



HAL
open science

Gaussian stationary processes over graphs, general frame and maximum likelihood identification

Thibault Espinasse, Fabrice Gamboa, Jean-Michel Loubes

► **To cite this version:**

Thibault Espinasse, Fabrice Gamboa, Jean-Michel Loubes. Gaussian stationary processes over graphs, general frame and maximum likelihood identification. 2011. hal-00586905v1

HAL Id: hal-00586905

<https://hal.science/hal-00586905v1>

Submitted on 18 Apr 2011 (v1), last revised 21 Mar 2012 (v3)

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Gaussian stationary processes over graphs, general frame and maximum likelihood identification

T. Espinasse, F. Gamboa and J-M. Loubes

April 19, 2011

Abstract

In this paper, we give a general construction of stationary Gaussian processes indexed on graphs. This construction relies on spectral theory of Hilbertian operators defined on a graph. We then extend natural maximum likelihood estimators of the parameters of the corresponding spectral density and provide their asymptotic behaviour.

Contents

1	Definitions and useful properties for spectral analysis and Toeplitz operators	3
2	Analytic construction	8
3	Convergence of approximate likelihood maximum estimators	9
4	Algebraical construction of covariance operator on a graph	17
5	Homogeneity and spectral measure	21
6	Simulations	23
7	Appendix	24
	7.1 Szegő Lemmas	24
	7.2 Proofs of theorem 3.1	29
	7.3 Proof of the technical lemmas	30

Introduction

In the past few years, many interest has been paid to the study of random fields over graphs, driven by the growing needs for both theoretical and practical results for data defined on graphs. On one side, the definition of graphical models by J.N. Darroch, S.L. Lauritzen and T.P. Speed in 1980 [7] introduced a new regard on Markov fields, and many tools have been developed in this direction (see, for instance [18] and [19]). On another side, the industrial demand linked to graphical problems had rose up with the apparition of new technologies. In very particular the Internet and social networks provide a huge field of applications, but biology, economy, geography or image analysis benefit also from models taking into account a graph structure.

The analysis of road traffic is at the root of this work. Actually, prediction of road traffic deals the forecast of the speed of the vehicles which may be seen as a spatial random fields over the traffic network. If some work has been done without taking into account the particular graph structure of the speed process, see in particular [8] and [14] for the related statistical issues, in this paper we build a model of a class of Gaussian random fields over graphs and study the particular properties of such stochastic process.

A random field over graph is a spatial process indexed by the vertices of a graph, namely $X_i, i \in G$. Many models already exist in the probabilist literature, ranging from Markov fields to autoregressive processes, which are based on two general kinds of construction. On the one hand, graphical models following the ideas of Markov fields described for instance in [17], are built by specifying a dependency structure for X_i and X_j , conditionally to the other variables, as soon as the locations i and j are neighbors. This particular dependency structure must be chosen by the statistician. We refer for instance to [7] and references therein. On the other hand, the graph itself, through the adjacency operator, can provide the dependency. This is the case for autoregressive models (see [11]) using strongly the local form of the graph for statistical inference.

We will provide in this paper a new framework taking into account both advantages of these two points of views. Hence, we extend some classical results from time series to spatial fields over general graph, and we will tackle the issue of building ARMA processes on graphs and provide some conditions on the graph. For this, we will make use of spectral analysis of time series and extend to this framework classical results such as the problem of identification of the maximum likelihood estimator, through an approximation of the Gaussian contrast called the Whittle's approximation and usual Szegö Lemmas. This will enable us to construct a maximum

likelihood estimate for the parameters of the spectral density of processes on the graph that we will define. We point out that we will compare throughout all the paper our new framework with the processes indexed by \mathbb{Z}^d with $d \geq 1$.

The paper falls into the following parts. Section 1 is devoted to recall some definitions of graphs and spectral theory for time series. Then we state the definition of general spatial *ARMA* processes over a graph in Section 2, and the main theorem is given in Section 3. Section 4 provides some way to modify the graph himself, in order to consider a larger class of processes, which leads to the construction of a large class of Markov fields. Section 5 gives the definition of the spectral measure for the graph, and a sufficient condition for its existence. Some simulations are provided in Section 6. The last section provides all the tools necessary to prove the theorem, in particular Szegő Lemmas are given in Section 7.1, while the proofs of the technical Lemmas are postponed in Section 7.3.

1 Definitions and useful properties for spectral analysis and Toeplitz operators

In the whole paper, we will consider a Gaussian spatial process $(X_i)_{i \in G}$ indexed by the vertices of an infinite undirected weighted graph.

We will call $\mathbf{G} = (G, W)$ this graph, where

- G is the set of vertex. \mathbf{G} is said to be infinite as soon as G is infinite (but countable).
- $W \in [-1, 1]^{G \times G}$ is the symmetric weighted adjacency operator. That is, $1 \geq |W_{ij}| \neq 0$ when $i \in G$ and $j \in G$ are connected.

We chose W symmetric ($W_{ij} = W_{ji}$, $i, j \in G$) since we deal only with undirected graphs.

For any vertex $i \in G$, a vertex $j \in G$ is said to be a neighbor of i if and only if $W_{ij} \neq 0$. The degree $\deg(i)$ of i is the number of neighbors, and the degree of the graph \mathbf{G} is defined as the maximum degree of the vertices of the graph \mathbf{G} :

$$\deg(G) := \max_{i \in G} \deg(i).$$

From now, assume the degree of the graph \mathbf{G} is bounded :

$$\deg(\mathbf{G}) < +\infty.$$

When $G = \mathbb{Z}$, we will use the renormalized adjacency operator $W_{ij}^{(\mathbb{Z})} = \frac{1}{2} \mathbb{1}_{\{|i-j|=1\}}$, $(i, j \in \mathbb{Z})$. Here, $\deg(\mathbb{Z}) = 2$. This example will be used in all the paper to illustrate our talk.

Assuming that the entries of W belong to $[-1, 1]$ is not restrictive since renormalizing the adjacency operator does not change the objects introduced later. In particular, the spectral representation of Hilbertian operator is not sensitive to a renormalization, and it is a key tool used in the whole paper.

To introduce the spectral decomposition, consider the action of the adjacency operator on $l^2(G)$ as

$$\forall u \in l^2(G), (Wu)_i := \sum_{j \in G} W_{ij} u_j, (i \in G).$$

Notice that as the degree of G and the entries of W are both bounded, W lies in B_G the set of all bounded Hilbertian operator on $l^2(G)$. B_G is endowed with the classical operator norm

$$\forall A \in B_G, \|A\|_{2,op} := \sup_{u \in l^2(G), \|u\|_2} \|Au\|_2,$$

where $\|\cdot\|_2$ stands for the usual norm on $l^2(G)$.

Recall that for any bounded Hilbertian operator $A \in B_G$, the spectrum $\text{Sp}(A)$ is defined as the set of complex numbers λ where $\lambda \text{Id} - A$ is not invertible, where Id stands for the identity of B_G . Since W is bounded and self-adjoint, $\text{Sp}(W)$ is a compact non-empty subset of \mathbb{R} .

We aim now at providing a spectral representation of any bounded normal Hilbertian operator. For this, recall first the definition of a resolution of identity E :

Definition 1.1. *Let \mathcal{M} be a σ -algebra over a set Ω . We call identity resolution (on \mathcal{M}) an application*

$$E : \mathcal{M} \rightarrow B_G$$

such that, $\forall \omega, \omega' \in \mathcal{M}$,

1. $E(\emptyset) = 0, E(\Omega) = I$
2. Every $E(\omega)$ is a self-adjoint projection operator.
3. $E(\omega \cap \omega') = E(\omega)E(\omega')$
4. If $\omega \cap \omega' = \emptyset$, then $E(\omega \cup \omega') = E(\omega) + E(\omega')$

5. For all $x \in l^2(G)$ and $y \in l^2(G)$, the functional $E_{x,y}$ defined by

$$E_{xy}(\omega) = \langle E(\omega)x, y \rangle_{l^2(G)}$$

is a complex measure on \mathcal{M} .

We can now state the fundamental theorem

Theorem 1.1 (Spectral decomposition). *If $A \in B_G$ is normal (A commutes with its adjoint), then there exists a unique identity resolution E over all Borelian subsets of the spectrum of A denoted $\text{Sp}(A)$, such that*

$$A = \int_{\text{Sp}(A)} \lambda dE(\lambda)$$

Moreover, for any $U \in B_S$ such that $UA = AU$, every projector $E(\omega), \omega \in \text{Sp}(A)$ commutes with U

Since W is self-adjoint in our case, it is a normal operator, so Theorem 1.1 may be applied.

We obtain the spectral representation of the adjacency operator W thanks to the identity resolution E over $\text{Sp}(W)$

$$W = \int_{\text{Sp}(W)} \lambda dE(\lambda).$$

Moreover, with this decomposition, we can give a spectral representation for the powers of W , $W^k, k \in \mathbb{Z}$. Define first, for any $i \in G$, the sequences δ_i in $l^2(G)$ by

$$\delta_i := (\mathbb{1}_{k=i})_{k \in G}.$$

For any $i, j \in G$, the sequences δ_i and δ_j define a real measure μ_{ij} by

$$\forall \omega \subset \text{Sp}(W), \mu_{ij}(\omega) := E_{\delta_i, \delta_j}(\omega) = \langle E(\omega)\delta_i, \delta_j \rangle_{l^2(G)}.$$

Hence, we can write :

$$\forall k \in \mathbb{N}, \forall i, j \in G, (W^k)_{ij} = \int_{\text{Sp}(W)} \lambda^k d\mu_{ij}.$$

In the usual case of \mathbb{Z} , an explicit expression for μ_{ij} can be given. Denote $T_k(X)$ the k^{th} -Tchebychev polynomial, $k \in \mathbb{N}$. We can provide the spectral decomposition of $W^{(\mathbb{Z})}$:

$$\forall i, j \in \mathbb{Z}, \left((W^{(\mathbb{Z})})^k \right)_{ij} = \int_{[-1,1]} \lambda^k \frac{T_{|j-i|}(\lambda)}{\sqrt{1-\lambda^2}} d\lambda$$

This shows that, in this case, for any $i, j \in G$, the measure $d\mu_{ij}$ is absolutely continuous with respect to the Lebesgue measure, and its density is given by

$$\frac{d\mu_{ij}}{d\lambda} = \frac{T_{|j-i|}(\lambda)}{\sqrt{1-\lambda^2}}$$

Note that we recover the usual spectral decomposition by setting :

$$\forall i, j \in G, d\hat{\mu}_{ij}(t) := \frac{1}{2\pi} \cos((j-i)t) dt.$$

We get

$$\forall i, j \in \mathbb{Z}, \left((W^{(\mathbb{Z})})^k \right)_{ij} = \int_{[0, 2\pi]} \cos(t)^k d\hat{\mu}_{ij}(t)$$

Note that this corresponds to choosing another identity resolution, given by the family of measure $(\hat{\mu}_{ij})_{i,j \in G}$. Further on, this enables to handle the usual case of processes indexed by \mathbb{Z} . Since our aim is to study some kind of stationary processes indexed by the vertex G of the graph \mathbf{G} , recall once again what happen for the usual case of \mathbb{Z} .

Let $\mathbf{X} = (X_i)_{i \in \mathbb{Z}}$ be a stationary Gaussian process indexed by \mathbb{Z} . Since X is Gaussian, stationarity is equivalent to second order stationarity, that is, $\forall i, k \in \mathbb{Z}, \text{Cov}(X_i, X_{i+k})$ does not depend on i . Thus, we can define

$$r_k := \text{Cov}(X_i, X_{i+k})$$

In this example, we assume that $(r_k)_{k \in \mathbb{Z}} \in l^1(\mathbb{Z})$. This leads to a particular form of the covariance operator K defined on $l^2(\mathbb{Z})$ by

$$\forall i, j \in \mathbb{Z}, K_{ij} := r_{i-j}.$$

Recall that $B_{\mathbb{Z}}$ denotes here the set of bounded Hilbertian operators on $l^2(\mathbb{Z})$. Notice that, since $(r_k)_{k \in \mathbb{Z}} \in l^1(\mathbb{Z})$, $K \in B_{\mathbb{Z}}$ (see for instance [6] for more details). This bounded operator is constant over each diagonals, and so called a Toeplitz operator (see also [5] for a general introduction to Toeplitz operators).

Toeplitz operators enjoy the following representation,

$$\forall i, j \in \mathbb{Z}, T(f)_{ij} := K_{ij} = \frac{1}{2\pi} \int_{[0, 2\pi]} f(t) \cos((i-j)t) dt,$$

where f is the spectral density of the process \mathbf{X} , defined by

$$f(t) := 2 \sum_{k \in \mathbb{N}^*} r_k \cos(kt) + r_0$$

This expression can be written, using the Tchebychev polynomials $(T_k)_{k \in \mathbb{N}}$,

$$f(t) := 2 \sum_{k \in \mathbb{N}^*} r_k T_k(\cos(t)) + r_0 T_0(\cos(t)).$$

This falls in the framework of Theorem 3.1, by setting

$$\forall \lambda \in [-1, 1], g(\lambda) := 2 \sum_{k \in \mathbb{N}^*} r_k T_k(\lambda) + r_0 T_0(\lambda).$$

We get, using the family $(\hat{\mu}_{ij})_{i,j \in \mathbb{Z}}$ defined above,

$$\forall i, j \in \mathbb{Z}, K_{ij} = \int_{[0, 2\pi]} g(\cos(t)) d\hat{\mu}_{ij}(t).$$

Notice that the last expression may also be written $K = g(W^{(\mathbb{Z})})$, and the convergence is ensured by the boundedness of $W^{(\mathbb{Z})}$ together with the boundedness of the Tchebychev polynomials, $T_k([-1, 1]) \subset [-1, 1], \forall k \in \mathbb{Z}$, and the summability of the sequence $(r_k)_{k \in \mathbb{Z}}$.

This remark is fundamental for the construction we provide in the following section, Definition 2.1.

Finally, we will extend usual MA processes to any graph, using this theory. This will be the purpose of Section 2.

Recall here some properties about the moving average representation MA_∞ of the process, which exists as soon as $\log(f)$ is integrable. In this case, there exists a sequence $(a_k)_{k \in \mathbb{N}}$, with $a_0 = 1$, and a white noise $\epsilon = (\epsilon_k)_{k \in \mathbb{Z}}$, such that the process \mathbf{X} may be written as

$$\forall i \in \mathbb{Z}, X_i = \sum_{k \in \mathbb{N}} a_k \epsilon_{i-k}$$

Defining the function h over the unit circle \mathcal{C} by

$$\forall x \in \mathcal{C}, h(x) = \sum_{k \in \mathbb{N}} a_k x^k,$$

we recover, with a few computations, the spectral decomposition of the covariance operator K of \mathbf{X} :

$$\forall i, j \in \mathbb{Z}, K_{ij} = \int_{[0, 2\pi]} |h(e^{it})|^2 d\hat{\mu}_{ij}(t),$$

which shows the equality

$$g(\cos(t)) = |h(e^{it})|^2$$

Recall that when h is a polynomial of degree p , the process is said to be MA_p . In this case, g is also a polynomial of degree k . Reciprocally, if g is a real polynomial of degree p , and as soon as $g(\cos(t))$ is even, and non-negative for any $t \in [0, 2\pi]$, the Fejer-Riesz theorem provides a factorization of $g(\cos(t))$ such that $g(\cos(t)) = |h(e^{it})|^2$ (see for instance [13]). This proves that \mathbf{X} is MA_p if, and only if, its covariance operator may be written $g(A)$, with g is polynomial of degree p .

2 Analytic construction

In this section, we will define moving average and autoregressive processes over the graph \mathbf{G} . In the whole section, we deal with the adjacency weighted operator W of the graph \mathbf{G} .

As explained in the last section, since W is bounded and self-adjoint, $\text{Sp}(W)$ is a non-empty compact subset of \mathbb{R} , which admits a spectral decomposition thanks to an identity resolution E , given by

$$W = \int_{\text{Sp}(W)} \lambda dE(\lambda).$$

We define here MA and AR Gaussian processes, with respect to the operator W , by defining the correspondent classes of covariance operators, since the covariance characterized fully any Gaussian process.

Definition 2.1. *Let $(X_i)_{i \in G}$ be a Gaussian process, indexed by the vertices G of the graph \mathbf{G} , and K its covariance operator.*

If there exists an analytic function g defined on the convex hull of $\text{Sp}(W)$, such that

$$K = \int_{\text{Sp}(W)} g(\lambda) dE(\lambda),$$

we will say that X is

- $MA_q^{(W)}$ if g is a polynomial of degree q .
- $AR_p^{(W)}$ if $\frac{1}{g}$ is a polynomial of degree p which has no root in the convex hull of $\text{Sp}(W)$.
- $ARMA_{p,q}^{(W)}$ if $g = \frac{P}{Q}$ with P a polynomial of degree p and Q a polynomial of degree q with no roots in the convex hull of $\text{Sp}(W)$.

Otherwise, we will talk about the $MA_\infty^{(W)}$ representation of the process \mathbf{X} .

Notice that, in the usual case of \mathbb{Z} , and for finite order *ARMA*, we recover the usual definition as shown in the end of the previous section. So, the last definition may be seen as an extension of isotropic *ARMA* for any graph \mathbf{G} .

This kind of modeling is interesting when the interactions are propagated time by time and locally (that may be for instance a good modeling for traffic problems).

Now, we get a representation of moving average processes over any graph \mathbf{G} . The following section gives the main result of this paper, that is the maximum likelihood identification, and section 4 proposes a general definition of stationary processes indexed by the vertices of a graph, and shows the stationarity of *MA* representation thanks to this definition.

3 Convergence of approximate likelihood maximum estimators

In this section let $\mathbf{G} = (G, W)$ be a graph with bounded degree ($\deg(\mathbf{G}) < +\infty$). Let also $(X_i)_{i \in G}$ be a Gaussian spatial process indexed by the vertices of \mathbf{G} with spectral density f_{θ_0} depending of an unknown parameter $\theta_0 \in \Theta$. We aim at estimating θ_0 . For this, we will generalize maximum likelihood estimation for parameters of time series.

We will also develop a Whittle approximation for *ARMA* processes indexed by the vertices of a graph. That is an approximation of the likelihood which provides the convergence of maximum likelihood estimate. We follow here the guidelines of the proof given in [2] for the usual case of time series.

Let us now specify the framework of our study. Let $(\mathbf{G}_n)_{n \in \mathbb{N}}$ be a growing sequence of nested subgraphs, which means that if $\mathbf{G}_n = (G_n, W_n)$, we have $G_n \subset G_{n+1} \subset G$ and for any $i, j \in G_n$, it holds that $W_n(i, j) = W(i, j)$.

We will use the following notations. Let $m_n = \text{Card}(G_n)$. We set also $\delta_n = \text{Card} \{i \in G_n, \exists j \in G \setminus G_n, W_{ij} \neq 0\}$.

The sequence $(m_n)_{n \in \mathbb{Z}}$ may actually be seen as the ‘‘volume’’ of the graph \mathbf{G}_n , and δ_n as the size of the boundary of G_n . For the special case $G = \mathbb{Z}^d$ and $G_n = [-n, n]^d$, we get $m_n = (2n + 1)^d$ and $\delta_n = 2d(2n + 1)^{d-1}$.

The ratio $\frac{\delta_n}{m_n}$ is a natural quantity associated to the expansion of the graph which also appears isoperimetrical [16] and expander issues. We will assume here that this ratio tends to 0 with the size of the graph, and also that, even W has to be renormalized.

Assumption 3.1.

- $\sup_{i,j \in S} W_{ij} \leq \frac{1}{D}$

- $\delta_n = o(m_n)$

Notice here that the first assumption implies that $\text{Sp}(W) \subset [-1, 1]$. The second assumption is a non-expansion criteria which is satisfied for the last examples $G = \mathbb{Z}^d$ and $G_n = [-n, n]^d$, but not for an homogeneous tree, whatever the choice of the sequence of subgraphs $(\mathbf{G}_n)_{n \in \mathbb{N}}$.

We will now chose a parametric family of covariance operators of $MA^{(W)}$ processes as defined in the last section. First, let Θ be a compact subspace of \mathbb{R} .

We point out that for sake of simplicity, we choose a one-dimensional parameter space Θ . Nevertheless, all the results could be extend to the case $\Theta \subset \mathbb{R}^k, k \geq 1$.

Define \mathcal{F} as the set of analytic functions over the convex hull of $\text{Sp}(W)$.

Let also be a $(f_\theta)_{\theta \in \Theta}$ be a parametric family of functions of \mathcal{F} . They define a parametric set of covariances on G by

$$K(f_\theta) = f_\theta(W).$$

As in [2], we will need a strong regularity for this family of spectral densities.

Let us introduce a regularity factor for any analytic function $f \in \mathcal{F}$, $f(x) = \sum_k f_k x^k (x \in \text{Sp}(W))$, with

$$\alpha(f) := \sum_{h \in \mathbb{N}} |f_h| (h + 1).$$

Define further,

$$\mathcal{F}_\rho := \{f \in \mathcal{F}, \alpha(\log(f)) \leq \rho\}, \rho > 0.$$

Notice that for any $f \in \mathcal{F}_\rho$, we have $\alpha(f) \leq e^\rho, \alpha(\frac{1}{f}) \leq e^\rho$.

Assumption 3.2.

- *The application $\theta \rightarrow f_\theta$ is injective.*
- *For any $\lambda \in \text{Sp}(W)$, the application $\theta \rightarrow f_\theta(\lambda)$ is continuous.*
- $\forall \theta \in \Theta, f_\theta \in \mathcal{F}_\rho$.

From now, fix $\theta_0 \in \Theta$. Let \mathbf{X} be a centered Gaussian MA_∞ process over \mathbf{G} with covariance operator $K(f_{\theta_0})$ (see section 2).

We observe the restriction of this process on the subgraph \mathbf{G}_n defined before. Our aim is to compute the maximum likelihood estimator of θ_0 . Let $X_n = (\mathbf{X}_i)_{i \in G_n}$ be the observed process and $K_n(f_\theta)$ be its covariance :

$$X_n \sim \mathcal{N}(0, K_n(f_{\theta_0})).$$

The corresponding log-likelihood at θ is

$$L_n(\theta) := -\frac{1}{2} \left(m_n \log(2\pi) + \log \det (K_n(f_\theta)) + X_n^T (K_n(f_\theta))^{-1} X_n \right).$$

Recall what happens in the usual case of time series ($G = \mathbb{Z}$, $G_n = [-n, n]$). Using the notations of Section 1, we assume that \mathbf{X} is a stationary process indexed by \mathbb{Z} with spectral density f , such that $\log(f)$ is integrable. Recall that, if g is such that $f(t) = g(\cos(t))$, we have

$$K(g) = g(W^{(\mathbb{Z})}).$$

It is usual to maximize an approximation of the likelihood. The classical approximation is Whittle's one ([9]), where the $\log \det (K_n(f_\theta))$ is replaced by

$$\frac{1}{2\pi} \int_{[0, 2\pi]} \log (g(\cos(t))) dt.$$

Back to the general case, we aim at performing the same kind of approximation. So that, we will need the following assumption to ensure the convergence of $\log \det (K_n(f_\theta))$ (see Section 2 for the definition of μ_{xx}) :

Assumption 3.3. *There exists a positive measure μ , such that*

$$\frac{1}{m_n} \sum_{x \in G_n} \mu_{xx} \xrightarrow{\mathcal{D}} \mu,$$

as n goes to infinity.

Here, \mathcal{D} stands for the convergence in distribution

The limit measure μ is classically called spectral measure of \mathbf{G} with respect to the sequence of subgraph \mathbf{G}_n (see [15] for example). In Section 5, we give some sufficient condition to ensure the existence of the spectral measure.

As in the case of time series (for $G = \mathbb{Z}$), we can approximate the log-likelihood. It avoids an inversion of matrix and a computation of a determinant. Indeed, we will consider the two following approximations.

$$\bar{L}_n(\theta) := -\frac{1}{2} \left(m_n \log(2\pi) + m_n \int \log(f_\theta(x)) d\mu(x) + X_n^T (K_n(f_\theta))^{-1} X_n \right)$$

$$\tilde{L}_n(\theta) := -\frac{1}{2} \left(m_n \log(2\pi) + m_n \int \log(f_\theta(x)) d\mu(x) + X_n^T \left(K_n \left(\frac{1}{f_\theta} \right) \right) X_n \right)$$

Generally, approximate maximum likelihood estimators are not asymptotically normal. Indeed, the derivate of the approximated log-likelihood has to be asymptotically unbiased [2].

To overcome this problem in \mathbb{Z}^d , the tapered periodogram can be used [17]. The following corresponds to graph extensions of standard time series models.

This is tractable only in two cases :

- The MA_K case : If there exists $K > 0$ such that the true spectral density f_{θ_0} is a polynomial of degree bounded by K .
- The AR_K case : If there exists $K > 0$ such that *all* the spectral densities (for any $\theta \in \Theta$) of the parametric set are such that $\frac{1}{f_\theta}$ is a polynomial of degree bounded by K .

So, to define the unbiased approximate log-likelihood, we first introduce the unbiased periodogram in one of this cases.

We first count, for any couple of vertices $(i, j) \in G^2$, the number of path (counted with their weight) going from i to j with length $k = 0, \dots, K$.

For this, define the K -type of a couple of vertex $(i, j) \in G^2$ as the $(K+1)$ -tuple given by

$$t(i, j) := \left(W_{ij}^{(p)} \right)_{p \leq K}.$$

Then, denote

$$V_K := \left\{ \left(W_{ij}^{(p)} \right)_{p \leq K}, i, j \in G \right\}.$$

This set gives among any $i, j \in G$ all possible $(K+1)$ -tuple of the number of path (counted with their weight) going from i to j with length $k = 0, \dots, K$. We highlight that it is finite as soon as the entries of the weighted adjacency operator W take a finite number of values, and since the degree of \mathbf{G} is bounded.

In the following, we chose n large enough to ensure that

Assumption 3.4.

$$\forall v \in V_K, \exists (i, j) \in G_n^2, \left(W_{ij}^{(p)} \right)_{p \leq K} = v.$$

This is possible since V_K is finite, and since $t(i, j) = \{0\}^{K+1}$ as soon as the shortest path from i to j is longer than K .

Define now the matrix $B^{(n)}$ (the dependency on K is omitted, for clarity) by

$$\begin{aligned} B_{ij}^{(n)} &:= \frac{\text{Card} \{(k, l) \in G_n \times G, t(k, l) = t(i, j)\}}{\text{Card} \{(k, l) \in G_n \times G_n, t(k, l) = t(i, j)\}}, \text{ if } t(i, j) \neq \{0\}^{K+1} \\ &:= 1 \text{ if } t(i, j) = \{0\}^{K+1}. \end{aligned}$$

The matrix $B^{(n)}$ gives a correction of the boundary problem, comparing, for any $v \in V_K$ the frequency of the interior couples of vertices of K -type v with the boundary couples of vertices of K -type v .

The explicit construction for $G = \mathbb{Z}^2$ and $K = 2$ is given at the end of the section.

Back to the general case, we can define the unbiased periodogram as

$$X_n^T Q_n \left(\frac{1}{g} \right) X_n$$

where $Q_n(g) := B^{(n)} \odot K_n(g)$, and \odot denotes the Hadamard product for the matrix i.e. $(M \odot N)_{ij} = M_{ij} N_{ij}$. Notice here that thanks to the example given just below, it is actually a way to extend the tapered periodogram (see for instance [10]) in the general case.

Let us write now the unbiased empirical log-likelihood :

$$L_n^{(u)}(\theta) := -\frac{1}{2} \left(m_n \log(2\pi) + m_n \int \log(f_\theta(x)) d\mu(x) + X_n^T \left(Q_n \left(\frac{1}{f_\theta} \right) \right) X_n \right)$$

We denote $\theta_n, \tilde{\theta}_n, \bar{\theta}_n, \theta^{(u)}$ the maximum likelihood estimators associated to respectively $L_n, \tilde{L}_n, \bar{L}_n, L_n^{(u)}$.

We will need the following assumption, which holds for instance for \mathbb{Z}^d :

Assumption 3.5. *There exists a positive sequence $(u_n)_{n \in \mathbb{N}}$ such that, $u_n \rightarrow 0$, $\sup_{ij} \left| B_{ij}^{(n)} - 1 \right| \leq u_n$*

We can now state the fundamental theorem :

Theorem 3.1. *Under Assumptions 3.2 and 3.3, the estimators $\theta_n, \bar{\theta}_n, \tilde{\theta}_n$ converge, as n goes to infinity, $P_{f_{\theta_0}}$ -a.s. to the true value θ_0 for any $\theta_0 \in \Theta$. If moreover Assumption 3.5 is true, we get the convergence of $\theta_n^{(u)}$.*

Proof. The proof follows the guidelines of the usual proof that can be found in [2]. We highlight the main changes due to the graph case. We first define the Kullback information on G_n of f with respect to g , by

$$\mathbb{I}\mathbb{K}_n(f, g) := \mathbb{E}_{P_f} \left[-\log\left(\frac{dP_g}{dP_f}\right) \right]$$

and the asymptotic information (on \mathbf{G}) by

$$\mathbb{I}\mathbb{K}(f, g) = \lim_n \frac{1}{m_n} \mathbb{I}\mathbb{K}_n(f, g)$$

when it is finite.

The convergence of the estimators of the maximum approximate likelihood relies on the following lemmas :

Lemma 3.1. *Under Assumptions 3.2 and 3.3, the asymptotic Kullback information exists and verifies*

$$\mathbb{I}\mathbb{K}(f, g) = \frac{1}{2} \int \left(-\log\left(\frac{f}{g}\right) - 1 + \frac{f}{g} \right) d\mu$$

Furthermore, if we denote $l_n(\theta, X) = \frac{1}{m_n} L_n(\theta, X)$, we have $P_{f_{\theta_0}}$ -almost surely,

$$l_n(\theta_0) - l_n(\theta) \rightarrow \mathbb{I}\mathbb{K}(f_{\theta_0}, f_{\theta})$$

uniformly as n tends to infinity.

This property is still true if we change L_n into \bar{L}_n or \tilde{L}_n

Furthermore, for both the AR_K or the MA_K case (see above), this still holds when L_n is changed into $L_n^{(u)}$.

Lemma 3.2. *Let f_{θ_0} be the true spectral density, and l_n a deterministic sequence of continuous functions such that*

$$\forall x \in \Theta, l_n(\theta_0) - l_n(x) \rightarrow \mathbb{I}\mathbb{K}(f_{\theta_0}, f_x)$$

uniformly as n tends to infinity. Then, if $x_n = \arg \max_{\theta} l_n(x)$, we obtain

$$x_n \rightarrow \theta_0$$

□

We require additional assumptions to provide the asymptotic normality and efficiency of the last estimator $\theta_n^{(u)}$, in both the MA_k , or the AR_k cases.

Assumption 3.6. *Assume that*

- There exists a positive sequence $(v_n)_{n \in \mathbb{N}}$ such that $v_n = o(\sqrt{m_n})$ and

$$\left| \frac{1}{m_n} \text{Tr}(K_{G_n}(g)) - \int g d\mu \right| \leq \alpha(g)v_n$$

- $\forall \theta \in \Theta, \frac{d}{d\theta}(f_\theta) \in \mathcal{F}_\rho, \frac{d^2}{d\theta^2}(f_\theta) \in \mathcal{F}_\rho$

The first assumption means that the convergence to the spectral measure μ is faster than $o(\sqrt{m_n})$. The second assumption is more classical, and same as in the usual case of \mathbb{Z} (see [2]).

Theorem 3.2. For any $\theta_0 \in \overset{\circ}{\Theta}$, in both the AR_K or MA_K cases, and under all previous assumptions 3.3, 3.2, 3.5, 3.6, the estimator $\theta_n^{(u)}$ of θ_0 is asymptotically normal :

$$\sqrt{m_n}(\theta_n^{(u)} - \theta_0) \xrightarrow{\mathcal{D}} \mathcal{N} \left(0, \left(\frac{1}{2} \int \left(\frac{f'_{\theta_0}}{f_{\theta_0}} \right)^2 d\mu \right)^{-1} \right), \text{ as } n \rightarrow \infty.$$

Furthermore, this asymptotic variance is the Fisher information, so this estimator is asymptotically efficient.

Proof. Here again, we mimic the usual proof by extending the result of [2] to the graph case. First, Taylor's formula gives, since $\theta_0 \in \overset{\circ}{\Theta}$, the existence of $\check{\theta}_n \in [\theta_n^{(u)}, \theta_0]$, such that

$$(l_n^{(u)})'(\theta_0) = (l_n^{(u)})'(\theta_n^{(u)}) + (\theta_0 - \theta_n^{(u)})(l_n^{(u)})''(\check{\theta}_n).$$

As $\theta_n^{(u)} = \arg \max l_n^{(u)}$, we get $(l_n^{(u)})'(\theta_n^{(u)}) = 0$ And so,

$$\sqrt{m_n}(\theta_0 - \theta_n^{(u)}) = \left((l_n^{(u)})''(\check{\theta}_n) \right)^{-1} \sqrt{m_n}(l_n^{(u)})'(\theta_0).$$

The theorem relies on three lemmas :

Lemma 3.3 provides the asymptotic normality for $\sqrt{m_n}(l_n^{(u)})'(\theta_0)$. Combined with Lemma 3.4, we get the asymptotic normality for $\sqrt{m_n}(\theta_0 - \theta_n^{(u)})$. Finally, Lemma 3.5 gives the Fisher information, and this proves that the estimator $\theta_n^{(u)}$ realize the Cramér-Rao bound. This provides the efficiency of this estimator.

Lemma 3.3.

$$\sqrt{m_n}(l_n^{(u)})'(\theta_0) \xrightarrow{n \rightarrow \infty} \mathcal{N} \left(0, \frac{1}{2} \int \left(\frac{f'_{\theta_0}}{f_{\theta_0}} \right)^2 d\mu \right)$$

Lemma 3.4.

$$\left((l_n^{(u)})''(\check{\theta}_n) \right)^{-1} \xrightarrow{n \rightarrow \infty} 2 \left(\int \left(\frac{f'_{\theta_0}}{f_{\theta_0}} \right)^2 d\mu \right)^{-1}$$

Lemma 3.5. *The asymptotic Fisher information may be written as :*

$$J(\theta_0) = \frac{1}{2} \int \left(\frac{f'_{\theta_0}}{f_{\theta_0}} \right)^2 d\mu$$

□

Let us now explain what happens for \mathbb{Z}^2 , for $K = 2$. We define $W^{(\mathbb{Z}^2)}$ by

$$\forall i, j, k, l \in \mathbb{Z}, W^{(\mathbb{Z}^2)}((i, j), (k, l)) := \frac{1}{4} \mathbb{1}_{|i-j|+|k-l|=1}.$$

We denote by u, v the vertices of \mathbb{Z}^2 and the canonical generator of \mathbb{Z}^2 by e_1, e_2 . We mean that if $u = (i, j), i, j \in \mathbb{Z}$, then we can write $v = u + ke_1 + le_2, k, l \in \mathbb{Z}$ for $v = (i + k, j + l)$. Actually there exist only five 2-type of vertices (see figure 3 for the meaning of this construction) : $\forall u \in \mathbb{Z}^2$

- $t(u, u) = (0, 0, \frac{1}{4}) := t_1$
- $t(u, u + e_1) = t(u, u - e_1) = t(u, u + e_2) = t(u, u - e_2) = (0, 1, 0) := t_2$
- $t(u, u + e_1 + e_2) = t(u, u + e_1 - e_2) = t(u, u - e_2 + e_2) = t(u, u - e_1 - e_2) = (0, 0, \frac{1}{8}) := t_3$
- $t(u, u + 2e_1) = t(u, u - 2e_1) = t(u, u + 2e_2) = t(u, u - 2e_2) = (0, 0, \frac{1}{16}) := t_4$
- $t(u, v) = (0, 0, 0) := t_5$ in all the other cases

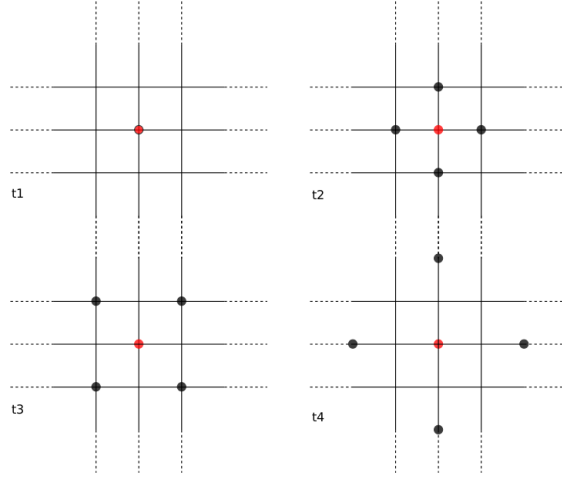
In this example, we set $G_n = [1, n]^2$, and we can compute that

- $\frac{\text{Card}\{(k, l) \in G_n \times G, t(k, l) = t_1\}}{\text{Card}\{(k, l) \in G_n \times G_n, t(k, l) = t_1\}} = \frac{n^2}{n^2} = 1$
- $\frac{\text{Card}\{(k, l) \in G_n \times G, t(k, l) = t_2\}}{\text{Card}\{(k, l) \in G_n \times G_n, t(k, l) = t_2\}} = \frac{4n(n-1)}{4n^2}$
- $\frac{\text{Card}\{(k, l) \in G_n \times G, t(k, l) = t_3\}}{\text{Card}\{(k, l) \in G_n \times G_n, t(k, l) = t_3\}} = \frac{4(n-1)^2}{n^2}$
- $\frac{\text{Card}\{(k, l) \in G_n \times G, t(k, l) = t_4\}}{\text{Card}\{(k, l) \in G_n \times G_n, t(k, l) = t_4\}} = \frac{4n(n-2)}{4n^2}$

One can notice that

$$\sup_{ij} \left| B_{ij}^{(n)} - 1 \right| \xrightarrow{n \rightarrow \infty} 0.$$

The assumption 3.5 ensure that this property holds for the graph we consider.



4 Algebraical construction of covariance operator on a graph

In Section 2, we dealt with the case of $ARMA^{(W)}$ processes indexed by the vertex G of a graph \mathbf{G} . We built all processes with the weighted adjacency operator without specifying a general notion of stationarity. In this section, we give a definition of stationarity for Gaussian processes on a graph, compatible with the construction of $ARMA^{(W)}$ processes given in Section 2.

This leads to a large class of covariance operators. In this class, we lay a special interest to general $ARMA$, built with operators different from W . As said in Section 2, this leads to a modification of the graph and we can consider working on the new graph. Actually, this section is devoted to modify the weights in an “isotropic” sense, and locally, for finite order operators. Now, let us give a sense to this “isotropic” modification.

For clarity, each time we will define a quantity of interest we will illustrate it on the classical graph \mathbb{Z} .

To define a notion of stationarity (with respect to W) for Gaussian processes indexed by G , a first idea is to use the set of all automorphisms of \mathbf{G} . Recall that, a permutation σ on G is an automorphism if it leaves W invariant :

$$\forall i, j \in G, W_{\sigma(i)\sigma(j)} = W_{ij}.$$

In the case of \mathbb{Z} the automorphisms are the symmetry and the translation operators. Stationarity is defined through invariance by these transformations of the covariance function. More generally, the definition of stationarity on \mathbb{Z}^d , homogeneous trees or distance transitive graph may be set in the same

way [10], [1], [12]. Unfortunately, as the generic situation for a graph is to have a trivial set of automorphisms (reduced to the identity operator), the last way to stationarity is a dead-end, since any covariance operator would be stationary. To get away from this dead-end we choose to take another path defining stationary covariance operators as images of W by adequate invariant functions. We recall here that $W \in B_G$ where B_G is the set of bounded Hilbertian operators on $l^2(G)$. We denote also by Σ_G the set of permutation of G . Define the operator M_σ attached to the permutation $\sigma \in \Sigma_G$ by

$$\forall i, j \in G, (M_\sigma)_{ij} = \mathbb{1}_{i=\sigma(j)}$$

Define F_G the set of all continuous linear operators from $l^1(G)$ to $l^\infty(G)$, and let us define the class of invariant functions.

Definition 4.1. *We call invariant, a real function*

$$\Phi : \text{Dom}(\Phi) \subset B_G \mapsto F_G,$$

that satisfies the following assumptions:

- Dom_Φ is stable by any permutations and by the transposition :

$$\forall A \in \text{Dom}(\Phi), \forall \sigma \in \Sigma_G, M_\sigma^{-1} A M_\sigma \in \text{Dom}(\Phi), A^T \in \text{Dom}(\Phi)$$

- Φ commutes with the conjugation by M_σ , for any permutation σ of G :

$$\forall \sigma \in \Sigma_G, \forall A \in \text{Dom}(\Phi), \Phi(M_\sigma^{-1} A M_\sigma) = M_\sigma^{-1} \Phi(A) M_\sigma$$

- Φ commutes with the transposition.

$$\forall A \in \text{Dom}(\Phi), \Phi(A^T) = \Phi(A)^T.$$

We will denote by \mathcal{I}_G the set of invariant functions from B_G to F_G .

Remark : Notice that Φ is given by a family of functions ϕ_{ij} all defined on $\text{Dom}(\Phi)$. Actually, thanks to the invariance, these functions are completely determined by the domain $\text{Dom}(\Phi)$, and two real-valued functions $\phi := \phi_{11}$ and $\psi := \phi_{12}$ from B_G to \mathbb{R} such that

- For any σ such that $\sigma(1) = k$,

$$\forall k \in G, \forall A \in \text{Dom}(\Phi), \Phi(A)(k, k) = \phi(M_\sigma^{-1} A M_\sigma)$$

- For any σ such that $\sigma(1) = k, \sigma(2) = l$,

$$\forall k, l \in G, k \neq l, \forall A \in \text{Dom}(\Phi), \Phi(A)(k, l) = \psi(M_\sigma^{-1} A M_\sigma)$$

- If τ_{12} is the simple permutation, permuting 1 and 2,

$$\psi(A) = \psi(M_{\tau_{12}} A^T M_{\tau_{12}})$$

Notice that this condition implies that

- For any σ letting 1 invariant,

$$\forall k \in G, \forall A \in \text{Dom}(\Phi), \phi(A) = \phi(M_\sigma^{-1} A M_\sigma)$$

- For any σ letting 1 and 2 invariant,

$$\forall k, l \in G, k \neq l, \forall A \in \text{Dom}(\Phi), \psi(A) = \psi(M_\sigma^{-1} A M_\sigma)$$

An example may be given by the discrete Laplacian. Recall that the discrete Laplacian $L^{(W)}$ of W on the graph is defined by :

$$\forall i, j \in G, L_{ij}^{(W)} := \mathbb{1}_{i=j} \sum_{k \in G} W_{ik} - W_{ij}$$

Set, for $A \in B_G$, $\phi_L(A) = \sum_{k \in G} A_{1k}$ and $\psi_L(A) = -A_{12}$, where ϕ and ψ are defined in the last remark. We get

$$\Phi_L(A)_{ij} = \mathbb{1}_{(i=j)} \sum_{k \in G} A_{ik} - A_{ij} = L_{ij}^{(A)}$$

That proves that $\Phi_L \in \mathcal{I}_G$, and its domain is the set of the operators $A \in B_G$ such that, for any $i \in G$, the sequence $(A_{ik})_{k \in G}$ is summable.

Back to the general construction, we can make another remark about the previous construction of invariant functions. Actually, one can imagine that the function ϕ_{ii} given by an invariant function Φ is built exploring the graph from the vertex $i \in G$. We can then define the order of an invariant function either finite or infinite, to measure if this exploration is local. The distance d denotes the natural distance on the graph \mathbf{G} , that is, for $i, j \in G$, the length of the shortest path going from i to j .

For any $r > 0$, let us define $B^{(W)}(i, r)$ the ball of radius r (for the natural distance d) centered on a vertex $i \in G$:

$$B^{(W)}(i, r) := \{j \in G, d(i, j) \leq r\}.$$

This leads to the following definition :

Definition 4.2 (Order of an invariant function). *Let Φ be an invariant function in \mathcal{I}_G , we will say that Φ is r -local for some $r \geq 0$ if, with the notations of the previous remark and for any $W \in \text{Dom}(\Phi)$, $\phi(W)$ depends only on $(W_{jk}, j, k \in B^{(W)}(1, r))$, and $\psi(W)$ depends only on*

$$(W_{jk}, j, k \in B^{(W)}(1, r) \cup B^{(W)}(2, r)).$$

The order of Φ is define as the smallest $r \geq 0$ such that Φ is r -local.

The order of an invariant function is a very important notion because it provides some smaller classes of invariant functions and will enable us to classify stationary processes.

Back to the general construction, we recall that F_G is the set of linear continuous operators from $l^1(G)$ into $l^\infty(G)$. Any covariance operator is in this class, so we can define stationarity over the graph \mathbf{G} with some subclasses of F_G using the invariant functions \mathcal{I}_G .

We are now able to state an extension of ‘‘isotropy’’ for any graph, and gives some way to modify the adjacency operator.

Definition 4.3. *We say that a Gaussian process $(X_i)_{i \in S}$ is stationary of order k , if its covariance operator K verifies*

$$K \in \mathcal{S}_+^k(W) := \{K = \Phi(W), \Phi \in I_G \text{ of order } r, K \text{ is positive definite} \}$$

We say that a operator W' is an admissible modification of the graph \mathbf{G} of order k if we have

$$W' \in \mathcal{S}_1^+(W) := \{K = \Phi(W), \Phi \in I_G \text{ of order } r, (W_{ij} = 0 \Rightarrow K_{ij} = 0)\}$$

To close this section, let us notice that if W takes a finite number of values, then any admissible modification of order k do the same, since there is a finite number of possible subgraphs of size bounded by D^k .

For the usual case of $G = \mathbb{Z}$ we are able to recover any covariance operator K , even non regular. Set

$$\forall i, j \in \mathbb{Z}, K_{ij} = r_{|i-j|},$$

and define Φ as

$$\forall W \in B_G, \Phi(W) = \sum_{p \geq 0} r_p \Phi^{(p)}(W).$$

with, for all $i, j \in \mathbb{Z}$,

$$\Phi^{(p)}(W)_{ij} = \sum_{k_1 \neq i, j \in \mathbb{Z}} \sum_{k_2 \neq i, j, k_1 \in \mathbb{Z}} \cdots \sum_{k_{p-1} \neq i, j, k_1, \dots, k_{p-2} \in \mathbb{Z}} W_{ik_1} W_{k_1 k_2} \cdots W_{k_{p-1} j}.$$

The general construction gives a large class of Gaussian graphical models, since any AR_1 built with a isotropic modification of \mathbf{G} provides a graphical model [7]. This falls into a very classical point of view about Gaussian fields, and allow us to use all general tools developed in this framework for our case.

5 Homogeneity and spectral measure

This section is entirely dedicated to understand the convergence of the spectral measure 3.3, and to several sufficient assumptions to ensure this convergence.

To this aim, we will choose an operator $A \in S_1$.

First, it is important to notice that Assumption 3.3 is equivalent to the following, under Assumption 3.1, thanks to Lemma 7.1.

Assumption 5.1. *Weak convergence of the spectral measure*

Let A_{G_n} denote the extracted operator from A over the subgraph G_n , and $\lambda_1, \dots, \lambda_{\#G_n}$ its corresponding eigenvalues, written with their multiplicity orders. We will say that the graph G admits the spectral measure μ with respect to the sequence G_n if

$$\frac{1}{\#G_n} \sum_{i=1}^{\#G_n} \delta_{\lambda_i} \xrightarrow{n \rightarrow \infty} \mu.$$

in the sense of the weak convergence.

Let us define

$$\mu_n^{(1)} = \frac{1}{\#G_n} \sum_{i=1}^{\#G_n} \delta_{\lambda_i}.$$

and

$$\mu_n^{(2)} = \frac{1}{\#G_n} \sum_{x \in G_n} \mu_{x,x}.$$

To prove this equivalence, we just have to notice that :

$$\begin{aligned} \int_{\text{Sp}(W)} \lambda^k d\mu_n^{(1)}(\lambda) &= \int_{\text{Sp}(W)} \lambda^k d\mu_n^{(2)}(\lambda) \\ &= \frac{1}{\#G_n} \sum_{i=1}^{\#G_n} \lambda_i^k - \frac{1}{\#G_n} \sum_{x \in G_n} (W^k)_{xx} \\ &= \frac{1}{\#G_n} \text{Tr}((W_{G_n})^k) - \frac{1}{\#G_n} \text{Tr}((W^k)_{G_n}). \end{aligned}$$

So, we get the result by the quasi-homomorphism lemma 7.1.

This is a classical graph theory assumption, usually verified in the frame of random graph (in which G_n is not a growing sequence of subgraphs, but a graph of size n whose spectral measure converge weakly to a particular law.). For instance, the spectral measure of some random graphs with given degree distribution converge to the spectral measure of regular tree (see [3] [4]), whose is well known. [15].

In the frame of road traffic, we provide also another assumption about the sequence of graphs, which ensures the convergence of the spectral measure.

We define first V_l as

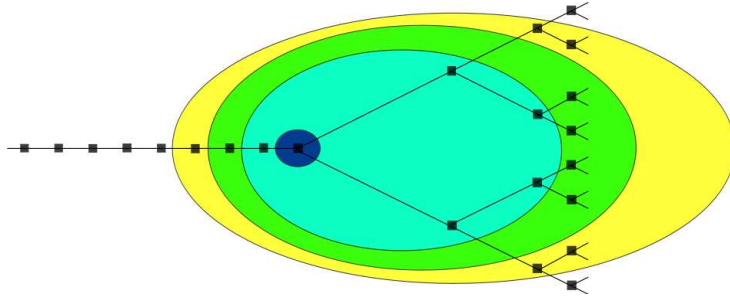
$$\forall l \geq 0, V_l := \{W^l_{kk}, k \in G\}.$$

Notice first that since W is of finite order, V_l is finite for any $l \geq 0$.

Assumption 5.2 (Homogeneity assumption). *Assume that the sequence $(G_n)_{n \in \mathbb{N}}$ and the operator W verify :*

$$\forall l \geq 0, \forall n \geq 1, \forall v \in V_l, \frac{\#\{j \in G_n, (A^l)_{jj} = v\}}{\#G_n} \xrightarrow{n \rightarrow \infty} p_v^{(l)}.$$

Remark : This last assumption may be understood as an homogeneity in frequency of the type (see section 3) of vertices chosen in the subgraphs G_n . To illustrate this talk, figure 5 shows an example of admissible sequences of subgraphs.



Proposition 5.1. *Under homogeneity assumption, we obtain the convergence of the weak convergence of the spectral measure :*

$$\frac{1}{\#G_n} \sum_{i=1}^{\#G_n} \delta_{\lambda_i} \rightarrow \mu.$$

Proof. Under homogeneity assumption, we get

$$\frac{1}{\#G_n} \text{Tr}((W^l)_{G_n}) \rightarrow \sum_{v \in J_l} p_v^{(l)} v.$$

Define

$$m_l := \sum_{v \in J_l} p_v^{(l)} v.$$

Since $\sup_{i,j \in S} |W_{ij}| \leq \frac{1}{d}$, we have

$$\|W\|_{2, \text{in}} \leq 1.$$

So

$$\forall l \geq 0, J_l \subset \text{Sp}(W) \subset [-1, 1].$$

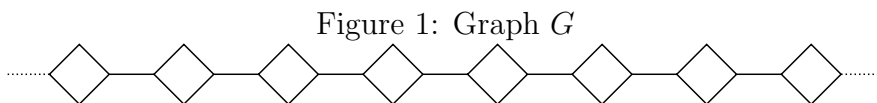
Then

$$m_l \leq 1.$$

Since the series $t \mapsto \sum m_l \frac{t^l}{l!}$ has a positive convergence radius, there exists a probability measure μ with moments $(m_l)_{l \geq 0}$. The weak convergence of the measure is derived from the convergence of the moments. This achieves the proof of the last proposition. \square

6 Simulations

In this section, we give some simulations over a very simple case, where the graph G is built taking some rhombus connected by a simple edge on the left and right. See Figure 3



The sequence of nested subgraphs chosen here is the growing neighborhood sequence (we chose a point x and we take $G_n = \{y \in G, d(x, y) \leq n\}$). We want to fit an AR_2 model, such that,

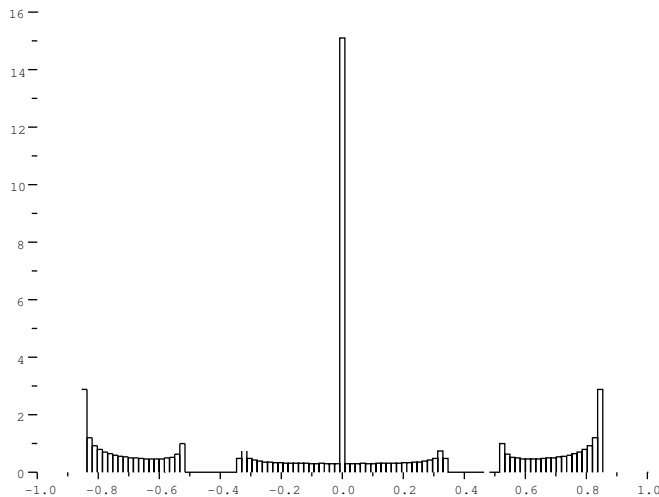
$$\Theta =]-1, 1[$$

$$\forall \theta \in \Theta, f_\theta(x) = \left(\frac{1}{1 - \theta x} \right)^2$$

Here, we take for A the adjacency operator of G normalized in order to get $\|A\|_{\text{op}, \infty} \leq 1$. We chose $\theta_0 = \frac{1}{2}$, $\#G_n = 724$, and we approximate

the spectral measure of G with Monte-Carlo methods. Figure 4 shows the empirical spectrum of the graph G with respect to the sequence of subgraph G_n .

Figure 2: Empirical spectrum



To compute $(K_n(f_\theta))^{-1}$, we use the power series representation of f_θ , and truncate this expression after the 15 first coefficient. This choice is quite arbitrary, but it ensure that the simulations error are neglectible with respect to the theoretical ones.

Figure 5 gives the empirical distribution of

$$\sqrt{\#G_n} \sqrt{\int_{\text{Sp}(A)} \left(\frac{f'_\theta}{f_\theta} \right)^2 (\tilde{\theta}_n - \theta_0)}.$$

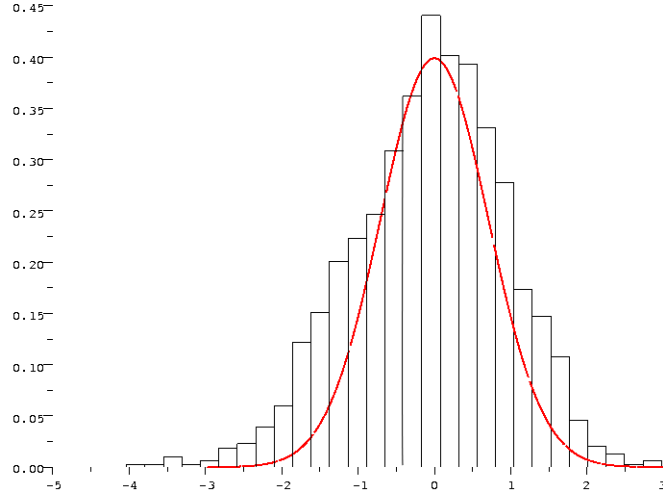
We observe that given the size of the subgraphs chosen, the error is a little less concentrated than the limit error (in red) which should be a $\mathcal{N}(0, 1)$.

7 Appendix

7.1 Szegő Lemmas

Szegő Lemmas are useful in time series analysis since they provide the good approximation to choose for the likelihood. As said in Section 3, this approximations of the likelihood are easier to compute.

Figure 2: Empirical distribution



In this section, we generalize a weak version of the Szegő Lemmas, for a general graph, under Assumption 3.1 (non expansion criteria for G_n), and Assumption 3.3 (existence of the spectral measure μ).

For any matrix $(B_{ij})_{i,j \in G_n}$, we define the norm block :

$$b_N(B) = \frac{1}{\delta_N} \sum_{i,j \in G_N} |B_{ij}|$$

We can state the equivalent version of the first Szegő lemma for time-series

Lemma 7.1. *Asymptotic homomorphism*

For any k, n positive integers, and g_1, \dots, g_k analytic functions over $[-1, 1]$ such that the regularity factor α is finite for all this functions ($\alpha(g_i) < +\infty$), we have :

$$b_n(K_{G_n}(g_1) \cdots K_{G_n}(g_k) - K_{G_n}(g_1 \cdots g_k)) \leq \frac{k-1}{2} \alpha(g_1) \cdots \alpha(g_k)$$

Corollary 7.1.

$$\frac{1}{m_n} \log \det(K_n(g)) \xrightarrow{n \rightarrow \infty} \int \log(g) d\mu$$

Proof. This proof follows again the one of [2]. We will prove the result by induction on k . First we deal with the case $k = 2$. Let f and g analytic functions over $[-1, 1]$ such that $\alpha(f) < +\infty$ and $\alpha(g) < +\infty$. We write

$$\begin{aligned} & b_n(K_{G_n}(f)K_{G_n}(g) - K_{G_n}(fg)) \\ &= \frac{1}{\delta_n} \sum_{i,j \in G_n} \left| \sum_{k \in G_n} (K_{G_n}(f))_{ik} (K_{G_n}(g))_{kj} - \sum_{k \in G} (K_{G_n}(f))_{ik} (K_{G_n}(g))_{kj} \right| \\ &= \frac{1}{\delta_n} \sum_{i,j \in G_n} \sum_{k \in G \setminus G_n} |K(f)_{ik}| |K(g)_{kj}| \end{aligned}$$

Using $K(g) = \sum_{h=0}^{\infty} g_h W^h$, Fubini's theorem gives, since all the previous sequences are $l^1(\mathbb{G})$,

$$\begin{aligned} & b_n(K_{G_n}(f)K_{G_n}(g) - K_{G_n}(fg)) \\ &\leq \frac{1}{\delta_n} \sum_{i,j \in G_n} \sum_{k \in G \setminus G_n} \left| (K_{G_n}(f))_{ik} (K_{G_n}(g))_{kj} \right| \\ &\leq \left(\sup_{k \in G \setminus G_n} \sum_{i \in G_n} |K(h)_{ik}| \right) \times \frac{1}{\delta_n} \sum_{k \in G \setminus G_n} \sum_{j \in G_n} \sum_{h=0}^{\infty} |g_h| |(W^h)_{kj}| \\ &\leq \left(\sup_{k \in G} \sum_{i \in G} |K(f)_{ik}| \right) \times \sum_{h=0}^{\infty} |g_h| \frac{1}{\delta_n} \sum_{k \in G \setminus G_n} \sum_{j \in G_n} |(W^h)_{kj}| \end{aligned}$$

Introducing

$$\Delta_h = \sup_{N \in \mathbb{N}} \frac{1}{\delta_N} \sum_{k \in G \setminus G_N} \sum_{j \in G_N} |(W^h)_{kj}|,$$

we get

$$b_n(K_{G_n}(f)K_{G_n}(g) - K_{G_n}(fg)) \leq \sup_{k \in G} \sum_{i \in G} |K(f)_{ik}| \sum_{h=0}^{\infty} |g_h| \Delta_h.$$

The coefficient Δ_h can be seen as a porosity factor, measuring the weight of the paths going from the interior of G_n to outside, following path of length h .

Note that $\Delta_h \leq h$, so we get

$$\sum_{h=0}^{\infty} |g_h| \Delta_h \leq \alpha(g)$$

Notice that the following define a norm on B_G :

$$\forall B \in B_G, \|B\|_{\infty, in} := \sup_{k \in G} \sum_{i \in G} |B_{ik}|.$$

We get

$$\begin{aligned} \|K(f)\|_{\infty, in} &= \sup_{k \in G} \sum_{i \in G} |K(f)_{ik}| \\ &\leq \sum_{h=0}^{\infty} |f_h| \|A^h\|_{\infty, in} \\ &\leq \sum_{h=0}^{\infty} |f_h| \|A\|_{\infty, in}^h \\ &\leq \sum_{h=0}^{\infty} |f_h| := \|g\|_{1, pol} \end{aligned}$$

Finally, we get

$$b_n(K_{G_n}(f)K_{G_n}(g) - K_{G_n}(fg)) \leq \|g\|_{1, pol} \alpha(g)$$

To conclude the proof of the lemma, by symmetrization of the last inequality, and since $1 \leq (h+1)$,

$$b_n(K_{G_n}(f)K_{G_n}(g) - K_{G_n}(fg)) \leq \frac{1}{2} \alpha(f) \alpha(g)$$

We need the following inequalities, to perform the inductive step :

$$\begin{aligned} \alpha(fg) &\leq \alpha(f) \alpha(g) \\ b_n(BC) &\leq \|B\|_{\infty, in} b_n(C) \\ b_n(B+C) &\leq b_n(B) + b_n(C) \\ \|K_{G_n}(f)\|_{\infty, in} &= \|f\|_{1, pol} \leq \alpha(f) \end{aligned}$$

Under the previous assumptions, we get,

$$\begin{aligned} b_n(K_{G_n}(g_1) \times \cdots \times K_{G_n}(g_k) - K_{G_n}(g_1 \cdots g_k)) \\ &\leq \|K_{G_n}(g_1)\|_{\infty, in} b_n(K_{G_n}(g_2) \cdots K_{G_n}(g_k) - K_{G_n}(g_2 \cdots g_k)) \\ &\quad + b_n(K_{G_n}(g_1)K_{G_n}(g_2 \cdots g_k) - K_{G_n}(g_1 \cdots g_k)) \\ &\leq \alpha(g_1) \frac{k-2}{2} \alpha(g_2) \cdots \alpha(g_k) + \frac{1}{2} \alpha(g_1) \alpha(g_2 \cdots g_k) \\ &\leq \frac{k-1}{2} \alpha(g_1) \cdots \alpha(g_k), \end{aligned}$$

which proves the result. \square

The following lemma enables to replace $K_n(g)$ by the unbiased version $Q_n(g)$.

Lemma 7.2. *Under assumptions 3.1, 3.3 and 3.5, and if f or g is polynomial with degree K , we have*

$$\frac{1}{m_n} \text{Tr} \left((K_n(f)K_n(g))^p - (K_n(f)Q_n(g))^p \right) \leq u_n \alpha(f)^p \alpha(g)^p$$

Proof. We define, for any f ,

$$f_{abs}(x) = \sum_k |f_k| x^k$$

Actually, the proof is based of the following idea : as soon as f or g is a polynomial of degree less or equal to K , we have to control only path of length less or equal to K . For any $p \geq 0$, we get,

$$\begin{aligned} & \frac{1}{m_n} \text{Tr} \left(\left(K_{G_n}(f) K_{G_n}\left(\frac{1}{g}\right) \right)^p - \left(K_{G_n}(f) Q_{G_n}\left(\frac{1}{g}\right) \right)^p \right) \\ & \leq \frac{1}{m_n} \sum_{i \in G_n} \sum_{i_0=i, i_1, \dots, i_{2p}=i} \prod_{l=0 \dots p-1} a_{i_{2l} i_{2l+1}} K_n\left(\frac{1}{g}\right)_{i_{2l} i_{2l+1}} K_n(f)_{i_{2l+1} i_{2l+2}} \\ & \quad - \frac{1}{m_n} \sum_{i \in G_n} \sum_{i_0=i, i_1, \dots, i_{2p}=i} \prod_{l=0 \dots p-1} K_n\left(\frac{1}{g}\right)_{i_{2l} i_{2l+1}} K_n(f)_{i_{2l+1} i_{2l+2}} \\ & \leq \frac{1}{m_n} \sup_{i_1, i_2, \dots, i_{2p+1}} \left(\prod_{l=0 \dots p-1} a_{i_{2l+1} i_{2l+2}} - 1 \right) \\ & \quad \times \sum_{i \in G_n} \sum_{i_0=i, i_1, \dots, i_{2p}=i} \prod_{l=0 \dots p-1} \left| K_n\left(\frac{1}{g}\right)_{i_{2l} i_{2l+1}} K_n(f)_{i_{2l+1} i_{2l+2}} \right| \\ & \leq \frac{1}{m_n} \sup_{i_1, i_2, \dots, i_{2p+1}} \left(\prod_{l=0 \dots p-1} a_{i_{2l+1} i_{2l+2}} - 1 \right) \\ & \quad \times \sum_{i \in G_n} \sum_{i_0=i, i_1, \dots, i_{2p}=i} \prod_{l=0 \dots p-1} K_n\left(\left(\frac{1}{g}\right)_{abs}\right)_{i_{2l} i_{2l+1}} K_n(f_{abs})_{i_{2l+1} i_{2l+2}} \\ & \leq \sup_{i_1, i_2, \dots, i_{2p+1}} \left| \prod_{l=0 \dots p-1} a_{i_{2l+1} i_{2l+2}} - 1 \right| \left\| \left(K_{G_n}(f_{abs}) K_{G_n}\left(\left(\frac{1}{g}\right)_{abs}\right) \right)^p \right\|_{2, in} \\ & \leq \sup_{i_1, i_2, \dots, i_{2p+1}} \left| \prod_{l=0 \dots p-1} a_{i_{2l+1} i_{2l+2}} - 1 \right| \alpha(f)^p \alpha\left(\frac{1}{g}\right)^p \\ & \leq u_n \alpha(f)^p \alpha\left(\frac{1}{g}\right)^p 3.5 \end{aligned}$$

This achieves the proof of the Lemma. This shows that the unbiased quadratic form Q_n may actually be understood as a correction of the error between $K_n(f)K_n(g)$ and $K_n(fg)$. \square

7.2 Proofs of theorem 3.1

Recall that the theorem relies on two lemmas. Lemma 3.2 states a condition for deterministic sequences to provide the convergence of the maximizer of this sequence.

Proof. of Lemma 3.2

Let x_∞ be an accumulation point of the sequence x_n , and x_{n_k} be a subsequence of limit x_∞ . As the function $x \mapsto \frac{1}{2} \int -\log\left(\frac{f_{\theta_0}}{f_x}\right) - 1 + \frac{f_{\theta_0}}{f_x} d\mu$ is continuous on Θ , and the convergence is uniform in x , we have

$$\ell_{n_k}(\theta_0) - \ell_{n_k}(x_{n_k}) \xrightarrow{k \rightarrow \infty} \frac{1}{2} \int -\log\left(\frac{f_{\theta_0}}{f_{x_\infty}}\right) - 1 + \frac{f_{\theta_0}}{f_{x_\infty}} d\mu.$$

But we can notice that, thanks to the definition of x_n , $\ell_{n_k}(\theta_0) - \ell_{n_k}(x_{n_k}) \leq 0$. So, since the function $x \mapsto -\log(x) + x - 1 \geq 0$ and is equal to zero only when $x = 1$, we get that $f_{\theta_0} = f_{x_\infty}$. By injectivity of the function $x \rightarrow f_x$, we get $x_\infty = \theta_0$, for any accumulation point x_∞ of the sequence $(x_n)_{n \in \mathbb{N}}$, which ends the proof of this first lemma. \square

Lemma 3.1 provides the uniform convergence of the contrast to the Kullback information. The proof may be cut into several lemmas.

Proof. of Lemma 3.1

First, notice that by construction, we have,

$$IK(f, g) = \lim_n \mathbb{E} \left[\frac{1}{m_n} (L_n(f, X_{G_n}) - L_n(g, X_{G_n})) \right]$$

when it exists. Then, we can compute

$$\begin{aligned} l_n(f, X) - l_n(g, X) &= -\frac{1}{2m_n} (\log \det(K_{G_n}(f)) - \log \det(K_{G_n}(g))) \\ &\quad - \frac{1}{2m_n} (X^T K_{G_n}(f)^{-1} X - X^T K_{G_n}(g)^{-1} X) \end{aligned}$$

Let $f, g \in \mathcal{F}_\rho, \rho > 0$, we will now provide several lemmas which, together ensure, the desired convergences.

Lemma 7.3 (Weak convergence of the measure). *For any ϕ a continuous bounded function from $\text{Im}(g)$ in \mathbb{R} , we get that*

$$\frac{1}{m_n} (\text{Tr}(\phi(K_n(g))) - K_n(\phi(g))) \xrightarrow{n \rightarrow \infty} 0.$$

Corollary 7.2 (Determinant lemma). *It holds that*

$$\left| \frac{1}{m_n} \log(\det(K_{G_n}(g))) - \int \log(g) d\mu \right| \xrightarrow{n \rightarrow \infty} 0.$$

Lemma 7.4 (Concentration lemma). *We get, the following convergence, for respectively $\Gamma = K_{G_n}(\frac{1}{g})$, $\Gamma = (K_{G_n}(g))^{-1}$ and $\Gamma = Q_n(\frac{1}{g})$ the unbiased operator.*

$$P_{f_{\theta_0}} - a.s, X^T \Gamma X \xrightarrow{n \rightarrow \infty} \int \frac{f}{g} d\mu$$

Lemma 7.4 ensures the convergence of $\tilde{l}_n(f) - \tilde{l}_n(g)$, $\bar{l}_n(f) - \bar{l}_n(g)$ and $l_n^{(u)}(f) - l_n^{(u)}(g)$ to $IK(f, g)$. Combined with lemma 7.2, it leads to the convergence of $l_n(f) - l_n(g)$ to $IK(f, g)$. Now, only the uniform convergence of this expressions remains to be shown. It is sufficient to provide the equicontinuity of the sequence of function.

Lemma 7.5 (Equicontinuity lemma). *For all $n \geq 0$, the functions $l_n(f) - l_n(g)$ are equicontinuous over $\{g_\theta, \theta \in \Theta\}$. This property remains if we substitute l_n by either \bar{l}_n, \tilde{l}_n or $l_n^{(u)}$*

We can now achieve the proof of the theorem :

First, notice that the considered function space is compact. So there exists a sequence $g_{\theta_p}, \theta_p \in \Theta$ dense in this space. All the lemmas above ensure the convergence almost sure of $l_n(f) - l_n(g_{\theta_p})$ to $IK(f, g_{\theta_p})$. The same result holds for \bar{l}_n, \tilde{l}_n and $l_n^{(u)}$.

The equicontinuity of this sequence leads to, since $(g_{\theta_p})_{p \in \mathbb{N}}$ is dense, the convergence for every $\theta \in \Theta$, of $l_n(f) - l_n(g_\theta)$ to $IK(f, g_\theta)$, and the corresponding difference with \bar{l}_n, \tilde{l}_n and $l_n^{(u)}$.

By compactity and equicontinuity of $\{g_\theta, \theta \in \Theta\}$, this convergence is furthermore uniform on Θ . This ends the proof of the main lemma. \square

7.3 Proof of the technical lemmas

Proof. of Lemma 7.3 Since the functions have compact support, it is sufficient to prove the result for $\phi(x) = x^k, k > 0$. Using $\alpha(g) \leq \rho < +\infty$ and Lemma

7.1, we get

$$\frac{1}{m_n} \text{Tr}((K_n(g))^k - K_n(g^k)) \leq \frac{\delta_n}{m_n} k \alpha(g)^k \xrightarrow{k \rightarrow \infty} 0.$$

□

Proof. of Lemma 7.4 First, consider the case $\Gamma_n = K_n\left(\frac{1}{g}\right)$. We compute the Laplace transform of $X_n^T \Gamma_n X_n$:

$$\begin{aligned} & \mathbb{E}_{P_f} \left[e^{\lambda X_n^T K_n\left(\frac{1}{g}\right) X_n} \right] \\ &= \frac{1}{(\sqrt{2\pi})^{m_n} \sqrt{\det(K_n(f))}} \int e^{\frac{1}{2} X_n^T ((K_n(f))^{-1} - 2\lambda K_n\left(\frac{1}{g}\right)) X_n} \\ &= \frac{1}{(\sqrt{2\pi})^{m_n} \sqrt{\det(K_n(f))}} \sqrt{\det \left(\left[(K_n(f))^{-1} - 2\lambda K_n\left(\frac{1}{g}\right) \right]^{-1} \right)} (\sqrt{2\pi})^{m_n} \\ &= \frac{1}{\sqrt{\det \left(I_{G_n} - 2\lambda K_n\left(\frac{1}{g}\right) K_n(f) \right)}}. \end{aligned}$$

These equalities hold as soon as $I_{G_n} - 2\lambda K_n\left(\frac{1}{g}\right) K_n(f)$ is positive, which is ensured for λ small enough.

Then, we can write

$$\begin{aligned} \phi_n(\lambda) &= \frac{1}{m_n} \log \left(\mathbb{E}_{P_f} \left[e^{\lambda X_n^T K_n\left(\frac{1}{g}\right) X_n} \right] \right) \\ &= -\frac{1}{2m_n} \log \det \left(I_{G_n} - 2\lambda K_n\left(\frac{1}{g}\right) K_n(f) \right). \end{aligned}$$

And we obtain, using $\gamma(\log(f)) < +\infty$, $\gamma(\log(g)) < +\infty$,

$$\lim_n \phi_n(\lambda) = -\frac{1}{2} \lim_n \frac{1}{m_n} \log \left(\mathbb{E}_{P_f} \left[e^{\lambda X_n^T K_n\left(\frac{1}{g}\right) X_n} \right] \right) = -\frac{1}{2} \int \log \left(1 - 2\lambda \frac{f}{g} \right) = \phi(\lambda).$$

We can also compute

$$\phi''(\lambda) = \int \frac{2\left(\frac{f}{g}\right)^2}{\left(1 - 2\lambda \frac{f}{g}\right)^2} d\mu > 0.$$

So, as soon as ϕ is convex, $\phi^*(\lambda) > 0$, for any $\lambda \neq \phi'(0) = \int \frac{f}{g} d\mu$.

We can now write

$$\begin{aligned}
\frac{1}{m_n} \log(\mathbb{P}(\frac{1}{m_n} X_n^T \Gamma_n X_n \geq t)) &= \frac{1}{m_n} \log(\mathbb{P}(e^{\lambda X_n^T \Gamma_n X_n} \geq e^{m_n \lambda t})) \\
&\leq \frac{1}{m_n} \log(e^{-m_n \lambda t}) + \frac{1}{m_n} \log(\mathbb{E}[e^{\lambda X_n^T \Gamma_n X_n}]) \\
&\leq -\lambda t + \phi_n(\lambda).
\end{aligned}$$

Then we get, $\forall t > \int \frac{f}{g} d\mu$,

$$\limsup_n \left(\frac{1}{m_n} \log(\mathbb{P}(\frac{1}{m_n} X_n^T \Gamma_n X_n \geq t)) \right) \leq -\lambda t + \phi(\lambda) = \phi^*(\lambda) > 0.$$

We can obtain the same bound for $t < \int \frac{f}{g}$. By Borel-Cantelli theorem, we get the \mathbb{P}_f -almost sure convergence of $\frac{1}{m_n} X_n^T \Gamma_n X_n$ to $\int \frac{f}{g} d\mu$.

To prove the same convergence with $\Gamma_n = (K_n(g))^{-1}$, we have to prove that the difference between the spectral empirical measure of $K_n(f)K_n(\frac{1}{g})$ and $K_n(f)(K_n(g))^{-1}$ converges weakly to zero. It is sufficient to control the convergence of every moment, because they have compact support.

For this, we make use of the Schatten norms defined, if $s_k(A)$ are the singular values of A , by

$$\|A\|_{Sch,p} = \left(\sum s_k(A)^p \right)^{\frac{1}{p}}.$$

Note that

$$|Tr(AB)| \leq \|AB\|_{Sch,1} \leq \|A\|_{Sch,1} \|B\|_{Sch,\infty}.$$

Hence, we can compute, for any $p \geq 1$,

$$\begin{aligned}
\frac{1}{m_n} \left| Tr \left(K_n^p(\frac{1}{g}) K_n^p(f) - K_n^{-p}(g) K_n^p(f) \right) \right| \\
\leq \frac{1}{m_n} \|K_n(g)^{-p} K_n^p(f)\|_{Sch,\infty} \left\| \left(K_n^p(\frac{1}{g}) K_n^p(g) - I_{G_n} \right) \right\|_{Sch,1} \\
\leq \frac{\delta_n}{m_n} \frac{M^{2p}}{m^{2p}} \alpha(g)^{2p} \alpha(\frac{1}{g})^{2p} \rightarrow 0.
\end{aligned}$$

To prove the same bound with $\Gamma_n = Q_n(\frac{1}{g})$, we have to prove that the difference between the spectral empirical measures of $K_n(f)K_n(\frac{1}{g})$ and $K_n(f)Q_n(\frac{1}{g})$ converge weakly to zero. It is a direct consequence of lemma 7.2. So, we get

$$X_n^T \Gamma_n X_n \rightarrow 0, \text{ a.s.}$$

□

Proof. of Lemma 7.5 We can first write :

$$\begin{aligned}
\frac{1}{m_n} |\log \det(K_n(f)) - \log \det(K_n(g))| &\leq \frac{1}{m_n} \sum |\log(\lambda_i) - \log(\mu_i)| \\
&\leq \frac{1}{m} \sup |\lambda_i - \mu_i| \\
&\leq \frac{1}{m} \|K_n(f) - K_n(g)\|_{2,op} \\
&\leq \frac{M}{m} \|f - g\|_\infty.
\end{aligned}$$

and

$$\left| \int \log(f) d\mu - \int \log(g) d\mu \right| \leq \frac{1}{m} \|f - g\|_\infty.$$

Then, notice that we have

$$\frac{1}{m_n} |X_n^T B X_n| \leq \frac{1}{m_n} \|B\|_{2,op} |X^T X|.$$

So it is sufficient to bound the quadratic form in norm 2 to ensure the equicontinuity of the sequences of function above.

So let us compute

$$\left\| K_n\left(\frac{1}{f}\right) - K_n\left(\frac{1}{g}\right) \right\|_{2,op} \leq \frac{1}{m^2} \|f - g\|_\infty.$$

and

$$\|(K_n(f))^{-1} - (K_n(g))^{-1}\|_{2,op} \leq \|(K_n(f))^{-1}(K_n(g))^{-1}\|_{2,op} \|(K_n(f)) - (K_n(g))\|_{2,op} \frac{1}{m^2} \|f - g\|_\infty.$$

and

$$\begin{aligned}
\left\| Q_n\left(\frac{1}{f}\right) - Q_n\left(\frac{1}{g}\right) \right\|_{2,op} &\leq \left\| Q_n\left(\frac{1}{f}\right) - Q_n\left(\frac{1}{g}\right) \right\|_{1,op} \\
&\leq \left(1 + C \frac{\delta_n}{m_n}\right) \sum_k |\psi_k - \phi_k|.
\end{aligned}$$

Here, we denote $\psi = \frac{1}{f}$ and $\phi = \frac{1}{g}$.

Since the map $g \mapsto \frac{1}{g}$ is continuous over \mathcal{F}_ρ , which is a compact subspace of l_1 , we get the uniform equicontinuity of the map $g \mapsto X^T Q_n\left(\frac{1}{g}\right) X$. We have proved this lemma, and the equicontinuity of the sequence of functions $l_n(f) - l_n(g)$. \square

Proof. of Lemma 3.3

To prove the assertion, it is sufficient to prove that

$$\lim_n \mathbb{E} \left[\exp \left(i \sqrt{m_n} t \left((l_n^{(u)})'(\theta_0) \right) \right) \right] = \exp \left(- \int \frac{1}{4} t^2 \frac{(f'_{\theta_0})^2(t)}{f_{\theta_0}^2(t)} d\mu(t) \right), \text{ with}$$

$$(l_n^{(u)})'(\theta) = -\frac{1}{2} \int \frac{f'_\theta}{f_\theta} d\mu + \frac{1}{2m_n} X^T Q_n \left(\frac{f'_\theta}{f_\theta} \right) X.$$

So

$$\begin{aligned} \sqrt{m_n} \mathbb{E} \left[(l_n^{(u)})'(\theta_0) \right] &= \sqrt{m_n} \left(-\frac{1}{2} \int \frac{f'_{\theta_0}}{f_{\theta_0}} d\mu + \frac{1}{2m_n} \text{Tr} \left(K_n(f_{\theta_0}) Q_n \left(\frac{f'_{\theta_0}}{f_{\theta_0}} \right) \right) \right) \\ &= \sqrt{m_n} \left(-\frac{1}{2} \int \frac{f'_{\theta_0}}{f_{\theta_0}} d\mu + \frac{1}{2m_n} \text{Tr} \left(K_n \left(f_{\theta_0} \frac{f'_{\theta_0}}{f_{\theta_0}} \right) \right) \right) \\ &\leq C v_n \sqrt{m_n} \xrightarrow{n \rightarrow \infty} 0. \end{aligned}$$

The equality holds here only if f_{θ_0} is polynomial, or if all the $f_\theta, \theta \in \Theta$ are polynomial. This brings out that the second theorem holds only for the AR_K or MA_K case. It also explains the denomination of unbiased estimator of the approximate likelihood for $\theta^{(u)}$. The equality can be obtained by derivation of the equality, true in both AR_K and MA_K cases,

$$K_n(f_{\theta_0}) Q_n \left(\frac{1}{f_\theta} \right) = K_n \left(\frac{f_{\theta_0}}{f_\theta} \right).$$

We can define

$$Z_n = t \frac{1}{2m_n} X^T Q_n \left(\frac{f'_\theta}{f_\theta} \right) X.$$

We have proven that, if

$$Z = t \frac{1}{2} \int \frac{f'_\theta}{f_\theta} d\mu.$$

then

$$\sqrt{m_n} (\mathbb{E} [Z_n] - Z) \rightarrow 0.$$

Then, it is sufficient to prove

$$\lim_n \mathbb{E} \left[\exp \left(i \sqrt{m_n} (Z_n - \mathbb{E} [Z_n]) \right) \right] = \exp \left(- \int \frac{1}{4} t^2 \frac{(f'_{\theta_0})^2(t)}{f_{\theta_0}^2(t)} d\mu(t) \right).$$

If τ_k denotes the eigenvalues of the symmetric matrix

$$M_n := \frac{t}{2} K_n(f_{\theta_0})^{\frac{1}{2}} Q_n \left(\frac{f'_{\theta_0}}{f_{\theta_0}} \right) K_n(f_{\theta_0})^{\frac{1}{2}},$$

then we can write

$$Z_n = \frac{1}{m_n} \sum_{k=1}^{m_n} \tau_k Y_k^2.$$

where $(Y_k)_{k \in G_n}$ is a standard Gaussian vector.

The independence of Y_k leads to

$$\log(\mathbb{E}[\exp(i\sqrt{m_n}(Z_n - \mathbb{E}[Z_n]))]) = - \sum_{k=1}^{m_n} \left(i \frac{\tau_k}{\sqrt{m_n}} + \frac{1}{2} \log(1 - 2i \frac{\tau_k}{\sqrt{m_n}}) \right).$$

The τ_k are bounded, thanks to the following inequality :

$$\begin{aligned} \|M_n\|_{2,op} &= \left\| \frac{t}{2} K_n(f_{\theta_0})^{\frac{1}{2}} Q_n\left(\frac{f'_{\theta_0}}{f_{\theta_0}^2}\right) K_n(f_{\theta_0})^{\frac{1}{2}} \right\|_{2,op} \\ &\leq \left\| \frac{t}{2} K_n(f_{\theta_0})^{\frac{1}{2}} \right\|_{2,op} \left\| Q_n\left(\frac{f'_{\theta_0}}{f_{\theta_0}^2}\right) \right\|_{2,op} \left\| K_n(f_{\theta_0})^{\frac{1}{2}} \right\|_{2,op} \\ &\leq \left\| \frac{t}{2} K_n(f_{\theta_0})^{\frac{1}{2}} \right\|_{2,op} \left\| Q_n\left(\frac{f'_{\theta_0}}{f_{\theta_0}^2}\right) \right\|_{1,op} \left\| K_n(f_{\theta_0})^{\frac{1}{2}} \right\|_{2,op} \\ &\leq M \alpha(f'_{\theta_0}) \alpha(f_{\theta_0})^2 \left(1 + \frac{\delta_n}{m_n}\right). \end{aligned}$$

The Taylor expansion of $\log(1 - 2\frac{\tau_k}{\sqrt{m_n}})$ gives

$$\log(\mathbb{E}[\exp(i\sqrt{m_n}(Z_n - \mathbb{E}[Z_n]))]) = -\frac{1}{m_n} \sum_{k=1}^{m_n} \tau_k^2 + R_n.$$

With $|R_n| \leq C \frac{1}{m_n \sqrt{m_n}} \sum_{k=1}^{m_n} |\tau_k|^3$

Since the τ_k are bounded the assertion is proven if we prove that

$$\frac{1}{m_n} \text{Tr}(M_n^2) = \frac{1}{m_n} \sum_{k=1}^{m_n} \tau_k^2 \xrightarrow{n \rightarrow \infty} \int \frac{1}{4} t^2 \frac{(f'_{\theta_0})^2(t)}{f_{\theta_0}^2(t)} d\mu(t)$$

The convergence is due to Lemmas 7.1 and 7.2. □

Proof. of Lemma 3.4 We have

$$(l_n^{(u)})''(\theta) = -\frac{1}{2m_n} \left(\int \frac{f''_{\theta} f_{\theta} - (f'_{\theta})^2}{f_{\theta}^2} d\mu + X^T Q_n \left(\frac{2(f'_{\theta})^2 - f''_{\theta} f_{\theta}}{f_{\theta}^3} \right) X \right),$$

which leads to

$$(l_n^{(u)})''(\theta) \xrightarrow{n \rightarrow \infty} \frac{1}{2} \int \left(\frac{f''_{\theta} f_{\theta} - (f'_{\theta})^2}{f_{\theta}^2} + \frac{f_{\theta_0} (2(f'_{\theta})^2 - f''_{\theta} f_{\theta})}{f_{\theta}^3} \right) d\mu. \quad (1)$$

Since the sequence $l_n^{(u)}$ is equicontinuous and $\check{\theta}_n \xrightarrow{n \rightarrow \infty} \theta_0$, we get the following convergence :

$$(l_n^{(u)})''(\check{\theta}_n) \xrightarrow{n \rightarrow \infty} \frac{1}{2} \int \left(\frac{(f'_{\theta_0})^2}{f_{\theta_0}^2} \right) d\mu.$$

□

Proof. of Lemma 3.5 As usually, it is sufficient to compute

$$\frac{1}{m_n} \text{Var} (L'_n(\theta_0)) = \lim \frac{1}{2m_n} \text{Tr}(M_n(\theta_0)^2),$$

where $M_n(\theta) = K_n(f_\theta)^{-1} K_n(f'_\theta) K_n(f_\theta)^{-1} K_n(f_{\theta_0})$.

This leads to

$$\frac{1}{m_n} \text{Var} (L'_n(\theta_0)) \rightarrow \frac{1}{2} \int \frac{(f'_{\theta_0})^2}{f_{\theta_0}^2} d\mu.$$

That ends the proof of this lemma and the theorem.

□

References

- [1] Jean-Pierre Arnaud. Fonctions sphériques et fonctions définies positives sur l'arbre homogène. *C. R. Acad. Sci. Paris Sér. A-B*, 290(2):A99–A101, 1980.
- [2] Robert Azencott and Didier Dacunha-Castelle. *Series of irregular observations*. Applied Probability. A Series of the Applied Probability Trust. Springer-Verlag, New York, 1986. Forecasting and model building.
- [3] Charles Bordenave and Marc Lelarge. Resolvent of large random graphs. *Random Structures Algorithms*, 37(3):332–352, 2010.
- [4] Charles Bordenave, Marc Lelarge, and Justin Salez. The rank of diluted random graphs, 2009.
- [5] Albrecht Böttcher and Bernd Silbermann. *Analysis of Toeplitz operators*. Springer Monographs in Mathematics. Springer-Verlag, Berlin, second edition, 2006. Prepared jointly with Alexei Karlovich.
- [6] Peter J. Brockwell and Richard A. Davis. *Introduction to time series and forecasting*. Springer Texts in Statistics. Springer-Verlag, New York, second edition, 2002. With 1 CD-ROM (Windows).

- [7] J. N. Darroch, S. L. Lauritzen, and T. P. Speed. Markov fields and log-linear interaction models for contingency tables. *Ann. Statist.*, 8(3):522–539, 1980.
- [8] Fabrice Gamboa, Jean-Michel Loubes, and Elie Maza. Semi-parametric estimation of shifts. *Electron. J. Stat.*, 1:616–640, 2007.
- [9] Ulf Grenander and Gábor Szegő. *Toeplitz forms and their applications*. Chelsea Publishing Co., New York, second edition, 1984.
- [10] Xavier Guyon. Parameter estimation for a stationary process on a d -dimensional lattice. *Biometrika*, 69(1):95–105, 1982.
- [11] Xavier Guyón. Modelling for spatial statistics. *Investigación Oper.*, 31(1):1–33, 2010.
- [12] Herbert Heyer. Stationary random fields over graphs and related structures. In *Wavelets, multiscale systems and hypercomplex analysis*, volume 167 of *Oper. Theory Adv. Appl.*, pages 157–171. Birkhäuser, Basel, 2006.
- [13] M. G. Kreĭn and A. A. Nudel'man. *The Markov moment problem and extremal problems*. American Mathematical Society, Providence, R.I., 1977. Ideas and problems of P. L. Čebyšev and A. A. Markov and their further development, Translated from the Russian by D. Louvish, Translations of Mathematical Monographs, Vol. 50.
- [14] Jean-Michel Loubes, Élie Maza, Marc Lavielle, and Luis Rodríguez. Road trafficking description and short term travel time forecasting, with a classification method. *Canad. J. Statist.*, 34(3):475–491, 2006.
- [15] Bojan Mohar and Wolfgang Woess. A survey on spectra of infinite graphs. *Bull. London Math. Soc.*, 21(3):209–234, 1989.
- [16] Christophe Pittet. On the isoperimetry of graphs with many ends. *Colloq. Math.*, 78(2):307–318, 1998.
- [17] GUYON (X.). *Champs aléatoires sur un réseau*. Masson, 1992.
- [18] N. Verzelen and F. Villers. Tests for Gaussian graphical models. *Comput. Statist. Data Anal.*, 53(5):1894–1905, 2009.
- [19] Nicolas Verzelen. Adaptive estimation of stationary Gaussian fields. *Ann. Statist.*, 38(3):1363–1402, 2010.