Active Mean Fields for Probabilistic Image Segmentation: Connections with Chan-Vese and Rudin-Osher-Fatemi Models

Marc Niethammer¹, Kilian M. Pohl², Firdaus Janoos³, and William M. Wells III⁴

¹University of North Carolina at Chapel Hill, Department of Computer Science, Biomedical Research Imaging Center (BRIC), mn@cs.unc.edu
²Center for Health Sciences, SRI International & Department of Psychiatry and Behavioral Sciences, Stanford University, kilian.pohl@stanford.edu
³Two Sigma Investments, firdaus.janoos@gmail.com
⁴Harvard Medical School and Brigham and Women’s Hospital, sw@bwh.harvard.edu

Draft from: January 26, 2015

Abstract

Image segmentation is a fundamental task for extracting semantically meaningful regions from an image. The goal is to assign object labels to each image location. Due to image-noise, shortcomings of algorithms and other ambiguities in the images, there is uncertainty in the assigned labels. In multiple application domains, estimates of this uncertainty are important. For example, object segmentation and uncertainty quantification is essential for many medical application, including tumor segmentation for radiation treatment planning. While a Bayesian characterization of the label posterior provides estimates of segmentation uncertainty, Bayesian approaches can be computationally prohibitive for practical applications. On the other hand, typical optimization based algorithms are computationally very efficient, but only provide maximum a-posteriori solutions and hence no estimates of label uncertainty. In this paper, we propose Active Mean Fields (AMF), a Bayesian technique that uses a mean-field approximation to derive an efficient segmentation and uncertainty quantification algorithm. This model, which allows combining any label-likelihood measure with a boundary length prior, yields a variational formulation that is convex. A specific implementation of that model is the Chan–Vese segmentation model (CV), which formulates the binary segmentation problem through Gaussian likelihoods combined with a boundary-length regularizer. Furthermore, the Euler–Lagrange equations derived from the AMF model are equivalent to those of the popular Rudin-Osher-Fatemi (ROF) model for image de-noising. Solutions to the AMF model can thus be implemented by directly utilizing highly-efficient ROF solvers on log-likelihood ratio fields. We demonstrate the approach using synthetic data, as well as real medical images (for heart and prostate segmentations), and on standard computer vision test images.

Contents

1 Introduction 2
   1.1 Motivations 2
   1.2 Contributions 3
   1.3 Background 3
   1.4 Structure of the Article 5
2 Active Mean Fields (AMF) 5
Segmentation approaches rarely provide measures of segmentation uncertainties. In fact, most existing and probabilistically-motivated segmentation approaches only compute maximum a posteriori (i.e., MAP) labels of images [Kolmogorov & Zabih, 2004; Li, 2009; Bresson et al., 2007; Cremers et al., 2007; Paragios & Deriche, 2002]. Thus, using these models to segment ambiguous boundaries is problematic especially for applications where confidence in object boundaries impacts analysis. For example, many radiation treatment plans base dose dispensation on the boundaries of tumors segmented from medical images with low signal to noise ratio. The precision of the planning could especially be improved around highly-confident tumor boundaries thereby reducing the risk of damaging healthy tissue in those areas. As significant information about label uncertainty is contained in the posterior distribution it is natural to go beyond computing a MAP solution and instead compute the posterior distribution itself or a computationally efficient approximation.

This paper develops a method for such an efficient approximation of the posterior distribution on labels. Furthermore, it connects this method to the Rudin-Osher-Fatemi (ROF) model for image denoising [Rudin et al., 1992, Vogel & Oman, 1996; Beck & Teboulle, 2009] and previously existing level-set segmentation approaches [Osher & Fedkiw, 2003], in particular the Chan-Vese segmentation model [Chan et al., 2000]. Due to these connections we can (i) make use of the efficient solvers for the ROF model to approximate the posterior distribution on labels. Furthermore, we can (ii) compute the solution to the Chan-Vese model through the MAP realization of our approximation to the posterior distribution. I.e., our model is more general and subsumes the Chan-Vese model. In contrast to the implicit style of active-contour methods that represent labels by way of zero level-sets, we use a dense logit (“log odds”), representation of label probabilities. This is akin to the convex approaches for active contours [Appleton & Talbot, 2006], but in a probabilistic formulation.
provides such posterior distributions. In addition, we previously found the logit transformation of label probabilities (log odds) to be a useful representation for segmentation: it represents label probabilities in a linear vector space, it facilitates PCA on probabilistic segmentations, and has, empirically, attractive interpolation behavior (Pohl et al., 2007b).

In previous work, we described an Active Mean Fields (AMF) approach to image segmentation that used a variational mean field method (VMF) approach along with a logit representation to construct a segmentation system similar to the one described here (Pohl et al., 2007a). This method empirically showed good segmentation and convergence properties.

In the present work, we revisit the AMF formalism. Here we use a more principled approximation that leads to a different Euler-Lagrange (EL) equation. With the new formulation, we make strong connections to the Chan-Vese (Chan et al., 2000) and Rudin-Osher-Fatemi (Rudin et al., 1992) models of segmentation and de-noising, and we show good approximation properties in comparison to the “exact” distribution.

1.2 Contributions

The main contributions of this article are:

- It derives a variational mean-field approximation, active mean fields (AMF), for binary image segmentation under a boundary length prior, that allows computationally efficient estimation of an approximate posterior distribution on the label-map.
- It establishes strong connections between the proposed AMF model, active-contour models and total-variation (TV) denoising. In particular, the model retains the global optimality of convex active-contours while estimating a level-set function that has a direct interpretation as an approximate posterior on the segmentation. This is in contrast to level-set techniques which use the zero level-set only as a means for representing the object boundary with no (probabilistic) interpretation of the non-zero level-sets.
- It demonstrates how the Rudin-Osher-Fatemi (ROF) TV denoising model can be used to efficiently compute solutions of the AMF model. Hence, given the widespread availability of high-performance ROF-solvers, the AMF model is quite simple to implement and will be immediately usable by the community with little effort.

1.3 Background

The earliest and simplest probabilistic image segmentation approaches frequently used pixel-wise independent Maximum Likelihood (ML) or MAP classifiers (Vannier et al., 1985), that could be as simple as image thresholding. Better performance, in the face of noise, motivated the use of regularization, or prior probability models on the label fields that discouraged fragmentation (Besag, 1986), leading to the wide-spread application of Markov Random Field (MRF) models (Held et al., 1997; Zhang et al., 2001). Image segmentation with MRF models was initially thought to be computationally complex, which motivated approximations, including the mean field approach from statistical physics (Kapur et al., 1998; Chandler, 1987). Moreover recently, fast solvers using graph cuts, belief propagation and linear programming techniques that can find globally optimal solutions for certain classes of problems have been developed (Szeliski et al., 2008).

Typically, the segmentation problem is posed as minimization of energy or negative log-probability which incorporates an image likelihood or misfit term and a regularization term or prior on the boundaries of segmented objects. This regularization may be specified either: (1) directly on the boundary (explicitly as a parametric curve or surface, or implicitly through the use of level-set functions); or (2) by representing objects via indicator functions, where discontinuities in those functions identify boundaries.

The direct boundary representation is attractive because it reduces complexity as only objects of codimension one need to be dealt with (i.e., a curve in 2D, a surface in 3D, etc.). The price for this reduction in complexity is that, usually, minimization becomes non-convex, hence can get trapped in poor local minima in the absence of good initializations. In the
snakes approach (Kass et al., 1988), a popular example of explicit boundary representation, the boundary curve represented by knot-points is evolved such that it captures the object of interest (for example, by getting attracted to edges) while assuring regularity of the boundary by penalizing rapid boundary changes through elasticity and rigidity terms. Although computationally efficient, explicit parametric representations cannot easily deal with topological changes and have numerical issues due to their fixed object parametrization (e.g. when an initial curve grows or shrinks drastically). Furthermore, though not an intrinsic problem of explicit parameterizations, such methods are typically not geometric, making evolutions dependent on curve parameterizations.

In contrast, level-set representations (Osher & Fedkiw, 2003; Malladi et al., 1995) and active-contour methods (Caselles et al., 1997; Kichenas-samy et al., 1995) do not suffer from these topological and parametrization issues. These methods use implicit representations of the label-field, where an object’s boundary is, for example represented through the zero level-set of a function. This level-set function is then evolved instead of evolving a parametric boundary representation directly. As the level-set function is by construction either strictly positive or negative (depending on convention used) inside the object and strictly negative or positive on the outside of the object, a labeling can be obtained by simple thresholding. Level-set approaches for image segmentation make use of advanced numerical methods to solve the associated partial differential equations (Osher & Fedkiw, 2003; Sapiro, 2001). To assure boundary regularization segmentation energies typically penalize boundary length/surface area.

While initial curve and surface evolution methods focused on energy minimization based on boundary regularity and boundary misfit, later approaches such as the Chan-Vese model (Chan et al., 2000), added terms that encoded statistics about the regions inside and outside the segmentation boundary. Such models can be as simple as isotropic Gaussian likelihoods with a specified (but distinct) mean, as in the Chan-Vese case, but can be as complex as trying to maximally separate intensity or feature distributions inside and outside an object (Georgiou et al., 2007). While region-based models are less sensitive to initialization, they are still non-convex when combined with weighted curve-length terms for regularization. Hence, a global optimum cannot be guaranteed by numerical optimization for such formulations.

The dependency on curve/surface initializations popularized the formulation of energy minimization methods through graphs. In the context of image segmentation, the idea is to create a graph with added source and sink nodes in such a way that a minimum cut of the graph implies a variable configuration which minimizes the original image segmentation energy (Boykov et al., 2001). For a large class of binary segmentation problems such a graph-cut approach allows for the efficient computation of globally optimal solutions through max-flow algorithms (Kolmogorov & Zabih, 2004). In particular, discrete versions of the active-contour and Chan-Vese models (with fixed means) can be formulated. To avoid computing trivial solutions for the boundary-only active contours, graph-cut formulations add seed-points, specifying fixed background and foreground pixels/voxels. While conceptually attractive, graph-cut approaches suffer from the need to build the graphs and the necessity to specify discrete neighborhood structures which may negatively affect the regularity of the obtained solution by creating so-called metrication artifacts.

Recently the focus has shifted away from level-set and graph representations to formulations of active contours and related models by means of indicator functions (Appleton & Talbot, 2006; Bresson et al., 2007) defined in the continuum and allowing for convex formulations. These methods are closely related to segmentation via graph-cuts, but avoid the construction of graphs and can alleviate metrication artifacts. A key insight here is that the boundary-length / area term can be achieved through the total-variation (TV) of the indicator function, which becomes convex when followed by relaxation of the indicator functions to the [0, 1] interval. Hence these approaches strike an attractive balance between Partial Differential Equation (PDE)-based level-set formulations and the global properties of graph-cut methods. Furthermore as they are specified via PDEs, high-accuracy, fast, parallel and GPU implementa-
tions for their solvers are available (Pock et al., 2008). As these convex formulations make use of total-variation terms, they are conceptually related to total-variation (TV) image-denoising. The use of TV regularization for denoising was pioneered by Rudin, Osher and Fatemi (the ROF model (Rudin et al., 1992)). The ROF model uses quadratic (i.e. \( \ell_2 \)) coupling to the image intensities and total-variation (TV) for edge-preserving noise-removal (Burger & Osher, 2013). Approaches with \( \ell_1 \) coupling yielding a form of geometric scale-space have also been proposed (Chan & Esedoglu, 2005).

Segmentation approaches based on energy-optimization as discussed above typically either have a probabilistic interpretation (as negative log-likelihoods) or have been explicitly derived from probabilistic considerations. The reader is referred to Cremers et al. (Cremers et al., 2007) for a review of recent developments in probabilistic level-set segmentation. All these techniques, while probabilistic in nature, seek optimal labels and do not directly provide information about the posterior distribution or uncertainty in these solutions.

In contrast, our goal for AMF segmentation is to take the Bayesian approach of characterizing posterior distributions on label-maps by combining explicit representations of label likelihoods along with a boundary length prior. As we show, our approach makes strong connections to ROF-denoising, and convex active-contour and probabilistic active-contour formulations. In particular, we show that using ROF-denoising on the logit field of label probabilities results in a “denoised” logit transform from which a label-map probability function can easily be obtained through a sigmoid transformation. More importantly, this probability function is a good approximation of the posterior of the segmentation under a curve-length prior and can be very efficiently computed using standard ROF-solver technology.

1.4 Structure of the Article

In Section 2 we specify a discrete-space probabilistic formulation of segmentation with the goal of finding the posterior distribution of labels, given an input image. We use use the VMF approach, along with a linearization approximation that serves to simplify the problem. This results in an optimization problem for determining the parameters of an approximation to the posterior distribution on pixel labels. In the style of Chan and Vese (Chan et al., 2000) and many others, we shift from discrete to continuous space facilitating use of the calculus of variations for the optimization problem, yielding the Euler-Lagrange equations for the AMF model.

Section 3 describes some properties of AMF. We show that for a one-parameter family of realizations, the approximated and exact posteriors agree ratio-metrically, and we explore their agreement for more general realizations. In addition, we show that the AMF problem is convex, and is unbiased in a particular sense.

In Section 4 we show that the AMF Euler-Lagrange equations for the zero level-set correspond to those of a special case of the Chan-Vese model (Chan et al., 2000), and that the AMF “approximate posterior” has the same mode, or MAP realization, as the exact posterior distribution. Subsequently we show that the AMF Euler-Lagrange equations have the same form as those of the ROF model of image de-noising, and we discuss methods that may be used for solving AMF.

Section 5 shows experimental results for examples that include intensity ambiguities, demonstrating the quality of the AMF in approximating the true posterior via Gibbs sampling, and illustrating the approach on real ultrasound images of the heart and the prostate as well as two standard computer vision test images.

Finally, Section 6 concludes with a summary and an outlook on future work. Detailed derivations of the approximation properties can be found in the Appendix.

2 Active Mean Fields (AMF)

This section introduces the basic discrete-space probabilistic model (Section 2.1), that includes a conventional conditionally independent likelihood term and a prior that penalizes the boundary length of the labeling. The variational mean field (VMF) approach
is used (Section 2.2), along with a Plefka approximation (Section 2.3), to construct a factorized distribution that, given image data, approximates the posterior distribution on labelings. The resulting optimization problem for determining the parameters of the variational distribution has a KL-divergence data attachment term and a total-variation regularizer. The objective function is converted to continuous space (Section 2.4), yielding the Euler-Lagrange optimizer. The objective function is converted to continuous space (Section 2.4), yielding the Euler-Lagrange equations of the AMF model (Section 2.5), that involve logit label probabilities and likelihoods along with a curvature term.

In the following sections, we use upper-case $P$ and $Q$ to represent probability mass functions and lower-case $p$ and $q$ to represent probability density functions.

2.1 Original Probability Model

Define the space of images as a compact domain $\mathcal{X} \subset \mathbb{R}^2$ indexed by $x \in \mathbb{R}^2$ and let $i : x_i \in \mathcal{X}$ denote the indices of the lattice of image pixels, and let $Z$ denote a binary random field defined on the pixel lattice whose realizations $z$ are the binary labelings of a real-valued image $y$ on the pixel lattice.

In the following sections, we use upper-case $P$ and $Q$ to represent probability mass functions and lower-case $p$ and $q$ to represent probability density functions.

2.1 Original Probability Model

Define the space of images as a compact domain $\mathcal{X} \subset \mathbb{R}^2$ indexed by $x \in \mathbb{R}^2$ and let $i : x_i \in \mathcal{X}$ denote the indices of the lattice of image pixels, and let $Z$ denote a binary random field defined on the pixel lattice whose realizations $z$ are the binary labelings of a real-valued image $y$ on the pixel lattice.

We use the usual assumption that the likelihood term, the probability density of observing intensities conditioned on labels, is conditionally iid, id;

$$p(y|z) = \prod_{i \in \mathcal{X}} p(y_i | z_i)$$

where $\psi$ is the logit transform of the image likelihood

$$\psi_i = \ln \frac{p(y_i|1)}{p(y_i|0)},$$

$y$ denotes the image and $z$ the label field; $y_i$ and $z_i$ denote the image and the label at pixel $i$. The definition of $p(y_i|z_i)$ is problem specific and is assumed to be given (for example, specified parametrically or obtained through kernel density estimation on a given feature space; we will not address this issue here).

Next, we apply a prior that penalizes the length $L(z)$ of the boundaries of the label-map,

$$P(z) \propto \exp(-\lambda L(z)).$$

Here, $\lambda \in \mathbb{R}^+$ is a constant. The larger $\lambda$ the more irregular segmentation boundaries are penalized and therefore discouraged. We defer discussion of the length functional $L(\cdot)$ to Section 2.4.

By Bayes' rule the posterior probability of the label-map given the image is

$$P(z|y) \propto p(y|z)P(z).$$

2.2 Mean Field Approximation

The mean-field method approximates the joint distribution of a countable family of random-variables as a product of univariate distributions and the variational mean-field (VMF) approximation, which is widely used in machine learning and other fields (Wainwright & Jordan, 2008), uses variational techniques to estimate this approximation.

For the binary segmentation problem, we define the mean-field approximation $Q(z; \theta)$ of the posterior distribution $P(z|y)$ as a field of independent Bernoulli random variables $z_i$ defined on the lattice with probability $\theta_i$:

$$Q(z; \theta) \triangleq \prod_{i \in \mathcal{X}} \theta_i^{z_i} (1 - \theta_i)^{1 - z_i}$$

$$= \exp \left\{ \sum_{i \in \mathcal{X}} [z_i \phi_i + \ln (1 - \theta_i)] \right\}$$

where $\phi_i \triangleq \ln \frac{\theta_i}{1 - \theta_i}$. The parameter field $\hat{\theta}$ of the Bernoulli random-field is selected to minimize the KL-divergence between the original probability model and the mean-field approximation:

$$\hat{\theta} = \arg \min_{\theta} \text{KL}[Q(z; \theta) || P(z|y)]$$

$$= \arg \min_{\theta} \mathbb{E}_Q [\ln Q(z; \theta) - \ln P(z|y)]$$

$$= \arg \min_{\theta} \mathbb{E}_Q [\ln Q(z; \theta) - \ln P(y|z) + \lambda L(z)].$$

With minor abuse of KL-divergence notation:

$$\hat{\theta} = \arg \min_{\theta} \text{KL}[Q(z; \theta) || P(y|z)] + \mathbb{E}_Q [\lambda L(z)].$$
In other words, the VMF approximation selects the parameters of a variational distribution \( Q(Z; \theta) \) that captures information contained in the image likelihood function which is local to each pixel along with the expected value of the segmentation boundary length, which is a global property that regularizes the solution.

2.3 Plefka’s Approximation

Although minimizing the KL-divergence term in eqn. (9) with respect to \( \theta \) is relatively straightforward, minimizing \( E_Q[L(z)] \) is not, as it entails a sum over all configurations of \( z \). In the mean-field literature, difficult expectations of functions of random-fields have been approximated using Plefka’s method \cite{Plefka1982}.

Plefka’s approximation involves a first (or higher) order Taylor expansion of the function about the most probable field. Specifically, using the first order expansion of \( L(z) \approx L(z^*) + (z - z^*) \cdot \nabla L(z^*) \), then, 
\[
E_Q[L(z)] \approx L(z^*) + (E_Q[z] - z^*) \nabla L(z^*) \approx L(E_Q[z])
\]
so that an approximation of eqn. (9) is
\[
\hat{\theta} = \arg \min_{\theta} KL[Q(z; \theta)||P(y|z)] + \lambda L(\theta). \tag{10}
\]

Moreover, \( E_q[L(Z)] \geq L(\theta) \) which follows from Jensen’s inequality \( E_q[TV[Z]] > TV[E_q[Z]] = TV[\theta] \) and because \( E_q[z] = \theta \). Therefore, the Plefka approximation of eqn. (10) is a lower bound to the original objective function of eqn. (9). While this is not directly useful for our purposes, there has been some work on “converse Jensen inequalities” \cite{Dragomir2004} that may provide useful bound relationships. In the end, approximations are justified by the quality of their results; in Section 3 we describe some good approximation properties of the AMF formulation.

2.4 Continuous Variant of of Variational Problem

In the previous section, we showed how the problem computing the posterior distribution of a label-field under an (unspecified) boundary-length prior results in solving an optimization problem of eqn. (10).

To solve this problem using computationally efficient PDE optimization techniques, we first replace the random-field defined on a discrete lattice by one defined on continuous space.

Expanding Equation 10 using Equations 1 and 5 we get:
\[
\hat{\theta} = \arg \min_{\theta} E_Q \left[ \sum_{i \in I_N} z_i \phi_i + \ln (1 - \theta_i) - z_i \psi_i - \ln p(y_i|0) \right] + \lambda L(\theta)
\]
\[
= \arg \min_{\theta} \sum_{i \in I_N} \left[ \theta_i \phi_i + \ln (1 - \theta_i) - \theta_i \psi_i - \ln p(y_i|0) \right] + \lambda L(\theta). \tag{11}
\]

Extending \( \theta \) to the continuum, dropping the constant in \( \theta \) term, and re-arranging, we replace summation over the lattice with integration,
\[
\hat{\theta} = \arg \min_{\theta} \int_X \theta(x)(\phi(x) - \psi(x)) + \ln (1 - \theta(x)) dx + \lambda L(\theta)
\]
where \( v \) is the area/volume of a lattice element and
\[
L(\theta) = \psi(x) \triangleq \ln \frac{p(y(x)|z(x) = 1)}{p(y(x)|z(x) = 0)} \quad \text{and} \quad \phi(x) = \ln \frac{\theta(x)}{1 - \theta(x)}
\]
are the logit transforms of the likelihood and variational probability functions, respectively. The corresponding probability is obtained as \( \theta = \sigma(\phi) \) where \( \sigma(x) \triangleq (1 + \exp(-x))^{-1} \) is the sigmoid or logistic transform.

By the co-area formula \cite{BethuelGhidaglia1994}, the length of the boundaries of a binary image defined on the continuum is equal to its total-variation:
\[
L(z) = TV[z(x)] = \int_X ||\nabla z(x)||_2 dx \tag{13}
\]
where \( \nabla z \) is the (weak) gradient of \( z \).

Therefore putting it all together, the continuous variant of the variational problem is:
\[
\hat{\theta} = \arg \min_{\theta} \int_X \theta(x)(\phi(x) - \psi(x)) + \ln (1 - \theta(x)) + \lambda ||\nabla \theta(x)||_2 dx, \tag{14}
\]
which we call the Active Mean Field approximation.
2.5 Euler-Lagrange (EL) Equations

Defining the curvature operator,
\[ \kappa(x) \triangleq \nabla \cdot \left( \frac{\nabla \theta(x)}{||\nabla \theta(x)||} \right), \tag{15} \]
it can be shown that the Euler-Lagrange equation describing the stationary points of eqn. (14) are given by:
\[ \phi(x) - \psi(x) - v \lambda \kappa(\theta(x)) = 0, \tag{16} \]
and because the logit transform is monotonic,
\[ \phi(x) - \psi(x) - v \lambda \kappa(\phi(x)) = 0. \tag{17} \]

As the objective function is strictly convex (see Section 3.2) in \(\theta\), the stationary point is the unique global optimum. In summary, the distribution \(Q(x; \theta)\) approximates the the “exact” distribution, \(P(x|y)\), in the KL-divergence sense when \(\phi\) (the logit transform of the parameter \(\theta\)) satisfies the Euler-Lagrange equation of the AMF model; we will refer to eqn. (17) as the “AMF Equation.”

3 Properties of AMF

In this section we summarize some approximation properties of AMF (Section 3.1), show the objective function to be convex (Section 3.2), and show that AMF is unbiased in a specific sense (Section 3.3).

3.1 Approximation Properties

Our goal is an efficient yet accurate approximation, \(Q(z; \theta)\) to the exact posterior distribution \(P(z|y)\) for general realizations of \(z\). Because the normalizer, \(P(y)\), for \(P(z|y)\) is not available we compare \(\frac{P(z|y)}{P(z_0|y)}\) and \(\frac{Q(z; \theta)}{Q(z_0; \theta)}\), where \(z_0\) is the most probable realization under both distributions. For convenience, we only summarize the results of some of the approximation properties of the AMF model here. Mathematical details can be found in the Appendix. In particular, the Appendix shows that

a) The zero level-set of \(\phi\) is the boundary of the most probable realization \((z_0)\) of \(Q(z; \theta)\) as well as the boundary of the MAP realization under \(P(z|y)\). This is not generally the case for mean field approximations.

b) The probability ratios are not only in agreement for the zero level-set, but also for for realizations that are bounded by any level-set of \(\phi\).

c) The probability ratios approximately agree for realization whose boundary normals are close in direction to \(\nabla \phi\).

d) If neither a) nor b) hold the probability ratio for \(Q\) will be larger than that for \(P\), i.e., it under-estimates the length penalty associated with the prior.

3.2 Convexity

A nice property of the AMF model is that its energy is strictly convex and therefore we can find a unique global minimizer. This is in contrast to the related total variation based segmentation models which are generally convex (but not strictly so) and therefore may have multiple non-unique optima.

To show convexity, we consider the continuum formulation (Bresson et al., 2007) which can be rewritten as a function of \(\theta(x) \in [0, 1]\), as:
\[ E_{amf}(\theta) = \int_X -\theta \psi + v \lambda \|\nabla \theta\|^2 \]
\[ + (1 - \theta) \log(1 - \theta) + \theta \log \theta \, dx \] \tag{18}
where dependencies on space are dropped only for notational convenience (i.e., \(\theta = \theta(x)\) and \(\psi = \psi(x)\)) and we expressed \(\Phi\) in terms of \(\theta\). The term
\[ \int_X -\theta \psi + \lambda \|\nabla \theta\|^2 \, dx \] \tag{19}
is convex in \(\theta\) as the first summand is linear in \(\theta\), the 2-norm is convex, \(\nabla\) is a linear operator and both terms are summed with a positive weight \(\lambda\). To see that the rest of the integrand is also convex, consider a function of the form
\[ f(\theta) = (1 - \theta) \log(1 - \theta) + \theta \log \theta. \]
which implies that
\[ f''(\theta) = \frac{1}{\theta(1 - \theta)} > 0 \quad \text{for} \quad \theta \in (0, 1). \]
Therefore, \( f_X(1-\theta) \log(1-\theta)+\theta \log \theta \, dx \) is strictly convex. Because the sum of convex and strictly convex functions is strictly convex, the overall AMF energy is strictly convex in \( \theta \) and therefore has a unique global minimizer (see (Boyd & Vandenberghe, 2004) for details on convexity preserving operations). In particular, we note that for a non-informative data term, i.e., pixels are locally equally likely to be foreground or background (\( \psi = 0 \)), \( \theta(x) = \frac{1}{2} \) is the globally optimal solution. For the related standard total-variation segmentation model (Appleton & Talbot, 2006), which would only minimize Equation (19), any constant solution would be a global minimizer.

### 3.3 Unbiased in Homogeneous Regions

In this section we analyze the behavior of the AMF estimator over homogeneous (i.e., constant intensity) patches of an image. The AMF objective function, eqn. (14) can be written:

\[
\hat{\theta} = \arg \min_\theta \int_X KL(Q(x; \theta))||P(y(x)|x(x))||dx + v\lambda TV[z].
\]

where \( Q(x; \theta) \) is a Bernoulli distribution over the binary random variable \( z(x) \) with parameter \( \theta(x) \). It is easy to see that over a patch \( X \) of constant intensity, i.e., \( \psi(x) = \psi_0 \), the optimum will be attained at \( \ln \theta(x)/(1-\theta(x)) = \phi(x) = \psi_0 \), as both the KL and TV terms will vanish. This in turn implies that the regularizer does not interact in homogeneous regions and an unbiased probability estimate is obtained.

In contrast, other Bayesian segmentation approaches, e.g. the Ising model (Ising, 1925), lack this “unbiased in homogeneous regions” property and because of this interaction with the regularizer, setting the regularization parameter \( \lambda \) in such cases can be tricky.

To illustrate this point, consider a VMF treatment of the Ising model that parallels the approach and notation used for AMF. Defining an Ising model where neighbors of \( x_i \) are specified by \( N(i) \),

\[
P(z|y) \propto p(y|z)P(z) \text{ where } P(z) \propto e^{-U(z)}
\]

and the neighborhood potential term is

\[
U(z) \propto \lambda \sum_i \sum_{j \in N(i)} z_i(1-z_j) + (1-z_i)z_j. \tag{21}
\]

Using the VMF approximation, we obtain:

\[
\hat{\theta} = \arg \min_\theta \left\{ KL(Q(Z; \theta)||P(y|z)) + \lambda E_{Q(Z; \theta)}[U(Z)] \right\}
\]

which yields the following stationary-point equation:

\[
\phi_i - \psi_i - 4\lambda \sum_{j \in N(i)} \left[ \theta_j - \frac{1}{2} \right] = 0, \tag{22}
\]

which are the consistency equations that characterize the solution of the VMF approximation to the Ising model. It is clear from Equation (22) that the regularization term will only be zero when \( \theta(x) = \frac{1}{2} \) while in other cases the unbiasedness property will not apply.

### 4 Connections to Chan-Vese and ROF

In this section we establish the connection between the AMF model and the Chan-Vese segmentation model (Section 4.1) as well as the ROF denoising model (Section 4.2). In particular, we show that the Chan Vese Euler-Lagrange equations (26) correspond to those of the zero level-set of the AMF model, so a Chan-Vese segmentation can be obtained as the zero level-set of the AMF solution. We also show that the AMF Euler-Lagrange equations (28) have the same form as the ones of the ROF model (29). Therefore, the solver technologies that have been developed for the ROF model may be deployed for AMF, as discussed in Section 4.3.

#### 4.1 Connection to Chan-Vese

The energy for a generalized Chan-Vese model using a relaxed indicator function (i.e., \( \theta \in [0,1] \)) as
suggested in [Bresson et al., 2007] can be written as:
\[
\mathcal{L}_{cv}(\theta) = \int_X -\theta(x)\psi(x) + \nu\lambda\|\nabla \theta(x)\|_2^2 \, dx, \tag{23}
\]
where the second term encodes the boundary length prior. Such a length prior is essential to encourage large, contiguous segmentation areas. Note that without the length prior, Chan-Vese segmentation becomes Otsu-thresholding (Otsu, 1975) which cannot suppress image fragmentation and irregularity. Furthermore, the importance of the length-prior becomes especially clear in the context of the Mumford-Shah model (Mumford & Shah, 1989) (of which the Chan-Vese model is a special case). In the absence of a length prior and assuming that each pixel has constant image intensity, Mumford-Shah segmentation will allocate a separate segmentation region to each pixel.

The standard Chan-Vese model without area prior can be recovered from this formulation for the special case that the class conditional intensity model is Gaussian, \(y_i|x_i = 1 \sim N(\mu_1, \sigma_1^2)\) and \(y_i|x_i = 0 \sim N(\mu_0, \sigma_0^2)\). Then, ignoring constants,
\[
\psi(x) = -\frac{1}{2\sigma_0^2} \left(y(x) - \mu_0\right)^2 + \frac{1}{2\sigma_1^2} \left(y(x) - \mu_1\right)^2. \tag{24}
\]

The Euler-Lagrange equations of the generalized Chan-Vese energy eqn. (23) can be obtained as:
\[
-\psi(x) - \nu\lambda\kappa(\theta(x)) = 0. \tag{25}
\]
This is identical to the AMF Euler-Lagrange equation at the zero level-set \(\phi(x) = 0\). By construction, the zero level-set of a level-set implementation for the generalized Chan-Vese model has to agree with the solution obtained from the Euler-Lagrange equations of the generalized Chan-Vese model using indicator functions as both minimize the same energy function just using different parameterizations. Consequently, also the zero level-sets of both the AMF model and the level-set implementation of Chan-Vese need to agree.

In contrast to the generalized Chan-Vese model described above, the original Chan-Vese model (Chan et al., 2000) contains an area penalty. Such penalties cannot effectively be added in the indicator-function based approaches to the Chan-Vese/active-contour models proposed by Appleton et al. (Appleton & Talbot, 2006) and Bresson et al. (Bresson et al., 2007) as their goal is to capture a binary segmentation result through a relaxed indicator function, (i.e., \(\theta \in [0, 1]\) instead of \(\theta \in \{0, 1\}\)). It can be shown that, in certain instances, adding such area terms with a relaxed indicator function produces undesirable results (Nickhammer & Zach, 2013). A curve evolution approach, as proposed in the context of a level-set evolution in the original Chan-Vese approach, does not suffer from such shortcomings as it does not use a relaxation. The corresponding Euler-Lagrange equations are, on the zero level-set (Chan et al., 2000),
\[
-\psi(x) - \nu\lambda\kappa(\phi(x)) + \nu = 0, \tag{26}
\]
where \(\nu\) stems from the energy term
\[
E_{\text{area}}(C) = \nu \text{Area}(\text{inside}(C))
\]
that penalizes the area enclosed by the evolving curve \(C\) which is implicitly described by the zero level-set of \(\phi\). Now, examining the \(\nu\) level-set of the AMF model, so that \(\phi(x) = \nu\), we notice that this level-set satisfies the same Euler-Lagrange equation as the zero level-set of the Chan-Vese model with a specified non-zero value of \(\nu\). In other words, the level-sets of the dense AMF solution provide a family of solutions for the Chan-Vese problem for a continuum of values of the area/volume penalty.

### 4.2 Connection between AMF and ROF Models

In their seminal paper, Rudin, Osher and Fatemi (Rudin et al., 1992) proposed a de-noising method for, e.g., intensity images \(u_0(x)\),
\[
u^*(x) = \arg \min_u \int_X \|\nabla u(x)\|_2^2 dx \\
\quad \text{s.t.} \quad \int_X (u(x) - u_0(x))^2 dx = \sigma^2. \tag{27}
\]
As discussed by Vogel and Oman (Vogel & Oman, 1996), this is equivalent to the following uncon-
strained problem,

$$u^*(x) = \arg \min_u \left[ \int_X (u(x) - u_0(x))^2 dx + \frac{\alpha}{2} \int_X \|\nabla u(x)\|_2^2 dx \right], \quad (28)$$

for a suitable choice of $\alpha$ (and setting their parameter $\beta$ to zero). They refer to this formulation as “TV penalized least squares.”

The corresponding Euler-Lagrange equation is

$$u(x) - u_0(x) - \alpha \kappa(u(x)) = 0. \quad (29)$$

Interestingly, this equation has the same form as the Euler-Lagrange equations of the AMF model,

$$\phi(x) - \psi(x) - \nu \lambda \kappa(\phi(x)) = 0. \quad (30)$$

In this equivalence, the de-noised intensity of the ROF model, $u$, corresponds to the logit parameter field of the AMF distribution, $\phi$, while the input intensity of ROF, $u_0$, corresponds to the logit likelihood of the labels in the AMF problem, $\psi$. Furthermore, if the class conditional intensity model is homoscedastic Gaussian, then (from Equation 24) $\psi(x)$ is linear in the observed intensity, and the AMF solution is equivalent to solving an ROF problem that is effectively de-noising the observed intensities, with scaling and offset so that the result is the estimated logit parameter of the posterior.

### 4.3 Solving the AMF model via ROF

Optimize

Because of the equivalence of the Euler-Lagrange equations of the AMF and the ROF models established in Section 4.2, the considerable technology that has been developed for solving the ROF model may be brought to bear on AMF problems. Both the AMF and the ROF models are strictly convex so a unique global minimizer will be obtained. In particular, this implies that the AMF model can be solved by solving the ROF model regarding $\phi$ as the image reconstruction and $\psi$ as the original image to be reconstructed. As there is a direct mapping from $\phi$ to $\theta$, this allows computing a solution for $\theta$ that fulfills the AMF Euler-Lagrange equations. As the solution is unique this will be the unique solution for the AMF model.

The ROF model was initially solved (Rudin et al., 1992) using a gradient descent method, and this may still be a reasonable option if AMF solving is embedded in an outer, e.g. expectation-maximization, iteration. The difficulty in computing the optimum of the ROF energy is due to the total-variation term that is not differentiable everywhere. Initial approaches, therefore, replaced the total-variation term by $\sqrt{\| \nabla u \|^2 + \beta^2}$ (Chan et al., 1999) that makes the energy everywhere differentiable, but slightly changes the optimization problem. To allow for better discretization of the total-variation term, primal-dual (Chan et al., 1999), and fully dual methods (Chambolle, 2004) have been explored. More recently, methods based on accelerated proximal gradient descent (FISTA) (Beck & Teboulle, 2009) and split Bregman iterations (Goldstein & Osher, 2009) have been applied to solve the ROF model. We use FISTA for the experiments described below. More specifically, given an ROF solver (ROFsolve) that minimizes

$$\mathcal{L}_{ROF}(u; u_0, \lambda) = \int_X (u(x) - u_0(x))^2 + \frac{\lambda}{2} \|\nabla u\|_2^2 dx$$

such that

$$u^* = \text{ROFsolve}(u_0, \lambda) = \arg \min_u \mathcal{L}_{rof}(u; u_0, \lambda), \quad (32)$$

solving the AMF problem for a given $\psi$ and $\lambda$ then simply becomes

$$\theta^* = \sigma(\text{ROFsolve}(\psi, \lambda)). \quad (33)$$

Equation (33) is the central result concerning practical implementations as it connects the optimal AMF solution to a straightforward ROF denoising problem. Hence, the solution of the AMF model becomes one line of source-code given an ROF solver.

### 5 Experiments

This section illustrates the behavior of the proposed AMF model using synthetic and real datasets. Sec-
Section 5.1 shows how the AMF model deals with ambiguous segmentation regions in comparison to a standard Chan-Vese model. Section 5.2 investigates how well the AMF model agrees with the original probability model without approximations. Finally, Section 5.3 applies the AMF model to real ultrasound data of the heart and the prostate and two computer vision test images: the cameraman and Fabio.

5.1 Segmentation with Ambiguity

A goal of AMF is to provide label probabilities from which the MAP solution for the segmentation can be obtained, but which also allow the assessment of segmentation uncertainty. To test this behavior, we generated a highly ambiguous segmentation scenario, depicted in Figure 1. The intensity distribution of the circle in the center of the image was chosen such that half of the circle has intensities that lie exactly in the middle between the foreground and background. In particular, the intensity distributions were obtained by a mixture of Gaussians with means \( \mu = \{30, 50, 70\} \) and corresponding standard deviations \( \sigma = \{5, 10, 5\} \). This is also illustrated by the class-conditional distribution on intensities in Figure 1(right). The results were obtained by assuming we know the correct means and standard deviations for the foreground and background classes; likelihoods were computed based on the noisy data. The regularization term was weighted with \( \lambda = 5 \).

Figure 2 (upper two images) shows the local label probabilities for the noisy input image and for the noise-free image (that will not be available in practice). Figure 2 (lower two images) shows the label probabilities after running the AMF model (left) and after thresholding (binarization) at \( P = 0.5 \) (right) that also corresponds to the MAP solution. As desired, the AMF model captures the segmentation uncertainty by estimating the upper part of the circle at a probability close to \( P = 0.5 \). At the same time, due to spatial regularization, the AMF model removes noise effects. The MAP solution captures the most likely foreground area, but completely loses the ambiguous area.

Figure 3 shows the estimated label probabilities and their true local counterparts along with a subtraction. The AMF method has effectively estimated the true label probabilities. Note that the true local label probabilities do not incorporate the effect of regularization. Hence, these two probabilities will slightly disagree at the segmentation boundaries.

5.2 Agreement with the Original Probability Model

In order to evaluate agreement between the original probability model

\[
\ln P(z|y) \approx v^{-1} \int_{\mathcal{X}} z(x)\psi(x) + \ln p(y(x)|0) \, dx - \lambda L(z) + \text{const.}
\] (34)

and the AMF approximation, we conducted the following set of experiments on synthetic images. A binary random field was generated by sampling on a 100×100 grid from a Gaussian process with Matérn covariance function (Cressie, 1991) with order parameter \( p \) and scale parameter \( l \), that provides fine-grained control over the smoothness of the field. This continuous valued image was then thresholded at a quantile value selected uniformly at random to create the ground truth binary label map \( \hat{z} \) to which Gaussian noise is added to create a noisy image \( y \). For our experiments, we set \( p = 1 \) while varying length scale parameter \( l = 1, 3, 5 \) and noise standard deviation \( \sigma = 0.25 \) and 0.4. Increasing \( l \) produces label maps with smoother edges and larger contiguous regions. Single realizations of \( \hat{z} \) for \( l = 1, 3, 5 \) are shown in Fig. 4(a–c). Corresponding noisy images for \( \sigma = 0.4 \) are shown Fig. 4(d–f).

For each setting of \( l \) and \( \sigma \), 50 ground truth binary label maps \( \hat{z} \) were realized and for each case, the AMF approximating distribution \( Q \) was computed by solving the ROF equations on the logit maps of the corresponding images \( y \). The original probability model \( P \) was also explored using Gibbs sampling with 5 chains of \( N = 10^5 \) particles each, temperature \( T = 1 \) and thinning factor=0.1. Convergence was tested using the Gelman and Rubin diagnostic (Gelman & Rubin, 1992) resulting in approximately \( 2 \times 10^4 \) particles being retained and the partition function was computed from the sample.
Figure 1: Ambiguous segmentation scenario. Left: original image, middle: noisy image, right: class conditional distributions. Distributions clearly overlap which should result in a segmentation ambiguity for the upper part of the circle which was deliberately chosen to have intensities in between the background and the foreground (bottom part of the circle).

Figure 2: Top left: noisy label probabilities based on the noisy image. Top right: noise-free label probabilities based on the noise-free image (which is not available in practice). Bottom left: estimated label probabilities using the AMF model. Bottom right: estimated MAP solution (binarization at $P = 0.5$) from the AMF-estimated label probabilities. Clearly, the AMF model captures more information – the MAP solution completely loses the ambiguity of the upper part of the circle.
Figure 3: Left: estimated label probabilities by the AMF model. Middle: noise-free label probabilities. Right: difference between the probabilities. Differences exist primarily at the segmentation boundaries, which is expected since the AMF model includes spatial regularization effects while the noise-free label probabilities are computed strictly locally. Overall, there is a good agreement between the probabilities.

The probability mass of each particle, according to $P$ and $Q$ for the different cases of $l$ and $\sigma$, are plotted in Fig. 5 (row 1). To improve clarity of the figure, values less than a tenth were discarded and only a tenth of the remaining particles (through random decimation) are shown. The X-axis is the probability value $P$ as evaluated by Gibbs sampling from the original probability model. The Y-axis is the AMF probability $Q$ of that sample. We observe a linear trend between the AMF and original models ($r^2 = 0.22$ for $\sigma = 0.40$ and $r^2 = 0.43$ for $\sigma = 0.25$). Smoother images are harder to approximate - because of an increase of non-local interactions that cannot be well approximated by a local mean-field distribution. In addition, increasing noise causes greater misapproximation.

Fig. 5 (row 2) shows representative cross plots of the mean area of the label-map estimated by Gibbs sampling from $P$, and from the closed-form evaluation of $Q$. Again, good agreement is achieved that deteriorates with image smoothness and noise. Row 3 shows a similar cross plot of the variance of the area of the label-map estimated by Gibbs sampling from $P$, and from the closed-form evaluation of $Q$. Although there are noticeable differences between the estimated variances under $P$ and $Q$, the variance estimates under $Q$ seem surprisingly good, given the approximation through the mean field $Q$.

5.3 Evaluation on Real Data

We used ultrasound images of the heart and the prostate as well as the cameraman image (as provided in Matlab) and the Fabio image (Needell & Ward, 2013) to assess the behavior of the AMF model on real data. Our goal in this section is not to beat state-of-the art segmentation methods for these applications (which may for example, use shape models to improve segmentation results for the medical images), but to illustrate the AMF approach in the context of challenging datasets. Note, however, that the AMF model can be based on any foreground and background probability map. Therefore, it is able to augment other more sophisticated pre-processing to obtain foreground and background probabilities. We limit ourselves to simple intensity distributions for the heart example and show results based on a probability map created by a prototype machine-learning approach for the prostate example. To avoid computational issues, probabilities were clamped to be in $[1e-5, 1e-5]$. Convergence for FISTA was left at the default value of $1e-4$ of the Matlab package by Amir Beck and Marc Teboulle. The maximum number of iteration steps was set to 10,000 but was never reached. Image size for the heart example was 314 by 350 pixels, for the prostate example 257 by 521 pixels, for the cameraman image 256 by 256 pixels, and for the Fabio image 253 by 254 pixels.
Figure 4 shows simulated label maps and corresponding noisy images. Figure 6 shows an ultrasound image of the heart (left), an expert segmentation into blood pool, myocardium, and valves (middle) and the intensity distribution for the blood pool and outside the heart (right). These intensity distributions clearly overlap. We initialized the AMF model with this user-defined intensity distribution by sampling from the image followed by kernel-density estimation of the intensities. We re-estimated the intensity-distributions during the optimization. Specifically, given an intensity distribution, we compute the AMF solution, from that we obtain the binarized MAP solution that we used to re-estimate the intensity distributions using kernel-density estimation. We alternate AMF solution and density estimation to convergence.

Figure 7 shows the results of the AMF model for the estimation of label probabilities. The intensity ambiguity is captured in the estimated label probabilities of the AMF model. Regularization behaves as expected: low regularization results in noisy label probability maps. Moderate to high regularization allows capturing of the blood pool (for the MAP solution) while declaring other regions ambiguous or low-probability. Very large regularization declares the full image ambiguous, as expected, because the model will, in this case, prefer overly large segmentation regions.

Figure 8 shows an ultrasound image of the prostate (left) and the corresponding results of an experimental prostate segmentation system (right). The prototype system analyzed Radio Frequency (RF) ultrasound data using deep learning and random forest classification to generate label probabilities. (Alternating optimization, as in the heart example, was not used.) Figure 9 shows the results of the AMF model. The same conclusions as for the heart example apply. More regularity yields cleaner looking probability fields. Changes are not as drastic as for the heart example as the initial probability map is already substantially more regular.

Figures 10 and 12 show the original cameraman and Fabio images, including their segmentations.
based on a modified version of Otsu thresholding (where foreground and background classes can have distinct means and standard deviations) and their corresponding intensity histograms. Both images can be separated reasonably well using intensity information alone.

Figures 11 and 13 show the corresponding AMF results. We obtained these results by initializing AMF using the modified Otsu-thresholding procedure and then followed the same alternating optimization approach as for the heart ultrasound segmentation. Clearly, larger regularization puts the emphasis on larger image structures.

6 Conclusions

We described a method for binary image segmentation which allows efficient estimation of approximate label probabilities through a variational mean field approximation. We carefully analyzed the theoretical properties of the model and tested its behavior on synthetic and real datasets. A particularly useful feature of our model is that it has strong connections to the Chan-Vese segmentation model and the ROF image-denoising model. This allows for an extremely easy implementation of the model by using off-the-shelf solvers for the ROF model. We believe our model to be an attractive alternative to Chan-Vese-like segmentation as it provides more information about a segmentation combined with a simple, efficient way to compute solutions. We deliberately focused on the binary segmentation problem to highlight connections to other approaches. Future work will extend our approach to the multi-label case.

7 Acknowledgements

This work was supported by NIH grants P41RR013218, P41EB015898, P41EB015902, NIAID R24 AI067039, R01CA111288 and NSF grants EECS-1148870, EECS-0925875. We also gratefully acknowledge Tassilo Klien and Andrey Fedorov for providing prostate data, and Dr. Theodore Abraham for the heart data.

Appendix: Comparison of Approximate and Exact Distributions

This section will compare the approximated and exact distributions, $Q(Z; \theta)$ and $P(Z|y)$, respectively, for general realizations of $z$. Because the normalizer for $P(Z|y)$ is not available, and for convenience, we will compare $\ln P(z|y)$ and $\ln Q(z; \theta)$, where $z_0$ is the most probable realization under both distributions.

For calculating the log probability ratio of $P$, we return to the original probability model. From Equations 1 - 3,

$$\ln P(z|y) = \sum_{i \in \mathcal{I}} z_i \psi_i - \lambda L(z) + \text{const} \quad (35)$$

$$\approx v^{-1} \int_{\mathcal{X}} z(x) \psi(x) dx - \lambda L(z) + \text{const}. \quad (36)$$

Then the log probability ratio for the exact posterior, $P$, is

$$\ln \frac{P(z|y)}{P(z_0|y)} \approx v^{-1} \int_{\mathcal{X}} (z(x) - z_0(x)) \psi(x) dx - \lambda L(z) + \lambda L(z_0).$$

Working towards the probability ratio for the AMF approximate posterior, $Q$, using Equation 4 and similar technique,

$$\ln Q(z; \theta) \approx v^{-1} \int_{\mathcal{X}} z(x) \phi(x) dx + \text{const}.$$ 

Here it is easy to see that the most probable realization under $Q$ is bounded by the zero level-set of $\phi$.

We may now write the log probability ratio for $Q$,

$$\ln \frac{Q(z; \theta)}{Q(z_0; \theta)} \approx v^{-1} \int_{\mathcal{X}} (z(x) - z_0(x)) \phi(x) dx.$$

Subtracting the two probability ratios,

$$\ln \frac{P(z|y)}{P(z_0|y)} - \ln \frac{Q(z; \theta)}{Q(z_0; \theta)} = v^{-1} \int_{\mathcal{X}} (z(x) - z_0(x))(\psi(x) - \phi(x)) dx - \lambda L(z) + \lambda L(z_0). \quad (37)$$
We now make use of the AMF equation, \( \phi(x) - \psi(x) - v \lambda \kappa(\phi(x)) = 0 \), to establish relationships among the log probability ratios of \( p \) and \( q \). We obtain

\[
\ln \frac{P(z|y)}{P(z_0|y)} - \ln \frac{Q(z; \theta)}{Q(z_0; \theta)} = \lambda \left[ - \int_X (z(x) - z_0(x)) \kappa(\phi(x)) dx - L(z) + L(z_0) \right]
\]

\[
= \lambda \left[ - \int_{X:x(x)=1} \nabla \cdot \left( \frac{\nabla \phi(x)}{|\nabla \phi(x)|} \right) dx
+ \int_{X:x_0(x)=1} \nabla \cdot \left( \frac{\nabla \phi(x)}{|\nabla \phi(x)|} \right) dx - L(z) + L(z_0) \right]
\]

\[
= \lambda \left[ - \int_{c(s)} N(x) \cdot \left( \frac{\nabla \phi(x)}{|\nabla \phi(x)|} \right) ds
+ \int_{c_0(s)} N(x) \cdot \left( \frac{\nabla \phi(x)}{|\nabla \phi(x)|} \right) ds - L(z) + L(z_0) \right].
\]

The last line uses the divergence theorem. \( c(s) \) is the boundary of \( z(x) \) oriented so that the outward normal points from \( z(x) = 1 \) towards \( z(x) = 0 \), and similarly for \( c_0(x) \) and \( z_0(x) \). \( N(x) \) is the outward normal vector to the curve in question (see Figure 14).

Then

\[
\ln \frac{P(z|y)}{P(z_0|y)} - \ln \frac{Q(z; \theta)}{Q(z_0; \theta)} = \lambda \left[ \int_{c(s)} \beta(x)ds - \int_{c_0(s)} \beta(x)ds - L(z) + L(z_0) \right],
\]

where

\[
\beta(x) = N(x) \cdot \left( \frac{-\nabla \phi(x)}{|\nabla \phi(x)|} \right)
\]

is the dot product of two unit vectors, the outward normal to the curve and the negative direction of the gradient of \( \phi \).

On the curve \( c_0 \), \( \beta(x) = 1 \), because the boundary of \( z_0 \) is a level-set of \( \phi(x) \). In that case the second and fourth terms cancel. Re-writing the third term as an integral over \( c \),

\[
\ln \frac{P(z|y)}{P(z_0|y)} - \ln \frac{Q(z; \theta)}{Q(z_0; \theta)} = \lambda \left[ \int_{c(s)} \beta(x)ds - \int_{c(s)} 1ds \right]
\]

\[
= \lambda \int_{c(s)} (\beta(x) - 1)ds. \quad (39)
\]

Because \( \beta(x) \) is the dot product of two unit vectors, we may write \( \beta(x) = \cos(\alpha(x)) \), where \( \alpha \) is the angle between the normal to the curve and the negative gradient direction of \( \phi(x) \). Then, using \( \cos(\alpha) - 1 = -2\sin^2\left(\frac{\alpha}{2}\right) \),

\[
\ln \frac{P(z|y)}{P(z_0|y)} - \ln \frac{Q(z; \theta)}{Q(z_0; \theta)} = -2\lambda \int_{c(s)} \sin^2\left(\frac{\alpha(x)}{2}\right) ds.
\]

See also Figure 14 for a graphical illustration.

Summarizing the comparison of the probability ratios of the exact and approximate distributions, \( P \) and \( Q \), respectively we see the following:

- For realizations that are bounded by level-sets of \( \phi \), \( \alpha \) is zero, so the probability ratios agree.
- For realizations whose boundaries are in direction “close” to level-sets of \( \phi \), the probability ratios approximately agree (the disagreement is quadratic in \( \alpha \)).
• For curves where $\alpha$ is not small, the probability ratio for $Q$ will be larger than for $P$, i.e., $Q$ underestimates the length penalty of $P$.

We saw above that the zero level-set of $\phi$ is the boundary of the most probable realization under the approximate distribution, $Q(z; \theta)$ (and it is unique). Hence, we can conclude that it is also the the boundary of the MAP realization under $P(z|y)$. In summary, $z_0$, whose boundary is the zero level-set of $\phi$, satisfies

$$z_0 = H(\phi(x)) = \arg \max_z Q(z; \theta) = \arg \max_z P(z|y).$$

**References**

Appleton, B., & Talbot, H. 2006. Globally minimal surfaces by continuous maximal flows. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 28(1), 106–118.

Beck, A., & Teboulle, M. 2009. Fast gradient-based algorithms for constrained total variation image denoising and deblurring problems. *Image Processing, IEEE Transactions on*, 18(11), 2419–2434.

Besag, Julian. 1986. On the Statistical Analysis of Dirty Pictures. *J Royal Stat Soc.. Series B (Methodological)*, 48(3), 259–302.

Bethuel, F., & Ghidaglia, J.-M. 1994. *Geometry in Partial Differential Equations*. World Scientific. Chap. 1, pages 1–17.

Boyd, S., & Vandenberghe, L. 2004. *Convex optimization*. Cambridge university press.

Boykov, Y., Veksler, O., & Zabih, R. 2001. Fast approximate energy minimization via graph cuts. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 23(11), 1222–1239.

Bresson, X., Esedoglu, S., Vandergheynst, P., Thiran, J.-P., & Osher, S. 2007. Fast Global Minimization of the Active Contour.Snake Model. *Journal of Mathematical Imaging and Vision*, 28, 151–167.

Burger, M., & Osher, S. 2013. A guide to the TV zoo. Pages 1–70 of: *Level Set and PDE Based Reconstruction Methods in Imaging*. Springer.

Caselles, V., Kimmel, R., & Sapiro, G. 1997. *Geodesic active contours*. *International journal of computer vision*, 22(1), 61–79.

Chambolle, A. 2004. An algorithm for total variation minimization and applications. *Journal of Mathematical Imaging and Vision*, 20(1-2), 89–97.

Chan, T. F., & Esedoglu, S. 2005. Aspects of total variation regularized L1 function approximation. *SIAM Journal on Applied Mathematics, 65*(5), 1817–1837.

Chan, T. F., & Esedoglu, S. 2005. Aspects of total variation regularized L1 function approximation. *SIAM Journal on Applied Mathematics, 65*(5), 1817–1837.

Chan, T. F., Golub, G. H., & Mulet, P. 1999. A non-linear primal-dual method for total variation-based image restoration. *SIAM Journal on Scientific Computing, 20*(6), 1964–1977.

Chan, T. F., Sandberg, B. Y., & Vese, L. A. 2000. Active contours without edges for vector-valued images. *Journal of Visual Communication and Image Representation*, 11(2), 130–141.

Chandler, D. 1987. *Introduction to Modern Statistical Mechanics*. Oxford University Press.

Cremers, D., Rousson, M., & Deriche, R. 2007. A review of statistical approaches to level set segmentation: integrating color, texture, motion and shape. *International journal of computer vision*, 72(2), 195–215.

Cressie, N. 1991. *Statistics for Spatial Data*. John Wiley & Sons.

Dragomir, S. S. 2004. A converse of the Jensen inequality for convex mappings of several variables and applications. *Acta Mathematica Vietnamica*, 29(1), 77–88.

Gelman, A., & Rubin, D. B. 1992. Inference from Iterative Simulation Using Multiple Sequences. *Statistical Science*, 7, 457–472.
Georgiou, T., Michailovich, O., Rathi, Y., Malcolm, J., & Tannenbaum, A. 2007. Distribution metrics and image segmentation. *Linear algebra and its applications*, 425(2), 663–672.

Goldstein, T., & Osher, S. 2009. The split Bregman method for L1-regularized problems. *SIAM Journal on Imaging Sciences*, 2(2), 323–343.

Held, K, Kops, Elena Rota, Krause, BJ, III, William Wells, Kikinis, Ron, & Müller-Grtner, HW. 1997. Markov Random Field Segmentation of Brain MR Images. 16(6), 878–886.

Ising, Ernst. 1925. Beitrag zur Theorie des Ferromagnetismus. *Zeitschrift f"ur Physik A Hadrons and Nuclei*, 31(1), 253–258.

Kapur, Tina, Eric, W, Grimson, L, Kikinis, Ron, & Wells, William M. 1998. Enhanced spatial priors for segmentation of magnetic resonance imagery. Pages 457–468 of: *Medical Image Computing and Computer-Assisted InterventionMICCAI98*. Springer.

Kass, M., Witkin, A., & Terzopoulos, D. 1988. Snakes: Active contour models. *International journal of computer vision*, 1(4), 321–331.

Kichenassamy, S., Kumar, A., Olver, P., Tannenbaum, A., & Yezzi, A. 1995. Gradient flows and geometric active contour models. Pages 810–815 of: *Computer Vision, 1995. Proceedings., Fifth International Conference on*. IEEE.

Kolmogorov, V., & Zabih, R. 2004. What energy functions can be minimized via graph cuts? *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 26(2), 147–159.

Li, S. Z. 2009. *Markov Random Field Modeling in Image Analysis*. Springer.

Malladi, R., Sethian, J. A., & Vemuri, B. C. 1995. Shape modeling with front propagation: A level set approach. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 17(2), 158–175.

Mumford, D., & Shah, J. 1989. Optimal approximations by piecewise smooth functions and associated variational problems. *Communications on pure and applied mathematics*, 42(5), 577–685.

Needell, D., & Ward, R. 2013. Stable image reconstruction using total variation minimization. *SIAM Journal on Imaging Sciences*, 6(2), 1035–1058.

Niethammer, M., & Zach, C. 2013. Segmentation with area constraints. *Medical image analysis*, 17(1), 101–112.

Osher, S., & Fedkiw, R. 2003. *Level Set Methods and Dynamic Implicit Surfaces*. Springer.

Otsu, N. 1975. A threshold selection method from gray-level histograms. *Automatica*, 11(285-296), 23–27.

Paragios, N., & Deriche, R. 2002. Geodesic active regions: A new framework to deal with frame partition problems in computer vision. *Journal of Visual Communication and Image Representation*, 13(1), 249–268.

Pleka, T. 1982. Convergence condition of the TAP equation for the infinite-ranged Ising spin glass model. *Journal of Physics A: Mathematical and general*, 15(6), 1971.

Pock, Thomas, Unger, Markus, Cremers, Daniel, & Bischof, Horst. 2008. Fast and exact solution of total variation models on the GPU. Pages 1–8 of: *Computer Vision and Pattern Recognition Workshops, 2008. CVPR’08. IEEE Computer Society Conference on*. IEEE.

Pohl, K. M., Fisher, J., Grimson, W. E. L., Kikinis, R., & III, W. M. Wells. 2006. A Bayesian model for joint segmentation and registration. *NeuroImage*, 31(1), 228–239.

Pohl, K. M., Kikinis, R., & III, W. M. Wells. 2007a. Active Mean Fields: Solving the Mean Field Approximation in the Level Set Framework. Pages 26–37 of: Karssmeijer, Nico, &
Lelieveldt, Boudewijn (eds), *Information Processing in Medical Imaging*. Lecture Notes in Computer Science, vol. 4584. Springer Berlin Heidelberg.

Pohl, K. M., Fisher, J., Bouix, S., Shenton, M., McCarley, R. W., Grimson, W. E. L., Kikinis, R., & III, W. M. Wells. 2007b. Using the logarithm of odds to define a vector space on probabilistic atlases. *Medical Image Analysis*, 11(5), 465–477.

Rudin, L. I., Osher, S., & Fatemi, E. 1992. Nonlinear total variation based noise removal algorithms. *Physica D: Nonlinear Phenomena*, 60(1), 259–268.

Sapiro, G. 2001. *Geometric Partial Differential Equations and Image Analysis*. Cambridge.

Szeliski, R., Zabih, R., Scharstein, D., Veksler, O., Kolmogorov, V., Agarwala, Aseem, Tappen, M., & Rother, C. 2008. A Comparative Study of Energy Minimization Methods for Markov Random Fields with Smoothness-Based Priors. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 30(6), 1068–1080.

Vannier, Michael W, Butterfield, Robert L, Jordan, Douglas, Murphy, William A, Levitt, Robert G, & Gado, Mokhtar. 1985. Multispectral analysis of magnetic resonance images. *Radiology*, 154(1), 221–224.

Vogel, C. R., & Oman, M. E. 1996. Iterative methods for total variation denoising. *SIAM Journal on Scientific Computing*, 17(1), 227–238.

Wainwright, Martin J, & Jordan, Michael I. 2008. Graphical models, exponential families, and variational inference. *Foundations and Trends® in Machine Learning*, 1(1-2), 1–305.

Zhang, Yongyue, Brady, Michael, & Smith, Stephen. 2001. Segmentation of brain MR images through a hidden Markov random field model and the expectation-maximization algorithm. *Medical Imaging, IEEE Transactions on*, 20(1), 45–57.
Figure 5: Row 1 - shows the probability of a label-map under the original model assessed by Gibbs sampling versus the AMF model. Row 2: Mean area of the foreground under the two models. Row 3: Variance of the areas of the foreground under the two models.
Figure 6: Left: ultrasound image of the heart. Middle: ultrasound image of the heart with overlaid expert segmentations of blood pool (red), myocardium (blue) and valves (yellow). Right: intensity distributions for the blood pool and the areas outside of the heart for the intensity-normalized image ($I \in [0, 1]$). Intensity distributions clearly overlap making an intensity-only segmentation challenging.
Figure 7: Intensity-based segmentation results of the heart from an ultrasound image for the AMF model. Increased regularization captures increasingly consistent regions. Moderate to high regularization retain high probabilities of the blood pool while estimating low probabilities for the surroundings. Very large regularizations yield ambiguous label probabilities throughout the complete image. Magenta contour indicates expert segmentation of the blood-pool, blue contour indicates the 0.5 probability isocontour of the AMF solution. (Best viewed zoomed.)
Figure 8: Left: ultrasound image of the prostate. Right: prostate probability map obtained by a machine-learning approach.

Figure 9: Probability-map-based segmentation results of the prostate from an ultrasound image for the AMF model. Increased regularization captures increasingly consistent regions. Moderate to high regularization retain high probabilities of the prostate while estimating low probabilities for the surroundings. Very large regularizations yield ambiguous label probabilities throughout the complete image. Blue contour indicates the 0.5 probability isocontour of the AMF solution. (Best viewed zoomed.)

Figure 10: Left: Cameraman image. Middle: Otsu-thresholded cameraman image. Right: intensity distributions for the intensity-normalized image ($I \in [0, 1]$) based on the classes determined by Otsu thresholding.
Figure 11: Intensity-based segmentation results for the cameraman image for the AMF model.
Figure 12: Left: Fabio image. Middle: Otsu-thresholded Fabio image. Right: intensity distributions for the intensity-normalized image ($I \in [0, 1]$) based on the classes determined by Otsu thresholding.
Figure 13: Intensity-based segmentation results for the Fabio image for the AMF model.