

BOUNDS ON THE NORM OF WIGNER-TYPE RANDOM MATRICES

LÁSZLÓ ERDŐS AND PETER MÜHLBACHER

IST AUSTRIA, A-3400 KLOSTERNEUBURG

ABSTRACT. We consider a Wigner-type ensemble, i.e. large hermitian $N \times N$ random matrices $H = H^*$ with centered independent entries and with a general matrix of variances $S_{xy} = \mathbb{E}|H_{xy}|^2$. The norm of H is asymptotically given by the maximum of the support of the self-consistent density of states. We establish a bound on this maximum in terms of norms of powers of S that substantially improves the earlier bound $2\|S\|_\infty^{1/2}$ given in [1]. The key element of the proof is an effective Markov chain approximation for the contributions of the weighted Dyck paths appearing in the iterative solution of the corresponding Dyson equation.

1. INTRODUCTION AND THE MAIN RESULT

Large hermitian random matrices with independent entries tend to exhibit deterministic patterns. In particular, the empirical density of eigenvalues typically converges to a deterministic density profile, ρ , called the *self-consistent density of states*, that can be determined by solving a system of quadratic equations. Under very general conditions, ρ is compactly supported and the largest eigenvalue of the random matrix is asymptotically given by the maximum of the support of ρ .

In the simplest case of $N \times N$ Wigner matrices, i.e. when H has centered, identically distributed entries that are independent (up to the symmetry constraint $H = H^*$), the self-consistent density of states is given explicitly by the Wigner semicircle law. Under the customary normalization $\mathbb{E}|H_{xy}|^2 = N^{-1}$, the semicircle distribution is supported in $[-2, 2]$. With very high probability (and also almost surely) the Euclidean matrix norm tends to 2, i.e. $\|H\|_2 \rightarrow 2$ as $N \rightarrow \infty$, assuming the fourth moment of $\sqrt{N}H_{xy}$ is finite [2].

In this paper we consider *Wigner type* matrices introduced in [1]. These are generalizations of the Wigner ensemble where independence of the matrix elements is retained but their distribution may vary within the matrix. We assume $\mathbb{E}H = 0$. The relevant parameter of the model is the matrix of variances

$$S = (S_{xy})_{x,y=1}^N, \quad S_{xy} := \mathbb{E}|H_{xy}|^2.$$

E-mail addresses: lerdos@ist.ac.at, peter@muehlbacher.me.

Date: February 15, 2018.

2010 Mathematics Subject Classification. 60B20, 60J10.

Partially supported by ERC Advanced Grant RANMAT No. 338804.

The self-consistent density of states is obtained via the solution of a system of quadratic equations

$$-\frac{1}{m_x} = z + \sum_{y=1}^N S_{xy} m_y \quad x = 1, 2, \dots, N, \quad (1.1)$$

where $z \in \mathbb{H}$ is a complex spectral parameter in the upper half plane. This equation was extensively studied in [3, 4]. Under the additional condition that $\Im m_x > 0$, the solution to (1.1) is unique and it depends analytically on $z \in \mathbb{H}$. Its average,

$$m(z) := \frac{1}{N} \sum_{x=1}^N m_x(z) \quad (1.2)$$

is the Stieltjes transform of a probability density ρ . This relation defines the self-consistent density of states measure that can be obtained by inverting the Stieltjes transform as

$$\rho(d\tau) := \lim_{\eta \downarrow 0} \frac{1}{\pi N} \sum_{x=1}^N \Im m_x(\tau + i\eta) d\tau. \quad (1.3)$$

A simple symmetry argument shows that $m(-\bar{z}) = -\overline{m(z)}$ and thus ρ is an even measure. We remark that under additional assumptions on S , the measure ρ is absolutely continuous with a Hölder continuous density except at $\tau = 0$, where it may have a Dirac delta component (Corollary 7.4 [4]). Note that m_x, m, ρ as well as S depend on N , i.e. $m_x = m_x^{(N)}$ etc., but this dependence will sometimes be omitted from the notation.

Under very general conditions on S and some higher moment assumption of H , it is well known that the empirical density of eigenvalues of H is asymptotically given by ρ . This holds not only on the global scale, but even on very small scales slightly above the typical eigenvalue spacing; these statements are called *local laws* for Wigner-type matrices (Theorem 1.7 of [1]).

Local laws are typically not sensitive to individual eigenvalues except at the spectral edges, where a stronger version of the local law holds. Therefore, the maximum of the support of ρ correctly describes the largest eigenvalue or the norm of H (see [5] for a quite general setup). In particular, the norm of a Wigner matrix with the above normalization converges to 2. The speed of convergence has been addressed in several papers in increasing generality, see e.g. [2, 6, 7, 8]. Similarly, the norm of a Wigner-type matrix converges to $\max \text{supp } \rho$. For a general variance matrix S , neither ρ nor its support can be computed explicitly; our current goal is to give a good bound on $\max \text{supp } \rho$. A relatively simple argument (see Proposition 2.1 [1]) gives

$$\max \text{supp } \rho \leq 2\|S\|^{1/2}, \quad (1.4)$$

where for any matrix M we let $\|M\| := \|M\|_\infty := \max_x \sum_y |M_{xy}|$ denote the matrix norm induced by the maximum norm on $\mathbb{C}^N$. In this paper $\|\cdot\|$ always denotes this maximum norm.

Our main theorem considerably improves the bound (1.4) and it is still expressed in terms of norms of powers of S .

Theorem 1.1. *Let S be a variance matrix, i.e. a symmetric $N \times N$ matrix with nonnegative entries. Let ρ be the self-consistent density of states obtained from the unique solution of (1.1) via inverse Stieltjes transform (1.3). Set*

$$z_j := \frac{\|S^j\|}{\|S\|^j} \quad (1.5)$$

for any $j \in \mathbb{N}$. Then for any fixed $J \in \mathbb{N} \cup \{\infty\}$ we have

$$\max \text{supp } \rho \leq 2 \frac{\|S\|^{\frac{1}{2}}}{w_c(J)}, \quad (1.6)$$

where $w_c(J)$ is the smallest positive root of the function

$$\phi_J(w) := 1 - \frac{w}{2} \left(1 + \sum_{j=1}^J \left(\frac{w}{2} \right)^j z_j + \sum_{j>J} \left(\frac{w}{2} \right)^j \right). \quad (1.7)$$

It is easy to see that $w_c(J)$ is an increasing function of J , so a choice of larger J yields a better bound. In particular $J = \infty$ is the best. On the other hand, larger J is more computation intensive as it requires to compute norms of higher powers of S .

Comparing (1.6) with (1.4), notice that the main source of the improvement is the simple fact that the inequality $\|S^j\| \leq \|S\|^j$ rarely saturates. Indeed, it is easy to see that w_c is a strictly monotonically decreasing function of all z_j . If all $\|S^j\|$ norms were replaced with $\|S\|^j$, i.e. we set $z_j = 1$, then $w_c(J) = 1$ for any J and the two bounds were identical. Once $z_j < 1$ for some j , we have $w_c > 1$.

In the Appendix we illustrate in an example the effect of the improvement and compare it with the exact value of $\max \text{supp } \rho$ obtained numerically.

Combining Theorem 1.1 with Corollary 2.3 of [9] (or Theorem 4.7 of [5]) on the convergence of the largest eigenvalue of the Wigner type matrix and using that $w_c(J)$ depends continuously and monotonically on z_j , we immediately obtain the following

Corollary 1.2. *Let $H = H^{(N)}$ be a sequence of hermitian $N \times N$ Wigner type matrices, with centered entries and matrix of variances $S_{xy} = S_{xy}^{(N)} = \mathbb{E}|H_{xy}|^2$. Assume that $S_{xy} \leq C^*/N$ for some constant C^* , independent of N . Further, we assume a finite moment condition on the matrix elements, i.e. that for any $q \in \mathbb{N}$ there is a constant C_q , independent of N , such that*

$$\max_{x,y,N} \mathbb{E}(\sqrt{N}|H_{xy}|)^q \leq C_q.$$

Set

$$z_j := \limsup_{N \rightarrow \infty} \frac{\|[S^{(N)}]^j\|}{\|S^{(N)}\|^j}$$

and for any $J \in \mathbb{N} \cup \{\infty\}$ let $w_c(J)$ be the smallest positive root of ϕ_J defined in (1.7). Then for any $\epsilon > 0$ (small) and any $D > 0$ (large) we have the following bound on the largest eigenvalue of H :

$$\mathbb{P}\left(|\lambda_{\max}(H^{(N)})| \geq 2 \frac{\|S^{(N)}\|^{\frac{1}{2}}}{w_c(J)} + \epsilon\right) \leq C(\epsilon, D)N^{-D}$$

for some constant $C(\epsilon, D)$ depending only on C^* and the sequence of constants C_q , in addition to ϵ and D .

Instead of Wigner type matrices, one may also consider Gram matrices, i.e., matrices of the form $H = XX^*$ where X is an $M \times N$ matrix with centered independent entries (without any symmetry condition) and $S_{ij} := \mathbb{E}|x_{ij}|^2$ is the matrix of variances. The spectral radius of H is the square of the spectral radius of the linearized matrix

$$\mathcal{X} := \begin{pmatrix} 0 & X \\ X^* & 0 \end{pmatrix} \quad \text{with variance matrix} \quad \mathcal{S} := \begin{pmatrix} 0 & S \\ S^t & 0 \end{pmatrix}.$$

Since $\mathcal{X}$ is a Wigner type matrix, Theorem 1.1 and its corollary directly applies. The norm $\|\mathcal{S}^j\|$ can be trivially expressed in terms of the norms of matrices of the form $SS^*SS^* \dots$ and $S^*SS^*S \dots$.

We remark that very similar questions were studied independently in a recent work of M. Ottolini [10] who derived a variational formula for $\max \text{supp } \rho$ and proved the convergence of the largest eigenvalue to it. This formula is exact, but not explicit in terms of S as it still requires to solve a variational problem. It is an open question to establish connections between the two approaches, especially find an explicit formula, if possible, in terms of S for the solution of Ottolini's variational problem.

We now explain the main novelty of our approach. We introduce a tree-graph expansion for representing the solution to (1.1). Unlike in the traditional proof of the Wigner semi-circle law via the moment method, in our case the graphs do not contribute equally; they are weighted by factors of S_{xy} assigned to edges. This defines an S -dependent measure $\mathcal{P}_S$ on the space of trees.

We then estimate the contribution of each tree by chopping it up into possibly long linear segments. Along the linear pieces, we can perform the summation $\sum_{xyz \dots uv} S_{xy} S_{yz} \dots S_{uv} = (S^j)_{xv}$ explicitly. This enables us to use the stronger bound $\|S^j\|$ instead of the trivial one $\|S\|^j$. We present an algorithm for a good chopping. We then compute the expected value of the corresponding contributions with respect to the measure $\mathcal{P}_S$. It turns out that the relevant regime is the limit as the size of the trees goes to infinite. In this limit we approximate the measure $\mathcal{P}_S$ by a Markov chain for the purpose of computing the weighted contributions of all graphs. The approximate Markov structure becomes apparent as we identify the tree graphs with Dyck paths. Finally, in the Markov model we can compute the answer explicitly.

2. TREES AND DYCK PATHS

We start with a simple observation that allows us to express $\max \text{supp } \rho$ in terms of the radius of convergence of the Laurent series expansion of the Stieltjes transform of ρ .

Let ρ be a compactly supported, symmetric probability measure on the real line with $\text{supp } \rho \subset [-r, r]$ for $r := \max \text{supp } \rho > 0$. Clearly its Stieltjes transform

$$m(z) := \int_{\mathbb{R}} \frac{\rho(d\tau)}{\tau - z}$$

is analytic on $\mathbb{C} \setminus \text{supp } \rho$. Its Laurent series is written as

$$m(z) = -\frac{1}{z} \sum_{k=0}^{\infty} \left(\frac{1}{z}\right)^k \mu_k, \quad \mu_k := \int_{\mathbb{R}} \tau^k \rho(d\tau).$$

By the Cauchy-Hadamard theorem on the radius of convergence of this power series we immediately obtain

$$\max \text{supp } \rho = \limsup_{k \rightarrow \infty} \mu_k^{1/k}. \quad (2.1)$$

We can apply the relation (2.1) not only for $m(z)$ defined in (1.2), but also for each $m_x(z)$ since it is the Stieltjes transform of some probability measure ρ_x (see, e.g. Theorem 2.1 [4]). Clearly ρ_x is also symmetric and $\rho = \frac{1}{N} \sum_x \rho_x$, thus

$$\text{supp } \rho = \bigcup_{x=1}^N \text{supp } \rho_x. \quad (2.2)$$

In particular, if $|z| > \max \text{supp } \rho$, then not only $m(z)$ has a convergent Laurent series, but each m_x as well:

$$m_x(z) = -\frac{1}{z} \sum_{k=0}^{\infty} \left(\frac{1}{z}\right)^k \mu_{x,k}, \quad \mu_{x,k} := \int_{\mathbb{R}} \tau^k \rho_x(d\tau). \quad (2.3)$$

Similarly, we have

$$\max \text{supp } \rho = \max_x \max \text{supp } \rho_x = \max_x \limsup_{k \rightarrow \infty} \mu_{x,k}^{1/k}. \quad (2.4)$$

To estimate $\limsup_{k \rightarrow \infty} \mu_{x,k}^{1/k}$, we will express $\mu_{x,k}$ in terms of sums of products of the matrix elements of S . Similarly to the standard proof of the Wigner semicircle law by the moment method (e.g. Section 2.1 [11]), we represent these sums diagrammatically, via an expansion in terms of Dyck paths. Since the paths are weighted by S , the estimate is not a simple combinatorial enumeration of the Dyck paths. We will see that these weights substantially distort the uniform counting measure on the set of Dyck paths. In the next sections we develop a formalism to bookkeep and effectively estimate these weights. In what follows we will use the notations

$$[a, b] = \{i \in \mathbb{N} : a \leq i \leq b\}, \quad [a, b) = \{i \in \mathbb{N} : a \leq i \leq b-1\}.$$

2.1. Dyck Path. We start with recalling the definition of the Dyck paths:

Definition 2.1. *Dyck paths of length $2k$ are paths $\pi : [0, 2k] \rightarrow \mathbb{N}$ such that $\pi(0) = \pi(2k) = 0$, $|\pi(i) - \pi(i+1)| = 1$. Denote the set of Dyck paths of length $2k$ by D_{2k} . We say that the i -th step is an *up-run* if $\pi(i) < \pi(i+1)$ and a *down-run* otherwise.*

Alternatively, Dyck paths encode algebraically legitimate bracketing of a product of $2k$ non-associative symbols in a straightforward manner. It is sufficient to bookkeep the brackets only. Thus we consider a string consisting of k opening and k closing brackets in such a way that for every *up-run* we append a "(" to the string, for every *down-run* append a ")".

We now recall the *tree representation* of the Dyck paths. Let $\mathcal{T}_k$ denote the set of planar, rooted, undirected trees $\Gamma = (V(\Gamma), E(\Gamma))$ with $|E(\Gamma)| = k$.

We always draw a tree in the plane in such a way that the vertices at the same distance from the root are drawn at the same height (horizontal level) relative to the root and the root is the lowest point (Figure 1). The height function is denoted by $h(v)$ for all $v \in V(\Gamma)$, $\Gamma = (V(\Gamma), E(\Gamma)) \in \mathcal{T}_k$; we set $h(\text{root}) = 0$. In this way every vertex v (apart from the root) has a unique *father*, i.e. an adjacent vertex of height $h(v) - 1$, and may have some *children*; these are adjacent vertices with height $h(v) + 1$, whose number we denote by $c(v)$. Unless v is a root, we have $c(v) = d(v) - 1$, where $d(v)$ is the *degree* of v , i.e. the number of adjacent vertices. Vertices with no children are called *leaves*.

Every edge $e \in E(\Gamma)$ has two vertices adjacent to it, denoted by $e_-, e_+ \in V(\Gamma)$, the sign indicating their relative height, i.e. $h(e_-) < h(e_+)$. We also extend the height function to edges by setting $h(e) := h(e_+)$.

The planarity imposes an orientation on every tree. In particular, it is possible to walk around the outer boundary of Γ (say, in clockwise direction) starting and arriving at the root. In this way, for any element Γ of $\mathcal{T}_k$ we can assign an element $\pi(\Gamma)$ of D_k in such a way that we set $\pi(i)$ to be the distance to the root at the i -th step of this walk. This map is clearly a bijection for each fixed k . We define $\Gamma : \bigcup_{k=0}^{\infty} D_{2k} \rightarrow \bigcup_{k=0}^{\infty} \mathcal{T}_k$ to be the inverse of this map. For any Dyck path π , we call $\Gamma(\pi)$ “the tree corresponding to π ” (see Figure 1 for an example).

A finite collection of several disjoint trees is called *forest*. The set of forests with a total of k edges is denoted by $\mathcal{F}_k := \{\{\Gamma_i\}_i : \Gamma_i \in \mathcal{T}_{m_i} \text{ with } \sum_i m_i = k\}$. Every component Γ_i has a single root which is drawn as its lowest vertex. For any forest Γ , the set of roots is denoted by $R(\Gamma) \subseteq V(\Gamma)$. For general forests not all roots will be drawn at the same horizontal level. The vertices are drawn as “bullets” with the convention that roots are unfilled ($\circ$) and filled ($\bullet$) otherwise.

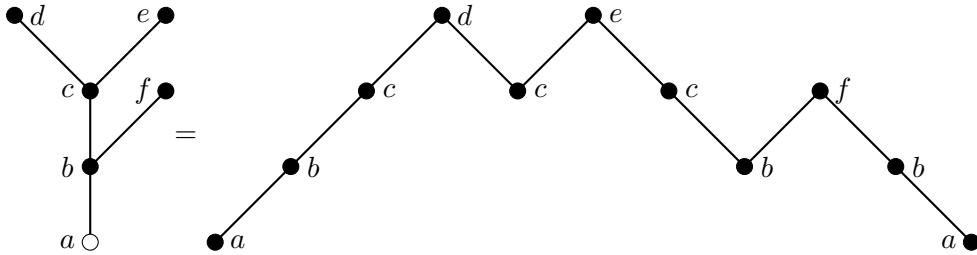


FIGURE 1. Tree and Dyck path representation of $\Gamma = \Gamma("((()())())") \in \mathcal{T}_5$.

2.2. Graphical Representation of the Dyck Path Expansion. We now introduce a graphical representation to rewrite $\mu_{x,k}$. The same expansion in a slightly different presentation was also used in [10].

Lemma 2.2 (Graphical representation of $m_x(z)$). *For every $\Gamma = (V(\Gamma), E(\Gamma)) \in \mathcal{T}_k$, and for every $x \in \{1, \dots, N\}$ set*

$$val_x(\Gamma) := \left(\prod_{v \in V(\Gamma)} \sum_{x_v=1}^N \right) \left[\delta_{x_{root}=x} \prod_{e \in E(\Gamma)} S_{x_{e_-} x_{e_+}} \right]. \quad (2.5)$$

Then for the Laurent series (2.3) we have

$$m_x(z) = -\frac{1}{z} \sum_{k=0}^{\infty} \left[\sum_{\Gamma \in \mathcal{T}_k} val_x(\Gamma) \right] z^{-2k}, \quad |z| > \max \text{supp } \rho_x. \quad (2.6)$$

Proof. Introduce $u_x(z) := -zm_x(z)$ and note that the QVE (1.1) is equivalent to

$$u_x = 1 + z^{-2} \sum_{y=1}^N S_{xy} u_x u_y. \quad (2.7)$$

Using (2.3), $u_x(z)$ admits a Laurent series expansion for $|z|$ large enough. By the symmetry of the measure ρ_x , the odd coefficients vanish and with $c_{x,k} := \mu_{x,2k}$ we have

$$u_x(z) = \sum_{k=0}^{\infty} c_{x,k} z^{-2k}, \quad (2.8)$$

for large $|z|$. Now plugging (2.8) into (2.7) and comparing coefficients we get the following recursion:

$$c_{x,k} = \sum_{y=1}^N S_{xy} \sum_{n=0}^{k-1} c_{x,k-n-1} c_{y,n}. \quad (2.9)$$

To show that $c_{x,k} = \sum_{\Gamma \in \mathcal{T}_k} val_x(\Gamma)$ we proceed by induction on k . The base case is clear. Assume that for all n with $0 \leq n < k$ we already know that

$$c_{x,n} = \sum_{\Gamma \in \mathcal{T}_n} val_x(\Gamma) \quad \text{for } x = 1, \dots, N. \quad (2.10)$$

We identify every $\Gamma \in \mathcal{T}_k$ with its Dyck path $\pi(\Gamma) \in D_{2k}$. We define for every $\Gamma \in \mathcal{T}_k$ the numbers $n_1 = n_1(\Gamma)$, $n_2 = n_2(\Gamma)$ by

$$2n_1 := \max\{t \in [0, 2k) : \pi(\Gamma)(t) = 0\},$$

and $n_2 := k - n_1 - 1$. Now every $\Gamma \in \mathcal{T}_k$ can *uniquely* be written as $\Gamma = \Gamma_1 \oplus \Gamma_2$ with $\Gamma_1 \in \mathcal{T}_{n_1}$, $\Gamma_2 \in \mathcal{T}_{n_2}$, where we define $\Gamma_1 \oplus \Gamma_2$ via its Dyck path representation as follows:

$$\pi(\Gamma_1 \oplus \Gamma_2)(t) := \begin{cases} \pi(\Gamma_1)(t), & \text{if } t = 0, 1, \dots, 2|E(\Gamma_1)| \\ \pi(\Gamma_2)(t) + 1, & \text{if } t = 2|E(\Gamma_1)| + 1, \dots, 2|E(\Gamma_1)| + 2|E(\Gamma_2)| + 1 \\ 0, & \text{if } t = 2|E(\Gamma_1)| + 2|E(\Gamma_2)| + 2. \end{cases}$$

See Figure 2 for an illustration. By definition (2.5), we have

$$\sum_{y=1}^N S_{xy} val_x(\Gamma_1) val_y(\Gamma_2) = val_x(\Gamma_1 \oplus \Gamma_2). \quad (2.11)$$

Plugging in (2.10) into (2.9) and using (2.11) we see that

$$c_{x,k} = \sum_{y=1}^N S_{xy} \sum_{n=0}^{k-1} \text{val}_x(\Gamma_1) \text{val}_y(\Gamma_2) \stackrel{(2.11)}{=} \sum_{n=0}^{k-1} \sum_{\Gamma_1 \in \mathcal{T}_n} \sum_{\Gamma_2 \in \mathcal{T}_{k-n-1}} \text{val}_x(\Gamma_1 \oplus \Gamma_2). \quad (2.12)$$

Since for every $\Gamma \in \mathcal{T}_k$ there exists exactly one pair $n_1(\Gamma), n_2(\Gamma)$ such that $\Gamma = \Gamma_1 \oplus \Gamma_2$ with uniquely determined $\Gamma_i \in \mathcal{T}_{n_i}$, we have that (2.12) is just $\sum_{\Gamma \in \mathcal{T}_k} \text{val}_x(\Gamma)$. $\square$

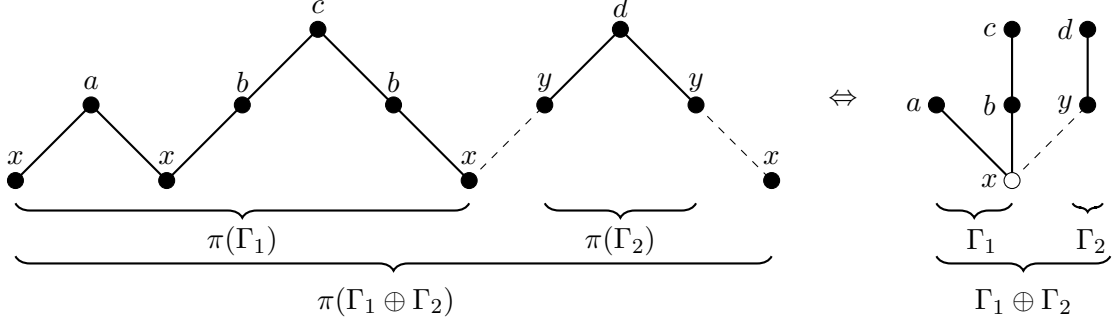


FIGURE 2. Illustration of $\pi(\Gamma_1 \oplus \Gamma_2)$ and $\Gamma_1 \oplus \Gamma_2$.

Combining (2.4) with Lemma 2.2 we get

$$\max \text{supp } \rho \leq \limsup_k \left| \sum_{\Gamma \in \mathcal{T}_k} \text{val}(\Gamma) \right|^{\frac{1}{2k}}, \quad (2.13)$$

for

$$\text{val} := \max_{x=1, \dots, N} \text{val}_x. \quad (2.14)$$

Remark 2.3 (Obtaining the already known bound $2\|S\|^{\frac{1}{2}}$). *Ignoring the internal structure of the trees $\Gamma \in \mathcal{T}_k$ and successively summing up the labels starting from the leaves of the tree, using $\sum_y S_{xy} \leq \|S\|$, we can bound $\text{val}(\Gamma)$ by $\|S\|^k$. Noting that $|\mathcal{T}_k| = |D_{2k}| = O(2^{2k})$ as $k \rightarrow \infty$ we easily get the claimed bound from (2.13).*

2.3. Intuition for improvement. We explain in a simple example how to improve the previous trivial bound. For example, for Γ as in Figure 1 we have

$$\text{val}(\Gamma) = \max_x \sum_{yzuvw} S_{xy} S_{yz} S_{zu} S_{zv} S_{yw}.$$

Instead of simply bounding it by $\|S\|^5$ we could bound it by

$$\text{val}(\Gamma) \leq \max_x \sum_{yzu} S_{xy} S_{yz} S_{zu} \max_{z'} \sum_v S_{z'v} \max_{y'} \sum_w S_{y'w} \leq \|S^3\| \|S\|^2 \quad (2.15)$$

or by

$$\text{val}(\Gamma) \leq \max_x \sum_{yw} S_{xy} S_{yw} \max_{y'} \sum_{zv} S_{y'z} S_{zv} \max_{z'} \sum_u S_{z'u} \leq \|S^2\|^2 \|S\|, \quad (2.16)$$

both of which are less or equal than $\|S\|^5$ since the norm is submultiplicative.

It is easy to see that such a process always gives a bound of the form $\text{val}(\Gamma) \leq \prod_{i \leq k} \|S^i\|^{p_i}$ with some sequence of natural numbers $(p_i)_i$ such that $\sum_i ip_i = k$, depending on Γ . For certain Γ 's the improvement over the trivial bound is meagre or even non-existent. While we have a certain freedom in “chopping up” the multiple summation for $\text{val}(\Gamma)$, in general we cannot obtain all bounds of the form $\text{val}(\Gamma) \leq \prod_{i \leq k} \|S^i\|^{p_i}$. Our current choice of Γ , for example, does not admit the bounds $\|S^4\| \|S\|$ or $\|S^5\|$; these would require a path of length 4 or 5, respectively, from the root. The worst case for $k = 5$, the tree corresponding to $\pi = ()()()()$ (every node is connected to the root), does not admit any bound other than $\|S\|^5$.

In the next sections we first formalise the above process of chopping up val . Then we quantify how trees between the two extreme cases (the completely linear tree with $d(v) \leq 2$ for every vertex v and the tree where $d(v_{\text{root}}) = k$) typically look like for large k and which chopping up gives the “best” (smallest) weight.

It is also worth noting that for general S we cannot say that one bound is always better than the other one. Depending on S either $\|S^3\| \|S\|^2$ or $\|S^2\|^2 \|S\|$ may be preferable. In what follows we simply choose a fixed method (independent of S) to obtain our bound.

2.4. The Chopping Up Process.

2.4.1. Introducing the chopping-up operation. Recall that $\mathcal{F}_k$ denotes the set of forests (collection of rooted trees drawn in the plane according to the convention of Section 2.1) with k edges in total.

Chopping up is an operation $\mathcal{F}_k \rightarrow \mathcal{F}_k$ for every k where some vertices of $\Gamma \in \mathcal{F}_k$ are *split* but the edge set remains unchanged. *Splitting of a vertex v* is an operation that creates a few new copies of v and disconnects some (or all) edges emanating from v in the upward direction in such a way that these edges will emanate from a new copy of v . The new vertices (called *copies* of v in the splitting), together with the original v that is kept, are drawn next to each other in an oriented fashion to keep the planarity of the graph, see Figure 3 for possible splittings of the tree in Figure 1. In particular all copies

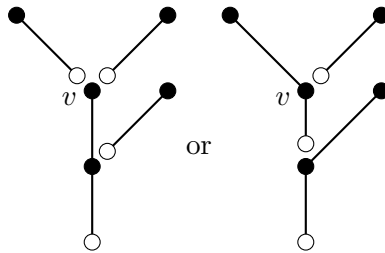


FIGURE 3. Two possible splittings of the tree in Figure 1. The vertex v is completely split in the left graph, and it is split in the leftmost way (almost completely) in the right graph.

are drawn at the same horizontal level as v . There is at most a single edge emanating from v downwards (connecting v to its father); this edge will never be separated from the original v . Some children of v , however, may disconnect from v and connect instead to a

copy of v . These copies become the roots of a new component. Thus the vertex v remains filled or unfilled, but all new copies of v will be unfilled, and they are the lowest point of their connected component (in the new graph). Splitting is applied only to vertices with degree at least two (we do not split leaves or roots with only one child).

If Γ' is obtained by chopping up Γ , then we indicate this fact by $\Gamma \prec \Gamma'$. We clearly have transitivity, i.e. if $\Gamma, \Gamma', \Gamma'' \in \mathcal{F}_k$ with $\Gamma \prec \Gamma'$ and $\Gamma' \prec \Gamma''$, then

$$\Gamma \prec \Gamma''. \quad (2.17)$$

Now we define a particularly useful subset of possible splittings:

Definition 2.4. Fix any vertex v of $\Gamma \in \mathcal{F}_k$ with number of children $c(v) \geq 1$. A splitting at the vertex v is called *complete* if it yields $c(v)$ copies.

A splitting at v which yields $c(v) - 1$ copies is called *almost complete*. In this latter case, if the edge connecting v to its remaining child was the *leftmost** one (out of all the edges connecting v to its children), we call the splitting *leftmost* and if it was the *rightmost* one, we call the splitting *rightmost*.

We call a forest Γ *linear*, if graph-theoretically it is a union of paths, i.e. the degree of every vertex is at most two. Notice that we obtain a linear Γ' in chopping up Γ , if we split every vertex v of Γ with $c(v) \geq 1$ either *completely* or *almost completely*.

2.4.2. Monotonicity of val along chopping. We now extend the previously introduced concept of a *value* from trees ((2.5), (2.14)) to any $\Gamma \in \mathcal{F}_k$.

Definition 2.5. Every vertex v gets a label $x_v \in \{1, \dots, N\}$ and labels assigned to non-root vertices are summed up, while we take the maximum over labels assigned to roots:

$$\text{val}(\Gamma) := \left(\prod_{v \in R(\Gamma)} \max_{x_v} \right) \left(\prod_{v \in V(\Gamma) \setminus R(\Gamma)} \sum_{x_v=1}^N \right) \left[\prod_{e \in E(\Gamma)} S_{x_{e_-} x_{e_+}} \right]. \quad (2.18)$$

where $R(\Gamma) \subseteq V(\Gamma)$ is the set of roots.

Note that this coincides with (2.14) for trees, i.e. when there is only one root. See Figure 4 for some chopped up trees and their values.

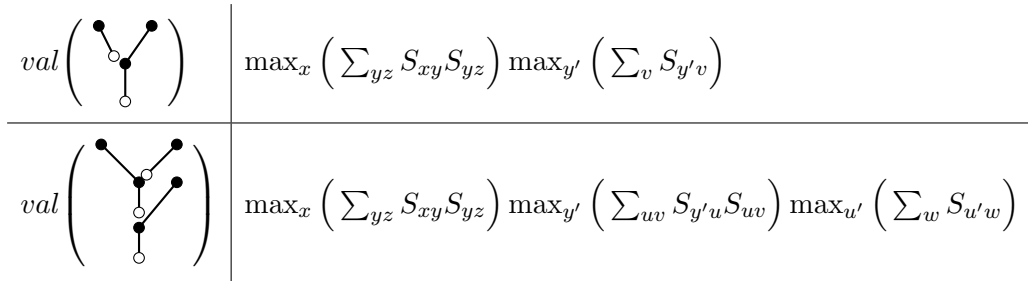


FIGURE 4. Chopped-up trees and their values.

*Since the forest Γ is drawn in the plane in a specific way, right and left are meaningful concepts.

Lemma 2.6. *The value function is monotonic along the chopping up operation:*

$$val(\Gamma) \leq val(\Gamma') \quad \text{if} \quad \Gamma \prec \Gamma'. \quad (2.19)$$

Proof. The proof of this statement is an easy induction on subsequent splitting of vertices. It is based upon the trivial estimate

$$\max_a \sum_{b,c} S_{ab} S_{bc} \leq \max_a \sum_b S_{ab} \max_{b'} \sum_c S_{b'c} = \|S\|^2, \quad (2.20)$$

which, in our graphical language, can also be written as:

$$val\left(\begin{array}{c} \bullet \\ | \\ \bullet \\ | \\ \bullet \\ | \\ \circ \end{array}\right) \leq val\left(\begin{array}{c} \bullet \\ | \\ \circ \\ | \\ \bullet \\ | \\ \circ \end{array}\right)$$

Here the original vertex with label b was split, the copy received a new label b' . For more complicated graphs the proof is similar. $\square$

2.4.3. Bounding $val(\Gamma)$ using the chopping process. Now we will fix $N, k \in \mathbb{N}$, $\Gamma \in \mathcal{T}_k$ and its corresponding Dyck path $\pi = \pi(\Gamma)$. We will chop up Γ , i.e. construct a linear chopped-up tree $\Gamma' \in \mathcal{F}_k$ with $\Gamma \prec \Gamma'$ with the minimal amount of chopping.

As an example, Figure 5 shows the leftmost and rightmost almost complete splittings. They give rise to the bounds (2.15) and (2.16), respectively.

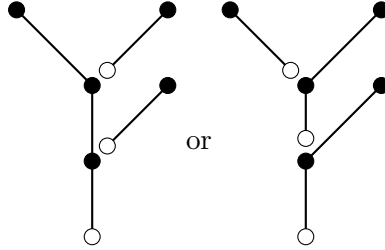


FIGURE 5. Splitting up Γ from Figure 1 in the leftmost and the rightmost way.

Remark 2.7 (Structure of bounds from the Dyck path representation of Γ). *Notice that with the leftmost choice, we made a gain on the monotonically increasing parts (consecutive up-runs) of the corresponding Dyck path π , while with the rightmost choice we gained on the monotonically decreasing parts (consecutive down-runs). Here “gain” means that we did not chop up the corresponding monotonic segments into pieces of length one; this allowed us to use the norms of higher powers of S instead of trivially estimating them by higher powers of $\|S\|$.*

In what follows we want to quantify this gain. Recall the definition of z_j from (1.5). For any sequence $T = (T_1, T_2, \dots, T_J)$ of nonnegative integers we set the notation

$$z^T := \prod_{j=1}^J z_j^{T_j}.$$

The upper cutoff J is a fixed parameter in Theorem 1.1. Note that T_1 does not influence z^T since $z_1 = 1$.

Define for any fixed $J \in \mathbb{N}$ and any path π (i.e. any sequence $\pi = (\pi(i))_{i \in [a,b]}$ with $\pi(i) \in \mathbb{N}$, $|\pi(i+1) - \pi(i)| = 1$) the J -tuples $U(\pi)$ and $D(\pi)$ by

$$U(\pi)_j := \#\{\text{up-runs of length } j\} \quad \text{and} \quad D(\pi)_j := \#\{\text{down-runs of length } j\} \quad (2.21)$$

for $j = 1, 2, \dots, J$. The observation from Remark 2.7 proves the following:

Lemma 2.8. *Let $\Gamma \in \mathcal{T}_k$ and let $\pi = \pi(\Gamma)$ be the Dyck path corresponding to Γ . With the rightmost choice we get*

$$\frac{\text{val}(\Gamma)}{\|S\|^k} \leq z^{U(\pi)}, \quad (2.22)$$

while the leftmost choice gives

$$\frac{\text{val}(\Gamma)}{\|S\|^k} \leq z^{D(\pi)}, \quad (2.23)$$

□

In our concrete case (compare Figures 1 and 5) we have $U = (2, 0, 1, 0, \dots)$ and $D = (1, 2, 0, \dots)$.

2.4.4. *Introducing our choice of chopping up Γ .* We will need a mixture of the two estimates (2.22) and (2.23), and it will be more convenient to work with the Dyck path $\pi = \pi(\Gamma) \in D_{2k}$ corresponding to $\Gamma \in \mathcal{T}_k$. Namely, above a certain threshold height ℓ we will to follow the rightmost choice, below that level the leftmost choice (see Figure 6 for a naive sketch).

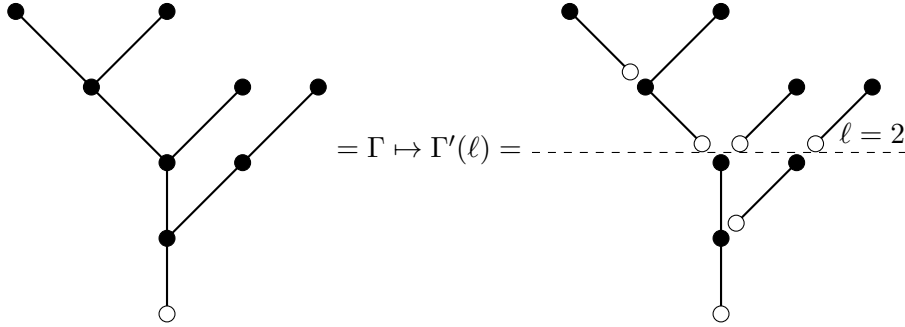


FIGURE 6. Graph splitting at level $\ell = 2$ (Above ℓ : rightmost splitting, below ℓ : leftmost splitting, at ℓ : complete splitting.)

Moreover, this choice will be determined not by the actual height of the vertex, but by the height of π at certain coarse-grained cutoff times in order to avoid that the rightmost and leftmost choices alternate too often. These requirements necessitate a slightly more refined construction.

Choose a (small) parameter ε and define the sequence of *cutoff times*

$$t_j := \lfloor 2k\varepsilon j \rfloor, \quad j = 0, 1, \dots, 1/\varepsilon$$

(we assume that $1/\varepsilon$ is an integer). These cutoff times naturally split any path π into $1/\varepsilon$ segments of equal^{*} length, i.e.

$$\pi = \pi_{[0,2k]} = \bigcup_{j=0}^{1/\varepsilon-1} \pi_j, \quad \pi_j := \pi_{[t_j, t_{j+1}]}.$$

Even though π_j is defined as the restriction of π , with a slight abuse of notation we will shift its argument starting at 0, i.e. strictly speaking $\pi_j = \pi_{[t_j, t_{j+1}]} \circ s_{-t_j}$ where $s_{-t_j}(t) := t - t_j$ is the shift operator. The height $\pi(t_j)$ of the beginning of each segment will be called the j -th *cutoff height*. For every integer $i \in [0, 2k]$ there is a unique j such that $i \in [t_j, t_{j+1})$ and the *cutoff height of i* is defined to be $\pi(t_j)$, i.e. the cutoff height of any index i is determined by the initial point of its segment.

Given the parameters $\varepsilon > 0$, $\ell > 1$, $\ell \in \mathbb{N}$, and given a $\Gamma \in \mathcal{T}_k$, and hence the corresponding Dyck path $\pi \in D_{2k}$, we now define a specific chopped-up graph $\Gamma'(\varepsilon, \ell)$ with $\Gamma \prec \Gamma'(\varepsilon, \ell)$.

First we define a subset R of the (integer) time variables in $[0, 2k]$ as follows:

$$R := \bigcup_{j=0}^{1/\varepsilon-1} R_j,$$

with

$$R_j := \begin{cases} \{i \in [t_j, t_{j+1}) : \pi(i) < \pi(i+1)\} & \text{if } \pi(t_j) \leq 2k\varepsilon(\ell-1) \\ \{i \in [t_j, t_{j+1}) : \pi(i) > \pi(i+1)\} & \text{if } \pi(t_j) \geq 2k\varepsilon(\ell+1), \\ \emptyset & \text{if } 2k\varepsilon(\ell-1) < \pi(t_j) < 2k\varepsilon(\ell+1). \end{cases}$$

The set R contains those times $i \in \mathbb{N}$ when the path goes upwards whenever its cutoff height is below the lower threshold $2k\varepsilon(\ell-1)$ as well as those times when the path goes downwards if the cutoff height is above the upper threshold $2k\varepsilon(\ell+1)$. For any $i \in R$, we *tag* the edges between $(i, \pi(i))$ and $(i+1, \pi(i+1))$ of the Dyck path and draw them bold (see Figure 7).

Moreover, define

$$P := \{i \in R : (\pi(i+1) - \pi(i))(\pi(i) - \pi(i-1)) > 0\},$$

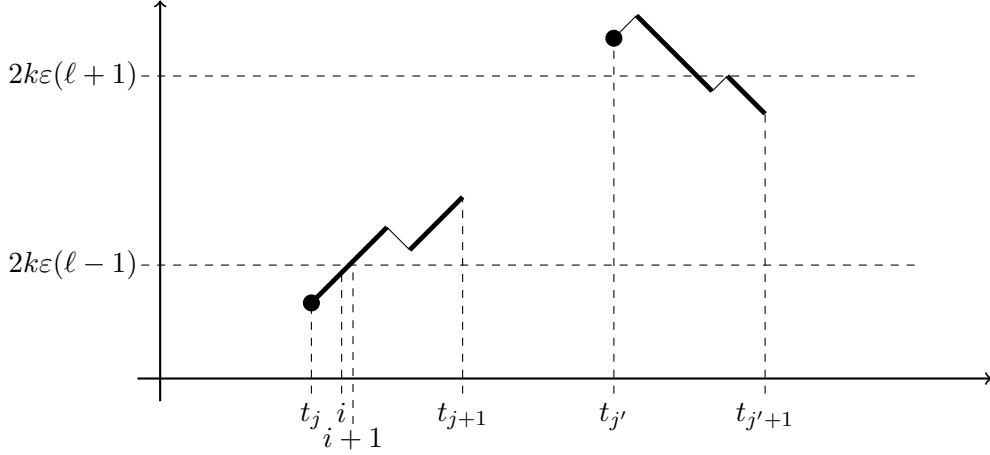
i.e. these are the indices $i \in R$ such that $(i, \pi(i))$ is in the middle of a monotonic segment of length at least 2. In particular, we have the following property:

Lemma 2.9 (Property **P**). *If $i \in P$, then both edges of the Dyck path adjacent to the point $(i, \pi(i))$ are tagged. Moreover, if $i \in [t_j, t_{j+1})$ for some j , then exactly one of the following two options holds:*

- *either: $\pi(t_j) \leq 2k\varepsilon(\ell-1)$ and $\pi(i-1) < \pi(i) < \pi(i+1)$*
- *or: $\pi(t_j) \geq 2k\varepsilon(\ell+1)$ and $\pi(i-1) > \pi(i) > \pi(i+1)$.*

□

^{*}Equal up to ± 1 , which will not matter as $k \rightarrow \infty$. Henceforth we will assume that $t_j \stackrel{(\text{def})}{=} \lfloor 2k\varepsilon j \rfloor = 2k\varepsilon j$.

FIGURE 7. Sketch for tagged (bold) edges defined by R .

This construction (together with the fact that the maximal height difference within each segment π_j is at most $2k\varepsilon$) implies the following observations:

- (i) Below a security layer of width $4k\varepsilon$ around the fixed level $2k\varepsilon\ell$ all up-runs of π are tagged, above the security layer all down-runs of π are tagged.
- (ii) At any given level all tagged edges are of the same type (upward or downward).
- (iii) The choice whether the up-runs or the down-runs are tagged is decided at the cutoff times t_j and this choice is valid for the entire path segment π_j .

Now we are ready to define the chopped up graph $\Gamma'(\varepsilon, \ell)$ that we will actually use.

Definition 2.10 (Definition of $\Gamma'(\varepsilon, \ell)$). *Fix $k \in \mathbb{N}$, $\Gamma \in \mathcal{T}_k$, and let $\pi = \pi(\Gamma)$ be the corresponding Dyck path. We define $\Gamma'(\varepsilon, \ell)$ by the following procedure that determines how we split the vertices $V(\Gamma)$:*

We walk around Γ starting from the root in clockwise direction and we successively mark all edges to be split either completely or almost completely in the leftmost or the rightmost way (but we do not split them yet). The marking is determined by the following rules:

Step 1: *We mark the root for leftmost splitting.*

Step 2: *Now consider the i -th step (for $i > 0$, the root has been dealt with) and fix j such that $i \in [t_j, t_{j+1})$. Let $v(i) \in V(\Gamma)$ denote the vertex reached at the i -th step.*

- (a) *If $i = t_j$, $\pi(t_j) \leq 2k\varepsilon(\ell - 1)$, and $v(i)$ is visited for the first time^{*}, then we mark the vertex $v(i)$ for complete splitting.*
- (b) *If $i = t_j$, $\pi(t_j) \geq 2k\varepsilon(\ell + 1)$, and $v(i)$ is visited for the last time, then we also mark $v(i)$ for complete splitting.*
- (c) *If $i \neq t_j$, $i \in P$, then we mark $v(i)$ for almost complete splitting either in the leftmost or the rightmost way, depending on whether $\pi(t_j) \leq 2k\varepsilon(\ell - 1)$ or $\pi(t_j) \geq 2k\varepsilon(\ell + 1)$, respectively.*

^{*}I.e. there is no $j < i$ with $v(j) = v(i)$.

Step 3: Consider all vertices that were left unmarked* in Step 1 and mark them for complete splitting. Finally, we perform the prescribed splittings.

Lemma 2.11. *The procedure described in Definition 2.10 is well-defined, i.e. every vertex receives an unambiguous marking.*

Proof. The root is always marked for leftmost splitting by Step 1. Since the root is visited the first time at $i = 0$, rule (a) together with $i > 0$ does not mark the root. The conditions of rule (b) also exclude the root since for the root $\pi(i) = 0$. Finally, rule (c) applies only to vertices in the middle of a monotonic segment ($i \in P$), hence it does not apply to the root either, thus Step 1 is not in conflict with Step 2.

Now consider the vertices visited at times $i = t_j$ for some j . It is easy to see from the definition of P there is no other time $i' \neq i$ with $v(i') = v(t_j)$ s.t. (c) marks $v(i')$, hence there is no conflict between (c) and (a), (b). There is no conflict between (a) and (b) due to the mutually exclusive conditions on $\pi(t_j)$.

It remains to show that rule (c) is applied to the same vertex $v(i)$ at most once. When walking around Γ , the same vertex $v \in V(\Gamma)$ is visited several (even number of) times, say $v(i_1) = v(i_2) = \dots = v(i_{2m}) = v$. However, we claim that only at most one of the time indices $i_1, i_2, \dots, i_{2m}$ can be in P (and if v is the root, then clearly none can be in P); in other words, the above procedure triggers a splitting of v at most at one of the times $i_1, i_2, \dots, i_{2m}$. Assuming to the contrary that there exist $i, i' \in P$ with $v(i) = v(i')$ and $i \neq i'$, then by Property P both $(i, \pi(i))$ and $(i', \pi(i'))$ are joining two marked edges of π of the same monotonicity type. By the observation (ii) above, either both are increasing: $\pi(i-1) < \pi(i) < \pi(i+1)$ and $\pi(i'-1) < \pi(i') < \pi(i'+1)$, or both are decreasing: $\pi(i-1) > \pi(i) > \pi(i+1)$ and $\pi(i'-1) > \pi(i') > \pi(i'+1)$. However, the construction of the graph Γ from π excludes $v(i) = v(i')$ in both cases, which is a contradiction, proving the original claim. In particular, to every v that is split by rule (c) (but not by (a) or (b)) along the procedure above, there is a unique time $i = i_v$, when it was split. $\square$

One may also arrive at $\Gamma'(\varepsilon, \ell)$ as follows. Split the root almost completely in the leftmost way. Consider any $i > 0$ and fix $j = j(i)$ such that $i \in [t_j, t_{j+1})$. We first split almost completely all those vertices $v(i) \in V(\Gamma)$ where $h(t_j) \notin [2k\varepsilon(\ell-1), 2k\varepsilon(\ell+1)]$ that are either in an up-run, i.e. $\pi(i-1) < \pi(i) < \pi(i+1)$, in case $\pi(t_j) \leq 2k\varepsilon(\ell-1)$ or in a down-run, i.e. $\pi(i-1) > \pi(i) > \pi(i+1)$, in case $\pi(t_j) \geq 2k\varepsilon(\ell+1)$. The corresponding splitting is in the leftmost or the rightmost way, respectively. Next we split some of these vertices even further, namely those vertices $v(t_j)$ that are in an up-run (if $\pi(t_j) \leq 2k\varepsilon(\ell-1)$) and those in a down-run (if $\pi(t_j) \geq 2k\varepsilon(\ell+1)$) we split completely. Finally, in the last step, all unsplit vertices are split completely.

Lemma 2.12. *Fix ε, ℓ . Let $\pi \in D_{2k}$ and consider its j -th subpath $\pi_j \stackrel{(def)}{=} \pi_{[t_j, t_{j+1}]}$. Set*

$$\tilde{T}_{\ell, \varepsilon}(\pi) := \sum_{j=0}^{1/\varepsilon-1} \left[U(\pi_j) \cdot \mathbf{1}(\pi(t_j) \leq 2k\varepsilon(\ell-1)) + D(\pi_j) \cdot \mathbf{1}(\pi(t_j) \geq 2k\varepsilon(\ell+1)) \right]. \quad (2.24)$$

*Note that, in particular, this was the case for $i \in [t_j, t_{j+1})$ such that $2k\varepsilon(\ell-1) < \pi(t_j) < 2k\varepsilon(\ell+1)$.

Let $\Gamma'(\varepsilon, \ell)$ be the chopping-up of the tree $\Gamma = \Gamma(\pi)$ given in Definition 2.10. Then we have

$$\frac{\text{val}(\Gamma)}{\|S\|^k} \leq z^{\tilde{T}_{\ell, \varepsilon}(\pi)}. \quad (2.25)$$

Proof. Recalling that by Definition 2.10 $\Gamma'(\varepsilon, \ell)$ is a set of linear trees, we have

$$\frac{\text{val}(\Gamma)}{\|S\|^k} \stackrel{(2.19)}{\leq} \frac{\text{val}(\Gamma'(\varepsilon, \ell))}{\|S\|^k} \leq \prod_{l \in \Gamma'(\varepsilon, \ell)} z_{|E(l)|}, \quad (2.26)$$

where $|E(l)|$ denotes the number of edges in the linear tree l . The second inequality in (2.26) holds since by the definition in (2.18) the value of a forest is just the product over values of its (tree) components and the value of a linear tree of length n is estimated by $\|S^n\| \stackrel{(\text{def})}{=} z_n \|S\|^n$.

The r.h.s. of (2.26) is equal to $z^{\tilde{T}_{\ell, \varepsilon}(\pi)}$. To see this, note that rules (a) and (b) allow us to consider the subpaths π_j independently of each other (like in the definition of $\tilde{T}_{\ell, \varepsilon}$) and rule (c) ensures we do not “overcount”. Note that every edge of $\Gamma'(\varepsilon, \ell)$ gives rise to exactly two edges of π (at the same height). By not overcounting we mean that we need to make sure to use at most one of these two edges in $\tilde{T}_{\ell, \varepsilon}$. This is obvious since one of these edges is going up, one is going down and the characteristic functions in (2.24) take only segments going up *or* going down at any given height into account. $\square$

Instead of working with $\tilde{T}_{\ell, \varepsilon}$, we would prefer to work with something more tractable without technical restrictions of security layers, similar to what we sketched in Figure 6. To this end we define the J -sequences with a threshold at $2k\varepsilon\ell$ without security zone, i.e. we set

$$T_{\ell, \varepsilon}(\pi) := \sum_{j=0}^{1/\varepsilon-1} T_{\ell, \varepsilon}(\pi_j), \quad T_{\ell, \varepsilon}(\pi_j) := U(\pi_j) \cdot \mathbf{1}(\pi(t_j) \leq 2k\varepsilon\ell) + D(\pi_j) \cdot \mathbf{1}(\pi(t_j) > 2k\varepsilon\ell).$$

We also define

$$\Delta_{\ell, \varepsilon}(\pi) := \#\{j \in [0, 1/\varepsilon) : |\pi(t_j) - 2k\varepsilon\ell| \leq 2k\varepsilon\},$$

the number of cutoff times when the cutoff height is close to the threshold $2k\varepsilon\ell$.

We introduce the shorthand notation to denote the expectation w.r.t. the uniform measure on $\mathcal{T}_k$

$$\mathbb{E}f(\Gamma) := \frac{1}{|\mathcal{T}_k|} \sum_{\Gamma' \in \mathcal{T}_k} f(\Gamma'),$$

for any $f : \mathcal{T}_k \rightarrow \mathbb{R}$. We use a similar convention for $f : D_{2k} \rightarrow \mathbb{R}$, using the bijection between D_{2k} and $\mathcal{T}_k$.

Now we quantify at what cost we can consider $z^{T_{\ell, \varepsilon}}$ instead of $z^{\tilde{T}_{\ell, \varepsilon}}$ in (2.25). Given $\Gamma \in \mathcal{T}_k$, its corresponding Dyck path $\pi \in D_{2k}$, $\varepsilon > 0$, $\ell > 1$ integer, we clearly have

$$\frac{\text{val}(\Gamma)}{\|S\|^k} \leq \frac{z^{T_{\ell, \varepsilon}(\pi)}}{(\min_{j \leq J} z_j^{1/j})^{\Delta_{\ell, \varepsilon}(\pi) 2k\varepsilon}}. \quad (2.27)$$

For any fixed ℓ there are some Γ for which the bound given by (2.27) is very bad. Namely, if most $\pi(t_j)$ are close to $2k\varepsilon\ell$, i.e. the path spends a lot of time in the security layer, then $\Delta_{\ell,\varepsilon}$ is large and the estimate (2.27) is weak. To prevent this, we will choose the security layer depending on the path in a coarse-grained fashion in the next lemma.

Lemma 2.13. *Fix two parameters L and M , then*

$$[\mathbb{E} \text{val}(\Gamma)]^{1/2k} \leq \frac{L^{1/2k}}{(\min_{j \leq J} z_j^{1/j})^{1/L}} \|S\|^{1/2} \cdot \max_{m \leq L} [\mathbb{E} z^{T_{M+2m,\varepsilon}(\pi)}]^{1/2k}. \quad (2.28)$$

Proof. By the pigeonhole principle, for any fixed $\pi \in D_{2k}$, there exists an $\ell = \ell(\pi)$ of the form $\ell = M + 2m$ with $m \in \{1, 2, \dots, L\}$ such that

$$\Delta_{\ell,\varepsilon}(\pi) \leq \frac{1}{\varepsilon L}.$$

Choosing this $\ell = \ell(\pi)$ in the estimate (2.27), we obtain

$$\frac{\text{val}(\Gamma)}{\|S\|^k} \leq \frac{z^{T_{\ell(\pi),\varepsilon}(\pi)}}{(\min_{j \leq J} z_j^{1/j})^{2k/L}}.$$

Summing up for all possible values of $\ell(\pi)$, we obtain the following bound:

$$\frac{\text{val}(\Gamma)}{\|S\|^k} \leq \frac{1}{(\min_{j \leq J} z_j^{1/j})^{2k/L}} \sum_{m=1}^L z^{T_{M+2m,\varepsilon}(\pi)}.$$

Now we take expectation $\mathbb{E}$ and take the $2k$ -th root, to finish the proof. $\square$

We will take the limits in the following order

$$\lim_{L,M} \lim_{\varepsilon} \limsup_k$$

which makes the prefactor in (2.28) one, hence negligible. So it is sufficient to estimate

$$\lim_{\varepsilon} \limsup_k \mathbb{E} z^{T_{\ell,\varepsilon}(\pi)} = \lim_{\varepsilon} \limsup_k \mathbb{E} \prod_{j=0}^{1/\varepsilon-1} z^{T_{\ell,\varepsilon}(\pi_j)} \quad (2.29)$$

for $\ell = M + 2m$ fixed.

3. DISTRIBUTION OF DYCK PATHS

The uniform measure on the set of Dyck paths is equivalent to an inhomogeneous Markov chain that we describe now.

Fix k and note that because of the boundary conditions no Dyck path can leave the triangle

$$\Delta_k := \{(t, h) \in \mathbb{N}^2 : h \leq t \text{ and } h \leq 2k - t\}.$$

Introduce furthermore

$$\begin{aligned} \Delta_{k,\delta} &:= \Delta_k \cap \{(t, h) : t + 2k\varepsilon \leq 2k(1 - \delta)\} & \Delta_{k,\delta}^{top} &:= \Delta_{k,\delta} \cap \{(t, h) : h > 2k\varepsilon\ell\} \\ & & \Delta_{k,\delta}^{bot} &:= \Delta_{k,\delta} \cap \{(t, h) : h \leq 2k\varepsilon\ell\} \end{aligned}$$

as illustrated in Figure 8. For any $(t, h) \in \Delta_k$, let

$$p_{t,h} := \mathbb{P}(\pi(t+1) = h+1 | \pi(t) = h)$$

be the conditional probability, w.r.t. $\mathbb{P}$, the uniform measure on D_{2k} , that the path goes up at time t and height h . In [12] (Eq. (4)) it is shown that

$$p_{t,h} = \frac{1}{2} \frac{h+2}{h+1} \frac{2k-t-h}{2k-t}. \quad (3.1)$$

The Markov property allows us to consider parts of the path (of length $2k\varepsilon$, say) separately. We will encode these subpaths by their increments. To this end introduce $\Omega = \{+1, -1\}^{2k\varepsilon}$ and equip it with the natural σ -algebra $\mathcal{F} = 2^\Omega$.

Fix $(t, h) \in \Delta_k$ such that $t + 2k\varepsilon \leq 2k$, and an $i \in [0, 2k\varepsilon]$ integer. Define for $\omega = (\omega(i))_{i=1}^{2k\varepsilon}$ the *absolute height at relative time i* as

$$h_i^{abs} = h_i^{abs}(\omega) := h + \sum_{j=1}^i \omega(j),$$

and the *absolute time* $t_i^{abs} := t + i$. Furthermore, for $i \in [0, 2k\varepsilon]$ let

$$\begin{aligned} p_{i+1} = p_{i+1}(\omega) &:= \frac{1}{2} \frac{h_i^{abs} + 2}{h_i^{abs} + 1} \frac{2k - t_i^{abs} - h_i^{abs}}{2k - t_i^{abs}}, & p_{i+1}^{top} = p_{i+1}^{top}(\omega) &:= \frac{1}{2} \frac{2k - t_i^{abs} - h_i^{abs}}{2k - t_i^{abs}}, \\ p_{i+1}^{bot} = p_{i+1}^{bot}(\omega) &:= \frac{1}{2} \frac{h_i^{abs} + 2}{h_i^{abs} + 1}, \end{aligned}$$

as well as

$$\begin{aligned} \Pi_{t,h} &:= \Pi_{t,h}^{top} \cap \Pi_{t,h}^{bot}, & \Pi_{t,h}^{top} &:= \{\omega \in \Omega : h_i^{abs}(\omega) \leq 2k - t_i^{abs}, \forall i \in [0, 2k\varepsilon]\}, \\ & & \Pi_{t,h}^{bot} &:= \{\omega \in \Omega : h_i^{abs}(\omega) \geq 0, \forall i \in [0, 2k\varepsilon]\}. \end{aligned}$$

The set $\Pi_{t,h}$ encodes the paths of length $2k\varepsilon$ that are legitimate continuations as a Dyck path starting from $(t, h) \in \Delta_k$. On $\Pi_{t,h} \subseteq \Omega$ we define the probability measure $\mathbb{P}_{(t,h)}$ defined by

$$\mathbb{P}_{(t,h)}(\omega) := \prod_{i=1}^{2k\varepsilon} (p_i(\omega))^{\mathbf{1}(\omega(i)=1)} (1 - p_i(\omega))^{\mathbf{1}(\omega(i)=-1)}. \quad (3.2)$$

Similarly, on $\Pi_{t,h}^{top}$ and $\Pi_{t,h}^{bot}$ we define the probability measures $\mathbb{P}_{(t,h)}^{top}$ and $\mathbb{P}_{(t,h)}^{bot}$ with p_i replaced by p_i^{top} and p_i^{bot} in (3.2), respectively. We naturally extend the probability measures $\mathbb{P}_{(t,h)}, \mathbb{P}_{(t,h)}^{top}, \mathbb{P}_{(t,h)}^{bot}$ to the entire measure space $(\Omega, \mathcal{F})$ by setting them zero for ω not in $\Pi_{t,h}, \Pi_{t,h}^{top}$, and $\Pi_{t,h}^{bot}$, respectively.

The result (3.1) shows that $p_i(\omega)$ is the transition probability (under the uniform distribution on D_{2k}) that the path starting at (t, h) goes up in the i -th step after t , conditioned that it passed through at (t, h) and its evolution between times t and $t + i - 1$ was given by $(\omega_1, \omega_2, \dots, \omega_{i-1})$.

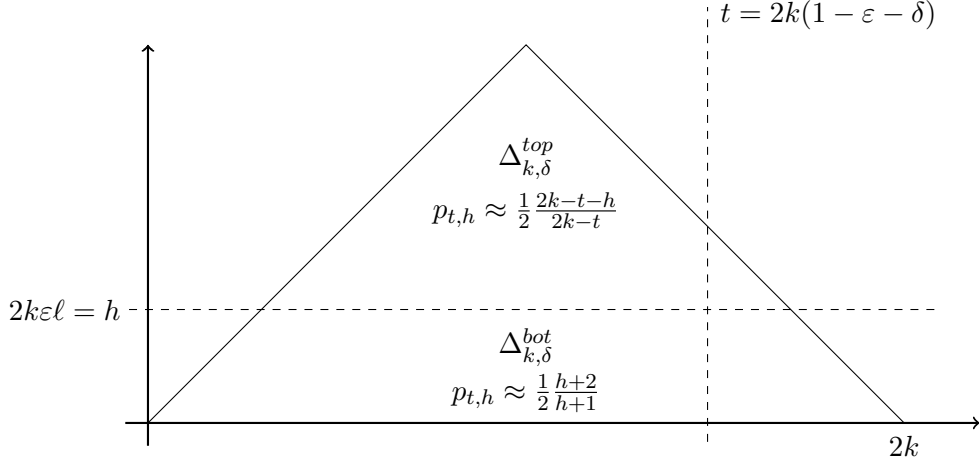


FIGURE 8. Landscape of the transition probability.

Define the random variable $\pi^h : \omega \mapsto (h_i^{abs}(\omega))_{i=0}^{2k\varepsilon}$.^{*} Then clearly for every fixed $\omega \in \Omega$ we have that $\mathbb{P}_{(t,h)}(\omega)$ is the probability of $\{\tilde{\pi}_{[t,t+2k\varepsilon]} = \pi^h(\omega)\}$ for $\tilde{\pi}$ being sampled from the uniform distribution on D_{2k} , conditioned on $\{\tilde{\pi}(t) = h\}$, i.e.

$$\mathbb{P}_{(t,h)}(\omega) = \mathbb{P}\left(\tilde{\pi}_{[t,t+2k\varepsilon]} = \pi^h(\omega) \mid \tilde{\pi}(t) = h\right) \quad \forall \omega \in \Omega \text{ fixed.}$$

Writing $\mathbb{E}_{(t,h)}$ for the expectation under $\mathbb{P}_{(t,h)}$, we make use of the Markov property to write the last term on the r.h.s. of (2.28) as

$$\mathbb{E} z^{T_{\ell,\varepsilon}(\pi)} = \mathbb{E} \prod_{j=0}^{1/\varepsilon-1} z^{T_{\ell,\varepsilon}(\pi_j)} = \mathbb{E}_{(0,0)} \left[z^{T_{\ell,\varepsilon}(\pi_0)} \mathbb{E}_{(t_1,\pi_0(t_1))} \left[z^{T_{\ell,\varepsilon}(\pi_1)} \mathbb{E}_{(t_2,\pi_1(t_1))} [\dots] \right] \right], \quad (3.3)$$

where $\ell = M + 2m$. We also recall the convention that $\pi_j(i) = \pi(t_j + i)$ for all integers $i \in [0, 2k\varepsilon]$, i.e. we start to count the time variable from zero for every resampled path. In the following we will find an upper bound on $\mathbb{E}_{(t_j,h)} z^{T_{\ell,\varepsilon}(\pi_j)}$ independent of j or h . In fact, we bound $\mathbb{E}_{(t,h)} z^{T_{\ell,\varepsilon}(\pi)}$ for any path $\pi = \pi^h$ starting from h . Recall that $T_{\ell,\varepsilon}(\pi) = U(\pi)$ or $T_{\ell,\varepsilon}(\pi) = D(\pi)$, counting up-runs or down-runs, depending on whether $\pi^h(0) = h \leq 2k\varepsilon\ell$ or $h > 2k\varepsilon\ell$, respectively.

3.1. Bound by Simple Random Walk. Let $\mathbb{E}_\mu$ be the expectation under a time-homogeneous random walk starting from 0 and with probability μ of going up and $1 - \mu$ of going down. In particular, $\mathbb{E}_{\frac{1}{2}}$ refers to the expectation under the simple random walk.

Lemma 3.1. *Fix some (large) integers $\ell \geq 2$ and k , and (small) $\varepsilon, \delta > 0$ such that $\varepsilon/\delta < (\ell + 2)^{-1}$. Then for all $(t, h) \in \Delta_{k,\delta}$ we have*

$$\mathbb{E}_{(t,h)} z^{T_{\ell,\varepsilon}(\pi^h)} \leq (1 + \eta(k, \varepsilon, \delta))^{2k\varepsilon} \mathbb{E}_{\frac{1}{2}} z^{U(\pi^h)}$$

^{*}Since π^h is a bijection for every fixed h we will sometimes abuse notation to use ω and paths of length $2k\varepsilon$ (most prominently the subpaths $\pi_j^{(\text{def})} = \pi_{[t_j, t_{j+1}]}$) interchangeably.

with some error term $\eta(k, \varepsilon, \delta) = \eta_{t,h,\ell}(k, \varepsilon, \delta)$ that satisfies

$$|\eta(k, \varepsilon, \delta)| \leq \left((\ell + 1) \frac{\varepsilon}{\delta} + \frac{1}{k\varepsilon} \right). \quad (3.4)$$

In particular, for any fixed ℓ , we have the following limit uniformly in t and h :

$$\lim_{\substack{\varepsilon, \delta \rightarrow 0: \\ \varepsilon/\delta \rightarrow 0}} \limsup_{k \rightarrow \infty} \eta(k, \varepsilon, \delta) = 0.$$

The intuition behind this estimate is that, for h below some threshold $2k\varepsilon\ell$ and t away from the endpoint $2k$ the probability measure $\mathbb{P}_{(t,h)}(\cdot)$ favours going up and hence we expect longer up-runs than in, say, a simple random walk. This is an effect of the repulsive boundary condition at 0 that forces $\pi(i) \geq 0$ for all i . Longer up-runs clearly give us smaller $z^{U(\pi^h)}$ and in that regime $T_{\ell,\varepsilon}$ counts the up-runs. Similarly, for the region where $h > 2k\varepsilon\ell$, the measure $\mathbb{P}_{(t,h)}(\cdot)$ favours going down due to the constraint $\pi(2k) = 0$ at the endpoint and in this regime $T_{\ell,\varepsilon}$ counts the down-runs. Note that the distribution of $U(\pi^h)$ and $D(\pi^h)$ are the same for the simple random walk, hence $\mathbb{E}_{\frac{1}{2}} z^{U(\pi^h)} = \mathbb{E}_{\frac{1}{2}} z^{D(\pi^h)}$.

To formalise this intuition we recall Holley's inequality from [13] (Theorems 2.1 and 2.6, as well as the remark after the statement of Theorem 2.1):

Lemma 3.2 (Holley's inequality). *Let $\Omega := \{-1, 1\}^E$ for some finite E and $\mathcal{F} = 2^\Omega$ be the discrete σ -algebra. Consider the partial order $\leq$ on Ω , given by $\omega \leq \omega'$ iff $\omega(e) \leq \omega'(e)$ for all $e \in E$.*

Let X be an increasing random variable from the measure space $(\Omega, \mathcal{F})$ to $\mathbb{R}$, i.e. $X(\omega) \leq X(\omega')$ for any $\omega \leq \omega' \in \Omega$. Let μ_1, μ_2 be probability measures on $(\Omega, \mathcal{F})$ satisfying

- (1) $\mu_1(\omega^e) \mu_2(\omega_e) \leq \mu_1(\omega_e) \mu_2(\omega^e)$, and
- (2) $\mu(\omega_{ef}^e) \mu(\omega_e^f) \geq \mu(\omega_f^e) \mu(\omega_e^f)$ for $\mu = \mu_1$ or $\mu = \mu_2$,

where ω^e and ω_e are defined by $\omega^e(i) = \omega_e(i) := \omega(i)$ if $i \neq e$ and $\omega^e(i) := 1, \omega_e(i) := -1$ if $i = e$. Furthermore we set $\omega_f^e := (\omega_e)^f, \omega_{ef} := (\omega_e)_f$, and $\omega^{ef} := (\omega^e)^f$. Then we have*

$$\mu_1(X) \leq \mu_2(X). \quad \square$$

To apply Holley's inequality in the proof of Lemma 3.1 we would like to approximate the measure $\mathbb{P}_{(t,h)}$ by a simpler Markov chain. We define

$$\widehat{\mathbb{P}}_{(t,h)}(\omega) := \begin{cases} \mathbb{P}_{(t,h)}^{top}(\omega), & \text{if } h > 2k\varepsilon\ell \\ \mathbb{P}_{(t,h)}^{bot}(\omega), & \text{if } h \leq 2k\varepsilon\ell \end{cases} \quad (3.5)$$

and let $\widehat{\mathbb{E}}_{(t,h)}$ denote the expectation w.r.t. $\widehat{\mathbb{P}}_{(t,h)}$.

Lemma 3.3. *Fix some (large) integers $\ell \geq 2$ and k , some (small) $\varepsilon, \delta > 0$ such that $\varepsilon/\delta < (\ell + 2)^{-1}$, and $(t, h) \in \Delta_{k,\delta}$. Then we have*

$$\mathbb{E}_{(t,h)} z^{T_{\ell,\varepsilon}} \leq (1 + \eta(k, \varepsilon, \delta))^{2k\varepsilon} \widehat{\mathbb{E}}_{(t,h)} z^{T_{\ell,\varepsilon}}, \quad (3.6)$$

*The order indicated by the bracket breaks the symmetry; one may have defined $\omega_f^e := (\omega^e)_f$. However, this notation only occurs in the combination $\mu(\omega_f^e) \mu(\omega_e^f)$ which is independent of this choice.

where the error term $\eta(k, \varepsilon, \delta) = \eta_{t,h,\ell}(k, \varepsilon, \delta)$ is chosen to be the same as in Lemma 3.1; in particular the bound (3.4) holds.

The bound (3.6) controls the measure $\mathbb{P}_{(t,h)}$ by a simpler measure $\widehat{\mathbb{P}}_{(t,h)}$. We could have made an approximation with a Markov chain with constant transition rates (on scale of $2k\varepsilon$) in the regime that is far away from the boundary of Δ_k . This possibility is indicated in Figure 8, but we will not need it in our proof.

Proof of Lemma 3.3. We show that for all $(t, h) \in \Delta_{k,\delta}$ and all $\omega \in \Omega$, we have

$$\mathbb{P}_{(t,h)}(\omega) \leq (1 + \eta(k, \varepsilon, \delta))^{2k\varepsilon} \widehat{\mathbb{P}}_{(t,h)}(\omega), \quad (3.7)$$

then (3.6) will follow. To see (3.7), it clearly suffices to show that both

$$\max_{1 \leq i \leq 2k\varepsilon} \left| 1 - \frac{p_i(\omega)}{p_i^{top}(\omega)} \right| \quad \text{and} \quad \max_{1 \leq i \leq 2k\varepsilon} \left| 1 - \frac{p_i(\omega)}{p_i^{bot}(\omega)} \right|$$

satisfy the same bound (3.4) as $\eta(k, \varepsilon, \delta)$ does, uniformly for all (t, h) in $\Delta_{k,\delta}^{top}$ and $\Delta_{k,\delta}^{bot}$, respectively, as well as uniformly for all $\omega \in \Omega$.

Elementary calculations using the assumptions on (t, h) , in particular $t \leq 2k(1 - \delta - \varepsilon)$, and the fact that $|h - h_i^{abs}(\omega)| \leq 2k\varepsilon$ as well as $0 \leq t_i^{abs} - t \leq 2k\varepsilon$ give

$$0 \leq \frac{p_i(\omega)}{p_i^{top}(\omega)} - 1 = \frac{1}{h_i^{abs} + 1} \leq \frac{1}{2k\varepsilon(\ell - 1) + 1} \leq \frac{1}{k\varepsilon}, \quad \text{for } \ell \geq 2$$

as well as

$$0 \leq 1 - \frac{p_i(\omega)}{p_i^{bot}(\omega)} = \frac{h_i^{abs}}{2k - t_i^{abs}} \leq (\ell + 1)\varepsilon \frac{1}{1 - \varepsilon - \frac{t}{2k}} \leq (\ell + 1)\frac{\varepsilon}{\delta}.$$

Noting that these bounds do not depend on t, h, i or ω we have proven the claim. $\square$

Before we prove Lemma 3.1 we also show that the time-homogeneous approximation $\mathbb{P}_{(t,h)}^{bot}(\omega)$ can be calculated explicitly in terms of the “relative height difference” of the path induced by ω .

Lemma 3.4. Fix $(t, h) \in \Delta_k$ and $\omega \in \Omega = \{+1, -1\}^{2k\varepsilon}$. Let $\Delta\omega := \sum_{i=1}^{2k\varepsilon} \omega(i)$ be the relative height difference of the path $\pi^h(\omega)$. Then we have

$$\mathbb{P}_{(t,h)}^{bot}(\omega) \stackrel{(def)}{=} \prod_{i=1}^{2k\varepsilon} \left(p_i^{bot}(\omega) \right)^{\mathbf{1}(\omega(i)=1)} \left(1 - p_i^{bot}(\omega) \right)^{\mathbf{1}(\omega(i)=-1)} = \left(\frac{1}{2} \right)^{2k\varepsilon} \frac{h + 1 + \Delta\omega}{h + 1}.$$

Proof. Recall that by definition we have $h_0^{abs} = h, h_1^{abs} = h_0^{abs} + \omega(1), \dots$. Hence

$$\begin{aligned} 2^{2k\varepsilon} \mathbb{P}_{(t,h)}^{bot} &\stackrel{(def)}{=} \prod_{i=0}^{2k\varepsilon-1} \left(\frac{h_i^{abs} + 2}{h_i^{abs} + 1} \right)^{\mathbf{1}(\omega(i+1)=1)} \left(\frac{h_i^{abs}}{h_i^{abs} + 1} \right)^{\mathbf{1}(\omega(i+1)=-1)} \\ &= \prod_{i=0}^{2k\varepsilon-1} \frac{h_i^{abs} + \omega(i+1) + 1}{h_i^{abs} + 1} \\ &= \frac{h_{2k\varepsilon} + 1}{h + 1}, \end{aligned}$$

where the last step followed by noting that $h_{i+1}^{abs} = h_i^{abs} + \omega(i+1)$ and a telescoping product argument. Since $h_{2k\varepsilon} = h + \Delta\omega$ this proves the claim. $\square$

Proof of Lemma 3.1. To use Holley's inequality we consider $\Omega = \{+1, -1\}^{2k\varepsilon} = \{(\cdot)\}^{2k\varepsilon}$, where every $\omega \in \Omega$ is naturally identified with a part of a Dyck path π^h via its bracket notation. This induces a partial ordering on the set of subpaths of length $2k\varepsilon$ and allows for ω^e or ω_e to be interpreted as the increments of the path where the e -th step is replaced with an up or down, respectively. Then $\omega \leq \omega'$ if and only if $\omega_i \leq \omega'_i$ at every position i , i.e. if and only if $\pi(\omega)$ goes down every time $\pi(\omega')$ goes down. Furthermore let $\mathbb{P}_{\frac{1}{2}}$ be the probability measure corresponding to $\mathbb{E}_{\frac{1}{2}}$, i.e.

$$\mathbb{P}_{\frac{1}{2}}(\omega) \stackrel{(\text{def})}{=} 2^{-2k\varepsilon}.$$

Now fix $(t, h) \in \Delta_{k, \delta}$. By Lemma 3.3 it suffices to show that

$$\widehat{\mathbb{E}}_{(t, h)} z^{T_{\ell, \varepsilon}(\pi)} \leq \mathbb{E}_{\frac{1}{2}} z^{U(\pi)}. \quad (3.8)$$

By definition of $T_{\ell, \varepsilon}$ and $\widehat{\mathbb{P}}_{(t, h)}$ we deal with two different cases depending on h .

Case: $h \leq 2k\varepsilon\ell$. In this regime we have, by definition of $T_{\ell, \varepsilon}$, that $z^{T_{\ell, \varepsilon}} = z^U$ and $\widehat{\mathbb{P}}_{(t, h)}(\omega) \stackrel{(\text{def})}{=} \mathbb{P}_{(t, h)}^{bot}(\omega)$. Now set*

$$X(\omega) := -z^{U(\pi(\omega))},$$

and we claim that X is increasing. Indeed, this easily follows by induction (for every pair $\omega \leq \omega'$ introduce a sequence $\omega = \omega^{(0)} \leq \omega^{(1)} \leq \omega^{(2)} \leq \dots \leq \omega^{(I)} = \omega'$ such that $\omega^{(i)} = \omega_{e_i}^{(i+1)}$ for some e_i), using the fact that $z^{U(\pi(\omega^e))} \leq z^{U(\pi(\omega_e))}$. This inequality is a consequence of the submultiplicativity of the norm and the definition of z_j in (1.5).

To get (3.8) we will apply Holley's inequality to X and $\mu_1 = \mathbb{P}_{\frac{1}{2}}, \mu_2 = \mathbb{P}_{(t, h)}^{bot}$. Hence it suffices to check that

- (1) $\mathbb{P}_{\frac{1}{2}}(\omega^e) \mathbb{P}_{(t, h)}^{bot}(\omega_e) \leq \mathbb{P}_{\frac{1}{2}}(\omega_e) \mathbb{P}_{(t, h)}^{bot}(\omega^e)$, and
- (2) $\mu(\omega_{ef}) \mu(\omega_e^f) \geq \mu(\omega_f^e) \mu(\omega_e^f)$ for $\mu = \mathbb{P}_{\frac{1}{2}}$ or $\mu = \mathbb{P}_{(t, h)}^{bot}$.

To check condition (1), notice that $\mathbb{P}_{\frac{1}{2}}(\omega_e) = \mathbb{P}_{\frac{1}{2}}(\omega^e)$, so it suffices to prove

$$\frac{\mathbb{P}_{(t, h)}^{bot}(\omega_e)}{\mathbb{P}_{(t, h)}^{bot}(\omega^e)} \leq 1. \quad (3.9)$$

Without loss of generality we may assume $e = 1$ since changing the e -th entry in ω does not change the contribution from the first $e - 1$ terms in the product in

$$\mathbb{P}_{(t, h)}^{bot}(\omega) \stackrel{(\text{def})}{=} \prod_{i=1}^{2k\varepsilon} \left(p_i^{bot}(\omega) \right)^{\mathbf{1}(\omega(i)=1)} \left(1 - p_i^{bot}(\omega) \right)^{\mathbf{1}(\omega(i)=-1)}.$$

Now pick any $\omega = \omega^e$, notice that $\Delta\omega^e = \Delta\omega_e + 2$, and apply Lemma 3.4 to see that (3.9) holds and hence condition (1) in Holley's inequality is satisfied.

*Note that $U(\pi)$ is independent of the initial height $h = \pi^h(0) = \pi(0)$, hence we will suppress h in the notation.

Condition (2) is trivially fulfilled by choosing $\mu = \mathbb{P}_{\frac{1}{2}}$ for which we have equality. Hence Holley's inequality in this setup gives (3.8) for $h \leq 2k\varepsilon\ell$.

Case: $h > 2k\varepsilon\ell$. Similarly to the previous case we now apply Holley's inequality to $X := z^{D(\pi)}$, $\mu_1 = \widehat{\mathbb{P}}_{(t,h)}(\omega) \stackrel{(\text{def})}{=} \mathbb{P}_{(t,h)}^{\text{top}}$, and $\mu_2 = \mathbb{P}_{\frac{1}{2}}$. As before, X is increasing and condition (2) of Holley's inequality is trivially fulfilled by $\mu_2 = \mathbb{P}_{\frac{1}{2}}$. To show condition (1), i.e.

$$\mathbb{P}_{(t,h)}^{\text{top}}(\omega^e) \mathbb{P}_{\frac{1}{2}}(\omega_e) \leq \mathbb{P}_{(t,h)}^{\text{top}}(\omega_e) \mathbb{P}_{\frac{1}{2}}(\omega^e) \quad \Leftrightarrow \quad \mathbb{P}_{(t,h)}^{\text{top}}(\omega^e) \leq \mathbb{P}_{(t,h)}^{\text{top}}(\omega_e),$$

we fix any $e \in [1, 2k\varepsilon)$ and consider for some $\omega = \omega^e$ the ratio

$$\frac{\mathbb{P}_{(t,h)}^{\text{top}}(\omega^e)}{\mathbb{P}_{(t,h)}^{\text{top}}(\omega_e)} = \underbrace{\frac{q_{t_{e-1}, h_{e-1}}^{\text{abs}}}{1 - q_{t_{e-1}, h_{e-1}}^{\text{abs}}}}_{\leq 1} \prod_{i=e}^{2k\varepsilon-1} \underbrace{\left(\frac{q_{t_i^{\text{abs}}, h_i^{\text{abs}}}}{q_{t_i^{\text{abs}}, h_i^{\text{abs}}-2}} \right)^{\mathbf{1}(\omega(i)=1)}}_{\leq 1} \underbrace{\left(\frac{1 - q_{t_i^{\text{abs}}, h_i^{\text{abs}}}}{1 - q_{t_i^{\text{abs}}, h_i^{\text{abs}}-2}} \right)^{\mathbf{1}(\omega(i)=-1)}}_{\geq 1}, \quad (3.10)$$

where

$$q_{t,h} := \frac{1}{2} \frac{2k - t - h}{2k - t} \quad 1 - q_{t,h} = \frac{1}{2} \frac{2k - t + h}{2k - t}. \quad (3.11)$$

Note that $h_i^{\text{abs}(\text{def})} = h_i^{\text{abs}}(\omega) = h_i^{\text{abs}}(\omega^e)$ since we assumed $\omega = \omega^e$ and $h_i^{\text{abs}}(\omega^e) = h_i^{\text{abs}}(\omega_e) + 2 \cdot \mathbf{1}(i \geq e)$.

Now we show that (3.10) is less or equal than 1 for all choices of e and ω . Since the first factor in the product in (3.10) is less or equal than 1 we can bound

$$\begin{aligned} \frac{\mathbb{P}_{(t,h)}^{\text{top}}(\omega^e)}{\mathbb{P}_{(t,h)}^{\text{top}}(\omega_e)} &\leq \frac{q_{t_{e-1}, h_{e-1}}^{\text{abs}}}{1 - q_{t_{e-1}, h_{e-1}}^{\text{abs}}} \prod_{i=e}^{2k\varepsilon-1} \left(\frac{1 - q_{t_i^{\text{abs}}, h_i^{\text{abs}}(\omega)}}{1 - q_{t_i^{\text{abs}}, h_i^{\text{abs}}(\omega)-2}} \right)^{\mathbf{1}(\omega(i)=-1)} \\ &\leq \left(\max_{\substack{j \in [0, 2k\varepsilon), \\ \tilde{\omega} \in \Omega}} \frac{q_{t_j^{\text{abs}}, h_j^{\text{abs}}(\tilde{\omega})}}{1 - q_{t_j^{\text{abs}}, h_j^{\text{abs}}(\tilde{\omega})}} \right) \prod_{i=1}^{2k\varepsilon-1} \frac{1 - q_{t_i^{\text{abs}}, h_i^{\text{abs}}(\omega')}}{1 - q_{t_i^{\text{abs}}, h_i^{\text{abs}}(\omega')-2}}. \end{aligned} \quad (3.12)$$

where $\omega' := (1, -1, -1, \dots, -1)$. The second inequality holds because setting $e = 1$ in the product gives more factors that are greater or equal to one and because all factors for $i \geq e$ can be directly compared (for different ω) using

$$1 \leq \frac{1 - q_{t, h+c}}{1 - q_{t, h-2+c}} \leq \frac{1 - q_{t, h}}{1 - q_{t, h-2}}, \quad \text{for all } c \geq 0, (t, h) \in \Delta_k.$$

Since $q \mapsto \frac{q}{1-q}$ is monotonically increasing for $0 \leq q < 1$, and $(t, h) \mapsto q_{t,h}$ is monotonically decreasing in both variables, we have

$$\max_{\substack{j \in [0, 2k\varepsilon), \\ \tilde{\omega} \in \Omega}} \frac{q_{t_j^{\text{abs}}, h_j^{\text{abs}}(\tilde{\omega})}}{1 - q_{t_j^{\text{abs}}, h_j^{\text{abs}}(\tilde{\omega})}} \leq \max_{\substack{\Delta t \in [0, m], \\ \Delta h \in [-m, m]}} \frac{q_{t+\Delta t, h+\Delta h}}{1 - q_{t+\Delta t, h+\Delta h}} = \frac{q_{t, h-m}}{1 - q_{t, h-m}} = \frac{2k - t - h + m}{2k - t + h - m}$$

for $m := 2k\varepsilon - 1$. To show that (3.12) is less or equal than 1, it suffices (using the formulas (3.11) and $t_i^{\text{abs}(\text{def})} = t + i, h_i^{\text{abs}}(\omega') = h - i + 2$) to see that

$$\frac{2k - t - h + m}{2k - t + h - m} \prod_{i=1}^m \frac{2k - t + h + 2 - 2i}{2k - t + h - 2i} \leq 1.$$

This is easy to see by a telescoping product argument and using that $h \geq 2m$ (since $4k\varepsilon \leq 2k\varepsilon\ell < h$ by assumption). Thus $\mathbb{P}_{(t,h)}^{\text{top}}(\omega^e) \leq \mathbb{P}_{(t,h)}^{\text{top}}(\omega_e)$ so that condition (1) of

Holley's inequality is satisfied. Hence we can apply Holley's inequality to get (3.8) for $h > 2k\varepsilon\ell$. This completes the proof of the lemma. $\square$

3.2. Calculating the lim sup. Now we are ready to prove the main result:

Proof of Theorem 1.1. Start with (2.28) and note that the expectation on the r.h.s. can be written as in (3.3). We apply the trivial bound $z^{T_{\ell,\varepsilon}(\pi)} \leq 1$ for the first δ/ε innermost terms in (3.3) and Lemma 3.1 $\frac{1}{\varepsilon}(1 - \delta)$ times to the remaining ones, yielding

$$\begin{aligned} & [\mathbb{E} \text{val}(\Gamma)]^{1/2k} \\ & \leq \frac{L^{1/2k}}{(\min_{j \leq J} z_j^{1/j})^{1/L}} \|S\|^{1/2} \cdot \max_{\substack{(t,h) \in \Delta_{k,\delta} \\ \ell \in [M, M+2L]}} \left[\prod_{j=0}^{\frac{1-\delta}{\varepsilon}-1} \left((1 + \eta_{t,h,\ell}(k, \varepsilon, \delta))^{2k\varepsilon} \mathbb{E}_{\frac{1}{2}} z^{U(\pi_j)} \right) \right]^{1/2k} \\ & = \frac{L^{1/2k}}{(\min_{j \leq J} z_j^{1/j})^{1/L}} \|S\|^{1/2} \cdot \left(1 + (M + 2L + 1) \frac{\varepsilon}{\delta} + \frac{1}{k\varepsilon} \right)^{1-\delta} \left[\mathbb{E}_{\frac{1}{2}} z^{U(\pi_0)} \right]^{\frac{1-\delta}{2k\varepsilon}}, \end{aligned}$$

where π_0 , the random variable over which we are taking expectation, is a path of length $2k\varepsilon$. There we used that since $\ell = M + 2m$ is between M and $M + 2L$, we have $\eta_{t,h,\ell}(k, \varepsilon, \delta) \leq (M + 2L + 1) \frac{\varepsilon}{\delta} + \frac{1}{k\varepsilon}$ from (3.4), uniformly in t, h .

After taking limits in the following order

$$\lim_{L, M \rightarrow \infty} \lim_{\substack{\varepsilon, \delta \rightarrow 0: \\ \varepsilon/\delta \rightarrow 0}} \limsup_{k \rightarrow \infty},$$

using that $\lim_k |\mathcal{T}_k|^{1/2k} = 2$ and the change of variables $n = 2k\varepsilon$ we have for $\pi^{(n)}$ a simple random walk of length n , as in (2.13):

$$\max \supp \rho = \limsup_k [|\mathcal{T}_k| \cdot \mathbb{E} \text{val}(\Gamma)]^{\frac{1}{2k}} \leq 2 \|S\|^{\frac{1}{2}} \limsup_{n \rightarrow \infty} \left(\mathbb{E}_{\frac{1}{2}} z^{U(\pi^{(n)})} \right)^{\frac{1}{n}}. \quad (3.13)$$

To estimate $(\mathbb{E}_{\frac{1}{2}} z^{U(\pi^{(n)})})^{\frac{1}{n}}$ for large n , we introduce a randomised stopping time n^* with geometric distribution $\mathbb{P}(n^* = m) = w^{m-1}(1 - w)$, where w is a new parameter to be optimised later. Denote the expectation over n^* by $\mathbb{E}^*$. We set $U^* := U(\pi^{(n^*-1)})$. Following Theorem 2 in [14] we find that for $J \in \mathbb{N}$ we have

$$\mathbb{E}^* \mathbb{E}_{\frac{1}{2}} z^{U^*} = (1 - w) \frac{1 + \sum_{j=1}^J \left(\frac{w}{2}\right)^j z_j + \sum_{j>J} \left(\frac{w}{2}\right)^j}{1 - \frac{w}{2} \left(1 + \sum_{j=1}^J \left(\frac{w}{2}\right)^j z_j + \sum_{j>J} \left(\frac{w}{2}\right)^j\right)}, \quad (3.14)$$

as well as

$$\mathbb{E}^* \mathbb{E}_{\frac{1}{2}} z^{U^*} = (1 - w) \sum_{n \geq 0} w^n \mathbb{E}_{\frac{1}{2}} z^{U(\pi^{(n)})},$$

as in equation (10) in [14]. Interpreting $\mathbb{E}_{\frac{1}{2}} z^{U(\pi^{(n)})}$ as the coefficients of the power series (in w) of $\mathbb{E}^* \mathbb{E}_{\frac{1}{2}} z^{U^*}$, it suffices to find (the inverse of) its radius of convergence to get the lim sup in (3.13) by Cauchy-Hadamard.

Considering the explicit formula (3.14), note that this radius of convergence is equal to the minimum of 2 and w_c being defined as the smallest (in absolute value) root

of the denominator in (3.14), i.e. the function ϕ_J defined in (1.7). It is easy to see that the smallest (in absolute value) root of ϕ_J is positive and smaller than 2. Thus $\limsup_{n \rightarrow \infty} \left(\mathbb{E}_{\frac{1}{2}} z^{U(\pi(n))} \right)^{\frac{1}{n}} = \frac{1}{w_c}$, proving Theorem 1.1. $\square$

APPENDIX A. NUMERICS

For $N = 500$, $J = 50$, and $S_{ij} := e^{\frac{i+j}{N}}$ the trivial bound $2\|S\|^{1/2} \approx 4.316$ and the empirical average (number of samples = 10) of the largest eigenvalue (in absolute value) is ≈ 3.677 (with empirical standard deviation of ≈ 0.047). Our method improves the trivial bound to ≈ 3.870 , a factor of improvement of $w_c \approx 1.115$.

REFERENCES

- [1] O. Ajanki, L. Erdős, T. Krüger, *Universality for general Wigner-type matrices*. Prob. Theor. Rel. Fields **169** no. 3-4, 667727 (2017)
- [2] Z.D. Bai, Y.Q. Yin, *Necessary and sufficient conditions for almost sure convergence of the largest eigenvalue of a Wigner matrix*. The Annals of Probability, 16(4):1729-1741 (1988)
- [3] O. Ajanki, L. Erdős, T. Krüger, *Singularities of solutions to quadratic vector equations on complex upper half-plane*. Comm. Pure Appl. Math. **70**(9), 16721705 (2017)
- [4] O. Ajanki, L. Erdős, T. Krüger, *Quadratic vector equations on complex upper half-plane*. Preprint [arxiv:1506.05095](#). Accepted to Memoirs of AMS.
- [5] J. Alt, L. Erdős, T. Krüger, Y. Nemish, *Location of the spectrum of Kronecker random matrices*. Preprint. [arxiv:1706.08343](#)
- [6] V.H. Vu, *Spectral norm of random matrices*. In Harold N. Gabow and Ronald Fagin, editors, STOC, pages 423-430. ACM (2005)
- [7] Z.D. Bai, J.W. Silverstein, *Spectral analysis of large dimensional random matrices*. Springer Verlag, 2 edition (2012)
- [8] L. Erdős, A. Knowles, H.-T. Yau, J. Yin, *The local semicircle law for a general class of random matrices*. Electron. J. Probab. 18 no. 59, 1-58. (2013)
- [9] L. Erdős, T. Krüger, D. Schröder, *Random matrices with slow correlation decay*. Preprint. [arxiv:1705.10661](#)
- [10] M. Ottolini, *Spectral norm of random matrices with non-identically distributed entries*. Master Thesis, University Pisa (2017)
- [11] Anderson, G., Guionnet, A., Zeitouni, O., *An Introduction to Random Matrices*. Studies in advanced mathematics, **118**, Cambridge University Press (2009)
- [12] D. B. Arnold, M. R. Sleep, *Uniform Random Generation of Balanced Parenthesis Strings*, ACM Transactions on Programming Languages and Systems (TOPLAS) 2.1 (1980): 122-128.
- [13] G. Grimmet, *The Random-Cluster Model*. Vol. 333. Springer Science & Business Media (2006)
- [14] L. Holst, T. Konstantopoulos, *Runs in coin tossing: a general approach for deriving distributions for functionals*. [arxiv:1407.6831](#)