

Tail dependence of recursive max-linear models with regularly varying noise variables

Nadine Gissibl, Claudia Klüppelberg, Moritz Otto

Center for Mathematical Sciences, Technical University of Munich, Boltzmannstrasse 3, Garching 85748, Germany

Abstract

Recursive max-linear structural equation models with regularly varying noise variables are considered. Their causal structure is represented by a directed acyclic graph (DAG). The problem of identifying a recursive max-linear model and its associated DAG from its matrix of pairwise tail dependence coefficients is discussed. For example, it is shown that if a causal ordering of the associated DAG is additionally known, then the minimum DAG representing the recursive structural equations can be recovered from the tail dependence matrix. For the relevant subclass of recursive max-linear models, identifiability of the associated minimum DAG from the tail dependence matrix and the initial nodes is shown. Algorithms find the associated minimum DAG for the different situations. Furthermore, given a tail dependence matrix, an algorithm outputs all compatible recursive max-linear models and their associated minimum DAGs.

Keywords: causal inference, directed acyclic graph, graphical model, max-linear model, max-stable model, regular variation, structural equation model, extreme value theory, tail dependence coefficient

2010 MSC: 60G70, 05C75, 62-09, 65S05

1. Introduction

Causal inference is fundamental in virtually all areas of science. Examples for concepts established over the last years to understand causal inference include structural equation modeling (see e.g. [Bollen, 1989](#); [Pearl, 2009](#)) and graphical modeling (see e.g. [Lauritzen, 1996](#); [Spirtes et al., 2000](#); [Koller and Friedman, 2009](#)).

In extreme risk analysis it is especially important to understand causal dependencies. We consider *recursive max-linear models* (RMLMs), which are max-linear structural equation models whose causal structure is represented by a *directed acyclic graph* (DAG). Such models are directed graphical models ([Pearl, 2009](#), Theorem 1.4.1); i.e., the DAG encodes conditional independence relations in the distribution via the (directed global) Markov property. RMLMs were introduced and studied in [Gissibl and Klüppelberg \(2017\)](#). They may find their application in situations when extreme risks play an essential role and may propagate through a network, for example, when modeling water levels or pollution concentrations in a river or when modeling risks in a large industrial structure. In [Einmahl et al. \(2017\)](#) a RMLM was fitted to data from the EURO STOXX 50 Index, where the DAG structure was assumed to be known.

In this paper we assume *regularly varying* noise variables. This leads to models treated in classical multivariate extreme value theory. The books by [Beirlant et al. \(2004\)](#), [de Haan and Ferreira \(2006\)](#), and [Resnick \(1987, 2007\)](#) provide a detailed introduction into this field. A RMLM with regularly varying noise variables is in the *maximum domain of attraction* of an *extreme value (max-stable) distribution*. The spectral measure of the limit distribution, which describes the dependence structure given by the DAG, is discrete. Every max-stable random vector with discrete spectral measure is *max-linear* (ML), and every multivariate max-stable

Email addresses: n.gissibl@tum.de (Nadine Gissibl), cklu@tum.de (Claudia Klüppelberg), moritz.otto@kit.edu (Moritz Otto)

distribution can be approximated arbitrarily well via a ML model (e.g. [Yuen and Stoev, 2014](#), Section 2.2). This demonstrates the important role of ML models in extreme value theory. They have been investigated, generalized, and applied to real world problems by many researchers; see e.g. [Schlather and Tawn \(2002\)](#), [Wang and Stoev \(2011\)](#), [Falk et al. \(2015\)](#), [Strokorb and Schlather \(2015\)](#), [Einmahl et al. \(2012\)](#), [Cui and Zhang \(2017\)](#), and [Kiriliouk \(2017\)](#).

One main research problem that is addressed for restricted recursive structural equation models, where the functions are required to belong to a specified function class, is the *identifiability* of the coefficients and the DAG from the observational distribution. Recently, particular attention in this context has been given to recursive structural equation models with additive Gaussian noise; see e.g. [Peters et al. \(2014\)](#), [Ernest et al. \(2016\)](#), and references therein. For RMLMs this problem is investigated in [Gissibl et al. \(2017\)](#). In the present paper we discuss the identifiability of RMLMs from their (*upper*) *tail dependence coefficients* (TDCs).

The TDC, which goes back to [Sibuya \(1960\)](#), measures the extremal dependence between two random variables and is a simple and popular dependence measure in extreme value theory. Methods to construct multivariate max-stable distributions with given TDCs have been proposed, for example, by [Schlather and Tawn \(2002\)](#), [Falk \(2005\)](#), [Falk et al. \(2015\)](#), and [Strokorb and Schlather \(2015\)](#). Somehow related we identify all RMLMs with the same given TDCs.

1.1. Problem description and important concepts

First we briefly review RMLMs and introduce the TDC formally. We then describe the idea of this work in more detail and state the main results.

Max-linear models on DAGs

Consider a RMLM $\mathbf{X} = (X_1, \dots, X_d)$ on a DAG $\mathcal{D} = (V, E)$ with nodes $V = \{1, \dots, d\}$ and edges $E = \{(k, i) : i \in V \text{ and } k \in \text{pa}(i)\}$:

$$X_i = \bigvee_{k \in \text{pa}(i)} c_{ki} X_k \vee c_{ii} Z_i, \quad i = 1, \dots, d, \quad (1)$$

where $\text{pa}(i)$ denotes the parents of node i in \mathcal{D} and $c_{ki} > 0$ for $k \in \text{pa}(i) \cup \{i\}$; the noise variables Z_1, \dots, Z_d , represented by a generic random variable Z , are assumed to be independent and identically distributed with support $\mathbb{R}_+ := (0, \infty)$ and *regularly varying* with index $\alpha \in \mathbb{R}_+$, abbreviated by $Z \in \text{RV}(\alpha)$. Denoting the distribution function of Z by F_Z , the latter means that

$$\lim_{t \rightarrow \infty} \frac{1 - F_Z(xt)}{1 - F_Z(t)} = x^{-\alpha}$$

for every $x \in \mathbb{R}_+$. Examples for F_Z include Cauchy, Pareto, and log-gamma distributions. For details and background on regular variation, see e.g. [Resnick \(1987, 2007\)](#).

The properties of the noise variables imply the existence of a normalizing sequence $a_n \in \mathbb{R}_+$ such that for independent copies $\mathbf{X}^{(1)}, \dots, \mathbf{X}^{(n)}$ of \mathbf{X} ,

$$a_n^{-1} \bigvee_{\nu=1}^n \mathbf{X}^{(\nu)} \xrightarrow{d} \mathbf{M}, \quad n \rightarrow \infty, \quad (2)$$

where \mathbf{M} is a non-degenerate random vector with distribution function denoted by G and all operations are taken componentwise. Thus \mathbf{X} is in the maximum domain of attraction of G ; we write $\mathbf{X} \in \text{MDA}(G)$. The limit vector \mathbf{M} (its distribution function G) is necessarily max-stable: for all $n \in \mathbb{N}$ and independent copies $\mathbf{M}^{(1)}, \dots, \mathbf{M}^{(n)}$ of \mathbf{M} , the distributional equality $\mathbf{a}_n \mathbf{M} + \mathbf{b}_n \stackrel{d}{=} \bigvee_{\nu=1}^n \mathbf{M}^{(\nu)}$ holds for appropriately chosen normalizing sequences $\mathbf{a}_n \in \mathbb{R}_+^d$ and $\mathbf{b}_n \in \mathbb{R}^d$. In the present situation we have $\mathbf{a}_n = (n^{1/\alpha}, \dots, n^{1/\alpha})$ and $\mathbf{b}_n = (0, \dots, 0)$. Furthermore, \mathbf{M} is again a RMLM on \mathcal{D} , with the same weights in (1) as \mathbf{X} and standard α -Fréchet distributed noise variables, i.e.,

$$F_Z(x) = \Phi_\alpha(x) = \exp\{-x^{-\alpha}\}, \quad x \in \mathbb{R}_+.$$

A proof of (2) as well as an explicit formula for G and its univariate and bivariate marginal distributions can be found in Appendix A.2, Proposition A.2.

In what follows we summarize the most important properties of \mathbf{X} presented in Gissibl and Klüppelberg (2017) which are needed throughout the paper. Every component of \mathbf{X} can be written as a max-linear function of its ancestral noise variables:

$$X_i = \bigvee_{j \in \text{An}(i)} b_{ji} Z_j, \quad i = 1, \dots, d, \quad (3)$$

where $\text{An}(i) = \text{an}(i) \cup \{i\}$ and $\text{an}(i)$ are the ancestors of i in \mathcal{D} (Gissibl and Klüppelberg, 2017, Theorem 2.2). For $i \in V$, $b_{ii} = c_{ii}$. For $j \in \text{an}(i)$, b_{ji} can be determined by a path analysis of \mathcal{D} as explained in the following. Throughout we write $k \rightarrow i$ whenever \mathcal{D} has an edge from k to i . With every path $p = [j = k_0 \rightarrow k_1 \rightarrow \dots \rightarrow k_n = i]$ we associate a weight, which we define to be the product of the edge weights along p multiplied by c_{jj} . The coefficient b_{ji} is then the maximum weight of all paths from j to i . In summary, we have for $i \in V$ and $j \in \text{an}(i)$,

$$b_{ji} = \bigvee_{p \in P_{ji}} d_{ji}(p) \quad \text{with} \quad d_{ji}(p) := c_{k_0 k_0} \prod_{\nu=0}^{n-1} c_{k_\nu k_{\nu+1}}, \quad (4)$$

where P_{ji} is the set of all paths from j to i . For all $i \in V$ and $j \in V \setminus \text{An}(i)$ we set $b_{ji} = 0$. We call the coefficients b_{ji} *ML coefficients* (MLCs) and summarize them in the *ML coefficient matrix* (MLCM) $B = (b_{ij})_{d \times d}$. For the *reachability matrix* R of \mathcal{D} , whose ji -th entry is one if $j \in \text{An}(i)$ and zero else, we find

$$R = \text{sgn}(B), \quad (5)$$

where sgn denotes the signum function and is taken componentwise. As a consequence, the ancestors and descendants of every node in \mathcal{D} can be obtained from B .

Not all paths are needed for computing b_{ji} in (4). We call a path p from j to i *max-weighted path from j to i* if it realizes the maximum in (4), i.e., if $b_{ji} = d_{ji}(p)$. The concept of max-weighted paths is essential. This has been worked out in Gissibl and Klüppelberg (2017). For example, max-weighted paths may lead to more conditional independence relations in the distribution of \mathbf{X} than those encoded by \mathcal{D} via the Markov property (Gissibl and Klüppelberg, 2017, Remark 3.9). RMLMs where all paths are max-weighted play a central role in this paper; we call them *recursive max-weighted models* (RMWMs).

Further DAGs and weights may exist such that \mathbf{X} satisfies (1); for a detailed characterization of these DAGs and weights, see Theorem 5.4 of Gissibl and Klüppelberg (2017). The smallest DAG of this kind is the one that has an edge $k \rightarrow i$ if and only if (iff) this is the only max-weighted path from k to i in \mathcal{D} (Gissibl and Klüppelberg, 2017, Remark 5.2(ii) and Theorem 5.4(a)). We call this DAG \mathcal{D}^B , the *minimum ML DAG of \mathbf{X}* . It can be determined from B (Gissibl and Klüppelberg, 2017, Theorem 5.3). The other DAGs representing \mathbf{X} in the sense of (1) are those that have at least the edges of \mathcal{D}^B and the same reachability matrix. For edges contained in \mathcal{D}^B , the weights from (1) are uniquely defined by B . From these weights the weights for the other edges can be derived.

Remark 1.1. The random vector \mathbf{X} and its distribution are characterized by the distribution F_Z of the noise variables and the max-linear dependence structure induced by \mathcal{D} . So computing the max-stable limit distribution G concerns only the marginal limits, whereas the max-linear dependence structure remains always the same (cf. also the proof of Proposition A.2). This restrictive dependence structure of \mathbf{X} can be generalized naturally within the framework of multivariate regular variation. See Resnick (1987, 2007) for background on multivariate regular variation.

In the literature various equivalent formulations of regular variation for random vectors can be found. The extent of a possible generalization can be probably best understood when considering an equivalent representation of the dependence in a regularly varying vector. A random vector $\mathbf{X} \in \mathbb{R}_+^d$ is regularly varying with index $\alpha \in \mathbb{R}_+$ iff there exists a random vector Θ with values in $\mathbb{S}^{d-1} = \{\mathbf{x} \in \mathbb{R}_+^d : \|\mathbf{x}\| = 1\}$, where $\|\cdot\|$ is any norm in \mathbb{R}_+^d , such that for every $x \in \mathbb{R}_+$,

$$\frac{\mathbb{P}(\|\mathbf{X}\| > tx, \mathbf{X}/\|\mathbf{X}\| \in \cdot)}{\mathbb{P}(\|\mathbf{X}\| > t)} \xrightarrow{v} x^{-\alpha} \mathbb{P}(\Theta \in \cdot), \quad t \rightarrow \infty. \quad (6)$$

The notation \xrightarrow{v} stands for vague convergence on the Borel σ -algebra of \mathbb{S}^{d-1} . We immediately find from (6) that the dependence structure of \mathbf{X} is for moderate values of $\|\mathbf{X}\|$ arbitrary; only when $\|\mathbf{X}\|$ becomes large, the dependence structure becomes that of Θ . When assuming that the dependence structure in the limit is max-linear given by \mathcal{D} and the marginal limits are α -Fréchet (with an appropriate scale parameter), then $\mathbf{X} \in \text{MDA}(G)$ with G still as in Proposition A.2; hence, \mathbf{X} would have the same TDCs as in the present less general framework. So similarly to the flexibility of the margins, expressed by $Z \in \text{RV}(\alpha)$, there would also be flexibility in the dependence structure.

In this paper the restriction to the limiting max-linear dependence provides a sufficient model as the focus lies on the causal structure in terms of the DAGs. This allows for a more concise notation and makes the focus of the paper more transparent. \square

The tail dependence matrix of \mathbf{X}

For $i \in V$ we denote the distribution function of component X_i of the RMLM \mathbf{X} by F_i and its generalized inverse by $F_i^{\leftarrow}(u) = \inf\{x \in \mathbb{R}_+ : F_i(x) \geq u\}$ for $0 < u < 1$. The TDC between X_i and X_j is then given by the limit

$$\chi(i, j) = \lim_{u \uparrow 1} \mathbb{P}(X_i > F_i^{\leftarrow}(u) \mid X_j > F_j^{\leftarrow}(u)).$$

We summarize all TDCs in the *tail dependence matrix* (TDM) $\chi = (\chi(i, j))_{d \times d}$.

Defining the *standardized MLCM* of \mathbf{X} by

$$\overline{B} = (\overline{b}_{ij})_{d \times d} := \left(\frac{b_{ij}^\alpha}{\sum_{k \in \text{An}(j)} b_{kj}^\alpha} \right)_{d \times d}, \quad (7)$$

the TDC between X_i and X_j can be computed as

$$\chi(i, j) = \chi(j, i) = \sum_{k \in \text{An}(i) \cap \text{An}(j)} \overline{b}_{ki} \wedge \overline{b}_{kj}. \quad (8)$$

By (5) and (7) it is the sum of the pairwise minima of the i -th and j -th column of \overline{B} . A proof of (8) is given in Appendix A.2. There we implicitly show that \mathbf{X} and the limit vector \mathbf{M} from (2) have the same TDM χ .

The TDC $\chi(i, j)$ is zero iff i and j do not have common ancestors. Therefore, the *initial nodes* of \mathcal{D} (i.e., the nodes without parents) constitute a set V_0 of maximum cardinality such that $\chi(i, j)$ is zero for all distinct $i, j \in V_0$. This property turns out to be helpful when identifying from χ . We also show that $\chi(i, j)$ is zero iff X_i and X_j are independent, which is reminiscent of the multivariate Gaussian distribution with its equivalence between independence and zero correlation.

Obviously, when investigating χ , understanding the structure of \overline{B} is essential. Not surprisingly, \overline{B} inherits structural properties from B . For example, \overline{B} is again a MLCM of a RMLM on the same DAG \mathcal{D} , and its columns add up to one. Properties of \overline{B} , which we use throughout this paper, are summarized in Appendix A.1, Lemma A.1.

Identifiability from χ

The main goal of this paper is to investigate how far the dependence structure of \mathbf{X} and the DAG \mathcal{D} can be recovered from the TDM χ . We call two RMLMs that have the same TDM χ -*equivalent*. For example, \mathbf{X} and the limit vector \mathbf{M} from (2) are χ -equivalent. The set

$$\{(\tilde{b}_{ij})_{d \times d} \in \mathbb{R}_+^{d \times d} : \tilde{b}_{ij} = \beta_j \overline{b}_{ij}^{1/\tilde{\alpha}} \text{ for all } i, j \in V \text{ and } \beta_j \in \mathbb{R}_+\} \quad (9)$$

contains the MLCMs of all RMLMs that have the same standardized MLCM \overline{B} as \mathbf{X} and regularly varying noise variables with index $\tilde{\alpha} \in \mathbb{R}_+$; this can be verified by using Theorem 5.7 of Gissibl and Klüppelberg (2017). Obviously, all the corresponding RMLMs are also χ -equivalent to \mathbf{X} . Therefore, given χ only, we can never identify the true representations (1) and (3) of \mathbf{X} and the DAG \mathcal{D} .

The RMLM \mathbf{X} has the same minimum ML DAG \mathcal{D}^B as every RMLM with MLCM \overline{B} (Lemma A.1(e)). As a consequence, \mathcal{D}^B can be determined from \overline{B} (cf. Gissibl and Klüppelberg, 2017, Theorem 5.3). This raises the question of whether \overline{B} and, hence, the minimum ML DAG of \mathbf{X} are identifiable from χ . The answer is generally no, quite simply due to the symmetry of χ .

Example 1.2. [\overline{B} is not identifiable from χ]

Consider two RMLMs on the DAGs \mathcal{D}_1 and \mathcal{D}_2 with standardized MLCMs

$$\mathcal{D}_1 \quad \textcircled{1} \longrightarrow \textcircled{2} \quad \overline{B}_1 = \begin{pmatrix} 1 & b \\ 0 & 1-b \end{pmatrix} \quad \text{and} \quad \overline{B}_2 = \begin{pmatrix} 1-b & 0 \\ b & 1 \end{pmatrix} \quad \textcircled{1} \longleftarrow \textcircled{2} \quad \mathcal{D}_2$$

for some $b \in (0, 1)$. For both we find the same TDCs: $\chi(1, 1) = \chi(2, 2) = 1$ and $\chi(1, 2) = \chi(2, 1) = b$. \square

We show, however, that \overline{B} can be computed recursively from χ and some additional information on the DAG \mathcal{D} . This may be its reachability matrix R but also only a *causal ordering* σ ; i.e., σ is a permutation on $V = \{1, \dots, d\}$ such that $\sigma(j) < \sigma(i)$ for all $i \in V$ and $j \in \text{an}(i)$. If \mathbf{X} is max-weighted, then \overline{B} is identifiable from χ and the initial nodes V_0 of \mathcal{D} .

The question also arises which RMLMs are all χ -equivalent to \mathbf{X} and what their minimum ML DAGs are. Since by (9) every MLCM of a RMLM with TDM χ can be obtained from its particular standardized version, it suffices to clarify which the standardized MLCMs of all RMLMs with TDM χ are. To this end we use the identifiability results mentioned above to develop an algorithm that computes these matrices from χ . The proposed procedure can be considerably simplified for RMWMs.

Another interesting point is how DAGs of χ -equivalent RMLMs relate to each other. Here we also investigate the RMWMs as a relevant subclass of RMLMs separately. For example, an initial node in a DAG of a RMWM is again an initial node in a DAG of a χ -equivalent RMWM or it must be a terminal node (i.e., a node without descendants).

Our paper is organized as follows. We provide some basic results in Section 2. For a RMLM \mathbf{X} we investigate its TDM χ and link it to its standardized MLCM \overline{B} and its associated DAG \mathcal{D} . Here we discuss the situations when two components of \mathbf{X} have zero tail dependence. We also introduce the important concept of χ -cliques, which allows us to identify potential initial node sets in \mathcal{D} from χ . Section 3 is devoted to RMWMs. We point out the specific properties of χ which lead to the identifiability of \overline{B} from χ and the initial nodes. We also present necessary and sufficient conditions on a matrix to be the TDM of a RMWM. In Section 4 we then study different identifiability problems based on χ . We propose algorithms to compute \overline{B} from χ and some further information on \mathcal{D} such as a causal ordering. We also explain how the standardized MLCMs of all RMLMs that have TDM χ can be determined. In Section 5 we consider χ -equivalent RMLMs and analyze relationships between them and their DAGs. We use these results to investigate whether RMWMs on different DAGs can be χ -equivalent at all and if so under which conditions. Section 6 concludes.

Note that all recursion formulas presented in the paper are well-defined, since we work with DAGs. Throughout we illustrate our findings with examples for the (standardized) MLCM of a RMLM on a given DAG. It can be verified by Theorem 4.2 or Corollary 4.3(a) of Gissibl and Klüppelberg (2017) that the presented matrices are indeed MLCMs of RMLMs on the particular DAGs. Moreover, we use the following notation throughout the paper. We denote the ancestors, parents, and descendants of node i in \mathcal{D} by $\text{an}(i)$, $\text{pa}(i)$, and $\text{de}(i)$, respectively. We define $\text{An}(i) := \text{an}(i) \cup \{i\}$, $\text{Pa}(i) := \text{pa}(i) \cup \{i\}$, and $\text{De}(i) := \text{de}(i) \cup \{i\}$. For (possibly random) $a_i \in \mathbb{R}$ we set $\bigvee_{i \in \emptyset} a_i = 0$ and $\sum_{i \in \emptyset} a_i = 0$. We generally consider statements for $i \in \emptyset$ as invalid.

2. The recursive max-linear model and its tail dependence matrix

In this section for a RMLM \mathbf{X} on a DAG \mathcal{D} , we highlight some relations between its TDM χ , its standardized MLCM \overline{B} , and the DAG \mathcal{D} . They prove particularly useful when we identify the RMLMs that are χ -equivalent to \mathbf{X} in Section 4.4 or investigate DAGs of χ -equivalent RMLMs in Section 5.

2.1. The tail dependence coefficients and max-weighted paths

We start with lower and upper bounds for the TDC between two components of \mathbf{X} such that in \mathcal{D} the two corresponding nodes are connected by a path. We also show that max-weighted paths lead to simple expressions for the TDCs and to nice relationships between them. It is precisely these properties that motivate us to consider RMWMs in detail later on.

Lemma 2.1. *Let $i \in V$ and $j \in \text{an}(i)$.*

- (a) *We have $0 < \frac{\bar{b}_{ji}}{\bar{b}_{jj}} \leq \chi(j, i)$ with equality iff there is a max-weighted path from every $k \in \text{An}(j)$ to i passing through j . In that case, $\chi(j, i) = \sum_{k \in \text{An}(j)} \bar{b}_{ki}$.*
- (b) *We have $\chi(j, i) \leq \sum_{k \in \text{An}(j)} \bar{b}_{ki} < 1$.*
- (c) *Let $k \in \text{de}(j) \cap \text{an}(i)$. If there is a max-weighted path from every $\ell \in \text{An}(j)$ to k and from every $\ell \in \text{An}(j)$ to i passing through j as well as from every $\ell \in \text{An}(k)$ to i passing through k , then*

$$\chi(j, i) = \chi(j, k)\chi(k, i) < \chi(j, k) \wedge \chi(k, i). \quad (10)$$

Proof. As $\text{An}(j) \subseteq \text{An}(i)$, we have by (8), $\chi(j, i) = \sum_{k \in \text{An}(j)} \bar{b}_{ki} \wedge \bar{b}_{kj}$.

(a) For $k \in \text{An}(j)$, by Lemma A.1(d), (f), $\frac{\bar{b}_{kj}\bar{b}_{ji}}{\bar{b}_{jj}} \leq \bar{b}_{ki} \wedge \bar{b}_{kj}$ with equality iff there is a max-weighted path from k to i passing through j . With this, using also Lemma A.1(b), (a), we obtain $\chi(j, i) \geq \frac{\bar{b}_{ji}}{\bar{b}_{jj}} \sum_{k \in \text{An}(j)} \bar{b}_{kj} = \frac{\bar{b}_{ji}}{\bar{b}_{jj}} > 0$ with equality iff there is a max-weighted path from every $k \in \text{An}(j)$ to i passing through j . In that case Lemma A.1(d) yields $\chi(j, i) = \sum_{k \in \text{An}(j)} \frac{\bar{b}_{kj}\bar{b}_{ji}}{\bar{b}_{jj}} = \sum_{k \in \text{An}(j)} \bar{b}_{ki}$.

(b) As $\text{An}(j) \subsetneq \text{An}(i)$, by Lemma A.1(a), (b) we find $\chi(j, i) \leq \sum_{k \in \text{An}(j)} \bar{b}_{ki} < \sum_{k \in \text{An}(i)} \bar{b}_{ki} = 1$.

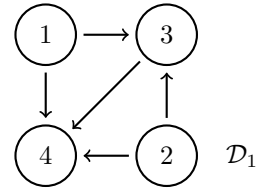
(c) The equality in (10) follows from (a) and Lemma A.1(d), the inequality then from the strict inequality in (b). \square

In the proof of Lemma 2.1 we have used that for $i \in V$, $k \in \text{an}(i)$, and $j \in \text{an}(k)$, \mathcal{D} has a max-weighted path from j to i passing through k iff $\bar{b}_{ji} = \frac{\bar{b}_{jk}\bar{b}_{ki}}{\bar{b}_{kk}}$ (Lemma A.1(d)). As to the equality in (10), one could expect that the MLCs can be replaced by the corresponding TDCs. The following example disproves this. In particular, it proves that the converse of Lemma 2.1(c) is not true in general and also that we may have the equality in (10) although $k \notin \text{de}(j) \cap \text{an}(i)$.

Example 2.2. $[\chi(j, i) = \chi(j, k)\chi(k, i)]$ is neither necessary nor sufficient for $\bar{b}_{ji} = \frac{\bar{b}_{jk}\bar{b}_{ki}}{\bar{b}_{kk}}$

- (1) Consider a RMLM on \mathcal{D}_1 with standardized MLCM

$$\bar{B} = \begin{pmatrix} 1 & 0 & 0.4 & 0.3 \\ 0 & 1 & 0.4 & 0.25 \\ 0 & 0 & 0.2 & 0.125 \\ 0 & 0 & 0 & 0.325 \end{pmatrix}.$$



As $\bar{b}_{24} = \frac{\bar{b}_{23}\bar{b}_{34}}{\bar{b}_{33}}$, the path $[2 \rightarrow 3 \rightarrow 4]$ is max-weighted. Computing χ we find $\chi(2, 4) < \chi(2, 3)\chi(3, 4)$.

That is, $\chi(2, 4) \neq \chi(2, 3)\chi(3, 4)$ although there is a max-weighted path from 2 to 4 passing through 3.

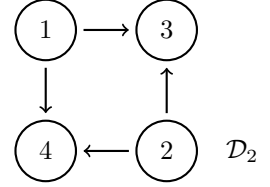
- (2) Now consider a RMLM on \mathcal{D}_1 with standardized MLCM

$$\bar{B} = \begin{pmatrix} 1 & 0 & 0.1 & 0.085 \\ 0 & 1 & 0.8 & 0.5 \\ 0 & 0 & 0.1 & 0.04 \\ 0 & 0 & 0 & 0.375 \end{pmatrix}.$$

The path $[2 \rightarrow 3 \rightarrow 4]$ is not max-weighted, since $\frac{\bar{b}_{23}\bar{b}_{34}}{\bar{b}_{24}} \neq \bar{b}_{24}$. However, we have $\chi(2,3)\chi(3,4) = \chi(2,4)$. In summary, $\chi(2,3)\chi(3,4) = \chi(2,4)$ although there is no max-weighted path from 2 to 4 passing through 3.

(3) Finally, consider a RMLM on \mathcal{D}_2 with standardized MLCM

$$\bar{B} = \begin{pmatrix} 1 & 0 & 1/3 & 1/6 \\ 0 & 1 & 1/3 & 1/3 \\ 0 & 0 & 1/3 & 0 \\ 0 & 0 & 0 & 1/2 \end{pmatrix}.$$



Here we find $\chi(1,3)\chi(3,4) = \chi(1,4)$; but 3 is not an ancestor of 4. According to this the equality in (10) may hold although $k \notin \text{de}(j) \cap \text{an}(i)$. \square

2.2. The tail dependence coefficients and the initial nodes

In this section we mainly investigate how χ and \mathcal{D} relate to each other.

Two components of \mathbf{X} are independent iff the TDC between them is zero. We link these two properties with the relationship between the two corresponding nodes in \mathcal{D} .

Theorem 2.3. *Let \mathbf{X} be a RMLM on a DAG $\mathcal{D} = (V, E)$ with TDM χ and $i, j \in V$. Then the following statements are equivalent:*

- (a) X_i and X_j are independent.
- (b) $\text{An}(i) \cap \text{An}(j) = \emptyset$.
- (c) $\chi(i, j) = 0$.

Proof. The equivalence between (a) and (b) follows from representation (3) for X_i and X_j and the distributional properties of the noise variables. The one between (b) and (c) is immediate by (8) and Lemma A.1(a). \square

Remark 2.4. (i) Let R be the reachability matrix of \mathcal{D} . The ij -th (ji -th) entry of $R^T R$ equals the cardinality of $\text{An}(i) \cap \text{An}(j)$. Thus by Theorem 2.3, $\text{sgn}(\chi) = \text{sgn}(R^T R)$. That is, we learn from $\chi(i, j) > 0$ only that $\text{An}(i) \cap \text{An}(j) \neq \emptyset$ but not whether i and j are connected by a path as is the case for the (standardized) MLCs (Lemma A.1(a) and (5), respectively).

- (ii) In the more general framework of Remark 1.1, parts (a) and (b) of Theorem 2.3 would have to be replaced by
 - (a') X_i and X_j are asymptotically independent; i.e., the corresponding components of the limit vector in (2) are independent.
 - (b') The dependence structure in the limit is given by a DAG, in which $\text{An}(i) \cap \text{An}(j) = \emptyset$.
The equivalence between (a') and (c) is a well-known result in extreme value theory; see e.g. Theorem 6.2.3 and the subsequent remark in de Haan and Ferreira (2006). \square

In what follows we investigate the relationship between χ and the initial nodes V_0 of \mathcal{D} . This is motivated by the fact that a RMLM is recursively defined by the structure of \mathcal{D} . For example, to obtain representation (3) of \mathbf{X} from its representation (1) recursively, we would start with representation (3) of the components X_i with $i \in V_0$. Then by proceeding iteratively we would substitute the parental variables in (1) by their representation (3). Such an iterative procedure starting with the initial nodes could also identify all RMLMs which have (the given) TDM χ .

The TDC between two components of \mathbf{X} simplifies considerably when in \mathcal{D} one of the corresponding nodes is an initial node. If both nodes are initial nodes, then the TDC between them is zero. We provide these and further related results.

Lemma 2.5. (a) For distinct $i, j \in V_0$, $\chi(i, j) = 0$.
(b) Let $W \subseteq V$ such that $\chi(i, j) = 0$ for all distinct $i, j \in W$. Then $|W| \leq |V_0|$.

- (c) For $i \in V$ and $j \in V_0$, $\text{An}(i) \cap V_0 = \{k \in V_0 : \chi(k, i) > 0\}$ and $\text{De}(j) = \{k \in V : \chi(j, k) > 0\}$.
(d) For $i \in V$ and $j \in V_0$, $\chi(j, i) = \bar{b}_{ji}$.

Proof. (a) and (c) follow from the fact that initial nodes have no ancestors and [Theorem 2.3](#).

(b) Assume that $|W| > |V_0|$. Since for every $i \in V$ there is some $j \in \text{An}(i) \cap V_0$, we have $j \in \text{An}(i_1) \cap \text{An}(i_2)$ for some $j \in V_0$ and distinct $i_1, i_2 \in W$. As $\text{An}(i_1) \cap \text{An}(i_2) \neq \emptyset$, again by [Theorem 2.3](#), $\chi(i_1, i_2) \neq 0$. This is, however, a contradiction to the fact that $\chi(i_1, i_2) = 0$ as $i_1, i_2 \in W$. Hence, $|W| \leq |V_0|$.

(d) As $\text{An}(j) = \{j\}$, we obtain from (8) by [Lemma A.1\(a\)](#), (f), $\chi(j, i) = \sum_{k=1}^d \bar{b}_{ki} \wedge \bar{b}_{kj} = \bar{b}_{ji} \wedge \bar{b}_{jj} = \bar{b}_{ji}$. \square

From [Lemma 2.5\(a\)](#), (b) we learn that V_0 is one of the node sets of maximum cardinality such that for every two distinct nodes, the TDC between their corresponding components of \mathbf{X} is zero. We introduce a concept which allows us to determine these sets from χ by a graph. For an illustration of these notions, we refer to [Example 4.12](#) below.

Definition 2.6. Let χ be the TDM of a RMLM on a DAG \mathcal{D} .

(a) We call the undirected graph that has nodes V and an edge between k and i iff $\chi(k, i) > 0$, χ -graph. Let \mathcal{D}^χ be the complement of the χ -graph (for the definition of the complement of an (undirected) graph, see e.g. [Diestel, 2010](#), Chapter 1.1) and $W \subseteq V$.

(b) We call W a χ -clique if it is a clique in \mathcal{D}^χ (for the definition of a clique in a graph, see e.g. [Koller and Friedman, 2009](#), Definition 2.13).

(c) We call W a *maximum χ -clique* if it is a maximum clique of \mathcal{D}^χ ; i.e., W is a clique in \mathcal{D}^χ such that no clique in \mathcal{D}^χ with higher cardinality exists. \square

The χ -graph associated with the TDM χ of \mathbf{X} corresponds to the covariance graph of the random vector \mathbf{X} introduced in [Cox and Wermuth \(1993\)](#), in which two (distinct) nodes are connected by an edge iff their corresponding components are dependent (cf. [Theorem 2.3](#)). In the non-Gaussian case, however, the name covariance graph is misleading.

The following theorem is an immediate consequence of [Definition 2.6](#) and [Lemma 2.5\(a\)](#), (b).

Theorem 2.7. *Let \mathbf{X} be a RMLM on a DAG \mathcal{D} with TDM χ . Then the set V_0 is a maximum χ -clique.*

[Theorem 2.7](#) raises the question of how V_0 is related to possible other maximum χ -cliques.

Lemma 2.8. *Let W be a maximum χ -clique.*

(a) *There is only one bijection $\varphi : V_0 \rightarrow W$ such that for every $j \in V_0$, $\chi(j, \varphi(j)) > 0$ and $\chi(j, i) = 0$ for all $i \in W \setminus \{\varphi(j)\}$.*

(b) *Let φ be the bijection from (a). Then for $j \in V_0$, $\text{An}(\varphi(j)) \cap V_0 = \{j\}$ and $\text{De}(j) \cap W = \{\varphi(j)\}$. In particular, if $j \neq \varphi(j)$, then \mathcal{D} has a path from j to $\varphi(j)$.*

(c) *Let $i, j \in V \setminus W$. If $\chi(i, j) < \sum_{k \in W} \chi(k, i) \wedge \chi(k, j)$, then $V_0 \neq W$.*

Proof. (a) Since maximum χ -cliques have the same cardinality, we know from [Theorem 2.7](#) that $|V_0| = |W|$. As for every $i \in W$, $\text{An}(i) \cap V_0 \neq \emptyset$, it suffices by [Lemma 2.5\(c\)](#) to show that $|\text{De}(j) \cap W| = 1$ for $j \in V_0$. We first assume that $|\text{De}(j) \cap W| > 1$. Using [Theorem 2.3](#) similarly as in the proof of [Lemma 2.5\(b\)](#) yields a contradiction. Hence, $|\text{De}(j) \cap W| \leq 1$. As $|V_0| = |W|$, $|\text{De}(j) \cap W| = 1$ must hold.

(b) follows from (a) and [Lemma 2.5\(c\)](#).

(c) Assume that $V_0 = W$. Using [Lemma A.1\(a\)](#) and [Lemma 2.5\(d\)](#) we obtain from (8)

$$\chi(i, j) = \sum_{k=1}^d \bar{b}_{ki} \wedge \bar{b}_{kj} \geq \sum_{k \in W} \chi(k, i) \wedge \chi(k, j).$$

Since this contradicts the conditions of (c), V_0 and W must be different. \square

3. The recursive max-weighted model and its tail dependence matrix

In this section we focus on RMWMs, i.e., RMLMs where all paths are max-weighted. We first present some structural properties of a RMWM \mathbf{X} on a DAG \mathcal{D} with standardized MLCM \bar{B} . We then investigate

its TDM χ and show that the assumption of all paths in \mathcal{D} being max-weighted involves simple relations between the TDCs and the (standardized) MLCs. Finally, we give necessary and sufficient conditions on a matrix to be the TDM of a RMWM on a given DAG.

3.1. Some structural properties of a recursive max-weighted model

All RMLMs on polytrees are RMWMs simply because in a polytree there is at most one path between every two (distinct) nodes (see also [Gissibl and Klüppelberg, 2017](#), Example 3.2). Furthermore, a RMWM can be constructed on every DAG, as the following example shows. Note the particularly simple structure of the RMLM introduced by it.

Example 3.1. [The homogeneous model]

Let $\mathcal{D} = (V, E)$ be a DAG with $V = \{1, \dots, d\}$ and Z_1, \dots, Z_d as in (1). Consider the RMLM defined by

$$X_i := \frac{1}{|\text{An}(i)|^{1/\alpha}} \left(\bigvee_{k \in \text{pa}(i)} |\text{An}(k)|^{1/\alpha} X_k \vee Z_i \right), \quad i = 1, \dots, d.$$

We find that every path p from j to i has the same weight $d_{ji}(p) = |\text{An}(i)|^{-1/\alpha}$. As a consequence, every path is max-weighted and \mathbf{X} is a RMWM. Its representation (3) is given by

$$X_i = \frac{1}{|\text{An}(i)|^{1/\alpha}} \bigvee_{j \in \text{An}(i)} Z_j, \quad i = 1, \dots, d.$$

For the TDC from (8) between X_i and X_j , we have

$$\chi(i, j) = \sum_{k \in \text{An}(i) \cap \text{An}(j)} \frac{1}{|\text{An}(i)|} \wedge \frac{1}{|\text{An}(j)|} = \frac{|\text{An}(i) \cap \text{An}(j)|}{|\text{An}(i) \vee \text{An}(j)|}.$$

If $j \in \text{an}(i)$, then this reduces to $\chi(i, j) = |\text{An}(j)|/|\text{An}(i)|$. Finally, by [Proposition A.2](#) the components of the limit vector \mathbf{M} introduced in (2) are standard α -Fréchet distributed. \square

Recall from the Introduction the prominent role of the minimum ML DAG \mathcal{D}^B of \mathbf{X} , which equals the minimum ML DAG $\mathcal{D}^{\bar{B}}$ of a RMLM with MLCM \bar{B} ([Lemma A.1\(e\)](#)). The fact that \mathbf{X} is max-weighted ensures that $\mathcal{D}^{\bar{B}}$ only depends on $\text{sgn}(\bar{B})$ but not on the precise values of the standardized MLCs. Since $\text{sgn}(\bar{B})$ is the reachability matrix of \mathcal{D} ((5) and [Lemma A.1\(a\)](#)), $\mathcal{D}^{\bar{B}}$ can be determined from pure graph theoretical properties. To clarify this we introduce a basic concept in graph theory, which goes back to [Aho et al. \(1972\)](#).

Definition 3.2. Let \mathcal{D} be a DAG.

- (a) An edge $k \rightarrow i$ is redundant if \mathcal{D} has another path from k to i .
- (b) The DAG \mathcal{D}^{tr} obtained from \mathcal{D} by deleting its redundant edges is called transitive reduction of \mathcal{D} . \square

Since $\mathcal{D}^{\bar{B}}$ has an edge $k \rightarrow i$ iff this is the only max-weighted path from k to i in \mathcal{D} , the fact that \mathcal{D} has only max-weighted paths yields part (i) of the following remark. By [Definition 3.2](#) and [Lemma A.1\(a\)](#) (ii) is a consequence of (i).

Remark 3.3. Let $\underline{\mathcal{D}}^{\text{tr}}$ be the transitive reduction of \mathcal{D} .

- (i) The DAGs $\mathcal{D}^{\bar{B}}$ and \mathcal{D}^{tr} coincide.
- (ii) $\mathcal{D}^{\bar{B}}$ is the DAG with the minimum number of edges that has reachability matrix $\text{sgn}(\bar{B})$.
- (iii) Even if \mathbf{X} is a RMLM but not max-weighted, it may happen that $\mathcal{D}^{\bar{B}} = \mathcal{D}^{\text{tr}}$ with all paths max-weighted in $\mathcal{D}^{\bar{B}}$. In that case all results presented in this section hold with respect to \mathcal{D}^{tr} . \square

3.2. Properties of the tail dependence coefficients of a recursive max-weighted model

The following result points out the simple structure of χ . It follows from [Lemma 2.1](#)(a), (c), since in \mathcal{D} all paths are max-weighted.

Lemma 3.4. *Let $i \in V$.*

- (a) For $j \in \text{An}(i)$, $\chi(j, i) = \frac{\bar{b}_{ji}}{\bar{b}_{jj}} = \sum_{k \in \text{An}(j)} \bar{b}_{ki} = \sum_{k \in \text{An}(j)} \bar{b}_{kk} \chi(k, i)$.
- (b) For $k \in \text{an}(i)$ and $j \in \text{an}(k)$, $\chi(j, i) = \chi(j, k) \chi(k, i) < \chi(j, k) \wedge \chi(k, i)$.
- (c) For $j \in \text{an}(i)$ and some path $[j = k_0 \rightarrow k_1 \rightarrow \dots \rightarrow k_n = i]$, $\chi(j, i) = \prod_{\nu=0}^{n-1} \chi(k_\nu, k_{\nu+1})$.

The equality $\chi(j, i) = \chi(j, k) \chi(k, i)$ for some $j \in \text{An}(i) \cap \text{An}(k)$ does not necessarily imply that $k \in \text{An}(i)$ (cf. part (3) of [Example 2.2](#)). For RMWMs, however, whenever these products hold for all $j \in \text{An}(i) \cap \text{An}(k) \cap V_0$, where V_0 are again the initial nodes in \mathcal{D} , we can conclude that $k \in \text{An}(i)$.

Proposition 3.5. *For $i, k \in V$, $k \in \text{An}(i)$ iff $\chi(j, i) = \chi(j, k) \chi(k, i)$ for all $j \in \text{An}(i) \cap \text{An}(k) \cap V_0$.*

Proof. Assume that $\chi(j, i) = \chi(j, k) \chi(k, i)$ for all $j \in \text{An}(i) \cap \text{An}(k) \cap V_0$. We first show that $\chi(\ell, i) \leq \chi(\ell, k)$ for every $\ell \in \text{An}(i) \cap \text{An}(k)$. We obtain for $j \in \text{An}(\ell) \cap V_0$, using the assumptions and [Lemma 3.4](#)(b),

$$\chi(k, i) = \frac{\chi(j, i)}{\chi(j, k)} = \frac{\chi(j, \ell) \chi(\ell, i)}{\chi(j, \ell) \chi(\ell, k)} = \frac{\chi(\ell, i)}{\chi(\ell, k)}.$$

Hence, $\chi(\ell, i) = \chi(\ell, k) \chi(k, i)$ and $\chi(\ell, i) \leq \chi(\ell, k)$. Together with [Lemma 3.4](#)(a) we then find from (8)

$$\chi(k, i) = \sum_{\ell \in \text{An}(k) \cap \text{An}(i)} \bar{b}_{\ell\ell} (\chi(\ell, k) \wedge \chi(\ell, i)) = \sum_{\ell \in \text{An}(k) \cap \text{An}(i)} \bar{b}_{\ell\ell} \chi(\ell, i) = \chi(k, i) \sum_{\ell \in \text{An}(k) \cap \text{An}(i)} \bar{b}_{\ell\ell} \chi(\ell, k).$$

By the assumptions and [Theorem 2.3](#) $\chi(k, i) > 0$ so that $\sum_{\ell \in \text{An}(k) \cap \text{An}(i)} \bar{b}_{\ell\ell} \chi(\ell, k) = 1$. As $1 = \sum_{\ell \in \text{An}(k)} \bar{b}_{\ell\ell} \chi(\ell, k)$ ([Lemma 3.4](#)(a)) and $\bar{b}_{\ell\ell} \chi(\ell, k) > 0$ for all $\ell \in \text{An}(k)$ ([Lemma A.1](#)(a) and [Theorem 2.3](#)), we have $\text{An}(i) \cap \text{An}(k) = \text{An}(k)$. This finally implies that $\text{An}(k) \subseteq \text{An}(i)$, equivalently $k \in \text{An}(i)$.

The converse statement holds due to [Lemma 3.4](#)(b). \square

In [Lemma 3.4](#)(a) we have written the positive standardized MLCs as functions of themselves and TDCs. We now present expressions for them only in terms of TDCs.

Proposition 3.6. *For $i \in V$ and $j \in \text{An}(i)$,*

$$\bar{b}_{ji} = \chi(j, i) - \sum_{k \in \text{an}(j)} \lambda_{jk} \chi(k, i) \quad \text{with} \quad \lambda_{jk} = 1 - \sum_{\ell \in \text{de}(k) \cap \text{an}(j)} \lambda_{j\ell}. \quad (11)$$

Proof. As by [Lemma 3.4](#)(a) $\bar{b}_{ji} = \chi(j, i) - \sum_{k \in \text{an}(j)} \bar{b}_{ki}$, it suffices to show that $\sum_{k \in \text{an}(j)} \lambda_{jk} \chi(k, i) = \sum_{k \in \text{an}(j)} \bar{b}_{ki}$. Using again [Lemma 3.4](#)(a) yields

$$\sum_{k \in \text{an}(j)} \lambda_{jk} \chi(k, i) = \sum_{k \in \text{an}(j)} \lambda_{jk} \sum_{\ell \in \text{An}(k)} \bar{b}_{\ell i}.$$

Noting that $k \in \text{an}(j)$ and $\ell \in \text{An}(k)$ iff $\ell \in \text{an}(j)$ and $k \in \text{De}(\ell) \cap \text{an}(j)$, we can interchange the two summation operators to obtain

$$\sum_{k \in \text{an}(j)} \lambda_{jk} \sum_{\ell \in \text{An}(k)} \bar{b}_{\ell i} = \sum_{\ell \in \text{an}(j)} \bar{b}_{\ell i} \sum_{k \in \text{De}(\ell) \cap \text{an}(j)} \lambda_{jk} = \sum_{\ell \in \text{an}(j)} \bar{b}_{\ell i} (\lambda_{j\ell} + \sum_{k \in \text{de}(\ell) \cap \text{an}(j)} \lambda_{jk}) = \sum_{\ell \in \text{an}(j)} \bar{b}_{\ell i},$$

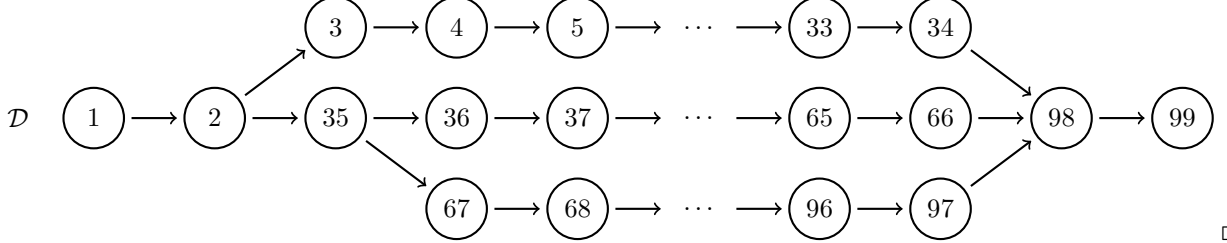
where we have used the definition of $\lambda_{j\ell}$ for the last equality. \square

Before we give an example of representation (11), we summarize some characteristics of the coefficients λ_{jk} . Denoting by $\text{pa}^{\text{tr}}(j)$ the parents of j in the transitive reduction \mathcal{D}^{tr} of \mathcal{D} , we have $\lambda_{jk} = 1$ for $k \in \text{pa}^{\text{tr}}(j)$ as $\text{de}(k) \cap \text{an}(j) = \emptyset$. For $k \in \text{an}(j) \setminus \text{pa}^{\text{tr}}(j)$ it can be verified that $\lambda_{jk} \neq 0$ iff there exists no $\tilde{k} \in \text{de}(k) \cap \text{an}(j)$ such that $|\text{De}(\tilde{k}) \cap \text{pa}^{\text{tr}}(j)| = |\text{De}(k) \cap \text{pa}^{\text{tr}}(j)|$.

Example 3.7. [On representation (11)]

Consider a RMWM \mathbf{X} on the DAG \mathcal{D} depicted below, and note that here $\mathcal{D} = \mathcal{D}^{\text{tr}}$. We determine, as an example, representation (11) for the MLCs $\bar{b}_{36,66}$ and $\bar{b}_{98,99}$:

$$\bar{b}_{36,66} = \chi(36, 66) - \chi(35, 66), \quad \bar{b}_{98,99} = \chi(98, 99) - \chi(34, 99) - \chi(66, 99) - \chi(97, 99) + \chi(2, 99) + \chi(35, 99).$$



□

We address again the interrelations between the TDCs and prove that every TDC can be written as linear combination of minima of two TDCs.

Proposition 3.8. For $i, j \in V$,

$$\chi(i, j) = \sum_{k \in \text{An}(i) \cap \text{An}(j)} \mu_{ij,k} (\chi(k, i) \wedge \chi(k, j)) \quad \text{with} \quad \mu_{ij,k} = 1 - \sum_{\ell \in \text{de}(k) \cap \text{An}(i) \cap \text{An}(j)} \mu_{ij,\ell}. \quad (12)$$

Proof. Applying Lemma 3.4(a) and Lemma A.1(b), (d) we obtain for $k \in \text{An}(i) \cap \text{An}(j)$,

$$\chi(k, i) \wedge \chi(k, j) = \frac{\bar{b}_{ki}}{\bar{b}_{kk}} \wedge \frac{\bar{b}_{kj}}{\bar{b}_{kk}} = \left(\frac{\bar{b}_{ki}}{\bar{b}_{kk}} \wedge \frac{\bar{b}_{kj}}{\bar{b}_{kk}} \right) \left(\sum_{\ell \in \text{An}(k)} \bar{b}_{\ell k} \right) = \sum_{\ell \in \text{An}(k)} \frac{\bar{b}_{\ell k} \bar{b}_{ki}}{\bar{b}_{kk}} \wedge \frac{\bar{b}_{\ell k} \bar{b}_{kj}}{\bar{b}_{kk}} = \sum_{\ell \in \text{An}(k)} \bar{b}_{\ell i} \wedge \bar{b}_{\ell j}.$$

With this we then have

$$\sum_{k \in \text{An}(i) \cap \text{An}(j)} \mu_{ij,k} (\chi(k, i) \wedge \chi(k, j)) = \sum_{k \in \text{An}(i) \cap \text{An}(j)} \mu_{ij,k} \sum_{\ell \in \text{An}(k)} \bar{b}_{\ell i} \wedge \bar{b}_{\ell j}.$$

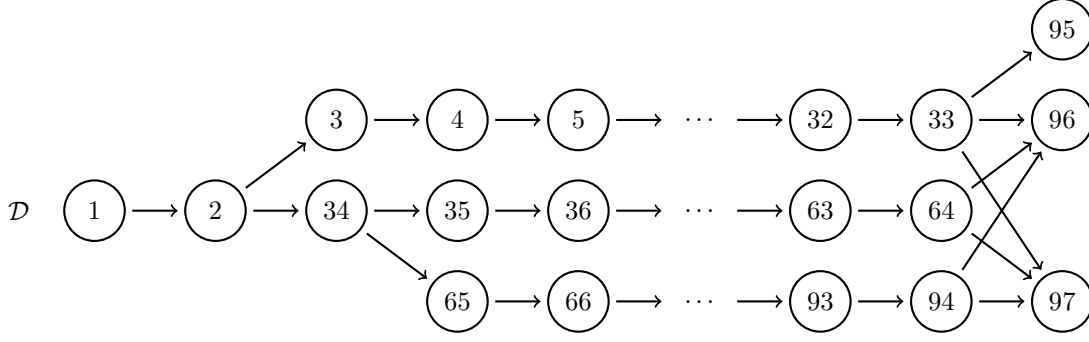
Using that $k \in \text{An}(i) \cap \text{An}(j)$ and $\ell \in \text{An}(k)$ iff $\ell \in \text{An}(i) \cap \text{An}(j)$ and $k \in \text{De}(\ell) \cap \text{An}(i) \cap \text{An}(j)$ to interchange the summation operators similarly as in the proof of Proposition 3.6 and the definition of $\mu_{ij,\ell}$ similarly as the one of $\lambda_{j\ell}$ there, we finally find (12). □

For $i, j \in V$ denote by $\text{lca}(i, j)$ the lowest common ancestors of i and j ; i.e., $k \in \text{lca}(i, j)$ iff $k \in \text{An}(i) \cap \text{An}(j)$ and \mathcal{D} has no path from k to another node in $\text{An}(i) \cap \text{An}(j)$. For $\mu_{ij,k}$ from (12) we have $\mu_{ij,k} = 1$ for $k \in \text{lca}(i, j)$ as in that case $\text{de}(k) \cap \text{An}(i) \cap \text{An}(j) = \emptyset$. It can be verified that $\mu_{ij,k} = 0$ for $k \in (\text{An}(i) \cap \text{An}(j)) \setminus \text{lca}(i, j)$ iff there exists some $\tilde{k} \in \text{de}(k) \cap \text{An}(i) \cap \text{An}(j)$ such that $|\text{De}(\tilde{k}) \cap \text{lca}(i, j)| = |\text{De}(k) \cap \text{lca}(i, j)|$. With this, if $j \in \text{An}(i)$, then $\mu_{ij,j} = 1$ and $\mu_{ij,k} = 0$ for $k \in \text{an}(j)$. Thus in that case the right-hand side of the first equality in (12) is equal to $\chi(j, i) \wedge \chi(j, j) = \chi(j, i)$, and representation (12) is trivial. Note the analogy of the coefficients $\mu_{ij,k}$ to the coefficients λ_{jk} in (11).

Example 3.9. [On representation (12)]

Consider a RMWM on the DAG \mathcal{D} depicted below. We present, as an example, representation (12) for the TDCs $\chi(95, 96)$ and $\chi(96, 97)$:

$$\begin{aligned} \chi(95, 96) &= \chi(33, 95) \wedge \chi(33, 96), \\ \chi(96, 97) &= \chi(33, 96) \wedge \chi(33, 97) + \chi(64, 96) \wedge \chi(64, 97) + \chi(94, 96) \wedge \chi(94, 97) \\ &\quad - \chi(34, 96) \wedge \chi(34, 97) - \chi(2, 96) \wedge \chi(2, 97). \end{aligned}$$



We conclude this section with necessary and sufficient conditions on a matrix to be the TDM of a RMWM on a given DAG \mathcal{D} . To be such a matrix, the ij -th (ji -th) entry of the matrix must satisfy a property depending on the relationship between i and j in \mathcal{D} . For example, based on [Theorem 2.3](#), it must be zero iff $\text{An}(i) \cap \text{An}(j) = \emptyset$. By [Lemma A.1\(e\)](#), [Remark 3.3\(i\)](#), and [Theorem 5.4 of Gissibl and Klüppelberg \(2017\)](#), a RMWM on \mathcal{D} is a RMWM on every DAG that has reachability matrix R of \mathcal{D} . Consequently, it would be sufficient to specify R and to require the four conditions below for any DAG with reachability matrix R such as the transitive reduction \mathcal{D}^{tr} of \mathcal{D} .

Theorem 3.10. *Let $\mathcal{D} = (V, E)$ be a DAG with nodes $V = \{1, \dots, d\}$ and reachability matrix R . Let $\chi = (\chi(i, j))_{d \times d}$ be a symmetric matrix with ones on the diagonal. For $i \in V$ define $\bar{b}_{ii} := 1 - \sum_{k \in \text{an}(i)} \bar{b}_{kk} \chi(k, i)$ recursively. Then χ is the TDM of a RMWM \mathbf{X} on \mathcal{D} iff the following conditions hold:*

- (a) $\text{sgn}(\chi) = \text{sgn}(R^T R)$.
- (b) For all $i \in V$, $\bar{b}_{ii} > 0$.
- (c) For all $i \in V$, $j \in \text{an}(i)$, and $k \in \text{de}(j) \cap \text{pa}(i)$, $\chi(j, i) = \chi(j, k) \chi(k, i)$.
- (d) For all $i, j \in V$ such that $i \notin \text{An}(j)$ and $j \notin \text{An}(i)$ but $\text{An}(i) \cap \text{An}(j) \neq \emptyset$,

$$\chi(i, j) = \sum_{k \in \text{An}(i) \cap \text{An}(j)} \bar{b}_{kk} (\chi(k, i) \wedge \chi(k, j)).$$

In that case \bar{b}_{ii} is the i -th diagonal entry of the standardized MLCM \bar{B} of \mathbf{X} . Furthermore, for $i, j \in V$, $\bar{b}_{ji} = 0$ if $j \in V \setminus \text{An}(i)$, and $\bar{b}_{ji} = \bar{b}_{jj} \chi(j, i)$ if $j \in \text{an}(i)$.

Proof. Assume that χ is the TDM of a RMWM \mathbf{X} on \mathcal{D} . The statements (a) and (c) follow from [Remark 2.4\(i\)](#) and [Lemma 3.4\(b\)](#). By [Lemma 3.4\(a\)](#) \bar{b}_{ii} is the i -th diagonal entry of the standardized MLCM \bar{B} of \mathbf{X} . Since all \bar{b}_{ii} are positive according to [Lemma A.1\(a\)](#), assertion (b) holds. The representation of $\chi(i, j)$ in (d) is again a consequence of [Lemma 3.4\(a\)](#).

Assume now that (a)-(d) hold. For every $i \in V$ define $\bar{b}_{ji} := \bar{b}_{jj} \chi(j, i)$ for all $j \in \text{an}(i)$ and for all $j \in V \setminus \text{An}(i)$, $\bar{b}_{ji} := 0$. We first show that $\bar{B} = (\bar{b}_{ij})_{d \times d}$ is the MLCM of a RMWM on \mathcal{D} , where weights from its representation (1) are given by $c_{ii} := \bar{b}_{ii}$ and $c_{ki} := \frac{\bar{b}_{ki}}{\bar{b}_{kk}} = \chi(k, i)$ for $i \in V$ and $k \in \text{pa}(i)$. As $\text{sgn}(\chi) = \text{sgn}(R^T R)$ and $\bar{b}_{ii} > 0$, the weights c_{ki} for $i \in V$ and $k \in \text{Pa}(i)$ are positive, which is a necessary condition for them by the definition of a RMLM in (1). Let $p = [j = k_0 \rightarrow k_1 \rightarrow \dots \rightarrow k_n = i]$ be a path in \mathcal{D} . Using (c) iteratively yields

$$d_{ji}(p) = c_{jj} \prod_{\nu=0}^{n-1} c_{k_\nu, k_{\nu+1}} = \bar{b}_{jj} \prod_{\nu=0}^{n-1} \chi(k_\nu, k_{\nu+1}) = \bar{b}_{jj} \chi(j, k_2) \prod_{\nu=2}^{n-1} \chi(k_\nu, k_{\nu+1}) = \dots = \bar{b}_{jj} \chi(i, j) = \bar{b}_{ji}.$$

This implies that $\bar{B} = (\bar{b}_{ij})_{d \times d}$ is the MLCM of a RMWM \mathbf{X} . Since it suffices to specify one RMLM that has TDM χ , we may assume that $Z \in \text{RV}(1)$. Denoting the TDM of \mathbf{X} by $\bar{\chi} = (\bar{\chi}(i, j))_{d \times d}$, it remains to show that $\bar{\chi} = \chi$. Since the diagonal entries of χ equal one, the equality of the diagonal entries is obvious. For $i, j \in V$ such that $\text{An}(i) \cap \text{An}(j) = \emptyset$, the ij -th (ji -th) entries of χ and $\bar{\chi}$ are zero and, hence, equal due to condition (a) and [Theorem 2.3](#). The matrix \bar{B} is the standardized MLCM of \mathbf{X} as $\alpha = 1$ and $\bar{b}_{ii} = 1 - \sum_{k \in \text{an}(i)} \bar{b}_{ki}$ for every $i \in V$. Thus for $i \in V$ and $j \in \text{an}(i)$ we have by [Lemma 3.4\(a\)](#) and the definition

of \bar{B} that $\bar{\chi}(j, i) = \frac{\bar{b}_{ji}}{\bar{b}_{jj}} = \chi(i, j)$. Finally, for $i, j \in V$ such that $j \notin \text{An}(i)$ and $i \notin \text{An}(j)$ but $\text{An}(i) \cap \text{An}(j) \neq \emptyset$, using [Lemma 3.4\(a\)](#), the result shown before, and condition (d), we obtain

$$\bar{\chi}(i, j) = \sum_{k \in \text{An}(i) \cap \text{An}(j)} \bar{b}_{kk} (\bar{\chi}(k, i) \wedge \bar{\chi}(k, j)) = \sum_{k \in \text{An}(i) \cap \text{An}(j)} \bar{b}_{kk} (\chi(k, i) \wedge \chi(k, j)) = \chi(i, j).$$

□

In [Example 5.5](#) below we present a possible application of [Theorem 3.10](#).

Remark 3.11. In [Theorem 3.10](#) the coefficients \bar{b}_{ii} can also be defined by $1 - \sum_{k \in \text{an}(i)} \lambda_{ik} \chi(k, i)$ with λ_{ik} as in [\(11\)](#). We give a sketch of a proof of this assertion: we show that $\lambda_{ik} = 1 - \sum_{\ell \in \text{de}(k) \cap \text{an}(i)} \lambda_{\ell k}$ and use this to verify that if (c) holds, then the assertion is valid as well. Moreover, condition (d) can be replaced by

(d') For all $i, j \in V$ such that $i \notin \text{An}(j)$ and $j \notin \text{An}(i)$ but $\text{An}(i) \cap \text{An}(j) \neq \emptyset$,

$$\chi(i, j) = \sum_{k \in \text{An}(i) \cap \text{An}(j)} \mu_{ij,k} (\chi(k, i) \wedge \chi(k, j)) \quad \text{with } \mu_{ij,k} \text{ as in (12).}$$

By going through the proof of [Theorem 3.10](#), we observe that this can be done due to the representation of $\chi(i, j)$ in [\(12\)](#). □

4. Identifiability problems based on the tail dependence matrix of a recursive max-linear model

Throughout this section we assume that the TDM $\underline{\chi}$ of a RMLM \mathbf{X} on a DAG \mathcal{D} with standardized MLCM \bar{B} is given. We first show the identifiability of \bar{B} from χ and the reachability matrix R of \mathcal{D} . We then assume that the reachability relation of \mathcal{D} is not fully known but only a causal ordering σ . This still leads to identifiability of \bar{B} from χ . We also investigate whether \bar{B} can be recovered from χ and the initial nodes V_0 of \mathcal{D} . It turns out that this is generally not possible, but we verify it for RMWMs. We prove the different identifiability results by providing algorithms which compute \bar{B} from χ and the additionally known information on \mathcal{D} . Finally, based on these results we present an approach, which finds the standardized MLCMs of all RMLMs with TDM χ . Since this method simplifies for RMWMs considerably, we give an adapted and modified version for this subclass of RMLMs.

4.1. Identifiability from the tail dependence matrix and the reachability matrix

The following algorithm computes \bar{B} from χ and R recursively. The rows of \bar{B} are filled up successively until \bar{B} is obtained, where the number of ancestors determines the order in which the rows are treated. The existence of such an algorithm proves the identifiability of \bar{B} from χ and R .

Algorithm 4.1. [Find \bar{B} from χ and R]

For $\nu = 0, \dots, d-1$,

for $j \in V$ such that $|\text{an}(j)| = \nu$, set

$$\bar{b}_{ji} = 0 \text{ for all } i \in V \setminus \text{De}(j) \quad \text{and} \quad \bar{b}_{ji} = \chi(j, i) - \sum_{k \in \text{an}(j)} \bar{b}_{ki} \wedge \bar{b}_{kj} \text{ for all } i \in \text{De}(j). \quad (13)$$

Eq. [\(13\)](#) follows from [Lemma A.1\(a\)](#), [\(8\)](#), and [Lemma A.1\(f\)](#). If \mathbf{X} is max-weighted, then by [Lemma 3.4\(a\)](#) [\(13\)](#) can be replaced by

$$\bar{b}_{ji} = 0 \text{ for all } i \in V \setminus \text{De}(j), \quad \bar{b}_{jj} = 1 - \sum_{k \in \text{an}(j)} \bar{b}_{kj}, \quad \text{and} \quad \bar{b}_{ji} = \bar{b}_{jj} \chi(j, i) \text{ for all } i \in \text{de}(j). \quad (14)$$

To avoid the iterative loop, we can also use [\(11\)](#) for computing the diagonal entries of \bar{B} . Note, however, that this requires to calculate the coefficients λ_{jk} appearing in [\(11\)](#) recursively as well.

4.2. Identifiability from the tail dependence matrix and a causal ordering

So far we have dealt with the identifiability from χ and the reachability matrix R of \mathcal{D} . Here we investigate the identifiability from χ and a causal ordering σ of \mathcal{D} . If R is given, then we know for every two (distinct) $i, j \in V$ whether there is a path from j to i ; but from σ we only learn that there is no path from j to i if $\sigma(j) > \sigma(i)$.

There exists a causal ordering for every DAG due to the acyclicity (see also [Diestel, 2010](#), Appendix A). However, it is not necessarily unique. For example, the DAG \mathcal{D}_1 from [Example 2.2](#) has the identity function on $V = \{1, 2, 3, 4\}$ and the permutation $\tilde{\sigma}$ on V given by $\tilde{\sigma}(2) = 1, \tilde{\sigma}(1) = 2, \tilde{\sigma}(3) = 3, \tilde{\sigma}(4) = 4$ as causal orderings.

The DAG \mathcal{D} has a causal ordering which can be completely described by its initial nodes V_0 and χ as follows.

Lemma 4.2. *We denote the initial nodes by $V_0 = \{i_1, \dots, i_{|V_0|}\}$ and define $V_0^i := \{k \in V_0 : \chi(k, i) > 0\}$ for $i \in V$. Then \mathcal{D} has a causal ordering σ such that*

$$\sigma(i_\nu) = \nu \text{ for } \nu = 1, \dots, |V_0| \text{ and for all } i, j \in V, \sigma(j) < \sigma(i) \text{ whenever } |V_0^j| < |V_0^i|. \quad (15)$$

Proof. Recall from [Lemma 2.5\(c\)](#) that $V_0^j = \text{An}(j) \cap V_0$ and $V_0^i = \text{An}(i) \cap V_0$. With this it is not difficult to see that \mathcal{D} has such a causal ordering. \square

Now we give an iterative procedure which computes \overline{B} from χ and σ . Obviously, this proves the identifiability of \overline{B} from χ and σ . Here the rows of \overline{B} are also filled up successively, where the order of the nodes given by σ defines the order in which the rows are treated.

Algorithm 4.3. [Find \overline{B} from χ and σ]

For $\nu = 1, \dots, d$,

for $j \in V$ such that $\sigma(j) = \nu$, set

$$\begin{aligned} \overline{b}_{ji} &= 0 \text{ for all } i \in V \text{ such that } \sigma(j) > \sigma(i), \\ \overline{b}_{ji} &= \chi(j, i) - \sum_{k: \sigma(k) < \sigma(j)} \overline{b}_{ki} \wedge \overline{b}_{kj} \text{ for all } i \in V \text{ such that } \sigma(j) \leq \sigma(i). \end{aligned} \quad (16)$$

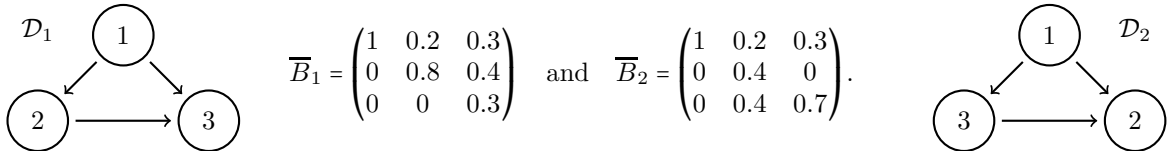
Eq. (16) can be obtained from (8) by using [Lemma A.1\(a\)](#), the definition of a causal ordering, and [Lemma A.1\(f\)](#).

4.3. Identifiability of recursive max-weighted models from the tail dependence matrix and the initial nodes

In what follows we assume \mathbf{X} to be max-weighted. Then recalling [Lemma 2.5\(c\)](#), [Proposition 3.5](#) involves a procedure to determine R from χ and V_0 . Since [Algorithm 4.1](#) computes \overline{B} from χ and R , we can identify \overline{B} from χ and V_0 . This is usually not possible outside the class of RMWMs.

Example 4.4. [\overline{B} is generally not identifiable from χ and V_0]

Consider two RMLMs on \mathcal{D}_1 and \mathcal{D}_2 with standardized MLCMs \overline{B}_1 and \overline{B}_2 given by



We find by [Lemma A.1\(d\)](#) that none of the two models is max-weighted. Since both have the same χ and \mathcal{D}_1 and \mathcal{D}_2 share the same initial node $V_0 = \{1\}$, we cannot distinguish between \overline{B}_1 and \overline{B}_2 based on χ and V_0 . \square

Proceeding as suggested by [Proposition 3.5](#) to recover R from χ and V_0 is very tedious, since many conditions may need to be verified. Therefore, we introduce an alternative method which computes \overline{B} from χ and V_0 : we first determine a causal ordering σ of \mathcal{D} and apply then [Algorithm 4.3](#) to obtain \overline{B} . From the next proposition we learn how a causal ordering σ of \mathcal{D} can be computed from χ and V_0 ; note that we encountered property (i) in (15).

Proposition 4.5. *Let V_0^i for $i \in V$ be as in [Lemma 4.2](#). Every permutation σ on V such that for all $i, j \in V$,*

- (i) $\sigma(j) < \sigma(i)$ whenever $|V_0^j| < |V_0^i|$ and
- (ii) $\sigma(j) < \sigma(i)$ whenever $|V_0^j| = |V_0^i|$ and $\max_{k \in V_0^i} \chi(k, i) < \max_{k \in V_0^j} \chi(k, j)$

is a causal ordering of \mathcal{D} .

Proof. Assume that σ is no causal ordering of \mathcal{D} , i.e., $\sigma(j) > \sigma(i)$ for some $i \in V$ and $j \in \text{an}(i)$. Recall from [Lemma 2.5\(c\)](#) that $V_0^j = \text{An}(j) \cap V_0$ and $V_0^i = \text{An}(i) \cap V_0$. As $j \in \text{an}(i)$, $V_0^j \subseteq V_0^i$. But then because of the properties of σ , $V_0^j = V_0^i$ and $\max_{k \in V_0^j} \chi(k, j) \leq \max_{k \in V_0^j} \chi(k, i)$. Assume now that $j \in V_0^j$, and note that $i \notin V_0^j$ as $j \in \text{an}(i)$. Then, since for $i_1, i_2 \in V$ the TDC $\chi(i_1, i_2) = 1$ iff $i_1 = i_2$ (cf. (8) and [Lemma A.1\(a\)](#)), we find $1 = \max_{k \in V_0^j} \chi(k, j) \leq \max_{k \in V_0^j} \chi(k, i) < 1$. This contradiction proves that $j \notin V_0^j$, which implies again that $V_0^j = \text{an}(j) \cap V_0$. As $\max_{k \in V_0^j} \chi(k, j) \leq \max_{k \in V_0^j} \chi(k, i)$, $\chi(k, j) \leq \chi(k, i)$ for some $k \in \text{an}(j) \cap V_0$. Observe from [Lemma 3.4\(b\)](#) that $j \notin \text{an}(i)$, since otherwise $\chi(k, i) < \chi(k, j)$. This, however, contradicts our original assumption, and σ must be a causal ordering of \mathcal{D} . \square

Finally, we clarify the precise steps of our approach to determine \overline{B} from χ and V_0 .

Algorithm 4.6. [Modification of [Algorithm 4.3](#) for RMWMs: find \overline{B} from χ and V_0]

1. Find a causal ordering σ of \mathcal{D} from χ and V_0 :

for $\nu = 1, \dots, |V_0|$,

find all $j \in V$ such that $|V_0^j| = |\{k \in V_0 : \chi(k, j) > 0\}| = \nu$ and summarize them in the set A_ν ;

sort the nodes $k_1, \dots, k_{|A_\nu|}$ from A_ν such that

$$\max_{\ell \in V_0} \chi(\ell, k_1) \geq \max_{\ell \in V_0} \chi(\ell, k_2) \geq \dots \geq \max_{\ell \in V_0} \chi(\ell, k_{|A_\nu|});$$

for $\mu = 1, \dots, |A_\nu|$,

set $\sigma(k_\mu) = \sum_{\lambda=1}^{\nu-1} |A_\lambda| + \mu$, where $\sum_{\lambda=1}^0 := 0$.

2. Apply [Algorithm 4.3](#) to obtain \overline{B} from χ and σ .

Observe from [Proposition 4.5](#) that every permutation σ on V which can be chosen in step 1. is indeed a causal ordering of \mathcal{D} .

4.4. Identifiability from the tail dependence matrix

We now combine the previous results to find the standardized MLCMs of all RMLMs that have TDM χ . In the first part we deal with general RMLMs. Because of the identifiability properties derived in [Section 4.3](#), we assume in the second part that χ is the TDM of a RMWM. We provide an algorithm, which outputs the standardized MLCMs of all RMWMs that have TDM χ .

(General) recursive max-linear models

Every permutation $\tilde{\sigma}$ on $V = \{1, \dots, d\}$ is a causal ordering of a DAG with nodes V but not necessarily of a DAG that corresponds to a RMLM with TDM χ . But if this is the case, then applying [Algorithm 4.3](#) with $\sigma = \tilde{\sigma}$ yields the corresponding standardized MLCM \overline{B} . This suggests the following procedure to prove the existence of a RMLM which has TDM χ and whose associated DAG has causal ordering $\tilde{\sigma}$: first apply [Algorithm 4.3](#) with $\sigma = \tilde{\sigma}$, and check then whether the obtained matrix \overline{B} is the standardized MLCM of a RMLM which has TDM χ and whose associated DAG has causal ordering $\tilde{\sigma}$. In the second step it is enough to verify that \overline{B} is the MLCM of a RMLM, which can be done by [Theorem 5.7 of Gissibl and Klüppelberg \(2017\)](#).

Lemma 4.7. *Let $\tilde{\sigma}$ be a permutation on V and \overline{B} the matrix obtained by applying [Algorithm 4.3](#) with $\sigma = \tilde{\sigma}$. If \overline{B} is the MLCM of a RMLM (RMWM), then \overline{B} is the standardized MLCM of a RMLM (RMWM) which has TDM χ and whose associated DAG has causal ordering $\tilde{\sigma}$.*

Proof. Let \mathbf{X} be the RMLM (RMWM) with MLCM \overline{B} and $Z \in RV(1)$. Its existence is guaranteed as \overline{B} is the MLCM of a RMLM (RMWM). We show that \mathbf{X} has standardized MLCM \overline{B} and TDM χ as well as that its associated DAG \mathcal{D} has causal ordering $\tilde{\sigma}$. Recall from [\(5\)](#) that $\text{sgn}(\overline{B})$ is the reachability matrix of \mathcal{D} . Thus by [\(16\)](#) $\tilde{\sigma}$ is a causal ordering of \mathcal{D} and $\overline{b}_{ii} = 1 - \sum_{k \in \text{an}(i)} \overline{b}_{ki}$ for every $i \in V$. As the latter holds and $\alpha = 1$, \overline{B} is the standardized MLCM of \mathbf{X} . The fact that \mathbf{X} has TDM χ also follows from [\(16\)](#). \square

Lemma 4.7 suggests a “naive” method to find the standardized MLCMs of all RMLMs that have TDM χ : for every permutation on V compute the matrix \overline{B} from [Algorithm 4.3](#), and check whether it is the MLCM of a RMLM; if so, then \overline{B} is the standardized MLCM of a RMLM with TDM χ . However, the number of permutations on V to be investigated can often be significantly reduced. By [Theorem 2.7](#) and [Lemma 2.8\(c\)](#) the set of all maximum χ -cliques W (see [Definition 2.6](#)) such that $\chi(i, j) \geq \sum_{k \in W} \chi(k, i) \wedge \chi(k, j)$ for all $i, j \in V \setminus W$ contains the initial node sets of all DAGs underlying RMLMs with TDM χ . So it suffices to investigate the causal orderings of DAGs that have such initial nodes W . But also the number of causal orderings to be investigated for every such set W can be reduced further by [Lemma 4.2](#): it is enough to consider those permutations on V , which satisfy the properties σ has in [\(15\)](#) with $V_0 = W$. The following algorithm describes the precise steps of an approach to find the standardized MLCMs of all RMLMs with TDM χ .

Algorithm 4.8. [Find all \overline{B} from χ]

1. Find all maximum χ -cliques:
 - (a) find the complement \mathcal{D}^x of the χ -graph;
 - (b) find all maximum cliques of \mathcal{D}^x .
2. For every maximum χ -clique $W = \{i_1, \dots, i_{|W|}\}$,
 - (a) check $\chi(i, j) \geq \sum_{k \in W} \chi(k, i) \wedge \chi(k, j)$ for all $i, j \in V \setminus W$;
if not, then there is no RMLM with TDM χ on a DAG with initial nodes W ;
else,
 - (b) for every permutation $\tilde{\sigma}$ on $V = \{1, \dots, d\}$ such that
$$\tilde{\sigma}(i_\nu) = \nu \text{ for } \nu = 1, \dots, |W| \text{ and}$$

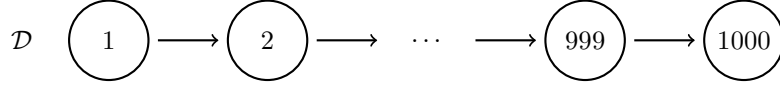
$$\tilde{\sigma}(j) < \tilde{\sigma}(i) \text{ whenever } |\{k \in W : \chi(k, j) > 0\}| < |\{k \in W : \chi(k, i) > 0\}|,$$
 - i. apply [Algorithm 4.3](#) with $\sigma = \tilde{\sigma}$;
 - ii. check whether \overline{B} obtained in i. is the MLCM of a RMLM; for instance using [Theorem 5.7](#) of [Gissibl and Klüppelberg \(2017\)](#);
if not, then there is no RMLM with TDM χ on a DAG with causal ordering $\tilde{\sigma}$;
else, \overline{B} is the standardized MLCM of a RMLM with TDM χ .

When the algorithm returns a standardized MLCM \overline{B} of a RMLM with TDM χ in step ii., then it is not necessary to perform steps i., ii. for further permutations on V which are causal orderings of DAGs with reachability matrix $\text{sgn}(\overline{B})$, since all of them would lead to the same \overline{B} . For the application of [Algorithm 4.8](#), we have assumed so far that χ is the TDM of a RMLM. If this is not the case, [Algorithm 4.8](#) would not produce any output. The same applies to [Algorithm 4.11](#) below if χ is not the TDM of a RMWM.

One could drop step 2.(a) and perform step 2.(b) for all maximum χ -cliques. However, the performance of step 2.(a) can be very effective.

Example 4.9. [Not all maximum χ -cliques are initial node sets]

Consider the TDM χ of a RMLM on the DAG \mathcal{D} depicted below. Note that such a RMLM is max-weighted, since \mathcal{D} is a polytree (cf. [Section 3.1](#)). [Theorem 2.3](#) yields that the sets $\{1\}, \dots, \{1000\}$ are the maximum χ -cliques. For every $k \in \{2, \dots, 999\}$ we know from [Lemma 3.4\(b\)](#) that $\chi(1, 1000) < \chi(1, k) \wedge \chi(k, 1000)$. The property tested in step 2.(a) is therefore not fulfilled for the maximum χ -cliques $W \in \{\{2\}, \dots, \{999\}\}$. However, we can verify by [Lemma 3.4\(b\)](#) that it is fulfilled for $W \in \{\{1\}, \{1000\}\}$. Consequently, step 2.(b) needs only be performed for $W \in \{\{1\}, \{1000\}\}$ and not for the other 998 maximum χ -cliques.



□

It is indeed necessary to perform step ii., i.e., to verify that a matrix \overline{B} obtained in i. is a MLCM of a RMLM.

Example 4.10. [Not every \overline{B} obtained in ii. belongs to a RMLM]
Consider the TDM

$$\chi = \begin{pmatrix} 1 & 1/10 & 1/3 \\ 1/10 & 1 & 13/30 \\ 1/3 & 13/30 & 1 \end{pmatrix}.$$

Performing steps i. and ii. of [Algorithm 4.8](#) with $\tilde{\sigma}$ being the identity function on $V = \{1, 2, 3\}$ and also with $\tilde{\sigma}$ given by $\tilde{\sigma}(1) = 1$, $\tilde{\sigma}(3) = 2$, $\tilde{\sigma}(2) = 3$ (note that these permutations are really tested in step 2.(b)), we find

$$\overline{B}_1 = \begin{pmatrix} 1 & 1/10 & 1/3 \\ 0 & 9/10 & 1/3 \\ 0 & 0 & 1/3 \end{pmatrix} \quad \text{and} \quad \overline{B}_2 = \begin{pmatrix} 1 & 1/10 & 1/3 \\ 0 & 17/30 & 0 \\ 0 & 1/3 & 2/3 \end{pmatrix}.$$

As can be verified by Theorem 4.2 of [Gissibl and Klüppelberg \(2017\)](#), the matrix \overline{B}_1 is the MLCM of a RMLM on the DAG \mathcal{D}_1 depicted in [Example 4.4](#). Although $\text{sgn}(\overline{B}_2)$ is the reachability matrix of a DAG, namely of the DAG \mathcal{D}_2 from [Example 4.4](#), which is a necessary property of a matrix to be the MLCM of a RMLM according to (5), it is no MLCM of a RMLM. □

Recursive max-weighted models

Assume now that χ is the TDM of a RMWM. We modify and adapt [Algorithm 4.8](#) to obtain a procedure which outputs the standardized MLCMs of all RMWMs with TDM χ . Among the maximum χ -cliques which we find in step 2.(a) of [Algorithm 4.8](#) are the initial node sets of the DAGs underlying the RMWMs that have TDM χ . We learn from [Proposition 4.5](#) and [Lemma 4.7](#) that a maximum χ -clique is such an initial node set iff the matrix \overline{B} obtained by [Algorithm 4.6](#) is the MLCM of a RMWM. In that case, \overline{B} is obviously the standardized MLCM of a RMWM with TDM χ . These observations lead to the following procedure.

Algorithm 4.11. [Modification of [Algorithm 4.8](#) for RMWMs: find all \overline{B} from χ]

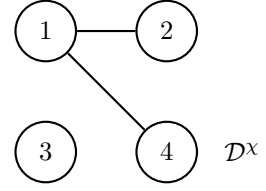
1. Find all maximum χ -cliques (cf. step 1. of [Algorithm 4.8](#)).
 2. For every maximum χ -clique W ,
 - (a) check $\chi(i, j) \geq \sum_{k \in W} \chi(k, i) \wedge \chi(k, j)$ for all $i, j \in V \setminus W$;
if not, then there is no RMWM with TDM χ on a DAG with initial nodes W ;
else,
 - i. apply [Algorithm 4.6](#) with $V_0 = W$;
 - ii. check the following properties for the matrix \overline{B} obtained in i.:
 - $\text{sgn}(\overline{B})$ is the reachability matrix of a DAG
 - for all $i \in V$, $j \in \text{an}(i)$, and $k \in \text{de}(j) \cap \text{pa}(i)$, $\overline{b}_{ji} = \frac{\overline{b}_{jk}\overline{b}_{ki}}{\overline{b}_{kk}}$
- if not, then there is no RMWM with TDM χ on a DAG with initial nodes W ;
else, \overline{B} is the standardized MLCM of a RMWM with TDM χ .

That the properties we verify for the matrix \overline{B} in step ii. are sufficient for \overline{B} to be the MLCM of a RMWM can be verified by Corollary 4.3(a) of [Gissibl and Klüppelberg \(2017\)](#).

To conclude this section, we highlight the essential steps of [Algorithm 4.11](#) with an example.

Example 4.12. [The class of RMWMs is not closed under χ -equivalence]
Consider the TDM

$$\chi = \begin{pmatrix} 1 & 0 & 0.2 & 0 \\ 0 & 1 & 0.6 & 0.5 \\ 0.2 & 0.6 & 1 & 0.5 \\ 0 & 0.5 & 0.5 & 1 \end{pmatrix}.$$



We read from the complement \mathcal{D}^χ of the χ -graph that the sets $W_1 = \{1, 2\}$ and $W_2 = \{1, 4\}$ are the maximum χ -cliques. Applying [Algorithm 4.6](#) with $V_0 = W_1$ and $V_0 = W_2$, we get the matrices

$$\mathcal{D}_1 \quad \begin{array}{c} \textcircled{1} \quad \textcircled{2} \\ \downarrow \quad \swarrow \\ \textcircled{3} \quad \textcircled{4} \end{array} \quad \bar{B}_1 = \begin{pmatrix} 1 & 0 & 0.2 & 0 \\ 0 & 1 & 0.6 & 0.5 \\ 0 & 0 & 0.2 & 0 \\ 0 & 0 & 0 & 0.5 \end{pmatrix} \quad \text{and} \quad \bar{B}_2 = \begin{pmatrix} 1 & 0 & 0.2 & 0 \\ 0 & 0.5 & 0.1 & 0 \\ 0 & 0 & 0.2 & 0 \\ 0 & 0.5 & 0.5 & 1 \end{pmatrix} \quad \begin{array}{c} \textcircled{1} \quad \textcircled{4} \\ \downarrow \quad \swarrow \\ \textcircled{3} \quad \textcircled{2} \end{array} \quad \mathcal{D}_2$$

The matrix \bar{B}_1 is the MLCM of a RMWM on \mathcal{D}_1 , whereas \bar{B}_2 is not the MLCM of a RMWM, but it is the MLCM of a RMLM on \mathcal{D}_2 . Therefore, all RMWMs with TDM χ have the same standardized MLCM \bar{B}_1 , and \mathcal{D}_1 is their associated DAG. Furthermore, all these models are χ -equivalent to the RMLMs with standardized MLCM \bar{B}_2 . \square

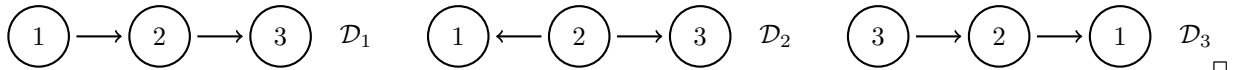
5. χ -equivalent recursive max-linear models and their DAGs

In this section we mainly present interrelations between DAGs of χ -equivalent RMLMs.

One of the best known equivalence relations on the set of DAGs is certainly the Markov equivalence: two DAGs are Markov equivalent if they entail the same conditional independence relations through the Markov property; for a characterization of such DAGs, see e.g. [Verma and Pearl \(1990\)](#). The associated DAG of a recursive linear Gaussian structural equation model can be identified from the distribution only up to a Markov equivalence class (under the assumption of faithfulness; see e.g. [Spirtes and Zhang \(2016\)](#)). In the following example we discuss the relation between χ -equivalence of RMLMs and Markov equivalence of their associated DAGs.

Example 5.1. [The difference between χ -equivalence of RMLMs and Markov equivalence of their DAGs]

- (1) Undirected graphs underlying Markov equivalent DAGs coincide. [Example 4.12](#) clarifies that this does not hold for DAGs of χ -equivalent RMLMs. Such DAGs are therefore not necessarily Markov equivalent.
- (2) For the TDCs of a RMLM \mathbf{X} on \mathcal{D}_1 , which is always a RMWM, we have by [Lemma 3.4\(b\)](#) that $\chi(1, 3) < \chi(1, 2) \wedge \chi(2, 3)$. Since \mathcal{D}_2 has initial node 2, by [Lemma 2.8\(c\)](#) there cannot be a RMLM that is χ -equivalent to \mathbf{X} on \mathcal{D}_2 . Thus although the DAGs \mathcal{D}_1 and \mathcal{D}_2 are Markov equivalent, there exist no χ -equivalent RMLMs on \mathcal{D}_1 and \mathcal{D}_2 .
- (3) As can be verified by [Theorem 3.10](#), RMLMs on the Markov equivalent DAGs \mathcal{D}_1 and \mathcal{D}_3 are always χ -equivalent. This shows that there can be χ -equivalent RMLMs on Markov equivalent DAGs.



DAGs of χ -equivalent RMLMs have the same number of initial nodes, since the initial node sets of such DAGs are maximum χ -cliques, which have the same cardinality by definition. We learn from [Algorithm 4.3](#) that if the standardized MLCMs of two χ -equivalent RMLMs differ, then the causal orderings of their associated DAGs must also differ. So for these two DAGs there exist nodes $i, j \in V$ such that one DAG has a path from j to i and the other has one from i to j . We provide further properties of two DAGs underlying χ -equivalent RMLMs.

Proposition 5.2. Let \mathbf{X} and $\tilde{\mathbf{X}}$ be χ -equivalent RMLMs on DAGs \mathcal{D} and $\tilde{\mathcal{D}}$, respectively. We denote the initial nodes in \mathcal{D} and $\tilde{\mathcal{D}}$ by V_0 and \tilde{V}_0 , the ancestors of i by $\text{an}(i)$ and $\tilde{\text{an}}(i)$, and the descendants of i by $\text{de}(i)$ and $\tilde{\text{de}}(i)$.

(a) There is only one bijection $\varphi: V_0 \rightarrow \tilde{V}_0$ such that for every $j \in V_0$, $\chi(j, \varphi(j)) > 0$ and $\chi(j, \tilde{j}) = 0$ for all $\tilde{j} \in \tilde{V}_0 \setminus \{\varphi(j)\}$.

Let φ be the bijection from (a) and $j \in V_0$.

(b) We have $\text{An}(\varphi(j)) \cap V_0 = \tilde{\text{De}}(\varphi(j)) \cap V_0 = \{j\}$. In particular, if $j \neq \varphi(j)$, then \mathcal{D} has a path from j to $\varphi(j)$, and $\tilde{\mathcal{D}}$ has one from $\varphi(j)$ to j .

(c) We have $\text{De}(j) = \tilde{\text{De}}(\varphi(j))$.

(d) For $i \in V$, $\tilde{\text{An}}(i) \cap \tilde{V}_0 = \{\varphi(j) : j \in \text{An}(i) \cap V_0\}$.

Proof. (a) is immediate by Lemma 2.8(a), since \tilde{V}_0 is a maximum χ -clique.

(b) Since \tilde{V}_0 is a maximum χ -clique, according to Lemma 2.8(b), $\text{An}(\varphi(j)) \cap V_0 = \{j\}$. Note that for every $\tilde{j} \in \tilde{V}_0$, $\chi(\tilde{j}, \varphi^{-1}(\tilde{j})) > 0$ and $\chi(\tilde{j}, j) > 0$ for all $j \in V_0 \setminus \{\varphi^{-1}(\tilde{j})\}$, where $\varphi^{-1}: \tilde{V}_0 \rightarrow V_0$ denotes the inverse of φ . As V_0 is a maximum χ -clique, we therefore have again by Lemma 2.8(b) that $\tilde{\text{De}}(i) \cap V_0 = \{\varphi^{-1}(i)\}$ with $i = \varphi(j)$, which is obviously equivalent to $\tilde{\text{De}}(\varphi(j)) \cap V_0 = \{j\}$.

(c) Let $i \in \text{De}(j)$. By (b) $j \in \text{An}(\varphi(j)) \cap \text{An}(i)$ and, consequently, by Theorem 2.3 $\chi(\varphi(j), i) > 0$. Lemma 2.5(c) then yields that $i \in \tilde{\text{De}}(\varphi(j))$. Hence, $\text{De}(j) \subseteq \tilde{\text{De}}(\varphi(j))$. From this, by reversing the roles of \mathcal{D} and $\tilde{\mathcal{D}}$ and noting that $\chi(\tilde{j}, \varphi^{-1}(\tilde{j})) > 0$ for all $\tilde{j} \in \tilde{V}_0$, we observe that $\tilde{\text{De}}(\varphi(j)) \subseteq \text{De}(j)$.

(d) can be verified by (c). \square

Recursive max-weighted models

Now we consider χ -equivalent RMWMs and investigate their DAGs. Because of Theorem 2.7, Algorithm 4.6, and Lemma A.1(e), if a TDM χ of a RMWM has one maximum χ -clique W , all RMWMs with TDM χ (the models are then χ -equivalent by definition) have the same standardized MLCM and, hence, the same minimum ML DAG, which again has initial nodes W . By Algorithm 4.6 the initial nodes of DAGs of χ -equivalent RMLMs with different standardized MLCMs must also differ. We present further interrelationships between DAGs of χ -equivalent RMWMs with regard to their initial nodes.

Theorem 5.3. Let \mathbf{X} and $\tilde{\mathbf{X}}$ be χ -equivalent RMWMs on DAGs \mathcal{D} and $\tilde{\mathcal{D}}$, respectively. We denote by V_0 and \tilde{V}_0 the initial nodes in \mathcal{D} and $\tilde{\mathcal{D}}$ and by V_∞ and \tilde{V}_∞ their terminal nodes. Let $\varphi: V_0 \rightarrow \tilde{V}_0$ be the bijection from Proposition 5.2(a) and $j \in V_0$ such that $j \neq \varphi(j)$.

(a) We have $\varphi(j) \in V_\infty$. In particular, $\tilde{V}_0 \subseteq (V_0 \cap \tilde{V}_0) \cup V_\infty$.

(b) If $p = [j = k_0 \rightarrow k_1 \rightarrow \dots \rightarrow k_{n-1} \rightarrow k_n = \varphi(j)]$ is a path in the transitive reduction \mathcal{D}^{tr} of \mathcal{D} , then $\tilde{p} = [\varphi(j) = k_n \rightarrow k_{n-1} \rightarrow \dots \rightarrow k_1 \rightarrow k_0 = j]$ is a path in the transitive reduction $\tilde{\mathcal{D}}^{\text{tr}}$ of $\tilde{\mathcal{D}}$.

Proof. We denote by $\text{an}(i)$ and $\tilde{\text{an}}(i)$ the ancestors of i in \mathcal{D} and $\tilde{\mathcal{D}}$ and by $\text{de}(i)$ and $\tilde{\text{de}}(i)$ its descendants.

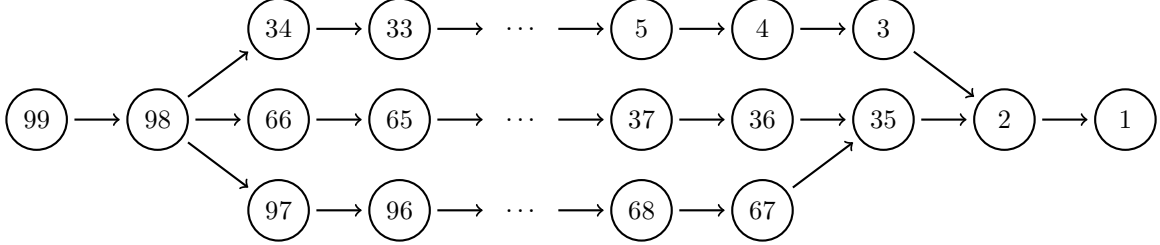
(a) Assume that $\varphi(j) \notin V_\infty$. Consequently, by Proposition 5.2(b) \mathcal{D} has a path from j to some $i \neq \varphi(j)$ passing through $\varphi(j)$. Replacing V_0 by \tilde{V}_0 , we learn from the the proof of Lemma 2.8(c) that $\chi(j, i) \geq \chi(\varphi(j), j) \wedge \chi(\varphi(j), i)$. But this contradicts Lemma 3.4(b). Hence, $\varphi(j) \in V_\infty$.

(b) Let p be a path in \mathcal{D}^{tr} . To prove that \tilde{p} is a path in $\tilde{\mathcal{D}}^{\text{tr}}$, because of the properties of $\tilde{\mathcal{D}}^{\text{tr}}$, it suffices to show that for $\nu = 0, \dots, n-1$, $k_{\nu+1} \in \tilde{\text{an}}(k_\nu)$ and $\tilde{\text{de}}(k_{\nu+1}) \cap \tilde{\text{an}}(k_\nu) \neq \emptyset$. Recalling from Proposition 5.2(b) that $\text{An}(\varphi(j)) \cap V_0 = \{j\}$, we observe that $\text{An}(k_\nu) \cap \text{An}(k_{\nu+1}) \cap V_0 = \{j\}$. We then obtain from Proposition 5.2(d) that $\tilde{\text{An}}(k_\nu) \cap \tilde{\text{An}}(k_{\nu+1}) \cap \tilde{V}_0 = \{\varphi(j)\}$. By Lemma 3.4(b) we have $\chi(k_\nu, \varphi(j)) = \chi(k_\nu, k_{\nu+1})\chi(k_{\nu+1}, \varphi(j))$. As $\tilde{\text{An}}(k_\nu) \cap \tilde{\text{An}}(k_{\nu+1}) \cap \tilde{V}_0 = \{\varphi(j)\}$, using Proposition 3.5 then proves that $k_{\nu+1} \in \tilde{\text{an}}(k_\nu)$. To show that $\tilde{\text{de}}(k_{\nu+1}) \cap \tilde{\text{an}}(k_\nu) \neq \emptyset$, assume the converse. Let $\ell \in \tilde{\text{de}}(k_{\nu+1}) \cap \tilde{\text{an}}(k_\nu)$. By reversing the roles of \mathcal{D}^{tr} and $\tilde{\mathcal{D}}^{\text{tr}}$ and noting that for every $\tilde{j} \in \tilde{V}_0$, $\chi(\tilde{j}, \varphi^{-1}(\tilde{j})) > 0$ and $\chi(\tilde{j}, j) > 0$ for all $j \in V_0 \setminus \{\varphi^{-1}(\tilde{j})\}$, where $\varphi^{-1}: \tilde{V}_0 \rightarrow V_0$ denotes the inverse of φ , we know from above that then $k_\nu \in \text{an}(\ell)$ and $\ell \in \text{an}(k_{\nu+1})$, i.e., $\text{de}(k_\nu) \cap \text{an}(k_{\nu+1}) \neq \emptyset$. But this is in contradiction to the fact that p is a path in \mathcal{D}^{tr} . Hence, $\tilde{\mathcal{D}}^{\text{tr}}$ must contain \tilde{p} . \square

In the next example we use Theorem 5.3 to find RMWMs that are χ -equivalent to a given one.

Example 5.4. [Continuation of [Example 3.7](#): find χ -equivalent RMWMs]

By [Theorem 2.3](#) the sets $\{1\}, \dots, \{99\}$ are the maximum χ -cliques. Since 99 is the only terminal node in \mathcal{D} , it may be the only initial node of a DAG that underlies a potential RMWM with the same TDM χ as \mathbf{X} and differs from \mathcal{D} . Thus the DAG



is the transitive reduction $\tilde{\mathcal{D}}^{\text{tr}}$ of such a DAG. To verify the existence of a RMWM with TDM χ on a DAG whose transitive reduction is $\tilde{\mathcal{D}}^{\text{tr}}$, we may compute the matrix \bar{B} from [\(14\)](#) and check then whether it is the MLCM of a RMWM. \square

We conclude this section with an example investigating whether a RMWM on a known DAG is χ -equivalent to a RMWM on another given DAG.

Example 5.5. [The existence of χ -equivalent RMWMs on given DAGs]

We consider a RMWM \mathbf{X} with TDM χ on \mathcal{D}_1 and clarify when \mathbf{X} is χ -equivalent to a RMWM on \mathcal{D}_2 . Note that all RMLMs on \mathcal{D}_1 and on \mathcal{D}_2 are max-weighted. By [Theorem 3.10](#) we find

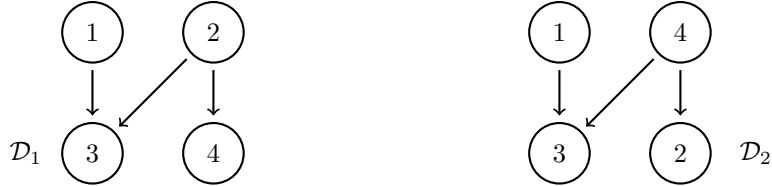
$$\chi(1,2) = 0, \chi(1,4) = 0, \chi(1,3) > 0, 1 - \chi(1,3) - \chi(2,3) > 0, 1 - \chi(2,4) > 0, \chi(3,4) = \chi(2,3) \wedge \chi(2,4) > 0$$

and also that χ is the TDM of a RMWM on \mathcal{D}_2 iff

$$\chi(1,2) = 0, \chi(1,4) = 0, \chi(1,3) > 0, 1 - \chi(1,3) - \chi(3,4) > 0, 1 - \chi(2,4) > 0, \chi(2,3) = \chi(2,4) \wedge \chi(3,4) > 0.$$

This implies that \mathbf{X} is χ -equivalent to a RMWM on \mathcal{D}_2 iff $\chi(2,3) = \chi(3,4)$.

As shown in [Example 4.12](#) the matrix χ given therein is the TDM of a RMWM on \mathcal{D}_1 . As $\chi(2,3) = 0.6 \neq \chi(3,4) = 0.5$, such a model cannot be χ -equivalent to a RMWM on \mathcal{D}_2 . Of course, we already know this from [Example 4.12](#).



\square

6. Conclusion

A RMLM is not restricted to heavy-tailed noise variables, but is defined in [Gissibl and Klüppelberg \(2017\)](#) for independent noise variables with support \mathbb{R}_+ . Only, if the noise variables are heavy-tailed, the TDM is meaningful (not identical to 0) for modeling the dependence structure in a RMLM.

In this heavy-tailed setting, we considered the problem of identifying a RMLM \mathbf{X} on a DAG \mathcal{D} from its TDM χ . Simply because of the symmetry of χ , the identifiability of \mathbf{X} is not possible in general. RMLMs with arbitrary index of regular variation and MLCM whose column sums are also arbitrary have TDM χ . As our focus was on the causal structure of \mathbf{X} represented by \mathcal{D} , we concentrated on the standardized model, where the index of regular variation is one and the columns of its MLCM \bar{B} add up to one. We showed that \bar{B} can be recovered from χ and some additional information on \mathcal{D} such as the full reachability relation or only a causal ordering. In these situations we can also determine the minimum ML DAG \mathcal{D}^B of \mathbf{X} , the smallest DAG which represents the recursive max-linear dependence structure of \mathbf{X} . We developed an

algorithm which outputs the standardized MLCMs of all RMLMs having TDM χ . Moreover, we found the RMWMs as a relevant subclass of RMLMs. The simple structure of their TDMS allows for identifiability of \overline{B} and \mathcal{D}^B from χ and the initial nodes of \mathcal{D} . This led to a simpler approach to find the standardized MLCMs of all RMWMs with TDM χ .

Future work will focus on statistical properties of RMLMs.

Acknowledgements

We thank Steffen Lauritzen and Jonas Peters for interesting discussions and Zhongwei Zhang for his careful reading of our manuscript. MO and NG thank the International Graduate School of Science and Engineering (IGSSE) of the Technical University of Munich for support.

References

References

- Aho, A.V., Garey, M.R., Ullman, J.D., 1972. The transitive reduction of a directed graph. *SIAM Journal on Computing* 1, 131–137.
- Beirlant, J., Goegebeur, Y., Segers, J., Teugels, J., 2004. *Statistics of Extremes: Theory and Applications*. Wiley, Chichester.
- Bollen, K.A., 1989. *Structural Equations with Latent Variables*. Wiley, New York.
- Cox, D.R., Wermuth, N., 1993. Linear dependencies represented by chain graphs. *Statistical Science* 8, 204–218.
- Cui, Q., Zhang, Z., 2017. Max-linear competing factor models. *Journal of Business & Economic Statistics* 0, 1–13.
- Diestel, R., 2010. *Graph Theory. Graduate Texts in Mathematics, Vol. 173*. 4th ed., Springer, Heidelberg.
- Eimmahl, J.H., Kiriliouk, A., Segers, J., 2017. A continuous updating weighted least squares estimator of tail dependence in high dimensions. *ArXiv:1601.04826*, *Extremes*, to appear.
- Eimmahl, J.H., Krajina, A., Segers, J., 2012. An M-estimator for tail dependence in arbitrary dimensions. *The Annals of Statistics* 40, 1764–1793.
- Ernest, J., Rothenhäusler, D., Bühlmann, P., 2016. Causal inference in partially linear structural equation models: identifiability and estimation. *ArXiv:1607.05980*.
- Falk, M., 2005. On the generation of a multivariate extreme value distribution with prescribed tail dependence parameter matrix. *Statistics & Probability Letters* 75, 307–314.
- Falk, M., Hofmann, M., Zott, M., 2015. On generalized max-linear models and their statistical interpolation. *Journal of Applied Probability* 52, 736–751.
- Gissibl, N., Klüppelberg, C., 2017. Max-linear models on directed acyclic graphs. *ArXiv:1512.07522*, *Bernoulli*, to appear.
- Gissibl, N., Klüppelberg, C., Lauritzen, S., 2017. Estimation, identifiability, and structure learning of recursive max-linear models. In preparation.
- de Haan, L., Ferreira, A., 2006. *Extreme Value Theory: An Introduction*. Springer, New York.
- Kiriliouk, A., 2017. Hypothesis testing for tail dependence parameters on the boundary of the parameter space with application to generalized max-linear models. *ArXiv:1708.07019*.
- Koller, D., Friedman, N., 2009. *Probabilistic Graphical Models: Principles and Techniques*. MIT Press, New York.
- Lauritzen, S.L., 1996. *Graphical Models*. Oxford University Press, Oxford.
- Pearl, J., 2009. *Causality: Models, Reasoning, and Inference*. 2nd ed., Cambridge University Press, New York.
- Peters, J., Mooij, J.M., Janzing, D., Schölkopf, B., 2014. Causal discovery with continuous additive noise models. *Journal of Machine Learning Research* 15, 2009–2053.
- Resnick, S.I., 1987. *Extreme Values, Regular Variation, and Point Processes*. Springer, New York.
- Resnick, S.I., 2007. *Heavy-Tail Phenomena: Probabilistic and Statistical Modeling*. Springer, New York.
- Schlather, M., Tawn, J., 2002. Inequalities for the extremal coefficients of multivariate extreme value distributions. *Extremes* 5, 87–102.
- Sibuya, M., 1960. Bivariate extreme statistics. *Annals of the Institute of Statistical Mathematics* 11, 195–210.
- Spirtes, P., Glymour, C., Scheines, R., 2000. *Causation, Prediction, and Search*. 2nd ed., MIT Press, New York.
- Spirtes, P., Zhang, K., 2016. Causal discovery and inference: concepts and recent methodological advances. *Applied Informatics* 3, 1–28.
- Strokorb, K., Schlather, M., 2015. An exceptional max-stable process fully parameterized by its extremal coefficients. *Bernoulli* 21, 276–302.
- Verma, T., Pearl, J., 1990. Equivalence and synthesis of causal models, in: *Proceedings of the 6th Conference on Uncertainty in Artificial Intelligence*, Cambridge, MA. pp. 220–227.
- Wang, Y., Stoev, S.A., 2011. Conditional sampling for spectrally discrete max-stable random fields. *Advances in Applied Probability* 43, 461–483.
- Yuen, R., Stoev, S.A., 2014. CRPS M-estimation for max-stable models. *Extremes* 17, 387–410.

A. Appendix

A.1. Properties of the standardized max-linear coefficient matrix of a recursive max-linear model

We summarize some properties of the standardized MLCM \bar{B} defined in (7), which are used throughout the paper.

Lemma A.1. *Let \mathbf{X} be a RMLM on a DAG \mathcal{D} with MLCM B and standardized MLCM \bar{B} .*

- (a) *We have $\text{sgn}(\bar{B}) = \text{sgn}(B)$.*
- (b) *For $i \in V$, $\sum_{k \in \text{An}(i)} \bar{b}_{ki} = \sum_{k=1}^d \bar{b}_{ki} = 1$.*
- (c) *The matrix \bar{B} is the MLCM of a RMLM on \mathcal{D} .*
- (d) *For $i \in V$, $k \in \text{an}(i)$, and $j \in \text{an}(k)$, $\bar{b}_{ji} \geq \frac{\bar{b}_{jk}\bar{b}_{ki}}{\bar{b}_{kk}}$ with equality iff there is a max-weighted path from j to i passing through k .*
- (e) *The minimum ML DAGs \mathcal{D}^B and $\mathcal{D}^{\bar{B}}$ coincide.*
- (f) *For distinct $i, j \in V$, $\bar{b}_{jj} > \bar{b}_{ji}$.*

Proof. (a) and (b) are immediate consequences of the definition of \bar{B} and (5).

(c) can be verified by Theorem 4.2 of Gissibl and Klüppelberg (2017).

(d) The inequality follows from (c) and Corollary 3.12 of Gissibl and Klüppelberg (2017) and the rest of the statement from Theorem 3.10(a) of Gissibl and Klüppelberg (2017) and by observing that $\bar{b}_{ji} = \frac{\bar{b}_{jk}\bar{b}_{ki}}{\bar{b}_{kk}}$ iff $b_{ji} = \frac{b_{jk}b_{ki}}{b_{kk}}$.

(e) is a consequence of Theorem 5.3 of Gissibl and Klüppelberg (2017) and the definition of \bar{B} .

(f) For $j \in V \setminus \text{An}(i)$ we have immediately by (a) that $\bar{b}_{ji} = 0 < \bar{b}_{jj}$. For $j \in \text{An}(i)$ we obtain by parts (b) and (d),

$$1 = \sum_{k \in \text{An}(j)} \bar{b}_{ki} + \sum_{k \in \text{An}(i) \setminus \text{An}(j)} \bar{b}_{ki} \geq \frac{\bar{b}_{ji}}{\bar{b}_{jj}} \sum_{k \in \text{An}(j)} \bar{b}_{kj} + \sum_{k \in \text{An}(i) \setminus \text{An}(j)} \bar{b}_{ki} = \frac{\bar{b}_{ji}}{\bar{b}_{jj}} + \sum_{k \in \text{An}(i) \setminus \text{An}(j)} \bar{b}_{ki}.$$

Since $\text{An}(i) \setminus \text{An}(j) \neq \emptyset$ and $\bar{b}_{ki} > 0$ for all $k \in \text{An}(i) \setminus \text{An}(j)$, we find $1 > \frac{\bar{b}_{ji}}{\bar{b}_{jj}}$, equivalently $\bar{b}_{jj} > \bar{b}_{ji}$. \square

A.2. Derivation of the tail dependence matrix of a recursive max-linear model

We first prove (2) and specify G and its univariate and bivariate marginal distributions.

Proposition A.2. *Let \mathbf{X} be a RMLM on a DAG \mathcal{D} with MLCM B . Then $\mathbf{X} \in \text{MDA}(G)$ with*

$$G(\mathbf{x}) = \exp \left\{ - \sum_{j=1}^d \bigvee_{i \in \text{De}(j)} \left(\frac{b_{ji}}{x_i} \right)^\alpha \right\}, \quad \mathbf{x} = (x_1, \dots, x_d) \in \mathbb{R}_+^d.$$

Let $\mathbf{M} = (M_1, \dots, M_d)$ be a random vector with distribution function G . Then for $i, j \in V$ the distribution functions of M_i and (M_i, M_j) are given by

$$G_i(x_i) = \exp \left\{ - x_i^{-\alpha} \sum_{j \in \text{An}(i)} b_{ji}^\alpha \right\} \quad \text{and} \quad G_{ij}(x_i, x_j) = \exp \left\{ - \sum_{k \in \text{An}(i) \cup \text{An}(j)} \left(\frac{b_{ki}}{x_i} \right)^\alpha \vee \left(\frac{b_{kj}}{x_j} \right)^\alpha \right\}.$$

Proof. As $Z \in \text{RV}(\alpha)$, there exists a normalizing sequence $a_n \in \mathbb{R}_+$ such that for every $x \in \mathbb{R}_+$,

$$\lim_{n \rightarrow \infty} F_Z^n(a_n x) = \Phi_\alpha(x) \tag{A.1}$$

(e.g. Resnick, 1987, Proposition 1.11). Using (3), the independence of the noise variables, and (A.1), we obtain for $\mathbf{x} \in \mathbb{R}_+^d$,

$$\begin{aligned} [\mathbb{P}(\mathbf{X} \leq a_n \mathbf{x})]^n &= [\mathbb{P}(\bigvee_{j \in \text{An}(i)} b_{ji} Z_j \leq a_n x_i, i \in V)]^n \\ &= [\mathbb{P}(Z_j \leq a_n \bigwedge_{i \in \text{De}(j)} \frac{x_i}{b_{ji}}, j \in V)]^n \\ &= \prod_{j=1}^d F_Z^n(a_n \bigwedge_{i \in \text{De}(j)} \frac{x_i}{b_{ji}}) \\ &\xrightarrow{n \rightarrow \infty} \prod_{j=1}^d \Phi_\alpha(\bigwedge_{i \in \text{De}(j)} \frac{x_i}{b_{ji}}) = G(\mathbf{x}). \end{aligned}$$

This proves that $\mathbf{X} \in \text{MDA}(G)$ (cf. Eq. (2)). Finally, the distribution functions of M_i and (M_i, M_j) are obtained by letting all other components of \mathbf{x} in G tend to ∞ and recalling (5). \square

Proof of (8). For every $k \in V$ we have $n(1 - F_k(a_{k,n})) \rightarrow 1$ as $n \rightarrow \infty$ with $a_{k,n} := F_k^\leftarrow(1 - \frac{1}{n}) = (\frac{1}{1 - F_k})^\leftarrow(n)$. Thus,

$$\begin{aligned} \chi(i, j) &= \lim_{n \rightarrow \infty} \frac{\mathbb{P}(X_i > a_{i,n}, X_j > a_{j,n})}{1 - F_j(a_{j,n})} \\ &= \lim_{n \rightarrow \infty} n[1 - F_i(a_{i,n}) + 1 - F_j(a_{j,n}) - 1 + \mathbb{P}(X_i \leq a_{i,n}, X_j \leq a_{j,n})] \\ &= 2 - \lim_{n \rightarrow \infty} n[1 - \mathbb{P}(X_i \leq a_{i,n}, X_j \leq a_{j,n})]. \end{aligned}$$

By Proposition 5.10(b), whose conditions are satisfied according to Proposition A.2, and Eq. (5.38) of Resnick (1987), we find

$$\chi(i, j) = 2 + \log G_{ij}((-1/\log G_i)^\leftarrow(1), (-1/\log G_j)^\leftarrow(1)),$$

where $(-1/\log G_i)^\leftarrow$ and $(-1/\log G_j)^\leftarrow$ denote the generalized inverses of the functions $-1/\log G_i$ and $-1/\log G_j$. With the representations for G_i , G_j , and G_{ij} from Proposition A.2, we then obtain by a simple calculation

$$\chi(i, j) = 2 - \sum_{k \in \text{An}(i) \cup \text{An}(j)} \bar{b}_{ki} \vee \bar{b}_{kj}.$$

Finally, using Lemma A.1(b), (a) yields

$$\begin{aligned} \chi(i, j) &= \sum_{k \in \text{An}(i) \cup \text{An}(j)} \bar{b}_{ki} + \sum_{k \in \text{An}(i) \cup \text{An}(j)} \bar{b}_{kj} - \sum_{k \in \text{An}(i) \cup \text{An}(j)} \bar{b}_{ki} \vee \bar{b}_{kj} \\ &= \sum_{k \in \text{An}(i) \cup \text{An}(j)} \bar{b}_{ki} \wedge \bar{b}_{kj} = \sum_{k \in \text{An}(i) \cap \text{An}(j)} \bar{b}_{ki} \wedge \bar{b}_{kj}. \end{aligned}$$

We learn from this proof that \mathbf{X} and the limit vector \mathbf{M} from (2) have the same TDM, since $\mathbf{M} \in \text{MDA}(G)$. \square