Exact MAP inference in general higher-order graphical models using linear programming

Ikhlef Bechar IKHLEF.BECHAR@GMAIL.COM, IKHLEF.BECHAR@UNIV-TLN.FR
Aix Marseille Univ, Université de Toulon, CNRS, ENSAM, LSIS
La Garde 83957, France

Abstract

This paper is concerned with the problem of exact MAP inference in general higher-order graphical models by means of a traditional linear programming relaxation approach. In fact, the proof that we have developed in this paper is a rather simple algebraic proof being made straightforward, above all, by the introduction of two novel algebraic tools. Indeed, on the one hand, we introduce the notion of delta-distribution which merely stands for the difference of two arbitrary probability distributions, and which mainly serves to alleviate the sign constraint inherent to a traditional probability distribution. On the other hand, we develop an approximation framework of general discrete functions by means of an orthogonal projection expressing in terms of linear combinations of function margins with respect to a given collection of point subsets, though, we rather exploit the latter approach for the purpose of modeling locally consistent sets of discrete functions from a global perspective. After that, as a first step, we develop from scratch the expectation optimization framework which is nothing else than a reformulation, on stochastic grounds, of the convex-hull approach, as a second step, we develop the traditional LP relaxation of such an expectation optimization approach, and we show that it enables to solve the MAP inference problem in graphical models under rather general assumptions. Last but not least, we describe an algorithm which allows to compute an exact MAP solution from a perhaps fractional optimal (probability) solution of the proposed LP relaxation.

1. Introduction

The MAP inference problem in general higher-order graphical models (Bayesian models, Markov random field models (MRFs), and beyond), also, called the discrete higher-order multiple-partitioning (or multi-label) problem (HoMPP) can be stated as follows. Given:

1. a discrete domain of points (sites) assumed, without loss of generality, to be the integer set $\Omega = \{1, \ldots, n\}$, where n stands for an integer which is greater than, or equal to 2,
2. a discrete label-set assumed, without loss of generality, to be the integer set $\mathcal{L} = \{0, \ldots, L-1\}$, where L stands for an integer which is greater than, or equal to 2,
3. a hypersite-set $\mathcal{S}$ consisting of subsets of Ω with cardinality greater than, or equal to 1,
4. a set of real-valued local functions $\{g_s : \mathcal{L}^{|s|} \longrightarrow \mathbb{R}, \forall s \in \mathcal{S}\}$.

then, the goal is to find a multi-label function of the form:

$$\begin{array}{rccc} \tilde{x}: & \Omega & \longrightarrow & \mathcal{L} \\ & i & \longmapsto & \tilde{x}(i) \end{array}$$

in a way which either minimizes, or maximizes a higher-order cost function defined for all multi-label function $\tilde{x}$ as:

$$g(\tilde{x}) = \sum_{s \in \mathcal{S}} g_s\big(\tilde{x}(s)\big) \tag{1}$$

where it has been assumed that, $\forall s \in \mathcal{S}$, one has $\tilde{x}(s) = \big(\tilde{x}(i)\big)_{i \in s}$.

For the sake of convenience in the remainder, we propose to encode the multi-label function $\tilde{x}(\cdot)$ by means of a n-dimensional integer vector $x = (x_1, \ldots, x_n) \in \mathcal{L}^n$, while simply bearing in mind that $\forall i \in \Omega$, one has $\tilde{x}(i) = x_i$, and we refer throughout to x as the multi-label vector (MLV). The problem then amounts to finding an integer vector solution in $\mathcal{L}^n$ which either globally solves the following minimization problem:

$$\inf_{x \in \mathcal{L}^n} \Big\{ g(x) = \sum_{s \in \mathcal{S}} g_s\big(x_s\big) \Big\} \tag{2}$$

or globally solves the following maximization problem:

$$\sup_{x \in \mathcal{L}^n} \Big\{ g(x) = \sum_{s \in \mathcal{S}} g_s\big(x_s\big) \Big\} \tag{3}$$

More generally, one might be interested in finding both modes (i.e.; the minimum and the maximum solutions) of g, and we propose to denote such a problem by:

$$\operatorname*{Modes}_{x \in \mathcal{L}^n} \Big\{ g(x) = \sum_{s \in \mathcal{S}} g_s\big(x_s\big) \Big\} \tag{4}$$

Furthermore, in order to rule out any trivial instances of the HoMPP, therefore, we make throughout the following mild assumptions:

- $\cup_{s \in \mathcal{S}} \{i \in s\} = \Omega$,
- g and $g_s, \forall s \in \mathcal{S}$ are non-constant functions.

For the sake of clarity in the remainder, we shall be referring to minimization problem (2), maximization problem (3) and modes finding problem (4) using the acronyms MinMPP, MaxMPP, and ModesMPP, respectively. Furthermore, we want to emphasize that in practice, such a higher-order function $g(x)$ often arises as minus the log-likelihood of a instance of a graphical model (e.g.; a Bayesian model, a MRF model, and so forth) given the observed data (up to the minus log of a normalization constant). Then, depending on the application, either one may only be interested in a MAP solution of the HoMPP (i.e.; a one which maximizes the likelihood, equivalently, which minimizes $g(x)$), or in both modes of the likelihood. In fact, one of the contributions of this paper is that it also shows that both modes of g are intimately related to each other (please refer to section 8, especially to the discussion which follows Theorem 11 for more details).

The remainder of this paper is structured as follows. After reviewing some the existing literature on the MAP inference problem in graphical models, we first reformulate, on stochastic grounds, both MinMPP (2) and MaxMPP (3) as expectation minimization and maximization linear programs (LPs), respectively. After that, we introduce the notion of delta-distribution, and we reformulate ModesMPP (4) as a delta-expectation minimization LP. Next, we introduce the ortho-marginal framework as a general discrete function approximation by means of an orthogonal projection in terms of linear combinations of function margins with respect to a given hypersite-set, though, as mentioned in the abstract, we rather use the latter for the purpose of modeling local consistency from a global perspective. Then, we proceed in a traditional way for obtaining useful LP relaxations of the HoMPP, by merely enforcing locally the probability and the delta-probability axioms, respectively. Having in mind the two mathematical tools above, namely, the notion of delta-distribution and the ortho-marginal framework, we reformulate the proposed LP relaxations from a global viewpoint, and we show that their optimal solutions coincide with the ones of their original (hard) versions. Last but not least, since one is only guaranteed to recover a set of optimal marginal distributions (resp. a set of optimal marginal delta-distributions) of the HoMPP, we also develop an algorithm allowing to compute modes of g from a perhaps fractional solution of its LP relaxation. Before moving to the crux of the approach, we want to emphasize that the present paper is self-contained, moreover, all the presented results throughout are shown using rather simple algebraic techniques widely accessible to anyone who is familiar with the basic concepts of linear algebra, linear programming, and probability theory.

2. Related work

Maximum-A-Posteriori (MAP) estimation in higher-order probabilistic graphical models (Lauritzen, 1991; Bishop, 2006; Wainwright & Jordan, 2008), also referred to in the operational research and computer vision literatures as the higher-order multiple-partitioning (or the multi-label) problem (HoMPP), has been, for many decades, a central topic in the literature of AI and related fields (statistics, machine learning, data-mining, natural language processing, computer vision, coding theory, operations research, computational biology, to name a few). In the culture of mathematical programming, the HoMPP is nothing else than unconstrained integer programming (Nemhauser & Wolsey, 1988; Grootschel, Lovasz, & Schrijver, 1993), whereas in the culture of data science, the HoMPP often arises as an inference (or an inverse) problem, in the sense that one is interested in finding the most likely configuration of model parameters which explains the observed data. The choice of a graphical model for a given practical situation may be motivated by the nature of the (random) process which generates the data, but may also be severely constrained by available computing resources and/or real-time considerations. Thus, factorable graphical models have arisen as an almost inescapable AI tool both for modeling and solving a variety of AI problems, above all, due to their modularity, flexibility as well as their ability to model a variety of real-world problems. In this regard, two popular classes of graphical models are the Bayesian graphs (or the directed graphical models) (Pearl, 1982; Pearl & Russell, 2002), and the Markov random field (MRF) graphs (or the undirected graphical models) (Hammersley & Clifford, 1971; Kinderman & Snell, 1980). Historically, MRFs had long been known in the field of statistical physics (Ising, 1925; Ashkin & Teller, 1943; Potts, 1952), before they were first introduced in computer science (Besag, 1974) and later popularized by many other authors (Geman & Geman, 1984; Besag, 1986; Geman & Graffigne, 1986; Li, 1995). Nowadays, graphical models are a branch in its own right of statistical and probability theories, and of which use in AI is ubiquitous.

With that being said, exact MAP inference, or even approximate MAP inference in general graphical models is a hard combinatorial problem (Karp, 1972; Cooper, 1990; Dagum & Luby, 1993; Shimony, 1994; Roth, 1996; Chickering, 1996; Cipra, 2000; Megretski, 1996; Boykov, Veksler, & Zabih, 2001; Park & Darwiche, 2004; Cohen, Cooper, Jeavons, & Krokhin, 2006). As a matter of fact, unless $P = NP$, one may not even hope achieving an approximate polynomial-time algorithm for computing the modes of an arbitrary instance of a graphical model. Therefore, except in particular cases which are known to be solvable exactly and in polynomial-time (Hammer, 1965; Greig, Porteous, & Seheult, 1989; Boykov, Veksler, & Zabih, 1998; Ishikawa, 2003; Schlesinger, 2007; Osokin, Vetrov, & Kolmogorov, 2011), the MAP inference problem in graphical models has been mostly dealt with, so far, by using heuristical approaches, and which may be ranked in three main categories. First, probability-sampling-based approaches also called the Markov Chain Monte Carlo (MCMC) methods (Hastings, 1970; Green, 1995; Gelfand & Smith, 1990) have been among the firstly used MAP estimation algorithms in MRF models (Geman & Geman, 1984; Besag, 1986), and their good practical performances both in terms of computational efficiency and accuracy have been, extensively, reported in the literature (Baddeley & Van Lieshout, 1993; Winkler, 1995; Descombes, 2011). Graph-theory based approaches which are mostly variants of the graph-cut algorithm have been also extensively used for optimizing a plethora of MRF instances which are mainly encountered in computer vision (Boykov et al., 2001; Kolmogorov & Zabih, 2004; Liu & Veksler, 2010; Veksler, 2012). More recently, fostered by the important breakthroughs in linear programming (Chvàtal, 1983; Dantzig, 1990; Karmarkar, 1984; Bertsimas & Tsitsiklis, 1997) and, more generally, in convex programming (Ye, 1989; Nesterov & Nemirovsky, 1994; Nesterov, 2004, 2009; Lesaja, 2009; Beck & Teboulle, 2009), as well as by the important recent surge in high-performance computing, such as multi-processor and parallel computing (GPU) technologies (Bolz, Farmer, Grinspun, & Schrooder, 2003; Li, Lu, Hu, & Jiang, 2011), linear and convex programming relaxation approaches–including spatially-continuous continuous approaches (Nikolova, Esedoglu, & Chan, 2006; Cremers, Pock, Kolev, & Chambolle, 2011; Lellmann & Schnorr, 2011; Nieuwenhuis, Toeppe, & Cremers, 2013; Zach, Hane, & Pollefeys, 2014) and spatially-discrete ones (Schlesinger, 1976; Hummel & Zucker, 1983; Hammer, Hansen, & Simeone, 1984; Pearl, 1988; Sherali & Adams, 1990; Koster, Van Hoesel, & Kolen, 1998; Chekuri, Khanna, Naor, & Zosin, 2005; Kingsford, Chazelle, & Singh, 2005; Kolmogorov, 2006a; Werner, 2007; Cooper, 2012)– have arisen as a promising alternative both to graph-theory based and MCMC based MAP estimation approaches in graphical models. Generally speaking, the latter category of approaches may also be seen as an approximate marginal inference approach in graphical models (Wainwright, Jaakkola, & Willsky, 2005; Wainwright & Jordan, 2008), in the sense that, one generally attempts to optimize the objective over a relaxation of the marginal polytope constraints, in such a way that, an approximate MAP solution may be found by a mere rounding procedure, or by means of a more sophisticated message passing algorithm (Wainwright et al., 2005; Kolmogorov, 2006b; Komodakis, Paragios, & Tziritas, 2011; Sontag & Jaakkola, 2008). In fact, the approach which is described in this paper belongs to the latter category of approaches, yet, it may solve the MAP inference problem in an arbitrary graphical model instance.

3. The HoMPP expectation optimization framework

The goal of this section is to transform both MinMPP (2) and MaxMPP (3) into equivalent continuous optimization problems, and, eventually, into linear programs by means of the expectation-optimization framework.

Therefore, we first assume MinMPP (2), and in order to fix ideas once and for all throughout, we propose to develop from scratch the expectation minimization (EM) approach, allowing to recast any instance of MinMPP (2) as a linear program (LP). In the introduction section, we have assumed that the labeling process is purely deterministic, but unknown. Therefore in this section, we rather advocate a random multi-label process, consisting in randomly drawing vector samples $x \in \mathcal{L}^n$ with a certain probability, then assigning to each site $i \in \Omega$ realization x_i of its random label. Let us stress that randomization serves here only temporarily for developing the EM approach which is deterministic. Therefore, suppose a random multi-label vector (RMLV) $\mathcal{X} = \big(\mathcal{X}_i\big)_{i\in\Omega}$, with value domain $\mathcal{L}^n$, and consider the stochastic (random) version of the objective function of MinMPP (2) expressing as:

$$g\big(\mathcal{X}\big) = \sum_{s\in\mathcal{S}} g_s\big(\mathcal{X}_s\big)$$

Then, one writes the expectation of $g\big(\mathcal{X}\big)$ as:

$$\begin{aligned}\mathbb{E}\big[g\big(\mathcal{X}\big)\big] &= \textstyle\sum_{x\in\mathcal{L}^n} g(x)\,\mathbb{P}\big(\mathcal{X}=x\big) = \sum_{x\in\mathcal{L}^n}\big(\sum_{s\in\mathcal{S}} g_s(x_s)\big)\,\mathbb{P}\big(\mathcal{X}=x\big)\\ &= \textstyle\sum_{s\in\mathcal{S}}\sum_{x_s\in\mathcal{L}^{|s|}} g_s(x_s)\,\mathbb{P}\big(\mathcal{X}_s=x_s\big)\end{aligned}$$

Please observe that $\mathbb{E}\big[g\big(\mathcal{X}\big)\big]$ expresses solely in terms of the marginal distributions of the random vectors $\mathcal{X}_s, \forall s \in \mathcal{S}$. Next, suppose that one is rather given a set (or a family) $\mathscr{P}$ of candidate probability distributions of RMLV $\mathcal{X}$ [1], such that:

$$\forall p \in \mathscr{P},\ \begin{cases} p : \mathcal{L}^n \longrightarrow \mathbb{R},\\ p(x) \geq 0,\ \forall x \in \mathcal{L}^n,\\ \sum_{x\in\mathcal{L}^n} p(x) = 1.\end{cases}$$

and the goal is to choose among $\mathscr{P}$ the joint distribution of RMLV $\mathcal{X}$ which solves the following minimization problem:

$$\min_{p\in\mathscr{P}} \quad \Big\{ \mathbb{E}_p\big[g\big(\mathcal{X}\big)\big] = \sum_{x\in\mathcal{L}^n} g(x)\,p(x) \Big\} \tag{5}$$

We refer to minimization problem (5) as EMinMPP, standing for expectation minimization multiple-partitioning problem. Now, in order to see how EMinMPP (5) relates to MinMPP (2), one may write $\forall x \in \mathcal{L}^n$: $g(x) = \sum_{y\in\mathcal{L}^n} g(y)\,\mathbf{1}_x(y) = \mathbb{E}_{\mathbf{1}_x}\big[g\big(\mathcal{X}\big)\big]$, where $\mathbf{1}_x(\cdot)$ stands for the indicator function of x, defined as:

$$\forall y \in \mathcal{L}^n,\ \mathbf{1}_x(y) = \begin{cases} 1, & \text{if } y = x,\\ 0, & \text{else.}\end{cases}$$

1. But more formally speaking, one would rather consider a set of independent copies of $\mathcal{X}$, each of which is endowed with its own distribution in $\mathscr{P}$.

in such a way that, by denoting by $\mathscr{I} = \{\mathbf{1}_x(\cdot),\, x \in \mathcal{L}^n\}$ which stands for the set of indicator functions of the integer vector set $\{x \in \mathcal{L}^n\}$, one may completely reformulate MinMPP (2) as an instance of EMinMPP (5), with $\mathscr{P} = \mathscr{I}$. Furthermore, since $\forall p \in \mathscr{P}$, $\mathbb{E}_p\big[g(\mathcal{X})\big]$ writes as some convex combination of the elements of the set $\{g(x),\, x \in \mathcal{L}^n\}$, one derives immediately that:

$$\min_{p \in \mathscr{P}} \Big\{ \mathbb{E}_p\big[g(\mathcal{X})\big] \Big\} \geq \inf_{x \in \mathcal{L}^n} \{ g(x) \} \tag{6}$$

which means that EMinMLP (5) is an upper-bound for MinMPP (2). Then, Theorem 1 below gives a sufficient condition under which EMinMPP (5) exactly solves MinMPP (2), and how one may obtain, accordingly, an optimal vector solution of MinMPP (2) from a perhaps fractional optimal probability solution of EMinMPP (5).

Theorem 1 *Suppose that $\mathscr{P} \supseteq \mathcal{I}$. Then EMinMPP (5) achieves an optimal objective value equal to $\inf_{x \in \mathcal{L}^n} \{ g(x) \}$. Furthermore, if p^* is an optimal (probability) solution of EMinMPP (5), then any $x \in \mathcal{L}^n$, such that, $p^*(x) > 0$, is an optimal solution of MinMPP (2).*

Proof 1 *The assumption that $\mathscr{P} \supseteq \mathcal{I}$ guarantees that a strict equality is achieved in formula (6). Moreover, if a distribution p^* of RMLV $\mathcal{X}$ is optimal for problem (5), then so must be any indicator function which is expressed with a strictly positive coefficient in the convex combination of p^* in terms of indicator functions of the set $\mathcal{L}^n$, in other words, any integer vector sample of p^* must also be optimal for MinMPP (2).*

Clearly, Theorem 1 is nothing else than the probabilistic counterpart of the well-known convex hull reformulation in integer programming (Sherali & Adams, 1990; Grootschel et al., 1993; Bertsimas & Tsitsiklis, 1997; Wainwright & Jordan, 2008). Having said that, in the remainder, we will assume that $\mathscr{P}$ stands for the entire convex set of candidate joint distributions of $\mathcal{X}$ which is given by:

$$\mathscr{P} = \Big\{ p : \mathcal{L}^n \to [0,1], \text{ s.t.}, \sum_{x \in \mathcal{L}^n} p(x) = 1 \Big\} \tag{7}$$

Clearly, one has $\mathscr{P} \supseteq \mathcal{I}$, and $\mathscr{P}$ coincides with the convex hull of $\mathscr{I}$. One may reexpress EMinMPP (5), accordingly, as a linear program (LP) as follows:

$$\begin{array}{l} \min \Big\{ \sum_{s \in \mathcal{S}} \sum_{x_s \in \mathcal{L}^{|s|}} g_s(x_s)\, p_s(x_s) \Big\} \\ \begin{cases} p_s(x_s) = \sum_{i \notin s} \sum_{x_i \in \mathcal{L}} p(x_1, \ldots, x_i, \ldots, x_n),\ \forall x_s \in \mathcal{L}^{|s|},\ \forall s \in \mathcal{S} \\ \sum_{x \in \mathcal{L}^n} p(x) = 1 \\ p(x) \geq 0, \forall x \in \mathcal{L}^n \end{cases} \end{array} \tag{8}$$

Equally, one finds that the following LP:

$$\begin{array}{l} \max \Big\{ \sum_{s \in \mathcal{S}} \sum_{x_s \in \mathcal{L}^{|s|}} g_s(x_s)\, p_s(x_s) \Big\} \\ \begin{cases} p_s(x_s) = \sum_{i \notin s} \sum_{x_i \in \mathcal{L}} p(x_1, \ldots, x_i, \ldots, x_n),\ \forall x_s \in \mathcal{L}^{|s|},\ \forall s \in \mathcal{S} \\ \sum_{x \in \mathcal{L}^n} p(x) = 1 \\ p(x) \geq 0, \forall x \in \mathcal{L}^n \end{cases} \end{array} \tag{9}$$

completely solves MaxMPP (3). Throughout, we shall refer to LP (8) and LP (9) using the acronyms EMinMLP and EMaxMLP, respectively. We conclude this section by merely saying that both EMinMLP (8) and EMaxMLP (9) are untractable in their current form, and the goal in the remainder of this paper is to develop their efficient LP relaxations.

4. The HoMPP delta-expectation minimization framework

In this section, we develop the delta-expectation minimization framework for addressing ModesMPP (4), in other words, both MinMPP (2) and MaxMPP (3) in a common minimization framework.

4.1 Joint delta-distribution

Definition 1 (Joint delta-distribution) *We call a joint delta-distribution of RMLV $\mathcal{X}$ any function $q : \mathcal{L}^n \to \mathbb{R}$ which can write in terms of the difference of two (arbitrary) joint distributions of RMLV $\mathcal{X}$ as:*

$$q(x) = p(x) - p'(x), \, \forall x \in \mathcal{L}^n$$

where both p and p' stand for (ordinary) joint distributions of RMLV $\mathcal{X}$.

Theorem 2 provides a useful alternative definition of a joint delta-distribution of RMLV $\mathcal{X}$, without resorting to its ordinary joint distributions.

Theorem 2 *A function $q : \mathcal{L}^n \to \mathbb{R}$ defines a joint delta-distribution of RMLV $\mathcal{X}$, if and only if, q satisfies the following two formulas:*

1. $\sum_{x \in \mathcal{L}^n} q(x) = 0$,
2. $\sum_{x \in \mathcal{L}^n} |q(x)| \leq 2$.

The proof of Theorem 2 is detailed in Appendix section A.1.

Interestingly, one has managed to get rid of the pointwise sign constraint of ordinary distributions of RMLV $\mathcal{X}$ by means of its joint delta-distributions. One then notes that the decomposition of a joint delta-distribution of RMLV $\mathcal{X}$ in terms of the difference of its two ordinary joint distributions is, generally, non-unique, hence Proposition 1 which fully characterizes joint delta-distributions of RMLV $\mathcal{X}$ admitting such a unique decomposition.

Proposition 1 *A joint delta-distribution q of RMLV $\mathcal{X}$ admits a unique decomposition of the form $q = p - p'$, where both p and p' stand for joint distributions of RMLV $\mathcal{X}$, if and only if, one has $\sum_{x \in \mathcal{L}^n} |q(x)| = 2$, in which case, p and p' are uniquely given by:*

$$\begin{cases} p(x) = \sup\big\{0, q(x)\big\}, \, \forall x \in \mathcal{L}^n \\ p'(x) = \sup\big\{0, -q(x)\big\}, \, \forall x \in \mathcal{L}^n \end{cases}$$

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

Last but not least, Proposition 2 below establishes that any zero-mean function $q : \mathcal{L}^n \to \mathbb{R}$ defines, at worst, up to a multiplicative scale, a joint delta-distribution of RMLV $\mathcal{X}$.

Proposition 2 *Suppose a nonzero function* $q : \mathcal{L}^n \to \mathbb{R}$, *such that,* $\sum_{x\in\mathcal{L}^n} q(x) = 0$. *Then, there exists* $\lambda_q = \frac{2}{\sum_{x\in\mathcal{L}^n}|q(x)|} > 0$, *such that,* $\forall\lambda \in [0, \lambda_q]$, *the normalized function* $\tilde{q}_\lambda : \mathcal{L}^n \to \mathbb{R}$ *defined as:*

$$\tilde{q}_\lambda(x) = \lambda\, q(x), \forall x \in \mathcal{L}^n$$

defines a joint delta-distribution of RMLV $\mathcal{X}$.

The proof of Proposition 2 is sketched in Appendix section A.2.

4.2 Reformulation of a HoMPP as a delta-expectation minimization problem

We begin by introducing the notion of delta-expectation of a real-valued random function of RMLV $\mathcal{X}$.

Definition 2 (Delta-expectation) *Suppose* q *is a joint delta-distribution of RMLV* $\mathcal{X}$, *and suppose a real-valued function* $f : \mathcal{L}^n \to \mathbb{R}$. *Then, one defines the delta-expectation of random function* $f\big(\mathcal{X}\big)$ *as:*

$$\Delta\mathbb{E}_q\big[f\big(\mathcal{X}\big)\big] = \sum_{x\in\mathcal{L}^n} f(x)\, q(x) \tag{10}$$

Next, similarly to the EM framework, one rather assumes a set of candidate joint delta-distributions of RMLV $\mathcal{X}$ denoted by $\mathscr{Q}$, and considers the delta-expectation minimization problem:

$$\min_{q\in\mathscr{Q}} \Big\{ \Delta\mathbb{E}_q\Big[g\big(\mathcal{X}\big)\Big] \Big\} \tag{11}$$

In the remainder, we take $\mathscr{Q}$ as the entire (convex) set of joint delta-distributions of RMLV $\mathcal{X}$ which, according to Theorem 2, is defined as:

$$\mathscr{Q} = \Big\{ q : \mathcal{L}^n \to \mathbb{R},\ \text{s.t.,} \sum_{x\in\mathcal{L}^n} q(x) = 0,\ \text{and,} \sum_{x\in\mathcal{L}^n} |q(x)| \leq 2 \Big\} \tag{12}$$

thus enabling delta-expectation minimization problem (11) to be expressed as a LP as follows:

$$\begin{array}{l} \min \Big\{ \sum_{x\in\mathcal{L}^n} g(x)\, q(x) \Big\} \\ \begin{cases} \sum_{x\in\mathcal{L}^n} q(x) = 0 \\ \sum_{x\in\mathcal{L}^n} |q(x)| \leq 2 \end{cases} \end{array} \tag{13}$$

which may also expand, using the marginal delta-distributions of $\mathcal{X}_s, \forall s \in \mathcal{S}$, as:

$$\begin{array}{l} \min \Big\{ \sum_{s\in\mathcal{S}} \sum_{x_s\in\mathcal{L}^{|s|}} g_s(x_s)\, q_s(x_s) \Big\} \\ \begin{cases} q_s(x_s) = \sum_{i\notin s} \sum_{x_i\in\mathcal{L}} q(x_1, \ldots, x_i, \ldots, x_n),\ \forall x_s \in \mathcal{L}^{|s|},\ \forall s \in \mathcal{S} \\ \sum_{x\in\mathcal{L}^n} q(x) = 0 \\ \sum_{x\in\mathcal{L}^n} |q(x)| \leq 2 \end{cases} \end{array} \tag{14}$$

In the remainder, we refer to problem (13) using the acronym DEMinMLP. Theorem 3 below may be seen as the delta-distribution analog of Theorem 1.

Theorem 3 *Suppose q^* is an optimal solution of DEMinMLP (13). It follows that:*

1. *q^* achieves an optimal objective value which is equal to:* $\inf_{x\in\mathcal{L}^n}\big\{g(x)\big\}-\sup_{x\in\mathcal{L}^n}\big\{g(x)\big\}$,
2. $\forall x \in \mathcal{L}^n$, $q^*(x) > 0 \Rightarrow g(x) = \inf_{y\in\mathcal{L}^n}\big\{\, g(y)\,\big\}$,
3. $\forall x \in \mathcal{L}^n$, $q^*(x) < 0 \Rightarrow g(x) = \sup_{y\in\mathcal{L}^n}\big\{\, g(y)\,\big\}$.

Moreover, q^* satisfies $\sum_{x\in\mathcal{L}^n}|q^*(x)| = 2$, thereby, admitting a unique decomposition of the form $q^* = p^* - p'^*$, where both p^* and p'^* stand for two joint distributions of RMLV $\mathcal{X}$ which are given by:

$$\begin{cases} p^*(x) = \sup\{0, q^*(x)\},\ \forall x \in \mathcal{L}^n \\ p'^*(x) = \sup\{0, -q^*(x)\},\ \forall x \in \mathcal{L}^n \end{cases}$$

and which are optimal for EMinMLP (8) and EMaxMLP (9), respectively.

The proof of Theorem 3 is easily established by using the definition of a delta-distribution followed by the use of the result of Theorem 1.

5. The ortho-marginal framework

In this section, we describe an algebraic approach (called the ortho-marginal framework) for general discrete function approximation via an orthogonal projection in terms of linear combinations of function margins with respect to a given hypersite-set. Nevertheless, the main usefulness of such an approach in the present paper is that it enables to model any set of locally constant functions (see Definition 8) in terms of a global (yet non-unique) mother function $f : \mathcal{L}^n \to \mathbb{R}$. Therefore, in order to fix ideas once and for all in the remainder, subsection 5.1 is devoted to the introduction of all the useful definitions to the development of the ortho-marginal framework, and subsection 5.2 is devoted to the description of its main results.

Beforehand, we want to note that, throughout this section, we assume that $\mathcal{C}$ is a hypersite-set with respect to Ω, moreover, we assume some order (e.g.; a lexicographic order) on the elements of $\mathcal{C}$ which means that, whatever $c, c' \in \mathcal{C}$, if $c \neq c'$, then either one has $c < c'$, or one has $c' < c$.

5.1 Definitions

Definition 3 (Maximal hypersite-set) *One says that $\mathcal{C}$ is maximal, if and only if:*

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

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

Definition 4 (Frontier hypersite-set) *One defines the frontier of $\mathcal{C}$, denoted by $Front\big(\mathcal{C}\big)$, as the smallest maximal hypersite-set which is contained in $\mathcal{C}$. In plain words, $Front\big(\mathcal{C}\big)$ is the hypersite-set which contains all the hypervertices in $\mathcal{C}$ which are not included in any of its other hypervertices.*

Definition 5 (Ancestor hypersite) *Suppose a hypersite* $c \in \mathcal{C}/Front\big(\mathcal{C}\big)$ *(if any). Then, we call an ancestor hypersite of* c*, any hypersite* $\tilde{c} \in Front\big(\mathcal{C}\big)$*, such that,* $c \subset \tilde{c}$*.*

Definition 6 (Ancestry function) *We call an ancestry function with respect to* $\mathcal{C}$*, any function:*

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

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

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

Remark 1 *It does not take much effort to see that higher-order function* g *(1) may rewrite solely in terms of local functions with respect* $Front(\mathcal{S})$ *as:*

$$g(x) = \sum_{s \in Front(\mathcal{S})} {g'}_s(x_s), \, \forall x \in \mathcal{L}^n$$

by simply merging each term g_s *of* g *with respect to any* $s \in \mathcal{S}/Front(\mathcal{S})$ *with the term corresponding to any of its ancestors in* $Front(\mathcal{S})$*.*

Definition 7 (Margin) *Suppose a function* $u : \mathcal{L}^n \to \mathbb{R}$*, and a hypersite* $c \in \mathcal{C}$*. Then, one defines the margin of* u *with respect to* c *as the function* $u_c : \mathcal{L}^{|c|} \to \mathbb{R}$ *defined as:*

$$u_c(x_c) = \textstyle\sum_{i \in \Omega/c} \sum_{x_i \in \mathcal{L}} u(x_1, \ldots, x_i, \ldots, x_n), \, \forall x_c \in \mathcal{L}^{|c|} \tag{15}$$

Definition 8 (Pseudo-marginal) *One says that a set of local functions of the form* $\big\{u_c : \mathcal{L}^{|c|} \to \mathbb{R}, \forall c \in \mathcal{C}\big\}$ *is a pseudo-marginals-set (or a set of locally consistent functions) with respect to* $\mathcal{C}$*, if and only if, it satisfies the following identities:*

$$\begin{cases} \forall c, t \in Front(\mathcal{C}), \, c \cap t \neq \emptyset \Rightarrow \sum_{i \in c/t} \sum_{x_i \in \mathcal{L}} u_c(x_c) = \sum_{i \in t/c} \sum_{x_i \in \mathcal{L}} u_t(x_t) \\ \forall c \in \mathcal{C}/Front(\mathcal{C}), \, u_c(x_c) = \sum_{i \in anc(c)/c} \sum_{x_i \in \mathcal{L}} u_{anc(c)}(x_{anc(c)}), \, \forall x_c \in \mathcal{L}^{|c|} \end{cases} \tag{16}$$

where c/t *stands for the hypersite of which sites belong to* c *but do not belong to* t*, and* $anc(\cdot)$ *stands for an arbitrary ancestor function with respect to* $\mathcal{C}$ *(see Definition 6).*

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

Convention 1 *We abuse of notation by denoting by* $\emptyset$ *the empty hypersite (i.e.; a one which does not contain any site), and we 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 real quantity* $u_\emptyset = \sum_{x \in \mathcal{L}^n} u(x)$*.*

Definition 9 (Frontier-closure of a hypersite-set) *One defines the frontier-closure of* $\mathcal{C}$ *as the hypersite-set with respect to* Ω*, denoted by* $Fclos_\cap\big(\mathcal{C}\big)$*, such that:*

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

Algorithm 1 in Appendix section B.16 then shows how one may iteratively construct the frontier-closure of an arbitrary hypersite-set.

5.2 Main results of the ortho-marginal framework

First of all, Theorem 4 below establishes that marginalization of any function $f : \mathcal{L}^n \to \mathbb{R}$ with respect to $\mathcal{C}$ is intimately related to an orthogonal projection of f.

Theorem 4 *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 $\mathcal{C}$ of u coincides with the one of f,*
2. *all the margins of v with respect to $\mathcal{C}$ are identically equal to zero.*
3. *the closed-form expression of function u is given by:*

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

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

$$\rho_c = \begin{cases} 1 & , \text{ if } c \in Front(\mathcal{C}), \\ 1 - \sum_{t \in Fclos_{\cap}\left(\mathcal{C}\right) \text{ s.t. } c \subset t} \rho_t & , \text{ if } c \in Fclos_{\cap}\left(\mathcal{C}\right)/Front(\mathcal{C}) \end{cases} \tag{17}$$

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

$$\left(\mathscr{O}_{\mathcal{C}} f\right)(x) = \sum_{c \in Fclos_{\cap}\left(\mathcal{C}\right)} \tfrac{\rho_c\, f_c(x_c)}{L^{n-|c|}}, \,\forall x \in \mathcal{L}^n \tag{18}$$

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

The proof of Theorem 4 is sketched in A.4.

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

Theorem 5 below builds on the result of Theorem 4 for establishing that any pseudo-marginals-set with respect to $\mathcal{C}$ may be viewed as the actual margins-set with respect to $\mathcal{C}$ of a global, yet non-unique, function : $\mathcal{L}^n \to \mathbb{R}$.

Theorem 5 *Suppose $\left\{u_c : \mathcal{L}^{|c|} \to \mathbb{R}, \forall c \in \mathcal{C}\right\}$ is a pseudo-marginals-set, thus verifying identities (16). Then, whatever a function $v : \mathcal{L}^n \to \mathbb{R}$, the function $u : \mathcal{L}^n \to \mathbb{R}$ defined as:*

$$u(x) = \big(v(x) - (\mathscr{O}_{\mathcal{C}} v)(x)\big) + \sum_{c \in Fclos_{\cap}\left(\mathcal{C}\right)} \frac{\rho_c\, u_c(x_c)}{L^{n-|c|}}, \,\forall x \in \mathcal{L}^n \tag{19}$$

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

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

where the linear coefficients $\rho_c, \forall c \in Fclos_{\cap}\left(\mathcal{C}\right)$ are defined according to formula (17) above.

The proof of Theorem 5 is sketched in A.5.

Definition 10 (Ortho-marginal space) *The ortho-marginal space with respect to $\mathcal{C}$ denoted by $\mathscr{M}_{\mathcal{C}}$, is defined as the linear function space which is given by:*

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

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

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

Remark 2 *One notes that any function $u \in \mathscr{M}_{\mathcal{C}}$, reflexively, writes in terms of its margins with respect to $Fclos_{\cap}\big(\mathcal{C}\big)$ as:*

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

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

Proposition 3 *Suppose a real-valued function $h : \mathcal{L}^n \to \mathbb{R}$. Then, one has $h \in \mathscr{M}_{\mathcal{C}}$, if and only if, there exists a set of local functions $\big\{h_c : \mathcal{L}^{|c|} \to \mathbb{R}, \forall c \in \mathcal{C}\big\}$ (not to be confused here with the margins of h with respect to $\mathcal{C}$), such that:*

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

The proof of Proposition 3 is sketched in A.6.

Proposition 4 *One has:*

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

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

Proof 2 *The proof of Proposition 4 follows immediately from the definition of $\mathscr{M}_{\mathcal{C}}$, since if $h, h' \in \mathscr{M}$, then both h and h' write as a linear combination of their respective margins with respect to $Fclos_{\cap}\big(\mathcal{C}\big)$, which then must coincide if $h \equiv h'$, and vice-versa.*

6. LP relaxation of the HoMPP over the local marginal-polytope

In order to fix ideas throughout, thus, this section consists of subsection 6.1 in which we introduce some (or better said, we recall some already known) useful definitions, and subsection 6.2 where we develop the LP relaxation approach of the HoMPP.

6.1 Definitions

Definition 11 (Pseudo-marginal probability set) *Suppose* $\mathcal{P} := \{p_s : \mathcal{L}^{|s|} \to \mathbb{R}, \forall s \in \mathcal{S}\}$ *is a pseudo-marginals-set which, thus, satisfies identities (16). If, moreover,* $\mathcal{P}$ *verifies the following identities:*

$$\forall s \in \mathit{Front}(\mathcal{S}), \begin{cases} \sum_{x_s \in \mathcal{L}^{|s|}} p_s(x_s) = 1 \\ p_s(x_s) \geq 0, \, \forall x_s \in \mathcal{L}^{|s|} \end{cases} \tag{20}$$

then $\mathcal{P}$ *is called a pseudo-marginal probability set with respect to* $\mathcal{S}$*.*

Definition 12 (Pseudo-marginal polytope) *The pseudo- (or the local-) marginal polytope with respect to* $\mathcal{S}$ *denoted by* $\tilde{\mathscr{P}}_{\mathcal{S}}$ *is defined as the space of all the pseudo-marginal probability sets with respect to* $\mathcal{S}$*.*

Definition 13 (Pseudo-marginal delta-probability set) *Suppose* $\mathcal{Q} = \{q_s : \mathcal{L}^{|s|} \to \mathbb{R}, \forall s \in \mathcal{S}\}$ *is a pseudo-marginals-set which, thus, satisfies identities (16). If, moreover,* $\mathcal{Q}$ *verifies the identities:*

$$\forall s \in \mathit{Front}(\mathcal{S}), \begin{cases} \sum_{x_s \in \mathcal{L}^{|s|}} q_s(x_s) = 0 \\ \sum_{x_s \in \mathcal{L}^{|s|}} \big|q_s(x_s)\big| \leq 2 \end{cases} \tag{21}$$

then $\mathcal{Q}$ *is called a pseudo-marginal delta-probability set with respect to* $\mathcal{S}$*.*

Definition 14 (pseudo-marginal delta-polytope) *The pseudo-marginal delta-polytope with respect to* $\mathcal{S}$ *denoted by* $\tilde{\mathscr{Q}}_{\mathcal{S}}$ *is defined as the space of all the pseudo-marginal delta-probability sets with respect to* $\mathcal{S}$*.*

Remark 3 *Let us note that the system of identities that defines either a pseudo-marginal probability set, or a pseudo-marginal delta-probability set necessarily presents many redundancies, thus, making it prone to further simplifications. For the sake of example, by taking into account the developed arguments in section 5, one may see immediately that the identities of the form* $\sum_{x_s \in \mathcal{L}^{|s|}} p_s(x_s) = 1, \forall s \in \mathit{Front}(\mathcal{S})$ *may be reduced to a single identity of the form* $\sum_{x_{s_0} \in \mathcal{L}^{|s_0|}} p_{s_0}(x_{s_0}) = 1$*, equally, the identities of the form* $\sum_{x_s \in \mathcal{L}^{|s|}} q_s(x_s) = 0, \forall s \in \mathit{Front}(\mathcal{S})$ *may be reduced to a single identity of the form* $\sum_{x_{s_0} \in \mathcal{L}^{|s_0|}} q_{s_0}(x_{s_0}) = 0$*, with* s_0 *standing for an arbitrary hypersite in* $\mathcal{S}$*, and so on. Nevertheless, for the sake of simplicity, we will not proceed to such simplifications in this paper, though, the latter may turn out to be desirable in practice, above all, for bigger values of* n*.*

6.2 Relaxation

One proceeds in a traditional way for obtaining LP relaxations of EMinMLP (8) and EMaxMLP (9), hence of MinMPP (2) and MaxMPP (3), respectively, by just enforcing locally the probability axioms, as follows:

$$\min_{\{p_s : \mathcal{L}^{|s|} \to \mathbb{R}, \forall s \in \mathcal{S}\} \in \tilde{\mathscr{P}}_{\mathcal{S}}} \Big\{ \sum_{s \in \mathcal{S}} \sum_{x_s \in \mathcal{L}^{|s|}} g_s(x_s) \, p_s(x_s) \Big\} \tag{22}$$

and

$$\begin{array}{c}\max\Big\{\ \sum_{s\in\mathcal{S}}\sum_{x_s\in\mathcal{L}^{|s|}} g_s(x_s)\, p_s(x_s)\ \Big\}\\ \big\{p_s:\mathcal{L}^{|s|}\to\mathbb{R},\forall s\in\mathcal{S}\big\}\in\tilde{\mathscr{P}}_{\mathcal{S}}\end{array}\tag{23}$$

where $\tilde{\mathscr{P}}_{\mathcal{S}}$ stands for the pseudo-marginal polytope (see Definition 12).

Equally, one may obtain a useful LP relaxation of DEMinMLP (13), hence of ModesMPP (4), by just enforcing just enforcing locally the delta-probability axioms, as follows:

$$\begin{array}{c}\min\Big\{\ \sum_{s\in\mathcal{S}}\sum_{x_s\in\mathcal{L}^{|s|}} g_s(x_s)\, q_s(x_s)\ \Big\}\\ \big\{q_s:\mathcal{L}^{|s|}\to\mathbb{R},\forall s\in\mathcal{S}\big\}\in\tilde{\mathscr{Q}}_{\mathcal{S}}\end{array}\tag{24}$$

where $\tilde{\mathscr{Q}}_{\mathcal{S}}$ stands for the pseudo-marginal delta-polytope (see Definition 14.

In the remainder, we refer to LP (22), LP (23), and LP (24) using the acronyms PseudoEMinMLP, PseudoEMaxMLP, and PseudoΔEMinMLP, respectively. One then easily checks that all of PseudoEMinMLP (22), PseudoEMaxMLP (23), and PseudoΔEMinMLP (24) are bounded, moreover, they constitute a lower-bound for EMinMLP (8), an upper bound for EMaxMLP (9), and a lower bound for DEMinMLP (13), respectively.

7. Optimality study of the LP relaxations

This section is divided into two main subsections. First, subsection 7.1 develops equivalent global reformulations of the described LP relaxations in section 5, thereby, setting the stage for their optimality study in subsection 7.2.

7.1 Global reformulation of the LP relaxations

The main result in this section regarding the equivalent global reformulation of PseudoEMinMLP (22), PseudoEMaxMLP (23), and PseudoΔEMinMLP (24) is highlighted in Theorem 6 below.

Theorem 6 *1. PseudoEMinMLP (22) is equivalent to the following LP:*

$$\begin{array}{l}\min\Big\{\ \sum_{x\in\mathcal{L}^n} g(x)\,p(x)\ \Big\}\\ \begin{cases}\sum_{i\notin s}\sum_{x_i\in\mathcal{L}} p(x_1,\ldots,x_i,\ldots,x_n)\geq 0,\ \forall x_s\in\mathcal{L}^{|s|},\ \forall s\in \mathit{Front}(\mathcal{S})\\ \sum_{x\in\mathcal{L}^n} p(x)=1\end{cases}\end{array}\tag{25}$$

2. PseudoEMaxMLP (23) is equivalent to the following LP:

$$\begin{array}{l}\max\Big\{\ \sum_{x\in\mathcal{L}^n} g(x)\,p(x)\ \Big\}\\ \begin{cases}\sum_{i\notin s}\sum_{x_i\in\mathcal{L}} p(x_1,\ldots,x_i,\ldots,x_n)\geq 0,\ \forall x_s\in\mathcal{L}^{|s|},\ \forall s\in \mathit{Front}(\mathcal{S})\\ \sum_{x\in\mathcal{L}^n} p(x)=1\end{cases}\end{array}\tag{26}$$

3. PseudoΔEMinMLP (24) is equivalent to the following LP:

$$\begin{array}{l}\min\Big\{\ \sum_{x\in\mathcal{L}^n} g(x)\,q(x)\ \Big\}\\ \begin{cases}\sum_{x_s\in\mathcal{L}^n}\big|\sum_{i\notin s}\sum_{x_i\in\mathcal{L}} q(x_1,\ldots,x_i,\ldots,x_n)\big|\leq 2,\ \forall s\in \mathit{Front}(\mathcal{S})\\ \sum_{x\in\mathcal{L}^n} q(x)=0\end{cases}\end{array}\tag{27}$$

in the sense that any of the global LP reformulations above:

1. *achieves the same optimal objective value as its local reformulation counterpart,*
2. *the margins set with respect to $\mathcal{S}$ of any of its feasible solutions is a feasible solution of its local reformulation counterpart,*
3. *conversely, whatever a feasible solution of its local counterpart, any function $f : \mathcal{L}^n \to \mathbb{R}$ of which margins set with the respect to $\mathcal{S}$ is its feasible solution and achieves an objective value equal to the one achieved by the former in its local counterpart.*

The proof of Theorem 6 is sketched in Appendix section A.7

Throughout, we refer to LP (25), LP (26), and LP (27) using the acronyms GlbPseudoEMinMLP, GlbPseudoEMaxMLP, and GlbPseudo∆EMinMLP, respectively.

7.2 Main optimality results

One begins by observing an interesting phenomenon which is as follows. First of all, consider the LP which stands for the difference of GlbPseudoEMinMLP (25) and GlbPseudoEMaxMLP (26), in that order, as follows:

$$\min\Big\{ \textstyle\sum_{x\in\mathcal{L}^n} g(x)\,\big(p(x)-p'(x)\big) \Big\}$$
$$\begin{cases} \sum_{i\notin s}\sum_{x_i\in\mathcal{L}} p(x) \geq 0,\ \forall x_s \in \mathcal{L}^{|s|},\ \forall s \in \mathrm{Front}(\mathcal{S}) \\ \sum_{i\notin s}\sum_{x_i\in\mathcal{L}} p'(x) \geq 0,\ \forall x_s \in \mathcal{L}^{|s|},\ \forall s \in \mathrm{Front}(\mathcal{S}) \\ \sum_{x\in\mathcal{L}^n} p(x) = 1 \\ \sum_{x\in\mathcal{L}^n} p'(x) = 1 \end{cases} \tag{28}$$

Clearly, solving both GlbPseudoEMinMLP (25) and GlbPseudoEMaxMLP (26) amounts to solving LP (28) once, and vice-versa, this is on the one hand. On the other hand, suppose q is a feasible solution of GlbPseudo∆EMinMLP (27), and p and p' are feasible solutions of GlbPseudoEMinMLP (25) and GlbPseudoEMaxMLP (26), respectively, hence of LP (28) too. It follows, by Proposition 2, that both q and $p-p'$ are, at worst, up to a multiplicative scale (greater than, or equal to 1), feasible solutions of DEMinMLP (13). But, since only their respective orthogonal-projection parts, namely $\mathscr{O}_{\mathcal{S}}q$ and $\mathscr{O}_{\mathcal{S}}(p-p')$ are, in fact, effective in GlbPseudo∆EMinMLP (27) and LP (28), respectively, as one may write:

$$\begin{cases} \langle g, q\rangle = \langle g, \mathscr{O}_{\mathcal{S}}q\rangle \\ \langle g, p-p'\rangle = \langle g, \mathscr{O}_{\mathcal{S}}(p-p')\rangle \end{cases}$$

plus, by Theorem 5, the margins with respect to $\mathrm{Front}(\mathcal{S})$ of $\mathscr{O}_{\mathcal{S}}q$ coincide with the ones of q, and the margins with respect to $\mathrm{Front}(\mathcal{S})$ of $\mathscr{O}_{\mathcal{S}}(p-p')$ coincide with the ones of $p-p'$, then, in the light of the result of Theorem 3, one would want to know to which extent at least one of the following two max-min problems:

$$\max_{p,p'}\Big\{ \min_q \big\{ \textstyle\sum_{x\in\mathcal{L}^n} |q(x)| \big\} \Big\}$$
$$\begin{cases} p \text{ is a feasible of GlbPseudoEMinMLP (25)} \\ p' \text{ is a feasible of GlbPseudoEMaxMLP (26)} \\ \mathscr{O}_{\mathcal{S}}q = \mathscr{O}_{\mathcal{S}}(p-p') \end{cases} \tag{29}$$

and

$$\begin{gathered}\max_{q'}\Big\{\ \min_q\big\{\ \textstyle\sum_{x\in\mathcal{L}^n}|q(x)|\ \big\}\ \Big\}\\ \begin{cases} q' \text{ is a feasible of GlbPseudo}\Delta\text{EMinMLP (27)}\\ \mathscr{O}_{\mathcal{S}}q=\mathscr{O}_{\mathcal{S}}q'\end{cases}\end{gathered} \tag{30}$$

achieves an optimal objective which is equal to 2, as by Theorem 3, this would immediately imply that one may efficiently solve any HoMPP instance by means of its LP relaxation. Furthermore, it is easy to check that max-min problem (29) is an upper bound for max-min problem (30), implying that, if the latter achieves an optimal objective value which is equal to 2, then the former will also achieve an optimal objective value which is equal to 2. But nevertheless, we will establish, hereafter, two separate results for each of max-min problems (29) and (30) above, just in order to stress on the fact that, for finding the modes of g, one actually has the choice between solving two LP instances, namely, PseudoEMinMLP (22) and PseudoEMaxMLP (23), or solving a single LP instance, namely, PseudoΔEMinMLP (24), as both choices above turn out to be equivalent.

Theorem 7 *The optimal objective value of max-min problem (29) is equal to* 2.

Theorem 8 *The optimal objective value of max-min problem (30) is equal to* 2.

The proofs of Theorem 8 and Theorem 7 are described in Appendix sections A.8 and A.9, respectively.

In short, Theorem 7 and Theorem 8 establish exactness of the claim that we have just made above which is that both feasible sets of PseudoΔEMinMLP (24) and LP (28) are within the "tolerance interval" which is allowed by Theorem 3 in order to hope solving the HoMPP by means of its LP relaxation. Said otherwise, by taking into account the arguments that we have developed above, either result of Theorem 7 or of Theorem 8 is enough to guarantee that one may completely solve the HoMPP by means of PseudoEMinMLP (22) (equivalently, by means of PseudoEMaxMLP (23)) or by means of PseudoΔEMinMLP (24). Therefore, we summarize the latter findings in Theorem 9 and Theorem 10 below.

Theorem 9 *PseudoEMinMLP (22) completely solves EMinMLP (8), in the sense that:*

1. *they both achieve the same optimal objective value equal to* $\inf_{x\in\mathcal{L}^n}\big\{g(x)\big\}$,
2. *any optimal solution* $\big\{p_s^*,\forall s\in\mathcal{S}\big\}$ *of PseudoEMinMLP (22) defines an actual marginals-set with respect to* $\mathcal{S}$ *which is originated from a joint distribution* p^* *of RMLV* $\mathcal{X}$ *which is optimal for EMinMLP (8).*

Similar conclusions as above are, obviously, drawn regarding PseudoEMaxMLP (23), on the one hand, and EMaxMLP (9), on the other hand, which then achieve an optimal objective value equal to $\sup_{x\in\mathcal{L}^n}\big\{g(x)\big\}$.

The proof of Theorem 9 is described in Appendix section A.10.

Theorem 10 *PseudoΔEMinMLP (24) exactly solves DEMinMLP (13), in the sense that:*

1. *PseudoΔEMinMLP (24) achieves the same optimal objective value as DEMinMLP (13), which is equal to* $\inf_{x\in\mathcal{L}^n}\big\{g(x)\big\} - \sup_{x\in\mathcal{L}^n}\big\{g(x)\big\}$,

2. *any optimal solution* $\big\{q_s^*, \forall s\in\mathcal{S}\big\}$ *of PseudoΔEMinMLP (24) defines an actual delta-marginals-set with respect to* $\mathcal{S}$ *which is originated from a joint delta-distribution* q^* *of RMLV* $\mathcal{X}$ *which is optimal for DEMinMLP (13).*

The proof of Theorem 10 is described in Appendix section A.11.

8. Computation of a full integral MAP solution of the HoMPP

It might be the case that a HoMPP instance has multiple MAP solutions (i.e.; g might have multiple minima and/or multiple maxima), thus, the resolution either of PseudoEMinMLP (22), or PseudoEMaxMLP (23) (resp. of PseudoΔEMinMLP (24)) might only yield the marginals with respect to $\mathcal{S}$ of a fractional (i.e.; non-binary) optimal distribution (resp. a fractional (i.e.; non-signed binary) delta-distribution) happening to be some convex combination of optimal binary distributions (resp. delta distributions). Thus in such a case, one moreover needs join the pieces in order to obtain an full MAP solution of a HoMPP instance. Therefore, the goal in the remainder of this section is to address the latter problem under general assumptions about a HoMPP instance.

8.1 Theory

For the sake of example, assume PseudoEMinMLP (22) of which resolution has yielded an optimal solution denoted by $\mathcal{P}^* = \big\{p_s^*, \forall s\in\mathcal{S}\big\}$. Then by Theorem 9, $\mathcal{P}^*$ stands for a set of marginal distributions of $\mathcal{X}_s, \forall s\in\mathcal{S}$ being originated from a joint distribution of RMLV $\mathcal{X}$ denoted by p^* which is, thus, optimal for EMinMLP (8). Moreover, by Theorem 1, obtaining a full optimal solution of MinMPP (2) amounts to obtaining a sample x^0 from p^*, however, for the sake of computational efficiency, one wants to avoid accessing p^* (which is hard). Therefore, in the remainder of this section, we describe an approach for computing a sample of p^* directly from $\mathcal{P}^*$.

Then, a first naive (yet, polynomial-time) algorithm for achieving the aforementioned goal is based on the result of Proposition 5 below.

Proposition 5 *Suppose* $s\in$ *Front*$(\mathcal{S})$ *and* $x_s^0\in\mathcal{L}^{|s|}$*, such that,* $\mathbb{P}\big(\mathcal{X}_s = x_s^0\big) > 0$*. Then, there exists* $x\in\mathcal{L}^n$*, such that,* $x_s = x_s^0$ *and* $\mathbb{P}\big(\mathcal{X} = x\big) > 0$*.*

Proof 3 *Such a result of Proposition 5 follows immediately from the identity:*

$$\mathbb{P}\big(\mathcal{X}_s = x_s\big) = \textstyle\sum_{i\notin s}\sum_{x_i\in\mathcal{L}}\mathbb{P}\big(\mathcal{X} = x\big),\, \forall x_s\in\mathcal{L}^{|s|}$$

as otherwise, i.e.; if $\forall x\in\mathcal{L}^n$*, such that,* $x_s = x_s^0$*, one had* $\mathbb{P}\big(\mathcal{X} = x\big) = 0$*, then one would have* $\mathbb{P}\big(\mathcal{X}_s = x_s^0\big) = 0$*, which is a contradiction with the assumption that* $\mathbb{P}\big(\mathcal{X}_s = x_s^0\big) > 0$*.*

Based on such a result of Proposition 5, one may proceed as follows. Suppose $s\in$ Front$(\mathcal{S})$ and $x_s^0\in\mathcal{L}^{|s|}$, such that, $\mathbb{P}\big(\mathcal{X}_s = x_s^0\big) = p_s^*(x_s^0) > 0$. Thus, if one replaced in g the value of x_s with its optimal value x_s^0, solved a new instance of PseudoEMinMLP (22) accordingly, and repeated this procedure with respect to some $s^{(1)} \in \text{Front}(\mathcal{S})/\{s\}$, then with respect to some $s^{(2)} \in \text{Front}(\mathcal{S})/\{s, s^{(1)}\}$, and so on, until all the variables x_i of g are exhausted, one would be guaranteed to ultimately obtain in polynomial-time a full mode of g. Obviously, such an algorithm is utterly slow, as it requires solving multiple instances of PseudoEMinMLP (22), successively (yet, with less variables each time). On the other hand, sampling from a general probability distribution by sole access to its marginals is not a straightforward procedure. Fortunately, as it will be shown hereafter, distributions of RMLV $\mathcal{X}$ which are candidates for optimality in EMinMLP (8) (equivalently, in EMaxMLP (9)) are not any (see Proposition 6 below), thereby, making it possible to efficiently compute their samples by sole access to their marginals-sets with respect to $\text{Front}(\mathcal{S})$.

Let us then begin by introducing the sign function defined as:

$$\forall a \in \mathbb{R}, \, \text{sign}(a) = \begin{cases} -1, & \text{if } a < 0, \\ +1, & \text{if } a > 0, \\ 0, & \text{if } a = 0. \end{cases}$$

as well as the indicator functions:

$$\forall a \in \mathbb{R}, \, \mathbf{1}_0(a) = \begin{cases} 1, & \text{if } a = 0, \\ 0, & \text{else.} \end{cases}, \, \mathbf{1}_+(a) = \begin{cases} 1, & \text{if } a > 0, \\ 0, & \text{else.} \end{cases}, \, \mathbf{1}_-(a) = \begin{cases} 1, & \text{if } a < 0, \\ 0, & \text{else.} \end{cases}$$

Proposition 6 *Suppose q^* is an optimal solution of DEMinMLP (13). Then, whatever a function $v : \mathcal{L}^n \to \mathbb{R}$, one has:*

$$\big\langle \big|v_{\bar{\mathscr{M}}_{\mathcal{S}}}\big| \, , \, \mathbf{1}_0(q^*) \big\rangle \geq \big| \big\langle v_{\bar{\mathscr{M}}_{\mathcal{S}}} \, , \, sign(q^*) \big\rangle \big|$$

where $\langle \cdot \, , \, \cdot \rangle$ stands for the scalar product, $v_{\bar{\mathscr{M}}_{\mathcal{S}}} \equiv v - \mathscr{O}_{\mathcal{S}} v$, finally, the functions $\mathbf{1}_0(q^*)$ and $sign(q^*)$ are defined as:

$$\forall x \in \mathcal{L}^n, \, \begin{cases} \mathbf{1}_0(q^*)(x) = \mathbf{1}_0\big(q^*(x)\big) \\ sign(q^*)(x) = sign\big(q^*(x)\big) \end{cases}$$

The proof of Proposition 6 is described in Appendix section A.13.

In a nutshell, such a result of Proposition 6 says that only "sparse enough" joint delta-distributions of RMLV $\mathcal{X}$ are potential candidates for optimality in DEMinMLP (13) (please note that it may be easily shown that such a statement would not apply if g was not a non-higher-order function). Furthermore, in the light of the result of Theorem 3 which establishes that any optimal solution of DEMinMLP (13) denoted by q^* decomposes as the difference of two joint distributions denoted by p^{+*} and p^{-*} having disjoint supports, and which are respectively optimal for EMinMLP (8) and EMaxMLP (9), one derives that either joint distribution (i.e.; p^{+*} or p^{-*}) is necessarily more sparse than q^*. With that being said, the remainder of this section is devoted to the elaboration of an efficient deterministic sampling algorithm from an optimal distribution (resp. an optimal delta-distribution) of RMLV $\mathcal{X}$ by sole access to its marginals-set with respect to $\text{Front}(\mathcal{S})$.

We shall then make a mild assumption in the remainder about higher-order function g, without undermining anyhow generality, as such an assumption just attempts to make sure that there is no redundancy between the local functions $g_s, \forall s \in \text{Front}(\mathcal{S})$ which form g. Said otherwise, for $\forall s \in \text{Front}(\mathcal{S})$, g_s cannot fully express in terms of remaining terms of g, i.e.; $g_{s'}, \forall s' \in \text{Front}(\mathcal{S})/\{s\}$. Nevertheless, let us claim upfront that the presence of such redundancy in g does not undermine anyhow generality of the MAP solution finding algorithm that we develop further in this section, as such redundancy may be easily unraveled and coped with accordingly as it shall be explained in more details, hereafter[2].

Definition 15 (Atomicity of a local function) *Suppose a hypersite $s \in \mathit{Front}(\mathcal{S})$, and introduce the hypersite-set:*

$$\mathcal{C}^{(s)} = \left\{ s' \cap s, \forall s' \in \mathit{Front}(\mathcal{S})/\{s\} \right\}$$

standing for the hypersite-set which consists of all the nonempty intersections between s and each of the remaining hypersites in $\mathit{Front}(\mathcal{S})/\{s\}$. Furthermore, if $\mathcal{C}^{(s)}$ is a nonempty hypersite-set, then introduce the ortho-marginal operator $\mathscr{O}_{\mathcal{C}^{(s)}}$ defined over the (local) function space $\left\{f : \mathcal{L}^{|s|} \to \mathbb{R}\right\}$, and denote by $\mathscr{M}_{\mathcal{C}^{(s)}}$ the ortho-marginal space with respect to $\mathcal{C}^{(s)}$, i.e.; :

$$\mathscr{M}_{\mathcal{C}^{(s)}} = \left\{f : \mathcal{L}^{|s|} \to \mathbb{R},\ s.t., \mathscr{O}_{\mathcal{C}^{(s)}} f = f\right\}$$

Then, one says that a local function $h_s : \mathcal{L}^{|s|} \to \mathbb{R}$ is atomic in $\mathit{Front}(\mathcal{S})$, if and only if, either $\mathcal{C}^{(s)}$ is the empty hypersite-set, or $\mathcal{C}^{(s)}$ is a nonempty hypersite-set and $h_s \notin \mathscr{M}_{\mathcal{C}^{(s)}}$.

Assumption 1 *Higher-order function g (1) verifies that $\forall s \in \mathit{Front}(\mathcal{S})$, g_s is atomic in $\mathit{Front}(\mathcal{S})$.*

Theorem 11 *Suppose $\left\{q_s^*, \forall s \in \mathcal{S}\right\}$ is an optimal solution of PseudoΔEMinMLP (24). Then, under Assumption 1, one has:*

$$\textstyle\sum_{x_s \in \mathcal{L}^{|s|}} |q_s^*(x_s)| = 2,\ \forall s \in \mathit{Front}(\mathcal{S}) \tag{31}$$

The proof of Theorem 11 is detailed in appendix section A.12.

In a nutshell, such a result of Theorem 11 says that under Assumption 1, and at optimality of PseudoΔEMinMLP (24), all the marginal delta-distributions with respect to $\text{Front}(\mathcal{S})$ must admit a unique decomposition in terms of the difference of two marginal distributions (see Proposition 1 for more details). Still under the same assumption, an immediate consequence of such a result of Theorem 11 is that, whatever a minimum solution $x^{(inf)}$ of g, and whatever its maximum solution $x^{(sup)}$, one necessarily has $\forall s \in \text{Front}(\mathcal{S})$, "$x_s^{(inf)} \neq x_s^{(sup)}$" standing for "$\exists i \in s$, such that, $x_i^{(inf)} \neq x_i^{(sup)}$".

Theorem 12 *Let p^{+*} and p^{-*} be two distributions of RMLV $\mathcal{X}$ which are optimal for EMinMLP (8) and EMaxMLP (9), respectively, and denote by $\left\{p_s^{+*}, \forall s \in \mathcal{S}\right\}$ and $\left\{p_s^{-*}, \forall s \in \mathcal{S}\right\}$ their respective marginals-sets with respect to $\mathcal{S}$. Then, under Assumption 1:*

- *in order for any $x \in \mathcal{L}^n$ to be a sample of p^{+*} (i.e.; a minimum solution of g, such that, $p^{+*}(x) > 0$), it is necessary and sufficient that:*

$$p_s^{+*}(x_s) > 0,\ \forall s \in \mathit{Front}(\mathcal{S})$$

2. Let us say, upfront, by simply amending $\text{Front}(\mathcal{S})$, i.e.; by removing from it all such hypersites corresponding to the terms of g which present such redundancy.

- *in order for any* $x \in \mathcal{L}^n$ *to be a sample of* p^{-*} *(i.e.; a maximum solution of* g*, such that,* $p^{-*}(x) > 0$*), it is necessary and sufficient that:*

$$p_s^{-*}(x_s) > 0, \, \forall s \in \mathit{Front}(\mathcal{S})$$

Theorem 13 *Suppose* q^* *is a delta-distribution of RMLV* $\mathcal{X}$ *which is optimal for DEMinMLP (13), and denote by* $\{q_s^*, \forall s \in \mathcal{S}\}$ *its delta-marginals-set with respect to* $\mathcal{S}$*. Then, under Assumption 1:*

- *in order for any* $x \in \mathcal{L}^n$ *to be a Inf-sample of* q^* *(i.e.; a minimum solution of* g*, such that,* $q^*(x) > 0$*), it is necessary and sufficient that:*

$$q_s^*(x_s) > 0, \, \forall s \in \mathit{Front}(\mathcal{S})$$

- *in order for any* $x \in \mathcal{L}^n$ *to be a Sup-sample of* q^* *(i.e.; a maximum solution of* g*, such that,* $q^*(x) < 0$*), it is necessary and sufficient that:*

$$q_s^*(x_s) < 0, \, \forall s \in \mathit{Front}(\mathcal{S})$$

- *in order for any* $x \in \mathcal{L}^n$ *to be a sample of* q^* *(i.e.; a mode of* g*, such that,* $q^*(x) \neq 0$*), it is necessary and sufficient that:*

$$q_s^*(x_s) \neq 0, \, \forall s \in \mathit{Front}(\mathcal{S})$$

Theorem 12 and 13 are jointly shown in Appendix section A.14, as they turn out to be both sides of a same coin under Assumption 1.

8.2 Algorithm

Let us introduce the following definition.

Definition 16 (Consistent sub-samples) *Suppose* $\mathcal{C}$ *is hypersite-set with respect to* Ω*, and denote by* $\mathit{Front}(\mathcal{C})$ *its frontier hypersite-set. Furthermore, suppose* $\mathcal{P} = \{p_s, \forall s \in \mathcal{C}\}$ *is a set of probability marginals with respect to* $\mathcal{C}$*. Then, one says that a set of integer vectors of the form* $\{x_s^{(s)} \in \mathcal{L}^{|s|}, \forall s \in \mathit{Front}(\mathcal{C})\}$ *is a consistent sub-samples-set with respect to* $\mathcal{P}$*, if and only if:*

1. $p_s(x_s^{(s)}) > 0, \, \forall s \in \mathit{Front}(\mathcal{C})$,
2. $\exists x \in \mathcal{L}^n$*, such that,* $x_s = x_s^{(s)}, \, \forall s \in \mathit{Front}(\mathcal{C})$*, which also amounts to saying that:* $\forall s_1, s_2 \in \mathit{Front}(\mathcal{C}),\, s_1 \neq s_2 \wedge s_1 \cap s_2 \neq \emptyset \Rightarrow x_{s_1 \cap s_2}^{(s_1)} = x_{s_1 \cap s_2}^{(s_2)}$.

Now, suppose $\mathcal{P}^{+*} = \{p_s^{+*}, \forall s \in \mathcal{S}\}$ and $\mathcal{P}^{-*} = \{p_s^{-*}, \forall s \in \mathcal{S}\}$ are the optimal solutions of PseudoEMinMLP (22) and PseudoEMaxMLP (23), respectively, and denote by p^{+*} and p^{-*} the respective joint distributions of RMLV $\mathcal{X}$ from which $\mathcal{P}^{+*}$ and $\mathcal{P}^{-*}$ are originated. Clearly, by Theorem 12, generation of samples from p^{+*} (resp. from p^{-*}) which also happen to be minimum (resp. maximum) solutions of g is equivalent to finding consistent sub-samples-sets with respect to $\mathcal{P}^{+*}$ (resp. $\mathcal{P}^{-*}$). Then, Theorem 14 below builds on both results of Proposition 6 and Theorem 12 for establishing that one may achieve such a goal in a totally greedy and causal way.

For the sake of clarity in the remainder, we shall assume some (arbitrary) order on the hypersites of $\mathrm{Front}(\mathcal{S})$, and let us denote $\mathrm{Front}(\mathcal{S}) = \{s_1, \ldots, s_m\}$, with $m = |\mathrm{Front}(\mathcal{S})|$.

Theorem 14 *Let $\mathcal{P}^*$ either stand for $\mathcal{P}^{+*}$ or $\mathcal{P}^{-*}$, and put $\mathcal{P}^* = \{p_s^*, \forall s \in \mathcal{S}\}$. Furthermore, suppose $k \in \{1, \ldots, m-1\}$, and let $\mathcal{P}_k = \{p_{s_i}^*, \forall i = 1, \ldots, k\}$ and $\mathcal{P}_{k+1} = \{p_{s_i}^*, \forall i = 1, \ldots, k+1\}$ stand for two marginals subset of $\mathcal{P}^*$, moreover, suppose that $\{x_{s_i}^{(i)} \in \mathcal{L}^{|s_i|}, \forall i = 1, \ldots, k\}$ is a consistent sub-samples-set with respect to $\mathcal{P}_k$. Then, under Assumption 1, there exists $x_{s_{k+1}}^{(k+1)} \in \mathcal{L}^{|s_{k+1}|}$, such that, $\{x_{s_i}^{(i)} \in \mathcal{L}^{|s_i|}, \forall i = 1, \ldots, k+1\}$ is a consistent sub-samples-set with respect to $\mathcal{P}_{k+1}$.*

The proof of Theorem 14 is described in Appendix section A.15.

In plain words, such a result of Theorem 14 says:

- first, that under Assumption 1, one may construct in a totally greedy, yet, causal way samples of any optimal joint distribution of RMLV $\mathcal{X}$, solely, by means of its marginals-set with respect to $\mathrm{Front}(\mathcal{S})$,

- second, that one may terminate the search as soon as one finds a consistent sub-samples-set with respect to a marginals subset, in this case, $\mathcal{P}_1^* = \{p_s, \forall s \in \mathcal{C}\}$, for some $\mathcal{C} \subseteq \mathrm{Front}(\mathcal{S})$, such that, $\cup_{s \in \mathcal{C}} \{i \in s\} = \Omega$, because, all the variables $x_i, \forall i \in \Omega$ are then set.

Then, Algorithm 2 which is described in Appendix section B.17 sketches such a greedy computation of either mode of g, solely, by using the marginals-set $\mathcal{P}^*$ which is originated from an optimal joint distribution of RMLV $\mathcal{X}$. Clearly, if one chooses to solve instead PseudoΔEMinMLP (24), then such an algorithm 2 can be effortlessly transcribed in terms of the delta-marginals-set which is originated from an optimal joint delta-distribution of RMLV $\mathcal{X}$.

9. Conclusion

We have introduced two novel algebraic concepts, namely, the notion of joint delta-distribution, and the ortho-marginal framework, which later have enabled us to show the main result in this paper regarding the exact resolution of the MAP inference problem in general higher-order graphical models by means of a traditional LP relaxation over the local marginal polytope. Furthermore, we have thoroughly studied the case where the found optimal solution is not originated from a binary distribution (resp. signed binary delta-distribution), and proposed an efficient deterministic sampling algorithm of (integral) MAP solutions directly from perhaps non-binary optimal marginal distributions (resp. optimal marginal delta-distributions). Let us emphasize, nevertheless, that the proposed LP relaxation in this paper is still prone to further accelerations, but which we have not reported in this paper because of lack of space. Such accelerations may concern, for instance, the (logarithmic) reduction of the number of LP variables for bigger values of the number of labels L, as well as other possible variable simplifications, assuming for instance the MRF framework. Therefore, as future work, we envisage to investigate such possible method accelerations, and to report them, accordingly, in subsequent papers. Last but not least, another possible opening of the main result in this paper concerns method extension to the continuous-domain case for solving some classes of continuous computer vision and pattern recognition problems.

Proofs

A.1 Proof of Theorem 2

Proof 4 *Let us first show the necessary condition of Theorem 2. Thus, suppose q is a joint delta-distribution of RMLV $\mathcal{X}$, hence admitting a decomposition of the form:*

$$q(x) = p(x) - p'(x),\ \forall x \in \mathcal{L}^n \tag{32}$$

where both p and p' stand for joint distributions of RMLV $\mathcal{X}$. Then, one has

$$\begin{aligned} \textstyle\sum_{x\in\mathcal{L}^n} q(x) &= \textstyle\sum_{x\in\mathcal{L}^n} \big(p(x) - p'(x)\big) = \sum_{x\in\mathcal{L}^n} p(x) - \sum_{x\in\mathcal{L}^n} p'(x) = 0 \\ \textstyle\sum_{x\in\mathcal{L}^n} |q(x)| &= \textstyle\sum_{x\in\mathcal{L}^n} \big|p(x) - p'(x)\big| \leq \sum_{x\in\mathcal{L}^n} p(x) + \sum_{x\in\mathcal{L}^n} p'(x) = 2 \end{aligned}$$

Second, in order to show the sufficient condition of Theorem 2, suppose a function $q : \mathcal{L}^n \to \mathbb{R}$ verifying:

$$\begin{cases} \sum_{x\in\mathcal{L}^n} q(x) = 0 \\ \sum_{x\in\mathcal{L}^n} |q(x)| \leq 2 \end{cases}$$

We need the following two identities for the remaining part of the proof:

$$\forall a, b \in \mathbb{R},\ \inf\big\{a, b\big\} = \frac{a + b - |a - b|}{2},\ \sup\big\{a, b\big\} = \frac{a + b + |a - b|}{2} \tag{33}$$

Let us first show the following three claims:

1. $|q(x)| \leq 1,\ \forall x \in \mathcal{L}^n$,
2. $\sum_{x\in\mathcal{L}^n} \sup\big\{0, -q(x)\big\} \leq 1$,
3. $\sum_{x\in\mathcal{L}^n} \inf\big\{1, 1 - q(x)\big\} \geq 1$.

First, suppose $x \in \mathcal{L}^n$. Then, using the fact that $\sum_{y\in\mathcal{L}^n} q(y) = 0$, one finds that one has $q(x) = -\sum_{y\in\mathcal{L}^n/\{x\}} q(y)$, in such a way that, one may write:

$$|q(x)| \ = \frac{|q(x)| + |-\sum_{y\in\mathcal{L}^n/\{x\}} q(y)|}{2} \leq \frac{\sum_{y\in\mathcal{L}^n} |q(y)|}{2} \leq 1$$

Second, using the identity of the sup in formula (33), one has:

$$\textstyle\sum_{x\in\mathcal{L}^n} \sup\big\{0, -q(x)\big\} \ = \frac{1}{2} \sum_{x\in\mathcal{L}^n} (q(x) + |q(x)|) = \sum_{x\in\mathcal{L}^n} \frac{|q(x)|}{2} \leq 1$$

Third and last, using the identity of the inf in formula (33), one has:

$$\sum_{x\in\mathcal{L}^n} \inf\big\{1, 1-q(x)\big\} \quad = \sum_{x\in\mathcal{L}^n} \tfrac{2-q(x)-|q(x)|}{2} = L^n - \sum_{x\in\mathcal{L}^n} \tfrac{|q(x)|}{2} \geq L^n - 1 \geq 1$$

as, by assumption, both L and n are integers greater than 1, which thus shows the three claims above. Now, suppose two functions $p : \mathcal{L}^n \to \mathbb{R}$, and $p' : \mathcal{L}^n \to \mathbb{R}$, such that, $p(x) - p'(x) = q(x)$, $\forall x \in \mathcal{L}^n$. Thus, one may write: $p'(x) = p(x) + q(x)$, $\forall x \in \mathcal{L}^n$ and one easily checks that: $\sum_{x\in\mathcal{L}^n} p'(x) = \sum_{x\in\mathcal{L}^n} p(x)$, in such a way that, p and p′ define two joint distributions of RMLV $\mathcal{X}$, if and only if, the following linear system:

$$\begin{cases} 0 \leq p'(x) + q(x) \leq 1, \forall x \in \mathcal{L}^n \\ 0 \leq p'(x) \leq 1, \forall x \in \mathcal{L}^n \\ \sum_{x\in\mathcal{L}^n} p'(x) = 1 \end{cases}$$

equivalently, the following linear system:

$$\begin{cases} \sup\big\{0, -q(x)\big\} \leq p'(x) \leq \inf\big\{1, 1-q(x)\big\}, \forall x \in \mathcal{L}^n \\ \sum_{x\in\mathcal{L}^n} p'(x) = 1 \end{cases} \tag{34}$$

(where the unknown is p′) has a solution. First, since we have already shown that $\forall x \in \mathcal{L}^n$, $|q(x)| \leq 1$, by using both the identities of the sup and the inf in formula (33), one derives that one has $\forall x \in \mathcal{L}^n$:

$$\sup\big\{0, -q(x)\big\} - \inf\big\{1, 1-q(x)\big\} = \frac{(-q(x) + |q(x)|) - (2 - q(x) - |q(x)|)}{2} = |q(x)| - 1 \leq 0,$$

and one concludes that the system of linear inequalities:

$$\sup\big\{0, -q(x)\big\} \leq p'(x) \leq \inf\big\{1, 1-q(x)\big\}, \forall x \in \mathcal{L}^n$$

has a solution. Furthermore, in order for a solution p′ of the latter system to satisfy that $\sum_{x\in\mathcal{L}^n} p'(x) = 1$, it is necessary and sufficient that one has:

$$\begin{cases} \sum_{x\in\mathcal{L}^n} \sup\big\{0, -q(x)\big\} & \leq 1 \\ \sum_{x\in\mathcal{L}^n} \inf\big\{1, 1-q(x)\big\} & \geq 1 \end{cases}$$

which has already been shown to be true above, hence finally the proof of Theorem 2.

A.2 Proof of Proposition 2

Proof 5 *Suppose a nonzero function $q : \mathcal{L}^n \to \mathbb{R}$, such that, $\sum_{x\in\mathcal{L}^n} q(x) = 0$. Furthermore, suppose $\lambda > 0$, and denote $q_\lambda(x) = \lambda\, q(x)$, $\forall x \in \mathcal{L}^n$. One first checks that:*

$$\begin{cases} \sum_{x\in\mathcal{L}^n} \tilde{q}_\lambda(x) = \lambda \sum_{x\in\mathcal{L}^n} q(x) = 0 \\ \sum_{x\in\mathcal{L}^n} |\tilde{q}_\lambda(x)| = \lambda \sum_{x\in\mathcal{L}^n} |q(x)| \end{cases}$$

in such a way that, $\forall \lambda \in [0, \frac{2}{\sum_{x\in\mathcal{L}^n} |q(x)|}]$, one has: $\sum_{x\in\mathcal{L}^n} |q_\lambda(x)| \leq \frac{2\sum_{x\in\mathcal{L}^n} |q(x)|}{\sum_{x\in\mathcal{L}^n} |q(x)|} = 2$, which means that $\tilde{q}_\lambda$ does define a valid joint delta-distribution of RMLV $\mathcal{X}$, hence the proof of Proposition 2.

A.3 Proof of Proposition 1

Proof 6 *Suppose q is a non-zero joint delta-distribution of RMLV $\mathcal{X}$, and let us put $q = p - p'$, where p and p' stand for two ordinary joint distributions of RMLV $\mathcal{X}$. Then, we have already shown (see the proof in Appendix section A.1) that all the solutions of the couple (p, p') are given by:*

1. *p' is any solution of the linear system (34),*
2. *once p' is known, p is uniquely defined as: $p = p' + q$.*

Thus, denote $|q| = \sum_{x\in\mathcal{L}^n} |q(x)|$. One finds that the general formula of p' is given by:

$$\begin{cases} \forall x \in \mathcal{L}^n, \begin{cases} p'(x) & = \tau(x)\sup\big\{0, -q(x)\big\} + (1-\tau(x))\inf\big\{1, 1-q(x)\big\} \\ & = \big(1-|q(x)|\big)\tau(x) + \frac{-q(x)+|q(x)|}{2} \\ \tau(x) & \in [0,1] \end{cases} \\ \sum_{x\in\mathcal{L}^n} \big(1-|q(x)|\big)\,\tau(x) = 1 - \frac{|q|}{2} \end{cases} \tag{35}$$

in such a way that, since $|q| \le 2$, one has:

- *if $|q| = 2$, then all the possible solutions for τ are given by:*

$$\begin{cases} \big(1-|q(x)|\big)\,\tau(x) = 0,\ \forall x \in \mathcal{L}^n \\ \tau(x) \in [0,1],\ \forall x \in \mathcal{L}^n \end{cases}$$

  and one finds that the unique solution of (p, p') is given $\forall x \in \mathcal{L}^n$ by $p'(x) = \frac{-q(x)+|q(x)|}{2} = \sup\{0, -q(x)\}$, and $p(x) = \frac{q(x)+|q(x)|}{2} = \sup\{0, q(x)\}$.

- *if $|q| < 2$, then suppose an arbitrary $x_0 \in \mathcal{L}^n$, and one finds, for instance, that $\forall x \in \mathcal{L}^n$, $\tau_1(x) = \tau_1 = \frac{1-\frac{|q|}{2}}{L^n - |q|}$, and $\tau_2(x) = \begin{cases} \frac{1-\frac{|q|}{2}}{L^n - 1 + |q(x_0)| - |q|}, \text{if } x \neq x_0, \\ 0, \text{if } x = x_0. \end{cases}$ are two possible solutions of τ, and hence $p_1(x) = \sup\{0, q(x)\} + \tau_1\big(1-|q(x)|\big)$, $p_1'(x) = \sup\{0, -q(x)\} + \tau_1\big(1-|q(x)|\big)$, and $p_2(x) = \sup\{0, q(x)\} + \tau_2(x)\,\big(1-|q(x)|\big)$, $p_2'(x) = \sup\{0, -q(x)\} + \tau_2(x)\,\big(1-|q(x)|\big)$, are two possible solutions of (p, p').*

hence the proof of Proposition 1.

A.4 Proof of Theorem 4

We show Theorem (4) by construction, using successive projections of an arbitrary function via marginalization. For the sake of clarity, let us put $m = |\mathrm{Front}(\mathcal{C})|$, and assume an arbitrary (e.g.; a lexicographic order) on the elements of $\mathrm{Front}(\mathcal{C})$, in a way which enables to write $\mathrm{Front}(\mathcal{C}) = \{c_1, \ldots, c_m\}$. Next, suppose an arbitrary function $f : \mathcal{L}^n \to \mathbb{R}$, and let us write:

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

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:

$$f^{(i+1)}(x) = f^{(i)}(x) - \frac{f^{(i)}_{c_{i+1}}(x_{c_{i+1}})}{L^{n-|c_{i+1}|}},\ \forall x \in \mathcal{L}^n,\ \forall i \in \big\{1,\dots,m-1\big\} \tag{36}$$

where $\forall i \in \big\{1,\dots,m-1\big\}$, $f^{(i)}_{c_{i+1}}$ 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:

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

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:

$$f^{(i)}_{c_j}(x_{c_j}) = 0,\ \forall x_{c_j} \in \mathcal{L}^{|c_j|},\ \forall j \in \big\{1,\dots,i\big\} \tag{37}$$

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

$$f^{(i+1)}_{c_j}(x_{c_j}) = 0,\ \forall x_{c_j} \in \mathcal{L}^{|c_j|},\ \forall j \in \big\{1,\dots,i+1\big\} \tag{38}$$

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

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

Next, suppose $j \in \big\{1,\dots,i\big\}$. Then, by using formula (37) above, and the commutativity property of marginalization, one finds:

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

hence, the proof of formula (38). 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:

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

where $\forall c \in \mathrm{Fclos}_{\cap}\big(\mathcal{C}\big)$, 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

(36), one finds that there exists constant coefficients ρ_c, $\forall c \in \mathrm{Fclos}_\cap\big(\mathcal{C}\big)$, such that, u writes as some linear combination of the margins of f with respect to $\mathrm{Fclos}_\cap\big(\mathcal{C}\big)$ as follows:

$$u(x) = \sum_{c \in \mathrm{Fclos}_\cap\left(\mathcal{C}\right)} \frac{\rho_c\, f_c(x_c)}{L^{n-|c|}}, \forall x \in \mathcal{L}^n \tag{39}$$

Now, in order to derive an iterative expression for the coefficients ρ_c, $\forall c \in \mathrm{Fclos}_\cap\big(\mathcal{C}\big)$ in formula (39) above, then, first by construction, one has:

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

Second, suppose $c \in \mathrm{Fclos}_\cap\big(\mathcal{C}\big)/\mathrm{Front}\big(\mathcal{C}\big)$. Then, by marginalization of both sides of formula (39) with respect to all the variables with indices in Ω/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 (39) after marginalization with respect to c correspond to all the hypersite $t \in \mathrm{Fclos}_\cap\big(\mathcal{C}\big)$, such that, $c \subseteq t$, and one derives (since f is assumed to be arbitrary) that:

$$1 = \rho_c + \sum_{t \in \mathrm{Fclos}_\cap\left(\mathcal{C}\right) \text{ s.t. } c \subset t} \rho_t$$

hence:

$$\rho_c = 1 - \sum_{t \in \mathrm{Fclos}_\cap\left(\mathcal{C}\right) \text{ s.t. } c \subset t} \rho_t$$

where the symbol $\subset$ stands for the strict inclusion, which thus terminates the proof of the first part of Theorem 4.

Now, in order to show the second part of Theorem 4, let us put:

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

where we have denoted by f_c the margin of f with respect to c, $\forall c \in \mathrm{Fclos}_\cap\big(\mathcal{C}\big)$. Then, on the one hand, we have already shown above that $\forall c \in \mathrm{Fclos}_\cap\big(\mathcal{C}\big)$, 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 $\forall c \in \mathrm{Fclos}_\cap\big(\mathcal{C}\big)$ by g_c the margin of g with respect to c. Then, one has:

$$\begin{aligned} \sum_{x \in \mathcal{L}^n} \bar{f}(x)\,(\mathscr{O}_{\mathcal{C}} g)(x) &= \sum_{x \in \mathcal{L}^n} \bar{f}(x)\,\Big(\sum_{c \in \mathrm{Fclos}_\cap\left(\mathcal{C}\right)} \frac{\rho_c\, g_c(x_c)}{L^{n-|c|}}\Big) \\ &= \sum_{c \in \mathrm{Fclos}_\cap\left(\mathcal{C}\right)} \sum_{x_c \in \mathcal{L}^{|c|}} \frac{\rho_c\, \bar{f}_c(x_c)\, g_c(x_c)}{L^{n-|c|}} = 0 \end{aligned}$$

and since both f and g are arbitrary, one concludes that the operator $\mathscr{O}_{\mathcal{C}}$ is an orthogonal projection, thus, establishing the proof of the second part of Theorem 4.

A.5 Proof of Theorem 5

We will show Theorem 5 by construction, in the spirit of the proof of Theorem 4. Therefore, suppose $\big\{u_c, \forall c \in \mathcal{C}\big\}$ is a pseudo-marginals-set. Then, we will iteratively construct a function $\tilde{u} : \mathcal{L}^n \to \mathbb{R}$ which ultimately develops as:

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

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

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

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:

$$\tilde{u}^{(i+1)}(x) = \tilde{u}^{(i)}(x) + \frac{u_{c_{i+1}}(x_{c_{i+1}})}{L^{n-|c_{i+1}|}} - \frac{\tilde{u}^{(i)}_{c_{i+1}}(x_{c_{i+1}})}{L^{n-|c_{i+1}|}}, \forall x \in \mathcal{L}^n \tag{40}$$

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:

$$\tilde{u}^{(i)}_{c_j}(x_{c_j}) = u_{c_j}(x_{c_j}), \forall x_{c_j} \in \mathcal{L}^{|j|}, \forall j \in \{1, \ldots, i\}, \forall i \in \{1, \ldots, m\} \tag{41}$$

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:

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

Second, by assuming that $j \in \{1, \ldots, i\}$, then by marginalization of both sides of formula (40) with respect to all the variables with indices in $\Omega/x_{c_{i+1}}$, finally by using the assumption that the set $\{u_c, \forall c \in \mathcal{C}\}$ defines a pseudo-marginals-set, one finds:

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

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 (41).

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

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

where $\beta_c, \forall c \in \mathrm{Fclos}_\cap(\mathcal{C})$ stand for some real coefficients, such that, $\tilde{u}^{(m)}_c = u_c, \forall c \in \mathrm{Front}(\mathcal{C})$, implying that one has $\tilde{u}^{(m)}_c \equiv u_c, \forall c \in \mathcal{C}$. Next, one proceeds in exactly the same way as we have done in the proof of Theorem 4 for deriving an iterative formula for the coefficients $\beta_c, \forall c \in \mathrm{Fclos}_\cap(\mathcal{C})$, and one finds:

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

where the coefficients $\rho_c, \forall c \in \mathrm{Fclos}_\cap(\mathcal{C})$ are given by formula (17). Finally, by using Theorem 4, one derives that whatever a function $v : \mathcal{L}^n \to \mathbb{R}$, the function defined as:

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

where $\forall c \in \mathrm{Fclos}_\cap(\mathcal{C})$, v_c stands for the margin of v with respect to c, satisfies $\tilde{u}_c = u_c, \forall c \in \mathcal{C}$, thus, establishing the proof of Theorem 5.

A.6 Proof of Proposition 3

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

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

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

$$\begin{array}{rcl} \big\langle h, v \big\rangle &=& \sum_{x\in\mathcal{L}^n} h(x)\, v(x) = \sum_{c\in\mathcal{C}} \sum_{x_c \in \mathcal{L}^{|c|}} h_c\big(x_c\big)\, v_c(x_c) \\ &=& \sum_{c\in\mathcal{C}} \big\langle h_c, \mathscr{O}_\mathcal{C} v_c \big\rangle = \sum_{c\in\mathcal{C}} \big\langle \mathscr{O}_\mathcal{C} h_c, v_c \big\rangle \\ &=& \sum_{c\in\mathcal{C}} \big\langle \mathscr{O}_\mathcal{C} h_c, v \big\rangle = \big\langle \mathscr{O}_\mathcal{C}\big(\sum_{c\in\mathcal{C}} h_c\big), v \big\rangle = \big\langle \mathscr{O}_\mathcal{C} h, v \big\rangle \end{array}$$

hence $\big\langle h - \mathscr{O}_\mathcal{C} h, v \big\rangle = 0$. But, since v is assumed to be arbitrary, one finally derives that $\mathscr{O}_\mathcal{C} h \equiv h$. The converse is obvious by definition of $\mathscr{M}_\mathcal{C}$, hence the proof of Proposition 3.

A.7 Proof of Theorem 6

Proof 7 *We settle for showing equivalence between PseudoEMinMLP (22) and GlbPseudoEMinMLP (25), as one may proceed in exactly the same way for showing equivalence between PseudoEMaxMLP (23) and GlbPseudoEMaxMLP (26), on the one hand, and equivalence between PseudoΔEMinMLP (24) and GlbPseudoΔEMinMLP (27), on the other hand. Therefore, suppose $\mathcal{P} = \{p_s : \mathcal{L}^{|s|} \longrightarrow \mathbb{R}, \forall s \in \mathcal{S}\}$ is a feasible solution of PseudoΔEMinMLP (24) which, by definition, stands for a pseudo-marginal probability set. Since $\mathcal{P}$ is also a pseudo-marginals-set, by Theorem 6, one finds that there exists a function $p : \mathcal{L}^n \to \mathbb{R}$ of which set of margins with respect to $\mathcal{S}$ coincides with $\mathcal{P}$, in other words, one has:*

$$p_s(x_s) = \sum_{i \notin s} \sum_{x_i \in \mathcal{L}} p(x),\ \forall x_s \in \mathcal{L}^{|s|},\ \forall s \in \mathcal{S}$$

It follows since $p_s(x_s) \geq 0$, $\forall x_s \in \mathcal{L}^{|s|}$, $\forall s \in \mathcal{S}$ that:

$$\sum_{i \notin s} \sum_{x_i \in \mathcal{L}} p(x) \geq 0,\ \forall x_s \in \mathcal{L}^{|s|},\ \forall s \in \mathcal{S}$$

Furthermore, one has:

$$1 = \sum_{x_s \in \mathcal{L}^{|s|}} p_s(x_s) = \sum_{i \in s} \sum_{x_i \in \mathcal{L}} \big(\sum_{i \notin s} \sum_{x_i \in \mathcal{L}} p(x) \big) = \sum_{x \in \mathcal{L}^n} p(x),\ \forall x_s \in \mathcal{L}^{|s|},\ \forall s \in \mathcal{S}$$

Finally, one has:

$$\begin{array}{rcl} \sum_{s \in \mathcal{S}} \sum_{x_s \in \mathcal{L}^{|s|}} g_s(x_s)\, p_s(x_s) & = & \sum_{s \in \mathcal{S}} \sum_{x_s \in \mathcal{L}^{|s|}} g_s(x_s) \big(\sum_{i \notin s} \sum_{x_i \in \mathcal{L}} p(x) \big) \\ & = & \sum_{x \in \mathcal{L}^n} \big(\sum_{s \in \mathcal{S}} g_s(x_s) \big)\, p(x) = \sum_{x \in \mathcal{L}^n} g(x)\, p(x) \end{array}$$

But, since $\mathcal{P}$ is arbitrary, one derives that whatever a feasible solution $\mathcal{P}$ of PseudoEMinMLP (22), there exists a feasible solution p of GlbPseudoEMinMLP (25), such that, the objective value of $\mathcal{P}$ in the former is equal to the objective value of p in the latter, and one derives that GlbPseudoEMinMLP (25) is a lower bound for PseudoEMinMLP (22). Conversely, suppose p is a feasible solution of GlbPseudoEMinMLP (25), and $\forall s \in \mathcal{S}$, denote by p_s the margin of p with respect to s, and put $\mathcal{P} = \{p_s : \mathcal{L}^{|s|} \longrightarrow \mathbb{R}, \forall s \in \mathcal{S}\}$. One then easily checks that $\mathcal{P}$ defines a pseudo-marginal probability set, which is thus a feasible solution of PseudoEMinMLP (22), and the objective value of p in the former is equal to the objective value of $\mathcal{P}$ in the latter, and one derives that GlbPseudoEMinMLP (25) is a lower bound for PseudoEMinMLP (22). One finally concludes that PseudoEMinMLP (22) and GlbPseudoEMinMLP (25) are equivalent, in the sense that, they achieve the same optimal objective value, furthermore, if $\mathcal{P}^* = \{p^*_s : \mathcal{L}^{|s|} \longrightarrow \mathbb{R}, \forall s \in \mathcal{S}\}$ is an optimal solution of PseudoEMinMLP (22), then any function $p^* : \mathcal{L}^n \to \mathbb{R}$ of which set of margins with respect to $\mathcal{S}$ coincides with $\mathcal{P}^*$ is an optimal solution of GlbPseudoEMinMLP (25), conversely, if p^* is an optimal solution of GlbPseudoEMinMLP (25), then the set of margins of p^* with respect to $\mathcal{S}$ denoted by $\{p^*_s : \mathcal{L}^{|s|} \longrightarrow \mathbb{R}, \forall s \in \mathcal{S}\}$ is an optimal solution of PseudoEMinMLP (22), thus, establishing the proof of Theorem 6.

A.8 Proof of Theorem 7

Proof 8 *We first have the following lemma which is shown in Appendix subsection A.8.1.*

Lemma 1 *Suppose $\big\{q_s : \mathcal{L}^{|s|} \longrightarrow \mathbb{R}, \forall s \in \mathit{Front}(\mathcal{S})\big\}$ is a pseudo-marginals-set, such that:*

$$\textstyle\sum_{x_s \in \mathcal{L}^{|s|}} q_s(x_s) = 0,\ \forall s \in \mathit{Front}(\mathcal{S})$$

said otherwise, $\big\{q_s : \mathcal{L}^{|s|} \longrightarrow \mathbb{R}, \forall s \in \mathit{Front}(\mathcal{S})\big\}$ is the set of margins with respect to $\mathit{Front}(\mathcal{S})$ of a function $q : \mathcal{L}^n \to \mathbb{R}$, verifying $\sum_{x \in \mathcal{L}^n} q(x) = 0$. Then, the minimization problem:

$$\begin{array}{c} \min_Q \Big\{ \sum_{x \in \mathcal{L}^n} |Q(x)| \Big\} \\ \sum_{i \notin s} \sum_{x_i \in \mathcal{L}} Q(x) = q_s(x_s),\ \forall x_s \in \mathcal{L}^{|s|},\ \forall s \in \mathit{Front}(\mathcal{S}) \end{array} \tag{42}$$

and its following LP relaxation:

$$\begin{array}{l} \qquad\qquad \min \Big\{ 2\tau \Big\} \\ \begin{cases} \forall s \in \mathit{Front}(\mathcal{S}),\ \forall x_s \in \mathcal{L}^{|s|}, \begin{cases} p_s^+(x_s) - p_s^-(x_s) = q_s(x_s) \\ p_s^+(x_s) \geq 0 \\ p_s^-(x_s) \geq 0 \end{cases} \\ \sum_{x_s \in \mathcal{L}^{|s|}} p_s(x_s) = \tau,\ \forall s \in \mathit{Front}(\mathcal{S}) \\ \big\{p_s^+ : \mathcal{L}^{|s|} \to \mathbb{R}, \forall s \in \mathcal{S}\big\} \text{ and } \big\{p_s^- : \mathcal{L}^{|s|} \to \mathbb{R}, \forall s \in \mathcal{S}\big\} \text{ are pseudo-marginals-sets} \end{cases} \end{array} \tag{43}$$

achieve the same optimal objective value.

Next, let $\mathcal{P}^+ = \big\{p_s^+ : \mathcal{L}^{|s|} \longrightarrow \mathbb{R}, \forall s \in \mathit{Front}(\mathcal{S})\big\}$ and $\mathcal{P}^- = \big\{p_s^- : \mathcal{L}^{|s|} \longrightarrow \mathbb{R}, \forall s \in \mathit{Front}(\mathcal{S})\big\}$ be a feasible solution of PseudoEMinMLP (22) and a feasible solution of PseudoEMaxMLP (23), respectively, which thus stand for two pseudo-marginal probability sets verifying $1 = \sum_{x_s \in \mathcal{L}^{|s|}} p_s^+(x_s) = \sum_{x_s \in \mathcal{L}^{|s|}} p_s^-(x_s)$, $\forall s \in \mathit{Front}(\mathcal{S})$, and put:

$$q_s(x_s) = p_s^+(x_s) - p_s^-(x_s),\ \forall s \in \mathit{Front}(\mathcal{S}),\ \forall x_s \in \mathcal{L}^{|s|}$$

Then, one checks that:

$$\textstyle\sum_{x_s \in \mathcal{L}^{|s|}} q_s(x_s) = \sum_{x_s \in \mathcal{L}^{|s|}} p_s^+(x_s) - \sum_{x_s \in \mathcal{L}^{|s|}} p_s^-(x_s) = 0,\ \forall s \in \mathit{Front}(\mathcal{S})$$

Now, assume minimization problem (42). It follows, since, at least, $\big(\mathcal{P}^+, \mathcal{P}^-, \tau = 1\big)$ is a feasible solution of problem (43), that minimization problem (42) achieves an optimal objective value which is less than, or equal to 2. But, since both $\mathcal{P}^+$ and $\mathcal{P}^-$ are arbitrary, one concludes that max-min problem (29) achieves an optimal objective value which is, at most, equal to 2, this is on the one hand. On the other hand, since all the joint distributions of RMLV $\mathcal{X}$ are feasible both for GlbPseudoEMinMLP (25) and GlbPseudoEMaxMLP (26), thus such an optimal value of 2 must be achieved by max-min problem (30) anyway, as otherwise, one would be able to construct a joint delta-distribution of RMLV $\mathcal{X}$ as: $q = p_0^+ - p_0^-$, where p_0^+ and p_0^- stand for two respective feasible solutions of GlbPseudoEMinMLP (25) and GlbPseudoEMaxMLP (26), such that $\sum_{x \in \mathcal{L}^n} |q(x)| < 2$ yet achieving an objective value in DEMinMLP (13) which is less than, or equal to the optimal objective value of the latter and which, by Theorem 3, is impossible. One concludes that the optimal objective value of max-min problem (29) is 2, hence the proof of Theorem 7.

A.8.1 Proof of Lemma 1

Proof 9 *In order to show Lemma 1, first suppose Q is a feasible solution of problem (42). Since one has $\sum_x Q(x) = \sum_x q(x) = 0$, by Proposition 2, one finds that there exists two nonnegative functions $p^+ : \mathcal{L}^n \to \mathbb{R}^+$ and $p^- : \mathcal{L}^n \to \mathbb{R}^+$, such that:*

$$\sum_{x \in \mathcal{L}^n} |Q(x)| = 2 \sum_{x \in \mathcal{L}^n} p^+(x) = 2 \sum_{x \in \mathcal{L}^n} p^-(x)$$

and one concludes that problem (42) is equivalent to the LP:

$$\begin{aligned} &\min_{p^+,p^-} \left\{ 2 \sum_{x \in \mathcal{L}^n} p^+(x) \right\} \\ &\begin{cases} \sum_{i \notin s} \sum_{x_i \in \mathcal{L}} p^+(x) - \sum_{i \notin s} \sum_{x_i \in \mathcal{L}} p^-(x) = q_s(x_s), \, \forall x_s \in \mathcal{L}^{|s|}, \, \forall s \in \mathit{Front}(\mathcal{S}) \\ p^+(x) \geq 0, \, p^-(x) \geq 0, \, \forall x \in \mathcal{L}^n \end{cases} \end{aligned} \tag{44}$$

Next, consider the following LP relaxation of problem (44):

$$\begin{aligned} &\min_{p^+,p^-} \left\{ 2 \sum_{x \in \mathcal{L}^n} p^+(x) \right\} \\ &\begin{cases} \forall s \in \mathit{Front}(\mathcal{S}), \, \forall x_s \in \mathcal{L}^{|s|}, \begin{cases} \sum_{i \notin s} \sum_{x_i \in \mathcal{L}} p^+(x) - \sum_{i \notin s} \sum_{x_i \in \mathcal{L}} p^-(x) = q_s(x_s) \\ \sum_{i \notin s} \sum_{x_i \in \mathcal{L}} p^+(x) \geq 0 \\ \sum_{i \notin s} \sum_{x_i \in \mathcal{L}} p^-(x) \geq 0 \end{cases} \\ p^+(x) \in \mathbb{R}, \, p^-(x) \in \mathbb{R}, \, \forall x \in \mathcal{L}^n \end{cases} \end{aligned} \tag{45}$$

and let us show that problem (44), hence problem (42), may be completely solved by LP (45). First of all, one easily checks that problem (45) is bounded, since whatever its feasible solution (p^+, p^-), one has $\forall s \in \mathit{Front}(\mathcal{S})$:

$$\sum_{x \in \mathcal{L}^n} p^+(x) = \sum_{i \in s} \sum_{x_i \in \mathcal{L}} \left(\sum_{i \notin s} \sum_{x_i \in \mathcal{L}} p^+(x) \right) \geq 0$$

For the sake of simplicity in the remainder of the proof, thus whatever a function $u : \mathcal{L}^n \longrightarrow \mathbb{R}$, and $\forall s \in \mathit{Front}(\mathcal{S})$, we simply denote by u_s the margin of u with respect to s. Now, we need the following Lemma which is shown in Appendix subsection A.8.2.

Lemma 2 *Problem (45) and the following two (Lagrangian-like) problems:*

$$\begin{aligned} &\min_{p^+,p^-} \left\{ 2 \sum_{x \in \mathcal{L}^n} p^+(x) - 2 \inf_{s \in \mathit{Front}(\mathcal{S})} \left\{ \inf_{x_s \in \mathcal{L}^{|s|}} \left\{ \inf \left\{ L^{|s|} \, p_s^+(x_s), L^{|s|} \, p_s^-(x_s) \right\} \right\} \right\} \\ &\begin{cases} \forall s \in \mathit{Front}(\mathcal{S}), \, \forall x_s \in \mathcal{L}^{|s|}, \begin{cases} p_s^+(x_s) - p_s^-(x_s) = q_s(x_s) \\ p_s^+(x_s) \geq 0 \\ p_s^-(x_s) \geq 0 \end{cases} \\ p^+(x) \in \mathbb{R}, \, p^-(x) \in \mathbb{R}, \, \forall x \in \mathcal{L}^n \end{cases} \end{aligned} \tag{46}$$

and

$$\begin{aligned} &\min_{p^+,p^-} \left\{ 2 \sum_{x \in \mathcal{L}^n} p^+(x) - 2 \inf_{s \in \mathit{Front}(\mathcal{S})} \left\{ \inf_{x_s \in \mathcal{L}^{|s|}} \left\{ \inf \left\{ L^{|s|} \, p_s^+(x_s), L^{|s|} \, p_s^-(x_s) \right\} \right\} \right\} \\ &\begin{cases} p_s^+(x_s) - p_s^-(x_s) = q_s(x_s), \, \forall x_s \in \mathcal{L}^{|s|}, \, \forall s \in \mathit{Front}(\mathcal{S}) \\ p^+(x) \geq 0, \, p^-(x) \geq 0, \, \forall x \in \mathcal{L}^n \end{cases} \end{aligned} \tag{47}$$

achieve the same optimal objective value.

Now, let us resume with the rest of the proof of Lemma 1. First, observe that problems (45) and (46) are defined over the same feasible set, and that the objective function of the former is greater than the objective function of the latter, yet by Lemma 2, they achieve the same optimal objective value. One derives that problems (45) and (46) are equivalent, in the sense that, they achieve the same optimal objective value, plus their respective optimal solutions coincide. As an aside, one also establishes that any optimal solution (p^{+*}, p^{-*}) *of problem (45), hence, of problem (46) has to satisfy that:*

$$-2\ \inf_{s \in Front(\mathcal{S})} \Big\{ \ \inf_{x_s \in \mathcal{L}^{|s|}} \big\{ \ \inf \big\{ L^{|s|}\, p_s^{+*}(x_s), L^{|s|}\, p_s^{-*}(x_s) \big\} \Big\} = 0$$

This is on the one hand. On the other hand, since by Lemma 2, problems (46) and (47) achieve the same optimal objective value, their objective functions are equal, and the feasible set of the latter (which happens to be the same feasible set of the original problem (44)) is included in the feasible set of problem (45), one derives that problem (45) is minimized over the feasible set of problem (44), which also means that the former completely solves the latter. Finally, by observing that problem (45) happens to be nothing else than a global reformulation of problem (43), one finally establishes the proof of Lemma 1.

A.8.2 Proof of Lemma 2

Proof 10 *First of all, one may rewrite problem (45) as:*

$$\min_{p^+, p^-, M} \Big\{ \ 2 \sum_{x \in \mathcal{L}^n} p^+(x) \Big\}$$
$$\begin{cases} \forall s \in Front(\mathcal{S}),\ \forall x_s \in \mathcal{L}^{|s|}, \begin{cases} p_s^+(x_s) - p_s^-(x_s) = q_s(x_s) \\ p_s^+(x_s) \geq 0 \\ p_s^-(x_s) \geq 0 \end{cases} \\ p^+(x) \geq -M,\ p^-(x) \geq -M,\ \forall x \in \mathcal{L}^n \\ M \geq 0 \end{cases} \tag{48}$$

as one exactly recovers problem (45) for a big enough value of M *in problem (48). Next, by doing change of variables* $p^+ \leftarrow p^+ + M$*,* $p^- \leftarrow p^- + M$ *in problem (48), one finds that problem (45) is equivalent to the problem:*

$$\min_{p^+, p^-, M} \Big\{ \ 2 \sum_{x \in \mathcal{L}^n} p^+(x) - 2M\, L^n \Big\}$$
$$\begin{cases} \forall s \in Front(\mathcal{S}),\ \forall x_s \in \mathcal{L}^{|s|}, \begin{cases} p_s^+(x_s) - p_s^-(x_s) = q_s(x_s) \\ p_s^+(x_s) \geq M\, L^{n-|s|} \\ p_s^-(x_s) \geq M\, L^{n-|s|} \end{cases} \\ p^+(x) \geq 0,\ p^-(x) \geq 0,\ \forall x \in \mathcal{L}^n \\ M \geq 0 \end{cases} \tag{49}$$

Since the following system of linear inequalities:

$$\begin{cases} \forall x_s \in \mathcal{L}^{|s|},\ \forall s \in Front(\mathcal{S}), \begin{cases} p_s^+(x_s) \geq M\, L^{n-|s|} \\ p_s^-(x_s) \geq M\, L^{n-|s|} \end{cases} \\ M \geq 0 \end{cases}$$

is equivalent to the following one:

$$0 \leq M \leq \frac{1}{L^n} \inf_{s \in \mathit{Front}(\mathcal{S})} \Big\{ \inf_{x_s \in \mathcal{L}^{|s|}} \big\{ \inf \big\{ L^{|s|} \, p_s^+(x_s), L^{|s|} \, p_s^-(x_s) \big\} \big\} \Big\}$$

and by observing that the right hand side of the latter formula is always nonnegative, and that the objective function of problem (49) achieves its smaller values for bigger values of M, thus by substituting M in problem (49) with its optimal value in terms of the variables p_s^+, and p_s^- which is:

$$\frac{1}{L^n} \inf_{s \in \mathit{Front}(\mathcal{S})} \Big\{ \inf_{x_s \in \mathcal{L}^{|s|}} \big\{ \inf \big\{ L^{|s|} \, p_s^+(x_s), L^{|s|} \, p_s^-(x_s) \big\} \big\} \Big\}$$

one exactly retrieves problem (47), and one concludes that problem (45) achieves the same optimal objective value as problem (47). Second, problem (46) may also write as:

$$\min_{p^+, p^-} \Big\{ 2 \sum_{x \in \mathcal{L}^n} p^+(x) - 2 \inf_{s \in \mathit{Front}(\mathcal{S})} \Big\{ \inf_{x_s \in \mathcal{L}^{|s|}} \big\{ \inf \big\{ L^{|s|} \, p_s^+(x_s), L^{|s|} \, p_s^-(x_s) \big\} \big\} \Big\}$$
$$\begin{cases} \forall s \in \mathit{Front}(\mathcal{S}),\, \forall x_s \in \mathcal{L}^{|s|}, \begin{cases} p_s^+(x_s) - p_s^-(x_s) = q_s(x_s) \\ p_s^+(x_s) \geq 0 \\ p_s^-(x_s) \geq 0 \end{cases} \\ p^+(x) \geq -M,\, p^-(x) \geq -M,\, \forall x \in \mathcal{L}^n \\ M \geq 0 \end{cases} \tag{50}$$

in such a way that, by doing the changes of variables $p^+ \leftarrow p^+ + M$, $p^- \leftarrow p^- + M$, one may rewrite problem (50) as:

$$\min_{p^+, p^-} \Big\{ 2 \sum_{x \in \mathcal{L}^n} p^+(x) - 2 \inf_{s \in \mathit{Front}(\mathcal{S})} \Big\{ \inf_{x_s \in \mathcal{L}^{|s|}} \big\{ \inf \big\{ L^{|s|} \, p_s^+(x_s), L^{|s|} \, p_s^-(x_s) \big\} \big\} \Big\}$$
$$\begin{cases} \forall s \in \mathit{Front}(\mathcal{S}),\, \forall x_s \in \mathcal{L}^{|s|}, \begin{cases} p_s^+(x_s) - p_s^-(x_s) = q_s(x_s) \\ p_s^+(x_s) \geq M \, L^{n-|s|} \\ p_s^-(x_s) \geq M \, L^{n-|s|} \end{cases} \\ p^+(x) \geq 0,\, p^-(x) \geq 0,\, \forall x \in \mathcal{L}^n \\ M \geq 0 \end{cases} \tag{51}$$

One then observes that the variable M does not appear in the objective function of problem (51), and the latter could only increase with larger values of M (as larger values of M constrain more the possible values of (p^+, p^-)), one finds that the optimal value of M in problem (51) is 0, in such a way that, after substituting M with its optimal value which is 0 in problem (51), and by noticing that the resulting constraints of the form: $\forall s \in \mathit{Front}(\mathcal{S})$, $\forall x_s \in \mathcal{L}^{|s|}$, $\begin{cases} p_s^+(x_s) \geq 0 \, L^{n-|s|} = 0 \\ p_s^-(x_s) \geq 0 \, L^{n-|s|} = 0 \end{cases}$ become redundant, one retrieves problem (47). One concludes that problem (47) and problem (46) achieve the same optimal objective value, thus establishing the proof of Lemma 2.

A.9 Proof of Theorem 8

Proof 11 *First, we have the following lemma which is shown in Appendix subsection A.9.1.*

Lemma 3 *The optimal objective value of minimization problem (42) is equal to:*

$$\sup_{s\in Front(\mathcal{S})}\Big\{\sum_{x_s\in\mathcal{L}^{|s|}}|q_s(x_s)|\Big\}$$

But, since any feasible solution q of GlbPseudoΔEMinMLP (27) verifies:

$$\sup_{s\in Front(\mathcal{S})}\Big\{\sum_{x_s\in\mathcal{L}^{|s|}}|q_s(x_s)|\Big\}\leq 2$$

where $\forall s\in Front(\mathcal{S})$, q_s stands for the margin of q with respect to s, by applying the result of Lemma 3, one first derives that the optimal objective value of max-min problem (30) is less than, or equal to 2, but since the latter value is, at least, achieved by any feasible solution of DEMinMLP (13) writing in the form $\mathbf{1}_x(\cdot)-\mathbf{1}_y(\cdot)$, and verifying that, $\exists s\in Front(\mathcal{S})$, such that, $x_s\neq y_s$, one finally concludes that the optimal objective value of max-min problem (30) is equal to 2, hence the proof of Theorem 8.

A.9.1 Proof of Lemma 3

Assume LP (45) which, by Lemma 1, achieves the same optimal objective value as minimization problem (42). Next, by observing that one may fully express the variables $p_s^-(x_s)$ in terms of the variables $p_s^+(x_s)$, as:

$$p_s^-(x_s)=p_s^+(x_s)-q_s(x_s),\ \forall x_s\in\mathcal{L}^{|s|},\ \forall s\in\mathrm{Front}(\mathcal{S})$$

which then define nonnegative pseudo-marginals as far as the following condition:

$$p_s^+(x_s)\geq\sup\{0,q_s(x_s)\},\ \forall x_s\in\mathcal{L}^{|s|},\ \forall s\in\mathrm{Front}(\mathcal{S})$$

is satisfied, one may rewrite problem (45) as:

$$\begin{array}{l}\min_{p^+}\Big\{2\sum_{x\in\mathcal{L}^n}p^+(x)\Big\}\\ \begin{cases}p_s^+(x_s)\geq\sup\{0,q_s(x_s)\},\ \forall x_s\in\mathcal{L}^{|s|},\ \forall s\in\mathrm{Front}(\mathcal{S})\\ p^+(x)\in\mathbb{R},\ \forall x\in\mathcal{L}^n\end{cases}\end{array}\tag{52}$$

where it has been assumed that, $\forall s\in\mathrm{Front}(\mathcal{S})$, p_s stands for the margin of p with respect to s. One finds that the dual of LP (52) is given by:

$$\begin{array}{l}\max\Big\{E=2\sum_{s\in\mathrm{Front}(\mathcal{S})}\sum_{x_s\in\mathcal{L}^{|s|}}\mu_s(x_s)\sup\{0,q_s(x_s)\}\Big\}\\ \begin{cases}\sum_{s\in\mathrm{Front}(\mathcal{S})}\mu_s(x_s)=1,\forall x\in\mathcal{L}^n\\ \mu_s(x_s)\geq 0,\ \forall x_s\in\mathcal{L}^{|s|},\ \forall s\in\mathrm{Front}(\mathcal{S})\end{cases}\end{array}\tag{53}$$

Let E^* be the optimal value of dual LP (53) which is also the optimal objective value of problem (45), and denote:

$$s_0=Argsup_{s\in\mathrm{Front}(\mathcal{S})}\{\ \sum_{x_s\in\mathcal{L}^{|s|}}\sup\{0,q_s(x_s)\}\ \}$$

then define the solution $\mu^{(0)} = \big(\mu_s^{(0)}(x_s)\big)_{x_s \in \mathcal{L}^{|s|}, s \in \mathrm{Front}(\mathcal{S})}$ as:

$$\forall s \in \mathrm{Front}(\mathcal{S}),\, \forall x_s \in \mathcal{L}^{|s|},\, \mu_s^{(1)}(x_s) = \begin{cases} 1, & \text{if } s = s_0, \\ 0, & \text{else.} \end{cases}$$

Then, one checks that $\mu^{(0)}$ is a feasible solution of dual LP (53), and achieves an objective value which is equal to $\sup_{s \in \mathrm{Front}(\mathcal{S})} \big\{ \sum_{x_s \in \mathcal{L}^{|s|}} |q_s(x_s)| \big\}$. One derives immediately that:

$$E^* \geq \sup_{s \in \mathrm{Front}(\mathcal{S})} \big\{ \textstyle\sum_{x_s \in \mathcal{L}^{|s|}} |q_s(x_s)| \big\} \tag{54}$$

This is on the one hand. On the other hand, surprisingly, we show hereafter that dual LP (53) may also be seen as an expectation maximization reformulation of a multi-label problem with $\mathcal{L}^n$ as domain, and $\mathrm{Front}(\mathcal{S})$ as label-set. In fact, let us introduce the random couple $\big(S, X\big)$, where S stands for a random variable taking values in $\mathrm{Front}(\mathcal{S})$, and $X = (X_1, \ldots, X_m)$ is a m-dimensional vector of integer random variables taking values in $\mathcal{L}$, and assumed to be independent and identically distributed, in such a way that, $\forall x \in \mathcal{L}^n$, one has $\mathbb{P}\big(X = x\big) = \Pi_{i=1}^m \mathbb{P}(X_i = x_i)$, and let us put $\mathbb{P}(X_i = x_i) = \frac{1}{L}, \forall i = 1, \ldots,$ $\forall x_i = 0, \ldots, L-1$, and $\forall i = 1, \ldots, m$. Furthermore, introduce two real-value functions f and $F : \big(\mathrm{Front}(\mathcal{S}), \mathcal{L}^n\big) \to \mathbb{R}$ defined as: $f(s;x) = \frac{\sup\{0, q_s(x_s)\}}{L^{n-|s|}}, \forall x \in \mathcal{L}^n$, and $F(s;x) = 2\, L^n\, f(s;x), \forall x \in \mathcal{L}^n$.

Lemma 4 *Dual LP (53) is equivalent to the expectation maximization problem:*

$$\begin{aligned} &\max_{\mathbb{P}} \Big\{ \mathbb{E}_{\mathbb{P}}\Big[F(S;X)\Big] = 2 \textstyle\sum_{s \in Front(\mathcal{S})} \sum_{x \mathcal{L}^n} f(s;x)\, \mathbb{P}(S = s/X = x) \Big\} \\ &\begin{cases} \mathbb{P}(S = s/X = x) = \mathbb{P}(S = s/X_s = x_s),\, \forall (s,x) \in Front(\mathcal{S}) \times \mathcal{L}^n \\ \mathbb{P}(X = x) = \frac{1}{L^n},\, \forall x \in \mathcal{L}^n \end{cases} \end{aligned} \tag{55}$$

The proof of Lemma 4 is sketched in Appendix subsection A.9.2 (please refer also to Remark 4 for further explanations concerning the modeling of $\mathbb{P}(X = x)$ as a uniform distribution). Now, denote $p(s,x) = \mathbb{P}(S = s, X = x)$, $\forall (s,x) \in \mathrm{Front}(\mathcal{S}) \times \mathcal{L}^n$ standing for the joint distribution of the couple $\big(S, X\big)$, and $p(s/x) = \mathbb{P}(S = s/X = x)$ standing for its conditional distribution. Then, by using the fact that one has $s \in \mathrm{Front}(\mathcal{S})$:

$$\mathbb{P}(S = s, X_s = x_s) = \textstyle\sum_{i \notin s} \sum_{x_i \in \mathcal{L}} \mathbb{P}(S = s, X = x) = \sum_{i \notin s} \sum_{x_i \in \mathcal{L}} \mathbb{P}(S = s/X = x)\mathbb{P}(X = x)$$

hence

$$\mathbb{P}(S = s/X_s = x_s) = \textstyle\sum_{i \notin s} \sum_{x_i \in \mathcal{L}} \mathbb{P}(S = s/X = x)\, p_{\bar{s}}^0$$

it has been assumed that $p_{\bar{s}}^0 = \frac{1}{L^{n-|s|}}$ which is a constant, because $\mathbb{P}(X = x)$ is assumed to be uniform, in such a way that, one may reexpress problem (55) as a LP as follows (while simply bearing in mind that one has $\mathbb{P}(X = x) = \frac{1}{L^n}$, $\forall x \in \mathcal{L}^n$):

$$\begin{aligned} &\max_p \Big\{ \mathbb{E}_p\Big[F(S;X)\Big] = 2 \textstyle\sum_{s \in \mathrm{Front}(\mathcal{S})} \sum_{x \mathcal{L}^n} f(s;x)\, p(s/x) \Big\} \\ &\begin{cases} p(s/x) - \sum_{i \notin s} \sum_{x_i \in \mathcal{L}} p(s/x)\, p_{\bar{s}}^0 = 0,\, \forall s \in \mathrm{Front}(\mathcal{S}),\, \forall x \in \mathcal{L}^n \\ \sum_{s \in \mathrm{Front}(\mathcal{S})} p(s/x) = 1,\, \forall x \in \mathcal{L}^n \\ p(s/x) \geq 0,\, \forall s \in \mathrm{Front}(\mathcal{S}), \forall x_s \in \mathcal{L}^{|s|} \end{cases} \end{aligned} \tag{56}$$

The dual of LP (56) writes as:

$$\begin{array}{l} \min_{p^+,p^-} \left\{ 2\sum_{x\mathcal{L}^n} p^+(x) \right\} \\ \begin{cases} p^+(x) \geq p^-(x) + \Big(\sup\{0, q_s(x_s)\} - p^-_s(x_s) \Big)\, p^0_{\bar{s}},\ \forall s \in \mathrm{Front}(\mathcal{S}),\ \forall x \in \mathcal{L}^n \\ p^+(x) \in \mathbb{R},\ p^-(x) \in \mathbb{R},\ \forall x \in \mathcal{L}^n \end{cases} \end{array} \tag{57}$$

where $\forall s \in \mathrm{Front}(\mathcal{S})$, p^-_s stands for the margin of p^- with respect to s. Denote by (p^{+*}, p^{-*}) the optimal solution of LP (57), thus verifying:

$$p^{+*}(x) = \sup_{s\in\mathrm{Front}(\mathcal{S})} \Big\{ p^{-*}(x) + \Big(\sup\{0, q_s(x_s)\} - p^{-*}_s(x_s) \Big)\, p^0_{\bar{s}} \Big\},\ \forall x \in \mathcal{L}^n$$

One finds that there exists $s_1 \in \mathrm{Front}(\mathcal{S})$, such that:

$$\begin{cases} p^{+*}(x) = p^{-*}(x) + \Big(\sup\{0, q_{s_1}(x_{s_1})\} - p^-_{s_1}(x_{s_1}) \Big)\, p^0_{\bar{s}_1},\ \forall x \in \mathcal{L}^n \\ p^{+*}_{s_1}(x_{s_1}) = \sup\{0, q_s(x_{s_1})\},\ \forall x_{s_1} \in \mathcal{L}^{|s_1|} \\ 2\sum_{x\in\mathcal{L}^n} p^{+*}(x) = 2\sum_{x_{s_1}\in\mathcal{L}^{|s_1|}} \sup\{0, q_s(x_{s_1})\} = \sum_{x_{s_1}\in\mathcal{L}^{|s_1|}} \big| q_{s_1}(x_{s_1}) \big| \end{cases}$$

One finally derives that:

$$E^* \leq \sup_{s\in\mathrm{Front}(\mathcal{S})} \left\{ \sum_{x_s\in\mathcal{L}^{|s|}} |q_s(x_s)| \right\} \tag{58}$$

Combining formulas (54) and (58), one concludes that:

$$E^* = \sup_{s\in\mathrm{Front}(\mathcal{S})} \left\{ \sum_{x_s\in\mathcal{L}^{|s|}} |q_s(x_s)| \right\}$$

hence the proof of Lemma 3.

A.9.2 Proof of Lemma 4

First, suppose $\mathbb{P}(S = s/X = x)$ is a conditional probability of the couple (S, X) verifying $\mathbb{P}(S = s/X = x) = \mathbb{P}(S = s/X_s = x_s), \forall s \in \mathrm{Front}(\mathcal{S}), \forall x \in \mathcal{L}^n$, and put $\mu_s(x_s) = \mathbb{P}(S = s/X_s = x_s)$, $\forall s \in \mathrm{Front}(\mathcal{S}), \forall x_s \in \mathcal{L}^{|s|}$. Then, one checks that:

$$\begin{cases} \mu_s(x_s) \geq 0,\ \forall s \in \mathrm{Front}(\mathcal{S}), \forall x_s \in \mathcal{L}^{|s|} \\ \sum_{s\in\mathrm{Front}(\mathcal{S})} \mu_s(x_s) = \sum_{s\in\mathrm{Front}(\mathcal{S})} \mathbb{P}(S = s/X = x) = 1,\ \forall x \in \mathcal{L}^n \\ 2\sum_{s\in\mathrm{Front}(\mathcal{S})} \sum_{x_s\in\mathcal{L}^{|s|}} \mu_s(x_s)\ \sup\{0, q_s(x_s)\} = \mathbb{E}_{\mathbb{P}}\Big[F(S; X) \Big] \end{cases}$$

Conversely, suppose $\big(\mu_s(x_s)\big)_{x_s\in\mathcal{L}^{|s|}, s\in\mathrm{Front}(\mathcal{S})}$ is a feasible solution of dual LP (53), and put:

$$\begin{cases} \mathbb{P}(X = x) = \frac{1}{L^n},\ \forall x \in \mathcal{L}^n \\ \mathbb{P}(S = s, X = x) = \mathbb{P}(X = x)\, \mu_s(x_s),\ \forall s \in \mathrm{Front}(\mathcal{S}),\ \forall x \in \mathcal{L}^n \end{cases}$$

Then, one checks that:

$$\begin{cases} \mathbb{P}(S = s/X = x) \geq 0,\ \forall s \in \mathrm{Front}(\mathcal{S}),\ \forall x \in \mathcal{L}^n \\ \sum_{s\in\mathrm{Front}(\mathcal{S})} \sum_{x\in\mathcal{L}^n} \mathbb{P}(S = s, X = x) = \sum_{x\in\mathcal{L}^n} \big(\sum_{s\in\mathrm{Front}(\mathcal{S})} \mu_s(x_s) \big) = \sum_{x\in\mathcal{L}^n} P(x) = 1 \\ \mathbb{P}(S = s/X = x) = \mathbb{P}(S = s/X_s = x_s),\ \forall s \in \mathrm{Front}(\mathcal{S}),\ \forall x \in \mathcal{L}^n \\ \mathbb{E}_{\mathbb{P}}\Big[F(S; X) \Big] = 2\sum_{s\in\mathrm{Front}(\mathcal{S})} \sum_{x_s\in\mathcal{L}^{|s|}} \mu_s(x_s)\ \sup\{0, q_s(x_s)\} \end{cases}$$

One concludes that dual LP (53) and expectation maximization problem (55) are equivalent.

Remark 4 *Please, note that we have chosen $\mathbb{P}(X = x)$ as the uniform distribution as $\mathbb{P}(X = x) = \frac{1}{L^n}, \forall x \in \mathcal{L}^n$ just for the sake of obtention of a nice formula of the objective function of problem (55) expressing as the expectation of $F(S; X)$. But in fact, such an equivalence relationship between problems (53) and (55) is guaranteed independently of the choice of $\mathbb{P}(X = x)$, as far as one has $\mathbb{P}(X = x) > 0, \forall x \in \mathcal{L}^n$. This is, because, problem (55) solely depends upon the conditional probabilities $\mathbb{P}(S/X)$, such that, $\mathbb{P}(S = s/X = x) = \mathbb{P}(S = s/X = x_s)$.*

A.10 Proof of Theorem 9

First of all, denote by ΔLP the LP standing for the difference of GlbPseudoEMinMLP (25) and GlbPseudoEMaxMLP (26), in that order. Obviously, ΔLP is a lower bound for DEMinMLP (13), this is on the one hand. On the other hand, let p and p' be two respective feasible solutions of GlbPseudoEMinMLP (25) and GlbPseudoEMaxMLP (26), and let us put $q = p - p'$. Since Theorem 7 guarantees that the margins of q with respect to Front($\mathcal{S}$) are originated from a joint delta-distribution q' of RMLV $\mathscr{X}$, for instance:

$$q' = \mathscr{O}_{\mathcal{S}} q + Argmin_{\bar{q} \in \bar{\mathscr{M}}_{\mathcal{S}}} \big\{ \textstyle\sum_{x \in \mathcal{L}^n} |q(x) + \bar{q}(x)| \big\}$$

thus verifying $\sum_{x \in \mathcal{L}^n} |q'(x)| \leq 2$, Theorem 2 then guarantees that:

$$\big\langle g, p - p' \big\rangle = \big\langle g, q \big\rangle = \big\langle g, q' \big\rangle \geq \inf_{x \in \mathcal{L}^n} \big\{ g(x) \big\} - \sup_{x \in \mathcal{L}^n} \big\{ g(x) \big\}$$

which means that ΔLP is an upper bound for DEMinMLP (13). Thus, one derives that ΔLP exactly solves DEMinMLP (13), implying that GlbPseudoEMinMLP (25), hence, PseudoEMinMLP (22), must solve EMinMLP (8), and GlbPseudoEMaxMLP (26), hence, PseudoEMaxMLP (23) must solve EMaxMLP (9). Finally, by Theorem 5, the proof of the remaining claim of Theorem 9 concerning coincidence, on the one hand, of the optimal pseudo-marginals of PseudoEMinMLP (22) (resp. the optimal pseudo-marginals of PseudoEMaxMLP (23)), on the other hand, of the ordinary margins with respect to Front($\mathcal{S}$) of an optimal solution of EMinMLP (8) (resp. an optimal solution of EMaxMLP (9)) follows immediately, hence the proof of Theorem 9.

A.11 Proof of Theorem 10

The proof of Theorem 10 is quasi-identical to the proof of Theorem 9, first, by taking q directly as a feasible solution of GlbPseudoΔEMinMLP (27), next, by using Theorem 8 instead of Theorem 7 for establishing that GlbPseudoEMinMLP (25), hence, PseudoΔEMinMLP (24) exactly solves DEMinMLP (13), finally, by using Theorem 5 for establishing coincidence of the optimal pseudo-marginals of PseudoΔEMinMLP (24) on the one hand, and the marginals of an optimal solution of DEMinMLP (13), on the other hand.

A.12 Proof of Theorem 11

Proof 12 *Denote throughout by $g^{(inf)} = \inf_{x \in \mathcal{L}^n} \big\{ g(x) \big\}$, and $g^{(sup)} = \sup_{x \in \mathcal{L}^n} \big\{ g(x) \big\}$. We need Lemma 5 below which is shown in Appendix subsection A.12.1.*

Lemma 5 *Assume GlbPseudoΔEMinMLP (27). Then, under Assumption 1, there exists real constants μ and $\lambda_s > 0, s \in Front(\mathcal{S})$, such that, any optimal solution of GlbPseudoΔEMinMLP (27) is also an optimal solution of the following Lagrangian functional:*

$$\mathscr{L}\big(q;\mu,\lambda\big) = \sum_{x\in\mathcal{L}^n} g(x)\,q(x) + \mu\sum_{x\in\mathcal{L}^n} q(x) + \sum_{s\in Front(\mathcal{S})}\lambda_s\,\big(\sum_{x_s\in\mathcal{L}^{|s|}}|q_s(x_s)| - 2\big) \tag{59}$$

defined $\forall q : \mathcal{L}^n \to \mathbb{R}$.

Now, assume q^* is an optimal solution of GlbPseudoΔEMinMLP (27), and denote $\forall s \in Front(\mathcal{S})$ by q^*_s the margin of q^* with respect to s. It follows, by Lagrangian duality, that:

$$\lambda_s\,\big(\sum_{x_s\in\mathcal{L}^{|s|}}|q^*_s(x_s)| - 2\big) = 0,\ \forall s \in Front(\mathcal{S})$$

and we have already shown that one has $\lambda_s > 0, \forall s \in Front(\mathcal{S})$, one derives:

$$\sum_{x_s\in\mathcal{L}^{|s|}}|q^*_s(x_s)| = 2,\ \forall s \in Front(\mathcal{S})$$

hence the proof of Theorem 11.

A.12.1 Proof of Lemma 5

Proof 13 *We will make use of Lemma 6 below which is shown in Appendix subsection A.12.2.*

Lemma 6 *Suppose a hypersite $s_0 \in Front(\mathcal{S})$, and define the hypersite:*

$$\tilde{s}_0 = \big\{\, i \in s_0, s.t., \forall s \in Front(\mathcal{S})/\{s_0\}, i \notin s \,\big\}$$

where $Front(\mathcal{S})/\{s_0\}$ stands for the hypersite-set which contains all the hypersites in $Front(\mathcal{S})$, except s_0, then introduce:

$$\mathcal{T}^{(s_0)} = \big\{s \cap s_0,\ \forall s \in Front(\mathcal{S})/\{s_0\}\big\}$$

standing for hypersite-set consisting of all the hypersites which are intersections between s_0 and any of the remaining hypersites in $Front(\mathcal{S})$. Furthermore, suppose the local function space $\mathscr{F}^{(s_0/\tilde{s}_0)} = \big\{f : \mathcal{L}^{|s_0/\tilde{s}_0|} \to \mathbb{R}\big\}$, and consider the ortho-marginal operator $\mathscr{O}_{\mathcal{T}^{(s_0)}}$ defined over $\mathscr{F}^{(\tilde{s}_0/s_0)}$, finally, denote by $\mathscr{M}^{(s_0/\tilde{s}_0)}$ the ortho-marginal space which is a function subspace of $\mathscr{F}^{(s_0/\tilde{s}_0)}$ and which is induced by $\mathscr{O}_{\mathcal{T}^{(s_0)}}$ as:

$$\mathscr{M}^{(s_0/\tilde{s}_0)} = \big\{f : \mathcal{L}^{|s_0/\tilde{s}_0|} \to \mathbb{R},\ s.t., f = \mathscr{O}_{\mathcal{T}^{(s_0)}} f\big\}$$

Then, $s_0/\tilde{s}_0$ is "strictly included" in s_0, moreover, whatever a function $u : \mathcal{L}^n \to \mathbb{R}$, one has:

1. $\big(\mathscr{O}_{Front(\mathcal{S})/\{s_0\}}u\big)_{s_0} \in \mathscr{M}^{(s_0/\tilde{s}_0)}$,
2. $\big(\mathscr{O}_{Front(\mathcal{S})/\{s_0\}}u\big)_{c} \equiv u_c,\ \forall c \in \mathcal{T}^{(s_0)}$.

where $\mathscr{O}_{Front(\mathcal{S})/\{s_0\}}$ stands for the ortho-marginal operator with respect to the hypersite-set $Front(\mathcal{S})/\{s_0\}$, and $\forall c \in \mathcal{T}^{(s_0)} \cup \{s_0\}$, $\big(\mathscr{O}_{Front(\mathcal{S})/\{s_0\}}u\big)_c$ stands for the margin of function $\big(\mathscr{O}_{Front(\mathcal{S})/\{s_0\}}u\big)(x)$ with respect to c.

Then, in order to show Lemma 5, assume GlbPseudoΔEMinMLP (27), and denote:

$$\mathbb{Q} = \Big\{ q : \mathcal{L}^n \to \mathbb{R}, s.t., \sum_{x \in \mathcal{L}^n} q(x) = 0, \, and \, \sum_{x_s \in \mathcal{L}^n} |q_s(x_s)| \leq 2, \, \forall s \in \mathcal{S} \Big\}$$

standing for the feasible set of GlbPseudoΔEMinMLP (27), where it has been again assumed that $\forall q \in \mathbb{Q}$, and $\forall s \in \mathcal{S}$, q_s stands for the margin of q with respect to s. Next, since GlbPseudoΔEMinMLP (27) is a convex minimization problem satisfying the Slater condition, for there exists, at least, its feasible solution $q^0 = 0$ satisfying: $\sum_{x_s \in \mathcal{L}^{|s|}} |q^0_s(x_s)| = 0 < 2$, $\forall s \in$ Front($\mathcal{S}$), one derives that there exists $\mu \in \mathbb{R}$, and $\lambda = (\lambda_s)_{s \in Front(\mathcal{S})} \in \mathbb{R}^{+|Front(\mathcal{S})|}$, such that, the Lagrangian functional of GlbPseudoΔEMinMLP (27) defined $\forall q : \mathcal{L}^n \to \mathbb{R}$ as:

$$\mathscr{L}\big(q; \mu, \lambda\big) = \sum_{x \in \mathcal{L}^n} g(x)\, q(x) + \mu \sum_{x \in \mathcal{L}^n} q(x) + \sum_{s \in Front(\mathcal{S})} \lambda_s \left(\sum_{x_s \in \mathcal{L}^{|s|}} |q_s(x_s)| - 2 \right)$$

satisfies $\min_q \big\{ \mathscr{L}\big(q; \mu, \lambda\big) \big\} = \min_{q \in \mathbb{Q}} \big\{ \sum_{x \in \mathcal{L}^n} g(x)\, q(x) \big\}$, Now, assume that $\exists\, s_0 \in$ Front($\mathcal{S}$), such that, $\lambda_{s_0} = 0$, an let us show that this leads to an absurdity. Begin by denoting $\tilde{s}_0 = \big\{ i \in s_0, s.t., \forall s \in Front(\mathcal{S})/\{s_0\}, i \notin s \big\}$ standing for the subset of s_0 consisting of sites in s_0 which do not belong to any $s \in$ Front($\mathcal{S}$)$/\{s_0\}$. Furthermore, assume a function $u : \mathcal{L}^n \to \mathbb{R}$, and put: $q(x) = u(x) - \big(\mathscr{O}_{Front(\mathcal{S})/\{s_0\}} u\big)(x)$, $\forall x \in \mathcal{L}^n$, where $\mathscr{O}_{Front(\mathcal{S})/\{s_0\}}$ stands for the ortho-marginal operator with respect to Front($\mathcal{S}$)$/\{s_0\}$. Then, by Theorem 4, one finds:

$$\forall s \in Front(\mathcal{S}), \, \forall x_s \in \mathcal{L}^{|s|}, \, q_s(x_s) = \begin{cases} 0, \; \textit{if } s \in Front(\mathcal{S})/\{s_0\}, \\ u_{s_0}(x_{s_0}) - \big(\mathscr{O}_{Front(\mathcal{S})/\{s_0\}} u\big)_{s_0}(x_{s_0}), \; \textit{if } s = s_0. \end{cases}$$

where $\forall s \in$ Front($\mathcal{S}$), u_s and q_s respectively stand for the margins of u and q with respect to s, and observe that $\sum_{x_{s_0} \in \mathcal{L}^{|s_0|}} q_{s_0}(x_{s_0}) = 0$. One derives:

$$\begin{array}{rcl} \mathscr{L}\big(q, \mu, \lambda\big) & = & \sum_{x_{s_0} \in \mathcal{L}^{|s_0|}} \big(g_{s_0}(x_{s_0}) + \mu\big)\, q_{s_0}(x_{s_0}) \\ & = & \sum_{x_{s_0} \in \mathcal{L}^{|s_0|}} \big(g_{s_0}(x_{s_0}) + \mu\big) \big(u_{s_0}(x_{s_0}) - \big(\mathscr{O}_{Front(\mathcal{S})/\{s_0\}} u\big)_{s_0}(x_{s_0})\big) \\ & = & \sum_{x_{s_0} \in \mathcal{L}^{|s_0|}} g_{s_0}(x_{s_0}) \big(u_{s_0}(x_{s_0}) - \big(\mathscr{O}_{Front(\mathcal{S})/\{s_0\}} u\big)_{s_0}(x_{s_0})\big) \end{array}$$

Then, since minimization problem $\min_{q \in \mathbb{Q}} \big\{ \sum_{x \in \mathcal{L}^n} g(x) q(x) \big\}$ is bounded, and u is assumed to be arbitrary, one derives:

$$\sum_{x_{s_0} \in \mathcal{L}^{|s_0|}} g_{s_0}(x_{s_0}) \big(u_{s_0}(x_{s_0}) - \big(\mathscr{O}_{Front(\mathcal{S})/\{s_0\}} u\big)_{s_0}(x_{s_0})\big) = 0, \, \forall u : \mathcal{L}^n \to \mathbb{R}$$

which, by Lemma 6, is possible, if and only if, $g_{s_0} \in \mathscr{M}^{(s_0/\tilde{s}_0)}$, and which is an absurdity, as by assumption, g_{s_0} is atomic in Front($\mathcal{S}$), and one derives that $\lambda_{s_0} > 0$. But, since s_0 is assumed to be arbitrary, one finally derives that $\lambda_s > 0$, $\forall s \in$ Front($\mathcal{S}$), hence the proof of Lemma 5.

A.12.2 Proof of Lemma 6

Proof 14 *First, the statement that $s_0/\tilde{s}_0$ is "strictly included" in s_0 follows immediately by definition of Front($\mathcal{S}$). Second, suppose a function $u : \mathcal{L}^n \to \mathbb{R}$, and observe that:*

1. $\forall x \in \mathcal{L}^n$, $\big(\mathscr{O}_{\mathit{Front}(\mathcal{S})/\{s_0\}}u\big)(x)$ *is only a function of* $x_{\Omega/\tilde{s}_0}$ *standing for the sub-vector of* x *with indices in* $\Omega/\tilde{s}_0$*,*

2. $\forall y \in \mathcal{L}^{|s_0|}$, $\big(\mathscr{O}_{\mathit{Front}(\mathcal{S})/\{s_0\}}u\big)_{s_0}(y)$ *is only a function of* $y_{s_0/\tilde{s}_0}$ *standing for the sub-vector of* y *with indices which are in* $s_0/\tilde{s}_0$*,*

3. *By Theorem 4, one has:*

$$\big(\mathscr{O}_{\mathit{Front}(\mathcal{S})/\{s_0\}}u\big)_s \equiv u_s,\ \forall s \in \mathit{Front}(\mathcal{S})/\{s_0\} \tag{60}$$

Thus, one the hand, the function $\big(\mathscr{O}_{\mathit{Front}(\mathcal{S})/\{s_0\}}u\big)_{s_0}(x)$ *only depends on* $x_{s_0/\tilde{s}_0}$*, and one finds that there exists real coefficients* $\beta_c^{(s_0)}, \forall c \in \mathcal{T}^{(s_0)}$*, such that:*

$$\big(\mathscr{O}_{\mathit{Front}(\mathcal{S})/\{s_0\}}u\big)_{s_0}(y) \quad = \quad \textstyle\sum_{c\in\mathcal{T}^{(s_0)}} \beta_c^{(s_0)}\, u_c(y_c),\ \forall y \in \mathcal{L}^{|s_0/\tilde{s}_0|}$$

where $\forall c \in \mathcal{T}^{(s_0)}$, u_c *stands for the margin of* u *with respect to* c*. Thus, one may use the result of Proposition 3 for establishing that* $\big(\mathscr{O}_{\mathit{Front}(\mathcal{S})/\{s_0\}}u\big)_{s_0} \in \mathscr{M}^{(s_0/\tilde{s}_0)}$*. On the other hand, by taking the margins of both hand sides of formula (60) with respect to* $s_0 \cap s, \forall s \in \mathit{Front}(\mathcal{S})/\{s_0\}$*, one establishes that:*

$$\big(\mathscr{O}_{\mathit{Front}(\mathcal{S})/\{s_0\}}u\big)_{s_0\cap s} \equiv u_{s_0\cap s},\ \forall s \in \mathit{Front}(\mathcal{S})/\{s_0\}$$

hence:

$$\big(\mathscr{O}_{\mathit{Front}(\mathcal{S})/\{s_0\}}u\big)_c \equiv u_c,\ \forall c \in \mathcal{T}^{(s_0)}$$

thus, proving Lemma 6.

A.13 Proof of Proposition 6

Proof 15 *Denote by* $\mathscr{M}_\mathcal{S}$ *the ortho-marginal function space with respect to* $\mathscr{M}_\mathcal{S}$ *and by* $\bar{\mathscr{M}}_\mathcal{S}$ *its complement space, and let* q^* *stand for a joint delta-distribution of RMLV* $\mathcal{X}$ *which is optimal for DEMinMLP (13). Clearly,* q^* *has to verify* $q^* = \mathit{Argmin}_{\bar{q}\in\bar{\mathscr{M}}_\mathcal{S}}\big\{\ \sum_{x\in\mathcal{L}^n} |q^*(x) + \bar{q}(x)|\ \big\}$*, as otherwise, one would have* $\min_{\bar{q}\in\bar{\mathscr{M}}_\mathcal{S}}\big\{\ \sum_{x\in\mathcal{L}^n} |q^*(x) + \bar{q}(x)|\ \big\} < 2$ *which, then, is a contradiction with the assumption that* q^* *is optimal for DEMinMLP (13). Therefore, let us study the conditions under which the optimal solution of the problem:*

$$\min_{\bar{q}\in\bar{\mathscr{M}}_\mathcal{S}} \Big\{ \textstyle\sum_{x\in\mathcal{L}^n} |q^*(x) + \bar{q}(x)| \Big\} \tag{61}$$

denoted by $\bar{q}^*$ *verifies that* $\bar{q}^* = 0$*. Thus, suppose an infinitesimal function increment* $v : \mathcal{L}^n \to \mathbb{R}$*, and denote* $v_{\bar{\mathscr{M}}_\mathcal{S}} = \big(\mathit{Id} - \mathscr{O}_\mathcal{S}\big)v$*, where Id stands for the identity operator. Then, one may write:*

$$\textstyle\sum_{x\in\mathcal{L}^n} \big|q^*(x) + v_{\bar{\mathscr{M}}_\mathcal{S}}(x)\big| \quad = \quad \sum_{x\in\mathcal{L}^n} \big|v_{\bar{\mathscr{M}}_\mathcal{S}}(x)\big|\, \mathbf{1}_0\big(q^*(x)\big) + \sum_{x\in\mathcal{L}^n} v_{\bar{\mathscr{M}}_\mathcal{S}}(x)\, \mathit{sign}\big(q^*(x)\big)$$

and observe that, $\forall \alpha \in \mathbb{R}^+$*, one has:*

$$\begin{aligned} &\textstyle\sum_{x\in\mathcal{L}^n} \big|(\alpha v)_{\bar{\mathscr{M}}_\mathcal{S}}(x)\big|\, \mathbf{1}_0\big(q^*(x)\big) + \sum_{x\in\mathcal{L}^n} (\alpha v)_{\bar{\mathscr{M}}_\mathcal{S}}(x)\, \mathit{sign}\big(q^*(x)\big) \\ &= \alpha\Big(\textstyle\sum_{x\in\mathcal{L}^n} \big|v_{\bar{\mathscr{M}}_\mathcal{S}}(x)\big|\, \mathbf{1}_0\big(q^*(x)\big) + \sum_{x\in\mathcal{L}^n} v_{\bar{\mathscr{M}}_\mathcal{S}}(x)\, \mathit{sign}\big(q^*(x)\big)\Big) \end{aligned}$$

Thus, one may relax, henceforth, the assumption that v is an infinitesimal function, and establish that $\bar{q}^* = 0$, if and only if, whatever a function $v : \mathcal{L}^n \to \mathbb{R}$, one has:

$$\begin{cases} \langle |v_{\bar{\mathscr{M}}_\mathcal{S}}|, \mathbf{1}_0(q^*) \rangle \geq \langle v_{\bar{\mathscr{M}}_\mathcal{S}}, sign(q^*) \rangle \\ \langle |v_{\bar{\mathscr{M}}_\mathcal{S}}|, \mathbf{1}_0(q^*) \rangle \geq -\langle v_{\bar{\mathscr{M}}_\mathcal{S}}, sign(q^*) \rangle \end{cases}$$

where the lower inequality is derived by applying the upper inequality with $-v$ in the place of v. But, since minimization problem (61) is convex, which means that its local and global optimality conditions are equivalent, one finally derives that $\bar{q}^* = 0$, if and only if, whatever a function $v : \mathcal{L}^n \to \mathbb{R}$, one has:

$$\langle |v_{\bar{\mathscr{M}}_\mathcal{S}}|, \mathbf{1}_0(q^*) \rangle \geq |\langle v_{\bar{\mathscr{M}}_\mathcal{S}}, sign(q^*) \rangle|$$

hence, the proof of Proposition 6.

A.14 Proof of Theorem 12 and Theorem 13

Proof 16 *Let us first show that, under Assumption 1, both results of Theorem 12 and Theorem 13 are equivalent in a sense that we clarify shortly. First of all, Theorem 3 guarantees that there is a bijection between:*

- *on the one hand, the set of couples of joint distributions of RMLV $\mathcal{X}$ of the form (p^{+*}, p^{-*}), where p^{+*} is optimal for EMinMLP (8), and p^{-*} is optimal for EMaxMLP (9),*
- *on the other hand, the set of joint delta-distributions of RMLV $\mathcal{X}$ which are optimal for DEMinMLP (13).*

Second, let $\mathcal{P}^{+*} = \{p_s^{+*}, \forall s \in \mathcal{S}\}$ and $\mathcal{P}^{-*} = \{p_s^{-*}, \forall s \in \mathcal{S}\}$ stand for an optimal solution of PseudoEMinMLP (22) and an optimal solution of PseudoEMaxMLP (23), respectively. Then, Theorem 9 guarantees that:

- *$\mathcal{P}^{+*}$ defines a marginals-set with respect to $\mathcal{S}$ of a joint distribution of RMLV $\mathcal{X}$ denoted by p^{+*} which is optimal for EMinMLP (8),*
- *$\mathcal{P}^{-*}$ defines a marginals-set with respect to $\mathcal{S}$ of a joint distribution of RMLV $\mathcal{X}$ denoted by p^{-*} which is optimal for EMaxMLP (9).*

Let us put $q^* = p^{+*} - p^{-*}$ of which delta-marginals-sets with respect to $\mathcal{S}$ denoted by $\mathcal{Q}^*$ is given by $\mathcal{Q}^* = \{q_s^* := p_s^{+*} - p_s^{-*}, \forall s \in \mathcal{S}\}$. Therefore, q^* stands (by definition) for a joint delta-distribution of RMLV $\mathcal{X}$ which, by Theorem 3, is optimal for DEMinMLP (13), and hence Theorem 10 guarantees that $\mathcal{Q}^*$ is, in turn, optimal for PseudoΔEMinMLP (24). Furthermore, under Assumption 1 and by Theorem 11, one has:

$$\textstyle\sum_{x_s \in \mathcal{L}^{|s|}} |q_s^*(x_s)| = \sum_{x_s \in \mathcal{L}^{|s|}} |p_s^{+*} - p_s^{-*}| = 2, \; \forall s \in Front(\mathcal{S})$$

and by Proposition 1, one derives:

$$\forall s \in Front(\mathcal{S}), \, \forall x_s \in \mathcal{L}^{|s|}, \, \begin{cases} p_s^{+*} > 0 \Rightarrow p_s^{-*} = 0, \\ p_s^{-*} > 0 \Rightarrow p_s^{+*} = 0. \end{cases}$$

One finally derives:

$$\forall s \in \mathit{Front}(\mathcal{S}),\, \forall x_s \in \mathcal{L}^{|s|},\, \begin{cases} q_s^*(x_s) > 0 \Leftrightarrow p_s^{+*} > 0, \\ q_s^*(x_s) > 0 \Rightarrow p_s^{-*} = 0, \\ q_s^*(x_s) < 0 \Leftrightarrow p_s^{-*} > 0, \\ q_s^*(x_s) < 0 \Rightarrow p_s^{+*} = 0. \end{cases}$$

hence the proof of equivalence between Theorem 12 and Theorem 13. Consequently, one may settle for showing the sufficient condition of Theorem 12 and the necessary condition of Theorem 13.

Thus, let us first show the necessary condition of Theorem 12. First, by using the identity: $\mathbb{P}\big(\mathcal{X}_s = x_s\big) = \sum_{i \notin s} \sum_{x_i \in \mathcal{L}} \mathbb{P}\big(\mathcal{X} = x\big) = 0$, *one finds:*

$$\forall s \in \mathit{Front}(\mathcal{S}),\, \forall x_s \in \mathcal{L}^{|s|},\, \mathbb{P}\big(\mathcal{X}_s = x_s\big) = 0 \Rightarrow \mathbb{P}\big(\mathcal{X} = y\big) = 0, \forall y \in \mathcal{L}^n,\ \mathit{s.t.},\ y_s = x_s \tag{62}$$

and one immediately derives:

$$\forall x \in \mathcal{L}^n,\, \mathbb{P}\big(\mathcal{X} = x\big) > 0 \Rightarrow \forall s \in \mathit{Front}(\mathcal{S}),\, \mathbb{P}\big(\mathcal{X}_s = x_s\big) > 0$$

as otherwise, i.e.; if there existed $x \in \mathcal{L}^n$, *and* $s \in \mathit{Front}(\mathcal{S})$, *such that,* $\mathbb{P}\big(\mathcal{X} = x\big) > 0$, *and* $\mathbb{P}\big(\mathcal{X}_s = x_s\big) = 0$, *then, by formula (62), one would have* $\forall y \in \mathcal{L}^n$, *such that,* $y_s = x_s$, $\mathbb{P}\big(\mathcal{X} = y\big) = 0$, *in particular,* $\mathbb{P}\big(\mathcal{X} = x\big) = 0$, *which is in contradiction with the assumption that* $\mathbb{P}\big(\mathcal{X} = x\big) > 0$, *hence the proof of the necessary condition of Theorem 12. Next, in order to show the sufficient condition of Theorem 12, assume GlbPseudoΔEMinMLP (27), and let us study optimality of the Lagrangian functional we have introduced in formula (59) (see Lemma 5 in Appendix section A.12). First of all, since* $\mathscr{L}\big(q;\mu,\lambda\big)$ *is a convex function of* q, *one concludes that global minimality and local minimality of* $\mathscr{L}\big(q;\mu,\lambda\big)$ *are equivalent. Thus, suppose a function* $q : \mathcal{L}^n \to \mathbb{R}$, *and assume an infinitesimal functional increment* $v : \mathcal{L}^n \to \mathbb{R}$. *Then, one has:*

$$\begin{aligned} &\mathscr{L}\big(q+v;\mu,\lambda\big) - \mathscr{L}\big(q;\mu,\lambda\big) = \textstyle\sum_{x \in \mathcal{L}^n} \big(g(x) + \mu\big)\, v(x) \\ &+ \textstyle\sum_{s \in \mathit{Front}(\mathcal{S})} \lambda_s \Big(\sum_{x_s \in \mathcal{L}^{|s|}} \big(\big|q_s(x_s) + v_s(x_s)\big| - \big|q_s(x_s)\big| \big) \Big) \end{aligned}$$

where $\forall s \in \mathit{Front}(\mathcal{S})$, q_s *and* v_s *stand for the margins of* v *and* q *with respect to* s, *respectively. Since, by assumption,* v *is infinitesimal, one finds that* $\forall s \in \mathit{Front}(\mathcal{S})$, *and* $\forall x_s \in \mathcal{L}^{|s|}$:

$$\big|q_s(x_s) + v_s(x_s)\big| \;=\; |q_s(x_s)| + |v_s(x_s)|\, \mathbf{1}_0\big(q_s(x_s)\big) + v_s(x_s)\, \mathit{sign}\big(q_s(x_s)\big)$$

One may write accordingly:

$$\begin{aligned} &\mathscr{L}\big(q+v;\mu,\lambda\big) - \mathscr{L}\big(q;\mu,\lambda\big) = \textstyle\sum_{x \in \mathcal{L}^n} \big(g(x) + \mu\big)\, v(x) \\ &+ \textstyle\sum_{s \in \mathit{Front}(\mathcal{S})} \lambda_s \sum_{x_s \in \mathcal{L}^{|s|}} v_s(x_s)\, \mathit{sign}\big(q_s(x_s)\big) + \sum_{s \in \mathit{Front}(\mathcal{S})} \lambda_s \sum_{x_s \in \mathcal{L}^{|s|}} |v_s(x_s)|\, \mathbf{1}_0\big(q_s(x_s)\big) \end{aligned}$$

Let us now put:

$$\begin{aligned} &h(q,v) = \textstyle\sum_{x \in \mathcal{L}^n} \big(g(x) + \mu\big)\, v(x) \\ &+ \textstyle\sum_{s \in \mathit{Front}(\mathcal{S})} \lambda_s \sum_{x_s \in \mathcal{L}^{|s|}} v_s(x_s)\, \mathit{sign}\big(q_s(x_s)\big) + \sum_{s \in \mathit{Front}(\mathcal{S})} \lambda_s \sum_{x_s \in \mathcal{L}^{|s|}} |v_s(x_s)|\, \mathbf{1}_0\big(q_s(x_s)\big) \end{aligned}$$

and observe that $\forall \alpha \in \mathbb{R}^+$*, one has* $h(q, \alpha\, v) = \alpha\, h(q, v)$*. Thus, one may relax in the remainder the assumption that* v *is infinitesimal, and establish that* q^* *is a globally minimal solution of* $\mathscr{L}\big(q; \mu, \lambda\big)$*, if and only if, one has:* $h(q^*, v) \geq 0$, $\forall v : \mathcal{L}^n \to \mathbb{R}$*, in other words, if and only if,* $\forall v : \mathcal{L}^n \to \mathbb{R}$*, one has:*

$$\begin{aligned}&\textstyle\sum_{x \in \mathcal{L}^n} \big(g(x) + \mu\big)\, v(x) + \sum_{s \in \mathit{Front}(\mathcal{S})} \lambda_s \sum_{x_s \in \mathcal{L}^{|s|}} v_s(x_s)\, \mathit{sign}\big(q_s^*(x_s)\big)\\ &+ \textstyle\sum_{s \in \mathit{Front}(\mathcal{S})} \lambda_s \sum_{x_s \in \mathcal{L}^{|s|}} |v_s(x_s)|\, \mathbf{1}_0\big(q_s^*(x_s)\big) \geq 0\end{aligned}$$

In particular, by applying the latter inequality, first, with $v(y) = \mathbf{1}_x\big(y\big), \forall y \in \mathcal{L}^n$*, second, with* $v(y) = -\mathbf{1}_x\big(y\big), \forall x \in \mathcal{L}^n$ *with respect to all* $x \in \mathcal{L}^n$*, one derives:*

$$\forall x \in \mathcal{L}^n,\ \begin{cases} g(x) + \mu + \sum_{s \in \mathit{Front}(\mathcal{S})} \lambda_s\, \mathit{sign}\big(q_s^*(x_s)\big) + \sum_{s \in \mathit{Front}(\mathcal{S})} \lambda_s\, \mathbf{1}_0\big(q_s^*(x_s)\big) \geq 0 \\ -g(x) - \mu - \sum_{s \in \mathit{Front}(\mathcal{S})} \lambda_s\, \mathit{sign}\big(q_s^*(x_s)\big) + \sum_{s \in \mathit{Front}(\mathcal{S})} \lambda_s\, \mathbf{1}_0\big(q_s^*(x_s)\big) \geq 0 \end{cases}$$

in such a way that:

$$\forall x \in \mathcal{L}^n,\, q_s^*(x_s) \neq 0,\, \forall s \in \mathit{Front}(\mathcal{S}) \Rightarrow g(x) + \mu + \textstyle\sum_{s \in \mathit{Front}(\mathcal{S})} \lambda_s\, \mathit{sign}\big(q_s^*(x_s)\big) = 0$$

In particular, since one has $q_s^*(x_s^{(inf)}) > 0, \forall s \in \mathit{Front}(\mathcal{S})$*, and* $q_s^*(x_s^{(sup)}) < 0, \forall s \in \mathit{Front}(\mathcal{S})$ *(which is the necessary condition that we have shown earlier in this Appendix section), by applying the latter formula, first, with* $x^{(inf)}$*, second, with* $x^{(sup)}$*, one finds:*

$$\begin{cases} g(x^{(inf)}) + \mu + \sum_{s \in \mathit{Front}(\mathcal{S})} \lambda_s = 0 \\ g(x^{(sup)}) + \mu - \sum_{s \in \mathit{Front}(\mathcal{S})} \lambda_s = 0 \end{cases}$$

hence:

$$\begin{cases} \mu + \sum_{s \in \mathit{Front}(\mathcal{S})} \lambda_s = -g(x^{(inf)}) \\ \mu - \sum_{s \in \mathit{Front}(\mathcal{S})} \lambda_s = -g(x^{(sup)}) \end{cases}$$

Now, suppose $x^0 \in \mathcal{L}^n$*, such that,* $q_s^*(x_s^0) > 0, \forall s \in \mathit{Front}(\mathcal{S})$*. One finds* $g(x^0) + \mu + \sum_{s \in \mathit{Front}(\mathcal{S})} \lambda_s = 0$*, hence,* $g(x^0) = g(x^{(inf)})$*, which means that* x^0 *is a Inf-sample of* q^**. Also, suppose* $x^1 \in \mathcal{L}^n$*, such that,* $q_s^*(x_s^1) < 0, \forall s \in \mathit{Front}(\mathcal{S})$*. One finds* $g(x^1) + \mu - \sum_{s \in \mathit{Front}(\mathcal{S})} \lambda_s = 0$*, hence,* $g(x^1) = g(x^{(sup)})$*, which means that* x^1 *is a Sup-sample of* q^**, which thus establishes the proof of the first and second statements both of Theorem 12 and Theorem 13. The third and last statement of Theorem 13 is a mere consequence of its first and second statements, and the result of Theorem 11.*

A.15 Proof of Theorem 14

Proof 17 *We show Theorem 14 by induction. Therefore, assume* $x_{s_1}^{(1)} \in \mathcal{L}^{|s_1|}$*, such that,* $p_{s_1}^*(x_{s_1}^{(1)}) > 0$*. Then, by Proposition 5, there exists* $x \in \mathcal{L}^n$ *which is a sample of* p^* *and, such that,* $x_{s_1} = x_{s_1}^{(1)}$*, this is on the one hand. On the other hand, since* x *is a sample of* p^**, then by Theorem 12, one has* $p_s^*(x_s) > 0, \forall s \in \mathit{Front}(\mathcal{S})$*, but since one has* $x_{s_1} = x_{s_1}^{(1)}$*, one derives:*

$$\exists x \in \mathcal{L}^n,\ \text{s.t.},\ \big(x_{s_1} = x_{s_1}^{(1)}\big) \wedge \big(p_s^*(x_s) > 0,\, \forall s \in \mathit{Front}(\mathcal{S})\big) \tag{63}$$

Next, suppose $x^{(2)}_{s_2} \in \mathcal{L}^{|s_2|}$*, such that,* $p^*_{s_2}(x^{(2)}_{s_2}) > 0$*, and if* $s_1 \cap s_2 \neq \emptyset$*, then one has* $x^{(s_2)}_{s_1\cap s_2} = x^{(s_1)}_{s_1\cap s_2}$*. Clearly, by formula (63), such a vector* $x^{(2)}_{s_2}$ *exists, moreover, verifies that* $\exists x \in \mathcal{L}^n$*, such that,* $x_{s_2} = x^{(2)}_{s_2} \wedge p^*_s(x_s) > 0, \forall s \in Front(\mathcal{S})/\{s_2\}$*. Then, there exists two possible cases: the case where* $x_{s_1} = x^{(1)}_{s_1}$*, and the case where* $x_{s_1} \neq x^{(1)}_{s_1}$*. Therefore, suppose the latter case, and define the vector* $x' \in \mathcal{L}^n$*, such that,* $x'_{s_1} = x^{(1)}_{s_1}$*,* $x'_{s_2} = x^{(2)}_{s_2}$*, and* $x'_i = x_i, \forall i \notin s_1 \cup s_2$*, and one derives that* x' *verifies* $p^*_s(x'_s) > 0, \forall s \in Front(\mathcal{S})$*. Thus, in both cases, one finds:*

$$\exists x \in \mathcal{L}^n,\ \text{s.t.},\ \big(x_{s_1} = x^{(1)}_{s_1} \wedge x_{s_2} = x^{(2)}_{s_2}\big) \wedge \big(p^*_s(x_s) > 0, \forall s \in Front(\mathcal{S})\big) \tag{64}$$

Now, let $k \in \{2, \ldots, m-1\}$*, and suppose that one has a series of vectors* $x^{(i)}_{s_i} \in \mathcal{L}^{|s_i|}$*,* $i = 1, \ldots, k$*, such that:*

$$\exists x \in \mathcal{L}^n,\ \text{s.t.},\ \big(\wedge_{i=1}^{k} x_{s_i} = x^{(i)}_{s_i}\big) \wedge \big(p^*_s(x_s) > 0, \forall s \in Front(\mathcal{S})\big) \tag{65}$$

and let us show that:

1. $\exists x^{(k+1)}_{s_{k+1}} \in \mathcal{L}^{|s_{k+1}|}$*, such that:*

$$\begin{cases} p^*_{s_{k+1}}(x^{(k+1)}_{s_{k+1}}) > 0 \\ \forall i = 1, \ldots, k,\ s_{k+1} \cap s_i \neq \emptyset \Rightarrow x^{(k+1)}_{s_{k+1}\cap s_i} = x^{(i)}_{s_{k+1}\cap s_i} \end{cases} \tag{66}$$

2. $\forall x^{(k+1)}_{s_{k+1}} \in \mathcal{L}^{|s_{k+1}|}$ *verifying formula (66), one has:*

$$\exists x \in \mathcal{L}^n,\ \text{s.t.},\ \big(\wedge_{i=1}^{k+1} x_{s_i} = x^{(i)}_{s_i}\big) \wedge \big(p^*_s(x_s) > 0, \forall s \in Front(\mathcal{S})\big) \tag{67}$$

First of all, existence of $x^{(k+1)}_{s_{k+1}} \in \mathcal{L}^{|s_{k+1}|}$ *verifying formula (66) above immediately follows, by assumption, from formula (65). Second, suppose* $x^{(k+1)}_{s_{k+1}} \in \mathcal{L}^{|s_{k+1}|}$ *verifying formula (66). Then, by formula (63), one finds that there exists* $y \in \mathcal{L}^n$*, such that,* $y_{s_{k+1}} = x^{(k+1)}_{s_{k+1}}$*, and* $p^*_s(y_s) > 0, \forall s \in Front(\mathcal{S})$*. Then, one distinguishes the following two cases:*

1. $y_{s_i} = x^{(i)}_{s_i}, \forall i = 1, \ldots, k,$
2. $\exists i \in \{1, \ldots, k\}, s.t., y_{s_i} \neq x^{(i)}_{s_i}.$

In the former case, the proof is established immediately. Therefore, suppose the latter case, and define the vector $x' \in \mathcal{L}^n$*, such that,* $x'_{s_i} = x^{(i)}_{s_i}, \forall i = 1, \ldots, k+1$*, and* $x'_{s_i} = y_{s_i}, \forall i = k+2, \ldots, m$*. Then, by formula (66), one finds that such a vector* x' *exists, this is on the one hand. On the other hand, by construction, one has* $p^*_s(x'_s) > 0, \forall s \in Front(\mathcal{S})$*, thus establishing the proof of Theorem 14.*

ALGORITHM 1: Computation of the frontier-closure of a hypersite-set $\mathcal{C}$

Input: Hypersite-set $\mathcal{C}$.
Output: $\mathrm{Fclos}_\cap(\mathcal{C})$.
Construct $\mathrm{Front}(\mathcal{C})$;
Put $m := |\mathrm{Front}(\mathcal{C})|$;
Order $\mathrm{Front}(\mathcal{C})$ as: $\mathrm{Front}(\mathcal{C}) = \{c_1, \ldots, c_m\}$;
Put $\mathrm{Fclos}_\cap(\mathcal{C}) := \{c_1\}$;
$i := 2$;
repeat
    $\mathrm{TmpHypersiteSet} := \{c \cap c_i, \text{s.t.}, c \in \mathrm{Fclos}_\cap(\mathcal{C}) \wedge c \cap c_i \neq \emptyset \wedge c \cap c_i \notin \mathrm{Fclos}_\cap(\mathcal{C})\}$;
    $\mathrm{Fclos}_\cap(\mathcal{C}) := \mathrm{Fclos}_\cap(\mathcal{C}) \cup \{c_i\} \cup \mathrm{TmpHypersiteSet}$;
    i ++;
until $i == m$;
Finally, put $\mathrm{Fclos}_\cap(\mathcal{C}) := \mathrm{Fclos}_\cap(\mathcal{C}) \cup \{\emptyset\}$;

Algorithms

B.16 Construction of the frontier-closure of a hypersite-set

B.17 Obtention of a mode of g from an optimal marginals-set

References

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

Baddeley, A., & Van Lieshout, M. (1993). Stochastic geometry models in high-level vision. *Journal of Applied Statistics*, *1*, 231–256.

Beck, A., & Teboulle, M. (2009). A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM J. Imaging Sciences*, *2*, 183–202.

Bertsimas, D., & Tsitsiklis, J. (1997). *Introduction to Linear Optimization*. Belmont, MA: Athena Scientific.

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

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

Bishop, C. (2006). *Pattern Recognition and Machine Learning*. New York: Springer.

Bolz, J., Farmer, I., Grinspun, E., & Schrooder, P. (2003). Sparse matrix solvers on the gpu: Conjugate gradients and multigrid. *ACM Transactions on Graphics*, *22*, 917–924.

Boykov, Y., Veksler, O., & Zabih, R. (1998). Markov random fields with efficient approximations. In *Proceedings of CVPR*.

Boykov, Y., Veksler, O., & Zabih, R. (2001). Fast approximate energy minimization via graph cuts. *IEEE Trans. Pattern Anal. Mach. Intell.*, *23*, 1222–1239.

ALGORITHM 2: Computation of a MAP solution from an optimal marginals-set

Input: Optimal probability marginals-set $\mathcal{P}^* = \big\{p_s^*, \forall s \in \mathcal{S}\big\}$
Output: One sample $x \in \mathcal{L}^n$ of mother distribution p^*
Construct Front$\big(\mathcal{S}\big)$;
Order Front$\big(\mathcal{S}\big)$ as: Front$\big(\mathcal{S}\big) = \big\{s_1, \ldots, s_m\big\}$;
Put $IS\_ATOMIC\_ALL := FALSE$;
repeat
  Put k:=1 ;
  Put $CONTINUE := TRUE$;
  repeat
    if g_{s_k} *is NOT atomic in Front*$\big(\mathcal{S}\big)$*;*
    then
      Front$\big(\mathcal{S}\big)$:= Front$\big(\mathcal{S}\big)/\{s_k\}$;
      Put $m := m$-1 ;
      Reorder Front$\big(\mathcal{S}\big)$ accordingly as Front$\big(\mathcal{S}\big) = \big\{s_1, \ldots, s_m\big\}$;
      Put $CONTINUE := FALSE$;
    end
  until $CONTINUE == FALSE$ *or* $k == m$;
  if $CONTINUE == TRUE$ *;*
  then
    Put $IS\_ATOMIC\_ALL := TRUE$;
  end
  k ++;
until $IS\_ATOMIC\_ALL = TRUE$;
Choose any $x_{s_1}^{(1)} \in \mathcal{L}^{|s_1|}$ s.t. $p_s^*(x_{s_1}^{(1)}) > 0$;
Put $x_{s_1} := x_{s_1}^{(1)}$;
Put $W := \big\{i \in s_1\big\}$;
Put $k := 2$;
repeat
  Put $t := \big\{i \in \Omega \text{ s.t. } i \in s_k \wedge i \notin W\big\}$;
  if $t \neq \emptyset$ *;*
  then
    Choose any $x_{s_k}^{(k)} \in \mathcal{L}^{|s_k|}$ s.t. $p_{s_k}^*\big(x_{s_k}^{(k)}\big) > 0 \wedge_{i=1}^{k-1} \big(x_{s_i \cap s_k}^{(i)} == x_{s_i \cap s_k}^{(k)}\big)$;
    Update x as: $x_t := x_t^{(k)}$;
    Update W as: $W := W \cup \{i \in s_k\}$;
  end
  k ++;
until $W == \Omega$;

Chekuri, C., Khanna, S., Naor, J., & Zosin, L. (2005). Linear programming formulation and approximation algorithms for the metric labeling problem. *SIAM Journal on Discrete Mathematics*, *18*, 608–625.

Chickering, D. (1996). Learning bayesian networks is np-complete. *Lecture Notes in Statistics*, *112*, 21130.

Chvàtal, V. (1983). *Linear Programming*. New York: W.H. Freeman and Company.

Cipra, B. (2000). The ising model is np-complete. *SIAM News*, *33*, 6.

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

Cooper (2012). Tractable triangles and cross-free convexity in discrete optimisation. *Journal of Artificial Intelligence Research*, *44*, 455–490.

Cooper, G. (1990). The computational complexity of probabilistic inference using bayesian belief networks. *Artificial Intelligence*, *42*, 393–405.

Cremers, D., Pock, T., Kolev, K., & Chambolle, A. (2011). *Convex Relaxation Techniques for Segmentation, Stereo and Multiview Reconstruction*. MIT Press.

Dagum, P., & Luby, M. (1993). Approximating probabilistic inference in bayesian belief networks is np-hard. *Artificial Intelligence*, *60*, 141–153.

Dantzig, G. (1990). *Origins of the simplex method*. ACM Press.

Descombes, X. (2011). *Stochastic Geometry for Image Analysis*. France: Wiley-ISTE.

Gelfand, A., & Smith, A. (1990). Sampling-based approaches to calculating marginal densities. *Journal of the American Statistical Association*, *85*, 398–409.

Geman, S., & Geman, D. (1984). Stochastic relaxation, gibbs distributionsand the bayesian restoration of images. *IEEE Trans. Pattern Anal. Mach. Intell.*, *6*, 721–741.

Geman, S., & Graffigne, C. (1986). Markov random field image models and their applications to computer vision. In *Proceedings of the International Congress of Mathematicians*.

Green, P. (1995). Reversible jump markov chains monte carlo computation and bayesian model determination. *Biometrika*, *82*, 711–732.

Greig, D., Porteous, B., & Seheult, A. (1989). Exact a posteriori maximum estimation in binary images. *J. Royal Statistical Soc. Series B*, *5*, 271–279.

Grootschel, M., Lovasz, L., & Schrijver, A. (1993). *Geometric Algorithms and Combinatorial Optimization*. Springer-Verlag.

Hammer, P. L., Hansen, P., & Simeone, B. (1984). Roof duality, complementation and persistency in quadratic 0-1 optimization. *Mathematical Programming*, *28*, 121–155.

Hammer, P. (1965). Some network flow problems solved with pseudo-boolean programming. *Operations Research*, *13*, 388–399.

Hammersley, J., & Clifford, P. (1971). *Markov fields on finite graphs and lattices*. Unpublished.

Hastings, W. (1970). Monte carlo sampling methods using markov chains and their applications. *Biometrika*, *97-109*, 57.

Hummel, R., & Zucker, S. (1983). On the foundations of relaxation labeling processes. *IEEE Trans. Pattern Anal. Mach. Intell.*, *5*, 267–287.

Ishikawa, H. (2003). Exact optimization for markov random fields with convex priors. *IEEE Trans. Pattern Anal. Mach. Intell.*, *25*, 1333–1336.

Ising, E. (1925). Beitrag zur theorie des ferromagnetismus. *Z. Phys.*, *31*, 253–258.

Karmarkar, N. (1984). A new polynomial-time algorithm for linear programming. In *Proceedings of STOC*.

Karp, R. (1972). *Reducibility among combinatorial problems*. New York: Plenum Press.

Kinderman, R., & Snell, J. (1980). Markov random fields and their applications. *American Mathematical Society*, *1*, 1–142.

Kingsford, C., Chazelle, B., & Singh, M. (2005). Solving and analyzing side-chain positioning problems using linear and integer programming. *Bioinformatics*, *21*, 1028–1039.

Kolmogorov, V. (2006a). Convergent tree-reweighted message passing for energy minimization. *IEEE Trans. on Patt. Anal. and Mach. Intell.*, *28*, 1568–1583.

Kolmogorov, V. (2006b). Convergent tree-reweighted message passing for energy minimization. *IEEE Trans. Pattern Analysis and Machine Intelligence*, *28*, 1568–1583.

Kolmogorov, V., & Zabih, R. (2004). What energy functions can be minimized via graph cuts?. *IEEE Trans. Pattern Anal. Mach. Intell.*, *26*, 1222–1239.

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

Koster, A., Van Hoesel, C. P. M., & Kolen, A. (1998). The partial constraint satisfaction problem: Facets and lifting theorems. *Operations Research Letters*, *23*, 89–97.

Lauritzen, S. (1991). *Graphical models*. Oxford Science Publications.

Lellmann, J., & Schnorr, C. (2011). Continuous multiclass labeling approaches and algorithms. *SIAM J. Imaging Sci.*, *4*, 10491096.

Lesaja, G. (2009). Introducing interior-point methods for introductory operations research courses and/or linear programming courses. *Open Operational Research Journal*, *3*, 1–12.

Li, J., Lu, R., Hu, X., & Jiang, Z. (2011). A gpu-based parallel algorithm for large scale linear programming problem. *Intelligent Decision Technologies. Smart Innovation, Systems and Technologies*, *10*, 37–46.

Li, S. (1995). *Markov Random Field Modeling in Computer Vision*. Springer-Verlag.

Liu, J. S. X., & Veksler, O. (2010). Order preserving moves for graph cut based optimization. *IEEE Trans. Pattern Anal. Mach. Intell.*, *32*, 1182–1196.

Megretski, A. (1996). *Learning Bayesian Networks is NP-Complete*. New York: Plenum Press.

Nemhauser, G., & Wolsey, L. (1988). *Integer and combinatorial optimization*. Wiley.

Nesterov, Y. (2004). *Introductory lectures on convex optimization*. SIAM.

Nesterov, Y. (2009). Primal-dual subgradient methods for convex problems. *Mathematical programming*, *120*, 221–259.

Nesterov, Y., & Nemirovsky, A. (1994). *Interior-point polynomial methods in convex programming*. SIAM.

Nieuwenhuis, C., Toeppe, E., & Cremers, D. (2013). A survey and comparison of discrete and continuous multi-label optimization approaches for the potts model. *International Journal of Computer Vision*, *104*, 223–240.

Nikolova, M., Esedoglu, S., & Chan, T. (2006). Algorithms for finding global minimizers of image segmentation and denoising models. *SIAM Journal of Applied Mathematics*, *66*, 1632–1648.

Osokin, A., Vetrov, D., & Kolmogorov, V. (2011). Sub-modular decomposition framework for inference in associative markov networks with global constraints. In *Proceedings of CVPR*.

Park, J., & Darwiche, A. (2004). Complexity results and approximation strategies for map explanations. *Journal of Artificial Intelligence Research*, *21*, 101133.

Pearl, J. (1982). Reverend bayes on inference engines: A distributed hierarchical approach. In *Proceedings of ec. Nat. Conf. Artif. Intell., AAAI-82*.

Pearl, J. (1988). *Probabilistic reasoning in intelligent systems*. San Mateo:Morgan Kauffmann.

Pearl, J., & Russell, S. (2002). *Bayesian Networks*. MIT Press.

Potts, R. (1952). Some generalized order-disorder transformations. In *Proceedings of the Cambridge Philosophical Society*.

Roth, D. (1996). On the hardness of approximate reasoning. *Artificial Intelligence*, *273-302*, 82.

Schlesinger, D. (2007). Exact solution of permuted submodular minsum problems. In *Proceedings of EMMCVPR*.

Schlesinger, M. (1976). Syntactic analysis of two-dimensional visual signals in noisy conditions. *Cybernetics and Systems Analysis*, *12*, 612–628.

Sherali, H., & Adams, W. (1990). A hierarchy of relaxations between the continuous and convex hull representations for zero-one programming problems. *SIAM J. Discrete Math.*, *3*, 411–430.

Shimony, Y. (1994). Finding the maps for belief networks is np-hard. *Aritifical Intelligence*, *68*, 399–410.

Sontag, D., & Jaakkola, T. (2008). New outer bounds on the marginal polytope. *Advances in Neural Information Processing Systems, MIT Press*, *21*, 33.

Veksler, O. (2012). Multi-label moves for mrfs with truncated convex priors. *International Journal of Computer Vision*, *98*, 1–14.

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

Wainwright, M., & Jordan, M. (2008). *Graphical models, exponential families, and variational inference*. Foundations and Trends in Machine Learning.

Werner, T. (2007). A linear programming approach to max-sum problem: A review. *IEEE Trans. Pattern Anal. Mach. Intell.*, *29*, 1165–1179.

Winkler, G. (1995). *Image Analysis, Random Fields and Markov Chain Monte Carlo Methods: A Mathematical Introduction*. Springer.

Ye, Y. (1989). *An extension of Karmarkars algorithm and the trust region method for convex quadratic programming*. Springer-Verlag.

Zach, C., Hane, C., & Pollefeys, M. (2014). What is optimized in convex relaxations for multilabel problems: Connecting discrete and continuously inspired map inference. *IEEE Trans. Pattern Anal. Mach. Intell.*, *36*, 157–170.