THÈSE

En vue de l'obtention du
DOCTORAT DE L'UNIVERSITÉ DE TOULOUSE

Délivré par :
Université Toulouse III Paul Sabatier (UT3 Paul Sabatier)

Discipline ou spécialité :
Mathématiques appliquées

Présentée et soutenue par
Thibault Espinasse
le : 5 décembre 2011

Titre :
Champs et processus gaussiens indexés par des graphes, estimation et prédiction

---

École doctorale :
Mathématiques Informatique Télécommunications (MITT)

Unité de recherche :
UMR 5219

Directeurs de thèse :
Fabrice Gamboa
Jean-Michel Loubes

Rapporteurs :
Jean-Marc Bardet
Mauro Piccioni

Membres du jury :
Benoît Cadre
Serge Cohen

"C'est véritablement utile, puisque c'est joli."

St Exupéry

Remerciements

Je tiens tout d'abord à remercier mes directeurs de thèse, Fabrice Gamboa et Jean-Michel Loubes. Travailler sous leur direction fut un grand plaisir pendant ces trois années, tant au niveau personnel que mathématique. Ils ont su me guider, me motiver, me secourir parfois, tout en me laissant une grande liberté sur les nouvelles pistes et sujets à explorer. Merci à eux pour leur aide, leur patience infinie, leur disponibilité et leur confiance.
Merci aussi à Jean-Marc Bardet et Mauro Piccioni, d'avoir accepté de rapporter sur cette thèse. Je tiens à les remercier pour leur lecture très minutieuse, et pour toutes leurs remarques qui m'ont permis à la fois de mettre en perspective certains résultats, et de corriger plusieurs erreurs. Je suis trés honoré qu'ils aient accepté.
Je remercie aussi Benoit Cadre et Serge Cohen, qui ont accepté de faire partie du jury de cette thèse, malgré un emploi du temps très chargé. Je leur en suis très reconnaissant.
Je tiens aussi à remercier Chichi pour son accueil chaleureux, sa disponibilité, et les discussions passionnantes que nous avons pu partager. Je regrette de ne pas avoir pu finir le travail à temps pour qu'il apparaisse ici.
Je voudrais aussi remercier Jean-Noël pour les échanges fructueux, et pour le travail qu'il a fait sur des données réelles. Je le remercie aussi, ainsi que Philippe Goudal de m'avoir laisser utiliser ses résultats dans ce manuscrit.
Merci aussi à toute l'équipe de l'ESP pour son accueil pendant ces trois ans. Merci à tous les permanents pour leur disponibilité et leur patience. Merci en particulier à Charles, Jean-Marc, Nicolas, Patrick, Serge, les Xaviers, et tous les autres. Merci aussi à Agnès, Delphine, Françoise et Marie-Laure qui m'ont sorti plusieurs fois des labyrinthes administratifs. Je remercie bien sûr tous les autres thésards, en particulier les collocs Blandine, Guillaume, Laury-Anne, Manu, Nadine, Yohann, ainsi que Adil, Adrien, Alice, Amic, Benjamin, Chloé, Flavia, Gouic, Hung, Julie, Marianne, Matthieu, Meuleu, Nabil, Paul, Rim, Sandrine, Victor et tous les autres, pour tous les bons moments passés ensemble.
Je voudrais aussi remercier mes professeurs de l'ENS et d'Orsay pour m'avoir guidé dans mes choix d'orientation, W. Werner, J.F. Le Gall, P. Massart, Gilles et Patricia, et mes amis, en particulier Jérémie, Marion, Robin, JP, Julien, Nicolas, Alice, Gilles...
Je tiens à remercier chaleureusement ma famille : mes parents et ma soeur, qui me supportent depuis 27 ans, et qui m'ont toujours soutenu, peu importe la situation. Merci pour tout. Un petit clin d'oeil à Enzo, Malik, Hayden, Lilya et aux bébés en cours. Merci à Ninou et Eme de m'avoir soutenu quand je tenais plus, et de m'avoir hébergé pendant la rédaction du manuscrit. Je remercie aussi toute la famille Barriacoise, Pam et Maureen, Manu D. et mes amis Toulousains de longue date (Flo, Julie, Manon, Thom, Yannick...).

Enfin, en vrac, je voudrais aussi saluer Bernard et Anne-Hélène, ILTG, Pierre D.M. pour son soutien dans l'adversité, Frankie, Brigitte, Jimmy et les deux Fred pour leurs explications toujours claires (en économie comme en sciences).

Champs et processus gaussiens indexés par des graphes, estimation et prédiction

L'objet de cette thèse est l'étude de processus gaussiens indexés par des graphes. Le but est de fournir des outils pour la modélisation, l'estimation, et la prédiction de tels champs ou processus, utilisant fortement la structure du graphe. Dans un premier travail, nous nous intéressons au problème de prédiction *aveugle* de séries chronologiques et montrons que le biais de l'erreur de prédiction décroît à une vitesse qui dépend de la régularité de la densité spectrale, sous une hypothèse de courte mémoire.
Nous utilisons ensuite la structure spectrale du graphe pour proposer des modèles de covariance pour des champs gaussiens indexés par ce graphe. Cela fournit immédiatement une représentation spectrale, qui permet d'étendre l'approximation de Whittle et l'estimation par quasi-maximum de vraissemblance à ce cadre.
Enfin, cette construction et le lemme de Szegö peuvent être étendus au cas spatio-temporel. Cela permet de mettre en pratique la théorie sur des données réelles.

Gaussian fields and processes indexed by graphs, estimation and prediction

In this work, we study Gaussian processes indexed by graphs. We aim at providing tools for modelisation, estimation, and prediction, that uses the structure of the underlying graphs. In the first Chapter, we deal with the *blind* prediction problem, and compute, in the case of short range dependancy, the rate of convergence of the bias in the prediction error. This rate depends on the regularity of the spectral density of the process.
Then, we use the eigenstructure of the adjacency operator of a graph to propose some models for covariance operators of Gaussian fields indexed by this graph. It leads to a spectral representation for this operator, that can be used to extend Whittle approximation, and quasi-maximum likelihood estimation.
Finally, this construction may be extended to the spatio-temporal case, where the Szegö lemma still holds.

Table des matières

Introduction générale

Un des axes de recherche fondamentaux en statistiques est la création de modèles pour expliquer le comportement de phénomènes aléatoires, et *comprendre le hasard* et réduire les incertitudes. Les processus aléatoires ont ainsi fourni des modèles pour analyser de nombreuses données issues de l'économie, la biologie, la sociologie, la météorologie... En particulier, plusieurs méthodes de prédiction ont été développées pour les séries chronologiques. Plusieurs généralisations ont plus récemment permis de définir des processus indexés par des espaces autres que $\mathbb{Z}$ (par exemple par $\mathbb{Z}^d$ [43],$\mathbb{R}^d$ [60], par des arbres [5], ou par des variétés [27]).
L'objet de cette thèse est l'étude statistique de processus gaussiens indexés par des graphes. Cette étude est motivée par un problème industriel de prédiction (en temps réel) du trafic routier. L'objectif est de prédire spatialement et temporellement, à partir de l'observation d'un unique échantillon de vitesses données sur le réseau routier à différents instants, des données manquantes ou futures.

En pratique, le réseau routier est découpé en portions de routes (arcs), connectées entre elles par des noeuds (les intersections). Des mesures de vitesses sont enregistrées :
– Á intervalles de temps réguliers sur certains arcs. Elles sont collectées par des capteurs physiques disposés le long des tronçons (boucles magnétiques)
– Á intervalles de temps irréguliers sur d'autres arcs. Ici les informations sont recueillies et transmises par une flotte de véhicules (Flotte de Taxis et de livreurs, Abonnés Coyotte...).
Nous disposons donc de données de vitesses $(X_{i,t})_{(i,t)\in \mathrm{Obs}}$, où Obs désigne un sous-ensemble de $G \times \mathbb{Z}$. Ici G désigne l'ensemble des arcs du réseau routier.

Notre but est d'utiliser la structure spatiale du réseau routier (un graphe) afin de spécifier la structure de covariance du processus observé. Pour cela, on modélise le processus de vitesses comme un processus gaussien indexé par les sommets G d'un graphe $\mathbf{G}$ et le temps est discret.
Le processus est choisi recentré afin d'intégrer dans sa moyenne les données exogènes du problème (jour de la semaine, taille des routes...).

Dans cette thèse, nous nous proposons d'étendre certains résultats existants pour les séries chronologiques au cas d'un graphe général.
La première étape est de construire un modèle pour l'opérateur de covariance du processus spatial, à temps fixé. Nous souhaitons utiliser fortement la structure de graphe. En outre, nous désirons que ce modèle soit *stationnaire* et *isotrope* au sens où, si le réseau routier *est semblable* en deux points, alors la structure de covariance entre ces deux points *se ressemble* aussi.
S'inspirant de la décomposition de Wold des séries chronologiques régulières, le

processus spatial est modélisé comme une moyenne mobile isotrope d'ordre infini (MA_∞). Ce type de modèle satisfait en effet la contrainte de *stationnarité* dans la mesure où toutes les arrêtes jouent le même rôle.
D'autre part, l'analyse spectrale de l'opérateur d'adjacence fournit une représentation spectrale pour les covariances de processus MA_∞. Dans toute la thèse, nous considèrons ce type de modèles spatiaux. Le modèle spatio-temporel découle de ce modèle spatial, en considérant une suite stationnaire (au sens litteral) de champs spatiaux ainsi construits.

La thèse est articulée en quatre chapitres, qui répondent à quatre questions successives :

1. Le fait d'utiliser le même échantillon pour l'étape d'estimation et celle de prédiction est-il préjudiciable à la qualité de la prévision ?
2. Quel type de modèle pour les processus indexés par un graphe permet d'utiliser la structure sous-jacente pour faire de l'inférence statistique ?
3. Comment peut-on généraliser la construction précédente (obtenue à temps fixé) à des processus spatio-temporels ?
4. Quelles quantités contiennent l'information pertinente sur la structure du graphe ?

La question 1 est traitée dans le chapitre 2, dans le cas d'une série chronologique $(X_i)_{i\in\mathbb{Z}}$. Dans ce chapitre, on se propose, à partir d'un unique échantillon $(X_i)_{i=-N,\cdots,-1}$, de prédire les données futures lorsque l'opérateur de covariance est inconnu.
Pour cela, on estime l'opérateur de prédiction à partir des coefficients de covariance empiriques. On contrôle ensuite le biais de l'erreur de prédiction commise, lorsque le même échantillon est utilisé pour la prédiction. Plus précisemment, la prédiction est faite à partir d'un sous-échantillon

$$(X_i)_{i=-K(N),\cdots,-1},$$

pour une suite $(K(N))_{N\in\mathbb{N}}$ bien choisie. On obtient une borne pour ce risque et la convergence du prédicteur correspondant dès que le nombre de données utilisées pour l'estimation est assez grand devant le nombre de données utilisées pour la prédiction ($K(N) << N$).

Le point 2 est l'objet du chapitre 3, partie centrale de la thèse. Un modèle de covariance de processus (spatiaux) réguliers est proposé. Ce modèle est construit avec l'objectif de satisfaire les critères suivants :
– Dans les cas classiques ($G = \mathbb{Z}^d, d \geq 1$, G distance-transitif...), il existe des modèles dans la littérature [43], [45]. On souhaite que le modèle proposé coincide avec ceux-là.

– Le modèle proposé devra également aussi fournir les outils classiques qui existent sur les exemples $G = \mathbb{Z}^d, d \geq 1$, G distance-transitif... En particulier, ce modèle doit assurer l'existence d'une représentation spectrale. De plus, on souhaite que celle-ci coïncide avec la représentation spectrale usuelle des cas classiques [5], [45]...
– Enfin, on veut étendre l'approximation de Whittle à ce modèle.

Il se trouve que tous ces points peuvent être satisfaits en considérant des processus MA_∞ isotropes. Dans le chapitre 3, on considère donc les modèles MA_∞ indexés par des graphes, et on étend l'approximation de Whittle à de tels processus.
Cette approximation a d'abord été developpée pour les séries chronologiques [67], [68]. Elle permet en particulier de faciliter le calcul de l'estimateur du maximum de vraisemblance en en maximisant une version approchée. L'extension aux graphes est possible sous des hypothèses de régularité sur la densité spectrale, d'homogénéité et d'amenabilité sur le graphe et pour un choix judicieux de suites de sous-graphes sur lesquels le processus est observé.
On montre alors la consistance de l'estimateur de vraisemblance approchée (type Whittle). Pour obtenir la normalité et l'efficacité asymptotiques, il est nécessaire, comme dans le cas de $\mathbb{Z}^d$, de "déformer" le périodogramme. L'estimateur *débiaisé* proposé dans la chapitre 3 (pour des graphes quasi-transitifs par exemple) est basé sur cette idée. Nous prouvons que cet estimateur est asymptotiquement normal et efficace.

Pour répondre à la question 3, il suffit d'étendre la construction du point précédent à des processus spatio-temporels. On recherche donc une structure donnant une suite de champs (spatiaux) gaussiens, MA_∞ sur le graphe, et stationnaire temporellement. En fait, il suffit de construire des processus stationnaires anisotropes sur $\mathbb{Z} \times G$. C'est l'objet du chapitre 4. Ce modèle est actuellement testé sur des données réelles issue de l'entreprise Mediamobile, spécialisée dans la prédiction du trafic routier. Malheureusement les premiers résultats ne nous sont pas encore parvenus à ce jour.

Le chapitre 5 explique l'importance de la notion de mesure spectrale pour les graphes. L'information sur la structure spatiale de graphe peut en effet être lue dans sa mesure spectrale.
Cette mesure apparait sous différentes formes dans la littérature. Elle peut être interprétée comme une transformation de Fourier, ou comme une diagonalisation de l'opérateur d'adjacence. Ce dernier point de vue permet de retrouver, des résultats classiques sur la représentation spectrale d'opérateurs invariants par automorphismes, dans plusieurs exemples.
C'est la structure spectrale de l'opérateur d'adjacence du graphe qui fournit les outils statistiques pour l'étude de processus indexés par des graphes (approxima-

tion de Whittle, prédiction...).
Le chapitre 5 contient surtout une partie bibliographique sur la mesure spectrale. Nous donnons plusieurs exemples classiques sur lesquels cette mesure peut-être calculée.
Nous espérons que notre thèse montre que la classe de MA isotropes introduite est une généralisation très naturelle des séries chronologiques (cas $G = \mathbb{Z}$), tant au niveau des outils mathématiques nécessaires à leur construction que des propriétés dont ils ont hérités. En particulier, lorsque le graphe intervient directement dans l'évolution du processus physique (comme c'est le cas pour le trafic routier), les modèles MA consituent d'excellents candidats pour modéliser et ainsi étudier le processus considéré, dès lors qu'une notion de *stationnarité* est choisie.
Enfin, nous signalons que les chapitres sont indépendants (à l'exception du chapitre 5). Les objets mathématiques considérés sont donc réintroduits dans chaque chapitre.

Chapitre 1

Introduction

Dans ce chapitre, nous présentons quelques définitions, notations et résultats préliminaires. Cette thèse s'articule autour de l'idée suivante : le développement d'outils statistiques pour l'étude des processus gaussiens indexés par des graphes doit s'inspirer à la fois des outils existants pour $\mathbb{Z}$ (graphe infini le plus simple que l'on puisse considérer), et de méthodes multidimensionelles, existant essentiellement pour des champs indexés par des variétés.
De nombreux outils ont déjà été généralisés au cas $\mathbb{Z}^d, d \geq 1$, ainsi qu'à d'autres structures de graphes possédant de nombreuses symétries. Nous nous proposons ici de poursuivre ces extensions, en étudiant les modèles $ARMA$. Cette idée permet de s'affranchir d'outils algébriques, en les remplaçant par des outils hilbertiens. Ainsi, certaines définitions et propriétes pourront être étendues à des graphes non nécessairement symétriques.
Nous présentons ici toutes les définitions et les résultats qui ont été utilisés dans notre travail.

1.1 Prédiction

Notre objectif est de développer des outils de prédiction pour des processus gaussiens indexés par des graphes. Dans cette section, nous présentons le problème de filtrage sous le point de vue le plus général, et rappelons la définition et quelques propriétés élémentaires du meilleur prédicteur linéaire.
Soit G un ensemble fini ou dénombrable quelconque. Soit $\mathbf{X}$ un processus gaussien **centré** (non dégénéré) indexé par cet ensemble :

$$\mathbf{X} := (X_i)_{i \in G} .$$

Dans tout le mémoire, on note, pour tout sous-ensemble G' de G,

$$X_{G'} := (X_i)_{i \in G'} ,$$

la restriction du processus $\mathbf{X}$ aux positions $i \in G'$.
En toute généralité, le problème de prédiction se présente de la façon suivante. Soient O, M deux sous-ensembles (finis ou dénombrables) de G. On suppose que l'on observe le processus $\mathbf{X}$ en toutes les positions $j \in O$, et on cherche le meilleur prédicteur de X_M (i.e. de $\mathbf{X}$ aux positions $i \in M$) en fonction des observations X_O. Autrement dit, on cherche la statistique $\hat{X}_M$, solution du problème d'optimisation suivant :

$$\hat{X}_M := \arg \min_{Z_M \in \sigma(X_O)^M} \mathbb{E}\left[(Z_M - X_M)^2\right],$$

où $\sigma(X_O)$ désigne la tribu engendrée par X_O.
Par définition de l'espérance conditionnelle, le meilleur prédicteur de X_M observant X_O n'est autre que l'espérance conditionnelle de X_M sachant X_O :

$$\hat{X}_M = \mathbb{E}\left[X_M | X_O\right].$$

Dans le cas gaussien centré, ce prédicteur s'exprime explicitement comme une fonction linéaire des observations.
Pour donner son expression, notons, pour tout G_1, G_2 sous-ensembles de G, $\langle X_{G_1}, X_{G_2} \rangle$ l'opérateur (ou la matrice) de covariance entre X_{G_1} et X_{G_2}. Lorsque O, M sont finis, on a alors

$$\hat{X}_M = \langle X_M, X_O \rangle \Big(\langle X_O, X_O \rangle \Big)^{-1} X_O.$$

En outre, dans le cas gaussien, le mode conditionnel coincide avec l'espérance conditionnelle. Cela nous permet de remarquer que le meilleur prédicteur est aussi le plus vraisemblable :

$$\hat{X}_M = \arg \max_{Z_M \in \mathbb{R}^M} - \begin{bmatrix} Z_M^T & X_O^T \end{bmatrix} \Big(\langle X_{M \cup O}, X_{M \cup O} \rangle \Big)^{-1} \begin{bmatrix} Z_M \\ X_O \end{bmatrix}.$$

Enfin, d'un point de vue prédiction, modéliser les données comme une réalisation d'un processus gaussien revient à supposer une forme de régularité sur celle-ci. En effet, on peut trouver une formulation équivalente à tout ce qui précède dans le cadre des RKHS [6]. Nous n'introduisons pas ici tout le formalisme correspondant (voir par exemple [2]). Remarquons seulement que si l'opérateur de covariance de $\mathbf{X}$ possède un inverse continu, alors cet inverse fournit un produit scalaire permettant de construire l'espace de Hilbert des fonctions $l^2(G)$ (pour ce produit scalaire). Si le processus est assez régulier, cet espace de Hilbert est un RKHS (admettant l'opérateur de covariance comme noyau).
Dès lors, les résultats précédents peuvent être interprétés de façon entièrement déterministe, comme

– Un problème de projection orthogonale pour le point de vue

$$\hat{X}_M := \arg \min_{Z_M \in \sigma(X_O)^M} \mathbb{E}\left[(Z_M - X_M)^2\right],$$

– Un problème de minimisation sous contrainte pour le point de vue

$$\hat{X}_M = \arg \max_{Z_M \in \mathbb{R}^M} - \begin{bmatrix} Z_M^T & X_O^T \end{bmatrix} \Big(\langle X_{M\cup O}, X_{M\cup O}\rangle\Big)^{-1} \begin{bmatrix} Z_M \\ X_O \end{bmatrix}.$$

Il paraît alors naturel d'introduire le problème de minimisation pénalisé par une fonction de coût l :

$$\hat{X} = \arg \min_{Z_M, Z_O \in \mathbb{R}^M \times \mathbb{R}^O} \begin{bmatrix} Z_M^T & Z_O^T \end{bmatrix} \Big(\langle X_{M\cup O}, X_{M\cup O}\rangle\Big)^{-1} \begin{bmatrix} Z_M \\ Z_O \end{bmatrix} + \lambda l\left(Z_O, X_O\right), \lambda > 0$$

Du point de vue des processus gaussiens, ce problème peut se justifier en considérant que les observations ont été bruitées. Il peut aussi être interprété comme un problème de régularisation de Tikhonov, décrit par exemple dans [34], [62]).
Revenons à la forme du meilleur prédicteur

$$\hat{X}_M = \langle X_M, X_O\rangle \left(\langle X_O, X_O\rangle\right)^{-1} X_O.$$

Son calcul effectif requiert la connaissance de l'opérateur de covariance de $\mathbf{X}$. Dans la pratique,
– Cet opérateur doit être estimé ;
– Son inverse doit être estimé ;
– Le prédicteur doit être calculable.
Chacun de ces points pose un problème. Pour pouvoir estimer l'opérateur de covariance, il faut fournir un modèle, qui dépend de la structure de G. Dans notre travail, nous nous intéressons au cas où G est un graphe. Le cas le plus simple $G = \mathbb{Z}$ est traité dans le chapitre 2, avec un modèle non paramétrique. Dans les chapitres suivants, nous étudions un modèle paramétrique de covariance pour des processus indexés par des graphes.
Une fois le modèle établi, il faut trouver un moyen de calculer (ou d'estimer) l'inverse de l'opérateur de covariance. Pour cela, plusieurs outils seront proposés, inspirés de méthodes classiques, que l'on étendra au cas des graphes.
Enfin, pour traiter les questions de calculabilité, on étendra sur ces modèles l'approximation de Whittle de la vraisemblance, qui permet de maximiser une fonction beaucoup plus rapide à calculer en pratique.

1.2 Modèles pour les opérateurs de covariance

Séries chronologiques

Présentons tout d'abord le cas des séries chronologiques, c'est-à-dire le cas $G = \mathbb{Z}$. On s'interesse à un processus stationnaire (toujours gaussien centré) $\mathbf{X} = (X_i)_{i\in\mathbb{Z}}$. Dans le cas gaussien, la stationnarité au sens strict (invariance des lois multidimensionnelles par translation du temps) est caractérisée par la stationnarité $\mathbb{L}^2$. Cela signifie que la covariance entre $X_i, X_{i+k}, i, k \in \mathbb{Z}$ ne dépend pas de i. On peut dès lors définir, pour tout $k \in \mathbb{Z}$ l'autocovariance r_k par

$$r_k := \mathrm{Cov}\,(X_0, X_k)\,.$$

Lorsque la suite $(r_k)_{k\in\mathbb{Z}}$ appartient à $l^2(\mathbb{Z})$, on peut aussi définir la série de Fourier associée à la suite $(r_k)_{k\in\mathbb{Z}}$ par

$$\forall t \in]-\pi, \pi[, f(t) := \sum_{k\in\mathbb{Z}} r_k e^{-ikt}.$$

Cette fonction mesurable $\mathbb{L}^2(]-\pi, \pi[)$ est paire et positive. Elle est appelée densité spectrale du processus.
L'opérateur de covariance Γ du processus $\mathbf{X}$ est donc un opérateur de Toeplitz (voir par exemple [23]) associé à la densité f. On note

$$\Gamma = \mathcal{T}(f).$$

Cela fournit un modèle non paramétrique pour la covariance, et des outils pour l'étudier.
Dans le premier chapitre, on s'intéresse au problème de prédiction aveugle. Autrement dit, à partir d'un unique échantillon fini $X_O, O \subset \mathbb{Z}$, on cherche un prédicteur $\hat{X}_M$ de données manquantes X_M $(M \subset \mathbb{Z})$. En réalité, il s'agit d'estimer l'opérateur de projection, puis de l'appliquer aux données sur lesquelles l'estimateur est consruit. La difficulté du problème aveugle est donc induite par la dépendance entre les données permettant de construire l'estimateur et les données utilisées pour la prédiction.

Extension aux graphes

Notre but a été ensuite d'étendre à des processus indexés par des graphes des résultats existant pour les séries chronologiques. La première étape est de fournir un modèle de covariance pour des champs aléatoires indexés par des graphes. Ce modèle sera obtenu par extension à des graphes quelconques d'une classe de champs stationnaires, définis sur des graphes symétriques dont on rappelle ici quelques exemples.

Définitions

De façon générale, on définit un graphe pondéré $\mathbf{G} = (G, W)$ par
– G désigne l'ensemble des sommets (au plus dénombrable).
– $W \in [-1, 1]^{G \times G}$ est un opérateur symétrique de poids.
L'opérateur W sera appelé opérateur d'adjacence (non pondéré) lorsque

$$\forall i, j \in G, W_{ij} \in \{0, 1\},$$

et opérateur *pondéré* d'adjacence sinon. Deux sommets $i, j \in G$ sont dits voisins ($i \sim j$) si, et seulement si, $W_{ij} \neq 0$.
On notera d la distance usuelle sur le graphe (pour $i, j \in G$, $d(i, j)$ désigne la longueur du plus court chemin allant de i à j). Le graphe sera supposé connexe :

$$\forall i, j \in G, d(i, j) < +\infty.$$

Le degré d'un sommet $i \in G$ est défini comme le nombre de voisins de i. Un graphe est dit q-régulier ($q \in \mathbb{N}$) si le degré de chaque sommet est égal à q.
Considerons maintenant l'action de l'opérateur pondéré d'adjacence sur $l^2(G)$ donnée par :

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

Supposons que le graphe est de degré borné, c'est-à-dire que le degré des sommets du graphe est borné. Alors l'opérateur pondéré d'adjacence est un opérateur hilbertien borné (voir par exemple [55], Theorème 3.1). Son spectre $\mathrm{Sp}(W)$ (i.e. l'ensemble des nombres complexes λ tels que $\lambda \,\mathrm{Id} - A$ n'est pas inversible, Id désignant l'identité sur $l^2(G)$) est un compact non vide de $\mathbb{R}$.
L'opérateur hilbertien W, est continu et normal (car symétrique). Il admet donc une représentation spectrale par rapport à une résolution de l'identité E (voir chapitre 3 pour la définition, et [58] pour plus de détails) donnée par

$$W = \int_{\mathrm{Sp}(W)} \lambda \mathrm{d}E(\lambda).$$

Cette représentation sera utile pour comprendre le lien entre les modèles MA_∞ que nous allons considérer et la représentation spectrale usuelle existante dans les cas classiques. Enfin, on appelle *automorphisme du graphe* $\mathbf{G}$ une permutation σ de l'ensemble des sommets G qui laisse W invariant :

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

Avant d'introduire notre modèle, rappelons rapidement les définitions dans le cas de $\mathbb{Z}^d$ et des arbres homogènes, de la représentation spectrale.

L'exemple de $\mathbb{Z}^d$

Le cas de $\mathbb{Z}^d$ est très proche de celui de $\mathbb{Z}$. Un processus gaussien centré $\mathbf{X} = (X_i)_{i\in\mathbb{Z}^d}$ est dit stationnaire si son opérateur de covariance vérifie :

$$\langle X_i, X_{i+k}\rangle, i, k \in \mathbb{Z}^d$$

ne dépend que de k. On peut alors définir, comme pour les séries chronologiques, le $k^{\text{ième}}$ coefficient de covariance par

$$r_k := \langle X_{0_{\mathbb{Z}^d}}, X_k\rangle,$$

ainsi que la densité spectrale associée, dès lors que $(r_k)_{k\in\mathbb{Z}^d} \in \mathbb{L}^2(\mathbb{Z}^d)$, par

$$\forall (t_1, \cdots, t_d) \in [0, 2\pi[^d, f(t_1, \cdots, t_d) = \sum_{(k_1,\cdots,k_d)\in\mathbb{Z}^d} r_{(k_1,\cdots,k_d)} e^{-i(k_1t_1+\cdots+k_dt_d)}.$$

Le cas de l'arbre q-régulier

Considérons maintenant le cas de l'arbre $(q+1)$-homogène $\mathcal{A}_{q+1}$. Ce graphe est défini comme le seul graphe infini connexe $(q+1)$-régulier sans cycles. On peut aussi définir une notion de stationnarité sur cette structure (voir par exemple [5].). Un processus $\mathbf{X} = (X_i)_{i\in\mathcal{A}_q}$ est dit stationnaire si son opérateur de covariance Γ vérifie

$$\exists \phi \in \mathbb{R}^{\mathbb{N}}, \forall i, j \in G, \Gamma_{ij} = \phi(d(i, j)).$$

Dans ce cas là, il existe encore une représentation spectrale de la covariance. En effet, il existe une unique mesure positive $\nu_{\mathbf{X}}$ (voir par exemple [5]) telle que :

$$\phi(n) = \int_{[-2\sqrt{q},2\sqrt{q}]} P_n(t)\mathrm{d}\nu_{\mathbf{X}}(t),$$

où $(P_n)_{n\in\mathbb{N}}$ est une suite de polynômes définie par
– $P_0(t) = 1$,
– $P_1(t) = t$,
– $\forall n \geq 1, (q+1)tP_n(t) = qP_{n+1}(t) + \frac{1}{q+1}P_{n-1}(t)$.
Notons que pour $q = 1$, on retrouve le cas de $\mathbb{Z}$, et les polynômes de Tchebychev (à un changement de variable près), voir Chapitre 5.
En réalité, la positivité de Γ équivaut à la positivité de $\nu_{\mathbf{X}}$ car cette représentation correspond à une représentation spectrale de l'opérateur de covariance Γ lui même (voir le chapitre 5).
Remarquons enfin que, à la fois dans le cas de $\mathbb{Z}^d$, celui des arbres, mais aussi pour tous les exemples classiques de la littérature (graphe distance-transitif, réseau...), la stationnarité $\mathbb{L}^2$ d'un processus est définie par l'invariance de son opérateur de covariance par tout automorphisme de graphe.

Dans la majorité de ces cas là, il se trouve que l'on peut retrouver tous les processus stationnaires assez réguliers (i.e. à courte mémoire) en considérant les modèles MA_∞. Plus de détails seront donnés sur des exemples dans le chapitre 5.

Modèles MA_∞

Soit **G** un graphe quelconque. Dans le chapitre 3, on considère des opérateurs de covariance $\Gamma = \mathcal{K}(f)$ définis à partir d'une fonction mesurable positive f régulière (par exemple analytique sur l'enveloppe convexe de $\mathrm{Sp}(W)$) de la façon suivante (relation qui définit l'opérateur $\mathcal{K}$) :

$$\Gamma = \mathcal{K}(f) := \int_{\mathrm{Sp}(W)} f(\lambda)\mathrm{d}E(\lambda).$$

L'égalité précédente peut aussi être interprétée au sens de la convergence normale de la série entière associée. Si $\forall x \in \mathrm{Sp(W)}, f(x) = \sum_{k\in\mathbb{N}} f_k x^k$, alors on a

$$\mathcal{K}(f) = \sum_{k\in\mathbb{N}} f_k W^k.$$

La notion de modèle MA_∞ s'interprète au travers de l'égalité précédente. Lorsque f est un polynôme de degré p, on parlera de modèle MA_p. De même, si $\frac{1}{f}$ est un polynôme de degré q, on parlera de modèle AR_q.
En fait, cela correspond à une façon naturelle d'étendre à un graphe quelconque les processus (réguliers) définis sur $\mathbb{Z}, \mathbb{Z}^d$, ou sur $\mathcal{A}_q$ à partir de modèles MA_∞ isotropes (et non causaux).
En effet, les représentations spectrales obtenues ci-dessus pour $\mathbb{Z}^d$ et $\mathcal{A}_q$ correspondent à celle d'un processus MA_∞ sur le graphe correspondant. Cette assertion est détaillée à travers les exemples du Chapitre 5. C'est encore le cas pour les graphes distance-transitifs (voir [65] pour la définition et la forme générale de ces graphes). En effet, toute fonction invariante par les automorphismes, et suffisamment régulière peut s'écrire sous cette forme (voir [65]).
Le choix d'un tel modèle sur un graphe quelconque s'explique donc par une volonté d'étendre cette notion, et les outils spectraux correspondants, au cas non transitif. En fait, ces modèles MA sont liés à l'algèbre d'adjacence [14] du graphe. La représentation spectrale découle alors de l'étude de cette algèbre (pour plus de détails, voir chapitre 5).

Cas anisotrope

La construction précédente peut-être généralisée à des processus spatio-temporels. C'est l'objet d'une partie appliquée de notre travail développée au Chapitre 4. En

réalité, il suffit de comprendre que dans le cas le plus général, un processus stationnaire en temps (discret : $\mathbb{Z}$) et MA en espace (graphe : $\mathbf{G}$) peut se représenter comme un processus indexé par $\mathbb{Z} \times G$, stationnaire, mais anisotrope.
C'est pourquoi nous introduisons, à partir d'un modèle *jouet*, des processus MA anisotropes. Dans le chapitre 4, ils sont construits sur $\mathbb{Z} \times G$, mais la construction est la même sur $G_1 \times G_2$, si $\mathbf{G}_1$ et $\mathbf{G}_2$ sont des graphes (de degré borné) quelconques. Notons E^{G_1} et E^{G_2} les résolutions de l'identité associées aux opérateurs d'ajacence W^{G_1} et W^{G_2}. Alors le modèle général de processus MA anisotropes indexés par $G_1 \times G_2$ s'écrit (pour ψ analytique sur l'enveloppe convexe de $\mathrm{Sp}(G_1) \times \mathrm{Sp}(G_2)$) :

$$\mathcal{K}_{st}(\psi) := \int_{\mathrm{Sp}(G_1)\times\mathrm{Sp}(G_2)} \psi(x,t) \mathrm{d}E^{G_1}(x) \otimes \mathrm{d}E^{G_2}(t).$$

Si on écrit

$$\forall (x,t) \in \mathrm{Sp}(G_1) \times \mathrm{Sp}(G_2), \psi(x,t) = \sum_{k,l\in\mathbb{N}} \psi_{kl} x^k t^l,$$

cette construction peut-être encore interprétée comme une moyenne mobile infinie anisotrope :

$$\mathcal{K}_{st}(\psi) = \sum_{k,l\in\mathbb{N}} \psi_{kl} \left(W^{G_1}\right)^k \otimes \left(W^{G_2}\right)^l,$$

où $\otimes$ désigne le produit tensoriel.
On peut retrouver ainsi la construction correspondant à $\mathbb{Z}^d$. De plus, remarquons maintenant que

– Si le processus est MA au sens précédent sur le produit cartésien $G_1 \times G_2$ (souvent noté $G_1 \square G_2$, voir par exemple [39]), alors il existe une fonction f qui factorise ψ par
$$\psi(x,t) = f(x+t).$$
– Si le processus est MA sur le produit de Kronecker $G_1 \otimes G_2$ (voir par exemple [39] pour la définition), alors il existe une fonction f qui factorise ψ par
$$\psi(x,t) = f(xt).$$

Ce modèle de covariance est utilisé dans toute la thèse. Dans les chapitres 3 et 4, on se donnera un modèle paramétrique de densités dans un espace de fonctions bien choisi, qui fournira un modèle paramétrique d'opérateurs de covariance.

1.3 Estimation

Une fois les modèles établis, on cherche à estimer l'opérateur de covariance et son inverse, afin d'estimer le meilleur prédicteur linéaire.
Deux méthodes sont proposées ici, l'une dans un cadre non paramétrique, l'autre dans un cadre paramétrique.

Séries chronologiques

La première méthode, utilisée dans le chapitre 2, permet d'estimer l'opérateur de covariance d'une série chronologique stationnaire. Elle repose sur un résultat de concentration dû à Comte [28] sur les coefficients empiriques non biaisés de covariance.
S'inspirant ensuite d'articles de Bickel et Levina [13], [12] et de Dahlhaus [29], on utilise l'estimateur empirique non biaisé de la covariance pour estimer l'inverse de cet opérateur.
Bickel prouve que lorsque l'estimation est faite sur plusieurs échantillons, régulariser l'inverse de l'estimateur empirique non biaisé de la covariance permet d'estimer de manière consistante l'inverse de l'opérateur de covariance.
Dans le cadre de la prédiction aveugle, où un seul échantillon est disponible, on estime l'opérateur de covariance sur une petite partie des données paramétrée par une sous-fenêtre. Cela permet d'estimer aussi correctement l'inverse de l'opérateur de covariance (après une étape de régularisation). Cela mène à un prédicteur n'utilisant qu'une partie des données observées. Le résultat principal de la première partie donne le taux de convergence de ce prédicteur.
Plus précisemment, on suppose que l'on observe $X_{-N}, \cdots, X_{-1}, N \in \mathbb{N}$. On utilise l'ensemble de ces données pour l'estimation de la covariance et de son inverse, et seulement une partie $X_{-K(N)}, \cdots X_{-1}$ pour la prédiction, pour une suite $(K(N))_{N\in\mathbb{N}}$ bien choisie. Cela permet de construire un prédicteur empirique $\hat{P}^{(N)}_{O_K B_K}$ des données $B_K = \left\{X_1, \cdots X_{K(N)}\right\}$ (futur proche) en fonctions des observations $O_K = \left\{X_{-K(N)}, \cdots X_{-1}\right\}$. Alors, on prouve le théorème suivant :

Théorème 1 *Sous des hypothèses de régularité sur le processus et sur la densité spectrale (type Sobolev), et pour N assez grand, on a*

$$\sqrt{\mathcal{R}(\hat{P}^{(N)}_{O_K B_K})} \leq C_1 \frac{K(N)^2\sqrt{\log(K(N))}}{\sqrt{N}} + C_2 \frac{1}{K(N)^{\frac{2s-1}{2}}},$$

Les constantes C_1 et C_2, les hypothèses exactes, et la définition du risque $\mathcal{R}$ utilisé apparaissent dans le Chapitre 2.

Ce théorème, ainsi que les lemmes donnant indépendamment le contrôle du biais et de la variance de ce prédicteur empirique, permettent en pratique de choisir la suite $(K(N))_{N\in\mathbb{N}}$ de façon à ce que le risque ci-dessus soit de l'ordre de grandeur de l'innovation.
Nous donnons donc ici une justification théorique de procédés pratiques très utilisés. En effet, il est courant d'estimer souvent l'opérateur de covariance sur l'échantillon même qui servira à la prédiction. Nous montrons en particulier, sous des hypothèses de faible dépendance, qu'il est raisonnable d'utiliser le même échantillon.

En effet, l'estimateur empirique du meilleur prédicteur linéaire est consistant dès que l'on utilise assez de données pour l'estimation.
Remarquons enfin que dans le cadre d'un modèle paramétrique, on sait que, au moins pour des densités spectrales assez régulières, l'estimateur de maximum de vraisemblance est consistant. Nous ne donnons pas plus de détails ici, mais notre extension aux graphes du cas des séries chronologiques s'est inspirée de la preuve de Azencott et Dacunha-Castelle [8].

Extension aux graphes

Dans les chapitres suivants (3 et 4), on considère un graphe $\mathbf{G}$ et on se donne un modèle paramérique de fonctions $(f_\theta)_{\theta\in\Theta}$, où Θ est un intervalle compact de $\mathbb{R}^d, d \geq 1$. Ces fonctions sont supposées analytiques sur l'enveloppe convexe de $\mathrm{Sp}(W)$.
Cela fournit un modèle paramétrique d'opérateurs de covariance $(\mathcal{K}(f_\theta))_{\theta\in\Theta}$. Soit $\theta_0 \in \Theta$, et $\mathbf{X} = (X_i)_{i\in G}$ un champ gaussien centré de covariance $\mathcal{K}(f_{\theta_0})$. On suppose que l'on observe ce processus sur une suite de sous-graphes $(G_n)_{n\in\mathbb{N}}$.

$$X_{G_n} := (X_i)_{i\in G_n} .$$

Si on note $\mathcal{K}_n(f_{\theta_0})$ la covariance du processus X_n, on peut estimer θ_0 par maximum de vraisemblance. C'est l'objet du chapitre 3. Dans le chapitre 4, ce modèle paramétrique est étendu au cas spatio-temporel.
Sous des hypothèses de régularité sur la classe de fonctions, sur le graphe, et sur la suite de sous-graphes sur lequel le processus est observé, on obtient la consistance de l'estimateur de maximum de vraisemblance.
De plus, ce résultat reste valide dans le cas anisotrope, et donc dans le cas spatio-temporel.

1.4 Calcul

Les problèmes de calculabilité numérique posés dans le premier chapitre et dans les suivants ne sont pas du même ordre.
Dans le chapitre 2, le calcul de la vitesse de convergence nécessite de savoir calculer le prédicteur sur le passé infini. Plus précisemment, pour prouver la convergence de l'estimateur empirique du meilleur prédicteur linéaire, il faut obtenir une autre expression de ce prédicteur.
Dans les chapitres suivants, la question se pose pour des soucis de calcul numérique. On souhaite fournir un estimateur rapide à calculer en pratique. Pour cela, on va étendre l'approximation de Whittle au cas des graphes.

Séries chronologiques

Dans le problème de prédiction aveugle de séries chronologiques, pour calculer le biais du prédicteur par rappport au prédicteur sur le passé infini, il faut utiliser une expression alternative du prédicteur sur le passé infini, permettant d'inverser une matrice finie au lieu d'un opérateur hilbertien. Cette expression est fournie par une inversion de Schur par blocs, donnant :

Proposition 1 *Soit Γ l'opérateur de covariance de la série chronologique $\mathbf{X}$ et Λ son inverse. Le meilleur prédicteur linéaire sur $(X_i)_{i\in A}$ ($A \subset \mathbb{Z}$) peut s'écrire des deux façons suivantes :*

$$\begin{aligned} p_A &= \begin{bmatrix} Id_A & \Gamma_A^{-1}\Gamma_{AA^C} \\ 0 & 0 \end{bmatrix}. \\ &= \begin{bmatrix} \mathrm{Id}_A & -\Lambda_{AA^C}\Lambda_{A^C}^{-1} \\ 0 & 0 \end{bmatrix} \end{aligned}$$

L'erreur de prédiction est donnée par

$$\mathbb{E}\left[(P_A Y - Y)^2\right] = u^T \Lambda_{A^C}^{-1} u,$$

où $Y = u^T X$. (voir chapitre 2 pour plus de détails sur les notations).

L'inversion par bloc utilisant le complément de Schur peut être utilisée en pratique pour la prédiction sur une longue séquence de données. Lorsque $A = \mathbb{Z}^-$, l'expression et l'erreur obtenues sont celles fournies par Bondon dans [16]. Calculer l'opérateur correspondant plutôt que ses coefficients permet une généralisation à des sous-ensembles quelconques $A \subset \mathbb{Z}$.

Extension aux graphes

Intéressons-nous maintenant au cadre paramétrique. Dans le cas des séries chronologiques comme dans le cas de champs aléatoires indexés par des graphes, il est difficile en pratique de calculer l'estimateur du maximum de vraisemblance (temps de calcul, instabilité de l'inverse...).
Il requiert en effet le calcul d'un déterminant et l'inversion d'une matrice, pour maximiser l'expression :

$$L_n(\theta) := -\frac{1}{2}\left(\sharp G_n \log(2\pi) + \log\det\left(\mathcal{K}_n(f_\theta)\right) + X_n^T\left(\mathcal{K}_n(f_\theta)\right)^{-1} X_n\right).$$

Dans le cas des séries chronologiques, pour éviter ces étapes lourdes en temps de calcul, on utilise en pratique les approximations du déterminant et de l'inverse de l'opérateur de covariance suivantes :

Si $\mathcal{T}(f_\theta)$ désigne l'opérateur de Toeplitz associé à f_θ,

$$\frac{1}{N}\log\det\left(\mathcal{T}_N(f_\theta)\right) \approx \frac{1}{\pi}\int_{[-\pi,\pi]}\log\left(f_\theta(\lambda)\right)\mathrm{d}\lambda.$$

$$\frac{1}{N}X_N^T\Big(\mathcal{T}_N(f_\theta)\Big)^{-1}X_N \approx \frac{1}{N}X_N^T\mathcal{T}_N(\frac{1}{f_\theta})X_N.$$

Cette approximation est due à Whittle [67], et apparait de très nombreuses fois dans la litterature ([68], [9], [10]).
On peut faire la même approximation dans le cas des graphes, avec le modèle de covariance donné par l'opérateur $\mathcal{K}$ introduit précédemment.
Pour cela, il est nécessaire d'introduire une mesure spectrale μ, dépendant du graphe $\mathbf{G}$ ainsi que de la suite de sous-graphes $(G_n)_{n\in\mathbb{N}}$ sur lesquels le processus est observé. Cette mesure est la limite de la mesure locale moyenne à la racine lorsqu'on enracine le sous-graphe G_n uniformément. Le Chapitre 5 donne des détails sur cette construction.
L'existence de cette mesure spectrale dépend à la fois d'une forme d'homogéné-ité sur le graphe G, et de l'exploration de cette homogénéité par la suite des sous-graphes $(G_n)_{n\in\mathbb{N}}$. Plus de détails seront donnés dans les chapitres 3 et 5.
La mesure spectrale globale μ apparait sous differentes formes dans la littérature (voir [45], [65], [5], et l'article de revue de Mohar et Woess [55] qui utilise la même approche).
Cela mène à une vraisemblance approchée $\tilde{L}_n$ définie par

$$\tilde{L}_n(\theta) := -\frac{1}{2}\left(\sharp G_n\log(2\pi) + \sharp G_n\int\log(f_\theta(x))\mathrm{d}\mu(x) + X_n^T\left(\mathcal{K}_n\left(\frac{1}{f_\theta}\right)\right)X_n\right).$$

On définit ensuite l'estimateur du maximum de vraisemblance approchée par

$$\tilde{\theta}_n := \arg\max\tilde{L}_n(\theta).$$

Le principal théorème de la partie 3 stipule que cet estimateur de vraisemblance approchée est consistant, sous certaines hypothèses de régularité et de géométrie des graphes considérés.

Théorème 2 *Sous des hypothèses de géométrie sur les graphes, et de régularité sur les fonctions considérées (voir chapitre 4), la suite $(\tilde{\theta}_n)_{n\in\mathbb{N}}$ converge, avec n, $P_{f_{\theta_0}}$-p.s. vers la vraie valeur θ_0.*

La preuve de ce théorème repose essentiellement sur une extension d'un théorème de Szegö [41], qui fournit un contrôle sur la quantité

$$\mathrm{Tr}\left(\mathcal{T}_N(f)\mathcal{T}_N(g) - \mathcal{T}_N(fg)\right).$$

Ce contrôle est donné en fonction de la régularité des fonctions f, g.
Dans le Chapitre 3, s'inspirant de la version de [8] de ce théorème (où beaucoup de régularité sur les fonctions est nécessaire), nous étendons ce résultat au cas des graphes, obtenant un contrôle de la quantité

$$\mathrm{Tr}\left(\mathcal{K}_n(f)\mathcal{K}_n(g) - \mathcal{K}_n(fg)\right).$$

Des hypothèses supplémentaires sur la suite de sous-graphes sont exigées. La différence $\mathrm{Tr}\left(\mathcal{K}_n(f)\mathcal{K}_n(g) - \mathcal{K}_n(fg)\right)$ apparait en effet comme une mesure des effets de bord (voir la preuve dans le chapitre 4 pour plus de détails).
Nous avons donc besoin ici d'hypothèses sur les bords de G_n. Plus précisémment, la validité du théorème requiert, outre des hypothèses de régularité sur les fonctions f, g, que le ratio $\frac{\sharp \delta G_n}{\sharp G_n}$ tende vers 0 (ici δG_n désigne le bord de G_n, i.e. les points de G_n voisins de $G \setminus G_n$). Il faut en particulier que le graphe soit amenable (i.e. une suite de sous-graphes $(G_n)_{n\in\mathbb{N}}$ vérifiant $\frac{\sharp \delta G_n}{\sharp G_n} \to 0$ existe). Cela exclut malheureusement les arbres de notre travail.
Encore une fois, ceci peut se généraliser au cas anisotrope, et donc au cas spatio-temporel. Bien que nous n'ayons pas encore écrit toutes les preuves (seulement le lemme de Szegö est écrit dans le cas spatio-temporel, voir chapitre 4), elles s'adaptent parfaitement.

Remarquons maintenant que, dans le chapitre 2, l'estimateur utilisé n'est pas la covariance empirique classique, mais une version non biaisée.
Pour l'estimation par maximum de vraisemblance, on peut aussi utiliser une version "déformée" du périodogramme permettant d'obtenir un score asymptotiquement non biaisé. C'est ce que nous faisons dans le cas des graphes. D'une certaine façon, cette construction ressemble à celle du périodogramme conique (ou raboté). Pour cela, on s'inspire de travaux de Guyon [42], [44] et Dahlhaus et Kunsch [30]. Dans le chapitre 3, on construit ce périodogramme conique $X_n^T \mathcal{Q}_n(g) X_n$ pour des graphes possédant des invariances algébriques.
On peut alors définir une nouvelle approximation de la vraisemblance

$$L_n^{(u)}(\theta) := -\frac{1}{2}\left(\sharp G_n \log(2\pi) + \sharp G_n \int \log(f_\theta(x))\mathrm{d}\mu(x) + X_n^T\left(\mathcal{Q}_n(\frac{1}{f_\theta})\right)X_n\right).$$

Notons

$$\theta_n^{(u)} := \arg\max L_n^{(u)}(\theta).$$

On obtient alors

Théorème 3 *Dans les cas AR_P ou MA_P ($P < +\infty$), et sous des hypothèses de géométrie pour les graphes, et de régularité sur les fonctions (voir Chapitre 3), l'estimateur $\theta_n^{(u)}$ de θ_0 est convergent et asymptotiquement normal :*

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

De plus, l'information de Fisher asymptotique (au sens de la croissance de la suite de sous-graphes $(G_n)_{n\in\mathbb{N}}$) est donnée par :

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

Par conséquent, l'estimateur est asymptotiquement efficace

L'extension de ce théorème à des classes plus générales de graphes est un problème techniquement difficile, et fait partie des perspectives importantes. En particulier, la construction proposée ici est encore difficile à mettre en oeuvre en pratique. Le calcul numérique de cet objet étant un de nos principaux objectifs, nous espérons obtenir dans nos prochains travaux une expression plus simple à mettre en oeuvre, et s'adaptant à des classes plus générales de graphes.

Chapitre 2

Estimation error for blind Gaussian time series prediction

Dans ce chapitre, on s'intéresse au problème de prédiction aveugle de séries chronologiques gaussiennes. On construit un projecteur empirique sur les observations en injectant l'operateur de covariance empirique dans une décomposi-tion de Schur du projecteur sur les observations. L'erreur de prédiction est ensuite calculée, en fonction de la taille de l'échantillon utilisé pour la prédiction, et de la régularité de la densité spectrale [1].

Introduction

In many concrete situations the statistician observes a finite path $X_1, \ldots, X_n$ of a real temporal phenomenon which can be modeled as realizations of a stationary process $\mathbf{X} := (X_t)_{t\in\mathbb{Z}}$ (we refer, for example, to [23], [60] and references therein). Here we consider a second order weakly stationary process, which implies that its mean is constant and that $\mathbb{E}(X_t X_s)$ only depends on the distance between t and s. In the sequel, we will assume that the process is Gaussian, which implies that it is also strongly stationary, in the sense that, for any $t, n \in \mathbb{Z}$,

$$(X_1, \cdots, X_n) \stackrel{\mathcal{L}}{=} (X_{t+1}, \cdots, X_{t+n}),\ (t \in \mathbb{Z}, n \in \mathbb{N}).$$

Our aim is to predict this series when only a finite number of past values are observed. Moreover, we want a sharp control of the prediction error. For this, recall that, for Gaussian processes, the best predictor of $X_t, t \geq 0$, when observing $X_{-N}, \cdots, X_{-1}$, is obtained by a suitable linear combination of the $(X_i)_{i=-N,\cdots,-1}$. This predictor, which converges to the predictor onto the infinite past, depends

1. Published in MMS : [35]

on the unknown covariance of the time series. Thus, this covariance has to be estimated. Here, we are facing a blind filtering problem, which is a major difficulty with regards to the usual prediction framework.
Kriging methods often impose a parametric model for the covariance (see [61], [8], [60]). This kind of spatial prediction is close to our work. Nonparametric estimation may be done in a functional way (see [15], [3], [4]). This approach is not efficient in the blind framework. Here, the blind problem is bypassed using an idea of Bickel and Levina [13] for the estimation of the inverse of the covariance. He shows that the inverse of the empirical estimate of the covariance is a good choice when many samples are at hand.
We propose in this chapter a new methodology, when only a path of the process is observed. For this, following Comte [28], we build an accurate estimate of the projection operator. Finally this estimated projector is used to build a predictor for the future values of the process. Asymptotic properties of these estimators are studied.
The chapter falls into the following parts. In Section 2.1, definitions and technical properties of time series are given. Section 2.2 is devoted to the construction of the empirical projection operator whose asymptotic behavior is stated in Section 2.3. Finally, we build a prediction of the future values of the process in Section 2.4. All the proofs are gathered in Section 2.5.

2.1 Notations and preliminary definitions

In this section, we present our general frame, and recall some basic properties about time series, focusing on their predictions.
Let $\mathbf{X} = (X_k)_{k\in\mathbb{Z}}$ be a zero-mean Gaussian stationary process. Observing a finite past $X_{-N}, \cdots, X_{-1}$ $(N \geq 1)$ of the process, we aim at predicting the present value X_0 without any knowledge on the covariance operator.
Since X is stationary, let $r_{i-j} := \mathrm{Cov}(X_i, X_j), (i, j \in \mathbb{Z})$ be the covariance between X_i and X_j. Here we will consider short range dependent processes, and thus we assume that

$$\sum_{k\in\mathbb{Z}} r_k^2 < +\infty,$$

So that there exists a measurable function $f^\star \in \mathbb{L}_2([0, 2\pi))$ defined by

$$f^\star(t) := \sum_{k=-\infty}^{\infty} r_k e^{ikt}, (a.e.)$$

This function is the so-called spectral density of the time series. It is real, even and non negative. As $\mathbf{X}$ is Gaussian, the spectral density conveys all the information

on the process distribution.

Define the covariance operator Γ of the process $\mathbf{X}$, by setting

$$\forall i, j \in \mathbb{Z}, \Gamma_{ij} = \mathrm{Cov}(X_i, X_j).$$

Note that Γ is the Toeplitz operator associated to $f^\star$. It is usually denoted by $\mathcal{T}(f^\star)$ (for a thorough overview on the subject, we refer to [20]). This Hilbertian operator acts on $l^2(\mathbb{Z})$ as follows

$$\forall u \in l^2(\mathbb{Z}), i \in \mathbb{Z}, (\Gamma u)_i := \sum_{j\in\mathbb{Z}} \Gamma_{ij} u_j = \sum_{j\in\mathbb{Z}} r_{i-j} u_j = (\mathcal{T}(f^\star)u)_i.$$

For sake of simplicity, we shall from now denote Hilbertian operators as infinite matrices.
Recall that for any bounded Hilbertian operator A, the spectrum $\mathrm{Sp}(A)$ is defined as the set of complex numbers λ such that $\lambda \,\mathrm{Id} - A$ is not invertible (here Id stands for the identity on $l^2(\mathbb{Z})$).
The spectrum of any Toeplitz operator, associated with a bounded function, satisfies the following property (see, for instance [23]) :

$$\forall f \in \mathbb{L}_\infty\left([0, 2\pi)\right), \mathrm{Sp}(\mathcal{T}(f)) \subset \left[\min(f), \max(f)\right].$$

Now consider the main assumption of this chapter :

Assumption 2.1.1

$$\exists m, m' > 0, \forall t \in [0, 2\pi), m < f^\star(t) < m'.$$

This assumption ensures the invertibility of the covariance operator, since $f^\star$ is bounded away from zero (shot range dependancy). As a positive definite operator, we can define its square-root $\Gamma^{\frac{1}{2}}$. Let Q be any linear operator acting on $l^2(\mathbb{Z})$, consider the operator norm $\|Q\|_{2,op} := \sup_{u\in l_2(\mathbb{Z}), \|u\|_2=1} \|Qu\|_2$, and define the warped operator norm as

$$\|Q\|_\Gamma := \sup_{u\in l_2(\mathbb{Z}), \left\|\Gamma^{\frac{1}{2}}u\right\|_2=1} \left\|\Gamma^{\frac{1}{2}}Qu\right\|_2.$$

Note that, under Assumption (2.1.1) $\|\Gamma\|_{2,op} \leq m'$, hence the warped norm $\|.\|_\Gamma$ is well defined and equivalent to the classical one

$$\frac{m}{m'}\|Q\|_{2,op} \leq \|Q\|_\Gamma \leq \frac{m'}{m}\|Q\|_{2,op}.$$

Finally, both the covariance operator and its inverse are continuous with respect to the previous norms.

The warped norm is actually the natural induced norm over the Hilbert space

$$H = (l_2(\mathbb{Z}), \langle ., . \rangle_\Gamma),$$

where

$$\langle x, y \rangle_\Gamma := x^T \Gamma y = \sum_{i,j \in \mathbb{Z}} x_i \Gamma_{ij} y_j.$$

From now on, all the operators are defined on H. Set

$$\mathbb{L}^2(\mathbb{P}) := \left\{ Y \in \overline{\text{Span}}\left((X_i)_{i\in\mathbb{Z}}\right), \mathbb{E}[Y^2] < +\infty \right\}$$

The following proposition (see for instance [23]) shows the particular interest of H :

Proposition 2.1.2 *The map*

$$\begin{aligned} \Phi : \quad & H \to \mathbb{L}^2(\mathbb{P}) \\ & u \to u^T \mathbf{X} = \sum_{i\in\mathbb{Z}} u_i X_i. \end{aligned}$$

defines a canonical isometry between H and $\mathbb{L}_2(\mathbb{P})$.

The isometry will enable us to consider, in the proofs, alternatively sequences $u \in H$ or the corresponding random variables $Y \in \mathbb{L}_2(\mathbb{P})$.

We will use the following notations : recall that Γ is the covariance operator and denote, for any $A, B \subset \mathbb{Z}$, the corresponding minor (A, B) by

$$\Gamma_{AB} := (\Gamma_{ij})_{i\in A, j\in B}.$$

Note that, when A and B are finite, Γ_{AB} is the covariance matrix between $(X_i)_{i\in A}$ and $(X_j)_{j\in B}$. Diagonal minors will be simply written $\Gamma_A := \Gamma_{AA}$, for any $A \in \mathbb{Z}$. In our prediction framework, let $O \subset \mathbb{Z}$ and assume that we observe the process $\mathbf{X}$ at times $i \in O$. It is well known that the best linear prediction of a random variable Y by observed variables $(X_i)_{i\in O}$ is also the best prediction, defined by $P_O(Y) := \mathbb{E}[Y|(X_i)_{i\in O}]$. Using the isometry, there exist unique $u \in H$ and $v \in H$ with $Y = \Phi(u)$ and $P_O(Y) = \Phi(v)$. Hence, we can define a projection operator acting on H, by setting $p_O(u) := v$. This corresponds to the natural projection in H onto the set $\{\Phi^{-1}(X_i),\ i \in O\}$. Note that this projection operator may be written by block

$$p_O u := \begin{bmatrix} \Gamma_O^{-1} \Gamma_{O\mathbb{Z}} \\ 0 \end{bmatrix} u.$$

The operator Γ_O^{-1} is well defined since $f^\star \geq m > 0$. Finally, the best prediction observing $(X_i)_{i\in O}$ is

$$\mathbb{E}\left[Y = \Phi(u)|(X_i)_{i\in O}\right] = P_O(\Phi(u)) = \Phi(p_O u).$$

This provides an expression of the projection when the covariance Γ is known. Actually, in many practical situations, Γ is unknown and need to be estimated from the observations. Recall that we observe $X_{-N}, \ldots, X_{-1}$. We will estimate the covariance with this sample and use a subset of these observations for the prediction. This last subset will be $\left\{(X_i)_{i\in O_{K(N)}}\right\}$, with $O_{K(N)} := [-K(N), \cdots -1]$. Here $(K(N))_{N\in\mathbb{N}}$ is a suitable growing sequence. Hence, the predictor $\hat{Y}$ will be here

$$\hat{Y} = \hat{P}_{O_{K(N)}} Y,$$

where $\hat{P}_{O_{K(N)}}$ denotes some estimator of the projection operator onto $O_{K(N)}$, built with the full sample $(X_i)_{i=-N,\cdots,-1}$.
As usual, we estimate the accuracy of the prediction by the quadratic error

$$\mathrm{MSE}(\hat{Y}) = \mathbb{E}\left[\left(\hat{Y} - Y\right)^2\right].$$

The bias-variance decomposition gives

$$\begin{aligned}\mathbb{E}\left[\left(\hat{Y} - Y\right)^2\right] &= \mathbb{E}\Big[\left(\hat{P}_{O_{K(N)}} Y - P_{O_{K(N)}} Y\right)^2\Big] + \mathbb{E}\Big[\left(P_{O_{K(N)}} Y - P_{\mathbb{Z}^-} Y\right)^2\Big] \\ &\quad + \mathbb{E}\Big[\left(P_{\mathbb{Z}^-} Y - Y\right)^2\Big],\end{aligned}$$

where

$$\hat{P}_{O_{K(N)}} Y = \hat{Y},$$

$$P_{O_{K(N)}} Y = \mathbb{E}\Big[Y|(X_i)_{i\in O_{K(N)}}\Big],$$

and

$$\hat{P}_{\mathbb{Z}^-} Y = \mathbb{E}\Big[Y|(X_i)_{i<0}\Big].$$

This error can be divided into three terms

- The last term $\mathbb{E}\Big[\left(P_{\mathbb{Z}^-} Y - Y\right)^2\Big]$ is the prediction with infinite past error. It is induced by the variance of the unknown future values, and may be easily computed using the covariance operator. This variance does not go to zero as N tends to infinity. It can be seen as an additional term that does not depend on the estimation procedure and thus will be omitted in the error term.
- The second term $\mathbb{E}\Big[\left(P_{O_{K(N)}} Y - P_{\mathbb{Z}^-} Y\right)^2\Big]$ is a bias induced by the temporal threshold on the projector.

– The first term $\mathbb{E}\Big[\left(\hat{P}_{O_{K(N)}}Y - P_{O_{K(N)}}Y\right)^2\Big]$ is a variance, due to the fluctuations of the estimation, and decreases to zero as soon as the estimator is consistent. Note that to compute this error, we have to handle the dependency between the prediction operator and the variable Y we aim to predict.

Finally, the natural risk is obtained by removing the prediction with infinite past error :

$$\begin{aligned} R(\hat{Y} = \hat{P}_{O_{K(N)}}Y) &:= \mathbb{E}\Big[\left(\hat{P}_{O_{K(N)}}Y - P_{O_{K(N)}}Y\right)^2\Big] + \mathbb{E}\Big[\left(P_{O_{K(N)}}Y - P_{\mathbb{Z}_*^-}Y\right)^2\Big] \\ &= \mathbb{E}\left[\left(\hat{Y} - \mathbb{E}\left[Y|(X_i)_{i<0}\right]\right)^2\right]. \end{aligned}$$

The global risk will be computed by taking the supremum of $R(\hat{Y})$ among of all random variables Y in a suitable set (growing with N). This set will be defined in the next section.

2.2 Construction of the empirical projection operator

Recall that the expression of the empirical unbiased covariance estimator is given by (see for example [8])

$$\forall\ 0 < p < N,\ \hat{r}^{(N)}(p) = \frac{1}{N-p}\sum_{k=-N}^{-p-1} X_k X_{k+p}.$$

Notice that, when p is close to N, the estimation is hampered since we only sum $N-p$ terms. Hence, we will not use the complete available data but rather use a cut-off.

Recall that $O_{K(N)} := [-K(N), -1]$ denotes the indices of the subset used for the prediction step. We define the empirical spectral density as

$$\hat{f}_K^{(N)}(t) = \sum_{p=-K(N)}^{K(N)} \hat{r}^{(N)}(p) e^{ipt}. \tag{2.1}$$

We now build an estimator for $p_{O_{K(N)}}$ (see Section 2.1 for the definition of $p_{O_{K(N)}}$). First, we divide the index space $\mathbb{Z}$ into $M_K \cup O_K \cup B_K \cup F_K$ where :

– $M_K = \{\cdots, -K-2, -K-1\}$ denotes the index of the past data that will not be used for the prediction (missing data)
– $O_K = -K, \cdots, -1$ the index of the data used for the prediction (observed data)
– $B_K = 0, \cdots, K-1$ the index of the data we currently want to forecast (blind data)

– $F_K = K, K+1, \cdots$ the remaining index (future data)
In the following, we omit the dependency on N to simplify the notations.
As discussed in Section 2.1, the projection operator p_{O_K} may be written by blocks as :

$$p_{O_K} = \begin{bmatrix} (\Gamma_{O_K})^{-1}\Gamma_{O_K\mathbb{Z}} \\ 0 \end{bmatrix}.$$

Since, we will apply this operator only to sequences with support in B_K, we may consider

$$\forall u \in l^2(\mathbb{Z}), \mathrm{Supp}(u) \subset B_K, p_{O_K B_K} u := \begin{bmatrix} (\Gamma_{O_K})^{-1}\Gamma_{O_K B_K} & 0 \\ 0 & 0 \end{bmatrix} u.$$

The last expression is given using the following block decomposition, if B_K^C denotes the complement of B_K in $\mathbb{Z}$:

$$\begin{bmatrix} O_K B_K & O_K B_K^C \\ O_K^C B_K & O_K^C B_K^C \end{bmatrix}.$$

Hence, the two quantities $\Gamma_{O_K B_K}$ and $(\Gamma_{O_K})^{-1}$ have to be estimated. On the one hand, a natural estimator of the first matrix is given by $\hat{\Gamma}_{O_K B_K}$ defined as

$$\left(\hat{\Gamma}^{(N)}_{O_K B_K}\right)_{ij} = \hat{r}^{(N)}(|j-i|), i \in O_K, j \in B_K.$$

On the other hand, a natural way to estimate $(\Gamma_{O_K})^{-1}$ could be to use $(\hat{\Gamma}^{(N)}_{O_K})$ (defined as $\left(\hat{\Gamma}^{(N)}_{O_K}\right)_{ij} = \hat{r}^{(N)}(|j-i|), i,j \in O_K$) and invert it. However, it is not sure that this matrix is invertible. So, we will consider an empirical regularized version by setting

$$\tilde{\Gamma}^{(N)}_{O_K} = \hat{\Gamma}^{(N)}_{O_K} + \hat{\alpha}\,\mathrm{Id}_{O_K},$$

for a well chosen $\hat{\alpha}$.
Set

$$\hat{\alpha} = -\min \hat{f}_K^{(N)} \mathbb{1}_{\min \hat{f}_K^{(N)} \leq 0} + \frac{m}{4} \mathbb{1}_{\min \hat{f}_K^{(N)} \leq \frac{m}{4}}.$$

so that $\left\|(\tilde{\Gamma}^{(N)}_{O_K})^{-1}\right\|_{2,op} \leq \frac{m}{4}$. Remark that $\tilde{\Gamma}^{(N)}_{O_K}$ is the Toeplitz matrix associated to the function $\tilde{f}_K^{(N)} = \hat{f}_K^{(N)} + \hat{\alpha}$, that has been tailored to ensure that $\tilde{f}^{(N)}$ is always greater than $\frac{m}{4}$, yielding the desired control to compute $\tilde{\Gamma}^{-1}_{O_K}$. Other regularization schemes could have been investigated. Nevertheless, note that adding a translation factor makes computation easier than using, for instance, a threshold on $\hat{f}_K^{(N)}$. Indeed, with our perturbation, we only modify the diagonal coefficients of the covariance matrix.

Finally, we will consider the following estimator, for any $Y \in \mathcal{B}_K := \text{Span}\left((X_i)_{i\in B_K}\right)$:

$$\hat{Y} := \hat{P}^{(N)}_{O_K B_K}(Y) = \Phi\left(\hat{p}^{(N)}_{O_K B_K}\Phi^{-1}(Y)\right),$$

where the estimator $\hat{p}^{(N)}_{O_K B_K}$ of $p_{\mathbb{Z}^- B_K}$, with window $K(N)$, is defined as follows

$$\hat{p}^{(N)}_{O_K B_K} = \left(\tilde{\Gamma}^{(N)}_{O_K}\right)^{-1}\hat{\Gamma}^{(N)}_{O_K B_K}. \tag{2.2}$$

2.3 Asymptotic behavior of the empirical projection operator

In this section, we give the rate of convergence of the estimator built previously (see Section 2.2). We will bound uniformly the bias of prediction error for random variables in the close future.
First, let us give some conditions on the sequence $(K(N))_{N\in\mathbb{N}}$) :

Assumption 2.3.1 *The sequence* $(K(N))_{N\in\mathbb{N}}$ *satisfies*
– $\lim K(N) \xrightarrow{N\to\infty} +\infty$.
– $\lim \frac{K(N)^2\log(K(N))}{N} \xrightarrow{N\to\infty} 0$.

Recall that the pointwise risk in $Y \in \mathbb{L}^2(\mathbb{P})$ is defined by

$$R(\hat{Y}) = \mathbb{E}\left[\left(\hat{Y} - \mathbb{E}\left[Y|(X_i)_{i<0}\right]\right)^2\right].$$

The global risk for the window $K(N)$ is defined by taking the supremum of the pointwise risk over all random variables $Y \in \mathcal{B}_K = \text{Span}\left((X_i)_{i\in B_K}\right)$

$$\mathcal{R}_{K(N)}\left(\hat{P}^N_{O_K B_K}\right) = \sup_{\substack{Y\in\mathcal{B}_K,\\ \text{Var}(Y)\le 1}} R(\hat{P}^N_{O_K}(Y)).$$

Notice that we could have chosen to evaluate the prediction quality only on X_0. Here, we choose another path, that is to predict all random variables from the close future. Our result is then quite stronger than if we had dealt only with prediction of X_0.
To get a control on the bias of the prediction, we need some regularity assumption. We consider Sobolev's type regularity by setting

$$\forall s > 1, W_s := \left\{ g \in \mathbb{L}_2([0,2\pi)), g(t) = \sum_{k\in\mathbb{Z}} a_k e^{ikt}, \sum_{k\in\mathbb{Z}} k^{2s}a_k^2 < \infty \right\}.$$

and define

$$\forall g \in W_s, g(t) = \sum_{k\in\mathbb{Z}} a_k e^{ikt}, \|g\|_{W_s} := \inf\left\{ M, \sum_{k\in\mathbb{Z}} k^{2s} a_k^2 \leq M \right\}.$$

Assumption 2.3.2 *There exists $s \geq 1$ such that $f^\star \in W_s$.*

We can now state our results. The following lemmas may be used in other frameworks than the blind problem. More precisely, if the blind prediction problem is very specific, the control of the loss between prediction with finite and infinite past is more classical, and the following lemmas may be applied for that kind of questions. The case where independent samples are available may also be tackled with the last estimators, using rates of convergences given in operator norms.
The bias is given by the following lemma

Lemma 2.3.3 *For N large enough, the following upper bound holds,*

$$\|p_{O_K B_k} - p_{\mathbb{Z}^- B_K}\|_\Gamma \leq C_2 \frac{1}{K(N)^{\frac{2s-1}{2}}},$$

where $C_2 = \left\|\frac{1}{f^\star}\right\|_{W_{2s}} m'(1+\frac{m'}{m})$.

In the last lemma, we assume regularity in terms of Sobolev's classes. Nevertheless, the proof may be written with some other kind of regularity. The proof is given in appendix, and is essentially based on Proposition 2.4.1. This last proposition provides the Schur block inversion of the projection operator.
The control for the variance is given in the following lemma :

Lemma 2.3.4

$$\begin{aligned}\int_0^\infty \mathbb{P}\left(\left\|\hat{p}^N_{O_K B_K} - p_{O_K B_K}\right\|_\Gamma^4 > t\right) \mathrm{d}t \;&\leq\; C_0^4 K(N)^4 (\frac{\log(K(N))}{N})^2 \\ &\quad + o(K(N)^4 (\frac{\log(K(N))}{N})^2),\end{aligned}$$

where $C_0 = 4m'(\frac{6m'}{m^2} + \frac{4}{m} + 2)$

Again, we choose this concentration formulation to deal with the dependency of the blind prediction problem, but this result gives immediately a control of the variance of the estimator whenever independent samples are observed (one for the estimation, and another one for the prediction).
The proof of this lemma is given in Section 2.5. It is based on a concentration inequality of the estimators $\hat{r}_p^{(N)}$ (see Comte [28]).
Integrating this rate of convergence over the blind data, we get our main theorem.

Theorem 2.3.5 *Under Assumptions 2.1.1, 2.3.1 and 2.3.2, for N large enough, the empirical estimator satisfies*

$$\sqrt{\mathcal{R}(\hat{P}^{(N)}_{O_K B_K})} \leq C_1 \frac{K(N)^2 \sqrt{\log(K(N))}}{\sqrt{N}} + C_2 \frac{1}{K(N)^{\frac{2s-1}{2}}},$$

where C_1 and C_2 are given in Appendix.

Again, the proof of this result is given in Section 2.5. It is quite technical. The main difficulty is induced by the blindness. Indeed, in this step, we have to deal with the dependency between the data and the empirical projector.
Obviously, the best rate of convergence is obtained by balancing the variance and the bias and finding the best window $K(N)$. Indeed, the variance increases with $K(N)$ while the bias decreases. Define $\hat{P}^{(N)}_\star$ as the projector $\hat{P}^{(N)}_{K^\star(N)}$ associated to the sequence $K^\star(N)$ that minimizes the bound in the last theorem. We get :

Corollary 2.3.6 (Rate of convergence of the prediction estimator) *Under Assumptions* 1.1 *and* 2.1*, for N large enough and choosing $K(N) = \left\lfloor (\frac{N}{\log N})^{\frac{1}{(2s+3)}} \right\rfloor$, we get*

$$\sqrt{\mathcal{R}(\hat{P}^{(N)}_\star)} \leq O\left(\left(\frac{\log N}{N} \right)^{\frac{2s-1}{2(2s+3)}} \right). \tag{2.3}$$

Notice that, in real life issues, it would be more natural to balance the risk given in Theorem 2.3.5, with the macroscopic term of variance given by

$$\mathbb{E}\Big[Y - \mathbb{E}\left[Y | (X_i)_{i<0} \right] \Big].$$

This leads to a larger $K(N)$. Nevertheless, Corollary 2.3.6 has a theoretical interest. Indeed, it recovers the classical semi-parametric rate of convergence, and provides a way to get away from dependency. Notice that, the estimation rate increases with the regularity s of the spectral density $f^\star$. More precisely, if $s \to \infty$, we obtain $(\frac{\log N}{N})^{\frac{1}{2}}$. This is, up to the log-term, the optimal speed. As a matter of fact, in this case, estimating the first coefficients of the covariance matrix is enough. Hence, the bias is very small. Proving a lower bound on the mean error (that could lead to a minimax result), is a difficult task, since the tools used to design the estimator are far from the usual estimation methods.

2.4 Projection onto finite observations with known covariance

We aim at providing an exact expression for the projection operator. For this, we generalize the expression given by Bondon ([16], [17]) for a projector onto infinite

past.

$$p_A = \begin{bmatrix} Id_A & \Gamma_A^{-1}\Gamma_{AA^C} \\ 0 & 0 \end{bmatrix}.$$

Denote also $\Lambda := \Gamma^{-1} = T(\frac{1}{f^\star})$ the inverse of the covariance operator, the following proposition provides an alternative expression of any projection operators.

Proposition 2.4.1 *One has*

$$p_A = \begin{bmatrix} \mathrm{Id}_A & -\Lambda_{AA^C}\Lambda_{A^C}^{-1} \\ 0 & 0 \end{bmatrix}$$

Furthermore, the prediction error verifies

$$\mathbb{E}\left[(P_A Y - Y)^2\right] = u^T \Lambda_{A^C}^{-1} u,$$

where $Y = \Phi(u) = u^T X$.

The proof of this proposition is given in Appendix. We point out that this proposition is helpful for the computation of the bias. Indeed, it gives a way to calculate the norm of the difference between two inverses operators.

2.5 Appendix

Proof of Proposition 2.4.1

Proof. For the proof of Proposition 2.4.1, let us choose

$$A \subset \mathbb{Z},$$

and denote the complement of A in $\mathbb{Z}$ by

$$M := A^C$$

First of all, $\Lambda = \Gamma^{-1}$ is a Toeplitz operator over H with eigenvalues in $[\frac{1}{m'}; \frac{1}{m}]$. Λ_M may be inverted as a principal minor of Λ. Let us define the Schur complement of Λ on sequences with support in M : $S = \Lambda_A - \Lambda_{AM}\Lambda_M^{-1}\Lambda_{MA}$. The next lemma provides an expression of Γ_A^{-1} (see for instance [69]).

Lemma 2.5.1

$$\begin{aligned} \Gamma_A^{-1} &= S \\ &= \Lambda_A - \Lambda_{AM}\Lambda_M^{-1}\Lambda_{MA}. \end{aligned}$$

Proof. of Lemma 2.5.1
One can check

$$\begin{aligned}
&\begin{bmatrix} \Lambda_A & \Lambda_{AM} \\ \Lambda_{MA} & \Lambda_M \end{bmatrix} \begin{bmatrix} S^{-1} & -S^{-1}\Lambda_{AM}\Lambda_M^{-1} \\ -\Lambda_M^{-1}\Lambda_{MA}S^{-1} & \Lambda_M^{-1} + \Lambda_M^{-1}\Lambda_{MA}S^{-1}\Lambda_{AM}\Lambda_M^{-1} \end{bmatrix} \\
= &\begin{bmatrix} \Lambda_A S^{-1} - \Lambda_{AM}\Lambda_M^{-1}\Lambda_{MA}S^{-1} & -\Lambda_A S^{-1}\Lambda_{AM}\Lambda_M^{-1} + \Lambda_{AM}(\Lambda_M^{-1} + \Lambda_M^{-1}\Lambda_{MA}S^{-1}\Lambda_{AM}\Lambda_M^{-1}) \\ \Lambda_{MA}S^{-1} - \Lambda_M\Lambda_M^{-1}\Lambda_{MA}S^{-1} & -\Lambda_{MA}S^{-1}\Lambda_{AM}\Lambda_M^{-1} + \Lambda_M(\Lambda_M^{-1} + \Lambda_M^{-1}\Lambda_{MA}S^{-1}\Lambda_{AM}\Lambda_M^{-1}) \end{bmatrix} \\
= &\begin{bmatrix} SS^{-1} & (\Lambda_{AM}\Lambda_M^{-1}\Lambda_{MA}S^{-1} + I_A - \Lambda_A S^{-1})\Lambda_{AM}\Lambda_M^{-1} \\ \Lambda_{MA}S^{-1} - \Lambda_{MA}S^{-1} & -\Lambda_{MA}S^{-1}\Lambda_{AM}\Lambda_M^{-1} + I_M + \Lambda_{MA}S^{-1}\Lambda_{AM}\Lambda_M^{-1} \end{bmatrix} \\
= &\begin{bmatrix} I_A & 0 \\ 0 & I_M \end{bmatrix}.
\end{aligned}$$

Since the matrix are symmetric, we can transpose the last equality. We obtain that

$$\begin{aligned}
\begin{bmatrix} S^{-1} & -S^{-1}\Lambda_{AM}\Lambda_M^{-1} \\ -\Lambda_M^{-1}\Lambda_{MA}S^{-1} & \Lambda_M^{-1} + \Lambda_M^{-1}\Lambda_{MA}S^{-1}\Lambda_{AM}\Lambda_M^{-1} \end{bmatrix} &= \Lambda^{-1} \\
&= \Gamma.
\end{aligned}$$

So that $\Gamma_A = S^{-1}$. ∎
We now compute the projection operator :

$$\begin{aligned}
p_A &= \begin{bmatrix} Id_A & \Gamma_A^{-1}\Gamma_{AM} \\ 0 & 0 \end{bmatrix} \\
&= \begin{bmatrix} Id_A & S\Gamma_{AM} \\ 0 & 0 \end{bmatrix} \\
&= \begin{bmatrix} Id_A & (\Lambda_A - \Lambda_{AM}\Lambda_M^{-1}\Lambda_{MA})\Gamma_{AM} \\ 0 & 0 \end{bmatrix} \\
&= \begin{bmatrix} Id_A & \Lambda_A\Gamma_{AM} - \Lambda_{AM}\Lambda_M^{-1}(Id_M - \Lambda_M\Gamma_M) \\ 0 & 0 \end{bmatrix} \\
&= \begin{bmatrix} Id_A & \Lambda_A\Gamma_{AM} - \Lambda_{AM}\Lambda_M^{-1} + \Lambda_{AM}\Gamma_M \\ 0 & 0 \end{bmatrix} \\
&= \begin{bmatrix} Id_A & -\Lambda_{AM}\Lambda_M^{-1} \\ 0 & 0 \end{bmatrix},
\end{aligned}$$

where we have used $\Lambda\Gamma = Id$ in the last two lines.
Now consider Q the quadratic error operator. It is defined as

$$\forall u \in l^2(\mathbb{Z}), u^T Q u := \left\| (p_A u - u)^2 \right\|_\Gamma = \mathbb{E}\left[(\Phi(u) - P_A\Phi(u))^2 \right].$$

This operator Q can be obtained by a direct computation (writing the product right above), but it is easier to use the expression of the variance of a projector in

the Gaussian case given for instance by [60].

$$Q = \Gamma_M - \Gamma_{MA}\Gamma_A^{-1}\Gamma_{AM}$$

Again, notice that Q is the Schur complement of Γ on sequences with support in A, and thanks to Lemma 2.5.1 applied to Λ instead of Γ, we get

$$Q = \Lambda_M^{-1}.$$

This ends the proof of Proposition 2.4.1. ∎

Proof of Theorem 2.3.5

Proof. of Theorem 2.3.5
Recall that we aim at providing a bound on $\sqrt{\mathcal{R}(\hat{P}^{(N)}_{O_K B_K})}$.
Notice first that we have

$$\sqrt{\mathcal{R}(\hat{P}^{(N)}_{O_K B_K})} \leq \sqrt{\sup_{\substack{Y\in\mathcal{B}_K,\\ \mathrm{Var}(Y)\leq 1}} \mathbb{E}\left[(\hat{P}^{(N)}_{O_K B_K}(Y) - P_{O_K B_K}(Y))^2\right]} + \sqrt{\mathcal{R}(P_{O_K B_K}))}.$$

Using Lemma 2.3.3 for a sequence $(K(N))_{N\in\mathbb{N}}$ and a centered random variable $Y \in \mathrm{Span}\left((X_i)_{i\in B_K}\right)$ such that $\mathbb{E}[Y^2] = 1$, we have

$$\begin{aligned}\sqrt{\mathcal{R}(P_{O_K B_K}))} &\leq \|p_{O_K B_K} - p_{\mathbb{Z}^- B_K}\|_\Gamma \sqrt{\mathbb{E}[Y^2]} \\ &\leq C_2 \frac{1}{K(N)^{\frac{2s-1}{2}}}.\end{aligned}$$

For the variance, we first notice that $Y = \Phi(u) = u^T\mathbf{X}$,

$$1 = \mathbb{E}\left[Y^2\right] = u^T\Gamma_{B_K}u \geq mu^Tu = m\sum_{i=0}^{K(N)-1} u_i^2,$$

Denote $A = \hat{p}^{(N)}_{O_K B_K} - p_{O_K B_K}$. We can write, by applying twice Cauchy-Schwarz's inequality,

$$\begin{aligned}\mathbb{E}\left[\left(\hat{P}^{(N)}_{O_K B_K}Y - P_{O_K B_K}Y\right)^2\right] &= \int_\omega \Big(\sum_{i=-K(N)}^{-1}\sum_{j=0}^{K(N)-1} A_{ij}(\omega)u_jX_i(\omega)\Big)^2 \mathrm{d}\mathbb{P}(\omega) \\ &\leq \int_\omega \sum_{i=-K(N)}^{-1}\left(\sum_{j=0}^{K(N)-1} A_{ij}(\omega)u_j\right)^2 \sum_{i=-K(N)}^{-1} X_i^2(\omega)\mathrm{d}\mathbb{P}(\omega) \\ &\leq \int_\omega \sum_{i=-K(N)}^{-1}\sum_{j=0}^{K(N)-1} A_{ij}^2(\omega)\sum_{j=0}^{K(N)-1} u_j^2 \sum_{i=-K(N)}^{-1} X_i^2(\omega)\mathrm{d}\mathbb{P}(\omega).\end{aligned}$$

So that,

$$\mathbb{E}\left[\left(\hat{P}^{(N)}_{O_K B_K} Y - P_{O_K B_K} Y\right)^2\right] \leq \int_\omega \sum_{i=-K(N)}^{-1} \sum_{j=0}^{K(N)-1} A_{ij}^2(\omega) \frac{1}{m} \sum_{i=-K(N)}^{-1} X_i^2 \mathrm{d}\mathbb{P}(\omega).$$

Using the following equivalence between two norms for finite matrices with size (n, m) (see for instance [59]),

$$\sqrt{\sum_{i=1}^{n}\sum_{j=1}^{m} A_{ij}^2} \leq \sqrt{n}\,\|A\|_{2,op}\,,$$

we obtain

$$\mathbb{E}\left[\left(\hat{P}^{(N)}_{O_K B_K} Y - P_{O_K B_K} Y\right)^2\right] \leq \frac{K(N)}{m} \int_\omega \|A(\omega)\|_{2,op}^2 \sum_{i=-K(N)}^{-1} X_i^2(\omega) \mathrm{d}\mathbb{P}(\omega).$$

Further,

$$\begin{aligned}
\mathbb{E}\left[\left(\hat{P}^{(N)}_{O_K B_K} Y - P_{O_K B_K} Y\right)^2\right] &\leq \frac{K(N)}{m} \int_\omega \|A(\omega)\|_{2,op}^2 \sum_{j=-K(N)}^{-1} X_j^2(\omega) \mathrm{d}\mathbb{P}(\omega) \\
&\leq \frac{K(N)}{m} \sqrt{\int_\omega \|A(\omega)\|_{2,op}^4 \,\mathrm{d}\mathbb{P}(\omega)} \sqrt{\int_\omega \left(\sum_{j=-K(N)}^{-1} X_j^2(\omega)\right)^2 \mathrm{d}\mathbb{P}(\omega)} \\
&\leq \frac{K(N)}{m} \sqrt{\int_{\mathbb{R}^+} \mathbb{P}\left(\|A\|_{2,op}^4 > t\right) \mathrm{d}t} \sqrt{K(N)^2 \int_\omega \left(X_j^4\right) \mathrm{d}\mathbb{P}(\omega)},
\end{aligned}$$

We have used here again Cauchy-Schwarz's inequality and the fact that, for all nonnegative random variable Y,

$$\mathbb{E}[Y] = \int_{\mathbb{R}^+} \mathbb{P}(Y > t)\,\mathrm{d}t.$$

Since X_0 is Gaussian, its moment of order four r_4 is finite. Then Lemma 2.3.4 yields that, for N large enough,

$$\mathbb{E}\left[\left(\hat{P}^{(N)}_{O_K B_K} Y - P_{O_K B_K} Y\right)^2\right] \leq \frac{C_0^2 \sqrt{r_4} K(N)^4 \log(K(N)))}{mN}.$$

So that,

$$\sqrt{\sup_{\substack{Y \in \mathcal{B}_K, \\ \mathrm{Var}(Y) \leq 1}} \mathbb{E}\left[(\hat{P}^{(N)}_{O_K}(Y) - P_{O_K}(Y))^2\right]} \leq \frac{C_1 K(N)^2 \sqrt{\log(K(N))}}{\sqrt{N}},$$

with $C_1 = \frac{C_0 \sqrt[4]{r_4}}{\sqrt{m}}$. This ends the proof of the theorem. ∎

Proofs of concentration and regularity lemmas

First, we compute the bias and prove Lemma 2.3.3 :

Proof. of Lemma 2.3.3

Recall that we aim to obtain a bound on $\|p_{O_K B_K} - p_{\mathbb{Z}^- B_K}\|_\Gamma$. Using Proposition 2.4.1, we can write (looking at the following operators as operators on $l^2(\mathbb{Z})$)

$$\begin{aligned}\|p_{O_K B_K} - p_{\mathbb{Z}^- B_K}\|_\Gamma &\leq \|p_{O_K \mathbb{Z}^+} - p_{\mathbb{Z}^- \mathbb{Z}^+}\|_\Gamma \\ &\leq \left\| \begin{bmatrix} (\Gamma_{O_K})^{-1}\Gamma_{O_K\mathbb{Z}^+} \\ 0 \end{bmatrix} - \begin{bmatrix} -\Lambda_{O_K\mathbb{Z}^+}(\Lambda_{\mathbb{Z}^+})^{-1} \\ -\Lambda_{M_K\mathbb{Z}^+}(\Lambda_{\mathbb{Z}^+})^{-1} \end{bmatrix} \right\|_\Gamma .\end{aligned}$$

So that, using the norms equivalence,

$$\begin{aligned}\|p_{O_K B_K} - p_{\mathbb{Z}^- B_K}\|_\Gamma &\leq \frac{m'}{m} \left\| \begin{bmatrix} (\Gamma_{O_K})^{-1}\Gamma_{O_K\mathbb{Z}^+} \\ 0 \end{bmatrix} - \begin{bmatrix} -\Lambda_{O_K\mathbb{Z}^+}(\Lambda_{\mathbb{Z}^+})^{-1} \\ -\Lambda_{M_K\mathbb{Z}^+}(\Lambda_{\mathbb{Z}^+})^{-1} \end{bmatrix} \right\|_{2,op} \\ &\leq \frac{m'}{m} \left\| \begin{bmatrix} (\Gamma_{O_K})^{-1}\Gamma_{O_K\mathbb{Z}^+} + \Lambda_{O_K\mathbb{Z}^+}(\Lambda_{\mathbb{Z}^+})^{-1} \\ \Lambda_{M_K\mathbb{Z}^+}(\Lambda_{\mathbb{Z}^+})^{-1} \end{bmatrix} \right\|_{2,op} \\ &\leq \frac{m'}{m} \left\| \begin{bmatrix} (\Gamma_{O_K})^{-1}\Gamma_{O_K\mathbb{Z}^+}\Lambda_{\mathbb{Z}^+} + \Lambda_{O_K\mathbb{Z}^+} \\ \Lambda_{M_K\mathbb{Z}^+} \end{bmatrix} \right\|_{2,op} \left\| (\Lambda_{\mathbb{Z}^+})^{-1} \right\|_{2,op} \\ &\leq \frac{m'}{m} \left\| (\Lambda_{\mathbb{Z}^+})^{-1} \right\|_{2,op} \\ &\quad \times \left(\left\| (\Gamma_{O_K})^{-1}\Gamma_{O_K\mathbb{Z}^+}\Lambda_{\mathbb{Z}^+} + \Lambda_{O_K\mathbb{Z}^+} \right\|_{2,op} + \|\Lambda_{M_K\mathbb{Z}^+}\|_{2,op} \right) .\end{aligned}$$

The last step follows from the inequality :

$$\left\| \begin{matrix} A \\ B \end{matrix} \right\|_{2,op} \leq \left\| \begin{matrix} A \\ 0 \end{matrix} \right\|_{2,op} + \left\| \begin{matrix} 0 \\ B \end{matrix} \right\|_{2,op} = \|A\|_{2,op} + \|B\|_{2,op} .$$

But, since $\Lambda = \Gamma^{-1}$,

$$\Gamma_{O_K\mathbb{Z}^+}\Lambda_{\mathbb{Z}^+} + \Gamma_{O_K}\Lambda_{O_K\mathbb{Z}^+} = -\Gamma_{O_K M_K}\Lambda_{M_K\mathbb{Z}^+}.$$

So, we obtain,

$$\begin{aligned}\|p_{O_K B_K} - p_{\mathbb{Z}^- B_K}\|_\Gamma &\leq \frac{m'}{m} \left\| (\Lambda_{\mathbb{Z}^+})^{-1} \right\|_{2,op} \left(\left\| (\Gamma_{O_K})^{-1} \left(-\Gamma_{O_K M_K}\Lambda_{M_K\mathbb{Z}^+} \right) \right\|_{2,op} + \|\Lambda_{M_K\mathbb{Z}^+}\|_{2,op} \right) \\ &\leq \frac{m'}{m} \left\| (\Lambda_{\mathbb{Z}^+})^{-1} \right\|_{2,op} \left(\left\| (\Gamma_{O_K})^{-1} \right\|_{2,op} \| -\Gamma_{O_K M_K}\Lambda_{M_K\mathbb{Z}^+} \|_{2,op} + \|\Lambda_{M_K\mathbb{Z}^+}\|_{2,op} \right) \\ &\leq \frac{m'}{m} \left\| (\Lambda_{\mathbb{Z}^+})^{-1} \right\|_{2,op} \left(\left\| (\Gamma_{O_K})^{-1} \right\|_{2,op} \|\Gamma_{O_K M_K}\|_{2,op} \|\Lambda_{M_K\mathbb{Z}^+}\|_{2,op} + \|\Lambda_{M_K\mathbb{Z}^+}\|_{2,op} \right) \\ &\leq \frac{m'}{m} \left\| (\Lambda_{\mathbb{Z}^+})^{-1} \right\|_{2,op} \left(\left\| (\Gamma_{O_K})^{-1} \right\|_{2,op} \|\Gamma_{O_K M_K}\|_{2,op} + 1 \right) \|\Lambda_{M_K\mathbb{Z}^+}\|_{2,op} .\end{aligned}$$

But, we have,

$$\left\|(\Lambda_{\mathbb{Z}^+})^{-1}\right\|_{2,op} \leq m',$$

as the inverse of a principal minor of Λ.

$$\left\|(\Gamma_{O_K})^{-1}\right\|_{2,op} \leq \frac{1}{m},$$

since it is the inverse of a principal minor of Γ.

$$\|\Gamma_{O_K M_K}\|_{2,op} \leq m',$$

as an extracted operator of Γ.
Thus, we get

$$\|p_{O_K B_K} - p_{\mathbb{Z}^- B_K}\|_\Gamma \leq C_4 \|\Lambda_{M_K \mathbb{Z}^+}\|_{2,op},$$

where $C_4 = \frac{m'^2}{m}(1 + \frac{m'}{m})$. Since $f^\star \in W_s$ (Assumption 2.1), and $f^\star \geq m > 0$, we have also $\frac{1}{f^\star} \in W_s$ (see for instance [58]). If we denote $p(k) = \Lambda_{i,i+k}$ the Fourier coefficient of $\frac{1}{f^\star}$, we get

$$\begin{aligned}
\|\Lambda_{M_K \mathbb{Z}^+}\|_{2,op} &\leq \|\Lambda_{M_K \mathbb{Z}^+}\|_2 \\
&\leq \sqrt{\sum_{i \leq -K(N); 0 \leq j} p(j-i)^2} \\
&\leq \sqrt{\sum_{i=K(N)}^{\infty} \sum_{j=i}^{\infty} p(j)^2} \\
&\leq \sqrt{\sum_{i=K(N)}^{\infty} \frac{\left\|\frac{1}{f^\star}\right\|_{W_s}}{i^{2s}}} \\
&\leq \sqrt{\left\|\frac{1}{f^\star}\right\|_{W_s} \frac{1}{K(N)^{2s-1}}}.
\end{aligned}$$

■

So that the lemma is proved and the bias is given by

$$\|p_{O_K B_K} - p_{\mathbb{Z}^- B_K}\|_\Gamma \leq C_4 \sqrt{\left\|\frac{1}{f^\star}\right\|_{W_s}} \frac{1}{K(N)^{\frac{2s-1}{2}}}.$$

Actually, the rate of convergence for the bias is given by the regularity of the spectral density, since it depends on the coefficients far away from the principal diagonal.

Now, we prove Lemma 2.3.4, which achieves the proof of the theorem.
Proof. of Lemma 2.3.4

Recall that $A = \hat{p}^{(N)}_{O_K B_K} - p_{O_K B_K}$. We aim at proving that

$$\int_0^\infty \mathbb{P}\left(\|A\|_\Gamma^4 > t\right) \mathrm{d}t \leq C_0^4 K(N)^4 (\frac{\log(K(N))}{N})^2 + o(K(N)^4 (\frac{\log(K(N))}{N})^2).$$

To simplify the notations, the dependency on N is omitted.
First,

$$\begin{aligned}
\|A\|_{2,op} &= \left\|(\tilde{\Gamma}_{O_K})^{-1}\hat{\Gamma}_{O_K B_K} - (\Gamma_{O_K})^{-1}\Gamma_{O_K B_K}\right\|_{2,op} \\
&\leq \|\Gamma_{O_K B_K}\|_{2,op} \left\|(\tilde{\Gamma}_{O_K})^{-1} - (\Gamma_{O_K})^{-1}\right\|_{2,op} \\
&\quad + \left\|(\tilde{\Gamma}_{O_K})^{-1}\right\|_{2,op} \left\|\hat{\Gamma}_{O_K B_K} - \Gamma_{O_K B_K}\right\|_{2,op} \\
&\leq \|\Gamma_{O_K B_K}\|_{2,op} \left\|(\tilde{\Gamma}_{O_K})^{-1}\right\|_{2,op} \left\|(\Gamma_{O_K})^{-1}\right\|_{2,op} \left\|\tilde{\Gamma}_{O_K} - \Gamma_{O_K}\right\|_{2,op} \\
&\quad + \left\|(\tilde{\Gamma}_{O_K})^{-1}\right\|_{2,op} \left\|\hat{\Gamma}_{O_K B_K} - \Gamma_{O_K B_K}\right\|_{2,op}.
\end{aligned}$$

But, we have,

$$\|\Gamma_{O_K B_K}\|_{2,op} \leq m',$$

as an extracted operator of Γ.

$$\left\|(\Gamma_{O_K})^{-1}\right\|_{2,op} \leq \frac{1}{m},$$

as the inverse of a principal minor of Γ.

$$\left\|(\tilde{\Gamma}_{O_K})^{-1}\right\|_{2,op} \leq \frac{4}{m},$$

thanks to the regularization.
Now, note that, for any matrix $B \in M_{n,m}(\mathbb{R})$,

$$\|B\|_{2,op} \leq \sqrt{\sup_{i=1,\cdots,n} \sum_{j=1}^{m} |B_{ij}|} \sqrt{\sup_{j=1,\cdots,m} \sum_{i=1}^{n} |B_{ij}|}.$$

So,

$$\left\|\tilde{\Gamma}_{O_K} - \Gamma_{O_K}\right\|_{2,op} \leq K(N) \sup_{p \leq 2K(N)} \left\{\left|\hat{r}^{(N)}(p) - r(p)\right|\right\} + |\hat{\alpha}|.$$

And we also get,

$$\left\|\hat{\Gamma}_{O_K B_K} - \Gamma_{O_K B_K}\right\|_{2,op} \leq K(N) \sup_{p \leq 2K(N)} \left\{\left|\hat{r}^{(N)}(p) - r(p)\right|\right\}.$$

Recall the definition of $\hat{\alpha}$,

$$\hat{\alpha} = -\min \hat{f}_K^N \mathbb{1}_{\min \hat{f}_K^N \leq 0} + \frac{m}{4} \mathbb{1}_{\min \hat{f}_K^N \leq \frac{m}{4}}.$$

We will now use the following lemma, proved in the next section.

Lemma 2.5.2 *The empirical spectral density is such that, for N large enough*

$$\left\| \hat{f}_{K(N)}^N - f^\star \right\|_\infty \leq (2K(N)+1) \sup_{p \leq 2K(N)} \left\{ \left| \hat{r}^{(N)}(p) - r(p) \right| \right\} + \frac{m}{4}.$$

Thanks to Lemma 2.5.2, we have

$$|\hat{\alpha}| \leq (2K(N)+1) \sup_{p \leq 2K(N)} \left\{ \left| \hat{r}^{(N)}(p) - r(p) \right| \right\} + \frac{m}{4} \mathbb{1}_{\min \hat{f}_K^N \leq \frac{m}{4}}.$$

So, we obtain,

$$\begin{aligned}
\|A\|_{2,op} &\leq \frac{4m'}{m^2} \left(K(N) \sup_{p \leq 2K(N)} \left\{ \left| \hat{r}^{(N)}(p) - r(p) \right| \right\} + |\hat{\alpha}| \right) \\
&\quad + \frac{4}{m} K(N) \left(\sup_{p \leq 2K(N)} \left\{ \left| \hat{r}^{(N)}(p) - r(p) \right| \right\} \right) \\
&\leq \left(\frac{12m'}{m^2} + \frac{4}{m} + \frac{4m'}{m^2 K(N)} \right) K(N) \left(\sup_{p \leq 2K(N)} \left\{ \left| \hat{r}^{(N)}(p) - r(p) \right| \right\} \right) \\
&\quad + \frac{m'}{m} \mathbb{1}_{\min \hat{f}_K^N \leq \frac{m}{4}}.
\end{aligned}$$

We will use here some other technical lemmas. Their proofs are also postponed to the last section. The first one gives an uniform concentration result on the estimator $\hat{r}^{(N)}(p)$:

Lemma 2.5.3 *Assume that Assumption 2.3.1 holds. Then, there exists N_0 such that, for all $N \geq N_0$, and $x \geq 0$,*

$$\forall p \leq 2K(N), \left| \hat{r}^{(N)}(p) - r(p) \right| \leq 4m' \left(\sqrt{\frac{(\log(K(N)) + x)}{N}} + \frac{x}{N} \right),$$

with probability at least $1 - e^{-x}$

For ease of notations, we set $C_0 = 4m' \left(\frac{12m'}{m^2} + \frac{4}{m} \right)$ and $C_3 = \frac{m'}{m}$. For the computation of the mean, the interval $[0, +\infty[$ will be divided into three parts, where only the first contribution is significant, thanks to the exponential concentration. We will prove that the two other parts are negligible.

We obtain, for all $x \geq 0$

$$\|A\|_{2,op} \leq (C_0 + o(1))K(N)\left(\sqrt{\frac{\log(K(N)) + x}{N}} + \frac{x}{N}\right) + C_3 \mathbb{1}_{\min \hat{f}_K^N \leq \frac{m}{4}},$$

with probability at least $1 - e^{-x}$
Set $t_1 = \left(C_0 K(N)\sqrt{\frac{\log(K(N))}{N}}\right)^4$. For $t \in [0, t_1]$, we use the inequality

$$\mathbb{P}\left(\|A\|_{2,op}^4 > t\right) \leq 1.$$

We obtain the first contribution to the integral. This is also the non negligible part.

$$\int_0^{t_1} \mathbb{P}\left(\|A\|_{2,op}^4 > t\right) \mathrm{d}t \leq \left(C_0 K(N)\sqrt{\frac{\log(K(N))}{N}}\right)^4.$$

Now, set $t_2 = \left(C_0 K(N)\sqrt{\frac{\log(K(N))+N}{N}} + C_3\right)^4$. For $t \in [t_1, t_2]$, we use

$$\begin{aligned}\mathbb{P}\Bigg(\|A\|_{2,op}^4 > \sup\left(C_0^4 K(N)^4 \left(\frac{\log(K(N)) + x}{N}\right)^2, C_0^4 K(N)^4 \left(\frac{x}{N}\right)^4\right)\Bigg) \\ \leq e^{-x} + \mathbb{P}\left(\min \hat{f}_K^N \leq \frac{m}{4}\right).\end{aligned}$$

Notice that the last lemma provides

$$\mathbb{P}\Big(2K(N) \sup_{p \leq 2K(N)} \left\{\left|\hat{r}^{(N)}(p) - r(p)\right|\right\} > \frac{m}{2}\Big) \leq e^{-\frac{Nm^2}{(64K(N)m')^2}}.$$

Indeed, set $x_0(N) = \frac{Nm^2}{(64K(N)m')^2}$.
One can compute that with probability at least $1 - e^{-x_0(N)}$,

$$\begin{aligned}\sup_{p \leq 2K(N)} \left\{\left|\hat{r}^{(N)}(p) - r(p)\right|\right\} &\leq 4m'\left(\sqrt{\frac{\log(K(N)) + x_0(N)}{N}} + \frac{x_0(N)}{N}\right) \\ &\leq 4m'\left(\sqrt{\frac{\log(K(N))}{N} + \frac{m^2}{(64K(N)m')^2}} + \frac{m^2}{(64K(N)m')^2}\right) \\ &\leq 4m'\left(\sqrt{\frac{\log(K(N))}{N}} + \sqrt{\frac{m^2}{(64K(N)m')^2}} + \frac{m^2}{(64K(N)m')^2}\right) \\ &\leq 4m'\left(\sqrt{\frac{\log(K(N))}{N}} + \frac{m}{(64K(N)m')} + \frac{m^2}{(64K(N)m')^2}\right) \\ &\leq \frac{m}{8K(N)},\end{aligned}$$

for N large enough. Hence,

$$\mathbb{P}\left(\min \hat{f}_K^N \leq \frac{m}{4}\right) \leq e^{-\frac{Nm^2}{(64K(N)m')^2}}.$$

So, we have

$$\mathbb{P}\left(\|A\|_{2,op}^4 > \max\left(C_0^4 K(N)^4 \left(\frac{\log(K(N))+x}{N}\right)^2, C_0^4 K(N)^4 \left(\frac{x}{N}\right)^4\right)\right) \leq e^{-x}+e^{-\frac{Nm^2}{(64K(N)m')^2}}.$$

Finally, the following lemma (the proof is again postponed in Appendix) will be useful to transform a probability inequality into a $\mathbb{L}^2$ inequality.

Lemma 2.5.4 *Let X be a nonnegative random variable such that there exists two one to one maps f_1 and f_2 and a $C > 0$ with*

$$\forall x \geq 0, \mathbb{P}\left(X > \sup(f_1(x), f_2(x))\right) \leq e^{-x} + C.$$

Then, for all $t \in \mathrm{Im}(f_1) \cap \mathrm{Im}(f_2)$,

$$\mathbb{P}\left(X > t\right) \leq e^{-f_1^{-1}(t)} + e^{-f_2^{-1}(t)} + C.$$

So, thanks to lemma 2.5.4, we have

$$\mathbb{P}\left(\|A\|_{2,op}^4 > t\right) \leq e^{-N\sqrt{\frac{t}{C_0^4 K(N)^4}}-\log(K(N))} + e^{-N\sqrt[4]{\frac{t}{C_0^4 K(N)^4}}} + e^{-\frac{Nm^2}{(64K(N)m')^2}}.$$

Now, we will prove that each term can be neglected. Integrating by part, we obtain

$$\begin{aligned}
\int_{t_1}^{t_2} e^{-\sqrt{\frac{t}{C_0^4 K(N)^4}+\log(K(N))}}\mathrm{d}t &\leq \int_{t_1}^{\infty} e^{-N\sqrt{\frac{t}{C_0^4 K(N)^4}}-\log(K(N))}\mathrm{d}t \\
&\leq \left[\frac{-2\sqrt{t}C_0^2 K(N)^2}{N} e^{-N\sqrt{\frac{t}{C_0^4 K(N)^4}}-\log(K(N))}\right]_{t_1}^{\infty} \\
&\quad + \int_{t_1}^{\infty} \frac{C_0^2 K(N)^2}{N\sqrt{t}} e^{-N\sqrt{\frac{t}{C_0^4 K(N)^4}}-\log(K(N))}\mathrm{d}t \\
&\leq \frac{2\log(K(N))C_0^4 K(N)^4}{N^2} + \frac{2C_0^4 K(N)^4}{N^2} \\
&= o\left(\left(C_0 K(N)\sqrt{\frac{\log(K(N))}{N}}\right)^4\right).
\end{aligned}$$

Then,

$$\begin{aligned}\int_{t_1}^{t_2} e^{-\sqrt[4]{\frac{Nt}{C_0^4K(N)^4}}}\mathrm{d}t &\leq t_2 e^{-N\sqrt[4]{\frac{t_1}{C_0^4K(N)^4}}} \\ &= o\left(\left(C_0K(N)\sqrt{\frac{\log(K(N))}{N}}\right)^4\right).\end{aligned}$$

So that,

$$\begin{aligned}\int_{t_1}^{t_2} e^{-x_0(N)}\mathrm{d}t &\leq t_2 e^{-\frac{Nm^2}{(64K(N)m')^2}} \\ &= o\left(\left(C_0K(N)\sqrt{\frac{\log(K(N))}{N}}\right)^4\right).\end{aligned}$$

Leading to

$$\int_{t_1}^{t_2}\mathbb{P}\left(\|A\|_{2,op}^4 > t\right)\mathrm{d}t = o\left(\left(C_0K(N)\sqrt{\frac{\log(K(N))}{N}}\right)^4\right),$$

Finally, for $t \in [t_2, +\infty[$, we use

$$\mathbb{P}\left(\|A\|_{2,op}^4 > \max\left(\left(2C_0K(N)\sqrt{\frac{\log(K(N))+x}{N}}+2C_3\right)^4, \left(2C_0K(N)\frac{x}{N}+2C_3\right)^4\right)\right) \leq e^{-x}.$$

Thanks to lemma 2.5.4, we get

$$\mathbb{P}\left(\|A\|_{2,op}^4 > t\right) \leq e^{-N\left(\frac{\sqrt[4]{t}-C_3}{C_0K(N)}\right)^2+\log(K(N))} + e^{-N\frac{(\sqrt[4]{t}-C_3)}{C_0K(N)}}.$$

So, integrating by part once more, we obtain

$$\begin{aligned}\int_{t_2}^{+\infty} e^{-N\left(\frac{\sqrt[4]{t}-C_3}{C_0K(N)}\right)^2+\log(K(N))}\mathrm{d}t &\leq \int_{\sqrt[4]{t_2}-C_3}^{+\infty} 4(u+C_3)^3 e^{-N\left(\frac{u}{C_0K(N)}\right)^2+\log(K(N))}\mathrm{d}u \\ &\leq \left[P_1(u,N,K(N))e^{-N\left(\frac{u}{C_0K(N)}\right)^2+\log(K(N))}\right]_{\sqrt[4]{t_2}-C_3}^{+\infty} \\ &\leq P_1(u,N,K(N))e^{-N} \\ &\leq o\left(\left(C_0K(N)\sqrt{\frac{\log(K(N))}{N}}\right)^4\right).\end{aligned}$$

Here, $P_1(u, N, K(N))$ is a polynomial of degree 3 in u and is rational function in N and $K(n)$.
Furthermore,

$$\begin{aligned}\int_{t_2}^{+\infty} e^{-N\frac{(\sqrt[4]{t}-C_3)}{C_0K(N)}}\mathrm{d}t &\leq \int_{\sqrt[4]{t_2}-C_3}^{+\infty} 4(u+C_3)^3 e^{-N\frac{u}{C_0K(N)}}\mathrm{d}u \\ &\leq \left[P_2(u,N,K(N))e^{-N\frac{u}{C_0K(N)}}\right]_{\sqrt[4]{t_2}-C_3}^{+\infty} \\ &\leq P_2(u,N,K(N))e^{-\sqrt{N(\log(K(N))+N)}} \\ &\leq o\left(\left(C_0K(N)\sqrt{\frac{\log(K(N))}{N}}\right)^4\right),\end{aligned}$$

where $P_2(u, N, K(N))$ is a polynomial of degree 3 in u and is rational function in N and $K(n)$.
We proved here

$$\int_0^{\infty}\mathbb{P}\left(\|A\|_{2,op}^4 > t\right) \leq C_0^4K(N)^4(\frac{\log(K(N))}{N})^2 + o(K(N)^4(\frac{\log(K(N))}{N})^2).$$

This ends the proof. ∎

Technical lemmas

We prove now the technical lemmas :
Proof. of Lemma 2.5.3
Recall that, for any $f \in \mathbb{L}^2([0, 2\pi))$, the matrix $\mathcal{T}_N(f)$ is the Toeplitz matrix associated to the density f. Define $g_p(t) = \frac{N}{N-p}\cos(pt)$. We have
Hence, we get $\hat{r}^{(N)}(p) = \frac{1}{N}X^T\mathcal{T}_N(g_p)X$.
We use the following proposition from Laurent and Massart [49]. Let $X_1, \cdots, X_N$ be a centered Gaussian stationary sequence with spectral density f and g a bounded function such that $\mathcal{T}_N(g)$ is a symmetric non negative matrix. Then the following concentration inequality holds for $Z_N(g) = \frac{1}{N}\left(X^T\mathcal{T}_N(g)X - \mathbb{E}[X^T\mathcal{T}_N(g)X]\right)$:

$$\mathbb{P}\left(|Z_N(g)| \geq 2\|f\|_\infty\left(\|g\|_2\sqrt{x} + \|g\|_\infty x\right)\right) \leq 2e^{-Nx}.$$

We obtain

$$\mathbb{P}\left(\left|\hat{r}^{(N)}(p) - r(p)\right| > 2m'\frac{N}{N-p}(\sqrt{x}+x)\right) \leq 2e^{-Nx}.$$

or, equivalently,

$$\left|\hat{r}^{(N)}(p) - r(p)\right| > 2m'\frac{N}{N-p}(\sqrt{\frac{x+\log(K(N))+2\log(2)}{N}} + \frac{x+\log(K(N))+2\log(2)}{N}),$$

with probability lower than $\frac{e^{-x}}{2K(N)}$. By taking an equivalent, we obtain that there exists N_0 such that, for all $N \geq N_0$, for all $p \leq 2K(N)$

$$\mathbb{P}\left(\left|\hat{r}^{(N)}(p) - r(p)\right| > 4m'\sqrt{\frac{x + \log(K(N))}{N}} + \frac{x}{N}\right) \leq \frac{e^{-x}}{2K(N)}.$$

■

Proof. of Lemma 2.5.4
We set $t = \sup(f_1(x), f_2(x))$ If $t = f_1(x)$ then

$$\mathbb{P}\left(X > t\right) \leq e^{-f_1^{-1}(t)} + C \leq e^{-f_1^{-1}(t)} + e^{-f_2^{-1}(t)} + C.$$

Symmetrically, if $t = f_2(x)$ we have

$$\mathbb{P}\left(X > t\right) \leq e^{-f_1^{-1}(t)} + e^{-f_2^{-1}(t)} + C.$$

■

Proof. of Lemma 2.5.2 It is sufficient to ensure that the bias is small enough. Choose N_0 such that

$$2\left\|f^\star\right\|_{W_s} K(N)^{-s+1} \leq \frac{m}{4}.$$

Then we use

$$\begin{aligned}
\left\|\hat{f}^N_{K(N)} - f^\star\right\|_\infty &\leq \sum_{p=-K(N)}^{K(N)} \left|\hat{r}^{(N)}(p) - r(p)\right| + 2\sum_{p>K(N)} |r(p)| \\
&\leq (2K(N)+1) \sup_{p\leq 2K(N)} \left\{\hat{r}^{(N)}(p) - r(p)\right\} + 2\left\|f^\star\right\|_{W_s} K(N)^{-s+1} \\
&\leq (2K(N)+1) \sup_{p\leq 2K(N)} \left\{\hat{r}^{(N)}(p) - r(p)\right\} + \frac{m}{4}.
\end{aligned}$$

This ends the proof of the last lemma. ■

Acknowledgement :
The authors wish to thank the anonymous referee for a careful reading of the manuscript and for providing very useful comments.

Chapitre 3

Parametric estimation for Gaussian fields indexed by graphs

Dans ce chapitre, nous utilisons la théorie spectrale des opérateurs de Hilbert, afin d'étudier des processus $ARMA$ gaussiens indexés par des graphes. Nous étendons l'approximation de Whittle pour l'estimation des paramètres de la densité spectrale associée et nous prouvons l'optimalité asymptotique des estimateurs ainsi construits.

Introduction

In the past few years, much interest has been paid to the study of random fields over graphs. It has been driven by the growing needs for both theoretical and practical results for data indexed by graphs. On the one hand, the definition of graphical models by J.N. Darroch, S.L. Lauritzen and T.P. Speed in 1980 [32] fostered new interest in Markov fields, and many tools have been developed in this direction (see, for instance [64] and [63]). On the another hand, the industrial demand linked to graphical problems has risen 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 also benefit 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 with the forecast of speed of vehicles which may be seen as a spatial random field over the traffic network. Some work has been done without taking into account the particular graph structure of the speed process (see for example [36] and [52] for related statistical issues). In this paper, we build a new model for Gaussian random fields over graphs and study statistical properties of such

stochastic processes.

A random field over a graph is a spatial process indexed by the vertices of a graph, namely $(X_i)_{i\in G}$, where G is a given graph. Many models already exist in the probabilistic literature, ranging from Markov fields to autoregressive processes, which are based on two general kinds of construction. On the one hand, graphical models are defined as Markov fields (see for instance [43]), with a particular dependency structure. Actually, they are built by specifying a dependency structure for X_i and X_j, conditionally to the other variables, as soon as the locations $i \in G$ and $j \in G$ are connected. For graphical models, we refer for instance to [32] and references therein. On the other hand, the graph itself, through the adjacency operator, can provide the dependency. This is the case, for example, of autoregressive models on $\mathbb{Z}^d$ (see [43]). Here, the local form of the graph is strongly used for statistical inference.

More precisely, the usual purpose of graphical models is to design an underlying graph which reflects the dependency of the data. This method has to be applied when this graph is not easily known (for instance social networks) or when it plays the role of a model which helps understanding the correlations between high complex data (for instance for biological purpose). Our approach differs since, in our case, the graph is known, and we aim at using a model with *stationary* properties. Indeed, in the case of road traffic, we can consider that the correlations of the process depend mainly on the local structure of the network. This assumption is commonly accepted among professionals of road trafficking speaking of capacity of the road.

In this paper, we extend some classical results from time series to spatial fields over general graphs and provide a new definition for regular $ARMA$ processes on graphs. For this, we will make use of spectral analysis and extend to our framework some classical results of time series. In particular, the notion of spectral density may be extended to graphs. This will enable us to construct a maximum likelihood estimate for parametric models of spectral densities. This also leads to an extension of the Whittle's approximation (see [41], [8]). Actually, many extensions of this approximation have been performed, even in non-stationary cases (see [31], [57], [38]). The extension studied here concerns general $ARMA$ processes over graphs. We point out that we will compare throughout all the paper our new framework with the case $G = \mathbb{Z}^d, d \geq 1$.

Section 3.1 is devoted to some definitions of graphs and spectral theory for time series. Then we state the definition of general $ARMA$ processes over a graph in Section 3.2. The convergence of the Whittle maximum likelihood estimate and its asymptotic efficiency are given in Theorems 3.3.7 and 3.3.10 in Section 3.3. Sec-

tion 3.4 is devoted to a short discussion on potential applications and perspectives. Some simulations are provided in Section 3.5. The last section provides all necessary tools to prove the main theorems, in particular Szegö's Lemmas for graphs are given in Section 3.6, while the proofs of the technical Lemmas are postponed in Section 3.6.

3.1 Definitions and useful properties for spectral analysis and Toeplitz operators

Graphs, adjacency operator, and spectral representation

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 vertices. $\mathbf{G}$ 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, $|W_{ij}| \neq 0$ when $i \in G$ and $j \in G$ are connected.
We assume that W is 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 of the vertex i, and the degree of the graph $\mathbf{G}$ is defined as the maximum degree of the vertices of the graph $\mathbf{G}$:

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

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

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

Assume now that W is renormalized : its entries belong to $[-\frac{1}{\deg(\mathbf{G})}, \frac{1}{\deg(\mathbf{G})}]$. This is not restrictive since re-normalizing the adjacency operator does not change the objects introduced later. In particular, the spectral representation of Hilbertian operator is not sensitive to a renormalization.
Notice that in the classical case $G = \mathbb{Z}$, the renormalized adjacency operator is

$$W_{ij}^{(\mathbb{Z})} = \frac{1}{2}\mathbb{1}_{\{|i-j|=1\}}, (i,j \in \mathbb{Z}). \tag{3.1}$$

Here, $\deg(\mathbb{Z}) = 2$. This case will be used in all the paper as an illustration example.

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

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

We denote by B_G the set of all bounded Hilbertian operators on $l^2(G)$ (the set of square sommable real sequences indexed by G). The operator space B_G will be endowed with the classical operator norm

$$\forall A \in B_G, \|A\|_{2,op} := \sup_{u \in l^2(G), \|u\|_2 \leq 1} \|Au\|_2 ,$$

where $\|.\|_2$ stands for the usual norm on $l^2(G)$.
Notice that, as the degree of $\mathbf{G}$ and the entries of W are both bounded, W lies in B_G, and we have

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

Recall that for any bounded Hilbertian operator $A \in B_G$, the spectrum $\mathrm{Sp}(A)$ is defined as the set of all complex numbers λ such that $\lambda \,\mathrm{Id} - A$ is not invertible (here Id stands for the identity on $l^2(G)$). Since W is bounded and symmetric, $\mathrm{Sp}(W)$ is a non-empty compact subset of $\mathbb{R}$ [58].
We aim now at providing a spectral representation of any bounded normal Hilbertian operator. For this, first recall the definition of a resolution of identity (see for example [58]) :

Definition 3.1.1 *Let $\mathcal{M}$ be a σ-algebra over a set Ω. We call identity resolution (on $\mathcal{M}$) a map*

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

such that,

1. *$E(\emptyset) = 0, E(\Omega) = I$.*
2. *For any $\omega \in \mathcal{M}$, the operator $E(\omega)$ is a projection operator.*
3. *For any $\omega, \omega' \in \mathcal{M}$, we have*

$$E(\omega \cap \omega') = E(\omega)E(\omega') = E(\omega')E(\omega).$$

4. *For any $\omega, \omega' \in \mathcal{M}$ such that $\omega \cap \omega' = \emptyset$, we have*

$$E(\omega \cup \omega') = E(\omega) + E(\omega').$$

We can now recall the fundamental decomposition theorem (see for example [58])

Theorem 3.1.2 (Spectral decomposition) *If $A \in B_G$ is symmetric, then there exists a unique identity resolution E over all Borelian subsets of* $\mathrm{Sp}(A)$*, such that*

$$A = \int_{\mathrm{Sp}(A)} \lambda \mathrm{d}E(\lambda).$$

From the last theorem, we obtain the spectral representation of the adjacency operator W thanks to an identity resolution E over the Borelians of $\mathrm{Sp}(W)$

$$W = \int_{\mathrm{Sp}(W)} \lambda \mathrm{d}E(\lambda).$$

Obviously, we have

$$W^k = \int_{\mathrm{Sp}(W)} \lambda^k \mathrm{d}E(\lambda), k \in \mathbb{N}.$$

Define now, 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 the real measure μ_{ij} by

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

Hence, we can write :

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

This family of measures $\mu_{ij}, i, j \in G$ will be used in the whole paper. They convey both spectral information of the adjacency operator, and combinatorial information on the number of path and loops in $\mathbf{G}$. Indeed, the quantity $\left(W^k\right)_{ij}$ is the number of path (counted with their weights) going from i to j with length k.
Note also that all diagonals measures $\mu_{ii}, i \in G$ are probability measures.

The adjacency operator of $\mathbb{Z}$ and its spectral decomposition

In the usual case of $\mathbb{Z}$, an explicit expression for μ_{ij} can be given.
Denote $T_k(X)$ the k^{th}-Chebychev polynomial ($k \in \mathbb{N}$). We can provide the spectral decomposition of $W^{(\mathbb{Z})}$ ($W^{(\mathbb{Z})}$ has been defined in Equation 3.1).

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

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

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

Notice that we recover the usual spectral decomposition pushing forward μ_{ij} by the function cos :

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

We get

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

Time series, spectral representation, and MA_∞

Our aim is to study some kind of stationary processes indexed by the vertices G of the graph $\mathbf{G}$. To begin with, let us recall the usual case of $\mathbb{Z}$. In particular, let us introduce Toeplitz operators associated to stationary time series.
Let $\mathbf{X} = (X_i)_{i\in\mathbb{Z}}$ be a stationary Gaussian process indexed by $\mathbb{Z}$. Since $\mathbf{X}$ is Gaussian, stationarity is equivalent to second order stationarity, that is, $\forall i,k \in \mathbb{Z}, \mathrm{Cov}(X_i, X_{i+k})$ does not depend on i. Thus, we can define

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

Aassume further that $(r_k)_{k\in\mathbb{Z}} \in l^1(\mathbb{Z})$. This leads to a particular form of the covariance operator Γ defined on $l^2(\mathbb{Z})$ by

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

Recall that $B_{\mathbb{Z}}$ denotes here the set of bounded Hilbertian operators on $l^2(\mathbb{Z})$. Notice that, since $(r_k)_{k\in\mathbb{Z}} \in l^1(\mathbb{Z})$, we have $\Gamma \in B_{\mathbb{Z}}$ (see for instance [23] for more details). This bounded operator is constant over each diagonals, and is therefore called a Toeplitz operator (see also [20] for a general introduction to Toeplitz operators).
As $(r_k)_{k\in\mathbb{Z}} \in l^1(\mathbb{Z})$, we have

$$\forall i,j \in \mathbb{Z}, \mathcal{T}(g)_{ij} := \Gamma_{ij} = \frac{1}{2\pi}\int_{[0,2\pi]} g(t)\cos\left((i-j)t\right)\mathrm{d}t,$$

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

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

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

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

Let, for $\lambda \in [-1,1]$,

$$f(\lambda) := 2\sum_{k\in\mathbb{N}^*} r_k T_k(\lambda) + r_0 T_0(\lambda). \tag{3.2}$$

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

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

Notice that the last expression may also be written as $\Gamma = f(W^{(\mathbb{Z})})$, and the convergence of the operator valued series defined by Equation 3.2 is ensured by the boundedness of $W^{(\mathbb{Z})}$ and of the Chebychev polynomials $(T_k([-1,1]) \subset [-1,1], \forall k \in \mathbb{Z})$, together with the summability of the sequence $(r_k)_{k\in\mathbb{Z}}$.
We will extend usual MA processes to any graph, using this previous remark. This will be the purpose of Section 3.2.
Let us recall some properties about the moving average representation MA_∞ of a process on $\mathbb{Z}$. This representation exists as soon as the log of the spectral density is integrable (see for instance [23] for the Wold decomposition of a stationary sequence). In this case, there exists a sequence $(a_k)_{k\in\mathbb{N}}$, with $a_0 = 1$, and a Gaussian 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 Γ of $\mathbf{X}$:

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

This implies the equality

$$f\left(\cos(t)\right) = \left|h(e^{it})\right|^2.$$

Recall that when h is a polynomial of degree p (with non null first coefficient), the process is said to be MA_p. In this case, f is also a polynomial of degree p. Reciprocically, if f is a real polynomial of degree p, and as soon as $f\left(\cos(t)\right)$ is even, and non-negative for any $t \in [0,2\pi]$, the Fejér-Riesz theorem provides a factorization of $f\left(\cos(t)\right)$ such that $f\left(\cos(t)\right) = |h(e^{it})|^2$ (see for instance [48]). This proves that $\mathbf{X}$ is MA_p if, and only if, its covariance operator may be written $f(W^{(\mathbb{Z})})$, where f is a polynomial of degree p.
This remark is fundamental for the construction we provide in the following section (see Definition 3.2.1).

Whittle maximum likelihood estimation for time series

Here, we recall briefly the Whittle's approximation for time series. Let Θ be a compact interval of $\mathbb{R}^d, d \geq 1$, and $(f_\theta)_{\theta\in\Theta}$ be a parametric family of spectral densities. Let $\theta_0 \in \Theta$, and assume that $(X_i)_{i\in\mathbb{Z}}$ is a Gaussian time series whith spectral density f_{θ_0}.
If we observe $\mathbf{X}_n := (X_i)_{i=1,\cdots n}, n > 0$, we can define the maximum likelihood estimate $\hat{\theta}_n$ of θ_0 as :

$$\hat{\theta}_n := \arg\max L_n(\theta, \mathbf{X}_n),$$

where

$$L_n(\theta, \mathbf{X}_n) := -\frac{1}{2}\left(n\log(2\pi) + \log\det\left(\mathcal{T}_n(f_\theta)\right) + \mathbf{X}_n^T\Big(\mathcal{T}_n(f_\theta)\Big)^{-1}\mathbf{X}_n\right).$$

This estimator is consistent as soon as the spectral densities are regular enough, and under assumptions on the function $\theta \mapsto f_\theta$ (see for instance [8]). However, in practical situations, it is hard to compute. The Whittle's estimate is built by maximizing an approximation of the likelihood instead of the likelihood itself :

$$\tilde{\theta}_n := \arg\max \tilde{L}_n(\theta, \mathbf{X}_n),$$

where

$$\tilde{L}_n(\theta, \mathbf{X}_n) := -\frac{1}{2}\left(n\log(2\pi) + n\int_{[0,2\pi]}\log\left(f_\theta(\lambda)\right)\mathrm{d}\lambda + \mathbf{X}_n^T\mathcal{T}_n(\frac{1}{f_\theta})\mathbf{X}_n\right).$$

The Whittle estimate is also consistent and asymptotically normal and efficient, as soon as the spectral densities are regular enough.
The consistency of the Whittle estimate relies on the Szegö's Lemma, which provide a bound on the error between $\frac{1}{n}\log\det\left(\mathcal{T}_n(f_\theta)\right)$ and $\int_{[0,2\pi]}\log\left(f_\theta(\lambda)\right)$. There exists many versions of this Lemma (see for instance [8], [41]).
In this work, we are interested in a weak version given by Azencott and Dacunha-Castelle in [8]. The lemma relies on the following fondamental inequality : Let $f(x) = \sum_{k\in\mathbb{N}} f_k x^k$ and $g(x) = \sum_{k\in\mathbb{N}} g_k x^k$ be two analytic functions on the complex unitar disk. Then we have

$$\sum_{i,j=1,\cdots,N}\left|\Big(\mathcal{T}_N(f)\mathcal{T}_N(g) - \mathcal{T}_N(fg)\Big)_{ij}\right| \leq \frac{1}{2}\sum_{k\in\mathbb{N}}(k+1)f_k\sum_{k\in\mathbb{N}}(k+1)g_k. \tag{3.3}$$

In the following, we aim at developing the same kind of tools for processes indexed by a graph.

3.2 Spectral definition of $ARMA$ processes

In this section, we will define moving average and autoregressive processes over the graph $\mathbf{G}$.
As explained in the last section, since W is bounded and self-adjoint, $\mathrm{Sp}(W)$ is a non-empty compact subspace of $\mathbb{R}$, and W admits a spectral decomposition thanks to an identity resolution E, given by

$$W = \int_{\mathrm{Sp}(W)} \lambda \mathrm{d}E(\lambda).$$

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

Definition 3.2.1 *Let $(X_i)_{i\in G}$ be a Gaussian process, indexed by the vertices G of the graph $\mathbf{G}$, and Γ its covariance operator.*
If there exists an analytic function f defined on the convex hull of $\mathrm{Sp}(W)$, such that

$$\Gamma = \int_{\mathrm{Sp}(W)} f(\lambda)\mathrm{d}E(\lambda),$$

we will say that X is
- *MA_q if f is a polynomial of degree q.*
- *AR_p if $\frac{1}{f}$ is a polynomial of degree p which has no root in the convex hull of $\mathrm{Sp}(W)$.*
- *$ARMA_{p,q}$ if $f = \frac{P}{Q}$ with P a polynomial of degree p and Q a polynomial of degree q with no roots in the convex hull of $\mathrm{Sp}(W)$.*

*Otherwise, we will talk about the MA_∞ representation of the process $\mathbf{X}$. We call f the **spectral density** of the process $\mathbf{X}$, and denote its corresponding covariance operator by*

$$\Gamma = \mathcal{K}(f).$$

Remark Actually, this last construction may also be understood as

$$\Gamma = \mathcal{K}(f) = f(W),$$

in the sense of normal convergence of the associated power series. However, the spectral representation will be useful in the following. Even if we consider only regular processes in this works, the definition using the spectral representation allows weaker regularity than the definition using the normal convergence of the associated power series.

This kind of modeling is interesting when the interactions are locally propagated (that may be for instance a good modeling for traffic problems.).
The notation $\mathcal{K}(.)$ has to be understood by analogy with the notation $\mathcal{T}(.)$ used for Toeplitz operators.
Notice that, in the usual case of $\mathbb{Z}$, and for finite order $ARMA$, we recover the usual definition as shown in Subsection 3.1. So, the last definition may be seen as an extension of isotropic $ARMA$ for any graph $\mathbf{G}$. Besides, note that this extension is given by the equivalence, for any $g \in \mathbb{L}^2([0, 2\pi])$, such that $\int_{[0,2\pi]} \log(g) < +\infty$,

$$\forall f \in \mathbb{L}^2([-1, 1]), (g(t) = f(\cos(t)) \Leftrightarrow \mathcal{T}(g) = \mathcal{K}(f)).$$

This means that, in the usual case $\mathbf{G} = \mathbb{Z}$, the definition of spectral density in our framework is the usual one, up to a change of variable $\lambda = \cos(t)$ (see Subection 3.1).
Now, we get a representation of moving average processes over any graph $\mathbf{G}$. The following section gives the main result of this paper. It deals with the maximum likelihood identification.

3.3 Convergence of maximum approximated likelihood estimators

In this section as before, $\mathbf{G} = (G, W)$ is a graph with bounded degree. 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} (defined in Section 3.2) depending on an unknown parameter $\theta_0 \in \Theta$. We aim at estimating θ_0. For this, we will generalize classical maximum likelihood estimation of time series.
We will also develop a Whittle's approximation for $ARMA$ processes indexed by the vertices of a graph. We follow here the guidelines of the proof given in [8] for the usual case of time series.

Framework and Assumptions

Let us now specify the framework of our study. Let $(\mathbf{G}_n)_{n\in\mathbb{N}}$ be a growing sequence of finite nested subgraphs (induced by $\mathbf{G}$). This means that if $\mathbf{G}_n = (G_n, W_n)$, we have $G_n \subset G_{n+1} \subset G$ and for any $i, j \in G_n$, $W_n(i, j) = W(i, j)$.
Let $m_n = \text{Card}(G_n)$. We set also

$$\delta_n = \text{Card}\{i \in G_n, \exists j \in G\backslash G_n, W_{ij} \neq 0\}.$$

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 that also appears in isoperimetrical [56] and graph expander issues. We will assume here that this ratio goes to 0 when the size of the graph goes to infinity. In short, we set

Assumption 3.3.1 $\delta_n = o(m_n)$

This assumption is a non-expansion criterion. The graph has to be amenable, which is satisfied for the last examples $G = \mathbb{Z}^d$ and $G_n = [-n, n]^d$, but not for a homogeneous tree, whatever the choice of the sequence of subgraphs $(\mathbf{G_n})_{n\in\mathbb{N}}$ is.
We will now choose a parametric family of covariance operators of MA processes as defined in the last section. First, let Θ be a compact interval 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 easily extended to the case $\Theta \subset \mathbb{R}^k, k \geq 1$.
Define $\mathcal{F}$ as the set of positive analytic functions over the convex hull of $\mathrm{Sp}(W)$. Let also $(f_\theta)_{\theta\in\Theta}$ be a parametric family of functions of $\mathcal{F}$. They define a parametric set of covariances on G (see Section 3.2) by

$$\mathcal{K}(f_\theta) = f_\theta(W).$$

As in [8], we will need a strong regularity for this family of spectral densities.
Let us introduce a regularity factor for any analytic function

$$f \in \mathcal{F}, \forall x \in \mathrm{Sp}(W), f(x) = \sum_k f_k x^k,$$

by setting

$$\alpha(f) := \sum_{k\in\mathbb{N}} |f_k|\,(k+1). \tag{3.4}$$

Now, let $\rho > 0$ and define,

$$\mathcal{F}_\rho := \{f \in \mathcal{F}, \alpha(\log(f)) \leq \rho\}\,. \tag{3.5}$$

Notice that for any $f \in \mathcal{F}_\rho$, we have $\alpha(f) \leq \sum \frac{\alpha(f)^k}{k!} \leq e^\rho, \alpha(\frac{1}{f}) \leq e^\rho$.
We need the following assumption

Assumption 3.3.2
– *The map $\theta \to f_\theta$ is injective.*
– *For any $\lambda \in Sp(W)$, the map $\theta \to f_\theta(\lambda)$ is continuous.*
– $\forall\theta \in \Theta, f_\theta \in \mathcal{F}_\rho$.

From now on, consider $\theta_0 \in \mathring{\Theta}$. Let $\mathbf{X}$ be a centered Gaussian MA_∞ process over $\mathbf{G}$ with covariance operator $\mathcal{K}(f_{\theta_0})$ (see Section 3.2).
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 $\mathcal{K}_n(f_\theta)$ be its covariance :

$$X_n \sim \mathcal{N}\left(0, \mathcal{K}_n(f_{\theta_0})\right).$$

The corresponding log-likelihood at θ is

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

As discussed before, in the case $G = \mathbb{Z}$, it is usual to maximize an approximation of the likelihood. The classical approximation is the Whittle's one ([41]), where

$$\frac{1}{n}\log\det\left(\mathcal{T}_n(g)\right)$$

is replaced by

$$\frac{1}{2\pi}\int_{[0,2\pi]} \log\left(g\left(t\right)\right) \mathrm{d}t.$$

Back to the general case, we aim at performing the same kind of approximation. For this, we will need the following assumption to ensure the convergence of $\log\det\left(\mathcal{K}_n(f_\theta)\right)$ (see Section 3.1 for the definition of μ_{ii}) :

Assumption 3.3.3 *There exist a positive measure μ, such that*

$$\frac{1}{m_n}\sum_{i\in G_n} \mu_{ii} \underset{n\to\infty}{\overset{\mathcal{D}}{\to}} \mu.$$

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

The limit measure μ is classically called the spectral measure of $\mathbf{G}$ with respect to the sequence of subgraphs $(\mathbf{G}_n)_{n\in\mathbb{Z}}$ (see [55] for example).
Actually, under Assumption 3.3.1, Assumption 3.3.3 is equivalent to the convergence of the empirical distribution of eigenvalues of W_{G_n} (here, W_{G_n} denotes the restriction of W over the subgraph G_n).
That is, if $\lambda_1^{(n)}, \cdots, \lambda_{m_n}^{(n)}$ denote the eigenvalues (written with their multiplicity orders) of W_{g_n}, Define

$$\mu_n^{[1]} := \frac{1}{m_n}\sum_{i=1}^{m_n} \delta_{\lambda_i^{(n)}},$$

and

$$\mu_n^{[2]} = \frac{1}{m_n}\sum_{i\in G_n} \mu_{ii}.$$

Then, under Assumption 3.3.1, the convergence of $\mu_n^{[1]}$ to μ (i.e. Assumption 3.3.3) is equivalent to the convergence of $\mu_n^{[2]}$ to μ.
To prove this equivalence we just have to notice that :

$$\begin{aligned}\int_{\mathrm{Sp}(W)} \lambda^k \mathrm{d}\mu_n^{(1)}(\lambda) &- \int_{\mathrm{Sp}(W)} \lambda^k \mathrm{d}\mu_n^{(2)}(\lambda) \\ &= \frac{1}{m_n}\sum_{i=1}^{m_n}\left(\lambda^{(n)}\right)_i^k - \frac{1}{m_n}\sum_{i\in G_n}(W^k)_{ii} \\ &= \frac{1}{m_n}\mathrm{Tr}\left((W_{G_n})^k\right) - \frac{1}{m_n}\mathrm{Tr}\left((W^k)_{G_n}\right).\end{aligned}$$

So that, we get the result by Lemma 3.6.1 (see Section 3.6).
As in the case of time series (for $G = \mathbb{Z}$), we can approximate the log-likelihood. It avoids an inversion of a matrix and a computation of a determinant. Indeed, we will consider the two following approximations.

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

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

Notice that approximated maximum likelihood estimators are not asymptotically normal in general (see for instance [42] for $\mathbb{Z}^d$). Indeed, the score associated to the approximated log-likelihood has to be asymptotically unbiased [8].
To overcome this problem in $\mathbb{Z}^d$, the tapered periodogram can be used (see [43], [42], [30]).
Let us consider graph extensions of standard time series models :
– The MA_P case : There exists $P > 0$ such that the true spectral density f_{θ_0} is a polynomial of degree bounded by P.
– The AR_P case : There exists $P > 0$ 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 P.

So, to define the good approximated log-likelihood, we first introduce the unbiased periodogram in each of the last cases. Now, let $P > 0$.
Define a subset V_P of signed measures on $\mathbb{R}$ as

$$V_P := \{\mu_{ij}, i, j \in G, d_{\mathbf{G}}(i,j) \leq P\},$$

where $d_{\mathbf{G}}(i,j), i, j \in G$ stands for the usual distance on the graph $\mathbf{G}$, i.e. the length of the shortest path going from i to j.
We will need the following assumption

Assumption 3.3.4 *The set V_P of possible local measures over G is finite, and n is large enough to ensure that*

$$\forall v \in V_P, \exists (i,j) \in G_n^2, \mu_{ij} = v.$$

Remark This assumption is quite strong, and holds for instance for quasi-transitive graphs (i.e. such that the quotient of the graph with its automorphism group is finite). This assumption may be relaxed, but it is a hard and technical work that will be the issue of a forthcoming paper.

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

$$\begin{aligned} B_{ij}^{(n)} &:= \frac{\text{Card}\left\{(k,l) \in G_n \times G, \mu_{kl} = \mu_{ij}\right\}}{\text{Card}\left\{(k,l) \in G_n \times G_n, \mu_{kl} = \mu_{ij}\right\}}, \text{ if }, d_{\mathbf{G}}(k,l) \leq P \\ &:= 1 \text{ if } d_{\mathbf{G}}(k,l) > P. \end{aligned}$$

The matrix $B^{(n)}$ gives a boundary correction, comparing, for any $v \in V_P$ the frequency of the interior couples of vertices with local measure v with the boundary couples of vertices with local measure v. Actually, this way to deal with the edge effect is very similar to the one used for $\mathbf{G} = \mathbb{Z}^d$ (see [30], [42]).
As example, let us now describe the case $G = \mathbb{Z}^2$, for $P = 2$. In this case $W^{(\mathbb{Z}^2)}$ is

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

In this example, we set $G_n = [1,n]^2$, and we can compute the matrix $B^{(n)}$. Indeed, it is only needed to notice that

$$\mu_{(i_1,j_1),(i_1+k,j_1+l)} = \mu_{(i_2,j_2),(i_2+\epsilon_1 k, j_2+\epsilon_2 l)}, i_1, i_2, j_1, j_2, k, l \in \mathbb{Z}, \epsilon_1, \epsilon_2 \in \{-1,1\}.$$

This means that the local measure of a couple of vertices only depends of their relative positions (stationarity and isotropy of this set of measure). So, we need to count the configurations given by Figure 3.1 since we consider only couples of vertices $u, v \in \mathbb{Z}^2$ such that $d_{\mathbb{Z}^2}(u,v) \leq 2$.
We get, for any $i,j \in \mathbb{Z}$,
– $B^{(n)}_{(i,j),(i,j)} = \frac{n^2}{n^2} = 1$.
– $B^{(n)}_{(i,j),(i,j\pm 1)} = B^{(n)}_{(i,j),(i\pm 1,j)} = \frac{4n^2}{4n(n-1)}$.
– $B^{(n)}_{(i,j),(i\pm 1,j\pm 1)} = \frac{n^2}{4(n-1)^2}$.
– $B^{(n)}_{(i,j),(i,j\pm 2)} = B^{(n)}_{(i,j),(i\pm 2,j)} = \frac{4n^2}{4n(n-2)}$
One can notice that

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

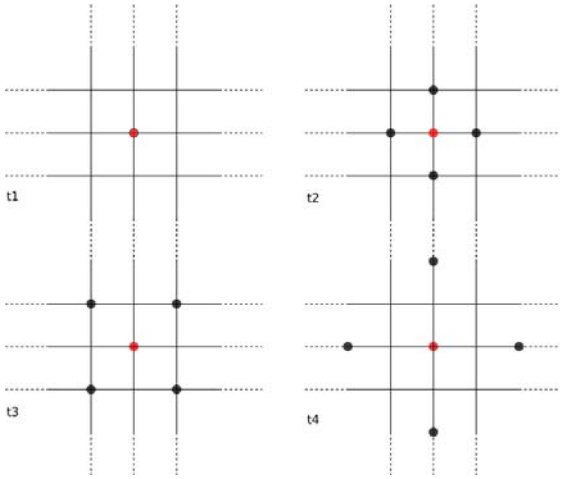


FIGURE 3.1: Possible configurations for couple of vertices

Assumption 3.3.5 ensure that this property holds for the graph we consider.
Back to the general case, let $f \in \mathcal{F}_\rho$. We define the unbiased periodogram as

$$X_n^T \mathcal{Q}_n(\frac{1}{f}) X_n.$$

where

$$\mathcal{Q}_n(f) := B^{(n)} \odot \mathcal{K}_n(f).$$

Here, the operation $\odot$ denotes the Hadamard product for matrices, that is

$$\forall i,j \in G_n, \left(B^{(n)} \odot \mathcal{K}_n(f)\right)_{ij} = \left(B^{(n)}\right)_{ij} \mathcal{K}_n(f)_{ij}.$$

Notice that this is actually a way to extend the so called tapered periodogram (see for instance [42]).
We now define the unbiased empirical log-likelihood, for any $\theta \in \Theta$

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

We denote by $\hat{\theta}_n$, $\tilde{\theta}_n$, $\bar{\theta}_n$, $\theta^{(u)}$ the maximum likelihood estimators associated to L_n, $\tilde{L}_n$, $\bar{L}_n$, $L_n^{(u)}$, respectively.
We will need the following assumption,

Assumption 3.3.5 *There exists a positive sequence $(u_n)_{n\in\mathbb{N}}$ such that,*

$$u_n \underset{n\to\infty}{\to} 0,$$

and

$$\sup_{ij} \left| B_{ij}^{(n)} - 1 \right| \leq u_n.$$

Notice that the last assumption holds for example in the case $\mathbf{G} = \mathbb{Z}^d, d > 1$.
To prove asymptotic normality and efficiency of the estimator $\theta_n^{(u)}$, we will also need the following assumption.

Assumption 3.3.6 *Assume that*
– *There exists a positive sequence* $(v_n)_{n\in\mathbb{N}}$ *such that* $v_n = o(\frac{1}{\sqrt{m_n}})$ *and*

$$\forall f \in \mathcal{F}_\rho, \left| \frac{1}{m_n} \operatorname{Tr}(\mathcal{K}_{G_n}(f)) - \int f \mathrm{d}\mu \right| \leq \alpha(f) v_n.$$

– *For any* $\theta \in \Theta$, f_θ *is twice differentiable on* Θ *and*

$$\frac{\mathrm{d}}{\mathrm{d}\theta}(f_\theta) \in \mathcal{F}_\rho, \frac{\mathrm{d}^2}{\mathrm{d}\theta^2}(f_\theta) \in \mathcal{F}_\rho.$$

The first assumption means that the convergence of the empirical distribution of eigenvalues of $\mathcal{K}(f)$ to the spectral measure μ is faster than $\frac{1}{\sqrt{m_n}}$. It holds for instance for quasi-transitives graphs, with a suitable sequence of subgraphs. The second assumption is more classical. For example it is required in the case $\mathbf{G} = \mathbb{Z}$ (see [8]).

Convergence and asymptotic optimality

Let $\rho > 0$. We can now state one of our main result :

Theorem 3.3.7 *Under Assumptions 3.3.1, 3.3.2 and 3.3.3, the sequences* $(\hat{\theta}_n)_{n\in\mathbb{N}}$, $(\bar{\theta}_n)_{n\in\mathbb{N}}$, $(\tilde{\theta}_n)_{n\in\mathbb{N}}$ *converge, as* n *goes to infinity,* $P_{f_{\theta_0}}$*-a.s. to the true value* θ_0*. If moreover Assumption 3.3.5 holds, this is also true for* $(\theta_n^{(u)})_{n\in\mathbb{N}}$*.*

Proof. The proof follows the guidelines of [8]. We highlight the main changes performed here. First, we define the Kullback information on G_n of f_{θ_0} with respect to $f \in \mathcal{F}_\rho$, by

$$\mathbb{IK}_n(f_{\theta_0}, f) := \mathbb{E}_{P_{f_{\theta_0}}} \left[-\log\left(\frac{\mathrm{d}P_f^{(n)}}{\mathrm{d}P_{f_{\theta_0}}^{(n)}}\right) \right].$$

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

$$\mathbb{IK}(f_{\theta_0}, f) = \lim_n \frac{1}{m_n} \mathbb{IK}_n(f_{\theta_0}, f)$$

whenever it is finite.

The convergence of the estimators of the maximum approximated likelihood is a direct consequence of the following lemmas :

Lemma 3.3.8 *For any $f \in \mathcal{F}_\rho$, and under Assumptions 3.3.1, 3.3.2 and 3.3.3, the asymptotic Kullback information exists and may be written as*

$$\mathbb{IK}(f_{\theta_0}, f) = \frac{1}{2}\int \left(-\log(\frac{f_{\theta_0}}{f}) - 1 + \frac{f_{\theta_0}}{f} \right) \mathrm{d}\mu.$$

Furthermore, if we set $l_n(\theta, X_n) = \frac{1}{m_n} L_n(\theta, X_n)$, we have that $P_{f_{\theta_0}}$-a.s.,

$$l_n(\theta_0, X_n) - l_n(\theta, X_n) \underset{n\to\infty}{\to} \mathbb{IK}(f_{\theta_0}, f_\theta)$$

uniformly in $\theta \in \Theta$.
This property also holds for $\bar{l}_n := \frac{1}{m_n}\bar{L}_n$ and $\tilde{l}_n := \frac{1}{m_n}\tilde{L}_n$
Furthermore, for $P > 0$, and for both the AR_P or the MA_P case (see above), this also holds for $l_n^{(u)} := \frac{1}{m_n} L_n^{(u)}$.

Lemma 3.3.9 *Let f_{θ_0} be the true spectral density, and $(\ell_n)_{n\in\mathbb{N}}$ be a deterministic sequence of continuous functions such that*

$$\forall \theta \in \Theta, \ell_n(\theta_0) - \ell_n(\theta) \underset{n\to\infty}{\to} \mathbb{IK}(f_{\theta_0}, f_\theta)$$

uniformly as n tends to infinity. Then, if $\theta_n = \arg\max_\theta \ell_n(\theta)$, we have

$$\theta_n \underset{n\to\infty}{\to} \theta_0.$$

The proofs of these lemmas are postponed in Appendix (Subsection 3.6). ∎

Theorem 3.3.10 *In both the AR_P or MA_P cases, and and under all previous assumptions 3.3.1, 3.3.2, 3.3.3, 3.3.4, 3.3.5, 3.3.6, the estimator $\theta_n^{(u)}$ of θ_0 is asymptotically normal :*

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

Furthermore, the Fisher information of the model is

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

Hence, the previous estimator is asymptoticly efficient.

Proof. Here again, we mimic the usual proof by extending the result of [8] to the graph case.
Using a Taylor expansion, we get

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

where $\breve{\theta}_n \in \left]\theta_n^{(u)}, \theta_0\right[$. As $\theta_n^{(u)} = \arg\max l_n^{(u)}$, we have

$$(l_n^{(u)})'(\theta_n^{(u)}) = 0.$$

So that,

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

The end of the proof relies on three lemmas :
Lemma 3.3.11 provides the asymptotic normality for $\sqrt{m_n}(l_n^{(u)})'(\theta_0)$. Combined with Lemma 3.3.12, we get the asymptotic normality for $\sqrt{m_n}(\theta_0 - \theta_n^{(u)})$. Finally, Lemma 3.3.13 gives the asymptotic Fisher information. This information is defined as

$$J(\theta) = \lim_{n\to\infty} \mathbb{E}_{\mathbb{P}_{f_{\theta_0}}} \left[\left(\frac{\partial}{\partial\theta} l_n(X_n, \theta)\right)^2\right]$$

Lemma 3.3.11 *Under assumptions of Theorem 3.3.10,*

$$\sqrt{m_n}(l_n^{(u)})'(\theta_0) \underset{n\to\infty}{\overset{\mathcal{D}}{\to}} \mathcal{N}\left(0, \frac{1}{2}\int \left(\frac{f'_{\theta_0}}{f_{\theta_0}}\right)^2 \mathrm{d}\mu\right).$$

Lemma 3.3.12 *Under assumptions of Theorem 3.3.10,*

$$\left((l_n^{(u)})''(\breve{\theta}_n)\right)^{-1} \underset{n\to\infty}{\to} 2\left(\int \left(\frac{f'_{\theta_0}}{f_{\theta_0}}\right)^2 \mathrm{d}\mu\right)^{-1}, P_{f_{\theta_0}} - \ a.s.$$

Lemma 3.3.13 *Under assumptions of Theorem 3.3.10, the asymptotic Fisher information is :*

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

The proofs of these lemmas are postponed in Appendix (Subsection 3.6)

■

3.4 Discussion

Note first that Theorem 3.3.7 provides consistency of the estimators under weak conditions on the graph. Indeed, amenability ensures Assumption 3.3.1, for a suitable sequence of subgraphs. Assumption 3.3.3 holds as soon as there is a kind of homogeneity in the graph. The simplest application is quasi-transitives graph. Note that if $\mathbf{G}$ is "close" to be quasi-transitive, Assumption 3.3.3 is still true. We also could adapt notions of unimodularity [1] or stationarity [11] to our framework and prove the existence of a spectral measure. Furthermore, Assumption 3.3.3 holds for the real traffic network (this will be explained in a forthcomming paper).
To build the estimator $\theta_n^{(u)}$, stronger assumptions on the graph $\mathbf{G}$ are needed. Let us discuss two very special cases. First, Theorem 3.3.10 may be applied in the $\mathbb{Z}^d$ case with holes, that is in the presence of missing data, up to the condition that they remain few enough. Actually, Assumption 3.3.1 is required, so the boundary of the subgraphs (counting the holes) has to be small in front of the volume of this subgraphs.
We need furthermore a kind of homogeneity for these holes. For instance, we can assume that the data are missing completely at random. This particular case is interesting for prediction issues.
Another strong potential application is quasi-transitive graphs, as mentioned above. Indeed, take for instance a finite graph (the pattern) and reproduce it at each vertex of an infinite (amenable) vertex-transitive graph. The final graph is then quasi-transitive, and all the previous assumptions hold.
This seems to be a natural extension of what happens for $\mathbb{Z}^d$. Furthermore, in this situation as in $\mathbb{Z}^d$, our work may also be applied to a process with missing values.

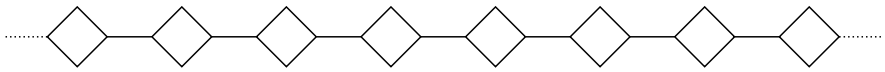
Note also that conditions of both amenability of the graphs and regularity of spectral densities seem natural, looking at the Szegö's Lemmas (see Section 3.6). Indeed, the difference computed in Lemma 3.6.1 is only due to edge effects.

Thus, there are two ways for relaxing this conditions. On the one hand, it could be interesting to deal with lower regularity (for instance to study long memory processes) for the spectral densities. On the other hand, it could be also interesting to relax conditions on the graph, for instance for more regular densities. In particular, we could investigate the case of random graphs, and try to pick up homogeneity conditions into the random structure. As mentioned above, another natural extension of this work could be done to graphs "close" to be quasi-transitive.
These two limits of our present work are actually two of our main perspectives in this framework.

3.5 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 both on the left and right (see Figure 3.2).

FIGURE 3.2: Graph G



The sequence of nested subgraphs chosen here is the growing neighborhood sequence (we chose a point x and we take $G_n = \{y \in G, d_{\mathbf{G}}(x,y) \leq n\}$). We study an AR_2 model, where,

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

Here, we take for W the adjacency operator of G normalized in order to get $\sup_{i,j\in G} W_{ij} \leq \frac{1}{\deg(G)}$. We choose $\theta_0 = \frac{1}{2}$, $m_n = 724$. We approximate the spectral measure of G by the spectral measure of a very large graph (around 10000 vertices) built in the same way. Figure 3.3 shows the empirical spectrum of the graph G with respect to the sequence of subgraphs $(G_n)_{n\in\mathbb{N}}$.
To compute $(\mathcal{K}_n(f_\theta))^{-1}$, we use the power series representation of f_θ, and truncate this expression after the 15 first coefficient. This choice ensures that the simulation errors are neglectible with respect to the theoretical ones.
Figure 3.4 gives the empirical distribution of

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

FIGURE 3.3: Empirical spectrum

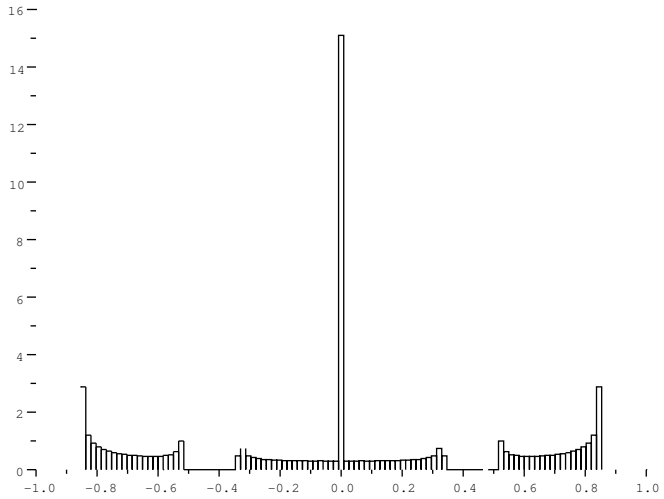
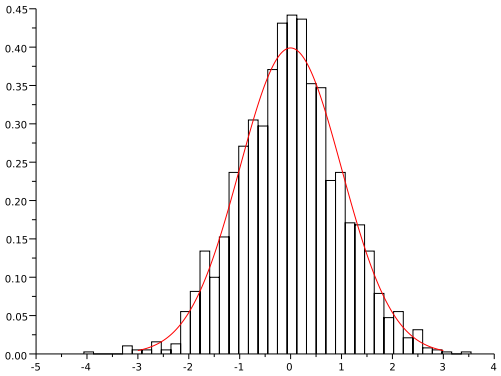


FIGURE 3.4: Empirical distribution of $\sqrt{m_n}\sqrt{\int_{\mathrm{Sp}(A)}\left(\frac{f'_\theta}{f_\theta}\right)^2}\left(\tilde{\theta}_n - \theta_0\right)$.



3.6 Appendix

Szegö's Lemmas

Szegö's Lemmas [41] are useful in time series analysis. Indeed, they provide good approximations for the likelihood. As explained in Section 3.3, these approximations of the likelihood are easier to compute.
In this section, we generalize a weak version of the Szegö Lemmas, for a general graph, under Assumption 3.3.1 (non expansion criterion for G_n), and Assumption 3.3.3 (existence of the spectral measure μ).
For any matrix $(B_{ij})_{i,j\in G_n}$, we define the block norm

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

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

Lemma 3.6.1 *Asymptotic homomorphism*
Let k, n be positive integers, and let $g_1, \cdots, g_k$ be analytic functions over $[-1,1]$ having finite regularity factors (i.e. $\alpha(g_i) < +\infty, i = 1, \cdots, k$). Then,

$$b_n \left(\mathcal{K}_n(g_1)\cdots\mathcal{K}_n(g_k) - \mathcal{K}_n(g_1\cdots g_k)\right) \leq \frac{k-1}{2}\alpha(g_1)\cdots\alpha(g_k).$$

Corollary 3.6.2 *For any $g \in \mathcal{F}_\rho$ (see the first page of Subsection 3.3 for the definition), and under Assumptions 3.3.1 and 3.3.3,*

$$\frac{1}{m_n}\log\det(\mathcal{K}_n(g)) \underset{n\to\infty}{\to} \int \log(g)\mathrm{d}\mu.$$

Proof. of Lemma 3.6.1 This proof follows again the one of [8]. 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

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

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

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

Introducing

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

we get

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

The coefficient Δ_h is a porosity factor. It measures the weight of the paths of length h going from the interior of G_n to outside.
Note that $\Delta_h \leq h + 1$, so we get

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

Now, we define another norm on B_G :

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

We thus obtain

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

Finally, we get

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

To conclude the proof of the lemma, define, for $f \in \{_\rho$,

$$\tilde{\alpha}(f) = \alpha(f) - f_0.$$

Notice that $\Delta_0 = 0$. Hence, by symmetrization of the last inequality, and since $1 \leq (h+1)$, we have,

$$\begin{aligned}
b_n\left(K(f)K(g) - K(fg)\right) &\leq \frac{1}{2}\left(|f|_{1,pol}\, \tilde{\alpha(g)} - |g|_{1,pol}\, \tilde{\alpha}(f)\right) \\
&\leq \frac{1}{2}\sum_{k,j\geq 1} |f_k|\,|g_j| \left((k+1)+(j+1)\right) + \frac{1}{2}\,|f_0|\,\tilde{\alpha}(g) + \frac{1}{2}\,|g_0|\,\tilde{\alpha}(f) \\
&\leq \frac{1}{2}\sum_{k,j\geq 1} |f_k|\,|g_j| \left((k+1)(j+1)\right) + \frac{1}{2}\,|f_0|\,\tilde{\alpha}(g) + \frac{1}{2}\,|g_0|\,\tilde{\alpha}(f) \\
&\leq \frac{1}{2}\alpha(f)\alpha(g)
\end{aligned}$$

$$b_n\left(\mathcal{K}_n(f)\mathcal{K}_n(g) - \mathcal{K}_n(fg)\right) \leq \frac{1}{2}\alpha(f)\alpha(g). \tag{3.6}$$

To perform the inductive step, we need the following inequalities [59] :

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

Let $k > 1$, and assume that for all $j \leq k-1$, Lemma 3.6.1 holds. Under the previous assumptions, and the inductive hypothesis for $k-1$ we get,

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

which completes the induction step and proves the result. ■

Proof. of Corollary 3.6.2
Let $g \in \mathcal{F}_\rho$, and k be a positive integer. Using Lemma 3.6.1, we have

$$\mathrm{Tr}\left(\mathcal{K}_n(g)^k - \mathcal{K}_n(g^k)\right) \leq \frac{\delta_n}{m_n} b_n \left(\mathcal{K}_n(g)^k - \mathcal{K}_n(g^k)\right). \tag{3.7}$$

Thus, we have, thanks to Assumption 3.3.1

$$\frac{1}{m_n}\mathrm{Tr}\left(\mathcal{K}_n(g)^k - \mathcal{K}_n(g^k)\right) \underset{n\to+\infty}{\to} 0.$$

Denote $\mu_g^{[1]}$ the real measure whose k^{th}-moment is given by

$$\int x^k \mathrm{d}\mu_g^{[1]} = \lim_n \frac{1}{m_n}\mathrm{Tr}\left(\mathcal{K}_n(g)^k\right),$$

and $\mu_g^{[2]}$ the real measure whose k^{th}-moment is given by

$$\int x^k \mathrm{d}\mu_g^{[2]} = \lim_n \frac{1}{m_n}\mathrm{Tr}\left(\mathcal{K}_n(g^k)\right).$$

Notice that both of these measures have support between $\inf g \geq e^{-\rho} > 0$ and $\sup g \leq e^{\rho} < +\infty$, since $\alpha(\log(g)) < \rho$ (see Section 3.3). Therefore, the equality of the moments given by Equation 3.7 gives the equality of the measures $\mu_g^{[1]}$ and $\mu_g^{[2]}$.
So that, we get

$$\frac{1}{m_n}\log\left(\det\left(\mathcal{K}_n(g)\right)\right) - \frac{1}{m_n}\mathrm{Tr}\left(\mathcal{K}_n\left(\log(g)\right)\right) \underset{n\to+\infty}{\to} 0. \tag{3.8}$$

Assumption 3.3.3 completes the proof of the Corollary since it implies that

$$\frac{1}{m_n}\mathrm{Tr}\left(\mathcal{K}_n\left(\log(g)\right)\right) \underset{n\to+\infty}{\to} \int \log(g)\mathrm{d}\mu.$$

■

The following lemma enables to replace $\mathcal{K}_n(g)$ by the unbiased version $\mathcal{Q}_n(g)$ (see Section 3.3 for the definition).

Lemma 3.6.3 *Under Assumptions 3.3.1,3.3.3, 3.3.4 and 3.3.5, and if f or g is a polynomial having degree less than or equal to P, we have*

$$\left|\frac{1}{m_n}\mathrm{Tr}\left((\mathcal{K}_n(f)\mathcal{K}_n(g))^p - (\mathcal{K}_n(f)\mathcal{Q}_n(g))^p\right)\right| \leq 2^p u_n \alpha(f)^p \alpha(g)^p.$$

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 having degree less than or equal to P, we have to control only the number of paths of length less than or equal to P (counted with their weights).
Let p be a positive number. Recall that $\mathcal{Q}_n(\frac{1}{g}) = B^{(n)} \odot \mathcal{K}_n(\frac{1}{g})$ (see Section 3.3), we have,

$$\begin{aligned}
\tfrac{1}{m_n} & \left| \mathrm{Tr}\left(\left(\mathcal{K}_n(f)\mathcal{K}_n(\frac{1}{g}) \right)^p - \left(\mathcal{K}_n(f)\mathcal{Q}_n(\frac{1}{g}) \right)^p \right) \right| \\
\leq & \frac{1}{m_n} \left| \sum_{i\in G_n} \sum_{i_0=i,i_1,\cdots,i_{2p}=i} \prod_{l=0\cdots p} B^{(n)}_{i_{2l}i_{2l+1}} \mathcal{K}_n(\frac{1}{g})_{i_{2l}i_{2l+1}} \mathcal{K}_n(f)_{i_{2l+1}i_{2l+2}} \right. \\
& \left. - \frac{1}{m_n} \sum_{i\in G_n} \sum_{i_0=i,i_1,\cdots,i_{2p}=i} \prod_{l=0\cdots p} \mathcal{K}_n(\frac{1}{g})_{i_{2l}i_{2l+1}} \mathcal{K}_n(f)_{i_{2l+1}i_{2l+2}} \right|.
\end{aligned}$$

Then,

$$\begin{aligned}
\tfrac{1}{m_n} & \left| \mathrm{Tr}\left(\left(\mathcal{K}_n(f)\mathcal{K}_n(\frac{1}{g}) \right)^p - \left(\mathcal{K}_n(f)\mathcal{Q}_n(\frac{1}{g}) \right)^p \right) \right| \\
\leq & \frac{1}{m_n} \sup_{i_1,i_2,\cdots,i_{2p+1}} \left| \prod_{l=0\cdots p-1} B^{(n)}_{i_{2l+1}i_{2l+2}} - 1 \right| \\
& \times \sum_{i\in G_n} \sum_{i_0=i,i_1,\cdots,i_{2p}=i} \prod_{l=0\cdots p} \left| \mathcal{K}_n(\frac{1}{g})_{i_{2l}i_{2l+1}} \mathcal{K}_n(f)_{i_{2l+1}i_{2l+2}} \right| \\
\leq & \frac{1}{m_n} \sup_{i_1,i_2,\cdots,i_{2p+1}} \left| \prod_{l=0\cdots p-1} B^{(n)}_{i_{2l+1}i_{2l+2}} - 1 \right| \\
& \times \sum_{i\in G_n} \sum_{i_0=i,i_1,\cdots,i_{2p}=i} \prod_{l=0\cdots p} \mathcal{K}_n((\frac{1}{g})_{abs})_{i_{2l}i_{2l+1}} \mathcal{K}_n(f_{abs})_{i_{2l+1}i_{2l+2}} \\
\leq & \sup_{i_1,i_2,\cdots,i_{2p+1}} \left| \prod_{l=0\cdots p-1} B^{(n)}_{i_{2l+1}i_{2l+2}} - 1 \right| \left\| \left(K_{G_n}(f_{abs}) K_{G_n}((\frac{1}{g})_{abs}) \right)^p \right\|_{2,in} \\
\leq & \sup_{i_1,i_2,\cdots,i_{2p+1}} \left| \prod_{l=0\cdots p-1} B^{(n)}_{i_{2l+1}i_{2l+2}} - 1 \right| \alpha(f)^p \alpha(\frac{1}{g})^p.
\end{aligned}$$

Using Assumption 3.3.5, we get,

$$\begin{aligned}
\tfrac{1}{m_n} & \left|\mathrm{Tr}\left(\left(\mathcal{K}_n(f)\mathcal{K}_n(\frac{1}{g})\right)^p - \left(\mathcal{K}_n(f)\mathcal{Q}_n(\frac{1}{g})\right)^p\right)\right| \\
\leq & \left|(1+u_n)^p - 1\right| \alpha(f)^p \alpha(\frac{1}{g})^p \\
\leq & \left|(1+u_n-1)\left((1+u_n)^{p-1} + (1+u_n)^{p-2} + \cdots + 1\right)\right| \alpha(f)^p \alpha(\frac{1}{g})^p \\
\leq & \left|u_n\left(2^p - 1\right)\right| \alpha(f)^p \alpha(\frac{1}{g})^p \\
\leq & u_n 2^p \alpha(f)^p \alpha(\frac{1}{g})^p.
\end{aligned}$$

This ends the proof of the Lemma. ■

Finally, the following lemma explains the choice of $B^{(n)}$. The unbiased quadratic form $\mathcal{Q}_n$ is no more than a correction of the error between $\mathcal{K}_n(f)\mathcal{K}_n(g)$ and $\mathcal{K}_n(fg)$.

Lemma 3.6.4 (Exact correction) *Let $f, g \in \mathcal{F}_\rho$, and assume that either f or g is a polynomial of degree less than or equal to P (see Section 3.3). Then, the unbiased quadratic form $\mathcal{Q}_n(f_\theta)$ verify*

$$\mathrm{Tr}\left(\mathcal{K}_n(f)\mathcal{Q}_n(g)\right) = \mathrm{Tr}\left(\mathcal{K}_n(fg)\right).$$

Proof. of Lemma 3.6.4
First, notice that

$$\mathrm{Tr}\left(\mathcal{K}_n(f)\mathcal{Q}_n(g)\right) = \sum_{i,j \in G_n} \mathcal{K}_n(f)_{ij}\mathcal{K}_n(g)_{ij}B^{(n)}_{ij}.$$

Since this expression is symmetric on f, g, we can now consider the case where f is a polynomial of degree less than or equal to P.
Actually, since f is a polynomial, $\mathcal{K}_n(f)_{ij} = 0$ as soon as $d(i,j) > P$ $(i,j \in G)$. Then, if $i, j, k, l \in G$ are such that $\mu_{ij} = \mu_{kl}$, we have

$$\mathcal{K}_n(f)_{ij}\mathcal{K}_n(g)_{ij} = \mathcal{K}_n(f)_{kl}\mathcal{K}_n(g)_{kl}.$$

So that, we may here denote, for convenience, $K(f)_{\mu_{ij}}$.

Using Assumption 3.3.4, this leads to

$$\begin{aligned}
\mathrm{Tr}\left(\mathcal{K}_n(f)\mathcal{Q}_n(g)\right) &= \sum_{i,j\in G_n} \mathcal{K}_n(f)_{ij}\mathcal{K}_n(g)_{ij}B_{ij}^{(n)} \\
&= \sum_{v\in V_P} \sum_{\substack{i,j\in G_n \\ \mu_{ij}=v, d_{\mathbf{G}}(i,j)\leq P}} \mathcal{K}_n(f)_v\mathcal{K}_n(g)_v B_v^{(n)} \\
&= \sum_{v\in V_P} \mathcal{K}_n(f)_v\mathcal{K}_n(g)_v \mathrm{Card}\left\{(i,j)\in G_n\times G_n, \mu_{ij}=v\right\} \\
&\quad \times \frac{\mathrm{Card}\left\{(i,j)\in G_n\times G, \mu_{ij}=v\right\}}{\mathrm{Card}\left\{(i,j)\in G_n\times G_n, \mu_{ij}=v\right\}}, \\
&= \sum_{v\in V_P} \sum_{\substack{(i,j)\in G_n\times G, \\ \mu_{ij}=v, d_{\mathbf{G}}(i,j)\leq P}} \mathcal{K}_n(f)_v\mathcal{K}_n(g)_v B_v^{(n)} \\
&= \sum_{(i,j)\in G_n\times G} \mathcal{K}_n(f)_{ij}\mathcal{K}_n(g)_{ij}B_{ij}^{(n)} \\
&= \mathrm{Tr}\left(\mathcal{K}_n(fg)\right).
\end{aligned}$$

That ends the proof of Lemma 3.6.4. ∎

Proofs of the lemmas of Theorem 3.3.7

Recall that the theorem relies on two lemmas. Lemma 3.3.9 states a condition on deterministic sequences to provide the convergence of the maximizer of these sequences.

Proof. of Lemma 3.3.9 Recall that f_{θ_0} denotes the true spectral density. Let $(\ell_n)_{n\in\mathbb{N}}$ be a deterministic sequence of continuous functions such that

$$\forall\theta\in\Theta, \ell_n(\theta_0)-\ell_n(\theta) \underset{n\to\infty}{\to} \frac{1}{2}\int\left(-\log(\frac{f_{\theta_0}}{f_\theta})-1+\frac{f_{\theta_0}}{f_\theta}\right)\mathrm{d}\mu, \tag{3.9}$$

uniformly as n tends to infinity. Denotes moreover $\theta_n = \arg\max_\theta \ell_n(\theta)$. We aim at proving that

$$\theta_n \underset{n\to\infty}{\to} \theta_0.$$

Using the compactness of Θ, let θ_∞ be an accumulation point of the sequence $(\theta_n)_{n\in\mathbb{N}}$, and $(\theta_{n_k})_{k\in\mathbb{N}}$ be a subsequence converging to θ_∞. As the function

$$\theta\mapsto\frac{1}{2}\int\left(-\log(\frac{f_{\theta_0}}{f_\theta})-1+\frac{f_{\theta_0}}{f_\theta}\right)\mathrm{d}\mu$$

is continuous on Θ, and the convergence of $(\ell_n(\theta_0)-\ell_n(\theta))_{n\in\mathbb{N}}$ is uniform in θ, we have

$$\ell_{n_k}(\theta_0)-\ell_{n_k}(\theta_{n_k}) \xrightarrow{k\to\infty} \frac{1}{2}\int -\log(\frac{f_{\theta_0}}{f_{\theta_\infty}})-1+\frac{f_{\theta_0}}{f_{\theta_\infty}}\mathrm{d}\mu. \tag{3.10}$$

But we can notice that, thanks to the definition of θ_n, $\ell_{n_k}(\theta_0) - \ell_{n_k}(\theta_{n_k}) \leq 0$ So, since the function $x \mapsto -\log(x) + x - 1$ is non negative and vanishes if, and only if, $x = 1$, we get that $f_{\theta_0} = f_{\theta_\infty}$. By injectivity of the function $\theta \to f_\theta$, we get $\theta_\infty = \theta_0$, for any accumulation point θ_∞ of the sequence $(\theta_n)_{n\in\mathbb{N}}$, which ends the proof of this first lemma. ■

Lemma 3.3.8 provides the uniform convergence of the contrasts of maximum likelihood and approximated maximum likelihood to the Kullback information. The proof may be cut into several lemmas.

Proof. of Lemma 3.3.8

First, notice that by construction, we have, for any $\theta \in \Theta$,

$$\mathbb{IK}(f_{\theta_0}, f_\theta) = \lim_n \mathbb{E}\left[\frac{1}{m_n}\left(L_n(f_{\theta_0}, X_n) - L_n(f_\theta, X_n)\right)\right], \tag{3.11}$$

when it exists. Then, we can compute

$$\begin{aligned} l_n(f_{\theta_0}, X_n) - l_n(f_\theta, X_n) &= -\frac{1}{2m_n}\left(\log\det(\mathcal{K}_n(f_{\theta_0})) - \log\det(\mathcal{K}_n(f_\theta))\right) \\ &\quad -\frac{1}{2m_n}\left(X_n^T\mathcal{K}_n(f_{\theta_0})^{-1}X_n - X_n^T\mathcal{K}_n(f_\theta)^{-1}X_n\right) \end{aligned}$$

Corollary 3.6.2 of Lemma 3.6.1 provides the following convergence

$$\frac{1}{m_n}\left(\log\det(\mathcal{K}_n(f_{\theta_0})) - \log\det(\mathcal{K}_n(f_\theta))\right) \underset{n\to\infty}{\to} \int \log\left(\frac{f_{\theta_0}}{f_\theta}\right)\mathrm{d}\mu. \tag{3.12}$$

To prove the existence of $\mathbb{IK}(f_{\theta_0}, f_\theta)$, it only remains to prove the $\mathbb{P}_{f_{\theta_0}}$-a.s. convergence of $\frac{1}{m_n}X_n^T\mathcal{K}_n(f_\theta)^{-1}X_n$ to $\int \frac{f_{\theta_0}}{f_\theta}\mathrm{d}\mu$ as n goes to infinity.
This is ensured by the following Lemma.

Lemma 3.6.5 (Convergence lemma) *For respectively* $\Lambda = \mathcal{K}_n(\frac{1}{f_\theta})$, $\Lambda = (\mathcal{K}_n(f_\theta))^{-1}$ *or* $\Lambda = \mathcal{Q}_n(\frac{1}{f_\theta})$, *we have,*

$$\frac{1}{m_n}X_n^T\Lambda X_n \underset{n\to\infty}{\to} \int \frac{f_{\theta_0}}{f_\theta}\mathrm{d}\mu, \mathbb{P}_{f_{\theta_0}} - a.s.$$

Lemma 3.6.5 combined with Corollary 3.6.2 ensures the $\mathbb{P}_{f_{\theta_0}}$ − a.s. convergence of $\tilde{l}_n(f_{\theta_0}) - \tilde{l}_n(f_\theta)$, $\bar{l}_n(f_{\theta_0}) - \bar{l}_n(f_\theta)$ to $\mathbb{IK}(f_{\theta_0}, f_\theta)$. It provides also the $\mathbb{P}_{f_{\theta_0}}$ − a.s. convergence of $l_n^{(u)}(f_{\theta_0}) - l_n^{(u)}(f_\theta)$ to $\mathbb{IK}(f_{\theta_0}, f_\theta)$ in the AR_P or MA_P cases (see Section 3.3). To complete the assertion of Lemma 3.3.8, it only remains to show the uniform convergences on Θ of the last quantities. This will be done using an equicontinuity argument given by the following Lemma.

Lemma 3.6.6 (Equicontinuity lemma) *For all $n \geq 0$, the sequences of functions*

$$(l_n(f_{\theta_0}, X_n) - l_n(f_\theta, X_n))_{n\in\mathbb{N}}$$

is an $\mathbb{P}_{f_{\theta_0}}$-a.s. equicontinuous sequence on $(\{f_\theta, \theta \in \Theta\}, \|.\|_\infty)$. This property also holds for $\bar{l_n}$,$\tilde{l_n}$. Furthermore, the sequence $\left(l_n^{(u)}(f_{\theta_0}, X_n - l_n^{(u)}(f_\theta, X_n)\right)_{n\in\mathbb{N}}$ is also $\mathbb{P}_{f_{\theta_0}}$-a.s. equicontinuous, on $\left(\{f_\theta, \theta \in \Theta\}, \|.\|_{1,pol}\right)$.

We can now end the proof of Lemma 3.3.8 :
First, notice that the space $\{f_\theta, \theta \in \Theta\}$ is compact for the topology of the uniform convergence. This also holds for $\left(\{f_\theta, \theta \in \Theta\}, \|.\|_{1,pol}\right)$. So, there exists a dense sequence $(f_{\theta_p})_{p\in\mathbb{N}}$. Then, using Lemma 3.6.1 and Corollary 3.6.2, the sequence $\left(l_n(f_{\theta_0}, X_n) - l_n(f_{\theta_p}, X_n)\right)_{n\in\mathbb{N}}$ converges $\mathbb{P}_{f_{\theta_0}}$-a.s. to $\mathbb{IK}(f_{\theta_0}, f_{\theta_p})$.
If a sequence of functions is equicontinuous and converges pointwise on a dense subset of its domain, and if its co-domain is a complete space, then the sequence converges pointwise on all the domain [58].
Using this well known property, we obtain, $\mathbb{P}_{f_{\theta_0}}$-a.s., the pointwise convergence of

$$(l_n(f_{\theta_0}, X_n) - l_n(f_\theta, X_n))_{n\in\mathbb{N}}$$

to $\mathbb{IK}(f_{\theta_0}, f_\theta)$, for any $\theta \in \Theta$.
Furthermore, if a sequence of functions is equicontinuous and converges pointwise on its domain, then this convergence is uniform on any compact subspace of the domain [58].
Thus, we get, $\mathbb{P}_{f_{\theta_0}}$-a.s., the uniform convergence on Θ of the sequence

$$(l_n(f_{\theta_0}, X_n) - l_n(f_\theta, X_n))_{n\in\mathbb{N}}$$

to $\mathbb{IK}(f_{\theta_0}, f_\theta)$.
Using the same kind of arguments, this uniform convergence also holds for $\bar{l_n}$,$\tilde{l_n}$ and $l_n^{(u)}$. This concludes the proof of Lemma 3.3.8.

■

Proof of the technical lemmas

Proof. of Lemma 3.6.5
Let $\theta \in \Theta$. First, consider the case $\Lambda_n = \mathcal{K}_n\left(\frac{1}{f_\theta}\right)$. We aim at proving that

$$\frac{1}{m_n} X_n^T \Lambda_n X_n \underset{n\to\infty}{\to} \int \frac{f_{\theta_0}}{f_\theta} \mathrm{d}\mu, \mathbb{P}_{f_{\theta_0}} - \text{a.s.}.$$

To do that, we make use of classical tools of large deviation (see [33]). We compute the Laplace transform of $X_n^T \Lambda_n X_n$:

$$\begin{aligned}
\mathbb{E}_{\mathbb{P}_{f_{\theta_0}}} & \left[e^{\lambda X_n^T \mathcal{K}_n(\frac{1}{f_\theta}) X_n} \right] \\
= & \frac{1}{(\sqrt{2\pi})^{m_n} \sqrt{\det(\mathcal{K}_n(f_{\theta_0}))}} \int e^{\frac{1}{2} X_n^T \left(\left(\mathcal{K}_n(f_{\theta_0})\right)^{-1} - 2\lambda \mathcal{K}_n(\frac{1}{f_\theta}) \right) X_n} \\
= & \frac{1}{\sqrt{\det(\mathcal{K}_n(f_{\theta_0}))}} \sqrt{\det\left(\left[\left(\mathcal{K}_n(f_{\theta_0})\right)^{-1} - 2\lambda \mathcal{K}_n(\frac{1}{f_\theta}) \right]^{-1} \right)} \\
= & \frac{1}{\sqrt{\det\left(I_{G_n} - 2\lambda \mathcal{K}_n(f_{\theta_0})^{\frac{1}{2}} \mathcal{K}_n(\frac{1}{f_\theta}) \mathcal{K}_n(f_{\theta_0})^{\frac{1}{2}} \right)}}.
\end{aligned}$$

These last equalities hold as soon as $I_{G_n} - 2\lambda \mathcal{K}_n(f_{\theta_0})^{\frac{1}{2}} \mathcal{K}_n(\frac{1}{f_\theta}) \mathcal{K}_n(f_{\theta_0})^{\frac{1}{2}}$ is positive. This is true whenever $\lambda \leq 0$ or small enough.
Now, for $\lambda \leq 0$, define

$$\phi_n(\lambda) := \frac{1}{m_n} \log \left(\mathbb{E}_{\mathbb{P}_{f_{\theta_0}}} \left[e^{\lambda X_n^T \mathcal{K}_n(\frac{1}{f_\theta}) X_n} \right] \right),$$

This function verifies

$$\phi_n(\lambda) = -\frac{1}{2m_n} \log\det \left(I_{G_n} - 2\lambda \mathcal{K}_n(f_{\theta_0})^{\frac{1}{2}} \mathcal{K}_n(\frac{1}{f_\theta}) \mathcal{K}_n(f_{\theta_0})^{\frac{1}{2}} \right).$$

Define also

$$\phi(\lambda) = \lim_n \phi_n(\lambda),$$

We get, using Corollary 3.6.2,

$$\phi(\lambda) = -\frac{1}{2} \int \log \left(1 - 2\lambda \frac{f_{\theta_0}}{f_\theta} \right).$$

We can also compute

$$\phi''(\lambda) = \int \frac{2(\frac{f_{\theta_0}}{f_\theta})^2}{(1 - 2\lambda \frac{f_{\theta_0}}{f_\theta})^2} \mathrm{d}\mu > 0.$$

As usually, we define the convex conjugate of ϕ by

$$\phi^*(t) := \sup_{\lambda \in \mathbb{R}^-} \left[\lambda t - \phi(\lambda) \right], t \in \mathbb{R}.$$

As soon as ϕ is strictly convex, $\phi^*(t) > \phi(0) = 0$, for any $t \neq \phi'(0) = \int \frac{f}{g} \mathrm{d}\mu$.
We can now write, for $\lambda \leq 0$,

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

Then we get, $\forall t > \int \frac{f_{\theta_0}}{f_\theta} \mathrm{d}\mu$,

$$\limsup_n \left(\frac{1}{m_n} \log(\mathbb{P}(\frac{1}{m_n} X_n^T \Lambda_n X_n \geq t))\right) \leq -\lambda t + \phi(\lambda)$$

So that, taking the infimum on λ, we get

$$\limsup_n \left(\frac{1}{m_n} \log(\mathbb{P}(\frac{1}{m_n} X_n^T \Lambda_n X_n \geq t))\right) \leq -\phi^*(t) < 0$$

We can obtain the same bound for $t < \int f_{\theta_0} f_\theta \mathrm{d}\mu$. By Borel-Cantelli theorem, we get the $\mathbb{P}_{f_{\theta_0}}$-almost sure convergence of $\frac{1}{m_n} X_n^T \Lambda_n X_n$ to $\int f_{\theta_0} f_\theta \mathrm{d}\mu$. To prove the same convergence with $\Lambda_n = (\mathcal{K}_n(f_\theta))^{-1}$, we have to show that the difference between the spectral empirical measure of $\mathcal{K}_n(f_{\theta_0})^{\frac{1}{2}} \mathcal{K}_n(\frac{1}{f_\theta}) \mathcal{K}_n(f_{\theta_0})^{\frac{1}{2}}$ and $\mathcal{K}_n(f_{\theta_0})^{\frac{1}{2}} \mathcal{K}_n(f_\theta)^{-1} \mathcal{K}_n(f_{\theta_0})^{\frac{1}{2}}$ converges weakly to zero. It is sufficient to control the convergence of every moment, because these two last measures both have compact support.
For this, we make use of the Schatten norms. For any A, B matrices of $M_{m_n}(\mathbb{R})$, we define

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

where $s_k(A)$ are the singular values of A.
Note that

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

Recall that since $f_\theta \in \mathcal{F}_\rho$, we have $e^{-\rho} \leq f_\theta \leq e^\rho$. Hence, for any $p \geq 1$,

$$\begin{aligned}&\frac{1}{m_n} \left|\mathrm{Tr}\left(\mathcal{K}_n^p(\frac{1}{f_\theta}) \mathcal{K}_n^p(f_{\theta_0}) - \mathcal{K}_n^{-p}(f_\theta) \mathcal{K}_n^p(f_{\theta_0})\right)\right| \\ &\qquad \leq \frac{1}{m_n} \left\|\mathcal{K}_n(f_\theta)^{-p} \mathcal{K}_n^p(f_{\theta_0})\right\|_{Sch,\infty} \left\|\left(\mathcal{K}_n^p(\frac{1}{\theta}) \mathcal{K}_n^p(f_\theta) - I_{G_n}\right)\right\|_{Sch,1} \\ &\qquad \leq \frac{\delta_n}{m_n} \frac{e^{2\rho p}}{e^{-2\rho p}} \alpha(f_\theta)^{2p} \alpha(\frac{1}{f_\theta})^{2p} \underset{n \to \infty}{\to} 0.\end{aligned}$$

To obtain the same bound with $\Lambda_n = \mathcal{Q}_n(\frac{1}{f_\theta})$, we have to prove that the difference between the spectral empirical measures of $\mathcal{K}_n(f_{\theta_0})^{\frac{1}{2}}\mathcal{K}_n(\frac{1}{f_\theta})\mathcal{K}_n(f_{\theta_0})^{\frac{1}{2}}$ and $\mathcal{K}_n(f_{\theta_0})^{\frac{1}{2}}\mathcal{Q}_n(\frac{1}{f_\theta})\mathcal{K}_n(f_{\theta_0})^{\frac{1}{2}}$ converge weakly to zero. This last assertion is a direct consequence of Lemma 3.6.3. So, we get

$$\frac{1}{m_n}X_n^T\Lambda_n X_n \to \int \frac{f_{\theta_0}}{f_\theta}, \mathbb{P}_{f_{\theta_0}} - \text{ a.s.}$$

■

Proof. of Lemma 3.6.6
Recall that we aim at proving that, $\mathbb{P}_{f_{\theta_0}}$-a.s., the sequence of functions

$$(l_n(f_{\theta_0}, X_n) - l_n(f_\theta, X_n))_{n\in\mathbb{N}}$$

is equicontinuous on $\{f_\theta, \theta \in \Theta\}$, and that this property also holds for $\bar{l_n}$,$\tilde{l}_n$ and $l_n^{(u)}$.
First, we will prove the equicontinuity of the sequence

$$\left(\frac{1}{m_n}\log\det(\mathcal{K}_n(f_\theta))\right)_{n\in\mathbb{N}}.$$

Let $\theta, \theta' \in \Theta$.
Denote λ_i the eigenvalues of $\mathcal{K}_n(f_{\theta'})^{-1}\left(\mathcal{K}_n(f_{\theta'}) - \mathcal{K}_n(f_\theta)\right)$. Since $f_\theta \in \mathcal{F}_\rho$, we have $e^{-\rho} \le f_\theta \le e^\rho$.
Notice that we have

$$\begin{aligned}\sup_{i=1,\cdots,n}|\lambda_i| &= \left\|\mathcal{K}_n(f_{\theta'})^{-1}\left(\mathcal{K}_n(f_{\theta'}) - \mathcal{K}_n(f_\theta)\right)\right\|_{2,op} \\ &\le e^\rho \left\|f_{\theta'} - f_\theta\right\|_\infty.\end{aligned}$$

So that, to prove the equicontinuity, we may assume that θ is close enough to θ' to ensure that $\sup_{i=1,\cdots,n}|\lambda_i| \le \frac{1}{2}$.
We have

$$\begin{aligned}&\frac{1}{m_n}\left|\log\det(\mathcal{K}_n(f_{\theta'})) - \log\det(\mathcal{K}_n(f_\theta))\right| \\ &\qquad = \frac{1}{m_n}\left|\log\det\left(I_{G_n} - \mathcal{K}_n(f_{\theta_0})^{-1}\left(\mathcal{K}_n(f_{\theta'}) - \mathcal{K}_n(f_\theta)\right)\right)\right| \\ &\qquad \le \frac{1}{m_n}\sum_{i\in G_n}|\log(1+\lambda_i)| \\ &\qquad \le \sup_{i\in G_n}|\log(1+\lambda_i)| \\ &\qquad \le \log(2)\sup_{i\in G_n}|\lambda_i| \\ &\qquad \le \log(2)e^\rho\left\|f_{\theta'} - f_\theta\right\|_\infty.\end{aligned}$$

Furthermore, the sequence $(\int \log(f_\theta)\mathrm{d}\mu)_{n\in\mathbb{N}}$ is also equicontinuous since, using a Taylor formula,

$$\int \left|\log(f_{\theta'})\mathrm{d}\mu - \int \log(f_\theta)\mathrm{d}\mu\right| \leq e^{\rho} \left\|f_{\theta'} - f_\theta\right\|_\infty .$$

Now we tackle the equicontinuity of the sequences

$$\left(X_n^T \mathcal{K}_n(f_\theta)^{-1} X_n\right)_{n\in\mathbb{N}},$$

$$\left(X_n^T \mathcal{K}_n(\frac{1}{f_\theta}) X_n\right)_{n\in\mathbb{N}}$$

and

$$\left(X_n^T \mathcal{Q}_n(\frac{1}{f_\theta}) X_n\right)_{n\in\mathbb{N}}.$$

Notice first that, for any matrix $B \in M_n(\mathbb{R})$,

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

It is thus sufficient to prove the equicontinuity of the sequences

$$(\mathcal{K}_n(f_\theta)^{-1})_{n\in\mathbb{N}},$$

$$(\mathcal{K}_n(\frac{1}{f_\theta}))_{n\in\mathbb{N}}$$

and

$$(\mathcal{Q}_n(f_\theta)^{-1})_{n\in\mathbb{N}},$$

for the norm $\|.\|_{2,op}$
Note that

$$\begin{aligned}\left\|\mathcal{K}_n(\frac{1}{f_{\theta'}}) - \mathcal{K}_n(\frac{1}{f_\theta})\right\|_{2,op} &\leq \left|\frac{1}{f_{\theta'}} - \frac{1}{f_\theta}\right|_\infty \\ &\leq e^{2\rho}\left\|f_{\theta'} - f_\theta\right\|_\infty .\end{aligned}$$

Then,

$$\begin{aligned}\left\|(\mathcal{K}_n(f_{\theta'}))^{-1} - (\mathcal{K}_n(f_\theta))^{-1}\right\|_{2,op} &\leq \left\|(\mathcal{K}_n(f_{\theta'}))^{-1}(\mathcal{K}_n(f_\theta))^{-1}\right\|_{2,op}\left\|(\mathcal{K}_n(f_{\theta'})) - (\mathcal{K}_n(f_\theta))\right\|_{2,op} \\ &\leq e^{2\rho}\left\|f_{\theta'} - f_\theta\right\|_\infty .\end{aligned}$$

Then, recall that, for any symmetric matrix $B \in M_n(\mathbb{R})$, we have

$$\left\|B\right\|_{2,op} \leq \left\|B\right\|_{\infty,op}.$$

Recall also that $\mathcal{Q}_n(f_\theta) = B^{(n)} \odot \mathcal{K}_n(f_\theta)$. Denote

$$\begin{aligned}
\left\| \mathcal{Q}_n(\frac{1}{f_{\theta'}}) - \mathcal{Q}_n(\frac{1}{f_\theta}) \right\|_{2,op} &\leq \left\| \mathcal{Q}_n(\frac{1}{f_{\theta'}}) - \mathcal{Q}_n(\frac{1}{f_\theta}) \right\|_{\infty,op} \\
&\leq \sup_{i,j=1,\cdots n} \left| B_{ij}^{(n)} \right| \left\| \mathcal{K}_n(\frac{1}{f_{\theta'}}) - \mathcal{K}_n(\frac{1}{f_\theta}) \right\|_{\infty,op} \\
&\leq (1+u_n) \left\| \frac{1}{f_{\theta'}} - \frac{1}{f_\theta} \right\|_{1,pol} \text{ (see Assumption 3.3.5).}
\end{aligned}$$

Since the map $f_\theta \mapsto \frac{1}{f_\theta}$ is continuous over $\mathcal{F}_\rho$, which is compact, we get the uniform equicontinuity of the map $f_\theta \mapsto X_n^T \mathcal{Q}_n(\frac{1}{f_\theta}) X_n$ (for the norm $\|.\|_{1,pol}$).
This concludes the proof of Lemma 3.6.6 ∎

Proof. of Lemma 3.3.11
We aim at proving the asymptotic normality of $\sqrt{m_n}(l_n^{(u)})'(\theta_0)$.
Using the Fourier transform, it is sufficient to prove that

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

Recall that we have

$$(l_n^{(u)})'(\theta) = -\frac{1}{2}\int \frac{f'_\theta}{f_\theta}\mathrm{d}\mu + \frac{1}{2m_n} X_n^T \mathcal{Q}_n(\frac{f'_\theta}{f_\theta^2}) X_n.$$

We can compute

$$\begin{aligned}
\sqrt{m_n}\mathbb{E}\left[(l_n^{(u)})'(\theta_0)\right] &= \sqrt{m_n}\left(-\frac{1}{2}\int \frac{f'_{\theta_0}}{f_{\theta_0}}\mathrm{d}\mu + \frac{1}{2m_n}\operatorname{Tr}\left(\mathcal{K}_n(f_{\theta_0})\mathcal{Q}_n(\frac{f'_\theta}{f_\theta^2})\right)\right) \\
&= \sqrt{m_n}\left(-\frac{1}{2}\int \frac{f'_{\theta_0}}{f_{\theta_0}}\mathrm{d}\mu + \frac{1}{2m_n}\operatorname{Tr}\left(\mathcal{K}_n\left(f_{\theta_0}\frac{f'_{\theta_0}}{f_{\theta_0}^2}\right)\right)\right) \text{ (see Lemma 3.6.4)} \\
&\leq C v_n \sqrt{m_n} \underset{n\to\infty}{\to} 0 \text{ (see Assumption 3.3.6).}
\end{aligned}$$

If we define

$$Z_n = t\frac{1}{2m_n} X^T \mathcal{Q}_n(\frac{f'_\theta}{f_\theta^2}) X,$$

and

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

the last equality means that

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

This holds only if f_{θ_0} is a polynomial, or if all the $f_\theta, \theta \in \Theta$ are polynomials. This brings out that the second theorem holds for the AR_P or MA_P case. It also explains the term 'unbiased estimator' used for $\theta^{(u)}$.
Then, it is sufficient to show

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

If τ_k denotes the eigenvalues of the symmetric matrix

$$M_n := \frac{t}{2}\mathcal{K}_n(f_{\theta_0})^{\frac{1}{2}}\mathcal{Q}_n(\frac{f'_{\theta_0}}{f^2_{\theta_0}})\mathcal{K}_n(f_{\theta_0})^{\frac{1}{2}},$$

then we can write

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

where $(Y_k)_{k\in G_n}$ has the standard Gaussian distribution on $\mathbb{R}^{m_n}$.
The independence of Y_k leads to

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

The τ_k are bounded, thanks to the following inequality :

$$\begin{aligned}
\|M_n\|_{2,op} &= \left\|\frac{t}{2}\mathcal{K}_n(f_{\theta_0})^{\frac{1}{2}}\mathcal{Q}_n(\frac{f'_{\theta_0}}{f^2_{\theta_0}})\mathcal{K}_n(f_{\theta_0})^{\frac{1}{2}}\right\|_{2,op} \\
&\leq \left\|\frac{t}{2}\mathcal{K}_n(f_{\theta_0})^{\frac{1}{2}}\right\|_{2,op}\left\|\mathcal{Q}_n(\frac{f'_{\theta_0}}{f^2_{\theta_0}})\right\|_{2,op}\left\|\mathcal{K}_n(f_{\theta_0})^{\frac{1}{2}}\right\|_{2,op} \\
&\leq \left\|\frac{t}{2}\mathcal{K}_n(f_{\theta_0})^{\frac{1}{2}}\right\|_{2,op}\left\|\mathcal{Q}_n(\frac{f'_{\theta_0}}{f^2_{\theta_0}})\right\|_{1,op}\left\|\mathcal{K}_n(f_{\theta_0})^{\frac{1}{2}}\right\|_{2,op} \\
&\leq e^{\rho}\alpha(f'_{\theta_0})\alpha(f_{\theta_0})^2(1+u_n).
\end{aligned}$$

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

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

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 will be proved if we show that

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

This last convergence is a consequence of Lemmas 3.6.1 and 3.6.3.
This provides the asymptotic normality of $\sqrt{m_n}(l_n^{(u)})'(\theta_0)$ and concludes the proof of Lemma 3.3.11 :

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

■

Proof. of Lemma 3.3.12
We aim now at proving the $P_{f_{\theta_0}}$-a.s. following convergence :

$$\left((l_n^{(u)})''(\breve{\theta}_n)\right)^{-1} \underset{n\to\infty}{\to} \frac{1}{2}\left(\int \frac{(f'_{\theta_0})^2}{f_{\theta_0}^2}\mathrm{d}\mu\right)^{-1}$$

We have

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

which leads to

$$(l_n^{(u)})''(\theta) \underset{n\to\infty}{\to} \frac{1}{2}\int \left(\frac{f''_\theta f_\theta - (f'_\theta)^2}{f_\theta^2} + \frac{f_{\theta_0}\left(2(f'_\theta)^2 - f''_\theta f_\theta\right)}{f_\theta^3}\right)\mathrm{d}\mu, P_{f_{\theta_0}}\text{-a.s.}$$

Since the sequence $l_n^{(u)}$ is equicontinuous and $\breve{\theta}_n \underset{n\to\infty}{\to} \theta_0$, we obtain the desired convergence :

$$(l_n^{(u)})''(\breve{\theta}_n) \underset{n\to\infty}{\to} \frac{1}{2}\int \left(\frac{(f'_{\theta_0})^2}{f_{\theta_0}^2}\right)\mathrm{d}\mu, P_{f_{\theta_0}}\text{-a.s.}$$

■

Proof. of Lemma 3.3.13
We want to compute the asymptotic Fisher information. As usual, it is sufficient to compute

$$\frac{1}{m_n}\mathrm{Var}\left(L'_n(\theta_0)\right) = \lim_n \frac{1}{2m_n}\mathrm{Tr}(M_n(\theta_0)^2),$$

where $M_n(\theta) = \mathcal{K}_n(f_\theta)^{-1}\mathcal{K}_n(f'_\theta)\mathcal{K}_n(f_\theta)^{-1}\mathcal{K}_n(f_{\theta_0})$.
This leads, together with Lemma 3.6.1, and Assumption 3.3.3 to

$$\frac{1}{m_n}\mathrm{Var}\left(L'_n(\theta_0)\right) \to \frac{1}{2}\int \frac{(f'_{\theta_0})^2}{f_{\theta_0}^2}\mathrm{d}\mu.$$

This ends the proof of the last lemma. ■

3.7 Extensions

In the present chapter, we provided tools to study $ARMA$ processes on graphs. Such processes have a very specific covariance structure. Indeed, as we built the covariance as a power series of the adjacency operators, every edges play the same role.
Therefore, $ARMA$ models may be considered as too much restrictive. For instance, on $\mathbb{Z}^d, d > 1$, any stationary process (even isotropic) is not necessary an $ARMA$ process. In this section, we aim at giving a more general framework, studying larger class of processes. However, we would like that, in the usual case when there exists an universal notion of stationary processes ($\mathbb{Z}^d$, the homogeneous tree...) we recover these notions with the definitions proposed here.
Note that we could have generated $ARMA$ covariance with other operator than the adjacency one. For instance, we could have chosen the discrete Laplacian L (see for instance [26]) of the graph $\mathbf{G}$ and made all the previous work with covariance operators of the form

$$\Gamma = f(L).$$

This underlines a limit of the previous work : the generator has been chosen as W, but this choice is not necessarily the best. On the contrary, rather than W, the generator should come from physical real-life models. The first goal of this section is therefore to take into consideration this remark.
Another motivation for this work comes from graphical models. Indeed, as explained in the previous sections, our motivation was far away from such issues. But the only Markov processes we can obtain with $ARMA$ models are AR_1. Indeed, the inverse of the covariance operator has to vanish out of the edges. On the other hand, graphical models are also restrictive since, in many cases, the underlying graph is nothing else than a mere model for the correlation.
Actually, graphical models may be always seen as AR_1 models, built with other operators than the adjacency one. For instance, we can choose the inverse of the covariance operator Γ as the generator, and the function $\frac{1}{x}$ as the density. This trivial remark leads to an important question :
Why do we restrict ourselves to order 1 ?
More precisely, in many situations, the Markov assumption is a model. It certainly provides conditional independency of variables X_i, X_j indexed by vertices $i, j \in G$ which are not neighbors (with respect to all the other variables). But in some real cases, where the underlying graph exists (for instance social networks), the meaning of the Markov hypothesis is not obvious.
Here we will present the beginning of a work, established with the goal of providing a link between $ARMA$ models and graphical models. Consider that we observe a process $(X_i)_{i \in G}$. Many interesting statistical questions arise :

– The graph $\mathbf{G}$ may be unknown. We have to estimate it. To this aim, we need some assumptions on the correlation structure of the process. This is classically achieved by considering that the process is Markovian. This is the natural framework of graphical models, and a lot of work has been done in this direction for many years (for instance see [64]).
– Assume now that the graph is known, and that the process is still Markovian. Another natural question which arises is the estimation of the covariance structure of the process. This question is still close to graphical models issues. But now, instead of modeling the process as a Markov process, we can extend this to $ARMA$ models (not necessary generated by the adjacency operator), and try to estimate the generator. This will not be done here, but it is clearly one of the main perspective for the future works.
– Assume now that the generator is also known. We aim at estimating the covariance operator of the process. This work has been done in the last section, when the generator is the adjacent operator W. In this section, we extend the last work to other generators. Indeed, only a few assumptions on W were needed. This extension to other generators explains the term 'admissible modification of the graph' which appears in the following.

Finally, this extension also leads to a very formal notion of stationarity that does not use the invariance by automorphisms. We underline here (without proofs) that this notion is not exactly equivalent to the usual one (invariance of the covariance operators by automorphisms), because it changes, in some way, the condition of isomorphic to isospectral.

So this section is devoted to modify the weights in an "isotropic" way, and locally, for finite order operators (see Definition 3.7.2). Now, let us give a sense to this "isotropic" modification.

For sake of clarity, each time we will define a quantity of interest we will illustrate it for the case $\mathbf{G} = \mathbb{Z}$.

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 graphs may be set in the same way [42], [5], [45]. Unfortunately, as the generic situation for a graph is to a have a trivial set of automorphisms (reduced to the identity operator), this way to define stationarity is a dead-end. Indeed, in this approach any covariance operator would be station-

ary. 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 all bounded Hilbertian operators on $l^2(G)$. We denote also by Σ_G the set of permutation of G. Let us define the operator M_σ attached to the permutation $\sigma \in \Sigma_G$ by

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

Let F_G be the set of all continuous linear operators from $l^1(G)$ to $l^\infty(G)$. We first define the class of invariant functions.

Definition 3.7.1 *We call invariant a function*

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

that satisfies the following assumptions :
– $\mathrm{Dom}(\Phi)$ *is stable by any permutations and by transposition*

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

– *For any permutation $\sigma \in \Sigma_G$, Φ commutes with the conjugation by M_σ*

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

– Φ *commutes with the transposition*

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

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

Remark Notice that any invariant function $\Phi \in \mathcal{I}_G$ is given by a family of functions Φ_{ij} all defined on a subset $\mathrm{Dom}(\Phi)$ of B_G. Let $1_G, 2_G \in G$ be two vertices of **G**. Actually, thanks to the invariance, the functions $(\Phi_{ij})_{i,j \in G}$ are completely determined by their domain $\mathrm{Dom}(\Phi)$, and two real-valued functions $\phi := \Phi_{1_G 1_G}$ and $\psi := \Phi_{1_G 2_G}$ from B_G to $\mathbb{R}$ such that
– For any $k \in G$, for all $A \in \mathrm{Dom}(\Phi)$, and for any permutation $\sigma \in \Sigma_G$ such that $\sigma(1_G) = k$,

$$\Phi_{kk}(A) = \phi(M_\sigma^{-1} A M_\sigma).$$

– For any $k, l \in G, k \neq l$, for all $A \in \mathrm{Dom}(\Phi)$, and for any permutation $\sigma \in \Sigma_G$ such that $\sigma(1_G) = k, \sigma(2_G) = l$,

$$\Phi_{kl}(A) = \psi(M_\sigma^{-1} A M_\sigma).$$

– For all $A \in \mathrm{Dom}(\Phi)$, and for any permutation $\sigma \in \Sigma_G$ permuting 1_G and 2_G

$$\psi(A) = \psi(M_\sigma^{-1} A^T M_\sigma).$$

Note that we also get

– For any $k \in G$, for all $A \in \mathrm{Dom}(\Phi)$, and for any permutation $\sigma \in \Sigma_G$ letting 1_G invariant

$$\phi(A) = \phi(M_\sigma^{-1} A M_\sigma).$$

– For any $k \in G$, for all $A \in \mathrm{Dom}(\Phi)$, and for any permutation $\sigma \in \Sigma_G$ letting 1_G and 2_G invariant,

$$\psi(A) = \psi(M_\sigma^{-1} A M_\sigma).$$

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

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

Set, for $A \in B_G$,

$$\phi_L(A) := \frac{1}{2} \left(\sum_{k \in G} A_{k 1_G} + \sum_{k \in G} A_{1_G k} \right),$$

and

$$\psi_L(A) := -A_{1_G 2_G},$$

where ϕ and ψ are defined in the last remark. We get

$$\Phi_L(A)_{ij} = \mathbb{1}_{(i=j)} \frac{1}{2} \left(\sum_{k \in G} A_{ki} + \sum_{k \in G} A_{ik} \right) - A_{ij} = L_{ij}^{(A)}.$$

The domain of Φ_L is the set of the operators $A \in B_G$ such that, for any $i \in G$, the sequences $(A_{ik})_{k \in G}$ and $(A_{ki})_{k \in G}$ are summable. This domain is stable by permutations and transposition. Moreover, we can verify the invariance property, writing, for any permutation $\sigma \in \Sigma_G$,

$$\begin{aligned}
\Phi_L(A)_{\sigma(i)\sigma(j)} &= \mathbb{1}_{(\sigma(i)=\sigma(j))} \frac{1}{2} \left(\sum_{k \in G} A_{k\sigma(i)} + \sum_{k \in G} A_{\sigma(i)k} \right) - A_{\sigma(i)\sigma(j)} \\
&= \mathbb{1}_{i=j} \frac{1}{2} \left(\sum_{\sigma(k) \in G} A_{\sigma(k)\sigma(i)} + \sum_{\sigma(k) \in G} A_{\sigma(i)\sigma(k)} \right) - A_{\sigma(i)\sigma(j)} \\
&= \Phi_L(M_\sigma^{-1} A M_\sigma)_{ij},
\end{aligned}$$

and

$$\begin{aligned}\forall \sigma \in \Sigma_G, \Phi_L(A^T)_{ij} &= \mathbb{1}_{i=j}\frac{1}{2}\left(\sum_{k\in G} A_{ik} + \sum_{k\in G} A_{ki}\right) - A_{ji} \\ &= \mathbb{1}_{j=i}\sum_{k\in G}\frac{1}{2}\left(\sum_{k\in G} A_{jk} + \sum_{k\in G} A_{kj}\right) - A_{ji} \\ &= \Phi_L(M_\sigma^{-1} A M_\sigma)_{ij}.\end{aligned}$$

Then, this shows that

$$\Phi_L \in \mathcal{I}_{\mathcal{G}}.$$

We go back to the general construction. One can have the intuition that the functions $\Phi_{ii}, i \in G$ associated to an invariant function Φ contains information on the exploration of the graph from the vertex $i \in G$. This leads to the definition of the order of an invariant function. This order may be either finite or infinite and roughly speaking depends on the exploration size. Recall that the natural distance $d_{\mathbf{G}}$ on $\mathbf{G}$ measure the length of the shortest path going a vertex to another. This distance depends only on the edges, and not on the weights.
For any $r > 0$, let us define $B^{(W)}(i, r)$ as the ball of radius r (for the natural distance $d_{\mathbf{G}}$) centered on a vertex $i \in G$:

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

We now precisely define the order of an invariant function.

Definition 3.7.2 *Let Φ be an invariant function in $\mathcal{I}_G$, we will say that Φ is r-local for some $r \geq 0$ if for any $W \in \mathrm{Dom}(\Phi)$, $\phi(W)$ depends only on $\left(W_{jk}, j, k \in B^{(W)}(1_G, r)\right)$, and $\psi(W)$ depends only on*

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

where the functions ϕ and ψ have been defined in the previous Remark. 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 an important notion. Indeed, Definition 3.7.1 builds a large class of invariant functions. This class of functions will lead to a class of covariance operators of stationary processes (see Definition 3.7.3). Therefore, this last class of stationary covariance operators will be very large too. Defining the order of an invariant function gives a way 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 lies in this set (as soon

as the variance is bounded), 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 isotropic stationary Gaussian processes to any graph.

Definition 3.7.3 *We say that a Gaussian process $(X_i)_{i\in G}$ is stationary of order r, if its covariance operator Γ verifies*

$$\Gamma \in \mathcal{S}_+^r(W) := \begin{Bmatrix} \Gamma = \Phi(W), \Phi \in \mathcal{I}_G, \Phi \text{ has order } r \\ \Gamma \text{ is positive definite} \end{Bmatrix}$$

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

$$W' \in \mathcal{S}_1^r(W) := \begin{Bmatrix} W' = \Phi(W), \Phi \in \mathcal{I}_G, \Phi \text{ has order } r \\ \Big(W_{ij} = 0 \Rightarrow W'_{ij} = 0\Big) \end{Bmatrix}$$

Remark A stationary process of finite order $r > 0$ has the following property.
Let $(i_1, j_1), (i_2, j_2) \in G^2$ be two couples of vertices such that $B^{(W^{(G)})}(i_1, r)$ is isomorphic to $B^{(W^{(G)})}(i_2, r)$ and $B^{(W^{(G)})}(j_1, r)$ is isomorphic to $B^{(W^{(G)})}(j_2, r)$. Then for any stationary covariance operator Γ of order less or equal to r,

$$\Gamma_{i_1 j_1} = \Gamma_{i_2 j_2}.$$

That means that, if the graph is locally isomorphic in two different regions, then the correlations will be also identical in these areas.

To close this section, let us notice that if the entries of W takes their values in a finite set, then the entries of any admissible modification of finite order $r > 0$ take also their values in a finite set, since there is a finite number of possible subgraphs of size bounded by $\deg(\mathbf{G})^r$.

Remark All the previous work may be applied to a admissible modification W' of the graph instead of W.
In very particular, any AR_1 model built with an isotropic modification of $\mathbf{G}$ provides a graphical model [32]. That was one of the main motivation for this section. Indeed the definition of stationarity we propose here leads, on one hand, to usual stationary processes when it is well defined ($\mathbb{Z}^d$, the homogeneous tree...).
On the other hand, for a graph with a trivial set of automorphism, we recover many classical kind of modeling as $ARMA$ processes and Markov fields. This falls into a very classical point of view on Gaussian fields, and allow us to used all general tools developed in this framework for our case.

To end this section, let us deal with an example with the usual case of $G = \mathbb{Z}$. In this case, the last construction recovers any covariance operator Γ, even non regular. Indeed, set

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

and define, for any $A \in B_{\mathbb{Z}}$ such that the following is well defined, and for any $p \in \mathbb{N}, i, j \in \mathbb{Z}$,

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

We can define $\Phi \in \mathcal{I}_G$, for any A such it is well defined, as

$$\Phi(A) = \sum_{p \geq 0} r_p \Phi^{(p)}(A).$$

Notice that $W^{(\mathbb{Z})} \in \mathrm{Dom}(\Phi)$. Thus we get that

$$\Gamma = \Phi(W^{(\mathbb{Z})}),$$

which proves the statement.

Remark Back to the general case, if Γ is the covariance of an $ARMA$ process over the graph $\mathbf{G}$, built with any admissible modification W' of the graph, then $\Gamma \in \mathcal{S}_+^\infty$, so $ARMA$ processes are always stationary processes with respect to Definition 3.7.3.
In the framework of road traffic modeling, as in other modeling issues, W' may be seen as a structural generating operator. That means that the physical evolution of the underlying temporal process (diffusion...) is given by this operator. The parameters of the $ARMA$ process gives the particular state of the field at a given time.

Chapitre 4

Spatio-temporal anisotropic case

Dans ce chapitre, nous étendons la construction précédente au cas spatio-temporel, en vue d'applications avec des données réelles. Nous prouvons le Lemme de Szegö pour le cas spatio-temporel. Les démonstrations sont très proches de celle du cas uniquement spatial (à temps fixé).

Introduction

In this chapter, we extend the construction of $ARMA$ processes indexed by graphs and Whittle-type estimation of the parameters of the *spectral density* to the spatio-temporal case. Actually, we perform this extension in view of applications of our work on real datas.
In the framework of road traffic, the vertices are the locations of speed sensors on stretches of roads. This vertices are neighbors as soon as the corresponding stretches of roads are connected by a node.
We aim to use the physical characteristics of this network to specify a covariance structure for the data. The global street network will be modeled as an infinite graph $\mathbf{G}_\infty$.
In this Chapter, we will assume that the speed process is only observed in a few locations. That is, in a subset G of the vertices set G_∞, and at many times. More precisely, we model the road traffic speed process as a Gaussian spatio-temporal process $\mathbf{X} := (X_{i,t})_{i\in G_\infty, t\in\mathbb{Z}}$ (the time here is discrete). Indeed, we aim at filtering some missing or corrupted values. For this, we will use kriging methods (see for instance [60]) for processes indexed by graphs.
In the most general frame, our observation is $(X_{i,t})_{(i,t)\in O}$, where O is a subset of $G_\infty \times \mathbb{Z}$.
From a theoretical point of view, the set of observations should be a growing sequence of subgraphs of $G_\infty \times \mathbb{Z}$. In practical situations, one has only to chose O

large enough.
The process $\mathbf{X}$ will be assumed to have zero-mean. Indeed, our data have already been centered with regression methods. In this work, we are interested essentially in the correlations. Indeed, our aim is twofold :
– On the one hand, we chose a fixed time t_0, and deal with a spatial field. In this case, the observation set is $O = G \times \{t_0\}$, where $G \subset G_\infty$. In this case, we can forget the temporal dimension, and we provide a way to perform spatial filtering for the Gaussian field.
– On the other hand, we generalize this construction, to spatio-temporal processes, and build a generalized spatio-temporal predictor.
In both cases, we need to learn the covariance structure of the process from the data. This is performed using a parametric model and the estimators built in Chapter 3. This work has been established in view of application to road traffic problems.
This parametric estimation relies on a Whittle's approximation (see for instance [67], [68]...) of the maximum likelihood method. Our construction requires the definition of the spectral measure of a graph. Then, as in the classical setting of time series, we plug it into the Maximum Likelihood Estimator.

4.1 Notations and theoretical background

In this section, we give the theoretical background required to understand the estimation procedure developed in the paper. More precisely, we recall the general framework of a Gaussian field indexed by a graph (here the time is fixed). Further, we will extend the tools to the spatio-temporal case.

Gaussian field

Let $\mathbf{G}_\infty = (G_\infty, W^{G_\infty})$ be a countable graph. That is,
– G_∞ is the set of vertices (modeled as infinite, but countable).
– $W^{G_\infty} \in [-1, 1]^{G_\infty \times G_\infty}$ is the symmetric weighted adjacency operator.
For any vertex $i \in G_\infty$, a vertex $j \in G_\infty$ is said to be a neighbor of i if, and only if, $W_{ij}^{G_\infty} \neq 0$. The degree $\deg(i)$ of i is the number of neighbors of the vertex i, and the degree of the graph $\mathbf{G}_\infty$ is defined as the maximum degree of the vertices of the graph $\mathbf{G}_\infty$:

$$\deg(\mathbf{G}_\infty) := \max_{i \in G_\infty} \deg(i).$$

From now, we assume that this degree is finite. For the real data, we have $\deg(G_\infty) < 10$. Indeed, there exists no stretch of road linked with more than 10 others.

In the most usual frame, we choose for W^{G_∞} the adjacency operator A^{G_∞}. Recall that the adjacency operator is defined, for $i, j \in G_\infty$, by $A_{ij} = 1$ if i, j are neighbors, and $A_{ij} = 0$ otherwise.
However, we can also modify in an isotropic way this weights (see Chapter 3, Section 3.7) in order to get a better model for the process (since W^{G_∞} will be the generator of the AR spatial field). For instance, for diffusion processes, one can choose for W^{G_∞} the discrete Laplacian (see for instance [26]) instead of the adjacency operator.
Anyway, we renormalize this operator to get

$$\sup_{i,j \in G_\infty} \left| W_{ij}^{G_\infty} \right| \leq 1.$$

Now, consider the action of the adjacency operator W^{G_∞} on $l^2(G_\infty)$ as

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

Denote B_{G_∞} the set of Hilbertian operators on $l^2(G_\infty)$.
The Hilbertian operator W^{G_∞} is continuous with respect to the the classical operator norm $\|.\|_{2,op}$ over B_{G_∞}, defined by

$$\forall A \in B_{G_\infty}, \|A\|_{2,op} := \sup_{u \in l^2(G_\infty), \|u\|_2 \leq 1} \|Au\|_2 ,$$

where $\|.\|_2$ stands for the usual norm on $l^2(G_\infty)$.
Define a growing sequence of nested subgraphs as a sequence $(\mathbf{G}_N)_{N \in \mathbb{N}}, \mathbf{G}_N = (G_N, W^{(G_N)}), N \in \mathbb{N}$ that verifies $G_N \subset G_{N+1} \subset G_\infty$ and, for any $i, j \in G_N$, $W^{(G_N)}(i, j) = W^{G_\infty}(i, j)$.

Spectral measure

The adjacency operator W^{G_∞} is a bounded normal Hilbertian operator. Thus, it admits a spectral decomposition, with respect to an identity resolution $\mathrm{d}E^{G_\infty}$ (see for instance [58]).

$$W^{G_\infty} = \int_{\mathrm{Sp}(W^{G_\infty})} x \mathrm{d}E^{G_\infty}(x).$$

Let, for $i \in G_\infty$, $\delta^{(i)} \in l^2(G_\infty)$ be the sequence defined by

$$\forall j \in G_\infty, \delta_j^{(i)} = \mathbb{1}_{i=j}.$$

We can define the local measure $\mu_{ij}^{G_\infty}$ of G_∞ at the couple of vertices $(i, j) \in G_\infty^2$ as

$$\mu_{ij}^{G_\infty} := \langle \delta^{(i)}, E^{G_\infty} \delta^{(j)} \rangle_{l^2(G_\infty)}.$$

(Here, $\langle ., . \rangle_{l^2(G_\infty)}$ denotes the classical scalar product on $l^2(G_\infty)$.)
Actually, it is the only compact measure on $\mathbb{R}$ that verifies

$$\forall l \geq 0, \left((W^{G_\infty})^l\right)_{ij} = \int_{\mathbb{R}} \lambda^l \mathrm{d}\mu_{ij}^{G_\infty}(\lambda).$$

Let now $(\mathbf{G}_N)_{N\in\mathbb{N}}$ be a growing sequence of nested subgraphs of $\mathbf{G}_\infty$.
The global spectral measure μ^{G_∞} with respect to this sequence is defined, when it exists, as the limit of the mean local measure over this sequence of subgraphs :

$$\mu^{G_\infty} := \lim_{N\to\infty} \frac{1}{\sharp G_n} \sum_{i\in G_n} \mu_{ii}^{G_\infty},$$

for the topology of the weak convergence.
The meaning of this spectral measure is easier to understand through the convergence of the moments :

$$\forall l \geq 0, \frac{1}{\sharp G_N} \left((W^{G_\infty})^l\right) \xrightarrow{N\to\infty} \int_{\mathbb{R}} \lambda^l \mu^{G_\infty}(\lambda).$$

Recall that for any bounded Hilbertian operator $A \in B_G$, the spectrum $\mathrm{Sp}(A)$ is defined as the set of complex numbers λ such that $\lambda \,\mathrm{Id}_G - A$ is not invertible (here Id_G stands for the identity on $l^2(G)$). Since W is bounded and self-adjoint, $\mathrm{Sp}(W)$ is a compact non-empty subset of $\mathbb{R}$ [58].
Actually, the measure μ^{G_∞} is supported by $\mathrm{Sp}(W^{G_\infty})$.

Model

Using the spectral decomposition of the adjacency operator, we may now recall the model for the covariance operators of autoregressive processes.
First, we denote by $\mathcal{D}_\mathcal{K}$ the set of all real functions, continuous over $\mathrm{Sp}(W^{G_\infty})$, whose inverse is a polynomial, with no roots in the convex hull of the spectrum of W^{G_∞}. Then, for any non negative $f \in \mathcal{D}_\mathcal{K}$, we define the covariance operator $\mathcal{K}(f)$ associated to the spectral density f by :

$$\mathcal{K}(f) = \int_{\mathrm{Sp}(W^{G_\infty})} f(x) \mathrm{d}E^{G_\infty}(x).$$

Note that any function $f \in \mathcal{D}_\mathcal{K}$ may be written as a power series :

$$\forall x \in \mathrm{Sp}(W^{G_\infty}), f(x) = \sum_{k\in\mathbb{N}} f_k x^k.$$

Actually, in the spatial case, the operator $\mathcal{K}(f)$ may also be written

$$\mathcal{K}(f) = f(W^{G_\infty}),$$

in the sense of the normal convergence of the corresponding power series :

$$\mathcal{K}(f) = \sum_{k\in\mathbb{N}} f_k (W^{G_\infty})^k.$$

Denote also by $\mathcal{K}_{G_N}(f)$ the restriction of the operator $\mathcal{K}(f)$ to the subgraph G_N.

Szegö Lemma

As in the case of time series, the fundamental properties of the operator $\mathcal{K}$ are given by the Szegö Lemma.
Define the boundary of the graph G_N as

$$\delta(G_N) := \left\{ i \in G_N, \exists j \in G \backslash G_N, W_{ij}^{G_\infty} \neq 0 \right\}.$$

Then, we have, for any $f, g \in \mathcal{D}_{\mathcal{K}}$,

$$\frac{1}{\delta(G_N)} \sup_{i,j\in G_N} \left| \left(\mathcal{K}_{G_N}(f)\mathcal{K}_{G_N}(g) - \mathcal{K}_{G_N}(fg) \right)_{ij} \right| \leq C_{fg},$$

where C_{fg} is given in Chapter 3.
Assume now that

$$\frac{\sharp\delta(G_N)}{\sharp G_N} \underset{N\to\infty}{\to} 0.$$

Assume also that the spectral measure μ^{G_∞} of G_∞ with respect to the sequence $(G_N)_{N\in\mathbb{N}}$ exists. This two assumptions imply that

$$\frac{1}{\sharp G_N} \log\det(\mathcal{K}_{G_N}(f)) \to \int_{\mathrm{Sp}(W^{G_\infty})} \log(f(x)) \mathrm{d}\mu^{G_\infty}(x).$$

Moreover, the Szegö Lemma gives also the following convergence

$$\frac{1}{\sharp G_N} \left\| \mathcal{K}_{G_N}(f)^{-1} - \mathcal{K}_{G_N}(\frac{1}{f}) \right\|_{2,op} \underset{N\to\infty}{\to} 0.$$

This partially justifies the Whittle approximation explained in the next subsection.

Estimation

In this section, we assume that Θ is a compact subset of $\mathbb{R}^d, d \geq 1$, and that we have a parametric family of spectral densities $(f_\theta)_{\theta\in\Theta}$, such that

$$\forall \theta \in \Theta, f_\theta \in \mathcal{D}_{\mathcal{K}}.$$

Then, let $\theta_0 \in \mathring{\Theta}$, and assume that $\mathbf{Y}$ is a centered Gaussian field with covariance $\mathcal{K}(f_{\theta_0})$.

Again, let $(\mathbf{G}_N)_{N\in\mathbb{N}}$ be a growing sequence of nested subgraphs. Assume that the observation is the Gaussian vector $\mathbf{Y}_{G_N}$, that is the restriction of the field $\mathbf{Y}$ to the subgraph G_N (so that, the asymptotic is meant as N goes to infinity). That is,

$$\mathbf{Y}_{G_N} \sim \mathcal{N}\Big(0, \mathcal{K}_{G_N}(f)\Big).$$

Now, we can write the normalized log-likelihood of the model :

$$l_N(\theta, \mathbf{Y}_{G_N}) = -\frac{1}{2}\left(\log(2\pi) + \frac{1}{\sharp G_N}\log\det\Big(\mathcal{K}_{G_N}(f_\theta)\Big) + \frac{1}{\sharp G_N}\mathbf{Y}_{G_N}^T\Big(\mathcal{K}_{G_N}(f_\theta)\Big)^{-1}\mathbf{Y}_{G_N}\right).$$

The Whittle's approximation consists in a modification of this expression, as in the time series case.
We get

$$\tilde{l}_N(\theta, \mathbf{Y}_{G_N}) = -\frac{1}{2}\left(\log(2\pi) + \int \log(f_\theta(x))\mathrm{d}\mu(x) + \frac{1}{\sharp G_N}\mathbf{Y}_{G_N}^T\mathcal{K}_{G_N}\left(\frac{1}{f_\theta}\right)\mathbf{Y}_{G_N}\right).$$

Finally, we estimate θ_0 by maximizing the last expression :

$$\hat{\theta}_N := \arg\max \tilde{l}_N(\theta, \mathbf{Y}_{G_N}).$$

In Chapter 3, we proved the convergence of this estimator.

Prediction

Now, we may consider that the covariance operator is known. In practical situations, this means that the estimation has already been done with an independent sample. Assume that we observe the process $\mathbf{Y}$ at some locations $G_o \subset G_\infty$. We wish to predict some missing values $(Y_i)_{i\in G_m}$ using the observed values $(Y_i)_{i\in G_o}$.
Consider the two Gaussian vectors

$$\mathbf{Y}_{G_m} := (Y_i)_{i\in G_m},$$

$$\mathbf{Y}_{G_o} := (Y_i)_{i\in G_o},$$

The prediction is built by taking the best linear regression of $\mathbf{Y}_{G_m}$ over $\mathbf{Y}_{G_o}$. Actually, this corresponds also to the vector $\hat{Y}_{G_m}$ which minimizes the quadratic form $\mathbf{Y}_G\mathcal{K}_G(f_{\theta_0})^{-1}\mathbf{Y}_G$.

$$\hat{Y}_{G_m} := \arg\min_{Z_m\in\mathbb{R}^{G_m}} \begin{bmatrix} Z_m & Y_{G_o} \end{bmatrix} \mathcal{K}_G(f_{\theta_0})^{-1} \begin{bmatrix} Z_m \\ Y_{G_o} \end{bmatrix}.$$

Indeed, in the Gaussian case, the conditional mode is also the conditional expectation, so that, the best linear predictor.

Note that, in practical situations, we may use here the Szegö Lemma to approximate $\mathcal{K}_G(f_{\theta_0})^{-1}$ by $\mathcal{K}_G(\frac{1}{f_{\theta_0}})$.
This best linear prediction may easily be expressed using the covariance operator. Denote, for any X, Y Gaussian vectors living in the same Gaussian space, $\langle X, Y\rangle$ the covariance operator (or the covariance matrix, depending on the dimension) between X, and Y. That is, in our case,

$$\langle \mathbf{Y}_{G_1}, \mathbf{Y}_{G_2}\rangle := (\mathrm{Cov}(Y_i, Y_j))_{i\in G_1, j\in G_2}\,.$$

The optimal prediction of the missing values $\mathbf{Y}_{G_m}$ using the observed values $\mathbf{Y}_{G_o}$ is given by :

$$\hat{Y}_{G_m} = \langle \mathbf{Y}_{G_m}, \mathbf{Y}_{G_o}\rangle \left(\langle \mathbf{Y}_{G_o}, \mathbf{Y}_{G_o}\rangle\right)^{-1} \mathbf{Y}_{G_o}.$$

Regularization

Assume now that there may be some observation noise, or some corrupted values. Thus, we want to regularize the observed field. This may be done by maximizing a penalized likelihood instead of the likelihood itself. It is close to the prediction problem, except that the constraint $\hat{Y}_{G_o} = \mathbf{Y}_{G_o}$ turns into a penalization on $\left\|\hat{Y}_{G_o} - \mathbf{Y}_{G_o}\right\|_2$. Denote $G = G_o \cup G_m$.
We obtain the following regularized process :

$$\hat{Y}_G^{(r)} := \arg\max_{Z\in\mathbb{R}^G} -Z_G^T \mathcal{K}_G(f_\theta)^{-1} Z_G + \lambda \left(Z_{G_o} - \mathbf{Y}_{G_o}\right)^T \left(Z_{G_o} - \mathbf{Y}_{G_o}\right).$$

Here again, we can use a Whittle approximation, and define instead the regularized process as

$$\hat{Y}_G^{(r)} := \arg\max_{Z\in\mathbb{R}^{\sharp G}} Z_G^T \mathcal{K}_G(\frac{1}{f_\theta}) Z_G + \lambda \left(Z_{G_o} - \mathbf{Y}_{G_o}\right)^T \left(Z_{G_o} - \mathbf{Y}_{G_o}\right).$$

This method may be used to filter the corrupted values. However, the parameter λ has to be adjusted. In practical situation, we choose λ from a cross validation procedure.

Gaussian spatio-temporal process

Now, we develop the same kind of tools for a Gaussian spatio-temporal process. Actually, the major part of the tools used before remains available in the spatio-temporal case.
We will explain the general construction of covariance operators for spatio-temporal processes. For this, we first introduce a toy-model to explain the general form of admissible covariances. Then, we prove a Szegö Lemma for this model. However,

since the proofs are pretty close to the ones performed in Chapter 3, some step are skipped. Nevertheless, we claim that the Szegö Lemma is sufficient, almost in the autoregressive model, to provide the convergence of approximated maximum likelihood estimators.

A toy model

Let us introduce a natural model for the space-time process $\mathbf{X}$. Consider a space-time process, causal-AR_1 in time, and non-causal autoregressive in space. This construction will use the non-causal autoregressive structure given in the last section.
Recall that a causal temporal AR_1 process $(Z_n)_{n\in\mathbb{Z}}$ on $\mathbb{Z}$ verifies :

$$Z_n = aZ_{n-1} + \epsilon_n, |a| < 1.$$

Here ϵ is a white noise of variance σ^2.
In the spatio-temporal case, we can use, in place of ϵ, a sequence $(H_n)_{n\in\mathbb{Z}}$ of i.i.d Gaussian spatial fields of covariance $\mathcal{K}(f)$, where $f \in \mathcal{D}_{\mathcal{K}}$ (let say $f = \frac{1}{P}$, where P is a polynomial of degree p).
Then we can mimic the last construction, and take the following model for $\mathbf{X}$:

$$X_{.,n} := LX_{.,n-1} + H_n,$$

where L is a generator, that will be also chosen as the inverse of a polynomial of the adjacency operator W^{G_∞} (for instance $L = \mathcal{K}(\frac{1}{Q})$, where Q is a polynomial of degree q). Note here that the problem of existence of such processes will be tackled later, using the spectral representation of this process.
This leads to a particular form of covariance. Denote by B the shift operator on $\mathbb{Z}$ and Γ the covariance operator of the process $\mathbf{X}$.
Denote also by $W^{\mathbb{Z}}$ the normalized adjacency operator of the graph $\mathbb{Z}$. That is

$$W^{\mathbb{Z}}_{nm} = \frac{1}{2}\mathbb{1}_{|n-m|=1}, n, m \in \mathbb{Z}.$$

We get

$$\Gamma = \left(I_{G_\infty\otimes\mathbb{Z}} - \left(Q(W^{G_\infty})\right)^{-1}\otimes B^T\right)^{-1}\left(P(W^{G_\infty})^{-1}\otimes I_{\mathbb{Z}}\right)\left(I_{G_\infty\otimes\mathbb{Z}} - \left(Q(W^{G_\infty})\right)^{-1}\otimes B\right)^{-1}.$$

(Here, ^{T} stands for the transposition)
This computation is exactly the same as in the time series case (in this case, we obtain $\sigma^2\left(\mathrm{Id}_{\mathbb{Z}} - aB^T\right)^{-1}(\mathrm{Id}_{\mathbb{Z}} - aB)^{-1}$). Here, $P(W^{G_\infty})^{-1}\otimes I_{\mathbb{Z}}$ stands for σ^2.

Indeed, it is the variance of the process H viewed as a spatio-temporal process. Thus, $\Big(I_{G_\infty\otimes\mathbb{Z}} - (Q(W^{G_\infty}))^{-1}\otimes B\Big)^{-1}$ stands for $(\mathrm{Id}_{\mathbb{Z}} - aB)^{-1}$.
Using the commutativity of $Q(W^{G_\infty})$ and $(P(W^{G_\infty})$, Γ may also be written as :

$$\Gamma = P(W^{G_\infty})^{-1}\otimes I_{\mathbb{Z}}\Big(I_{G_\infty\otimes\mathbb{Z}} - 2Q(W^{G_\infty})\otimes W^{(\mathbb{Z})} + Q(W^{G_\infty})^2\otimes I_{\mathbb{Z}}\Big)^{-1}.$$

This example helps to understand the general structure of the process $\mathbf{X}$. Let us now discuss the general frame, using again a notion of spectral density.

Spectral measure

As above, we work with the renormalized adjacency operator (or with an isotropic modification of the graph, if it may have a better physical sense).
As we hope to build spatio-temporal processes, with discrete time, we have to consider also the graph $\mathbb{Z}$, that will correspond to the indices of the time ($n\in\mathbb{Z}$). We will use both the spectral decomposition of the graph G_∞ and $\mathbb{Z}$. Write

$$W^{G_\infty} = \int_{\mathrm{Sp}(W^{G_\infty})} x\mathrm{d}E^{G_\infty}(x),$$
$$W^{\mathbb{Z}} = \int_{\mathrm{Sp}(W^{\mathbb{Z}})} t\mathrm{d}E^{\mathbb{Z}}(t).$$

Here, E^{G_∞} denotes, as in Chapter 3 the identity resolution associated to the operator W^{G_∞}, and $E^{\mathbb{Z}}$ the identity resolution associated to the operator $W^{\mathbb{Z}}$.
As for the graph G_∞, we can define, for any $n,m\in\mathbb{Z}$, the local measure of $\mathbb{Z}$ at the couple of vertices n,m. That is

$$\mu^{\mathbb{Z}}_{nm} := \langle\delta^{(n)}, E^{\mathbb{Z}}\delta^{(m)}\rangle_{l^2(\mathbb{Z})}.$$

Actually, this family of measure is well-known :

$$\forall n,m\in\mathbb{Z}, \mathrm{d}\mu^{\mathbb{Z}}_{nm}(t) = \frac{T_{|n-m|}(t)}{\sqrt{1-t^2}}\mathrm{d}\lambda_{[-1,1]}(t),$$

where T_k denotes the k^{th} Chebychev's polynomial, and $\mathrm{d}\lambda_{[-1,1]}$ denotes the restriction of the Lebesgue measure to $[-1,1]$ (see Chapter 5). In particular, the measure $\mu^{\mathbb{Z}}_{nn}, n\in\mathbb{Z}$ is constant and equal to the arc-sinus law, at any vertex $n\in\mathbb{Z}$.
So that, the global spectral measure, with respect to any growing sequence of nested subgraphs of $\mathbb{Z}$, is also the arc-sinus law :

$$\mathrm{d}\mu^{\mathbb{Z}} = \frac{1}{\sqrt{1-t^2}}\mathrm{d}\lambda_{[-1,1]}(t).$$

Model

Back to the toy model, we can now write in the spectral domain the covariance operator obtained in this case.
Using

$$\Gamma = P(W^{G_\infty})^{-1} \otimes I_{\mathbb{Z}} \Big(I_{G_\infty \otimes \mathbb{Z}} - 2Q(W^{G_\infty}) \otimes W^{(\mathbb{Z})} + Q(W^{G_\infty})^2 \otimes I_{\mathbb{Z}} \Big)^{-1}.$$

We get

$$\Gamma = \int_{x \in \mathrm{Sp}(W^{G_\infty})} \int_{t \in \mathrm{Sp}(W^{\mathbb{Z}})} \frac{1}{P(x)} \frac{1}{1 - 2Q(x)t + Q(x)^2 t^2} \mathrm{d}E^{G_\infty}(x) \otimes \mathrm{d}E^{\mathbb{Z}}(t),$$

where $\otimes$ denotes the tensor product.
Hence, if we denote by $\mathcal{D}^{\mathrm{st}}_{\mathcal{K}}$ the set of functions whose inverse is a polynomial of two variables, with no roots in the closure of the convex hull of $\mathrm{Sp}(W^{G_\infty}) \times \mathrm{Sp}(W^{\mathbb{Z}})$, it is natural to define, for any non negative $\phi \in \mathcal{D}^{\mathrm{st}}_{\mathcal{K}}$, the covariance operator $\mathcal{K}^{\mathrm{st}}(\phi)$ associated to the spectral density ϕ by :

$$\mathcal{K}^{\mathrm{st}}(\phi) = \int_{x \in \mathrm{Sp}(W^{G_\infty})} \int_{t \in \mathrm{Sp}(W^{\mathbb{Z}})} \phi(x,t) \mathrm{d}E^{G_\infty}(x) \otimes \mathrm{d}E^{\mathbb{Z}}(t).$$

Actually, for the monomial $\phi(x,t) = x^k t^l$, the corresponding covariance operator is nothing else than the spatial operator $\mathcal{K}$ applied to the identity function (which is not nonnegative, but the definition may be extended to non covariance operators) over the tensor product $(W^{G_\infty})^k \otimes (W^{\mathbb{Z}})^l$. That is

$$\mathcal{K}^{\mathrm{st}}(x^k t^l) = (W^{G_\infty})^k \otimes (W^{\mathbb{Z}})^l.$$

In particular, when $\phi(x,t) = f(xt)$, the corresponding covariance is the covariance associated to an MA_∞ process on the Kronecker product $W^{G_\infty} \otimes W^{\mathbb{Z}}$. In this case, we have

$$\mathcal{K}^{\mathrm{st}}(\phi) = \mathcal{K}^{W^{G_\infty} \otimes W^{\mathbb{Z}}}(f).$$

Finally, denote by $\mathcal{K}^{\mathrm{st}}_O(\phi)$ the restriction of the operator $\mathcal{K}^{\mathrm{st}}(\phi)$ to a subset O of $G_\infty \times \mathbb{Z}$.
Note that this construction is coherent with the stationarity. If we fix a vertex $i_0 \in G_\infty$ and consider the corresponding process $\mathbf{X}_{i_0,.}$, it is a stationary time series. Its spectral density may be compute by an integration of ϕ over the first variable x with the measure $\mu_{i_0 i_0}$. Symmetrically, fix a time t_0, the corresponding Gaussian field is a classical autoregressive spatial process.

Szegö Lemma

In this section, we prove only the Szegö Lemma for the space-time processes defined in the last subsection, but not the whole convergence theorem. It avoids some technical proofs, very close to the ones of Chapter 3. We admit that it is sufficient to ensure the convergence of the corresponding estimators. Actually, the proof is very close to the one in the spatial case. So that, we mimic our last proof.
Consider the cylinder $C_{N,T} = G_N \times [1,T]$. Consider also its complementary

$$F_{N,T} = (G \times \mathbb{Z}) \setminus (G_N \times [1,T]).$$

Then, denote $S_{N,T}$ the surface area of the cylinder $C_{N,T}$.

$$S_{N,T} = 2\sharp G_N + T\sharp\delta(G_N).$$

For any matrix $(B_{(i,n),(j,m)})_{(i,n),(i',n')\in C_{N,T}}$, we define the block norm

$$b_{N,T}(B) = \frac{1}{S_{N,T}} \sum_{(i,n),(i',n')\in C_{N,T}} \left|B_{(i,n),(i',n')}\right|.$$

We can state the equivalent version of the first Szegö lemma

Lemma 4.1.1 *Asymptotic homomorphism*
Let $\phi, \psi \in \mathcal{D}^{st}_{\mathcal{K}}$, then, we have the following control

$$b_{N,T}\left(\mathcal{K}^{st}_{G_N\times[1,T]}(\phi)\mathcal{K}^{st}_{G_N\times[1,T]}(\psi) - \mathcal{K}^{st}_{G_N\times[1,T]}(\phi\psi)\right) \leq C_{\phi\psi}.$$

Proof. of Lemma 4.1.1
Let $\phi, \psi \in \mathcal{D}^{\mathrm{st}}_{\mathcal{K}}$.
We can write $\forall (x,t) \in \mathrm{Sp}(W^{G_\infty}) \times \mathrm{Sp}(W^{\mathbb{Z}}), \psi(x,t) = \sum_{k,l\in\mathbb{N}} \psi_{kl} x^k t^l$, where the ψ_{kl} decrease exponentially, because $\frac{1}{\psi}$ is a polynomial, with no roots in $\mathrm{Sp}(W^{G_\infty}) \times \mathrm{Sp}(W^{\mathbb{Z}})$.
We write

$$\begin{aligned} & b_{N,T}\left(\mathcal{K}^{\mathrm{st}}_{G_N\times[1,T]}(\phi)\mathcal{K}^{\mathrm{st}}_{G_N\times[1,T]}(\psi) - \mathcal{K}^{\mathrm{st}}_{G_N\times[1,T]}(\phi\psi)\right) \\ = & \frac{1}{S_{N,T}} \sum_{(i,n),(i',n')\in C_{N,T}} \sum_{(j,m)\in F_{N,T}} \left|\mathcal{K}^{\mathrm{st}}_{G_N\times[1,T]}(\phi)_{(i,n),(j,m)}\right| \left|\mathcal{K}^{\mathrm{st}}_{G_N\times[1,T]}(\psi)_{(j,m),(i',n')}\right|. \end{aligned}$$

Using $\mathcal{K}^{\mathrm{st}}(\psi) = \sum_{k,l=0}^{\infty} \psi_{kl} (W^{G_\infty})^k \otimes (W^{\mathbb{Z}})^l$, Fubini's theorem gives, since all the previous sequences are in $l^1(G_\infty \times \mathbb{Z})$,

$$
\begin{aligned}
& b_{N,T}\Big(\mathcal{K}^{\mathrm{st}}_{G_N\times[1,T]}(\phi)\mathcal{K}^{\mathrm{st}}_{G_N\times[1,T]}(\psi) - \mathcal{K}^{\mathrm{st}}_{G_N\times[1,T]}(\phi\psi)\Big) \\
\leq\ & \frac{1}{S_{N,T}} \sum_{(i,n),(i',n')\in C_{N,T}} \sum_{(j,m)\in F_{N,T}} \left| \left(\mathcal{K}^{\mathrm{st}}_{G_N\times[1,T]}(\phi)\right)_{(i,n),(j,m)} \left(\mathcal{K}^{\mathrm{st}}_{G_N\times[1,T]}(\psi)\right)_{(j,m),(i',n')} \right| \\
\leq\ & \left(\sup_{(j,m)\in F_{N,T}} \sum_{(i,n)\in C_{N,T}} \left| \mathcal{K}^{\mathrm{st}}_{G_N\times[1,T]}(\phi)_{(i,n),(j,m)} \right| \right) \\
& \times \frac{1}{S_{N,T}} \sum_{(j,m)\in F_{N,T}} \sum_{(i',n')\in C_{N,T}} \sum_{k,l=0}^{\infty} |\psi_{k,l}| \left| (W^{G_\infty})^k_{ji'} (W^{\mathbb{Z}})^l_{mn'} \right| \\
\leq\ & \left(\sup_{(j,m)\in F_{N,T}} \sum_{(i,n)\in C_{N,T}} \left| \mathcal{K}^{\mathrm{st}}_{G_N\times[1,T]}(\phi)_{(i,n),(j,m)} \right| \right) \\
& \times \sum_{k,l=0}^{\infty} |\psi_{kl}| \frac{1}{S_{N,T}} \sum_{(j,m)\in F_{N,T}} \sum_{(i',n')\in C_{N,T}} \left| (W^{G_\infty})^k_{ji'} (W^{\mathbb{Z}})^l_{mn'} \right| .
\end{aligned}
$$

Introducing

$$
\Delta_{k,l} = \sup_{N,T\in\mathbb{N}} \frac{1}{S_{N,T}} \sum_{(j,m)\in F_{N,T}} \sum_{(i',n')\in C_{N,T}} \left| (W^{G_\infty})^k_{ji'} (W^{\mathbb{Z}})^l_{mn'} \right| ,
$$

we get

$$
\begin{aligned}
& b_{N,T}\Big(\mathcal{K}^{\mathrm{st}}_{G_N\times[1,T]}(\phi)\mathcal{K}^{\mathrm{st}}_{G_N\times[1,T]}(\psi) - \mathcal{K}^{\mathrm{st}}_{G_N\times[1,T]}(\phi\psi)\Big) \\
& \qquad \leq \sup_{(j,m)\in G_\infty\times\mathbb{Z}} \sum_{(i,n)\in G_\infty\times\mathbb{Z}} \left| \mathcal{K}^{\mathrm{st}}(\phi)_{(i,n),(j,m)} \right| \sum_{k,l=0}^{\infty} |\psi_{k,l}| \, \Delta_{k,l}.
\end{aligned}
$$

The coefficient $\Delta_{k,l}$ is a porosity factor. It measures the weight of the paths of length (k,l) going from the interior of the cylinder $G_N \times [1,N]$ to outside.
Note that $\Delta_{k,l} \leq kl + 1$, so we get

$$
\sum_{kl=0}^{\infty} |\psi_{kl}| \, \Delta_{kl} \leq \sum_{kl=0}^{\infty} |\psi_{kl}| \, (kl+1).
$$

Thanks to the exponential decreasing, this last expression is finite.
Now, we define another norm on $B_{G_\infty\times\mathbb{Z}}$:

$$
\|B\|_{\infty,in} := \sup_{(i,n)\in G_\infty\times\mathbb{Z}} \sum_{(j,m)\in G_\infty\times\mathbb{Z}} \left| B_{(i,n),(j,m)} \right| , (B \in B_{G_\infty\times\mathbb{Z}}) .
$$

We thus obtain

$$\begin{aligned}
\left\|\mathcal{K}^{\mathrm{st}}(\phi)\right\|_{\infty,in} &= \sup_{(i,n)\in G_\infty\times\mathbb{Z}} \sum_{(j,m)\times\mathbb{Z}} \left|\mathcal{K}^{\mathrm{st}}(\phi)_{(i,n)(j,m)}\right| \\
&\leq \sum_{k,l=0}^{\infty} |\phi_{kl}| \left\|(W^{G_\infty})^k\right\|_{\infty,in} \left\|(W^{\mathbb{Z}})^l\right\|_{\infty,in} \\
&\leq \sum_{k,l=0}^{\infty} |\phi_{kl}| \left\|W^{G_\infty}\right\|_{\infty,in}^k \left\|W^{G_\infty}\right\|_{\infty,in}^k \left\|W^{\mathbb{Z}}\right\|_{\infty,in}^l \\
&\leq \sum_{k,l=0}^{\infty} |\phi_{k,l}| := \|\phi\|_{1,pol}\,.
\end{aligned}$$

Finally, we get

$$b_{N,T}\left(\mathcal{K}^{\mathrm{st}}_{G_N\times[1,T]}(\phi)\mathcal{K}^{\mathrm{st}}_{G_N\times[1,T]}(\psi) - \mathcal{K}^{\mathrm{st}}_{G_N\times[1,T]}(\phi\psi)\right) \leq \|\phi\|_{1,pol} \sum_{k,l=0}^{\infty} |\psi_{kl}|\,(kl+1).$$

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

$$b_{N,T}\left(\mathcal{K}^{\mathrm{st}}_{G_N\times[1,T]}(\phi)\mathcal{K}^{\mathrm{st}}_{G_N\times[1,T]}(\psi) - \mathcal{K}^{\mathrm{st}}_{G_N\times[1,T]}(\phi\psi)\right) \leq \frac{1}{2}\sum_{k,l=0}^{\infty} |\phi_{kl}|\,(kl+1) \sum_{kl=0}^{\infty} |\psi_{kl}|\,(kl+1).$$

Now, the exponential decreasing of the coefficients gives

$$\frac{1}{2}\sum_{k,l=0}^{\infty} |\phi_{kl}|\,(kl+1) \sum_{kl=0}^{\infty} |\psi_{kl}|\,(kl+1) \leq C_{\phi,\psi} < \infty.$$

This finishes the proof of the Lemma ∎

Estimation

Now, the Szegö Lemma is proven. We admit now, that it is sufficient to get the convergence of the approximated maximum likelihood estimator built in the last subsection . However, we did not tackle the question of asymptotic normality and efficiency. In this case, build an unbiased periodogram could be very tricky. Hopefully, it is not very important in prediction or regularization issues. Indeed, in such schemes, the variance of the underlying white noise (corresponding to the innovation in the case of time series) makes the prediction error much larger than the bias of the estimator.

As in the last section, assume that Θ is a compact subset of $\mathbb{R}^d, d \geq 1$, and that we have a parametric family of spectral densities $(\phi_\theta)_{\theta\in\Theta}$, such that $\forall\theta \in \Theta, \phi_\theta \in \mathcal{D}^{\text{st}}_{\mathcal{K}}$. Then, let $\theta_0 \in \mathring{\Theta}$, and assume that $\mathbf{X}$ is a centered Gaussian space-time process with covariance $\mathcal{K}^{\text{st}}(\phi_{\theta_0})$.
Again, let $(G_N)_{N\in\mathbb{N}}$ be a growing sequence of nested subgraphs of G_∞. Assume that we observe $\mathbf{X}_{G_N,T}$, the restriction of the process $\mathbf{X}$ to the cylinder $G_N \times [1, T]$ (so that, the asymptotic is meant as N, T go to infinity, and we need to specify the joint rate). That is,

$$\mathbf{X}_{G_N,T} \sim \mathcal{N}\left(0, \mathcal{K}^{\text{st}}_{G_N\times[1,T]}(\phi)\right).$$

Now, we can write the normalized log-likelihood of the model :

$$l_N(\theta, \mathbf{X}_{G_N,T}) = -\frac{1}{2}\Bigg(\log(2\pi) + \frac{1}{T\sharp G_N}\log\det\left(\mathcal{K}^{\text{st}}_{G_N,[1,T]}(\phi_\theta)\right) + \frac{1}{T\sharp G_N}\mathbf{X}^T_{G_N,T}\left(\mathcal{K}^{\text{st}}_{G_N,[1,T]}(\phi_\theta)\right)^{-1}\mathbf{X}_{G_N,T}\Bigg).$$

The Whittle's approximation consists in a modification of this expression, as in the spatial case.
We get

$$\tilde{l}_N(\theta, \mathbf{X}_{G_N,T}) = -\frac{1}{2}\Bigg(\log(2\pi) + \int\log(\phi_\theta(x))\mathrm{d}\mu(x) + \frac{1}{T\sharp G_N}\mathbf{X}^T_{G_N,T}\mathcal{K}_{G_N,[1,T]}\left(\frac{1}{\phi_\theta}\right)\mathbf{X}_{G_N,T}\Bigg).$$

Finally, we estimate θ_0 by maximizing the last expression :

$$\hat{\theta}_N := \arg\max \tilde{l}_N(\theta, \mathbf{X}_{G_N,T}).$$

We admit that if G_N, T are such that

$$\frac{\sharp C_{N,T}}{S_{N,T}} \xrightarrow{N,T\to\infty} 0,$$

then this estimator is consistent.

Prediction and regularization

Actually, there is nothing more than in the spatial case. We have only to specify the framework for the prediction and the regularization. Once we know the covariance

estimator (for instance after an estimation over an independent sample), we can plug it into the projection operator.
Here, the observation indices is a subset $O \subset G_\infty \times \mathbb{Z}$. The procedure is the same as in the spatial case.

4.2 Estimation of the spectral measure

In this section, we compute the spectral measure of the traffic network.
Indeed, in practical cases, the global measure μ^{G_∞} is unknown and has to be estimated.
The assumption $\sharp\delta(G_N) = o(\sharp G_N)$ is really reasonable for the traffic framework, and holds in all simulations done for this work. As said before, the traffic network is embedded into $\mathbb{R}^2$ without accumulation points, and so that, it is amenable. That means that we can choose some sequences of nested subgraphs such that the assumption holds.
Now, let us introduce the procedure performed here to estimate the spectral measure. As a matter of fact, thanks to the homogeneity of the traffic network, this limit does not depend on the choice of the sequence of subgraphs. We point out that we will discuss the structure of this measure in a forthcoming paper.
Figure 4.1 shows different area of the French road traffic network where the spectral measure has been estimated. Actually, the data and the results come from Mediamobile, a French firm specialized in prediction of road traffic. Some other experiments (estimation, prediction, regularization) using the methods developed here are still in progress.
Figure 4.2 gives the results of this estimation step, and an goodness-of-fit χ^2 test between all this distributions. The null-hypothesis is rejected in none of the simulations done. This means that we recover the same pattern in every studied area. Therefore, Whittle type estimation may be a good framework for road-traffic forecasting.

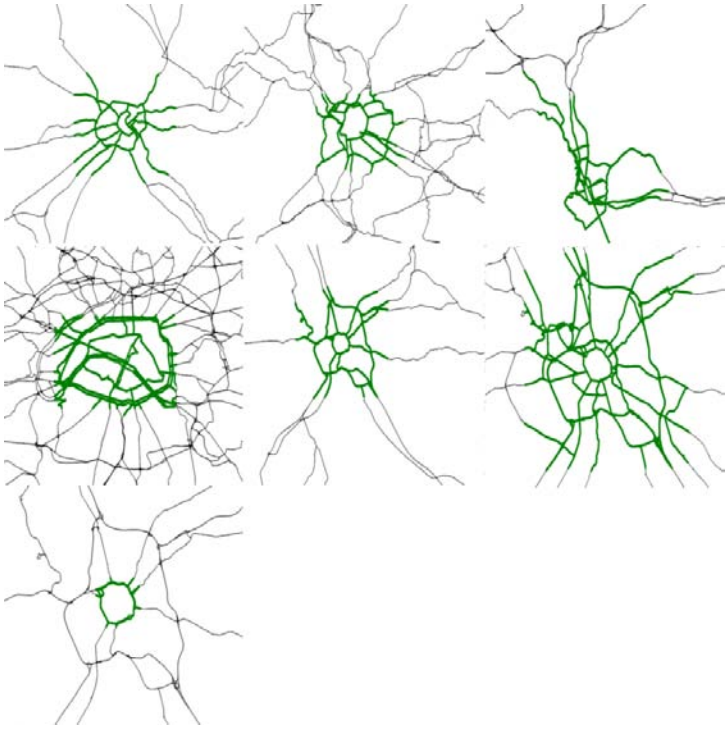


FIGURE 4.1: Selected Areas

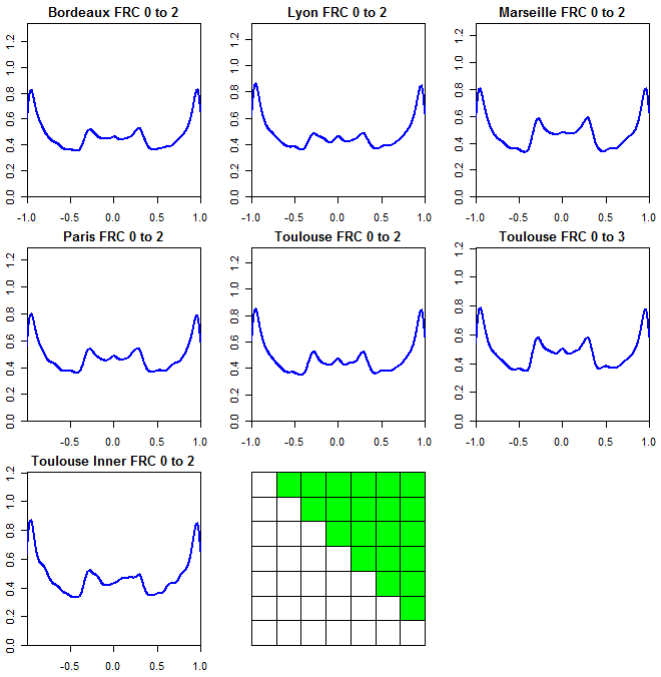


FIGURE 4.2: Goodness-of-fit χ^2 test. Green is for non rejected null hypothesis

Chapitre 5

Spectral measure : a little survey with the Hilbertian point of view

Cette partie propose une discussion et une revue bibliographique autour de la notion de mesure spectrale d'un graphe. Nous présentons quelques propriétés de la mesure locale, pour une classe de graphes très particulière, puis donnons plusieurs exemples de graphes sur lesquels cette mesure peut être calculée explicitement.

Introduction

In this chapter, we highlight the notion of spectral measure which is at the core of our study. To this aim, we first recall some well known spectral properties of several structured graphs, and then show how they can be recovered using simple Hilbertian tools.
We will work with a few examples of the literature : the square lattice $\mathbb{Z}^d, d \geq 1$, the homogeneous tree $\mathcal{A}_{q+1}$, any distance-transitive graph, which is more general than the homogeneous tree (see for instance [39] for the definition), and the semi-homogeneous tree.
Notice that the case of the time series $G = \mathbb{Z}$ has already been explained in Chapter 3.
In the first Section, we give a way to compute the local measure at the root, in a specific class of rooted graphs.
Then, Section 5.2 is devoted to recover the spectral measure on several examples of graphs.
Actually, all of the following work is more or less well known. In particular, this

work follows both the ideas of Mohar [55] and Bordenave [19]. A very close approach appears in all the recent works of Obata [51], [46]. This last work deals with spectral analysis, with a quantum point of view, of growing graphs. It is interesting for our purpose of statistical inference on graphs.
We underline here the importance of the adjacency algebra, for statistical inference (in particular for the Whittle approximation), and of its spectral representation. Indeed, $MA_P, P < +\infty$ processes are nothing more than processes whose covariance operators lies in the adjacency algebra of the graph. Studying the properties of this kind of processes means studying this algebra.

5.1 Simple computation of the local measure at the root

In this section, we study a very **specific class** of rooted graphs, and give an elementary way to understand the measure at the root. In particular, this computation does not involve Gelfand's pair or any other tools from harmonic analysis [45], [50], [65]. This relies only on the spectral decomposition of Hilbertian operators. This point of view is also the one chosen by Mohar and Woess to present the results in their survey on spectra of infinite graphs [55], and is close to the tools used by Bordenave [18] . Indeed, the induction proposed here is equivalent, in the case of trees, to an induction on the resolvent function over the size of the tree.
Our computation gives another way to understand the spectral decomposition of time series and spatial processes and maybe useful in other frameworks.
Let us first introduce the class of rooted graphs that is considered in this work.
A rooted graph $(\mathbf{G}, o)$ is a graph $\mathbf{G} = (G, W)$ (for the definition, see Chapter 1) given with a root $o \in G$. Here, we choose unweighted graphs, for a sake of simplicity. Then, W denotes the adjacency operator (or matrix, if $\sharp G < +\infty$).
We assume that $\mathbf{G}$ has bounded degree. Hence, $\mathrm{Sp}(W)$ is a compact subset of $\mathbb{R}$ (see Chapter 1).
Then, we denote by S_k the k^{th}-stratum of the graph. That is

$$S_k := \{j \in G, d(o, j) = k\}.$$

Here, d stands for the natural distance on the graph G (see the Introduction).
Denote also, for any $i \in G$, and for all $l \geq 0$,

$$s_l(i) := \{j \in S_l, j \sim i\}.$$

Note that, if $i \in S_k, k \geq 0$, then $s_l(i) = \emptyset$ as soon as $|k - l| > 1$.
Finally, denote, for $i \in S_k, k \geq 0$ by $a_k(i) = \sharp s_{k-1}(i)$, $b_k(i) = \sharp s_k(i)$ and $c_k(i) = \sharp s_{k+1}(i)$ (by convention, $s_{-1}(o) = \emptyset$). Denote also $d_k = \sharp S_k$.
From now, we make the following assumption on G :

Assumption 5.1.1 *For any $k \geq 0$, a_k, b_k and c_k do not depend on $i \in S_k$. This means that any vertex at the same stratum has the same number of neighbors on the inferior stratum (a_k), the same stratum (b_k), and the superior stratum (c_k). Figure 5.1 shows a part of an example of such graph.*

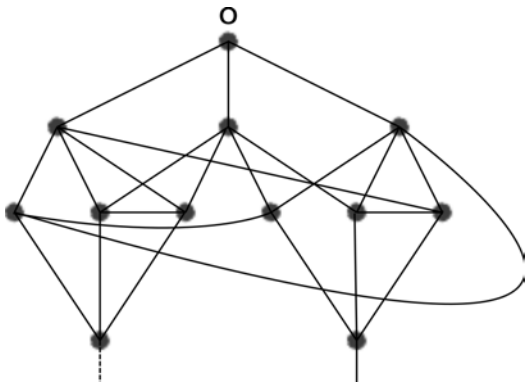


FIGURE 5.1: Exemple of graph which verifies Assumption 5.1.1

As in Chapter 3, we consider the spectral decomposition of W, with respect to an identity resolution E :

$$W = \int_{\mathrm{Sp}(W)} \lambda \mathrm{d}E(\lambda).$$

Consider also the family of (signed) measures $(\mu_{ij})_{i,j\in G}$ defined in the Chapter 3 by

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

where, for any $i \in G$, the sequences δ_i in $l^2(G)$ is defined by

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

Denote $\nu_0 = \mu_{00}$, and $\nu_k = \frac{1}{\sharp S_k} \sum_{i\in S_k} \mu_{0i}$. Note that, these measures have compact support subset of $\mathrm{Sp}(W)$.
Note also that the measure $\mu_{ii}, i \in G$ (and it is true for a diagonal local measure of any bounded Hilbertian normal operator) is always a probability measure.
We can show that :

Proposition 5.1.2 *The family of measures $(\nu_k)_{k\in\mathbb{N}}$ verifies :*
- *For any $k \geq 0$, the measure ν_k is absolutely continuous with respect to ν_0 ;*
- *For any $k \geq 0$, the density $P_k := \frac{\mathrm{d}\nu_k}{\mathrm{d}\nu_0}$ is a polynomial of degree k ;*
- *The family of polynomials $(P_k)_{k\in\mathbb{N}}$ verifies the following inductive equalities :*

1. $\forall x \in \mathrm{Sp}(W), P_0(x) = 1.$
2. $\forall x \in \mathrm{Sp}(W), P_1(x) = \frac{x}{c_0}.$
3. $\forall x \in \mathrm{Sp}(W), xP_k(x) = a_k P_{k-1}(x) + b_k P_k(x) + c_k P_{k+1}(x).$

Note here that the sequences $(a_k)_{k\geq 0}$ and $(c_k)_{k\geq 0}$ are linked. Indeed, counting the total number of edges going from stratum S_{k-1} to stratum S_k gives :

$$a_k d_k = c_{k-1} d_{k-1}.$$

Proof. Let us prove this result by induction : First, we have

$$P_0 := \frac{\mathrm{d}\nu_0}{\mathrm{d}\nu_0} := 1,$$

by definition.
Let ϕ be a polynomial. Let $\Gamma = \phi(W)$. We can write

$$(W\Gamma)_{oo} = \int_{\mathrm{Sp}(W)} x\phi(x)\mathrm{d}\nu_0(x).$$

But, using $W\Gamma_{oo} = \sum_{j\sim o} \Gamma_{jo} = \sum_{j\in S_1} \Gamma_{jo}$, we also have

$$(W\Gamma)_{oo} = d_1 \int_{\mathrm{Sp}(W)} \phi(x)\mathrm{d}\nu_1.$$

Therefore, the equality holds for any polynomial ϕ, and we get that ν_1 is absolutely continuous with respect to ν_0, and that

$$\forall x \in \mathrm{Sp}(W), P_1(x) := \frac{\mathrm{d}\nu_1}{\mathrm{d}\nu_0}(x) = \frac{x}{d_1} = \frac{x}{c_0}.$$

Now, let $k > 1$. To perform the inductive step, assume that, for any $j \leq k$, ν_j is absolutely continuous with respect to ν_0, and that the polynomials

$$P_j = \frac{\mathrm{d}\nu_j}{\mathrm{d}\nu_0}, j \leq k,$$

verify the inductive equality given in the proposition.
We can write :

$$\frac{1}{d_k} \sum_{i\in S_k} (W\Gamma)_{oi} = \int_{\mathrm{Sp}(W)} x\phi(x)\mathrm{d}\nu_k(x).$$

On the other hand, we have,

$$\sum_{i\in S_k} (W\Gamma)_{oi} = \int_{\mathrm{Sp}(W)} \phi(x)\Big(c_{k-1}d_{k-1}\mathrm{d}\nu_{k-1} + b_k d_k \mathrm{d}\nu_k + a_{k+1}d_{k+1}\mathrm{d}\nu_{k+1}\Big).$$

Since this equality is true for any polynomial ϕ, once again, this leads to the absolute continuity of ν_{k+1} with respect to ν_0. Furthermore, we get that the density $P_{k+1} := \frac{\mathrm{d}\nu_{k+1}}{\mathrm{d}\nu_0}$ is a polynomial of degree $k+1$ which verifies :

$$\forall x \in \mathrm{Sp}(W), d_k x P_k(x) = c_{k-1} d_{k-1} P_{k-1}(x) + b_k d_k P_k(x) + a_{k+1} d_{k+1} P_{k+1}(x).$$

Hence, using $a_{k+1} d_{k+1} = c_k d_k$, we get

$$\forall x \in \mathrm{Sp}(W), x P_k(x) = a_k P_{k-1}(x) + b_k P_k(x) + c_k P_{k+1}(x).$$

This ends the proof of the proposition. ∎

Now, we are looking for a family of polynomials $(Q_k)_{k\in\mathbb{N}}$ such that,

$$\forall k, p \in \mathbb{N}, \sum_{g\in S_p} (Q_k(W))_{og} = \mathbb{1}_{p=k}.$$

With another induction, we can obtain also the following proposition.

Proposition 5.1.3 *The family of polynomials $(P_k)_{k\in\mathbb{N}}$ defined in Proposition 5.1.2 verifies the equality :*

$$\forall k, p \in \mathbb{N}, \sum_{g\in S_p} (P_k(W))_{og} = \mathbb{1}_{p=k}.$$

Proof.
Again, we will prove the proposition by induction on k. For the initialization, let $p \geq 0$. Since $P_0 = 1$, we have $P_0(W) = \mathrm{Id}$ and so

$$\sum_{g\in S_p} (P_0(W))_{og} = \mathbb{1}_{p=0}.$$

Then recall that,

$$P_1(x) = \frac{x}{d_1}.$$

We can write, for any $p \geq 0$,

$$\sum_{g\in S_p} (P_1(W))_{og} = \sum_{g\in S_p} \frac{W_{og}}{d_1} = \mathbb{1}_{p=1},$$

by definition of S_1 and d_1.
Let $k > 1$. To perform the inductive step, let us write, for any $p \geq 0$,

$$\begin{aligned}\sum_{g\in S_p}(P_k(W)W)_{og} &= \sum_{g\in S_p}\sum_{j\sim g}(P_k(W))_{oj}\\ &= c_{p-1}\sum_{j\in S_{p-1}}(P_k(W))_{oj} + b_p\sum_{j\in S_p}(P_k(W))_{oj} + a_{p+1}\sum_{j\in S_{p+1}}(P_k(W))_{oj}\\ &= c_k\mathbb{1}_{p=k+1} + b_k\mathbb{1}_{p=k} + a_k\mathbb{1}_{p=k-1}.\end{aligned}$$

Then, using

$$\forall x\in \mathrm{Sp}(W), xP_k(x) = a_kP_{k-1}(x) + b_kP_k(x) + c_kP_{k+1}(x),$$

we may compute

$$\begin{aligned}c_k\sum_{g\in S_p}P_{k+1}(W) &= \sum_{g\in S_p}(P_k(W)W - b_kP_k(W) - a_kP_{k-1}(W))_{og}\\ &= c_k\mathbb{1}_{p=k+1} + b_k\mathbb{1}_{p=k} + a_k\mathbb{1}_{p=k-1} - b_k\mathbb{1}_{p=k} - a_k\mathbb{1}_{p=k-1}\\ &= c_k\mathbb{1}_{p=k+1}.\end{aligned}$$

This ends the proof of this proposition. ■
Now, using both propositions 5.1.2 and 5.1.3, we obtain

Proposition 5.1.4 *The family of polynomials $(P_k)_{k\in\mathbb{N}}$ is orthogonal with respect to the measure ν_0.*

Proof. Indeed, on the one hand, we have, for any $k,p\geq 0$,

$$\mathbb{1}_{k=p} = \sum_{g\in S_p}(P_k(W))_{og} = d_p\int_{\mathrm{Sp}(W)}P_k(x)\mathrm{d}\nu_p.$$

But, using $\frac{\mathrm{d}\nu_p}{\mathrm{d}\nu_0} = P_p$, we get

$$\frac{1}{d_p}\sum_{g\in S_p}(P_k(W))_{og} = \int_{\mathrm{Sp}(W)}P_k(x)P_p(x)\mathrm{d}\nu_0.$$

So that,

$$\int_{\mathrm{Sp}(W)}P_k(x)P_p(x)\mathrm{d}\nu_0 = \frac{1}{d_p}\mathbb{1}_{k=p}.$$

■

Now, we can use the Stieltjes transform to compute the spectral measure ν_0. This will be tackled in a few classical examples in the next section.
We will consider some very structured graphs. For this, define an automorphism of the rooted graph (G,o) as an automorphism of the graph G which stabilized the root.

Consider now a rooted graph (G, o) such that its quotient with its set of automorphisms is isomorphic to $\mathbb{N}$. This means that all vertices in the same stratum play the same role. This implies in particular that, for two vertices $j, g \in G$ which belong to the same stratum, one has $\mu_{og} = \mu_{oj}$. This leads to, for $j \in S_k, k \geq 0$,

$$\mu_{oj} = \nu_k.$$

5.2 A few examples

In this section, we give some simple applications of the last computation, and a simple way to recover classical spectral measures. Note that, in all the examples below (except for the semi-homogeneous tree) the graphs are vertex-transitive. This means that any vertex can be send to any other by an automorphism. It implies that the local measure is equal at any vertex, and then that it is also equal to the global spectral measure μ (defined in Chapter 3), whathever the choice of the sequence of subgraphs $(G_n)_{n\in\mathbb{N}}$ is. Recall that the global measure is defined as the weak limit of the mean local measure on G_n, uniformly rooted :

$$\mu := \lim_{n\to\infty} \frac{1}{\sharp G_n} \sum_{o\in G_n} \mu_{oo}, \text{ in the sense of the weak convergence.}$$

The two way infinite path $G = \mathbb{Z}$

Let us chose the origin as 0, and consider the rooted graph $(\mathbb{Z}, 0)$. It verifies Assumptions 5.1.1, with the sequences

$$\begin{aligned} a_k &= 1, k \geq 1, \\ b_k &= 0, \forall k \geq 0, \\ c_0 &= 2, \\ c_k &= 1, k \geq 1. \end{aligned}$$

Denote by $(P_k)_{k\in\mathbb{N}}$ the family of polynomials which verifies that the local measure at the root is orthogonal with respect to this measure. This family verifies the relations

$$\begin{aligned} P_0(x) &= 1, \\ P_1(x) &= x/2, \\ \forall k \geq 1, xP_k(x) &= P_{k+1}(x) + P_{k-1}(x). \end{aligned}$$

Now, consider the change of variables $t = \frac{x}{2}$, and define, for any $k \geq 0$,

$$T_k(t) := P_k(x).$$

We get

$$\begin{aligned} T_0(t) &= 1 \\ T_1(t) &= t, \\ \forall k \geq 1, 2tT_k(t) &= T_{k+1}(t) + T_{k-1}(t). \end{aligned}$$

We recognize the Chebychev polynomials, and using the properties of this family of polynomials, we recover

– The measure $\mathrm{d}\mu_{00}^{Z}$ is absolutely continuous with respect to the Lebesgue measure $\mathrm{d}\lambda$
– The density verifies

$$\forall x \in [-2, 2], \frac{\mathrm{d}\mu_{00}^{\mathbb{Z}}}{\mathrm{d}\lambda}(x) = \frac{1}{\pi\sqrt{1 - \frac{x^2}{4}}}.$$

– Moreover, we get

$$\forall x \in [-2, 2], \forall k \in \mathbb{Z}, \frac{\mathrm{d}\mu_{0k}^{\mathbb{Z}}}{\mathrm{d}\mu_{00}^{\mathbb{Z}}}(x) = T_{|k|}(\frac{x}{2}).$$

Finally, let g be a positive measurable function, analytic over $[-2, 2]$, and define

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

The operator Γ is positive definite. Let $\mathbf{X}$ be a Gaussian process of covariance Γ. Here g denotes the spectral density of the process $\mathbf{X}$, in the sense of our framework. Define futher, for any $k \in \mathbb{Z}$,

$$r_k := \Gamma_{0k} = \langle X_0, X_k \rangle.$$

Using the definition of $\mu_{0k}^{\mathbb{Z}}$, we have

$$\begin{aligned} r_k &= \int_{[-2,2]} g(x) \mathrm{d}\mu_{0k}^{\mathbb{Z}} \\ &= \frac{1}{\pi} \int_{[-2,2]} g(x) T_{|k|}(\frac{x}{2}) \frac{1}{\sqrt{1 - \frac{x^2}{4}}} \mathrm{d}\lambda(x) \end{aligned}$$

Define f by the relation

$$\forall t \in [-\pi, \pi], f(t) = g(2\cos(t)).$$

We get, using the change of variables $2\cos(t) = x$,

$$r_k = -\frac{2}{\pi} \int_{[\pi,0]} f(t) \cos(kt) \mathrm{d}\lambda(t)$$

Example 1 (The two way infinite path)
Using the notations introduced before, we obtain

$$r_k = \frac{1}{\pi}\int_{[-\pi,\pi]} f(t)\cos(kt)\mathrm{d}\lambda(t)$$

We recover the usual spectral framework for time series, with the notations given in Introduction.

The square lattice $G = \mathbb{Z}^d$

The computation of the spectral measure of $\mathbb{Z}^d$ does not require the proposition of the last section. Indeed, the knowledge of what happens for $\mathbb{Z}$ is sufficient to do the work.
Given two graphs G_1 and G_2, we can write the Cartesian product $G_1 \times G_2$ (see [39] for the definition), using the Kronecker product (denoted by $\otimes$, see [39] for the definition) :

$$W^{G_1\times G_2} = I_{G_1}\otimes W^{G_2} + W^{G_1}\otimes I_{G_2}.$$

This gives the spectral decomposition of $W^{G_1\times G_2}$, using the tensor product $\otimes$:

$$W^{G_1\times G_2} = \int_{\mathrm{Sp}(G_1)\times\mathrm{Sp}(G_2)} (x+t)\mathrm{d}E^{G_1}(x)\otimes \mathrm{d}E^{G_2}(t).$$

It gives immediately the spectral decomposition of $\mathbb{Z}^d$:

$$W^{\mathbb{Z}^d} = \int_{[-2,2]^d} (\lambda_1+\lambda_2+\cdots+\lambda_d)\left(\mathrm{d}E^{\mathbb{Z}}\right)^{\otimes d}(\lambda_1,\cdots,\lambda_d).$$

So that, for any $k \geq 0$,

$$\left((W^{\mathbb{Z}^d})^k\right)_{0_{\mathbb{Z}^d}0_{\mathbb{Z}^d}} = \int_{[-2,2]^d} (\lambda_1+\lambda_2+\cdots+\lambda_d)^k \mathrm{d}\mu^{\mathbb{Z}}_{00}(\lambda_1)\cdots\mathrm{d}\mu^{\mathbb{Z}}_{00}(\lambda_d).$$

Furthermore, we can compute $\mu^{\mathbb{Z}^d}_{kl}, k, l \in \mathbb{Z}^d$. Indeed, notice first that, using an automorphism sending (k,l) onto $(0_{\mathbb{Z}^d}, l-k)$, it is sufficient to compute $\mu^{\mathbb{Z}^d}_{0_{\mathbb{Z}^d}k}, k \in \mathbb{Z}^d$. To simplify the notations, we denote

$$\mu^{\mathbb{Z}^d}_k := \mu^{\mathbb{Z}^d}_{0_{\mathbb{Z}^d}k}.$$

As in the previous section, let g be a positive measurable function, analytic over $[-2d, 2d]$, and define

$$\Gamma := \mathcal{K}^{\mathbb{Z}^d}(\psi) = g(W^{\mathbb{Z}^d}).$$

The operator Γ is positive definite. Let $\mathbf{X}$ be a Gaussian process of covariance Γ.

Here again g denotes the spectral density of the process $\mathbf{X}$, in the sense of our framework.
Define futher, for any $k = (k_1, \cdots, k_d) \in \mathbb{Z}^d$,

$$r_k := \Gamma_{0k} = \langle X_0, X_k \rangle.$$

Using the definition of $\mu_k^{\mathbb{Z}^d}$, we have

$$\begin{aligned} r_k &= \int_{[-2d,2d]} g(x) \mathrm{d}\mu_k^{\mathbb{Z}^d} \\ &= \int_{[-2,2]^d} g(\lambda_1 + \lambda_2 + \cdots + \lambda_d) \mathrm{d}\mu_{0k_1}^{\mathbb{Z}}(\lambda_1) \cdots \mathrm{d}\mu_{0k_d}^{\mathbb{Z}}(\lambda_d). \end{aligned}$$

Define ψ by the relation

$$\forall (t_1, \cdots, t_d) \in [-\pi, \pi]^d, \psi(t_1, \cdots, t_d) = g(2\cos(t_1) + \cdots + 2\cos(t_d)).$$

With the change of variables $2\cos(t_i) = \lambda_i, i = 1 \cdots d$, we get the following expression.

Example 2 (The square lattice)
Using the notations introduced before, we obtain

$$r_k = \frac{1}{\pi^d} \int_{[-\pi,\pi]^d} \psi(t_1, \cdots, t_d) \cos(k_1 t_1) \cdots \cos(k_d t_d) \mathrm{d}\lambda(t_1) \cdots \mathrm{d}\lambda(t_d).$$

We recover the usual spectral framework for $\mathbb{Z}^d$ (see for instance [43] and [44]).
Notice also that, we can build anisotropic processes, choosing a function ψ which can not be written as

$$\psi(t_1, \cdots, t_d) = g(\cos(t_1) + \cdots + \cos(t_d)).$$

In particular, for the Kronecker product $\mathbb{Z}^{\otimes d}$, (and it is the case for the Kronecker product of any other graphs) the previous spectral representation holds, with ψ such that, for some function g,

$$\psi(t_1, \cdots, t_d) = g\left(\prod_{i=1}^{d} \cos(t_i)\right).$$

The explicit computation of the global spectral measure $\mu^{\mathbb{Z}^d}$ is not easy. Therefore, the Fourier representation is used instead of the measure $\mu^{\mathbb{Z}^d}$ itself.

The homogeneous tree $G = \mathcal{A}_{q+1}$

The $(q+1)$-homogeneous tree $\mathcal{A}_{q+1}$ is also vertex transitive. So that, as in the case $\mathbf{G} = \mathbb{Z}$, the measure at the root is the same at any vertex, and also equal to the global measure.
First choose a vertex $o \in \mathcal{A}_{q+1}$ as the root. The rooted graph $(\mathcal{A}_{q+1}, o)$ verifies Assumptions 5.1.1, with the sequences

$$\begin{aligned} a_k &= 1, k \geq 1, \\ b_k &= 0, \forall k \geq 0, \\ c_0 &= q+1 \\ c_k &= q, k \geq 1. \end{aligned}$$

So that the measure $\mu_{oo}^{\mathcal{A}_{q+1}}$ is such that the sequence of polynomials $(P_k)_{k\in\mathbb{N}}$ defined below, is an orthogonal family, with respect to this measure :

$$\begin{aligned} P_0(x) &= 1, \\ P_1(x) &= \frac{x}{q+1}, \\ \forall k \geq 1, xP_k(x) &= qP_{k+1} + P_{k-1}. \end{aligned}$$

This sequence of polynomial is, up to a different normalization, the sequence introduced by Cartier, Dunau and Arnaud ([25], [5]).
We can compute the measure $\mu_{oo}^{\mathcal{A}_{q+1}}$, using its Stieltjes transform.
Define the resolvent operator R, for z large enough, by

$$R(z) := (z\,\mathrm{Id} - W^{\mathcal{A}_{q+1}})^{-1}.$$

Then, the Stieltjes transform s of $\mu_{oo}^{\mathcal{A}_{q+1}}$ admits a continued fraction expansion (see for instance [66]) :

$$s(z) = R(z)_{oo} = \frac{1}{z - (q+1)\frac{1}{z - q\frac{1}{z-\cdots}}}.$$

For the regular computation, see [18]. We recall the main steps.
Define

$$\tilde{s}(z) = \frac{1}{z - q\frac{1}{z - q\frac{1}{z-\cdots}}}.$$

We have

$$\tilde{s}(z) = \frac{1}{z - q\tilde{s}(z)}.$$

Hence, we have (see for instance [18])

$$\tilde{s}(z) = \frac{z - \sqrt{z^2 - 4q}}{2q}.$$

Thus,

$$\begin{aligned} s(z) &= \frac{1}{z-(q+1)\tilde{s}(z)} \\ &= \frac{2q}{(q-1)z+(q+1)\sqrt{z^2-4q}} \\ &= -\frac{1}{2}\frac{-(q+1)z+(q+1)\sqrt{z^2-4q}}{(q+1)^2-z^2}. \end{aligned}$$

Using the Stieltjes inversion formula (see [18]), we get the absolute continuity of $\mu_{oo}^{\mathcal{A}_{q+1}}$ with respect to the Lebesgue measure, with the following density :

$$f(x) = \frac{(q+1)}{2\pi}\frac{\sqrt{4q-x^2}}{(q+1)^2-x^2}\mathbb{1}_{[-2\sqrt{q},2\sqrt{q}]}.$$

This formula is due to McKay [54] and Kesten [47].
Now, let g be a positive measurable function, analytic over $[-2\sqrt{q},2\sqrt{q}]$, and define

$$\Gamma := \mathcal{K}^{\mathcal{A}_{q+1}}(\psi) = g(W^{\mathcal{A}_{q+1}}).$$

The operator Γ is positive definite. Let $\mathbf{X}$ be a Gaussian process of covariance Γ. Here again g denotes the spectral density of the process $\mathbf{X}$, in the sense of our framework.
Define futher, for any $i,j \in \mathcal{A}_{q+1}$,

$$r_{d(i,j)} := \Gamma_{ij} = \langle X_i, X_j \rangle.$$

Using the definition of $\mu_{ij}^{\mathbb{Z}^d}$, we have

$$\begin{aligned} r_{d(i,j)} &= \int_{[-2\sqrt{q},2\sqrt{q}]} g(x)\mathrm{d}\mu_{ij}^{\mathcal{A}_{q+1}}(x) \\ &= \int_{[-2\sqrt{q},2\sqrt{q}]} g(x)P_{d(i,j)}(x)\mathrm{d}\mu_{oo}^{\mathcal{A}_{q+1}}(x) \end{aligned}$$

Example 3 (The homogeneous tree)
Using the notations introduced before, we obtain

$$r_k = \int_{[-2\sqrt{q},2\sqrt{q}]} g(x)P_k(x)f(x)\mathrm{d}\lambda(x).$$

We recover the classical representation of the covariance of a stationary process indexed by a homogeneous tree.

In the paper of Arnaud [5], they also obtain a spectral representation for the process himself. This representation relies on a factorization of the polynomials $P_n, n \in \mathbb{Z}$. Furthermore, this representation is linked to the ends of the tree.
The same representation may be obtained, using tools of harmonic analysis, computing the Green function and the Poisson kernel on a tree. This leads to the same factorization, and the explicit formula depends on probability of returns of the isotropic random walk on the tree. These two approachs are equivalent since the transition kernel P of the random walk is colinear with the adjacency operator W.
We do not give further details here, since it is far away from our purpose, but one may refer to [53], [24]...
We can now deal with the more general example of distance-transitives graphs.

Distance-transitive graphs

Recall that the graph G is said distance-transitive if for any $i, j, k, l \in G$ such that $d(i, j) = d(k, l)$, there exists an automorphism sending i on k and j on l.
Infinite distance-transitive graphs with bounded degree are of the following form (see for instance [65]) : there exists two integers $a, b \in \mathbb{N}^*$ such that every vertex is in a b-clique, and these cliques are assembled in a a-regular tree-like way (each vertex is in a different copies of complete graphs of b vertices.)
We will denote this graph by $\mathcal{DT}_{a,b}$. See for instance figure 5.2 for the construction of $\mathcal{DT}_{2,3}$.
Note that, thanks to the definition of $\mathcal{A}_q$,

$$\mathcal{DT}_{a,2} = \mathcal{A}_a.$$

Now choose a vertex $o \in \mathcal{DT}_{a,b}$ as the root. Here again, the measure at the root is equal to the global specral measure μ (thanks to the vertex transitivity).
The rooted graph $(\mathcal{DT}_{a,b}, o)$ verifies Assumption 5.1.1, with the sequences

$$\begin{aligned} a_k &= 1, k \geq 1, \\ b_k &= b - 2, \forall k \geq 1, \\ c_0 &= a(b-1), \\ c_k &= (a-1)(b-1), k \geq 1. \end{aligned}$$

Define the sequence of polynomials $(P_k)_{k\in\mathbb{N}}$ by

$$\begin{aligned} P_0(x) &= 1, \\ P_1(x) &= \frac{x}{a(b-1)}, \\ \forall k \geq 1, xP_k(x) &= (a-1)(b-1)P_{k+1}(x) + (b-2)P_k(x) + P_{k-1}(x). \end{aligned}$$

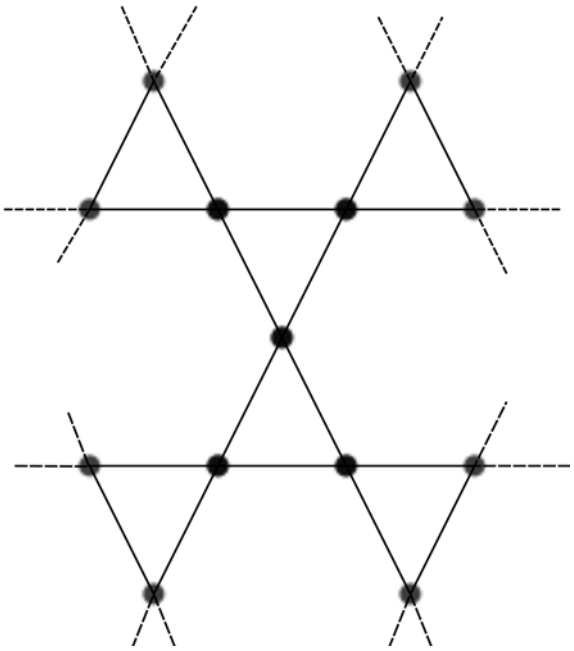


FIGURE 5.2: The graph $\mathcal{DT}_{2,3}$

Therefore, thanks to Proposition 5.1.2, the family of polynomials $(P_k)_{k\in\mathbb{N}}$ is orthogonal with respect to the measure at the root $\mu_{oo}^{\mathcal{DT}_{a,b}}$. This polynomials appears in [65] and [45] under the name of Cartier polynomials.
The following presentation follows the ideas of Voit [65] and Heyer [45]. However, our approach is again inspired by the one which appears in the survey of Mohar and Woess [55].
Again, using the Stieltjes transform, we may compute the resolvent

$$R(z) := (z\,\mathrm{Id} - W^{\mathcal{DT}_{a,b}})^{-1},$$

using

$$R(z)_{00} = \frac{1}{z - a(b-1)\frac{1}{z-(b-2)-(a-1)(b-1)\frac{1}{z-\cdots}}}.$$

Define

$$\tilde{s}(z) = \frac{1}{z-(b-2)-(a-1)(b-1)\frac{1}{z-(b-2)-(a-1)(b-1)\frac{1}{z-\cdots}}},$$

we obtain

$$\tilde{s}(z) = \frac{(z-b+2) - \sqrt{(z-b+2)^2 - 4(a-1)(b-1)}}{2(a-1)(b-1)}.$$

Further,

$$R(z)_{00} = \frac{1}{z - (z-(b-2))\frac{a}{2(a-1)} - \frac{a\sqrt{(z-(b-2))^2-4(a-1)(b-1)}}{2(a-1)}}.$$
$$= \frac{2(a-1)}{(a-2)z - a(b-2) - a\sqrt{(z-(b-2))^2 - 4(a-1)(b-1)}}.$$

Denote

$$I_{a,b} = [(b-2) - 2\sqrt{(a-1)(b-1)}, (b-2) + 2\sqrt{(a-1)(b-1)}].$$

Using again a Stieltjes inversion formula, we obtain (see [55] for the result, and [7] for the computation),

$$\mathrm{d}\mu_{oo}^{\mathcal{DT}_{a,b}}(x) = \frac{a}{2\pi}\frac{\sqrt{4(a-1)(b-1) - (x-(b-2))^2}}{(a(b-1)-x)(a+x)}\mathrm{d}\lambda_{I_{a,b}},$$
$$\text{if } a \geq b$$

$$\mathrm{d}\mu_{oo}^{\mathcal{DT}_{a,b}}(x) = \frac{a}{2\pi}\frac{\sqrt{4(a-1)(b-1) - (x-(b-2))^2}}{(a(b-1)-x)(a+x)}\mathrm{d}\lambda_{I_{a,b}} + \frac{b-a}{b}\delta_{-a}(x),$$
$$\text{if } a < b$$

Note that, when $b = 2$, we recover the example of a homogeneous tree.
As above, let $a, b \in \mathbb{N}$, and g be a positive measurable function, analytic over

$$[\min\left((b-2) - 2\sqrt{(a-1)(b-1)}, -a\right), (b-2) + 2\sqrt{(a-1)(b-1)}],$$

and define

$$\Gamma := \mathcal{K}^{\mathcal{DT}_{a,b}}(\psi) = g(W^{\mathcal{DT}_{a,b}}).$$

The operator Γ is positive definite. Let $\mathbf{X}$ be a Gaussian process of covariance Γ. Define futher, for any $i, j \in \mathcal{DT}_{a,b}$,

$$r_{d(i,j)} := \Gamma_{ij} = \langle X_i, X_j \rangle.$$

Using the definition of $\mu_{ij}^{\mathbb{Z}^d}$, we have, if $k = d(i,j)$

$$r_k = \int_{[-2\sqrt{q},2\sqrt{q}]} g(x)\mathrm{d}\mu_{ij}^{\mathcal{DT}_{a,b}}(x)$$

Example 4

Using the notations introduced before, we obtain

$$r_k = \int_{[-2\sqrt{q},2\sqrt{q}]} g(x) P_{d(i,j)}(x) \mathrm{d}\mu_{oo}^{\mathcal{DT}^{a,b}}(x).$$

We recover the classical representation of the covariance of a stationary process indexed by a distance-transitive graph (see for instance [45]).

Semi-homogeneous tree

Finally, we deal with the semi-homegeneous tree $\mathcal{A}_{a,b}$, in order to show what happens in the case of a non vertex-transitive graph. Define the semi-homogeneous tree as the only connected graph without cycles such that
– Each vertex has degree a or b,
– All neighbors have different degree.
This graph have been studied by many authors [21], [37]...
If we choose a vertex o_a of degree a, and a vertex o_b of degree b, then both $(\mathcal{A}_{a,b}, o_a)$ and $(\mathcal{A}_{a,b}, o_b)$ verify Assumption 5.1.1.
Proposition 5.1.2 may be applied to $(\mathcal{A}_{a,b}, o_a)$ with the sequences

$$\begin{aligned} a_k &= 1, k \geq 1, \\ b_k &= 0, \forall k \geq 0, \\ c_0 &= a \\ c_{2k+1} &= b-1, k \geq 0. \\ c_{2k} &= a-1, k \geq 0. \end{aligned}$$

Define a family $(P_k)_{k\in\mathbb{N}}$ of polynomials by

$$\begin{aligned} P_0(x) &= 1, \\ P_1(x) &= \frac{x}{a}, \\ \forall k \geq 1, xP_{2k}(x) &= (a-1)P_{2k+1} + P_{k-1}, \\ \forall k \geq 1, xP_{2k-1}(x) &= (b-1)P_{2k} + P_{k-1}, \end{aligned}$$

Then the spectral local measure $\mu_{o_a o_a}^{\mathcal{A}_{a,b}}$ at the root is orthogonal with respect to this family of polynomials. Permutting a, b, we obtain the same result for $\mu_{o_b o_b}^{\mathcal{A}_{a,b}}$ with a

family $(Q_k)_{k\in\mathbb{N}}$ defined by

$$\begin{aligned} Q_0(x) &= 1, \\ Q_1(x) &= \frac{x}{b}, \\ \forall k \geq 1, xQ_{2k}(x) &= (a-1)Q_{2k+1} + Q_{k-1}, \\ \forall k \geq 1, xQ_{2k-1}(x) &= (b-1)Q_{2k} + Q_{k-1}, \end{aligned}$$

Actually, we have

$$\forall k \geq 1, Q_k(x) = \frac{x}{b} P_{k-1}(x).$$

Then, to compute the measure $\mu^{\mathcal{A}_{a,b}}_{o_a o_a}$, we have to note that

$$\forall k \geq 0, \left(W^{\mathcal{A}_{a,b}}\right)^{2k}_{o_a o_a} = \left(W^{\mathcal{DT}_{a,b}} + a\,\mathrm{Id}\right)^{k}_{oo}.$$

Indeed, this equality may be proven by induction, or undertood as shown by the construction of figure 5.2.

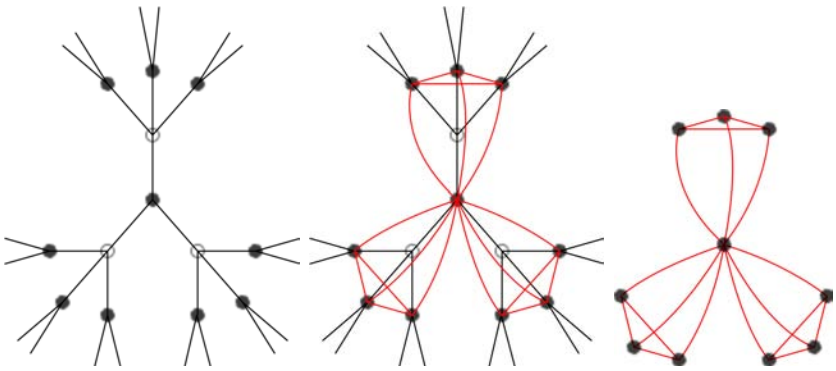


FIGURE 5.3: Equivalence between semi-homogeneous tree and distance-transitive graphs

Furthermore, since $\mathcal{A}_{a,b}$ is a bipartite graph, we have

$$\left(W^{\mathcal{A}_{a,b}}\right)^{2k+1}_{o_a o_a} = 0.$$

Hence, we get that the spectral measure $\mu^{\mathcal{A}_{a,b}}_{o_a o_a}$ is the only symmetric measure on $\mathbb{R}$ which verifies :

$$\forall k \geq 0, \int_{\mathbb{R}} x^{2k} \mathrm{d}\mu^{\mathcal{A}_{a,b}}_{o_a o_a}(x) = \int_{\mathbb{R}} (y+a)^k \mathrm{d}\mu^{\mathcal{DT}_{a,b}}_{o_a o_a}(y).$$

Denote

$$I^{(2)}_{a,b} := [\left|\sqrt{a-1} - \sqrt{b-1}\right|, \sqrt{a-1} + \sqrt{b-1}]$$

With the change of variables $y = x^2 - a$, and a few calculations (using the fact that the measure is symmetric), we get (see [40])

$$\mathrm{d}\mu^{\mathcal{A}_{a,b}}_{o_a o_a}(x) = \frac{a}{\pi}\frac{\sqrt{4(a-1)(b-1) - \Big(x^2 - \Big((a-1)+(b-1)\Big)\Big)^2}}{\Big(ab - x^2\Big)\,|x|}\mathbb{1}_{|x|\in I^{(2)}_{a,b}}\mathrm{d}\lambda(x),$$

if $a \geq b$

$$\mathrm{d}\mu^{\mathcal{A}_{a,b}}_{o_a o_a}(x) = \frac{a}{\pi}\frac{\sqrt{4(a-1)(b-1) - \Big(x^2 - \Big((a-1)+(b-1)\Big)\Big)^2}}{\Big(ab - x^2\Big)\,|x|}\mathbb{1}_{|x|\in I^{(2)}_{a,b}}\mathrm{d}\lambda(x) + \frac{b-a}{2b}\delta_0(x),$$

if $a < b$

Of course, we can do the same computation for $\mu^{\mathcal{A}_{a,b}}_{o_b o_b}$, getting the same result. Note that only the dirac mass and a multiplicative constant changes between this two cases.
Note also that, when $a = b$, we obtain

$$\begin{aligned}\mathrm{d}\mu^{\mathcal{A}_{a,a}}_{o_a o_a}(x) &= \frac{a}{\pi}\frac{\sqrt{4(a-1)^2 - \big(x^2 - 2(a-1)\big)^2}}{\Big(a^2 - x^2\Big)\,|x|}\mathrm{d}\lambda_{[-2\sqrt{a-1},2\sqrt{a-1}]},\\ &= \frac{a}{\pi}\frac{\sqrt{-x^4 + 4(a-1)x^2}}{\Big(a^2 - x^2\Big)\,|x|}\mathrm{d}\lambda_{[-2\sqrt{a-1},2\sqrt{a-1}]}.\end{aligned}$$

We recover, as whished, the spectral measure of an homogeneous tree.
As above, let now $a, b \in \mathbb{N}$, and g be a positive measurable function, analytic over

$$[-\left(\sqrt{a-1} + \sqrt{b-1}\right), \sqrt{a-1} + \sqrt{b-1}],$$

and define

$$\Gamma := \mathcal{K}^{\mathcal{A}_{a,b}}(\psi) = g(W^{\mathcal{A}_{a,b}}).$$

The operator Γ is positive definite. Let $\mathbf{X}$ be a Gaussian process of covariance Γ.

Example 5

Using the notations introduced before, we obtain

$$\begin{aligned}\Gamma_{i,j} &= \int_{[-2\sqrt{q},2\sqrt{q}]} g(x)P_{d(i,j)}(x)\mathrm{d}\mu^{\mathcal{A}_{a,b}}_{o_a o_a}(x), \ \textit{if } i \textit{ or } j = a,\\ &= \int_{[-2\sqrt{q},2\sqrt{q}]} g(x)P_{d(i,j)}(x)\mathrm{d}\mu^{\mathcal{A}_{a,b}}_{o_b o_b}(x), \ \textit{otherwise.}\end{aligned}$$

We recover the classical representation of the covariance of a stationary process indexed by a distance-transitive graph (see for instance [45]).
Finally, the global measure $\mu^{\mathcal{A}_{a,b}}$ may be easily obained for any sequences of subgraphs $(G_n))_{n\in\mathbb{N}}$. However, it depends on the choice of this sequence.
With the notations used by Mohar in [40], if c_a, c_b denotes the asymptotic proportion of vertices of degree a (resp. b) in G_n, then we have

$$\mu^{\mathcal{A}_{a,b}} = c_a\mu^{\mathcal{A}_{a,b}}_{o_a o_a} + c_b\mu^{\mathcal{A}_{a,b}}_{o_b o_b}.$$

We stop here the examples, but we could study also a lot of other graphs (for instance a lot of particular lattices). Finally, note that, except for the semi-homogeneous tree, all the measures obtained in this section belongs to the class of Free-Meixner measures, since the sequences $(a_k)_{k\in\mathbb{N}}, (b_k)_{k\in\mathbb{N}}, (c_k)_{k\in\mathbb{N}}$ are constant (see for instance [22]).

5.3 Stationary random graphs and global measure

In this last section, we recall some properties about the convergence of the spectral distribution of a finite graph to the global spectral measure. This will be done with two different approachs, which provides the same results. In particular, this gives a way to understand the convergence of the spectrum of the adjacency operator of a random regular graph to the measure of a homogeneous tree.
The first approach appears in [55]. It considers directly the convergence of the identity resolution. The second is due to Aldous and Lyons [1], and shows the convergence of the spectral measure as a consequence of a local convergence.
Actually, both approachs relies on local convergence of graphs. For instance, in the case of convergence of random regular graph to the homogeneous tree, an assumption under the number of small loops is needed (see for instance [55]). Assuming that there is, with a large probability, a small number of short loops, is a way to say that *locally*, the graph is isomorphic to a tree.

Convergence of the identity resolution

First consider a sequence of graphs $(\mathbf{G}_n)$ such that $\mathbf{G}_n = \left(G, W^{(n)}\right)$ (without loss of generality, we may assume that the graphs have the same vertex set).
In [40], a proposition states that if $\left(W^{(n)}\right)$ converges weakly to W, that is, for any $x \in l^2(G)$,

$$W^{(n)}x \to Wx,$$

then the associated identity resolution $\left(E^{(n)}(\lambda)\right)_{n\in\mathbb{N}}$ converges to $E(\lambda)$ for every λ where E is continuous.
Further, the weak convergence of $\left(W^{(n)}\right)$ to W implies the weak convergence of the sequence of measures $\left(\mu_{ij}^{(n)}\right)_{n\in\mathbb{N}}, i, j \in G$ to μ_{ij} on the set of continuity of E.
Consider now a sequence of q-regular random graphs $(\mathbf{G}_n)$ such that the number of small loops decrease to 0 a.s. as n tends to infinity (for the explicit assumption, see [54]). Then, it appears that the spectral measure $\left(\mu_{ij}^{(n)}\right)_{n\in\mathbb{N}}, ij \in G$ converges weakly (and in probability) to the spectral measure μ_{ij} of a q-regular tree, since we can prove the weak convergence of the adjcency operator of $\mathbf{G}_n$ to the adjacency operator W of $\mathcal{A}_q$. The assumption under the number of loops holds for instance for Erdös-Rényi graphs conditionned to be q-regular.
Actually, the same kind of results can be proven for any quasi-transitive graphs (i.e. graphs such that their quotient with their automophism group is finite).
The assumption under the weak convergence of the adjacency operator is also satisfied by a sequence of subgraphs of $\mathbf{G}$. For instance, let us take a vertex $o \in G$ and $\mathbf{G}_n$ as the ball of radius n and center o. Then the spectral measure $\mu_{oo}^{(n)}$ converges weakly to μ_{oo}.

Weak local convergence of the graph

Another point of view has been developped for random graphs. It deals with the local convergence of random rooted graphs. The usual framework and notations appear for instance in [1] and [11]. Denote by $\mathcal{G}_\bullet$ the set of isomorphism classes of locally finite ($\deg(G) < +\infty$) rooted graphs endowed with the distance $\mathbf{d}$ defined by $\mathbf{d}\Big((\mathbf{G}, o), (\mathbf{G}', o')\Big) = \frac{1}{r+1}$, where r is the largest integer such that the ball of center o and radius r in $\mathbf{G}$ is isomorphic to the ball of center o' and radius r in $\mathbf{G}'$. This define a metric on $\mathcal{G}_\bullet$. This metric space is then Polish set, and we can study convergence properties in this metric.
Actually, when the limit of the sequence of rooted graph has no automorphisms, the induced topology is equivalent to the topology given for $\mathbf{G}$ by the last notion of convergence. This allows us to recover some results.

In [1], the authors consider random weak limits of finite graphs by rooting uniformly at random the graph G_n. In the examples considered above, the limit measure ν on $\mathcal{G}_\bullet$ is atomic. Indeed, if $(G_n)_{n\in\mathbb{N}}$ denotes a sequence of q-regular graphs with no small loops, then rooting this graph at random provides the weak convergence of this sequence to the homogeneous tree $\mathcal{A}_q$, in the sense of [1]. The limit measure verifies, for any $o \in \mathcal{A}_q$, $\nu(\mathcal{A}_q, o) = 1$.
For our purpose, instead of looking to

$$\nu^{(n)} = \frac{1}{\sharp G_n} \sum_{o\in G_n} \delta_{(G_n,o)},$$

we should consider the measure

$$\tilde{\nu}^{(n)} = \frac{1}{\sharp G_n} \sum_{o\in G_n} \delta_{(G,o)},$$

where $(G_n)_{n\in\mathbb{N}}$ is a sequence of growing subgraphs of G.
The difference here is essential. Indeed, choose for instance $(G_n)_{n\in\mathbb{N}}$ as a sequence of growing connected subgraphs of the q-homogeneous tree $\mathcal{A}_q$. In this case, the measure $\nu^{(n)}$ converges weakly, but not to the Dirac distribution on $(\mathcal{A}_q, o), o \in \mathcal{A}_q$. Indeed, there is some edge effects due to the exponential growth. However, the measure $\tilde{\nu}^{(n)}$ converge to $\delta_{(G,o)}$.
Now, the convergence of the trace measure (which appears both in [1] and [18]) is a consequence of the random weak convergence of the rooted graphs.
Indeed, if $\tilde{\nu}^{(n)}$ converges to $\tilde{\nu}$, then $\mu^{(n)} = \frac{1}{\sharp G_n} \sum_{o\in G_n} \mu_{oo}$ converges to μ, and we have

$$\mu = \mathbb{E}_{\tilde{\nu}}[\mu_{oo}^{(G)}].$$

We do not gives further details here, but we need to keep in mind that this notion of local convergence, and associated definitions of stationary sequences of random graphs [11] or unimodular graphs [1] may be a good framework to study spectral properties of the limit graph. Furthermore, it allows us to deal with random sequences of finite graphs.
Finally, in the framework of road traffic, we just state a combinatorial assumption about the sequence of subgraphs, that ensures the convergence of the spectral measure.
We first define, for any $l \geq 0$, the l-type of a vertex $k \in G$ by

$$t_l(k) = W_{kk}^l.$$

The l-type of a vertex is the number of loops around k, counted with their weights. Define also a subset U_l of $\mathbb{R}$ as

$$U_l := \{t_l(k), k \in G\}.$$

This set gives all the possible values among the vertices $k \in G$ of the l-type of k. Notice first that when the entries of W take their values in a finite set, U_l is finite for any $l \geq 0$.

Assumption 5.3.1 (Homogeneity assumption) *The sequence $(G_n)_{n\in\mathbb{N}}$ and the operator W satisfy :*

$$\forall l \geq 0, \forall n \geq 1, \forall v \in U_l, \frac{\sharp\{j \in G_n, t_l(j) = v\}}{\sharp G_n} \underset{n\to\infty}{\to} p_v^{(l)}.$$

Remark : This last assumption may be understood as an homogeneity hypothesis on the frequency of the type for the vertices of G_n. Figure 5.4 shows an example of admissible sequences of subgraphs.

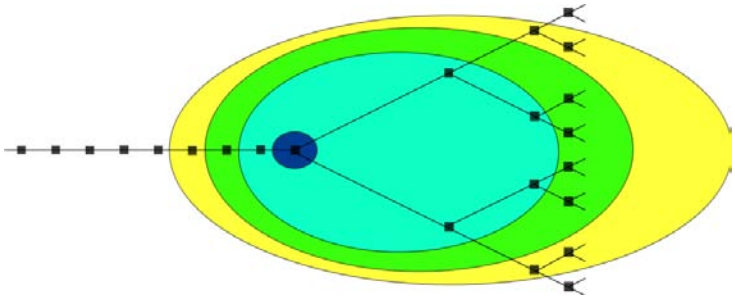


FIGURE 5.4: Example of admissible sequences of subgraphs.

Proposition 5.3.2 *Under Assumption 5.3.1, the spectral measure is convergent. That is, there exists μ with*

$$\frac{1}{\sharp G_n} \sum_{i\in G_n} \mu_{ii} \underset{n\to\infty}{\overset{\mathcal{D}}{\to}} \mu,$$

Proof. Under homogeneity assumption, we have

$$\frac{1}{\sharp G_n} \mathrm{Tr}\left((W^l)_{G_n}\right) \underset{n\to\infty}{\to} \sum_{v\in U_l} p_v^{(l)} v.$$

Define

$$\mu^{(l)} := \sum_{v\in U_l} p_v^{(l)} v.$$

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

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

So that

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

And

$$\mu^{(l)} \leq 1.$$

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

Bibliographie

[1] D. Aldous and R. Lyons. Processes on unimodular random networks. *Electron. J. Probab.*, 12 :no. 54, 1454–1508, 2007.

[2] A. Antoniadis. Statistics on Banach space valued Gaussian random variables. In *Probability in Banach spaces, III (Medford, Mass., 1980)*, volume 860 of *Lecture Notes in Math.*, pages 1–8. Springer, Berlin, 1981.

[3] A. Antoniadis, E. Paparoditis, and T. Sapatinas. A functional wavelet-kernel approach for time series prediction. *J. R. Stat. Soc. Ser. B Stat. Methodol.*, 68(5) :837–857, 2006.

[4] A. Antoniadis, E. Paparoditis, and T. Sapatinas. Bandwidth selection for functional time series prediction. *Statist. Probab. Lett.*, 79(6) :733–740, 2009.

[5] J.-P. Arnaud. Fonctions sphériques et fonctions définies positives sur l'arbre homogène. *C. R. Acad. Sci. Paris Sér. A-B*, 290(2) :A99–A101, 1980.

[6] N. Aronszajn. Theory of reproducing kernels. *Trans. Amer. Math. Soc.*, 68 :337–404, 1950.

[7] R. Askey and J. Wilson. Some basic hypergeometric orthogonal polynomials that generalize Jacobi polynomials. *Mem. Amer. Math. Soc.*, 54(319) :iv+55, 1985.

[8] R. Azencott and D. Dacunha-Castelle. *Series of irregular observations.* Applied Probability. A Series of the Applied Probability Trust. Springer-Verlag, New York, 1986. Forecasting and model building.

[9] J.-M. Bardet and P. R. Bertrand. A non-parametric estimator of the spectral density of a continuous-time Gaussian process observed at random times. *Scand. J. Stat.*, 37(3) :458–476, 2010.

[10] J.-M. Bardet, P. Doukhan, and J. R. León. Uniform limit theorems for the integrated periodogram of weakly dependent time series and their applications to Whittle's estimate. *J. Time Ser. Anal.*, 29(5) :906–945, 2008.

[11] I. Benjamini and N. Curien. Ergodic theory on stationary random graphs. Technical Report arXiv :1011.2526, Nov 2010.

[12] P. J. Bickel and E. Levina. Covariance regularization by thresholding. *Ann. Statist.*, 36(6) :2577–2604, 2008.

[13] P. J. Bickel and E. Levina. Regularized estimation of large covariance matrices. *Ann. Statist.*, 36(1) :199–227, 2008.

[14] N. Biggs. *Algebraic graph theory*. Cambridge Mathematical Library. Cambridge University Press, Cambridge, second edition, 1993.

[15] J. Bigot, R. Biscay, J.-M. Loubes, and L. Muniz Alvarez. Nonparametric estimation of covariance functions by model selection, 2009.

[16] P. Bondon. Prediction with incomplete past of a stationary process. *Stochastic Process. Appl.*, 98(1) :67–76, 2002.

[17] P. Bondon. Influence of missing values on the prediction of a stationary time series. *J. Time Ser. Anal.*, 26(4) :519–525, 2005.

[18] C. Bordenave and M. Lelarge. Resolvent of large random graphs. *Random Structures Algorithms*, 37(3) :332–352, 2010.

[19] C. Bordenave, M. Lelarge, and J. Salez. The rank of diluted random graphs, 2009.

[20] A. Böttcher and B. Silbermann. *Analysis of Toeplitz operators*. Springer Monographs in Mathematics. Springer-Verlag, Berlin, second edition, 2006. Prepared jointly with Alexei Karlovich.

[21] F. Bouaziz-Kellil. Représentations sphériques des groupes agissant transitivement sur un arbre semi-homogène. *Bull. Soc. Math. France*, 116(3) :255–278, 1988.

[22] M. Bozejko and W. Bryc. On a class of free Lévy laws related to a regression problem. *J. Funct. Anal.*, 236(1) :59–77, 2006.

[23] P. J. Brockwell and R. A. Davis. *Introduction to time series and forecasting*. Springer Texts in Statistics. Springer-Verlag, New York, second edition, 2002. With 1 CD-ROM (Windows).

[24] R. Burioni and D. Cassi. Random walks on graphs : ideas, techniques and results. *J. Phys. A*, 38(8) :R45–R78, 2005.

[25] P. Cartier. Harmonic analysis on trees. In *Harmonic analysis on homogeneous spaces (Proc. Sympos. Pure Math., Vol. XXVI, Williams Coll., Williamstown, Mass., 1972)*, pages 419–424. Amer. Math. Soc., Providence, R.I., 1973.

[26] F. R. K. Chung. *Spectral graph theory*, volume 92 of *CBMS Regional Conference Series in Mathematics*. Published for the Conference Board of the Mathematical Sciences, Washington, DC, 1997.

[27] S. Cohen, C. Lacaux, and M. Ledoux. A general framework for simulation of fractional fields. *Stochastic Process. Appl.*, 118(9) :1489–1517, 2008.

[28] F. Comte. Adaptive estimation of the spectrum of a stationary Gaussian sequence. *Bernoulli*, 7(2) :267–298, 2001.

[29] R. Dahlhaus. Approximations for the inverse of Toeplitz matrices with applications to stationary processes. *Linear Algebra Appl.*, 127 :27–40, 1990.

[30] R. Dahlhaus and H. Künsch. Edge effects and efficient parameter estimation for stationary random fields. *Biometrika*, 74(4) :877–882, 1987.

[31] R. Dahlhaus and W. Polonik. Nonparametric quasi-maximum likelihood estimation for Gaussian locally stationary processes. *Ann. Statist.*, 34(6) :2790–2824, 2006.

[32] 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.

[33] A. Dembo and O. Zeitouni. *Large deviations techniques and applications*, volume 38 of *Stochastic Modelling and Applied Probability*. Springer-Verlag, Berlin, 2010. Corrected reprint of the second (1998) edition.

[34] H. W. Engl, M. Hanke, and A. Neubauer. *Regularization of inverse problems*, volume 375 of *Mathematics and its Applications*. Kluwer Academic Publishers Group, Dordrecht, 1996.

[35] T. Espinasse, F. Gamboa, and J. Loubes. Estimation error for blind gaussian time series prediction. *Mathematical Methods of Statistics*, 20 :206–223, 2011. 10.3103/S1066530711030033.

[36] F. Gamboa, J.-M. Loubes, and E. Maza. Semi-parametric estimation of shifts. *Electron. J. Stat.*, 1 :616–640, 2007.

[37] P. Gérardin and K. F. Lai. Asymptotic behaviour of eigenfunctions on semi-homogeneous tree. *Pacific J. Math.*, 196(2) :415–427, 2000.

[38] L. Giraitis and P. M. Robinson. Whittle estimation of arch models. *Econometric Theory*, 17(03) :608–631, June 2001.

[39] C. Godsil and G. Royle. *Algebraic graph theory*, volume 207 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, 2001.

[40] C. D. Godsil and B. Mohar. Walk generating functions and spectral measures of infinite graphs. In *Proceedings of the Victoria Conference on Combinatorial Matrix Analysis (Victoria, BC, 1987)*, volume 107, pages 191–206, 1988.

[41] U. Grenander and G. Szegő. *Toeplitz forms and their applications*. Chelsea Publishing Co., New York, second edition, 1984.

[42] X. Guyon. Parameter estimation for a stationary process on a d-dimensional lattice. *Biometrika*, 69(1) :95–105, 1982.

[43] X. Guyon. *Champs aléatoires sur un réseau*. Masson, 1992.

[44] X. Guyon. Modelling for spatial statistics. *Investigación Oper.*, 31(1) :1–33, 2010.

[45] H. 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.

[46] A. Hora and N. Obata. Asymptotic spectral analysis of growing regular graphs. *Trans. Amer. Math. Soc.*, 360(2) :899–923, 2008.

[47] H. Kesten. Symmetric random walks on groups. *Trans. Amer. Math. Soc.*, 92 :336–354, 1959.

[48] 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.

[49] B. Laurent and P. Massart. Adaptive estimation of a quadratic functional by model selection. *Ann. Statist.*, 28(5) :1302–1338, 2000.

[50] G. Letac. Problèmes classiques de probabilité sur un couple de Gel′fand. In *Analytical methods in probability theory (Oberwolfach, 1980)*, volume 861 of *Lecture Notes in Math.*, pages 93–120. Springer, Berlin, 1981.

[51] S. Liang, N. Obata, and S. Takahashi. Asymptotic spectral analysis of generalized erdös-rényi random graphs. In *Noncommutative harmonic analysis with applications to probability*, volume 78 of *Banach Center Publ.*, pages 211–229. Polish Acad. Sci. Inst. Math., Warsaw, 2007.

[52] J.-M. Loubes, E. Maza, M. Lavielle, and L. Rodríguez. Road trafficking description and short term travel time forecasting, with a classification method. *Canad. J. Statist.*, 34(3) :475–491, 2006.

[53] L. Lovász. Random walks on graphs : a survey. In *Combinatorics, Paul Erdős is eighty, Vol. 2 (Keszthely, 1993)*, volume 2 of *Bolyai Soc. Math. Stud.*, pages 353–397. János Bolyai Math. Soc., Budapest, 1996.

[54] B. D. McKay. The expected eigenvalue distribution of a large regular graph. *Linear Algebra Appl.*, 40 :203–216, 1981.

[55] B. Mohar and W. Woess. A survey on spectra of infinite graphs. *Bull. London Math. Soc.*, 21(3) :209–234, 1989.

[56] C. Pittet. On the isoperimetry of graphs with many ends. *Colloq. Math.*, 78(2) :307–318, 1998.

[57] P. M. Robinson. Multiple local whittle estimation in stationary systems. *Annals of Statistics*, 36 :2508–2530, 2008.

[58] W. Rudin. *Functional analysis*. International Series in Pure and Applied Mathematics. McGraw-Hill Inc., New York, second edition, 1991.

[59] G. A. F. Seber. *A matrix handbook for statisticians*. Wiley Series in Probability and Statistics. Wiley-Interscience [John Wiley & Sons], Hoboken, NJ, 2008.

[60] M. L. Stein. *Interpolation of spatial data*. Springer Series in Statistics. Springer-Verlag, New York, 1999. Some theory for Kriging.

[61] M. L. Stein and M. S. Handcock. Some asymptotic properties of kriging when the covariance function is misspecified. *Math. Geol.*, 21(2) :171–190, 1989.

[62] A. N. Tikhonov. On the stability of inverse problems. *C. R. (Doklady) Acad. Sci. URSS (N.S.)*, 39 :176–179, 1943.

[63] N. Verzelen. Adaptive estimation of stationary Gaussian fields. *Ann. Statist.*, 38(3) :1363–1402, 2010.

[64] N. Verzelen and F. Villers. Tests for Gaussian graphical models. *Comput. Statist. Data Anal.*, 53(5) :1894–1905, 2009.

[65] M. Voit. A product formula for orthogonal polynomials associated with infinite distance-transitive graphs. *J. Approx. Theory*, 120(2) :337–354, 2003.

[66] H. S. Wall. *Analytic Theory of Continued Fractions*. D. Van Nostrand Company, Inc., New York, N. Y., 1948.

[67] P. Whittle. The analysis of multiple stationary time series. *J. Roy. Statist. Soc. Ser. B.*, 15 :125–139, 1953.

[68] P. Whittle. Estimation and information in stationary time series. *Ark. Mat.*, 2 :423–434, 1953.

[69] F. Zhang, editor. *The Schur complement and its applications*, volume 4 of *Numerical Methods and Algorithms*. Springer-Verlag, New York, 2005.

Champs et processus gaussiens indexés par des graphes, estimation et prédiction

L'objet de cette thèse est l'étude de processus gaussiens indexés par des graphes. Le but est de fournir des outils pour la modélisation, l'estimation, et la prédiction de tels champs ou processus, utilisant fortement la structure du graphe. Dans un premier travail, nous nous intéressons au problème de prédiction *aveugle* de séries chronologiques et montrons que le biais de l'erreur de prédiction décroît à une vitesse qui dépend de la régularité de la densité spectrale, sous une hypothèse de courte mémoire.
Nous utilisons ensuite la structure spectrale du graphe pour proposer des modèles de covariance pour des champs gaussiens indexés par ce graphe. Cela fournit immédiatement une représentation spectrale, qui permet d'étendre l'approximation de Whittle et l'estimation par quasi-maximum de vraissemblance à ce cadre.
Enfin, cette construction et le lemme de Szegö peuvent être étendus au cas spatio-temporel. Cela permet de mettre en pratique la théorie sur des données réelles.

Gaussian fields and processes indexed by graphs, estimation and prediction

In this work, we study Gaussian processes indexed by graphs. We aim at providing tools for modelisation, estimation, and prediction, that uses the structure of the underlying graphs. In the first Chapter, we deal with the *blind* prediction problem, and compute, in the case of short range dependancy, the rate of convergence of the bias in the prediction error. This rate depends on the regularity of the spectral density of the process.
Then, we use the eigenstructure of the adjacency operator of a graph to propose some models for covariance operators of Gaussian fields indexed by this graph. It leads to a spectral representation for this operator, that can be used to extend Whittle approximation, and quasi-maximum likelihood estimation.
Finally, this construction may be extended to the spatio-temporal case, where the Szegö lemma still holds.