Connective Constants on Nested Fractal Graphs

HUA QIU AND YIFAN WANG

Abstract

We study self-avoiding walks on the canonical one-sided graphs of Lindstrøm nested fractals. We prove that the connective constant μ exists and identify $\log \mu$ with the critical inverse temperature of a finite-dimensional boundary-state renormalization. If the boundary-state partition vectors are bounded at criticality, then the fixed-length counts c_n satisfy two-sided polynomial bounds around μ^n. We also prove that h-flexibility implies $c_{n+h}/c_n \to \mu^h$. For regular polygonal N-gaskets, we derive exact crossing recursions, determine the smallest flexibility step h, and obtain explicit algebraic connective constants for the 6- and 9-gaskets. The Vicsek graph has no flexibility step, and its successive ratios do not converge.

Contents

1 Introduction **2**

2 Nested fractal graphs **7**

3 Existence of the connective constant **9**
3.1 A convolution inequality 9
3.2 Submultiplicativity and the root limit 12

4 The critical inverse temperature **14**
4.1 Boundary-state recursions 15
4.2 Comparison with rooted walks 17
4.3 Polynomial bounds 19

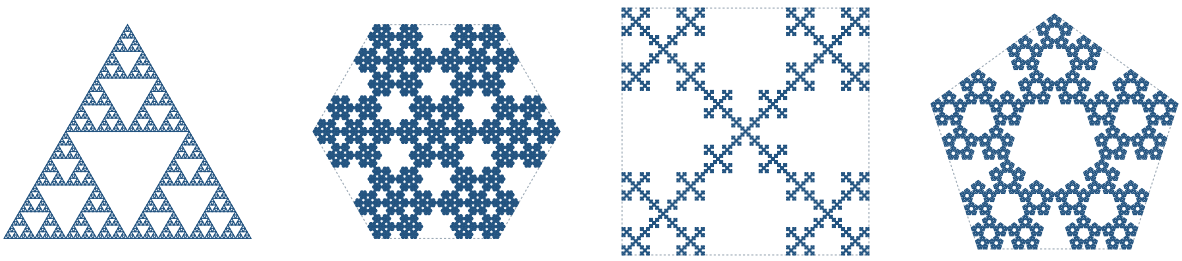
5 Polygonal N-gaskets **20**
5.1 Geometry and crossing recursions 21
5.2 Critical values and connective constants 24

6 Ratio limits **25**
6.1 Local surgery and canonical patterns 26
6.2 Ratio estimates 28
6.3 The Sierpiński gasket 32
6.4 Flexibility of polygonal N-gaskets 33
6.5 The Vicsek counterexample 35

Appendix A. Dynamics of the polygonal crossing recursions **37**

Appendix B. Off-critical rates **47**

2020 Mathematics Subject Classification. Primary 82B41, Secondary 28A80.
Key words and phrases. self-avoiding walk, connective constant, nested fractal graph, ratio limit.
The research of Hua Qiu was supported by the National Natural Science Foundation of China (grants 12471087 and 12531004).



(a) Sierpiński gasket (b) Lindstrøm snowflake (c) Vicsek set (d) Pentagasket

Figure 1: Four standard nested fractals.

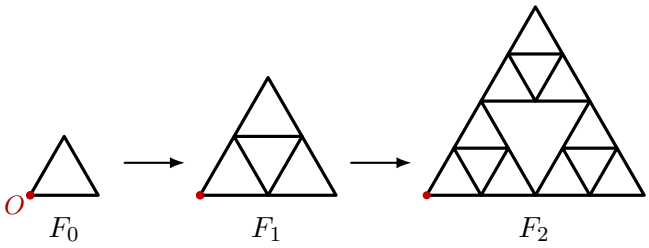


Figure 2: The one-sided blow-up $F_0 \subset F_1 \subset F_2$ for the Sierpiński gasket. The red vertex is the fixed root O.

1 Introduction

Self-avoiding walks arose as lattice models for long-chain polymers. Early systematic mathematical study and numerical estimation began with Hammersley and Morton [6], and Hammersley later isolated their exponential growth rate, now called the *connective constant* [5]. It has since become one of the basic quantities in the subject; see Madras and Slade [19]. Let c_n denote the number of n-step self-avoiding walks from the origin on $\mathbb{Z}^d$, with $d \geq 2$. Submultiplicativity shows that $\mu = \lim_{n\to\infty} c_n^{1/n}$ exists and that $\mu^n \leq c_n$. Hammersley and Welsh proved [7]

$$c_n \leq \mu^n \exp\{O(\sqrt{n})\}$$

and Hutchcroft later replaced $O(\sqrt{n})$ by $o(\sqrt{n})$ [15]. The precise subexponential correction remains unknown in dimensions $2, 3, 4$; for $d \geq 5$, the lace expansion gives $c_n \sim A\mu^n$ [8, 9]. The convergence of fixed-step counting ratios is a separate and more delicate problem. On $\mathbb{Z}^d$, Kesten's two-step ratio limit $c_{n+2}/c_n \to \mu^2$ is closely tied to local pattern replacement [16]. Exact values are exceptionally rare: among the standard two-dimensional lattices, the only rigorously known one is the honeycomb value $\mu = \sqrt{2+\sqrt{2}}$, proved by Duminil-Copin and Smirnov [3].

Nested fractal graphs are self-similar but not transitive. A walk rooted at one vertex therefore cannot usually be split into two walks rooted in the same way. Finite ramification provides a different tool: the connections made by a walk at the boundary of a finite graph cell can be recorded by finitely many states. Ben-Avraham and Havlin used this idea to study self-avoiding walks on finitely ramified fractals [2]. Hattori, Hattori and Kusuoka developed an exact renormalization for self-avoiding paths on the pre-Sierpiński gasket [11]. Building on this special two-variable crossing recursion, Hattori and Kusuoka established the existence of the connective constant for the Sierpiński gasket [13]. See also the construction of a self-avoiding process as a renormalization limit on the Sierpiński gasket [10], and its higher-dimensional extensions [12, 14].

On a general nested fractal, there may be more boundary vertices, and the part of a walk inside one graph cell may have several path components. The Sierpiński gasket recursion therefore does not extend directly. In this paper, our first result is a geometric proof that the connective constant exists on nested fractal graphs. The argument combines graph cells and the Lindstrøm

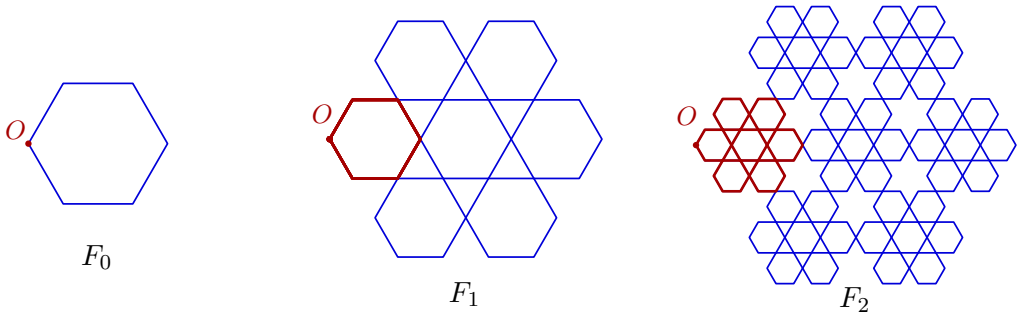


Figure 3: The first three finite graphs of the hexagonal snowflake. The red vertex is the fixed root O, and the red subgraphs in F_1 and F_2 identify the copy of the preceding graph. The one-sided construction therefore expands away from O.

reflection symmetries with a finite convolution inequality, which bounds walks from arbitrary starting vertices by finite products of walk counts from the distinguished root. In particular, the proof requires no dynamical analysis.

For the Sierpiński gasket, Hattori and Kusuoka obtained polynomial upper and lower corrections to the exponential growth of the fixed-length walk counts [13]. Their upper bound localizes a walk by the smallest graph cell containing it and estimates the corresponding critical partition functions. The rooted counts on the gasket are not submultiplicative, so the direct lower bound available on $\mathbb{Z}^d$ has no immediate analogue. Hattori and Kusuoka instead derived it from a substantially more detailed analysis of the special two-variable renormalization dynamics, including the limiting distribution of the critical path length and the regularity of its density. We retain the scale-localization idea in the upper bound, but derive the lower bound directly from the finite convolution estimate used in our existence proof, without a separate analysis of the critical dynamics. This is our second result: the boundary-state renormalization identifies its critical inverse temperature with the logarithm of the connective constant and, whenever the partition vectors remain bounded at criticality, gives corresponding two-sided polynomial bounds for the fixed-length walk counts.

Our third result is a fixed-step ratio theorem. We show that h-flexibility forces the ratio of the walk counts at lengths $n+h$ and n to converge to the h-th power of the connective constant.

For regular polygonal N-gaskets, which include the Sierpiński gasket as the case $N = 3$, we can say considerably more. We derive exact crossing recursions, identify the crossing critical inverse temperature, and determine the smallest flexibility step h. As on the honeycomb lattice [3], these recursions yield explicit algebraic connective constants for $N = 6, 9$.

We begin with the homogeneous iterated function system used throughout. Fix $d \geq 1$, let $\mathcal{I} = \{0, \dots, M-1\}$, where $M \geq 2$, and consider contractive similitudes

$$\Psi_i(x) = \lambda^{-1}\mathsf{U}_i x + \nu_i, \qquad i \in \mathcal{I}, \qquad \lambda > 1, \tag{1}$$

where each U_i is orthogonal. Their attractor is the unique nonempty compact set K satisfying

$$K = \bigcup_{i=0}^{M-1} \Psi_i(K).$$

The compact set K is the self-similar fractal generated by the iterated function system. Let p_i be the fixed point of Ψ_i. A fixed point p_i is *essential* if there exist a fixed point p_j and distinct indices $k, \ell \in \mathcal{I}$ such that $\Psi_k(p_i) = \Psi_\ell(p_j)$. Write

$$V_0 = \{q_0, \dots, q_{b-1}\}$$

for the set of essential fixed points, where $b = \#V_0 \geq 2$.

For a word $w = w_1 \cdots w_r \in \mathcal{I}^r$ with $r \geq 1$, write

$$\Psi_w = \Psi_{w_1} \circ \cdots \circ \Psi_{w_r}, \qquad K_w = \Psi_w(K), \qquad \partial K_w = \Psi_w(V_0).$$

For $r = 0$, let $\emptyset$ be the empty word, set $\mathcal{I}^0 = \{\emptyset\}$, and put $\Psi_\emptyset = \mathrm{id}$ and $K_\emptyset = K$. The set K_w is a *level-r fractal cell*, and ∂K_w is its *boundary*. The term *fractal cell* will always refer to a subset of K; finite cells of the graph will be called graph cells. For distinct $x, y \in V_0$, let $\mathscr{R}_{x,y}$ be reflection in the affine hyperplane $\{z \in \mathbb{R}^d : |z - x| = |z - y|\}$.

Definition 1.1 (Lindstrøm nested fractal [18])**.** The attractor K of the similitudes in (1) is a *nested fractal* if the following conditions hold.

(NF1) There is a nonempty bounded open set $\mathcal{U}$ such that $\Psi_i(\mathcal{U}) \cap \Psi_j(\mathcal{U}) = \varnothing$ for $i \neq j$ and $\bigcup_i \Psi_i(\mathcal{U}) \subset \mathcal{U}$.

(NF2) The set K is connected.

(NF3) For every distinct $x, y \in V_0$, the reflection $\mathscr{R}_{x,y}$ preserves V_0 and, for each $r \geq 1$, permutes the level-r fractal cells together with their boundaries.

(NF4) For distinct $i, j \in \mathcal{I}$,

$$\Psi_i(K) \cap \Psi_j(K) = \Psi_i(V_0) \cap \Psi_j(V_0).$$

Four standard examples are shown in Figure 1.

Definition 1.1 is the homogeneous form of Lindstrøm's definition. Iterating (NF4) shows that the same boundary-intersection identity holds for any two distinct fractal cells of the same level. Much of the work on nested fractals adopts the *one-point intersection condition* as a standing axiom:

$$\#\big(\Psi_u(V_0) \cap \Psi_v(V_0)\big) \leq 1 \qquad (u \neq v,\ |u| = |v|). \tag{2}$$

This condition appears in [4, Proposition 2.2(4)] and as assumption (A5) in [17]. Together with (NF4), it says that two distinct fractal cells of the same level are either disjoint or meet at one common boundary point. Whether it follows from the Lindstrøm axioms is unknown; see [1, Remark 5.25].

On V_0, join the pairs at the smallest positive Euclidean distance. Thus, with

$$\delta_0 = \min\{|x - y| : x, y \in V_0,\ x \neq y\},$$

let

$$F_0 = (V_0, E_0), \qquad E_0 = \big\{\{x, y\} \subset V_0 : |x - y| = \delta_0\big\}.$$

Choose $O \in V_0$, relabel the maps if necessary, and assume that $\Psi_0(O) = O$. The finite approximating graphs and the canonical one-sided infinite graph are

$$F_m = \bigcup_{w \in \mathcal{I}^m} \big(\Psi_0^{-m} \circ \Psi_w\big)(F_0), \qquad F_\infty = \bigcup_{m \geq 0} F_m. \tag{3}$$

Here powers of Ψ_0 denote its iterates or inverse iterates. Put $V_m = V(F_m)$, $V_\infty = V(F_\infty)$, $v_1 = \#V_1$, and

$$\partial F_m = \Psi_0^{-m}(V_0).$$

This one-sided blow-up is the graph-theoretic analogue of zooming into the fixed point O.

Figures 2 and 3 show how the construction produces, respectively, the familiar one-sided Sierpiński gasket graph and the hexagonal snowflake graph. In both examples the distinguished root remains fixed while a new collection of isometric graph pieces is added at every step.

A walk of length n is a sequence $w = (w(0), \dots, w(n))$ of vertices such that $\{w(j-1), w(j)\} \in E(F_\infty),\ 1 \le j \le n$. We write $L(w) = n$, and call w *self-avoiding* if its vertices are distinct. For $a \in V_\infty$, define

$$\begin{aligned} \mathcal{W}_{a,n} &= \{w : w(0) = a,\ L(w) = n,\ w \text{ is self-avoiding}\}, \\ c_{a,n} &= \#\mathcal{W}_{a,n}, \qquad c_n = c_{O,n}, \qquad u_n = \sup_{a \in V_\infty} c_{a,n}. \end{aligned} \tag{4}$$

We set $c_{a,0} = c_0 = u_0 = 1$, and write

$$D = \sup_{x \in V_\infty} \deg_{F_\infty}(x).$$

The two counting sequences c_n and u_n have the same exponential growth rate, even though F_∞ need not be transitive.

Theorem 1. *The limits of the rooted and uniform self-avoiding walk counts exist and are equal:*

$$\mu := \lim_{n\to\infty} c_n^{1/n} = \lim_{n\to\infty} u_n^{1/n} = \inf_{n \ge 1} u_n^{1/n}.$$

Moreover,

$$1 \le \mu \le D - 1 < \infty.$$

Thus μ is the connective constant of F_∞.

At scale m, finite ramification leaves only finitely many possible connection patterns at ∂F_m. For $\beta \ge 0$, let $\mathbf{Z}_m(\beta)$ be the vector of their partition functions, where a connection using ℓ edges has weight $e^{-\beta\ell}$. Define

$$\mathcal{B} = \left\{ \beta \ge 0 : \sup_{m \ge 0} \|\mathbf{Z}_m(\beta)\|_\infty < \infty \right\}, \qquad \beta_{\mathrm{c}} = \inf \mathcal{B}.$$

The precise boundary states are defined in Section 4. Their partition functions characterize the connective constant and, when they remain bounded at $\beta = \beta_{\mathrm{c}}$, control the remaining subexponential factor in the walk counts.

Theorem 2. (i) *The critical inverse temperature β_{c} is finite, and $\mu = e^{\beta_{\mathrm{c}}}$.*

(ii) *If $\beta_{\mathrm{c}} = \min \mathcal{B}$, then there are constants $A \ge 1$ and $p \ge 0$ such that*

$$A^{-1}\mu^n n^{-p} \le c_n \le A\mu^n n^p, \qquad n \ge 1.$$

We also study regular polygonal N-gaskets. Starting from a regular N-gon, we place a homothetic copy of the polygon at each vertex and choose the common contraction ratio at which neighboring copies first meet. When $N \ge 3$ and $4 \nmid N$, the resulting attractor is a nested fractal. Let $F_\infty^{(N)}$ be its one-sided graph. Let μ_N and $\beta_{\mathrm{c},N}$ be its connective constant and boundary-state critical inverse temperature, respectively. At each scale, the boundary-to-boundary crossings fall into finitely many types according to the boundary vertices they visit. We collect their partition functions in the crossing vector $\mathbf{C}_m^{(N)}(\beta)$ (see the exact definition in Section 5).

These crossing functions identify the crossing critical inverse temperature directly and give explicit connective constants when $N = 6$ or 9.

Theorem 3. *Let $N \ge 3$ and $4 \nmid N$. The crossing vector of the regular polygonal N-gasket satisfies a closed finite-dimensional polynomial recursion. Define its* crossing critical inverse temperature *by*

$$\beta_{\mathrm{c},N}^{\mathrm{cr}} := \inf \left\{ \beta \ge 0 : \sup_{m \ge 0} \|\mathbf{C}_m^{(N)}(\beta)\|_\infty < \infty \right\}.$$

Then $\mu_N = e^{\beta_{c,N}^{cr}}$.

For $N \in \{6, 9\}$, *put* $\tau = (\sqrt{5}-1)/2$ *and* $d_N = N/3$. *Then*

$$\mu_N = \left(\frac{\sqrt{1 + 4\tau^{1/d_N}} - 1}{2} \right)^{-1/d_N}. \tag{5}$$

In particular, μ_6 *and* μ_9 *are algebraic numbers.*

The crossing recursions contain more detailed information than is needed for Theorem 3. Appendix A gives the complete behavior of these N-gasket recursions above, below, and at the critical value, and Appendix B determines their off-critical rates. These results are stated there as a *complete phase diagram theorem* (Theorem 5) and an *off-critical rate theorem* (Theorem 6). For the Sierpiński gasket, this dynamical viewpoint originates in the work of Hattori, Hattori and Kusuoka [11]; among the remaining polygonal recursions, the $N = 5$ case requires the most detailed argument in our treatment, owing in part to its four-variable crossing recursion.

Remark 1.2. For every regular polygonal N-gasket, Theorem 5 and Proposition 5.4 show that the boundary-state partition vectors are bounded at $\beta_{c,N}$. By Definition 4.2, the infimum defining $\beta_{c,N}$ is therefore a minimum; hence Theorem 2(ii) applies to these graphs.

We call the graph *h-flexible* if, at some finite scale, every realizable boundary connection has two internal forest realizations whose numbers of edges differ by h. The formal definition is given in Section 6.

This condition yields a fixed-step ratio limit. For regular polygonal N-gaskets, its smallest possible step can be determined explicitly, while the Vicsek graph shows what can happen without flexibility.

Theorem 4. (i) *If the nested fractal graph is h-flexible, then*

$$\lim_{n\to\infty} \frac{c_{n+h}}{c_n} = \mu^h. \tag{6}$$

In particular, 1*-flexibility implies*

$$\lim_{n\to\infty} \frac{c_{n+1}}{c_n} = \mu.$$

(ii) *For the regular polygonal* N*-gasket with* $N \geq 3$ *and* $4 \nmid N$, *write*

$$N = 4k_N + \ell_N, \qquad \ell_N \in \{1, 2, 3\}, \qquad s_N = k_N + 1,$$

and put

$$h_N = \gcd\big(N - 2s_N,\, 2s_N - (N - 2s_N)\big).$$

Then h_N *is the smallest flexibility step. In fact, the graph is* h_N*-flexible at scale* 2, *and*

$$h_N = \begin{cases} 2, & N \equiv 2 \pmod 4, \\ 3, & N \equiv 9 \pmod{12}, \\ 1, & \textit{otherwise}. \end{cases}$$

Consequently,

$$\lim_{n\to\infty} \frac{c_{n+h_N}}{c_n} = \mu_N^{h_N}.$$

(iii) *The standard Vicsek graph is not h-flexible for any $h \geq 1$. Its connective constant is $\mu = \sqrt{2}$, but, for every $k \geq 2$,*

$$\begin{aligned} c_{3^k-2} &= 2^{(3^k+1)/2}, \\ c_{3^k-1} &= 3 \cdot 2^{(3^k-3)/2}, \\ c_{3^k} &= 3 \cdot 2^{(3^k-1)/2}. \end{aligned} \tag{7}$$

Consequently,

$$\frac{c_{3^k-1}}{c_{3^k-2}} = \frac{3}{4}, \qquad \frac{c_{3^k}}{c_{3^k-1}} = 2,$$

and the successive ratios do not converge.

The rest of the paper is organized as follows. Section 2 develops the geometry of graph cells and establishes the basic properties of F_∞. Section 3 proves the existence of the connective constant stated in Theorem 1. Section 4 introduces the boundary-state recursion and proves both the critical identity and the two-sided fixed-length bounds in Theorem 2. Section 5 studies regular polygonal N-gaskets and proves Theorem 3. Section 6 proves the ratio-limit results, determines the smallest flexibility step of each regular polygonal N-gasket, and gives the Vicsek counterexample in Theorem 4. Appendix A gives the complete phase diagram of the crossing recursions for regular polygonal N-gaskets, and Appendix B gives their off-critical rates.

2 Nested fractal graphs

The reflections in (NF3) preserve Euclidean distance and V_0, and therefore act by automorphisms of the nearest-neighbor graph F_0. In this section, we record the graph-theoretic consequences of the nested fractal axioms.

Proposition 2.1. *The graph F_0 is connected. Moreover, $F_m \subset F_{m+1}$ for every m, and F_∞ is an infinite, connected graph of uniformly bounded degree.*

Proof. Choose $u, v \in V_0$ with $|u - v| = \delta_0$. Every $x \in V_0$ has a neighbor in F_0: this is v when $x = u$, and otherwise it is $\mathscr{R}_{x,u}(v)$, since

$$|\mathscr{R}_{x,u}(v) - x| = |\mathscr{R}_{x,u}(v) - \mathscr{R}_{x,u}(u)| = |v - u| = \delta_0.$$

If F_0 were disconnected, choose x, y in different components with $\delta_* = |x - y|$ minimal, and let z be a neighbor of x. Then $|x - z| = \delta_0 < \delta_*$. Let H be the perpendicular bisector of z and y, and set

$$d_y = \operatorname{dist}(y, H), \qquad d_x = \operatorname{dist}(x, H).$$

Since $|x - z| < |x - y|$, the point x lies on the same side of H as z. Projection onto the line perpendicular to H therefore gives $\delta_* \geq d_y + d_x$. Since z and y are also in different components,

$$2d_y = |z - y| \geq \delta_*.$$

It follows that $2d_x \leq \delta_*$. Equality would force x and z to be the same point on the line perpendicular to H, which is impossible. Thus $2d_x < \delta_*$. Since $\mathscr{R}_{z,y}$ is an automorphism of F_0 and sends z to y, it maps the component containing x onto the component containing y. Therefore $x' = \mathscr{R}_{z,y}(x)$ is in the component of y, whereas

$$|x - x'| = 2d_x < \delta_*.$$

This contradicts the minimal choice of x, y. Hence F_0 is connected.

For $i \in \mathcal{I}$, set

$$S_{m,i} = \Psi_0^{-(m+1)} \Psi_i \Psi_0^m.$$

Then $S_{m,i}$ is an isometry, $S_{m,0}$ is the identity, and (3) gives

$$F_{m+1} = \bigcup_{i=0}^{M-1} S_{m,i}(F_m).$$

Since $S_{m,0}$ is the identity, $F_m \subset F_{m+1}$. Since K is connected by (NF2), the incidence graph of the level-m fractal cells is connected; here its vertices are the cells, and two vertices are adjacent when the corresponding cells intersect. Otherwise, the union of the cells would split into two disjoint nonempty compact sets. By (NF4) and (2), two adjacent cells meet at a common boundary vertex. Since each copy of F_0 is connected, their union F_m is connected. Hence F_∞ is connected. Moreover,

$$\operatorname{diam}(\partial F_m) = \lambda^m \operatorname{diam}(V_0),$$

so F_∞ is infinite.

Finally, we prove that the degrees are uniformly bounded. For $r \geq 0$ and $\alpha \in \mathcal{I}^r$, call the subgraph $\Psi_0^{-r}\Psi_\alpha(F_0)$ an *elementary copy*; it is an isometric copy of F_0 in F_∞. Choose a closed ball $\overline{B}(x_{\mathcal{U}}, \rho) \subset \mathcal{U}$. To each elementary copy $\Psi_0^{-r}\Psi_\alpha(F_0)$, associate the ball $\Psi_0^{-r}\Psi_\alpha(B(x_{\mathcal{U}}, \rho))$. Because the two similarities have reciprocal ratios, all these associated balls have radius ρ.

Fix $x \in V_\infty$, and consider any finite family of distinct elementary copies containing x. If $s \geq r$, then

$$\Psi_0^{-r}\Psi_\alpha(F_0) = \Psi_0^{-s}\Psi_0^{s-r}\Psi_\alpha(F_0).$$

The composition $\Psi_0^{s-r}\Psi_\alpha$ equals Ψ_β for a word $\beta \in \mathcal{I}^s$, so the right-hand side is an elementary copy. Choose s no smaller than the parameters r of all the copies in the family. The above identity then represents every member of the family with the same parameter s. Their associated balls are then pairwise disjoint by (NF1). If

$$R = \max_{y \in K} |x_{\mathcal{U}} - y|,$$

then the center of every associated ball is at distance at most R from x. Thus all these radius-ρ balls lie in $B(x, R + \rho)$. Comparing d-dimensional volumes shows that the family contains at most

$$C = \left(\frac{R + \rho}{\rho}\right)^d < \infty.$$

This bound is independent of x and of the chosen family. Every edge incident with x belongs to an elementary copy, and one copy contributes at most $\max_{y \in V_0} \deg_{F_0}(y)$ incident edges. Hence

$$\deg_{F_\infty}(x) \leq C \max_{y \in V_0} \deg_{F_0}(y),$$

uniformly in x. □

We shall repeatedly use finite copies of F_m inside F_∞. For $m \geq 0$, $r \geq m$, and $\alpha \in \mathcal{I}^{r-m}$, define

$$S_{r,\alpha}^{(m)} = \Psi_0^{-r}\Psi_\alpha\Psi_0^m, \qquad \Delta_{m;r,\alpha} = S_{r,\alpha}^{(m)}(F_m), \qquad \partial\Delta_{m;r,\alpha} = \Psi_0^{-r}\Psi_\alpha(V_0). \tag{8}$$

The map $S_{r,\alpha}^{(m)}$ has similarity ratio one and is an isometry. Let

$$\mathcal{T}_m = \{\Delta_{m;r,\alpha} : r \geq m,\ \alpha \in \mathcal{I}^{r-m}\},$$

where repeated copies of the same subgraph are identified. An element of $\mathcal{T}_m$ is a *graph m-cell*. When no confusion with a fractal cell can arise, we simply call it an *m-cell*. Thus a graph m-cell is a finite subgraph of F_∞ isometric to F_m, including all vertices and edges of that copy, and its boundary consists of the $\#V_0$ vertices specified in (8).

Proposition 2.2. *For every $m \geq 0$, F_m is an m-cell. Two distinct m-cells are either disjoint or meet at one common boundary vertex. Consequently, every edge of F_∞ lies in exactly one m-cell. If $m \geq 1$, every m-cell is the union of M $(m-1)$-cells, and the union of the boundaries of these M subcells contains v_1 vertices.*

Proof. The representation in (8) shows that every m-cell is a copy of F_m. If two distinct m-cells meet, the iterated form of (NF4) says that their intersection is contained in the intersection of their boundary sets. By (2), this intersection has at most one vertex. Thus two distinct m-cells cannot share an edge. Conversely, every edge of F_∞ belongs to an elementary copy $\Psi_0^{-r}\Psi_\alpha(F_0)$ for some $r \geq 0$. If $r \geq m$, write $\alpha = \alpha'\alpha''$, where α' has length $r-m$. Since $\Psi_{\alpha''}(F_0) \subset \Psi_0^m(F_m)$, we have

$$\Psi_0^{-r}\Psi_\alpha(F_0) \subset \Psi_0^{-r}\Psi_{\alpha'}\Psi_0^m(F_m) = \Delta_{m;r,\alpha'}.$$

Thus the edge is contained in the m-cell $\Delta_{m;r,\alpha'}$. If $r < m$, the elementary copy is one of the copies in the union defining F_r, hence is contained in $F_r \subset F_m = \Delta_{m;m,\emptyset}$. Thus the edge again lies in an m-cell. Hence every edge belongs to exactly one m-cell.

If $\Delta = \Delta_{m;r,\alpha}$ and $m \geq 1$, the decomposition in (8) gives

$$\Delta = \bigcup_{i=0}^{M-1} \Delta_{m-1;r,\alpha i}.$$

The graphs on the right are the M $(m-1)$-subcells of Δ. Their boundary union is the image of $\bigcup_i \Psi_i(V_0)$ under the injective similarity $\Psi_0^{-r}\Psi_\alpha$, and hence its cardinality is

$$\#\left(\Psi_0^{-1}\left(\bigcup_i \Psi_i(V_0)\right)\right) = \#V_1 = v_1.$$

□

Proposition 2.1 shows that $2 \leq D < \infty$. Since F_∞ is infinite, connected, and locally finite, König's infinity lemma gives an infinite self-avoiding ray starting at O. Consequently,

$$1 \leq c_n \leq u_n \leq D(D-1)^{n-1}, \qquad n \geq 1. \tag{9}$$

3 Existence of the connective constant

We prove Theorem 1 in two steps. First, the reflection axiom gives a local operation which moves any boundary point of a finite cell to the distinguished root. We then combine the resulting uniform comparison with submultiplicativity.

3.1 A convolution inequality

Lemma 3.1. *For every $r \geq 0$ and every $x \in \partial F_r$, there is a Euclidean isometry $A_{r,x}$ whose restriction to F_r is a graph automorphism and which satisfies $A_{r,x}(x) = O$.*

Proof. Write $x = \Psi_0^{-r}(q)$ with $q \in V_0$. If $q = O$, take the identity. Otherwise define

$$A_{r,x} = \Psi_0^{-r}\mathscr{R}_{q,O}\Psi_0^r. \tag{10}$$

When $r = 0$, the reflection $\mathscr{R}_{q,O}$ preserves V_0 by (NF3) and is therefore an automorphism of the nearest-neighbor graph F_0. When $r \geq 1$, (NF3) in Definition 1.1 says that $\mathscr{R}_{q,O}$ permutes the level-r fractal cells and their boundaries. Since it preserves Euclidean distances, it also transports every copy of the nearest-neighbor edge set E_0 to the corresponding edge set. Hence the map in (10) preserves F_r in either case. Finally,

$$A_{r,x}(x) = \Psi_0^{-r}\mathscr{R}_{q,O}(q) = \Psi_0^{-r}(O) = O,$$

because $\mathscr{R}_{q,O}$ interchanges q and O, and O is fixed by Ψ_0. □

Lemma 3.2. *Let*

$$\eta = (x_0, \dots, x_\ell), \qquad \ell \geq 1,$$

be a finite polygonal path in $\mathbb{R}^d$, with consecutive vertices distinct. For $y \in V_\infty$, let $\mathscr{E}_y(\eta)$ be the set of distinct vertex sequences

$$(T(x_0), \dots, T(x_\ell))$$

such that T is a Euclidean isometry of $\mathbb{R}^d$, $T(x_0) = y$, and

$$\{T(x_{j-1}), T(x_j)\} \in E(F_\infty), \qquad 1 \leq j \leq \ell.$$

Then

$$\#\mathscr{E}_y(\eta) \leq D^d.$$

The same estimate holds if $T(x_\ell) = y$ is prescribed in place of $T(x_0) = y$.

Proof. Set

$$r(\eta) = \dim \mathrm{Aff}(x_0, \dots, x_\ell),$$

where $\mathrm{Aff}(x_0, \dots, x_\ell)$ is the smallest affine subspace of $\mathbb{R}^d$ containing the vertices of η. For $0 \leq j \leq \ell$, put

$$\mathsf{A}_j = \mathrm{Aff}(x_0, \dots, x_j).$$

We construct a possible image sequence from left to right. Suppose that the images

$$T(x_0), \dots, T(x_j)$$

have already been fixed. If $x_{j+1} \in \mathsf{A}_j$, there are coefficients $\lambda_0, \dots, \lambda_j$ such that

$$x_{j+1} = \sum_{i=0}^{j} \lambda_i x_i, \qquad \sum_{i=0}^{j} \lambda_i = 1.$$

Every Euclidean isometry is affine. Hence every isometry compatible with these previously fixed images must satisfy

$$T(x_{j+1}) = \sum_{i=0}^{j} \lambda_i T(x_i).$$

Thus the next image $T(x_{j+1})$ is uniquely determined.

If $x_{j+1} \notin \mathsf{A}_j$, then $T(x_{j+1})$ must be a neighbor of the already fixed vertex $T(x_j)$. There are at most D choices. At exactly such a step,

$$\dim \mathsf{A}_{j+1} = \dim \mathsf{A}_j + 1.$$

Since $\dim \mathsf{A}_0 = 0$ and $\dim \mathsf{A}_\ell = r(\eta)$, the dimension increases exactly $r(\eta)$ times. A factor D occurs only at those times, and this proves the desired estimate. If the image of the terminal vertex is prescribed, apply the result just proved to the reversed path $(x_\ell, \dots, x_0)$. □

Set

$$Q = v_1 + 1, \qquad C = D^d.$$

Both constants depend only on the fixed nested fractal graph. The next proposition is the main combinatorial estimate in the proof of the existence of the connective constant.

Proposition 3.3. *For every $a \in V_\infty$ and $n \geq 1$,*

$$c_{a,n} \leq \sum_{q=1}^{\min\{Q,n\}} C^q \sum_{\substack{\ell_1 + \cdots + \ell_q = n \\ \ell_i \geq 1}} \prod_{i=1}^{q} c_{\ell_i}. \tag{11}$$

The terms in (11) have a direct interpretation. A walk is divided into q pieces of positive lengths $\ell_1, \dots, \ell_q$. After a suitable isometry, each piece becomes a self-avoiding walk from O, which accounts for the product $\prod_i c_{\ell_i}$. The number of pieces is at most Q, and at most C^q original walks can give the same q-tuple of walks from O.

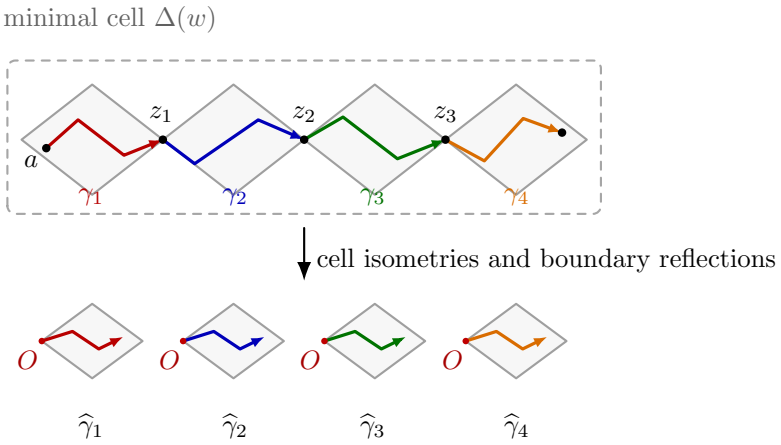


Figure 4: Decomposition inside the smallest cell containing the walk. The upper walk is split whenever two consecutive edges lie in different children. The first piece is reversed, and Lemma 3.1 then moves the initial point of every piece to O.

Figure 4 illustrates the decomposition and the normalization of its pieces.

Proof. Fix $a \in V_\infty$ and $w \in \mathcal{W}_{a,n}$. Since w has only finitely many edges, it is contained in some F_s. By Proposition 2.2, F_s is an s-cell. Hence

$$m(w) = \min\{m \geq 0 : \text{some } \Delta \in \mathcal{T}_m \text{ contains every edge of } w\}$$

is well defined. There is only one $m(w)$-cell containing all the edges of w. Indeed, two such cells would both contain the first edge of w, which is impossible by Proposition 2.2. Denote this unique cell by $\Delta(w)$.

Suppose first that $m(w) = 0$. Every vertex of the 0-cell $\Delta(w)$ belongs to its boundary. Choose a representation of $\Delta(w)$ in (8) and use the inverse of its isometry to move w to F_0. Then apply Lemma 3.1 to move its initial point to O. This produces one walk in $\mathcal{W}_{O,n}$.

Now suppose that $m = m(w) \geq 1$. By Proposition 2.2, the cell $\Delta(w)$ has M children of level $m-1$, and every edge of w lies in exactly one of them. Divide w into maximal consecutive pieces for which all edges lie in the same child. Write

$$w = \gamma_1 \circ \cdots \circ \gamma_q, \qquad \gamma_i \subset \Lambda_i, \qquad \Lambda_i \neq \Lambda_{i+1}, \qquad L(\gamma_i) = \ell_i \geq 1, \qquad \sum_{i=1}^{q} \ell_i = n.$$

Here Λ_i is the child containing the edges of γ_i. The minimality of m implies $q \geq 2$, because otherwise all edges of w would lie in one $(m-1)$-cell.

For $1 \leq i < q$, let z_i be the common endpoint of γ_i and γ_{i+1}. The point z_i lies in both Λ_i and Λ_{i+1}. These children are distinct, so Proposition 2.2 shows that z_i is a boundary vertex of both. Moreover, the vertices $z_1, \ldots, z_{q-1}$ are distinct because w is self-avoiding. The union of the boundaries of all children of $\Delta(w)$ has v_1 vertices, again by Proposition 2.2. It follows that

$$q - 1 \leq v_1, \qquad q \leq Q.$$

This remains true if w leaves a child and later returns to it: each change from one child to another occurs at a different vertex of w.

The first piece need not start at a boundary vertex, but it ends at $z_1 \in \partial\Lambda_1$. We therefore reverse γ_1. For $i \geq 2$, the original orientation is retained, and γ_i starts at $z_{i-1} \in \partial\Lambda_i$. Thus every oriented piece now starts at a boundary vertex of its child. For each child, choose a representation in (8), move the corresponding piece to F_{m-1} by the inverse of its isometry, and then use Lemma 3.1. This gives a walk

$$\widehat{\gamma}_i \in \mathcal{W}_{O,\ell_i}, \qquad 1 \leq i \leq q.$$

Consequently, w determines q, the positive composition $\ell_1 + \cdots + \ell_q = n$, and the tuple $(\widehat{\gamma}_1, \ldots, \widehat{\gamma}_q)$.

It remains to estimate how many original walks can produce a fixed tuple. Fix a, q, the lengths, and the walks $\widehat{\gamma}_1, \ldots, \widehat{\gamma}_q$. Each oriented original piece is the image of the corresponding $\widehat{\gamma}_i$ under a Euclidean isometry: the required isometry is the composition of the inverse boundary automorphism from Lemma 3.1 and the corresponding isometry in (8).

When $m \geq 1$, the first normalized walk represents the reverse of γ_1. Its terminal image must therefore be a. By Lemma 3.2, there are at most $C = D^d$ possible image sequences for this piece. Each choice determines its other endpoint, which is z_1. The image of $\widehat{\gamma}_2$ must start at this prescribed vertex z_1, so the same lemma gives at most C choices for the second piece. Each of those choices determines z_2. Repeating the argument piece by piece gives at most C^q possible sequences of pieces. When $m = 0$, Lemma 3.2 is applied with the initial image fixed at a, and gives the same bound C for the single piece.

Once the images of all pieces are fixed, their concatenation determines the entire walk w, after the first piece is returned to its original orientation. Some of the image sequences counted above may not lie in the required children, and some concatenations may fail to be self-avoiding. Keeping these extra possibilities only enlarges the upper bound. For fixed q and fixed lengths, there are $\prod_{i=1}^{q} c_{\ell_i}$ possible tuples of walks from O. Finally, summing over $1 \leq q \leq \min\{Q, n\}$ and over all positive compositions of n proves (11). □

Remark 3.4. Every product on the right-hand side of (11) contains at most Q rooted factors; hence at least one factor has length at least $\lceil n/Q \rceil$. This is essential for the lower bound in Theorem 2(ii).

3.2 Submultiplicativity and the root limit

The numbers c_n are tied to the distinguished vertex O, and there is no evident submultiplicative inequality for them. The uniform counts u_n, however, do satisfy such an inequality.

Lemma 3.5. *For all $r, s \geq 0$,*

$$u_{r+s} \leq u_r u_s. \tag{12}$$

Consequently, the limit

$$\gamma = \lim_{n\to\infty} \frac{1}{n} \log u_n = \inf_{n\geq 1} \frac{1}{n} \log u_n$$

exists and belongs to $[0, \log(D-1)]$.

Proof. Fix $a \in V_\infty$, and cut a walk in $\mathcal{W}_{a,r+s}$ after its first r edges. There are at most u_r choices for the first part. Once that part is fixed, its endpoint is fixed as well. If we ignore the condition that the remaining s edges must avoid the vertices already visited, there are at most u_s choices for the second part. Thus $c_{a,r+s} \leq u_r u_s$. Taking the supremum over a proves (12).

It follows that $\log u_n$ is subadditive. Fekete's lemma therefore gives the asserted limit and the formula for its infimum. Finally, (4) and the degree estimate in Proposition 2.1 give $1 \leq u_n \leq D(D-1)^{n-1}$, which yields $0 \leq \gamma \leq \log(D-1)$. □

We now return to the walks from O. Taking the supremum over the starting vertex a in Proposition 3.3, specifically in (11), gives

$$u_n \leq \sum_{q=1}^{\min\{Q,n\}} C^q \sum_{\substack{\ell_1+\cdots+\ell_q=n \\ \ell_i\geq 1}} \prod_{i=1}^{q} c_{\ell_i}. \tag{13}$$

The important point in (13) is that the number of factors is bounded by the fixed constant Q, independently of n. We use this to compare the exponential growth of c_n with that of u_n.

Lemma 3.6. *The convolution estimate (13), together with Lemma 3.5, implies*

$$\lim_{n\to\infty} \frac{1}{n} \log c_n = \gamma.$$

Proof. Fix $\varepsilon > 0$. By Lemma 3.5, $j^{-1} \log u_j \to \gamma$. Hence the finitely many small values of j can be absorbed into a constant $A_\varepsilon \geq 1$ so that

$$u_j \leq A_\varepsilon \exp\big((\gamma + \varepsilon)j\big) \qquad j \geq 0.$$

For a fixed q, there are $\binom{n-1}{q-1}$ positive compositions of n into q parts. Consequently, the sum of all coefficients multiplying the products on the right-hand side of (13) is

$$G_n = \sum_{q=1}^{\min\{Q,n\}} C^q \binom{n-1}{q-1} \leq QC^Q n^{Q-1}, \tag{14}$$

where we used $C \geq 1$ and $\binom{n-1}{q-1} \leq n^{Q-1}$. By (13), not every product on its right-hand side can be smaller than u_n/G_n. Therefore there are $q \leq Q$ and positive integers $\ell_1, \ldots, \ell_q$, with $\ell_1 + \cdots + \ell_q = n$, such that

$$\prod_{i=1}^{q} c_{\ell_i} \geq \frac{u_n}{G_n}.$$

Let $k = \max_i \ell_i$, and choose one factor of this length. Since there are at most Q parts,

$$k \geq \left\lceil \frac{n}{q} \right\rceil \geq \left\lceil \frac{n}{Q} \right\rceil.$$

For every other part, use $c_{\ell_i} \leq u_{\ell_i}$ and the preceding upper bound for u_{ℓ_i}. The sum of those other lengths is $n - k$. We obtain

$$c_k \geq \frac{u_n}{G_n A_\varepsilon^{q-1} \exp\big((\gamma + \varepsilon)(n-k)\big)}. \tag{15}$$

We next compare this long rooted walk with a rooted walk of any shorter length. If $0 \leq t \leq k$, cutting a walk in $\mathcal{W}_{O,k}$ after t edges gives

$$c_k \leq c_t u_{k-t}.$$

We now choose $n = Qt$. This choice guarantees that the length k selected above satisfies $k \geq \lceil n/Q \rceil = t$, so the last inequality can be applied. Combining it with (15) and with

$$u_{k-t} \leq A_\varepsilon \exp\big((\gamma + \varepsilon)(k-t)\big)$$

gives

$$c_t \geq \frac{u_{Qt}}{G_{Qt} A_\varepsilon^{Q} \exp\big((\gamma + \varepsilon)(Q-1)t\big)}. \tag{16}$$

Indeed, the two exponential losses combine because

$$(Qt - k) + (k - t) = (Q-1)t$$

and the powers of A_ε are bounded by A_ε^Q, since $q \leq Q$.

Taking logarithms in (16) and dividing by t gives

$$\frac{1}{t} \log c_t \geq \frac{1}{t} \log u_{Qt} - \frac{1}{t} \log G_{Qt} - \frac{Q}{t} \log A_\varepsilon - (Q-1)(\gamma + \varepsilon).$$

By Lemma 3.5,

$$\frac{1}{t}\log u_{Qt} \longrightarrow Q\gamma.$$

Moreover, (14) shows that G_{Qt} grows polynomially in t, and hence

$$\frac{1}{t}\log G_{Qt} \longrightarrow 0, \qquad \frac{Q}{t}\log A_\varepsilon \longrightarrow 0.$$

Taking the lower limit in the preceding inequality therefore yields

$$\liminf_{t\to\infty} \frac{1}{t}\log c_t \geq Q\gamma - (Q-1)(\gamma+\varepsilon) = \gamma - (Q-1)\varepsilon.$$

Letting $\varepsilon \downarrow 0$ gives

$$\liminf_{t\to\infty} \frac{1}{t}\log c_t \geq \gamma.$$

On the other hand, $c_t \leq u_t$ by (4), and Lemma 3.5 therefore gives

$$\limsup_{t\to\infty} \frac{1}{t}\log c_t \leq \gamma.$$

The two inequalities prove the lemma. □

Proof of Theorem 1. By Lemmas 3.5 and 3.6,

$$\lim_{n\to\infty} c_n^{1/n} = \lim_{n\to\infty} u_n^{1/n} = e^\gamma = \inf_{n\geq 1} u_n^{1/n}.$$

Moreover, Proposition 2.1 and König's infinity lemma give an infinite self-avoiding ray from O, so $c_n \geq 1$. The same proposition gives the degree bound $c_n \leq D(D-1)^{n-1}$. Taking n-th roots proves $1 \leq \mu \leq D-1 < \infty$. □

4 The critical inverse temperature

In this section, Subsections 4.1 and 4.2 introduce finite boundary states and compare their partition functions with the generating function for walks from O, thereby proving Theorem 2(i). Subsection 4.3 combines this comparison with the finite convolution bound from Section 3 to prove Theorem 2(ii).

For $m \geq 0$, identify the boundary of F_m with V_0 by

$$\iota_m : V_0 \longrightarrow \partial F_m, \qquad \iota_m(q) = \Psi_0^{-m}(q).$$

Definition 4.1 (Boundary states)**.** A *linear forest* is an acyclic graph whose vertices have degree at most two. A *boundary forest* in F_m is a nonempty edge subgraph $\Gamma \subset F_m$ such that

(i) Γ is a linear forest;

(ii) every vertex of degree one belongs to ∂F_m;

(iii) every connected component contains at least one edge.

Thus every component is a self-avoiding path with two distinct endpoints in ∂F_m.

An *abstract boundary forest* is a nonempty linear forest whose vertex set is contained in V_0 and which has no isolated vertices; equivalently, each connected component is a path containing at least one edge. Let $\mathcal{S}$ be the finite set of all such abstract boundary forests. Thus an element $\sigma \in \mathcal{S}$ may consist of one path or of several vertex-disjoint paths; it records only a possible finite combination of boundary connections, not an actual walk in F_m.

To pass from the actual boundary forest $\Gamma \subset F_m$ to its abstract boundary pattern, consider each component of Γ in F_m. List the boundary vertices in the order in which the path meets

them. Keep only these listed vertices, replace each portion between two consecutive listed vertices by one abstract edge, and relabel the retained vertices by ι_m^{-1}. This compression operation is called taking the *boundary trace*. The resulting nonempty linear forest on V_0, hence an element of $\mathcal{S}$, is the *boundary state* of Γ, denoted by $\mathrm{tr}_m(\Gamma)$. It records which boundary vertices are visited and how consecutive visits are connected along each path; it does not record the internal vertices or the lengths of the individual portions.

For $\sigma \in \mathcal{S}$, put

$$\mathcal{C}_m(\sigma) = \{\Gamma \subset F_m : \Gamma \text{ is a boundary forest and } \mathrm{tr}_m(\Gamma) = \sigma\},$$

and, for $\beta \geq 0$, define

$$Z_m^\sigma(\beta) = \sum_{\Gamma \in \mathcal{C}_m(\sigma)} e^{-\beta \# E(\Gamma)}, \qquad \mathbf{Z}_m(\beta) = \big(Z_m^\sigma(\beta)\big)_{\sigma \in \mathcal{S}}. \tag{17}$$

We call $\mathbf{Z}_m(\beta)$ the *boundary-state partition vector*. We adjoin the *empty state* $\varnothing$ and set $Z_m^\varnothing(\beta) = 1$.

Definition 4.2 (Critical inverse temperature)**.** Let

$$\mathcal{B} = \left\{\beta \geq 0 : \sup_{m \geq 0} \|\mathbf{Z}_m(\beta)\|_\infty < \infty\right\}, \qquad \beta_c = \inf \mathcal{B}, \tag{18}$$

where the infimum is initially understood in the extended half-line $[0, \infty]$. The number β_c is the *critical inverse temperature* of the boundary-state renormalization.

4.1 Boundary-state recursions

The boundary-state construction in Definition 4.1 retains every possible connection pattern at the boundary of a cell. This is necessary when $\#V_0 \geq 4$, because the intersection of one global self-avoiding walk with a cell may have more than one path component.

The order retained in Definition 4.1 matters. For example, a component whose boundary trace is $q_0 - q_1 - q_2$ has degree two at q_1, whereas a component with trace $q_0 - q_2$ does not use q_1.

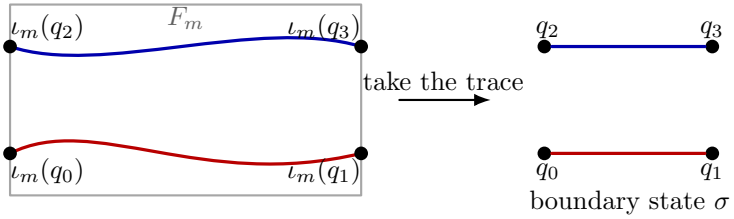


Figure 5: A boundary forest and its boundary state. Internal vertices and the shapes of the two paths are suppressed, while their boundary connections are retained.

Figure 5 records precisely the information kept by a boundary state.

Proposition 4.3. *There is a polynomial map*

$$\Phi : \mathbb{R}_+^{\mathcal{S}} \longrightarrow \mathbb{R}_+^{\mathcal{S}}$$

with nonnegative integer coefficients such that

$$\mathbf{Z}_{m+1}(\beta) = \Phi\big(\mathbf{Z}_m(\beta)\big), \qquad m \geq 0, \quad \beta \geq 0. \tag{19}$$

The map depends only on the incidence pattern of the first-level fractal cells and the identifications of their boundary vertices.

Proof. Let $\mathcal{S}_0 = \mathcal{S} \cup \{\varnothing\}$. For a tuple $\boldsymbol{\sigma} = (\sigma_0, \dots, \sigma_{M-1}) \in \mathcal{S}_0^M$, transport σ_i to the boundary $\Psi_i(V_0)$ of the first-level fractal cell $\Psi_i(K)$, and identify vertices which represent the same point. This produces an abstract graph whose vertex set is contained in

$$\bigcup_{i=0}^{M-1} \Psi_i(V_0).$$

We call the tuple *admissible* if this graph is acyclic, has maximum degree at most two, and every vertex of degree one belongs to V_0, the boundary of the parent cell. Suppressing all vertices outside V_0 then produces a well-defined state in $\mathcal{S}$. Let $\mathfrak{C}_\sigma$ be the finite set of admissible tuples whose resulting state is σ.

We claim that

$$Z_{m+1}^{\sigma}(\beta) = \sum_{\boldsymbol{\sigma} \in \mathfrak{C}_\sigma} \prod_{i=0}^{M-1} Z_m^{\sigma_i}(\beta), \qquad Z_m^{\varnothing}(\beta) = 1. \tag{20}$$

Indeed, Proposition 2.2 assigns every edge of F_{m+1} to a unique m-subcell. Restricting a boundary forest to these subcells therefore gives a unique tuple in $\mathcal{S}_0^M$. Any degree-one vertex created by the restriction lies on the boundary of the corresponding subcell. After the portions of the forest inside each child copy of F_m have been compressed to their child states, the resulting abstract union records the original forest, including the junctions at the junction vertices in the interior of the parent cell that arise from the boundaries of its m-subcells. Suppressing these internal junctions gives exactly the boundary state of the original forest. Hence the tuple is admissible.

Conversely, let $\boldsymbol{\sigma} = (\sigma_0, \dots, \sigma_{M-1}) \in \mathfrak{C}_\sigma$. For each $i \in \{0, \dots, M-1\}$, choose a forest $\Gamma_i \in \mathcal{C}_m(\sigma_i)$ when $\sigma_i \neq \varnothing$, and set $\Gamma_i = \varnothing$ when $\sigma_i = \varnothing$. Distinct child copies of F_m have disjoint edge sets and meet only at boundary vertices by Proposition 2.2. More precisely, admissibility means that $\bigcup_i \Gamma_i$ is acyclic, every vertex of this union has degree at most two, and every degree-one vertex belongs to the boundary V_0 of the parent cell. Therefore $\bigcup_i \Gamma_i$ is a boundary forest with state σ.

These two operations are inverse to each other. Moreover, the edge sets of distinct child copies are disjoint. Hence the number of edges in the union is the sum of the edge counts in the children, while its weight is the product of their weights. Summing over all admissible tuples proves (20).

The first-level incidence pattern is the same at every scale, so the finite sets $\mathfrak{C}_\sigma$ do not depend on m. Collecting (20) over $\sigma \in \mathcal{S}$ gives (19). □

We also require states with one free endpoint. A *marked boundary forest* in F_m is a pair (Γ, ξ), where Γ satisfies conditions (i) and (iii) of Definition 4.1, ξ is a distinguished degree-one vertex, and every degree-one vertex other than ξ belongs to ∂F_m. The point ξ is allowed to lie on ∂F_m. We introduce a new formal symbol $\star$, which is not a vertex of F_m or V_0, to represent a marked endpoint lying in the interior of F_m. The *marked boundary trace* $\mathrm{tr}_m^\bullet(\Gamma, \xi)$ is defined by applying the same compression procedure to (Γ, ξ), with the following modification: if $\xi \in \partial F_m$, the corresponding vertex $\iota_m^{-1}(\xi) \in V_0$ is marked; if $\xi \notin \partial F_m$, the symbol $\star$ is marked instead. In both cases the trace records which component contains the marked point. Let $\mathcal{S}^\bullet$ be the resulting finite set of marked states, and set

$$A_m^{\sigma^\bullet}(\beta) = \sum_{\substack{(\Gamma,\xi) \text{ marked} \\ \mathrm{tr}_m^\bullet(\Gamma,\xi) = \sigma^\bullet}} e^{-\beta \# E(\Gamma)}, \qquad \mathbf{A}_m(\beta) = \big(A_m^{\sigma^\bullet}(\beta)\big)_{\sigma^\bullet \in \mathcal{S}^\bullet}.$$

Proposition 4.4. *There is a matrix $\mathcal{L}(\mathbf{x})$, whose entries are polynomials in $\mathbf{x} = (x_\sigma)_{\sigma \in \mathcal{S}}$ with nonnegative integer coefficients, such that*

$$\mathbf{A}_{m+1}(\beta) = \mathcal{L}\big(\mathbf{Z}_m(\beta)\big)\mathbf{A}_m(\beta). \tag{21}$$

Consequently, if $\sup_m \|\mathbf{Z}_m(\beta)\|_\infty < \infty$*, then there is* $C_\beta < \infty$ *such that*

$$\|\mathbf{A}_m(\beta)\|_\infty \le C_\beta^{m+1}, \qquad m \ge 0. \tag{22}$$

Proof. The marked vertex ξ has degree one in Γ, so exactly one edge of Γ is incident with ξ. By Proposition 2.2, this edge belongs to exactly one child copy of F_m. We put the mark on that child restriction. Every other nonempty child restriction is an unmarked boundary forest, even if its copy shares the boundary vertex ξ.

For a fixed marked state of the parent cell, whether a given collection of child states can be joined to form that parent state is determined solely by the child states, as in the proof of Proposition 4.3. Therefore each parent marked partition function is a finite sum of terms consisting of exactly one marked factor and at most $M-1$ unmarked factors. This gives the exact linear recursion (21).

If $\|\mathbf{Z}_m(\beta)\|_\infty \le K_\beta$ for all m, the finitely many entries of $\mathcal{L}$ are bounded on the cube $[0, K_\beta]^{\mathcal{S}}$. Hence

$$\|\mathbf{A}_{m+1}(\beta)\|_\infty \le C_\beta \|\mathbf{A}_m(\beta)\|_\infty$$

for some $C_\beta < \infty$. Iteration, with one further increase of C_β to include $\|\mathbf{A}_0(\beta)\|_\infty$, proves (22). □

Since all terms in (17) are nonnegative, $\mathbf{Z}_m(\beta') \le \mathbf{Z}_m(\beta)$ coordinatewise whenever $\beta' \ge \beta$. Hence $\beta \in \mathcal{B}$ and $\beta' \ge \beta$ imply $\beta' \in \mathcal{B}$. Proposition 4.3 gives the fixed finite-dimensional polynomial recursion for the boundary-state partition functions, while Proposition 4.4 gives the corresponding linear recursion and growth estimate for marked partition functions. We use these two recursions below to analyze bounded orbits and the critical inverse temperature in Definition 4.2.

4.2 Comparison with rooted walks

Lemma 4.5. *There are constants* $c_0, c_1 < \infty$*, depending only on the graph, such that every self-avoiding walk of length* n *starting at* O *is contained in* $F_{m(n)}$*, where*

$$m(n) \le c_0 + c_1 \log(n+1).$$

Proof. By the definition of δ_0, every edge has length δ_0, and distinct points of V_0 are at least δ_0 apart. The degree of O is finite and $F_\infty = \bigcup_m F_m$, so there is m_0 such that every edge incident with O belongs to F_{m_0}. For $n \ge 1$, choose

$$m(n) = \max\left\{m_0, \lceil \log_\lambda(n+1) \rceil\right\}.$$

Suppose that a self-avoiding walk w from O has an edge outside $F_{m(n)}$, and consider its first such edge. Its initial vertex is a point at which $F_{m(n)}$ meets a different $m(n)$-cell. By Proposition 2.2, this point belongs to $\partial F_{m(n)}$. It cannot be O: an exit through O on the first step is excluded by the choice of m_0, and a later exit through O would revisit the initial vertex. The walk must therefore reach $\iota_{m(n)}(q)$ for some $q \in V_0 \setminus \{O\}$ before it exits. The length of the part already traversed is at least

$$\frac{|\iota_{m(n)}(q) - O|}{\delta_0} = \frac{\lambda^{m(n)}|q - O|}{\delta_0} \ge \lambda^{m(n)} \ge n+1 > n,$$

which is impossible. Thus $w \subset F_{m(n)}$. The asserted logarithmic bound follows immediately from the definition of $m(n)$. □

For $\beta \ge 0$, let

$$\chi(\beta) = \sum_{n=0}^{\infty} c_n e^{-\beta n}, \tag{23}$$

be the rooted self-avoiding walk *susceptibility*, or equivalently its length-generating function.

Proposition 4.6. *For every $\beta \geq 0$, the following implications hold.*

(i) *If $\beta \in \mathcal{B}$, then $\chi(\beta') < \infty$ for every $\beta' > \beta$.*

(ii) *If $\chi(\beta) < \infty$, then $\beta \in \mathcal{B}$.*

Proof. First suppose that $\beta \in \mathcal{B}$. By Proposition 4.4, there is $C_\beta \geq 1$ such that (22) holds. Fix $n \geq 1$, and use Lemma 4.5 to choose $m = m(n)$. Every $w \in \mathcal{W}_{O,n}$, marked at its terminal vertex, is a marked boundary forest in F_m, because its other endpoint is $O \in \partial F_m$. Consequently,

$$c_n e^{-\beta n} \leq \sum_{\sigma^\bullet \in \mathcal{S}^\bullet} A_m^{\sigma^\bullet}(\beta) \leq \#\mathcal{S}^\bullet \, C_\beta^{m+1} \leq C(n+1)^{\kappa_\beta}$$

for constants $C, \kappa_\beta < \infty$ which may depend on β, but not on n. If $\beta' > \beta$, it follows that

$$c_n e^{-\beta' n} \leq C(n+1)^{\kappa_\beta} e^{-(\beta'-\beta)n}.$$

The series on the right is summable, and hence $\chi(\beta') < \infty$. This proves (i).

For (ii), suppose that $\chi(\beta) < \infty$, and fix $\sigma \in \mathcal{S}$. Let k be the number of components of σ. The state σ specifies the two endpoints of each component and determines which boundary vertices belong to the same component. Order the components using the fixed order $q_0, \ldots, q_{b-1}$ of V_0, and orient each one from the endpoint with the smaller index to the other endpoint. For $\Gamma \in \mathcal{C}_m(\sigma)$, write its components as $\Gamma^{(1)}, \ldots, \Gamma^{(k)}$. The initial vertex of every oriented component belongs to ∂F_m. Applying Lemma 3.1 to all components produces an ordered k-tuple of nonempty self-avoiding walks starting at O, with the same respective lengths.

The map from Γ to this tuple is injective. Indeed, the initial boundary vertex of each component is fixed by σ, so the inverse of the corresponding automorphism in Lemma 3.1 recovers that component. If $\ell_i = \#E(\Gamma^{(i)})$, then

$$\#E(\Gamma) = \ell_1 + \cdots + \ell_k.$$

Using (17) and dropping the requirements that the recovered components be mutually disjoint and have state σ, we obtain

$$Z_m^\sigma(\beta) \leq \sum_{\ell_1, \ldots, \ell_k \geq 1} \prod_{i=1}^{k} c_{\ell_i} e^{-\beta \ell_i} = \big(\chi(\beta) - 1\big)^k.$$

The number k is bounded by b, and there are only finitely many states. Hence $\sup_m \|\mathbf{Z}_m(\beta)\|_\infty < \infty$, so $\beta \in \mathcal{B}$, as required. □

Proof of Theorem 2(i). By (9), $\chi(\beta) < \infty$ whenever $\beta > \log(D-1)$. Proposition 4.6(ii) therefore shows that $\mathcal{B} \neq \varnothing$, so $\beta_{\mathrm{c}} < \infty$.

Define the susceptibility threshold

$$\beta_\chi = \inf\{\beta \geq 0 : \chi(\beta) < \infty\}.$$

If $\beta \in \mathcal{B}$, Proposition 4.6(i) gives $\chi(\beta') < \infty$ for every $\beta' > \beta$. Hence $\beta_\chi \leq \beta$, and taking the infimum over $\beta \in \mathcal{B}$ yields

$$\beta_\chi \leq \beta_{\mathrm{c}}.$$

Conversely, if $\chi(\beta) < \infty$, then Proposition 4.6(ii) gives $\beta \in \mathcal{B}$. Taking the infimum over such β yields

$$\beta_{\mathrm{c}} \leq \beta_\chi.$$

Thus $\beta_{\mathrm{c}} = \beta_\chi$.

Finally, Theorem 1 gives

$$\lim_{n\to\infty} \big(c_n e^{-\beta n}\big)^{1/n} = \mu e^{-\beta}.$$

The root test applied to (23) shows that $\chi(\beta)$ converges for $\beta > \log \mu$ and diverges for $\beta < \log \mu$. Therefore $\beta_\chi = \log \mu$, and hence $\beta_{\mathrm{c}} = \log \mu$. Equivalently, $\mu = e^{\beta_{\mathrm{c}}}$, as claimed. □

4.3 Polynomial bounds

We now prove Theorem 2(ii), assuming that the infimum in (18) is attained. The upper-bound argument is the boundary-state form of the scale-localization method used by Hattori and Kusuoka on the Sierpiński gasket [13]. Their lower bound used a detailed analysis of the critical two-variable recursion; the lower bound below instead follows from the finite convolution estimate (13).

Proof of Theorem 2(ii). Assume that $\beta_{\mathrm{c}} = \min \mathcal{B}$. By Theorem 2(i), $\mu = e^{\beta_{\mathrm{c}}}$.

Upper bound. Since $\beta_{\mathrm{c}} \in \mathcal{B}$, Proposition 4.4 applies at $\beta = \beta_{\mathrm{c}}$. Together with Lemma 4.5, it gives constants $K \geq 1$ and $a \geq 0$ such that, for $n \geq 1$,

$$c_n e^{-\beta_{\mathrm{c}} n} \leq \#\mathcal{S}^{\bullet}\, \|\mathbf{A}_{m(n)}(\beta_{\mathrm{c}})\|_{\infty} \leq K(n+1)^a.$$

Since $n+1 \leq 2n$ for $n \geq 1$, increasing K gives

$$c_n \leq K \mu^n n^a, \qquad n \geq 1.$$

Lower bound. Recall the constants C, Q from Proposition 3.3 and the quantity G_n from (14). Put

$$s = Q - 1 + aQ, \qquad B = QC^Q K^Q.$$

The upper bound just proved, together with (13) and (14), gives

$$u_j \leq G_j K^Q \mu^j j^{aQ} \leq B \mu^j j^s, \qquad j \geq 1.$$

Fix $t \geq 1$ and set $n = Qt$. Since $\mu = \inf_{j \geq 1} u_j^{1/j}$ by Theorem 1, we have $u_n \geq \mu^n$. By (13) and (14), there are $q \leq Q$ and positive integers $\ell_1, \dots, \ell_q$, with $\ell_1 + \cdots + \ell_q = n$, such that

$$\prod_{i=1}^{q} c_{\ell_i} \geq \frac{u_n}{G_n} \geq \frac{\mu^n}{QC^Q n^{Q-1}}.$$

Let $k = \max_i \ell_i$. Then $k \geq \lceil n/Q \rceil = t$. Applying the upper bound to all factors except one of length k yields

$$\prod_{i \neq i_*} c_{\ell_i} \leq K^{q-1} \mu^{n-k} n^{a(q-1)}, \qquad \ell_{i_*} = k.$$

Since $q \leq Q$, the definitions of B and s therefore give

$$c_k \geq B^{-1} \mu^k n^{-s}.$$

On the other hand, cutting a rooted k-step walk after t edges gives

$$c_k \leq c_t u_{k-t} \leq B c_t \mu^{k-t} (k - t + 1)^s.$$

Here the last inequality follows from the preceding bound when $k > t$, and from $u_0 = 1$ when $k = t$, after increasing B if necessary. Since $k \leq n = Qt$, both n and $k - t + 1$ are at most Qt. Combining the last two estimates gives

$$c_t \geq B^{-2} Q^{-2s} \mu^t t^{-2s}.$$

Taking $p = 2s$ and increasing a single constant $A \geq 1$ to cover both bounds proves the assertion. □

We conclude this section with a useful criterion for the uniform boundedness of $\mathbf{Z}_m(\beta)$.

Remark 4.7. For distinct $i, j \in \{0, \dots, b-1\}$, let $e_{ij} \in \mathcal{S}$ be the abstract boundary forest consisting of the single edge $\{q_i, q_j\}$, and put

$$P_m^{ij}(\beta) = Z_m^{e_{ij}}(\beta).$$

Then

$$\sup_{m\geq 0} \|\mathbf{Z}_m(\beta)\|_\infty < \infty \quad \Longleftrightarrow \quad \sup_{m\geq 0} \max_{i\neq j} P_m^{ij}(\beta) < \infty.$$

Indeed, put $P_m(\beta) = \max_{i\neq j} P_m^{ij}(\beta)$. Fix $\sigma \in \mathcal{S}$ and $\Gamma \in \mathcal{C}_m(\sigma)$. Cutting each component of Γ at its successive visits to ∂F_m produces, for every abstract edge $\{q_i, q_j\}$ of σ, a self-avoiding path in $\mathcal{C}_m(e_{ij})$. These paths partition the edge set of Γ. The resulting tuple determines Γ, and its weights multiply. Hence

$$Z_m^\sigma(\beta) \leq \prod_{\{q_i,q_j\} \text{ an edge of } \sigma} P_m^{ij}(\beta).$$

Since σ is a linear forest on at most b vertices, it has at most $b-1$ edges. Consequently,

$$P_m(\beta) \leq \|\mathbf{Z}_m(\beta)\|_\infty \leq \max\{1, P_m(\beta)^{\,b-1}\}.$$

5 Polygonal N-gaskets

This section proves Theorem 3. We first determine the contact geometry of the first-level fractal cells and the exact crossing recursions. We then compare the crossing critical inverse temperature with the critical inverse temperature of the boundary-state renormalization and compute the connective constants for $N = 6$ and $N = 9$. The complete dynamics of the recursions is given in Appendix A.

Definition 5.1 (Regular polygonal N-gaskets)**.** Let $N \geq 3$ with $4 \nmid N$. Identify $\mathbb{R}^2$ with $\mathbb{C}$, and put

$$q_j = \exp(2\pi \mathrm{i} j/N), \qquad V_0^{(N)} = \{q_j : j \in \mathbb{Z}/N\mathbb{Z}\}.$$

Let $F_0^{(N)}$ be the cycle on $V_0^{(N)}$. Write

$$k_N = \lfloor N/4 \rfloor, \qquad \ell_N = N - 4k_N, \qquad s_N = k_N + 1,$$

and define

$$\rho_N = \frac{\sin(\pi/N)}{\sin(\pi/N) + \sin((2s_N - 1)\pi/N)}, \qquad \Psi_j^{(N)}(x) = q_j + \rho_N(x - q_j).$$

The attractor of the N homotheties $\{\Psi_j^{(N)} : j \in \mathbb{Z}/N\mathbb{Z}\}$ is the *regular polygonal N-gasket*. Its finite graphs and its one-sided graph are denoted by $F_m^{(N)}$ and $F_\infty^{(N)}$. In the construction (3), the indices are relabeled by $\mathbb{Z}/N\mathbb{Z}$, the root is $O = q_0$, and $\Psi_0^{(N)}$ is the map fixing O. The number ρ_N is the *first-contact ratio* [20, Section 7]. When $4 \mid N$, two consecutive first-level fractal cells in the first-contact construction share a non-degenerate line segment. Since $V_0^{(N)}$ is finite, this intersection cannot equal the intersection of their boundary vertex sets. Thus (NF4) fails, so this construction is not a nested fractal and is excluded from the definition.



(a) the 5-gasket (b) the 6-gasket

Figure 6: First-level polygonal hulls for two polygonal N-gaskets. The dashed polygon has vertex set $V_0^{(N)}$, and each blue polygon is the hull of one first-level fractal cell. In (a), ζ_0 marks the contact point between two adjacent first-level fractal cells.

Figure 6 illustrates the first-contact construction. For a fixed polygonal N-gasket, put

$$a_{j,m} = (\Psi_0^{(N)})^{-m}(q_j), \qquad j \in \mathbb{Z}/N\mathbb{Z}, \quad m \geq 0.$$

A *crossing* of $F_m^{(N)}$ is a self-avoiding walk between two prescribed distinct vertices of $\partial F_m^{(N)}$.

If $N \notin \{3, 5\}$, define

$$\begin{aligned} X_m^{(N)}(\beta) &= \sum_{\substack{w: a_{0,m} \to a_{s_N,m} \\ w \text{ self-avoiding in } F_m^{(N)}}} e^{-\beta L(w)}, \\ Y_m^{(N)}(\beta) &= \sum_{\substack{w: a_{-s_N,m} \to a_{s_N,m} \\ w \text{ self-avoiding in } F_m^{(N)}}} e^{-\beta L(w)}. \end{aligned} \tag{24}$$

Thus $X_m^{(N)}$ joins $a_{0,m}$ to $a_{s_N,m}$, whereas $Y_m^{(N)}$ joins $a_{-s_N,m}$ to $a_{s_N,m}$.

For $N = 3$, let $x_m(\beta)$ and $y_m(\beta)$ be the partition functions of crossings between two fixed boundary vertices which, respectively, avoid and visit the third boundary vertex. For $N = 5$, let $\phi_{1,m}(\beta)$ and $\phi_{2,m}(\beta)$ count crossings from $a_{0,m}$ to $a_{2,m}$ which, respectively, avoid and visit $a_{-2,m}$; let $\theta_{1,m}(\beta)$ and $\theta_{2,m}(\beta)$ count crossings from $a_{-2,m}$ to $a_{2,m}$ which, respectively, avoid and visit $a_{0,m}$. In each case the weight of a walk is $e^{-\beta L(w)}$. Define the *crossing vector*

$$\mathbf{C}_m^{(N)}(\beta) = \begin{cases} (x_m(\beta), y_m(\beta)), & N = 3, \\ (\phi_{1,m}(\beta), \phi_{2,m}(\beta), \theta_{1,m}(\beta), \theta_{2,m}(\beta)), & N = 5, \\ (X_m^{(N)}(\beta), Y_m^{(N)}(\beta)), & N \notin \{3, 5\}, \end{cases}$$

and its *crossing critical inverse temperature* by

$$\beta_{\mathrm{c},N}^{\mathrm{cr}} = \inf\left\{\beta \geq 0 : \sup_{m \geq 0} \|\mathbf{C}_m^{(N)}(\beta)\|_\infty < \infty\right\}. \tag{25}$$

The infimum is understood in $[0, \infty]$. Let μ_N and $\beta_{\mathrm{c},N}$ denote, respectively, the connective constant and the critical inverse temperature of $F_\infty^{(N)}$, with the latter defined by Definition 4.2.

Throughout this section, fix $N \geq 3$ with $4 \nmid N$. We use the polygonal N-gasket from Definition 5.1 and omit the superscript (N) from $F_m^{(N)}$ and $F_\infty^{(N)}$, and also from $X_m^{(N)}$ and $Y_m^{(N)}$.

5.1 Geometry and crossing recursions

The following proposition describes the contact pattern that determines the crossing recursions.

For a first-level fractal cell $\Psi_j^{(N)}(K)$, the fixed point q_j of $\Psi_j^{(N)}$ is called its *outer vertex*. More generally, when a parent fractal cell is decomposed into its immediate fractal subcells, the corresponding image of q_j is called the outer vertex of the j-th subcell.

Proposition 5.2. *The following statements hold.*

(i) *The polygonal N-gasket satisfies* (NF1)–(NF4) *and the one-point intersection condition* (2).

(ii) *The incidence graph of its first-level fractal cells is the cycle graph C_N on $\mathbb{Z}/N\mathbb{Z}$, with an edge between j and $j+1$ for each j modulo N.*

(iii) *The outer vertex of a first-level fractal cell is its unique vertex in $V_0^{(N)}$. Fix such a fractal cell and identify its boundary with $V_0^{(N)}$ by applying the inverse cell map followed by a dihedral relabeling that sends its outer vertex to q_0. Under this identification, its contact points with the previous and next first-level fractal cells in the cyclic order correspond to q_{-s_N} and q_{s_N}, respectively.*

Proof. We first prove (ii) and (iii). The standard first-contact calculation for two consecutive homothetic regular N-gons gives

$$\Psi_0^{(N)}(q_{s_N}) = \Psi_1^{(N)}(q_{1-s_N}).$$

For $4 \nmid N$, this is the unique contact of the two fractal cells. Rotating the identity by one step in the opposite direction also gives

$$\Psi_{-1}^{(N)}(q_{s_N-1}) = \Psi_0^{(N)}(q_{-s_N}).$$

Thus, under $(\Psi_0^{(N)})^{-1}$, the contact points of the first-level fractal cell $\Psi_0^{(N)}(K)$ with $\Psi_{-1}^{(N)}(K)$ and $\Psi_1^{(N)}(K)$ correspond to q_{-s_N} and q_{s_N}, respectively. A dihedral symmetry gives the same normalized description for every first-level fractal cell.

Rotating the configuration gives one contact for every consecutive pair, while nonconsecutive polygonal hulls are disjoint at the first-contact ratio [20, Section 7]. Hence the incidence graph is C_N. Moreover, each first-level polygonal hull contains only its fixed point from $V_0^{(N)}$, so this fixed point is the unique outer vertex of the corresponding fractal cell. This proves (ii) and (iii), as well as the first-level form of (NF4) and the one-point intersection condition.

It remains to prove (i). The interiors of the first-level polygonal hulls give (NF1), their cyclic contact pattern gives (NF2), and the dihedral symmetries of the regular polygon give (NF3). The preceding contact description gives (NF4) at the first level, and iteration gives the one-point intersection condition at every level. □

The symbols $\mathcal{S}$, $\mathcal{C}_m(\sigma)$, and $\mathbf{Z}_m$ refer to the boundary states, boundary forests, and boundary-state partition vector of this polygonal N-gasket. The resulting recursions are as follows. Every quantity in the proposition is evaluated at the same fixed $\beta \geq 0$.

Proposition 5.3. *If $N = 4k_N + \ell_N \notin \{3,5\}$, where $\ell_N \in \{1,2,3\}$, then the functions in* (24) *satisfy*

$$\begin{aligned} X_0 &= e^{-(k_N+1)\beta} + e^{-(3k_N+\ell_N-1)\beta}, & Y_0 &= e^{-2(k_N+1)\beta} + e^{-(2k_N+\ell_N-2)\beta}, \\ X_{m+1} &= X_m^2\big(Y_m^{k_N} + Y_m^{3k_N+\ell_N-2}\big), & Y_{m+1} &= X_m^2\big(Y_m^{2k_N+\ell_N-3} + Y_m^{2k_N+1}\big). \end{aligned} \tag{26}$$

For $N = 3$,

$$\begin{aligned} x_0 &= e^{-\beta}, & y_0 &= e^{-2\beta}, \\ x_{m+1} &= (x_m + y_m)^2 + x_m^2(x_m + 2y_m), & y_{m+1} &= x_m y_m (x_m + 2y_m). \end{aligned} \tag{27}$$

For $N = 5$, put

$$\bar{\phi}_m = \phi_{1,m} + \phi_{2,m}, \qquad \bar{\theta}_m = \theta_{1,m} + \theta_{2,m}.$$

Then

$$\begin{aligned}(\phi_{1,0},\phi_{2,0},\theta_{1,0},\theta_{2,0}) &= (e^{-2\beta},e^{-3\beta},e^{-\beta},e^{-4\beta}),\\ \phi_{1,m+1} &= \bar{\phi}_m^2\bar{\theta}_m(1+\theta_{1,m}),\\ \phi_{2,m+1} &= \bar{\phi}_m^2\bar{\theta}_m\theta_{2,m},\\ \theta_{1,m+1} &= \bar{\phi}_m^2+\bar{\theta}_m^2\theta_{1,m}\phi_{1,m}(\phi_{1,m}+2\phi_{2,m}),\\ \theta_{2,m+1} &= \bar{\theta}_m^2\theta_{2,m}\phi_{1,m}(\phi_{1,m}+2\phi_{2,m}).\end{aligned} \tag{28}$$

Proof. At level zero, $F_0 = C_N$, so the two paths between any prescribed pair of vertices are the two polygonal arcs. Their lengths give all the initial conditions in (26)–(28).

Inside an $(m+1)$-cell, label the m-subcells $\Delta_0,\dots,\Delta_{N-1}$ so that Δ_j contains the outer vertex $a_{j,m+1}$, and put

$$\zeta_j = \Delta_j \cap \Delta_{j+1}.$$

By Proposition 5.2(i)–(ii) and self-similarity, ζ_j is the unique contact between two consecutive subcells, and each subcell meets the rest of the parent only at its two contacts ζ_{j-1} and ζ_j. By Proposition 2.2, each edge belongs to a unique m-subcell. For a crossing, its *subcell itinerary* is the sequence of m-subcells containing its successive edges, with consecutive repetitions deleted. Once a crossing enters and leaves an internal subcell, both contacts have been used, so self-avoidance prevents a later return. An endpoint subcell cannot be re-entered either, since the other endpoint lies in a different subcell. Hence the subcell itinerary is one of the two simple arcs of C_N. Dihedral symmetry identifies every local orientation of a piece with the corresponding partition function.

We first prove (26). An X-crossing has two endpoint pieces of type X_m. Its short arc has $s_N - 1 = k_N$ interior pieces of type Y_m, and its long arc has

$$N - s_N - 1 = 3k_N + \ell_N - 2$$

such pieces. This gives the recursion for X_{m+1}. For a Y-crossing, the two arcs have

$$2s_N - 1 = 2k_N + 1 \quad \text{and} \quad N - 2s_N - 1 = 2k_N + \ell_N - 3$$

interior cells. This gives the recursion for Y_{m+1}. Since $s_N = \lfloor N/4 \rfloor + 1$, the endpoint cells of an X-crossing are adjacent only when $N = 3$. For a Y-crossing, they are adjacent precisely when $N - 2s_N = 1$, equivalently when $N \in \{3,5\}$. Thus, in the present case $N \notin \{3,5\}$, neither pair of endpoint cells is adjacent, and multiplication of the child partition functions is exact.

For $N = 3$, consider a crossing from $a_{0,m+1}$ to $a_{1,m+1}$. The direct itinerary (Δ_0,Δ_1) avoids $a_{2,m+1}$, and each of its two pieces may either avoid or visit its contact with Δ_2. The two contacts are distinct, so the total contribution is $(x_m + y_m)^2$.

The long itinerary is $(\Delta_0,\Delta_2,\Delta_1)$. Its two endpoint subcells also meet at one common contact. Either endpoint piece may visit that contact, but not both, since the resulting walk would otherwise repeat the vertex. The total weight of the allowed endpoint choices is therefore

$$(x_m + y_m)^2 - y_m^2 = x_m(x_m + 2y_m).$$

The middle piece has type x_m when the parent crossing avoids $a_{2,m+1}$, and type y_m when it visits that vertex. The long itinerary consequently contributes $x_m^2(x_m + 2y_m)$ to x_{m+1} and $x_m y_m(x_m + 2y_m)$ to y_{m+1}. This proves (27), the two-variable recursion introduced in [11].

It remains to prove (28). Consider first a crossing from $a_{0,m+1}$ to $a_{2,m+1}$. Its short itinerary $(\Delta_0,\Delta_1,\Delta_2)$ has two endpoint pieces of total weight $\bar{\phi}_m^2$ and one interior piece of weight $\bar{\theta}_m$. It avoids $a_{3,m+1}$, so its contribution to $\phi_{1,m+1}$ is $\bar{\phi}_m^2\bar{\theta}_m$. The long itinerary is $(\Delta_0,\Delta_4,\Delta_3,\Delta_2)$. The piece in Δ_4 has total weight $\bar{\theta}_m$, whereas the piece in Δ_3 has weight $\theta_{1,m}$ or $\theta_{2,m}$, according as it avoids or visits its outer vertex $a_{3,m+1}$. Hence the long itinerary contributes $\bar{\phi}_m^2\bar{\theta}_m\theta_{1,m}$ to $\phi_{1,m+1}$ and $\bar{\phi}_m^2\bar{\theta}_m\theta_{2,m}$ to $\phi_{2,m+1}$.

Next, consider a crossing from $a_{-2,m+1} = a_{3,m+1}$ to $a_{2,m+1}$. Its short itinerary (Δ_3, Δ_2) contributes $\bar{\phi}_m^2$ and avoids $a_{0,m+1}$, so it contributes $\bar{\phi}_m^2$ to $\theta_{1,m+1}$. Its long itinerary is $(\Delta_3, \Delta_4, \Delta_0, \Delta_1, \Delta_2)$. The pieces in Δ_4 and Δ_1 contribute $\bar{\theta}_m^2$. The middle piece in Δ_0 has type $\theta_{1,m}$ or $\theta_{2,m}$, according as the parent crossing avoids or visits $a_{0,m+1}$. The endpoint cells of the long itinerary share one common contact. Both endpoint pieces may visit it, but not simultaneously. The total weight of the allowed endpoint choices is

$$\bar{\phi}_m^2 - \phi_{2,m}^2 = \phi_{1,m}(\phi_{1,m} + 2\phi_{2,m}).$$

Multiplication of these factors gives the two θ-recursions and completes the proof of (28). □

Proposition 5.4. *For every $\beta \geq 0$,*

$$\sup_{m\geq 0} \|\mathbf{C}_m^{(N)}(\beta)\|_\infty < \infty \quad \Longleftrightarrow \quad \sup_{m\geq 0} \|\mathbf{Z}_m(\beta)\|_\infty < \infty.$$

Consequently, $\beta_{c,N}^{cr} = \beta_{c,N}$.

Proof. Each partition function defining $\mathbf{C}_m^{(N)}(\beta)$ is a finite sum of partition functions in $\mathbf{Z}_m(\beta)$: a crossing is a one-component boundary forest, and its boundary state records the additional boundary vertices, if any, which the crossing visits. Thus

$$\sup_{m\geq 0} \|\mathbf{Z}_m(\beta)\|_\infty < \infty \quad \Longrightarrow \quad \sup_{m\geq 0} \|\mathbf{C}_m^{(N)}(\beta)\|_\infty < \infty.$$

Conversely, suppose that $\|\mathbf{C}_m^{(N)}(\beta)\|_\infty \leq K$ for every m. Define the two total crossing weights by

$$\mathsf{O}_m = \begin{cases} x_m + y_m, & N = 3, \\ \bar{\phi}_m, & N = 5, \\ X_m, & N \notin \{3,5\}, \end{cases} \qquad \mathsf{P}_m = \begin{cases} x_m + y_m, & N = 3, \\ \bar{\theta}_m, & N = 5, \\ Y_m, & N \notin \{3,5\}. \end{cases}$$

Both quantities are at most $2K$.

For distinct $i, j \in \mathbb{Z}/N\mathbb{Z}$, let $P_m^{ij}(\beta)$ be the one-edge partition function from Remark 4.7. The same contact argument used in the proof of Proposition 5.3 shows that a path counted by $P_m^{ij}(\beta)$ has one of the two simple arcs of C_N as its subcell itinerary. Let $r \in \{1, \ldots, N-1\}$ be determined by $j - i \equiv r \pmod N$. Discarding possible intersections between the pieces gives, for $m \geq 1$,

$$P_m^{ij}(\beta) \leq \mathsf{O}_{m-1}^2 \left(\mathsf{P}_{m-1}^{r-1} + \mathsf{P}_{m-1}^{N-r-1} \right).$$

Since N is fixed, the right-hand side is bounded uniformly in m, i, j; the finitely many level-zero values are bounded as well. Remark 4.7 now gives $\sup_m \|\mathbf{Z}_m(\beta)\|_\infty < \infty$, which proves the reverse implication. Taking the infima in (25) and (18) proves the desired equality of critical inverse temperatures. □

5.2 Critical values and connective constants

Proof of Theorem 3. Proposition 5.2(i) places $F_\infty^{(N)}$ in the class of nested fractal graphs considered in Theorems 1 and 2. The closed polynomial recursions are given in Proposition 5.3. By Proposition 5.4 and Theorem 2(i),

$$\beta_{c,N}^{cr} = \beta_{c,N} = \log \mu_N.$$

It remains to compute this value when $N = 6$ or $N = 9$. In these two cases the initial crossing variables in (26) are equal, and the recursion preserves their equality. Writing their common value as t_m, we have

$$t_0 = e^{-d_N \beta} + e^{-2d_N \beta}, \qquad t_{m+1} = F(t_m), \qquad F(u) = u^{d_N+1} + u^{2d_N+1}.$$

The fixed points of F in $[0,\infty)$ are 0 and

$$t_* = \tau^{1/d_N},$$

because, for $u > 0$, $F(u) = u$ is equivalent to $u^{d_N} + u^{2d_N} = 1$, whose unique solution is t_* since $\tau + \tau^2 = 1$. Moreover,

$$F(u) - u = u\big(u^{d_N} + u^{2d_N} - 1\big), \qquad \operatorname{sgn}\big(F(u) - u\big) = \operatorname{sgn}(u - t_*) \quad (u > 0),$$

since $u \mapsto u^{d_N} + u^{2d_N}$ is strictly increasing. If $t_0 < t_*$, induction gives $0 < t_{m+1} < t_m < t_*$; its limit is a fixed point below t_*, hence $t_m \to 0$. If $t_0 = t_*$, then $t_m = t_*$ for every m. If $t_0 > t_*$, induction gives $t_{m+1} > t_m > t_*$; if the sequence were bounded, its limit would be a fixed point above t_*, which is impossible. Hence $t_m \to \infty$.

Consequently, the crossing orbit is bounded exactly when

$$e^{-d_N\beta} + e^{-2d_N\beta} \le \tau^{1/d_N}.$$

Since the left-hand side decreases continuously from 2 to 0 as β runs from 0 to ∞, (25) shows that $\beta_{\mathrm{c},N}^{\mathrm{cr}}$ is the unique positive solution of

$$e^{-d_N\beta_{\mathrm{c},N}^{\mathrm{cr}}} + e^{-2d_N\beta_{\mathrm{c},N}^{\mathrm{cr}}} = \tau^{1/d_N}.$$

Setting $z = e^{-d_N\beta_{\mathrm{c},N}^{\mathrm{cr}}}$ gives

$$z + z^2 = \tau^{1/d_N}, \qquad z = \frac{\sqrt{1 + 4\tau^{1/d_N}} - 1}{2}.$$

Since $\mu_N = e^{\beta_{\mathrm{c},N}^{\mathrm{cr}}}$, this is precisely (5). □

6 Ratio limits

We prove Theorem 4(i) as follows. In Subsection 6.1, Lemma 6.2 gives the state-preserving cell surgery, and Proposition 6.3 gives the positive-density estimate for canonical cell patterns. In Subsection 6.2, the marked identities of Lemma 6.4 convert this abundance, under the flexibility hypothesis, into the regularity estimates of Proposition 6.5, and Lemma 6.6 then gives ratio convergence. Theorem 4(i) is a finite-ramification version of the local replacement mechanism in ratio theorems such as Kesten's [16]. Subsection 6.3 verifies the hypothesis for the Sierpiński gasket. Subsection 6.4 determines the smallest flexibility step for every polygonal N-gasket, thereby proving Theorem 4(ii), and Subsection 6.5 proves part (iii).

Definition 6.1 (Finite-scale flexibility)**.** For $r \ge 1$, let

$$\mathcal{S}_r = \{\sigma \in \mathcal{S} : \mathcal{C}_r(\sigma) \ne \varnothing\}.$$

Given an integer $h \ge 1$, the graph is *h-flexible at scale r* if, for every $\sigma \in \mathcal{S}_r$, there are forests

$$\Gamma_\sigma^-, \Gamma_\sigma^+ \in \mathcal{C}_r(\sigma) \quad \text{such that} \quad \#E(\Gamma_\sigma^+) = \#E(\Gamma_\sigma^-) + h.$$

Thus every realizable way of connecting the boundary of an r-cell has two internal realizations whose lengths differ by exactly h. The graph is *h-flexible* if it is h-flexible at some finite scale; such an h is called a *flexibility step*.

6.1 Local surgery and canonical patterns

Fix a scale $r \geq 1$. For each r-cell Δ, fix one of the isometries

$$S_\Delta : F_r \longrightarrow \Delta$$

provided by (8). If w traverses an edge of Δ, let $\Gamma_\Delta(w)$ be the edge subgraph of Δ consisting of all edges of w that belong to Δ.

Lemma 6.2. *Let w be a self-avoiding walk, and let Δ be an r-cell which contains neither endpoint of w. If $\Gamma_\Delta(w) \neq \varnothing$, then $S_\Delta^{-1}\Gamma_\Delta(w)$ is a boundary forest in F_r. Moreover, it may be replaced by any other boundary forest with the same boundary state. The replacement yields a self-avoiding walk with the same endpoints as w. Such replacements may be performed simultaneously in any finite collection of distinct r-cells.*

Proof. Since Δ contains neither endpoint of w, every degree-one vertex of $\Gamma_\Delta(w)$ lies in $\partial\Delta$, so its pullback is a boundary forest. By Proposition 2.2, the unchanged part of w meets Δ only at boundary vertices, whose degrees and connections are recorded by the boundary state. A same-state replacement therefore reconnects the unchanged pieces in the same order without a new intersection. Since distinct r-cells have disjoint edges and share only boundary vertices, the replacements may be performed simultaneously. □

The next proposition gives the positive-density estimate for canonical cell patterns.

Proposition 6.3. *Fix $r \geq 1$. For each $\sigma \in \mathcal{S}_r$, fix a forest $\Gamma_\sigma \in \mathcal{C}_r(\sigma)$. For a walk w and an r-cell Δ that contains neither endpoint of w and satisfies $\Gamma_\Delta(w) \neq \varnothing$, put*

$$\sigma_\Delta(w) = \mathrm{tr}_r\big(S_\Delta^{-1}\Gamma_\Delta(w)\big).$$

We call Δ canonical for w *if*

$$S_\Delta^{-1}\Gamma_\Delta(w) = \Gamma_{\sigma_\Delta(w)}.$$

Let $\mathsf{m}_r(w)$ be the number of canonical r-cells of w. There is a constant $a > 0$ such that

$$\limsup_{n\to\infty} \#\{w \in \mathcal{W}_{O,n} : \mathsf{m}_r(w) \leq \lfloor an \rfloor\}^{1/n} < \mu. \tag{29}$$

Proof. Put $e_r = \#E(F_r)$, and let J_r be the number of nonempty edge subgraphs of F_r; it is enough to use the bound $J_r \leq 2^{e_r}$. Every edge of a walk belongs to one r-cell by Proposition 2.2, and one r-cell contains at most e_r of its edges. Hence an n-step walk traverses edges in at least $\lceil n/e_r \rceil$ distinct r-cells.

A vertex belongs to at most D distinct r-cells. Indeed, if it belongs to more than one, it is a boundary vertex of each of them, and each cell contributes a different incident edge; the degree is at most D. After discarding the cells which contain one of the two endpoints of the walk, at least

$$\left\lceil \frac{n}{e_r} \right\rceil - 2D$$

cells remain. Thus, with $\varpi_r = (2e_r)^{-1}$, every sufficiently long walk has at least $\lceil \varpi_r n \rceil$ endpoint-free cells with a nonempty restriction.

Let $\mathcal{B}_n(a)$ denote the cardinality of the set in (29). Choose $a, \delta > 0$ with $a + \delta < \varpi_r$, and put $k = \lfloor \delta n \rfloor$. From each walk counted by $\mathcal{B}_n(a)$, choose k of its endpoint-free cells whose restrictions are not canonical. There are at least

$$\binom{\lfloor (\varpi_r - a)n \rfloor}{k}$$

choices. In each chosen cell replace the restriction by the canonical forest having the same boundary state. By Lemma 6.2, all k replacements can be made simultaneously and the output is a self-avoiding walk with the same endpoints. Each replacement changes the length by at most e_r, so the output length lies between $n - e_r k$ and $n + e_r k$.

The chosen cells were noncanonical before replacement, and no unchosen restriction is changed. Hence the output has exactly k more canonical cells than the input, and therefore at most $k + \lfloor an \rfloor$ canonical cells. Given an output, the modified cells must be chosen among these canonical cells; there are at most

$$\binom{k + \lfloor an \rfloor}{k}$$

ways. In each selected cell there are at most J_r possible original restrictions. Keeping choices which do not reconstruct a valid walk only enlarges the bound. We have proved

$$\mathcal{B}_n(a) \binom{\lfloor (\varpi_r - a)n \rfloor}{k} \le J_r^k \binom{k + \lfloor an \rfloor}{k} \sum_{\substack{j \ge 0 \\ |j-n| \le e_r k}} c_j. \tag{30}$$

Let $\mathcal{H}(t) = -t \log t - (1-t)\log(1-t)$, with $\mathcal{H}(0) = \mathcal{H}(1) = 0$. For every $\lambda > \mu$, Theorem 1 gives a constant $C_\lambda < \infty$ such that

$$c_j \le C_\lambda \lambda^j, \qquad j \ge 0.$$

Apply $\binom{q}{p} \le \exp\bigl(q\mathcal{H}(p/q)\bigr)$ to the binomial coefficient

$$\binom{k + \lfloor an \rfloor}{k}$$

on the right-hand side of (30). Apply the Stirling lower bound $\binom{q}{p} \ge \exp\bigl(q\mathcal{H}(p/q) - o(q)\bigr)$ to the binomial coefficient

$$\binom{\lfloor (\varpi_r - a)n \rfloor}{k}$$

on the left-hand side of (30). Using $c_j \le C_\lambda \lambda^j$ and $\sum_{|j-n| \le e_r k} c_j \le C'_{\lambda,r} \lambda^{n + e_r k}$, divide (30) by its left-hand binomial coefficient and take n-th roots. Since

$$\frac{k}{n} \longrightarrow \delta, \qquad \frac{k + \lfloor an \rfloor}{n} \longrightarrow \delta + a, \qquad \frac{\lfloor (\varpi_r - a)n \rfloor}{n} \longrightarrow \varpi_r - a,$$

the factor J_r^k contributes $\delta \log J_r$ to the logarithm of the resulting upper bound. The two binomial coefficients contribute, respectively,

$$(\delta + a)\mathcal{H}\left(\frac{\delta}{\delta + a}\right) \quad \text{and} \quad -(\varpi_r - a)\mathcal{H}\left(\frac{\delta}{\varpi_r - a}\right),$$

while the sum over j contributes at most $(1 + e_r\delta)\log\lambda$. It follows that

$$\limsup_{n \to \infty} \mathcal{B}_n(a)^{1/n} \le \lambda^{1 + e_r\delta} \exp\left\{ \delta \log J_r + (\delta + a)\mathcal{H}\left(\frac{\delta}{\delta + a}\right) - (\varpi_r - a)\mathcal{H}\left(\frac{\delta}{\varpi_r - a}\right) \right\}.$$

Set $a = \delta^2$. Then

$$(\delta + \delta^2)\mathcal{H}\left(\frac{\delta}{\delta + \delta^2}\right) = O\left(\delta^2 \log(1/\delta)\right).$$

Moreover, $\delta \log J_r = O(\delta)$, whereas

$$(\varpi_r - a)\mathcal{H}\left(\frac{\delta}{\varpi_r - a}\right) = \delta \log(1/\delta) + O(\delta).$$

Thus the expression in braces is $-\delta \log(1/\delta) + O(\delta)$. Since $e_r \delta \log \mu = O(\delta)$, for all sufficiently small δ the expression in braces is smaller than $-e_r \delta \log \mu$. Fix such a δ. The continuous expression on the right of the preceding estimate has value strictly below μ at $\lambda = \mu$. Hence it remains below μ for some $\lambda > \mu$ sufficiently close to μ, for which the estimate is valid. This proves (29). □

6.2 Ratio estimates

Suppose from now on that the graph is h-flexible at scale r. For each $\sigma \in \mathcal{S}_r$, choose and fix forests $\Gamma_\sigma^-, \Gamma_\sigma^+ \in \mathcal{C}_r(\sigma)$, as in Definition 6.1, with Γ_σ^+ having exactly h more edges than Γ_σ^-. For a walk w and an endpoint-free r-cell Δ with $\Gamma_\Delta(w) \neq \varnothing$, write

$$\sigma_\Delta(w) = \mathrm{tr}_r\big(S_\Delta^{-1}\Gamma_\Delta(w)\big).$$

Let $m_-(w)$ and $m_+(w)$ be the numbers of such cells for which, respectively,

$$S_\Delta^{-1}\Gamma_\Delta(w) = \Gamma^-_{\sigma_\Delta(w)} \quad \text{or} \quad S_\Delta^{-1}\Gamma_\Delta(w) = \Gamma^+_{\sigma_\Delta(w)}.$$

Cells satisfying the first and second conditions are called *minus occurrences* and *plus occurrences*, respectively. Replacing a minus occurrence by the corresponding plus occurrence is a *switch*: it preserves the boundary state and increases the walk length by exactly h. Apply Proposition 6.3 first with $\Gamma_\sigma = \Gamma_\sigma^-$ and then with $\Gamma_\sigma = \Gamma_\sigma^+$. Together with Theorem 1, this gives constants $\alpha > 0$, $\vartheta \in (0,1)$, and n_0 such that, for each $\varepsilon \in \{-,+\}$,

$$\frac{\#\{w \in \mathcal{W}_{O,n} : m_\varepsilon(w) < \lfloor \alpha n \rfloor\}}{c_n} \leq \vartheta^n, \qquad n \geq n_0. \tag{31}$$

Indeed, take α smaller than both constants supplied by Proposition 6.3. For all sufficiently large n, each set in (31) is then contained in the corresponding exceptional set from that proposition. The strict inequality between the exceptional exponential growth rates and μ gives a common ϑ. A walk in which at least one switch type occurs fewer than $\lfloor \alpha n \rfloor$ times belongs to one of the two exceptional sets in (31). Hence

$$\frac{\#\{w \in \mathcal{W}_{O,n} : m_-(w) < \lfloor \alpha n \rfloor \text{ or } m_+(w) < \lfloor \alpha n \rfloor\}}{c_n} \leq 2\vartheta^n, \qquad n \geq n_0.$$

Thus, except for an exponentially small proportion of walks, both switch types occur at least $\lfloor \alpha n \rfloor$ times.

The next lemma counts marked switches in two ways.

Lemma 6.4. *For $j = 0, 1$, let*

$$\mathcal{E}_s^{(j)} = \#\{w \in \mathcal{W}_{O,s} : m_+(w) \leq j\}.$$

Then, for every $n \geq 1$,

$$c_{n+h} = \mathcal{E}_{n+h}^{(0)} + \sum_{w \in \mathcal{W}_{O,n}} \frac{m_-(w)}{m_+(w)+1}, \tag{32}$$

and

$$c_{n+2h} = \mathcal{E}_{n+2h}^{(1)} + \sum_{w \in \mathcal{W}_{O,n}} \frac{m_-(w)(m_-(w)-1)}{(m_+(w)+1)(m_+(w)+2)}. \tag{33}$$

Proof. For (32), mark one minus occurrence Δ in an n-step walk and replace $\Gamma^-_{\sigma_\Delta(w)}$ by $\Gamma^+_{\sigma_\Delta(w)}$. By Lemma 6.2, this produces a self-avoiding walk of length $n+h$, and

$$m_-(w') = m_-(w) - 1, \qquad m_+(w') = m_+(w) + 1.$$

Replacing the marked plus occurrence in w' by the corresponding minus occurrence recovers w. Thus the operation is reversible.

Assign the marked walk w the weight $1/(m_+(w)+1)$. For a fixed w, summing over its $m_-(w)$ possible marks gives

$$\frac{m_-(w)}{m_+(w)+1}.$$

Conversely, a walk w' with $m_+(w') \geq 1$ has $m_+(w')$ possible marked plus occurrences at which the switch can be reversed. Since $m_+(w') = m_+(w) + 1$, each such choice has weight $1/m_+(w')$, and their total weight is one. Therefore every $(n+h)$-step walk with at least one plus occurrence contributes one to the weighted sum. The remaining $(n+h)$-step walks are counted by $\mathcal{E}^{(0)}_{n+h}$, which proves (32).

For (33), mark an ordered pair of distinct minus occurrences and switch both cells. By Lemma 6.2, this produces a self-avoiding walk w' of length $n+2h$, with $m_+(w') = m_+(w) + 2$. Replacing the two marked plus occurrences by the corresponding minus occurrences recovers w.

Assign each marked walk the weight

$$\frac{1}{(m_+(w)+1)(m_+(w)+2)}.$$

For a fixed w, summing over its $m_-(w)(m_-(w)-1)$ ordered pairs gives the summand in (33). Conversely, a walk w' with $m_+(w') \geq 2$ has $m_+(w')(m_+(w')-1)$ ordered pairs at which both switches can be reversed. Each pair has weight

$$\frac{1}{(m_+(w')-1)m_+(w')},$$

so their total weight is one. Therefore every $(n+2h)$-step walk with at least two plus occurrences contributes one to the weighted sum. The remaining walks are counted by $\mathcal{E}^{(1)}_{n+2h}$, which proves (33). □

Proposition 6.5. *Put*

$$R_n = \frac{c_{n+h}}{c_n}.$$

There are constants $b > 0$, $B < \infty$, and n_1 such that

$$R_n \geq b, \qquad R_{n+h} - R_n \geq -\frac{B}{n}, \qquad n \geq n_1. \tag{34}$$

Proof. We first estimate the two error terms in (32) and (33). An $(n+j)$-step walk has an n-step initial part. After this part is fixed, there are at most D choices at each of the remaining j steps. Therefore

$$c_{n+j} \leq D^j c_n, \qquad j \geq 0.$$

For sufficiently large n, $0 < \lfloor \alpha(n+h) \rfloor$ and $1 < \lfloor \alpha(n+2h) \rfloor$. A walk counted by $\mathcal{E}^{(0)}_{n+h}$ has $m_+(w) = 0 < \lfloor \alpha(n+h) \rfloor$, while a walk counted by $\mathcal{E}^{(1)}_{n+2h}$ has $m_+(w) \leq 1 < \lfloor \alpha(n+2h) \rfloor$. Applying (31) at these two lengths gives

$$\mathcal{E}^{(0)}_{n+h} \leq \vartheta^{n+h} c_{n+h} \leq D^h \vartheta^{n+h} c_n,$$
$$\mathcal{E}^{(1)}_{n+2h} \leq \vartheta^{n+2h} c_{n+2h} \leq D^{2h} \vartheta^{n+2h} c_n.$$

Since h is fixed, there are $C_0 < \infty$ and n_2 such that

$$\frac{\mathcal{E}^{(0)}_{n+h} + \mathcal{E}^{(1)}_{n+2h}}{c_n} \leq C_0 \vartheta^n, \qquad n \geq n_2. \tag{35}$$

We next prove the lower bound for R_n. Since distinct r-cells have disjoint edge sets,

$$m_-(w) \leq n, \qquad m_+(w) \leq n.$$

For all sufficiently large n, if $m_-(w) \geq \lfloor \alpha n \rfloor$, then

$$\frac{m_-(w)}{m_+(w)+1} \geq \frac{\lfloor \alpha n \rfloor}{n+1} \geq \frac{\alpha}{2}.$$

By (31), such walks form a proportion at least $1-\vartheta^n$ of $\mathcal{W}_{O,n}$. Since $\mathcal{E}^{(0)}_{n+h} \ge 0$, equation (32) therefore gives, for all sufficiently large n,

$$R_n \ge \frac{\alpha}{2}(1-\vartheta^n) \ge \frac{\alpha}{3}.$$

This proves the first inequality in (34) with $b = \alpha/3$.

It remains to estimate $R_{n+h} - R_n$. Put

$$\mathcal{M}_{1,n} = \frac{1}{c_n} \sum_{w \in \mathcal{W}_{O,n}} \frac{m_-(w)}{m_+(w)+1},$$
$$\mathcal{M}_{2,n} = \frac{1}{c_n} \sum_{w \in \mathcal{W}_{O,n}} \frac{m_-(w)(m_-(w)-1)}{(m_+(w)+1)(m_+(w)+2)}.$$

By Lemma 6.4, equations (32) and (33) give

$$R_n = \mathcal{M}_{1,n} + \varepsilon_{1,n}, \qquad R_n R_{n+h} = \mathcal{M}_{2,n} + \varepsilon_{2,n},$$

where

$$\varepsilon_{1,n} = \frac{\mathcal{E}^{(0)}_{n+h}}{c_n}, \qquad \varepsilon_{2,n} = \frac{\mathcal{E}^{(1)}_{n+2h}}{c_n}.$$

Thus $\varepsilon_{1,n}, \varepsilon_{2,n} \ge 0$, and (35) gives $\varepsilon_{1,n} + \varepsilon_{2,n} \le C_0 \vartheta^n$. Also $0 \le \mathcal{M}_{1,n} \le n$. Choose $\vartheta_1 \in (\vartheta, 1)$. Since $n\vartheta^n = O(\vartheta_1^n)$, there is $C_1 < \infty$ such that

$$2\mathcal{M}_{1,n}\varepsilon_{1,n} + \varepsilon_{1,n}^2 \le C_1 \vartheta_1^n.$$

Using $\varepsilon_{2,n} \ge 0$, we obtain

$$R_n R_{n+h} - R_n^2 \ge \mathcal{M}_{2,n} - \mathcal{M}_{1,n}^2 - C_1 \vartheta_1^n.$$

Write $\omega(w) = m_-(w)/(m_+(w)+1)$. The Cauchy–Schwarz inequality gives

$$\mathcal{M}_{1,n}^2 \le \frac{1}{c_n} \sum_{w \in \mathcal{W}_{O,n}} \omega(w)^2.$$

For each w, a direct calculation gives

$$\frac{m_-(w)(m_-(w)-1)}{(m_+(w)+1)(m_+(w)+2)} - \omega(w)^2 = -\frac{m_-(w)(m_-(w)+m_+(w)+1)}{(m_+(w)+1)^2(m_+(w)+2)}.$$

It follows that

$$R_n R_{n+h} - R_n^2 \ge -\frac{1}{c_n} \sum_{w \in \mathcal{W}_{O,n}} \frac{m_-(w)(m_-(w)+m_+(w)+1)}{(m_+(w)+1)^2(m_+(w)+2)} - C_1 \vartheta_1^n.$$

For all sufficiently large n, $\lfloor \alpha n \rfloor \ge \alpha n/2$. Hence, for a walk satisfying $m_+(w) \ge \lfloor \alpha n \rfloor$, the fraction inside the sum is at most

$$\frac{n(2n+1)}{\lfloor \alpha n \rfloor^3} \le \frac{24}{\alpha^3 n}.$$

For every walk, the same fraction is at most $3n^2$. By (31), walks with $m_+(w) < \lfloor \alpha n \rfloor$ form a proportion at most ϑ^n. Consequently,

$$\frac{1}{c_n} \sum_{w \in \mathcal{W}_{O,n}} \frac{m_-(w)(m_-(w)+m_+(w)+1)}{(m_+(w)+1)^2(m_+(w)+2)} \le \frac{24}{\alpha^3 n} + 3n^2\vartheta^n.$$

Since $n^2\vartheta^n = O(1/n)$ and $\vartheta_1^n = O(1/n)$, the preceding estimates give a constant $C_2 < \infty$ such that

$$R_n R_{n+h} - R_n^2 \geq -\frac{C_2}{n}$$

for all sufficiently large n. Since the left-hand side equals $R_n(R_{n+h} - R_n)$ and $R_n \geq b$, division by R_n proves the second inequality in (34), with $B = C_2/b$. □

The following lemma converts the estimates in Proposition 6.5 into convergence of the ratios.

Lemma 6.6. *Let $\mathsf{a}_n > 0$, and suppose that*

$$\lim_{n\to\infty} \mathsf{a}_n^{1/n} = \lambda > 0.$$

Fix $h \geq 1$, and put $\mathcal{R}_n = \mathsf{a}_{n+h}/\mathsf{a}_n$. If there are $b > 0$, $K < \infty$, and n_0 such that

$$\mathcal{R}_n \geq b, \qquad \mathcal{R}_{n+h} - \mathcal{R}_n \geq -\frac{K}{n}, \qquad n \geq n_0,$$

then

$$\lim_{n\to\infty} \mathcal{R}_n = \lambda^h.$$

Proof. Fix $s \in \{0, \dots, h-1\}$, and set

$$\mathsf{a}_k^{(s)} = \mathsf{a}_{s+kh}, \qquad \varrho_k = \frac{\mathsf{a}_{k+1}^{(s)}}{\mathsf{a}_k^{(s)}}.$$

Put $L = \lambda^h$, $C = K/h$, and $\kappa = C/b$. Then

$$\left(\mathsf{a}_k^{(s)}\right)^{1/k} \longrightarrow L, \qquad \varrho_k = \mathcal{R}_{s+kh} \geq b, \qquad \varrho_{k+1} - \varrho_k \geq -\frac{C}{k}$$

for all sufficiently large k; since s is arbitrary, it is enough to prove that $\varrho_k \to L$ for this fixed s.

For fixed $\mathfrak{m} \geq 1$, set

$$Q_{\mathfrak{m},q} = \left(\frac{\mathsf{a}_{(\mathfrak{m}+1)q+1}^{(s)}}{\mathsf{a}_{\mathfrak{m}q}^{(s)}} \right)^{1/(q+1)}.$$

The root limit gives $Q_{\mathfrak{m},q} \to L$, since $\log \mathsf{a}_k^{(s)} = k \log L + o(k)$ and the two indices in the quotient differ by $q+1$. For sufficiently large q and $0 \leq k \leq q$,

$$\begin{aligned} \varrho_{\mathfrak{m}q+k} &\leq \varrho_{(\mathfrak{m}+1)q} + C \sum_{\ell=\mathfrak{m}q+k}^{(\mathfrak{m}+1)q-1} \frac{1}{\ell} \\ &\leq \varrho_{(\mathfrak{m}+1)q} \left(1 + \frac{\kappa}{\mathfrak{m}}\right). \end{aligned}$$

Multiplying over $k = 0, \dots, q$ yields

$$\liminf_{q\to\infty} \varrho_{(\mathfrak{m}+1)q} \geq \frac{L}{1 + \kappa/\mathfrak{m}}.$$

If $j = (\mathfrak{m}+1)q \leq k < j + \mathfrak{m} + 1$, then

$$\varrho_k \geq \varrho_j - \frac{C\mathfrak{m}}{j}.$$

Hence the same lower bound holds for $\liminf_{k\to\infty} \varrho_k$. Letting $\mathfrak{m} \to \infty$ gives

$$\liminf_{k\to\infty} \varrho_k \geq L.$$

For the reverse inequality, fix an integer $\mathfrak{m} > \kappa$. For sufficiently large q and $0 \leq k \leq q$,

$$\begin{aligned} \varrho_{\mathfrak{m}q+k} &\geq \varrho_{\mathfrak{m}q} - C \sum_{\ell=\mathfrak{m}q}^{\mathfrak{m}q+k-1} \frac{1}{\ell} \\ &\geq \varrho_{\mathfrak{m}q} \left(1 - \frac{\kappa}{\mathfrak{m}}\right). \end{aligned}$$

Multiplication and the limit of $Q_{\mathfrak{m},q}$ give

$$\limsup_{q\to\infty} \varrho_{\mathfrak{m}q} \leq \frac{L}{1 - \kappa/\mathfrak{m}}.$$

For any large k, let j be the smallest multiple of $\mathfrak{m}$ with $j \geq k$. Then

$$\varrho_k \leq \varrho_j + \frac{C\mathfrak{m}}{k}.$$

Thus the same upper bound holds for $\limsup_{k\to\infty} \varrho_k$. Letting $\mathfrak{m} \to \infty$ gives

$$\limsup_{k\to\infty} \varrho_k \leq L.$$

Together with the lower bound, this proves $\varrho_k \to L$ and completes the proof. □

Proof of Theorem 4(i). Theorem 1 gives $\lim_n c_n^{1/n} = \mu$. Proposition 6.5, through (34), verifies the two hypotheses of Lemma 6.6 for $\mathsf{a}_n = c_n$, which gives (6). □

6.3 The Sierpiński gasket

The flexibility condition can be checked directly on the Sierpiński gasket.

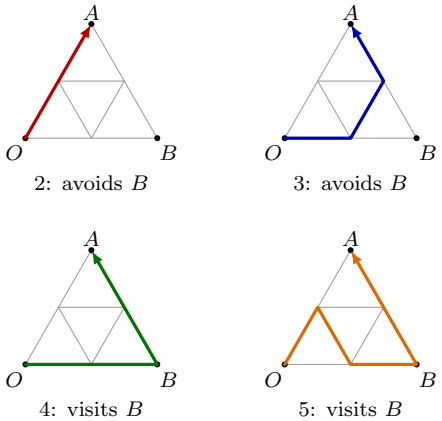


Figure 7: Representative pairs of boundary forests in F_1 of the Sierpiński gasket. The two forests in each pair have the same boundary state and lengths differing by one. Their symmetry images cover all six states in $\mathcal{S}_1$.

Proposition 6.7. *The Sierpiński gasket graph is* 1*-flexible at scale* 1.

Proof. We work in F_1. There are six states in $\mathcal{S}_1$. Every boundary forest in F_1 has one component: two nonempty components would require at least four degree-one vertices in ∂F_1, whereas $\#\partial F_1 = 3$. Thus a state either joins two boundary vertices without using the third, or visits all three boundary vertices with one of them in the middle. The first kind has realizations in F_1 of lengths 2 and 3, and the second has realizations in F_1 of lengths 4 and 5. Figure 7 illustrates one pair for each of these two kinds. The symmetries of F_1 transport these pairs to every choice of endpoints and middle vertex. Hence every state in $\mathcal{S}_1$ has two representatives in F_1 whose lengths differ by one. Therefore, the Sierpiński gasket graph is 1-flexible at scale 1. □

Combining Proposition 6.7 with Theorem 4(i) gives $\lim_{n\to\infty} c_{n+1}/c_n = \mu$ for the Sierpiński gasket.

6.4 Flexibility of polygonal N-gaskets

We now determine the smallest flexibility step for every polygonal N-gasket. By self-similarity, Proposition 5.2 applies within every parent cell. Thus each immediate subcell has an outer vertex on the boundary of its parent and two contacts with its neighboring subcells. A nonempty restriction of a boundary forest to this subcell is a union of paths whose endpoints belong to these three vertices. It has only one component, since two path components would require four endpoints. Its local behavior therefore has six possibilities: it either joins two of the three vertices and avoids the third, or visits all three with one of them in the middle.

Keep the notation of Definition 5.1, and abbreviate

$$k = k_N, \qquad \ell = \ell_N, \qquad s = s_N = k + 1, \qquad g = N - 2s = 2k + \ell - 2.$$

Set

$$d = 2s - g = 4 - \ell, \qquad h_N = \gcd(g, d).$$

Inside $F_1^{(N)}$, let

$$P_N = \{a_{-s,1}, a_{0,1}, a_{s,1}\}.$$

For a path in $F_1^{(N)}$ with endpoints in P_N, record only the vertices of P_N in the order in which the path visits them. A sequence and its reversal are regarded as the same *P_N-pattern.* Thus there are six patterns: three choices of two visited vertices, and three choices of the middle vertex when all three are visited.

Lemma 6.8. *For each of the six P_N-patterns, there are two self-avoiding paths in $F_1^{(N)}$ which have that pattern and whose lengths differ by h_N.*

Proof. By the graph construction and Proposition 5.2(ii)–(iii), $F_1^{(N)}$ is the union of N copies of $F_0^{(N)}$, arranged in a cycle. For this proof, write $a_i := a_{i,1}$, and let ζ_i be the contact between the i-th and $(i+1)$-st copies; all indices are understood modulo N. In the i-th copy, the arcs from a_i to either contact, avoiding the other contact, have length s. The arc between the two contacts, avoiding a_i, has length g. We may therefore represent this copy by the weighted triangle

$$a_i\zeta_{i-1}\zeta_i, \qquad i \in \mathbb{Z}/N\mathbb{Z},$$

whose three edge weights are s, s, g. Each weighted edge represents the corresponding arc in the original copy of $F_0^{(N)}$.

Inside one weighted triangle there are three useful local changes:

$$\begin{aligned} \zeta_{i-1}\zeta_i &\longleftrightarrow \zeta_{i-1}a_i\zeta_i, \\ a_i\zeta_i &\longleftrightarrow a_i\zeta_{i-1}\zeta_i, \qquad a_i\zeta_{i-1} \longleftrightarrow a_i\zeta_i\zeta_{i-1}. \end{aligned}$$

The first change has length difference $2s - g = d$. Each of the two changes in the second line has length difference g; they are the two symmetric forms of the same operation. We call these operations a *d-switch* and a *g-switch*, respectively.

Now $P_N = \{a_{-s}, a_0, a_s\}$. The three copies carrying these vertices divide the cyclic chain into three sections of lengths s, s, g, measured by the number of steps from one copy to the next. The interiors of the three sections are disjoint, and neighboring sections share an endpoint copy.

Fix a P_N-pattern and choose one of its two orientations. Between each consecutive pair in this ordered list, follow the section that avoids the remaining vertex of P_N. In the first and last copies of the section, use the length-s arc leading from the prescribed outer vertex to the contact belonging to that section. In every copy strictly between them, use the length-g arc between the two contacts. If the pattern visits all three vertices, the two chosen sections have disjoint intermediate copies and share only the copy carrying the middle vertex. Inside that copy, the incoming and outgoing length-s arcs meet only at the middle vertex, so the two paths can be concatenated. In either case, the result is a simple path γ with the prescribed pattern.

We shall use at most one g-switch, at one endpoint of γ. If that endpoint is a_i, the path uses one contact of the i-th copy. The switch first goes from a_i to the other contact and then to the contact already used by γ. The other contact lies on the unused section of the chain, so the new path is still simple and has the same P_N-pattern.

A d-switch may be made once in any intermediate copy of the section or sections used by γ. In the i-th such copy, it replaces the length-g arc between the two contacts by the two length-s arcs through a_i. Since this copy is not an endpoint copy, $a_i \notin P_N$, so the pattern is unchanged. Any set of distinct intermediate copies may be switched simultaneously: their interiors are disjoint, and every switch keeps the two contacts through which γ enters and leaves its copy. Thus one g-switch adds g to the length, while r d-switches add rd.

It remains to choose the switches so that the length difference is exactly h_N. If $\ell = 3$ and $N > 3$, then $d = h_N = 1$, and every path γ constructed above contains at least one intermediate copy, so one d-switch suffices. If $N = 3$, then $g = h_N = 1$, so one g-switch suffices. If $\ell = 2$, then $g = 2k$ and $d = h_N = 2$; again one d-switch suffices.

Suppose finally that $\ell = 1$. Then $g = 2k - 1$ and $d = 3$. Thus $h_N = 3$ when $k \equiv 2 \pmod 3$, and $h_N = 1$ otherwise. If $k = 3j$, compare a path with one g-switch to a path with $2j$ distinct d-switches; their added lengths differ by

$$2j \cdot 3 - (2k - 1) = 6j - (6j - 1) = 1.$$

If $k = 3j + 1$, the same two choices have length difference

$$(2k - 1) - 2j \cdot 3 = (6j + 1) - 6j = 1.$$

If $k = 3j + 2$, then both $g = 6j + 3$ and $d = 3$ are divisible by 3, and one d-switch gives a difference of 3.

The required d-switches are always available. The path γ either uses a section of length s, which has $s - 1 = k$ intermediate copies, or uses only the section of length g, which has $g - 1$ intermediate copies. For $k = 3j$ or $k = 3j + 1$, both $2j \le k$ and $2j \le g - 1$. For $k = 3j + 2$, only one d-switch is required, and both k and $g - 1$ are positive. Thus the construction applies to all six patterns. □

Proof of Theorem 4(ii). We first prove that the polygonal N-gasket is h_N-flexible at scale 2. Fix $\sigma \in \mathcal{S}_2$ and $\Gamma \in \mathcal{C}_2(\sigma)$. Choose a copy Δ of $F_1^{(N)}$ in $F_2^{(N)}$ which contains an edge of Γ. The copy Δ has one outer vertex on $\partial F_2^{(N)}$ and two contacts with its neighbors. Every endpoint of $\Gamma \cap \Delta$ is one of these three vertices. It cannot have two components, because that would require four distinct endpoints. Thus $\Gamma \cap \Delta$ is a single path. Self-similarity and a dihedral relabeling identify Δ with $F_1^{(N)}$ and these three vertices with P_N. The path $\Gamma \cap \Delta$ consequently has one of the six patterns in Lemma 6.8.

Replace this path, in turn, by the two paths supplied by that lemma, and leave Γ unchanged outside Δ. Because the two replacements visit the same vertices of P_N in the same order, both resulting graphs are boundary forests with state σ. Their edge counts differ by h_N. Since σ was arbitrary, the graph is h_N-flexible at scale 2.

We next show that no smaller step is possible. If $\ell = 2$, then $N = 4k + 2$. The cycle $F_0^{(N)}$ is bipartite, and its two contacts q_{-s} and q_s have the same color because their cyclic distance $2s$ is even. The copies can therefore be colored consistently at every scale, so each $F_r^{(N)}$ is bipartite. In a bipartite graph, the parity of the length of a path is determined by the colors of its endpoints. A boundary state fixes the endpoints of every component and their pairing, so the edge counts of all forests with that state have the same parity. Every flexibility step is therefore even. Since

$$h_N = \gcd(2k, 2) = 2,$$

the step constructed above is minimal.

Now suppose that $\ell = 1$ and $k \equiv 2 \pmod 3$, equivalently $N \equiv 9 \pmod{12}$. Then

$$3 \mid s = k + 1, \qquad 3 \mid g = 2k - 1.$$

Decompose a boundary forest into its pieces inside the copies of $F_0^{(N)}$. Each piece is a concatenation of arcs of lengths s, s, g, and hence has length divisible by 3. Since different copies have disjoint edge sets, the whole forest also has length divisible by 3. Thus every flexibility step is divisible by 3. Here $h_N = \gcd(g, 3) = 3$, so the constructed step is minimal.

In all remaining cases $h_N = 1$, which is automatically the smallest positive step. This proves the stated classification and minimality. Finally, the ratio limit follows from Theorem 4(i). □

6.5 The Vicsek counterexample

The Vicsek graph is not h-flexible for any $h \geq 1$. At every scale, all self-avoiding paths between a fixed pair of diagonally opposite boundary vertices have the same length. By Theorem 1, its connective constant exists, but its successive ratios have two different subsequential limits. Let

$$p_0 = (0, 0), \quad p_1 = (1, 0), \quad p_2 = (1, 1), \quad p_3 = (0, 1), \quad p_4 = (1/2, 1/2),$$

and define

$$\Psi_i(x) = p_i + \frac{1}{3}(x - p_i), \qquad 0 \leq i \leq 4.$$

The attractor of this iterated function system is the standard Vicsek nested fractal. Its essential fixed points are p_0, p_1, p_2, p_3, so F_0 is the four-cycle $p_0p_1p_2p_3p_0$, and we take $O = p_0$. Since Ψ_0 fixes O, the graphs F_m are defined by (3). The four corner cells meet the central cell at four distinct vertices and do not meet one another, so the one-point intersection condition (2) holds.

For a polynomial $G(z)$, write $[z^n]G(z)$ for the coefficient of z^n in $G(z)$.

Proposition 6.9. *Let*

$$H_m(z) = \sum_{\substack{w \text{ is a self-avoiding walk in } F_m \\ w(0)=O}} z^{L(w)}.$$

The number of n-step self-avoiding walks in F_m starting at O is $[z^n]H_m(z)$. Let $P_m(z)$ be the sum of $z^{L(w)}$ over all self-avoiding paths joining a fixed pair of diagonally opposite boundary vertices, and define $Q_m(z)$ in the same way for a fixed adjacent pair. Dihedral symmetry makes these polynomials independent of the chosen pair. They satisfy

$$\begin{aligned} P_{m+1} &= P_m^3, \\ Q_{m+1} &= P_m^2 Q_m, \\ H_{m+1} &= H_m + P_m(1 + 2Q_m + P_m)(H_m - 1), \end{aligned} \tag{36}$$

with

$$H_0 = 1 + 2z + 2z^2 + 2z^3, \qquad P_0 = 2z^2, \qquad Q_0 = z + z^3.$$

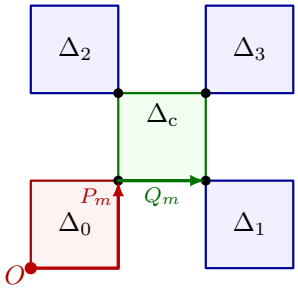


Figure 8: The five-copy decomposition of the Vicsek graph F_{m+1}. Each Δ_j is a copy of F_m; Δ_0 is the rooted copy and Δ_c is the central copy. A walk which leaves Δ_0 can enter at most one further corner copy.

Proof. The graph F_{m+1} consists of five copies of F_m: four corner copies and one central copy. Each corner copy meets the central copy at one vertex, and no two corner copies meet. Figure 8 illustrates this decomposition.

A path joining diagonally opposite boundary vertices crosses three copies in order: the initial corner, the central copy, and the terminal corner. Within each copy it again joins diagonally opposite boundary vertices, so $P_{m+1} = P_m^3$. For adjacent boundary vertices, the path joins diagonally opposite boundary vertices in each of the two corner copies and adjacent boundary vertices in the central copy. Hence $Q_{m+1} = P_m^2 Q_m$.

Now consider a walk counted by H_{m+1}. If it stays in the corner copy containing O, its contribution is H_m. Otherwise it first crosses that copy to the central copy, contributing P_m. From the contact vertex it either makes a nonempty walk inside the central copy and stops there, or crosses the central copy into one of the other three corner copies and makes a nonempty terminal walk in that copy. The three exits consist of two adjacent crossings and one diagonally opposite crossing. Therefore the contribution of walks leaving the rooted corner is

$$P_m\big((H_m - 1) + (2Q_m + P_m)(H_m - 1)\big).$$

Adding the walks that stay in the rooted corner gives the third recursion in (36). The initial polynomials follow by listing the self-avoiding paths in F_0. □

Proof of Theorem 4(iii). The first two recursions in Proposition 6.9 give

$$P_m(z) = 2^{3^m} z^{2\cdot 3^m}, \qquad Q_m(z) = 2^{3^m-1}\left(z^{2\cdot 3^m-1} + z^{2\cdot 3^m+1}\right).$$

Fix $r \geq 1$, choose a path counted by P_r, and let σ be its boundary state. Every forest in $\mathcal{C}_r(\sigma)$ consists of one self-avoiding path joining the same pair of diagonally opposite boundary vertices, and is therefore counted by P_r. The formula for P_r shows that every such forest has exactly $2 \cdot 3^r$ edges. Thus $\mathcal{C}_r(\sigma)$ contains no two forests whose numbers of edges differ by a positive integer. Since r was arbitrary, the Vicsek graph is not h-flexible at any finite scale for any $h \geq 1$.

The third recursion in (36), together with $\deg H_0 = 3$, gives by induction

$$\deg H_m = 2 \cdot 3^m + m + 1.$$

Moreover, every term in $H_{m+1} - H_m$ has degree at least $2 \cdot 3^m + 1$. Hence $[z^n]H_m$ no longer changes once $2 \cdot 3^m \geq n$, and its stable value is c_n.

At level 1, (36) gives

$$H_1 = 1 + 2z + 2z^2 + 6z^3 + 12z^4 + 20z^5 + 24z^6 + 16z^7 + 8z^8.$$

Moreover,

$$P_1 = 8z^6, \qquad Q_1 = 4z^5 + 4z^7.$$

The term $P_1(2Q_1 + P_1)(H_1 - 1)$ starts in degree 12. Therefore, in degrees $7, 8, 9$, only $H_1 + P_1(H_1 - 1)$ contributes to H_2. Using $P_1 = 8z^6$, we obtain

$$[z^7]H_2 = 16 + 8 \cdot 2 = 32, \qquad [z^8]H_2 = 8 + 8 \cdot 2 = 24, \qquad [z^9]H_2 = 0 + 8 \cdot 6 = 48.$$

Now let $k \geq 3$ and $j \in \{-2, -1, 0\}$. Since $\deg H_{k-1} < 3^k - 2$, $[z^{3^k+j}]H_{k-1} = 0$. Moreover, every term in

$$P_{k-1}(2Q_{k-1} + P_{k-1})(H_{k-1} - 1)$$

has degree at least $4 \cdot 3^{k-1} > 3^k$. Thus only $P_{k-1}(H_{k-1} - 1)$ contributes to the coefficient of z^{3^k+j} in H_k, and

$$[z^{3^k+j}]H_k = 2^{3^{k-1}}[z^{3^{k-1}+j}]H_{k-1}.$$

Starting with the three coefficients of H_2 above and applying this identity repeatedly proves (7).

The first formula in (7) gives

$$\mu = \lim_{k\to\infty} c_{3^k-2}^{1/(3^k-2)} = \sqrt{2}.$$

The other two formulas in (7) give

$$\frac{c_{3^k-1}}{c_{3^k-2}} = \frac{3}{4}, \qquad \frac{c_{3^k}}{c_{3^k-1}} = 2.$$

Since these two subsequential limits are different, c_{n+1}/c_n does not converge. □

Appendix A Dynamics of the polygonal crossing recursions

The two-variable recursion for the Sierpiński gasket and its critical behavior were introduced by Hattori, Hattori and Kusuoka [11]. The same renormalization viewpoint motivates the analysis below, which combines coordinatewise monotonicity with compactness of the critical orbit.

Theorem 5. *Let $N \geq 3$ and $4 \nmid N$. Every entry of $\mathbf{C}_m^{(N)}(\beta)$ tends to zero when $\beta > \beta_{c,N}$, and every entry tends to infinity when $\beta < \beta_{c,N}$. At $\beta = \beta_{c,N}$, the crossing vector has the following limits. Put $\tau = (\sqrt{5} - 1)/2$.*

(i) *If $N = 3$, then*

$$(x_m, y_m) \longrightarrow (\tau, 0).$$

(ii) *If $N = 5$, then*

$$(\phi_{1,m}, \phi_{2,m}, \theta_{1,m}, \theta_{2,m}) \longrightarrow (\phi_*, 0, \theta_*, 0),$$

where

$$\theta_*^5 + 2\theta_*^4 = 1, \qquad \phi_* = \frac{1}{\theta_*(1 + \theta_*)}.$$

(iii) *If $N \in \{6, 9\}$, then*

$$(X_m, Y_m) \longrightarrow (t_*, t_*), \qquad t_* = \tau^{1/d_N}, \qquad d_N = \frac{N}{3}.$$

(iv) *If* $N = 4k + \ell \notin \{3, 5, 6, 9\}$, *where* $\ell \in \{1, 2, 3\}$, *then*

$$(X_m, Y_m) \longrightarrow (X_{*,N}, Y_{*,N}),$$

where $Y_{*,N} \in (0, 1)$ *is the unique solution of*

$$Y_{*,N}^{N} + 2Y_{*,N}^{2k+2} = 1$$

and

$$X_{*,N} = \left(Y_{*,N}^{k} + Y_{*,N}^{3k+\ell-2}\right)^{-1}.$$

We prove Theorem 5 by treating the generic two-variable recursion first. The recursions for $N = 3$ and $N = 5$ require separate arguments; the cases $N = 6$ and $N = 9$ then follow from the scalar recursion used in the proof of Theorem 3.

Proposition A.1. *Let* $N = 4k + \ell \notin \{3, 5, 6, 9\}$, *where* $\ell \in \{1, 2, 3\}$. *There is a unique value* $\beta_{*,N} > 0$ *such that*

$$\begin{aligned} \beta > \beta_{*,N}: &\quad (X_m, Y_m) \longrightarrow (0, 0), \\ \beta < \beta_{*,N}: &\quad (X_m, Y_m) \longrightarrow (+\infty, +\infty). \end{aligned}$$

At $\beta = \beta_{*,N}$,

$$(X_m, Y_m) \longrightarrow (X_{*,N}, Y_{*,N}),$$

where $Y_{*,N} \in (0, 1)$ *is the unique solution of*

$$Y_{*,N}^{N} + 2Y_{*,N}^{2k+2} = 1$$

and

$$X_{*,N} = \left(Y_{*,N}^{k} + Y_{*,N}^{3k+\ell-2}\right)^{-1}.$$

Consequently, $\beta_{*,N} = \beta_{\mathrm{c},N}^{\mathrm{cr}} = \beta_{\mathrm{c},N}$.

Proof. Define

$$A(y) = y^k + y^{3k+\ell-2}, \qquad B(y) = y^{2k+\ell-3} + y^{2k+1},$$

so that the recursion can be written as

$$\mathcal{R}(x, y) = \left(x^2 A(y), x^2 B(y)\right).$$

By Proposition 5.3, $(X_{m+1}, Y_{m+1}) = \mathcal{R}(X_m, Y_m)$. Set

$$\alpha = k + 1, \qquad \delta = k + \ell - 3.$$

In the present case $k \geq 1$, $\delta \geq 1$, and $\alpha > \delta$. For every $x, y > 0$,

$$\frac{\mathcal{R}_2(x, y)}{\mathcal{R}_1(x, y)} = f(y) := \frac{y^\delta + y^\alpha}{1 + y^{\alpha+\delta}}. \tag{37}$$

Since

$$1 - f(y) = \frac{(1 - y^\delta)(1 - y^\alpha)}{1 + y^{\alpha+\delta}},$$

we have $0 < f(y) \leq 1$ for every $y > 0$, with equality only at $y = 1$. The ratio in (37) therefore gives

$$Y_m \leq X_m, \qquad m \geq 1. \tag{38}$$

Moreover,

$$f'(y) = \frac{\delta y^{\delta-1}(1 - y^{2\alpha}) + \alpha y^{\alpha-1}(1 - y^{2\delta})}{(1 + y^{\alpha+\delta})^2} > 0, \qquad 0 < y < 1.$$

Thus f is strictly increasing on $(0,1)$.

We first determine the positive fixed points of $\mathcal{R}$. The fixed-point equations are

$$x = A(y)^{-1}, \qquad yA(y)^2 = B(y).$$

Substituting

$$A(y) = y^k(1+y^{\alpha+\delta}), \qquad B(y) = y^{k+\delta}(1+y^{\alpha-\delta}),$$

into the second equation gives

$$y^N + 2y^{2k+2} = 1. \tag{39}$$

The left-hand side of (39) is strictly increasing from 0 to 3 on $(0,1)$. Hence $(X_{*,N}, Y_{*,N})$ is the unique positive fixed point.

We next locate the parameter at which the orbit changes from divergence to decay. Let

$$\mathcal{E} = \{\beta \geq 0 : Y_m(\beta) > 1 \text{ for some } m \geq 1\}.$$

If $Y_m > 1$, then (38) gives $X_m \geq Y_m > 1$. The recursion then gives $X_{m+1} > X_m^2$ and $Y_{m+1} > Y_m^{2k+\ell-1}$, so both coordinates tend to infinity.

The map $\mathcal{R}$ is strictly increasing in each variable, while both initial values are strictly decreasing functions of β. Consequently, if $\beta \in \mathcal{E}$, then every smaller nonnegative parameter also belongs to $\mathcal{E}$. Thus $\mathcal{E}$ is downward closed in $[0,\infty)$. At $\beta = 0$, the orbit enters the region $X_m, Y_m > 1$, so $0 \in \mathcal{E}$. Put $r = k+2 > 1$. For $z \in [0,\infty)^2$ with $\|z\|_\infty \leq 1$,

$$\|\mathcal{R}(z)\|_\infty \leq 2\|z\|_\infty^r.$$

Hence a sufficiently small square about the origin is invariant and every orbit starting there tends to the origin. Since the initial point tends to the origin as $\beta \to \infty$, sufficiently large β does not belong to $\mathcal{E}$. Finally, every finite iterate depends continuously on β, and hence $\mathcal{E}$ is open. There is therefore a unique $\beta_{*,N} > 0$ such that

$$\mathcal{E} = [0, \beta_{*,N}). \tag{40}$$

We now study the orbit at the endpoint $\beta_{*,N}$. By (40), it cannot have $Y_m > 1$. It cannot have $Y_m = 1$ either: in that case (38) gives $X_m \geq 1$, while (26) gives $Y_{m+1} = 2X_m^2 \geq 2$. Hence

$$0 < Y_m < 1, \qquad m \geq 1.$$

To bound the other coordinate, put

$$a_m = \log X_m, \qquad b_m = -\log Y_m, \qquad c = 2k+\ell-3.$$

For $m \geq 1$, the recursion gives

$$\begin{aligned} a_{m+1} &= 2a_m - kb_m + \varepsilon_m, \\ b_{m+1} &= -2a_m + cb_m - \eta_m, \end{aligned}$$

where the error terms are

$$\varepsilon_m = \log\bigl(1 + Y_m^{2k+\ell-2}\bigr), \qquad \eta_m = \log\bigl(1 + Y_m^{4-\ell}\bigr).$$

Since $0 < Y_m < 1$ and both exponents are positive,

$$0 \leq \varepsilon_m, \eta_m \leq \log 2.$$

The polynomial

$$P(t) = t^2 - (c+2)t + 2\delta$$

satisfies $P(0) > 0$ and $P(1) = \ell - 4 < 0$. Let $\lambda \in (0,1)$ be its smaller zero. Then $(2, 2-\lambda)$ is a positive left eigenvector of

$$\begin{pmatrix} 2 & -k \\ -2 & c \end{pmatrix}$$

with eigenvalue λ. Hence

$$W_m = 2a_m + (2-\lambda)b_m$$

satisfies

$$\begin{aligned} W_{m+1} &= \lambda W_m + 2\varepsilon_m - (2-\lambda)\eta_m \\ &\le \lambda W_m + 2\log 2. \end{aligned}$$

Iterating this scalar inequality shows that W_m is bounded above. Since $b_m \ge 0$ and $2 - \lambda > 0$, we have $2a_m \le W_m$. Thus a_m, and therefore X_m, is bounded above.

The critical orbit cannot enter the small invariant square constructed above. Otherwise, continuity of a finite iterate would put an orbit with $\beta < \beta_{*,N}$ in the same square, contradicting (40). Suppose that $Y_m \to 0$ along a subsequence. Since X_m is bounded, both coordinates of the next iterate tend to zero, which is impossible by the preceding paragraph. If $X_m \to 0$ along a subsequence, then $Y_m \le X_m$ by (38), and the same contradiction follows. Finally, suppose that $Y_{m_j} \to 1$. Then $X_{m_j} \ge Y_{m_j}$, and therefore

$$Y_{m_j+1} = X_{m_j}^2 B(Y_{m_j}) \ge Y_{m_j}^2 B(Y_{m_j}) \longrightarrow 2,$$

contrary to $Y_{m_j+1} < 1$. The critical orbit consequently stays in a compact subset of $(0,\infty)\times(0,1)$.

Let L be its *omega-limit set*, that is, the set of limits of convergent subsequences of the critical orbit. Since the orbit remains in a compact set and $\mathcal{R}$ is continuous, $\mathcal{R}(L) \subseteq L$. Conversely, if $z_{m_j} \to q \in L$, compactness gives a convergent subsequence of $(z_{m_j-1})_j$; its limit $p \in L$ satisfies $\mathcal{R}(p) = q$. Hence $\mathcal{R}(L) = L$.

Every monomial in $\mathcal{R}$ has total degree at least $r = k + 2 > 1$, and hence

$$\mathcal{R}(tz) \ge t^r \mathcal{R}(z), \qquad z \in (0,\infty)^2, \quad t \ge 1. \tag{41}$$

No two distinct points of L are comparable in the coordinatewise order. Indeed, if distinct $p, q \in L$ satisfy $p \le q$, strict monotonicity and $\mathcal{R}(L) = L$ allow us, after replacing them by their images, to choose $t > 1$ such that $q \ge tp$. Monotonicity and (41) then give by induction

$$\mathcal{R}^m(q) \ge t^{r^m} \mathcal{R}^m(p), \qquad m \ge 0.$$

This is impossible because the compact set $L \subset (0,\infty)\times(0,1)$ is bounded away from the coordinate axes.

The function f in (37) is strictly increasing on $(0,1)$. It follows that $\mathcal{R}$ is injective on $(0,\infty)\times(0,1)$: equality of two images first gives equality of their second input coordinates from the ratio in (37), and then equality of their first input coordinates. Order the points of L by their first coordinate. If $p_1 < q_1$, the preceding incomparability gives $p_2 > q_2$. If one also had $\mathcal{R}_1(p) \ge \mathcal{R}_1(q)$, then the strict increase of f, together with (37), would imply $\mathcal{R}_2(p) > \mathcal{R}_2(q)$. The two images would then be comparable, a contradiction. Hence $\mathcal{R}_1(p) < \mathcal{R}_1(q)$: the restriction of $\mathcal{R}$ to L preserves the order of the first coordinates.

Since $\mathcal{R}(L) = L$, the restriction of $\mathcal{R}$ to L is an order-preserving bijection. It therefore fixes the points of L with the smallest and largest first coordinates. Both are positive fixed points of $\mathcal{R}$, so the uniqueness proved in (39) forces them to coincide. Thus

$$L = \{(X_{*,N}, Y_{*,N})\},$$

and the critical orbit converges to this point.

If $\beta > \beta_{*,N}$, choose $0 < t < 1$ such that the initial point at β is bounded above by t times the critical initial point. Since every monomial has degree at least r, the counterpart of (41) for $0 < t < 1$, followed by induction, gives

$$(X_m(\beta), Y_m(\beta)) \le t^{r^m}(X_m(\beta_{*,N}), Y_m(\beta_{*,N})),$$

so the orbit tends to the origin. If $\beta < \beta_{*,N}$, (40) gives $Y_m > 1$ for some m. Then (38) and (26) imply that both coordinates tend to infinity. Since $\mathbf{C}_m^{(N)}(\beta) = (X_m(\beta), Y_m(\beta))$ in the present case, we have proved

$$\left\{\beta \ge 0 : \sup_{m\ge 0} \|\mathbf{C}_m^{(N)}(\beta)\|_\infty < \infty\right\} = [\beta_{*,N}, \infty).$$

The definition (25) and Proposition 5.4 therefore give $\beta_{*,N} = \beta_{\mathrm{c},N}^{\mathrm{cr}} = \beta_{\mathrm{c},N}$. □

For completeness, we also treat the case of the Sierpiński gasket studied in [11].

Proposition A.2. *For $N = 3$, the recursion (27) has a unique critical value $\beta_{*,3} > 0$. Its orbit tends to $(0,0)$ for $\beta > \beta_{*,3}$, both coordinates tend to infinity for $\beta < \beta_{*,3}$, and*

$$(x_m, y_m) \longrightarrow (\tau, 0), \qquad \tau = \frac{\sqrt{5}-1}{2},$$

at $\beta = \beta_{*,3}$. Consequently, $\beta_{*,3} = \beta_{\mathrm{c},3}^{\mathrm{cr}} = \beta_{\mathrm{c},3}$.

Proof. Put $r_m = y_m/x_m$. The recursion gives

$$r_{m+1} = r_m \frac{x_m(1+2r_m)}{(1+r_m)^2 + x_m(1+2r_m)}. \tag{42}$$

Thus $0 < r_{m+1} < r_m$. If the orbit is bounded, then x_m is bounded and $0 < r_m \le r_0 \le 1$. The fraction in (42) is consequently bounded above by a constant strictly smaller than one. Hence $r_m \to 0$.

If $x_m > 1$, then

$$x_{m+1} \ge x_m^2(1 + x_m) > x_m,$$

and $x_m \to \infty$. Since

$$\frac{y_{m+1}}{y_m} = x_m(x_m + 2y_m),$$

the second coordinate then tends to infinity as well.

Define

$$\mathcal{E}_3 = \{\beta \ge 0 : x_m(\beta) > 1 \text{ for some } m \ge 0\}.$$

The recursion is strictly increasing in both variables, while $(x_0, y_0) = (e^{-\beta}, e^{-2\beta})$ is strictly decreasing in β. Consequently, if $\beta \in \mathcal{E}_3$, then every smaller nonnegative parameter also belongs to $\mathcal{E}_3$; thus $\mathcal{E}_3$ is downward closed in $[0,\infty)$. At $\beta = 0$, one has $x_0 = y_0 = 1$ and $x_1 = 7 > 1$, so $0 \in \mathcal{E}_3$. For all sufficiently large β, the initial point lies in a small invariant neighborhood of the origin whose orbits tend to the origin, so such a parameter does not belong to $\mathcal{E}_3$. Since each finite iterate depends continuously on β, the set $\mathcal{E}_3$ is open. It follows that $\mathcal{E}_3 = [0, \beta_{*,3})$ for a unique $\beta_{*,3} > 0$.

At $\beta = \beta_{*,3}$, one has $x_m < 1$ for every m, since equality would imply $x_{m+1} > 1$. Also $y_m = r_m x_m \le x_m$, so the critical orbit is bounded. It cannot converge to the origin. Otherwise, a finite critical iterate would lie in a small invariant neighborhood of the origin; continuity would then put an orbit with $\beta < \beta_{*,3}$ in the same neighborhood, contrary to the definition of $\mathcal{E}_3$.

Since $r_m \to 0$, the first recursion may be written as

$$x_{m+1} = x_m^2\big((1+r_m)^2 + x_m(1+2r_m)\big).$$

The limiting map is $g(x) = x^2(1+x)$. Its only nonnegative fixed points are 0 and τ; moreover,

$$g(x) < x \quad (0 < x < \tau), \qquad g(x) > x \quad (x > \tau).$$

The critical orbit cannot have $x_m > \tau$. Indeed, the recursion gives $x_{m+1} > g(x_m) > x_m$ in that case. The subsequent values would then increase and, since $x_m < 1$, converge to a limit in $(\tau, 1]$. Passing to the limit and using $r_m \to 0$ would give $x = g(x)$, which is impossible in this interval.

Now fix $\varepsilon \in (0, \tau)$. At $r = 0$, the quotient x_{m+1}/x_m is $x(1+x)$. Since $\tau(1+\tau) = 1$,

$$x(1+x) \le (\tau - \varepsilon)(1 + \tau - \varepsilon) < 1, \qquad 0 < x \le \tau - \varepsilon.$$

By continuity, there are $q < 1$ and $r_0 > 0$ such that

$$x\big((1+r)^2 + x(1+2r)\big) \le q$$

whenever $0 < x \le \tau - \varepsilon$ and $0 \le r \le r_0$. Since $r_m \to 0$, this applies to every sufficiently late iterate. If such an iterate satisfied $x_m \le \tau - \varepsilon$, then $x_{m+1} \le q x_m \le \tau - \varepsilon$, and induction would give $x_{m+j} \le q^j x_m$ for every $j \ge 0$. The full orbit would then converge to the origin, contradicting the preceding paragraph. Therefore

$$\tau - \varepsilon < x_m \le \tau$$

for all sufficiently large m. Since ε is arbitrary, $x_m \to \tau$, and $y_m = r_m x_m \to 0$.

The scaling comparison used in (41), now with minimum degree 2, shows that every initial point corresponding to $\beta > \beta_{*,3}$ tends to zero. Every initial point corresponding to $\beta < \beta_{*,3}$ belongs to $\mathcal{E}_3$. Since $\mathbf{C}_m^{(3)}(\beta) = (x_m(\beta), y_m(\beta))$, we have proved

$$\left\{\beta \ge 0 : \sup_{m \ge 0} \|\mathbf{C}_m^{(3)}(\beta)\|_\infty < \infty\right\} = [\beta_{*,3}, \infty).$$

The definition (25) and Proposition 5.4 therefore give $\beta_{*,3} = \beta_{\mathrm{c},3}^{\mathrm{cr}} = \beta_{\mathrm{c},3}$, completing the proof. □

The pentagasket recursion differs from those in Propositions A.1 and A.2: it has four variables, and the coordinate quotient of its limiting planar map is not monotone on the whole positive half-line. We first isolate the planar argument needed to identify the critical limit.

Lemma A.3. *Consider*

$$\mathcal{G}(x, y) = \big(y^2(1+x^3), y^2 x(1+x)\big), \qquad x, y > 0.$$

Its unique positive fixed point is (θ_*, ϕ_*), where

$$\theta_*^5 + 2\theta_*^4 = 1, \qquad \phi_* = \frac{1}{\theta_*(1+\theta_*)}.$$

If a nonempty compact set $L \subset (0,\infty)^2$ satisfies $\mathcal{G}(L) = L$, then

$$L = \{(\theta_*, \phi_*)\}.$$

Proof. At a positive fixed point, the second equation gives $y = [x(1+x)]^{-1}$, and the first then becomes $x^5 + 2x^4 = 1$. Its left-hand side is strictly increasing, so the fixed point is unique and $\theta_* \in (1/2, 1)$.

We next describe the set separating decay from escape. Let

$$\mathcal{B}_0 = \{z \in (0,\infty)^2 : \mathcal{G}^m(z) \to (0,0)\}.$$

For sufficiently small $\delta > 0$, the square $(0,\delta)^2$ is invariant and every orbit in it tends to the origin, since $\|\mathcal{G}(z)\|_\infty \le 2\|z\|_\infty^2$ there. Thus $\mathcal{B}_0$ is open and downward closed in the coordinatewise order. For every fixed $x > 0$, the point (x, y) belongs to $\mathcal{B}_0$ when y is small enough, whereas it escapes to infinity when y is large enough. Hence

$$p(x) = \sup\{y > 0 : (x, y) \in \mathcal{B}_0\}$$

is finite and positive, and $(x, p(x)) \in \partial\mathcal{B}_0$. Here large y gives escape because, once both coordinates exceed one, their minimum grows at least by its fourth power. The escaping set is therefore open.

We need one elementary bound for an orbit $(x_n, y_n) = \mathcal{G}^n(z)$ starting on $\partial\mathcal{B}_0$. Such an orbit neither tends to the origin nor escapes. Since

$$\frac{y_{n+1}}{x_{n+1}} = \varphi(x_n), \qquad \varphi(x) = \frac{x(1+x)}{1+x^3} = \frac{x}{x^2 - x + 1} \le 1,$$

we have $y_n \le x_n$ for $n \ge 1$. Put $q_n = x_n y_n$. If $x_n > 1$ and $q_n > 1$, then

$$x_{n+1} = q_n^2(x_n + x_n^{-2}), \qquad y_{n+1} = q_n^2(1 + x_n^{-1}),$$

and both coordinates increase until the orbit escapes. Thus $x_n > 1$ implies $q_n \le 1$. If also $x_{n+1} > 1$, applying the same implication at time $n + 1$ gives

$$1 \ge x_{n+1}y_{n+1} \ge q_n^4 x_n, \qquad x_{n+1} \le x_n^{1/2} + x_n^{-5/2}.$$

If $x_n \le 1$, then $y_n \le x_n$, and hence

$$x_{n+1} = y_n^2(1 + x_n^3) \le x_n^2(1 + x_n^3) \le 2.$$

Together with the case $x_{n+1} \le 1$, this estimate shows that (x_n), and hence (y_n), is bounded above. Neither coordinate can approach zero along a subsequence: for x_n this would put a later iterate in $(0,\delta)^2$, while for y_n it follows from $y_n = x_n\varphi(x_{n-1})$. Every boundary orbit therefore has a compact tail in $(0,\infty)^2$.

The scaling inequality

$$\mathcal{G}(tz) \ge t^2\mathcal{G}(z), \qquad t \ge 1,$$

now shows that no two distinct boundary points are comparable. In fact, strict monotonicity would give boundary points $u < v$ and $t > 1$ such that $\mathcal{G}(v) \ge t\mathcal{G}(u)$; iteration would then give $\mathcal{G}^{m+1}(v) \ge t^{2^m}\mathcal{G}^{m+1}(u)$, contradicting the compact-tail bound. It follows that

$$\partial\mathcal{B}_0 = \{(x, p(x)) : x > 0\},$$

where p is strictly decreasing. It is also continuous: monotonicity makes p nonincreasing, while any jump would produce two comparable boundary points on the same vertical line. Moreover, if $y > p(x)$, comparison with the boundary orbit through $(x, p(x))$ and the same scaling inequality show that (x, y) escapes. Thus every point lies either below this curve, on it, or above it, according as its orbit tends to zero, stays on the curve, or escapes. In particular, $\mathcal{G}(\partial\mathcal{B}_0) \subseteq \partial\mathcal{B}_0$: an image below the curve would put the original point in $\mathcal{B}_0$, whereas an image above it would make the original point escape.

Put $h(x) = p(x)/x$, and let $T(x)$ be the first coordinate of $\mathcal{G}(x, p(x))$. The image again lies on the boundary curve, and hence

$$h(T(x)) = \varphi(x).$$

Since h is continuous and strictly decreasing, T is continuous. Since

$$\varphi'(x) = \frac{1 - x^2}{(x^2 - x + 1)^2}, \qquad \varphi(x) = \varphi(1/x),$$

the map T is decreasing on $(0,1)$, increasing on $(1,\infty)$, and satisfies $T(x) = T(1/x)$.

To locate its fixed point, set

$$Q(x) = \big(x, x\varphi(x)\big).$$

A direct calculation gives

$$\mathcal{G}(Q(x)) = R(x)Q(x), \qquad R(x) = \frac{x^3(1+x)^2}{1+x^3}.$$

The function R is strictly increasing because

$$\frac{R'(x)}{R(x)} = \frac{3}{x(1+x^3)} + \frac{2}{1+x} > 0,$$

and $R(x) = 1$ precisely when $x = \theta_*$. It follows by monotone iteration that the orbit of $Q(x)$ tends to the origin for $x < \theta_*$ and escapes for $x > \theta_*$. Comparing $Q(x)$ with the boundary curve gives

$$T(x) > x \quad (x < \theta_*), \qquad T(x) < x \quad (x > \theta_*).$$

The interval $[1/2, 1]$ attracts every orbit of T. Indeed,

$$\mathcal{G}(1/2, 1/2) = (9/32, 3/16) < (1/2, 1/2),$$

and its orbit decreases monotonically to the origin, so $p(1/2) > 1/2$. Also $\mathcal{G}(1, 2/3) = (8/9, 8/9)$, whose next iterate has both coordinates larger than one, so $p(1) < 2/3$. Therefore

$$h(1) < \frac{2}{3} \le \varphi(x) \le 1 < h(1/2), \qquad \frac{1}{2} \le x \le 1,$$

and the decrease of h gives $T([1/2, 1]) \subset (1/2, 1)$. Below $1/2$, the iterates increase while they remain below $1/2$; they cannot remain there forever, since their limit would be a fixed point below $1/2$. If $x > 1$, then $T(x) = T(1/x) > 1/2$: this follows from the invariance of $[1/2, 1]$ when $1/x \ge 1/2$, and from the decrease of T on $(0, 1)$ otherwise. The iterates decrease while they remain above one and must therefore enter $[1/2, 1]$, since T has no fixed point above one.

Finally, $\mathcal{G}$ has no nontrivial positive two-cycle. If (x, y) and (u, v) formed one, the coordinate quotients and the first-coordinate equations would give

$$\frac{v}{u} = \varphi(x), \qquad \frac{y}{x} = \varphi(u),$$

and then

$$J(x) = J(u), \qquad J(x) = x(x^2 - x + 1)^2(1 + x^3).$$

But

$$\frac{J'(x)}{J(x)} = \frac{5x^2 - 3x + 1}{x(x^2 - x + 1)} + \frac{3x^2}{1 + x^3} > 0,$$

so $x = u$ and then $y = v$. Since T is decreasing on $[1/2, 1]$, its even and odd iterates converge to a possible two-cycle. The preceding calculation excludes a nontrivial one, and therefore every orbit on the boundary curve converges to (θ_*, ϕ_*).

Now let $L \subset (0, \infty)^2$ be nonempty and compact, with $\mathcal{G}(L) = L$. Compactness excludes both decay to the origin and escape, so L lies on the boundary curve. The eventual-entry property and compactness give a common entry time for all first coordinates: the sets of points entering the forward-invariant open interval $(1/2, 1)$ by time n form an increasing open cover of the compact projection of L. Invariance of L then shows that the first coordinates already lie in $(1/2, 1)$. If I is their compact set of first coordinates, then the first-coordinate action of $\mathcal{G}(L) = L$ gives $T(I) = I$. Since T is decreasing,

$$T(\min I) = \max I, \qquad T(\max I) = \min I.$$

The absence of a nontrivial two-cycle gives $\min I = \max I = \theta_*$, and the conclusion follows. $\square$

Proposition A.4. *For $N = 5$, the recursion (28) has a unique critical value $\beta_{*,5} > 0$. All four variables tend to zero for $\beta > \beta_{*,5}$, and all four tend to infinity for $\beta < \beta_{*,5}$. At the critical value,*

$$(\phi_{1,m}, \phi_{2,m}, \theta_{1,m}, \theta_{2,m}) \longrightarrow (\phi_*, 0, \theta_*, 0),$$

where (θ_*, ϕ_*) is the fixed point in Lemma A.3. Consequently, $\beta_{*,5} = \beta_{\mathrm{c},5}^{\mathrm{cr}} = \beta_{\mathrm{c},5}$.

Proof. Put

$$a_m = \phi_{1,m}, \quad b_m = \phi_{2,m}, \quad \xi_m = \theta_{1,m}, \quad \widehat{\xi}_m = \theta_{2,m},$$

and introduce the two ratios

$$r_m = \frac{b_m}{a_m}, \qquad s_m = \frac{\widehat{\xi}_m}{\xi_m}.$$

The recursion gives

$$\begin{aligned} r_{m+1} &= s_m \frac{\xi_m}{1+\xi_m}, \\ s_{m+1} &= s_m \frac{H_m}{1+H_m}, \\ H_m &= \xi_m^3 \frac{(1+s_m)^2(1+2r_m)}{(1+r_m)^2}. \end{aligned} \tag{43}$$

Since $r_0 = e^{-\beta} \le 1$ and $s_0 = e^{-3\beta} \le 1$, (43) implies

$$0 < r_m \le 1, \qquad 0 < s_m \le 1.$$

The main pair satisfies

$$\begin{aligned} \xi_{m+1} &= a_m^2 \left[(1+r_m)^2 + \xi_m^3 (1+s_m)^2 (1+2r_m)\right], \\ a_{m+1} &= a_m^2 \xi_m (1+r_m)^2 (1+s_m)(1+\xi_m). \end{aligned} \tag{44}$$

When $r_m = s_m = 0$, the right-hand side is $\mathcal{G}(\xi_m, a_m)$, with $\mathcal{G}$ as in Lemma A.3.

We first separate escaping from bounded orbits. If $a_m > 1$ and $\xi_m > 1$, then (44) keeps both variables above one and makes them tend to infinity. In this case $\xi_{m+1} \ge \xi_m^3$ and

$$H_m \ge \frac{1}{4}\xi_m^3, \qquad \frac{s_{m+1}}{s_m} = 1 - \frac{1}{1+H_m}.$$

Thus $\sum_m (1+H_m)^{-1} < \infty$, so s_m, and then r_m, converges to a positive limit. Consequently, $b_m = r_m a_m$ and $\widehat{\xi}_m = s_m \xi_m$ also tend to infinity. We call such an orbit *escaping*.

Every orbit from the prescribed initial conditions that does not escape is uniformly bounded. Indeed, it must satisfy

$$a_m \xi_m \le 1, \qquad a_m^2 \xi_m \le 1.$$

To see this, if $\xi_m > 1$ and $a_m \xi_m > 1$, or if $\xi_m \le 1 < a_m$ and $a_m^2 \xi_m > 1$, the next main pair is already in $(1, \infty)^2$. The remaining cases give the two inequalities directly. Hence the second equation in (44) yields

$$a_{m+1} = (1+r_m)^2 (1+s_m)\big(a_m^2 \xi_m + (a_m \xi_m)^2\big) \le 16.$$

Since $a_0 = e^{-2\beta} \le 1$, it follows that $a_m \le 16$ for every m. If $\xi_m \le 1$, the first equation in (44) gives

$$\xi_{m+1} \le 16^2 (4+12) = 4096.$$

If $\xi_m > 1$ and $\xi_{m+1} > 1$, put $q_m = a_m \xi_m$. Because the orbit does not escape, the inequalities at time $m+1$ give

$$1 \ge a_{m+1} \xi_{m+1} \ge q_m^4 \xi_m,$$

and therefore

$$\xi_{m+1} \le 4\xi_m^{-5/2} + 12\xi_m^{1/2} \le 16\xi_m^{1/2}.$$

In the last case, $\xi_m \le 4096$ implies $\xi_{m+1} \le 16\sqrt{4096} = 1024$. Together with the cases $\xi_m \le 1$ and $\xi_{m+1} \le 1$, induction from $\xi_0 = e^{-\beta} \le 1$ therefore gives

$$a_m \le 16, \qquad \xi_m \le 4096,$$

and $r_m, s_m \le 1$ then bound all four variables.

Write $t = e^{-\beta}$, so the initial point is

$$(a_0, b_0, \xi_0, \widehat{\xi}_0) = (t^2, t^3, t, t^4).$$

The recursion is strictly increasing in each variable, and the initial point is strictly increasing in t. Let $\mathcal{E}_5$ be the set of $t \in (0,1]$ for which the orbit escapes. The preceding argument shows that escape occurs exactly when $a_m > 1$ and $\xi_m > 1$ for some m. Thus $\mathcal{E}_5$ is open and upward closed. It contains 1, because $a_1 = \xi_1 = 16$ when $t = 1$, and it excludes all sufficiently small t, whose initial points lie in an invariant neighborhood of the origin. Therefore

$$\mathcal{E}_5 = (t_{*,5}, 1]$$

for a unique $t_{*,5} \in (0,1)$. Set $\beta_{*,5} = -\log t_{*,5}$.

The critical orbit does not escape and is therefore bounded. It cannot tend to the origin, since continuity of a finite iterate would then put an orbit with $t > t_{*,5}$ in the same invariant neighborhood, contrary to the definition of $t_{*,5}$. Since ξ_m is bounded, H_m in (43) is bounded. Consequently, if $H_* = \sup_m H_m$ and $q = H_*/(1+H_*) < 1$, then

$$s_m \le s_0 q^m \le q^m, \qquad r_{m+1} \le s_m.$$

Thus the two ratios decay geometrically. By (44),

$$(\xi_{m+1}, a_{m+1}) = \mathcal{G}(\xi_m, a_m) + e_m, \qquad \|e_m\| \le C(r_m + s_m)$$

for some $C < \infty$, and hence $\sum_m \|e_m\| < \infty$.

The critical main pair stays away from both coordinate axes. If $a_{m_j} \to 0$, boundedness of ξ_m, r_m, s_m in (44) gives $(\xi_{m_j+1}, a_{m_j+1}) \to (0,0)$. If $\xi_{m_j} \to 0$, the second equation first gives $a_{m_j+1} \to 0$, while ξ_{m_j+1} remains bounded; applying the preceding argument at time $m_j + 1$ gives $(\xi_{m_j+2}, a_{m_j+2}) \to (0,0)$. Since $b_m \le a_m$ and $\widehat{\xi}_m \le \xi_m$, either alternative would put a full iterate in the invariant neighborhood of the origin, which is impossible at $t_{*,5}$.

Let L be the omega-limit set of (ξ_m, a_m). It is a nonempty compact subset of $(0,\infty)^2$. Since $e_m \to 0$, limits of shifted subsequences give $\mathcal{G}(L) \subseteq L$. Conversely, taking a convergent subsequence of the corresponding predecessors gives $\mathcal{G}(L) \supseteq L$. Hence $\mathcal{G}(L) = L$, and Lemma A.3 gives

$$L = \{(\theta_*, \phi_*)\}.$$

Therefore

$$(\xi_m, a_m) \longrightarrow (\theta_*, \phi_*),$$

while $b_m = r_m a_m \to 0$ and $\widehat{\xi}_m = s_m \xi_m \to 0$. This proves the critical limit.

If $t < t_{*,5}$, choose $\lambda \in (0,1)$ so that its initial point is at most λ times the critical initial point. Every monomial in (28) has degree at least two, so induction bounds the corresponding orbit by λ^{2^m} times the bounded critical orbit; it therefore tends to zero. If $t > t_{*,5}$, all four variables tend to infinity by the definition of $\mathcal{E}_5$. Since $t = e^{-\beta}$,

$$\left\{\beta \ge 0 : \sup_{m \ge 0} \|\mathbf{C}_m^{(5)}(\beta)\|_\infty < \infty\right\} = [\beta_{*,5}, \infty).$$

The definition (25) and Proposition 5.4 now give $\beta_{*,5} = \beta_{\mathrm{c},5}^{\mathrm{cr}} = \beta_{\mathrm{c},5}$. □

Proof of Theorem 5. For $N \notin \{3,5,6,9\}$, the conclusion follows from Proposition A.1; for $N = 3$ and $N = 5$, it follows from Propositions A.2 and A.4, respectively. In each case the critical value appearing there is $\beta_{c,N}$.

It remains to consider $N = 6$ and $N = 9$. The proof of Theorem 3 gives $X_m = Y_m = t_m$, where

$$t_{m+1} = F(t_m), \qquad F(u) = u^{d_N+1} + u^{2d_N+1},$$

whose unique positive fixed point is $t_* = \tau^{1/d_N}$. The computation in that proof and Proposition 5.4 show that the initial value equals t_* precisely when $\beta = \beta_{c,N}$, and the orbit is then constant. If $\beta > \beta_{c,N}$, the initial value is smaller than t_*; since $0 < F(u) < u$ for $0 < u < t_*$, the orbit decreases to zero. If $\beta < \beta_{c,N}$, the initial value is larger than t_*; since $F(u) > u$ for $u > t_*$, the orbit increases to infinity. This proves all four cases. □

Appendix B Off-critical rates

Away from criticality, every crossing variable tends either to zero or to infinity. Taking logarithms turns each polynomial recursion into a linear recursion with a bounded error.

The following lemma gives the asymptotic estimate used later in Theorem 6.

Lemma B.1. *Let* M *be a primitive nonnegative* 2×2 *matrix with Perron–Frobenius eigenvalue* $\Lambda > 1$*, and let*

$$v_{m+1} = \mathsf{M} v_m + e_m,$$

where (e_m) *is bounded. If both entries of* v_m *tend to infinity, then there exist a positive right eigenvector* v_+ *of* M *and a constant* $c > 0$ *such that*

$$\Lambda^{-m} v_m \longrightarrow c v_+.$$

Proof. Let v_+ and w_+ be positive right and left eigenvectors, normalized by $w_+^{\mathsf{T}} v_+ = 1$, and put $s_m = w_+^{\mathsf{T}} v_m$. Multiplying the recursion by w_+^{T} gives the scalar relation

$$s_{m+1} = \Lambda s_m + w_+^{\mathsf{T}} e_m.$$

Since (e_m) is bounded and $\Lambda > 1$, the sequence

$$\Lambda^{-m} s_m = \Lambda^{-m_0} s_{m_0} + \sum_{j=m_0}^{m-1} \Lambda^{-j-1} w_+^{\mathsf{T}} e_j$$

has a finite limit c.

The second eigenvalue Λ_2 of M satisfies $|\Lambda_2| < \Lambda$. Let

$$\Pi_+ = v_+ w_+^{\mathsf{T}}, \qquad \Pi_\perp = I - \Pi_+.$$

Then Π_+ is the projection onto the Perron–Frobenius direction and $\mathsf{M}\Pi_\perp = \Lambda_2 \Pi_\perp$. Applying $\Pi_\perp$ and iterating gives

$$\Pi_\perp v_m = \Lambda_2^{m-m_0} \Pi_\perp v_{m_0} + \sum_{j=m_0}^{m-1} \Lambda_2^{m-1-j} \Pi_\perp e_j.$$

Because (e_j) is bounded and $|\Lambda_2| < \Lambda$, division by Λ^m makes the right-hand side tend to zero. Combining this with the limit of $\Lambda^{-m} s_m$ gives

$$\Lambda^{-m} v_m \longrightarrow c v_+.$$

It remains to show that $c > 0$. Since both entries of v_m tend to infinity and $w_+ > 0$, one has $s_m \to \infty$, while the preceding limit gives $c \geq 0$. If $c = 0$, the scalar recursion could be solved backwards as

$$s_m = -\sum_{j=m}^{\infty} \Lambda^{m-1-j} w_+^{\mathsf{T}} e_j,$$

so (s_m) is bounded. This contradicts the positivity of w_+ and the assumption that both entries of v_m tend to infinity. Therefore $c > 0$. □

Theorem 6. *For $N \geq 3$ with $4 \nmid N$, define*

$$\Lambda_N^+ = \begin{cases} 3, & N = 3, \\ \dfrac{5+\sqrt{17}}{2}, & N = 5, \\ k_N + \dfrac{3}{2} + \sqrt{\left(k_N + \dfrac{3}{2}\right)^2 + 2(k_N + \ell_N - 3)}, & N \notin \{3, 5\}, \end{cases}$$

$$\Lambda_N^- = \begin{cases} 2, & N = 3, \\ 1+\sqrt{3}, & N = 5, \\ k_N + \dfrac{\ell_N - 1}{2} + \sqrt{\left(k_N + \dfrac{\ell_N - 1}{2}\right)^2 - 2(k_N + \ell_N - 3)}, & N \notin \{3, 5\}. \end{cases} \tag{45}$$

If $\beta < \beta_{\mathrm{c},N}$, then for every entry $\mathsf{c}_m(\beta)$ of $\mathbf{C}_m^{(N)}(\beta)$,

$$0 < \lim_{m\to\infty} (\Lambda_N^+)^{-m} \log \mathsf{c}_m(\beta) < \infty.$$

If $\beta > \beta_{\mathrm{c},N}$, then

$$0 < \lim_{m\to\infty} (\Lambda_N^-)^{-m} \big(-\log \mathsf{c}_m(\beta)\big) < \infty.$$

Proof. The phase diagram in Theorem 5 ensures that all crossing variables are eventually larger than one when $\beta < \beta_{\mathrm{c},N}$, and eventually smaller than one when $\beta > \beta_{\mathrm{c},N}$. Thus all logarithms below have the stated sign for sufficiently large m.

Suppose first that $N = 4k + \ell \notin \{3, 5\}$. On the divergent side, the larger power of Y_m in each sum in (26) gives

$$\begin{pmatrix} \log X_{m+1} \\ \log Y_{m+1} \end{pmatrix} = \mathsf{M}_N^+ \begin{pmatrix} \log X_m \\ \log Y_m \end{pmatrix} + e_m^+, \qquad \mathsf{M}_N^+ = \begin{pmatrix} 2 & 3k+\ell-2 \\ 2 & 2k+1 \end{pmatrix},$$

where e_m^+ is bounded; the error consists only of logarithms of factors of the form $1 + Y_m^{-a}$, with $a > 0$. On the side converging to zero, the smaller power of Y_m gives

$$\begin{pmatrix} -\log X_{m+1} \\ -\log Y_{m+1} \end{pmatrix} = \mathsf{M}_N^- \begin{pmatrix} -\log X_m \\ -\log Y_m \end{pmatrix} + e_m^-, \qquad \mathsf{M}_N^- = \begin{pmatrix} 2 & k \\ 2 & 2k+\ell-3 \end{pmatrix},$$

where e_m^- is bounded for the same reason, now with factors of the form $1 + Y_m^a$. Both matrices are primitive. Direct calculation of their characteristic polynomials gives the two Perron–Frobenius eigenvalues $\Lambda_N^{\pm}$ in (45). In particular, $\det \mathsf{M}_N^- = 2(k + \ell - 3)$, which produces the minus sign under the square root in the formula for Λ_N^-. Lemma B.1 now gives the required limit for both crossing variables.

For $N = 3$, put $r_m = y_m/x_m$ as in (42). On the escaping side, r_m decreases to a positive limit. Indeed, once $x_m > 1$,

$$0 < 1 - \frac{r_{m+1}}{r_m} = \frac{(1+r_m)^2}{(1+r_m)^2 + x_m(1+2r_m)} \leq \frac{4}{x_m},$$

whereas $x_{m+1} \geq x_m^3$. Therefore $\sum_m (1 - r_{m+1}/r_m) < \infty$, and the product $\prod_m (r_{m+1}/r_m)$ is positive. Since r_m is decreasing, it follows that r_m converges to a positive limit. The logarithmic recursion is consequently

$$\begin{pmatrix} \log x_{m+1} \\ \log y_{m+1} \end{pmatrix} = \begin{pmatrix} 3 & 0 \\ 2 & 1 \end{pmatrix} \begin{pmatrix} \log x_m \\ \log y_m \end{pmatrix} + O(1).$$

On the side converging to zero, $r_m \to 0$ and

$$\begin{pmatrix} -\log x_{m+1} \\ -\log y_{m+1} \end{pmatrix} = \begin{pmatrix} 2 & 0 \\ 2 & 1 \end{pmatrix} \begin{pmatrix} -\log x_m \\ -\log y_m \end{pmatrix} + O(1).$$

In both recursions, the first component has the form $v_{m+1} = \Lambda v_m + O(1)$, where $v_m \to \infty$ and $\Lambda = 3$ or 2. The elementary one-dimensional version of Lemma B.1 gives a finite positive limit for $\Lambda^{-m} v_m$; denote it by $c_v > 0$. If ω_m denotes the second component, then

$$\omega_{m+1} = \omega_m + 2v_m + O(1),$$

and hence

$$\omega_m = \omega_{m_0} + 2 \sum_{j=m_0}^{m-1} v_j + O(m).$$

It follows that

$$\Lambda^{-m} \omega_m \longrightarrow \frac{2c_v}{\Lambda - 1} > 0.$$

This proves $\Lambda_3^+ = 3$ and $\Lambda_3^- = 2$ in (45).

Finally, let $N = 5$, and use the notation $a_m, b_m, \xi_m, \widehat{\xi}_m, r_m, s_m$ from Proposition A.4. On the escaping side, the proof of Proposition A.4 shows that s_m has a positive limit and that r_m has the same limit. Taking logarithms in (44) therefore gives

$$\begin{pmatrix} \log a_{m+1} \\ \log \xi_{m+1} \end{pmatrix} = \begin{pmatrix} 2 & 2 \\ 2 & 3 \end{pmatrix} \begin{pmatrix} \log a_m \\ \log \xi_m \end{pmatrix} + O(1).$$

On the side converging to zero, $\xi_m \to 0$. Since $r_m, s_m \leq 1$, (43) gives

$$H_m \leq 12\xi_m^3 \longrightarrow 0, \qquad s_{m+1} \leq H_m \longrightarrow 0, \qquad r_{m+1} \leq s_m \longrightarrow 0.$$

Thus $r_m, s_m \to 0$, and (44) gives

$$\begin{pmatrix} -\log a_{m+1} \\ -\log \xi_{m+1} \end{pmatrix} = \begin{pmatrix} 2 & 1 \\ 2 & 0 \end{pmatrix} \begin{pmatrix} -\log a_m \\ -\log \xi_m \end{pmatrix} + O(1).$$

The two Perron–Frobenius eigenvalues are $(5 + \sqrt{17})/2$ and $1 + \sqrt{3}$, respectively. Lemma B.1 gives the stated limits for a_m and ξ_m. On the escaping side, the positive limits of r_m and s_m give the same limits for b_m and $\widehat{\xi}_m$. On the side converging to zero, (43) gives

$$-\log s_{m+1} = -\log s_m + 3(-\log \xi_m) + O(1), \qquad -\log r_{m+1} = -\log s_m - \log \xi_m + O(1).$$

Write

$$(\Lambda_5^-)^{-m}(-\log \xi_m) \longrightarrow \gamma_\xi > 0.$$

The first relation is a scalar inhomogeneous recursion. Summing it and using the limit for $-\log \xi_m$ gives

$$(\Lambda_5^-)^{-m}(-\log s_m) \longrightarrow \frac{3\gamma_\xi}{\Lambda_5^- - 1} > 0.$$

Substitution in the second relation then gives a finite positive limit for $(\Lambda_5^-)^{-m}(-\log r_m)$. Since

$$-\log b_m = -\log a_m - \log r_m, \qquad -\log \widehat{\xi}_m = -\log \xi_m - \log s_m,$$

the two remaining crossing variables have finite positive normalized limits as well. This completes the proof. □

References

[1] M. T. Barlow, Diffusions on fractals, in *Lectures on Probability Theory and Statistics (Saint-Flour, 1995)*, Lecture Notes in Math. **1690**, Springer, Berlin, 1998, 1–121.

[2] D. Ben-Avraham and S. Havlin, Self-avoiding walks on finitely ramified fractals, *Phys. Rev. A* **29** (1984), R2309–R2311.

[3] H. Duminil-Copin and S. Smirnov, The connective constant of the honeycomb lattice equals $\sqrt{2+\sqrt{2}}$, *Ann. of Math. (2)* **175** (2012), no. 3, 1653–1665.

[4] P. J. Fitzsimmons, B. M. Hambly and T. Kumagai, Transition density estimates for Brownian motion on affine nested fractals, *Comm. Math. Phys.* **165** (1994), 595–620.

[5] J. M. Hammersley, Percolation processes. II. The connective constant, *Proc. Cambridge Philos. Soc.* **53** (1957), 642–645.

[6] J. M. Hammersley and K. W. Morton, Poor man's Monte Carlo, *J. Roy. Statist. Soc. Ser. B* **16** (1954), 23–38.

[7] J. M. Hammersley and D. J. A. Welsh, Further results on the rate of convergence to the connective constant of the hypercubical lattice, *Quart. J. Math. Oxford Ser. (2)* **13** (1962), 108–110.

[8] T. Hara and G. Slade, Self-avoiding walk in five or more dimensions. I. The critical behaviour, *Comm. Math. Phys.* **147** (1992), 101–136.

[9] T. Hara and G. Slade, The lace expansion for self-avoiding walk in five or more dimensions, *Rev. Math. Phys.* **4** (1992), 235–327.

[10] K. Hattori and T. Hattori, Self-avoiding process on the Sierpiński gasket, *Probab. Theory Related Fields* **88** (1991), 405–428.

[11] K. Hattori, T. Hattori and S. Kusuoka, Self-avoiding paths on the pre-Sierpiński gasket, *Probab. Theory Related Fields* **84** (1990), 1–26.

[12] K. Hattori, T. Hattori and S. Kusuoka, Self-avoiding paths on the three-dimensional Sierpiński gasket, *Publ. Res. Inst. Math. Sci.* **29** (1993), 455–509.

[13] T. Hattori and S. Kusuoka, The exponent for the mean square displacement of self-avoiding random walk on the Sierpiński gasket, *Probab. Theory Related Fields* **93** (1992), 273–284.

[14] T. Hattori and T. Tsuda, Renormalization group analysis of the self-avoiding paths on the d-dimensional Sierpiński gaskets, *J. Stat. Phys.* **109** (2002), 39–66.

[15] T. Hutchcroft, The Hammersley–Welsh bound for self-avoiding walk revisited, *Electron. Commun. Probab.* **23** (2018), no. 5, 1–8.

[16] H. Kesten, On the number of self-avoiding walks, *J. Math. Phys.* **4** (1963), 960–969.

[17] T. Kumagai, Short time asymptotic behaviour and large deviation for Brownian motion on some affine nested fractals, *Publ. Res. Inst. Math. Sci.* **33** (1997), no. 2, 223–240.

[18] T. Lindstrøm, *Brownian Motion on Nested Fractals*, *Mem. Amer. Math. Soc.* **83** (1990), no. 420.

[19] N. Madras and G. Slade, *The Self-Avoiding Walk*, Probability and Its Applications, Birkhäuser, Boston, MA, 1993.

[20] J. T. Tyson and J.-M. Wu, Quasiconformal dimensions of self-similar fractals, *Rev. Mat. Iberoam.* **22** (2006), no. 1, 205–258.

School of Mathematics, Nanjing University, Nanjing, 210093, P. R. China.
Email address: huaqiu@nju.edu.cn

School of Mathematics, Nanjing University, Nanjing, 210093, P. R. China.
Email address: 602025210014@smail.nju.edu.cn