

# Large deviations for correlated random variables described by a matrix product ansatz

Florian Angeletti,<sup>1,2</sup> Hugo Touchette,<sup>1,2</sup> Eric Bertin,<sup>3,4</sup> and Patrice Abry<sup>3</sup>

<sup>1</sup> National Institute for Theoretical Physics (NITheP), Stellenbosch 7600, South Africa

<sup>2</sup> Institute of Theoretical Physics, University of Stellenbosch, Stellenbosch 7600, South Africa

<sup>3</sup> Laboratoire de physique, UMR CNRS 5672, École Normale Supérieure de Lyon, 46 Allée d'Italie, Lyon 69007, France

<sup>4</sup> Laboratoire Interdisciplinaire de Physique, Université Joseph Fourier Grenoble, CNRS UMR 5588, BP 87, F-38402 Saint-Martin d'Hères, France

E-mail: [florianangeletti@sun.ac.za](mailto:florianangeletti@sun.ac.za), [htouchette@sun.ac.za](mailto:htouchette@sun.ac.za), [eric.bertin@ujf-grenoble.fr](mailto:eric.bertin@ujf-grenoble.fr), [patrice.abry@ens-lyon.fr](mailto:patrice.abry@ens-lyon.fr)

**Abstract.** We study the large deviations of sums of correlated random variables described by a matrix product ansatz, which generalizes the product structure of independent random variables to matrices whose non-commutativity is the source of correlations. We show with specific examples that different large deviation behaviors can be found with this ansatz. In particular, it is possible to construct sums of correlated random variables that violate the Law of Large Numbers, the Central Limit Theorem, as well as sums that have nonconvex rate functions or rate functions with linear parts or plateaux.

## 1. Introduction

The study of stationary states and fluctuations of nonequilibrium systems using concepts and methods from large deviation theory has become an active topic in statistical mechanics [1–4], following the successful application of this theory to equilibrium systems [5–8]. For both types of systems, it is known that the calculation of rate functions, the central object of large deviation theory characterizing the likelihood of fluctuations, is in general equivalent to, and therefore as difficult as, calculating partition functions [3]. From this point of view, the simplest systems for which large deviations can be obtained are systems of independent particles, which correspond in probabilistic terms to independent and identically distributed (i.i.d) random variables. Next come linear chains of particles interacting via first-neighbor potentials, which can be mapped to Markov chains. In this case, results such as the Gärtner-Ellis Theorem can be used to obtain rate functions by solving an eigenvalue problem, which is essentially a transfer operator problem [3–5].

Large deviations have been obtained for other models of correlated systems: e.g., random and Gibbs fields [9–12], random matrices [13–17], random walks in random environments [18–20], hidden Markov processes [21–23], and processes that explicitly evolve in non-Markovian way [24]. However, as one goes beyond Markov processes

the applications of large deviation techniques becomes very difficult and often leads to non-explicit results. Not much, in particular, is known on the application of the Gärtner-Ellis Theorem for sums of strongly correlated random variables – a problem mirrored again in statistical mechanics in the difficulty of calculating partition functions for strongly correlated particles.

The goal of this contribution is to present a class of correlated random variables, defined via a matrix product ansatz, for which the large deviations of the sample mean can be obtained explicitly. This class of random variables was proposed recently in [25, 26] as a generalization of recent results on stationary distributions of interacting particle models, in particular, the asymmetric exclusion process, which has been actively studied recently [27–33]. The basis of this ansatz, presented in the next section, is to express the joint probability distribution of a sequence of random variables as a product of matrices whose non-commutativity is the source of correlations between these random variables.

Our goal here is to obtain large deviation results for the sample mean of these “matrix-correlated” random variables using the Gärtner-Ellis Theorem, and to show that the application of this theorem leads in this context to a simple product structure for generating functions, similar to that of i.i.d random variables, but involving matrices rather than scalar functions. The difference is crucial as it leads to large deviation behaviors that do not arise for i.i.d random variables and ergodic finite Markov chains. We will show, for example, that sample means having nonconvex rate functions are possible for matrix-correlated random variables, as are rate functions with linear parts or plateaux. These examples are related, as will be explained, to extensions of the Law of Large Numbers and Central Limit Theorem, recently studied in [34]. Here we validate and complement this study from the point of view of large deviation theory.

## 2. Definitions and model

We study the sum

$$S_n = \sum_{i=1}^n X_i \quad (1)$$

of  $n$  random variables  $X_1, \dots, X_n$  whose correlations are described by the joint probability density function (pdf)  $P(x_1, \dots, x_n)$ .<sup>‡</sup> The general model of correlation that we consider is defined by the following form for the joint pdf, referred to as the *matrix product ansatz*:

$$P(x_1, \dots, x_n) = \frac{1}{\mathcal{L}(\mathcal{E}^n)} \mathcal{L}(\mathcal{R}(x_1) \dots \mathcal{R}(x_n)) \quad (2)$$

where  $\mathcal{R}(x)$  is a positive  $d \times d$  matrix function,  $\mathcal{L}$  is a linear form defined by

$$\mathcal{L}(M) = \text{tr } \mathcal{A}^T M \quad (3)$$

with  $\mathcal{A}$  a positive  $d \times d$  matrix, and  $\mathcal{E}$  is the *structure* matrix defined by  $\mathcal{E} = \int \mathcal{R}(x) dx$  and such that  $\mathcal{L}(\mathcal{E}^n) \neq 0$ .

This ansatz was proposed and studied in [26, 34], following similar forms of joint pdfs appearing in the context of nonequilibrium particle models, such as the exclusion

<sup>‡</sup> We consider throughout real random variables, but discrete random variables are also possible.

process [31–33]. Similar matrix product state ansatz also appear in quantum many-body physics [35, 36]. The main property of this model is that the correlations in  $X_1, \dots, X_n$  are controlled by the structure matrix  $\mathcal{E}$ . To be more precise, let  $\nu_1, \dots, \nu_m$  with  $m \leq d$  denote the *distinct* eigenvalues of  $\mathcal{E}$ , ordered in descending order of their real part, and let  $J_{k,l}$  be the  $l$ -th Jordan block associated with the eigenvalue  $\nu_k$  in a Jordan basis  $B$  of  $\mathcal{E}$ :

$$\mathcal{E} = B \begin{pmatrix} J_{1,1} & & 0 \\ & \ddots & \\ 0 & & J_{m,r} \end{pmatrix} B^{-1}, \quad J_{k,l} = \begin{pmatrix} \nu_k & 1 & & 0 \\ & \ddots & \ddots & \\ & & \ddots & 1 \\ 0 & & & \nu_k \end{pmatrix}. \quad (4)$$

From the Perron-Frobenius Theorem for non-negative matrices, the dominant eigenvalue  $\nu_1$  of  $\mathcal{E}$  is positive real. The Jordan blocks  $J_{k,l}$  determine the type of correlation in  $X_1, \dots, X_n$  according to the following cases (which are not mutually exclusive) [26]:

- (i) If there is at least one block  $J_{k,l}$  with  $k \neq 1$  (i.e., there exists at least one non-dominant eigenvalue), then the sequence of random variables  $X_1, \dots, X_n$  exhibits exponential, *short-range* correlation, as is typically the case for ergodic Markov chains. An example of matrix  $\mathcal{E}$  that falls into this case are irreducible aperiodic matrices, such that, for some  $k > 0$ ,  $(\mathcal{E}^k)_{i,j} > 0$  for any  $i, j$ .
- (ii) If the dominant eigenvalue  $\nu_1$  has more than one Jordan blocks, then the sequence  $X_1, \dots, X_n$  exhibits (generically non-zero) constant correlations, as arises, for example, in non-ergodic Markov chains. The identity matrix  $\mathbb{I}$  provides the simplest subclass of matrices with this kind of Jordan normal form.
- (iii) If there is at least one block  $J_{1,l}$  associated with  $\nu_1$  with dimension greater than 1, then  $X_1, \dots, X_n$  exhibits polynomial long-range correlation in the sense that  $\mathbb{E}[X_k X_l] \approx Q(k/n, l/n)$  with  $Q(X, Y)$  a polynomial function. The simplest example of such matrices are the so-called linear irreversible subclass

$$\mathcal{E} = \mathbb{I} + U, \quad U = \begin{pmatrix} 0 & 1 & 0 \\ \vdots & \ddots & 1 \\ 0 & \dots & 0 \end{pmatrix}. \quad (5)$$

As shown in [26], these three cases can also be understood by noting that the product ansatz admits a representation in terms of hidden Markov chains, consisting of a hidden Markov chain layer  $\Gamma_1, \dots, \Gamma_{n+1} \in \{1, \dots, d\}$  determining the visible layer  $X_1, \dots, X_n$ . The transition matrix of the  $\Gamma$  Markov chain is obtained from the structure matrix  $\mathcal{E}$  according to

$$P(\Gamma_1 = i, \Gamma_{n+1} = f) = \mathcal{A}_{if} \frac{(\mathcal{E}^n)_{if}}{\mathcal{L}(\mathcal{E}^n)}, \quad (6)$$

$$P(\Gamma_{k+1} = j | \Gamma_k = i, \Gamma_{n+1} = f) = \mathcal{E}_{ij} \frac{(\mathcal{E}^{n-k})_{jf}}{(\mathcal{E}^{n-k+1})_{if}}. \quad (7)$$

This Markov chain is non-homogeneous and nonstandard, due to the dependence on the final state  $\Gamma_{n+1}$ . In particular for  $k = n$ , the transition rate  $P(\Gamma_{k+1} = j | \Gamma_k = i, \Gamma_{n+1} = f)$  equals 1 if  $j = f$  and 0 otherwise, meaning that the last step of  $\Gamma$  is

deterministic. For a given sequence  $\Gamma_1, \dots, \Gamma_{n+1}$ , the random variables  $X_1, \dots, X_n$  are then independent but non-identically distributed, with a pdf depending on  $\Gamma$ :

$$P(x_1, \dots, x_n | \Gamma) = \prod_{k=1}^n \mathcal{P}_{\Gamma_k \Gamma_{k+1}}(x_k), \quad (8)$$

where

$$\mathcal{P}_{i,j}(x) = \frac{\mathcal{R}_{i,j}(x)}{\mathcal{E}_{i,j}} \quad (9)$$

represents the conditional pdf associated with transitions from states  $i$  to  $j$  in the hidden layer. With this result, the exponentially-decaying and constant correlation cases, mentioned above, can be understood as arising from similar correlations at the hidden Markov chain  $\Gamma$  level, while the polynomial correlation case is more unusual and originates from the non-homogeneous nature of  $\Gamma$ . At this point, it is important to note that the states of the Markov chain  $\Gamma$  are not the values of the random variables  $X_1, \dots, X_n$ : the former has  $d$  discrete states, as seen from (6) and (7), while the  $X_i$ 's are again real random variables with joint pdf (8).

The hidden Markov chain representation of matrix-correlated random variables is useful for designing and synthesizing random variables with prescribed statistical properties, such as fixed marginal pdfs, correlation functions, or higher order dependencies [26]. It can also be used to derive analogues of the Law of Large Numbers and the Central Limit Theorem [34]. In this case, it has been found that non-standard limit laws appear for random variables with polynomial or constant correlation, illustrating the difference between these two kinds of long-range correlation and the exponentially-decaying kind, for which standard laws typically apply.

Here, we are interested in extending these results by studying the large fluctuations of the sum  $S_n$  that are of order  $O(n)$  with respect to  $S_n$  or, equivalently, order  $O(1)$  with respect to the sample mean  $S_n/n$ . These fluctuations are the focus of the theory of large deviations [3–8] and are known to be characterized by the following exponential pdf:

$$P(S_n/n = s) = e^{-nI(s)+o(n)} \quad (10)$$

where  $o(n)$  denotes corrections growing slower than linearly in  $n$ . We say that  $S_n/n$  satisfies a *large deviation principle* (LPD) if its pdf has the form above or, equivalently, if the following limit exists:

$$I(s) = \lim_{n \rightarrow \infty} -\frac{1}{n} \ln P(S_n/n = s). \quad (11)$$

The function  $I(s)$  defined by this limit is called the *rate function*; it governs according to (10) the rate at which  $P(S_n/n)$  decays to 0 when  $n \rightarrow \infty$  and so the rate at which this pdf concentrates exponentially with  $n$  around the typical values of  $S_n/n$  corresponding to the zeros of  $I(s)$ . This will be studied in more detail below.

The main result that we will use to study the large deviations of  $S_n/n$  is the Gärtner-Ellis Theorem [5, 37], which enables one to obtain the rate function  $I(s)$  from the so-called *scaled cumulant generating function* (SCGF) defined as

$$\lambda(w) = \lim_{n \rightarrow \infty} \frac{1}{n} \ln g_n(w), \quad (12)$$

where

$$g_n(w) = \mathbb{E} [e^{wS_n}] = \int_{\mathbb{R}} dx_1 \cdots \int_{\mathbb{R}} dx_n P(x_1, \dots, x_n) e^{w \sum_{i=1}^n x_i} \quad (13)$$

is the generating function of  $S_n$ . The Gärtner-Ellis Theorem states in simplified form that, if  $\lambda(w)$  exists and is differentiable everywhere, then  $S_n/n$  satisfies an LDP with rate function  $I(s)$  given by the Legendre-Fenchel transform of  $\lambda(w)$ :

$$I(s) = \lambda^*(s) = \sup_w \{ws - \lambda(w)\}. \quad (14)$$

In the next section, we will see that the differentiability property of  $\lambda(w)$  is not always satisfied, a sign that  $I(s)$  is either nonconvex or has linear parts [3]. Interestingly, both cases can arise for sums of matrix-correlated random variables, as will be shown with explicit examples in Sec. 4, and must be treated with a local version of the Gärtner-Ellis Theorem or other results [3], such as the complex integral method described in [38]. The reason why this theorem does not apply for these cases has to do essentially with the fact that a nonconvex  $I(s)$  and its convex hull have the same  $\lambda^*(s)$ , which is nondifferentiable somewhere; see Sec. 4.4 of [3] for more details.

### 3. Large deviation results

The Gärtner-Ellis Theorem takes a simple form for i.i.d random variables due to the fact that the generating function  $g_n(w)$  factorizes into a product of identical marginal generating functions,  $g(w) = \mathbb{E} [e^{wX}]$ , so that

$$\lambda(w) = \lim_{n \rightarrow \infty} \frac{1}{n} \ln g(w)^n = \ln g(w). \quad (15)$$

Using the product structure of the matrix product ansatz, we show in this section that a similar result holds for matrix-correlated random variables by replacing  $g(w)$  with a matrix generating function  $\mathcal{G}(w)$ . Specific examples of rate functions obtained from this matrix generalization of the Gärtner-Ellis Theorem are presented in the next section.

#### 3.1. Scaled cumulant generating function

Since  $\mathcal{L}$  is a linear form, we can directly expand (13) in the definition of the generating function  $g_n(w)$  to obtain

$$g_n(w) = \int_{\mathbb{R}} \frac{1}{\mathcal{L}(\mathcal{E}^n)} \mathcal{L}(e^{wx_1} \mathcal{R}(x_1) dx_1 \dots e^{wx_n} \mathcal{R}(x_n) dx_n) = \frac{\mathcal{L}(\mathcal{G}^n(w))}{\mathcal{L}(\mathcal{E}^n)}, \quad (16)$$

where

$$\mathcal{G}(w) \equiv \int_{\mathbb{R}} e^{wx} \mathcal{R}(x) dx \quad (17)$$

is the matrix generating function of the  $\mathcal{R}$  matrix. Each component of  $\mathcal{G}_{i,j}(w)$  can be decomposed as

$$\mathcal{G}_{i,j}(w) = \mathcal{E}_{i,j} g_{i,j}(w) \quad (18)$$

where  $g_{i,j}(w)$  is the generating function associated with the well defined (positive and normalizable) pdf  $\mathcal{P}_{i,j}(x)$  defined in (9). Substituting (16) in the limit defining the SCGF, we then obtain

$$\lambda(w) = \lim_{n \rightarrow \infty} \frac{1}{n} \ln \mathcal{L}(\mathcal{G}(w)^n) - K(\mathcal{E}), \quad (19)$$

where

$$K(\mathcal{E}) = \lim_{n \rightarrow \infty} \frac{1}{n} \ln \mathcal{L}(\mathcal{E}^n), \quad (20)$$

is a constant independent of  $w$ . In fact, this constant can be taken to be 0, since for any structure matrix  $\mathcal{E}$  there is a constant  $\eta > 0$  such that  $K(\eta\mathcal{E}) = 0$  and the pdf (2) is invariant by scalar multiplication of the structure matrix. The SCGF is given accordingly only by the limit involving the matrix  $\mathcal{G}(w)$ .

Our result (19) for the SCGF has an obvious similarity with the product structure of the i.i.d result of (15), as well as with the case of sample means of Markov-correlated random variables, for which we have

$$\lambda(w) = \lim_{n \rightarrow \infty} \frac{1}{n} \ln \text{tr} \rho \pi(w)^{n-1}, \quad (21)$$

where  $\pi(w)_{i,j} = \pi_{i,j} e^{wj}$  is the so-called tilted matrix obtained from the transition matrix  $\pi_{i,j}$  of the (assumed ergodic) Markov chain,  $\rho$  is the initial pdf of the Markov chain and  $\text{tr} v = \sum_i v_i$  is the vector component sum [3]. The similarity with Markov chains stems directly from the hidden Markov representation mentioned earlier. However, it is important to note that  $\pi(w)$  and  $\mathcal{G}(w)$  arise from different contexts and follow different sets of constraints. First,  $\pi(w)$  is not properly speaking a generating function whereas  $\mathcal{G}(w)$  is. Moreover,  $\pi(0) = \pi$  is a stochastic matrix (i.e.,  $\sum_j \pi_{i,j} = 1$ ), whereas  $\mathcal{G}(0) = \mathcal{E}$  is only a non-negative matrix. The matrix function  $\mathcal{G}(w)$  is therefore less constrained than  $\pi(w)$  in general. For instance, the dominant Jordan blocks  $J_{1,k}$  of  $\pi$  are necessarily of size 1 because  $\pi$  describes a probability flow, which has to vanish between different stationary states. The structure matrix  $\mathcal{E}$  is not subject to this restriction because it represents an affinity (i.e., a non-normalized probability) rather than a probability flow.

To find the SCGF  $\lambda(w)$ , we expand  $\mathcal{G}(w)^n$  using its Jordan decomposition. We slightly extend our previous notation to now denote by  $J_k(w)$  the Jordan block associated with the  $m(w)$  distinct eigenvalues  $\nu_k(w)$  of  $\mathcal{G}(w)$ . Then

$$\mathcal{G}(w) = \sum_{k=1}^{m(w)} \tilde{J}_k(w) = \sum_{k=1}^{m(w)} (\tilde{J}_k(w) - \nu_k(w)\mathbb{I}) + \nu_k(w)\mathbb{I} \quad (22)$$

where  $\tilde{J}_k(w)$  is the Jordan matrix  $J_k(w)$  completed with zeros outside of the block  $k$ :

$$\tilde{J}_k(w) = B \begin{pmatrix} 0 & & & 0 \\ & \ddots & & \\ & & J_k(w) & \\ 0 & & & \ddots & 0 \end{pmatrix} B^{-1}. \quad (23)$$

Using  $[\mathbb{I}, \tilde{J}_k] = 0$  and

$$(\tilde{J}_k(w) - \nu_k(w)\mathbb{I})^d = 0, \quad (24)$$

we can expand  $\mathcal{G}(w)^n$  as

$$\mathcal{G}(w)^n = \sum_{k=1}^{m(w)} \sum_{r=1}^d \nu_k(w)^n \binom{n}{r} \left( \frac{\tilde{J}_k(w)}{\nu_k(w)} - \nu_k(w)\mathbb{I} \right)^r. \quad (25)$$

Substituting (25) in (19) then leads to

$$\lim_{n \rightarrow \infty} \frac{\ln \mathcal{L}(\mathcal{G}(w)^n)}{n} = \max_l \{\ln |\nu_l(w)|\}, \quad (26)$$

where the maximum is on the eigenvalues  $\nu_l(w)$  such that  $\mathcal{L}(\tilde{J}_l(w)) \neq 0$ . This condition eliminates eigenvalues coming from  $\mathcal{E}$  that are unreachable in the hidden chain  $\Gamma$  due to a particular choice of  $\mathcal{A}$ . The conclusion that we reach from (26) is that the limit defining  $\lambda(w)$  exists for any matrix representation  $\mathcal{R}(x)$  and reads

$$\lambda(w) = \max_l \{\lambda_l(w)\} = \ln \nu_1(w), \quad (27)$$

with  $\lambda_l(w) = \ln |\nu_l(w)|$ . This result is similar to the Markov chain case, for which  $\lambda(w)$  is given by the dominant eigenvalue of the matrix  $\pi(w)$ . Thus, although  $\pi(w)$  and  $\mathcal{G}(w)$  are rather different matrices, they yield similar SCGFs in the  $n \rightarrow \infty$  limit because of the Perron-Frobenius Theorem. In both cases, a lot of structure contained in the generating function is in fact lost in the SCGF.

### 3.2. Rate function

To apply the Gärtner-Ellis Theorem, we need to study the differentiability of  $\lambda(w)$ . As in the Markov case, if the dominant eigenvalue is unique, then  $\lambda(w)$  is as smooth as  $\mathcal{G}(w)$ . However, if we have an eigenvalue collision for some  $w$ , i.e., if for some  $w$ , say  $w = w_0$ , the dominant eigenvalue  $\nu_1(w_0)$  has a multiplicity greater than 1, then  $\lambda(w_0)$  may be nondifferentiable because of the maximum in (27). Physically, this non-differentiability is interpreted as a dynamical first-order phase transition at  $w_0$  [39–42]. Around such a collision point, the colliding log-eigenvalues  $\lambda_r(w)$  of  $\mathcal{G}(w)$  can be generically approximated as

$$\lambda_r(w_0 + w) = \lambda(w_0) + \mu_r w + O(w^2), \quad (28)$$

so that, by (27), we have

$$\lambda(w_0 + w) = \lambda(w_0) + \begin{cases} \max\{\mu_r\} w + O(w^2) & \text{if } w > 0 \\ \min\{\mu_r\} w + O(w^2) & \text{if } w < 0. \end{cases} \quad (29)$$

Thus, we see that if the  $\mu_r$  are different, then  $\lambda(w)$  is nondifferentiable at  $w_0$ . In this case, we can still apply the Gärtner-Ellis Theorem but only locally at points where  $\lambda(w)$  is differentiable; see Sec. 4.4. of [3] for more details. Here, this means that we can apply this theorem at all  $w$  except  $w_0$ . Doing so yields  $I(s)$  as the Legendre transform of  $\lambda(w)$  for  $s \leq \min \mu_r$  and  $s \geq \max \mu_r$ , but not for  $s \in (\min \mu_r, \max \mu_r)$  because of the nondifferentiability of  $\lambda(w)$  at  $w_0$  and the fact that  $\lambda'(w_0 - 0) = \min \mu_r$  and  $\lambda'(w_0 + 0) = \max \mu_r$ . On this open interval,  $I(s)$  can be nonconvex or linear, but this cannot be determined from  $\lambda(w)$ , as explained in more detail in [3, 7].

Nonconvex rate functions typically appear in non-ergodic Markov chains [8, 43, 44] in addition to mixtures of i.i.d sample means [3, 45], whereas rate functions with linear branches are known to arise in Markov chains with absorbing states. All these cases can lead to extensions of the Law of Large Numbers involving more than one concentration points. In our case, we expect this sort of extensions to arise whenever long-range correlations are present, since this case of correlations appears when the structure matrix  $\mathcal{E}$  has a multiple dominant eigenvalue at  $w = 0$ . This will be investigated in Sec. 4.

3.3. Connection with the Law of Large Numbers and the Central Limit Theorem

The rate function  $I(s)$  provides information not only about the large deviations of the sample mean  $S_n/n$ , but also about its small deviations and its most probable values corresponding to the zeros and global minima of  $I(s)$ . In the case where  $I(s)$  has only one global minimum at  $s = \mu$ , then the most probable value is also the typical value, in the sense that  $S_n/n$  converges almost surely to  $\mu$ . In this case, the sample mean thus concentrates to the mean, in accordance with the Law of Large Numbers.

Using the Gärtner-Ellis Theorem, we can express the concentration point  $\mu$  as  $\mu = \lambda'(0)$ , assuming that  $\lambda(w)$  is differentiable at  $w = 0$ . In our case, we have  $\lambda(w) = \ln \nu_1(w)$  and the normalization condition  $K(\mathcal{E}) = 0$  implies (see (20)) that  $\nu_1(0) = 1$ , so that  $\lambda'(0) = \nu_1'(0)$ . Using classical perturbation theory [46] for the eigenvalue  $\nu_1(w)$  then yields

$$\mu = \nu_1'(0) = \sum_{i,j} \rho_i \mathcal{G}'_{i,j}(0) v_j \tag{30}$$

where  $\rho$  and  $v$  are respectively the left- and right-eigenvectors of  $\mathcal{E}$  associated with  $\nu_1(0)$ , normalized so that  $\text{tr } \nu = 1$  and  $\nu^T \rho = 1$ . Since the matrix  $\mathcal{G}'(0)$  can be decomposed as

$$\mathcal{G}'_{i,j}(0) = \mathcal{E}_{i,j} \left( \partial_w \int_{\mathbb{R}} e^{-wx} \mathcal{P}_{i,j}(x) dx \right) \Big|_{w=0} = \mathcal{E}_{i,j} \mu_{i,j} \tag{31}$$

where  $\mu_{i,j}$  is the mean of the pdf  $\mathcal{P}_{i,j}(x)$ , we also have

$$\mu = \sum_{i,j} (\rho_i \mathcal{E}_{i,j} v_j) \mu_{i,j}. \tag{32}$$

Therefore, we see that the concentration point of  $S_n/n$  is a weighted average of the mean of  $\mathcal{P}_{i,j}(x)$ . This should hold in general whenever  $\mathcal{E}$  leads to short-range correlations.

As before, we can understand this result probabilistically by appealing to the hidden Markov chain representation. For collision-free structure matrices  $\mathcal{E}$ , the hidden Markov chain is known to converge towards a stationary state almost surely [34, 47]. In this state, the mean of  $S_n/n$  is then a weighted mixture of the mean of different distributions  $\mathcal{P}_{i,j}(x)$ , where the weights are the stationary probabilities, corresponding to  $\rho_i \mathcal{E}_{i,j} v_j$ , of observing a transition from state  $i$  to  $j$  in the hidden Markov chain.

To close this section, let us study the fluctuations of  $S_n$  around its concentration point, characterized by the behavior of  $I(s)$  around its minimum. It is often stated that if  $I(s)$  has a quadratic minimum, then the sum  $S_n$  satisfies the Central Limit Theorem in the sense that

$$\frac{S_n - \lambda'(0)n}{\sqrt{\lambda''(0)n}} \tag{33}$$

converges in distribution towards the normal distribution. However, this relation is not rigorously valid: to obtain the Central Limit Theorem requires further conditions, such as  $\lambda(w)$  to be holomorphic [48].

In our case,  $\lambda(w)$  is holomorphic if  $\mathcal{G}(w)$  is collision-free and holomorphic. The matrix function  $\mathcal{G}(w)$  is holomorphic whenever the cumulant generating functions of the pdf  $\mathcal{P}_{i,j}(x)$  are holomorphic. This holomorphic condition is stronger than the differentiability condition needed to derive the rate function  $I(s)$  which is nevertheless



satisfied for a large class of  $\mathcal{P}_{i,j}(x)$  §. In the presence of a collision, it is possible to obtain a SCGF that is differentiable but non-holomorphic in 0. To see this, assume that the colliding log-eigenvalues  $\lambda_r(w)$  are twice differentiable and  $\lambda'_r(0) = 0$ . Then

$$\lambda_r(w) = \frac{\sigma_r^2}{2}w^2 + O(w^3) \tag{34}$$

and

$$\lambda(w) = \frac{\max\{\sigma_r^2\}}{2}w^2 + O(w^3), \tag{35}$$

so that  $\lambda(w)$  is differentiable at 0. However, considering now the argument of the function  $\lambda_r(w)$  as a complex variable  $z$  and rewriting these expressions in the direction  $z = \imath\mathbb{R}$  yields

$$\lambda_r(z = \imath w) = -\frac{\sigma_r^2}{2}w^2 + O(w^3) \tag{36}$$

and

$$\lambda(z = \imath w) = -\frac{\min\{\sigma_r^2\}}{2}z^2 + O(z^3), \tag{37}$$

showing that  $\lambda(w)$  is not holomorphic if the constants  $\sigma_r$  are not equal. In this case, the Central Limit Theorem might not hold for the rescaled sum  $S_n/\sqrt{n}$ , consistently with the results of [34].

#### 4. Long-range correlation examples

We now give illustrations of the two most interesting cases of correlation obtained with the matrix ansatz, namely, polynomial and constant correlations, and obtain the rate functions of the sample mean  $S_n/n$  for both cases. For simplicity, we consider the case of two-dimensional matrices ( $d=2$ ), which already shows interesting large deviation behavior.

##### 4.1. Rate function with flat part

For  $d = 2$ , the only structure matrix (up to some trivial transformations) leading to polynomial correlation is  $\mathcal{E} = \mathbb{I} + U$  [47]. This kind of correlation structure does not exist for finite Markov chain, so it is interesting to determine the rate function  $I(s)$  in this case.

For simplicity, we consider the projection matrix

$$\mathcal{A} = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \tag{38}$$

for which the generating function  $g_n(w)$  can be computed exactly as

$$g_n(w) = \mathcal{G}_{1,2}(w) \frac{\mathcal{G}_{1,1}^n(w) - \mathcal{G}_{2,2}^n(w)}{\mathcal{G}_{1,1}(w) - \mathcal{G}_{2,2}(w)}, \tag{39}$$

where  $\mathcal{G}_{i,j}(w)$  represents the (scalar) generating function of the pdf  $\mathcal{P}_{i,j}(x)$ . Due to the choice of  $\mathcal{E}$ , the matrix  $\mathcal{G}(w)$  is triangular superior. The diagonal coefficients  $\mathcal{G}_{k,k}$

§ Log-normal distributions are a noteworthy example of distributions which do not satisfy this condition, even if sums of log-normal i.i.d random variables do converge towards a normal distribution.

are therefore the eigenvalues  $\nu_k(w)$  of  $\mathcal{G}(w)$ . Defining  $c(w) = \mathcal{G}_{1,2}(w)$  and using the notation  $\lambda_k(w) = \ln \nu_k(w)$  introduced in the previous section then yields

$$g_n(w) = c(w) \frac{e^{n\lambda_1(w)} - e^{n\lambda_2(w)}}{e^{\lambda_1(w)} - e^{\lambda_2(w)}}, \quad (40)$$

The expression of  $g_n(w)$  is at this point quite general and covers many cases of large deviations. To be more concrete, we now make a number of assumptions leading to a rate function having a flat part. For this, we assume that

$$\lambda'_1(0) = \mu_1 < \lambda'_2(0) = \mu_2 \quad (41)$$

and that  $\lambda_1(w) = \lambda_2(w)$  has the only solution  $w = 0$ . In this case, it can be verified from (40) that  $\lambda(w)$  is differentiable everywhere except at  $w = 0$ . The left- and right-derivatives of  $\lambda(0)$  are  $\lambda'(0^-) = \mu_1$  and  $\lambda'(0^+) = \mu_2$ , respectively. Following our discussion of the Gärtner-Ellis Theorem of the previous section, the rate function  $I(s)$  can therefore be obtained from the Legendre transform of  $\lambda(w)$  for  $s \in (-\infty, \mu_1] \cup [\mu_2, \infty)$  but not for the complementary interval  $(\mu_1, \mu_2)$ . In the latter interval,  $I(s)$  could be flat or nonconvex, but as mentioned before the knowledge of  $\lambda(w)$  is not sufficient to discriminate between these two cases.

However, since we have the exact expression of the generating function for any  $n$  and therefore more information than the sole limit  $\lambda(w)$ , we can obtain the pdf of  $S_n/n$  by inverting the Laplace transform

$$P(S_n/n = s) = \frac{1}{2i\pi} \int_{r-i\infty}^{r+i\infty} g_n(w) e^{-nws} dw, \quad (42)$$

where  $r$  is an arbitrary real constant in the region of convergence of  $g_n(w)$ . From this integral, involving the so-called Bromwich contour, we can obtain the rate function of  $S_n/n$  following the method proposed in [38] by expanding  $g_n(w)$  in series form and by applying a saddlepoint approximation to each term in the series, taking care of any pole when transforming the integration contour to reach the saddlepoints.

In our case, the series is simply  $g_n(w) = g_{n,1}(w) - g_{n,2}(w)$ , where

$$g_{n,k}(w) = c(w) \frac{e^{n\lambda_k(w)}}{e^{\lambda_1(w)} - e^{\lambda_2(w)}}. \quad (43)$$

The two functions  $g_{n,1}(w)$  and  $g_{n,2}(w)$  have a pole in 0, since  $\lambda_1(0) = \lambda_2(0) = 0$ , whose residue is

$$\text{Res}_{g_{n,1}(w)}(0) = \frac{1}{\mu_1 - \mu_2} = \text{Res}_{g_{n,2}(w)}(0). \quad (44)$$

With this, we now split the integral of the inverse Laplace transform in two parts:

$$P(S_n/n = s) = p_{n,1}(s) - p_{n,2}(s), \quad (45)$$

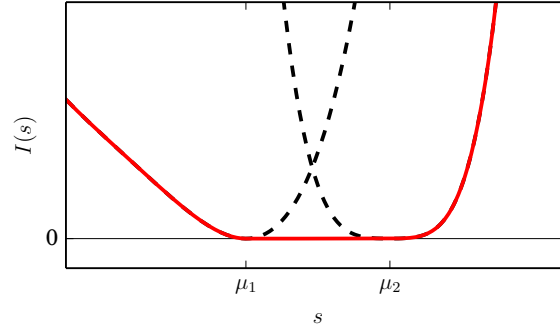
where

$$p_{n,k}(s) = \frac{1}{2i\pi} \int_{r-i\infty}^{r+i\infty} g_{n,k}(w) e^{-nws} dw, \quad (46)$$

and apply the saddlepoint approximation to each term, starting with  $r \neq 0$ , by deforming the Bromwich contour to go through the saddlepoint  $w_k^*$  of the exponential term satisfying

$$\lambda'_k(w_k^*) = s. \quad (47)$$

Two different situations then arise:



**Figure 1.** Sketch of the rate function for linear irreversible structure matrix. Dashed lines: underlying rate function  $I_1(s)$  and  $I_2(s)$ . Red solid line: resulting rate function  $I(s)$ .

- If  $w_k^*$  has the same sign as  $r$ , the integration path can be deformed without going through the pole at  $w = 0$ , which leads to

$$p_{n,k}(s) \approx e^{-n[s w_k^* - \lambda_k(w_k^*)]} = e^{-n \lambda_k^*(s)} \quad (48)$$

where  $\lambda_k^*$  is the Legendre transform of  $\lambda_k$ .

- The deformation of the contour to the saddlepoint must cross the pole 0, in which case the residue must be included:

$$p_{n,k}(s) \approx \text{Res}_{g_{n,k}(w)}(0) + e^{-n \lambda_k^*(s)}. \quad (49)$$

Here, the transition between these two regimes happens for  $s = \mu_k$  since  $\lambda_k'(0) = \mu_k$ . For instance, if we choose  $r < 0$ , then

$$p_{n,k}(s) \approx e^{-n \lambda_k^*(s)} + \begin{cases} 0 & s < \mu_k, \\ \text{Res}_{g_{n,k}(w)}(0) & \mu_k < s. \end{cases} \quad (50)$$

Outside the interval  $[\mu_1, \mu_2]$ , the residue terms cancel each other, so that combining  $p_{n,1}(s)$  and  $p_{n,2}(s)$  yields

$$I(s) = \begin{cases} I_1(s) & s < \mu_1, \\ 0 & \mu_1 < s < \mu_2, \\ I_2(s) & \mu_2 < s. \end{cases} \quad (51)$$

Here  $I_k(s) = \lambda_k^*(s)$  actually corresponds to the rate function of a sample mean  $S_n/n$  of i.i.d random variables distributed with pdf  $\mathcal{P}_{k,k}(x)$ .

The result for  $I(s)$  is sketched in Fig. 1. As clearly seen, the effect of the pole in each term of the generating function is that  $I(s)$  is flat for  $x \in (\mu_1, \mu_2)$ . This automatically implies that there is no Law of Large Numbers for  $S_n/n$ , that is,  $S_n/n$  does not concentrate to a Dirac-delta pdf in the limit  $n \rightarrow \infty$ . It may instead converge to a pdf that scales slower than exponentially for  $s \in (\mu_1, \mu_2)$  or to some stationary distribution that does not scale with  $n$ .

To find out, we can use the hidden Markov representation. For the structure matrix  $\mathcal{E} = \mathbb{I} + U$ , the corresponding hidden Markov chain is a highly non-stationary Markov chain [34], which for  $d = 2$  stays in its initial state for a random time uniformly distributed on  $\{0, 1, \dots, n\}$  before jumping to its final state. Consequently, the Markov chain does not converge almost surely towards a stationary state, and in this case, it can be shown that the sample mean actually converges towards the uniform distribution on the interval  $[\mu_1, \mu_2]$  [34]. Thus, we have a concentration phenomenon on a whole interval rather than on a point, explaining the flat branch in the rate function  $I(s)$ .

Our particular choice for the matrices  $\mathcal{E}$  and  $\mathcal{A}$  simplifies the computations leading to this flat branch, but is otherwise not significant. Using the Perron-Frobenius decomposition of the matrix  $\mathcal{E}$  into irreducible blocks, it is possible to show that  $g_n(w)$  has a form similar to (39) whenever  $\mathcal{E}$  leads to long-range correlations, so that the calculation steps given above are representative of this case. Consequently, we expect flat branches in rate functions to be a generic phenomenon for matrix-correlated random variables  $X_1, \dots, X_n$  exhibiting polynomial long-range correlation.

#### 4.2. Nonconvex rate function

We now consider a model with constant correlation. For  $d = 2$ , the only structure matrix with such correlation is

$$\mathcal{E} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}. \tag{52}$$

Choosing, for simplicity,

$$\mathcal{A} = \begin{pmatrix} \frac{1}{2} & 0 \\ 0 & \frac{1}{2} \end{pmatrix}, \tag{53}$$

the computation of the generating function is then straightforward and leads to

$$g_n(w) = \frac{1}{2} \left( e^{n\lambda_1(w)} + e^{n\lambda_2(w)} \right). \tag{54}$$

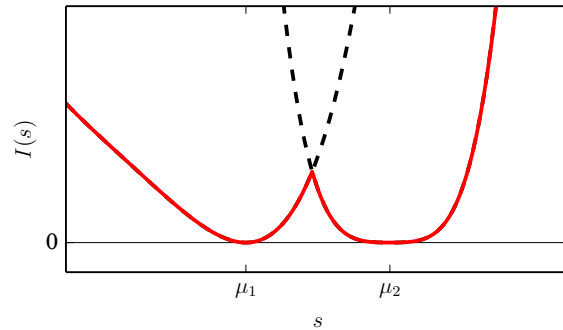
As before, we have  $\lambda_1(0) = \lambda_2(0) = 0$  and the SCGF is not differentiable assuming  $\lambda_1'(0) \neq \lambda_2'(0)$ . In this case, we calculate the rate function using the inverse Laplace transform method, and, since there is now no pole, the calculation is much simpler and leads to

$$I(s) = \min\{I_1(s), I_2(s)\} \tag{55}$$

where  $I_k(s) = \lambda_k^*(s)$  corresponds again to the rate function of a sample mean of i.i.d random variables with pdf  $\mathcal{P}_{k,k}(x)$ .

The full rate function  $I(s)$  is sketched in Fig. 2. Its main feature is that it is nonconvex with our assumption that  $\lambda_1'(0) \neq \lambda_2'(0)$ , which implies for this case that there are two distinct concentration points for  $S_n/n$ . These two concentration points can be understood again using the hidden Markov chain representation. For the structure matrix  $\mathcal{E} = \mathbb{I}$ , the hidden Markov chain stays in the same state from its beginning to its end. The concentration and fluctuations of  $S_n/n$  therefore simply depend on the choice of initial state, similarly to the case of non-ergodic Markov chains.

This example is simple, but should be relevant to other cases of matrix-correlated variables leading to constant correlation. In this case, it is possible to show, using the Perron-Frobenius decomposition of the matrix  $\mathcal{E}$  mentioned before, that decoupled terms such as those appearing in the right-hand side of (54) are present in the expression of  $g_n(w)$ . As these decoupled terms are responsible for the nonconvex part of  $I(s)$ , we expect nonconvex rate functions to be generic in this case.



**Figure 2.** Sketch of the rate function for identity structure matrix. Dashed lines: underlying rate function  $I_1(s)$  and  $I_2(s)$ . Red solid line: resulting rate function  $I(s)$ .

## 5. Conclusion

We have shown in this contribution how to obtain large deviations for sums of random variables described by a matrix product ansatz. These random variables are also described by a complementary hidden Markov model, which is probably the more natural setting for studying limits laws, such as the Law of Large Numbers and Central Limit Theorem, as was done in [34]. However, as shown here, the matrix product representation offers a natural starting point for generalizing large deviations results of i.i.d variables in the context of the Gärtner-Ellis Theorem by introducing a matrix cumulant generating function.

For system with short-range correlation, the Gärtner-Ellis Theorem was used to show that analogues of the Law of Large Number or the Central Limit Theorem hold. In the presence of long-range correlation, the direct calculation of the rate function for two specific examples has shown that flat and nonconvex rate functions are possible. Nonconvex rate functions are associated with constant correlation and are similar in nature to nonconvex rate functions appearing in non-ergodic Markov chains. By contrast, polynomial correlation leads to flat rate functions, which cannot appear for sample means defined on finite Markov chains.

## Acknowledgments

HT thanks Stefano Ruffo and Thierry Dauxois for supporting a visit to ENS Lyon with the ANR grant LORIS (ANR-10-CEXC-010-01).

## References

- [1] L. Bertini, A. De Sole, D. Gabrielli, G. Jona-Lasinio, and C. Landim. Stochastic interacting particle systems out of equilibrium. *J. Stat. Mech.*, 2007(07):P07014, 2007.
- [2] B. Derrida. Non-equilibrium steady states: Fluctuations and large deviations of the density and of the current. *J. Stat. Mech.*, 2007(07):P07023, 2007.
- [3] H. Touchette. The large deviation approach to statistical mechanics. 478(1-3):1–69, 2009.
- [4] R. J. Harris and H. Touchette. Large deviation approach to nonequilibrium systems. In R. Klages, W. Just, and C. Jarzynski, editors, *Nonequilibrium Statistical Physics of Small*

- Systems: Fluctuation Relations and Beyond*, volume 6 of *Reviews of Nonlinear Dynamics and Complexity*, pages 335–360, Weinheim, 2013. Wiley-VCH.
- [5] R. S. Ellis. *Entropy, Large Deviations, and Statistical Mechanics*. Springer, New York, 1985.
  - [6] Y. Oono. Large deviation and statistical physics. *Prog. Theoret. Phys. Suppl.*, 99:165–205, 1989.
  - [7] R. S. Ellis. An overview of the theory of large deviations and applications to statistical mechanics. *Scand. Actuarial J.*, 1:97–142, 1995.
  - [8] R. S. Ellis. The theory of large deviations: From Boltzmann’s 1877 calculation to equilibrium macrostates in 2D turbulence. *Physica D*, 133:106–136, 1999.
  - [9] H.-O. Georgii. *Gibbs Measures and Phase Transitions*, volume 9. Walter de Gruyter & Co., Berlin, 1988.
  - [10] H. Föllmer and S. Orey. Large deviations for the empirical field of a Gibbs measure. *Ann. Prob.*, 16(3):961–977, 1988.
  - [11] H.-O. Georgii. Large deviations and maximum entropy principle for interacting random fields on  $\mathbf{Z}^d$ . *Ann. Prob.*, 21(4):1845–1875, 1993.
  - [12] A. Eizenberg, Y. Kifer, and B. Weiss. Large deviations for  $\mathbf{Z}^d$  actions. *Comm. Math. Phys.*, 164:433–454, 1994.
  - [13] G. W. Anderson, A. Guionnet, and O. Zeitouni. *An Introduction to Random Matrices*. Cambridge University Press, Cambridge, 2010.
  - [14] P. Vivo, S. N. Majumdar, and O. Bohigas. Large deviations and random matrices. *Acta Phys. Polonica B*, 38(13):4139–4151, 2007.
  - [15] A. Guionnet. Large deviations and stochastic calculus for large random matrices. *Prob. Surveys*, 1:72–172, 2004.
  - [16] D. S. Dean and S. N. Majumdar. Large deviations of extreme eigenvalues of random matrices. *Phys. Rev. Lett.*, 97(16):160201, 2006.
  - [17] E. Katzav and I. Pérez Castillo. Large deviations of the smallest eigenvalue of the Wishart-Laguerre ensemble. *Phys. Rev. E*, 82:040104, 2010.
  - [18] A. Greven and F. den Hollander. Large deviations for a random walk in random environment. *Ann. Prob.*, 22(3):1381–1428, 1994.
  - [19] F. den Hollander. *Large Deviations*. Fields Institute Monograph. Amer. Math. Soc., Providence, R.I., 2000.
  - [20] O. Zeitouni. Random walks in random environments. *J. Phys. A*, 39(40):R433–R464, 2006.
  - [21] Y. Ephraim and N. Merhav. Hidden Markov processes. *IEEE Trans. Info. Th.*, 48(6):1518–1569, 2002.
  - [22] V. Kargin. A large deviation inequality for vector functions on finite reversible Markov chains. *Ann. Appl. Prob.*, 17(4):1202–1221, 2007.
  - [23] M. del Greco Fabiola. Applications of large deviations to hidden Markov chains estimation. In A. Di Ciaccio, M. Coli, and J. M. Angulo Ibanez, editors, *Advanced Statistical Methods for the Analysis of Large Data-Sets*, Studies in Theoretical and Applied Statistics, pages 279–285. Springer, New York, 2012.
  - [24] R. J. Harris and H. Touchette. Current fluctuations in stochastic systems with long-range memory. *J. Phys. A*, 42(34):342001, 2009.
  - [25] F. Angeletti, E. Bertin, and P. Abry. Matrix products for the synthesis of stationary time series with a priori prescribed joint distributions. In *Proceeding of the IEEE Int. Conf. on Acoust. Speech and Sig. Proc. (ICASSP)*, pages 3897 – 3900, 2012.
  - [26] F. Angeletti, E. Bertin, and P. Abry. Random vector and time series definition and synthesis from matrix product representations: From statistical physics to hidden markov models. *IEEE Transactions on Signal Processing*, 61:5389 – 5400, 2013.
  - [27] V. Hakim and J. P. Nadal. Exact results for 2D directed animals on a strip of finite width. *J. Phys. A*, 16(7):L213, 1983.
  - [28] N. Crampe, E. Ragoucy, and D. Simon. Matrix coordinate Bethe ansatz: applications to XXZ and ASEP models. *J. Phys. A*, 44(40):405003, 2011.
  - [29] A. Lazarescu and K. Mallick. An exact formula for the statistics of the current in the TASEP with open boundaries. *J. Phys. A*, 44(31):315001, 2011.
  - [30] A. Lazarescu. Matrix ansatz for the fluctuations of the current in the ASEP with open boundaries. *J. Phys. A*, 46(14):145003, 2013.
  - [31] B. Derrida, M. R. Evans, V. Hakim, and V. Pasquier. Exact solution of a 1D asymmetric exclusion model using a matrix formulation. *J. Phys. A*, 26:1493–1517, 1993.
  - [32] K. Mallick and S. Sandow. Finite dimensional representations of the quadratic algebra: Applications to the exclusion process. *J. Phys. A*, 30:4513, 1997.
  - [33] R. A. Blythe and M. R. Evans. Nonequilibrium steady states of matrix-product form: a solver’s guide. *J. Phys. A*, 40(46):R333–R441, 2007.

- [34] F. Angeletti, E. Bertin, and P. Abry. Statistics of sums of correlated variables described by a matrix product ansatz. *European Physics Letters*, 104(5):50009, 2013.
- [35] F. Verstraete, V. Murg, and J.I. Cirac. Matrix product states, projected entangled pair states, and variational renormalization group methods for quantum spin systems. *Advances in Physics*, 57(2):143–224, 2008.
- [36] I. Lesanovsky, M. van Horssen, M. Guță, and J. P. Garrahan. Characterization of dynamical phase transitions in quantum jump trajectories beyond the properties of the stationary state. *Phys. Rev. Lett.*, 110:150401, 2013.
- [37] J. Gärtner. On large deviations from the invariant measure. *Th. Prob. Appl.*, 22:24–39, 1977.
- [38] H. Touchette, R. J. Harris, and J. Tailleur. First-order phase transitions from poles in asymptotic representations of partition functions. *Phys. Rev. E*, 81(3):030101, 2010.
- [39] C. Appert-Rolland, B. Derrida, V. Lecomte, and F. van Wijland. Universal cumulants of the current in diffusive systems on a ring. *Phys. Rev. E*, 78:021122, 2008.
- [40] R. Lefevere, M. Mariani, and L. Zambotti. Macroscopic fluctuation theory of aerogel dynamics. *J. Stat. Mech.*, 2010(12):L12004, 2010.
- [41] R. Lefevere, M. Mariani, and L. Zambotti. Large deviations of the current in stochastic collisional dynamics. *Journal of Mathematical Physics*, 52(3):–, 2011.
- [42] V. Lecomte, J. P. Garrahan, and F. van Wijland. Inactive dynamical phase of a symmetric exclusion process on a ring. *J. Phys. A*, 45(17):175001, 2012.
- [43] I. H. Dinwoodie and S. L. Zabell. Large deviations for exchangeable random vectors. *Ann. Prob.*, 20:1147–1166, 1992.
- [44] I. H. Dinwoodie. Identifying a large deviation rate function. *Ann. Prob.*, 21:216–231, 1993.
- [45] D. Ioffe. Two examples in the theory of large deviations. *Stat. Prob. Lett.*, 18:297–300, 1993.
- [46] T. Kato. *Perturbation Theory for Linear Operators*. Classics in mathematics. Springer, 1995.
- [47] F. Angeletti, E. Bertin, and P. Abry. General statistics of sums of correlated variables described by a matrix product ansatz. *Work in progress*, 2014.
- [48] W. Bryc. A remark on the connection between the large deviation principle and the central limit theorem. *Stat. Prob. Lett.*, 18(4):253–256, 1993.