A Dirichlet Energy Criterion for Graph-Based Image Segmentation

Dominique Zosso and Stanley J. Osher
Department of Mathematics
University of California, Los Angeles
Los Angeles, CA 90095-1555, USA.
{zosso,sjo}@math.ucla.edu

Braxton Osting
Department of Mathematics
University of Utah, Salt Lake City
Salt Lake City, UT 84112, USA.
osting@math.utah.edu

***Abstract*—We consider a graph-based approach for image segmentation. We introduce several novel graph construction models which are based on graph-based segmentation criteria extending beyond—and bridging the gap between—segmentation approaches based on edges and homogeneous regions alone. The resulting graph is partitioned using a criterion based on the sum of the minimal Dirichlet energies of partition components. We propose an efficient primal-dual method for computing the Dirichlet energy ground state of partition components and a rearrangement algorithm is used to improve graph partitions. The method is applied to a number of example segmentation problems. We demonstrate the graph partitioning method on the five-moons toy problem, and illustrate the various image-based graph constructions, before successfully running a variety of region-, edge-, hybrid, and texture-based image segmentation experiments. Our method seamlessly generalizes region- and edge-based image segmentation to the multi-phase case and can intrinsically deal with image bias as well as more interesting image features such as texture descriptors.**

***Keywords*—Image segmentation, graph partitioning, active contours, Dirichlet energy, primal-dual hybrid gradients.**

I. Introduction

We are interested in the classical problem of finding the boundaries of objects in images. This problem has a huge spectrum of applications; for example, in medical imaging we would like to detect and localize a tumour in MRI slices or CT volumes, as a vital step for correct diagnosis, treatment planning, and evaluation. Other prominent examples are surveillance and remote sensing (vehicle detection, number plate recognition, satellite image analysis), computer vision, or microscopy: Briefly, whenever the partitioning is required to perform downstream analysis and processing on distinct subregions of interest of the input image.

This object delineation task is often formulated as an image segmentation problem as follows: Image segmentation is the task of partitioning the image domain Ω into homogeneous regions corresponding to individual objects, $\Omega = \sqcup_i \Omega_i$, or by duality, to find the contours Γ that define the boundaries $\partial\Omega_i$ of these objects. Classical models for image segmentation, which are briefly surveyed in Section II-A, are very successful in cases where edges are very prominent or regions neatly homogeneous; however they are prone to fail on more complicated images which are affected by intensity bias (uneven illumination). Such bias is common in imaging, for example as field inhomogeneity in MRI, or sample non-uniformity in atomic force microscopy.

To exploit the structure of images better than by just looking at global statistics or very local features, non-local and graph-based approaches have proven very successful. Indeed, images often have repetitive character, so that pixels do not only correlate very locally, but also with other pixels much farther away [1]–[3]. These non-local dependencies are naturally handled using a graph-based approach, allowing for pairwise comparisons to be made between pixels. Moreover, if pixels are considered as vertices of a graph, and the edge weights model pixel similarity between (arbitrary) pairs of pixels, then *the segmentation is equivalent to a graph partitioning problem.* The quality of the obtained segmentation, however, critically depends on the appropriate construction of this graph on the one hand, and the model employed for its partitioning on the other hand.

Therefore, in this paper, we propose: (i) a geometric graph partitioning model based on a Dirichlet energy criterion and its efficient optimization, (ii) the formulation of the image segmentation problem as such a graph partitioning problem by defining appropriate strategies for graph construction to model existing segmentation models, and (iii) a novel definition of graph-based segmentation criteria extending beyond segmentation approaches based purely on edges or homogeneous regions alone. In this work, we combine the strengths of localized region-based models, as interpolators of edge- and region-based segmentation, with the proven benefits of graph-based non-locality in a single, geometric graph partitioning approach. Our algorithm *can* easily incorporate semi-supervised information (transductive learning), but does not require it. Also, once our image-based graphs have been constructed, the underlying image information can be discarded; in particular, we do not require any region statistics to be updated during the iterations. Finally, our model intrinsically includes the multiphase case and thus seamlessly generalizes beyond the standard image bi-partition.

Outline: The remainder of this paper is structured as follows. In Section II, we discuss related work. In Sections III and IV, we introduce the graph partitioning model and an algorithm for finding local minima, respectively. In Section V we discuss graph construction. In Section VI we present the results from several numerical experiments. We conclude in Section VII with a brief discussion.

This work is supported by NSF DMS-1461138, NSF DMS-1118971, UC Lab Fees grant, ONR grant N00014-14-0444, and the W. M. Keck Foundation.

II. Related Work

A. Image segmentation and active contours

Image segmentation is most commonly formulated variationally as the problem of optimizing a parameterization of the regions or their contours towards specific segmentation criteria encoded in the objective functional [4].

Let $I\colon \Omega \subset \mathbb{R}^n \to \mathbb{R}$ denote the input image to be segmented, where Ω denotes the image domain. The prime example of edge-based image segmentation are geodesic active contours (GAC), which relies on Euclidean curve shortening [5]. Its essential ingredient is the conformal metric g (edge detector) modulating the Euclidean arc-length of the segmenting contour Γ, to attract the curve to image features of interest (*e.g.*, strong image gradients $|\nabla I|$):

$$J_{\text{GAC}}[\Gamma] = \int_\Gamma g\Big(\big|\nabla I(\Gamma)\big|\Big), \tag{1}$$

with, *e.g.*, $g(u) := (1+\lambda u^2)^{-1}$, $\lambda > 0$.

The canonical region-based image segmentation model is the Chan-Vese model (CV), [6]. The CV model is defined as the *cartoon-limit* of the Mumford-Shah functional (MS) [7], such that the recovered image is required to be piecewise constant, and regions are thus characterized by a single representative color each, μ_i. The reduction of the model to two phases and using the level-set representation [8], [9] of these two phases leads to the classical formulation [6]:

$$\begin{aligned} J_{\text{CV}}[\mu_1,\mu_2,\phi] = \lambda_1 \int_\Omega (\mu_1 - I)^2 H(\phi) \\ + \lambda_2 \int_\Omega (\mu_2 - I)^2 (1 - H(\phi)) + \beta \int_\Omega |\nabla H(\phi)|, \end{aligned} \tag{2}$$

where the levelset function $\phi\colon \Omega \to \mathbb{R}$ is positive in object regions, negative in background regions, zero on the object boundaries, H is the Heaviside function, and $\lambda_1, \lambda_2, \beta > 0$ are parameters. The last, total variation term is the co-area-formula equivalent of the boundary length.

The above models are based either on very local image gradients, or on global intensity statistics. While certainly appropriate in some cases, "natural" images are typically more complicated and exhibit structure that cannot be characterized at the local or global level, respectively. Therefore, inspired by more recent "localized" region-based models [10], [11], our work aims at a model that seamlessly interpolates the criterion between edge- and region-based segmentation.

Moreoever, a graph representation of images is more appropriate to both capture the discrete, sampled nature of image pixels, as well as the multiscale and multirange correlations observed in common images. Another great advantage of such methods is that the edge weight can be based on almost any distance (or similarity) measure between pixels, *e.g.*, as a function of texture or color. Several formulations of graph-based image segmentation have been proposed; examples in literature include ratio cuts, normalized cuts, graph cuts, spectral clustering, and random walkers [12]–[18].

More recently, the image segmentation problem, and in particular the established Chan-Vese and geodesic active contour models, have been re-considered under a pairwise, "non-local" perspective [19]. The authors of [20] extend the existing models by considering non-local interface regularization through non-local total variation and non-local region-based data-terms relating to [11]. The work of [21] can be considered an extension of [20], mainly written in the language of patch-based non-local methods [22], [23] and by including a broader variety of distance-metrics, such as Gabor features and the Wasserstein distance [24]. These three examples all work with "traditional" levelset functions or phase-field approximations, and not within a graph-partitioning setting. The model in [25] performs CV-like segmentation of data defined on graphs; it corresponds to the non-local-TV model in [20], but using MBO-like threshold dynamics for optimization. In particular, this approach still requires computing the mean statistics for each region at every iteration. Finally, Ginzburg-Landau diffuse interface models and MBO threshold dynamics [26], [27] have been used to bi-partition a graph constructed based on image features; however, these models require semi-supervised input as forcing term to avoid trivial partitions.

B. Graph partitioning

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 are that datasets are typically large-scale and high-dimensional, common formulations of the problem lead to NP-hard problems, and that the criterion of optimality is application dependent.

A common criterion is the generalized Cheeger cut (balanced graph cut). That is, one seeks a k-partition, such that

$$\min_{V=\sqcup_{\ell=1}^k V_\ell} \sum_{\ell=1}^k \frac{\text{per}(V_\ell)}{\min\left(|V_\ell|, |V_\ell^c|\right)}, \quad \text{where} \quad \text{per}(S) := \sum_{i\in S, j\in S^c} w_{ij} \tag{3}$$

is the graph perimeter of $S \subset V$ [28], [29], and S^c is the subgraph complement $S^c := V \setminus S$. This problem is NP-hard, and spectral clustering methods [30], which use a sub-level-set of the Fiedler vector, were the first attempt at relaxing it [12].

Many methods have now been developed for the efficient solution of (3). These include diffuse interface methods [26], graph Merriman-Bence-Osher (MBO) flows [31], graph curvature flow [32]–[34], and proximal splitting methods [35]–[37].

Recently, an alternative to the criterion in (3), based on the Dirichlet energies of the partition components was proposed [38]. There, a non-convex optimality criterion is given by the sum of the Dirichlet-Laplacian eigenvalues of partition components:

$$\min_{V=\sqcup_{\ell=1}^k V_\ell} \sum_{\ell=1}^k \lambda(V_\ell), \quad \text{where } \lambda(S) := \inf_{\psi} \|\nabla\psi\|_{w,E}^2 \quad \text{s.t.} \quad \|\psi\|_{V,2} = 1,\ \psi|_{S^c} = 0 \tag{4}$$

is the Dirichlet energy of a subset $S \subset V$:

$$\|\nabla\psi\|_{w,E}^2 := \sum_{(i,j)\in E} w_{ij}(\psi_i - \psi_j)^2,$$
$$\|\psi\|_{S,2}^2 := \sum_{i\in S} d_i^r \psi_i^2, \quad d_i := \sum_j w_{ij}, \quad \text{and} \quad r \in [0,1]. \quad (5)$$

The value $\lambda = \lambda(S)$ in (4) satisfies the Dirichlet eigenvalue problem, $\Delta_r\psi = \lambda\psi$, in $S \subset V$, where $\psi = \psi(S)$, is the corresponding eigenvector (or eigenfunction), satisfying the Dirichlet boundary condition $\psi = 0$ on S^c. Here, $\Delta_r := D^{1-r} - D^{-r/2}WD^{-r/2}$, where $D = \mathrm{diag}(W\mathbb{1})$ and W are the degree and weighted similarity matrix, respectively.

This eigenvalue partitioning problem (4) has an analogous continuum formulationwhich is highly interesting as well, and various results have been proven, such as existence of optimal partitions, regularity, and the asymptotic behavior of optimal partitions as the number of partitions tends to infinity, as well as computational methods for finding locally optimal partitions [39]–[46].

III. Dirichlet Energy Criterion for Graph Partitioning

A. A discrete calculus for graphs

In this section, we recall several definitions of graph-based differential operators. Let $G = (V, E, w)$ be an undirected 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 $w\colon E \to \mathbb{R}$. For notational convenience, we extend w to $V \times V$ by setting $w(i,j) = 0$ if $(i,j) \notin E$.

For functions $f, g\colon V \to \mathbb{R}$, 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, $\|f\|_2^2 := \langle f, f\rangle_{L^2(V)}$. More generally, for $f\colon V \to \mathbb{R}$, we define the partial L^p norm on a subset $S \subseteq V$, $\|f\|_{S,p}^p := \left(\sum_{i\in S} |f_i|^p\right)$.

For skew-symmetric functions $u, v\colon E \to \mathbb{R}$, 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 $\|v\|_{L^2(E)}^2 := \langle v, v\rangle_{L^2(E)}$.

For $f\colon V \to \mathbb{R}$, we define the graph gradient as edge function $\nabla_w f\colon E \to \mathbb{R}$ consisting of weighted vertex differences,

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

For skew-symmetric $v\colon E \to \mathbb{R}$, let $\mathrm{div}_w\, v\colon V \to \mathbb{R}$ denote the divergence

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

It is easy to verify that the graph gradient and divergence are adjoint using the vertex and edge L^2 inner products, *i.e.*, for $f \in L^2(V)$ and $v \in L^2(E)$,

$$\langle \nabla_w f, v\rangle_{L^2(E)} = \langle f, \mathrm{div}_w\, v\rangle_{L^2(V)}. \quad (6)$$

The graph Laplacian, $\Delta_w\colon L^2(V) \to L^2(V)$ is defined as the composition of graph divergence and graph gradient,

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

Finally, the Dirichlet energy of $f\colon V \to \mathbb{R}$ is defined as

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

B. Proposed Dirichlet energy graph partitioning criterion

Let $S \subset V$ be a vertex subset. On S, we consider the vertex function $u^\star$, which has unit $L^1(V)$ norm[1], vanishes on the complement $S^c := V\backslash S$, and has minimal Dirichlet energy, *i.e.*, satisfies

$$J^\star[S] := \min_u \; D[u]$$
$$\text{s.t.} \quad u\colon V \to [0,1], \; \|u\|_{V,1} = 1, \; \text{and } u|_{S^c} = 0. \quad (7)$$

We refer to the unique minimizer, $u^\star$, in (7) as the ground state associated with the subset S.

To a vertex partition of the graph, $V = \sqcup_{\ell=1}^k V_\ell$, we associate an energy, which is the sum of the minimal Dirichlet energies, $\sum_{\ell=1}^k J^\star[V_\ell]$. We then consider the *Dirichlet graph partitioning problem*,

$$\min_{V=\sqcup_\ell V_\ell} \sum_{\ell=1}^k J^\star[V_\ell]. \quad (8)$$

We call $V_\ell^\star$ the optimal partitions and $u_\ell^\star$ their ground states.

A semi-supervised version for transductive learning [37] or label transfer [26], [27] can be obtained simply by imposing prescribed labels for some of the graph vertices.

C. Relaxation of Dirichlet boundary conditions

An interesting route of rendering (8) accessible to efficient computational optimization has been proposed for the closely related Dirichlet and Beltrami energy graph partitioning problem in [38], [42], [47]. 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:

$$u|_{S^c} = 0 \;\Leftrightarrow\; \|u\|_{S^c,q}^q = 0. \quad (9)$$

Instead of imposing this constraint strictly, one can penalize violations by including the partial L^q norm in the optimization problem. Natural choices include $q = \{1, 2\}$.

The relaxed partitioning model then becomes:

$$\min_{V=\sqcup_\ell V_\ell} \sum_{\ell=1}^k J^{\alpha,\star}[V_\ell] \quad (10)$$

where the inner model (7) is relaxed to

$$J^{\alpha,\star}[S] := \min_u \left\{ D[u] + \frac{\alpha}{q}\|u\|_{S^c,q}^q \right\}$$
$$\text{s.t.} \quad u\colon V \to [0,1], \; \|u\|_{V,1} = 1. \quad (11)$$

A variation of this penalty term based on the relaxation studied in [40] has also been implemented as a numerical method in [48].

[1]This is a significant difference from [38], because it results in a convex set of admissible functions, enabling fast convex optimization methods.

IV. Algorithms

A. Rearrangement algorithm for problem (10)

With the relaxation of the boundary condition in place, the bi-level optimization problem can be solved by a rearrangement algorithm extended from [38].

1) Given partitions $V = \sqcup_\ell V_\ell$, compute a minimizer $u_\ell^\star$ for satisfying (11) with $S = V_\ell$ for each $\ell = 1, \ldots, k$. (This is the inner problem for (10).)
2) Given minimizers $u_\ell^\star$, reassign partitions by the "winner takes all" rule,

$$V_\ell \ni v \iff \ell = \arg\max_l \{u_l^\star(v)\}.$$

In [38], it was shown that the solution to the relaxed partitioning model (10) exists and the rearrangement algorithm converges to a local minimum in a finite number of steps.

B. Primal-dual methods for solution of the inner problem

The computation of the minimizers $u_\ell^\star$ satisfying (11) is the computational bottleneck of the proposed optimization scheme.

We note that all functionals in (11) are proper, closed and convex, and the set of admissible u, the probability simplex

$$U := \{u \mid u \colon V \to [0,1],\ \|u\|_{V,1} = 1\}, \tag{12}$$

is closed and convex, as well. Therefore, using the Legendre-Fenchel transform (convex conjugate) of the Dirichlet energy,

$$D^*[\phi] := \sup_u \{\langle \nabla_w u, \phi\rangle - D[u]\} = \frac{1}{2}\|\phi\|^2_{L^2(E)}, \tag{13}$$

the inner problem (11) is rewritten as the primal-dual problem:

$$\min_{u\in U}\max_{\phi\in L^2(E)} \left\{ \langle \nabla_w u, \phi\rangle_{L^2(E)} - \frac{1}{2}\|\phi\|^2_{L^2(E)} + \frac{\alpha}{q}\|u\|^q_{S^c,q} \right\}. \tag{14}$$

To solve this saddle point problem efficiently, we adapt the general primal-dual hybrid gradients algorithm [47], [49]–[51] to (14) as follows:

$$\phi^{n+1} = \arg\min_{\phi\in L^2(E)} \left\{ -\langle \nabla_w \bar{u}^n, \phi\rangle_{L^2(E)} + \frac{1}{2}\|\phi\|^2_{L^2(E)} + \frac{1}{2r_1}\|\phi^n - \phi\|_2^2 \right\} \tag{15a}$$

$$u^{n+1} = \arg\min_{u\in U} \left\{ \langle u, \operatorname{div}_w \phi\rangle_{L^2(V)} + \frac{\alpha}{q}\|u\|^q_{S^c,q} + \frac{1}{2r_2}\|u^n - u\|_2^2 \right\} \tag{15b}$$

$$\bar{u}^{n+1} = 2u^{n+1} - u^n \tag{15c}$$

Problems (15a) and (15b) are proximal updates, while (15c) is a simple extragradient step "predicting" the update of the primal variable u for the update of the dual variable ϕ. Note the appearance of the predicted value $\bar{u}$ instead of u in (15a).

C. Dual variable update

We first address the update of the *dual* variable ϕ. The solution to (15a) is immediately given by

$$\phi^{n+1}(i,j) = \frac{\phi^n(i,j) + r_1(\nabla_w \bar{u}^n)(i,j)}{1 + r_1}, \quad (i,j) \in E. \tag{16}$$

We discuss the elimination of the dual variable in §IV-E, below.

D. Primal variable update

We now turn our attention to the minimization problem (15b), 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 $\tilde{u}_q$, for the unconstrained problem with $u \in L^2(V)$. To find the optimal solution within U, we then project $\tilde{u}_q$ onto U,

$$u^{n+1} = P_U(\tilde{u}_q), \tag{17}$$

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 [49], [52], modulo the backward update used here instead of explicit, forward gradient descent. To this end, we first have to find the respective $\tilde{u}_q$.

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

$$\min_u \left\{ \langle u, \operatorname{div}_w \phi\rangle_{L^2(V)} + \frac{\alpha}{2}\|u\|^2_{S^c,2} + \frac{1}{2r_2}\|u^n - u\|_2^2 \right\} \tag{18}$$

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

$$(\tilde{u}_2)_i = \frac{u_i^n - r_2(\operatorname{div}_w \phi^{n+1})_i}{1 + \alpha r_2(1 - \chi_i)}, \quad \forall i \in V, \tag{19}$$

where $\chi \colon V \to \{0,1\}$ is the characteristic function of the subset $S \subset V$, the current partition, and the update can be computed by simple vertex-wise operations.

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:

$$\min_u \left\{ \langle u, \operatorname{div}_w \phi\rangle_{L^2(V)} + \alpha\|u\|_{S^c,1} + \frac{1}{2r_2}\|u^n - u\|_2^2 \right\}. \tag{20}$$

An equivalent minimization problem is given by

$$\min_u \left\{ \alpha r_2\|u\|_{S^c,1} + \frac{1}{2}\|u - (u^n - r_2 \operatorname{div}_w \phi^{n+1})\|_2^2 \right\}, \tag{21}$$

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

$$(\tilde{u}_1)_i = \operatorname{shrink}(u_i^n - r_2(\operatorname{div}_w \phi^{n+1})_i, \alpha r_2(1 - \chi_i)), \tag{22}$$

for $i \in V$, where the soft-thresholding operator is defined as

$$\operatorname{shrink}(z,\tau) := \begin{cases} z + \tau & z < -\tau, \\ 0 & |z| \le \tau, \\ z - \tau & z > \tau. \end{cases} \tag{23}$$

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

3) Probability simplex projection: Here, $P_U \colon L^2(V) \to U$ denotes projection on the probability simplex:

$$P_U(u) := \arg\min_{v\in U} \|v - u\|_2^2 \quad \text{i.e., such that} \quad \|v\|_{V,1} = 1 \ \text{ and } \ \forall i \in V \colon v_i \ge 0. \tag{24}$$

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^+ [53]. On the other hand, if $\|u^+\|_{V,1} \leq 1$, then the projection is obtained by distributing the lacking mass over all vertices, uniformly:

$$(P_U(u))_i = \begin{cases} \text{shrink}((u^+)_i, \lambda) & \text{if } \|u^+\|_{V,1} > 1, \\ (u^+)_i + \frac{1-\|u^+\|_{V,1}}{|V|} & \text{if } \|u^+\|_{V,1} \leq 1, \end{cases} \tag{25}$$

for $i \in V$, where $|V|$ denotes the number of vertices in V. An efficient strategy for finding the appropriate thresholding parameter λ is given in [53].

E. Elimination of the dual variable

In the update of the primal variable, the dual variable only appears through its divergence, $\text{div}_w\, \phi^{n+1}$. It is easy to see that

$$(\text{div}_w\, \phi^{n+1})_i = \frac{(\text{div}_w\, \phi^n)_i + r_1(\Delta_w \bar{u}^n)_i}{1 + r_1}, \quad \forall i \in V. \tag{26}$$

Assuming $\phi^0(i,j) = 0$ for all $(i,j) \in E$, we thus get

$$(\text{div}_w\, \phi^{n+1})_i = \sum_{\eta=0}^{n} \frac{r_1(\Delta_w \bar{u}^{n-\eta})_i}{(1+r_1)^{\eta+1}}, \quad \forall i \in V. \tag{27}$$

This allows rewriting the primal variable updates without the dual variable, effectively eliminating the dual variable from the scheme altogether. After elimination, the $q = 2$ primal update based on (19) becomes

$$(\tilde{u}_2)_i = \frac{u_i^n - \sum_{\eta=0}^{n} \frac{r_1 r_2 (\Delta_w \bar{u}^{n-\eta})_i}{(1+r_1)^{\eta+1}}}{1 + \alpha r_2 (1 - \chi_i)}, \tag{28}$$

and the $q = 1$ update (22) turns into

$$(\tilde{u}_1)_i = \text{shrink}\left(u_i^n - \sum_{\eta=0}^{n} \frac{r_1 r_2 (\Delta_w \bar{u}^{n-\eta})_i}{(1+r_1)^{\eta+1}}, \alpha r_2 (1 - \chi_i) \right). \tag{29}$$

F. Impact of dual variable elimination and speed up

After eliminating the dual variable, the resulting update equations (28) and (29) seem to perform an alternating sequence of explicit heat diffusion, probability-simplex projection (17), and extra gradient step (15c) (see algorithm 1). But the explicit heat diffusion is well-known to be critical regarding time-step selection and therefore dreadfully slow. *How could this be a reasonable thing to do, then*?

The answer needs closer inspection: if for a moment we forget about the summation and only look at the first term $\eta = 0$, then that is indeed an explicit heat update step, subject to the usual constraint on the time step, here $r_1 r_2 \leq \rho(\Delta_w)$. However, thanks to the sum, previous heat updates continue to be taken into account. Away from the solution, subsequent heat updates point towards the same direction and the sum lets the history of updates interfere constructively, effectively greatly increasing the time step of the explicit heat update. This is more pronounced for $r_1 \ll 1$, reducing the discounting of previous time steps (*i.e.*, increasing the memory horizon). On the other hand, close to the solution, subsequent heat steps have opposing sign; here, the summation leads to destructive interference of subsequent updates, dampens the oscillations and effectively stabilizes the explicit heat update. Larger values of r_1 reduce the memory horizon and lets the system transition from boosting, constructive interference to stabilizing, destructive interference more quickly, thereby reducing overshooting.

Algorithm 1 Dirichlet energy algorithm for graph partitioning

Input: Graph $G(V, E, w)$, initial $\{V_\ell\}_{\ell=1}^k$; k, α, q, r_1, r_2.

while $\{V_\ell\}_{\ell=1}^k$ not converged **do**
  for all $\ell = 1 \dots k$ **do**
    Initialize u_ℓ^0, $\bar{u}_\ell^0 \leftarrow u_\ell^0$, χ from V_ℓ.
    while u_ℓ^n not converged **do**
      $\tilde{u}_q$ using (28)/(29) ▷ Diffusion
      $u_\ell^{n+1} \leftarrow P_U(\tilde{u}_q)$ ▷ L^1 projection
      $\bar{u}_\ell^{n+1} \leftarrow 2u_\ell^{n+1} - u_\ell^n$ ▷ Extra-gradient
    end while
  end for
  $V_\ell \ni v \iff \ell = \arg\max_l \{u_l^\infty(v)\}$ ▷ Rearrangement
end while

Output: Partitions $\{V_\ell^\star\}_{\ell=1}^k$, ground states $\{u_\ell^\star\}_{\ell=1}^k$.

The elimination of the dual variable, besides highlighting the boosting and stabilizing properties of the scheme, also provides great computational speed up over the original scheme including both primal and dual variable updates. These primal ((19) and (22)) and dual updates (16) involve the non-local operators $\nabla_w u$ and $\text{div}_w\, \phi$. Performing their composition, Δ_w, in a single step after the elimination of the dual variable is computationally much more interesting.

G. Comparison to Merriman-Bence-Osher (MBO) dynamics

The rearrangement algorithm presented in Section IV-A for solving the bi-level optimization problem (10) has many attributes in common with the diffusion-generated Merriman-Bence-Osher (MBO) flow [27], [34], [54]–[56]. Recall that the MBO flow consists of alternatively diffusing the characteristic function associated with each partition component for a short time τ and then assigning each vertex the label of the diffused state which is maximal at that vertex.

Meanwhile, the ground state for the ℓ-th partition component satisfying (11) is proportional to the long-time limit of the solution to

$$\dot{u}_\ell = [\Delta - \alpha(1 - \chi_\ell)]\, u_\ell, \quad u(0) = u_0 \tag{30}$$

for any generic initial condition u_0. One diffusive step of the MBO method can be seen as one iteration of a split-step method for solving (30) for initial condition given by a characteristic function. The rearrangement steps for the two methods are then exactly the same.

In this light, the qualitative difference between MBO flow and the method studied here is the following. On each partition component, the MBO method diffuses the characteristic function for a short time τ—and hence only the vertices near the partition interface are significantly changed. Therefore the thresholding step only changes the partition membership of vertices near the interface. On the other hand, the ground states satisfying (11), have effectively evolved according to (30) for longer time and could (for smaller values of α) have significant mass on the complement of their respective partition components. Viewed as the ground state of a Schrödinger operator with potential localized on each component, this corresponds to the states having significantly tunneled, in the quantum sense, into their neighboring partition components. A small value of α roughly corresponds to choosing a larger time step τ in the MBO flow.

V. Graph Construction for Image Segmentation

In the following subsections, we briefly present different strategies for the construction of graphs, the partitioning of which solves the image segmentation problem. We first build graphs that mimic the well-known edge- and region-based image segmentation models, before interpolating them in a semi-local hybrid model. Lastly, we briefly describe how more complicated pixel (dis-)similarities can be used to construct these graphs, *e.g.*, for texture-based segmentation.

A. Domain partitioning

As starting point we consider a rectangular image domain with free boundary conditions to be partitioned, without considering image information at all. To this end one constructs a rectangular grid and a graph where each vertex is connected to its nearest neighbors, $N(i)$. Weights w of these edges are taken to be 1:

$$w(i,j) := \begin{cases} 1, & \text{if } j \in N(i) \\ 0, & \text{otherwise} \end{cases} \qquad \forall i,j \in V. \tag{31}$$

Partitioning this graph corresponds to a "blind" segmentation of the image domain with regular interfaces.

B. Edge-based image segmentation

Edge-based image segmentation according to (1) minimizes the interface length modulated by a metric based on image gradients. In terms of a corresponding graph construction for partitioning, the edge-based model achieves image segmentation by lessening the interface cost at locations of strong image gradients. The edge-detector can take many different forms. Here, we suggest to use the Perona-Malik edge detector [57], in accordance to the original GAC paper [5]. We use the following weight construction:

$$w(i,j) := \begin{cases} \frac{1}{\epsilon+\lambda\|I(j)-I(i)\|^2}, & \text{if } j \in N(i) \\ 0, & \text{otherwise} \end{cases} \tag{32}$$

$\forall i,j \in V$, for some $\epsilon, \lambda > 0$, where $I(i) \in \mathbb{R}^d$ denotes the feature value of pixel i, and $\|I(j)-I(i)\|^2$ is the squared L^2-distance of the pixels in feature space. The driving force for segmentation is thus obtained through a *weakening of edges between regions*. Note that because of image noise it might be useful to compute the edge-detector at a different, coarser scale by pre-filtering the image, for example with a Gaussian low-pass filter, such as in [58]. Also, the formulation is intrinsically capable of dealing with non-scalar image features I, such as color, hyperspectral, or other multi-channel features.

C. Region-based image segmentation

On the other hand, the Chan-Vese segmentation model (2) is based on a combination of regular interfaces and homogeneous regions. A simple graph partitioning model corresponding to region-based image segmentation (including the multiphase case $k > 2$) can be obtained through extension of the domain partitioning graph (31) by *strengthening the weight between spatially distant but feature-wise similar pixels*. We introduce the coefficient p related to λ/β in (2) as a balancing factor between the homogeneity and interface regularity terms:

$$w(i,j) := \begin{cases} \frac{p}{\epsilon+\lambda\|I(j)-I(i)\|^2} + (1-p), & \text{if } j \in N(i) \\ \frac{p}{\epsilon+\lambda\|I(j)-I(i)\|^2}, & \text{otherwise} \end{cases} \tag{33}$$

$\forall i,j \in V$, and $p \in (0,1)$. The optimal partition aligns the ground states u_ℓ with homogeneous regions in the image, and gradients of u_ℓ with gradients in I.

Instead of computing the pixel distance from its region's statistics, μ, as in the CV model (2), here the range-proximity of a pixel to its peers is determined stochastically through pairwise comparisons. In particular, given the typical size of an image to be segmented, a fully connected graph is obviously not practicable and we uniformly subsample the non-local connectivities at a very low rate. This is tantamount to estimating the region statistics from the statistics of some of its samples.

D. Graph construction for hybrid GAC/CV interpolation

Region-based image segmentation typically fails when images are affected by bias, while edge-based image segmentation suffers from "myopia". We thus want to interpolate between the two models through a smooth family of graph constructions. This is most easily achieved by working the region-statistics-from-sampling analogy. Indeed, think of a graph-edge sampling that is not uniform, but whose sampling radius can be tuned through a scaling parameter. A localized sampling of edges allows overcoming the myopia of the purely edge-based model, while limiting the comparison of local pixel intensity to localized region statistics. This is the graph analog of local region-statistics and region-scalable fitting models [10], [11].

In our implementation, we propose to sample the neighbors according to a 2D Gaussian distribution centered at the pixel of interest and with prescribed standard deviation. To this end, we actually perform a much simpler strategy in terms of implementation efficiency: we uniformly sample the image domain to get candidates j_k, as for the region-based construction. For each such candidate pixel j_k we also draw a sample from an r-Rayleigh distributed random variable $r_k \sim \mathcal{R}(r)$ (the Euclidean distance of a 2D sample from the mean of its Gaussian distribution is Rayleigh distributed). If the distance of the uniform sample exceeds the value of the Rayleigh sample, $\|i-j_k\| > r_k$, then the candidate pixel j_k is rejected. The distribution of the retained samples j_k will be Gaussian

with isotropic standard deviation r in each dimension; from pixel i they approximately have median distance $r\sqrt{2\ln 2}$ and mean distance $r\sqrt{\frac{\pi}{2}}$. The distribution is naturally bounded by the extent of the image domain, skewing the distribution accordingly; so these statistics are only reliable far from the boundaries and for reasonably small r. It is easy to see, that for $r \to \infty$ the model performs uniform sampling and corresponds to region-based graph construction (33); equally, for r very small and $p = 1$ the graph approaches the edge-based construction (32).

A potential alternative to non-uniform sampling for semi-local graph construction is uniform sampling localized by kernel weighting. This less stochastic approach was chosen in [20], [21], and can also be related to bilateral filtering [59].

E. Graph construction based on textural features

The world of image segmentation holds many examples where region affiliation of a pixel is not characterized simply by its intensity or intensity gradient. Instead, region homogeneity or image "edges" can equally be defined in terms of texture. Texture-region-based models for example make use of Gabor features [60]–[62], while edge-based models have been defined, *e.g.*, in terms of generalized edge-detectors [63].

More recently, the Wasserstein distance between local image histograms was used as a texture-based dissimilarity measure of pixels [21], [24]. To this end, one can estimate the local image histogram by sampling a patch (neighborhood) around the pixel of interest i, denoted as vector $\mathcal{P}_i$. For practical purposes, and considering scalar images, the L^2-Wasserstein distance (or Vasershtein metric, "earth-mover distance") between pixels i and j is then computed as the least Euclidean distance over all possible histogram permutations, P_π:

$$W(i,j) = \min_{P_\pi} \|\mathcal{P}_i - P_\pi \mathcal{P}_j\|_2. \tag{34}$$

In this case, the optimal permutation is the one that matches equally ranked samples; in practice, we therefore compute the L^2-distance between sorted image patch-vectors [21]. At this point, it is straightforward to see how this can easily be built into the proposed segmentation-through-graph-partitioning algorithm: Instead of using the pixel intensity as pixel feature $I(i)$ in the graph construction, we pre-process the image by extracting patches around each pixel as sorted feature vectors.

VI. Numerical Experiments

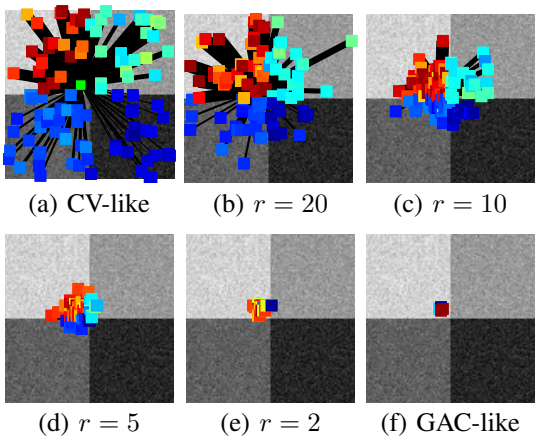
In this section, we consider several image segmentation problems. All numerical results are based on a Matlab implementation of the algorithms described in Sections III-V and run on a standard desktop computer[2].

A. Five moons toy problem

We start by illustrating the proper functioning of the proposed Dirichlet criterion for general graph partitioning. A classical example is the five-moons partitioning problem, illustrated in Figure 1. To examine the robustness of the graph partitioning model, we initialize it multiple times: Out of 100 independent, unsupervised runs with different random noise and random initialization, $41\,\%$ of results exhibited at least $99\,\%$ partition purity. The algorithm required approximately 10-15 outer iterations to converge.

[2]Code available at http://www.math.ucla.edu/~zosso/code.html.

Fig. 1. Five moons. Parameters are: $k = 5$, $\alpha = 0.2$, $q = 2$. Purity $99.88\,\%$.



(a) CV-like (b) $r = 20$ (c) $r = 10$

(d) $r = 5$ (e) $r = 2$ (f) GAC-like

Fig. 2. Image-based graph construction illustration. Edge-weight is indicated by linewidth and pixel "heat". **(a)** Region-based. **(b)–(e)** Hybrid with decreasing sampling radius r. **(f)** Edge-based.

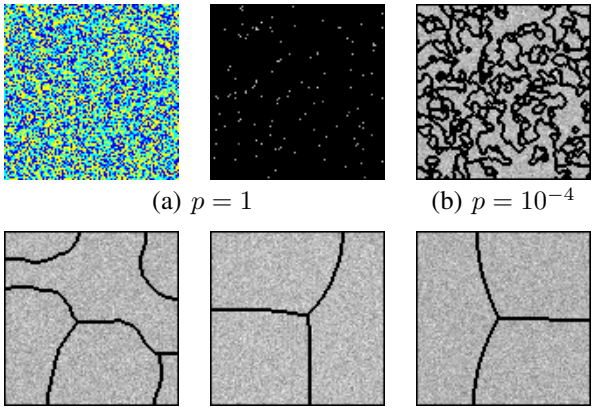
B. Image-based graph construction

We now turn to the construction of graphs based on underlying images. As a visualization, we consider a 100×100 grayscale image, with four different quadrants and affected by additive Gaussian white noise, see Figure 2. For a pixel of interest located at $(45, 45)$ in the upper left, bright quadrant, we show its edges and weight according to region, edge, and hybrid graph construction. This highlights the CV/GAC-interpolating nature of the proposed hybrid graph construction.

C. Partitioning noise

Next, we look at the segmentation of a purely white noise image, *i.e.*, we are interested in the system response in the absence of meaningful image stimulus. We partition a 100×100 Gaussian random noise image using both region and edge-based graph construction. The results are shown in Figure 3.

For the region-based, CV-like model we note the following: In the absence of regularization ($p = 1$), the pixels are clustered by intensity only, and the partition entirely lacks spatial coherence. This essentially reproduces basic intensity-based thresholding schemes and can be related to k-means clustering. For smaller values of p, thus increasing the interface regularity weight, spatial coherence increases; the resulting partitions have coarser structure. In the limit $p \to 0$, we end up with an image-independent domain partitioning problem, as given by (31).



(a) $p = 1$ (b) $p = 10^{-4}$ (c) $p = 10^{-5}$ (d) $p = 10^{-6}$ (e)

Fig. 3. Segmenting noise. **(a)–(d)** CV-like region-based graph construction for various values of p. $k = 3$, $\alpha = 10^{-4}$, 20 edges per vertex, $q = 2$, $\epsilon = 0.001$, $\lambda = 15$. For the fine-grained $p = 1$ case we show both the partition labels and the boundaries overlaid on the input image, for the other cases the boundaries are sufficiently intelligible. **(e)** Segmenting the noise with GAC-like edge-based graph construction. Same parameters, except $\alpha = 10$.

The edge-based problem does not have a regularity balancing term. Due to the i.i.d. nature of the noise, the metric is stochastically flat; as a result the interface location is not influenced by the "image" (noise), and we also effectively solve a classical domain partitioning problem.

This experiment highlights the intrinsic drive of the model to achieve about equally sized partitions in the absence of strong image cues, and in particular so in the edge-based case.

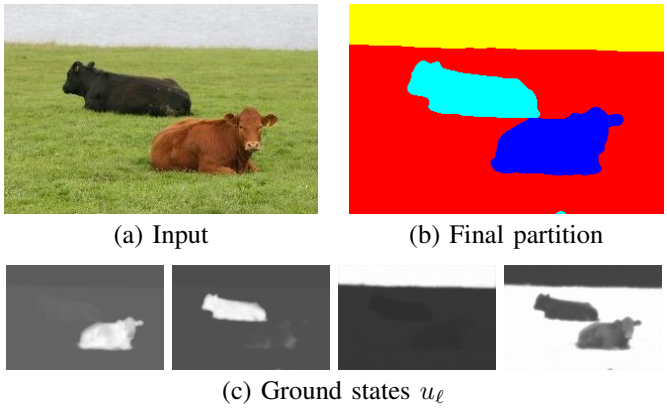
D. Region-based

As a first actual image segmentation example we consider segmenting the 318×212 RGB input image, $I \colon V \to [0,1]^3$, shown in Figure 4(a). From this image, we construct a graph using (33) with parameters $p = 7 \cdot 10^{-3}$, $\epsilon = 0.001$, and $\lambda = 15$. The long range connections are sub-sampled at about 20 edges per pixel. We apply the graph partitioning algorithm with parameters $k = 4$, $\alpha = 0.1$, and $q = 2$. The resulting segmented image is shown in Figure 4(b) and the associated ground states u_ℓ are given in Figure 4(c).

E. Edge-based and hybrid

To explore the edge-based and hybrid graph constructions, let us now consider images affected by bias. Two examples are provided in figures 5 and 6. The first example has dimensions 96×96, while the Yin-Yang image is 134×134.

As is to be expected, in both examples the region-based segmentation attempt fails due to bias (left column, respectively). In contrast, the edge-based approach successfully captures the outline of the structures, presenting strong edges despite overall bias, but fails to "see" the holes due to the model's inherent myopia. The hybrid approach, however, turns out to be an ideal compromise of both worlds: it is sufficiently local so as to not be affected by bias thanks to the limited visibility, yet the visibility is far enough and "semi-region-based" to capture the holes.



(a) Input (b) Final partition (c) Ground states u_ℓ

Fig. 4. Region-based image segmentation as a graph partitioning problem: **(a)** 318×212 RGB input image. **(b)–(c)** Image segmentation results using graph partitioning parameters $k = 4$, $\alpha = 0.1$, $q = 2$.

Fig. 5. Image segmentation affected by bias: Partition and ground states. **left**: CV-like region-based, **center**: GAC-like edge-based, and **right**: hybrid graph construction. Parameters: $k = 2$, $q = 2$, $\alpha = 2$, 60 edges per vertex, $\epsilon = 0.001$, $p = 0.1$, $\lambda = 250$; hybrid sampling radius $r = 8$.

F. Texture segmentation based on Wasserstein distance

Finally, we make use of the Wasserstein-distance region-graph construction (34) to tackle the segmentation of images based on texture. In Figure 7 we show the successful results for a cheetah (256×256) and a zebra image (254×157), and in Figure 8 we apply the model to a Brodatz texture composite.

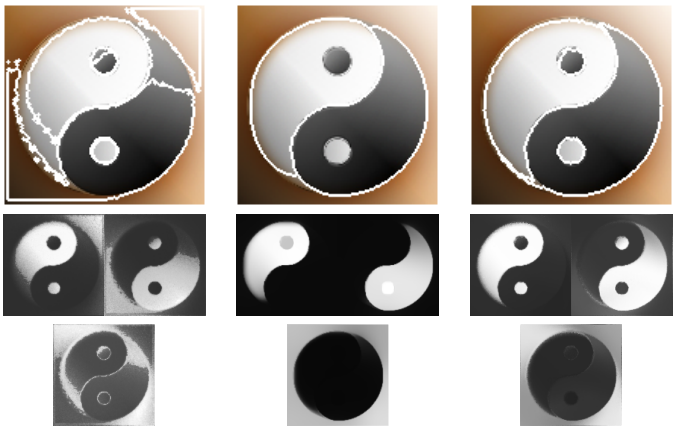


Fig. 6. Yin-Yang: Tri-partition and ground states. **left**: CV-like region-based, **center**: GAC-like edge-based, and **right**: hybrid graph construction. Parameters: $k = 3$, $q = 2$, $\alpha = 2$, 50 edges per vertex, $\epsilon = 0.001$, $p = 0.1$, $\lambda = 25$; hybrid sampling radius $r = 13$.

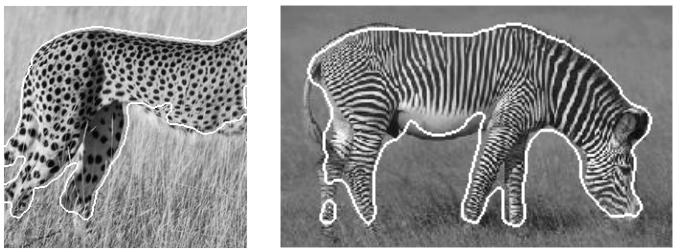


Fig. 7. Wasserstein-distance based graph construction and Dirichlet 2-partition from purely random initialization. **left:** Cheetah. $k = 2$, $q = 2$, $\alpha = 0.1$, 60 edges per vertex, $\epsilon = 0.001$, $p = 0.15$, $\lambda = 25$, Wasserstein patch size 13×13. **right:** Zebra. Same parameters, except $p = 0.3$.

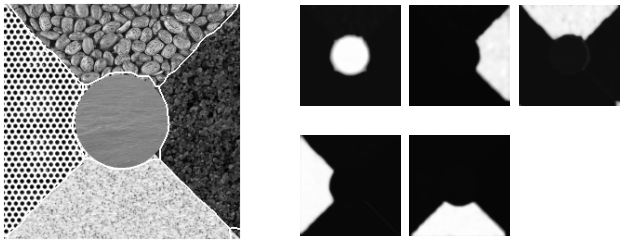


Fig. 8. Wasserstein-distance based graph construction and 5-phase segmentation from random initialization: Partition **(left)** and first five ground states **(right)**. Parameters: $k = 5$, $q = 2$, $\alpha = 0.5$, 60 edges per vertex, $\epsilon = 0.001$, $p = 0.1$, $\lambda = 25$, Wasserstein patch size 13×13.

VII. Discussion and Conclusions

In this paper, we considered a graph-based approach for image segmentation. We describe several graph construction models and the resulting graphs are partitioned using a criterion based on the sum of the minimal Dirichlet energies of partition components. We propose an efficient primal-dual method for computing the Dirichlet energy of partition components and a rearrangement algorithm is proposed to find graph partitions. The inner problem is particularly efficiently solved thanks to a possible elimination of the dual variable, resulting in a boosted explicit heat diffusion scheme. The proposed algorithm is applied to a number of example segmentation problems, both region- and edge-based, binary and multiphase. We demonstrate advantages of the hybrid semi-local graph construction in the presence of image bias, and show the seamless generalization to texture features through the use of a patch-based Wasserstein distance in the graph construction.

The *simple* region- and edge-based image segmentation problem shown here are, of course, more efficiently solved with specific state-of-the-art segmentation models. The strength of the proposed approach is its versatility and flexibility in how the graph construction can generalize to any interesting feature distance, (*e.g.*, patch-based or using other texture descriptors). 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, such as [10], [11], [64]. Finally, our model intrinsically includes multi-phase segmentation problems, and incorporates semi-supervised input in a straightforward way.

The contributions in this paper for graph construction and graph partitioning can be viewed as independent. Namely, the graph construction method described here can be used with your favorite graph partitioning method, while the graph partitioning method described here can be used for other learning tasks involving graph partitioning. Finally, the primal-dual methods proposed for the computation of the ground states for each partition component are very efficient and will be applied to a variety of problems for elliptic PDE in [65].

Acknowledgment

The authors would like to thank Jing An and Mandy (Mengqi) Xia.

References

[1] A. Buades, B. Coll, and J.-M. Morel, "A non-local algorithm for image denoising," in *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*. IEEE, 2005, pp. 60–65.

[2] ——, "Nonlocal image and movie denoising," *Int. J. Comput. Vis.*, vol. 76, pp. 123–139, 2008.

[3] D. Tschumperlé and L. Brun, "Image denoising and registration by PDE's on the space of patches," in *International Workshop on Local and Non-Local Approximation in Image Processing (LNLA'08)*, 2008.

[4] T. F. Chan, J. J. Shen, and L. A. Vese, "Variational PDE models in image processing," *Notices of the AMS*, vol. 50, no. 1, pp. 14–26, 2003.

[5] V. Caselles, R. Kimmel, and G. Sapiro, "Geodesic active contours," *Int. J. Comput. Vis.*, vol. 22, no. 1, pp. 61–79, 1997.

[6] T. F. Chan and L. A. Vese, "Active contours without edges," *IEEE Transactions on Image Processing*, vol. 10, no. 2, pp. 266–277, 2001.

[7] D. Mumford and J. Shah, "Optimal approximations by piecewise smooth functions and associated variational problems," *Communications on Pure and Applied Mathematics*, vol. 42, no. 5, pp. 577–685, Jul. 1989.

[8] S. Osher and J. A. Sethian, "Fronts propagating with curvature-dependent speed: algorithms based on Hamilton-Jacobi formulations," *Journal of Computational Physics*, vol. 79, no. 1, pp. 12–49, Nov. 1988.

[9] M. Burger and S. J. Osher, "A survey on level set methods for inverse problems and optimal design," *European journal of applied mathematics*, vol. 16, no. 02, pp. 263–301, 2005.

[10] 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*, vol. 84, no. 2, pp. 184–193, Jul. 2008.

[11] 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*, vol. 17, no. 10, pp. 1940–1949, Oct. 2008.

[12] J. Shi and J. Malik, "Normalized cuts and image segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 8, pp. 888–905, 2000.

[13] S. Wang and J. M. Siskind, "Image segmentation with ratio cut," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 25, no. 6, pp. 675–690, Jun. 2003.

[14] Y. Boykov and V. Kolmogorov, "Computing geodesics and minimal surfaces via graph cuts," in *Proceedings Ninth IEEE International Conference on Computer Vision*. IEEE, 2003, pp. 26–33 vol.1.

[15] P. F. Felzenszwalb and D. P. Huttenlocher, "Efficient graph-based image segmentation," *International Journal of Computer Vision*, vol. 59, no. 2, pp. 167–181, Sep. 2004.

[16] C. Fowlkes, S. Belongie, F. Chung, and J. Malik, "Spectral grouping using the Nyström method." *IEEE transactions on pattern analysis and machine intelligence*, vol. 26, no. 2, pp. 214–25, Feb. 2004.

[17] Y. Boykov and G. Funka-Lea, "Graph cuts and efficient N-d image segmentation," *International Journal of Computer Vision*, vol. 70, no. 2, pp. 109–131, Nov. 2006.

[18] L. Grady, "Random walks for image segmentation." *IEEE transactions on pattern analysis and machine intelligence*, vol. 28, no. 11, pp. 1768–83, Nov. 2006.

[19] L. Bertelli, B. Sumengen, B. S. Manjunath, and F. Gibou, "A variational framework for multiregion pairwise-similarity-based image segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 30, no. 8, pp. 1400–1414, 2008.

[20] X. Bresson and T. Chan, "Fast dual minimization of the vectorial total variation norm and applications to color image processing," *Inverse Problems and Imaging*, vol. 2, no. 4, pp. 455–484, Nov. 2008.

[21] M. Jung, G. Peyré, and L. D. Cohen, "Nonlocal active contours," *SIAM Journal on Imaging Sciences*, vol. 5, no. 3, pp. 1022–1054, 2012.

[22] G. Peyré, S. Bougleux, and L. D. Cohen, "Non-local regularization of inverse problems," in *ECCV 2008*, 2008, pp. 57–68.

[23] S. Bougleux, G. Peyré, and L. Cohen, "Non-local regularization of inverse problems," *Inverse Problems and Imaging*, vol. 5, no. 2, pp. 511–530, May 2011.

[24] K. Ni, X. Bresson, T. Chan, and S. Esedoglu, "Local histogram based segmentation using the Wasserstein distance," *International Journal of Computer Vision*, vol. 84, no. 1, pp. 97–111, 2009.

[25] H. Hu, J. Sunu, and A. L. Bertozzi, "Multi-class graph Mumford-Shah model for plume detection using the MBO scheme," in *Energy Minimization Methods in Computer Vision and Pattern Recognition*, ser. Lecture Notes in Computer Science, X.-C. Tai, E. Bae, T. F. Chan, and M. Lysaker, Eds. Springer International Publishing, 2015, vol. 8932, pp. 209–222.

[26] A. L. Bertozzi and A. Flenner, "Diffuse interface models on graphs for classification of high dimensional data," *Multiscale Modeling & Simulation*, vol. 10, no. 3, pp. 1090–1118, 2012.

[27] E. Merkurjev, T. Kostic, and A. L. Bertozzi, "An MBO scheme on graphs for segmentation and image processing," *SIAM J. Imaging Sciences*, vol. 6, no. 4, pp. 1903–1930, 2013.

[28] F. R. K. Chung, *Spectral Graph Theory*. AMS, 1997.

[29] U. von Luxburg, "A tutorial on spectral clustering," *Statistics and Computing*, vol. 17, no. 4, pp. 395–416, 2007.

[30] D. A. Spielman and S.-H. Teng, "Spectral partitioning works: planar graphs and finite element meshes," in *37th Annual Symposium on Foundations of Computer Science (Burlington, VT, 1996)*. IEEE Comput. Soc. Press, 1996, pp. 96–105.

[31] 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*, vol. 36, no. 8, pp. 1600–1613, 2014.

[32] A. Chambolle, "An algorithm for mean curvature motion," *Interfaces and Free Boundaries*, vol. 6, no. 2, pp. 195–218, 2004.

[33] A. Chambolle and J. Darbon, "On total variation minimization and surface evolution using parametric maximum flows," *International Journal of Computer Vision*, vol. 84, no. 3, pp. 288–307, 2009.

[34] 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*, vol. 82, no. 1, pp. 3–65, Apr. 2014.

[35] A. Szlam and X. Bresson, "Total variation and Cheeger cuts," in *27th International Conference on Machine Learning*, vol. 1, no. 4, 2010, pp. 1039–1046.

[36] X. Bresson, T. Laurent, D. Uminsky, and J. von Brecht, "Multiclass total variation clustering," in *Advances in Neural Information Processing Systems*, 2013, pp. 1421–1429.

[37] X. Bresson, X.-C. Tai, T. F. Chan, and A. Szlam, "Multi-class transductive learning based on $\ell 1$ relaxations of Cheeger cut and Mumford-Shah-Potts model," *Journal of Mathematical Imaging and Vision*, vol. 49, no. 1, pp. 191–201, May 2014.

[38] B. Osting, C. D. White, and E. Oudet, "Minimal Dirichlet energy partitions for graphs," *SIAM Journal on Scientific Computing*, vol. 36, no. 4, pp. A1635–A1651, Aug. 2014.

[39] D. Bucur, G. Butazzo, and A. Henrot, "Existence results for some optimal partition problems," *Adv. Math. Sci. Appl.*, vol. 8, pp. 571–579, 1998.

[40] L. A. Cafferelli and F. H. Lin, "An optimal partition problem for eigenvalues," *J. Sci. Comp.*, vol. 31, no. 1-2, pp. 5–18, 2007.

[41] V. Bonnaillie-Noël, B. Helffer, and G. Vial, "Numerical simulations for nodal domains and spectral minimal partitions," *Preprint Hal 00150455*, 2007.

[42] B. Bourdin, D. Bucur, and E. Oudet, "Optimal partitions for eigenvalues," *SIAM Journal on Scientific Computing*, vol. 31, no. 6, pp. 4100–4114, 2010.

[43] B. Helffer and T. Hoffmann-Ostenhof, "Remarks on two notions of spectral minimal partitions," *preprint*, 2009.

[44] B. Helffer, "On spectral minimal partitions: a survey," *Milan J. Math.*, vol. 78, pp. 575–590, 2010.

[45] B. Helffer, T. Hoffmann-Ostenhof, and S. Terracini, "On spectral minimal partitions: the case of the sphere," *Around the Research of Vladimir Maz'ya III*, pp. 153–178, 2010.

[46] D. Bucur and B. Velichkov, "Multiphase shape optimization problems," *submitted*, 2013.

[47] D. Zosso and B. Osting, "A minimal surface criterion for graph partitioning," *submitted to: AIMS J. Inverse Problems and Imaging*, p. 30, 2015.

[48] C. M. Elliott and T. Ranner, "A computational approach to an optimal partition problem on surfaces," Tech. Rep., 2014.

[49] M. Zhu and T. Chan, "An efficient primal-dual hybrid gradient algorithm for total variation image restoration," UCLA CAM Report 08-34, Tech. Rep., 2008.

[50] A. Chambolle and T. Pock, "A first-order primal-dual algorithm for convex problems with applications to imaging," *Journal of Mathematical Imaging and Vision*, vol. 40, no. 1, pp. 120–145, Dec. 2011.

[51] D. Zosso and A. Bustin, "A primal-dual projected gradient algorithm for efficient Beltrami regularization," UCLA CAM Report 14-52, Tech. Rep., 2014.

[52] K. J. Arrow, L. Hurwicz, and H. Uzawa, *Studies in linear and non-linear programming*. Cambridge Univ. Press, 1958.

[53] J. Duchi, S. Shalev-Shwartz, Y. Singer, and T. Chandra, "Efficient projections onto the l1-ball for learning in high dimensions," in *Proceedings of the 25th international conference on Machine learning - ICML '08*. New York, New York, USA: ACM Press, Jul. 2008, pp. 272–279.

[54] B. Merriman, J. K. Bence, and S. Osher, "Diffusion generated motion by mean curvature," UCLA CAM Report 92-18, Tech. Rep., 1992.

[55] ——, "Diffusion generated motion by mean curvature," *AMS Selected Letters, Crystal Grower's Workshop*, pp. 73–83, 1993.

[56] S. Esedoglu and F. Otto, "Threshold dynamics for networks with arbitrary surface tensions," *Communications on Pure and Applied Mathematics*, vol. 68, no. 5, pp. 808–864, 2015.

[57] P. Perona and J. Malik, "Scale-space and edge detection using anisotropic diffusion," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 12, no. 7, pp. 629–639, 1990.

[58] F. Catte, P.-L. Lions, J.-M. Morel, and T. Coll, "Image selective smoothing and edge detection by nonlinear diffusion," *SIAM Journal on Numerical Analysis*, vol. 29, no. 1, p. 182, Jul. 1992.

[59] C. Tomasi and R. Manduchi, "Bilateral filtering for gray and color images," in *ICCV 1998*. Narosa Publishing House, 1998, pp. 839–846.

[60] A. Jain and F. Farrokhnia, "Unsupervised texture segmentation using Gabor filters," in *IEEE International Conference on Systems, Man, and Cybernetics Conference Proceedings*. IEEE, 1990, pp. 14–19.

[61] N. Paragios and R. Deriche, "Geodesic active regions and level set methods for supervised texture segmentation," *International Journal of Computer Vision*, vol. 46, no. 3, pp. 223–247, 2002.

[62] B. Sandberg, T. F. Chan, and L. A. Vese, "A level-set and Gabor-based active contour algorithm for segmenting textured images," UCLA CAM Report 02-39, Tech. Rep., 2002.

[63] C. Sagiv, N. A. Sochen, and Y. Y. Zeevi, "Integrated active contours for texture segmentation," *IEEE Transactions on Image Processing*, vol. 15, no. 6, pp. 1633–1646, 2006.

[64] 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*, p. 24, 2015.

[65] D. Zosso, B. Osting, and S. J. Osher, "Elliptic PDE and primal-dual optimization methods," *in preparation*, 2015.