Statistical and Learning Techniques in Computer Vision
Lecture 7: Approximate Inference

Jens Rittscher and Chuck Stewart

1 Overview

- Motivation
- Iterated Conditional Modes (ICM)
- Model Paramters
- Pseudolikelihood

2 Motivation

Any property of the posterior distribution $p(\mathbf{x}|\mathbf{y})$, or in the case of image restoration $p(\mathbf{I}|I_0)$, can be be simulated by running a Gibbs sampler as proposed by Geman and Geman [GG84]. Consider the following:

- While Geman and Geman [GG84] make only very weak assumptions to ensure convergence. Any asynchronous method of updating, whether deterministic or stochastic, is acceptable, provided each pixel is visited infinitely often.
- The MAP estimate can be estimated from a single realization of the chain. But it is not quite clear on how long the chain needs to be.
- The MCMC approach demands a prohibitive amount of computational resources.

In order to make use of such complicated models in practice we need to consider some alternatives

3 Iterated Conditional Modes (ICM)

In the specific case of image restoration the posterior was given by

$$p(\mathbf{I}|I_0) = p(I_0|\mathbf{I})\ p(\mathbf{I})\ , \tag{1}$$

where I_0 was the given noisy image. Assume that $\hat{I}$ is a provisional estimate of the de-noised image . By making use of the local Markov property we have

$$p(\mathbf{i}_p|I_0, \hat{I}_{S\backslash p}) \approx p(i_p^0|\mathbf{i}_p)\ p(\mathbf{i_p}|\hat{I}_{\delta p})\ .$$

For any locally-dependent MRF this expression will only involve a small number of random variables.

Let us review that this conditional probability has a simple form. The potential representation of the posterior implied was given as

$$p(\mathbf{I}|I_0) = \exp\left(\sum_{<s,p>} G(s,t)\right) \ . \tag{2}$$

Hence

$$\begin{aligned} p(\mathbf{i}_p|I_0, \hat{I}_{S\setminus p}) &= \frac{p(\mathbf{i}_p|I_0, \hat{I}_{S\setminus p})}{p(I_0, \hat{I}_{S\setminus p})} \\ &= \frac{\exp\left(\sum_{<p,q>} G(p,q)\right) \ \exp\left(\sum_{\substack{<s,t> \\ s\neq p, t\neq p}} G(s,t)\right)}{\exp\left(\sum_{p}\sum_{<p,q>} G(p,q)\right) \ \exp\left(\sum_{\substack{<s,t> \\ s\neq p, t\neq p}} G(s,t)\right)} \\ &= \frac{\exp\left(\sum_{<p,q>} G(p,q)\right)}{\sum_{p} \exp\left(\sum_{<p,q>} G(p,q)\right)} \ . \end{aligned} \tag{3}$$

which will only depend on a small neighbourhood of the site p. In certain cases it would possible to determine the MAP estimate of the *conditional mode*

$$i_p^* = \max_{i_p} \ p(\mathbf{i}_p|I_0, \hat{I}_{S\setminus p})$$

in closed form. Applying this estimate to all modes of the MRF leads to an algorithm that is known as *Iterated Conditional Modes* or ICM.

- **Initialization.** Pick $\hat{I}$.
- **Iterate for** $j = 1, 2, 3, \ldots J$
  - **ICM Step.** For each $s \in S$ determine

$$i_s^{(j)} = \max_{i_s} \ p(\mathbf{i}_s|I_0, \hat{I}_{S\setminus s}) \ .$$

The separate ICM step does not need to be restricted to individual sites, pixel locations in our case. An extension of the algorithm known as *Block ICM* updates a subset of random variables $T \subset S$ in one step.

4 Model Parameters

In general our models will of course depend on certain parameters, i.e.

$$p(\mathbf{x};\theta) = \exp\left(\sum_{C\in\mathcal{C}} \phi_C(\mathbf{x}_C,\theta)\right) \ . \tag{4}$$

In practice the optimal parameter $\hat{\theta}$ will be estimated using a training set $(x_1,\ldots,x_T)$. The parameter θ is a hidden parameter in the same way we introduced a set of hidden parameters for the estimation of the Gaussian mixture model using EM.

Example 4.1 The posterior for the problem of images restauration is

$$p(\mathbf{I}|I_0) \propto exp\left(-\alpha \sum_{<s,p>} |\mathbf{i}_s - \mathbf{i}_p|^2 - \beta \sum_s |\mathbf{i}_p - \mathbf{i}_p^0|^2\right) \ . \tag{5}$$

Clearly α and β are free parameters which we have not discussed so far. In case we choose $\alpha \gg \beta$ very large our smoothess prior will, for example, dominate. The other extrem $\alpha \ll \beta$ would indicate that our reconstructed image is very very close to the noisy image I_0.

Given a training example (I_0, I^*) we need to estimate those parameters. By learning the parameters α and β we would in way characterize the imaging process. In case we are looking at a different processs we would need to relearn these paramerts.

5 Pseudolikelihood

Similarly to the idea of ICM we use the conditional modes for parameter estimation. Besag introduced the notion of a *pseudolikelihood function* for $T \subset S$ is given by

$$PL_T(x,\theta) = \log\left(\prod_{s\in T} p(x_s|x_{S\setminus s};\theta)\right) \tag{6}$$

Assume that that we now have a training sample x_0. We can use the pseudolikelihood function to formulate the *maximum pseudolikelihood estimator* (MPLE) estimator as

$$\theta^* = \max_\theta \sum_{s\in T}\left(\log p(x_s^0|x_{S\setminus s}^0;\theta) - \log\sum_{x_s^0} p(x_p^0|x_{S\setminus s}^0;\theta)\right) \tag{7}$$

The maximum pseudolikelihood estimator for the observation window T maximises $PL(\cdot)$. Under certain conditions the MPLE estimator is unbiased

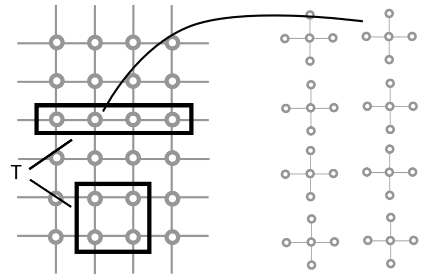

Figure 1: **Pseudolikelihood.** *Given a subset $T \subset S$ the pseudolikeihood function treats each site in T as an independent sample.*

and is asymptotically consistent (meaning: increasing size of the window T) [Win95, Theorem 14.3.1].

In order to compute the MPLE estimator in practice it is desirable that the the potential only depends on the parameters linearly, i.e.

$$p(\mathbf{x};\theta) = \exp\left(\sum_{C\in\mathcal{C}} \phi_C(\mathbf{x}_C,\theta)\right) = \exp\left(\sum_{C\in\mathcal{C}} < \theta, \phi'_C(\mathbf{x}_C) > \right)$$

In this particular case the pseudolikelihood function is concave [Win95] and the gradient can be computed as

$$\nabla PL_T(x^0;\theta) = \sum_{s\in T}\left[\phi(x_s^0, x_{\delta(s)}^0) - \mathbf{E}[\phi(x_s^0, x_{\delta(s)}^0)|x_{S\setminus s}^0;\theta]\right] \ , \tag{8}$$

where $\mathbf{E}[\phi(x_s^0, x_{\delta(s)})|x_{S\setminus s}^0;\theta]$ denotes the conditional expectation with respect to the distribution $p(x_s|x_{\delta(s)};\theta)$.

6 Further Reading

- Perez gives a very nice overview of Markov random fields for images [**?**]. This is a very good tutorial paper which summarizes the ideas presented so far.
- Although ICM was introduced earlier common references to both the algorithms and the pseudolikelihood estimation are written by Besag [**?**, Bes86]. For a good overview of these ideas read Besag [Bes86].
- Winkler [Win95] gives a detailed mathematical background of the convergence properties of the pseudoliklihood estimator.

References

[Bes86] J. Besag. On the statistical analysis of dirty pictures. *Journal of the Royal Statistical Society. Series B*, 48(3):259–302, 1986.

[GG84] S. Geman and D. Geman. Stochastic relaxation, gibbs distribtutions, and the bayesian restauration of images. *IEEE. Trans. Pattern Anal*, 6(6):721–741, November 1984.

[Win95] G. Winkler. *Image analysis, random fields and dynamics Monte Carlo methods: a mathematical introduction*. Spinger, 1995.