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

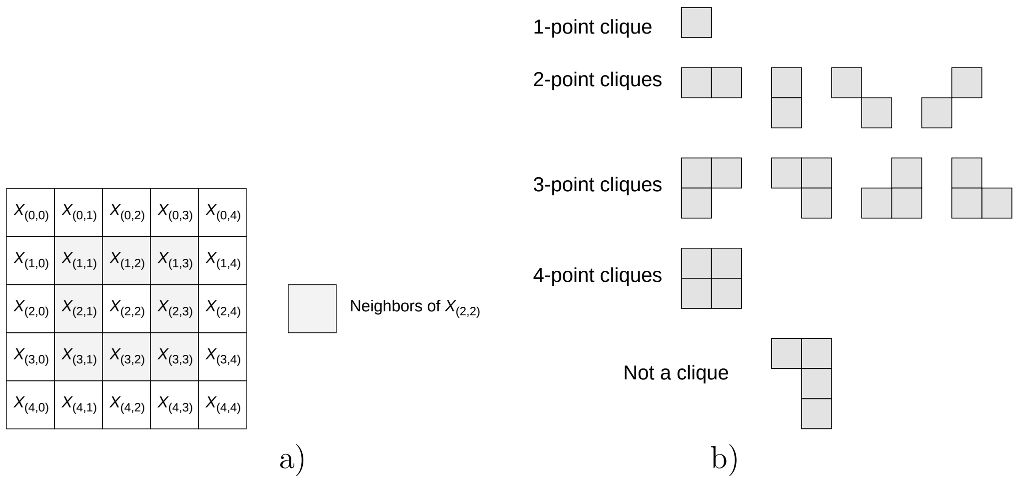

Figure 1.1: An eight point a) neighborhood system, and b) its associated cliques.

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 $s, r \in S$

$$r \in \partial s \Rightarrow s \in \partial r \quad \textit{and} \quad s \notin \partial s$$

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 $x \in \Omega$

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

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 Stationary Distributions of Markov Chains

In this section, we present the basic results of Markov chain theory that we will need to analyze simulation methods [3]. The fundamental concept is that a well behaved Markov chain will eventually reach a stable stationary distribution after the transient behavior dies away. The objective will be to define the technical conditions that ensure that this happens.

Let $\{X_n\}_{n=0}^{\infty}$ be a discrete valued and discrete state homogeneous Markov chain taking values in the set Ω. Define the notation

$$\begin{aligned} \pi_j^{(n)} &\triangleq P\{X_n = j\} \\ P_{i,j} &\triangleq P\{X_n = j | X_{n-1} = i\} \ , \end{aligned}$$

and define $\pi^{(n)}$ to be the corresponding $1 \times |\Omega|$ row vector, and P to be the corresponding $|\Omega| \times |\Omega|$ matrix. A fundamental property of Markov chains is that the marginal probability density for time $n+1$ can be expressed as

$$\pi_j^{(n+1)} = \sum_{i \in \Omega} \pi_i^{(n)} P_{i,j} \ . \tag{1.1}$$

In matrix notation, this is equivalent to

$$\pi^{(n+1)} = \pi^{(n)} P \ . \tag{1.2}$$

Repeated application of (1.2) results in the equation

$$\pi^{(n+m)} = \pi^{(n)} P^m \ , \tag{1.3}$$

or equivalently

$$\pi_j^{(n+m)} = \sum_{i \in \Omega} \pi_i^{(n)} P_{i,j}^m \ . \tag{1.4}$$

where $P_{i,j}^m$ is defined by the recursion

$$P_{i,j}^{m+1} = \sum_{k \in \Omega} P_{i,k}^m P_{k,j} \ .$$

More generally, the Chapman-Kolmogorov relation states that

$$P_{i,j}^{m+k} = \sum_{k \in \Omega} P_{i,k}^{m} P_{k,j}^{k} \ .$$

We next define a number of properties for homogeneous Markov chains that we will need.

Definition 5 *Communicating states*
The states $i, j \in \Omega$ of a Markov chain are said to communicate *if there exists integers $n > 0$ and $m > 0$ such that $P_{i,j}^n > 0$ and $P_{j,i}^m > 0$.*

Intuitively, two states communicate if it is possible to transition between the two states. It is easily shown that communication is an equivalence property, so it partitions the set of states into disjoint sets that all communicate with each other. This leads to a natural definition for irreducible Markov chains.

Definition 6 *Irreducible Markov Chain*
A discrete time discrete state homogeneous Markov chain is said to be irreducible if for all $i, j \in \Omega$ i and j communicate.

So a Markov chain is irreducible if it is possible to change from any initial state to any other state in finite time.

In some cases, a state of a Markov chain my repeat periodically. This type of periodic repetition can last indefinitely.

Definition 7 *Periodic state*
We denote the period of a state $i \in \Omega$ by the value $d(i)$ where $d(i)$ is the largest integer so that $P_{i,i}^n = 0$ whenever n is not divisible by $d(i)$. If $d(i) > 1$, then we say that the state i is periodic.

It can be shown that states of a Markov chain that communicate must have the same period. Therefore, all the states of an irreducible Markov chain must have the same period. We say that the an irreducible Markov chain is aperiodic if all the states have period 1.

Using these definitions, we may now state a theorem which gives basic conditions for convergence of the distribution of the Markov chain.

Theorem 2 *Limit Theorem for Markov Chains*
Let $X_n \in \Omega$ be a discrete-state discrete-time homogeneous Markov chain with transition probabilities $P_{i,j}$ and the following additional properties

- *Ω is a finite set*
- *The Markov chain is irreducible*
- *The Markov chain is aperiodic*

There exists a unique stationary distribution π, which for all states i is given by

$$\pi_j = \lim_{n\rightarrow\infty} P^n_{i,j} > 0 \tag{1.5}$$

and which is the unique solution to the following set of equations.

$$1 = \sum_{i\in\Omega} \pi_i \tag{1.6}$$

$$\pi_j = \sum_{i\in\Omega} \pi_i P_{i,j} \ . \tag{1.7}$$

The relations of (1.7) are sometimes called the full balance equations (FBE). Any probability density which solves the FBE is guaranteed to be the stationary distribution of the Markov chain. Furthermore, in the limit as $n \rightarrow \infty$, the Markov chain is guaranteed to converged to this stationary distribution independently of the initial state. Markov chains that have this property of (1.5) are said to be *ergodic*. It can be shown that for ergodic Markov chains, expectations of state variables can be replace by time averages, which will be very useful in later sections.

Theorem 2 gives relatively simple conditions to establish that a Markov chain has a stationary distribution. However, while it may be known that a stationary distribution exists, it may be very difficult to compute the solution of the FBEs to determine the precise form of that distribution. It is often useful to use the property of reversibility as a method to solve this problem. First we must show that the time reverse of a Markov chain is itself a Markov chain.

Proposition 1 *Time Reverse of Markov Chains*
Let $\{X_n\}_{n=-\infty}^{\infty}$ be a Markov Chain. Then the time reversed process $Y_n = X_{-n}$ is also a Markov chain.

Since the time reversal of a Markov chain is also a Markov chain, it must also have a transition distribution which we denote at $Q_{i,j}$. Therefore, we know that

$$P\{X_n = i, X_{n+1} = j\} = \pi_i^n P_{i,j} = \pi_j^{n+1} Q_{j,i} \ .$$

If the Markov chain has a stationary distribution, then $\pi = \pi^n = \pi^{n+1}$, and we have that

$$\pi_i P_{i,j} = \pi_j Q_{j,i} \ .$$

Furthermore, if the Markov chain is reversible, then we know that $P_{i,j} = Q_{i,j}$. This yields the so-called detailed balance equations (DBE).

$$\pi_i P_{i,j} = \pi_j P_{j,i} \tag{1.8}$$

The DBE specify that the rate of transitions from state i to state j equals the rate of transitions from state j to i. This is always the case when a Markov chain is reversible.

Definition 8 *Reversible Markov Chain*
A homogeneous Markov chain with transition probabilities $P_{i,j}$ is said to be reversible if there exists a stationary distribution which solves the detailed balance equations of (1.8).

Notice that if one finds a solution to the DBE, then this solution must also be a solution to the FBE, and is therefore the stationary distribution of an ergodic Markov chain. To see this

$$\begin{aligned}
\sum_{i\in\Omega} \pi_i P_{i,j} &= \sum_{i\in\Omega} \pi_j P_{j,i} \\
&= \pi_j \sum_{i\in\Omega} P_{j,i} \\
&= \pi_j
\end{aligned}$$

Finally, it is useful to study the convergence behavior of Markov chains. Let us consider the case when the Markov chains state Ω is finite. In this case, we may use a matrix representation for P. We know that any matrix P may be diagonalized using eigen decomposition and expressed in the form

$$P = E^{-1} \Lambda E$$

where the rows of E are the left hand eigenvectors of P, and Λ is a diagonal matrix of eigenvalues. Using this decomposition, we can see that

$$\begin{aligned}
P^m &= P^{m-2} \, E^{-1} \Lambda E^{-1} \, E \Lambda E \\
&= P^{m-2} \, E^{-1} \Lambda^2 E \\
&= E^{-1} \Lambda^m E
\end{aligned}$$

So the distribution at time n is given by

$$\pi^{(n)} = \pi^{(0)} E^{-1} \Lambda^m E$$

When P corresponds to a irreducible and aperiodic Markov chain, then it must have a stationary distribution. In this case, exactly one of the eigenvalues is 1, and the remaining eigenvalues have magnitude strictly less then 1. In this way, we can see that the distribution of $\pi^{(n)}$ converges geometrically to its stationary distribution π.

1.4.2 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

$$p(x) = \frac{1}{z} \exp\{-u(x)\} \tag{1.9}$$

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, (1.9) 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

$$p\left(x_s | x_i \, i \neq s\right) = \frac{\exp\{-u\left(x_s | x_i \, i \neq s\right)\}}{\sum_{x'_s \in \Omega} \exp\{-u\left(x'_s | x_i \, i \neq s\right)\}} \tag{1.10}$$

We can generate samples from the distribution of (1.9) 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 (1.9).

Theorem 3 *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 4 *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

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.

[3] S. M. Ross. *Stochastic Processes*. John Wiley & Sons, New York, 1983.