Optimality Bounds for a Variational Relaxation of the Image Partitioning Problem

Jan Lellmann, Frank Lenzen, and Christoph Schnörr

Image and Pattern Analysis Group & HCI
Dept. of Mathematics and Computer Science, University of Heidelberg

Abstract. Variational relaxations can be used to compute approximate minimizers of optimal partitioning and multiclass labeling problems on continuous domains. While the resulting relaxed convex problem can be solved globally optimal, in order to obtain a discrete solution a rounding step is required, which may increase the objective and lead to suboptimal solutions. We analyze a probabilistic rounding method and prove that it allows to obtain discrete solutions with an *a priori* upper bound on the objective, ensuring the quality of the result from the viewpoint of optimization. We show that the approach can be interpreted as an approximate, multiclass variant of the coarea formula.

1 Introduction

In a series of papers [1–3], several authors have recently proposed convex relaxations of multiclass labeling problems in a variational framework. We consider the problem formulation

$$\inf_{u \in \mathcal{C}_{\mathcal{E}}} f(u), \quad f(u) := \int_{\Omega} \langle u(x), s(x) \rangle dx + \int_{\Omega} \Psi(Du)\,, \tag{1}$$
$$\mathcal{C}_{\mathcal{E}} := \{u \in \mathrm{BV}(\Omega)^l, u(x) \in \mathcal{E} := \{e^1, \ldots, e^l\} \text{ a.e.}\}, \quad \Omega = (0,1)^d,$$

for finding an optimal *labeling function* u that is of bounded variation [4]. Here e^i denotes the i-th unit vector representing the i-th label, $s \in L^1(\Omega), s \geqslant 0$ are the local costs representing the data term, and $\Psi : \mathbb{R}^{d \times l} \to \mathbb{R}_{\geqslant 0}$ is positively homogeneous, convex and continuous, and defines the regularizer. This formulation covers problems such as color segmentation, denoising, inpainting, depth from stereo and many more; see [5] for the definition of a class of regularizers Ψ relevant to various applications.

Problem (1) can also be seen as the problem of finding an *optimal partition* of Ω into l (not necessarily connected) sets $P_i := u^{-1}(e^i)$. It constitutes a hard problem due to the discrete decision at each point. However, formulation (1) permits a convenient relaxation to a convex problem:

$$\inf_{u \in \mathcal{C}} f(u), \quad \mathcal{C} := \{u \in \mathrm{BV}(\Omega)^l, u(x) \in \Delta_l \text{ a.e.}\}, \tag{2}$$

where $\Delta_l := \{x \in \mathbb{R}^l | x \geqslant 0, \sum_i x_i = 1\}$ is the convex hull of $\mathcal{E} := \{e^1, \ldots, e^l\}$, i.e. the l-dimensional unit simplex. Problem (2) is convex and can thus be solved globally optimal. However, the minimizer u^* of the relaxed problem may not lie in $\mathcal{C}_\mathcal{E}$, i.e. it is not necessarily *discrete*. In order to obtain a true partition of Ω, some rounding process is thus required to generate a discrete labeling $\bar{u}^*$. This may may increase the objective, and lead to a suboptimal solution of the original problem (1). While plausible deterministic methods exist [5], we are not aware of a method that allows to *bound the objective* of the *obtained* discrete solution $\bar{u}^*$ with respect to the objective of the (unknown) *optimal* discrete solution $u^*_\mathcal{E}$ in the spatially continuous setting for general regularizers.

Contribution. In this work, we consider a probabilistic rounding approach and derive a probabilistic bound of the form (see Thm. 2 below)

$$\mathbb{E}f(\bar{u}^*) \leqslant (1+\varepsilon)f(u^*_\mathcal{E}), \tag{3}$$

where $\bar{u}^*$ is the solution obtained by applying a custom probabilistic rounding method to the solution u^* of the convex relaxed problem (2), and $u^*_\mathcal{E}$ is the solution of the original partitioning problem (1). The approach is based on the work of Kleinberg and Tardos [6], who derive similar bounds in an LP relaxation framework. However their results are restricted in that they assume a grid discretization of the image domain and extensively make use of the fact that the number of grid points is finite. The bounds derived in Thm. 2 are compatible with their bounds, as well as the ones derived for the graph cut-based α-expansion in [7]. However, our results hold in the spatially continuous setting *without assuming a particular problem discretization.*

In the continuous setting, in [8] a similar bound was announced for the special case of the uniform metric. The approach is based on a continuous extension of the α-expansion method, which requires to solve a *sequence* of problems. In contrast, our approach only requires to solve a *single* convex problem, and provides valid bounds for a broad class of regularizers [5].

For an overview of generic approaches for solving integer problems using relaxation techniques we also refer to [9]. As these known approaches only apply in finite-dimensional spaces, deriving similar results for functions on continuous domains requires considerable additional mathematical work. Due to space restrictions we will only provide an outline of the proofs, and refer to an upcoming report for the technical details.

Notation. Superscripts v^i usually denote a collection of vectors or elements of a sequence, while subscripts v_k denote vector components. We denote $\mathbb{N} = \{1,2,\ldots\}$, $e = (1,\ldots,1)$. $\|\cdot\|_2$ is the usual Euclidean resp. the Frobenius norm, and $\mathcal{B}_r(x)$ denotes the ball of radius r in x. For a set $\mathcal{S}$, we define $1_\mathcal{S}(x) = 1$ iff $x \in \mathcal{S}$ and $1_\mathcal{S}(x) = 0$ otherwise.

Regarding measure-theoretic notations and functions of bounded variation we refer to [4]. In particular, we will use the d-dimensional Lebesgue measure $\mathcal{L}^d$, the k-dimensional Hausdorff measure $\mathcal{H}^k$, the distributional gradient Du and the total variation $\mathrm{TV}(u) = |Du|(\Omega)$. For some $\mathcal{L}^d$-measurable set $E \subseteq \Omega$, we denote its volume $|E| = \mathcal{L}^d(E)$, the measure-theoretic interior $(E)^1$ and exterior $(E)^0$, the reduced boundary $\mathcal{F}E$ with generalized inner normal ν_E, and the perimeter $\mathrm{Per}(E) = \mathrm{TV}(1_E)$. $Du \llcorner E$ is the restriction of Du to E, and $\Psi(Du)$ denotes the measure $\Psi(Du/|Du|)|Du|$, i.e. Ψ transforms the density of the measure Du with respect to its total variation measure $|Du|$. For $u \in \mathrm{BV}(\Omega)^l$ we denote by $\tilde{u}$ its approximate limit, and by $u^+_{\mathcal{F}E}$ and $u^-_{\mathcal{F}E}$ its one-sided limits [4, Thm. 3.77] on the reduced boundary of a set of finite perimeter E, i.e. $\mathrm{Per}(E) < \infty$.

2 Probabilistic Rounding and the Coarea Formula

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. [4]). Assume that $u' \in \mathrm{BV}(\Omega)$ and $u'(x) \in [0,1]$ for a.e. $x \in \Omega$, then the coarea formula states that its total variation can be represented by summing the boundary lengths of its superlevelsets:

$$\mathrm{TV}(u') = \int_0^1 \mathrm{TV}(1_{\{u'>\alpha\}})d\alpha. \tag{4}$$

The coarea formula provides a connection between problem (1) and the relaxation (2) in the two-class case, where $\mathcal{E} = \{e^1, e^2\}$, $u \in \mathcal{C}_{\mathcal{E}}$ and therefore $u_1 = 1 - u_2$: As noted in [5], $\mathrm{TV}(u) = \|e^1 - e^2\| \,\mathrm{TV}(u_1) = \sqrt{2}\,\mathrm{TV}(u_1)$, therefore the coarea formula (4) can be rewritten as

$$\mathrm{TV}(u) = \sqrt{2}\int_0^1 \mathrm{TV}(1_{\{u_1>\alpha\}})d\alpha = \int_0^1 \mathrm{TV}(e^1 1_{\{u_1>\alpha\}} + e^2 1_{\{u_1\leqslant\alpha\}})d\alpha \tag{5}$$

$$= \int_0^1 \mathrm{TV}(\bar{u}_\alpha)d\alpha\,, \quad \bar{u}_\alpha := e^1 1_{\{u_1>\alpha\}} + e^2 1_{\{u_1\leqslant\alpha\}} \tag{6}$$

Consequently, the total variation of u can be computed by taking the mean over the total variations of a set of *discrete* labelings $\{\bar{u}_\alpha \in \mathcal{C}_{\mathcal{E}} | \alpha \in [0,1]\}$, obtained by *rounding u at different thresholds α*. We now adopt a *probabilistic* view of (6): We regard the mapping

$$(u, \alpha) \in \mathcal{C} \times [0,1] \mapsto \bar{u}_\alpha \in \mathcal{C}_{\mathcal{E}} \quad (\text{for a.e. } \alpha \in [0,1]) \tag{7}$$

as a *parameterized, deterministic* rounding algorithm, that depends on u and on an additional parameter α. From this, we obtain a *probabilistic* (randomized) rounding algorithm by assuming α to be a uniformly distributed random variable. Under these assumptions, the coarea formula (6) can be written as

$$\mathrm{TV}(u) = \mathbb{E}_\alpha \,\mathrm{TV}(\bar{u}_\alpha). \tag{8}$$

This has the probabilistic interpretation 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 (2). A well-known implication is that if $u = u^*$, i.e. u minimizes (2), then almost every $\bar{u}_\alpha = \bar{u}^*_\alpha$ is a minimizer of (1) [10].

Unfortunately, property (8) is intrinsically restricted to the two-class case with TV regularizer. In the general case, one would hope to obtain a relation

$$f(u) = \int_\Gamma f(\bar{u}_\gamma)d\mu(\gamma) = \mathbb{E}_\gamma f(\bar{u}_\gamma) \tag{9}$$

Algorithm 1 Continuous Probabilistic Rounding

Require: $u \in \mathcal{C}$
1: $u^0 \leftarrow u$, $U^0 \leftarrow \Omega$, $c^0 \leftarrow (1,\ldots,1) \in \mathbb{R}^l$.
2: **for** $k = 1, 2, \ldots$ **do**
3: Randomly choose $\gamma^k := (i^k, \alpha^k)$ uniformly from $\{1,\ldots,l\} \times [0,1]$
4: $M^k \leftarrow U^{k-1} \cap \{x \in \Omega | u^{k-1}_{i^k}(x) > \alpha^k\}$
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: $c^k_j \leftarrow \begin{cases} \min\{c^{k-1}_j, \alpha^k\}, & \text{if } j = i^k, \\ c^{k-1}_j, & \text{otherwise.} \end{cases}$
8: **end for**

for some probability space (Γ, μ). For $l = 2$ and $\Psi(x) = \|\cdot\|_2$, (8) shows that (9) holds with $\gamma = \alpha$, $\Gamma = [0,1]$, μ the Lebesgue measure, and $\bar{u}_\gamma : \mathcal{C} \times \Gamma \to \mathcal{C}_\mathcal{E}$ as defined in (7).

In the multiclass case, the difficulty lies in providing a suitable probability space (Γ, μ) and parameterized rounding step $(u, \gamma) \mapsto \bar{u}_\gamma$. Unfortunately, obtaining a relation such as (8) 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 (2) and subsequent rounding.

In this work we will derive a bound of the form

$$(1+\varepsilon) f(u) \geqslant \int_\Gamma f(\bar{u}_\gamma) d\mu(\gamma) = \mathbb{E}_\gamma f(u_\gamma). \tag{10}$$

This can be seen as an *approximate* variant of the coarea formula. While (10) is not sufficient to provide a bound on $f(\bar{u}_\gamma)$ for *particular* γ, it permits a *probabilistic* bound in the sense of (3): For any minimizer u^* of the relaxed problem (2),

$$\mathbb{E}_\gamma f(\bar{u}^*_\gamma) \leqslant (1+\varepsilon) f(u^*) \leqslant (1+\varepsilon) f(u^*_\mathcal{E}), \tag{11}$$

holds, i.e. the ratio between the objective of the *rounded relaxed solution* and the *optimal discrete solution* is bounded – in a probabilistic sense – by $(1+\varepsilon)$.

In the following sections, we will construct a suitable parameterized rounding method and probability space in order to obtain an approximate coarea formula of the form (10).

3 Probabilistic Rounding for Multiclass Image Partitions

We consider the probabilistic rounding approach based on [6] as defined in Alg. 1. The algorithm proceeds in a number of phases. At each iteration, a label and a threshold $(i^k, \alpha^k) \in \Gamma' := \{1,\ldots,l\} \times [0,1]$ are randomly chosen (step 3), and label i^k is assigned to all yet unassigned points where $u^{k-1}_{i^k} > \alpha^k$ holds (step 5). In contrast to the two-class case considered above, the randomness is provided by a *sequence* (γ^k) of uniformly distributed random variables, i.e. $\Gamma = (\Gamma')^\mathbb{N}$.

After iteration k, all points in the set $U^k \subseteq \Omega$ have not yet been assigned a label, while all points in $\Omega \setminus U^k$ have been assigned a discrete label in iteration k or in a previous iteration. Iteration $k+1$ potentially modifies points only in the set U^k. The variable c_j^k stores the lowest threshold α chosen for label j up to and including iteration k.

For fixed input u, the algorithm can be seen as mapping a *sequence* of parameters (or instances of random variables) $\gamma = (\gamma^k) \in \Gamma$ into a *sequence* of states $(u_\gamma^k)_{k=1}^\infty$, $(U_\gamma^k)_{k=1}^\infty$ and $(c_\gamma^k)_{k=1}^\infty$. We drop the subscript γ if it does not create ambiguities.

In order to define the parameterized rounding step $(u, \gamma) \mapsto \bar{u}_\gamma$, we observe that, once $U_\gamma^{k'} = \emptyset$ occurs for some k', the sequence (u_γ^k) becomes stationary at $u_\gamma^{k'}$. In this case the algorithm may be terminated, with output $\bar{u}_\gamma := u_\gamma^{k'}$:

Definition 1. *Let $u \in \mathrm{BV}(\Omega)^l$, and denote $\Gamma := (\Gamma')^{\mathbb{N}}$. For some $\gamma \in \Gamma$, if $U_\gamma^{k'} = \emptyset$ for some $k' \in \mathbb{N}$, we denote $\bar{u}_\gamma := u_\gamma^{k'}$. For a functional $f : \mathrm{BV}(\Omega)^l \to \mathbb{R}$, define*

$$\begin{aligned} f(\bar{u}_{(\cdot)}) &: \Gamma \to \mathbb{R} \cup \{+\infty\} \\ \gamma \in \Gamma &\mapsto f(\bar{u}_\gamma) := \begin{cases} f(u_\gamma^{k'}), \; U_\gamma^{k'} = \emptyset \text{ and } u_\gamma^{k'} \in \mathrm{BV}(\Omega)^l, \\ +\infty, \quad \text{otherwise.} \end{cases} \end{aligned} \tag{12}$$

Denote by $f(\bar{u})$ the corresponding random variable induced by assuming γ to be uniformly distributed on Γ.

Note that $f(\bar{u}_\gamma)$ is well-defined: if $U_\gamma^{k'} = \emptyset$ for some (γ, k') then $u_\gamma^{k'} = u_\gamma^{k''}$ for all $k'' \geqslant k'$. In the remainder of this work, we will show that the expectation of $f(\bar{u}_\gamma)$ over all sequences γ can be bounded according to

$$\mathbb{E} f(\bar{u}) = \mathbb{E}_\gamma f(\bar{u}_\gamma) \leqslant (1+\varepsilon) f(\bar{u}) \tag{13}$$

for some $\varepsilon \geqslant 0$, cf. (10). Consequently, the rounding process may only increase the average objective in a controlled way.

We first show that almost surely Alg. 1 generates (in a finite number of iterations) a *discrete* labeling function $\bar{u}_\gamma \in \mathcal{C}_{\mathcal{E}}$.

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

$$\mathbb{P}(f(\bar{u}) < \infty) = 1. \tag{14}$$

Proof. Due to space restrictions we can only provide a sketch of the proof. The first part is to show that (u^k) becomes stationary almost surely, i.e.

$$\mathbb{P}(\exists k \in \mathbb{N} : U^k = \emptyset) = 1. \tag{15}$$

Define $n_j^k \in \mathbb{N}_0$ the number of $k' \in \{1, \ldots, k\}$ s.t. $i^{k'} = j$. Then the vector n^k is multinomially distributed, $n^k \sim \mathrm{Multinomial}(k; 1/l, \ldots, 1/l)$. Accordingly, the probability that all $c_j^k, j = 1, \ldots, l$ are smaller than $1/l$ is

$$\mathbb{P}(c^k < l^{-1} e) = \sum_{n_1^k + \ldots + n_l^k = k} \frac{k!}{n_1^k! \cdot \ldots \cdot n_l^k!} \left(\frac{1}{l}\right)^k \prod_{j=1}^{l} \left(1 - \left(1 - \frac{1}{l}\right)^{n_j^k}\right), \tag{16}$$

which can be shown to converge to 1 for $k \to \infty$. Since $u(x) \in \Delta_l$, the condition $c^k < l^{-1}e$ implies $U^k = \emptyset$. Therefore (15) follows from

$$1 \geqslant \mathbb{P}(\exists k \in \mathbb{N} : e^\top c^k < 1) \geqslant \mathbb{P}(c^k < l^{-1}e) \overset{k\to\infty}{\to} 1. \tag{17}$$

The second part of the proof consists in showing that almost surely all iterates u^k are contained in $\mathrm{BV}(\Omega)^l$, for which it suffices to show that

$$\mathbb{P}\left(u^k \in \mathrm{BV}(\Omega)^l \ \forall k \in \mathbb{N}\right) = 1. \tag{18}$$

This can be seen using induction to show that $u^k \in \mathrm{BV}(\Omega)^l$ and $\mathrm{Per}(U^k) < \infty$ almost surely for all $k \in \mathbb{N}_0$. For $u^{k-1} \in \mathrm{BV}(\Omega)^l$, by [4, Thm. 3.40] it holds that $\mathrm{Per}(\{x \in \Omega | u^{k-1}_{i^k}(x) \leqslant \alpha^k\}) < \infty$ for $\mathcal{L}^1$-a.e. $\alpha^k \in [0,1]$ (and all i^k), therefore

$$\mathbb{P}(\mathrm{Per}(U^k) < \infty | \, \mathrm{Per}(U^{k-1}) < \infty) = 1. \tag{19}$$

The statement for u^k follows, since for the same reason $\mathrm{Per}(M^k) < \infty$ almost surely (cf. Alg. 1 for the definition of M^k), and [4, Thm. 3.84] ensures

$$\mathrm{Per}(M^k) < \infty, u^{k-1} \in \mathrm{BV}(\Omega)^l \Rightarrow u^k = e^{i^k} 1_{M^k} + u^{k-1} 1_{\Omega \setminus M^k} \in \mathrm{BV}(\Omega)^l. \tag{20}$$

□

4 A Probabilistic A Priori Optimality Bound

In the previous sections we have shown that the rounding process induced by Alg. 1 is well-defined in the sense that it returns a discrete solution $\bar{u}_\gamma \in \mathrm{BV}(\Omega)^l$ almost surely. We now return to proving an upper bound for the expectation of $f(\bar{u})$ as in the approximate coarea formula (3).

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

Proposition 1. *The sequence* (u^k) *generated by Alg. 1 satisfies*

$$\mathbb{E}(\langle u^k, s\rangle) = \langle u, s\rangle \quad \forall k \in \mathbb{N}. \tag{21}$$

Proof. In Alg. 1, instead of step 5 we consider the update

$$u^k \leftarrow e^{i^k} 1_{\{u^{k-1}_{i^k} > \alpha^k\}} + u^{k-1} 1_{\{u^{k-1}_{i^k} \leqslant \alpha^k\}}, \tag{22}$$

which yields exactly the same iterates. Denote $\gamma' := (\gamma^1, \ldots, \gamma^{k-1})$ and $u^{\gamma'} := u^{k-1}_\gamma$. We use an induction argument on k: For $k \geqslant 1$,

$$\begin{aligned} \mathbb{E}_\gamma \langle u^k_\gamma, s\rangle &= \mathbb{E}_{\gamma'} \frac{1}{l} \sum_{i=1}^{l} \int_0^1 \sum_{j=1}^{l} s_j \cdot \left(e^i 1_{\{u^{\gamma'}_i > \alpha\}} + u^{\gamma'} 1_{\{u^{\gamma'}_i \leqslant \alpha\}}\right)_j d\alpha \qquad (23) \\ &= \mathbb{E}_{\gamma'} \frac{1}{l} \sum_{i=1}^{l} \int_0^1 \left(s_i \cdot 1_{\{u^{\gamma'}_i > \alpha\}} + \left(1 - 1_{\{u^{\gamma'}_i > \alpha\}}\right) \langle u^{\gamma'}, s\rangle\right) d\alpha. \end{aligned}$$

Now we take into account the relation [4, Prop. 1.78],

$$\int_0^1 \int_\Omega s_i(x) \cdot 1_{u_i > \alpha}(x) dx d\alpha = \int_\Omega s_i(x) u_i(x) dx = \langle u_i, s_i \rangle. \tag{24}$$

This leads to

$$\begin{aligned} \mathbb{E}_\gamma \langle u_\gamma^k, s \rangle \quad &= \quad \mathbb{E}_{\gamma'} \frac{1}{l} \sum_{i=1}^{l} \left(s_i u_i^{\gamma'} + \langle u^{\gamma'}, s \rangle - u_i^{\gamma'} \langle u^{\gamma'}, s \rangle \right) d\alpha \\ &\overset{u^{\gamma'}(x) \in \Delta_l}{=} \mathbb{E}_{\gamma'} \langle u^{\gamma'}, s \rangle \quad = \quad \mathbb{E}_\gamma \langle u_\gamma^{k-1}, s \rangle. \end{aligned} \tag{25}$$

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

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

$$U_{\gamma^k} := \{x \in \Omega | u_{i^k}(x) \leqslant \alpha^k\}, \quad V_{\gamma^k} := \left(U_{\gamma^k}\right)^1, \quad V^k := (U^k)^1. \tag{26}$$

As the measure-theoretic interior is invariant under $\mathcal{L}^d$-negligible modifications, given some fixed sequence γ 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 $L^1(\Omega)^l$.

We use (without proof) the fact that the measure-theoretic interior satisfies $(E \cap F)^1 = (E)^1 \cap (F)^1$ for any $\mathcal{L}^d$-measurable sets E, F. Some calculations yield

$$U^k = U_{\gamma^1} \cap \ldots \cap U_{\gamma^k}, \quad V^k = V_{\gamma^1} \cap \ldots \cap V_{\gamma^k} \quad (k \geqslant 1), \tag{27}$$

$$U^{k-1} \setminus U^k = U_{\gamma^1} \cap \left(\left(U_{\gamma^2} \cap \ldots \cap U_{\gamma^{k-1}} \right) \setminus \left(U_{\gamma^2} \cap \ldots \cap U_{\gamma^k} \right) \right) \quad (k \geqslant 2),$$

$$V^{k-1} \setminus V^k = V_{\gamma^1} \cap \left(\left(V_{\gamma^2} \cap \ldots \cap V_{\gamma^{k-1}} \right) \setminus \left(V_{\gamma^2} \cap \ldots \cap V_{\gamma^k} \right) \right) \quad (k \geqslant 2), \tag{28}$$

$$\Omega \setminus V^k = \bigcup_{k'=1}^{k} \left(V^{k'-1} \setminus V^{k'} \right) \quad (k \geqslant 1). \tag{29}$$

Moreover (again without proof), since V^k is the measure-theoretic interior of U^k, both sets are equal up to an $\mathcal{L}^d$-negligible set.

We now prepare for an induction argument on the expectation of the regularizing term when restricted to the sets $V^{k-1} \setminus V^k$. We first state an intermediate result required for the proofs.

Proposition 2. *Let $u, v \in \mathcal{C}$, $\Psi \leqslant \rho_u \| \cdot \|_2$, and $E \subseteq \Omega$ s.t. $\mathrm{Per}(E) < \infty$. Then $w := u 1_E + v 1_{\Omega \setminus E} \in \mathrm{BV}(\Omega)^l$,*

$$Dw = Du \llcorner (E)^1 + Dv \llcorner (E)^0 + \nu_E \left(u_{\mathcal{F}E}^+ - v_{\mathcal{F}E}^- \right)^\top \mathcal{H}^{d-1} \llcorner (\mathcal{F}E \cap \Omega), \tag{30}$$

and, for some Borel set $A \subseteq \Omega$,

$$\int_A \Psi(Dw) \leqslant \sqrt{2} \rho_u \, \mathrm{Per}(E) + \int_{A \cap (E)^1} \Psi(Du) + \int_{A \cap (E)^0} \Psi(Dv). \tag{31}$$

Proof. We omit the details of the proof due to space restrictions. It relies on [4, Thm. 3.84], [4, Prop. 2.37] and the fact that

$$\begin{aligned}\int_{A\cap\mathcal{F}E\cap\Omega}\Psi(Dw) &= \int_{A\cap\mathcal{F}E\cap\Omega}\Psi(\nu_E(w^+_{\mathcal{F}E}(x)-w^-_{\mathcal{F}E}(x))^\top)d\mathcal{H}^{d-1} \\ &\leqslant \int_{A\cap\mathcal{F}E\cap\Omega}\rho_u\|\nu_E(w^+_{\mathcal{F}E}(x)-w^-_{\mathcal{F}E}(x))^\top\|_2 d\mathcal{H}^{d-1} \\ &\leqslant \sqrt{2}\rho_u\operatorname{Per}(E).\end{aligned} \tag{32}$$

□

The following proposition provides the initial step for $k=1$.

Proposition 3. *Let* $\rho_l\|\cdot\|_2 \leqslant \Psi \leqslant \rho_u\|\cdot\|_2$. *Then*

$$\mathbb{E}\int_{V^0\setminus V^1}\Psi(D\bar{u}) \leqslant \frac{2}{l}\frac{\rho_u}{\rho_l}\int_\Omega \Psi(Du). \tag{33}$$

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

$$\bar{u}_\gamma = 1_{V_{(i,\alpha)}}e^i + 1_{\Omega\setminus V_{(i,\alpha)}}\bar{u}_\gamma \quad \mathcal{L}^d - a.e. \tag{34}$$

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

$$\int_{V^0\setminus V^1}\Psi(D\bar{u}_\gamma) = \int_{\Omega\setminus V_{(i,\alpha)}}\Psi(D\bar{u}_\gamma) = \int_{\Omega\setminus V_{(i,\alpha)}}\Psi\left(D\left(1_{V_{(i,\alpha)}}e^i + 1_{\Omega\setminus V_{(i,\alpha)}}\bar{u}_\gamma\right)\right).$$

Since $u\in \mathrm{BV}(\Omega)^l$, we know that $\operatorname{Per}(V_{(i,\alpha)})<\infty$ holds for $\mathcal{L}^1$-a.e. α and any i [4, Thm. 3.40]. Therefore we conclude from Prop. 2 that (for $\mathcal{L}^1$-a.e. α),

$$\begin{aligned}&\int_{\Omega\setminus V_{(i,\alpha)}}\Psi(D\bar{u}_\gamma) \leqslant \rho_l\sqrt{2}\operatorname{Per}\left(V_{(i,\alpha)}\right) + \\ &\int_{\left(\Omega\setminus V_{(i,\alpha)}\right)\cap\left(\Omega\setminus V_{(i,\alpha)}\right)^1}\Psi\left(De^i\right) + \int_{\left(\Omega\setminus V_{(i,\alpha)}\right)\cap\left(\Omega\setminus V_{(i,\alpha)}\right)^0}\Psi\left(D\bar{u}_\gamma\right).\end{aligned} \tag{35}$$

Both of the integrals are zero, since $De^i=0$ and $(\Omega\setminus V_{(i,\alpha)})^0 = (V_{(i,\alpha)})^1 = V_{(i,\alpha)}$, therefore $\int_{\Omega\setminus V_{(i,\alpha)}}\Psi(D\bar{u}_\gamma)\leqslant \rho_l\sqrt{2}\operatorname{Per}(V_{(i,\alpha)})$. This implies

$$\mathbb{E}_\gamma\int_{\Omega\setminus V_{(i,\alpha)}}\Psi(D\bar{u}_\gamma) \leqslant \frac{1}{l}\sum_{i=1}^{l}\int_0^1 \rho_u\sqrt{2}\operatorname{Per}(V_{(i,\alpha)})d\alpha. \tag{36}$$

Also, $\operatorname{Per}(V_{(i,\alpha)})=\operatorname{Per}(U_{(i,\alpha)})$ since the perimeter is invariant under $\mathcal{L}^d$-negligible modifications. The assertion then follows using the coarea formula [4, Thm. 3.40]:

$$\mathbb{E}_\gamma\int_{V^0\setminus V^1}\Psi(D\bar{u}_\gamma) \leqslant \frac{1}{l}\sum_{i=1}^{l}\int_0^1 \rho_u\sqrt{2}\operatorname{Per}(U_{(i,\alpha)})d\alpha \tag{37}$$

$$\overset{\text{coarea}}{=} \frac{\sqrt{2}}{l}\rho_u\sum_{i=1}^{l}\operatorname{TV}(u_i) \leqslant \frac{2}{l}\rho_u\int_\Omega\|Du\|_2 \leqslant \frac{2}{l}\frac{\rho_u}{\rho_l}\int_\Omega\Psi(Du).$$

□

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

Proposition 4. *Let $\Psi \leqslant \rho_u \|\cdot\|_2$. Then, for any $k \geqslant 2$,*

$$F := \mathbb{E} \int_{V^{k-1} \setminus V^k} \Psi(D\bar{u}) \leqslant \frac{(l-1)}{l} \mathbb{E} \int_{V^{k-2} \setminus V^{k-1}} \Psi(D\bar{u}). \tag{38}$$

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

$$W_{\gamma'} := V_{\gamma'}^{k-2} \setminus V_{\gamma'}^{k-1} = \left(V_{\gamma^2} \cap \ldots \cap V_{\gamma^{k-1}}\right) \setminus \left(V_{\gamma^2} \cap \ldots \cap V_{\gamma^k}\right). \tag{39}$$

By Prop. 1 we may assume that, under the expectation, $\bar{u}_\gamma$ exists and is an element of $\mathrm{BV}(\Omega)^l$. We denote $\gamma^1 = (i, \alpha)$, then $V^{k-1} \setminus V^k = V_{(i,\alpha)} \cap W_{\gamma'}$ due to (28), and

$$F = \frac{1}{l} \sum_{i=1}^{l} \int_0^1 \left(\mathbb{E}_{\gamma'} \int_{V_{(i,\alpha)} \cap W_{\gamma'}} \Psi(D\bar{u}_{((i,\alpha),\gamma')}) \right) d\alpha. \tag{40}$$

We now use (without proof) the fact that if two functions v, w in $\mathrm{BV}(\Omega)$ coincide (in L^1, i.e. $\mathcal{L}^d$-a.e.) on a set E with $\mathrm{Per}(E) < \infty$, then the measures $\Psi(Dv) \llcorner (E)^1$ and $\Psi(Dw) \llcorner (E)^1$ coincide. In particular, 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 substitute $D\bar{u}_{((i,\alpha),\gamma')}$ by $D\bar{u}_{\gamma'}$ in (40):

$$F = \frac{1}{l} \sum_{i=1}^{l} \int_0^1 \left(\mathbb{E}_{\gamma'} \int_{W_{\gamma'}} 1_{V_{(i,\alpha)}} \Psi(D\bar{u}_{\gamma'}) \right) d\alpha. \tag{41}$$

By definition of the measure-theoretic interior, $1_{V_{(i,\alpha)}}$ is bounded from above by the density function of $U_{(i,\alpha)}$, $\Theta_{U_{(i,\alpha)}}(x) := \lim_{\delta \searrow 0} |\mathcal{B}_\delta(x) \cap U_{(i,\alpha)}| / |\mathcal{B}_\delta(x)|$ [4, Def. 2.55], which exists $\mathcal{H}^{d-1}$-a.e. on Ω by [4, Thm. 3.61]. Therefore, denoting by $\mathcal{B}_\delta(\cdot)$ the mapping $x \in \Omega \mapsto \mathcal{B}_\delta(x)$,

$$F \leqslant \frac{1}{l} \sum_{i=1}^{l} \int_0^1 \left(\mathbb{E}_{\gamma'} \int_{W_{\gamma'}} \left(\lim_{\delta \searrow 0} \frac{|\mathcal{B}_\delta(\cdot) \cap U_{(i,\alpha)}|}{|\mathcal{B}_\delta(\cdot)|} \right) \Psi(D\bar{u}_{\gamma'}) \right) d\alpha. \tag{42}$$

Rearranging the integrals and the limit, which can be justified by dominated convergence using $\Psi \leqslant \rho_u \|\cdot\|_2$ and $\mathrm{TV}(\bar{u}_{\gamma'}) < \infty$ almost surely, we get

$$\begin{aligned} F &\leqslant \frac{1}{l} \mathbb{E}_{\gamma'} \lim_{\delta \searrow 0} \int_{W_{\gamma'}} \left(\sum_{i=1}^{l} \int_0^1 \left(\frac{|\mathcal{B}_\delta(\cdot) \cap U_{(i,\alpha)}|}{|\mathcal{B}_\delta(\cdot)|} \right) d\alpha \right) \Psi(D\bar{u}_{\gamma'}) \\ &= \frac{1}{l} \mathbb{E}_{\gamma'} \lim_{\delta \searrow 0} \int_{W_{\gamma'}} \frac{1}{|\mathcal{B}_\delta(\cdot)|} \left(\sum_{i=1}^{l} \int_0^1 \int_{\mathcal{B}_\delta(\cdot)} 1_{\{u_i(y) \leqslant \alpha\}} dy d\alpha \right) \Psi(D\bar{u}_{\gamma'}). \end{aligned} \tag{43}$$

We again apply [4, Prop. 1.78] to the two innermost integrals, which leads to

$$F \leqslant \frac{1}{l} \mathbb{E}_{\gamma'} \lim_{\delta \searrow 0} \int_{W_{\gamma'}} \frac{1}{|\mathcal{B}_\delta(\cdot)|} \left(\sum_{i=1}^{l} \int_{\mathcal{B}_\delta(\cdot)} (1 - u_i(y)) dy \right) \Psi(D\bar{u}_{\gamma'}) \,. \tag{44}$$

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

$$F \leqslant \frac{l-1}{l}\mathbb{E}_{\gamma'} \int_{W_{\gamma'}} \Psi(D\bar{u}_{\gamma'}) = \frac{l-1}{l}\mathbb{E}_{\gamma'} \int_{V_{\gamma'}^{k-2} \setminus V_{\gamma'}^{k-1}} \Psi(D\bar{u}_{\gamma'})\,. \tag{45}$$

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

$$F \leqslant \frac{l-1}{l}\mathbb{E}_{\gamma} \int_{V_{\gamma}^{k-1} \setminus V_{\gamma}^{k}} \Psi(D\bar{u}_{\gamma})\,. \tag{46}$$

The following theorem is the main result of this work, and provides an approximate coarea formula as in (10).

Theorem 2. *Let $s : \Omega \to [0,\infty)$ s.t. $s \in L^1(\Omega)^l$, $\Psi : \mathbb{R}^{d\times l} \to \mathbb{R}_{\geqslant 0}$ positively homogeneous, convex and continuous with $\rho_l \|z\|_2 \leqslant \Psi(z) \leqslant \rho_u \|z\|_2$ $\forall z \in \mathbb{R}^{d\times l}$, and $u \in \mathcal{C}$. Then Alg. 1. generates a discrete labeling $\bar{u} \in \mathcal{C}_\mathcal{E}$ almost surely, and*

$$\mathbb{E} f(\bar{u}) \leqslant 2\frac{\rho_u}{\rho_l} f(u). \tag{47}$$

Proof. The first part follows from Thm. 1. Therefore there almost surely exists $k' := k'(\gamma) \geqslant 1$ s.t. $U^{k'} = \emptyset$ and $\bar{u}_\gamma = u_\gamma^{k'}$. The stationarity implies

$$\langle \bar{u}_\gamma, s\rangle = \langle u_\gamma^{k'}, s\rangle = \lim_{k\to\infty} \langle u_\gamma^k, s\rangle \text{ and } \Omega = \bigcup_{k=1}^{\infty} \left(V^{k-1} \setminus V^k\right) \tag{48}$$

almost surely (cf. (29)). Thus

$$\mathbb{E}_\gamma f(\bar{u}_\gamma) = \mathbb{E}_\gamma \left(\lim_{k\to\infty} \langle u_\gamma^k, s\rangle \right) + \mathbb{E}_\gamma \left(\sum_{k=1}^{\infty} \int_{V^{k-1}\setminus V^k} \Psi(D\bar{u}_\gamma) \right) \tag{49}$$

$$= \lim_{k\to\infty} \left(\mathbb{E}_\gamma \langle u_\gamma^k, s\rangle\right) + \sum_{k=1}^{\infty} \mathbb{E}_\gamma \int_{V^{k-1}\setminus V^k} \Psi(D\bar{u}_\gamma) \tag{50}$$

The first term is equal to $\langle u, s\rangle$ due to Prop. 1. An induction argument using Prop. 3 and Prop. 4 shows

$$\int_{V^{k-1}\setminus V^k} \Psi(D\bar{u}_\gamma) \leqslant \sum_{k=1}^{\infty} \left(\frac{l-1}{l}\right)^{k-1} \frac{2}{l}\frac{\rho_u}{\rho_l} \int_\Omega \Psi(Du) = 2\frac{\rho_u}{\rho_l} \int_\Omega \Psi(Du)\,, \tag{51}$$

therefore

$$\mathbb{E}_\gamma f(\bar{u}_\gamma) \leqslant \langle u, s\rangle + 2\frac{\rho_u}{\rho_l} \int_\Omega \Psi(Du)\,. \tag{52}$$

Since $s \geqslant 0, \rho_u \geqslant \rho_l$ and therefore $\langle u, s\rangle \leqslant 2(\rho_u/\rho_l)\langle u, s\rangle$, this proves the assertion (47). Swapping the integral and limits in (50) can be justified retrospectively by the dominated convergence theorem, using $0 \leqslant \langle u, s\rangle \leqslant \infty$ and $\int_\Omega \Psi(Du) \leqslant \rho_u \operatorname{TV}(u) < \infty$. □

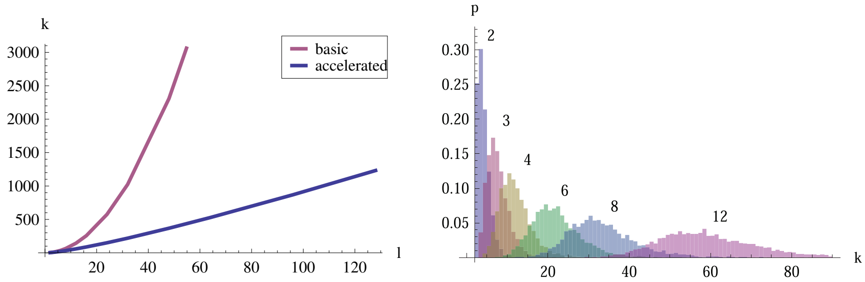

Fig. 1. Left: Label count l vs. mean number of iterations k of the probabilistic rounding algorithm. The improved sampling of α^k greatly accelerates the method. Empirically, $k \approx 2l\ln(l)$ for the accelerated method. As a result, runtime is comparable to the deterministic rounding methods. **Right:** Histogram (probability density scale) of the number of iterations k over 5000 runs for $2-12$ labels.

Corollary 1. *Under the conditions of Thm. 2, if u^* minimizes f over $\mathcal{C}$, $u^*_{\mathcal{E}}$ minimizes f over $\mathcal{C}_{\mathcal{E}}$, and $\bar{u}^*$ denotes the output of Alg. 1 applied to u^*, then*

$$\mathbb{E} f\left(\bar{u}^*\right) \leqslant 2\frac{\rho_u}{\rho_l} f(u^*_{\mathcal{E}}). \tag{53}$$

Proof. This follows immediately from Thm. 2 using $f(u^*) \leqslant f(u^*_{\mathcal{E}})$, cf. (11). □

We have demonstrated that the proposed approach allows to recover, from the solution u^* of the convex *relaxed* problem (2), an approximate *discrete* solution $\bar{u}^*$ of the nonconvex *original* problem (1), with an upper bound on the objective.

The bound in (53) is of the same order as the known bounds for finite-dimensional metric labeling [6] and α-expansion [7], however it extends these results to problems on continuous domains for a broad class of regularizers [5].

5 Experiments

Although the main purpose of Alg. 1 is to provide a basis for deriving the bound in Thm. 2, we will briefly point out some of its empirical characteristics.

Expected number of iterations. In practice, choosing $\alpha^k \in [0,1]$ leads to an unnecessary large number of iterations, as no point is assigned a label in iteration k unless $\alpha^k < c^{k-1}_{i^k}$. The method can be accelerated without affecting the derived bounds by choosing $\alpha^k \in [0, c^{k-1}_{i^k}]$ instead, thereby skipping the redundant iterations.

Fig. 1 shows the mean number of iterations k until $e^\top c^k < 1$, over 5000 runs per label count. From the proof of Thm. 1 it can be seen that this provides a worst-case upper bound for the expected number of iterations until $\bar{u}_\gamma$ is obtained. For the accelerated method, k is almost perfectly proportional to $l\ln(l)$; we conjecture that asymptotically $k = 2l\ln(l)$.

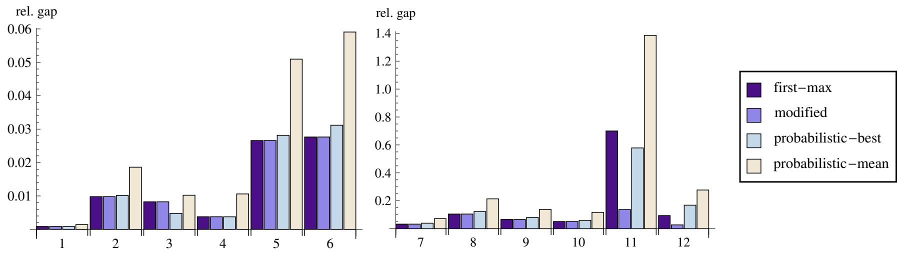

Fig. 2. Relative gap (*a posteriori* bound) ε' of the rounded solution for the test problems using deterministic "first-max" and "modified" rounding [5], and best and mean gap obtained using the proposed probabilistic method. While the energy increase through probabilistic rounding is usually slightly larger than for the deterministic methods, it is well below the *a priori* bound of $\varepsilon = 2\rho_u/\rho_l - 1$ derived in Cor. 1 (Table 1).

Optimality. In order to evaluate the tightness of the bound (53) in Thm. 2 in practice, we selected 12 prototypical multiclass labeling problems with $3-64$ labels each. For each we computed the relaxed solution u^* and the mean as well as the best objective of the rounded solution $\bar{u}^*$ during 10000 iterations of Alg. 1.

The employed primal-dual optimization approach provides a lower bound $f_D(v^*) \leqslant f(u^*)$ via the dual objective f_D and a dual feasible point v^*. This allows to compute the *relative gap* $\varepsilon' := (f(\bar{u}^*) - f_D(v^*))/f_D(v^*)$, which provides an *a posteriori* upper bound for the optimality w.r.t. the *discrete* solution $u^*_{\mathcal{E}}$,

$$\frac{f(\bar{u}^*) - f(u^*_{\mathcal{E}})}{f(u^*_{\mathcal{E}})} \leqslant \varepsilon', \tag{54}$$

in contrast to the theoretical, *a priori* upper bound $\varepsilon = 2\rho_u/\rho_l - 1$ derived from Cor. 1. In practice, the *a posteriori* bound stayed well below the theoretical bound (Table 1), which is consistent with the good practical performance of the α-expansion method that has a similar *a priori* bound.

Relative Performance. We compared the probabilistic approach to two deterministic rounding methods: While the "first-max" method assigns to each point the first label i s.t. $u_i(x) = \max_j u_j(x)$, the "modified" method [5] chooses the unit vector e^i that is closest to $u(x)$ with respect to a norm defined by Ψ. Compared to these methods, Alg. 1 usually leads to a slightly larger energy increase (Fig. 2). For problems 11 and 12, where ρ_u/ρ_l is large, the solution is clearly inferior to the one obtained using the "modified" rounding. This can be attributed to the fact that the latter takes into account the detailed structure of Ψ, which is neither required nor used in order to obtain the bounds in Thm. 2.

However, for problems that are inherently difficult for convex relaxation approaches, we found that the probabilistic approach often generated better solutions. An example is the "inverse triple junction" inpainting problem (second row in Fig. 3), which has at least 3 distinct discrete solutions. A variant of this problem, formulated on graphs, was used as a worst-case example to show the tightness of the LP relaxation bound in [6].

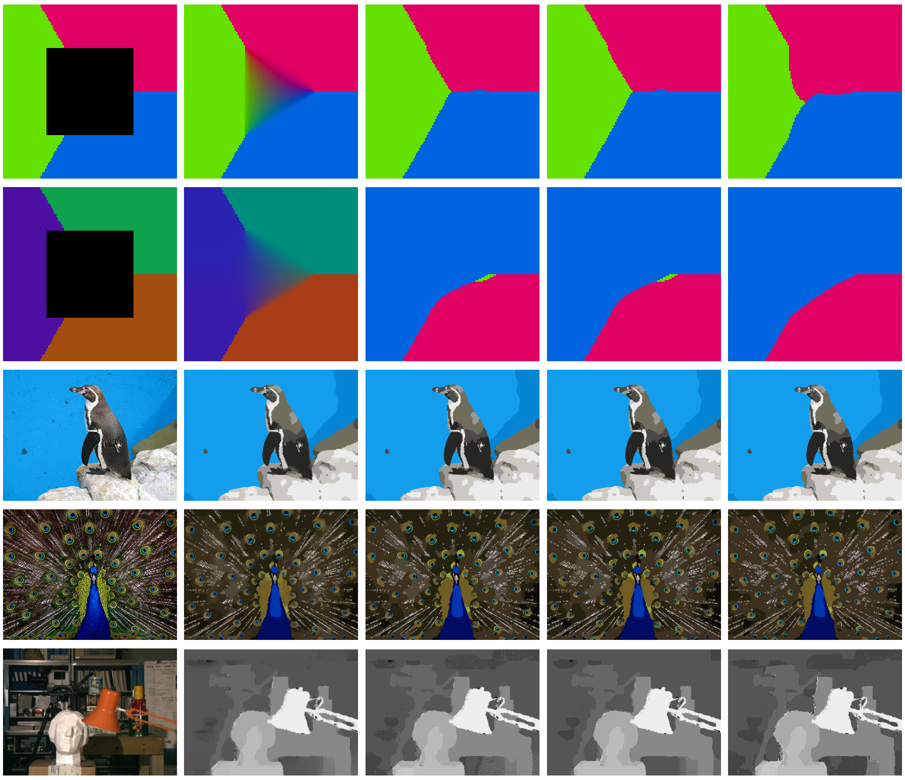

Fig. 3. Top to bottom: Problems 2,3,5,8,11 of the test set. **Left to right:** Input, relaxed solution, discrete solutions obtained by deterministic "first-max" and "modified" rounding [5], result of the probabilistic rounding. In specially crafted situations, the probabilistic method may perform slightly worse (first row) or better (second row). On real-world data, results are very similar (rows 3–5). In contrast to the deterministic approaches, the proposed method provides true *a priori* optimality bounds.

We would like to emphasize that the purpose of these experiments is not to demonstrate a practical superiority of the proposed method compared to other techniques, but rather to provide an illustration on what bounds can be *expected in practice* compared to the *a priori* bounds in Thm. 2.

6 Conclusion

We presented a probabilistic rounding method for recovering approximate solutions of multiclass labeling or image partitioning problems from solutions of convex relaxations. To our knowledge, this is the first fully convex approach that is both formulated in the spatially continuous setting and provides an *a priori* bound on the optimality of the generated discrete solution. We showed that the approach can also be interpreted as an approximate variant of the coarea formula. Numerical experiments confirm the theoretical bounds.

problem	1	2	3	4	5	6	7	8	9	10	11	12
N	76800	14400	14400	129240	76800	86400	86400	76800	86400	76800	110592	21838
l	3	3	3	4	8	12	12	12	12	12	16	64
k	7.1	6.9	5.0	11.0	27.2	47.5	47.0	43.6	46.5	46.0	70.7	335.0
bound	1.	1.	1.	1.	1.	1.	1.	1.	1.	1.	253.275	375.836
rel. gap	0.0014	0.0186	0.0102	0.0106	0.0510	0.0591	0.0722	0.2140	0.1382	0.1172	1.4072	0.2772

Table 1. Number of pixels N, number of labels l, mean number of iterations k, predicted *a priori* bound $\varepsilon = 2\rho_u/\rho_l - 1$, and mean relative gap (*a posteriori* bound) ε'. The *a posteriori* bound is well below the bound predicted by Thm. 2. Problems $1 - 10$ are color segmentation/inpainting problems with $\Psi = \| \cdot \|_2$. The depth-from-stereo resp. inpainting problems 11 and 12 use an approximated cut-linear metric as in [5].

Future work may include extending the results to non-homogeneous regularizers, and improving the tightness of the bound. Also, the connection to recent convex relaxation techniques [11, 12] for solving nonconvex variational problems should be further explored.

References

1. 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)
2. 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. Volume 5567 of LNCS. (2009) 150–162
3. Pock, T., Chambolle, A., Cremers, D., Bischof, H.: A convex relaxation approach for computing minimal partitions. In: Comp. Vis. Patt. Recogn. (2009)
4. Ambrosio, L., Fusco, N., Pallara, D.: Functions of Bounded Variation and Free Discontinuity Problems. Clarendon Press (2000)
5. 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)
6. Kleinberg, J.M., Tardos, E.: Approximation algorithms for classification problems with pairwise relationships: Metric labeling and Markov random fields. In: Found. Comp. Sci. (1999) 14–23
7. Boykov, Y., Veksler, O., Zabih, R.: Fast approximate energy minimization via graph cuts. Patt. Anal. Mach. Intell. **23** (2001) 1222–1239
8. Olsson, C., Byröd, M., Overgaard, N.C., Kahl, F.: Extending continuous cuts: Anisotropic metrics and expansion moves. In: Int. Conf. Comp. Vis. (2009)
9. Bertsimas, D., Weismantel, R.: Optimization over Integers. Dynamic Ideas (2005)
10. Chan, T.F., Esedoḡlu, S., Nikolova, M.: Algorithms for finding global minimizers of image segmentation and denoising models. J. Appl. Math. **66** (2006) 1632–1648
11. Alberti, G., Bouchitté, G., Dal Maso, G.: The calibration method for the Mumford-Shah functional and free-discontinuity problems. Calc. Var. Part. Diff. Eq. **16** (2003) 299–333
12. Pock, T., Cremers, D., Bischof, H., Chambolle, A.: Global solutions of variational models with convex regularization. J. Imaging Sci. **3** (2010) 1122–1145