, M. J., and Haussecker, H. W. Image statistics and anisotropic diffusion. In *Proceedings Ninth IEEE Internat.Conf. on Computer Vision* (Oct 2003), pp. 840–847 vol.2.

Setzer, S., and Steidl, G. *Variational Methods with Higher Order Derivatives in Image Processing, Approximation XII*. Nashboro Press, Brentwood, 2008.

Stein, C. M. Estimation of the Mean of a Multivariate Normal Distribution. *The Annals of Statistics* 9, 6 (1981), 1135 – 1151.

Strong, D., and Chan, T. Edge-preserving and scale-dependent properties of total variation regularization. *Inverse Problems* 19, 6 (nov 2003), S165–S187.

Stuart, A. M. Inverse problems: a Bayesian perspective. *Acta Numer.* 19 (2010), 451–559.

Tovey, R., Benning, M., Brune, C., Lagerwerf, M. J., Collins, S. M., Leary, R. K., Midgley, P. A., and Schönlieb, C. B. Directional sinogram inpainting for limited angle tomography. *Inverse Problems* 35, 2 (jan 2019).

Vidal, A. F., De Bortoli, V., Pereyra, M., and Durmus, A. 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, 4 (2020), 1945–1989.

Wang, Y., Yin, W., and Zeng, J. Global convergence of ADMM in nonconvex nonsmooth optimization. *J. Sci. Comput.* 78, 1 (2019), 29–63.

Wang, Z., Bovik, A., Sheikh, H. R., and Simoncelli, E. P. Image quality assessment: from error visibility to structural similarity. *IEEE Trans. Image Proc.* 13 (2004), 600–612.

Weickert, J. *Anisotropic Diffusion in Image Processing*. B.G. Teubner, Stuttgart, 1998.

Weickert, J., and T. Brox. Diffusion and regularization of vector- and matrix-valued images. In *Inverse Problems, Image Analysis, and Medical Imaging* (Dec 2002), AMS, pp. 251–268.

Wu, C., and Tai, X.-C. Augmented lagrangian method, dual methods, and split bregman iteration for ROF, vectorial TV, and high order models. *SIAM J. Imaging Sci.* 3 (2010), 300–339.

Xie, Q., Zeng, D., Zhao, Q., Meng, D., Xu, Z., Liang, Z., and Ma, J. Robust low-dose ct sinogram preprocessing via exploiting noise-generating mechanism. *IEEE Trans. Med. Imaging* 36 (2017), 2487–2498.

Zhang, H., and Wang, Y. Edge adaptive directional total variation. *The Journal of Engineer.* 2013, 11 (2013), 61–62.

Zhang, J., Gelb, A., and Scarnati, T. Empirical bayesian inference using joint sparsity. *arXiv:2103.15618* (2021).

Zhao, H., Liu, Y., Huang, C., and Wang, T. Hybrid-weighted total variation and nonlocal low-rank-based image compressed sensing reconstruction. *IEEE Access* 8 (2020), 23002–23010.