

# Optimality Bounds for a Variational Relaxation of the Image Partitioning Problem

Jan Lellmann · Frank Lenzen · Christoph Schnörr

Published online: 9 November 2012  
© Springer Science+Business Media New York 2012

**Abstract** We consider a variational convex relaxation of a class of optimal partitioning and multiclass labeling problems, which has recently proven quite successful and can be seen as a continuous analogue of Linear Programming (LP) relaxation methods for finite-dimensional problems. While for the latter several optimality bounds are known, to our knowledge no such bounds exist in the infinite-dimensional setting. We provide such a bound by analyzing a probabilistic rounding method, showing that it is possible to obtain an integral solution of the original partitioning problem from a solution of the relaxed problem with an *a priori* upper bound on the objective. The approach has a natural interpretation as an approximate, multiclass variant of the celebrated coarea formula.

**Keywords** Convex relaxation · Multiclass labeling · Approximation bound · Combinatorial optimization · Total variation · Linear programming relaxation

---

J. Lellmann (✉) · F. Lenzen · C. Schnörr  
Image and Pattern Analysis Group & HCI, Dept. of Mathematics  
and Computer Science, University of Heidelberg, Heidelberg,  
Germany  
e-mail: [j.lellmann@damtp.cam.ac.uk](mailto:j.lellmann@damtp.cam.ac.uk)

F. Lenzen  
e-mail: [lenzen@iwr.uni-heidelberg.de](mailto:lenzen@iwr.uni-heidelberg.de)

C. Schnörr  
e-mail: [schnoerr@math.uni-heidelberg.de](mailto:schnoerr@math.uni-heidelberg.de)

*Present address:*  
J. Lellmann  
Dept. of Applied Mathematics and Theoretical Physics,  
University of Cambridge, Centre for Mathematical Sciences,  
Wilberforce Road, Cambridge CB3 0WA, UK

## 1 Introduction and Background

### 1.1 Convex Relaxations of Partitioning Problems

In this work, we will be concerned with a class of *variational* problems used in image processing and analysis for formulating multiclass image partitioning problems, which are of the form

$$\inf_{u \in \mathcal{C}_{\mathcal{E}}} f(u) := \int_{\Omega} \langle u(x), s(x) \rangle dx + \int_{\Omega} d\Psi(Du), \quad (1)$$

$$\mathcal{C}_{\mathcal{E}} := \text{BV}(\Omega, \mathcal{E}) \quad (2)$$

$$= \{u \in \text{BV}(\Omega)^l \mid u(x) \in \mathcal{E} \text{ for a.e. } x \in \Omega\}, \quad (3)$$

$$\mathcal{E} := \{e^1, \dots, e^l\}. \quad (4)$$

The *labeling function*  $u : \Omega \rightarrow \mathbb{R}^l$  assigns to each point in the image domain  $\Omega \subset \mathbb{R}^d$  a label  $i \in \mathcal{I} := \{1, \dots, l\}$ , which is represented by one of the  $l$ -dimensional unit vectors  $e^1, \dots, e^l$ . Since the labeling function is piecewise constant and therefore cannot be assumed to be differentiable, the problem is formulated as a *free discontinuity problem* in the space  $\text{BV}(\Omega, \mathcal{E})$  of functions of bounded variation; see [2] for an overview. We generally assume  $\Omega$  to be a bounded Lipschitz domain.

The objective function  $f$  consists of a data term and a regularizer. The data term is given in terms of the nonnegative  $L^1$  function  $s(x) = (s_1(x), \dots, s_l(x)) \in \mathbb{R}^l$ , and assigns to the choice  $u(x) = e^i$  the “penalty”  $s_i(x)$ , in the sense that

$$\int_{\Omega} \langle u(x), s(x) \rangle dx = \sum_{i=1}^l \int_{\Omega_i} s_i(x) dx, \quad (5)$$

where  $\Omega_i := u^{-1}(\{e^i\}) = \{x \in \Omega \mid u(x) = e^i\}$  is the *class region* for label  $i$ , i.e., the set of points that are assigned

the  $i$ -th label. The data term generally depends on the input data—such as color values of a recorded image, depth measurements, or other features—and promotes a good fit of the minimizer to the input data. While it is purely local, there are no further restrictions such as continuity, convexity etc., therefore it covers many interesting applications such as segmentation, stitching, inpainting, multi-view 3D reconstruction and optical flow [23].

### 1.2 Convex Regularizers

The *regularizer* is defined by the positively homogeneous, continuous and convex function  $\Psi : \mathbb{R}^{d \times l} \rightarrow \mathbb{R}_{\geq 0}$  acting on the distributional derivative  $Du$  of  $u$ , and incorporates additional prior knowledge about the “typical” appearance of the desired output. For piecewise constant  $u$ , it can be seen that the definition in (1) amounts to a weighted penalization of the *discontinuities* of  $u$ :

$$\int_{\Omega} d\Psi(Du) = \int_{J_u} \Psi(v_u(x)(u^+(x) - u^-(x))^{\top}) d\mathcal{H}^{d-1}(x), \tag{6}$$

where  $J_u$  is the jump set of  $u$ , i.e., the set of points where  $u$  has well-defined right-hand and left-hand limits  $u^+$  and  $u^-$  and (in an infinitesimal sense) jumps between the values  $u^+(x), u^-(x) \in \mathbb{R}^l$  across a hyperplane with normal  $v_u(x) \in \mathbb{R}^d, \|v_u(x)\|_2 = 1$ . We refer to [2] for the precise definitions.

A particular case is to set  $\Psi = (1/\sqrt{2})\|\cdot\|_2$ , i.e., the scaled Frobenius norm. In this case  $J(u)$  is just the scaled total variation of  $u$ , and, since  $u^+(x)$  and  $u^-(x)$  assume values in  $\mathcal{E}$  and cannot be equal on the jump set  $J_u$ , it holds that

$$J(u) = \frac{1}{\sqrt{2}} \int_{J_u} \|u^+(x) - u^-(x)\|_2 d\mathcal{H}^{d-1}(x), \tag{7}$$

$$= \mathcal{H}^{d-1}(J_u). \tag{8}$$

Therefore, for  $\Psi = (1/\sqrt{2})\|\cdot\|_2$  the regularizer just amounts to penalizing the total length of the interfaces between class regions as measured by the  $(d - 1)$ -dimensional Hausdorff measure  $\mathcal{H}^{d-1}$ , which is known as *uniform metric* or *Potts regularization*.

A general regularizer was proposed in [19], based on [5]: given a metric *distance*  $d : \{1, \dots, l\}^2 \rightarrow \mathbb{R}_{\geq 0}$ , (not to be confused with the ambient space dimension), define

$$\Psi_d(z) := \sup_{v \in \mathcal{D}_{\text{loc}}^d} \langle z, v \rangle, \quad z = (z^1, \dots, z^l) \in \mathbb{R}^{d \times l}, \tag{9}$$

$$\mathcal{D}_{\text{loc}}^d := \left\{ (v^1, \dots, v^l) \in \mathbb{R}^{d \times l} \mid \dots \|v^i - v^j\|_2 \leq d(i, j) \right\}$$

$$\forall i, j \in \{1, \dots, l\}, \dots, \left. \sum_{k=1}^l v^k = 0 \right\}. \tag{10}$$

It was then shown that

$$\Psi_d(v(e^j - e^i)^{\top}) = d(i, j), \tag{11}$$

therefore in view of (6) the corresponding regularizer is *non-uniform*: the boundary between the class regions  $\Omega_i$  and  $\Omega_j$  is penalized by its length, multiplied by the weight  $d(i, j)$  depending on the labels of both regions.

However, even for the comparably simple regularizer (7), the model (1) is a (spatially continuous) *combinatorial* problem due to the integral nature of the constraint set  $\mathcal{C}_{\mathcal{E}}$ , therefore optimization is nontrivial. In the context of multiclass image partitioning, a first approach can be found in [20], where the problem was posed in a level set-formulation in terms of a labeling function  $\phi : \Omega \rightarrow \{1, \dots, l\}$ , which is subsequently relaxed to  $\mathbb{R}$ . Then  $\phi$  is replaced by polynomials in  $\phi$ , which coincide with the indicator functions  $u_i$  for the case where  $\phi$  assumes integral values. However, the numerical approach involves several nonlinearities and requires to solve a sequence of nontrivial subproblems.

The representation (1) suggests a more straightforward convex approach: replace  $\mathcal{E}$  by its convex hull, which is the unit simplex in  $l$  dimensions,

$$\Delta_l := \text{conv}\{e^1, \dots, e^l\} = \left\{ a \in \mathbb{R}^l \mid a \geq 0, \sum_{i=1}^l a_i = 1 \right\}, \tag{12}$$

and solve the *relaxed* problem

$$\inf_{u \in \mathcal{C}} f(u), \tag{13}$$

$$\mathcal{C} := \text{BV}(\Omega, \Delta_l) \tag{14}$$

$$= \{u \in \text{BV}(\Omega)^l \mid u(x) \in \Delta_l \text{ for a.e. } x \in \Omega\}. \tag{15}$$

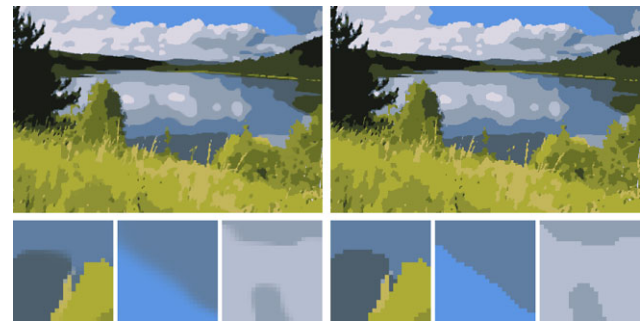
Sparked by a series of papers [5, 17, 30], recently there has been much interest in problems of this form, since they—although generally nonsmooth—are convex and therefore can be solved to global optimality, e.g., using primal-dual techniques. The approach has proven useful in a wide range of applications [10, 11, 14, 29].

### 1.3 Finite-Dimensional vs. Continuous Approaches

Many of these applications have been tackled before in a finite-dimensional setting, where they can be formulated as combinatorial problems on a grid graph, and solved using combinatorial optimization methods such as  $\alpha$ -expansion and related integer linear programming (ILP) methods



**Fig. 1** Segmentation of an image into 12 classes using a combinatorial method. *Top left:* Input image, *Top right:* Result obtained by solving a combinatorial discretized problem with 4-neighborhood. *The bottom row* shows detailed views of the marked parts of the image. The minimizer of the combinatorial problem exhibits blocky artifacts caused by the choice of discretization



**Fig. 2** Segmentation obtained by solving a finite-differences discretization of the relaxed spatially continuous problem. *Left:* Non-integral solution obtained as a minimizer of the discretized relaxed problem. *Right:* Integral labeling obtained by rounding the fractional labels in the solution of the relaxed problem to the nearest integral label. The rounded result is almost free of geometric artifacts

[4, 15]. These methods have been shown to yield an integral labeling  $u' \in \mathcal{C}_\mathcal{E}$  with the *a priori* bound

$$f(u') \leq 2 \frac{\max_{i \neq j} d(i, j)}{\min_{i \neq j} d(i, j)} f(u_\mathcal{E}^*), \tag{16}$$

where  $u_\mathcal{E}^*$  is the (unknown) solution of the integral problem (1). They therefore permit to compute a suboptimal solution to the—originally NP-hard [4]—combinatorial problem with an *upper bound* on the objective. No such bound is yet available for methods based on the spatially continuous problem (13).

Despite these strong theoretical and practical results available for the finite-dimensional combinatorial energies, the function-based, infinite-dimensional formulation (1) has several unique advantages:

- The energy (1) is truly isotropic, in the sense that for a proper choice of  $\Psi$  it is invariant under rotation of the coordinate system. Pursuing finite-dimensional “discretize-first” approaches generally introduces artifacts due to the inherent anisotropy, which can only be avoided by increasing the neighborhood size, thereby reducing sparsity and severely slowing down the graph cut-based methods.

In contrast, properly discretizing the *relaxed* problem (13) and solving it as a *convex* problem with subsequent thresholding yields much better results without compromising the sparse structure (Figs. 1 and 2, [13]). This can be attributed to the fact that solving the discretized problem as a *combinatorial* problem in effect discards much of the information about the problem structure that is contained in the nonlinear terms of the discretized objective.

- Present combinatorial optimization methods [4, 15] are inherently sequential and difficult to parallelize. On the other hand, parallelizing primal-dual methods for solving the relaxed problem (13) is straight-forward, and GPU implementations have been shown to outperform state-of-the-art graph cut methods [30].

- Analyzing the problem in a fully functional-analytic setting gives valuable insight into the problem structure, and is of theoretical interest in itself.

### 1.4 Optimality Bounds

However, one possible drawback of the spatially continuous approach is that the solution of the relaxed problem (13) may assume *fractional* values, i.e., values in  $\Delta_l \setminus \mathcal{E}$ . Therefore, in applications that require a true partition of  $\Omega$ , some rounding process is needed in order to generate an integral labeling  $\bar{u}^*$ . This may increase the objective, and lead to a suboptimal solution of the original problem (1).

The regularizer  $\Psi_d$  as defined in (9) is *tight* in the sense that it majorizes all other regularizers that can be written in integral form and satisfy (11). Therefore it is in a sense “optimal”, since it introduces as few fractional solutions as possible. In practice, this forces solutions of the relaxed problem to assume integral values in most points, and rounding is only required in a few small regions.

However, the rounding step may still increase the objective and generate suboptimal integral solutions. Therefore the question arises whether the approach allows to recover “good” integral solutions of the original problem (1).

In the following, we are concerned with the question whether it is possible to obtain, using the convex relaxation (13), *integral* solutions with an upper bound on the objective. We focus on inequalities of the form

$$f(\bar{u}^*) \leq C f(u_\mathcal{E}^*) \tag{17}$$

for some constant  $C \geq 1$ , which provide an *upper bound on the objective* of the *rounded* integral solution  $\bar{u}^*$  with respect to the objective of the (unknown) *optimal* integral solution  $u_\mathcal{E}^*$  of (1). Note that if the relaxation is not exact, it is only possible to show (17) for some  $C$  strictly larger than one.

The reverse inequality

$$f(u_{\mathcal{E}}^*) \leq f(\bar{u}^*) \tag{18}$$

always holds since  $\bar{u}^* \in \mathcal{C}_{\mathcal{E}}$  and  $u_{\mathcal{E}}^*$  is an optimal integral solution. An alternative interpretation of (17) is

$$\frac{f(\bar{u}^*) - f(u_{\mathcal{E}}^*)}{f(u_{\mathcal{E}}^*)} \leq C - 1, \tag{19}$$

which provides a bound on the relative gap to the optimal objective of the combinatorial problem.

For many convex problems one can find a *dual* representation of the problem in terms of a *dual objective*  $f_D$  and a *dual feasible set*  $\mathcal{D}$  such that

$$\min_{u \in \mathcal{C}} f(u) = \max_{v \in \mathcal{D}} f_D(v), \tag{20}$$

see [25] for the general case and [18, 19] for results on the specific problem (13).

If such a representation exists,  $C$  can be obtained *a posteriori* by actually computing (or approximating)  $\bar{u}^*$  and a dual feasible point: Assume that a feasible primal-dual pair  $(u, v) \in \mathcal{C} \times \mathcal{D}$  is known, where  $u$  approximates  $u^*$ , and assume that some integral feasible  $\bar{u} \in \mathcal{C}_{\mathcal{E}}$  has been obtained from  $u$  by a rounding process. Then the pair  $(\bar{u}, v)$  is feasible as well since  $\mathcal{C}_{\mathcal{E}} \subset \mathcal{C}$ , and we obtain an *a posteriori* optimality bound with respect to the optimal *integral* solution  $u_{\mathcal{E}}^*$ :

$$\begin{aligned} \frac{f(\bar{u}) - f_D(u_{\mathcal{E}}^*)}{f_D(u_{\mathcal{E}}^*)} &\leq \frac{f(\bar{u}) - f_D(u_{\mathcal{E}}^*)}{f_D(v)} \\ &\leq \frac{f(\bar{u}) - f_D(v)}{f_D(v)} =: \delta, \end{aligned} \tag{21}$$

which amounts to setting  $C' := \delta + 1$  in (19). However, this requires that the primal and dual objectives  $f$  and  $f_D$  can be accurately evaluated, and requires to compute a minimizer of the problem for the specific input data, which is generally difficult, especially in the infinite-dimensional formulation.

In contrast, true *a priori* bounds do not require knowledge of a solution and apply uniformly to all problems of a class, irrespective of the particular input. When considering rounding methods, one generally has to discriminate between

- *deterministic* vs. *probabilistic* methods, and
- *spatially discrete* (finite-dimensional) vs. *spatially continuous* (infinite-dimensional) methods.

To our knowledge, most *a priori* approximation results hold only in the finite-dimensional setting, and are usually proven using graph-based pairwise formulations, see [28] for an overview. In contrast, we assume an “optimize first” perspective due to the reasons outlined in the introduction. Unfortunately, the proofs for the finite-dimensional results

often rely on pointwise arguments that cannot directly be transferred to the continuous setting. Deriving similar results for continuous problems therefore requires considerable additional work.

### 1.5 Contribution and Main Results

In this work we prove that using the regularizer (9), the *a priori* bound (16) can be carried over to the spatially continuous setting. Preliminary versions of these results with excerpts of the proofs have been announced as conference proceedings [18]. We extend these results to provide the exact bound (16), and supply the full proofs.

As the main result, we show that it is possible to construct a rounding method parametrized by  $\gamma \in \Gamma$ , where  $\Gamma$  is an appropriate parameter space:

$$R : \mathcal{C} \times \Gamma \rightarrow \mathcal{C}_{\mathcal{E}}, \tag{22}$$

$$u \in \mathcal{C} \mapsto \bar{u}_{\gamma} := R_{\gamma}(u) \in \mathcal{C}_{\mathcal{E}}, \tag{23}$$

such that for a suitable probability distribution on  $\Gamma$ , the following theorem holds for the expectation  $\mathbb{E}f(\bar{u}) := \mathbb{E}_{\gamma} f(\bar{u}_{\gamma})$ :

**Theorem 1** *Let  $u \in \mathcal{C}$ ,  $s \in L^1(\Omega)^l$ ,  $s \geq 0$ , and let  $\Psi : \mathbb{R}^{d \times l} \rightarrow \mathbb{R}_{\geq 0}$  be positively homogeneous, convex and continuous. Assume there exists a lower bound  $\lambda_l > 0$  such that, for  $z = (z^1, \dots, z^l)$ ,*

$$\Psi(z) \geq \lambda_l \frac{1}{2} \sum_{i=1}^l \|z^i\|_2 \quad \forall z \in \mathbb{R}^{d \times l}, \quad \sum_{i=1}^l z^i = 0. \tag{24}$$

*Moreover, assume there exists an upper bound  $\lambda_u < \infty$  such that, for every  $v \in \mathbb{R}^d$  satisfying  $\|v\|_2 = 1$ ,*

$$\Psi(v(e^i - e^j)^{\top}) \leq \lambda_u \quad \forall i, j \in \{1, \dots, l\}. \tag{25}$$

*Then Algorithm 1 generates an integral labeling  $\bar{u} \in \mathcal{C}_{\mathcal{E}}$  almost surely, and*

$$\mathbb{E}f(\bar{u}) \leq 2 \frac{\lambda_u}{\lambda_l} f(u). \tag{26}$$

---

#### Algorithm 1 Continuous Probabilistic Rounding

---

- 1:  $u^0 \leftarrow u, U^0 \leftarrow \Omega, c^0 \leftarrow (1, \dots, 1) \in \mathbb{R}^l$ .
  - 2: **for**  $k = 1, 2, \dots$  **do**
  - 3: Randomly choose  $\gamma^k = (i^k, \alpha^k) \in \mathcal{I} \times [0, 1]$  uniformly.
  - 4:  $M^k \leftarrow U^{k-1} \cap \{x \in \Omega \mid u_{i^k}^{k-1}(x) > \alpha^k\}$ .
  - 5:  $u^k \leftarrow e^{i^k} 1_{M^k} + u^{k-1} 1_{\Omega \setminus M^k}$ .
  - 6:  $U^k \leftarrow U^{k-1} \setminus M^k$ .
  - 7:  $c_j^k \leftarrow \begin{cases} \min\{c_j^{k-1}, \alpha^k\}, & j = i^k, \\ c_j^{k-1}, & \text{otherwise.} \end{cases}$
  - 8: **end for**
-

We refer to Sect. 3.1 for a description of the individual steps of the algorithm. Note that always  $\lambda_u \geq \lambda_l$ , since (25) and (24) imply

$$\lambda_u \geq \Psi(v(e^i - e^j)^\top) \geq \frac{\lambda_l}{2} (\|v\|_2 + \|v\|_2) = \lambda_l \tag{27}$$

for every  $v$  with  $\|v\|_2 = 1$ .

The proof of Theorem 1 (Sect. 4) is based on the work of Kleinberg and Tardos [12], which is set in an LP relaxation framework. However their results are restricted in that they assume a graph-based representation and extensively rely on the finite dimensionality. In contrast, our results hold in the continuous setting without assuming a particular problem discretization.

Theorem 1 guarantees that—in a probabilistic sense—the rounding process may only increase the energy in a controlled way, with an upper bound depending on  $\Psi$ . An immediate consequence is

**Corollary 1** *Under the conditions of Theorem 1, if  $u^*$  minimizes  $f$  over  $\mathcal{C}$ ,  $u_{\mathcal{E}}^*$  minimizes  $f$  over  $\mathcal{C}_{\mathcal{E}}$ , and  $\bar{u}^*$  denotes the output of Algorithm 1 applied to  $u^*$ , then*

$$\mathbb{E}f(\bar{u}^*) \leq 2 \frac{\lambda_u}{\lambda_l} f(u_{\mathcal{E}}^*). \tag{28}$$

Therefore the proposed approach allows to recover, from the solution  $u^*$  of the convex *relaxed* problem (13), an approximate *integral* solution  $\bar{u}^*$  of the nonconvex *original* problem (1) with an upper bound on the objective.

In particular, for the tight relaxation of the regularizer as in (9), we obtain

$$\mathbb{E}f(\bar{u}^*) \leq 2 \frac{\lambda_u}{\lambda_l} = 2 \frac{\max_{i \neq j} d(i, j)}{\min_{i \neq j} d(i, j)} \tag{29}$$

(cf. Proposition 13), which is exactly the same bound as has been achieved for the combinatorial  $\alpha$ -expansion method (16).

To our knowledge, this is the first bound available for the fully spatially convex relaxed problem (13). Related is the work of Olsson et al. [21, 22], where the authors consider an infinite-dimensional analogue to the  $\alpha$ -expansion method known as continuous binary fusion [27], and claim that a bound similar to (16) holds for the corresponding fixed points when using the separable regularizer

$$\Psi_A(z) := \sum_{j=1}^l \|Az^j\|_2, \quad z \in \mathbb{R}^{d \times l}, \tag{30}$$

for some  $A \in \mathbb{R}^{d \times d}$ , which implements an anisotropic variant of the uniform metric. However, a rigorous proof in the BV framework was not given.

In [3], the authors propose to solve the problem (1) by considering the dual problem to (13) consisting of  $l$  coupled maximum-flow problems, which are solved using a log-sum-exp smoothing technique and gradient descent. In case the dual solution allows to unambiguously recover an integral primal solution, the latter is necessarily the unique minimizer of  $f$ , and therefore a global *integral* minimizer of the combinatorial problem (1). This provides an *a posteriori* bound, which applies if a dual solution can be computed. While useful in practice as a certificate for global optimality, in the spatially continuous setting it requires explicit knowledge of a dual solution, which is rarely available since it depends on the regularizer  $\Psi$  as well as the input data  $s$ .

In comparison, the *a priori* bound (28) holds uniformly over all problem instances, does not require knowledge of any primal or dual solutions and covers also non-uniform regularizers.

## 2 A Probabilistic View of the Coarea Formula

### 2.1 The Two-Class Case

As a motivation for the following sections, we first provide a probabilistic interpretation of a tool often used in geometric measure theory, the *coarea formula* (cf. [2]). Given a scalar function  $u' \in \text{BV}(\Omega, [0, 1])$ , the coarea formula states that its total variation can be computed by summing the boundary lengths of its super-levelsets:

$$\text{TV}(u') = \int_0^1 \text{TV}(1_{\{u' > \alpha\}}) d\alpha. \tag{31}$$

Here  $1_A$  denotes the characteristic function of a set  $A$ , i.e.,  $1_A(x) = 1$  iff  $x \in A$  and  $1_A(x) = 0$  otherwise. The coarea formula provides a connection between problem (1) and the relaxation (13) in the two-class case, where  $\mathcal{E} = \{e^1, e^2\}$ , and  $u \in \mathcal{C}_{\mathcal{E}}$  implies  $u_1 = 1 - u_2$ : as noted in [16],

$$\text{TV}(u) = \|e^1 - e^2\|_2 \text{TV}(u_1) = \sqrt{2} \text{TV}(u_1), \tag{32}$$

therefore the coarea formula (31) can be rewritten as

$$\text{TV}(u) = \sqrt{2} \int_0^1 \text{TV}(1_{\{u_1 > \alpha\}}) d\alpha \tag{33}$$

$$= \int_0^1 \text{TV}(e^1 1_{\{u_1 > \alpha\}} + e^2 1_{\{u_1 \leq \alpha\}}) d\alpha \tag{34}$$

$$= \int_0^1 \text{TV}(\bar{u}_{\alpha}) d\alpha, \quad \text{where} \tag{35}$$

$$\bar{u}_{\alpha} := e^1 1_{\{u_1 > \alpha\}} + e^2 1_{\{u_1 \leq \alpha\}}. \tag{36}$$

Consequently, the total variation of  $u$  can be expressed as the *mean* over the total variations of a set of *integral* labelings



$\{\bar{u}_\alpha \in \mathcal{C}_\mathcal{E} \mid \alpha \in [0, 1]\}$ , obtained by *rounding*  $u$  at different thresholds  $\alpha$ . We now adopt a probabilistic view of (36). We regard the mapping

$$R : (u, \alpha) \in \mathcal{C} \times [0, 1] \mapsto \bar{u}_\alpha \in \mathcal{C}_\mathcal{E} \quad (\text{a.e. } \alpha \in [0, 1]) \quad (37)$$

as a parametrized *deterministic* rounding algorithm that depends on  $u$  and on an additional parameter  $\alpha$ . From this we obtain a *probabilistic* (randomized) rounding algorithm by assuming  $\alpha$  to be a uniformly distributed random variable. With these definitions the coarea formula (36) can be written as

$$\text{TV}(u) = \mathbb{E}_\alpha \text{TV}(\bar{u}_\alpha). \quad (38)$$

This states that applying the probabilistic rounding to (arbitrary, but fixed)  $u$  does—in a probabilistic sense, i.e., in the mean—not change the objective. It can be shown that this property extends to the full functional  $f$  in (13): in the two-class case, the “coarea-like” property

$$f(u) = \mathbb{E}_\alpha f(\bar{u}_\alpha) \quad (39)$$

holds. Functions with property (39) are also known as *levelable functions* [8, 9] or *discrete total variations* [6] and have been studied in [26]. A well-known implication is that if  $u = u^*$ , i.e.,  $u$  minimizes the relaxed problem (13), then in the two-class case almost every  $\bar{u}^* = \bar{u}_\alpha^*$  is an integral minimizer of the original problem (1), i.e., the optimality bound (17) holds with  $C = 1$  [7].

### 2.2 The Multi-Class Case and Generalized Coarea Formulas

Generalizing these observations to more than two labels hinges on a property similar to (39) that holds for *vector-valued*  $u$ . In a general setting, the question is whether there exist

- a probability space  $(\Gamma, \mu)$ , and
- a *parametrized rounding method*, i.e., for  $\mu$ -almost every  $\gamma \in \Gamma$ :

$$R : \mathcal{C} \times \Gamma \rightarrow \mathcal{C}_\mathcal{E}, \quad (40)$$

$$u \in \mathcal{C} \mapsto \bar{u}_\gamma := R_\gamma(u) \in \mathcal{C}_\mathcal{E} \quad (41)$$

satisfying  $R_\gamma(u') = u'$  for all  $u' \in \mathcal{C}_\mathcal{E}$ ,

such that a “multiclass coarea-like property” (or *generalized coarea formula*)

$$f(u) = \int_\Gamma f(\bar{u}_\gamma) d\mu(\gamma) \quad (42)$$

holds. The equivalent probabilistic interpretation is

$$f(u) = \mathbb{E}_\gamma f(\bar{u}_\gamma). \quad (43)$$

For  $l = 2$  and  $\Psi(x) = \|\cdot\|_2$ , (38) shows that (43) holds with  $\gamma = \alpha$ ,  $\Gamma = [0, 1]$ ,  $\mu = \mathcal{L}^1$ , and  $R : \mathcal{C} \times \Gamma \rightarrow \mathcal{C}_\mathcal{E}$  as defined in (37). However, property (38) is intrinsically restricted to the two-class case and the TV regularizer.

In the multiclass case, the difficulty lies in providing a suitable combination of a probability space  $(\Gamma, \mu)$  and a parametrized rounding step  $(u, \gamma) \mapsto \bar{u}_\gamma$ . Unfortunately, obtaining a relation such as (38) for the full functional (1) is unlikely, as it would mean that solutions to the (after discretization) NP-hard problem (1) could be obtained by solving the convex relaxation (13) and subsequent rounding, which can be achieved in polynomial time.

Therefore we restrict ourselves to an *approximate* variant of the generalized coarea formula:

$$Cf(u) \geq \int_\Gamma f(\bar{u}_\gamma) d\mu(\gamma) = \mathbb{E}_\gamma f(\bar{u}_\gamma). \quad (44)$$

While (44) is not sufficient to provide a bound on  $f(\bar{u}_\gamma)$  for *particular*  $\gamma$ , it permits a probabilistic bound: for any minimizer  $u^*$  of the relaxed problem (13), Eq. (44) implies

$$\mathbb{E}_\gamma f(\bar{u}_\gamma^*) \leq Cf(u^*) \leq Cf(u_\mathcal{E}^*), \quad (45)$$

and thus the ratio between the objective of the *rounded relaxed solution* and the *optimal integral solution* is bounded—in a probabilistic sense—by the constant  $C$ .

In the following sections we construct a suitable parametrized rounding method and probability space in order to obtain an approximate generalized coarea formula of the form (44).

## 3 Probabilistic Rounding for Multiclass Image Partitions

### 3.1 Approach

We consider the probabilistic rounding approach based on [12] as defined in Algorithm 1.

The algorithm proceeds in a number of phases. At each iteration, a label and a threshold

$$\gamma^k := (i^k, \alpha^k) \in \Gamma' := \mathcal{I} \times [0, 1]$$

are randomly chosen (step 3), and label  $i^k$  is assigned to all yet unassigned points  $x$  where  $u_{i^k}^{k-1}(x) > \alpha^k$  holds (step 5). In contrast to the two-class case considered above, the randomness is provided by a *sequence*  $(\gamma^k)$  of uniformly distributed random variables, i.e.,  $\Gamma = (\Gamma')^\mathbb{N}$ .

After iteration  $k$ , all points in the set  $U^k \subseteq \Omega$  are still *unassigned*, while all points in  $\Omega \setminus U^k$  have been assigned an (integral) label in iteration  $k$  or in a previous iteration. It-

eration  $k + 1$  potentially modifies points only in the set  $U^k$ . The variable  $c_j^k$  stores the lowest threshold  $\alpha$  chosen for label  $j$  up to and including iteration  $k$ , and is only required for the proofs.

For any  $u \in L^1(\Omega, \Delta_l)$  and fixed  $\gamma$ , the sequences  $(u^k)$ ,  $(M^k)$  and  $(U^k)$  are unique up to  $\mathcal{L}^d$ -negligible sets, and therefore the sequence  $(u^k)$  is well-defined when viewed as elements of  $L^1$ .

In an actual implementation, the algorithm could be terminated as soon as all points in  $\Omega$  have been assigned a label, i.e.,  $|U^k| := \mathcal{L}^d(U^k) = 0$ . However, in our framework used for analysis the algorithm never terminates explicitly. Instead, for fixed input  $u$  we regard the algorithm as a mapping between sequences of parameters (or instances of random variables)  $\gamma = (\gamma^k) \in \Gamma$  and sequences of states  $(u_\gamma^k)$ ,  $(U_\gamma^k)$  and  $(c_\gamma^k)$ . We drop the subscript  $\gamma$  if it does not create ambiguities. The elements of the sequence  $(\gamma^{(k)})$  are independently uniformly distributed, therefore choosing  $\gamma$  can be seen as sampling from the product space.

In order to define the parametrized rounding step  $(u, \gamma) \mapsto \bar{u}_\gamma$ , we observe that once  $|U_\gamma^{k'}| = 0$  occurs for some  $k' \in \mathbb{N}$ , the sequence  $(u_\gamma^k)$  becomes stationary at  $u_\gamma^{k'}$ . In this case the output of the algorithm is defined as  $\bar{u}_\gamma := u_\gamma^{k'}$ :

**Definition 1** Let  $u \in \text{BV}(\Omega)^l$  and  $f : \text{BV}(\Omega)^l \rightarrow \mathbb{R}$ . For arbitrary, fixed  $\gamma \in \Gamma$ , let  $(u_\gamma^k)$  be the sequence generated by Algorithm 1 and define  $\bar{u}_\gamma : \Omega \rightarrow \bar{\mathbb{R}}^l$  as

$$\bar{u}_\gamma(x)_j := \begin{cases} u_\gamma^{k'}(x)_j, & \exists k' \in \mathbb{N} : |U_\gamma^{k'}| = 0, \\ +\infty, & \text{otherwise.} \end{cases} \tag{46}$$

We extend  $f$  to all functions  $u' : \Omega \rightarrow \bar{\mathbb{R}}^l$  by setting  $f(u') := +\infty$  if  $u' \notin \text{BV}(\Omega, \Delta_l)$  and consider the induced mapping  $f(\bar{u}_{(\cdot)}) : \Gamma \rightarrow \mathbb{R} \cup \{+\infty\}$ ,  $\gamma \in \Gamma \mapsto f(\bar{u}_\gamma)$ , i.e.,

$$f(\bar{u}_\gamma) = \begin{cases} f(u_\gamma^{k'}), & \bar{u}_\gamma \in \text{BV}(\Omega, \Delta_l), \\ +\infty, & \text{otherwise.} \end{cases} \tag{47}$$

We denote by  $f(\bar{u})$  the random variable induced by assuming  $\gamma$  to be uniformly distributed on  $\Gamma$ , and by  $\mu$  the uniform probability measure on  $\Gamma$ .

In the following we often use  $\mathbb{P} = \mu$  where it does not create ambiguities. Measures are generally understood to be extended to the completion of the underlying  $\sigma$ -algebra, i.e., all subsets of zero sets are measurable.

As indicated above,  $f(\bar{u}_\gamma)$  is well-defined—indeed, if  $|U_\gamma^{k'}| = 0$  for some  $(\gamma, k')$  then  $u_\gamma^{k'} = u_\gamma^{k''}$  for all  $k'' \geq k'$ . Instead of focusing on local properties of the random sequence  $(u_\gamma^k)$  as in the proofs for the finite-dimensional case, we derive our results directly for the sequence  $(f(u_\gamma^k))$ . In

particular, we show that the expectation of  $f(\bar{u})$  over all sequences  $\gamma$  can be bounded according to

$$\mathbb{E} f(\bar{u}) = \mathbb{E}_\gamma f(\bar{u}_\gamma) \leq C f(\bar{u}) \tag{48}$$

for some  $C \geq 1$ , cf. (44). Consequently, the rounding process may only increase the average objective in a controlled way.

### 3.2 Termination Properties

Theoretically, the algorithm may produce a sequence  $(u_\gamma^k)$  that does not become stationary, or becomes stationary with a solution that is not an element of  $\text{BV}(\Omega)^l$ . In Theorem 2 below we show that this happens only with zero probability, i.e., almost surely Algorithm 1 generates (in a finite number of iterations) an integral labeling function  $\bar{u}_\gamma \in \mathcal{C}_\mathcal{E}$ . The following two propositions are required for the proof. We use the definition  $e := (1, \dots, 1)$ .

**Proposition 1** For the sequence  $(c^k)$  generated by Algorithm 1,

$$\begin{aligned} \mathbb{P}(e^\top c^k < 1) & \geq \sum_{p \in \{0, 1\}^l} (-1)^{e^\top p} \left( \sum_{j=1}^l \frac{1}{l} \left( \left( 1 - \frac{1}{l} \right)^{p_j} \right) \right)^k \end{aligned} \tag{49}$$

holds. In particular,

$$\mathbb{P}(e^\top c^k < 1) \xrightarrow{k \rightarrow \infty} 1. \tag{50}$$

*Proof* Denote by  $n_j^k \in \mathbb{N}_0$  the number of  $k' \in \{1, \dots, k\}$  such that  $i^{k'} = j$ , i.e., the number of times label  $j$  was selected up to and including the  $k$ -th step. Then

$$(n_1^k, \dots, n_l^k) \sim \text{Multinomial} \left( k; \frac{1}{l}, \dots, \frac{1}{l} \right), \tag{51}$$

i.e., the probability of a specific instance is

$$\mathbb{P}((n_1^k, \dots, n_l^k)) = \begin{cases} \frac{k!}{n_1^k! \dots n_l^k!} \left( \frac{1}{l} \right)^k, & \sum_j n_j^k = k, \\ 0, & \text{otherwise.} \end{cases} \tag{52}$$

Therefore,

$$\begin{aligned} \mathbb{P}(e^\top c^k < 1) & = \sum_{n_1^k, \dots, n_l^k} \mathbb{P}(e^\top c^k < 1 \mid (n_1^k, \dots, n_l^k)) \\ & \quad \cdot \mathbb{P}((n_1^k, \dots, n_l^k)) \end{aligned} \tag{53}$$

$$\begin{aligned} & = \sum_{n_1^k + \dots + n_l^k = k} \frac{k!}{n_1^k! \dots n_l^k!} \left( \frac{1}{l} \right)^k \\ & \quad \cdot \mathbb{P}(e^\top c^k < 1 \mid (n_1^k, \dots, n_l^k)). \end{aligned} \tag{54}$$

Since  $c_1^k, \dots, c_l^k < \frac{1}{l}$  is a sufficient condition for  $e^\top c < 1$ , we may bound the probability according to

$$\mathbb{P}(e^\top c < 1) \geq \sum_{n_1^k + \dots + n_l^k = k} \frac{k!}{n_1^k! \dots n_l^k!} \left(\frac{1}{l}\right)^k \cdot \mathbb{P}\left(c_j^k < \frac{1}{l} \forall j \in \mathcal{I} \mid (n_1^k, \dots, n_l^k)\right). \tag{55}$$

We now consider the distributions of the components  $c_j^k$  of  $c^k$  conditioned on the vector  $(n_1^k, \dots, n_l^k)$ . Given  $n_j^k$ , the probability of  $\{c_j^k \geq t\}$  is the probability that in each of the  $n_j^k$  steps where label  $j$  was selected the threshold  $\alpha$  was randomly chosen to be *at least as large as*  $t$ . For  $0 < t < 1$ , we conclude

$$\mathbb{P}(c_j^k < t \mid (n_1^k, \dots, n_l^k)) = \mathbb{P}(c_j^k < t \mid n_j^k) \tag{56}$$

$$= 1 - \mathbb{P}(c_j^k \geq t \mid n_j^k) \tag{57}$$

$$\stackrel{0 < t < 1}{=} 1 - (1 - t)^{n_j^k}. \tag{58}$$

The above formulation also covers the case  $n_j^k = 0$  (note that we assumed  $0 < t < 1$ ). For fixed  $k$  the distributions of the  $c_j^k$  are independent when conditioned on  $(n_1^k, \dots, n_l^k)$ . Therefore we obtain from (55) and (58)

$$\mathbb{P}(e^\top c < 1) \stackrel{(55)}{\geq} \sum_{n_1^k + \dots + n_l^k = k} \frac{k!}{n_1^k! \dots n_l^k!} \left(\frac{1}{l}\right)^k \cdot \prod_{j=1}^l \mathbb{P}\left(c_j^k < \frac{1}{l} \mid (n_1^k, \dots, n_l^k)\right) \tag{59}$$

$$\stackrel{(58)}{=} \sum_{n_1^k + \dots + n_l^k = k} \frac{k!}{n_1^k! \dots n_l^k!} \left(\frac{1}{l}\right)^k \cdot \prod_{j=1}^l \left(1 - \left(1 - \frac{1}{l}\right)^{n_j^k}\right). \tag{60}$$

Expanding the product and swapping the summation order, we derive

$$\mathbb{P}(e^\top c^k < 1) \tag{61}$$

$$\geq \sum_{n_1^k + \dots + n_l^k = k} \frac{k!}{n_1^k! \dots n_l^k!} \left(\frac{1}{l}\right)^k \cdot \sum_{p \in \{0,1\}^l} \prod_{j=1}^l \left(-\left(1 - \frac{1}{l}\right)^{n_j^k}\right)^{p_j} \tag{62}$$

$$= \sum_{p \in \{0,1\}^l} (-1)^{e^\top p} \sum_{n_1^k + \dots + n_l^k = k} \frac{k!}{n_1^k! \dots n_l^k!}$$

$$\cdot \prod_{j=1}^l \left(\frac{1}{l} \left(1 - \frac{1}{l}\right)^{p_j}\right)^{n_j^k}. \tag{63}$$

Using the multinomial summation formula, we conclude

$$\mathbb{P}(e^\top c^k < 1) \geq \sum_{p \in \{0,1\}^l} (-1)^{e^\top p} \underbrace{\left(\sum_{j=1}^l \frac{1}{l} \left(1 - \frac{1}{l}\right)^{p_j}\right)^k}_{=: q_p}, \tag{64}$$

which proves (49). Note that in (64) the  $n_j^k$  do not occur explicitly anymore. To show the second assertion (50), we use the fact that, for any  $p \neq 0$ ,  $q_p$  can be bounded by  $0 < q_p < 1$ . Therefore

$$\mathbb{P}(e^\top c^k < 1) \geq q_0 + \sum_{p \in \{0,1\}^l, p \neq 0} (-1)^{e^\top p} (q_p)^k \tag{65}$$

$$= 1 + \sum_{p \in \{0,1\}^l, p \neq 0} (-1)^{e^\top p} \underbrace{(q_p)^k}_{\xrightarrow{k \rightarrow \infty} 0} \tag{66}$$

$$\xrightarrow{k \rightarrow \infty} 1, \tag{67}$$

which proves (50).  $\square$

We now show that Algorithm 1 generates a sequence in  $\text{BV}(\Omega)^l$  almost surely. The *perimeter* of a set  $A$  is defined as the total variation of its characteristic function  $\text{Per}(A) := \text{TV}(1_A)$  in  $\Omega$ .

**Proposition 2** For the sequences  $(u^k)$ ,  $(U^k)$  generated by Algorithm 1, define

$$A := \bigcap_{k=1}^{\infty} \{\gamma \in \Gamma \mid \text{Per}(U_\gamma^k) < \infty\}. \tag{68}$$

Then

$$\mathbb{P}(A) = 1. \tag{69}$$

If  $\text{Per}(U_\gamma^k) < \infty$  for all  $k$ , then  $u_\gamma^k \in \text{BV}(\Omega)^l$  for all  $k$  as well. Moreover,

$$\mathbb{P}(u^k \in \text{BV}(\Omega)^l \wedge \text{Per}(U^k) < \infty \forall k \in \mathbb{N}) = 1, \tag{70}$$

i.e., the algorithm almost surely generates a sequence of BV functions  $(u^k)$  and a sequence of sets of finite perimeter  $(U^k)$ .

*Proof* We first show that if  $\text{Per}(U^{k'}) < \infty$  for all  $k' \leq k$ , then  $u^k \in \text{BV}(\Omega)^l$  for all  $k' \leq k$  as well. For  $k = 0$ , the assertion holds since  $u^0 = u \in \text{BV}(\Omega)^l$  by assumption. For



$k \geq 1$ ,

$$u^k = e^{i^k} 1_{M^k} + u^{k-1} 1_{\Omega \setminus M^k}. \tag{71}$$

Since  $M^k = U^{k-1} \cap (\Omega \setminus U^k)$ , and  $U^k, U^{k-1}$  are assumed to have finite perimeter,  $M^k$  also has finite perimeter. Applying [2, Theorem 3.84] together with the boundedness of  $u^{k-1}$  and  $u^{k-1} \in \text{BV}(\Omega)^l$  and an induction argument then provide  $u^k \in \text{BV}(\Omega)^l$ .

We now denote

$$I^k := \{\gamma \in \Gamma \mid \text{Per}(U_\gamma^k) = \infty\}, \tag{72}$$

and the event that the first set with non-finite perimeter is encountered at step  $k \in \mathbb{N}_0$  by

$$B^k := I^k \cap (\Gamma \setminus I^{k-1}) \cap \dots \cap (\Gamma \setminus I^0). \tag{73}$$

Note that  $U^0 = \Omega$ , therefore  $\text{Per}(U^0) = \text{TV}(1_{U^0}) = 0 < \infty$  and  $\mathbb{P}(B^0) = 0$ . For  $k \geq 1$ , we use the basic inequality  $\mathbb{P}(E \cap F) \leq \mathbb{P}(E \mid F)$  and obtain

$$\begin{aligned} \mathbb{P}(B^k) &= \mathbb{P}(\text{Per}(U^k) = \infty \wedge \text{Per}(U^{k'}) < \infty \forall k' < k) \\ &\leq \mathbb{P}(\text{Per}(U^k) = \infty \mid \text{Per}(U^{k'}) < \infty \forall k' < k) \\ &= \mathbb{P}(\text{Per}(U^{k-1} \cap \{x \in \Omega \mid u_{i^k}^{k-1}(x) \leq \alpha^k\}) \\ &= \infty \mid \text{Per}(U^{k'}) < \infty \forall k' < k). \end{aligned} \tag{74}$$

By the argument from the beginning of the proof, we know that  $u^{k-1} \in \text{BV}(\Omega)^l$  under the condition on the perimeter  $\text{Per}(U^{k'})$ , therefore from [2, Theorem 3.40] we conclude that  $\text{Per}(\{x \in \Omega \mid u_{i^k}^{k-1}(x) \leq \alpha^k\})$  is finite for  $\mathcal{L}^1$ -a.e.  $\alpha^k$  and all  $i^k$ , i.e., for fixed  $i^k$  the set

$$\{\alpha^k \in [0, 1] \mid \text{Per}(\{x \in \Omega \mid u_{i^k}^{k-1}(x) \leq \alpha^k\}) = \infty\} \tag{75}$$

is contained in an  $\mathcal{L}^1$ -zero set. As the sets of finite perimeter are closed under finite intersection, and since the  $\alpha^k$  are drawn from a uniform distribution, this implies that

$$\mathbb{P}(\text{Per}(U^k) < \infty \mid \text{Per}(U^{k-1}) < \infty) = 1. \tag{76}$$

Together with (74) we arrive at

$$\mathbb{P}(B^k) = 0, \tag{77}$$

which implies the assertion,

$$\mathbb{P}(A) = 1 - \mathbb{P}\left(\bigcup_{k=0}^{\infty} B^k\right) \geq 1 - \sum_{k=0}^{\infty} \mathbb{P}(B^k) = 1. \tag{78}$$

Equation (70) follows immediately.

Measurability of the sets involved follows from a similar recursive argument starting from (75) and using the fact that all sets or their complements are contained in a zero set,

and are therefore measurable with respect to their respective (complete) probability measures.  $\square$

Using these propositions, we now formulate the main result of this section: Algorithm 1 almost surely generates an integral labeling that is of bounded variation.

**Theorem 2** *Let  $u \in \text{BV}(\Omega)^l$  and  $f(\bar{u})$  as in Definition 1. Then*

$$\mathbb{P}(f(\bar{u}) < \infty) = 1. \tag{79}$$

*Proof* The first part is to show that  $(u^k)$  becomes stationary almost surely, i.e.,

$$\mathbb{P}(\exists k \in \mathbb{N} : |U^k| = 0) = 1. \tag{80}$$

Assume there exists  $k$  such that  $e^\top c^k < 1$ , and assume further that  $|U^k| > 0$ , i.e.,  $U^k$  contains a non-negligible subset where  $u_j(x) \leq c_j^k$  for all labels  $j$ . But then  $e^\top u(x) \leq e^\top c^k < 1$  on that set, which is a contradiction to  $u(x) \in \Delta_l$  almost everywhere. Therefore  $U^k$  must be a zero set. From this observation and Proposition 1 we conclude, for all  $k' \in \mathbb{N}$ ,

$$1 \geq \mathbb{P}(\exists k \in \mathbb{N} : |U^k| = 0) \geq \mathbb{P}(e^\top c^{k'} < 1) \xrightarrow{k' \rightarrow \infty} 1, \tag{81}$$

which proves (80).

In order to show that  $f(\bar{u}_\gamma) < \infty$  with probability 1, it remains to show that the result is almost surely in  $\text{BV}(\Omega)^l$ . A sufficient condition is that almost surely all iterates  $u^k$  are elements of  $\text{BV}(\Omega)^l$ , i.e.,

$$\mathbb{P}(u^k \in \text{BV}(\Omega)^l \forall k \in \mathbb{N}) = 1. \tag{82}$$

This is shown by Proposition 2. Then

$$\mathbb{P}(f(\bar{u}) < \infty) \tag{83}$$

$$\geq \mathbb{P}(\{\exists k \in \mathbb{N} : |U^k| = 0\} \wedge \{u^k \in \text{BV}(\Omega)^l \forall k \in \mathbb{N}\})$$

$$= \mathbb{P}(\{u^k \in \text{BV}(\Omega)^l \forall k \in \mathbb{N}\})$$

$$- \mathbb{P}(\{\forall k \in \mathbb{N} : |U^k| > 0\})$$

$$\wedge \{u^k \in \text{BV}(\Omega)^l \forall k \in \mathbb{N}\}) \tag{84}$$

$$\stackrel{(80)}{=} \mathbb{P}(\{u^k \in \text{BV}(\Omega)^l \forall k \in \mathbb{N}\}) - 0 \tag{85}$$

$$\stackrel{(82)}{=} 1. \tag{86}$$

Thus  $\mathbb{P}(f(\bar{u}) < \infty) = 1$ , which proves the assertion.  $\square$

#### 4 Proof of the Main Theorem

In order to show the bound (48) and Theorem 1, we first need several technical propositions regarding the composition of

two BV functions along a set of finite perimeter. We denote by  $(E)^1$  and  $(E)^0$  the measure-theoretic interior and exterior of a set  $E$ , see [2],

$$(E)^t := \left\{ x \in \Omega \mid \lim_{\rho \searrow 0} \frac{|\mathcal{B}_\rho(x) \cap E|}{|\mathcal{B}_\rho(x)|} = t \right\}, \quad t \in [0, 1]. \tag{87}$$

Here  $\mathcal{B}_\rho(x)$  denotes the ball with radius  $\rho$  centered in  $x$ , and  $|A| := \mathcal{L}^d(A)$  the Lebesgue content of a set  $A \subseteq \mathbb{R}^d$ .

**Proposition 3** *Let  $\Psi$  be positively homogeneous and convex, and satisfy the upper-boundedness condition (25). Then*

$$\Psi(v(z^1 - z^2)^\top) \leq \lambda_u \quad \forall z^1, z^2 \in \Delta_l. \tag{88}$$

Moreover, there exists a constant  $C < \infty$  such that

$$\Psi(w) \leq C \|w\|_2 \quad \forall w \in W, \tag{89}$$

$$W := \left\{ w = (w^1 | \dots | w^l) \in \mathbb{R}^{d \times l} \mid \sum_{i=1}^l w^i = 0 \right\}. \tag{90}$$

*Proof* See the Appendix. □

**Proposition 4** *Let  $E, F \subseteq \Omega$  be  $\mathcal{L}^d$ -measurable sets. Then*

$$(E \cap F)^1 = (E)^1 \cap (F)^1. \tag{91}$$

*Proof* See the Appendix. □

In the following proposition we denote by  $u_{\mathcal{F}E}^+$  and  $v_{\mathcal{F}E}^-$  the one-sided approximate limits of  $u$  and  $v$  on the reduced boundary  $\mathcal{F}E$  (traces in the sense of [2, Theorem 3.77]), and by  $\nu_E$  the generalized inner normal of the set  $E$  [2, Definition 3.54].

The measure  $\Psi(Du)$  is defined as (cf. [2, (2.26), Proposition 3.23])

$$\Psi(Du)(A) := \int_A \Psi\left(\frac{Du}{|Du|}\right) d|Du|. \tag{92}$$

**Proposition 5** *Let  $u, v \in \text{BV}(\Omega, \Delta_l)$  and  $E \subseteq \Omega$  such that  $\text{Per}(E) < \infty$ . Define*

$$w := u1_E + v1_{\Omega \setminus E}. \tag{93}$$

Then  $w \in \text{BV}(\Omega, \Delta_l)^1$ , and

$$Dw = Du_{\mathcal{L}(E)^1} + Dv_{\mathcal{L}(E)^0} + \nu_E(u_{\mathcal{F}E}^+ - v_{\mathcal{F}E}^-)^\top \mathcal{H}^{d-1}_{\mathcal{L}(\mathcal{F}E \cap \Omega)}. \tag{94}$$

Moreover, for continuous, convex and positively homogeneous  $\Psi$  satisfying the upper-boundedness condition (25) and any Borel set  $A \subseteq \Omega$ ,

$$\int_A d\Psi(Dw) \leq \int_{A \cap (E)^1} d\Psi(Du)$$

$$+ \int_{A \cap (E)^0} d\Psi(Dv) + \lambda_u \text{Per}(E). \tag{95}$$

*Proof* See the Appendix. □

**Proposition 6** *Let  $u, v \in \text{BV}(\Omega, \Delta_l)$ ,  $E \subseteq \Omega$  such that  $\text{Per}(E) < \infty$ , and*

$$u|_{(E)^1} = v|_{(E)^1} \quad \mathcal{L}^d\text{-a.e.} \tag{96}$$

Then  $(Du)_{\mathcal{L}(E)^1} = (Dv)_{\mathcal{L}(E)^1}$ , and  $\Psi(Du)_{\mathcal{L}(E)^1} = \Psi(Dv)_{\mathcal{L}(E)^1}$ . In particular,

$$\int_{(E)^1} d\Psi(Du) = \int_{(E)^1} d\Psi(Dv). \tag{97}$$

The result also holds when  $(E)^1$  is replaced by  $(E)^0$ . Moreover, the condition (96) is equivalent to

$$u|_E = v|_E \quad \mathcal{L}^d\text{-a.e.} \tag{98}$$

*Proof* See the Appendix. □

**Remark 1** Note that taking the measure-theoretic interior  $(E)^1$  is of central importance. Proposition 6 does not hold when replacing the integral over  $(E)^1$  with the integral over  $E$ , as can be seen from the example of the closed unit ball, i.e.,  $E = \mathcal{B}_1(0)$ ,  $u = 1_E$  and  $v \equiv 1$ .

### 4.1 Proof of Theorem 1

In Sect. 3.2 we have shown that the rounding process induced by Algorithm 1 is well-defined in the sense that it returns an integral solution  $\bar{u}_\gamma \in \text{BV}(\Omega)^l$  almost surely. We now return to proving an upper bound for the expectation of  $f(\bar{u})$  as in the approximate coarea formula (44).

We first establish measurability and show that the expectation of the linear part (data term) of  $f$  is invariant under the rounding process.

**Proposition 7** *Let  $(u_\gamma^k)$  be the sequence generated by Algorithm 1. Then for every  $k \geq 1$  the mappings*

$$g^k : \Gamma \times \Omega \rightarrow \mathbb{R}^l, \quad (\gamma, x) \mapsto u_\gamma^k(x) \tag{99}$$

and

$$h : \Gamma \times \Omega \rightarrow \bar{\mathbb{R}}^l, \quad (\gamma, x) \mapsto \bar{u}_\gamma(x) \tag{100}$$

are  $(\mu \times \mathcal{L}^d)$ -measurable.

*Proof* In Algorithm 1, instead of step 5 we consider the simpler update

$$u^k \leftarrow e^{i^k} 1_{\{u_{i^k}^{k-1} > \alpha^k\}} + u^{k-1} 1_{\{u_{i^k}^{k-1} \leq \alpha^k\}}. \tag{101}$$

This yields exactly the same sequence  $(u^k)$ , since if  $u_{i^k}^{k-1}(x) > \alpha^k$ , then either  $x \in U^{k-1}$ , or  $u_{i^k}^{k-1}(x) = 1$ . In both algorithms, points that are assigned a label  $e^{i^k}$  at some point in the process will never be assigned a *different* label at a later point. This is made explicit in Algorithm 1 by keeping track of the set  $U^k$  of yet unassigned points. In contrast, using the step (101), a point may be contained in several of the sets  $\{u_{i^k}^{k-1} \leq \alpha^k\}$  of points that get assigned label  $i^k$  in step  $k$ , but once assigned its label cannot change during a later iteration.

For the measurability of the  $g^k$  it suffices to show measurability of the mapping

$$(\gamma^1, \dots, \gamma^k, x) \in (\Gamma^k) \times \mathbb{R} \mapsto u_{(\gamma^1, \dots, \gamma^k)}^k(x). \tag{102}$$

From the update (101) we see that  $u_{(\gamma^1, \dots, \gamma^k)}^k$  is a finite sum of functions of the form  $e^{i^k} \cdot 1_{A^1} \cdots 1_{A^l}$  and  $u \cdot 1_{A^1} \cdots 1_{A^l}$ , for some  $l \leq k$ , where each  $A^m, m \leq l$  is either the set  $\{(\gamma^1, \dots, \gamma^k, x) \mid u(x)_{i^m} > \alpha^m\}$  or its complement. Each of these indicator functions is jointly measurable in  $(\gamma, x)$ : every component of  $u$  is again measurable, and for any measurable scalar-valued function  $v$ , the set  $B := \{(\alpha, x) \mid v(x) > \alpha\}$  is the countable union of measurable sets,

$$B = \bigcup_{t \in \mathbb{Q}} (-\infty, t] \times \{v^{-1}((t, +\infty))\}, \tag{103}$$

and therefore  $(\alpha, x) \mapsto 1_B(x)$  is jointly measurable in  $(\alpha, x)$ . Consequently,  $u_{\gamma}^k$  is the finite sum of products of functions that are jointly measurable in  $(\gamma, x)$ , which shows the first assertion.

Regarding the second assertion, Theorem 2 shows that  $h(\gamma, x) = \lim_{k \rightarrow \infty} g^k(\gamma, x)$ , except possibly for a negligible set of  $\gamma$  where the sequence  $(u_{\gamma}^k)$  does not become stationary. Since all  $g^k$  are measurable, their pointwise limit and therefore  $h$  are measurable as well.  $\square$

**Proposition 8** For every  $k \geq 1$  the mappings

$$g^{ik} : \Gamma \rightarrow \mathbb{R}, \quad \gamma \mapsto \int_{\Omega} \langle u_{\gamma}^k, s \rangle dx \tag{104}$$

and

$$h' : \Gamma \rightarrow \mathbb{R}, \quad \gamma \mapsto \int_{\Omega} \langle \bar{u}_{\gamma}, s \rangle dx \tag{105}$$

are  $\mu$ -measurable.

*Proof* The first assertion follows directly from Proposition 7 and  $(\mu \times \mathcal{L}^d)$ -measurability of the map  $(\gamma, x) \mapsto s(x)$ . For each fixed  $\gamma$  the sequence  $(g^{ik}(\gamma))_k$  is bounded since  $s \in L^1(\Omega)$  and  $u$  is essentially bounded. Together with Theorem 2 this implies

$$h'(\gamma) = \lim_{k \rightarrow \infty} g^{ik}(\gamma) \quad \text{for } \mu\text{-a.e. } \gamma \in \Gamma, \tag{106}$$

therefore  $h'$  is measurable as well, as it is the limit of measurable functions.  $\square$

**Proposition 9** The sequence  $(u^k)$  generated by Algorithm 1 satisfies

$$\mathbb{E} \int_{\Omega} \langle u^k, s \rangle dx = \int_{\Omega} \langle u, s \rangle dx \quad \forall k \in \mathbb{N}. \tag{107}$$

*Proof* Proposition 8 shows that the expectation is well-defined. Integrability on  $\Gamma \times \mathbb{R}^d$  again holds because  $u_{\gamma}^k$  is in  $L^1(\Omega, \Delta_I)$  and therefore essentially bounded,  $s \in L^1(\Omega)$ , and  $\Omega$  is bounded, which uniformly bounds the inner integral over all  $\gamma$ .

Assume  $\gamma \in \Gamma$  is arbitrary but fixed, and denote  $\gamma' := (\gamma^1, \dots, \gamma^{k-1})$  and  $u^{\gamma'} := u_{\gamma'}^{k-1}$ . We apply induction on  $k$ : For  $k \geq 1$ ,

$$\mathbb{E}_{\gamma} \int_{\Omega} \langle u_{\gamma}^k, s \rangle dx \tag{108}$$

$$= \mathbb{E}_{\gamma'} \frac{1}{l} \sum_{i=1}^l \int_0^1 \int_{\Omega} \sum_{j=1}^l s_j \cdot (e^i 1_{\{u_i^{\gamma'} > \alpha\}} + q u^{\gamma'} 1_{\{u_i^{\gamma'} \leq \alpha\}})_j dx d\alpha \tag{109}$$

$$= \mathbb{E}_{\gamma'} \frac{1}{l} \sum_{i=1}^l \int_0^1 \int_{\Omega} (s_i \cdot 1_{\{u_i^{\gamma'} > \alpha\}} + u^{\gamma'} 1_{\{u_i^{\gamma'} \leq \alpha\}} \langle u^{\gamma'}, s \rangle) dx d\alpha \tag{110}$$

$$= \mathbb{E}_{\gamma'} \frac{1}{l} \sum_{i=1}^l \int_0^1 \int_{\Omega} (s_i \cdot 1_{\{u_i^{\gamma'} > \alpha\}} + (1 - 1_{\{u_i^{\gamma'} > \alpha\}}) \langle u^{\gamma'}, s \rangle) dx d\alpha. \tag{111}$$

We take into account the property [2, Proposition 1.78], which is a direct consequence of Fubini’s theorem, and also used in the proof of the thresholding theorem for the two-class case [7]:

$$\int_0^1 \int_{\Omega} s_i(x) \cdot 1_{\{u_i > \alpha\}}(x) dx d\alpha \tag{112}$$

$$= \int_{\Omega} s_i(x) u_i(x) dx. \tag{113}$$

This leads to

$$\begin{aligned} \mathbb{E}_{\gamma} \int_{\Omega} \langle u_{\gamma}^k, s \rangle dx &= \mathbb{E}_{\gamma'} \frac{1}{l} \sum_{i=1}^l \int_{\Omega} (s_i u_i^{\gamma'} + \langle u^{\gamma'}, s \rangle - u_i^{\gamma'} \langle u^{\gamma'}, s \rangle) dx \end{aligned} \tag{114}$$

and therefore, using  $u^{\gamma'}(x) \in \Delta_I$  a.e.,

$$\mathbb{E}_\gamma \int_\Omega \langle u_\gamma^k, s \rangle dx = \mathbb{E}_{\gamma'} \int_\Omega \langle u^{\gamma'}, s \rangle dx \tag{115}$$

$$= \mathbb{E}_\gamma \int_\Omega \langle u_\gamma^{k-1}, s \rangle dx. \tag{116}$$

Since  $\langle u^0, s \rangle = \langle u, s \rangle$ , the assertion follows by induction.  $\square$

*Remark 2* Proposition 9 shows that the data term is—in the mean—not affected by the probabilistic rounding process, i.e., it satisfies an exact coarea-like formula, even in the multiclass case.

Bounding the regularizer is more involved: For  $\gamma^k = (i^k, \alpha^k)$ , define

$$U_{\gamma^k} := \{x \in \Omega \mid u_{i^k}(x) \leq \alpha^k\}, \tag{117}$$

$$V_{\gamma^k} := (U_{\gamma^k})^1, \tag{118}$$

$$V^k := (U^k)^1. \tag{119}$$

As the measure-theoretic interior is invariant under  $\mathcal{L}^d$ -negligible modifications, given some fixed sequence  $\gamma$  the sequence  $(V^k)$  is invariant under  $\mathcal{L}^d$ -negligible modifications of  $u = u^0$ , i.e., it is uniquely defined when viewing  $u$  as an element of  $L^1(\Omega)^l$ . Some calculations yield

$$U^k = U_{\gamma^1} \cap \dots \cap U_{\gamma^k}, \quad k \geq 1, \tag{120}$$

$$U^{k-1} \setminus U^k = U_{\gamma^1} \cap ((U_{\gamma^2} \cap \dots \cap U_{\gamma^{k-1}}) \setminus (U_{\gamma^2} \cap \dots \cap U_{\gamma^k})), \quad k \geq 2. \tag{121}$$

From these observations and Proposition 4,

$$V^k = V_{\gamma^1} \cap \dots \cap V_{\gamma^k}, \quad k \geq 1, \tag{122}$$

$$V^{k-1} \setminus V^k = V_{\gamma^1} \cap ((V_{\gamma^2} \cap \dots \cap V_{\gamma^{k-1}}) \setminus (V_{\gamma^2} \cap \dots \cap V_{\gamma^k})), \quad k \geq 2, \tag{123}$$

$$\Omega \setminus V^k = \bigcup_{k'=1}^k (V^{k'-1} \setminus V^{k'}), \quad k \geq 1. \tag{124}$$

The last equality can be shown by induction: For the base case  $k = 1$ , we have  $V^0 = (U^0)^1 = (\Omega)^1 = \Omega$ , where the last equality can be shown by mutual inclusion, using the fact that  $\Omega$  is open and has a Lipschitz boundary by assumption. For  $k \geq 2$ ,

$$\bigcup_{k'=1}^k V^{k'-1} \setminus V^{k'} \tag{125}$$

$$= (V^{k-1} \setminus V^k) \cup \bigcup_{k'=1}^{k-1} (V^{k'-1} \setminus V^{k'}) \tag{126}$$

$$= (V^{k-1} \setminus V^k) \cup (\Omega \setminus V^{k-1}) \tag{127}$$

$$\stackrel{V^k \subseteq V^{k-1}}{=} \Omega \setminus V^k, \tag{128}$$

which shows (124).

Moreover, since  $V^k$  is the measure-theoretic interior of  $U^k$ , both sets are equal up to an  $\mathcal{L}^d$ -negligible set (cf. (197)). Again we first show measurability of the involved mappings.

**Proposition 10** For every  $k \geq 1$  the mappings

$$g''^k : \Gamma \rightarrow \mathbb{R}, \quad \gamma \mapsto \int_{V^{k-1} \setminus V^k} d\Psi(D\bar{u}_\gamma) \tag{129}$$

and

$$h'' : \Gamma \rightarrow \mathbb{R}, \quad \gamma \mapsto \int_\Omega d\Psi(D\bar{u}_\gamma) \tag{130}$$

are  $\mu$ -measurable.

*Proof* We only sketch the proof. Let  $k \geq 1$  be arbitrary but fixed. Using a similar argument as in the proof of Proposition 8 (see also the proof of Theorem 1) one can see that  $h''(\gamma) = \sum_{k=1}^\infty g''^k(\gamma)$ , therefore it suffices to show measurability of the  $g''^k$ .

We note that  $g''^k$  can be written, up to a  $\mu$ -negligible set, as the sum

$$g''^k(\gamma) = \sum_{k'=1}^\infty \mathbf{1}_{\{\gamma \mid e^\top c^{k'} < 1 \leq e^\top c^{k'-1}\}} p^{k'}(\gamma), \tag{131}$$

$$p^{k'}(\gamma) := \int_{V^{k-1} \setminus V^k} d\Psi(Du_\gamma^{k'}).$$

The key is that  $u_\gamma^{k'} = \bar{u}_\gamma$  once  $e^\top c^{k'} < 1$ . Each  $p^{k'}$  depends only on a finite number of  $\gamma^i$ , and since the indicator function is measurable, it is enough to show measurability of the mappings  $p^{k'}$  in their respective finite-dimensional subsets of  $\Gamma$  for all  $k' \in \mathbb{N}$ .

Choose a fixed but arbitrary  $k'$ . With the definition  $E_\gamma := U_{\gamma^{k'}}$  we obtain from Proposition 4

$$V^{k-1} \setminus V^k = V^{k-1} \cap (\Omega \setminus (E_\gamma)^1), \tag{132}$$

which together with [2, Theorem 3.84] leads to

$$p^{k'}(\gamma) = \int_{V^{k-1} \cap \mathcal{F}E_\gamma} d\Psi(Du_\gamma^{k'}) \tag{133}$$

$$= \int_\Omega \Psi((\nu_{E_\gamma})((u_\gamma^{k'})_{\mathcal{F}E_\gamma}^+ - (u_\gamma^{k'})_{\mathcal{F}E_\gamma}^-))^\top \cdot \mathbf{1}_{V^{k-1}} d|D1_{E_\gamma}|, \tag{134}$$

where  $\nu_{E_\gamma}(x) := (D1_{E_\gamma}/|D1_{E_\gamma}|)(x)$  on  $\mathcal{F}E_\gamma$ . Measurability of the  $p^{k'}$  can be shown using a result about measure-valued mappings [2, Proposition 2.26]. This first requires

to show that the mapping  $\gamma \mapsto |D1_{E_\gamma}|(B)$  is  $\mu$ -measurable for every open set  $B \subseteq \Omega$ , which is a corollary of the coarea formula [2, Theorem 3.40].

The second requirement is that the integrand in (134) is bounded and  $(\mathcal{B}_\mu \times \mathcal{B}(\Omega))$ -measurable. For the indicator function this follows from the definitions in a straightforward way. The normal mapping can be rewritten as

$$(\gamma, x) \mapsto 1_{\mathcal{F}E_\gamma} \lim_{\rho \rightarrow 0} D1_{E_\gamma}(\mathcal{B}_\rho(x)) / |D1_{E_\gamma}(\mathcal{B}_\rho(x))|. \tag{135}$$

Using a slight modification of [2, Proposition 2.26] one can show the  $(\mathcal{B}_\mu \times \mathcal{B}(\Omega))$ -measurability of the mappings  $(\gamma, x) \mapsto D1_{E_\gamma}(\mathcal{B}_\rho(x))$  and  $(\gamma, x) \mapsto |D1_{E_\gamma}(\mathcal{B}_\rho(x))|$ , and therefore of  $1_{\mathcal{F}E_\gamma}$  and of the normal mapping in (135). Together with Proposition 7 this ensures  $(\mathcal{B}_\mu \times \mathcal{B}(\Omega))$ -measurability of the normal and trace terms in (134), and, since  $\Psi$  is continuous, of the whole integrand.

Therefore all assumptions of [2, Proposition 2.26] are fulfilled, and we obtain the  $\mu$ -measurability of all  $p^{k'}$  and finally of  $g''^k$  and  $h''$ .  $\square$

We now prepare for an induction argument on the expectation of the regularizing term when restricted to the sets  $V^{k-1} \setminus V^k$ . The following proposition provides the initial step ( $k = 1$ ).

**Proposition 11** *Assume that  $\Psi$  satisfies the lower- and upper-boundedness conditions (24) and (25). Then*

$$\mathbb{E} \int_{V^0 \setminus V^1} d\Psi(D\bar{u}) \leq \frac{2}{l} \frac{\lambda_u}{\lambda_l} \int_{\Omega} d\Psi(Du). \tag{136}$$

*Proof* Denote  $(i, \alpha) = \gamma^1$ . Since  $1_{U(i,\alpha)} = 1_{V(i,\alpha)}$   $\mathcal{L}^d$ -a.e., we have

$$\bar{u}_\gamma = 1_{\Omega \setminus V(i,\alpha)} e^i + 1_{V(i,\alpha)} \bar{u}_\gamma \quad \mathcal{L}^d\text{-a.e.} \tag{137}$$

Therefore, since  $V^0 = (U^0)^1 = (\Omega)^1 = \Omega$ ,

$$\begin{aligned} & \int_{V^0 \setminus V^1} d\Psi(D\bar{u}_\gamma) \\ &= \int_{\Omega \setminus V(i,\alpha)} d\Psi(D\bar{u}_\gamma) \\ &= \int_{\Omega \setminus V(i,\alpha)} d\Psi(D(1_{\Omega \setminus V(i,\alpha)} e^i + 1_{V(i,\alpha)} \bar{u}_\gamma)). \end{aligned} \tag{138}$$

Since  $u \in \text{BV}(\Omega)^l$ , we know that  $\text{Per}(V(i,\alpha)) < \infty$  holds for  $\mathcal{L}^1$ -a.e.  $\alpha$  and any  $i$  [2, Theorem 3.40]. Therefore we conclude from Proposition 5 that for  $\mathcal{L}^1$ -a.e.  $\alpha$ ,

$$\begin{aligned} & \int_{\Omega \setminus V(i,\alpha)} d\Psi(D\bar{u}_\gamma) \\ & \leq \lambda_u \text{Per}(V(i,\alpha)) \end{aligned}$$

$$\begin{aligned} & + \int_{(\Omega \setminus V(i,\alpha)) \cap (\Omega \setminus V(i,\alpha))^1} d\Psi(De^i) \\ & + \int_{(\Omega \setminus V(i,\alpha)) \cap (\Omega \setminus V(i,\alpha))^0} d\Psi(D\bar{u}_\gamma). \end{aligned} \tag{139}$$

Both of the integrals are zero, since  $De^i = 0$  and

$$(\Omega \setminus V(i,\alpha))^0 = (V(i,\alpha))^1 = V(i,\alpha), \tag{140}$$

therefore

$$\int_{\Omega \setminus V(i,\alpha)} d\Psi(D\bar{u}_\gamma) \leq \lambda_u \text{Per}(V(i,\alpha)). \tag{141}$$

Since Proposition 10 provides measurability the bound carries over to the expectation,

$$\mathbb{E}_\gamma \int_{\Omega \setminus V(i,\alpha)} d\Psi(D\bar{u}_\gamma) \leq \frac{1}{l} \sum_{i=1}^l \int_0^1 \lambda_u \text{Per}(V(i,\alpha)) d\alpha.$$

Also  $\text{Per}(V(i,\alpha)) = \text{Per}(U(i,\alpha))$  since the perimeter is invariant under  $\mathcal{L}^d$ -negligible modifications. The assertion then follows using  $V^0 = \Omega$ ,  $V^1 = V(i,\alpha)$  and the coarea formula:

$$\mathbb{E}_\gamma \int_{V^0 \setminus V^1} d\Psi(D\bar{u}_\gamma) \tag{142}$$

$$\leq \frac{1}{l} \sum_{i=1}^l \int_0^1 \lambda_u \text{Per}(U(i,\alpha)) d\alpha \tag{143}$$

$$\stackrel{\text{coarea}}{=} \frac{\lambda_u}{l} \sum_{i=1}^l \text{TV}(u_i) = \frac{\lambda_u}{l} \int_{\Omega} \sum_{i=1}^l d\|Du_i\|_2 \tag{144}$$

$$\stackrel{(24)}{\leq} \frac{2}{l} \frac{\lambda_u}{\lambda_l} \int_{\Omega} d\Psi(Du). \tag{145}$$

$\square$

We now take care of the induction step for the regularizer bound.

**Proposition 12** *Let  $\Psi$  satisfy the upper-boundedness condition (25). Then, for any  $k \geq 2$ ,*

$$F := \mathbb{E} \int_{V^{k-1} \setminus V^k} d\Psi(D\bar{u}) \tag{146}$$

$$\leq \frac{(l-1)}{l} \mathbb{E} \int_{V^{k-2} \setminus V^{k-1}} d\Psi(D\bar{u}). \tag{147}$$

*Proof* Define the shifted sequence  $\gamma' = (\gamma'^k)_{k=1}^\infty$  by  $\gamma'^k := \gamma^{k+1}$ , and let

$$W_{\gamma'} := V_{\gamma'}^{k-2} \setminus V_{\gamma'}^{k-1} \tag{148}$$

$$= (V_{\gamma^2} \cap \dots \cap V_{\gamma^{k-1}}) \setminus (V_{\gamma^2} \cap \dots \cap V_{\gamma^k}). \tag{149}$$



By Proposition 2 and Proposition 10 we may assume that  $\bar{u}_\gamma$  exists  $\mu$ -a.e. and is an element of  $BV(\Omega)^l$ , and that the expectation is well-defined. We denote  $\gamma^1 = (i, \alpha)$ , then  $V^{k-1} \setminus V^k = V_{(i,\alpha)} \cap W_{\gamma'}$  due to (123). For each pair  $(i, \alpha)$  we denote by  $((i, \alpha), \gamma')$  the sequence obtained by prepending  $(i, \alpha)$  to the sequence  $\gamma'$ . Then

$$F = \frac{1}{l} \sum_{i=1}^l \int_0^1 \mathbb{E}_{\gamma'} \int_{V_{(i,\alpha)} \cap W_{\gamma'}} d\Psi(D\bar{u}_{((i,\alpha),\gamma')}) d\alpha. \tag{150}$$

Since in the first iteration of the algorithm no points in  $U_{(i,\alpha)}$  are assigned a label,  $\bar{u}_{((i,\alpha),\gamma')} = \bar{u}_{\gamma'}$  holds on  $U_{(i,\alpha)}$ , and therefore  $\mathcal{L}^d$ -a.e. on  $V_{(i,\alpha)}$ . Therefore we may apply Proposition 6 and substitute  $D\bar{u}_{((i,\alpha),\gamma')}$  by  $D\bar{u}_{\gamma'}$  in (150):

$$F = \frac{1}{l} \sum_{i=1}^l \int_0^1 \left( \mathbb{E}_{\gamma'} \int_{V_{(i,\alpha)} \cap W_{\gamma'}} d\Psi(D\bar{u}_{\gamma'}) \right) d\alpha \tag{151}$$

$$= \frac{1}{l} \sum_{i=1}^l \int_0^1 \left( \mathbb{E}_{\gamma'} \int_{W_{\gamma'}} 1_{V_{(i,\alpha)}} d\Psi(D\bar{u}_{\gamma'}) \right) d\alpha. \tag{152}$$

By definition of the measure-theoretic interior (87), the indicator function  $1_{V_{(i,\alpha)}}$  is bounded from above by the density function  $\Theta_{U_{(i,\alpha)}}$  of  $U_{(i,\alpha)}$ ,

$$1_{V_{(i,\alpha)}}(x) \leq \Theta_{(i,\alpha)}(x) := \lim_{\delta \searrow 0} \frac{|\mathcal{B}_\delta(x) \cap U_{(i,\alpha)}|}{|\mathcal{B}_\delta(x)|}, \tag{153}$$

which exists  $\mathcal{H}^{d-1}$ -a.e. on  $\Omega$  by [2, Proposition 3.61]. Therefore, denoting by  $\mathcal{B}_\delta(\cdot)$  the mapping  $x \in \Omega \mapsto \mathcal{B}_\delta(x)$ ,

$$F \leq \frac{1}{l} \sum_{i=1}^l \int_0^1 \mathbb{E}_{\gamma'} \int_{W_{\gamma'}} \lim_{\delta \searrow 0} \frac{|\mathcal{B}_\delta(\cdot) \cap U_{(i,\alpha)}|}{|\mathcal{B}_\delta(\cdot)|} d\Psi(D\bar{u}_{\gamma'}) d\alpha.$$

Rearranging the integrals and the limit, which can be justified by  $TV(\bar{u}_{\gamma'}) < \infty$  almost surely and dominated convergence using (25), we get

$$\begin{aligned} F &\leq \frac{1}{l} \mathbb{E}_{\gamma'} \lim_{\delta \searrow 0} \int_{W_{\gamma'}} \sum_{i=1}^l \int_0^1 \frac{|\mathcal{B}_\delta(\cdot) \cap U_{(i,\alpha)}|}{|\mathcal{B}_\delta(\cdot)|} d\alpha d\Psi(D\bar{u}_{\gamma'}) \\ &= \frac{1}{l} \mathbb{E}_{\gamma'} \lim_{\delta \searrow 0} \int_{W_{\gamma'}} \frac{1}{|\mathcal{B}_\delta(\cdot)|} \\ &\quad \cdot \left( \sum_{i=1}^l \int_0^1 \int_{\mathcal{B}_\delta(\cdot)} 1_{\{u_i(y) \leq \alpha\}} dy d\alpha \right) d\Psi(D\bar{u}_{\gamma'}). \end{aligned} \tag{154}$$

We again apply [2, Proposition 1.78] to the two innermost integrals (alternatively, use Fubini’s theorem), which leads to

$$F \leq \frac{1}{l} \mathbb{E}_{\gamma'} \lim_{\delta \searrow 0} \int_{W_{\gamma'}} \frac{1}{|\mathcal{B}_\delta(\cdot)|} \tag{155}$$

$$\cdot \left( \sum_{i=1}^l \int_{\mathcal{B}_\delta(\cdot)} (1 - u_i(y)) dy \right) d\Psi(D\bar{u}_{\gamma'}). \tag{156}$$

Using the fact that  $u(y) \in \Delta_l$ , this collapses according to

$$\begin{aligned} F &\leq \frac{1}{l} \mathbb{E}_{\gamma'} \lim_{\delta \searrow 0} \int_{W_{\gamma'}} \frac{1}{|\mathcal{B}_\delta(\cdot)|} \left( \int_{\mathcal{B}_\delta(\cdot)} (l-1) dy \right) d\Psi(D\bar{u}_{\gamma'}) \\ &= \frac{1}{l} \mathbb{E}_{\gamma'} \lim_{\delta \searrow 0} \int_{W_{\gamma'}} (l-1) d\Psi(D\bar{u}_{\gamma'}) \end{aligned} \tag{157}$$

$$= \frac{l-1}{l} \mathbb{E}_{\gamma'} \int_{W_{\gamma'}} d\Psi(D\bar{u}_{\gamma'}) \tag{158}$$

$$= \frac{l-1}{l} \mathbb{E}_{\gamma'} \int_{V_{\gamma'}^{k-2} \setminus V_{\gamma'}^{k-1}} d\Psi(D\bar{u}_{\gamma'}). \tag{159}$$

Reverting the index shift and using  $\bar{u}_{\gamma'} = \bar{u}_\gamma$  concludes the proof:

$$F \leq \frac{l-1}{l} \mathbb{E}_\gamma \int_{V_\gamma^{k-1} \setminus V_\gamma^k} d\Psi(D\bar{u}_\gamma). \tag{160}$$

□

We are now ready to prove the main result, Theorem 1, as stated in the introduction.

*Proof of Theorem 1* The fact that the algorithm provides  $\bar{u} \in \mathcal{C}_\mathcal{E}$  almost surely follows from Theorem 2. Therefore there almost surely exists  $k' := k'(\gamma) \geq 1$  such that  $|U^{k'}| = 0$  and  $\bar{u}_\gamma = u_\gamma^{k'}$ . On one hand, this implies

$$\int_\Omega \langle \bar{u}_\gamma, s \rangle dx = \int_\Omega \langle u_\gamma^{k'}, s \rangle dx = \lim_{k \rightarrow \infty} \int_\Omega \langle u_\gamma^k, s \rangle dx \tag{161}$$

almost surely. On the other hand,  $V^{k'} = (U^{k'})^1 = \emptyset$  and therefore

$$\bigcup_{k=1}^{k'} V^{k-1} \setminus V^k \stackrel{(124)}{=} \Omega \setminus V^{k'} = \Omega \tag{162}$$

almost surely. From (161) and (162) we obtain

$$\begin{aligned} \mathbb{E}_\gamma f(\bar{u}_\gamma) &= \mathbb{E}_\gamma \left( \lim_{k \rightarrow \infty} \int_\Omega \langle u_\gamma^k, s \rangle dx \right) \\ &\quad + \mathbb{E}_\gamma \left( \sum_{k=1}^\infty \int_{V^{k-1} \setminus V^k} d\Psi(D\bar{u}_\gamma) \right). \end{aligned} \tag{163}$$

In the first term, the  $u_\gamma^k$  are elements of  $BV(\Omega, \Delta_l)$  and therefore  $L^\infty(\Omega, \mathbb{R}^l)$  except possibly on a negligible set of  $\gamma$ . Since  $s \in L^1(\Omega)$  this means  $\gamma \mapsto \langle u_\gamma^k, s \rangle = |\langle u_\gamma^k, s \rangle|$  is bounded from above by a constant outside a negligible

set (by Proposition 8 it is also measurable) and the dominated convergence theorem applies. The second term satisfies the requirements for monotone convergence, since all summands exist, are nonnegative almost surely, and measurable by Proposition 10. Therefore the integrals and limits can be swapped,

$$\mathbb{E}_\gamma f(\bar{u}_\gamma) = \lim_{k \rightarrow \infty} \left( \mathbb{E}_\gamma \int_\Omega \langle u_\gamma^k, s \rangle dx + \sum_{k=1}^\infty \mathbb{E}_\gamma \int_{V^{k-1} \setminus V^k} d\Psi(D\bar{u}_\gamma) \right) \tag{164}$$

The first term in (164) is equal to  $\int_\Omega \langle u, s \rangle dx$  due to Proposition 9. An induction argument using Proposition 11 and 12 shows that the second term can be bounded according to

$$\sum_{k=1}^\infty \mathbb{E}_\gamma \int_{V^{k-1} \setminus V^k} d\Psi(D\bar{u}_\gamma) \tag{165}$$

$$\leq \sum_{k=1}^\infty \left( \frac{l-1}{l} \right)^{k-1} \frac{2\lambda_u}{l\lambda_l} \int_\Omega d\Psi(Du) \tag{166}$$

$$= 2 \frac{\lambda_u}{\lambda_l} \int_\Omega d\Psi(Du), \tag{167}$$

therefore

$$\mathbb{E}_\gamma f(\bar{u}_\gamma) \leq \int_\Omega \langle u, s \rangle dx + 2 \frac{\lambda_u}{\lambda_l} \int_\Omega d\Psi(Du). \tag{168}$$

Since  $s \geq 0$  and  $\lambda_u \geq \lambda_l$ , and therefore the linear term is bounded by  $\int_\Omega \langle u, s \rangle dx \leq 2(\lambda_u/\lambda_l) \int_\Omega \langle u, s \rangle dx$ , this proves the assertion.  $\square$

Corollary 1 (see Sect. 1) follows immediately using  $f(u^*) \leq f(u_\epsilon^*)$ , cf. (45). We have demonstrated that the proposed approach allows to recover, from the solution  $u^*$  of the convex relaxed problem (13), an approximate integral solution  $\bar{u}^*$  of the nonconvex original problem (1) with an upper bound on the objective.

For the specific case  $\Psi = \Psi_d$  as in (9), we have

**Proposition 13** *Let  $d : \mathcal{I}^2 \rightarrow \mathbb{R}_{\geq 0}$  be a metric and  $\Psi = \Psi_d$ . Then one may set*

$$\lambda_u = \max_{i,j \in \{1, \dots, l\}} d(i, j) \quad \text{and} \quad \lambda_l = \min_{i \neq j} d(i, j).$$

*Proof* From the remarks in the introduction we obtain (cf. [19])

$$\Psi_d(v(e^i - e^j)^\top) = d(i, j),$$

which shows the upper bound. For the lower bound, take any  $z \in \mathbb{R}^{d \times l}$  satisfying  $ze = 0$  as in (24), set  $c := \min_{i \neq j} d(i, j)$ ,

$v^i := \frac{c}{2} \frac{z^i}{\|z^i\|_2}$  if  $z^i \neq 0$  and  $v^i := 0$  otherwise, and  $v := v'(I - \frac{1}{7}ee^\top)$ . Then  $v \in \mathcal{D}_{\text{loc}}^d$ , since  $\|v^i - v^j\|_2 = \|v^i - v^j\|_2 \leq c$  and  $ve = v'(I - \frac{1}{7}ee^\top)e = 0$ . Therefore,

$$\Psi_d(z) \geq \langle z, v \rangle = \langle z, v' \rangle \tag{169}$$

$$= \sum_{i=1, \dots, l, z^i \neq 0} \left\langle z^i, \frac{c}{2} \frac{z^i}{\|z^i\|_2} \right\rangle = \frac{c}{2} \sum_{i=1}^l \|z^i\|_2, \tag{170}$$

proving the lower bound.  $\square$

Finally, for  $\Psi_d$  we obtain the factor

$$2 \frac{\lambda_u}{\lambda_l} = 2 \frac{\max_{i,j} d(i, j)}{\min_{i \neq j} d(i, j)}, \tag{171}$$

determining the optimality bound, as claimed in the introduction (29). The bound in (28) is the same as the known bounds for finite-dimensional metric labeling [12] and  $\alpha$ -expansion [4], however it extends these results to problems on continuous domains for a broad class of regularizers.

### 5 Conclusion

In this work we considered a method for recovering approximate solutions of image partitioning problems from solutions of a convex relaxation. We proposed a probabilistic rounding method motivated by the finite-dimensional framework, and showed that it is possible to obtain *a priori* bounds on the optimality of the integral solution obtained by rounding a solution of the convex relaxation.

The obtained bounds are compatible with known bounds for the finite-dimensional setting. However, to our knowledge, this is the first fully convex approach that is both formulated in the spatially continuous setting and provides a true *a priori* bound. We showed that the approach can also be interpreted as an approximate variant of the coarea formula.

A peculiar property of the presented approach is that it provides a bound of two for the uniform metric even in the two-class case, where the relaxation is known to be exact. The question remains how to prove an optimal bound.

While the results apply to a quite general class of regularizers, they are formulated for the homogeneous case. Non-homogeneous regularizers constitute an interesting direction for future work. In particular, such regularizers naturally occur when applying convex relaxation techniques [1, 24] in order to solve nonconvex variational problems.

With the increasing computational power, such techniques have become quite popular recently. For problems where the convexity is confined to the data term, they permit to find a global minimizer. A proper extension of the results outlined in this work may provide a way to find good approximate solutions of problems where also the regularizer is nonconvex.

**Acknowledgements** This publication is partly based on work supported by Award No. KUK-I1-007-43, made by King Abdullah University of Science and Technology (KAUST).

**Appendix**

*Proof of Proposition 3* In order to prove the first assertion (88), note that the mapping  $w \mapsto \Psi(vw^\top)$  is convex, therefore it must assume its maximum on the polytope  $\Delta_l - \Delta_l := \{z^1 - z^2 \mid z^1, z^2 \in \Delta_l\}$  in a vertex of the polytope. Since the polytope  $\Delta_l - \Delta_l$  is the difference of two polytopes, its vertex set is at most the difference of their vertex sets,  $V := \{e^j - e^i \mid i, j \in \{1, \dots, l\}\}$ . On this set, the bound  $\Psi(vw^\top) \leq \lambda_u$  holds for  $w \in V$  due to the upper-boundedness condition (25), which proves (88).

The second equality (90) follows from the fact that  $G := \{b^{ik} := e^k(e^i - e^{i+1})^\top \mid 1 \leq k \leq d, 1 \leq i \leq l - 1\}$  is a basis of the linear subspace  $W$ , satisfying  $\Psi(b^{ik}) \leq \lambda_u$ , and  $\Psi$  is positively homogeneous and convex, and thus subadditive. Specifically, there is a linear transform  $T : W \rightarrow \mathbb{R}^{d \times (l-1)}$  such that  $w = \sum_{i,k} b^{ik} \alpha_{ik}$  for  $\alpha = T(w)$ . Then

$$\Psi(w) = \Psi\left(\sum_{i,k} b^{ik} \alpha_{ik}\right) \tag{172}$$

$$\leq \Psi\left(\sum_{i,k} |\alpha_{ik}| \operatorname{sgn}(\alpha_{ik}) b^{ik}\right) \tag{173}$$

$$\leq \sum_{i,k} |\alpha_{ik}| \Psi(\operatorname{sgn}(\alpha_{ik}) b^{ik}). \tag{174}$$

Since (25) ensures  $\Psi(\pm b^{ik}) \leq \lambda_u$ , we obtain

$$\Psi(w) \leq \lambda_u \sum_{i,k} |\alpha_{ik}| \leq \lambda_u \|T\| \|w\|_2 \tag{175}$$

for any suitable operator norm  $\|\cdot\|$  and any  $w \in W$ .  $\square$

*Proof of Proposition 4* Denote  $\mathcal{B}_\delta := \mathcal{B}_\delta(x)$ . We prove mutual inclusion:

“ $\subseteq$ ”: From the definition of the measure-theoretic interior,

$$x \in (E \cap F)^1 \Rightarrow \lim_{\delta \searrow 0} \frac{|\mathcal{B}_\delta \cap E \cap F|}{|\mathcal{B}_\delta|} = 1. \tag{176}$$

Since  $|\mathcal{B}_\delta| \geq |\mathcal{B}_\delta \cap E| \geq |\mathcal{B}_\delta \cap E \cap F|$  (and vice versa for  $|\mathcal{B}_\delta \cap F|$ ), it follows by the “sandwich” criterion that both  $\lim_{\delta \searrow 0} |\mathcal{B}_\delta \cap E|/|\mathcal{B}_\delta|$  and  $\lim_{\delta \searrow 0} |\mathcal{B}_\delta \cap F|/|\mathcal{B}_\delta|$  exist and are equal to 1, which shows  $x \in E^1 \cap F^1$ .

“ $\supseteq$ ”: Assume that  $x \in E^1 \cap F^1$ . Then

$$1 \geq \limsup_{\delta \searrow 0} \frac{|\mathcal{B}_\delta \cap E \cap F|}{|\mathcal{B}_\delta|} \tag{177}$$

$$\geq \liminf_{\delta \searrow 0} \frac{|\mathcal{B}_\delta \cap E \cap F|}{|\mathcal{B}_\delta|} \tag{178}$$

$$= \liminf_{\delta \searrow 0} \frac{|\mathcal{B}_\delta \cap E| + |\mathcal{B}_\delta \cap F| - |\mathcal{B}_\delta \cap (E \cup F)|}{|\mathcal{B}_\delta|}. \tag{179}$$

We obtain equality,

$$1 \geq \liminf_{\delta \searrow 0} \frac{|\mathcal{B}_\delta \cap E \cap F|}{|\mathcal{B}_\delta|} \tag{180}$$

$$\geq \liminf_{\delta \searrow 0} \frac{|\mathcal{B}_\delta \cap E|}{|\mathcal{B}_\delta|} + \liminf_{\delta \searrow 0} \frac{|\mathcal{B}_\delta \cap F|}{|\mathcal{B}_\delta|} + \liminf_{\delta \searrow 0} \left( - \frac{|\mathcal{B}_\delta \cap (E \cup F)|}{|\mathcal{B}_\delta|} \right) \tag{181}$$

$$= 2 - \limsup_{\delta \searrow 0} \frac{|\mathcal{B}_\delta \cap (E \cup F)|}{|\mathcal{B}_\delta|} \geq 1, \tag{182}$$

$\leq 1$

from which we conclude that

$$\limsup_{\delta \searrow 0} \frac{|\mathcal{B}_\delta \cap E \cap F|}{|\mathcal{B}_\delta|} = \liminf_{\delta \searrow 0} \frac{|\mathcal{B}_\delta \cap E \cap F|}{|\mathcal{B}_\delta|} = 1,$$

i.e.,  $x \in (E \cap F)^1$ .  $\square$

*Proof of Proposition 5* First note that

$$\int_{\mathcal{F}E \cap \Omega} \|w_{\mathcal{F}E}^+ - w_{\mathcal{F}E}^-\|_2 d\mathcal{H}^{d-1} \tag{183}$$

$$\leq \sup\{\|w_{\mathcal{F}E}^+(x) - w_{\mathcal{F}E}^-(x)\|_2 \mid x \in \mathcal{F}E \cap \Omega\} \cdot \mathcal{H}^{d-1}(\mathcal{F}E \cap \Omega) \tag{184}$$

$$\stackrel{(*)}{\leq} \operatorname{esssup}\{\|w(x) - w(y)\|_2 \mid x, y \in \Omega\} \cdot \operatorname{TV}(1_E) \tag{185}$$

$$\stackrel{(**)}{\leq} \sqrt{2} \operatorname{TV}(1_E) \tag{186}$$

$$= \sqrt{2} \operatorname{Per}(E) < \infty. \tag{187}$$

The inequality (\*) is a consequence of the definition of  $w_{\mathcal{F}E}^\pm$  and [2, Theorem 3.77], and (\*\*) follows directly from  $w(x), w(y) \in \Delta_l$  a.e. on  $\Omega$ . The upper bound (187) permits applying [2, Theorem 3.84] on  $w$ , which provides  $w \in \operatorname{BV}(\Omega)^l$  and (94). Due to [2, Proposition 3.61], the sets  $(E)^0, (E)^1$  and  $\mathcal{F}E$  form a (pairwise disjoint) partition of  $\Omega$ , up to an  $\mathcal{H}^{d-1}$ -zero set. Therefore, since  $\Psi(Du) \ll |Du| \ll \mathcal{H}^{d-1}$  by construction, from [2, Theorem 2.37, 3.84] we obtain, for any Borel set  $A$ ,

$$\int_A d\Psi(Dw) \tag{188}$$

$$= \int_{A \cap (E)^1} d\Psi(Dw) + \int_{A \cap (E)^0} d\Psi(Dw)$$

$$+ \int_{A \cap \mathcal{F}E \cap \Omega} \Psi(v_E(w_{\mathcal{F}E}^+(x) - w_{\mathcal{F}E}^-(x))^\top) d\mathcal{H}^{d-1}. \tag{189}$$

Since  $w(x) \in \Delta_l$  a.e. by assumption, we conclude that  $w_{\mathcal{F}E}^+$  and  $w_{\mathcal{F}E}^-$  must have values in  $\Delta_l$  as well, see [2, Theorem 3.77]. Therefore we can apply Proposition 3 to obtain

$$\int_A d\Psi(Dw) \tag{190}$$

$$\leq \int_{A \cap (E)^1} d\Psi(Dw) + \int_{A \cap (E)^0} d\Psi(Dw) + \int_{A \cap \mathcal{F}E \cap \Omega} \lambda_u d\mathcal{H}^{d-1} \tag{191}$$

$$\leq \int_{A \cap (E)^1} d\Psi(Dw) + \int_{A \cap (E)^0} d\Psi(Dw) + \lambda_u \text{Per}(E). \tag{192}$$

We rewrite  $\Psi(Dw)$  using (94),

$$\Psi(Dw) = \Psi(Du_{\perp}(E)^1 + Dv_{\perp}(E)^0 + v_E(u_{\mathcal{F}E}^+ - v_{\mathcal{F}E}^-)^\top \mathcal{H}^{d-1} \llcorner (\mathcal{F}E \cap \Omega)). \tag{193}$$

From [2, Proposition 2.37] we obtain that  $\Psi$  is additive on mutually singular Radon measures  $\mu, \nu$ , i.e., if  $|\mu| \perp |\nu|$ , then

$$\int_B d\Psi(\mu + \nu) = \int_B d\Psi(\mu) + \int_B d\Psi(\nu) \tag{194}$$

for any Borel set  $B \subseteq \Omega$ . This holds in particular for the three measures in (193), therefore

$$\Psi(Dw) = \Psi(Du_{\perp}(E)^1) + \Psi(Dv_{\perp}(E)^0) + \Psi(v_E(u_{\mathcal{F}E}^+ - v_{\mathcal{F}E}^-)^\top \mathcal{H}^{d-1} \llcorner (\mathcal{F}E \cap \Omega)). \tag{195}$$

Since  $Du_{\perp}(E)^1 \ll |Du_{\perp}(E)^1| = |Du|_{\perp}(E)^1$ , we conclude  $\Psi(Dw)_{\perp}(E)^1 = \Psi(Du)_{\perp}(E)^1$  and  $\Psi(Dw)_{\perp}(E)^0 = \Psi(Dv)_{\perp}(E)^0$ . Substitution into (192) proves the remaining assertion,

$$\int_A d\Psi(Dw) \leq \int_{A \cap (E)^1} d\Psi(Du) + \int_{A \cap (E)^0} d\Psi(Dv) + \lambda_u \text{Per}(E). \tag{196}$$

□

*Proof of Proposition 6* We first show (98). It suffices to show that

$$\{x \in (E)^1 \Leftrightarrow x \in E\} \text{ for } \mathcal{L}^d\text{-a.e. } x \in \Omega. \tag{197}$$

This can be seen by considering the precise representative  $\tilde{1}_E$  of  $1_E$  [2, Definition 3.63]: Starting with the definition,

$$x \in (E)^1 \Leftrightarrow \lim_{\delta \searrow 0} \frac{|E \cap \mathcal{B}_\delta(x)|}{|\mathcal{B}_\delta(x)|} = 1, \tag{198}$$

the fact that  $\lim_{\delta \searrow 0} \frac{|\Omega \cap \mathcal{B}_\delta(x)|}{|\mathcal{B}_\delta(x)|} = 1$  implies

$$x \in (E)^1 \Leftrightarrow \lim_{\delta \searrow 0} \frac{|(\Omega \setminus E) \cap \mathcal{B}_\delta(x)|}{|\mathcal{B}_\delta(x)|} = 0 \tag{199}$$

$$\Leftrightarrow \lim_{\delta \searrow 0} \frac{1}{|\mathcal{B}_\delta(x)|} \int_{\mathcal{B}_\delta(x)} |1_E - 1| dy = 0 \tag{200}$$

$$\Leftrightarrow \tilde{1}_E(x) = 1. \tag{201}$$

Substituting  $E$  by  $\Omega \setminus E$ , the same equivalence shows that  $x \in (E)^0 \Leftrightarrow \tilde{1}_{\Omega \setminus E}(x) = 1 \Leftrightarrow \tilde{1}_E(x) = 0$ . As  $\mathcal{L}^d(\Omega \setminus ((E)^0 \cup (E)^1)) = 0$ , this shows that  $1_{E^1} = \tilde{1}_E$   $\mathcal{L}^d$ -a.e. Using the fact that  $\tilde{1}_E = 1_E$  [2, Proposition 3.64], we conclude that  $1_{(E)^1} = 1_E$   $\mathcal{L}^d$ -a.e., which proves (197) and therefore the assertion (98).

Since the measure-theoretic interior  $(E)^1$  is defined over  $\mathcal{L}^d$ -integrals, it is invariant under  $\mathcal{L}^d$ -negligible modifications of  $E$ . Together with (197) this implies

$$((E)^1)^1 = (E)^1, \quad \mathcal{F}(E)^1 = \mathcal{F}E, \quad ((E)^1)^0 = (E)^0. \tag{202}$$

To show the relation  $(Du)_{\perp}(E)^1 = (Dv)_{\perp}(E)^1$ , consider

$$Du_{\perp}(E)^1 = D(1_{\Omega \setminus (E)^1}u + 1_{(E)^1}u)_{\perp}(E)^1 \tag{203}$$

$$\stackrel{(*)}{=} D(1_{\Omega \setminus (E)^1}u + 1_{(E)^1}v)_{\perp}(E)^1. \tag{204}$$

The equality (\*) holds due to the assumption (96), and due to the fact that  $Df = Dg$  if  $f = g$   $\mathcal{L}^d$ -a.e. (see, e.g., [2, Proposition 3.2]). We continue from (204) via

$$Du_{\perp}(E)^1 \tag{205}$$

$$\stackrel{\text{Proposition 5}}{=} \{Du_{\perp}((E)^1)^0 + Dv_{\perp}((E)^1)^1\} \tag{206}$$

$$+ v_{(E)^1}(u_{\mathcal{F}E^1}^+ - v_{\mathcal{F}E^1}^-)^\top \mathcal{H}^{d-1} \llcorner (\mathcal{F}(E)^1 \cap \Omega) \}_{\perp}(E)^1$$

$$\stackrel{(202)}{=} (Du_{\perp}(E)^0 + Dv_{\perp}(E)^1)_{\perp}(E)^1$$

$$+ (v_{(E)^1}(u_{\mathcal{F}E^1}^+ - v_{\mathcal{F}E^1}^-)^\top \mathcal{H}^{d-1} \llcorner (\mathcal{F}E \cap \Omega))_{\perp}(E)^1 \tag{207}$$

$$\begin{aligned}
 &= Du_{\mathcal{L}}((E)^0 \cap (E)^1) + Dv_{\mathcal{L}}((E)^1 \cap (E)^1) \\
 &\quad + v_{(E)^1}(u_{\mathcal{F}E^1}^+ - v_{\mathcal{F}E^1}^-)^{\top} \mathcal{H}^{d-1}_{\mathcal{L}}(\mathcal{F}E \cap \Omega \cap (E)^1)
 \end{aligned} \tag{208}$$

$$= Dv_{\mathcal{L}}(E)^1. \tag{209}$$

Therefore  $Du_{\mathcal{L}}(E)^1 = Dv_{\mathcal{L}}(E)^1$ . Then,

$$\begin{aligned}
 &\Psi(Du)_{\mathcal{L}}(E)^1 \\
 &= \Psi(Du_{\mathcal{L}}(E)^1 + Du_{\mathcal{L}}(\Omega \setminus (E)^1))_{\mathcal{L}}(E)^1
 \end{aligned} \tag{210}$$

$$\stackrel{(*)}{=} \Psi(Du_{\mathcal{L}}(E)^1)_{\mathcal{L}}(E)^1 + \Psi(Du_{\mathcal{L}}(\Omega \setminus (E)^1))_{\mathcal{L}}(E)^1. \tag{211}$$

In the equality (\*) we used the additivity of  $\Psi$  on mutually singular Radon measures [2, Proposition 2.37]. By definition of the total variation,  $|\mu|_A = |\mu|_{\mathcal{L}}A$  holds for any measure  $\mu$ , therefore  $|Du_{\mathcal{L}}(\Omega \setminus (E)^1)| = |Du|_{\mathcal{L}}(\Omega \setminus (E)^1)$  and  $|Du_{\mathcal{L}}(\Omega \setminus (E)^1)|((E)^1) = 0$ , which together with (again by definition)  $\Psi(\mu) \ll |\mu|$  implies that the second term in (211) vanishes. Since all observations equally hold for  $v$  instead of  $u$ , we conclude

$$\Psi(Du)_{\mathcal{L}}(E)^1 = \Psi(Du_{\mathcal{L}}(E)^1)_{\mathcal{L}}(E)^1 \tag{212}$$

$$\stackrel{(209)}{=} \Psi(Dv_{\mathcal{L}}(E)^1)_{\mathcal{L}}(E)^1 \tag{213}$$

$$= \Psi(Dv)_{\mathcal{L}}(E)^1. \tag{214}$$

Equation (97) follows immediately. □

**References**

1. Alberti, G.: The calibration method for the Mumford-Shah functional and free-discontinuity problems. *Calc. Var. Partial Differ. Equ.* **16**(3), 299–333 (2003)
2. Ambrosio, L., Fusco, N., Pallara, D.: *Functions of Bounded Variation and Free Discontinuity Problems*. Clarendon, Oxford (2000)
3. Bae, E., Yuan, J., Tai, X.C.: Global minimization for continuous multiphase partitioning problems using a dual approach. *Int. J. Comput. Vis.* **92**, 112–129 (2011)
4. Boykov, Y., Veksler, O., Zabih, R.: Fast approximate energy minimization via graph cuts. *IEEE Trans. Pattern Anal. Mach. Intell.* **23**(11), 1222–1239 (2001)
5. Chambolle, A., Cremers, D., Pock, T.: A convex approach for computing minimal partitions. *Tech. Rep. 649*, Ecole Polytechnique CMAP (2008)
6. Chambolle, A., Darbon, J.: On total variation minimization and surface evolution using parametric maximum flows. *Int. J. Comput. Vis.* **84**, 288–307 (2009)
7. Chan, T.F., Esedoğlu, S., Nikolova, M.: Algorithms for finding global minimizers of image segmentation and denoising models. *J. Appl. Math.* **66**(5), 1632–1648 (2006)
8. Darbon, J., Sigelle, M.: Image restoration with discrete constrained total variation part I: fast and exact optimization. *J. Math. Imaging Vis.* **26**(3), 261–276 (2006)

9. Darbon, J., Sigelle, M.: Image restoration with discrete constrained variation part II: levelable functions, convex priors and non-convex cases. *J. Math. Imaging Vis.* **26**(3), 277–291 (2006)
10. Delaunoy, A., Fundana, K., Prados, E., Heyden, A.: Convex multi-region segmentation on manifolds. In: *Int. Conf. Comp. Vis* (2009)
11. Goldstein, T., Bresson, X., Osher, S.: Global minimization of Markov random field with applications to optical flow. *CAM Report 09-77*, UCLA (2009)
12. Kleinberg, J.M., Tardos, E.: Approximation algorithms for classification problems with pairwise relationships: metric labeling and Markov random fields. In: *Found. Comp. Sci.*, pp. 14–23 (1999)
13. Klodt, M., Schoenemann, T., Kolev, K., Schikora, M., Cremers, D.: An experimental comparison of discrete and continuous shape optimization methods. In: *Europ. Conf. Comp. Vis*, Marseille, France (2008)
14. Kolev, K., Klodt, M., Brox, T., Cremers, D.: Continuous global optimization in multiview 3d reconstruction. *Int. J. Comput. Vis.* **84**(1) (2009). doi:[10.1007/s11263-009-0233-1](https://doi.org/10.1007/s11263-009-0233-1)
15. Komodakis, N., Tziritas, G.: Approximate labeling via graph cuts based on linear programming. *IEEE Trans. Pattern Anal. Mach. Intell.* **29**(8), 1436–1453 (2007)
16. Lellmann, J., Becker, F., Schnörr, C.: Convex optimization for multi-class image labeling with a novel family of total variation based regularizers. In: *Int. Conf. Comp. Vis* (2009)
17. Lellmann, J., Kappes, J., Yuan, J., Becker, F., Schnörr, C.: Convex multi-class image labeling by simplex-constrained total variation. In: *Scale Space and Var. Meth. LNCS*, vol. 5567, pp. 150–162 (2009)
18. Lellmann, J., Lenzen, F., Schnörr, C.: Optimality bounds for a variational relaxation of the image partitioning problem. In: *Energy Min. Meth. Comp. Vis. Patt. Recogn.* (2011)
19. Lellmann, J., Schnörr, C.: Continuous multiclass labeling approaches and algorithms. *SIAM J. Imaging Sci.* (2011). doi:[10.1137/100805844](https://doi.org/10.1137/100805844)
20. Lysaker, M., Tai, X.C.: Iterative image restoration combining total variation minimization and a second-order functional. *Int. J. Comput. Vis.* **66**(1), 5–18 (2006)
21. Olsson, C.: *Global optimization in computer vision: convexity, cuts and approximation algorithms*. Ph.D. Thesis, Lund Univ. (2009)
22. Olsson, C., Byröd, M., Overgaard, N.C., Kahl, F.: Extending continuous cuts: anisotropic metrics and expansion moves. In: *Int. Conf. Comp. Vis* (2009)
23. Paragios, N., Chen, Y., Faugeras, O. (eds.): *The Handbook of Mathematical Models in Computer Vision*. Springer, Berlin (2006)
24. Pock, T., Cremers, D., Bischof, H., Chambolle, A.: Global solutions of variational models with convex regularization. *J. Imaging Sci.* **3**(4), 1122–1145 (2010)
25. Rockafellar, R.T., Wets, R.J.B.: *Variational Analysis*, 2nd edn. Springer, Berlin (2004)
26. Strandmark, P., Kahl, F., Overgaard, N.C.: Optimizing parametric total variation models. In: *Int. Conf. Comp. Vis.* (2009)
27. Trobin, W., Pock, T., Cremers, D., Bischof, H.: Continuous energy minimization by repeated binary fusion. In: *Europ. Conf. Comp. Vis.*, vol. 4, pp. 667–690 (2008)
28. Vazirani, V.V.: *Approximation Algorithms*. Springer, Berlin (2010)
29. Yuan, J., Bae, E., Tai, X.C., Boykov, Y.: A continuous max-flow approach to Potts model. In: *Europ. Conf. Comp. Vis*, pp. 379–392 (2010)
30. Zach, C., Gallup, D., Frahm, J.M., Niethammer, M.: Fast global labeling for real-time stereo using multiple plane sweeps. In: *Vis. Mod. Vis.* (2008)





**Jan Lellmann** received a diploma degree in computer science and a diploma degree in mathematics from the University of Karlsruhe in 2006 and 2007, and a doctoral degree in mathematics from the University of Heidelberg in 2011. He is currently working as a research associate in the image analysis group at the Department of Applied Mathematics and Theoretical Physics (DAMTP) of the University of Cambridge. His research interests include variational methods in image processing, convex relaxation strategies and related numerical optimization methods.



**Frank Lenzen** is a postdoctoral researcher at the Heidelberg Collaboratory for Image Processing (HCI) at the Heidelberg University in Germany. His research focuses on variational methods and optimization with applications in image processing and computer vision. He received a diploma in mathematics from the University of Bonn in 2002 and a Ph.D. in natural sciences from the University of Innsbruck, Austria, in 2006. He is coauthor of the book ‘Variational methods in imaging’.



**Christoph Schnörr** is full professor at the Faculty of Mathematics and Computer Science of the Heidelberg University, where he is heading the Image and Pattern Analysis Group (<http://ipa.iwr.uni-heidelberg.de/>) and co-directing an industry-on-campus center for image processing (<http://hci.iwr.uni-heidelberg.de/>). His research interests include image analysis and computer vision and related aspects of mathematical modelling and optimization.