HAL open science

Approximate and exact solutions of intertwining equations through random spanning forests

Luca Avena, Fabienne Castell, Alexandre Gaudillière, Clothilde Mélot

To cite this version:

Luca Avena, Fabienne Castell, Alexandre Gaudillière, Clothilde Mélot. Approximate and exact solutions of intertwining equations through random spanning forests. Vares M.E.; Fernández R.; Fontes L.R.; Newman C.M. In and Out of Equilibrium 3: Celebrating Vladas Sidoravicius, 77 (77), Birkäuser; Springer International Publishing, pp.27-69, 2021, Progress in probability, 10.1007/978-3-030-60754-8_3 . hal-01470573v2

HAL Id: hal-01470573

https://hal.science/hal-01470573v2

Submitted on 14 Dec 2020

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

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

Approximate and exact solutions of intertwining equations through random spanning forests

Luca Avena, Fabienne Castell, Alexandre Gaudillière and Clothilde Mélot

Abstract For different reversible Markov kernels on finite state spaces, we look for families of probability measures for which the time evolution almost remains in their convex hull. Motivated by signal processing problems and metastability studies we are interested in the case when the size of such families is *smaller* than the size of the state space, and we want such distributions to be with "small overlap" among them. To this aim we introduce a *squeezing* function to measure the common overlap of such families, and we use random forests to build random approximate solutions of the associated intertwining equations for which we can bound from above the expected values of both squeezing and total variation errors. We also explain how to modify some of these approximate solutions into exact solutions by using those eigenvalues of the associated Laplacian with the largest size.

Keywords: Intertwining, Markov process, finite networks, multiresolution analysis, metastability, random spanning forests.

2010 Mathematical Subject Classification: 05C81; 15A15, 60J28.

Luca Avena
Universiteit Leiden, Niels Bohrweg 1, 2333 CA Leiden, Netherlands,
e-mail: l.avena@math.leidenuniv.nl

Fabienne Castell
Aix-Marseille Université, CNRS, Centrale Marseille, I2M UMR CNRS 7373, 39, rue Joliot Curie, 13 453 Marseille Cedex 13, France, e-mail: fabienne.castell@univ-amu.fr

Alexandre Gaudillière
Aix-Marseille Université, CNRS, Centrale Marseille, I2M UMR CNRS 7373, 39, rue Joliot Curie, 13 453 Marseille Cedex 13, France, e-mail: alexandre.gaudilliere@math.cnrs.fr

Clothilde Mélot
Aix-Marseille Université, CNRS, Centrale Marseille, I2M UMR CNRS 7373, 39, rue Joliot Curie, 13 453 Marseille Cedex 13, France, e-mail: clothilde.melot@univ-amu.fr

1 Main results, motivations and heuristic

The aim of this work is to build exact and approximate solutions of certain intertwining equations between Markov kernels on finite state spaces. The intertwining equations we look at are related to the two following problems. First, we want to build wavelet-like multiresolution schemes for signal processing on arbitrary weighted graphs. Second, we want to make sense of the notion of metastability without asymptotics, in a finite setup where no large-volume or low-temperature limits are in place. We will partially address these problems by giving "good approximate solutions" of the intertwining equations, making use of random spanning forests.

1.1 Intertwining equations

The basic object in this paper is an irreducible stochastic matrix P on a finite state space $\mathscr{X}$. P is associated (see Section 1.3.1 for precise definitions) with the generator $\mathscr{L}$ of a continuous time process X on $\mathscr{X}$ defined by

$$\mathscr{L}f(x) = \sum_{y\in\mathscr{X}} w(x,y)\big[f(y)-f(x)\big], \qquad f:\mathscr{X}\to\mathbb{R}, \qquad x\in\mathscr{X},$$

or, equivalently, with a connected edge weighted graph $G=(\mathscr{X},\mathscr{E},w)$, or $G=(\mathscr{X},w)$, with $\mathscr{X}$ as vertex set, and

$$\mathscr{E} = \Big\{(x,y)\in\mathscr{X}\times\mathscr{X} : w(x,y)>0\Big\}$$

as edges set. We will assume throughout the paper that P (or $\mathscr{L}$) is reversible with respect to some probability measure μ on $\mathscr{X}$:

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

We look at solutions $(\Lambda,\bar{P})$ of the intertwining equations

$$\Lambda P = \bar{P}\Lambda, \tag{2}$$

and, for $q'>0$,

$$\Lambda K_{q'} = \bar{P}\Lambda, \tag{3}$$

where

- $\bar{P}$ is a stochastic matrix defined on some finite state space $\bar{\mathscr{X}}$;
- $\Lambda:\bar{\mathscr{X}}\times\mathscr{X}\to[0,1]$ is a rectangular stochastic matrix;

and $K_{q'}$ is the transition kernel on $\mathscr{X}$ given by

$$K_{q'}(x,y) := P_x(X(T_{q'})=y) = q'(q'\,\mathrm{Id}-\mathscr{L})^{-1}(x,y)\,, \tag{4}$$

with $T_{q'}$ an exponential random variable with parameter q' that is independent of X.

Solving Equation (2) amounts to find a family of probability measures $\nu_{\bar{x}} = \Lambda(\bar{x},\cdot)$ on $\mathscr{X}$ and such that, for some stochastic matrix $\bar{P}$,

$$\nu_{\bar{x}}P = \Lambda P(\bar{x},\cdot) = \bar{P}\Lambda(\bar{x},\cdot) = \sum_{\bar{y}\in\bar{\mathscr{X}}} \bar{P}(\bar{x},\bar{y})\nu_{\bar{y}}, \qquad \bar{x}\in\bar{\mathscr{X}}. \tag{5}$$

In other words the one step evolution of the $\nu_{\bar{x}}$'s have to remain in their convex hull. Solving Equation (3) is the same, except that the "one step evolution" has now to be considered in continuous time and on time scale $1/q'$. In both cases a trivial solution is always given by taking all the $\nu_{\bar{x}}$ equal to the equilibrium measure μ.

Related literature

Intertwining relations, restricted to measures $\nu_{\bar{x}}$ with disjoint support, appeared in the context of diffusion processes in the paper by Rogers and Pitman [20], as a tool to state identities in laws. This method was later successfully applied to many other examples (see for instance [5], [11], [15]). In the context of Markov chains, intertwining was used by Diaconis and Fill [6] without the disjoint support restriction to build strong stationary times and to control convergence rates to equilibrium. This approach initiated in [6] is intimately related with metastability, as will be made clearer in Sections 1.2.2 and 1.2.3, and it has been recently developed in different directions, see e.g. [16] and [14]. However, contrary to our setup, in these references intertwining relations have mainly been considered with an absorbing point for $\bar{P}$ in $\bar{\mathscr{X}}$ and with size m of $\bar{\mathscr{X}}$ being (much) larger than or equal to the size n of $\mathscr{X}$. At present, applications of intertwining include random matrices [7], particle systems [24], spectral clustering [1] ...

Our contribution

Motivated by signal processing and metastability problems (see Section 1.2), in this paper we are instead interested in the case where

(R1) the size m of $\bar{\mathscr{X}}$ is smaller than the size n of $\mathscr{X}$,
(R2) $\bar{P}$ is irreducible,
(R3) the probability measures $\left(\nu_{\bar{x}} : \bar{x}\in\bar{\mathscr{X}}\right)$ are linearly independent and have small "joint overlap".

We will define the *squeezing* of a collection of probability measures to control this overlap (see Section 1.3.2) and a small "joint overlap" will correspond to little squeezed probability measures. We will see in Section 2.2 that, for any reversible stochastic kernel P with non-negative eigenvalues and for any positive $m < n$, non-degenerate solutions of Equation (2) with $|\bar{\mathscr{X}}| = m$ always exist. By "non-degenerate solutions" we mean linearly independent probability measures such that Equation (5) holds for some irreducible $\bar{P}$. But we will argue that exact solutions

tend to be squeezed solutions. Then, rather than looking at the less squeezed solutions in the large space of all solutions for a given m, we will first consider approximate solutions with small squeezing. To this aim we will make use of random spanning forests to build random approximate solutions for which we will be able to bound both the expected value of an error term in intertwining Equation (2) and the expected value of the squeezing (Theorem 1). Then we will use the same random forests to build random approximate solutions of Equation (3) with no overlap, i.e., with disjoint support (Theorem 2). Assuming knowledge of the $n-m$ largest eigenvalues of $-\mathscr{L}$, we will finally see how to modify such an approximate solution of (3) with m probability measures $\nu_{\bar{x}}$ into exact solutions for q' small enough (Theorem 3).

Structure of the paper

In the rest of this section, we detail our motivations, linking signal processing and metastability studies, and we give some heuristics in Section 1.2. After having fixed some notation in Section 1.3.1, we define the squeezing of a probability measure family in Section 1.3.2, we introduce random forests in Section 1.3.3, and state our main results in Sections 1.3.4, 1.3.5 and 1.3.6. In Section 2 we prove some preliminary results, and we give the proofs of our three main theorems in the three last sections. We conclude with an appendix that contains the proof of the main statement that links metastability studies with Equation (5).

1.2 Motivations and heuristics.

Before stating precise results, we would like to explain why we are interested in solutions to (2) and (3) satisfying requirements (R1–3). These come from two motivating problems we describe now, the first one being the construction of a multiresolution analysis for signals on graphs, the second one being a proposal of metastability results without asymptotics.

1.2.1 Pyramidal algorithms in signal processing

First we are interested in extending classical pyramidal algorithms of signal processing on the discrete torus

$$\mathscr{X} = \mathscr{X}_0 = \mathbb{Z}_n = \mathbb{Z}/n\mathbb{Z}$$

to the case of signals on generic edge-weighted graphs. Such algorithms are used for example to analyze or compress a given signal

$$f = f_0 : \mathscr{X}_0 \to \mathbb{R}$$

through *filtering* and *subsampling* operations. A *filter* is a linear operator which is diagonal in the same base as the discrete Laplacian $\mathscr{L}$. A *low-pass filter* K has eigenvalues of order 1 for low frequency modes, i.e., eigenvectors that are associated with small eigenvalues of $-\mathscr{L}$, and it has small eigenvalues for high frequency modes, i.e., eigenvectors that are associated with large eigenvalues of $-\mathscr{L}$. Assuming that n is an even number, a pyramidal algorithm first computes $m = n/2$ *approximation coefficients* by

- computing a low-pass filtered version Kf of the original signal f,
- subsampling Kf by keeping one in each two of its n values, those in some $\mathscr{X}_1 = \bar{\mathscr{X}} \subset \mathscr{X}$, for example the $n/2$ values in the even sites of $\mathbb{Z}_n$.

In doing so it defines a function

$$\bar{f} : \bar{x} \in \bar{\mathscr{X}} \mapsto Kf(\bar{x}) \in \mathbb{R}$$

that can naturally be seen as a signal $f_1 : \mathbb{Z}_{n/2} \to \mathbb{R}$ on a twice smaller torus. It then computes an *approximation* $\tilde{f}$ of f on $\mathscr{X}$ as a function of the approximation coefficients, and a *detail function* $\tilde{g} = f - \tilde{f}$, which in turn can be encoded into $n-m$ *detail coefficients*. Wavelet decomposition algorithms are of this kind. It then applies a similar treatment to f_1, to define f_2, then f_3, ... up to reaching a simple signal defined on a small torus made of a few points only. The reason why this can be useful for compression is that, for well chosen filters, many of the detail coefficients obtained at the different levels are very small or negligible for a large class of smooth signals f. And one just has to store the few non-negligible detail coefficients together with the coarsest approximation's coefficients to reconstruct a good approximation of the original signal f. The point is then to find "good" filters, i.e. "good" $\varphi_{\bar{x}}$ in $\ell_2(\mu)$ (in this case μ is the uniform measure on $\mathscr{X}$, the reversible measure of the simple random walk associated with the discrete Laplacian) so that, for all $f \in \ell_2(\mu)$,

$$\bar{f}(\bar{x}) = \langle \varphi_{\bar{x}}, f \rangle = Kf(\bar{x}).$$

And a basic requirement for good filters is that, for each $\bar{x}$, $\varphi_{\bar{x}}$ is *localized* around $\bar{x}$. Even though the measures $d\nu_{\bar{x}} = \varphi_{\bar{x}} d\mu$ (so that $\bar{f}(\bar{x}) = \langle \nu_{\bar{x}} | f \rangle$) are usually signed measures and not measures, this is the reason why we want to think of the computation of the approximation coefficients $\bar{f}(\bar{x})$ as computation of *local means*. K being a low-pass filter, $\varphi_{\bar{x}}$ needs also to be "localized" in Fourier space (written in the diagonalizing basis of $\mathscr{L}$, it must have small coefficients on high-frequency modes). Thus the difficulty comes from Heisenberg principle, which roughly says that no function $\varphi_{\bar{x}}$ can be well localized both in Fourier space and around $\bar{x}$. Part of the art of wavelet design lies in the ability to make a good compromise with Heisenberg principle (see for example Chapter 7 in [23] for more details on this point).

When moving to the case of signal processing for generic edge-weighted graph, there are three main issues one has immediately to address to build pyramidal algorithms:

(Q1) What kind of subsampling should one use? What could "one every second node" mean?
(Q2) Which kind of filter should one use? How to compute local means?
(Q3) On which (weighted) graph should the approximation coefficients $\bar{f}(\bar{x})$ be defined to iterate the procedure?

On a general weighted finite graph $G = (\mathscr{X}, \mathscr{E}, w)$, none of these questions has a canonical answer. Several attempts to tackle these issues and to generalize wavelet constructions have been proposed: see [22, 19] for recent reviews on this subject and [9] for one of the most popular method. A good starting point to partially answer questions (Q2) and (Q3), is to look for a solution $(\Lambda, \bar{P})$ to intertwining Equation (2), since any row $v_{\bar{x}}$ of Λ automatically belongs to an eigenspace of P, and is therefore frequency localized. Moreover, $\bar{P}$ is a candidate to define the graph structure on $\bar{\mathscr{X}}$. Requirements (R1) (R3) on $(\Lambda, \bar{P})$ reflect then the need of a subsampling procedure (with $m = |\bar{\mathscr{X}}|$ and n of the same order), and of space localization of the $v_{\bar{x}}$. (R2) is more technical, and essentially ensures that we can deal with $\bar{P}$ at the next level in the pyramidal algorithm in the same way we deal with P. We could however continue the pyramidal algorithm with a signal defined on unconnected graphs. Question (Q1) is left apart for the time being. This is where the random forest comes into the play, and we will come back to this question in Section 1.2.4.

Based on Theorem 1, we developed in [2] a novel wavelet transform. To our knowledge, our approach is the first one based on the solution of intertwining equations.

1.2.2 Metastability and intertwining

Our second motivation stems from metastability studies, where it is common to build a coarse-grained version $\bar{X}$ of a Markov process X, possibly by seeing $\bar{X}$ as a measure-valued process on a small state space, these measures being probability measures on the large state space $\mathscr{X}$, on which X is defined. For example, when we want to describe the crystallisation of a slightly supersaturated vapor, we can do it in the following way. Vapor and crystal are defined by probability measures concentrated on very different parts of a very large state space. On this space a Markov process describing the temporal evolution of a microscopic configuration evolves, and this Markovian evolution has to be "macroscopically captured" by a new two-state Markov process evolving from gas (a probability measure on the large state space) to crystal (another probability measure on the same space almost non-overlapping with the previous one). And this evolution is such that the gas should appear as a *local equilibrium* left only to reach a more stable crystalline equilibrium. This is usually done in some asymptotic regime (e.g. large volume or low temperature asymptotic) and we refer to [17] and [4] for mathematical accounts on the subject.

But we are here outside any asymptotic regime: we are given a finite graph $(\mathscr{X}, w)$ or a Markov process X and we want to define a finite coarse-grained version of this graph and Markov process, $(\bar{\mathscr{X}}, \bar{w})$ and $\bar{X}$. Solving the intertwining equation $\Lambda P = \bar{P}\Lambda$, with the size of $\bar{P}$ smaller than the size of P, provides a clean way to do

so. In this equation P is given and stands for the transition kernel of a discrete time skeleton $\hat{X}$ of X (see Section 1.3.1 for a precise definition) and we look for an $m \times m$ stochastic matrix $\bar{P}$ together with a collection of m probability measures $\nu_{\bar{x}}$ on $\mathscr{X}$ that defines the rectangular matrix Λ by

$$\Lambda(\bar{x}, x) = \nu_{\bar{x}}(x), \qquad \bar{x} \in \bar{\mathscr{X}}, \qquad x \in \mathscr{X}.$$

This equation reads

$$\nu_{\bar{x}} P = \sum_{\bar{y} \in \bar{\mathscr{X}}} \bar{P}(\bar{x}, \bar{y}) \nu_{\bar{y}} \tag{6}$$

for all $\bar{x}$ in $\bar{\mathscr{X}}$ and it suggests that the evolution of X can be roughly described through that of $\bar{X}$, associated with the transition kernel $\bar{P}$: from state or local equilibrium $\nu_{\bar{x}}$ the process X evolves towards a new state or local equilibrium $\nu_{\bar{y}}$ which is chosen according to the Markovian kernel $\bar{P}$. This can be turned into a rigorous and powerful mathematical statement by the following proposition, which is a partial rewriting of Section 2.4 of [6] in the spirit of [13], and whose proof is given in appendix.

Proposition 1. *If Equation* (6) *is in force for some $\bar{x}$ in $\bar{\mathscr{X}}$, then there are a filtration $\mathscr{F}$ for which $\hat{X}$ is $\mathscr{F}$-adapted, a $\mathscr{F}$-stopping time $T_{\bar{x}}$ and a $\mathscr{F}_{T_{\bar{x}}}$-mesurable random variable $\bar{Y}_{\bar{x}}$ with value in $\bar{\mathscr{X}} \setminus \{\bar{x}\}$ such that, for $\hat{X}$ started in $\nu_{\bar{x}}$:*

1. *$T_{\bar{x}}$ is geometric with parameter $1 - \bar{P}(\bar{x}, \bar{x})$;*
2. *$\nu_{\bar{x}}$ is stationary up to $T_{\bar{x}}$, i.e., for all $t \geq 0$,*

$$P_{\nu_{\bar{x}}}\left(\hat{X}(t) = \cdot \;\middle|\; t < T_{\bar{x}}\right) = \nu_{\bar{x}}; \tag{7}$$

3. *$P_{\nu_{\bar{x}}}\left(\bar{Y}_{\bar{x}} = \bar{y}\right) = \frac{\bar{P}(\bar{x}, \bar{y})}{1 - \bar{P}(\bar{x}, \bar{x})}$ for all $\bar{y}$ in $\bar{\mathscr{X}} \setminus \{\bar{x}\}$;*
4. *$P_{\nu_{\bar{x}}}\left(\hat{X}(T_{\bar{x}}) = \cdot \;\middle|\; \bar{Y}_{\bar{x}} = \bar{y}\right) = \nu_{\bar{y}}(\cdot)$;*
5. *$\left(\bar{Y}_{\bar{x}}, \hat{X}(T_{\bar{x}})\right)$ and $T_{\bar{x}}$ are independent.*

Notice the slight abuse of notation. In fact, in the above statement, $P_{\nu_{\bar{x}}}$ captures also the extra-randomness of the random variables $T_{\bar{x}}$ and $\bar{Y}_{\bar{x}}$.

As far as metastability is concerned, a possibly more natural approach is, instead of $\Lambda P = \bar{P}\Lambda$, to look for a solution of $\Lambda K_{q'} = \bar{P}\Lambda$ for a small q' and with $K_{q'}$ the transition kernel associated with our process X looked along a Poisson process of intensity q' (see Equation (4) of Section 1.1). It is indeed on a "long" time scale $1/q'$ that one is usually looking at a coarse-grained Markovian version of X. But whatever the equation we are looking at, $\Lambda P = \bar{P}\Lambda$ or $\Lambda K_{q'} = \bar{P}\Lambda$, again we want solutions $\nu_{\bar{x}}$ that are localized in well distinct part of the state space, that is solutions satisfying (R3). Concerning (R1), in metastability studies, we are often interested in cases where m is very small with respect to n. However if one implements a complete pyramidal algorithm, one will solve at the same time intertwining equations with very different m and n by transitivity of the coarse-graining procedure.

1.2.3 Heisenberg principle, approximate solutions and related work

There is actually at least a fourth question without canonical answer that arises when going from classical pyramidal or wavelet algorithms to signal processing for generic weighted graphs: what is a "Heisenberg principle" limiting the localisation of our $\nu_{\bar{x}}$? We do not have an answer to this question, but, although we explained why we are interested in localized, non-overlapping, little squeezed solutions of the intertwining equations, we will see in Section 2 that exact solutions of intertwining equations are strongly localized in Fourier domain, then, a priori, poorly localized in space. This is the main difficulty faced by the present approach and this is one of the two reasons why we turned to approximate solutions of intertwining equations. We will also see in the next section that one needs a detailed knowledge of the spectrum and the eigenvectors of the Laplacian $\mathscr{L}$ to build exact solutions of intertwining equations. From an algorithmic point of view this can be very costly, and this is the other reason why we turned to approximate solutions.

In [2] we analyse the full pyramidal algorithm, including a wavelet basis construction, rather than simply focusing on intertwining equations of a one-step reduction. But we are still looking for a generalised Heisenberg principle that could serve as a guideline for similar constructions. And our results suggest that such a Heisenberg principle should degenerate in presence of a gap in the spectrum (see 1.3.4).

Before concluding these introductory part on intertwining equations, let us note that Proposition 1 can still be used to make sense of *approximate* intertwining. We will show in Section 2, denoting by d_{TV} the total variation distance:

Proposition 2. *If $\hat{X}$ and $\bar{X}$ are discrete time Markov chains on finite spaces $\mathscr{X}$ and $\bar{\mathscr{X}}$ with transitions kernels P and $\bar{P}$, if, for each $\bar{x}$ in $\bar{\mathscr{X}}$, $\nu_{\bar{x}}$ is a probability measure on $\mathscr{X}$, then, setting*

$$\varepsilon = \max_{\bar{x}\in\bar{\mathscr{X}}} d_{\mathrm{TV}}\left(\nu_{\bar{x}}P, \sum_{\bar{y}\in\bar{\mathscr{X}}} \bar{P}(\bar{x},\bar{y})\nu_{\bar{y}}\right)$$

and for any convex combination

$$\nu = \sum_{\bar{x}\in\bar{\mathscr{X}}} \bar{\nu}(\bar{x})\nu_{\bar{x}},$$

there is a coupling between $\hat{X}^{\nu}$ and $\bar{X}^{\bar{\nu}}$, i.e., $\hat{X}$ and $\bar{X}$ started from ν and $\bar{\nu}$, a Markov chain $(Z,\bar{Z})$ on some product space $\mathscr{X}\times\tilde{\mathscr{X}}$, with $\bar{\mathscr{X}}\subset\tilde{\mathscr{X}}$, and two geometric random times T and $\bar{T}$ with mean value $1/\varepsilon$ such that for all $k\geq 0$,

$$P\big(Z_k = x \,\big|\, \bar{Z}_k = \bar{x}\big) = \nu_{\bar{x}}(x), \qquad x\in\mathscr{X}, \qquad \bar{x}\in\bar{\mathscr{X}},$$

$\hat{X}_k^{\nu} = Z_k$ conditionally to $\{T>k\}$ and $\bar{X}_k^{\bar{\nu}} = \bar{Z}_k$ conditionally to $\{\bar{T}>k\}$. In particular it holds, for all $k\geq 0$,

$$d_{\mathrm{TV}}\left(P\big(\hat{X}_k^{\nu}=\cdot\big),\sum_{\bar{x}\in\bar{\mathscr{X}}}P\big(\bar{X}_k^{\bar{\nu}}=\bar{x}\big)\nu_{\bar{x}}\right)\leq P\big(T\leq k\big)+P\big(\bar{T}\leq k\big)$$
$$=2\big(1-(1-\varepsilon)^k\big)\leq 2k\varepsilon. \tag{8}$$

Comment: It is then possible to use approximate intertwining and the coarse-grained version $\bar{X}$ of $\hat{X}$ to control, for example, the mixing time of $\hat{X}$ from that of $\bar{X}$: if $1/\varepsilon$ is large with respect to the latter, one can upper bound the mixing time of $\hat{X}$ by adding that of $\bar{X}$ to the time k needed for all the $\delta_x P^k$ —distribution of $\hat{X}_k$ when $\hat{X}$ is started in x— to be close to the convex hull of the $\nu_{\bar{x}}$. Note that the latter will be related with the squeezing of the $\nu_{\bar{x}}$ in the sense of Section 1.3.2.

1.2.4 Some heuristics on the subsampling question: well distributed points, renormalization and determinantal processes

We now go back to the subsampling question (Q1), i.e. the issue of finding m points, a fraction of n, that are in some sense well distributed in $\mathscr{X}$. This question turns out to be much simpler than (Q2) and (Q3), and a random solution is proposed in [3]. This solution is based on a random spanning forest Φ (i.e. a random collection of oriented rooted trees on the graph $G=(\mathscr{X},w)$ exhausting $\mathscr{X}$), whose law depends on a real parameter $q>0$. We denote by $\rho(\Phi)$ the set of tree roots of Φ. This forest will be precisely described in Section 1.3.3, but we review at once some of its features related to question (Q1). Let us denote, for any subset A of $\mathscr{X}$, by H_A and H_A^+ the hitting time of and the return time to A for the process X:

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

$$H_A^+ := \inf\{t\geq \sigma_1, X(t)\in A\},$$

with σ_1 the first time of the Poisson process that links $\hat{X}$ with X (see Section 1.3.1). For each x in $\mathscr{X}$ the mean hitting time $E_x[H_{\rho(\Phi)}]$ is a random variable, since so are Φ and $\rho(\Phi)$ (E_x being the expectation w.r.t the law of X starting from x). And it turns out that its expected value, with or without conditioning on the size of $\rho(\Phi)$, does not depend on x. In this sense the roots of the random forest are "well spread" on $\mathscr{X}$. More precisely, denoting by $\mathbb{E}_{x,q}$ the expectation w.r.t to the joint law of the Markov process X and on the random forest Φ, and by $\mathbb{E}_q$ expectation w.r.t to the law $\mathbb{P}_q$ of Φ, we have (see [3]):

Proposition 3. *For any $x\in\mathscr{X}$ and $m\in\{1,\cdots,n\}$ it holds*

$$\mathbb{E}_{x,q}\left[H_{\rho(\Phi)}\right]=\frac{\mathbb{P}_q\left[|\rho(\Phi)|>1\right]}{q}; \tag{9}$$

$$\mathbb{E}_{x,q}\left[H_{\rho(\Phi)}\;\middle|\;|\rho(\Phi)|=m\right]=\frac{\mathbb{P}_q\left[|\rho(\Phi)|=m+1\right]}{q\mathbb{P}_q\left[|\rho(\Phi)|=m\right]}; \tag{10}$$

$$\mathbb{E}_q\left[\frac{1}{m}\sum_{\bar{x}\in\rho(\Phi)} E_{\bar{x}}\left[H^+_{\rho(\Phi)} \;\middle|\; |\rho(\Phi)|=m\right]\right] = \frac{n}{\alpha m}\,. \tag{11}$$

This suggests to take $\bar{\mathscr{X}} = \rho(\Phi)$.

This is in line, in the context of very low temperature metastability systems, with Scoppola's renormalization introduced in [21] and with Freidlin and Wentzell's W-graphs [8]. Renormalization consists in individuating a sequence of smaller and smaller subsets of $\mathscr{X}$ with strong recurrence properties on longer and longer time scales. The coarse-grained Markov processes of this approach are the traces of the original process on these subsets. These subsets are naturally built as the roots of forests, or W-graphs in [8], made of bigger and bigger trees. These forests arise in this context of very low temperature systems as almost deterministic limits of our random forests Φ, and local equilibria reduce to (unsqueezed) Dirac masses. In moving away from this asymptotic regime through intertwining equations we consider dealing with more squeezed local equilibria.

As a consequence of Burton and Pemantle's transfer current Theorem, $\rho(\Phi)$ is a determinantal process on $\mathscr{X}$, and its kernel is $K_q = q(q\mathrm{Id}-\mathscr{L})^{-1}$ (see [3]):

Proposition 4. *For any subset A of $\mathscr{X}$,*

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

where $\det_A$ *applied to some matrix is the minor defined by the rows and columns corresponding to A.*

By using reversibility, one can see that the determinant of $(K_q(x,y))_{x,y\in A}$ is, up to a multiplicative factor $\prod_{x\in A}\mu(x)$, the Gram matrix of the distributions $\big(P_x\big(X(\tilde{T}_q)=\cdot\big), x\in A\big)$, with $\tilde{T}_q$ the square of an independent centered Gaussian variable with variance $1/(2q)$ (in such a way that the sum of two independent copies of $\tilde{T}_q$ has the same law as T_q). This means that a family of nodes $\bar{x}$ is unlikely to be part of $\rho(\Phi)$ if the volume of the parallelepiped formed by these distributions is small. It suggests that the distributions $\big(P_{\bar{x}}\big(X(\tilde{T}_q)=\cdot\big), \bar{x}\in\rho(\Phi)\big)$ are typically little squeezed (i.e. well space-localized) and so should be the distributions $\big(K_q(\bar{x},\cdot), \bar{x}\in\rho(\Phi)\big)$, which are easier to deal with. To have a trade-off between squeezing and approximation error in intertwining equations, it will be convenient to introduce a second parameter $q'>0$ and set $\nu_{\bar{x}} = K_{q'}(\bar{x},\cdot)$ for $\bar{x}$ in $\rho(\Phi)$. At this point the choice made for $\bar{P}$ in 1.3.4 may be the most natural one.

Finally, when dealing with metastability issues, building local equilibria $\nu_{\bar{x}}$ from single "microscopic configurations" $\bar{x}$ in $\rho(\Phi)$ seems rather unnatural. In our previous example, no special microscopic configuration should play a role in defining what a metastable vapor should be. One should better look for larger structures associated with Φ, like the partition $\mathscr{A}(\Phi)$ of $\mathscr{X}$ defined by the trees of Φ, rather than $\rho(\Phi)$. Then, in view of the following proposition from [3], the unsqueezed measures $\mu_{A(\bar{x})}$ appear to be natural candidates for giving approximate solutions of (3):

Proposition 5. *Conditional law of the roots, given the partition.*
Let m be fixed, and $A_1,\ldots,A_m$ be a partition of $\mathscr{X}$. For any $x_1\in A_1,\cdots,x_m\in A_m$,

$$\mathbb{P}_q\left[\rho(\Phi)=\{x_1,\cdots,x_m\}\;\middle|\;\mathscr{A}(\phi)=(A_1,\cdots,A_m)\right]=\prod_{i=1}^{m}\mu_{A_i}(x_i)\,, \tag{12}$$

where μ_A is the invariant measure μ conditioned to A ($\mu_A(B)=\mu(A\cap B)/\mu(A)$). Hence, given the partition, the roots are independent, and distributed according to the invariant measure.

Again, in the context of very low temperature metastable systems, this is in line with the so-called cycle decomposition ([8], [17]).

1.2.5 About the reversibility assumption.

Concerning signal processing issues, the reversibility assumption (1) is rather standard. Actually, the classical multiresolution analysis of signals defined on the regular grid assumes the "reversibility of the grid", even in the case of audio signal where $\mathscr{X}$ is a time interval. When considering oriented and non-reversible graphs, the question of building a suitable Fourier analysis is already a delicate one, beyond the scope of the present paper.

In metastability studies, both reversible and non-reversible settings have been considered. Common approaches are usually initiated in the former context, where a richer palette of techniques is available. In our case, looking at metastable issues through intertwining equations and random forests does not rely on reversibility hypotheses, but our squeezing analysis is based on the description of determinantal processes associated with self-adjoint kernels: these are mixture of determinantal processes with a deterministic size and associated with a projector. Without such a reversibility hypotheses the full description of determinantal process kernels is still an open question. It is worth noting that the root process we use for subsampling is an example of such a non-reversible determinantal process, and that our total variation estimates in Theorem 1 still hold in this context.

1.3 Notations and main results.

We describe now our main results, and for this purpose, introduce notations used throughout the paper.

1.3.1 Functions, measures, Markov kernel and generator

Let $\mathscr{X}$ be a finite space with cardinality $|\mathscr{X}|=n$. We consider an irreducible continuous time Markov process $(X(t),t\geq 0)$ on $\mathscr{X}$, with generator $\mathscr{L}$:

$$\mathscr{L}f(x):=\sum_{y\in\mathscr{X}}w(x,y)(f(y)-f(x)), \tag{13}$$

where $f : \mathscr{X} \to \mathbb{R}$ is an arbitrary function, and $w : \mathscr{X} \times \mathscr{X} \to [0,+\infty[$ gives the transition rates. For $x \in \mathscr{X}$, let

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

Note that $\mathscr{L}$ acts on functions as the matrix, still denoted by $\mathscr{L}$, the entries of which are:

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

Let $\alpha > 0$ be defined by

$$\alpha = \max_{x \in \mathscr{X}} w(x)\,. \tag{14}$$

Hence, $P := \mathscr{L}/\alpha + \mathrm{Id}$ 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 point process $(\sigma_i, i > 0)$ on $\mathbb{R}^+$ with rate α. At each point, or time, in the Poisson process, X moves according to the trajectory of $\hat{X}$, i.e., with $\sigma_0 = 0$:

$$X(t) = \sum_{i=0}^{+\infty} \hat{X}_i \mathbb{1}_{\sigma_i \leq t < \sigma_{i+1}}\,.$$

We assume that X is reversible with respect to the probability measure μ on $\mathscr{X}$, (i.e. (1)). The process X being irreducible, μ is strictly positive. The operator $-\mathscr{L}$ is self-adjoint and positive; we denote by $(\lambda_i; i = 0, \cdots, n-1)$ the real eigenvalues of $-\mathscr{L}$ in increasing order. It follows from the fact that P is irreducible that

$$0 = \lambda_0 < \lambda_1 \leq \lambda_2 \cdots \leq \lambda_{n-1} \leq 2\alpha\,. \tag{15}$$

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

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

The scalar product of two functions f and g in $\ell_2(\mu)$ is

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

The corresponding norm is denoted by $\|\cdot\| = \|\cdot\|_2$. When f is a function and ν is a signed measure, the duality bracket between ν and f is

$$\langle \nu | f \rangle = \sum_{x \in \mathscr{X}} \nu(x) f(x)\,.$$

$\ell_p^*(\mu)$ denotes the dual space of $\ell_p(\mu)$ with respect to $\langle \cdot | \cdot \rangle$. It is the space of signed measures endowed with the norm:

$$\|\nu\|_p^* = \left(\sum_{x \in \mathscr{X}} \left| \frac{\nu(x)}{\mu(x)} \right|^{p^*} \mu(x) \right)^{1/p^*}$$

where p^* is the conjugate exponent of p: $1/p + 1/p^* = 1$. $\ell_p^*(\mu)$ is identified with $\ell_{p^*}(\mu)$ through the isometry: $\nu \in \ell_p^*(\mu) \mapsto \nu^* \in \ell_{p^*}(\mu)$, where $\nu^*(x) = \nu(x)/\mu(x)$ is the density of ν with respect to μ. The inverse of this isometry is still denoted by *. It associates to a function $f \in \ell_p(\mu)$, the signed measure $f^* \in \ell_{p^*}^*(\mu)$ whose density with respect to μ is f: $f^*(A) = \sum_{x \in A} \mu(x) f(x)$ for all subset A of $\mathscr{X}$. $\ell_2^*(\mu)$ is an Euclidean space whose scalar product is denoted by:

$$\langle \nu, \rho \rangle^* := \langle \nu^*, \rho^* \rangle = \sum_{x \in \mathscr{X}} \nu(x) \rho(x) \frac{1}{\mu(x)} .$$

The corresponding norm is denoted by $\|\cdot\|^*$. For $\nu \in \ell_2^*(\mu)$ and $f \in \ell_2(\mu)$, one gets

$$\langle \nu | f \rangle = \langle \nu, f^* \rangle^* = \langle \nu^*, f \rangle .$$

1.3.2 Squeezing of a collection of probability measures

For some finite space $\bar{\mathscr{X}}$ of size $m \leq n$, let $(\nu_{\bar{x}} : \bar{x} \in \bar{\mathscr{X}})$ be a collection of m probability measures on $\mathscr{X}$ which is identified with the matrix Λ, the row vectors of which are the $\nu_{\bar{x}}$'s: $\Lambda(\bar{x}, \cdot) = \nu_{\bar{x}}$ for each $\bar{x}$ in $\bar{\mathscr{X}}$. Since these measures form acute angles between them ($\langle \nu_{\bar{x}}, \nu_{\bar{y}} \rangle^* \geq 0$ for all $\bar{x}$ and $\bar{y}$ in $\bar{\mathscr{X}}$) and have disjoint supports if and only if they are orthogonal, one could use the volume of the parallelepiped they form to measure their "joint overlap". The square of this volume is given by the determinant of the Gram matrix:

$$\mathrm{Vol}(\Lambda) = \sqrt{\det(\Gamma)},$$

with Γ the square matrix on $\bar{\mathscr{X}}$ with entries $\Gamma(\bar{x}, \bar{y}) = \left\langle \nu_{\bar{x}}, \nu_{\bar{y}} \right\rangle^*$, that is

$$\Gamma := \Lambda D(1/\mu) \Lambda^t , \tag{16}$$

where $D(1/\mu)$ is the diagonal matrix with entries given by $(1/\mu(x), x \in \mathscr{X})$, and Λ^t is the transpose of Λ. Loosely speaking, the less overlap, the largest the volume.

We will instead use the *squeezing* of Λ, that we define by

$$\mathscr{S}(\Lambda) := \begin{cases} +\infty & \text{if } \det(\Gamma) = 0, \\ \sqrt{\mathrm{Trace}\left(\Gamma^{-1}\right)} \in]0, +\infty[& \text{otherwise,} \end{cases} \tag{17}$$

to measure this "joint overlap". We call it "squeezing" because the $\nu_{\bar{x}}$ and the parallelepiped they form are squeezed when $\mathscr{S}(\Lambda)$ is large. This is also the half diameter of the rectangular parallelepiped that circumscribes the ellipsoid defined by the Gram matrix Γ : this ellipsoid is squeezed too when $\mathscr{S}(\Lambda)$ is large. We note

finally that our squeezing controls the volume of Λ. Indeed, by comparison between harmonic and geometric mean applied to the eigenvalues of the Gram matrix, small squeezing implies large volume: $\mathrm{Vol}(\Lambda)^{1/n}\mathscr{S}(\Lambda) \geq \sqrt{n}$. We will also show in Section 2:

Proposition 6. *Let $(\nu_{\bar{x}}, \bar{x} \in \bar{\mathscr{X}})$ be a collection of m probability measures on $\mathscr{X}$.*

1. *We have*
$$\mathscr{S}(\Lambda) \geq \sqrt{\sum_{\bar{x}\in\bar{\mathscr{X}}} \frac{1}{\|\nu_{\bar{x}}\|^{*2}}}\,. \tag{18}$$
Equality holds if and only if the $(\nu_{\bar{x}}, \bar{x} \in \bar{\mathscr{X}})$ are orthogonal.
2. *Assume that μ is a convex combination of the $(\nu_{\bar{x}}, \bar{x} \in \bar{\mathscr{X}})$. Then,*
$$\mathscr{S}(\Lambda) \geq 1\,.$$
Equality holds if and only if the $(\nu_{\bar{x}}, \bar{x} \in \bar{\mathscr{X}})$ are orthogonal.

Comment: $\mathscr{S}(\Lambda)$ is thus maximal when the $\nu_{\bar{x}}$, $\bar{x} \in \bar{\mathscr{X}}$, are linearly dependent, and minimal when they are orthogonal. Moreover, we know the minimal value of $\mathscr{S}(\Lambda)$, when μ is a convex combination of the $(\nu_{\bar{x}}, \bar{x} \in \bar{\mathscr{X}})$. Note that this is necessarily the case if the convex hull of the $\nu_{\bar{x}}$ is stable under P, i.e. when $\Lambda P = \bar{P}\Lambda$ for some stochastic $\bar{P}$. Indeed it is then stable under $e^{t\mathscr{L}}$ for any $t > 0$ and the rows of $\Lambda e^{t\mathscr{L}}$ converge to μ when t goes to infinity. Note also that we are using "$\ell_2(\mu)$ computations" (through the Gram matrix) to define the squeezing of measures that are normalized in $\ell_1(\mu) \sim \ell_\infty^*(\mu)$ (these are *probability* measures). This proposition shows that such a mixture of norms is not meaningless.

1.3.3 Random forests

Note that the weight function w induces a structure of oriented graph on $\mathscr{X}$, $e = (x,y)$ being an oriented edge if and only if $w(e) := w(x,y) > 0$. Let $\mathscr{E}$ be the set of oriented edges, and $G = (\mathscr{X}, \mathscr{E})$ the oriented graph just defined. An oriented forest ϕ on $\mathscr{X}$ is a collection of rooted trees that are subgraphs of G, oriented from their leaves towards their root. A spanning oriented forest (s.o.f.) on $\mathscr{X}$ is an oriented forest which exhausts the points in $\mathscr{X}$. The set of roots of a spanning oriented forest ϕ is denoted by $\rho(\phi)$.

We introduce now a real parameter $q > 0$, and associate to each oriented forest a weight
$$w_q(\phi) := q^{|\rho(\phi)|} \prod_{e\in\phi} w(e)\,. \tag{19}$$
These weights can be renormalized to define a probability measure on the set of spanning oriented forest,
$$\pi_q(\phi) := \frac{w_q(\phi)}{Z(q)}\,, \tag{20}$$
where the partition function $Z(q)$ is given by

$$Z(q) := \sum_{\phi \text{ s.o.f.}} w_q(\phi)\,. \tag{21}$$

We can sample from π_q by using Wilson's algorithm ([25], [18]) which can be described as follows. Let Φ_c be the current state, an oriented forest, of the spanning oriented forest being constructed. At the beginning, Φ_c has no nodes or edges. While Φ_c is not spanning, i.e., while there is a vertex in $\mathscr{X}$ which is not in the vertex set $V(\Phi_c)$ of Φ_c, perform the following steps:

- Choose a point x in $\mathscr{X} \setminus V(\Phi_c)$, in any deterministic or random way.
- Let evolve the Markov process $(X(t), t \geq 0)$ from x, and stop it at $T_q \wedge H_{V(\Phi_c)}$ with T_q an independent exponential time of parameter q and $H_{V(\Phi_c)}$ the hitting time of $V(\Phi_c)$.
- Erase the loops, in order of appearance, of the trajectory drawn by X to 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 is also a powerful tool to study it. The main strength of this algorithm is the freedom one has in choosing the starting points x's of X.

In the sequel, Φ will denote a random variable defined on some probability space $(\Omega_f, \mathscr{A}_f, \mathbb{P}_q)$, having distribution π_q. The corresponding expectation will be denoted by $\mathbb{E}_q$. We will often work with two independent sources of randomness: the Markov process X, and the random forest Φ. Integration with respect to X starting from x will be denoted by P_x and E_x. When X is started with an initial measure π, we will use the notations P_π and E_π. When we integrate over both randomness, we will use the notations $\mathbb{E}_{x,q}, \mathbb{E}_{\pi,q}$ and $\mathbb{P}_{x,q}, \mathbb{P}_{\pi,q}$. The random forest Φ defines a partition of $\mathscr{X}$, two points being in the same set of the partition if they belong to the same tree. This partition will be denoted by $\mathscr{A}(\Phi)$. A point $x \in \mathscr{X}$ being fixed, τ_x is the tree of Φ containing x, ρ_x its root, and $A(x)$ the unique element of $\mathscr{A}(\Phi)$ containing x.

A theorem of Kirchhoff [10] gives in this context that

$$Z(q) = \det(q\,\mathrm{Id} - \mathscr{L}) = \prod_{j<n} (q + \lambda_j), \tag{22}$$

and this implies (see for example [3] for more details, a proof of (22) and the following proposition):

Proposition 7. *For all $k \in \{0, \cdots, n\}$,*

$$\mathbb{P}_q\left[|\rho(\Phi)| = k\right] = \sum_{\substack{J \subset \{0,\cdots,n-1\} \\ |J|=k}} \prod_{j \in J} \frac{q}{q+\lambda_j} \prod_{j \notin J} \frac{\lambda_j}{q+\lambda_j}\,.$$

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

1.3.4 Approximate solution of $\Lambda P = \bar{P}\Lambda$

Assume that we sampled Φ from π_q for some parameter $q > 0$. For $q' > 0$ we then set

- $\bar{\mathscr{X}} := \rho(\Phi)$;
- For any $\bar{x} \in \bar{\mathscr{X}}$, $\nu_{\bar{x}}(\cdot) := K_{q'}(\bar{x}, \cdot)$ (cf. Equation (4)), i.e. $\Lambda = K_{q'}|_{\bar{\mathscr{X}} \times \mathscr{X}}$;
- $\bar{P}(\bar{x}, \bar{y}) := P_{\bar{x}}\left[X(H^+_{\bar{\mathscr{X}}}) = \bar{y}\right]$ with, for any $A \subset \mathscr{X}$,

$$H^+_A := \inf\{t \geq \sigma_1, X(t) \in A\}\,. \tag{23}$$

  H^+_A is in other words the return time in A, and $\bar{P}$ is the (irreducible and reversible) Markovian kernel associated with the trace chain of X on $\bar{\mathscr{X}}$.

Here $\bar{\mathscr{X}}$ is a random subset of $\mathscr{X}$, and so is its cardinality. If we want to keep approximately m points from $\mathscr{X}$, we have to ensure that

$$\mathbb{E}_q\left[\left|\bar{\mathscr{X}}\right|\right] = \sum_{i=0}^{n-1} \frac{q}{q+\lambda_i} \approx m\,. \tag{24}$$

This can be obtained, starting from any q to sample Φ, by updating q according to $q \leftarrow qm/|\rho(\Phi)|$ before re-sampling Φ and going so up to getting a satisfactory number of roots (see [3] for more details).

Let us remind the definition (14) of α, and let us define

$$p_j := \frac{q}{q+\lambda_j}\,, \quad p'_j := \frac{q'}{q'+\lambda_j}\,, \qquad j < n,$$

and denote by d_{TV} the total variation distance: if ν and ν' are two probability measures on $\mathscr{X}$,

$$d_{TV}(\nu, \nu') = \frac{1}{2}\sum_{x \in \mathscr{X}} \left|\nu(x) - \nu'(x)\right|\,.$$

Theorem 1. *For all $m \in \{1, \cdots, n\}$,*

$$\mathbb{E}_q\left[\sum_{\bar{x} \in \bar{\mathscr{X}}} d_{TV}(\Lambda P(\bar{x}, \cdot), \bar{P}\Lambda(\bar{x}, \cdot)) \,\middle|\, \left|\bar{\mathscr{X}}\right| = m\right] \leq \frac{q'(n-m)}{\alpha}\,, \tag{25}$$

and

$$\mathbb{E}_q\left[\sum_{\bar{x} \in \bar{\mathscr{X}}} d_{TV}(\Lambda P(\bar{x}, \cdot), \bar{P}\Lambda(\bar{x}, \cdot))\right] \leq \frac{q'}{\alpha}\sum_{i=1}^{n-1} \frac{\lambda_i}{q+\lambda_i}\,. \tag{26}$$

In addition, with

$$S_n := \sum_{j=1}^{n-1} p_j'^2 (1-p_j)^2\,;\; T_n := \sum_{j=1}^{n-1} \frac{p_j^2}{p_j'^2}\,;\; V_n = \sum_{j=1}^{n-1} p_j(1-p_j)\,,$$

it holds

$$\mathbb{E}_q\left[\mathscr{S}(\Lambda)\,\middle|\,|\bar{\mathscr{X}}|=m\right] \leq \frac{\min\left\{\sqrt{1+\sqrt{\frac{T_n}{S_n}}}\,\exp\left(\sqrt{S_nT_n}-V_n\right);\sqrt{1+T_n}\exp\left(\frac{(1+S_nT_n))}{2}-V_n\right)\right\}}{\mathbb{P}_q\left[|\bar{\mathscr{X}}|=m\right]} \tag{27}$$

for any $m\in\{1,\cdots,n\}$.

Proof: See Section 3.
Comment: Our upper bounds depend on $\mathscr{L}$ through its spectrum only. They show that if there is a gap in this spectrum —that is if for some $1<m<n$ it holds $\lambda_{m-1}\ll\lambda_m$— then we can have asymptotically exact solutions with small squeezing by choosing $\lambda_{m-1}\ll q\ll q'\ll\lambda_m$. We then have indeed $q'\ll\alpha$ since $\lambda_m\leq 2\alpha$ and $p_j\sim p'_j\sim 1$ for $j<m$, while $p_j\ll p'_j\ll 1$ for $j\geq m$. We can then have a vanishing error in the approximation, see (26). In addition we can have $V_n\ll 1$, $S_n\ll 1$, $T_n\sim m-1$, $\mathbb{P}_q[|\bar{\mathscr{X}}|=m]\sim 1$ (recall Proposition 7) and an upper bound on the mean value of $\mathscr{S}(\Lambda)$ that goes like $\sqrt{m}$. This upper bound has to be compared with the lower bounds of Proposition 6, i.e. with 1 if we have asymptotic solutions of intertwining equations. For some simple low temperature metastable systems as quickly mentioned in Section 1.2.4, there is such a gap in the spectrum and this construction will give indeed asymptotic solutions with $\mathscr{S}(\Lambda)$ going to 1. There is room for improvement in the sense that our approximate solutions can be even less squeezed that what is ensured by the theorem.

1.3.5 Approximate solutions of $\Lambda K_{q'}=\bar{P}\Lambda$

Assume once again that we sampled Φ from π_q for some parameter $q>0$. But let us modify our choices for $\bar{\mathscr{X}}$, Λ and $\bar{P}$, by using this time the partition $\mathscr{A}(\Phi)$. Set:

- $\bar{\mathscr{X}}:=\rho(\Phi)$ (one could rather think that $\bar{\mathscr{X}}$ is the set of the different pieces forming the partition $\mathscr{A}(\Phi)$ but the notation will be simpler by using the set of roots, which obviously is in one to one correspondence through the map $A:\bar{x}\in\rho(\Phi)\mapsto A(\bar{x})$);
- for any $\bar{x}\in\bar{\mathscr{X}}$, $\nu_{\bar{x}}(\cdot):=\mu_{A(\bar{x})}(\cdot)$, with, for any $A\subset\mathscr{X}$, μ_A being defined by the probability μ conditioned to A: $\mu_A:=\mu(\cdot|A)$;
- for any $\bar{x},\bar{y}\in\bar{\mathscr{X}}$, $\bar{P}(\bar{x},\bar{y}):=P_{\mu_{A(\bar{x})}}\left[X(T_{q'})\in A(\bar{y})\right]$, with $T_{q'}$ being as previously an exponential random variable of parameter q' that is independent from X. Irreducibility and reversibility of $\bar{P}$ are then inherited from those of P.

It follows from Proposition 6 that the squeezing of $\left\{\nu_{\bar{x}},\bar{x}\in\bar{\mathscr{X}}\right\}$ is minimal and equal to one.

To bound the distance between $\Lambda K_{q'}$ and $\bar{P}\Lambda$, we introduce another random forest Φ' distributed as $\pi_{q'}$ and independent of Φ and X. For any $x\in\mathscr{X}$, t'_x is the tree

containing x in Φ', ρ'_x its root, $A'(x)$ the unique element of $\mathscr{A}(\Phi')$ containing x, and Γ'_x is the path going from x to ρ'_x in Φ'. By Wilson algorithm started at x, Γ'_x is the trajectory of a loop-erased random walk started from x and stopped at an exponential time $T_{q'}$. We denote by $|\Gamma'_x|$ its length, that is the number of edges to be crossed in Φ' to go from x to ρ'_x.

Theorem 2. *Let $p \geq 1$, and p^* its conjugate exponent, so that $\frac{1}{p} + \frac{1}{p^*} = 1$. Then,*

$$\mathbb{E}_q\left[\sum_{\bar{x}\in\bar{\mathscr{X}}} d_{TV}(\Lambda K_{q'}(\bar{x},\cdot),\bar{P}\Lambda(\bar{x},\cdot))\right] \leq (\mathbb{E}_q\left[|\rho(\Phi)|\right])^{1/p}\left(\frac{q'}{q}\sum_{x\in\mathscr{X}}\mathbb{E}_{q'}\left[|\Gamma'_x|\right]\right)^{1/p^*}.$$

Proof: See Section 4.
Comment: Note that

$$q'\mathbb{E}_{q'}\left[|\Gamma'_x|\right] = \alpha\frac{\mathbb{E}_{q'}\left[|\Gamma'_x|\right]}{\alpha/q'}$$

is, up to the factor α, the ratio between the mean number of steps of the loop-erased random walk and the mean number of steps of the simple random walk up to time $T_{q'}$, that is the time fraction spent outside loops up to time $T_{q'}$. As a consequence "the more recurrent is X on time scale $1/q'$", the smaller is this ratio.

1.3.6 Exact solutions of $\Lambda K_{q'} = \bar{P}\Lambda$

We finally modify the previous random measures $\mu_{A(\bar{x})}$ to build exact solution of Equation (3) for q' small enough. We will use to this end a result due to Micchelli and Willoughby [12]: for any $m > 0$

$$MW_m := \prod_{j\geq m}\frac{1}{\lambda_j}(\mathscr{L} + \lambda_j \mathrm{Id})$$

is a Markovian kernel (one can see [3] for a probabilistic insight into the proof of this result). Assume then that we sampled Φ from π_q for some parameter $q > 0$, let us keep $\bar{\mathscr{X}} = \rho(\Phi)$, but let us now set

$$\nu_{\bar{x}} = \mu_{A(\bar{x})}MW_m, \qquad \bar{x} \in \bar{\mathscr{X}},$$

with $m = |\bar{\mathscr{X}}|$.

Theorem 3. *If the $\nu_{\bar{x}}$ have finite squeezing, then for q' small enough, the $\nu_{\bar{x}}K_{q'}$ are in the convex hull of the $\nu_{\bar{x}}$.*

Proof: See Section 5.
Comment: Since we do not give quantitative bounds on how small q' has to be for the thesis to hold, and we do not bound the squeezing of these $\nu_{\bar{x}}$, Theorem 3 is at first not a very insightful result. However the proof we will give suggests that the $\nu_{\bar{x}}$ are natural candidates for not too squeezed solution associated with some

non-very small q'. It will also give further motivation to use squeezing to measure joint overlap. We actually got to our squeezing definition by looking for quantitative bounds for this theorem.

2 Preliminary results

2.1 Proof of Proposition 6

If Γ is not invertible, points (1) and (2) are obviously true. We assume therefore that Γ is invertible. Let $\tilde{\Lambda} := \Gamma^{-1}\Lambda$, and let $(\tilde{v}_{\bar{x}}, \bar{x} \in \bar{\mathscr{X}})$ be the row vectors of $\tilde{\Lambda}$. Note that

$$\tilde{\Lambda} D(1/\mu)\Lambda^t = \Gamma^{-1}\Lambda D(1/\mu)\Lambda^t = \Gamma^{-1}\Gamma = \mathrm{Id}\,.$$
$$\tilde{\Lambda} D(1/\mu)\tilde{\Lambda}^t = \Gamma^{-1}\Lambda D(1/\mu)\Lambda^t\Gamma^{-1} = \Gamma^{-1}\,.$$

Hence, for all $\bar{x}, \bar{y} \in \mathscr{X}$, $\left\langle \tilde{v}_{\bar{x}}, v_{\bar{y}} \right\rangle^* = \delta_{\bar{x}\bar{y}}$ and $\|\tilde{v}_{\bar{x}}\|^{*2} = (\Gamma^{-1})(\bar{x},\bar{x})$.

1. We have $\mathscr{S}(\Lambda)^2 = \sum_{\bar{x}\in\bar{\mathscr{X}}} \|\tilde{v}_{\bar{x}}\|^{*2} \geq \sum_{\bar{x}\in\bar{\mathscr{X}}} \frac{\langle \tilde{v}_{\bar{x}}, v_{\bar{x}}\rangle^{*2}}{\|v_{\bar{x}}\|^{*2}} = \sum_{\bar{x}\in\bar{\mathscr{X}}} \frac{1}{\|v_{\bar{x}}\|^{*2}}$.
   Assume now that the $v_{\bar{x}}$'s, $\bar{x} \in \bar{\mathscr{X}}$ are orthogonal. $\Gamma = \mathrm{diag}(\|v_{\bar{x}}\|^{*2})$, so that $\mathrm{Trace}(\Gamma^{-1}) = \sum_{\bar{x}\in\bar{\mathscr{X}}} \frac{1}{\|v_{\bar{x}}\|^{*2}}$.
   In the opposite direction, assume instead that $\mathrm{Trace}(\Gamma^{-1}) = \sum_{\bar{x}\in\bar{\mathscr{X}}} \frac{1}{\|v_{\bar{x}}\|^{*2}}$. Then for any $\bar{x} \in \bar{\mathscr{X}}$, $\left|\langle \tilde{v}_{\bar{x}}, v_{\bar{x}}\rangle^*\right| = \|\tilde{v}_{\bar{x}}\|^* \|v_{\bar{x}}\|^*$. This implies that for all $\bar{x} \in \bar{\mathscr{X}}$, there exists a real number $\alpha(\bar{x}) \neq 0$ such that $\tilde{v}_{\bar{x}} = \alpha(\bar{x}) v_{\bar{x}}$. Taking the scalar product with $v_{\bar{y}}$ leads to $\delta_{\bar{x}\bar{y}} = \left\langle \tilde{v}_{\bar{x}}, v_{\bar{y}} \right\rangle^* = \alpha(\bar{x}) \left\langle v_{\bar{x}}, v_{\bar{y}} \right\rangle^*$. Hence $(v_{\bar{x}}, \bar{x} \in \bar{\mathscr{X}})$ are orthogonal.
2. Let us write μ as a convex combination of the $(v_{\bar{x}}, \bar{x} \in \bar{\mathscr{X}})$:

$$\mu = \sum_{\bar{x}\in\bar{\mathscr{X}}} \alpha(\bar{x}) v_{\bar{x}}, \qquad \alpha(\bar{x}) \geq 0, \qquad \sum_{\bar{x}\in\bar{\mathscr{X}}} \alpha(\bar{x}) = 1.$$

   Note that for any probability measure ν, $\langle \mu, \nu\rangle^* = \sum_{x\in\mathscr{X}} \mu(x)\nu(x)/\mu(x) = 1$. As a special case, for any $\bar{y} \in \bar{\mathscr{X}}$,

$$1 = \left\langle \mu, v_{\bar{y}} \right\rangle^* = \sum_{\bar{x}\in\bar{\mathscr{X}}} \alpha(\bar{x}) \left\langle v_{\bar{x}}, v_{\bar{y}} \right\rangle^* \geq \alpha(\bar{y}) \left\|v_{\bar{y}}\right\|^{*2}. \tag{28}$$

   By point (1), we deduce that

$$\mathscr{S}(\Lambda)^2 \geq \sum_{\bar{x}\in\bar{\mathscr{X}}} \frac{1}{\|v_{\bar{x}}\|^{*2}} \geq \sum_{\bar{x}\in\bar{\mathscr{X}}} \alpha(\bar{x}) = 1\,.$$

Equality holds if and only if (28) and (18) are equalities. By point (1), this implies that the $(\nu_{\bar{x}}, \bar{x} \in \bar{\mathscr{X}})$ are orthogonal. In the opposite direction, when the $\nu_{\bar{x}}$, for $\bar{x} \in \bar{\mathscr{X}}$, are orthogonal, (28) and (18) are equalities, and $\mathscr{S}(\Lambda) = 1$.

2.2 Elementary observations on intertwining equations

Consider Equation (2) for any reversible and irreducible stochastic kernel P, and assume an $m \times n$ rectangular stochastic matrix $\Lambda = (\Lambda(\bar{x}, x))_{\bar{x} \in \bar{\mathscr{X}}, x \in \mathscr{X}}$ to be a solution for some $\bar{P}$ with $m \leq n$. Let us write $(\theta_j)_{j<n} = (1 - \lambda_j/\alpha)_{j<n}$ for the n eigenvalues of P in decreasing order:

$$1 = \theta_0 > \theta_1 \geq \cdots \geq \theta_{n-1} \geq -1.$$

We also set $[n] = \{0, 1, 2, \ldots, n-1\}$, call μ the reversible measure of P, and write $\nu_{\bar{x}} = \Lambda(\bar{x}, \cdot)$ for the rows of Λ.

Lemma 1. *Assume Equation* (2) *is in force. If Λ is non-degenerate, i.e., if Λ is of rank m, then there is an orthonormal basis of left eigenvectors $(\mu_j : 0 \leq j < n)$ of P such that*

$$\mu_j P = \theta_j \mu_j, \qquad j < n,$$

there is a subset J of $[n]$ such that $0 \in J$ and $|J| = m$ and there is an invertible matrix $C = (C(\bar{x}, j))_{\bar{x} \in \bar{\mathscr{X}}, j \in J}$ with $C(\bar{x}, 0) = 1$ for all $\bar{x}$ in $\bar{\mathscr{X}}$, such that

$$\nu_{\bar{x}} = \sum_{j \in J} C(\bar{x}, j) \mu_j, \qquad \bar{x} \in \bar{\mathscr{X}}, \tag{29}$$

and

$$\bar{P} C(\cdot, j) = \theta_j C(\cdot, j), \qquad j \in J. \tag{30}$$

In particular, the spectrum of $\bar{P}$ is contained in that of P, with eigenvalue multiplicities that do not exceed the corresponding ones for P.

Proof. Let V be the subspace of $\ell_2^*(\mu)$ spanned by the $\nu_{\bar{x}}$. Since Λ is non-degenerate, V is of dimension m. Since $\Lambda P = \bar{P} \Lambda$, the $\nu_{\bar{x}} P$ are convex combinations of the $\nu_{\bar{x}}$ and V is stable by the self-adjoint operator P. It follows that there is an orthonormal basis of left eigenvectors μ_j, with $\mu_0 = \mu$, a subset $J \subset [n]$ of size m, and an invertible matrix C such that (29) holds. Since for $j > 0$ one has $\langle \mu, \mu_j \rangle^* = 0$, by computing the scalar product with μ of both sides of equations (29), it follows that 0 belongs to J and $C(\bar{x}, 0) = 1$ for each $\bar{x}$.

Now, applying P on both sides of (29) we obtain

$$\sum_{j \in J} \sum_{\bar{y} \in \bar{\mathscr{X}}} \bar{P}(\bar{x}, \bar{y}) C(\bar{y}, j) \mu_j = \sum_{j \in J} \theta_j C(\bar{x}, j) \mu_j, \qquad \bar{x} \in \bar{\mathscr{X}}.$$

By identifying the decomposition coefficients in the basis of the μ_j's, this gives (30). Since the m column vectors $C(\cdot, j)$ are linearly independent, they form a basis of the

functions on $\bar{\mathscr{X}}$. This is why equations (30) completely describe the spectrum of $\bar{P}$ and we can conclude that the spectrum of $\bar{P}$ is contained in that of P with the multiplicity constraint. □

The previous lemma shows on the one hand a localisation property in Fourier space of exact solutions of intertwining equations: the $\nu_{\bar{x}}$ have to be with no component on $n-m$ eigenvectors of the Laplacian $\mathscr{L}$ (see equations (29)). On the other hand, it shows that finding exact solutions of intertwining equation implies to have a detailed knowledge of the eigenvectors of the Laplacian.

Conversely, it is now possible to describe all the non-degenerate solutions of the intertwining equations in terms of, on the one hand, the eigenvectors and eigenvalues of P and, on the other hand, the set of diagonalizable stochastic matrices $\bar{P}$ with a given spectrum contained in that of P, and satisfying the multiplicity constraint. Any right eigenvector basis $(C(\cdot,j) : j \in J)$ —satisfying (30) and with $C(\cdot,0) \equiv 1$— of such a $\bar{P}$ will provide, through equations (29) and possibly after rescaling, a non-degenerate solution of the intertwining equations. The only delicate point to check is indeed the non-negativity of the $\nu_{\bar{x}}$. But if this fails, and since $\mu = \mu_0$ charges all points in $\mathscr{X}$, one just has to replace the $C(\cdot,j)$ for positive j in J, by some $\delta_j C(\cdot,j)$ for some small enough δ_j.

At this point we just have to give sufficient conditions for the set of diagonalisable stochastic matrices with a given spectrum to ensure that our intertwining equations do have solutions. The next lemma shows that, if P has non-negative eigenvalues, then we will find solutions with $\bar{\mathscr{X}}$ of any size $m < n$. We further note that this hypothesis will always be fulfilled if instead of considering P we consider its lazy version $(P+Id)/2$.

Lemma 2. *For any*

$$1 = \theta_0 > \theta_1 \geq \theta_2 \geq \cdots \geq \theta_{m-1} \geq 0$$

there always exists a reversible and irreducible *stochastic matrix $\bar{P}$ with such a spectrum. In particular, if P is a reversible and irreducible stochastic matrix that admits $(\theta_j : j < m)$ as a subsequence of its ordered spectrum with multiplicities, then the Markov chains associated with P and $\bar{P}$ are intertwined.*

Proof. Let us set

$$A = \begin{pmatrix} 1 & -1 & 0 & \cdots & \cdots & 0 \\ 1 & 1 & -2 & \ddots & & \vdots \\ 1 & 1 & 1 & -3 & \ddots & \vdots \\ \vdots & \vdots & \vdots & \ddots & \ddots & 0 \\ \vdots & \vdots & \vdots & & \ddots & -(m-1) \\ 1 & 1 & 1 & \cdots & \cdots & 1 \end{pmatrix},$$

a matrix with orthogonal rows, and introduce the diagonal matrices

$$D_\theta = \begin{pmatrix} \theta_0 & & & & \\ & \theta_1 & & & \\ & & \ddots & & \\ & & & \ddots & \\ & & & & \theta_{m-1} \end{pmatrix}, \qquad D_{\bar\mu} = \begin{pmatrix} \frac{1}{1\times 2} & & & & \\ & \frac{1}{2\times 3} & & & \\ & & \ddots & & \\ & & & \frac{1}{(m-1)m} & \\ & & & & \frac{1}{m} \end{pmatrix},$$

the second one being such that $Q = D_{\bar\mu}^{1/2} A$ is orthogonal. We compute

$$\bar P = D_{\bar\mu}^{-1/2} Q D_\theta Q^t D_{\bar\mu}^{1/2} = A D_\theta A^t D_{\bar\mu}$$

to find

$$\bar P = \begin{pmatrix} \frac{\Sigma_1+\theta_1}{1\times 2} & \frac{\Sigma_1-\theta_1}{2\times 3} & \frac{\Sigma_1-\theta_1}{3\times 4} & \cdots & \frac{\Sigma_1-\theta_1}{(m-1)m} & \frac{\Sigma_1-\theta_1}{m} \\ \frac{\Sigma_1-\theta_1}{1\times 2} & \frac{\Sigma_2+2^2\theta_2}{2\times 3} & \frac{\Sigma_2-2\theta_2}{3\times 4} & \cdots & \frac{\Sigma_2-2\theta_2}{(m-1)m} & \frac{\Sigma_2-2\theta_2}{m} \\ \frac{\Sigma_1-\theta_1}{1\times 2} & \frac{\Sigma_2-2\theta_2}{2\times 3} & \frac{\Sigma_3+3^2\theta_3}{3\times 4} & \cdots & \frac{\Sigma_3-3\theta_3}{(m-1)m} & \frac{\Sigma_3-3\theta_3}{m} \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ \frac{\Sigma_1-\theta_1}{1\times 2} & \frac{\Sigma_2-2\theta_2}{2\times 3} & \frac{\Sigma_3-3\theta_3}{3\times 4} & \cdots & \frac{\Sigma_{m-1}+(m-1)^2\theta_{m-1}}{(m-1)m} & \frac{\Sigma_{m-1}-(m-1)\theta_{m-1}}{m} \\ \frac{\Sigma_1-\theta_1}{1\times 2} & \frac{\Sigma_2-2\theta_2}{2\times 3} & \frac{\Sigma_3-3\theta_3}{3\times 4} & \cdots & \frac{\Sigma_{m-1}-(m-1)\theta_{m-1}}{(m-1)m} & \frac{\Sigma_m}{m} \end{pmatrix} \tag{31}$$

with, for all $1 \le k \le m$, $\Sigma_k = \sum_{j<k} \theta_j$. $\bar P$ is stochastic, irreducible and reversible with respect to $\bar\mu$ defined by

$$\bar\mu(k) = \begin{cases} \frac{1}{k(k+1)} & \text{if } k < m, \\ \frac{1}{m} & \text{if } k = m. \end{cases}$$

It also has the desired spectrum. □

Comment: The proof actually shows that the positivity hypothesis on the θ_j's can be slightly relaxed: we only have to require the numerators of the diagonal coefficients in (31) to be non-negative.

We conclude this section by observing that the universal solution we just provided is not fully satisfactory. First, it requires a detailed knowledge of the spectrum that can be practically unavailable. Second, we can expect such a universal solution to produce very squeezed solutions. Indeed, the coefficients $C(\bar x, j)$ in (30) will be given by the matrix $C = D_{\bar\mu}^{-1/2} Q = A$ or by $C = A D_\delta$ with D_δ a rescaling diagonal matrix

$$D_\delta = \begin{pmatrix} 1 & & & \\ & \delta_1 & & \\ & & \ddots & \\ & & & \delta_{m-1} \end{pmatrix}$$

ensuring the non-negativity of the $\nu_{\bar x}$. The fact that the δ_i's may have to be chosen very small can be the source of very strong squeezing.

2.3 Proof of Proposition 2

Let $\xi_{\bar{x}}^+$ and $\xi_{\bar{x}}^-$ be, for each $\bar{x}$ in $\bar{\mathscr{X}}$, the positive and negative part of the signed measure $\xi_{\bar{x}} = \xi_{\bar{x}}^+ - \xi_{\bar{x}}^-$ such that

$$\nu_{\bar{x}} P = \xi_{\bar{x}} + \sum_{\bar{y} \in \bar{\mathscr{X}}} \bar{P}(\bar{x}, \bar{y}) \nu_{\bar{y}}. \tag{32}$$

Since $\nu_{\bar{x}} P$ and the convex combination of the $\nu_{\bar{y}}$ are both probability measures, $\xi_{\bar{x}}^+$ and $\xi_{\bar{x}}^-$ have the same mass

$$\varepsilon_{\bar{x}} = d_{\mathrm{TV}} \left(\nu_{\bar{x}} P, \sum_{\bar{y} \in \bar{\mathscr{X}}} \bar{P}(\bar{x}, \bar{y}) \nu_{\bar{y}} \right).$$

Adding $\xi_{\bar{x}}^-$ on both sides of Equation (32) and dividing by $1 + \varepsilon_{\bar{x}}$, we get

$$\frac{1}{1+\varepsilon_{\bar{x}}} \nu_{\bar{x}} P + \frac{\varepsilon_{\bar{x}}}{1+\varepsilon_{\bar{x}}} \pi_{\bar{x}}^- = \frac{\varepsilon_{\bar{x}}}{1+\varepsilon_{\bar{x}}} \pi_{\bar{x}}^+ + \frac{1}{1+\varepsilon_{\bar{x}}} \sum_{\bar{y} \in \bar{\mathscr{X}}} \bar{P}(\bar{x}, \bar{y}) \nu_{\bar{y}}, \tag{33}$$

where $\pi_{\bar{x}}^+$ and $\pi_{\bar{x}}^-$ are the probability measures obtained by normalization from $\xi_{\bar{x}}^+$ and $\xi_{\bar{x}}^-$.

Let us build a new set $\tilde{\mathscr{X}}$ by associating some $\bar{x}'$ with each $\bar{x}$ in $\bar{\mathscr{X}}$, calling $\bar{\mathscr{X}}'$ the set of these associated $\bar{x}'$ and setting $\tilde{\mathscr{X}} = \bar{\mathscr{X}} \cup \bar{\mathscr{X}}'$ to get a twice as large set. We can then read Equation (33) as an exact intertwining equation at site $\bar{x}$, between a Markov chain $Z_{\bar{x}}$ with values in $\mathscr{X}$ and transition probabilities

$$P_{\bar{x}}(x, y) = \frac{1}{1+\varepsilon_{\bar{x}}} P(x, y) + \frac{\varepsilon_{\bar{x}}}{1+\varepsilon_{\bar{x}}} \pi_{\bar{x}}^-(y), \qquad x, y \in \mathscr{X},$$

on the one hand, and on the other hand a Markov chain $\bar{Z}$ with values in the augmented set $\tilde{\mathscr{X}}$ and transition probabilities

$$\tilde{P}(\bar{x}, \bar{y}) = \frac{1}{1+\varepsilon_{\bar{x}}} \bar{P}(\bar{x}, \bar{y}), \qquad \bar{x} \in \bar{\mathscr{X}}, \qquad \bar{y} \in \bar{\mathscr{X}},$$

$$\tilde{P}(\bar{x}, \bar{y}') = \frac{\varepsilon_{\bar{x}}}{1+\varepsilon_{\bar{x}}} \mathbb{1}_{\{\bar{y}' = \bar{x}'\}}, \qquad \bar{x} \in \bar{\mathscr{X}}, \qquad \bar{y}' \in \bar{\mathscr{X}}',$$

$$\tilde{P}(\bar{x}', \tilde{y}) = \mathbb{1}_{\{\tilde{y} = \bar{x}'\}}, \qquad \bar{x}' \in \bar{\mathscr{X}}', \qquad \tilde{y} \in \tilde{\mathscr{X}}.$$

The linking probabilities at our $\bar{x}$ in Equation (33) are the $\nu_{\bar{y}}$ and $\nu_{\bar{x}'} = \pi_{\bar{x}}^+$. The process $Z_{\bar{x}}$ can be constructed with a sequence of independent uniform random variables $(U_k, k \geq 1)$. Assuming that at time k, $Z_{\bar{x}}(k) = z$ and that $U_{k+1} > \frac{\varepsilon_{\bar{x}}}{1+\varepsilon_{\bar{x}}}$, $Z_{\bar{x}}(k+1)$ is sampled with $P(z, \cdot)$, while if $U_{k+1} \leq \frac{\varepsilon_{\bar{x}}}{1+\varepsilon_{\bar{x}}}$, $Z_{\bar{x}}(k+1)$ is sampled with $\pi_{\bar{x}}^-$.

By Proposition 1, we get a stopping time $T_{\bar{x}}$ and a random variable $\tilde{Y}_{\bar{x}}$, with values in $\{\bar{x}'\} \cup \bar{\mathscr{X}} \setminus \{\bar{x}\}$, such that, conditionally to $\tilde{Y}_{\bar{x}}$, the law of $Z_{\bar{x}}(T_{\bar{x}})$ is $\nu_{\tilde{Y}_{\bar{x}}}$. For $\nu = \nu_{\bar{x}}$ we define then the Markov chain $(Z,\bar{Z})$ on $\mathscr{X} \times \bar{\mathscr{X}}$ in the following way:

- the law of $(Z,\bar{Z})(0)$ is $\nu_{\bar{x}} \otimes \delta_{\bar{x}}$;
- for $k < T_{\bar{x}}$, $(Z,\bar{Z})(k) = (Z_{\bar{x}}(k),\bar{x})$;
- $(Z,\bar{Z})(T_{\bar{x}}) = (Z_{\bar{x}}(T_{\bar{x}}),\tilde{Y}_{\bar{x}})$.
- If $\tilde{Y}_{\bar{x}} = \bar{x}'$ then we set $(Z,\bar{Z})(k) = (Z,\bar{Z})(T_{\bar{x}})$ for all $k \geq T_{\bar{x}}$. Otherwise $\tilde{Y}_{\bar{x}}$ plays now the previous role of $\bar{x}$.

This construction is naturally adapted to $\bar{Z}$ started in $\bar{\nu}$ and Z in $\nu_{\bar{Z}}$. We get the desired equality on the law of Z_k conditioned on $\bar{Z}_k$ as a consequence of properties (2) and (4) of Proposition 1.

By properties (1), (3) and (5) of Proposition 1, $\bar{Z}$ and $\bar{X}^{\bar{\nu}}$ have the same law before the absorbing time

$$\bar{K} = \min\left\{k > 0 : \bar{Z}_k \notin \bar{\mathscr{X}}\right\}.$$

We can then set $\bar{X}_k^{\bar{\nu}} = \bar{Z}_k$ for $k < \bar{K}$, and build $\bar{X}^{\bar{\nu}}$ independently from $(Z,\bar{Z})$ for $k \geq \bar{K}$. Setting

$$K = \min\left\{k < \bar{K} : U_k \leq \frac{\varepsilon_{\bar{Z}_k}}{1+\varepsilon_{\bar{Z}_k}}\right\},$$

with the usual convention that the minimum of the empty set is $+\infty$, we can also set $\hat{X}_k^{\nu} = Z_k$ for $k < K$, and build $\hat{X}^{\nu}$ independently of $(Z,\bar{Z})$ for $k \geq K$. We simply conclude this coupling construction by observing that K and $\bar{K}$ stochastically dominate two (correlated) geometric random variables T and $\bar{T}$ with success probability

$$\varepsilon \geq \max_{\bar{x}\in\bar{\mathscr{X}}} \frac{\varepsilon_{\bar{x}}}{1+\varepsilon_{\bar{x}}}.$$

Let us finally explain why this implies our claimed upper bound on the total variation distance between the law of $\hat{X}_k^{\nu}$ and

$$\xi_k = \sum_{\bar{x}\in\bar{\mathscr{X}}} P\big(\bar{X}_k^{\bar{\nu}} = \bar{x}\big)\nu_{\bar{x}}$$

for any $k \geq 0$. For any $A \subset \mathscr{X}$ it holds

$$\begin{aligned}
\xi_k(A) &= \sum_{\bar{x}\in\bar{\mathscr{X}}} P\big(\bar{X}_k^{\bar{\nu}} = \bar{x}, \bar{T} \leq k\big)\nu_{\bar{x}}(A) + \sum_{\bar{x}\in\bar{\mathscr{X}}} P\big(\bar{X}_k^{\bar{\nu}} = \bar{x}, \bar{T} > k\big)\nu_{\bar{x}}(A) \\
&\leq \sum_{\bar{x}\in\bar{\mathscr{X}}} P\big(\bar{X}_k^{\bar{\nu}} = \bar{x}, \bar{T} \leq k\big) + \sum_{\bar{x}\in\bar{\mathscr{X}}} P\big(\bar{Z}_k = \bar{x}\big)\nu_{\bar{x}}(A) \\
&= P\big(\bar{T} \leq k\big) + \sum_{\bar{x}\in\bar{\mathscr{X}}} P\big(\bar{Z}_k = \bar{x}, Z_k \in A\big) \\
&\leq P\big(\bar{T} \leq k\big) + P\big(Z_k \in A\big) \\
&\leq P\big(\bar{T} \leq k\big) + P\big(T \leq k\big) + P\big(\hat{X}_k^{\nu} \in A\big).
\end{aligned}$$

The same inequality holds with the complementary of A and this concludes the proof.

3 Proof of Theorem 1

3.1 Total variation estimates

Inequality (26) is a direct consequence of Inequality (25) and Proposition 7. Indeed,

$$\begin{aligned}
&\mathbb{E}_q\left[\sum_{\bar{x}\in\bar{\mathscr{X}}} d_{TV}(\Lambda P(\bar{x},\cdot),\bar{P}\Lambda(\bar{x},\cdot))\right]\\
&\quad=\sum_{i=1}^{n}\mathbb{E}_q\left[\sum_{\bar{x}\in\bar{\mathscr{X}}} d_{TV}(\Lambda P(\bar{x},\cdot),\bar{P}\Lambda(\bar{x},\cdot))\,\middle|\,\left|\bar{\mathscr{X}}\right|=i\right]\mathbb{P}_q\left[\left|\bar{\mathscr{X}}\right|=i\right]\\
&\quad\le\sum_{i=1}^{n}\frac{q'(n-i)}{\alpha}\mathbb{P}_q\left[\left|\bar{\mathscr{X}}\right|=i\right]\\
&\quad=\frac{q'}{\alpha}\mathbb{E}_q\left[n-\left|\bar{\mathscr{X}}\right|\right].
\end{aligned}$$

It remains thus to prove (25). Applying Markov property at time σ_1, we get

$$\bar{P}(\bar{x},\cdot)=\sum_{y\in\mathscr{X}}P(\bar{x},y)P_y\left[X(H_{\bar{\mathscr{X}}})=\cdot\right].$$

Moreover, set $\delta_{\bar{x}}$ the Dirac measure at $\bar{x}$, seen both as a probability measure on $\mathscr{X}$ and as a row vector of dimension n. Then, we can rewrite

$$\begin{aligned}
\Lambda P(\bar{x},\cdot)=\delta_{\bar{x}}K_{q'}P(\cdot)=\delta_{\bar{x}}PK_{q'}(\cdot)&=\sum_{y\in\mathscr{X}}P(\bar{x},y)P_y\left[X(T_{q'})=\cdot\right]\\
&=\sum_{y\in\mathscr{X}}P(\bar{x},y)P_y\left[H_{\bar{\mathscr{X}}}<T_{q'};X(T_{q'})=\cdot\right]\\
&\quad+\sum_{y\in\mathscr{X}}P(\bar{x},y)P_y\left[H_{\bar{\mathscr{X}}}\ge T_{q'};X(T_{q'})=\cdot\right]\\
&=\sum_{y\in\mathscr{X},\bar{z}\in\bar{\mathscr{X}}}P(\bar{x},y)P_y\left[H_{\bar{\mathscr{X}}}<T_{q'};X(H_{\bar{\mathscr{X}}})=\bar{z}\right]P_{\bar{z}}\left[X(T_{q'})=\cdot\right]\\
&\quad+\sum_{y\in\mathscr{X}}P(\bar{x},y)P_y\left[H_{\bar{\mathscr{X}}}\ge T_{q'};X(T_{q'})=\cdot\right]\\
&=\bar{P}\Lambda(\bar{x},\cdot)-\sum_{y\in\mathscr{X},\bar{z}\in\bar{\mathscr{X}}}P(\bar{x},y)P_y\left[H_{\bar{\mathscr{X}}}\ge T_{q'};X(H_{\bar{\mathscr{X}}})=\bar{z}\right]P_{\bar{z}}\left[X(T_{q'})=\cdot\right]\\
&\qquad+\sum_{y\in\mathscr{X}}P(\bar{x},y)P_y\left[H_{\bar{\mathscr{X}}}\ge T_{q'};X(T_{q'})=\cdot\right].
\end{aligned}$$

Therefore,

$$
\begin{aligned}
&d_{TV}(\Lambda P(\bar{x},\cdot),\bar{P}\Lambda(\bar{x},\cdot)) \\
&\quad= \frac{1}{2}\sum_{x\in\mathscr{X}} |\Lambda P(\bar{x},x)-\bar{P}\Lambda(\bar{x},x)| \\
&\quad\le \frac{1}{2}\sum_{x\in\mathscr{X},y\in\mathscr{X},\bar{z}\in\bar{\mathscr{X}}} P(\bar{x},y)P_y\left[H_{\bar{\mathscr{X}}}\ge T_{q'};X(H_{\bar{\mathscr{X}}})=\bar{z}\right]P_{\bar{z}}\left[X(T_{q'})=x\right] \\
&\qquad+\frac{1}{2}\sum_{x\in\mathscr{X},y\in\mathscr{X}} P(\bar{x},y)P_y\left[H_{\bar{\mathscr{X}}}\ge T_{q'};X(T_{q'})=x\right] \\
&\quad= \sum_{y\in\mathscr{X}} P(\bar{x},y)P_y\left[H_{\bar{\mathscr{X}}}\ge T_{q'}\right] \\
&\quad= \sum_{y\in\mathscr{X}} P(\bar{x},y)E_y\left[1-e^{-q'H_{\bar{\mathscr{X}}}}\right] Ł \\
&\quad\le \sum_{y\in\mathscr{X}} P(\bar{x},y)E_y\left[q'H_{\bar{\mathscr{X}}}\right]=q'E_{\bar{x}}\left[H^+_{\bar{\mathscr{X}}}-\sigma_1\right].
\end{aligned}
$$

We now take the expectation with respect to $\mathbb{E}_q$.

$$
\begin{aligned}
&\mathbb{E}_q\left[\sum_{\bar{x}\in\bar{\mathscr{X}}} d_{TV}(\Lambda P(\bar{x},\cdot),\bar{P}\Lambda(\bar{x},\cdot))\,\middle|\,\left|\bar{\mathscr{X}}\right|=m\right] \\
&\qquad\le q'\mathbb{E}_q\left[\sum_{\bar{x}\in\bar{\mathscr{X}}} E_{\bar{x}}\left[H^+_{\bar{\mathscr{X}}}-\sigma_1\right]\,\middle|\,\left|\bar{\mathscr{X}}\right|=m\right].
\end{aligned}
$$

Formula (11) gives then the desired result.

3.2 Squeezing estimates

We now prove the quantitative upper bounds on the squeezing of Λ stated in (27). We begin with the following lemma:

Lemma 3. *For any $m\in\{1,\cdots,n\}$,*

$$
\begin{aligned}
&\mathbb{E}_q\left[\mathscr{S}(\Lambda)\,\middle|\,\left|\bar{\mathscr{X}}\right|=m\right] \\
&\qquad\le \frac{\sqrt{\sum_{|J|=m-1}\prod_{j\in J}p_j'^2}\sqrt{\sum_{|J|=m}\prod_{j\in J}p_j'^{-2}\prod_{j\in J}p_j^2\prod_{j\notin J}(1-p_j)^2}}{\sum_{|J|=m}\prod_{j\in J}p_j\prod_{j\notin J}(1-p_j)}. \qquad (34)
\end{aligned}
$$

Proof. Note first that

$$\mathscr{S}(\Lambda)^2 = \sum_{\bar{x}\in\bar{\mathscr{X}}} \Gamma^{-1}(\bar{x},\bar{x}) = \sum_{\bar{x}\in\bar{\mathscr{X}}} \frac{\det_{\bar{\mathscr{X}}\setminus\{\bar{x}\}}(\Gamma)}{\det(\Gamma)} = \sum_{\bar{x}\in\bar{\mathscr{X}}} \frac{\mathrm{Vol}^2(v_{\bar{y}};\bar{y}\in\bar{\mathscr{X}},\bar{y}\neq\bar{x})}{\mathrm{Vol}^2(v_{\bar{y}};\bar{y}\in\bar{\mathscr{X}})}.$$

Hence,

$$\begin{aligned}\mathbb{E}_q&\left[\mathscr{S}(\Lambda)\,\middle|\,\left|\bar{\mathscr{X}}\right|=m\right]\\ &= \sum_{|R|=m}\mathbb{P}_q\left[\bar{\mathscr{X}}=R\,\middle|\,\left|\bar{\mathscr{X}}\right|=m\right]\frac{\sqrt{\sum_{\bar{x}\in R}\mathrm{Vol}^2(v_{\bar{y}};\bar{y}\in R,\bar{y}\neq\bar{x})}}{\sqrt{\mathrm{Vol}^2(v_{\bar{y}};\bar{y}\in R)}}.\end{aligned} \tag{35}$$

From Proposition 4, $\bar{\mathscr{X}}=\rho(\Phi)$ is a determinantal process associated to the kernel K_q. Remind that for all $j\in\{0,\cdots,n-1\}$, $\mu_j(-\mathscr{L})=\lambda_j\mu_j$. The μ_j are orthogonal by symmetry of $-\mathscr{L}$, and we assume that for all $j\in\{0,\cdots,n-1\}$, $\left\|\mu_j\right\|^*=1$, so that $\mu_0=\mu$. Hence, we get $\mu_j K_q=\frac{q}{q+\lambda_j}\mu_j$. One way to construct $\rho(\Phi)$, the number of roots being fixed equal to m, is to choose m eigenvectors of K_q, according to Bernoulli random variables with parameters p_j, and then to choose $\bar{\mathscr{X}}$ according to the determinantal process associated to the projector operator onto the m chosen eigenvectors. More formally,

$$\begin{aligned}\mathbb{P}_q&\left[\bar{\mathscr{X}}=R\,\middle|\,\left|\bar{\mathscr{X}}\right|=m\right]\\ &= \frac{1}{Z_{m,q}}\sum_{|J|=m}\prod_{j\in J}\frac{q}{q+\lambda_j}\prod_{j\notin J}\frac{\lambda_j}{q+\lambda_j}\det{}^2\left(\left\langle\frac{\delta_{\bar{x}}}{\|\delta_{\bar{x}}\|^*};\mu_j\right\rangle^*_{\bar{x}\in R,j\in J}\right),\end{aligned} \tag{36}$$

where $Z_{m,q}$ is a normalizing constant ($Z_{m,q}=\mathbb{P}_q\left[\left|\bar{\mathscr{X}}\right|=m\right]$). We go back to (35) and turn to the term $\mathrm{Vol}^2(v_{\bar{y}};\bar{y}\in R)$. It follows from Cauchy-Binet formula that

$$\mathrm{Vol}^2(v_{\bar{y}};\bar{y}\in R)=\sum_{|J|=m}\det{}^2\left(\left\langle v_{\bar{y}},\mu_j\right\rangle^*,\bar{y}\in R,j\in J\right).$$

Note that

$$v_{\bar{y}}=\delta_{\bar{y}}K_{q'}=\sum_{j=0}^{n-1}\left\langle\delta_{\bar{y}};\mu_j\right\rangle^*\mu_j K_{q'}=\sum_{j=0}^{n-1}p'_j\left\langle\delta_{\bar{y}};\mu_j\right\rangle^*\mu_j.$$

. Thus $\left\langle v_{\bar{y}},\mu_j\right\rangle^*=p'_j\left\langle\delta_{\bar{y}};\mu_j\right\rangle^*$. We obtain then

$$\mathrm{Vol}^2(v_{\bar{y}};\bar{y}\in R)=\sum_{|J|=m}\prod_{j\in J}p'^2_j\det{}^2\left(\left\langle\delta_{\bar{y}};\mu_j\right\rangle^*,\bar{y}\in R,j\in J\right). \tag{37}$$

Putting (36) and (37) into (35), we are led to

$$\mathbb{E}_q\left[\mathscr{S}(\Lambda)\,\middle|\,\left|\bar{\mathscr{X}}\right|=m\right]=\frac{1}{Z_{m,q}}\sum_{|R|=m}\frac{\sqrt{\sum_{\bar{x}\in R}\mathrm{Vol}^2(v_{\bar{y}};\bar{y}\in R,\bar{y}\neq\bar{x})}}{\prod_{\bar{x}\in R}\|\delta_{\bar{x}}\|^{*2}}$$
$$\times\frac{\sum_{|J|=m}\prod_{j\in J}p_j\prod_{j\notin J}(1-p_j)\det^2\left(\left\langle\delta_{\bar{x}};\mu_j\right\rangle^*,\bar{x}\in R,j\in J\right)}{\sqrt{\sum_{|J|=m}\prod_{j\in J}p_j'^2\det^2\left(\left\langle\delta_{\bar{y}};\mu_j\right\rangle^*,\bar{y}\in R,j\in J\right)}}.$$

Cauchy-Schwartz inequality then yields

$$\frac{\sum_{|J|=m}\prod_{j\in J}p_j\prod_{j\notin J}(1-p_j)\det^2\left(\left\langle\delta_{\bar{x}};\mu_j\right\rangle^*,\bar{x}\in R,j\in J\right)}{\sqrt{\sum_{|J|=m}\prod_{j\in J}p_j'^2\det^2\left(\left\langle\delta_{\bar{y}};\mu_j\right\rangle^*,\bar{y}\in R,j\in J\right)}}$$
$$\leq\sqrt{\sum_{|J|=m}\prod_{j\in J}\frac{p_j^2}{p_j'^2}\prod_{j\notin J}(1-p_j)^2\det^2\left(\left\langle\delta_{\bar{x}};\mu_j\right\rangle^*,\bar{x}\in R,j\in J\right)}$$

and

$$\mathbb{E}_q\left[\mathscr{S}(\Lambda)\,\middle|\,\left|\bar{\mathscr{X}}\right|=m\right]$$
$$\leq\frac{1}{Z_{m,q}}\sqrt{\sum_{|R|=m}\frac{\sum_{\bar{x}\in R}\mathrm{Vol}^2(v_{\bar{y}};\bar{y}\in R,\bar{y}\neq\bar{x})}{\prod_{\bar{x}\in R}\|\delta_{\bar{x}}\|^{*2}}}$$
$$\times\sqrt{\sum_{\substack{|R|=m\\|J|=m}}\prod_{j\in J}\frac{p_j^2}{p_j'^2}\prod_{j\notin J}(1-p_j)^2\det^2\left(\left\langle\frac{\delta_{\bar{x}}}{\|\delta_{\bar{x}}\|^*};\mu_j\right\rangle^*;\bar{x}\in R,j\in J\right)}.$$

Using again Cauchy-Binet formula, we get

$$\sum_{|R|=m}\det^2\left(\left\langle\frac{\delta_{\bar{x}}}{\|\delta_{\bar{x}}\|^*};\mu_j\right\rangle^*,\bar{x}\in R,j\in J\right)=\mathrm{Vol}^2(\mu_j,j\in J)=1\,,$$

so that the term in the second square root is equal to

$$\sum_{|J|=m}\prod_{j\in J}\frac{p_j^2}{p_j'^2}\prod_{j\notin J}(1-p_j)^2\,.$$

We turn now to the term in the first square root, which can be rewritten, by using twice the Cauchy-Binet formula, as

$$\begin{aligned}
&\sum_{\bar{x}\in\mathscr{X}} \frac{1}{\|\delta_{\bar{x}}\|^{*2}} \sum_{|R|=m,\bar{x}\in R} \frac{\mathrm{Vol}^2(v_{\bar{y}};\bar{y}\in R,\bar{y}\neq\bar{x})}{\prod_{\bar{x}\in R\setminus\{\bar{x}\}} \|\delta_{\bar{x}}\|^{*2}} \\
&\quad= \sum_{\bar{x}\in\mathscr{X}} \mu(\bar{x}) \sum_{R\subset\mathscr{X}\setminus\{\bar{x}\},|R|=m-1} \frac{\mathrm{Vol}^2(v_{\bar{y}};\bar{y}\in R)}{\prod_{\bar{y}\in R} \|\delta_{\bar{y}}\|^{*2}} \\
&\quad= \sum_{\bar{x}\in\mathscr{X}} \mu(\bar{x}) \sum_{R\subset\mathscr{X}\setminus\{\bar{x}\},|R|=m-1} \sum_{|J|=m-1} \prod_{j\in J} p_j'^2 \det{}^2\left(\left\langle \frac{\delta_{\bar{y}}}{\|\delta_{\bar{y}}\|^*};\mu_j\right\rangle^*;\bar{y}\in R, j\in J\right) \\
&\quad\leq \sum_{\bar{x}\in\mathscr{X}} \mu(\bar{x}) \sum_{|J|=m-1} \prod_{j\in J} p_j'^2 \sum_{|R|=m-1} \det{}^2\left(\left\langle \frac{\delta_{\bar{y}}}{\|\delta_{\bar{y}}\|^*};\mu_j\right\rangle^*,\bar{y}\in R, j\in J\right) \\
&\quad= \sum_{|J|=m-1} \prod_{j\in J} p_j'^2 \,\mathrm{Vol}^2(\mu_j, j\in J) \\
&\quad= \sum_{|J|=m-1} \prod_{j\in J} p_j'^2 .
\end{aligned}$$

To end the proof of the lemma, it is sufficient to note that

$$Z_{m,q} = \mathbb{P}_q\left(\left|\bar{\mathscr{X}}\right| = m\right) = \sum_{|J|=m} \prod_{j\in J} p_j \prod_{j\notin J}(1-p_j).$$

□

We can now conclude the proof of (27) and of Theorem 1. For any $t>0$ it holds

$$\sum_{|J|=m-1} \prod_{j\in J} p_j'^2 \leq \frac{1}{t^{m-1}} \prod_{j=0}^{n-1}(1+tp_j'^2) = \frac{1+t}{t^{m-1}} \prod_{j=1}^{n-1}(1+tp_j'^2).$$

since the left-hand is the coefficient of t^{m-1} in the product $\prod_{j=0}^{n-1}(1+tp_j'^2)$. In the same way, for any $x>0$,

$$\begin{aligned}
&\sum_{|J|=m-1,J\subset\{1,\cdots,n-1\}} \prod_{j\in J} \frac{p_j^2}{p_j'^2} \prod_{j\in\{1,\cdots,n-1\}\setminus J} (1-p_j)^2 \\
&\qquad= \prod_{j=1}^{n-1}(1-p_j)^2 \sum_{|J|=m-1,J\subset\{1,\cdots,n-1\}} \prod_{j\in J} \frac{p_j^2}{p_j'^2(1-p_j)^2} \\
&\qquad\leq \frac{\prod_{j=1}^{n-1}(1-p_j)^2}{x^{m-1}} \prod_{j=1}^{n-1}\left(1+x\frac{p_j^2}{p_j'^2(1-p_j)^2}\right)
\end{aligned}$$

$$= \frac{1}{x^{m-1}} \prod_{j=1}^{n-1} \left((1-p_j)^2 + x \frac{p_j^2}{p_j'^2} \right).$$

Hence, for any $x, t > 0$, for any $m \in \{1, \cdots, n\}$,

$$\begin{aligned}
&\mathbb{E}_q \left[\mathscr{S}(\Lambda) \,\middle|\, |\bar{\mathscr{X}}| = m \right] \\
&\quad \leq \frac{1}{\mathbb{P}_q \left[|\bar{\mathscr{X}}| = m \right]} \frac{\sqrt{1+t}}{(tx)^{\frac{m-1}{2}}} \sqrt{\prod_{j=1}^{n-1} (1+tp_j'^2) \left((1-p_j)^2 + x \frac{p_j^2}{p_j'^2} \right)}.
\end{aligned}$$

One can check that

$$(1+tp_j'^2) \left((1-p_j)^2 + x \frac{p_j^2}{p_j'^2} \right) = \left(1 + (\sqrt{xt} - 1) p_j\right)^2 + \left(\sqrt{t} p_j' (1-p_j) - \sqrt{x} \frac{p_j}{p_j'} \right)^2.$$

Take now $xt = 1$. We obtain that for any $t > 0$, for any $m \in \{1, \cdots, n\}$,

$$\begin{aligned}
\mathbb{E}_q \left[\mathscr{S}(\Lambda) \mathbb{1}_{|\bar{\mathscr{X}}| = m} \right] &\leq \sqrt{1+t} \sqrt{\prod_{j=1}^{n-1} \left(1 + \left(\sqrt{t} p_j' (1-p_j) - \frac{1}{\sqrt{t}} \frac{p_j}{p_j'} \right)^2 \right)} \\
&\leq \sqrt{1+t} \exp \left(\frac{1}{2} \sum_{j=1}^{n-1} \left(\sqrt{t} p_j' (1-p_j) - \frac{1}{\sqrt{t}} \frac{p_j}{p_j'} \right)^2 \right) \\
&= \sqrt{1+t} \exp \left(\frac{t}{2} \sum_{j=1}^{n-1} p_j'^2 (1-p_j)^2 + \frac{1}{2t} \sum_{j=1}^{n-1} \frac{p_j^2}{p_j'^2} - \sum_{j=1}^{n-1} p_j (1-p_j) \right).
\end{aligned}$$

Optimizing the exponential term in t and choosing $t = T_n$ lead to (27).

4 Proof of Theorem 2

Let us first rewrite $K_{q'}$ in terms of μ.

Lemma 4. *For $x \in \mathscr{X}$,*

$$K_{q'}(x, \cdot) = \mathbb{E}_{q'} \left[\mu_{A'(x) \cap A(\rho_x')}(\cdot) \right], \mathbb{P}_q \, a.s..$$

Proof. Starting Wilson's algorithm from x to construct Φ', we get

$$\begin{aligned}
K_{q'}(x, y) &= P_x \left[X(T_{q'}) = y \right] \\
&= \mathbb{P}_{q'} \left[\rho_x' = y \right] \\
&= \mathbb{E}_{q'} \left[\mathbb{P}_{q'} \left[\rho_x' = y \,\middle|\, \mathscr{A}(\Phi') \right] \right]
\end{aligned}$$

$$= \mathbb{E}_{q'}\left[\mu_{A'(x)}(y)\right] ,$$

where the last equality comes from Proposition 5. Hence, $\mathbb{P}_q$ a.s.,

$$\begin{aligned}
K_{q'}(x,y) &= \sum_{\bar{x}\in\bar{\mathscr{X}}} \mathbb{E}_{q'}\left[\mu_{A'(x)\cap A(\bar{x})}(y)\,\mu_{A'(x)}(A(\bar{x}))\right] \\
&= \sum_{\bar{x}\in\bar{\mathscr{X}}} \mathbb{E}_{q'}\left[\mu_{A'(x)\cap A(\bar{x})}(y)\mathbb{P}_{q'}\left[\rho'_x \in A(\bar{x}) \mid \mathscr{A}(\Phi')\right]\right] \\
&= \sum_{\bar{x}\in\bar{\mathscr{X}}} \mathbb{E}_{q'}\left[\mu_{A'(x)\cap A(\bar{x})}(y)\mathbb{1}_{A(\bar{x})}(\rho'_x)\right] \\
&= \mathbb{E}_{q'}\left[\mu_{A'(x)\cap A(\rho'_x)}(y)\right] .
\end{aligned}$$

□

Lemma 5. *For any $x \in \mathscr{X}$, set $\tilde{K}_{q'}(x,\cdot) = \mathbb{E}_{q'}\left[\mu_{A(\rho'_x)}(\cdot)\right]$. Then, $\mathbb{P}_q$ a.s.,*

$$\Lambda\tilde{K}_{q'} = \bar{P}\Lambda .$$

Proof. $\mathbb{P}_q$ a.s., for any $x,y \in \mathscr{X}$,

$$\begin{aligned}
\tilde{K}_{q'}(x,y) = \mathbb{E}_{q'}\left[\mu_{A(\rho'_x)}(y)\right] &= \sum_{\bar{y}\in\bar{\mathscr{X}}} \mu_{A(\bar{y})}(y)\mathbb{P}_{q'}\left[\rho'_x \in A(\bar{y})\right] \\
&= \sum_{\bar{y}\in\bar{\mathscr{X}}} \nu_{\bar{y}}(y)P_x\left[X(T_{q'}) \in A(\bar{y})\right] .
\end{aligned}$$

Hence, $\mathbb{P}_q$ a.s., for any $\bar{x} \in \bar{\mathscr{X}}$, and $y \in \mathscr{X}$,

$$\begin{aligned}
\nu_{\bar{x}}\tilde{K}_{q'}(y) &= \sum_{x\in\mathscr{X}}\sum_{\bar{y}\in\bar{\mathscr{X}}} \nu_{\bar{x}}(x)\nu_{\bar{y}}(y)P_x\left[X(T_{q'}) \in A(\bar{y})\right] \\
&= \sum_{\bar{y}\in\bar{\mathscr{X}}} \nu_{\bar{y}}(y)P_{\nu_{\bar{x}}}\left[X(T_{q'}) \in A(\bar{y})\right] \\
&= \bar{P}\Lambda(\bar{x},y) .
\end{aligned}$$

□

Therefore, $\mathbb{P}_q$ a.s., for any $\bar{x} \in \bar{\mathscr{X}}$,

$$\begin{aligned}
d_{TV}(\Lambda K_{q'}(\bar{x},\cdot),\bar{P}\Lambda(\bar{x},\cdot)) &= d_{TV}(\Lambda K_{q'}(\bar{x},\cdot),\Lambda\tilde{K}_{q'}(\bar{x},\cdot)) \\
&\le \sum_{x\in\mathscr{X}} \nu_{\bar{x}}(x)\, d_{TV}(K_{q'}(x,\cdot),\tilde{K}_{q'}(x,\cdot)) \\
&\le \sum_{x\in\mathscr{X}} \nu_{\bar{x}}(x)\mathbb{E}_{q'}\left[d_{TV}(\mu_{A'(x)\cap A(\rho'_x)},\mu_{A(\rho'_x)})\right] .
\end{aligned}$$

When B is a subset of C, one has $d_{TV}(\mu_B,\mu_C) = \mu_C(B^c)$. This yields

$$d_{TV}(\Lambda K_{q'}(\bar{x},\cdot),\bar{P}\Lambda(\bar{x},\cdot)) \le \sum_{x\in\mathscr{X}} \nu_{\bar{x}}(x)\mathbb{E}_{q'}\left[\mu_{A(\rho'_x)}(A'(x)^c)\right]$$

$$= \sum_{x\in\mathscr{X}} \nu_{\bar{x}}(x)\mathbb{E}_{q'}\left[\mathbb{P}_q\left[\rho_{\rho'_x}\notin A'(x)|\mathscr{A}(\Phi)\right]\right]$$

Note that

$$\sum_{\bar{x}\in\bar{\mathscr{X}}} \nu_{\bar{x}}(x) = \sum_{\bar{x}\in\bar{\mathscr{X}}} \frac{\mu(x)}{\mu(A(\bar{x}))}\mathbb{1}_{A(\bar{x})}(x) = \sum_{\bar{x}\in\bar{\mathscr{X}}} \frac{\mu(x)}{\mu(A(x))}\mathbb{1}_{A(\bar{x})}(x)$$
$$= \frac{\mu(x)}{\mu(A(x))}\sum_{\bar{x}\in\bar{\mathscr{X}}} \mathbb{1}_{A(\bar{x})}(x) = \frac{\mu(x)}{\mu(A(x))}\,.$$

Summing on $\bar{x}$ and integrating w.r.t. $\mathbb{E}_q$, leads to

$$\mathbb{E}_q\left[\sum_{\bar{x}\in\bar{\mathscr{X}}} d_{TV}(\Lambda K_{q'}(\bar{x},\cdot),\bar{P}\Lambda(\bar{x},\cdot))\right] \le \sum_{x\in\mathscr{X}} \mathbb{E}_{q,q'}\left[\mu_{A(x)}(x)\mathbb{1}_{A'(x)^c}(\rho_{\rho'_x})\right]\,.$$

Let $p\ge 1$ and p^* its conjugate exponent. Using Hölder's inequality, we get

$$\begin{aligned}
&\mathbb{E}_q\left[\sum_{\bar{x}\in\bar{\mathscr{X}}} d_{TV}(\Lambda K_{q'}(\bar{x},\cdot),\bar{P}\Lambda(\bar{x},\cdot))\right]\\
&\quad\le \left(\sum_{x\in\mathscr{X}} \mathbb{E}_{q,q'}\left[\mu_{A(x)}(x)^p\right]\right)^{1/p}\left(\sum_{x\in\mathscr{X}} \mathbb{P}_{q,q'}\left[\rho_{\rho'_x}\notin A'(x)\right]\right)^{1/p^*}\\
&\quad\le \left(\sum_{x\in\mathscr{X}} \mathbb{E}_q\left[\mu_{A(x)}(x)\right]\right)^{1/p}\left(\sum_{x\in\mathscr{X}} \mathbb{P}_{q,q'}\left[\rho_{\rho'_x}\notin A'(x)\right]\right)^{1/p^*}\,.
\end{aligned}$$

Note that

$$\sum_{x\in\mathscr{X}} \mathbb{E}_q\left[\mu_{A(x)}(x)\right] = \sum_{x\in\mathscr{X}} \mathbb{P}_q\left[\rho_x = x\right] = \sum_{x\in\mathscr{X}} \mathbb{P}_q\left[x\in\rho(\Phi)\right] = \mathbb{E}_q\left[|\rho(\Phi)|\right].$$

Therefore,

$$\begin{aligned}
&\mathbb{E}_q\left[\sum_{\bar{x}\in\bar{\mathscr{X}}} d_{TV}(\Lambda K_{q'}(\bar{x},\cdot),\bar{P}\Lambda(\bar{x},\cdot))\right]\\
&\quad\le \left(\mathbb{E}_q\left[|\rho(\Phi)|\right]\right)^{1/p}\left(\sum_{x\in\mathscr{X}} \mathbb{P}_{q,q'}\left[\rho_{\rho'_x}\notin A'(x)\right]\right)^{1/p^*}\,. \qquad (38)
\end{aligned}$$

To conclude the proof of our theorem we evaluate $\mathbb{P}_{q,q'}\left[\rho_{\rho'_x}\notin A'(x)\right]$ for x any given point in $\mathscr{X}$.

Lemma 6. *For any $x\in\mathscr{X}$, let Γ'_x be the path going from x to ρ'_x in Φ'. Then,*

$$\mathbb{P}_{q,q'}\left[\rho_{\rho'_x}\notin A'(x)\right] \le \frac{q'}{q}\mathbb{E}_{q'}\left[\left|\Gamma'_x\right|\right]\,.$$

Proof. To decide whether $\rho_{\rho'_x}$ is in $A'(x)$ or not, we do the following construction:

1. We begin the construction of Φ' using Wilson's algorithm starting from x. Thus, we let evolve the Markov process starting from x until an exponential time of parameter q', and erase the loop. The result is an oriented path γ' $(=\Gamma'_x)$ without loops from x to a point y $(=\rho'_x)$.
2. We go on with the construction of Φ with Wilson's algorithm starting from y. We let evolve the Markov process starting from y until an exponential time T_q of parameter q. The Markov process stops at a point v $(=\rho_{\rho'_x})$.
3. Finally, we continue the construction of Φ' using Wilson's algorithm starting from v. We let evolve the Markov process starting from v, and we stop it after an exponential time $T_{q'}$ of parameter q', or when it reaches the already constructed path γ'. At this point, we are able to decide whether $\rho_{\rho'_x}$ is in $A'(x)$ or not, since $\rho_{\rho'_x} \in A'(x)$ if and only if $T_{q'}$ is bigger than the hitting time of γ'.

Using this construction, we get that for any self-avoiding path γ' from x to y,

$$\mathbb{P}_{q,q'}\left[\rho_{\rho'_x} \notin A'(x) | \Gamma'_x = \gamma'; \rho'_x = y\right] = P_y\left[T_{q'} < H_{\gamma'} \circ \theta_{T_q}\right],$$

where θ_t denotes the time shift. Recall that σ_1 is the first time of the clock process on which X is build from $\hat{X}$, and let S_i be the successive return times to γ':

$$S_0 = 0\,,\ S_1 = \inf\left\{t \geq \sigma_1; X(t) \in \gamma'\right\} = H^+_{\gamma'}\,,\ S_{i+1} = S_i + S_1 \circ \theta_{S_i}\,.$$

Then,

$$P_y\left[T_{q'} < H_{\gamma'} \circ \theta_{T_q}\right] = \sum_{i=0}^{\infty} P_y\left[S_i \leq T_q < S_{i+1}; T_{q'} < H_{\gamma'} \circ \theta_{T_q}\right].$$

Now, if $S_i \leq T_q < S_i + \sigma_1 \circ \theta_{S_i}$, $X(T_q) \in \gamma'$ and $H_{\gamma'} \circ \theta_{T_q} = 0 < T_{q'}$. If $T_q \geq S_i + \sigma_1 \circ \theta_{S_i}$ and $T_q < S_{i+1}$, $X(T_q) \notin \gamma'$ and $H_{\gamma'} \circ \theta_{T_q} = S_{i+1} - T_q$. Therefore,

$$\begin{aligned} P_y\left[T_{q'} < H_{\gamma'} \circ \theta_{T_q}\right] &= \sum_{i=0}^{\infty} P_y\left[S_i + \sigma_1 \circ \theta_{S_i} \leq T_q < T_{q'} + T_q < S_{i+1}\right] \\ &= \sum_{i=0}^{\infty} \sum_{z \in \gamma'} P_y\left[S_i \leq T_q; X(S_i) = z\right] P_z\left[\sigma_1 \leq T_q < T_{q'} + T_q < H^+_{\gamma'}\right], \end{aligned}$$

using Markov property at time S_i. Set $\tilde{G}_q(y,z,\gamma') = E_y\left[\sum_{i=0}^{+\infty} \mathbb{1}_{S_i \leq T_q; X(S_i)=z}\right]$. Since $z \in \gamma'$, $\tilde{G}_q(y,z,\gamma')$ is the mean number of visits to the point z up to time T_q. We have obtained that

$$P_y\left[T_{q'} < H_{\gamma'} \circ \theta_{T_q}\right] = \sum_{z \in \gamma'} \tilde{G}_q(y,z,\gamma') P_z\left[\sigma_1 \leq T_q < T_q + T_{q'} < H^+_{\gamma'}\right].$$

We now use Markov property at time σ_1 to write

$$P_z\left[\sigma_1 \leq T_q < T_q + T_{q'} < H^+_{\gamma'}\right]$$

$$
\begin{aligned}
&= \sum_{u\notin\gamma'} P_z\left[\sigma_1 \le T_q, X(\sigma_1)=u\right] P_u(T_q < T_{q'} + T_q < H_{\gamma'}) \\
&\le \sum_{u\notin\gamma'} \frac{\alpha}{q+\alpha} P(z,u) P_u(T_{q'} < H_{\gamma'}) \\
&= \sum_{u\notin\gamma'} \frac{1}{q+\alpha} w(z,u) \mathbb{P}_{q'}\left[\rho'_u \neq y \mid \Gamma'_x = \gamma'\right] ,
\end{aligned}
$$

using that $\alpha P(z,u) = \mathscr{L}(z,u) = w(z,u)$ for $z \neq u$. Integrating over γ' and y, we are led to

$$
\mathbb{P}_{q,q'}\left[\rho_{\rho'_x} \notin A'(x)\right] \le \sum_{y\in\mathscr{X}} \sum_{\gamma: x\rightsquigarrow y} \sum_{z\in\gamma} \sum_{u\notin\gamma} \frac{\tilde{G}_q(y,z,\gamma)}{q+\alpha} w(z,u) \mathbb{P}_{q'}\left[\rho'_u \neq y; \Gamma'_x = \gamma; \rho'_x = y\right]
$$

where the sum over γ' is the sum on all self-avoiding paths going from x to y. Now, introducing for any such path γ

$$
\mathscr{F}_1(y,\gamma,u) := \{\phi \; s.o.f.\,; y \in \rho(\phi), \gamma \subset \phi, \rho_u \neq y\} ,
$$

this can be rewritten, with $w(\phi) = \prod_{e\in\phi} w(e)$, as

$$
\mathbb{P}_{q,q'}\left[\rho_{\rho'_x} \notin A'(x)\right] = \sum_{y\in\mathscr{X}} \sum_{\gamma: x\rightsquigarrow y} \sum_{z\in\gamma} \sum_{u\notin\gamma} \sum_{\phi\in\mathscr{F}_1(y,\gamma,u)} \frac{\tilde{G}_q(y,z,\gamma)}{q+\alpha} w(z,u) \frac{(q')^{|\rho(\phi)|} w(\phi)}{Z(q')} .
$$

Lemma 7. *Let $G_q(y,z) = E_y\left[\int_0^{T_q} \mathbb{1}_{X(s)=z} ds\right]$. Then $G_q(y,z) = \tilde{G}_q(y,z,\gamma)/(q+\alpha)$ for any self-avoiding path γ that contains z and goes from x to y.*

Proof. Let V_i be the successive return times to z:

$$
V_0 = 0 \,,\; V_1 = \inf\{t \ge \sigma_1; X(t) = z\} \;,\; V_{i+1} = V_i + V_1 \circ \theta_{V_i} .
$$

Then $\tilde{G}_q(y,z,\gamma) = \delta_y(z) + \sum_{i=1}^{+\infty} E_y[\mathbb{1}_{V_i \le T_q}]$. Moreover, using Markov's property at time V_i,

$$
\begin{aligned}
G_q(y,z) &= \sum_{i=0}^{\infty} E_y\left[\int_{V_i}^{V_{i+1}} \mathbb{1}_{T_q \ge s} \mathbb{1}_{X(s)=z} ds\right] \\
&= \sum_{i=0}^{\infty} E_y\left[\mathbb{1}_{V_i \le T_q} E_{X(V_i)}\left[\int_0^{V_1} \mathbb{1}_{T_q \ge s} \mathbb{1}_{X(s)=z} ds\right]\right] \\
&= E_y\left[\int_0^{V_1} \mathbb{1}_{T_q \ge s} \mathbb{1}_{X(s)=z} ds\right] + \sum_{i=1}^{\infty} E_y\left[\mathbb{1}_{V_i \le T_q}\right] E_z\left[\int_0^{V_1} \mathbb{1}_{T_q \ge s} \mathbb{1}_{X(s)=z} ds\right] \\
&= \left(\delta_y(z) + \sum_{i=1}^{\infty} E_y\left[\mathbb{1}_{V_i \le T_q}\right]\right) E_z\left[\int_0^{V_1} \mathbb{1}_{T_q \ge s} \mathbb{1}_{X(s)=z} ds\right] \\
&= \tilde{G}_q(y,z,\gamma) E_z\left[\int_0^{V_1} \mathbb{1}_{T_q \ge s} \mathbb{1}_{X(s)=z} ds\right] .
\end{aligned}
$$

Now, $E_z\left[\int_0^{V_1} \mathbb{1}_{T_q\geq s}\mathbb{1}_{X(s)=z}\,ds\right] = E_z\left[\int_0^{\sigma_1} \mathbb{1}_{T_q\geq s}\,ds\right] = E\left[\sigma_1\wedge T_q\right] = \frac{1}{q+\alpha}$. $\square$

Hence,

$$\mathbb{P}_{q,q'}\left[\rho_{\rho'_x}\notin A'(x)\right] \leq \sum_{y\in\mathscr{X}}\sum_{\gamma:x\rightsquigarrow y}\sum_{z\in\gamma}\sum_{u\notin\gamma}\sum_{\phi\in\mathscr{F}_1(y,\gamma,u)} G_q(y,z)w(z,u)\frac{(q')^{|\rho(\phi)|}w(\phi)}{Z(q')}\,.$$

We fix y, γ and z and want to perform the summations over u and ϕ. With any pair (u,ϕ), with $u\notin\gamma$ and $\phi\in\mathscr{F}_1(y,\gamma,u)$, we associate a new forest $\tilde{\phi}=\tilde{\phi}(u,\phi)$ in the following way:

1. we reverse the edges from z to y along γ;
2. we add the edge (z,u).

The forest $\tilde{\phi}$ is such that:

- $\left|\rho(\tilde{\phi})\right| = |\rho(\phi)|-1$;
- $z\notin\rho(\tilde{\phi})$.
- the piece $\gamma_{x\rightsquigarrow z}$ of the path γ going from x to z belongs to $\tilde{\phi}$;
- the path $\overleftarrow{\gamma}_{y\rightsquigarrow z}$ consisting of the reversed path γ from z to y, belongs to $\tilde{\phi}$.

Using reversibility, one has $\mu(z)\prod_{e\in\gamma_{z\rightsquigarrow y}} w(e) = \mu(y)\prod_{e\in\overleftarrow{\gamma}_{y\rightsquigarrow z}} w(e)$, and

$$w(z,u)w(\phi) = w(\tilde{\phi})\mu(y)/\mu(z)\,.$$

Set $\mathscr{F}_2(y,z,\gamma) = \left\{\phi\ s.o.f.\,;\, z\notin\rho(\phi), \gamma_{x\rightsquigarrow z}\subset\phi, \overleftarrow{\gamma}_{y\rightsquigarrow z}\subset\phi\right\}$. Note that the function

$$(u,\phi)\in\{(u,\phi), u\notin\gamma, \phi\in\mathscr{F}_1(y,\gamma,u)\}\mapsto\tilde{\phi}\in\mathscr{F}_2(y,z,\gamma)$$

is one to one. Indeed, given $\tilde{\phi}$ in $\mathscr{F}_2(y,z,\gamma)$, u is the "ancestor" of z in $\tilde{\phi}$, and once we know u, ϕ is obtained by cutting the edge (z,u), and by reversing the path $\overleftarrow{\gamma}_{y\rightsquigarrow z}$. Therefore, we obtain

$$\begin{aligned}&\sum_{u\notin\gamma}\sum_{\phi\in\mathscr{F}_1(y,\gamma,u)} G_q(y,z)w(z,u)\frac{(q')^{|\rho(\phi)|}w(\phi)}{Z(q')}\\ &\quad= \sum_{\phi\in\mathscr{F}_2(y,z,\gamma)} G_q(y,z)\frac{\mu(y)}{\mu(z)}\frac{(q')^{|\rho(\phi)|+1}w(\phi)}{Z(q')} = \sum_{\phi\in\mathscr{F}_2(y,z,\gamma)} G_q(z,y)\frac{(q')^{|\rho(\phi)|+1}w(\phi)}{Z(q')}\end{aligned}$$

by reversibility. At this point, we are led to

$$\mathbb{P}_{q,q'}\left[\rho_{\rho'_x}\notin A'(x)\right] \leq \sum_{y\in\mathscr{X}}\sum_{\gamma:x\rightsquigarrow y}\sum_{z\in\gamma}\sum_{\phi\in\mathscr{F}_2(y,z,\gamma)} G_q(z,y)\frac{(q')^{|\rho(\phi)|+1}w(\phi)}{Z(q')}\,.$$

We now perform the summations over z and γ and ϕ, y being fixed. Note that if $\phi\in\mathscr{F}_2(y,z,\gamma)$ for some z and γ, x and y are in the same tree ($\tau_x=\tau_y$ using the notations of Section 1.3.3), and z is their first common ancestor $a(x,y)$ in that tree. Let us then denote

$$\mathscr{F}_3(y,x) = Ł\left\{\phi\, s.o.f.\,;\, \tau_x = \tau_y, a(x,y) \notin \rho(\phi)\right\}.$$

Then,

$$\cup_{\gamma: x\rightsquigarrow y} \cup_{z\in\gamma} \mathscr{F}_2(y,z,\gamma) \subset \mathscr{F}_3(y,x).$$

In addition, given a forest $\phi \in \mathscr{F}_3(y,x)$, there is a unique $\gamma : x \rightsquigarrow y$, and $z \in \gamma$ such that $\phi \in \mathscr{F}_2(y,z,\gamma)$: z is the first common ancestor $a(x,y)$ of x and y, whereas γ is the concatenation of the path going from x to $a(x,y)$ and the reversed path from y to $a(x,y)$. Therefore,

$$\begin{aligned}\sum_{\gamma: x\rightsquigarrow y} \sum_{z\in\gamma} \sum_{\phi\in\mathscr{F}_2(y,z,\gamma)} & G_q(z,y)\frac{(q')^{|\rho(\phi)|+1}w(\phi)}{Z(q')} \\ &= \sum_{\phi\in\mathscr{F}_3(y,x)} G_q(a(x,y),y)\frac{(q')^{|\rho(\phi)|+1}w(\phi)}{Z(q')}.\end{aligned}$$

It remains to sum over y. When moving y in τ_x, $a(x,y)$ moves along the path γ_x going from x to the root of τ_x. Hence,

$$\begin{aligned}&\sum_{y\in\mathscr{X}} \sum_{\phi\in\mathscr{F}_3(y,x)} G_q(a(x,y),y)\frac{(q')^{|\rho(\phi)|+1}w(\phi)}{Z(q')} \\ &= \sum_{\phi\, s.o.f.} \sum_{z\in\gamma_x, z\neq\rho_x} \sum_{y\in\tau_x; a(x,y)=z} G_q(z,y)\frac{(q')^{|\rho(\phi)|+1}w(\phi)}{Z(q')} \\ &\leq \frac{q'}{q} \sum_{\phi\, s.o.f.} \sum_{z\in\gamma_x, z\neq\rho_x} \pi_{q'}(\phi) \leq \frac{q'}{q}\mathbb{E}_{q'}\left[\left|\Gamma'_x\right|\right].\end{aligned}$$

□

5 Proof of Theorem 3

Let us first rewrite our approximate solutions of Equation (3) with error terms. There are signed measures $\varepsilon_{\bar{x},q'}$ such that, for all $\bar{x}$ in $\bar{\mathscr{X}}$,

$$\mu_{A(\bar{x})}K_{q'} = \sum_{\bar{y}\in\bar{\mathscr{X}}} P_{\mu_{A(\bar{x})}}\big(X(T_{q'}) \in A(\bar{y})\big)\,\mu_{A(\bar{y})} + \varepsilon_{\bar{x},q'}.$$

Let us now apply the "'low-pass filter" MW_m on both sides of the equations. On the one hand, $K_{q'}$ and MW_m commute. On the other hand, our linear independence (i.e. finite squeezing) hypothesis implies that the $\varepsilon_{\bar{x},q'}MW_m$ are linear combinations of the $\mu_{A(\bar{x})}MW_m$. Indeed, since the image $\mathrm{im}(MW_m)$ of MW_m is a vector space of dimension m that contains the m linearly independent $\nu_{\bar{x}}$, the latter should span $\mathrm{im}(MW_m)$. We then get, by using the notation of the proof of Proposition 6,

$$\nu_{\bar{x}} K_{q'} = \sum_{\bar{y}\in\mathscr{X}} \left(P_{\mu_{A(\bar{x})}}\big(X(T_{q'}) \in A(\bar{y})\big) + \langle \tilde{\nu}_{\bar{y}}, \varepsilon_{\bar{x},q'} MW_m \rangle^* \right) \nu_{\bar{y}}.$$

Now, when q' goes to 0, $P_{\mu_{A(\bar{x})}}\big(X(T_{q'}) \in A(\bar{y})\big)$ converges to $\mu(A(\bar{y})) > 0$, and, by Theorem 2, $\varepsilon_{\bar{x},q'}$ goes to zero. Since our $\nu_{\bar{x}}$ do not depend on q', this concludes the proof of the theorem.

Let us list what would be needed to give quantitative bounds on q' to ensure that we can build in this way exact solutions of (3). We would need:

1. upper bounds on the $\varepsilon_{\bar{x},q'}$;
2. upper bounds on the $\|\tilde{\nu}_{\bar{x}}\|$;
3. lower bounds on the $P_{\mu_{A(\bar{x})}}\big(X(T_{q'}) \in A(\bar{y})\big)$.

The latter are out of reach in such a general framework, the first ones are provided by Theorem 2, the second ones would be a consequence of upper bounds on the squeezing. This is the reason why we introduce the squeezing to measure joint overlap. We note that given Proposition 5 and Equation (27) in Theorem 1, we are not so far of getting such bounds. But no convexity inequality leads here to the conclusion.

Appendix: Proof of proposition 1

If such random variables exist then, for all $\bar{x}$, $\bar{y} \neq \bar{x}$ and y,

$$\begin{aligned} P_{\nu_{\bar{x}}}\left(T_{\bar{x}} = 1, \bar{Y}_{\bar{x}} = \bar{y} \,\middle|\, \hat{X}(1) = y\right) &= \frac{P_{\nu_{\bar{x}}}\left(T_{\bar{x}} = 1, \bar{Y}_{\bar{x}} = \bar{y}, \hat{X}(1) = y\right)}{(\nu_{\bar{x}} P)(y)} \\ &= \frac{(1-\bar{P}(\bar{x},\bar{x}))\frac{\bar{P}(\bar{x},\bar{y})}{1-\bar{P}(\bar{x},\bar{x})}\nu_{\bar{y}}(y)}{(\nu_{\bar{x}} P)(y)} \\ &= \frac{\bar{P}(\bar{x},\bar{y})\nu_{\bar{y}}(y)}{(\nu_{\bar{x}} P)(y)}. \end{aligned}$$

By summing on $\bar{y}$ we get

$$P_{\nu_{\bar{x}}}\left(T_{\bar{x}} = 1 \,\middle|\, \hat{X}(1) = y\right) = \frac{(\nu_{\bar{x}} P)(y) - \bar{P}(\bar{x},\bar{x})\nu_{\bar{x}}(y)}{(\nu_{\bar{x}} P)(y)} = 1 - \frac{\bar{P}(\bar{x},\bar{x})\nu_{\bar{x}}(y)}{(\nu_{\bar{x}} P)(y)}.$$

We also have

$$\begin{aligned} P_{\nu_{\bar{x}}}\left(\bar{Y}_{\bar{x}} = \bar{y} \,\middle|\, \hat{X}(1) = y, T_{\bar{x}} = 1\right) &= \frac{P_{\nu_{\bar{x}}}\left(\bar{Y}_{\bar{x}} = \bar{y}, T_{\bar{x}} = 1 \,\middle|\, \hat{X}(1) = y\right)}{P_{\nu_{\bar{x}}}\left(T_{\bar{x}} = 1 \,\middle|\, \hat{X}(1) = y\right)} \\ &= \frac{\bar{P}(\bar{x},\bar{y})\nu_{\bar{y}}(y)}{(\nu_{\bar{x}} P)(y) - \bar{P}(\bar{x},\bar{x})\nu_{\bar{x}}(y)}. \end{aligned}$$

We are then led to build $T_{\bar{x}} \geq 1$ and $\bar{Y}_{\bar{x}}$ in the following way.

1. At $t=1$ we set $T_{\bar{x}}=1$ with probability $1-\bar{P}(\bar{x},\bar{x})\nu_{\bar{x}}(\hat{X}(1))/(\nu_{\bar{x}}P)(\hat{X}(1))$ by using a uniform random variable U_1 which is independent of $\hat{X}$ —it holds

$$\bar{P}(\bar{x},\bar{y})\nu_{\bar{x}}(y)/(\nu_{\bar{x}}P)(y)\leq 1$$

   for all y in $\mathscr{X}$, as a consequence of Equation (6).
2. If we just set $T_{\bar{x}}=1$ we then set $\bar{Y}_{\bar{x}}=\bar{y}\neq\bar{x}$ with a probability given by the ratio $\bar{P}(\bar{x},\bar{y})\nu_{\bar{y}}(\hat{X}(1))/[(\nu_{\bar{x}}P)(\hat{X}(1))-\bar{P}(\bar{x},\bar{x})\nu_{\bar{x}}(\hat{X}(1))]$ by using a uniform random variable U_1' that is independent of U_1 and $\hat{X}$. (Once again (6) ensures that these are positive quantities summing to one.)
3. If for all $s<t$ we did not decide to set $T_{\bar{x}}=s$ then we set in the same way $T_{\bar{x}}=t$ with probability $1-\bar{P}(\bar{x},\bar{x})\nu_{\bar{x}}(\hat{X}(t))/(\nu_{\bar{x}}P)(\hat{X}(t))$, in which case we set $\bar{Y}_{\bar{x}}=\bar{y}\neq\bar{x}$ with probability $\bar{P}(\bar{x},\bar{y})\nu_{\bar{y}}(\hat{X}(t))/[(\nu_{\bar{x}}P)(\hat{X}(t))-\bar{P}(\bar{x},\bar{x})\nu_{\bar{x}}(\hat{X}(t))]$. This is naturally done by using uniform random variable that are independent of $\hat{X}$ and $U_1, U_1', U_2, U_2', \ldots, U_{t-1}, U_{t-1}'$.

At this point, the key property to check is the stationarity of $\nu_{\bar{x}}$ up to $T_{\bar{x}}$. To this end it suffices to check Equation (7) with $t=1$. And one has

$$\begin{aligned}
P_{\nu_{\bar{x}}}\left(\hat{X}(1)=y\,\middle|\,T_{\bar{x}}>1\right) &= \frac{P_{\nu_{\bar{x}}}\left(\hat{X}(1)=y,T_{\bar{x}}>1\right)}{P_{\nu_{\bar{x}}}\left(T_{\bar{x}}>1\right)}\\
&= \frac{P_{\nu_{\bar{x}}}\left(\hat{X}(1)=y\right)-P_{\nu_{\bar{x}}}\left(\hat{X}(1)=y,T_{\bar{x}}=1\right)}{1-P_{\nu_{\bar{x}}}\left(T_{\bar{x}}=1\right)}\\
&= \frac{\nu_{\bar{x}}P(y)-\nu_{\bar{x}}P(y)\left(1-\frac{\bar{P}(\bar{x},\bar{x})\nu_{\bar{x}}(y)}{\nu_{\bar{x}}P(y)}\right)}{1-\sum_z\nu_{\bar{x}}P(z)\left(1-\frac{\bar{P}(\bar{x},\bar{x})\nu_{\bar{x}}(z)}{\nu_{\bar{x}}P(z)}\right)}\\
&= \frac{\bar{P}(\bar{x},\bar{x})\nu_{\bar{x}}(y)}{1-\sum_z\nu_{\bar{x}}P(z)+\sum_z\bar{P}(\bar{x},\bar{x})\nu_{\bar{x}}(z)}=\nu_{\bar{x}}(y).
\end{aligned}$$

Points (1)–(5) immediately follow.

References

1. Andersen, R., Gharan, S. O., Peres, Y., Trevisan, L.: Almost optimal local graph clustering using evolving sets, Journal of the ACM **63**, 1–31 (2016)
2. Avena, L., Castell, F., Gaudillière, A., Mélot, C.: Intertwining wavelets or multiresolution analysis through random forests on graphs, Appl. Comput. Harmon. Anal. **48**, 949–992 (2020)
3. Avena, L., Gaudillière, A.: Two Applications of Random Spanning Forests, J. Theor. Probab. **31**, 1975–2004 (2018)
4. Bovier, A., den Hollander, F.: Metastability: a potential-theoretic approach. Springer, Cham (2015)
5. Carmona, P., Petit, F., Yor, M.: Beta-gamma random variables and intertwining relations between certain Markov processes, Rev. Mat. Iberoamericana **14**, 311–367 (1998)

6. Diaconis, P., Fill, J. A.: Strong stationary times via a new form of duality, Ann. Probab. **18**, 1483–1522 (1990)
7. Donati-Martin, C., Doumerc, Y., Matsumoto, H., Yor, M.: Some properties of the Wishart processes and a matrix extension of the Hartman-Watson laws, Publ. Res. Inst. Math. Sci. **40**, 1385–1412 (2004)
8. Freidlin, M. I., Wentzell, A. D.: Random pertubations of dynamical systems. Springer-Verlag, Berlin, Heidelberg (1984).
9. Hammond, D. K., Vandergheynst, P., Gribonval, R.: Wavelets on graphs via spectral graph theory: Applied and Computational Harmonic Analysis **30**, 129–150 (2009)
10. Kirchhoff, G.: Über die Auflösung der Gleichungen, auf welche man bei der Untersuchung der lineraren Verteilung galvanisher Ströme geführt wird, Ann. Phys. chem **72**, 497–508 (1847)
11. Matsumoto, H., Yor, M.: An analogue of Pitman's $2M - X$ theorem for exponential Wiener functionals, I. A time-inversion approach, Nagoya Math. J. **159**, 125–166 (2000)
12. Micchelli, C. A., Willoughby, R. A.: On functions which preserve the class of Stieltjes matrices, Lin. Alg. Appl. **23**, 141–156 (1979)
13. Miclo, L.: On absorption times and Dirichlet eigenvalues, ESAIM: Probability and Statistics **14**, 117–150 (2010)
14. Miclo, L.: On the construction of set-valued dual processes, Electron. J. Probab. **25**, 1–64 (2020)
15. Miclo, L., Patie, P.: On intertwining relations between Ehrenfest, Yule and Ornstein-Uhlenbeck processes, ArXiv:1904.10693 (2019).
16. Morris, B., Peres, Y.: Evolving sets, mixing and heat kernel bounds, Probab. Theory Relat. Fields **133**, 245–266 (2005)
17. Olivieri, E., Vares, M. E.: Large deviations and metastability. Cambridge University Press (2005)
18. Propp, J., Wilson, D.: How to get a perfectly random sample from a generic Markov chain and generate a random spanning tree of a directed graph, Journal of Algorithms **27**, 179–217 (1998)
19. Ricaud, B., Borgnat, P., Tremblay, N., Goncalves, P., Vandergheynst, P.: Fourier could be a data scientist: from graph Fourier transform to signal processing on graphs, C. R. Physique **20**, 474–488 (2019)
20. Rogers, L. C. G., Pitman, J. W.: Markov functions, Ann. Probab. **9**, 573–582 (1981)
21. Scoppola, E.: Renormalization group for Markov chains and application to metastability, Journal of Statistical Physics **73**, 83–121 (1993)
22. Shuman, D. I., Narang, S. K., Frossard, P., Ortega, A., Vandergheynst, P.: The Emerging Field of Signal Processing on Graphs: Extending High-Dimensional Data Analysis to Networks and Other Irregular Domains, IEEE Signal Processing Magazine **30**, 83–98 (2013)
23. Vetterli, M., Kovacevic, J., Goyal, V. K.: Foundations of Signal Processing. Cambridge University Press (2014)
24. Warren, J.: Dyson's Brownian motions, intertwining and interlacing, Electron. J. Probab. **12**, 573–590 (2007)
25. Wilson, D.: Generating random spanning trees more quickly than the cover time, Proceedings of the twenty-eighth annual ACM symposium on the theory of computing, 296–303 (1996)