Statistical Techniques in Robotics (16-831, F12) Lecture#07 (Wednesday September 19)

Gibbs Field & Markov Random Field

Lecturer: Drew Bagnell *Scribe:Minghao Ruan* [1]

1 SLAM (continued from last lecture)

Figure 1 depicts a typical SLAM system representd as a Bayes Network. x_i represents the robot state at time i; z_i is the corresponding measurement at instance i; and l_is are some landmarks.

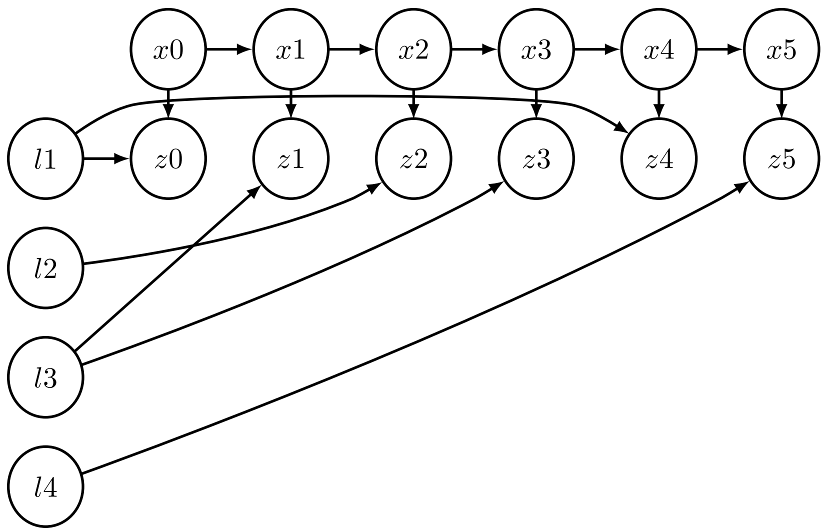


Figure 1: Represent the SLAM problem as a Bayes Network.

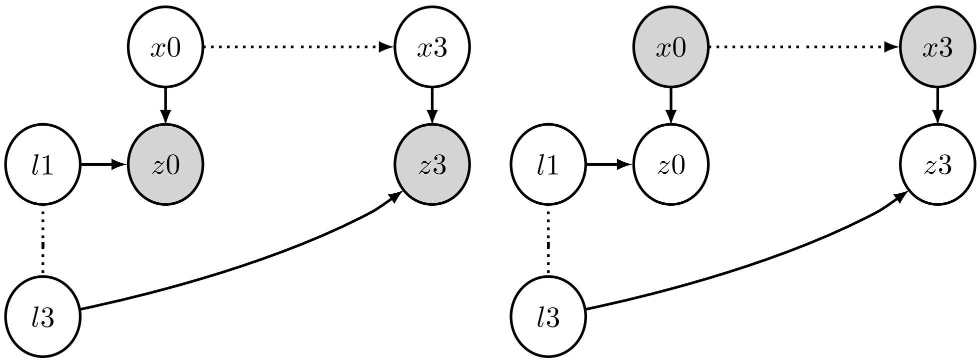
It should be noted, however, this diagram is quite strange in the following sense:

- A real robot may not be absolutely certain about which landmark it observes at any instance. (We basically assume the landmarks are "broadcasting" their names)
- Not seeing a certain landmark could also provide useful information.
- We also assume one landmark per observation. This is because we can break up multiple landmarks into two observations with no motion between the two instances.

Despite these facts, we still prefer four 2D filters(for each landmark) to one 8D filter which is much more expensive to maintain. Therefore we are interested in the question: given all the observations, are the landmarks conditionally independent of each other?

The answer is No. For example $l_1 \perp l_3 \,|\, z_0 \ldots z_n$ is not true becuase observing z_0 unblockes the path between the two converging arrows (as shown in Fig 3(a)).

[1] Some content adapted from previous scribes: Bradford Neuman, Byron Boots



(a) Observation unblockes the path (b) Knowning the states blocks the path

If this is a mapping problem, however, where the robot states x_i are known, the landmarks are indeed conditionally independent because knowledge about x_i blocks the flow of information as shown in Fig 3(b).

2 Undirected Graph

2.1 Gibbs Field

Like Bayes Network, Gibbs Field also represents the relationship of a set of random variables, but with undirected connections. A simple Gibbs Field is shown in Figure 2;

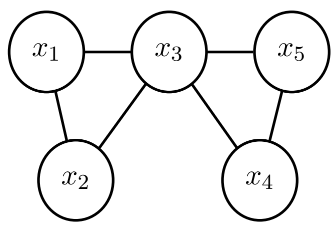


Figure 2: A simple Gibbs Field with 5 random variables.

The first important structure in the graph is a *clique*, which is a subset of the vertices in which any two nodes are connected by an edge. The simplest clique is an individual node, for example x_1, x_3 and x_6. Often we are interested in the maximum cliques which are not proper subsets of any other cliques. In Fig 2, the two maximum cliques are $c_1 = \{x_1,\ x_2,\ x_3\}$ and $c_2 = \{x_3,\ x_4,\ x_5\}$.

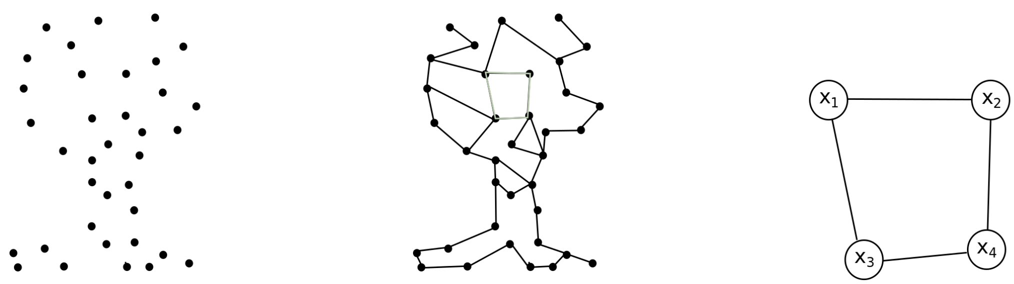
For any Gibbs Field, we can define the joint probability as follows:

$$p(\vec{x}) = \frac{1}{Z} \prod_{c_i \in C} \Phi_{c_i}(\overrightarrow{x_{c_i}}) \tag{1}$$

where Φ_{c_i} is the ith clique potential function which only depends on its member nodes in c_i. Each potential function has to be strictly positive, but unlike probability distribution, they need not sum to 1. For this reason, the final product has to be normalized by summing over all possible outcomes to make it a valid probability distribution. Since we can easily write a clique potential function as a product of all potential functions of its sub-cliques, it is convenient to define the joint probability over only the maximum cliques.

2.1.1 A Concrete Example

Suppose we have a point cloud from a laser scan and we'd like to label each point as either $Vegetation(= 1)$ or $Ground(= 0)$. Suppose the raw cloud has been converted to some mesh grid as in Fig 3(b), and we zoom in to a small patch as in Fig 3(c).



(a) Laser scan points (b) Connect points in a mesh (c) One clique of the graph

We can then define a joint probability destribution for this configuration in Fig 3(c):

$$p(\vec{x}) = \frac{1}{Z}\Phi_{12}(x_1,\ x_2)\Phi_{13}(x_1,\ x_3)\Phi_{23}(x_2,\ x_3)\Phi_{14}(x_1,\ x_4) \tag{2}$$

where

$$\Phi_{ij}(x_i,\ x_j) = \begin{cases} 1 & \text{if } x_i = x_j \\ 1/2 & \text{otherwise} \end{cases}$$

Intuitively, the maximum probability can be achieved when all vertices have the same label. In real problems, Eq. 2 also takes into account of prior information. For example, a modified version of Eq. 2 may look like:

$$p(\vec{x}) = \frac{1}{Z}\Phi_{12}(x_1,\ x_2)\Phi_{13}(x_1,\ x_3)\Phi_{23}(x_2,\ x_3)\Phi_{14}(x_1,\ x_4)\Phi_1(x_1)\Phi_2(x_2)\Phi_3(x_3)\Phi_4(x_4) \tag{3}$$

where

$$\Phi_i(x_i) = \begin{cases} 1 & \text{Vegetation} \\ 0 & \text{Ground} \end{cases}$$

2.1.2 Clique Potential as Energy Function

As the name suggests, a *potential function* can be related to the energy of a state, which is an indicator of the likelihood of a particular configuation associated with the clique. Higher energy means "unstable" or "unlikely", whereas lower energy means "stable" or "more likely". The energy functions usually take the following form:

$$p(\vec{x}) = \frac{1}{Z}\, e^{-\sum f_c(\vec{x})} \tag{4}$$

3 Markov Random Field(MRF)

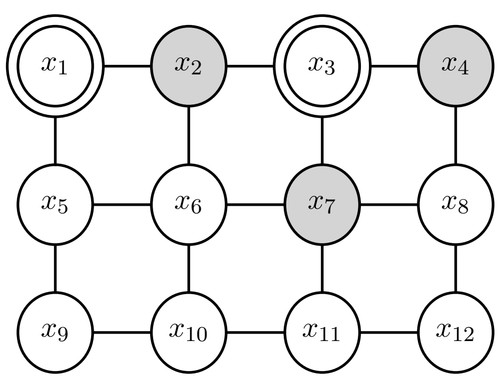
Sometimes we are only interested in finding out given a graph, whether two sets of random variables are conditionally independent if a third set of variables are observed. Becuase of its simple rule of finding conditional independence, Markov Random Field is most commonly used to express such relationship between any given random variables. Much like the structure of Bayes Network makes it suitable for logical reasoning, the graphical representation of the MRF makes it particular useful for inference in physical space.

3.1 Rules of Conditional Independence

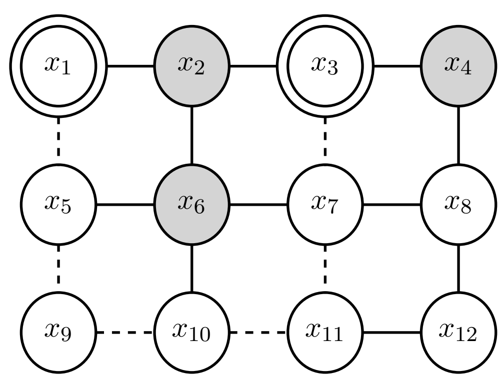
The simple answer is: two variables are conditionally independent if all paths between two variables are blocked (by observed nodes).

The formal rules are the following:

- **Local rule:** A variable is conditionally independent of *all* other variables given all its neigbors. Fig. 3(d) shows the *Markov blanket* as all paths into and out of x_3 are "blanketed".
- **Global rule:** If all paths between two sets of variables A and B pass through another set of nodes S. Then any two subsets of A and B are conditionally independent is S is observed.



(d) $x_1 \perp x_3 \mid x_2,\ x_4,\ x_7$



(e) $x_1 \not\perp x_3 \mid x_2,\ x_4,\ x_6$

Figure 3: Markov Random Field example.

3.2 Hammersley-Clifford theorem

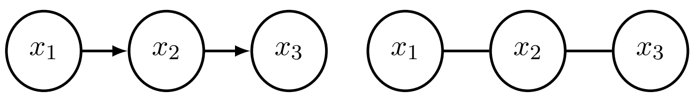
Both Gibbs Field and MRF can be derived from a given graph but they each reveals a different aspect of the problem. Like Bayes Network, Gibbs Field also defines a joint probability, whereas Markov Random Field is mainly for listing conditional independence between variables.

Both theories developed independently for a while until Hammersley-Clifford theorem states that under certain conditions, the Gibbs Field and Markov Random Field are equivalent. Though the proof is not a trivial task, the conclusion is quite straightforward:

- Gibbs field on graph G has all the conditional independence relationships of an MRF on G.
- Every MRF can be written as a Gibbs Field on the same graph provided $p(\vec{x}) > 0$ everywhere.

4 Converting Bayes Net to Gibbs Field/MRF

Let's first consider the toy problem in Fig. 4, where the the conversion is done by directly removing the arrows.



(a) Bayes Net representation of a set of RVs

(b) Gibbs Field/MRF of the same set of RVs

Figure 4: Converting a simple Bayes Network to a Gibbs Field/MRF

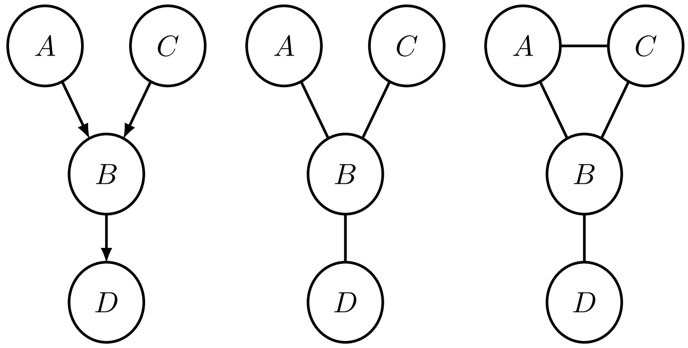
By definition, the joint probability for the Bayes Network is $p(\vec{x}) = p(x_1)\, p(x_2|x_1)\, p(x_3|x_2)$. Similarly, the joint probability for the Gibbs Field/MRF can be written as product of potential functions over cliques: $p(\vec{x}) = \frac{1}{Z}\, \Phi_{12}(x_1, x_2)\, \Phi_{23}(x_2, x_3)\, \Phi_1(x_1)\, \Phi_2(x_2)\, \Phi_3(x_3)$.

To convert the Bayes Net to Gibbs Field/MRF, we only have to come up with the right potential functions that are consistent with the Bayes Network description. In this case, the following assignment would satisfy the requirement:

$$\begin{cases} \Phi_1(x_1) \;=\; p(x_1) \\ \Phi_{12}(x_1,\, x_2) \;=\; p(x_1 \,|\, x_2) \\ \Phi_{23}(x_2,\, x_3) \;=\; p(x_2 \,|\, x_3) \\ \Phi_2(x_2) \;=\; \Phi_3(x_3) \;=\; \Phi_{13}(x_1,\, x_3) \;=\; 1 \end{cases}$$

What if we have a more complicated graph? For instance, a converging triple as in Fig. 5(a)

The intuitive answer would also be simply removing the arrows, but this is insufficient. Let's examine what happens after removing the arrows in Fig. 5(a). Observing B blocks the path in Gibbs Field/MRF in Fig. 5(b), but the same action would have unblocked the path between A and C in Bayes net in Fig. 5(a), therefore simply removing the arrows results in a contradiction.



(a) Bayes' Net (b) Incorrect! (c) Correct, with moralizing

Figure 5: Moralizing process.

The solution is to "Moralize" the parents whenever there is a converging child. The downside is the loss of structure in the graph (recall from last lecture, fewer arrows means more structure). In this case, A and C are no longer independent even if B is not observed.