

## **Uncertainty quantification in image segmentation using the Ambrosio–Tortorelli approximation of the Mumford–Shah energy**

Michael Hintermüller<sup>1,2</sup>, Steven-Marian Stengl<sup>1,2</sup>, Thomas M. Surowiec<sup>3</sup>

submitted: March 12, 2020

<sup>1</sup> Weierstrass Institute  
Mohrenstr. 39  
10117 Berlin  
Germany  
E-Mail: michael.hintermueller@wias-berlin.de  
steven-marian.stengl@wias-berlin.de

<sup>2</sup> Humboldt-Universität zu Berlin  
Unter den Linden 6  
10099 Berlin  
Germany  
E-Mail: hint@math.hu-berlin.de  
stengl@math.hu-berlin.de

<sup>3</sup> Department of Mathematics and Computer Science  
Philipps-Universität Marburg  
Hans-Meerwein-Str. 6  
35032 Marburg  
Germany  
E-Mail: surowiec@mathematik.uni-marburg.de

No. 2703  
Berlin 2020



---

2010 *Mathematics Subject Classification.* 62D99, 65N75, 68U10, 65K10.

*Key words and phrases.* Image segmentation, Mumford–Shah, Ambrosio–Tortorelli, measurable selection, Monte-Carlo, sampling.

The first two authors acknowledge financial support by the DFG under the grant HI 1466/10-1 and the third author under SU 963/1-1, both associated to the project “Generalized Nash Equilibrium Problems with Partial Differential Operators: Theory, Algorithms, and Risk Aversion” within the priority programme SPP1962 “Non-smooth and Complementarity-based Distributed Parameter Systems: Simulation and Hierarchical Optimization”. Moreover, the first author received funding by the Einstein Center Mathematics research project (ECMath CH12) as well as by the Berlin Mathematics Research Center MATH+.

Edited by  
Weierstraß-Institut für Angewandte Analysis und Stochastik (WIAS)  
Leibniz-Institut im Forschungsverbund Berlin e. V.  
Mohrenstraße 39  
10117 Berlin  
Germany

Fax: +49 30 20372-303  
E-Mail: [preprint@wias-berlin.de](mailto:preprint@wias-berlin.de)  
World Wide Web: <http://www.wias-berlin.de/>

# Uncertainty quantification in image segmentation using the Ambrosio–Tortorelli approximation of the Mumford–Shah energy

Michael Hintermüller, Steven-Marian Stengl, Thomas M. Surowiec

## Abstract

The quantification of uncertainties in image segmentation based on the Mumford–Shah model is studied. The aim is to address the error propagation of noise and other error types in the original image to the restoration result and especially the reconstructed edges (sharp image contrasts). Analytically, we rely on the Ambrosio–Tortorelli approximation and discuss the existence of measurable selections of its solutions as well as sampling-based methods and the limitations of other popular methods. Numerical examples illustrate the theoretical findings.

## 1 Introduction

In the use of modern imaging techniques in medicine and embedded artificial intelligence systems for an automated investigation of anatomical structure or the classification of objects (human, tree, cloud, etc.), one of the fundamental steps is the division of a given image into distinct areas with characteristic properties, see, e.g., [Par10, Chapter 2] and [Ban08, Part II]. In this context, image segmentation is concerned with the problem of identifying regions with approximately homogeneous features (such as color or gray values, texture, etc.) within given image data. The boundary curve separating such homogeneous features is called the *edge set* or simply the *edge*.

Over the years, different approaches have been proposed in order to accomplish this task by solving, e.g., suitably defined partial differential equations or by minimizing appropriate energies. The latter technique leads to variational models; see, e.g., [KWT87, CKSS97, GG84, MS89, Cha95, Coh97, HR03, HR04, HL09] and the monographs [MS95], [AFK08, Chapter 4] or [CS05, Chapter 7] as well as the references therein for an overview.

In this work, we focus on the Mumford–Shah energy (model)

$$MS^g(u, \Gamma) := \frac{1}{2} \int_{\Omega \setminus \Gamma} |\nabla u|^2 dx + \alpha \mathcal{H}^1(\Gamma) + \frac{\beta}{2} \|u - g\|_{L^2(\Omega)}^2,$$

which was introduced in [MS89]. Here,  $g \in L^2(\Omega)$  is a given grayscale image, possibly degraded by noise, where  $\Omega \subset \mathbb{R}^2$  represents the image domain, and  $\alpha, \beta > 0$  are given parameters. Further,  $\mathcal{H}^1$  denotes the one-dimensional Hausdorff-measure. In order to discuss the structure of  $MS^g$ , we note that the third term is a least-squares data fitting term involving our desired reconstruction  $u$  and the data  $g$ . It is inspired by a maximum likelihood argument regarding the statistical properties of the underlying noise. Since the Mumford–Shah model aims to establish a division in monochromatic regions the (generalized) gradient of the reconstructed image is minimized (penalized when viewed together with the choices of  $\alpha$  and  $\beta$ ) over the complement of the edge set  $\Gamma$  in  $\Omega$ . This motivates the first term in  $MS^g$ . In order to achieve a certain robustness in finding  $\Gamma$  with respect to noise the length of the edge set is penalized in the second term. Moreover, penalizing the length of  $\Gamma$  secures a Caccioppoli property of the edge set.

The aim associated with the above Mumford–Shah model typically is to find a restored image  $u : \Omega \rightarrow \mathbb{R}$  and an edge set  $\Gamma$ , which separates homogeneous regions of interest, by minimizing  $MS^g$ ; see (MS) below. In this context, the Hausdorff-term in  $MS^g$  is one ingredient yielding existence of a minimizing pair  $(u_g, \Gamma_g)$ . We note that such a minimizer is attached to the specific realization of the noise  $\xi$  contained in the data, i.e.,  $g = g_0 + \xi$  with  $g_0$  representing the underlying true image.

Obviously, the location of  $\Gamma_g$  within  $\Omega$  is influenced by the specific realization  $\xi$  of an associated random process and might not properly reflect the location of the true edge  $\Gamma_{g_0}$  pertinent to  $g_0$ . Indeed, rather than computing  $\Gamma_g$  one is merely interested in estimating the uncertainty in the location of the reconstructed edge set, given statistical properties of the underlying random process. This brings us into the realm of *uncertainty quantification* of geometric objects.

In general, one of the fundamental components of predictive estimation in uncertainty quantification is to analyze how uncertainty propagates through a given mathematical model; also known as model prediction [Smi13, Chap. 9]. We observe that the vast majority of the existing literature on the propagation of uncertainty is concerned with well-defined systems whose (unique) solutions are given by real numbers, vectors, matrices, or distributed parameters; see e.g., [Smi13, Chap. 9, Chap. 10] or for an introduction with emphasis on methods also [Sul15].

However, within the context of image processing (segmentation, in particular), variational models of crack propagation, or free discontinuity problems in general, the output of the model is a *geometric* quantity, which itself does not even live in a linear space; see [AFP00] and the references therein. Due to the associated complexity, we emphasize the role played by the geometric variable (the edge set). In fact, of the existing methods for the propagation of uncertainty: direct evaluation of the mean and variance of a quantity of interest, sampling methods, perturbation methods, and spectral representations, only sampling methods appear to be possible.

In contrast to inverse problems, where Bayesian techniques admitting probability densities to characterize the true value are utilized, we are here confronted rather with a kind of forward problem. It is also worth mentioning that the model considered in this work yields solution pairs consisting of the restored image and associated edge set, which is not guaranteed to be unique due to a lack of strict convexity of the objective in the Mumford–Shah problem.

Our intention here is to therefore draw attention to an important but somewhat neglected class of problems for uncertainty quantification. As a means of investigating these models, in particular the edge set, we later propose the notion of generalized cumulative distribution function and pointwise quantiles. Of course, pointwise averages, variances, etc. are also conceivable. However, we believe that an estimation of the probability of an edge appearing within a given image patch (or a patch of pixels in the discrete setting) is the more valuable information.

Let us now return to the Mumford–Shah model and introduce the associated variational framework. Given sufficient Sobolev regularity of  $u$  on the set  $\Omega \setminus \Gamma$ , this leads us to the following minimization problem:

$$\begin{aligned} & \text{minimize } MS^g(u, \Gamma) \text{ over } (u, \Gamma) \in \mathcal{A}, \\ & \text{with } \mathcal{A} := \{(u, \Gamma) : \Gamma \subset \Omega \text{ closed and } u \in H^1(\Omega \setminus \Gamma)\}. \end{aligned} \tag{MS}$$

Note that (MS) constitutes the renown Mumford–Shah problem. The interested reader is referred to the original paper [MS89] for more details, and for a condensed overview including a discussion of the existence of solutions we point to [AFP00] and the references therein. Uniqueness of a solution cannot be expected in general as it can be seen in a one-dimensional counter example in [CS05, Section 7.4.5]. The analytic difficulties arise due to the set-variable  $\Gamma$  as well as  $u$  being an element of a space depending on this set. Moreover, we have the Hausdorff measure directly appearing in

the functional. These aspects not only significantly complicate the analysis of (MS), but they also challenge the numerical treatment.

An elegant way to address these issues is due to Ambrosio and Tortorelli (cf. [AT92]). Driven by the analytical notion of  $\Gamma$ -convergence, the main idea is to treat the edge set along with its Hausdorff measure by a *phase field* approach. In fact, the so-called *Ambrosio–Tortorelli functional*  $AT_\varepsilon^g : H^1(\Omega) \times K \rightarrow \mathbb{R}$  is defined by

$$AT_\varepsilon^g(u, z) := \frac{1}{2} \int_\Omega z^2 |\nabla u|^2 dx + \frac{\alpha}{2} \left( \varepsilon \int_\Omega |\nabla z|^2 dx + \frac{1}{4\varepsilon} \int_\Omega (z - 1)^2 dx \right) + \frac{\beta}{2} \int_\Omega |u - g|^2 dx + \frac{\eta}{2} \int_\Omega |\nabla u|^2 dx,$$

together with the constraint set

$$K := \{y \in H^1(\Omega) : 0 \leq y \leq 1 \text{ almost everywhere (a.e.) on } \Omega\}.$$

As before, the quantity  $u \in H^1(\Omega)$  denotes the restored image. The geometric quantity  $\Gamma$ , on the other hand, is approximated by the function  $z \in H^1(\Omega)$ , which can be interpreted as an edge indicator. In this context, the zero-level set of  $z$  is associated with an edge and the level set associated with  $z = 1$  relates to the absence of such an edge. From this we see that the first term in  $AT_\varepsilon^g$  refers to the first term of (MS). The second and third terms approximate the Hausdorff measure. The remaining terms are the data fitting term as well as an additional regularization for the phase field variable  $z$  with weighting factor  $\eta > 0$ . This term ensures coercivity. In [AT92], it is shown that the sequence of minimization problems

$$\text{minimize } AT_\varepsilon^g(u, z) \text{ over } (u, z) \in H^1(\Omega) \times K, \quad (\text{AT})$$

$\Gamma$ -converges in the strong  $L^1(\Omega; \mathbb{R}^2)$ -topology to the original Mumford–Shah problem (MS) as  $\varepsilon \rightarrow 0$  and  $\eta = o(\varepsilon)$ , with  $\frac{o(t)}{t} \rightarrow 0$  as  $t \rightarrow 0$ .

As motivated above, given noisy observations  $g$ , we are interested in studying the influence of random errors and other data transformation effects in the image data on the reconstruction of edges in the image. The degradation model underlying our study is given by

$$g = Lg_0 + \xi,$$

where  $L$  is a linear continuous operator from  $L^2(\Omega)$  to  $L^2(\Omega)$ , i.e.,  $L \in \mathcal{L}(L^2(\Omega))$ . Note that  $L$  may model image blur, subsampling or Fourier and wavelet transforms, respectively, to mention only a few. Moreover,  $\xi \in L^2(\Omega)$  is an oscillating map with zero mean. From a stochastic point of view  $L$  and  $\xi$  constitute realizations of stochastic processes. We also assume that  $L$  does not annihilate constant functions, i.e.  $L1 \neq 0$ .

Since  $\Gamma$  in (MS) is a set, a mathematically meaningful notion of uncertainty propagation of this object is not immediately clear. In view of this, the Ambrosio–Tortorelli formulation offers a transformation of this geometric feature to a functional variable which is more amenable to the study of the propagation of uncertainty. Taking account of the additional data transformation operator  $L$  in our model  $AT_\varepsilon^g$  becomes

$$AT_\varepsilon^g(u, z) := \frac{1}{2} \int_\Omega z^2 |\nabla u|^2 dx + \frac{\alpha}{2} \left( \varepsilon \int_\Omega |\nabla z|^2 dx + \frac{1}{4\varepsilon} \int_\Omega (z - 1)^2 dx \right) + \frac{\beta}{2} \int_\Omega |Lu - g|^2 dx + \frac{\eta}{2} \int_\Omega |\nabla u|^2 dx,$$

which constitutes the segmentation model under investigation in this paper.

The rest of the paper is organized as follows. In Section 2 we provide a collection of important results on the Ambrosio–Tortorelli functional concerning properties of the functional as well as first-order conditions for its minimizers. These will be used in Section 3 to give a mathematical justification of *random* processed image and edge indicators in the sense of measurable selections based on the theorem of Kuratowski–Ryll–Nardzewski. In Section 4 we briefly discuss frontiers of popular techniques in uncertainty quantification before we turn our attention to sampling based methods. We conclude our work in Section 5 with numerical examples.

## 2 Properties of the deterministic problem

Now we collect several properties of the Ambrosio–Tortorelli functional and its minimizers. In order to reduce technicalities in the subsequent proofs we establish the following norm equivalence. Below,  $\|\cdot\|_{L^2(\Omega)}$  denotes the usual  $L^2(\Omega)$ -norm on a bounded domain  $\Omega \subset \mathbb{R}^d$  and  $H^1(\Omega)$  is the Sobolev space of functions in  $L^2(\Omega)$  with generalized gradient in  $L^2(\Omega)^d$ , see e.g. [AF03].

**Lemma 1.** *Let  $L \in \mathcal{L}(L^2(\Omega))$  with  $L1 \neq 0$ . Then the norm*

$$\|u\|^2 := \|\nabla u\|_{L^2(\Omega)}^2 + \|Lu\|_{L^2(\Omega)}^2$$

*and the usual  $H^1$ -norm, i.e.,  $\|u\|_{H^1(\Omega)}^2 = \|\nabla u\|_{L^2(\Omega)}^2 + \|u\|_{L^2(\Omega)}^2$ , are equivalent on  $H^1(\Omega)$ .*

*Proof.* Let  $u \in H^1(\Omega)$  be chosen arbitrarily. At first we see

$$\|u\|^2 = \|\nabla u\|_{L^2(\Omega)}^2 + \|Lu\|_{L^2(\Omega)}^2 \leq \|\nabla u\|_{L^2(\Omega)}^2 + \|L\|_{\mathcal{L}(L^2(\Omega))}^2 \|u\|_{L^2(\Omega)}^2$$

and hence  $\|u\| \leq \max(1, \|L\|_{\mathcal{L}(L^2)}) \|u\|_{H^1(\Omega)}$ , where  $\|\cdot\|_{\mathcal{L}(L^2)}$  denotes the operator norm for linear bounded operators from  $L^2(\Omega)$  to  $L^2(\Omega)$ .

For proving the inverse inequality we define the quantity  $m := \frac{1}{\lambda^d(\Omega)} \int_{\Omega} u dx \in \mathbb{R}$ , where  $\lambda^d$  denotes the  $d$ -dimensional Lebesgue measure, and let  $C_P > 0$  be the Poincaré constant, which is the (smallest) real number fulfilling the Poincaré inequality

$$\|u - m\|_{L^2(\Omega)} \leq C_P \|\nabla u\|_{L^2(\Omega)} \text{ for all } u \in H^1(\Omega),$$

see e.g. [ABM14, Corollary 5.4.1]. Using the above we find

$$\begin{aligned} \|u\|_{H^1(\Omega)}^2 &= \|\nabla u\|_{L^2(\Omega)}^2 + \|u\|_{L^2(\Omega)}^2 = \|\nabla u\|_{L^2(\Omega)}^2 + \|u - m\|_{L^2(\Omega)}^2 + \lambda^d(\Omega)m^2 \\ &\leq (1 + C_P^2) \|\nabla u\|_{L^2(\Omega)}^2 + \lambda^d(\Omega) \frac{1}{\|L1\|_{L^2(\Omega)}^2} \|mL1\|_{L^2(\Omega)}^2 \\ &\leq (1 + C_P^2) \|\nabla u\|_{L^2(\Omega)}^2 + 2\lambda^d(\Omega) \frac{1}{\|L1\|_{L^2(\Omega)}^2} \left( \|L(u - m)\|_{L^2(\Omega)}^2 + \|Lu\|_{L^2(\Omega)}^2 \right) \\ &\leq \left( 1 + C_P^2 + 2\lambda^d(\Omega)C_P^2 \frac{\|L\|_{\mathcal{L}(L^2(\Omega))}^2}{\|L1\|_{L^2(\Omega)}^2} \right) \|\nabla u\|_{L^2(\Omega)}^2 + 2\lambda^d(\Omega) \frac{1}{\|L1\|_{L^2(\Omega)}^2} \|Lu\|_{L^2(\Omega)}^2 \\ &\leq \max \left( 1 + C_P^2 + 2\lambda^d(\Omega)C_P^2 \frac{\|L\|_{\mathcal{L}(L^2(\Omega))}^2}{\|L1\|_{L^2(\Omega)}^2}, 2\lambda^d(\Omega) \frac{1}{\|L1\|_{L^2(\Omega)}^2} \right) \|u\|^2, \end{aligned}$$

which completes the proof. □

From now on we equip the Sobolev space  $H^1(\Omega)$  for the processed image  $u$  with the new norm  $\|\cdot\|$ , whereas the norm of the  $z$ -component remains unchanged. Next we consider continuity properties of the Ambrosio–Tortorelli functional.

**Theorem 2.** *The mapping  $AT_\varepsilon^g : H^1(\Omega) \times K \rightarrow \mathbb{R}$  is weakly lower semi-continuous.*

*Proof.* Let  $(u_n, z_n) \rightharpoonup (u, z)$  be a weakly convergent sequence in  $H^1(\Omega) \times K$ . Then there exists a subsequence (not relabeled) such that the limes inferior is attained and  $z_n \rightarrow z$  converges additionally pointwise almost everywhere (Fischer-Riesz). Using Lebesgue's dominated convergence theorem we obtain  $z_n \nabla u_n \rightharpoonup z \nabla u$  in  $L^2(\Omega; \mathbb{R}^d)$ . Eventually we obtain by the weak lower semi-continuity of the  $L^2$ -norm that

$$\begin{aligned} AT_\varepsilon^g(u, z) &= \frac{1}{2} \|z \nabla u\|_{L^2(\Omega; \mathbb{R}^d)}^2 + \frac{\eta}{2} \|\nabla u\|_{L^2(\Omega; \mathbb{R}^d)}^2 + \frac{\beta}{2} \|Lu - g\|_{L^2(\Omega)}^2 \\ &\quad + \frac{\alpha}{8\varepsilon} \|z - 1\|_{L^2(\Omega)}^2 + \frac{\alpha\varepsilon}{2} \|\nabla z\|_{L^2(\Omega)}^2 \\ &\leq \liminf_{n \rightarrow \infty} \frac{1}{2} \|z_n \nabla u_n\|_{L^2(\Omega; \mathbb{R}^d)}^2 + \liminf_{n \rightarrow \infty} \frac{\eta}{2} \|\nabla u_n\|_{L^2(\Omega; \mathbb{R}^d)}^2 + \liminf_{n \rightarrow \infty} \frac{\beta}{2} \|Lu_n - g\|_{L^2(\Omega)}^2 \\ &\quad + \liminf_{n \rightarrow \infty} \frac{\alpha}{8\varepsilon} \|z_n - 1\|_{L^2(\Omega)}^2 + \liminf_{n \rightarrow \infty} \frac{\alpha\varepsilon}{2} \|\nabla z_n\|_{L^2(\Omega)}^2 \\ &\leq \liminf_{n \rightarrow \infty} AT_\varepsilon^g(u_n, z_n), \end{aligned}$$

which proves the lower-semicontinuity of  $AT_\varepsilon^g$ . □

Using the direct method of calculus of variations in combination with the two previous lemmas, it is not difficult to derive the existence of a solution of (AT). Due to the nonconvexity of  $AT_\varepsilon^g$  uniqueness cannot be expected in general.

For the sake of characterizing the set of minimizers we turn our attention to first-order optimality conditions. The derivation of a corresponding system is challenged by the nonsmoothness of  $AT_\varepsilon^g$  with respect to the  $H^1(\Omega) \times H^1(\Omega)$ -topology together with the presence of the constraint set  $K$ . We address these issues by using a truncation argument. In fact, for all  $u, z \in H^1(\Omega)$  with  $z \in L^\infty(\Omega)$  it holds that  $AT_\varepsilon^g(u, z) < \infty$  as well as

$$AT_\varepsilon^g(u, \max(\min(z, 1), 0)) \leq AT_\varepsilon^g(u, z).$$

The constraint set is then substituted by  $z \in X$  with  $X := H^1(\Omega) \cap L^\infty(\Omega)$  together with the norm  $\|\cdot\|_X^2 := \|\cdot\|_{H^1(\Omega)}^2 + \|\cdot\|_{L^\infty(\Omega)}^2$ . Consequently, we obtain

$$\min_{(u,z) \in H^1(\Omega) \times K} AT_\varepsilon^g(u, z) = \min_{(u,z) \in H^1(\Omega) \times X} AT_\varepsilon^g(u, z), \quad (1)$$

and the respective sets of minimizers coincide. The advantage of the righthand side formulation in (1) lies in the fact that the Ambrosio–Tortorelli functional is continuously Fréchet-differentiable on  $H^1(\Omega) \times X$  with derivative

$$\begin{aligned} \langle DAT_\varepsilon^g(u, z), (\varphi, \psi) \rangle_{H^1 \times X} &= \int_{\Omega} ((z^2 + \eta) \nabla u \nabla \varphi + \beta(Lu - g)L\varphi) \, dx \\ &\quad + \int_{\Omega} \left( z\psi |\nabla u|^2 + \alpha\varepsilon \nabla z \nabla \psi + \frac{\alpha}{4\varepsilon} (z - 1)\psi \right) \, dx \\ &=: a(u, z; \varphi) + b(u, z; \psi). \end{aligned}$$

Here,  $\langle \cdot, \cdot \rangle$  denotes the dual pairing between  $H^1(\Omega) \times X$  and its dual. Then the stationarity system associated with the problem on  $H^1(\Omega) \times X$  in (1) in its weak form reads

$$a^g(u, z; \varphi) = \int_{\Omega} (z^2 + \eta) \nabla u \nabla \varphi dx + \beta \int_{\Omega} (Lu - g) L \varphi dx = 0 \text{ for all } \varphi \in H^1(\Omega), \quad (2a)$$

$$b(u, z; \psi) = \int_{\Omega} |\nabla u|^2 z \psi dx + \alpha \varepsilon \int_{\Omega} \nabla z \nabla \psi dx + \frac{\alpha}{4\varepsilon} \int_{\Omega} (z - 1) \psi dx = 0 \text{ for all } \psi \in X. \quad (2b)$$

Utilizing coercivity properties of the involved inner products, one readily obtains the a priori bounds

$$\min(\eta, \beta) \|u\| \leq \beta \|g\|_{L^2(\Omega)}, \quad \text{and} \quad (3a)$$

$$\min\left(\alpha \varepsilon, \frac{\alpha}{4\varepsilon}\right) \|z\|_{H^1(\Omega)} \leq \frac{\alpha}{4\varepsilon} \lambda^d(\Omega)^{\frac{1}{2}}. \quad (3b)$$

Moreover, we deduce  $z \in K$ , which results from (2b) by testing with  $\psi = -(z - 1)^+ + (-z)^+$  with  $(\cdot)^+ = \max\{0, \cdot\}$  in a pointwise sense; see also [BOS10, Prop. 1.3].

After clarifying existence of solutions and first-order necessary conditions we next draw our attention to the following convergence result.

**Theorem 3.** *Suppose there exists a bounded sequence  $((u_j, z_j))_{j \in \mathbb{N}} \subset H^1(\Omega) \times X$  such that*

$$|a^g(u_j, z_j; \varphi)| \leq \mu_j \|\varphi\| \quad \forall \varphi \in H^1(\Omega), \quad (4a)$$

$$|b(u_j, z_j; \psi)| \leq \nu_j \|\psi\|_{H^1(\Omega)} \quad \forall \psi \in X. \quad (4b)$$

*Further, let  $(\mu_j, \nu_j) \rightarrow 0$  and  $z_j \in K$  for all  $j \in \mathbb{N}$ . Then there exist a point  $(u^*, z^*)$  and a subsequence such that*

$$(u_{j_k}, z_{j_k}) \rightarrow (u^*, z^*) \text{ in } H^1(\Omega; \mathbb{R}^2), \quad \text{and } (u^*, z^*) \text{ solves (2).}$$

*Proof.* The result follows from a straight forward adaptation of the proof in [BOS10, Theorem 4.1].  $\square$

The above result is useful when studying properties of solution sets of (2) and for the convergence analysis of a splitting-type algorithm in function space for iteratively solving (2). We continue here by studying such a splitting method and postpone the former aspect to Section 3.

---

#### Algorithm 1: Splitting Method

---

**Data:** parameters  $\alpha, \beta, \varepsilon, \eta$ , image data  $g$ , starting values  $u_0 \in H^1(\Omega), z_0 \in K$

**Result:**  $u \in H^1(\Omega), z \in K$  fulfilling (2)

```

1 for  $j = 0, 1, 2, \dots$  do
2    $u_{j+1} = \operatorname{argmin}_{u \in H^1(\Omega)} AT_{\varepsilon}^g(u, z_j);$ 
3    $z_{j+1} = \operatorname{argmin}_{z \in K} AT_{\varepsilon}^g(u_{j+1}, z);$ 
4 end
```

---

We note that the algorithm splits the overall nonconvex minimization problem into the iterative solution of two subsequent problems. Indeed within one iteration, in each step a convex quadratic minimization problem needs to be solved, which is equivalent to solving a linear partial differential equation (PDE), respectively. The convergence of this iteration scheme can then be obtained by applying a result due to [Hae10] concerning the decay of the distance of two consecutive image iterates.

**Corollary 4.** *Let  $(u_j, z_j)_{j \in \mathbb{N}}$  be the sequence generated by the splitting algorithm. Then there exists a stationary point  $(u^*, z^*) \in H^1(\Omega) \times X$  and a subsequence such that*

$$(u_{j_k}, z_{j_k}) \rightarrow (u^*, z^*) \text{ (strongly) in } H^1(\Omega; \mathbb{R}^2).$$

*Proof.* Adapting [Hae10, Lemma 2.3.9 and Theorem 2.3.10], we obtain

$$\sum_{j=1}^{\infty} \|u_{j+1} - u_j\|^2 < \infty, \text{ which implies } \lim_{j \rightarrow \infty} \|u_{j+1} - u_j\| = 0.$$

Since  $u_{j+1} = \operatorname{argmin}_{u \in H^1(\Omega)} AT_\varepsilon^g(u, z_j)$  it holds that

$$a^g(u_{j+1}, z_j; \varphi) = 0 \text{ for all } \varphi \in H^1(\Omega).$$

Hence we obtain for every  $\varphi \in H^1(\Omega)$  that

$$\begin{aligned} |a^g(u_j, z_j; \varphi)| &= |a^g(u_{j+1}, z_j; \varphi) - a^g(u_j, z_j; \varphi)| \\ &= \left| \int_{\Omega} (z_j^2 + \eta)(\nabla u_{j+1} - \nabla u_j) \nabla \varphi dx + \beta \int_{\Omega} (Lu_{j+1} - Lu_j) L\varphi dx \right| \\ &\leq (1 + \eta) \|\nabla u_{j+1} - \nabla u_j\|_{L^2(\Omega)} \|\nabla \varphi\|_{L^2(\Omega)} + \beta \|Lu_{j+1} - Lu_j\|_{L^2(\Omega)} \|L\varphi\|_{L^2(\Omega)} \\ &\leq \max(1 + \eta, \beta) \|u_{j+1} - u_j\| \cdot \|\varphi\|. \end{aligned}$$

Since  $b(u_j, z_j; \psi) = 0$  for all  $\psi \in X$ , we see that the conditions of Theorem 3 are fulfilled with  $\mu_j = \max(1 + \eta, \beta) \|u_{j+1} - u_j\|$  and  $\nu_j = 0$ . Hence, the assertion follows.  $\square$

We point out that the last estimate can also be used as a numerically useful stopping criterion for the splitting algorithm.

With these results at hand we are now able to derive finer properties of the dependence of the solutions of the minimization problem and the first-order system on the given image data, respectively.

### 3 Existence of measurable selections

Next we introduce the concept of *uncertain edges*. As motivated earlier, we assume that the real image is inaccessible due to corruption by (random) noise and a deterministic degradation operator. The resulting data are therefore modeled by a random variable  $g$ . Rather than addressing the issue of finding a reconstruction of the image along with an edge set, we seek to study the propagation of noise into the solution. For the reasons mentioned above, we focus on the Ambrosio–Tortorelli approximation (AT) of the Mumford–Shah energy. The main issue now is to establish existence of random variables of solutions as well as their characterization. For this purpose we define the following set-valued solution operators  $\mathcal{S}_{\min}, \mathcal{S}_{\text{stat}} : L^2(\Omega) \rightrightarrows H^1(\Omega; \mathbb{R}^2)$  defined by

$$\begin{aligned} \mathcal{S}_{\min}(g) &:= \operatorname{argmin}\{AT_\varepsilon^g(u, z) : (u, z) \in H^1(\Omega) \times K\}, \\ \mathcal{S}_{\text{stat}}(g) &:= \{(u, z) \in H^1(\Omega) \times K : (u, z) \text{ solves (2) for } g\}. \end{aligned}$$

We start by providing structural properties of these operators.

**Theorem 5.** *Let  $\mathcal{S} \in \{\mathcal{S}_{\min}, \mathcal{S}_{\text{stat}}\}$ . Then the following properties hold true:*

(i)  $\mathcal{S}$  has nonempty and compact values.

(ii) For all closed sets  $C \subset H^1(\Omega) \times K$  it holds that  $\mathcal{S}^{-1}(C) \subset L^2(\Omega)$  is closed,

where we use  $\mathcal{S}^{-1}(C) := \{g \in L^2(\Omega) : \mathcal{S}(g) \cap C \neq \emptyset\}$ .

*Proof.* We split the proof into several steps. Below  $C \subset H^1(\Omega) \times K$  always denotes a closed set.

*Step 1 (Nonemptiness).* Since we know that the Ambrosio–Tortorelli problem has a minimizer for every  $g \in L^2(\Omega)$ , it holds that  $\mathcal{S}_{\min}(g) \neq \emptyset$ . Since every minimizer fulfils the stationarity system,  $\mathcal{S}_{\text{stat}}(g) \neq \emptyset$  follows, as well.

*Step 2 ( $\mathcal{S}_{\min}(g)$  is closed).* Let  $m := \min_{(u,z) \in H^1(\Omega) \times K} AT_\varepsilon^g(u, z)$ . Then we rewrite

$$\begin{aligned} \mathcal{S}_{\min}(g) &= \operatorname{argmin}\{AT_\varepsilon^g(u, z) : (u, z) \in H^1(\Omega) \times K\} \\ &= \{(u, z) \in H^1(\Omega) \times K : AT_\varepsilon^g(u, z) \leq m\} =: \operatorname{lev}_m AT_\varepsilon^g. \end{aligned}$$

Since according to Theorem 2 the Ambrosio–Tortorelli functional is (weakly) l.s.c. on  $H^1(\Omega) \times K$  with respect to the  $H^1(\Omega; \mathbb{R}^2)$ -topology, we find that  $\mathcal{S}_{\min}(g)$  is closed.

*Step 3 ( $\mathcal{S}_{\text{stat}}(g)$  and  $\mathcal{S}_{\min}(g)$  are compact).* Let  $(u_n, z_n)_{n \in \mathbb{N}} \subset \mathcal{S}_{\text{stat}}(g)$  be a sequence of stationary points. Then this sequence is bounded in  $H^1(\Omega) \times X$  by the a priori estimate (3), and  $z_n \in K$  holds true, as well. We apply Theorem 3 with  $\mu_j = \nu_j = 0$  and get the existence of a stationary point together with a strongly  $H^1(\Omega; \mathbb{R}^2)$ -convergent subsequence. Thus, the compactness of  $\mathcal{S}_{\text{stat}}$  as well as the compactness of  $\mathcal{S}_{\min}(g)$  (using Step 2) are established.

*Step 4 (Proof of (ii) for  $\mathcal{S}_{\text{stat}}$ ).* Let  $(g_n)_{n \in \mathbb{N}} \subset \mathcal{S}_{\text{stat}}^{-1}(C)$  be a (strongly)  $L^2(\Omega)$ -convergent sequence with limit  $g \in L^2(\Omega)$ . We show  $g \in \mathcal{S}_{\text{stat}}^{-1}(C)$ . For this purpose, let  $(u_n, z_n)_{n \in \mathbb{N}} \subset C$  be a sequence of stationary points for  $g_n$ ,  $n \in \mathbb{N}$ . Then,  $0 \leq z_n \leq 1$  holds almost everywhere and

$$\begin{aligned} a^{g_n}(u_n, z_n; \varphi) &= 0 \quad \forall \varphi \in H^1(\Omega), \\ b(u_n, z_n; \psi) &= 0 \quad \forall \psi \in X, \end{aligned}$$

for every  $n \in \mathbb{N}$ . Again, we get the boundedness of  $((u_n, z_n))_{n \in \mathbb{N}}$  in  $H^1(\Omega) \times X$  by the uniform boundedness of  $\|g_n\|_{L^2(\Omega)}$ ,  $n \in \mathbb{N}$ , combined with the a priori bound (3). Moreover, we find

$$\begin{aligned} |a^g(u_n, z_n; \varphi)| &= |a^g(u_n, z_n; \varphi) - a^{g_n}(u_n, z_n; \varphi)| = \left| \beta \int_{\Omega} (g_n - g) L \varphi dx \right| \\ &\leq \beta \|g_n - g\|_{L^2(\Omega)} \|L \varphi\|_{L^2(\Omega)}. \end{aligned}$$

Since  $g_n \rightarrow g$  in  $L^2(\Omega)$ , the conditions of Theorem 3 are fulfilled for  $g$  with  $\mu_n = \beta \|g_n - g\|_{L^2(\Omega)}$  and  $\nu_n = 0$ . Hence, there exists a stationary point  $(u^*, z^*)$  together with a strongly  $H^1(\Omega; \mathbb{R}^2)$ -convergent subsequence of  $(u_n, z_n)_{n \in \mathbb{N}}$ . Since  $(u_n, z_n) \subset C$  for all  $n \in \mathbb{N}$  and  $C$  is closed, we obtain  $(u^*, z^*) \in C$ . Therefore,  $\mathcal{S}_{\text{stat}}(g) \cap C \neq \emptyset$  and, hence,  $g \in \mathcal{S}_{\text{stat}}^{-1}(C)$ .

*Step 5 (Proof of (ii) for  $\mathcal{S}_{\min}$ ).* Let  $(g_n)_{n \in \mathbb{N}} \subset \mathcal{S}_{\min}^{-1}(C)$  be a strongly  $L^2(\Omega)$ -convergent sequence with limit  $g \in L^2(\Omega)$ . We show  $g \in \mathcal{S}_{\min}^{-1}(C)$ . Take  $(u_n, z_n)_{n \in \mathbb{N}} \subset C$  as a corresponding sequence of minimizers for  $g_n$ ,  $n \in \mathbb{N}$ . For all  $n \in \mathbb{N}$  and all  $(u, z) \in H^1(\Omega) \times K$  it thus holds that

$$AT_\varepsilon^{g_n}(u_n, z_n) \leq AT_\varepsilon^{g_n}(u, z).$$

Since every minimizer is a stationary point, we obtain from the arguments of step 4 a stationary point  $(u^*, z^*) \in C$  and a strongly  $H^1(\Omega; \mathbb{R}^2)$ -convergent subsequence  $(u_{n_k}, z_{n_k})_{k \in \mathbb{N}}$ . Then  $(u^*, z^*)$  is also a minimizer. In fact, we have

$$\begin{aligned} AT_\varepsilon^g(u^*, z^*) &\leq \liminf_{k \rightarrow \infty} AT_\varepsilon^g(u_{n_k}, z_{n_k}) \\ &= \liminf_{k \rightarrow \infty} \left( AT_\varepsilon^{g_{n_k}}(u_{n_k}, z_{n_k}) + \beta(Lu_{n_k} - g_{n_k}, g_{n_k} - g)_{L^2(\Omega)} + \frac{\beta}{2} \|g_{n_k} - g\|_{L^2(\Omega)}^2 \right) \\ &= \liminf_{k \rightarrow \infty} AT_\varepsilon^{g_{n_k}}(u_{n_k}, z_{n_k}) \leq \lim_{k \rightarrow \infty} AT_\varepsilon^{g_{n_k}}(u, z) = AT_\varepsilon^g(u, z). \end{aligned}$$

Since  $(u, z)$  was arbitrary, we conclude that  $(u^*, z^*)$  is also a minimizer for  $g$ . Hence  $\mathcal{S}_{\min}^{-1}(C)$  is closed. This completes the proof.  $\square$

The aim is to construct a random variable, whose values are the solutions of (AT) for the given image and for almost every realization of noise. This quantity then defines an associated reconstructed image and edge indicator, respectively. Therefore we aim at using the *Kuratowski–Ryll–Nardzewski measurable selection theorem*, which we state here for ease of reference and refer the interested reader to [Bee93, Theorem 6.6.7].

**Theorem 6** (Kuratowski and Ryll–Nardzewski). *Let  $Y$  be a Polish space,  $\mathcal{B}(Y)$  the Borel algebra on  $Y$ ,  $(\Xi, \mathcal{F})$  a measurable space and  $\Psi : \Xi \rightrightarrows Y$  a set-valued operator such that*

- (i)  $\Psi(\xi)$  is nonempty and closed for all  $\xi \in \Xi$ , and
- (ii) for all  $U \subset Y$  open, it holds that  $\Psi^{-1}(U) = \{\xi \in \Xi : \Psi(\xi) \cap U \neq \emptyset\} \in \mathcal{F}$ .

Then there exists a measurable selection, i.e. a mapping  $\Psi_{\text{sel}} : \Xi \rightarrow Y$  such that  $\Psi_{\text{sel}}(\xi) \in \Psi(\xi)$  for all  $\xi \in \Xi$ .

Given the notation of Theorem 6 we note that the fulfilment of the condition  $\Psi^{-1}(C) \in \mathcal{F}$  for all closed subsets  $C \subset Y$  implies assumption (ii). Indeed, let  $U \subset Y$  be an arbitrary open set  $U \subset Y$ . By  $U^c$  we denote the complement of  $U$  in  $Y$ . Rewriting  $U$  gives

$$\begin{aligned} U &= \{y \in Y : \text{dist}(y, U^c) > 0\} = \{y \in Y : \exists n \in \mathbb{N} : \text{dist}(y, U^c) \geq \frac{1}{n}\} \\ &= \bigcup_{n=1}^{\infty} \{y \in Y : \text{dist}(y, U^c) \geq \frac{1}{n}\} = \bigcup_{n=1}^{\infty} C_n, \end{aligned}$$

with  $C_n := \{y \in Y : \text{dist}(y, U^c) \geq \frac{1}{n}\} = (\text{dist}(\cdot, U^c))^{-1}([\frac{1}{n}, +\infty))$  as a closed set for every  $n \in \mathbb{N}$ . Then we obtain

$$\begin{aligned} \Psi^{-1}(U) &= \{\xi \in \Xi : \Psi(\xi) \cap U \neq \emptyset\} = \{\xi \in \Xi : \exists n \in \mathbb{N} : \Psi(\xi) \cap C_n \neq \emptyset\} \\ &= \bigcup_{n=1}^{\infty} \Psi^{-1}(C_n). \end{aligned}$$

Since  $\Psi^{-1}(C_n) \in \mathcal{F}$  by assumption, we have  $\Psi^{-1}(U) \in \mathcal{F}$ .

Based on this observation, we are now able to prove the existence of measurable selections for the operators  $\mathcal{S}_{\min}$  and  $\mathcal{S}_{\text{stat}}$ .

**Corollary 7.** *Let  $\mathcal{S} \in \{\mathcal{S}_{\min}, \mathcal{S}_{\text{stat}}\}$ . Then there exist Borel functions  $u_{\text{sel}} : L^2(\Omega) \rightarrow H^1(\Omega)$  and  $z_{\text{sel}} : L^2(\Omega) \rightarrow K$  such that  $(u_{\text{sel}}(g), z_{\text{sel}}(g)) \in \mathcal{S}(g)$  for all  $g \in L^2(\Omega)$ .*

*Proof.* Set  $\Psi \in \{\mathcal{S}_{\min}, \mathcal{S}_{\text{stat}}\}$ ,  $\Xi = L^2(\Omega)$ ,  $\mathcal{F} = \mathcal{B}(L^2(\Omega))$  be the Borel algebra and  $Y = H^1(\Omega) \times K$ . Since  $H^1(\Omega; \mathbb{R}^2)$  is a separable Hilbert space it is a Polish space. As closed subsets of Polish spaces are Polish as well (see [TSE16, Proposition A.1] or use [AB06, Corollary 3.5]) also  $Y \subset H^1(\Omega)^2$  is a Polish space and hence the assumptions of Theorem 6 are fulfilled.  $\square$

For the moment we focus on the denoising case, i.e.,  $L = \text{id}_{L^2(\Omega)}$ . Due to the analytical difficulties concerning the edge set as a geometric quantity, we have drawn our attention to the Ambrosio–Tortorelli model. In the following we temporarily return to the original Mumford–Shah model. In fact, it is possible to derive a result similar to the one above for the processed image only. For this sake we first discuss a reformulation of the Mumford–Shah problem (MS) based on *special functions of bounded variations*.

For this purpose we recall that the space of functions of bounded variation  $BV(\Omega)$  is defined as the set of  $L^1$ -functions with bounded total variation. The latter quantity is given by

$$|Du|(\Omega) := \sup \left\{ \int_{\Omega} u \operatorname{div} p \, dx : p \in C_0^1(\Omega; \mathbb{R}^d), |p(x)|_{\infty} \leq 1 \text{ for all } x \in \Omega \right\}.$$

Accordingly we have  $BV(\Omega) = \{u \in L^1(\Omega) : |Du|(\Omega) < +\infty\}$ . Equipped with the norm  $\|u\|_{BV} = \|u\|_{L^1(\Omega)} + |Du|(\Omega)$  the space  $BV(\Omega)$  is a Banach space. The distributional derivative  $Du$  of  $u \in BV(\Omega)$  is a Radon measure which can be decomposed according to

$$Du = \nabla u \cdot \lambda^d + (u^+ - u^-) \nu_{S_u} \mathcal{H}^{d-1} \llcorner S_u + \mathcal{C}_u.$$

Here, the first term is the part of  $Du$  that is absolutely continuous with respect to the Lebesgue measure with  $\nabla u$  denoting its density. The terms  $u^+$ ,  $u^-$  are the so-called upper respectively lower limits of  $u$  and  $S_u := \{x \in \Omega : u^+(x) > u^-(x)\}$  represents the *jump set*. The second term denotes the part of the measure that is absolutely continuous with respect to the Hausdorff measure restricted to  $S_u$ , which we denote here as  $\mathcal{H}^{d-1} \llcorner S_u$ . The remaining part is singular with respect to both the Lebesgue and Hausdorff measure and is referred to as the *Cantor part* of  $Du$ . For the definition of approximate limits as well as further details the reader is referred to [ABM14, Section 10],[EG91, Section 5] as well as to the monograph [AFP00].

The space of *special functions of bounded variation* is then defined as

$$SBV(\Omega) := \{u \in BV(\Omega) : \mathcal{C}_u = 0\},$$

which is the subspace of  $BV$ -functions with vanishing Cantor part.

Using the above notation it is possible to provide the following equivalent reformulation of the Mumford–Shah problem, originally proposed by [DGCL89], as a problem with the processed image contained in the  $SBV$ -space, i.e.,

$$\begin{aligned} & \text{minimize } MS_{SBV}^g(u) \text{ over } u \in SBV(\Omega) \text{ with} \\ & MS_{SBV}^g(u) := \frac{1}{2} \int_{\Omega} |\nabla u|^2 dx + \alpha \mathcal{H}^{d-1}(S_u) + \frac{\beta}{2} \|u - g\|_{L^2(\Omega)}^2. \end{aligned} \quad (\text{MS-SBV})$$

We note that a solution must fulfil  $u \in L^2(\Omega)$  and  $\nabla u \in L^2(\Omega; \mathbb{R}^d)$  for the value of the functional to be finite. Observe further that this formulation has the advantage that the edge set does not enter

as an explicit variable, but it is rather encoded in the function  $u$  through a jump. On the other hand, a potential drawback of this approach is given by the fact that now  $u$  is contained in a space that is neither reflexive nor separable. For details on the equivalence of the two formulations as well as the existence proof for  $g \in L^\infty(\Omega)$  the reader is referred to [DGCL89] or to [AFP00, Section 7].

For (MS-SBV) we are now able to derive an analogous version of Theorem 5, when we consider the solution operator as a mapping from  $L^\infty(\Omega)$  to  $L^2(\Omega)$ . Let us briefly discuss the requirement on  $g$ . In fact, from a practical view point an image signal is the product of a technical measurement process producing data in a fixed and finite range. Therefore the image signal will typically be essentially bounded. From the mathematical viewpoint we seek to apply a result by Ambrosio addressing SBV functions as well as Theorem 6. In order to do so, we need bounded sequences in  $L^\infty(\Omega)$  and a separable solution space  $L^2(\Omega) \supset SBV(\Omega)$ .

**Theorem 8.** *Let  $T : L^\infty(\Omega) \rightrightarrows L^2(\Omega)$  be the solution mapping of the Mumford–Shah problem (MS-SBV), i.e.,  $T(g) := \operatorname{argmin} \{MS_{SBV}^g(u) : u \in SBV(\Omega)\}$ . Then the following properties hold:*

(i)  *$T$  has nonempty and compact values.*

(ii) *For all closed sets  $C \subset L^2(\Omega)$  it holds that  $T^{-1}(C) \subset L^2(\Omega)$  is closed.*

Before we provide the proof, we also state that there exists a measurable selection of minimizers, i.e., a Borel function  $u_{\text{sel}} : L^\infty(\Omega) \rightarrow L^2(\Omega)$  with  $u_{\text{sel}}(g) \in T(g)$  for all  $g \in L^\infty(\Omega)$ .

*Proof.* Let  $g \in L^\infty(\Omega)$ .

*Step 1* (The set  $T(g) \subset L^2(\Omega)$  is compact). We know that the Mumford–Shah problem has a solution (cf. [Amb90, Example 5.2]). Hence  $T(g) \neq \emptyset$ . Let  $(u_n)_{n \in \mathbb{N}} \subset T(g) \subset SBV(\Omega)$  be a sequence of solutions. By the same truncation argument as in [Amb90] we obtain  $\|u_n\|_{L^\infty(\Omega)} \leq \|g\|_{L^\infty(\Omega)}$  for all  $n \in \mathbb{N}$ . Thus,  $(u_n)_{n \in \mathbb{N}}$  is uniformly bounded and by using

$$m := \min_{v \in SBV(\Omega)} MS_{SBV}^g(v) = MS_{SBV}^g(u_n) \geq \frac{1}{2} \int_{\Omega} |\nabla u_n|^2 dx + \alpha \mathcal{H}^{d-1}(S_{u_n}),$$

we obtain that the conditions of the compactness result of Ambrosio [Amb89] are fulfilled. Hence, there exist a subsequence  $(u_{n_k})_{k \in \mathbb{N}}$  as well as  $u \in SBV(\Omega) \cap L^\infty(\Omega)$  such that  $u_{n_k} \rightarrow u$  almost everywhere and the gradients  $\nabla u_{n_k} \rightharpoonup \nabla u$  converge weakly in  $L^2(\Omega; \mathbb{R}^d)$ . Due to the  $L^\infty$ -boundedness we obtain  $u_{n_k} \rightarrow u$  in  $L^2(\Omega)$  by dominated convergence. Moreover, we get by the lower semi-continuity of the norm and the second part of Ambrosio’s compactness result that

$$m \leq MS_{SBV}^g(u) \leq \liminf_{n \rightarrow \infty} MS_{SBV}^g(u_n) = m$$

and eventually  $u \in T(g)$ . This proves the compactness of  $T(g)$  in  $L^2(\Omega)$ .

*Step 2* ( $T^{-1}(C) \subset L^\infty(\Omega)$  is closed for all closed subsets  $C \subset L^2(\Omega)$ ). Let  $C \subset L^2(\Omega)$  be closed. We choose an arbitrary sequence  $(g_n)_{n \in \mathbb{N}} \subset T^{-1}(C)$  convergent in  $L^\infty(\Omega)$  towards  $g \in L^\infty(\Omega)$ . Then for all  $n \in \mathbb{N}$  there exists  $u_n \in T(g_n)$  that minimizes the corresponding Mumford–Shah functional. It thus holds that  $MS_{SBV}^{g_n}(u_n) \leq MS_{SBV}^{g_n}(v)$  for all  $v \in SBV(\Omega)$ . By plugging in  $v = 0$  and using the uniform boundedness of  $g_n$  in  $L^\infty$  we recognize that the conditions of Ambrosio’s compactness result are fulfilled, and therefore a subsequence  $(u_{n_k})_{k \in \mathbb{N}}$  as well as a limit  $u$  exist. By the above arguments we see  $u_{n_k} \rightarrow u$  in  $L^2(\Omega)$  and by the closedness assumption also  $u \in C$ . Together with the convergence of  $(g_{n_k})_{k \in \mathbb{N}}$ , we find analogously to Step 1 that

$$MS_{SBV}^g(u) \leq \liminf_{n \rightarrow \infty} MS_{SBV}^{g_{n_k}}(u_{n_k}) \leq \liminf_{n \rightarrow \infty} MS_{SBV}^{g_{n_k}}(v) = MS_{SBV}^g(v)$$

for all  $v \in SBV(\Omega)$ . Thus,  $u \in T(g) \cap C$ , and therefore  $T^{-1}(C)$  is a closed subset of  $L^\infty(\Omega)$ . We obtain the existence of a measurable selection again by the theorem of Kuratowski–Ryll–Nardzewski.  $\square$

While the above result provides some theoretical insights into the behavior of the computed image, it does not yield a way of how to treat the jump sets respectively the edge sets. We therefore return to the Ambrosio–Tortorelli approximation.

So far we have discussed measurable selections of operators on function space with respect to their topologies and induced Borel algebras. We consider now an abstract probability space  $(\Xi, \mathcal{F}, \mathbb{P})$  together with a random variable  $g : (\Xi, \mathcal{F}, \mathbb{P}) \rightarrow (L^2(\Omega), \mathcal{B}(L^2(\Omega)))$ . If we consider now a measurable selection  $g \mapsto (u_{\text{sel}}(g), z_{\text{sel}}(g))$  of minimizers (stationary points), then the composition  $\xi \mapsto (u_{\text{sel}}(g(\xi)), z_{\text{sel}}(g(\xi)))$  is a random variable of minimizers (stationary points). We refer to the respective components simply as  $\xi \mapsto (u(\xi), z(\xi))$ .

After establishing the existence of measurable selections, the next step is to derive a variational characterization, i.e. we seek to interpret selections of stationary points as solutions of a stochastic partial differential equation (SPDE) and, respectively, selections of minimizers as minimizer of an optimization problem. For this purpose we need to introduce elements of Bochner space theory.

First we equip the (abstract) measurable space  $(\Xi, \mathcal{F})$  with a probability measure  $\mathbb{P}$ . The resulting measure space induces a Lebesgue integral for  $\mathbb{R}$ -valued random variables denoted by

$$\mathbb{E}[\cdot] := \int_{\Xi} (\cdot) \, d\mathbb{P},$$

provided the argument is Lebesgue integrable. For a measurable set  $M \in \mathcal{F}$ , we define the characteristic function as

$$\mathbb{1}_M(\xi) := \begin{cases} 1 & \text{if } \xi \in M, \\ 0 & \text{else,} \end{cases}$$

and remark that  $\mathbb{E}[\mathbb{1}_M] = \mathbb{P}[M]$  holds. Let now  $V$  be a separable Banach space and  $p \in [1, +\infty]$ . The space of *Bochner  $p$ -integrable functions* is denoted by

$$L^p(\Xi, \mathcal{F}, \mathbb{P}; V) = \{v : \Xi \rightarrow V : v \text{ is measurable with } \|v\|_V \in L^p(\Xi, \mathcal{F}, \mathbb{P})\}.$$

Since we do not change  $(\Xi, \mathcal{F})$ , we also denote the above space by  $L^p_{\mathbb{P}}(V)$ . For all  $v \in L^p_{\mathbb{P}}(V)$  the *Bochner integral*

$$\mathbb{E}[v] := \int_{\Xi} v \, d\mathbb{P}$$

exists and is referred to as the *expectation*. Since the Lebesgue integral is just a special case of the Bochner integral with  $V = \mathbb{R}$ , there is no risk of confusion. For an overview on Bochner spaces the interested reader is referred to e.g. [HvNVW16, Section 1]. For  $V = H^1(\Omega)$  we also write  $H_{\mathbb{P}} := L^2_{\mathbb{P}}(H^1(\Omega))$  as well as

$$X_{\mathbb{P}} := \{v \in H_{\mathbb{P}} : \exists C \in (0, \infty) : |v| \leq C \text{ holds } \lambda^d \otimes \mathbb{P}\text{-a.e.}\},$$

where we write  $\lambda^d \otimes \mathbb{P}$  for the product measure on  $\Omega \times \Xi$ , equipped with its product  $\sigma$ -algebra  $\mathcal{B}(\Omega) \otimes \mathcal{F}$ . The spaces are equipped with the norms

$$\begin{aligned} \|\cdot\|_{H_{\mathbb{P}}}^2 &:= \mathbb{E}[\|\cdot\|_{H^1(\Omega)}^2], \text{ and} \\ \|\cdot\|_{X_{\mathbb{P}}}^2 &:= \mathbb{E}[\|\cdot\|_{H^1(\Omega)}^2] + \|\cdot\|_{L^\infty(\Omega \times \Xi, \lambda^d \otimes \mathbb{P})}^2 \end{aligned}$$

respectively. In any case, for our random edge indicator  $z \in X_{\mathbb{P}}$  we obtain  $0 \leq z \leq 1$  holding  $\lambda^d \otimes \mathbb{P}$ -almost everywhere, and in combination with the a priori estimate (3b) the boundedness in  $X_{\mathbb{P}}$  and  $L_{\mathbb{P}}^{\infty}(H^1(\Omega))$ . If we, moreover, assume that  $g \in L_{\mathbb{P}}^p(L^2(\Omega))$ , then also  $u \in L_{\mathbb{P}}^p(H^1(\Omega))$ . From now on let  $g \in L_{\mathbb{P}}^2(L^2(\Omega))$ .

Our first target is the SPDE-characterization. For this purpose, let  $(u, z)$  be a random variable of stationary points. Choosing arbitrary  $\varphi \in H_{\mathbb{P}}$  and  $\psi \in X_{\mathbb{P}}$  we prove the measurability of  $\xi \mapsto a^{g(\xi)}(u(\xi), z(\xi); \varphi(\xi))$  as well as  $\xi \mapsto b(u(\xi), z(\xi); \psi(\xi))$ . The proof is only carried out for the latter one as the proof for the former is analogous.

Indeed, for  $c \in \mathbb{R}$  we consider the preimage of  $(-\infty, c]$  under  $b$ , which is

$$B := \{(u', z', \psi') \in H^1(\Omega) \times K \times B_R^{L^{\infty}}(0) : b(u', z'; \psi') \leq c\},$$

where  $R > 0$  and  $B_R^{L^{\infty}}(0) := \{\tilde{\psi} \in X : \|\tilde{\psi}\|_{L^{\infty}} \leq R\}$ . We seek to show the closedness of  $B$  in  $H^1(\Omega; \mathbb{R}^3)$ . Let  $(u'_n, z'_n, \psi'_n) \rightarrow (u', z', \psi')$  converge in  $H^1(\Omega; \mathbb{R}^3)$ . By dominated convergence there exists a subsequence (not relabeled) such that  $z'_n$  and  $\psi'_n$  converge pointwise almost everywhere. It also holds that

$$\begin{aligned} & \left| \int_{\Omega} |\nabla u'_n|^2 z'_n \psi'_n dx - \int_{\Omega} |\nabla u'|^2 z' \psi' dx \right| \\ &= \left| \int_{\Omega} (|\nabla u'_n|^2 - |\nabla u'|^2) z'_n \psi'_n dx - \int_{\Omega} |\nabla u'|^2 (z' \psi' - z'_n \psi'_n) dx \right| \\ &\leq R \|\nabla u'_n - \nabla u'\|_{L^2(\Omega)} \cdot \|\nabla u'_n + \nabla u'\|_{L^2(\Omega)} + \left| \int_{\Omega} |\nabla u'|^2 (z' \psi' - z'_n \psi'_n) dx \right|. \end{aligned}$$

By the strong convergence of  $\nabla u'_n$  the first part tends to zero, and by the pointwise convergence of  $z'_n \psi'_n$  and  $|\nabla u'|^2 |z'_n \psi'_n - z' \psi'| \leq 2R |\nabla u'|^2$  the second part approaches zero by dominated convergence, as well. Hence we obtain

$$b(u', z'; \psi') = \lim_{n \rightarrow \infty} b(u'_n, z'_n; \psi'_n) \leq c,$$

yielding  $(u', z', \psi') \in B$ . By the  $H^1$ -measurability of  $u, z$  and  $\psi$  we get with  $R := \|\psi\|_{X_{\mathbb{P}}}$  that

$$\{\xi \in \Xi : b(u(\xi), z(\xi), \psi(\xi)) \leq c\} = \{\xi \in \Xi : (u(\xi), z(\xi), \psi(\xi)) \in B\} \in \mathcal{F}$$

holds for all  $c \in \mathbb{R}$ . This implies the measurability of  $\xi \mapsto b(u(\xi), z(\xi); \psi(\xi))$ . Analogously we obtain the measurability of  $a^{g(\cdot)}(u(\cdot), z(\cdot); \varphi(\cdot))$ .

Using the Cauchy-Schwarz inequality we obtain for  $u, \varphi \in H_{\mathbb{P}}$  and  $z, \psi \in X_{\mathbb{P}}$

$$\mathbb{E} [|a^g(u, z; \varphi)|], \mathbb{E} [|b(u, z; \psi)|] < \infty.$$

Hence, every measurable selection of stationary points fulfils the SPDE-system (in weak form)

$$\mathbb{E} [a^g(u, z; \varphi)] = 0 \quad \text{for all } \varphi \in H_{\mathbb{P}}, \tag{5a}$$

$$\mathbb{E} [b(u, z; \psi)] = 0 \quad \text{for all } \psi \in X_{\mathbb{P}}. \tag{5b}$$

Consequently, every measurable selection of stationary points is also a solution of the SPDE (5). The converse holds true as well. Let  $\varphi^* \in H^1(\Omega)$  and  $\psi^* \in X$  arbitrary. Then we define the following events

$$A_{\Xi} := \{\xi \in \Xi : a^{g(\xi)}(u(\xi), z(\xi); \varphi^*) \geq 0\},$$

$$B_{\Xi} := \{\xi \in \Xi : b(u(\xi), z(\xi), \psi^*) \geq 0\}.$$

Plugging  $\varphi = \varphi^* \mathbb{1}_{A_{\Xi}} - \varphi^* \mathbb{1}_{A_{\Xi}^c}$  and  $\psi = \psi^* \mathbb{1}_{B_{\Xi}} - \psi^* \mathbb{1}_{B_{\Xi}^c}$  into (5) we obtain

$$0 = \mathbb{E} [a^g(u, z; \varphi)] = \mathbb{E} [a^g(u, z; \varphi^*) \mathbb{1}_{A_{\Xi}} - a^g(u, z; \varphi^*) \mathbb{1}_{A_{\Xi}^c}] = \mathbb{E} [|a^g(u, z; \varphi^*)|]$$

and analogously  $0 = \mathbb{E} [|b(u, z; \psi^*)|]$ , from which we deduce  $a(u, z; \varphi^*) = 0$   $\mathbb{P}$ -a.s. as well as  $b(u, z; \psi^*) = 0$   $\mathbb{P}$ -a.s. for all  $\varphi^* \in H^1(\Omega)$  and  $\psi \in X$ . From this we see the equivalence of the scenariowise interpretation of (2) and the SPDE-system (5).

After the characterization of stationary points we want to characterize measurable selections of minimizers in a manner similar to the interchange of minimization and integration in [RWW09, Chapter 14, Section F]. For this purpose we define the constraint set

$$K_{\mathbb{P}} := \{v \in H_{\mathbb{P}} : 0 \leq v \leq 1 \text{ holds } \lambda^d \otimes \mathbb{P}\text{-almost everywhere}\}$$

and formulate the following minimization problem

$$\text{minimize } \mathbb{E} [AT_{\varepsilon}^g(u, z)] \text{ over } (u, z) \in H_{\mathbb{P}} \times K_{\mathbb{P}}. \quad (6)$$

Using the same arguments as for the deterministic Ambrosio–Tortorelli problem we obtain the following result.

**Theorem 9.** *The problem (6) admits a solution, and one can relax the constraint  $z \in K_{\mathbb{P}}$  to  $z \in X_{\mathbb{P}}$ , i.e.,*

$$\min_{(u, z) \in H_{\mathbb{P}} \times X_{\mathbb{P}}} \mathbb{E} [AT_{\varepsilon}^g(u, z)] = \min_{(u, z) \in H_{\mathbb{P}} \times K_{\mathbb{P}}} \mathbb{E} [AT_{\varepsilon}^g(u, z)].$$

Moreover, the functional  $\mathbb{E}[AT_{\varepsilon}^g(\cdot, \cdot)] : H_{\mathbb{P}} \times X_{\mathbb{P}} \rightarrow \mathbb{R}$  is Fréchet differentiable, and for every minimizer  $(u, z) \in H_{\mathbb{P}} \times X_{\mathbb{P}}$  the stationarity system (5) is satisfied.

Let  $(u^*, z^*)$  be a measurable selection of minimizers and  $(\bar{u}, \bar{z})$  a minimizer of (6). From this we obtain  $AT_{\varepsilon}^g(u^*, z^*) \leq AT_{\varepsilon}^g(\bar{u}, \bar{z})$   $\mathbb{P}$ -a.s. Taking the expectation yields  $\mathbb{E}[AT_{\varepsilon}^g(u^*, z^*)] \leq \mathbb{E}[AT_{\varepsilon}^g(\bar{u}, \bar{z})]$ . Since  $(\bar{u}, \bar{z})$  minimizes (6), it holds that  $\mathbb{E}[AT_{\varepsilon}^g(\bar{u}, \bar{z})] \leq \mathbb{E}[AT_{\varepsilon}^g(u^*, z^*)]$ . Hence  $(u^*, z^*)$  is a solution of (6) and it follows by the above scenariowise inequality that  $AT_{\varepsilon}^g(u^*, z^*) = AT_{\varepsilon}^g(\bar{u}, \bar{z})$   $\mathbb{P}$ -a.s. The solution of the minimization problem for the space of Banach space valued random variables is therefore identical to the solution in the sense of  $\mathbb{P}$ -almost every scenario. We summarize our findings on the characterization of measurable selections in the following theorem.

**Theorem 10.** *Let  $g \in L_{\mathbb{P}}^2(L^2(\Omega))$ . A random variable  $(u, z) \in H_{\mathbb{P}} \times K_{\mathbb{P}}$  is a solution of (6) if and only if it is a  $\mathbb{P}$ -a.s. measurable selection of minimizers, and it solves (5) if and only if it is a  $\mathbb{P}$ -a.s. measurable selection of stationary points.*

The purpose of this section is to give a mathematical meaning to random processed image and edges. The approximation of the Mumford–Shah problem yields the existence of random variables that solve the nondeterministic problem scenario-wise. In this sense, a measurable selection of minimizers serves the purpose of a random reconstruction and a random edge. Our numerical treatment of noise propagation in the following section is based on this observation.

## 4 Numerical methods

As we are concerned with computing the influence of the signal's randomness on the edge, our problem class falls into the realm of *forward problems* in the terminology of uncertainty quantification (UQ). For the sake of accessibility, we briefly summarize some of the main challenges also when designing numerical solution schemes.

- 1 The solution may be non-unique.
- 2 Due to the lack of convexity of the solution sets, we are unable to use the *Michael selection theorem* (cf. [Bee93, Theorem 6.6.3]) to establish the existence of *continuous* selections.

These facts challenge the use of many popular methods in UQ. Indeed:

- 3 The Wiener Chaos expansion is significantly troubled by the local constraint  $0 \leq z \leq 1$ . The global orthonormal expansion is not compatible to the pointwise inequality constraints. For a computational approach in this direction we refer to [PKP13].
- 4 Stochastic collocation is essentially an interpolation method and necessarily needs point evaluations and continuous dependence on the parameter.

The above techniques seek to approximate the entire random variable by discrete objects. A different approach is based on the definition of a *quantity of interest* and its approximation using the techniques of statistics and numerics. In this paper we employ Monte-Carlo based techniques, which rely on the *strong law of large numbers* in a Hilbert space setting. The latter one is stated here for the sake of convenience. For a proof see, e.g., [CZ11].

**Theorem 11.** *Let  $H$  be a separable Hilbert space and  $(Z_n)_{n \in \mathbb{N}}$  a sequence of pairwise independent, identically distributed (iid.) random variables in  $L^1_{\mathbb{P}}(H)$  with  $\mathbb{E}[Z_1] = m$ . Then*

$$\hat{m}_n := \frac{1}{n} \sum_{i=1}^n Z_i \rightarrow m \text{ converges } \mathbb{P}\text{-a.s.},$$

*i.e.,  $\mathbb{P}[\lim_{n \rightarrow \infty} \|\hat{m}_n - m\|_H = 0] = 1$  holds true. If, additionally,  $(Z_n)_{n \in \mathbb{N}}$  is a family in  $L^2_{\mathbb{P}}(H)$ , then  $\hat{m}_n \rightarrow m$  in  $L^2_{\mathbb{P}}(H)$ , as well.*

## 4.1 Generalized confidence intervals

In our context, one important quantity of interest is concerned with the identification of regions, where it is more likely to encounter an edge. In this sense, we would like to obtain a notion of *confidence intervals* for the edge indicator. To achieve this, we first provide a pointwise generalization of cumulative distributional functions (cdfs) in the following definition.

**Definition 12.** *Let  $Z$  be an  $L^2(\Omega)$ -valued random variable. The function  $F : L^2(\Omega) \rightarrow L^\infty(\Omega)$  is defined pointwise almost everywhere by*

$$F(\tau)(x) := \mathbb{P}[Z(x) \leq \tau(x)]$$

*for  $\tau \in L^2(\Omega)$  and is called the generalized cumulative distribution function (gcdf) of  $Z$ .*

Note that the above pointwise definition is justified by the Fubini theorem on integration on product spaces. Using this, we choose a *level of significance*  $s \in (0, 1)$  and define for  $z$  – interpreted as a random variable in  $L^2_{\mathbb{P}}(L^2(\Omega))$  only – the pointwise quantile  $\tau_s$  for (almost all)  $x \in \Omega$  by

$$\tau_s(x) := \inf\{t \geq 0 : \mathbb{P}[z(x) \leq t] \geq s\}.$$

This defines a *one-sided* confidence interval (the lower bound is zero), and two questions immediately arise: (i) Why is this object measurable, and (ii) how can this object be estimated in practice? Both questions are answered by using the following method.

---

**Algorithm 2:** Bisection method
 

---

**Data:** generalized cdf  $F$ , level of significance  $s \in (0, 1)$

**Result:** pointwise quantile  $\tau_s$

- 1 Set  $t_0^\ell = 0$  and  $t_0^u = 1$  as  $t_0^\ell, t_0^u \in L^2(\Omega)$ ;
  - 2 **for**  $n = 0, 1, 2, \dots$  **do**
    - 3 Calculate the mid point:  
 $t_n^{\text{mid}} = \frac{1}{2}(t_n^\ell + t_n^u)$ ;
    - 4 Update the upper/ lower bound:  
 $t_{n+1}^\ell = \mathbb{1}_{\{F(t_n^{\text{mid}}) \geq s\}} t_n^\ell + \mathbb{1}_{\{F(t_n^{\text{mid}}) < s\}} t_n^{\text{mid}}$ ;
    - 5  $t_{n+1}^u = \mathbb{1}_{\{F(t_n^{\text{mid}}) < s\}} t_n^u + \mathbb{1}_{\{F(t_n^{\text{mid}}) \geq s\}} t_n^{\text{mid}}$ ;
  - 6 **end**
- 

By definition the above sequences are nondecreasing, respectively, nonincreasing and almost everywhere it holds that  $t_n^\ell \leq t_{n+1}^\ell$  ( $t_n^u \geq t_{n+1}^u$ ) and, moreover,  $F(t_n^\ell) < s \leq F(t_n^u)$  for all  $n \in \mathbb{N}$ . By the monotonicity we also obtain the inequality  $0 \leq t_n^\ell < t_n^u \leq 1$  and hence the convergence  $t_n^\ell \nearrow t_*$  and  $t_n^u \searrow t^*$  pointwise almost everywhere on  $\Omega$ . Due to the measurability of  $t_n^\ell, t_n^u$ , we obtain by the pointwise convergence, that the limits are measurable as well. Moreover, by construction we get  $|t_{n+1}^\ell - t_{n+1}^u| \leq \frac{1}{2}|t_n^\ell - t_n^u|$  and inductively  $|t_n^\ell - t_n^u| \leq 2^{-n}$  almost everywhere. Therefore we get  $t_* = t^*$ .

In addition,  $t^*$  is the pointwise quantile of interest, which is equivalent to stating that for all  $\tilde{t} \leq t^*$  with  $\lambda^d(A) > 0$  with  $A := \{x \in \Omega : \tilde{t}(x) < t^*(x)\}$  also  $\mathbb{P}[z < \tilde{t}] < s$  holds a.e. on  $A$ . Indeed, by the choice of  $\tilde{t}$  there exists for all  $x \in A$  a number  $n = n(x) \in \mathbb{N}$  with  $\tilde{t}(x) < t_n^\ell(x)$ . Hence, by defining  $A_n := \{x \in \Omega : \tilde{t} < t_n^\ell\}$  we see that  $A_n \subset A_{n+1}$  and  $A = \bigcup_{n=1}^{\infty} A_n$ . By the construction of  $(t_n^\ell)_{n \in \mathbb{N}}$  it holds that  $\mathbb{P}[z < \tilde{t}] \leq \mathbb{P}[z \leq t_n^\ell] < s$  almost everywhere on  $A_n$ . We then see the required inequality  $\mathbb{P}[z < \tilde{t}] < s$  almost everywhere on  $A$  and have thus shown  $t^* = \tau_s$ .

In practice the exact gcdf is unknown and hence an approximation is needed. For this sake we define the *empirical quantiles* in full analogy to the classical case of  $\mathbb{R}$ -valued random variables.

**Definition 13.** Let  $Z$  be an  $L^2(\Omega)$ -valued random variable. Consider an iid. family of samples  $(Z_i)_{i \in \mathbb{N}}$  with distribution  $\mathbb{P}_Z$ . The empirical gcdf is defined as the mapping

$$\tau \mapsto F_n(\tau) := \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{\{Z_i \leq \tau\}} \quad (\in L^\infty(\Omega) \text{ for every realization}).$$

For proving convergence of  $(F_n)_{n \in \mathbb{N}}$  as introduced in Definition 13, we recall the statistical theory for  $\mathbb{R}$ -valued random variables and refer the reader to [vdV00] for proofs and additional information.

**Lemma 14.** Let  $(F_n)_{n \in \mathbb{N}}$  be a sequence of cdfs of associated  $\mathbb{R}$ -valued random variables and let  $F$  be the cdf of an  $\mathbb{R}$ -valued random variable. We assume  $F_n(t) \rightarrow F(t)$  for all  $t \in \mathbb{R}$ . Let  $s \in (0, 1)$  such that the quantile function  $F^{-1}$  is continuous in  $s$ . Then we have  $F_n^{-1}(s) \rightarrow F^{-1}(s)$ .

*Proof.* See [vdV00, Lemma 21.2 in a stronger form]. □

Based on the preceding lemma we now prove a convergence result for the  $s$ -quantiles of the empirical gcdfs.

**Theorem 15.** *Let  $Z$  be an  $\mathbb{R}$ -valued random variable with cdf  $F : \mathbb{R} \rightarrow [0, 1]$  and  $(Z_i)_{i \in \mathbb{N}}$  be an iid. family of samples with distribution  $\mathbb{P}_Z$ . For  $n \in \mathbb{N}$  we denote by  $F_n$  the empirical cdf. For a level of significance  $s \in (0, 1)$ , chosen such that the quantile function  $F^{-1}$  is continuous in  $s$ , the empirical quantiles  $F_n^{-1}(s) \rightarrow F^{-1}(s)$  converge with probability 1. If, moreover,  $0 \leq Z \leq 1$  holds  $\mathbb{P}$ -a.s., then in addition we get  $F_n^{-1}(s) \rightarrow F^{-1}(s)$  in  $L^p(\Xi, \mathcal{F}, \mathbb{P})$  for all  $p \in [1, \infty)$ .*

*Proof.* Using the strong law of large numbers we get  $F_n(t) \rightarrow \mathbb{E}[F_n(t)] = F(t)$   $\mathbb{P}$ -a.s for all  $t \in \mathbb{R}$ . By Lemma 14 we know that for every realization of  $F_n$  the corresponding quantiles converge and therefore  $F_n^{-1}(s) \rightarrow F^{-1}(s)$  with probability 1. If we assume  $0 \leq Z \leq 1$ , we get immediately  $0 \leq F_n^{-1}(s), F^{-1}(s) \leq 1$ . By dominated convergence we see that for all  $p \in [1, \infty)$  also  $\mathbb{E}[|F_n^{-1}(s) - F^{-1}(s)|^p] \rightarrow 0$  converges.  $\square$

After the above preparations, we can formulate a function space valued version of the previous theorem. For this, we need to extend the continuity assumption for the quantile in Lemma 14 on the level of significance to our situation. Note further that the gcdf induces for almost all  $x \in \mathbb{R}$  a cdf of an  $\mathbb{R}$ -valued random variable reading as  $\mathbb{R} \ni t \mapsto \mathbb{P}[z(x) \leq t] = F(t)(x)$ . Hence one can interpret the quantile function in a pointwise fashion as a function  $s \mapsto (F(\cdot)(x))^{-1}(s)$  for almost all  $x \in \Omega$ . This observation leads to the following assumption.

**Assumption 16.** *Let  $s \in (0, 1)$  be fixed such that*

$$N := \{x \in \Omega : s \mapsto (F(\cdot)(x))^{-1}(s) \text{ is not continuous in } s\}$$

*has Lebesgue measure zero.*

Utilizing Theorem 15 in a pointwise fashion, we obtain the following result.

**Theorem 17.** *Let  $z \in K_{\mathbb{P}}$  and  $(z_i)_{i \in \mathbb{N}}$  be iid. samplings according to the distribution  $\mathbb{P}_z$ . Moreover, fix the level of significance  $s \in (0, 1)$  fulfilling Assumption 16 with respect to the gcdf of  $z$ . Then the sequence of pointwise empirical  $s$ -quantiles  $\tau_{s,n}$  converges to the  $s$ -quantile  $\tau_s$  for almost all points with probability 1 and in  $L^p(\Omega \times \Xi)$  for all  $p \in [1, \infty)$ , i.e.*

$$\|\tau_{s,n} - \tau_s\|_{L^p(\Omega \times \Xi)} \rightarrow 0.$$

*Proof.* Since for almost all  $x \in \Omega$  the function  $t \mapsto \mathbb{P}[z(x) \leq t] = F(t)(x)$  is a cdf, we can apply Theorem 15 pointwise and obtain the convergence of  $F_n^{-1}(s) \rightarrow F^{-1}(s)$  pointwise  $\lambda^d \otimes \mathbb{P}$ -almost everywhere. Due to the pointwise restriction of  $z$  we know almost surely that  $\tau_{s,n} \in [0, 1]$  a.e. on  $\Omega$ . Since  $|\tau_{s,n} - \tau_s| \leq 1$  holds  $\lambda^d \otimes \mathbb{P}$ -a.e. we derive by dominated convergence

$$\|\tau_{s,n} - \tau_s\|_{L^p(\Omega \times \Xi)}^p = \int_{\Omega} \mathbb{E}[|\tau_{s,n} - \tau_s|^p] dx \rightarrow 0,$$

which ends the proof.  $\square$

In this subsection we have encountered the pointwise quantile as quantity of interest for the investigation of random edges. The associated convergence analysis yields a foundation for a sample-based approximation method for the gcdf and the pointwise quantile. The bisection method from the beginning of this section can then be used to efficiently approximate the empirical quantile for a given sample size.

## 4.2 Variance reduction by control variates

In contrast to the previous section, where we focus on quantiles, we discuss here the expected value as a quantity of interest. Therefore we consider the random edge indicator as an object in  $L^2_{\mathbb{P}}(L^2(\Omega))$  and aim at employing the Monte-Carlo method here as well. Since the convergence order is always  $\frac{1}{2}$ , the convergence speed is merely depending on the variance. Associated acceleration methods aim at the formulation of a new random variable with the same expectation as the original one, but a smaller variance. Several such methods do exist mainly in finite dimensions (cf. [Owe13] for an overview). We pursue an approach based on *control variates*. For convenience we first review the one-dimensional case aiming at an infinite-dimensional generalization.

Let  $Z \in L^2(\Xi, \mathcal{F}, \mathbb{P})$  be an  $\mathbb{R}$ -valued random variable. Our aim is to estimate  $\mathbb{E}[Z]$ . Suppose, we have another random variable  $Y$ , called the *control variate*, together with its (exact) expected value. Then we sample from the variable

$$Z_{\lambda} = Z + \lambda(Y - \mathbb{E}[Y])$$

for a yet to be chosen  $\lambda \in \mathbb{R}$ . Obviously, it holds that  $\mathbb{E}[Z_{\lambda}] = \mathbb{E}[Z]$  and the variance of  $Z_{\lambda}$  is minimal with respect to the weighting parameter  $\lambda$  for the choice

$$\lambda = -\frac{\text{Cov}(Z, Y)}{\text{Var}(Y)}$$

where  $\text{Cov}(Z, Y) := \mathbb{E}[(Z - \mathbb{E}[Z]) \cdot (Y - \mathbb{E}[Y])]$  and  $\text{Var}(Y) := \mathbb{E}[(Y - \mathbb{E}[Y])^2]$ . The minimal variance for  $Z_{\lambda}$  then becomes

$$\text{Var}(Z_{\lambda}) = \text{Var}(Z) - \frac{\text{Cov}(Z, Y)^2}{\text{Var}(Y)},$$

and, when  $\text{Cov}(Z, Y) \neq 0$ , a variance reduction is achieved (when compared to  $Z$ ). Of course, this concept can be extended to  $m$  control variates  $(Y^{(j)})_{j=1}^m$  by defining

$$Z_{\lambda} = Z + \sum_{j=1}^m \lambda_j (Y^{(j)} - \mathbb{E}[Y^{(j)}]).$$

The optimal  $\lambda = (\lambda_1, \dots, \lambda_m)^{\top}$  then reads

$$\lambda = -C_Y^+ c_Z,$$

where  $C_Y = (\text{Cov}(Y^{(i)}, Y^{(j)}))_{i,j=1,\dots,m} \in \mathbb{R}^{m \times m}$  denotes the covariance matrix of  $Y$  and the quantity  $c_Z = (\text{Cov}(Z, Y^{(j)}))_{j=1,\dots,m} \in \mathbb{R}^m$  denotes the covariance vector of the control variates and the desired target. For a matrix  $A \in \mathbb{R}^{m \times m}$  we denote its *Moore-Penrose-inverse* by  $A^+$ . In the same way as in the previous section we extend the above methodology in a pointwise fashion and introduce for this aim the following notation.

For  $U, V \in L^2_{\mathbb{P}}(L^2(\Omega))$  we define the *pointwise covariance* as

$$\mathbb{C}[U, V] = \mathbb{E}[(U - \mathbb{E}[U])(V - \mathbb{E}[V])] \in L^1(\Omega),$$

where  $\mathbb{E}$  is the expectation defined via the Bochner integral. Further we define the *pointwise variance* as

$$\mathbb{V}[U] := \mathbb{C}[U, U] = \mathbb{E}[(U - \mathbb{E}[U])^2]$$

in contrast to the *variance of  $U$* , which reads

$$\text{Var}(U) = \mathbb{E} \left[ \|U - \mathbb{E}[U]\|_{L^2(\Omega)}^2 \right] = \int_{\Omega} \mathbb{V}[U] \, dx.$$

For a random variable  $Z \in L^2_{\mathbb{P}}(L^2(\Omega))$ , a family  $(Y^{(j)})_{j=1,\dots,m} \subset L^2_{\mathbb{P}}(L^2(\Omega))$  of control variates, and a measurable function  $\lambda : \Omega \rightarrow \mathbb{R}^m$  we consider the random variable

$$Z_{\lambda} = Z + \sum_{j=1}^m \lambda_j (Y^{(j)} - \mathbb{E}[Y^{(j)}]).$$

The pointwise minimization of  $\mathbb{V}[Z_{\lambda}]$  with respect to  $\lambda$  gives us the minimizer of  $\text{Var}(Z_{\lambda})$  due to the Fubini theorem. Next we prove its measurability.

**Lemma 18.** *The mapping induced by the Moore-Penrose inverse  $(\cdot)^+ : \mathbb{R}^{m \times m} \rightarrow \mathbb{R}^{m \times m}$  with  $A \mapsto A^+$  is measurable.*

*Proof.* By Theorem 3.4 in [Alb72] it holds that

$$A^+ = \lim_{\delta \searrow 0} (A^{\top} A + \delta^2 I_m)^{-1} A^{\top}.$$

Let  $GL_m(\mathbb{R})$  denote the set of invertible matrices. Since the mapping  $(\cdot)^{-1} : GL_m(\mathbb{R}) \rightarrow GL_m(\mathbb{R})$  is continuous, the functions  $f_k : \mathbb{R}^{m \times m} \rightarrow \mathbb{R}^{m \times m}$  defined by

$$A \mapsto f_k(A) = \left( A^{\top} A + \frac{1}{k} I_m \right)^{-1} A^{\top}$$

are continuous and thus measurable. Due to the pointwise convergence  $A^+ = \lim_{k \rightarrow \infty} f_k(A)$  also the mapping  $(\cdot)^+$  is measurable.  $\square$

Hence we have proven the following result.

**Proposition 19.** *The mapping  $\lambda : \Omega \rightarrow \mathbb{R}^m$ , defined by  $\lambda = -C_Y^+ c_Z$ , is measurable.*

*Proof.* Since the composition as well as the multiplication of measurable functions are measurable as well, we obtain the measurability of  $\lambda$  using the preceding Lemma.  $\square$

## 5 Numerical examples

Our numerical methods are next applied to a selection of real world images. In this respect, the given image data is a (discrete) pixel matrix  $(G_{jk})_{j,k=1}^m$  of size  $m \times m$  with values between 0 (black) and 255 (white). In order to utilize our continuous framework we interpret this pixel mask as a function in  $L^{\infty}(\Omega)$  with  $\Omega := (0, m)^2$ . This is done by decomposing  $\Omega$  (up to null sets) into open squares of size 1 and rewriting

$$g_0 = \sum_{j,k=1}^m G_{jk} \mathbb{1}_{Q_{jk}},$$

with  $Q_{jk} := (0, 1)^2 + (j-1, m-k)$ . Concerning the underlying degradation models, we focus on two cases: (i) *white noise* and (ii) *randomized motion blur*.

## 5.1 White noise model

In this setting we assume  $L := id_{L^2(\Omega)}$  and that the pixel values are corrupted by a family of independent, normally distributed random variables, whose standard deviation is  $\sigma > 0$  for all pixels. Then the random variable reads as

$$\xi = \sum_{j,k=1}^m \sigma \cdot \xi_{jk} \mathbb{1}_{Q_{jk}},$$

where  $\xi_{jk} \sim \mathcal{N}(0, 1)$  are iid. with standard normal distribution. In our experiments we set  $\sigma = 40$ .

## 5.2 Motion blur model

In addition to the noise model above we now also consider the *motion blur* operator  $\widehat{L}_\ell \in \mathcal{L}(L^2(\Omega))$  with horizontal movement defined by

$$\widehat{L}_\ell g_0(x) := \frac{1}{2} \int_{-1}^1 g_0(x + t\ell v) dt \text{ with } \ell \sim \mathcal{U}(0, \bar{\ell}),$$

and  $v = (1, 0)^\top$  the direction of motion. The parameter  $\ell$  controls the velocity of the motion, respectively the strength of the blurring effect. Since the integration domain exceeds  $\Omega$  we extend the argument  $g_0$  by zero outside its domain. In our model we assume the parameter  $\ell \sim \mathcal{U}(0, \bar{\ell})$  to be a uniformly distributed random variable, where we choose  $\bar{\ell} = \frac{m}{4}$  to be one fourth of the image size. In view of our underlying degradation framework, we decompose the above object into

$$g = \widehat{L}_\ell g_0 = Lg_0 + \xi,$$

with  $L \in \mathcal{L}(L^2(\Omega))$  the linear operator such that  $Lg_0 = \mathbb{E}[\widehat{L}_\ell g_0]$  holds for all  $g_0 \in L^2(\Omega)$  and  $\xi := \widehat{L}_\ell g_0 - Lg_0$  as centered random variable. The expected value is the subject of the following auxiliary result.

**Lemma 20.** *Let  $\widehat{L}_\ell g(x) := \frac{1}{2} \int_{-1}^1 g(x + t\ell v) dt$  as above and  $\ell \sim \mathcal{U}(0, \bar{\ell})$  be a uniformly distributed random variable with parameter  $\bar{\ell} > 0$ . Then the operator  $L$  has the form*

$$Lg(x) = \mathbb{E}[\widehat{L}_\ell g](x) = \frac{1}{2} \int_{-1}^1 g(x + t\bar{\ell}v) (-\log|t|) dt \text{ for } x \in \Omega.$$

*Proof.* It is easily seen that for each  $g \in C_0^\infty(\Omega)$  the mapping  $\ell \mapsto \widehat{L}_\ell g$  is (Lipschitz-)continuous. Using  $\|\widehat{L}_\ell\|_{\mathcal{L}(L^2(\Omega))} \leq 1$  for all  $\ell \in \mathbb{R}$  (shown similarly to the proof of  $\|L\|_{\mathcal{L}(L^2(\Omega))} \leq 1$  below) and the density of  $C_0^\infty(\Omega)$  in  $L^2(\Omega)$  we obtain by the Banach–Steinhaus theorem the continuity of  $\ell \mapsto \widehat{L}_\ell g$  for all  $g \in L^2(\Omega)$  and hence the measurability of the above random variable.

For the calculation of the expected value choose again  $g \in C_0^\infty(\mathbb{R}^d)$  arbitrarily. Then we obtain by direct computation

$$\begin{aligned} \mathbb{E}[\widehat{L}_\ell g](x) &= \frac{1}{\bar{\ell}} \int_0^{\bar{\ell}} \frac{1}{2} \int_{-1}^1 g(x + t\ell v) dt d\ell = \frac{1}{2} \int_0^1 \int_{-1}^1 g(x + t\ell \cdot \bar{\ell}v) dt d\ell \\ &= \frac{1}{2} \int_0^1 \frac{1}{\ell} \int_{-\ell}^{\ell} g(x + t \cdot \bar{\ell}v) dt d\ell = \frac{1}{2} \int_M \frac{1}{\ell} g(x + t \cdot \bar{\ell}v) d^2(t, \ell) =: (*). \end{aligned}$$

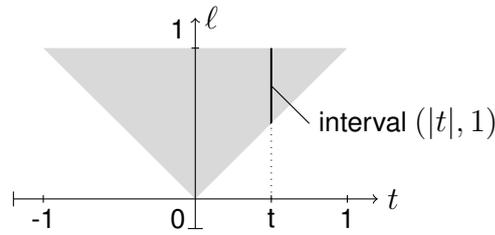


Figure 1: Depiction of the set of integration  $M$  with a vertical slice at  $t$ .

The set of integration  $M$  is depicted in Figure 1. Now we change the order of integration by slicing the set vertically instead of horizontally. A slice at  $t \in (-1, 1)$  reads then as the interval  $(|t|, 1)$ . Hence we proceed and obtain

$$\begin{aligned} (*) &= \frac{1}{2} \int_{-1}^1 \int_{|t|}^1 \frac{1}{\ell} g(x + t \cdot \bar{\ell}v) d\ell dt = \frac{1}{2} \int_{-1}^1 g(x + t \cdot \bar{\ell}v) \int_{|t|}^1 \frac{1}{\ell} d\ell dt \\ &= \frac{1}{2} \int_{-1}^1 g(x + t \cdot \bar{\ell}v) (-\log |t|) dt. \end{aligned}$$

Finally, we obtain the boundedness of  $L$  by

$$\begin{aligned} \|Lg\|_{L^2(\Omega)}^2 &\leq \int_{\mathbb{R}^d} |Lg|^2(x) dx \\ &= \frac{1}{4} \int_{\mathbb{R}^d} \int_{-1}^1 \int_{-1}^1 g(x + t \cdot \bar{\ell}v) g(x + s \cdot \bar{\ell}v) (-\log |t|) (-\log |s|) dt ds dx \\ &= \frac{1}{4} \int_{-1}^1 \int_{-1}^1 \left( \int_{\mathbb{R}^d} g(x + t \cdot \bar{\ell}v) g(x + s \cdot \bar{\ell}v) dx \right) (-\log |t|) (-\log |s|) ds dt \\ &\leq \frac{1}{4} \|g\|_{L^2(\Omega)}^2 \left( \int_{-1}^1 (-\log |t|) dt \right)^2 = \|g\|_{L^2(\Omega)}^2, \end{aligned}$$

from which we deduce the relation for  $L^2(\Omega)$  by using again the density of  $C_0^\infty(\Omega)$  in  $L^2(\Omega)$ .  $\square$

### 5.3 Experimental setup

As set of test images we take the image of the cameraman as well as an MRI scan of a human head in side view (see Figure 2 below). Since both of our methods are sample based, we need to minimize the Ambrosio–Tortorelli functional for each sample. This task can be parallelized. Of course we are unable to solve the minimization problem exactly and hence employ the finite element method (FEM) proposed in [BOS10] in combination with the splitting method for the corresponding discretization. In order to obtain an FE mesh, we divide each of the above mentioned pixel squares into two triangles by inserting alternately a diagonal; see Figure 3. Since the FEM approximation needs resources not depending on the image (like parts of the local stiffness matrices), we are able to share them among the minimization processes in order to reduce computational overhead. Our implementation is done in MATLAB. For the numerical solution of the PDEs we use the software package AFEM [AFE].

In the white noise case we use the parameters  $\alpha = 75, \beta = 1, \varepsilon = 10^{-2}$  and  $\eta = \varepsilon^{1.1}$ , and in the motion blur case we change to the parameters  $\alpha = 50, \beta = 10$  for the cameraman and

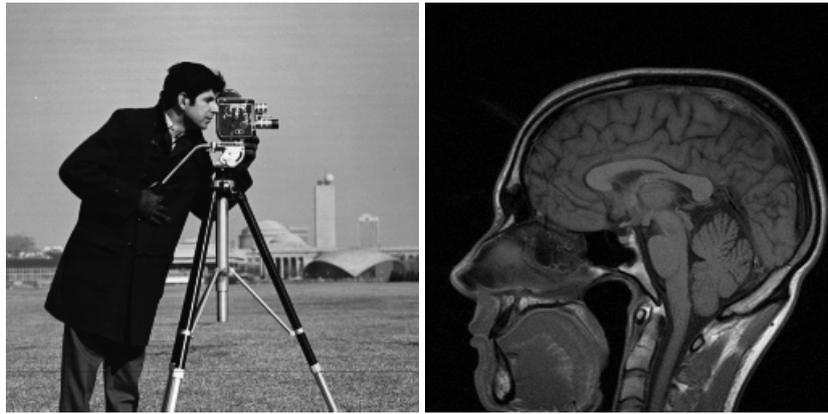


Figure 2: Left: image of the camera man, Right: MRI of the human head.

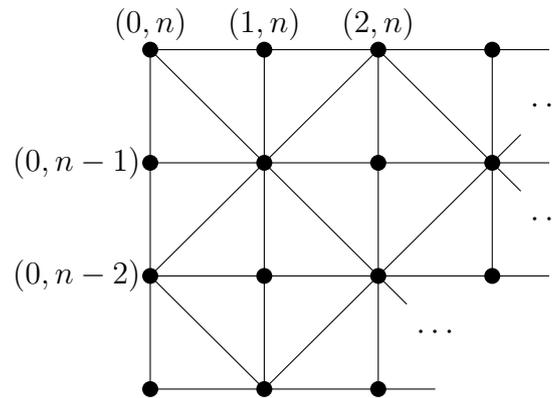


Figure 3: Depiction of the used triangulation in the upper left corner of the domain  $\Omega = (0, m) \times (0, n)$ .

$\alpha = 20, \beta = 10$  for the MRI. Since in the latter case the noise component  $\xi$  is depending on the original image, a change of parameters is justified.

Moreover, we emphasize that it is difficult to obtain the mass matrix with entries  $(L\lambda_i, L\lambda_j)_{L^2(\Omega)}$  exactly, where  $\lambda_i$  are nodal basis functions. Hence we are working with the interpolations  $I_{\mathcal{T}}L\lambda_i$  instead, where  $I_{\mathcal{T}} : C(\bar{\Omega}) \rightarrow C(\bar{\Omega})$  denotes the nodal interpolation operator defined by

$$(I_{\mathcal{T}}y)(x) := \sum_{i=1}^{(n+1)(m+1)} y(x_i)\lambda_i,$$

with  $(x_i)_{i=1, \dots, (n+1)(m+1)}$  being the coordinates of the nodes.

## 5.4 Pointwise quantiles

Firstly we consider the pointwise quantiles in order to detect areas, where an edge is *likely* to be found. As level of significance we take  $s = 0.9$  for the noise model cases. For the motion blur case we take  $s = 0.5$  – obtaining the median – and  $s = 0.1$  to better study the behavior of the quantiles. The visual results are depicted in Figure 4.

We note that the quantiles of experiments for the white noise case do not differ significantly from the edge indicators of the original images. In fact, on samples of edge indicators one can see little dark spots generated by the algorithm. Since they are not very distinct and their position is random, the

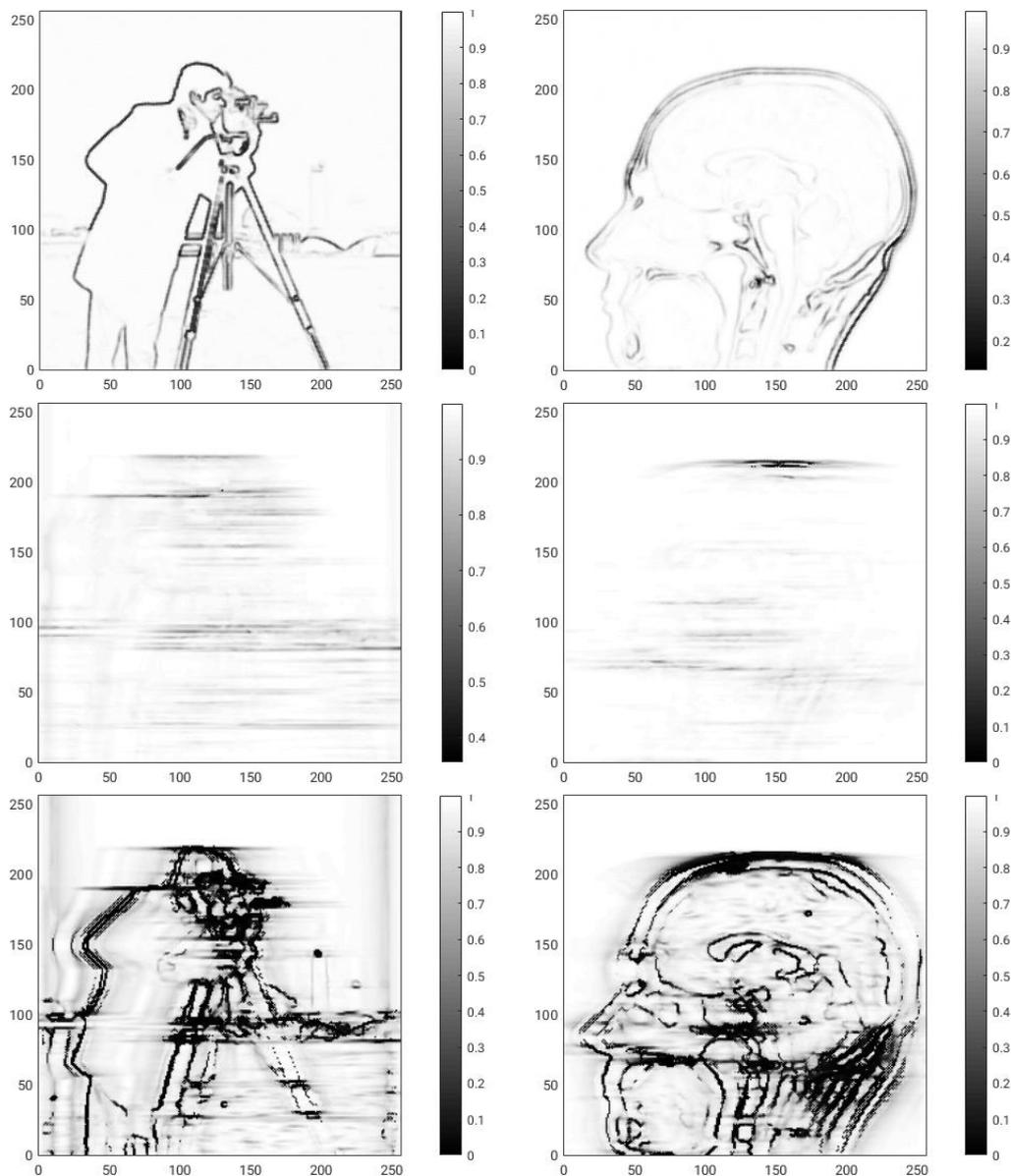


Figure 4: Numerical results for the quantiles. Top row: white noise model with  $s = 0.9$ , Middle row: motion blur model with  $s = 0.5$ , Bottom: motion blur model with  $s = 0.1$  Left: Cameraman, Right: MRI.

quantile does not react sensitive to them. A more sensitive reaction can be seen for the expected values in the following section.

In the motion blur model the reconstruction behavior has an influence. Since our blurring operator and the actually applied  $\widehat{L}_\ell$  differ, the reconstruction part tries to transport intensity where it would normally not belong to. This behavior leads to high oscillations (especially for small  $\ell$ ) as well as doubling effects (especially for big  $\ell$ ) in the corresponding reconstructed image. This results in the horizontal motion of strongly indicated edges as well as weak, blurred edges. Only very small parts of the edges – mainly horizontal ones – remain at the same position.

In order to further clarify this effect, we resort to an additional synthetic test image similar to the one from [HS06] for the parameters  $\alpha = 20$ ,  $\beta = 10$  and evaluate the quantiles using only 200 samples for a descending series of levels of significance. The results can be found in Figure 5.

One observes that for high levels of significance ( $s = 0.9$  and  $s = 0.75$ ) the only clearly recognizable edges are the horizontal ones (upper and lower edge of the bar, the lower edge of the triangle and at the upper and lower poles of the circle). When reducing  $s$  the moving edges gain importance, which results in gray areas around the places corresponding to the circle and the triangle. As it can be seen in the result for  $s = 0.3$ , the presence of distinct edges next to each other leads – due to the motion blur – to an overlap. This results in the detection of an edge at a position where for the original image no edge occurs at all. For low levels of significance  $s = 0.2$  and  $s = 0.1$  the edges corresponding to the original image are duplicated in horizontal direction around their original positions. In the last image this results in a broad band of edges.

Coming now back to our images considered in Figure 4 we can build on the above observations to get an improved understanding of the mechanisms leading to these results. As we have seen, edges in different directions occur in the quantile for very different values of  $s$ . Thus, we use here two levels  $s = 0.5$  and  $s = 0.1$  to get a better understanding. Again, we observe for the first value only gray areas as well as prolonged horizontal edges (e.g the shoulder of the cameraman and the top of the head in the MRI example). For the lower value of  $s = 0.1$  we observe doubled edges around the ones for the original images. In the cameraman image on the left and right boundary, grey horizontal ‘oscillations’ can be observed. This effect occurs since the image function has been extended by 0 outside of  $\Omega$ . Since the background is originally grey we generate here an edge (in contrast to the MRI image, where the background was already black in the first place). This also leads to some oscillatory effects on the left side due to “collisions” of the boundary and the man. We note that since our synthetic test image is an extreme example we do not have here such large bands of edges, but we still find duplications of edges.

## 5.5 Expected values

In this section we utilize the control variates technique to obtain the expected values of the edge indicators. Besides the mean values themselves we also want to access the variance reduction effect. In our experiments we use a total of 3 (respectively 4) control variates ( $y_j$ ) for the noise-only model (respectively motion blur model). Their choice is heuristic and specified in Table 1 below. For the noise case we choose a different edge detection approach using filter-based methods which is way cheaper than the Mumford–Shah or Ambrosio–Tortorelli based methods, respectively. Therefore we use the so-called *Marr–Hildreth* filter, which reads

$$\Delta(k_\sigma \star g)(x) = \Delta \left( \int_{\mathbb{R}^2} k_\sigma(\cdot - y)g(y)dy \right) (x) = \int_{\mathbb{R}^2} (\Delta k_\sigma)(x - y)g(y)dy,$$

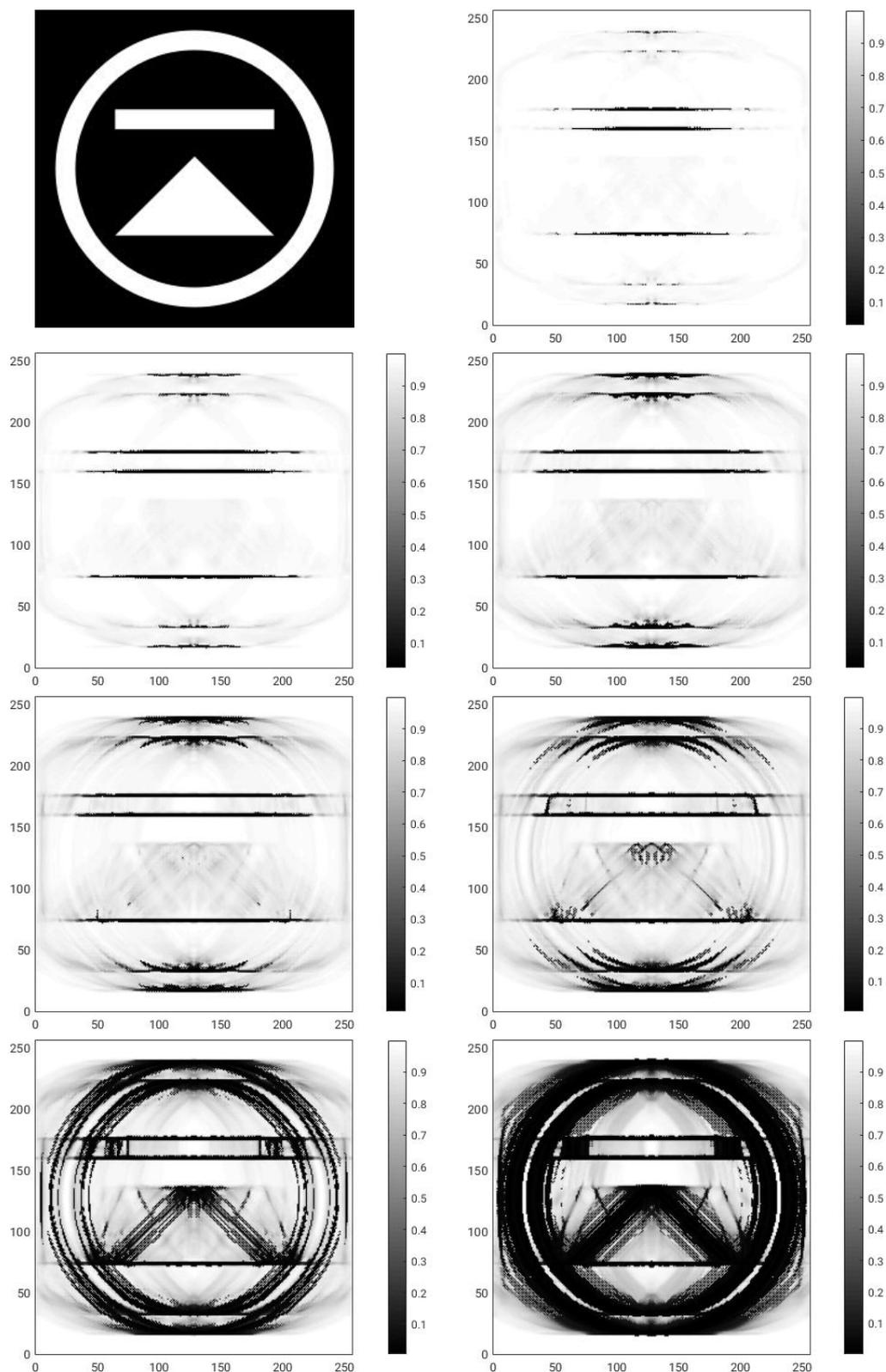


Figure 5: Numerical results of the pointwise quantiles applied to the synthetic test image in the upper left corner. Results for the levels of significance (from top to bottom, from left to right; starting in the upper right corner)  $s = 0.9, 0.75, 0.5, 0.4, 0.3, 0.2$  and  $0.1$ .

noise model	motion blur model
$y_1 :=  \Delta(k_1 \star g) $	$y_1 :=  \Delta(k_1 \star v_1) $
$y_2 := \exp(-0.005 \cdot y_1)$	$y_2 :=  \Delta(k_1 \star v_2) $
$y_3 := g$	$y_3 := \exp(-0.005 \cdot y_1)$
	$y_4 := \exp(-0.005 \cdot y_2)$

Table 1: Overview of the applied control variates.

where  $k_\sigma := \exp(-\frac{1}{2\sigma}|x|^2)$  denotes the Gaussian kernel. The idea is to apply a smoothing step to reduce on the one hand the possible influence of noise and, on the other hand, to increase the regularity from an image signal in  $L^2(\mathbb{R}^2)$  to a smoothed signal in  $C^\infty(\mathbb{R}^2)$ . The application of the Laplacian seeks to find sources and valleys in the gradient map, which resort to rapid changes in the gradient. Due to its structure the Marr-Hildreth filter is also called *Laplacian of Gaussian* in the literature. For more details on this the interested reader is referred to [CS05, section 2.6.2] and the references therein.

As control variates we then choose its absolute value as well as the exponential of its weighted negative. Especially the latter has a chance to behave at least locally similar to the edge indicator obtained by our model. Additionally we use the original image to get the noise applied to the image itself.

For motion blur, a direct application of the above technique would be inappropriate. Due to the presence of the operator  $L$  we first apply a reconstruction step to the image based on the following consideration: Having the edge indicator given, one would just need to solve the linear equation (2a) to obtain the restored image. Since the real edge indicator is not given and expensive to calculate, we perform the reconstruction with respect to the two extreme cases,  $z = 1$  corresponding to the a priori assumption of no edge occurring in the image at all, and  $z = 0$  corresponding to the case of edges appearing everywhere in the image. Based on this, we consider the functions  $v_1, v_2 \in H^1(\Omega)$  as solutions of the PDEs

$$\begin{aligned} -\eta\Delta v_1 + \beta L^* L v_1 &= \beta L^* g \text{ in } \Omega, & \frac{\partial v_1}{\partial \nu} &= 0 \text{ on } \partial\Omega, \\ -(1 + \eta)\Delta v_2 + \beta L^* L v_2 &= \beta L^* g \text{ in } \Omega, & \frac{\partial v_2}{\partial \nu} &= 0 \text{ on } \partial\Omega, \end{aligned}$$

where  $L^*$  is the adjoint of  $L$  and  $\nu$  denotes the outward unit normal to  $\partial\Omega$ . In this sense we obtain two reconstructions of the image with respect to two different parameter choices. As for the noise model case we apply again the Marr-Hildreth filter to  $v_1, v_2$  and use the absolute values as well as an exponential of its weighted negative.

To study the variance reduction effect, we compare the variances of the edge indicator  $z$  with the ones of the modified variable  $z_{cv}$ . Since our approach is based on the pointwise application of control variates we moreover investigate *where* the variance has been reduced. For this sake we plot the difference of the pointwise variances  $\mathbb{V}[z] - \mathbb{V}[z_{cv}]$ . We remark here that mathematically the difference can never be negative, but since our multiplier and our variances are just estimations, it can happen that scattered negative values do occur. Hence the plots have been truncated at zero from below to give a clearer view. Note that in the Figure 7 the colormap is inverted for better visibility.

An overall number of 1000 samples has been generated, where 100 of them have been used to approximate the multipliers for the control variates. Consequently, for the estimation of mean, variance and difference of pointwise variances the other 900 were used.

For the noise case we see in Figure 6 that the background is considerably darker in comparison to

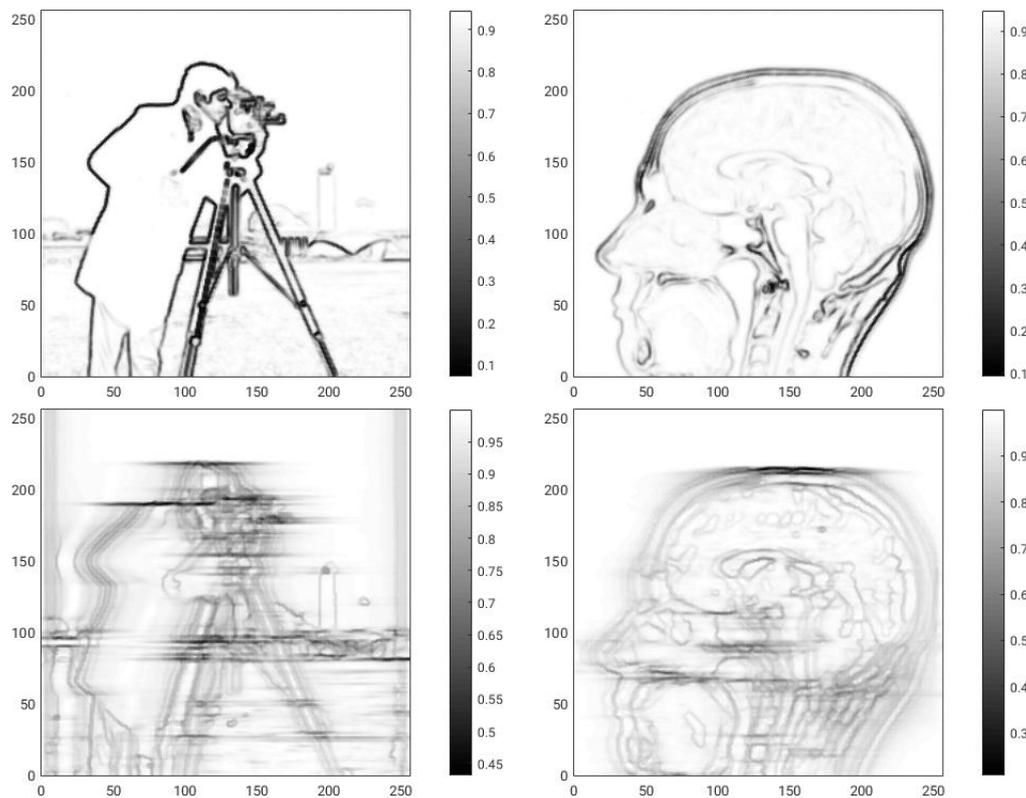


Figure 6: Numerical Results for the expected values. Top: white noise model, Bottom: motion blur model. Left: Cameraman, Right: MRI.

image	degradation type	$\text{Var}(z)$	$\text{Var}(z_{cv})$	ratio
cameraman	noise	305.808	185.297	1.650
MRI	noise	285.167	171.889	1.659
cameraman	motion blur	1580.497	616.192	2.565
MRI	motion blur	1739.972	601.837	2.891

Table 2: Variance reduction effects.

the quantiles. This behavior is caused by the isotropic distribution of little dark spots generated by the noise. Besides that, there is again a strong similarity to the edge indicator of the original image. Moreover, a noticeable variance reduction can be observed. By considering the plots of the difference of the pointwise variances we see a bigger improvement in regions, where many edges are close to each other.

For the motion blur model we see horizontal edges pulled longer. By closer inspection one finds several edges parallel and near to the original ones, such that they form a weak blurred band. This is caused by the motion of edges in horizontal direction described in the previous section. A significantly higher variance reduction effect can be observed, which is not surprising as the Mumford–Shah model has been designed to reduce the effect of noise. Regarding the difference of the pointwise variances we find the highest reductions in close proximity to the edges of the original image.

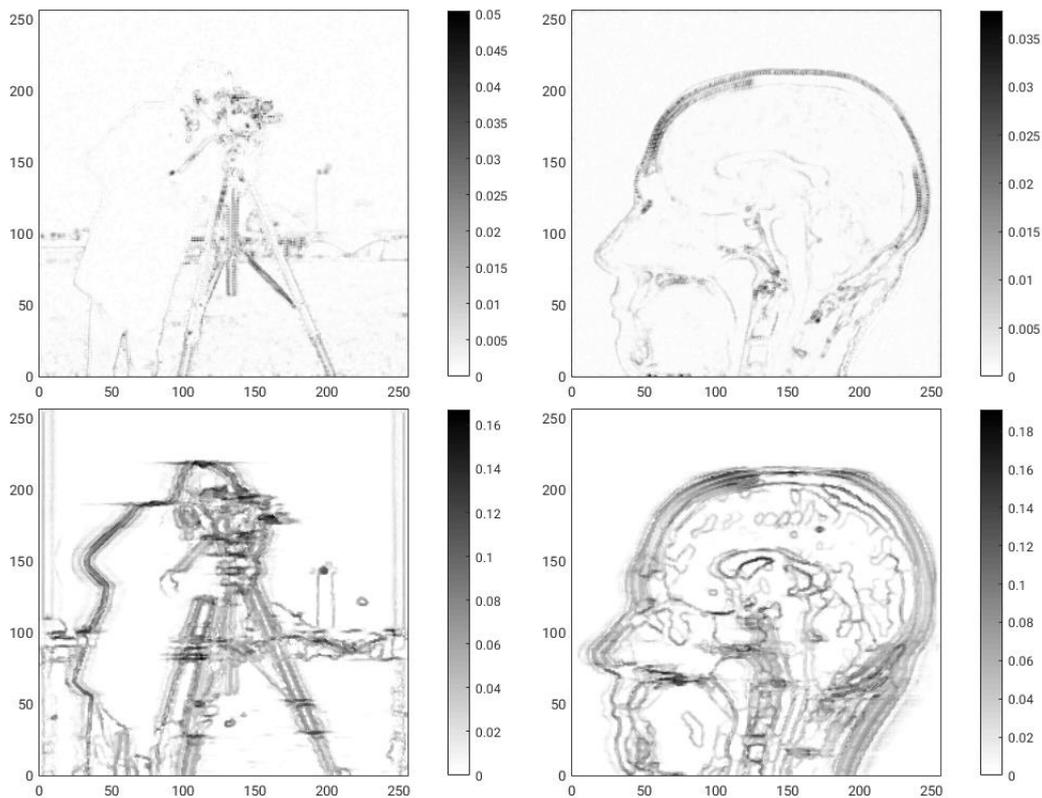


Figure 7: Differences of variance operators  $\mathbb{V}[z] - \mathbb{V}[z_{cv}]$  Top: white noise model, Bottom: motion blur model. Left: Cameraman, Right: MRI.

## 6 Conclusion

In this paper, we studied the influence of errors in images on edges resulting from an image segmentation procedure. Due to the inherent difficulty of performing uncertainty quantification for a geometric quantity we transitioned to the Ambrosio–Tortorelli problem to profit from the accessible vector space structure. Despite the lack of convexity or unique solutions we were able to establish a rigorous theoretical foundation for selections of solutions. To perform a quantitative treatment we proposed numerical methods addressing quantiles and the expected values and applied them to practically meaningful images and instances of the considered degradation model.

Since uncertainty quantification of geometric objects is a relatively new and scarcely investigated topic we offer here a possible access point for researchers dealing with problems having a similar geometric structure. From the practical viewpoint, our results might be of importance to obtain a qualitative understanding of the uncertainty associated with (reconstructed) edges in certain medical or machine vision applications.

## 7 A word of caution at the end

As an addendum we want to emphasize that the white noise model considered in this work *cannot* be interpreted as a pixelwise projection of a random variable with values in  $L^2(\Omega)$ . In the following we will make use of techniques from probability theory, which can for example be found in [Kle13]. In contrast to before let  $\Omega = (0, 1)^2$  and define for  $n \in \mathbb{N}$  a uniform grid  $Q^n = (Q_{jk}^n)_{j,k=1}^n$  such that

the area of a square around every pixel is  $\frac{1}{n^2}$ . Precisely speaking, we are going to show that there exists *no* random variable  $\xi^* \in L^2_{\mathbb{P}}$  such that the white noise with respect to this grid  $\xi_Q$  has the same distribution as  $\Pi_Q \xi^*$ , where the operator  $\Pi_Q \in \mathcal{L}(L^2(\Omega))$  is the piecewise projection onto the pixel grid defined by

$$\Pi_Q g := \sum_{j,k=1}^n \frac{1}{|Q_{jk}|} \int_{Q_{jk}} g dx \cdot \mathbb{1}_{Q_{jk}}.$$

Having  $\lambda^2(Q_{jk}^n) = \frac{1}{n^2}$  we introduce the orthonormal system  $(e_{jk}^n)_{j,k=1}^n$  with  $e_{jk}^n := n \mathbb{1}_{Q_{jk}^n}$  and rewrite the projection  $\Pi_{Q^n}$  as

$$\Pi_{Q^n} g = \sum_{j,k=1}^n (e_{jk}^n, g)_{L^2(\Omega)} \cdot e_{jk}^n.$$

The white noise reads as

$$\xi^n = \xi_{Q^n} = n \sum_{j,k=1}^n \xi_{jk} \mathbb{1}_{Q_{jk}^n} = \sum_{j,k=1}^n \xi_{jk} e_{jk}^n,$$

with  $(\xi_{jk})_{j,k=1}^n$  a family of independent, identically distributed random variables with standard normal distribution. It is straight forward to show that  $\Pi_{Q^n} \xi^{2n} = \xi^n$  holds for all  $n$ . For proving the non-existence of  $\xi^*$  as described above we assume the contrary. Using the fact that  $\text{diam}(Q_{jk}^n) = \frac{\sqrt{2}}{n}$  goes to zero for  $n \rightarrow \infty$  one can show by the density of smooth functions in  $L^2(\Omega)$  combined with a version of the Banach–Steinhaus theorem (cf. [Wer06, Korollar IV 2.5]) that the operators  $\Pi_{Q^n}$  converge pointwise to the identity on  $L^2(\Omega)$ .

Hence  $\Pi_{Q^n} \xi^* \rightarrow \xi^*$  converges  $\lambda^2 \otimes \mathbb{P}$ -a.e. In particular this yields that  $\xi_n$  – having the same distribution as  $\Pi_{Q^n} \xi^*$  – must converge *in distribution* to  $\xi^*$  meaning that the corresponding probability distributions  $\mu_n$  are weakly convergent in the sense of measures. This implies especially the pointwise convergence of the respective characteristic functions, reading for an arbitrary  $v \in L^2(\Omega)$  as

$$\varphi_n(v) := \int_{L^2(\Omega)} \exp(i_{\mathbb{C}}(w, v)_{L^2(\Omega)}) d\mu_n(w) = \mathbb{E} [\exp(i_{\mathbb{C}}(\xi^n, v)_{L^2(\Omega)})],$$

towards the characteristic function of  $\xi^*$ . The computation of  $\varphi_n$  yields

$$\begin{aligned} \varphi_n(v) &:= \mathbb{E} [\exp(i_{\mathbb{C}}(\xi^n, v)_{L^2(\Omega)})] = \mathbb{E} \left[ \exp \left( i_{\mathbb{C}} \sum_{j,k=1}^n \xi_{jk} (e_{jk}^n, v)_{L^2(\Omega)} \right) \right] \\ &= \mathbb{E} \left[ \prod_{j,k=1}^n \exp(i_{\mathbb{C}} \xi_{jk} (e_{jk}^n, v)_{L^2(\Omega)}) \right] = \prod_{j,k=1}^n \mathbb{E} [\exp(i_{\mathbb{C}} \xi_{jk} (e_{jk}^n, v)_{L^2(\Omega)})] \\ &= \prod_{j,k=1}^n \exp \left( -\frac{1}{2} (e_{jk}^n, v)_{L^2(\Omega)}^2 \right) = \exp \left( -\frac{1}{2} \sum_{j,k=1}^n (e_{jk}^n, v)_{L^2(\Omega)}^2 \right) \\ &= \exp \left( -\frac{1}{2} \|\Pi_{Q^n} v\|_{L^2(\Omega)}^2 \right). \end{aligned}$$

As we have seen above the linear operators  $\Pi_{Q^n}$  converge pointwise to the identity, such that we deduce the convergence

$$\varphi_n(v) \rightarrow \exp \left( -\frac{1}{2} \|v\|_{L^2(\Omega)}^2 \right) \text{ for all } v \in L^2(\Omega).$$

Since we have assumed that the  $\mu_n$  converge weakly, this implies that the characteristic function of the distribution of  $\xi^*$  must be  $\exp\left(-\frac{1}{2}\|v\|_H^2\right)$ . But according to [MR04, Proposition 1.2.11] there exists no probability measure with this particular characteristic function. This yields the anticipated contradiction.

Hence we deduce that the term *white noise* is only justified with respect to a given pixel grid.

## References

- [AB06] C.D. Aliprantis and K.C. Border. *Infinite Dimensional Analysis: A Hitchhiker's Guide*. Springer Berlin Heidelberg, 2006.
- [ABM14] H. Attouch, G. Buttazzo, and G. Michaille. *Variational Analysis in Sobolev and BV Spaces: Applications to PDEs and Optimization, Second Edition*. MOS-SIAM Series on Optimization. Society for Industrial and Applied Mathematics, 2014.
- [AF03] R.A. Adams and J.J.F. Fournier. *Sobolev Spaces*. Pure and Applied Mathematics. Elsevier Science, 2003.
- [AFE] AFEM software package. [www.math.hu-berlin.de/~ccafm/teachingAdvanced/Tutorials/afem.php](http://www.math.hu-berlin.de/~ccafm/teachingAdvanced/Tutorials/afem.php). last visited 10th of May 2019.
- [AFK08] G. Aubert, O. Faugeras, and P. Kornprobst. *Mathematical Problems in Image Processing: Partial Differential Equations and the Calculus of Variations*. Applied Mathematical Sciences. Springer New York, 2008.
- [AFP00] L. Ambrosio, N. Fusco, and D. Pallara. *Functions of Bounded Variation and Free Discontinuity Problems*. Oxford Science Publications. Clarendon Press, 2000.
- [Alb72] A. Albert. *Regression and the Moore-Penrose Pseudoinverse*. Mathematics in Science and Engineering. Elsevier Science, 1972.
- [Amb89] L. Ambrosio. A compactness theorem for a new class of functions of bounded variation. *Bollettino della Unione Matematica Italiana*, 3(4):857–881, 1989.
- [Amb90] L. Ambrosio. Existence theory for a new class of variational problems. *Archive for Rational Mechanics and Analysis*, 111(4):291–322, 1990.
- [AT92] L. Ambrosio and V. M. Tortorelli. On the approximation of free discontinuity problems. *Boll. Unione Mat. Ital., VII. Ser., B*, 6(1):105–123, 1992.
- [Ban08] I. Bankman. *Handbook of Medical Image Processing and Analysis*. Academic Press series in biomedical engineering. Elsevier Science, 2 edition, 2008.
- [Bee93] G. Beer. *Topologies on Closed and Closed Convex Sets*. Mathematics and Its Applications. Springer, 1993.
- [BOS10] S. Burke, C. Ortner, and E. Süli. An adaptive finite element approximation of a variational model of brittle fracture. *SIAM Journal on Numerical Analysis*, 48(3):980–1012, 2010.
- [Cha95] A. Chambolle. Image segmentation by variational methods: Mumford and Shah functional and the discrete approximations. *SIAM J. Appl. Math.*, 55(3):827–863, 1995.

- [CKSS97] V. Caselles, R. Kimmel, G. Sapiro, and C. Sbert. Minimal surfaces based object segmentation. *IEEE Trans. on Pattern Anal. Machine Intell.*, 19(4):394–398, April 1997.
- [Coh97] L. D. Cohen. Avoiding local minima for deformable curves in image analysis. In A. Le Méhauté, C. Rabut, and L. L. Schumaker, editors, *Curves and Surfaces with Applications in CAGD*, pages 77–84. Vanderbilt University Press, Nashville, TN, 1997.
- [CS05] T. F. Chan and J. Shen. *Image Processing and Analysis: Variational, PDE, Wavelet, and Stochastic Methods*. Other titles in applied mathematics. Society for Industrial and Applied Mathematics, 2005.
- [CZ11] Y.-X. Chen and W.-J. Zhu. Note on the strong law of large numbers in a Hilbert space. *General Mathematics*, 19(3), 2011.
- [DGCL89] E. De Giorgi, M. Carriero, and A. Leaci. Existence theorem for a minimum problem with free discontinuity set. *Archive for Rational Mechanics and Analysis*, 108(4):195–218, Nov 1989.
- [EG91] L.C. Evans and R.F. Gariepy. *Measure Theory and Fine Properties of Functions*. Studies in Advanced Mathematics. Taylor & Francis, 1991.
- [GG84] D. Geman and S. Geman. Stochastic relaxation, Gibbs distribution, and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6(6):721–741, 1984.
- [Hae10] J. Haehnle. *Numerical Analysis of the Mumford-Shah and Mumford-Shah-Euler Functionals for Sphere-Valued Functions, and applications to Numerical Image Processing*. dissertation, Universität Tübingen, 2010.
- [HL09] M. Hintermüller and A. Laurain. Multiphase image segmentation and modulation recovery based on shape and topological sensitivity. *Journal of Mathematical Imaging and Vision*, 35(1):1–22, 2009.
- [HR03] M. Hintermüller and W. Ring. A second order shape optimization approach for image segmentation. *SIAM J. Appl. Math.*, 64(2):442–467, 2003.
- [HR04] M. Hintermüller and W. Ring. An inexact Newton-CG-type active contour approach for the minimization of the Mumford-Shah functional. *J. Math. Imag. Vision*, 20:19–42, 2004.
- [HS06] M. Hintermüller and G. Stadler. An infeasible primal-dual algorithm for total bounded variation–based inf-convolution-type image restoration. *SIAM Journal on Scientific Computing*, 28(1):1–23, 2006.
- [HvNVW16] T. Hytönen, J. van Neerven, M. Veraar, and L. Weis. *Analysis in Banach Spaces: Volume I: Martingales and Littlewood-Paley Theory*. Ergebnisse der Mathematik und ihrer Grenzgebiete. 3. Folge / A Series of Modern Surveys in Mathematics. Springer International Publishing, 2016.
- [Kle13] A. Klenke. *Probability Theory: A Comprehensive Course*. Universitext. Springer London, 2013.
- [KWT87] M. Kass, A. Witkin, and D. Terzopoulos. Snakes; active contour models. *Int. J. of Computer Vision*, 1:321–331, 1987.

- [MR04] S. Maniglia and A. Rhandi. Gaussian measures on separable Hilbert spaces and applications. *Quaderni del Dipartimento di Matematica dell'Università del Salento*, 2004(1), 2004.
- [MS89] D. Mumford and J. Shah. Optimal approximations by piecewise smooth functions and associated variational problems. *Communications on pure and applied mathematics*, 42(5):577–685, 1989.
- [MS95] J.-M. Morel and S. Solimini. *Variational Methods in Image Segmentation*, volume 14 of *Progress in Nonlinear Differential Equations and their Applications*. Birkhäuser Boston Inc., Boston, MA, 1995. With seven image processing experiments.
- [Owe13] A. B. Owen. *Monte Carlo theory, methods and examples*. not published yet, see <http://statweb.stanford.edu/~owen/mc/>, 2013.
- [Par10] J.R. Parker. *Algorithms for Image Processing and Computer Vision*. EBL-Schweitzer. Wiley, 2 edition, 2010.
- [PKP13] T. Pätz, R. M. Kirby, and T. Preusser. Ambrosio-Tortorelli segmentation of stochastic images: Model extensions, theoretical investigations and numerical methods. *International Journal of Computer Vision*, 103(2):190–212, 2013.
- [RWW09] R.T. Rockafellar, M. Wets, and R.J.B. Wets. *Variational Analysis*. Grundlehren der mathematischen Wissenschaften. Springer Berlin Heidelberg, 2009.
- [Smi13] R.C. Smith. *Uncertainty Quantification: Theory, Implementation, and Applications*. Computational Science and Engineering. SIAM, 2013.
- [Sul15] T.J. Sullivan. *Introduction to Uncertainty Quantification*. Texts in Applied Mathematics. Springer International Publishing, 2015.
- [TSE16] F. Trèves, P.A. Smith, and S. Eilenberg. *Topological Vector Spaces, Distributions and Kernels: Pure and Applied Mathematics*. Number Bd. 25. Elsevier Science, 2016.
- [vdV00] A.W. van der Vaart. *Asymptotic Statistics*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, 2000.
- [Wer06] D. Werner. *Funktionalanalysis*. Springer-Lehrbuch. Springer Berlin Heidelberg, 2006.