CrossMark

The Semi-infinite Asymmetric Exclusion Process: Large Deviations via Matrix Products

Horacio González Duhart[1] · Peter Mörters[1] ·
Johannes Zimmer[1]

Abstract We study the totally asymmetric exclusion process on the positive integers with a single particle source at the origin. Liggett (Trans. Am. Math. Soc. **213**, 237–261, 1975) has shown that the long term behaviour of this process has a phase transition: If the particle production rate at the source and the original density are below a critical value, the stationary measure is a product measure, otherwise the stationary measure is spatially correlated. Following the approach of Derrida et al. (J. Phys. A **26**(7), 1493, 1993) it was shown by Grosskinsky (2004) that these correlations can be described by means of a matrix product representation. In this paper we derive a large deviation principle with explicit rate function for the particle density in a macroscopic box based on this representation. The novel and rigorous technique we develop for this problem combines spectral theoretical and combinatorial ideas and is potentially applicable to other models described by matrix products.

✉ Peter Mörters
maspm@bath.ac.uk

Horacio González Duhart
h.duhart@bath.edu

Johannes Zimmer
J.Zimmer@bath.ac.uk

[1] University of Bath, Bath, UK

1 Introduction

Many natural systems are not in thermodynamic equilibrium, which loosely speaking means that there is a permanent exchange of energy or matter of the system with its surroundings or within the system itself. In statistical physics, the asymmetric exclusion process is often considered the paradigm of such a system out of equilibrium. In the absence of a general theory for systems out of equilibrium, it has been argued that large deviation rate functions play an important role as a replacement for the thermodynamical potential [4]. The principal aim of the present paper is to develop a rigorous mathematical technique to derive such rate functions from a particular type of representation of the stationary state of the system, the matrix products, which twenty years after the pioneering work of Derrida et al. [11] is available for a wide range of particle systems out of equilibrium, see for example Blythe and Evans [6] for a survey.

We present our method in the case of the totally asymmetric exclusion process (TASEP) on the positive integers with a single particle source at the origin, a case which has apparently not been treated in the literature so far. In this Markovian model, particles are positioned on the sites of the semi-infinite lattice $\mathbb{N} = \{1, 2, \dots\}$ in such a way that no site carries more than one particle. The dynamics of the model can be informally described as follows: A particle source carries a Poisson clock with intensity $\alpha > 0$. If this clock rings, the source attempts to inject a particle at site one. If this site is vacant the injection takes place, otherwise it is suppressed and nothing happens. Also, every particle in the system carries an independent Poisson clock with rate one, and when the clock rings the particle tries to jump to the neighbouring site on its right. If this site is vacant the jump takes place, otherwise it is suppressed. Note that the exclusion interaction originating from the suppression of jumps and injections ensures that no site ever carries more than one particle.

The exclusion interaction in this model has a profound effect on the behaviour of the system. Most notably the detailed balance equations for this Markov chain have no non-trivial solution. Hence the system is not reversible, in other words it is out of equilibrium. The long-term behaviour of the process shows local convergence to a stationary measure which depends on the initial configuration of the system. Assuming that initially particles are iid Bernoulli with density ρ, this stationary measure has an interesting phase transition described by Liggett [16]. If the injection rate α satisfies $\alpha \leq \frac{1}{2}$ and $\rho \leq 1 - \alpha$ the system does not feel the interaction and the stationary measure is the product measure with density α. If however $\alpha > \frac{1}{2}$ or $\rho > 1 - \alpha$, the exclusion of particles leads to spatial correlations in the stationary measure, which is no longer a product measure. In this case, the overall particle density at stationarity is the maximum of $1/2$ and the initial density ρ, independently of the injection rate α.

There have been considerable efforts to describe the long range correlations of the stationary measures and the microscopic transition kernels in the exclusion process explicitly. For instance, Sasamoto and Williams [19] and Tracy and Widom [21] derive explicit formulas from combinatorial identities, and Sasamoto [18] uses an ansatz based on orthogonal polynomials. A particularly successful approach to describe spatial correlations is the matrix product ansatz first suggested in 1993 by Derrida, Evans, Hakim and Pasquier [11] and refined and extended in a large number of papers, see [10, 13, 15] for a few further examples. Large deviation principles have been derived for the hydrodynamical limits of a range of boundary driven exclusion processes by Bertini and coauthors [3, 5] and the method should be extendable to our case. In principle, large deviation principles for the particle density in a macroscopic box then follow from these results by contraction, see [7]. However, the optimisation in path space, which is required to get an explicit rate function, is often unwieldy and technical as Bahadoran's paper [2] readily testifies.

In the light of these difficulties it is a natural idea to try and derive large deviation principles directly from the matrix product ansatz. This plan was carried out by Derrida et al. [12] in the case of an asymmetric exclusion process on a finite interval of sites. Key to their method is a saddle point argument, which allows to derive an additivity formula which compares the stationary measure on the interval with stationary measures on complementary subintervals. From this formula an explicit rate function for the particle density is derived. The paper [12] was a spectacular success, but we have not been able to implement this method in the case of a semi-infinite lattice. In a different development, Angeletti et al. [1] show that already for matrix product representations with finite matrices the large deviation principles that arise from this exhibit a rich phenomenology. Finite matrix representations have the advantage that they can be studied using the Perron-Frobenius theory, which is unavailable for infinite matrices. Physical examples, however, are almost always based on representations by infinite matrices.

In this paper we present a rigorous and novel approach to calculate large deviations for the macroscopic particle density in the semi-infinite totally asymmetric exclusion process. We use the matrix product representation as a starting point, and base the analysis on the Gärtner-Ellis theorem. To study the asymptotics of the cumulant generating function of the particle density, we use quite different approaches for the lower and upper bounds. The lower bound is based on the spectral theory of Toeplitz operators in a suitable weighted sequence space, while the upper bound exploits combinatorial identities coming directly from the matrix product ansatz. As our method is not too technical, we believe that it is very promising to deal with a wide range of other particle systems whose stationary measure can be described by a matrix product representation.

The paper is organised as follows. In Section 2 we give a rigorous definition of the model, state background results and formulate and interpret our main result. Section 3 discusses the matrix product representation in this case and describes our approach to the large deviation problem. The proof of the upper bound for the cumulant generating function is carried out in Section 4, while the lower bound is derived in Section 5. The proof is completed in Section 6, in which we also provide some further comments on our technique.

2 The Semi-infinite TASEP

2.1 Background

To give a formal definition of the model, we first define the auxiliary switching and swapping functions $\sigma^x, \sigma^{x,y} \colon \{0, 1\}^{\mathbb{N}} \to \{0, 1\}^{\mathbb{N}}$ by

$$(\sigma^{x,y}\eta)_z = \begin{cases} \eta_y & \text{if } z = x \\ \eta_x & \text{if } z = y \\ \eta_z & \text{if } z \notin \{x, y\}, \end{cases} \qquad (\sigma^x \eta)_z = \begin{cases} 1 - \eta_x & \text{if } z = x \\ \eta_z & \text{if } z \neq x. \end{cases}$$

A semi-infinite totally asymmetric exclusion process (TASEP) with injection rate $\alpha \in (0, 1)$ is a Markov process $\{\xi(t)\}_{t \geq 0}$ in continuous time with state space $\{0, 1\}^{\mathbb{N}}$ and semigroup $S(t)$ identified by its infinitesimal generator G defined by

$$\begin{aligned} Gf(\eta) = \ &\alpha(1 - \eta_1)\left(f(\sigma^1 \eta) - f(\eta)\right) \\ &+ \sum_{k \in \mathbb{N}} \eta_k(1 - \eta_{k+1})\left(f(\sigma^{k,k+1}\eta) - f(\eta)\right), \end{aligned} \qquad (1)$$

where $f : \{0, 1\}^{\mathbb{N}} \to \mathbb{R}$ is a function that depends only on a finite number of sites.

Denote by ν_α the *product measure* with constant density α, that is

$$\nu_\alpha\{\eta \in \{0, 1\}^{\mathbb{N}} : \eta_{j_1} = 1, \eta_{j_2} = 1, \ldots, \eta_{j_n} = 1\} = \alpha^n$$

for all distinct choices of $j_1, j_2, \ldots, j_n \in \mathbb{N}$ and all $n \in \mathbb{N}$. We say a measure μ_ρ on $\{0, 1\}^{\mathbb{N}}$ is *asymptotically product* with density ρ if

$$\lim_{k \to \infty} \mu_\rho\{\eta \in \{0, 1\}^{\mathbb{N}} : \eta_{j_1+k} = 1, \eta_{j_2+k} = 1, \ldots, \eta_{j_n+k} = 1\} = \rho^n$$

for all distinct choices of $j_1, j_2, \ldots, j_n \in \mathbb{N}$ and all $n \in \mathbb{N}$. It is known that the semi-infinite TASEP is not an ergodic process and there is no uniqueness of the stationary measure as proved in Theorem 1.8 of [16].

Theorem 2.1 [16, Theorem 1.8] *Let μ be a product measure on $\{0, 1\}^{\mathbb{N}}$ for which $\rho :=$ $\lim_{k \to \infty} \mu\{\eta : \eta_k = 1\}$ exists. Then there exist probability measures μ_ϱ^α defined if*

- *either $\alpha \leq \frac{1}{2}$ and $\varrho > 1 - \alpha$,*
- *or $\alpha > \frac{1}{2}$ and $\frac{1}{2} \leq \varrho \leq 1$,*

which are asymptotically product with density ϱ, such that

$$if \, \alpha \leq \tfrac{1}{2} \, then \, \lim_{t \to \infty} \mu S(t) = \begin{cases} \nu_\alpha & if \, \rho \leq 1 - \alpha \\ \mu_\rho^\alpha & if \, \rho > 1 - \alpha, \end{cases}$$

$$and \, if \, \alpha > \tfrac{1}{2} \, then \, \lim_{t \to \infty} \mu S(t) = \begin{cases} \mu_{1/2}^\alpha & if \, \rho \leq \tfrac{1}{2} \\ \mu_\rho^\alpha & if \, \rho > \tfrac{1}{2}. \end{cases}$$

Observe the notation we adopt throughout this paper: The *initial* particle density for the TASEP is denoted ρ, whereas ϱ is the *stationary* particle density under μ_ϱ^α. Convergence in Theorem 2.1 is not uniform and slows down the further sites are from the origin. The initial density ρ determines the stationary measure the process will converge to. In particular, starting from all sites empty, if the injection rate satisfies $\alpha \leq \frac{1}{2}$, the distribution of the process converges to the product measure with constant density α. If $\alpha > \frac{1}{2}$, the distribution of the process converges to $\mu_{\frac{1}{2}}^\alpha$, which has spatial correlations and an asymptotic density equal to $\frac{1}{2}$. Observe that the injection mechanism is not able to produce a stationary particle density larger than $\frac{1}{2}$ unless there is initially a high density of particles in the system.

We will explore matrix representations of the measures μ_ϱ^α in Section 3.

2.2 Main Result

Our problem at hand is to find the rate function for a large deviation principle of the empirical density on the first n sites

$$Z_n = \frac{1}{n} \sum_{k=1}^{n} \eta_k \tag{2}$$

under the stationary measure, as n goes to infinity. Recall from Theorem 2.1 that the answer to this problem depends in a subtle way on the spatial correlations occurring in the case $\alpha > \frac{1}{2}$.

The theory of large deviations analyses the exponential decay of probabilities of increasingly unlikely events. Formally, a sequence of random variables $\{Z_n\}_{n \in \mathbb{N}}$ taking values in

[0, 1] satisfies a large deviation principle with rate function $I : [0, 1] \to [0, \infty]$ under the probability measure $\mathbb{P}$ if

(i)   the function I is lower semicontinuous,
(ii) for all open sets $G \subset [0, 1]$ we have

$$\liminf_{n \to \infty} \frac{1}{n} \log \mathbb{P}\{Z_n \in G\} \geq - \inf_{z \in G} I(z),$$

(iii) and for all closed sets $F \subset [0, 1]$ we have

$$\limsup_{n \to \infty} \frac{1}{n} \log \mathbb{P}\{Z_n \in F\} \leq - \inf_{z \in F} I(z).$$

The main result of this paper is the following large deviation principle.

Theorem 2.2 *Let $\{Z_n\}_{n \in \mathbb{N}}$ be the sequence of random variables defined as the empirical density* (2) *of a semi-infinite TASEP with injection rate $\alpha \in (0, 1)$ and initial asymptotically product measure for which ρ as defined in Theorem 2.1 exists. Then, under the stationary probability measure given by Theorem 2.1, $\{Z_n\}_{n \in \mathbb{N}}$ satisfies a large deviation principle with convex rate function $I : [0, 1] \to [0, \infty]$ given as follows.*

(a) *If $\alpha \leq \frac{1}{2}$ and $\rho < 1 - \alpha$, then*

$$I(z) = z \log \frac{z}{\alpha} + (1 - z) \log \frac{1 - z}{1 - \alpha}.$$

(b) *If $\alpha > \frac{1}{2}$ and $0 \leq \rho \leq \frac{1}{2}$, then*

$$I(z) = \begin{cases} z \log \dfrac{z}{\alpha} + (1 - z) \log \dfrac{1 - z}{1 - \alpha} + \log(4\alpha(1 - \alpha)) & \text{if } 0 \leq z \leq 1 - \alpha, \\ 2\left[z \log z + (1 - z) \log(1 - z) + \log 2\right] & \text{if } 1 - \alpha < z \leq \frac{1}{2}, \\ z \log z + (1 - z) \log(1 - z) + \log 2 & \text{if } \frac{1}{2} < z \leq 1. \end{cases}$$

(c) *If $\alpha > \frac{1}{2}$ and $\frac{1}{2} < \rho < \alpha$, then*

$$I(z) = \begin{cases} z \log \dfrac{z}{\alpha} + (1 - z) \log \dfrac{1 - z}{1 - \alpha} + \log \dfrac{\alpha(1 - \alpha)}{\rho(1 - \rho)} & \text{if } 0 \leq z \leq 1 - \alpha, \\ 2\left[z \log z + (1 - z) \log(1 - z) - \log \sqrt{\rho(1 - \rho)}\right] & \text{if } 1 - \alpha < z \leq 1 - \rho, \\ z \log \dfrac{z}{\rho} + (1 - z) \log \dfrac{1 - z}{1 - \rho} & \text{if } 1 - \rho < z \leq 1. \end{cases}$$

Recall that part (a) is well-known and included for completeness. It implies the weak law of large numbers, saying that the empirical density converges in probability to α if $\alpha \leq \frac{1}{2}$, see Fig. 1. The unique zero of the rate function moves from 0 to $\frac{1}{2}$ with the value of α, at the same time it is getting easier to achieve any given density larger than $\frac{1}{2}$.

Part (b) shows that the empirical density converges to $\frac{1}{2}$ if $\alpha > \frac{1}{2}$ and $\rho < \frac{1}{2}$. Looking at Fig. 1 we see that the rate function is still convex and its zero stays fixed at $\frac{1}{2}$. Now reaching high densities has always the same cost regardless of the value of α, but low densities become increasingly expensive as the value of α increases. Note that the rate function is non-analytic at the value $z = 1 - \alpha$, which reveals a dynamical phase transition in the sense of [17].
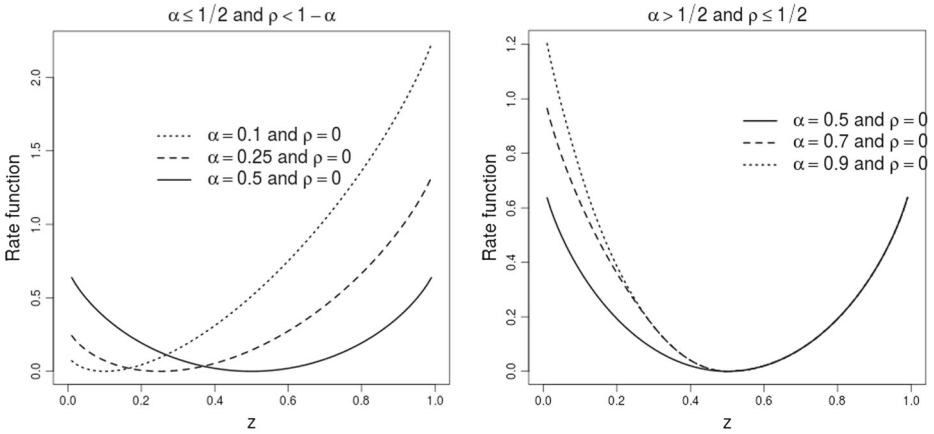
Fig. 1 Rate functions for case *(a)* and *(b)* of Theorem 2.2

For part (c), the minimum of the rate function is now at ρ. Low densities still become increasingly expensive as α increases; yet, high densities now become cheaper, see the left diagram of Fig. 2. In this case, we observe that both α and ρ play a role in the rate function, see the right diagram of Fig. 2 to appreciate the joint effect. The phase transitions are seen at $z = 1 - \alpha$, as in the previous case, and $z = 1 - \rho$. As $\rho \to \alpha$ we recover the rate function of Bernoulli product measures.

In parts (b) and (c), the cost in the first regime, when $z \leq 1 - \alpha$, is up to a shift by a negative constant equal to the cost of changing the boundary density. The rates in the third regime, when $1 - z$ is smaller than the typical density, represent the cost of replacing the typical density in the asymptotic Bernoulli product measure by the desired value, so the cost in this regime is bulk dominated. In the second regime the cost is larger than in the third, indicating that both a bulk and a boundary cost have to be paid.

The regimes when $\alpha \leq \frac{1}{2}$ and $\rho \geq 1 - \alpha$, or when $\alpha > \frac{1}{2}$ and $\rho \geq \alpha$ are not covered by the techniques of this paper and remain open, see Fig. 3. In the former case we have no matrix product representation.
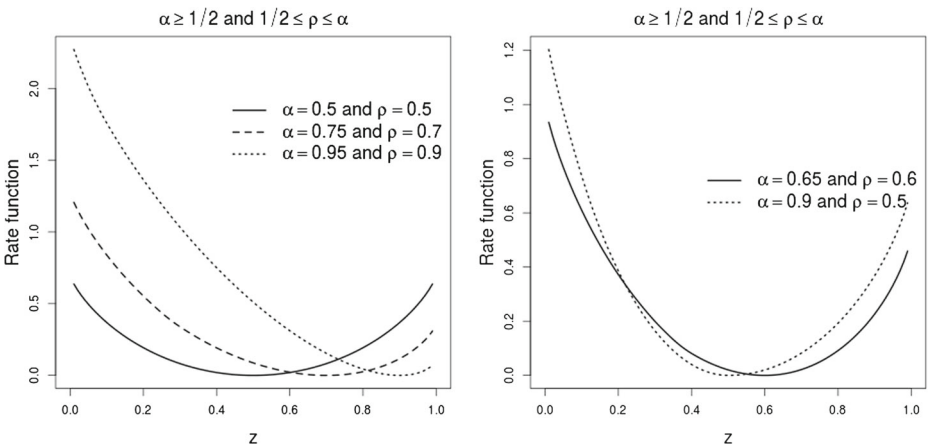

Fig. 2 Rate functions for case *(c)* of Theorem 2.2
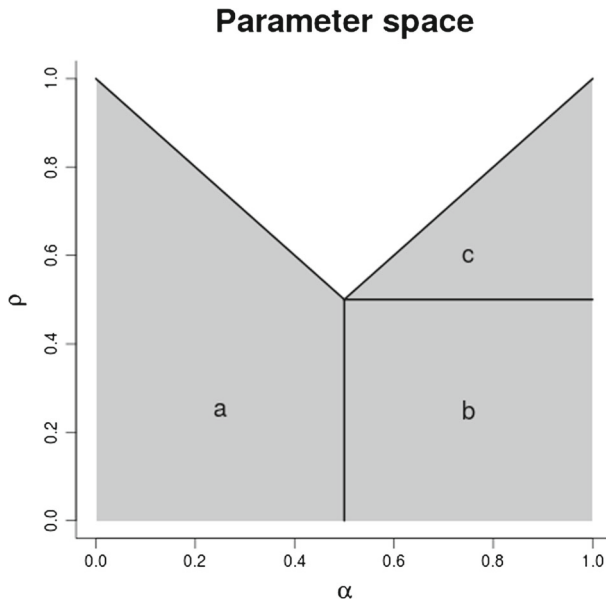
Parameter space

Fig. 3 The range of validity of Theorem 2.2 in parameter space is shaded in *grey*

If $\alpha > \frac{1}{2}$ and $\rho < \alpha$ it is worth comparing our large deviation result for the semi-infinite TASEP with that for the finite TASEP studied in [12]. The stationary measure on the semi-infinite TASEP can be obtained as a limit of the stationary measures on the finite TASEP with n sites and boundary densities chosen as α on the left boundary, and $\min\{1 - \rho, \frac{1}{2}\}$ on the right boundary, see [16, Section 3]. However, the transition rates across bonds in the semi-infinite TASEP are not equal to $\min\{1 - \rho, \frac{1}{2}\}$. It is therefore somewhat surprising that the large deviation rate of the sequence of densities in the finite TASEP with increasing system size obtained in [12, (3.12)] still agrees with the rate we have obtained for the average density over increasing blocks in the infinite TASEP at stationarity.

3 Matrix Product Ansatz

Grosskinsky [14], following the seminal work of [11], has given a characterisation of the long range dependence in μ_ϱ^α with a matrix product ansatz.

Theorem 3.1 [14, Theorem 3.2] *Suppose there exist (possibly infinite) nonnegative matrices D, E and vectors w and v, fulfilling the algebraic relations*

$$DE = D + E, \tag{3a}$$
$$\alpha w^\mathsf{T} E = w^\mathsf{T}, \tag{3b}$$
$$c(D + E)v = v, \tag{3c}$$

for some c > 0. Then

(a)   *the probability measure $\bar{v}_c^\alpha$ defined by*

$$\bar{v}_c^\alpha\{\zeta \in \{0, 1\}^\mathbb{N} : \zeta_1 = \eta_1, \ldots, \zeta_n = \eta_n\} = \frac{w^\mathsf{T} \left(\prod_{k=1}^n \eta_k D + (1 - \eta_k)E\right)v}{w^\mathsf{T}(D + E)^n v} \quad (4)$$

is invariant for the generator (1) *if and only if*

- *either* $\alpha \leq \frac{1}{2}$ *and* $0 \leq c \leq \alpha(1 - \alpha)$
- *or* $\alpha > \frac{1}{2}$ *and* $0 \leq c \leq \frac{1}{4}$.

(b) *The measure* $\bar{v}_c^\alpha$ *has stationary current* $\mathbb{E}_{\bar{v}_c^\alpha}[\eta_k(1 - \eta_{k+1})] = c$, *for all* $k \geq 1$. *It equals* v_α *if* $c = \alpha(1 - \alpha)$ *and* $\alpha \leq \frac{1}{2}$, *and otherwise it is asymptotically product with density* ϱ *given as the solution of* $c = \varrho(1 - \varrho)$ *which satisfies* $\varrho \geq \frac{1}{2}$.

Very often, to apply Theorem 3.1, no explicit solution of Eq. 3 is needed. Below we only use the recursive structure of these equations to show that the measures $\bar{v}_c^\alpha$ and μ_ϱ^α agree under certain conditions. Note that this does not follow directly from Theorem 2.1 as this result does not describe the long-term behaviour of the TASEP started in $\bar{v}_c^\alpha$, which is not necessarily a product measure.

Proposition 3.2 *If* $\alpha \geq \frac{1}{2}$, $\varrho \geq \frac{1}{2}$ *and* $c = \varrho(1 - \varrho)$, *then the measures* $\bar{v}_c^\alpha$ *and* μ_ϱ^α *agree.*

Proof By part (e) in [16, Theorem 3.10] the measure μ_ϱ^α is uniquely determined by the following two properties, numbered as in [16],

(c) If $u, n \in \mathbb{N}$ with $1 < u < u + 1 < n$, and $\eta \in \{0, 1\}^n$ with $\eta_u = 1$, $\eta_{u+1} = 0$, then

$$\mu_\varrho^\alpha\{\zeta : \zeta_k = \eta_k \text{ for } k \leq n\}$$
$$= c \mu_\varrho^\alpha\{\zeta : \zeta_k = \eta_k \text{ for } k \leq u - 1, \zeta_k = \eta_{k+1} \text{ for } u + 1 \leq k \leq n - 1\}.$$

(d) If $n > 1$ and $\eta \in \{0, 1\}^n$ with $\eta_1 = 0$, then

$$\alpha \mu_\varrho^\alpha\{\zeta : \zeta_k = \eta_k \text{ for } k \leq n\} = c \mu_\varrho^\alpha\{\zeta : \zeta_k = \eta_{k+1} \text{ for } k \leq n - 1\}.$$

We show that $\bar{v}_c^\alpha$ satisfies these properties. Under the assumptions of (c) we get from properties (3a) in the second equality and Eq. 3c in the third one

$$\bar{v}_c^\alpha\{\zeta : \zeta_k = \eta_k \text{ for } k \leq n\}$$
$$= \frac{w^\mathsf{T} \left(\prod_{k=1}^{u-1} \eta_k D + (1 - \eta_k)E\right) DE \left(\prod_{k=u+2}^n \eta_k D + (1 - \eta_k)E\right)v}{w^\mathsf{T}(D + E)^n v}$$
$$= \frac{w^\mathsf{T} \left(\prod_{k=1}^{u-1} \eta_k D + (1 - \eta_k)E\right) (D + E) \left(\prod_{k=u+2}^n \eta_k D + (1 - \eta_k)E\right)v}{w^\mathsf{T}(D + E)^n v}$$
$$= c \bar{v}_c^\alpha\{\zeta : \zeta_k = \eta_k \text{ for } k \leq u - 1, \zeta_k = \eta_{k+1} \text{ for } u + 1 \leq k \leq n - 1\}.$$

Under the assumptions of (d) we get from conditions (3b) in the second equality and Eq. 3c in the third one,

$$\alpha \bar{v}_c^\alpha\{\zeta : \zeta_k = \eta_k \text{ for } k \leq n\} = \alpha \frac{w^\mathsf{T} E \left(\prod_{k=2}^n \eta_k D + (1 - \eta_k)E\right)v}{w^\mathsf{T}(D + E)^n v}$$
$$= \frac{w^\mathsf{T} \left(\prod_{k=2}^n \eta_k D + (1 - \eta_k)E\right)v}{w^\mathsf{T}(D + E)^n v}$$
$$= c \bar{v}_c^\alpha\{\zeta : \zeta_k = \eta_{k+1} \text{ for } k \leq n - 1\}.$$

Hence $\bar{v}_c^\alpha$ satisfies (c) and (d) and therefore agrees with μ_ϱ^α. $\square$

We now explain our approach to find a large deviation principle of the empirical density under this measure. We will approach this via the Gärtner-Ellis theorem, see Theorem V.6 in [9]. Here we state the conditions specific for our case.

Theorem 3.3 *(Gärtner-Ellis) Let $\{Z_n\}_{n\in\mathbb{N}}$ be a sequence of random variables on a probability space $(\Omega, \mathcal{A}, \mathbb{P})$, where Ω is a non-empty subset of $\mathbb{R}$. If the limit cumulant generating function $\Lambda\colon \mathbb{R} \to \mathbb{R}$ defined by*

$$\Lambda(\theta) = \lim_{n\to\infty} \frac{1}{n} \log \mathbb{E}[e^{n\theta Z_n}]$$

exists and is differentiable on all $\mathbb{R}$, then $\{Z_n\}_{n\in\mathbb{N}}$ satisfies a large deviation principle with rate function $I\colon \Omega \to [-\infty, \infty]$ defined by

$$I(z) = \sup_{\theta\in\mathbb{R}}\{z\theta - \Lambda(\theta)\}.$$

To calculate the moment generating function $M_n(\theta)$ of Z_n we use Theorem 3.1 and Proposition 3.2 in the third equality and condition (3c) in the fifth one, to get

$$M_n(\theta) = v\mathbb{E}\left[e^{n\theta Z_n}\right] = \mathbb{E}\left[\exp\left(\theta \sum_{k=1}^{n} \xi_k\right)\right]$$

$$= \sum_{\eta\in\{0,1\}^n} \bar{v}_c^{\alpha}\{\xi : \xi_k = \eta_k \text{ for } k \le n\} \exp\left(\theta \sum_{k=1}^{n} \eta_k\right)$$

$$= \frac{w^{\mathsf{T}}(e^\theta D + E)^n v}{w^{\mathsf{T}}(D + E)^n v} = \frac{c^n}{w^{\mathsf{T}}v} w^{\mathsf{T}}(e^\theta D + E)^n v,$$

and the cumulant generating function simplifies to

$$\Lambda(\theta) = \lim_{n\to\infty} \frac{1}{n} \log M_n(\theta)$$

$$= \lim_{n\to\infty} \frac{1}{n} \log w^{\mathsf{T}}(e^\theta D + E)^n v + \log c. \tag{5}$$

If D and E were finite matrices, we could identify this limit using the Perron-Frobenius theorem as the spectral radius of the matrix $e^\theta D + E$. However in our example (and in almost all physically interesting examples) the matrices solving Eq. 3 are necessarily infinite. A first idea would be to truncate the matrices to finite size, calculate the spectral radius and taking a limit, but this turns out to lead to a wrong result, as it neglects the important information contained in the vectors v and w.

We will look at lower and upper bounds in Eq. 5 separately. For the *upper bound* we exploit that matrices D and E, as well as the vectors v and w, solving Eq. 3 are explicitly known. We introduce weighted ℓ^2 spaces, denoted ℓ_s^2, and interpret the matrix $e^\theta D + E$ as an operator on these spaces. If the weights are such that v is an element of ℓ_s^2, and w^{T} an element of its dual, we can get a bound on Eq. 5 from the spectral radius of the operator, which can be optimised by minimising the bound over all admissible weights. In order to obtain the spectral radius we use a simple isomorphism between weighted and unweighted ℓ^2 spaces. Acting on the unweighted spaces, the operator has a Toeplitz structure and from the general theory of Toeplitz operators on ℓ^2 an explicit formula for the spectral radius is available. This estimate will be carried out in detail in Section 4.

The technique for the *lower bound* only relies on the structure of the equations (3). These provide an algorithmic way to reduce arbitrary products of D and E to linear combinations

of monomials of the form $E^{p-j}D^j$ with $0 \leq j \leq p$. We expand the product $(e^{\theta}D + E)^n$ into a polynomial with nonnegative coefficients $f^n_{p,j}(\theta)$, and use Eq. 3 to derive a recursion formula for the coefficients. While it seems to be too complicated to fully resolve this recursion, we focus on selected key terms which can be derived explicitly. Note that for a lower bound we can drop all inaccessible terms in the expansion. In our case, to obtain the growth rate it suffices to include the fastest growing terms $f^n_{p,p}D^p$ and $f^n_{p,0}E^p$ corresponding to all summands in the product which can be reduced to monomials of just one variable. It is quite a typical phenomenon that only a small number of boundary summands contribute to the growth of the sum, and that these coefficients can be identified without solving the entire system of equations. This estimate will be carried out explicitly in Section 5.

4 Upper Bound: Spectral Theory of Toeplitz Operators

4.1 Weighted l^2 Spaces

To find an upper bound for the cumulant generating function we consider the weighted spaces

$$\ell^2_s = \left\{ x = (x_k)_{k \in \mathbb{N}} : \sum_{k=1}^{\infty} |x_k|^2 s^k < \infty \right\}$$

with $s > 0$. Note that imposing $s = 1$ recovers the usual ℓ^2 space. Moreover for fixed s, ℓ^2_s with its corresponding norm

$$|x|^2_{\ell^2_s} = \sum_{k=1}^{\infty} |x_k|^2 s^k$$

is a Banach space. The next lemma will help us to translate classic ℓ^2 theory to ℓ^2_s.

Lemma 4.1 *The function* $T_s : \ell^2 \to \ell^2_s$ *defined by*

$$(T_s x)_k = \frac{x_k}{s^{k/2}}$$

for $s > 0$ is a bijective isometry.

Proof We can define the inverse $T_s^{-1} : \ell^2_s \to \ell^2$ by $(T_s^{-1}x)_k = x_k s^{k/2}$ and hence T_s is bijective. We just need to prove it is an isometry, so let $x \in \ell^2_s$ and calculate

$$|T_s^{-1}x|^2_{\ell^2} = \sum_{k=1}^{\infty} |(T_s^{-1}x)_k|^2 = \sum_{k=1}^{\infty} |x_k s^{k/2}|^2 = |x|^2_{\ell^2_s}.$$

Analogously, for $x \in \ell^2$ we have $|T_s x|_{\ell^2_s} = |x|_{\ell^2}$. $\square$

Lemma 4.2 *The dual space* ℓ^{2*}_s *can be identified with* $\ell^2_{s^{-1}}$.

Proof Define the dual product $\langle \cdot, \cdot \rangle_D : \ell^2_{s^{-1}} \times \ell^2_s \to \mathbb{R}$ by $\langle y, x \rangle_D = \langle T_{s^{-1}}^{-1}y, T_s^{-1}x \rangle$, where $\langle \cdot, \cdot \rangle$ is the usual inner product in ℓ^2. We first prove that for each vector $y \in \ell^2_{s^{-1}}$ there exists a function $f_y \in \ell^{2*}_s$ such that $f_y(x) = \langle y, x \rangle_D$. To this end, let $y \in \ell^2_{s^{-1}}$ and define $f_y : \ell^2_s \to \mathbb{R}$ by

$$f_y(x) = \langle y, x \rangle_D = \sum_{k \in \mathbb{N}} x_k y_k.$$

The linearity of f_y follows easily from the definition; the Cauchy-Schwarz inequality in ℓ^2 shows it is also bounded,

$$
\begin{aligned}
|f_y(x)| = \left| \sum_{k \in \mathbb{N}} x_k y_k \right| &= \left| \sum_{k \in \mathbb{N}} x_k s^{\frac{k}{2}} y_k s^{-\frac{k}{2}} \right| \\
&\leq \left(\sum_{k \in \mathbb{N}} |x_k|^2 s^k \right)^{\frac{1}{2}} \left(\sum_{k \in \mathbb{N}} |y_k|^2 s^{-k} \right)^{\frac{1}{2}} = |x|_{\ell_s^2} |y|_{\ell_{s^{-1}}^2}.
\end{aligned}
$$

Conversely, let $f \in \ell_s^{2*}$. Define $g \colon \ell^2 \to \mathbb{R}$ by $g(x) = (f \circ T_s)(x)$. Since f and T_s are both linear, so is g, and since f is bounded,

$$
|g(x)| = |(f \circ T_s)(x)| \leq |f|_{\ell_s^{2*}} |T_s(x)|_{\ell_s^2} = |f|_{\ell_s^{2*}} |x|_{\ell^2} < \infty.
$$

Hence, $g \in \ell^{2*}$ and by the Riesz Representation theorem there exists a unique $\tilde{y} \in \ell^2$ such that $g(x) = \langle x, \tilde{y} \rangle$ for all $x \in \ell^2$. Let $y = T_{s^{-1}} \tilde{y} \in \ell_{s^{-1}}^2$. Since T_s is invertible we have that for all $x \in \ell_s^2$

$$
f(x) = (g \circ T_s^{-1})(x) = \langle T_s^{-1} x, \tilde{y} \rangle = \sum_{k \in \mathbb{N}} (T_s^{-1} x)_k (T_{s^{-1}}^{-1} y)_k = \sum_{k \in \mathbb{N}} x_k y_k = \langle y, x \rangle_D,
$$

whence $f \in \ell_s^{2*}$ is represented by $y \in \ell_{s^{-1}}^2$. □

We now need an explicit solution for Eq. 3, so we first define the values

$$
\lambda_1 = \frac{1 - 2c + \sqrt{1 - 4c}}{2c}, \tag{6}
$$

$$
\lambda_2 = \frac{1 - 2c - \sqrt{1 - 4c}}{2c}. \tag{7}
$$

Elementary calculations show that the matrices D, E and the vectors v and w defined by

$$
D = \begin{pmatrix} 1 & 1 & 0 & 0 & \cdots \\ 0 & 1 & 1 & 0 & \cdots \\ 0 & 0 & 1 & 1 & \cdots \\ 0 & 0 & 0 & 1 & \ddots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix}, \qquad E = \begin{pmatrix} 1 & 0 & 0 & 0 & \cdots \\ 1 & 1 & 0 & 0 & \cdots \\ 0 & 1 & 1 & 0 & \cdots \\ 0 & 0 & 1 & 1 & \ddots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix}, \tag{8a}
$$

$$
w^{\mathsf{T}} = \left(1, \frac{1}{\alpha} - 1, \left(\frac{1}{\alpha} - 1 \right)^2, \cdots \right) \quad \text{and} \quad v = \frac{1}{\lambda_1 - \lambda_2} \begin{pmatrix} \lambda_1 - \lambda_2 \\ \lambda_1^2 - \lambda_2^2 \\ \lambda_1^3 - \lambda_2^3 \\ \vdots \end{pmatrix} \tag{8b}
$$

satisfy the matrix product conditions (3). In the boundary case $c = \frac{1}{4}$ we have $\lambda_1 = 1 = \lambda_2$ and we take $v^{\mathsf{T}} = (1, 2, 3, \ldots)$.

To simplify notation, define for fixed $\theta \in \mathbb{R}$ the operator $A(\theta) \colon \ell_s^2 \to \ell_s^2$ with the infinite matrix representation

$$A(\theta) = e^{\theta} D + E = \begin{pmatrix} 1 + e^{\theta} & e^{\theta} & 0 & 0 & \cdots \\ 1 & 1 + e^{\theta} & e^{\theta} & 0 & \cdots \\ 0 & 1 & 1 + e^{\theta} & e^{\theta} & \cdots \\ 0 & 0 & 1 & 1 + e^{\theta} & \ddots \\ \vdots & \vdots & \vdots & \ddots & \ddots \end{pmatrix} \tag{9}$$

and then the k-th component of the vector $A(\theta)x$ is

$$(A(\theta)x)_k = \begin{cases} x_1(1 + e^{\theta}) + x_2 e^{\theta} & \text{if } k = 1 \\ x_{k-1} + x_k(1 + e^{\theta}) + x_{k+1} e^{\theta} & \text{if } k > 1. \end{cases}$$

Proposition 4.3 *The operator* $A(\theta) \colon \ell_s^2 \to \ell_s^2$ *is bounded.*

Proof Let $x \in \ell_s^2$. Using Cauchy-Schwarz in $\mathbb{R}^2$ and $\mathbb{R}^3$ for each term of $A(\theta)x$ gives

$$|A(\theta)x|_{\ell_s^2}^2 = \sum_{k=1}^{n} |(A(\theta)x)_k|^2 s^k$$

$$= |x_1(1 + e^{\theta}) + x_2 e^{\theta}|^2 s + \sum_{k=2}^{n} |x_{k-1} + x_k(1 + e^{\theta}) + x_{k+1} e^{\theta}|^2 s^k$$

$$\leq (x_1^2 + x_2^2)((1 + e^{\theta})^2 + e^{2\theta}) + \sum_{k=2}^{n} (x_{k-1}^2 + x_k^2 + x_{k+1}^2)(1 + (1 + e^{\theta})^2 + e^{2\theta}) s^k$$

$$\leq C_s |x|_{\ell_s^2}^2,$$

where $C_s > 0$ is a constant independent of x and hence we see that $A(\theta)$ is a bounded linear operator. $\qquad \square$

Lemma 4.4 *Let* $L \in \mathcal{L}(\ell_s^2)$, *that is a bounded linear operator from* ℓ_s^2 *to itself. The operator* $\tilde{L} = T_s^{-1} \circ L \circ T_s$ *satisfies* $\tilde{L} \in \mathcal{L}(\ell^2)$.

Proof Take $x \in \ell^2$. Then by Lemma 4.1,

$$|\tilde{L}x|_{\ell^2} \leq |T_s^{-1}|_{\mathcal{L}(\ell_s^2, \ell^2)} |L|_{\mathcal{L}(\ell_s^2)} |T_s|_{\mathcal{L}(\ell^2, \ell_s^2)} |x|_{\ell^2} < \infty.$$

By Lemma 4.1 we conclude that $|\tilde{L}|_{\mathcal{L}(\ell^2)} \leq |L|_{\mathcal{L}(\ell_s^2)}$. Analogously, since $L = T_s \circ L \circ T_s^{-1}$, we have that $|\tilde{L}|_{\mathcal{L}(\ell^2)} = |L|_{\mathcal{L}(\ell_s^2)}$. $\qquad \square$

The tilde operator commutes with exponentiation.

Lemma 4.5 *Let* $L \in \mathcal{L}(\ell_s^2)$, *then* $\widetilde{L^n} = \tilde{L}^n$.

Proof We proceed by induction over n. For $n = 1$, the proposition is a tautology. We assume the proposition true for n, let $x \in \ell^2$ and calculate

$$\tilde{L}^{n+1} x = \tilde{L} \circ \tilde{L}^n x = T_s^{-1} \circ L \circ T_s \circ T_s^{-1} \circ L^n \circ T_s x = T_s^{-1} \circ L^{n+1} \circ T_s x = \widetilde{L^{n+1}} x.$$

$\qquad \square$

Recall from Eq. 8 the explicit form of w and v and note that if $s \in (0, 1)$

$$|v|_{\ell_s^2}^2 = \sum_{k \in \mathbb{N}} k^2 s^k = \frac{s(1+s)}{(1-s)^3} < \infty. \tag{10}$$

On the other hand, if $s > \left(\frac{1}{\alpha} - 1\right)^2$,

$$|w|_{\ell_{s^{-1}}^2}^2 = \sum_{k \in \mathbb{N}} \left(\frac{1}{\alpha} - 1\right)^{2(k-1)} s^{-k} = \frac{1}{s - \left(\frac{1}{\alpha} - 1\right)^2} < \infty. \tag{11}$$

Therefore, for $s \in \left((\frac{1}{\alpha} - 1)^2, 1\right)$ we have that $v \in \ell_s^2$ and $w \in \ell_{s^{-1}}^2$.

4.2 Toeplitz Operators

Before stating the main result of this section, we need to review some properties of Toeplitz operators. Let $a = \{a_k\}_{k \in \mathbb{Z}} \in \ell^2(\mathbb{C})$, that is, a double sequence of complex numbers such that $\sum_{k \in \mathbb{Z}} |a_k|^2 < \infty$. A Toeplitz operator A defined by the double sequence $a \in \ell^2(\mathbb{C})$ is an infinite matrix with the structure

$$A = \begin{pmatrix} a_0 & a_{-1} & a_{-2} & \cdots \\ a_1 & a_0 & a_{-1} & \cdots \\ a_2 & a_1 & a_0 & \cdots \\ \vdots & \vdots & \ddots & \ddots \end{pmatrix}.$$

The *symbol* $\kappa : \{z \in \mathbb{C} : |z| = 1\} \to \mathbb{C}$ of a Toeplitz operator is defined by

$$\kappa(z) = \sum_{k \in \mathbb{Z}} a_k z^k.$$

We recall Theorem 7.1 in [22] that deals with spectra of Toeplitz operators.

Theorem 4.6 [22, Theorem 7.1] *Let A be a Toeplitz operator. If A has a continuous symbol κ, then its spectrum is given by the image of the unit circle under κ together with all the points enclosed by this curve with non-zero winding number.*

For fixed $\theta \in \mathbb{R}$, the operator $A(\theta)$ defined by Eq. 9 is by Proposition 4.3 in $\mathcal{L}(\ell_s^2)$. By Lemma 4.4 the operator $\tilde{A}(\theta)$ is a Toeplitz operator in ℓ^2 with its symbol κ given by

$$\kappa(\zeta) = \frac{e^\theta}{\zeta \sqrt{s}} + 1 + e^\theta + \zeta \sqrt{s}.$$

Writing $\zeta = e^{i\varphi}$ as an element of the unit circle,

$$\kappa(e^{i\varphi}) = 1 + e^\theta + \left(\sqrt{s} + \frac{e^\theta}{\sqrt{s}}\right) \cos\varphi + \left(\sqrt{s} - \frac{e^\theta}{\sqrt{s}}\right) i \sin\varphi,$$

which we recognise as a parametrised ellipse centred at $1 + e^\theta$, with major axis of length $\sqrt{s} + \frac{e^\theta}{\sqrt{s}}$ along the real line, and minor axis of length $|\sqrt{s} - \frac{e^\theta}{\sqrt{s}}|$. Therefore, the spectral radius is found when $z = 1$ and

$$\rho(\tilde{A}(\theta)) = \kappa(1) = 1 + \sqrt{s} + e^\theta\left(1 + \frac{1}{\sqrt{s}}\right). \tag{12}$$

We now state the main result of this section: the upper bound for the cumulant generating function Λ.

Proposition 4.7 *For Λ defined by Eq. 5, an upper bound is*

$$\Lambda(\theta) \leq \begin{cases} \log\left(\frac{e^{\theta}}{1-\alpha} + \frac{1}{\alpha}\right) + \log c & \text{if } -\infty < \theta \leq 2\log\left(\frac{1}{\alpha} - 1\right), \\ \log\left(1 + e^{\theta/2}\right)^2 + \log c & \text{if } 2\log\left(\frac{1}{\alpha} - 1\right) < \theta \leq -2\log\lambda_1, \\ \log\left(1 + \lambda_1 e^{\theta}\right) + \log\left(1 + \frac{1}{\lambda_1}\right) + \log c & \text{if } -2\log\lambda_1 < \theta < \infty. \end{cases}$$

Proof Recall from Eq. 10 that $v \in \ell_s^2$; by Proposition 4.3 $A(\theta) \in \mathcal{L}(\ell_s^2)$ and therefore $A(\theta)v \in \ell_s^2$. Also $w \in \ell_{s-1}^2$ from Eq. 11. Hence, by Eq. 5 and Cauchy-Schwarz,

$$\Lambda(\theta) = \lim_{n \to \infty} \frac{1}{n} \log(w^{\mathsf{T}} A(\theta)^n v) + \log c$$

$$\leq \lim_{n \to \infty} \frac{1}{n} \log\left(|w|_{\ell_{s-1}^2} |A(\theta)^n v|_{\ell_s^2}\right) + \log c.$$

The norm of w does not contribute to the limit since it does not depend on n. By Lemma 4.4, Lemma 4.1, and Lemma 4.5 we can continue the previous estimate

$$= \lim_{n \to \infty} \frac{1}{n} \log(|T_s \circ \widetilde{A(\theta)^n} \circ T_s^{-1} v|_{\ell^2}) + \log c$$

$$\leq \lim_{n \to \infty} \frac{1}{n} \log(|\tilde{A}(\theta)^n|_{\mathcal{L}(\ell^2)} |T_s^{-1} v|_{\ell^2}) + \log c;$$

once again, the norm of $T_s^{-1} v$ does not contribute to the limit since it does not depend on n. We insert the factor $\frac{1}{n}$ to the logarithm by continuity and use the definition of spectral radius

$$= \lim_{n \to \infty} \log\left(|\tilde{A}(\theta)^n|_{\mathcal{L}(\ell^2)}^{\frac{1}{n}}\right) + \log c$$

$$= \log \rho(\tilde{A}(\theta)) + \log c.$$

We now use Eq. 12 to find the spectral radius

$$= \log\left[1 + \sqrt{s} + e^{\theta}\left(1 + \frac{1}{\sqrt{s}}\right)\right] + \log c.$$

Since the left hand side does not depend on s, it is a lower bound on the right hand side for s, so we take the infimum over the interval $\left((\frac{1}{\alpha} - 1)^2, \frac{1}{\lambda_1^2}\right)$.

$$\Lambda(\theta) \leq \inf_{s \in \left(\left(\frac{1}{\alpha} - 1\right)^2, \frac{1}{\lambda_1^2}\right)} \log\left[1 + \sqrt{s} + e^{\theta}\left(1 + \frac{1}{\sqrt{s}}\right)\right] + \log c.$$

Given θ, the value of s that reaches the infimum of this function is given by

$$s^* = \begin{cases} \left(\frac{1}{\alpha} - 1\right)^2 & \text{if } -\infty < \theta \leq 2\log\left(\frac{1}{\alpha} - 1\right), \\ e^{\theta} & \text{if } \log\left(\frac{1}{\alpha} - 1\right)^2 < \theta \leq -2\log\lambda_1, \\ \frac{1}{\lambda_1^2} & \text{if } -2\log\lambda_1 < \theta < \infty. \end{cases}$$

Plugging s^* into the formula gives the result of the lemma. $\square$

5 Lower Bound: A Combinatorial Approach

In order to find the lower bound we use a completely different approach. We will focus on the powers of $e^\theta D + E$.

Proposition 5.1 *There exists a sequence of polynomials* $f_{p,j}^n(\theta)$ *in* e^θ *such that*

$$(e^\theta D + E)^n = \sum_{p=1}^n \sum_{j=0}^p f_{p,j}^n(\theta) E^{p-j} D^j \tag{13}$$

and they can be defined recursively in two ways: Starting with

$$f_{1,0}^1(\theta) = 1, \qquad f_{1,1}^1(\theta) = e^\theta,$$

the first characterisation for $f_{p,j}^n$ *with* $n > 1$ *is*

$$f_{p,j}^n(\theta) = \begin{cases} \sum_{k=1}^{n-1} f_{k,k}^{n-1}(\theta) & \text{if } n > 1, \ p = 1, \qquad j \in \{0,1\} \\ f_{p-1,0}^{n-1}(\theta) + \sum_{k=p}^{n-1} f_{k,k-p+1}^{n-1}(\theta) & \text{if } n > 1, \ 1 < p < n, \ j = 0 \\ e^\theta f_{p-1,j-1}^{n-1}(\theta) + \sum_{k=p}^{n-1} f_{k,k-p+j}^{n-1}(\theta) & \text{if } n > 1, \ j \leq p < n, \ j > 0 \\ f_{n-1,0}^{n-1}(\theta) & \text{if } n > 1, \ p = n, \qquad j = 0 \\ e^\theta f_{n-1,j-1}^{n-1}(\theta) & \text{if } n > 1, \ p = n, \qquad 0 < j \leq p; \end{cases} \tag{14}$$

and the second characterisation is

$$f_{p,j}^n(\theta) = \begin{cases} \sum_{k=1}^{n-1} e^\theta f_{k,0}^{n-1}(\theta) & \text{if } n > 1, \ p = 1, \qquad j \in \{0,1\} \\ f_{p-1,j}^{n-1}(\theta) + \sum_{k=p}^{n-1} e^\theta f_{k,j}^{n-1}(\theta) & \text{if } n > 1, \ 1 < p < n, \ 0 \leq j < p \\ e^\theta f_{p-1,p-1}^{n-1}(\theta) + \sum_{k=p}^{n-1} e^\theta f_{k,p-1}^{n-1}(\theta) & \text{if } n > 1, \ 1 < p < n, \ j = p \\ f_{n-1,j}^{n-1}(\theta) & \text{if } n > 1, \ p = n, \qquad 0 \leq j < n \\ e^\theta f_{n-1,n-1}^{n-1}(\theta) & \text{if } n > 1, \ p = n, \qquad j = n. \end{cases} \tag{15}$$

Proof We prove this by induction. For $n = 1$ we have $(e^\theta D + E)^1 = e^\theta D + E$ which settles the initial values $f_{1,0}^1(\theta) = 1$ and $f_{1,1}^1(\theta) = e^\theta$. To find the recursion we assume the induction hypothesis:

$$(e^\theta D + E)^n = \sum_{p=1}^n \sum_{j=0}^p f_{p,j}^n(\theta) E^{p-j} D^j$$

and expand the next power. However, there are two ways we can use to expand, namely $(e^\theta D + E)^{n+1} = (e^\theta D + E)^n (e^\theta D + E)$ or $(e^\theta D + E)(e^\theta D + E)^n$. The former will give Eq. 14, the latter Eq. 15. The functions $f_{p,j}^n$ are all polynomials in e^θ because this holds for the induction hypothesis and the operations in the induction step are only multiplications and additions of polynomials with positive coefficients. □

We now state an auxiliary result.

Lemma 5.2 *For* $n \geq 2$ *and* $1 \leq p \leq r \leq n - 1$, *we have the identity*

$$\sum_{k=p}^r k \binom{n-1-k}{r-k} = \frac{np - pr + r}{n-r+1} \binom{n-p}{r-p}. \tag{16}$$

Proof First note that the cases $p = r$ and $r = n - 1$ are easy to check directly. We prove the general case by induction over n. The case $n = 2$ is again easy to see. We now assume that Eq. 16 holds for fixed $n \geq 2$ and all $1 \leq p \leq r \leq n - 1$.

To show the result for $n + 1$ we may assume $1 \leq p < r \leq n - 1$, ignoring the easy cases settled at the beginning of the proof. Starting from the left hand side for $n + 1$ and using the induction hypothesis on the third equality,

$$
\begin{aligned}
\sum_{k=p}^{r} k \binom{n-k}{r-k} &= \sum_{k=p}^{r-1} k \left[\binom{n-k-1}{r-k-1} + \binom{n-k-1}{r-k} \right] + r \\
&= \sum_{k=p}^{r-1} k \binom{n-k-1}{r-k-1} + \sum_{k=p}^{r} k \binom{n-k-1}{r-k} \\
&= \frac{np - p(r-1) + (r-1)}{n - (r-1) + 1} \binom{n-p}{(r-1)-p} + \frac{np - pr + r}{n - r + 1} \binom{n-p}{r-p} \\
&= \frac{(n+1)p - pr + r}{(n+1) - r + 1} \binom{(n+1) - p}{r - p}.
\end{aligned}
$$

$\square$

We now identify the coefficients $f_{p,p}^{n}(\theta)$, for $1 \leq p \leq n$.

Proposition 5.3 *For the coefficients defined in Proposition 5.1,*

$$
f_{p,p}^{n}(\theta) = \begin{cases} \sum_{r=p}^{n-1} \frac{p}{n} \binom{n-p-1}{r-p} \binom{n}{r} e^{r\theta} & \text{if } 1 \leq p < n, \\ e^{n\theta} & \text{if } p = n. \end{cases} \tag{17}
$$

Proof Putting $j = p$ in Eq. 14 of Proposition 5.1 we get a simplified recursion:

$$
f_{1,1}^{1}(\theta) = e^{\theta}
$$

and

$$
f_{p,p}^{n}(\theta) = \begin{cases} \sum_{k=1}^{n-1} f_{k,k}^{n-1}(\theta) & \text{if } n > 1, \ p = 1 \\ e^{\theta} f_{p-1,p-1}^{n-1}(\theta) + \sum_{k=p}^{n-1} f_{k,k}^{n-1}(\theta) & \text{if } n > 1, \ 1 < p < n \\ e^{\theta} f_{n-1,n-1}^{n-1}(\theta) & \text{if } n > 1, \ p = n. \end{cases} \tag{18}
$$

If $p = n$, it is easy to see by induction that

$$
f_{n,n}^{n}(\theta) = e^{n\theta}. \tag{19}
$$

Now, if $p < n$, we proceed again by induction. Here the base of induction has to be $n = 2$. The recursion Eq. 18 gives $f_{1,1}^{2}(\theta) = f_{1,1}^{1}(\theta) = e^{\theta}$, as required by formula (17). We now assume that Eq. 17 holds for fixed $n \geq 2$ and all $p < n$. We first consider the branch of Eq. 18, referring to the case $p = 1$. Using Eq. 19 and the induction hypothesis we obtain

$$
f_{1,1}^{n+1}(\theta) = \sum_{k=1}^{n} f_{k,k}^{n}(\theta) = \sum_{k=1}^{n-1} f_{k,k}^{n}(\theta) + f_{n,n}^{n}(\theta) = \sum_{k=1}^{n-1} \sum_{r=k}^{n-1} \frac{k}{n} \binom{n-k-1}{r-k} \binom{n}{r} e^{r\theta} + e^{n\theta},
$$

changing the order of summation and using Lemma 5.2 gives

$$
= \sum_{r=1}^{n-1} \sum_{k=1}^{r} \frac{k}{n} \binom{n-k-1}{r-k} \binom{n}{r} e^{r\theta} + e^{n\theta} = \sum_{r=1}^{n-1} \frac{1}{n} \binom{n}{r} e^{r\theta} \sum_{k=1}^{r} k \binom{n-k-1}{r-k} + e^{n\theta}
$$

$$
= \sum_{r=1}^{n-1} \frac{1}{n} \binom{n}{r} e^{r\theta} \frac{n}{n-r+1} \binom{n-1}{r-1} + e^{n\theta} = \sum_{r=1}^{n} \frac{1}{n+1} \binom{n-1}{r-1} \binom{n+1}{r} e^{r\theta}.
$$

Since this is the result required by the induction step, the case $p = 1$ is settled. We can therefore turn our attention to the remaining branch of Eq. 18, covering $1 < p \le n$. We obtain from the induction hypothesis

$$
f_{p,p}^{n+1}(\theta) = e^{\theta} f_{p-1,p-1}^{n}(\theta) + \sum_{k=p}^{n} f_{k,k}^{n}(\theta)
$$

$$
= e^{\theta} \sum_{r=p-1}^{n-1} \frac{p-1}{n} \binom{n-p}{r-p+1} \binom{n}{r} e^{r\theta} + \sum_{k=p}^{n-1} \sum_{r=k}^{n-1} \frac{k}{n} \binom{n-k-1}{r-k} \binom{n}{r} e^{r\theta} + e^{n\theta};
$$

changing the summation order and grouping the terms by powers of e^{θ} yields

$$
= \sum_{r=p}^{n} \frac{p-1}{n} \binom{n-p}{r-p} \binom{n}{r-1} e^{r\theta} + \sum_{r=p}^{n-1} \sum_{k=p}^{r} \frac{k}{n} \binom{n-k-1}{r-k} \binom{n}{r} e^{r\theta} + e^{n\theta}
$$

$$
= \sum_{r=p}^{n-1} \left[\frac{p-1}{n} \binom{n-p}{r-p} \binom{n}{r-1} + \sum_{k=p}^{r} \frac{k}{n} \binom{n-k-1}{r-k} \binom{n}{r} \right] e^{r\theta}
$$

$$
+ \left[\frac{p-1}{n} \binom{n-p}{n-p} \binom{n}{n-1} + 1 \right] e^{n\theta};
$$

simplifying and using Lemma 5.2 in the second line gives

$$
= \sum_{r=p}^{n-1} \left[\frac{r(p-1)}{n-r+1} \binom{n-p}{r-p} + \sum_{k=p}^{r} k \binom{n-k-1}{r-k} \right] \frac{1}{n} \binom{n}{r} e^{r\theta} + p e^{n\theta}
$$

$$
= \sum_{r=p}^{n-1} \left[\frac{r(p-1)}{n-r+1} \binom{n-p}{r-p} + \frac{np-pr+r}{n-r+1} \binom{n-p}{r-p} \right] \frac{1}{n} \binom{n}{r} e^{r\theta} + p e^{n\theta}
$$

$$
= \sum_{r=p}^{n-1} \left[\frac{r(p-1)}{n-r+1} + \frac{np-pr+r}{n-r+1} \right] \frac{1}{n} \binom{n-p}{r-p} \binom{n}{r} e^{r\theta} + p e^{n\theta};
$$

grouping that last term with the rest of the terms in the sum finally results in

$$
= \sum_{r=p}^{n} \frac{p}{n+1} \binom{n-p}{r-p} \binom{n+1}{r} e^{r\theta}.
$$

$\square$

Proposition 5.4 *For all $n \in \mathbb{N}$, $1 \le p \le n$ and $0 \le j \le p$ we have the symmetry*

$$
f_{p,j}^{n}(\theta) = e^{n\theta} f_{p,p-j}^{n}(-\theta).
$$

Proof Defining the polynomials $g_{p,j}^n(\theta) := e^{n\theta} f_{p,p-j}^n(-\theta)$, we can write

$$f_{p,j}^n(\theta) = e^{n\theta} g_{p,p-j}^n(-\theta),$$

and by the definition Eq. 13 of the polynomials $f_{p,j}^n(\theta)$ we obtain by changing the summation index

$$(e^{\theta} D + E)^n = \sum_{p=1}^n \sum_{j=0}^p f_{p,j}^n(\theta) E^{p-j} D^j = \sum_{p=1}^n \sum_{j=0}^p e^{n\theta} g_{p,p-j}^n(-\theta) E^{p-j} D^j$$

$$= \sum_{p=1}^n \sum_{j=0}^p e^{n\theta} g_{p,j}^n(-\theta) E^j D^{p-j}. \tag{20}$$

Evaluating this expression for $n = 1$, we find

$$e^{\theta} D + E = e^{\theta} g_{1,0}^1(-\theta) D + e^{\theta} g_{1,1}^1(-\theta) E,$$

and hence $g_{1,1}^1(\theta) = e^{\theta}$ and $g_{1,0}^1(\theta) = 1$. Next we find a recursive relation for these polynomials by expanding and employing Eq. 20,

$$(e^{\theta} D + E)^{n+1} = (e^{\theta} D + E)^n (e^{\theta} D + E)$$

$$= \left(\sum_{p=1}^n \sum_{j=0}^p e^{n\theta} g_{p,j}^n(-\theta) E^j D^{p-j} \right) (e^{\theta} D + E)$$

and equating the coefficients to

$$(e^{\theta} D + E)^{n+1} = \sum_{p=1}^{n+1} \sum_{j=0}^p e^{n\theta} g_{p,j}^n(-\theta) E^j D^{p-j}$$

we find that the polynomials $g_{p,j}^n$ satisfy the following recursion:

$$g_{1,0}^1(\theta) = 1, \qquad g_{1,1}^1(\theta) = e^{\theta}$$

$$g_{p,j}^n(\theta) = \begin{cases} \sum_{k=1}^{n-1} e^{\theta} g_{k,0}^{n-1}(\theta) & \text{if } n > 1, \ p = 1, \qquad j \in \{0, 1\} \\ g_{p-1,j}^{n-1}(\theta) + \sum_{k=p}^{n-1} e^{\theta} g_{k,j}^{n-1}(\theta) & \text{if } n > 1, \ 1 < p < n, \ 0 \le j < p \\ e^{\theta} g_{p-1,p-1}^{n-1}(\theta) + \sum_{k=p}^{n-1} e^{\theta} g_{k,p-1}^{n-1}(\theta) & \text{if } n > 1, \ 1 < p < n, \ j = p \\ g_{n-1,j}^{n-1}(\theta) & \text{if } n > 1, \ p = n, \qquad 0 \le j < n \\ e^{\theta} g_{n-1,n-1}^{n-1}(\theta) & \text{if } n > 1, \ p = n, \qquad j = n. \end{cases}$$

This is the recursion Eq. 15 of Proposition 5.1, and hence $f_{p,j}^n(\theta) = g_{p,j}^n(\theta)$. Thus from the definition of $g_{p,j}^n(\theta)$ we conclude that $f_{p,j}^n(\theta) = e^{n\theta} f_{p,p-j}^n(-\theta)$ as claimed. $\qquad \square$

From the symmetry rule of Proposition 5.4 we obtain a simple characterisation of the coefficients $f_{p,0}^n(\theta)$.

Corollary 5.5

$$f_{p,0}^n(\theta) = \begin{cases} \sum_{r=1}^{n-p} \dfrac{p}{n} \dbinom{n-p-1}{r-1} \dbinom{n}{r} e^{r\theta} & \text{if } 1 \le p < n \\ 1 & \text{if } p = n. \end{cases}$$

Proof The result follows by combining Proposition 5.4 for $j = p$, and Proposition 5.3. $\quad \square$

We plug the expansion (13) into Eq. 5 and obtain

$$\Lambda(\theta) = \lim_{n\to\infty} \frac{1}{n} \log \left\{ \sum_{p=1}^{n} \sum_{j=0}^{p} f_{p,j}^{n}(\theta) w^{\mathsf{T}} E^{p-j} D^{j} v \right\} + \log c. \tag{21}$$

Since all terms are positive, we can find lower bounds by considering only the terms for which $j = 0$ or $j = p$. This is the content of the next couple of results.

Proposition 5.6 *For the cumulant generating function Λ defined by* Eq. 5,

$$\Lambda(\theta) \geq \begin{cases} 2\log\left(1 + e^{\theta/2}\right) + \log c & \text{if } \theta \leq -2\log\lambda_1, \\ \log\left(1 + \lambda_1 e^{\theta}\right) + \log\left(1 + \frac{1}{\lambda_1}\right) + \log c & \text{if } \theta > -2\log\lambda_1. \end{cases}$$

Proof From the explicit form of D, w, and v in Eq. 8 it can be shown by induction that

$$w^{\mathsf{T}} D^{p} v = (1 + \lambda_1)^{p} \frac{\alpha}{\lambda_1 - \lambda_2} \left\{ \frac{\lambda_1}{\alpha - \lambda_1(1-\alpha)} - \left(\frac{1+\lambda_2}{1+\lambda_1}\right)^{p} \frac{\lambda_2}{\alpha - \lambda_2(1-\alpha)} \right\}. \tag{22}$$

We consider only the terms for which $j = p$ in Eq. 21 and note that the expression in the square parenthesis in Eq. 22 vanishes in the limit taken in Eq. 21. Hence, using the Laplace principle and Proposition 5.3,

$$\Lambda(\theta) \geq \lim_{n\to\infty} \frac{1}{n} \log \left\{ \sum_{p=1}^{n} f_{p,p}^{n}(\theta)(1+\lambda_1)^{p} \right\} + \log c$$

$$= \lim_{n\to\infty} \sup_{1\leq p\leq r\leq n} \frac{1}{n} \log \left\{ \frac{p}{n} \binom{n-p-1}{r-p} \binom{n}{r} e^{r\theta} (1+\lambda_1)^{p} \right\} + \log c.$$

We now use Stirling's formula and a change of variables $\varepsilon = \frac{r}{n}$ and $\delta = \frac{p}{n}$ to obtain

$$\Lambda(\theta) \geq \sup_{0<\delta\leq\varepsilon\leq 1} \left[\lim_{n\to\infty} \frac{1}{n} \log \left\{ \delta\binom{n(1-\delta)}{n(\varepsilon-\delta)} \binom{n}{n\varepsilon} \right\} + \varepsilon\theta + \delta\log(1+\lambda_1) \right] + \log c$$

$$= \sup_{0<\varepsilon\leq 1} \left[\sup_{0<\delta\leq\varepsilon} \left\{ -(\varepsilon-\delta)\log(\varepsilon-\delta) + (1-\delta)\log(1-\delta) + \delta\log(1+\lambda_1) \right\} \right.$$

$$\left. + \varepsilon\theta - 2(1-\varepsilon)\log(1-\varepsilon) - \varepsilon\log\varepsilon \right] + \log c.$$

The inner problem, when ε is fixed, is solved by choosing $\delta^{\max} = 0$, if $\varepsilon \leq \frac{1}{1+\lambda_1}$, and $\delta^{\max} = \frac{(1+\lambda_1)\varepsilon - 1}{\lambda_1}$, if $\varepsilon > \frac{1}{1+\lambda_1}$. So now we have

$$\Lambda(\theta) \geq \max \left\{ \sup_{0<\varepsilon\leq\frac{1}{1+\lambda_1}} \left[\varepsilon\theta - 2\varepsilon\log\varepsilon - 2(1-\varepsilon)\log(1-\varepsilon) \right], \right.$$

$$\left. \sup_{\frac{1}{1+\lambda_1}<\varepsilon\leq 1} \left[\varepsilon\theta - \varepsilon\log\varepsilon - (1-\varepsilon)\log[\lambda_1(1-\varepsilon)] + \log(1+\lambda_1) \right] \right\} + \log c.$$

This problem is solved by choosing

$$\varepsilon^{\max} = \begin{cases} \frac{e^{\theta/2}}{1+e^{\theta/2}} & \text{if } \theta \leq -2\log\lambda_1 \\[2mm] \frac{\lambda_1 e^{\theta}}{1+\lambda_1 e^{\theta}} & \text{if } \theta > -2\log\lambda_1. \end{cases}$$

Plugging this value of $\varepsilon^{\max}$ yields the result of the proposition. □

Proposition 5.7 *For the cumulant generating function Λ defined by* Eq. 5,

$$\Lambda(\theta) \geq \begin{cases} \log\left(\frac{e^{\theta}}{1-\alpha} + \frac{1}{\alpha}\right) + \log c & \text{if } \theta \leq 2\log\left(\frac{1}{\alpha} - 1\right), \\[2mm] 2\log\left(1 + e^{\theta/2}\right) + \log c & \text{if } \theta > 2\log\left(\frac{1}{\alpha} - 1\right). \end{cases}$$

Proof We follow the same technique as in the previous proposition, but now consider only those values for which $j = 0$ in Eq. 13. Hence, using Corollary 5.5

$$\Lambda(\theta) \geq \lim_{n\to\infty} \frac{1}{n} \log \left\{ \sum_{p=1}^{n} f_{p,0}^{n}(\theta) \frac{1}{\alpha^p} \right\} + \log c$$

$$= \lim_{n\to\infty} \sup_{\substack{1\leq p\leq n \\ 0\leq r\leq n-p}} \frac{1}{n} \log \left\{ \frac{p}{n}\binom{n-p-1}{r-1}\binom{n}{r} e^{r\theta}\alpha^{-p} \right\} + \log c$$

With the same change of variables as before, $\varepsilon = \frac{r}{n}$ and $\delta = \frac{p}{n}$, we have

$$\Lambda(\theta) \geq \sup_{\substack{0<\delta\leq 1 \\ 0\leq\varepsilon\leq 1-\delta}} \left[\lim_{n\to\infty} \frac{1}{n} \log\left\{ \delta\binom{n(1-\delta)}{n\varepsilon}\binom{n}{n\varepsilon} \right\} + \varepsilon\theta - \delta\log\alpha \right] + \log v$$

$$= \sup_{\substack{0\leq\delta\leq 1 \\ 0\leq\varepsilon\leq 1-\delta}} \{(1-\delta)\log(1-\delta) - (1-\delta-\varepsilon)\log(1-\delta-\varepsilon) - \delta\log\alpha$$

$$- 2\varepsilon\log\varepsilon - (1-\varepsilon)\log(1-\varepsilon) + \varepsilon\theta\} + \log c.$$

Splitting the problem in two, for a fixed ε, we find the optimal δ as $\delta^{\max} = 0$, if $1 - \alpha \leq \varepsilon \leq 1$, and $\delta^{\max} = 1 - \frac{\varepsilon}{1-\alpha}$, if $0 \leq \varepsilon < 1 - \alpha$. And the remaining problem is solved by choosing the optimum ε as

$$\varepsilon^{\max} = \begin{cases} \frac{e^{\theta/2}}{1+e^{\theta/2}} & \text{if } \theta \leq 2\log\left(\frac{1}{\alpha} - 1\right), \\[2mm] \frac{\alpha}{\alpha+e^{-\theta}(1-\alpha)} & \text{if } \theta > 2\log\left(\frac{1}{\alpha} - 1\right). \end{cases}$$

With these values of ε we get the desired result. □

Corollary 5.8 *For Λ defined by* Eq. 5,

$$\Lambda(\theta) \geq \begin{cases} \log\left(\frac{e^{\theta}}{1-\alpha} + \frac{1}{\alpha}\right) + \log c & \text{if } -\infty < \theta \leq 2\log\left(\frac{1}{\alpha} - 1\right), \\[2mm] 2\log\left(1 + e^{\theta/2}\right) + \log c & \text{if } 2\log\left(\frac{1}{\alpha} - 1\right) < \theta \leq -2\log\lambda_1, \\[2mm] \log\left(1 + \lambda_1 e^{\theta}\right) + \log\left(1 + \frac{1}{\lambda_1}\right) + \log c & \text{if } -2\log\lambda_1 < \theta < \infty. \end{cases}$$

Proof The bounds from Propositions 5.6 and 5.7 are the same in the interval $2\log\left(\frac{1}{\alpha} - 1\right) \leq \theta \leq -2\log\lambda_1$. In the other intervals, a comparison of the bounds establishes the claim. □

6 The Rate Function

Summarising, we have the following result.

Corollary 6.1 *For the cumulant generating function Λ defined by* Eq. 5,

$$\Lambda(\theta) = \begin{cases} \log\left(\frac{e^\theta}{1-\alpha} + \frac{1}{\alpha}\right) + \log c & \text{if } -\infty < \theta \leq 2\log\left(\frac{1}{\alpha} - 1\right), \\ 2\log\left(1 + e^{\theta/2}\right) + \log c & \text{if } 2\log\left(\frac{1}{\alpha} - 1\right) < \theta \leq -2\log\lambda_1, \\ \log\left(1 + \lambda_1 e^\theta\right) + \log\left(1 + \frac{1}{\lambda_1}\right) + \log c & \text{if } -2\log\lambda_1 < \theta < \infty. \end{cases} \tag{23}$$

Proof This follows from the fact that the upper and lower bounds from Proposition 4.7 and Corollary 5.8, respectively, are the same. □

Finally we have the necessary tools to prove Theorem 2.2.

Proof The rate function in the case *(a)* $\alpha \leq 1/2$ and $\rho < 1 - \alpha$ is known from Cramér's theorem, see e.g. Exercise 2.2.23 (b) in [8]. For the case *(c)*, $\alpha > 1/2$ and $\frac{1}{2} < \rho < \alpha$, we show that the function Λ defined by Eq. 5, given explicitly in Corollary 6.1, satisfies the hypotheses of the Gärtner-Ellis theorem 3.3. Note that Λ is defined for all real numbers. An evaluation at the boundaries of the domains gives

$$\Lambda\left(2\log\left(\frac{1}{\alpha} - 1\right)\right) = -2\log\alpha + \log c = \lim_{h\to 0^+} \Lambda\left(2\log\left(\frac{1}{\alpha} - 1\right) - h\right)$$

and

$$\Lambda\left(-2\log\lambda_1\right) = 2\log\left(1 + \frac{1}{\lambda_1}\right) + \log c = \lim_{h\to 0^+} \Lambda\left(-2\log\lambda_1 + h\right),$$

which implies that it is continuous in all $\mathbb{R}$. Moreover,

$$\lim_{h\to 0^\pm} \frac{\Lambda\left(2\log(\frac{1}{\alpha} - 1) + h\right) - \Lambda\left(2\log\left(\frac{1}{\alpha} - 1\right)\right)}{h} = 1 - \alpha$$

and

$$\lim_{h\to 0^\pm} \frac{\Lambda\left(-2\log\lambda_1 + h\right) - \Lambda\left(-2\log\lambda_1\right)}{h} = \frac{1}{1 + \lambda_1}.$$

Therefore, Λ is differentiable in $\mathbb{R}$. By the Gärtner-Ellis theorem we find the rate function,

$$I(z) = \sup_{\theta\in\mathbb{R}} \{z\theta - \Lambda(\theta)\} \qquad \text{for } z \in (0, 1).$$

For fixed $z \in (0, 1)$ the function $\theta \mapsto z\theta - \Lambda(\theta)$ is well-defined, continuous, and differentiable in all $\mathbb{R}$. It is also a concave function and hence the maximum is reached at a value of θ where the derivative vanishes. Since

$$\frac{d}{d\theta}(z\theta - \Lambda(\theta)) = \begin{cases} z - \frac{\alpha e^\theta}{\alpha e^\theta + 1 - \alpha} & \text{if } -\infty < \theta \leq \log\left(\frac{1}{\alpha} - 1\right)^2, \\ z - \frac{e^{\theta/2}}{1 + e^{\theta/2}} & \text{if } \log\left(\frac{1}{\alpha} - 1\right)^2 < \theta \leq -2\log\lambda_1, \\ z - \frac{\lambda_1 e^\theta}{1 + \lambda_1 e^\theta} & \text{if } -2\log\lambda_1 < \theta < \infty, \end{cases}$$

we get $\frac{d}{d\theta}(z\theta - \Lambda(\theta)) = 0$ if and only if

$$\theta = \log \frac{z(1-\alpha)}{\alpha(1-z)} \quad \text{and} \quad \theta \leq 2\log\left(\frac{1}{\alpha} - 1\right),$$

$$\text{or} \quad \theta = 2\log \frac{z}{1-z} \quad \text{and} \quad 2\log\left(\frac{1}{\alpha} - 1\right) < \theta \leq -2\log\lambda_1,$$

$$\text{or} \quad \theta = \log \frac{z}{\lambda_1(1-z)} \quad \text{and} \quad \theta > -2\log\lambda_1.$$

This means that

$$\theta = \log \frac{z(1-\alpha)}{\alpha(1-z)} \quad \Leftrightarrow \quad 0 < z < 1 - \alpha,$$

$$\theta = 2\log \frac{z}{1-z} \quad \Leftrightarrow \quad 1 - \alpha < z \leq \frac{1}{1+\lambda_1},$$

$$\theta = \log \frac{z}{\lambda_1(1-z)} \quad \Leftrightarrow \quad \frac{1}{1+\lambda_1} < z < 1.$$

Since the value of θ that satisfies $\frac{d}{d\theta}(z\theta - \Lambda(\theta)) = 0$ is unique it must be the maximum. By plugging in this value in Eq. 23, we reach the desired result. The remaining case *(b)* is obtained simply by taking the limit of case *(c)* as $c \to \frac{1}{4}$. □

An alternative approach to the problems studied here is based on translation into a random walk problem. Rewriting the infinite matrix $e^\theta D + E$ as $2(1 + e^\theta)Q_\theta$, one can see that Q_θ can be interpreted as the probability transition matrix of a discrete time random walk $\{X_n\}_{n\geq 0}$ on $\mathbb{N}_0$ killed at the origin. By the form of the vector w^T, shown in Eq. 8, the product $w^\mathsf{T} Q_\theta^n$ is, up to a normalising constant, the sub-probability distribution of the n-th step of the random walk starting from a geometric distribution with parameter $2 - \frac{1}{\alpha}$. If T denotes the killing time of the random walk, then Eq. 23 becomes

$$\Lambda(\theta) = \lim_{n\to\infty} \frac{1}{n} \log w^\mathsf{T}(e^\theta D + E)^n v + \log c$$

$$= \log\left(2c(1 + e^\theta)\right) + \lim_{n\to\infty} \frac{1}{n} \log \mathbb{E}\left[(\lambda_1^{X_n} - \lambda_2^{X_n})\mathbb{1}_{\{T>n\}}\right],$$

using the form of v given in Eq. 8. The latter rate may be evaluated using a functional large deviation principle for the random walk. This alternative method hinges on the availability of the large deviation result and the resolvability of the variational problems that come out of its application, which may be more complicated than our original approach. In principle, however, the method seems suited to not only reveal large deviations for the asymptotic density, but also for a density profile depending on a macroscopic space variable, as done by Derrida et al. in [12] in the case of the finite TASEP.

By contrast, the technique presented in this paper is more elementary and direct. It therefore seems to be more flexible, giving hope to produce large deviation principles for a range of other systems with spatially correlated distributions given by a matrix product representation. In particular there are several natural variants even of the example of a semi-infinite TASEP considered in this paper: For example, we would be interested in identifying a large deviation principle for the semi-infinite TASEP process starting from a configuration given by a value of ρ in Theorem 2.1 with $\rho > \alpha > \frac{1}{2}$, which was so far excluded for technical reasons. It also seems feasible to generalise the large deviation principle for the overall density to a large deviation principle for a density profile. Finally, it would be interesting to analyse the non-stationary TASEP using a time dependent matrix representation, as given by Stinchcombe and Schütz in [20].

References

1. Angeletti, F., Touchette, H., Bertin, E., Abry, P.: Large deviations for correlated random variables described by a matrix product ansatz. J. Stat. Mech. Theory Exp. **2014**(2), P02003, 17 (2014)
2. Bahadoran, C.: A quasi-potential for conservation laws with boundary conditions. arXiv:1010.3624 (2010)
3. Bertini, L., De Sole, A., Gabrielli, D., Jona-Lasinio, G., Landim, C.: Large deviations for the boundary driven symmetric simple exclusion process. Math. Phys. Anal. Geom. **6**(3), 231–267 (2003)
4. Bertini, L., De Sole, A., Gabrielli, D., Jona-Lasinio, G., Landim, C.: Large deviation approach to non equilibrium processes in stochastic lattice gases. Bull. Braz. Math. Soc. **37**(4), 611–643 (2006)
5. Bertini, L., Landim, C., Mourragui, M.: Dynamical large deviations for the boundary driven weakly asymmetric exclusion process. Ann. Probab. **37**(6), 2357–2403 (2009)
6. Blythe, R.A., Evans, M.R.: Nonequilibrium steady states of matrix-product form: a solver's guide. J. Phys. A **40**(46), R333–R441 (2007)
7. Bodineau, T., Giacomin, G.: From dynamic to static large deviations in boundary driven exclusion particle systems. Stoch. Proc. Appl. **110**(1), 67–81 (2004)
8. Dembo, A., Zeitouni, O.: Large Deviations Techniques and Applications, Volume 38 of Stochastic Modelling and Applied Probability. Springer, Berlin (2010)
9. den Hollander, F.: Large Deviations, Volume 14 of Fields Institute Monographs. American Mathematical Society, Providence (2000)
10. Derrida, B.: Matrix ansatz large deviations of the density in exclusion processes. In: International Congress of Mathematicians, vol. III, pp. 367–382. Eur. Math. Soc., Zürich (2006)
11. Derrida, B., Evans, M.R., Hakim, V., Pasquier, V.: Exact solution of a 1d asymmetric exclusion model using a matrix formulation. J. Phys. A **26**(7), 1493–1517 (1993)
12. Derrida, B., Lebowitz, J.L., Speer, E.R.: Exact large deviation functional of a stationary open driven diffusive system: the asymmetric exclusion process. J. Statist. Phys. **110**(3–6), 775–810 (2003)
13. Evans, M.R., Ferrari, P.A., Mallick, K.: Matrix representation of the stationary measure for the multispecies TASEP. J. Stat. Phys. **135**(2), 217–239 (2009)
14. Grosskinsky, S.: Phase transitions in nonequilibrium stochastic particle systems with local conservation laws. PhD thesis, TU Munich (2004)
15. Hinrichsen, H.: Matrix product ground states for exclusion processes with parallel dynamics. J. Phys. A **29**(13), 3659–3667 (1996)
16. Liggett, T.M.: Ergodic theorems for the asymmetric simple exclusion process. Trans. Am. Math. Soc. **213**, 237–261 (1975)
17. Nyawo, P.T., Touchette, H.: A minimal model of dynamical phase transition. EPL (Europhys. Lett.) **116**(5), 50009 (2016)
18. Sasamoto, T.: One-dimensional partially asymmetric simple exclusion process with open boundaries: orthogonal polynomials approach. J. Phys. A **32**(41), 7109–7131 (1999)
19. Sasamoto, T., Williams, L.: Combinatorics of the asymmetric exclusion process on a semi-infinite lattice. arXiv:1204.1114 (2012)
20. Stinchcombe, R.B., Schütz, G.M.: Operator algebra for stochastic dynamics and the Heisenberg chain. EPL (Europhys. Lett.) **29**(9), 663 (1995)
21. Tracy, C.A., Widom, H.: The asymmetric simple exclusion process with an open boundary. J. Math. Phys. **54**(10), 103301, 16 (2013)
22. Trefethen, L.N., Embree, M.: Spectra and Pseudospectra. Princeton University Press, Princeton (2005)