

1.2.4

Markov-Gibbs Equivalence

An MRF is characterized by its local property (the Markovianity) whereas a GRF is characterized by its global property (the Gibbs distribution). The Hammersley-Clifford theorem [Hammersley and Clifford_1971] establishes the equivalence of these two types of properties. The theorem states that *F is an MRF on \mathcal{S} with respect to \mathcal{N} if and only if F is a GRF on \mathcal{S} with respect to \mathcal{N}* . Many proofs of the theorem exist, *e.g.* in [Besag_1974 ; Moussouris_1974 ; Kindermann and Snell_1980].

A proof that a GRF is an MRF is given as follows. Let $P(\mathbf{f})$ be a Gibbs distribution on \mathcal{S} with respect to the neighborhood system \mathcal{N} . Consider the conditional probability

$$P(f_i \mid f_{\mathcal{S}-\{i\}}) = \frac{P(f_i, f_{\mathcal{S}-\{i\}})}{P(f_{\mathcal{S}-\{i\}})} = \frac{P(\mathbf{f})}{\sum_{\mathbf{f}'_i \in \mathcal{L}} P(\mathbf{f}')} \tag{1.30}$$

where $\mathbf{f}' = \{f_1, \dots, f_{i-1}, f'_i, \dots, f_m\}$ is any configuration which agrees with \mathbf{f} at all sites except possibly \mathbf{i} . Writing $P(\mathbf{f}) = Z^{-1} \times \mathrm{e}^{-\sum_{\mathbf{c} \in \mathcal{C}} V_{\mathbf{c}}(\mathbf{f})}$ out gives (This also provides a formula for calculating the conditional probability $P(f_i \mid f_{\mathcal{N}_i}) = P(f_i \mid f_{\mathcal{S}-\{i\}})$ from potential functions.)

$$P(f_i \mid f_{\mathcal{S}-\{i\}}) = \frac{\mathrm{e}^{-\sum_{\mathbf{c} \in \mathcal{C}} V_{\mathbf{c}}(\mathbf{f})}}{\sum_{\mathbf{f}'_i} \mathrm{e}^{-\sum_{\mathbf{c} \in \mathcal{C}} V_{\mathbf{c}}(\mathbf{f}')}} \tag{1.31}$$

Divide \mathcal{C} into two set \mathcal{A} and \mathcal{B} with \mathcal{A} consisting of cliques containing \mathbf{i} and \mathcal{B} cliques not containing \mathbf{i} . Then the above can be written as

$$P(f_i \mid f_{\mathcal{S}-\{i\}}) = \frac{\left[\mathrm{e}^{-\sum_{\mathbf{c} \in \mathcal{A}} V_{\mathbf{c}}(\mathbf{f})} \right] \left[\mathrm{e}^{-\sum_{\mathbf{c} \in \mathcal{B}} V_{\mathbf{c}}(\mathbf{f})} \right]}{\sum_{\mathbf{f}'_i} \left\{ \left[\mathrm{e}^{-\sum_{\mathbf{c} \in \mathcal{A}} V_{\mathbf{c}}(\mathbf{f}')} \right] \left[\mathrm{e}^{-\sum_{\mathbf{c} \in \mathcal{B}} V_{\mathbf{c}}(\mathbf{f}')} \right] \right\}} \tag{1.32}$$

Because $V_{\mathbf{c}}(\mathbf{f}) = V_{\mathbf{c}}(\mathbf{f}')$ for any clique \mathbf{c} that does not contain \mathbf{i} , $\mathrm{e}^{-\sum_{\mathbf{c} \in \mathcal{B}} V_{\mathbf{c}}(\mathbf{f})}$ cancels from both the numerator and denominator. Therefore, this probability depends only on the potentials of the cliques containing \mathbf{i} ,

$$P(f_i \mid f_{\mathcal{S}-\{i\}}) = \frac{\mathrm{e}^{-\sum_{\mathbf{c} \in \mathcal{A}} V_{\mathbf{c}}(\mathbf{f})}}{\sum_{\mathbf{f}'_i} \mathrm{e}^{-\sum_{\mathbf{c} \in \mathcal{A}} V_{\mathbf{c}}(\mathbf{f}')}} \tag{1.33}$$

that is, it depends on labels at \mathbf{i} 's neighbors. This proves that a Gibbs random field is a Markov random field. The proof that an MRF is a GRF is much more involved; a result to be described in the next subsection, which is about the unique GRF representation [Griffeath_1976], provides such a proof.

The practical value of the theorem is that it provides a simple way of specifying the joint probability. One can specify the joint probability $P(\mathbf{F} = \mathbf{f})$ by specifying the clique potential functions $V_{\mathbf{c}}(\mathbf{f})$ and chooses appropriate potential functions for desired system behavior. In this way, he encodes the *a priori* knowledge or preference about interactions between labels.

How to choose the forms and parameters of the potential functions for a proper encoding of constraints is a major topic in MRF modeling. The forms of the potential functions determine the form of the Gibbs distribution. When all the parameters involved in the potential functions are specified, the Gibbs distribution is completely defined. Defining the functional forms is the theme in Chapters 2--5 while estimating parameters is the subject in Chapters 6 and 7.

To calculate the joint probability of an MRF, which is a Gibbs distribution, it is necessary to evaluate the partition function (1.25). Because it is the sum over a combinatorial number of configurations in \mathbb{F} , the computation is usually intractable. The explicit evaluation can be avoided in maximum-probability based MRF vision models when $U(\mathbf{f})$ contains no unknown parameters, as we will see subsequently. However, this is not true when the parameter estimation is also a part of the problem. In the latter case, the energy function $U(\mathbf{f}) = U(\mathbf{f} \mid \boldsymbol{\theta})$ is also a function of parameters $\boldsymbol{\theta}$ and so is the partition function $Z = Z(\boldsymbol{\theta})$. The evaluation of $Z(\boldsymbol{\theta})$ is required. To circumvent the formidable difficulty therein, the joint probability is often approximated in practice. Several approximate formulae will be introduced in Chapter 6 where the problem of MRF parameter estimation is the subject.