An Estimation Theoretical View on Ambrosio-Tortorelli Image Segmentation

Kai Krajsek¹, Ines Dedovic¹, and Hanno Scharr¹

¹ IBG-2: Plant Sciences
Forschungszentrum Jülich
52425 Jülich, Germany

In this paper, we examine the Ambrosio-Tortorelli (AT) functional [1] for image segmentation from an estimation theoretical point of view. Instead of considering a single point estimate, i.e. the maximum-a-posteriori (MAP) estimate, we adopt a wider estimation theoretical viewpoint, meaning we consider images to be random variables and investigate their distribution. We derive an effective block-Gibbs-sampler for this posterior probability density function (PDF) based on the theory of Gaussian Markov random fields (GMRF) [2].

1 Introduction

In this paper, we examine the theoretical properties of a block Gibbs sampler [2] for the Ambrosio-Tortorelli (AT) image segmentation problem [1] which has previously been published in [3]. Gibbs sampling on discrete state spaces has been introduced by Geman and Geman [4] and theoretically studied by [5, 6] on discrete as well as continuous state spaces. Markov chain Monte Carlo on continuous state spaces can be traced back to the work of Harris [7]. The Mumford-Shah (MS) [8] introduced by Geman and Geman [4] and theoretically studied by [5, 6] on discrete as well as continues state spaces. Markovian chains on continuous state spaces can be traced back to the work of Harris [7].

2 Bayesian Estimation Theory

Let $(\Theta, \mathcal{F}, P)$ denote the probability space consisting of the sample space $\Theta$, the $\sigma$-algebra $\mathcal{F}$ of events and a probability measure $P$ and let $(X, \mathcal{E})$ denote the state space with corresponding $\sigma$-algebra $\mathcal{E}$. The state space in this paper will be the entire Euclidean space $\mathbb{R}^N$ for some $N \in \mathbb{N}$ with corresponding Borel $\sigma$-algebra $\mathcal{E}$. A real-valued random variable $z : \Theta \rightarrow \mathbb{R}^N$ is a function from the sample space $\Theta$ to the state space $\mathbb{R}^N$ which is $(\mathcal{F}, \mathcal{E})$ measurable. To each subset $A \in \mathcal{E}$ a probability $P(A) := P(z^{-1}(A))$ can then be assigned describing the chance to find a realization of the random event $z$ within $A$. We formulate our AT estimation problem on a regular discrete image domain $\mathcal{G}_h$ with grid size $h$. Then, images can be represented as column vectors $u, v, g \in \mathbb{R}^N$ and $z = (u, v)$, respectively. Gradients $\nabla u, \nabla v$ are approximated by finite difference operators that can be described by matrices acting on the column vectors. Bayesian estimators are characterized by means of their risk $R(\hat{z}) = E[L]$ defined by the expectation of a loss function $L : \mathbb{R}^N \rightarrow \mathbb{R}_+^1, e \rightarrow L(e)$ with respect to the posterior PDF $p(z|g)$. Prominent estimators are given by the quadratic loss function $L(e) = \|e\|^2$ leading to the minimum mean squared error estimator (MMSEE)

$$\hat{z} = \int z p(z|g) dz ,$$

and the (vector) hit and miss loss function $L(e) = 0$ for $\|e\| \leq \delta$ and $L(e) = 1$ for $\|e\| > \delta$ for sufficient small $\delta$, leading to the MAP estimator $\hat{z} = \arg\max_z p(z|g)$ which is usually applied in AT image segmentation. We approximate the MMSEE

* Corresponding author: e-mail k.krajsek@fz-juelich.de
by generating $n$ samples $z^n$ from the posterior PDF $p(z|g)$. The MMSE is then simply obtained by its sample mean. This means that we transform the original AT image segmentation problem into an estimation problem without the need of any optimization step for a highly non-convex energy functional.

3 Markov Chain Monte Carlo Approximation

Markov Chain Monte Carlo denote techniques generating a stochastic process, the Markov chain (MC), $z^0, z^1, \ldots$ which approximate samples from a target distribution $p(z|g)$ (cmp. with [10]). These samples obey the Markovian property: the transition probability of the next sample $z^{n+1}$ depends only on the present sample $z^n$, i.e. $P(z_{j+1} \in A|z^n, z^{j-1}, \ldots, z^0) = P(z_{j+1} \in A|z^n)$, without any influence of the past events $z^{j-1}, \ldots, z^0$. Let $\mu$ be an arbitrary probability measure. The left multiplication with a Markov kernel is defined by $\mu P(A) = \int P(A|z) \mu(dz)$ and the multiplication of two kernels by $(P_1 P_2)(A|z) = \int P_2(A|y)P_1(dy|z)$. In particular, we obtain the $n$-time transition Markov kernel by recursion $P^n(A|z) = \int P(A|y)P^{n-1}(dy|z)$. Important properties of MCs are irreducibility and Harris recurrence. A Markov kernel is denoted as $P(|g)-$irreducible if for every $z \in X$ and each $A \in \mathcal{E}$ with $P(A|g) > 0$ there exist at least one $n_0$ such that $P^{n_0}(A|z) > 0$ holds. Loosely speaking, each subset $A$ with positive posterior probability $P(A|g) > 0$ can be reached from any point in $X$. A probability measure $\mu$ is denoted as invariant with respect to the Markov kernel if $\mu P = \mu$. The core idea of MCMC is to generate samples from a Markov chain whose invariant distribution is the target distribution which is unique in case of a $P(|g)-$irreducible Markov kernel. A Markov chain is denoted as positive Harris (recurrent) if it is $P(|g)-$irreducible and each subset $A$ with positive measure $P(A|g) > 0$ is visited infinitely often with probability one. We have now gathered all necessary terms in order to formulate the main proposition needed for our purpose. It follows directly by combining Theorem 17.0.1 and Theorem 17.1.6 in [10].

Proposition 1 Let $z^0, z^1, \ldots$ be a positive Harris Markov chain with invariant distribution $P(|g)$ and with arbitrary initial distribution $\mu$. If the expectation of the absolute value of each component $z_i, i = 1, 2, \ldots, 2N$ is finite, i.e. $\int |z_i| p(z|g)dz < \infty$, the law of large numbers holds:

$$
\lim_{n \to \infty} \frac{1}{n} \sum_{j=0}^{n-1} z_j = \int z p(z|g)dz \quad \text{a.s. . (3)}
$$

As Proposition 1 holds for any initial distribution, it holds in particular for initial distributions concentrated at a single point $z_0$, i.e. $\mu = \delta_{z_0}$. Consequently, we can start our MC at any point in the state space to approximate the MMSE estimator with the sample mean. The first step of our approach is to approximate the AT energy functional (1) on a discrete grid. By investigating the posterior PDF $p(u,v|g)$ of the AT energy, we recognize that by fixing $u$ or $v$ the resulting AT energy becomes a quadratic function of the other variable and consequently the corresponding conditional PDFs become Gaussian distributions from which samples can be obtained directly. The two step block-Gibbs-sampler [2] partitions the vector $z = (u, v)$ such that we can sample from the conditional PDFs $u^{j+1} \sim p(u|v^{j+1}, g)$ and $v^{j+1} \sim p(v|u^{j+1}, g)$ in turn. Before applying Proposition 1 we have to assure to fulfill its conditions: The target distribution $P(A|g)$ is per construction an invariant distribution of the block Gibbs Markov kernel [2]. The irreducibility of the block Gibbs sampler follows directly from the fact that all Markov kernels are Gaussian distributions, i.e. the conditional PDFs are nonnegative over the whole state space $X$. Consequently, there is a chance to reach each subset $A$ with $P(A|g) > 0$ in each step of the block Gibbs sampler. Our Markov chain is positive Harris (recurrent), which follows directly from Corrolary 1 in [6]. Finally, the condition $\int |z_i| p(z|g)dz < \infty$ can always be accomplished by introducing an appropriate Gaussian prior for $u$ and $v$, respectively.

Acknowledgements The research leading to these results has received funding from the European Communities Seventh Framework Programme FP7/2007-2013 Challenge 2 Cognitive Systems Interaction Robotics under grant agreement No 247947 GARNICS

References