A Flexible Space-Variant Anisotropic Regularization for Image Restoration with Automated Parameter Selection*

Luca Calatroni†, Alessandro Lanza‡, Monica Pragliola‡, and Fiorella Sgallari‡

Abstract. We propose a new space-variant anisotropic regularization term for variational image restoration, based on the statistical assumption that the gradients of the target image distribute locally according to a bivariate generalized Gaussian distribution. The highly flexible variational structure of the corresponding regularizer encodes several free parameters which hold the potential for faithfully modeling the local geometry in the image and describing local orientation preferences. For an automatic estimation of such parameters, we design a robust maximum likelihood approach and report results on its reliability on synthetic data and natural images. For the numerical solution of the corresponding image restoration model, we use an iterative algorithm based on the alternating direction method of multipliers. A suitable preliminary variable splitting together with a novel result in multivariate nonconvex proximal calculus yield a very efficient minimization algorithm. Several numerical results showing significant quality improvement of the proposed model with respect to some related state-of-the-art competitors are reported, in particular, in terms of texture and detail preservation.

Key words. image reconstruction, multivariate generalized Gaussian distribution, space-variant regularization, anisotropic modeling, nonconvex variational modeling, ADMM

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

DOI. 10.1137/18M1227937

1. Introduction. Image restoration is the task of recovering a clean and sharp image from a noisy, and potentially blurred, observation. In mathematical terms, let Ω be a rectangular image domain of size $d_1 \times d_2$ and let $n := d_1 d_2$ be the total number of image pixels in Ω. For a given blurred and noisy image $g \in \mathbb{R}^n$, the typical image restoration inverse problem can be written as

$$(1.1) \qquad \text{find } u \in \mathbb{R}^n \qquad \text{such that} \qquad g = \mathcal{T}(Ku),$$

where $K \in \mathbb{R}^{n \times n}$ is a known linear blurring operator, while $\mathcal{T}(\cdot)$ denotes the operator modeling the presence of noise in g in a nondeterministic and very likely nonlinear way.

†CMAP, École Polytechnique, Palaiseau, 91128, Route de Saclay, France (luca.calatroni@polytechnique.edu).

‡Department of Mathematics, University of Bologna, Piazza di Porta San Donato 5, Bologna, Italy (alessandro.lanza2@unibo.it, monica.pragliola2@unibo.it, fiorella.sgallari@unibo.it).

Due to the ill-posedness of the problem (1.1), it is in general impossible to find u from (1.1) due to the lack of stability and/or uniqueness properties. Therefore, in practice, the task can be reformulated as the problem of finding an estimate u^* of the desired u as accurate as possible via a well-posed problem. In particular, variational regularization methods compute the restored image $u^* \in \mathbb{R}^n$ as a minimizer of a cost functional $\mathcal{J} : \mathbb{R}^n \to \mathbb{R}^+$ such that the problem can be formulated as

$$(1.2) \qquad \text{find } u^* \in \underset{u \in \mathbb{R}^n}{\arg\min} \ \left\{ \mathcal{J}(u) := R(u) + \mu F(Ku; g) \right\}.$$

The functionals R and F are commonly referred to as the *regularization* and the *data fidelity* terms, respectively. While R encodes prior information on the desired image u (such as its regularity and its sparsity patterns), F is a data term which measures the "distance" between the given image g and u after the action of the operator K with respect to some norm corresponding to the noise statistics in the data; cf., e.g., [47]. The regularization parameter $\mu > 0$ controls the trade-off between the two terms.

A very popular choice for R is the total variation (TV) seminorm [42, 11, 49], which is defined in the discrete setting as

$$(1.3) \qquad R(u) \ = \ \mathrm{TV}(u) := \sum_{i=1}^{n} \|(\nabla u)_i\|_2 \,,$$

where for each $i = 1, \ldots, n$, by $(\nabla u)_i \in \mathbb{R}^2$ we denote the discrete gradient of image u at pixel i. The choice of TV-type regularizations for image restoration problems became very popular in the last three decades due mainly to its convexity and, most importantly, its edge-preservation capability.

As mentioned above, the choice of F depends on the noise distribution in the data. In this paper, we are particularly interested in the case of additive (zero-mean) white Gaussian noise (AWGN), i.e., we consider the following form for the degradation model in (1.1):

$$g = \mathcal{T}(Ku) = Ku + b,$$

where the additive corruption $b \in \mathbb{R}^n$ stands for a vector of independent realizations drawn from the same univariate Gaussian distribution with zero mean and variance σ^2. Note that such noise distribution is indeed fully described by the unique scalar parameter $\sigma > 0$. Other similar noise models appearing in applications are the additive white Laplacian noise and the impulsive salt and pepper noise (SPN), which can also be fully described by a unique scalar parameter, it being either the standard deviation or the probability of a pixel being corrupted, respectively.

Finally, the regularization parameter μ in (1.2) plays a crucial role in the reconstruction results since its size balances the smoothing provided by the regularization and the trust in the data. Very often, μ is chosen empirically by brute-force optimization with respect to some fixed image quality measure (such as the signal-to-noise ratio (SNR) or the structural similarity index (SSIM)). However, for AWGN data, when the noise level σ is known, effective techniques based on discrepancy principles or the L-curve can be used [17]. More recently, similar approaches with "adaptive" discrepancy principles have been proposed for possibly

combined AWGN and SPN noise models in [28], while learning approaches based on the use of training sets and not requiring any prior knowledge of the noise level have been studied for optimal parameter selection in [9].

It is well known that a statistically consistent data fidelity term modeling the presence of AWGN in the data is the squared L_2 norm of the residual image, which, combined with the TV regularizer (1.3) results in the popular TV-L_2—or Rudin–Osher–Fatemi (ROF) [42]—image restoration model:

$$(1.4) \qquad \text{find } u^* \quad \text{such that} \quad u^* \in \underset{u \in \mathbb{R}^n}{\arg\min} \left\{ \text{TV}(u) \, + \, \frac{\mu}{2} \|Ku - g\|_2^2 \right\}.$$

Due to the presence of the TV regularizer, model (1.4) is nonsmooth, a fundamental feature which guarantees the desirable property of edge preservation. Furthermore, its convexity makes it appealing for several efficient optimization methods—see [12] for a review—and it is often used as a reference model for the study of either higher-order regularizations (e.g., the total generalized variation [8]) or of non-Gaussian [3, 35, 15, 44] and possibly combined [10, 33] noise distributions.

However, in addition to the well-known reconstruction drawbacks such as the *staircasing effect*, the TV regularizer in (1.3) suffers from additional limitations. First of all, it is *global* or *space invariant*, i.e., its local regularization contribution at each pixel takes exactly the same form and, as a result, it cannot adapt its functional shape to local image structures. Furthermore, it is not adapted to situations where clear local directional texture may appear, as it happens for instance in fiber and seismic imaging applications. For the mentioned problems the use of some dominant [26, 2] or local [55] anisotropy information can strongly improve the quality of the reconstruction.

The intrinsic limits of the TV regularizer have been discussed in great detail in [32] from a statistical point of view. There, the authors point out how the use of TV regularization implicitly corresponds to consider a space-invariant one-parameter half-Laplacian distribution for the gradient magnitudes of u, which is in general too restrictive to model the actual distribution of gradient magnitudes in real images. To overcome this issue, in [32] the authors propose the more general half-generalized Gaussian distribution as a prior which results in the following TV$_p$ regularization model

$$(1.5) \qquad \text{TV}_p(u) := \sum_{i=1}^{n} \|(\nabla u)_i\|_2^p, \quad p \in (0, 2].$$

The exponent p appearing in (1.5) is a free parameter which provides the TV$_p$ regularizer with higher flexibility than the TV regularizer. The parameter p, however, is fixed over the whole image domain and, hence, does not allow one to capture locality in the image.

In [30, 29] the authors consider a space-variant extension of the TV$_p$ regularizer in (1.5) which can better adapt to local image smoothness upon suitable parameter estimation. The new TV$_{\alpha,p}^{\text{sv}}$ regularizer is defined there by

$$(1.6) \qquad \text{TV}_{\alpha,p}^{\text{sv}}(u) := \sum_{i=1}^{n} \alpha_i \|(\nabla u)_i\|_2^{p_i}, \quad p_i \in (0, 2], \quad \alpha_i > 0 \ \ \forall \, i = 1, \dots, n,$$

and shown to be effective on several image restoration problems.

1.1. Contribution. In this paper we propose a *space-variant* and directional image regularizer denoted by $\mathrm{DTV}_p^{\mathrm{sv}}$ to extend even further the $\mathrm{TV}_{\alpha,p}^{\mathrm{sv}}$ regularization model (1.6) as

$$(1.7) \qquad \mathrm{DTV}_p^{\mathrm{sv}}(u) := \sum_{i=1}^{n} \|\Lambda_i R_{\theta_i} (\nabla u)_i\|_2^{p_i}, \qquad p_i > 0 \ \ \forall\, i = 1, \ldots, n.$$

For every $i = 1, 2, \ldots, n$, the weighting and rotation matrices $\Lambda_i, R_{\theta_i} \in \mathbb{R}^{2\times 2}$ are defined, respectively, by

$$(1.8) \quad \Lambda_i := \begin{pmatrix} \lambda_i^{(1)} & 0 \\ 0 & \lambda_i^{(2)} \end{pmatrix}, \quad \lambda_i^{(1)} \geq \lambda_i^{(2)} > 0, \qquad R_{\theta_i} := \begin{pmatrix} \cos\theta_i & -\sin\theta_i \\ \sin\theta_i & \cos\theta_i \end{pmatrix}, \quad \theta_i \in [0, 2\pi),$$

so that θ_i has to be understood as the local image orientation, while the parameters $\lambda_i^{(1)}$ and $\lambda_i^{(2)}$ at any point weight the TV-like smoothing along the direction θ_i and its orthogonal, respectively.

Under these definitions, we can then define our space-variant, anisotropic (or directional), and possibly nonconvex $\mathrm{DTV}_p^{\mathrm{sv}}$-$\mathrm{L}_2$ variational model for image restoration:

$$(1.9) \qquad \text{find} \ \ u^* \in \underset{u\in\mathbb{R}^n}{\arg\min} \left\{ \mathcal{J}(u) := \mathrm{DTV}_p^{\mathrm{sv}}(u) + \frac{\mu}{2} \|Ku - g\|_2^2 \right\}, \quad \mu > 0.$$

Note here that the nonconvexity arises whenever $0 < p_i < 1$ for at least one $i = 1, \ldots, n$.

The proposed $\mathrm{DTV}_p^{\mathrm{sv}}$ regularizer (1.7)–(1.8) is highly flexible as it potentially adapts to local smoothness and directional properties of the image at hand, provided that a reliable estimation of the parameters $\lambda_i^{(1)}, \lambda_i^{(2)}, \theta_i$, and p_i is given. In fact, in comparison to the previous work by the authors in [30, 29], our proposal extends the $\mathrm{TV}_{\alpha,p}^{\mathrm{sv}}$ regularization model (1.6) so as to accommodate further local directional information, which can significantly improve the restoration results in the case, for instance, of textured and/or high-detailed images.

The statistical rationale of our approach relies on a prior assumption on the distribution of the gradient magnitudes of the desired image u which we assume to be space variant and locally drawn from a bivariate generalized Gaussian distribution (BGGD) [6, 46, 45].

The main contribution of this work is twofold: on one side, we propose the highly flexible $\mathrm{DTV}_p^{\mathrm{sv}}$ regularizer in (1.7)–(1.8) and justify its variational form via MAP estimation. On the other hand, to guarantee its actual applicability on image restoration problems, we propose an automated efficient method for the robust estimation of the model parameters from the observed image g by means of a maximum likelihood (ML) estimation approach. The effectiveness of such estimation is confirmed numerically on several synthetic and natural examples, showing a good agreement with local geometrical structures in the images considered in terms of their "local" shape. From a numerical point of view, we solve the optimization problem (1.9) by means of an efficient iterative minimization algorithm based on the alternating direction method of multipliers (ADMM) [7] and apply it to several test images under different degradation levels, comparing the results with other relevant competing approaches. Finally, in order to get a fully automated image restoration approach, the regularization parameter μ in our model (1.9) is automatically adjusted along the ADMM iterations as described in [20],

such that the computed solution u^* satisfies the discrepancy principle [54], i.e., it belongs to the discrepancy set

$$(1.10) \qquad \mathcal{D} := \left\{ u \in \mathbb{R}^n : \|Ku - g\|_2 \leq \delta := \tau\sigma\sqrt{n} \right\}.$$

In (1.10), the discrepancy threshold value δ depends on the a priori known or estimated noise level σ, the number of pixels n, and the discrepancy parameter τ, which is typically chosen to be slightly greater than one, in order to avoid underestimation of the noise.

1.2. Organization of the paper. First, in section 2 we draw some analogies between the proposed $\mathrm{DTV}_p^{\mathrm{sv}}$ discrete regularizer (1.7)–(1.8) and some related previous studies on its infinite-dimensional correspondent. Then, in section 3 we show that the $\mathrm{DTV}_p^{\mathrm{sv}}$ regularization model can be derived via standard MAP estimation by assuming that the image gradients are drawn locally from a space-variant BGGD. In section 4 we describe in detail the ML procedure used for automatically estimating the local parameters appearing in the $\mathrm{DTV}_p^{\mathrm{sv}}$ regularizer from the observed corrupted image g. The existence of global minimizers for the total $\mathrm{DTV}_p^{\mathrm{sv}}$-$\mathrm{L}_2$ objective functional in (1.9) is then proved in section 5 via standard arguments. Next, in section 6 we describe in detail the ADMM algorithm used to compute such minimizers and present a novel useful result in multivariate nonconvex proximal calculus. As far as our numerical tests are concerned, we report in section 7 the results of the ML approach described above for a robust estimation of the parameter maps. In section 8 we report the results obtained by the $\mathrm{DTV}_p^{\mathrm{sv}}$-$\mathrm{L}_2$ image restoration model applied to some image deblurring/denoising problems observing its good performance in terms, mainly, of texture and detail preservation. Finally, we conclude our work with some outlook for future research directions in section 9.

2. Formulation in function spaces. The formulation of the $\mathrm{DTV}_p^{\mathrm{sv}}$ regularizer (1.7)–(1.8) in an infinite-dimensional function space defined over a regular $\Omega \subset \mathbb{R}^2$ reads as

$$(2.1) \qquad \mathcal{DTV}_{p(\cdot)}^{\mathrm{sv}}(u) := \int_\Omega |\mathbf{M}_{\boldsymbol{\lambda},\boldsymbol{\theta}}(x)\nabla u(x)|^{p(x)} \ dx,$$

where $p : \Omega \to (0, \infty)$ stands for the variable exponent and the tensor $\mathbf{M}_{\boldsymbol{\lambda},\boldsymbol{\theta}}$ is defined for any $x \in \Omega$ in terms of analogous weighting and rotation operators as in (3.6) by:

$$\mathbf{M}_{\boldsymbol{\lambda},\theta}(x) := \boldsymbol{\Lambda}_{\boldsymbol{\lambda}}(x)\mathbf{R}_\theta^T(x),$$

where $\boldsymbol{\lambda} = (\lambda_1, \lambda_2) \in L^\infty(\Omega; \mathbb{R}_+^2)$ and $\theta : \Omega \to [0, 2\pi)$. Note that whenever $\mathbf{M}_{\boldsymbol{\lambda},\theta} = \mathbf{I}$, there is no directionality encoded in the problem. In the following, we will refer to this special case as an *isotropic* model.

Several well-known image regularization models can be cast in a functional form similar to (2.1) or in their corresponding PDE counterparts.

2.1. Constant exponent $p \geq 1$. In the convex and constant case $p(x) = p \in [1, \infty)$ for every $x \in \Omega$, (2.1) can be thought of as an anisotropic image regularizer where images are chosen as elements in the Sobolev space $W^{1,p}(\Omega)$ or, more generally, modeled as Radon measures in some subspace of $\mathrm{BV}(\Omega)$, the space of functions of bounded variation [1].

In the special case $p = 2$ the functional (2.1) can be rewritten for $u \in H^1(\Omega)$ as

$$(2.2) \qquad \mathcal{DTV}_2^{\mathrm{sv}}(u) = \int_\Omega \|\nabla u(x)\|_{\mathbf{W}_{\boldsymbol{\lambda},\theta}}^2 \ dx,$$

where $\|e\|_{\mathbf{W}_{\boldsymbol{\lambda},\theta}} := \sqrt{\langle e, \mathbf{W}_{\boldsymbol{\lambda},\theta} e \rangle}$ is a scalar product and $\mathbf{W}_{\boldsymbol{\lambda},\theta} := \mathbf{M}_{\boldsymbol{\lambda},\theta}^T \mathbf{M}_{\boldsymbol{\lambda},\theta}$ is a symmetric and positive semidefinite anisotropic tensor. By taking the L^2-gradient flow of the energy in (2.2), we can easily draw connections between this choice and the standard anisotropic diffusion PDE models proposed by Weickert in [51] and Weickert and Scharr in [52]. Indeed, by endowing Ω with Neumann boundary conditions we get that minimizing $\mathcal{DTV}_2^{\mathrm{sv}}$ corresponds to computing the stationary solution of

$$(2.3) \qquad \begin{cases} u_t = \mathrm{div}\left(\mathbf{W}_{\boldsymbol{\lambda},\theta} \nabla u\right) & \text{on } \Omega \times (0, \infty], \\ u(x, 0) = f(x) & \text{on } \Omega, \\ \langle \mathbf{W}_{\boldsymbol{\lambda},\theta} \nabla u, \boldsymbol{n} \rangle = 0 & \text{on } \partial\Omega \times (0, T], \end{cases}$$

where $\boldsymbol{n}$ stands for the outward normal vector on $\partial\Omega$. The Cauchy problem (2.3) is a reference model for anisotropic PDE approaches for image restoration. The tensor $\mathbf{W}_{\boldsymbol{\lambda},\theta}$ stands for a space-dependent diffusivity matrix which can introduce nonlinearities in the model [53] or classically related to a structure-tensor modeling as in [51, 41, 43].

Recently, a similar formalism has also been employed in [2, 55, 27, 26] in the case $p(x) = p = 1$ for "dominant" fixed principal direction $\theta(x) = \bar{\theta} \in [0, 2\pi)$ and adapted in [48, 16] to local directionalities in the context of medical imaging. In such a case the functional in (2.1) reads

$$\mathcal{DTV}_1^{\mathrm{sv}}(u) := \int_\Omega |\mathbf{M}_{\boldsymbol{\lambda},\theta}(x) \nabla u(x)| \ dx,$$

which can be seen as a directional version of TV regularization, and thereafter called DTV regularization. In [27] a higher-order directional total generalized variation regularizer is also studied to promote smoother reconstructions and a full analysis in function spaces is performed. The choice of considering only the main direction $\bar{\theta}$ in the image restricts the authors to study very simple images with regular stripe patterns. For this purpose, an algorithm estimating such direction and some experiments on its robustness/sensitivity to noise are presented.

2.2. Variable exponent $p(x) \geq 1$. In the isotropic case, variable exponent models have been considered, e.g., in [4] under the modeling assumption $p(x) = p(|\nabla u(x)|)$ with

$$(2.4) \qquad \lim_{s \to 0} p(s) = 2, \qquad \lim_{s \to +\infty} p(s) = 1.$$

Heuristically, such conditions correspond to considering a quadratic smoothing in correspondence with flat areas (small gradients) and a TV-type in correspondence with edges (large gradients). To overcome the difficulties arising from the theoretical analysis of such general modeling, in [13, 34] some easier variable exponent models have been considered. There, the

image regularizer takes the following form:

$$(2.5) \qquad \mathcal{R}_{p(\cdot)}(u) := \int_\Omega \frac{1}{p(x)} |\nabla u(x)|^{p(x)} \, dx,$$

where for every $x \in \Omega$ the exponent function $p : \Omega \to [1,2]$ is defined via the following explicit formula,

$$(2.6) \qquad p(x) = 1 + \frac{1}{1 + k|G_\varsigma * \nabla g(x)|}, \quad \varsigma, \ k > 0,$$

where G_ς is a convolution kernel of parameter ς and g is the given corrupted image. Under such a choice, the conditions (2.4) are satisfied and all the possible intermediate values are allowed. In [13, 34] the regularizer (2.5)–(2.6) is combined with L^2-fidelity and shown to reduce staircasing compared to constant exponent models.

2.3. Nonconvex models with constant exponents $0 < p < 1$. More recently, some nonconvex image regularization models in the form (2.1) with constant exponent $0 < p < 1$ have been considered. In [22, 21], for instance, nonconvex TV$_p$-type regularizers have been shown to indeed be preferable for some applications in comparison to convex ($p \geq 1$) models as the ones described above. The analysis in function spaces covered by the authors is motivated by the use of discrete models such as the ones proposed previously in [36, 39]. For this type of regularization and upon an appropriate Huber-type smoothing, efficient trust-region-based optimization solvers are designed.

Generally speaking, nonconvex regularizers in the form (2.1) with constant exponent $p < 1$ are nowadays well known to promote stronger sparsity in the data, improving significantly the reconstruction obtained in terms of structure preservation and edge sharpness. On the other hand, such methods may result in an overcomplication of the problem in correspondence with homogeneous image regions, where a plain isotropic smoothing may still be preferable. For this reason, a space-variant image regularizer adapting its convexity to the local geometrical structures sounds desirable and appealing for imaging applications.

We stress that the fine analysis of (2.1) in a functional setting becomes very challenging in the case $p < 1$, since the extension of such regularizing functionals to spaces similar to $BV(\Omega)$ is not trivial at all. For some theoretical considerations in this direction we refer the reader to [21]. To simplify the difficulties arising in such framework, our model is studied in a purely discrete setting. Its extension and analysis in a functional framework is left for future research.

3. Statistical derivation via MAP estimation. A common statistical paradigm in image restoration is the MAP approach by which the restored image is obtained as a global minimizer of the negative log-likelihood distribution given the observed image g and the known blurring operator K combined with some prior probability on the unknown target image u; see, e.g., [25, 57]. In formulas,

$$(3.1) \qquad u^* \in \arg\max_{u \in \mathbb{R}^n} P(u|g; K) = \arg\min_{u \in \mathbb{R}^n} \left\{ -\log P(g|u; K) - \log P(u) \right\}.$$

The equality above comes from the application of the Bayes' formula after dropping the normalization term $P(g)$.

In the case of AWGN the likelihood term in (3.1) takes the following special form

$$(3.2) \qquad P(g|u; K) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(Ku-g)_i^2}{2\sigma^2}\right) = \frac{1}{W} \exp\left(-\frac{\|Ku-g\|_2^2}{2\sigma^2}\right),$$

where $\sigma > 0$ denotes the AWGN standard deviation and $W > 0$ is a normalization constant.

As far as the unknown image u is concerned, a standard choice consists in its modeling via a Markov random field (MRF) such that its prior $P(u)$ takes the form of a Gibbs prior, whose general form reads

$$(3.3) \qquad P(u) = \frac{1}{Z} \prod_{i=1}^{n} \exp\left(-\alpha V_{\mathcal{N}_i}(u)\right) = \frac{1}{Z} \exp\left(-\alpha \sum_{i=1}^{n} V_{\mathcal{N}_i}(u)\right),$$

where $\alpha > 0$ is the MRF parameter and, for every $i = 1, \ldots, n$, $\mathcal{N}_i$ denotes the set of all neighboring pixels of i (also known as a "clique"), $V_{\mathcal{N}_i}$ stands for the potential function on $\mathcal{N}_i$, and Z is the normalizing partition function not depending on u. Such MRF modeling has been widely explored in the context of Bayesian models for imaging and combined in [40] with learning strategies to design data-driven filters over extended neighborhoods.

For more model-oriented approaches, by setting for any $i = 1, \ldots, n$, $V_{\mathcal{N}_i}(u) := \|(\nabla u)_i\|_2$, the Gibbs prior in (3.3) reduces to the TV prior: $P(u) = \frac{1}{Z} \exp\left(-\alpha \sum_{i=1}^{n} \|(\nabla u)_i\|_2\right)$ or, equivalently, interpreted by saying that each $\|(\nabla u)_i\|_2$ is distributed according to a half-Laplacian distribution with parameter $\alpha > 0$. Via similar considerations, in [32] the authors have shown how such a one-parameter model is in fact too restrictive to describe the statistical distribution of the gradient in real images.

For this reason, we proceed differently and model the joint distribution of the two partial derivatives of the gradient vector $(\nabla u)_i$ at any pixel by a BGGD [6]. Namely, for all $i = 1, \ldots, n$, we assume that

$$(3.4) \qquad P((\nabla u)_i; p_i, \Sigma_i) = \frac{1}{2\pi |\Sigma_i|^{1/2}} \frac{p_i}{\Gamma(2/p_i)\, 2^{2/p_i}} \exp\left(-\frac{1}{2}((\nabla u)_i^T \Sigma_i^{-1} (\nabla u)_i)^{p_i/2}\right),$$

where Γ stands for the Gamma function, the *covariance matrices* $\Sigma_i \in \mathbb{R}^{2\times 2}$ are symmetric positive definite with determinant $|\Sigma_i|$ and $p_i/2$ is often referred to as the *shape parameter*. Note that in (3.4) when $p_i = 2$ for every $i = 1, \ldots, n$, then the BGGD reduces to a standard bivariate Gaussian distribution with pixelwise covariance matrices Σ_i.

Proceeding similarly as above, we can then deduce the expression of the corresponding prior under such assumption. It reads

$$(3.5) \qquad P(u) = \frac{1}{Z} \exp\left(-\frac{1}{2} \sum_{i=1}^{n} \left((\nabla u)_i^T \Sigma_i^{-1} (\nabla u)_i\right)^{p_i/2}\right).$$

The symmetric positive definite matrices Σ_i contain information on both the directionality and the scale of the BGGD at pixel i. To see that explicitly, we consider their eigenvalue decomposition:

$$(3.6) \qquad \Sigma_i = V_i^T E_i V_i, \quad E_i = \begin{pmatrix} e^{(1)}{}_i & 0 \\ 0 & e_i^{(2)} \end{pmatrix}, \quad e^{(1)}{}_i \geq e_i^{(2)} > 0, \quad V_i^T V_i = V_i V_i^T = I,$$

where for every $i = 1, \ldots, n$, $e^{(1)}{}_i$, $e_i^{(2)}$ are the (positive) eigenvalues of Σ_i, V_i is the orthonormal (rotation) modal matrix, and I denotes the 2×2 identity matrix. We then rewrite the terms in the sum appearing in (3.5) as

$$\left((\nabla u)_i^T \Sigma_i^{-1} (\nabla u)_i \right)^{\frac{p_i}{2}} = \left((\nabla u)_i^T V_i^T E_i^{-1} V_i (\nabla u)_i \right)^{\frac{p_i}{2}} = \left\| E_i^{-1/2} V_i (\nabla u)_i \right\|_2^{p_i},$$

whence by setting

$$(3.7) \qquad\qquad\qquad \Lambda_i := E_i^{-1/2}, \quad R_{\theta_i} := V_i,$$

and after recalling the definition of the $\mathrm{DTV}_p^{\mathrm{sv}}$ regularizer given in (1.7)–(1.8) we observe that the prior in (3.5) can indeed be expressed as

$$(3.8) \qquad\qquad\qquad P(u) = \frac{1}{Z} \exp\left(-\frac{1}{2} \mathrm{DTV}_p^{\mathrm{sv}}(u) \right).$$

By plugging the expression of the Gaussian likelihood (3.2) and the BGGD prior (3.8) in the MAP inference formula (3.1) and after dropping the constant terms, we finally obtain the $\mathrm{DTV}_p^{\mathrm{sv}}$-$\mathrm{L}_2$ image restoration model (1.9) for blur and AWGN removal by setting $\mu = 2/\sigma^2$.

4. Automatic estimation of the $\mathrm{DTV}_p^{\mathrm{sv}}$ parameters. The very high flexibility of the proposed space-variant anisotropic $\mathrm{DTV}_p^{\mathrm{sv}}$ regularizer (1.7)–(1.8) would be useless without an effective procedure for automatically and reliably estimating all its parameters from the observed corrupted data g.

In this section we propose a statistical optimization strategy for the estimation of the covariance matrices Σ_i and the parameters p_i of the BGGD defined in (3.4) when a collection of samples is available. Similar strategies estimating model parameters for directional regularizers from statistical priors have been proposed, e.g., in [38] for anisotropic PDEs in the form (2.3). For simplicity, in the following we will drop the dependence on i of the quantities appearing in (3.4) and denote by $x := \nabla u \in \mathbb{R}^2$ the local gradient of the image u. First, we observe that the requirement for Σ to be symmetric positive definite means

$$\Sigma = \begin{bmatrix} \sigma_1 & \sigma_3 \\ \sigma_3 & \sigma_2 \end{bmatrix} \qquad \text{with} \qquad \begin{cases} \sigma_1 > 0, \\ |\Sigma| = \sigma_1 \sigma_2 - \sigma_3^2 > 0. \end{cases}$$

As suggested in [46, 45, 37], it is possible to decouple the spread and the directionality of the BGGD by introducing a further *scale parameter* $m > 0$, so that (3.4) takes the following form:

$$P(x; p, \Sigma, m) = \frac{1}{\pi \Gamma\left(\frac{2}{p}\right) 2^{\frac{2}{p}}} \frac{p}{2m|\Sigma|^{\frac{1}{2}}} \exp\left(-\frac{1}{2m^{p/2}} (x^T \Sigma^{-1} x)^{p/2} \right).$$

By imposing that the trace of the covariance matrix Σ is fixed and equal to the dimension of the ambient space, i.e., $\mathrm{tr}(\Sigma) = d = 2$, we easily get the following expression of the constraint

set $\mathcal{C}$ for the parameters $p, m, \sigma_1, \sigma_2, \sigma_3$ to be well-defined:

$$
(4.1) \qquad \mathcal{C} := \begin{cases} p > 0, \\ m > 0, \\ \sigma_1 + \sigma_2 = 2, \\ \sigma_1 \sigma_2 - \sigma_3^2 > 0, \end{cases} \qquad \longrightarrow \qquad \mathcal{C} = \begin{cases} p > 0, \\ m > 0, \\ \sigma_1^2 + \sigma_3^2 - 2\sigma_1 < 0. \end{cases}
$$

The set $\mathcal{C}$ is an open (unbounded) semicylinder in $\mathbb{R}^4$. After a change of coordinates which shifts the center of the circle in the $\sigma_1 - \sigma_3$ plane to the origin, we obtain the following expression of Σ^{-1},

$$
(4.2) \qquad \tilde{\sigma}_1 := 1 - \sigma_1 \qquad \longrightarrow \qquad \Sigma^{-1} = \frac{1}{1 - \tilde{\sigma}_1^2 - \sigma_3^2} \begin{bmatrix} 1 + \tilde{\sigma}_1 & -\sigma_3 \\ -\sigma_3 & 1 - \tilde{\sigma}_1 \end{bmatrix}.
$$

To avoid heavy notation, in the following we will still denote by σ_1 the same variable after this change of coordinates.

4.1. ML estimation of the BGGD parameters.

For any point Ω let $\mathcal{N}(x) = \{x_1, \dots, x_N\} \subset \mathbb{R}^{2 \times N}$ denote the neighborhood centered in x of N independent and identically distributed samples drawn from a BGGD with unknown parameters $p, \sigma_1, \sigma_3, m \in \mathcal{C}$. Then, the corresponding *likelihood* function reads

$$
(4.3) \qquad \mathcal{L}(p, \Sigma, m; x) = \prod_{x_j \in \mathcal{N}(x)} P(x_j; p, \Sigma, m) = \prod_{j=1}^{N} P(x_j; p, \Sigma, m)
$$

$$
= \left[\frac{1}{|\Sigma|^{1/2}} \frac{p}{2\pi \Gamma\left(\frac{2}{p}\right) 2^{2/p} m} \right]^N \exp\left(-\frac{1}{2m^{p/2}} \sum_{j=1}^{N} (x_j^T \Sigma^{-1} x_j)^{p/2} \right).
$$

We now look for $(p^*, \Sigma^*, m^*) \in \mathcal{C}$ maximizing $\mathcal{L}$. Equivalently, by taking the negative logarithm, we aim to solve

$$
(4.4) \qquad (p^*, \Sigma^*, m^*) \longleftarrow \underset{p, \Sigma, m \,\in\, \mathcal{C}}{\arg\min} \; \{\mathcal{F}(p, \Sigma, m; x) := -\log \, \mathcal{L}(p, \Sigma, m; x)\},
$$

whence, by recalling the fundamental property $\Gamma(z+1) = z\Gamma(z)$ for every $z \in \mathbb{R}$, we deduce

$$
\mathcal{F}(p, \Sigma, m; x) = -\left[N \log\left(\frac{1}{|\Sigma|^{1/2}} \frac{1}{\pi \Gamma\left(\frac{2}{p} + 1\right) 2^{2/p} m} \right) - \frac{1}{2m^{p/2}} \sum_{j=1}^{N} (x_j^T \Sigma^{-1} x_j)^{p/2} \right]
$$

$$
= N \log\left(|\Sigma|^{1/2} \pi \Gamma\left(\frac{2}{p} + 1\right) 2^{2/p} \right) + N \log m + \frac{1}{2m^{p/2}} \sum_{j=1}^{N} (x_j^T \Sigma^{-1} x_j)^{p/2}.
$$

Note that $\mathcal{F}$ is differentiable on $\mathcal{C}$. Therefore, by simply imposing the first-order optimality condition for m, we can find a closed formula for m^* as follows:

$$
(4.5) \qquad \frac{\partial \mathcal{F}}{\partial m} = \frac{N}{m} - \frac{p}{4m^{\frac{p}{2}+1}} \sum_{j=1}^{N} (x_j^T \Sigma^{-1} x_j)^{p/2} \qquad \longrightarrow \qquad m^* = \left(\frac{p}{4N} \sum_{j=1}^{N} (x_j^T \Sigma^{-1} x_j)^{p/2} \right)^{\frac{2}{p}}.
$$

We now substitute this formula into the expression of $\mathcal{F}$, thus getting

$$(4.6) \quad \mathcal{F}(p, \Sigma; x) = N \log \left(|\Sigma|^{1/2} \pi \Gamma \left(\frac{2}{p} + 1 \right) 2^{2/p} \right) + \frac{2N}{p} \log \left(\frac{p}{4N} \sum_{j=1}^{N} (x_j^T \Sigma^{-1} x_j)^{p/2} \right) + \frac{2N}{p}.$$

By making explicit the dependence of $\mathcal{F}$ on the entries of Σ, we have that (4.6) turns into

$$(4.7) \qquad \mathcal{F}(p, \sigma_1, \sigma_2, \sigma_3; x) = N \log \left(\frac{1}{|\Sigma|^{1/2}} \pi \Gamma \left(\frac{2}{p} + 1 \right) 2^{2/p} \right) + \frac{2N}{p} + \frac{2N}{p} \log \frac{p}{4N}$$
$$+ \frac{2N}{p} \log \left(\sum_{j=1}^{N} (\sigma_2 x_{j,1}^2 + \sigma_1 x_{j,2}^2 - 2\sigma_3 x_{j,1} x_{j,2})^{p/2} \right).$$

We now study the behavior of $\mathcal{F}$ expressed as above as $(p, \sigma_1, \sigma_2, \sigma_3)$ approach the boundary of the set $\mathcal{C}$ defined in (4.1). Thanks to the formula for m^* derived in (4.5), we start noticing that $\mathcal{C}$ can in fact be expressed as a subset in $\mathbb{R}^3$ defined by the variables p, σ_1, and σ_3 only. By further switching to polar coordinates in the $\sigma_1 - \sigma_3$ plane, we get

$$(\sigma_1, \sigma_3) = \varrho(\cos \phi, \sin \phi), \quad 0 \leq \varrho < 1, \quad \phi \in [0, 2\pi),$$

so that the matrices in (4.2) take the following form:

$$(4.8) \qquad \Sigma = \begin{bmatrix} 1 - \varrho \cos \phi & \varrho \sin \phi \\ \varrho \sin \phi & 1 + \varrho \cos \phi \end{bmatrix}, \qquad \Sigma^{-1} = \frac{1}{1 - \varrho^2} \begin{bmatrix} 1 + \varrho \cos \phi & -\varrho \sin \phi \\ -\varrho \sin \phi & 1 - \varrho \cos \phi \end{bmatrix},$$

and the functional $\mathcal{F}$ in (4.7) becomes

$$\mathcal{F}(p, \phi, \varrho; x) = N \log \left(\Gamma \left(\frac{2}{p} + 1 \right) \frac{\pi}{\sqrt{1 - \varrho^2}} \left(\frac{p}{2N} \right)^{2/p} \right) + \frac{2N}{p} + \frac{2N}{p} \log \frac{p}{4N}$$
$$(4.9) \qquad + \frac{2N}{p} \log \left[\sum_{j=1}^{N} ((1 + \varrho \cos \phi) x_{j,1}^2 + (1 - \varrho \cos \phi) x_{j,2}^2 - 2\varrho \sin \phi \ x_{j,1} x_{j,2})^{p/2} \right].$$

As a conclusion, we can finally rewrite the ML problem (4.4) as the following constrained optimization problem,

$$(4.10) \qquad\qquad (p^*, \phi^*, \varrho^*) \in \underset{\substack{p \in (0, \infty), \\ \phi \in [0, 2\pi), \\ \varrho \in [0, 1)}}{\arg \min} \ \mathcal{F}(p, \phi, \varrho).$$

Note, that since the problem (4.10) is formulated over a noncompact set of $\mathbb{R}^3$, the existence of its solution is in general not guaranteed.

4.2. Reformulation on a compact set. One possible way to overcome this problem consists in characterizing explicitly the configurations of the samples $x_1, \ldots, x_N \in \mathcal{N}(x)$ for which the functional $\mathcal{F}$ in (4.9) does not attain its minimum inside $\mathcal{C}$. To do so, let us first rename the last term in (4.9) as

$$A(\phi, \varrho) := \frac{2N}{p} \log \left[\sum_{i=j}^{N} ((1 + \varrho \cos \phi) x_{j,1}^2 + (1 - \varrho \cos \phi) x_{j,2}^2 - 2\varrho \sin \phi \ x_{j,1} x_{j,2})^{p/2} \right].$$

For any $p \in (0, +\infty)$, if $A(\phi, \varrho)$ is bounded as $\varrho \to 1^-$, then the functional $\mathcal{F}$ in (4.9) tends to $+\infty$ and the minimum is necessarily attained in the interior of $\mathcal{C}$. However, if $A(\phi, \varrho)$ is unbounded as $\varrho \to 1^-$, nothing can be said about the behavior of $\mathcal{F}$ 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 $x_1, \ldots, x_N \in \mathcal{N}(x)$ for which $\mathcal{F}$ tends to $-\infty$ at the boundary. In order to characterize such configurations, note that as $\varrho \to 1^-$ we have that by continuity

$$A(\phi, \varrho) \to \frac{2N}{p} \log \left[\sum_{j=1}^{N} (\sqrt{1 + \cos \phi} \ x_{j,1} - \sqrt{1 - \cos \phi} \ x_{j,2})^p \right],$$

which tends to $-\infty$ if and only if the argument of the logarithm tends to zero, i.e., when

(4.11)
$$x_{j,2} = \sqrt{\frac{\sqrt{1 + \cos \phi}}{\sqrt{1 - \cos \phi}}} \ x_{j,1} \qquad \forall j = 1, \ldots, N.$$

This situation corresponds to the case when the samples x_j all lie on the line passing through the origin with slope $\sqrt{\frac{\sqrt{1+\cos \phi}}{\sqrt{1-\cos \phi}}}$.

A possible way to guarantee the existence of solutions of the problem (4.10) is to reformulate the problem over a compact subset of $\mathbb{R}^3$. Although this may sound a little bit artificial, note that for imaging applications such an assumption makes perfect sense for different reasons. First, as far as the range for the parameter ϱ is concerned, note that the degenerate configurations (4.11) happening as ϱ approaches 1^- are easily detectable in a preprocessing step and, in practice, very unlikely for natural images since they would correspond to situations where gradient components are linearly correlated for any sample $j = 1, \ldots, N$. Therefore, provided we can perform such a preliminary check, the case $\varrho = 1$ becomes admissible since no other possible configurations are allowed under this choice.

Regarding the admissible values for p, we notice that the more we enforce sparsity (i.e., the closer p is to zero), the more the BGGD will tend to a Dirac delta distribution, making the estimation of local anisotropy in a neighborhood of the point considered very hard (see section 4 for more details). Additionally, as was commonly done in previous work for variable exponent models for imaging, an upper bound for such values—typically chosen as $\bar{p} \geq 2$—can be fixed. Therefore, in practice, we can fix lower and upper bounds $0 < \varepsilon < \bar{p}$ for the exponent range.

After these observation, we can then reformulate the problem (4.10) as follows:

(4.12) $$(p^*, \phi^*, \varrho^*) \longleftarrow \underset{p, \theta, \varrho}{\arg \min} \, \mathcal{F}(p, \phi, \varrho; x)$$

$$\text{s.t. } p \in [\varepsilon, \bar{p}], \ \ 0 \leq \varrho \leq 1, \ \ 0 \leq \phi \leq 2\pi,$$

where now the constraint set is compact, which, combined with the continuity of $\mathcal{F}$, guarantees that the minimization problem admits a minimum.

Before carrying on with our discussion, we recall once again that the ML procedure described above is *local*, i.e., it has to be repeated for any pixel in the image domain, thus resulting in the estimation of the parameter map $(p_i^*, \phi_i^*, \varrho_i^*)$ for $i = 1, \ldots, n$.

For each pixel $i = 1, \ldots, n$, the triple of estimated parameters is involved in the computation of the matrices Λ_i, R_{θ_i} defining the regularizer in (1.7). Relying on (4.8), the eigenvalues $e_i^{(1)}, e_i^{(2)}$ can be easily computed. Observe that, due to the normalization condition on the trace introduced in (4.1), the minimum eigenvalue $e_i^{(2)}$ can be directly derived by the maximum eigenvalue $e_i^{(1)}$:

$$e_i^{(1)} = 1 + \varrho_i, \quad e_i^{(2)} = 2 - e_i^{(1)} = 1 - \varrho_i.$$

Therefore, recalling (3.7), the matrix Λ_i is obtained as follows:

(4.13) $$\Lambda_i := \begin{pmatrix} \lambda_i^{(1)} & 0 \\ 0 & \lambda_i^{(2)} \end{pmatrix} = \begin{pmatrix} \frac{1}{\sqrt{e_i^{(1)}}} & 0 \\ 0 & \frac{1}{\sqrt{e_i^{(2)}}} \end{pmatrix}.$$

Once $e_i^{(1)}$ is available, its corresponding eigenvector $(v_1)_i$, satisfying $\Sigma_i (v_1)_i = e^{(1)}{}_i (v_1)_i$, can be further calculated using the formula

$$(v_1)_i = \sqrt{\frac{1 + \cos \phi_i}{2}} \begin{bmatrix} \frac{\sin \phi_i}{\sqrt{1 + \cos \phi_i}} \\ 1 \end{bmatrix}.$$

As a consequence, the local angle θ_i describing the local orientation is computed by

$$\theta_i = \arctan \frac{\sqrt{1 + \cos \phi_i}}{\sin \phi_i},$$

and the rotation matrix R_{θ_i} is given as in (1.8).

Furthermore, it is helpful to represent the estimated BGGD to visualize its shape in the plane $((D_h u)_i, (D_v u)_i)$. In order to draw the corresponding level curves, we only need the maximum eigenvalue $e_i^{(1)}$ and the rotation angle θ_i. Such curves are the ellipses having semi-axes a_i, b_i, and eccentricity ϵ_i given by

$$a_i := \sqrt{e^{(1)}{}_i}, \quad b_i := \sqrt{e_i^{(2)}}, \quad \epsilon_i := \frac{\sqrt{a_i^2 - b_i^2}}{a_i} = \frac{\sqrt{e^{(1)}{}_i - e_i^{(2)}}}{\sqrt{e^{(1)}{}_i}}.$$

An illustrative drawing of the anisotropy ellipses described above is reported in Figure 1.
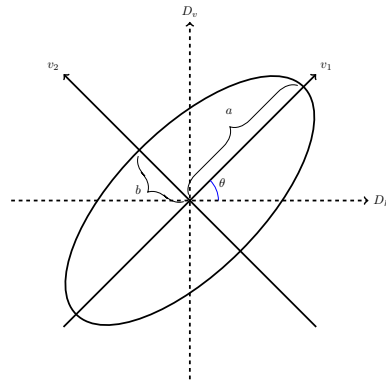
Figure 1. *Representation of anisotropy ellipses describing BGGD level lines in the plane $D_h - D_v$ in terms of the eigenvalues and eigenvectors of the estimated matrix Σ.*

5. Existence of solutions. In this section, we provide an existence result for the solutions of the proposed $\mathrm{DTV}_p^{\mathrm{sv}}$-$\mathrm{L}_2$ variational model (1.7)–(1.9). In general, the $\mathrm{DTV}_p^{\mathrm{sv}}$-$\mathrm{L}_2$ functional is not convex, therefore, it is not guaranteed to admit a unique global minimizer. However, by applying a general lemma whose proof can be found in [14, Lemma 2.7.1] we will prove that existence of global minimizers is guaranteed. In the following, we will use the notations $\mathrm{null}(M)$, $\mathrm{span}(v_1, \ldots, v_m)$, I_m, 0_m, and 1_m to denote the null space of the linear operator M, the linear span of the set of vectors $v_1, \ldots, v_m$, the identity matrix of order m, and the all-zeros and all-ones m-dimensional vectors, respectively. We have that the following lemma holds true [14].

Lemma 5.1. *Let $A_1 \in \mathbb{R}^{m \times n}$, $A_2 \in \mathbb{R}^{q \times n}$ be two linear operators satisfying*

$$\mathrm{null}(A_1) \cap \mathrm{null}(A_2) = \{0_n\},$$

and let $f_1 : \mathbb{R}^m \to [-\infty, +\infty]$ and $f_2 : \mathbb{R}^q \to [-\infty, +\infty]$ be two proper, lower semicontinuous, and coercive functions. Then, the function $h : \mathbb{R}^n \to [-\infty, +\infty]$ defined by

$$h(x) := f_1(A_1 x) + f_2(A_2 x)$$

is lower semicontinuous and coercive.

We now apply this result to the $\mathrm{DTV}_p^{\mathrm{sv}}$-$\mathrm{L}_2$ model.

Proposition 5.2. *The $\mathrm{DTV}_p^{\mathrm{sv}}$-$\mathrm{L}_2$ functional $\mathcal{J} : \mathbb{R}^n \to \mathbb{R}$ defined in (1.7)–(1.9) is continuous, bounded from below by zero, and coercive, hence, it admits global minimizers.*

Proof. Let $A_1 \in \mathbb{R}^{2n \times n}$ be the matrix defined by

$$(5.1) \qquad A_1 = LD, \quad L = \mathrm{diag}\,(L_1, L_2, \ldots, L_n), \quad L_i = \Lambda_i R_{\theta_i} \in \mathbb{R}^2, \;\; i = 1, \ldots, n,$$

with $\Lambda_i, R_{\theta_i} \in \mathbb{R}^2$ the full rank matrices in (1.8) and $D \in \mathbb{R}^{2n \times n}$ a finite difference operator discretizing the image gradient, let $A_2 = K$, and let $f_1 : \mathbb{R}^{2n} \to \mathbb{R}$, $f_2 : \mathbb{R}^n \to \mathbb{R}$ be the

functions defined by

$$(5.2) \qquad \begin{aligned} f_1(y) &:= \sum_{i=1}^{n} \|(y_{2i-1}, y_{2i})\|_2^{p_i}, & y \in \mathbb{R}^{2n}, \\ f_2(z) &:= \frac{\mu}{2} \|z - g\|_2^2, & z \in \mathbb{R}^n. \end{aligned}$$

Then, the DTV_p^{sv}-L_2 energy functional in (1.7)–(1.9) can be written as

$$(5.3) \qquad \mathcal{J}(u) = f_1(A_1 u) + f_2(A_2 u).$$

As the block diagonal matrix L in (5.1) has full rank (all matrices L_i have full rank), the linear operator A_1 has the same null space as the discrete gradient operator D. It follows that

$$(5.4) \qquad \big(\text{null}(A_1) = \text{null}(D) = \text{span}(1_n)\big) \cap \big(\text{null}(A_2) = \text{null}(K)\big) = \{0_n\},$$

in fact constant images do not belong to the null space of the linear blur operator K. Furthermore, functions f_1 and f_2 in (5.2) are clearly continuous, bounded from below by zero, and coercive. It thus follows from Lemma 5.1 that the DTV_p^{sv}-L_2 functional $\mathcal{J}$ in (5.3) is continuous, bounded from below by zero, and coercive, hence, it admits at least one global minimizer. ∎

Uniqueness of solutions is in general not guaranteed. However, if the functional is strictly convex, this trivially holds.

Corollary 5.3. *Let $\mathcal{J} : \mathbb{R}^n \to \mathbb{R}$ be the DTV_p^{sv}-L_2 functional defined in (1.7)–(1.9). If $p_i > 1$ for every $i = 1, \dots, n$, then $\mathcal{J}$ is strongly convex. Hence it admits a unique global minimizer.*

Note, however, that as we discussed in the introduction, in this work we are more interested in the nonconvex case, e.g., when there exists at least one $i \in \{1, \dots, n\}$ such that $p_i < 1$, since, in this, better regularization properties are enforced in DTV_p^{sv}-L_2. Therefore, in our applications uniqueness will in general not be guaranteed and we will generally be dealing with the case of local minima.

6. Numerical solution by ADMM. We can now describe the ADMM-based iterative algorithm [7] used to numerically solve the proposed DTV_p^{sv}-L_2 model (1.7)–(1.9) once the values of all the parameters $p_i, \theta_i, \lambda_i^{(i)}, \lambda_i^{(2)}$, $i = 1, \dots, n$, which define the regularizer, have been set according to the procedure illustrated in section 4. To this purpose, we first introduce two auxiliary variables $r \in \mathbb{R}^n$ and $t \in \mathbb{R}^{2n}$ and rewrite model (1.7)–(1.9) in the following equivalent constrained form:

$$(6.1) \qquad \{u^*, r^*, t^*\} \leftarrow \underset{u,r,t}{\arg\min} \left\{ \sum_{i=1}^{n} \|\Lambda_i R_{\theta_i} t_i\|_2^{p_i} + \frac{\mu}{2} \|r\|_2^2 \right\}$$

$$(6.2) \qquad \text{subject to}: \ r = Ku - g, \ t = Du,$$

where $D := (D_h^T, D_v^T)^T \in \mathbb{R}^{2n \times n}$ denotes the discrete gradient operator with D_h, $D_v \in \mathbb{R}^{n \times n}$ two linear operators representing finite difference discretizations of the first-order partial derivatives of the image u in the horizontal and vertical direction, respectively, and where $t_i := \big((D_h u)_i, (D_v u)_i\big)^T \in \mathbb{R}^2$ stands for the discrete gradient of u at pixel i. We notice

that the auxiliary variable t is introduced to transfer the discrete gradient operator out of the possibly nonconvex nonsmooth regularizer whereas the variable r is aimed to adjust the regularization parameter μ along the ADMM iterations such that the computed solution u^* satisfies the discrepancy principle [54], i.e., belongs to the discrepancy set $\mathcal{D}$ in (1.10).

In order to solve problem (6.1)–(6.2) via ADMM, we start defining the augmented Lagrangian functional as follows:

$$
\mathcal{L}(u, r, t; \rho_r, \rho_t) := \sum_{i=1}^{n} \|\Lambda_i R_{\theta_i} t_i\|_2^{p_i} + \frac{\mu}{2} \|r\|_2^2 - \langle \rho_t, t - Du \rangle + \frac{\beta_t}{2} \|t - Du\|_2^2
$$

$$
(6.3) \qquad\qquad - \langle \rho_r, r - (Ku - g) \rangle + \frac{\beta_r}{2} \|r - (Ku - g)\|_2^2,
$$

where $\beta_r, \beta_t > 0$ are the scalar penalty parameters, while $\rho_r \in \mathbb{R}^n$, $\rho_t \in \mathbb{R}^{2n}$ are the vectors of Lagrange multipliers associated with the linear constraints $r = Ku - g$ and $t = Du$ in (6.2), respectively.

By setting for simplicity $x := (u, r, t)$, $y := (\rho_r, \rho_t)$, $X := \mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R}^{2n}$, and $Y := \mathbb{R}^n \times \mathbb{R}^{2n}$, we observe that solving (6.1)–(6.2) amounts to seeking the solutions of the following saddle point problem:

$$
\text{find } (x^*; y^*) \in X \times Y
$$

$$
(6.4) \qquad \text{such that } \mathcal{L}(x^*; y) \leq \mathcal{L}(x^*; y^*) \leq \mathcal{L}(x; y^*) \quad \forall (x; y) \in X \times Y,
$$

where the augmented Lagrangian functional $\mathcal{L}$ is defined in (6.3).

Upon suitable initialization, and for any $k \geq 0$, the kth iteration of the ADMM iterative algorithm applied to solve the saddle-point problem (6.4) reads as follows:

$$
(6.5) \qquad u^{(k+1)} \leftarrow \underset{u \in \mathbb{R}^n}{\arg\min} \ \mathcal{L}(u, r^{(k)}, t^{(k)}; \rho_r^{(k)}, \rho_t^{(k)}),
$$

$$
(6.6) \qquad r^{(k+1)} \leftarrow \underset{r \in \mathbb{R}^n}{\arg\min} \ \mathcal{L}(u^{(k+1)}, r, t^{(k)}; \rho_r^{(k)}, \rho_t^{(k)}),
$$

$$
(6.7) \qquad t^{(k+1)} \leftarrow \underset{t \in \mathbb{R}^{2n}}{\arg\min} \ \mathcal{L}(u^{(k+1)}, r^{(k+1)}, t; \rho_r^{(k)}, \rho_t^{(k)}),
$$

$$
(6.8) \qquad \rho_r^{(k+1)} \leftarrow \rho_r^{(k)} - \beta_r \left(r^{(k+1)} - (Ku^{(k+1)} - g) \right),
$$

$$
(6.9) \qquad \rho_t^{(k+1)} \leftarrow \rho_t^{(k)} - \beta_t \left(t^{(k+1)} - Du^{(k+1)} \right).
$$

We notice that subproblems (6.5) and (6.6) for the primal variables u and r admit solutions based on formulas given in [29] for identical subproblems. In particular, subproblem (6.5) for u reduces to the solution of the following $n \times n$ system of linear equations

$$
(6.10) \qquad \left(D^T D + \frac{\beta_r}{\beta_t} K^T K \right) u = D^T \left(t^{(k)} - \frac{1}{\beta_t} \rho_t^{(k)} \right) + \frac{\beta_r}{\beta_t} K^T \left(r^{(k)} - \frac{1}{\beta_r} \rho_r^{(k)} + g \right),
$$

which is solvable since

$$
\text{null} \left(D^T D + \frac{\beta_r}{\beta_t} K^T K \right) = \text{null} \left(D^T D \right) \cap \text{null} \left(K^T K \right) = \text{null} \left(D \right) \cap \text{null} \left(K \right) = \{0_n\},
$$

where the last equality has been previously stated in (5.4). Assuming periodic boundary conditions for u—such that that both $D^T D$ and $K^T K$ are block circulant matrices with circulant blocks—the linear system (6.10) can be solved efficiently by one application of the forward two-dimensional (2D) fast Fourier transform (FFT) and one application of the inverse 2D FFT, each at a cost of $O(n \log n)$.

The solution of the subproblem (6.6) for r is obtained by first computing the vector

$$(6.11) \qquad w^{(k+1)} = Ku^{(k+1)} - g + \frac{1}{\beta_r}\rho_r^{(k)},$$

and then, recalling [29] and the definition of the discrepancy set in (1.10), by jointly computing the new values of both the regularization parameter μ and the variable r as follows:

$$(6.12) \quad \begin{aligned} \|w^{(k+1)}\|_2 \le \delta &\implies \mu^{(k+1)} = 0, & r^{(k+1)} = w^{(k+1)}, \\ \|w^{(k+1)}\|_2 > \delta &\implies \mu^{(k+1)} = \beta_r\big(\|w^{(k+1)}\|_2/\delta - 1\big), & r^{(k+1)} = \delta w^{(k+1)}/\|w^{(k+1)}\|_2. \end{aligned}$$

As far as the minimization subproblem for t in (6.7) is concerned, after simple algebraic manipulations, we deduce that it can be rewritten as follows:

$$t^{(k+1)} \leftarrow \operatorname*{arg\,min}_{t \in \mathbb{R}^{2n}} \ \sum_{i=1}^{n} \left\{ \|\Lambda_i R_{\theta_i} t_i\|_2^{p_i} + \frac{\beta_t}{2}\left\| t_i - \left(\left(Du^{(k+1)}\right)_i + \frac{1}{\beta_t}\left(\rho_t^{(k)}\right)_i\right)\right\|_2^2 \right\}.$$

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

$$(6.13) \qquad t_i^{(k+1)} \leftarrow \operatorname*{arg\,min}_{t_i \in \mathbb{R}^2} \left\{ \|\Lambda_i R_{\theta_i} t_i\|_2^{p_i} + \frac{\beta_t}{2}\left\| t_i - q_i^{(k+1)}\right\|_2^2 \right\}, \quad i = 1,\ldots,n,$$

where the vectors $q_i^{(k+1)} \in \mathbb{R}^2$ are defined explicitly at any iteration by

$$q_i^{(k+1)} := \left(Du^{(k+1)}\right)_i + \frac{1}{\beta_t}\left(\rho_t^{(k)}\right)_i, \quad i = 1,\ldots,n.$$

The solutions of the n bivariate optimization problems in (6.13) require the computation of a special proximal mapping operator. We dedicate the following section 6.1 to carefully discuss the solution of this optimization problem and show that it can be eventually rewritten as a one-dimensional optimization problem and thus solved efficiently.

To summarize, we report in Algorithm 6.1 the pseudocode of the proposed ADMM iterative scheme used to solve the saddle-point problem (6.3)–(6.4).

Over the last decades, the ADMM algorithm has been applied to a wide range of convex and nonconvex optimization problems arising in several areas of signal and image processing. In convex settings, several convergence results have been established for ADMM-type algorithms; see, for example, [19] and references therein. Such convergence results cover the proposed DTV$_p^{\mathrm{sv}}$-L$_2$ model in the special convex case when $p_i \ge 1$ for every $i = 1,\ldots,n$. However, very few studies on the convergence properties of ADMM in nonconvex regimes have been performed. To the best of our knowledge, provable convergence results of ADMM

in nonconvex regimes are still very limited to particular classes of problems and under certain conditions; see, e.g., [23, 50, 5]. Nevertheless, from an empirical point of view, the ADMM works extremely well for various applications involving nonconvex objectives, thus suggesting heuristically its good performance in such cases as well.

Algorithm 6.1 ADMM scheme for the solution of problem (1.7)–(1.9).

---

inputs: observed image $g \in \mathbb{R}^n$, noise standard deviation $\sigma > 0$

parameters: discrepancy parameter $\tau \simeq 1$, ADMM penalty parameters $\beta_r, \beta_t > 0$

output: approximate solution $u^* \in \mathbb{R}^n$ of (1.7)–(1.9)

  1.   **Initialization:**

  2.   · estimate model parameters $p_i, R_{\theta_i}, \Lambda_i, i = 1, \ldots, n$, by ML approach in section 4

  3.   · set $\delta = \tau \sigma \sqrt{n}$, $u^{(0)} = g$, $r^{(0)} = Ku^{(0)} - g$, $t^{(0)} = Du^{(0)}$, $\rho_r^{(0)} = \rho_t^{(0)} = 0$, $k = 0$

  4.   **while** not converging **do:**

  5.   · **update primal variables:**

  6.   · compute $u^{(k+1)}$ by solving (6.10)

  7.   · compute $r^{(k+1)}$ by applying (6.11), (6.12)

  8.   · compute $t^{(k+1)}$ see section 6.1

  9.   · **update dual variables:**

10.   · compute $\rho_r^{(k+1)}$, $\rho_t^{(k+1)}$ by applying (6.8), (6.9)

11.   · $k = k + 1$

12.   **end for**

13. $u^* = u^{(k+1)}$

6.1. A nonconvex proximal mapping solving (6.13).

6.1. A nonconvex proximal mapping solving (6.13). In this section, we describe a novel result in multivariate nonconvex proximal calculus which is crucial to efficiently solve step 8 in the ADMM Algorithm 6.1, i.e., the problem (6.13). Such a problem can be interpreted as the calculation of a nonconvex proximal mapping; see [18]. We then start recalling its definition.

Definition 6.1 (proximal map for nonconvex functions). *Let $f : \mathbb{R}^n \to \mathbb{R}$ be a proper, lower semicontinuous, and possibly nonconvex function and let $\beta > 0$. The proximal map of f with parameter β is the set-valued function $\mathrm{prox}_{\beta f} : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ defined for any $q \in \mathbb{R}^n$ by*

$$(6.14) \qquad \mathrm{prox}_{\beta f}(q) := \arg\min_{t \in \mathbb{R}^n} \left\{ f(t) + \frac{\beta}{2} \|t - q\|_2^2 \right\}.$$

Note that under such a definition the set $\mathrm{prox}_{\beta f}(q)$ is in general not a singleton. Furthermore, for some particular choices of $\beta > 0$ it may also be empty.

We present in the following the results concerned with the computation of the proximal map $\text{prox}_{\beta f}$ in (6.14), in the case when $f : \mathbb{R}^2 \to \mathbb{R}$ is the function

(6.15) $\qquad f(t) := \left(t^T A\, t\right)^{p/2}, \quad t \in \mathbb{R}^2, \quad A \in \mathbb{R}^{2 \times 2}$ symmetric positive definite, $\quad p > 0$.

The ADMM substep (6.13) will then be a special instance of (6.14) under the choice of f as above, $\beta = \beta_t$, $t = t_i$, $A = R_{\theta_i}^T \Lambda_i^2 R_{\theta_i}$, $p = p_i$, and $q = q_i^{(k+1)}$ for $i = 1, \dots, n$ and $k \geq 0$.

We now ensure that under the choice (6.15) above the minimization problem (6.14) admits solutions. Then, assuming that A has condition number $\kappa > 1$ we show how the calculation of the proximal map can be reduced to the solution of a one-dimensional problem, whose form depends on the input q and the matrix A. Note that the case $\kappa = 1$ boils down to considering a scalar and diagonal matrix A, which simplifies the problem and for which the results discussed in [32] can be used.

Proposition 6.2. *Under the choice* (6.15), *the optimization problem* (6.14) *admits at least one solution.*

Proof. Under the choice (6.15), both terms in the objective function in (6.14) are continuous, bounded from below by zero, and coercive over the entire domain $\mathbb{R}^2$. It clearly follows that the total objective function is continuous, bounded from below by zero, and coercive, hence, it admits at least one global minimizer. ∎

In the following, for $v, w \in \mathbb{R}^n$ we denote by $v \circ w$, $|v|$, and $\text{sign}(v)$ the componentwise (or Hadamard) product between v and w and the componentwise absolute value and sign of v, respectively.

Proposition 6.3. *Let* p, $\beta > 0$, $q \in \mathbb{R}^2$, *and let* $A \in \mathbb{R}^{2 \times 2}$ *be a symmetric positive definite matrix with condition number* $\kappa > 1$ *and eigenvalue decomposition*

(6.16) $\qquad A = V^T \Lambda V, \quad V^T V = V V^T = I_2, \quad \Lambda = \text{diag}(\lambda_1, \lambda_2), \ \lambda_1 > \lambda_2 > 0.$

Let us further define

(6.17) $\qquad \tilde{q} := V q, \quad s := \text{sign}(\tilde{q}), \quad \bar{q} := |\tilde{q}|, \quad \overline{\beta} := \dfrac{\beta}{\lambda_2^{p/2}}, \quad \overline{\Lambda} := \text{diag}(\kappa, 1), \quad \kappa = \dfrac{\lambda_1}{\lambda_2}.$

Then, any solution $t^* \in \mathbb{R}^2$ *of the problem*

(6.18) $\qquad \qquad t^* \in \underset{t \in \mathbb{R}^2}{\arg\min} \left\{ F(t) := (t^T A t)^{p/2} + \dfrac{\beta}{2} \, \|t - q\|_2^2 \right\}$

can be expressed as

(6.19) $\qquad \qquad t^* = V^T (s \circ z^*), \quad z^* \in \underset{z \in \mathcal{H}_1}{\arg\min} \ H(z),$

where the objective function $H : \mathbb{R}^2 \to \mathbb{R}$ *and the feasible set* $\mathcal{H}_1 \subset \mathbb{R}^2$ *are defined by*

(6.20) $\qquad H(z) := \left(z^T \overline{\Lambda}\, z\right)^{p/2} + \dfrac{\overline{\beta}}{2} \, \| z - \bar{q} \|_2^2, \qquad \mathcal{H}_1 := \mathcal{H} \cap \left([0, \bar{q}_1] \times [0, \bar{q}_2]\right)$

with $\mathcal{H}$ being the rectangular hyperbola defined by

$$(6.21) \quad \mathcal{H} := \left\{ (z_1, z_2) \in \mathbb{R}^2 : (z_1 - c_1)(z_2 - c_2) = c_1 c_2, \quad c_1 = -\frac{\bar{q}_1}{\kappa - 1}, \ c_2 = \frac{\kappa \bar{q}_2}{\kappa - 1} \right\}.$$

Proof. We start by noticing that the matrix Λ in (6.16) can be factorized as $\Lambda = \lambda_2 \overline{\Lambda}$, where $\overline{\Lambda}$ is defined in (6.17). By substituting such factorization into (6.16), we can reformulate problem (6.18) as

$$(6.22) \qquad t^* \in \underset{t \in \mathbb{R}^2}{\arg\min} \left\{ \lambda_2^{p/2} \left(t^T V^T \overline{\Lambda} V t \right)^{p/2} + \frac{\beta}{2} \| t - q \|_2^2 \right\}.$$

After introducing the bijective linear change of variable

$$y := V t \quad \Longleftrightarrow \quad t = V^T y,$$

we have that problem (6.22) can be equivalently expressed as

$$(6.23) \qquad t^* = V^T y^*,$$

$$(6.24) \qquad y^* \in \underset{y \in \mathbb{R}^2}{\arg\min} \left\{ G(y) := \left(y^T \overline{\Lambda} y \right)^{p/2} + \frac{\overline{\beta}}{2} \| y - \tilde{q} \|_2^2 \right\},$$

where $\overline{\beta}$ and $\tilde{q}$ are defined in (6.17).

If $\tilde{q}_1 = \tilde{q}_2 = 0$ then one can trivially show that clearly $y^* = (0,0) \Longrightarrow t^* = (0,0)$. We can then assume that $\tilde{q} \in \mathbb{R}^2 \setminus \{0\}$ and exploit symmetries of the function G in (6.24) to restrict the optimization problem to the case where $\tilde{q}$ lies in the first quadrant only. First, we notice that, for any given $a \in \mathbb{R}$ and $b \in \mathbb{R} \setminus \{0\}$, we have

$$(6.25) \qquad a^2 = (\text{sign}(b))^2 a^2 = (\text{sign}(b) a)^2,$$

$$(a - b)^2 = (a - \text{sign}(b)|b|)^2 = \left(\text{sign}(b) \left(\frac{a}{\text{sign}(b)} - |b| \right) \right)^2$$

$$(6.26) \qquad = (\text{sign}(b))^2 (\text{sign}(b) a - |b|)^2 = (\text{sign}(b) a - |b|)^2.$$

By now recalling definitions of function G in (6.24) and of matrix $\overline{\Lambda}$ in (6.17), and then using (6.25)–(6.26), we can write

$$G(y) = \left(\kappa y_1^2 + y_2^2 \right)^{p/2} + \frac{\overline{\beta}}{2} \left((y_1 - \tilde{q}_1)^2 + (y_2 - \tilde{q}_2)^2 \right)$$

$$= \left(\kappa \left(\text{sign}(\tilde{q}_1) y_1 \right)^2 + \left(\text{sign}(\tilde{q}_2) y_2 \right)^2 \right)^{p/2} + \frac{\overline{\beta}}{2} \left(\left(\text{sign}(\tilde{q}_1) y_1 - |\tilde{q}_1| \right)^2 + \left(\text{sign}(\tilde{q}_2) y_2 - |\tilde{q}_2|^2 \right) \right).$$

By setting $S := \text{diag}\left(\text{sign}\left(\tilde{q}_1 \right), \text{sign}\left(\tilde{q}_2 \right) \right)$ we can now set

$$z := S y \quad \Longleftrightarrow \quad y = S^{-1} z,$$

which is a linear bijective change of variable since $\tilde{q}_1, \tilde{q}_2 \in \mathbb{R} \setminus \{0\} \Longrightarrow \text{sign}\left(\tilde{q}_1 \right), \text{sign}\left(\tilde{q}_2 \right) \in \{-1, 1\}$. Recalling the definition of s and $\overline{q}$ in (6.17), we thus get that the optimization
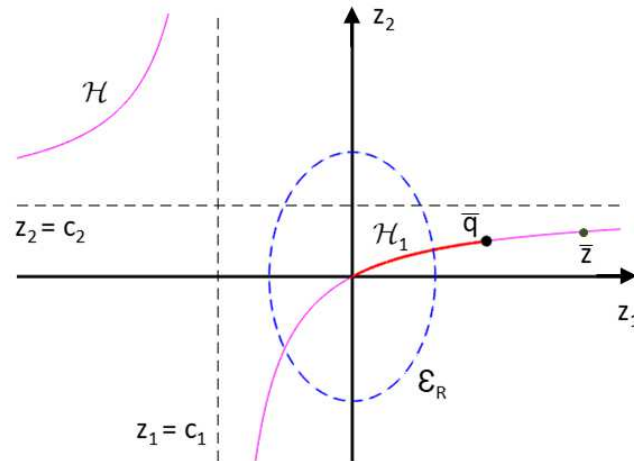
Figure 2. *Graphical representation for the bivariate minimization problem* (6.28).

problem (6.24) is equivalent to

$$(6.27) \qquad y^* = s \circ z^*,$$

$$(6.28) \qquad z^* \in \operatorname*{arg\,min}_{z \in \mathbb{R}^2} \left\{ H(z) := \left(z^T \overline{\Lambda} z \right)^{p/2} + \frac{\overline{\beta}}{2} \| z - \overline{q} \|_2^2 \right\},$$

where the vector $\overline{q} = (|\tilde{q}_1|, |\tilde{q}_1|)$ now lies in the first (open) quadrant $(0, +\infty)^2$.

We now prove that the solutions z^* in (6.28) belong to the arc of hyperbola $\mathcal{H}_1$ defined in (6.20). To this aim, we consider the following one-parameter family of ellipses depending on a parameter $R > 0$:

$$\mathcal{E}_R := \left\{ (z_1, z_2) \in \mathbb{R}^2 : \ z^T \overline{\Lambda} z = R^2 \right\} = \left\{ (z_1, z_2) \in \mathbb{R}^2 : \ \kappa z_1^2 + z_2^2 = R^2 \right\}$$

$$(6.29) \quad = \left\{ (z_1, z_2) \in \mathbb{R}^2 : \ z_1 = z_1(\theta; R) = \frac{R}{\sqrt{\kappa}} \cos \theta, \ z_2 = z_2(\theta; R) = R \sin \theta, \ \theta \in [0, 2\pi[\right\}$$

and, as a start, we show that the minimizers of the restriction of the function H in (6.28) to any ellipse $\mathcal{E}_R$ in (6.29) lie on the hyperbola $\mathcal{H}$ in (6.21). In Figure 2 we show the hyperbola $\mathcal{H}$ (magenta solid line) with its two orthogonal asymptotes, the arc $\mathcal{H}_1$ defined in (6.20) (red solid thick line), and one ellipse $\mathcal{E}_R$ (blue dashed line) as in (6.29).

Let us observe first that when restricted to an ellipse $\mathcal{E}_R$ of the form in (6.29), the objective function H depends only on θ (R can be regarded as a fixed parameter). The restriction $H_R : \mathbb{R} \to \mathbb{R}$ then takes the following form

$$H_R(\theta; R) = R^p + \frac{\overline{\beta}}{2} \left(\left(\frac{R}{\sqrt{\kappa}} \cos \theta - \overline{q}_1 \right)^2 + \left(R \sin \theta - \overline{q}_2 \right)^2 \right).$$

For any $R > 0$, the function H_R above is clearly periodic with period 2π, bounded (from below and above), and infinitely many times differentiable in θ, hence, the minimizers of H_R can be sought for among its stationary points in the interval $[0, 2\pi)$. The first-order derivative

of H_R is as follows:

$$H'_R(\theta; R) = \overline{\beta}\left(-\frac{R}{\sqrt{\kappa}}\sin\theta\left(\frac{R}{\sqrt{\kappa}}\cos\theta - \overline{q}_1\right) + R\cos\theta\left(R\sin\theta - \overline{q}_2\right)\right)$$

$$(6.30) \qquad = \overline{\beta}\,\frac{\kappa-1}{\sqrt{\kappa}}\,\left((z_1(\theta; R) - c_1)(z_2(\theta; R) - c_2) - c_1 c_2\right),$$

where (6.30) follows after some simple algebraic manipulations from the parametrization in (6.29) with c_1, c_2 constants defined in (6.21). Since $\beta > 0$, $\kappa > 1$ by assumption, the scalar quantity $\overline{\beta}(\kappa-1)/\sqrt{\kappa}$ in (6.30) is positive, hence we have

$$(6.31) \quad H'_R(\theta; R) = 0 \,(> 0, < 0) \iff (z_1(\theta; R) - c_1)(z_2(\theta; R) - c_2) - c_1 c_2 = 0 \,(> 0, < 0).$$

It thus follows that, for any fixed $R > 0$ (that is, for any ellipse $\mathcal{E}_R$ in (6.29)), any stationary point $z(\theta_R^* : R)$ of H_R satisfies

$$\left(z_1\left(\theta_R^*; R\right),\, z_2\left(\theta_R^*; R\right)\right) \in \mathcal{E}_R \cap \mathcal{H},$$

i.e., it belongs to the set of intersection points between the ellipse $\mathcal{E}_R$ and the hyperbola $\mathcal{H}$ (see the two intersection points in Figure 2). It also follows from (6.31) that the intersection point in the first quadrant is the global minimizer for H_R, whereas the one in the third quadrant is the global maximizer. Since previous considerations hold true for any ellipse $\mathcal{E}_R$, then any global minimizer z^* of the unrestricted objective function H in (6.28) must belong to the restriction of the hyperbola $\mathcal{H}$ in (6.21) to the first quadrant.

Finally, it is easy to further shrink the locus of potential global minimizers z^* to the arc $\mathcal{H}_1$ defined in (6.20). Let us argue by contradiction and suppose there exists a global minimizer $\overline{z}$ belonging to the restriction of the hyperbola $\mathcal{H}$ to the first quadrant but not to $\mathcal{H}_1$; see Figure 2. We have

$$H(\overline{z}) - H(\overline{q}) = \underbrace{\left(\overline{z}^T\overline{\Lambda}\,\overline{z}\right)^{p/2} - \left(\overline{q}^T\overline{\Lambda}\,\overline{q}\right)^{p/2}}_{>0} + \underbrace{\frac{\overline{\beta}}{2}\left(\|\overline{z} - \overline{q}\|_2^2 - \|\overline{q} - \overline{q}\|_2^2\right)}_{>0} > 0,$$

whence $\overline{z}$ cannot be a global minimizer for the function H. ∎

In the following corollary we exploit and complete the results in the previous Proposition 6.3 by showing how the bivariate minimization problem in (6.19) can be reduced to an equivalent univariate problem.

Corollary 6.4. *The minimizers $z^* \in \mathbb{R}^2$ in (6.19) can be obtained as follows:*

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

with c_1, $c_2 \in \mathbb{R}$ defined in (6.21) and $z_1^* \in \mathbb{R}$ the solution(s) of the following one-dimensional constrained minimization problem:

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

$$h_1(\xi) = \xi^2\left(\kappa + \frac{c_2^2}{(\xi - c_1)^2}\right), \quad h_2(\xi) = \xi\,(\kappa - 1)\left(\xi - 2c_1 + 2\,\frac{c_2^2}{\kappa(\xi - c_1)}\right).$$

Proof. The proof is immediate by deriving the expression of z_2 as a function of z_1 from the definition of the hyperbola $\mathcal{H}$ in (6.21), then substituting this expression in the objective function H in (6.20) and, finally, carrying out some algebraic manipulations. ∎

7. Parameters estimation results. In this section, an extensive evaluation on the accuracy of the ML estimation procedure described in section 4 is carried out.

In order to assess the quality of the estimation, we introduce in the following some useful statistical notions.

Definition 7.1. *Let $\omega > 0$ be an unknown parameter of a fixed probability distribution p_ω and for $\ell > 0$ let ω_j, $j = 1, \ldots, \ell$, be estimates of ω obtained by a given estimation procedure. The sample estimator $\hat{\omega}$ of ω is defined as the average:*

$$\hat{\omega} := \frac{\sum_{j=1}^{\ell} \omega_j}{\ell}.$$

We can then define the relative bias $\mathcal{B}_{\hat{\omega}}$, *the* empirical variance $\mathcal{V}_{\hat{\omega}}$, *and the* relative root mean square error *$rmse_{\hat{\omega}}$ of the estimator $\hat{\omega}$ as*

$$\mathcal{B}_{\hat{\omega}} := \frac{\mathbb{E}(\hat{\omega} - \omega)}{\omega}, \quad \mathcal{V}_{\hat{\omega}} := \frac{1}{\ell - 1} \sum_{j=1}^{\ell} (\omega_j - \hat{\omega})^2, \quad rmse_{\hat{\omega}} := \frac{\sqrt{\mathcal{V}_{\hat{\omega}} + \mathcal{B}_{\hat{\omega}}^2}}{\omega}.$$

In the following, the accuracy and the precision of the estimator is evaluated by analyzing its performance on the estimation of the parameters $(p, e^{(1)}, \theta)$. As discussed in section 4.2, parameters $e^{(1)}, \theta$ can be derived from ϱ, ϕ and from (4.13), we recall that $e^{(1)} = (\frac{1}{\lambda^{(1)}})^2$. In addition, we also consider how the quality of the estimation of $(p, e^{(1)}, \theta)$ affects the estimation of the scale parameter m, which is computed directly via the formula (4.5) as a nonlinear function of (p, ϕ, ϱ) or, equivalently, of $(p, e^{(1)}, \theta)$, as well as of the samples. The nonlinearity may affect the accuracy of its estimation.

7.1. Parameter estimation: Accuracy and precision. We now perform some tests assessing the accuracy and the precision of the ML estimation procedure proposed in section 4 in terms of the quantities defined above. As a first test we compare the results obtained by applying the ML procedure to estimate a BGGD of parameters $(\bar{p}, \bar{e}^{(1)}, \bar{\theta}, \bar{m}) = (1, 1.4, 45°, 0.3)$. We run our tests for an increasing number $N \in \{10, 10^2, 10^3, 10^4, 10^5, 10^6\}$ of samples drawn from the distribution. For each value of N, the estimation procedure is run $\ell = 200$ times. For any $j = 1, \ldots, \ell$ we estimate the parameter triple $(p^*, \phi^*, \varrho^*)_j$ and consider the corresponding estimators of the true parameters as defined in Definition 7.1. The results are shown in Figures 3–5.

For all parameters (including the scale parameter m), the behavior of relative bias, variance, and relative root mean square error as the number of samples increases reveals good precision and accuracy. In particular, low values of such error quantities are already obtained when $N \approx 10^2$.

7.2. Parameter estimation on synthetic neighborhoods. We now test the ML estimation procedure on a simple synthetic image reported in Figure 6(a). Here, the goal is to evaluate the effectiveness of the estimation when discriminating between different image regions such
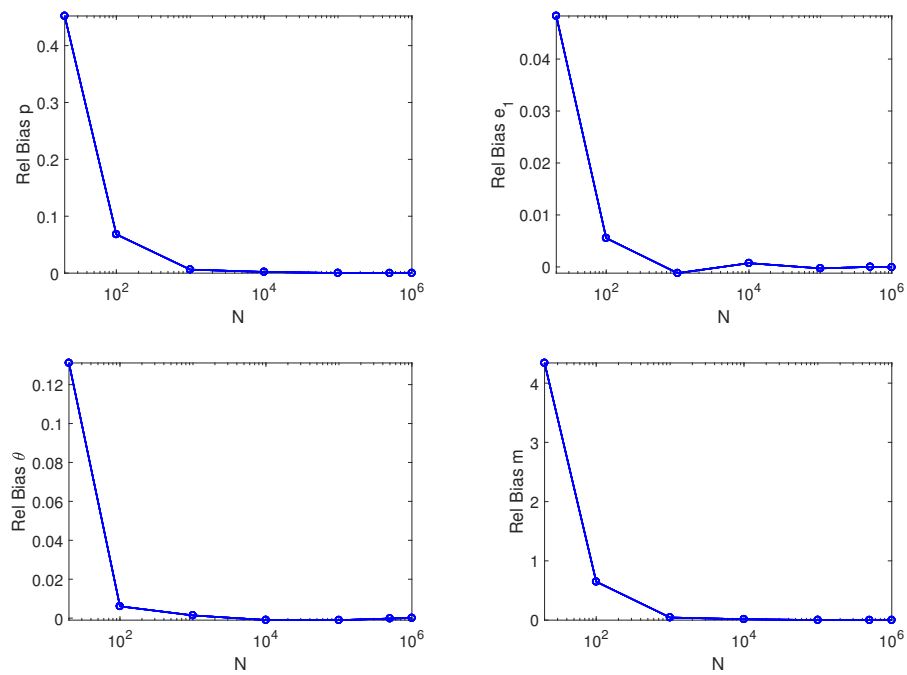
Figure 3. *Plots of relative bias for estimated $(p^*, e^{(1)*}, \theta^*, m^*)$ in semilogarithmic scale on x-axis.*
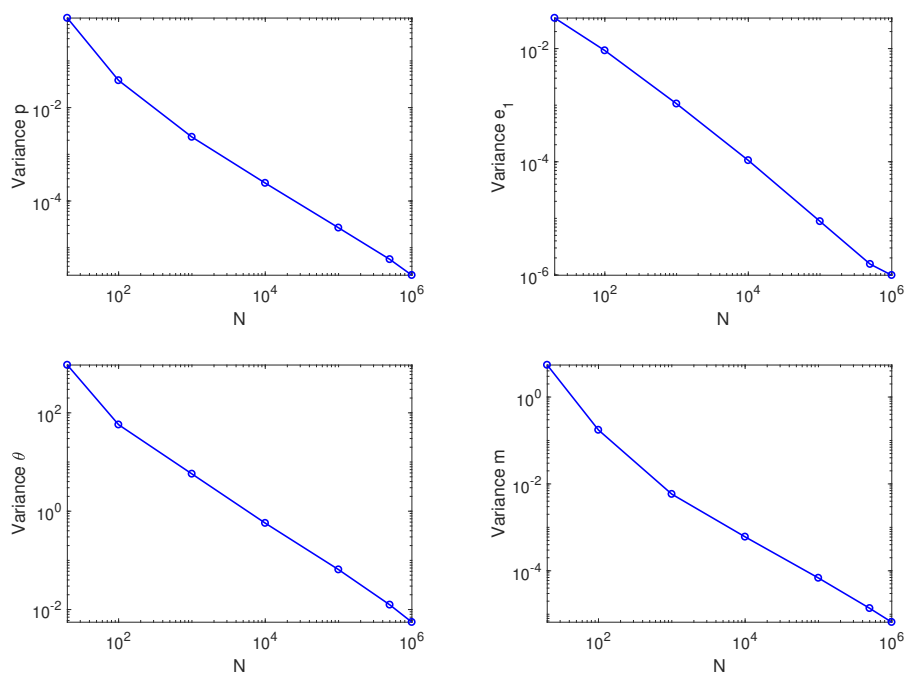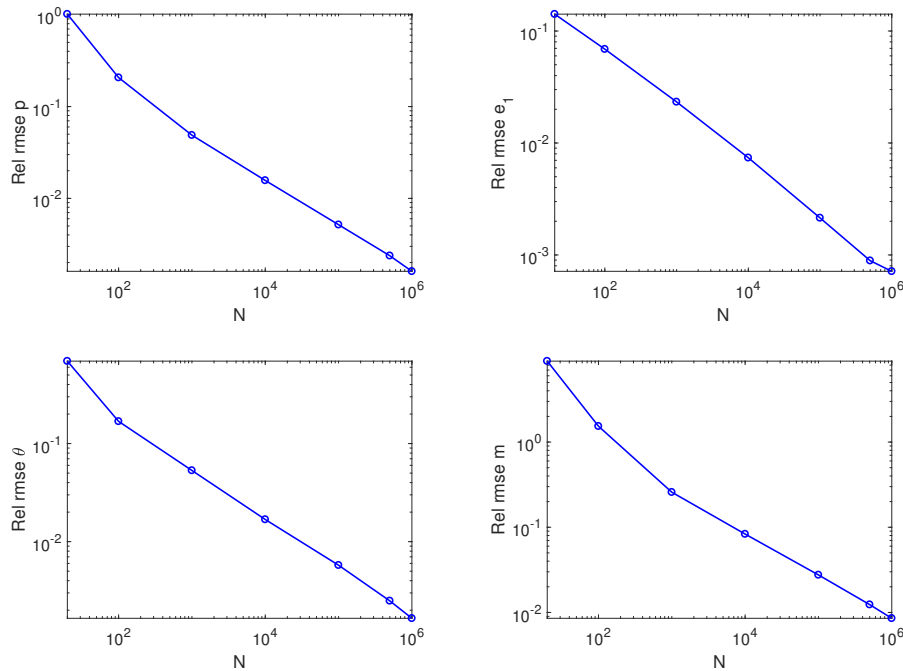

Figure 4. *Plot of the empirical variance for estimated $(p^*, e^{(1)*}, \theta^*, m^*)$ in semilogarithmic scale on x-axis.*

Figure 5. *Plot of relative root mean square error for estimated* $(p^*, e^{(1)^*}, \theta^*, m^*)$ *in semilogarithmic scale on x-axis.*

as edges, corners, and circular profiles in terms of the functional shape of the estimated BGGD. In the following test, we estimate the parameters of the unknown BGGD in three different situations where a pixel surrounded by a 11×11 neighborhood is chosen to lie on a vertical edge (Figure 6), a corner (Figure 7), and on a circular profile (Figure 8). In order to avoid degenerate configurations of the gradients, such as the ones described in (4.11), we preliminarily corrupt the image by a small AWGN with $\sigma = 0.03$.

Edge points. In Figure 6(b), we report the scatter plot of the gradients of the edge points in the red-bordered region encountered in Figure 6(a), which, as expected, shows its distribution along the x-axis. The parameter estimation procedure of the BGGD at one of such edge points is run by taking 121 samples of gradients in the 11×11 neighborhood. The estimation procedure results in the following parameters $(p^*, e^{(1)^*}, \theta^*, m^*) = (0.07, 1.60, -177.82°, 2*10^{-5})$. Note that the low value of the parameter p leads to a very-fat-tail distribution, as shown in Figure 6(c). The orientation and the eccentricity of the level curves are in line with the clear directionality of the samples as can be seen in Figure 6(d).

Corner points. For the corner example in Figure 7, the scatter plot of the gradients is reported in Figure 7(b). The ML procedure results in this case in the estimation $(p^*, e^{(1)^*}, \theta^*, m^*) = (0.07, 1.08, 72.49°, 3*10^{-7})$. The estimated PDF is reported in Figure 7(c). Similarly as before, note that a very-fat-tail distribution is estimated. On the other hand, since $e^{(1)^*} \approx 1$, we also have $e^{(2)^*} \approx 1$ and the eccentricity of the ellipse $\epsilon \approx 0$. We can conclude that, in this case, the distribution is almost isotropic and the angle θ has a negligible influence on the orientation of the level curves as can be seen in Figure 6(d).
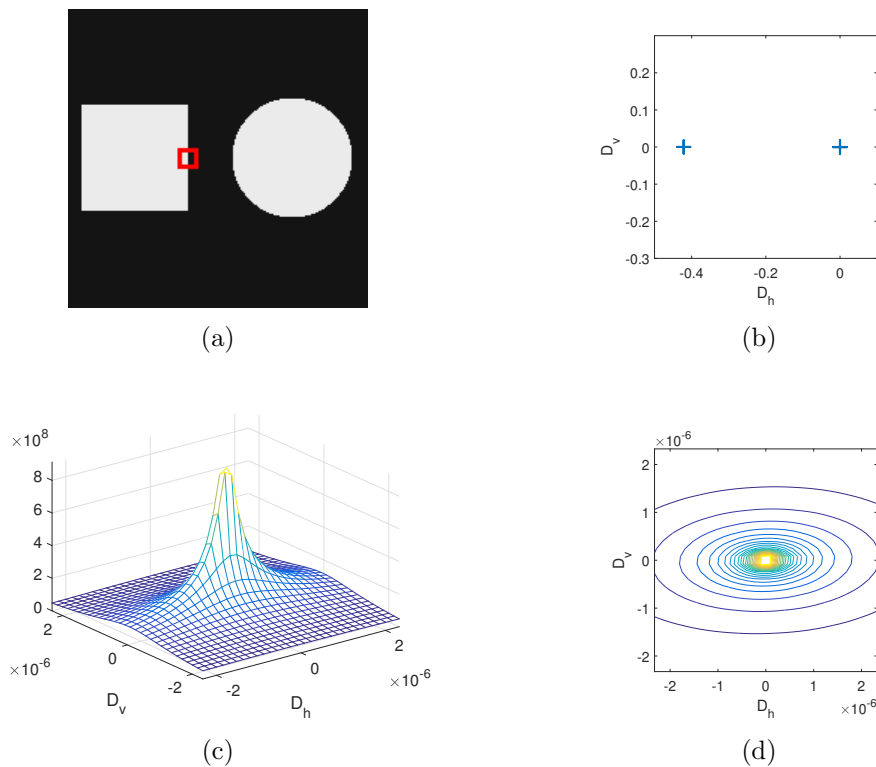
Figure 6. (a): *BGGD parameter estimation for a synthetic geometrical image. Test for edge image pixel.* (b): *Scatter plot of the gradients in the read-bordered region.* (c): *PDF with estimated parameters* $(p^*,\ e^{(1)^*},\ \theta^*,\ m^*) = (0.07,\ 1.60,\ -177.82°,\ 2*10^{-5})$. (c): *Level curves of the estimated PDF.*

Circle points. Finally, we consider the ML parameter estimation procedure in correspondence with a pixel lying on a circular profile; see Figure 8. In this case, the estimated parameters are $(p^*, e^{(1)^*}, \theta^*, m^*) = (0.08, 1.44, 49.28°, 2*10^{-6})$. The values obtained for $e^{(1)^*}$ and θ^* reflect the spatial distribution of the gradients in Figure 8(b).
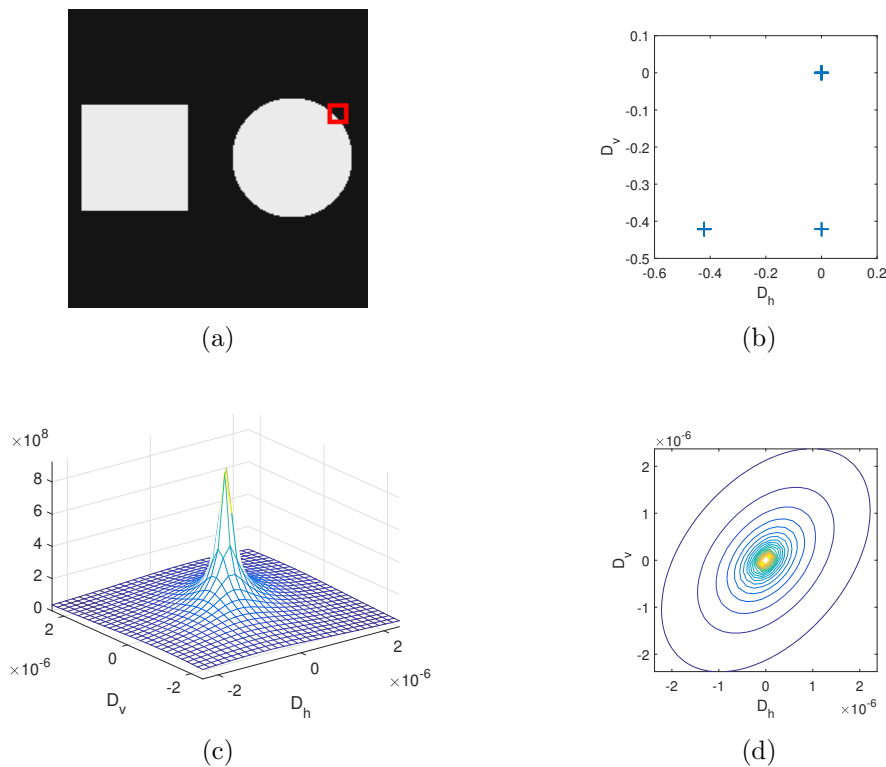
7.3. Parameter estimation on synthetic images. Motivated by the good results above, we report in this section the numerical experiments concerned with the estimation of the four parameters $(p^*, e^{(1)^*}, \theta^*, m^*)$ at any image pixel. For the following estimations, we fix a neighborhood of 3×3 pixels, It is worth remarking here that the tests in section 7.1 have been computed on samples directly drawn from a BGGD. For such examples, we remarked on how a large number of samples reflects on a reliable estimation of the BGGD parameters. When dealing with real images, however, our goal rather consists in estimating the parameters of the BGGD of the local gradient from the surrounding ones, since, clearly, one single sample is not sufficient to get a reliable estimate. However, the samples involved in the estimation procedure are in general not drawn from the same BGGD as their parameters may be different. Thus, their number has to be limited in order to reduce modeling errors as much as possible. In conclusion, the size of the neighborhood is a trade-off between the local properties of the image and the robustness of the estimate procedure, the former requiring small neighborhoods,

(a)

(b)

(c)

(d)
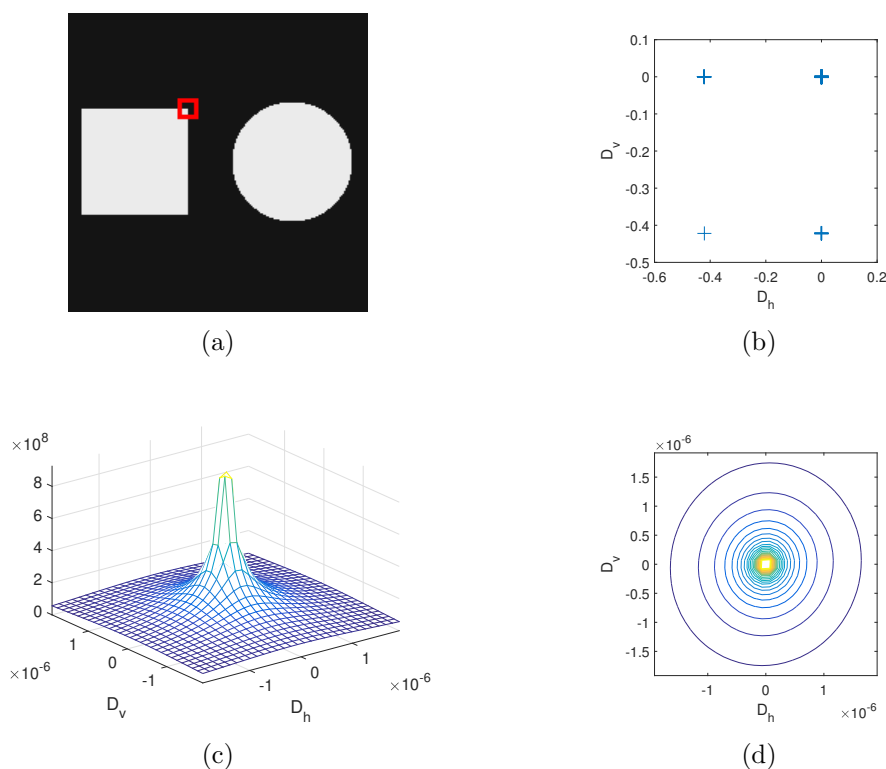
Figure 7. (a): *BGGD parameter estimation for a synthetic geometrical image. Test for corner image pixel.* (b): *Scatter plot of the gradients in the read-boarded region.* (c): *PDF with estimated parameters* $(p^*,\ e^{(1)^*},\ \theta^*,\ m^*) = (0.07,\ 1.08,\ 72.49°,\ 3*10^{-7})$. (d): *Level curves of the estimated PDF.*

the latter requiring larger ones. In order to avoid degenerate configurations, we corrupt the images by AWGN with $\sigma = 0.03$. Moreover, the search interval for the shape parameter p is set equal to $[0.1, 5]$. We start considering the synthetic test image used already in the experiment above, i.e., Figure 9(a). Here we perform the estimation of the parameters at any pixel and report the local parameter maps in Figure 9(c), 9(d), 9(e), and 9(f). Furthermore, we report in Figure 9(b) the anisotropy ellipses representing the level curves of the estimated PDF, drawn as described in section 4.2, whose orientation, given by the θ-map in Figure 9(e), is in line with what we expected and with the test proposed in the previous subsection (see Figures 6–7). One can also observe that the higher values in the $e^{(1)}$-map are estimated to be along the edges, describing the strong anisotropy of the level curves there, while the higher values in the p-map are in the piecewise constant regions. This can be explained by saying that in these regions the estimation procedure detects a plain bivariate Gaussian distribution characterized by a shape parameter $p = 2$. This is of course due to the presence of AWGN.

The same experiments are proposed for a `geometric` test image in Figure 10(a). Even though such an image presents edges displaced along different orientations and details on different scales; the results showed in Figures 10(b)–10(f) confirm the robustness of the estimator in distinguishing between different image regions.

Figure 8. (a): *BGGD Parameter estimation for a synthetic geometrical image. Test on image pixel lying on circular profile.* (b): *Scatter plot of the gradients in the red-boarded region.* (c): *PDF with estimated parameters* $(p^*, e^{(1)^*}, \theta^*, m^*) = (0.08, 1.44, 49.28°, 2*10^{-6})$. (d): *Level curves of the estimated PDF.*

Remark 7.2. In order to generate the samples used in the parameter map estimation above, one has to choose a suitable discretization of the image gradient. Here, we considered central differences schemes. Compared to standard forward/backward difference schemes, this choice avoids the undesired correlation between the horizontal and the vertical components. As preliminary numerical tests showed, such a correlation may indeed result in a deviation between the estimated θ^* from the one estimated above.

8. Applications to image denoising and deblurring. In this section, we evaluate the performance of the $\mathrm{DTV}_p^{\mathrm{sv}}$-$\mathrm{L}_2$ image reconstruction model (1.7)–(1.9) applied to the restoration of gray-scale images corrupted by (known) blur and AWGN.

Denoting by u the ground-truth image, the quality of the given corrupted images g and of the restored images u^* is measured by means of standard image quality measures, i.e., the blurred SNR

$$\mathrm{BSNR}(u^*, u) := 10 \log_{10} \frac{\|Ku - \overline{Ku}\|_2^2}{\|u^* - Ku\|_2^2},$$

where by $\overline{Ku}$ we have denoted the average intensity of the blurred image Ku, and the improved

(a) Test image.

(b) Anisotropy ellipses.

(c) p map.

(d) $e^{(1)}$ map.

(e) θ map.

(f) m map.
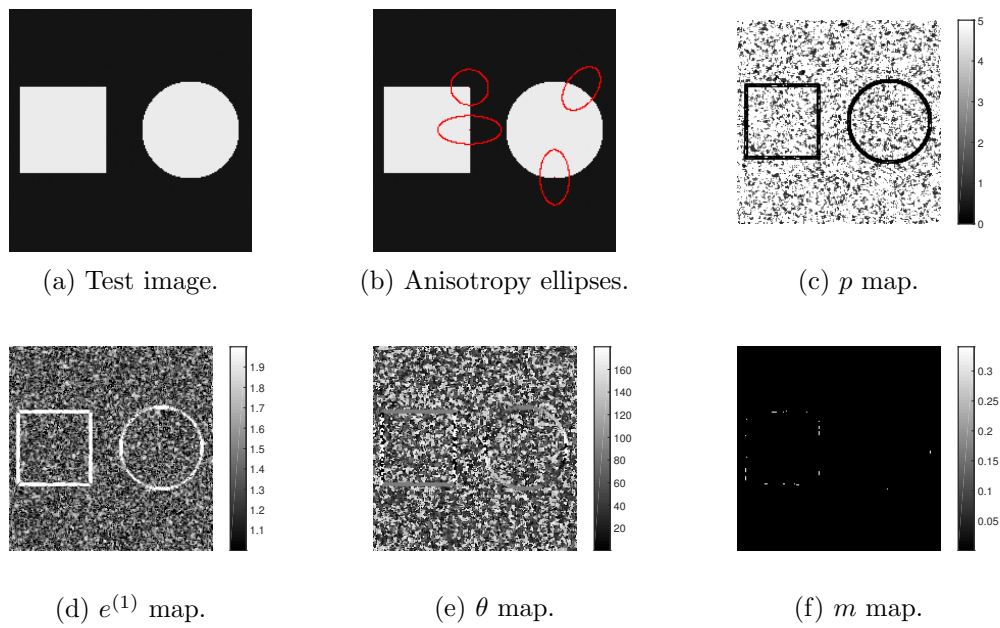
Figure 9. *Test on* `synthetic` *image.*

(a) Test image.

(b) Anisotropy ellipses.

(c) p map.

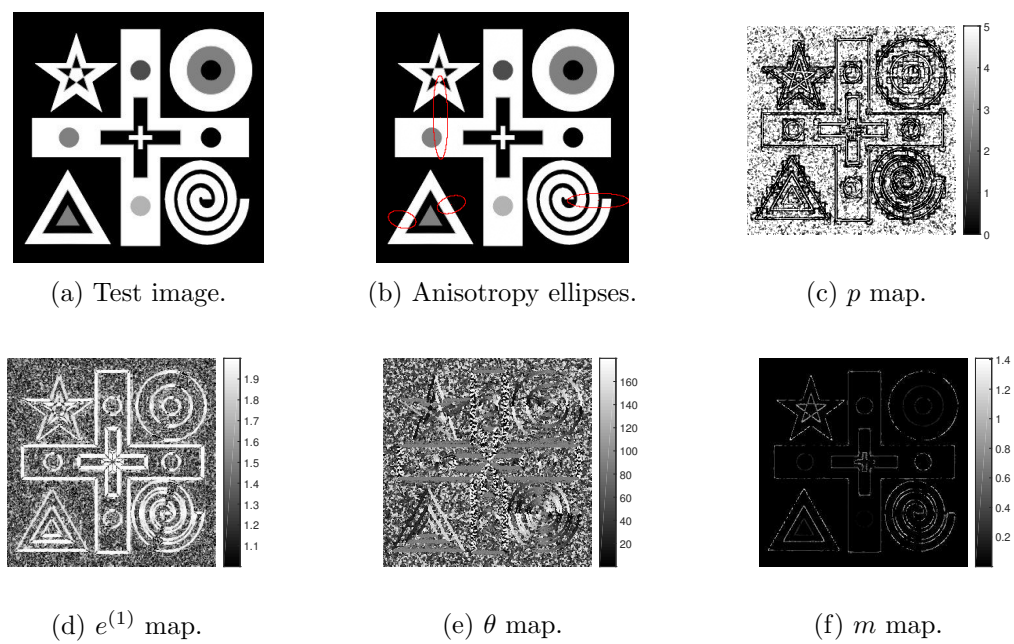(d) $e^{(1)}$ map.

(e) θ map.

(f) m map.

Figure 10. *Test on* `geometric` *image.*

signal-to-noise ratio

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

defined also in terms of the given noisy g. The larger the BSNR and the ISNR values, the higher the quality of restoration. For a more visually inspired standard quality measure, we will also quantify our results in terms of the standard SSIM, [56].

The $\mathrm{DTV}_p^{\mathrm{sv}}$-$\mathrm{L}_2$ model will be compared with the following ones:

- The ROF or TV-L_2 model [42];
- the TV$_p$-L_2 model, with constant $p \in (0, 2]$; see [32];
- the TV$_{\alpha,p}^{\mathrm{sv}}$-$\mathrm{L}_2$ model with space variant $p_i \in (0, 2]$, $i \in \{1, \dots, n\}$ [30, 29].

We stress that in order to compute the following results, an accurate and reliable estimation of the parameters appearing in the $\mathrm{DTV}_p^{\mathrm{sv}}$ needs to be performed. We do that by means of the ML procedure described in section 4 whose accuracy has been extensively confirmed by the tests in section 7.

For the numerical solution of the $\mathrm{DTV}_p^{\mathrm{sv}}$-$\mathrm{L}_2$ model we use the ADMM-based Algorithm 6.1 where for all tests we manually set the penalty parameters β_t and β_r. Iterations are stopped whenever the following stopping criterion is verified:

$$\frac{\left\| u^{(k)} - u^{(k-1)} \right\|_2}{\left\| u^{(k-1)} \right\|_2} < 10^{-4}.$$

Finally, the parameter μ is set based on the discrepancy principle, stated in (1.10).

Barbara image. We start testing the reconstruction algorithm on a zoom of a high resolution (1024×1024) `barbara` test image with size 471×361, characterized by the joint presence of texture and cartoon regions. The image here has been corrupted by a Gaussian blur of `band`$= 9$ and `sigma`$= 2$ and AWGN resulting in BSNR values equal to 20 dB, 15 dB, and 10 dB. The original image and the observed image, as well as the four parameter maps, computed considering a neighborhood of size 7×7, are shown in Figure 11. In order to avoid inaccurate estimations of the parameters due to the presence of possibly large noise, the parameter p^* in the TV$_p$-L_2 model as well as the local maps of the parameters in the $\mathrm{DTV}_p^{\mathrm{sv}}$-$\mathrm{L}_2$ have been computed after a few iterations (usually 5) of the TV-L_2 model. Furthermore, as discussed in section 4.2, the p parameter has been computed by restricting the admissible range to $[0.1, 2]$. In Tables 1 and 2 the ISNR and SSIM values achieved by the TV-L_2, TV$_p$-L_2 (with estimated global $p = 0.92$), TV$_{\alpha,p}^{\mathrm{sv}}$-$\mathrm{L}_2$ (with space variant parameters estimated as in [29]), and $\mathrm{DTV}_p^{\mathrm{sv}}$-$\mathrm{L}_2$ models for different values of the initial BSNR are reported. We note that the proposed model outperforms the competing ones. As shown in Figure 12, the flexibility of the $\mathrm{DTV}_p^{\mathrm{sv}}$ regularizer strongly improves the reconstruction quality mainly in terms of better texture preservation.

Natural image. As a second test, we compared the performance of a $\mathrm{DTV}_p^{\mathrm{sv}}$-$\mathrm{L}_2$ restoration model on a 500×500 portion of a high resolution (1024×1024) natural test image characterized by fine-scale textures of different types. As in the previous example, we similarly corrupt the image by AWGN and a Gaussian blur of `band` $= 9$ and `sigma` $= 2$ with BSNR $= 20$ dB, 15 dB, and 10 dB. The original and the observed images, as well as the four parameter maps computed considering neighborhoods of size 3×3, are shown in Figure 13. Similarly as for the

(a) Zoom of original u.

(b) Zoom of g.

(c) p map.

(d) $e^{(1)}$ map.

(e) θ map.

(f) m map.
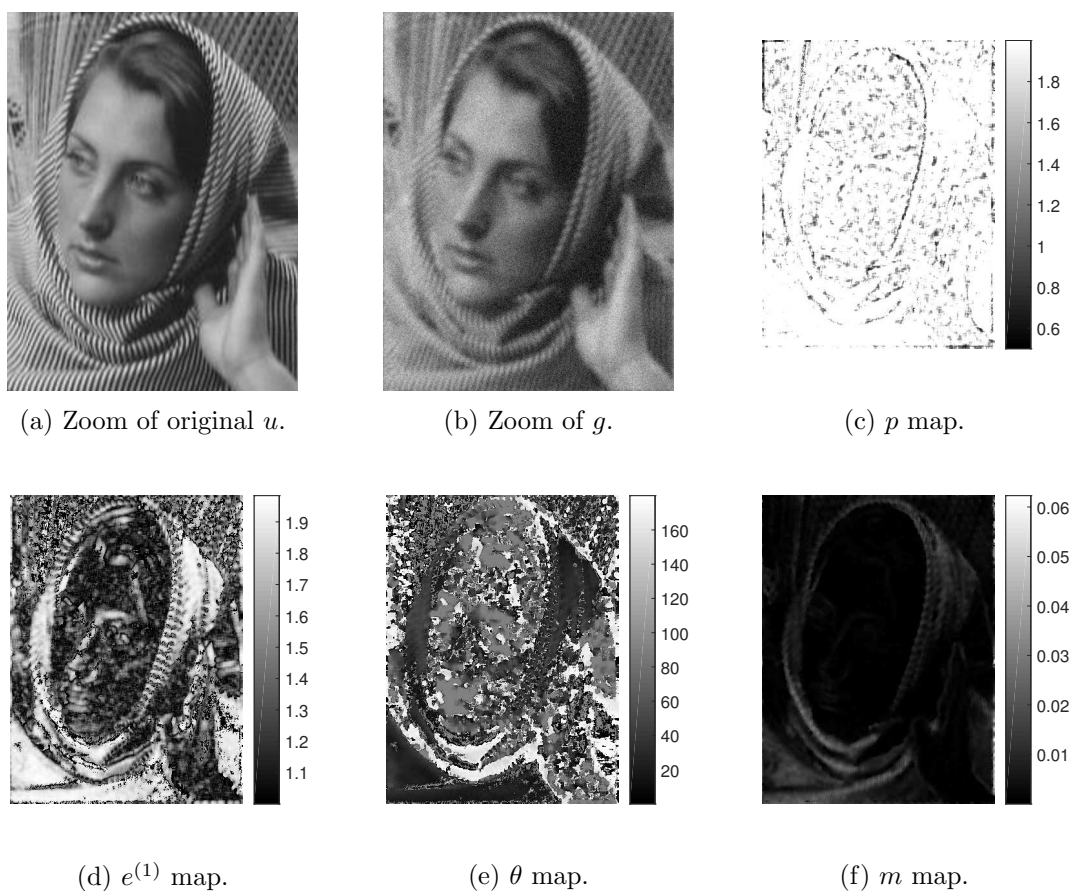
Figure 11. *Parameter maps for a zoom of the barbara test image. Image is corrupted by AWGN and blur for a $BSNR = 10$ dB.*

Table 1
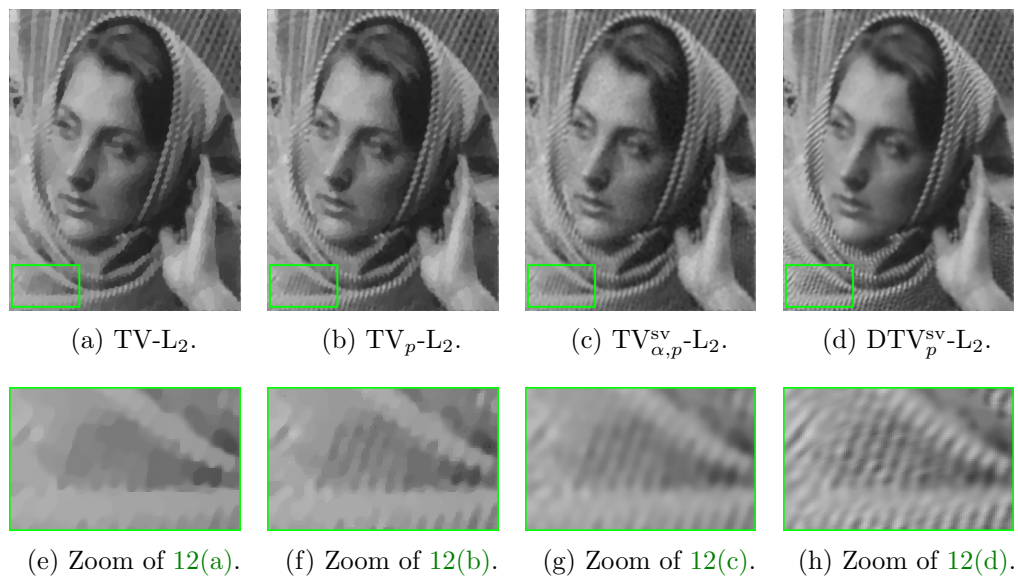*ISNR values for the barbara test image for decreasing $BSNR = 20, 15, 10$ dB.*

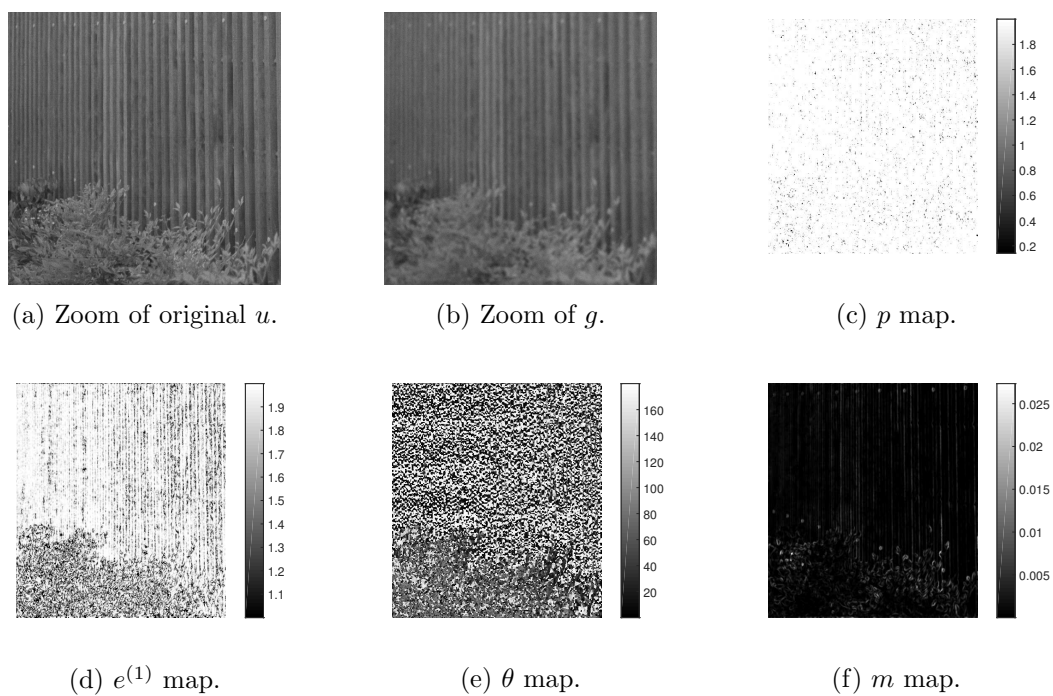| BSNR	TV-L$_2$	TV$_p$-L$_2$	TV$^{sv}_{\alpha,p}$-L$_2$	DTV$^{sv}_p$-L$_2$
20	2.46	3.14	3.23	**3.61**
15	1.74	1.99	2.14	**2.79**
10	1.59	2.02	2.13	**2.90**

Table 2
SSIM values for the barbara test image for decreasing $BSNR = 20, 15, 10$ dB.

BSNR	TV-L$_2$	TV$_p$-L$_2$	TV$^{sv}_{\alpha,p}$-L$_2$	DTV$^{sv}_p$-L$_2$
20	0.80	0.83	0.83	**0.85**
15	0.74	0.75	0.77	**0.80**
10	0.65	0.68	0.69	**0.74**

(a) TV-L$_2$.   (b) TV$_p$-L$_2$.   (c) TV$^{\mathrm{sv}}_{\alpha,p}$-L$_2$.   (d) DTV$^{\mathrm{sv}}_p$-L$_2$.

(e) Zoom of 12(a).   (f) Zoom of 12(b).   (g) Zoom of 12(c).   (h) Zoom of 12(d).

Figure 12. *Detail of reconstruction of* `barbara` *image in Figure* 11. *Texture components are much better preserved by encoding directional information.*

(a) Zoom of original u.   (b) Zoom of g.   (c) p map.

(d) $e^{(1)}$ map.   (e) θ map.   (f) m map.

Figure 13. *Parameter maps for a zoom of a natural test image. Image is corrupted by AWGN and blur for a $BSNR = 10$ dB.*

Table 3
ISNR values for the test image in Figure 13 *for* $BSNR = 20, 15, 10$ *dB.*

BSNR	TV-L$_2$	TV$_p$-L$_2$	TV$_p^{\mathrm{sv}}$-L$_2$	DTV$_p^{\mathrm{sv}}$-L$_2$
20	2.07	2.43	2.53	**2.78**
15	1.83	2.06	2.26	**2.56**
10	0.94	1.55	1.86	**2.45**

Table 4
SSIM values for the test image in Figure 13 *for* $BSNR = 20, 15, 10$ *dB.*

BSNR	TV-L$_2$	TV$_p$-L$_2$	TV$_p^{\mathrm{sv}}$-L$_2$	DTV$_p^{\mathrm{sv}}$-L$_2$
20	0.78	0.79	0.80	**0.81**
15	0.76	0.77	0.78	**0.79**
10	0.70	0.72	0.74	**0.76**

numerical test above, few preliminary iterations of TV-L$_2$ are performed before computing the parameter maps. The research interval for the p parameter has been set equal to $[0.1, 2]$. It is worth remarking that the very small neighborhood size used for the parameter estimation is the one yielding the best restoration results for this test. We believe that this is motivated by the very fine scale of details in the test image. In Tables 3 and 4, the ISNR and SSIM values achieved by the TV-L$_2$, the TV$_p$-L$_2$ (with estimated global $p = 0.7$), the TV$_{\alpha,p}^{\mathrm{sv}}$-L$_2$ (with space-variant parameters estimated as in [29]), and the DTV$_p^{\mathrm{sv}}$-L$_2$ models for different values of BSNR are reported. Also in this case, the proposed model outperforms the competing ones. Note that the improvement is actually more significant in correspondence of higher noise levels. In Figure 14, a visual comparison between the reconstructions obtained by the different models for BSNR= 10 dB is proposed.

9. Conclusions and outlook. We presented a new space-variant anisotropic image regularization term for image restoration problems based on the statistical assumption that the gradients of the target image are distributed locally according to a BGGD. This leads to a highly flexible regularizer characterized by four per-pixel free parameters. For their automatic and effective selection, we propose a neighborhood-based estimation procedure relying on the ML approach. We empirically show the good asymptotic properties of the estimator and its consistency with the geometric intuition about the behavior of the BGGD in various image regions (edges, corners, and homogeneous areas). In terms of such parameters, we then study the corresponding space-variant and directional energy functional and apply it to the problem of image restoration in the case of AWGN. Numerically, the restored image is computed efficiently by means of an iterative algorithm based on ADMM. The proposed regularizer is shown to outperform other space-variant restoration models and it is shown to achieve high quality restoration results, even when dealing with high levels of blur and noise. The directional feature of the regularization considered results, in particular, in a better preservation of texture and details.

Future research directions include, first, the design of numerical algorithms other than ADMM with proved convergence properties also in the nonconvex case such as, e.g., some

(a) TV-L$_2$.   (b) TV$_p$-L$_2$.   (c) TV$^{\mathrm{sv}}_{\alpha,p}$-L$_2$.   (d) DTV$^{\mathrm{sv}}_p$-L$_2$.

(e) Zoom of 14(a).   (f) Zoom of 14(b).   (g) Zoom of 14(c).   (h) Zoom of 14(d).
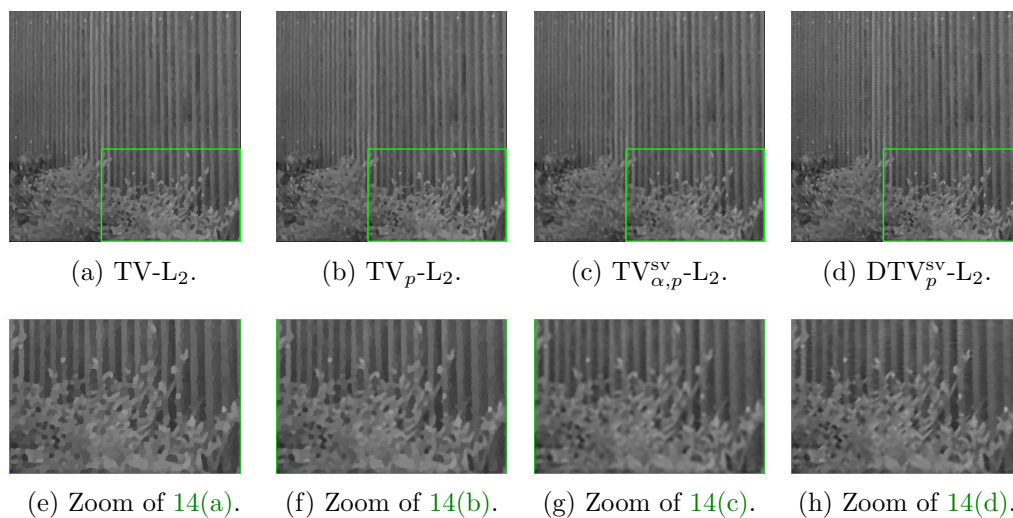
Figure 14. *Detail of reconstruction of natural test image Figure* 13. *Texture components are much better preserved by encoding directional information.*

suitable adaptation of the generalized Krylov subspace approaches proposed in [31, 24]. Then, automatic selection from the observed image of the "optimal" neighborood size for the preliminary parameter estimation step is a matter worthy to be investigated. Finally, it would be very interesting to couple the proposed regularizer with other data fidelity terms, so as to deal with noises other than additive Gaussian.

REFERENCES

[1] L. Ambrosio, N. Fusco, and D. Pallara, *Functions of bounded variation and free discontinuity problems*, Oxford University Press, Oxford, 2000.

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

[3] M. Benning and M. Burger, *Error estimates for general fidelities*, Electron. Trans. Numer. Anal., 38 (2011), pp. 44–68, http://etna.mcs.kent.edu/vol.38.2011/pp44-68.dir/pp44-68.pdf.

[4] P. Blomgren, T. F. Chan, P. Mulet, and C. K. Wong, *Total variation image restoration: Numerical methods and extensions*, in Proceedings of International Conference on Image Processing, Vol. 3, IEEE Computer Society, Los Alamitos, CA, 1997, pp. 384–387 https://doi.org/10.1109/ICIP.1997.632128.

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

[6] Z. Boukouvalas, S. Said, L. Bombrun, Y. Berthoumieu, and T. Adali, *A new Riemannian averaged fixed-point algorithm for MGGD parameter estimation*, IEEE Signal Process. Lett., 22 (2015), pp. 2314–2318, https://doi.org/10.1109/LSP.2015.2478803.

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

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

[9] L. CALATRONI, C. CHUNG, J. C. DE LOS REYES, C.-B. SCHÖNLIEB, AND T. VALKONEN, *Bilevel approaches for learning of variational imaging models*, in Variational Methods in Imaging and Geometric Control, Radon Ser. Comput. Appl. Math. 18, De Gruyter, Berlin, 2017, 40.

[10] L. CALATRONI, J. C. DE LOS REYES, AND C.-B. SCHÖNLIEB, *Infimal convolution of data discrepancies for mixed noise removal*, SIAM J. Imaging Sci., 10 (2017), pp. 1196–1233, https://doi.org/10.1137/16M1101684.

[11] A. CHAMBOLLE AND P.-L. LIONS, *Image recovery via total variation minimization and related problems*, Numer. Math., 76 (1997), pp. 167–188, https://doi.org/10.1007/s002110050258.

[12] A. CHAMBOLLE AND T. POCK, *An introduction to continuous optimization for imaging*, Acta Numer., 25 (2016), pp. 161–319, https://doi.org/10.1017/S096249291600009X.

[13] Y. CHEN, S. LEVINE, AND M. RAO, *Variable exponent, linear growth functionals in image restoration*, SIAM J. Appl. Math., 66 (2006), pp. 1383–1406, https://doi.org/10.1137/050624522.

[14] R. CIAK, *Coercive Functions from a Topological Viewpoint and Properties of Minimizing Sets of Convex Functions Appearing in Image Restoration*, Doctoral thesis, Technische Universität Kaiserslautern, Kaiserslautern, Germany, 2015, http://nbn-resolving.de/urn:nbn:de:hbz:386-kluedo-41000.

[15] V. DUVAL, J.-F. AUJOL, AND Y. GOUSSEAU, *The TVL1 model: A geometric point of view*, Multiscale Model. Simul., 8 (2009), pp. 154–189, https://doi.org/10.1137/090757083.

[16] M. J. EHRHARDT AND M. M. BETCKE, *Multicontrast MRI reconstruction with structure-guided total variation*, SIAM J. Imaging Sci., 9 (2016), pp. 1084–1106, https://doi.org/10.1137/15M1047325.

[17] H. ENGL, M. HANKE, AND A. NEUBAUER, *Regularization of Inverse Problems*, Math. Appl., Springer, Dordrecht, The Netherlands, 2000.

[18] W. HARE AND C. SAGASTIZÁBAL, *Computing proximal points of nonconvex functions*, Math. Program., 116 (2009), pp. 221–258, https://doi.org/10.1007/s10107-007-0124-6.

[19] B. HE AND X. YUAN, *On the $O(1/n)$ convergence rate of the Douglas–Rachford alternating direction method*, SIAM J. Numer. Anal., 50 (2012), pp. 700–709, https://doi.org/10.1137/110836936.

[20] C. HE, C. HU, W. ZHANG, AND B. SHI, *A fast adaptive parameter estimation for total variation image restoration*, IEEE Trans. Image Process., 23 (2014), pp. 4954–4967, https://doi.org/10.1109/TIP.2014.2360133.

[21] M. HINTERMÜLLER, T. VALKONEN, AND T. WU, *Limiting aspects of nonconvex TV^ϕ models*, SIAM J. Imaging Sci., 8 (2015), pp. 2581–2621, https://doi.org/10.1137/141001457.

[22] M. HINTERMÜLLER AND T. WU, *Nonconvex TV^q-models in image restoration: Analysis and a trust-region regularization–based superlinearly convergent solver*, SIAM J. Imaging Sci., 6 (2013), pp. 1385–1415, https://doi.org/10.1137/110854746.

[23] M. HONG, Z.-Q. LUO, AND M. RAZAVIYAYN, *Convergence analysis of alternating direction method of multipliers for a family of nonconvex problems*, SIAM J. Optim., 26 (2016), pp. 337–364, https://doi.org/10.1137/140990309.

[24] G. HUANG, A. LANZA, S. MORIGI, L. REICHEL, AND F. SGALLARI, *Majorization–minimization generalized Krylov subspace methods for ℓ_p-ℓ_q optimization applied to image restoration*, BIT, 57 (2017), pp. 351–378, https://doi.org/10.1007/s10543-016-0643-8.

[25] J. HUANG AND D. MUMFORD, *Statistics of natural images and models*, in Proceedings. 1999 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Vol. 1, IEEE Computer Society, Los Alamitos, CA, 1999, pp. 541–547, https://doi.org/10.1109/CVPR.1999.786990.

[26] R. KONGSKOV AND Y. DONG, *Directional total generalized variation regularization for impulse noise removal*, in Scale Space and Variational Methods in Computer Vision, F. Lauze, Y. Dong, and A. B. Dahl, eds., 2017, Springer, Cham, Switzerland, pp. 221–231, https://doi.org/10.1007/978-3-319-58771-4_18.

[27] R. D. KONGSKOV, Y. DONG, AND K. KNUDSEN, *Directional total generalized variation regularization*, BIT, to appear, https://doi.org/10.1007/s10543-019-00755-6.

[28] A. LANGER, *Automated parameter selection for total variation minimization in image restoration*, J. Math. Imaging Vision, 57 (2017), pp. 239–268, https://doi.org/10.1007/s10851-016-0676-2.

[29] A. LANZA, S. MORIGI, M. PRAGLIOLA, AND F. SGALLARI, *Space-variant generalised Gaussian regularisation for image restoration*, Comput. Methods Biomech. Biomed. Eng. Imaging Vis., to appear, https://doi.org/10.1080/21681163.2018.1471620.

[30] A. Lanza, S. Morigi, M. Pragliola, and F. Sgallari, *Space-variant TV regularization for image restoration*, in VipIMAGE 2017, J. M. R. Tavares and R. Natal Jorge, eds., Springer, Cham, Switzerland, 2018, pp. 160–169, https://doi.org/10.1007/978-3-319-68195-5_17.

[31] A. Lanza, S. Morigi, L. Reichel, and F. Sgallari, *A generalized Krylov subspace method for ℓ_p-ℓ_q minimization*, SIAM J. Sci. Comput., 37 (2015), pp. S30–S50, https://doi.org/10.1137/140967982.

[32] A. Lanza, S. Morigi, and F. Sgallari, *Constrained TV_p-ℓ^2 model for image restoration*, J. Sci. Comput., 68 (2016), pp. 64–91, https://doi.org/10.1007/s10915-015-0129-x.

[33] A. Lanza, S. Morigi, F. Sgallari, and Y.-W. Wen, *Image restoration with Poisson-Gaussian mixed noise*, Comput. Methods Biomech. Biomed. Eng. Imaging Vis., 2 (2014), pp. 12–24, https://doi.org/10.1080/21681163.2013.811039.

[34] F. Li, Z. Li, and L. Pi, *Variable exponent functionals in image restoration*, Appl. Math. Comput., 216 (2010), pp. 870–882, https://doi.org/10.1016/j.amc.2010.01.094.

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

[36] M. Nikolova, M. K. Ng, and C. P. Tam, *Fast nonconvex nonsmooth minimization methods for image restoration and reconstruction*, IEEE Trans. Image Process., 19 (2010), pp. 3073–3088, https://doi.org/10.1109/TIP.2010.2052275.

[37] F. Pascal, L. Bombrun, J. Tourneret, and Y. Berthoumieu, *Parameter estimation for multivariate generalized Gaussian distributions*, IEEE Trans. Signal Process., 61 (2013), pp. 5960–5971, https://doi.org/10.1109/TSP.2013.2282909.

[38] P. Peter, J. Weickert, A. Munk, T. Krivobokova, and H. Li, *Justifying tensor-driven diffusion from structure-adaptive statistics of natural images*, in Energy Minimization Methods in Computer Vision and Pattern Recognition, X.-C. Tai, E. Bae, T. F. Chan, and M. Lysaker, eds., Springer, Cham, Switzerland, 2015, pp. 263–277.

[39] P. Rodriguez, *Multiplicative updates algorithm to minimize the generalized total variation functional with a non-negativity constraint*, 2010 IEEE International Conference on Image Processing, IEEE, Piscataway, NJ, 2010, pp. 2509–2512, https://doi.org/10.1109/ICIP.2010.5654074.

[40] S. Roth and M. J. Black, *Fields of experts*, Int. J. Comput. Vis., 82 (2009), 205, https://doi.org/10.1007/s11263-008-0197-6.

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

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

[43] H. Scharr, M. J. Black, and H. W. Haussecker, *Image statistics and anisotropic diffusion*, in Proceedings Ninth IEEE International Conference on Computer Vision, Vol. 2, IEEE Computer Society, Los Alamitos, CA, 2003, pp. 840–847, https://doi.org/10.1109/ICCV.2003.1238435.

[44] F. Sciacchitano, Y. Dong, and T. Zeng, *Variational approach for restoring blurred images with Cauchy noise*, SIAM J. Imaging Sci., 8 (2015), pp. 1894–1922, https://doi.org/10.1137/140997816.

[45] K. Sharifi and A. Leon-Garcia, *Estimation of shape parameter for generalized Gaussian distributions in subband decompositions of video*, IEEE Trans. Circuits Syst. Video Tech., 5 (1995), pp. 52–56, https://doi.org/10.1109/76.350779.

[46] K.-S. Song, *A globally convergent and consistent method for estimating the shape parameter of a generalized Gaussian distribution*, IEEE Trans. Inform. Theory, 52 (2006), pp. 510–527, https://doi.org/10.1109/TIT.2005.860423.

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

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

[49] L. Vese, *A study in the BV space of a denoising–deblurring variational problem*, Appl. Math. Optim., 44 (2001), pp. 131–161, https://doi.org/10.1007/s00245-001-0017-7.

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

[51] J. WEICKERT, *Anisotropic Diffusion in Image Processing*, Teubner, Stuttgart, 1998, https://www.mia.uni-saarland.de/weickert/Papers/book.pdf.

[52] J. WEICKERT AND H. SCHARR, *A scheme for coherence-enhancing diffusion filtering with optimized rotation invariance*, J. Vis. Commun. Image Represent., 13 (2002), pp. 103–118, https://doi.org/10.1006/jvci.2001.0495.

[53] J. WEICKERT AND T. BROX, *Diffusion and regularization of vector- and matrix-valued images*, in Inverse Problems, Image Analysis, and Medical Imaging, AMS, Providence, RI, 2002, pp. 251–268, http://lmb.informatik.uni-freiburg.de/Publications/2002/Bro02a.

[54] Y. W. WEN AND R. H. CHAN, *Parameter selection for total-variation-based image restoration using discrepancy principle*, IEEE Trans. Image Process., 21 (2012), pp. 1770–1781, https://doi.org/10.1109/TIP.2011.2181401.

[55] H. ZHANG AND Y. WANG, *Edge adaptive directional total variation*, J. Eng., 2013 (2013), pp. 61–62, https://doi.org/10.1049/joe.2013.0116.

[56] W. ZHOU, A. BOVIK, H. SHEIKH, AND E. SIMONCELLI, *Image quality assessment: From error visibility to structural similarity.*, IEEE Trans. Image Process., 13 (2004), https://doi.org/10.1109/TIP.2003.819861.

[57] S. C. ZHU, Y. WU, AND D. MUMFORD, *Filters, random fields and maximum entropy (FRAME): Towards a unified theory for texture modeling*, Int. J. Comput. Vis., 27 (1998), pp. 107–126, https://doi.org/10.1023/A:1007925832420.