Decomposability and selection of graphical models for multivariate time series

BY ROLAND FRIED *

Department of Statistics, University of Dortmund, 44221 Dortmund, Germany

fried@statistik.uni-dortmund.de

AND VANESSA DIDELEZ

Department of Statistical Science, University College London,

London WC1E 6BT, U.K.

vanessa@stats.ucl.ac.uk

SUMMARY

We derive conditions for decomposition and collapsibility of graphical interaction models for multivariate time series. These properties enable us to perform stepwise model selection under certain restrictions. For illustration, we apply the results to a multivariate time series describing the haemodynamic system as monitored in intensive care.

Some key words: Multivariate time series; Partial spectral coherency; Graphical interaction model; Marginalization; Model selection.

1 Introduction

In multivariate data, usually a multitude of relations among the variables can be found. However, many of them may be spurious or indirect, i.e. induced by others. Graphical models display the essential relations between the variables in graphical form. The variables are represented as vertices and those pairwise associations that persist when removing the effects of the other variables are shown as edges. In most cases the statistical meaning of association is some kind of conditional dependence. Thus, a missing edge indicates *conditional independence* of the corresponding

*Corresponding author, phone ++49 - 231 - 755 - 3129, fax ++49 - 231 - 755 - 5305.

variables given all the remaining variables. Dawid (1979) discusses conditional independence as a basic tool for statistical inference, and the monographs by Whittaker (1990), Cox & Wermuth (1996), Lauritzen (1996) and Edwards (2000) give broad reviews on graphical models.

Brillinger (1996) proposes a suitable modification of graphical models for analysing the associations among the components of a multivariate time series, further developed by Dahlhaus (2000). Their approach is based on the partial spectral coherency, which measures the linear dependence between two components of a multivariate time series after removing the linear effects of the remaining components (cf. Brillinger, 1981, section 8.3). This provides a method to detect associations due to partial linear, possibly time–lagged, relations between the variables of a multivariate time series. The graphical representation is thus termed 'partial correlation graph'. For a multivariate Gaussian process, conditional independence of two component processes and zero partial correlations at all time lags are equivalent, whereas the former is a stronger property in general. The usefulness of partial correlation graphs has already been proved in medical applications (Timmer et al., 2000, Gather et al., 2002).

In this paper, we focus on dependence structures of subprocesses derived from a high–dimensional time series by marginalising with respect to some of the components. The results that we present are useful for several reasons: In some situations we might wish to discard some of the components because their measurement is difficult, inaccurate, or simply incomplete. It is then helpful to know whether an omission induces spurious or misleading associations among the remaining, fully observed, components. A similar idea is exploited by Didelez & Pigeot (1999) in order to cope with missing values in a non–dynamic setting. Another reason for our investigations is the desire to reduce computational complexity without distorting the underlying dependence structures. In fact, simplification of complex manipulations of multivariate distributions is one of the main benefits of conditional independence graphs and exploited e.g. by probabilistic expert systems (Cowell et al., 1999). Finally, let us mention that properties such as decomposition and collapsibility can facilitate more sophisticated selection strategies of graphical models for time series. We advocate that one should take advantage of the graphical independence structure, as far as possible, to obtain a more refined and reliable selection procedure. This latter aspect will be investigated in more detail when having presented the theoretical results. While conditions for decomposability of conditional independence graphs for independent observations are well–known (Frydenberg, 1990), they have not yet been derived for graphical interaction models for dynamic data.

We proceed as follows. In Section 2 we introduce the necessary terminology and review partial correlation graphs for multivariate time series. In Section 3 we derive conditions for decomposability and collapsibility of these models that allow the use of standard methods to perform the estimations under specific restrictions. A model selection procedure that exploits these properties is proposed in Section 4 and illustrated in Section 5 by an application to a multivariate time series describing the haemodynamic system of a critically ill patient. We close with some conclusions in Section 6.

2 Graphical Models for Multivariate Time Series

2.1 Graph Notations

A *graph* $G = (V, E)$ consists of a finite set of *vertices* V and a set of *edges* $E \subseteq V \times V$, that are ordered pairs of vertices. It can be visualised by drawing a circle for each vertex and connecting each pair a, b of vertices whenever $(a, b) \in E$ or $(b, a) \in E$ by an edge. In this paper we focus exclusively on *undirected graphs* where $(a, b) \in E$ implies $(b, a) \in E$. The edges are therefore simply represented by lines. Directed graphs typically encode different dependence structures, subject to the kind of graph. In the context of stochastic processes, directed edges have been used to depict asymmetric dependencies (Eichler, 2000, Didelez, 2000) such as influences from past events on the presence or future. However, in the undirected graphs considered here, an edge stands for a symmetric association.

In *conditional independence graphs*, the vertices represent univariate random variables. The *pairwise Markov property* for undirected graphical models then states that two variables a and b are conditionally independent given all remaining variables if they are not connected by an edge, i.e. if $(a, b), (b, a) \notin E$. In a *partial correlation graph* for a multivariate time series each vertex stands for a component of that series and a missing edge indicates a zero partial correlation between a and b at all time lags after removing the linear effects of the remaining variables (Dahlhaus, 2000). A more formal definition is given in the next section.

If two variables a and b are *connected* by a *path*, i.e. if vertices $a = a_0, \ldots, a_l = b$,

$l \geq 1$, exist such that then there is an edge between each pair of successive vertices, there is some (linear) relation between them (possibly mediated by other variables). A *connectivity component* of an undirected graph is a maximal subset of pairwise connected variables (for an undirected graph connectivity is an equivalence relation on V). A graph that includes all edges is called *complete*. It typically represents the saturated model allowing for all associations. If we eliminate some vertices retaining only a subset A and eliminate all edges (a, b) not contained in $A \times A$ we get the *subgraph* G_A induced by A. The *boundary* $bd(A)$ of $A \subset V$ consists of all vertices $v \in V \setminus A$ that are joined by an edge to some vertex $a \in A$. The *closure* of A is $A \cup bd(A)$.

The following notion of a decomposition will be useful for dimension reduction: Let (A, B, S) be a partition of V. Then, we say that G can be decomposed into subgraphs $G_{A \cup S}$ and $G_{B \cup S}$ if (i) S separates A and B in G, i.e. if every path between any $a \in A$ and $b \in B$ necessarily contains at least one vertex $s \in S$, and (ii) G_S is complete. The triple (A, B, S) is then called a *decomposition of* G. Obviously, for dimension reduction, a decomposition has to be *proper*, i.e. $A \neq \emptyset$ and $B \neq \emptyset$.

2.2 Partial Correlation Graphs

Let us now address in more detail the statistical models induced by graphs as they are considered in this paper. Let $\mathrm{X}(t) = \{(X_1(t), \ldots, X_k(t))', \, t \in \mathbb{Z}\}$, be a multivariate stationary time series of dimension k. Suppose that the autocovariance fuction

$$\gamma_{ab}(h) = cov\{X_a(t+h), X_b(t)\}, \; h \in \mathbb{Z},$$

is absolutely summable with respect to all time lags h for all pairs $a, b \in V = \{1, \ldots, k\}$. Then the *cross–spectrum* between two components $\{X_a(t), t \in \mathbb{Z}\}$ and $\{X_b(t), t \in \mathbb{Z}\}$ of the time series is defined as the Fourier–transform of their covariance function

$$f_{ab}(\lambda) = f_{X_a X_b}(\lambda) = \frac{1}{2\pi} \sum_{h=-\infty}^{\infty} \gamma_{ab}(h) \exp(-i\lambda h), \quad \lambda \in [0, \pi]$$

(see Brillinger, 1981, p. 232). This decomposes the covariance function γ_{ab} into periodic functions of frequencies λ. The variables X_a and X_b are uncorrelated at all time lags h iff $f_{ab}(\lambda)$ equals zero at all frequencies. Similarly, we can define the *cross–spectrum* between two multivariate time series $\{\mathrm{X}_A(t), t \in \mathbb{Z}\}$ and $\{\mathrm{X}_B(t), t \in \mathbb{Z}\}$ to be the (component wise) Fourier–transform of their covariance matrix function

$\{\Gamma_{AB}(h), h \in \mathbb{Z}\}$. For ease of notation we will also use X_A instead of $\{\mathrm{X}_A(t), t \in \mathbb{Z}\}$ in the following.

Let (A, B, C) be a partition of the observed variables. In order to assess the *partial* linear relations between the variables in A and the variables in B we eliminate the linear effects of X_C from X_A and X_B. Thus, we have to determine a vector $\mu^\star_{A,C}$ and a filter $\{\mathrm{d}^\star_{A,C}(h), h \in \mathbb{Z}\}$ minimising the expectation of the vector product

$$E\left[\left\{\mathrm{X}_A(t) - \mu_{A,C} - \sum_h \mathrm{d}_{A,C}(h)\mathrm{X}_C(t-h)\right\}\left\{\mathrm{X}_A(t) - \mu_{A,C} - \sum_h \mathrm{d}_{A,C}(h)\mathrm{X}_C(t-h)\right\}'\right].$$

If the spectral density matrix $f_{VV}(\lambda)$ for the whole process is regular at all frequencies the solution is unique (Brillinger, 1981, Theorem 8.3.1) and we can define the residual series $\{\varepsilon_{A\cdot C}(t), t \in \mathbb{Z}\}$ by

$$\varepsilon_{A\cdot C}(t) = \mathrm{X}_A(t) - \mu^\star_{A,C} - \sum_{h=-\infty}^{\infty} \mathrm{d}^\star_{A,C}(h)\mathrm{X}_C(t-h)\,.$$

The *partial cross–spectrum* between X_A and X_B given X_C can then be defined as the cross–spectrum between $\{\varepsilon_{A\cdot C}(t), t \in \mathbb{Z}\}$ and $\{\varepsilon_{B\cdot C}(t), t \in \mathbb{Z}\}$

$$f_{AB\cdot C}(\lambda) = f_{\varepsilon_{A\cdot C}\varepsilon_{B\cdot C}}(\lambda)\,,\ \lambda \in [0, \pi]\,.$$

Brillinger (1981, p. 296) shows that the partial cross–spectrum can be calculated using

$$f_{AB\cdot C}(\lambda) = f_{AB}(\lambda) - f_{AC}(\lambda)\{f_{CC}(\lambda)\}^{-1}f_{CB}(\lambda)\,, \tag{1}$$

where the entries of the matrices $f_{AC}(\lambda)$, $f_{CB}(\lambda)$ and $f_{CC}(\lambda)$ are ordinary cross–spectra between the corresponding variables.

The *partial spectral coherency* between component processes X_a and X_b is a standardization of their partial cross–spectrum,

$$R_{ab\cdot V\setminus\{a,b\}}(\lambda) = \frac{f_{ab\cdot V\setminus\{a,b\}}(\lambda)}{\{f_{aa\cdot V\setminus\{a,b\}}(\lambda)f_{bb\cdot V\setminus\{a,b\}}(\lambda)\}^{1/2}}\,, \tag{2}$$

that measures partial correlation as a function of the frequency, while the *(partial) phase–spectrum* between components a, b is defined via the Eulerian representation

$$f_{ab}(\lambda) = |f_{ab}(\lambda)|\exp\{i\phi_{ab}(\lambda)\}\,.$$

It can be interpreted as angle between the residual components of frequency λ.

Zero partial correlation at all time lags defines an *orthogonality relation* $\perp\!\!\!\perp$ as follows:

$$\begin{aligned} \mathrm{X}_A \perp\!\!\!\perp \mathrm{X}_B \mid \mathrm{X}_C \quad &\Longleftrightarrow \quad cov\{\varepsilon_{A\cdot C}(t), \varepsilon_{B\cdot C}(t+h)\} = 0 \quad \forall\, h \in \mathbb{Z} \\ &\Longleftrightarrow \quad f_{AB\cdot C}(\lambda) = 0 \quad \forall\, \lambda \in [0, \pi] \\ &\Longleftrightarrow \quad R_{AB\cdot C}(\lambda) = 0 \quad \forall\, \lambda \in [0, \pi], \end{aligned} \tag{3}$$

where $R_{AB\cdot C}(\lambda)$ denotes the matrix with entries $R_{ab\cdot C}$, $a \in A$, $b \in B$. For ease of notation we also write $A \perp\!\!\!\perp B|C$ instead of $\mathrm{X}_A \perp\!\!\!\perp \mathrm{X}_B|\mathrm{X}_C$ identifying the components of subvectors with their indices.

The connection between a graph and the partial correlation structure of a multivariate time series is formally established as follows. Let the vertices $V = \{1, \ldots, k\}$ of G represent the components of a time series X_V. Then we say that G is the *partial correlation graph* of X_V if

$$(a, b), (b, a) \notin E \quad \Longrightarrow \quad a \perp\!\!\!\perp b \mid V \setminus \{a, b\}, \tag{4}$$

i.e. $f_{ab\cdot V\setminus\{a,b\}}(\lambda) = 0$ for all frequencies $\lambda \in [0, \pi]$ and vice versa.

Partial correlation graphs for multivariate time series defined like this generalise conditional independence graphs for a multivariate normal distribution since for an independent sample the matrix of cross-spectra is a constant multiple of the covariance matrix.

Let $\{\varepsilon_{B\cdot C}(t), t \in \mathbb{Z}\}$ be calculated analogously from X_B. Note, that both past and future effects of X_C are eliminated here as we use the resulting partial correlations to define undirected graphical models. It would be natural to define directed graphical models for multivariate time series by eliminating past effects only (Eichler, 2000).

While the foregoing definition of partial correlation graphs refers to a pairwise property it implies much more vanishing partial correlations than are apparent at first glance. Dahlhaus (2000) proves for partial correlation graphs that, under the assumption of the spectral matrix being regular at all frequencies, the above pairwise Markov property (4) for partial correlation graphs implies the so–called *global Markov property.* The latter is generally stronger than the pairwise property and states that two subprocesses, A and B, have zero partial correlations at all time lags given the linear effects of a subprocess $C \subseteq V\backslash(A \cup B)$ whenever C separates A and B in G. Put differently, the variables in A and B are not associated if the linear effects of a separating subset C are controlled. Note that A, B, C do not need to be a partition of V, i.e. zero partial correlations can be retained even when discarding the components in $V\backslash(A \cup B \cup C)$, nor is C necessarily unique. The requirement that the spectral density matrix be regular at all frequencies essentially implies that there is no linear relationship among the components of the multivariate time series at any frequency. This would mean that some of the components would carry no new information making them redundant at a frequency where the spectral matrix is not regular.

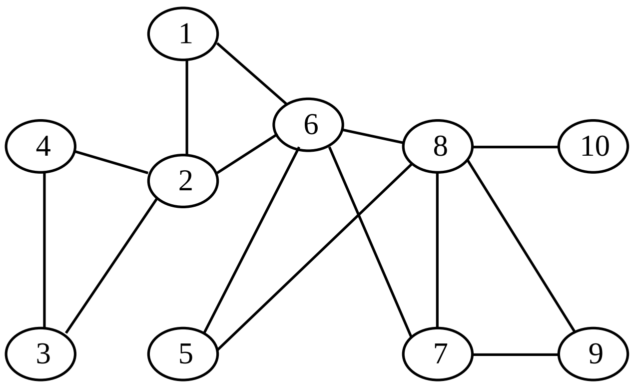

Figure 1: True partial correlation graph for a special VAR(1)-process.

In order to illustrate partial correlation graphs for time series we use a VAR(1)-process $\mathrm{X} = (X_1, \ldots, X_{10})$ with ten variables

$$\mathrm{X}(t) \quad = \quad \Phi \mathrm{X}(t-1) + \varepsilon(t)\,,$$

where $\{\varepsilon(t) : t \in \mathbb{Z}\}$ is a white noise process consisting of $N(0, I)$-distributed errors and $\Phi = (\phi_{i,j})$ with $\phi_{i,i} = 0.3, i = 1, \ldots, 10$, $\phi_{1,2} = \phi_{2,1} = \phi_{2,6} = \phi_{3,2} = \phi_{3,4} = \phi_{6,5} = \phi_{6,8} = \phi_{7,6} = \phi_{9,7} = \phi_{9,8} = \phi_{10,8} = \phi_{8,10} = 0.3$, and the remaining $\phi_{k,l} = 0$. As proved by Dahlhaus (2000), in the partial correlation graph for such a VAR(1)-process two component processes a and b have to be connected by an edge iff $\phi_{a,b} \neq 0$, or $\phi_{b,a} \neq 0$, or if a and b jointly affect a third component, i.e. if there exists a $c \in V \setminus \{a, b\}$ such that $\phi_{c,a} \neq 0$ and $\phi_{c,b} \neq 0$. The resulting partial correlations graph for this multivariate process is shown in Figure 1. The edges induced by two components jointly affecting a third one are: $(2, 4)$, $(5, 8)$, and $(7, 8)$. A conditional independencies that can be read off the graph is, for instance, $1 \perp\!\!\!\perp 3 \mid \{2, 4, 5, 6, 7, 8, 9, 10\}$ because there is no edge between X_1 and X_3 (pairwise Markov property). Further we get from the global Markov property that conditioning on X_2 suffices to make X_1 and X_3 independent, i.e. $1 \perp\!\!\!\perp 3|2$, due to the corresponding graph separation.

3 Decomposition of Partial Correlation Graphs

In the following we derive some properties of partial correlation graphs that aim at reducing complexity in several ways. In particular, we are interested in the partial correlation structure of subprocesses X_A, $A \subseteq V$, or, more formally, in the partial cross–spectra $f_{ab \cdot A \setminus \{a,b\}}(\lambda)$, $a, b \in A$. Note that the pattern of vanishing partial correlations in such a submodel does not necessarily have to match the subgraph G_A of the partial correlation graph G for the whole process. Marginalising with respect to $\mathrm{X}_{V \setminus A}$ might induce other associations or independencies than shown in G_A. Such phenomena are known as Simpson's Paradox (Simpson, 1951).

Throughout this section we assume that $G = (V, E)$ is the partial correlation graph of a multivariate stationary time series X_V and that the spectral density matrix of X_V is regular at all frequencies λ.

PROPOSITION 1 (decomposition). If (A, B, S) is a decomposition of G then the subgraph $G_{A \cup S}$ is not smaller than the partial correlation graph of the subprocess $\mathrm{X}_{A \cup S}$ in the sense that all edges missing in $G_{A \cup S}$ are also missing in the partial correlation graph of this subprocess.

Proof. This result can be proved in the same way as Proposition 12.2.1 in Whittaker (1990) using the global Markov property shown by Dahlhaus (2000). □

The condition in Proposition 1 ensures that missing edges in a subgraph can still be regarded as zero partial correlations within the corresponding subprocess after marginalising over the remaining components. This applies for instance to the AR(1)-process introduced in the last section choosing the subgraph $G_{\{1,2,3,4,5,6\}}$ since $(\{1, 2, 3, 4\}, \{7, 8, 9, 10\}, \{5, 6\})$ is a decomposition.

The next proposition transfers a result on collapsibility, which is well–known in the case of conditional independence graphs (cf. Whittaker, 1990, proposition 12.5.1) to partial correlation graphs for multivariate time series. It uses proposition 1 to characterise subsets that can be neglected without inducing additional dependencies among the remaining components.

PROPOSITION 2 (collapsibility). If the boundary of each connectivity component of $B \subseteq V$ is complete then $G_{V\backslash B}$ is not smaller than the partial correlation graph of the subprocess $\mathrm{X}_{V\backslash B}$. We say that G is *collapsible* onto $V\backslash B$ (or over B).

Note, that the condition of each connectivity component of $B \subseteq V$ being complete corresponds to the notion of a *simplicial* subset as introduced by Frydenberg (1990).

Proof of Proposition 2. First suppose that B consists of one connectivity component only. We define $A = V \setminus \{B \cup bd(B)\}$. Since the boundary of B is complete and separates A from B we have that $(A, B, bd(B))$ is a decomposition. Thus, the result is a simple consequence of proposition 1. If there are several connectivity components the result follows from repeated application of this argument by successively discarding the connectivity components of B. □

In our example, we can apply this proposition e.g. to $B = \{1, 7, 8, 9, 10\}$ so that the missing edges in $G_{\{2,3,4,5,6\}}$ correspond to zero partial correlations within $X_{\{2,3,4,5,6\}}$.

We will now show that zero partial correlations have their counterpart in zero regression coefficients in dynamic regression. Here, the separating subset can be incomplete as we consider relations between the separated sets A and B only.

PROPOSITION 3 (dynamic regression). Let (A, B, S) be a partition of V. A and B are partially uncorrelated at all time lags, $A \perp\!\!\!\perp B|S$, iff all entries of the optimal linear filter $\mathrm{d}^{\star}_{A,B\cup S}(h)$ being coefficients of components in B equal zero at all time lags h.

Proof. The optimal vector $\mu^{\star}_{A,B\cup S}$ and the optimal linear filter $\mathrm{d}^{\star}_{A,B\cup S}$ can be calculated from a multiple regression of X_A on $\mathrm{X}_{B\cup S}$ in the frequency domain considering each frequency individually (Brillinger, 1981, Theorem 8.3.1):

$$\mu^{\star}_{A,B\cup S} = E\left[X_A(t) - \left\{\sum_h \mathrm{d}^{\star}_{A,B\cup S}(h)\right\} E\{\mathrm{X}_{B\cup S}(t)\}\right] \tag{5}$$

$$\mathrm{d}^{\star}_{A,B\cup S}(h) = \frac{1}{2\pi}\int_0^{2\pi} \mathrm{A}(\lambda)\exp(ih\lambda)d\lambda \tag{6}$$

$$\text{with } \mathrm{A}(\lambda) = (f_{AB}(\lambda), f_{AS}(\lambda)) \begin{pmatrix} f_{BB}(\lambda) & f_{BS}(\lambda) \\ f_{SB}(\lambda) & f_{SS}(\lambda) \end{pmatrix}^{-1}. \tag{7}$$

It is easily verified that this inverse can be written as

$$\begin{pmatrix} N^{-1} & -N^{-1}M \\ -PN^{-1} & \{f_{SS}(\lambda)\}^{-1} + PN^{-1}M \end{pmatrix},$$

where $M = f_{BS}(\lambda)\{f_{SS}(\lambda)\}^{-1}$, $N = f_{BB}(\lambda) - f_{BS}(\lambda)\{f_{SS}(\lambda)\}^{-1}f_{SB}(\lambda)$ and $P = \{f_{SS}(\lambda)\}^{-1}f_{SB}(\lambda)$.
Assume now that all entries of $\mathrm{d}^{\star}_{A,B\cup S}(h)$ which correspond to components in B equal zero at all time lags h. This is equivalent to all functions in the Fourier transform $\mathrm{A}(\lambda)$ of the filter, that correspond to elements in B, being identical to zero, cf. (6). From (7) we see that this is equivalent to

$$\begin{aligned} 0 &= \{f_{AB}(\lambda) - f_{AS}(\lambda)\{f_{SS}(\lambda)\}^{-1}f_{SB}(\lambda)\}N^{-1} \\ \Longleftrightarrow 0 &= f_{AB}(\lambda) - f_{AS}(\lambda)\{f_{SS}(\lambda)\}^{-1}f_{SB}(\lambda)\,. \end{aligned}$$

From formula (1) we see, in turn, that this is equivalent to $f_{AB\cdot S}(\lambda)$ being equal to zero for all frequencies which defines the orthogonality $A \perp\!\!\!\perp B|S$, cf. (3).
Further, we see immediately from formula (5) that $\mu^{\star}_{A,B\cup S}$ does not depend on the variables in B in this case. $\square$

Since the above argumentation is symmetric in A and B, we also find that all entries of the optimal linear filter $\mathrm{d}^{\star}_{B,A\cup S}(h)$ being coefficients for components in A equal zero at all time lags, too.

For the AR(1)-process introduced in section 1, we see from proposition 3 e.g. that the optimal linear filter for X_1 given the other variables depends on $X_{2,6}$ only. In principle, proposition 3 can be used to construct optimal filters for a model satisfying the restrictions by a partial correlation graph. The next proposition provides a computationally more efficient solution for the estimation of the partial spectral coherencies in special cases. It is similar to proposition 1. However, it is stronger than the latter since it shows that not only the zero partial spectral coherencies but all partial spectral coherencies within the subprocesses X_A or X_B are retained if (A, B, S) is a partition of the graph such that $A \perp\!\!\!\perp B|S$.

PROPOSITION 4. Let (A, B, S) be a partition of V such that $A \perp\!\!\!\perp B|S$. Then

a) the partial spectral coherency of $a, \tilde{a} \in A$ w.r.t. X_V is the same as the partial spectral coherency of these components w.r.t. $\mathrm{X}_{A\cup S}$, i.e.

$$R_{a\tilde{a}\cdot V\setminus\{a,\tilde{a}\}}(\lambda) = R_{a\tilde{a}\cdot(A\cup S)\setminus\{a,\tilde{a}\}}(\lambda), \quad \lambda \in [0, \pi] .$$

b) the partial cross-spectrum of $a \in A$ and $s \in S$ w.r.t. X_V is the same as the partial cross-spectrum of these components w.r.t. $\mathrm{X}_{A\cup S}$, i.e.

$$f_{as\cdot V\setminus\{a,s\}}(\lambda) = f_{as\cdot(A\cup S)\setminus\{a,s\}}(\lambda), \quad \lambda \in [0, \pi] .$$

Proof. a) Since S separates A and B in G, it is obvious that $(S \cup A) \setminus \{a, \tilde{a}\}$ separates $\{a, \tilde{a}\}$ from B in this graph. Using proposition 3 we see that in the optimal filter $\mathrm{d}^\star_{a,V\setminus\{a,\tilde{a}\}}$ all coefficients of variables in B equal zero. Thus, the non–zero elements of the optimal filter for a taking all components into account are the same as when taking only $A \cup S$ into account, and the optimal constants $\mu^\star_{a,V\setminus\{a,\tilde{a}\}}$ and $\mu^\star_{a,(A\cup S)\setminus\{a,\tilde{a}\}}$ are also the same. This implies that the residual series $\varepsilon_{a\cdot V\setminus\{a,\tilde{a}\}}$ and $\varepsilon_{a\cdot(A\cup S)\setminus\{a,\tilde{a}\}}$ are identical. The same is true for $\tilde{a}$, of course, and hence a and $\tilde{a}$ have the same partial spectral coherencies in both processes.
b) For $s \in S$ the residual series $\varepsilon_{s\cdot V\setminus\{a,s\}}$ and $\varepsilon_{s\cdot(A\cup S)\setminus\{a,s\}}$ are usually not identical. However, the difference is a linear transform of $\mathrm{X}_{V\setminus\{a,s\}}$ as

$$\begin{aligned}
\varepsilon_{s\cdot(A\cup S)\setminus\{a,s\}}(t) - \varepsilon_{s\cdot V\setminus\{a,s\}}(t) \quad &= \quad \mu^\star_{s,V\setminus\{a,s\}} + \sum_{u=-\infty}^{\infty} \mathrm{d}^\star_{s,V\setminus\{a,s\}}(u)\mathrm{X}_{V\setminus\{a,s\}}(t-u) \\
&- \quad \mu^\star_{s,(A\cup S)\setminus\{a,s\}} - \sum_{u=-\infty}^{\infty} \mathrm{d}^\star_{s,(A\cup S)\setminus\{a,s\}}(u)\mathrm{X}_{(A\cup S)\setminus\{a,s\}}(t-u) .
\end{aligned}$$

Since $\varepsilon_{a\cdot V\setminus\{a,s\}} \perp\!\!\!\perp \mathrm{X}_{V\setminus\{a,s\}}$, cf. proposition 3a), we have

$$\begin{aligned}
& cov\{\varepsilon_{a\cdot(A\cup S)\setminus\{a,s\}}(t), \varepsilon_{s\cdot(A\cup S)\setminus\{a,s\}}(t+h)\} \\
= \; & cov\{\varepsilon_{a\cdot V\setminus\{a,s\}}(t), \varepsilon_{s\cdot V\setminus\{a,s\}}(t+h)\} \\
& + cov\{\varepsilon_{a\cdot V\setminus\{a,s\}}(t), \varepsilon_{s\cdot(A\cup S)\setminus\{a,s\}}(t+h) - \varepsilon_{s\cdot V\setminus\{a,s\}}(t+h)\} \\
= \; & cov\{\varepsilon_{a\cdot V\setminus\{a,s\}}(t), \varepsilon_{s\cdot V\setminus\{a,s\}}(t+h)\} + 0
\end{aligned}$$

for all time lags $h \in \mathbb{Z}$. The partial cross-spectrum between a and s is simply the Fourier transform of these covariances which proves the result. □

Given the partial correlation graph G of X_V and a partition (A, B, S) such that $A \perp\!\!\!\perp B | S$, the above result allows to estimate the partial spectral coherency between variables $a \in A$ and $\tilde{a} \in A$ discarding the variables in B using the subprocess $A \cup S$, only, without any loss of information. Amongst other things, we may then also test zero partial spectral coherencies within X_A w.r.t. the whole time series X_V using only subprocesses in $A \cup S$. We may then argue that the latter tests will usually be more powerful as we do not adjust for irrelevant variables. This is particularly interesting when the subgraph induced by $A \cup S$ is complete since then simple computational formula exist, cf. (1). For a given partial correlation graph, we can estimate the partial spectral coherency between variables a and b using the unrestricted process $\mathrm{X}_{cl(a,b)}$, since $bd(a, b)$ separates $\{a, b\}$ from $V \setminus cl(a, b)$. In our running example, it is sufficient to observe $\mathrm{X}_{\{2,3,4\}}$ for estimating the partial spectral coherency among X_3 and X_4.

In most practical applications, however, we do not know the partial correlation graph beforehand so that the above ideas cannot directly be applied. In the following section we therefore suggest a stepwise procedure, where we first estimate a preliminary graph the separation properties of which are then exploited to find more reliable estimates based on subprocesses. Note that the empirical partial spectral coherencies will rarely have exact zero entries when the underlying partial correlation is zero. Therefore, the estimates based on the whole process and those utilising the subprocess $\mathrm{X}_{A \cup S}$ will typically differ even though the theoretical ones do not.

4 Selection of Partial Correlation Graphs

4.1 Motivation

In applications, when investigating research hypotheses by empirical analysis of multivariate time series data, one can first estimate the cross–spectra from the data and then use the empirical versions of equations (1) and (2) to estimate the partial spectral coherencies. Thereafter, a decision has to be made whether the underlying theoretical partial spectral coherency may be zero — obviously, sampling variability will cause estimates to be distinct from zero. The program *Spectrum* developed by Dahlhaus & Eichler (2000) estimates the cross–spectrum using a nonparametric kernel estimator. In addition, it constructs an approximate bound for the 95%–quantile of the maximal squared estimated partial spectral coherency in the saturated model (corresponding to the complete graph) assuming that the considered true partial spectral coherency is zero. Thus partial uncorrelatedness of each pair of variables can be tested simultaneously at the same approximate local significance level by comparing the estimated partial spectral coherency with this bound. Such a procedure results in an one–step selection of a partial correlation graph: Edges (a, b) have to be included whenever the partial spectral coherency between X_a and X_b is significantly different from zero.

In many applications it is known that associations may have different strengths and it seems straightforward to include this information into the graph. The strength of an association can heuristically be regarded as an expected relative change in one of the variables when the other one changes by a certain relative amount adjusting for all remaining ones. A classification of the strength may be based on the area under the partial spectral coherency $R_{ab\cdot V\setminus\{a,b\}}$ as measured by the *partial mutual information*

$$-\frac{1}{2\pi}\int \log\{1 - |R_{ab\cdot V\setminus\{a,b\}}(\lambda)|^2\}d\lambda,$$

between the time series X_a and X_b (Brillinger, 1996, Granger & Hatanaka, 1964), or by variants of this. In a clinical context, Gather et al. (2002) use gradually distinct edges to classify the strength of the associations estimated in the saturated model. Note, however, that all properties of the dependence structure that can be read off the graph rely on the missing edges, only, i.e. on zero partial correlations, regardless of the strength of the non–vanishing associations.

Although reporting the strength of the associations provides information on ambiguous edges it is still a heuristic approach. When calculating the partial spectral coherencies between two variables w.r.t. X_V we eliminate the linear influences of all other variables and not only the linear effects of those variables that are really relevant. It seems plausible to suspect that this may hide some associations by reducing the power to reject zero partial correlations. Such a mistake might be prevented by conditioning on smaller sets whenever possible.

More refined selection strategies often proceed stepwise such as backward or forward procedures. Application of this kind of strategies to time series data requires the estimation in models where some of the partial spectral coherencies are restricted to zero. However, to the best of our knowledge, there is no general estimation theory for such a task available yet. We run into problems even if we do not use a non–parametric approach but assume a parametric model like a Gaussian vector

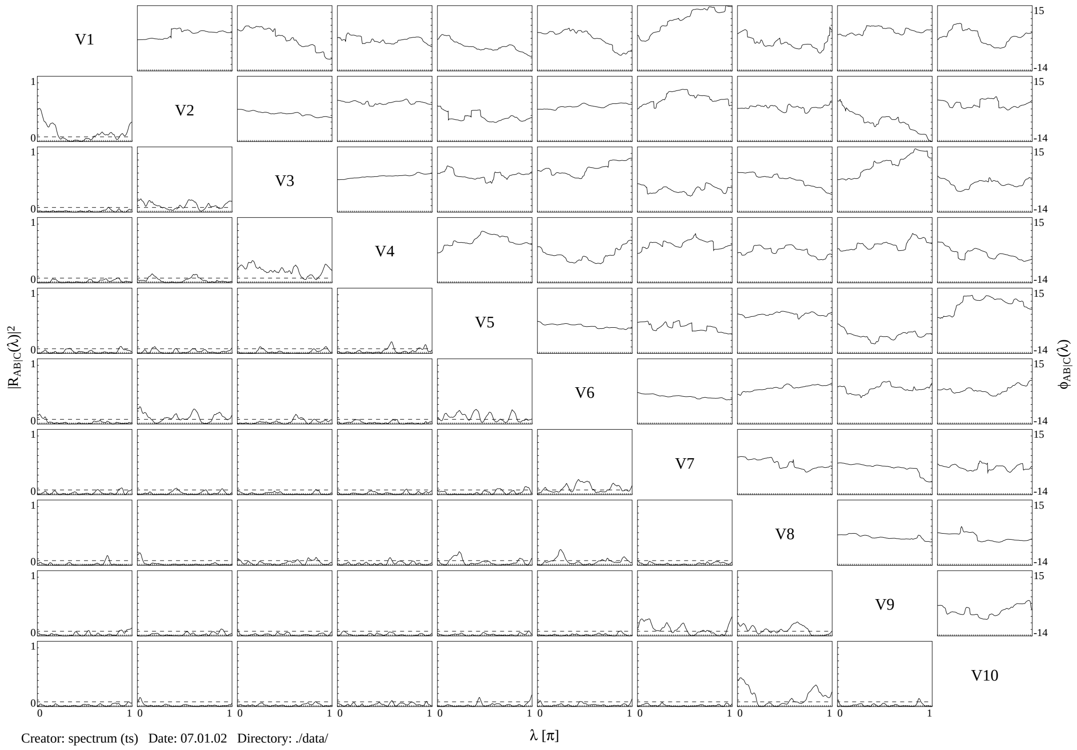

Figure 2: AR(1)-process: Squared estimated partial spectral coherencies (below the diagonal) and partial phase spectra (above the diagonal). Some strong partial correlations are evident from the partial spectral coherencies, cf. e.g. (1,2), while other partial spectral coherencies are close to zero, e.g (1,3).

autoregressive model (Dahlhaus & Eichler, 2001).

Instead, we suggest to utilise the properties of partial correlation graphs derived in the previous section and estimate the partial spectral coherencies non–parametrically taking at least those restrictions into account that come along with a separation of the graph. Starting by estimating the partial spectral coherencies in the saturated model we then eliminate associations for which the area below the curve representing the squared partial spectral coherency which is above the approximate joint 95% confidence bound is very small. The separation properties of the graph found in this first step may in turn be exploited in a second step to double–check missing and dubious interactions by restricting the computations to appropriate submodels exploiting propositions 2 and 4. This double check can repeatedly be applied if the graph changes in the second step.

For illustration we simulate 1000 observations from the VAR(1)-process introduced in section 2. Applying the program *Spectrum* to the simulated time series reveals strong partial correlations among e.g. $(1,2)$ and $(3,4)$, cf. Figure 2. However, the partial spectral coherencies between e.g. $(2,4)$ and $(1,6)$ exceed the critical bound only slightly at a few frequencies, just like the partial spectral coherency for $(4,5)$. In contrast, the partial spectral coherency for $(7,8)$ does not even cross the confidence bound even though there is an edge in the model. Therefore, the one–step selection would probably either neglect or include all of $(2,4),(1,6),(4,5)$ and certainly neglect $(7,8)$. Figure 3 depicts the preliminary model for the partial correlation graph. Ambiguous edges are represented by dashed lines.

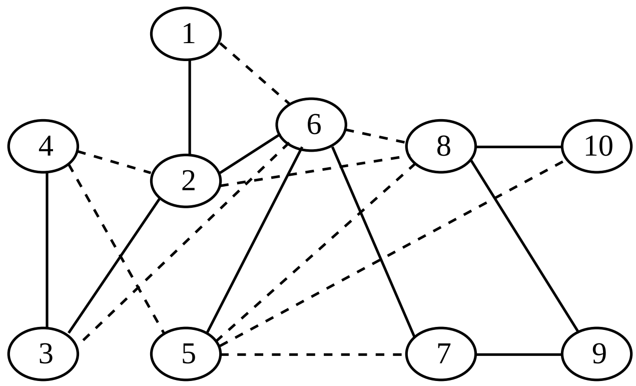

Figure 3: Partial correlation graph for data simulated from a special VAR(1)-process, one-step selection.

Partitions (A,B,S) of G with $A \perp\!\!\!\perp B|S$ and $|A|>1$ allow to double–check the importance of ambiguous and missing edges within A using the subprocess $\mathrm{X}_{A\cup S}$, only, cf. proposition 4a). In principle we could consider all these separations and include an edge if it is significant for any of them. However, in case of many variables there may be many separations and using all of them is computationally expensive. Therefore, we suggest using only the separation $(\{a,b\}, V\setminus cl(a,b), bd(a,b))$ to verify the absence or presence of an edge (a,b) as $cl(a,b)$ is the minimal set for which the partial spectral coherency among (a,b) is the same as for the whole process if we assume that the given partial correlation graph is true. For the missing edge $(7,8)$ e.g. we have $cl(7,8)=\{2,5,6,7,8,9,10\}$. Again, *Spectrum* can be used to estimate the partial spectral coherencies within the saturated model for $A\cup S$. Although this does not use the information provided by the absence of some of the edges within A, it does use the information on the absence of associations between A and B.

Therefore we consider the results obtained in this way to be in better agreement with the current model than the estimates calculated in the first step. In the simulated example, however, the estimated partial spectral coherency of $(7, 8)$ does not change much confirming the absence of $(7, 8)$. For the edges $(2, 4)$ and $(4, 5)$ we have $cl(2, 4) = \{1, 2, 3, 4, 5, 6, 8\}$ and $cl(4, 5) = \{2, 3, 4, 5, 6, 7, 8, 10\}$. Alternatively, we could also check the presence of $(4, 5)$ within the subgraph $cl(2, 4)$ exploiting proposition 4b). Applying *Spectrum* to this subgraph results in an decrease of the estimated partial spectral coherency between $(4, 5)$. Similarly, the estimated partial spectral coherency between $(5, 7)$ decreases when we consider the submodel for $\{4, 5, 6, 7, 8, 9, 10\}$. Therefore, we could decide that these edges have to be omitted, hence modifying the result of the one–step model selection in the second step. In this particular example, exploiting the information on collapsible subsets in the selection process thus leads to a result which is closer to the original graph.

4.2 Selection Strategy

Before describing the selection procedure in more detail let us mention some general principles that should guide the model selection. Firstly, we propose to adhere to the principle of *coherent* model selection (cf. Edwards, 2000, p. 165), which implies in our situation that once a zero partial correlation can clearly be rejected this should not be doubted in further steps. However, it might still be considered worthwhile to investigate whether some of the edges which in previous steps have been found to be ambiguous can be deleted despite the coherence principle. Secondly, it is important to realise that if, in a stepwise selection, the edges of the current graph G' are a superset of those in the true graph G any separation found in the former will still hold in the latter. Hence we start with the complete graph and focus on deleting edges in further steps. Thirdly, and related to the foregoing aspects, expert knowledge on existing and absent associations should be included whenever possible to improve the selection results and to concentrate on the unsettled problems.

In more detail, the model selection we propose proceeds as follows.

1. In the first step, we measure the area below the squared absolute partial spectral coherencies which is above the 95% confidence bound in the complete model for all pairs of variables and utilise these values to classify all possible edges according to the following categories:

(I) Edges that clearly have to be included.

(A) Ambiguous edges that possibly have to be included.

(O) Edges that may be omitted.

2. Since false omission of an edge is worse than wrong inclusion of an edge because it creates separations that might not be reliable, we next examine the edges in (O) in a second (forward) step using separation properties of the graph where only edges belonging to category (O) are omitted. Every edge (a, b) contained in (O) with $V \setminus cl(a, b) \neq \emptyset$ is reclassified based on the estimated partial spectral coherency calculated in the saturated model for $\mathrm{X}_{cl(a,b)}$. If $cl(a, b) = V$ we can possibly apply proposition 4b) to a separation (A, B, S) with $a \in A$ and $b \in S$. We suggest using only separations where $|A \cup S|$ is minimal. This procedure is repeated until the results no longer change.

3. Thereafter, in a third (backward) step, we examine all edges belonging to the updated (A) proceeding as explained above. If the graph can be reduced we may possibly find more separations. In particular, if an edge (a, b) is reclassified as (O) we get the possibility to reclassify the edges (a, c) and (b, c), $c \in V$, that are not contained in (I) using a reduced set $cl(a, c)$ and $cl(b, c)$ respectively. We repeat steps two and three in this case.

The stepwise search stops when the classifications are stable. Note however, that it might happen that this is not the case. We suggest to construct the final partial correlation graph based on the classification obtained in the last step using distinct edges for the edges in category (I) and (A) and omitting the edges in category (O).

The sensitivity of the resulting graph to the classifications obtained in the first step can be examined by performing the same stepwise search strategy starting with a preliminary graph containing only the edges in (I). This usually increases the number of separations and allows further model checks. However, the results from this analysis have to be treated with caution as the starting graph is likely to have too few edges. In general, neither separation nor collapsibility found in a too sparse graph hold in the true graph.

5 Application to Physiological Time Series

The previous results are now illustrated with a multivariate time series describing the haemodynamic system of a critically ill patient measured in intensive care. The data investigated here was collected by a Unix–based clinical information system (CIS) at the surgical intensive care unit of the Community Hospital Dortmund. We consider time series of the heart rate (HR), arterial mean pressure (APM), pulmonary arterial mean pressure ($PAPM$), central venous pressure (CVP), pulsoxymetry ($SpO2$) and blood temperature ($Temp$). All variables are automatically recorded at 1–minute time intervals by bedside medical devices. These measurements provide crucial information on the clinical status of the patient. Over all, 2440 observations measured at subsequent points in time are available for the analysis.

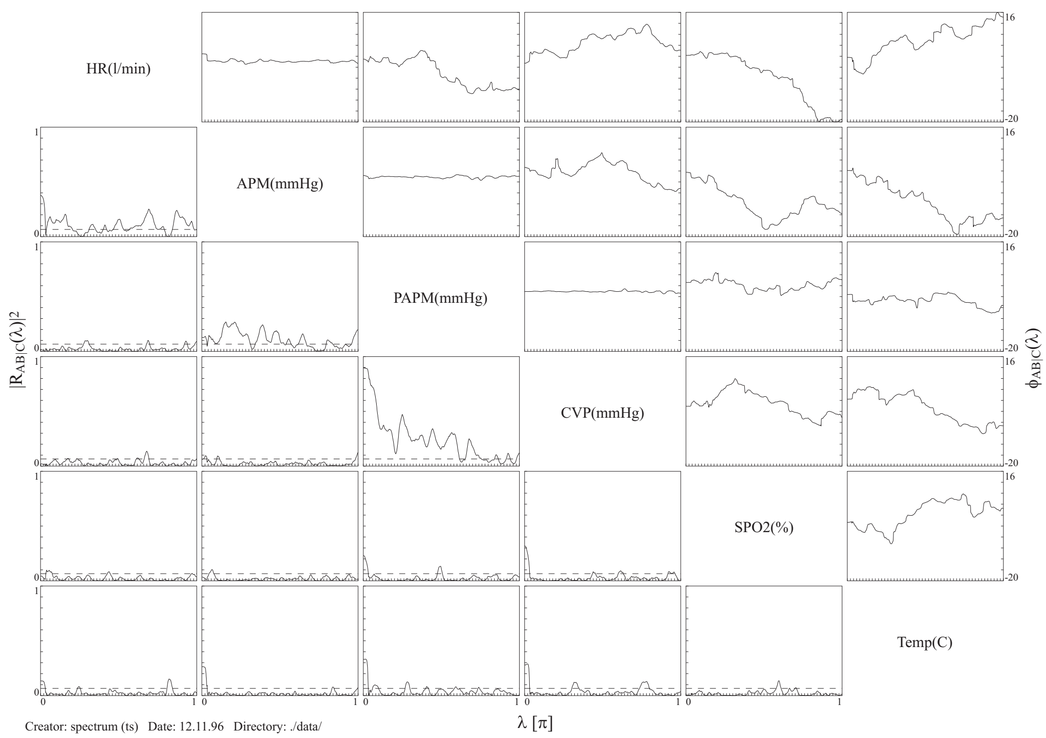

Figure 4: Squared estimated partial spectral coherencies (below the diagonal) within the haemodynamic system of one patient. Strong partial correlations are found between HR and APM, APM and $PAPM$, as well as between $PAPM$ and CVP.

We expect 'empirical associations' found by statistical analysis to reflect 'physiological associations'. The latter mean, in the clinical context, that a change in one physiological variable affects another physiological variable. Note that the term physiological association does not distinguish between causal, linear or non–linear or in some sense directed relations. Therefore we do not aim at finding directed associations but apply graphical interaction models based on (symmetric) partial correlations. On the one hand, this can be used to investigate whether well–known physiological associations among the variables are detected and thus helps evaluating the selection strategy. On the other hand, it gives an impression of how current medical knowledge is supported by the data.

Figure 4 shows the squared pairwise partial spectral coherencies between all variables estimated in the saturated model, i.e. for every pair the linear effects of all remaining components have been removed. For HR and $PAPM$, HR and CVP, HR and $SpO2$, HR and $Temp$, APM and CVP, APM and $SpO2$, as well as for $SpO2$ and $Temp$ the estimate exceeds the approximate 5% confidence bound only at a few frequencies by a slight amount. Therefore we classify the corresponding edges to belong to category (O). For APM and $Temp$, $PAPM$ and $SpO2$, $PAPM$ and $Temp$, CVP and $SpO2$, as well as for CVP and $Temp$ the estimate exceeds the bound in a few regions by a rather large amount. Thus, we classify these edges into category (A). For HR and APM, APM and $PAPM$, and $PAPM$ and CVP the estimate is significantly distinct from zero over the whole range. Therefore, we classify these edges into category (I), cf. Figure 5.

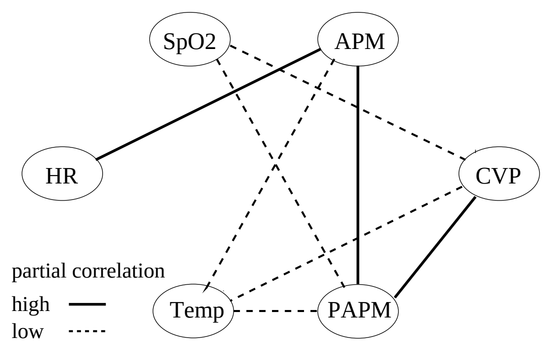

Figure 5: Partial correlation graph for the haemodynamic system, one-step selection.

To refine or validate this model we perform a stepwise search as outlined in the previous section. In the second step, we check the absence of the missing edges using suitable separations (A, B, S) of the partial correlation graph shown in Figure 5, which can be viewed as a summary of all partial spectral coherencies calculated in the first step. We use these partitions to estimate the partial spectral coherencies among the components in $A \cup S$ applying the program *Spectrum* to estimate the partial spectral coherencies for the complete subgraph corresponding to $X_{A\cup S}$.

The absence of $(Temp, SpO2)$ can be checked using $cl(Temp, SpO2) = \{APM, PAPM, CVP, Temp, SpO2\}$. We find only minor changes of the estimated partial spectral coherency between $Temp$ and $SpO2$ as compared to Figure 4. Similarly, the absence of $(HR, Temp)$ and $(HR, SpO2)$ can be validated by estimation of the partial spectral coherencies from the corresponding closures. Proposition 4a) cannot be used to check $(HR, PAPM)$ since $cl(HR, PAPM) = V$. Instead, we consider the decomposition (A, B, S) with $A = \{HR\}$, $B = \{CVP, Temp, SpO2\}$ and $S = \{APM, PAPM\}$ applying proposition 4b). We do not find this edge to be significant as the estimated partial spectral coherency remains small. Repeated application of proposition 2 to check (APM, CVP) using $A = \{APM, CVP, PAPM, Temp\}$, however, results in an increase of the estimated partial spectral coherency, cf. Figure 6. Therefore this edge becomes ambiguous. This reclassification is confirmed by the fact that we already noticed an increase of the estimated partial spectral coherency among APM and CVP before when considering $cl(Temp, SpO2)$. Further checking using proposition 4b) validates the absence of (HR, CVP) and $(APM, SpO2)$.

In the third step we check the edges which have been found to be ambiguous before. We can check $(PAPM, Temp)$, $(CVP, Temp)$, $(CVP, SpO2)$, $(PAPM, SpO2)$, $(APM, Temp)$ and (APM, CVP) using the subprocess $\{APM, PAPM, CVP, Temp, SpO2\}$ and find the estimates of the partial spectral coherencies of $Temp$ to increase slightly in comparison to the estimates obtained from the saturated model for all variables, while the estimate for $PAPM$ and $SpO2$ remains about the same. Hence, we do not reclassify these edges. Note that we apply proposition 4b) for checking $(APM, Temp)$ and (APM, CVP) here. Checking $(APM, Temp)$ using $cl(APM, Temp) = \{APM, Temp, HR, CVP, PAPM\}$ we also find this edge to be ambiguous. We do not get new separations in the third step so that the stepwise search finishes here.

Overall we have found evidence that, in addition to the results of the first step, the edge (APM, CVP) should be included in the partial correlation graph of the haemodynamic system due to a weak association. This agrees with medical knowl-

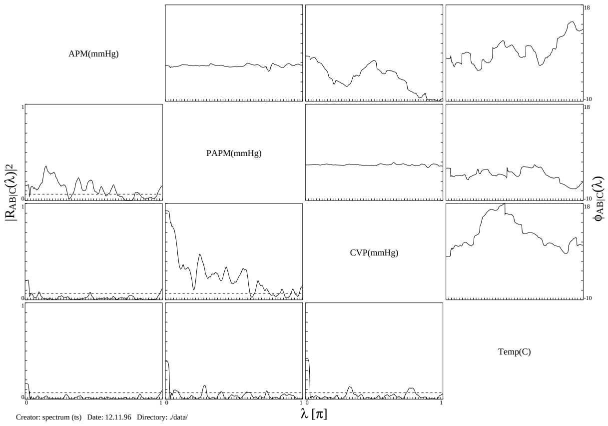

Figure 6: Estimated partial spectral coherencies (below the diagonal) within APM, $PAPM$, CVP and $Temp$. Here, the partial spectral coherency among APM and CVP is weakly significant.

edge since both APM and CVP measure blood pressure and their association is not expected to be mediated by the other variables in the analysis.

In order to examine the sensitivity of the previous results on the classification found in the first step we run the procedure again but start with the graph containing only the edges in (I) from the first step. We do not report the details of the analysis as the results essentially support those above. However, this analysis that starts with a sparser graph does indeed provide some evidence for edges in (A) being spurious. Discarding CVP and $PAPM$ does not increase the partial spectral coherencies among $\{HR, Temp, SpO2\}$. This can be regarded as indicator for some of the weak partial associations between either $Temp$ and $\{PAPM, CVP\}$ or $SpO2$ and $\{PAPM, CVP\}$ being spurious since $Temp$ and $SpO2$ should be expected to show an association when these mediating variables are omitted. At least the absence of the latter pair of associations is additionally supported by medical knowledge: There is no obvious reason why pulsoximetry $SpO2$ should be associated with any of the other variables. Omitting the edges $(SpO2, PAPM)$ and $(SpO2, CVP)$ does not affect any of the earlier results. We may therefore delete them from the final graph justifying this with the medical background knowledge.

Figure 7 depicts the final partial correlation graph derived from our stepwise search strategy. The graph reveals the associations among the different measures of blood pressures with the association among CVP and $PAPM$ and $PAPM$ and

APM being stronger than among APM and CVP. This agrees with the fact that CVP and $PAPM$ as well as $PAPM$ and APM are measured very close by each other, while the distance in the blood circuit between the measurement of APM and CVP is rather large. Moreover, we see a strong association among the heart rate and APM, while HR is not associated directly with $PAPM$ and CVP. This might be due to the current clinical state of the patient (cf. Gather et al., 2002). The weak associations between the blood temperature and all blood pressures agree with existing medical knowledge.

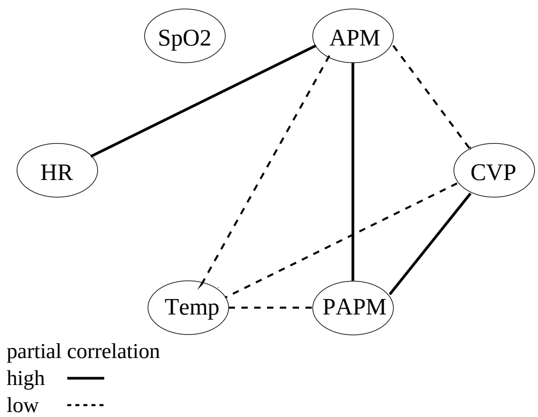

Figure 7: Partial correlation graph for the haemodynamic system derived from application of a stepwise search strategy.

6 Conclusion

We have derived conditions for decomposition and collapsibility of partial correlation graphs for multivariate time series. The results enable us to restrict the estimation of the partial spectral coherencies to subprocesses so that problems arising from missing data or zero restrictions for some of the partial spectral coherencies can be dealt with. These theoretical findings correspond to those well–known and important results for conditional independence graphs. Our results thus underpin the usefulness of the generalization of graphical models to dynamic data proposed by Dahlhaus (2000).

There are still some open problems regarding the general applicability of partial correlation graphs to multivariate time series. For model selection strategies based on stepwise elimination or deviance comparison, as commonly used in case of independent data, we lack appropriate methods in the time series context or would require further restrictive assumptions (Dahlhaus & Eichler, 2001). The results provided here, however, solve the problem of non–parametric estimation of the partial spectral coherencies under specific restrictions on the dependence structure.

As to the medical application considered in section 5, a clinical study (Gather et al., 2002) provides evidence that distinct clinical states like pulmonary hypertension, congestive heart failure or septic shock come along with distinct partial correlation structures. Therefore, we expect to gain new insights into the causes of clinical complications and the effects of medical interventions by application of graphical interaction models. Reliable strategies for model selection and model checking as presented here are very valuable in this regard.

ACKNOWLEDGEMENTS

We thank Rainer Dahlhaus and Michael Eichler for providing the program *Spectrum.* Financial support by the Deutsche Forschungsgemeinschaft (SFB 475, 'Reduction of complexity in multivariate data structures') is gratefully acknowledged.

REFERENCES

BRILLINGER, D.R. (1981). *Time Series. Data Analysis and Theory.* San Francisco: Holden Day.

BRILLINGER, D.R. (1996). Remarks concerning graphical models for time series and point processes. *Revista de Econometria* **16**, 1–23.

COWELL, R.G., DAWID, A.P., LAURITZEN, S.L. & SPIEGELHALTER, D.J. (1999). *Probabilistic Networks and Expert Systems.* New York: Springer.

COX, D.R. & WERMUTH, N. (1996). *Multivariate Dependencies.* London: Chapman & Hall.

DAHLHAUS, R. (2000). Graphical interaction models for multivariate time series. *Metrika* **51**, 157–172.

DAHLHAUS, R. & EICHLER, M. (2000). Spectrum. Computer program is available at http://www.statlab.uni–heidelberg.de/projects/graphical.models/

DAHLHAUS, R. & EICHLER, M. (2001). Causality and graphical models in time series analysis. Preprint, Department of Mathematics, University of Heidelberg, Germany.

DAWID, A.P. (1979). Conditional independence in statistical theory. *J. R. Statist. Soc.* Series B **41**, 1–31.

DIDELEZ, V. (2000). *Graphical Models for Event History Data based on Local Independence.* Berlin: Logos–Verlag.

DIDELEZ, V. & PIGEOT, I. (1999). Maximum likelihood estimation in graphical models with missing values. *Biometrika* **85**, 960–966.

EDWARDS, D. (2000). *Introduction to Graphical Modelling.* Second edition. New York: Springer.

EICHLER, M. (2000). *Graphical Models in Time Series Analysis.* Unpublished Ph.D–thesis, Department of Statistics, University of Heidelberg.

FRYDENBERG, M. (1990). Marginalization and collapsibility in graphical interaction models. *Annals of Statistics* **18**, 790–805.

GATHER, U., IMHOFF, M. & FRIED, R. (2002). Graphical models for multivariate time series from intensive care monitoring. *Statistics in Medicine*, to appear.

GRANGER, C.W.J. & HATANAKA, M. (1964). *Spectral Analysis of Economic Time Series.* Princeton: Princeton Press.

LAURITZEN, S.L. (1996). *Graphical Models.* Oxford: Clarendon Press.

SIMPSON, C.H. (1951). The interpretation of interaction in contingency tables. *J. R. Statist. Soc.* B **13**, 238–241

TIMMER, J., LAUK, M., HÄUSSLER, S., RADT, V., KÖSTER, B., HELLWIG, B., GUSCHLBAUER, B., LÜCKING, C.H., EICHLER, M. & DEUSCHL, G. (2000). Cross-spectral analysis of tremor time series. *Int. J. Bif. Chaos* **10**, 2595–2610.

WHITTAKER, J. (1990). *Graphical Models in Applied Multivariate Statistics.* Chichester: Wiley.