On Functions of Markov Random Fields

Bernhard C. Geiger
Know-Center GmbH, 8010, Graz, Austria
Email: geiger@ieee.org

Ali Al-Bashabsheh
Beijing Advanced Inst. on Big Data and Brain
Computing, Beijing, China (entropyali@gmail.com)

***Abstract*—We derive two sufficient conditions for a function of a Markov random field (MRF) on a given graph to be a MRF on the same graph. The first condition is information-theoretic and parallels a recent information-theoretic characterization of lumpability of Markov chains. The second condition, which is easier to check, is based on the potential functions of the corresponding Gibbs field. We illustrate our sufficient conditions at the hand of several examples and discuss implications for practical applications of MRFs. As a side result, we give a partial characterization of functions of MRFs that are information preserving.**

***Index Terms*—Markov random field, Gibbs field, lumpability, hidden Markov random field**

I. Introduction

Since the late 1950s, researchers have actively investigated properties of functions of Markov chains. In particular, considerable effort has been devoted to obtain sufficient and necessary conditions for *lumpability*, the rare scenario in which a function of a Markov chain has the Markov property [1]–[3].

In this work, we extend the concept of lumpability and its investigation to Markov random fields (MRFs). Specifically, given a MRF $X := (X_1, \dots, X_N)$ on a graph $\mathcal{G}$, we determine conditions for a set of functions $\{g_1, \dots, g_N\}$ such that the transformation $Y := (g_1(X_1), \dots, g_N(X_N))$ is a MRF on $\mathcal{G}$. In other words, the problem we investigate asks the question under which functions an independence structure (i.e., a collections of independence statements) remains valid.

Aside from being an interesting problem in its own, it is also practically motivated from an inference perspective. For instance, multidimensional data X is often modeled as a hidden MRF, i.e., the data X is inaccessible and may only be inferred from some observed random variable $Z := (Z_1, \dots, Z_N)$, where each Z_i is conditionally independent of X given X_i. In some scenarios, however, not X is of interest but its transformation Y. E.g., in image processing, the graph $\mathcal{G}$ is often taken as a lattice with a distance-based neighborhood and the random variables X and Z are, respectively, taken as the true and observed pixel values. In this case, Y may denote the pixel values in a subsampled version of the image, labels of image segments, or it may denote quantized pixel values for the sake of identifying regions with similar intensities, etc. Transforming X to Y potentially creates additional or breaks existing dependencies, i.e., the graph $\mathcal{G}_Y$ w.r.t. which Y is a MRF is generally different from $\mathcal{G}$. Rather than inferring X from the observed Z and subsequently computing Y via the known transformations, in this work, we are interested in scenarios where Y is directly inferred from Z. This is computationally tractable if (Y, Z) turns out to be a hidden MRF itself. Among other things, this requires determining the graph $\mathcal{G}_Y$ w.r.t. which Y is a MRF.

The remainder of this paper can be summarized as follows. Section II introduces notation and basic definitions, and Section III formulates the problem and provides some examples. Section IV places the current work in context with previous results on stochastic transformations of MRFs [4, Sec. IV] and subfields of MRFs [4]–[6]. Section V gives two sufficient conditions for Y to be a MRF on the same graph as X, i.e., for $\mathcal{G}_Y = \mathcal{G}$. The first condition is based on the characterization of MRFs via clique potentials, while the second is information-theoretic and resembles the information-theoretic characterization of Markov chain lumpability [3, Th. 2]. As a side result, Section VI presents necessary and sufficient conditions for the transformation Y to have the same information content as X. For the sake of readability, and due to space limitations, all proofs are in [7].

II. Notation and Preliminaries

Let $\mathcal{G} := (\mathcal{V}, E)$ be an undirected graph with vertices $\mathcal{V} := \{1, \dots, N\}$ and edges $E \subseteq [\mathcal{V}]^2$, where $[A]^2$ is the set of two-element subsets of A. We call $\mathcal{G}$ complete if $E = [\mathcal{V}]^2$, chordal if every induced cycle of $\mathcal{G}$ has length three, a tree if $\mathcal{G}$ is connected and acyclic, and a path if there is a permutation $v_1, \dots, v_N$ of the vertices such that $E = \{\{v_i, v_{i+1}\},\ i = 1, \dots, N-1\}$. If $\{i, j\} \in E$, then the vertices i and j are neighbors, and we use $\mathcal{N}_i$ to denote the neighbors of i, i.e.,

$$\mathcal{N}_i := \{j \in \mathcal{V} \setminus \{i\} \colon \{i, j\} \in E\}. \tag{1}$$

A set $C \subseteq \mathcal{V}$ is called a clique if it is a singleton or if $[C]^2 \subseteq E$. We use $\mathcal{C}$ to denote the set of cliques of $\mathcal{G}$.

We denote random variables (RVs) by upper case letters, e.g., X, alphabets by calligraphic letters, e.g., $\mathcal{X}$, and realizations by lower case letters, e.g., x. We assume that all our RVs are defined on a common probability space $(\Omega, \mathcal{T}, \mathbb{P})$. Specifically, let X_i be a discrete RV with alphabet $\mathcal{X}_i$ that is associated with vertex $i \in \mathcal{V}$. For a set $A \subseteq \mathcal{V}$, we write $X_A := (X_i,\ i \in A)$ and $\mathcal{X}_A := \prod_{i \in A} \mathcal{X}_i$. We furthermore use the abbreviations $X := X_{\mathcal{V}}$ and $X_{\not{i}} := X_{\mathcal{V} \setminus \{i\}}$, and similarly for the alphabets of these RVs. The RV X_A is characterized by its probability mass function (PMF)

$$p_{X_A}(x_A) := \mathbb{P}(\{\omega \in \Omega \colon X_A(\omega) = x_A\}), \forall x_A \in \mathcal{X}_A. \tag{2}$$

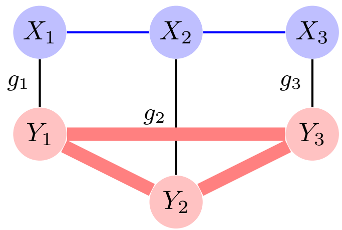

Fig. 1. The problem of lumpability. The blue vertices and edges correspond to the original $(\mathcal{G}, p_X)$-MRF X, the black labeled edges correspond to functions through which the RVs in the MRF are observed, thus defining $Y := (g_1(X_1), g_2(X_2), g_3(X_3))$. In general, the minimal graph w.r.t. which Y is a MRF is complete (see red vertices and edges): By observing a Markov path X_1—X_2—X_3 through a non-injective function, the Markov property is lost in general. The lumpability problem seeks conditions on p_X and $\{g_i\}$ such that the minimal graph for Y is equivalent to the original graph $\mathcal{G}$ or a subgraph of $\mathcal{G}$.

Definition 1. Let $\mathcal{G} := (\mathcal{V}, E)$ be a graph and $X := (X_i, i \in \mathcal{V})$ be a RV with PMF p_X, then X is a *Markov random field* (MRF) on $\mathcal{G}$, abbreviated X is a $(\mathcal{G}, p_X)$-MRF, if

$$\forall i \in \mathcal{V}: \quad p_{X_i|X_{\not i}} = p_{X_i|X_{\mathcal{N}_i}}, \tag{3}$$

i.e., if the distribution of X_i depends on the remaining RVs only via the RVs neighboring i. If p_X is unspecified, but known to belong to a family of distributions for which (3) holds for every member, then we say that X is a $\mathcal{G}$-MRF.

For any $A, B \subseteq \mathcal{V}$, the entropy of X_A is defined as

$$H(X_A) := -\sum_{x_A \in \mathcal{X}_A} p_{X_A}(x_A) \log p_{X_A}(x_A) \tag{4}$$

and the conditional entropy of X_A given X_B as $H(X_A|X_B) := H(X_{A\cup B}) - H(X_B)$. With this notation, the lemma below follows immediately from Definition 1.

Lemma 1. *X is a $\mathcal{G}$-MRF if and only if (iff), for every $i \in \mathcal{V}$, $H(X_i|X_{\not i}) = H(X_i|X_{\mathcal{N}_i})$.*

Note that if X is a $\mathcal{G}$-MRF, then it is a MRF on every graph on $\mathcal{V}$ whose edge set contains E. Trivially, every X is a MRF on the complete graph. Of particular interest is thus the *minimal* graph w.r.t. which X is a MRF. We will assume throughout this paper that the graph $\mathcal{G}$ w.r.t. which X is a MRF is minimal.

III. Problem Statement and Motivating Examples

In this work, we consider functions of MRFs. Specifically, let $\{g_i, i \in \mathcal{V}\}$ (subsequently abbreviated as $\{g_i\}$ to simplify notation) be a set of functions $g_i\colon \mathcal{X}_i \to \mathcal{Y}_i$ indexed by the vertices $i \in \mathcal{V}$, and let $Y_i := g_i(X_i)$. For $A \subseteq \mathcal{V}$, we define the function $g_A\colon \mathcal{X}_A \to \mathcal{Y}_A$ as the functions g_i, $i \in A$, applied to X_A coordinate-wise, i.e., $g_A(X_A) := (g_i(X_i), i \in A) = Y_A$, and, as before, use the abbreviation $g(X) := g_{\mathcal{V}}(X) = Y$. We call a set of functions $\{g_i\}$ *non-trivial* if at least one function g_i is non-injective. Given a $(\mathcal{G}, p_X)$-MRF X and a set of functions $\{g_i\}$, we call the tuple $(\mathcal{G}, p_X, \{g_i\})$ the *lumping* of X. We will focus on the following two problems:

Problem 1 (Lumpability)**.** Determine conditions on the lumping $(\mathcal{G}, p_X, \{g_i\})$ so that Y is a MRF w.r.t. $\mathcal{G}$, where in this case we say $(\mathcal{G}, p_X, \{g_i\})$ is *lumpable*, see Fig. 1. By the remark below Lemma 1, $(\mathcal{G}, p_X, \{g_i\})$ is lumpable whenever it does not introduce new edges, i.e., whenever Y is a $(\mathcal{G}', p_Y)$-MRF with $\mathcal{G}' := (\mathcal{V}, E')$ and $E' \subseteq E$

Problem 2 (Information Preservation)**.** Determine conditions on the lumping $(\mathcal{G}, p_X, \{g_i\})$ so that $H(Y) = H(X)$, where in this case we say $(\mathcal{G}, p_X, \{g_i\})$ is *information preserving*.

Throughout this work we assume the set of functions $\{g_i\}$ is non-trivial. Otherwise, if all the functions g_i are injective, then X and Y would have the same distribution since $\{g_i\}$ is simply a relabeling of the distribution's domain, and so the lumping would be trivially lumpable and information preserving. We also assume that $\mathcal{G}$ is connected, which is w.l.o.g. since the RVs of different components of the graph are independent, and this independence is retained for any set of functions $\{g_i\}$.

To get some intuition on why a function of a MRF may not be a MRF on the same graph, note that X_i and $X_{\not i}$ are conditionally independent given $X_{\mathcal{N}_i}$ only when $X_{\mathcal{N}_i}$ contains all the information about X_i that is available in $X_{\not i}$. Taking a function of $X_{\mathcal{N}_i}$ may reduce this information to a point where $Y_{\mathcal{N}_i}$ no longer contains all the information about Y_i that is available in $Y_{\not i}$, which effectively introduces edges in the minimal graph for Y that have not been present in $\mathcal{G}$. This parallels the fact that a function of a Markov chain rarely results in a Markov chain [1, Th. 31]. (A Markov chain is a $\mathcal{G}$-MRF where $\mathcal{G}$ is the infinite path graph, i.e., with the natural numbers $\mathbb{N}$ as the set of vertices and $\{\{i, i+1\} : i \in \mathbb{N}\}$ as the set of edges.) Regarding information preservation, a lumping is information preserving iff $\{g_i\}$ maps the support of p_X injectively. Thus, both lumpability and information preservation appear to be the exception rather than the rule. The following examples demonstrate different lumpability and information preservation scenarios and give some intuition on the corresponding lumpings $(\mathcal{G}, p_X, \{g_i\})$.

Example 1 (Neither Information Preserving nor Lumpable)**.** Let X_1—X_2—X_3 be a Markov path, i.e., a $\mathcal{G}$-MRF on the path graph $\mathcal{G} := (\{1,2,3\}, \{\{1,2\},\{2,3\}\})$, where each RV X_i takes values from $\{0,1,2\}$. Suppose that $p_{X_1|X_2}(1|0) = p_{X_3|X_2}(1|2) = 0$, $p_{X_1|X_2}(1|2) = p_{X_3|X_2}(1|0) = p > 0$, and $p_{X_2}(0) = p_{X_2}(2) \in (0, 0.5)$. For all other configurations, assume $p_{X_1|X_2}$ and $p_{X_3|X_2}$ are positive. Let $g_i(x_i) = \mathrm{mod}(x_i, 2)$ for every i, then one can verify that $p_{Y_1|Y_2}(1|0) = p/2 = p_{Y_3|Y_2}(1|0)$, while $p_{Y_1,Y_3|Y_2}(1,1|0) = 0$. Thus, Y_1 and Y_3 are not conditionally independent given Y_2, and so the minimal graph for Y contains the new edge $\{1,3\}$, i.e., the lumping $(\mathcal{G}, p_X, \{g_i\})$ is not lumpable. (In this example the minimal graph for Y is the complete graph, see Fig. 1.) Furthermore, since, e.g., both configurations $x := (0,0,0)$ and $x' := (0,0,2)$ have positive probabilities, but are mapped to the same $y := (0,0,0)$, the lumping is not information preserving.

Example 2 (Information Preserving but not Lumpable)**.** Let $X_1 := X_2 + Z_1$ and $X_3 := X_2 + Z_3$, where $Z_1 \in \{0,1\}$, $X_2 \in$

$\{-1,1\}$, and $Z_3 \in \{-1,0\}$ are mutually independent RVs. It follows that X_1—X_2—X_3 is a Markov path as in the previous example with edges $E = \{\{1,2\},\{2,3\}\}$. Assume g_1 and g_3 are the identity functions and $g_2 \equiv 0$. Since Y_2 is constant, Y_1 and Y_3 are conditionally independent given Y_2 iff Y_1 and Y_3 are independent, which is not true due to the coupling through X_2. (Assuming p_{X_2} is strictly positive.) Hence, the lumping $(\mathcal{G}, p_X, \{g_i\})$ is not lumpable since the minimal graph for Y must contain the edge $\{1,3\}$, which is not in E. (Indeed, Y is a MRF w.r.t. the graph $(\{1,2,3\},\{\{1,3\}\})$.) Furthermore, one can show that $X_2 = 1$ iff $X_1 > 0$ and that $X_2 = -1$ iff $X_3 < 0$, hence $Y = (X_1, 0, X_3)$ contains the same information as X, i.e., the lumping is information preserving.

Example 3 (Lumpable and Information Preserving)**.** Let $X_2 := (X_1, Z_2, X_3)$, where X_1, Z_2, and X_3 are mutually independent. Then, we have the Markov path X_1—X_2—X_3 again, where the PMF p_X satisfies

$$p_X(x_1,(z_1,z_2,z_3),x_3) = \begin{cases} p_{X_1}(x_1)p_{Z_2}(z_2)p_{X_3}(x_3), & x_1 = z_1, x_3 = z_3 \\ 0, & \text{else.} \end{cases} \tag{5}$$

Now suppose that g_1 and g_3 are the identity mappings and that g_2 is such that $g_2(z_1,z_2,z_3) = z_2$. Obviously, the thus defined RVs Y_1, Y_2, and Y_3 are independent, i.e., Y is a MRF on the empty graph, and so $(\mathcal{G}, p_X, \{g_i\})$ is lumpable. Furthermore, it is clear that $H(g(X)) = H(X)$, and so the lumping is information preserving.

IV. Previous Work on MRFs

Yeung et al. characterized MRFs using the I-measure [5], [6]. Specifically, if X is a $\mathcal{G}$-MRF and $A \subseteq \mathcal{V}$, they investigated the minimal graph $\mathcal{G}_A := (A, E_A)$ on which X_A is a MRF. They showed that E_A contains $\{i,j\} \in [A]^2$ if either $\{i,j\} \in E$ or if there is a path between i and j in $\mathcal{G}$ of which all intermediate vertices lie in $\mathcal{V} \setminus A$, see [5, Th. 5] or [6, Th. 8]. More generally, Sadeghi [8] characterized probabilistic graphical models, admitting mixed graphs $\tilde{\mathcal{G}}$ with directed, doubly-directed, and undirected edges, and presented an algorithm that generates a corresponding graph for a subset $A \subseteq \mathcal{V}$ of the vertices of $\tilde{\mathcal{G}}$, cf. [8, Algorithm 1]. With the restriction to undirected graphs, this algorithm terminates with $\mathcal{G}_A$ as discussed in [6]. Much earlier, Pérez and Heitz investigated the problem above (for $(\mathcal{G}, p_X)$-MRFs) from a Gibbs field perspective, i.e., using potential functions, where they showed that X_A is a $(\mathcal{G}_A, p_{X_A})$-MRF [4, Th. 2], but $\mathcal{G}_A$ is only minimal if additional conditions are fulfilled [4, Th. 3].

To clarify some connections between previous works and the current one, consider the following setup. Given a MRF X w.r.t. a graph $\mathcal{G}_X$ and a transformation $p_{Y|X}$ of X to Y, assume the joint RV (X,Y) with PMF $p_{X,Y}$ is a MRF on a graph $\mathcal{G}_{X,Y}$. The vertex set of this graph is the disjoint union of the vertices of $\mathcal{G}_X$ and a set of vertices associated with Y, and the edge set is obtained from the edges of $\mathcal{G}_X$ and the transformation $p_{Y|X}$.[1] Now, the previous and current works are instantiations of this setup with the following specific assumptions: [5], [6] leave $p_{X,Y}$ unspecified and only consider the graph structure $\mathcal{G}_{X,Y}$,[2] [4] assumes $p_{Y|X}$ is strictly positive, and this work assumes

$$p_{Y|X}(y|x) = \prod_{i\in\mathcal{V}} \mathbb{I}[g_i(x_i) = y_i], \tag{6}$$

where $\mathbb{I}[\cdot]$ is the indicator function, i.e., that Y_i is a deterministic function of X_i. Each instantiation is non-trivial and interesting on its own, with applications indicated in its corresponding work cited above. (Problem 1 is motivated by Markov chain lumpability, see Section I and, e.g., [2, §6.3].)

Now, Problem 1 cannot be solved with the instantiation in [4] since the conditional distribution (6) is not strictly positive, nor can it be solved using [5], [6] since the framework therein finds a graph $\mathcal{G}_Y$ that is minimal for any $p_X \cdot p_{Y|X}$ in the family of distributions specified by $\mathcal{G}_{X,Y}$. In contrast, in Problem 1 we are given a fixed set of transformations $\{g_i\}$ and (often) a fixed distribution p_X. Indeed, if $\mathcal{G}_X$ is connected, then [5, Th. 5] states that one can find a PMF p_X and a set of transformations $\{p_{Y_i|X_i}\}$ such that Y does not satisfy any conditional independence statements. (See Example 1 for an explicit choice of p_X and $\{g_i\}$ that results in the complete graph in the case of the Markov path.)

Little work has been done regarding information-preserving lumpings of a MRF, see Problem 2. A work in a related direction is [9], which shows that under certain conditions the entropy $H(X_A)$, for $A \subseteq \mathcal{V}$, can be bounded from above by the entropy of a MRF w.r.t. the subgraph of $\mathcal{G}$ induced by A.

V. Sufficient Conditions for MRF Lumpability

Below we investigate Problem 1, namely, we determine sufficient conditions for the lumping $(\mathcal{G}, p_X, \{g_i\})$ to be lumpable. Note that according to Problem 1, $(\mathcal{G}, p_X, \{g_i\})$ is lumpable if Y is a $(\mathcal{G}, p_Y)$-MRF, even if $\mathcal{G}$ is not minimal for Y. We further assume within this section that $p_X(x) > 0$ for every $x \in \mathcal{X}$, and that g is surjective, i.e., $\mathcal{Y}$ is the image of $\mathcal{X}$ under g. This allows the characterization of a MRF via its connection to Gibbs fields. (Despite this assumption, the joint distribution p_{XY} is not strictly positive, see (6).) Specifically, let $\psi_A\colon \mathcal{X}_A \to \mathbb{R}$ be a potential function. We abuse notation and extend the domain of ψ_A to $\mathcal{X}$, i.e., for $x := (x_1,\dots,x_N) \in \mathcal{X}$ we write $\psi_A(x) := \psi_A(x_A)$, where $x_A := (x_i,\ i \in A)$. The following lemma gives the characterization required in this section.

Lemma 2 (Hammersley-Clifford [10])**.** *X is a $(\mathcal{G}, p_X)$-MRF satisfying $p_X(x) > 0$ for every $x \in \mathcal{X}$ iff there exists a family of potential functions $\{\psi_C,\ C \in \mathcal{C}\}$ such that*

$$\forall x \in \mathcal{X}\colon \quad p_X(x) = \frac{1}{Z}\prod_{C\in\mathcal{C}} \psi_C(x), \tag{7a}$$

[1] When $p_{Y|X}$ is a degenerate PMF, we have a deterministic mapping. For coordinatewise mappings, $p_{Y|X}$ factors as a product of $p_{Y_i|X_i}$, i.e., in $\mathcal{G}_{X,Y}$, each Y_i has one incident edge, which is the one connecting it to X_i. See Fig. 1.

[2] As mentioned earlier, the minimal graph $\mathcal{G}_Y$ w.r.t. which Y is a MRF can be found by applying [5, Th. 5] or [4, Th. 2 & 3] to $\mathcal{G}_{X,Y}$.

where $\mathcal{C}$ is the set of cliques of $\mathcal{G}$ and

$$Z := \sum_{x\in\mathcal{X}} \prod_{C\in\mathcal{C}} \psi_C(x). \tag{7b}$$

Since the potential functions in the family $\{\psi_C,\ C\in\mathcal{C}\}$ are defined on cliques, we call ψ_C a *clique potential*. Note that the choice of $\{\psi_C,\ C\in\mathcal{C}\}$ is not unique. Indeed, Lemma 2 may be satisfied with a subset of potential functions being identically one.

For a non-trivial set of functions $\{g_i\}$, Y is a $(\mathcal{G}, p_Y)$-MRF iff we can find a family of potential functions $\{U_C,\ C\in\mathcal{C}\}$ such that, for every $y\in\mathcal{Y}$

$$\begin{aligned} Z\cdot p_Y(y) &= Z\cdot \sum_{x\in g^{-1}(y)} p_X(x) \\ &= \sum_{x\in g^{-1}(y)} \prod_{C\in\mathcal{C}} \psi_C(x) = \prod_{C\in\mathcal{C}} U_C(y) \end{aligned} \tag{8}$$

where Z is the partition function from (7b). Such a family can obviously be found if, for all $y\in\mathcal{Y}$, the family $\{\psi_C,\ C\in\mathcal{C}\}$ is constant on the preimage $g^{-1}(y) := \{x\in\mathcal{X} : g(x)=y\}$. Specifically, if for every $C\in\mathcal{C}$ and for every $y\in\mathcal{Y}$ we have

$$\psi_C(x) = \psi_C(x'),\ \forall x,x'\in g^{-1}(y), \tag{9}$$

then we can choose $U_C(y)$ as this common value multiplied by the cardinality of the preimage $g^{-1}(y)$ to satisfy (8). The remainder of this section will give milder conditions than (9) that gurantee lumpability.

For any clique C that contains vertex i, we say ψ_C depends on x_i *only via* y_i if for all $y_i\in\mathcal{Y}_i$ and $x_i, x_i'\in g_i^{-1}(y_i)$

$$\psi_C(x_{\not{i}}, x_i) = \psi_C(x_{\not{i}}, x_i'), \forall x_{\not{i}}\in\mathcal{X}_{\not{i}} \tag{10}$$

otherwise, we say ψ_C *strictly* depends on x_i. The following result will assume that for every vertex i there is at most one clique potential that is allowed to strictly depend on x_i. For all i, let $C'(i)$ denote the corresponding clique. (If no potential function strictly depends on x_i then $C'(i)$ is chosen as any clique involving i.) We can view this as a mapping $C'\colon \mathcal{V}\to\mathcal{C}$ that assigns to each vertex i the unique clique that may strictly depend on x_i, which in effect partitions $\mathcal{V}$ into equivalence classes $\mathcal{V}_\ell$, $\ell=1,\dots,L$, such that all the vertices $i\in\mathcal{V}_\ell$ are assigned the same clique $C'(i)$. For convenience, the clique $C'(i)$, common to all $i\in\mathcal{V}_\ell$, will be denoted $C'(\mathcal{V}_\ell)$.

Proposition 1. *Assume X is a $(\mathcal{G}, p_X)$-MRF characterized by a family $\{\psi_C,\ C\in\mathcal{C}\}$ of potential functions such that, for all $i\in\mathcal{V}$, there is at most one clique whose potential may strictly depend on x_i, then Y is a $(\mathcal{G}, p_Y)$-MRF.*

Moreover, with C' and $\mathcal{V}_1,\dots,\mathcal{V}_L$ as above, the $(\mathcal{G}, p_Y)$-MRF is characterized by the family $\{U_C,\ C\in\mathcal{C}\}$ of potential functions, where

$$U_{C'(\mathcal{V}_\ell)}(g(x)) = \sum_{x'_{\mathcal{V}_\ell}\in g^{-1}_{\mathcal{V}_\ell}(g_{\mathcal{V}_\ell}(x_{\mathcal{V}_\ell}))} \psi_{C'(\mathcal{V}_\ell)}(x'_{\mathcal{V}_\ell}, x_{\mathcal{V}\setminus\mathcal{V}_\ell}) \tag{11a}$$

for $\ell = 1,\dots,L$, and

$$U_C(g(x)) = \psi_C(x), \forall C\in\mathcal{C}\setminus\cup_{j\in\mathcal{V}} C'(j). \tag{11b}$$

Corollary 1. *If* (9) *holds, then Proposition 1 is trivially fulfilled. In this case, $C'(i)$ is any clique of which i is a member and* (11a) *simplifies to*

$$U_{C'(\mathcal{V}_\ell)}(g(x)) = |g^{-1}_{\mathcal{V}_\ell}(g_{\mathcal{V}_\ell}(x_{\mathcal{V}_\ell}))|\cdot\psi_{C'(\mathcal{V}_\ell)}(x). \tag{12}$$

Since, even for a fixed joint PMF p_X, the family of potential functions is not unique, Y is a $(\mathcal{G}, p_Y)$-MRF if we can find at least one family of potential functions that characterizes p_X and for which Proposition 1 holds.

Example 4. Let X_1—X_2—X_3 be a Markov path and fix a set of functions $\{g_1, g_2, g_3\}$. Suppose that $\psi_{\{i\}}$, for $i=1,2,3$, are arbitrary, and $\psi_{\{1,2\}}(x_1,x_2) = U_{\{1,2\}}(g_1(x_1), g_2(x_2))$ and $\psi_{\{2,3\}}(x_2,x_3) = U_{\{2,3\}}(g_2(x_2), g_3(x_3))$ for some $U_{\{1,2\}}$ and $U_{\{2,3\}}$. Thus, only $\psi_{\{i\}}$ may strictly depend on x_i, and so Proposition 1 applies. Now, the same PMF p_X can be characterized using the potentials $\psi'_{\{1,2\}} = \psi_{\{1,2\}}\cdot\sqrt{\psi_{\{2\}}}$, $\psi'_{\{2,3\}} = \psi_{\{2,3\}}\cdot\sqrt{\psi_{\{2\}}}$, $\psi'_{\{1\}} = \psi_{\{1\}}$, $\psi'_{\{2\}} = 1$, and $\psi'_{\{3\}} = \psi_{\{3\}}$. Assuming ψ_2 strictly depends on x_2, then both $\psi_{\{1,2\}}$ and $\psi_{\{2,3\}}$ strictly depend on x_2, and so the condition in Proposition 1 is violated.

We now complement Proposition 1, which is based on clique potentials, by a sufficient condition for lumpability based on conditional entropies. This condition follows from Lemma 1, the data processing inequality, and the fact that conditioning reduces entropy.

Proposition 2. *Let X be a $\mathcal{G}$-MRF. If, for every $i\in\mathcal{V}$,*

$$H(Y_i|Y_{\mathcal{N}_i}) = H(Y_i|X_{\mathcal{N}_i}) \tag{13}$$

then Y is a $\mathcal{G}$-MRF.

Equation (13) gives an intuitive interpretation for lumpability of MRFs: If (but not only if, see Example 5 below) the neighbors of X_i are not more informative about the outcome of Y_i than the function of these neighbors, then Y is a $\mathcal{G}$-MRF. In other words, Y is a $\mathcal{G}$-MRF if the lumping is such that $Y_{\mathcal{N}_i}$ captures all information in $X_{\mathcal{N}_i}$ that is relevant to Y_i.

Example 5. Let $X := (X_1, X_2)$ be a Markov path, i.e., $\mathcal{V} = \{1,2\}$ and $E=\{1,2\}$. Trivially, since $\mathcal{G}$ is the complete graph, Y is a $\mathcal{G}$-MRF for every set of functions $\{g_1, g_2\}$. However, one can construct examples for p_X and $\{g_1,g_2\}$ such that there exists $y\in\mathcal{Y}$ and a pair $x_1, x_1'\in g_1^{-1}(y_1)$ such that

$$p_{Y_2|X_1}(y_2|x_1) \neq p_{Y_2|X_1}(y_2|x_1'). \tag{14}$$

Thus, the condition of Proposition 2 does not hold, showing that it is only sufficient but not necessary.

There is some similarity between (13) and an information-theoretic sufficient condition for the lumpability of an irreducible and aperiodic Markov chain X_1—X_2—X_3—$\cdots$ (see [2] for terminology). Suppose that X is stationary, i.e., the alphabets of X_i are all the same, $p_{X_{i+1}|X_i} = p_{X_{j+1}|X_j}$ for every $i,j\in\mathbb{N}$, and the initial distribution p_{X_1} coincides with the unique distribution invariant under the one-step conditional distribution $p_{X_{i+1}|X_i}$. If further all the functions g_i are identical, i.e., $g_i = g_0$, $i \in \mathcal{V}$, for some function g_0, then one can show that the tuple $(\mathcal{G}, p_X, g_0)$ is lumpable if [3, Th. 2]

$$H(Y_i|X_{i-1}) = H(Y_i|Y_{i-1}). \tag{15}$$

(By stationarity, it suffices that (15) holds for any i.) Indeed, the main difference between (13) and (15) is that the latter is conditioned on only a subset of the neighbors, which corresponds to the case in which $\mathcal{G}$ is directed, i.e., for $X_1 \to X_2 \to \cdots$. Proposition 2 shows that, for undirected graphs, (13) takes the place of (15) in a sufficient condition for lumpability.

VI. Information-Preserving MRF Lumpings

We next briefly talk about information-preserving lumpings of MRFs, see Problem 2. A lumping can only be information preserving if g maps the support of p_X injectively. If the support of p_X coincides with $\mathcal{X}$, then only trivial sets of functions $\{g_i\}$, in which every g_i is injective, can be information preserving. In this section, we therefore drop the assumption that p_X is positive on $\mathcal{X}$. However, while it is clear that $H(X) = H(Y)$ iff g is injective on the support of p_X, this does not imply that every g_i is injective on the support of p_{X_i}. In other words, a lumping $(\mathcal{G}, p_X, \{g_i\})$ can be information preserving even if some or all of the functions g_i are non-injective, i.e., even if $H(X_i) > H(Y_i)$ for some $i \in \mathcal{V}$.

Proposition 3. *Let X be a $(\mathcal{G}, p_X)$-MRF.*

- *For any graph $\mathcal{G}$, if the lumping $(\mathcal{G}, p_X, \{g_i\})$ is information preserving, then*

$$\forall i \in \mathcal{V}: \quad H(X_i|Y_i, X_{\mathcal{N}_i}) = 0. \tag{16a}$$

- *For any chordal graph $\mathcal{G}$, the lumping $(\mathcal{G}, p_X, \{g_i\})$ is information preserving if there exist a vertex permutation $v_1, \dots, v_N$ and sets $A_{v_i} = \mathcal{N}_{v_i} \cap \{v_1, \dots, v_{i-1}\}$ such that*

$$\forall i \in \mathcal{V}: \quad H(X_{v_i}|Y_{v_i}, X_{A_{v_i}}) = 0. \tag{16b}$$

Example 6. Let $X_1 = X_2$, i.e., X is a MRF on a path, which is a cordal graph. Assume that $g_1 \equiv g_2$ and that $g = (g_1, g_2)$ is non-injective on the support of p_X. Thus, $H(g(X)) < H(X)$.

Since $H(X_1|X_2) = 0$ and $H(X_2|X_1) = 0$, we have $H(X_1|g_1(X_1), X_2) = 0$ and $H(X_2|g_2(X_2), X_1) = 0$, i.e., the necessary condition for information preservation (16a) holds. However, we have that $H(X_1|Y_1) > 0$ due to the non-injectivity of g_1, and so (16b) does not hold for the permutation $(v_1, v_2) = (1, 2)$. A similar argument holds for the permutation $(v_1, v_2) = (2, 1)$. Thus, the sufficient condition for chordal graphs (16b) is violated.

Remark 1. Let $X_1 — X_2 — X_3 — \cdots$ be an irreducible, aperiodic, and stationary Markov chain. The graph w.r.t. which X is a MRF is an (infinite) path, which is chordal. Since $A_i = \{i-1\}$ (under the choice $v_i = i$ for all i) and due to stationarity, the sufficient condition in (16b) simplifies to $H(X_2|Y_2, X_1) = 0$. We thus recover [3, Prop. 4].

While the condition that g maps the support of p_X injectively is an equivalent characterization of information preservation, the conditions in Proposition 3 (that are only necessary or sufficient) have practical justification. Indeed, for alphabets $\mathcal{X}_i$ with fixed cardinality, the support of p_X grows exponentially in the number N of vertices. In contrast, (16a) requires checking whether g_i maps the support of $p_{X_i|X_{\mathcal{N}_i}}$ injectively for every i; the number of parameters characterizing this conditional PMF is exponential only in the size of the neighborhood of i, which is much smaller than N for sparse graphs. Thus, rather than checking g *globally*, which is exponential in N, it suffices to check a computationally less expensive *local* condition for each g_i.

We finally remark that Proposition 3 holds regardless whether $(\mathcal{G}, p_X, \{g_i\})$ is lumpable or not. A better understanding of the interactions between lumpability and information preservation, i.e., between Problems 1 and 2, seems to be of practical and theoretical interest. Thus, a closer investigation of these interactions shall be the subject of future work.

Acknowledgments

The authors thank Raymond W. Yeung for valuable discussions and for pointing out [8]. The work of Bernhard C. Geiger has been supported by the HiDALGO project and has been partly funded by the European Commission's ICT activity of the H2020 Programme under grant agreement number 824115. The Know-Center is funded within the Austrian COMET Program - Competence Centers for Excellent Technologies - under the auspices of the Austrian Federal Ministry of Climate Action, Environment, Energy, Mobility, Innovation and Technology, the Austrian Federal Ministry of Digital and Economic Affairs, and by the State of Styria. COMET is managed by the Austrian Research Promotion Agency FFG.

References

[1] L. Gurvits and J. Ledoux, "Markov property for a function of a Markov chain: a linear algebra approach," *Linear Algebra Appl.*, vol. 404, pp. 85–117, 2005.

[2] J. G. Kemeny and J. L. Snell, *Finite Markov Chains*, 2nd ed. Springer, 1976.

[3] B. C. Geiger and C. Temmel, "Lumpings of Markov chains, entropy rate preservation, and higher-order lumpability," *J. Appl. Probab.*, vol. 51, no. 4, pp. 1114–1132, Dec. 2014.

[4] P. Perez and F. Heitz, "Restriction of a Markov random field on a graph and multiresolution statistical image modeling," *IEEE Trans. Inf. Theory*, vol. 42, no. 1, pp. 180–190, Jan. 1996.

[5] R. W. Yeung, A. Al-Bashabsheh, C. Chen, Q. Chen, and P. Moulin, "Information-theoretic characterizations of Markov random fields and subfields," in *Proc. IEEE Int. Symp. on Inf. Theory (ISIT)*, Jun. 2017, pp. 3040–3044.

[6] R. W. Yeung, A. Al-Bashabsheh, C. Chen, Q. Chen, and P. Moulin, "On information-theoretic characterizations of Markov random fields and subfields," *IEEE Trans. Inf. Theory*, vol. 65, no. 3, pp. 1493–1511, 2019.

[7] B. C. Geiger and A. Al-Bashabsheh, "On functions of Markov random fields," May 2020, preprint `arXiv:2005.13908 [cs.IT]`.

[8] K. Sadeghi, "Marginalization and conditioning for LWF chain graphs," *Ann. Statist.*, vol. 44, no. 4, pp. 1792–1816, Aug. 2016.

[9] M. G. Reyes and D. L. Neuhoff, "Entropy bounds for a Markov random subfield," in *Proc. IEEE Int. Symp. on Inf. Theory (ISIT)*, Seoul, Jul. 2009, pp. 309–313.

[10] J. Hammersley and P. Clifford, "Markov fields on finite graphs and lattices," 1971, unpublished manuscript.