ALEA, Lat. Am. J. Probab. Math. Stat. 19, 51–68 (2022)
DOI: 10.30757/ALEA.v19-03

Bound on the running maximum of a random walk with small drift

Ofer Busani and Timo Seppäläinen

Ofer Busani, University of Bristol, School of Mathematics, Fry Building, Woodland Rd., Bristol BS8 1UG, UK.
E-mail address: o.busani@bristol.ac.uk
URL: https://people.maths.bris.ac.uk/∼di18476/

Timo Seppäläinen, University of Wisconsin-Madison, Mathematics Department, Van Vleck Hall, 480 Lincoln Dr., Madison WI 53706-1388, USA.
E-mail address: seppalai@math.wisc.edu
URL: http://www.math.wisc.edu/∼seppalai

Abstract. We derive a lower bound for the probability that a random walk with i.i.d. increments and small negative drift μ exceeds the value $x > 0$ by time N. When the moment generating functions are bounded in an interval around the origin, this probability can be bounded below by $1 - O(x|\mu| \log N)$. The approach is elementary and does not use strong approximation theorems.

Contents

Received by the editors July 29th, 2021; accepted September 21st, 2021.
2010 *Mathematics Subject Classification.* 60G50, 60G17.
Key words and phrases. random walk, running maximum, lower bound.
O. Busani was supported by EPSRC's EP/R021449/1 Standard Grant.
T. Seppäläinen was partially supported by National Science Foundation grant DMS-1854619 and by the Wisconsin Alumni Research Foundation.

1. Introduction

1.1. *Background.* This paper arose from the need of a random walk estimate for the authors' article Busani and Seppäläinen (2020) on directed polymers. This estimate is a *positive* lower bound on the running maximum of a random walk with a small *negative* drift. Importantly, the bound had to come with sufficient control over its constants so that it would apply to an infinite sequence of random walks whose drift scales to zero as the maximum is taken over expanding time intervals. The natural approach via a Brownian motion embedding appeared to not give either the desired precision or the uniformity. Hence we resorted to a proof from scratch. For possible wider use we derive the result here under general hypotheses on the distribution of the step of the walk.

The polymer application of the result pertains to the exactly solvable log-gamma polymer on the plane. The objective of Busani and Seppäläinen (2020) is to prove that there are no bi-infinite polymer paths on the planar lattice $\mathbb{Z}^2$. The technical result is that there does not exist any nontrivial Gibbs measures on bi-infinite paths that satisfy the Dobrushin-Lanford-Ruelle (DLR) equations under the Gibbsian specification defined by the quenched polymer measures. In terms of limits of finite polymer distributions, this means that as the southwest and northeast endpoints of a random polymer path are taken to opposite infinities, the middle portion of the path escapes. This is proved by showing that in the limit the probability that the path crosses the y-axis along a given edge decays to zero. This probability in turn is controlled by considering stationary polymer processes from the two endpoints to an interval along the y-axis. The crossing probability can be controlled in terms of a maximum of a random walk. In the case of the log-gamma polymer, the steps of this random walk are distributed as the difference of two independent log-gamma variables. The case needed for Busani and Seppäläinen (2020) is treated in Example 2.7 below.

1.2. *The question considered.* We seek a lower bound on the probability that the running maximum of a random walk with negative drift reaches a level $x > 0$. To set the stage, we discuss the matter through Brownian motion. Let $S_n^N = \sum_{i=1}^n X_i^N$ be a random walk with drift $\mathbb{E}(X_1^N) = \mu_N = \mu N^{-1/2} < 0$, and such that the random walks S^N converge weakly to a Brownian motion with drift $\mu < 0$. The probability of the event

$$\sup_{1\leqslant n\leqslant N} S_n^N > x$$

should be approximately the same as that of

$$\sup_{0\leqslant s\leqslant 1} (B_s + s\mu) > xN^{-1/2}. \tag{1.1}$$

This latter can be computed (see (3.7)) to be

$$\mathbb{P}\big\{\sup_{0\leqslant s\leqslant 1} (B_s + s\mu) > xN^{-1/2}\big\} = 1 - O(|\mu| xN^{-1/2}). \tag{1.2}$$

This suggests that we should aim for an estimate of the type

$$\mathbb{P}\Big(\sup_{1\leqslant n\leqslant N} S_n^N > x\Big) \geqslant 1 - O(|\mu_N| x). \tag{1.3}$$

To reach this precision weak convergence is not powerful enough, for a weak approximation of random walk by Brownian motion reaches only a precision of $O(N^{-1/4})$ (Sawyer, 1972; Fraser, 1973). Our estimate (2.4) below does almost capture (1.3): we have to allow an additional $\log N$ factor inside the $O(\cdot)$ and consider x of order at least $(\log N)^2$.

The by-now classical Komlós-Major-Tusnády (KMT) coupling (Komlós et al., 1976) gives a strong approximation of random walk with Brownian motion with a discrepancy that grows logarithmically in time. This precision is sufficient for us, as we illustrate in Section 3. The problem is now the control of the constants in the approximation. Uniformity of the constants is necessary for our application in Busani and Seppäläinen (2020). But verifying this uniformity from the original work

Komlós et al. (1976) appeared to be a highly nontrivial task. In the end it was more efficient to derive the estimate (Theorem 2.2 below) from the ground up.

The difficulty of the original KMT proof has motivated several recent attempts at simplification and better understanding of the result, such as Bhattacharjee and Goldstein (2016), Chatterjee (2012), and Krishnapur (2020). There is another strong approximation result due to Sakhanenko (1984) which, according to p. 232 of Chatterjee (2012), "is so complex that some researchers are hesitant to use it".

1.3. *Sketch of the proof.* Our proof is elementary. The most sophisticated result used is the Berry-Esseen theorem. Given a random walk of small drift $\mu < 0$, our approach can be summarized in two main steps:

(1) Up to the time the random walk hits the level $-\varepsilon|\mu|^{-1}$ it behaves like an unbiased random walk.
(2) By the time the random walk hits the level $-\varepsilon|\mu|^{-1}$ it will have had about $\log_2(\varepsilon|\mu|^{-1}x^{-1})$ independent opportunities to hit the level x. By the previous step this implies that the probability on the left-hand side of (1.3) is of order $1-(1/2)^{\log_2(\varepsilon|\mu|^{-1}x^{-1})} = 1-O(|\mu|\varepsilon^{-1}x)$.

As we will take $\varepsilon = (\log N)^{-1}$ in the proof, we will obtain the right order in (1.3) up to a logarithmic factor (Theorem 2.2). After the statement of the theorem we illustrate it with examples. Then we demonstrate that even if we knew that the constants in the KMT approximation can be taken uniform, the result would not be essentially stronger in the regime in which we apply our result.

2. Main result

For each $N \in \mathbb{Z}_{>0}$, let $\{X_i^N\}_{i\geqslant 1}$ be a sequence of non-degenerate i.i.d. random variables. Denote their moment generating function by

$$M_N(\theta) = \mathbb{E}\big(e^{\theta X_1^N}\big).$$

Write $M_N^{(0)} = M_N$ and $M_N^{(i+1)} = (d/d\theta)M_N^{(i)}$.

Assumption 2.1. *We assume that the random variables $\{X_i^N\}_{i\geqslant 1}$ satisfy the following:*

(i) *There exists an open interval $(-\theta_0, \theta_0)$ around the origin on which each moment generating function M_N is finite. Furthermore, there exists a finite constant C_M and $\theta_1 > 0$ such that we have the uniform bounds*

$$|M_N^{(i)}(\theta)| \leqslant C_M \quad \text{for all } N,\ 0 \leqslant i \leqslant 3, \text{ and } \theta \in [-\theta_1, \theta_1] \tag{2.1}$$

for the compact interval $[-\theta_1, \theta_1] \subset (-\theta_0, \theta_0)$.

(ii) *There exists a finite constant $\sigma_* > 0$ such that*

$$\mathbb{E}\big[(X_1^N)^2\big] \geqslant \sigma_N^2 = \mathbb{V}\mathrm{ar}(X_1^N) \geqslant \sigma_*^2 \quad \text{for all } N. \tag{2.2}$$

(iii) *There exists a finite constant $D_\mu > 0$ such that the expectations $\mu_N = \mathbb{E}\big(X_1^N\big)$ satisfy*

$$-D_\mu(\log N)^{-3} \leqslant \mu_N \leqslant 0 \quad \text{for all } N. \tag{2.3}$$

The conditions in Assumption 2.1 are fairly natural. Note that (2.1) has to be checked only for $i = 0$ at the expense of shrinking the interval $[-\theta_1, \theta_1]$ and increasing C_M. To make a positive maximum possible, condition (2.2) ensures enough diffusivity and condition (2.3) limits the strength of the negative drift. The bound (2.4) below shows that D_μ has to be vanishingly small in order for the result to be nontrivial.

For $m \geqslant 1$ let $S_m^N = \sum_{k=1}^m X_k^N$ be the random walk associated with the steps $\{X_i^N\}_{i\geqslant 1}$. Here is the main theorem.

Theorem 2.2. *There exist finite constants C and N_0 that depend on $\theta_1, \sigma_*^2, D_\mu$ and C_M such that, for every $N \geqslant N_0$ and $x \geqslant (\log N)^2$,*

$$\mathbb{P}\Big(\max_{1\leqslant m\leqslant N} S_m^N \leqslant x\Big) \leqslant Cx(\log N)(|\mu_N| \vee N^{-1/2}). \tag{2.4}$$

Remark 2.3 (The constants in the theorem). The constant C in the upper bound (2.4) is given by

$$C = 4\exp\big\{4(C_0 + 4c_\tau^{-1})(1 + D_\mu) + 8\theta_1^{-1}\big(1 + \log(e^{\theta_1}C_M + 1)\big) + 4C_M\big\} + 3 \tag{2.5}$$

where

$$C_0 = 2(\sigma_*^{-1} + 9e^{\theta_1}\sigma_*^{-3}C_M^{5/2}) + 2e^{8C_M\sigma_*^{-4}+8\sigma_*^{-2}}c_\tau^{-1} + 12\sigma_*^{-2}\,, \tag{2.6}$$

$$c_\tau = \log\frac{2}{1 + \Phi_{\sigma_*^2}[-2, 2]}\,, \tag{2.7}$$

and $\Phi_{\sigma_*^2}$ is the mean zero Gaussian distribution with variance σ_*^2.

Throughout the proof we state explicitly the various conditions $N \geqslant N_0$ required along the way. Let us assume that $N \geqslant 2$ so that $\log N$ does not vanish. Then all the conditions $N \geqslant N_0$ can be combined into a single condition of the form

$$f(C_M, D_\mu, \sigma_*^2, \theta_1, N) \geqslant 1 \tag{2.8}$$

where the function f is a strictly positive continuous function on $\mathbb{R}^4_{>0} \times \mathbb{R}_{\geqslant 2}$, nondecreasing in θ_1, nonincreasing in C_M and D_μ, but depends on σ_*^2 in both directions. When $(C_M, D_\mu, \sigma_*^2, \theta_1)$ is restricted to a compact subset $\mathcal{K}$ of $\mathbb{R}^4_{>0}$, there exists a finite index $N_{\mathcal{K}}$ such that $f(C_M, D_\mu, \sigma_*^2, \theta_1, N)$ is a nondecreasing function of $N \geqslant N_{\mathcal{K}}$ for any fixed $(C_M, D_\mu, \sigma_*^2, \theta_1) \in \mathcal{K}$, and

$$\lim_{N\to\infty}\ \inf_{(C_M, D_\mu, \sigma_*^2, \theta_1)\in\mathcal{K}} f(C_M, D_\mu, \sigma_*^2, \theta_1, N) = \infty.$$

In particular, for each compact subset $\mathcal{K} \subset \mathbb{R}^4_{>0}$ there exists a finite index $N_{0,\mathcal{K}}$ such that (2.8) holds for all $N \geqslant N_{0,\mathcal{K}}$ and all $(C_M, D_\mu, \sigma_*^2, \theta_1) \in \mathcal{K}$. Furthermore, it is evident from (2.5)-(2.7) that C is a continuous function of $(C_M, D_\mu, \sigma_*^2, \theta_1) \in \mathbb{R}^4_{>0}$. We conclude with the following local uniformity statement.

Corollary 2.4. *For each compact subset $\mathcal{K} \subset \mathbb{R}^4_{>0}$ there exist finite constants $C_{0,\mathcal{K}}$ and $N_{0,\mathcal{K}}$ such that the following holds: the estimate (2.4) with $C = C_{0,\mathcal{K}}$ on the right-hand side is valid whenever $N \geqslant N_{0,\mathcal{K}}$, simultaneously for all walks $\{S_m^N\}_{m\geqslant 1}$ that satisfy Assumption 2.1 with parameters $(C_M, D_\mu, \sigma_*^2, \theta_1) \in \mathcal{K}$.*

We illustrate the result with some examples.

Example 2.5 (Gaussian random walk). Let B_t be a Brownian motion, $\mu < 0$, and define the random walk $S_m^N = B_m + mN^{-1/2}\mu$. We can verify that the bound (2.4) is off by a logarithmic factor in this case, by comparison with the running maximum of the Brownian motion. For $x > 0$ and large enough N

$$\begin{aligned}\mathbb{P}\big(\max_{1\leqslant m\leqslant N} S_m^N \leqslant x\big) &\geqslant \mathbb{P}\big(\sup_{0\leqslant t\leqslant N} B_t + tN^{-1/2}\mu \leqslant x\big)\\ &\geqslant 1 - e^{-2xN^{-1/2}|\mu|} \geqslant x|\mu_N| = xN^{-1/2}|\mu|.\end{aligned} \tag{2.9}$$

where the middle inequality follows from (3.7) with $\mu(N) = \mu$ and $b(N) = xN^{-1/2}$.(2.9) shows that the optimal error is at most $O(x|\mu_N|)$, and that Theorem 2.2, if not optimal, is only $\log N$ away from being so.

A natural way to produce examples is to take X_i^N as the difference of two independent random variables whose means come closer as N grows and whose variances stay bounded and bounded away from zero.

Example 2.6 (Exponential walk). Consider a random walk $S_n = \sum_{k=1}^n X_k$ with step distribution $X_k \overset{d}{=} Y_\alpha - Y_\beta$ where Y_α and Y_β are two independent exponential random variables with rates α and β, respectively. $\mu = \mathbb{E}[X_k] = \frac{\beta-\alpha}{\alpha\beta}$ so we assume that $\alpha > \beta$. The distribution of the supremum of S_n is well-known and also feasible to compute (Example (b) in Section XII.5 of Feller (1971)): for $x > 0$,

$$\mathbb{P}\big(\sup_{n\geqslant 0} S_n \leqslant x\big) = 1 - \frac{\beta}{\alpha} e^{-(\alpha-\beta)x} = \beta|\mu|\big(1+\beta x\big) + O(\mu^2 x^2)$$

where we assume $|\mu|x$ small and expand $e^s = 1 + s + O(s^2)$. We obtain a lower bound:

$$\begin{aligned}\mathbb{P}\big(\max_{1\leqslant n\leqslant N} S_n \leqslant x\big) &\geqslant P\big(\sup_{n\geqslant 0} S_n \leqslant x\big) = \beta|\mu|\big(1+\beta x\big) + O(\mu^2 x^2)\\ &\geqslant \beta^2|\mu|x + O(\mu^2 x^2).\end{aligned}$$

Thus for $|\mu| \geqslant N^{-1/2}$ and small $x|\mu|$, the upper bound (2.4) loses only a logarithmic factor.

That $\max_{1\leqslant n\leqslant N} S_n$ is close to the overall maximum $\sup_{n\geqslant 0} S_n$ in the case $|\mu| \geqslant N^{-1/2}$ is a consequence of the fact that the overall maximum is taken at a random time of order μ^{-2}. This claim is seen conveniently through ladder intervals $\{T_i\}_{i\geqslant 1}$. These are the intervals $T_i = \tau_i - \tau_{i-1}$ between successive ladder epochs defined by $\tau_0 = 0$ and

$$\tau_i = \inf\{n > \tau_{i-1} : S_n > S_{\tau_{i-1}}\}.$$

The distribution of T_i is given by

$$\mathbb{P}(T_i = \infty) = 1 - \frac{\beta}{\alpha} \quad \text{and} \quad \mathbb{P}(T_i = n) = C_{n-1}\frac{\alpha^{n-1}\beta^n}{(\alpha+\beta)^{2n-1}} \quad \text{for } n \in \mathbb{Z}_{>0},$$

where $\{C_k\}_{k\geqslant 0}$ are the Catalan numbers. (This calculation can be found in Lemma B.3 in the appendix of Fan and Seppäläinen (2020).) Set $T_0 = 0$ and let $N = \max\{n \geqslant 0 : T_n < \infty\}$ be the number of finite ladder intervals. The maximum $\sup_{n\geqslant 0} S_n$ is taken at time $\zeta = \sum_{i=1}^N T_i$. One calculates $\mathbb{E}[\zeta] = \frac{1}{\alpha\beta}\mu^{-2}$ and $\mathbb{V}\mathrm{ar}[\zeta] = c_{\alpha,\beta}\mu^{-4}$. Thus for large enough k, $\mathbb{P}(\zeta > k\mu^{-2}) \leqslant C_{\alpha,\beta}k^{-2}$.

Example 2.7 (Log-gamma walk). This is the application of Theorem 2.2 used in Busani and Seppäläinen (2020).

Let G^λ denote generically a parameter λ gamma random variable, that is, G^λ has density function $f(x) = \Gamma(\lambda)^{-1}x^{\lambda-1}e^{-x}$ on $\mathbb{R}_{>0}$. For $\alpha, \beta > 0$ let $S_m^{\alpha,\beta} = \sum_{i=1}^m X_i^{\alpha,\beta}$ denote the random walk where the distribution of the i.i.d. steps $\{X_i^{\alpha,\beta}\}_{i\geqslant 1}$ is specified by

$$X_1^{\alpha,\beta} \overset{d}{=} \log G^\alpha - \log G^\beta$$

with two independent gamma variables G^α and G^β on the right.

Let $\psi_0(s) = \Gamma'(s)/\Gamma(s)$ be the digamma function and $\psi_1(s) = \psi_0'(s)$ the trigamma function on $\mathbb{R}_{>0}$. Their key properties are that ψ_0 is strictly increasing with $\psi_0(0+) = -\infty$ and $\psi_0(\infty) = \infty$, while ψ_1 is strictly decreasing and strictly convex with $\psi_1(0+) = \infty$ and $\psi_1(\infty) = 0$.

Fix a compact interval $[\rho_{\min}, \rho_{\max}] \subset (0, \infty)$. Fix a positive constant a_0 and let $\{s_N\}_{N\geqslant 1}$ be a sequence of nonnegative reals such that $0 \leqslant s_N \leqslant a_0(\log N)^{-3}$. Define a set of admissible pairs

$$\mathcal{S}_N = \{(\alpha, \beta) : \alpha, \beta \in [\rho_{\min}, \rho_{\max}],\ -s_N \leqslant \alpha - \beta \leqslant 0\}.$$

For $(\alpha, \beta) \in \mathcal{S}_N$, the mean step satisfies

$$\begin{aligned}\mu_{\alpha,\beta} &= \mathbb{E}[X_1^{\alpha,\beta}] = \mathbb{E}[\log G^\alpha] - \mathbb{E}[\log G^\beta] = \psi_0(\alpha) - \psi_0(\beta)\\ &= \psi_1(\lambda)(\alpha - \beta) \ \in\ [-a_0\psi_1(\rho_{\min})(\log N)^{-3}, 0]\end{aligned} \tag{2.10}$$

where we used the mean value theorem with some $\lambda \in (\rho_{\min}, \rho_{\max})$. We take $D_\mu = a_0\psi_1(\rho_{\min})$.

The MGF of $X_1^{\alpha,\beta}$ is

$$M_{\alpha,\beta}(\theta) = \mathbb{E}\big[e^{\theta X_1^{\alpha,\beta}}\big] = \mathbb{E}\big[(G^\alpha)^\theta\big]\,\mathbb{E}\big[(G^\beta)^{-\theta}\big] = \frac{\Gamma(\alpha+\theta)\Gamma(\beta-\theta)}{\Gamma(\alpha)\Gamma(\beta)} \tag{2.11}$$

for $\theta \in (-\alpha, \beta)$. For the interval in assumption (2.1) we can take $[-\theta_1, \theta_1] = [-\frac{1}{2}\rho_{\min}, \frac{1}{2}\rho_{\min}]$. Now (2.1) holds with a single constant $C_M = C_M(\rho_{\min}, \rho_{\max})$ for all choices of $\alpha, \beta \in [\rho_{\min}, \rho_{\max}]$.

The variance satisfies

$$\mathbb{V}\mathrm{ar}(X_1^{\alpha,\beta}) = \mathbb{V}\mathrm{ar}(\log G^\alpha) + \mathbb{V}\mathrm{ar}(\log G^\beta) = \psi_1(\alpha) + \psi_1(\beta) \geqslant 2\psi_1(\rho_{\max}) = \sigma_*^2.$$

The constants $(C_M, D_\mu, \sigma_*^2, \theta_1)$ have been fixed and they work simultaneously for all $(\alpha, \beta) \in \mathcal{S}_N$ and all $N \geqslant 1$. Define C through (2.5)–(2.7). Choose N_0 so that (2.8) holds for all $N \geqslant N_0$. Now C and N_0 are entirely determined by $(a_0, \rho_{\min}, \rho_{\max})$. We state the result as a corollary of Theorem 2.2.

Corollary 2.8. *In the setting described above the bound below holds for all $N \geqslant N_0$, $(\alpha, \beta) \in \mathcal{S}_N$, and $x \geqslant (\log N)^2$:*

$$\mathbb{P}\Big\{\max_{1\leqslant m\leqslant N} S_m^{\alpha,\beta} \leqslant x\Big\} \leqslant Cx(\log N)(\mu_{\alpha,\beta} \vee N^{-1/2}).$$

3. Comparison with the KMT coupling

As a counterpoint to our Theorem 2.2 we derive here an estimate for a single random walk with the Komlós-Major-Tusnády (KMT) (Komlós et al., 1976) coupling with Brownian motion. We emphasize though that *Theorem 3.1 below is not an alternative to our Theorem 2.2 because we do not know how the constants C, K, λ below depend on the distribution of the walk. Hence without further work we cannot apply the resulting estimate (3.2) to an infinite family of random walks.*

However, this section does illustrate that in a certain regime of vanishing drift the estimates (2.4) and (3.2) are essentially equivalent, as explained below in Remark 3.2. So even if one were to conclude that the constants C, K, λ below can be taken uniform, the result remains the same.

Let $\overline{S}_n = \sum_{k=1}^n \overline{X}_k$ be a mean-zero random walk with i.i.d. steps $\{\overline{X}_k\}$ and unit variance $E[\overline{X}^2] = 1$. The KMT coupling (Theorem 1 in Komlós et al. (1976)) constructs this walk together with a standard Brownian motion $B_\bullet$ on a probability space such that the following bound holds:

$$P\Big(\max_{1\leqslant k\leqslant N} |\overline{S}_k - B_k| \geqslant C\log N + z\Big) \leqslant Ke^{-\lambda z} \qquad \text{for all } N \in \mathbb{Z}_{>0} \text{ and } z > 0, \tag{3.1}$$

where C, K, λ are finite positive constants determined by the distribution of $\overline{X}_k$.

We apply this to the running maximum of a random walk with a negative drift.

Theorem 3.1. *Let $S_n = \sum_{i=1}^n X_i$ be a random walk with i.i.d. steps $\{X_i\}$ that satisfy $E[e^{tX}] < \infty$ for $t \in (-\delta, \delta)$ for some $\delta > 0$. Assume the drift is negative: $\mu = EX_1 < 0$, and the variance $\sigma^2 = E[(X_1 - \mu)^2] > 0$. Then there exists a constant C_1 determined by the distribution of the normalized variable $\overline{X}_1 = \sigma^{-1}(X_1 - \mu)$ such that, for all real $x > 0$ and integers $N > e^4$,*

$$\begin{aligned} P\Big\{\max_{0\leqslant k\leqslant N} S_k < x\Big\} \leqslant C_1\Big(N^{1-(\log N)/2} + \frac{\sigma x + \sigma^2 \log N}{N^{3/2}\mu^2}\, e^{(\sigma^{-1}x + \log N)\sigma^{-1}\mu}\Big) \\ + 1 - e^{2(\sigma^{-1}x + C_1 \log N)\sigma^{-1}\mu}. \end{aligned} \tag{3.2}$$

Remark 3.2. To compare this estimate with Theorem 2.2, imagine that we can let μ vary as a function of N while preserving the constant C_1 in (3.2). Consider the regime where σ^2 is constant, $x > \log N$ and $|\mu|$ vanishes fast enough so that $x|\mu|$ stays bounded. Then the first parenthetical expression on the right of (3.2) is dominated by a constant multiple of $xN^{-3/2}\mu^{-2}$. To the last part apply $1 - e^s \leqslant |s|$ for $s < 0$. The bound (3.2) becomes

$$P\{\max_{0\leqslant k\leqslant N} S_k < x\} \leqslant C_2 x\big(N^{-3/2}\mu^{-2} + |\mu|\big). \tag{3.3}$$

The bound (2.4) is worse than the one above by at most a $\log N$ factor, and not at all if μ vanishes fast enough. In particular, for the application in Busani and Seppäläinen (2020), the KMT bound cannot give anything substantially better than Theorem 2.2.

Proof of Theorem 3.1: Apply (3.1) to the mean-zero unit-variance normalized walk $\overline{S}_N = \sigma^{-1}(S_N - N\mu)$. To simplify some steps below we can assume that $C \geqslant 1 \vee \lambda^{-1}$. Let $x > 0$ and $z = \lambda^{-1}\log N$.

$$\begin{aligned} P\{\max_{0\leqslant k\leqslant N} S_k < x\} &= P\{\max_{0\leqslant k\leqslant N} \big(\overline{S}_k + k\sigma^{-1}\mu\big) < \sigma^{-1}x\} \\ &\leqslant Ke^{-\lambda z} + P\{\max_{0\leqslant k\leqslant N} \big(B_k + k\sigma^{-1}\mu\big) < \sigma^{-1}x + C\log N + z\}. \end{aligned} \tag{3.4}$$

Let $M_k = \sup_{0\leqslant s\leqslant 1}(B_{k+s} - B_k)$. Since $\mu < 0$,

$$\sup_{0\leqslant t\leqslant N} \big(B_t + t\sigma^{-1}\mu\big) \leqslant \max_{0\leqslant k\leqslant N}\big(B_k + k\sigma^{-1}\mu\big) + \max_{0\leqslant k\leqslant N-1} M_k.$$

With this we continue from above.

$$\begin{aligned} \text{line (3.4)} \;\leqslant\; & Ke^{-\lambda z} + P\{\sup_{0\leqslant t\leqslant N}\big(B_t + t\sigma^{-1}\mu\big) < \sigma^{-1}x + 2C\log N + z\} \\ & + P\{\max_{0\leqslant k\leqslant n-1} M_k > C\log N\}. \end{aligned} \tag{3.5}$$

We bound the two probabilities above separately. Recall that $C \geqslant 1$. For the running maximum of standard Brownian motion, by (2.8.4) on page 96 of Karatzas and Shreve (1991),

$$\begin{aligned} P\{\max_{0\leqslant k\leqslant N-1} M_k > \log N\} &\leqslant NP\{\sup_{0\leqslant s\leqslant 1} B_s > \log N\} = N\sqrt{2/\pi}\int_{\log N}^{\infty} e^{-y^2/2}\,dy \\ &\leqslant \frac{N\sqrt{2/\pi}}{\log N}\int_{\log N}^{\infty} ye^{-y^2/2}\,dy = \frac{\sqrt{2/\pi}}{\log N}N^{1-(\log N)/2}. \end{aligned} \tag{3.6}$$

For the running maximum of Brownian motion with drift, use first Brownian scaling, and then the density of the hitting time $T_{b(N)}$ of the point $b(N) = N^{-1/2}(\sigma^{-1}x + 2C\log N + z)$ with drift $\mu(N) = \sigma^{-1}N^{1/2}\mu < 0$ from (3.5.12–3.5.13) on page 197 of Karatzas and Shreve (1991).

$$\begin{aligned} & P\{\sup_{0\leqslant t\leqslant N}\big(B_t + t\sigma^{-1}\mu\big) < \sigma^{-1}x + 2C\log N + z\} \\ &= P\{\sup_{0\leqslant t\leqslant 1}\big(B_t + t\sigma^{-1}N^{1/2}\mu\big) < N^{-1/2}(\sigma^{-1}x + 2C\log N + z)\} \\ &= P\{\sup_{0\leqslant t\leqslant 1}\big(B_t + t\mu(N)\big) < b(N)\} = P^{(\mu(N))}\{T_{b(N)} > 1\} \\ &= b(N)\int_1^{\infty}\frac{1}{\sqrt{2\pi s^3}}e^{-(b(N)-\mu(N)s)^2/2s}\,ds + P^{(\mu(N))}\{T_{b(N)} = \infty\} \\ &= b(N)e^{b(N)\mu(N)}\int_1^{\infty}\frac{1}{\sqrt{2\pi s^3}}e^{-\frac12 b(N)^2 s^{-1} - \frac12\mu(N)^2 s}\,ds + 1 - e^{2b(N)\mu(N)} \\ &\leqslant 2e^{b(N)\mu(N)}\frac{b(N)}{\mu(N)^2} + 1 - e^{2b(N)\mu(N)} \\ &\leqslant 2e^{(\sigma^{-1}x + 3C\log N)\sigma^{-1}\mu}\frac{\sigma x + 3C\sigma^2\log N}{N^{3/2}\mu^2} + 1 - e^{2(\sigma^{-1}x+3C\log N)\sigma^{-1}\mu}. \end{aligned} \tag{3.7}$$

The second last inequality dropped the denominator $2\pi s^3 \geqslant 1$ and the term $-\frac{1}{2}b(N)^2 s^{-1}$ from the exponent, and then integrated. The last inequality substituted in $z = \lambda^{-1}\log N \leqslant C\log N$ to bound

$$N^{-1/2}(\sigma^{-1}x + \log N) \leqslant b(N) \leqslant N^{-1/2}(\sigma^{-1}x + 3C\log N).$$

The conclusion (3.2) follows from substituting into (3.5) the bounds from above. □

4. Auxiliary facts

Before starting the proof proper, we record some simple facts. First, assumptions (2.1) and (2.2) gives these bounds:

$$\begin{aligned}
&0 < \sigma_*^2 \leqslant \mu_{2,N} \equiv \mathbb{E}\big[(X_1^N)^2\big] = M_N^{(2)}(0) \leqslant C_M,\\
&|\mu_{3,N}| \equiv |\mathbb{E}\big[(X_1^N)^3\big]| = |M_N^{(3)}(0)| \leqslant C_M, \qquad (4.1)\\
&\mathbb{P}(X_1^N > t) \leqslant C_M e^{-\theta_1 t}.
\end{aligned}$$

Lemma 4.1. *Let $\{Y_i\}$ be i.i.d. random variables with common marginal distribution ν. Assume that, for two constants $0 < c_1, C_1 < \infty$,*

$$\mathbb{E}(e^{tY_1}) \leqslant C_1 \quad \textit{for} \quad t \in [0, c_1]. \tag{4.2}$$

Then

$$\mu_{\max} = \mu_{\max}(\nu, n) \equiv \mathbb{E}\big[\max\{0, Y_1, ..., Y_n\}\big] \leqslant c_1^{-1}\log(C_1 n + 1).$$

Proof: For $0 < t \leqslant c_1$,

$$e^{t\mu_{\max}} \leqslant \mathbb{E}\big(e^{t(0\vee\max_{1\leqslant i\leqslant n} Y_i)}\big) \leqslant 1 + \mathbb{E}\Big(\sum_{i=1}^n e^{tY_i}\Big) = 1 + n\mathbb{E}\big(e^{tY_1}\big) \leqslant C_1 n + 1,$$

and the claim follows by taking $t = c_1$. □

Since $M_N'' > 0$ there is a unique minimizer

$$\theta_0^N = \arg\min\{M_N(\theta)\}. \tag{4.3}$$

Lemma 4.2. *Let N_0 be such that $C_M|\mu_N| \leqslant \frac{1}{3}\sigma_*^2$ for $N \geqslant N_0$ and set $c_M = 2\sigma_*^{-2}$. Then for $N \geqslant N_0$,*

$$0 \leqslant \theta_0^N \leqslant c_M|\mu_N|.$$

Proof: If $M_N'(0) = \mu_N = 0$ then the minimum is taken at $\theta_0^N = 0$.

So suppose $M_N'(0) = \mu_N < 0$. Expansion for $\theta \in (0, \theta_1)$ gives, with some $\theta' \in (0, \theta)$,

$$M_N'(\theta) = \mu_N + \mu_{2,N}\theta + \tfrac{1}{2}M_N^{(3)}(\theta')\theta^2 \geqslant \mu_N + \mu_{2,N}\theta - \tfrac{1}{2}C_M\theta^2.$$

Since M' is strictly increasing and $c_M = 2/\sigma_*^2 \geqslant 2\mu_{2,N}^{-1}$, by the choice of N_0 we have for $N \geqslant N_0$

$$M_N'(-c_M\mu_N) \geqslant M_N'\Big(-\frac{2\mu_N}{\mu_{2,N}}\Big) \geqslant -\mu_N - 2C_M\frac{\mu_N^2}{\mu_{2,N}^2} \geqslant -\mu_N\big(1 - \tfrac{2}{3}\sigma_*^2/\mu_{2,N}^2\big) > 0.$$

It follows that there exists a unique $\theta_0^N \in (0, c_M|\mu_N|)$ such that $M_N'(\theta_0^N) = 0$. □

Define a tilted measure $Q(d\omega) = f^{\theta_0^N}_{N,n}(\omega)\mathbb{P}(d\omega)$ in terms of the Radon-Nikodym derivative

$$f^{\theta_0^N}_{N,n}(\omega) = \frac{e^{\theta_0^N S_n^N}}{\mathbb{E}\big(e^{\theta_0^N S_n^N}\big)}.$$

Denote the expectation under Q by $\mathbb{E}^Q$. Increase N_0 further so that $N \geqslant N_0$ implies $\theta_0^N \in [-\theta_1/2, \theta_1/2]$ and $-\mu_N \leqslant 2$. Then for $0 \leqslant i \leqslant 3$ and $\theta \in (-\theta_1/2, \theta_1/2)$, the MGF under Q satisfies

$$\begin{aligned} M^{(i)}_{Q,N}(\theta) &= \mathbb{E}^Q\big((X_1^N)^i e^{\theta X_N^1}\big) = M_N(\theta_0^N)^{-1}\mathbb{E}\big((X_1^N)^i e^{(\theta+\theta_0^N)X_N^1}\big) \\ &= M_N(\theta_0^N)^{-1} M_N^{(i)}(\theta_0^N + \theta) \leqslant e^{-\mu_N\theta_0^N} C_M \leqslant e^{\theta_1} C_M, \end{aligned} \tag{4.4}$$

where the first inequality used Jensen's inequality and (2.1). From this we get moment bounds under Q: for $0 \leqslant i \leqslant 3$,

$$\mathbb{E}^Q\big((X_1^N)^i\big) = M^{(i)}_{Q,N}(0) \leqslant e^{\theta_1} C_M. \tag{4.5}$$

For $|\theta| \leqslant \theta_1$, there exists $\theta' \in (-\theta_1, \theta_1)$

$$M_N^{(2)}(\theta) = \mu_{2,N} + M_N^{(3)}(\theta')\theta$$

Increase N_0 further if necessary so that $\theta_0^N \leqslant \frac{\sigma_*^2}{2C_M}$ for $N \geqslant N_0$ and we can write

$$M_N^{(2)}(\theta_0^N) \geqslant \sigma_*^2 - C_M\theta_0^N \geqslant \frac{\sigma_*^2}{2}.$$

Then from $\mathbb{E}^Q(X_1^N) = 0$ and the third equation in (4.4),

$$\mathbb{V}\mathrm{ar}^Q(X_1^N) = \mathbb{E}^Q\big((X_1^N)^2\big) = M^{(2)}_{Q,N}(0) = M_N(\theta_0^N)^{-1} M_N^{(2)}(\theta_0^N) \geqslant C_M^{-1}\frac{\sigma_*^2}{2}. \tag{4.6}$$

5. Proof of the main theorem

To lighten the notation we omit the label N from $\mu = \mu_N$ and $\theta_0 = \theta_0^N$, and from some other notation that obviously depend on N. For $y > 0$ let

$$\tau_y = \inf\{m \geqslant 1 : |S_m^N| \geqslant y\}$$

denote the first hitting time of the cylinder of width $2y$. Let Φ_{σ^2} denote the centered Gaussian distribution with variance σ^2.

Lemma 5.1. *For real $k \geqslant 0$ and $y \geqslant y_0$ we have $\mathbb{P}(\tau_y > ky^2) \leqslant 2e^{-c_\tau k}$, where*

$$y_0 = 1 \vee \frac{6C_M\sigma_*^{-3}}{1 - \Phi_{\sigma_*^2}[-2,2]} \qquad \text{and} \qquad c_\tau = \log\frac{2}{1 + \Phi_{\sigma_*^2}[-2,2]} \in (0, \log 2). \tag{5.1}$$

Proof: Let $\bar{S}_m^N = S_m^N - m\mu$ be the centered walk. Consider an integer $k \geqslant 1$ and a real $y \geqslant 1$. Look at the process along time increments of size $\lfloor y\rfloor^2$:

$$\begin{aligned} \mathbb{P}(\tau_y > ky^2) &\leqslant \mathbb{P}(\tau_y > k\lfloor y\rfloor^2) \leqslant \mathbb{P}(\,|S^N_{m\lfloor y\rfloor^2}| \leqslant y \text{ for } m = 1, \dots, k\,) \\ &\leqslant \mathbb{P}(\,|S^N_{m\lfloor y\rfloor^2} - S^N_{(m-1)\lfloor y\rfloor^2}| \leqslant 2y \text{ for } m = 1, \dots, k\,) \\ &= \big(\mathbb{P}\{\,S^N_{\lfloor y\rfloor^2} \in [-2y, 2y]\}\big)^k = \big(\mathbb{P}\{\,\lfloor y\rfloor^{-1}\overline{S}^N_{\lfloor y\rfloor^2} \in [-2 - \mu\lfloor y\rfloor, 2 - \mu\lfloor y\rfloor]\}\big)^k \\ &\leqslant \Big(\Phi_{\sigma_N^2}[-2 - \mu\lfloor y\rfloor, 2 - \mu\lfloor y\rfloor] + 3\frac{\mu_{3,N}}{\sigma^3}\lfloor y\rfloor^{-1}\Big)^k \leqslant \big(\Phi_{\sigma_N^2}[-2,2] + 6C_M\sigma_*^{-3}y^{-1}\big)^k. \end{aligned}$$

The penultimate inequality is the Berry-Esseen Theorem. We use the version from Durrett (2010, Section 3.4.4) where the constant is 3. The last inequality is a simple property of a centered Gaussian. Now for $y \geqslant y_0$ and c_τ as above,

$$\begin{aligned}\sup_{\sigma_*^2 \leqslant \sigma^2 \leqslant C_M} \Phi_{\sigma^2}[-2,2] + 3C_M\sigma_*^{-3}y^{-1} &\leqslant \Phi_{\sigma_*^2}[-2,2] + 3C_M\sigma_*^{-3}y_0^{-1} \\ &\leqslant \tfrac{1}{2}(1+\Phi_{\sigma_*^2}[-2,2]) = e^{-c_\tau}.\end{aligned}$$

We have proved $\mathbb{P}(\tau_y > ky^2) \leqslant e^{-c_\tau k}$ for $k \in \mathbb{Z}_{\geqslant 0}$. Extend this to real $k \in \mathbb{R}_{\geqslant 0}$:

$$\mathbb{P}(\tau_y > ky^2) \leqslant \mathbb{P}(\tau_y > \lfloor k \rfloor y^2) \leqslant e^{-c_\tau \lfloor k \rfloor} \leqslant e^{-c_\tau(k-1)} = 2(1+\Phi_{\sigma_*^2}[-2,2])^{-1}e^{-c_\tau k} < 2e^{-c_\tau k}. \qquad \square$$

Let $H_N = \mu^{-2} \wedge N$. By (2.3)

$$D_\mu^{-2}(\log N)^6 \leqslant H_N \leqslant N. \tag{5.2}$$

Define the truncated version of τ_y

$$\hat{\tau}_y = \tau_y \wedge H_N.$$

The following result shows that although the random walk S_m^N has negative drift, up to times of order H_N it behaves similarly to an unbiased random walk in the following sense: if $y > 0$ is not too small, but small compared to $H_N^{1/2}$, the probability that the random walk reaches level y before level $-y$ is close to $1/2$. Our choice of H_N can be justified by decomposing the random walk into

$$S_n^N = \sum_{i=1}^{n}\left(X_i^N - \mu\right) + n\mu.$$

For $\varepsilon > 0$ small and $|\mu| \geqslant N^{-1/2}$ (so that $H_N = \mu^{-2}$),

$$(\varepsilon H_N)^{-1/2} S_{\varepsilon H_N}^N = (\varepsilon H_N)^{-1/2}\sum_{i=1}^{\varepsilon H_N}\left(X_i^N - \mu\right) + \varepsilon^{1/2}. \tag{5.3}$$

As

$$(\varepsilon H_N)^{-1/2}\sum_{i=1}^{\varepsilon H_N}\left(X_i^N - \mu\right) \overset{d}{\approx} N(0,\sigma),$$

we see that the left hand side of (5.3) is dominated by the first term on the right hand side. That is, up to time εH_N the random walk S^N behaves approximately like an unbiased random walk.

Lemma 5.2. *Let y_0 be as in (5.1). There exist finite constants N_0 and C_0 such that, for $N \geqslant N_0$ and $y_0 \leqslant y \leqslant (\log N)^{-1}H_N^{1/2}$,*

$$\mathbb{P}(S_{\hat{\tau}_y} \geqslant y) \geqslant \tfrac{1}{2}\Big[1 - C_0 H_N^{-\frac{1}{2}}\big(y + (\log H_N)^2\big) - \frac{2}{\theta_1 y}\log(e^{\theta_1}C_M H_N + 1)\Big].$$

C_0 depends on θ_1, σ_*^2 and C_M while N_0 depends on θ_1, σ_*^2, D_μ and C_M.

Proof: The constant C_0 comes as follows in terms of the constants previously introduced above and new constants C_2, C_3, C_4 introduced below in the course of the proof:

$$\begin{aligned}C_0 = C_2 + C_4 &= 2(\sigma_*^{-1} + 9e^{\theta_1}\sigma_*^{-3}C_M^{5/2}) + 2C_3 c_\tau^{-1} + 6c_M \\ &= 2(\sigma_*^{-1} + 9e^{\theta_1}\sigma_*^{-3}C_M^{5/2}) + 2e^{2C_M c_M^2 + 4c_M}c_\tau^{-1} + 6c_M.\end{aligned} \tag{5.4}$$

Under the measure Q, S_n is a mean-zero random walk and hence a martingale. Furthermore, $\hat\tau_y$ is a bounded stopping time. From this,

$$0 = \int S_{\hat\tau_y} dQ = \int_{S_{\hat\tau_y}\geqslant y} S_{\hat\tau_y} dQ + \int_{S_{\hat\tau_y}\leqslant -y} S_{\hat\tau_y} dQ + \int_{S_{\hat\tau_y}\in(-y,y)} S_{\hat\tau_y} dQ.$$

On the event $S_{\hat\tau_y} \geqslant y$, we have $\hat\tau_y = \tau_y$ and $S_{\hat\tau_y-1} < y \leqslant S_{\hat\tau_y} = S_{\hat\tau_y-1} + X^N_{\hat\tau_y} \leqslant y + X^N_{\hat\tau_y}$ and so

$$\begin{aligned}\int_{S_{\hat\tau_y}\geqslant y} S_{\hat\tau_y}\, dQ &\leqslant \int_{S_{\hat\tau_y}\geqslant y} \bigl(y + X^N_{\hat\tau_y}\bigr)\, dQ \leqslant \int_{S_{\hat\tau_y}\geqslant y} \bigl(y + 0 \vee \max_{1\leqslant i\leqslant H_N} X^N_i\bigr)\, dQ \\ &\leqslant yQ(S_{\hat\tau_y} \geqslant y) + \mu_{\max}(Q, H_N) \;\leqslant\; yQ(S_{\hat\tau_y} \geqslant y) + 2\theta_1^{-1}\log(e^{\theta_1}C_M H_N + 1),\end{aligned}$$

where we applied Lemma 4.1 under the distribution Q with $C_1 = e^{\theta_1}C_M, c_1 = \frac{1}{2}\theta_1$ from (4.4). Combine the displays above to obtain

$$\begin{aligned}Q(S_{\hat\tau_y} \geqslant y) &\geqslant -y^{-1}\int\limits_{S_{\hat\tau_y}\leqslant -y} S_{\hat\tau_y} dQ \;-\; y^{-1}\int\limits_{S_{\hat\tau_y}\in(-y,y)} S_{\hat\tau_y} dQ \;-\; 2\theta_1^{-1}y^{-1}\log(e^{\theta_1}C_M H_N + 1) \\ &\geqslant Q(S_{\hat\tau_y} \leqslant -y) - Q(S_{\hat\tau_y} \in (-y,y)) - 2\theta_1^{-1}y^{-1}\log(e^{\theta_1}C_M H_N + 1).\end{aligned}$$

Use

$$Q(S_{\hat\tau_y} \leqslant -y) = 1 - Q(S_{\hat\tau_y} \geqslant y) - Q(S_{\hat\tau_y} \in (-y,y))$$

to rewrite the above as

$$Q(S_{\hat\tau_y} \geqslant y) \geqslant \tfrac{1}{2}[1 - 2Q(S_{\hat\tau_y} \in (-y,y)) - 2\theta_1^{-1}y^{-1}\log(e^{\theta_1}C_M H_N + 1)]. \tag{5.5}$$

It remains to bound the probability on the right. $S_{\hat\tau_y} \in (-y,y)$ forces $\hat\tau_y = H_N$ and thereby another application of the Berry-Esseen theorem, while using (4.5), (4.6) and $y \geqslant y_0 \geqslant 1$, gives

$$\begin{aligned}Q\bigl\{S_{\hat\tau_y} \in (-y,y)\bigr\} &= Q\bigl\{H_N^{-1/2}S_{H_N} \in (-H_N^{-1/2}y, H_N^{-1/2}y)\bigr\} \\ &\leqslant \Phi_{\sigma_*^2}(-H_N^{-1/2}y, H_N^{-1/2}y) + 3\frac{e^{\theta_1}C_M}{2^{-3/2}C_M^{-3/2}\sigma_*^3}H_N^{-\frac{1}{2}} \\ &\leqslant 2(2\pi\sigma_*^2)^{-1/2}yH_N^{-1/2} + 9\frac{e^{\theta_1}C_M}{C_M^{-3/2}\sigma_*^3}H_N^{-\frac{1}{2}} \leqslant (\sigma_*^{-1} + 9e^{\theta_1}\sigma_*^{-3}C_M^{5/2})yH_N^{-\frac{1}{2}} \\ &\equiv \tfrac{1}{2}C_2 yH_N^{-\frac{1}{2}}.\end{aligned}$$

Rewrite (5.5) as

$$Q(S_{\hat\tau_y} \geqslant y) \geqslant \tfrac{1}{2}[1 - C_2 yH_N^{-\frac{1}{2}} - 2y^{-1}\theta_1^{-1}\log(e^{\theta_1}C_M H_N + 1)]. \tag{5.6}$$

It remains to switch from Q back to the original distribution $\mathbb{P}$. Recall the Radon-Nikodym derivative $f_n^\theta = M(\theta)^{-n}e^{\theta S_n}$. Introduce a temporary quantity $G_0 > 1$ to be chosen precisely below. Decompose according to the value of $\hat\tau_y$ and use Cauchy-Schwarz:

$$\begin{aligned}Q(S_{\hat\tau_y} \geqslant y) &= \mathbb{E}\bigl[f^{\theta_0}_{\hat\tau_y}(\mathbf{1}_{S_{\hat\tau_y}\geqslant y, \hat\tau_y\leqslant G_0} + \mathbf{1}_{S_{\hat\tau_y}\geqslant y, \hat\tau_y > G_0})\bigr] \\ &\leqslant \mathbb{E}\bigl[f^{\theta_0}_{\hat\tau_y}\mathbf{1}_{S_{\hat\tau_y}\geqslant y, \hat\tau_y\leqslant G_0}\bigr] + \Bigl(\mathbb{E}\bigl[(f^{\theta_0}_{\hat\tau_y})^2\bigr]\Bigr)^{\frac{1}{2}}\Bigl(\mathbb{P}\{S_{\hat\tau_y} \geqslant y, \hat\tau_y > G_0\}\Bigr)^{\frac{1}{2}}.\end{aligned} \tag{5.7}$$

Let us first bound the second term on line (5.7). Note that f_n^θ is a $\mathbb{P}$-martingale and $\hat\tau_y$ is a stopping time bounded by H_N. Hence $(f_n^\theta)^2$ is a submartingale and we have

$$\begin{aligned}&\Big(\mathbb{E}[(f^{\theta_0}_{\hat\tau_y})^2]\Big)^{\frac12}\Big(\mathbb{P}\{S_{\hat\tau_y}\geqslant y,\hat\tau_y>G_0\}\Big)^{\frac12}\\ &\leqslant\Big(\mathbb{E}[(f^{\theta_0}_{H_N})^2]\Big)^{\frac12}\Big(\mathbb{P}\{S_{\hat\tau_y}\geqslant y,\hat\tau_y>G_0\}\Big)^{\frac12}=\Big(\frac{M(2\theta_0)}{M(\theta_0)^2}\Big)^{H_N/2}\Big(\mathbb{P}\{S_{\hat\tau_y}\geqslant y,\hat\tau_y>G_0\}\Big)^{\frac12}.\end{aligned}$$

To bound the M-factor on the right, expand M and use (2.1), (4.1) and $\mu<0$. In the numerator, for some $\eta\in(0,2\theta_0)$,

$$M(2\theta_0)=1+\mu 2\theta_0+2\mu_2\theta_0^2+\tfrac86 M^{(3)}(\eta)\theta_0^3\leqslant 1+2\mu_2\theta_0^2+\tfrac43 C_M\theta_0^3$$

and similarly in the denominator:

$$\begin{aligned}&\Big[M(2\theta_0)M(\theta_0)^{-2}\Big]^{H_N/2}\leqslant\Big(1+2\mu_2\theta_0^2+\tfrac43 C_M\theta_0^3\Big)^{H_N/2}\Big(1+\mu\theta_0+\tfrac12\mu_2\theta_0^2-\tfrac16 C_M\theta_0^3\Big)^{-H_N}\\ &\leqslant\Big(1+2C_Mc_M^2\mu^2+\tfrac43 C_Mc_M^3|\mu|^3\Big)^{\frac12\mu^{-2}}\Big(1-c_M\mu^2-C_Mc_M^3|\mu|^3\Big)^{-\mu^{-2}}\\ &\leqslant e^{2C_Mc_M^2+4c_M}\equiv C_3.\end{aligned}\tag{5.8}$$

Above we used $H_N\leqslant\mu^{-2}$ and increased N_0 once more so that $N\geqslant N_0$ guarantees $\frac23 c_M|\mu|\leqslant 1$, $c_M\mu^2+C_Mc_M^3|\mu|^3\leqslant\frac12$ and $C_Mc_M^3|\mu|\leqslant 1$. Then we applied the bounds

$$\begin{aligned}&\Big(1+2C_Mc_M^2\mu^2+\tfrac43 C_Mc_M^3|\mu|^3\Big)^{\frac12\mu^{-2}}\leqslant e^{C_Mc_M^2(1+\frac23 c_M|\mu|)}\leqslant e^{2C_Mc_M^2},\\ &\Big(1-c_M\mu^2-C_Mc_M^3|\mu|^3\Big)^{-\mu^{-2}}\leqslant\Big(1+2c_M\mu^2\big(1+C_Mc_M^3|\mu|\big)\Big)^{\mu^{-2}}\leqslant e^{2c_M(1+C_Mc_M^3|\mu|)}\leqslant e^{4c_M},\end{aligned}$$

where the second line also used $(1-a)^{-1}\leqslant 1+2a$ for $a\in[0,\frac12]$. Put (5.8) back up, set $G_0=yH_N^{1/2}$, and apply Lemma 5.1 (for which we use the assumption $y\geqslant y_0$):

$$\Big(\mathbb{E}[(f^{\theta_0}_{\hat\tau_y})^2]\Big)^{\frac12}\Big(\mathbb{P}\{S_{\hat\tau_y}\geqslant y,\hat\tau_y>G_0\}\Big)^{\frac12}\leqslant C_3\big(\mathbb{P}\{\hat\tau_y>G_0\}\big)^{\frac12}\leqslant 2C_3e^{-c_\tau H_N^{1/2}y^{-1}}.\tag{5.9}$$

Next we bound the first term on line (5.7). Use $M(\theta_0)\leqslant 1$. Let $\mathcal{M}_n=\max_{1\leqslant i\leqslant n}X_i^N$.

$$\begin{aligned}&\mathbb{E}\big[f^{\theta_0}_{\hat\tau_y}\mathbf{1}_{S_{\hat\tau_y}\geqslant y,\hat\tau_y\leqslant G_0}\big]=\mathbb{E}\Big[\frac{e^{\theta_0S_{\hat\tau_y}}}{M(\theta_0)^{\hat\tau_y}}\mathbf{1}_{S_{\hat\tau_y}\geqslant y,\hat\tau_y\leqslant G_0}\Big]\\ &\leqslant\mathbb{E}\Big[\frac{e^{\theta_0S_{\hat\tau_y}}}{M(\theta_0)^{G_0}}\mathbf{1}_{S_{\hat\tau_y}\geqslant y,\mathcal{M}_{H_N}\leqslant(\log H_N)^2}\Big]+\mathbb{E}\Big[\frac{e^{\theta_0S_{\hat\tau_y}}}{M(\theta_0)^{\hat\tau_y}}\mathbf{1}_{S_{\hat\tau_y}\geqslant y,\mathcal{M}_{H_N}>(\log H_N)^2}\Big]\\ &\leqslant\mathbb{E}\Big[\frac{e^{\theta_0(y+\mathcal{M}_{H_N})}}{M(\theta_0)^{G_0}}\mathbf{1}_{S_{\hat\tau_y}\geqslant y,\mathcal{M}_{H_N}\leqslant(\log H_N)^2}\Big]+\mathbb{E}\Big[\frac{e^{\theta_0S_{\hat\tau_y}}}{M(\theta_0)^{\hat\tau_y}}\mathbf{1}_{S_{\hat\tau_y}\geqslant y,\mathcal{M}_{H_N}>(\log H_N)^2}\Big]\end{aligned}\tag{5.10}$$

Let us first bound the second term. Using Cauchy-Schwarz, the bound

$$\mathbb{P}(\mathcal{M}_{H_N}>t)\leqslant H_NC_Me^{-\theta_1t},$$

the bound (5.8), and the tail bound in (4.1), it follows that

$$\begin{aligned}\mathbb{E}\Big[\frac{e^{\theta_0S_{\hat\tau_y}}}{M(\theta_0)^{\hat\tau_y}}\mathbf{1}_{S_{\hat\tau_y}\geqslant y,\mathcal{M}_{H_N}>(\log H_N)^2}\Big]&\leqslant\Big(\mathbb{E}[(f^{\theta_0}_{H_N})^2]\Big)^{\frac12}\Big(\mathbb{P}\{\mathcal{M}_{H_N}>(\log H_N)^2\}\Big)^{\frac12}\\ &\leqslant C_3H_N^{1/2}C_M^{1/2}e^{-\frac12\theta_1(\log H_N)^2}.\end{aligned}$$

The first term on the last line of (5.10) is bounded as follows, with $G_0 = yH_N^{1/2}$.

$$\begin{aligned}
&\mathbb{E}\Big[\frac{e^{\theta_0(y+\mathcal{M}_{H_N})}}{M(\theta_0)^{G_0}}\mathbf{1}_{S_{\hat\tau_y}\geqslant y,\mathcal{M}_{H_N}\leqslant(\log H_N)^2}\Big] \leqslant \mathbb{E}\Big[\frac{e^{\theta_0(y+(\log H_N)^2)}}{M(\theta_0)^{G_0}}\mathbf{1}_{S_{\hat\tau_y}\geqslant y}\Big]\\
&\leqslant \mathbb{P}(S_{\hat\tau_y}\geqslant y)e^{c_M H_N^{-1/2}[y+(\log H_N)^2]}M(\theta_0)^{-yH_N^{1/2}} \leqslant \mathbb{P}(S_{\hat\tau_y}\geqslant y)e^{c_M H_N^{-1/2}[y+(\log H_N)^2]}e^{c_M y H_N^{-1/2}}\\
&= \mathbb{P}(S_{\hat\tau_y}\geqslant y)e^{c_M H_N^{-1/2}[2y+(\log H_N)^2]}\\
&\leqslant \mathbb{P}(S_{\hat\tau_y}\geqslant y)\big[1+2c_M H_N^{-1/2}[2y+(\log H_N)^2]\big] \leqslant \mathbb{P}(S_{\hat\tau_y}\geqslant y)+2c_M H_N^{-1/2}[2y+(\log H_N)^2].
\end{aligned}$$

We used above Jensen's inequality in the form $M(\theta_0)^{-yH_N^{1/2}} \leqslant e^{-\theta_0\mu y H_N^{1/2}}$, the definition of H_N in the form $|\mu|H_N^{1/2} \leqslant 1$, and then $\theta_0 \leqslant c_M|\mu| \leqslant c_M H^{-1/2}$. Furthermore, by (5.2) and our assumption $y \leqslant (\log N)^{-1}H_N^{1/2}$ we have

$$c_M H_N^{-1/2}[2y+(\log H_N)^2] \leqslant c_M(2+D_\mu)(\log N)^{-1} \leqslant \log 2$$

where we choose N_0 large enough so that the last inequality holds for $N \geqslant N_0$. Then we applied the inequality $e^x \leqslant 1+2x$ for $x \in [0,\log 2]$.

Going back to (5.10), for $N \geqslant N_0$,

$$\begin{aligned}
E\big[f_{\hat\tau_y}^{\theta_0}\mathbf{1}_{S_{\hat\tau_y}\geqslant y,\hat\tau_y\leqslant G_0}\big] &\leqslant \mathbb{P}(S_{\hat\tau_y}\geqslant y)+2c_M H_N^{-1/2}[2y+(\log H_N)^2]+C_3H_N^{1/2}C_M^{1/2}e^{-\frac12\theta_1(\log H_N)^2}\\
&\leqslant \mathbb{P}(S_{\hat\tau_y}\geqslant y)+3c_M H_N^{-1/2}[2y+(\log H_N)^2].
\end{aligned}$$

The second inequality is guaranteed for example by choosing N_0 large enough so that $N \geqslant N_0$ implies

$$D_\mu^{-2}(\log N)^6 \geqslant e^{\theta_1^{-1}} \qquad\text{and}\qquad c_M C_3^{-1}C_M^{-1/2}\big(\log[D_\mu^{-2}(\log N)^6]\big)^2 \geqslant e^{\frac12\theta_1^{-1}}.$$

This works due to the lower bound (5.2) on H_N and because the function $f(x) = xe^{-\frac12\theta_1(\log x)^2}$ achieves its maximum $e^{\frac12\theta_1^{-1}}$ at $x = e^{\theta_1^{-1}}$ after which it decreases.

Combine the above with (5.9) on line (5.7) to get this upper bound:

$$\begin{aligned}
Q(S_{\hat\tau_y}\geqslant y) &\leqslant \mathbb{P}(S_{\hat\tau_y}\geqslant y)+2C_3e^{-c_\tau H_N^{1/2}y^{-1}}+3c_M H_N^{-1/2}[2y+(\log H_N)^2]\\
&\leqslant \mathbb{P}(S_{\hat\tau_y}\geqslant y)+2C_3c_\tau^{-1}H_N^{-1/2}y+3c_M H_N^{-1/2}[2y+(\log H_N)^2] \qquad (5.11)\\
&\leqslant \mathbb{P}(S_{\hat\tau_y}\geqslant y)+C_4H_N^{-1/2}[y+(\log H_N)^2]
\end{aligned}$$

where $C_4 = 2C_3c_\tau^{-1}+6c_M$. The second inequality above came from $xe^{-x} \leqslant e^{-1}$ for $x \geqslant 0$. Put (5.11) and (5.6) together to obtain the claim of the lemma. ☐

By adjusting a constant we can replace $\hat\tau_y$ with τ_y in the previous estimate.

Corollary 5.3. *Under the assumptions of Lemma 5.2, with* $C_{10} = C_0+2c_\tau^{-1}$,

$$\mathbb{P}(S_{\tau_y}\geqslant y) \geqslant \tfrac12\Big[1-C_{10}H_N^{-\frac12}\big(y+(\log H_N)^2\big)-\frac{2}{\theta_1 y}\log(e^{\theta_1}C_MH_N+1)\Big] \qquad (5.12)$$

Proof: The assumption $y_0 \leqslant y \leqslant H_N^{1/2}$ implies that Lemma 5.1 applies to give

$$\mathbb{P}(\tau_y > H_N) \leqslant e^{-c_\tau H_N y^{-2}} \leqslant e^{-c_\tau H_N^{1/2}y^{-1}} \leqslant c_\tau^{-1}H_N^{-1/2}y. \qquad (5.13)$$

The claim then comes from Lemma 5.2 and $\mathbb{P}(S_{\tau_y}\geqslant y) \geqslant \mathbb{P}(S_{\hat\tau_y}\geqslant y)-\mathbb{P}(\tau_y > H_N)$. ☐

For $w > 0$ truncate:

$$\widehat X_i^{N,w} = X_i^N\mathbf{1}_{\{X_i^N\geqslant -w\}}-w\mathbf{1}_{\{X_i^N<-w\}} \qquad\text{and}\qquad \widehat S_n^{N,w} = \sum_{i=1}^n \widehat X_i^{N,w}.$$

Define

$$t_y = \inf\{m \geqslant 1 : |\widehat{S}_m^{N,w}| \geqslant y\}.$$

We transfer bound (5.12) to the truncated walk $\widehat{S}$. The reason is that the proof of the forthcoming Lemma 5.5 is easier for the truncated RW.

Corollary 5.4. *Under the assumptions of Lemma 5.2, with $C_{11} = C_0 + 4c_\tau^{-1}$,*

$$\mathbb{P}(\widehat{S}_{t_y}^{N,w} \geqslant y) \geqslant \tfrac{1}{2}\Big[1 - C_{11}H_N^{-\frac{1}{2}}\big(y + (\log H_N)^2\big) - \frac{2}{\theta_1 y}\log(e^{\theta_1}C_M H_N + 1) - H_N C_M e^{-\theta_1 w}\Big].$$

Proof: Note that

$$\begin{aligned}&\mathbb{P}\big(\widehat{S}_m^{N,w} \neq S_m^N \quad \text{for some } 1 \leqslant m \leqslant H_N\big) = \mathbb{P}\big(\widehat{X}_i^{N,w} \neq X_i^N \quad \text{for some } 1 \leqslant i \leqslant H_N\big) \\ &= \mathbb{P}\big(\inf_{1\leqslant i\leqslant H_N} X_i^N < -w\big) \leqslant H_N C_M e^{-\theta_1 w}.\end{aligned} \tag{5.14}$$

Moreover,

$$\begin{aligned}\mathbb{P}(\widehat{S}_{t_y}^{N,w} \geqslant y) &\geqslant \mathbb{P}(S_{\tau_y}^N \geqslant y,\ \tau_y \leqslant H_N,\ \widehat{S}_m^{N,w} = S_m^N \quad \text{for all } 1 \leqslant m \leqslant H_N) \\ &\geqslant \mathbb{P}(S_{\tau_y}^N \geqslant y) - \mathbb{P}(\tau_y > H_N) - \mathbb{P}(\widehat{S}_m^{N,w} \neq S_m^N \quad \text{for some } 1 \leqslant m \leqslant H_N) \\ &\geqslant \mathbb{P}(S_{\tau_y}^N \geqslant y) - c_\tau^{-1}H_N^{-1/2}y - H_N C_M e^{-\theta_1 w},\end{aligned} \tag{5.15}$$

where we used (5.14) and (5.13). Combine the above with (5.12) to obtain the result. ☐

We turn to the main argument of the proof of Theorem 2.2, that is, to show that the probability of the random walk $\widehat{S}_m$ to hit the level x before hitting the level $-\varepsilon H^{1/2}$ is close to $x|\mu|$. This gives rise to the error term in (2.4). We sketch the reasoning.

Let us try to hit the level $x > 0$ starting from the origin. By Corollary 5.4 there is a probability $\approx 1/2$ to hit x before hitting $-x$. Suppose we failed and hit $-x$ first. We have another chance to hit x by going $2x$ upward from the level $-x$. By Corollary 5.4 the probability of going $2x$ up to the level x before going $2x$ down to the level $-4x$ is $\approx 1/2$. We continue this way until we either hit the level x or the level $-\varepsilon H_N^{1/2}$. How many trials to hit x do we have before we hit $-\varepsilon H_N^{1/2}$? Approximately $K = \log_2(x^{-1}\varepsilon H_N^{1/2})$. The trials are independent and so the probability of hitting the level $-\varepsilon H_N^{1/2}$ before hitting the level x is $\approx 2^{-K} = Cx|\mu|$, which is what we seek.

We introduce the notation to make the sketch precise. See Figure 5.1 for an illustration.

Define $K = \lfloor \log_2(x^{-1}(\log N)^{-1}H_N^{1/2})\rfloor - 2$. For $i \geqslant 0$ set $L_i = 2^{i+2} - 3$. Inductively these satisfy $L_0 = 1$ and $L_i = 2L_{i-1} + 3$. Furthermore,

$$xL_K \leqslant (\log N)^{-1}H_N^{1/2}.$$

Define the stopping times

$$T_0 = \inf\{n : |\widehat{S}_n^{N,x}| \geqslant xL_0\}$$
$$\text{and} \qquad T_i = \inf\{n \geqslant T_{i-1} : \widehat{S}_n^{N,x} \leqslant -xL_i \text{ or } \widehat{S}_n^{N,x} \geqslant x\}.$$

Note that $T_i = T_{i-1}$ is possible.

Lemma 5.5. *There exist finite constants C_{12} and N_0 such that for $N \geqslant N_0$ and $x \geqslant (\log N)^2$,*

$$\mathbb{P}\big(\max_{1\leqslant m\leqslant T_K} \widehat{S}_m^{N,x} < x\big) \leqslant C_{12}\,x(\log N)H_N^{-1/2},$$

where

$$C_{12} = 4\exp\big\{4(C_0 + 4c_\tau^{-1})(1 + D_\mu) + 8\theta_1^{-1}\big(1 + \log(e^{\theta_1}C_M + 1)\big) + 4C_M\big\}$$

and C_0 in the expression above is from (5.4).

Proof: Since $C_0 \geqslant 2$, we have $C_{12} \geqslant 4e^8 \geqslant 2^{10}$. Then we can assume that $x \leqslant 2^{-10}(\log N)^{-1}H_N^{1/2}$, for otherwise the bound on the probability is > 1. This guarantees that $K \geqslant 8$. It also implies that unless $|\mu| \leqslant 2^{-10}(\log N)^{-3}$, the result is trivial.

Since $\widehat{X}_i^{N,x} \geqslant -x$,

$$\{\widehat{S}_{T_i}^{N,x} \leqslant -xL_i\} = \{-x(L_i+1) < \widehat{S}_{T_i}^{N,x} \leqslant -xL_i\} \tag{5.16}$$

and

$$\{\widehat{S}_{T_i}^{N,x} \leqslant -xL_i, \widehat{S}_{T_{i+1}}^{N,x} \leqslant -xL_{i+1}\} \subseteq \{T_i < T_{i+1}\}.$$

Note that

$$E \equiv \Big\{ \max_{1 \leqslant m \leqslant T_K} \widehat{S}_m^{N,x} < x \Big\} \subseteq \bigcap_{1 \leqslant i \leqslant K} \{\widehat{S}_{T_i}^{N,x} \leqslant -xL_i\}. \tag{5.17}$$

Due to (5.16)

$$\begin{aligned}
&\mathbb{P}\big(\widehat{S}_{T_0}^{N,x} \leqslant -xL_0, ..., \widehat{S}_{T_{i-1}}^{N,x} \leqslant -xL_{i-1}, \widehat{S}_{T_i}^{N,x} \leqslant -xL_i\big) \\
&= \mathbb{P}\big(\widehat{S}_{T_i}^{N,x} \leqslant -xL_i | \widehat{S}_{T_0}^{N,x} \leqslant -xL_0, ..., \widehat{S}_{T_{i-1}}^{N,x} \leqslant -xL_{i-1}\big) \\
&\qquad \cdot \mathbb{P}\big(\widehat{S}_{T_0}^{N,x} \leqslant -xL_0, ..., \widehat{S}_{T_{i-1}}^{N,x} \leqslant -xL_{i-1}\big) \\
&\leqslant \mathbb{P}\big(\widehat{S}_{T_i}^{N,x} \leqslant -xL_i | \widehat{S}_{T_0}^{N,x} \leqslant -xL_0, ..., \widehat{S}_{T_{i-1}}^{N,x} = -x(L_{i-1}+1)\big) \\
&\qquad \cdot \mathbb{P}\big(\widehat{S}_{T_0}^{N,x} \leqslant -xL_0, ..., \widehat{S}_{T_{i-1}}^{N,x} \leqslant -xL_{i-1}\big) \\
&= \mathbb{P}\big(\widehat{S}_{t_{x(L_{i-1}+2)}}^{N,x} \leqslant -x(L_{i-1}+2)\big)\mathbb{P}\big(\widehat{S}_{T_0}^{N,x} \leqslant -xL_0, ..., \widehat{S}_{T_{i-1}}^{N,x} \leqslant -xL_{i-1}\big).
\end{aligned} \tag{5.18}$$

The last equality used the definition of the stopping time t_y, the definition of L_i, and the Markov property. For $1 \leqslant i \leqslant K$ define the events

$$A_i^N = \{\widehat{S}_{t_{x(L_{i-1}+2)}}^{N,x} \leqslant -x(L_{i-1}+2)\}.$$

Applying (5.18) to (5.17) repeatedly,

$$\mathbb{P}(E) \leqslant \mathbb{P}\Big(\bigcap_{1 \leqslant i \leqslant K} \{\widehat{S}_{T_i}^{N,x} \leqslant -xL_i\} \Big) \leqslant \prod_{1 \leqslant i \leqslant K} \mathbb{P}(A_i^N).$$

Let $x \geqslant (\log N)^2$. Recall that by (5.2), $(\log H_N)^2 H_N^{-1/2} \leqslant D_\mu (\log N)^{-1}$ and $H_N \leqslant N$. Apply Corollary 5.4 with $w = x$ and $y_i = x(L_{i-1}+2) \in [x, (\log N)^{-1}H_N^{1/2}]$ for $i = 1, \ldots, K$ and $N \geqslant N_0$ to get this estimate:

$$\begin{aligned}
\mathbb{P}(A_i^N) &\leqslant \tfrac{1}{2}\Big[1 + C_{11}H_N^{-\frac{1}{2}}\big(y_i + (\log H_N)^2\big) + \frac{2}{\theta_1 y_i}\log(e^{\theta_1}C_M H_N + 1) + H_N C_M e^{-\theta_1 x}\Big] \\
&\leqslant \tfrac{1}{2}\Big[1 + C_{11}(1+D_\mu)(\log N)^{-1} + \frac{2\log(e^{\theta_1}C_M N + 1)}{\theta_1(\log N)^2} + C_M N^{1-\theta_1(\log N)^2}\Big] \\
&\leqslant \tfrac{1}{2}\big[1 + C_A(\log N)^{-1}\big],
\end{aligned}$$

where we set

$$C_A = C_{11}(1+D_\mu) + 2\theta_1^{-1}\big(1 + \log(e^{\theta_1}C_M + 1)\big) + C_M$$

and if necessary we increase N_0 further so that $N^{1-\theta_1(\log N)^2} \leqslant (\log N)^{-1}$ for $N \geqslant N_0$.

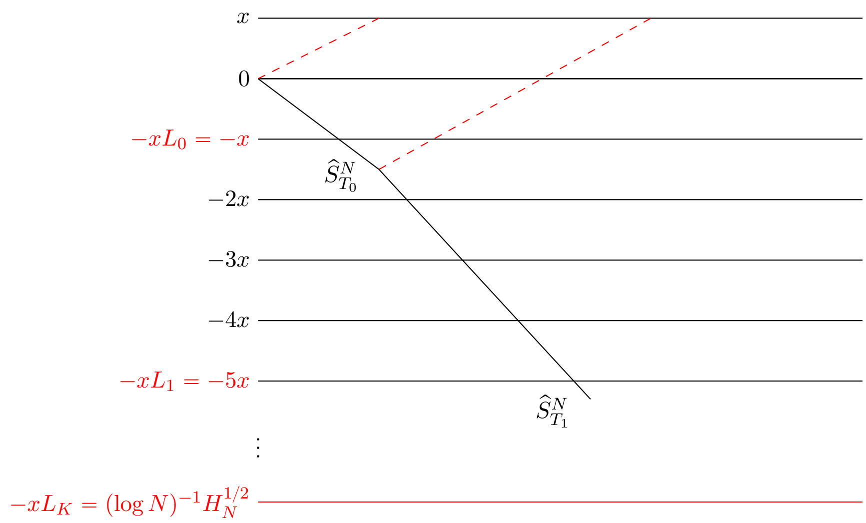

FIGURE 5.1. By the time the random walk $\widehat{S}^N$ exits the cylinder of radius $(\log N)^{-1}H_N^{1/2}$ it has had about K independent opportunities to hit the level x, each with probability close to $1/2$.

Continue with the above estimate,

$$\begin{aligned}
\mathbb{P}(E) &\leqslant \prod_{i=1}^{K}\mathbb{P}(A_i^N) \leqslant \Big(\tfrac{1}{2}\big[1 + C_A(\log N)^{-1}\big]\Big)^K \\
&= \Big(\tfrac{1}{2}\big[1 + C_A(\log N)^{-1}\big]\Big)^{\lfloor \log_2(x^{-1}(\log N)^{-1}H_N^{1/2})\rfloor - 2} \\
&\leqslant 4x(\log N)H_N^{-1/2}\big[1 + C_A(\log N)^{-1}\big]^{\log_2 N} \\
&\leqslant 4e^{4C_A}x(\log N)H_N^{-1/2} = 4e^{4C_A}x(\log N)(|\mu| \vee N^{-1/2}),
\end{aligned}$$

where we used $\log_2 N = \frac{\log N}{\log 2} \leqslant 4\log N$. □

We are ready to prove Theorem 2.2. By Lemma 5.5, by the time $\widehat{S}$ hits the level $(\log N)^{-1}H_N^{1/2}$, with high probability it has hit level x as well. It remains to verify the two points below.

(i) $\widehat{S}$ is close to S on the time interval $[1, N]$. This follows from a union bound and the exponential tail of X_1^N.

(ii) With high probability by time N we hit the boundary of the cylinder of width $(\log N)^{-1}H_N^{1/2}$. This follows from Lemma 5.1.

Proof of Theorem 2.2: Consider $x \geqslant (\log N)^2$. Observe that

$$\Big\{\max_{1\leqslant m\leqslant N}|\widehat{S}_m^{N,x}| \geqslant (\log N)^{-1}H_N^{1/2},\ \max_{1\leqslant m\leqslant T_K}\widehat{S}_m^{N,x} > x\Big\} \subseteq \Big\{\max_{1\leqslant m\leqslant N}\widehat{S}_m^{N,x} \geqslant x\Big\}.$$

Indeed, on the event $\max_{1\leqslant m\leqslant N}|\widehat{S}_m^{N,x}| \geqslant (\log N)^{-1}H_N^{1/2} \geqslant xL_K$ we have $T_K \leqslant N$.

Next,

$$\begin{aligned}
\mathbb{P}\big(\widehat{S}_m^{N,x} \neq S_m^N \quad \text{for some } 1 \leqslant m \leqslant N\big) &= \mathbb{P}\big(X_i^N < -x \quad \text{for some } 1 \leqslant i \leqslant N\big) \\
&\leqslant C_M N e^{-\theta_1 x} \leqslant C_M N e^{-\theta_1(\log N)^2}.
\end{aligned} \tag{5.19}$$

By Lemma 5.1, (5.19) and Lemma 5.5,

$$\begin{aligned}
&\mathbb{P}\big(\max_{1\leqslant m\leqslant N} \widehat{S}_m^{N,x} \geqslant x\big) \geqslant 1 - \mathbb{P}\big(\max_{1\leqslant m\leqslant N} |\widehat{S}_m^{N,x}| < (\log N)^{-1} H_N^{1/2}\big) - \mathbb{P}\big(\max_{1\leqslant m\leqslant T_K} \widehat{S}_m^{N,x} \leqslant x\big)\\
&\geqslant 1 - \Big[\mathbb{P}\big(\max_{1\leqslant m\leqslant N} |S_m^N| < (\log N)^{-1} H_N^{1/2}\big) + \mathbb{P}\big(\widehat{S}_m^{N,x} \neq S_m^N \quad \text{for some } 1 \leqslant m \leqslant N\big)\Big]\\
&\qquad - \mathbb{P}\big(\max_{1\leqslant m\leqslant T_K} \widehat{S}_m^{N,x} \leqslant x\big)\\
&= 1 - \Big[\mathbb{P}\big(\tau_{(\log N)^{-1} H_N^{1/2}} > N\big) + \mathbb{P}\big(\widehat{S}_m^{N,x} \neq S_m^N \quad \text{for some } 1 \leqslant m \leqslant N\big)\Big]\\
&\qquad - \mathbb{P}\big(\max_{1\leqslant m\leqslant T_K} \widehat{S}_m^{N,x} < x\big)\\
&\geqslant 1 - \big[2e^{-c_\tau N (\log N)^2 H_N^{-1}} + C_M N e^{-\theta_1 (\log N)^2}\big] - C_{12} x (\log N) H_N^{-1/2}\\
&\geqslant 1 - 2e^{-c_\tau (\log N)^2} - C_M N e^{-\theta_1 (\log N)^2} - C_{12} x (\log N) H_N^{-1/2}\\
&\geqslant 1 - (C_{12} + 2) x (\log N) H_N^{-1/2}.
\end{aligned}$$

To get the inequalities above for $N \geqslant N_0$ we increase N_0 if necessary so that $N \geqslant N_0$ guarantees $(\log N)^{-1} H_N^{1/2} \geqslant y_0$ to apply Lemma 5.1, and furthermore so that $2e^{-c_\tau (\log N)^2} \vee C_M N e^{-\theta_1 (\log N)^2} \leqslant (\log N)^3 N^{-1/2}$ to get the last inequality.

Now the final inequality:

$$\begin{aligned}
\mathbb{P}\big(\max_{1\leqslant m\leqslant N} S_m^N \geqslant x\big) &\geqslant \mathbb{P}\big(\max_{1\leqslant m\leqslant N} \widehat{S}_m^{N,x} \geqslant x\big) - \mathbb{P}\big(\widehat{S}_m^{N,x} \neq S_m^N \quad \text{for some } 1 \leqslant m \leqslant N\big)\\
&\geqslant 1 - (C_{12} + 2) x (\log N) H_N^{-1/2} - C_M N e^{-\theta_1 (\log N)^2}\\
&\geqslant 1 - (C_{12} + 3) x (\log N) (|\mu| \vee N^{-1/2}).
\end{aligned}$$

Theorem 2.2 has been proved. □

References

Bhattacharjee, C. and Goldstein, L. On strong embeddings by Stein's method. *Electron. J. Probab.*, **21**, Paper No. 15, 30 (2016). MR3485357.

Busani, O. and Seppäläinen, T. Non-existence of bi-infinite polymer Gibbs measures in the inverse-gamma polymer model. *ArXiv Mathematics e-prints* (2020). arXiv: 2010.11279, to appear in *Electron. J. Probab.*

Chatterjee, S. A new approach to strong embeddings. *Probab. Theory Related Fields*, **152** (1-2), 231–264 (2012). MR2875758.

Durrett, R. *Probability: theory and examples*, volume 31 of *Cambridge Series in Statistical and Probabilistic Mathematics*. Cambridge University Press, Cambridge, fourth edition (2010). ISBN 978-0-521-76539-8. MR2722836.

Fan, W.-T. L. and Seppäläinen, T. Joint distribution of Busemann functions in the exactly solvable corner growth model. *Probability and Mathematical Physics*, **1** (1), 55–100 (2020). DOI: 10.2140/pmp.2020.1.55.

Feller, W. *An introduction to probability theory and its applications. Vol. II.* John Wiley & Sons, Inc., New York-London-Sydney, second edition (1971). MR0270403.

Fraser, D. F. The rate of convergence of a random walk to Brownian motion. *Ann. Probability*, **1**, 699–701 (1973). MR362424.

Karatzas, I. and Shreve, S. E. *Brownian motion and stochastic calculus*, volume 113 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, second edition (1991). ISBN 0-387-97655-8. MR1121940.

Komlós, J., Major, P., and Tusnády, G. An approximation of partial sums of independent RV's, and the sample DF. II. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete*, **34** (1), 33–58 (1976). MR402883.

Krishnapur, M. One idea and two proofs of the KMT theorems. *ArXiv Mathematics e-prints* (2020). arXiv: 2008.03287.

Sakhanenko, A. I. Rate of convergence in the invariance principle for variables with exponential moments that are not identically distributed. In *Limit theorems for sums of random variables*, volume 3 of *Trudy Inst. Mat.*, pp. 4–49. "Nauka" Sibirsk. Otdel., Novosibirsk (1984). MR749757.

Sawyer, S. Rates of convergence for some functionals in probability. *Ann. Math. Statist.*, **43**, 273–284 (1972). MR301782.