

A MINIMAL SURFACE CRITERION FOR GRAPH PARTITIONING

DOMINIQUE ZOSSO

University of California, Los Angeles, Department of Mathematics
520 Portola Plaza, Box 951555, Los Angeles, CA 90095-1555, USA

BRAXTON OSTING

University of Utah, Salt Lake City, Department of Mathematics
155 S. 1400 E, Rm 233, Salt Lake City, UT 84112, USA

(Communicated by the associate editor name)

ABSTRACT. We consider a geometric approach to graph partitioning based on the graph Beltrami energy, a discrete version of a functional that appears in classical minimal surface problems. More specifically, the optimality criterion is given by the sum of the minimal Beltrami energies of the partition components. Since the Beltrami energy interpolates between the Total Variation and Dirichlet energies, various results for optimal partitions for these two energies can be recovered. We adapt primal-dual convex optimization methods to solve for the minimal Beltrami energy for each component of a given partition. A rearrangement algorithm is proposed to find the graph partition to minimize a relaxed version of the objective. The method is applied to several clustering problems on graphs constructed from manifold discretizations, synthetic data, the MNIST handwritten digit dataset, and image segmentation. The model has a semisupervised extension and provides a natural representative for the clusters as well.

1. Introduction. Given a graph $G = (V, E)$ with non-negative edge weights $\{w_{ij}\}_{(i,j) \in E}$, we consider the problem of “optimally” partitioning the vertex set, V , into k subsets. The challenges to partitioning are that datasets are typically large-scale and high-dimensional, common mathematical formulations of the problem lead to NP-hard problems, and the criterion of optimality is application dependent. Here we propose a model for graph partitioning based on an extension of the Beltrami energy to graphs. To motivate this model, we consider the following isoperimetric problems.

1.1. Motivating minimal surface problems. Plateau’s problem is to find a minimal surface with a prescribed boundary condition [43]. That is, if $\Omega \subset \mathbb{R}^2$ is a bounded open subset of the plane and $f: \partial\Omega \rightarrow \mathbb{R}$, Plateau’s problem is to find the parametric surface, $u(\mathbf{x})$, with minimal surface area that satisfies the prescribed boundary conditions $u(\mathbf{x}) = f(\mathbf{x})$ for $\mathbf{x} \in \partial\Omega$. Physically, this solution can be interpreted as finding the soap film that covers a closed wire, described by the set $\{(\mathbf{x}, f(\mathbf{x})) : \mathbf{x} \in \partial\Omega\} \subset \mathbb{R}^3$. One approach to solving this problem is to introduce the functional

$$J[u] = \int_{\Omega} \sqrt{1 + |\nabla u|^2} \, d\mathbf{x} + \alpha \int_{\partial\Omega} |u - f| \, d\mathbf{x},$$

2010 *Mathematics Subject Classification.* 65K15, 49Q05, 65Dxx, 68U10, 53C44.

Key words and phrases. graph partition, clustering, Beltrami energy, minimal surface, primal-dual hybrid gradients, convex optimization, bilevel optimization, threshold dynamics, rearrangement algorithm.

This work is supported by NSF DMS-1461138, the Swiss National Science Foundation under grants PBELP2-137727 and P300P2-147778, and the W. M. Keck Foundation.

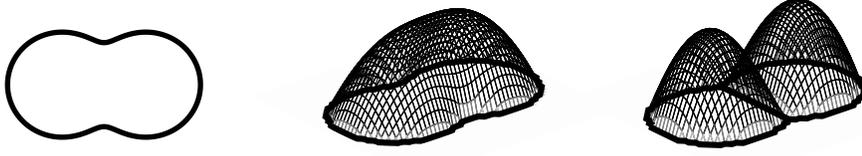


FIGURE 1. **(left)** A “pinched oval” domain, Ω . **(center)** The solution to (1). **(right)** For the same domain, the solution to (3) with $k = 2$. On each component of the optimal partitioning, $\Omega = \sqcup_{i=1}^k \Omega_i$, we plot the minimal area surface u_i^* .

where $\alpha > 0$, and minimize over functions of bounded variation,

$$\min_{u \in BV(\Omega)} J[u].$$

The first term of J is the surface area of the surface $(\mathbf{x}, u(\mathbf{x}))$ while the second term penalizes where u violates the boundary condition, but allows for a discontinuity at the boundary. It can be shown that the infimum of this problem is attained and for convex domains, some regularity results of the minimizer can be shown [25].

A related problem is to find the surface with minimal area that encloses a volume V above the plane and is zero on the boundary, $u(\mathbf{x}) = 0$ for $\mathbf{x} \in \partial\Omega$. This problem can be similarly expressed as the optimization problem

$$S(\Omega, V) := \min_{u \in BV(\Omega)} \int_{\Omega} \sqrt{1 + |\nabla u|^2} \, d\mathbf{x} + \alpha \int_{\partial\Omega} |u| \, d\mathbf{x} \quad (1a)$$

$$\text{such that } \int_{\Omega} |u(\mathbf{x})| \, d\mathbf{x} = V \quad \text{and} \quad u \geq 0. \quad (1b)$$

Again we can interpret this problem in terms of a soap film. We imagine a closed wire in the shape of $\partial\Omega$ that lies flat on the Euclidean plane. The wire is covered in a film of soap and a person slowly inflates (“blows on”) the soap film to form a bubble enclosing volume V . The shape of the bubble is then described by the parametric surface $(\mathbf{x}, u(\mathbf{x})) \in \mathbb{R}^{2+1}$. Of course, if the person blows too much air at the soap film, then the soap film will “bubble out” and no longer be a parametric surface of this form. At this value of V , the regularity of the solution to (1) breaks down and there is a discontinuity at the boundary. We don’t dwell on these issues here, as we will consider a discrete version of (1), where the questions of well-posedness and regularity are much simpler. As an illustration, for $V = 1$ and a domain Ω in the shape of a “pinched oval” (see Figure 1(left)), the optimal surface is plotted in Figure 1(center). Details of this computation are discussed in Section 5.4.

We now informally consider a “foam” consisting of k interacting bubbles of equal volume that are confined to a domain Ω . If bubble i is contained in Ω_i with $\Omega = \sqcup_{i=1}^k \Omega_i$, satisfying $\Omega_i \cap \Omega_j = \emptyset$ for $i \neq j$, then we could solve (1) on each subdomain, Ω_i to obtain the shape of bubble i , say $u_i^*(\mathbf{x})$ for $\mathbf{x} \in \Omega_i$. The total surface area of the foam is given by

$$\sum_{i=1}^k \int_{\Omega_i} \sqrt{1 + |\nabla u_i^*(\mathbf{x})|^2} \, d\mathbf{x}. \quad (2)$$

Of course, in this scenario the bubbles are also going to exert a force on one another to try to reduce the overall surface tension. For example, if one of the Ω_i is very small, then bubble i will exert a force on its neighbors and the domain Ω_i will enlarge as to decrease the overall surface tension. To model this behavior, we consider minimizing the total surface

area (2) over all k partitions of the domain $\Omega = \sqcup_{i=1}^k \Omega_i$ such that u_i^* is the solution to (1) on each subdomain, Ω_i .

This motivates the following isoperimetric question: what k partition of the domain minimizes the sum of the minimal surface areas, (2)? That is, we seek to solve the optimization problem,

$$P(\Omega, V, k) := \min_{\Omega = \sqcup_i \Omega_i} \sum_{i=1}^k S(\Omega_i, V). \quad (3)$$

Here, the minimum is taken over all disjoint k -partitions of Ω . For the pinched oval domain, we plot the optimal partition for $k = 2$ in Figure 1(right). On each subdomain, Ω_i^* , we also plot the surface $u^*(\mathbf{x})$ which attains the minimum in the isoperimetric problem (1).

There are two interesting limits for (3).

1. In the limit that $V \rightarrow 0$, the first term in the objective function in (1) is equivalent to the Dirichlet energy. Optimal Dirichlet partitions have been considered in a variety of papers [5, 6, 8, 9, 10, 28, 29, 40, 41]. One difference is the constraint; we consider the L^1 constraint and not the L^2 constraint since it is convex and has a physical interpretation in terms of the minimal surface problem.
2. For large V , the optimal solution is almost flat on each partition component and there is a large discontinuity at the interfaces of the partition components. The problem limits to the classical problem of finding a partition of k equal-volume components for which the surface area of the interface between components is minimal [23, 26, 42, 50].

In this paper, we formulate graph analogues of (1) and (3) and show that these problems are well-posed. To solve the bi-level optimization problem efficiently, we propose a primal-dual hybrid gradients method for the computation of the ground state of each partition (inner problem) and a rearrangement algorithm to find the optimal partitions (outer problem). We prove that the primal-dual hybrid gradients method is convergent and that the iterations of the rearrangement algorithm do not increase the objective value. We apply the proposed graph-partitioning method to a variety of example problems in graph partitioning coming from manifold discretizations, clustering and transductive learning, and image segmentation.

1.2. Outline. In Section 2, we give a brief background on the use of the Beltrami energy in image analysis and introduce terminology in order to formulate the graph analogues of (1) and (3). In Section 3, we introduce the graph analogue of the Beltrami energy, formulate a graph partitioning problem based on this energy, and give a rearrangement algorithm for solving the outer problem. In Section 4, we show how the inner problem can be solved using primal-dual optimization methods. In Sections 5 and 6, we apply these methods to a variety of graph partitioning examples. Finally, we conclude in Section 7 with a discussion.

2. Background. In this section, we briefly review how the Beltrami framework has been used previously for image analysis and introduce some graph terminology in preparation for the introduction of the Beltrami graph functional.

2.1. Beltrami framework and minimal surfaces. The Beltrami framework was introduced in the late 1990s for geometric image regularization. Adopting this geometric viewpoint amounts to interpreting objects such as images, shapes, or vector fields as surfaces embedded in some higher dimensional space [51]. The Beltrami framework associates the spatial coordinates along with the features by defining a Monge surface, say $X: (x, y) \mapsto (x, y, I(x, y))$ for a 2D scalar image. The Beltrami energy then measures the area of this

surface and its minimization amounts to regularizing the image without losing important edge details.

More formally, the *Beltrami embedding* associates the spatial coordinates with the features. For a scalar function $I: \Omega \rightarrow \mathbb{R}$, we define the diffeomorphism $X: \Omega \subset \mathbb{R}^n \rightarrow \Omega \times \mathbb{R}$, by $\mathbf{x} \mapsto (\mathbf{x}, I(\mathbf{x}))$. The embedding space of the manifold $\mathcal{M} \subset \Omega \times \mathbb{R}$ has coordinates with dimensions of a different nature, namely spatial components versus feature components. The relative scaling between these components is defined by a *metric tensor* h_{ij} that incorporates tuning of the aspect ratio, by a factor β , $h_{ij} := \text{diag}(1, \dots, 1, \beta^2)$. A metric tensor $g_{\mu\nu}$ associated with the original domain Ω is obtained through a pullback, $g_{\mu\nu} := \beta^2(\nabla I \otimes \nabla I) + \text{Id}_n$, where \otimes denotes the outer (tensor) product and Id_n is the $n \times n$ identity matrix. The determinant of the metric tensor is $g(I) := \det g_{\mu\nu} = 1 + \beta^2 |\nabla I|^2$, and can be seen as a generalized edge detector. Finally, the Beltrami energy integrates this edge detector to compute the volume of the manifold \mathcal{M} :

$$J_{\beta, \Omega}[I] := \int_{\Omega} \sqrt{g(I)} = \int_{\Omega} \sqrt{1 + \beta^2 |\nabla I|^2}. \quad (4)$$

The Beltrami energy is a measure of image regularity and generalizes to more complicated (co)domains [32]. The Beltrami functional is superficially similar to the regularized TV functional [46]. However, while regularized TV can be viewed as the mere smoothing of the TV functional, e.g., $\int_{\Omega} |\nabla u| \approx \varepsilon \int_{\Omega} \sqrt{1 + \varepsilon^{-2} |\nabla u|^2}$ with $\varepsilon \ll 1$, for numerical purposes, the ability of the Beltrami functional to interpolate between the TV and Dirichlet functionals has proven beneficial in many instances [3, 16, 22, 31, 34, 36, 44, 45, 47, 51, 52, 54, 56].

For $\beta = 1$, $J_{\beta, \Omega}[u]$ is the area of the Monge surface given by the graph $(\mathbf{x}, u(\mathbf{x})) \in \mathbb{R}^{d+1}$, where $\mathbf{x} \in \Omega \subset \mathbb{R}^d$. Consequently, properties of $J_{1, \Omega}$ are central to the study of minimal surfaces [19, 25].

2.2. A discrete calculus for graphs. In this section we recall several definitions of graph-based differential operators, which we then need to generalize the Beltrami energy to graphs.

Let $G = (V, E, w)$ be an undirected weighted graph with vertex set $V = \{i\}_{i=1}^N$, edge set $E = \{(i, j)\}$ for some $i, j \in V$, and non-negative edge-weights $w: E \rightarrow \mathbb{R}$. For notational convenience, we extend w to $V \times V$ by setting $w(i, j) = 0$ if $(i, j) \notin E$.

2.2.1. Inner products and norms on graphs.

Definition 2.1. For functions $f, g: V \rightarrow \mathbb{R}$, we define the L^2 inner product,

$$\langle f, g \rangle_{L^2(V)} := \sum_{i \in V} f_i g_i$$

and derived L^2 norm, $\|f\|_2 := \langle f, f \rangle_{L^2(V)}^{\frac{1}{2}} = \sqrt{\sum_{i \in V} f_i^2}$.

Definition 2.2. For skew-symmetric¹ edge functions $u, v: E \rightarrow \mathbb{R}$, we define the L^2 inner product,

$$\langle u, v \rangle_{L^2(E)} := \sum_{(i, j) \in E} u(i, j) v(i, j)$$

and derived L^2 norm $\|v\|_{L^2(E)} := \langle v, v \rangle_{L^2(E)}^{\frac{1}{2}} = \sqrt{\sum_{(i, j) \in E} v(i, j)^2}$.

¹A function $u: E \rightarrow \mathbb{R}$ is skew-symmetric if $u(i, j) = -u(j, i)$.

Definition 2.3 (Partial $L^p(V)$ norm). More generally, for $f: V \rightarrow \mathbb{R}$, we define the partial L^p norm on a subset $S \subseteq V$,

$$\|f\|_{S,p} := \left(\sum_{i \in S} |f_i|^p \right)^{\frac{1}{p}}.$$

We recover $L^2(V)$ for $S = V$ and $p = 2$.

Definition 2.4 (Magnitude). We define the magnitude $|v| \in L^2(V)$ of $v \in L^2(E)$,

$$|v|_i := \sqrt{\sum_{j: (i,j) \in E} v(i,j)^2}, \quad i \in V.$$

2.2.2. *Differential operators on graphs.* We define the following operators on graphs, which can be viewed as discrete differential operators.

Definition 2.5 (Graph gradient). For $f: V \rightarrow \mathbb{R}$, define the graph gradient to be the edge vector, $\nabla_w f: E \rightarrow \mathbb{R}$, with entries consisting of weighted vertex differences,

$$(\nabla_w f)(i,j) := \sqrt{w(i,j)}(f_i - f_j), \quad (i,j) \in E.$$

Definition 2.6 (Graph divergence). For skew-symmetric $v: E \rightarrow \mathbb{R}$, the divergence $\text{div}_w v: V \rightarrow \mathbb{R}$ is defined

$$(\text{div}_w v)_i := \sum_{j: (i,j) \in E} \sqrt{w(i,j)}v(i,j), \quad i \in V.$$

It is easy to verify that the graph gradient and divergence are adjoint using the vertex and edge L^2 inner products,

$$\langle \nabla_w f, v \rangle_{L^2(E)} := \langle f, \text{div}_w v \rangle_{L^2(V)}, \quad f \in L^2(V), v \in L^2(E). \quad (5)$$

Definition 2.7 (Graph Laplacian). The graph Laplacian, $\Delta_w: L^2(V) \rightarrow L^2(V)$ is defined as the composition of graph divergence and graph gradient,

$$(\Delta_w f)_i := (\text{div}_w \circ \nabla_w f)_i = \sum_{j: (i,j) \in E} w_{ij}(f_i - f_j).$$

Definition 2.8 (Dirichlet Energy). The Dirichlet energy of $f: V \rightarrow \mathbb{R}$ is defined

$$D[f] := \|\nabla_w f\|_{L^2(E)}^2 = \sum_{(i,j) \in E} |(\nabla_w f)(i,j)|^2 = \frac{1}{2} \sum_{i \in V} |\nabla_w f|_i^2.$$

Definition 2.9 (Total Variation). The isotropic total variation of $f: V \rightarrow \mathbb{R}$ is defined

$$TV[f] := \sum_{i \in V} |\nabla_w f|_i.$$

The anisotropic total variation of $f: V \rightarrow \mathbb{R}$ is defined

$$TV_a[f] := \sum_{(i,j) \in E} |(\nabla_w f)(i,j)|.$$

In this paper, we will use the isotropic total variation, not the anisotropic total variation.

3. Proposed geometric graph partitioning model. With the definitions of inner products and differential operators on graphs from Section 2 in hand, we next define the Beltrami energy on graphs.

In this section, we assume that $G = (V, E, w)$ is a given weighted graph, where the weights are determined by the application. These weights then propagate to the graph objects defined in Section 2.

3.1. Definition and properties of the graph Beltrami energy.

Definition 3.1. We define the graph Beltrami energy of $u: V \rightarrow \mathbb{R}$:

$$\begin{aligned} J_\beta[u] &:= \sum_{i \in V} \sqrt{1 + \beta^2 |\nabla_w u|_i^2} \\ &= \sum_{i \in V} \sqrt{1 + \beta^2 \sum_{j: (i,j) \in E} w_{ij} (u_i - u_j)^2}. \end{aligned} \quad (6)$$

Lemma 3.2. *The graph Beltrami energy, J_β , defined in (6), is a proper, closed, convex function on $L^2(V)$.*

Proof. Let $\mathbb{R}_+^N = \{x \in \mathbb{R}^N \mid x_i \geq 0\}$, and identify $L^2(V)$ with \mathbb{R}^N . Then define

$$f: \mathbb{R}_+^N \rightarrow \mathbb{R}, f(\mathbf{x}) = \sum_{i=1}^N \sqrt{1 + \beta^2 x_i^2}.$$

The function f is non-decreasing and strictly convex on \mathbb{R}_+^N . Further, define

$$g: \mathbb{R}^N \rightarrow \mathbb{R}^N, g_i(u) = \sqrt{\sum_{j=1}^N w_{ij} (u_i - u_j)^2}.$$

If we can show that g_i for $i = 1 \dots N$, are convex functions, then J_β is convex since $J_\beta = f \circ g$.

To see that $g_i: \mathbb{R}^N \rightarrow \mathbb{R}$ is convex, we define

$$h_i: \mathbb{R}_+^N \rightarrow \mathbb{R}, h_i(\mathbf{y}) = \sqrt{\sum_{j=1}^N w_{ij} y_j^2}.$$

This h_i is non-decreasing and convex on \mathbb{R}_+^N . We then define $\gamma_i: \mathbb{R}^N \rightarrow \mathbb{R}^N$, with $(\gamma_i)_j(\mathbf{x}) = |x_i - x_j|$. Since $(\gamma_i)_j$ for $j = 1, \dots, N$, are convex and $g = h \circ \gamma$, then g is convex. \square

Remark 1. For a graph which corresponds to a square lattice, the graph Beltrami energy approximates the Beltrami energy as considered in Section 1. Namely, let $\Omega \subset \mathbb{R}^2$, V correspond to the set of points in a square lattice, $\{x_i, y_i\}_{i \in \mathbb{Z}^2}$, which are a subset of Ω , and the edge set, E , connect vertices with their four nearest neighbors with weight $w(i, j) = \frac{1}{2}h^{-1}$, where h is the lattice mesh size. Then, for $\beta = 1$, the graph Beltrami energy approximates the surface area of the discretized parameterized surface (x_i, y_i, u_i) where $u_i = u(x_i, y_i)$ for some $u \in H^1(\Omega)$.

Remark 2. We note that varying β allows interpolating between the graph Dirichlet energy ($\beta \rightarrow 0$) and graph total variation ($\beta \rightarrow \infty$),

$$J_\beta[u] \sim \begin{cases} |V| + \beta^2 \cdot D[u] & \beta \rightarrow 0 \\ \beta \cdot TV[u] & \beta \rightarrow \infty. \end{cases} \quad (7)$$

This graph-based definition makes the benefits of Beltrami regularization available for data defined on graphs, and applies, *e.g.*, to clustering or segmentation, with numerous applications in machine learning.

3.2. Minimal Beltrami energy. Motivated by the minimal surface problems discussed in the introduction (Section 1), we consider the following isoperimetric graph problem. Let $S \subset V$ be a vertex subset. On S , we consider the vertex function u^* , which has unit $L^1(V)$ norm, vanishes on the complement $S^c := V \setminus S$, and has minimal Beltrami energy, *i.e.*, satisfies

$$J_\beta^*[S] := \min J_\beta[u] \quad \text{s.t.} \quad u: V \rightarrow [0, 1], \|u\|_{V,1} = 1, \text{ and } u|_{S^c} = 0. \quad (8)$$

Here J_β is defined in (6).

Lemma 3.3. *The set of vertex functions,*

$$\begin{aligned} U &:= \{u: V \rightarrow \mathbb{R} \mid u \geq 0, \|u\|_{V,1} = 1\} \\ U_0 &:= U \cap \{u: u|_{S^c} = 0\} \end{aligned}$$

are both closed convex.

Proof. The set U is simply the standard simplex and thus closed convex. U_0 is closed convex since it is the intersection of closed convex sets. \square

It now follows from Lemmata 3.2 and 3.3 that (8) is a convex optimization problem, so there exists at least one solution.

Remark 3. Note that the substitution $z = \beta u$ in (8), shows the equivalence to the problem

$$\min J_1[z] \quad \text{s.t.} \quad z: V \rightarrow [0, \beta], \|z\|_{V,1} = \beta, \text{ and } u|_{S^c} = 0.$$

The respective minimizers of these two problems are related by $z^* = \beta u^*$, while the optimal values are not equal.

In terms of the graph (V, E, w) described in Remark 1, we interpret $J_\beta^*[S]$ for $\beta = 1$ to be the minimal surface area of the unit-mass distributed over the set $S \subset V$.

3.3. Minimal Beltrami energy graph partitioning. We now consider a vertex partition of the graph, $V = \sqcup_{\ell=1}^k V_\ell$. To this vertex partition, we associate an energy, which is the sum of the minimal Beltrami energies, $\sum_{\ell=1}^k J_\beta^*[V_\ell]$. We then consider the *Beltrami graph partitioning problem*,

$$\min_{V = \sqcup_{\ell} V_\ell} \sum_{\ell=1}^k J_\beta^*[V_\ell]. \quad (9)$$

Remark 4. Since there are only a finite number of partitions, the minimum in (9) is attained by some partition.

3.4. Boundary conditions relaxation. An interesting route of rendering (9) accessible to efficient computational optimization has been proposed for the closely related Dirichlet energy graph partitioning problem in [6, 41]. The approach consists of relaxing the Dirichlet boundary conditions of the inner minimization problem by replacing them with a penalty.

Indeed, the boundary conditions are equivalently obtained by imposing the partial L^q norm of u to vanish on S^c :

$$u|_{S^c} = 0 \Leftrightarrow \|u\|_{S^c, q}^q = 0. \quad (10)$$

Instead of imposing this constraint strictly, one can penalize violations by including the partial L^q norm in the optimization problem. Natural choices include $q = 2$ (quadratic penalty) and $q = 1$ (absolute error).

The relaxed partitioning model then becomes:

$$\min_{V = \sqcup_{\ell} V_\ell} \sum_{\ell=1}^k J_\beta^{\alpha, *}[V_\ell] \quad (11)$$

where the inner model (8) is relaxed to

$$J_{\beta}^{\alpha*}[S] := \min \left\{ J_{\beta}[u] + \frac{\alpha}{q} \|u\|_{S^c, q}^q \right\} \quad \text{s.t.} \quad u: V \rightarrow [0, 1], \|u\|_{V, 1} = 1. \quad (12)$$

Proposition 1. Fix $\beta > 0$ and $S \subset V$. For any $\alpha > 0$, the relaxed problem (12) is a convex optimization problem and thus has a minimizer. As $\alpha \rightarrow \infty$, the minimum value of J_{β}^{α} converges to the minimum value of J_{β} ,

$$\lim_{\alpha \rightarrow \infty} J_{\beta}^{\alpha*}[S] = J_{\beta}^*[S]. \quad (13)$$

Also there exists a sequence of values α_k with corresponding minimizers u_k^* of (12) such that the u_k^* converge to a minimizer of (8).

Proof. The relaxed boundary condition term, $\frac{\alpha}{q} \|u\|_{S^c, q}^q$, is closed convex, J_{β} is closed convex (Lemma 3.2), and the constraint set U is convex (Lemma 3.3), so (12) is a convex optimization problem.

We compute

$$\frac{d}{d\alpha} J_{\beta}^{\alpha*} = \frac{1}{q} \|u^*\|_{S^c, q}^q \geq 0,$$

where u_{α}^* is a minimizer for J_{β}^{α} . For any finite α , we have that $J_{\beta}^{\alpha*}[S] \leq J_{\beta}^*[S]$ since $U_0 \subset U$ and $J_{\beta}^{\alpha} \equiv J_{\beta}$ on U_0 . Consequently $\lim_{\alpha \rightarrow \infty} J_{\beta}^{\alpha*}[S]$ exists and satisfies $\lim_{\alpha \rightarrow \infty} J_{\beta}^{\alpha*}[S] \leq J_{\beta}^*[S]$.

For the reverse inequality, we consider an increasing sequence $\{\alpha_k\}_k$. Let $\{u_k^*\}_k$ be a minimizer of $J_{\beta}^{\alpha}[S]$ with $\alpha = \alpha_k$. There exists a \tilde{u} and a subsequence of $\{u_k^*\}_k$ (labeled using the same index) such that $u_k^* \rightarrow \tilde{u}$. Passing to the limit we see that $\|\tilde{u}\|_{S^c, q}^q = 0$, so $u \in U_0$ (admissible for (8)). Thus we have that $\lim_{\alpha \rightarrow \infty} J_{\beta}^{\alpha*}[S] = J_{\beta}(\tilde{u}) \geq J_{\beta}^*[S]$. This holds for any limit point of a sequence of minimizers. \square

3.5. Rearrangement algorithm for the bi-level optimization problem (11). With the relaxation of the boundary condition in place, the bi-level optimization problem can be solved by a rearrangement algorithm extended from [41].

1. Given partitions $V = \sqcup_{\ell} V_{\ell}$, compute a minimizer u_{ℓ}^* for satisfying (12) with $S = V_{\ell}$ for each $\ell = 1, \dots, k$. (This is the inner problem for (11).)
2. Given minimizers u_{ℓ}^* , reassign partitions by the ‘‘winner takes all’’ rule,

$$V_k \ni v \iff k = \arg \max_{\ell} \{u_{\ell}^*(v)\}.$$

This two-step iterative algorithm is also reminiscent of the structure of popular clustering algorithms like k-means [35] and expectation-maximization (E-M) for Gaussian mixture models [17]. This iterative algorithm also shares many attributes with diffusion generated motion by mean curvature [20, 24, 37, 38, 39, 55].

Proposition 2. The iterations of the rearrangement algorithm do not increase the objective function in (11).

Proof. Let $\sqcup_{\ell} V_{\ell}$ and $\sqcup_{\ell} V_{\ell}^+$ denote the old and updated vertex partitions, respectively, and u_{ℓ} and u_{ℓ}^+ the corresponding minimizers of the inner problem (12). The old objective satisfies

$$\sum_{\ell} J_{\beta}[u_{\ell}] + \frac{\alpha}{q} \|u_{\ell}\|_{V_{\ell}, q}^q \geq \sum_{\ell} J_{\beta}[u_{\ell}] + \frac{\alpha}{q} \|u_{\ell}\|_{V_{\ell}^+, q}^q \geq \sum_{\ell} J_{\beta}[u_{\ell}^+] + \frac{\alpha}{q} \|u_{\ell}^+\|_{V_{\ell}^+, q}^q, \quad (14)$$

which is the value of the updated objective. The first inequality holds by construction of the algorithm and the second inequality holds by the optimality of u_{ℓ}^+ in (12). \square

In practice, we observe that the rearrangement algorithm converges in a small number of iterations.

4. Primal-dual methods for solution of the inner problem. The computation of the minimizers u_ℓ^* satisfying (12) with $S = V_\ell$ for each $\ell = 1, \dots, k$ is the computational bottleneck of the proposed optimization scheme in Section 3.5. Thus, it is important to make this step as efficient as possible. By Proposition 1 these inner problems are convex. There is a huge corpus of algorithms that apply to these convex optimization problems. Recently, primal-dual splitting algorithms have gained particular attraction [19], most notably so in the context of TV and L^1 -type problems in imaging, *e.g.*, [11, 21, 59, 60]. Zosso and Bustin have applied the theory to the particular case of the Beltrami functional [51], used as an interesting regularizer in imaging problems that interpolates between the classical H^1 and TV regularizers [62].

In Subsections 4.1 we show how the inner problem (12) can be formulated as a saddle point problem. In Subsection 4.2 we summarize a general family of algorithms which we adapt to our problem in Subsection 4.3-4.5. In Subsection 4.6 we present a semi-supervised extension to our algorithm.

4.1. Saddle point formulation. We recall the definition of the convex conjugate (a.k.a. Legendre-Fenchel transform) of a function:

Definition 4.1 (Legendre-Fenchel transform). The convex conjugate of a function f is the function f^* defined by

$$f^*(s) = \sup_{x \in \mathbb{R}^n} \left\{ \langle s, x \rangle - f(x) \right\}, \quad s \in \mathbb{R}^n. \quad (15)$$

The convex conjugate of a closed convex function is again a closed convex function. Also, the biconjugate, $f^{**} := (f^*)^*$, is the largest closed convex function with $f^{**} \leq f$. As a result, $f = f^{**}$ iff f is closed convex (Fenchel-Moreau Theorem).

As the following Theorem shows, convex problems of a certain form can be rewritten as a primal-dual problem using the Legendre-Fenchel transform. This saddle point formulation of the problem will be exploited by the numerical methods employed here.

Theorem 4.2. *Let $F: W \rightarrow \mathbb{R}$ be a closed and convex functional on the set W , $G: X \rightarrow \mathbb{R}$ a closed and convex functional, and $K: X \rightarrow W$ be a continuous linear operator. Then we have the following equivalence:*

$$\underbrace{\min_{x \in X} \left\{ F(Kx) + G(x) \right\}}_{\text{Primal}} = \min_{x \in X} \underbrace{\max_{\phi \in W^*} \left\{ \langle Kx, \phi \rangle - F^*(\phi) + G(x) \right\}}_{\text{Primal-Dual}} \quad (16)$$

where x and ϕ are the primal and dual variables, respectively, F^* is the convex conjugate of F , and W^* is the dual space of W .

For a proof of Theorem 4.2, see [19, p. 178]. Comparing (12) with the primal-dual model in Theorem 4.2, we make the following identifications:

$$F(Ku) = \sum_{i \in V} \sqrt{1 + \beta^2 |\nabla_w u|_i^2} \quad \text{and} \quad G(u) = \frac{\alpha}{q} \|u\|_{S^c, q}^q.$$

In particular, we thus define

$$K: X \rightarrow W, \quad u \mapsto \nabla_w u \quad \text{and} \quad F: W \rightarrow \mathbb{R}, \quad x \mapsto \sum_{i \in V} \sqrt{1 + \beta^2 |x|_i^2},$$

with the reflexive spaces $X = U$ and $W = L^2(E)$. The Legendre-Fenchel transform F^* of the function F is found as:

$$F^*: W^* \rightarrow \mathbb{R}, \quad \phi \mapsto - \sum_{i \in V} \sqrt{1 - |\phi|_i^2 / \beta^2}, \quad (17)$$

where $W^* = L^2(E) \cap \{\phi: |\phi|_i^2 \leq \beta^2\}$, which is closed convex. In Lemma 3.2 it was shown that F is convex. Thus, using Theorem 4.2, we now get the equivalent primal-dual problem:

$$\min_{u \in U} \max_{\phi \in W^*} \left\{ \langle \nabla_w u, \phi \rangle_{L^2(E)} + \sum_{i \in V} \sqrt{1 - |\phi|_i^2 / \beta^2} + \frac{\alpha}{q} \|u\|_{S^{c,q}}^q \right\}. \quad (18)$$

Further, using the adjoint relation (5) of the graph differential operators, (18) can be rewritten:

$$\min_{u \in U} \max_{\phi \in W^*} \left\{ \langle u, \operatorname{div}_w \phi \rangle_{L^2(V)} + \sum_{i \in V} \sqrt{1 - |\phi|_i^2 / \beta^2} + \frac{\alpha}{q} \|u\|_{S^{c,q}}^q \right\}. \quad (19)$$

Thus, by Theorem 4.2, (19) and (12) are equivalent.

4.2. Alternating direction minimization. Having established the general primal-dual saddle point problem (16), the saddle-point can now be found iteratively, for example by the backward-backward splitting algorithm by Chambolle and Pock [11]. This algorithm has also been classified as a modified primal-dual hybrid gradient method (PDHGM) in [21]. The iterative steps are:

$$\phi^{n+1} = \arg \min_{\phi \in W^*} \left\{ -\langle Kx^n, \phi \rangle + F^*(\phi) + \frac{1}{2r_1} \|\phi^n - \phi\|_2^2 \right\} \quad (20a)$$

$$x^{n+1} = \arg \min_{x \in V} \left\{ \langle x, K^* \phi^{n+1} \rangle + G(x) + \frac{1}{2r_2} \|x^n - x\|_2^2 \right\} \quad (20b)$$

$$\bar{x}^{n+1} = x^{n+1} + \theta(x^{n+1} - x^n). \quad (20c)$$

The first two steps are the proximal update of the dual and primal variable, respectively. The third, extra-gradient step is an overrelaxation of the primal update in order to overcome the stepsize shortening typical of first order methods. Relaxation parameter $\theta = 0$ corresponds to the Arrow-Hurwitz algorithm, whereas faster convergence can be achieved for $\theta = 1$ [11]. In general, $O(1/N)$ convergence, where N is the number of iterations, has been shown for fixed r_1, r_2 , such that

$$r_1 r_2 \|K\|^2 \leq 1, \quad (21)$$

where $\|K\|$ is the operator norm/induced norm of the linear operator K , or equivalently $\|K\|^2 = \|K^* K\|$ is the induced norm of the composition $K^* K$.

It should be mentioned that the above PDHG-algorithm was preceded by TV-minimization by Zhu and Chan [59, 60], which solves the minimization steps only approximately, through an explicit step of gradient ascent/descent followed by projection onto the convex sets W^* and V , respectively, and without the extra-gradient step:

$$\phi^{n+1} = P_{W^*} (\phi^n + r_1 (Kx^n - \partial_\phi F^*(\phi^n))) \quad (22a)$$

$$x^{n+1} = P_V (x^n + r_2 (K^* \phi^{n+1} - \partial_x G(x^n))) \quad (22b)$$

where $P_S(\cdot) = \arg \min_{s \in S} \|s - \cdot\|_2^2$ denotes Euclidean projection on the closed convex set S . This is an instance of Arrow-Hurwicz-Uzawa algorithm [1]. Our algorithm will combine Chambolle and Pock's implicit minimization steps with Zhu and Chan's projections onto convex sets.

4.3. Primal-dual hybrid gradients for the inner problem. The saddle-point problem of the inner optimization is given in (18). We adapt the general primal-dual hybrid gradients algorithm in (20) to (18) as follows:

$$\phi^{n+1} = \arg \min_{\phi \in W^*} \left\{ -\langle \nabla_w \bar{u}^n, \phi \rangle_{L^2(E)} - \sum_{i \in V} \sqrt{1 - |\phi|_i^2 / \beta^2} + \frac{1}{2r_1} \|\phi^n - \phi\|_2^2 \right\} \quad (23a)$$

$$u^{n+1} = \arg \min_{u \in U} \left\{ \langle u, \operatorname{div}_w \phi^{n+1} \rangle_{L^2(V)} + \frac{\alpha}{q} \|u\|_{S^c, q}^q + \frac{1}{2r_2} \|u^n - u\|_2^2 \right\} \quad (23b)$$

$$\bar{u}^{n+1} = u^{n+1} + \theta(u^{n+1} - u^n) \quad (23c)$$

We identify in our case $K = \nabla_w$, and more importantly $K^*K = \Delta_w$. The general $O(1/N)$ convergence guarantee (21) of [11] applies if

$$r_1 r_2 \leq \rho(\Delta_w)^{-1}, \quad (24)$$

where $\rho(\Delta_w)$ denotes the spectral radius of the graph Laplacian.

In the following subsections, we show how the respective minimization sub-problems (23a) and (23b) can be solved efficiently.

4.4. Dual variable update. We first address the update of the *dual* variable ϕ . The minimization problem of the dual variable update (23a) can be rewritten as

$$\phi^{n+1} = \arg \min_{\phi \in W^*} \left\{ -\langle \nabla_w \bar{u}^n, \phi \rangle_{L^2(E)} - \sum_{i \in V} \frac{1 - |\phi|_i^2 / \beta^2}{\sqrt{1 - |\phi|_i^2 / \beta^2}} + \frac{1}{2r_1} \|\phi^n - \phi\|_2^2 \right\}. \quad (25)$$

We propose to solve this in an iteratively reweighted least squares (IRLS) approach as follows. We observe that the objective in (25) without the square-root in the denominator is quadratic in ϕ . We fix the current estimate of ϕ in the square-root to obtain a weighted least squares problem, this weight is then updated, and the process is repeated until convergence:

$$\psi^0 = \phi^n \quad (26a)$$

$$\psi^{t+1} = \arg \min_{\psi \in W^*} \left\{ -\langle \nabla_w \bar{u}^n, \psi \rangle_{L^2(E)} - \sum_{i \in V} \frac{\beta^2 - |\psi|_i^2}{\beta \sqrt{\beta^2 - |\psi^t|_i^2}} + \frac{1}{2r_1} \|\phi^n - \psi\|_2^2 \right\} \quad (26b)$$

$$\phi^{n+1} = \lim_{t \rightarrow \infty} \psi^t \quad (26c)$$

Our focus is on the solution of the fixed-point minimization step (26b).

Lemma 4.3. *For small enough r_1 , the solution to (26b) is given by*

$$\psi^{t+1}(i, j) = \frac{\varepsilon_i^t}{\varepsilon_i^t + 2r_1/\beta} [\phi^n(i, j) + r_1 (\nabla_w \bar{u}^n)(i, j)], \quad \forall (i, j) \in E, \quad (27)$$

where $\varepsilon_i^t := \sqrt{\beta^2 - |\psi^t|_i^2}$.

Proof. The solution to the unconstrained optimization problem

$$\arg \min_{\psi \in L^2(V)} \left\{ -\langle \nabla_w \bar{u}^n, \psi \rangle_{L^2(E)} - \sum_{i \in V} \frac{\beta^2 - |\psi|_i^2}{\beta \varepsilon_i^t} + \frac{1}{2r_1} \|\phi^n - \psi\|_2^2 \right\}. \quad (28)$$

is given by (27). It remains to show that, for sufficiently small r_1 , $\psi^{t+1} \in W^*$ and is therefore the solution of the constrained problem (26b) as well. In (27), for each i , the coefficient in front of the square brackets is less than 1. A geometric illustration of the contraction is

given in Figure 2(a). A sufficient condition for $\Psi^{t+1} \in W^*$ is $\Phi^n + r_1(\nabla_w \bar{u}^n) \in W^*$, or equivalently:

$$|\Phi^n + r_1(\nabla_w \bar{u}^n)|_i \leq \beta, \quad \forall i \in V. \quad (29)$$

We compute for each $i \in V$:

$$|\Phi^n + r_1(\nabla_w \bar{u}^n)|_i^2 \leq |\Phi^n|_i^2 + r_1^2 |\nabla_w \bar{u}^n|_i^2 \quad (30a)$$

$$= |\Phi^n|_i^2 + r_1^2 \sum_{j, (i,j) \in E} w(i,j) (\bar{u}_i^n - \bar{u}_j^n)^2 \quad (30b)$$

$$\leq |\Phi^n|_i^2 + r_1^2 \sum_{j, (i,j) \in E} w(i,j) \quad (30c)$$

$$\leq |\Phi^n|_i^2 + r_1^2 \max_{o \in V} \sum_{j, (o,j) \in E} w(o,j). \quad (30d)$$

Since $\Phi^n \in W^*$, for

$$r_1 \leq \sqrt{\frac{\beta^2 - \max_i |\Phi^n|_i^2}{\max_i d_i}},$$

where $d_i = \sum_{j, (i,j) \in E} w(i,j)$ is the degree of vertex i in the graph, the condition is met. \square

Lemma 4.4. *If $r_1 > 0.08 \cdot \beta^2$, the iterative scheme (26) converges to a fixed point in W^* .*

Proof. Using Lemma 4.3 and denoting $\alpha(i,j) := \Phi^n(i,j) + r_1(\nabla_w \bar{u}^n)(i,j)$, for $(i,j) \in E$, the iterates in (26) can be written

$$\Psi^{t+1}(i,j) = \frac{\epsilon_i^t}{\epsilon_i^t + 2r_1/\beta} \alpha(i,j).$$

We then have that

$$\epsilon_i^{t+1} = \sqrt{\beta^2 - |\Psi^{t+1}|_i^2} = \sqrt{\beta^2 - \left(\frac{\epsilon_i^t}{\epsilon_i^t + 2r_1/\beta} \right)^2 |\alpha|_i^2}.$$

Thus, in the iterative scheme (26) there are no dependencies between the vertices. We claim that the mappings

$$x \mapsto f_i(x) := \sqrt{\beta^2 - \left(\frac{x}{x + 2r_1/\beta} \right)^2 |\alpha|_i^2}$$

for every $i \in V$ are contraction mappings on the interval $[0, \beta]$ and hence by the contraction mapping theorem, each has a unique fixed point in $[0, \beta]$. To show that f_i is a contraction mapping, we compute

$$|f'_i(x)| = \frac{|\alpha|_i^2 x 2r_1/\beta}{(x + 2r_1/\beta)^3} \frac{1}{\sqrt{\beta^2 - \left(\frac{x}{x + 2r_1/\beta} \right)^2 |\alpha|_i^2}} \leq \frac{|\alpha|_i^2 2r_1}{(2r_1/\beta)^3} \frac{1}{\sqrt{\beta^2 - \left(\frac{\beta}{\beta + 2r_1/\beta} \right)^2 |\alpha|_i^2}}.$$

From (29), we have that $|\alpha|_i \leq \beta$ and thus

$$|f'_i(x)| \leq \frac{\beta^2 2r_1}{(2r_1/\beta)^3} \frac{1}{\sqrt{\beta^2 - \left(\frac{\beta}{\beta + 2r_1/\beta} \right)^2 \beta^2}} = \frac{\beta^2}{2r_1} \left[\frac{\xi}{(1 + \xi)^3} \frac{1}{\sqrt{1 - \left(\frac{\xi}{1 + \xi} \right)^2}} \right] \equiv \frac{\beta^2}{2r_1} g(\xi)$$

where $\xi = \frac{\beta x}{2r_1}$. A plot of the function $g(\xi)$ reveals that it is bounded by 0.16 for all $\xi \geq 0$.

We have that $|f'_i(x)| \leq 0.08 \frac{\beta^2}{r_1} < 1$ for every x , implying that f is a contraction mapping. \square

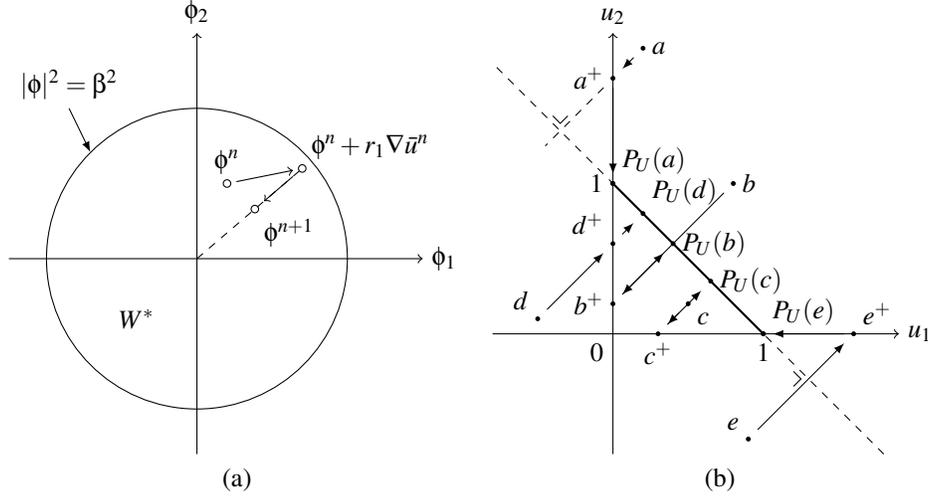


FIGURE 2. Geometric illustration of dual and primal variable updates. **(a)** The dual variable update step involves the current estimate ϕ^n and a contribution from the primal gradient. The sum $\phi^n + r_1 \nabla_w \bar{u}^n$ is then shortened towards the origin; the resulting point is guaranteed to be in W^* when the intermediate $\phi^n + r_1 \nabla_w \bar{u}^n$ is in W^* . In this case, the dual variable update does not involve any projection. **(b)** The projection of the primal variable update on the probability simplex U is achieved in two steps. First, the variable is shifted along the main diagonal such that its smallest component has value 0 ($u \rightarrow u^+$). Then, this non-negative u^+ is projected on to the simplex by uniform distribution of the mass deficit (if $|u^+|_{V,1} < 1$), or appropriately shrunk (soft-thresholding) if the current mass exceeds 1. The points a – e represent different cases. See Sections 4.4 and 4.5.3.

In practice we take the value of r_1 to be much smaller than the sufficient condition given in Lemma 4.4 and have not encountered any convergence issues.

4.5. Primal variable update. We now turn our attention to the minimization problem (23b), associated with the proximal update of the *primal* variable u . There are two cases to be considered, corresponding to the two choices $q = 2$ and $q = 1$ of the penalty term of the relaxed Dirichlet boundary condition. As will be shown shortly, both problems admit a simple closed form solution u_q^* , for the unconstrained problem with $u \in L^2(V)$. To approximate the optimal solution within U , we then project u_q^* onto U ,

$$u^{n+1} = P_U(u^*), \quad (31)$$

which is a reasonable approach for practical purposes, given the convexity of the involved functionals and the set U . Indeed, this mimics the projected gradient steps in the primal update of the Arrow-Hurwitz-Uzawa algorithm (22b), modulo the backward update used here instead of explicit, forward gradient descent. We first have to find the respective u_q^* .

4.5.1. Primal variable update for $q = 2$. Let us consider the relaxed problem $u \in L^2(V)$:

$$u^{n+1} = \arg \min_{u \in L^2(V)} \left\{ \langle u, \operatorname{div}_w \phi^{n+1} \rangle_{L^2(V)} + \frac{\alpha}{2} \|u\|_{S^{c,2}}^2 + \frac{1}{2r_2} \|u^n - u\|_2^2 \right\} \quad (32)$$

The minimization problem is quadratic and admits the following closed form solution:

$$(u_2^*)_i = \frac{u_i^n - r_2(\operatorname{div}_w \phi^{n+1})_i}{1 + \alpha r_2(1 - \chi_i)},$$

where χ is the characteristic function of the current partitioning, and the update can be computed by simple vertex-wise operations.

4.5.2. *Primal variable update for $q = 1$.* Let us in turn consider the relaxed problem $u \in L^2(V)$ of the $q = 1$ based problem:

$$u^{n+1} = \arg \min_{u \in L^2(V)} \left\{ \langle u, \operatorname{div}_w \phi \rangle_{L^2(V)} + \alpha \|u\|_{S^c, 1} + \frac{1}{2r_2} \|u^n - u\|_2^2 \right\}. \quad (33)$$

We first note that an equivalent minimization problem is given by

$$u_1^* = \arg \min_u \left\{ \alpha r_2 \|u\|_{S^c, 1} + \frac{1}{2} \|u - (u^n - r_2 \operatorname{div}_w \phi^{n+1})\|_2^2 \right\}, \quad (34)$$

which is an L^1 -type problem solved by shrinkage:

$$(u_1^*)_i = \operatorname{shrink}(u_i^n - r_2(\operatorname{div}_w \phi^{n+1})_i, \alpha r_2(1 - \chi_i)), \quad (35)$$

where the soft-thresholding operator is defined as

$$\operatorname{shrink}(z, \tau) := \begin{cases} z + \tau & z < -\tau, \\ 0 & |z| \leq \tau, \\ z - \tau & z > \tau. \end{cases} \quad (36)$$

Again, we see that all operations are simple and acting vertex-wise.

4.5.3. *Probability simplex projection.* Here, $P_U : L^2(V) \rightarrow U$ denotes projection on the probability simplex:

$$P_U(u) := \arg \min_v \|v - u\|_2^2 \quad \text{s.t.} \quad \|v\|_{V, 1} = 1 \quad \text{and} \quad \forall i \in V : v_i \geq 0 \quad (37)$$

To avoid special considerations regarding negative components, we first uniformly shift the function u so that it is non-negative at each vertex:

$$u^+ := u - \min_{i \in V} (u_i).$$

Such a translation happens along a direction orthogonal to the probability simplex, and the projection is thereby not altered. The actual projection algorithm then practically distinguishes between two cases, depending on the L^1 -norm of the non-negative $u^+ \in L^2(V)$: If $\|u^+\|_{V, 1} > 1$, then $P_U(u)$ is obtained by an appropriate soft-thresholding of u^+ [18]. On the other hand, if $\|u^+\|_{V, 1} \leq 1$, then the projection is obtained by distributing the lacking mass over all vertices, uniformly:

$$(P_U(u))_i = \begin{cases} \operatorname{shrink}((u^+)_i, \lambda) & \text{if } \|u^+\|_{V, 1} > 1, \\ (u^+)_i + \frac{1 - \|u^+\|_{V, 1}}{|V|} & \text{if } \|u^+\|_{V, 1} \leq 1, \end{cases} \quad i \in V, \quad (38)$$

where $|V|$ denotes the number of vertices in V . Different cases of this projection are illustrated for a 2-dimensional u in Figure 2 (b). An efficient strategy for finding the appropriate thresholding parameter λ is given in [18].

4.6. Semi-supervised extension. The proposed partitioning algorithm lends itself to a straightforward semi-supervised extension. Indeed, in Section 3.5 we defined the rearrangement algorithm as the outer step of the bi-level optimization scheme: Given the k minimizing u_ℓ^* , $\ell = 1, \dots, k$, each vertex is reassigned by the “winner takes all” rule. It is easy to modify this rearrangement step to keep fixed the given assignments of a subset of vertices. The motivation for such a semi-supervised extension is twofold:

First, the semi-supervised extension allows to integrate prior knowledge in to the partitioning problem. In the context of data clustering, this is known as *transductive learning*, where the dataset to be clustered consists of both labeled and unlabeled points.

Second, when a graph is a discretization of a manifold with boundary, it is desirable to enforce extra conditions on some “boundary vertices”, *e.g.*, Dirichlet boundary conditions. A Dirichlet boundary condition can be weakly enforced on a set of vertices using the same method; this is explained with an example in Section 5.4.

5. Experiments I: graphs from discretized manifolds. In this first experimental section, we focus on the partitioning of graphs which are discretizations of manifolds. We consider several examples that have been previously studied and explore a few new problems. These graphs are local in the sense that vertices are adjacent only to their nearest neighbors. These graphs are particularly valuable examples because they are easy to visualize and the existence of analytical solutions in some cases provides instructional insight into the nature of the proposed method. Graphs which are not discretizations of manifolds are considered in Section 6.

The proposed experiments gradually increase in complexity. We start by partitioning a small graph representing a discretized line segment. We then increase the problem dimension and partition the discretization of a square domain with both Neumann and periodic boundary conditions. We then partition barbell and ring-like subdomains of the square graph using Dirichlet boundary conditions, to illustrate the use of the semi-supervised extension. Finally, we bi-partition a discretization of the cube with periodic boundary conditions. In each case, along with the locally optimal partition $\sqcup_\ell V_\ell$ and a composite of ground states, $\max_\ell u_\ell$, we also report the value of the objective function minus the value $N \cdot k$,

$$o := \sum_{\ell=1}^k J_\beta^{\alpha,*}[V_\ell] - N \cdot k = \sum_{\ell=1}^k \sum_{i \in V} \left(\sqrt{1 + \beta^2 |\nabla_w u_\ell^*|_i^2} - 1 \right) + \frac{\alpha}{q} \|u_\ell^*\|_{V_\ell^e, q}^q.$$

In this section and in Section 6, we present results based on a Matlab implementation of the algorithms described in Sections 3 and 4. All computations were done on a 16-core 2.90 GHz Intel® Xeon® desktop computer with 128 GB of RAM.

5.1. Line graph. We start by constructing a line graph—the discretization of a line segment—with 20 vertices. We fix the parameter $\beta = 1$ and consider the bi-partition $k = 2$. A partition $\sqcup_\ell V_\ell$ is randomly initialized and our partitioning algorithm is applied for different values of α and q , respectively. The evolution of the partitions and their minimizing u_ℓ is illustrated in Figure 3. These results suggest that the parameter q , which is the norm which penalizes the violation of the Dirichlet boundary condition, does not heavily affect the minimal partition, while the parameter α , which is the coefficient of the penalization, plays an important role: too large an α effectively pins the minimizing u_ℓ in local minima, while too small an α doesn’t enforce the desired boundary conditions. In this simple example, the iterative algorithm converges in only 5 outer iterations.

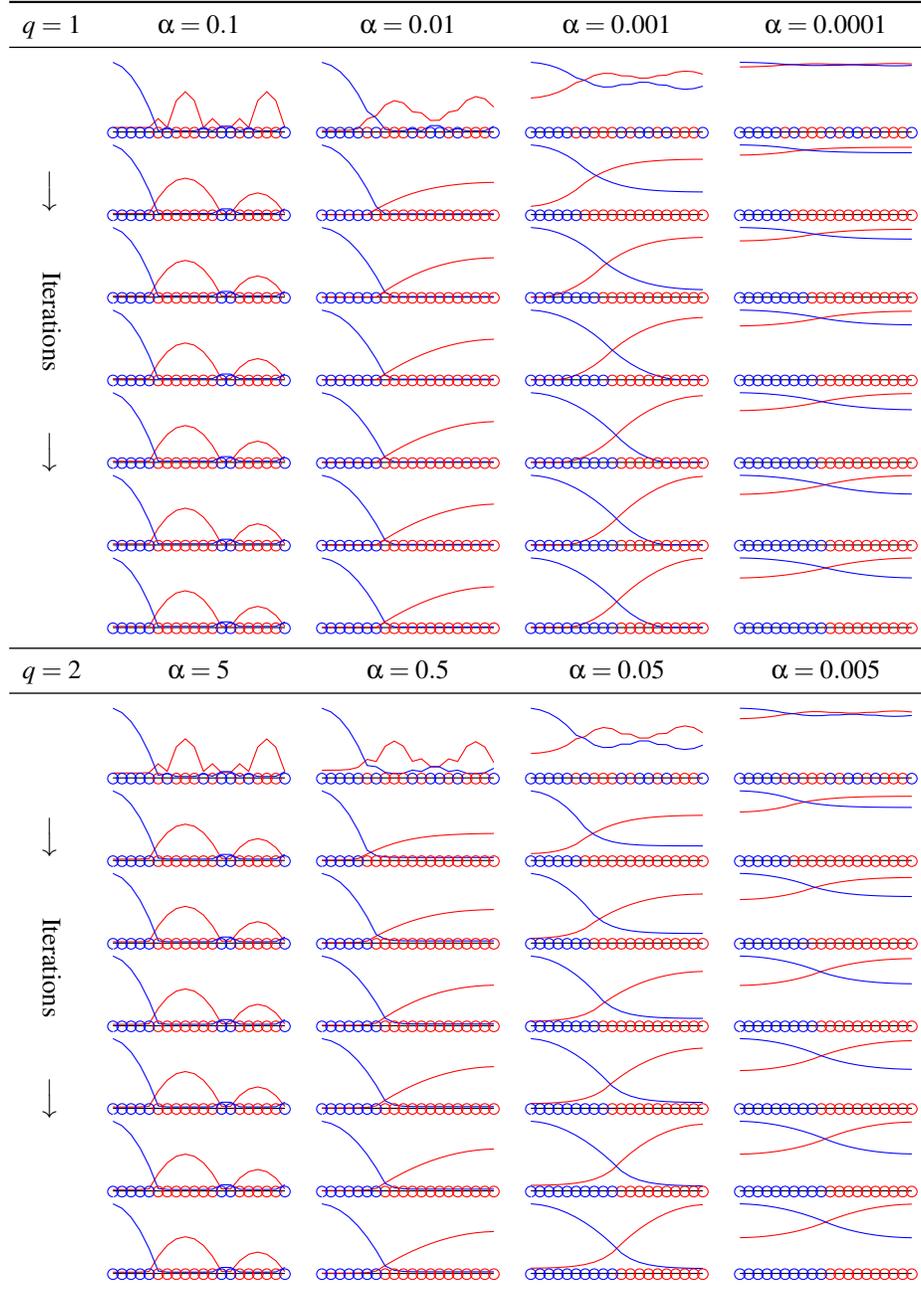


FIGURE 3. Partitioning a line segment, for different α and q . In each column, the topmost subplot shows the initial (random) partition (colored circles) and the corresponding optimal u_i . From one row to the next, the partition is rearranged according to §3.5, and new eigenfunctions u_i are shown. For large α , the partition remains stuck in a local minima. For α too small, the u_i become too flat. Results for $q = 1$ and $q = 2$ are qualitatively similar. See §5.1.

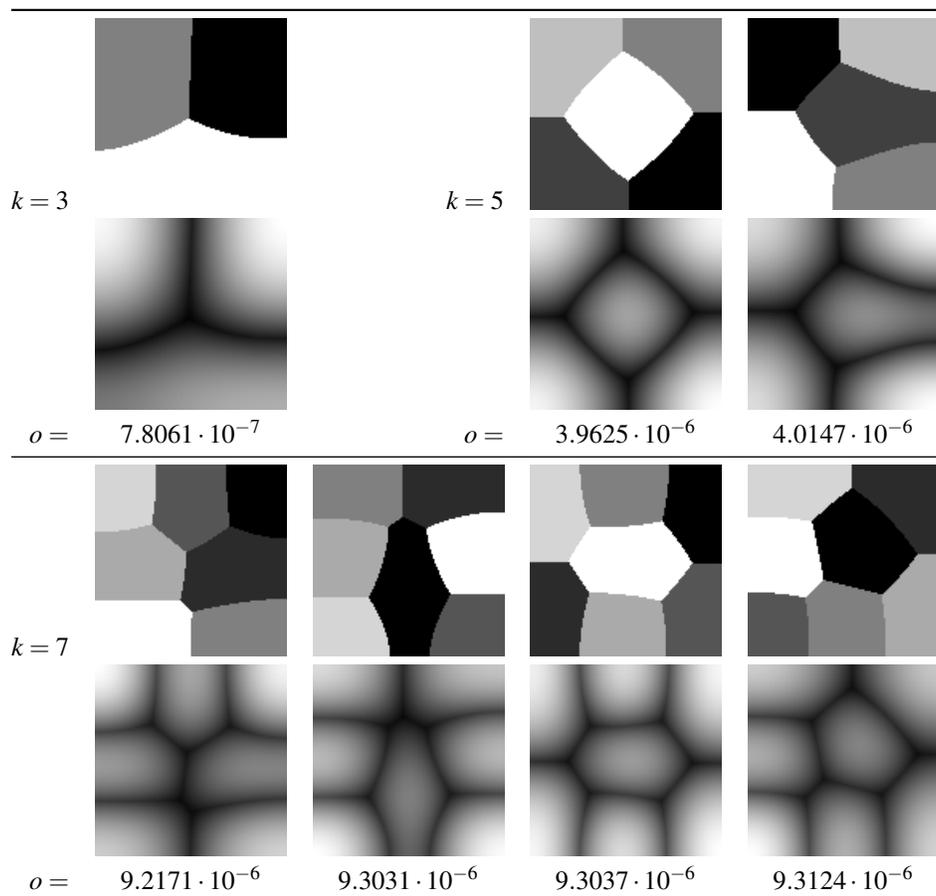


FIGURE 4. Locally optimal partitions (top) and the associated ground state composites (bottom) for the square lattice graph for various values of k . Here $\beta = 2$, $q = 1$ for all k and $\alpha = 5 \cdot 10^{-7}$ for $k = 3$, and $\alpha = 10^{-6}$ for $k = 5, 7$. See §5.2.

5.2. Square lattice graph. We consider a square domain with free boundary conditions. To this end we construct a 120×120 square grid and a graph where each (non-boundary) vertex is connected to its four nearest neighbors; vertices at the edge are connected to their three nearest neighbors, and the four corner vertices are connected to their two nearest neighbors. Edge weights w are taken to be constant. For this graph, the optimal partitions are non-trivial. In Figure 4, we show some (locally) optimal partitions for $k \in \{3, 5, 7\}$, computed with our algorithm. We use constant $\beta = 4$ and $q = 1$ for all k , while the parameter α increases with the number of partitions.

These results can be compared to the numerical experiments conducted in [4, 27] for optimal spectral partitions ($\lim \beta \rightarrow 0$). Interestingly, we find that the partitions obtained for finite β are very similar to the optimal spectral partitions. Namely, for $k = 3$, there appears to be a continuous family of 3-partitions with all components meeting at a triple-junction with similar values of energy. The partition for $k = 5$ with smallest energy is similar to a spectral partition in [27]. Some of the obtained partitions also bear striking

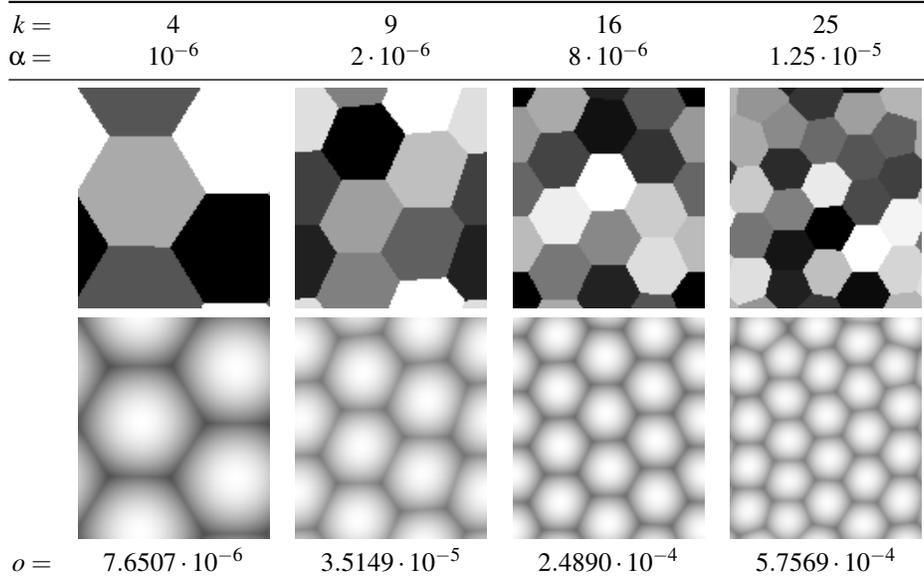


FIGURE 5. Locally optimal partitions of a torus graph (top) and the associated ground states (bottom). In all cases, $\beta = 4$ and $q = 1$. See §5.3.

resemblance to steady-state configurations of Renyi-entropy minimizing space partitioning processes [14, Fig. 5].

5.3. Torus graph (square lattice graph with periodic boundary conditions). We consider the 110×127 torus graph, obtained from a truncated square lattice by identifying opposing “boundary vertices”. The aspect ratio $2 : \sqrt{3}$ is chosen such that regular hexagons can potentially tile the manifold with correct periodicity. The edge weights, w , are taken to be constant. We initialize the partition with a Voronoi tessellation for randomly chosen generators. In Figure 5, we plot the $k = 4, 9, 16$, and 25 partitions computed for constant $\beta = 4$ and $q = 1$, while the parameter α increases with the number of partitions.

To illustrate the periodicity of the minimizers, we also show a tiling of the obtained partitions in Figure 6. Indeed, for $k \in (2\mathbb{N})^2$, the two-periodic square partitioning problem admits a tiling of regular hexagons and our numerical experiments suggest that this is the optimal partition for the Beltrami energy. It has been proven that the hexagonal tiling minimizes perimeter subject to a fixed volume constraint [26]. It has also been conjectured that the optimal partitions for the Dirichlet objective function are hexagonal [10]; see also [6, 41] for numerical support of this conjecture.

5.4. Discretizations of a barbell and annulus. After having considered domain partitioning with free and periodic boundary conditions, we now make use of the semi-supervised extension to partition subdomains of the square with Dirichlet boundary conditions. To this end, we again construct a square grid with edge weights as in Section 5.2. In addition, we assign the label $0 \notin 1, \dots, k$ to vertices outside the intended domain, and keep these labels unchanged during the rearrangement step.

As a first example, we look at the “pinched oval” or “barbell” domain as illustrated in Figure 1. We compute locally optimal partitions for $k = 3, 4$ and 5 . Representative local minimizers are shown in Figure 7.

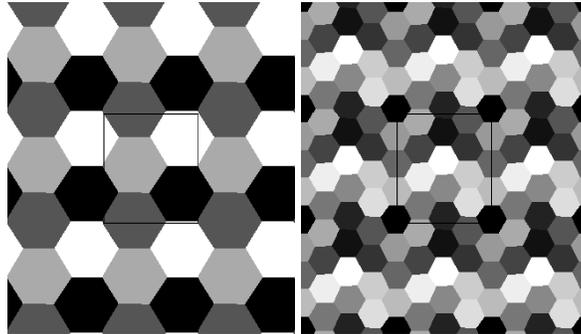


FIGURE 6. Lattice partitioning as tilings of torus partitions, based on $k = 4, 16$. For $k \in (2\mathbb{N})^2$, the two-periodic (torus) partition problem mimics the two-dimensional lattice problem. See §5.3.

Our second example is a punctured disk: an annulus with varying size hole. Depending on the relative size of the hole, the optimal partitions differ significantly, as shown in Figure 8, for $k = 8$. For a relatively large hole radius $r = 0.5$ (half the disk radius), the eight locally optimal partitions are simply radially distributed on the ring. For very small puncture $r = 0.05$, on the other hand, we find a less symmetric partition. As a matter of fact, for small hole size and too small α , a single partition can “surround” the puncture. The most interesting locally optimal partitions are obtained for intermediate hole sizes around $r = 0.1$ and $r = 0.2$, with non-trivial symmetrical patterns emerging.

5.5. Periodic volume partitioning. We consider partitions of a triply-periodic volume. We recall the triply-periodic minimal surface found by Hermann Schwarz [49] referred to as the *P*-surface (“primitive”) and *approximated* implicitly by the zero level-set of the function

$$(x, y, z) \mapsto \cos(x) + \cos(y) + \cos(z). \quad (39)$$

This minimal surface bi-partitions the volume into two equally-sized components such that the interface is triply periodic and necessarily has zero mean curvature.

We consider the $26 \times 26 \times 26$ torus graph, corresponding to the three-dimensional extension of the two-dimensional torus graph previously considered in Section 5.3. The edge weights, w , are taken to be constant. Starting from a randomly initialized partition, our algorithm typically converges to one of two local optima, shown in Figure 9: a planar bisection of the cube, or a surface very similar to the *P*-surface. The *P*-surface minimum has smaller objective function value than the planar bisection. For comparison, we also plot the zero level-set of the function in (39), which very well approximates the *P*-surface.

This result is interesting because the *P*-surface is a stationary point for an optimization problem which only agrees with our proposed partitioning model in the TV limit ($\beta \rightarrow \infty$); see (7). The fact that our computations suggest that this surface is also optimal for finite values of β suggests that the *P*-surface is a minimizer of a larger class of optimization problems.

6. Experiments II: Data graphs. We now turn our attention to the clustering of data-sets via the partitioning of associated weighted (non-lattice) graphs.

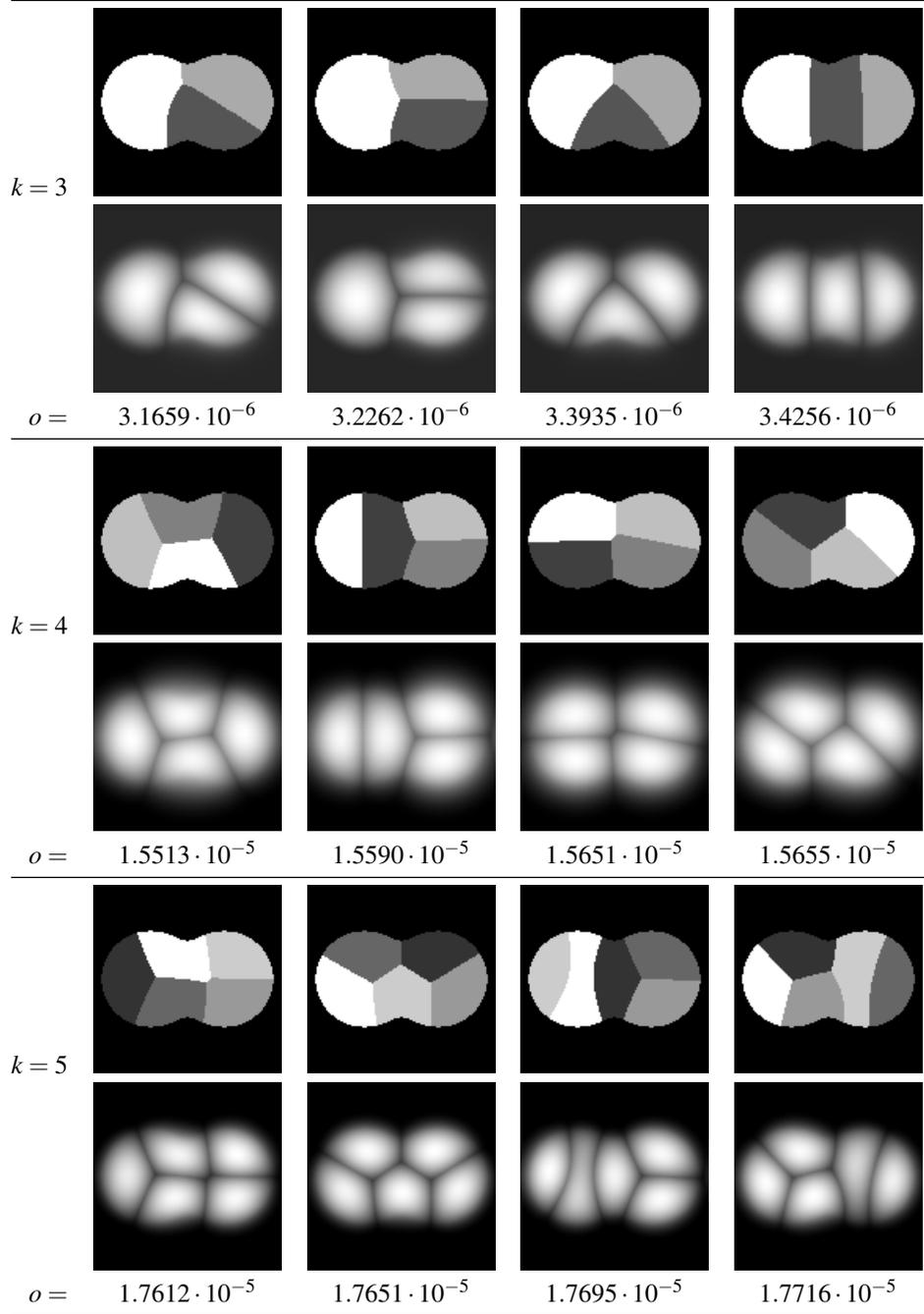


FIGURE 7. Minimizing partitions (top) and a composite of ground state functions (bottom) of a discretized “barbell” domain with Dirichlet boundary conditions. First row: $k = 3$, $\alpha = 0.02$, $q = 2$, $\beta = 2$. Middle row: $k = 4$, $\alpha = 2 \cdot 10^{-6}$, $q = 1$, $\beta = 2$. Bottom row: $k = 5$, $\alpha = 2 \cdot 10^{-6}$, $q = 1$, $\beta = 1$. See §5.4.

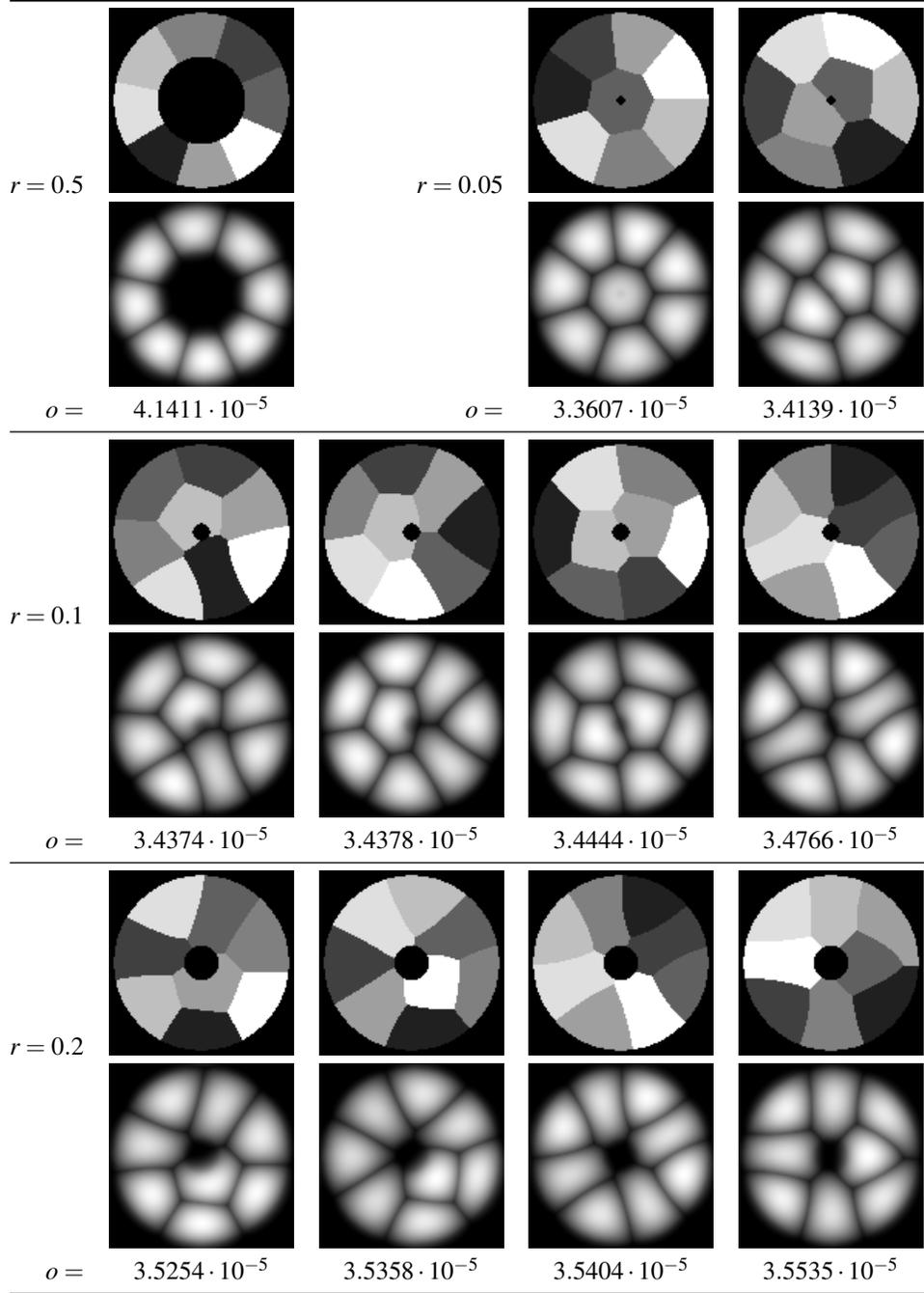


FIGURE 8. Minimizing partitions (top) and a composite of ground state functions (bottom) of a discretized annulus domain with Dirichlet boundary conditions and varying inner radius r . Here $k = 8$, $\alpha = 3 \cdot 10^{-6}$, $q = 1$, and $\beta = 1$. See §5.4.

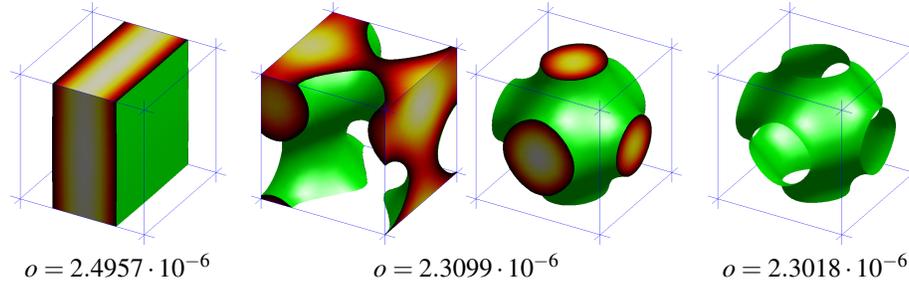


FIGURE 9. Partitions of a $26 \times 26 \times 26$ discretization of a triply-periodic three-dimensional volume. Here $k = 2$, $\alpha = 10^{-6}$, $q = 1$, and $\beta = 1$. **(left)** Interface (green) and one of the ground states, u , (heat map) of a planar bisection. **(center)** Interface and one of the ground states, u , of a “P-surface-like” partition, before and after re-centering for visualization. **(right)** The zero level-set of $\cos(x) + \cos(y) + \cos(z)$, which very well approximates the P-surface. See §5.5.

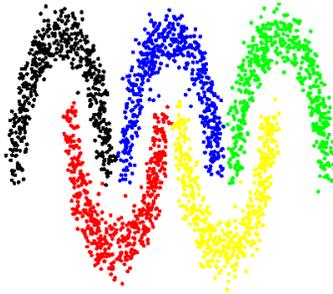


FIGURE 10. Optimal partitions for the five moons problem. Here $q = 2$, $\beta = 2$, $\alpha = 0.1$, and a 99.6% purity is attained. See §6.1.

6.1. Five moons. We begin by considering the five moons partitioning problem, a collection of points in the shape of five moons. We assign edge weights using a Gaussian kernel, $w(i, j) = e^{-d^2(x_i, x_j)/\sigma^2}$, where $d(\cdot, \cdot)$ is Euclidean distance and $\sigma = 1$. In Figure 10, we give a scatter plot of the points where colors represent the labels assigned to each point by the algorithm. The method attains a 99.6% purity value.

6.2. MNIST handwritten digits. The MNIST handwritten digit dataset consists of 70,000 28×28 greyscale images of handwritten digits 0 to 9. As input we used the similarity matrix for the MNIST dataset obtained from the website of Z. Yang [57, 58]. We symmetrize this matrix via $\tilde{W}_{ij} = \max\{W_{ij}, W_{ji}\}$.

First, we perform naive clustering with our proposed algorithm. The purity obtained, as defined in [57], differs a lot between individual runs, and we show the best result (in terms of purity) over several dozen runs without supervision. In each case, the algorithm converges in approximately 20 iterations. For parameters $q = 1$, $\beta = 2$, $\alpha = 10^{-5}$, the best observed purity (without supervision) is 96.23% which is comparable to the performance of state-of-the-art clustering algorithms.

Figure 11 is a graphical display of the quality of the output. The first row shows the blurry averaged images within each cluster, while the second row indicates the representative images for each cluster (where each ground state achieves its maximum). In this particular run, the representative image of the ‘3’ cluster is actually an ‘8’, indicative of the strong ambiguity between these classes in terms of graph separation. The second best representative of this class (where the ground state is second largest), is truly a ‘3’ and shown as 3bis. To further illustrate the diversity of handwriting samples, for each cluster we show the misclassified image with largest ground state value along with its true label. Visual inspection shows that the misclassification is “intuitively explainable”.

In Table 1, we display the confusion matrix for the obtained clusters, both in absolute and relative numbers. The columns represent the ground truth labels and the rows represent the labels assigned by the proposed algorithm. Each column represents the distribution of the true labels across our partitions. Again the confusion matrix highlights the difficulties in the classification seen above.

In addition to the unsupervised clustering, we perform instances of transductive learning by making use of the semi-supervised algorithm extension (§4.6) and the true labels for a small portion of the digit images. We apply the rearrangement algorithm to the MNIST dataset, using $n = 20$ random initializations and an increasing percentage of labelled data points. See Figure 12 for a scatter plot of attained purity and objective function, and Table 2 for numerical values. For each percentage of labels, we report the lowest, average, and highest purity, and the lowest, average, and highest objective function value, as well as the purity of the lowest objective function partition (which would be the reasonable best guess). The objective function value for the ground truth labels is $1.8913 \cdot 10^{-4}$. We observe that in the unsupervised case, state-of-the-art purity can be achieved, but overall the results are very diverse. As soon as a few true labels are provided, the spread is dramatically reduced and average purity greatly improved. We note that low objective function value does not necessarily signify high purity. In particular, the ground truth labels have higher objective function value than any of the 2.5, 5, or 10% semi-supervised results.

6.3. Image Segmentation. We finally consider the formulation of image segmentation as a graph partitioning problem. Automatic analysis of images often requires the identification of coherent regions of interest via efficient and robust algorithms, a task called *image segmentation*. Due to the non-locally repetitive nature of the patterns in natural images, correlations between pairs of pixels extend well beyond just neighboring pixels. Therefore, graphs are a natural representation: we identify pixels of the image with vertices V , while edges E model the correlations between any pairs of pixels. Thereby, image segmentation effectively becomes a graph partitioning task of the graph $G = (V, E)$.

Region-based image segmentation looks for homogeneity (read: least intensity variance) within the partitions. The canonical region-based image segmentation model is the Chan-Vese model (CV), [12], which strives for region homogeneity while keeping the region’s interface short:

$$J_{CV}[\mu_1, \mu_2, \phi] = \int_{\Omega} (\mu_1 - I)^2 H(\phi) + \int_{\Omega} (\mu_2 - I)^2 (1 - H(\phi)) + \beta \int_{\Omega} |\nabla H(\phi)|, \quad (40)$$

where μ_i denotes the mean feature of each region, the levelset function $\phi: \Omega \rightarrow \mathbb{R}$ is positive in object regions, negative in background regions, zero on the object boundaries, and H is the Heaviside function. The last, total variation term is the co-area-formula equivalent of the boundary length, and its coefficient β balances the data-fidelity versus boundary regularity.

	0	1	2	3	4	5	6	7	8	9
0	6841	1	38	3	3	22	25	2	21	19
1	5	7812	56	19	49	4	17	88	73	23
2	6	22	6692	44	3	3	1	22	20	7
3	1	6	18	6845	1	137	0	1	138	92
4	1	9	6	2	6510	6	5	20	29	57
5	9	0	7	88	1	6019	41	2	188	11
6	28	4	12	4	25	69	6785	0	27	4
7	4	12	121	52	12	8	0	7040	16	150
8	4	2	34	52	4	11	2	4	6248	24
9	4	9	6	32	216	34	0	114	65	6571
Σ	6903	7877	6990	7141	6824	6313	6876	7293	6825	6958
	0	1	2	3	4	5	6	7	8	9
0	99.10	0.01	0.54	0.04	0.04	0.35	0.36	0.03	0.31	0.27
1	0.07	99.17	0.80	0.27	0.72	0.06	0.25	1.21	1.07	0.33
2	0.09	0.28	95.74	0.62	0.04	0.05	0.01	0.30	0.29	0.10
3	0.01	0.08	0.26	95.85	0.01	2.17	0.00	0.01	2.02	1.32
4	0.01	0.11	0.09	0.03	95.40	0.10	0.07	0.27	0.42	0.82
5	0.13	0.00	0.10	1.23	0.01	95.34	0.60	0.03	2.75	0.16
6	0.41	0.05	0.17	0.06	0.37	1.09	98.68	0.00	0.40	0.06
7	0.06	0.15	1.73	0.73	0.18	0.13	0.00	96.53	0.23	2.16
8	0.06	0.03	0.49	0.73	0.06	0.17	0.03	0.05	91.55	0.34
9	0.06	0.11	0.09	0.45	3.17	0.54	0.00	1.56	0.95	94.44
Σ	100 %	100 %	100 %	100 %	100 %	100 %	100 %	100 %	100 %	100 %

TABLE 1. MNIST confusion matrix. The columns represent the ground truth labels and the rows represent the labels assigned by the proposed algorithm. Each column sums to the number of digits in each classe/one and represents the distribution of the true labels across our partitions. We see that the algorithm does generally well, but exhibits some confusion among the digits (3, 5, 8), (4, 9) and (2, 7). See §6.2.

fixed labels	purity [%]			objective function [10^{-4}]			guess
	min.	avg.	max.	min.	avg.	max.	
0 %	64.31	84.32	95.84	1.6260	1.8833	1.9992	78.16
1 %	78.66	88.74	96.24	1.6577	1.8823	1.9331	87.64
2.5 %	95.88	96.21	96.32	1.8873	1.8883	1.8897	96.28
5 %	96.37	96.41	96.47	1.8874	1.8878	1.8884	96.42
10 %	96.61	96.64	96.70	1.8870	1.8878	1.8884	96.66
GT: 100 %	100.00			1.8913			n/a

TABLE 2. Purity and objective function for MNIST data set, with $n = 20$ runs per fraction of fixed labels. $q = 1$, $\beta = 2$, $\alpha = 10^{-5}$. The *guess*-column shows the purity of the run with least objective function value. See §6.2.

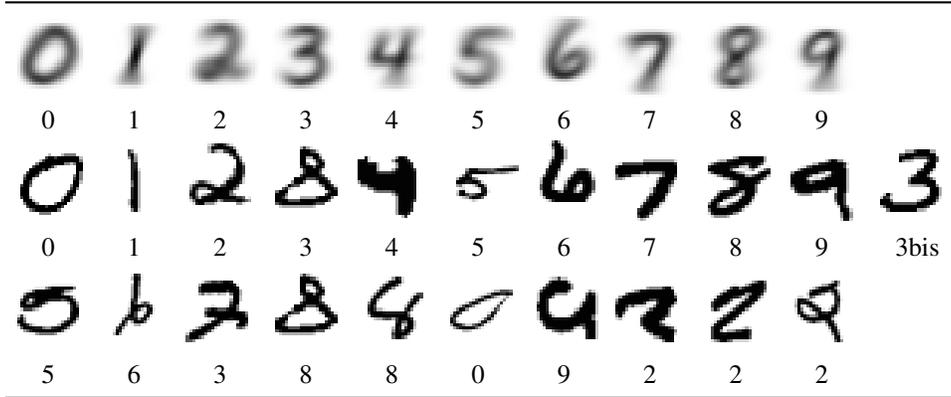


FIGURE 11. MNIST handwritten digits; each image is 28×28 pixels. **(top)** The cluster means. **(middle)** Results of best fully unsupervised clustering: $q = 1$, $\beta = 2$, $\alpha = 10^{-5}$, purity= 96.23%. Representative images for each of the $k = 10$ clusters (where u_i is maximized). The representative image for the ‘3’ cluster is actually an ‘8’; the best true ‘3’ is shown at the far right. See §6.2. **(bottom)** Examples of confused images, wrongly assigned to a cluster (true label given below).

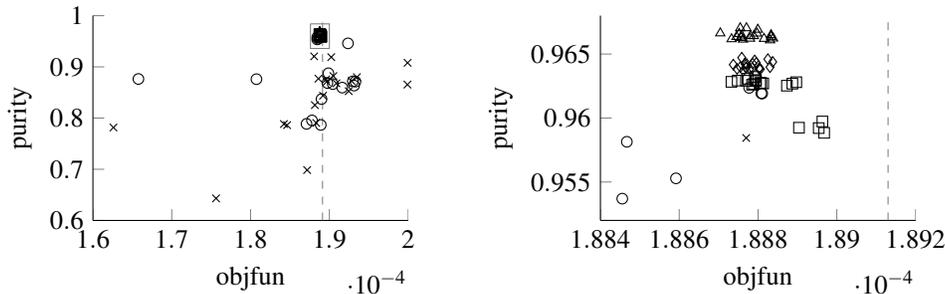


FIGURE 12. MNIST purity versus objective function o , for $q = 1$, $\beta = 2$, $\alpha = 10^{-5}$. **(left)** Purity and objective function value for 20 separate runs for each different fraction of semisupervised vertex labels: $\times = 0\%$, $\circ = 1\%$, $\square = 2.5\%$, $\diamond = 5\%$, $\triangle = 10\%$. **(right)** Detail of gray box region. Vertical dashed line indicates objective function of ground truth partition. See §6.2.

A simple graph partitioning model corresponding to such region-based image segmentation (including the multiphase case $k > 2$) can be obtained through extension of the square-lattice graph by *strengthening the weight between spatially distant but feature-wise similar pixels*:

$$w_{ij} := \begin{cases} \frac{p}{\varepsilon + \lambda \|I(j) - I(i)\|^2} + (1 - p), & \text{if } j \in N(i) \\ \frac{p}{\varepsilon + \lambda \|I(j) - I(i)\|^2}, & \text{otherwise} \end{cases} \quad \forall i, j \in V, \quad (41)$$

where $N(i)$ denotes the 4-connected von Neumann neighborhood of pixel i , $I(i) \in \mathbb{R}^d$ denotes the value of pixel i , $\|I(j) - I(i)\|^2$ is the squared L^2 -distance of the pixels in feature space, and $p \in (0, 1)$ balances the relative weight of the region-homogeneity term vs versus

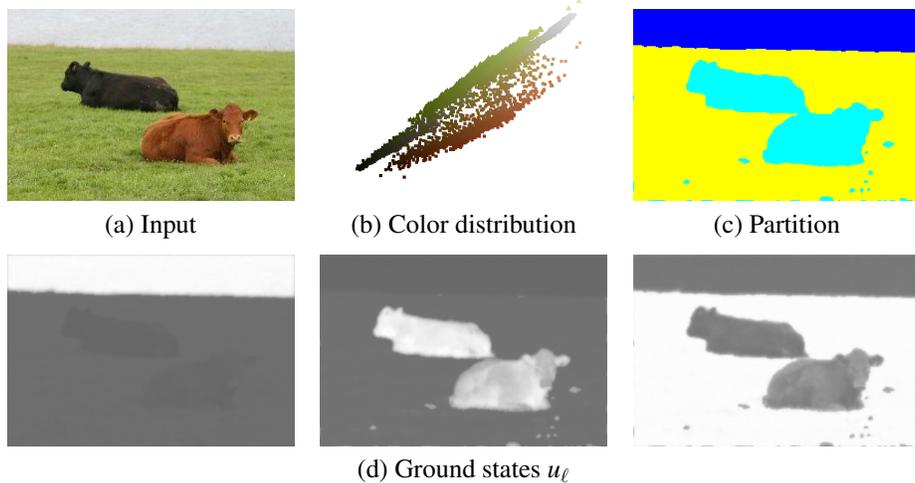


FIGURE 13. Region-based image segmentation as a graph partitioning problem: (a) 318×212 RGB input image. (b) Scatterplot of the pixel distribution in color space. The green (\triangle) and gray (\circ) background form small, dense clusters, while the cows black and brown shades are much wider spread (\square). (c)–(d) Image segmentation results using graph partitioning parameters $k = 3$, $\alpha = 0.1$, $\beta = 1$, $q = 2$. See §6.3

the boundary length regularization stemming from the lattice graph. The optimal partition aligns flat regions of u with homogeneous regions in the image, and gradients of u with gradients in I . A fully connected graph is not practicable and we uniformly subsample the non-local connectivities at a very low rate.

As an example we consider segmenting the 318×212 RGB input image, $I: V \rightarrow [0, 1]^3$, shown in Figure 13(a). From this image, we construct a graph using (41) with parameters $p = 10^{-4}$, $\varepsilon = 0.01$, and $\lambda = 15$. The long range connections are sub-sampled at about 80 edges per pixel. In Figure 13(b), we give a scatterplot of the pixel distribution in color space. We apply the graph partitioning algorithm with parameters $k = 3$, $\alpha = 0.1$, $\beta = 1$, and $q = 2$. The resulting segmented image is shown in Figure 13(c) and the associated ground states are shown in Figure 13(d).

The color-based image segmentation problem shown here is, of course, more efficiently solved with state-of-the-art segmentation models. However, it is easy to see how the current graph construction (41) can be generalized to more interesting feature distances (*e.g.*, patch-based or using other texture descriptors) that greatly extend beyond the Chan-Vese model. Also, by non-uniformly sub-sampling the long-range connectivities, we can render the region-based segmentation more local and can thus seamlessly deal with images affected by bias, for example as in [7, 33, 61].

7. Discussion and further directions. In this paper we introduced a novel non-convex graph partitioning objective based on the sum of the minimal discrete Beltrami energies of the partition components. This objective is a discrete analogue of an energy that appears in classical minimal surface problems. We adapt primal-dual convex optimization methods to solve for the minimal Beltrami energy for each component of a given partition and a rearrangement algorithm is proposed to find locally optimal graph partitions. The

methods are illustrated for a variety of example graph partitioning problems in Sections 5 and 6. Similar to partitions obtained from minimizing the Dirichlet energy sum of partition components [41], the objective proposed here naturally provides confidences for label assignments and consequently produces a representative for each cluster.

We view the Beltrami energy considered here as an interesting generalization and interpolation between the Total Variation and Dirichlet energies that have previously been used for graph partitioning (see (7)). In fact, we observe in computational experiments that several of the optimal partitions obtained are insensitive to the interpolation parameter β ; our experiments support the hypothesis that certain partitions are optimal for a class of objective functions. This phenomena has recently been studied for certain point-set configurations on a sphere which minimize a class of energies which generalize the inverse squared energy appearing in Thompson’s problem [13]. Configurations that minimize such classes of energies are referred to as “universally optimal”. It wouldn’t be surprising if certain partitions were also minimizers for a class of objective functions including the Beltrami energy.

It is interesting to observe that many of the manifold partitions in Figures 4–8 feature 120° triple junctions, even in cases where a 90° quadruple junction would be more symmetrical; see, *e.g.*, $k = 4$ in Figure 7. This behavior is well known in crystal grain growth [30] and has typically been associated with stable configurations of interface motion by mean curvature [38, 39, 48]. This observation further strengthens the intuition that the bi-level rearrangement algorithm employed in this work very closely relates to the MBO scheme and motion by mean-curvature stemming from diffusion-threshold dynamics more generally.

The proposed computational methods for solving the bi-level optimization problem (11) are based on primal-dual optimization methods for the inner problem and a rearrangement algorithm for the outer problem. We view this application as an interesting demonstration of the power and flexibility of primal-dual methods for solving a broad class of convex optimization problems. However, in the present application, the inner problem is the computational bottleneck and it seems doubtful that algorithmic advances will make partitioning methods requiring the (approximate) solution of such non-local energies competitive with more local strategies [2, 55]. However, it is plausible that many local heuristics for partitioning can be viewed as approximations of partitions which minimize a non-local energy, such as the one studied here.

Many interesting research directions remain. One obvious direction would be to formulate the analogous partitioning problem in the continuum. Proving the existence and studying properties of optimal partitions of surfaces is likely to be challenging. The rearrangement algorithm proposed in Section 3.5 has the same structure as expectation-maximization (EM) algorithms. Making this connection more concrete, *i.e.*, defining the correct probability distribution, is a very interesting direction of pursuit. Finally, there is an intriguing relationship between the Dirichlet partitioning problem and an antagonistically-interacting random walker Model [14, 15, 40, 41]. It would be of interest if there was a connection between a non-linear diffusion process, such as the one studied in [53], with the Beltrami partitioning problem.

Acknowledgments. The authors would like to gratefully acknowledge valuable conversations with Stan Osher, Luminita Vese, Jing An, Chris D. White, and Mengqi (Mandy) Xia.

REFERENCES

- [1] K. J. Arrow, L. Hurwicz and H. Uzawa, *Studies in linear and non-linear programming*, Cambridge Univ. Press, 1958.
- [2] David A. Bader, Henning Meyerhenke, Peter Sanders and Dorothea Wagner (eds.), *Graph partitioning and graph clustering*, American Mathematical Society, 2013.
- [3] D. Barash, Fundamental relationship between bilateral filtering, adaptive smoothing, and the nonlinear diffusion equation, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **24** (2002), 844–847.
- [4] V. Bonnaillie-Noël and B. Helffer, Numerical analysis of nodal sets for eigenvalues of Aharonov-Bohm Hamiltonians on the square with application to minimal partitions, *Experimental Mathematics*, **20** (2011), 304–322.
- [5] V. Bonnaillie-Noël, B. Helffer and G. Vial, Numerical simulations for nodal domains and spectral minimal partitions, *ESAIM: Control, Optimisation and Calculus of Variations*, **16** (2008), 221–246.
- [6] B. Bourdin, D. Bucur and E. Oudet, Optimal partitions for eigenvalues, *SIAM Journal on Scientific Computing*, **31** (2010), 4100–4114.
- [7] T. Brox and D. Cremers, On local region models and a statistical interpretation of the piecewise smooth Mumford-Shah functional, *International Journal of Computer Vision*, **84** (2008), 184–193.
- [8] D. Bucur, G. Butazzo and A. Henrot, Existence results for some optimal partition problems, *Adv. Math. Sci. Appl.*, **8** (1998), 571–579.
- [9] D. Bucur and B. Velichkov, Multiphase shape optimization problems, *SIAM Journal on Control and Optimization*, **52** (2014), 3556–3591.
- [10] L. A. Cafferelli and F. H. Lin, An optimal partition problem for eigenvalues, *Journal of Scientific Computing*, **31** (2007), 5–18.
- [11] A. Chambolle and T. Pock, A first-order primal-dual algorithm for convex problems with applications to imaging, *Journal of Mathematical Imaging and Vision*, **40** (2011), 120–145.
- [12] T. F. Chan and L. A. Vese, Active contours without edges, *IEEE Transactions on Image Processing*, **10** (2001), 266–277.
- [13] H. Cohn and A. Kumar, Universally optimal distribution of points on spheres, *Journal of the American Mathematical Society*, **20** (2006), 50.
- [14] O. Cybulski, V. Babin and R. Holyst, Minimization of the Renyi entropy production in the space-partitioning process, *Physical Review E*, **71** (2005), 046130.
- [15] O. Cybulski and R. Holyst, Three-dimensional space partition based on the first Laplacian eigenvalues in cells, *Physical Review E*, **77** (2008), 056101.
- [16] L. Dascal, G. Rosman and R. Kimmel, Efficient Beltrami filtering of color images via vector extrapolation, in *SSVM'07: Proceedings of the 1st international conference on Scale space and variational methods in computer vision*, Springer-Verlag, 2007, 92–103.
- [17] A. P. Dempster, N. M. Laird and D. B. Rubin, Maximum likelihood from incomplete data via the EM algorithm, *Journal of the Royal Statistical Society. Series B (Methodological)*, **39** (1977), 1–38.
- [18] J. Duchi, S. Shalev-Shwartz, Y. Singer and T. Chandra, Efficient projections onto the l_1 -ball for learning in high dimensions, in *Proceedings of the 25th international conference on Machine learning - ICML '08*, ACM Press, New York, New York, USA, 2008, 272–279.
- [19] I. Ekeland and R. Temam, *Convex analysis and variational problems*, North-Holland Publishing Company, Amsterdam, 1976.
- [20] S. Eshedoglu and F. Otto, Threshold dynamics for networks with arbitrary surface tensions, *Communications on Pure and Applied Mathematics*, **68** (2015), 808–864.
- [21] E. Esser, X. Zhang and T. F. Chan, A general framework for a class of first order primal-dual algorithms for convex optimization in imaging science, *SIAM Journal on Imaging Sciences*, **3** (2010), 1015–1046.
- [22] M. Feigin and N. Sochen, Anisotropic regularization for inverse problems with application to the Wiener filter with Gaussian and impulse noise, in *Scale Space and Variational Methods in Computer Vision*, vol. 5567 of Lecture Notes in Computer Science, Springer Berlin / Heidelberg, 2009, 319–330.
- [23] R. Gabbriellini, A new counter-example to Kelvin’s conjecture on minimal surfaces, *Philosophical Magazine Letters*, **89** (2009), 483–491.
- [24] C. Garcia-Cardona, E. Merkurjev, A. L. Bertozzi, A. Flenner and A. G. Percus, Multiclass data segmentation using diffuse interface methods on graphs, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **36** (2014), 1600–1613.
- [25] E. Giusti, *Minimal surfaces and functions of bounded variation*, Springer Science & Business Media, 1984.
- [26] T. C. Hales, The honeycomb conjecture, *Discrete & Computational Geometry*, **25** (2001), 1–22.
- [27] B. Helffer, On spectral minimal partitions: a survey, *Milan J. Math.*, **78** (2010), 575–590.

- [28] B. Helffer and T. Hoffmann-Ostenhof, Remarks on two notions of spectral minimal partitions, *Adv. Math. Sci. Appl.*, **20** (2010), 249–263.
- [29] B. Helffer, T. Hoffmann-Ostenhof and S. Terracini, On spectral minimal partitions: the case of the sphere, 2010, 153–178.
- [30] C. Herring, Some theorems on the free energies of crystal surfaces, *Physical Review*, **82** (1951), 87–93.
- [31] R. Kaftory, N. A. Sochen and Y. Y. Zeevi, Variational blind deconvolution of multi-channel images, *Int. J. Imaging Syst. Technol.*, **15** (2005), 56–63.
- [32] R. Kimmel, R. Malladi and N. Sochen, Images as embedded maps and minimal surfaces: movies, color, texture, and volumetric medical images, *Int. J. Comput. Vis.*, **39** (2000), 111–129.
- [33] C. Li, C.-Y. Kao, J. C. Gore and Z. Ding, Minimization of region-scalable fitting energy for image segmentation, *IEEE Transactions on Image Processing*, **17** (2008), 1940–1949.
- [34] Z. Liang and Y. Li, Beltrami flow in Hilbert space with applications to image denoising, *Journal of Electronic Imaging*, **21** (2012), 043019 1–11.
- [35] S. Lloyd, Least squares quantization in PCM, *IEEE Transactions on Information Theory*, **28** (1982), 129–137.
- [36] L. Lopez-Perez, R. Deriche and N. Sochen, The Beltrami flow over triangulated manifolds, in *CVAMIA and MMBIA Workshop at ECCV 2004*, 2004, 135–144.
- [37] E. Merkurjev, T. Kostic and A. L. Bertozzi, An MBO scheme on graphs for segmentation and image processing, *SIAM J. Imaging Sciences*, **6** (2013), 1903–1930.
- [38] B. Merriman, J. K. Bence and S. Osher, Diffusion generated motion by mean curvature, *UCLA CAM Report 92-18*, 1–12.
- [39] B. Merriman, J. K. Bence and S. Osher, Diffusion generated motion by mean curvature, *AMS Selected Letters, Crystal Grower’s Workshop*, 73–83.
- [40] B. Osting and C. D. White, Nonnegative matrix factorization of transition matrices via eigenvalue optimization, in *NIPS OPT*, 2013.
- [41] B. Osting, C. D. White and E. Oudet, Minimal Dirichlet energy partitions for graphs, *SIAM Journal on Scientific Computing*, **36** (2014), A1635–A1651.
- [42] E. Oudet, Approximation of partitions of least perimeter by Γ -convergence: around Kelvin’s conjecture, *Experimental Mathematics*, **20** (2011), 260–270.
- [43] J. Plateau, *Statique expérimentale et théorique des liquides soumis aux seules forces moléculaires*, Gauthier-Villars, Paris, 1873.
- [44] G. Rosman, X.-C. Tai, L. Dascal and R. Kimmel, *Polyakov action for efficient color image processing*, Technical Report 623, Technion CIS-2010-04, 2010.
- [45] A. Roussos and P. Maragos, Tensor-based image diffusions derived from generalizations of the total variation and Beltrami functionals, in *2010 IEEE International Conference on Image Processing*, IEEE, 2010, 4141–4144.
- [46] L. Rudin, S. Osher and E. Fatemi, Nonlinear total variation based noise removal algorithms, *Physica D*, **60** (1992), 259–268.
- [47] C. Sagiv, N. A. Sochen and Y. Y. Zeevi, Gabor feature space diffusion via the minimal weighted area method, in *Energy Minimization Methods in Computer Vision and Pattern Recognition*, 2001, 621–635.
- [48] H. Schaeffer and L. Vese, Variational dynamics of free triple junctions, *Journal of Scientific Computing*, **59** (2013), 386–411.
- [49] H. A. Schwarz, *Gesammelte mathematische Abhandlungen*, Springer Berlin Heidelberg, Berlin, Heidelberg, 1890.
- [50] W. Sir Thomson, On the division of space with minimum partitional area, *Acta Mathematica*, **11** (1887), 121–134.
- [51] N. Sochen, R. Kimmel and R. Malladi, A general framework for low level vision, *IEEE Transactions on Image Processing*, **7** (1998), 310–318.
- [52] N. Sochen, R. Deriche and L. Lopez-Perez, The Beltrami flow over implicit manifolds, in *9th IEEE ICCV*, 2003, 832–839.
- [53] N. A. Sochen, Stochastic processes in vision: from Langevin to Beltrami, *Proceedings Eighth IEEE International Conference on Computer Vision. ICCV 2001*, **1** (2001), 288–293.
- [54] A. Spira, R. Kimmel and N. Sochen, A short-time Beltrami kernel for smoothing images and manifolds, *IEEE Transactions on Image Processing*, **16** (2007), 1628–36.
- [55] Y. van Gennip, N. Guillen, B. Osting and A. L. Bertozzi, Mean curvature, threshold dynamics, and phase field theory on finite graphs, *Milan Journal of Mathematics*, **82** (2014), 3–65.
- [56] A. Wetzler and R. Kimmel, Efficient Beltrami flow in patch-space, in *Scale Space and Variational Methods in Computer Vision 2011* (eds. A. M. Bruckstein, B. M. Haar Romeny, A. M. Bronstein and M. M. Bronstein),

- vol. 6667 of Lecture Notes in Computer Science, Springer Berlin Heidelberg, Berlin, Heidelberg, 2012, 134–143.
- [57] Z. Yang, T. Hao, O. Dikmen, X. Chen and E. Oja, Clustering by nonnegative matrix factorization using graph random walk, in *Advances in Neural Information Processing Systems 25*, 2012, 1079–1087.
- [58] Z. Yang, Clustering by nonnegative matrix factorization using graph random walk (NMFR), URL <http://users.ics.aalto.fi/rozyang/nmfr/index.shtml>.
- [59] M. Zhu and T. Chan, *An efficient primal-dual hybrid gradient algorithm for total variation image restoration*, Technical report, UCLA CAM Report 08-34, 2008.
- [60] M. Zhu, S. J. Wright and T. F. Chan, Duality-based algorithms for total-variation-regularized image restoration, *Computational Optimization and Applications*, **47** (2010), 377–400.
- [61] D. Zosso, J. An, J. Stevick, N. Takaki, M. Weiss, L. S. Slaughter, H. H. Cao, P. S. Weiss and A. L. Bertozzi, Image segmentation with dynamic artifacts detection and bias correction, *submitted to: AIMS J. Inverse Problems and Imaging*.
- [62] D. Zosso and A. Bustin, A primal-dual projected gradient algorithm for efficient Beltrami regularization, *submitted to: Computer Vision and Image Understanding – Special Issue on Discrete and Continuous Geometry in Computer Vision*.

Received July 3rd, 2015; revised xxxx 20xx.

E-mail address: zosso@math.ucla.edu

E-mail address: osting@math.utah.edu