# Infering Graphical Models from Time Series

## Modelling, Simulating, and Inference of Complex Biological Systems

Speaker: Sebastian Petry

Institut of Statistics
Ludwig-Maximilians-University Munich

07-14-06



## Outline





## Outline





## Graphical Models

**The marriage (Quotation from Preface of [1])**

Graphical models are a marriage between probability theory and graph theory. They provide a natural tool for dealing with two problems that occur throughout applied mathematics and engineering — uncertainty and complexity — and in particular they are playing an increasingly important role in design and analysis of machine learning. Fundamental to the idea of graphical model is the notion of modularity — a complex system is built by combining simpler parts.



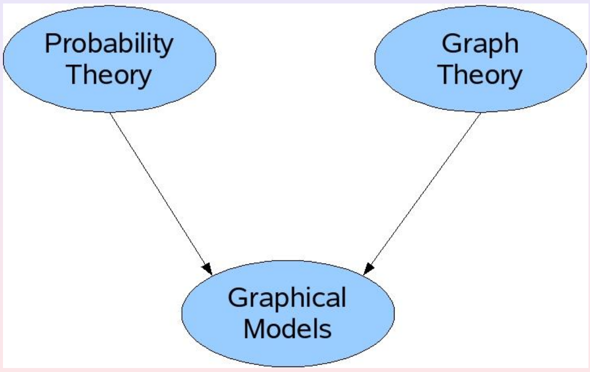
## Graphical Models

**The two parts of the marriage (Quotation from Preface of [1])**

*Probability theory* provides the glue whereby the parts are combined, ensuring that the systems as whole is consistent, and providing ways to interface models to data.

The *graph theoretic* side of graphical models provides both an intuitively appealing interface by which humans can model highly-interacting sets of variables as well as a data structure that lends itself naturally to the design of efficient general-purpose algorithms.








## Outline





## Outline

**Definition 1 (graph)**

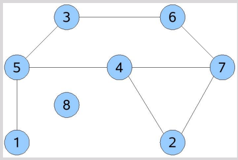
A *graph* $G$ is a tupel $G := (V, E)$ with a finite set $V \neq \emptyset$ and a subset $E \subseteq V \times V$ of two-elementic subsets of $V$. The elements of $V$ are called *vertices* and the elements of $E$ *edges*. The two vertex $v_i, v_j \in V$, $v_i \neq v_j$, of a edge $e = \{v_i, v_j\} \in E$ are called *end vertex* of $e$.

We call $v_i$ and $v_j$ with $e$ incident and $v_i$ and $v_j$ are adjacent.





**Example 2**




$$
\begin{aligned}
G &= (V, E)\\
V &= \{1, 2, 3, 4, 5, 6, 7, 8\}\\
E &= \{\{1, 5\}, \{5, 4\}, \{5, 3\},\\
&\quad \{4, 7\}, \{3, 6\}, \{2, 4\},\\
&\quad \{2, 7\}, \{6, 7\}\}
\end{aligned}
$$





**Definition 3 (digraph)**

A *directed graph* or *digraph* $G$ is a tupel $G := (V, E)$ with a finite set $V \neq \emptyset$ and a subset $E \subseteq V \times V$ of ordered paires $(v_i, v_j) \in V \times V$, $v_i \neq v_j$ and $\exists (v_i, v_j) \Rightarrow \nexists (v_j, v_i)$. The elements of $V$ are also called *vertices* and the elements of $E$ *edges* or *arcs*. For an arc $e = (v_i, v_j) \in E$ $v_i$ is called *tail* and $v_j$ called *head*.

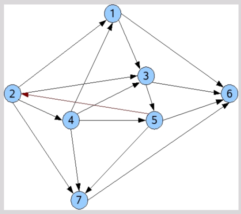
We call $v_i$ and $v_j$ with $e$ *incident* or $v_i$ and $v_j$ are *adjacent*.

We call $d_{in}(v)$ the number of arcs with head $v$ *indegree* and $d_{out}(v)$ the number of arcs with tail $v$ *outdegree*.





**Example 4**




$$
\begin{pmatrix}
0 & 0 & 1 & 0 & 0 & 1 & 0\\
1 & 0 & 1 & 1 & 0 & 0 & 1\\
0 & 0 & 0 & 0 & 1 & 1 & 0\\
1 & 0 & 1 & 0 & 1 & 0 & 1\\
0 & 1 & 0 & 0 & 0 & 1 & 1\\
0 & 0 & 0 & 0 & 0 & 0 & 0\\
0 & 0 & 0 & 0 & 0 & 1 & 0
\end{pmatrix}
$$

Adjacenty matrix: Tails in row and heads in column.

This digraph contains the directed cycle:

$$C := \{(2, 4), (4, 5), (5, 2)\}.$$





## Outline







**Definition 5 (dag)**

A digraph $G = (V, E)$ is called *directed acyclic graph (dag)*, iff there exists no sequence in each subset $E' \subseteq E$, with $E' = \{e'_1 = (v'_{1T}, v'_{1H}), ..., e'_n = (v'_{nT}, v'_{nH})\}$, $2 < n \leq |E|$, for which

- $v'_{1T} = v'_{nH}$,
- $v'_{iH} = v'_{(i+1)T}$, $\forall i = 2, ..., n-1$,

holds.





**Lemma 6**

*Let $G = (E, V)$ be a dag. Then there exists at least one vertex $v \in V$ for which $d_{in}(v) = 0$ holds.*

By Lemma 6 and Definition 5 the following theorem holds.

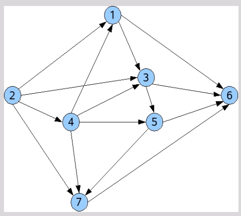
**Theorem 7**

*Every dag has a topological order.*

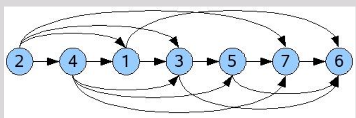




**Example 8 (A dag)**




$$
\begin{pmatrix}
0 & 0 & 1 & 0 & 0 & 1 & 0\\
1 & 0 & 1 & 1 & 0 & 0 & 1\\
0 & 0 & 0 & 0 & 1 & 1 & 0\\
1 & 0 & 1 & 0 & 1 & 0 & 1\\
0 & 0 & 0 & 0 & 0 & 1 & 1\\
0 & 0 & 0 & 0 & 0 & 0 & 0\\
0 & 0 & 0 & 0 & 0 & 1 & 0
\end{pmatrix}
$$



**Lemma 9**

*Every partial ordered set, i.e. $v_{iT} < v_{iH} \; \forall \, i$, can be embeded in a linear ordered set.*

In a lot of problems it is necessary to weight the edges. This leads to

**Definition 10 (network)**

Let $G = (V, E)$ a graph or digraph and $w : E \to \mathbb{R}$. Then a pair $(G, w)$ is called *network*.





## Outline







## Outline







## Basic Situation

Let $\mathbf{Y}^T = (Y_1, ..., Y_q)$, $\mathbf{X}^T = (X_1, ..., X_p)$, and $\mathbf{Z}^T = (Z_1, ..., Z_s)$ be random vectors with metric components. The random vector $(\mathbf{Y}^T, \mathbf{X}^T, \mathbf{Z}^T)^T$ possesses (if they exist) the mean

$$\mu = E((\mathbf{Y}^T, \mathbf{X}^T, \mathbf{Z}^T)^T) = (\mu_{\mathbf{Y}}^T, \mu_{\mathbf{X}}^T, \mu_{\mathbf{Z}}^T)^T$$

and the covariance(matrix)

$$\mathbf{\Sigma} = Cov((\mathbf{Y}^T, \mathbf{X}^T, \mathbf{Z}^T)^T) = \begin{pmatrix} \mathbf{\Sigma_{YY}} & \mathbf{\Sigma_{YX}} & \mathbf{\Sigma_{YZ}} \\ \mathbf{\Sigma_{XY}} & \mathbf{\Sigma_{XX}} & \mathbf{\Sigma_{XZ}} \\ \mathbf{\Sigma_{ZY}} & \mathbf{\Sigma_{ZX}} & \mathbf{\Sigma_{ZZ}} \end{pmatrix}.$$





## Three Kinds of Covariance

**We can distingush three kinds of covariance:**

1. marginal covariance
2. conditional covariance
3. partial covariance





## Marginal Covariance

**Definition 11 (marginal covariance)**

The *marginal covariance* between two random vectors $\mathbf{Y}$ and $\mathbf{X}$ is given by the submatrix

$$Cov(\mathbf{Y}, \mathbf{X}) = \mathbf{\Sigma_{YX}}$$

of $\mathbf{\Sigma}$.

**Theorem 12**

*$\mathbf{Y}$ and $\mathbf{X}$ are marginal uncorrelated, iff $\mathbf{\Sigma_{YX}} = \mathbf{0}$.*





## Conditional Covariance

**Definition 13 (conditional covariance)**

If the conditional density exists then the *conditional covariance* between $\mathbf{Y}$ and $\mathbf{X}$ given $\mathbf{Z} = \mathbf{z}$, $Cov(\mathbf{Y}, \mathbf{X}|\mathbf{Z} = \mathbf{z})$, is defined by the covariance of the conditional density of the random vector $\mathbf{Y}, \mathbf{X}|\mathbf{Z} = \mathbf{z}$.





## Partial Covariance

**Definition 14 (partial covariance)**

The *partial covariance* between $\mathbf{Y}$ and $\mathbf{X}$ given $\mathbf{Z}$, $Cov(\mathbf{Y}, \mathbf{X}|\mathbf{Z})$, is defined by

$$Cov(\mathbf{Y}, \mathbf{X}|\mathbf{Z}) = \mathbf{\Sigma_{YX}} - \mathbf{\Sigma_{YZ}}\mathbf{\Sigma_{ZZ}^{-1}}\mathbf{\Sigma_{XZ}^{T}} = \mathbf{\Sigma_{YX}} - \mathbf{\Sigma_{YZ}}\mathbf{\Sigma_{ZZ}^{-1}}\mathbf{\Sigma_{ZX}}. \quad (1)$$

## Something more about Covariances

**Theorem 15**

*Let* $\mathbf{\Sigma}^{-1} = \left( Cov\left(\left(\begin{pmatrix} \mathbf{Y} \\ \mathbf{X} \\ \mathbf{Z} \end{pmatrix}\right)\right)\right)^{-1} = \begin{pmatrix} \mathbf{C}^{\mathbf{YY}} & \mathbf{C}^{\mathbf{YX}} & \mathbf{C}^{\mathbf{YZ}} \\ \mathbf{C}^{\mathbf{XY}} & \mathbf{C}^{\mathbf{XX}} & \mathbf{C}^{\mathbf{XZ}} \\ \mathbf{C}^{\mathbf{ZY}} & \mathbf{C}^{\mathbf{ZX}} & \mathbf{C}^{\mathbf{ZZ}} \end{pmatrix}$ *the inverse of* $\mathbf{\Sigma}$*. Then* $Cov(\mathbf{Y}, \mathbf{X}|\mathbf{Z}) = \mathbf{0} \Leftrightarrow \mathbf{C}^{\mathbf{YX}} = \mathbf{0}$ *holds. Or in other words:* $\mathbf{Y}$ *and* $\mathbf{X}$ *given* $\mathbf{Z}$ *are partial uncorrelated, iff* $\mathbf{C}^{\mathbf{YX}} = \mathbf{0}$.

**Theorem 16**

*In the Gaussian case conditional and partial covariance are equivalent.*





## Outline







**The regression view of partial covariance**

Let $\mathbf{Y}$ be a $p$-dimensional random vector, $i$ and $j$ be two components of $\mathbf{Y}$, and $C$ the set of the remaining components of $\mathbf{Y}$ being not $i$ or $j$. Then the two regressions holds

$$\begin{aligned} Y_{i|C} &= Y_i - \mathbf{\Sigma}_{i,C}\mathbf{\Sigma}_{CC}^{-1}\mathbf{Y}_C, \\ Y_{j|C} &= Y_j - \mathbf{\Sigma}_{j,C}\mathbf{\Sigma}_{CC}^{-1}\mathbf{Y}_C, \end{aligned} \qquad (2)$$

where $\mathbf{\Sigma}_{i,C}\mathbf{\Sigma}_{CC}^{-1}$ are the submatrices of $\mathbf{\Sigma}$, because of $Cov(Y_{i|C}, Y_{j|C}) := Cov(Y_i, Y_j|\mathbf{Y}_C)$ (see (1)) holds.





**Interpretation and Comment**

1. $\mathbf{\Sigma}_{i,C}\mathbf{\Sigma}_{CC}^{-1}$ can be interpreted as a regression coefficient vector for $i$ given $C$.
2. Extention to the multivariate case $\mathbf{Y}_{i|C}$ is possible.

Intuitively it makes sense to avoid redundancies what leads to an orthogonal regression system.

In other words we are looking for a regression system of $Y_i$ on $Y_{r(i)}$ for $i = p-1, ..., 1$ with $r(i) = (i+1, ..., p)$. This defines a process of successive *orthogonalization*.





## Heuristic Illustration I ($p = 4$)

$$\epsilon = \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \epsilon_3 \\ \epsilon_4 \end{pmatrix} = \begin{pmatrix} 1 & -\beta_{1|2.34} & -\beta_{1|3.24} & -\beta_{1|4.23} \\ 0 & 1 & -\beta_{2|3.4} & -\beta_{2|4.3} \\ 0 & 0 & 1 & -\beta_{3|4} \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} Y_1 \\ Y_2 \\ Y_3 \\ Y_4 \end{pmatrix}$$

$$= \mathbf{AY}$$

$\beta_{i|k.C}$ denotes the coefficient of $Y_k$ in the regression of $Y_i$ on all variables after the conditioning sign. Computing the covariances matrix $Cov(\epsilon, \epsilon)$ leads because of independence to





## Heuristic Illustration II ($p = 4$)

$$\begin{aligned} Cov(\epsilon, \epsilon) &= E(\epsilon\epsilon^T) \\ &= diag(Var(\epsilon_i)) \\ &=: \Delta \\ \Delta &= E((\mathbf{AY})(\mathbf{AY})^T) \\ &= \mathbf{A}^T E(\mathbf{YY}^T)\mathbf{A} \\ &= \mathbf{A}^T \mathbf{\Sigma} \mathbf{A} \\ \Rightarrow \mathbf{\Sigma}^{-1} &= \mathbf{A}^T \Delta^{-1} \mathbf{A} \\ \text{and } \mathbf{\Sigma} &= \mathbf{A}^{-1} \Delta \mathbf{A}^{-T} \end{aligned}$$





## Heuristic Illustration III ($p = 4$)

By identifying (2) with the residual $\epsilon_i$ it follows

$$\begin{pmatrix} Y_{1|C} \\ Y_{2|C} \\ Y_{3|C} \\ Y_{4|C} \end{pmatrix} = \begin{pmatrix} 1 & -\beta_{1|2.34} & -\beta_{1|3.24} & -\beta_{1|4.23} \\ 0 & 1 & -\beta_{2|3.4} & -\beta_{2|4.3} \\ 0 & 0 & 1 & -\beta_{3|4} \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} Y_1 \\ Y_2 \\ Y_3 \\ Y_4 \end{pmatrix}$$

whereas the correspondending set $C$ for each $Y_{i|C}$ is given by the topological order of $\mathbf{A}$.





## Heuristic Illustration IV ($p = 4$)

The $\beta$-elements of $\mathbf{A}$ of each row of $\mathbf{A}$ are consequently given by

$$-a_{i,\,r(i)} = \mathbf{\Sigma}_{i,C}\mathbf{\Sigma}_{CC}^{-1}$$

and the diagonal elements of $\Delta$ by

$$\mathbf{\Sigma}_{i,i} - \mathbf{\Sigma}_{i,C}\mathbf{\Sigma}_{CC}^{-1}\mathbf{\Sigma}_{i,C}^T.$$

## Heuristic Illustration V ($p = 4$)

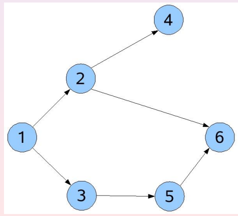
1. $\mathbf{A} = diag(1) - \mathbf{G}$: So $\mathbf{G}$ has the structure of a dag — or can interpreted as a network.
2. $\mathbf{\Sigma} = \mathbf{A}^{-1}\Delta\mathbf{A}^{-T}$: It is possible to induce the concentration matrices (and by inverting it the covariance matrix) by a network or a dag structure.
3. Under conditions you can get a dag network by knowing the concentration and the covariance matrix.



The following theorem holds:

**Theorem 17**

*Let $\mathbf{X}$ be an Gaussian random vector with covariance matrix $\mathbf{\Sigma}$. The variable $\mathbf{X}$, or the covariance matrix $\mathbf{\Sigma}$, is said to factorize in a dag, iff each component $X_i$ of $\mathbf{X}$ is independent from its nondescendant variables, given its parents.*








## Outline





**Definition 18 (multiple time series (MTS))**

Let $t \in \mathbb{Z}$ and $\mathbf{X}(t)^T = (X_1(t), ..., X_m(t))$, with $X_i(t) \in \mathbb{R}$, $\forall\, i \in \{1, ..., m\}$, be an $m$-dimensional random vector. Then the on $\mathbb{Z}$ ordered set $\mathbf{X} := \{\mathbf{X}(t)\}_{t\in\mathbb{Z}}$ is the random process and the realisation $\mathbf{x}$ of $\mathbf{X}$ is called *(multiple) time series (MTS)*.

**Definition 19 (autocovariance function (ACF))**

For a random process $\mathbf{X}$ the *autocovariance function (ACF)* is a matrix-valued mapping $\mathbb{Z} \times \mathbb{Z} \to \mathbb{R}^m \times \mathbb{R}^m$, $(t, s) \mapsto \mathbf{\Gamma}(t, s)$,

$$\mathbf{\Gamma}(t, s) = E[((\mathbf{X}(t) - E(\mathbf{X}(t)))(\mathbf{X}(s) - E(\mathbf{X}(s)))^T)].$$





**Definition 20 (stationarity)**

A random process $\mathbf{X}$ is called stationary, iff $\boldsymbol{\mu}(t) \equiv \boldsymbol{\mu}$[a] and the ACF only depends on $h = |t - s|$. So the ACF becomes a to matrix-valued mapping $\mathbb{Z} \to \mathbb{R}^m \times \mathbb{R}^m$, $h \mapsto \mathbf{\Gamma}(h)$,

$$\begin{aligned}\mathbf{\Gamma}(h) &= E(((\mathbf{X}(t+h) - \boldsymbol{\mu})(\mathbf{X}(t) - \boldsymbol{\mu})^T)) \\ &= E(((\mathbf{X}(h) - \boldsymbol{\mu})(\mathbf{X}(0) - \boldsymbol{\mu})^T)).\end{aligned}$$

[a] we only consider $\boldsymbol{\mu}(t) \equiv \mathbf{0}$



**Example 21 (ACF, $m = 3$ and $\boldsymbol{\mu} = \mathbf{0}$)**

$$\begin{aligned}\mathbf{\Gamma}(h) &= E(((\mathbf{X}(t+h))(\mathbf{X}(t))^T)) \\ &= E(\mathbf{X}(h)\mathbf{X}(0)^T) \\ &= E\begin{pmatrix} X_1(h)X_1(0) & X_1(h)X_2(0) & X_1(h)X_3(0) \\ X_2(h)X_1(0) & X_2(h)X_2(0) & X_2(h)X_3(0) \\ X_3(h)X_1(0) & X_3(h)X_2(0) & X_3(h)X_3(0) \end{pmatrix}\end{aligned}$$

**Lemma 22**

*For a stationary random process the ACF, $\mathbf{\Gamma}(h)$, is symetric and non-negativ.*

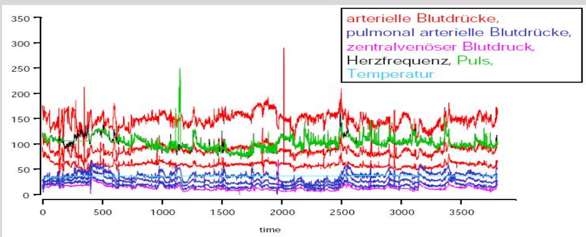




**Definition 23 (Gaussian time series)**

A time series is called *Gaussian*, iff all finite sets of marginals are jointly Gaussian.

**Example 24 (MTS)**






**Definition 25 (spectral density matrix (SDM))**

Given an $m$-dimensional stationary MTS for which $\sum_{h=-\infty}^{+\infty} ||\mathbf{\Gamma}(h)||_2 < +\infty$ holds. Then the spectral density matrix (SDM) is well defined as the matrix-valued mapping $\mathbb{R} \to \mathbb{R}^m \times \mathbb{R}^m$, $\omega \mapsto \mathbf{f}(\omega)$, with

$$\mathbf{f}(\omega) = \frac{1}{2\pi}\sum_{h=-\infty}^{+\infty} \mathbf{\Gamma}(h)\exp(-ih\omega)$$

For SDM the following holds:

1. $\mathbf{f}(\omega)$ is Hermitian for each $\omega \in \mathbb{R}$.
2. $\omega \mapsto \mathbf{f}(\omega)$ is $2\pi$-periodic.
3. for real-valued vectors $\mathbf{f}(\omega) = \mathbf{f}(-\omega)$ holds.
4. $\mathbf{\Gamma}(h) = \int_0^{2\pi} \mathbf{f}(\omega) \exp(ih\omega) d\omega$.





**Definition 26 (sample autocovariance function (SACF))**

The sample autocovariance function (SACF) is defined as

$$\hat{\mathbf{\Gamma}}(h) = \frac{1}{T} \sum_{t=0}^{T-h-1} (\mathbf{x}(t+h) - \bar{\mathbf{x}})(\mathbf{x}(t) - \bar{\mathbf{x}})^T, \; h \in [0, \; T-1]$$

with $\bar{\mathbf{x}} = \frac{1}{T} \sum_{t=0}^{T-1} \mathbf{x}(t)$ as the sample mean of data.

**Lemma 27**

*The SACF is a consistent estimator for the ACF and under weak assumtions normal distributed.*





**Definition 28 (periodogram)**

Let $\mathbf{d}(k) = \frac{1}{\sqrt{T}} \sum_{t=0}^{T-1} \mathbf{x}(t) \exp(-ikt)$ the discrete Fourier transform of data. At each frequencies $\omega_k = \frac{2\pi k}{T}$, $\omega \in [0, 2\pi)$ especially $k \in \{1, ..., T\}$, the *periodogram* is defined as

$$\mathbf{I}(\omega_k) = \frac{1}{2\pi} \mathbf{d}(k)\mathbf{d}(k)^*.$$





**Lemma 29**

*The periodogram does not provide a consistent estimator of SDM. Smoothing can solve this problem:*

$$\hat{\mathbf{f}}^r(\omega_k) = \sum_{j=-\infty}^{+\infty} \mathbf{W}_r(j)\mathbf{I}(\omega_{j+k}),$$

*with the weight-matrix $\mathbf{W}_r(.)$ and $r$ as smoothing parameter.*

**Comment**

A skilled choice of the weight-matrix $\mathbf{W}_r(.)$ makes possible to optimize the periodogram $\hat{\mathbf{f}}^r(\omega_k)$ by $r$. For example use $\mathbf{W}_r(j) = \frac{1}{r\sqrt{2\pi}} \exp(\frac{-j^2}{2r^2})$ and minimize AIC with the Whittle approximation of the likelihood by $r$. (For details see [2])





## Outline







**Theorem 30**

*A Gaussian stationary stochastic process with an absolutely summable ACF has the spectral representation*

$$\mathbf{X}(t) = \int_0^{2\pi} \exp(it\omega) d\mathbf{Z}(\omega),$$

*where $\mathbf{Z}(\omega)$ is a random process with orthogonal increments such that $\omega_1 < \omega_2$, $Cov(\mathbf{Z}(\omega_2) - \mathbf{Z}(\omega_1)) = \int_{\omega_1}^{\omega_2} \mathbf{f}(\omega) d\omega$. In other words: $\mathbf{X}(t)$ is a superposition of infinite many independent random signals at different frequencies.*





The previous theorem 30 in symbiosis with the property that for $X \sim N(\mu, \sigma^2)$

$$E(exp(itX)) = exp\left(i\mu t - \frac{1}{2}\sigma^2 t^2\right)$$

holds, makes possible to use the theory of Gaussian random vector on Gaussian stationary MTS by replacing the covariance matrix with the SDM or better periodogram.





Now we apply the theory of Gaussian variables to Gaussian stationary MTS and it follows

**Lemma 31**

*Two time series $x_i$ and $x_j$ are marginally independent iff*

$$\forall \omega \in [0, 2\pi), \quad f_{ij}(\omega) = 0.$$

*The time series $x_i$ and $x_j$ are partially and therefore conditionally independent given all other times series $x_k$, $k \neq i, j$ iff*

$$\forall \omega \in [0, 2\pi), \quad (f(\omega)^{-1})_{ij} = 0.$$

## Outline







- Casting the structure learning as a model selection. The structure is a dag.
- Minimizing the AIC score (3) that is recovered by entropy rates and KL divergence.
- For details see [2].





$$J(G) = \sum_{i=1}^{m} J_i(\pi_i(G)) \tag{3}$$

where the local score is

$$J_i(\pi_i(G)) = \frac{-T}{4\pi}\int_0^{2\pi} \log \frac{\det(\hat{\mathbf{f}}_{\{i\}\cup\pi_i}(\omega))}{\det(\hat{\mathbf{f}}_{\pi_i}(\omega))} d\omega + (2|\pi| + 1)\frac{df}{2}. \tag{4}$$

(4) is approximated using the samples of $\hat{f}(\omega)$ as

$$J_i(\pi_i(G)) = \frac{-T}{2H}\sum_{k=0}^{H-1} \log \frac{\det((\mathbf{f}_k)_{\{i\}\cup\pi_i})}{\det((\mathbf{f}_k)_{\pi_i})} + (2|\pi| + 1)\frac{df}{2}. \tag{5}$$





## Comments to AIC

- We learn the structure of $G$ by minimize the AIC $J(G)$. This problem is numerically complex. Using the greedy algorithm and other mathematical tricks can lead to an efficient solving procedures. The problem is NP-complete!
- Often it can be convenient to restrict the number of parents. The dag structure makes this possible.
- One of the major gains from learing a spare structure for the SDM is that we can perform and optimize the smoothing perodogram locally on cliques of $G$. The AIC score is given on the next slide.





$$\begin{aligned} J(G, r) &= \frac{T}{2H}\sum_{k=0}^{H-1}\sum_{i=0}^{m}\left(\frac{\det((\mathbf{f}_k^{r_i})_{i\cup\pi_i})}{\det((\mathbf{f}_k^{r_i})_{\pi_i})} + tr\left\{(\mathbf{f}_k^{r_i})^{-1}_{\{i\}\cup\pi_i})\mathbf{I}_{\{i\}\cup\pi_i}(\omega_k)\right\}\right. \\ &\quad \left. + tr\left\{(\mathbf{f}_k^{r_i})^{-1}_{\pi_i})\mathbf{I}_{\pi_i}(\omega_k)\right\}\right) + \sum_{i=1}^{m}(2|\pi_i| + 1)\frac{df_i}{2}. \end{aligned}$$





## Outline







- Graphical models are special cases of networks.
- Applications are possible in different ways.
- Graphs and Topology
  - A dag is a topological orderable digraphs.
  - Strictly triangular matrices have the structure of a dag
- Multivariate Random Vectors and Times Series
  - The partial covariance has the structure of a dag
  - Using knowlegde about random vectors on MTS is in the Gaussian case possible by replacing the covariance matrix by SDM or periodogram
  - Possible to define the independence of MTS
- Searching the structure of independence MTS by using the knowledge of dags and probability theory





Some references

- Jordan, M.I. (1999). *Learing in Graphical Models*, MIT Press.
- Francis R., Bach and Jordan, M.I. (2004). *Learing in Graphical Models for Stationary Time Series*, IEEE Transactions on Signal Processing, Vol52, No.8, August 2004.
- Jungnickel, D. (1987). Graphen, Netzwerke und Algorithmen, B.I.-Wissenschaftsverlag.
- Wermuth, N., Cox, D.R., Marchetti, G.M.. *Covariance chains*.
- Murphy, K.P. (2001). *An introduction to graphical models*.