



Intertwining wavelets or multiresolution analysis on graphs through random forests



Luca Avena^a, Fabienne Castell^b, Alexandre Gaudillière^b, Clothilde Mélot^{b,*}

^a Leiden University, Netherlands

^b Aix Marseille Univ, CNRS, Centrale Marseille, I2M, Marseille, France

ARTICLE INFO

Article history:

Received 14 July 2017

Received in revised form 18 April 2018

Accepted 21 September 2018

Available online 5 October 2018

Communicated by Amit Singer

MSC:

94A12

05C81

05C85

15A15

60J20

60J28

Keywords:

Graph signal processing

Multiresolution analysis

Wavelet basis

Intertwining

Markov processes

Random spanning forests

Determinantal processes

ABSTRACT

We propose a new method for performing multiscale analysis of functions defined on the vertices of a finite connected weighted graph. Our approach relies on a random spanning forest to downsample the set of vertices, and on approximate solutions of Markov intertwining relation to provide a subgraph structure and a filter bank leading to a wavelet basis of the set of functions. Our construction involves two parameters q and q' . The first one controls the mean number of kept vertices in the downsampling, while the second one is a tuning parameter between space localization and frequency localization. We provide an explicit reconstruction formula, bounds on the reconstruction operator norm and on the error in the intertwining relation, and a Jackson-like inequality. These bounds lead to recommend a way to choose the parameters q and q' . We illustrate the method by numerical experiments.

© 2018 Elsevier Inc. All rights reserved.

1. Introduction

Graphs provide a flexible representation of geometric structures of irregular domains, and they are now a commonly used tool in numerous applications including neurosciences, social sciences, biology, transport, communications, etc. One edge between two vertices models an interaction between them, and one can also associate a weight to each edge to quantify levels of interaction, leading to a weighted graph. A graph signal associates data to each vertex of the graph. As a measure of the brain activity in distinct functional regions,

* Corresponding author.

E-mail addresses: l.avena@math.leidenuniv.nl (L. Avena), fabienne.castell@univ-amu.fr (F. Castell), alexandre.gaudilliere@math.cnrs.fr (A. Gaudillière), clothilde.melot@univ-amu.fr (C. Mélot).

functional magnetic resonance images are an instance of a such a graph signal. Therefore, signal processing on weighted graphs has raised significant interest in recent years (see [21] for a review on the subject). This paper is concerned with the problem of defining a multiresolution analysis for graph signals. To introduce the problem, let us recall the now classical method to perform such a multiresolution analysis on a regular domain.

1.1. Multiresolution analysis on regular grids

Let us consider a discrete periodic function $f : \mathbb{Z}_n = \mathbb{Z}/n\mathbb{Z} \rightarrow \mathbb{R}$, viewed as a vector in \mathbb{R}^n . The multiresolution analysis of f is based on wavelet analysis and the so-called multiresolution scheme. Performing the wavelet analysis amounts to compute an “approximation” $f_1 \in \mathbb{R}^{n/2}$ and a “detail” component $g_1 \in \mathbb{R}^{n/2}$ through classical operations in signal processing such as “filtering” and “downsampling”. Roughly speaking one wishes the approximation to give the main trends present in f whereas the detail would contain more refined information. This is done by splitting the frequency content of f into two components: f_1 focuses on the low frequency part of f whereas the high frequencies in f are contained in g_1 . In signal processing and image processing, “filtering” (i.e. computing the convolution $f \star k$ for some well chosen kernel k) allows to perform such frequency splittings. In our case $K^l(f) = f \star k^l$ yields $K^l(f)$ as a low frequency component of f and $K^h(f) = f \star k^h$ yields $K^h(f)$ as a high frequency version of f . The vectors f_1 and g_1 are “downsampled” versions of $K^l(f)$ and $K^h(f)$ by a factor of 2, which means that one keeps one coordinate of $K^l(f)$ and $K^h(f)$ out of two, to build f_1 and g_1 respectively. Thus the total length of the concatenation of the two vectors $[f_1, g_1]$ is exactly n , the length of f . To sum up we have

$$\begin{aligned} f_1(\bar{x}) &= K^l(f)(\bar{x}) = \langle \varphi_{\bar{x}}, f \rangle \\ g_1(\bar{x}) &= K^h(f)(\bar{x}) = \langle \psi_{\bar{x}}, f \rangle \end{aligned}$$

where

- \bar{x} belongs to the set of downsamples $\overline{\mathbb{Z}_n}$ isomorphic to $\mathbb{Z}/\frac{n}{2}\mathbb{Z}$.
- $\{\varphi_{\bar{x}}, \bar{x} \in \overline{\mathbb{Z}_n}\}$ is the set of functions such that the equality between linear forms $\langle \varphi_{\bar{x}}, \cdot \rangle = K^l(\cdot)(\bar{x})$ holds for all $\bar{x} \in \overline{\mathbb{Z}_n}$.
- In the same way $\{\psi_{\bar{x}}, \bar{x} \in \overline{\mathbb{Z}_n}\}$ is such that $\langle \psi_{\bar{x}}, \cdot \rangle = K^h(\cdot)(\bar{x})$ holds for all $\bar{x} \in \overline{\mathbb{Z}_n}$.

The choice of k^l and k^h is clearly crucial and is done in such a way that perfect reconstruction of f is possible from f_1 and g_1 , so that there is no lost information in the representation $[f_1, g_1]$. One can iterate the scheme to produce a sequence $f_{N_0} \in \mathbb{R}^{n/2^{N_0}}, g_{N_0} \in \mathbb{R}^{n/2^{N_0}}, \dots, g_1 \in \mathbb{R}^{n/2}$ such that the total length of the concatenated vectors $[f_{N_0}, g_{N_0}, \dots, g_1]$ is exactly n . This is the reason why this scheme yields a multiresolution representation of f . Remark that the perfect reconstruction condition amounts to have $\mathcal{B} = \{\varphi_{\bar{x}}, \psi_{\bar{x}}, \bar{x} \in \overline{\mathbb{Z}_n}\}$ a basis for the signals f on \mathbb{Z}_n . A famous construction by Ingrid Daubechies [6] derives several families of orthonormal compactly supported such basis \mathcal{B} . These families combine localization in space around the point \bar{x} and localization properties in frequency due to the filtering step they have been built from. Using this space-frequency localization one can derive key properties of the wavelet analysis of a signal which rely on the deep links between the local regularity properties of f and the behavior and decay properties of detail coefficients. Let us finish by saying that wavelet type algorithms led to developments of many powerful algorithms of compression, signal restauration, classification, and other kind of time-frequency transforms etc. We will not go into more detail and refer the interested reader to one of the numerous books on wavelet methods and their applications such as [6] or [12].

1.2. Signal processing on graphs

Our aim is to define a multiresolution analysis on a generic weighted finite graph $G = (\mathcal{X}, w, \mu)$, where:

- \mathcal{X} is the set of vertices of G with cardinality $|\mathcal{X}| = n$;
- $w : \mathcal{X} \times \mathcal{X} \mapsto \mathbb{R}^+$ is a weight function, which associates to each pair (x, y) a positive weight $w(x, y)$;
- μ is a measure on \mathcal{X} such that w is reversible w.r.t. μ , meaning that

$$\forall x, y \in \mathcal{X}, \quad \mu(x)w(x, y) = \mu(y)w(y, x). \quad (1)$$

The weight function w provides a graph structure on \mathcal{X} , (x, y) being an edge of the graph iff $w(x, y) > 0$. On such a graph, the operations of translation and downsampling are no longer canonically defined and several attempts to tackle these issues and generalize the wavelet constructions have been proposed: see [21] for a review on this subject and [11] for one of the most popular methods. We will describe some of them in section 1.5, but let us stress that when going to the case of a generic weighted graph, there are three main problems one has to face:

- (Q1): What kind of downsampling should we use? What is the meaning of “keep one point out of two”?
- (Q2): On which weighted graph should the approximation f_1 be defined to iterate the procedure?
- (Q3): Which kind of filters should one use? What is a good “local” mean?

This paper proposes a construction of a multiresolution analysis on a weighted graph which is based on a random downsampling method to answer question (Q1), and on Markov processes intertwining to answer questions (Q2) and (Q3). Markov processes enter naturally into the game through the graph Laplacian \mathcal{L} defined on $\ell_2(\mathcal{X}, \mu)$ by

$$\mathcal{L}f(x) := \sum_{y \in \mathcal{X}} w(x, y)(f(y) - f(x)), \quad (2)$$

where $f : \mathcal{X} \rightarrow \mathbb{R}$ is an arbitrary function, and $w(x, y)$ is now viewed as the transition rate from x to y . For $x \in \mathcal{X}$, let

$$w(x) := \sum_{y \in \mathcal{X} \setminus \{x\}} w(x, y).$$

Note that \mathcal{L} acts on functions as the matrix, still denoted by \mathcal{L} :

$$\mathcal{L}(x, y) = w(x, y) \text{ for } x \neq y; \quad \mathcal{L}(x, x) = -w(x).$$

\mathcal{L} is the generator of a continuous time Markov process $X = (X(t), t \geq 0)$, which jumps from x to a neighboring site y at random exponential times (see section 2.3). The law of X starting from x is denoted by P_x , and E_x is the expectation w.r.t. P_x . Apart from the symmetry assumption (1), we will assume throughout the paper that

$$\mathcal{L} \text{ is irreducible.} \quad (3)$$

Under this assumption, μ is the unique (up to multiplicative constants) invariant measure of the process $(X(t), t \geq 0)$ (i.e. $\mu\mathcal{L} = 0$), and $\mu(x) > 0$ for any $x \in \mathcal{X}$. Since \mathcal{X} is finite, we can assume without loss of generality that μ is a probability measure on \mathcal{X} .

1.3. Our approach for downsampling, weighting and filtering procedures on graphs

The first step in our proposal of a multiresolution scheme consists in answering (Q1) and constructing a random subset $\bar{\mathcal{X}}$ of \mathcal{X} , whose main feature is to be “well spread” on \mathcal{X} . In this respect, we use an adaptation of Wilson’s algorithm ([24]) studied in [2], whose output is a random rooted spanning forest Φ on G . Its law π_q depends on the weight function w and on a positive real q . This algorithm and the properties of the corresponding random forest will be described in section 3.3. The set of roots of Φ , denoted by $\rho(\Phi)$, has the nice property of being a determinantal process on \mathcal{X} with kernel given by the Green kernel:

$$K_q(x, y) := q(q\text{Id} - \mathcal{L})^{-1}(x, y) = P_x[X(T_q) = y], \quad (4)$$

where T_q is an exponential random variable with parameter q , independent of the process X . Being a determinantal process, the roots of Φ tend to repulse each other, and this repulsiveness property results in the noteworthy fact that the mean time needed by the process $(X(t), t \geq 0)$ to reach $\rho(\Phi)$, does not depend on the starting point in $\mathcal{X} \setminus \rho(\Phi)$. Therefore, this set of roots is a natural choice for $\bar{\mathcal{X}}$, especially as the parameter q can be tuned to control the mean number of points in $\rho(\Phi)$. See Section 3 for more details.

Once the downsampling $\bar{\mathcal{X}}$ is fixed, it remains to answer questions (Q2) and (Q3), i.e. we have to construct a weight function \bar{w} on $\bar{\mathcal{X}} \times \bar{\mathcal{X}}$, and to define a filtering method which gives the approximation f_1 and the detail g_1 of a given function f_0 in $\ell_2(\mathcal{X}, \mu)$. This will be achieved by looking for a solution $(\bar{\mathcal{L}}, \Lambda)$ to the intertwining relation

$$\bar{\mathcal{L}}\Lambda = \Lambda\mathcal{L}, \quad (5)$$

where

- $\bar{\mathcal{L}}$ is a Markov generator on $\bar{\mathcal{X}}$;
- $\Lambda : \bar{\mathcal{X}} \times \mathcal{X} \mapsto \mathbb{R}^+$ is a positive rectangular matrix.

An intertwining relation gives a natural link between two Markov processes living on different state spaces. It appeared in the context of diffusion processes in the paper by Rogers and Pitman [19], as a tool to state identities in laws, and was later successfully applied to many other examples (see for instance [3], [13], etc.). In the context of Markov chains, intertwining was used by Diaconis and Fill [7] to construct strong stationary times, and to control the rate of convergence to equilibrium. At the time being, applications of intertwining include random matrices [8], particle systems [23], etc. For our purpose, the intertwining relation can be very useful since

- $\bar{\mathcal{L}}$ provides a natural choice for the graph structure to be put on $\bar{\mathcal{X}}$;
- each row $\nu_{\bar{x}} := \Lambda(\bar{x}, \cdot)$ ($\bar{x} \in \bar{\mathcal{X}}$) defines a positive measure on \mathcal{X} , which can serve as a “local mean” around \bar{x} . f_0 will be approximated by the function f_1 defined on $\bar{\mathcal{X}}$ by

$$\forall \bar{x} \in \bar{\mathcal{X}}, f_1(\bar{x}) := \nu_{\bar{x}}(f_0) = \sum_{x \in \mathcal{X}} \Lambda(\bar{x}, x) f_0(x);$$

- it will serve as a basis for getting a Jackson-type inequality (see Proposition 18).

To be of any use in signal processing, the “filters” $(\nu_{\bar{x}}, \bar{x} \in \bar{\mathcal{X}})$ have to be well localized in space and frequency. In the graph context, frequency localization means that the filters belong to an eigenspace of the graph laplacian \mathcal{L} . Hence, we are interested in solutions to (5) such that the measures $(\nu_{\bar{x}}, \bar{x} \in \bar{\mathcal{X}})$ are linearly

independent measures tending to be non-overlapping (space localization), and contained in eigenspaces of \mathcal{L} . In addition, in order to iterate the procedure, we also need $\bar{\mathcal{L}}$ to be reversible on $\bar{\mathcal{X}}$.

Note that saying that $(\Lambda, \bar{\mathcal{L}})$ is an exact solution to (5) implies that the linear space spanned by the measures $(\nu_{\bar{x}}, \bar{x} \in \bar{\mathcal{X}})$ is stable by \mathcal{L} , and is therefore a direct sum of eigenspaces of \mathcal{L} , so that these measures provide filters which are frequency localized. Hence the error in the intertwining relation is a measure of frequency localization: the smaller the intertwining error, the better the frequency localization. Finding solutions to (5) is the purpose of [1], where an exact linearly independent solution to (5) is provided. However this solution tends to be overlapping. [1] provides also approximate solutions to (5) with small overlapping, and it is one of these approximate solutions we use in our multiresolution analysis scheme. In order to describe it, we first consider the trace process of $(X(t), t \geq 0)$ on $\bar{\mathcal{X}}$, i.e. the process $(X(t), t \geq 0)$ sampled at the passage times on $\bar{\mathcal{X}}$. This process is a Markov process on $\bar{\mathcal{X}}$, whose generator $\bar{\mathcal{L}}$ defines a weight function \bar{w} on $\bar{\mathcal{X}}$, symmetric with respect to the measure μ restricted on $\bar{\mathcal{X}}$. We will recall some general facts about Markov processes, and make the previous statements more precise in section 2.3.

Turning to the rectangular matrix Λ , it depends on a parameter q' and involves the Green kernel $K_{q'}$ (see definition (4)). The rectangular matrix Λ is just the restriction of $K_{q'}$ to $\bar{\mathcal{X}} \times \mathcal{X}$. Note that

- when q' goes to 0, for any $\bar{x} \in \bar{\mathcal{X}}$, $K_{q'}(\bar{x}, y)$ goes to $\mu(y)$ so that (5) is clearly satisfied. μ being the left-eigenvector of \mathcal{L} corresponding to the eigenvalue 0, the $K_{q'}(\bar{x}, y)$ are well frequency localized. However, the vectors $(K_{q'}(\bar{x}, \cdot), \bar{x} \in \bar{\mathcal{X}})$ become linearly dependent and very badly space localized.
- when q' goes to ∞ , $K_{q'}(\bar{x}, \cdot)$ goes to $\delta_{\bar{x}}$. Hence, the space localization is perfect. However, the frequency localization is lost, and the error in (5) tends to grow.

Hence, a compromise has to be made concerning the choice of q' , and we will discuss this point later.

To sum up our proposal, two parameters q and q' being fixed,

- take $\bar{\mathcal{X}} := \rho(\Phi)$, Φ being sampled with π_q . Set $\check{\mathcal{X}} = \mathcal{X} \setminus \bar{\mathcal{X}}$.
- $\bar{\mathcal{L}}$ is the generator of the trace process of X on $\bar{\mathcal{X}}$. It can be shown (see Lemma 8 section 4.1) that $\bar{\mathcal{L}}$ is irreducible and reversible w.r.t. the probability measure $\mu_{\bar{\mathcal{X}}}$, which is the measure μ conditioned to the set $\bar{\mathcal{X}}$: $\mu_{\bar{\mathcal{X}}}(A) = \mu(A \cap \bar{\mathcal{X}})/\mu(\bar{\mathcal{X}})$. $\bar{\mathcal{L}}$ is actually the Schur complement in \mathcal{L} , of \mathcal{L} restricted to $\check{\mathcal{X}}$, which was already used for instance in [11]. For any \bar{x}, \bar{y} in $\bar{\mathcal{X}}$, such that $\bar{x} \neq \bar{y}$, the weight function \bar{w} on $\bar{\mathcal{X}} \times \bar{\mathcal{X}}$ is then defined by $\bar{w}(\bar{x}, \bar{y}) = \bar{\mathcal{L}}(\bar{x}, \bar{y})$.
- $\Lambda(\bar{x}, \cdot) := K_{q'}(\bar{x}, \cdot)$.

$f (=f_0)$ being a function in $\ell_2(\mathcal{X}, \mu)$, we define the approximation of f as the function $\bar{f} (=f_1)$ defined on $\bar{\mathcal{X}}$ by

$$\forall \bar{x} \in \bar{\mathcal{X}}, \quad \bar{f}(\bar{x}) = \Lambda f(\bar{x}) = K_{q'} f(\bar{x}), \quad (6)$$

and its detail function as the function $\check{f} (=g_1)$ defined on $\check{\mathcal{X}}$ by

$$\forall \check{x} \in \check{\mathcal{X}}, \quad \check{f}(\check{x}) = (K_{q'} - \text{Id})f(\check{x}). \quad (7)$$

We proved in [1] that some localization property of the $(\nu_{\bar{x}} = \Lambda(\bar{x}, \cdot), \bar{x} \in \bar{\mathcal{X}})$ defined by (6) followed from the fact that $\bar{\mathcal{X}}$ is a determinantal process with kernel K_q . Since $\check{\mathcal{X}} = \mathcal{X} \setminus \bar{\mathcal{X}}$ is also a determinantal process, with kernel $\text{Id} - K_q$, this suggested the detail definition (7), in which the sign convention is chosen to have a self-adjoint analysis operator $U : f \mapsto (\bar{f}, \check{f})$ in $\ell_2(\mathcal{X}, \mu)$.

The process can then be iterated with $(\bar{\mathcal{X}}, \bar{\mathcal{L}}, \mu_{\bar{\mathcal{X}}})$ in place of $(\mathcal{X}, \mathcal{L}, \mu)$. This leads to a multiresolution scheme.

1.4. Description of the results

We give now a description of our main results. Note that whatever the choice of a subset $\bar{\mathcal{X}}$ of \mathcal{X} , one can still define $\bar{\mathcal{L}}$, Λ , \bar{f} and \check{f} as we just did it previously. Our first set of results assumes that $\bar{\mathcal{X}}$ is any subset of \mathcal{X} , and provides a reconstruction formula, and bounds on various operators including approximation, detail and analysis operators. They are expressed in terms of the hitting time of $\bar{\mathcal{X}}$:

$$H_{\bar{\mathcal{X}}} := \inf \{t \geq 0, X(t) \in \bar{\mathcal{X}}\}.$$

To state them, we introduce some notation. First of all, we define the maximal rate $\alpha > 0$ by

$$\alpha = \max_{x \in \mathcal{X}} w(x). \quad (8)$$

Hence, the matrix $P = \text{Id} + \mathcal{L}/\alpha$ is a stochastic matrix on \mathcal{X} . $\bar{\mathcal{X}}$ being any subset of \mathcal{X} , we define the two positive real numbers β and γ by

$$\frac{1}{\gamma} := \max_{\bar{x} \in \bar{\mathcal{X}}} E_{\bar{x}}[H_{\bar{\mathcal{X}}}], \quad \frac{1}{\beta} := \max_{\bar{x} \in \bar{\mathcal{X}}} \sum_{z \in \mathcal{X}} P(\bar{x}, z) E_z[H_{\bar{\mathcal{X}}}]. \quad (9)$$

$\bar{\alpha}$ denotes the maximal rate of $\bar{\mathcal{L}}$:

$$\bar{\alpha} := \max_{\bar{x} \in \bar{\mathcal{X}}} -\bar{\mathcal{L}}(\bar{x}, \bar{x}).$$

We stress the obvious fact that γ , β and $\bar{\alpha}$ are functions of the subset $\bar{\mathcal{X}}$.

Finally, when $(M(u, v), u \in \mathcal{U}, v \in \mathcal{V})$ is a rectangular matrix, when U is a subset of \mathcal{U} , and V a subset of \mathcal{V} , M_{UV} denotes the submatrix of M obtained by restricting the entries of M to U and V .

To begin with our results, we get a reconstruction formula:

Reconstruction formula (Proposition 10).

For any $f \in \ell_2(\mathcal{X}, \mu)$, let $\bar{f} \in \ell_2(\bar{\mathcal{X}}, \mu)$ and $\check{f} \in \ell_2(\check{\mathcal{X}}, \mu)$ be defined by (6) and (7). Then,

$$f = \bar{R}\bar{f} + \check{R}\check{f},$$

where \bar{R} and \check{R} are rectangular matrices indexed respectively by $\mathcal{X} \times \bar{\mathcal{X}}$ and $\mathcal{X} \times \check{\mathcal{X}}$, whose block decompositions are

$$\bar{R} = \begin{pmatrix} \text{Id}_{\bar{\mathcal{X}}} - \frac{1}{q'} \bar{\mathcal{L}} \\ (-\mathcal{L}_{\check{\mathcal{X}}\bar{\mathcal{X}}})^{-1} \mathcal{L}_{\check{\mathcal{X}}\bar{\mathcal{X}}} \end{pmatrix}, \quad \text{and} \quad \check{R} = \begin{pmatrix} \mathcal{L}_{\bar{\mathcal{X}}\check{\mathcal{X}}}(-\mathcal{L}_{\check{\mathcal{X}}\check{\mathcal{X}}})^{-1} \\ -\text{Id}_{\check{\mathcal{X}}} - q'(-\mathcal{L}_{\check{\mathcal{X}}\check{\mathcal{X}}})^{-1} \end{pmatrix}. \quad (10)$$

We provide also upper bounds on norms of various operators: the approximation operator \bar{R} , the detail operator \check{R} , and the intertwining error operator $\bar{\mathcal{L}}\Lambda - \Lambda\mathcal{L}$. These bounds are given here when these operators are seen as operators from one ℓ_∞ space to another one. They are stated in greater generality in Propositions 14, 11 and 16.

Approximation and detail operator norms (Propositions 14 and 11).

Let $\bar{\mathcal{X}}$ be any proper subset of \mathcal{X} (i.e. $\emptyset \subsetneq \bar{\mathcal{X}} \subsetneq \mathcal{X}$), $\check{\mathcal{X}} = \mathcal{X} \setminus \bar{\mathcal{X}}$, and let \bar{R} and \check{R} be the operators defined in (10). For any $\bar{f} \in \ell_\infty(\bar{\mathcal{X}}, \mu_{\bar{\mathcal{X}}})$ and any $\check{f} \in \ell_\infty(\check{\mathcal{X}}, \mu_{\check{\mathcal{X}}})$,

$$\|\bar{R}\bar{f}\|_{\infty, \mathcal{X}} \leq \left(1 + 2\frac{\bar{\alpha}}{q'}\right) \|\bar{f}\|_{\infty, \bar{\mathcal{X}}}, \quad \text{and} \quad \|\check{R}\check{f}\|_{\infty, \mathcal{X}} \leq \max\left(\frac{\alpha}{\beta}; 1 + \frac{q'}{\gamma}\right) \|\check{f}\|_{\infty, \check{\mathcal{X}}}. \quad (11)$$

Intertwining error norm (Proposition 16).

Let $\bar{\mathcal{X}}$ be any proper subset of \mathcal{X} , $\check{\mathcal{X}} = \mathcal{X} \setminus \bar{\mathcal{X}}$. For any $f \in \ell_\infty(\mathcal{X}, \mu)$,

$$\|(\bar{\mathcal{L}}\Lambda - \Lambda\mathcal{L})f\|_{\infty, \bar{\mathcal{X}}} \leq 2q' \frac{\alpha}{\beta} \|f\|_{\infty, \mathcal{X}}. \quad (12)$$

Note that these bounds reflect the competition between $\|(\bar{\mathcal{L}}\Lambda - \Lambda\mathcal{L})f\|_{\infty, \bar{\mathcal{X}}}$ (frequency localization) and $\|\bar{R}\bar{f}\|_{\infty, \mathcal{X}}$ (space localization). Actually, the term $\bar{\alpha}/q'$ appearing in (11) is a decreasing function of q' and an increasing function of $\bar{\mathcal{X}}$, while the term q'/β in (12) is increasing in q' and decreasing in $\bar{\mathcal{X}}$.

Other results are stated in the paper, including a bound on the norm of the detail \check{f} in terms of the norm of $\mathcal{L}f$ (Proposition 15), and a Jackson's type inequality on the approximation error after K steps of the mutiresolution scheme (Proposition 18).

The other set of results focuses on the case where $\bar{\mathcal{X}}$ is the set of roots of the random forest Φ . They provide estimates on its cardinality $|\bar{\mathcal{X}}|$, and on the quantities $\bar{\alpha}$, β and γ involved in the previous statements. $\bar{\mathcal{X}}$ being random, all these quantities are random ones, and the estimates we get are averaged ones. To state them, we will assume that the random variable Φ is defined on some probability space $(\Omega_f, \mathcal{A}_f, \mathbb{P}_q)$. The corresponding expectation will be denoted by \mathbb{E}_q .

Cardinality estimates (Proposition 6).

Let $\bar{\mathcal{X}} = \rho(\Phi)$ and $\check{\mathcal{X}} = \mathcal{X} \setminus \bar{\mathcal{X}}$. Let $r \in]0; 1[$, and define $\mathcal{R}_r = \{x \in \mathcal{X}; w(x) \geq r\alpha\}$ the set of “rapid” points. Then,

$$|\mathcal{X}| \frac{q}{q + \alpha} \leq \mathbb{E}_q [|\bar{\mathcal{X}}|] ; \mathbb{E}_q [|\bar{\mathcal{X}} \cap \mathcal{R}_r|] \leq |\mathcal{R}_r| \frac{q + (2 - r)\alpha}{q + 2\alpha}.$$

These estimates can be used to tune the parameter q in order to target a given proportion of kept points. Next, we obtain bounds on the quantities $\bar{\alpha}$, β and γ when $\bar{\mathcal{X}} = \rho(\Phi)$. Unfortunately, these bounds are just lower bounds, and are moreover averaged ones. Nevertheless, since they are expressed in terms of the cardinality of $\rho(\Phi)$, they can easily be estimated by Monte Carlo methods, and can serve as a guide for the choice of q and q' . These are then the key estimates where our particular random choice for $\bar{\mathcal{X}}$ plays a central role.

Estimates on $\bar{\alpha}$, β , γ (Propositions 20, 21 and 22).

Let $\bar{\mathcal{X}} = \rho(\Phi)$ and $\check{\mathcal{X}} = \mathcal{X} \setminus \bar{\mathcal{X}}$.

$$\mathbb{E}_q [\bar{\alpha}] \geq q \mathbb{E}_q \left[\frac{|\check{\mathcal{X}}|}{|\bar{\mathcal{X}}| + 1} \right], \quad \mathbb{E}_q \left[\frac{1}{\beta} \right] \geq \mathbb{E}_q \left[\frac{|\check{\mathcal{X}}|}{\alpha |\bar{\mathcal{X}}|} \right], \quad \mathbb{E}_q \left[\frac{1}{\gamma} \right] \geq \frac{1}{q} \mathbb{E}_q \left[\frac{|\mathcal{X}|}{|\check{\mathcal{X}}| + 1} \right]. \quad (13)$$

Discussion on the choice of q and q'

Based on these estimates, we argue in section 6.4 that the parameter q should be chosen in $[\theta_1\alpha; \theta_2\alpha]$ in order to ensure that the mean number of vertices of the subgraph is at least a given proportion $(\theta_1/(\theta_1 + 1))$ of the size of the original graph, and that a given proportion of the rapid points are decimated. In addition it should minimize the function $q \mapsto \mathbb{E}_q(\bar{\alpha}/\beta)$, for the approximation operator norm and the intertwining error to be small (see (11) and (12)). The Monte-Carlo estimation of this function is computationally costly, so that we propose to minimize the function $q \in [\theta_1\alpha; \theta_2\alpha] \mapsto q \mathbb{E}_q \left[\frac{|\check{\mathcal{X}}|}{(1 + |\bar{\mathcal{X}}|)} \right] \mathbb{E}_q \left[\frac{|\check{\mathcal{X}}|}{|\bar{\mathcal{X}}|} \right]$. This is possible in practice since, from [2], we can simultaneously sample a whole continuum of forests Φ_q for $q \in [\theta_1\alpha; \theta_2\alpha]$, each of them with the correct distribution π_q . Once q has been fixed, $\bar{\mathcal{X}}$ is sampled according to π_q , $\bar{\alpha}$ and β are computed and q' is chosen equal to $2\bar{\alpha} |\bar{\mathcal{X}}| / |\check{\mathcal{X}}|$, which will ensure numerical stability of the algorithm (see Section 6.4).

1.5. Related works

Previous authors have explored multiresolution analysis on graphs, and have proposed answers to questions (Q1-3). Far from being exhaustive, we describe some of these, and refer the interested reader to [21] for a more complete state of the art.

Concerning the downsampling procedure, also referred to as the graph coarsening problem, many approaches have been investigated. To mention a few, one can try to decompose the graph into bipartite graphs for which the notion of “one every two points” is clear. This is the way followed in [15,17], using either *coloring-based downsampling*, or *maximum spanning tree*. In [20], the partitioning of \mathcal{X} into two subsets, is based on the sign of the eigenfunction of $-\mathcal{L}$ associated to the maximal eigenvalue. Finally, the authors of [22] use a community detection algorithm maximizing the modularity, to partition the original graph into many connected subgraphs of small size. The vertices of the downsampled graph are the elements of the partition, and are not properly speaking selected points of the original graph.

Turning to the weighting procedure, bipartite graphs designed methods put an edge between two selected nodes, if they share at least a neighbor in the original graph, and the weight may be proportional to the number of shared neighbors. Starting from a bipartite graph, this leads immediately to a non bipartite subgraph, thus the need to “decompose” a graph by bipartite ones. The authors of [20] rely on the so-called *Kron reduction*, which is the same as computing the Schur complement. In [22], the weight between two communities is the sum of the weight of edges linking these communities.

Finally, as far as the filtering procedure is concerned, various filter banks have been proposed, leading sometimes to orthonormal basis, or just frames. The *graph wavelet filtering bank* of [16] is more specifically designed for bipartite graphs, and exploit the specific features of bipartite graph Laplacian. It produces an orthonormal basis. The *diffusion wavelets* of [4] are obtained by constructing orthonormal basis of the spaces $(V_{k-1} \ominus V_k, k = 1, \dots, K)$, where $V_k = \Im((- \mathcal{L})^{2^k})$, and are thus designed to form an orthonormal basis. The *spectral graph wavelets* of [11] are functions of type $g(-t\mathcal{L})\delta_n$ where g is a function localized around 0 for the low-pass filters, or away from 0 for the high-pass ones. Different values of t are used to select different frequencies, thus leading to a frame. Since the computation of $g(-t\mathcal{L})$ requires the knowledge of the spectral decomposition of \mathcal{L} , polynomial approximations are performed. In [22], the filters used are the eigenfunctions of the Laplacians restricted to each small community, naturally leading to an orthonormal basis. In [10], assuming that we are given a subsampling procedure encoding the geometry of the original graph, the authors construct a Haar basis adapted to this subsampling.

Compared to these works, our downsampling approach through Wilson’s algorithm, is a partitioning method in the spirit of [22], a community corresponding to a tree of the forest. The mean number of selected points can be adjusted through the parameter q .

The filters we use are a special case of spectral graph wavelets of [11], the function g being equal to $1/(1+x)$ for the scaling function, and $x/(1+x)$ for the wavelet, and t corresponding to $1/q'$ (note that $K_{q'} = q'(q'\text{Id} - \mathcal{L})^{-1} = (\text{Id} - \mathcal{L}/q')^{-1}$). This choice is very natural in the signal processing variational approach, since $K_{q'}f$ minimizes the function $g \mapsto \langle g, -\mathcal{L}g \rangle + q'\|g - f\|^2$. Since we use only one value of t at each step, the filter bank we obtain is a basis, which is not orthogonal, but tends to be so when $q' \gg 1$ (in this case $K_{q'}(\bar{x}, \cdot) \simeq \delta_{\bar{x}}$, while $(K_{q'} - \text{Id})(\check{x}, \cdot) \simeq \mathcal{L}(\check{x}, \cdot)/q'$).

The weighting procedure through Kron reduction has also already be used in [20]. But in our approach, weighting and filtering are linked together through the intertwining relation.

To sum up, the advantages of our approach are:

- to provide a new partitioning method, allowing to tune the mean number of kept vertices;
- to link the subgraph structure and the choice of the filters through the intertwining relation;
- to produce a filter bank leading to a basis;
- to mimic the steps of a classical wavelet analysis algorithm;

- to allow the computation of various error bounds;
- to suggest a systematic method to choose the parameters q and q' , and ensure numerical stability;
- to have a computational complexity similar to already existing methods. Actually, apart from the sampling of the random forest which is of order $\alpha |\mathcal{X}|/q$, our approach only requires the computation of $K_{q'}$ and of the Schur complement, already present in [11,20]. Starting from a sparse matrix \mathcal{L} , $K_{q'}$ can be efficiently approximated by polynomials of small order. Due to the good repartition property of our random $\bar{\mathcal{X}}$, such a polynomial approximation can also be implemented for the inversion involved in the computation of $\bar{\mathcal{L}}$. This results in a sparse $\bar{\mathcal{L}}$, however the level of sparsity obtained is not enough to go on with the algorithm for large graphs. Hence, as in [20], a sparsification step can be added after the computation of $\bar{\mathcal{L}}$. Our proposition for this sparsification will be guided by the intertwining error bounds that provide a Jackson-like inequality (see Section 7).

Finally, we would like to mention that our multiresolution scheme construct a basis of the space of signals, which can be used to analyze, compress, etc., any signal on the graph. When one wants to handle just one specific signal, adaptative multiresolution schemes as in [9] may be more appropriate.

1.6. Organization of the paper

We begin in section 2 with notations used throughout the paper. Section 3 is devoted to the description and the properties of the random forest, and to the downsampling procedure. The weighting and the filtering procedures, $\bar{\mathcal{X}}$ being any subset of \mathcal{X} , are discussed in section 4. We prove in this section the bounds on operator norms. We go on with the iterative scheme and Jackson's inequality in section 5.2. The discussion on the choice of q and q' and the estimates on $\bar{\alpha}$, β and γ , are given in section 6. We summarize the pyramidal algorithm and discuss computational issues in section 7, and we finally end with numerical experiments illustrating the method in section 8.

2. Notations and preliminary results

In this section, we give the notations used throughout the paper, and we state some useful results.

2.1. Sets of functions and measures

A function f on \mathcal{X} will be seen as a column vector, whereas a signed measure on \mathcal{X} will be seen as a row vector. For $p \geq 1$, $\ell_p(\mathcal{X}, \mu)$ is the space of functions on \mathcal{X} endowed with the norm

$$\|f\|_p = \left(\sum_{x \in \mathcal{X}} |f(x)|^p \mu(x) \right)^{1/p}.$$

The scalar product of two functions f and g in $\ell_2(\mathcal{X}, \mu)$ is

$$\langle f, g \rangle = \sum_{x \in \mathcal{X}} f(x)g(x)\mu(x)$$

When f is a function on \mathcal{X} , f^* will denote the signed measure whose density w.r.t. μ is f : $f^*(A) = \sum_{x \in A} \mu(x)f(x)$ for all subset A of \mathcal{X} . Similarly, when ν is a signed measure on \mathcal{X} , ν^* is the density of ν w.r.t. μ : $\nu^*(x) = \nu(x)/\mu(x)$ for any $x \in \mathcal{X}$.

2.2. Schur complement

Let M be a matrix of size $n = p + r$ and let

$$M = \begin{pmatrix} A & B \\ C & D \end{pmatrix}$$

be its block decomposition, A being a square matrix of size p and D a square matrix of size r . If D is invertible, the Schur complement of D in M is the square matrix of size p defined by

$$S_M(D) := A - BD^{-1}C.$$

We remind the reader the following standard results concerning the Schur complement (see for instance [25]):

Proposition 1. Assume that D is invertible.

- (1) $M = \begin{pmatrix} Id & B \\ 0 & D \end{pmatrix} \begin{pmatrix} S_M(D) & 0 \\ D^{-1}C & Id \end{pmatrix}.$
- (2) $\det(M) = \det(D)\det(S_M(D)).$
- (3) M is invertible if and only if $S_M(D)$ is invertible. In that case,

$$M^{-1} = \begin{pmatrix} S_M(D)^{-1} & -S_M(D)^{-1}BD^{-1} \\ -D^{-1}CS_M(D)^{-1} & D^{-1} + D^{-1}CS_M(D)^{-1}BD^{-1} \end{pmatrix}.$$

- (4) Assume that M defines a positive symmetric operator in $\ell^2(\mathcal{X}, \mu)$. Let $\lambda_{\max}(M)$ (respectively $\lambda_{\min}(M)$) be the largest (respectively the smallest) eigenvalue of M . Then, $S_M(D)$ is also positive symmetric and

$$\lambda_{\min}(M) \leq \lambda_{\min}(S_M(D)) \leq \lambda_{\max}(S_M(D)) \leq \lambda_{\max}(M).$$

2.3. Markov process

Consider an irreducible continuous time Markov process $(X(t), t \geq 0)$ on \mathcal{X} , with generator \mathcal{L} given by (2) satisfying (1) and (3). We recall the definition (8) of $\alpha > 0$ as the maximal rate, and that of the matrix $P := \mathcal{L}/\alpha + Id$. P is an irreducible stochastic matrix, and we denote by $(\hat{X}_k, k \in \mathbb{N})$ a discrete time Markov chain with transition matrix P . The process $(X(t), t \geq 0)$ can be constructed from $(\hat{X}_k, k \in \mathbb{N})$ and an independent Poisson process $(\tau_i, i \geq 0)$ on \mathbb{R}^+ with rate α . At each event of the Poisson process, X moves according to the trajectory of \hat{X} :

$$X(t) = \sum_{i=0}^{+\infty} \hat{X}_i \mathbf{1}_{\tau_i \leq t < \tau_{i+1}}.$$

By (1) and (3), μ is strictly positive. In addition, $-\mathcal{L}$ defines a positive symmetric operator on $\ell_2(\mathcal{X}, \mu)$, and we denote by $(\lambda_i; i = 0, \dots, n-1)$ the real eigenvalues of $-\mathcal{L}$ in increasing order. It follows from the fact that P is irreducible that

$$0 = \lambda_0 < \lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_{n-1} \leq 2\alpha. \quad (14)$$

The right eigenvector of $-\mathcal{L}$ associated to λ_i is denoted by e_i :

$$-\mathcal{L}e_i = \lambda_i e_i.$$

The $(e_i, i = 0, \dots, n-1)$ are normalized to have an ℓ_2 -norm equal to 1, and form an orthonormal basis of $\ell_2(\mathcal{X}, \mu)$. By construction, $e_0 = \mathbf{1}$. For any subset A of \mathcal{X} , H_A is the hitting time of A by the process X :

$$H_A := \inf \{t \geq 0, X(t) \in A\},$$

while H_A^+ is the return time to A by the process X :

$$H_A^+ := \inf \{t \geq \tau_1, X(t) \in A\}. \quad (15)$$

3. Random spanning forests, Wilson's algorithm and downsampling procedure

We spend now some time on the description and the properties of the spanning random forest Φ used in the downsampling procedure. Let us call \mathcal{E} the set of unoriented edges of G , that is the set of pairs $\{x, y\}$ such that $w(x, y) > 0$ (and $w(y, x) > 0$). A spanning *unoriented* forest is a graph without cycles, with \mathcal{X} as set of vertices, and a subset of \mathcal{E} as edge set. An unoriented tree is a connected component of such a forest. By choosing in each tree one specific vertex, which we call root, we define a rooted spanning forest. Note that the number of roots is the same as the number of trees. Orienting each edge of a tree toward its root, we obtain an spanning oriented forest (s.o.f.) ϕ . The set of roots of a spanning oriented forest ϕ is denoted by $\rho(\phi)$. If $e = (x, y)$ is an oriented edge, we will use $w(e)$ for $w(x, y)$, and say that $e \in \phi$ if e is an edge of a tree of the forest ϕ .

3.1. A probability measure on forests

We introduce now a real parameter $q > 0$, and associate to each oriented forest a weight

$$w_q(\phi) := q^{|\rho(\phi)|} \prod_{e \in \phi} w(e), \quad (16)$$

where $|\rho(\phi)|$ is the cardinality of $\rho(\phi)$, i.e. the number of trees in the forest ϕ . $w_1(\phi)$ will be denoted by $w(\phi)$ so that

$$w_q(\phi) := q^{|\rho(\phi)|} w(\phi).$$

These weights can be renormalized to define a probability measure on the set of spanning oriented forest:

$$\pi_q(\phi) := \frac{w_q(\phi)}{Z(q)}, \quad (17)$$

where the partition function $Z(q)$ is given by

$$Z(q) := \sum_{\phi \text{ s.o.f.}} w_q(\phi). \quad (18)$$

3.2. Wilson's algorithm

A way to sample a random s.o.f. Φ from π_q is given by the following iterative algorithm. Let Φ_c be the current state of the forest being constructed, and let \mathcal{V}_c be the set of vertices of Φ_c . At the beginning, \mathcal{V}_c is equal to \emptyset .

While $\mathcal{X} \setminus \mathcal{V}_c \neq \emptyset$, perform the following steps:

- Choose a point x at random in $\mathcal{X} \setminus \mathcal{V}_c$.
- Let evolve the Markov process $(X(t), t \geq 0)$ from x , and stop it either when it reaches \mathcal{V}_c , or after an independent exponential time of parameter q .
- Erase the loops of the trajectory drawn by X . We obtain a self-avoiding path C starting from x and oriented towards its end-point.
- Add C to Φ_c .

Each iteration of the “while loop” stopped by the exponential time, gives birth to another tree. Wilson’s algorithm is not only a way to sample π_q . It provides also a powerful tool to analyze it, the reason being that it does not depend on the way we “choose a point” in the first step of the “while” loop. In addition there exists a coupling of the probability measures π_q for different values of q . This means that we can construct the random forests for different values of q from the same set of random variables. This coupling is explained in [2].

3.3. Properties of the random forest

Using Wilson’s algorithm, and the explicit knowledge of the law of a loop-erased Markov process, the following statement, as all the results stated in this section, is proven in [2]:

Proposition 2 (Partition function). *The partition function $Z(q)$ is the characteristic polynomial of \mathcal{L} :*

$$Z(q) = \det(q \text{Id} - \mathcal{L}). \quad (19)$$

Some other important features of the random s.o.f. Φ are listed below.

Proposition 3 (Number of roots). *For all $k \in \{0, \dots, n\}$,*

$$\mathbb{P}_q[|\rho(\Phi)| = k] = \sum_{\substack{I \subset \{0, \dots, n-1\} \\ |I|=k}} \prod_{i \in I} \frac{q}{q + \lambda_i} \prod_{i \notin I} \frac{\lambda_i}{q + \lambda_i}.$$

Otherwise stated, the number of roots has the same law as $\sum_{i=0}^{n-1} B_i$ where B_0, \dots, B_{n-1} are independent, B_i having Bernoulli distribution with parameter $\frac{q}{q + \lambda_i}$.

Note that since $\lambda_0 = 0$, $B_0 = 1$ a.s. and we recover the fact that $|\rho(\Phi)| \geq 1$ a.s.

Proposition 4 (Set of roots). *$\rho(\Phi)$ is a determinantal process on \mathcal{X} with kernel K_q (see definition in (4)), i.e. for any subset A of \mathcal{X} ,*

$$\mathbb{P}_q(A \subset \rho(\Phi)) = \det_A(K_q),$$

where \det_A is the minor defined by the rows and columns corresponding to A .

The following statement involves two independent sources of randomness, the Markov process X starting from x defined on $(\Omega, \mathcal{A}, P_x)$, and the random forest Φ defined on $(\Omega_f, \mathcal{A}_f, \mathbb{P}_q)$. Integration on the product space $(\Omega \times \Omega_f, \mathcal{A} \otimes \mathcal{A}_f)$ w.r.t. the product measure $\mathbb{P}_{x,q} := P_x \otimes \mathbb{P}_q$ is denoted by $\mathbb{E}_{x,q}$.

Proposition 5 (Hitting time of $\rho(\Phi)$).

(1) For any $x \in \mathcal{X}$,

$$\mathbb{E}_{x,q} [H_{\rho(\Phi)}] = \frac{\mathbb{P}_q [|\rho(\Phi)| > 1]}{q}. \quad (20)$$

(2) For any $x \in \mathcal{X}$, and $m \in \{1, \dots, n\}$,

$$\mathbb{E}_{x,q} [H_{\rho(\Phi)} | |\rho(\Phi)| = m] = \frac{\mathbb{P}_q [|\rho(\Phi)| = m+1]}{q \mathbb{P}_q [|\rho(\Phi)| = m]}. \quad (21)$$

Note that these expressions do not depend on the starting point x , saying that in some sense, the roots of the random forest are “well spread” on \mathcal{X} .

3.4. The downsampling procedure

In view of the results of section 3.3, $\bar{\mathcal{X}} := \rho(\Phi)$ is a natural candidate as a downsampling of \mathcal{X} . We give now estimates on the mean of $|\rho(\Phi)|$ which unlike Proposition 3, do not depend on the knowledge of the eigenvalues λ_i . To this purpose, we introduce the mean value of the eigenvalues of $-\mathcal{L}/\alpha$:

$$m(\mathcal{L}) := \frac{1}{|\mathcal{X}|} \sum_{i=0}^{|\mathcal{X}|-1} \frac{\lambda_i}{\alpha} = \frac{\text{Trace}(-\mathcal{L})}{\alpha |\mathcal{X}|}.$$

Note that by (14), $m(\mathcal{L}) \in]0, 2]$. We get then

Proposition 6. Let $\bar{\mathcal{X}} = \rho(\Phi)$ and $\check{\mathcal{X}} = \mathcal{X} \setminus \bar{\mathcal{X}}$. Then

- (1) $\mathbb{E}_q [|\bar{\mathcal{X}}|] \geq |\mathcal{X}| \frac{q}{q+\alpha}.$
- (2) $\mathbb{E}_q [|\check{\mathcal{X}}|] \geq |\mathcal{X}| \frac{\alpha}{q+2\alpha} m(\mathcal{L}).$
- (3) For any $r \in [0, 1]$, set $\mathcal{R}_r := \{x \in \mathcal{X} \text{ s.t. } w(x) \geq r\alpha\}.$

$$\mathbb{E}_q [|\check{\mathcal{X}} \cap \mathcal{R}_r|] \geq |\mathcal{R}_r| \frac{r\alpha}{q+2\alpha}.$$

Remark. Hence, in a loose sense, $\bar{\mathcal{X}}$ contains a given proportion of points in \mathcal{X} , and $\check{\mathcal{X}}$ contains a given proportion of the “rapid” points in \mathcal{X} , i.e. points for which the rate of escape is high. Since they do not depend on the spectral decomposition of \mathcal{L} , these estimates can be helpful concerning the choice of q . Taking for instance $q \in [2\alpha/3; \alpha]$ ensures that the mean proportion of sampled points is greater than $2/5$, and that the mean proportion of decimated points is greater than $m(\mathcal{L})/3$.

The proof of Proposition 6 relies on the following lemma.

Lemma 7. For any $x \in \mathcal{X}$, and any $q > 0$,

$$K_q(x, x) \geq \frac{q}{q + w(x)}. \quad (22)$$

$$1 - K_q(x, x) \geq \frac{w(x)}{q + \lambda_{n-1}} \geq \frac{w(x)}{q + 2\alpha}. \quad (23)$$

Proof. Remind that $(e_i, i = 0, \dots, n-1)$ is an orthonormal basis of eigenfunctions of $-\mathcal{L}$ in $\ell_2(\mathcal{X}, \mu)$, and let $(\mathbf{1}_x, x \in \mathcal{X})$ be the canonical basis of $\mathbb{R}^{\mathcal{X}}$ ($\mathbf{1}_x(y) = 0$ for $y \neq x$ and $\mathbf{1}_x(x) = 1$). Note that

$$w(x) = -\mathcal{L}(x, x) = -\mathcal{L}\mathbf{1}_x(x) = \sum_{i=0}^{n-1} \lambda_i \langle e_i, \mathbf{1}_x \rangle e_i(x) = \sum_{i=0}^{n-1} \lambda_i e_i(x)^2 \mu(x).$$

In the same way,

$$K_q(x, x) = \sum_{i=0}^{n-1} \frac{q}{q + \lambda_i} e_i(x)^2 \mu(x).$$

Note that for any $x \in \mathcal{X}$,

$$\sum_{i=0}^{n-1} e_i(x)^2 \mu(x) = \sum_{i=0}^{n-1} \langle \mathbf{1}_x, e_i \rangle^2 \frac{1}{\mu(x)} = \frac{\|\mathbf{1}_x\|^2}{\mu(x)} = 1;$$

Hence, for any $x \in \mathcal{X}$, $(e_i^2(x)\mu(x); i = 0, \dots, n-1)$ is a probability distribution on $\{0, 1, \dots, n-1\}$. The function $\lambda \in \mathbb{R}^+ \mapsto q/(q + \lambda)$ being convex, it follows from Jensen's inequality, that for any $x \in \mathcal{X}$,

$$K_q(x, x) \geq \frac{q}{q + \sum_{i=0}^{n-1} \lambda_i e_i(x)^2 \mu(x)} = \frac{q}{q + w(x)}.$$

Moreover,

$$\begin{aligned} 1 - K_q(x, x) &= \sum_{i=0}^{n-1} \frac{\lambda_i}{q + \lambda_i} e_i(x)^2 \mu(x), \\ &\geq \frac{1}{q + \lambda_{n-1}} \sum_{i=0}^{n-1} \lambda_i e_i(x)^2 \mu(x), \\ &= \frac{w(x)}{q + \lambda_{n-1}}. \end{aligned}$$

This ends the proof of Lemma 7, since $\lambda_{n-1} \leq 2\alpha$. \square

We return now to the proof of Proposition 6.

Proof. (1) It follows from Proposition 4 that $\mathbb{E}_q[|\check{\mathcal{X}}|] = \sum_{x \in \mathcal{X}} \mathbb{P}_q[x \in \rho(\Phi)] = \sum_{x \in \mathcal{X}} K_q(x, x)$. Using (22) and the definition of α , we obtain that $\mathbb{E}_q[|\check{\mathcal{X}}|] \geq \frac{q}{q+\alpha} |\mathcal{X}|$.

(2) By Proposition 3, $\mathbb{E}_q[|\check{\mathcal{X}}|] = \sum_{i=0}^{n-1} \frac{\lambda_i}{q+\lambda_i} \geq \frac{1}{q+\lambda_{n-1}} \text{Trace}(-\mathcal{L})$.

(3) By (23), $\mathbb{E}_q[|\check{\mathcal{X}} \cap \mathcal{R}_r|] = \sum_{x \in \mathcal{R}_r} \mathbb{P}_q[x \notin \rho(\Phi)] = \sum_{x \in \mathcal{R}_r} (1 - K_q(x, x)) \geq \sum_{x \in \mathcal{R}_r} \frac{w(x)}{q+\lambda_{n-1}} \geq |\mathcal{R}_r| \frac{r\alpha}{q+\lambda_{n-1}}$. \square

4. The weighting and filtering procedures

This section aims to describe the weighting and the filtering procedures we use in the multiresolution scheme, once $\check{\mathcal{X}}$ has been chosen. Hence, we will assume throughout this section that $\check{\mathcal{X}}$ is any proper subset of \mathcal{X} .

4.1. The weighting procedure

Let us define for any $\bar{x}, \bar{y} \in \bar{\mathcal{X}}$,

$$\tilde{P}(\bar{x}, \bar{y}) := P_{\bar{x}}[X(H_{\check{\mathcal{X}}}^+) = \bar{y}]. \quad (24)$$

\tilde{P} is a stochastic matrix on $\bar{\mathcal{X}}$, and one can associate to \tilde{P} a weight function \bar{w} and a Markov generator $\bar{\mathcal{L}}$ on $\bar{\mathcal{X}}$, which are unique up to multiplicative constants. The next lemma explains how to compute $\bar{\mathcal{L}}$ from \mathcal{L} .

Lemma 8. *Let $\bar{\mathcal{X}}$ be any proper subset of \mathcal{X} , and set $\check{\mathcal{X}} = \mathcal{X} \setminus \bar{\mathcal{X}}$. Let \tilde{P} be the stochastic matrix on $\bar{\mathcal{X}}$ defined by (24). Then, $\mathcal{L}_{\check{\mathcal{X}}\check{\mathcal{X}}}$ is invertible. Let $\bar{\mathcal{L}}$ be the Schur complement of $\mathcal{L}_{\check{\mathcal{X}}\check{\mathcal{X}}}$ in \mathcal{L} . $\bar{\mathcal{L}}$ is an irreducible Markov generator on $\bar{\mathcal{X}}$, $\tilde{P} = \text{Id} + \bar{\mathcal{L}}/\alpha$, and for any $\bar{x}, \bar{y} \in \bar{\mathcal{X}}$, $\mu(\bar{x})\bar{\mathcal{L}}(\bar{x}, \bar{y}) = \mu(\bar{y})\bar{\mathcal{L}}(\bar{y}, \bar{x})$.*

Proof. Let $Q(x, \bar{x}) := P_x \left[X(H_{\bar{\mathcal{X}}}^+) = \bar{x} \right]$, so that $\tilde{P} = Q_{\bar{\mathcal{X}}\bar{\mathcal{X}}}$. Using Markov property at time τ_1 leads to

$$Q(x, \bar{x}) = \sum_{z \in \mathcal{X}} P(x, z) P_z [X(H_{\bar{\mathcal{X}}}) = \bar{x}] = P(x, \bar{x}) + \sum_{z \in \check{\mathcal{X}}} P(x, z) Q(z, \bar{x}).$$

Hence, since $P = \text{Id} + \mathcal{L}/\alpha$,

$$\begin{cases} Q_{\bar{\mathcal{X}}\bar{\mathcal{X}}} &= P_{\bar{\mathcal{X}}\bar{\mathcal{X}}} + P_{\bar{\mathcal{X}}\check{\mathcal{X}}} Q_{\check{\mathcal{X}}\bar{\mathcal{X}}} \\ Q_{\check{\mathcal{X}}\bar{\mathcal{X}}} &= P_{\check{\mathcal{X}}\bar{\mathcal{X}}} + P_{\check{\mathcal{X}}\check{\mathcal{X}}} Q_{\check{\mathcal{X}}\bar{\mathcal{X}}} \end{cases} \iff \begin{cases} \tilde{P} &= P_{\bar{\mathcal{X}}\bar{\mathcal{X}}} + P_{\bar{\mathcal{X}}\check{\mathcal{X}}} Q_{\check{\mathcal{X}}\bar{\mathcal{X}}} \\ -\mathcal{L}_{\check{\mathcal{X}}\check{\mathcal{X}}} Q_{\check{\mathcal{X}}\bar{\mathcal{X}}} &= \mathcal{L}_{\check{\mathcal{X}}\bar{\mathcal{X}}} \end{cases}. \quad (25)$$

We are now going to prove that $-\mathcal{L}_{\check{\mathcal{X}}\check{\mathcal{X}}}$ is invertible. Let $\phi_{\check{\mathcal{X}}}$ be a vector of $\text{Ker}(-\mathcal{L}_{\check{\mathcal{X}}\check{\mathcal{X}}})$, and let $\psi := (0_{\bar{\mathcal{X}}}, \phi_{\check{\mathcal{X}}})$. We get on one hand $\langle \psi, -\mathcal{L}\psi \rangle = 0$. On the other hand, $\langle \psi, -\mathcal{L}\psi \rangle = \sum_{i=0}^{n-1} \lambda_i \langle \psi, e_i \rangle^2$. Therefore, for any $i \in \{0, \dots, n-1\}$, $\lambda_i \langle \psi, e_i \rangle^2 = 0$. $\lambda_0 = 0$ being simple, $\langle \psi, e_i \rangle = 0$ for all $i \geq 1$. Hence, ψ is colinear to $e_0 = \mathbf{1}$. If $\check{\mathcal{X}} \neq \emptyset$ and $\bar{\mathcal{X}} \neq \emptyset$, this implies that $\phi_{\check{\mathcal{X}}} \equiv 0$. Hence $-\mathcal{L}_{\check{\mathcal{X}}\check{\mathcal{X}}}$ is invertible. Going back to (25), we obtain that

$$Q_{\bar{\mathcal{X}}\bar{\mathcal{X}}} = (-\mathcal{L}_{\check{\mathcal{X}}\check{\mathcal{X}}})^{-1} \mathcal{L}_{\check{\mathcal{X}}\bar{\mathcal{X}}}, \quad (26)$$

and

$$\tilde{P} = \text{Id}_{\bar{\mathcal{X}}} + \frac{1}{\alpha} \mathcal{L}_{\bar{\mathcal{X}}\bar{\mathcal{X}}} - \frac{1}{\alpha} \mathcal{L}_{\bar{\mathcal{X}}\check{\mathcal{X}}} (\mathcal{L}_{\check{\mathcal{X}}\check{\mathcal{X}}})^{-1} \mathcal{L}_{\check{\mathcal{X}}\bar{\mathcal{X}}} = \text{Id}_{\bar{\mathcal{X}}} + \frac{\bar{\mathcal{L}}}{\alpha}. \quad (27)$$

Therefore, $\forall \bar{x}, \bar{y} \in \bar{\mathcal{X}}, \bar{x} \neq \bar{y}, \bar{\mathcal{L}}(\bar{x}, \bar{y}) = \alpha \tilde{P}(\bar{x}, \bar{y}) \geq 0$, and $\bar{\mathcal{L}}\mathbf{1}_{\bar{\mathcal{X}}} = 0$. $\bar{\mathcal{L}}$ is thus a Markov generator on $\bar{\mathcal{X}}$. The fact that $\bar{\mathcal{L}}$ is reversible w.r.t. μ is a direct consequence of the fact that $\bar{\mathcal{L}}$ is a Schur complement, and of the reversibility of \mathcal{L} w.r.t. μ . Concerning the irreducibility statement, if \bar{x} and \bar{y} are two distinct elements of $\bar{\mathcal{X}}$, the irreducibility of \mathcal{L} implies that there exists a path in \mathcal{X} going from \bar{x} to \bar{y} with positive probability for P . Sampling this path on the passage times on $\bar{\mathcal{X}}$, we obtain a path in $\bar{\mathcal{X}}$ going from \bar{x} to \bar{y} with positive probability for \tilde{P} . \square

Once $\bar{\mathcal{L}}$ is defined, \bar{w} and $\bar{\alpha}$ are defined in the same way that w and α were defined from \mathcal{L} , i.e.

$$\forall \bar{x} \in \bar{\mathcal{X}}, \bar{w}(\bar{x}) := -\bar{\mathcal{L}}(\bar{x}, \bar{x}); \bar{\alpha} := \max_{\bar{x} \in \bar{\mathcal{X}}} \bar{w}(\bar{x}). \quad (28)$$

4.2. The filtering procedure

We have now to define the “approximation” and the “detail” of a given function f on \mathcal{X} , and to say how to reconstruct f from the approximation and the detail. This is the aim of this section.

4.2.1. The analysis operator

Let $q' > 0$ be fixed and define the collection $(\nu_{\bar{x}}, \bar{x} \in \bar{\mathcal{X}})$ of probability measures on \mathcal{X} by

$$\nu_{\bar{x}}(x) = K_{q'}(\bar{x}, x).$$

Let $(\rho_{\check{x}}, \check{x} \in \check{\mathcal{X}})$ be the collection of signed measures on \mathcal{X} defined by

$$\rho_{\check{x}}(x) = (K_{q'} - \text{Id})(\check{x}, x). \quad (29)$$

We associate to the collection of measures $(\nu_{\bar{x}}, \rho_{\check{x}})_{\bar{x} \in \bar{\mathcal{X}}, \check{x} \in \check{\mathcal{X}}}$ the corresponding collection of functions $(\phi_{\bar{x}}, \psi_{\check{x}})_{\bar{x} \in \bar{\mathcal{X}}, \check{x} \in \check{\mathcal{X}}}$:

$$\phi_{\bar{x}}(x) := \nu_{\bar{x}}^*(x) = \frac{\nu_{\bar{x}}(x)}{\mu(x)}; \psi_{\check{x}}(x) := \rho_{\check{x}}^*(x) = \frac{\rho_{\check{x}}(x)}{\mu(x)}. \quad (30)$$

The functions $(\psi_{\check{x}}, \check{x} \in \check{\mathcal{X}})$ play the role of “wavelets” in our proposition of a multiresolution analysis. Note that the functions $\phi_{\bar{x}}$ are naturally normalized in $\ell_1(\mathcal{X}, \mu)$, but this is not the case for the functions $\psi_{\check{x}}$ ($\|\psi_{\check{x}}\|_1 = 2(1 - K_{q'}(\check{x}, \check{x}))$). Note also that for any $\check{x} \in \check{\mathcal{X}}$, $\langle \psi_{\check{x}}, \mathbf{1} \rangle = 0$. In addition, we get the following result:

Lemma 9. *The family $(\phi_{\bar{x}}, \psi_{\check{x}})_{\bar{x} \in \bar{\mathcal{X}}, \check{x} \in \check{\mathcal{X}}}$ is a basis of $\ell_2(\mathcal{X}, \mu)$.*

Proof. Since $\mu(x) > 0$ for all $x \in \mathcal{X}$, it is equivalent to prove that the matrix M whose row vectors are given by $(\nu_{\bar{x}}, \rho_{\check{x}})_{\bar{x} \in \bar{\mathcal{X}}, \check{x} \in \check{\mathcal{X}}}$ is invertible. M can be rewritten in terms of its block matrices according to the rows and columns indexed by $\bar{\mathcal{X}}$ and $\check{\mathcal{X}}$:

$$M = \begin{pmatrix} K_{q'}|_{\bar{\mathcal{X}}, \bar{\mathcal{X}}} & K_{q'}|_{\bar{\mathcal{X}}, \check{\mathcal{X}}} \\ K_{q'}|_{\check{\mathcal{X}}, \bar{\mathcal{X}}} & K_{q'}|_{\check{\mathcal{X}}, \check{\mathcal{X}}} - \text{Id}_{\check{\mathcal{X}}} \end{pmatrix}.$$

Note that the eigenvalues of $K_{q'} - \text{Id}$ are the non positive real numbers $(-\frac{\lambda_j}{q' + \lambda_j}; j = 0, \dots, n-1)$. Therefore $\lambda_{\max}(K_{q'}|_{\check{\mathcal{X}}, \check{\mathcal{X}}} - \text{Id}_{\check{\mathcal{X}}}) \leq \lambda_{\max}(K_{q'} - \text{Id}) = 0$. Equality holds iff there exists an eigenfunction of $K_{q'} - \text{Id}$ associated to the eigenvalue 0, which is contained in $\text{Span}(\mathbf{1}_{\{\check{x}\}}; \check{x} \in \check{\mathcal{X}})$. But, 0 is a simple eigenvalue of $K_{q'} - \text{Id}$ associated to the function $\mathbf{1}$. Since $\check{\mathcal{X}} \neq \mathcal{X}$, we get that $\lambda_{\max}(K_{q'}|_{\check{\mathcal{X}}, \check{\mathcal{X}}} - \text{Id}_{\check{\mathcal{X}}}) < 0$. Thus, $K_{q'}|_{\check{\mathcal{X}}, \check{\mathcal{X}}} - \text{Id}_{\check{\mathcal{X}}}$ is invertible and by Proposition 1,

$$\det(M) = \det(K_{q'}|_{\check{\mathcal{X}}, \check{\mathcal{X}}} - \text{Id}_{\check{\mathcal{X}}}) \det(S_M(K_{q'}|_{\check{\mathcal{X}}, \check{\mathcal{X}}} - \text{Id}_{\check{\mathcal{X}}})) \quad (31)$$

Concerning the Schur complement in (31), note that

$$\begin{aligned} S_M(K_{q'}|_{\check{\mathcal{X}}, \check{\mathcal{X}}} - \text{Id}_{\check{\mathcal{X}}}) &= K_{q'}|_{\bar{\mathcal{X}}, \bar{\mathcal{X}}} - K_{q'}|_{\bar{\mathcal{X}}, \check{\mathcal{X}}}(K_{q'}|_{\check{\mathcal{X}}, \check{\mathcal{X}}} - \text{Id}_{\check{\mathcal{X}}})^{-1}K_{q'}|_{\check{\mathcal{X}}, \bar{\mathcal{X}}} \\ &= \text{Id}_{\bar{\mathcal{X}}} + S_{K_{q'} - \text{Id}}(K_{q'}|_{\check{\mathcal{X}}, \check{\mathcal{X}}} - \text{Id}_{\check{\mathcal{X}}}). \end{aligned}$$

$K_{q'} - \text{Id}$ is a symmetric negative operator in $\ell^2(\mathcal{X}, \mu)$ such that $\lambda_{\min}(K_{q'} - \text{Id}) = -\frac{\lambda_{n-1}}{q' + \lambda_{n-1}} > -1$. Using Proposition 1, we deduce that

$$\lambda_{\min}(S_{K_{q'} - \text{Id}}(K_{q'}|_{\check{\mathcal{X}}, \check{\mathcal{X}}} - \text{Id}_{\check{\mathcal{X}}})) \geq \lambda_{\min}(K_{q'} - \text{Id}) > -1.$$

Hence $\det(S_M(K_{q'}|_{\check{\mathcal{X}}, \check{\mathcal{X}}} - \text{Id}_{\check{\mathcal{X}}})) > 0$. We have thus proven that $\det(M) \neq 0$. \square

Once we have defined a basis $(\xi_x)_{x \in \mathcal{X}} := (\phi_{\bar{x}}, \psi_{\check{x}})_{\bar{x} \in \bar{\mathcal{X}}, \check{x} \in \check{\mathcal{X}}}$ of $\ell_2(\mathcal{X}, \mu)$, we are able to define the analysis operator

$$\begin{aligned} U : \ell_2(\mathcal{X}, \mu) &\rightarrow \mathbb{R}^{\mathcal{X}} \\ f &\mapsto (\langle \xi_x, f \rangle)_{x \in \mathcal{X}} = (\langle \phi_{\bar{x}}, f \rangle, \langle \psi_{\check{x}}, f \rangle)_{\bar{x} \in \bar{\mathcal{X}}, \check{x} \in \check{\mathcal{X}}} \end{aligned}.$$

The hope is that when f is “regular” (a notion still to be defined), the coefficients $(\langle f, \psi_{\check{x}} \rangle)_{\check{x} \in \check{\mathcal{X}}}$ are “small”. Actually, when f is a constant function, these coefficients are equal to zero. We stress the fact that these coefficients are not normalized. f can be reconstructed from the coefficients $U(f)$. Let $(\tilde{\xi}_x)_{x \in \mathcal{X}}$ be the dual basis of $(\xi_x)_{x \in \mathcal{X}}$ in $\ell_2(\mathcal{X}, \mu)$, defined by $\langle \tilde{\xi}_x, \xi_y \rangle = \delta_{xy}$. Then we get the obvious reconstruction formula

$$f = \sum_{x \in \mathcal{X}} \langle \xi_x, f \rangle \tilde{\xi}_x.$$

The approximation of f is then $\sum_{\bar{y} \in \bar{\mathcal{X}}} \langle \phi_{\bar{y}}, f \rangle \tilde{\xi}_{\bar{y}}$, while the detail of f is $\sum_{\check{y} \in \check{\mathcal{X}}} \langle \psi_{\check{y}}, f \rangle \tilde{\xi}_{\check{y}}$. The reconstruction of f from its coefficients $U(f)$ can here be made explicit. This is the content of the following proposition.

4.2.2. The reconstruction formula

Proposition 10 (Reconstruction formula). *For any $f \in \ell_2(\mathcal{X}, \mu)$, let $\bar{f} \in \ell_2(\bar{\mathcal{X}}, \mu)$ and $\check{f} \in \ell_2(\check{\mathcal{X}}, \mu)$ be defined by*

$$\begin{aligned} \forall \bar{x} \in \bar{\mathcal{X}}, \quad \bar{f}(\bar{x}) &= K_{q'} f(\bar{x}) = U(f)(\bar{x}); \\ \forall \check{x} \in \check{\mathcal{X}}, \quad \check{f}(\check{x}) &= (K_{q'} - \text{Id}) f(\check{x}) = U(f)(\check{x}). \end{aligned}$$

Then,

$$f = \begin{pmatrix} \text{Id}_{\bar{\mathcal{X}}} - \frac{1}{q'} \bar{\mathcal{L}} & \mathcal{L}_{\bar{\mathcal{X}}\check{\mathcal{X}}}(-\mathcal{L}_{\check{\mathcal{X}}\check{\mathcal{X}}})^{-1} \\ (-\mathcal{L}_{\check{\mathcal{X}}\check{\mathcal{X}}})^{-1} \mathcal{L}_{\check{\mathcal{X}}\bar{\mathcal{X}}} & q' \mathcal{L}_{\check{\mathcal{X}}\check{\mathcal{X}}}^{-1} - \text{Id}_{\check{\mathcal{X}}} \end{pmatrix} \begin{pmatrix} \bar{f} \\ \check{f} \end{pmatrix} = \bar{R} \bar{f} + \check{R} \check{f},$$

where

$$\bar{R} = \begin{pmatrix} \text{Id}_{\bar{\mathcal{X}}} - \frac{1}{q'} \bar{\mathcal{L}} \\ (-\mathcal{L}_{\check{\mathcal{X}}\check{\mathcal{X}}})^{-1} \mathcal{L}_{\check{\mathcal{X}}\bar{\mathcal{X}}} \end{pmatrix}, \quad \text{and} \quad \check{R} = \begin{pmatrix} \mathcal{L}_{\bar{\mathcal{X}}\check{\mathcal{X}}}(-\mathcal{L}_{\check{\mathcal{X}}\check{\mathcal{X}}})^{-1} \\ -\text{Id}_{\check{\mathcal{X}}} - q'(-\mathcal{L}_{\check{\mathcal{X}}\check{\mathcal{X}}})^{-1} \end{pmatrix}. \quad (32)$$

Proof. Note that $\begin{pmatrix} \bar{f} \\ \check{f} \end{pmatrix} = \left[K_{q'} - \begin{pmatrix} 0 & 0 \\ 0 & \text{Id}_{\check{\mathcal{X}}} \end{pmatrix} \right] f$. Since $K_{q'} = (\text{Id} - \frac{1}{q'} \mathcal{L})^{-1}$, we get

$$\left(\text{Id} - \frac{1}{q'} \mathcal{L} \right) \begin{pmatrix} \bar{f} \\ \check{f} \end{pmatrix} = \left[\text{Id} - \left(\text{Id} - \frac{1}{q'} \mathcal{L} \right) \begin{pmatrix} 0 & 0 \\ 0 & \text{Id}_{\check{\mathcal{X}}} \end{pmatrix} \right] f = \begin{pmatrix} \text{Id}_{\bar{\mathcal{X}}} & \mathcal{L}_{\bar{\mathcal{X}}\check{\mathcal{X}}}/q' \\ 0 & \mathcal{L}_{\check{\mathcal{X}}\check{\mathcal{X}}}/q' \end{pmatrix} f.$$

The inverse of $\begin{pmatrix} \text{Id}_{\bar{\mathcal{X}}} & \mathcal{L}_{\bar{\mathcal{X}}\check{\mathcal{X}}}/q' \\ 0 & \mathcal{L}_{\check{\mathcal{X}}\check{\mathcal{X}}}/q' \end{pmatrix}$ is $\begin{pmatrix} \text{Id}_{\bar{\mathcal{X}}} & \mathcal{L}_{\bar{\mathcal{X}}\check{\mathcal{X}}}(-\mathcal{L}_{\check{\mathcal{X}}\check{\mathcal{X}}})^{-1} \\ 0 & q'(\mathcal{L}_{\check{\mathcal{X}}\check{\mathcal{X}}})^{-1} \end{pmatrix}$. Hence,

$$\begin{aligned} f &= \begin{pmatrix} \text{Id}_{\bar{\mathcal{X}}} & \mathcal{L}_{\bar{\mathcal{X}}\check{\mathcal{X}}}(-\mathcal{L}_{\check{\mathcal{X}}\check{\mathcal{X}}})^{-1} \\ 0 & q'(\mathcal{L}_{\check{\mathcal{X}}\check{\mathcal{X}}})^{-1} \end{pmatrix} \begin{pmatrix} \text{Id}_{\bar{\mathcal{X}}} - \frac{1}{q'} \mathcal{L}_{\bar{\mathcal{X}}\bar{\mathcal{X}}} & -\frac{1}{q'} \mathcal{L}_{\bar{\mathcal{X}}\check{\mathcal{X}}} \\ -\frac{1}{q'} \mathcal{L}_{\check{\mathcal{X}}\bar{\mathcal{X}}} & \text{Id}_{\check{\mathcal{X}}} - \frac{1}{q'} \mathcal{L}_{\check{\mathcal{X}}\check{\mathcal{X}}} \end{pmatrix} \begin{pmatrix} \bar{f} \\ \check{f} \end{pmatrix}, \\ &= \begin{pmatrix} \text{Id}_{\bar{\mathcal{X}}} - \frac{1}{q'} \bar{\mathcal{L}} & \mathcal{L}_{\bar{\mathcal{X}}\check{\mathcal{X}}}(-\mathcal{L}_{\check{\mathcal{X}}\check{\mathcal{X}}})^{-1} \\ (-\mathcal{L}_{\check{\mathcal{X}}\check{\mathcal{X}}})^{-1} \mathcal{L}_{\check{\mathcal{X}}\bar{\mathcal{X}}} & q'(\mathcal{L}_{\check{\mathcal{X}}\check{\mathcal{X}}})^{-1} - \text{Id}_{\check{\mathcal{X}}} \end{pmatrix} \begin{pmatrix} \bar{f} \\ \check{f} \end{pmatrix}. \quad \square \end{aligned}$$

In the sequel, we will talk about \bar{R} as the approximation operator, and about \check{R} as the detail operator. In order to get stable numerical results in the analysis and the reconstruction of the signal f , it is important to control the operator norm of these operators. This is easy for the analysis operator U since for any $f \in \ell_p(\mathcal{X}, \mu)$,

$$\|U(f)\|_{p,\mathcal{X}} = \|K_{q'}(f) - f\mathbf{1}_{\check{\mathcal{X}}}\|_{p,\mathcal{X}} \leq \|K_{q'}(f)\|_{p,\mathcal{X}} + \|f\|_{p,\mathcal{X}}.$$

$K_{q'}$ being a probability kernel, symmetric w.r.t. μ , we get by Jensen's inequality,

$$\|K_{q'}f\|_{p,\mathcal{X}}^p = \sum_{x \in \mathcal{X}} \mu(x) |K_{q'}f(x)|^p \quad (33)$$

$$\leq \sum_{x \in \mathcal{X}} \mu(x) K_{q'}(|f|^p)(x) = \langle \mathbf{1}, K_{q'}(|f|^p) \rangle = \langle K_{q'}(\mathbf{1}), |f|^p \rangle = \|f\|_{p,\mathcal{X}}^p. \quad (34)$$

Hence for any $f \in \ell_p(\mathcal{X}, \mu)$,

$$\|U(f)\|_{p,\mathcal{X}} \leq 2\|f\|_{p,\mathcal{X}}. \quad (35)$$

The aim of the two following sections is to provide bounds on the norm of the reconstruction operator.

4.3. Detail operator norm

We give in this section a control on the norm of the detail operator \check{R} , in terms of β and γ defined in (9). Using Markov property at time τ_1 , one can express $1/\beta$ as

$$\frac{1}{\beta} = \max_{\bar{x} \in \check{\mathcal{X}}} E_{\bar{x}} \left[H_{\bar{\mathcal{X}}}^+ - \tau_1 \right].$$

In the sequel, when A is a subset of \mathcal{X} , and f is a function on A , $\|f\|_{p,A}$ is the norm of f in $\ell_p(A, \mu_A)$, where μ_A is the conditional probability μ on A : for any $B \subset A$, $\mu_A(B) = \mu(B)/\mu(A)$.

Proposition 11 (Detail operator norm). *Let $\bar{\mathcal{X}}$ be any proper subset of \mathcal{X} , $\check{\mathcal{X}} = \mathcal{X} \setminus \bar{\mathcal{X}}$, and let \check{R} be the operator defined in (10). For any $p \geq 1$, for any $f \in \ell_p(\check{\mathcal{X}}, \mu_{\check{\mathcal{X}}})$,*

$$\|\check{R}f\|_{p,\mathcal{X}} \leq \left[\left(\frac{\alpha}{\beta} \right)^{p/p^*} + \left(1 + \frac{q'}{\gamma} \right)^p \right]^{1/p} \mu(\check{\mathcal{X}})^{1/p} \|f\|_{p,\check{\mathcal{X}}}, \quad (36)$$

where p^* is the conjugate exponent of p .

The proof is a consequence of Lemmas 12–13 given below.

Lemma 12. *For any $x \in \mathcal{X}$, and $\check{x} \in \check{\mathcal{X}}$, we define*

$$G_{\bar{\mathcal{X}}}(x, \check{x}) = E_x \left[\int_0^{H_{\check{\mathcal{X}}}^+} \mathbf{1}_{X(s)=\check{x}} ds \right].$$

Then,

$$(-\mathcal{L}_{\check{\mathcal{X}}\check{\mathcal{X}}})^{-1} = (G_{\bar{\mathcal{X}}})_{\check{\mathcal{X}}\check{\mathcal{X}}} \text{ , and } \mathcal{L}_{\bar{\mathcal{X}}\bar{\mathcal{X}}}(-\mathcal{L}_{\check{\mathcal{X}}\check{\mathcal{X}}})^{-1} = \alpha(G_{\bar{\mathcal{X}}})_{\bar{\mathcal{X}}\bar{\mathcal{X}}}. \quad (37)$$

The adjoint operator of

$$(-\mathcal{L}_{\check{\mathcal{X}}\check{\mathcal{X}}})^{-1} : \ell_p(\check{\mathcal{X}}, \mu) \mapsto \ell_q(\check{\mathcal{X}}, \mu)$$

is the operator

$$(-\mathcal{L}_{\check{\mathcal{X}}\check{\mathcal{X}}})^{-1} : \ell_{q^*}(\check{\mathcal{X}}, \mu) \mapsto \ell_{p^*}(\check{\mathcal{X}}, \mu).$$

The adjoint operator of

$$\mathcal{L}_{\bar{\mathcal{X}}\check{\mathcal{X}}}(-\mathcal{L}_{\check{\mathcal{X}}\check{\mathcal{X}}})^{-1} : \ell_p(\check{\mathcal{X}}, \mu) \mapsto \ell_q(\bar{\mathcal{X}}, \mu)$$

is the operator

$$(-\mathcal{L}_{\check{\mathcal{X}}\check{\mathcal{X}}})^{-1} \mathcal{L}_{\check{\mathcal{X}}\bar{\mathcal{X}}} : \ell_{q^*}(\bar{\mathcal{X}}, \mu) \mapsto \ell_{p^*}(\check{\mathcal{X}}, \mu),$$

with

$$(-\mathcal{L}_{\check{\mathcal{X}}\check{\mathcal{X}}})^{-1} \mathcal{L}_{\check{\mathcal{X}}\bar{\mathcal{X}}}(\check{x}, \bar{x}) = P_{\bar{x}}[X(H_{\bar{\mathcal{X}}}) = \bar{x}] , \check{x} \in \check{\mathcal{X}}, \bar{x} \in \bar{\mathcal{X}}. \quad (38)$$

Proof.

$$\begin{aligned} G_{\bar{\mathcal{X}}}(x, \check{x}) &= E_x \left[\int_0^{\tau_1} \mathbf{1}_{X(s)=\check{x}} ds + \int_{\tau_1}^{H_{\bar{\mathcal{X}}}^+} \mathbf{1}_{X(s)=\check{x}} ds \right], \\ &= \delta_x(\check{x}) E_x(\tau_1) + \sum_{z \in \check{\mathcal{X}}} P(x, z) E_z \left[\int_0^{H_{\bar{\mathcal{X}}}} \mathbf{1}_{X(s)=\check{x}} ds \right] \text{ by Markov property at time } \tau_1, \\ &= \frac{1}{\alpha} \delta_x(\check{x}) + \sum_{z \in \check{\mathcal{X}}} P(x, z) E_z \left[\int_0^{H_{\bar{\mathcal{X}}}^+} \mathbf{1}_{X(s)=\check{x}} ds \right]. \end{aligned}$$

This can be rewritten as

$$\begin{aligned} \begin{cases} (G_{\bar{\mathcal{X}}})_{\bar{\mathcal{X}}\check{\mathcal{X}}} = P_{\bar{\mathcal{X}}\check{\mathcal{X}}}(G_{\bar{\mathcal{X}}})_{\check{\mathcal{X}}\check{\mathcal{X}}} \\ (G_{\bar{\mathcal{X}}})_{\check{\mathcal{X}}\check{\mathcal{X}}} = \frac{1}{\alpha} \text{Id}_{\check{\mathcal{X}}} + P_{\check{\mathcal{X}}\check{\mathcal{X}}}(G_{\bar{\mathcal{X}}})_{\check{\mathcal{X}}\check{\mathcal{X}}} \end{cases} &\iff \begin{cases} -\mathcal{L}_{\check{\mathcal{X}}\check{\mathcal{X}}}(G_{\bar{\mathcal{X}}})_{\check{\mathcal{X}}\check{\mathcal{X}}} = \text{Id}_{\check{\mathcal{X}}} \\ (G_{\bar{\mathcal{X}}})_{\bar{\mathcal{X}}\check{\mathcal{X}}} = \frac{1}{\alpha} \mathcal{L}_{\bar{\mathcal{X}}\check{\mathcal{X}}}(G_{\bar{\mathcal{X}}})_{\check{\mathcal{X}}\check{\mathcal{X}}} \end{cases} \\ &\iff \begin{cases} (-\mathcal{L}_{\check{\mathcal{X}}\check{\mathcal{X}}})^{-1} = (G_{\bar{\mathcal{X}}})_{\check{\mathcal{X}}\check{\mathcal{X}}} \\ \mathcal{L}_{\bar{\mathcal{X}}\check{\mathcal{X}}}(-\mathcal{L}_{\check{\mathcal{X}}\check{\mathcal{X}}})^{-1} = \alpha(G_{\bar{\mathcal{X}}})_{\bar{\mathcal{X}}\check{\mathcal{X}}} \end{cases}. \end{aligned}$$

Using the symmetry of \mathcal{L} with respect to μ , for any $g \in \ell_{q^*}(\check{\mathcal{X}}, \mu)$ and any $f \in \ell_p(\check{\mathcal{X}}, \mu)$,

$$\begin{aligned} \langle (-\mathcal{L}_{\check{\mathcal{X}}\check{\mathcal{X}}})^{-1} f, g \rangle_{\check{\mathcal{X}}} &= \sum_{\check{x}, \check{y} \in \check{\mathcal{X}}} \mu(\check{x}) (-\mathcal{L}_{\check{\mathcal{X}}\check{\mathcal{X}}})^{-1}(\check{x}, \check{y}) f(\check{y}) g(\check{x}) \\ &= \sum_{\check{x}, \check{y} \in \check{\mathcal{X}}} \mu(\check{y}) (-\mathcal{L}_{\check{\mathcal{X}}\check{\mathcal{X}}})^{-1}(\check{y}, \check{x}) f(\check{y}) g(\check{x}) = \langle f, (-\mathcal{L}_{\check{\mathcal{X}}\check{\mathcal{X}}})^{-1} g \rangle_{\check{\mathcal{X}}}. \end{aligned}$$

In the same way, for all $g \in \ell_{q^*}(\bar{\mathcal{X}}, \mu)$ and all $f \in \ell_p(\check{\mathcal{X}}, \mu)$,

$$\begin{aligned}
\langle \mathcal{L}_{\bar{\mathcal{X}}\check{\mathcal{X}}}(-\mathcal{L}_{\check{\mathcal{X}}\check{\mathcal{X}}})^{-1}f, g \rangle_{\bar{\mathcal{X}}} &= \sum_{\bar{x} \in \bar{\mathcal{X}}; \check{y}, \check{z} \in \check{\mathcal{X}}} \mu(\bar{x}) \mathcal{L}(\bar{x}, \check{y})(-\mathcal{L}_{\check{\mathcal{X}}\check{\mathcal{X}}})^{-1}(\check{y}, \check{z}) f(\check{z}) g(\bar{x}) \\
&= \sum_{\bar{x} \in \bar{\mathcal{X}}; \check{y}, \check{z} \in \check{\mathcal{X}}} \mu(\check{z}) \mathcal{L}(\check{y}, \bar{x})(-\mathcal{L}_{\check{\mathcal{X}}\check{\mathcal{X}}})^{-1}(\check{z}, \check{y}) f(\check{z}) g(\bar{x}) \\
&= \langle f, (-\mathcal{L}_{\check{\mathcal{X}}\check{\mathcal{X}}})^{-1} \mathcal{L}_{\check{\mathcal{X}}\bar{\mathcal{X}}} g \rangle_{\check{\mathcal{X}}} .
\end{aligned}$$

(38) was already stated in the proof of Lemma 8 (see (26)). \square

We are now able to give bounds on the norms of the operators involved in the definition of \bar{R} and \check{R} .

Lemma 13. *Let $\bar{\mathcal{X}}$ be any proper subset of \mathcal{X} , $\check{\mathcal{X}} = \mathcal{X} \setminus \bar{\mathcal{X}}$. β and γ being defined by (9), we get*

$$\|(-\mathcal{L}_{\check{\mathcal{X}}\check{\mathcal{X}}})^{-1} \mathcal{L}_{\check{\mathcal{X}}\bar{\mathcal{X}}} f\|_{p, \check{\mathcal{X}}} \leq \left(\frac{\alpha}{\beta} \frac{\mu(\bar{\mathcal{X}})}{\mu(\check{\mathcal{X}})} \right)^{1/p} \|f\|_{p, \bar{\mathcal{X}}} . \quad (39)$$

$$\|\mathcal{L}_{\bar{\mathcal{X}}\check{\mathcal{X}}}(-\mathcal{L}_{\check{\mathcal{X}}\check{\mathcal{X}}})^{-1} f\|_{p, \bar{\mathcal{X}}} \leq \left(\frac{\alpha}{\beta} \right)^{1/p^*} \left(\frac{\mu(\check{\mathcal{X}})}{\mu(\bar{\mathcal{X}})} \right)^{1/p} \|f\|_{p, \check{\mathcal{X}}} . \quad (40)$$

$$\|(-\mathcal{L}_{\check{\mathcal{X}}\check{\mathcal{X}}})^{-1} f\|_{p, \check{\mathcal{X}}} \leq \frac{1}{\gamma} \|f\|_{p, \check{\mathcal{X}}} . \quad (41)$$

Proof.

$$\begin{aligned}
\mu(\check{\mathcal{X}}) \|(-\mathcal{L}_{\check{\mathcal{X}}\check{\mathcal{X}}})^{-1} \mathcal{L}_{\check{\mathcal{X}}\bar{\mathcal{X}}} f\|_{p, \check{\mathcal{X}}}^p &= \sum_{\check{x} \in \check{\mathcal{X}}} \mu(\check{x}) |(-\mathcal{L}_{\check{\mathcal{X}}\check{\mathcal{X}}})^{-1} \mathcal{L}_{\check{\mathcal{X}}\bar{\mathcal{X}}} f(\check{x})|^p , \\
&= \sum_{\check{x} \in \check{\mathcal{X}}} \mu(\check{x}) |E_{\check{x}} [f(X(H_{\bar{\mathcal{X}}}))]|^p \text{ by (38),} \\
&\leq \sum_{\check{x} \in \check{\mathcal{X}}} \mu(\check{x}) E_{\check{x}} [|f(X(H_{\bar{\mathcal{X}}}))|^p] \text{ by Jensen's inequality,} \\
&= \sum_{\check{x} \in \check{\mathcal{X}}, \bar{y} \in \bar{\mathcal{X}}} \mu(\check{x}) ((-\mathcal{L}_{\check{\mathcal{X}}\check{\mathcal{X}}})^{-1} \mathcal{L}_{\check{\mathcal{X}}\bar{\mathcal{X}}})(\check{x}, \bar{y}) |f|^p(\bar{y}) \text{ by (38),} \\
&= \sum_{\check{x} \in \check{\mathcal{X}}, \bar{y} \in \bar{\mathcal{X}}} \mu(\bar{y}) (\mathcal{L}_{\bar{\mathcal{X}}\check{\mathcal{X}}}(-\mathcal{L}_{\check{\mathcal{X}}\check{\mathcal{X}}})^{-1})(\bar{y}, \check{x}) |f|^p(\bar{y}) \text{ by symmetry,} \\
&= \alpha \sum_{\bar{y} \in \bar{\mathcal{X}}} \mu(\bar{y}) |f(\bar{y})|^p E_{\bar{y}} \left[\int_0^{H_{\check{\mathcal{X}}}^+} \mathbf{1}_{\check{\mathcal{X}}}(X(s)) ds \right] \text{ by (37),} \\
&= \alpha \sum_{\bar{y} \in \bar{\mathcal{X}}} \mu(\bar{y}) |f(\bar{y})|^p E_{\bar{y}} [H_{\check{\mathcal{X}}}^+ - \tau_1] , \\
&\leq \alpha \max_{\bar{x} \in \bar{\mathcal{X}}} E_{\bar{x}} [H_{\check{\mathcal{X}}}^+ - \tau_1] \mu(\bar{\mathcal{X}}) \|f\|_{p, \bar{\mathcal{X}}}^p .
\end{aligned}$$

This gives (39). (40) follows from (39), since the operator $\mathcal{L}_{\bar{\mathcal{X}}\check{\mathcal{X}}}(-\mathcal{L}_{\check{\mathcal{X}}\check{\mathcal{X}}})^{-1}$ from $\ell_p(\check{\mathcal{X}}, \mu)$ to $\ell_p(\bar{\mathcal{X}}, \mu)$ is the adjoint operator of the operator $(-\mathcal{L}_{\check{\mathcal{X}}\check{\mathcal{X}}})^{-1} \mathcal{L}_{\check{\mathcal{X}}\bar{\mathcal{X}}}$ from $\ell_{p^*}(\bar{\mathcal{X}}, \mu)$ to $\ell_{p^*}(\check{\mathcal{X}}, \mu)$. Concerning (41), note that by Lemma 12, for any $\check{x}, \check{y} \in \check{\mathcal{X}}$, $(-\mathcal{L}_{\check{\mathcal{X}}\check{\mathcal{X}}})^{-1}(\check{x}, \check{y}) \geq 0$.

$$\begin{aligned}
\mu(\check{\mathcal{X}}) \|(-\mathcal{L}_{\check{\mathcal{X}}\check{\mathcal{X}}})^{-1} f\|_{p, \check{\mathcal{X}}}^p &= \sum_{\check{x} \in \check{\mathcal{X}}} \mu(\check{x}) |(-\mathcal{L}_{\check{\mathcal{X}}\check{\mathcal{X}}})^{-1} f(\check{x})|^p ,
\end{aligned}$$

$$\begin{aligned}
&= \sum_{\check{x} \in \check{\mathcal{X}}} \mu(\check{x}) \left| \sum_{\check{y} \in \check{\mathcal{X}}} (-\mathcal{L}_{\check{\mathcal{X}}\check{\mathcal{X}}})^{-1}(\check{x}, \check{y}) f(\check{y}) \right|^p, \\
&\leq \sum_{\check{x} \in \check{\mathcal{X}}} \mu(\check{x}) \left(\sum_{\check{y} \in \check{\mathcal{X}}} (-\mathcal{L}_{\check{\mathcal{X}}\check{\mathcal{X}}})^{-1}(\check{x}, \check{y}) \right)^{p/p^*} \left(\sum_{\check{y} \in \check{\mathcal{X}}} (-\mathcal{L}_{\check{\mathcal{X}}\check{\mathcal{X}}})^{-1}(\check{x}, \check{y}) |f(\check{y})|^p \right) \\
&\quad \text{by Hölder's inequality,} \\
&\leq \left(\max_{\check{x} \in \check{\mathcal{X}}} \sum_{\check{y} \in \check{\mathcal{X}}} (-\mathcal{L}_{\check{\mathcal{X}}\check{\mathcal{X}}})^{-1}(\check{x}, \check{y}) \right)^{p/p^*} \left(\sum_{\check{x} \in \check{\mathcal{X}}, \check{y} \in \check{\mathcal{X}}} \mu(\check{x}) (-\mathcal{L}_{\check{\mathcal{X}}\check{\mathcal{X}}})^{-1}(\check{x}, \check{y}) |f(\check{y})|^p \right), \\
&\leq \left(\max_{\check{x} \in \check{\mathcal{X}}} \sum_{\check{y} \in \check{\mathcal{X}}} (-\mathcal{L}_{\check{\mathcal{X}}\check{\mathcal{X}}})^{-1}(\check{x}, \check{y}) \right)^{p/p^*} \left(\sum_{\check{y} \in \check{\mathcal{X}}} \mu(\check{y}) |f(\check{y})|^p \sum_{\check{x} \in \check{\mathcal{X}}} (-\mathcal{L}_{\check{\mathcal{X}}\check{\mathcal{X}}})^{-1}(\check{y}, \check{x}) \right) \\
&\quad \text{by symmetry,} \\
&\leq \left(\max_{\check{x} \in \check{\mathcal{X}}} \sum_{\check{y} \in \check{\mathcal{X}}} (-\mathcal{L}_{\check{\mathcal{X}}\check{\mathcal{X}}})^{-1}(\check{x}, \check{y}) \right)^p \mu(\check{\mathcal{X}}) \|f\|_{p, \check{\mathcal{X}}}^p
\end{aligned}$$

This ends the proof of (41) since by (37),

$$\sum_{\check{y} \in \check{\mathcal{X}}} (-\mathcal{L}_{\check{\mathcal{X}}\check{\mathcal{X}}})^{-1}(\check{x}, \check{y}) = E_{\check{x}} \left[\int_0^{H_{\check{\mathcal{X}}}^+} \mathbf{1}_{\check{\mathcal{X}}}(X(s)) ds \right] = E_{\check{x}} [H_{\check{\mathcal{X}}}^+] = E_{\check{x}} [H_{\check{\mathcal{X}}}] . \quad \square$$

Proof of Proposition 11. For any $f \in \ell_p(\check{\mathcal{X}}, \mu_{\check{\mathcal{X}}})$,

$$\begin{aligned}
\|\check{R}f\|_{p, \mathcal{X}}^p &= \left\| \left(\begin{array}{c} \mathcal{L}_{\bar{\mathcal{X}}\check{\mathcal{X}}} (-\mathcal{L}_{\check{\mathcal{X}}\check{\mathcal{X}}})^{-1} \\ -\text{Id}_{\check{\mathcal{X}}} - q'(-\mathcal{L}_{\check{\mathcal{X}}\check{\mathcal{X}}})^{-1} \end{array} \right) f \right\|_{p, \mathcal{X}}^p \\
&= \mu(\bar{\mathcal{X}}) \|\mathcal{L}_{\bar{\mathcal{X}}\check{\mathcal{X}}} (-\mathcal{L}_{\check{\mathcal{X}}\check{\mathcal{X}}})^{-1} f\|_{p, \bar{\mathcal{X}}}^p + \mu(\check{\mathcal{X}}) \|(\text{Id}_{\check{\mathcal{X}}} + q'(-\mathcal{L}_{\check{\mathcal{X}}\check{\mathcal{X}}})^{-1}) f\|_{p, \check{\mathcal{X}}}^p \\
&\leq \left[\left(\frac{\alpha}{\beta} \right)^{p/p^*} \mu(\check{\mathcal{X}}) + \left(1 + \frac{q'}{\gamma} \right)^p \mu(\check{\mathcal{X}}) \right] \|f\|_{p, \check{\mathcal{X}}}^p \text{ by (40) and (41). } \quad \square
\end{aligned}$$

4.4. Approximation operator norm

This section aims to control the operator norm of \bar{R} defined in (10).

Proposition 14 (Approximation operator norm). *Let $\bar{\mathcal{X}}$ be any proper subset of \mathcal{X} , $\check{\mathcal{X}} = \mathcal{X} \setminus \bar{\mathcal{X}}$, and let \bar{R} be the operator defined in (10). For any $p \geq 1$, for any $f \in \ell_p(\check{\mathcal{X}}, \mu_{\bar{\mathcal{X}}})$,*

$$\|\bar{R}f\|_{p, \mathcal{X}} \leq \left[\left(1 + 2 \frac{\bar{\alpha}}{q'} \right)^p + \frac{\alpha}{\beta} \right]^{1/p} \mu(\bar{\mathcal{X}})^{1/p} \|f\|_{p, \bar{\mathcal{X}}}, \quad (42)$$

where $\bar{\alpha}$ is defined by (28).

Proof. For any $f \in \ell_p(\bar{\mathcal{X}}, \mu_{\bar{\mathcal{X}}})$,

$$\begin{aligned} \|\bar{R}f\|_{p,\mathcal{X}}^p &= \left\| \begin{pmatrix} \text{Id}_{\bar{\mathcal{X}}} - \frac{\bar{\mathcal{L}}}{q'} \\ (-\mathcal{L}_{\check{\mathcal{X}}\check{\mathcal{X}}})^{-1} \mathcal{L}_{\check{\mathcal{X}}\bar{\mathcal{X}}} \end{pmatrix} f \right\|_{p,\mathcal{X}}^p \\ &= \mu(\bar{\mathcal{X}}) \left\| \begin{pmatrix} \text{Id}_{\bar{\mathcal{X}}} - \frac{\bar{\mathcal{L}}}{q'} \end{pmatrix} f \right\|_{p,\bar{\mathcal{X}}}^p + \mu(\check{\mathcal{X}}) \|(-\mathcal{L}_{\check{\mathcal{X}}\check{\mathcal{X}}})^{-1} \mathcal{L}_{\check{\mathcal{X}}\bar{\mathcal{X}}} f\|_{p,\check{\mathcal{X}}}^p \\ &\leq \mu(\bar{\mathcal{X}}) \left\| \begin{pmatrix} \text{Id}_{\bar{\mathcal{X}}} - \frac{\bar{\mathcal{L}}}{q'} \end{pmatrix} f \right\|_{p,\bar{\mathcal{X}}}^p + \mu(\bar{\mathcal{X}}) \frac{\alpha}{\beta} \|f\|_{p,\bar{\mathcal{X}}}^p \text{ by (39)}. \end{aligned}$$

It just remains to prove that for any $f \in \ell_p(\bar{\mathcal{X}}, \mu_{\bar{\mathcal{X}}})$,

$$\|\bar{\mathcal{L}}f\|_{p,\bar{\mathcal{X}}} \leq 2\bar{\alpha} \|f\|_{p,\bar{\mathcal{X}}}.$$

By definition of $\bar{\alpha}$, the matrix $\tilde{P} = \text{Id}_{\bar{\mathcal{X}}} + \frac{\bar{\mathcal{L}}}{\bar{\alpha}}$ is a stochastic matrix, which is symmetric w.r.t. $\mu_{\bar{\mathcal{X}}}$. Therefore, it is a contraction operator from $\ell_p(\bar{\mathcal{X}}, \mu_{\bar{\mathcal{X}}})$ to $\ell_p(\bar{\mathcal{X}}, \mu_{\bar{\mathcal{X}}})$. Hence,

$$\|\bar{\mathcal{L}}f\|_{p,\bar{\mathcal{X}}} = \bar{\alpha} \|(\text{Id}_{\bar{\mathcal{X}}} - \tilde{P})f\|_{p,\bar{\mathcal{X}}} \leq 2\bar{\alpha} \|f\|_{p,\bar{\mathcal{X}}}. \quad \square$$

4.5. Size of the detail

In this section, we provide control of the unnormalized detail coefficients \check{f} in terms of the regularity of the signal f .

Proposition 15. For any $p \geq 1$ and any $f \in \ell_p(\mathcal{X}, \mu)$,

$$\|\check{f}\|_{p,\check{\mathcal{X}}} = \|(K_{q'} - \text{Id})f\|_{p,\check{\mathcal{X}}} \leq \frac{\max_x K_{q'}(x, \check{\mathcal{X}})^{1/p}}{q' \mu(\check{\mathcal{X}})^{1/p}} \|\mathcal{L}f\|_{p,\mathcal{X}}.$$

Proof. Note that by definition of $K_{q'}$, $K_{q'} - \text{Id} = \frac{1}{q'} K_{q'} \mathcal{L}$. Hence,

$$\begin{aligned} \mu(\check{\mathcal{X}}) \|(K_{q'} - \text{Id})f\|_{p,\check{\mathcal{X}}}^p &= \frac{1}{q'^p} \sum_{\check{x} \in \check{\mathcal{X}}} \mu(\check{x}) |E_{\check{x}}[\mathcal{L}f(X(T_{q'}))]|^p \leq \frac{1}{q'^p} \sum_{\check{x} \in \check{\mathcal{X}}} \mu(\check{x}) E_{\check{x}}[|\mathcal{L}f(X(T_{q'}))|^p], \\ &\leq \frac{1}{q'^p} \sum_{\check{x} \in \check{\mathcal{X}}, x \in \mathcal{X}} \mu(\check{x}) K_{q'}(\check{x}, x) |\mathcal{L}f(x)|^p, \\ &\leq \frac{1}{q'^p} \sum_{\check{x} \in \check{\mathcal{X}}, x \in \mathcal{X}} \mu(x) K_{q'}(x, \check{x}) |\mathcal{L}f(x)|^p \text{ by symmetry}, \\ &\leq \frac{\max_x K_{q'}(x, \check{\mathcal{X}})}{q'^p} \sum_{x \in \mathcal{X}} \mu(x) |\mathcal{L}f(x)|^p. \quad \square \end{aligned}$$

4.6. Link between $(\mathcal{X}, \mathcal{L})$ and $(\bar{\mathcal{X}}, \bar{\mathcal{L}})$

We give bounds on the error in the intertwining relation.

Proposition 16. Let $\bar{\mathcal{X}}$ be any proper subset of \mathcal{X} , $\check{\mathcal{X}} = \mathcal{X} \setminus \bar{\mathcal{X}}$. Recall the definition of β given in (9). For any $p \geq 1$, and any $f \in \ell_p(\mathcal{X}, \mu)$,

$$\|(\bar{\mathcal{L}}(K_{q'})_{\bar{\mathcal{X}}\mathcal{X}} - (K_{q'})_{\bar{\mathcal{X}}\mathcal{X}}\mathcal{L})f\|_{p,\bar{\mathcal{X}}} \leq 2q' \left(\frac{\alpha}{\beta}\right)^{1/p^*} \frac{1}{\mu(\bar{\mathcal{X}})^{1/p}} \|f\|_{p,\mathcal{X}}. \quad (43)$$

Proof. Writing that $\bar{\mathcal{L}}$ is a Schur complement and using that $K_{q'}\mathcal{L} = \mathcal{L}K_{q'}$, we get

$$\begin{aligned} \bar{\mathcal{L}}(K_{q'})_{\bar{\mathcal{X}}\mathcal{X}} - (K_{q'})_{\bar{\mathcal{X}}\mathcal{X}}\mathcal{L} &= \mathcal{L}_{\bar{\mathcal{X}}\bar{\mathcal{X}}}(K_{q'})_{\bar{\mathcal{X}}\mathcal{X}} + \mathcal{L}_{\bar{\mathcal{X}}\bar{\mathcal{X}}}(-\mathcal{L}_{\bar{\mathcal{X}}\bar{\mathcal{X}}})^{-1}\mathcal{L}_{\bar{\mathcal{X}}\bar{\mathcal{X}}}(K_{q'})_{\bar{\mathcal{X}}\mathcal{X}} - (\mathcal{L}K_{q'})_{\bar{\mathcal{X}}\mathcal{X}} \\ &= -\mathcal{L}_{\bar{\mathcal{X}}\bar{\mathcal{X}}}(K_{q'})_{\bar{\mathcal{X}}\mathcal{X}} + \mathcal{L}_{\bar{\mathcal{X}}\bar{\mathcal{X}}}(-\mathcal{L}_{\bar{\mathcal{X}}\bar{\mathcal{X}}})^{-1}\mathcal{L}_{\bar{\mathcal{X}}\bar{\mathcal{X}}}(K_{q'})_{\bar{\mathcal{X}}\mathcal{X}} \\ &= \mathcal{L}_{\bar{\mathcal{X}}\bar{\mathcal{X}}}(-\mathcal{L}_{\bar{\mathcal{X}}\bar{\mathcal{X}}})^{-1}(\mathcal{L}_{\bar{\mathcal{X}}\bar{\mathcal{X}}}(K_{q'})_{\bar{\mathcal{X}}\mathcal{X}} + \mathcal{L}_{\bar{\mathcal{X}}\bar{\mathcal{X}}}(K_{q'})_{\bar{\mathcal{X}}\mathcal{X}}) \\ &= \mathcal{L}_{\bar{\mathcal{X}}\bar{\mathcal{X}}}(-\mathcal{L}_{\bar{\mathcal{X}}\bar{\mathcal{X}}})^{-1}(\mathcal{L}K_{q'})_{\bar{\mathcal{X}}\mathcal{X}}. \end{aligned} \quad (44)$$

It has already been proven in Lemma 13 (see (40)), that for any $f \in \ell_p(\check{\mathcal{X}}, \mu_{\check{\mathcal{X}}})$,

$$\|\mathcal{L}_{\bar{\mathcal{X}}\bar{\mathcal{X}}}(-\mathcal{L}_{\bar{\mathcal{X}}\bar{\mathcal{X}}})^{-1}f\|_{p,\bar{\mathcal{X}}} \leq \left(\frac{\alpha}{\beta}\right)^{1/p^*} \left(\frac{\mu(\check{\mathcal{X}})}{\mu(\bar{\mathcal{X}})}\right)^{1/p} \|f\|_{p,\check{\mathcal{X}}}$$

To get (43), it just remains to prove that for any $f \in \ell_p(\mathcal{X}, \mu)$,

$$\|\mathcal{L}K_{q'}f\|_{p,\check{\mathcal{X}}} \leq \frac{2q'}{\mu(\check{\mathcal{X}})^{1/p}} \|f\|_{p,\mathcal{X}}. \quad (45)$$

But this is a direct consequence of the equality $\mathcal{L}K_{q'} = q'(K_{q'} - \text{Id})$, and of the inequality (34). \square

5. The multiresolution scheme

We assume in this section that we have at our disposal a decreasing sequence of proper subsets of \mathcal{X} :

$$\mathcal{X}_0 = \mathcal{X} \supsetneq \mathcal{X}_1 \supsetneq \cdots \supsetneq \mathcal{X}_k \supsetneq \emptyset,$$

and a sequence of non negative real parameters $\{q'_i; i = 0, \dots, k-1\}$. We will discuss later the choice of such sequences. To stick to the previous notations, we define for $i \geq 0$,

- $\bar{\mathcal{X}}_i = \mathcal{X}_{i+1}$;
- $\check{\mathcal{X}}_i = \mathcal{X}_i \setminus \mathcal{X}_{i+1}$;
- $\mathcal{L}_0 = \mathcal{L}$, and for $i \geq 0$, $\mathcal{L}_{i+1} = S_{\mathcal{L}_i}((\mathcal{L}_i)_{\check{\mathcal{X}}_i, \check{\mathcal{X}}_i})$. \mathcal{L}_i is a Markov generator on \mathcal{X}_i which is symmetric with respect to μ .
- $\alpha_i = \max_{x_i \in \mathcal{X}_i} (-\mathcal{L}_i)(x_i, x_i)$;
- $1/\beta_i = \max_{\bar{x}_i \in \mathcal{X}_{i+1}} E_{\bar{x}_i} \left[H_{\mathcal{X}_{i+1}}^{(i)+} - \tau_1^{(i)} \right]$, where $H_A^{(i)+}$ is the return time on the set A of a Markov process $X^{(i)}$ with generator \mathcal{L}_i , and $\tau_1^{(i)}$ is its first jump time (exponential time with parameter α_i);
- $1/\gamma_i = \max_{\bar{x}_i \in \check{\mathcal{X}}_i} E_{\bar{x}_i} \left[H_{\mathcal{X}_{i+1}}^{(i)} \right]$;
- for any $x_i, y_i \in \mathcal{X}_i$, $K_i(x_i, y_i) = q'_i(q'_i \text{Id}_{\mathcal{X}_i} - \mathcal{L}_i)^{-1}(x_i, y_i) = \mathbb{P}_{x_i} [X^{(i)}(T_{q'_i}) = y_i]$, where $T_{q'_i}$ is an exponential random variable with parameter q'_i independent of $X^{(i)}$.

Given a signal f_i defined on \mathcal{X}_i , we can define its approximation coefficients at scale $i+1$:

$$\forall x_{i+1} \in \mathcal{X}_{i+1}, f_{i+1}(x_{i+1}) = \bar{f}_i(x_{i+1}) = \bar{K}_i(f_i)(x_{i+1}), \text{ where } \bar{K}_i := (K_i)_{\mathcal{X}_{i+1}, \mathcal{X}_i},$$

and its detail coefficients at scale $i+1$:

$$\forall \check{x}_i \in \check{\mathcal{X}}_i, \quad g_{i+1}(\check{x}_i) = \check{f}_i(\check{x}_i) = \check{K}_i(f_i)(\check{x}_i) - f_i(\check{x}_i) \quad \text{where } \check{K}_i = (K_i)_{\check{\mathcal{X}}_i, \mathcal{X}_i},$$

so that

$$f_i = \bar{R}_i f_{i+1} + \check{R}_i g_{i+1},$$

with

$$\bar{R}_i = \begin{pmatrix} \text{Id}_{\mathcal{X}_{i+1}} - \frac{1}{q'_i} \mathcal{L}_{i+1} \\ ((-\mathcal{L}_i)_{\check{\mathcal{X}}_i, \check{\mathcal{X}}_i})^{-1} (\mathcal{L}_i)_{\check{\mathcal{X}}_i, \mathcal{X}_{i+1}} \end{pmatrix}, \quad \text{and } \check{R}_i = \begin{pmatrix} (\mathcal{L}_i)_{\mathcal{X}_{i+1}, \check{\mathcal{X}}_i} ((-\mathcal{L}_i)_{\check{\mathcal{X}}_i, \check{\mathcal{X}}_i})^{-1} \\ q'_i ((\mathcal{L}_i)_{\check{\mathcal{X}}_i, \check{\mathcal{X}}_i})^{-1} - \text{Id}_{\check{\mathcal{X}}_i} \end{pmatrix}.$$

Iterating the procedure, we get the usual multiresolution scheme:

$$\begin{array}{ccccccc} f_0 = f & \rightarrow & f_1 & \rightarrow & f_2 & \cdots & \rightarrow & f_k \\ & & \searrow & & \searrow & & \searrow & \\ & & g_1 & & g_2 & & g_k \end{array}$$

5.1. Analysis operator norm

For $k \geq 1$, let us consider the analysis operator U_k :

$$\begin{array}{ccc} U_k : \ell_p(\mathcal{X}) & \rightarrow & \ell_p(\mathcal{X}_k, \mu_{\mathcal{X}_k}) \times \ell_p(\check{\mathcal{X}}_{k-1}, \mu_{\check{\mathcal{X}}_{k-1}}) \times \cdots \times \ell_p(\check{\mathcal{X}}_0, \mu_{\check{\mathcal{X}}_0}) \\ f & \mapsto & [f_k, g_k, g_{k-1}, \dots, g_1] \end{array}$$

The space $\ell_p(\mathcal{X}_k, \mu_{\mathcal{X}_k}) \times \ell_p(\check{\mathcal{X}}_{k-1}, \mu_{\check{\mathcal{X}}_{k-1}}) \times \cdots \times \ell_p(\check{\mathcal{X}}_0, \mu_{\check{\mathcal{X}}_0})$ is endowed with the norm

$$\|[f_k, g_k, g_{k-1}, \dots, g_1]\|_p := \left(\mu(\mathcal{X}_k) \|f_k\|_{p, \mathcal{X}_k}^p + \sum_{i=1}^k \mu(\check{\mathcal{X}}_{i-1}) \|g_i\|_{p, \check{\mathcal{X}}_{i-1}}^p \right)^{1/p}.$$

Proposition 17 (Analysis operator norm). For any $f \in \ell_p(\mathcal{X}, \mu)$,

$$\|U_k(f)\|_p \leq 2^{1/p^*} (1+k)^{1/p} \|f\|_p. \quad (46)$$

Remark. To maintain localization properties of our wavelet basis we do not use any orthogonalization procedure. Then, working with a general graph, we do not expect a conditioning of the multilevel analysis or reconstruction operator that is independent from the graph size. However for $p = \infty$ our analysis operator norm is independent of the graph size, while, for $p < \infty$ it depends only on the number of levels, which is typically only logarithmic in the graph size. Making a careful analysis of the following proof, we will see that at least in the case $p = 1$ our estimation is optimal for the analysis operator we proposed. As far as the multilevel reconstruction operator is concerned, its norm is strongly related with the choice of the q'_i that we will discuss at the end of Section 6.

Proof.

$$\begin{aligned} \|U_k(f)\|_p^p &= \sum_{x_k \in \mathcal{X}_k} \mu(x_k) |f_k(x_k)|^p + \sum_{i=1}^k \sum_{\check{x}_{i-1} \in \check{\mathcal{X}}_{i-1}} \mu(\check{x}_{i-1}) |g_i(\check{x}_{i-1})|^p, \\ &= \sum_{x_k \in \mathcal{X}_k} \mu(x_k) |(\bar{K}_{k-1} \cdots \bar{K}_0)(f)(x_k)|^p \\ &\quad + \sum_{i=1}^k \sum_{\check{x}_{i-1} \in \check{\mathcal{X}}_{i-1}} \mu(\check{x}_{i-1}) \left| ((\check{K}_{i-1} - \text{Id}_{\check{\mathcal{X}}_{i-1}, \mathcal{X}_{i-1}}) \bar{K}_{i-2} \cdots \bar{K}_0)(f)(\check{x}_{i-1}) \right|^p, \end{aligned}$$

$$\begin{aligned}
&\leq \sum_{x_k \in \mathcal{X}_k} \mu(x_k) |(\bar{K}_{k-1} \cdots \bar{K}_0)(f)(x_k)|^p \\
&\quad + 2^{p-1} \sum_{i=1}^k \sum_{\check{x}_{i-1} \in \check{\mathcal{X}}_{i-1}} \mu(\check{x}_{i-1}) \left(|(\check{K}_{i-1} \bar{K}_{i-2} \cdots \bar{K}_0)(f)(\check{x}_{i-1})|^p \right. \\
&\quad \left. + |(\bar{K}_{i-2} \cdots \bar{K}_0)(f)(\check{x}_{i-1})|^p \right), \\
&\leq 2^{p-1} (a_k + b_k),
\end{aligned}$$

where

$$\begin{aligned}
a_k &= \sum_{x_k \in \mathcal{X}_k} \mu(x_k) (\bar{K}_{k-1} \cdots \bar{K}_0)(|f|^p)(x_k) + \sum_{i=1}^k \sum_{\check{x}_{i-1} \in \check{\mathcal{X}}_{i-1}} \mu(\check{x}_{i-1}) (\check{K}_{i-1} \bar{K}_{i-2} \cdots \bar{K}_0)(|f|^p)(\check{x}_{i-1}) \\
b_k &= \sum_{i=1}^k \sum_{\check{x}_{i-1} \in \check{\mathcal{X}}_{i-1}} \mu(\check{x}_{i-1}) (\bar{K}_{i-2} \cdots \bar{K}_0)(|f|^p)(\check{x}_{i-1}).
\end{aligned}$$

Note that

$$\begin{aligned}
&\sum_{x_k \in \mathcal{X}_k} \mu(x_k) (\bar{K}_{k-1} \cdots \bar{K}_0)(|f|^p)(x_k) + \sum_{\check{x}_{k-1} \in \check{\mathcal{X}}_{k-1}} \mu(\check{x}_{k-1}) (\check{K}_{k-1} \bar{K}_{k-2} \cdots \bar{K}_0)(|f|^p)(\check{x}_{k-1}) \\
&= \sum_{x_{k-1} \in \mathcal{X}_{k-1}} \mu(x_{k-1}) (\bar{K}_{k-1} \bar{K}_{k-2} \cdots \bar{K}_0)(|f|^p)(x_{k-1}) \\
&= \sum_{x_{k-1} \in \mathcal{X}_{k-1}} \mu(x_{k-1}) (\bar{K}_{k-2} \cdots \bar{K}_0)(|f|^p)(x_{k-1})
\end{aligned}$$

by symmetry. Therefore, $a_k = a_{k-1} = a_1 = \|f\|_{p, \mathcal{X}}^p$. Concerning b_k , using symmetry, it holds

$$\begin{aligned}
&\sum_{i=1}^k \sum_{\check{x}_{i-1} \in \check{\mathcal{X}}_{i-1}} \mu(\check{x}_{i-1}) (\bar{K}_{i-2} \cdots \bar{K}_0)(|f|^p)(\check{x}_{i-1}) \\
&= \sum_{x_0 \in \mathcal{X}_0} \mu(x_0) |f(x_0)|^p \sum_{i=1}^k \sum_{x_1 \in \mathcal{X}_1} \cdots \sum_{x_{i-2} \in \mathcal{X}_{i-2}} \sum_{\check{x}_{i-1} \in \check{\mathcal{X}}_{i-1}} K_0(x_0, x_1) \cdots K_{i-3}(x_{i-3}, x_{i-2}) K_{i-2}(x_{i-2}, \check{x}_{i-1}) \\
&\leq k \|f\|_{p, \mathcal{X}}^p.
\end{aligned}$$

This ends the proof of (46). \square

We finally give an example for which one can check that in the case $p = 1$ our estimation (46) is optimal. Consider the graph of size $n = l + 1$ with vertex set $\mathcal{X} = \{0, 1, \dots, l\}$ and transition rates

$$w(x, y) = \begin{cases} \varepsilon^{2x+1} & \text{if } y = x + 1, \\ \varepsilon^{2x} & \text{if } y = x - 1, \\ 0 & \text{otherwise,} \end{cases}$$

together with the signal

$$f(x) = \varepsilon^{2x-l}, \quad 0 \leq x \leq l,$$

for $\varepsilon \ll 1$. Choosing $q_i = q'_i = \varepsilon^{i+\frac{3}{2}}$ for $0 \leq i \leq k$, the random subsets $\mathcal{X}_0 \supset \mathcal{X}_1 \supset \mathcal{X}_2 \supset \cdots$ sampled through Wilson's algorithm are typically given by

$$\mathcal{X}_i = \{i, \dots, l\}, \quad i \leq k,$$

and the inequalities of the previous proof are turned into equalities when $p = 1$ and in the limit $\varepsilon \rightarrow 0$.

5.2. Approximation error

Given the coefficients $[f_k, g_k, g_{k-1}, \dots, g_1]$, the reconstruction of $f = f_0$ is

$$\begin{aligned} f &= f_0 = \bar{R}_0 f_1 + \check{R}_0 g_1 \\ &= \bar{R}_0 \bar{R}_1 f_2 + \bar{R}_0 \check{R}_1 g_2 + \check{R}_0 g_1 \\ &= \bar{R}_0 \bar{R}_1 \cdots \bar{R}_{k-1} f_k + \sum_{j=0}^{k-1} (\bar{R}_0 \cdots \bar{R}_{j-1}) \check{R}_j g_{j+1}. \end{aligned}$$

The approximation of f at scale k is thus $\bar{R}_0 \bar{R}_1 \cdots \bar{R}_{k-1} f_k$ and we have the following Jackson's type inequality:

Proposition 18 (Jackson's inequality). For $i \in \{0, \dots, k-1\}$ let $|\bar{R}_i|_p$ ($|\check{R}_i|_p$ respectively) denote the norm operator of $\bar{R}_i : \ell_p(\mathcal{X}_{i+1}, \mu_{\mathcal{X}_{i+1}}) \mapsto \ell_p(\mathcal{X}_i, \mu_{\mathcal{X}_i})$ ($\check{R}_i : \ell_p(\check{\mathcal{X}}_i, \mu_{\check{\mathcal{X}}_i}) \mapsto \ell_p(\mathcal{X}_i, \mu_{\mathcal{X}_i})$ respectively). We set

$$\bar{r}_{i,p} := \left(\frac{\mu(\mathcal{X}_i)}{\mu(\mathcal{X}_{i+1})} \right)^{1/p} |\bar{R}_i|_p, \quad \check{r}_{i,p} := \left(\frac{\mu(\mathcal{X}_i)}{\mu(\check{\mathcal{X}}_i)} \right)^{1/p} |\check{R}_i|_p, \quad \text{and} \quad e_{i,p} := \left(\frac{\mu(\mathcal{X}_{i+1})}{\mu(\mathcal{X}_i)} \right)^{1/p} |\mathcal{L}_{i+1} \bar{K}_i - \bar{K}_i \mathcal{L}_i|_p.$$

For any $p \geq 1$ and any $f \in \ell_p(\mathcal{X}, \mu)$,

$$\begin{aligned} \|f - \bar{R}_0 \bar{R}_1 \cdots \bar{R}_{k-1} f_k\|_{p,\mathcal{X}} &\leq \left\{ \sum_{j=0}^{k-1} \bar{r}_{0,p} \cdots \bar{r}_{j-1,p} \check{r}_{j,p} \frac{1}{q'_j} \right\} \|\mathcal{L}f\|_{p,\mathcal{X}} \\ &\quad + \left\{ \sum_{j=0}^{k-1} \bar{r}_{0,p} \cdots \bar{r}_{j-1,p} \check{r}_{j,p} \frac{1}{q'_j} \sum_{l=0}^{j-1} e_{l,p} \right\} \|f\|_{p,\mathcal{X}}. \end{aligned} \quad (47)$$

Note that it has been proven in Propositions 14, 11 and 16 that

$$\bar{r}_{i,p} \leq \left[\left(1 + 2 \frac{\alpha_{i+1}}{q'_i} \right)^p + \frac{\alpha_i}{\beta_i} \right]^{1/p}, \quad \check{r}_{i,p} \leq \left[\left(\frac{\alpha_i}{\beta_i} \right)^{p/p^*} + \left(1 + \frac{q'_i}{\gamma_i} \right)^p \right]^{1/p}, \quad \text{and} \quad e_{i,p} \leq 2q'_i \left(\frac{\alpha_i}{\beta_i} \right)^{1/p^*}.$$

Moreover, we would like to stress the fact that the term involving $\|f\|_{p,\mathcal{X}}$ in (47) is linked to the error in the intertwining relation, and would disappear if this relation was exact.

In case the intertwining relation error is small, we can interpret our inequality (47) as a Jackson type inequality. Indeed in the setting of approximation theory, Jackson type inequalities relate the smoothness of a function with the approximation properties of a multiscale basis. Let us give some more details on these classical results in the setting of wavelet analysis. Assume ϕ is a real valued scaling function, in the Schwartz class or compactly supported, i.e. a function such that one can find a wavelet ψ such that $\{\phi(\cdot - k), k \in \mathbb{Z}\} \cup \{2^{j/2} \psi(2^j \cdot - k), j \in \mathbb{N}, k \in \mathbb{Z}\}$ is an orthonormal basis of $L^2(\mathbb{R})$, a so-called wavelet

basis. The construction of such kind of basis has a deep relationship with the discrete scheme described in the introduction 1.1 as it is explained in [12] (see also [6] for details). We can define for $j \geq 0$, $V_j := \text{Span}\{\phi_{j,k} = 2^{j/2}\phi(2^j \cdot - k), k \in \mathbb{Z}\}$ and $P_j f$ the orthogonal projector on V_j .

The remarkable properties of wavelet basis make it possible to analyze functions in other functional spaces than $L^2(\mathbb{R})$. Suppose the function ψ of our basis has n vanishing moments with $n \in \mathbb{N}^*$ (by construction ψ has always at least one vanishing moment). Let $1 \leq p \leq \infty$ and $W^{n,p}$ denote the Sobolev space of functions f in L^p whose derivative of order n belongs to L^p . Let $|f|_{W^{n,p}} = \|\frac{d^n f}{dx^n}\|_{L^p}$.

We have indeed the following result in the classical setting (see [5] for a detailed proof).

Theorem 19 (Classical Jackson's inequality). *There exists a constant $C > 0$ such that for all $f \in W^{n,p}$ and $j \geq 0$,*

$$\|f - P_j f\|_{L^p} \leq C 2^{-nj} |f|_{W^{n,p}}. \quad (48)$$

As one can see, our Jackson's inequality (47) relates also the smoothness of the signal f on the graph (measured by $\|\mathcal{L}f\|_{p,\mathcal{X}}$) provided that the error in the intertwining equation is small. The smallest is $\|\mathcal{L}f\|_{p,\mathcal{X}}$, the closer to f should be the approximation part. As in the classical inequality (48), the constant in front of the Laplacian term depends on the level of approximation and gets worse when this level gets coarser.

We can now prove Proposition 18.

Proof.

$$\begin{aligned} \|f - \bar{R}_0 \bar{R}_1 \cdots \bar{R}_{k-1} f_k\|_{p,\mathcal{X}_0} &\leq \sum_{j=0}^{k-1} \left\| (\bar{R}_0 \cdots \bar{R}_{j-1}) \check{R}_j g_{j+1} \right\|_{p,\mathcal{X}_0}, \\ &\leq \sum_{j=0}^{k-1} \prod_{i=0}^{j-1} |\bar{R}_i|_p |\check{R}_j|_p \|g_{j+1}\|_{p,\check{\mathcal{X}}_j}, \\ &\leq \sum_{j=0}^{k-1} \prod_{i=0}^{j-1} |\bar{R}_i|_p |\check{R}_j|_p \left(\frac{\mu(\mathcal{X}_j)}{\mu(\check{\mathcal{X}}_j)} \right)^{1/p} \frac{\|\mathcal{L}_j f_j\|_{p,\mathcal{X}_j}}{q'_j}, \end{aligned}$$

by Proposition 15. Since $\mu(\mathcal{X}_0) = 1$, this leads to

$$\|f - \bar{R}_0 \bar{R}_1 \cdots \bar{R}_{k-1} f_k\|_{p,\mathcal{X}_0} \leq \sum_{j=0}^{k-1} \bar{r}_0 \cdots \bar{r}_{j-1} \check{r}_j \frac{\mu(\mathcal{X}_j)^{1/p}}{q'_j} \|\mathcal{L}_j f_j\|_{p,\mathcal{X}_j}.$$

Now,

$$\begin{aligned} \mathcal{L}_j f_j &= \mathcal{L}_j \bar{K}_{j-1} f_{j-1} \\ &= \mathcal{L}_j \bar{K}_{j-1} \cdots \bar{K}_0 f_0 \\ &= \bar{K}_{j-1} \cdots \bar{K}_0 \mathcal{L}_0 f_0 + \sum_{l=0}^{j-1} \bar{K}_{j-1} \cdots \bar{K}_{l+1} (\mathcal{L}_{l+1} \bar{K}_l - \bar{K}_l \mathcal{L}_l) \bar{K}_{l-1} \cdots \bar{K}_0 f_0. \end{aligned}$$

For any $l \in \{0, \dots, j-1\}$, \bar{K}_l is an operator from $\ell_p(\mathcal{X}_l, \mu_{\mathcal{X}_l})$ to $\ell_p(\mathcal{X}_{l+1}, \mu_{\mathcal{X}_{l+1}})$, and (34) states that $\|\bar{K}_l\|_p \leq \left(\frac{\mu(\mathcal{X}_l)}{\mu(\mathcal{X}_{l+1})} \right)^{1/p}$. Hence,

$$\|\mathcal{L}_j f_j\|_{p,\mathcal{X}_j} \leq \left(\frac{\mu(\mathcal{X}_0)}{\mu(\mathcal{X}_j)} \right)^{1/p} \|\mathcal{L}_0 f_0\|_{p,\mathcal{X}_0} + \sum_{l=0}^{j-1} \left(\frac{\mu(\mathcal{X}_{l+1})}{\mu(\mathcal{X}_j)} \right)^{1/p} |\mathcal{L}_{l+1} \bar{K}_l - \bar{K}_l \mathcal{L}_l|_p \left(\frac{\mu(\mathcal{X}_0)}{\mu(\mathcal{X}_l)} \right)^{1/p} \|f_0\|_{p,\mathcal{X}}$$

$$= \left(\frac{\mu(\mathcal{X}_0)}{\mu(\mathcal{X}_j)} \right)^{1/p} \left[\|\mathcal{L}_0 f_0\|_{p, \mathcal{X}_0} + \sum_{l=0}^{j-1} e_l \|f_0\|_{p, \mathcal{X}} \right].$$

This gives (47). \square

6. About the choice of the parameters q, q'

The aim of this section is to give a guideline in the choice of the various parameters involved in the multiresolution scheme. The requirements we would like to achieve are the following ones:

- (1) At each step of the downsampling, we would like to keep a fixed proportion of points in the current set. This would ensure that the number of steps k in the multiresolution scheme is of order $\log(|\mathcal{X}|)$. In view of Proposition 6, this could be achieved by taking for all $j \geq 0$, $q_j \in [\theta_1 \alpha_j, \theta_2 \alpha_j]$ for some θ_1, θ_2 . With this choice and the choice $\mathcal{X}_{j+1} = \rho(\Phi^{(j)})$ (where $\Phi^{(j)}$ is the random s.o.f. on \mathcal{X}_j associated to \mathcal{L}_j), one has for instance,

$$\forall j \geq 0, |\mathcal{X}_j| \frac{q_j}{q_j + \alpha_j} \leq \mathbb{E}_{q_j} [|\mathcal{X}_{j+1}| | \mathcal{X}_j] \leq |\mathcal{X}_j| - |\mathcal{R}_{j,r}| \left(1 - \frac{r \alpha_j}{q_j + 2 \alpha_j} \right),$$

where $\mathcal{R}_{j,r} = \{x \in \mathcal{X}_j; w_j(x) \geq r \alpha_j\}$.

- (2) To ensure numerical stability of the reconstruction operator, we would like to control at each step the norms of \bar{R}_j and \check{R}_j . Concerning \bar{R}_j (see (42)), this requires α_{j+1}/q'_j and α_j/β_j to be small. As for \check{R}_j (see (36)), one has moreover to ensure that q'_j/γ_j is small. Note that once \mathcal{X}_j has been fixed, α_{j+1} , β_j and γ_j only depend on the choice of \mathcal{X}_{j+1} , hence of q_j . Since α_{j+1}/q'_j and q'_j/γ_j should be small, the product $\frac{\alpha_{j+1}}{q'_j} \frac{q'_j}{\gamma_j} = \frac{\alpha_{j+1}}{\gamma_j}$ should also be small, and one possible choice for q_j is to minimize α_{j+1}/γ_j .
- (3) In order to get a good approximation error, one would also like the error in the intertwining problem between \mathcal{L}_j and \mathcal{L}_{j+1} to be small. Referring to (47), this is achieved if at each step $q'_j(\alpha_j/\beta_j)^{1/p^*}$ is small with respect to the natural unit α_j .

To sum up, \mathcal{X}_j being chosen, one would like to choose (q_j, q'_j) such that

- (C1): $q_j \in [\theta_1 \alpha_j, \theta_2 \alpha_j]$;
- (C2): α_j/β_j is small;
- (C3): α_{j+1}/q'_j is small;
- (C4): q'_j/γ_j is small;
- (C5): $q'_j \alpha_j/\beta_j$ is small (consider the case $p = +\infty$) with respect to α_j .

In this respect, we need some estimates on α_{j+1} , γ_j and β_j in terms of q_j . The next sections are devoted to this task.

6.1. Estimate on $\bar{\alpha}$

Proposition 20. Assume that $\bar{\mathcal{X}} = \rho(\Phi)$ and let $\check{\mathcal{X}} = \mathcal{X} \setminus \bar{\mathcal{X}}$. Then, for any $m \in \{1, \dots, |\mathcal{X}| + 1\}$,

$$\mathbb{E}_q \left[\bar{\omega} \mathbf{1}_{|\bar{\mathcal{X}}|=m} \right] \geq \mathbb{E}_q \left[\frac{1}{|\bar{\mathcal{X}}|} \sum_{\bar{x} \in \bar{\mathcal{X}}} \bar{w}(\bar{x}) \mathbf{1}_{|\bar{\mathcal{X}}|=m} \right] = q \frac{|\mathcal{X}| - m + 1}{m} \mathbb{P}_q [|\bar{\mathcal{X}}| = m - 1]; \quad (49)$$

$$\mathbb{E}_q [\bar{\alpha}] \geq \mathbb{E}_q \left[\frac{1}{|\bar{\mathcal{X}}|} \sum_{\bar{x} \in \bar{\mathcal{X}}} \bar{w}(\bar{x}) \right] = q \mathbb{E}_q \left[\frac{|\check{\mathcal{X}}|}{|\bar{\mathcal{X}}| + 1} \right]. \quad (50)$$

Note that (49) is trivially true for $m = |\mathcal{X}| + 1$. It is also trivially true for $m = 1$, since in this case $\bar{\mathcal{L}}$ is a scalar, which is equal to 0 since $\bar{\mathcal{L}}$ is a Markov generator.

Assuming that $\bar{\alpha} = \max_{\bar{x} \in \bar{\mathcal{X}}} \bar{w}(\bar{x})$ is not too far from the mean of the $(w(\bar{x}), \bar{x} \in \bar{\mathcal{X}})$, Proposition 20 can be used to get an idea of the dependence of $\bar{\alpha}$ with respect to q (or m). Indeed, the term $\frac{|\check{\mathcal{X}}|}{|\bar{\mathcal{X}}| + 1}$ can be numerically estimated as a function of q , since there is a coupling allowing to sample the random forest for all the values of q at the same time. Getting an idea of $\bar{\alpha}$ as a function of q from the simulations, is more time consuming since it requires the computation of a Schur complement for all values of q , and this cannot be done inductively since $\rho(\Phi)$ is not an increasing set w.r.t. q .

Proof. (50) is a direct consequence of (49) after summing over m .

$$\mathbb{E}_q \left[\frac{1}{|\bar{\mathcal{X}}|} \sum_{\bar{x} \in \bar{\mathcal{X}}} \bar{w}(\bar{x}) \mathbf{1}_{|\bar{\mathcal{X}}|=m} \right] = \frac{1}{m} \sum_{x \in \mathcal{X}} \mathbb{E}_q \left[\mathbf{1}_{x \in \bar{\mathcal{X}}} \bar{w}(x) \mathbf{1}_{|\bar{\mathcal{X}}|=m} \right].$$

Hence, the result is a consequence of the following identity: for any $x \in \mathcal{X}$, and any $m \in \{1, \dots, n+1\}$,

$$\mathbb{E}_q \left[\mathbf{1}_{x \in \bar{\mathcal{X}}} \bar{w}(x) \mathbf{1}_{|\bar{\mathcal{X}}|=m} \right] = q \mathbb{P}_q \left[|\bar{\mathcal{X}}| = m - 1; x \notin \bar{\mathcal{X}} \right]. \quad (51)$$

The rest of the proof is devoted to proving (51). Remind that

$$\begin{aligned} \bar{w}(x) &= -\bar{\mathcal{L}}(x, x) = \sum_{y \in \bar{\mathcal{X}}, y \neq x} \bar{\mathcal{L}}(x, y) \\ &= \sum_{y \in \bar{\mathcal{X}}, y \neq x} \mathcal{L}(x, y) + \mathcal{L}_{\bar{\mathcal{X}}\check{\mathcal{X}}}(-\mathcal{L}_{\check{\mathcal{X}}\check{\mathcal{X}}})^{-1} \mathcal{L}_{\check{\mathcal{X}}\bar{\mathcal{X}}}(x, y) \\ &= \sum_{y \in \bar{\mathcal{X}}, y \neq x} w(x, y) + \sum_{\substack{y \in \bar{\mathcal{X}}, y \neq x \\ z_1, z_2 \notin \bar{\mathcal{X}}}} w(x, z_1) (-\mathcal{L}_{\check{\mathcal{X}}\check{\mathcal{X}}})^{-1}(z_1, z_2) w(z_2, y) \end{aligned}$$

Remind also from Lemma 12 (see (37)) that for $z_1, z_2 \in \check{\mathcal{X}}$, $(-\mathcal{L}_{\check{\mathcal{X}}\check{\mathcal{X}}})^{-1}(z_1, z_2) = G_{\bar{\mathcal{X}}}(z_1, z_2)$. This gives an explicit expression of $\bar{w}(x)$ as a function of $\bar{\mathcal{X}}$. Therefore,

$$\mathbb{E}_q \left[\mathbf{1}_{x \in \bar{\mathcal{X}}} \bar{w}(x) \mathbf{1}_{|\bar{\mathcal{X}}|=m} \right] = \sum_{R \ni x, |R|=m} \mathbb{P}_q \left[\bar{\mathcal{X}} = R \right] \left\{ \sum_{y \in R, y \neq x} w(x, y) + \sum_{\substack{y \in R, y \neq x \\ z_1, z_2 \notin R}} w(x, z_1) G_R(z_1, z_2) w(z_2, y) \right\}$$

Now,

$$\mathbb{P}_q \left[\bar{\mathcal{X}} = R \right] = \frac{1}{Z(q)} \sum_{\phi \text{ s.o.f.}, \rho(\phi)=R} q^m w(\phi) = \frac{q^m}{Z(q)} Z_R(0), \quad (52)$$

where we set $Z_R(0) := \sum_{\phi \text{ s.o.f.}, \rho(\phi)=R} w(\phi)$. In addition, it is proven in [2] (see Lemma 3.1, or Appendix B of the preprint) that for any subset R of \mathcal{X} , and any $z_1, z_2 \notin R$,

$$G_R(z_1, z_2) = E_{z_1} \left[\int_0^{H_R^+} \mathbf{1}_{X(s)=z_2} ds \right] = \frac{1}{Z_R(0)} \sum_{\substack{\phi \text{ s.o.f. } \rho(\phi)=R \cup \{z_2\} \\ z_1 \rightsquigarrow_\phi z_2}} w(\phi), \quad (53)$$

where $z_1 \rightsquigarrow_\phi z_2$ means that z_1 is contained in the tree of ϕ , whose root is z_2 . Hence,

$$Z(q)q^{-m}\mathbb{E}_q \left[\mathbf{1}_{x \in \bar{\mathcal{X}}} \bar{w}(x) \mathbf{1}_{|\bar{\mathcal{X}}|=m} \right] = T_1 + T_2,$$

where

$$\begin{aligned} \bullet \quad T_1 &:= \sum_{R \ni x; |R|=m} Z_R(0) \sum_{y \in R; y \neq x} w(x, y) \\ &= \sum_{\substack{\phi, \rho(\phi) \ni x \\ |\rho(\phi)|=m}} \sum_{y \in \rho(\phi); y \neq x} w(\phi) w(x, y) \\ &= \sum_{(\phi, y) \in \mathcal{S}_1(m, x)} w(\phi) w(x, y) \\ &\quad \text{where we set } \mathcal{S}_1(m, x) := \{(\phi, y) \text{ such that } \rho(\phi) \ni x, |\rho(\phi)| = m, y \in \rho(\phi), y \neq x\}. \\ \bullet \quad T_2 &:= \sum_{R \ni x; |R|=m} \sum_{\substack{y \in R; y \neq x \\ z_1, z_2 \notin R}} \sum_{\substack{\phi \text{ s.o.f. } \rho(\phi)=R \cup \{z_2\} \\ z_1 \rightsquigarrow_\phi z_2}} w(x, z_1) w(\phi) w(z_2, y) \\ &= \sum_{(R, y, z_1, z_2, \phi) \in \mathcal{S}_2(m, x)} w(x, z_1) w(\phi) w(z_2, y) \\ &\quad \text{where we set} \end{aligned}$$

$$\mathcal{S}_2(m, x) := \left\{ (R, y, z_1, z_2, \phi) \text{ such that } \begin{array}{l} R \ni x, |R| = m, y \in R, y \neq x, \\ z_1 \notin R, z_2 \notin R, \rho(\phi) = R \cup \{z_2\}, z_1 \rightsquigarrow_\phi z_2 \end{array} \right\}.$$

Let us compute T_1 . To any $(\phi, y) \in \mathcal{S}_1(m, x)$, we associate the forest $\phi_1 = \phi \cup \{(x, y)\}$, so that $w(\phi_1) = w(\phi)w(x, y)$. ϕ_1 is the forest obtained from ϕ by hanging the tree with root x to the root y . Hence $|\rho(\phi_1)| = m - 1$, and the distance $d(x, \rho(\phi_1))$ from x to the roots of ϕ_1 , is equal to 1. Otherwise stated, $\mathcal{S}_1(m, x)$ is sent by this operation to

$$\mathcal{F}_1(m, x) := \{\phi_1 \text{ such that } |\rho(\phi_1)| = m - 1, d(x, \rho(\phi_1)) = 1\}.$$

Note that this correspondence is one to one. Starting from $\phi_1 \in \mathcal{F}_1(m, x)$, we recover y as the root of the tree containing x , and ϕ is then obtained by cutting the edge (x, y) . Thus,

$$T_1 = \sum_{\phi_1 \in \mathcal{F}_1(m, x)} w(\phi_1).$$

Let us now turn our attention to T_2 . To any $(R, y, z_1, z_2, \phi) \in \mathcal{S}_2(m, x)$, we associate the forest $\phi_2 = \phi \cup \{(x, z_1), (z_2, y)\}$, so that $w(\phi_2) = w(x, z_1)w(\phi)w(z_2, y)$. ϕ_2 is the forest obtained from ϕ by first hanging the tree with root x to z_1 , so that x is now in the tree with root z_2 , and then z_2 is attached to y . Hence $|\rho(\phi_2)| = |\rho(\phi)| - 2 = m - 1$, and $d(x, \rho(\phi_2)) \geq 2$ (it could happen that $z_1 = z_2$). Otherwise stated, $\mathcal{S}_2(m, x)$ is sent to

$$\mathcal{F}_2(m, x) := \{\phi_2 \text{ such that } |\rho(\phi_2)| = m - 1, d(x, \rho(\phi_2)) \geq 2\}.$$

Note that this correspondence is one to one. Starting from $\phi_2 \in \mathcal{F}_2(m, x)$, we recover y as the root of the tree containing x ; z_1 is the point following x in the path going from x to y ; z_2 is the point preceding y in the path going from x to y ; $R = \rho(\phi_2) \cup \{x\}$, and ϕ is obtained from ϕ_2 by cutting the edges (x, z_1) , (z_2, y) . Thus,

$$T_2 = \sum_{\phi_2 \in \mathcal{F}_2(m, x)} w(\phi_2).$$

It is clear that $\mathcal{F}_1(m, x)$ and $\mathcal{F}_2(m, x)$ are disjoint sets and that

$$\mathcal{F}_1(m, x) \cup \mathcal{F}_2(m, x) = \{\phi \text{ such that } |\rho(\phi)| = m - 1; x \notin \rho(\phi)\}.$$

This leads to

$$\begin{aligned} \mathbb{E}_q \left[\mathbf{1}_{x \in \bar{\mathcal{X}}} \bar{w}(x) \mathbf{1}_{|\bar{\mathcal{X}}|=m} \right] &= \frac{q^m}{Z(q)} \sum_{\substack{\phi, |\rho(\phi)|=m-1, \\ x \notin \rho(\phi)}} w(\phi) \\ &= q \mathbb{P}_q [|\rho(\Phi)| = m - 1, x \notin \rho(\Phi)] \end{aligned}$$

This ends the proof of (51) and of Proposition 20. \square

6.2. Estimate on β

Proposition 21. Assume that $\bar{\mathcal{X}} = \rho(\Phi)$ and let $\check{\mathcal{X}} = \mathcal{X} \setminus \bar{\mathcal{X}}$. Then, for any $m \in \{1, \dots, |\mathcal{X}|\}$,

$$\mathbb{E}_q \left[\frac{1}{\beta} \mathbf{1}_{|\bar{\mathcal{X}}|=m} \right] \geq \mathbb{E}_q \left[\frac{1}{|\bar{\mathcal{X}}|} \sum_{\bar{x} \in \bar{\mathcal{X}}} \sum_{z \in \check{\mathcal{X}}} P(\bar{x}, z) E_z(H_{\bar{\mathcal{X}}}) \mathbf{1}_{|\bar{\mathcal{X}}|=m} \right] = \frac{|\mathcal{X}| - m}{\alpha m} \mathbb{P}_q [|\bar{\mathcal{X}}| = m]; \quad (54)$$

$$\mathbb{E}_q \left[\frac{1}{\beta} \right] \geq \mathbb{E}_q \left[\frac{|\check{\mathcal{X}}|}{\alpha |\bar{\mathcal{X}}|} \right]. \quad (55)$$

Proof. (55) is a direct consequence of (54) after summing over m .

$$\mathbb{E}_q \left[\frac{1}{|\bar{\mathcal{X}}|} \sum_{\bar{x} \in \bar{\mathcal{X}}} \sum_{z \in \check{\mathcal{X}}} P(\bar{x}, z) E_z(H_{\bar{\mathcal{X}}}) \mathbf{1}_{|\bar{\mathcal{X}}|=m} \right] = \frac{1}{m} \sum_{x \in \mathcal{X}} \mathbb{E}_q \left[\mathbf{1}_{x \in \bar{\mathcal{X}}} \sum_{z \in \check{\mathcal{X}}} P(x, z) E_z(H_{\bar{\mathcal{X}}}) \mathbf{1}_{|\bar{\mathcal{X}}|=m} \right].$$

Hence, the result is a consequence of the following identity: for any $x \in \mathcal{X}$, and any $m \in \{1, \dots, |\mathcal{X}|\}$,

$$\mathbb{E}_q \left[\mathbf{1}_{x \in \bar{\mathcal{X}}} \sum_{z \in \check{\mathcal{X}}} P(x, z) E_z(H_{\bar{\mathcal{X}}}) \mathbf{1}_{|\bar{\mathcal{X}}|=m} \right] = \frac{1}{\alpha} \mathbb{P}_q [|\bar{\mathcal{X}}| = m; x \notin \bar{\mathcal{X}}]. \quad (56)$$

The rest of the proof is devoted to state (56). For $x \in \bar{\mathcal{X}}$ and $z \in \check{\mathcal{X}}$, $P(x, z) = w(x, z)/\alpha$, and summing over z_2 in (53) leads to:

$$\forall z \notin R, \quad E_z(H_R) = \frac{1}{Z_R(0)} \sum_{y \notin R} \sum_{\substack{\phi \text{ s.o.f. } \rho(\phi) = R \cup \{y\} \\ z \rightsquigarrow_{\phi} y}} w(\phi) \quad (57)$$

Therefore, using (52),

$$\begin{aligned}
& \mathbb{E}_q \left[\mathbf{1}_{x \in \bar{\mathcal{X}}} \sum_{z \in \check{\mathcal{X}}} P(x, z) E_z(H_{\bar{\mathcal{X}}}) \mathbf{1}_{|\bar{\mathcal{X}}|=m} \right] \\
&= \frac{q^m}{\alpha Z(q)} \sum_{R \ni x, |R|=m} \sum_{z \notin R; y \notin R} \sum_{\substack{\phi \text{ s.o.f. } \rho(\phi)=R \cup \{y\} \\ z \rightsquigarrow_{\phi} y}} w(x, z) w(\phi) \\
&= \frac{q^m}{\alpha Z(q)} \sum_{(R, y, z, \phi) \in \mathcal{S}_3(m, x)} w(x, z) w(\phi),
\end{aligned}$$

where

$$\mathcal{S}_3(m, x) := \{(R, y, z, \phi) \text{ such that } R \ni x, |R| = m, y \notin R, z \notin R; \rho(\phi) = R \cup \{y\}, z \rightsquigarrow_{\phi} y\}.$$

To any $(R, y, z, \phi) \in \mathcal{S}_3(m, x)$, we associate the forest $\phi_3 = \phi \cup \{(x, z)\}$, so that $w(\phi_3) = w(x, z)w(\phi)$. ϕ_3 is the forest obtained from ϕ by hanging the tree with root x to z , so that x is now in the tree with root y . Hence $|\rho(\phi_3)| = |\rho(\phi)| - 1 = m$, and $d(x, \rho(\phi_3)) \geq 1$ (it could happen that $z = y$). $\mathcal{S}_3(m, x)$ is thus sent to $\mathcal{F}_3(m, x) := \{\phi \text{ such that } |\rho(\phi)| = m, x \notin \rho(\phi)\}$. Note that this correspondence is one to one. Starting from $\phi_3 \in \mathcal{F}_3(m, x)$, we recover y as the root of the tree containing x in ϕ_3 ; z is the point following x in the path going from x to y in ϕ_3 ; ϕ is obtained from ϕ_3 by cutting the edge (x, z) so that x is a root of ϕ , and $R = \rho(\phi) \setminus \{y\}$. Thus,

$$\mathbb{E}_q \left[\mathbf{1}_{x \in \bar{\mathcal{X}}} \sum_{z \in \check{\mathcal{X}}} P(x, z) E_z(H_{\bar{\mathcal{X}}}) \mathbf{1}_{|\bar{\mathcal{X}}|=m} \right] = \frac{q^m}{\alpha Z(q)} \sum_{\phi \in \mathcal{F}_3(m, x)} w(\phi) = \frac{1}{\alpha} \mathbb{P}_q [|\rho(\Phi)| = m; x \notin \rho(\Phi)].$$

This ends the proof of (56) and of Proposition 21. \square

6.3. Estimate on γ

Proposition 22. Assume that $\bar{\mathcal{X}} = \rho(\Phi)$ and let $\check{\mathcal{X}} = \mathcal{X} \setminus \bar{\mathcal{X}}$. Then, for any $m \in \{0, \dots, |\mathcal{X}|\}$,

$$\mathbb{E}_q \left[\frac{1}{\gamma} \mathbf{1}_{|\bar{\mathcal{X}}|=m} \right] \geq \mathbb{E}_q \left[\frac{1}{|\check{\mathcal{X}}|} \sum_{\check{x} \in \check{\mathcal{X}}} E_{\check{x}}(H_{\bar{\mathcal{X}}}) \mathbf{1}_{|\bar{\mathcal{X}}|=m} \right] = \frac{|\mathcal{X}|}{|\mathcal{X}| - m} \frac{1}{q} \mathbb{P}_q [|\bar{\mathcal{X}}| = m + 1]; \quad (58)$$

$$\mathbb{E}_q \left[\frac{1}{\gamma} \right] \geq \frac{1}{q} \mathbb{E}_q \left[\frac{|\mathcal{X}|}{|\check{\mathcal{X}}| + 1} \right]. \quad (59)$$

Note that (58) is trivially true for $m = 0$, since in this case $1/\gamma = +\infty$. It is also true for $m = |\mathcal{X}|$ with the convention $0/0 = 0$ in the right hand side, since in this case $1/\gamma = 0$.

Proof. Here again, (59) is a direct consequence of (58) after summing over m .

$$\begin{aligned}
& \mathbb{E}_q \left[\frac{1}{|\check{\mathcal{X}}|} \sum_{\check{x} \in \check{\mathcal{X}}} E_{\check{x}}(H_{\bar{\mathcal{X}}}) \mathbf{1}_{|\bar{\mathcal{X}}|=m} \right] \\
&= \frac{1}{|\mathcal{X}| - m} \sum_{x \in \mathcal{X}} \mathbb{E}_q \left[\mathbf{1}_{x \notin \bar{\mathcal{X}}} E_x(H_{\bar{\mathcal{X}}}) \mathbf{1}_{|\bar{\mathcal{X}}|=m} \right] \\
&= \frac{1}{|\mathcal{X}| - m} \sum_{x \in \mathcal{X}} \sum_{R; x \notin R, |R|=m} \mathbb{P}_q(\bar{\mathcal{X}} = R) E_x(H_R) \\
&= \frac{q^m}{(|\mathcal{X}| - m) Z(q)} \sum_{x \in \mathcal{X}} \sum_{R; x \notin R, |R|=m} \sum_{\substack{\phi \text{ s.o.f. } \rho(\phi)=R \cup \{y\} \\ x \rightsquigarrow_{\phi} y}} w(\phi) \\
&= \frac{q^m}{(|\mathcal{X}| - m) Z(q)} \sum_{x \in \mathcal{X}} \sum_{\phi \text{ s.o.f. } |\rho(\phi)|=m+1} w(\phi)
\end{aligned}$$

$$\begin{aligned}
&= \frac{1}{q(|\mathcal{X}|-m)} \sum_{x \in \mathcal{X}} \sum_{\phi \text{ s.o.f. } |\rho(\phi)|=m+1} \mathbb{P}_q(\Phi = \phi) \\
&= \frac{|\mathcal{X}|}{q(|\mathcal{X}|-m)} \mathbb{P}_q(|\rho(\Phi)| = m+1).
\end{aligned}$$

This ends the proof of (58) and of Proposition 22. \square

6.4. Discussion on the choice of $\bar{\mathcal{X}}$, q , q'

As was said in the beginning of the section, at each step of the multiresolution scheme, one would want to choose (q, q') so as to fulfill the constraints (C1-5). If (C3) and (C4) are satisfied, then the quantity $\bar{\alpha}/\gamma = (\bar{\alpha}/q')(q'/\gamma)$ has to be small. In the same way, if (C3) and (C5) are satisfied, then $\bar{\alpha}/\beta$ is small. One has therefore to choose $\bar{\mathcal{X}}$ at each step in such a way that $\bar{\alpha}/\gamma$ and/or $\bar{\alpha}/\beta$ are small. A possible choice for q is then to minimize the functions $q \mapsto \mathbb{E}_q(\bar{\alpha}/\gamma)$, or $q \mapsto \mathbb{E}_q(\bar{\alpha}/\beta)$ in some interval $[\theta_1\alpha; \theta_2\alpha]$ (in accordance to (C1)). These functions could be estimated using a Monte Carlo method. However, this can be quite costly, since for instance, the computation of $\bar{\alpha}$ needs the computation of a Schur complement. If one can afford such a computation one time, doing it several times to get a Monte Carlo estimation is time consuming. Instead, in the simulations presented in section 8, we used the estimates (50) and (55). Denote

$$\tilde{\alpha}(q) := q\mathbb{E}_q \left[\frac{|\check{\mathcal{X}}|}{1 + |\bar{\mathcal{X}}|} \right]; \quad \frac{1}{\tilde{\beta}(q)} := \frac{1}{\alpha} \mathbb{E}_q \left[\frac{|\check{\mathcal{X}}|}{|\bar{\mathcal{X}}|} \right]; \quad \frac{1}{\tilde{\gamma}(q)} := \frac{1}{q} \mathbb{E}_q \left[\frac{|\mathcal{X}|}{1 + |\check{\mathcal{X}}|} \right].$$

These functions are quite easy to estimate by a Monte Carlo method. Indeed, a coalescence-fragmentation process has been proposed in [2], allowing to couple the random forests sampled according to π_q for various values of q . In this process, $t = 1/q$ is seen as a time variable. Starting from t_0 , the process begins to sample a random forest whose law is π_{1/t_0} , and this forest evolves in a Markovian way with fragmentations or coalescences of trees. The marginal of this forest process $(\Phi(t), t \geq t_0)$ is at each time t , equal to $\pi_{1/t}$. We refer the reader to [2] for further details on the construction. Using this coalescence-fragmentation process, it is easy to sample the random function $q \mapsto |\rho(\Phi(1/q))|$, and to estimate the functions $\tilde{\alpha}(q)$, $\tilde{\beta}(q)$, $\tilde{\gamma}(q)$ by empirical means over a sample. One can then choose q as a minimizer of $q \in [\theta_1\alpha; \theta_2\alpha] \mapsto \tilde{\alpha}(q)/\tilde{\beta}(q)$ or $q \in [\theta_1\alpha; \theta_2\alpha] \mapsto \tilde{\alpha}(q)/\tilde{\gamma}(q)$. It can however be proven that $\tilde{\alpha}(q)/\tilde{\gamma}(q)$ is a decreasing function of q ; and the reconstruction formulas show that numerical stability depends, just like our Jackson's inequality, on the product of the $\|\bar{R}_j\|$ (rather than $\|\check{R}_j\|$), which is associated with $1/\tilde{\beta}$ (rather than $1/\tilde{\gamma}$). Therefore we decided in the simulations of section 8, to minimize the function $q \in [\theta_1\alpha; \theta_2\alpha] \mapsto \tilde{\alpha}(q)/\tilde{\beta}(q)$, with $\theta_1 = 1/8$ and $\theta_2 = 1$.

Turning back to the choice of q' , we could choose it so that the two terms of the product $(\bar{\alpha}/q')(q'/\beta) = \bar{\alpha}/\beta$ are of the same order. This leads to $q' = \theta'_1 \sqrt{\bar{\alpha}\beta}$. With this choice, one gets, using (42), (50) and (55)

$$\|\bar{R}\|_{\infty, \infty} \leq 1 + \frac{2}{\theta'_1} \sqrt{\frac{\bar{\alpha}}{\beta}} \simeq 1 + \frac{2}{\theta'_1} \sqrt{\frac{q}{\alpha} \frac{|\check{\mathcal{X}}|}{1 + |\bar{\mathcal{X}}|} \frac{|\check{\mathcal{X}}|}{|\bar{\mathcal{X}}|}} \leq 1 + \frac{2\sqrt{\theta_2}}{\theta'_1} \frac{|\check{\mathcal{X}}|}{|\bar{\mathcal{X}}|} \leq \frac{|\mathcal{X}|}{|\bar{\mathcal{X}}|}$$

if we choose $\theta'_1 = 2\sqrt{\theta_2}$. This value of θ'_1 ensures that the multiplicative term appearing in Jackson's inequality (47) for $p = +\infty$ behaves nicely:

$$\prod_{i=0}^{j-1} \left(1 + 2 \frac{\alpha_{i+1}}{q'_i} \right) \leq \prod_{i=0}^{j-1} \frac{|\mathcal{X}_i|}{|\mathcal{X}_{i+1}|} = \frac{|\mathcal{X}_0|}{|\mathcal{X}_j|}. \quad (60)$$

Another possible choice (made in the simulations results of section 8) is to ensure (60) by setting

$$1 + 2 \frac{\alpha_{i+1}}{q'_i} = \frac{|\mathcal{X}_i|}{|\mathcal{X}_{i+1}|} \iff q'_i = 2\alpha_{i+1} \frac{|\mathcal{X}_{i+1}|}{|\check{\mathcal{X}}_i|}.$$

7. A summary of the procedure, and computational issues

7.1. Summary

Fix two real numbers θ_1, θ_2 ($\theta_1 = 1/8$ and $\theta_2 = 1$ in our simulations). Starting from $(\mathcal{X}, \mathcal{L}, \alpha, \mu)$, one iteration of our multiresolution scheme goes through the following steps:

- (1) Choice of q . Make N i.i.d. draws ($N = 1$ in our simulations) of the coalescence-fragmentation forest process $(\Phi_t^{(i)}; t \in [1/(\theta_2\alpha); 1/(\theta_1\alpha)], i = 1, \dots, N)$. From this, estimate the functions

$$\tilde{\alpha}(q) \simeq q \frac{1}{N} \sum_{i=1}^N \frac{|\mathcal{X}| - |\rho(\Phi_{1/q}^{(i)})|}{1 + |\rho(\Phi_{1/q}^{(i)})|},$$

and

$$\frac{1}{\tilde{\beta}(q)} \simeq \frac{1}{N\alpha} \sum_{i=1}^N \frac{|\mathcal{X}| - |\rho(\Phi_{1/q}^{(i)})|}{|\rho(\Phi_{1/q}^{(i)})|}.$$

Choose q_{opt} as the minimizer in $[\theta_1\alpha; \theta_2\alpha]$ of the estimation of $\frac{\tilde{\alpha}(q)}{\tilde{\beta}(q)}$.

- (2) Once q_{opt} is chosen, draw one forest Φ from $\pi_{q_{opt}}$. Take $\bar{\mathcal{X}} = \rho(\Phi)$, $\check{\mathcal{X}} = \mathcal{X} \setminus \bar{\mathcal{X}}$, $\bar{\mu} = \mu(\cdot|\bar{\mathcal{X}})$.
- (3) Compute $\bar{\mathcal{L}} = \mathcal{L}_{\bar{\mathcal{X}}, \bar{\mathcal{X}}} - \mathcal{L}_{\bar{\mathcal{X}}, \check{\mathcal{X}}} \left(\mathcal{L}_{\check{\mathcal{X}}, \check{\mathcal{X}}} \right)^{-1} \mathcal{L}_{\check{\mathcal{X}}, \bar{\mathcal{X}}}$, and $\bar{\alpha} = \max_{\bar{x} \in \bar{\mathcal{X}}} (-\bar{\mathcal{L}}(\bar{x}, \bar{x}))$.
- (4) Choose $q' = 2\bar{\alpha} |\check{\mathcal{X}}| / |\bar{\mathcal{X}}|$. Compute $K_{q'} = q'(q' \text{Id} - \mathcal{L})^{-1}$. This can be done by a polynomial approximation. The scaling functions are then

$$x \mapsto \phi_{\bar{x}}(x) = \frac{K_{q'}(\bar{x}, x)}{\mu(x)},$$

while the wavelets are

$$x \mapsto \psi_{\check{x}}(x) = \frac{(K_{q'} - \text{Id})(\check{x}, x)}{\mu(x)}.$$

Compute also the reconstruction operators \bar{R} and \check{R} .

- (5) Go on with $(\bar{\mathcal{X}}, \bar{\mathcal{L}}, \bar{\mu}, \bar{\alpha})$ in place of $(\mathcal{X}, \mathcal{L}, \mu, \alpha)$.

7.2. Numerical issues

The three main computational issues at each step of the algorithm are:

- (1) the N sampling of the coalescence-fragmentation forest process $(\Phi_t^{(i)})_{t \in [\frac{1}{\theta_2\alpha}; \frac{1}{\theta_1\alpha}]; i=1, \dots, N}$. As proved in [14], the mean time to sample Φ_q is equal to $\sum_{i=0}^{n-1} \frac{1}{q+\lambda_i}$, corresponding to a number of random

jumps for the Markov chain of order $\sum_{i=0}^{n-1} \frac{\alpha}{q+\lambda_i} \leq \frac{\alpha|\mathcal{X}|}{q}$. Hence, the number of operations to sample N realizations of the forest process is of order $N|\mathcal{X}|/\theta_1$.

(2) the computation of the Schur complement. This requires to compute the inverse $(-\mathcal{L}_{\tilde{\mathcal{X}},\tilde{\mathcal{X}}})^{-1}$. Note that

$$(-\mathcal{L}_{\tilde{\mathcal{X}},\tilde{\mathcal{X}}})^{-1} = \frac{1}{\alpha}(\text{Id}_{\tilde{\mathcal{X}}} - P_{\tilde{\mathcal{X}},\tilde{\mathcal{X}}})^{-1} = \frac{1}{\alpha} \sum_{k=0}^{+\infty} (P_{\tilde{\mathcal{X}},\tilde{\mathcal{X}}})^k. \quad (61)$$

Since $(P_{\tilde{\mathcal{X}},\tilde{\mathcal{X}}})^k(\tilde{x},\tilde{y})$ is the probability for the discrete time Markov Chain to go from \tilde{x} to \tilde{y} in k steps without entering $\tilde{\mathcal{X}}$, and since $\tilde{\mathcal{X}}$ is “well spread”, this probability is fast decaying in k , so that a good approximation of $(-\mathcal{L}_{\tilde{\mathcal{X}},\tilde{\mathcal{X}}})^{-1}$ is obtained by truncating the series in (61) to the first terms.

The computation of $(-\mathcal{L}_{\tilde{\mathcal{X}},\tilde{\mathcal{X}}})^{-1}$ is then fast if the matrix $-\mathcal{L}_{\tilde{\mathcal{X}},\tilde{\mathcal{X}}}$ is sparse. This may be satisfied by \mathcal{L} at the beginning of the algorithm, but even if \mathcal{L} is sparse, this is usually no more the case for $\tilde{\mathcal{L}}$ (see Fig. 2). Therefore, after the computation of $\tilde{\mathcal{L}}$, the authors of [11,20] perform a sparsification method. In our frame, it would be convenient that the sparsification step does not alter too much the error bound we get on $\|(K_{q'})_{\tilde{\mathcal{X}},\mathcal{X}}\mathcal{L} - \tilde{\mathcal{L}}(K_{q'})_{\tilde{\mathcal{X}},\mathcal{X}}\|$. Denote by $\tilde{\mathcal{L}}_s$ the sparsified version of $\tilde{\mathcal{L}}$, and let us consider the intertwining error:

$$\begin{aligned} & \|(K_{q'})_{\tilde{\mathcal{X}},\mathcal{X}}\mathcal{L} - \tilde{\mathcal{L}}_s(K_{q'})_{\tilde{\mathcal{X}},\mathcal{X}}\|_{\infty,\infty} \\ & \leq \|(K_{q'})_{\tilde{\mathcal{X}},\mathcal{X}}\mathcal{L} - \tilde{\mathcal{L}}(K_{q'})_{\tilde{\mathcal{X}},\mathcal{X}}\|_{\infty,\infty} + \|(\tilde{\mathcal{L}} - \tilde{\mathcal{L}}_s)(K_{q'})_{\tilde{\mathcal{X}},\mathcal{X}}\|_{\infty,\infty} \\ & \leq \|(K_{q'})_{\tilde{\mathcal{X}},\mathcal{X}}\mathcal{L} - \tilde{\mathcal{L}}(K_{q'})_{\tilde{\mathcal{X}},\mathcal{X}}\|_{\infty,\infty} + \|\tilde{\mathcal{L}} - \tilde{\mathcal{L}}_s\|_{\infty,\infty} \\ & \leq \|(K_{q'})_{\tilde{\mathcal{X}},\mathcal{X}}\mathcal{L} - \tilde{\mathcal{L}}(K_{q'})_{\tilde{\mathcal{X}},\mathcal{X}}\|_{\infty,\infty} + \max_{\tilde{x} \in \tilde{\mathcal{X}}} \sum_{\tilde{y} \in \tilde{\mathcal{X}}} |\tilde{\mathcal{L}}(\tilde{x},\tilde{y}) - \tilde{\mathcal{L}}_s(\tilde{x},\tilde{y})|. \end{aligned}$$

Therefore, the sparsification should for instance satisfy

$$\max_{\tilde{x} \in \tilde{\mathcal{X}}} \sum_{\tilde{y} \in \tilde{\mathcal{X}}} |\tilde{\mathcal{L}}(\tilde{x},\tilde{y}) - \tilde{\mathcal{L}}_s(\tilde{x},\tilde{y})| \leq \|(K_{q'})_{\tilde{\mathcal{X}},\mathcal{X}}\mathcal{L} - \tilde{\mathcal{L}}(K_{q'})_{\tilde{\mathcal{X}},\mathcal{X}}\|_{\infty,\infty}.$$

By building $\tilde{\mathcal{L}}_s$ with this constraint only, and replacing \mathcal{L} by \mathcal{L}_s , we can however locally deteriorate a good intertwining approximation, i.e., increase a lot a small error

$$\epsilon(\tilde{x}) = \sum_{x \in \mathcal{X}} |(K_{q'}\mathcal{L})(\tilde{x},x) - (\tilde{\mathcal{L}}(K_{q'})_{\tilde{\mathcal{X}},\mathcal{X}})(\tilde{x},x)|$$

by allowing this error to grow up to the order of $\epsilon = \max_{\tilde{x} \in \tilde{\mathcal{X}}} \epsilon(\tilde{x})$. Since, at least from a probabilistic point of view (see [1] and its appendix), the collection of these local errors $\epsilon(\tilde{x})$ is more meaningful than the global error ϵ we follow a more local and restrictive approach.

For $\theta \geq 1$ ($\theta = 4$ in our simulation, except in Fig. 3 where we show the effect of different choices of θ in our sparsification scheme), we set to 0 a maximal number of coefficients on each row \tilde{x} of $\tilde{\mathcal{L}}$ until the off-diagonal suppressed weight remains below a targeted error level $\epsilon(\tilde{x})\bar{\alpha}/2\theta\alpha$ (the coefficients of \mathcal{L} and $\tilde{\mathcal{L}}$ are naturally on scale α and $\bar{\alpha}$ respectively). But this has to be done in a symmetric way to ensure the symmetry of $\tilde{\mathcal{L}}_s$. In addition one has to make $\tilde{\mathcal{L}}_s$ a Markov generator. To satisfy all these constraints, we proceed in the following way. We first sort in non-decreasing order the off-diagonal pairs (\tilde{x},\tilde{y}) (with $\tilde{x} \neq \tilde{y}$) according to the quantity $\max(\bar{w}(\tilde{x},\tilde{y}), \bar{w}(\tilde{y},\tilde{x}))$. We then set $\delta_{\tilde{x}} = 0$ for all $\tilde{x} \in \tilde{\mathcal{X}}$ and screen the ordered off-diagonal pairs. Each time (\tilde{x},\tilde{y}) is such that $\delta_{\tilde{x}} + \bar{w}(\tilde{x},\tilde{y}) \leq \epsilon(\tilde{x})$ and $\delta_{\tilde{y}} + \bar{w}(\tilde{y},\tilde{x}) \leq \epsilon(\tilde{y})$, we set to 0 both $\bar{w}(\tilde{x},\tilde{y})$ and $\bar{w}(\tilde{y},\tilde{x})$. We also increase $\delta_{\tilde{x}}$ by $\bar{w}(\tilde{x},\tilde{y})$, and $\delta_{\tilde{y}}$ by $\bar{w}(\tilde{y},\tilde{x})$. We finally adjust the diagonal coefficients $\tilde{\mathcal{L}}(\tilde{x},\tilde{x})$ by subtracting them the associated removed weight $\delta_{\tilde{x}}$. This ensures

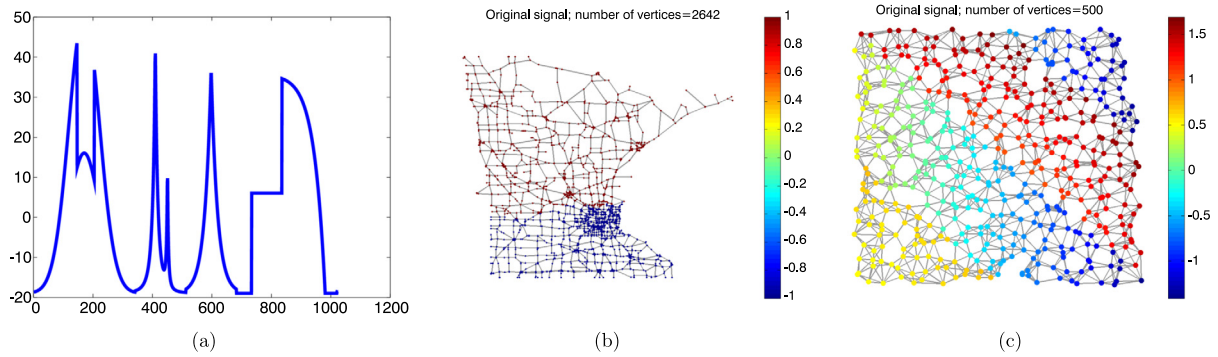


Fig. 1. Original signals: (a) Signal on the line presented in [12]. (b) Signal on the Minnesota roads network: sign of the first Fourier mode e_1 . (c) Signal from [20] on the sensor graph from the GSPBox. (For interpretation of the colors in the figure(s), the reader is referred to the web version of this article.)

$$\begin{aligned} \max_{\bar{x} \in \bar{\mathcal{X}}} \sum_{\bar{y} \in \bar{\mathcal{X}}} |\bar{\mathcal{L}}(\bar{x}, \bar{y}) - \bar{\mathcal{L}}_s(\bar{x}, \bar{y})| &\leq \frac{\bar{\alpha}}{\theta \alpha} \|(K_{q'})_{\bar{\mathcal{X}}, \mathcal{X}} \mathcal{L} - \bar{\mathcal{L}}(K_{q'})_{\bar{\mathcal{X}}, \mathcal{X}}\|_{\infty, \infty} \\ &\leq \|(K_{q'})_{\bar{\mathcal{X}}, \mathcal{X}} \mathcal{L} - \bar{\mathcal{L}}(K_{q'})_{\bar{\mathcal{X}}, \mathcal{X}}\|_{\infty, \infty}. \end{aligned}$$

In doing so, we can lose the irreducibility property, i.e., we can get a disconnected reduced graph associated with $(\bar{\mathcal{X}}, \bar{\mathcal{L}}_s)$. This actually does not raise any particular difficulty. The only use we made of irreducibility was to define without ambiguity our reference measure μ . When irreducibility is lost, we simply proceed by using the natural restriction of μ as reference measure.

- (3) the computation of $K_{q'}$. As in [20], this can be done by Chebyshev polynomial approximation. Apart from computational gain, using polynomial approximation has the advantage to produce filters $K_{q'}$ with a good space localization as soon as \mathcal{L} is nearly diagonal.

8. Numerical results

This section is devoted to numerical illustrations of our multiresolution scheme, referred to as “the intertwining wavelets multiresolution”. We show the results of some downsampling steps on Minnesota roads network (cf. Figs. 2 and 3) containing 2642 vertices, and use the multiresolution schemes to analyze and compress the three benchmark signals of Fig. 1.

8.1. Downsampling of the Minnesota roads network

Fig. 2 shows the result of two levels of forest’s roots sampling, combined with the weighting procedure through Schur’s complement computation without sparsification. It illustrates the loss of sparsity of the weighting procedure. In Fig. 3, we used the sparsification method proposed in section 7.2 with three values of the parameter θ . On these graphs, the width of one edge is proportional to its weight.

8.2. Analysis

8.2.1. The line

We analyze the signal of Fig. 1(a) using our multiresolution scheme. The results after one step are presented in Fig. 4, where we can see that the big detail coefficients are located at the discontinuities of the original signal. We also compare our scheme with a classical wavelet scheme involving Daubechies12 wavelets. The results are given in Fig. 5. After two steps, we end up with 370 approximation coefficients (f_2),

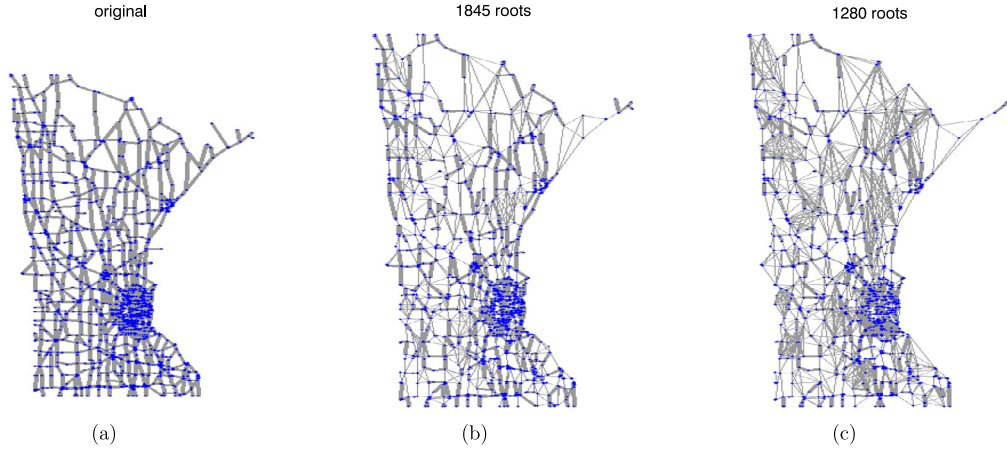


Fig. 2. One sequence of Minnesota's subgraphs, without sparsification. (a) Original graph. (b) First level of downsampling. (c) Second level of downsampling.

instead of 256 for the classical wavelets. In both cases, the number of non vanishing detail coefficients is small.

8.2.2. Minnesota graph

The analysis of the signal of Fig. 1(b) after two steps of the intertwining wavelets multiresolution is presented in Fig. 6. Here again, the big detail coefficients are located at discontinuities of the signal.

8.2.3. Sensor graph

The analysis of the signal of Fig. 1(c) after two steps of the intertwining wavelets multiresolution is presented in Fig. 7.

8.3. Compression

Now, we use the intertwining wavelets to compress the signals of Fig. 1. Since the intertwining wavelets $\psi_{\tilde{x}}$ are unnormalized, we normalize detail coefficients in the compression problem. More precisely, given a signal f , the unnormalized coefficients after one multiresolution step are \bar{f} and \check{f} , from which we can reconstruct f by

$$f = \bar{R}\bar{f} + \check{R}\check{f} = \sum_{\bar{x} \in \bar{\mathcal{X}}} \bar{f}(\bar{x})\tilde{\xi}_{\bar{x}} + \sum_{\check{x} \in \check{\mathcal{X}}} \check{f}(\check{x})\tilde{\xi}_{\check{x}},$$

where the dual basis $(\tilde{\xi}_{\bar{x}}, \tilde{\xi}_{\check{x}})_{\bar{x} \in \bar{\mathcal{X}}, \check{x} \in \check{\mathcal{X}}}$ corresponds to the columns of matrices \bar{R} and \check{R} . Unlike the classical wavelets, the basis $(\xi_x, x \in \mathcal{X})$ and its dual basis $(\tilde{\xi}_x, x \in \mathcal{X})$ are not orthonormal ones. Therefore to compress our signal, we truncate the “normalized coefficients” $\check{f}(\check{x}) \|\tilde{\xi}_{\check{x}}\|$. In a similar way, after k multiresolution steps, the unnormalized coefficients are $[f_k, g_k, g_{k-1}, \dots, g_1]$, from which we can reconstruct f by:

$$f = \bar{R}_0 \cdots \bar{R}_{k-1} f_k + \sum_{j=0}^{k-1} \bar{R}_0 \cdots \bar{R}_{j-1} \check{R}_j g_j \quad (62)$$

$$= \sum_{x_k \in \mathcal{X}_k} f_k(x_k) \tilde{\xi}_{x_k}^{(k,k)} + \sum_{j=0}^{k-1} \sum_{\check{x}_{j-1} \in \mathcal{X}_{j-1}} g_j(\check{x}_{j-1}) \tilde{\xi}_{\check{x}_{j-1}}^{(k,j)} \quad (63)$$

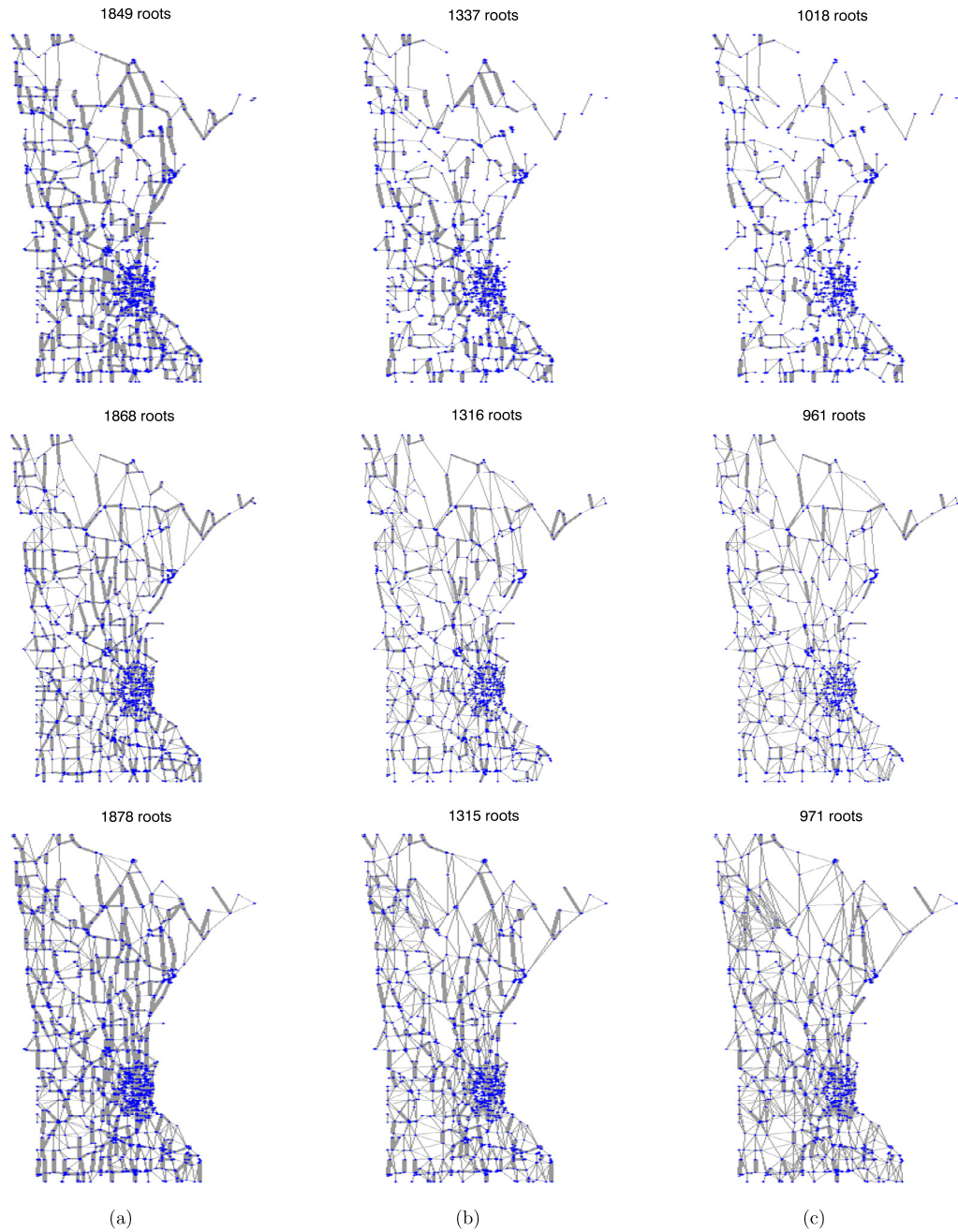


Fig. 3. One sequence of Minnesota's subgraphs, using sparsification. On the first line, $\theta = 2$, while $\theta = 4$ and $\theta = 8$ on the second and third ones. (a) First level of downsampling. (b) Second level of downsampling. (c) Third level of downsampling.

where $(\tilde{\xi}_{x_k}^{(k,k)}, x_k \in \mathcal{X}_k)$ are the columns of the matrix $\bar{R}_0 \cdots \bar{R}_{k-1}$, while $(\tilde{\xi}_{\check{x}_{j-1}}^{(k,j)}, \check{x}_{j-1} \in \check{\mathcal{X}}_{j-1})$ are the columns of the matrix $\bar{R}_0 \cdots \bar{R}_{j-1} \check{R}_j$ ($j = 0, \dots, k-1$). Given a threshold $\epsilon > 0$, the compressed version of f is

$$f_c = \sum_{x_k \in \mathcal{X}_k} f_k(x_k) \tilde{\xi}_{x_k}^{(k,k)} + \sum_{j=0}^{k-1} \sum_{\check{x}_{j-1} \in \check{\mathcal{X}}_{j-1}} g_j(\check{x}_{j-1}) \mathbf{1}_{|g_j(\check{x}_{j-1})| \|\tilde{\xi}_{\check{x}_{j-1}}^{(k,j)}\| \geq \epsilon} \tilde{\xi}_{\check{x}_{j-1}}^{(k,j)}.$$

Another way to compress f is to keep a fixed percentage of the highest (in absolute value) normalized detail coefficients. This is the way we have done our compression experiments.

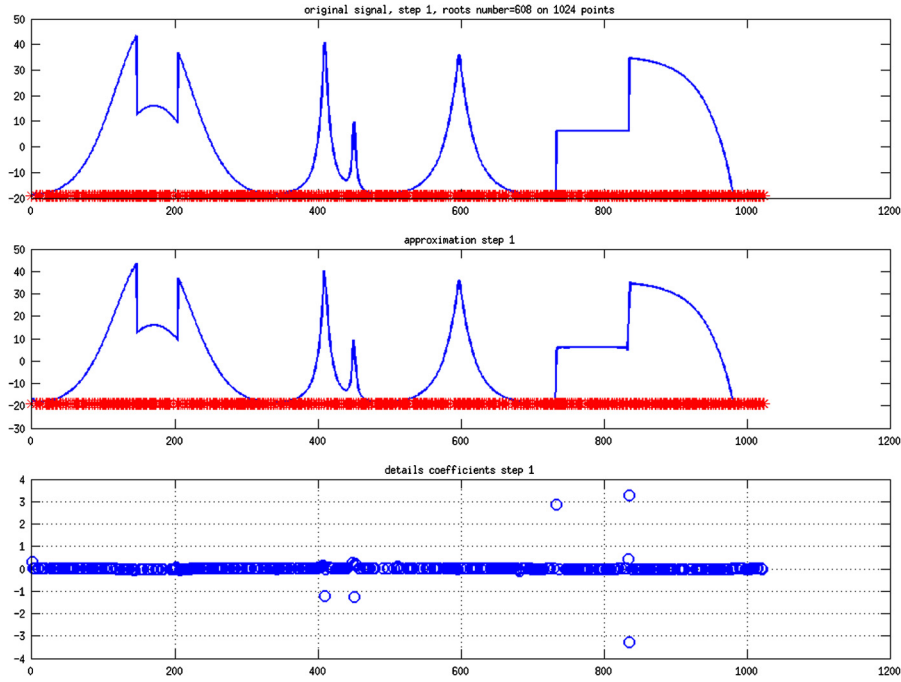


Fig. 4. One step of the analysis. The red crosses are the locations of the 608 chosen points in $\tilde{\mathcal{X}}$, while the blue circles correspond to the points of $\tilde{\mathcal{X}}$. The first figure is the original signal; the second one gives the approximations coefficients (f_1), and the third one the detail coefficients (g_1).

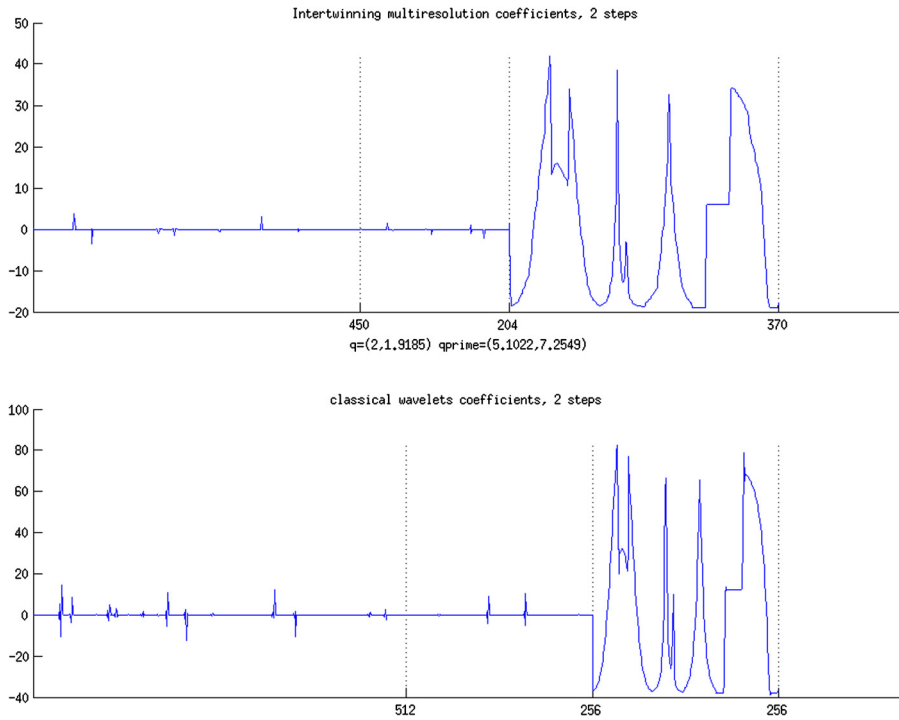


Fig. 5. Results of two steps of the multiresolution for the intertwining wavelets (first graph) and Daubechies 12 wavelets. The three parts of a graph give (g_1, g_2, f_2) .

8.3.1. The line

We compared the compression results of our method with those obtained using classical Daubechies12 wavelets. We let our algorithm evolve until we get an approximation of size less than 16. In the experiment,

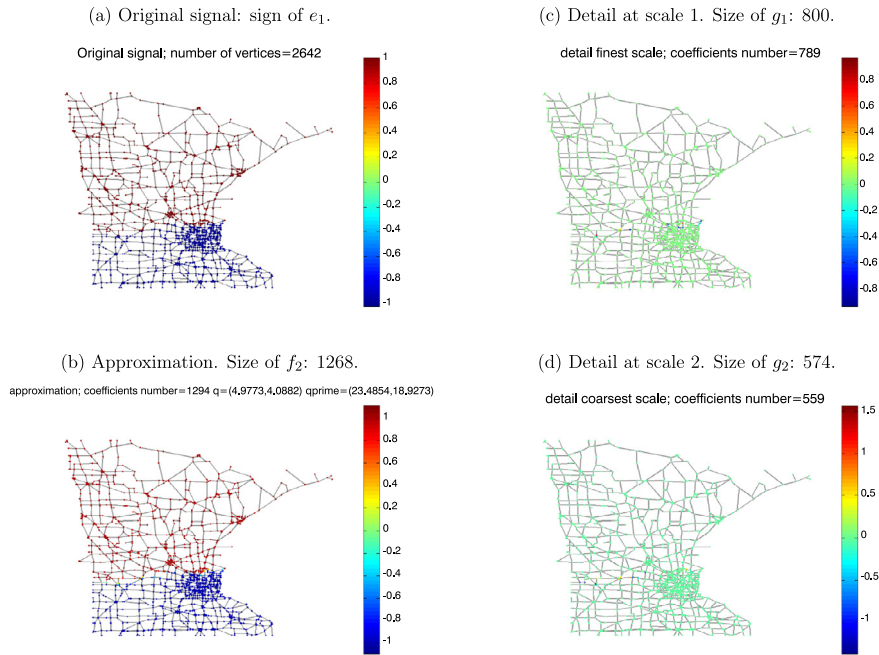


Fig. 6. Two steps of the intertwining wavelets multiresolution upsampling to the original graph: (a) f_0 ; (b) $\bar{R}_0 \bar{R}_1 f_2$; (c) $\check{R}_0 g_1$; (d) $\bar{R}_0 \check{R}_1 g_2$.

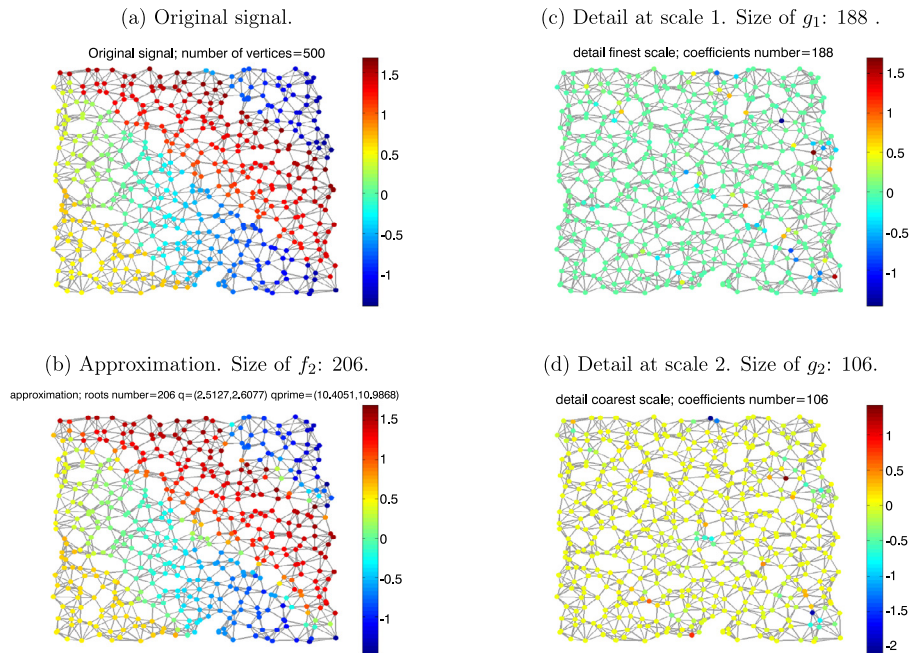


Fig. 7. Two steps of the intertwining wavelets multiresolution upsampling to the original graph: (a) f_0 ; (b) $\bar{R}_0 \bar{R}_1 f_2$; (c) $\check{R}_0 g_1$; (d) $\bar{R}_0 \check{R}_1 g_2$.

this was achieved after 20 steps, and led to 15 approximation coefficients. For classical wavelets, we get 16 approximation coefficients after 6 steps. For both methods, a given proportion p of the details coefficients are kept to compute the compressed signals f_c . Fig. 8 presents the relative errors $\|f - f_c\|_2 / \|f\|_2$ in terms of p , and shows the good behavior of the intertwining wavelets. The compressed signal computed with 10% of normalized detail coefficients is shown in Fig. 9.

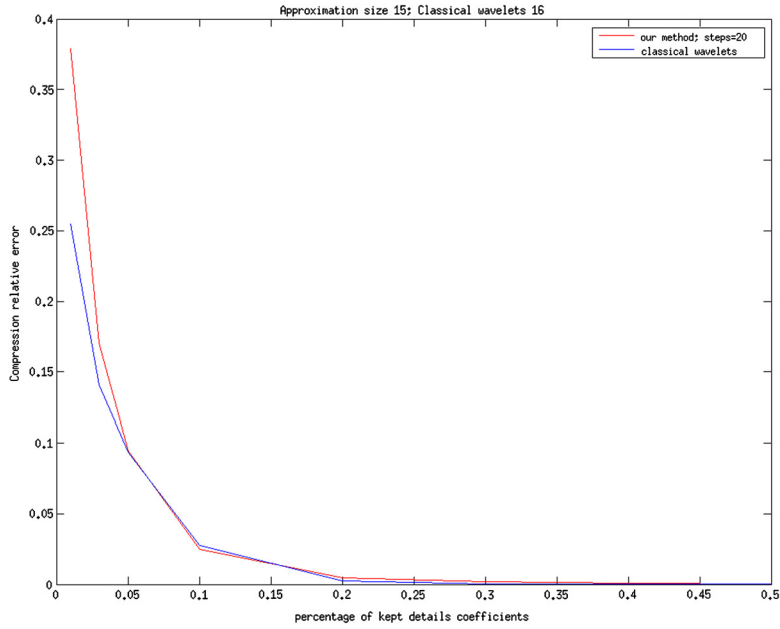


Fig. 8. Relative compression error of signal in Fig. 1(a) in terms of the percentage of kept normalized detail coefficients. In red, the error using intertwining wavelets. In blue, error using Daubechies12 wavelets.

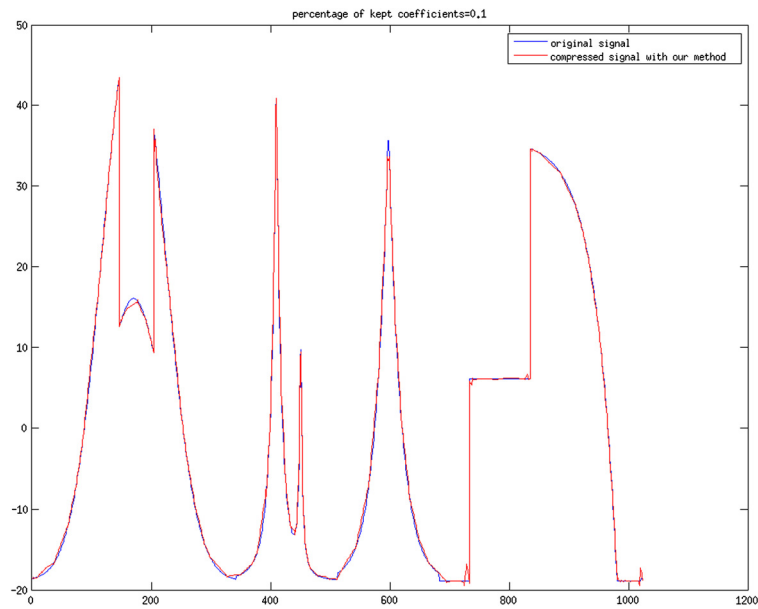


Fig. 9. Original signal (blue line) and compressed one (red line) keeping 10% of normalized detail coefficients.

8.3.2. The Minnesota graph

Fig. 10 gives the compressed signal computed with 10% of kept coefficients after 3 multiresolution steps.

In addition we compare the intertwining wavelets compression results with those obtained through the spectral graph wavelets pyramidal algorithm of [11,20]. For this purpose, we used the GSPBox¹ [18]. The main features of this pyramidal algorithm are the following ones:

¹ Available at <https://lts2.epfl.ch/gsp/>.

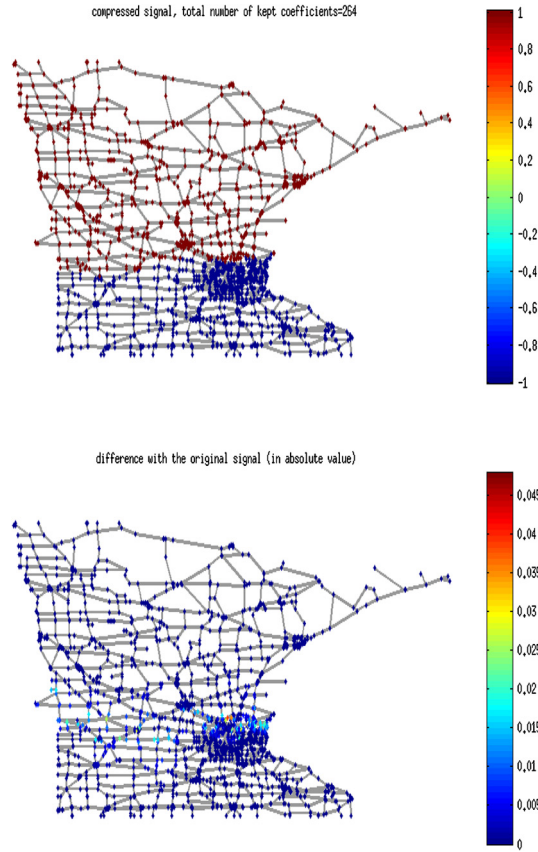


Fig. 10. Compression with 10% of kept coefficients. The first graph is the compressed signal, the second one is the error with the original one.

- (1) Subsampling: $\bar{\mathcal{X}}$ is chosen according to the sign of the highest frequency Fourier mode e_{n-1} .
- (2) Weighting: $\bar{\mathcal{L}}$ is computed by the Schur complement followed by a sparsification step.
- (3) Approximation coefficients: $\bar{f}(\bar{x}) = g(t\mathcal{L})(\bar{x})$, where g is a low-pass filter, and t is a positive real number. As in [20], we took $g(x) = 1/(1+x)$ and $t = 2$ to analyze the signal of Fig. 1(b). With this choice, $\bar{f}(\bar{x}) = K_{1/2}f(\bar{x})$.
- (4) The signal \bar{f} on $\bar{\mathcal{X}}$ is then interpolated on the whole of \mathcal{X} : $\epsilon > 0$ being fixed, the interpolation is defined as

$$f_{int}(x) = (\mathcal{L}^\epsilon)^{-1}_{\mathcal{X}\bar{\mathcal{X}}} \bar{\mathcal{L}}^\epsilon \bar{f}(x),$$

where $\mathcal{L}^\epsilon = \epsilon \text{Id} - \mathcal{L}$, and $\bar{\mathcal{L}}^\epsilon$ is the Schur complement of $\mathcal{L}^\epsilon_{\bar{\mathcal{X}},\bar{\mathcal{X}}}$ in \mathcal{L}^ϵ . Using Proposition 1, one can see at once that $f_{int}|_{\bar{\mathcal{X}}} = \bar{f}$.

- (5) The error prediction $y = f - f_{int}$ is stored.

At the end of one step, the signal $f \in \mathbb{R}^{\mathcal{X}}$ is encoded by $(\bar{f}, y) \in \mathbb{R}^{\bar{\mathcal{X}} \times \mathcal{X}}$, resulting in a redundant information. Fig. 11 presents the relative compression error in terms of the number of kept coefficients for the two methods. More precisely, we ran the spectral graph wavelets pyramidal algorithm for 7 steps, resulting in 20 approximation coefficients among approximately 2×2642 stored ones. We also ran our intertwining wavelets multiresolution until getting approximately the same number of approximation coefficients to get a fair comparison. This took 16 steps, resulting in 16 approximation coefficients. We kept then the same number of the biggest coefficients to construct the compressed version of the signal.

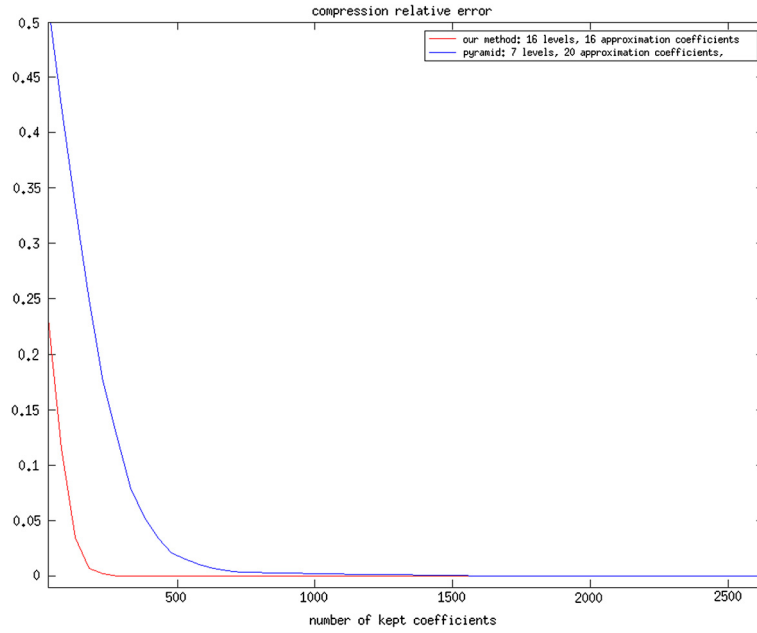


Fig. 11. Relative compression error of signal in Fig. 1(b), in terms of the number of kept coefficients. In red, the error using intertwining wavelets. In blue, error using the spectral graph wavelets pyramidal algorithm.

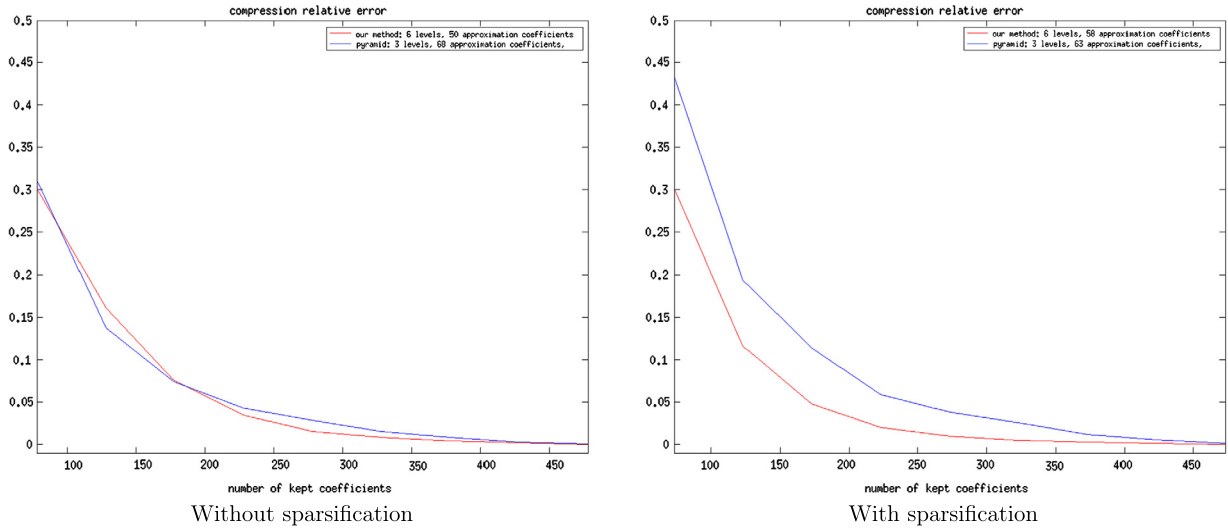


Fig. 12. Relative compression error of signal in Fig. 1(c), in terms of the number of kept coefficients. In red, the error using intertwining wavelets. In blue, error using the spectral graph wavelets pyramidal algorithm. In the first graph, we do not sparsify the graph, while we perform sparsification on the second.

8.3.3. Sensor graph

For the signal on the sensor graph we compare once again in the same way our method with the Pyramid algorithm on 3 steps. In this case we also included the results of our procedure *without* sparsification, since they surprisingly show that the sparsification improve the results. In Fig. 12, we compare the relative compression errors in terms of the number of kept coefficients for the Pyramid algorithm and our method with and without sparsification.

Acknowledgments

Marseille’s team thanks Leiden university for its hospitality along many cumulated weeks during which we discovered our wavelets. This was supported by Frank den Hollander’s ERC Advanced Grant 267356-VARIS. The four authors are especially grateful to Dominique Benielli (labex Archimède) who made possible the numerical work. L. Avena was partially supported by NWO Gravitation Grant 024.002.003-NETWORKS.

References

- [1] Luca Avena, Fabienne Castell, Alexandre Gaudillière, Clothilde Mélot, Approximate and exact solutions of intertwining equations through random forests, arXiv:1702.05992v1 [math.PR].
- [2] Luca Avena, Alexandre Gaudillière, Two applications of random spanning forests, *J. Theoret. Probab.* 31 (4) (2018) 1975–2004.
- [3] Philippe Carmona, Frédérique Petit, Marc Yor, Beta-gamma random variables and intertwining relations between certain Markov processes, *Rev. Mat. Iberoam.* 14 (2) (1998) 311–367.
- [4] Ronald R. Coifman, Mauro Maggioni, Diffusion wavelets, *Appl. Comput. Harmon. Anal.* 21 (1) (2006) 53–94.
- [5] Albert Cohen, Numerical Analysis of Wavelet Methods, *Stud. Math. Appl.*, Elsevier, Amsterdam, 2003.
- [6] Ingrid Daubechies, Ten Lectures on Wavelets, SIAM: Society for Industrial and Applied Mathematics, 1992.
- [7] Persi Diaconis, James Allen Fill, Strong stationary times via a new form of duality, *Ann. Probab.* 18 (4) (1990) 1483–1522.
- [8] Catherine Donati-Martin, Yan Doumerc, Hiroyuki Matsumoto, Marc Yor, Some properties of the Wishart processes and a matrix extension of the Hartman–Watson laws, *Publ. Res. Inst. Math. Sci.* 40 (4) (2004) 1385–1412.
- [9] Oren Elisha, Shai Dekel, Wavelet decompositions of random forests. Smoothness analysis, sparse approximation and applications, *J. Mach. Learn. Res.* 17 (2016) 198.
- [10] Matan Gavish, Nadler Boaz, Ronald R. Coifman, Multiscale wavelets on trees, graphs and high dimensional data: theory and applications to semi supervised learning, in: *Proceedings of the 27th International Conference on Machine Learning*, 2010, pp. 367–374.
- [11] David K. Hammond, Pierre Vandergheynst, Rémi Gribonval, Wavelets on graphs via spectral graph theory, *Appl. Comput. Harmon. Anal.* 30 (2) (2009) 129–150.
- [12] Stéphane Mallat, A Wavelet Tour of Signal Processing, Academic Press, 2008.
- [13] Hiroyuki Matsumoto, Marc Yor, An analogue of Pitman’s $2M - X$ theorem for exponential Wiener functionals. I. A time-inversion approach, *Nagoya Math. J.* 159 (2000) 125–166.
- [14] Philippe Marchal, Loop-erased random walks, spanning trees and Hamiltonian cycles, *Electron. Commun. Probab.* 5 (2000) 39–50.
- [15] Sunil K. Narang, Antonio Ortega, Perfect reconstruction two-channel wavelet filterbanks for graph structured data, *IEEE Trans. Signal Process.* 60 (6) (2012) 2786–2799.
- [16] Sunil K. Narang, Antonio Ortega, Compact support biorthogonal wavelet filterbanks for arbitrary undirected graphs, *IEEE Trans. Signal Process.* 61 (19) (2013) 4673–4685.
- [17] Ha Q. Nguyen, Minh N. Do, Downsampling of signal on graphs via maximum spanning tree, *IEEE Trans. Signal Process.* 63 (1) (2015) 182–191.
- [18] Nathanaël Perraudin, Johan Paratte, David I. Shuman, Vassilis Kalofolias, Pierre Vandergheynst, David K. Hammond, A toolbox for signal processing on graphs, ArXiv e-prints, <http://arxiv.org/abs/1408.5781>, Aug. 2014.
- [19] L.C.G. Rogers, J.W. Pitman, Markov functions, *Ann. Probab.* 9 (4) (1981) 573–582.
- [20] David I. Shuman, Mohamad J. Faraji, Pierre Vandergheynst, A multiscale pyramid transform for graph signals, *IEEE Trans. Signal Process.* 64 (8) (2016) 2119–2134.
- [21] David I. Shuman, Sunil K. Narang, Pascal Frossard, Antonio Ortega, Pierre Vandergheynst, The emerging field of signal processing on graphs: extending high-dimensional data analysis to networks and other irregular domains, *IEEE Signal Process. Mag.* 30 (3) (2013) 83–98.
- [22] Nicolas Tremblay, Pierre Borgnat, Subgraph-based filterbanks for graph signals, *IEEE Trans. Signal Process.* 64 (15) (2016) 3827–3840.
- [23] Jon Warren, Dyson’s Brownian motions, intertwining and interlacing, *Electron. J. Probab.* 12 (19) (2007) 573–590.
- [24] D. Wilson, Generating random spanning trees more quickly than the cover time, in: *Proceedings of the Twenty-Eighth Annual ACM Symposium on the Theory of Computing*, 1996, pp. 296–303.
- [25] Fuzhen Zhang (Ed.), *The Schur Complement and Its Applications*, Numer. Methods Algorithms, Springer, 2005.