Contents

Chapter 1

Discrete Valued Markov Random Fields

A serious disadvantage of Markov chain structures is that they lead to image models that are not isotropic. This is due to the fact that one must choose a 1-D ordering of the pixels. In fact for most applications, there is no natural 1-D ordering for the pixels in a plane.

Markov random fields (MRF) have been introduced as a class of image models that do not require a 1-D ordering of the image pixels, and therefore can produce more natural and isotropic image models. However, as we will see the disadvantage of MRF models is that problems such as parameter estimation can be much more difficult due to the intractable nature of the required normalizing constant. The key theorem required to work around this limitation is the Hammersley-Clifford Theorem which will be presented in detail. The following sections explain the theory and methods associated with discrete valued MRFs.

1.1 Definition of MRF and Gibbs Distributions

Before we can define an MRF, we must first define the concept of a neighborhood system. Let S be a set of lattice points with elements $s \in S$. Then we use the notation ∂s to denote the neighbors of s. Notice that ∂s is a subset of S, so the function ∂ is a mapping from S to the power set of S, or equivalently the set of all subsets of S denoted by 2^S.

However, not any mapping ∂s qualifies as a neighborhood system. In order

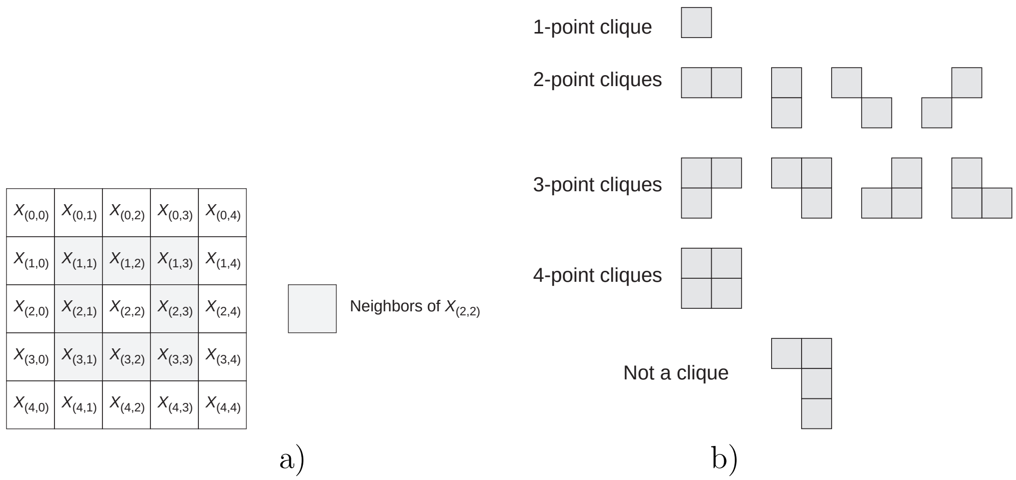

Figure 1.1: An eight point a) neighborhood system, and b) its associated cliques.

for ∂s to be a neighborhood system, it must meet the following symmetry constraint.

Definition 1 *Neighborhood system*
Let S be a set of lattice of points, then any mapping $\partial : S \rightarrow 2^S$ is a neighborhood system if for all $s, r \in S$

$$r \in \partial s \Rightarrow s \in \partial r \quad and \quad s \notin \partial s$$

In other words, if r is a neighbor of s, then s must be a neighbor of r; and it addition, s may not be a neighbor of itself. Notice that this definition is not restricted to a regular lattice. However, if the lattice S is a regular lattice, and the neighborhood is spatially invariant, then symmetry constraint necessitates that the neighbors of a point must be symmetrically distributed about each pixel. Figure 1.1a) shows such a symmetric 8-point neighborhood.

We may now give a general definition for MRFs.

Definition 2 *Discrete(Continuous) Markov Random Field*
Let $X_s \in \Omega$ be a discrete(continuous) valued random field defined on the lattice S with neighborhood system ∂s. Further assume that the X has probability mass(density) function $p(x)$. Then we say that X is a Markov random field (MRF) if its density function has the property that for all $x \in \Omega$

$$p(x_s | x_r \ for \ r \neq s) = p(x_s | x_{\partial r}) \ .$$

Notice that each pixel is only dependent on its neighbors.

A limitation of MRFs is that their definition does not yield a natural method for writing down the MRF's distribution. For this purpose, we will need to introduce the Gibbs distribution. We start by defining the concept of cliques which will be an integral part of the structure of Gibbs distributions.

Definition 3 *Clique*
Given a lattice S and neighborhood system ∂s, a clique is any set of lattice points $c \subset S$ such at for all $s, r \in c$, $r \in \partial s$.

Cliques are sets of point which are all neighbors of one another. Examples of cliques for an eight point neighborhood system on a rectangular are illustrated in Figure 1.1b) With this definition of cliques, we may now define the concept of a Gibbs distribution.

Definition 4 *Discrete (Continuous) Gibbs Distribution*
Let $p(x)$ be the probability mass(density) function of a discrete(continuous) valued random field $X_s \in \Omega$ defined on the lattice S with neighborhood system ∂s. Then we say that $p(x)$ is a Gibbs distribution if it can be written in the form

$$p(x) = \frac{1}{Z} \exp\left\{ -\sum_{c \in \mathcal{C}} V_c(x_c) \right\}$$

where Z is a normalizing constant known as the **partition function**, *$\mathcal{C}$ is the set of all cliques, x_c is the vector containing values of x on the set c, and $V_c(x_c)$ is any functions of x_c.*

We sometimes refer to the function $V_c(x_c)$ as a **potential function** and the function

$$U(x) = \sum_{c \in \mathcal{C}} V_c(x_c)$$

as the **energy function**.

The important result that relates MRFs and Gibbs distributions is the Hammersley-Clifford Theorem[1] stated below.

Theorem 1 *Hammersley-Clifford Theorem*
Let S be an N point lattice with neighborhood system ∂s, and X be a discrete(continuously) valued MRF on S with strictly positive probability mass(density) function $p(x) > 0$. Then X is an MRF if and only if $p(x)$ is a Gibbs distribution.

1.2 1-D MRFs and Markov Chains

1.3 The Ising Model and

1.4 Simulation

1.4.1 The Gibbs Sampler

In this section, we introduce the Gibbs sampler first presented in [2]. The Gibbs sample is a general method for producing samples from a distribution. It is particularly useful when the distribution being sampled is a Gibbs distribution, and the resulting samples form a Markov random field.

Let X_s be a finite dimensional random field that takes on values in a discrete and finite set Ω for all $s \in S$. If we assume that that the distribution of X is strictly positive, then without loss of generality, we know that the distribution of X can be written in the form

$$p(x) = \frac{1}{z} \exp\{-u(x)\} \tag{1.1}$$

where $u(x)$ is a real valued function. In fact, it is always possible to choose the neighbors of each pixel so that $\partial s = S$. In this case, the entire lattice S forms a clique, (1.1) is then a Gibbs distribution, and X is an MRF with a degenerate neighborhood system. In any case, the marginal distribution of a pixel can be written as

$$p\left(x_s | x_i \, i \neq s\right) = \frac{\exp\{-u\left(x_s | x_i \, i \neq s\right)\}}{\sum_{x'_s \in \Omega} \exp\{-u\left(x'_s | x_i \, i \neq s\right)\}} \tag{1.2}$$

We can generate samples from the distribution of (1.1) by using the following Gibbs sampler algorithm.

Gibbs Sampler Algorithm:

1. Set $N = \#$ of pixels
2. Order the N pixels as $N = s(0), \cdots, s(N-1)$
3. Repeat for $k = 0$ to ∞
   (a) Form $X^{(k+1)}$ from $X^{(k)}$ via

$$X_r^{(k+1)} = \begin{cases} W & \text{if } r = s(k) \\ X_r^{(k)} & \text{if } r \neq s(k) \end{cases}$$

where $W \sim p\left(x_{s(k)} \middle| X_i^{(k)}\, i \neq s(k)\right)$

We next show that the Gibbs sampler converges to the distribution of (1.1).

Theorem 2 *Stationary Distribution of Gibbs Sampler*
Let $p(x)$ be strictly positive distribution on Ω^N where Ω is a discrete and finite set. Then the Gibbs Sampler Algorithm converges to a stationary distribution with

$$p(x) = \lim_{k \to \infty} P\{X^{(k)} = x\} \ .$$

Let X have a strictly positive distribution

Notice that in the special case that $P_{i,j} > 0$ for all $i, j \in \Omega$, then the Markov chain is guaranteed to be both irreducible and aperiodic. This leads to the following useful corollary.

Theorem 3 *Limit Theorem 1 for Markov Chains Let X_n be a discrete-state discrete-time homogeneous Markov chain such that*

- *Ω is a finite set*
- *$P_{i,j} > 0$ for all $i, j \in \Omega$*

Then exists a unique stationary distribution

Chapter 1 Problems

1. Let $\{X_n\}_{n=1}^N$ be a 1D discrete valued MRF with neighborhood system
$$\partial n = \{n-1, n+1\} \cap \{1, \ldots, N\}$$
and strictly positive distribution. Prove that $\{X_n\}_{n=1}^N$ is also a Markov Chain.

2. Let X be a binary valued 2-D random field with $N \times N$ points. Assume that for $0 < i, j < N-1$
$$P\{X_{(0,j)} = 0\} = P\{X_{(i,0)} = 0\} = \frac{1}{2}$$
and for
$$\begin{aligned} & P\{X_{(i,j)} = x_{(i,j)} | X_r = x_r \ r < (i,j)\} \\ = \ & g(x_{(i,j)} | x_{(i-1,j)}, x_{(i,j-1)}) \\ = \ & \frac{1}{3}\delta(x_{(i,j)}, x_{(i-1,j)}) + \frac{1}{3}\delta(x_{(i,j)}, x_{(i,j-1)}) + \frac{1}{6} \end{aligned}$$

   a) Compute the complete density function for X.

   b) Show that X is a MRF. Give the cliques and and neighborhood for X.

3. Let $f : \mathbb{R}^N \to \mathbb{R}$ be a continuous convex function which is differentable every where and has a unique global minima. Let $x \in \{0,1\}^N$ be a binary valued vector with N components. We would like to compute
$$\hat{x} = \arg \min_{x \in \{0,1\}^N} f(x) \ .$$

   a) Is the minimum unique? Prove your answer or give a counter example.

   b) Does the Gauss-Seidel/Coordinate search method converge to a global minimum? Prove your answer or give a counter example.

Bibliography

[1] J. Besag. Spatial interaction and the statistical analysis of lattice systems. *Journal of the Royal Statistical Society B*, 36(2):192–236, 1974.

[2] S. Geman and D. Geman. Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, PAMI-6:721–741, November 1984.