Exact marginal inference in general Markov random field models using linear programming

Ikhlef Bechar

Aix Marseille Univ, Université de Toulon, CNRS, ENSAM, LSIS
83957 La Garde, France

Abstract

This paper addresses the problem of exact marginal inference in general higher-order Markov random field (MRF) models. This is a fundamental AI problem, yet, renowned for its hardness. Nevertheless, by introducing an algebraic framework (referred to as the ortho-marginal framework)–which turns out to be, at once, a general approximation framework of discrete functions with arbitrary accuracy by means of their sets of margins, as well as a a principled means for modeling locally consistent functions from a global perspective–, we are able to devise a linear programming approach which can solve the marginal inference problem for any instance of the MRF model both exactly and efficiently.

Keywords: Higher-order MRF model, marginal distribution, marginal inference, pseudo-marginal, ortho-marginal space, linear programming, polynomial-time solution.
2000 MSC: 220,
2000 MSC: 190

1. Introduction

Assume an undirected graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, with $\mathcal{V}$ and $\mathcal{E}$ standing for its vertex-set and edge-set, respectively, and let us call, throughout, a hypervertex

Email address:
ikhlef.bechar@gmail.com,ikhlef.bechar@univ-tln.fr,ikhlef.bechar@lsis.org (Ikhlef Bechar)

of $\mathcal{G}$ any subset of vertices of $\mathcal{G}$. Moreover, let $\mathcal{L}$ stand for a discrete label-set which we assume, without loss of generality, to be the integer set $\{0, \ldots, L-1\}$, with L standing for an integer which is greater than, or equal to 2, and whatever a hypervertex c of $\mathcal{G}$ and an integer vector $x \in \mathcal{L}^n$, denote $x_c = (x_i)_{i \in c}$. Next, suppose $\mathcal{X} = \big(\mathcal{X}_i\big)_{i \in \mathcal{V}}$ is a Markov random field taking values in $\mathcal{L}^n$ of which probability distribution, thus, verifies the strict positivity constraint expressing as $\mathbb{P}\big(\mathcal{X} = x\big) > 0, \forall x \in \mathcal{L}^n$, as well as the conditional property expressing as:

$$\mathbb{P}\big(\mathcal{X}_i = x_i / X_{\neq i} = x_{\neq i}\big) = \mathbb{P}\big(\mathcal{X}_i = x_i / \mathcal{X}_{\mathrm{ne}(i)} = x_{\mathrm{ne}(i)}\big),\, \forall x \in \mathcal{L}^n,\, \forall i \in \mathcal{V} \tag{1}$$

where we have denoted $x_{\neq i} = (x_j)_{j \in \mathcal{V}/\{i\}}$, $\forall i \in \mathcal{V}$, $\forall x \in \mathcal{L}^n$, and $\forall i \in \mathcal{V}$, $\mathrm{ne}(i)$ stands for the neighborhood of i in $\mathcal{G}$ defined as: $\mathrm{ne}(i) = \big\{j \in \mathcal{V}, \text{s.t.}, (i,j) \in \mathcal{E}\big\}$.

Denote by $\mathcal{C}$ the set of maximal cliques of $\mathcal{G}$. Then, one also has:

$$\mathrm{ne}(i) = \cup_{c \in \mathcal{C}, \text{s.t.}, i \in c} \big\{j \neq i \wedge j \in c\big\},\, \forall i \in \mathcal{V} \tag{2}$$

moreover, a completely equivalent definition of MRF $\mathcal{X}$ [1] is that there exists strictly positive functions $\phi_c : \mathcal{L}^{|c|} \longrightarrow \mathbb{R}, \forall c \in \mathcal{C}$, such that, the joint distribution of $\mathcal{X}$ writes as:

$$\mathbb{P}\big(\mathcal{X} = x\big) = \tfrac{1}{Z} \Pi_{c \in \mathcal{C}}\, \phi_c(x_c),\, \forall x \in \mathcal{L}^n \tag{3}$$

where Z (called the partition function) stands for a normalization constant which is defined, in such a way that, $\sum_{x \in \mathcal{L}^n} \mathbb{P}\big(\mathcal{X} = x\big) = 1$, but, which computation is generally hurdled from a practical point of view.

Now, given a hypervertex-set of $\mathcal{G}$ denoted by $\mathcal{S}$, the MRF marginal inference problem–which we will simply refer to in the remainder as "MRFMiP"– typically, consists in computing the marginal distributions of the random vectors $\mathcal{X}_s, \forall s \in \mathcal{S}$, and let us symbolize such a problem as:

$$\begin{aligned}
&\textit{Compute: } \big\{\, \mathbb{P}\big(\mathcal{X}_s = x_s\big),\, \forall x_s \in \mathcal{L}^{|s|},\, \forall s \in \mathcal{S} \,\big\} \\
&\text{s.t.,} \\
&\begin{cases}
\mathbb{P}\big(\mathcal{X} = x\big) = \tfrac{1}{Z} \Pi_{c \in \mathcal{K}}\, \phi_c(x_c),\, \forall x \in \mathcal{L}^n \\
\sum_{x \in \mathcal{L}^n} \big(\tfrac{1}{Z} \Pi_{c \in \mathcal{K}}\, \phi_c(x_c)\big) = 1
\end{cases}
\end{aligned} \tag{4}$$

For the sake of example, assume that $\mathcal{S} = \mathcal{V}$. Then, MRFMiP (4) reduces to the computation of the unary marginal probabilities of the form $\mathbb{P}\big(\mathcal{X}_i = x_i\big)$, $\forall x_i \in \mathcal{L}$, $\forall i \in \mathcal{V}$. Moreover, for the sake of computational tractability, one typically assumes, first, that $|\mathcal{S}|$ (standing for the cardinality of $\mathcal{S}$ in terms of number of hypervertices) is some polynomial function of n, second, the number $r = \sup_{s \in \mathcal{S}} \big\{ |s| \big\}$ referred to as the order of hypervertex-set $\mathcal{S}$ is small comparatively to n, for instance:

- either r is a constant (e.g.; 2, 3, etc), which means that r does not grow with n (think, for instance, of a $2D$ rectangular image grid of which neighborhood size is fixed independently of image size),
- or, at worst, r grows in a logarithmic way with n, i.e.; there exists a nonnegative real constant k, such that, $r = O(\log^k(n))$.

The remainder of this paper is structured as follows. After providing a background of marginal inference in MRF models, we introduce the ortho-marginal framework both for general discrete function approximation,as well as for global modeling of locally consistent sets of functions. After that, as a first step, we formulate the MRF marginal inference problem as a linear system of equations involving an exponential number of probabilities, and as a second step, we reduce the latter to a new linear system of equations, solely, involving a polynomial-time number of marginal probabilities with respect to a properly chosen hypervertex-set, and we show that the MRF marginal inference problem is completely solved by the latter. Last but not least, by using the ortho-marginal framework above, we show that one may completely reformulate the latter linear system of equations as a linear program which solves MRFMiP (4) exactly and efficiently.

2. Background of marginal inference in MRF models

Markov random fields (MRFs) turn out to be an important class of undirected graphical models enjoying the locality property [1, 2, 3, 4]. Historically, MRF models had long been known in the field of statistical physics [5, 6, 7, 8], before they have been brought to, and broadly popularized in the computer science literature [9, 10, 11]. Nowadays, MRFs happen to be a standard AI tool which finds several applications including inference in social and biological networks, model simulation and resolution in computational chemistry and computational biology, natural language processing and speech recognition, as well as a plethora of computer vision and pattern recognition applications, to name a few. The main strength of the MRF framework resides in the fact that it offers a principled means for modeling stochastic interactions between locally interacting network entities while being able to capture, at the same time, a realistic probabilistic model of a given network as a whole. Nevertheless, a major bottleneck of the MRF framework (more generally, of graphical models), resides in the fact that, generally speaking, inference (be it Maximum-A-Posteriori (MAP) inference, or marginal inference) in their arbitrary instances is NP hard [12, 13, 14, 15, 16, 17, 18, 19, 20, 21]. By the latter sentence, we mean that, unless $P = NP$, one may not hope achieving a polynomial-time algorithm for accessing, either exactly or, even, approximately the posterior likelihoods of arbitrary graphical models. As a matter of fact, both MAP inference and marginal inference in MRFs have been, traditionally, dealt with either using deterministic or stochastic heuristical approaches [4]. Since this paper is primarily concerned with marginal inference in MRF models, moreover, the literature on inference in graphical models is huge, we settle, hereafter, for mentioning some of the closest MRF marginal inference approaches to ours in the present paper.

Having said that, marginal inference in graphical models has become, since over a decade, an almost inescapable research topic in the literature of AI and beyond, and the reason for such enthusiasm is twofold. First of all, explosion of the internet technology (e.g.; social network technology), as well as of the computing power (e.g.; clusters of dozens of CPUs, GPU technology, etc), thus, resulting in an increasing need for treating increasingly bigger, more heterogeneous, and more complex volumes of computer-generated data (also known as big data), and hence, in the need for better performing and faster inference algorithms. Second, marginal inference often arises as a seducing alternative approach to MAP inference in graphical models, either for achieving approximate MAP inference in MRF instances which do not fall into any of the known solvable categories of MRFs using off-the-shelf MAP solvers (e.g.; graph-based algorithms, MPLP, etc), or, simply, for speeding up computations. In this regard, belief propagation (BP) [22, 23, 24] has been extensively used for optimizing a plethora of MRF instances [25, 26]. Although BP is exact for arbitrary tree-structured graphical models, it is known to perform poorly on graphical models presenting loopy-like structures such as MRFs, in such a way that, a BP algorithm may be stuck, indefinitely, in a loop, thereby, resulting in generally poor solutions. Therefore, for the purpose of overcoming such a limitation of BP, several improvements to the standard BP algorithm have been proposed in the literature, including the tree-reweighted belief propagation (TRBP) algorithm [27], the convergent TRBP [26], convergent message-passing based on geometric programming [28], the dual decomposition algorithm [29, 30], and so forth. A common point to all the aforementioned message passing like inference algorithms is that they all attempt to optimize over the pseudo-marginal polytope (i.e.; with respect to a locally consistent (or pseudo) marginals-set), moreover, it has been reported [31, 32] that tighter relaxations of the global marginal polytope may result in a significant improvement of the results by combining the conditional gradient method (also called Frank-Wolfe) [33, 34] with standard LP solvers. Building on the work in [31], it has also been shown that optimization of the TRW objective may be performed over the global marginal polytope using the Frank-Wolfe method [35], although such a method has been reported to suffer from two major drawbacks, including slowness and numerical instability in the vicinity of the feasible set (i.e.; the global marginal polytope) due to the explosion of the gradient of the entropy terms of the TRW objective, in such a way that, convergence of the Frank-Wolfe method may not be guaranteed, thus, an amendment of the latter method in a way which guarantees its global convergence has been proposed [36].

Therefore, in this paper, we attempt to take a step forward with respect to the existing state of the art methods, by proposing to solve the marginal inference problem in general MRF models in an exact and efficient way.

3. The ortho-marginal framework

In this section, we describe an algebraic approach–which we refer to as the ortho-marginal framework– of which usefulness in the present paper is twofold. First, it allows to view any set of locally constant functions (see Definition 6) as actual margins of a global (yet non-unique) mother function with respect to some hypervertex-set, thereby, allowing to apprehend problems– expressing in terms of the margins of a function with respect to a given hypervertex-set– from a global perspective. Second, since it induces an orthogonal space of functions, one may also view it as an approximation framework of functions $f : \mathcal{L}^n \to \mathbb{R}$ with arbitrary accuracy which, moreover, turns out to be, especially, useful for solving MRFMiP (4).

But beforehand, in order to fix ideas once and for all in the remainder, let us begin this section by introducing a few useful definitions. Therefore, throughout this section, we assume that $\mathcal{K}$ is a hypervertex-set with respect to $\mathcal{V}$, moreover, we assume some order (e.g.; a lexicographic order) on the elements of $\mathcal{K}$ which means that, whatever $c, c' \in \mathcal{K}$, if $c \neq c'$, then either one one has $c < c'$, or one has $c' < c$.

3.1. Definitions

Definition 1 (Maximal hypervertex-set). *One says that $\mathcal{K}$ is maximal, if and only if:*

$$\forall c, c' \in \mathcal{K}, c' \subseteq c \Rightarrow c' = c$$

or, in plain words, if one may not find in $\mathcal{K}$ both a hypervertex, and any of its subsets.

Definition 2 (Frontier hypervertex-set). *One defines the frontier of $\mathcal{K}$, denoted by Front$\big(\mathcal{K}\big)$, as the smallest maximal hypervertex-set which is contained* *in $\mathcal{K}$. In plain words, $Front(\mathcal{K})$ is the hypervertex-set which contains all the hypervertices in $\mathcal{K}$ which are not included in any of its other hypervertices.*

Definition 3 (Ancestor hypervertex). *Suppose a hypervertex $c \in \mathcal{K}/Front(\mathcal{K})$ (if any). Then, we call an ancestor hypervertex of c, any hypervertex $\tilde{c} \in Front(\mathcal{K})$, such that, $c \subset \tilde{c}$.*

Definition 4 (Ancestry function). *We call an ancestry function with respect to $\mathcal{K}$, any function:*

$$\begin{array}{rccc} anc: & \mathcal{K}/Front(\mathcal{K}) & \longrightarrow & Front(\mathcal{K}) \\ & c & \longmapsto & anc(c) \end{array}$$

such that, $\forall c \in Front(\mathcal{K})$, $anc(c)$ is an ancestor of c in $Front(\mathcal{K})$.

Please note that the ancestor of some hypervertex may not be unique, hence, the function anc$(\cdot)$ may not be unique too.

Definition 5 (Margin). *Suppose a function $u : \mathcal{L}^n \to \mathbb{R}$, and a hypervertex $c \in \mathcal{K}$. Then, one defines the margin of u with respect to c as the function $u_c : \mathcal{L}^{|c|} \to \mathbb{R}$ defined as:*

$$u_c(x_c) = \sum_{i \in \mathcal{V}/c} \sum_{x_i \in \mathcal{L}} u(x_1, \ldots, x_i, \ldots, x_n),\ \forall x_c \in \mathcal{L}^{|c|} \tag{5}$$

Definition 6 (Pseudo-marginal-set). *One says that a set of local functions of the form $\{u_c : \mathcal{L}^{|c|} \to \mathbb{R}, \forall c \in \mathcal{K}\}$ is a pseudo-marginals-set (or a set of locally consistent functions) with respect to $\mathcal{K}$, if and only if, it satisfies the following identities:*

$$\begin{cases} \forall c, t \in Front(\mathcal{K}),\ c \cap t \neq \emptyset \Rightarrow \sum_{i \in c/t} \sum_{x_i \in \mathcal{L}} u_c(x_c) = \sum_{i \in t/c} \sum_{x_i \in \mathcal{L}} u_t(x_t) \\ \forall c \in \mathcal{K}/Front(\mathcal{K}),\ u_c(x_c) = \sum_{i \in anc(c)/c} \sum_{x_i \in \mathcal{L}} u_{anc(c)}(x_{anc(c)}),\ \forall x_c \in \mathcal{L}^{|c|} \end{cases} \tag{6}$$

where c/t stands for the hypervertex of which sites belong to c but do not belong to t, and $anc(\cdot)$ stands for an arbitrary ancestor function with respect to $\mathcal{K}$ (see Definition 4).

Clearly, any set of actual margins with respect to $\mathcal{K}$ of an arbitrary function $: \mathcal{L}^n \to \mathbb{R}$ also defines a pseudo-marginals-set with respect to $\mathcal{K}$.

Convention 1. *Denote by* $\emptyset$ *the empty hypervertex (i.e.; a one which does not contain any hypervertex), and convene, henceforth, that whatever a function* $u : \mathcal{L}^n \to \mathbb{R}$*, the margin of* u *with respect to* $\emptyset$*, simply, denoted by* $u_\emptyset$*, is the quantity* $u_\emptyset = \sum_{x \in \mathcal{L}^n} u(x)$*.*

Definition 7 (Frontier-closure of a hypervertex-set). *One defines the frontier-closure of* $\mathcal{K}$ *as the hypervertex-set with respect to* $\mathcal{V}$*, denoted by* $Fclos_\cap(\mathcal{K})$*, such that:*

1. $Front(\mathcal{K}) \subseteq Fclos_\cap(\mathcal{K}), \quad \emptyset \subset Fclos_\cap(\mathcal{K})$,
2. $\forall c, c' \in Fclos_\cap(\mathcal{K}),\ c \cap c' \in Fclos_\cap(\mathcal{K})$.

3.2. Main results of the ortho-marginal framework

First of all, Theorem 1 below establishes that marginalization of any function $f : \mathcal{L}^n \to \mathbb{R}$ with respect to $\mathcal{K}$ is intimately related to an orthogonal projection of f.

Theorem 1. *Let* $f : \mathcal{L}^n \to \mathbb{R}$ *stand for an arbitrary real-valued function. Then,* f *may write as a direct sum of two functions* $u : \mathcal{L}^n \to \mathbb{R}$*, and* $v : \mathcal{L}^n \to \mathbb{R}$ *as:* $f = u \oplus v$*, such that:*

1. *the margins-set with respect to* $\mathcal{K}$ *of* u *coincides with the one of* f*,*
2. *all the margins of* v *with respect to* $\mathcal{K}$ *are identically equal to zero.*
3. *the closed-form expression of function* u *is given by:*

$$u(x) = \sum_{c \in Fclos_\cap(\mathcal{K})} \frac{\rho_c\, f_c(x_c)}{L^{n-|c|}},\ \forall x \in \mathcal{L}^n$$

where $\forall c \in Fclos_\cap(\mathcal{K})$*,* $f_c : \mathcal{L}^{|c|} \to \mathbb{R}$ *stands for the margin of* f *with respect to* c*, and the integer coefficients* $\rho_c, \forall c \in Fclos_\cap(\mathcal{K})$ *are iteratively given by:*

$$\rho_c = \begin{cases} 1 & ,\ \text{if } c \in Front(\mathcal{K}), \\ 1 - \sum_{t \in Fclos_\cap(\mathcal{K})\ s.t.\ c \subset t} \rho_t & ,\ \text{if } c \in Fclos_\cap(\mathcal{K}) / Front(\mathcal{K}) \end{cases} \quad (7)$$

Furthermore, introduce operator denoted by $\mathscr{O}_{\mathcal{K}}$ and defined as:

$$\big(\mathscr{O}_{\mathcal{K}} f\big)(x) = \textstyle\sum_{c \in Fclos_{\cap}\left(\mathcal{K}\right)} \frac{\rho_c\, f_c(x_c)}{L^{n-|c|}},\ \forall x \in \mathcal{L}^n \tag{8}$$

Then, $\mathscr{O}_{\mathcal{K}}$ is an orthogonal projection.

The proof of Theorem 1 is sketched in Appendix .1.

Notation 1. *We refer in the remainder to the operator $\mathscr{O}_{\mathcal{K}}$ as the ortho-marginal operator with respect to hypervertex-set $\mathcal{K}$.*

Theorem 2 below builds on the result of Theorem 1 for establishing that any pseudo-marginals-set with respect to $\mathcal{K}$ may be viewed as the actual margins-set with respect to $\mathcal{K}$ of a global, yet non-unique, function : $\mathcal{L}^n \to \mathbb{R}$.

Theorem 2. *Suppose $\big\{u_c : \mathcal{L}^{|c|} \to \mathbb{R}, \forall c \in \mathcal{K}\big\}$ is a pseudo-marginals-set, thus verifying identities (6). Then, whatever a function $v : \mathcal{L}^n \to \mathbb{R}$, the function $u : \mathcal{L}^n \to \mathbb{R}$ defined as:*

$$u(x) = \big(v(x) - (\mathscr{O}_{\mathcal{K}} v)(x)\big) + \sum_{c \in Fclos_{\cap}\left(\mathcal{K}\right)} \frac{\rho_c\, u_c(x_c)}{L^{n-|c|}},\ \forall x \in \mathcal{L}^n \tag{9}$$

verifies that its margins-set with respect to $\mathcal{K}$ coincides with the set $\big\{u_c, \forall c \in \mathcal{K}\big\}$, said otherwise, one has:

$$\sum_{i \in \mathcal{V}/c} \sum_{x_i \in \mathcal{L}} u(x_1, \ldots, x_i, \ldots, x_n) = u_c(x_c),\ \forall x_c \in \mathcal{L}^{|c|},\ \forall c \in \mathcal{K}$$

where the linear coefficients $\rho_c, \forall c \in Fclos_{\cap}\big(\mathcal{K}\big)$ are defined according to formula (7) above.

The proof of Theorem 2 is sketched in Appendix .2.

Definition 8 (Ortho-marginal space). *The ortho-marginal space with respect to $\mathcal{K}$ denoted by $\mathscr{M}_{\mathcal{K}}$, is defined as the linear function space which is given by:*

$$\mathscr{M}_{\mathcal{K}} = \Big\{u : \mathcal{L}^n \to \mathbb{R},\ s.t., \mathscr{O}_{\mathcal{K}} u \equiv u\Big\}$$

We also denote by $\bar{\mathscr{M}}$ the complement space of $\mathscr{M}$, defined as:

$$\bar{\mathscr{M}}_{\mathcal{K}} = \Big\{v : \mathcal{L}^n \to \mathbb{R},\ s.t., \mathscr{O}_{\mathcal{K}} v \equiv 0\Big\}$$

Remark 1. *One notes that any function* $u \in \mathscr{M}_{\mathcal{K}}$ *reflexively writes in terms of its margins with respect to* $Fclos_{\cap}\big(\mathcal{K}\big)$ *as:*

$$u(x) = \sum_{c \in Fclos_{\cap}\big(\mathcal{K}\big)} \frac{\rho_c \, u_c(x_c)}{L^{n-|c|}}, \, \forall x \in \mathcal{L}^n$$

where $\forall c \in Fclos_{\cap}\big(\mathcal{K}\big)$, u_c *stands for the margin of* u *with respect to* c.

Proposition 1. *Suppose a real-valued function* $h : \mathcal{L}^n \to \mathbb{R}$. *Then, one has* $h \in \mathscr{M}_{\mathcal{K}}$, *if and only if, there exists a set of local functions* $\big\{h_c : \mathcal{L}^{|c|} \to \mathbb{R}, \forall c \in \mathcal{K}\big\}$ *(not to be confused here with the margins of* h *with respect to* $\mathcal{K}$*), such that:*

$$h(x) = \textstyle\sum_{c \in \mathcal{K}} h_c(x_c), \, \forall x \in \mathcal{L}^n$$

The proof of Proposition 1 is sketched in Appendix .3.

Proposition 2. *One has:*

$$\forall h, h' \in \mathscr{M}_{\mathcal{K}}, \, h \equiv h' \Leftrightarrow h_c(x_c) = h'_c(x_c), \, \forall x_c \in \mathcal{L}^{|c|}, \, \forall c \in Front(\mathcal{K})$$

where $\forall c \in \mathcal{K}$, h_c *and* h'_c *stand for the margins with respect to* c *of* h *and* h', *respectively.*

Proof 1. *The proof of Proposition 2 follows immediately from the definition of* $\mathscr{M}_{\mathcal{K}}$, *since if* $h, h' \in \mathscr{M}$, *then both* h *and* h' *write as a linear combination of their respective margins with respect to* $Fclos_{\cap}\big(\mathcal{K}\big)$, *which then must coincide if* $h \equiv h'$, *and vice-versa.*

4. Reformulation of MRFMiP (4) in terms of marginal distributions

The goal of this section is to reformulate MRFMiP (4) as an equivalent problem expressing, solely, in terms of a set of marginal distributions with respect to an appropriately chosen hypervertex-set, thereby, setting the stage for a further LP relaxation of MRFMiP (4) (see section 5).

For the sake of notational convenience in the remainder, we begin this section by introducing the notions of augmented neighborhood and neighborhood-wise conditional probabilities.

Definition 9 (Augmented neighborhood). *We call the augmented neighborhood of any vertex $i \in \mathcal{V}$ the hypervertex denoted by $ne^{+}(i)$ and defined as $ne^{+}(i) = \{i\} \cup ne(i)$.*

Definition 10 (Neighborhood-wise conditional probabilities). *Let us introduce the local functions:*

$$\pi_i(x_{ne^{+}(i)}) \quad = \quad \mathbb{P}\big(\mathcal{X}_i = x_i / \mathcal{X}_{ne(i)} = x_{ne(i)}\big),\ \forall x_{ne^{+}(i)} \in \mathcal{L}^{|ne^{+}(i)|},\ \forall i \in \mathcal{V}$$

which we refer to in the remainder as the neighborhood-wise probabilities of MRF $\mathcal{X}$. Moreover, by using formula (3) above, the latter are explicitly given in closed-form by:

$$\pi_i(x_{ne^{+}(i)}) \quad = \quad \frac{\Pi_{c \in \mathcal{C}, s.t., i \in c} \phi_c(x_c)}{\sum_{x_i \in \mathcal{L}} \left(\Pi_{c \in \mathcal{C}, s.t., i \in c} \phi_c(x_c)\right)},\ \forall x_{ne^{+}(i)} \in \mathcal{L}^{|ne^{+}(i)|},\ \forall i \in \mathcal{V}$$

Clearly, by assumption, one has $\pi_i(x_i; x_{\mathrm{ne}(i)}) > 0$, $\forall x_{\mathrm{ne}^{+}(i)} \in \mathcal{L}^{|\mathrm{ne}^{+}(i)|}$, moreover, one has:

$$\textstyle\sum_{x_i \in \mathcal{L}} \pi_i(x_{\mathrm{ne}^{+}(i)}) \quad = \quad 1,\ \forall x_{\mathrm{ne}(i)} \in \mathcal{L}^{|\mathrm{ne}(i)|},\ \forall i \in \mathcal{V}$$

Convention 2. *Suppose a hypervertex c of $\mathcal{G}$ which we assume, without loss of generality, to be the integer subset $\{1, \ldots, k\}$, with $k = |c| \in \mathbb{N}^{+}$, and suppose an arbitrary function $f : \mathcal{L}^{|c|} \to \mathbb{R}$. Then, for sake of notational conciseness, we shall abuse notation throughout, by convening that $\forall i \in c$, and $\forall x_{c/\{i\}} \in \mathcal{L}^{|c|-1}$, one has:*

$$\textstyle\sum_{x_i \in \mathcal{L}} f(x_c) = \sum_{y_i \in \mathcal{L}} f(x_1, \ldots, x_{i-1}, y_i, x_{i+1}, \ldots, x_k)$$

Proposition 3. *The joint distribution of MRF $\mathcal{X}$ is the unique solution of the following linear system of equations:*

$$\begin{cases} p(x) - \pi_i(x_{ne^{+}(i)}) \sum_{x_i \in \mathcal{L}} p(x) = 0,\ \forall x \in \mathcal{L}^n,\ \forall i \in \mathcal{V} \\ \sum_{x \in \mathcal{L}^n} p(x) = 1 \\ p(x) \geq 0,\ \forall x \in \mathcal{L}^n \end{cases} \tag{10}$$

The proof of Proposition 3 is sketched in Appendix .4.

Remark 2. *As an aside, first, one easily checks that, whatever a probability distribution* $p(x), \forall x \in \mathcal{L}^n$*, and* $\forall i \in \mathcal{V}$*, the function* $q_i : \mathcal{L}^n \to \mathbb{R}^+$ *defined as:*

$$q_i(x) = \pi_i(x_{ne^+(i)}) \sum_{x_i \in \mathcal{L}} p(x), \, \forall x \in \mathcal{L}^n$$

also defines a joint distribution. Second, suppose $i \in \mathcal{V}$*, and introduce the two operators, respectively, denoted by* $\mathscr{A}_i$ *and* $\bar{\mathscr{A}_i}$ *and defined for any real-valued function* $u : \mathcal{L}^n \to \mathbb{R}$ *as:*

$$\begin{cases} (\mathscr{A}_i \, u)(x) = \pi_i(x_{ne^+(i)}) \sum_{x_i \in \mathcal{L}} u(x), \, \forall x \in \mathcal{L}^n \\ (\bar{\mathscr{A}_i} \, u)(x) = u(x) - \pi_i(x_{ne^+(i)}) \sum_{x_i \in \mathcal{L}} u(x), \, \forall x \in \mathcal{L}^n \end{cases}$$

Then, one also checks that $\forall i \in \mathcal{V}$*, both* $\mathscr{A}_i$ *and* $\bar{\mathscr{A}_i}$ *are projections (but, generally, not orthogonal ones), in such a way that, one may see a MRF model as the unique joint distribution lying on the intersection of the respective function spaces which are induced by the projections* $\mathscr{A}_i$*,* $\forall i \in \mathcal{V}$*, in such a way that, linear system (10) may rewrite as:*

$$\begin{cases} \mathscr{A}_i \, p \equiv p, \, \forall i \in \mathcal{V} \\ \sum_{x \in \mathcal{L}^n} p(x) = 1 \\ p(x) \geq 0, \, \forall x \in \mathcal{L}^n \end{cases}$$

equivalently, as the unique joint distribution which is projected as 0 *by* $\bar{\mathscr{A}_i}$*,* $\forall i \in \mathcal{V}$*, in such a way that, linear system (10) may also rewrite as:*

$$\begin{cases} \bar{\mathscr{A}_i} \, p \equiv 0, \, \forall i \in \mathcal{V} \\ \sum_{x \in \mathcal{L}^n} p(x) = 1 \\ p(x) \geq 0, \, \forall x \in \mathcal{L}^n \end{cases} \tag{11}$$

With that being said, by Proposition 3, solving MRFMiP (4) amounts to solving the system of linear equations:

$$\begin{cases} p(x) - \pi_i(x_{\mathrm{ne}^+(i)}) \sum_{x_i \in \mathcal{L}} p(x) = 0, \, \forall x \in \mathcal{L}^n, \, \forall i \in \mathcal{V} \\ p_s(x_s) = \sum_{i \in \mathcal{V}/s} \sum_{x_i \in \mathcal{L}} p(x), \, \forall x_s \in \mathcal{L}^{|s|}, \, \forall s \in \mathcal{S} \\ \sum_{x \in \mathcal{L}^n} p(x) = 1 \\ p(x) \geq 0, \, \forall x \in \mathcal{L}^n \end{cases} \tag{12}$$

Obviously, the resolution of MRFMiP (4) by means of linear system of equations (12) is utterly an inefficient enterprise. Therefore, the goal in the remainder of this section is to reformulate such a linear system of equations (12) as a new one writing, solely, in terms of a polynomial-time number of marginal probabilities with respect to a well-chosen hypervertex-set, thereby, setting the stage for its further LP relaxation (see section 5). Thus, let us first denote:

$$\mathcal{S}^+ = \left\{ s \cup \mathrm{ne}^+(i),\, \forall s \in \mathcal{S},\, \forall i \in \mathcal{V} \right\} \tag{13}$$

Convention 3. *For the sake of notational conciseness throughout, we convene that $\forall t \in \mathcal{S}^+$, the notation $t \doteq s \cup ne^+(i)$ is a shortcut for the statement "t is a hypervertex which is originated from the union of hypervertices s and $ne^+(i)$, with s standing for some hypervertex in $\mathcal{S}$, and $ne^+(i)$ standing for the augmented neighborhood of some vertex $i \in \mathcal{V}$".*

Then, instead of (12), consider the following linear system of equations:

$$\begin{cases} p_t(x_t) - \pi_i(x_{\mathrm{ne}^+(i)}) \sum_{x_i \in \mathcal{L}} p_t(x_t) = 0,\, \forall x_t \in \mathcal{L}^{|t|}, \forall t \doteq s \cup \mathrm{ne}^+(i) \in \mathcal{S}^+ \\ p_t(x_t) = \sum_{j \in \mathcal{V}/t} \sum_{x_j \in \mathcal{L}} p(x), \forall t \in \mathcal{S}^+ \\ \sum_{x \in \mathcal{L}^n} p(x) = 1 \\ p(x) \geq 0,\, \forall x \in \mathcal{L}^n \end{cases} \tag{14}$$

One then notes that, in contrast to (12), such a linear system of equations (14) fully expresses in terms of marginal distributions of the form $p_t(\cdot), \forall t \in \mathcal{S}^+$, with p standing for an arbitrary joint distribution. Moreover, the marginal distributions of the form $p_s(\cdot), \forall s \in \mathcal{S}$ may be derived as follows:

$$p_s(x_s) = \sum_{j \in t/s} \sum_{x_j \in \mathcal{L}} p_t(x_t),\, \forall x_s \in \mathcal{L}^{|s|},\, \forall t \doteq s \cup \mathrm{ne}^+(i) \in \mathcal{S}^+$$

Finally, one notes that the joint distribution of MRF $\mathcal{X}$ satisfies linear system of equations (14). Thus, in order to guarantee that the latter exactly solves MRFMiP (4), one needs, moreover, ensure that whatever a joint distribution $p(\cdot)$ satisfying linear system (14), one has:

$$p_s(x_s) = \mathbb{P}(\mathcal{X}_s = x_s),\, \forall x_s \in \mathcal{L}^{|s|}, \forall s \in \mathcal{S}$$

Theorem 3. *Whatever a joint distribution p of which marginals-set with respect to $\mathcal{S}^+$ satisfies linear system (14), one has:*

$$p_s(x_s) = \mathbb{P}(\mathcal{X}_s = x_s),\ \forall x_s \in \mathcal{L}^{|s|}, \forall s \in \mathcal{S}$$

where $\forall s \in \mathcal{S}$, p_s stands for the margin of p with respect to s.

The proof of Theorem 3 is sketched in Appendix .5.

In short, Theorem 3 guarantees that one may completely solve MRFMiP (4) by means of linear system of equations (14) which, in contrast to the original one (12), fully expresses in terms of the set of marginal distributions with respect to $\mathcal{S}^+$. Nevertheless, let us emphasize that the exact resolution of linear system (14) is not more tractable than the one of (14). This is, because, the former imposes on the unknown quantities involved, namely, $p_t(\cdot), \forall t \in \mathcal{S}^+$ to be marginal distributions of which (easy) enforcement is not a straightforward task [27].

5. Efficient linear programming reformulation of MRFMiP (4)

In the previous section, we have shown that one may completely reformulate MRFMiP (4) as a system of linear equations where the unknowns are marginal distributions with respect to $\mathcal{S}^+$. Therefore, in this section, we attempt to take a step further for completely reformulating MRFMiP (4) as a LP which, in contrast, can be solved efficiently.

Therefore, let us begin by introducing:

- the real-valued functions $\theta_t : \mathcal{L}^{|t|} \to \mathbb{R}$, $\forall t \in \mathcal{S}^+$ defined as:

$$\theta_t(x_t) = \frac{-1+L\, \pi_i(x_{\mathrm{ne}^+(i)})}{L^{|t|}},\ \forall x_t \in \mathcal{L}^{|t|},\ \forall t \in \mathcal{S}^+$$

  and please note that, $\forall t \in \mathcal{S}^+$, function $\theta_t(x_t)$ solely depends on $x_{\mathrm{ne}^+(i)}$,

- the signed Boolean functions $\nu_t : \mathcal{L}^{|t|} \to \{-1, 0, +1\}$, $\forall t \in \mathcal{S}^+$ defined as:

$$\forall t \in \mathcal{S}^+,\ \nu_t(x_t) = \begin{cases} 0, & \text{if } \theta_t(x_t) = 0, \\ +1, & \text{if } \theta_t(x_t) > 0, \\ -1, & \text{if } \theta_t(x_t) < 0. \end{cases} \tag{15}$$

- the three sets of integer vectors, respectively, denoted by Λ_t^0, Λ_t^+, and Λ_t^-, and defined as:

$$\forall t \in \mathcal{S}^+, \begin{cases} \Lambda_t^0 = \left\{x_t \in \mathcal{L}^{|t|},\ \text{s.t.}, \theta_t(x_t) = 0\right\}, \\ \Lambda_t^+ = \left\{x_t \in \mathcal{L}^{|t|},\ \text{s.t.}, \theta_t(x_t) > 0\right\}, \\ \Lambda_t^- = \left\{x_t \in \mathcal{L}^{|t|},\ \text{s.t.}, \theta_t(x_t) < 0\right\}. \end{cases}$$

and denote $\Lambda_t^\pm = \Lambda_t^+ \cup \Lambda_t^-$, $\forall t \in \mathcal{S}^+$. Then, consider the following LP:

$$\min \left\{ \tau + \sum_{t \doteq s \cup \text{ne}^+(i) \in \mathcal{S}^+} \sum_{x_t \in \mathcal{L}^{|t|}} \nu_t(x_t) \big(q_t(x_t) - \pi_i(x_{\text{ne}^+(i)}) \sum_{x_i \in \mathcal{L}} q_t(x_t) \big) \right\}$$
$$\begin{cases} \nu_t(x_t)\big(q_t(x_t) - \pi_i(x_{\text{ne}^+(i)}) \sum_{x_i \in \mathcal{L}} q_t(x_t)\big) \geq 0,\ \forall x_t \in \Lambda_t^\pm, \forall t \doteq s \cup \text{ne}^+(i) \in \mathcal{S}^+ \\ q_t(x_t) - \pi_i(x_{\text{ne}^+(i)}) \sum_{x_i \in \mathcal{L}} q_t(x_t) = 0,\ \forall x_t \in \Lambda_t^0, \forall t \doteq s \cup \text{ne}^+(i) \in \mathcal{S}^+ \\ \sum_{x_t \in \mathcal{L}^{|t|}} q_t(x_t) = \tau,\ \forall t \in \mathcal{S}^+ \\ \tau \geq 1 \\ \left\{q_t : \mathcal{L}^{|t|} \to \mathbb{R}^+, \forall t \in \mathcal{S}^+\right\} \text{ is a pseudo-marginals-set with respect to } \mathcal{S}^+ \end{cases} \tag{16}$$

Then, the main findings in this paper regarding the efficient exact resolution of MRFMiP (4) are described in Theorem 4 below.

Theorem 4. *Denote by* $\left(\tau^*, (q_t^*)_{\forall t \in \mathcal{S}^+}\right)$ *the optimal solution of LP (16). Then, one has:*

1. $\tau^* = 1$ *and* $\left\{q_t^*, \forall t \in \mathcal{S}^+\right\}$ *stands for a set of marginal distributions with respect to* $\mathcal{S}^+$*,*
2. $\forall t \doteq s \cup ne^+(i) \in \mathcal{S}^+$*, one has:*

$$\sum_{j \in \mathcal{V}/s} \sum_{x_j \in \mathcal{L}} q_t^*(x_t) = \mathbb{P}(\mathcal{X}_s = x_s),\ \forall x_s \in \mathcal{L}^{|s|}, \forall s \in \mathcal{S} \tag{17}$$

The proof of Theorem 4 is described in Appendix .6.

In a nutshell, Theorem 4 says that one may completely solve hard linear system of equations (14) (over the marginal polytope), hence, MRFMiP (4) by means of LP (16) (over the pseudo-marginal polytope). Moreover, Theorem 4 guarantees that the optimal solution of LP (16) denoted by $\left\{q_t^*, \forall t \in \mathcal{S}^+\right\}$ is an actual set of marginal distributions with respect to $\mathcal{S}^+$, however, it does not guarantee that the latter is originated from the MRF distribution of $\mathcal{X}$, in contrast, its set of marginal distributions with respect to $\mathcal{S}$ denoted by $\left\{q_s^*, \forall t \in \mathcal{S}\right\}$ and defined according to formula (17) above is.

6. Conclusion

We have described a linear programming approach which can solve in an exact and efficient way the MRF marginal inference problem under rather general assumptions both about MRF graph structure, and nature of clique-potentials involved. Let us emphasize that, in order to avoid overloading the present paper with too much technical details, we have omitted discussing further method acceleration. Indeed, an obvious acceleration of the proposed method concerns, for instance, the (logarithmic) reduction of the number of variables in the final LP (16) for bigger values of the number of labels L. Therefore, the latter along with other possible accelerations of the proposed method will be investigated and reported, accordingly, in future papers.

References

References

[1] J. Hammersley, P. Clifford, Markov fields on finite graphs and lattices, Unpublished.

[2] S. L. Lauritzen, Graphical models, Oxford Statistical Science Series, Oxford Science Publications 17.

[3] C. Bishop, Pattern recognition and machine learning, Springer, New York.

[4] M. Wainwright, M. Jordan, Graphical models, exponential families, and variational inference, Foundations and Trends in Machine Learning 1 (1–2) (2008) 1–305.

[5] E. Ising, Beitrag zur Theorie des Ferromagnetismus, Z. Phys. 31 (1) (1925) 253–258.

[6] J. Ashkin, E. Teller, Statistics of two-dimensional lattices with four components, Physical Review 64 (1943) 5–6.

[7] R. Potts, Some Generalized Order-Disorder Transformations, Proc. of the Cambridge Philosophical Society 48 (1952) 106–109.

[8] L. Landau, E. Lifshitz, Statistical physics: Course of theoretical physics, Elsevier 5.

[9] J. Besag, Spatial interaction and the statistical analysis of lattice systems (with discussion), J. Roy. Statist. Soc. Ser. B 36 (1974) 192–236.

[10] S. Geman, D. Geman, Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images, IEEE Trans. Patt. Anal. Mach. Intell. (T-PAMI) 6 (1984) 721–741.

[11] J. Besag, On the statistical analysis of dirty pictures, J. Roy. Statist. Soc. Ser. B 48 (3) (1986) 259–302.

[12] R. Karp, Reducibility among combinatorial problems, Complexity of Computer Computations, Plenum Press, New York (1972) 85–103.

[13] G. Cooper, The computational complexity of probabilistic inference using Bayesian belief networks, Artificial Intelligence 42 (1990) 393–405.

[14] P. Dagum, M. Luby, Approximating probabilistic inference in bayesian belief networks is NP–hard, Artificial Intelligence 60 (1993) 141–153.

[15] Y. Shimony, Finding the maps for belief networks is NP–hard, Aritifical Intelligence 68 (2) (1994) 399–410.

[16] D. Roth, On the hardness of approximate reasoning, Artificial Intelligence 82 (273–302).

[17] D. M. Chickering, Learning Bayesian networks is NP–Complete, Fisher, D. and Lenz, H., editors, Learning from Data: Artificial Intelligence and Statistics, Springer-Verlag 5 (121130).

[18] B. A. Cipra, The ising model is NP–Complete, SIAM News 33 (6).

[19] A. Megretski, Relaxations of quadratic programs in operator theory and system analysis, Systems, approximation, singular integral operators, and related topics 3.

[20] Y. Boykov, O. Veksler, R. Zabih, Fast approximate energy minimization via graph cuts, IEEE Trans. Patt. Anal. Mach. Intell. (T-PAMI) 23 (11) (2001) 1222–1239.

[21] D. Cohen, M. Cooper, P. Jeavons, A. Krokhin, The complexity of soft constraint satisfaction, Artificial Intelligence 170 (2006) 983–1016.

[22] J. Pearl, Reverend bayes on inference engines: A distributed hierarchical approach, Proc. Sec. Nat. Conf. Artif. Intell., AAAI-82, Pittsburgh, PA. Menlo Park, California (1982) 133–136.

[23] J. Yedidia, W. Freeman, Y. Weiss, Understanding Belief Propagation and its generalizations, Exploring Artificial Intelligence in the New Millennium (Chap. 8) (2003) 239–269.

[24] J. Yedidia, Y. Wang, S. Draper, Divide & Concur and difference-map BP decoders for ldpc codes, IEEE Trans. on Information Theory 57 (2) (2011) 786–802.

[25] V. Kolmogorov, M. Wainwright, On the optimality of tree-reweighted max-product message-passing, Uncertainty in Artificial Intelligence (UAI).

[26] V. Kolmogorov, Convergent tree-reweighted message passing for energy minimization, IEEE Trans. Patt. Anal. Mach. Intell. (T-PAMI) 28 (10) (2006) 1568–1583.

[27] M. Wainwright, T. Jaakkola, A. Willsky, Map estimation via agreement on (hyper)trees: Message-passing and linear-programming approaches, IEEE Trans. on Information Theory 51 (11) (2005) 3697–3717.

[28] A. Globerson, T. Jaakkola, Convergent propagation algorithms via oriented trees, Uncertainty in Artificial Intelligence (UAI).

[29] J. Jancsary, G. Matz, Convergent decomposition solvers for tree-reweighted free energies, AISTATS.

[30] N. Komodakis, N. Paragios, G. Tziritas, Mrf energy minimization and beyond via Dual Decomposition, IEEE Trans. Patt. Anal. Mach. Intell. (T-PAMI) 33 (3) (2011) 531–552.

[31] D. Sontag, T. Jaakkola, New outer bounds on the marginal polytope, Advances in Neural Information Processing Systems, MIT Press 21.

[32] D. Sontag, T. Meltzer, A. Globerson, T. Jaakkola, Y. Weiss, Tightening lp relaxations for MAP using Message Passing, Uncertainty in Artificial Intelligence (UAI).

[33] M. Jaggi, Revisiting Frank-Wolfe: Projection-free sparse convex optimization, Int. Conf. Mach. Learning (ICML).

[34] S. Lacoste-Julien, M. Jaggi, On the global linear convergence of Frank-Wolfe optimization variants, Adv. Neur. Inf. Proc. Sys. (NIPS).

[35] D. Belanger, D. Sheldon, A. McCallum, Marginal inference in mrfs using Frank-Wolfe, NIPS Workshop on Greedy Optimization, Frank-Wolfe and Friends.

[36] R. G. Krishnan, S. Lacoste-Julien, D. Sontag, Barrier Frank-Wolfe for Marginal Inference, Adv. Neur. Inf. Proc. Sys. (NIPS) 28.

Appendix .1. Proof of Theorem 1

We show Theorem (1) by construction, using successive projections of an arbitrary function via marginalization. For the sake of clarity, let us put $m = |\text{Front}(\mathcal{K})|$, and assume an arbitrary (e.g.; a lexicographic order) on the elements of $\text{Front}(\mathcal{K})$, in a way which enables to write $\text{Front}(\mathcal{K}) = \{c_1, \ldots, c_m\}$. Next, suppose an arbitrary function $f : \mathcal{L}^n \to \mathbb{R}$, and let us write:

$$f(x) = \frac{f_{c_1}(x_{c_1})}{L^{n-|c_1|}} + \left(f(x) - \frac{f_{c_1}(x_{c_1})}{L^{n-|c_1|}}\right), \forall x \in \mathcal{L}^n$$

where f_{c_1} stands for the margin of f with respect to c_1. Next, define the (residue) function:

$$f^{(1)}(x) = f(x) - \frac{f_{c_1}(x_{c_1})}{L^{n-|c_1|}}, \forall x \in \mathcal{L}^n$$

Then, one checks that:

$$f^{(1)}_{c_1}(x_{c_1}) = f_{c_1}(x_{c_1}) - L^{n-|c_1|}\,\frac{f_{c_1}(x_{c_1})}{L^{n-|c_1|}} = 0, \forall x_{c_1} \in \mathcal{L}^{|c_1|}$$

Now, consider the following function series:

$$f^{(i+1)}(x) = f^{(i)}(x) - \frac{f^{(i)}_{c_{i+1}}(x_{c_{i+1}})}{L^{n-|c_{i+1}|}}, \forall x \in \mathcal{L}^n, \forall i \in \{1, \ldots, m-1\} \tag{.1}$$

where $\forall i \in \{1, \ldots, m-1\}$, $f^{(i)}_{c_{i+1}}$ stands for the margin of $f^{(i)}$ with respect to c_{i+1}, and let us show by induction (with respect to i) the following statement:

$$\forall i \in \{1, \ldots, m\},\ f^{(i)}_{c_j}(x_{c_j}) = 0, \forall x_{c_j} \in \mathcal{L}^{|c_j|}, \forall j \in \{1, \ldots, i\}$$

First of all, we have already shown above that for $i = 1$, one has:

$$f^{(i)}_{c_j}(x_{c_j}) = 0, \forall x_{c_j} \in \mathcal{L}^{|c_j|}, \forall j \in \{1\}$$

Thus, next, suppose that:

$$f^{(i)}_{c_j}(x_{c_j}) = 0, \forall x_{c_j} \in \mathcal{L}^{|c_j|}, \forall j \in \{1, \ldots, i\} \tag{.2}$$

for some $i \in \{1, \ldots, m-1\}$, and show that one has:

$$f^{(i+1)}_{c_j}(x_{c_j}) = 0, \forall x_{c_j} \in \mathcal{L}^{|c_j|}, \forall j \in \{1, \ldots, i+1\} \tag{.3}$$

First, for $j = i+1$, one finds that:

$$f^{(i+1)}_{c_{i+1}}(x_{c_{i+1}}) = f^{(i)}_{c_{i+1}}(x_{c_{i+1}}) - L^{n-|c_{i+1}|}\,\frac{f^{(i)}_{c_{i+1}}(x_{c_{i+1}})}{L^{n-|c_{i+1}|}} = 0, \forall x_{c_{i+1}} \in \mathcal{L}^{|c_{i+1}|}$$

Next, suppose $j \in \{1, \ldots, i\}$. Then, by using formula (.2) above, and the commutativity property of marginalization, one finds that:

$$f^{(i+1)}_{c_j}(x_{c_j}) \quad = \quad f^{(i)}_{c_j}(x_{c_j}) - \frac{\left(f^{(i)}_{c_{i+1}}\right)_{c_j}(x_{c_j})}{L^{n-|c_{i+1}|}} = 0 - \frac{\left(f^{(i)}_{c_j}\right)_{c_{i+1}}(x_{c_j})}{L^{n-|c_{i+1}|}} = -\frac{0}{L^{n-|c_2|}} = 0, \forall x_{c_j} \in \mathcal{L}^{|c_j|}$$

hence, the proof of formula (.3). Next, by putting:

$$\forall x \in \mathcal{L}^n, \begin{cases} u(x) = f^{(m)}(x) \\ v(x) = f(x) - f^{(m)}(x) \end{cases}$$

one finds that f may write as $f = u + v$, such that:

$$\forall c \in \mathrm{Fclos}_\cap(\mathcal{K}), \forall x_c \in \mathcal{L}^{n-|c|}, \begin{cases} u_s(x_c) = f_c(x_c) \\ v_s(x_c) = 0 \end{cases}$$

where $\forall c \in \mathrm{Fclos}_\cap(\mathcal{K})$, f_c, u_c, and v_c respectively stand for the margin of f, the margin of u, and the margin of v with respect to c. Moreover, by mere induction with respect to formula (.1), one finds that there exists constant coefficients ρ_c, $\forall c \in \mathrm{Fclos}_\cap(\mathcal{K})$, such that, u writes as some linear combination of the margins of f with respect to $\mathrm{Fclos}_\cap(\mathcal{K})$ as follows:

$$u(x) = \sum_{c \in \mathrm{Fclos}_\cap(\mathcal{K})} \frac{\rho_c\, f_c(x_c)}{L^{n-|c|}}, \forall x \in \mathcal{L}^n \tag{.4}$$

Now, in order to derive an iterative expression for the coefficients ρ_c, $\forall c \in \text{Fclos}_\cap\big(\mathcal{K}\big)$ in formula (.4) above, then, first by construction, one has:

$$\rho_c = 1,\, \forall c \in \text{Front}\big(\mathcal{K}\big)$$

Second, suppose $c \in \text{Fclos}_\cap\big(\mathcal{K}\big)/\text{Front}\big(\mathcal{K}\big)$. Then, by marginalization of both sides of formula (.4) with respect to all the variables with indices in $\mathcal{V}/c$, and by mere identification (since we have already shown that $u_c = f_c$), one finds that the only margins which must still persist on the right handside of formula (.4) after marginalization with respect to c correspond to all the hypervertex $t \in \text{Fclos}_\cap\big(\mathcal{K}\big)$, such that, $c \subseteq t$, and one derives (since f is assumed to be arbitrary) that:

$$1 = \rho_c + \textstyle\sum_{t\in\text{Fclos}_\cap(\mathcal{K})\ \text{s.t.}\ c\subset t} \rho_t$$

hence:

$$\rho_c = 1 - \textstyle\sum_{t\in\text{Fclos}_\cap(\mathcal{K})\ \text{s.t.}\ c\subset t} \rho_t$$

where the symbol $\subset$ stands for the strict inclusion, which thus terminates the proof of the first part of Theorem 1.

Now, in order to show the second part of Theorem 1, let us put:

$$\bar{f}(x) = f(x) - (\mathscr{O}_\mathcal{K} f)(x) = f(x) - \textstyle\sum_{c\in\text{Fclos}_\cap(\mathcal{K})} \frac{\rho_c\, f_c(x_c)}{L^{n-|c|}},\, \forall x \in \mathcal{L}^n$$

where we have denoted by f_c the margin of f with respect to c, $\forall c \in \text{Fclos}_\cap\big(\mathcal{K}\big)$. Then, on the one hand, we have already shown above that $\forall c \in \text{Fclos}_\cap\big(\mathcal{K}\big)$, the margin of $\bar{f}$ with respect to c denoted by $\bar{f}_c$ is identically equal to 0. On the other hand, suppose a second arbitrary function $g : \mathcal{L}^n \to \mathbb{R}$, and denote $\forall c \in \text{Fclos}_\cap\big(\mathcal{K}\big)$ by g_c the margin of g with respect to c. On then has:

$$\begin{aligned}\textstyle\sum_{x\in\mathcal{L}^n} \bar{f}(x)\,(\mathscr{O}_\mathcal{K} g)(x) \quad &= \quad \textstyle\sum_{x\in\mathcal{L}^n} \bar{f}(x)\,\big(\sum_{c\in\text{Fclos}_\cap(\mathcal{K})} \frac{\rho_c\, g_c(x_c)}{L^{n-|c|}}\big)\\ &= \quad \textstyle\sum_{c\in\text{Fclos}_\cap(\mathcal{K})} \sum_{x_c\in\mathcal{L}^{|c|}} \frac{\rho_c\, \bar{f}_c(x_c)\, g_c(x_c)}{L^{n-|c|}} = 0\end{aligned}$$

and since both f and g are arbitrary, one concludes that the operator $\mathscr{O}_\mathcal{K}$ is an orthogonal projection, thus, establishing the proof of the second part of Theorem 1.

Appendix .2. Proof of Theorem 2

We will show Theorem 2 by construction, in the spirit of the proof of Theorem 1. Therefore, suppose $\big\{u_c, \forall c \in \mathcal{K}\big\}$ is a pseudo-marginals-set. Then, we will iteratively construct a function $\tilde{u} : \mathcal{L}^n \to \mathbb{R}$ which ultimately develops as:

$$\tilde{u}(x) = \sum_{c\in\text{Fclos}_\cap(\mathcal{K})} \frac{\beta_c\, u_c(x_c)}{L^{n-|c|}},\, \forall x \in \mathcal{L}^n$$

where $\beta_c, \forall c \in \text{Fclos}_\cap\big(\mathcal{K}\big)$ stand for some real coefficients and, such that, the set of margins of $\tilde{u}$ with respect to $\mathcal{K}$ coincides with $\big\{u_c, \forall c \in \mathcal{K}\big\}$. Therefore, let us first put $m = |\text{Front}\big(\mathcal{K}\big)|$, and assume an arbitrary (e.g.; a lexicographic) order on the elements of $\text{Front}\big(\mathcal{K}\big)$, in a way which enables to write $\text{Front}(\mathcal{K}) = \{c_1, \ldots, c_m\}$. Then, one starts by defining the following function:

$$\tilde{u}^{(1)}(x) = \frac{u_{c_1}(x_{c_1})}{L^{n-|c_1|}}, \forall x \in \mathcal{L}^n$$

and one easily checks that:

$$\tilde{u}^{(1)}_{c_1}(x_{c_1}) = L^{n-|c_1|}\left(\frac{u_{c_1}(x_{c_1})}{L^{n-|c_1|}}\right) = u_{c_1}(x_{c_1}), \forall x_{c_1} \in \mathcal{L}^{|s_1|}$$

where $\tilde{u}^{(1)}_{s_1}$ stands for the margin of $\tilde{u}^{(1)}$ with respect to s_1. Now, assume that $i \in \{1, \ldots, m-1\}$, and let us iteratively construct the function $\tilde{u}^{(i+1)}$ as follows:

$$\tilde{u}^{(i+1)}(x) = \tilde{u}^{(i)}(x) + \frac{u_{c_{i+1}}(x_{c_{i+1}})}{L^{n-|c_{i+1}|}} - \frac{\tilde{u}^{(i)}_{c_{i+1}}(x_{c_{i+1}})}{L^{n-|c_{i+1}|}}, \forall x \in \mathcal{L}^n \tag{.5}$$

where $\tilde{u}^{(i)}_{c_{i+1}}$ stands for the margin of $\tilde{u}^{(i)}$ with respect to c_{i+1}, then show, by induction (with respect to i), the following identities:

$$\tilde{u}^{(i)}_{c_j}(x_{c_j}) = u_{c_j}(x_{c_j}), \forall x_{c_j} \in \mathcal{L}^{|j|}, \forall j \in \{1, \ldots, i\}, \forall i \in \{1, \ldots, m\} \tag{.6}$$

First, for $i = 1$, we have already shown that:

$$\tilde{u}^{1}_{c_1}(x_{c_1}) = u_{c_1}(x_{c_1}), \forall x_{c_j} \in \mathcal{L}^{|c_j|}, \forall j \in \{1\}$$

Next, suppose that one has for some $i \in \{1, \ldots, m-1\}$:

$$\tilde{u}^{(i)}_{c_j}(x_{c_j}) = u_{c_j}(x_{c_j}), \forall x_{c_j} \in \mathcal{L}^{|c_j|}, \forall j \in \{1, \ldots, i\}$$

and show that one has:

$$\tilde{u}^{(i+1)}_{c_j}(x_{c_j}) = u_{c_j}(x_{c_j}), \forall x_{c_j} \in \mathcal{L}^{|c_j|}, \forall j \in \{1, \ldots, i+1\}$$

First, by assuming that $j = i + 1$, one finds that:

$$\tilde{u}^{(i+1)}_{c_{i+1}}(x_{c_{i+1}}) = \tilde{u}^{(i)}_{c_{i+1}}(x_{c_{i+1}}) + u_{c_{i+1}}(x_{c_{i+1}}) - \tilde{u}^{(i)}_{c_{i+1}}(x_{c_{i+1}}) = u_{c_{i+1}}(x_{c_{i+1}}), \forall x_{c_{i+1}} \in \mathcal{L}^{|c_{i+1}|}$$

Second, by assuming that $j \in \{1, \ldots, i\}$, then by marginalization of both sides of formula (.5) with respect to all the variables with indices in $\mathcal{V}/x_{c_{i+1}}$, finally by using the assumption that the set $\{u_c, \forall c \in \mathcal{K}\}$ defines a pseudo-marginals-set, one finds that:

$$\tilde{u}^{(i+1)}_{c_j}(x_{c_j}) = \tilde{u}^{(i)}_{c_j}(x_{c_j}) + \frac{u_{c_{i+1}\cap c_j}(x_{c_{i+1}\cap c_j})}{L^{|c_{i+1}\cap c_j|}} - \frac{\tilde{u}^{(i)}_{c_{i+1}\cap c_j}(x_{c_{i+1}\cap c_j})}{L^{|c_{i+1}\cap c_j|}}, \forall x_{c_j} \in \mathcal{L}^{|c_j|}$$

where $\tilde{u}^{(i+1)}_{c_j}$ stands for the margin of function $\tilde{u}^{(i+1)}$ with respect to c_j, $u_{c_{i+1}\cap c_j}$ stands for the margin of pseudo-marginal $u_{c_{i+1}}$ with respect to $c_{i+1} \cap c_j$, and $\tilde{u}^{(i)}_{c_{i+1}\cap c_j}$ stands for the margin of function $\tilde{u}^{(i)}$ with respect to $c_{i+1} \cap c_j$. Next, since by assumption, one has: $\tilde{u}^{(i)}_{c_j} = u_{c_j}$, thus one also has $\tilde{u}^{(i)}_{c_{i+1}\cap c_j} = u^{(i)}_{c_{i+1}\cap c_j}$, and one finally derives:

$$\tilde{u}^{(i+1)}_{c_j}(x_{c_j}) = u_{c_j}(x_{c_j}), \forall x_{c_j} \in \mathcal{L}^{|c_j|}$$

hence the proof of identities (.6).

Now, by assuming that $i = m$, thus one has constructed a function:

$$\tilde{u}^{(m)}(x) = \sum_{c \in \mathrm{Fclos}_{\cap}(\mathcal{K})} \frac{\beta_c \, u_c(x_c)}{L^{n-|c|}}, \, \forall x \in \mathcal{L}^n$$

where $\beta_c, \forall c \in \mathrm{Fclos}_{\cap}(\mathcal{K})$ stand for some real coefficients, such that, $\tilde{u}_c^{(m)} = u_c, \forall c \in \mathrm{Front}(\mathcal{K})$, implying that one has $\tilde{u}_c^{(m)} \equiv u_c, \forall c \in \mathcal{K}$. Next, one proceeds in exactly the same way as we have done in the proof of Theorem 1 for deriving a recursive formula for the coefficients $\beta_c, \forall c \in \mathrm{Fclos}_{\cap}(\mathcal{K})$, and one finds that:

$$\beta_c = \rho_c, \forall c \in \mathrm{Fclos}_{\cap}(\mathcal{K})$$

where the coefficients $\rho_c, \forall c \in \mathrm{Fclos}_{\cap}(\mathcal{K})$ are given by formula (7). Finally, by using Theorem 1, one derives that whatever a function $v : \mathcal{L}^n \to \mathbb{R}$, the function defined as:

$$\tilde{u}(x) = \Big(v(x) - \sum_{c \in \mathrm{Fclos}_{\cap}(\mathcal{K})} \frac{\rho_c \, v_c(x_c)}{L^{n-|c|}}\Big) + \sum_{c \in \mathrm{Fclos}_{\cap}(\mathcal{K})} \frac{\rho_c \, u_c(x_c)}{L^{n-|c|}}, \, \forall x \in \mathcal{L}^n$$

where $\forall c \in \mathrm{Fclos}_{\cap}(\mathcal{K})$, v_c stands for the margin of v with respect to c, satisfies that $\tilde{u}_c = u_c, \forall c \in \mathcal{K}$, thus, establishing the proof of Theorem 2.

Appendix .3. Proof of Proposition 1

Suppose the local functions $h_c : \mathcal{L}^{|c|} \to \mathbb{R}, \forall c \in \mathcal{K}$, and define the higher-order function $h : \mathcal{L}^n \to \mathbb{R}$ as:

$$h(x) = \textstyle\sum_{c \in \mathcal{K}} h_c(x_c), \, \forall x \in \mathcal{L}^n$$

Next, suppose an arbitrary function $v : \mathcal{L}^n \to \mathbb{R}$, moreover, $\forall c \in \mathrm{Front}(\mathcal{K})$, denote by v_c the margin of v with respect to c. Obviously, one has: $\mathscr{O}_{\mathrm{Front}(\mathcal{K})} v_c \equiv v_c$, $\forall c \in \mathrm{Front}(\mathcal{K})$, in such a way that, one may write:

$$\begin{aligned}
\langle h, v \rangle &= \textstyle\sum_{x \in \mathcal{L}^n} h(x) \, v(x) = \sum_{c \in \mathcal{K}} \sum_{x_c \in \mathcal{L}^{|c|}} h_c(x_c) \, v_c(x_c) \\
&= \textstyle\sum_{c \in \mathcal{K}} \langle h_c, \mathscr{O}_{\mathrm{Front}(\mathcal{K})} v_c \rangle = \sum_{c \in \mathcal{K}} \langle \mathscr{O}_{\mathrm{Front}(\mathcal{K})} h_c, v_c \rangle \\
&= \textstyle\sum_{c \in \mathcal{K}} \langle \mathscr{O}_{\mathrm{Front}(\mathcal{K})} h_c, v \rangle = \langle \mathscr{O}_{\mathrm{Front}(\mathcal{K})} \big(\sum_{c \in \mathcal{K}} h_c\big), v \rangle = \langle \mathscr{O}_{\mathrm{Front}(\mathcal{K})} h, v \rangle
\end{aligned}$$

hence $\langle h - \mathscr{O}_{\mathrm{Front}(\mathcal{K})} h, v \rangle = 0$. But, since v is assumed to be arbitrary, one finally derives that $\mathscr{O}_{\mathrm{Front}(\mathcal{K})} h \equiv h$. The converse is obvious by definition of $\mathscr{M}_{\mathcal{K}}$, hence the proof of Proposition 1.

Appendix .4. Proof of Proposition 3

Clearly, showing Proposition 3 amounts to showing that linear system (10) is equivalent to the following one:

$$\begin{cases}
p(x) - \pi_i(x_{\mathrm{ne}^+(i)}) \sum_{x_i \in \mathcal{L}} p(x) = 0, \, \forall x \in \mathcal{L}^n, \, \forall i \in \mathcal{V} \\
\sum_{x \in \mathcal{L}^n} p(x) = 1 \\
p(x) > 0, \, \forall x \in \mathcal{L}^n
\end{cases} \tag{.7}$$

because:

- at least, the solution which is given by $p(x) = \mathbb{P}(\mathcal{X} = x), \forall x \in \mathcal{L}^n$ satisfies linear system (.7),
- moreover, any solution p of linear system (.7) is a strictly nonnegative joint distribution satisfying the conditional property, which also means that p is a MRF distribution.

and since any MRF distribution is fully defined by its neighborhood-wise conditional probabilities, one concludes that the solution of linear system (.7) is unique and coincides with $\mathbb{P}(\mathcal{X} = x), \forall x \in \mathcal{L}^n$.

Therefore, let us assume the opposite of the latter claim by supposing that there exists a solution $p(x), \forall x \in \mathcal{L}^n$ of linear system (10), such that, there exists $x^0 \in \mathcal{L}^n$ with $p(x^0) = 0$, and let us show that this leads to a contradiction. Indeed, since $p(x^0)$ has to satisfy the identity:

$$p(x^0) = \pi_i(x^0_{\mathrm{ne}^+(i)}) \sum_{x^0_i \in \mathcal{L}} p(x^0),\ \forall i \in \mathcal{V}$$

one derives:

$$p(x^0_1, \ldots, x^0_{i-1}, x_i, x^0_{i+1}, \ldots, x^0_n) = 0,\ \forall x_i \in \mathcal{L},\ \forall i \in \mathcal{V}$$

Next, for the sake of notational convenience, let us convene that whatever a hypervertex c of $\mathcal{G}$, and whatever an integer vector $x_c \in \mathcal{L}^{|c|}$, the integer vector denoted by $y := (x^0_{\mathcal{V}/c}, x_c) \in \mathcal{L}^n$ is given by:

$$\forall i \in \mathcal{V},\ y_i = \begin{cases} x^0_i, \text{ if } i \in \mathcal{V}/c \\ x_i, \text{ if } i \in c \end{cases}$$

Then, suppose an arbitrary hypervertex c of $\mathcal{G}$, and let us show, by induction with respect to the size $|c|$ (in terms of number of vertices) of c, that one has $p(x) = 0, \forall x \in \mathcal{L}^n$. First, assume that $|c| = 1$, then we have already shown above that one has:

$$p(x^0_{\mathcal{V}/c}, x_c) = 0,\ \forall x_c \in \mathcal{L}^1,\ \forall c \in \mathcal{V} \tag{.8}$$

Second, assume that $|c| = 2$, and put $c = \{i_1, i_2\}$ with i_1 and i_2 standing for two distinct vertices of $\mathcal{G}$, moreover, assume an integer vector $(x^0_{\mathcal{V}/c}, x_c) \in \mathcal{L}^n$, and put $z = (x^0_{\mathcal{V}/\{i_2\}}, x_{i_1})$. Thus, first, by using formula (.8), one has $p(z) = 0$, moreover, since one also has:

$$p(z) = \pi_i(z_{\mathrm{ne}^+(i)}) \sum_{z_i \in \mathcal{L}} p(z),\ \forall i \in \mathcal{V}$$

one finds that $p(x^0_{\mathcal{V}/c}, x_c) = 0$, finally, c and x_c are assumed to be arbitrary, one concludes that whatever a hypervertex c of $\mathcal{G}$, such that, $|c| = 2$, and whatever an integer vector $(x^0_{\mathcal{V}/c}, x_c) \in \mathcal{L}^n$, one has $p(x^0_{\mathcal{V}/c}, x_c) = 0$. One repeats the latter procedure with respect to an arbitrary hypervertex c of $\mathcal{G}$, such that, $|c| = k$, $k = 3, 4, \ldots, n$, and one finds, successively, for $k = 3, \ldots, n$, that whatever a hypervertex c of $\mathcal{G}$, such that, $|c| = k$, one has:

$$p(x^0_{\mathcal{V}/c}, x_c) = 0,\ \forall x_c \in \mathcal{L}^{|c|}$$

In particular, for $k = n$, one finds that $p(x) = 0$, $\forall x \in \mathcal{L}^n$, hence, $\sum_{x \in \mathcal{L}^n} p(x) = 0 \neq 1$ which is a absurdity, hence, finally the proof of Proposition 3.

Appendix .5. Proof of Theorem 3

We need the following lemma which is shown in Appendix .5.1.

Lemma 1. *Suppose a higher-order function $h : \mathcal{L}^n \longrightarrow \mathbb{R}$ writing as:*

$$h(x) = \textstyle\sum_{s \in \mathcal{S}} h_s(x_s),\, \forall x \in \mathcal{L}^n$$

where $\forall s \in \mathcal{S}$, $h_s : \mathcal{L}^{|s|} \longrightarrow \mathbb{R}$ stands for an arbitrary local function (which is not to be confused here with the margin of h with respect to s). Then, one has:

$$\begin{aligned} \mathbb{E}\big[h(\mathcal{X})\big] &= \max_{\{w_t : \mathcal{L}^{|t|} \longrightarrow \mathbb{R}, \forall t \in \mathcal{S}^+\}} \inf_{x \in \mathcal{L}^n} \Big\{ h_s(x_s) + \sum_{t \doteq s \cup ne^+(i) \in \mathcal{S}^+} \big(w_t(x_t) - \sum_{x_i \in \mathcal{L}} \pi_i(x_{ne^+(i)})\, w_t(x_t) \big) \Big\} \\ &= \min_{\{w_t : \mathcal{L}^{|t|} \longrightarrow \mathbb{R}, \forall t \in \mathcal{S}^+\}} \sup_{x \in \mathcal{L}^n} \Big\{ h_s(x_s) + \sum_{t \doteq s \cup ne^+(i) \in \mathcal{S}^+} \big(w_t(x_t) - \sum_{x_i \in \mathcal{L}} \pi_i(x_{ne^+(i)})\, w_t(x_t) \big) \Big\} \end{aligned}$$

Then, in order to show Theorem 3, let $\mathscr{M}_{\mathcal{S}}$ stand for the ortho-marginal space with respect to $\mathcal{S}$ which is given by:

$$\mathscr{M}_{\mathcal{S}} = \big\{ f : \mathcal{L}^n \longrightarrow \mathbb{R}, \text{s.t.}, \mathscr{O}_{\mathcal{S}} f \equiv f \big\}$$

and suppose an arbitrary higher-order function $h \in \mathscr{M}_{\mathcal{S}}$ which, by Proposition 1, may write as:

$$h(x) = \textstyle\sum_{s \in \mathcal{S}} h_s(x_s),\, \forall x \in \mathcal{L}^n$$

where $\forall s \in \mathcal{S}$, $h_s : \mathcal{L}^{|s|} \longrightarrow \mathbb{R}$ stands for an arbitrary local function. After that, consider the following two LPs:

$$\min_p \Big\{ \textstyle\sum_{x \in \mathcal{L}^n} h(x)\, p(x) \Big\}$$
$$\begin{cases} p_t(x_t) - \pi_i(x_{\mathrm{ne}^+(i)}) \sum_{x_i \in \mathcal{L}} p_t(x_t) = 0,\, \forall x_t \in \mathcal{L}^{|t|}, \forall t \doteq s \cup \mathrm{ne}^+(i) \in \mathcal{S}^+ \\ p_t(x_t) = \sum_{j \in \mathcal{V}/t} \sum_{x_j \in \mathcal{L}} p(x), \forall t \in \mathcal{S}^+ \\ \sum_{x \in \mathcal{L}^n} p(x) = 1 \\ p(x) \geq 0,\, \forall x \in \mathcal{L}^n \end{cases} \tag{.9}$$

and

$$\max_p \Big\{ \textstyle\sum_{x \in \mathcal{L}^n} h(x)\, p(x) \Big\}$$
$$\begin{cases} p_t(x_t) - \pi_i(x_{\mathrm{ne}^+(i)}) \sum_{x_i \in \mathcal{L}} p_t(x_t) = 0,\, \forall x_t \in \mathcal{L}^{|t|}, \forall t \doteq s \cup \mathrm{ne}^+(i) \in \mathcal{S}^+ \\ p_t(x_t) = \sum_{j \in \mathcal{V}/t} \sum_{x_j \in \mathcal{L}} p(x), \forall t \in \mathcal{S}^+ \\ \sum_{x \in \mathcal{L}^n} p(x) = 1 \\ p(x) \geq 0,\, \forall x \in \mathcal{L}^n \end{cases} \tag{.10}$$

Then, by first assuming LP (.9), one finds that its dual writes as:

$$\max_{\mu,\, w_t : \mathcal{L}^{|t|} \longrightarrow \mathbb{R}, \forall t \in \mathcal{S}^+} \{ \mu \}$$
$$\begin{cases} \mu \leq h(x) + \sum_{t \doteq s \cup \mathrm{ne}^+(i) \in \mathcal{S}^+} \big(w_t(x_t) - \sum_{x_i \in \mathcal{L}} \pi_i(x_{\mathrm{ne}^+(i)})\, w_t(x_t) \big),\, \forall x \in \mathcal{L}^n \\ w_t(x_t) \in \mathcal{R},\, \forall x_t \in \mathcal{L}^{|t|},\, \forall t \in \mathcal{S}^+ \\ \mu \in \mathcal{R} \end{cases} \tag{.11}$$

in such a way that, by noticing that the optimal value of μ in dual LP (.11) in terms of functions h and $w_t, \forall t \in \mathcal{S}^+$ is given by:

$$\inf_{x \in \mathcal{L}^n} \left\{ h(x) + \sum_{t \doteq s \cup \mathrm{ne}^+(i) \in \mathcal{S}^+} \big(w_t(x_t) - \sum_{x_i \in \mathcal{L}} \pi_i(x_{\mathrm{ne}^+(i)})\, w_t(x_t) \big) \right\}$$

one may rewrite dual LP (.11), accordingly, as:

$$\max_{w_t : \mathcal{L}^{|t|} \longrightarrow \mathbb{R}, \forall t \in \mathcal{S}^+} \inf_{x \in \mathcal{L}^n} \left\{ h(x) + \sum_{t \doteq s \cup \mathrm{ne}^+(i) \in \mathcal{S}^+} \big(w_t(x_t) - \sum_{x_i \in \mathcal{L}} \pi_i(x_{\mathrm{ne}^+(i)})\, w_t(x_t) \big) \right\}$$

Second, by assuming LP (.10), and by proceeding in exactly the same way as we have done for deriving the dual of LP (.9), one finds that the dual of LP (.10) may write as:

$$\min_{w_t : \mathcal{L}^{|t|} \longrightarrow \mathbb{R}, \forall t \in \mathcal{S}^+} \sup_{x \in \mathcal{L}^n} \left\{ h(x) + \sum_{t \doteq s \cup \mathrm{ne}^+(i) \in \mathcal{S}^+} \big(w_t(x_t) - \sum_{x_i \in \mathcal{L}} \pi_i(x_{\mathrm{ne}^+(i)})\, w_t(x_t) \big) \right\}$$

Therefore, by Lemma 1, one derives that both LP (.9) and LP (.10) achieve the same optimal objective value which is equal to $\mathbb{E}\big[h(\mathcal{X})\big]$. Moreover, since in particular, the (MRF) solution which is given by $p(x) = \mathbb{P}\big(\mathcal{X} = x\big), \forall x \in \mathcal{L}^n$ is a feasible solution of LP (.9) (hence, a feasible solution of LP (.10) too), one derives that whatever a feasible solution p' of LP (.9), one has:

$$\sum_{x \in \mathcal{L}^n} h(x)\, p'(x) = \mathbb{E}\big[h(\mathcal{X})\big]$$

But, since h is assumed to be an arbitrary higher function in $\mathscr{M}_{\mathcal{S}}$, one concludes that whatever a real-valued function $f : \mathcal{L}^n \longrightarrow \mathbb{R}$, one has:

$$\big\langle \mathscr{O}_{\mathcal{S}} f, p' - p \big\rangle = \big\langle f, \mathscr{O}_{\mathcal{S}}(p' - p) \big\rangle = 0$$

where $\langle \cdot, \cdot \rangle$ stands for the scalar product, and one derives:

$$\big(\mathscr{O}_{\mathcal{S}} p'\big)(x) = \big(\mathscr{O}_{\mathcal{S}} p\big)(x),\ \forall x \in \mathcal{L}^n$$

which, by Proposition 2, is equivalent to:

$$p'_s(x_s) = \mathbb{P}\big(\mathcal{X}_s = x_s\big),\ \forall x_s \in \mathcal{L}^{|s|},\ \forall s \in \mathcal{S}$$

thus, establishing the proof of Theorem 3.

Appendix .5.1. Proof of Lemma 1

We need the following lemma which is shown in Appendix .5.2.

Lemma 2. *Whatever a real-valued function $f : \mathcal{L}^n \longrightarrow \mathbb{R}$, there exists n real-valued functions $v_i^* : \mathcal{L}^n \longrightarrow \mathbb{R}$, $\forall i \in \mathcal{V}$, such that:*

$$\begin{aligned}
\mathbb{E}\big[f(\mathcal{X})\big] &= f(x) + \sum_{i \in \mathcal{V}} \big(v_i^*(x) - \sum_{x_i \in \mathcal{L}} \pi_i(x_{ne^+(i)})\, v_i^*(x) \big),\ \forall x \in \mathcal{L}^n \\
&= \max_{\{v_i : \mathcal{L}^n \longrightarrow \mathbb{R}, \forall i \in \mathcal{V}\}} \inf_{x \in \mathcal{L}^n} \Big\{ f(x) + \sum_{i \in \mathcal{V}} \big(v_i(x) - \sum_{x_i \in \mathcal{L}} \pi_i(x_{ne^+(i)})\, v_i(x) \big) \Big\} \\
&= \min_{\{v_i : \mathcal{L}^n \longrightarrow \mathbb{R}, \forall i \in \mathcal{V}\}} \sup_{x \in \mathcal{L}^n} \Big\{ f(x) + \sum_{i \in \mathcal{V}} \big(v_i(x) - \sum_{x_i \in \mathcal{L}} \pi_i(x_{ne^+(i)})\, v_i(x) \big) \Big\}
\end{aligned}$$

Then, in order to show Lemma 1, suppose an arbitrary higher-order function $h : \mathcal{L}^n \longrightarrow \mathbb{R}$ writing as:

$$h(x) = \textstyle\sum_{s\in\mathcal{S}} h_s(x_s),\ \forall x \in \mathcal{L}^n$$

where $\forall s \in \mathcal{S}$, $h_s : \mathcal{L}^{|s|} \longrightarrow \mathbb{R}$ stands for an arbitrary local function. Next, suppose $s \in \mathcal{S}$. Then, on the one hand, by Lemma 2, one finds that there exists n real-valued functions $v_i^{(s*)} : \mathcal{L}^n \longrightarrow \mathbb{R}$, $\forall i \in \mathcal{V}$, such that, one has:

$$\mathbb{E}\big[h_s(\mathcal{X}_s)\big] \quad = \quad h_s(x_s) + \sum_{i\in\mathcal{V}} \big(v_i^{(s*)}(x) - \sum_{x_i\in\mathcal{L}} \pi_i(x_{\mathrm{ne}^+(i)})\, v_i^{(s*)}(x) \big),\ \forall x \in \mathcal{L}^n$$

hence:

$$\mathbb{E}\big[h_s(\mathcal{X}_s)\big] - h_s(x_s) = \textstyle\sum_{i\in\mathcal{V}} \big(v_i^{(s*)}(x) - \sum_{x_i\in\mathcal{L}} \pi_i(x_{\mathrm{ne}^+(i)})\, v_i^{(s*)}(x) \big),\ \forall x \in \mathcal{L}^n$$

and one finds that the function $u^{(s*)}(x) : \mathcal{L}^n \longrightarrow \mathbb{R}$ defined as:

$$u^{(s*)}(x) = \textstyle\sum_{i\in\mathcal{V}} \big(v_i^{(s*)}(x) - \sum_{x_i\in\mathcal{L}} \pi_i(x_{\mathrm{ne}^+(i)})\, v_i^{(s*)}(x) \big),\ \forall x \in \mathcal{L}^n$$

solely depends on x_s, in such a way that, $\forall x \in \mathcal{L}^n$, $u^{(s*)}(x)$ is also equal to the average of the set $\big\{u^{(s*)}(y), \forall y \in \mathcal{L}^n, \text{s.t., } y_{s\cup\mathrm{ne}^+(i)} = x_{s\cup\mathrm{ne}^+(i)}\big\}$, in other words, $\forall x \in \mathcal{L}^n$, one also has:

$$\begin{aligned}
u^{(s*)}(x) \quad &= \quad \tfrac{1}{L^{n-|s\cup\mathrm{ne}^+(i)|}} \textstyle\sum_{j\in\mathcal{V}/(s\cup\mathrm{ne}^+(i))} \sum_{x_j\in\mathcal{L}} u^{(s*)}(x) \\
&= \quad \tfrac{1}{L^{n-|s\cup\mathrm{ne}^+(i)|}} \textstyle\sum_{j\in\mathcal{V}/(s\cup\mathrm{ne}^+(i))} \sum_{x_j\in\mathcal{L}} \sum_{i\in\mathcal{V}} \big(v_i^{(s*)}(x) - \sum_{x_i\in\mathcal{L}} \pi_i(x_{\mathrm{ne}^+(i)})\, v_i^{(s*)}(x) \big) \\
&= \quad \tfrac{1}{L^{n-|s\cup\mathrm{ne}^+(i)|}} \textstyle\sum_{i\in\mathcal{V}} \Big(\sum_{j\in\mathcal{V}/(s\cup\mathrm{ne}^+(i))} \sum_{x_j\in\mathcal{L}} v_i^{(s*)}(x) - \\
&\qquad\qquad \textstyle\sum_{j\in\mathcal{V}/(s\cup\mathrm{ne}^+(i))} \sum_{x_j\in\mathcal{L}} \big(\sum_{x_i\in\mathcal{L}} \pi_i(x_{\mathrm{ne}^+(i)})\, v_i^{(s*)}(x) \big) \Big) \\
&= \quad \textstyle\sum_{i\in\mathcal{V}} \Big(\frac{\sum_{j\in\mathcal{V}/(s\cup\mathrm{ne}^+(i))} \sum_{x_j\in\mathcal{L}} v_i^{(s*)}(x)}{L^{n-|s\cup\mathrm{ne}^+(i)|}} - \sum_{x_i\in\mathcal{L}} \pi_i(x_{\mathrm{ne}^+(i)}) \frac{\sum_{j\in\mathcal{V}/(s\cup\mathrm{ne}^+(i))} \sum_{x_j\in\mathcal{L}} v_i^{(s*)}(x)}{L^{n-|s\cup\mathrm{ne}^+(i)|}} \Big) \\
&= \quad \textstyle\sum_{i\in\mathcal{V}} \big(w_i^{(s*)}(x_{s\cup\mathrm{ne}^+(i)}) - \sum_{x_i\in\mathcal{L}} \pi_i(x_{\mathrm{ne}^+(i)})\, w_i^{(s*)}(x_{s\cup\mathrm{ne}^+(i)}) \big)
\end{aligned}$$

where we have denoted:

$$w_i^{(s*)}(x_{s\cup\mathrm{ne}^+(i)}) = \tfrac{1}{L^{n-|s\cup\mathrm{ne}^+(i)|}} \textstyle\sum_{j\in\mathcal{V}/(s\cup\mathrm{ne}^+(i))} \sum_{x_j\in\mathcal{L}} v_i^{(s*)}(x),\ \forall x_{s\cup\mathrm{ne}^+(i)} \in \mathcal{L}^{|s\cup\mathrm{ne}^+(i)|}$$

Thus, one also has:

$$u^{(s*)}(x) \quad = \quad \textstyle\sum_{i\in\mathcal{V}} \big(w_i^{(s*)}(x_{s\cup\mathrm{ne}^+(i)}) - \sum_{x_i\in\mathcal{L}} \pi_i(x_{\mathrm{ne}^+(i)})\, w_i^{(s*)}(x_{s\cup\mathrm{ne}^+(i)}) \big),\ \forall x \in \mathcal{L}^n$$

and one finds that there exists n real-valued functions $w_i^{(s*)} : \mathcal{L}^{|s\cup\mathrm{ne}^+(i))|} \longrightarrow \mathbb{R}$, $\forall i \in \mathcal{V}$, such that, $\forall x \in \mathcal{L}^n$, one has:

$$\mathbb{E}\big[h_s(\mathcal{X}_s)\big] = h_s(x_s) + \sum_{i\in\mathcal{V}} \big(w_i^{(s*)}(x_{s\cup\mathrm{ne}^+(i)}) - \sum_{x_i\in\mathcal{L}} \pi_i(x_{\mathrm{ne}^+(i)})\, w_i^{(s*)}(x_{s\cup\mathrm{ne}^+(i)})\big) \qquad (.12)$$

Next, by using the identity $\mathbb{E}\big[h(\mathcal{X})\big] = \sum_{s\in\mathcal{S}} \mathbb{E}\big[h_s(\mathcal{X}_s)\big]$, and by adding up, side to side, equations (.12) with respect to all $s \in \mathcal{S}$, one finds that $\forall x \in \mathcal{L}^n$:

$$\begin{aligned}
\mathbb{E}\big[h(\mathcal{X})\big] \quad &= \quad h(x) + \textstyle\sum_{s\in\mathcal{S}} \Big(\sum_{i\in\mathcal{V}} \big(w_i^{(s*)}(x_{s\cup\mathrm{ne}^+(i)}) - \sum_{x_i\in\mathcal{L}} \pi_i(x_{\mathrm{ne}^+(i)})\, w_i^{(s*)}(x_{s\cup\mathrm{ne}^+(i)}) \big) \Big) \\
&= \quad h(x) + \textstyle\sum_{i\in\mathcal{V}} \Big(\big(\sum_{s\in\mathcal{S}} w_i^{(s*)}(x_{s\cup\mathrm{ne}^+(i)}) \big) - \sum_{x_i\in\mathcal{L}} \pi_i(x_{\mathrm{ne}^+(i)}) \big(\sum_{s\in\mathcal{S}} w_i^{(s*)}(x_{s\cup\mathrm{ne}^+(i)}) \big) \Big)
\end{aligned} \qquad (.13)$$

this is on the one hand. On the other hand, again, by Lemma 2, one finds that there exists n real-valued functions $v_i^* : \mathcal{L}^n \longrightarrow \mathbb{R}$, $\forall i \in \mathcal{V}$, such that:

$$\mathbb{E}\big[h(\mathcal{X})\big] \quad = \quad h(x) + \sum_{i \in \mathcal{V}} \big(\, v_i^*(x) - \sum_{x_i \in \mathcal{L}} \pi_i(x_{\mathrm{ne}^+(i)})\, v_i^*(x) \,\big) \tag{.14}$$

in such a way that, by identification of formulas (.13) and (.14), one finds that there exists n higher-order functions $v_i^* : \mathcal{L}^n \longrightarrow \mathbb{R}$, $\forall i \in \mathcal{V}$ of the form:

$$v_i^*(x) = \textstyle\sum_{s \in \mathcal{S}} w_i^{(s*)}(x_{s \cup \mathrm{ne}^+(i)}),\, \forall x \in \mathcal{L}^n,\, \forall i \in \mathcal{V}$$

such that:

$$\mathbb{E}\big[h(\mathcal{X})\big] \quad = \quad h(x) + \sum_{i \in \mathcal{V}} \big(\, v_i^*(x) - \sum_{x_i \in \mathcal{L}} \pi_i(x_{\mathrm{ne}^+(i)})\, v_i^*(x) \,\big)$$

Next, denote $\forall i \in \mathcal{V}$ by:

$$\mathscr{M}_i^+ = \big\{ v : \mathcal{L}^n \longrightarrow \mathbb{R}, \text{s.t.}, \exists v_s : \mathcal{L}^{|s \cup \mathrm{ne}^+(i)|} \longrightarrow \mathbb{R}, \forall s \in \mathcal{S}, \text{s.t.}, v(x) = \textstyle\sum_{s \in \mathcal{S}} v_s(x_{s \cup \mathrm{ne}^+(i)}), \forall x \in \mathcal{L}^n \big\}$$

One concludes that there exists n real-valued functions $v_i^* \in \mathscr{M}_i^+$, $\forall i \in \mathcal{V}$, such that, one has:

$$\begin{aligned}
\mathbb{E}\big[h(\mathcal{X})\big] &= h(x) + \textstyle\sum_{i \in \mathcal{V}} \big(\, v_i^*(x) - \sum_{x_i \in \mathcal{L}} \pi_i(x_{\mathrm{ne}^+(i)})\, v_i^*(x) \,\big),\, \forall x \in \mathcal{L}^n \\
&= \max_{\{v_i \in \mathscr{M}_i^+, \forall i \in \mathcal{V}\}} \inf_{x \in \mathcal{L}^n} \Big\{ h(x) + \textstyle\sum_{i \in \mathcal{V}} \big(\, v_i(x) - \sum_{x_i \in \mathcal{L}} \pi_i(x_{\mathrm{ne}^+(i)})\, v_i(x) \,\big) \Big\} \\
&= \min_{\{v_i : \mathcal{L}^n \longrightarrow \mathbb{R}, \forall i \in \mathcal{V}\}} \sup_{x \in \mathcal{L}^n} \Big\{ h(x) + \textstyle\sum_{i \in \mathcal{V}} \big(\, v_i(x) - \sum_{x_i \in \mathcal{L}} \pi_i(x_{\mathrm{ne}^+(i)})\, v_i(x) \,\big) \Big\}
\end{aligned}$$

equivalently, by Proposition 1, there exists $|\mathcal{S}^+|$ real-valued functions $w_t^* : \mathcal{L}^{|t|} \longrightarrow \mathbb{R}$, $\forall t \in \mathcal{S}^+$, such that, one has:

$$\begin{aligned}
\mathbb{E}\big[h(\mathcal{X})\big] &= h(x) + \sum_{t \doteq s \cup \mathrm{ne}^+(i) \in \mathcal{S}^+} \big(\, w_t^*(x_t) - \sum_{x_i \in \mathcal{L}} \pi_i(x_{\mathrm{ne}^+(i)})\, w_t^*(x_t) \,\big) \\
&= \max_{\{w_t : \mathcal{L}^{|t|} \longrightarrow \mathbb{R}, \forall t \in \mathcal{S}^+\}} \inf_{x \in \mathcal{L}^n} \Big\{ h(x) + \sum_{t \doteq s \cup \mathrm{ne}^+(i) \in \mathcal{S}^+} \big(\, w_t(x_t) - \sum_{x_i \in \mathcal{L}} \pi_i(x_{\mathrm{ne}^+(i)})\, w_t(x_t) \,\big) \Big\} \\
&= \min_{\{w_t : \mathcal{L}^{|t|} \longrightarrow \mathbb{R}, \forall t \in \mathcal{S}^+\}} \sup_{x \in \mathcal{L}^n} \Big\{ h(x) + \sum_{t \doteq s \cup \mathrm{ne}^+(i) \in \mathcal{S}^+} \big(\, w_t(x_t) - \sum_{x_i \in \mathcal{L}} \pi_i(x_{\mathrm{ne}^+(i)})\, w_t(x_t) \,\big) \Big\}
\end{aligned}$$

thus, establishing the proof of Lemma 1.

Appendix .5.2. Proof of Lemma 2

First of all, in order to show the first part of Lemma 2, assume an arbitrary function $f : \mathcal{L}^n \longrightarrow \mathbb{R}$. Then, since, by Proposition 3, MRF distribution $\mathbb{P}\big(\mathcal{X} = x\big), \forall x \in \mathcal{L}^n$ is the unique solution of system of linear equations (10), one derives, on the one hand, that $\mathbb{E}\big[f(\mathcal{X})\big]$ is the optimal objective value of the following two LPs:

$$\begin{aligned}
&\min_p \Big\{ \textstyle\sum_{x \in \mathcal{L}^n} f(x)\, p(x) \Big\} \\
&\begin{cases}
p(x) - \pi_i(x_{\mathrm{ne}^+(i)}) \textstyle\sum_{x_i \in \mathcal{L}} p(x) = 0,\, \forall x \in \mathcal{L}^n,\, \forall i \in \mathcal{V} \\
\textstyle\sum_{x \in \mathcal{L}^n} p(x) = 1 \\
p(x) \geq 0,\, \forall x \in \mathcal{L}^n
\end{cases}
\end{aligned} \tag{.15}$$

and

$$\max_{p}\Big\{\ \textstyle\sum_{x\in\mathcal{L}^n} f(x)\,p(x)\ \Big\}$$

$$\begin{cases} p(x)-\pi_i(x_{\mathrm{ne}^+(i)})\ \sum_{x_i\in\mathcal{L}} p(x)=0,\ \forall x\in\mathcal{L}^n,\ \forall i\in\mathcal{V} \\ \sum_{x\in\mathcal{L}^n} p(x)=1 \\ p(x)\geq 0,\ \forall x\in\mathcal{L}^n \end{cases} \tag{.16}$$

On the other hand, one first finds that the dual of LP (.15) writes as:

$$\max_{\mu,\ v_i:\mathcal{L}^n\longrightarrow\mathbb{R},\forall i\in\mathcal{V}}\{\ \mu\ \}$$

$$\begin{cases} \mu\leq f(x)+\sum_{i\in\mathcal{V}}\big(\,v_i(x)-\sum_{x_i\in\mathcal{L}}\pi_i(x_{\mathrm{ne}^+(i)})\,v_i(x)\,\big),\ \forall x\in\mathcal{L}^n \\ v_i(x)\in\mathcal{R},\ \forall x\in\mathcal{L}^n,\ \forall i\in\mathcal{V} \\ \mu\in\mathcal{R} \end{cases} \tag{.17}$$

in such a way that, by noticing that optimal value of μ in dual LP (.17) in terms of functions f and $v_i, \forall i\in\mathcal{V}$ is given by the expression:

$$\inf_{x\in\mathcal{L}^n}\{\ f(x)+\textstyle\sum_{i\in\mathcal{V}}\big(\,v_i(x)-\sum_{x_i\in\mathcal{L}}\pi_i(x_{\mathrm{ne}^+(i)})\,v_i(x)\,\big)\ \}$$

one may rewrite dual LP (.17), accordingly, as:

$$\max_{v_i:\mathcal{L}^n\longrightarrow\mathbb{R},\forall i\in\mathcal{V}}\ \inf_{x\in\mathcal{L}^n}\Big\{\ f(x)+\textstyle\sum_{i\in\mathcal{V}}\big(\,v_i(x)-\sum_{x_i\in\mathcal{L}}\pi_i(x_{\mathrm{ne}^+(i)})\,v_i(x)\,\big)\ \Big\}$$

Next, one proceeds in the same way as we have done above–by assuming LP (.16) instead of LP (.15)–for also establishing that the dual of LP (.16) may write as:

$$\min_{v_i:\mathcal{L}^n\longrightarrow\mathbb{R},\forall i\in\mathcal{V}}\ \sup_{x\in\mathcal{L}^n}\Big\{\ f(x)+\textstyle\sum_{i\in\mathcal{V}}\big(\,v_i(x)-\sum_{x_i\in\mathcal{L}}\pi_i(x_{\mathrm{ne}^+(i)})\,v_i(x)\,\big)\ \Big\}$$

One finally derives:

$$\begin{aligned} \mathbb{E}\big[f(\mathcal{X})\big] &= \max_{v_i:\mathcal{L}^n\longrightarrow\mathbb{R},\forall i\in\mathcal{V}}\ \inf_{x\in\mathcal{L}^n}\Big\{\ f(x)+\textstyle\sum_{i\in\mathcal{V}}\big(\,v_i(x)-\sum_{x_i\in\mathcal{L}}\pi_i(x_{\mathrm{ne}^+(i)})\,v_i(x)\,\big)\ \Big\} \\ &= \min_{v_i:\mathcal{L}^n\longrightarrow\mathbb{R},\forall i\in\mathcal{V}}\ \sup_{x\in\mathcal{L}^n}\Big\{\ f(x)+\textstyle\sum_{i\in\mathcal{V}}\big(\,v_i(x)-\sum_{x_i\in\mathcal{L}}\pi_i(x_{\mathrm{ne}^+(i)})\,v_i(x)\,\big)\ \Big\} \end{aligned}$$

thus, proving the first part of Lemma 2.

Second, in order to show the second part of Lemma 2, then, we have already shown above that there exists n real-valued functions $v_i^*:\mathcal{L}^n\longrightarrow\mathbb{R},\ \forall i\in\mathcal{V}$, such that:

$$\mathbb{E}\big[f(\mathcal{X})\big] = \inf_{x\in\mathcal{L}^n}\Big\{f(x)+\textstyle\sum_{i=1}^n\big(\,v_i^*(x)-\sum_{x_i\in\mathcal{L}}\pi_i(x_{\mathrm{ne}^+(i)})\,v_i^*(x)\,\big)\Big\}$$

hence:

$$\inf_{x\in\mathcal{L}^n}\Big\{f(x)-\mathbb{E}\big[f(\mathcal{X})\big]+\textstyle\sum_{i=1}^n\big(\,v_i^*(x)-\sum_{x_i\in\mathcal{L}}\pi_i(x_{\mathrm{ne}^+(i)})\,v_i^*(x)\,\big)\Big\}=0 \tag{.18}$$

Denote by:

$$\tilde{f}(x) = f(x)-\mathbb{E}\big[f(\mathcal{X})\big]+\textstyle\sum_{i=1}^n\big(\,v_i^*(x)-\sum_{x_i\in\mathcal{L}}\pi_i(x_{\mathrm{ne}^+(i)})\,v_i^*(x)\,\big),\ \forall x\in\mathcal{L}^n$$

Then, on the one hand, by using formula (.18), one has:

$$\tilde{f}(x)\geq 0, \forall x\in\mathcal{L}^n \tag{.19}$$

On the other hand, by using the identities:

$$\mathbb{P}\big(\mathcal{X}=x\big)=\pi_i(x_{\mathrm{ne}^+(i)})\,\mathbb{P}\big(\mathcal{X}_{\neq i}=x_{\neq i}\big),\ \forall x\in\mathcal{L}^n,\ \forall i\in\mathcal{V}$$

one finds that $\forall i\in\mathcal{V}$:

$$\begin{array}{rcl}\sum_{x\in\mathcal{L}^n}\big(\sum_{x_i\in\mathcal{L}}\pi_i(x_{\mathrm{ne}^+(i)})\,v_i^*(x)\big)\,\mathbb{P}\big(\mathcal{X}=x\big) & = & \sum_{x\in\mathcal{L}^n}\big(\pi_i(x_{\mathrm{ne}^+(i)})\,v_i^*(x)\big)\big(\sum_{x_i\in\mathcal{L}}\mathbb{P}\big(\mathcal{X}=x\big)\big)\\ & = & \sum_{x\in\mathcal{L}^n}v_i^*(x)\,\big(\pi_i(x_{\mathrm{ne}^+(i)})\,\mathbb{P}\big(\mathcal{X}_{\neq i}=x_{\neq i}\big)\,\big)\\ & = & \sum_{x\in\mathcal{L}^n}v_i^*(x)\,\mathbb{P}\big(\mathcal{X}=x\big)\\ & = & \mathbb{E}\big[v_i^*(\mathcal{X})\big]\end{array}$$

and one derives:

$$\begin{array}{rcl}\sum_{x\in\mathcal{L}^n}\tilde{f}(x)\,\mathbb{P}\big(\mathcal{X}=x\big) & = & \mathbb{E}\big[\tilde{f}(\mathcal{X})\big]\\ & = & \mathbb{E}\Big[f(\mathcal{X})-\mathbb{E}\big[f(\mathcal{X})\big]\Big]+\sum_{i=1}^n\big(\mathbb{E}\big[v_i^*(\mathcal{X})\big]-\mathbb{E}\big[v_i^*(\mathcal{X})\big]\big)\\ & = & 0\end{array}\qquad(.20)$$

in such a way that, by combining formulas (.19) and (.20) above, and since one has $\mathbb{P}\big(\mathcal{X}=x\big)>0, x\in\mathcal{L}^n$, one finally derives:

$$\tilde{f}(x)=f(x)-\mathbb{E}\big[f(\mathcal{X})\big]+\sum_{i=1}^n\big(v_i^*(x)-\sum_{x_i\in\mathcal{L}}\pi_i(x_{\mathrm{ne}^+(i)})\,v_i^*(x)\,\big)=0,\ \forall x\in\mathcal{L}^n$$

as otherwise, i.e.; if there existed $x\in\mathcal{L}^n$, such that, $\tilde{f}(x)>0$, then, one would find $\mathbb{E}\big[\tilde{f}(\mathcal{X})\big]>0$ which is an absurdity, hence:

$$\mathbb{E}\big[f(\mathcal{X})\big]=f(x)+\sum_{i=1}^n\big(v_i^*(x)-\sum_{x_i\in\mathcal{L}}\pi_i(x_{\mathrm{ne}^+(i)})\,v_i^*(x)\,\big),\ \forall x\in\mathcal{L}^n$$

thus, establishing the proof of Lemma 2.

Appendix .6. Proof of Theorem 4

First of all, one may completely reformulate linear system of equations (14) as a LP as follows:

$$\min_p\Big\{\ \sum_{x\in\mathcal{L}^n}p(x)+\sum_{t\doteq s\cup\mathrm{ne}^+(i)\in\mathcal{S}^+}\sum_{x_t\in\mathcal{L}^{|t|}}\nu_t(x_t)\big(p_t(x_t)-\pi_i(x_{\mathrm{ne}^+(i)})\sum_{x_i\in\mathcal{L}}p_t(x_t)\big)\ \Big\}$$
$$\begin{cases}\nu_t(x_t)\big(p_t(x_t)-\pi_i(x_{\mathrm{ne}^+(i)})\sum_{x_i\in\mathcal{L}}p_t(x_t)\big)\geq 0,\ \forall x_t\in\Lambda_t^{\pm},\forall t\doteq s\cup\mathrm{ne}^+(i)\in\mathcal{S}^+\\ p_t(x_t)-\pi_i(x_{\mathrm{ne}^+(i)})\sum_{x_i\in\mathcal{L}}p_t(x_t)=0,\ \forall x_t\in\Lambda_t^0,\forall t\doteq s\cup\mathrm{ne}^+(i)\in\mathcal{S}^+\\ \sum_{x_t\in\mathcal{L}^n}p(x)\geq 1\\ p(x)\geq 0,\ \forall x\in\mathcal{L}^n\end{cases}\qquad(.21)$$

as at optimality of the latter, one exactly recovers linear system of equations (14). Next, we need the following lemma which is shown in Appendix .6.1.

Lemma 3. *LP (16) is equivalent to the following one:*

$$
\min_{q}\Big\{ \textstyle\sum_{x\in\mathcal{L}^n} q(x) + \sum_{t\doteq s\cup ne^+(i)\in\mathcal{S}^+} \sum_{x_t\in\mathcal{L}^{|t|}} \nu_t(x_t)\big(q_t(x_t) - \pi_i(x_{ne^+(i)}) \sum_{x_i\in\mathcal{L}} q_t(x_t)\big)\Big\}
$$
$$
\begin{cases}
\nu_t(x_t)\big(q_t(x_t) - \pi_i(x_{ne^+(i)}) \sum_{x_i\in\mathcal{L}} q_t(x_t)\big) \geq 0,\ \forall x_t \in \Lambda_t^{\pm}, \forall t \doteq s\cup ne^+(i) \in \mathcal{S}^+\\
q_t(x_t) - \pi_i(x_{ne^+(i)}) \sum_{x_i\in\mathcal{L}} q_t(x_t) = 0,\ \forall x_t \in \Lambda_t^{0}, \forall t \doteq s\cup ne^+(i) \in \mathcal{S}^+\\
q_t(x_t) \geq 0,\ \forall x_t \in \mathcal{L}^{|t|},\ \forall t \in \mathcal{S}^+\\
q_t(x_t) = \sum_{j\in\mathcal{V}/t} \sum_{x_j\in\mathcal{L}} q(x_1,\ldots,x_j,\ldots,x_n),\ \forall x_t \in \mathcal{L}^{|t|},\ \forall t \in \mathcal{S}^+\\
\sum_{x_t\in\mathcal{L}^n} q(x) \geq 1\\
q(x) \in \mathbb{R},\ \forall x \in \mathcal{L}^n
\end{cases}
\tag{.22}
$$

in the sense that:

- *both LP (16) and LP (.22) achieve the same optimal objective value,*
- *any feasible solution of LP (16) coincides with the margins set with respect to $\mathcal{S}^+$ of a feasible solution of LP (.22), moreover, the objective value of the former in LP (16) is equal to the objective value of the latter in LP (.22),*
- *conversely, whatever a feasible solution of LP (.22), its margins set with respect to $\mathcal{S}^+$ is a feasible solution of LP (16), moreover, the objective value of the former in LP (.22) is equal to the objective value of the latter in LP (16).*

Then, in order to show Theorem 4, assume LP (.22) which, by Lemma 3, is equivalent to LP (16). First of all, one observes that LP (.21) and LP (.22) have the same objective function, moreover, the feasible set of LP (.21) is included in the feasible set of LP (16), and they achieve the same optimal objective value which is equal to 1. One concludes that any optimal solution of LP (.21) (which, then, is a probability distribution) is also an optimal solution of LP (16). Therefore, let us show, hereafter, the converse of the latter statement which is that, any optimal solution of LP (.22) is also an optimal (probability) solution of LP (.21). Let us then introduce the nonnegative unknown variable M, and rewrite LP (.22) as:

$$
\min_{q,M}\Big\{ \textstyle\sum_{x\in\mathcal{L}^n} q(x) + \sum_{t\doteq s\cup \mathrm{ne}^+(i)\in\mathcal{S}^+} \sum_{x_t\in\mathcal{L}^{|t|}} \nu_t(x_t)\big(q_t(x_t) - \pi_i(x_{\mathrm{ne}^+(i)}) \sum_{x_i\in\mathcal{L}} q_t(x_t)\big)\Big\}
$$
$$
\begin{cases}
\nu_t(x_t)\big(q_t(x_t) - \pi_i(x_{\mathrm{ne}^+(i)}) \sum_{x_i\in\mathcal{L}} q_t(x_t)\big) \geq 0,\ \forall x_t \in \Lambda_t^{\pm}, \forall t \doteq s\cup \mathrm{ne}^+(i) \in \mathcal{S}^+\\
q_t(x_t) - \pi_i(x_{\mathrm{ne}^+(i)}) \sum_{x_i\in\mathcal{L}} q_t(x_t) = 0,\ \forall x_t \in \Lambda_t^{0}, \forall t \doteq s\cup \mathrm{ne}^+(i) \in \mathcal{S}^+\\
q_t(x_t) \geq 0,\ \forall x_t \in \mathcal{L}^{|t|},\ \forall t \in \mathcal{S}^+\\
q_t(x_t) = \sum_{j\in\mathcal{V}/t} \sum_{x_j\in\mathcal{L}} q(x_1,\ldots,x_j,\ldots,x_n),\ \forall x_t \in \mathcal{L}^{|t|},\ \forall t \in \mathcal{S}^+\\
\sum_{x_t\in\mathcal{L}^n} q(x) \geq 1\\
q(x) \geq -\frac{M}{L^n},\ \forall x \in \mathcal{L}^n\\
M \geq 0
\end{cases}
\tag{.23}
$$

and please observe that one does exactly recover LP (.22) by means of LP (.23) when M achieves its optimal value which is any big enough positive value. After that, by doing the change of variable $p(x) = q(x) + \frac{M}{L^n}, \forall x \in \mathcal{L}^n$, and by noticing that:

$$\nu_t(x_t)\theta_t(x_t) = |\theta_t(x_t)|,\ \forall x_t \in \mathcal{L}^{|t|},\ \forall t \in \mathcal{S}^+$$

one may rewrite LP (.23) as:

$$\min_{p,M}\Big\{ \sum_{x\in\mathcal{L}^n} p(x) - \big(1 + \sum_{t\in\mathcal{S}^+} \sum_{x_t\in\mathcal{L}^{|t|}} |\theta_t(x_t)|\big) M + \sum_{t\doteq s\cup \mathrm{ne}^+(i)\in\mathcal{S}^+} \sum_{x_t\in\mathcal{L}^{|t|}} \nu_t(x_t)\big(p_t(x_t) - \pi_i(x_{\mathrm{ne}^+(i)}) \sum_{x_i\in\mathcal{L}} p_t(x_t)\big)\Big\}$$
$$\begin{cases} \nu_t(x_t)\big(p_t(x_t) - \pi_i(x_{\mathrm{ne}^+(i)}) \sum_{x_i\in\mathcal{L}} p_t(x_t)\big) \geq |\theta_t(x_t)|\, M,\ \forall x_t \in \Lambda_t^{\pm}, \forall t \doteq s \cup \mathrm{ne}^+(i) \in \mathcal{S}^+ \\ p_t(x_t) - \pi_i(x_{\mathrm{ne}^+(i)}) \sum_{x_i\in\mathcal{L}} p_t(x_t) = 0,\ \forall x_t \in \Lambda_t^0, \forall t \doteq s \cup \mathrm{ne}^+(i) \in \mathcal{S}^+ \\ p_t(x_t) \geq \frac{M}{L^{|t|}},\ \forall x_t \in \mathcal{L}^{|t|},\ \forall t \in \mathcal{S}^+ \\ p_t(x_t) = \sum_{j\in\mathcal{V}/t} \sum_{x_j\in\mathcal{L}} p(x_1,\dots,x_j,\dots,x_n),\ \forall x_t \in \mathcal{L}^{|t|},\ \forall t \in \mathcal{S}^+ \\ \sum_{x_t\in\mathcal{L}^n} p(x) - 1 \geq M \\ p(x) \geq 0,\ \forall x \in \mathcal{L}^n \\ M \geq 0 \end{cases} \tag{.24}$$

Next, by noticing that the optimal value of M in terms of any feasible solution p of LP (.24) is given by the expression:

$$\inf\Big\{ \sum_{x_t\in\mathcal{L}^n} p(x) - 1, \inf_{x_t\in\Lambda_t^{\pm}} \big\{ \frac{\nu_t(x_t)\big(p_t(x_t) - \pi_i(x_{\mathrm{ne}^+(i)}) \sum_{x_i\in\mathcal{L}} p_t(x_t)\big)}{|\theta_t(x_t)|} \big\}\Big\}$$

moreover, any feasible solution p of LP (.24) has to satisfy the following mixed system of linear equalities and inequalities:

$$\begin{cases} \nu_t(x_t)\big(p_t(x_t) - \pi_i(x_{\mathrm{ne}^+(i)}) \sum_{x_i\in\mathcal{L}} p_t(x_t)\big) \geq 0,\ \forall x_t \in \Lambda_t^{\pm}, \forall t \doteq s \cup \mathrm{ne}^+(i) \in \mathcal{S}^+ \\ p_t(x_t) - \pi_i(x_{\mathrm{ne}^+(i)}) \sum_{x_i\in\mathcal{L}} p_t(x_t) = 0,\ \forall x_t \in \Lambda_t^0, \forall t \doteq s \cup \mathrm{ne}^+(i) \in \mathcal{S}^+ \\ p_t(x_t) \geq 0,\ \forall x_t \in \mathcal{L}^{|t|},\ \forall t \in \mathcal{S}^+ \\ p_t(x_t) = \sum_{j\in\mathcal{V}/t} \sum_{x_j\in\mathcal{L}} p(x_1,\dots,x_j,\dots,x_n),\ \forall x_t \in \mathcal{L}^{|t|},\ \forall t \in \mathcal{S}^+ \\ \sum_{x_t\in\mathcal{L}^n} p(x) - 1 \geq 0 \\ p(x) \geq 0,\ \forall x \in \mathcal{L}^n \end{cases} \tag{.25}$$

in such a way that, the inequalities of the form:

$$p_t(x_t) \geq 0,\ \forall x_t \in \mathcal{L}^{|t|},\ \forall t \in \mathcal{S}^+$$

are redundant in system (.25), one may rewrite LP (.24), accordingly, as:

$$\begin{aligned}&\min_{p}\Big\{ \textstyle\sum_{x\in\mathcal{L}^n} p(x)+\\ &\textstyle\sum_{t\doteq s\cup \mathrm{ne}^+(i)\in\mathcal{S}^+}\sum_{x_t\in\mathcal{L}^{|t|}} \nu_t(x_t)\big(p_t(x_t)-\pi_i(x_{\mathrm{ne}^+(i)})\sum_{x_i\in\mathcal{L}} p_t(x_t)\big)-\\ &\Theta_0 \inf\Big\{ \textstyle\sum_{x_t\in\mathcal{L}^n} p(x)-1, \inf_{x_t\in\Lambda_t^{\pm}}\big\{\frac{\nu_t(x_t)\left(p_t(x_t)-\pi_i(x_{\mathrm{ne}^+(i)})\sum_{x_i\in\mathcal{L}} p_t(x_t)\right)}{|\theta_t(x_t)|}\big\}\Big\}\Big\}\\ &\begin{cases}\nu_t(x_t)\big(p_t(x_t)-\pi_i(x_{\mathrm{ne}^+(i)})\sum_{x_i\in\mathcal{L}} p_t(x_t)\big)\geq 0,\ \forall x_t\in\Lambda_t^{\pm}, \forall t\doteq s\cup\mathrm{ne}^+(i)\in\mathcal{S}^+\\ p_t(x_t)-\pi_i(x_{\mathrm{ne}^+(i)})\sum_{x_i\in\mathcal{L}} p_t(x_t)=0,\ \forall x_t\in\Lambda_t^0, \forall t\doteq s\cup\mathrm{ne}^+(i)\in\mathcal{S}^+\\ p_t(x_t)=\sum_{j\in\mathcal{V}/t}\sum_{x_j\in\mathcal{L}} p(x_1,\dots,x_j,\dots,x_n),\ \forall x_t\in\mathcal{L}^{|t|},\ \forall t\in\mathcal{S}^+\\ \sum_{x_t\in\mathcal{L}^n} p(x)\geq 1\\ p(x)\geq 0,\ \forall x\in\mathcal{L}^n\end{cases}\end{aligned} \tag{.26}$$

where we have denoted $\Theta_0 = 1+\sum_{t\in\mathcal{S}^+}\sum_{x_t\in\mathcal{L}^{|t|}}|\theta_t(x_t)|$. It follows that showing that any optimal solution of LP (.22) is nonnegative (hence, an optimal solution of LP (.21) too) amounts to showing that any optimal solution p^* of LP (.26) has to satisfy:

$$\inf\Big\{ \textstyle\sum_{x_t\in\mathcal{L}^n} p^*(x)-1, \inf_{x_t\in\Lambda_t^{\pm}}\big\{\frac{\nu_t(x_t)\left(p^*_t(x_t)-\pi_i(x_{\mathrm{ne}^+(i)})\sum_{x_i\in\mathcal{L}} p^*_t(x_t)\right)}{|\theta_t(x_t)|}\big\}\Big\}=0 \tag{.27}$$

Therefore, for the purpose of showing formula (.27) above, consider the following LP relaxation of LP (.26):

$$\begin{aligned}&\min_{q}\Big\{ \textstyle\sum_{x\in\mathcal{L}^n} q(x)+\\ &\textstyle\sum_{t\doteq s\cup \mathrm{ne}^+(i)\in\mathcal{S}^+}\sum_{x_t\in\mathcal{L}^{|t|}} \nu_t(x_t)\big(q_t(x_t)-\pi_i(x_{\mathrm{ne}^+(i)})\sum_{x_i\in\mathcal{L}} q_t(x_t)\big)-\\ &\Theta_0 \inf\Big\{ \textstyle\sum_{x_t\in\mathcal{L}^n} q(x)-1, \inf_{x_t\in\Lambda_t^{\pm}}\big\{\frac{\nu_t(x_t)\left(q_t(x_t)-\pi_i(x_{\mathrm{ne}^+(i)})\sum_{x_i\in\mathcal{L}} q_t(x_t)\right)}{|\theta_t(x_t)|}\big\}\Big\}\Big\}\\ &\begin{cases}\nu_t(x_t)\big(q_t(x_t)-\pi_i(x_{\mathrm{ne}^+(i)})\sum_{x_i\in\mathcal{L}} q_t(x_t)\big)\geq 0,\ \forall x_t\in\Lambda_t^{\pm}, \forall t\doteq s\cup\mathrm{ne}^+(i)\in\mathcal{S}^+\\ q_t(x_t)-\pi_i(x_{\mathrm{ne}^+(i)})\sum_{x_i\in\mathcal{L}} q_t(x_t)=0,\ \forall x_t\in\Lambda_t^0, \forall t\doteq s\cup\mathrm{ne}^+(i)\in\mathcal{S}^+\\ q_t(x_t)=\sum_{j\in\mathcal{V}/t}\sum_{x_j\in\mathcal{L}} q(x_1,\dots,x_j,\dots,x_n),\ \forall x_t\in\mathcal{L}^{|t|},\ \forall t\in\mathcal{S}^+\\ q_t(x_t)\geq 0,\ \forall x_t\in\mathcal{L}^{|t|},\ \forall t\in\mathcal{S}^+\\ \sum_{x_t\in\mathcal{L}^n} q(x)\geq 1\\ q(x)\in\mathbb{R},\ \forall x\in\mathcal{L}^n\end{cases}\end{aligned} \tag{.28}$$

Clearly, the objective function of LP (.28) coincides with the one of LP (.26), moreover, the feasible set of the latter is included in the one of the former. Thus, by introducing the nonnegative unknown variable M', one may rewrite LP (.28) as:

$$\min_{q,M'}\Big\{ \sum_{x\in\mathcal{L}^n} q(x)+ \\ \sum_{t\doteq s\cup \mathrm{ne}^+(i)\in\mathcal{S}^+} \sum_{x_t\in\mathcal{L}^{|t|}} \nu_t(x_t)\big(q_t(x_t)-\pi_i(x_{\mathrm{ne}^+(i)})\sum_{x_i\in\mathcal{L}} q_t(x_t)\big)- \\ \Theta_0 \inf\Big\{ \sum_{x_t\in\mathcal{L}^n} q(x)-1, \inf_{x_t\in\Lambda_t^{\pm}}\big\{\tfrac{\nu_t(x_t)\left(q_t(x_t)-\pi_i(x_{\mathrm{ne}^+(i)})\sum_{x_i\in\mathcal{L}} q_t(x_t)\right)}{|\theta_t(x_t)|}\big\}\Big\}\Big\}$$

$$\begin{cases} \nu_t(x_t)\big(q_t(x_t)-\pi_i(x_{\mathrm{ne}^+(i)})\sum_{x_i\in\mathcal{L}} q_t(x_t)\big)\geq 0,\ \forall x_t\in\Lambda_t^{\pm}, \forall t\doteq s\cup\mathrm{ne}^+(i)\in\mathcal{S}^+ \\ q_t(x_t)-\pi_i(x_{\mathrm{ne}^+(i)})\sum_{x_i\in\mathcal{L}} q_t(x_t)=0,\ \forall x_t\in\Lambda_t^0, \forall t\doteq s\cup\mathrm{ne}^+(i)\in\mathcal{S}^+ \\ q_t(x_t)=\sum_{j\in\mathcal{V}/t}\sum_{x_j\in\mathcal{L}} q(x_1,\dots,x_j,\dots,x_n),\ \forall x_t\in\mathcal{L}^{|t|},\ \forall t\in\mathcal{S}^+ \\ q_t(x_t)\geq 0,\ \forall x_t\in\mathcal{L}^{|t|},\ \forall t\in\mathcal{S}^+ \\ \sum_{x_t\in\mathcal{L}^n} q(x)\geq 1 \\ q(x)\geq -\frac{M'}{L^n},\ \forall x\in\mathcal{L}^n \\ M'\geq 0 \end{cases} \tag{.29}$$

again, please observe that one does exactly recover LP (.28) by means of LP (.29) when M' achieves its optimal value which is any big enough positive value. After that, by doing the change of variable $p(x)=q(x)+\frac{M'}{L^n}, \forall x\in\mathcal{L}^n$, and by noticing that the variable M' cancels out in the resulting objective function, one may rewrite LP (.29) as:

$$\min_{p,M'}\Big\{ \sum_{x\in\mathcal{L}^n} p(x)+ \\ \sum_{t\doteq s\cup \mathrm{ne}^+(i)\in\mathcal{S}^+} \sum_{x_t\in\mathcal{L}^{|t|}} \nu_t(x_t)\big(p_t(x_t)-\pi_i(x_{\mathrm{ne}^+(i)})\sum_{x_i\in\mathcal{L}} p_t(x_t)\big)- \\ \Theta_0 \inf\Big\{ \sum_{x_t\in\mathcal{L}^n} p(x)-1, \inf_{x_t\in\Lambda_t^{\pm}}\big\{\tfrac{\nu_t(x_t)\left(p_t(x_t)-\pi_i(x_{\mathrm{ne}^+(i)})\sum_{x_i\in\mathcal{L}} p_t(x_t)\right)}{|\theta_t(x_t)|}\big\}\Big\}\Big\}$$

$$\begin{cases} \nu_t(x_t)\big(q_t(x_t)-\pi_i(x_{\mathrm{ne}^+(i)})\sum_{x_i\in\mathcal{L}} p_t(x_t)\big)\geq |\theta_t(x_t)|\, M',\ \forall x_t\in\Lambda_t^{\pm}, \forall t\doteq s\cup\mathrm{ne}^+(i)\in\mathcal{S}^+ \\ p_t(x_t)-\pi_i(x_{\mathrm{ne}^+(i)})\sum_{x_i\in\mathcal{L}} p_t(x_t)=0,\ \forall x_t\in\Lambda_t^0, \forall t\doteq s\cup\mathrm{ne}^+(i)\in\mathcal{S}^+ \\ p_t(x_t)=\sum_{j\in\mathcal{V}/t}\sum_{x_j\in\mathcal{L}} p(x_1,\dots,x_j,\dots,x_n),\ \forall x_t\in\mathcal{L}^{|t|},\ \forall t\in\mathcal{S}^+ \\ p_t(x_t)\geq \frac{M}{L^{|t|}},\ \forall x_t\in\mathcal{L}^{|t|},\ \forall t\in\mathcal{S}^+ \\ \sum_{x_t\in\mathcal{L}^n} p(x)\geq 1+M' \\ p(x)\geq 0,\ \forall x\in\mathcal{L}^n \\ M'\geq 0 \end{cases} \tag{.30}$$

Next, by noticing that the objective function of LP (.30) does not depend on M', moreover, bigger values constrain more the set of feasible values of p, one deduces that the optimal value of M' is 0, in such a way that, LP (.30) is equivalent to:

$$\min_{p}\Big\{ \sum_{x\in\mathcal{L}^n} p(x) +$$
$$\sum_{t\doteq s\cup \mathrm{ne}^+(i)\in\mathcal{S}^+} \sum_{x_t\in\mathcal{L}^{|t|}} \nu_t(x_t)\big(p_t(x_t) - \pi_i(x_{\mathrm{ne}^+(i)})\sum_{x_i\in\mathcal{L}} p_t(x_t)\big) -$$
$$\Theta_0 \inf\Big\{ \sum_{x_t\in\mathcal{L}^n} p(x) - 1, \inf_{x_t\in\Lambda_t^{\pm}} \big\{ \frac{\nu_t(x_t)\big(p_t(x_t)-\pi_i(x_{\mathrm{ne}^+(i)})\sum_{x_i\in\mathcal{L}} p_t(x_t)\big)}{|\theta_t(x_t)|} \big\}\Big\}\Big\}$$
$$\begin{cases} \nu_t(x_t)\big(q_t(x_t) - \pi_i(x_{\mathrm{ne}^+(i)})\sum_{x_i\in\mathcal{L}} p_t(x_t)\big) \geq 0,\ \forall x_t\in\Lambda_t^{\pm}, \forall t \doteq s\cup \mathrm{ne}^+(i)\in\mathcal{S}^+ \\ p_t(x_t) - \pi_i(x_{\mathrm{ne}^+(i)})\sum_{x_i\in\mathcal{L}} p_t(x_t) = 0,\ \forall x_t\in\Lambda_t^{0}, \forall t \doteq s\cup \mathrm{ne}^+(i)\in\mathcal{S}^+ \\ p_t(x_t) = \sum_{j\in\mathcal{V}/t}\sum_{x_j\in\mathcal{L}} p(x_1,\ldots,x_j,\ldots,x_n),\ \forall x_t\in\mathcal{L}^{|t|},\ \forall t\in\mathcal{S}^+ \\ p_t(x_t)\geq 0,\ \forall x_t\in\mathcal{L}^{|t|},\ \forall t\in\mathcal{S}^+ \\ \sum_{x_t\in\mathcal{L}^n} p(x)\geq 1 \\ p(x)\geq 0,\ \forall x\in\mathcal{L}^n \end{cases} \tag{.31}$$

Finally, by dropping from the latter the redundant constraints of the form:

$$p_t(x_t)\geq 0,\ \forall x_t\in\mathcal{L}^{|t|},\ \forall t\in\mathcal{S}^+$$

one retrieves exactly LP (.26). To sum up, so far we have shown that:

- LP (.26) achieves the same optimal objective value as LP (.22),
- LP (.28) achieves the same optimal objective value as LP (.26).

and one derives, on the one hand, that LP (.28) achieves the same optimal objective value as LP (.22). On the other hand, since:

- the objective value of LP (.28) is smaller than the one of LP (.22),
- yet, the latter share the same feasible set and achieve the same optimal objective value.

one concludes that any optimal solution of LP (.22), hence of LP (.26) denoted by p^* has to satisfy formula (.27) above, which also means that the optimal solutions of LP (.22) are nonnegative (i.e.; probability) solutions, also, turning out to be optimal for LP (.21). Finally, by Lemma 3, one concludes that any optimal solution of LP (16) denoted by $\{q_t^*, \forall t\in\mathcal{S}^+\}$ stands for the set of marginal distributions which are a solution of linear system (14), thus, establishing the proof of Theorem 4.

Appendix .6.1. Proof of Lemma 3

Denote by $\mathscr{O}_{\mathcal{S}^+}$ the ortho-marginal operator with respect to $\mathcal{S}^+$, and suppose $\{q_t, \forall t\in\mathcal{S}^+\}$ is a pseudo-marginals-set which is feasible for LP (16). Then, by Theorem 2, $\{q_t, \forall t\in\mathcal{S}^+\}$ stands for the margins-set with respect to $\mathcal{S}^+$ of a (non-unique) function $q:\mathcal{L}^n\rightarrow\mathbb{R}$, moreover, there exists real coefficients $\rho_t, \forall t\in\mathrm{Fclos}_{\cap}(\mathcal{S}^+)$, such that, the general closed-form expression of such a function q is given by:

$$q(x) = v(x) - (\mathscr{O}_{\mathcal{S}^+}v)(x) + \sum_{t\in\mathrm{Fclos}_{\cap}(\mathcal{S}^+)} \frac{\rho_t\, q_t(x_t)}{L^{n-|t|}},\ \forall x\in\mathcal{L}^n \tag{.32}$$

with $v : \mathcal{L}^n \to \mathbb{R}$ being arbitrary, and one has:

$$\begin{cases} q_t(x_t) = \sum_{j \in \mathcal{V}/t} \sum_{x_j \in \mathcal{L}} q(x_1, \ldots, x_j, \ldots, x_n), \, \forall x_t \in \mathcal{L}^{|t|}, \, \forall t \in \mathcal{S}^+ \\ \sum_{x_t \in \mathcal{L}^{|t|}} q_t(x_t) = \sum_{x \in \mathcal{L}^n} q(x), \, \forall t \in \mathcal{S}^+ \end{cases} \tag{.33}$$

in such a way that, the rest of the proof of Lemma 3 follows immediately from the fact that LP (16) may be completely reformulated in terms of the margins-set with respect to $\mathcal{S}^+$ of a general function $q : \mathcal{L}^n \to \mathbb{R}$, said otherwise, as LP (.22), and vice-versa, thus, proving Lemma 3.