SIAM J. IMAGING SCIENCES
Vol. 14, No. 4, pp. 1749–1789

A Variational Approach to Additive Image Decomposition into Structure, Harmonic, and Oscillatory Components[*]

Martin Huska[†], Sung H. Kang[‡], Alessandro Lanza[†], and Serena Morigi[†]

Abstract. We propose a nonconvex variational decomposition model which separates a given image into piecewise-constant, smooth, and oscillatory components. This decomposition is motivated not only by image denoising and structure separation, but also by shadow and spot light removal. The proposed model clearly separates the piecewise-constant structure and smoothly varying harmonic part, thanks to having a separated oscillatory component. The piecewise-constant part is captured by TV-like nonconvex regularization, harmonic term via second-order regularization, and oscillatory (noise and texture) term via a H^{-1}-norm penalty. There are interesting interactions between these three regularization terms. We explore the effects of each regularization and the choice of parameters carefully. We propose an efficient alternating direction method of multipliers based minimization for fast numerical computation of the optimization problem. Various experiments are presented to show the robustness against a high level of noise, applications to soft spotlight and shadow removal, and the comparisons with other methods.

Key words. image decomposition, oscillatory-structure cartoon, convex nonconvex optimization

AMS subject classifications. 65K10, 65D18, 68U10

DOI. 10.1137/20M1355987

1. Introduction.

The task of decomposing images into their semantically different contents is of great interest in various image processing methods, such as image restoration, compression, segmentation, and object recognition. It can simplify the characteristics of the image to achieve better results for different imaging tasks. For example, in [9], the proposed model decomposes the given image into a piecewise-constant part and a harmonic part. This method has the advantage that the structured part, the piecewise-constant component, is separated extremely cleanly without any noise. However, even for low levels of noise, the harmonic part captures the general smooth field but including the noise, which eventually corrupts the image. The seminal work of Meyer and Lewis [20], where the G-norm function space was introduced, proposed to separate the image into a geometric part, e.g., using TV denoising [24], and oscillatory texture or noisy part. Various work has followed [2, 3, 4, 5, 14, 17, 23]. In [23], for example, the G-norm is approximated by a negative Sobolev norm of H^{-1}.

In this work, we propose an additive image decomposition model, which separates an observed image into a cartoon/structure component, a smooth part, and an oscillatory term.

---

[*] Received by the editors July 28, 2020; accepted for publication (in revised form) August 25, 2021; published electronically November 23, 2021.

https://doi.org/10.1137/20M1355987

Funding: Research by the first, third, and fourth authors was supported in part by the National Group for Scientific Computation (GNCS-INDAM), Research Projects 2021.

[†] Department of Mathematics, University of Bologna, 40126, Bologna, Italy (martin.huska@unibo.it, alessandro.lanza2@unibo.it, serena.morigi@unibo.it).

[‡] Mathematics, Georgia Institute of Technology, Atlanta, GA 30332-0160 USA (kang@math.gatech.edu).

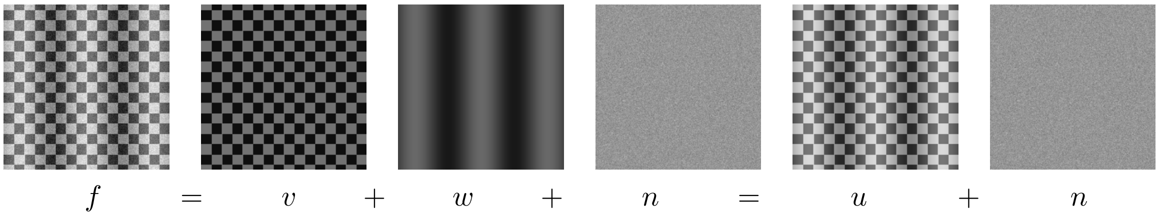


Figure 1.1. *A noisy image f decomposed following model* (1.1).

The structured part models homogeneous regions with sharp edges, the smooth part presents smoothly varying intensity characterized by a small high-derivative norm, and an oscillatory term is modeled by the use of an H^{-1}-norm penalty. We assume that the observed image f in vectorized form, $f \in \mathbb{R}^m$ of dimension $m = m_1 \times m_2$, is representable as the sum of a piecewise-constant component v, a smooth component w, and an oscillatory component n, such that

$$f = v + w + n = u + n\,, \tag{1.1}$$

where $u := v + w$ represents the underlying noise-/texture-free piecewise-smooth image. In case of noise, the oscillatory term $n \in \mathbb{R}^m$ in (1.1) is assumed to be independent and identically distributed (i.i.d.) additive white Gaussian noise, i.e., $n \sim Gauss(0, \sigma^2)$, with known variance σ^2. Figure 1.1 illustrates the objective of the separation task, where f is a balanced sum of v and w and a realization of Gaussian noise characterized by $\sigma = 10$. The v is the shadow-free image. The gradient magnitude of v is obviously sparse, and the first and second derivatives of w are very small in magnitude.

Typical image decomposition models separate the image into two components. Levine in [16] proposed an adaptive two-component decomposition model based on the edge detection of the input image, using the Huber penalty for the structure component and the L_p norm for the texture part. The authors in [14] used bounded mean oscillations space to model the oscillating patterns (texture). The model in [2] is extended to color images in [5], providing a novel way to solve the G-norm-related subproblem, since the projection algorithm proposed in [2] cannot be used for color images. The work [17] extended the H^{-1} space to the whole family of models using H^{-s} space. Later on in [29] the authors extended the TV-G decomposition model to manifold-valued images using the L_p approximation to the G-norm. A rather different additive two-component decomposition on manifolds into piecewise-constant (structure) and smooth components is presented in [9], while the additive white noise is handled via the L_2 residual. Recently, in [18] a nonconvex nonsmooth structure-texture decomposition model has been presented, where the structure component is modeled via a TV-like nonconvex penalty, while the H^{-1} space is used to approximate the G-norm for texture or noise separation. In [3], the authors proposed a three decomposition variational model of structure, texture, and noise, where the latter is captured by the negative Besov norm.

In addition to texture/noise removal, this work is closely related to retinex theory and shadow removal. Retinex is a theory simulating how human perception of image intensity gets distorted under certain uneven lighting conditions, i.e., the color-preservation property of the human vision system [21]. Shadow removal and retinex both have similar difficulties in the way that there is a large region of smooth gradient change independent of any edges and details of the image. The proposed method efficiently separates the shadow and soft light effects.

The main contributions of this work are summarized as follows:

- We propose an effective three-part additive decomposition nonconvex variational model to separate an image into piecewise-constant, smooth, and oscillatory components. This shows an advantage over structure separation even for a high level of noise, and shadow and spotlight removal.
- We characterize our proposed model in terms of convexity and coercivity, and we provide existence results for the nonconvexity, noncoercive variational problem.
- We develop a fast and efficient algorithm for this model, based on an alternating direction method of multipliers (ADMM)-based minimization algorithm, and also outline a simple strategy for an effective selection of all the parameters in the model and in the proposed ADMM algorithm.

This paper is organized in the following way. In section 2, we propose a new three-term variational model for image decomposition. Each of the three penalty terms is carefully considered in each subsection: subsection 2.1 details a TV-like nonconvex term for a piecewise-constant component, subsection 2.2 shows the new harmonic term, and subsection 2.3 investigates the statistical properties of the H^{-1}-norm approximation for Gaussian noise and the proposed H^{-1}-norm based regularization term. In section 3, insights on the effect of different penalty terms are presented, and a careful discussion of the parameter selection is discussed in section 4. In section 5, we discuss the existence of global minimizers for the proposed variational problem, and we describe an efficient ADMM-based numerical method for its solution. Various numerical examples are illustrated in section 6, which show an advantage over a high level of noise, and shadow and noise removal. Comparisons with other methods are presented. Conclusions are drawn in section 7.

2. The proposed nonconvex variational decomposition model.

We propose a model which separates a given image into piecewise-constant, smooth homogeneous, and oscillatory components. For this objective, we look into the following model problem, which decomposes a given image f into three components v, w, n having distinct features,

$$\{v^*, w^*, n^*\} \in \arg\min_{v,w,n} \left\{ \gamma \left(\|\nabla v\|_0 + \|\mathcal{H} w\|_2^2 + \|n\|_G \right) + \frac{1}{2}\|f - (v + w + n)\|_2^2 \right\}, \tag{2.1}$$

where the given image $f \in \mathbb{R}^m$ gets decomposed into three ideal components v^*, w^*, n^*. The ℓ_0 pseudonorm $\|x\|_0$ counts the nonzero components of a vector x to induce sparsity. The operator $\mathcal{H}(\cdot)$ represents a derivative of order higher than one, in particular, we use second-order derivative in this paper, and $\|\cdot\|_2^2$ represents the square of the l_2 norm. $\|\cdot\|_G$ is the G-norm that models the oscillatory component. The scalar value $\gamma > 0$ represents the regularization parameter balancing the regularization terms and the fidelity term.

The ℓ_0 pseudonorm of ∇v in (2.1) forces the gradient to be sparse encouraging the recovery of piecewise-constant components, but its combinatorial nature makes the minimization of (2.1) an NP-hard problem. Thus, it can be approximated by the sum of function values

$\phi_1(\cdot;a)$ as follows:

$$\|\nabla v\|_0 := \#\left\{i \,|\, |(\nabla v)_i| \neq 0, i = 1,\ldots,m\right\} \approx \sum_{i=1}^{m} \phi_1\left(|(\nabla v)_i|;a\right) := \mathcal{R}_1(v), \tag{2.2}$$

where $\phi_1(\cdot;a) : [0,+\infty) \to [0,1]$ is a nonconvex sparsity promoting penalty function, and

$$|(\nabla v)_i| = \sqrt{|(\partial_h v)_i|^2 + |(\partial_v v)_i|^2}$$

represents the ith gradient magnitude in terms of the ℓ_2 norm. Detailed characteristics of $\phi_1(\cdot;a)$ of this term will be given in section 2.1. The parameter a allows us to tune the degree of nonconvexity of $\phi_1(\cdot;a)$, such that $\phi_1(\cdot;a)$ tends to the ℓ_0 pseudonorm for $a \to \infty$. Compared with the classical convex TV prior, it promotes sparsity of gradient norms of the cartoon component more strongly, while better preserving sharp discontinuities [10, 11].

For the recovery of smooth functions, the use of TV-like penalties is not appropriate, since fine scale details are lost and smoothly varying features produce staircase effects. Therefore, we consider a second-order differential operator $\mathcal{H}(\cdot)$ in (2.1), and we define $\mathcal{R}_2(w)$ as

$$\mathcal{R}_2(w) := \sum_{j=1}^{m} \phi_2(|(\mathcal{H}w)_j|) = \sum_{j=1}^{m} |(\mathcal{H}w)_j|^2, \tag{2.3}$$

where $\phi_2 : [0,\infty) \to [0,\infty)$, $\phi_2(t) = t^2$. This induces w to be a smooth component with relatively small first- and second-order derivatives. We penalize deviations from a piecewise-constant model by constructing $\mathcal{R}_1(v)$ with a gradient operator, while to penalize model roughness or bumpiness (curvature), we use in $\mathcal{R}_2(w)$ a second-difference operator. Using this higher-order operator yields second-order Tikhonov regularization which favors “smooth” solutions.We present further details in subsection 2.2.

A usual convention through the Meyer’s space approach, introduced in [20], is to define the highly oscillating component n in terms of a vector field g, such that $n = \nabla \cdot g$, where an appropriate space is chosen for the field components $g = (g^h, g^v)$. In [3], the discrete version of the Meyer’s space G is introduced as

$$G = \left\{n \in \mathbb{R}^{m_1\times m_2} \;\middle|\; \exists g \in \mathbb{R}^{m_1\times m_2} \times \mathbb{R}^{m_1\times m_2} \text{ s.t. } n = \nabla \cdot g\right\},$$

which leads to the G-norm of n defined as

$$\|n\|_G = \inf\left\{\|g\|_\infty \;\middle|\; n = \nabla \cdot g,\; g = (g^h, g^v) \in \mathbb{R}^{m_1\times m_2} \times \mathbb{R}^{m_1\times m_2}\right\}; \tag{2.4}$$

here $\|g\|_\infty = \max_{i,j} |g_{i,j}|$, where $|g_{i,j}| = \sqrt{(g^h_{i,j})^2 + (g^v_{i,j})^2}$ for $g_{i,j}$ being the (i,j)th couple of the vector field g. The space G is a very good space to model oscillating patterns such as texture as well as noise, characterized by functions of zero mean, which attain a small norm in G space [3]. In order to overcome the computational difficulties derived by working with $\|n\|_G$, the authors in [28] proposed to replace the space G with $G_p = W^{-1,p}$ with $1 \leq p < +\infty$, furthermore simplified for $p = 2$, as proposed in [23]. In this case, the space G_2 is actually replaced by the negative Sobolev space H^{-1} which is the dual space of H_0^1, and is endowed with the following seminorm,

$$\|n\|_{H^{-1}} = \inf \left\{ \|g\|_2 = \sqrt{\sum_{i,j} |g_{i,j}|^2} \quad \middle| \; n = \nabla \cdot g \right\} , \tag{2.5}$$

which is proved to be bounded for oscillatory functions [23]. In this paper, we adopt the norm in (2.5) to model the oscillatory component, due to its potential to allow easier to handle regularizers w.r.t. the classical G-norm. Let g be defined in the image vectorized form as $g \in \mathbb{R}^{2m}$, $g := (g_1^h, g_2^h, \ldots, g_m^h, g_1^v, g_2^v, \ldots, g_m^v)^T$. We define the penalty $\phi_3 : [0, \infty) \to [0, \infty)$ to be

$$\mathcal{R}_3(g) := \phi_3 \left(\sum_{j=1}^{m} ((g_j^h)^2 + (g_j^v)^2) \right), \quad \text{where} \quad \phi_3(t) := t^2. \tag{2.6}$$

In summary, combining the approximations (2.2), (2.3), and (2.6) with a data fitting term $\mathcal{F}(v, w, g)$ in the variational model (2.1), and introducing three parameters $\gamma_1, \gamma_2, \gamma_3 \in \mathbb{R}_{++}$ with $\mathbb{R}_{++} = (0, \infty)$, to balance the regularization terms, we propose the following minimization model

$$\{v^*, w^*, g^*\} \in \arg \min_{v,w \in \mathbb{R}^m, g \in \mathbb{R}^{2m}} \mathcal{J}(v, w, g), \tag{2.7}$$

where

$$\begin{aligned}
\mathcal{J}(v, w, g) &:= \gamma_1 \mathcal{R}_1(v) + \frac{\gamma_2}{2} \mathcal{R}_2(w) + \frac{\gamma_3}{2} \mathcal{R}_3(g) + \frac{1}{2} \mathcal{F}(v, w, g) \\
&:= \gamma_1 \sum_{j=1}^{m} \phi_1(|(\nabla v)_j|; a) + \frac{\gamma_2}{2} \sum_{j=1}^{m} \phi_2(|(\mathcal{H} w)_j|) + \frac{\gamma_3}{2} \phi_3 \left(\sum_{j=1}^{m} ((g_j^h)^2 + (g_j^v)^2) \right) \\
&\quad + \frac{1}{2} \sum_{j=1}^{m} [f_j - (v_j + w_j + (\nabla \cdot g)_j)]^2 .
\end{aligned}$$

Here, $\mathcal{R}_1(v)$ is nonconvex, $\mathcal{R}_2(w)$, $\mathcal{R}_3(g)$, and $\mathcal{F}(v, w, g)$ are smooth and convex functions, and the overall model (2.7) is a nonconvex optimization problem. In this work, we denote by $\mathbb{R}_+$ and R_{++} the sets of nonnegative and positive real numbers, respectively. An accurate choice of the parameters γ_1, γ_2, γ_3, affects the results. Figure 2.1(a) shows the graphs of the penalty functions $\phi_1(t; a)$ in $\mathcal{R}_1(t)$ (solid line), $\phi_2(t)$ (dashed line) in $\mathcal{R}_2(t)$, and $\phi_3(t)$ (dot-dashed line) in $\mathcal{R}_3(t)$. As it will be discussed in sections 4 and 5, a good set of parameters, including a, allows for a suitable repositioning of the intersection points between these regularization terms to enforce the efficacy of the proposed decomposition model.

The rest of this section discusses each proposed regularizer, followed by the effects of different penalty terms in section 3, leading to a discussion on the parameter selection in section 4. In section 5 we point out some important characteristics of the cost functional (2.7) for its efficient minimization.

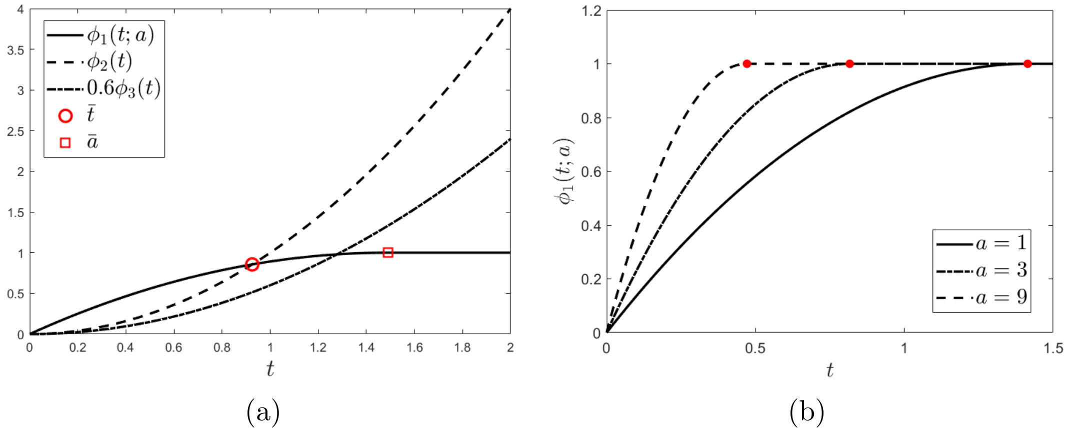


Figure 2.1. (a) *Graphs of* $\phi_1(t;a)$ *(solid line),* $\phi_2(t)$ *(dashed line), and* $\phi_3(t)$ *(dot-dashed line), used in* $\mathcal{R}_1(v)$*,* $\mathcal{R}_2(w)$*, and* $\mathcal{R}_3(g)$*, respectively. For illustration purposes, we plot the scaled version of* ϕ_3 *to underline how the intersection points between* ϕ_1*,* ϕ_2*, and* ϕ_3 *can be modified by tuning the regularization parameters (more details in section 4 in relation to (4.3)).* (b) *Plot of the regularization term* $\phi_1(t;a)$ *defined in (2.8) for different values of the concavity parameter* a*.*

2.1. $\mathcal{R}_1(v)$ penalty function.

The ℓ_0 pseudo-norm is often understood to be the ideal regularizer to induce sparsity. The alternative ℓ_1 norm is the convex relaxation of the ℓ_0 pseudo-norm, and plays a fundamental role in sparse image/signal processing. However, the ℓ_1 norm in sparsity-inducing regularizers can be considered, in general, to be overrelaxed. A substantial amount of recent work has argued for nonconvex regularizers in favor of their superior theoretical properties and excellent practical performances [22, 12, 25]. For the penalty $\mathcal{R}_1$ in (2.7), we require the following conditions for $\phi_1(\cdot;a)$, both for modeling and minimization algorithm:

(a) be nonconvex, such that the regularizer $\mathcal{R}_1$ promotes sparsity of the gradient magnitudes of the piecewise-constant component v more effectively than the classical isotropic TV regularizer;

(b) have the range in $[0,1]$ independent of the parameter a, so that the degree of nonconvexity represented by a can be freely tuned without affecting the ϕ_1 upper bound.

(c) have a form such that the associated multivariate proximity operator

$$\operatorname{prox}_{\phi}^{\alpha}(q) := \operatorname*{argmin}_{x\in\mathbb{R}^m} \left\{ \phi_1\big(\|x\|_2;a\big) + \frac{\alpha}{2}\,\|x-q\|_2^2 \right\}, \quad q\in\mathbb{R}^m,$$

admits a closed-form expression.

Following these requirements, we chose the regularizer to be a reparameterized and rescaled version of the minimax concave (MC) penalty [30], namely, a simple piecewise quadratic function defined by

$$\phi_1(t;a) = \begin{cases} -\dfrac{a}{2}t^2 + \sqrt{2a}\,t & \text{for } t\in\left[0,\sqrt{2/a}\right), \\ 1 & \text{for } t\in\left[\sqrt{2/a},+\infty\right). \end{cases} \tag{2.8}$$

In Figure 2.1(b), we show the plot of the MC penalty functions $\phi_1(t;a)$ defined in (2.8) for three different values $a\in\{1,3,9\}$ of the concavity parameter. The solid dots on the graphs represent the points $\big(\sqrt{2/a}, \phi_1\big(\sqrt{2/a}; a\big)\big)$ which separate the nonconvex quadratic piece of the penalty from the constant one.

The MC penalty satisfies the following properties:

$$\phi_1(t;a) \;\in\; C^1\big([0,\infty)\big) \;\cap\; C^\infty\big([0,\infty) \setminus \{\sqrt{2/a}\,\}\big)\,,$$

$$\phi_1'(t;a) = \begin{cases} \sqrt{2a} - at & \text{for } t \in \big[0, \sqrt{2/a}\big), \\ 0 & \text{for } t \in \big[\sqrt{2/a}, \infty\big), \end{cases} \qquad \phi_1''(t;a) = \begin{cases} -a & \text{for } t \in \big[0, \sqrt{2/a}\big), \\ 0 & \text{for } t \in \big(\sqrt{2/a}, \infty\big). \end{cases} \tag{2.9}$$

In particular, we notice that

$$a = \left| \min_t \phi_1''(t;a) \right| \,, \quad t \in [0, +\infty) \setminus \{\sqrt{2/a}\}\,,$$

such that the parameter a represents the degree of nonconvexity of ϕ_1 and, hence, can be referred to as the concavity parameter of ϕ_1. Note that for $a \to \infty$, the $\phi_1(\cdot;a)$ converges to the ℓ_0 pseudo-norm graph, which attains a constant 1 everywhere except at the origin, where it attains zero. This is stated by the following result together with the sparsity promoting property of $\mathcal{R}_1(v)$ inherited directly from the $\phi_1(\cdot;a)$ function.

Proposition 2.1. *Let $\phi_1(\cdot;a)$ be the function defined in* (2.9)*. Then for any vector $v \in \mathbb{R}^m$ we have*

$$\mathcal{R}_1(v) := \sum_j \phi_1(|(\nabla v)_j|; a) \le \|\nabla v\|_0. \tag{2.10}$$

Let $\mu := \min_{j:\,|(\nabla v)_j|>0} |(\nabla v)_j|$. If

$$\mu \ge \bar{a} := \sqrt{2/a}, \tag{2.11}$$

then the equality in (2.10) *holds.*

Proof. The result (2.10) follows from the fact that $0 \le \phi_1(\cdot;a) \le 1$. Moreover, from the properties of $\phi_1(\cdot;a)$ we have

$$\lim_{a\to\infty} \sum_j \phi_1(|(\nabla v)_j|; a) = \|\nabla v\|_0.$$

If the smallest nonvanishing gradient magnitude μ is greater than $\bar{a}$ defined in (2.11), then $\phi_1(|(\nabla v)_j|; a) = 1$ for every element j of nonzero gradient magnitude, thus equality holds. ■

We remark that (2.10) in Proposition 2.1 holds true for the vector $v \in \mathbb{R}^m$ itself, i.e., $\sum_j \phi_1(|v_j|; a) \le \|v\|_0$. However, we present the proposition in terms of the gradient to emphasize the "accountability" intensity changes for piecewise-constant parts.

The MC penalty in (2.8) provides a recognized alternative to any ℓ_p-norm based penalty, with p, $0 < p < 1$, and induces sparsity of the image gradient magnitudes more strongly than the ℓ_1 norm thus better favors piecewise-constant solutions.

2.2. $\mathcal{R}_2(w)$ penalty function.

The $\mathcal{R}_2(w)$ term is designed to capture the smooth component w. We promote sparsity of the second-order derivatives $\mathcal{H}(\cdot)$ and use $\phi_2(t) = t^2$. We define the regularizer $\mathcal{R}_2(\cdot)\colon \mathbb{R}^m \to [0,\infty)$ as follows:

(2.12)

$$\mathcal{R}_2(w) = \sum_{j=1}^{m} \phi_2\left(|(\mathcal{H}w)_j|\right) = \sum_{j=1}^{m} \left[(\mathcal{H}_{hh}w)_j^2 + (\mathcal{H}_{hv}w)_j^2 + (\mathcal{H}_{vh}w)_j^2 + (\mathcal{H}_{vv}w)_j^2\right] = \sum_{j=1}^{m} |(\mathcal{H}w)_j|^2.$$

The linear second-order operator $\mathcal{H}(\cdot)$ is applied to each pixel $j = 1,\dots,m$ of the image $w \in \mathbb{R}^m$, which results in a vector of second-order derivatives

$$(\mathcal{H}w)_j = ((\mathcal{H}_{hh}w)_j, (\mathcal{H}_{hv}w)_j, (\mathcal{H}_{vh}w)_j, (\mathcal{H}_{vv}w)_j)^T$$

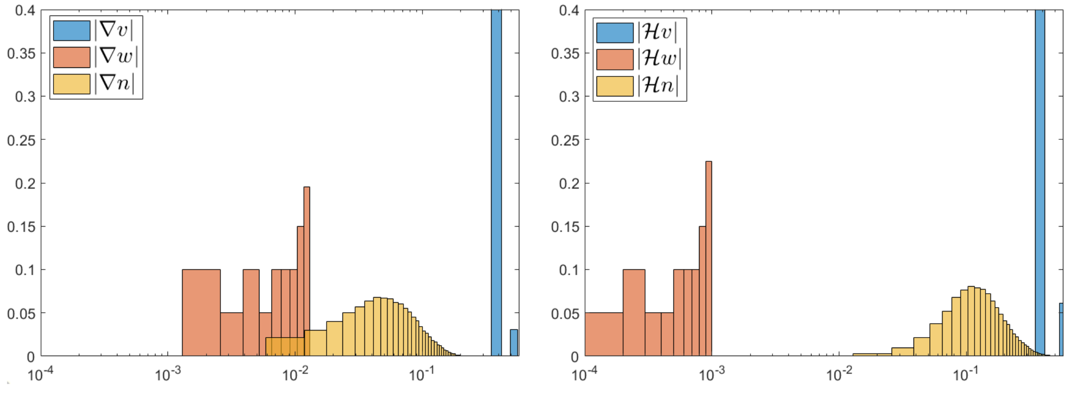
along the principal horizontal (h) and vertical (v) directions and their mixture. This regularizer $\mathcal{R}_2(\cdot)$ is an extension to second-order Tikhonov regularization, which extends the neighborhood Laplacian discretization by adding diagonal directions to the conventional horizontal-vertical Laplacian discretization.

The variational decomposition model in [9], which decomposes the image into piecewise-constant and smooth parts, uses the gradient norm to capture smooth functions, and in [8] a combination of first- and second-order derivatives is applied for reconstruction of piecewise-smooth signals. Another common way to include second-order derivatives as energy penalties is to consider the Hessian Schatten-norm regularizer, as proposed in [15].

In this paper, the proposed term $\mathcal{R}_2(w)$ in (2.12) which not only favors the smoothness of the image, but also the use of linear operators, allowed us to simplify the computation for w via normal equations, as detailed in section 5. Moreover, oscillations that usually characterize Gaussian noise present smaller magnitudes of $|\nabla(\cdot)|$ with respect to $|\mathcal{H}(\cdot)|$, unlike what happens for smooth regions. This motivates us to exploit the second-order derivatives to capture the pure smooth w image component and to increase the components' separation.

An insight into this separation is illustrated in Figure 2.2, where (a) is the histogram of the gradient magnitudes $|(\nabla(\cdot))_j|$ for every pixel $j = 1,\dots,m$, and (b) is the histogram of the second-order derivative magnitudes $|(\mathcal{H}(\cdot))_j|$ for the components v, w, and n of the image illustrated in Figure 1.1; here n represents a noise component. We notice that the minimum nonzero magnitude $|\mathcal{H}v|$ attains the same value as $|\nabla v|$ (where $\min_{j\colon |(\nabla v)_j|>0} |(\nabla v)_j| = 0.39$), and the maximum magnitude $|\mathcal{H}w|$ attains the value 0.001 that is approximately one order smaller than the maximum magnitude of $|\nabla w|$ (where $\max_j |(\nabla w)_j| = 0.012$). Therefore, the use of the second-order-based operator $\mathcal{H}(\cdot)$, instead of the gradient-based operator, $\nabla(\cdot)$, allows for a larger separation distance between w and v. For the oscillatory (noise) component n, the nonzero gradient magnitudes range in $[0.006, 0.29]$, while for the nonzero second-order derivatives magnitudes $|\mathcal{H}n| \in [0.01, 0.64]$. Again, using the $\mathcal{H}(\cdot)$ operator gives a larger separation distance between w and n components.

We note that other regularization terms can be utilized for each $\mathcal{R}_1(v)$, $\mathcal{R}_2(w)$, or $\mathcal{R}_3(g)$, as long as each regularization gives good separations, as illustrated in Figure 2.2. We chose simple terms, e.g., a linear operator for $\mathcal{R}_2(w)$ and an l_2-norm for $\mathcal{R}_3(g)$, for simpler and faster computation.



(a) (b)

Figure 2.2. *Effect of $\mathcal{R}_2(w)$: The components v, w, and n are from Figure 1.1.* (a) *Histograms of $|\nabla v|$, $|\nabla w|$, and $|\nabla n|$;* (b) *Histograms of $|\mathcal{H}w|$, $|\mathcal{H}v|$, and $|\mathcal{H}n|$. Logarithmic scale in the x-axis is applied. Notice a larger separation distance between w and v, as well as between w and n, when the new proposed $\mathcal{H}$ (second-order derivative) is used.*

2.3. $\mathcal{R}_3(g)$ penalty function and statistical characterization of the H^{-1}-norm for Gaussian noise. In this section, we explore a faithful bound estimation of the H^{-1}-norm approximation value $\|g\|_2$ of $\|n\|_{H^{-1}}$, as well as of the associated regularization term $\mathcal{R}_3(g)$. This is under the assumption that n is white additive Gaussian with known standard deviation σ.

The additive noise is modeled as an $m_1 \times m_2$ discrete random process $\mathcal{N} := \{N[i,j] : [i,j] \in \Omega\}$ (Ω is the image domain) with $N[i,j]$ denoting the scalar random variable modeling noise at pixel $[i,j]$; we assume $N[i,j]$ is zero-mean Gaussian with variance σ^2. Let $\mathcal{G}$ be an $m_1 \times m_2$ discrete random process $\mathcal{G} := \{G[i,j] = (G^h[i,j], G^v[i,j]) : [i,j] \in \Omega\}$ with $G[i,j]$ denoting the 2-dimensional vector random variable at pixel $[i,j]$ with bivariate Gaussian distribution having variance σ_g^2, that is,

$$G[i,j] \sim Gauss\left(\left[\begin{array}{c} 0 \\ 0 \end{array} \right], \left[\begin{array}{cc} \sigma_g^2 & 0 \\ 0 & \sigma_g^2 \end{array} \right] \right) . \tag{2.13}$$

We assume that the two components $G^h[i,j]$ and $G^v[i,j]$ are uncorrelated as indicated by the diagonal covariance matrix in (2.13), while the $G[i,j]$'s are identically distributed and correlated. Given a single realization $n := \{n[i,j] \in \mathbb{R} : [i,j] \in \Omega\} \in \mathbb{R}^{m_1 \times m_2}$ of the noise process $\mathcal{N}$, and a single realization $g := \{g[i,j] = (g^h[i,j], g^v[i,j]) \in \mathbb{R}^2 : [i,j] \in \Omega\} \in \mathbb{R}^{m_1 \times m_2}$ of the process $\mathcal{G}$, then the statistical representation of the relation $n = \nabla \cdot g$ reads as

$$n[i,j] = D_h g^h[i,j] + D_v g^v[i,j] \qquad \forall [i,j] \in \Omega \tag{2.14}$$

with D_h, D_v linear transformations, explicitly represented by

$$n[i,j] = \frac{1}{\delta} \Big(g^h[i+1,j] - g^h[i,j] + g^v[i,j+1] - g^v[i,j] \Big) \qquad \forall [i,j] \in \Omega \,, \tag{2.15}$$

with $\delta \in \mathbb{R}_+$ being the discretization step. However, since here we want to analyze the statistical characterization of the H^{-1}-norm we clearly will replace the realized, deterministic values $n[i,j]$ and $g[i,j]$ with the associated random variables $N[i,j]$ and $G[i,j]$ when dealing with relationships (2.14) and (2.15). The main result of the relation between the two discrete probabilistic distributions followed by $\mathcal{G}$ and $\mathcal{N}$ is summarized in Proposition 2.2, whereas results on the distribution of $\mathcal{R}_3(\mathcal{G})$ are presented in Proposition 2.3.

Proposition 2.2. *Let $\mathcal{N}$ be an $m_1 \times m_2$ i.i.d. Gaussian noise process with variance σ^2, and $\mathcal{G}$ be an $m_1 \times m_2$ identically distributed and correlated random process with bivariate Gaussian distribution having variance σ_g^2, which are related by (2.15). Then*

$$\sigma_g^2 = \frac{\delta^2 \sigma^2}{4(1+\rho_d)(1-\rho_g)} , \tag{2.16}$$

where ρ_g and ρ_d are scalar values representing Pearson's correlation coefficients between $G^h[i+1,j]$ and $G^h[i,j]$ as well as between $G^v[i,j+1]$ and $G^v[i,j]$, and between $D_hG^h[i,j]$ and $D_vG^v[i,j]$, respectively.

Proof. By applying the sum rule for normal variables to (2.15), we have

$$\frac{G^h[i+1,j]}{\delta}, \ \frac{G^h[i,j]}{\delta} \sim Gauss\left(0, \frac{\sigma_g^2}{\delta^2}\right) \implies D_hG^h[i,j] \sim Gauss(0,\sigma_d^2) ,$$

$$\frac{G^v[i,j+1]}{\delta}, \ \frac{G^v[i,j]}{\delta} \sim Gauss\left(0, \frac{\sigma_g^2}{\delta^2}\right) \implies D_vG^v[i,j] \sim Gauss(0,\sigma_d^2) ,$$

and σ_d^2 is related to σ_g^2 by the following

$$\sigma_d^2 = \frac{2\sigma_g^2}{\delta^2}(1-\rho_g) , \tag{2.17}$$

where ρ_g denotes Pearson's correlation coefficient between $G^h[i+1,j]$ and $G^h[i,j]$ as well as between $G^v[i,j+1]$ and $G^v[i,j]$. Applying again the sum rule to (2.14), we state

$$N[i,j] \sim Gauss(0,\sigma^2) \quad \text{with} \quad \sigma^2 = 2\sigma_d^2(1+\rho_d), \tag{2.18}$$

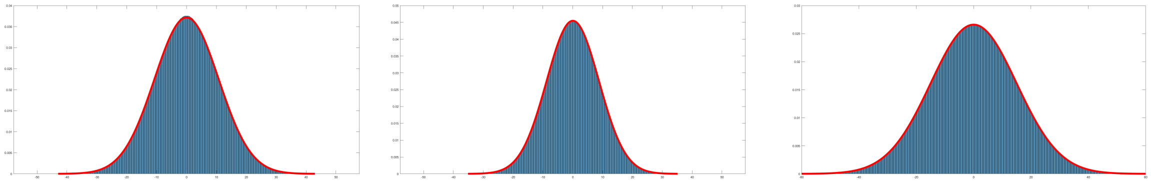
where ρ_d is Pearson's correlation coefficient between $D_hG^h[i,j]$ and $D_vG^v[i,j]$. Replacing (2.17) into (2.18), the relation in (2.16) between σ_g and σ follows. ■

In Proposition 2.2 it is assumed that $\mathcal{G}$ follows a bivariate Gaussian distribution. This result was proved experimentally by employing a Monte Carlo simulation. In particular, we generated 200,000 samples $n \in \mathbb{R}^{m_1 \times m_2}$ of white Gaussian noise processes with different noise standard deviations $\sigma = \{10, 15, 20\}$. First, we estimated the correlation coefficient between $g^h[i,j]$ and $g^v[i,j]$ resulting in a value of 0.07 which implies $G^h[i,j]$, $G^v[i,j]$ in (2.13) have very small to no correlation [1]. Then we estimated the coefficients ρ_d and ρ_g in (2.16).

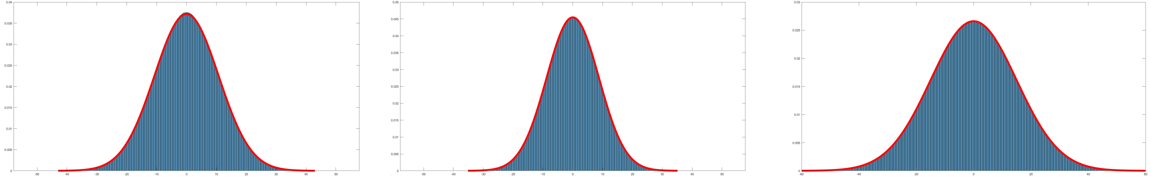
In Table 2.1, together with the estimated values of ρ_d, we report estimates of ρ_{g^h}, which is the correlation between the realizations $g^h[i+1,j]$ and $g^h[i,j]$, and ρ_{g^v}, which is the correlation between $g^v[i,j+1]$ and $g^v[i,j]$, for increasing image dimensions $m = \{128^2, 200^2, 256^2, 400^2, 512^2, 1024^2, 2048^2, 4000^2\}$. As expected, ρ_{g_h} and ρ_{g_v} are equal and we refer to them as ρ_g.

Table 2.1
Pearson's correlation coefficients for varying σ and image dimension m; ρ_g between adjacent pixels and ρ_d.

	$\sigma = 10$			$\sigma = 15$			$\sigma = 20$		
m	ρ_d	ρ_{g^h}	ρ_{g^v}	ρ_d	ρ_{g^h}	ρ_{g^v}	ρ_d	ρ_{g^h}	ρ_{g^v}
128^2	0.460	0.623	0.623	0.460	0.623	0.623	0.460	0.623	0.623
200^2	0.462	0.650	0.650	0.462	0.650	0.650	0.462	0.650	0.650
256^2	0.463	0.663	0.663	0.463	0.663	0.663	0.463	0.663	0.663
400^2	0.464	0.681	0.681	0.464	0.681	0.681	0.464	0.681	0.681
512^2	0.465	0.689	0.689	0.465	0.689	0.689	0.465	0.689	0.689
1024^2	0.466	0.701	0.701	0.466	0.701	0.701	0.466	0.701	0.701
2048^2	0.467	0.702	0.702	0.467	0.702	0.702	0.467	0.702	0.702
4000^2	0.467	0.702	0.702	0.467	0.702	0.702	0.467	0.702	0.702



$G^h \sim Gauss(0, \sigma_g^2)$ $D_h G^h \sim Gauss(0, \sigma_d^2)$ $D_h G^h + D_v G^v \sim Gauss(0, \sigma^2)$



$G^v \sim Gauss(0, \sigma_g^2)$ $D_v G^v \sim Gauss(0, \sigma_d^2)$ $N \sim Gauss(0, \sigma^2 = 15^2)$

Figure 2.3. *Histograms of sample distributions obtained by Monte Carlo simulations using* 20,000 *samples of Gaussian noise realizations n with standard deviation $\sigma = 15$. The red solid line is the associated theoretical probability distributions.*

The estimated values of ρ_d and ρ_g, for both g^h and g^v, remain constant for each value of the noise standard deviation σ, whereas with increasing dimensions m, the change follows a logarithmic growth. For an arbitrary image dimension m, we suggest the following polynomial regression for ρ_d and ρ_g in terms of the image dimension:

$$\rho_d(x = \log_2(m)) = 1.29 \times 10^{-5} x^3 - 8.35 \times 10^{-4} x^2 + 1.81 \times 10^{-2} x + 0.33 \,,$$

$$\rho_g(x = \log_2(m)) = 2.93 \times 10^{-6} x^5 - 2.53 \times 10^{-4} x^4 + 8.72 \times 10^{-3} x^3 - 0.15 x^2 + 1.34 x - 4.34 \,.$$

Assuming the number of samples sufficiently large, Monte Carlo simulations allowed the validation of the asymptotic behavior of the variables G^h, G^v, $D_h G^h$, $D_v G^v$, $D_h G^h + D_v G^v$. In particular, we generated 20,000 samples of Gaussian noise realizations n with the same standard deviations $\sigma = 15$. For each sample n the values g^h, g^v, $D_h g^h$, $D_v g^v$, $D_h g^h + D_v g^v$ were computed, the normalized histogram of these values was constructed and shown in Figure 2.3. The associated theoretical probability distributions are illustrated in a solid red line. The theoretical standard deviations σ_g and σ_d are computed by (2.16) and (2.17), respectively, and the Pearson's correlation coefficients ρ_d and ρ_g are taken from Table 2.1. We notice that all the variables, as expected, follow the Gaussian distributions asymptotically.

The following Proposition 2.3, which assumes i.i.d. Gaussian random processes $\mathcal{P}$, therefore uncorrelated, provides theoretically expected values for $\|\mathcal{P}\|_2$ and $\mathcal{R}_3(\mathcal{P})$. Even though the results in Proposition 2.3 cannot directly apply to $\mathcal{G}$ due to its correlated nature, we will use them to evaluate approximations of $\|\mathcal{G}\|_2$ and $\mathcal{R}_3(\mathcal{G})$ supported by encouraging simulation results. For further insight onto the sum of correlated squared Gaussian variables we refer the reader to [7].

Proposition 2.3. *Let $\mathcal{P}$ be an $m = m_1 \times m_2$ i.i.d. random process consisting of random variables $P[i,j]$ following a bivariate Gaussian distribution with variance s^2, then the following relationships hold:*

(i) $\|\mathcal{P}\|_2 \sim G\Gamma(2, 2m, s\sqrt{2})$ *has a generalized gamma distribution with expected value*

$$\mathbb{E}[\|\mathcal{P}\|_2] = \sqrt{2}s\frac{\Gamma(m+1/2)}{\Gamma(m)} = \sqrt{2}s\, e^{\ln\Gamma(m+1/2) - \ln\Gamma(m)};$$

(ii) $\mathcal{R}_3(\mathcal{P}) = \left(\|\mathcal{P}\|_2^2\right)^2 \sim G\Gamma(1/2, m/2, 4s^4)$ *has a generalized gamma distribution with scale parameter $1/2$, shape parameters $m/2$ and $4s^4$, and expected value*

$$\mathbb{E}\left[\mathcal{R}_3(\mathcal{P})\right] = 4s^4(m+1)m.$$

Proof. The sum of $2m$ independent squared Gaussian distributed variables of variance s^2 follows a Gamma distribution with shape parameter m and scale parameter $2s^2$:

$$\|\mathcal{P}\|_2^2 = \sum_{[i,j]\in\Omega} \|P[i,j]\|_2^2 \sim \Gamma(m, 2s^2). \tag{2.19}$$

The distribution of the $\|\mathcal{P}\|_2$ in (i) is then trivially obtained by taking the square root of the gamma distributed variable (2.19) which follows the generalized gamma distribution

$$\|\mathcal{P}\|_2 = \sqrt{\|\mathcal{P}\|_2^2} \sim G\Gamma(2, 2m, s\sqrt{2})$$

with shape parameters 2 and $2m$, and scale parameter $s\sqrt{2}$. The generalized gamma distribution represents a generalization of the χ distribution for non-standard-distributed variables. The expected value is given in terms of the gamma function $\Gamma(\cdot)$ as

$$\mathbb{E}[\|\mathcal{P}\|_2] = \sqrt{2}s\,\frac{\Gamma(m+1/2)}{\Gamma(m)}\ ,$$

thus concluding (i). From (2.19), the square of the gamma distributed variable results in

$$\mathcal{R}_3(\mathcal{P}) = \phi_3(\|\mathcal{P}\|_2^2) = \left(\|\mathcal{P}\|_2^2\right)^2 \sim G\Gamma(1/2, m/2, (2s^2)^2)\ ,$$

with expected value $\mathbb{E}[\mathcal{R}_3(\mathcal{P})] = 4s^4\frac{\Gamma(2(m/2+1))}{\Gamma(2(m/2))} = 4s^4(m+1)m$, thus proving (ii). ■

Table 2.2

Results on σ_g estimations and validation of the expected values of $\|g\|_2$ and of the model regularization term $\mathcal{R}_3(g)$.

$m = 200^2$	$\sigma =$	1	5	10	20	40	60
	σ_g	0.70	3.49	6.99	13.98	27.96	41.94
	$\bar{\sigma}_g$	0.70	3.52	7.04	14.08	28.15	42.25
	$\bar{\sigma}_g/\sigma_g$	1.007	1.007	1.007	1.007	1.007	1.007
	$\mathbb{E}[\|\mathcal{G}\|_2]$	0.78	3.88	7.75	15.51	31.01	46.52
	$\|g\|_2$	0.78	3.90	7.81	15.62	31.23	46.86
	$\|g\|_2/\mathbb{E}[\|\mathcal{G}\|_2]$	1.004	1.004	1.004	1.004	1.004	1.004
	$\mathbb{E}[\mathcal{R}_3(\mathcal{G})]$	0.36	225.82	3613	5.78e+04	9.25e+05	4.68e+06
	$\mathcal{R}_3(g)$	0.38	236.07	3786	6.08e+04	9.71e+05	4.92e+06
	$\mathcal{R}_3(g)/\mathbb{E}[\mathcal{R}_3(\mathcal{G})]$	1.05	1.05	1.05	1.05	1.05	1.05
$m = 400^2$	$\sigma =$	1	5	10	20	40	60
	σ_g	0.73	3.66	7.32	14.63	29.27	43.90
	$\bar{\sigma}_g$	0.74	3.68	7.36	14.71	29.42	44.15
	$\bar{\sigma}_g/\sigma_g$	1.005	1.005	1.005	1.005	1.005	1.005
	$\mathbb{E}[\|\mathcal{G}\|_2]$	1.62	8.12	16.23	32.46	64.92	97.38
	$\|g\|_2$	1.63	8.16	16.32	32.63	65.26	97.95
	$\|g\|_2/\mathbb{E}[\|\mathcal{G}\|_2]$	1.003	1.003	1.003	1.003	1.003	1.003
	$\mathbb{E}[\mathcal{R}_3(\mathcal{G})]$	6.94	4337	69400	1.11e+06	1.78e+07	8.99e+07
	$\mathcal{R}_3(g)$	7.17	4496	71903	1.15e+06	1.84e+07	9.33e+07
	$\mathcal{R}_3(g)/\mathbb{E}[\mathcal{R}_3(\mathcal{G})]$	1.04	1.04	1.04	1.04	1.04	1.04

Even though Proposition 2.3 cannot be directly applied for our process $\mathcal{G}$ due to the correlations between pixel variables $G[i,j]$ and $G[k,l]$ for $[i,j] \neq [k,l]$, for practical usage, we approximate the expected values of $\|\mathcal{G}\|_2$ and $\mathcal{R}_3(\mathcal{G})$ by results from Proposition 2.3 as

$$\mathbb{E}[\|\mathcal{G}\|_2] \approx \sqrt{2}\sigma_g \, \mathrm{e}^{\ln\Gamma(m+1/2) - \ln\Gamma(m)} \, , \tag{2.20}$$

$$\mathbb{E}[\mathcal{R}_3(\mathcal{G})] \approx 4\sigma_g^4 (m+1)m \, . \tag{2.21}$$

This is justified by the Monte Carlo simulations reported in Table 2.2, where we present a computational validation of Proposition 2.2 and approximations (2.20)–(2.21) with 20,000 realizations of a Gaussian noise process for each standard deviation σ in the "range $\sigma = \{1, 5, 10, 20, 40, 60\}$"/255 and sample dimension $m = m_1 \times m_2$ in the range $m = \{200^2, 400^2\}$. In this table, we also consider $\|g\|_2 = (\sum_{i=1}^{m} (g_i^h)^2 + (g_i^v)^2)^{1/2}$ for comparison with $\mathcal{R}_3(g)$. For each sample set, and for each increasing noise standard deviation value σ, we report in the first block in Table 2.2 the standard deviation σ_g computed according to formula (2.16) in Proposition 2.2, and the experimental standard deviation $\bar{\sigma}_g$ directly obtained from the samples. Analogously, in the second and third blocks of Table 2.2, we validate the approximation drawn from (i) and (ii) of Proposition 2.3, respectively. The approximate expected values $\mathbb{E}[\|\mathcal{G}\|_2]$ and $\mathbb{E}[\mathcal{R}_3(\mathcal{G})]$ in Table 2.2 were calculated using the σ_g value in (2.16). The reported ratios, being constant across different values of σ and close to one, indicate a low relative

Table 2.3

Comparison to $\sqrt{C}$ value computed via formula proposed by [3] *and via* (2.22) *for different image size m.*

m	32^2	64^2	128^2	256^2	512^2
$\sqrt{C}$ by [3]	0.843	0.906	0.965	1.021	1.073
$\sqrt{C}$ by (2.22)	0.840	0.904	0.964	1.021	1.073

error introduced as a consequence of neglecting the correlation between $\mathcal{G}[i,j]$ and $\mathcal{G}[k,l]$ for $[i,j] \neq [k,l]$.

Based on these statistical investigations, in the following we provide a formula which relates the approximation to $\|\mathcal{G}\|_2^2$ with the L_2 norm of the noise process $\mathcal{N}$.

Corollary 2.4. *Let $\mathcal{N} \in \mathbb{R}^m$ be a Gaussian noise process with known standard deviation σ, i.e., $\mathcal{N} \sim Gauss(0, \sigma^2 I_m)$. Then, under the approximation drawn from Proposition* 2.3, *we have*

$$\mathbb{E}[\|\mathcal{G}\|_2^2] \approx C \cdot \mathbb{E}[\|\mathcal{N}\|_2^2], \qquad C \approx \frac{2\sigma_g^2}{\sigma^2}. \tag{2.22}$$

Proof. The norm of a Gaussian noise process of arbitrary standard deviation σ distributes as $\|\mathcal{N}\|_2^2 \sim \Gamma(\frac{m}{2}, 2\sigma^2)$, while $\|\mathcal{G}\|_2^2$ approximately distributes as defined in (2.19). Computing the ratio of the respective expected values, we have

$$C \approx \frac{\mathbb{E}[\|\mathcal{G}\|_2^2]}{\mathbb{E}[\|\mathcal{N}\|_2^2]} \approx \frac{2m\sigma_g^2}{m\sigma^2}$$

which concludes the proof. ■

This provides a statistical insight into the distribution of the H^{-1}-norm approximation for Gaussian noise images. In [3], the authors proposed a formula to estimate the C value. An alternative formula is obtained by substituting σ_g as defined in (2.16) into (2.22) for $\delta = 1$, thus obtaining $C \approx \frac{1}{2(1+\rho_d)(1-\rho_g)}$. In Table 2.3, we report the calculated values of $\sqrt{C}$ computed by the formula (2.22) and by Proposition 3.5 in [3]. We note that these values are extremely similar and depend on the image dimension.

3. Effects of model penalty terms.

In this section, we investigate the potential of each penalty term in the proposed variational model (2.7), to discriminate between the different image components. Towards this aim, let us consider our idealized variational model (2.1) for additive decomposition, with terms $\|\nabla v\|_0$, $\|\mathcal{H}w\|_2$ with $\mathcal{H}(\cdot)$ of second order, and $\|n\|_{H^{-1}}$ in (2.4) evaluated by the approximation given by (2.5). A good model is given by a choice of spaces/norms so that, with the given desired properties of v, w, and n, we can obtain $\|\nabla v\|_0 << \|\nabla w\|_0$, $\|\nabla v\|_0 << \|\nabla n\|_0$, $\|\mathcal{H}w\|_2 << \|\mathcal{H}v\|_2$, $\|\mathcal{H}w\|_2 << \|\mathcal{H}n\|_2$, and finally $\|n\|_{H^{-1}} << \|v\|_{H^{-1}}$, $\|n\|_{H^{-1}} << \|w\|_{H^{-1}}$.

To illustrate this, six different synthetic test images u_1 to u_6 are presented, all having the same dimension $m = 512^2$ pixels, and defined as follows:

$$
\begin{aligned}
\text{Chessboard:} \quad & u_1(i,j) = \begin{cases} -80, & \sin(jz/512)\cos(iz/512) \leq 0, \\ 80, & \text{otherwise}, \end{cases} \\
\text{Diagonal stripes:} \quad & u_2(i,j) = \begin{cases} -80, & \sin((i+j)z/512) \leq 0, \\ 80, & \text{otherwise}, \end{cases} \\
\text{Blurry stripes:} \quad & u_3(i,j) = 80\sin(jz/512), \\
\text{Blurry chessboard:} \quad & u_4(i,j) = 80\sin(jz/512)\cos(iz/512), \\
\text{Shading stripes:} \quad & u_5(i,j) = j/2 + c\sin(jz/512), \\
\text{Noise image:} \quad & u_6(i,j) \sim Gauss(0,\sigma^2).
\end{aligned}
$$

The first two images u_1 and u_2 are piecewise-constant scalar fields with sharp edges; u_3 and u_4 represent smooth-gradient scalar fields with varying frequency of oscillations according to parameter z; u_5 is a combination of constant-gradient slope with oscillations, which represents a nonzero mean scalar field, and the image u_6 is a realization of Gaussian noise with different standard deviations σ. Figure 3.1 shows these test cases with varying parameters z, and Table 3.1 shows the associated norm values which characterize the three penalty terms in (2.1).

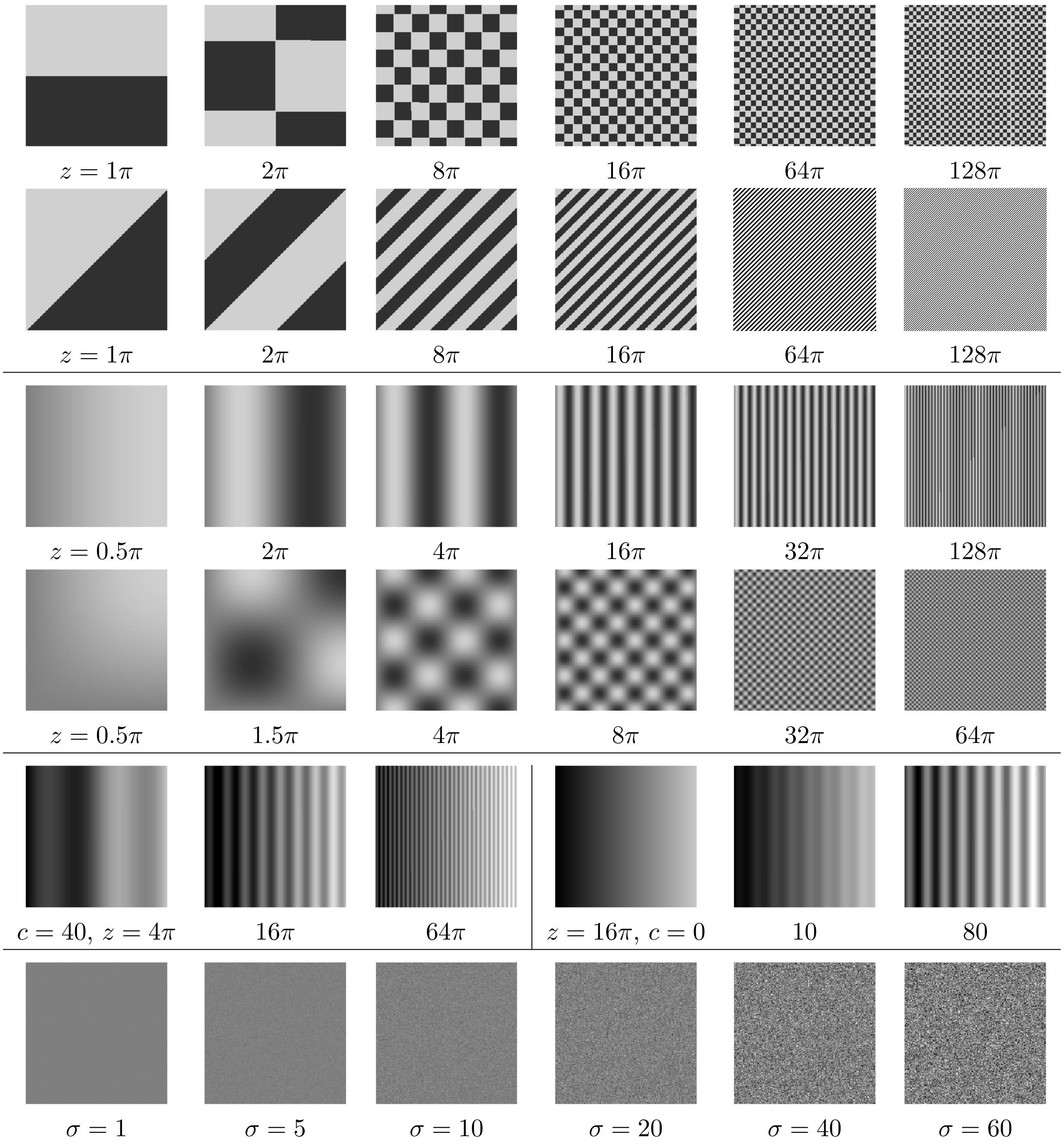
First, in Table 3.1, the piecewise-constant images such as u_1 and u_2 attain the smallest $\|\nabla u_*\|_0$ compared to the other two norms $\|\mathcal{H}u_*\|_2$, and $\|u_*\|_{H^{-1}}$, e.g., when $z = \pi$, $\|\nabla u_*\|_0$ values are around 500, while other norm values are much bigger. For u_1, $\|\nabla u_1\|_0$ is smaller compared to other norms until $z = 16\pi$, then for $z = 64\pi$ and 128π, $\|\mathcal{H}u_1\|_2$ and $\|u_1\|_{H^{-1}}$ values become more comparable with $\|\nabla u_1\|_0$. This effect is consistent with the visual effect, that for $z = 64\pi$ or 128π, the image no longer looks like a piecewise-constant image, but it looks more like a texture. In fact, for $z = 128\pi$, we notice that the approximated H^{-1}-norm values are the smallest. Across the images, $\|\nabla u_*\|_0$ for u_1 and u_2 are the smallest around 500, compared to $\|\nabla u_*\|_0$ of the other blurry or noisy images u_3 to u_6 above 130,000.

Second, for blurry smooth images u_3 and u_4, notice the small values of $\|\mathcal{H}u_*\|_2 < 0.5$, in comparison to the other norms above 100,000. Comparing across the images, Table 3.1 shows that the slope image u_5 also give small $\|\mathcal{H}u_*\|_0$ values, around 9 and 0, since this image is also dominated by a smooth component.

Third, as z increases, the details become finer in each row, and decreasing H^{-1}-norm values represent this fact accordingly. Also notice that in the last column of Table 3.1, H^{-1}-norm values are similar around 31,000–38,000 across the rows. This gives a clear idea what level of detail is preferred in the H^{-1}-norm.

Fourth, the mean of the image affects the norm values. Images u_5 are a combination of slope and sine functions with increasing oscillations and fixed magnitude $c = 40$ (first three columns), and increasing amplitude of the oscillating part fixing $z = 16\pi$ (next three columns). As oscillations increase from left to right $\|\mathcal{H}u_5\|_2$ values increase, but $\|\nabla u_5\|_0$ and $\|u_5\|_{H^{-1}}$ values stay relatively the same. This effect is due to the nonzero mean of the texture; in such cases H^{-1}-norm values remain large.

Fifth, for the noise images u_6, the H^{-1}-norm value is always smaller compared to $\|\nabla u_6\|_0$ and $\|\mathcal{H}u_6\|_2$. The H^{-1}-norm values increase as the noise level increases. Comparing the



$z = 1\pi$ 2π 8π 16π 64π 128π

$z = 1\pi$ 2π 8π 16π 64π 128π

$z = 0.5\pi$ 2π 4π 16π 32π 128π

$z = 0.5\pi$ 1.5π 4π 8π 32π 64π

$c = 40,\ z = 4\pi$ 16π 64π | $z = 16\pi,\ c = 0$ 10 80

$\sigma = 1$ $\sigma = 5$ $\sigma = 10$ $\sigma = 20$ $\sigma = 40$ $\sigma = 60$

Figure 3.1. *Case study images u_{1-6} for different z values, frequencies $z/2\pi$, and standard deviations σ.*

$\|u_6\|_{H^{-1}}$ values with other images, when $\sigma = 60$, the noisy u_6 gives a similar value of the H^{-1}-norm around 33,000, due to the image resolution, and the visual effect is similar across images with similar norm values. We can see that the H^{-1}-norm value stays small only for the noise images u_6, and relatively large for nontexture but oscillating images u_{1-5}.

Table 3.1

Case study: model norms approximating values corresponding to the images $u_1, \ldots, u_6$ in Figure 3.1.

Chessboard	$z =$	1π	2π	8π	16π	64π	128π
	$\|\nabla u_1\|_0 \approx$	513	1531	7540	15272	54944	89920
	$\|\mathcal{H}u_1\|_2 \approx$	5147	8868	19830	28507	57700	81730
	$\|u_1\|_{H^{-1}} \approx$	2.48e6	2.39e6	527582	255416	63056	31843
Diagonal stripes	$z =$	1π	2π	8π	16π	64π	128π
	$\|\nabla u_2\|_0 \approx$	511	1024	4096	8192	32768	65536
	$\|\mathcal{H}u_2\|_2 \approx$	10215	14449	28891	40856	81711	115556
	$\|u_2\|_{H^{-1}} \approx$	2.10e6	2.34e6	554548	272684	67252	33494
Blurry stripes	$z =$	0.5π	2π	4π	16π	32π	128π
	$\|\nabla u_3\|_0 \approx$	132257	232915	247452	260769	261632	261632
	$\|\mathcal{H}u_3\|_2 \approx$	0.27	4.36	17.45	278.93	1113	16950
	$\|u_3\|_{H^{-1}} \approx$	676799	1.99e6	1.14e6	296022	148488	38027
Blurry chessboard	$z =$	0.5π	1.5π	4π	8π	32π	64π
	$\|\nabla u_4\|_0 \approx$	142514	247336	259889	261389	261109	258047
	$\|\mathcal{H}u_4\|_2 \approx$	0.38	3.46	24.62	98.48	1571	6224
	$\|u_4\|_{H^{-1}} \approx$	719103	1.63e6	628639	305844	74601	37304
Shading stripes	$(c, z) =$	$(40, 4\pi)$	$(40, 16\pi)$	$(40, 64\pi)$	$(0, 16\pi)$	$(10, 16\pi)$	$(80, 16\pi)$
	$\|\nabla u_5\|_0 \approx$	227040	254455	261632	261632	226701	256769
	$\|\mathcal{H}u_5\|_2 \approx$	8.72	139.46	2204	0	34.87	278.93
	$\|u_5\|_{H^{-1}} \approx$	1.95e6	1.97e6	1.95e6	1.95e6	1.95e6	1.99e6
Noise image	$\sigma =$	1	5	10	20	40	60
	$\|\nabla u_6\|_0 \approx$	258999	261992	262108	262132	262139	262141
	$\|\mathcal{H}u_6\|_2 \approx$	2287	11396	22861	45657	91493	136451
	$\|u_6\|_{H^{-1}} \approx$	499.45	2796	5164	10581	22403	33071

Table 3.1 shows the norm values for images in the range $[0, 255]$. When the image intensity range changes, the reported values would change, but their relative behavior would be preserved.

4. Model parameter selection. Learning from the previous section, we propose simple strategies for an effective, automatic selection of all the parameters γ_1, γ_2, γ_3 that balance the energies in the minimizing function $\mathcal{J}$ in (2.7). The goal is to adjust the model parameters in such a way that each regularization term $\mathcal{R}_1(v)$, $\mathcal{R}_2(w)$, $\mathcal{R}_3(g)$ perform as well as possible. This can be achieved by balancing the energy contribution from each term to the total energy $\mathcal{J}$ with suitable γ_1, γ_2, γ_3 parameter values, such that

$$\gamma_1 \approx \frac{1}{\mathcal{R}_1(v)}, \qquad \gamma_2 \approx \frac{1}{\mathcal{R}_2(w)}, \qquad \gamma_3 \approx \frac{1}{\mathcal{R}_3(g)}.$$

We propose to estimate the values for $\mathcal{R}_1(v)$, $\mathcal{R}_2(w)$, $\mathcal{R}_3(g)$ using some a priori knowledge of the image components v, w, and n.

First of all, the estimate of γ_1 follows Proposition 2.1 which states that $\mathcal{R}_1(v)$ is a good approximation of the ℓ_0 pseudo-norm penalty. The value $\mathcal{R}_1(v)$, defined by the sum of sparsity inducing functions $\phi_1(\cdot; a)$, approximately represents the number of image pixels that form edges and can be estimated by simple edge detection filters on f. Similarly, the minimum jump (μ in Proposition 2.1) can be estimated, thus, in order to satisfy (2.11), the value $\bar{a}$ is set as

$$\bar{a} = \tau\mu = \tau \min_{j:\,|(\nabla v)_j|>0} |(\nabla v)_j| \,, \quad 0 < \tau < 1, \tag{4.1}$$

and consequently the value for a is given by (2.11). In general, for images with values in the range $[0,1]$, the value $\bar{a}$ is less than 0.1. The higher the γ_1 value, the fewer edges of different gradient magnitudes are captured in v.

Second, the penalty $\mathcal{R}_2(w)$ is a quadratic penalty term which relies on the convex function $\phi_2(\cdot)$. We assume that the second-order derivatives of w have small magnitudes; see, for example, the images u_{3-5} of Table 3.1. Having a priori knowledge on the smooth image component we expect, we could compute a value $\bar{t}$ as

$$\bar{t} := \max_j |(\mathcal{H}w)_j| \,. \tag{4.2}$$

Alternatively, we experimentally set $\bar{t} = 10^{-3}a$. Values smaller than $\bar{t}$ are captured by $\mathcal{R}_2(w)$, while higher values should be penalized by $\mathcal{R}_1(v)$. This suggests that $\bar{t}$ be set by the abscissa of the intersection point between the functions $\phi_1(\cdot)$ and $\phi_2(\cdot)$. This is illustrated in Figure 2.1(a) where $\bar{t}$ is marked by a circle. Then the parameter γ_2 is obtained by solving

$$\frac{\gamma_2}{2}\phi_2(\bar{t}) = \gamma_1\phi_1(\bar{t};a) \quad \Rightarrow \quad \gamma_2 = \frac{2\gamma_1\phi_1(\bar{t};a)}{\bar{t}^2} \,. \tag{4.3}$$

This procedure allows a control over what magnitudes should be favored either by $\mathcal{R}_1(v)$ or $\mathcal{R}_2(w)$.

Finally, in general, the value of $\mathcal{R}_3(g)$ depends on the frequency and magnitude of the oscillatory signal as shown in Table 3.1: $\mathcal{R}_3(g)$ increases with signal magnitude and decreases with increasing frequency. When the proposed variational model (2.7) is applied to noisy images, where n represents white Gaussian noise, assuming we know the noise level σ which degraded the observed image f, we can estimate the term $\mathcal{R}_3(g)$ by (2.21), and set γ_3 to be the inverse of its approximated expected value $\mathbb{E}[\mathcal{R}_3(g)]$, that is,

$$\gamma_3 = \frac{1}{4\sigma_g^4(m+1)m} \,, \tag{4.4}$$

where σ_g is defined in Proposition 2.2.

When instead the proposed variational model (2.7) is applied to images with a textured component, as in the example illustrated at the end of section 6, $\mathcal{R}_3(g)$ does not follow a generalized gamma distribution. An insight on the selection of parameter γ_3 is suggested by the decomposition results v and n w.r.t. the relative ratio γ_1/γ_3. Figure 4.1 illustrates the decomposition of the input image f consisting of six different frequencies with the same amplitude. Increasing the relative ratio γ_1/γ_3 shifts the oscillatory signal into the n^* component. Suitable choices for γ_3 w.r.t. a fixed γ_1 allow for good control on the scale of the texture to capture in the n^* component in terms of frequency and/or oscillations. The larger is the relative ratio γ_1/γ_3, the more oscillations of smaller frequencies are captured in n^*.

In summary, a general strategy for an automatic selection of the parameters γ_1, γ_2, and γ_3, when either γ_1 or γ_3 can be estimated, consists of

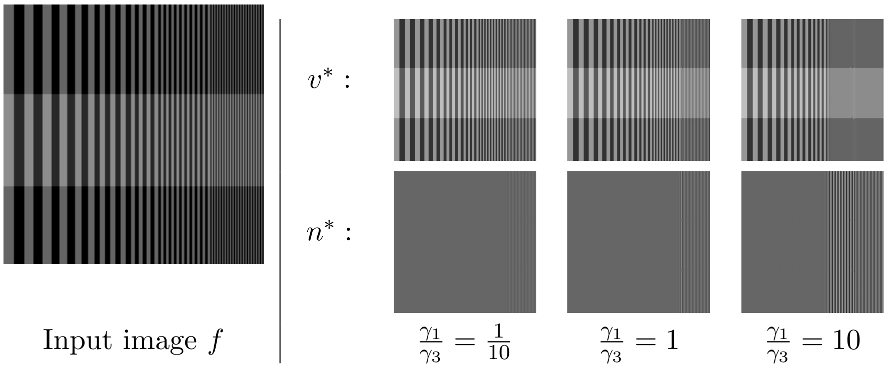


Figure 4.1. *Effect of the ratio γ_1/γ_3 on the dispensation of the piecewise-constant oscillations in f into v and n.*

1. computing the estimation for γ_3 (alternatively γ_1);
2. setting $\gamma_1 = \tau\gamma_3$ for $\tau > 1$ (alternatively, $\gamma_3 = \tau\gamma_1$ for $\tau < 1$);
3. estimating $\bar{a}$ via (4.1) and $\bar{t}$ via (4.2);
4. computing γ_2 via (4.3).

5. ADMM-based numerical solution.

In this section, we first provide details on the discretization of the differential operators. We analyze the existence and uniqueness of the optimization problem (2.7) and finally describe an efficient algorithm, based on the ADMM strategy, to obtain the numerical solution of the proposed variational model (2.7).

5.1. Discrete differential operators.

Given a grayscale image z represented as a matrix $m_1 \times m_2$, the first-order and the second-order differential operators at a pixel (i, j) are discretized using the following standard finite difference approximations: $(\nabla_h z)_{i,j} \approx z_{i,j+1} - z_{i,j}$, $(\nabla_v z)_{i,j} \approx z_{i+1,j} - z_{i,j}$, $(\mathcal{H}_{hh} z)_{i,j} \approx z_{i,j-1} - 2z_{i,j} + z_{i,j+1}$, $(\mathcal{H}_{vv} z)_{i,j} \approx z_{i-1,j} - 2z_{i,j} + z_{i+1,j}$, $(\mathcal{H}_{hv} z)_{i,j} = (\mathcal{H}_{vh} z)_{i,j} \approx z_{i,j} + z_{i+1,j+1} - z_{i,j+1} - z_{i+1,j}$ for $1 < i < m_1$, $1 < j < m_2$. Discretizations for boundary pixels come from assuming antireflective boundary conditions for z [26].

If the image z is represented in column-major form as an m-dimensional column vector with $m = m_1 m_2$, the discretized first- and second-order differential operators are represented by matrices $D := (D_h^T, D_v^T)^T \in \mathbb{R}^{2m \times m}$ and $H := (H_{hh}^T, H_{vv}^T, H_{hv}^T, H_{vh}^T)^T \in \mathbb{R}^{4m \times m}$, respectively, with $D_h, D_v, H_{hh}, H_{vv}, H_{hv}, H_{vh} \in \mathbb{R}^{m \times m}$ coefficient matrices of the finite difference operators approximating the first-order horizontal and vertical partial derivatives and the second-order horizontal, vertical, and mixed horizontal-vertical partial derivatives.

Using matrices D and H, the discretized gradient and the discretized second-order differential operator at a pixel j of the vectorized image z are defined as follows:

$$(\nabla x)_j \approx (Dx)_j := ((D_h x)_j, (D_v x)_j)^T \in \mathbb{R}^2 \,, \tag{5.1}$$

$$(\mathcal{H}(x))_j \approx (Hx)_j := ((H_{hh} x)_j, (H_{vv} x)_j, (H_{hv} x)_j, (H_{vh} x)_j)^T \in \mathbb{R}^4 \,. \tag{5.2}$$

In analogy with the continuous setting, we define the discrete version of the divergence operator in terms of the adjoint ∇^* of the gradient ∇ that, applied to a vector field $g = (g_h, g_v)$, is

$$\nabla \cdot g = -\nabla^* g \approx D^T g \,.$$

Using these discretizations of the differential operators and making all the penalty terms explicit, the minimization problem (2.7) reads as

$$\{v^*, w^*, g^*\} \in \arg \min_{v,w\in\mathbb{R}^m, g\in\mathbb{R}^{2m}} \mathcal{J}(v,w,g) \tag{5.3}$$

with

$$\mathcal{J}(v,w,g) = \gamma_1 \sum_{j=1}^{m} \phi_1\left(|(Dv)_j|; a\right) + \frac{\gamma_2}{2}\,\|Hw\|_2^2 + \frac{\gamma_3}{2}\,\|g\|_2^4 + \frac{1}{2}\left\|f - (v + w + D^T g)\right\|_2^2 . \tag{5.4}$$

5.2. Analysis of the model. In this section, we outline some important analytical properties of our cost functional $\mathcal{J}$ in (5.4), with particular focus on its convexity and coercivity, aiming at proving the existence of global minimizers. We remark that some of the reported results (or their proof) depend on the discretization choices outlined in section 5.1 for the differential operators D and H. However, analogous results could be obtained in a similar manner for other discretization schemes.

In the following Proposition 5.1, whose proof is provided in Appendix A, we analyze $\mathcal{J}$ with focus on its convexity. To simplify the notations, we introduce the total optimization variable $x := \left(v; w; g^h; g^v\right) \in \mathbb{R}^{4m}$.

Proposition 5.1. *For any $\gamma_1, \gamma_2, \gamma_3, a > 0$ and any $f \in \mathbb{R}^m$, the function $\mathcal{J}$ in (5.4) is proper, continuous, bounded from below by zero, and nonconvex in x. Moreover, $\mathcal{J}$ is strongly convex in w and strictly convex in g for any $\gamma_1, \gamma_2, \gamma_3, a > 0$ and any $f \in \mathbb{R}^m$, whereas it is convex (strongly convex) in v if parameters γ_1, a satisfy*

$$a \;\le\; (<)\; \frac{1}{\gamma_1\,\lambda_{\max}} \quad\Longleftrightarrow\quad a \;=\; \tau_c\,\frac{1}{\gamma_1\,\lambda_{\max}}\,, \quad \tau_c \in [0,1] \;\;(\tau_c \in [0,1))\,, \tag{5.5}$$

where $\lambda_{\max} \in \mathbb{R}_{++}$ denotes the maximum eigenvalue of matrix $D^T D$, which is equal to 8 for D defined in section 5.1.

Motivated by nonconvexity of the cost function $\mathcal{J}$, which does not allow us to obtain uniqueness results for its global minimizers, in the following we analyze the behavior of $\mathcal{J}$ at infinity, in particular coercivity, and demonstrate the existence of global minimizers.

First, in the following Lemma 5.2 we provide explicit forms for the null spaces of the finite difference matrices D and H defined in section 5.1. Then, in Proposition 5.3 we state the existence results.

Lemma 5.2. *Let $z \in \mathbb{R}^m$ be the vectorized (column-major) form of an $m_1 \times m_2$ image with $m = m_1 \times m_2$, and let $D \in \mathbb{R}^{2m\times m}$ and $H \in \mathbb{R}^{4m\times m}$ be the discretization of ∇ and $\mathcal{H}$ given in section 5.1 which compute the discrete first- and second-order partial derivatives of image z. Then, the null spaces of D and H are the 1-dimensional and 3-dimensional linear spaces of (vectorized) $m_1 \times m_2$ constant and affine images, respectively; in the formulas*

$$\text{null}(D) \;=\; \text{span}\left(h^{(1)}\right) \subset \mathbb{R}^m, \qquad \text{null}(H) \;=\; \text{span}\left(h^{(1)}, h^{(2)}, h^{(3)}\right) \subset \mathbb{R}^m \tag{5.6}$$

with basis vectors $h^{(1)}, h^{(2)}, h^{(3)} \in \mathbb{R}^m$ *defined by*

(5.7)
$$h^{(1)} = 1_m\,, \qquad h^{(2)} = \text{vec}\left.\begin{pmatrix} 1 & 2 & \dots & m_2 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & 2 & \dots & m_2 \end{pmatrix}\right\} m_1 \text{ rows}\,, \qquad h^{(3)} = \text{vec}\underbrace{\begin{pmatrix} 1 & \dots & 1 \\ 2 & \dots & 2 \\ \vdots & \dots & \vdots \\ m_1 & \dots & m_1 \end{pmatrix}}_{m_2 \text{ columns}}.$$

Proof. It follows from the definitions of matrices D and H given in section 5.1 that

$$\begin{aligned} \text{null}(D) &= \{z \in \mathbb{R}^m\colon\ (Dz)_j = 0_2 \ \ \forall j = 1,\dots,m\}\,, \\ \text{null}(H) &= \{z \in \mathbb{R}^m\colon\ (Hz)_j = 0_4 \ \ \forall j = 1,\dots,m\}\,. \end{aligned} \tag{5.8}$$

Recalling the finite difference choices outlined in section 5.1 and noting that here we are considering a vectorized image z, we have

$$(Dz)_j = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \iff \underbrace{\begin{pmatrix} -1 & 0 & 1 \\ -1 & 1 & 0 \end{pmatrix}}_{\overline{D}} \begin{pmatrix} z_j \\ z_{j+1} \\ z_{j+m_1} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \tag{5.9}$$

$$(Hz)_j = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix} \iff \underbrace{\begin{pmatrix} 1 & 0 & -2 & 0 & 1 & 0 \\ 0 & 1 & -2 & 1 & 0 & 0 \\ 0 & 0 & 1 & -1 & -1 & 1 \end{pmatrix}}_{\overline{H}} \begin{pmatrix} z_{j-m_1} \\ z_{j-1} \\ z_j \\ z_{j+1} \\ z_{j+m_1} \\ z_{j+1+m_1} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \tag{5.10}$$

where in (5.10) we neglected the fourth row of matrix $\overline{H}$ as it coincides with the third one. It is easy to prove that the solutions of underdetermined linear systems in (5.9) and (5.10), which correspond to the null spaces of coefficient matrices $\overline{D} \in \mathbb{R}^{2\times 3}$ and $\overline{H} \in \mathbb{R}^{3\times 6}$, respectively, read

$$\text{null}\big(\overline{D}\big) = \text{span}\big(\alpha_0\big) \subset \mathbb{R}^3\,, \qquad \text{null}\big(\overline{H}\big) = \text{span}\big(\alpha_1, \alpha_2, \alpha_3\big) \subset \mathbb{R}^6 \tag{5.11}$$

with basis vectors $\alpha_0 \in \mathbb{R}^3$ and $\alpha_1, \alpha_2, \alpha_3 \in \mathbb{R}^6$ given, e.g., by

(5.12)
$$\alpha_0 = (1,1,1)^T, \quad \alpha_1 = (1,1,1,1,1,1)^T, \quad \alpha_2 = (1,2,2,2,3,3)^T, \quad \alpha_3 = (2,1,2,3,2,3)^T.$$

A 2-dimensional visualization of the basis vectors is given in Figure 5.1.

Hence, conditions (5.9) and (5.10) are equivalent to saying that image z is locally (i.e., over the local stencils for D and H) constant and affine, respectively. In order to prove that such local properties extend to the whole image, we notice that, by shifting one pixel horizontally or vertically (i.e., moving from pixel j to pixels $j \pm 1$ or $j \pm m_1$), the intersection of the shifted

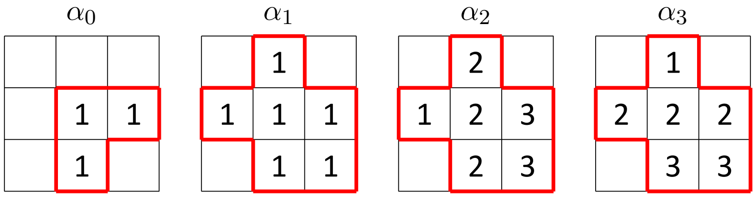


Figure 5.1. *2-dimensional visualization of basis vectors* $\alpha_0 \in \mathbb{R}^3$, $\alpha_1, \alpha_2, \alpha_3 \in \mathbb{R}^6$ *defined in* (5.12).

local stencil with the original one is made by one pixel for D and three pixels for H (see the red-bordered stencils depicted in Figure 5.1). This means that the local constant or affine configurations must be the same for all pixels and, hence, the null spaces of D and H are given by the 1-dimensional and 3-dimensional linear spaces of (vectorized) $m_1 \times m_2$ constant and affine images, respectively. In (5.7) we report one among the infinity of possible sets of basis vectors for the two null spaces. ■

Proposition 5.3. *For any* $\gamma_1, \gamma_2, \gamma_3, a > 0$ *and any* $f \in \mathbb{R}^m$*, the function* $\mathcal{J}$ *in* (5.4) *is not coercive in* x*, nevertheless it admits global minimizers.*

Proof. Proving that $\mathcal{J}$ is not coercive in x is immediate by considering how the restriction of $\mathcal{J}$ to the line $l(t) = (f; 0_m; 0_{2m}) + t\,(1_m; -1_m; ; 0_{2m})$, $t \in \mathbb{R}$, behaves at infinity. In fact, it follows from definition (5.4) that

(5.13)

$$\lim_{|t|\to\infty} \mathcal{J}\,(l(t)) \;=\; \lim_{|t|\to\infty} \left\{ \gamma_1 \sum_{j=1}^{m} \phi_1\left(|(Df)_j|; a\right) \right\} \le \gamma_1 m < +\infty \;\; \forall \gamma_1, \gamma_2, \gamma_3, a \in \mathbb{R}_{++} \; \forall\, f \in \mathbb{R}^m ,$$

where the second-last inequality comes from definition (2.8) of ϕ_1, namely, from $\max_{t\in\mathbb{R}_+} \phi_1(t; a) = 1 \; \forall\, a \in \mathbb{R}_{++}$. Proving that, in spite of noncoercivity, $\mathcal{J}$ admits global minimizers for any $\gamma_1, \gamma_2, \gamma_3, a \in \mathbb{R}_{++}$ and any $f \in \mathbb{R}^m$ is less straightforward. In fact, for f a constant image, the limit in (5.13) is equal to zero, which is the minimum value that the (nonnegative) function $\mathcal{J}$ can ever attain at (finite) domain points $x \in \mathbb{R}^{4m}$. We proceed as follows. First, we detect all the possible paths of noncoercivity for $\mathcal{J}$, namely, paths towards infinity in the domain $\mathbb{R}^{4m}$ of $\mathcal{J}$ along which the value of $\mathcal{J}$ does not tend towards $+\infty$. Then, we compute all the possible limit values of $\mathcal{J}$ along these paths and, finally, we prove that these or lower values are attained by $\mathcal{J}$ at (finite) domain points $x \in \mathbb{R}^{4m}$. This implies that, even if a global infimizer exists at infinity, then a corresponding, i.e., characterized by the same function value, global minimizer exists as well.

Paths of noncoercivity. We start by noting that the function $\mathcal{J}$ in (5.4) is given by the sum of four nonnegative terms. Hence, whenever the value of one of these terms goes to $+\infty$, the value of the total function $\mathcal{J}$ goes to $+\infty$ as well. Then, the term $(\gamma_3/2)\|g\|_2^4$ is coercive in g, hence possible paths of noncoercivity for $\mathcal{J}$ must be sought by keeping g bounded (that is, letting only $\|(v; w)\|_2$ approach $+\infty$). The nonconvex regularization term in $\mathcal{J}$ is bounded from above by $\gamma_1 m < +\infty$, hence it does not affect coercivity of $\mathcal{J}$. The sum of the two remaining terms in $\mathcal{J}$ is a quadratic function of $y := (v; w) \in \mathbb{R}^{2m}$ which can be written as

$$\mathcal{Q}(y) = \frac{\gamma_2}{2}\,\|Q_1 y\|_2^2 \;+\; \frac{1}{2}\,\|Q_2 y - z\|_2^2 \;=\; \frac{1}{2}\,y^T\big(\,\overbrace{\gamma_2 Q_1^T Q_1 + Q_2^T Q_2}^{Q\,\in\,\mathbb{R}^{2m\times 2m}}\,\big)\,y \;-\; z^T Q_2\,y \;+\; \frac{1}{2}\,\|z\|_2^2 \tag{5.14}$$

with

$$Q_1 = (0_m; H) \in \mathbb{R}^{2m\times 2m}, \quad Q_2 = (I_m, I_m) \in \mathbb{R}^{m\times 2m}, \quad z = f - D^T g \in \mathbb{R}^m\,. \tag{5.15}$$

The Hessian matrix Q in (5.14) is symmetric positive semidefinite and its null space, given by the intersection of the null spaces of Q_1 and Q_2, is given by

$$\operatorname{null}(Q) = \big\{(v; w) \in \mathbb{R}^{2m} :\; v = -w,\; w \in \operatorname{null}(H)\big\}\,.$$

The paths of noncoercivity for $\mathcal{Q}$ are thus only those approaching at infinity a direction parallel to $\operatorname{null}(Q)$. Along all other paths, $\mathcal{Q}$ and, hence, $\mathcal{J}$ tend to $+\infty$.

Limit values. In order to compute the limit values of $\mathcal{J}$ along all its paths of noncoercivity, it suffices to analyze the behavior at infinity of the restrictions $\mathcal{J}_{x_0}$ of $\mathcal{J}$ to the family of parameterized affine subspaces $\mathcal{S}_{x_0} \subset \mathbb{R}^{4m}$ with parameter $x_0 = \big(v_0; w_0; g_0\big) \in \mathbb{R}^{4m}$, of the form

$$\mathcal{S}_{x_0} = x_0 + \mathcal{V}, \quad \mathcal{V} = \operatorname{span}\Big(\nu^{(1)}, \nu^{(2)}, \nu^{(3)}\Big)\,, \quad \nu^{(i)} = \Big(h^{(i)}; -h^{(i)}; 0_{2m}\Big)\,, \quad i = 1, 2, 3\,, \tag{5.16}$$

where $h^{(i)} \in \mathbb{R}^m$, $i = 1, \ldots, 3$, are defined in (5.7) and represent the basis vectors of $\operatorname{null}(H)$, i.e., of the subspace of affine images (see Lemma 5.2). Based on (5.16) and on the definition of $\mathcal{J}$ in (5.4), the restrictions read

$$\begin{aligned}
\mathcal{J}_{x_0}(t_1, t_2, t_3) \;&=\; \mathcal{J}\Big(x_0 + t_1\,\nu^{(1)} + t_2\,\nu^{(2)} + t_3\,\nu^{(3)}\Big) \\
&=\; \mathcal{J}\Big(v_0 + t_1\,h^{(1)} + t_2\,h^{(2)} + t_3\,h^{(3)}\,,\; w_0 - t_1\,h^{(1)} - t_2\,h^{(2)} - t_3\,h^{(3)}\,,\; g_0\Big) \\
&=\; \overbrace{\gamma_1 \sum_{j=1}^{m} \phi_1\left(\left\|\Big(D\Big(v_0 + t_1\,h^{(1)} + t_2\,h^{(2)} + t_3\,h^{(3)}\Big)\Big)_j\right\|_2 ; a\right)}^{\widetilde{\mathcal{J}}_{x_0}(t_1,t_2,t_3)} \\
&\qquad \underbrace{+\,\frac{\gamma_2}{2}\,\|H w_0\|_2^2 + \frac{\gamma_3}{2}\,\|g_0\|_2^4 \;+\; \frac{1}{2}\,\big\|f - (v_0 + w_0 + D^T g_0)\big\|_2^2}_{\overline{\mathcal{J}}_{x_0}}\,, \qquad (t_1, t_2, t_3) \in \mathbb{R}^3\,,
\end{aligned} \tag{5.17}$$

where the latter term $\overline{\mathcal{J}}_{x_0}$ depends on x_0 but not on (t_1, t_2, t_3), hence the behavior of $\mathcal{J}_{x_0}$ at infinity, i.e., for $\|(t_1; t_2; t_3)\|_2 \to \infty$, depends mainly on the former term $\widetilde{\mathcal{J}}_{x_0}(t_1, t_2, t_3)$. It comes from the definitions of matrix D in section 5.1 and of vectors $h^{(1)}, h^{(2)}, h^{(3)}$ in (5.7) that $Dh^{(1)} = 0_{2m}$, $Dh^{(2)} = \big(1_m; 0_m\big)$, $Dh^{(3)} = \big(0_m; 1_m\big)$. It follows that

$$D\Big(v_0 + t_1\,h^{(1)} + t_2\,h^{(2)} + t_3\,h^{(3)}\Big) \;=\; Dv_0 + t_2(1_m; 0_m) + t_3(0_m; 1_m)$$

and, hence,

$$\left|\big(D\big(v_0+t_1\,h^{(1)}+t_2\,h^{(2)}+t_3\,h^{(3)}\big)\big)_j\right| = \left|(Dv_0)_j+(t_2;t_3)\right|, \qquad j=1,\dots,m\,. \tag{5.18}$$

Taking the limit, with a little abuse of notation (the limit does not formally exist), we have

$$\lim_{\|(t_1;t_2;t_3)\|_2\to\infty} \left|(Dv_0)_j+(t_2;t_3)\right| = \begin{cases} \qquad\qquad +\infty & \text{if } \|(t_2;t_3)\|_2\to\infty, \\ \left|(Dv_0)_j+(\bar{t}_2;\bar{t}_3)\right| < +\infty & \text{if } |t_1|\to\infty,\ (t_2;t_3)\to(\bar{t}_2;\bar{t}_3)\in\mathbb{R}^2, \end{cases} \tag{5.19}$$

$j=1,\dots,m$. Based on (5.17)–(5.19) and on the fact that $\lim_{t\to+\infty}\phi_1(t;a)=1$, $\forall\, a\in\mathbb{R}_{++}$, we can thus write

$$\lim_{\|(t_1;t_2;t_3)\|_2\to\infty} \mathcal{J}_{x_0}\left(t_1,t_2,t_3\right) \;=\; \begin{cases} l_{x_0}^{(1)} & \text{if } \|(t_2;t_3)\|_2\to\infty, \\ l_{x_0}^{(2)} & \text{if } |t_1|\to\infty,\ (t_2;t_3)\to(\bar{t}_2;\bar{t}_3)\in\mathbb{R}^2 \end{cases} \tag{5.20}$$

$l_{x_0}^{(1)}, l_{x_0}^{(2)}\in\mathbb{R}_+$ given by

$$l_{x_0}^{(1)} \;=\; \gamma_1 m + \overline{\mathcal{J}}_{x_0}\,, \qquad l_{x_0}^{(2)} \;=\; \gamma_1 \sum_{j=1}^{m} \phi_1\left(\left|(Dv_0)_j+(\bar{t}_2;\bar{t}_3)\right|;a\right) \;+\overline{\mathcal{J}}_{x_0}\,. \tag{5.21}$$

After noting that $l_{x_0}^{(2)}\leq l_{x_0}^{(1)}$, we complete the proof by demonstrating that, for any $\gamma_1,\gamma_2,\gamma_3$, $a\in\mathbb{R}_{++}$, $f\in\mathbb{R}^m$, $x_0\in\mathbb{R}^{4m}$, and any $(\bar{t}_1,\bar{t}_2)\in\mathbb{R}^2$, there exists a point $x=(v;w;g)\in\mathbb{R}^{4m}$ such that $\mathcal{J}(x)\leq l_{x_0}^{(2)}$. In fact, e.g., for

$$x \;=\; (v,w,g) \;=\; \big(v_0+D(\bar{t}_2h^{(2)}+\bar{t}_3h^{(3)})\,,\, w_0-D(\bar{t}_2h^{(2)}+\bar{t}_3h^{(3)})\,,\, g_0\big) \;\in\; \mathbb{R}^{4m}\,, \tag{5.22}$$

we have that $\mathcal{J}(x) \;=\; l_{x_0}^{(2)}$. ■

Noncoercivity of function $\mathcal{J}$ for v and w being opposite constant images is quite evident from the expression of $\mathcal{J}$ in (5.4). It is also evident that global (and also local) minimizers of $\mathcal{J}$ are defined modulo opposite constant offsets for v and w, namely, if $x^*=\big(v^*;w^*;g^*\big)$ is a minimizer for $\mathcal{J}$, then any $y^*(t)=\big(v^*+t\,1_m;w^*-t\,1_m;g^*\big)$, $t\in\mathbb{R}$, is also a minimizer. This property, which is typical of all variational decomposition models where the cartoon component is sought by promoting the sparsity of its gradients, makes our model in some way "redundant" and lets the employed optimization algorithm (when convergent) be a naive responsible for the selection of one among the infinity of equivalent minimizers. In order to make this selection more transparent, different strategies at the modeling level are possible. In Proposition 5.4 below we outline the most natural one, which ensures the equivalence between the original model and its modified version in terms of minimum cost function value.

Proposition 5.4. *For any* $\gamma_1,\gamma_2,\gamma_3,a>0$ *and any* $f\in\mathbb{R}^m$*, the function* $\mathcal{J}$ *in* (5.4) *is constant along straight lines in its domain* $\mathbb{R}^{4m}$ *of direction defined by the vector*

$$d \;:=\; \big(1_m;-1_m;0_{2m}\big)\,. \tag{5.23}$$

Hence, any constrained model of the form

$$x^* \in \operatorname*{argmin}_{x\,\in\,\mathcal{C}_{c,q}} \mathcal{J}(x) \tag{5.24}$$

with $\mathcal{C}_{c,q} \subset \mathbb{R}^{4m}$ *one among the infinity of* $(4m-1)$ *dimensional affine feasible sets defined by*

$$\mathcal{C}_{c,q} \;=\; \left\{ x \in \mathbb{R}^{4m} : \; c^T x \;=\; q \quad \text{with} \quad c^T d \neq 0 \right\}, \quad c \in \mathbb{R}^{4m}, \;\; q \in \mathbb{R}, \tag{5.25}$$

admits solutions and the solutions are equivalent to those of the unconstrained model (5.3)–(5.4) *in terms of (minimum) cost function value.*

Proof. Proving that $\mathcal{J}(x) = \mathcal{J}(x + t\,d)$ $\forall x \in \mathbb{R}^{4m}$ $\forall t \in \mathbb{R}$, with d defined in (5.23), is straightforward:

$$\mathcal{J}\big(x + t\,d\big) \;=\; \mathcal{J}(v + t\,1_m, w - t\,1_m, g + t\,0_{2m})$$

$$= \gamma_1 \sum_{j=1}^{m} \phi_1\left(|(Dv)_j + \cancel{(D(t\,1_m))_j}|; a \right) + \frac{\gamma_2}{2} |(Hw)_j - \cancel{(H(t\,1_m))_j}|^2 + \frac{\gamma_3}{2}\, \|g\|_2^4 \tag{5.26}$$

$$+ \frac{1}{2}\, \big\| f - (v + \cancel{t\,1_m} + w - \cancel{t\,1_m} + D^T g) \big\|_2^2 \;=\; \mathcal{J}(x) \quad \forall x \in \mathbb{R}^{4m}, \, \forall\, t \in \mathbb{R}\,, \tag{5.27}$$

where in (5.26) we fact (stated in Lemma 5.2) that constant images are mapped to the zero vector by matrices D and H. This implies that, if $x^* \in \mathbb{R}^{4m}$ is a minimizer of $\mathcal{J}$, the straight line $y^*(t) = x^* + t\,d$, $t \in \mathbb{R}$, contains an infinity of minimizers equivalent to x^* in terms of function value, namely, $\mathcal{J}\big(y^*(t)\big) = \mathcal{J}\big(x^*\big)$ $\forall\, t \in \mathbb{R}$.

We now notice that any $(4m-1)$-dimensional affine hyperplane $\mathcal{C}_{c,q}$ defined in (5.25) is not parallel to the vector d (due to condition $c^T d \neq 0$), hence it intersects any straight line of direction d in one and only one point. Therefore, if $y^*(t) = x^* + t\,d$, $t \in \mathbb{R}$, is a set of equivalent minimizers for $\mathcal{J}$, there is one and only one element of the set which also belongs to a feasible set $\mathcal{C}_{c,q}$ of the form in (5.25). Since according to Proposition 5.3 the function $\mathcal{J}$ admits global minimizers, it follows that the restriction of $\mathcal{J}$ to any feasible set $\mathcal{C}_{c,q}$ also admits global minimizers which are characterized by the same (minimum) cost function value. ■

Solving numerically constrained models of the form in (5.24)–(5.25) is slightly more complicated (and, in general, less efficient) than solving the unconstrained model (5.3)–(5.4). A second strategy that can be used to eliminate redundancy of the unconstrained model without imposing hard constraints consists in adding to our cost function $\mathcal{J}$ in (5.4) a "very small" regularization term capable of making $\mathcal{J}$ not constant on straight lines parallel to the vector d in (5.23). This is the strategy we use in the ADMM-based numerical solver presented in the following section 5.3, where we implicitly add (in the sense that we add the term by directly including a regularization matrix $\kappa\, I$ in the system of normal equations) the regularization term $(\kappa/2)\|x\|_2^2$ with $\kappa \in \mathbb{R}_{++}$ being a very small parameter.

5.3. Applying ADMM to the proposed model. In this section, we illustrate in detail the ADMM-based iterative algorithm used to numerically minimize the proposed unconstrained model (5.3)–(5.4) which presents a good separable structure. We first resort to the variable splitting technique to deal with the nondifferentiability of the nonconvex penalty term $\phi_1(\cdot; a)$. By introducing the auxiliary variable $t := Dv \in \mathbb{R}^{2m}$, we formulate the following constrained optimization problem,

$$\{v^*, w^*, g^*, t\} \;\leftarrow\; \arg\min_{v,w,g,t} \mathcal{J}(v, w, g, t) \quad \text{s.t.} \quad t = Dv,$$

with

$$\mathcal{J}(v,w,g,t) = \gamma_1 \sum_{j=1}^{m} \phi_1\left(|t_j|;a\right) + \frac{\gamma_2}{2} \sum_{j=1}^{m} |(Hw)_j|^2 + \frac{\gamma_3}{2}\|g\|_2^4 + \frac{1}{2}\left\| f - (v + w + D^T g)\right\|_2^2 .$$

The corresponding augmented Lagrangian functional for the optimization problem reads as

$$\begin{aligned} \mathcal{L}(v,w,g,t;\rho) = \gamma_1 \sum_{j=1}^{m} \phi_1\left(|t_j|;a\right) &+ \frac{\gamma_2}{2}\sum_{j=1}^{m} |(Hw)_j|^2 + \frac{\gamma_3}{2}\|g\|_2^4 \\ &- \langle \rho, t - Dv\rangle + \frac{\beta}{2}\|t - Dv\|_2^2 + \frac{1}{2}\left\| f - (v + w + D^T g)\right\|_2^2 , \end{aligned} \tag{5.28}$$

where $\beta > 0$ is a penalty scalar parameter, and $\rho \in \mathbb{R}^{2m}$ represents the vector of Lagrange multipliers associated with the linear constraint $t = Dv$.

To simplify notations, in the following we denote by $x := (v; w; g)$ the $(4m)$-dimensional column vector formed by stacking the three optimization variables v, $w \in \mathbb{R}^m$, $g \in \mathbb{R}^{2m}$. We then consider the following saddle-point problem:

$$\begin{aligned} &\text{find} \quad (x^*, t^*, \rho^*) \in \mathbb{R}^{4m} \times \mathbb{R}^{2m} \times \mathbb{R}^{2m} \\ &\text{s.t.} \;\; \mathcal{L}(x^*, t^*; \rho) \le \mathcal{L}\left(x^*, t^*; \rho^*\right) \le \mathcal{L}\left(x, t; \rho^*\right) \quad \forall (x,t,\rho) \in \mathbb{R}^{4m} \times \mathbb{R}^{2m} \times \mathbb{R}^{2m} . \end{aligned} \tag{5.29}$$

An ADMM-based iterative scheme is applied to approximate the solution of the saddle-point problem (5.28)–(5.29). Having zero-initialized vectors $t^{(0)}$ and $\rho^{(0)}$, the kth iteration of the proposed alternating iterative scheme reads as follows:

$$x^{(k+1)} = \operatorname*{argmin}_{x \in \mathbb{R}^{4m}} \mathcal{L}(x, t^{(k)}; \rho^{(k)}) \, , \tag{5.30}$$

$$t^{(k+1)} = \operatorname*{argmin}_{t \in \mathbb{R}^{2m}} \mathcal{L}(x^{(k+1)}, t; \rho^{(k)}) \, , \tag{5.31}$$

$$\rho^{(k+1)} = \rho^{(k)} - \beta(t^{(k+1)} - Dv^{(k+1)}) \, . \tag{5.32}$$

For the *x-subproblem* (5.30), the first-order optimality conditions read

$$\begin{pmatrix} (v^{(k+1)} + w^{(k+1)} + D^T g^{(k+1)} - f) + D^T \rho^{(k)} - \beta D^T (t^{(k)} - Dv^{(k)}) \\ (w^{(k+1)} + v^{(k+1)} + D^T g^{(k+1)} - f) + \gamma_2 H^T H w^{(k+1)} \\ D(D^T g^{(k+1)} + v^{(k+1)} + w^{(k+1)} - f) + 2\gamma_3 \left\| g^{(k+1)} \right\|_2^2 g^{(k+1)} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} . \tag{5.33}$$

By replacing the nonlinear term $\left\| g^{(k+1)} \right\|_2^2$ in the third equation with the value at the previous iteration k, (5.33) reduces to the following linear system of equations,

$$L x^{(k+1)} = y \, , \tag{5.34}$$

where

$$L = \begin{pmatrix} I + \beta D^T D & I & D^T \\ I & I + \gamma_2 H^T H & D^T \\ D & D & DD^T + 2\gamma_3 \left\| g^{(k)} \right\|_2^2 I \end{pmatrix}, \quad y = \begin{pmatrix} f + \beta D^T (t^{(k)} - \frac{1}{\beta}\rho^{(k)}) \\ f \\ Df \end{pmatrix} \tag{5.35}$$

which is solved for $x^{(k+1)} = (v^{(k+1)}; w^{(k+1)}; g^{(k+1)})^T$. The block in (5.35) containing the discretized operator $H^T H$ slightly worsens the conditioning of the linear system. A suitable approximate solution of (5.34) is determined by solving the following system of regularized equations,

$$(L + \kappa I)\, x^{(k+1)} = y \tag{5.36}$$

with a small scalar parameter $\kappa > 0$, that allows the system to be efficiently solved using iterative preconditioned conjugate gradient linear solver.

The *t-subproblem* (5.31) can be written omitting the constant terms as

$$t^{(k+1)} = \underset{t\in\mathbb{R}^{2m}}{\operatorname{argmin}} \left\{ \gamma_1 \sum_j \phi_1(|t_j|; a) - \langle \rho, t\rangle + \frac{\beta}{2}\|t - Dv\|_2^2 \right\}. \tag{5.37}$$

The minimization problem in (5.37), rewritten in componentwise form, is equivalent to the following m independent 2-dimensional problems of the form

$$t_j^{(k+1)} = \underset{t\in\mathbb{R}^{2}}{\operatorname{argmin}} \left\{ \frac{1}{\alpha}\phi_1(\|t\|_2; a) + \frac{1}{2}\|t - q_j\|_2^2 \right\} \tag{5.38}$$

with $j = 1, \dots, m$, $\alpha = \beta/\gamma_1$, and $q_j = (Dv^{(k)})_j + \rho_j^{(k)}/\beta$. Necessary and sufficient conditions for strong convexity of the cost functions in (5.38) are demonstrated in [9]. In particular, the problems in (5.38) are strongly convex if and only if the following condition holds:

$$a < \alpha \implies \beta > a\gamma_1 \implies \beta = \tau\,(a\gamma_1) \quad \text{for} \quad \tau > 1. \tag{5.39}$$

Under the assumption (5.39), the unique solutions of problems in (5.38) can be obtained in closed form as

$$t_j^{(k+1)} = \min(\max(\nu - \zeta/\|q_j\|_2, 0), 1)\, q_j,$$

where $\nu = \frac{\alpha}{\alpha - a}$ and $\zeta = \frac{\sqrt{2a}}{\alpha - a}$.

We remark that the condition on β in (5.28) only ensures the convexity conditions (5.39) of t-subproblem (5.38), but does not guarantee convergence of the overall ADMM scheme. In the case that convexity conditions (5.39) are satisfied, following [6], the convergence of the proposed two-block ADMM-based minimization algorithm could be investigated in future work.

In the numerical experiments, we set the coefficient τ in (5.39) to be $\tau = 5$, the β value to be 50, and κ in (5.36) to be $\kappa = 10^{-9}$. This setting has always guaranteed the ADMM convergence in our experiments.

6. Numerical examples.

In this section, we present experimental results on the additive decomposition of synthetic and real images. We first validate the proposed variational model (2.7) for the decomposition of images corrupted by an increasing level of additive Gaussian noise (Example 1) and different blending of the additive components (Example 2). Then, we investigate the performance of our model when applied to specific applications such as soft shadow removal and spotlight removal (Example 3). Finally, we compare our proposal to three interesting variational models, namely, [3], [9], and [18], proposed in literature for the additive decomposition of images (Example 4).

The code has been implemented in the MATLAB environment and Windows OS. When the ground truth images are known, quantitative measurements to evaluate the quality of the decomposition are provided by signal-to-noise ratio (SNR) values, defined by

$$\mathrm{SNR}(x^*, x) := 10 \log_{10} \left(\|x - E[x]\|_2^2 \;/\; \|x^* - x\|_2^2 \right)$$

with x^* the computed estimate of the original image x and $E[x]$ denoting the mean value of x.

For the oscillatory component n, we report the experimental standard deviation (ESD) of the signal defined as

$$ESD(n) = \frac{\|n\|_2}{\sqrt{m}} \,. \tag{6.1}$$

Relying on the discrepancy principle, given a known, or estimated, value for the noise standard deviation σ, the values of $ESD(n)$ should approach it. For all the experiments, we terminate the iterations of the ADMM algorithm as soon as two successive iterates satisfy either of the two following conditions:

$$k \,>\, 400\,, \qquad \left\| x^{(k+1)} - x^{(k)} \right\|_2 \,/\, \left\| x^{(k)} \right\|_2 \,<\, 10^{-6}\,. \tag{6.2}$$

Example 1: Degradation by different noise levels. We test the performance of our model (2.7) in decomposing a piecewise-smooth image f, shown in Figure 6.1, first column, obtained as a composition of a piecewise-constant rectangle shape v with smoothly varying illumination w, and different noise degradations n. The test image u, in Figure 6.1, top row, represents a nontrivial case in which, even if the w and v components of the noise-free image u are well-separable (in fact $|(Hw)_j| \in [0, 0.34]$ and $|(Dv)_j| \in [85.14, 120.41]$ for all j), the noise degradation corrupts not only the image but also the separability.

We set the model parameters following the discussion reported in section 4. In particular, we set $\bar{a} = 0.5 \min_j(|(Dv)_j|) = 42.57$ (see (2.11)); and consequently $a = 2/\bar{a}^2 = 0.001$, $\bar{t} = 10^{-3}\bar{a} = 0.043$; $\gamma_1 = 1/\#(|Dv|) = 1/1055$; $\gamma_2 = \gamma_2 \phi_1(\bar{t}, a)/\bar{t}^2 = 0.0012$; see (4.3). The parameter γ_3 is computed according to the formula (4.4), where σ_g is defined for the different levels of noise. In particular, the noisy images f, see Figure 6.1 first column, have been degraded by a Gaussian additive noise characterized by standard deviations in the range $\sigma = \{5, 25, 35, 100\}$ which lead to the following estimates of γ_3:

$$\begin{array}{llll}
\sigma \;= 5, & \sigma \;= 25, & \sigma \;= 35, & \sigma \;= 100, \\
\sigma_g = 3.49, & \sigma_g = 17.47, & \sigma_g = 24.46, & \sigma_g = 69.90, \\
\gamma_3 = 1/9.55e11, & \gamma_3 = 1/5.97e14, & \gamma_3 = 1/2.29e15, & \gamma_3 = 1/1.53e17.
\end{array}$$

The results are reported in each row of Figure 6.1 (from the second column) for increasing noise level σ. In each column, the resulting denoised image $u^* = v^* + w^*$, and its components

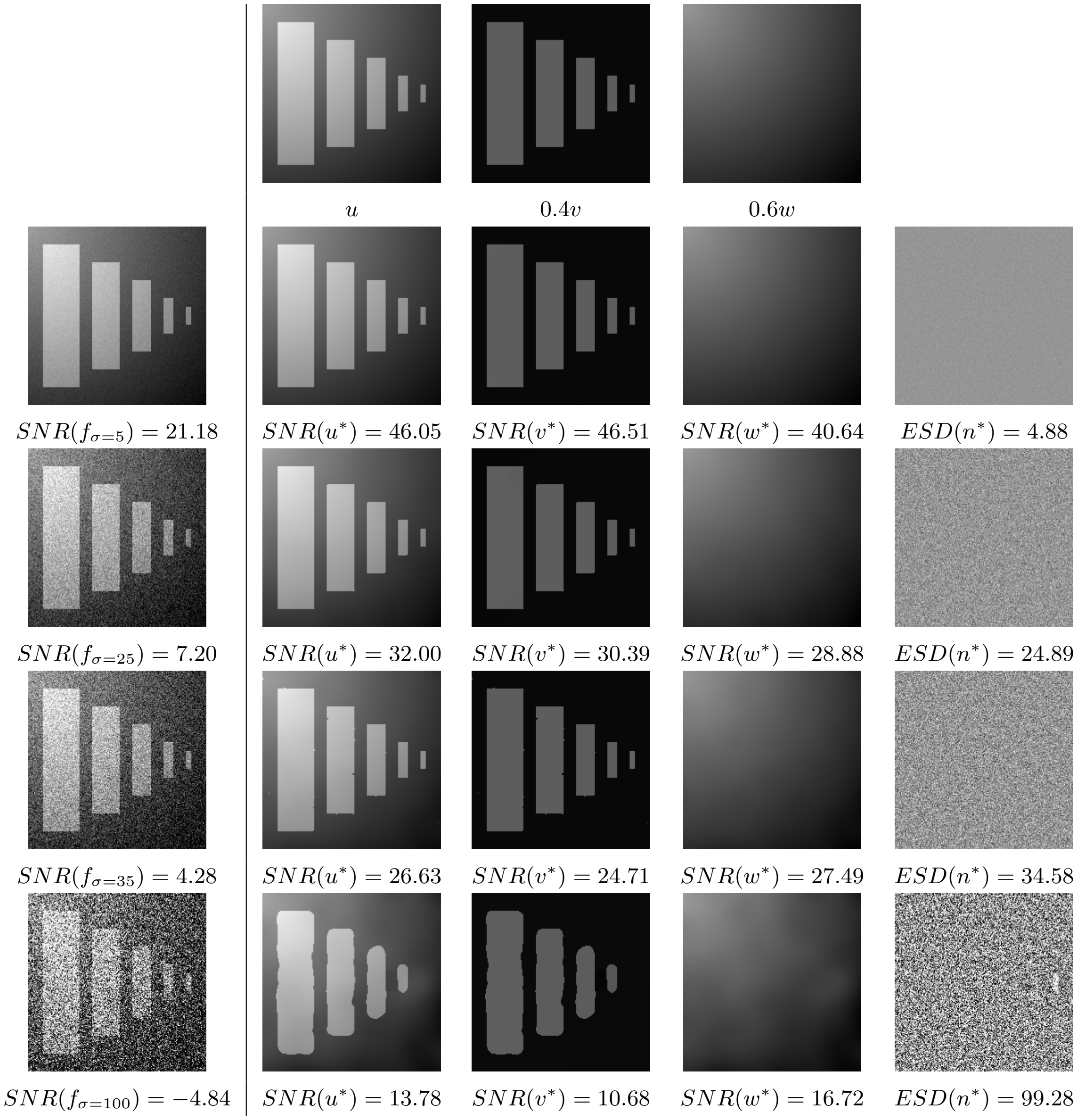
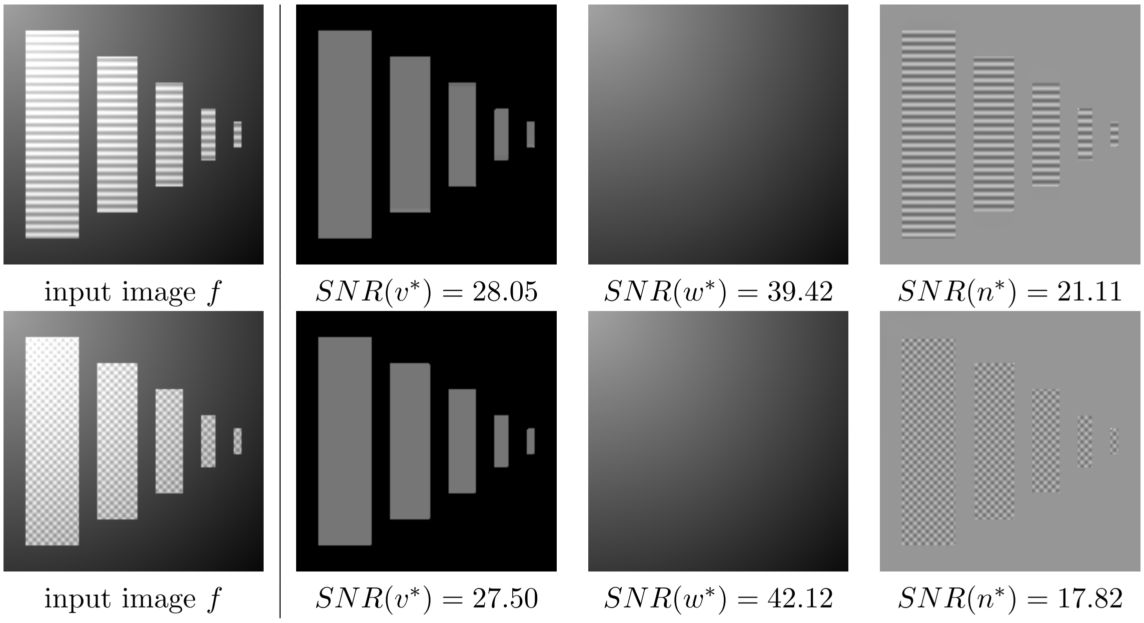


Figure 6.1. *Example* 1*: Decomposition results for different levels of additive Gaussian noise with* $\sigma = \{5, 25, 35, 100\}$*. The column* u^* *shows clean denoised image,* v^* *shows the piecewise-constant part,* w^* *the captured smooth part, and* n^* *the noise.*

v^*, w^*, and n^* are shown. Even for severely corrupted images, the effect of the proposed regularization term $\mathcal{R}_3(g)$, weighted by an appropriate parameter γ_3, well captures the noise oscillations. It achieves ESD values close to the noise standard deviation used to corrupt the original image. In the case of strong noise, as in the last row of Figure 6.1 where the noise is stronger than the edge magnitudes, the smallest rectangle goes into the n^* component.



input image f
$SNR(v^*) = 28.05$
$SNR(w^*) = 39.42$
$SNR(n^*) = 21.11$

input image f
$SNR(v^*) = 27.50$
$SNR(w^*) = 42.12$
$SNR(n^*) = 17.82$

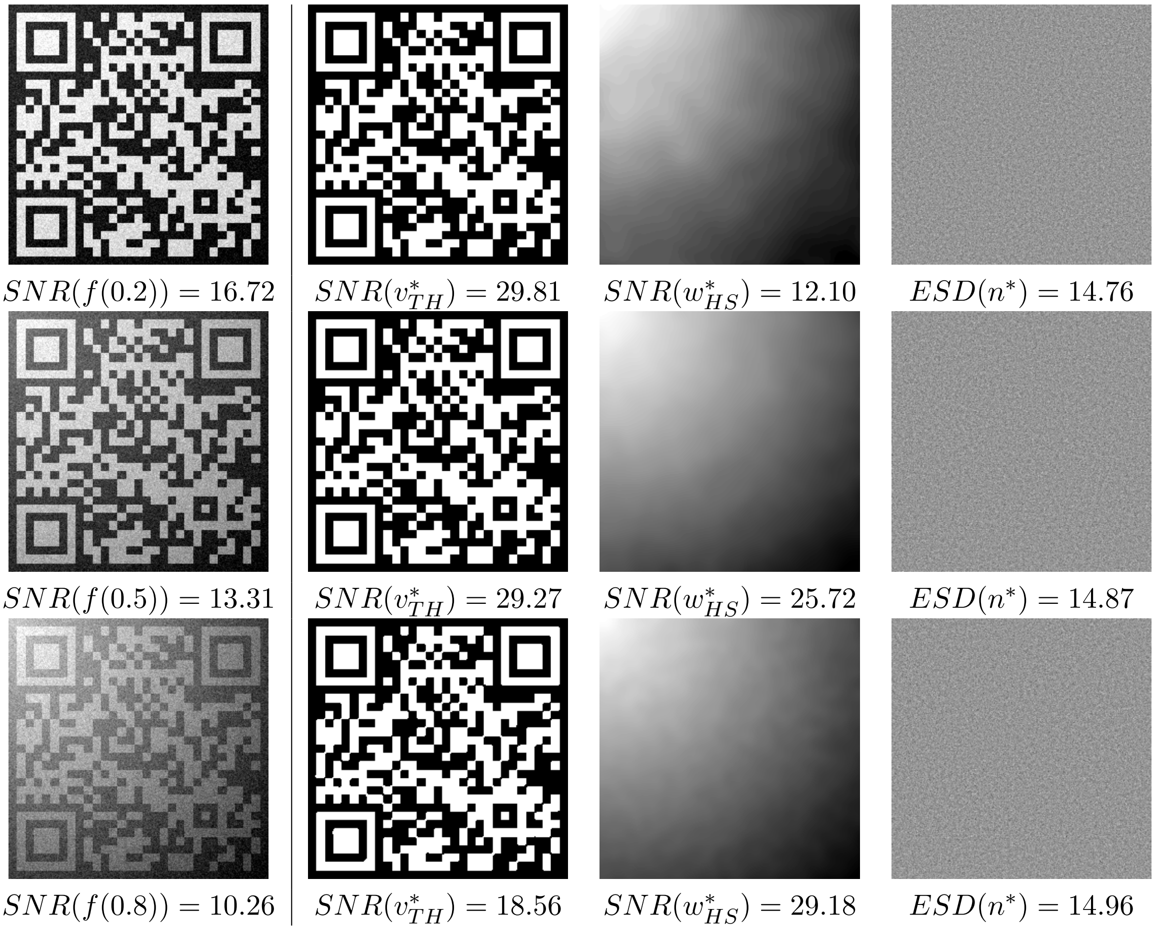
Figure 6.2. *Example 2: Decomposition results for different types of texture without noise. The column v^* shows the piecewise-constant part, w^* the captured smooth part, and the oscillatory texture is captured in n^*.*

Example 2: Different textures. In this example, we evaluate the performance of our decomposition model when the oscillatory function represents the texture component of a noise-free image (the first column of Figure 6.2). In particular, we kept the original components v and w as in the previous example adding a texture pattern inside the rectangle areas. In the first row of Figure 6.2 the texture is represented by horizontal stripes, while in the second row, the texture component is represented by a chessboard-like pattern. In the remaining columns of Figure 6.2 we report the resulting components v^*, w^*, and n^*, respectively. Both visually and quantitatively we can appreciate the texture captured entirely in n^* attaining the texture image $SNR(n^*) = \{21.11, 17.82\}$, respectively, keeping high quality reconstruction of both v^* and w^* as well.

Example 3: Different blending. We validate our model under different blendings of the v and w image components. This leads to different magnitudes of the first- and second-order derivatives and, consequently, affects the component separability. In Figure 6.3, the different images u are obtained by the linear mixtures of a piecewise component v, a QR code image, and a smooth component w, namely,

$$u(t) := (1-t)v + t\,w\,, \quad t \in [0,1]. \tag{6.3}$$

Then, they are corrupted by additive Gaussian noise of standard deviation $\sigma = 15$; the resulting degraded images $f(t)$ are shown in the first column of Figure 6.3 for $t = \{0.2, 0.5, 0.8\}$, respectively. For visualization purposes, the images illustrated in the second and third columns of Figure 6.3 have been slightly modified: the image v^*_{TH} is thresholded (TH) using the mean value of v^* as the threshold value, and the image w^*_{HS} is histogram stretched (HS).



$SNR(f(0.2)) = 16.72$ $SNR(v^*_{TH}) = 29.81$ $SNR(w^*_{HS}) = 12.10$ $ESD(n^*) = 14.76$

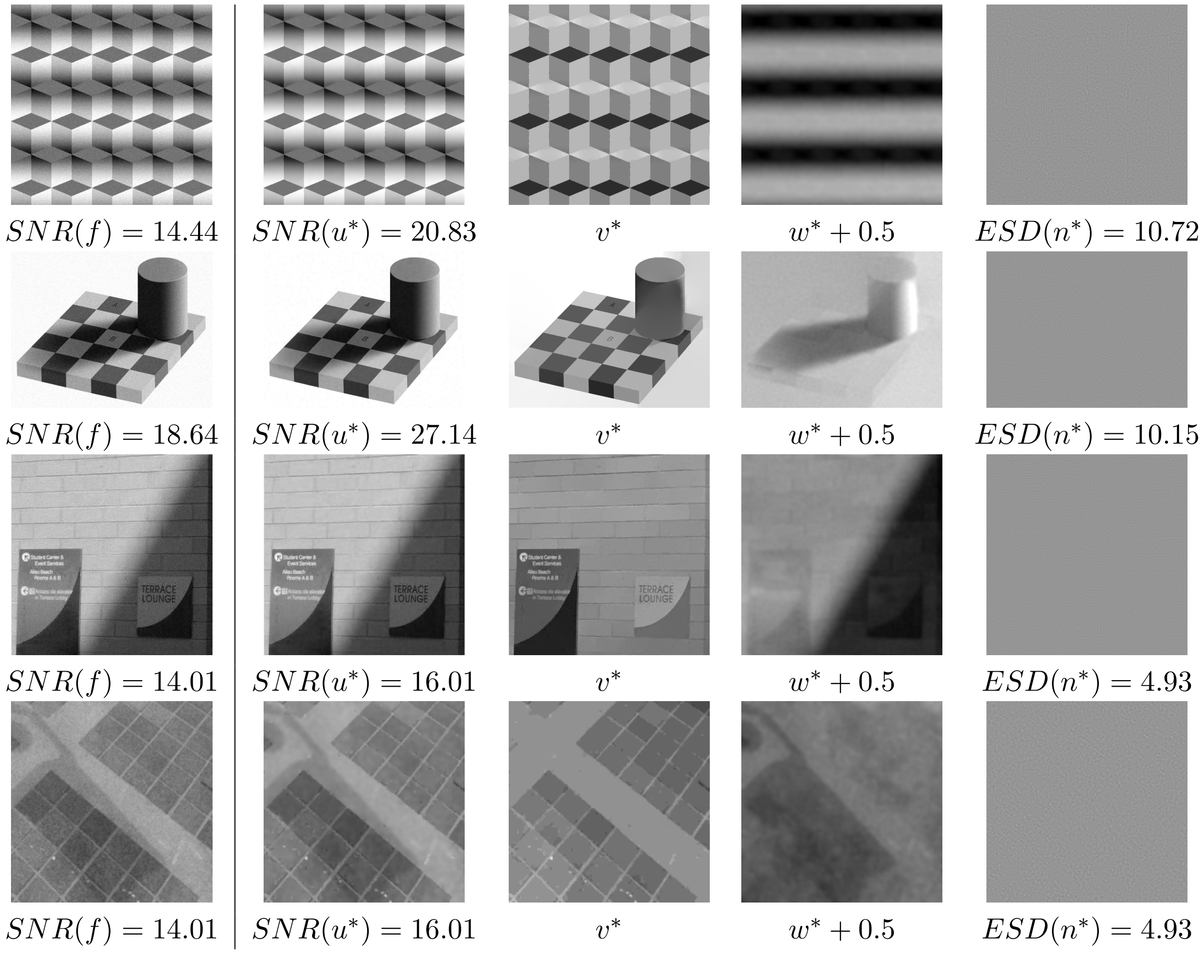
$SNR(f(0.5)) = 13.31$ $SNR(v^*_{TH}) = 29.27$ $SNR(w^*_{HS}) = 25.72$ $ESD(n^*) = 14.87$

$SNR(f(0.8)) = 10.26$ $SNR(v^*_{TH}) = 18.56$ $SNR(w^*_{HS}) = 29.18$ $ESD(n^*) = 14.96$

Figure 6.3. *Example 3: (first column) Noisy image f given by blending w and v according to (6.3) for $t = \{0.2, 0.5, 0.8\}$; (second column) v^* component, (third column) w^* component, (fourth column) n^* noise.*

For all the three blendings considered, the recovered noise component n^* approaches the standard deviation of the noise $\sigma = 15$, as illustrated in the last column. From top to bottom, the edges of the QR code images v become less significant with respect to the noise contribution. The noise interferes more with edges of v, thus achieving worse $SNR(v^*)$ results. On the other hand, the stronger the w component is, the better its recovery is, as highlighted by the increasing $SNR(w^*)$ values. This experiment demonstrates how the proposed model efficiently decomposes an image even when the contributions of each component in terms of gray-level intensities are not equally balanced.

Example 4: Soft light and shadow removal. Our decomposition model (2.7) can be successfully applied to remove soft shadows and soft light effects. Soft shadows and the "dual" effect of soft light are ubiquitous, but remain notoriously difficult to extract from photographs without damaging the underlying content. Currently shadow removal algorithms rely on shadow detection by an initial segmentation which turns out to be somewhat easy in natural scenes with umbra, but particularly difficult in the case of penumbra (soft shadow), even more



$SNR(f) = 14.44$ $SNR(u^*) = 20.83$ v^* $w^* + 0.5$ $ESD(n^*) = 10.72$

$SNR(f) = 18.64$ $SNR(u^*) = 27.14$ v^* $w^* + 0.5$ $ESD(n^*) = 10.15$

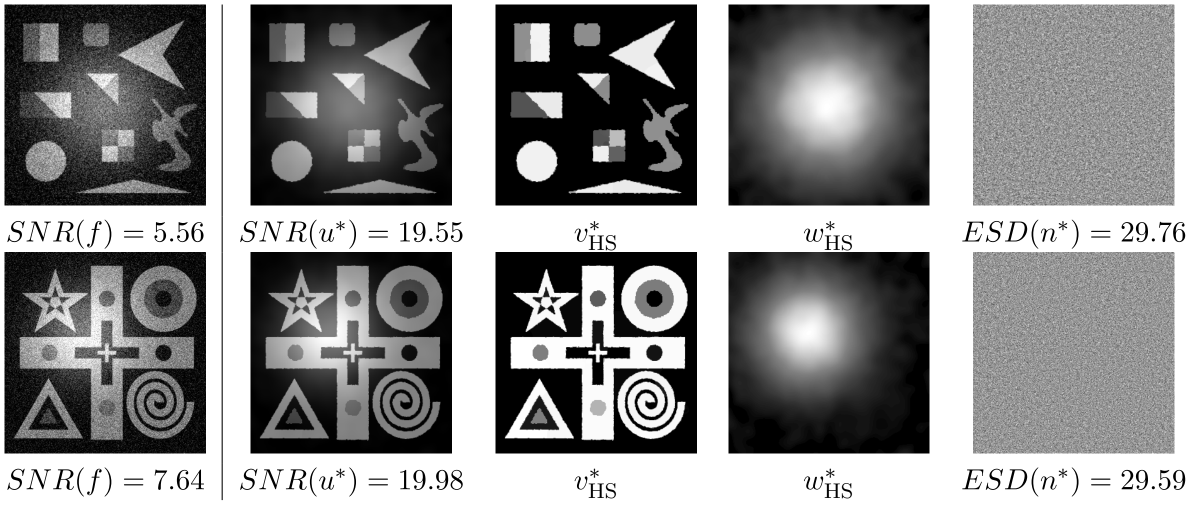
$SNR(f) = 14.01$ $SNR(u^*) = 16.01$ v^* $w^* + 0.5$ $ESD(n^*) = 4.93$

$SNR(f) = 14.01$ $SNR(u^*) = 16.01$ v^* $w^* + 0.5$ $ESD(n^*) = 4.93$

Figure 6.4. *Example 3: Soft shadow removal for noisy input images f in the first column. The second column shows denoised images u^*, and the third column shows the structure part of f in v^* components. The shadows are well separated in w^* and shown in the forth column, and the noise n^* is captured in the last column.*

challenging in the case of noisy images. For shadow removal, the proposed decomposition model (2.7) is directly applied to the noisy corrupted images, the smooth component w in (2.7) captures the shadow contribution, while the noise is separated in the n component, and the structures are enhanced in the v component.

In Figure 6.4 (first column) we show four images corrupted by additive Gaussian noise with standard deviation $\sigma = \{10, 10, 5, 5\}$, which represent the initial data f. In particular, the first two rows report images representing optical illusion examples. The change of lightning makes us perceive the shadowed objects/areas to be of different colors, even though in the image these parts have the same intensity. The first image is the Logvinenko illusion in which the top faces of the cubes have the same intensity value; and the second image is the Adelson's checkerboard illusion where the fields (squares) A and B have the same intensity value. These images are typically used in Retinex theory models [13, 19].



$SNR(f) = 5.56$ $SNR(u^*) = 19.55$ v^*_{HS} w^*_{HS} $ESD(n^*) = 29.76$

$SNR(f) = 7.64$ $SNR(u^*) = 19.98$ v^*_{HS} w^*_{HS} $ESD(n^*) = 29.59$

Figure 6.5. *Example 4: Soft light removal on noisy input images f in the first column. The second column u^* shows denoised image, and the third column v^* shows the structure part of f. The spotlight is well separated in w^* shown in the forth column.*

Figure 6.4 shows the separation result; the denoised image u^*, the piecewise-constant part v^*, the soft shadow in w^* where a constant 0.5 is added for visualization purposes, and the recovered noise n^*. As we expected the component w^* captured the soft shadow. The ESD values clearly indicate that the noise added to the images is accurately recovered. The soft cast shadow images illustrated in the last two rows of Figure 6.4 represent particularly difficult cases in which the input images f contain noise-like texture located on the tiles. Together with the image compression artifacts, the crevices “edges” between the tiles are numerically oversmoothed. The SNR values are thus slightly less significant in these cases; nonetheless, applying our decomposition algorithm produces visibly denoised u^*, and even correctly separates the soft shadow component w^*.

Finally, we demonstrate the performance of our decomposition model (2.7) on images f with a visible soft light effect which are corrupted by additive Gaussian noise with $\sigma = 30$ (first column of Figure 6.5). The result shows a clear separation: the denoised image $u^* = v^* + w^*$, the piecewise constant v^* and the spot light w^*, together with the noise component n^*. The proposed method separates the image in an excellent way, clearly showing the details in v^*, and a clear location of spotlights in w^*.

Example 5: Comparison to related works. We explore similarities and differences to some previous work, [3, 9, 18], where variational formulations are adopted for the additive decomposition of images.

In Figure 6.6, we compared our results with the two-component decomposition model introduced in [9], which similarly decomposes the observed noisy input f into v^* (cartoon) and w^* (smooth) components, treating the noise via the ℓ_2 fidelity residual. For comparison, we applied our decomposition model to the same synthetic image [9, Figure 8.5], illustrated in Figure 6.6 (first row) which has been corrupted by Gaussian noise. Increasing values of noise degradation, i.e., standard deviation $\sigma = \{5, 15\}$, are shown in Figure 6.6. The smaller

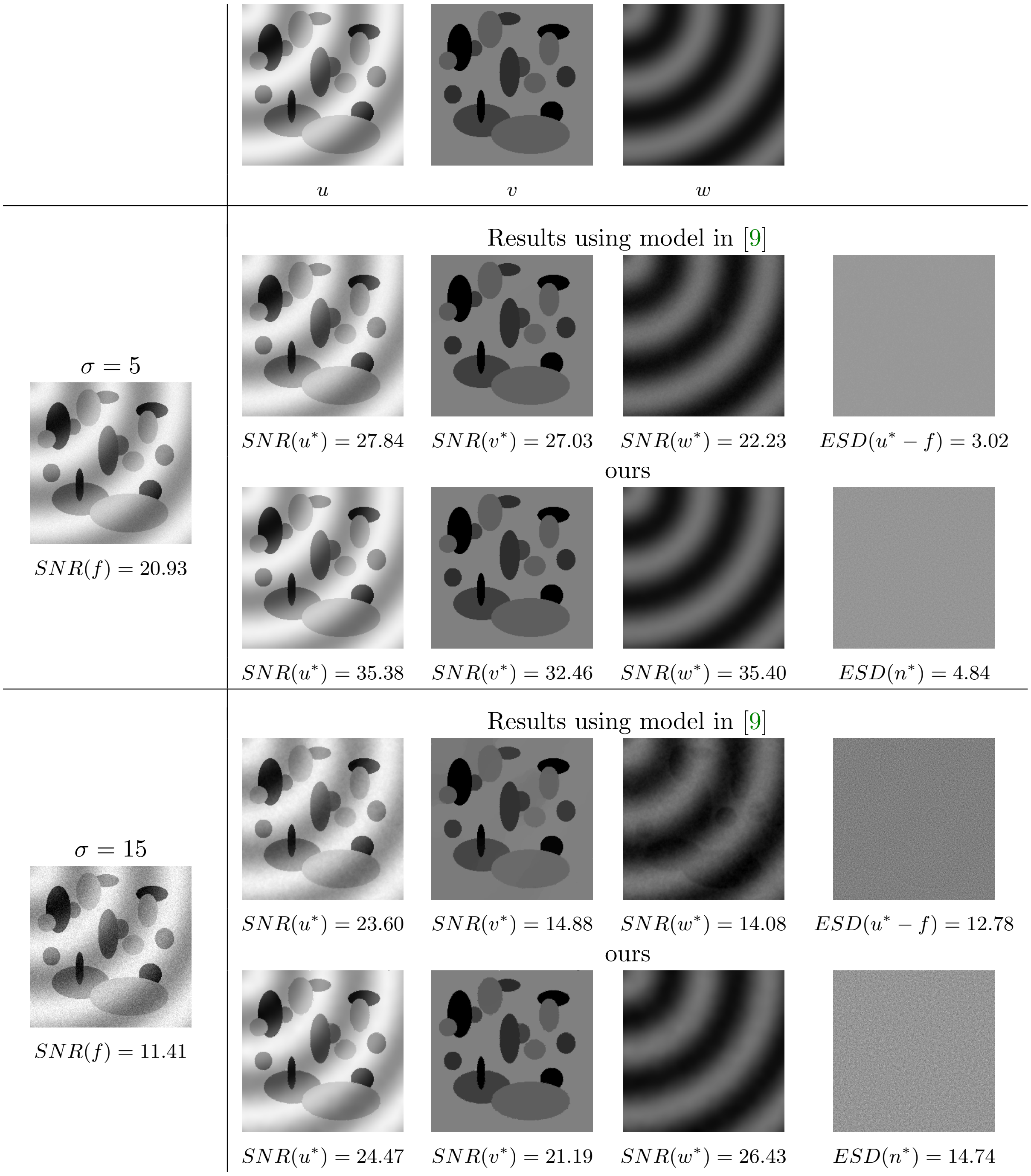
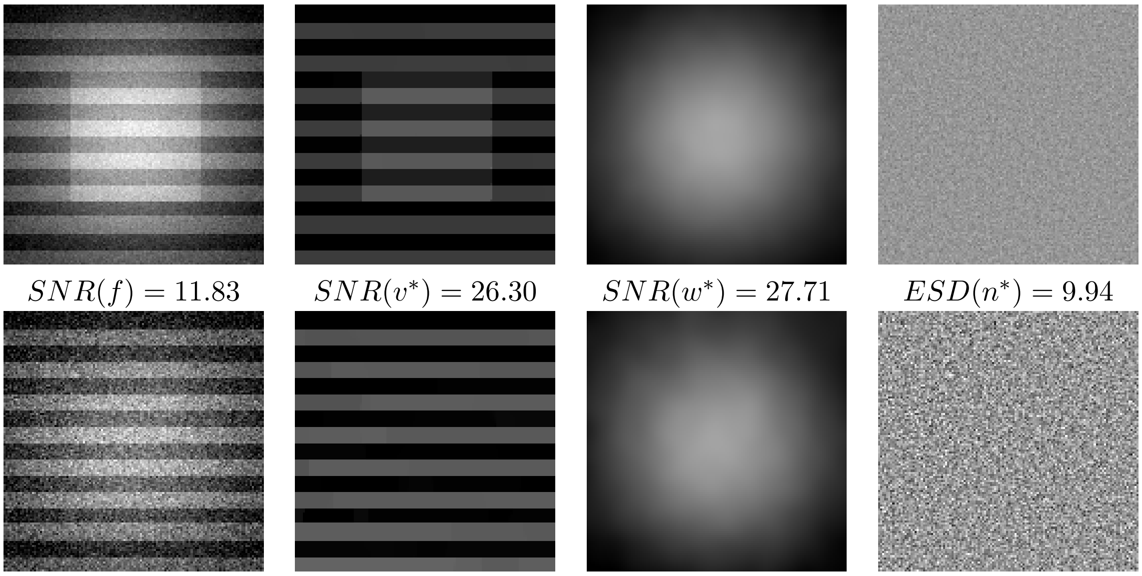


Figure 6.6. *Example 5: Comparison to [9] for $\sigma = \{5, 15\}$. For low levels of noise (top two rows), SNR values are higher for the proposed method. For a high level of noise $\gamma = 15$, not only are the SNR values higher for the proposed method, but also for [9] the residues are present in w^* and n^*.*



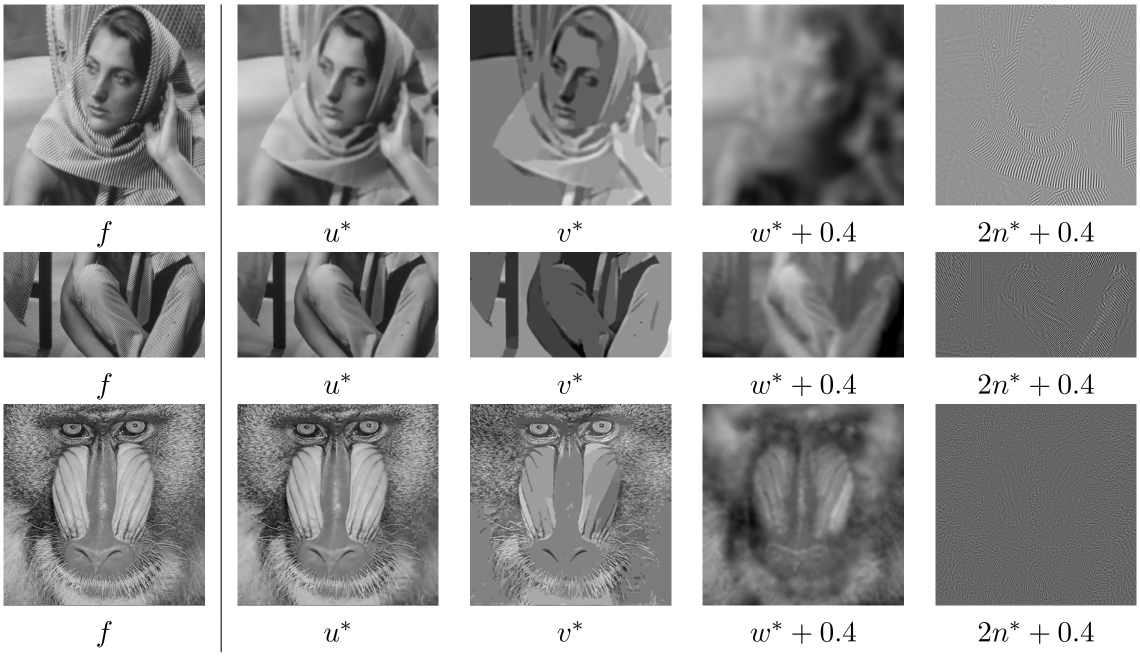
$SNR(f) = 11.83$ $SNR(v^*) = 26.30$ $SNR(w^*) = 27.71$ $ESD(n^*) = 9.94$

$SNR(f) = 4.97$ $SNR(v^*) = 19.05$ $SNR(w^*) = 17.70$ $ESD(n^*) = 34.29$

Figure 6.7. *Example* 5*: Comparison to* [3]*. These results are obtained by applying the proposed decomposition model. The model in* [3] *would decompose these images into structure, texture, and noise, thus the smooth function w^* would be incorporated into the structure component, resulting in staircase effects.*

level of noise $\sigma = 5$ corresponds to the one used in [9]. For low level of noise, results are well recovered also by the variational model in [9]. However, the SNR values attained by our model are higher and also the quality of the decomposition is improved: the third row of Figure 6.6 shows a much cleaner component w^*, validating the use of an H^{-1}-norm penalty to capture the oscillatory component instead of the ℓ_2 residual. For a higher level of noise, $\sigma = 15$, by applying [9] the noise component is not well-separable with respect to the v component, and the regularization term $\|\nabla w\|_2^2$ underperforms compared to the $\|\mathcal{H}w\|_2^2$ regularization term introduced in our model (2.7). This is confirmed by a visual comparison in the last two rows of Figure 6.6, where the smooth component w^* using [9] incorporates visible residuals from v and noise n. The combination of the higher-order regularization term on w and the H^{-1}-norm based term on n, poses an advantage of the proposed method.

In Figure 6.7, we represent the different separation capabilities of our model compared with the three-component variational model proposed in [3]. The model in [3] tackles the structure-texture-noise decomposition of an observed noisy input image, using TV, negative Sobolev, and negative Besov norms, respectively. We recreated the piecewise-constant synthetic image presented in [3] and added an additional smooth-gradient bell function w, as illustrated in the first row of Figure 6.7. The model in [3] decomposes the noisy corrupted image into the square in the center as the structure component, the horizontal stripes as the texture component, and a separate noise component. Our proposed model instead, separates the structure component v which combines the square and the stripes, the noise term n, and also the smooth component w. A similar example without the square box is shown in the second row of Figure 6.7, where v^* contains only horizontal stripes and a higher noise degradation. The proposed model



f u^* v^* $w^* + 0.4$ $2n^* + 0.4$

f u^* v^* $w^* + 0.4$ $2n^* + 0.4$

f u^* v^* $w^* + 0.4$ $2n^* + 0.4$

Figure 6.8. *Example 5: Structure-texture decomposition, comparison to* [18]. *From top to bottom we set* $\gamma_1/\gamma_3 = \{0.25, 0.25, 15\}$. *In these images,* n^* *captures the fine detailed textures.*

separates them clearly, while the model in [3] recovers the smooth function w into the v^* component producing inevitable staircases effects.

In the last example illustrated in Figure 6.8, we compare our decomposition model with the two-term decomposition model in [18]. The authors proposed a structure-texture decomposition based on [23] with the TV regularizer replaced by a parameterized nonconvex penalty function. Comparing the two models term by term, the structure component term in [18] is similar to our $\mathcal{R}_1(v)$ and their texture term is similar to the H^{-1}-norm based term $\mathcal{R}_3(g)$. The additional term $\mathcal{R}_2(w)$ makes the difference, allowing for a more significant structure decomposition $u = v + w$, a piecewise-smooth contribution, which better fits the mixture composition of real images. In Figure 6.8, the comparisons to [18] are shown. Ad hoc γ_3 values are selected to enable $\mathcal{R}_3(g)$ to capture the correct texture scales, while taking into account the influence of the relative ratio γ_1/γ_3 on the final decomposition result, as shown in section 4. For the three examples of piecewise-smooth images considered we set the parameter ratios $\gamma_1/\gamma_3 = \{0.25, 0.25, 15\}$. The first and second rows of Figure 6.8 show two results for two different parts of the Barbara image where we set $\gamma_1/\gamma_3 = 0.25$ which allows us to capture small scale texture in n^*, while in v^* only the main structure is kept. The image in the third row of Figure 6.8 is characterized by the amount of fur texture which overwhelms the face feature edges, thus even for the increased ratio $\gamma_1/\gamma_3 = 15$ the fur texture is not entirely captured in n^*.

The H^{-1}-norm term captures the oscillating patterns, this includes both noise and textures at a sufficiently small scale detail. This example demonstrates how to successfully apply our model to separate textured regions from noise-free images.

7. Conclusion.

We proposed a new method to decompose a given image f into piecewise-constant, smooth homogeneous, and oscillating components which can represent noise and/or textured parts. The variational model is composed of a TV-like nonconvex regularization, to capture the piecewise-constant part, a new harmonic term, and an H^{-1}-norm based penalty which captures the oscillating patterns. This includes both noise and textures at a sufficiently small scale detail. In-depth experimental analysis gave us an insight on the nature of the three different metrics (norms) involved in the proposed variational model, allowing for an interesting highlight on the interactions between these regularization terms. This led to automatic selection of free parameters in the cost functional. A theoretical analysis on the coercivity and convexity of the proposed variational model highlighted the existence of global minimizers. The variational model is then solved by an efficient ADMM-based algorithm which reduces the solution to a sequence of convex optimization subproblems. Various experiments are presented to show the robustness against a high level of noise, flexibility of decomposition for various applications such as soft light and soft shadow removal.

Appendix A. Proof of Proposition 5.1.

Proof. It comes immediately from the definition of our cost function $\mathcal{J}$ in (5.4) and of the penalty function ϕ_1 in (2.8) that $\mathcal{J}$ is proper, continuous, and bounded from below by zero.

In order to demonstrate that the function $\mathcal{J}$ defined in (5.4) can never be convex jointly in $(v; w; g)$ with $g = (g^h; g^v)$, it is sufficient to prove that the restriction of $\mathcal{J}$ to a straight line in its domain $\mathbb{R}^{4m}$ is not convex for any $\gamma_1, \gamma_2, \gamma_3, a \in \mathbb{R}_{++}$ and any $f \in \mathbb{R}^m$. In particular, we choose the straight line parameterized as follows:

$$\ell(t) \;=\; t\,\Big(\,\underbrace{1; -1; 0_{m-2}}_{v}\,;\,\underbrace{0_m}_{w}\,;\,\underbrace{0_m}_{g^h}\,;\,\underbrace{1; 0_{m-1}}_{g^v}\,\Big)\,, \qquad t \in \mathbb{R}\,. \tag{A.1}$$

To derive an explicit expression for the restriction $\mathcal{J}_\ell(t)$, we first introduce details on the two submatrices D_h, D_v of $D = (D_h^T, D_v^T)^T$ which, as described in section 5.1, represent the forward finite difference discretizations of the first-order horizontal and vertical partial derivatives of a vectorized $m_1 \times m_2$ image (with $m = m_1 m_2$), respectively. More precisely, matrices $D_h, D_v \in \mathbb{R}^{m \times m}$ are defined by

$$D_h \;=\; L_{m_1} \otimes I_{m_2}\,, \qquad D_v \;=\; I_{m_2} \otimes L_{m_1}\,, \qquad L_n \;=\; \begin{pmatrix} -1 & 1 & 0 & 0 & \ldots \\ 0 & -1 & 1 & 0 & \ldots \\ 0 & 0 & -1 & 1 & \\ \vdots & \vdots & & \ddots & \ddots \end{pmatrix}, \tag{A.2}$$

where $\otimes$ is the Kronecker product operator and $L_n \in \mathbb{R}^{n \times n}$ denotes the unscaled forward finite difference operator approximating the first-order derivative of an n-sample 1-dimensional signal. We have not specified the last row of L_n, which induces the boundary conditions for operators D_h and D_v, since the following proof holds true independently of the chosen boundary conditions.

We note that, for v, g^h, g^v and D_h, D_v defined as in (A.1) and (A.2), respectively, we have

$$(Dv)_1 = t\,(-1; -2)\,, \qquad (Dv)_2 = t\,(1; 1)\,, \qquad (Dv)_j = \big((D_h v)_j; (D_v v)_j\big) = 0 \;\; \text{for } j = 3, \ldots, M,$$

$$\|g\|_2^4 = \|g^v\|_2^4 = t^4\,, \qquad D^T g = D_v^T g^v = t\,\left(-1; 1; 0_{m-2}\right)\,. \tag{A.3}$$

Hence, placing (A.1) into the expression (5.4) of function $\mathcal{J}$ and then using (A.3), we have that the restriction $\mathcal{J}_\ell$ of $\mathcal{J}$ to the straight line in (A.1) reads

$$\begin{aligned}\mathcal{J}_\ell(t) &= \frac{1}{2}\|f\|_2^2 + \frac{\gamma_3}{2}t^4 + \gamma_1\left(\phi_1\left(\sqrt{5}\,|t|;a\right) + \phi_1\left(\sqrt{2}\,|t|;a\right)\right)\\ &= \frac{1}{2}\|f\|_2^2 + \frac{\gamma_3}{2}t^4 + \gamma_1 \times \begin{cases} -\dfrac{7\,a}{2}\,t^2 + \left(\sqrt{5}+\sqrt{2}\right)\sqrt{2a}\,|t| & \text{for } |t| \in \left[0, \sqrt{\frac{2}{5a}}\right),\\ 1 - a\,t^2 + 2\sqrt{2a}\,|t| & \text{for } |t| \in \left[\sqrt{\frac{2}{5a}}, \sqrt{\frac{1}{a}}\right),\\ 2 & \text{for } |t| \in \left[\sqrt{\frac{1}{a}}, +\infty\right),\end{cases}\end{aligned} \tag{A.4}$$

where (A.4) comes easily from the definition of ϕ_1 in (2.8). It is then immediate to note that $\mathcal{J}_\ell$ is infinitely many times differentiable for $|t| \in \left(0, \sqrt{2/(5a)}\right)$ and that the second-order derivative $\mathcal{J}_\ell''$ satisfies

$$\begin{aligned}&\forall\,\gamma_2 \in \mathbb{R}_{++}\,\forall\,f \in \mathbb{R}^m, \quad |t| \in \left(0, \sqrt{\frac{2}{5a}}\right)\\ &\quad\Longrightarrow\; \mathcal{J}_\ell''(t) \;=\; 6\,\gamma_3\,t^2 - 7\,\gamma_1\,a \;<\; 0 \quad \text{for} \quad |t| < \sqrt{\frac{7}{6}\,\frac{\gamma_1 a}{\gamma_3}}.\end{aligned} \tag{A.5}$$

Here it follows from (A.4) and (A.5) that

$$\begin{aligned}&\forall\,\gamma_1,\gamma_2,\gamma_3,a \in \mathbb{R}_{++}\,\forall\,f \in \mathbb{R}^m,\\ &\quad\exists\,\delta \in \mathbb{R}_{++}: \;\; \mathcal{J}_\ell''(t) < 0 \;\;\forall\,|t| \in (0,\delta)\,, \quad \delta = \min\left\{\sqrt{\frac{7}{6}\,\frac{\gamma_1\,a}{\gamma_3}}, \sqrt{\frac{2}{5a}}\right\}.\end{aligned} \tag{A.6}$$

From (8), we can conclude that the restriction $\mathcal{J}_\ell$ in (A.4) and, hence, the total function $\mathcal{J}$ in (5.4), are nonconvex (in the variables t and $(v;w;g)$, respectively) for any $\gamma_1,\gamma_2,\gamma_3,a \in \mathbb{R}_{++}$ and any $f \in \mathbb{R}^m$.

In order to analyze convexity of the function $\mathcal{J}$ in (5.4) separately with respect to the variables v, w, and g, first we rewrite $\mathcal{J}$ in the three following equivalent forms:

$$\mathcal{J}(v,w,g) \;=\; \overbrace{\gamma_1\sum_{j=1}^{m}\phi_1\left(|(Dv)_j|;a\right) + \frac{1}{2}\,\|v\|_2^2}^{\mathcal{J}_v(v)} \;+\; \mathcal{A}_v(v,w,g)\,, \tag{A.7}$$

$$\mathcal{J}(v,w,g) \;=\; \overbrace{\frac{\gamma_2}{2}\,\|Hw\|_2^2 + \frac{1}{2}\,\|w\|_2^2}^{\mathcal{J}_w(w)} \;+\; \mathcal{A}_w(v,w,g)\,, \tag{A.8}$$

$$\mathcal{J}(v,w,g) \;=\; \overbrace{\frac{\gamma_3}{2}\,\|g\|_2^4 + \frac{1}{2}\,\left\|D^T g\right\|_2^2}^{\mathcal{J}_g(g)} \;+\; \mathcal{A}_g(v,w,g)\,, \tag{A.9}$$

where $\mathcal{A}_v$, $\mathcal{A}_w$, $\mathcal{A}_g$, in (A.7)–(A.9) are affine functions of variables v, w, g, respectively. It follows that $\mathcal{J}$ is convex in the variables v, w, g if and only if $\mathcal{J}_v$, $\mathcal{J}_w$, $\mathcal{J}_g$, in (A.7)–(A.9) are convex in the same variables.

First, we notice that the quadratic function $\mathcal{J}_w$ in (A.8) is clearly strongly convex for any $\gamma_2 \in \mathbb{R}_{++}$, whereas the quartic function $\mathcal{J}_g$ in (A.9) is strictly convex (and coercive) for any $\gamma_3 \in \mathbb{R}_{++}$. Hence, the total function $\mathcal{J}$ is strongly convex in the variable w and strictly convex (and coercive) in the variable g for any $\gamma_1, \gamma_2, \gamma_3, a \in \mathbb{R}_{++}$ and any $f \in \mathbb{R}^m$.

Deriving convexity conditions for $\mathcal{J}_v$ in (A.7) is less straightforward. First, we rewrite the penalty function $\phi_1(\,\cdot\,; a) : \mathbb{R}_+ \to \mathbb{R}$ defined in (2.8) in the following equivalent form

$$\phi_1(t;a) \;=\; -\,\frac{a}{2}\,t^2 \;+\; q(t;a)\,, \qquad q(t;a) \;:=\; \begin{cases} \sqrt{2a}\,t & \text{for } \; t \in \big[\,0, \sqrt{2/a}\,\big), \\ \dfrac{a}{2}\,t^2 + 1 & \text{for } \; t \in \big[\,\sqrt{2/a}, +\infty\,\big) \end{cases} \tag{A.10}$$

where the introduced function $q(\,\cdot\,; a) : \mathbb{R}_+ \to \mathbb{R}$ is clearly continuous, convex, and monotonically increasing on its entire domain $\mathbb{R}_+$ for any $a \in \mathbb{R}_{++}$. Then, based on (A.10), we can rewrite the function $\mathcal{J}_v$ in (A.8) as follows,

$$\begin{aligned} \mathcal{J}_v(v) \;&=\; \frac{1}{2}\,\|v\|_2^2 \;+\gamma_1 \sum_{j=1}^{m} \Big(-\frac{a}{2}\,|(Dv)_j|^2 + q\left(|(Dv)_j|; a\right) \Big) \\ &=\; \frac{1}{2}\left(\|v\|_2^2 - \gamma_1 a\,\|Dv\|_2^2 \right) \;+\gamma_1 \sum_{j=1}^{m} q\left(|(Dv)_j|; a\right) \\ &=\; \underbrace{\frac{1}{2}\,v^T\left(I_m - \gamma_1 a\, D^T D\right) v}_{\mathcal{J}_v^{(1)}(v)} \;+\; \underbrace{\gamma_1 \sum_{j=1}^{m} q\left(z_j(v); a\right), \quad z_j(v) := |(Dv)_j|, \;\; j = 1, \ldots, m,}_{\mathcal{J}_v^{(2)}(v)} \end{aligned} \tag{A.11}$$

where the introduced functions $z_j : \mathbb{R}^m \to \mathbb{R}_+$, $j = 1, \ldots, m$, are all continuous and convex on their entire domain $\mathbb{R}^m$. Each term of the summation defining $\mathcal{J}_v^{(2)}$ in (A.11) is convex (as the composition of a convex function, z_j, and a convex monotonically increasing function, q). It follows that $\mathcal{J}_v^{(2)}$ is convex for any $\gamma_1 \in \mathbb{R}_{++}$ and, hence, the total function $\mathcal{J}_v$ in (A.11) is convex (strongly convex) if the quadratic function $\mathcal{J}_v^{(1)}$ is convex (strongly convex) or, equivalently, the Hessian matrix $I_m - \gamma_1 a\, D^T D$ of $\mathcal{J}_v^{(1)}$ is positive semidefinite (definite). By introducing the eigenvalue decomposition of the symmetric positive semidefinite matrix $D^T D \in \mathbb{R}^{m \times m}$,

$$D^T D \;=\; V^T \Lambda V\,, \qquad \Lambda \;=\; \mathrm{diag}(\lambda_1, \ldots, \lambda_m)\,, \qquad V^T V = V V^T = I_m \tag{A.12}$$

with λ_i, $i = 1, \ldots, m$, indicating the real nonnegative eigenvalues of $D^T D$, we then have

$$\begin{aligned} I_m - \gamma_1 a\, D^T D \;&=\; V^T V - \gamma_1 a\, V^T \Lambda V \;=\; V^T\left(I_m - \gamma_1 a\,\Lambda\right) V \\ &=\; V^T \mathrm{diag}\left(1 - \gamma_1 a\,\lambda_1, \ldots, 1 - \gamma_1 a\,\lambda_m\right)\, V\,. \end{aligned} \tag{A.13}$$

Since V is orthogonal, then the Hessian matrix in (8) is positive semidefinite (definite) if and only if the diagonal matrix in (8) is positive semidefinite (definite), that is, if and only if

(A.14)

$$1 - \gamma_1 a \, \lambda_j \ \geq \ (>) \ 0 \ \forall j = 1, \ldots, m \iff 1 - \gamma_1 a \, \lambda_{\max} \geq (>) \ 0 \iff a \leq (<) \ \frac{1}{\gamma_1 \lambda_{\max}} \, .$$

This proves condition (5.5), with the equivalent formulation in terms of the coefficient τ following straightforwardly. We finally note that for a gradient discretization matrix $D = (D_h^T, D_v^T)^T$ defined as in section 5.1, or, more formally, as in (A.2), the eigenvalues of matrix $D^T D$ are upper bounded by $\lambda_{\max} = 8$. This can be easily derived by using Gershgorin's theorem [27]; we refer the reader to [6] for details. ■

Acknowledgment. We greatly appreciate the anonymous referees for their valuable and useful comments.

REFERENCES

[1] A. G. Asuero, A. Sayago, and A. G. González, *The correlation coefficient: An overview*, Crit. Rev. Anal. Chem., 36 (2006), pp. 41–59, https://doi.org/10.1080/10408340500526766.

[2] J. Aujol, G. Aubert, L. Blanc-Féraud, and A. Chambolle, *Image decomposition into a bounded variation component and an oscillating component*, J. Math. Imaging Vision, 22 (2005), pp. 71–88, https://doi.org/10.1007/s10851-005-4783-8.

[3] J. Aujol and A. Chambolle, *Dual norms and image decomposition models*, Int. J. Comput. Vis., 63 (2005), pp. 85–104, https://doi.org/10.1007/s11263-005-4948-3.

[4] J. Aujol, G. Gilboa, T. Chan, and S. Osher, *Structure-texture image decomposition—modeling, algorithms, and parameter selection*, Int. J. Comput. Vis., 67 (2006), pp. 111–136, https://doi.org/10.1007/s11263-006-4331-z.

[5] J. Aujol and S. Kang, *Color image decomposition and restoration*, J. Vis. Commun. Image Represent., 17 (2006), pp. 916–928.

[6] R. Chan, A. Lanza, S. Morigi, and F. Sgallari, *Convex non-convex image segmentation*, Numer. Math., 138 (2018), pp. 635–680, https://doi.org/10.1007/s00211-017-0916-4.

[7] Y. Feng, M. Wen, J. Zhang, F. Ji, and G.-X. Ning, *Sum of arbitrarily correlated gamma random variables with unequal parameters and its application in wireless communications*, in Proceedings of the 2016 International Conference on Computing, Networking and Communications (ICNC), IEEE, Piscataway, NJ, 2016, https://doi.org/10.1109/ICCNC.2016.7440693.

[8] A. Gholami and S. Hosseini, *A balanced combination of Tikhonov and total variation regularizations for reconstruction of piecewise-smooth signals*, Signal Process., 93 (2013), pp. 1945–1960, https://doi.org/10.1016/j.sigpro.2012.12.008.

[9] M. Huska, A. Lanza, S. Morigi, and I. Selesnick, *A convex-nonconvex variational method for the additive decomposition of functions on surfaces*, Inverse Problems, 35 (2019), 124008, https://doi.org/10.1088/1361-6420/ab2d44.

[10] M. Huska, A. Lanza, S. Morigi, and F. Sgallari, *Convex non-convex segmentation of scalar fields over arbitrary triangulated surfaces*, J. Computational and Applied Mathematics, 349 (2019), pp. 438–451, https://doi.org/10.1016/j.cam.2018.06.048.

[11] A. Lanza, S. Morigi, I. Selesnick, and F. Sgallari, *Nonconvex nonsmooth optimization via convex–nonconvex majorization–minimization*, Numer. Math., 136 (2017), pp. 343–381, https://doi.org/10.1007/s00211-016-0842-x.

[12] A. Lanza, S. Morigi, and F. Sgallari, *Convex image denoising via non-convex regularization*, in Proc. Scale Space and Variational Methods in Computer Vision (SSVM), J.-F. Aujol, M. Nikolova, and N. Papadakis, eds., Springer, Cham, Switzerland, 2015, pp. 666–677, https://doi.org/10.1007/978-3-319-18461-6_53.

[13] H. Le and D. Samaras, *Shadow removal via shadow image decomposition*, in Proceedings of the IEEE International Conference on Computer Vision (ICCV), IEEE Computer Society, Los Alamitos, CA, 2019.

[14] T. M. Le and L. A. Vese, *Image decomposition using total variation and div*(BMO), Multiscale Model. Simul., 4 (2005), pp. 390–423, https://doi.org/10.1137/040610052.

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

[16] S. Levine, *An adaptive variational model for image decomposition*, in Energy Minimization Methods in Computer Vision and Pattern Recognition, A. Rangarajan, B. Vemuri, and A. L. Yuille, eds., Springer, Berlin, 2005, pp. 382–397.

[17] L. Lieu and L. Vese, *Image restoration and decomposition via bounded total variation and negative Hilbert-Sobolev spaces*, Appl. Math. Optim., 58 (2008), 167, https://doi.org/10.1007/s00245-008-9047-8.

[18] T. Liming, Z. Honglu, H. Chuanjiang, and F. Zhuang, *Non-convex and non-smooth variational decomposition for image restoration*, Appl. Math. Model., 69 (2019), pp. 355–377, https://doi.org/10.1016/j.apm.2018.12.021.

[19] W. Ma, J. Morel, S. Osher, and A. Chien, *An l1-based variational model for retinex theory and its application to medical images*, in CVPR 2011, IEEE, Piscataway, NJ, 2011, pp. 153–160, https://doi.org/10.1109/CVPR.2011.5995422.

[20] Y. Meyer and D. Lewis, *Oscillating Patterns in Image Processing and Nonlinear Evolution Equations: The Fifteenth Dean Jacqueline B. Lewis Memorial Lectures*, Mem. Am. Math. Soc., American Mathematical Society, Providence, RI, 2001.

[21] J. M. Morel, P. A. Belén, and C. Sbert, *A PDE formalization of retinex theory*, IEEE Trans. Image Process., 19 (2010), pp. 2825–2837.

[22] M. Nikolova, *Energy minimization methods*, in Handbook of Mathematical Methods in Imaging, O. Scherzer, ed., Springer, New York, 2011, pp. 138–186.

[23] S. Osher, A. Solé, and L. Vese, *Image decomposition and restoration using total variation minimization and the* H^{-1} *norm*, Multiscale Model. Simul., 1 (2003), pp. 349–370, https://doi.org/10.1137/S1540345902416247.

[24] L. 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.

[25] I. W. Selesnick and I. Bayram, *Sparse signal estimation by maximally sparse convex optimization*, IEEE Trans. Signal Process., 62 (2014), pp. 1078–1092.

[26] S. Serra-Capizzano, *A note on antireflective boundary conditions and fast deblurring models*, SIAM J. Sci. Comput., 25 (2004), pp. 1307–1325.

[27] R. Varga, *Matrix Iterative Analysis*, Springer Ser. Comput. Math., Springer, Berlin, 2000.

[28] L. A. Vese and S. Osher, *Modeling textures with total variation minimization and oscillating patterns in image processing*, J. Sci. Comput., 19 (2003), pp. 553–572, https://doi.org/10.1023/A:1025384832106.

[29] X. Wu, J. Zheng, C. Wu, and Y. Cai, *Variational structure-texture image decomposition on manifolds*, Signal Process., 93 (2013), pp. 1773–1784, https://doi.org/10.1016/j.sigpro.2013.01.019.

[30] C. Zhang, *Nearly unbiased variable selection under minimax concave penalty*, Ann. Statist., 38 (2010), pp. 894–942, https://doi.org/10.1214/09-AOS729.