GAUSSIAN FIELDS
Notes for Lectures

Ofer Zeitouni

Department of Mathematics
Weizmann Institute, Rehovot 76100, Israel
and
Courant institute, NYU, USA

December 14, 2017. Version 1.04g

DISCLAIMER: I have not (yet) properly acknowledged individually the sources for the material, especially in the beginning of the notes. These include Neveu's and Adler's courses on Gaussian processes [Ad90], the Ledoux-Talagrand book, and various articles.
Acknowledgments: To the participants of the topics in probability (Fall 2013) at NYU for many comments, and to Nathanael Berestycki for additional comments.

1 Gaussian random variables and vectors

1.1 Basic definitions and properties

Definition 1. *A random variable X is called Gaussian if its characteristic function is given by*
$$E(e^{i\theta X}) = e^{i\theta b - \frac{1}{2}\theta^2\sigma^2}\,,$$
for some $b \in \mathbb{R}$ and $\sigma^2 \geq 0$.

Note that we allow for $\sigma = 0$. If $\sigma^2 > 0$ then one has the pdf
$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x-b)^2/2\sigma^2}\,,$$
i.e. $EX = b$ and $\mathrm{Var}(X) = \sigma^2$. The reason for allowing $\sigma = 0$ is so that the next definition is not too restrictive.

Definition 2. *A random vector $\mathbf{X} = (X_1, \ldots, X_n)$ is called Gaussian if $\langle X, \nu \rangle$ is a Gaussian random variable for any deterministic $\nu \in \mathbb{R}^n$.*

Alternatively, $\mathbf{X}$ is a Gaussian random vector iff its characteristic function is given by

$$E(e^{i\langle \nu, X\rangle}) = e^{i\nu^T b - \frac{1}{2}\nu^T R\nu},$$

for some $b \in \mathbb{R}^d$ and R positive definite symmetric $d \times d$ matrix. In that case, $b = E\mathbf{X}$ and R is the covariance matrix of $\mathbf{X}$. (Check these claims!). Throughout, we use the term *positive* in the sense of not negative, i.e. a matrix is positive definite if it is symmetric and all its eigenvalues belong to $\mathbb{R}_+ = \{x \in \mathbb{R} : x \geq 0\}$.

We call random variables (vectors) *centered* if their mean vanishes.

Note: R may not be invertible, even if $\mathbf{X}$ is non-zero. But if $\det R = 0$, there exists a vector ν such that $\langle \nu, \mathbf{X}\rangle$ is deterministic.

The following easy facts are immediate from characteristic function computations.

Lemma 1. *If $\{X_n\}$ is a sequence of Gaussian random variables (vectors) that converge in probability to X, then X is Gaussian and the convergence takes place in L^p, any $p \in [1, \infty)$.*

Proof: (scalar case) Convergence of the characteristic function on compacts yield that X is Gaussian; it also gives that $b_n \to b$ and $R_n \to R$. In particular, since $E|X_n|^p$ is bounded by a continuous function of p, b_n, R_n, the L^p convergence follows from uniform integrability. $\square$

Lemma 2. *For any R symmetric and positive definite one can find a centered Gaussian vector $\mathbf{X}$ with covariance R.*

Proof: Take $\mathbf{Y}$ with i.i.d. centered standard Gaussian entries, and write $\mathbf{X} = R^{1/2}\mathbf{Y}$. $\square$

Lemma 3. *If $\mathbf{Z} = (\mathbf{X}\mathbf{Y})$ is a Gaussian vector and (with obvious block notation) $R_{X,Y} = 0$ then $\mathbf{X}$ is independent of $\mathbf{Y}$.*

Proof: Characteristic function factors. $\square$

The following is an important observation that shows that conditioning for Gaussian vectors is basically a linear algebra exercise.

Lemma 4. *If $\mathbf{Z} = (X, \mathbf{Y})$ is a centered Gaussian vector then $\hat{X}_Y := E[X|\mathbf{Y}]$ is a Gaussian random variable, and $\hat{X}_Y = T\mathbf{Y}$ for a deterministic matrix T. If $\det(R_{YY}) \neq 0$ then $T = R_{XY}R_{YY}^{-1}$.*

Proof: Assume first that $\det(R_{YY}) \neq 0$. Set $W = X - T\mathbf{Y}$. Then, since $T\mathbf{Y}$ is a linear combination of entries of $\mathbf{Y}$ and since $\mathbf{Z}$ is Gaussian, we have that $(W, \mathbf{Y})$ is a (centered) Gaussian vector. Now,

$$E(W\mathbf{Y}) = R_{XY} - TR_{YY} = 0\,.$$

Hence, by Lemma 3, W and $\mathbf{Y}$ are independent. Thus, $E[W|\mathbf{Y}] = EW = 0$, and the conclusion follows from the linearity of the conditional expectation.

In case $\det(R_{YY}) = 0$ and $\mathbf{Y} \neq 0$, let Q denote the projection to $\mathrm{range}(R_{YY})$, a subspace of dimension $d \geq 1$. Then $\mathbf{Y} = Q\mathbf{Y} + Q^{\perp}\mathbf{Y} = Q\mathbf{Y}$ since $\mathrm{Var}(Q^{\perp}\mathbf{Y}) = 0$. Changing bases, one thus finds a matrix B with $n-d$ zero rows so that $\mathbf{Y} = \hat{Q}B\mathbf{Y}$ for some matrix $\hat{Q}$, and the covariance matrix of the d dimensional vector of non-zero entries of $B\mathbf{Y}$ is non-degenerate. Now repeat the first part of the proof using the non-zero entries of $B\mathbf{Y}$ instead of $\mathbf{Y}$. $\square$

1.2 Gaussian vectors from Markov chains

Let $\mathcal{X}$ denote a finite state space on which one is given a (discrete time) irreducible, reversible Markov chain $\{S_n\}$. That is, with Q denoting the transition matrix of the Markov chain, there exists a (necessarily unique up to normalization) positive vector $\mu = \{\mu_x\}_{x\in\mathcal{X}}$ so that $\mu_x Q(x,y) = \mu_y Q(y,x)$. We often, but not always, normalize μ to be a probability vector.

Fix $\Theta \subset \mathcal{X}$ with $\Theta \neq \mathcal{X}$ and set $\tau = \min\{n \geq 0 : S_n \in \Theta\}$. Set, for $x, y \notin \Theta$,

$$G(x,y) = \frac{1}{\mu_y} E^x \sum_{n=0}^{\tau} \mathbf{1}_{\{S_n = y\}} = \frac{1}{\mu_y} \sum_{n=0}^{\infty} P^x(S_n = y, \tau > n)\,.$$

We also set $G(x,y) = 0$ if either $x \in \Theta$ or $y \in \Theta$. Note that, up to the multiplication by μ_y^{-1}, G is the Green function associated with the Markov chain killed upon hitting Θ. We now have the following.

Lemma 5. *G is symmetric and positive-definite.*

Proof: Let $Z_n(x,y)$ denote the collection of paths $\mathbf{z} = (z_0, z_1, \ldots, z_n)$ of length n that start at x, end at y and avoid Θ. We have

$$\begin{aligned} P^x(S_n = y, \tau > n) &= \sum_{\mathbf{z}\in Z_n(x,y)} \prod_{i=1}^{n-1} Q(z_i, z_{i+1}) = \sum_{\mathbf{z}\in Z_n(x,y)} \prod_{i=1}^{n-1} Q(z_{i+1}, z_i)\frac{\mu_{z_{i+1}}}{\mu_{z_i}} \\ &= \frac{\mu_y}{\mu_x} \sum_{\mathbf{z}\in Z_n(y,x)} \prod_{i=1}^{n-1} Q(z_i, z_{i+1}) = \frac{\mu_y}{\mu_x} P^y(S_n = x, \tau > n)\,. \end{aligned}$$

This shows that $G(x,y) = G(y,x)$. To see the positive definiteness, let $\hat{Q}$ denote the restriction of Q to $\mathcal{X} \setminus \Theta$. Then, $\hat{Q}$ is sub-stochastic, and due to irreducibility and the Perron-Frobenius theorem, its spectral radius is strictly smaller than 1. Hence, $I - \hat{Q}$ is invertible, and

$$(I - \hat{Q})^{-1}(x,y) = 1_{x=y} + \hat{Q}(x,y) + \hat{Q}^2(x,y) + \ldots = G(x,y)\mu_y\,.$$

In case μ_x is independent of x, this would imply that all eigenvalues of G are non-negative.

In the general case[1], introduce the bilinear form

$$E(f,g) = \sum \mu_x Q_{x,y}(f(y)-f(x))(g(y)-g(x)).$$

For functions that vanish on Θ, this can be written as

$$E(f,g) \sum_{x,y\in\mathcal{X}\setminus\Theta} \mu_x \hat{Q}_{x,y}(f(x)-f(y))(g(x)-g(y)) + \sum_{x\in\Theta,y\in\mathcal{X}\setminus\Theta} \mu_x Q_{x,y} f(y)g(y).$$

A bit of algebra shows that for any f,g,

$$E(f,g) = 2[\sum_x \mu_x f(x)g(x) - \sum \mu_x Q_{xy} f(x)g(y)].$$

Restricting to functions that vanish at Θ gives

$$\begin{aligned} E(f,g) &= 2[\sum \mu_x f(x)g(x) - \sum \mu_x \hat{Q}_{x,y} f(x)g(y)] \\ &= 2\sum f(x)g(y)\mu_x(I-\hat{Q})^{-1}_{x,y} = 2\sum f(x)g(x)G(x,y). \end{aligned}$$

Since $E(f,f) > 0$ if $f > 0$, the positivity of eigenvalues of G follows. □

From Lemmas 2 and 5 it follows that the function G is the covariance of some Gaussian vector.

Definition 3. *The (centered) Gaussian vector with covariance G (denoted $\{X(x)\}$) is called the Gaussian Free Field (GFF) associated with Q, Θ.*

The Green function representation allows one to give probabilistic representation for certain conditionings. For example, let $A \subset \mathcal{X}\setminus\Theta$ and set $\mathbf{X}_A = E[\mathbf{X}|X(x), x\in A]$. By Lemma 3 we have that $X_A(x) = \sum_{z\in A} a(x,z)X(z)$. We clearly have that for $x\in A$, $a(x,y) = 1_{x=y}$. On the other hand, because G_A (the restriction of G to A) is non-degenerate, we have that for $x\notin A$, $a(x,y) = \sum_{w\in A} G(x,w)G_A^{-1}(w,y)$. It follows that for any $y\in A$, $a(x,y)$ (as a function of $x\notin A$) is harmonic, i.e. $\sum Q(x,w)a(w,y) = a(x,y)$ for $x\notin A$. Hence, a satisfies the equations

$$\begin{cases} (I-Q)a(x,y) = 0,\ x\notin A\,, \\ a(x,y) = 1_{\{\mathbf{x}=\mathbf{y}\}}\,, \quad x\in A\,. \end{cases} \tag{1.2.1}$$

By the maximum principle, the solution to (1.2.1) is unique. On the other hand, one easily verifies that with $\tau_A = \min\{n\geq 0 : S_n\in A\}$, the function $\hat{a}(x,y) = P^x(\tau_A < \tau, S_{\tau_A} = y)$ satisfies (1.2.1). Thus, $a = \hat{a}$.

The difference $\mathbf{Y}_A = \mathbf{X} - \mathbf{X_A}$ is independent of $\{X_x\}_{x\in A}$ (see the proof of Lemma 4). What is maybe surprising is that $\mathbf{Y}_A$ can also be viewed as a GFF.

[1] thanks to Nathanael Berestycki for observing that we need to consider that case as well

Lemma 6. $\mathbf{Y}_A$ *is the GFF associated with* $(Q, \Theta \cup A)$.

Proof: Let G_A denote the Green function restricted to A (i.e., with $\tau_A \wedge \tau$ replacing τ). By the strong Markov property we have

$$G(x,y) = \sum_{y' \in A} a(x,y')G(y',y) + G_A(x,y)\,, \tag{1.2.2}$$

where the last term in the right side of (1.2.2) vanishes for $y \in A$. On the other hand,

$$E(Y_A(x)Y_A(x')) = G(x,x') - E(X(x)X_A(x')) - E(X(x')X_A(x)) + EX_A(x)X_A(x')\,.$$

Note that

$$EX(x)X_A(x') = \sum_{y \in A} a(x',y)G(x,y) = G(x',x) - G_A(x',x)$$

while

$$\begin{aligned} EX_A(x)X_A(x') &= \sum_{y,y' \in A} a(x,y)a(x',y)G(y,y') \\ &= \sum_{y' \in A} a(x,y')G(x',y') = G(x,x') - G_A(x,x')\,. \end{aligned}$$

Substituting, we get $E(Y_A(x)Y_A(x')) = G_A(x,x')$, as claimed. □

Another interesting interpretation of the GFF is obtained as follows. Recall that the GFF is the mean zero Gaussian vector with covariance G. Since G is invertible (see the proof of Lemma 5), the density of the vector $\{X_x\}_{x \in \mathcal{X} \setminus \Theta}$ is simply

$$p(\mathbf{z}) = \frac{1}{Z} \exp\left(-\mathbf{z}^T G^{-1} \mathbf{z}\right)$$

where Z is a normalization constant. Since $(I - \hat{Q})^{-1} = G\mu$, where μ denotes the diagonal matrix with entries μ_x on the diagonal, we get that $G^{-1} = \mu(I - \hat{Q})$. In particular, setting $\{z'_x\}_{x \in \mathcal{X}}$ with $z'_x = 0$ when $x \in \Theta$ and $z'_x = z_x$ if $x \in \mathcal{X} \setminus \Theta$, we obtain

$$p(\mathbf{z}) = \frac{1}{Z} \exp\left(-\sum_{x \neq y \in \mathcal{X}} (z'_x - z'_y)^2 C_{x,y}/2\right) \tag{1.2.3}$$

where $C_{x,y} = \mu_x \hat{Q}(x,y)$.

Exercise 1. Consider continuous time, reversible Markov chains $\{S_t\}_{t \geq 0}$ on a finite state space $\mathcal{X}$ with $G(x,y) = \frac{1}{\mu_y} E^x \int_0^\tau \mathbf{1}_{S_t = y}\,dt$, and show that the GFF can be associated also with that Green function.

Exercise 2. Consider a finite binary tree of depth n rooted at o and show that, up to scaling, the GFF associated with $\Theta = o$ and the simple random walk on the tree is the same (up to scaling) as assigning to each edge e an independent, standard Gaussian random variable Y_e and then setting

$$X_v = \sum_{e \in o \leftrightarrow v} Y_e \,.$$

Here, $o \leftrightarrow v$ denotes the geodesic connecting o to v. This is the model of a Gaussian binary branching random walk (BRW).

Exercise 3. Show that (1.2.3) has the following interpretation. Let $\mathcal{A} = \{\{x,y\} : \text{both } x \text{ and } y \text{ belong to } \Theta\}$. Let $g_{x,y} = 0$ if $\{x,y\} \in \mathcal{A}$ and let $\{g_{x,y}\}_{\{x,y\} \notin \mathcal{A}, Q_{x,y} > 0}$ be a collection of independent centered Gaussian variables, with $E g_{x,y}^2 = 1/C_{x,y}$. Set an arbitrary order on the vertices and define $g_{(x,y)} = g_{x,y}$ if $x < y$ and $g_{(x,y)} = -g_{x,y}$ if $x > y$. For a closed path $p = (x = x_0, x_1, \ldots, x_k = x_0)$ with vertices in $\mathcal{X}$ and $Q(x_i, x_{i+1}) > 0$ for all i, set $Y_p = \sum_{i=0}^{k-1} g_{(x_i, x_{i+1})}$, and let $\mathcal{P}$ denote the collection of all such closed paths. Let $\sigma_\Theta := \sigma(\{Y_p\}_{p \in \mathcal{P}})$. Let $\bar{g}_{(x,y)} = g_{(x,y)} - E(g_{(x,y)} | \sigma_\Theta)$, and recall that the collection of random variables $\bar{g}_{(x,y)}$ is independent of σ_Θ. Prove that $\{\bar{g}_{(x,y)}\}_{\{x,y\}: Q_{x,y} > 0}$ has the same law as $\{Z_{(x,y)} := (X_x - X_y)\}_{\{x,y\}: Q_{x,y} > 0}$ and deduce from this that the GFF can be constructed from sampling the collection of variables $\bar{g}_{(x,y)}$.

1.3 Spaces of Gaussian variables

Definition 4. *A Gaussian space is a closed subset of $L^2 = L^2(\Omega, \mathcal{F}, P)$ consisting of (equivalence classes of) centered Gaussian random variables.*

Note that the definition makes sense since the L^2 limit of Gaussian random variables is Gaussian.

Definition 5. *A Gaussian process (field, function) indexed by a set T is a collection of random variables $\{X_t\}_{t \in T}$ such that for any $(t_1, \ldots, t_k)$, the random vector $(X_{t_1}, \ldots, X_{t_k})$ is Gaussian.*

The closed subspace of L^2 generated by X_t is a Hilbert space with respect to the standard inner product, denoted H. With $\mathcal{B}(H)$ denoting the σ-algebra generated by H, we have that $\mathcal{B}(H)$ is the closure of $\sigma(X_t, t \in T)$ with respect to null sets. Any random variable measurable with respect to $\mathcal{B}(H)$ is called a *functional* on $\{X_t\}$.

Exercise 4. Let $\{H_i\}_{i \in I}$ be closed subsets of H and let $\mathcal{B}(H_i)$ denote the corresponding σ-algebras. Show that $\{\mathcal{B}(H_i)\}_{i \in I}$ is an independent family of σ-algebras iff the H_i are pairwise orthogonal.

As a consequence, if H' is a closed subset of H then $E(\cdot | \mathcal{B}(H'))$ is the orthogonal projection to H' in H.

1.4 The reproducing kernel Hilbert space associated with a centered Gaussian random process

Let $R(s,t) = EX_sX_t$ denote the covariance of the centered Gaussian process $\{X_t\}$. Note that R is symmetric and positive definite: $\infty > \sum a(s)a(t)R(s,t) \geq 0$ whenever the sum is over a finite set.

Define the map $u : H \to \mathbb{R}^T$ by

$$u(Z)(t) = E(ZX_t)\,.$$

Note in particular that $u(X_s)(\cdot) = R(s,\cdot)$.

Definition 6. *The space*

$$\mathcal{H} := \{g : g(\cdot) = u(Z)(\cdot), \text{some } Z \in H\}$$

equipped with the inner product $\langle f, g\rangle_{\mathcal{H}} = E(u^{-1}(f)u^{-1}(g))$, *is called the reproducing kernel Hilbert space (RKHS) associated with* $\{X_t\}$ *(or with* R*).*

We will see shortly the reason for the name.

Exercise 5. Check that $\mathcal{H}$ is a Hilbert space which is isomorphic to H (because the map u is injective and $\{X_t\}$ generates H).

Note: For $Z = \sum_{i=1}^k a_i X_{t_i}$ we have $u(Z)(t) = \sum_{i=1}^k a_i R(t_i, t)$. Thus $\mathcal{H}$ could also be constructed as the closure of such function under the inner product $\langle \sum_{i=1}^k a_i R(t_i,t), \sum_{i=1}^k b_i R(t_i,t)\rangle_{\mathcal{H}} = \sum a_i b_j R(t_i,t_j)$.

Now, for $h \in \mathcal{H}$ with $u^{-1}(h) =: Z$ we have, because $u^{-1}(R(t,\cdot)) = X_t$,

$$\langle h, R(t,\cdot)\rangle_{\mathcal{H}} = E(u^{-1}(h)X_t) = u(Z)(t) = h(t)\,.$$

Thus, $\langle h, R(t,\cdot)\rangle_{\mathcal{H}} = h(t)$, explaining the RKHS nomenclature. Further, since $R(t,\cdot) \in \mathcal{H}$ we also have that $\langle R(t,\cdot), R(s,\cdot)\rangle = R(s,t)$.

We can of course reverse the procedure.

Lemma 7. *Let* T *be an arbitrary set and assume* $K(\cdot,\cdot)$ *is a positive definite kernel on* $T \times T$. *Then there exists a closed Hilbert space* $\mathcal{H}$ *of functions on* T, *such that:*
- $K(t,\cdot) \in \mathcal{H}$ *and generates* $\mathcal{H}$.
- *for all* $h \in \mathcal{H}$, *one has* $h(t) = \langle h, K(t,\cdot)\rangle_{\mathcal{H}}$

Hint of proof: Start with finite combinations $\sum a_i K(s_i,\cdot)$, and close with respect to the inner product. Use the positivity of the kernel to show that $\langle h, h\rangle_{\mathcal{H}} \geq 0$ and then, by Cauchy-Schwarz and the definitions,

$$|h(t)|^2 = |\langle h, K(t,\cdot)\rangle_{\mathcal{H}}|^2 \leq \langle h, h\rangle_{\mathcal{H}} K(t,t)\,.$$

Thus, $\langle h, h\rangle_{\mathcal{H}} = 0$ implies $h = 0$. $\square$

Proposition 1. *Let T, K be as in Lemma 7. Then there exists a probability space and a centered Gaussian process $\{X_t\}_{t\in T}$ with covariance $R = K$.*

Proof: Let $\{h_i\}_{i\in J}$ be an orthonormal basis of $\mathcal{H}$. Let $\{Y_i\}_{i\in J}$ denote an i.i.d. collection of standard Gaussian random variables (exists even if J is not countable). Let $H = \{\sum a_i Y_i : \sum a_i^2 < \infty\}$ (sum over arbitrary countable subsets of J). H is a (not necessarily separable) Hilbert space. Now define the isomorphism of Hilbert spaces

$$\begin{aligned} &\mathcal{H} \xrightarrow{I} H \\ &h_i \mapsto Y_i \quad (i \in J) \end{aligned}$$

Set $X_t = I(K(t,\cdot))$. Now one easily checks that X_t satisfies the conditions, since $EX_sX_t = \langle K(t,\cdot), K(s,\cdot)\rangle_{\mathcal{H}} = K(s,t)$. $\square$

We discuss some continuity and separability properties of the Gaussian process $\{X_t\}$ in terms of its covariance kernel. In the rest of this section, we assume that T is a topological space.

Proposition 2. *The following are equivalent.*

- *The process $\{X_t\}_{t\in T}$ is L^2 continuous (i.e., $E(X_t - X_s)^2 \to_{|s-t|\to 0} 0$).*
- *The kernel $R : T \times T \to \mathbb{R}$ is continuous.*

Under either of these conditions, $\mathcal{H}$ is a subset of $C(T)$, the continuous functions on T. If T is separable, so is $\mathcal{H}$ and hence so is the process $\{X_t\}_{t\in T}$ in H.

Proof: If R is continuous we have $E(X_t - X_s)^2 = R(s,s) - 2R(s,t) + R(t,t)$, showing the L^2 continuity. Conversely,

$$\begin{aligned} &|R(s,t) - R(u,v)| = |E(X_sX_t - X_uX_v)| \\ &\le |E(X_s - X_u)(X_t - X_v)| + |EX_u(X_t - X_v)| + |E(X_s - X_u)X_v| \,. \end{aligned}$$

By Cauchy-Schwarz, the right side tends to 0 as $s \to u$ and $t \to v$.

Let $h \in \mathcal{H}$. By the RKHS representation, $h(t) = \langle h, R(t,\cdot)\rangle_{\mathcal{H}}$. Since $\{X_t\}$ is L^2 continuous, the isomorphism implies that $t \to R(t,\cdot)$ is continuous in $\mathcal{H}$, and then, from Cauchy-Schwarz and the above representation of h, one concludes that $t \to h(t)$ is continuous. Further, if T is separable then it has a dense subset $\{t_n\}$ and, by the continuity of K, we conclude that $\{R(t_n,\cdot)\}_n$ generates $\mathcal{H}$. From the isomorphism, it follows that $\{X_{t_n}\}_n$ generates H, i.e. $\{X_t\}$ is separable in H. $\square$

Exercise 6. Show that $\{X_t\}$ is bounded in L^2 iff $\sup_{t\in T} R(t,t) < \infty$, and that under either of these conditions, $\sup_{t\in T} |h(t)| \le \sqrt{\sup_T R(t,t)}\|h\|_{\mathcal{H}}$.

Let T be a separable topological space. We say that a stochastic process $\{X_t\}_{t\in T}$ is separable if there is a countable $D \subset T$ and a fixed null set $\Omega' \subset \Omega$ so that, for any open set $U \subset T$ and any closed set A,

$$\{X_t \in A, t \in D \cap U\} \setminus \{X_t \in A, t \in U\} \subset \Omega' .$$

$\{X_t\}_{t\in T}$ is said to have a separable version if there is a separable process $\{\tilde{X}_t\}_{t\in T}$ so that $P(X_t = \tilde{X}_t) = 1, \forall t \in T$. It is a fundamental result in the theory of stochastic process that if T is a separable metric space then $\{X_t\}_{t\in T}$ possesses a separable version. In the sequel, unless we state otherwise, we take T to be a compact second countable (and hence metrizable) Hausdorff space. This allows us to define a separable version of the process $\{X_t\}$, and in the sequel we always work with such version. When the covariance R is continuous, Proposition 3 below can be used to construct explicitely a separable version of the process (that actually works for any countable dense D).

Example 1. Let T be a finite set, and let $\{X_t\}_{t\in T}$ be a centered Gaussian process with non-degenerate covariance (matrix) R. Then, $\langle f, g\rangle_{\mathcal{H}} = \sum f_i g_j R^{-1}(i,j)$. To see that, check the RKHS property:

$$\langle f, R(t,\cdot)\rangle_{\mathcal{H}} = \sum_{i,j} f_i R(t,j) R^{-1}(i,j) = \sum_i f_i 1_{t=i} = f_t\,, \quad t \in T\,.$$

Example 2. Take $T = [0,1]$ and let X_t be standard Brownian motion. Then $R(s,t) = s \wedge t$. If $h(t) = \sum a_i R(s_i, t)$, $f(t) = \sum b_i R(s_i, t)$ then

$$\begin{aligned}\langle h, f\rangle_{\mathcal{H}} &= \sum_{i,j} a_i b_j R(s_i, s_j) = \sum_{i,j} a_i b_j (s_i \wedge s_j) \\ &= \sum_{i,j} a_i b_j \int_0^1 1_{[0,s_i]}(u) 1_{[0,s_j]}(u)du = \int_0^1 h'(u) f'(u) du\,.\end{aligned}$$

This hints that

$$\mathcal{H} = \{f : f(t) = \int_0^t f'(u)du, \int_0^1 (f'(u))^2 du < \infty\}\,,$$

with the inner product $\langle f, g\rangle_{\mathcal{H}} = \int_0^1 f'(s) g'(s) ds$. To verify that, need the RKHS property:

$$\langle R(t,\cdot), f(\cdot)\rangle_{\mathcal{H}} = \int_0^1 f'(u) 1_{[0,t]}(u) du = \int_0^t f'(s) ds = f(t)\,.$$

A useful aspect of the RKHS is that it allows one to rewrite X_t, with covariance R, in terms of i.i.d. random variables. Recall that we assume T to be second countable and we will further assume that R is continuous. Then, as we saw, both H and $\mathcal{H}$ are separable Hilbert spaces. Let $\{h_n\}$ be an orthonormal base of $\mathcal{H}$ corresponding to an i.i.d. basis of centered Gaussian variables $\{\xi_n\}$ in H.

Proposition 3. *With notation and assumption as above, we have*

$$R(s,\cdot) = \sum_n h_n(s) h_n(\cdot) \quad (\textit{equality in } \mathcal{H}) \tag{1.4.4}$$

$$X_s = \sum_n \xi_n h_n(s) \quad (\textit{equality in } H) \tag{1.4.5}$$

Further, the convergence in (1.4.4) *is uniform on compact subsets of* T.

Proof: We have

$$E(X_s \xi_n) = \langle R(s,\cdot), h_n \rangle_{\mathcal{H}} = h_n(s)\,,$$

where the first equality is from the definition of $\mathcal{H}$ and the second from the RKHS property. The equality (1.4.5) follows. The claim (1.4.4) then follows by writing $R(s,t) = E(X_s X_t)$ and applying (1.4.5). Note that one thus obtains $R(t,t) = \sum h_n(t)^2$.

To see the claimed uniform convergence, note that under the given assumptions, $h_n(\cdot)$ are continuous functions. The monotone convergence of the continuous functions $R_N(t) := \sum_{n \le N} h_n(t)^2$ toward the continuous function $R(t,t)$ is therefore, by Dini's theorem, uniform on compacts (indeed, fixing a compact $S \subset T$, the compact sets $S_N(\epsilon) := \{t \in S : R(t,t) \ge \epsilon + R_N(t)\}$ monotonically decrease to the empty set as $N \to \infty$, implying that there is a finite N with $S_N(\epsilon)$ empty). Thus, the sequence $f_N(s,\cdot) := \sum_{n \le N} h_n(s) h_n(\cdot)$ converges in $\mathcal{H}$ uniformly in s belonging to compacts. Now use again the RKHS property: $f_N(s,t) = \langle R(t,\cdot), f_N(s,\cdot) \rangle_{\mathcal{H}}$ to get

$$\begin{aligned}\sup_{s,t \in S \times S} |f_N(s,t) - R(s,t)| &= \sup_{s,t \in S \times S} |\langle R(t,\cdot), f_N(s,\cdot) - R(s,\cdot) \rangle_{\mathcal{H}}| \\ &\le \sup_{s \in S} \|f_N(s,\cdot) - R(s,\cdot)\|_{\mathcal{H}} \cdot \sup_{t \in S} \|R(t,\cdot)\|_{\mathcal{H}}\,. \end{aligned} \tag{1.4.6}$$

Since $\langle R(t,\cdot), R(t,\cdot) \rangle_{\mathcal{H}} = R(t,t)$, we have (by the compactness of S and continuity of $t \to R(t,t)$) that $\sup_{t \in S} \|R(t,\cdot)\|_{\mathcal{H}} < \infty$. Together with (1.4.6), this completes the proof of uniform convergence on compacts. $\square$

Remark: One can specialize the previous construction as follows. Start with a finite measure μ on a compact T with $\mathrm{supp}(\mu) = T$, and a symmetric positive definite continuous kernel $K(\cdot,\cdot)$ on T. Viewing K as an operator on $L^2_\mu(T)$, it is Hilbert-Schmidt, and its normalized eigenfunctions $\{h_n\}$ (with corresponding eigenvalues $\{\lambda_n\}$) form an orthonormal basis of $L^2_\mu(T)$, and in fact due to the uniform boundedness of K on T, one has $\sum \lambda_n < \infty$ (all these facts follow from the general form of Mercer's theorem). Now, one checks that $\{\sqrt{\lambda_n} h_n\}$ is an orthonormal base of $\mathcal{H}$, and therefore one can write $X_t = \sum \xi_n \sqrt{\lambda_n} h_n(t)$. This special case of the RKHS often comes under the name *Karhunen-Loeve* expansion. The Brownian motion example corresponds to μ Lebesgue on $T = [0,1]$.

As an application of the series representation in Proposition 3, we provide a $0-1$ law for the sample path continuity of Gaussian processes. Recall that we now work with T compact (and for convenience, metric) and hence $\{X_t\}$ is a separable process. Define the oscillation function

$$\mathrm{osc}_X(t) = \lim_{\epsilon \to 0} \sup_{u,v \in B(t,\epsilon)} |X_u - X_v|.$$

Here, $B(t,\epsilon)$ denotes the open ball of radius ϵ around t. Since $\{X_t\}$ is separable, the oscillation function is well defined as a random variable. The next theorem shows that it is in fact deterministic.

Theorem 1. *Assumptions as in the preceeding paragraph. Then there exists a deterministic function h on T, upper semicontinuous, so that*

$$P(\mathrm{osc}_X(t) = h(t), \quad \forall t \in T) = 1.$$

Proof: Let $B \subset T$ denote the closure of a non-empty open set. Define

$$\mathrm{osc}_X(B) = \lim_{\epsilon \to 0} \sup_{s,t \in B, d(s,t) < \epsilon} |X_s - X_t| ,$$

which is again well defined by the separability of $\{X_t\}$. Recall that (in the notation of Proposition 3), $X_t = \sum_{j=1}^\infty \xi_j h_j(t)$, where the functions $h_j(\cdot)$ are each uniformly continuous on the compact set T. Define

$$X_t^{(n)} = \sum_{j=n+1}^{\infty} \xi_j h_j(t) .$$

Since $X_t - X_t^{(n)}$ is uniformly continuous in t for each n, we have that $\mathrm{osc}_X(B) = \mathrm{osc}_{X^{(n)}}(B)$ a.s., with the null-set possibly depending on B and n. By Kolmogorov's $0-1$ law (applied to the sequence of independent random variables $\{\xi_n\}$), there exists a deterministic $h(B)$ such that $P(\mathrm{osc}_X(B) = h(B)) = 1$. Choose now a countable open base $\mathcal{B}$ for T, and set

$$h(t) = \inf_{B \in \mathcal{B}: t \in B} h(\overline{B}) .$$

Then h is upper-semicontinuous, and on the other hand

$$\mathrm{osc}_X(t) = \inf_{B \in \mathcal{B}: t \in B} \mathrm{osc}_X(\overline{B}) = \inf_{B \in \mathcal{B}: t \in B} h(\overline{B}) = h(t) ,$$

where the second equality is almost sure, and we used that $\mathcal{B}$ is countable. $\square$

The following two surprising corollaries are immediate:

Corollary 1. *TFAE:*
- $P(\lim_{s \to t} X_s = X_t,$ *for all* $t \in T) = 1.$
- $P(\lim_{s \to t} X_s = X_t) = 1,$ *for all* $t \in T.$

Corollary 2. $P(X.$ *is continuous on* $T) = 0$ *or* $1.$

Indeed, all events in the corollaries can be decided in terms of whether $h \equiv 0$ or not.

2 The Borell–Tsirelson-Ibragimov-Sudakov inequality

In what follows we always assume that T is compact and that $\{X_t\}$ is a centered Gaussian process on T with continuous covariance (we mainly assume the continuous covariance to ensure that $\{X_t\}$ is separable). We use the notation

$$X^{\sup} := \sup_{t \in T} X_t$$

noting that $X^{\sup}$ is *not* a norm.

Theorem 2 (Borell's inequality). *Assume that $X^{\sup} < \infty$ a.s.. Then, $EX^{\sup} < \infty$, and*

$$P\left(\mid X^{\sup} - EX^{\sup} \mid > x\right) \leq 2e^{-x^2/2\sigma_T^2}$$

where $\sigma_T^2 := \max_{t \in T} EX_t^2$.

The heart of the proof is a concentration inequality for standard Gaussian random variables.

Proposition 4. *Let $\mathbf{Y} = (Y_1, \ldots, Y_k)$ be a vector whose entries are i.i.d. centered Gaussians of unit variance. Let $f : \mathbb{R}^k \to \mathbb{R}$ be Lipschitz, i.e. $L_f := \sup_{x \neq y}(|f(x) - f(y)|/|x - y|) < \infty$. Then,*

$$P(|f(\mathbf{Y}) - Ef(\mathbf{Y})| > x) \leq 2e^{-x^2/2L_f^2}\,.$$

There are several proofs of Proposition 4. Borell's proof relied on the Gaussian isoperimetric inequality. In fact, Proposition 4 is an immediate consequence of the fact that the one dimensional Gaussian measure satisfies the log-Sobolev inequality, and that log-Sobolev inequalities are preserved when taking products of measures. Both these facts can be proved analytically, either from the Gaussian isoperimetry or directly from inequalities on Bernoulli variables (due to Bobkov). We will take a more probabilistic approach, following Pisier and/or Tsirelson and al.

Proof of Proposition 4: By homogeneity, we may and will assume that $L_f = 1$. Let $F(x, t) = E^x f(B_{1-t})$ where $B_\cdot$ is standard k-dimensional Brownian motion. The function $F(x, t)$ is smooth on $\mathbb{R} \times (0, 1)$ (to see that, represent it as integral against the heat kernel). Now, because the heat kernel and hence $F(x, t)$ is harmonic with respect to the operator $\partial_t + \frac{1}{2}\Delta$, we get by Ito's formula, with $I_t = \int_0^t (\nabla F(B_s, s), dB_s)$,

$$f(B_1) - Ef(B_1) = F(B_1, 1) - F(0, 0) = I_1\,. \tag{2.1.1}$$

Since f is Lipschitz(1), we have that $P_s f$ is Lipschitz(1) for any s and therefore $\|\nabla F(B_s, s)\|_2 \leq 1$, where $\|\cdot\|_2$ denotes here the Euclidean norm. On the other hand, since for any stochastic integral I_t with bounded integrand and any $\theta \in \mathbb{R}$ we have $1 = E(e^{\theta I_t - \theta^2 \langle I \rangle_t /2})$ where $\langle I \rangle_t$ is the quadratic variation process of I_t, we conclude that

$$1 = E(e^{\theta I_1 - \frac{\theta^2}{2} \int_0^1 \|\nabla F(B_s,s)\|_2^2 ds}) \geq E(e^{\theta I_1 - \frac{\theta^2}{2}}),$$

and therefore $E(e^{\theta I_1}) \leq e^{\theta^2/2}$. By Chebycheff's inequality we conclude that

$$P(|I_1| > x) \leq 2 \inf_\theta e^{\theta x + \theta^2/2} = 2e^{-x^2/2}.$$

Substituting in (2.1.1) yields the proposition. □

Proof of Theorem 2: We begin with the case where T is finite (the main point of the inequality is then that none of the constants in it depend on the cardinality of T); in that case, $\{X_t\}$ is simply a Gaussian vector $\mathbf{X}$, and we can write $\mathbf{X} = R^{1/2}\mathbf{Y}$ where $\mathbf{Y}$ is a vector whose components are i.i.d. standard Gaussians. Define the function $f : \mathbb{R}^{|T|} \to \mathbb{R}$ by $f(\mathbf{x}) = \max_{i \in T}(R^{1/2}\mathbf{x})_i$. Now, with e_i denoting the ith unit vector in $\mathbb{R}^{|T|}$,

$$\begin{aligned}|f(\mathbf{x}) - f(\mathbf{y})| &= |\max_{i\in T}(R^{1/2}\mathbf{x})_i - \max_{i \in T}(R^{1/2}\mathbf{y})_i| \leq \max_i |(R^{1/2}(\mathbf{x}-\mathbf{y}))_i| \\ &\leq \max_i \|e_i R^{1/2}\|_2 \|\mathbf{x}-\mathbf{y}\|_2 = \max_i (e_i R^{1/2} R^{1/2} e_i^T)^{1/2} \|\mathbf{x}-\mathbf{y}\|_2 \\ &= \max_i R_{ii}^{1/2} \|\mathbf{x}-\mathbf{y}\|_2 .\end{aligned}$$

Hence, f is Lipschitz(σ_T). Now apply Proposition 4 to conclude the proof of Theorem 2, in case T is finite.

To handle the case of infinite T, we can argue by considering a (dense) countable subset of T (here separability is crucial) and use monotone and then dominated convergence, as soon as we show that $EX^{\sup} < \infty$. To see that this is the case, we argue by contradiction. Thus, assume that $EX^{\sup} = \infty$. Let $T_1 \subset \ldots T_n \subset T_{n+1} \subset \ldots \subset T$ denote an increasing sequence of finite subsets of T such that $\cup_n T_n$ is dense in T. Choose M large enough so that $2e^{-M^2/2\sigma_T^2} < 1$. By the first part of the theorem,

$$\begin{aligned}1 > 2e^{-M^2/2\sigma_T^2} &\geq 2e^{-M^2/2\sigma_{T_n}^2} \geq P(\| X^{\sup}{}_{T_n} - EX^{\sup}{}_{T_n} \| > M) \\ &\geq P(EX^{\sup}{}_{T_n} - X^{\sup}{}_{T_n} > M) \geq P(EX^{\sup}{}_{T_n} - X^{\sup}{}_T > M).\end{aligned}$$

Since $EX^{\sup}{}_{T_n} \to EX^{\sup}{}_T$ by separability and monotone convergence, and since $X^{\sup}{}_T < \infty$ a.s., we conclude that the right side of the last display converges to 1 as $n \to \infty$, a contradiction. □

3 Slepian's inequality and variants

We continue with general tools for "nice" Gaussian processes; Borell's inequality allows one to control the maximum of a Gaussian process. Slepian's inequality allows one to compare two such processes. As we saw in the proof of Borell's inequality, once estimates are done (in a dimension-independent way) for Gaussian vectors, separability and standard convergence results allow one to transfer results to processes. Because of that, we focus our attention on Gaussian vectors, i.e. to the situation where $T = \{1, \dots, n\}$.

Theorem 3 (Slepian's lemma). *Let* $\mathbf{X}$ *and* $\mathbf{Y}$ *denote two* n*-dimensional centered Gaussian vectors. Assume the existence of subsets* $A, B \in T \times T$ *so that*

$$\begin{aligned} EX_iX_j &\le EY_iY_j, (i,j) \in A \\ EX_iX_j &\ge EY_iY_j, (i,j) \in B \\ EX_iX_j &= EY_iY_j, (i,j) \notin A \cup B. \end{aligned}$$

Suppose $f : \mathbb{R}^n \to \mathbb{R}$ *is smooth, with appropriate growth at infinity of* f *and its first and second derivatives (exponential growth is fine), and*

$$\begin{aligned} \partial_{ij} f &\ge 0, (i,j) \in A \\ \partial_{ij} f &\le 0, (i,j) \in B\,. \end{aligned}$$

Then, $Ef(\mathbf{X}) \le Ef(\mathbf{Y})$.

Proof: Assume w.l.o.g. that $\mathbf{X}, \mathbf{Y}$ are constructed in the same probability space and are independent. Define, for $t \in (0,1)$,

$$\mathbf{X}(t) = (1-t)^{1/2}\mathbf{X} + t^{1/2}\mathbf{Y}. \tag{3.1.1}$$

Then, with $'$ denoting differentiation with respect to t, we have $X_i'(t) = -(1-t)^{-1/2}X_i/2 + t^{-1/2}Y_i/2$. With $\phi(t) = Ef(\mathbf{X}(t))$, we get that

$$\phi'(t) = \sum_{i=1}^{n} E(\partial_i f(\mathbf{X}(t))X_i'(t))\,. \tag{3.1.2}$$

Now, by the independence of $\mathbf{X}$ and $\mathbf{Y}$,

$$EX_j(t)X_i'(t) = \frac{1}{2}E(Y_iY_j - X_iX_j)\,. \tag{3.1.3}$$

Thus, we can write (recall the conditional expectation representation and interpretation as orthogonal projection)

$$X_j(t) = \alpha_{ji}X_i'(t) + Z_{ji}\,, \tag{3.1.4}$$

where $Z_{ji} = Z_{ji}(t)$ is independent of $X_i'(t)$ and α_{ji} is proportional to the expression in (3.1.3). In particular, $\alpha_{ji} \ge 0, \le 0, = 0$ according to whether $(i,j) \in A, B, (A \cup B)^c$.

Using the representation (3.1.4), we can now write

$$
\begin{aligned}
E(X_i'(t)\partial_i f(\mathbf{X}(t))) &= E(X_i'(t)\partial_i f(\alpha_{1i}X_i'(t)+Z_{1i},\ldots,\alpha_{ni}X_i'(t)+Z_{ni})) \\
&=: M_i(\alpha_{1i},\ldots,\alpha_{ni};t)\,.
\end{aligned}
$$

We study the behavior of M as a function of the αs: note that

$$\frac{\partial M_i}{\partial \alpha_{ji}} = E(X_i'(t)^2 \partial_{ji} f(\cdots))$$

which is $\geq 0, \leq 0$ according to whether $(i,j)\in A$ or $(i,j)\in B$. Together with the computed signs of the αs, it follows that $M_i(\alpha_{1i},\ldots,\alpha_{ni}) \geq M_i(\mathbf{0})$. But due to the independence of the Z_{ij} on $X_i'(t)$, we have that $M(\mathbf{0})=0$. Hence, $\phi'(t)\geq 0$, implying $\phi(1)\geq\phi(0)$, and the theorem. □

Corollary 3 (Slepian's inequality). *Let $\mathbf{X},\mathbf{Y}$ be centered Gaussian vectors. Assume that $EX_i^2 = EY_i^2$ and $EX_iXj \geq EY_iY_j$ for all $i\neq j$. Then $\max_i X_i$ is stochastically dominated by $\max_i Y_i$, i.e., for any $x\in\mathbb{R}$,*

$$P(\max_i X_i > x) \leq P(\max_i Y_i > x)\,.$$

In particular, $E\max_i X_i \leq E\max_i Y_i$.

Of course, at the cost of obvious changes in notation and replacing max by sup, the result continue to hold for separable centered Gaussian processes.
Proof of Corollary 3: Fix $x\in\mathbb{R}$. We need to compute $E\prod_{i=1}^n f(X_i)$ where $f(y)=1_{y\leq x}$. Let f_k denote a sequence of smooth, monotone functions on $\mathbb{R}$ with values in $[0,1]$ that converge monotonically to f. Define $F_k(\mathbf{x}) = \prod_{i=1}^n f_k(x_i)$; then $\partial_{ij}F_k(\mathbf{x})\geq 0$ for $i\neq j$. By Slepian's lemma, with $\tilde F_k = 1-F_k$ we have that $E\tilde F_k(Y)\leq E\tilde F_k(X)$. Now, take limits as $k\to\infty$ and use monotone convergence to conclude the stochastic domination. The claim on the expectation is obtained by integration (or by using the fact that $\sup_i X_i$ and $\sup_i Y_i$ can now be constructed on the same probability space so that $\sup_i X_i \leq \sup_i Y_i$). □

It is tempting, in view of Corollary 3, to claim that Theorem 3 holds for f non-smooth, as long as its distributional derivatives satisfy the indicated constraints. In fact, in the Ledoux–Talagrand book, it is stated that way. However, that extension is false, as noted by Hoffman-Jorgensen. Indeed, it suffices to take $\mathbf{Y}$ a vector of independent standard centered Gaussian (in dimension $d\geq 2$), $\mathbf{X}=(X,\ldots,X)$ where X is standard centered Gaussian, and take $f=-1_D$ where $D=\{\mathbf{x}\in\mathbb{R}^d : x_1=x_2=\ldots=x_d\}$. Then $Ef(\mathbf{Y})=0$, $Ef(\mathbf{X})=-1$ but the mixed distributional derivatives of f vanish.

The condition on equality of variances in Slepian's inequality is sometimes too restrictive. When dealing with EX^{sup}, it can be dispensed with. This was done (independently) by Sudakov and Fernique. The proof we bring is due to S. Chatterjee; its advantage is that it provides a quantitative estimate on the gap in the inequality.

Proposition 5. *Let* $\mathbf{X}, \mathbf{Y}$ *be centered Gaussian vectors. Define* $\gamma_{ij}^X = E(X_i - X_j)^2$, $\gamma_{ij}^Y = E(Y_i - Y_j)^2$. *Let* $\gamma = \max |\gamma_{ij}^X - \gamma_{ij}^Y|$. *Then,*
- $|EX^{\text{sup}} - EY^{\text{sup}}| \leq \sqrt{\gamma \log n}$.
- *If* $\gamma_{ij}^X \leq \gamma_{ij}^Y$ *for all* i, j *then* $EX^{\text{sup}} \leq EY^{\text{sup}}$.

As a preliminary step in the proof, we provide a very useful Gaussian integration by parts.

Lemma 8. *Let* $\mathbf{X}$ *be a centered Gaussian vector and let* F *be a smooth function with at most polynomial growth at infinity of its first derivatives. Then*

$$E(X_i F(\mathbf{X})) = \sum_j E(X_i X_j) E(\partial_j F(\mathbf{X})) \,.$$

Proof of Lemma 8: Assume first that $\mathbf{X}$ has non-degenerate covariance. Then,

$$EX_i F(\mathbf{X}) = C \int x_i F(\mathbf{x}) e^{-\mathbf{x}^T R_X^{-1} \mathbf{x}/2} d\mathbf{x} \,. \tag{3.1.5}$$

We will integrate by parts: note that

$$\partial_j e^{-\mathbf{x}^T R_X^{-1} \mathbf{x}/2} = -\sum_k R_X^{-1}(j,k) x_k e^{-\mathbf{x}^T R_X^{-1} \mathbf{x}/2} = -(R_X^{-1}\mathbf{x})_j e^{-\mathbf{x}^T R_X^{-1} \mathbf{x}/2} \,.$$

Hence,

$$\nabla e^{-\mathbf{x}^T R_X^{-1} \mathbf{x}/2} = -R_X^{-1} \mathbf{x} e^{-\mathbf{x}^T R_X^{-1} \mathbf{x}/2} \,.$$

Integrating by parts in (3.1.5) and using the last display we get

$$\begin{aligned} \int \mathbf{x} F(\mathbf{x}) e^{-\mathbf{x}^T R_X^{-1} \mathbf{x}/2} d\mathbf{x} &= -R_X \int F(\mathbf{x}) \nabla e^{-\mathbf{x}^T R_X^{-1} \mathbf{x}/2} d\mathbf{x} \\ &= R_X \int \nabla F(\mathbf{x}) e^{-\mathbf{x}^T R_X^{-1} \mathbf{x}/2} d\mathbf{x} \,, \end{aligned}$$

completing the proof in case R_X is non-degenerate. To see the general case, replace R_X by the non-degenerate $R_X + \epsilon I$ (corresponding to adding an independent centered Gaussian of covariance ϵI to $\mathbf{X}$), and then use dominated convergence. $\square$

Proof of Proposition 5: Fix $\beta \in \mathbb{R}$, $\mathbf{X}, \mathbf{Y}$ independent, and set $F_\beta(\mathbf{x}) = \frac{1}{\beta} \log \sum_i e^{\beta x_i}$. Set $\mathbf{Z}(t) = (1-t)^{1/2} \mathbf{X} + \sqrt{t} \mathbf{Y}$, and define $\phi(t) = EF_\beta(\mathbf{Z}(t))$. Then,

$$\phi'(t) = E \sum_i \partial_i F_\beta(\mathbf{Z}(t)) \left(Y_i / 2\sqrt{t} - X_i / 2\sqrt{1-t} \right) .$$

Using Lemma 8 we get

$$\begin{aligned} E(X_i \partial_i F_\beta(\mathbf{Z}(t))) &= \sqrt{1-t} \sum_j R_X(i,j) E \partial_{ij}^2 F_\beta(\mathbf{Z}(t)) \,, \\ E(Y_i \partial_i F_\beta(\mathbf{Z}(t))) &= \sqrt{t} \sum_j R_Y(i,j) E \partial_{ij}^2 F_\beta(\mathbf{Z}(t)) \,. \end{aligned}$$

Therefore,

$$\phi'(t) = \frac{1}{2}\sum_{i,j} E\partial^2_{ij}F_\beta(\mathbf{Z}(t))(R_Y(i,j) - R_X(i,j))\,.$$

A direct computation reveals that

$$\partial_i F_\beta(\mathbf{x}) = \frac{e^{\beta x_i}}{\sum_j e^{\beta x_j}} =: p_i(\mathbf{x}) > 0\,,$$
$$\partial^2_{ij}F_\beta(\mathbf{x}) = \begin{cases} \beta(p_i(\mathbf{x}) - p_i^2(\mathbf{x})) & i = j \\ -\beta p_i(\mathbf{x})p_j(\mathbf{x}) & i \neq j\,. \end{cases}$$

Thus, $\phi'(t)$ equals the expectation of

$$-\frac{\beta}{2}\sum_{i,j} p_i(\mathbf{Z}(t))p_j(\mathbf{Z}(t))(R_Y(i,j) - R_X(i,j)) + \frac{\beta}{2}\sum_i p_i(\mathbf{Z}(t))(R_Y(i,i) - R_X(i,i))\,.$$

Because $\sum_i p_i(\mathbf{x}) = 1$, we get that the second term in the last display equals $\beta/4$ times

$$\sum_{i,j} p_i(\mathbf{Z}(t))p_j(\mathbf{Z}(t))(R_Y(i,i) - R_X(i,i) + R_Y(j,j) - R_X(j,j))\,.$$

Combining, we get that $\phi'(t)$ equals $\beta/4$ times the expectation of

$$\sum_{i,j} p_i(\mathbf{Z}(t))p_j(\mathbf{Z}(t))\left(R_Y(i,i) + R_Y(j,j) - 2R_Y(i,j) - R_X(i,i) - R_X(j,j) + 2R_X(i,j)\right)$$
$$= \sum_{i,j} p_i(\mathbf{Z}(t))p_j(\mathbf{Z}(t))\left(\gamma^Y_{ij} - \gamma^X_{ij}\right)\,.$$

Thus, if $\gamma^X_{ij} \le \gamma^Y_{ij}$ for all i, j, we get that $\phi'(t) \ge 0$. In particular, $\phi(0) \le \phi(1)$. Taking $\beta \to \infty$ yields the second point in the statement of the proposition. To see the first point, note that $\max_i x_i = \frac{1}{\beta}\log e^{\beta \max x_i}$ and therefore $\max x_i \le F_\beta(\mathbf{x}) \le \max_i x_i + (\log n)/\beta$. Since $|\phi(1) - \phi(0)| \le \beta\gamma/4$, and therefore

$$\left|EX^{\text{sup}} - EY^{\text{sup}}\right| \le \beta\gamma/4 + (\log n)/\beta \le \sqrt{\gamma \log n}$$

where in the last inequality we chose (the optimal) $\beta = 2\sqrt{(\log n)/\gamma}$. $\square$

Exercise 7. Prove Kahane's inequality: if $F : \mathbb{R}_+ \to \mathbb{R}$ is concave of polynomial growth, $EX_iX_j \le EY_iY_j$ for all i, j, and if $q_i \ge 0$, then

$$E(F(\sum_{i=1}^n q_i e^{X_i - \frac{1}{2}R_X(i,i)})) \ge E(F(\sum_{i=1}^n q_i e^{Y_i - \frac{1}{2}R_Y(i,i)}))\,.$$

Hint: Repeat the proof above of the Sudakov-Fernique inequality using F instead of F_β.

4 Entropy and majorizing measures

In view of Borell's inequality, an important task we still need to perform is the control of the expectation of the maximum (over the parameters in T) of a "nice" Gaussian process. A hint at the direction one could take is the following real analysis lemma of Garsia, Rodemich and Rumsey.

Lemma 9. *[Garsia-Rodemich-Rumsey lemma] Let $\Psi : \mathbb{R}_+ \to \mathbb{R}_+$ and $p : [0,1] \to \mathbb{R}_+$ be increasing functions with p continuous, $p(0) = 0$ and $\Psi(\infty) = \infty$. Set $\Psi^{-1}(u) = \sup\{v : \Psi(v) \le u\}$ $(u \ge \Psi(0))$ and $p^{-1}(x) = \max\{v : p(v) \le x\}$ $(x \in [0, p(1)])$. Let $f : [0,1] \to \mathbb{R}$ be continuous. Set*

$$I(t) = \int_0^1 \Psi\left(\frac{|f(t) - f(s)|}{p(|t-s|)}\right) ds\,.$$

Assume that $B := \int_0^1 I(t)dt < \infty$. Then,

$$|f(t) - f(s)| \le 8 \int_0^{|t-s|} \Psi^{-1}\left(\frac{4B}{u^2}\right) dp(u). \tag{4.1.1}$$

This classical lemma can be used to justify uniform convergence (see e.g. Varadhan's book for the stochastic processes course at NYU, for both a proof and application in the construction of Brownian motion). For completeness, we repeat the proof.

Proof of Lemma 9: By scaling, it is enough to prove the claim for $s = 0, t = 1$. Because $B < \infty$ we have $I(t_0) \le B$ for some $t_0 \in [0,1]$. For $n \ge 0$, set $d_n = p^{-1}(p(t_n)/2)$ (thus, $d_n < t_n$) and $t_{n+1} < d_n$ (thus, $d_{n+1} < d_n$ and $t_{n+1} < t_n$) converging to 0 so that

$$I(t_{n+1}) \le \frac{2B}{d_n} \text{ and therefore} \le \frac{2B}{d_{n+1}} \tag{4.1.2}$$

and

$$\Psi\left(\frac{|f(t_{n+1}) - f(t_n)|}{p(|t_{n+1} - t_n|)}\right) \le \frac{2I(t_n)}{d_n} \text{ and therefore } \le \frac{4B}{d_n d_{n-1}} \le \frac{4B}{d_n^2}\,. \tag{4.1.3}$$

Note that t_{n+1} can be chosen to satisfy these conditions since $\int_0^{d_n} I(s)ds \le B$ (and hence the Lebesgue measure of $s \le d_n$ so that the inequality in (4.1.2) is violated is strictly less than $d_n/2$) while similarly the set of $s \le d_n$ for which the inequality in (4.1.3) is violated is strictly less than $d_n/2$ for otherwise

$$\begin{aligned} I(t_n) &= \int_0^1 \Psi\left(\frac{|f(s) - f(t_n)|}{p(|s - t_n|)}\right) ds \ge \int_0^{d_n} \Psi\left(\frac{|f(s) - f(t_n)|}{p(|s - t_n|)}\right) ds \\ &> \frac{2I(t_n)}{d_n} \cdot \frac{d_n}{2} = I(t_n)\,. \end{aligned}$$

Hence, such a t_{n+1} can be found. Now, from (4.1.3),

$$|f(t_{n+1}) - f(t_n)| \leq p(t_n - t_{n+1})\Psi^{-1}\left(\frac{4B}{d_n^2}\right),$$

while

$$p(t_n - t_{n+1}) \leq p(t_n) = 2p(d_n) \text{ equality from the definition of } d_n\,.$$

Therefore, since $2p(d_{n+1}) = p(t_{n+1}) \leq p(d_n)$ and therefore $p(d_n) - p(d_{n+1}) \geq p(d_n)/2$, we get that

$$p(t_n - t_{n+1}) \leq 4[p(d_n) - p(d_{n+1})]\,.$$

We conclude that

$$\begin{aligned}
|f(0) - f(t_0)| &\leq 4\sum_{n=0}^{\infty}[p(d_n) - p(d_{n+1})]\Psi^{-1}\left(\frac{4B}{d_n^2}\right) \\
&\leq 4\sum_{n=0}^{\infty}\int_{d_{n+1}}^{d_n} dp(u)\Psi^{-1}\left(\frac{4B}{u^2}\right)du = 4\int_0^1 dp(u)\Psi^{-1}\left(\frac{4B}{u^2}\right)du\,,
\end{aligned}$$

where the first inequality is due to the monotonicity of Ψ^{-1}. Repeating the argument on $|f(1) - f(t_0)|$ yields the lemma. $\square$

The GRR lemma is useful because it gives a uniform modulus of continuity (e.g., on approximation of a Gaussian process on $[0,1]$ using the RKHS representation) as soon as an integrability condition is met (e.g., in expectation). For our needs, a particularly useful way to encode the information it provides on the supremum is in terms of the intrinsic metric determined by the covariance. Set $d(s,t) = \sqrt{E(X_s - X_t)^2}$, choose $p(u) = \max_{|s-t|\leq u} d(s,t)$, and set $\Psi(x) = e^{x^2/4}$ with $\Psi^{-1}(x) = 2\sqrt{\log x}$. Unraveling the definitions, we have the following.

Corollary 4. *There is a universal constant C with the following properties. Let $\{X_t\}$ be a centered Gaussian process on $T = [0,1]$ with continuous covariance. Assume that*

$$A := \int_0^1 \sqrt{\log(2^{5/4}/u)}dp(u) < \infty\,.$$

Then

$$E\sup_{t\in[0,1]} X_t \leq CA\,,$$

Remark: The constant $2^{5/4}$ is an artifact of the proof, and we will see later (toward the end of the proof of Theorem 4) that in fact one may replace it by 1 at the cost of modifying C.

Proof of Corollary 4: By considering $\bar{X}_t := X_t - X_0$, we may and will assume that $X_0 = 0$. Further, using the RKHS representation, one may consider only finite n approximations, with almost surely continuous sample paths (this requires an extra equicontinuity argument that I leave as exercise, and that is a consequence of the same computations detailed below). Set

$$Z := 2 \int_0^1 \int_0^1 \exp\left(\frac{(X_s - X_t)^2}{4p^2|s-t|}\right) dsdt \,.$$

Then, $EZ \le \sqrt{2}$. By Lemma 9, we have

$$X^{\sup} \le 16 \int_0^1 \sqrt{\log\left(\frac{4Z}{u^2}\right)} dp(u) \,,$$

and therefore, since the function $\sqrt{\log(4x/u^2)}$ is concave in x,

$$EX^{\sup} \le 16 \int_0^1 \sqrt{\log(4\sqrt{2}/u^2)} dp(u) \,.$$

The conclusion follows. $\square$

Corollary 4 is a prototype for the general bounds we will develop next. The setup will be of T being a Hausdorff space with continuous positive covariance kernel $R : T \times T \to \mathbb{R}$. Introduce as before the intrinsic metric $d(s,t) = \sqrt{E(X_s - X_t)^2}$. We assume that T is totally bounded in the metric d (the previous case of T being compact is covered by this, but the current assumption allows us also to deal e.g. with T being a countable set).

Definition 7. *A probability measure μ on T is called a majorizing measure if*

$$\mathcal{E}_\mu := \sup_{t \in T} \int_0^\infty \sqrt{\log(1/\mu(B_d(t,r)))} dr < \infty \,.$$

Note the resemblance to the definition of A in Corollary 4; choosing $p(u) = u$ and taking the one dimensional Lebesgue measure on $T = [0,1]$ maps between the expressions.

The following generalizes Corollary 4.

Theorem 4 (Fernique). *There exists a universal constant K such that, for any majorizing measure μ,*

$$EX^{\sup} \le K \mathcal{E}_\mu \,.$$

We will later see that Theorem 4 is optimal in that a complementary lower bound holds for *some* majorizing measure μ.

Proof: By scaling, we may and will assume that $\sup_{s,t \in T} d(s,t) = 1$. The first step of the proof is to construct an appropriate discrete approximation of T. Towards this end, let μ be given, and for any n, let $\{t_i^{(n)}\}_{i=1}^{r_n}$ be a finite collection of distinct points in T so that, with $B_{i,n} := B_d(t_i^{(n)}, 2^{-(n+2)})$

and $B^s_{i,n} := B_d(t_i^{(n)}, 2^{-(n+3)}) \subset B_{i,n}$, we have $T \subset \cup_i B^s_{i,n}$ and $\mu(B_{i,n}) \geq \mu(B_{i+1,n})$. (Thus, we have created a finite covering of T by d-balls of radii $2^{-(n+3)}$, with roughly decreasing μ-volume.) We now use these to extract disjoint subsets of T as follows. Set $C_1^{(n)} = B_{1,n}$ and for $i = 2, \ldots, n$, set

$$C_i^{(n)} = \begin{cases} \emptyset, & B_{i,n} \bigcap \left(\cup_{j=1}^{i-1} C_j^{(n)} \right) \neq \emptyset, \\ B_{i,n} \text{ otherwise.} \end{cases}$$

In particular, every ball $B_{i,n}$ intersects some C_j with $j \leq i$.

We define $\pi_n : T \to \{t_i^{(n)}\}_i$ by setting $\pi_n(t)$ to be the first $t_i^{(n)}$ for which $t \in B^s_{i,n}$ and $C_i^{(n)} \neq \emptyset$. If no such i exists (i.e., $C_i^{(n)} = \emptyset$ for all $B^s_{i,n}$ that cover t), then let $i(t)$ be the first index i for which $B^s_{i,n}$ covers t, and let $j < i(t)$ be the maximal index so that $C_j^{(n)} \cap B_{i(t),n} \neq \emptyset$; set then $\pi_n(t) = t_j^{(n)}$.

Let $\mathcal{T}_n$ denote the range of the map π_n and let $\mathcal{T} = \cup_n \mathcal{T}_n$. Note that by construction,

$$d(t, \pi_n(t)) \leq 2^{-(n+3)} + 2 \cdot 2^{-(n+2)} \leq 2^{-n} \,. \tag{4.1.4}$$

(In the first case in the construction of $\pi_n(t)$, we get $2^{-(n+3)}$.)

Set $\mu_t^{(n)} := \mu(B_{\pi_n(t),n})$. We now claim that

$$\mu_t^{(n)} \geq \mu(B(t, 2^{-(n+3)})). \tag{4.1.5}$$

Indeed, in the first case of the construction of $\pi_n(t)$ we have $d(t, \pi_n(t)) \leq 2^{-(n+3)}$ and therefore $\mu_t^{(n)} = \mu(B(\pi_n(t), 2^{-(n+2)})) \geq \mu(B(t, 2^{-(n+3)}))$. In the second case, we have $d(t, t_{i(t)}) \leq 2^{-(n+3)}$ and therefore, by the monotonicity of $\mu(B_{i,n})$,

$$\mu(B(t, 2^{-(n+3)})) \leq \mu(B(t_{i(t)}, 2^{-(n+2)})) \leq \mu(B(\pi_n(t), 2^{-(n+2)})) = \mu_t^{(n)} \,.$$

In either case, (4.1.5) holds.

The heart of the proof of the theorem is the construction of an auxilliary process whose distance dominates that defined by d, and then apply the Sudakov-Fernique inequality. Toward this end, attach to each $s \in \mathcal{T}_n$ an independent standard random variable $\xi_s^{(n)}$, and define the process

$$Y_t = \sum_{n=1}^{\infty} 2^{-n} \xi^{(n)}_{\pi_n(t)} \,.$$

We are going to study the process $\{Y_t\}$ for $t \in T$ (in fact, it would suffice to consider it for $t \in \mathcal{T}$). We have

$$E(X_s - X_t)^2 \leq 6E(Y_s - Y_t)^2 \,. \tag{4.1.6}$$

Indeed, let $N = N(s,t)$ be chosen such that $2^{-N} \leq d(s,t) < 2^{-N+1}$. Then, by (4.1.4), we have that $\pi_n(t) \neq \pi_n(s)$ for $n \geq N + 1$. In particular,

$$E(Y_t - Y_s)^2 \geq 2 \sum_{n=N+1}^{\infty} 2^{-2n} = 2^{-2N+2}/6 \geq d(s,t)^2/6\,.$$

Thus, by the Sudakov–Fernique inequality Proposition 5, we have that $EY^{\sup} \geq EX^{\sup}/\sqrt{6}$. We now evaluate $EY^{\sup}$.

The argument is somewhat surprising. For some M we will obtain the uniform bound $EY_\tau \leq M$ for any random variable τ with values in T *that may depend on* $\{Y_t\}$. Taking as the law of τ (an approximation of) the law of the maximum, this will imply that $EY^{\sup} \leq M$.

Let τ be a random variable with values in T of law ν and write

$$EY_\tau = \sum_{n=1}^{\infty} 2^{-n} \sum_{s \in \mathcal{T}_n} E(\xi_s^{(n)} 1_{\pi_n(\tau)=s})\,. \tag{4.1.7}$$

Now, set $g(u) = \sqrt{\log(1/u)}$ and, recalling that $\mu_s^{(n)} = \mu(B_{\pi_n(s),n})$, write

$$\begin{aligned}
E(\xi_s^{(n)} 1_{\pi_n(\tau)=s}) &= E(\xi_s^{(n)} 1_{\pi_n(\tau)=s} 1_{\xi_s^{(n)} > \sqrt{2} g(\mu_s^{(n)})}) + E(\xi_s^{(n)} 1_{\pi_n(\tau)=s} 1_{\xi_s^{(n)} \leq \sqrt{2} g(\mu_s^{(n)})}) \\
&\leq E(\xi_s^{(n)} 1_{\xi_s^{(n)} > \sqrt{2} g(\mu_s^{(n)})}) + \sqrt{2} g(\mu_s^{(n)}) P(\pi_n(\tau) = s) \\
&= \frac{\mu_s^{(n)}}{\sqrt{2\pi}} + \sqrt{2} g(\mu_s^{(n)}) P(\pi_n(\tau) = s)\,,
\end{aligned}$$

where the last equality follows from the Gaussian estimate

$$E(\xi_s 1_{\xi_s > \sqrt{2} g(a)}) = \frac{1}{\sqrt{2\pi}} \int_{\sqrt{2} g(a)}^{\infty} x e^{-x^2/2} dx = \frac{a}{\sqrt{2\pi}}\,.$$

Therefore, substituting in (4.1.7) we get

$$\begin{aligned}
EY_\tau &\leq \frac{1}{\sqrt{2\pi}} + \sum_n 2^{-n+1/2} \sum_{s \in \mathcal{T}_n} g(\mu_s^{(n)}) P(\pi_n(\tau) = s) \\
&\leq \frac{1}{\sqrt{2\pi}} + \sum_n 2^{-n+1/2} \int_T g(\mu(B(t, 2^{-(n+3)}))) \nu(dt) \\
&\leq \frac{1}{\sqrt{2\pi}} + \int_T \nu(dt) \sum_n 2^{-n+1/2} g(\mu(B(t, 2^{-(n+3)})))\,,
\end{aligned}$$

where the next to last inequality used (4.1.5). Due to the monotonicity of g, we have that

$$\sum_n 2^{-n+1/2} g(\mu(B(t, 2^{-(n+3)}))) \leq 8\sqrt{2} \int_0^\infty g(\mu(B(t,u))) du\,.$$

Thus,

$$EY_\tau \leq C(1 + \int_T \nu(dt) \int_0^\infty g(\mu(B(t,u)) du) \leq C(1 + \sup_t \int_0^\infty g(\mu(B(t,u)) du)\,,$$

for a universal constant C independent of ν. On the other hand, there are at least two distinct points s_1, s_2 in T with $d(s_1, s_2) > 1/2$ and therefore $B(s_1, u) \cap B(s_2, u) = \emptyset$ for $u \le 1/4$. Therefore, for t equal at least to one of s_1 or s_2 we have $\mu(B(t, u)) \le 1/2$ for all $u < 1/4$. Therefore,

$$\sup_{t \in T} \int_0^\infty g(\mu(B(t, u))du \ge \frac{1}{4}\sqrt{\log 2}\,. \tag{4.1.8}$$

Thus, $1+\sup_t \int_0^\infty g(\mu(B(t, u))du \le C' \sup_t \int_0^\infty g(\mu(B(t, u))du$ for a universal C'. This completes the proof. $\square$

While majorizing measures are useful (and, as we will see, capture the right behavior), it is rather awkward to work with them. A bound can be obtained using the notion of metric entropy, as follows. Consider (T, d) as a metric space, and let $N(\epsilon)$ denote the number of ϵ balls needed to cover T. Let $H(\epsilon) = \log N(\epsilon)$ denote the metric entropy of T. We have the following.

Proposition 6 (Dudley). *Assume that $\int_0^\infty \sqrt{H(u)}du < \infty$. Then, there exist a universal constant K and a majorizing measure μ such that,*

$$\mathcal{E}_\mu \le K \int_0^\infty \sqrt{H(u)}du\,.$$

Proof: By scaling we may and will assume that $\sup_{s,t} d(s, t) = 1$. Let $N_n = N(2^{-n})$. Choose, for each n positive integer, a partition $\{B_{n,k}\}_{k=1}^{N_n}$ of T where each element of the partition is contained in a ball of radius 2^{-n}, and fix arbitrary probability measures $\nu_{n,k}$ supported on $B_{n,k}$. Set

$$\mu(A) = \frac{1}{2}\sum_{n=0}^{\infty} 2^{-n}\frac{1}{N_n}\sum_k \nu_{n,k}(A \cap B_{n,k})\,.$$

Now, for any $t \in T$ and $u \in (2^{-(n+1)}, 2^{-n})$,

$$\mu(B(t, u)) \ge \frac{1}{2^{n+3}N_{n+2}}\,.$$

Thus,

$$\int_0^1 \sqrt{\log(1/\mu(B(t, u)))}du \le \sum_n 2^{-n}\sqrt{\log(2^{-n}N_n)} \le C + 2\int_0^1 H(u)^{1/2}du\,.$$

On the other hand, $\int_0^1 H(u)^{1/2}du \ge \sqrt{\log 2}$ (using the existence of a pair of points which cannot be covered by one ball of radius < 1). $\square$

As stated before, the majorizing measure bound is in fact tight (up to multiplicative constants). The result is contained in the following theorem.

Theorem 5 (Talagrand). *There exists a universal constant K such that for some majorizing measure μ,*

$$EX^{\sup} \ge K\mathcal{E}_\mu\,.$$

The proof of the theorem is instructive in that it will emphasize the role of an appropriate tree construction based on the covariance R. Before beginning, we introduce some notation.

Let $g(u) = \sqrt{\log(1/u)}$. Define

$$\mathcal{E}_\mu(A) = \sup_{t\in A} \int_0^\infty g(\mu(B(t,u)))du\,, \quad \mathcal{E}(A) = \inf_{\mu\in M_1(A)} \mathcal{E}_\mu(A)\,.$$

(Here, $M_1(X)$ denotes the set of probability measures on X; we have obviously that $\mathcal{E}_\mu = \mathcal{E}_\mu(T)$.) Note that $\mathcal{E}(\cdot)$ is not necessarily monotone, and therefore we set

$$\alpha(A) = \sup_{B\subset A} \mathcal{E}(B)\,, \quad \alpha = \sup_{\emptyset\neq V\subset T \text{ finite}} \alpha(V)\,.$$

By definition, $\alpha(\cdot)$ is monotone.

The proof of the theorem involves several steps. The first two are relatively easy and their proof is, at this stage, omitted .

Lemma 10. *Let* $\mathbf{X} = (X_1,\ldots,X_n)$*, be i.i.d. standard Gaussians. Then* $EX^{\sup} \sim \sqrt{2\log n}$*. In particular, there exist universal constants* C, c *such that* $c\sqrt{\log n} \le EX^{\sup} \le C\sqrt{\log n}$*.*

(The equivalence here is asymptotic as $n \to \infty$.) The proof of Lemma 10 proceeds by writing the CDF of $X^{\sup}$ as product of the CDFs of the X_is.

Lemma 11. *Let* (T,d) *be a metric space, of diameter* D*, and let* $\{A_i\}_{i=1}^n$ *be a partition of* T*. Then there exists a nonempty subset* $I \subset \{1,\ldots,n\}$ *such that, for all* $i \in I$*,*

$$\alpha(A_i) \ge \alpha(T) - D\sqrt{2\log(1+|I|)}\,.$$

The heart of the combinatorial argument in the proof of Theorem 5 employs Lemma 11 in order to introduce a hierarchical structure in (T,d) when the latter is finite. In what follows, set

$$\beta_i(A) = \alpha(A) - \sup_{t\in A} \alpha(A\cap B(t,6^{-(i+1)})) \ge 0\,,$$

which measures the minimal gain in α over restrictions to balls of radii $6^{-(i+1)}$.

Lemma 12. *Suppose* (X,d) *is finite of diameter* $D \le 6^{-i}$*. Then, one can find an integer* $I \le |X|$ *and sets* $\{B_k \subset X\}_{k=1}^I$ *so that the following hold.*
- $\mathrm{diam}(B_k) \le 6^{-(i+1)}$,
- $d(B_k, B_j) \ge 6^{-(i+2)}$ if $k \neq j$,

and

- $$\alpha(B_k) + \beta_{i+1}(B_k) \ge \alpha(X) + \beta_i(X) - 6^{-(i-1)}(2+\sqrt{\log I})\,. \qquad (4.1.9)$$

We will prove Lemma 12 after we see that it directly implies a domination of $EX^{\sup_V}$, where $X^{\sup_V} = \sup_{t\in V} X_t$.

Proposition 7. *Assume that* $\mathrm{diam}(T) = 1$. *Then there exists a universal constant* C *such that for any* $V \subset T$ *finite,*

$$EX^{\sup_V} \geq C\alpha(V)\,.$$

Going from Proposition 7 to Theorem 5 involves only a limiting and density argument, that we will discuss explicitely only after the proof of Proposition 7. It is in the proof of Proposition 7 that a tree structure in T will be most apparent.

Proof of Proposition 7: We assume Lemma 12 as well as the preliminary Lemmas 10 and 11. Note first that by a straightforward computation, if (X,Y) is a two dimensional centered Gaussian vector then $E\max(X,Y) = \sqrt{E(X-Y)^2/2\pi}$ and thus we may and will assume that $1 = \mathrm{diam}(V) \leq 6^{-3}\alpha(V)$ (the constant 6^3 is of course arbitrary).

Set $M = \min\{m : \inf_{s,t\in V, d(s,t)>0} d(s,t) > 6^{-m}\}$. Using Lemma 12, construct a collection of families $\{\mathcal{B}_i\}_{i=0}^M$, where $\mathcal{B}_i$ is non-empty and consists of disjoint sets of diameter at most 6^{-i} (in fact, actually smaller than $6^{-(i+1)}$, at least distanced by $6^{-(i+2)}$. Do that so that the sets are nested, i.e. for each $B \in \mathcal{B}_i$ and $j < i$ there is a $B' \in \mathcal{B}_j$ with $B \subset B'$, and such that (4.1.9) in Lemma 12 holds at each level.

For $t \in V$, if $t \in B \subset \mathcal{B}_i$, define $B_t^i := B$, and set

$$N_t^i = \left|\left\{B \in \mathcal{B}_{i+1} :\ B \subset B_t^i\right\}\right|\,.$$

We then have from (4.1.9) that for any $B \in \mathcal{B}_{i+1}$ with $B \subset B_t^i$,

$$\alpha(B) + \beta_{i+1}(B) \geq \alpha(B_t^i) + \beta_i(B_t^i) - 6^{-(i-1)}(2 + \sqrt{\log N_t^i})\,. \tag{4.1.10}$$

Set now $\tilde{V} = \cap_{i=0}^M (\cup_{B\in\mathcal{B}_i} B)$ and, for $t \in \tilde{V}$, set $\Psi_t^k = \sum_{i=k}^{M-1} 6^{-(i+1)}\sqrt{\log N_t^i}$. By construction, $\tilde{V}$ is non empty and whenever B_t^M is non-empty (which is equivalent to $t \in \tilde{V}$) one has that $B_t^M = \{t\}$. Therefore, $\alpha(B_t^M) = \beta_M(B_t^M) = 0$. On the other hand, $\alpha(B_t^0) = \alpha(V)$ while $\beta_0(B_t^0) \geq 0$. We thus get by telescoping

$$\alpha(V) \leq \inf_{t\in\tilde{V}} \sum_{i=0}^{M-1} \left(\alpha(B_t^i) + \beta_i(B_t^i) - \alpha(B_t^{i+1}) - \beta_{i+1}(B_t^{i+1})\right)\,. \tag{4.1.11}$$

Therefore, using (4.1.10), one gets

$$\alpha(V) \leq 6^2 \inf_{t\in\tilde{V}} \Psi_t^0 + 2\sum_i 6^{-(i-1)} \leq 6^2 \inf_{t\in\tilde{V}} \Psi_t^0 + \frac{\alpha(V)}{2}\,,$$

where in the last inequality we used the assumption that $\alpha(V) \geq 6^3$. We conclude that

$$\alpha(V) \le 72 \inf_{t\in\tilde V} \Psi^0_t. \tag{4.1.12}$$

We still need to compare to the supremum of a Gaussian vector, and here too we use a comparison with a tree-like process. With $\{\xi_B\}$ i.i.d. standard Gaussians, set, for $t \in \tilde V$, $\xi^i_t = \xi_{B^i_t}$ and $Y^k_t = \sum_{i=k+1}^M 6^{-i}\xi^i_t$. Then, for $s,t \in \tilde V$ with $d(s,t) > 0$ there is an $i_0 \in (0,\dots,M)$ such that $d(s,t) \in (6^{-(i_0-1)}, 6^{-i_0})]$ and therefore, arguing as in the proof of the upper bound, we deduce that

$$E(Y^0_t - Y^0_s)^2 \le 2\sum_{i_0-1}^{\infty} 6^{-2i} \le \frac{2\cdot 6^4}{35} d(s,t)^2\,.$$

Hence for an explicit universal constant C,

$$E\max_{t\in\tilde V} Y^0_t \le CE\max_{t\in\tilde V} X_t\,. \tag{4.1.13}$$

We are left with our last task, which is to control the expectation of the maximum of $E\max_{t\in\tilde V} Y^0_t$ *from below*. Toward this end, note that for $B \in \mathcal{B}_{M-1}$ so that $B\cap\tilde V \ne \emptyset$, we have from Lemma 10 that for a universal constant $c > 0$,

$$E\max_{t\in B} Y^{M-1}_t \ge c\min_{t\in B} \Psi^{M-1}_t\,.$$

We will propagate this estimate inductively (backward), claiming that for any $B \in \mathcal{B}_k$ with $B\cap\tilde V \ne \emptyset$,

$$E\max_{t\in B} Y^k_t \ge c\min_{t\in B} \Psi^k_t\,. \tag{4.1.14}$$

(the universal constant c does *not* depend on k and is the same as in Lemma 10). Once we prove that, we have that it holds with $k = 0$ and then, using (4.1.12) and (4.1.14) in the first inequality,

$$\alpha(V) \le CE\max_{t\in\tilde V} Y^0_t \le CE\max_{t\in\tilde V} X_t \le CEX^{\sup}{}_V\,.$$

So assume that (4.1.14) holds for k, we show it holds for $k-1$. Fix $B \in \mathcal{B}_{k-1}$ so that $B\cap\tilde V \ne \emptyset$, and define $\mathcal{C}^B_k = \{B\cap\tilde V : B \in \mathcal{B}_k\}$. For $C \in \mathcal{C}^B_k$, set $\Omega_C = \{\xi_C > \xi_{C'},\ \text{for all } C' \in \mathcal{C}^B_k, C' \ne C\}$. Then, define τ^k_C so that $\max_{t\in C} Y^k_t =: Y_{\tau_C}$, and note that

$$\sup_{t\in B} Y^{k-1}_t \ge \sum_{C\in\mathcal{C}^B_k} 1_{\Omega_C}\left(Y_{\tau_C} + 6^{-k}\xi_C\right) = \sum_{C\in\mathcal{C}^B_k} 1_{\Omega_C} Y_{\tau_C} + 6^{-k}\sup_{C\in\mathcal{C}^B_k}\xi_C\,. \tag{4.1.15}$$

Because 1_{Ω_C} is independent of the Y^k_ts, we have (using that the ξ_Cs are identically distributed and hence $P(\Omega_C) = 1/|\mathcal{C}^B_k|$),

$$E\sum_{C\in\mathcal{C}^B_k} 1_{\Omega_C} Y_{\tau_C} = \frac{1}{|\mathcal{C}^B_k|}\sum_{C\in\mathcal{C}^B_k} E\sup_{t\in C} Y^k_t\,.$$

Using the induction hypothesis (recall that c there is determined from Lemma 10 and in particular is universal), we thus obtain that

$$E \sum_{C\in\mathcal{C}_k^B} 1_{\Omega_C} Y_{\tau_C} \geq \frac{1}{|\mathcal{C}_k^B|} \sum_{C\in\mathcal{C}_k^B} c \inf_{t\in C} \Psi_t^k \geq c \inf_{t\in B} \Psi_t^k \,.$$

Substituting in (4.1.15) and using Lemma 10, we conclude that

$$\sup_{t\in B} Y_t^k \geq c \inf_{t\in B} \Psi_t^k + c6^{-k}\sqrt{\log(|\mathcal{C}_k^B|)} \geq c\Psi_t^{k-1} \,.$$

This completes the proof. $\square$

Proof of Theorem 5: Fix subsets $T_n = \{t_i^n\} \subset T$ so that T_n is $2^{-(n+4)}$ dense in T. Fix k. Since $\mathcal{E}(T_k) \leq \alpha(T_k) \leq \alpha$, we may find a measure $\mu_k \in M_1(T_k)$ so that

$$\sup_{t\in T_k} \sum_i 2^{-i} g(\mu_k(B(t, 2^{-i}))) \leq 2\alpha \,. \tag{4.1.16}$$

Apply the construction in the proof of the upper bound to yield a nested family of subsets $\{C_{i,k}^{(n)}\} = \{B(t_{i,k}^n, 2^{-(n+2)})\}$ (with a map $\pi_n^k : T \mapsto \{t_{i,k}^n\} \subset T_k$ that is 2^{-n} dense) so that

$$\mu_k(C_{i,k}^{(n)}(t)) \leq \mu_k(B(t, 2^{-(n+3)})). \tag{4.1.17}$$

Define $\mu_{n,k} \in M_1(T_n)$ by $\mu_{n,k}(t_{i,k}^n) = Z_{n,k}^{-1}\mu_k(C_{i,k}^{(n)}(t))$ where $Z_{n,k}$ is a normalization constant. Fixing n and taking $k \to \infty$ yields a sequence of measures on the finite set T_n, that converges on subsequences. Let $\bar\mu_n$ be a subsequential limit point.

Fix now $t \in T$ and fix a further subsequence (in k) such that $\pi_k^n(t)$ converges to a limit point $\pi_n(t) \in T_n$ (and thus, equals the limit from some k onward). Call this k_0. Let $\tau_k(t)$ denote the smallest in lexicographic order $t_i^{(k)}$ with $d(t, t_i^{(k)}) \leq 2^{-(k+4)}$; Such a τ_k exists because $\{t_i^{(k_0)}\}$ is $2^{-(k_0+4)}$-dense. We get

$$\begin{aligned}\sum_n 2^{-n} g(\bar\mu_n(\pi_n(t)) &\leq \sum_n 2^{-n} \liminf_k g(Z_{k,n}^{-1}\mu_k(C_k^{(n)}(t))) \\ &\leq \sum_n 2^{-n} \liminf_k g(\mu_k(C_k^{(n)}(t))) \,,\end{aligned}$$

because $Z_{k,n} \leq 1$. Therefore,

$$\sum_n 2^{-n} g(\bar\mu_n(\pi_n(t)) \leq 16 \sum_n 2^{-(n+4)} \liminf_k g(\mu_k(B(\tau_k(t), 2^{-(n+4)}) \leq 32\alpha \,, \tag{4.1.18}$$

where the first inequality is due to (4.1.17) and the last inequality follows from (4.1.16). Define $\mu \in M_1(T)$ so that $\mu(A) = \sum_n 2^{-n}\bar\mu_n(A)$. Use that

$g(\mu(\{\pi_n(t)\}) \leq g(2^{-n}\bar{\mu}_n(\{\pi_n(t)\}) \leq g(2^{-n}) + g(\bar{\mu}_n(\{\pi_n(t)\}) \leq g(2^{-n})S$ and that $\pi_n(t) \subset B(t, 2^{-n})$ and therefore $\mu(B(t, 2^{-n}) \geq \mu(\{\pi_n(t)\})$, see (4.1.4), to conclude that

$$\begin{aligned}
\int_0^1 g(\mu(B(t,u))du &\leq \sum_n 2^{-n} g(\mu(B(t, 2^{-n})) \\
&\leq \sum_n 2^{-n} g(\mu(\{\pi_n(t)\}) \\
&\leq \sum_n 2^{-n} g(\bar{\mu}_n(\{\pi_n(t)\})) + \sum_n 2^{-n} g(2^{-n}) \\
&\leq 32\alpha + 2\sqrt{\log 2},
\end{aligned}$$

where the last inequality is due to (4.1.18). Using again (4.1.8), we conclude that the right side of the last display is bounded by $C\alpha$. Applying Proposition 7 then completes the proof. $\square$

Proof of Lemma 12: We construct a (nested) sequence of subsets of X as follows: Set $X_1 = X$ and for $1 \leq k < \kappa$, $X_k = X \setminus \cup_{j<k} A_j$. We next show how to construct subsets $\emptyset \neq B_k \subseteq A_k \subseteq X_k$ so that the following conditions hold:

$$\operatorname{diam}(A_k) \leq 6^{-(i+1)}, d(B_k, X_k \setminus A_k) \geq 6^{-(i+2)}, \text{and} \tag{4.1.19}$$

either

$$\alpha(B_k) + \alpha(X_k) \geq 2\alpha(A_k) \text{ (such } k \text{ belongs to } I_1) \tag{4.1.20}$$

or

$$\alpha(B_k) + \beta_{i+1}(B_k) \geq \alpha(X_k) \text{ (such } k \text{ belongs to } I_2)\,. \tag{4.1.21}$$

Indeed, given X_k, if there exists $t \in X_k$ so that

$$\alpha(B(t, 6^{-(i+2)})) + \alpha(X_k) \geq 2\alpha(B(t, 2\cdot 6^{-(i+2)}) \cap X_k) \tag{4.1.22}$$

then set $A_k = B(t, 2\cdot 6^{-(i+2)}) \cap X_k$, $B_k = B(t, 6^{-(i+2)}) \cap X_k$, with obviously $k \in I_1$. Otherwise, let $t_0 \in X_k$ minimize $B(t, 2\cdot 6^{-(i+2)})$ and set $A_k = B(t, 3\cdot 6^{-(i+2)}) \cap X_k$, $B_k = B(t, 2\cdot 6^{-(i+2)}) \cap X_k$; this choice satisfies (4.1.21) (and hence $k \in I_2$) because

$$\beta_{i+1}(B_k) = \alpha(B_k) - \sup_{t \in B_k} \alpha(B(t, 6^{-(i+2)}) \cap X_k)$$

is larger than

$$\alpha(B_k) + \alpha(X_k) - 2\sup_{t \in B_k} \alpha(B(t, 6^{-(i+2)}) \cap X_k)$$

(just use that the inequality in (4.1.22) does not hold over X_k) and therefore

$$\beta_{i+1}(B_k) \geq \alpha(X_k) - \alpha(B_k)\,, \text{ for } k \in I_2\,.$$

With $\kappa = \min\{k : \alpha(X_k) < \alpha(X) - 2 \cdot 6^{-i}\}$, we have that $\alpha(X_k) \geq \alpha(X) - 2 \cdot 6^{i-1}$ for all $k < \kappa$. Set $C_\kappa = \cup_{k<\kappa} A_k$. Applying Lemma 11 with the two set partition (C_κ, X_κ) of X, one obtains that

$$\alpha(C_\kappa) \geq \alpha(X) - 6^{-i}\sqrt{2\log 2}\,. \tag{4.1.23}$$

(We used here that $\alpha(X_\kappa) < \alpha(X) - 2 \cdot 6^{-i} < \alpha(X) - 6^{-i}\sqrt{2\log 2}$.) Another application of Lemma 11 to C_k yields that there exists a subset I of $I_1 \cup I_2$ so that for all $k \in I$,

$$\alpha(A_k) \geq \alpha(C_k) - 6^{-i}\sqrt{2\log(1+|I|)} \geq \alpha(X) - 2 \cdot 6^{-i}(2 + \sqrt{\log(1+|I|)})\,. \tag{4.1.24}$$

(The last inequality used (4.1.23).) Therefore, for $k \in I \cap I_1$, using that (4.1.22), mplies that

$$\beta_i(X) \leq \alpha(X) - \alpha(B(t, 6^{-(i+2)})) \leq \alpha(X) - 2\alpha(A_k) + \alpha(X_k),$$

we conclude that

$$\begin{aligned}\alpha(B_k) + \beta_{i+1}(B_k) &\geq \alpha(B_k) \geq 2\alpha(A_k) - \alpha(X_k)\\ &\geq \alpha(X) + \beta_i(X) - 3(\alpha(X) - \alpha(A_k))\\ &\geq \alpha(X) + \beta_i(X) - 6^{-i}(2 + \sqrt{\log(1+|I|)})\,.\end{aligned}$$

(The last inequality follows from (4.1.24).) On the other hand, for $k \in I \cap I_2$, we get

$$\begin{aligned}\alpha(B_k) + \beta_{i+1}(B_k) &\geq \alpha(X_k) \geq \alpha(X) - 2 \cdot 6^{-i}\\ &\geq 2\alpha(X) - \alpha(A_k) - 2 \cdot 6^{-i}(4 + \sqrt{\log(1+|I|)})\\ &\geq \alpha(X) + \beta_i(X) - 2 \cdot 6^{-i}(4 + \sqrt{\log(1+|I|)})\,,\end{aligned}$$

where the first inequality is due to (4.1.21), the second because $k < \kappa$, the third from (4.1.24), and the last from the definition of $\beta_i(X)$ and the monotonicity of $\alpha(\cdot)$:

$$\begin{aligned}\beta_i(X) &= \alpha(X) - \sup_{t \in A} \alpha(X \cap B(t, 6^{-(i+1)}))\\ &\leq \alpha(X) - \sup_{t \in A_k} \alpha(X \cap B(t, 6^{-(i+1)}))\\ &\leq \alpha(X) - \alpha(X_k \cap B(t_0, 3 \cdot 6^{-(i+2)})) = \alpha(X) - \alpha(A_k)\,.\end{aligned}$$

This completes the proof. $\square$

Proof of Lemma 11: Order the A_is in decreasing order of their α measure. Fix $V \subset T$ and fix probability measures μ_i supported on $A_i \cap V$ and $a_i = 1/(i+1)^2$. Define μ supported on V by

$$\mu(A) = \frac{\sum_{i=1}^n a_i \mu_i(A \cap A_i)}{\sum_{i=1}^n a_i} \geq \sum_{i=1}^n a_i \mu_i(A).$$

Then, for $t \in A_i \cap V$,

$$\int_0^\infty g(\mu(B(t,d))dr \le \int_0^D g(a_i\mu_i(B(t,r)))dr \le Dg(a_i) + \int_0^D g(\mu_i(B(t,r)))dr\,.$$

In particular,

$$\mathcal{E}_\mu(V) \le \sup_i \left(Dg(a_i) + \mathcal{E}_{\mu_i}(A_i \cap V)\right) = Dg(a_j) + \mathcal{E}_{\mu_j}(A_j \cap V)$$

for some $j \in \{1,\ldots,n\}$. Thus,

$$\mathcal{E}(V) \le Dg(a_j) + \mathcal{E}(A_j \cap V) \le Dg(a_j) + \mathcal{E}(A_j)\,.$$

Taking $I = \{1,\ldots,j\}$ and using the monotonicity of $\alpha(A_i)$ completes the proof. $\square$

5 Branching Random Walks

Branching random walks (BRWs), and their continuous time counterparts, branching Brownian motions (BBMs), form a natural model that describe the evolution of a population of particles where spatial motion is present. Groundbreaking work on this, motivated by biological applications, was done in the 1930's by Kolmogorov-Petrovsky-Piskounov and by Fisher. The model itself exhibit a rich mathematical structures; for example, rescaled limits of such processes lead to the study of superprocesses, and allowing interactions between particles creates many challenges when one wants to study scaling limits.

Our focus is slightly different: we consider only particles in $\mathbb{R}$, and are mostly interested in the atypical particles that "lead the pack". Surprisingly, this innocent looking question turns out to show up in unrelated problems, and in particular techniques developed to handle it show up in the study of the two dimensional Gaussian Free Field, through an appropriate underlying tree structure. For this reason, and also because it simplifies many proofs, we will restrict attention to Gaussian centered increments.

5.1 Definitions and models

We begin by fixing notation. Let $\mathcal{T}$ be a tree rooted at a vertex o, with vertex set V and edge set E. We denote by $|v|$ the distance of a vertex v from the root, i.e. the length of the geodesic (=shortest path, which is unique) connecting v to o, and we write $o \leftrightarrow v$ for the collection of vertices on that geodesic (including o and v). With some abuse of notation, we also write $o \leftrightarrow v$ for the collection of *edges* on the geodesic connecting o and v. Similarly, for $v, w \in V$, we write $\rho(v,w)$ for the length of the unique geodesic connecting v and w, and define $v \leftrightarrow w$ similarly. The nth generation of the tree is the collection $D_n := \{v \in V : |v| = n\}$, while for $v \in D_m$ and $n > m$, we denote by

$$D_n^v = \{w \in D_n : \rho(w,v) = n - m\}$$

the collection of descendants of v in D_n. Finally, the degree of the vertex v is denoted d_v.

Let $\{X_e\}_{e\in E}$ denote a family of independent (real valued) random variables attached to the edges of the tree $\mathcal{T}$, of law μ. As mentioned above, we consider μ to be the standard centered Gaussian law. For $v \in V$, set $S_v = \sum_{e\in o\leftrightarrow v} X_e$. The *Branching Random Walk* (BRW) is simply the collection of random variables $\{S_v\}_{v\in V}$. We will be interested in the *maximal displacement* of the BRW, defined as

$$M_n = \max_{v\in D_n} S_v \,.$$

In our treatment, we always assume that the tree $\mathcal{T}$ is a k-ary tree, with $k \geq 2$: $d_o = k$ and $d_v = k+1$ for $v \neq o$.

As mentioned above, we will only discuss the Gaussian case, but whenever possible the statements will be phrased in a way that extends to more general distributions. With this in mind, introduce the large deviations rate function associated with the increments:

$$I(x) = \sup_{\lambda\in\mathbb{R}}(\lambda x - \Lambda(\lambda)) = x^2/2\,, \quad \left(\Lambda(\lambda) = \log E_\mu(e^{\lambda X}) = \lambda^2/2\right)\,, \quad (5.1.1)$$

which is strictly convex and has compact level sets. Set $x^* = \sqrt{2\log k}$ to be the unique point so that $x^* > E_\mu(X)$ and $I(x^*) = \log k$. We then have $I(x^*) = \lambda^* x^* - \Lambda(\lambda^*)$ where $x^* = \Lambda'(\lambda^*)$ and $x^* = I'(\lambda^*) = \lambda^*$.

5.2 Warm up: getting rid of dependence

We begin with a warm-up computation. Note that M_n is the maximum over a collection of k^n variables, that are not independent. Before tackling computations related to M_n, we first consider the same question when those k^n variables are independent. That is, let $\{\tilde{S}_v\}_{v\in D_n}$ be a collection of i.i.d. random variables, with $\tilde{S}_v$ distributed like S_v, and let $\tilde{M}_n = \max_{v\in D_n} \tilde{S}_v$. We then have the following. (The statement extends to the Non-Gaussian, non-lattice case by using the Bahadur-Rao estimate.)

Theorem 6. *With notation as above, there exists a constant C so that*

$$P(\tilde{M}_n \leq \tilde{m}_n + x) \to \exp(-Ce^{-I'(x^*)x})\,, \quad (5.2.1)$$

where

$$\tilde{m}_n = nx^* - \frac{1}{2I'(x^*)}\log n. \quad (5.2.2)$$

In what follows, we write $A \sim B$ if A/B is bounded above and below by two universal positive constants (that do not depend on n).

Proof. The key is the estimate, valid for $a_n = o(\sqrt{n})$,

$$P(\tilde{S}_v > nx^* - a_n) \sim \frac{C}{\sqrt{n}} \exp(-nI(x^* - a_n/n)), \tag{5.2.3}$$

which is trivial in the Gaussian case. We have

$$nI(x^* - a_n/n) = nI(x^*) - I'(x^*)a_n + o(1).$$

Therefore, recalling that $I(x^*) = \log k$,

$$\begin{aligned} P(\tilde{M}_n \leq nx^* - a_n) &\sim \left(1 - \frac{C}{k^n\sqrt{n}} e^{I'(x^*)a_n + o(1)}\right)^{k^n} \\ &\sim \exp(-Ce^{I'(x^*)a_n + o(1)}/\sqrt{n}). \end{aligned}$$

Choosing now $a_n = \log n/2I'(x^*) - x$, one obtains

$$P(\tilde{M}_n \leq m_n + x) \sim \exp(-Ce^{-I'(x^*)x + o(1)}).$$

The claim follows. □

Remark 1. With some effort, the constant C can also be evaluated to be $1/\sqrt{2\pi}x^*$, but this will not be of interest to us. On the other hand, the constant in front of the $\log n$ term will play an important role in what follows.

Remark 2. Note the very different asymptotics of the right and left tails: the right tail decays exponentially while the left tail is doubly exponential. This is an example of extreme distribution of the Gumbel type.

5.3 BRW: the law of large numbers

As a further warm up, we will attempt to obtain a law of large numbers for M_n. Recall, from the results of Section 5.2, that $\tilde{M}_n/n \to x^*$. Our goal is to show that the same result holds for M_n.

Theorem 7 (Law of Large Numbers). *We have that*

$$\frac{M_n}{n} \to_{n\to\infty} x^*, \quad \textit{almost surely} \tag{5.3.1}$$

Proof. While we do not really need in what follows, we remark that the almost sure convergence can be deduced from the subadditive ergodic theorem. Indeed, note that each vertex in D_n can be associated with a word $a_1 \dots a_n$ where $a_i \in \{1, \dots, k\}$. Introduce an arbitrary (e.g., lexicographic) order on the vertices of D_n, and define

$$v_m^* = \min\{v \in D_m : S_v = \max_{w \in D_m} S_w\}\,.$$

For $n > m$, write

$$M_n^m = \max_{w \in D_n^{v_m^*}} S_w - S_{v_m^*}\,.$$

We then have, from the definitions, that $M_n \geq M_m + M_n^m$, and it is not hard to check that M_n possesses all moments (see the first and second moment arguments below). One thus concludes, by applying the subadditive ergodic theorem (check the stationarity and ergodicity assumptions, which here follow from independence!), that $M_n/n \to c$, almost surely, for some constant c. Our goal is now to identify c.

The upper bound Let $\bar{Z}_n = \sum_{v \in D_n} \mathbf{1}_{S_v > (1+\epsilon)x^* n}$ count how many particles, at the nth generation, are at location greater than $(1+\epsilon)nx^*$. We apply a first moment method: we have, for any $v \in D_n$, that

$$E\bar{Z}_n = k^n P(S_v > n(1+\epsilon)x^*) \leq k^n e^{-nI((1+\epsilon)x^*)}\,,$$

where we applied Chebyshev's inequality in the last inequality. By the strict monotonicity of I at x^*, we get that $E\bar{Z}_n \leq e^{-nc(\epsilon)}$, for some $c(\epsilon) > 0$. Thus,

$$P(M_n > (1+\epsilon)nx^*) \leq E\bar{Z}_n \leq e^{-c(\epsilon)n}\,.$$

It follows that

$$\limsup_{n\to\infty} \frac{M_n}{n} \leq x^*\,, \quad \text{almost surely}\,.$$

The lower bound A natural way to proceed would have been to define

$$\underline{Z}_n = \sum_{v \in D_n} \mathbf{1}_{S_v > (1-\epsilon)x^* n}$$

and to show that with high probability, $\underline{Z}_n \geq 1$. Often, one handles this via the second moment method: recall that for any nonegative, integer valued random variable Z,

$$EZ = E(Z\mathbf{1}_{Z\geq 1}) \leq (EZ^2)^{1/2}(P(Z \geq 1))^{1/2}$$

and hence

$$P(Z \geq 1) \geq \frac{(EZ)^2}{E(Z^2)}. \tag{5.3.2}$$

In the case of independent summands, we obtain by this method that

$$P(\tilde{M}_n \geq (1-\epsilon)x^* n) \geq \frac{k^{2n} P(\tilde{S}_v \geq (1-\epsilon)x^* n)^2}{k^n(k^n-1)P(\tilde{S}_v \geq (1-\epsilon)x^* n)^2 + k^n P(\tilde{S}_v \geq (1-\epsilon)x^* n)}\,.$$

Since (e.g., by Cramer's theorem of large deviations theory),

$$\alpha_n := k^n P(\tilde{S}_v \geq (1-\epsilon)x^* n) \to \infty\,, \quad \text{exponentially fast}$$

one obtains that

$$P(\tilde{M}_n \geq (1-\epsilon)x^* n) \geq \frac{1}{\frac{k^n-1}{k^n} + 1/\alpha_n} \geq 1 - e^{-c'(\epsilon)n}\,,$$

implying that

$$\liminf \frac{\tilde{M}_n}{n} \geq x^*\,, \quad \text{almost surely.}$$

Any attempt to repeat this computation with M_n, however, fails, because the correlation between the events $\{S_v > nx^*(1-\epsilon)\}$ and $\{S_w > nx^*(1-\epsilon)\}$ with $v \neq w$ is too large (check this!). Instead, we will consider different events, whose probability is similar but whose correlation is much smaller. Toward this end, we keep track of the trajectory of the ancestors of particles at generation n. Namely, for $v \in D_n$ and $t \in \{0, \ldots, n\}$, we define the ancestor of v at levet t as $v_t := \{w \in D_t : \rho(v, w) = n - t\}$. We then set $S_v(t) = S_{v_t}$, noting that $S_v = S_v(n)$ for $v \in D_n$. We will later analyze in more detail events involving $S_v(t)$, but our current goal is only to prove a law of large numbers. Toward this end, define, for $v \in D_n$, the event

$$B_v^\epsilon = \{|S_v(t) - x^* t| \leq \epsilon n, t = 1, \ldots, n\}\,.$$

We now recall a basic large deviations result.

Theorem 8 (Varadhan, Mogulskii). *With notation and assumption as above,*

$$\lim_{\epsilon\to 0} \limsup_{n\to\infty} \frac{1}{n} \log P(B_v^\epsilon) = \lim_{\epsilon\to 0} \liminf_{n\to\infty} \frac{1}{n} \log P(B_v^\epsilon) = -I(x^*)\,.$$

Define now

$$Z_n = \sum_{v \in D_n} \mathbf{1}_{B_v^\epsilon}\,.$$

By theorem 8, we have that

$$EZ_n \geq e^{-c(\epsilon)n}\,. \tag{5.3.3}$$

To obtain an upper bound requires a bit more work. Fix a pair of vertices $v, w \in D_n$ with $\rho(v, w) = 2r$. Note that the number of such (ordered) pairs is $k^{n+r-1}(k-1)$. Now, using independence in the first equality, and homogenuity in the first inequality,

$$\begin{aligned}
P(B_v^\epsilon \cap B_w^\epsilon) &= E\Big(1_{\{|S_v(t)-x^*t| \leq \epsilon n, t=1,\ldots,n-r\}} \\
&\qquad \cdot \left(P\left(|S_v(t) - x^* t| \leq \epsilon n, t = n-r+1, \ldots, n | S_v(n-r)\right)\right)^2\Big) \\
&\leq P(|S_v(t) - x^* t| \leq \epsilon n, t = 1, \ldots, n-r) \\
&\qquad \cdot P(|S_v(t) - x^* t| \leq 2\epsilon n, t = 1, \ldots, r)^2\,.
\end{aligned}$$

Using Theorem 8, we then get that for all n large enough,

$$P(B_v^\epsilon \cap B_w^\epsilon) \leq e^{-(n-r)I(x^*)-2rI(x^*)+c(\epsilon)n},$$

where $c(\epsilon) \to_{\epsilon\to 0} 0$. Therefore,

$$EZ_n^2 \leq \sum_{r=0}^{n} k^{n+r} e^{-(n+r)I(x^*)+c(\epsilon)n} = e^{c(\epsilon)n}.$$

It follows from (5.3.2), (5.3.3) and the last display that, for any $\delta > 0$,

$$P(\exists v \in D_n : S_v \geq (1-\delta)x^* n) \geq e^{-o(n)}. \tag{5.3.4}$$

It seems that (5.3.4) is not quite enough to conclude. However, that turns out not to be the case. In the non-Gaussian case, one may proceed by truncating the tree at depth ϵn and use independence. In the Gaussian case, I will follow a suggestion of Eliran Subag, and directly apply Borell's inequality: indeed,

$$P(|M_n - EM_n| \geq \delta n) \leq 2e^{-\delta^2 n/2}.$$

Using (5.3.4), one deduces that $EM_n \geq (1-\delta)x^* n - \delta n$. Together with the upper bound and the arbitrariness of δ, this implies that $EM_n/n \to x^*$, and subadditivity (or directly, the use of Borell inequality together with the Borel-Cantelli lemma) then yield that also $M_n/n \to x^*$ almost surely. $\square$

5.4 A prelude to tightness: the Dekking-Host argument

The law of large number in Theorem 7 is weaker than the statement in Theorem 6 in two respects: first, no information is given in the latter concerning corrections from linear behavior, and second, no information is given, e.g., on the tightness of $M_n - EM_n$, let alone on its convergence in distribution. In this short section, we describe an argument, whose origin can be traced to [DH91], that will allow us to address the second point, once the first has been settled. Because it is a rather general argument, we drop in this subsection the assumption that the increments are Gaussian.

The starting point is the following recursion:

$$M_{n+1} \stackrel{d}{=} \max_{i=1}^{k}(M_{n,i} + X_i), \tag{5.4.1}$$

where $\stackrel{d}{=}$ denotes equality in distribution, $M_{n,i}$ are independent copies of M_n, and X_i are independent copies of X_e which are also independent of the collection $\{M_{n,i}\}_{i=1}^k$. Because of the independence and the fact that $EX_i = 0$, we have that

$$E\left(\max_{i=1}^{k}(M_{n,i} + X_i)\right) \geq E\left(\max_{i=1}^{k}(M_{n,i})\right).$$

Therefore,

$$EM_{n+1} \geq E\left(\max_{i=1}^{k}(M_{n,i})\right) \geq E\left(\max_{i=1}^{2}(M_{n,i})\right) .$$

Using the identity $\max(a,b) = (a+b+|a-b|)/2$, we conclude that

$$E(M_{n+1} - M_n) \geq \frac{1}{2} E|M_n - M'_n| , \tag{5.4.2}$$

where M'_n is an independent copy of M_n.

The importance of (5.4.2) cannot be over-estimated. First, suppose that there exists $K < \infty$ such that $X_e < K$, almost surely (this was the setup for which Dekking and Host invented this argument). In that case, we have that $EM_{n+1} - EM_n \leq K$, and therefore, using (5.4.2), we immediately see that the sequence $\{M_n - EM_n\}_{n\geq 1}$ is tight (try to prove this directly to appreciate the power of (5.4.2)). In making this assertion, we used the easy

Exercise 8. Prove that for every $C > 0$ there exists a function $f = f_C$ on R with $f(K) \to_{K\to\infty} 0$, such that if X, Y are i.i.d. with $E|X - Y| < C < \infty$, then $P(|X - EX| > K) \leq f(K)$.

However, (5.4.2) has implications even when one does not assume that $X_e < K$ almost surely for some K. First, it reduces the question of tightness to the question of computing an upper bound on $EM_{n+1} - EM_n$ (we will provide such a bound, of order 1, in the next section). Second, even without the work involved in proving such a bound, we have the following observation, due to [BDZ11].

Corollary 5. *For any $\delta > 0$ there exists a deterministic sequence $\{n_j^\delta\}_{j\geq 1}$ with $\limsup(n_j^\delta/j) \leq (1+\delta)$, so that the sequence $\{M_{n_j^\delta} - EM_{n_j^\delta}\}_{j\geq 1}$ is tight.*

Proof. Fix $\delta \in (0,1)$. We know that $EM_n/n \to x^*$. By (5.4.2), $EM_{n+1} - EM_n \geq 0$. Define $n_0^\delta = 0$ and $n_{j+1}^\delta = \min\{n \geq n_j^\delta : EM_{n+1} - EM_n \leq 2x^*/\delta\}$. We have that $n_{j+1}^\delta < \infty$ because otherwise we would have $\limsup EM_n/n \geq 2x^*/\delta$. Further, let $K_n = |\{\ell < n : \ell \notin \{n_j^\delta\}\}|$. Then, $EM_n \geq 2K_n x^*/\delta$, hence $\limsup K_n/n \leq \delta/2$, from which the conclusion follows. □

5.5 Tightness of the centered maximum

We continue to refine results for the BRW, in the spirit of Theorem 6; we will not deal with convergence in law, rather, we will deal with finer estimates on EM_n, as follows.

Theorem 9. *With notation and assumption as before, we have*

$$EM_n = nx^* - \frac{3}{2I'(x^*)} \log n + O(1) . \tag{5.5.1}$$

Remark 3. It is instructive to compare the logarithmic correction term in (5.5.1) to the independent case, see (5.2.2): the constant $1/2$ coming from the Bahadur-Rao estimate (5.2.3) is replaced by $3/2$. As we will see, this change is due to extra constraints imposed by the tree structure, and ballot theorems that are close to estimates on Brownian bridges conditioned to stay positive.

Theorem 9 was first proved by Bramson [Br78] in the context of Branching Brownian Motions. The branching random walk case was discussed in [ABR09], who stressed the importance of certain ballot theorems. Recently, Roberts [Ro11] significantly simplified Bramson's original proof. The proof we present combines ideas from these sources. To reduce technicalities, we consider only the case of Gaussian increments in the proofs.

Before bringing the proof, we start with some preliminaries related to Brownian motion and random walks with Gaussian increments.

Lemma 13. *Let $\{W_t\}_t$ denote a standard Brownian motion. Then*

$$P(W_t \in dx, W_s \geq -1 \text{ for } s \leq t) = \frac{1}{\sqrt{2\pi t}} e^{-x^2/2t}\left(1 - e^{-(x+2)/2t}\right) dx\,. \tag{5.5.2}$$

Note that the right side in (5.5.2) is of order $(x+2)/t^{3/2}$ for all $x = O(\sqrt{t})$ positive. Further, by Brownian scaling, for $y = O(\sqrt{t})$ positive,

$$P(W_t \in dx, W_s \geq -y \text{ for } s \leq t) = O\left(\frac{(x+1)(y+1)}{t^{3/2}}\right)\,. \tag{5.5.3}$$

Proof: This is D. André's reflection principle. Alternatively, the pdf in question is the pdf of a Brownian motion killed at hitting -1, and as such it solves the PDE $u_t = u_{xx}/2\,, u(t,-1) = 0$, with solution $p_t(0,x) - p_t(-2,x)$, where $p_t(x,y)$ is the standard heat kernel. □

Remark: An alternative approach to the proof of Lemma 13 uses the fact that a BM conditioned to remain positive is a Bessel(3) process. This is the approach taken in [Ro11].

We next bring a ballot theorem; for general random walks, this version can be found in [ABR08, Theorem 1]. We provide the proof in the case of Gaussian increments.

Theorem 10 (Ballot theorem). *Let X_i be iid random variables of zero mean, finite variance, with $P(X_1 \in (-1/2, 1/2)) > 0$. Define $S_n = \sum_{i=1}^n X_i$. Then, for $0 \leq k \leq \sqrt{n}$,*

$$P(k \leq S_n \leq k+1, S_i > 0, 0 < i < n) = \Theta\left(\frac{k+1}{n^{3/2}}\right)\,, \tag{5.5.4}$$

and the upper bound in (5.5.4) *holds for any $k \geq 0$.*

Here, we write that $a_n = \Theta(b_n)$ if there exist constanst $c_1, c_2 > 0$ so that

$$c_1 \le \liminf_{n\to\infty} \frac{a_n}{b_n} \le \limsup_{n\to\infty} \frac{a_n}{b_n} \le c_2 \,.$$

Proof (for standard Gaussian increments): The lower bound is an immediate consequence of Lemma 13. To see the upper bound, one approach would be to interpolate from the Brownian motion result. To make this work is however somewhat lengthy, so instead we provide a direct proof.

The idea is simple: staying positive for an interval of length $n/4$ when started at 0 has probability of order $c/\sqrt{n}$. When started at $z \in [k, k+1]$ (and looking at the reversed walk), the same event has probability of order $c(k+1)/\sqrt{n}$. Matching the two walks can't happen with probability larger than $c/\sqrt{n}$ by Gaussian estimates (for a random walk at time $n/2$).

To turn the above into a proof, we need the following facts. For $h \ge 0$, let $\tau_h = \min\{n \ge 1 : S_n \le -h\}$. Then,

$$P(\tau_h \ge n) \le c(h+1)/\sqrt{n}\,. \tag{5.5.5}$$

The case $h = 0$ in (5.5.5) is classical and in Feller's book, who uses the Sparre-Andersen bijection to compute the generating function of τ_0; in that case, one actually has that

$$\lim_{n\to\infty} n^{1/2} P(\tau_0 > n) = c\,.$$

The general case follows from this by choosing first α so that $P(S_{\alpha h^2} > h) \ge 1/3$ for all h. Then, by the FKG inequality,

$$P(S_{\alpha h^2} > h, \tau_0 > \alpha h^2) \ge P(S_{\alpha h^2} > h) P(\tau_0 > \alpha h^2) \ge c/h\,.$$

Therefore, from the Markov property,

$$\begin{aligned}\frac{c}{h} P(\tau_h > n) &\le P(S_{\alpha h^2} > h, \tau_0 > \alpha h^2, S_{\alpha h^2 + j} - S_{\alpha h^2} \ge -h, j = 1, \ldots, n) \\ &\le P(\tau_0 > \alpha h^2 + n) \le c n^{-1/2}\,,\end{aligned}$$

yielding (5.5.5).

To complete the proof of Theorem 10 in the case under consideration, define the reversed walk $S_i^r = S_n - S_{n-i}$, and consider the reverse hitting time $\tau_h^r = \min_{i: S_i^r < -h}$. Then, the event $\{S_i > 0, i = 1, \ldots, n, S_n \in [k, k+1]\}$ is contained in the event $\{\tau_0 > n/4, \tau_k^r > n/4, S_{3n/4} + S_{n/4}^r \in [k, k+1]\}$ and thus, using (conditional) independence,

$$\begin{aligned}&P(S_i > 0, i = 1, \ldots, n, S_n \in [k, k+1]\}) \\ &\le P(\tau_0 > n/4) P(\tau_k^r > n/4) \max_{y \in \mathbb{R}} P(S_{n/2} \in [y, y+1]) \le c \frac{k+1}{n^{3/2}}\,,\end{aligned}$$

as claimed. □

Exercise 9. Show that there exists a constant c_b so that

$$\lim_{x,y\to\infty} \lim_{n\to\infty} \frac{n^{3/2}}{xy} P^x(S_n \in [y, y+1], S_i > 0, i = 1, \ldots, n) = c_b\,. \tag{5.5.6}$$

A lower bound on the right tail of M_n Fix $y > 0$ independent of n and set

$$a_n = x^* n - \frac{3}{2I'(x^*)} \log n = x^* n - \frac{3}{2x^*} \log n.$$

For $v \in D_n$, define the event

$$A_v = A_v(y) = \{S_v \in [y + a_n - 1, y + a_n], S_v(t) \leq a_n t/n + y, t = 1, 2, \ldots, n\},$$

and set

$$Z_n = \sum_{v \in D_n} \mathbf{1}_{A_v}.$$

In deriving a lower bound on EM_n, we first derive a lower bound on the right tail of the distribution of M_n, using a second moment method. For this, we need to compute $P(A_v)$. Recall that we have $I(x^*) = \lambda^* x^* - \Lambda(\lambda^*) = \log k$, with $\lambda^* = I'(x^*) = x^*$. Introduce the new parameter λ_n^* so that

$$\lambda_n^* \frac{a_n}{n} - \Lambda(\lambda_n^*) = I(a_n/n) = (a_n/n)^2/2.$$

In the Gaussian case under consideration, we get that $\lambda_n^* = a_n/n$.

Define a new probability measure Q on $\mathbb{R}$ by

$$\frac{d\mu}{dQ}(x) = e^{-\lambda_n^* x + \Lambda(\lambda_n^*)},$$

and with a slight abuse of notation continue to use Q when discussing a random walk whose iid increments are distributed according to Q. Note that in our Gaussian case, Q only modifies the mean of P, not the variance.

We can now write

$$\begin{aligned} P(A_v) &= E_Q(e^{-\lambda_n^* S_v + n\Lambda(\lambda_n^*)} \mathbf{1}_{A_v}) \\ &\geq e^{-n[\lambda_n^*(a_n+y)/n - \Lambda(\lambda_n^*)]} Q(A_v) \qquad (5.5.7) \\ &= e^{-nI((a_n+y)/n)} Q(\tilde{S}_v \in [y-1, y], \tilde{S}_v(t) \geq 0, t = 1, 2, \ldots, n). \end{aligned}$$

where $\tilde{S}_v(t) = a_n t/n - S_v(t)$ is a random walk with iid increments whose mean vanishes under Q. Again, in the Gaussian case, the law of the increments is Gaussian and does not depend on n, and $\{\tilde{S}_v(t)\}_t$ is distributed like $\{S_v(t)\}_t$.

Applying Theorem 10, we get that

$$P(A_v) \geq C \frac{y+1}{n^{3/2}} e^{-nI((a_n+y)/n)}. \qquad (5.5.8)$$

Since

$$I((a_n + y)/n) = I(x^*) - I'(x^*) \left(\frac{3}{2I'(x^*)} \cdot \frac{\log n}{n} - \frac{y}{n} \right) + O\left(\left(\frac{\log n}{n} \right)^2 \right),$$

we conclude that

$$P(A_v) \geq C(y+1)k^{-n}e^{-I'(x^*)y}\,,$$

and therefore

$$EZ_n = k^n P(A_v) \geq c_1 e^{-I'(x^*)y}\,. \tag{5.5.9}$$

We next need to provide an upper bound on

$$EZ_n^2 = k^n P(A_v) + \sum_{v \neq w \in D_n} P(A_v \cap A_w) = EZ_n + k^n \sum_{s=1}^{n} k^s P(A_v \cap A_{v_s})\,, \tag{5.5.10}$$

where $v_s \in D_n$ and $\rho(v, v_s) = 2s$.

The strategy in computing $P(A_v \cap A_{v_s})$ is to condition on the value of $S_v(n-s)$. More precisely, with a slight abuse of notation, writing $I_{j,s} = a_n(n-s)/n + [-j, -j+1] + y$, we have that

$$\begin{aligned} & P(A_v \cap A_{v_s}) \\ \leq & \sum_{j=1}^{\infty} P(S_v(t) \leq a_n t/n + y, t = 1, 2, \ldots, n-s, S_v(n-s) \in I_{j,s}) \\ & \times \max_{z \in I_{j,s}} (P(S_v(s) \in [y + a_n - 1, y + a_n], S_v(t) \leq a_n(n-s+t)/n, t = 1, 2, \ldots, s | S_v(0) = z))^2\,. \end{aligned} \tag{5.5.11}$$

Repeating the computations leading to (5.5.8) (using time reversibility of the random walk) we conclude that

$$P(A_v \cap A_{v_s}) \leq \sum_{j=1}^{\infty} \frac{j^3(y+1)}{s^3(n-s)^{3/2}} e^{-(j+y)I'(x^*)} n^{3(n+s)/2n} k^{-(n+s)}\,. \tag{5.5.12}$$

Substituting in (5.5.10) and (5.5.11), and performing the summation over j first and then over s, we conclude that $EZ_n^2 \leq cEZ_n$, and therefore, using again (5.3.2),

$$P(M_n \geq a_n - 1) \geq P(Z_n \geq 1) \geq cEZ_n \geq c_0(y+1)e^{-I'(x^*)y} = c_0(y+1)e^{-x^* y}\,. \tag{5.5.13}$$

This completes the evaluation of a lower bound on the right tail of the law of M_n.

An upper bound on the right tail of M_n A subtle point in obtaining upper bounds is that the first moment method does not work directly - in the first moment one cannot distinguish between the BRW and independent random walks, and the displacement for these has a different logarithmic corrections (the maximum of k^n independent particles is larger).

To overcome this, note the following: a difference between the two scenarios is that at intermediate times $0 < t < n$, there are only k^t particles in the BRW setup while there are k^n particles in the independent case treated in Section 5.2. Applying the first moment argument at time t shows that there cannot be any BRW particle at time t which is larger than $x^*t + C\log n$, while this constraint disappears in the independent case. One thus expect that imposing this constraint in the BRW setup (and thus, pick up an extra $1/n$ factor from the ballot theorem 10) will modify the correction term.

Carrying out this program thus involves two steps: in the first, we consider an upper bound on the number of particles that never cross a barrier reflecting the above mentioned constraint. In the second step, we show that with high probability, no particle crosses the barrier. The approach we take combines arguments from [Ro11] and [ABR09]; both papers build on Bramson's original argument.

Turning to the actual proof, fix a large constant $\kappa > 0$, fix $y > 0$, and define the function

$$h(t) = \begin{cases} \kappa \log t, & 1 \le t \le n/2 \\ \kappa \log(n-t+1), & n/2 < t \le n\,. \end{cases} \tag{5.5.14}$$

Recall the definition $a_n = x^*n - \frac{3}{2I'(x^*)}\log n$ and let

$$\tau(v) = \min\{t > 0 : S_v(t) \ge a_n t/n + h(t) + y - 1\} \wedge n\,,$$

and $\tau = \min_{v\in D_n} \tau(v)$. (In words, τ is the first time in which there is a particle that goes above the line $a_n t/n + h(t) + y$.)

Introduce the events

$$B_v = \{S_v(t) \le a_n t/n + h(t) + y, 0 < t < n, S_v \in [y + a_n - 1, y + a_n]\}$$

and define $Y_n = \sum_{v\in D_n} \mathbf{1}_{B_v}$. We will prove the following.

Lemma 14. *There exists a constant c_2 independent of y so that*

$$P(B_v) \le c_2(y+1)e^{-I'(x^*)y}k^{-n}\,. \tag{5.5.15}$$

Proof of Lemma 14 (Gaussian case). Let $\beta_i = h(i) - h(i-1)$ (note that β_i is of order $1/i$ and therefore the sequence β_i^2 is summable). Define parameters $\tilde\lambda_n^*(i)$ so that

$$\tilde\lambda_n^*(i)\left(\frac{a_n}{n} + \beta_i\right) - \Lambda(\tilde\lambda_n^*(i)) = I\left(\frac{a_n}{n} + \beta_i\right) = \left(\frac{a_n}{n} + \beta_i\right)^2/2\,.$$

In the Gaussian case under consideration,

$$\tilde\lambda_n^*(i) = \lambda_n^* + \beta_i.$$

Define the new probability measures Q_i on $\mathbb{R}$ by

$$\frac{dP}{dQ_i}(x) = e^{-\tilde\lambda_n^*(i)x + \Lambda(\tilde\lambda_n^*(i))}\,,$$

and use $\tilde Q$ to denote the measure where X_i are independent of law Q_i. We have, similarly to (5.5.7),

$$P(B_v) = E_{\tilde{Q}}(e^{-\sum_{i=1}^n \tilde{\lambda}_n^*(i)X_i + \sum_{i=1}^n \Lambda(\tilde{\lambda}_n^*(i))}\mathbf{1}_{B_v})\,. \tag{5.5.16}$$

Using that $\sum_{i=1}^n \beta_i = 0$, one gets that on the event $S_n \in [y + a_n - 1, y + a_n]$,

$$\sum_{i=1}^n \tilde{\lambda}_n^*(i)X_i - \sum_{i=1}^n \Lambda(\tilde{\lambda}_n^*(i)) = nI((a_n + y)/n) + \sum_{i=1}^n \beta_i X_i + O(1)\,. \tag{5.5.17}$$

Substituting in (5.5.16), and using again that $\sum_{i=1}^n \beta_i = 0$, one gets

$$P(B_v) \leq Cn^{3/2}k^{-n}e^{-I'(x^*)y}E_{\tilde{Q}}(e^{-\sum_{i=1}^n \beta_i(X_i - a_n/n - \beta_i)}\mathbf{1}_{B_v})\,. \tag{5.5.18}$$

Using again that $\sum_{i=1}^n \beta_i = 0$, integration by parts yields $\sum \beta_i(X_i - (a_n + y)/n - \beta_i) = -\sum \tilde{S}(i)\tilde{\gamma}_i$, where under $\tilde{Q}$, $\tilde{S}(i)$ is a random walk with standard Gaussian increments, and $\tilde{\gamma}_i = \beta_{i+1} - \beta_i$. We thus obtain that

$$\begin{aligned} P(B_v) &\leq Cn^{3/2}k^{-n}e^{-I'(x^*)y}E_{\tilde{Q}}(e^{\sum_{i=1}^n \tilde{S}(i)\tilde{\gamma}_i}\mathbf{1}_{B_v}) \\ &\leq Cn^{3/2}k^{-n}e^{-I'(x^*)y}E_{\tilde{Q}}(e^{\sum_{i=1}^n \tilde{S}(i)\gamma_i}\mathbf{1}_{B_v})\,, \end{aligned} \tag{5.5.19}$$

where $\gamma_i = -\tilde{\gamma}_i = O(1/i^2)$. In terms of $\tilde{S}_i$, we can write

$$B_v = \{\tilde{S}(t) \leq y, \tilde{S}(n) \in y + [-1, 0]\}\,.$$

Without the exponential term, we have

$$\tilde{Q}(B_v) \leq c(y+1)n^{-3/2}\,.$$

Our goal is to show that the exponential does not destroy this upper bound. Let

$$\mathcal{C}_- = \{\exists t \in [(\log n)^4, n/2] : \tilde{S}(t) < -t^{2/3}\},$$
$$\mathcal{C}_+ = \{\exists t \in (n/2, n - (\log n)^4] : \tilde{S}(n) - \tilde{S}(t) < -(n-t)^{2/3}\}.$$

Then,

$$P(\mathcal{C}_- \cup \mathcal{C}_+) \leq 2\sum_{t=(\log n)^4}^{n/2} e^{-ct^{1/3}} \leq 2e^{-c(\log n)^{4/3}}\,.$$

Since $\gamma_i \sim 1/i^2$, one has $\sum_{(\log n)^4}^{n-(\log n)^4} \gamma_i \to_{n\to\infty} 0$ and further $\sum \tilde{S}(i)\gamma_i$ is Gaussian of zero mean and bounded variance. We thus obtain

$$\begin{aligned} &E_{\tilde{Q}}(e^{\sum_{i=1}^n \tilde{S}(i)\gamma_i}\mathbf{1}_{B_v}) \\ &\leq \frac{1}{n^2} + E_{\tilde{Q}}(e^{-\sum_{i=1}^{(\log n)^4}(\tilde{S}(i)\gamma_i + \tilde{S}(n-i)\gamma_{n-i})}\mathbf{1}_{B_v \cap \mathcal{C}_-^c \cap \mathcal{C}_+^c})\,. \end{aligned} \tag{5.5.20}$$

Denote by

$$B(z, z', t) = \{\tilde{S}(i) \leq z + y, i = 1, \ldots, n-t, \tilde{S}(n-t) \in [z'-1, z']\,.$$

We have, for $0 \le z, z', t < (\log n)^4$, by (5.5.6),

$$E_{\tilde{Q}}(B(z,z',t)) \le C \frac{(1+(z+y)_+)(1+(z+y-z')_+)}{n^{3/2}} .$$

We next decompose the expectation in the right side of (5.5.20) according to whether the curve $\tilde{S}(t)$ drops below the curve $\phi(t) = -(t^{2/3} \wedge (n-t)^{-2/3})$ or not. If it does not, then the exponential is of order 1. If it does, on the event $\mathcal{C}_-^c \cap \mathcal{C}_+^c$ it must do so either during the interval $[1, (\log n)^4]$ or $[n-(\log n)^4, n]$. Letting (t_-, z_-) denote the location and value of the first drop in the first interval and leting (t_+, z_+) denote the location and value of the last drop in the second interval, we then get

$$\begin{aligned}
&E_{\tilde{Q}}(e^{I''(a_n/n)\sum_{i=1}^{(\log n)^4}(\tilde{S}(i)\gamma_i + \tilde{S}(n-i)\gamma_{n-i})}\mathbf{1}_{B_v \cap \mathcal{C}_-^c \cap \mathcal{C}_+^c}) \\
&\le \frac{1}{n^2} + C\tilde{Q}(B_v) + \sum_{t_-,t_+=1}^{(\log n)^4} \sum_{z_- > t_-^{2/3}}^{(\log n)^4} \sum_{z_+ > t+^{-2/3}}^{(\log n)^4} e^{c(z_- + z_+)} e^{-cz_-^2/2t_-} e^{-cz_+^2/2t_+} \\
&\quad \times \max_{u,u' \in [0,1]} E_{\tilde{Q}}(B(u+z_-, u+z_+ + y, t_- + t_+)) \\
&\le C\frac{(y+1)}{n^{3/2}}.
\end{aligned}$$

Combined with (5.5.19), this completes the proof of Lemma 14. □

We need to consider next the possibility that $\tau = t < n$. Assuming that κ is large enough ($\kappa > 3/2I'(x^*)$ will do), an application of the lower bound (5.5.13) to the descendants of the parent of the particle v with $\tau_v < n$ reveals that for some constant c_3 independent of y,

$$E[Y_n | \tau < n] \ge c_3 .$$

(Recall that $Y_n = \sum_{v \in D_n} \mathbf{1}_{B_v}$.) We conclude that

$$P(\tau < n) \le \frac{E(Y_n)P(\tau < n)}{E(Y_n \mathbf{1}_{\tau<n})} = \frac{EY_n}{E(Y_n | \tau < n)} \le cEY_n . \tag{5.5.21}$$

One concludes from this and Lemma 14 that

$$P(M_n \ge a_n + y) \le P(\tau < n) + EY_n \le c_5(y+1)e^{-I'(x^*)y} . \tag{5.5.22}$$

In particular, this also implies that

$$EM_n \le x^* n - \frac{3}{2I'(x^*)} \log n + O(1) . \tag{5.5.23}$$

Remark 4. An alternative approach to the argument in (5.5.21), which is more in line with Bramson's original proof, is as follows. Note that

$$P(\tau \leq n - n^{\kappa'}) \leq \sum_{i=1}^{n-n^{\kappa'}} k^i P(S_n(i) \geq a_n i/n + h(i) + y) \leq C e^{-I'(x^*)y}\,,$$

where κ' can be taken so that $\kappa' \to_{\kappa\to\infty} 0$, and in particular for κ large we can have $\kappa' < 1$. Assume now κ large enough so that $\kappa' \leq 1/2$. For $t \geq n - n^{1/2}$, one repeats the steps in Lemma 14 as follows. Let N_t be the number of vertices $w \in D_t$ (out of k^t) whose path $S_w(s)$ crosses the barrier $(a_n s/n + h(s) + y - 1)$ at time $s = t$. We have

$$P(\tau = t) \leq EN_t \leq c(y+1)e^{-I'(x^*)yt/n} \frac{1}{(n-t)^{c_1\kappa - c_2}}$$

for appropriate constants c_1, c_2. Taking κ large enough ensures that

$$\sum_{t=n-n^{1/2}}^{n} EN_t \leq c(y+1)e^{-I'(x^*)y}\,.$$

Combining the last two displays leads to the same estimate as in the right side of (5.5.22), and hence to (5.5.23).

We finally prove a complementary lower bound on the expectation. Recall, see (5.5.13), that for any $y > 0$,

$$P(M_n \geq a_n(y)) \geq c(y+1)e^{-I'(x^*)y}\,,$$

where $a_n(y) = a_n + y$. In order to have a lower bound on EM_n that complements (5.5.23), we need only show that

$$\lim_{z\to-\infty} \limsup_{n\to\infty} \int_{-\infty}^{y} P(M_n \leq a_n(y))dy = 0\,. \tag{5.5.24}$$

Toward this end, fix $\ell > 0$ integer, and note that by the first moment argument used in the proof of the LLN (Theorem 7 applied to $\max_{w\in D_\ell}(-S_w)$), there exist positive constants c, c' so that

$$P(\min_{w\in D_\ell}(S_w) \leq -c\ell) \leq e^{-c'\ell}\,.$$

On the other hand, for each $v \in D_n$, let $w(v) \in D_\ell$ be the ancestor of v in generation ℓ. We then have, by independence,

$$P(M_n \leq -c\ell + (n-\ell)x^* - \frac{3}{2I'(x^*)}\log(n-\ell)) \leq (1-c_0)^{k^\ell} + e^{-c'\ell}\,,$$

where c_0 is as in (5.5.13). This implies (5.5.24). Together with (5.5.23), this completes the proof of Theorem 9. $\square$

5.6 Convergence of maximum and Gumbel limit law

We begin with a lemma, whose proof we only sketch.

Lemma 15. *There exists a constant $\bar{c}$ such that*

$$\lim_{y\to\infty}\limsup_{n\to\infty}\frac{e^{x^*y}}{y}P(M_n\geq m_n+y)=\lim_{y\to\infty}\liminf_{n\to\infty}\frac{e^{x^*y}}{y}P(M_n\geq m_n+y)=\bar{c}\,. \tag{5.6.25}$$

Note that the lemma is consistant with the upper and lower estimates on the right tail that we already derived. The main issue here is the convergence. *Proof (sketch):* The key new idea in the proof is a variance reduction step. To implement it, fix k (which will be taken function of y, going to infinity but so that $k << y$) and define, for any $v\in D_n$,

$$W_{v,k}=\max_{w\in D_k(v)}(S_w-S_v)\,.$$

Here, $D_k(v)$ denote the vertices in D_{n+k} that are descendants of $v\in D_n$. Now,

$$P(M_{n+k}>m_{n+k}+y)=P(\max_{v\in D_n}(S_v\geq m_n+(x^*k-W_{v,k}+y)))\,.$$

For each $v\in D_n$, we consider the event

$$A_v(n)=\{S_v(t)\leq tm_n/n+y,t=1,\ldots,n;S_v\geq m_n+(x^*k-W_{v,k}+y)\}\,.$$

Note that the event in $A_v(n)$ forces to have $W_{v,k}\geq x^*k+(m_n-S_v+y)\geq x^*k$, which (for k large) is an event of small probability. Now, one employs a curve $h(t)$ as described when deriving an upper bound on the right tail of M_n to show that

$$P(M_{n+k}>m_{n+k}+y)=(1+o_y(1))P(\cup_{v\in D_n}A_v(n))\,.$$

Next, using Exercise 9, one shows that

$$\lim_{y\to\infty}\lim_{n\to\infty}\frac{e^{x^*y}}{y}P(A_v(n))k^{-n}=\bar{c}\,,$$

for some constant $\bar{c}$. This is very similar to computations we already did.

Finally, note that conditionally on $\mathcal{F}_n=\sigma(S_v,v\in D_j,j\leq n)$, the events $\{W_{v,k}\geq x^*k+(m_n-S_v+y)\}_{v\in D_n}$ are independent. This introduces enough decorelation so that even when $v,w\in D_n$ are neighbors on the tree, one gets that

$$P(A_v(n)\cap A_w(n))\leq o_y(1)P(A_v(n))\,.$$

Because of that, defining $Z_n=\sum_{v\in D_n}1_{A_v(n)}$, one obtains that $EZ_n^2\leq(1+o_y(1))EZ_n+CEZ_n^2$ for some constant C and therefore, using that $\limsup_{n\to\infty}EZ_n\to_{y\to\infty}0$, one has

$$EZ_n \geq P(\cup_{v\in D_n} A_v(n)) \geq \frac{(EZ_n)^2}{EZ_n^2} \geq \frac{(EZ_n)^2}{EZ_n(1+o_y(1))} \geq EZ_n(1-o_y(1))\,.$$

Combining these three facts gives the lemma. □

We now finally are ready to state the following.

Theorem 11. *There exists a random variable Θ such that*

$$\lim_{n\to\infty} P(M_n \leq m_n + y) = E(e^{-\Theta e^{-\lambda^* y}})\,. \tag{5.6.26}$$

Thus, the law of $M_n - m_n$ converges to the law of a randomly shifted Gumbel distribution.

Remark: In fact, the proof we present will show that the random variable Θ is the limit in distribution of a sequence of random variables Θ_k. In reality, that sequence forms a martingale (the so called derivative martingale) with respect to $\mathcal{F}_k$, and the convergence is a.s.. We will neither need nor use that fact. For a proof based on the derivative martingale convergence, see Lalley and Sellke [LS87] for the BBM case and Aïdekon [Ai11] for the BRW case.

Proof (sketch): This time, we cut the tree at a fixed distance k from the root. Use that for n large, $\log(n+k) = \log(n) + O(1/n)$. Write

$$\begin{aligned}
P(M_{n+k} \leq m_{n+k} + y) &= E(\prod_{v\in D_k} 1_{(S_v + W_{v,k}(n)) \leq m_n + x^* k + y} \\
&\sim E(\prod_{v\in D_k} P(W_{v,k} \leq m_n + (x^* k - S_v) + y | S_v) \\
&\sim \epsilon(k) + E(\prod_{v\in D_k} \left(1 - \bar{c}(x^* k - S_v + y) e^{-\lambda^*(x^* k - S_v + y)}\right)\,,
\end{aligned}$$

where the symbol $a \sim b$ means that $a/b \to_{n\to\infty} 1$, and we used that with high probability $(1-\epsilon(k))$, $x^* k - S_v \geq x_k^* - S_k^*$ is of order $\log k$ and therefore we could apply Lemma 15 in the last equivalence. Fixing $\Theta_k = \bar{c}\sum_{v\in D_k}(x^* k - S_v)e^{-\lambda^*(x^* k - S_v)}$ and using that y is fixed while k is large, we conclude that

$$P(M_{n+k} \leq m_{n+k} + y) \sim \epsilon'(k) + E(e^{-\Theta_k e^{-\lambda^* y}})\,.$$

Since the right side does not depend on n, the convergence of the left side follows by taking $n \to \infty$ and then taking k large. Finally, the convergence also implies that the moment generating function of Θ_k converges, which in terms implies the convergence in distribution of Θ_k. □

5.7 Extremal process

We give a description of (a weak form of) a theorem due to [ABK11] and [ABBS11] in the Branching Brownian motion case and to [Ma11] in the (not necessarily Gaussian) BRW case, describing the distribution of the point process $\eta_n = \sum_{v\in D_n} \delta_{S_v - m_n}$. Our proof will follow the approach of Biskup and Louidor [BL13], and is tailored to the Gaussian setup we are considering.

We begin with a preliminary lemma. For a fixed constant R, set $\mathcal{M}_n(R) = \{v \in D_n : S_v > m_n - R\}$.

Lemma 16. *There exist functions $r(R) \to_{R\to\infty} \infty$ and $\epsilon(R) \to_{R\to\infty} 0$ so that*

$$\limsup_{n\to\infty} P(\exists u, v \in \mathcal{M}_n(R) : r(R) < d_T(u,v) < n - r(D)) \le \epsilon(R)\,, \quad (5.7.27)$$

where $d_T(u,v) = n - |a_{u,v}|$ is the tree distance between $u, v \in D_n$ and $a_{u,v}$ is the largest common ancestor of u and v.

The proof is immediate from the second moment computations we did; we omit details.

Fix now R (eventually, we will take $R \to \infty$ slowly with n) and define the *thinned* point process $\eta_n^s = \sum_{v\in D_n, S_v = \max_{w: d_T(v,w)\le R} S_w} \delta_{S_v - m_n}$. In words, η_n^s is the point process obtained by only keeping points that are leaders of their respective "clan", of depth R.

Theorem 12. *(a) The process η_n^s converges, as $n \to \infty$, to a random shift of a Poisson Point Process (PPP) of intensity $Ce^{-\lambda^* x}$, denoted η^s.*

(b) The process η_n converges, as $n \to \infty$, to a decorated version of η^s, which is obtained by replacing each point in η_s by a random cluster of points, independently, shifted around z.

A description of the decoration process is also available. We however will not bother with it. Instead, we will only sketch the proof of part (a) of Theorem 12.

Before the proof, we state a general result concerning invariant point processes, due to Liggett [Li78]. The setup of Liggett's theorem (narrowed to our needs; the general version replaces $\mathbb{R}$ by a locally compact second countable topological space) is a point process η on $\mathbb{R}$ (i.e., a random, integer valued measure on $\mathbb{R}$ which is finite a.s. on each compact), with each particle evolving individually according to a Markov kernel Q. For m a locally finite positive measure on $\mathbb{R}$, let μ_m denote the PPP of intensity m. For a random measure M on $\mathbb{R}$, we set $\bar\mu_M = \int \mu_m P(M \in dm)$ (a more suggestive notation would be $\bar\mu_M = E\mu_M$ where μ_M is, conditioned on M, a PPP of intensity M). We say that the law of a point process is invariant for Q if it does not change when each particle makes independently a move according to the Markov kernel Q.

One has the following. Througout, we assume that $Q^n(x,K) \to_{n\to\infty} 0$ uniformly in x for each fixed $K \subset\subset \mathbb{R}$.

Theorem 13 (Liggett [Li78]).

(a) $\bar\mu_M$ is invariant for Q iff $MQ = M$ in distribution.

(b) Every invariant probability measure is of the form $\bar{\mu}_M$ for some M.
(c) The extremal invariant probability measures for the point process are of the form μ_m with m satisfying $mQ = m$ iff $MQ = M$ in distribution implies $MQ = M$ a.s.
(d) In the special case where $Q(x, dy) = g(y-x)dy$ where g is a density function with finite exponential moments, condition (c) holds and all extreme invariant m are of the form $m(dx) = Ce^{-C'x}dx$, with C' depending on C and g.

(Part (d) of the theorem is an application of the Choquet-Deny theorem that characterizes the exponential distribution).
Proof of Theorem 12(a) (sketch): We write $\eta_n = \eta_n(\{S_v\})$ to emphasize that η_n depends on the Gaussian field $\{S_v\}_{v \in D_n}$. Note that due to the Gaussian structure,

$$\eta_n \stackrel{d}{=} \eta_n(\{\sqrt{1-1/n}S_v\} + \{\sqrt{1/n}S'_v\}), \tag{5.7.28}$$

where $\{S'_v\}$ is an independent copy of $\{S_v\}$ and the equality is in distribution. Now, $\{\sqrt{1/n}S'_v\}$ is a Gaussian field with variance *of order* 1, while $\sqrt{1-1/n}S_v = S_v - \frac{1}{2n}S_v + o(1)$.

Note that for any fixed $v \in D_n$, we have that $\max_{w:d_T(v,w)\le R}(S'_w - S'_v)/\sqrt{n} \le \delta(R)$ with probability going to 1 as $n \to \infty$, for an appropriate function $\delta(R) \to_{R\to\infty} 0$. By a diagonalization argument, one can then choose $R = R(n)$ so that

$$\max_{v,w\in D_n: S_v > m_n - R, d_T(v,w)\le R} (S'_w - S'_v)/\sqrt{n} \le \delta(R)$$

with probability going to 1 as $n \to \infty$.

Consider the right side of (5.7.28) as a (random) transformation on η_n; when restricting attention to the interval $(-R(n), \infty)$, which a.s. contains only finitely many points (first moment!), the transformation, with probability approaching 1 does the following:

- Replaces each point S_v by $S_v - x^*$.
- Adds to each *clan* an *independent* centered Gaussian random variable of variance 1.
- Adds a further small error of order $\delta(R_n)$

When thinning, one notes that the same transformation applies to the thinned process. Thus, any weak limit of the thinned process is invariant under the transformation that adds to each point an independent normal of mean $-x^*$. By the last point of Liggett's theorem, we conclude that any limit point of η_n^s is a random mixture of PPP with exponential intensity. The convergence of the maximum then determines both the exponent in the exponential intensity (must be λ^*) as well as the mixture (determined by the maximum). This completes the proof. □

6 The 2D discrete Gaussian Free Field

Take $V_N = ([0, N-1] \cap \mathbb{Z})^d$. Set $V_N^o = ((0, N-1) \cap \mathbb{Z})^d$ and identify all vertices in $\partial V_N = V_N \setminus V_N^o$, calling the resulting vertex the root of V_N. The collection of vertices thus obtained is denoted $\mathbf{V}_N$, and we take as edge set $\mathbf{E}_N$ the collection of all the (unordered) pairs (x, y) where either $x, y \in V_N^o$ and $|x-y|_1 = 1$ or $x \in V_N^o, y = o$ and there exists $z \in \partial V_N$ so that $|x-z|_1 = 1$. We thus obtain a sequence of graphs G_N where all vertices, except for the root, have degree $d^* = 2d$. The GFF on G_N is then defined as in Section 1, with a rescaling by $\sqrt{2d}$:

$$E\mathcal{X}_z^N \mathcal{X}_{z'}^N = E^z(\sum_{k=0}^{\tau-1} \mathbf{1}_{\{S_k = z'\}}) \,, \tag{6.1.1}$$

where $\{S_k\}$ is a simple random walk on G_N killed upon hitting o, with killing time τ. As before we set $\mathcal{X}_N^* = \max_{z \in \mathbf{V}_N} \mathcal{X}_z^N$.

Remark 5. As alluded to above, many authors, including the present one, refer to the field $\mathcal{X}_z^N$ as the GFF. I hope that this extra factor of $\sqrt{2d}$ will not cause too much confusion in what follows.

Recall from Borell's inequality that for $d \geq 3$, the sequence $\{\mathcal{X}_N^* - E\mathcal{X}_N^*\}_N$ is tight. On the other hand, for $d = 1$, the GFF is simply a random walk with standard Gaussian steps, conditioned to hit 0 at time N. In particular, $\mathcal{X}_N^*/\sqrt{N}$ scales like the maximum of a Brownian bridge, and thus $\mathcal{X}_N^* - E\mathcal{X}_N^*$ fluctuates at order $\sqrt{N}$. This leads us immediately to the question:

For $d = 2$, what is the order of $\mathcal{X}_N^*$ and are the fluctuations of order $O(1)$?

The rest of this section is devoted to the study of that question. In the rest of this subsection, we provide some a-priori comparisons and estimates.

Lemma 17. *For any $d \geq 1$, the sequence $E\mathcal{X}_N^*$ is monotone increasing in N.*

Proof. Let $N' > N$. For $z \in V_N^o$, write

$$\mathcal{X}_z^{N'} = E[\mathcal{X}_z^{N'} | \mathcal{F}_N] + \left(\mathcal{X}_z^{N'} - E[\mathcal{X}_z^{N'} | \mathcal{F}_N]\right) := A_z + B_z \,,$$

where $\mathcal{F}_N = \sigma(\mathcal{X}_{N'}^z : z \in V_{N'} \setminus V_N^o)$ and $\{A_z\}_{z \in V_N^o}$ and $\{B_z\}_{z \in V_N^o}$ are independent zero mean Gaussian fields. By the Markov property Lemma 6, we have that $\{B_z\}_{z \in V_N^o}$ is distributed like $\{\mathcal{X}_z^N\}_{z \in V_N^o}$. Therefore, since $E \max(X_\cdot + Y_\cdot) \geq E \max X_\cdot$, we conclude that $E\mathcal{X}_{N'}^* \geq E\mathcal{X}_N^*$. $\square$

The next lemma is an exercise in evaluating hitting probabilities for simple random walk.

Lemma 18 (GFF covariance, $d = 2$). *Fix $d = 2$. For any $\delta > 0$ there exists a $C = C(\delta)$ such that for any $v, w \in \mathbf{V}_N$ with $d(v, \partial V_N), d(w, \partial V_N) \geq \delta N$, one has*

$$\left| \mathbf{R}_{X^N}(v,w) - \frac{2}{\pi}\left(\log N - (\log \|v-w\|_2)_+\right) \right| \leq C\,. \tag{6.1.2}$$

Further,

$$\max_{x \in \mathbf{V}_N} \mathbf{R}_{X^N}(x,x) \leq (2/\pi)\log N + O(1)\,. \tag{6.1.3}$$

The proof of Lemma 18 can be found in [BDG01, Lemma 1] or [BZ11, Lemma 2.2].

Exercise 10. Using hitting estimates for simple random walks, prove Lemma 18.

6.2 The LLN for the 2D-GFF

We prove in this short section the Bolthausen-Deuschel-Giacomin LLN; our proof is shorter than theirs and involves comparisons with BRW.

Theorem 14. *Fix $d \geq 2$. Then,*

$$E\mathcal{X}_N^* \leq m_N + O(1)\,, \tag{6.2.1}$$

and

$$\lim_{N\to\infty} \frac{E\mathcal{X}_N^*}{m_N} = 1\,, \tag{6.2.2}$$

where

$$m_N = (2\sqrt{2/\pi})\log N - (3/4)\sqrt{2/\pi}\log\log N\,, \tag{6.2.3}$$

Further, for any $\epsilon > 0$ there exists a constant $c^* = c^*(\epsilon)$ so that for all large enough N,

$$P(|\mathcal{X}_N^* - m_N| \geq \epsilon m_N) \leq 2e^{-c^*(\epsilon)\log N}\,. \tag{6.2.4}$$

Proof. We note first that (6.2.4) follows from (6.2.2), (6.1.3) and Borell's inequality. Further, because of the monotonicity statement in Lemma 17, in the proof of (6.2.1) and (6.2.2) we may and will consider $N = 2^n$ for some integer n.

We begin with the introduction of a BRW that will be useful for comparison purposes. For $k = 0, 1, \ldots, n$, let $\mathcal{B}_k$ denote the collection of subsets of $\mathbb{Z}^2$ consisting of squares of side 2^k with corners in $\mathbb{Z}^2$, let $\mathcal{BD}_k$ denote the subset of $\mathcal{B}_k$ consisting of squares of the form $([0, 2^k - 1] \cap \mathbb{Z})^2 + (i2^k, j2^k)$. Note that the collection $\mathcal{BD}_k$ partitions $\mathbb{Z}^2$ into disjoint squares. For $x \in V_N$, let $\mathcal{B}_k(x)$ denote those elements $B \in \mathcal{B}_k$ with $x \in B$. Define similarly $\mathcal{BD}_k(x)$. Note that the set $\mathcal{BD}_k(x)$ contains exactly one element, whereas $\mathcal{B}_k(x)$ contains 2^{2k} elements.

Let $\{a_{k,B}\}_{k\geq 0, B\in\mathcal{BD}_k}$ denote an i.i.d. family of standard Gaussian random variables. The BRW $\{\mathcal{R}^N_z\}_{z\in V_N}$ is defined by

$$\mathcal{R}^N_z = \sum_{k=0}^n \sum_{B\in\mathcal{BD}_k(z)} a_{k,B}\,.$$

We again define $\mathcal{R}^*_N = \max_{z\in V_N} \mathcal{R}^N_z$. Note that $\mathcal{R}^N_z$ is a Branching random walk (with 4 descendants per particle). Further, the covariance structure of $\mathcal{R}^N_z$ respects a hierarchical structure on V^o_N: for $x,y\in V^o_N$, set $d_H(x,y) = \max\{k : y\notin \mathcal{BD}_k(x)\}$. Then,

$$\mathbf{R}_{\mathcal{R}_N}(x,y) = n - d_H(x,y) \leq n - \log_2 \|x-y\|_2\,. \tag{6.2.5}$$

We remark first that, as a consequence of the Markov property (see the computation in Lemma 17),

$$E\mathcal{X}^*_N \leq E \max_{x\in(N/2,N/2)+V_N} \mathcal{X}^{2N}_x\,.$$

Combined with Lemma 18 and the Sudakov-Fernique, we thus obtain that for some constant C independent of N,

$$E\mathcal{X}^*_N \leq \sqrt{\frac{2\log 2}{\pi}} E\mathcal{R}^*_N + C\,.$$

Together with computations for the BRW (the 4-ary version of Theorem 9), this proves (6.2.1).

To see (6.2.2), we dilute the GFF by selecting a subset of vertices in V_N. Fix $\delta>0$. Define $V^{\delta,1}_N = V_N$ and, for $k=2,\ldots,n-\log_2(1-\delta)n-1$, set

$$V^{\delta,k}_N = \{x\in V^{\delta,k-1}_N : |x-y|_\infty \geq \delta N/2^{n-k}, \forall y\in \cup_{B\in\mathcal{BD}_k}\partial B\}\,.$$

Note that $|V^{\delta,k}_N| \sim (1-\delta)^{2k}|V_N|$. We can now check that for $x,y\in V^{\delta,n(1-\log_2(1-\delta))}_N$, $\log_2|x-y|_2$ is comparable to $d_H(x,y)$. Obviously,

$$E\mathcal{X}^*_N \geq E(\max_{x\in V^{\delta,n(1-\log_2(1-\delta))}_N} \mathcal{X}^N_x)\,.$$

Applying the same comparison as in the upper bound, the right side is bounded below by the maximum of a diluted version of the BRW, to which the second moment argument used in obtaining the LLN for the BRW can be applied. (Unfortunately, a direct comparison with the BRW is not possible, so one has to repeat the second moment analysis. We omit further details since in Section 6.4 we will construct a better candidate for comparison, that will actually allow for comparison up to order 1.) We then get that for some universal constant C,

$$E\mathcal{X}^*_N \geq E(\max_{x\in V^{\delta,n(1-\log_2(1-\delta)}_N} \mathcal{X}^N_x) \geq \sqrt{\frac{2\log 2}{\pi}} E\mathcal{R}^*_{N(1-\delta)^{n+C}/C}\,.$$

This yields (6.2.2) after taking first $N\to\infty$ and then $\delta\to 0$. $\square$

6.3 A tightness argument: expectation is king

Our goal in this short section is to provide the following prelude to tightness, based on the Dekking–Host argument. It originally appeared in [BDZ11].

Lemma 19. *With $\mathcal{X}_N^{*'}$ an independent copy of $\mathcal{X}_N^*$, one has*

$$E|\mathcal{X}_N^{*'} - \mathcal{X}_N^*| \leq 2(E\mathcal{X}_{2N}^* - E\mathcal{X}_N^*)\,. \tag{6.3.1}$$

Note that by Lemma 17, the right side of (6.3.1) is positive. The estimate (6.3.1) reduces the issue of tightness of $\{\mathcal{X}_N^* - E\mathcal{X}_N^*\}_N$ to a question concerning precise control of $E\mathcal{X}_N^*$, and more specifically, to obtaining a lower bound on $E\mathcal{X}_{2N}^*$ which differs only by a constant from the upper bound (6.2.1) on $E\mathcal{X}_N^*$.

Exercise 11. Prove that if A_n is a sequence of random variables for which there exists a constant C independent of n so that $E|A_n - A_n'| \leq C$, where A_n' is an independent copy of A_n, then EA_n exists and the sequence $\{A_n - EA_n\}_n$ is tight.

In fact, Lemma 19 already yields a weak form of tightness.

Exercise 12. Combine Lemma 19 with the monotonicity statement (Lemma 17 and the LLN (Theorem 14) to deduce the existence of a deterministic sequence $N_k \to \infty$ so that $\{\mathcal{X}_{N_k^*} - E\mathcal{X}_{N_k}^*\}_k$ is tight.

(We eventually get rid of subsequences, but this requires extra estimates, as discussed in Lemma 20 below. The point of Exercise 12 is that tightness on subsequences is really a "soft" property.)
Proof of Lemma 19. By the Markov property of the GFF and arguing as in the proof of Lemma 17 (dividing the square V_{2N} into four disjoint squares of side N), we have

$$E\mathcal{X}_{2N}^* \geq E\max_{i=1}^{4} \mathcal{X}_N^{*,(i)} \geq E\max_{i=1}^{2} \mathcal{X}_N^{*,(i)}\,,$$

where $\mathcal{X}_N^{*,(i)}$, $i = 1, \ldots, 4$ are four independent copies of $\mathcal{X}_N^*$. Using again that $\max(a, b) = (a + b + |a - b|)/2$, we thus obtain

$$E\mathcal{X}_{2N}^* \geq E\mathcal{X}_N^* + E|\mathcal{X}_N^{*,(1)} - \mathcal{X}_N^{*,(2)}|/2\,.$$

This yields the lemma. □

We will use Lemma 19 in order to prove the following.

Theorem 15. *Let $\mathcal{X}_N^*$ denote the maximum of the two dimensional (discrete) GFF in a box of side N with Dirichlet boundary conditions. Then*

$$E\mathcal{X}_N^* = m_N + O(1)\,, \tag{6.3.2}$$

and the sequence $\{\mathcal{X}_N^* - E\mathcal{X}_N^*\}_N$ is tight.

The main task is the evaluation of *lower bounds* on $E\mathcal{X}_N^*$, which will be achieved by introducing a modified branching structure.

6.4 Expectation of the maximum: the modified BRW

We will now prove the following lemma, which is the main result of [BZ11].

Lemma 20. *With m_N as in* (6.2.3)*, one has*

$$E\mathcal{X}_N^* \geq m_N + O(1)\,. \tag{6.4.1}$$

Assuming Lemma 20, we have everything needed in order to prove Theorem 15.

Proof of Theorem 15. Combining Lemma 20 and (6.2.1), we have that $E\mathcal{X}_N^* = m_N + O(1)$. This yields (6.3.2). The tightness statement is now a consequence of Lemma 19, Exercise 11 and the fact that $m_{2N} - m_N$ is uniformly bounded. □

We turn to the main business of this section.

Proof of Lemma 20 (sketch). The main step is to construct a Gaussian field that interpolates between the BRW and the GFF, for which the second moment analysis that worked in the BRW case can still be carried out. Surprisingly, the new field is a very small variant of $\mathcal{R}_z^N$. We therefore refer to this field as the *modified branching random walk*, or in short MBRW.

We continue to consider $N = 2^n$ for some positive integer n and again employ the notation $\mathcal{B}_k$ and $\mathcal{B}_k(x)$. For $x, y \in \mathbb{Z}^2$, write $x \sim_N y$ if $x - y \in (N\mathbb{Z})^2$. Similarly, for $B, B' \subset V_N$, write $B \sim_N B'$ if there exist integers i, j so that $B' = B + (iN, jN)$. Let $\mathcal{B}_k^N$ denote the collection of subsets of $\mathbb{Z}^2$ consisting of squares of side 2^k with lower left corner in V_N. Let $\{b_{k,B}\}_{k\geq 0, B\in\mathcal{B}_k^N}$ denote a family of independent centered Gaussian random variables where $b_{k,B}$ has variance 2^{-2k}, and define

$$b_{k,B}^N = \begin{cases} b_{k,B}, & B \in \mathcal{B}_k^N\,, \\ b_{k,B'}, & B \sim_N B' \in \mathcal{B}_k^N\,. \end{cases}$$

The MBRW $\{\mathcal{S}_z^N\}_{z\in V_N}$ is defined by

$$\mathcal{S}_z^N = \sum_{k=0}^{n} \sum_{B\in\mathcal{B}_k(z)} b_{k,B}^N\,.$$

We will also need a truncated form of the MBRW: for any integer $k_0 \geq 0$, set

$$\mathcal{S}_z^{N,k_0} = \sum_{k=k_0}^{n} \sum_{B\in\mathcal{B}_k(z)} b_{k,B}^N\,.$$

We again define $\mathcal{S}_N^* = \max_{z\in V_N} \mathcal{S}_z^N$ and $\mathcal{S}_{N,k_0}^* = \max_{z\in V_N} \mathcal{S}_z^{N,k_0}$. The correlation structure of $\mathcal{S}$ respects a torus structure on V_N. More precisely, with $d^N(x,y) = \min_{z:\, z\sim_N y} \|x - z\|$, one easily checks that for some constant C independent of N,

$$|\mathbf{R}_{\mathcal{S}^N}(x,y) - (n - \log_2 d^N(x,y))| \leq C\,. \tag{6.4.2}$$

In particular, for points $x, y \in (N/2, N/2) + V_N$, the covariance of $\mathcal{S}_{2N}$ is comparable to that of $\mathcal{X}_{2N}$. More important, the truncated MBRW has the following nice properties. Define, for $x, y \in V_N$, $\rho_{N,k_0}(x,y) = E((\mathcal{S}_x^{N,k_0} - \mathcal{S}_y^{N,k_0})^2)$. The following are basic properties of ρ_{N,k_0}; verification is routine and omitted.

Lemma 21. *The function ρ_{N,k_0} has the following properties.*

$$\rho_{N,k_0}(x,y) \text{ decreases in } k_0. \tag{6.4.3}$$

$$\limsup_{k_0\to\infty} \limsup_{N\to\infty} \sup_{x,y\in V_N: d_N(x,y)\le 2^{\sqrt{k_0}}} \rho_{N,k_0}(x,y) = 0\,. \tag{6.4.4}$$

$$\begin{aligned} &\text{There is a function } g : \mathbb{Z}_+ \to \mathbb{R}_+ \text{ so that } g(k_0) \to_{k_0\to\infty} \infty \\ &\text{and, for } x, y \in V_N \text{ with } d^N(x,y) \ge 2^{\sqrt{k_0}}, \\ &\rho_{N,k_0}(x,y) \le \rho_{N,0}(x,y) - g(k_0)\,, \quad n > k_0. \end{aligned} \tag{6.4.5}$$

Equipped with Lemma 21, and using the Sudakov-Fernique Theorem, we have the following.

Corollary 6. *There exists a constant k_0 such that, for all $N = 2^n$ large,*

$$E\mathcal{X}_N^* \ge \sqrt{\frac{2\log 2}{\pi}} E\mathcal{S}_{N/4,k_0}^*\,. \tag{6.4.6}$$

Therefore, the proof of Lemma 20 reduces to the derivation of a lower bound on the expectation of the maximum of the (truncated) MBRW. This is contained in the following proposition, whose proof we sketch below.

Proposition 8. *There exists a function $f : \mathbb{Z}_+ \to \mathbb{R}_+$ such that, for all $N \ge 2^{2k_0}$,*

$$E\mathcal{S}_{N,k_0}^* \ge (2\sqrt{\log 2})n - (3/(4\sqrt{\log 2}))\log n - f(k_0)\,. \tag{6.4.7}$$

The proposition completes the proof of Lemma 20. □

Proof of Proposition 8 (sketch). Set $V_N' = V_{N/2} + (N/4, N/4) \subset V_N$ and define

$$\tilde{\mathcal{S}}_{N,k_0}^* = \max_{z\in V_N'} \mathcal{S}_z^{N,k_0}\,, \quad \tilde{\mathcal{S}}_N^* = \tilde{\mathcal{S}}_{N,0}^*\,.$$

Set

$$A_n = m_N\sqrt{\pi/2\log 2} = (2\sqrt{\log 2})n - (3/(4\sqrt{\log 2}))\log n\,.$$

An application of the second moment method (similar to what was done for the BRW, and therefore omitted) yields the following.

Proposition 9. *There exists a constant $\delta_0 \in (0,1)$ such that, for all N,*

$$P(\tilde{\mathcal{S}}_N^* \ge A_n) \ge \delta_0\,. \tag{6.4.8}$$

We now explain how to deduce Proposition 8 from Proposition 9. Our plan is to show that the left tail of $\tilde{\mathcal{S}}_N^*$ is decreasing exponentially fast; together with the bound (6.4.8), this will imply (6.4.7) with $k_0 = 0$. At the end of the proof, we show how the bound for $k_0 > 0$ follows from the case $k_0 = 0$. In order to show the exponential decay, we compare $\tilde{\mathcal{S}}_N^*$, after appropriate truncation, to four independent copies of the maximum over smaller boxes, and then iterate.

For $i = 1, 2, 3, 4$, introduce the four sets $W_{N,i} = [0, N/32)^2 + z_i$ where $z_1 = (N/4, N/4)$, $z_2 = (23N/32, N/4)$, $z_3 = (N/4, 23N/32)$ and $z_4 = (23N/32, 23N/32)$. (We have used here that $3/4 - 1/32 = 23/32$.) Note that $\cup_i W_{N,i} \subset V_N$, and that these sets are $N/4$-separated, that is, for $i \neq j$,

$$\min_{z \in W_{N,i}, z' \in W_{N,j}} d_\infty^N(x, y) > N/4 .$$

Recall the definition of $\mathcal{S}_z^N$ and define, for $n > 6$,

$$\bar{\mathcal{S}}_z^N = \sum_{k=0}^{n-6} \sum_{B \in \mathcal{B}_k(z)} b_{k,B}^N ;$$

note that

$$\mathcal{S}_z^N - \bar{\mathcal{S}}_z^N = \sum_{j=0}^{5} \sum_{B \in \mathcal{B}_{n-j}(z)} b_{n-j,B}^N .$$

Our first task is to bound the probability that $\max_{z \in V_N} (\mathcal{S}_z^N - \bar{\mathcal{S}}_z^N)$ is large. This will be achieved by applying Fernique's criterion in conjunction with Borell's inequality. We introduce some notation. Let $m(\cdot) = m_N(\cdot)$ denote the uniform probability measure on V_N (i.e., the counting measure normalized by $|V_N|$) and let $g : (0, 1] \to \mathbb{R}_+$ be the function defined by

$$g(t) = (\log(1/t))^{1/2} .$$

Set $\mathbf{G}_z^N = \mathcal{S}_z^N - \bar{\mathcal{S}}_z^N$ and

$$B(z, \epsilon) = \{z' \in V_N : E((\mathbf{G}_z^N - \mathbf{G}_{z'}^N)^2) \leq \epsilon^2\} .$$

Then, Fernique's criterion, Theorem 4, implies that, for some universal constant $K \in (1, \infty)$,

$$E(\max_{z \in V_N} \mathbf{G}_z^N) \leq K \sup_{z \in V_N} \int_0^\infty g(m(B(z, \epsilon))) d\epsilon . \tag{6.4.9}$$

For $n \geq 6$, we have, in the notation of Lemma 21,

$$E((\mathbf{G}_z^N - \mathbf{G}_{z'}^N)^2) = \rho_{N,n-5}(z, z') .$$

Therefore, there exists a constant C such that, for $\epsilon \geq 0$,

$$\{z' \in V_N : d_\infty^N(z,z') \le \epsilon^2 N/C\} \subset B(z,\epsilon)\,.$$

In particular, for $z \in V_N$ and $\epsilon > 0$,

$$m(B(z,\epsilon)) \ge ((\epsilon^4/C^2) \vee (1/N^2)) \wedge 1\,.$$

Consequently,

$$\int_0^\infty g(m(B(z,\epsilon)))d\epsilon \le \int_0^{\sqrt{C/N}} \sqrt{\log(N^2)}d\epsilon + \int_{\sqrt{C/N}}^{\sqrt{C}} \sqrt{\log(C^2/\epsilon^4)}d\epsilon < C_4\,,$$

for some constant C_4. So, from Fernique's criterion (6.4.9) we deduce that

$$E(\max_{z\in V_N}(\mathcal{S}_z^N - \bar{\mathcal{S}}_z^N)) \le C_4 K\,.$$

The expectation $E((\mathcal{S}_z^N - \bar{\mathcal{S}}_z^N)^2)$ is bounded in N. Therefore, using Borell's inequality, it follows that, for some constant C_5 and all $\beta > 0$,

$$P(\max_{z\in V_N}(\mathcal{S}_z^N - \bar{\mathcal{S}}_z^N) \ge C_4 K + \beta) \le 2e^{-C_5\beta^2}\,. \tag{6.4.10}$$

We also note the following bound, which is obtained similarly: there exist constants C_5, C_6 such that, for all $\beta > 0$,

$$P(\max_{z\in V'_{N/16}}(\bar{\mathcal{S}}_z^N - \mathcal{S}_z^{N/16}) \ge C_6 + \beta) \le 2e^{-C_7\beta^2}\,. \tag{6.4.11}$$

The advantage of working with $\bar{\mathcal{S}}^N$ instead of $\mathcal{S}^N$ is that the fields $\{\bar{\mathcal{S}}_z^N\}_{z\in W_{N,i}}$ are independent for $i = 1, \ldots, 4$. For every $\alpha, \beta > 0$, we have the bound

$$\begin{aligned} &P(\tilde{\mathcal{S}}_N^* \ge A_n - \alpha) \qquad\qquad (6.4.12)\\ &\ge P(\max_{z\in V'_N} \bar{\mathcal{S}}_z^N \ge A_n + C_4 - \alpha + \beta) - P(\max_{z\in V'_N}(\mathcal{S}_z^N - \bar{\mathcal{S}}_z^N) \ge C_4 + \beta)\\ &\ge P(\max_{z\in V'_N} \bar{\mathcal{S}}_z^N \ge A_n + C_4 - \alpha + \beta) - 2e^{-C_5\beta^2}\,, \end{aligned}$$

where (6.4.10) was used in the last inequality. On the other hand, for any $\gamma, \gamma' > 0$,

$$\begin{aligned} &P(\max_{z\in V'_N} \bar{\mathcal{S}}_z^N \ge A_n - \gamma) \ge P(\max_{i=1}^4 \max_{z\in W_{N,i}} \bar{\mathcal{S}}_z^N \ge A_n - \gamma)\\ &= 1 - (P(\max_{z\in W_{N,1}} \bar{\mathcal{S}}_z^N < A_n - \gamma))^4\\ &\ge 1 - \left(P(\max_{z\in V'_{N/16}} \mathcal{S}_z^{N/16} < A_n - \gamma + C_6 + \gamma') + 2e^{-C_7(\gamma')^2}\right)^4\,, \end{aligned}$$

where (6.4.11) was used in the inequality. Combining this estimate with (6.4.12), we get that, for any $\alpha, \beta, \gamma' > 0$,

$$
\begin{aligned}
&P(\tilde{\mathcal{S}}_N^* \geq A_n - \alpha) \\
&\geq 1 - 2e^{-C_5\beta^2} \\
&- \left(P(\max_{z \in V'_{N/16}} \mathcal{S}_z^{N/16} < A_n + C_4 + C_6 + \beta + \gamma' - \alpha) + 2e^{-C_7(\gamma')^2} \right)^4 .
\end{aligned} \tag{6.4.13}
$$

We now iterate the last estimate. Let $\eta_0 = 1 - \delta_0 < 1$ and, for $j \geq 1$, choose a constant $C_8 = C_8(\delta_0) > 0$ so that, for $\beta_j = \gamma'_j = C_8\sqrt{\log(1/\eta_j)}$,

$$\eta_{j+1} = 2e^{-C_5\beta_j^2} + (\eta_j + 2e^{-C_7(\gamma'_j)^2})^4$$

satisfies $\eta_{j+1} < \eta_j(1-\delta_0)$. (It is not hard to verify that such a choice is possible.) With this choice of β_j and γ'_j, set $\alpha_0 = 0$ and $\alpha_{j+1} = \alpha_j + C_4 + C_6 + \beta_j + \gamma'_j$, noting that $\alpha_j \leq C_9\sqrt{\log(1/\eta_j)}$ for some $C_9 = C_9(\delta_0)$. Substituting in (6.4.13) and using Proposition 8 to start the recursion, we get that

$$P(\tilde{\mathcal{S}}_N^* \geq A_n - \alpha_{j+1}) \geq 1 - \eta_{j+1} . \tag{6.4.14}$$

Therefore,

$$
\begin{aligned}
E\tilde{\mathcal{S}}_N^* &\geq A_n - \int_0^\infty P(\tilde{\mathcal{S}}_N^* \leq A_n - \theta)d\theta \\
&\geq A_n - \sum_{j=0}^\infty \alpha_j P(\tilde{\mathcal{S}}_N^* \leq A_n - \alpha_j) \\
&\geq A_n - C_9\sum_{j=0}^\infty \eta_j\sqrt{\log(1/\eta_j)} .
\end{aligned}
$$

Since $\eta_j \leq (1-\delta_0)^j$, it follows that there exists a constant $C_{10} > 0$ so that

$$E\mathcal{S}_N^* \geq E\tilde{\mathcal{S}}_N^* \geq A_n - C_{10} . \tag{6.4.15}$$

This completes the proof of Proposition 8 in the case $k_0 = 0$.

To consider the case $k_0 > 0$, define

$$\hat{\mathcal{S}}_{N,k_0}^* = \max_{z \in V'_N \cap 2^{k_0}\mathbb{Z}^2} \mathcal{S}_z^{N,k_0} .$$

Then, $\hat{\mathcal{S}}_{N,k_0}^* \leq \tilde{\mathcal{S}}_{N,k_0}^*$. On the other hand, $\hat{\mathcal{S}}_{N,k_0}^*$ has, by construction, the same distribution as $\tilde{\mathcal{S}}_{2^{-k_0}N,0}^* = \tilde{\mathcal{S}}_{2^{-k_0}N}^*$. Therefore, for any $y \in \mathbb{R}$,

$$P(\tilde{\mathcal{S}}_{N,k_0}^* \geq y) \geq P(\hat{\mathcal{S}}_{N,k_0}^* \geq y) \geq P(\tilde{\mathcal{S}}_{2^{-k_0}N}^* \geq y) .$$

We conclude that

$$E\mathcal{S}^*_{N,k_0} \geq E\tilde{\mathcal{S}}^*_{N,k_0} \geq E\tilde{\mathcal{S}}^*_{2^{-k_0}N}\,.$$

Application of (6.4.15) completes the proof of Proposition 8. □
Remark: J. Ding [Di11a] has improved on the proof of tightness by providing the following tail estimates.

Proposition 10. *The variance of $\mathcal{X}^*_N$ is uniformly bounded. Further, there exist universal constants c, C so that for any $x \in (0, (\log n)^{2/3})$, and with $\bar{\mathcal{X}}^*_N = \mathcal{X}^*_N - E\mathcal{X}^*_N$,*

$$ce^{-Cx} \leq P(\bar{\mathcal{X}}^*_N \geq x) \leq Ce^{-cx}\,, \quad ce^{-Ce^{Cx}} \leq P(\bar{\mathcal{X}}^*_N \leq -x) \leq Ce^{-ce^{cx}}\,,$$

It is interesting to compare these bounds with the case of BRW: while the bounds on the upper tail are similar, the lower tail exhibits quite different behavior, since in the case of BRW, just modifying a few variables near the root of the tree can have a significant effect on the maximum. On the other hand, the tail estimates in Proposition 10 are not precise enough for convergence proofs as we did in the BRW case. We will see that more can be said.

6.5 Tail estimates for GFF

`To complete`

6.6 Convergence of Maximum for GFF & extremal process

`To complete`

7 Isomorphism theorems and cover times

`To complete`

References

[ABR08] L. Addario-Berry and B. Reed, Ballot theorems for random walks with finite variance. arXiv:0802.2491v2. (Updated version: http://www.math.mcgill.ca/louigi/papers/ballot.pdf)

[ABR09] L. Addario-Berry and B. Reed, Minima in branching random walks, *Annals Probab.* **37** (2009), pp. 1044–1079.

[Ad90] R. Adler, *An Introduction to Continuity, Extrema and Related Topics for General Gaussian Processes*, Lecture Notes - Monograph Series, Institute Mathematical Statistics, Hayward, CA (1990). Errata, http://webee.technion.ac.il/people/adler/borell.pdf.

[Ai11] E. Aïdékon, Convergence in law of the minimum of a branching random walk. arXiv:1101.1810v3. To appear, *Annals Probab.*

[ABBS11] E. Aïdékon, M. Brunnet, J. Berestycki and Z. Shi, The branching Brownian motion seen from its tip, arXiv:1104.3738v2 (2011).

[ABK11] L.-P. Arguin, A. Bovier and N. Kistler, The extremal process of branching Brownian motion, arXiv:1103.2322v1 (2011).

[BL13] M. Biskup and O. Louidor, Extreme local extrema of two-dimensional discrete Gaussian free field, arXiv:1306.2602 (2013).

[BDG01] E. Bolthausen, J.-D. Deuschel and G. Giacomin, Entropic repulsion and the maximum of the two-dimensional harmonic crystal, *Annals Probab.* **29** (2001), pp. 1670–1692.

[BDZ11] E. Bolthausen, J.-D. Deuschel and O. Zeitouni, Recursions and tightness for the maximum of the discrete, two dimensional Gaussian free field, *Elec. Comm. Prob.* **16** (2011), pp. 114–119.

[Bo75] C. Borell, The Brunn–Minkowski inequality in Gauss space, *Invent. Math.* **30** (1975), pp. 207–216.

[Br78] M. Bramson, Maximal displacement of branching Brownian motion, *Comm. Pure Appl. Math.* **31** (1978), pp. 531–581.

[BZ09] M. Bramson and O. Zeitouni, Tightness for a family of recursion equations, *Annals Probab.* **37** (2009), pp. 615–653.

[BZ11] M. Bramson and O. Zeitouni, Tightness of the recentered maximum of the two–dimensional discrete Gaussian Free Field, *Comm. Pure App. Math* **65** (2011), pp. 1–20.

[DH91] F. M. Dekking and B. Host, Limit distributions for minimal displacement of branching random walks, *Probab. Theory Relat. Fields* **90** (1991), pp. 403–426.

[DZ98] A. Dembo and O. Zeitouni, *Large deviation techniques and applications*, 2nd edition, Springer, New-York (1998).

[Di11a] J. Ding, Exponential and double exponential tails for maximum of two-dimensional discrete Gaussian free field, arXiv:1105.5833v2 (2011).

[Fe75] X. Fernique, Regularité des trajectoires des fonctions aléatoires gaussiennes, Ecole d'Eté de Probabilités de Saint-Flour, IV-1974, pp. 1–96. *Lecture Notes in Math.* **480**, Springer, Berlin, (1975).

[LS87] S. P. Lalley and T. Sellke, A conditional limit theorem for the frontier of a branching Brownian motion, *Annals Probab.* **15** (1987), pp. 1052–1061.

[Li78] T. Liggett, Random invariant measures for Markov chains and independent particle systems, *Z. Wahr. verw. geb.* **45** (1978), pp. 297–313.

[Ma11] T. Madaule, Convergence in law for the branching random walk seen from its tip, arXiv:1107.2543 (2011).

[Ro11] M. Roberts, A simple path to asymptotics for the frontier of a branching Brownian motion, to appear, *Annals Probab.* (2012). arXiv:1106.4771v4.

[TIS76] B. S. Tsirelson, I. A. Ibragimov and V. N. Sudakov, Norms of Gaussian sample functions, *Lecture Notes Math.* **550**, Springer (1976).