Talagrand's inequality for interacting particle systems satisfying a log-Sobolev inequality

Florian Völlering

Fakultät II, Institut für Mathematik, TU Berlin, MA 7-5, Strasse des 17. Juni 136, 10623 Berlin, Germany. E-mail: voeller@math.tu-berlin.de

Abstract. Talagrand's inequality for independent Bernoulli random variables is extended to many interacting particle systems (IPS). The main assumption is that the IPS satisfies a log-Sobolev inequality. In this context it is also shown that a slightly stronger version of Talagrand's inequality is equivalent to a log-Sobolev inequality.

Additionally we also look at a common application, the relation between the probability of increasing events and the influences on that event by changing a single spin.

Résumé. Nous étendons l'inégalité de Talagrand pour des variables aléatoires de Bernoulli à une grande classe de systèmes de particules. L'hypothèse principale est que les systèmes de particules satisfont l'inégalité de log-Sobolev. Dans ce contexte nous démontrons également qu'une version plus forte de l'inégalité de Talagrand est équivalente à l'inégalité de log-Sobolev.

Nous considérons aussi comme application la relation entre la probablité d'un évènement croissant et les "influences" sur cet évènement du changement d'un seul spin.

MSC: Primary 60K35; secondary 60C05

Keywords: Talagrand's inequality; Russo's formula; Variance estimate; Dependent random variables; Log-Sobolev inequality; Interacting particle system

1. Introduction

In the famous paper [16], Talagrand proved that for Bernoulli measures μ_p, $0 < p < 1$, on the discrete hypercube $\{0, 1\}^N$ the following inequality holds:

$$\text{Var}_\mu(f) \leq C \sum_{i=1}^{N} \frac{\|\nabla_i f\|_{L^2(\mu_p)}^2}{1 + \log(\|\nabla_i f\|_{L^2(\mu_p)}/\|\nabla_i f\|_{L^1(\mu_p)})}. \tag{1}$$

Here f is an arbitrary function in $L^2(\mu_p)$, $\nabla_i f(\eta) = f(\eta^i) - f(\eta)$, η^i is obtained from η by changing the ith coordinate, and $C = C(p)$ is a constant dependent only on p.

This inequality has found many applications, for example in random graph theory [7] and percolation [1], and has as a consequence the widely used KKL-bound for influences of Boolean functions [10]. However the original proof is limited to product measures on the discrete hypercube. In [2] this restriction to independent coordinates was circumvented by encoding dependent random variables as functions of a larger set of independent random variables. It was shown that if the encoding is in some sense not too far from independent random variables, then an analogue to (1) holds. There are also versions of Talagrand's inequality and related inequalities like the KKL bound on continuous spaces [5,6,11]. In [6] Talagrand's inequality is proven under the assumption that there are Markovian dynamics

which satisfy the log-Sobolev inequality and a specific permutation relation between directional derivatives and the dynamics. While this approach is fruitful for products of Gaussian measures or the uniform measure on the sphere the required permutation relation is too restrictive for typical interacting particle systems.

In this paper we extend Talagrand's inequality to a wide class of interacting particle systems under the assumption of a log-Sobolev inequality. For example many finite range Gibbs measures are included. An important application of Talagrand's inequality is the study of influences on increasing events. By proving a version of Russo's formula for dependent random variables we can obtain similar estimates to the i.i.d. case.

We will prove under some conditions an equivalence between a version of Talagrand's inequality and a log-Sobolev inequality. Harder to prove, and from the point of view of applications more important, is the implication from a log-Sobolev inequality to Talagrand's inequality. Sufficient conditions for a log-Sobolev inequality are available in the context of interacting particle systems, and Talagrand's inequality as a consequence adds a useful variance estimate for the study of those.

2. Setting and results

2.1. Notation and definitions

Let E be a Polish space, $\mathcal{G}$ a finite or countable group and $\Omega := E^{\mathcal{G}}$. The most common choice is $E = \{0, 1\}$, $\mathcal{G} = \mathbb{Z}^d$. Elements of Ω we will denote by η, ξ, and elements of $\mathcal{G}$ by x, y, z. Given a probability measure μ on Ω and $p \geq 1$, we write

$$\|f\|_p = \|f\|_{L^p(\mu)} = \left(\int |f|^p \, \mathrm{d}\mu \right)^{1/p}.$$

By Φ we denote the convex function $\Phi(x) = \frac{x^2}{\log(e+|x|)}$. To Φ we consider the associated Orlicz norm,

$$\|f\|_{\Phi} := \inf \left\{ a > 0 \colon \int \Phi\left(\frac{f}{a} \right) \mathrm{d}\mu \leq 1 \right\}. \tag{2}$$

Let L be the generator of a Markov process on Ω, which we assume to be reversible with respect to an invariant and ergodic probability measure μ. The Dirichlet from associated to (L, μ) is given by $\mathcal{E}(f, g) = - \int f L g \, \mathrm{d}\mu$. The pair (L, μ) is said to satisfy a Poincaré inequality if there is a constant $\kappa > 0$ so that

$$\mathrm{Var}_{\mu}(f) = \int \left(f - \int f \, \mathrm{d}\mu \right)^2 \mathrm{d}\mu \leq \frac{1}{\kappa} \mathcal{E}(f, f), \quad \forall f \in L^2(\mu). \tag{3}$$

Similarly, (L, μ) satisfies a logarithmic Sobolev inequality if there is a constant $\rho > 0$ so that

$$\mathrm{Ent}_{\mu}(f^2) = \int f^2 \log\left(\frac{f^2}{\|f\|_2^2} \right) \mathrm{d}\mu \leq \frac{2}{\rho} \mathcal{E}(f, f), \quad \forall f \in L^2(\mu). \tag{4}$$

The best constants κ, ρ in (3) and (4) are called the spectral gap and the logarithmic Sobolev constant, which we will also call κ and ρ. It is well known that $\rho \leq \kappa$.

Let $(P_t)_{t \geq 0}$ be the semi-group generated by L. Both (3) and (4) imply contraction properties, (5) and (6), respectively:

$$\|P_t f\|_2 \leq \mathrm{e}^{-\kappa t} \|f\|_2; \tag{5}$$

$$\|P_t f\|_q \leq \|f\|_{p(t,q)}, \quad p(t, q) = 1 + (q-1)\mathrm{e}^{-2\rho t}. \tag{6}$$

The reader unfamiliar with these notions may be referred to [8]. In the context of interacting particle systems and Gibbs measures there are many interrelated notions of functional inequalities and mixing conditions on finite and infinite volume, which are also connected to decay of the semi-group. See [9,13,14] for a study of those.

We want to exploit some of the geometry available on Ω. To do so we assume that the generator is of the form

$$Lf(\eta) := \sum_{x \in \mathcal{G}} \int_E \big(f(\eta_{x \mapsto a}) - f(\eta)\big)\mu_{x,\eta}(\mathrm{d}a), \tag{7}$$

where $(\mu_{x,\eta})_{\eta \in \Omega, x \in \mathcal{G}}$ is a collection of probability measures on E, and $\eta_{x \mapsto a}$ is the configuration which is identical to η except at x, where it has value a. To avoid notation we will also write $\mu_{x,\eta}(\mathrm{d}\xi)$ with the understanding that $\xi = \eta_{x \mapsto a}$. We assume that the interaction is *finite range*: There exists a finite $\mathcal{N} \subset \mathcal{G}$ so that if $\eta'(y) = \eta(y)$ for all $y \in x + \mathcal{N}$, then $\mu_{x,\eta'} = \mu_{x,\eta}$. Glauber dynamics of interacting particle systems are a standard example for a generator of the form (7).

Based on the generator L and the measures $\mu_{x,\eta}$ we define for $x \in \mathcal{G}$ the linear operator Ψ_x on $L^p(\mu)$, $1 \leq p \leq \infty$, via

$$\Psi_x f(\eta) := \int f(\xi)\mu_{x,\eta}(\mathrm{d}\xi). \tag{8}$$

It is a consequence of the reversibility of L that

$$\int \Psi_x f \, \mathrm{d}\mu = \int f \, \mathrm{d}\mu. \tag{9}$$

From (9) follows an alternative way of writing the Dirichlet form:

$$\mathcal{E}(f, f) = \frac{1}{2} \sum_{x \in \mathcal{G}} \int \big[\Psi_x\big(f - f(\eta)\big)^2\big](\eta)\mu(\mathrm{d}\eta).$$

Define the derivative of a function f in direction x as

$$D_x f := \Psi_x f - f.$$

Note that this derivative is derived from the jump dynamics and does not necessarily agree with a possible natural derivative on E.

We will need that the derivatives of a function and the Dirichlet form are compatible in a certain sense. To make this concrete, a function $f \in L^2(\mu)$ is called *good with constant $K < \infty$*, if

$$\mathcal{E}(P_t f, P_t f) \leq K \sum_{x \in \mathcal{G}} \|D_x P_t f\|_2^2, \quad \forall t \geq 0. \tag{10}$$

To put (10) into perspective, observe that by Jensen's inequality we have $\sum_{x \in \mathcal{G}} \|D_x f\|_2^2 \leq 2\mathcal{E}(f, f)$. In this sense (10) is a form of a reverse Jensen's inequality, which may seem restrictive. However, under fairly mild conditions we have in fact that all functions in $L^2(\mu)$ are good.

Lemma 2.1. *Suppose E is finite and $\alpha := \inf\{\mu_{x,\eta}(e) \colon \eta \in \Omega, x \in \mathcal{G}, e \in E\} > 0$. Then any $f \in L^2(\mu)$ is good with constant α^{-3}.*

2.2. *Talagrand's inequality*

Theorem 2.2. *Assume that (L, μ) satisfies the log-Sobolev inequality (4) with constant $\rho > 0$, that L is of the form (7) and the interaction is finite range. Let $f \in L^2(\mu)$ be a good function with constant $K < \infty$. Then there exists a constant $C > 0$ so that*

$$\mathrm{Var}_\mu(f) \leq C \sum_{x \in \mathcal{G}} \|D_x f\|_\Phi^2. \tag{11}$$

The constant C is depends only on K, $|\mathcal{N}|$ and ρ.

Talagrand's inequality of the form (1) is a direct consequence from (11), as was already noted in [16].

Corollary 2.3. *Under the assumptions of Theorem* 2.2,

$$\mathrm{Var}_\mu(f) \leq c_1 C \sum_{x \in \mathcal{G}} \frac{\|D_x f\|_2^2}{1 + \log(\|D_x f\|_2 / \|D_x f\|_1)}, \tag{12}$$

where c_1 is a universal constant and terms on the right hand side where $D_x f$ is almost surely constant are treated as 0.

Proof. The claim follows directly by virtue of Lemma 2.5 in [16], which states that there is a universal constant C' so that

$$\|f\|_\Phi^2 \leq C' \frac{\|f\|_2^2}{1 + \log(\|f\|_2 / \|f\|_1)}.$$

$\square$

It is also possible to generalize Theorem 2.2 to product spaces.

Theorem 2.4. *Suppose we have, for $i = 1, \ldots, N$, the spaces $\Omega_i := E_i^{\mathcal{G}_i}$ with probability measures μ_i and dynamics generated by L_i satisfying the conditions of Theorem* 2.2. *Let μ be the product measure of the μ_i on $\Omega = \Omega_1 \times \cdots \times \Omega_N$. Let $L = L_1 + \cdots + L_N$ be the generator of the product semigroup P_t. Let $f \in L^2(\mu)$ be a good function with constant $K < \infty$ (where $\sum_{x \in \mathcal{G}}$ is replaced by $\sum_{i=1}^N \sum_{x \in \mathcal{G}_i}$ in the definition of a good function). Then there exists a constant $C > 0$ so that*

$$\mathrm{Var}_\mu(f) \leq C \sum_{i=1}^N \sum_{x \in \mathcal{G}_i} \|D_x f\|_\Phi^2.$$

The constant C is of the same form as in Theorem 2.2, *but with ρ replaced by the minimum over the ρ_i and $|\mathcal{N}|$ replaced by the maximum over the $|\mathcal{N}_i|$.*

If one compares (3), (4), (11) and (12) one observes that we have the following chain of implications:

$$(4) \Rightarrow (11) \Rightarrow (12) \Rightarrow (3).$$

It turns out that the first implication is in fact an equivalence. That is, not only does a log-Sobolev inequality imply Talagrand's inequality, but the converse is also true when using the version with the Orlicz norm.

Theorem 2.5. *Let L be of the form* (7). *Suppose that a function $f : \Omega \to \mathbb{R}$ satisfies*

$$\mathrm{Var}_\mu(f) \leq C \sum_{x \in \mathcal{G}} \|D_x f\|_\Phi^2 \tag{13}$$

for some $C \geq 1$. Then

$$\mathrm{Ent}_\mu(f) \leq c_2 C \mathcal{E}(f, f),$$

where c_2 is a universal constant.

Corollary 2.6. *Let L be of the form* (7) *with finite range. Assume that all $f \in L^2(\mu)$ are good with the same constant $K < \infty$. Then the following two statements are equivalent:*

(a) *(L, μ) satisfies a log-Sobolev inequality with constant $\rho > 0$;*

(b) *there is a constant $C > 0$ so that* (11) *is satisfied for all $f \in L^2(\mu)$.*

In [6] the proof of Talagrand's inequality consists of two key ingredients, namely hypercontractivity and a permutation relation between D_x and P_t. We will also use these two ingredients. However, the original permutation relation,

$$D_x P_t f \le e^{Kt} P_t D_x f, \quad 0 \le t \le t_0, \tag{14}$$

cannot hold in the context of interacting particle systems, as the following argument shows. Let $E = \{0, 1\}$, and $f(\eta) = \eta(x)$ for some $x \in \mathcal{G}$. Then $D_y f = 0$ for all $y \ne x$. But $D_y P_t f \ne 0$ for typical dynamics (with independent spin flips being an exception). Hence we need an alternative to (14) which respects the space–time structure generated by the Markov process. Most of the paper is devoted prove the following commutation property between the semigroup P_t and the derivative operator D_x:

Theorem 2.7. *Assume that the generator L is of the form* (7) *and the interaction is finite range. Then*

$$\sum_{x \in \mathcal{G}} \| D_x P_t f \|_2^2 \le \widetilde{C} 2^t \sum_{x \in \mathcal{G}} \| D_x f \|_{p(t)}^2, \tag{15}$$

with $\widetilde{C} = 2 e^{72 |\mathcal{N}|^2 (1 + |\mathcal{N}|)^2}$ and $p(t) = p(t, 2) = 1 + e^{-2\rho t}$. The constant ρ is again the log-Sobolev constant. However $\rho = 0$ is admitted as well, in which case we have the L^2-norm on both sides of (15).

The proof of Theorem 2.7 is at its core a recursive strategy. We will show that $\sum_{x \in \mathcal{G}} \| D_x P_t f \|_2^2$ can be estimated against $\sum_{x \in \mathcal{G}} \| D_x f \|_{p(t)}^2$ plus an error term of order 1. This error term can be estimated against $\sum_{x \in \mathcal{G}} \| D_x f \|_{p(t)}^2$ as well, but in doing so we introduce two error terms of order 2. This will then be iterated. The entire procedure is rather technical and is therefore done in the last sections of the paper.

2.3. *On influences and Russo's formula*

An important application of Talagrand's inequality is the study of the sensitivity of events to changes at a single site. We can obtain the same type of estimates as in [16] even in the context of dependent random variables. To this end we need to generalize is Russo's formula.

In this section we restrict ourselves to $E = \{0, 1\}$. We say a subset $A \subset \Omega = E^{\mathcal{G}}$ is increasing, if $\eta \in A$ and $\xi \ge \eta$ (coordinate wise) implies $\xi \in A$. Denote by A_x the event $\{\eta \in A, \eta^x \notin A\}$, where η^x is the configuration η flipped at x, that is a 1 at x is replaced by a 0 and vice versa.

Let μ_p, $p \in [a, b]$, be a family of measures on Ω. We assume that the corresponding dynamics given by the generator L_p from (7) are the heat bath Glauber dynamics. That is,

$$\mu_{x,\xi}^p(1) = \mu_p\big(\eta(x) = 1 | \eta = \xi \text{ off } x\big). \tag{16}$$

We also assume that the map $p \mapsto \mu_{x,\eta}^p(1)$ is increasing and differentiable for all $x \in \mathcal{G}$, $\eta \in \Omega$, and that all $\mu_{x,\eta}^p$ are finite range with the same neighborhood $\mathcal{N}$.

Theorem 2.8 (Russo's formula for dependent random variables). *Write $\beta_p := \inf_{\eta \in \Omega, x \in \mathcal{G}} \frac{d}{dp} \mu_{x,\eta}^p(1)$. Then, for any $p \in]a, b[$ and any increasing event A,*

$$\frac{d}{dp} \mu_p(A) \ge \beta_p \sum_{x \in \mathcal{G}} \frac{\mu_p(A_x)}{\sup_{\zeta \in \Omega} \mu_{x,\zeta}^p(1)} \ge \beta_p \sum_{x \in \mathcal{G}} \mu_p(A_x). \tag{17}$$

Remark. *In the case of Bernoulli product measures ν_p,* (17) *simplifies to $\frac{d}{dp} \nu_p(A) \ge \frac{1}{p} \sum_{x \in \mathcal{G}} \nu_p(A_x)$, which is in fact an equality and is the original form of Russo's formula.*

Corollary 2.9. *Fix an increasing event $A \subset \Omega$. Let ρ_p denote the log-Sobolev constant of (L_p, μ_p), which is assumed to be positive. Write $\delta_p := \sup\{\mu_{x,\eta}^p(A_x)\colon \eta \in \Omega, x \in \mathcal{G}\}$ and*

$$\alpha_p := \inf\{\mu_{x,\eta}^p(e)\colon \eta \in \Omega, x \in \mathcal{G}, e \in \{0,1\}\}.$$

If $\delta_p < e^2 \alpha_p^2$, then

$$\frac{\mathrm{d}}{\mathrm{d}p}\mu_p(A) \geq \frac{\beta_p \log(e^2\alpha_p^2\delta_p^{-1})}{4c_1 C}\mu_p(A)\big(1 - \mu_p(A)\big),$$

where c_1 and $C = C(\alpha_p^{-3}, |\mathcal{N}|, \rho_p)$ are as in Corollary 2.3 and β_p defined in Theorem 2.8.

Corollary 2.10. *With the notation of Corollary 2.9 in place, write $\delta := \sup_{p \in [a,b]} \delta_p$, $\alpha := \inf_{p \in [a,b]} \alpha_p$, $\beta := \inf_{p \in [a,b]} \beta_p$ and $\rho := \inf_{p \in [a,b]} \rho_p$. For $a \leq p_1 < p_2 \leq b$, we have*

$$\mu_{p_1}(A)\big(1 - \mu_{p_2}(A)\big) \leq \left(\frac{\delta}{e^2\alpha^2}\right)^{(p_2 - p_1)/C'},$$

where $C' = 4c_1 C(\alpha^{-3}, |\mathcal{N}|, \rho)/\beta$.

Finally we can also obtain an analogue to the KKL bound [10]. Here we slightly deviate from the setting of this section, in that we only have a single measure μ, and A can be any event in Ω. For such an event we define $\mathrm{supp}(A)$ as the set of $x \in \mathcal{G}$ for which $D_x \mathbb{1}_A$ is not constant almost surely.

Corollary 2.11. *Assume that $\alpha = \inf\{\mu_{x,\eta}(e)\colon \eta \in \Omega, x \in \mathcal{G}, e \in \{0,1\}\} > 0$ and (L, μ) satisfies a log-Sobolev inequality with constant $\rho > 0$. Write $R := |\mathrm{supp}(A)|/[\mu(A)(1 - \mu(A))]$, which we assume to be finite. Then*

$$\sup_{x \in \mathcal{G}} \mu(A_x) \geq \min\left(\frac{\log((\alpha^4/16)R)}{8CR}, \frac{e^2\alpha^2}{2}\right),$$

where $C = C(\alpha^{-3}, |\mathcal{N}|, \rho)$ is as in Theorem 2.2.

2.4. *Organization of the proofs*

The proofs are organized as follows: Section 3 proves Talagrand's inequality by using Theorem 2.7. In Section 4 we prove Russo's formula and related consequences plus Lemma 2.1. Then in Section 5 we prove Theorem 2.5. The remaining Sections 6 to 9 contain the proof of Theorem 2.7.

3. Proof of Talagrand's inequality

We will follow essentially the proof of Talagrand's inequality as in [6] with the use of the commutation property given by Theorem 2.7. Before proving Theorem 2.2 itself we need an auxiliary lemma.

Lemma 3.1. *Let $\Phi(x) = \frac{x^2}{\log(e + |x|)}$. There is a numerical constant $C > 0$ so that*

$$\int_1^2 \|f\|_r^2\, \mathrm{d}r \leq C\|f\|_\Phi^2.$$

Proof. By homogeneity we can assume $\|f\|_\Phi = 1$, which implies $\int \Phi(f)\, \mathrm{d}\mu \leq 1$. Write $f_n = |f|\mathbb{1}_{e^{n-1} < |f| \leq e^n}$, $n \geq 1$, and $g_0 = |f|\mathbb{1}_{|f| \leq 1}$. We have

$$\sum_{n=0}^{\infty} \frac{1}{n+1} \int f_n^2\, \mathrm{d}\mu \leq \sum_{n=0}^{\infty} \int \frac{f_n^2}{\log(e + f_n)}\, \mathrm{d}\mu \leq 1. \tag{18}$$

To prove the claim,

$$\int_1^2 \|f\|_r^2 \, dr = \int_1^2 \left(\sum_{n=0}^\infty \int f_n^r \, d\mu \right)^{2/r} dr$$

$$\leq \int_1^2 \left(\sum_{n=0}^\infty e^{(n-1)(r-2)} (n+1) \frac{\int f_n^2 \, d\mu}{n+1} \right)^{2/r} dr.$$

By convexity and (18), this is less than

$$\int_1^2 \sum_{n=0}^\infty e^{(2/r)(n-1)(r-2)} (n+1)^{2/r} \frac{\int f_n^2 \, d\mu}{n+1} \, dr.$$

Once we have shown that there is a constant C_1 so that

$$\int_1^2 e^{(2/r)(n-1)(r-2)} (n+1)^{2/r} \, dr \leq C_1 \tag{19}$$

uniformly in $n \in \mathbb{N}$ we can conclude that $\int_1^2 \|f\|_r^2 \, dr \leq C_1$. To show (19) we substitute $u = 2/r$ and then use the fact that $\int_1^2 e^{2x(1-u)} x^u \, du \leq 1$ for all $x \geq 1$:

$$\int_1^2 e^{(2/r)(n-1)(r-2)} (n+1)^{2/r} \, dr = \int_1^2 e^{2(n+1)(1-u)} (n+1)^u e^{-4(1-u)} \frac{2}{u^2} \, du \leq 2e^4. \qquad \square$$

Proof of Theorem 2.2. Without loss of generality we assume that $\int f \, d\mu = 0$. From the Poincaré inequality (5) it follows that

$$\mathrm{Var}_\mu(f) = \|f\|_2^2 \leq \|f\|_2^2 - \|P_T f\|_2^2 + e^{-2\kappa T} \|f\|_2^2$$

$$\leq \cdots \leq \frac{1}{1 - e^{-2\kappa T}} \left(\|f\|_2^2 - \|P_T f\|_2^2 \right)$$

$$= \frac{1}{1 - e^{-2\kappa T}} \int_0^T 2\mathcal{E}(P_t f, P_t f) \, dt. \tag{20}$$

Since f is a good function, we can estimate (20) using Theorem 2.7, to get

$$\mathrm{Var}_\mu(f) \leq \frac{4K e^{72|\mathcal{N}|^2(1+|\mathcal{N}|)^2}}{1 - e^{-2\kappa T}} \int_0^T 2^t \sum_{x \in \mathcal{G}} \|D_x f\|_{p(t)}^2 \, dt$$

$$\leq \frac{4K e^{72|\mathcal{N}|^2(1+|\mathcal{N}|)^2} 2^T}{1 - e^{-2\rho T}} \sum_{x \in \mathcal{G}} \int_0^T \|D_x f\|_{p(t)}^2 \, dt. \tag{21}$$

Choosing $T = \frac{1}{2\rho}$ and using the substitution $r = p(t)$, we have

$$\int_0^T \|D_x f\|_{p(t)}^2 \, dt \leq \frac{e}{2\rho} \int_1^2 \|D_x f\|_r^2 \, dr.$$

Using Lemma 3.1 we can conclude that

$$(21) \leq C(|\mathcal{N}|, \rho, K) \sum_{x \in \mathcal{G}} \|D_x f\|_\Phi^2. \qquad \square$$

Proof of Theorem 2.4. It is known that the log-Sobolev inequality tensorizes (see for example [8, Theorem 4.4]), that is, it generalizes to product spaces. The log-Sobolev constant on the product space is given by the minimum of the individual log-Sobolev constants. However, Ω is not of the form $E^{\mathcal{G}}$, so we cannot apply Theorem 2.2 directly. But the proofs are identical except for minor modifications to adjust to this more general setting. It suffices to replace the definition of I in Section 8 by

$$I(T, v, S) := \sum_{i=1}^{N} \sup_{X \in \mathcal{X}_T(\mathcal{G}_i)} \sum_{x \in \mathcal{G}_i} \||D_x \Psi_{S, x+X}(v, \infty) f\||^2_{p(\underline{S}(v))},$$

with $\mathcal{X}_T(\mathcal{G})$ being the maps from $\mathring{T}$ to $\mathcal{G}$, and also to replace $|N|$ and ρ by their respective maximum or minimum.   □

4. Russo's formula and other proofs

Proof of Lemma 2.1. Fix $x \in \mathcal{G}$ and $f \in L^2(\mu)$. Let η^* denote the maximizer (or one of the maximizers) of f out of the set $\{\xi \in \Omega : \xi = \eta \text{ off } x\}$. Similarly let η_* denote the minimizer. We have

$$\int (\Psi_x f - f)^2 \, d\mu \geq \int \left(\int (f(\xi) - f(\eta)) \mu_{x,\eta}(d\xi) \right)^2 \mathbb{1}_{\eta=\eta_*} \mu(d\eta)$$

$$\geq \int \left(\int_{\xi=\eta^*} (f(\xi) - f(\eta_*)) \mu_{x,\eta}(d\xi) \right)^2 \mathbb{1}_{\eta=\eta_*} \mu(d\eta)$$

$$\geq \alpha^2 \int (f(\eta^*) - f(\eta_*))^2 \mathbb{1}_{\eta=\eta_*} \mu(d\eta).$$

By (9) and the fact that η^* and η_* do not depend on $\eta(x)$, the above is equal to

$$\alpha^2 \int (f(\eta^*) - f(\eta_*))^2 \Psi_x(\mathbb{1}_{\eta=\eta_*}) \mu(d\eta) = \alpha^2 \int (f(\eta^*) - f(\eta_*))^2 \mu_{x,\eta}(\eta_*(x)) \mu(d\eta)$$

$$\geq \alpha^3 \int \Psi_x (f - f(\eta))^2 (\eta) \mu(d\eta).$$ □

Proof of Theorem 2.8. Fix $\varepsilon > 0$ so that $p - \varepsilon > a$. Since μ_p dominates $\mu_{p-\varepsilon}$, we can construct a coupling $\widehat{\mathbb{P}}$ of μ_p and $\mu_{p-\varepsilon}$ so that η and ξ have law μ_p and $\mu_{p-\varepsilon}$ respectively, and $\eta \geq \xi$ $\widehat{\mathbb{P}}$-a.s.

Fix a finite subset Λ of $\mathcal{G}$, and set $\Delta := \{x \in \mathcal{G} : \eta(x) \neq \xi(x)\}$. Since A is increasing,

$$\mu_p(A) - \mu_{p-\varepsilon}(A) = \widehat{\mathbb{P}}(\eta \in A, \xi \notin A) = \sum_{x \in \Lambda} \widehat{\mathbb{E}} \left(\frac{1}{|\Delta \cap \Lambda|} \mathbb{1}_{x \in \Delta} \mathbb{1}_{\eta \in A, \xi \notin A} \right)$$

$$\geq \sum_{x \in \Lambda} \widehat{\mathbb{E}} \left(\frac{1}{|\Delta \cap \Lambda|} \mathbb{1}_{x \in \Delta} \mathbb{1}_{\eta \in A, \eta^x \notin A} \right).$$

For $x \in \Lambda$ we have $\mathbb{1}_{x \in \Delta} \frac{1}{|\Delta \cap \Lambda|} \geq \mathbb{1}_{\Delta \cap \Lambda = \{x\}}$, and since A is increasing, $A_x = \{\eta \in A, \eta^x \notin A\}$ implies $\eta(x) = 1$. Hence,

$$\widehat{\mathbb{E}} \left(\frac{1}{|\Delta \cap \Lambda|} \mathbb{1}_{x \in \Delta} \mathbb{1}_{A_x} \right) \geq \widehat{\mathbb{E}} \left(\mathbb{1}_{A_x} \widehat{\mathbb{P}} [\Delta \cap \Lambda = \{x\} | \eta, \eta(x) = 1] \right)$$

$$\geq \widehat{\mathbb{P}}(A_x) \inf_{\zeta \in \Omega} \widehat{\mathbb{P}} (\Delta \cap \Lambda = \{x\} | \eta = \zeta \text{ off } x, \eta(x) = 1).$$

Fix $\zeta \in \Omega$ for now. By using (16) and assuming that x is such that $x + \mathcal{N} \subset \Lambda$, we get

$$
\widehat{\mathbb{P}}\big(\Delta \cap \Lambda = \{x\}|\eta = \zeta \text{ off } x, \eta(x) = 1\big)
$$

$$
= \frac{1}{\mu_{x,\zeta}^p(1)}\widehat{\mathbb{P}}\big(\Delta \cap \Lambda = \{x\}, \eta(x) = 1|\eta = \zeta \text{ off } x\big)
$$

$$
= \frac{1}{\mu_{x,\zeta}^p(1)}\widehat{\mathbb{E}}\big(\mathbb{1}_{\Delta \cap \Lambda \setminus \{x\} = \varnothing}\widehat{\mathbb{P}}\big(\eta(x) = 1, \xi(x) = 0|\eta = \xi = \zeta \text{ off } x\big)|\eta = \zeta \text{ off } x\big)
$$

$$
\geq \frac{1}{\mu_{x,\zeta}^p(1)}\widehat{\mathbb{E}}(\mathbb{1}_{\Delta \cap \Lambda \setminus \{x\} = \varnothing}|\eta = \zeta \text{ off } x) \inf_{\zeta' \in \Omega} \widehat{\mathbb{P}}\big(\eta(x) = 1, \xi(x) = 0|\eta = \xi = \zeta' \text{ off } x\big).
$$

By using (16) as well as $\eta(x) \geq \xi(x)$, we have

$$
\inf_{\zeta' \in \Omega} \widehat{\mathbb{P}}\big(\eta(x) = 1, \xi(x) = 0|\eta = \xi = \zeta' \text{ off } x\big) = \inf_{\zeta' \in \Omega}\big[\mu_{x,\zeta'}^p(1) - \mu_{x,\zeta'}^{p-\varepsilon}(1)\big] \geq \beta_p \varepsilon - \mathrm{o}(\varepsilon). \tag{22}
$$

By using the fact that Λ is finite and $\mu^{p-\varepsilon}$ converges to μ^p we also have

$$
\widehat{\mathbb{E}}(\mathbb{1}_{\Delta \cap \Lambda \setminus \{x\} = \varnothing}|\eta = \zeta \text{ off } x) = 1 - \mathrm{o}(1) \tag{23}
$$

uniformly in ζ. Combining (22) and (23), we get

$$
\inf_{\zeta \in \Omega} \widehat{\mathbb{P}}\big(\Delta \cap \Lambda = \{x\}|\eta = \zeta \text{ off } x, \eta(x) = 1\big) \geq \frac{1}{\sup_{\zeta \in \Omega} \mu_{x,\zeta}^p(1)}\beta\big(\varepsilon - \mathrm{o}(\varepsilon)\big)\mathbb{1}_{x+\mathcal{N} \subset \Lambda}.
$$

Therewith is

$$
\mu_p(A) - \mu_{p-\varepsilon}(A) \geq \sum_{x \in \Lambda:\, x+\mathcal{N} \in \Lambda} \mu_p(A_x)\frac{\beta}{\sup_{\zeta \in \Omega} \mu_{x,\zeta}^p(1)}\big(\varepsilon - \mathrm{o}(\varepsilon)\big),
$$

and by dividing by ε and then sending first ε to 0 and then Λ to $\mathcal{G}$ we finish the proof. $\qquad \square$

To combine Russo's formula and Talagrand's inequality we require the following lemma.

Lemma 4.1. *For any event A, any μ out of the measures considered in Section 2.3, any $q \geq 1$ and $x \in \mathcal{G}$,*

$$
\Big(\inf_{\eta \in \Omega} \mu_{x,\eta}(0)\Big)^q \mu(A_x) \leq \|D_x \mathbb{1}_A\|_q^q \leq 2\mu(A_x).
$$

Proof. For any $q \geq 1$, we use Jensen's inequality and drop the q:

$$
\|D_x \mathbb{1}_A\|_q^q \leq \iint |\mathbb{1}_{\xi \in A} - \mathbb{1}_{\eta \in A}|\mu_{x,\eta}(\mathrm{d}\xi)\mu(\mathrm{d}\eta) \leq \iint (\mathbb{1}_{\xi \in A_x} + \mathbb{1}_{\eta \in A_x})\mu_{x,\eta}(\mathrm{d}\xi)\mu(\mathrm{d}\eta).
$$

Since ξ is μ-distributed the upper bound follows.

The lower bound is obtained by the following calculation:

$$
\|D_x \mathbb{1}_A\|_q^q \geq \int \left|\int (\mathbb{1}_{\xi \in A} - \mathbb{1}_{\eta \in A})\mu_{x,\eta}(\mathrm{d}\xi)\right|^q \mathbb{1}_{\eta \in A_x}\mu(\mathrm{d}\eta)
$$

$$
= \int \mathbb{1}_{\eta \in A_x}\mu_{x,\eta}(0)^q \mu(\mathrm{d}\eta) \geq \inf_{\eta \in \Omega} \mu_{x,\eta}(0)^q \mu(A_x). \qquad \square
$$

Proof of Corollary 2.9. Lemma 2.1 guarantees that all conditions of Theorem 2.2 are satisfied. Combining Theorem 2.8, Lemma 4.1 and Corollary 2.3 yields the claim. $\qquad \square$

Proof of Corollary 2.10. Using Corollary 2.9, the proof is the same as Corollary 1.3 in [16].   □

Proof of Corollary 2.11. We can assume that $\sup_{x\in\mathcal{G}} \mu(A_x) \leq \alpha^2 e^2/2$, otherwise there is nothing to prove.

For the other estimate, we follow the argument from [6]. Let x^* be the (or one of the) maximizer(s) of $\sup_{x\in\mathcal{G}} \mu(A_x)$. If $\mu(A_{x^*}) > R^{-1/2}$, then the result holds as $C \geq 1$. So we assume $\mu(A_{x^*}) \leq R^{-1/2}$. By using Corollary 2.3 and Lemma 4.1,

$$\mu(A)\big(1 - \mu(A)\big) \leq C\big|\mathrm{supp}(A)\big| \frac{\|D_{x^*}\mathbb{1}_A\|_2^2}{1 + \log(\|D_{x^*}\mathbb{1}_A\|_2/\|D_{x^*}\mathbb{1}_A\|_1)}$$

$$\leq C\big|\mathrm{supp}(A)\big| \frac{2\mu(A_{x^*})}{1 + \log((\alpha/2)\mu(A_{x^*})^{-1/2})}.$$

By using the upper bounds on $\mu(A_{x^*})$, the result follows.   □

5. Log-Sobolev inequality from Talagrand's inequality

To prove Theorem 2.5 we follow the proof of Proposition 1 in [4], where the analogue statement was shown for continuous spaces. Our discrete derivative operators D_x force slightly more complicated arguments, but the general structure of the proof is unchanged. In contrast to the previous sections here we rely on various Orlicz norms and their properties. For basic properties of Orlicz spaces the reader is referred for example to [12].

We say $\varphi \colon \mathbb{R} \to \mathbb{R} \cup \{\infty\}$ is a Young function if it is convex, even, $\varphi(0) = 0$, and $\lim_{x\to\infty}\varphi(x) = \infty$. The norm $\|\cdot\|_\varphi$ is defined just as in (2), that is

$$\|f\|_\varphi := \inf\left\{a > 0 \colon \int \varphi\left(\frac{f}{a}\right)\mathrm{d}\mu \leq 1\right\}.$$

Lemma 5.1. *Let φ be a Young function and assume $\int \varphi(f)\,\mathrm{d}\mu \leq K$. Then $\|f\|_\varphi \leq \max(1, K)$.*

Proof. If $K \leq 1$, then $\int \varphi(f)\,\mathrm{d}\mu \leq 1$. If $K > 1$, by convexity,

$$\int \varphi\left(\frac{f}{K}\right)\mathrm{d}\mu \leq \frac{1}{K}\int \varphi(f)\,\mathrm{d}\mu \leq 1.$$ □

Lemma 5.2. *Let $f \in L^2(\mu)$. Then*

$$\|f\|_\Phi \leq 2\|f\|_2.$$

Proof. By homogeneity assume $\|f\|_2 = 1$. As $\Phi(x) \leq x^2 + |x|$, we have $\int \Phi(f)\,\mathrm{d}\mu \leq 2$, and Lemma 5.1 completes the proof.   □

Lemma 5.3. *For measurable $f, g \colon \widetilde{\Omega} \to \mathbb{R}$ on some Polish space $\widetilde{\Omega}$, $\|fg\|_\Phi \leq 24\|f\|_{e^{x^2}-1}\|g\|_2$.*

For a proof, see [4], Lemma 7.

Lemma 5.4. *Let $(\mu \times \mu_x)$ denote the probability measure on $\Omega \times \Omega$ given by $(\mu \times \mu_x)(\mathrm{d}\eta, \mathrm{d}\xi) = \mu_{\eta,x}(\mathrm{d}\xi)\mu(\mathrm{d}\eta)$, and for $f \colon \Omega \to \mathbb{R}$ let $f_2 \colon \Omega \times \Omega \to \mathbb{R}$, $f_2(\eta, \xi) = f(\xi) - f(\eta)$. Then*

$$\|D_x f\|_{\Phi;\mu} \leq \|f_2\|_{\Phi;(\mu\times\mu_x)},$$

where the Orlicz norms are to be understood on the measure spaces (Ω, μ) and $(\Omega \times \Omega, (\mu \times \mu_x))$ respectively.

Proof. By homogeneity assume $\|f_2\|_{\Phi;(\mu \times \mu_x)} = 1$, hence $\int \Phi(f_2)\,\mathrm{d}(\mu \times \mu_x) \leq 1$. For $D_x f$ we have, using Jensen's inequality,

$$
\int \Phi(D_x f)\,\mathrm{d}\mu = \int \Phi\left(\int \big(f(\xi) - f(\eta)\big)\,\mathrm{d}\mu_{x,\eta}\right)\mu(\mathrm{d}\eta)
$$
$$
\leq \iint \Phi\big(f(\xi) - f(\eta)\big)\mu_{x,\eta}(\mathrm{d}\xi)\mu(\mathrm{d}\eta)
$$
$$
= \int \Phi(f_2)\,\mathrm{d}(\mu \times \mu_x) = 1.
$$

Therefore $\|D_x f\|_{\Phi;\mu} \leq 1$. $\qquad\qquad\square$

Proof of Theorem 2.5. Let $\varphi(x) := x^2 \log(1 + x^2)$. In [3], Proposition 4.1, it is shown that

$$
\sup_{c \in \mathbb{R}} \mathrm{Ent}_\mu\big((f + c)^2\big) \leq \frac{13}{4}\left\|f - \int f\,\mathrm{d}\mu\right\|_\varphi^2.
$$

To prove (4) it therefore suffices to show that $\|f - \int f\,\mathrm{d}\mu\|_\varphi^2 \leq C' \mathcal{E}(f, f)$.

Let f be bounded, $\int f\,\mathrm{d}\mu = 0$ and $\mathcal{E}(f, f) = \frac{1}{2C}$, where C the constant from (13). By (13), Lemma 5.2 and Jensen's inequality,

$$
\int f^2\,\mathrm{d}\mu \leq C \sum_{x \in \mathcal{G}} \|D_x f\|_\Phi^2
$$
$$
\leq 4C \sum_{x \in \mathcal{G}} \|D_x f\|_2^2
$$
$$
\leq 8C\mathcal{E}(f, f) = 4. \tag{24}
$$

Define $g : \Omega \to \mathbb{R}$ as $g(\eta) = f(\eta)\sqrt{\log(1 + f(\eta)^2)}$. By Cauchy–Schwarz's inequality and (24),

$$
\left(\int g\,\mathrm{d}\mu\right)^2 \leq \int f^2\,\mathrm{d}\mu \int \log\big(1 + f^2\big)\,\mathrm{d}\mu
$$
$$
\leq \int f^2\,\mathrm{d}\mu \log\left(1 + \int f^2\,\mathrm{d}\mu\right) \leq 4\log 5. \tag{25}
$$

Since $|(x\sqrt{\log(1 + x^2)})'| \leq 2\sqrt{\log(1 + x^2)}$, we can use the mean value theorem to obtain the estimate

$$
\big|g(\eta) - g(\xi)\big| \leq 2\sqrt{\log\big(1 + \max\big(f(\eta)^2, f(\xi)^2\big)\big)}\big|f(\eta) - f(\xi)\big|. \tag{26}
$$

Using Lemma 5.4, (26) and Lemma 5.3,

$$
\|D_x g\|_\Phi \leq \|g_2\|_{\Phi;(\mu \times \mu_x)}
$$
$$
\leq \big\|2\sqrt{\log\big(1 + \max\big(f(\eta)^2, f(\xi)^2\big)\big)}\big|f(\xi) - f(\eta)\big|\big\|_{\Phi;(\mu \times \mu_x)}
$$
$$
\leq 48\big\|\sqrt{\log\big(1 + \max\big(f(\eta)^2, f(\xi)^2\big)\big)}
$$
$$
\cdot\big\|_{\mathrm{e}^{x^2}-1;(\mu \times \mu_x)}\big\|f(\xi) - f(\eta)\big\|_{L^2(\mu \times \mu_x)}.
$$

We have $\|f(\xi) - f(\eta)\|^2_{L^2(\mu \times \mu_x)} = \int \Psi_x(f - f(\eta))^2(\eta)\mu(\mathrm{d}\eta)$. To estimate the Orlicz norm, using $\max(a^2, b^2) \leq 2a^2 + 2(b-a)^2$, (24) and $\mathcal{E}(f, f) \leq \frac{1}{2C} < \frac{1}{2}$,

$$\int \left(e^{x^2} - 1\right) \circ \sqrt{\log\left(1 + \max(f(\eta)^2, f(\xi)^2)\right)}(\mu \times \mu_x)(\mathrm{d}\eta \, \mathrm{d}\xi)$$

$$= \iint \max\left(f(\eta)^2, f(\xi)^2\right)\mu_{\eta,x}(\mathrm{d}\xi)\mu(\mathrm{d}\eta)$$

$$\leq 2 \int f(\eta)^2 \mu(\mathrm{d}\eta) + 2 \iint \left(f(\xi) - f(\eta)\right)^2 \mu_{\eta,x}(\mathrm{d}\xi)\mu(\mathrm{d}\eta)$$

$$\leq 8 + 2\mathcal{E}(f, f) \leq 10.$$

Hence, by Lemma 5.1, $\|\sqrt{\log(1 + \max(f(\eta)^2, f(\xi)^2))}\|_{e^{x^2}-1;(\mu \times \mu_x)} \leq 10$. In total we obtain the estimate

$$\sum_{x \in \mathcal{G}} \|D_x g\|^2_{\Phi} \leq 480^2 \sum_{x \in \mathcal{G}} \int \Psi_x\left(f - f(\eta)\right)^2(\eta)\mu(\mathrm{d}\eta) = 2 \cdot 480^2 \mathcal{E}(f, f). \tag{27}$$

To finish the proof, using (25), (13) and (27),

$$\int \varphi(f) \, \mathrm{d}\mu = \int g^2 \, \mathrm{d}\mu = \mathrm{Var}_\mu(g) + \left(\int g \, \mathrm{d}\mu\right)^2$$

$$\leq 4 \log 5 + C \sum_{x \in \mathcal{G}} \|D_x g\|^2_{\Phi}$$

$$\leq 4 \log 5 + 2 \cdot 480^2 < 700^2.$$

By Lemma 5.1, $\|f\|_\varphi \leq 700^2$. For general f we therefore have

$$\mathrm{Ent}_\mu\left(f^2\right) \leq \frac{13}{4}\left\|f - \int f \, \mathrm{d}\mu\right\|^2_\varphi < \frac{13}{4}700^4 C\mathcal{E}(f, f). \qquad \square$$

6. Graphical construction

In preparation of the proof of Theorem 2.7 we start with the basic graphical construction of the dynamics. It is classical that a wide range if interacting particle systems and related models admit a graphical construction. Based on one a Poisson point process on $\mathcal{G}$ and additional randomness one can construct the law of the process. This connection is widely used. We will also make use of this underlying structure, which is one of the reasons we required that the generator L is of the special form (7). Intuitively speaking, the law of the Markov process is given by a Poisson point process of intensity 1, where at each point a resampling event takes place, that is a configuration η is replaced by a configuration ξ, where ξ is drawn independently from $\mu_{x,\eta}$. We will need some more subtle facts of the graphical construction, and in this section we will introduce the necessary notation and rigorous definition.

As a motivation, note that from (8) follows that the generator L can be written as

$$Lf = \sum_{x \in \mathcal{G}}[\Psi_x f - f].$$

Less immediate is the fact that the semi-group P_t can also be written in terms of Ψ_x. This is our aim, but to do so requires a few steps.

For a finite set $A = \{(x_i, t_i): 1 \leq i \leq n, t_1 < t_2 < \cdots < t_n\} \subset \mathcal{G} \times [0, T]$ we can define Ψ_A as

$$\Psi_A := \Psi_{x_1}\Psi_{x_2} \cdots \Psi_{x_n}.$$

To extend this definition to infinite sets A requires some more work, as we no longer have a well-ordered sequence. Before going into the details we state the aim of this section, which is how to represent the semi-group via a Poisson point process and Ψ:

Lemma 6.1. *Let $\overline{N}_{[0,t]}$ be a Poisson point process with unit intensity on $\mathcal{G} \times [0, t]$. The semi-group (P_t) is given by*

$$P_t f = \int \Psi_{\overline{N}_{[0,t]}} f \, d\overline{N}_{[0,t]}.$$

Let A be a countable subset of $\mathcal{G} \times [0, T]$ with no two points at the same time. A partial order $<_A$ on $\mathcal{G} \times [0, T]$ is defined as follows: $(x, t) <_A (y, s)$ iff either $x = y$ and $t < s$ or there exists a finite subset $\{(x_1, t_1), \ldots, (x_K, t_K)\} \subset A$ such that $t < t_1 < t_2 < \cdots < t_K \leq s$ and $x_{m-1} \in x_m + \mathcal{N}$, $2 \leq m \leq K$, as well as $x \in x_1 + \mathcal{N}$ and $x_K = y$. We write $A_{<x} := \{(y, t) \in A: (y, t) <_A (x, T)\}$. We call A *locally finite*, if $|A_{<x}| < \infty$ for all $x \in \mathcal{G}$.

For a locally finite $A \subset \mathcal{G} \times [0, T]$ and a local function $f : \Omega \to \mathbb{R}$ we define

$$\Psi_A f := \Psi_{A_f} f, \quad A_f := \bigcup_{x \in \text{supp}(f)} A_{<x}.$$

To extend this definition to non-local functions, note that Ψ_A induces a probability measure $\nu_{A,\eta}$ on Ω by $\int f \, d\nu_{A,\eta} = \Psi_A f(\eta)$, f local. We then define Ψ_A for non-local functions via

$$\Psi_A f(\eta) := \int f \, d\nu_{A,\eta}.$$

Lemma 6.2. *Let A, B be locally finite sets so that $A \leq_{A \cup B} B$, that is for no $a \in A, b \in B$ the relation $b <_{A \cup B} a$ holds. Then*

$$\Psi_{A \cup B} = \Psi_A \Psi_B.$$

Proof. Let f be a local function. By the nearest-neighbour property of Ψ_x and the fact that B is locally finite the function $\Psi_B f$ is local as well, with support given by

$$\text{supp}(\Psi_B f) = \{x \in \mathcal{G}: (x, 0) <_B (y, T), y \in \text{supp}(f)\}.$$

By construction

$$\Psi_A \Psi_B f = \Psi_{A_{\Psi_B f}} \Psi_{B_f} f.$$

From $A \leq_{A \cup B} B$ it then follows that

$$A_{\Psi_B f} = \{(x, s) \in A: (x, s) <_{A \cup B} (y, T), y \in \text{supp}(f)\}.$$

Hence

$$\Psi_A \Psi_B f = \Psi_{A_{\Psi_B f} \cup B_f} f = \Psi_{(A \cup B)_f} f = \Psi_{A \cup B} f. \qquad \square$$

We now have the tools to express the semi-group P_t via the Ψ_x.

Proof of Lemma 6.1. Define

$$\widetilde{P}_t := \int \Psi_{\overline{N}_{[0,t]}} f \, d\overline{N}_{[0,t]}.$$

Lemma 6.2 applied to $\overline{N}_{[0,t]}$ and $\overline{N}_{[t,t+s]}$ shows that the operator family is a semi-group, and easy calculations show that it is in fact a Markov semi-group whose generator is given by

$$\widetilde{L}f = \sum_{x \in \mathcal{G}}(\Psi_x f - f).$$

Since $\widetilde{L} = L$ the semi-groups must be the same as well.   □

7. Binary trees

In this short section we will introduce binary trees and associated notation. The way binary trees are tied to the proof of Theorem 2.7 is delegated to Sections 8 and 9.

A (rooted) *tree* is a finite connected cycle-free graph with a distinguished root vertex, which we denote by 0. Let d be the graph distance. Given a vertex v in a tree T, an adjacent vertex $w \in T$ is said to be a *child* of v if $d(w, 0) = d(v, 0) + 1$. Conversely, the *parent* of a non-root vertex v is the unique adjacent vertex w with $d(w, 0) = d(v, 0) - 1$. We call the *descendants* of a vertex v the set of all vertices w whose unique path to the root passes through v.

A *binary tree* is a tree whose vertices have at most 2 children, with the additional information whether a child is the *left* child or the *right* child (i.e., a vertex can have no children, a left child, a right child, or a left and a right child). We call a binary tree *full* if for each vertex the number of children is either 0 or 2. A vertex with no children is called a *leaf*. With ∂T we denote the set of all leaves of a tree T, and with $\overset{\circ}{T} = T \setminus \partial T$ the set of interior vertices.

From now on we will only consider full binary trees. On such a tree T we can introduce a well-ordering (see Figure 1). For any $v \in T$, the left child and all of its descendants are smaller than v, and the right child and all of its descendants are bigger than v.

From this well-ordering we also obtain an embedding $\iota_{\mathbb{Z}}$ of T into $\mathbb{Z}$, where we assume that $\iota_{\mathbb{Z}}$ is order preserving, $\iota_{\mathbb{Z}}(0) = 0$, and $\iota_{\mathbb{Z}}(T) = \{a, a + 1, \ldots, b\}$ for some $a, b \in \mathbb{Z}$. Observe that in this ordering leaf vertices and interior vertices alternate, with leaves getting mapped onto odd numbers (with the exception of the tree where the root is already a leaf).

Based on the embedding $\iota_{\mathbb{Z}}$ we define what we mean by $v + 1$:

$$v + 1 := \begin{cases} \iota_{\mathbb{Z}}^{-1}(\iota_{\mathbb{Z}}(v) + 1), & \iota_{\mathbb{Z}}(v) < b; \\ \infty, & \iota_{\mathbb{Z}}(v) = b. \end{cases}$$

Note that we simply write ∞ for a virtual vertex to the right of all of T. Similarly we define $v - 1$, with $-\infty$ for a virtual vertex to the left of T.

Let $\mathcal{T}_n$ denote the set of full binary trees with exactly n leaves, and $\mathcal{T} = \bigcup_{n \geq 1} \mathcal{T}_n$ the set of all full binary trees. Let $l(v)$ denote the left child of a vertex $v \in \overset{\circ}{T}$, and $r(v)$ its right child.

8. T-partitions and the basic iteration steps

Now we will use binary trees to develop the core steps of the iteration scheme. Given a full binary tree T and a real number $t \geq 0$ we define a T-partition of $[0, t]$ as a map $S : T \to [0, \infty[$ with $S(0) = t$ and $S(l(v)) + S(r(v)) = S(v)$ for all $v \in \overset{\circ}{T}$. Such a map represents a hierarchical partition: The original interval $[0, t]$ corresponds to the root
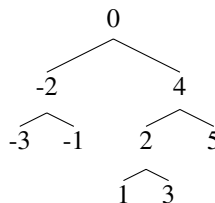

Fig. 1. A small tree with vertices identified by their embedding into $\mathbb{Z}$.

vertex, and the interval is then split into two pieces where their respective lengths are represented by the values $S(l(0))$ and $S(r(0))$. Those intervals are then split further, with each $v \in \mathring{T}$ corresponding to a further refinement of the partition. The leaves encode the final partition, and for a vertex $v \in T$ we write $\underline{S}(v) := \sum_{w \in \partial T, w < v} S(w)$ and $\overline{S}(v) := \sum_{w \in \partial T, w \leq v} S(w)$. In other words, the partition of the interval $[0, t]$ is given by $0 = \underline{S}(v_*) \leq \underline{S}(v_* + 2) \leq \underline{S}(v_* + 4) \leq \cdots \leq \underline{S}(v^*) \leq \underline{S}(\infty) = t$, where $v_* = \min\{v : v \in \partial T\}$, $v^* = \max\{v : v \in \partial T\}$.

The set of all possible T-partitions of $[0, t]$ we call $\mathcal{S}_{T,t}$. In addition to the T-partitions we also need maps $X : \mathring{T} \to \mathcal{G}$, which we collect in the set $\mathcal{X}_T$. The set $\mathcal{S}_{T,t}$ plays a big role in the structure of the proofs. The contribution of $\mathcal{X}_T$ is minor and is mostly used to simplify notation.

From now on we fix $t \geq 0$. Remember the operators Ψ_A from the graphical construction. Given a tree T as well as $v \in T$, $S \in \mathcal{S}_{T,t}$, $X \in \mathcal{X}_T$ and a realization of a Poisson point process $\overline{N}_{[0,t]} \subset \mathcal{G} \times [0, t]$ we define a linear operator $\Psi_{S,X}(v, \overline{N}_{[0,t]})$ on $L^p(\mu)$ by

$$\Psi_{S,X}(v, \overline{N}_{[0,t]}) f = \begin{cases} \Psi_{X(v)} f, & v \in \mathring{T}; \\ \Psi_{\overline{N}_{[\underline{S}(v), \overline{S}(v)[}} f, & v \in \partial T. \end{cases}$$

Based on $\Psi_{S,X}(v, \overline{N}_{[0,t]})$, we write

$$\Psi_{S,X}(v, w, \overline{N}_{[0,t]}) = \begin{cases} \Psi_{S,X}(v, \overline{N}_{[0,t]}) \Psi_{S,X}(v+1, \overline{N}_{[0,t]}) \cdots \Psi_{S,X}(w, \overline{N}_{[0,t]}), & v \leq w; \\ \mathrm{Id}, & v > w. \end{cases}$$

For notational convenience we set

$$\Psi_{S,X}(v, \infty, \overline{N}_{[0,t]}) = \Psi_{S,X}(v, \max\{w : w \in T\}, \overline{N}_{[0,t]}),$$

$$\Psi_{S,X}(-\infty, w, \overline{N}_{[0,t]}) = \Psi_{S,X}(\min\{v : v \in T\}, w, \overline{N}_{[0,t]})$$

and $\Psi_{S,X}(\infty, \infty, \overline{N}_{[0,t]}) = \mathrm{Id}$.

For a function $g(\overline{N}_{[0,t]}, \eta)$ of a realization $\overline{N}_{[0,t]}$ of a Poisson point process on $\mathcal{G} \times [0, t]$ and $\eta \in \Omega$, we define for $p \geq 1$ the norm

$$\||g\||_p := \left\| \int \left| g(\overline{N}_{[0,t]}, \cdot) \right| \mathrm{d}\overline{N}_{[0,t]} \right\|_p.$$

Now we define the key object to be used in the proof of Theorem 2.7. Given a tree T, as well as $v \in T$ and $S \in \mathcal{S}_{T,t}$, define

$$I(T, v, S) := \sup_{X \in \mathcal{X}_T} \sum_{x \in \mathcal{G}} \||D_x \Psi_{S,x+X}(v, \infty, \cdot) f\||^2_{p(\underline{S}(v))}.$$

For a tree $T \in \mathcal{T}_{n+1}$ the expression $I(T, v, S)$ represents an error term of order n. The vertex v is used to enumerate the different error terms which all correspond to the same tree T. Note that $I(T, \infty, S) = \sum_{x \in \mathcal{G}} \|D_x f\|^2_{p(t)}$, which is the goal for our estimation. The starting point, $\sum_{x \in \mathcal{G}} \|D_x P_t f\|^2_2$, can also be expressed by I:

$$\sum_{x \in \mathcal{G}} \|D_x P_t f\|^2_2 \leq I(\{0\}, 0, S) = I(\{0\}, -\infty, S). \tag{28}$$

This estimate follows from

$$|D_x P_t f| = \left| \int D_x \Psi_{\overline{N}_{[0,t]}} f \, \mathrm{d}\overline{N}_{[0,t]} \right| \leq \int |D_x \Psi_{\overline{N}_{[0,t]}} f| \, \mathrm{d}\overline{N}_{[0,t]},$$

which is obtained by Lemma 6.1 and Jensen's inequality.

The role of S is not yet evident at this point, but it will become more clear in Lemma 8.2 below.

In the following we will prove two lemmas which are the basic building blocks for the argument. Both lemmas estimate $I(T, v, S)$ by $I(T, v + 1, S)$, but the estimates are different for leaves and interior vertices. If v is an interior vertex, the statement is straightforward.

Lemma 8.1. *Let $T \in \mathcal{T}_n, n \geq 2$, and $v \in \mathring{T}$, $S \in \mathcal{S}_{T,t}$. Then*

$$I(T, v, S) \leq 9 I(T, v + 1, S).$$

The more complicated step is when v is a leaf. In that case we have to introduce an extra term with a tree of a higher order, which is the reason for introducing the trees in the first place.

Lemma 8.2. *Let $T \in \mathcal{T}_n$ and $v \in \partial T$, $S \in \mathcal{S}_{T,t}$. Then*

$$I(T, v, S) \leq 2 I(T, v + 1, S) + 2|\mathcal{N}|^2 S(v)\big(1 + S(v)|\mathcal{N}|\big)^2 \int_0^{S(v)} I\big(T_v', v' - 1, S_{r,v}'\big) \, dr,$$

where

- *T_v' is the tree in $\mathcal{T}_{n+1}$ obtained by appending two leaves to the vertex v;*
- *a vertex $w' \in T_v'$ represents the image of $w \in T$ when T is embedded canonically into T_v';*
- *$S_{r,v}' \in \mathcal{S}_{T_v',t}$ is given by $S_{r,v}'(v' - 1) = r$, $S_{r,v}'(v' + 1) = S(v) - r$ and $S_{r,v}'(w') = S(w)$ for $w \in T$.*

Corollary 8.3. *In the setting of Lemma 8.2, if $t \leq 1$, then*

$$I(T, v, S) \leq 2 I(T, v + 1, S) + 2|\mathcal{N}|^2\big(1 + |\mathcal{N}|\big)^2 S(v) \int_0^{S(v)} I\big(T_v', v' - 1, S_{r,v}'\big) \, dr.$$

The remainder of this section is dedicated to the proofs of Lemmas 8.1 and 8.2.

Proof of Lemma 8.1. As v is an interior vertex,

$$
\begin{aligned}
D_x \Psi_{S,X}(v, \infty, \cdot) f &= D_x \Psi_{X(v)} \Psi_{S,X}(v + 1, \infty, \cdot) f \\
&= \Psi_x \Psi_{X(v)} \Psi_{S,X}(v + 1, \infty, \cdot) f - \Psi_{X(v)} \Psi_{S,X}(v + 1, \infty, \cdot) f.
\end{aligned}
$$

By adding and subtracting the appropriate terms and using the triangle inequality,

$$\||\Psi_x \Psi_{X(v)} \Psi_{S,X}(v + 1, \infty, \cdot) f - \Psi_{X(v)} \Psi_{S,X}(v + 1, \infty, \cdot) f\||_p$$

$$\leq \||\Psi_x \Psi_{X(v)} \Psi_{S,X}(v + 1, \infty, \cdot) f - \Psi_x \Psi_{S,X}(v + 1, \infty, \cdot) f\||_p \tag{29}$$

$$+ \||\Psi_x \Psi_{S,X}(v + 1, \infty, \cdot) f - \Psi_{S,X}(v + 1, \infty, \cdot) f\||_p \tag{30}$$

$$+ \||\Psi_{X(v)} \Psi_{S,X}(v + 1, \infty, \cdot) f - \Psi_{S,X}(v + 1, \infty, \cdot) f\||_p. \tag{31}$$

Note that the terms in (30) and (31) are $D_x \Psi_{S,X}(v + 1, \infty, \cdot) f$ and $D_{X(v)} \Psi_{S,X}(v + 1, \infty, \cdot) f$ respectively. Furthermore, with $z = 0$ or $z = X(v)$,

$$
\begin{aligned}
&\sup_{X \in \mathcal{X}_T} \sum_{x \in \mathcal{G}} \||D_{x+z} \Psi_{S,x+X}(v + 1, \infty, \cdot) f\||_{p(\underline{S}(v))}^2 \\
&= \sup_{X \in \mathcal{X}_T} \sum_{x \in \mathcal{G}} \||D_x \Psi_{S,x-z+X}(v + 1, \infty, \cdot) f\||_{p(\underline{S}(v))}^2 \\
&\leq I(T, v + 1, S).
\end{aligned}
$$

The last inequality uses the fact that for interior vertices, $\underline{S}(v) = \underline{S}(v+1)$. Note that by (9) and Jensen's inequality, $\||\Psi_x g|\|_p \le \||g|\|_p$. Applying this fact to the function $D_{X(v)}\Psi_{S,X} f$ we can estimate (29) the same way as (30) and (31). To put everything together, we use above estimates and the fact that $(a+b+c)^2 \le 3(a^2+b^2+c^2)$:

$$\sup_{X \in \mathcal{X}_T} \sum_{x \in \mathcal{G}} \||D_x \Psi_{S,x+X}(v,\infty,\cdot)f|\|^2_{p(\underline{S}(v))}$$

$$\le 3 \sup_{X \in \mathcal{X}_T} \sum_{x \in \mathcal{G}} \||\Psi_x \Psi_{x+X(v)}\Psi_{S,x+X}(v+1,\infty,\cdot)f - \Psi_x \Psi_{S,x+X}(v+1,\infty,\cdot)f|\|^2_{p(\underline{S}(v))}$$

$$+ 3 \sup_{X \in \mathcal{X}_T} \sum_{x \in \mathcal{G}} \||\Psi_x \Psi_{S,x+X}(v+1,\infty,\cdot)f - \Psi_{S,x+X}(v+1,\infty,\cdot)f|\|^2_{p(\underline{S}(v))}$$

$$+ 3 \sup_{X \in \mathcal{X}_T} \sum_{x \in \mathcal{G}} \||\Psi_{x+X(v)}\Psi_{S,x+X}(v+1,\infty,\cdot)f - \Psi_{S,x+X}(v+1,\infty,\cdot)f|\|^2_{p(\underline{S}(v))}$$

$$\le 9 I(T, v+1, S). \qquad \square$$

Proof of Lemma 8.2. We start by studying

$$\int \left| D_x \Psi_{S,X}(v,\infty,\overline{N}_{[0,t]})f \right| d\overline{N}_{[0,t]}$$

$$= \int \left| D_x \Psi_{\overline{N}_{[\underline{S}(v),\overline{S}(v)[}} \Psi_{S,X}(v+1,\infty,\overline{N}_{[0,t]})f \right| d\overline{N}_{[0,t]}. \tag{32}$$

Denote by A the event that $\overline{N}_{[\underline{S}(v),\overline{S}(v)[} \cap ((x+\mathcal{N}) \times [\underline{S}(v),\overline{S}(v)[) = \varnothing$, i.e., the Poisson point process of the graphical construction has no points in a neighbourhood of x in the time interval $[\underline{S}(v),\overline{S}(v)[$. In that case D_x and $\Psi_{\overline{N}_{[\underline{S}(v),\overline{S}(v)[}}$ commute, and

$$(32) = \int \mathbb{1}_A \left| \Psi_{\overline{N}_{[\underline{S}(v),\overline{S}(v)[}} D_x \Psi_{S,X}(v+1,\infty,\overline{N}_{[0,t]})f \right| d\overline{N}_{[0,t]}$$

$$+ \int \mathbb{1}_{A^c} \left| D_x \Psi_{\overline{N}_{[\underline{S}(v),\overline{S}(v)[}} \Psi_{S,X}(v+1,\infty,\overline{N}_{[0,t]})f \right| d\overline{N}_{[0,t]}$$

$$\le \int \mathbb{1}_A \Psi_{\overline{N}_{[\underline{S}(v),\overline{S}(v)[}} \left| D_x \Psi_{S,X}(v+1,\infty,\overline{N}_{[0,t]})f \right| d\overline{N}_{[0,t]} \tag{33}$$

$$+ \int \mathbb{1}_{A^c} \left| D_x \Psi_{\overline{N}_{[\underline{S}(v),\overline{S}(v)[}} \Psi_{S,X}(v+1,\infty,\overline{N}_{[0,t]})f \right| d\overline{N}_{[0,t]}. \tag{34}$$

By construction $\Psi_{S,X}(v+1,\infty,\overline{N}_{[0,t]})$ does not depend on $\overline{N}_{[0,\overline{S}(v)[}$. Therefore we can split the integration into the time intervals $[0,\overline{S}(v)]$ and $[\overline{S}(v),t]$ and integrate separately:

$$(33) = \left(\int \mathbb{1}_A \Psi_{\overline{N}_{[\underline{S}(v),\overline{S}(v)[}} d\overline{N}_{[\underline{S}(v),\overline{S}(v)[} \right) \left[\int \left| D_x \Psi_{S,X}(v+1,\infty,\overline{N}_{[\underline{S}(v),t]})f \right| d\overline{N}_{[\underline{S}(v),t]} \right]$$

$$\le P_{S(v)} \left[\int \left| D_x \Psi_{S,X}(v+1,\infty,\overline{N}_{[0,t]})f \right| d\overline{N}_{[0),t]} \right].$$

Here we used the fact that the term in square brackets is a non-negative function, so that we can drop the indicator over A to increase the value and then use Lemma 6.1 to write the first term as the semi-group.

The estimate of (34) is more involved. Let M be Poisson($S(v)|\mathcal{N}|$)-distributed and $Q_1, Q_2, \ldots$ be independent and uniformly distributed on $(x+\mathcal{N}) \times [\underline{S}(v),\overline{S}(v)[$. Then $\overline{N}_{[\underline{S}(v),\overline{S}(v)[} \cap ((x+\mathcal{N}) \times [\underline{S}(v),\overline{S}(v)[)$ has the same distribution as $\bigcup_{k=1}^M \{Q_k\}$. The event A^c corresponds to the event $\{M \ge 1\}$. Therefore, at least the point Q_1 is present. The number of remaining points is no longer Poisson distributed, but given by a Poisson distribution conditioned on

at least one point, minus 1. However, this distribution has a density h with respect to the Poisson distribution which satisfies $\|h\|_\infty \leq 1 + S(v)|\mathcal{N}|$. Therefore, we can replace the distribution of the remaining points by a Poisson point process and obtain for any positive function g:

$$
\int \mathbb{1}_{A^c} g(\overline{N}_{[0,t]}) \, d\overline{N}_{[0,t]}
$$

$$
= \left(1 - e^{-|N|S(v)}\right) \int g(\overline{N}_{[0,t]}) \, d\left[\overline{N}_{[0,t]} \big| A^c\right]
$$

$$
\leq \frac{(1 - e^{-|N|S(v)})}{S(v)|\mathcal{N}|} \int_{\underline{S}(v)}^{\overline{S}(v)} \sum_{y \in \mathcal{N}} \int g\left(\overline{N}_{[0,t]} \cup \{(x+y, r)\}\right) h\left(\overline{N}_{[0,t]} \cup \{(x+y, r)\}\right) d\overline{N}_{[0,t]} \, dr
$$

$$
\leq 1 \cdot \left(1 + S(v)|\mathcal{N}|\right) \int_0^{S(v)} \sum_{y \in \mathcal{N}} \int g\left(\overline{N}_{[0,t]} \cup \{(x+y, \underline{S}(v) + r)\}\right) d\overline{N}_{[0,t]} \, dr. \tag{35}
$$

We now consider the function $g(\overline{N}_{[0,t]}) = |D_x \Psi_{\overline{N}_{[\underline{S}(v), \overline{S}(v)[}} \Psi_{S,X}(v+1, \infty, \overline{N}_{[0,t]}) f|$. Remember the definition of T'_v, $S'_{r,v}$ and define $X'_y \in \mathcal{X}_{T'_v}$ by $X'_y(v') = y$ and $X'_y(w') = X(w)$ for $w \in T$. Then we have for almost every $\overline{N}_{[0,t]}$

$$
g\left(\overline{N}_{[0,t]} \cup \{(x+y, \underline{S}(v) + r)\}\right) = \left| D_x \Psi_{\overline{N}_{[\underline{S}(v), \overline{S}(v)[} \cup \{(x+y, \underline{S}(v)+r)\}} \Psi_{S,X}(v+1, \infty, \overline{N}_{[0,t]}) f \right|
$$

$$
= \left| D_x \Psi_{\overline{N}_{[\underline{S}(v), \underline{S}(v)+r[}} \Psi_{x+y} \Psi_{\overline{N}_{[\underline{S}(v)+r, \overline{S}(v)[}} \Psi_{S,X}(v+1, \infty, \overline{N}_{[0,t]}) f \right|
$$

$$
= \left| D_x \Psi_{S'_{r,v}, X'_y}(v'-1, \infty, \overline{N}_{[0,t]}) f \right|.
$$

Using the above with (35) gives us

$$
(34) \leq \left(1 + S(v)|\mathcal{N}|\right) \int_0^{S(v)} \sum_{y \in \mathcal{N}} \int \left| D_x \Psi_{S'_{r,v}, X'_y}(v'-1, \infty, \overline{N}_{[0,t]}) f \right| d\overline{N}_{[0,t]} \, dr.
$$

Using all the previous estimates,

$$
I(T, v, S) \leq \sup_{X \in \mathcal{X}_T} \sum_{x \in \mathcal{G}} \bigg(\left\| \left| P_{S(v)} \big| D_x \Psi_{S,x+X}(v+1, \infty, \cdot) f \big| \right\| \right\|_{p(\underline{S}(v))}
$$

$$
+ \left(1 + S(v)|\mathcal{N}|\right) \int_0^{S(v)} \sum_{y \in \mathcal{N}} \left\| \left| D_x \Psi_{S'_{r,v}, x+X'_y}(v'-1, \infty, \cdot) f \right| \right\|_{p(\underline{S}(v))} dr \bigg)^2
$$

$$
\leq 2 \sup_{X \in \mathcal{X}_T} \sum_{x \in \mathcal{G}} \left\| \left| P_{S(v)} \big| D_x \Psi_{S,x+X}(v+1, \infty, \cdot) f \big| \right| \right\|_{p(\underline{S}(v))}^2 \tag{36}
$$

$$
+ 2|\mathcal{N}| S(v) \left(1 + S(v)|\mathcal{N}|\right)^2 \tag{37}
$$

$$
\times \int_0^{S(v)} \sum_{y \in \mathcal{N}} \sup_{X \in \mathcal{X}_T} \sum_{x \in \mathcal{G}} \left\| \left| D_x \Psi_{S'_{r,v}, x+X'_y}(v'-1, \infty, \cdot) f \right| \right\|_{p(\underline{S}(v))}^2 dr.
$$

Using hypercontractivity (6),

$$
(36) \leq 2 \sup_{X \in \mathcal{X}_T} \sum_{x \in \mathcal{G}} \left\| \left| D_x \Psi_{S,x+X}(v+1, \infty, \cdot) f \right| \right\|_{p(\overline{S}(v))}^2 = 2I(T, v+1, S).
$$

Since $\underline{S}(v) = S'_{r,v}(v'-1)$, and taking the supremum over $\mathcal{X}_{T'_v}$,

$$(37) \leq 2|\mathcal{N}|^2 S(v)\big(1 + S(v)|\mathcal{N}|\big)^2 \int_0^{S(v)} \sup_{X' \in \mathcal{X}_{T'_v}} \sum_{x \in \mathcal{G}} \||D_x \Psi_{S'_{r,v}, x+X'}\big(v'-1, \infty, \cdot\big)f\||^2_{P(\underline{S'_{r,v}}(v'-1))} \, \mathrm{d}r$$

$$= 2|\mathcal{N}|^2 S(v)\big(1 + S(v)|\mathcal{N}|\big)^2 \int_0^{S(v)} I\big(T'_v, v'-1, S'_{r,v}\big) \, \mathrm{d}r. \qquad \square$$

9. The measures $m_{T,t}$ and the iteration procedure

When looking at the error term introduced in Lemma 8.2 we see that there is some integration happening. When applying Lemma 8.2 (in the form of Corollary 8.3) to the integrand of the error term, we get another integral. All together, to describe the total influence of an error term we will have to integrate with respect to some measure which depends on the tree structure and the time interval $[0, t]$. We call this measure $m_{T,t}$.

Before defining the measure $m_{T,t}$ we have to introduce a bit of notation. Given a tree T, $T \neq \{0\}$, we denote by T_L the *left subtree* of T, which is the unique tree T_L for which a map $\iota_L : T_L \to T$ with the following properties exists:

- ι_L is injective and order preserving;
- $\iota_L(T_L) = \{v \in T : v < 0\}$;
- $\iota_L(0) = l(0)$.

Analogously we define T_R, the *right subtree* of T.

Fix a tree $T \neq \{0\}$ and $t \geq 0$. For $0 \leq s \leq t$, choose $S_L \in \mathcal{S}_{T_L, s}$ and $S_R \in \mathcal{S}_{T_R, t-s}$. From the partitions S_L and S_R on the subtrees we define a corresponding T-partition of $[0, t]$,

$$(S_L, t, S_R)(v) := \begin{cases} S_L(\iota_L^{-1}(v)), & v < 0; \\ t, & v = 0; \\ S_R(\iota_R^{-1}(v)), & v > 0. \end{cases}$$

For $T \in \mathcal{T}_n$ and $t \geq 0$ we define the measure $m_{T,t}$ on $\mathcal{S}_{T,t}$ recursively. If $T = \{0\}$, $m_{T,t} := \delta_{0 \mapsto t}$, the Dirac measure on the single element of $\mathcal{S}_{\{0\},t}$. If $T \neq \{0\}$,

$$m_{T,t} := t \int_0^t \int_{\mathcal{S}_{T_L,s}} \int_{\mathcal{S}_{T_R,t-s}} \delta_{(S_L,t,S_R)} m_{T_R,t-s}(\mathrm{d}S_R) m_{T_L,s}(\mathrm{d}S_L) \, \mathrm{d}s.$$

The structure of $m_{T,t}$ is chosen to mirror the structure of the second right hand term in Lemma 8.2 (for t small). The recursive nature makes both terms somewhat difficult to understand. However, we obtain a reasonable estimate on the mass of $m_{T,t}$.

Lemma 9.1. *For $T \in \mathcal{T}_n$, $t \geq 0$,*

$$\int_{\mathcal{S}_{T,t}} 1 m_{T,t}(\mathrm{d}S) \leq \frac{t^{2n-2}}{(2n-3)!!},$$

where $(2k-1)!! = \prod_{i=1}^k (2i-1)$ is the double factorial, with the convention of $(-1)!! = 1$.

Proof. The proof will be by induction over n. For $n = 1$ both sides of the statement equal 1. For the induction step we use the recursive definition of $m_{T,t}$. Assume that the statement holds for all trees in $\mathcal{T}_k$, $1 \leq k \leq n$ and all $t \geq 0$. For any $T \in \mathcal{T}_{n+1}$ and any $t \geq 0$,

$$\int_{\mathcal{S}_{T,t}} 1 m_{T,t}(\mathrm{d}S) = t \int_0^t \int_{\mathcal{S}_{T_L,s}} \int_{\mathcal{S}_{T_R,t-s}} 1 m_{T_R,t-s}(\mathrm{d}S_R) m_{T_L,s}(\mathrm{d}S_L) \, \mathrm{d}s$$

$$\leq t \int_0^t \frac{s^{2n_L-2}}{(2n_L-3)!!} \frac{(t-s)^{2n_R-2}}{(2n_R-3)!!} \, \mathrm{d}s, \tag{38}$$

where $T_L \in \mathcal{T}_{n_L}$ and $T_R \in \mathcal{T}_{n_R}$. Using partial integration,

$$(38) = \frac{(2n_L - 2)!(2n_R - 2)!}{(2n_L - 3)!!(2n_R - 3)!!(2n_L + 2n_R - 3)!} t^{2n_R + 2n_L - 2}.$$

A simple calculation shows that $\frac{(2n_L-2)!(2n_R-2)!}{(2n_L-3)!!(2n_R-3)!!(2n_L+2n_R-3)!} \leq \frac{1}{(2n_L+2n_r-3)!!}$, and since $n_L + n_R = n + 1$ this completes the proof. $\qquad\square$

To facilitate the recursive argument which will ultimately lead to a proof of Theorem 2.7 we need to look at the vertices of a tree in more detail. First, we denote by

$$\#v_- := \big|\{w \in \partial T \colon w < v\}\big|, \quad T \in \mathcal{T}, v \in T,$$

the number of leaves before a vertex v. Next we define

$$B(T) = \max\{v \in \mathring{T} \colon v - 1, v + 1 \in \partial T\}, \quad T \in \mathcal{T},$$

with the convention that the maximum of the empty set is $-\infty$. In other words, $B(T)$ is the last simple branching point in T, with simple referring to the fact that both its children are leaves.

Lemma 9.2. *For a tree T and T'_v as in Lemma 8.2, we have the following*:

(a) *Let $v \in \partial T$ be a leaf which satisfies $v \geq B(T) - 1$. Then in the expanded tree T'_v, $v' = B(T'_v)$. In other words, the embedding v' of v into T'_v is the last simple branching point in T'_v. Additionally, the number of leaves before v satisfies $\#v_- = \#B(T'_v)_- - 1$.*
(b) *The set $\mathcal{T}_{n+1}$ of full binary trees with exactly $n + 1$ leaves can be obtained as the disjoint union of expansions of the trees in $\mathcal{T}_n$: $\mathcal{T}_{n+1} = \mathring{\bigcup}_{T \in \mathcal{T}_n}\{T'_v \colon v \in \partial T, v \geq B(T) - 1\}$.*

Proof. If v is not a child of $B(T)$, then $v > B(T)$ and $v' = B(T'_v)$. If v is a child of $B(T)$, then $B(T)'$ is no longer a simple branching point, and hence $v' = B(T'_v)$. From this and $\#v'_- = \#v_- + 1$ follows $\#v_- = \#B(T'_v)_- - 1$.

For a tree $T \in \mathcal{T}_{n+1}$, let $T^* \in \mathcal{T}_n$ be the tree obtained by removing the two child leaves of $B(T)$. By slight abuse of notation we also write $B(T)$ for the leaf in T^* so that $(T^*)'_{B(T)} = T$. By construction of T^*, $B(T^*) < B(T)$ or $B(T)$ is a child of $B(T^*)$. Hence $\mathcal{T}_{n+1} \subset \bigcup_{T \in \mathcal{T}_n}\{T'_v \colon v \in \partial T, v \geq B(T) - 1\}$. The other inclusion is obvious. To show that the union is disjoined, we observe that $(T'_v)^* = T$ for any $v \in \partial T, v \geq B(T) - 1$. $\qquad\square$

We are now in state to use Lemmas 8.1 and 8.2 iteratively, providing us with a first commutation property between D_x and P_t. This iteration procedure is captured in the following theorem.

Theorem 9.3. *Suppose $t \leq 1$. For any $N \geq 0$,*

$$\sum_{x \in \mathcal{G}} \|D_x P_t f\|_2^2 \leq \left(\sum_{n=1}^{N} 2 \cdot 18^{n-1}\big(2 \cdot |\mathcal{N}|^2(1 + |\mathcal{N}|)^2\big)^{n-1} \frac{|\mathcal{T}_n|}{(2n-3)!!} t^{2n-2}\right) \sum_{x \in \mathcal{G}} \|D_x f\|_{p(t)}^2$$

$$+ \big(2 \cdot |\mathcal{N}|^2(1 + |\mathcal{N}|)^2\big)^N \sum_{T \in \mathcal{T}_{N+1}} 18^{\#(B(T)-1)_-} \int_{\mathcal{S}_{T,t}} I\big(T, B(T) - 1, S\big) m_{T,t}(\mathrm{d}S).$$

Proof. The proof will be by induction. For $N = 0$, we observe that

$$\sum_{T \in \mathcal{T}_1} 18^{\#(B(T)-1)_-} \int_{\mathcal{S}_{T,t}} I\big(T, B(T) - 1, S\big) m_{T,t}(\mathrm{d}S)$$

$$= I\big(\{0\}, -\infty, 0 \mapsto t\big) \geq \sum_{x \in \mathcal{G}} \|D_x P_t f\|_2^2,$$

where we used (28) for the last step. To show the induction step, it suffices to prove that

$$\sum_{T \in \mathcal{T}_N} 18^{\#(B(T)-1)_-} \int_{\mathcal{S}_{T,t}} I\big(T, B(T) - 1, S\big) m_{T,t}(\mathrm{d}S)$$

$$\leq 2 \cdot 18^{N-1} \frac{|\mathcal{T}_N|}{(2N-3)!!} t^{2N-2} \cdot \sum_{x \in \mathcal{G}} \|D_x f\|_{p(t)}^2$$

$$+ 2 \cdot |\mathcal{N}|^2 \big(1 + |\mathcal{N}|\big)^2 \sum_{T \in \mathcal{T}_{N+1}} 18^{\#(B(T)-1)_-} \int_{\mathcal{S}_{T,t}} I\big(T, B(T) - 1, S\big) m_{T,t}(\mathrm{d}S).$$

Fix $T \in \mathcal{T}_N$ and let $v \in \partial T$. Now we apply Corollary 8.3 to $I(T, v, S)$. If $v + 1 = \infty$, then

$$I(T, v, S) \leq 2 \sum_{x \in \mathcal{G}} \|D_x f\|_{p(t)}^2 + 2 \cdot |\mathcal{N}|^2 \big(1 + |\mathcal{N}|\big)^2 S(v) \int_0^{S(v)} I\big(T'_v, v' - 1, S'_{r,v}\big) \mathrm{d}r.$$

Otherwise, $v + 2 < \infty$ as well, and using Lemma 8.1 in addition to Corollary 8.3 yields

$$I(T, v, S) \leq 2I(T, v + 1, S) + 2 \cdot |\mathcal{N}|^2 \big(1 + |\mathcal{N}|\big)^2 S(v) \int_0^{S(v)} I\big(T'_v, v' - 1, S'_{r,v}\big) \mathrm{d}r$$

$$\leq 18 I(T, v + 2, S) + 2 \cdot |\mathcal{N}|^2 \big(1 + |\mathcal{N}|\big)^2 S(v) \int_0^{S(v)} I\big(T'_v, v' - 1, S'_{r,v}\big) \mathrm{d}r.$$

As $v + 2$ is again a leaf, we can iterate. Therefore,

$$I\big(T, B(T) - 1, S\big) \leq \sum_{\substack{v \in \partial T \\ v \geq B(T)-1}} 18^{\#v_- - \#(B(T)-1)_- } 2$$

$$\cdot |\mathcal{N}|^2 \big(1 + |\mathcal{N}|\big)^2 S(v) \int_0^{S(v)} I\big(T'_v, v' - 1, S'_{r,v}\big) \mathrm{d}r$$

$$+ 2 \cdot 18^{N-1-\#(B(T)-1)_-} \sum_{x \in \mathcal{G}} \|D_x f\|_{p(t)}^2.$$

Integrating with respect to $m_{T,t}$ and multiplying with $18^{\#(B(T)-1)_-}$, we get

$$18^{\#(B(T)-1)_-} \int_{\mathcal{S}_{T,t}} I(T, v, S) m_{T,t}(\mathrm{d}S)$$

$$\leq \sum_{\substack{v \in \partial T \\ v \geq B(T)-1}} 18^{\#v_- } 2 \cdot |\mathcal{N}|^2 \big(1 + |\mathcal{N}|\big)^2 \int_{\mathcal{S}_{T,t}} S(v) \int_0^{S(v)} I\big(T'_v, v' - 1, S'_{r,v}\big) \mathrm{d}r\, m_{T,t}(\mathrm{d}S)$$

$$+ 2 \cdot 18^{N-1} \int_{\mathcal{S}_{T,t}} 1\, m_{T,t}(\mathrm{d}S) \sum_{x \in \mathcal{G}} \|D_x f\|_{p(t)}^2.$$

Note that by Lemma 9.1,

$$\int_{\mathcal{S}_{T,t}} 1\, m_{T,t}(\mathrm{d}S) \leq \frac{t^{2N-2}}{(2N-3)!!}.$$

By construction of the measures $m_{T,t}$ and using the Lemma 9.2,

$$\int_{\mathcal{S}_{T,t}} S(v) \int_0^{S(v)} I\big(T_v', v'-1, S_{r,v}'\big)\,\mathrm{d}r\,m_{T,t}(\mathrm{d}S)$$
$$= \int_{\mathcal{S}_{T_v',t}} I\big(T_v', B\big(T_v'\big)-1, S'\big) m_{T_v',t}\big(\mathrm{d}S'\big).$$

Summing over $T \in \mathcal{T}_N$ completes the proof by using Lemma 9.2 again. $\qquad\square$

Proof of Theorem 2.7. The proof consists of two parts. In a first step, we will show that for $t \le 1$

$$\sum_{x\in\mathcal{G}} \|D_x P_t f\|_2^2 \le 2\mathrm{e}^{\lambda t^2} \sum_{x\in\mathcal{G}} \|D_x f\|_{p(t)}^2, \tag{39}$$

with $\lambda = 72 \cdot |\mathcal{N}|^2(1+|\mathcal{N}|)^2$.

It is known from combinatorics that $|\mathcal{T}_{n+1}| = C_n$, where $C_n = \frac{(2n)!}{(n+1)!n!}$ is the nth Catalan number [15]. Therewith

$$\frac{|\mathcal{T}_{n+1}|}{(2n-1)!!} = \frac{2^n}{(n+1)!},$$

and by Theorem 9.3,

$$\sum_{x\in\mathcal{G}} \|D_x P_t f\|_2^2 \le 2\sum_{n=0}^{\infty} \frac{1}{(n+1)!}\big(72|\mathcal{N}|^2\big(1+|\mathcal{N}|\big)^2 t^2\big)^n \le 2\mathrm{e}^{72|\mathcal{N}|^2(1+|\mathcal{N}|)^2 t^2},$$

showing (39). To obtain the claim (15) from (39) we first note that without changes in the proofs except for notation

$$\sum_{x\in\mathcal{G}} \|D_x P_t f\|_{p(s)}^2 \le 2\mathrm{e}^{\lambda t^2} \sum_{x\in\mathcal{G}} \|D_x f\|_{p(s+t)}^2 \tag{40}$$

holds as well for any $s \ge 0$. When we write $P_t = (P_{t/\lceil t\rceil})^{\lceil t\rceil}$, and apply (40) $\lceil t\rceil$ times, we get

$$\sum_{x\in\mathcal{G}} \|D_x P_t f\|_2^2 \le 2^{\lceil t\rceil}\mathrm{e}^{\lambda(t/\lceil t\rceil)^2} \sum_{x\in\mathcal{G}} \|D_x f\|_{p(t)}^2.$$

Using $\lceil t\rceil \le t+1$ and $t/\lceil t\rceil \le 1$ yields (15) with $\widetilde{C} = 2\mathrm{e}^{\lambda}$ to show the theorem. $\qquad\square$

References

[1] I. Benjamini, G. Kalai and O. Schramm. First passage percolation has sublinear distance variance. *Ann. Probab.* **31** (4) (2003) 1970–1978. MR2016607

[2] J. van den Berg and D. Kiss. Sublinearity of the travel-time variance for dependent first-passage percolation. *Ann. Probab.* **40** (2) (2012) 743–764. MR2952090

[3] S. G. Bobkov and F. Götze. Exponential integrability and transportation cost related to logarithmic Sobolev inequalities. *J. Funct. Anal.* **163** (1) (1999) 1–28. MR1682772

[4] S. G. Bobkov and C. Houdré. A converse Gaussian Poincaré-type inequality for convex functions. *Statist. Probab. Lett.* **44** (3) (1999) 281–290. MR1711613

[5] J. Bourgain, J. Kahn, G. Kalai, Y. Katznelson and N. Linial. The influence of variables in product spaces. *Israel J. Math.* **77** (1–2) (1992) 55–64. MR1194785

[6] D. Cordero-erausquin and M. Ledoux. Hypercontractive measures, Talagrand's inequality, and influences. In *Geometric Aspects of Functional Analysis* 169–189. *Lecture Notes in Math.* **2050**. Springer, Heidelberg, 2012. MR2985132

[7] E. Friedgut. Sharp thresholds of graph properties, and the k-sat problem. *J. Amer. Math. Soc.* **12** (4) (1999) 1017–1054. With an appendix by Jean Bourgain. MR1678031

[8] A. Guionnet and B. Zegarlinksi. Lectures on logarithmic Sobolev inequalities. In *Séminaire de Probabilités XXXVI* 1–134. J. Azéma, M. Émery, M. Ledoux and M. Yor (Eds). *Lecture Notes in Math.* **1801**. Springer, Berlin, 2003. MR1971582

[9] R. Holley and D. Stroock. Logarithmic Sobolev inequalities and stochastic Ising models. *J. Stat. Phys.* **46** (1987) 1159–1194. MR0893137

[10] J. Kahn, G. Kalai and N. Linial. The influence of variables on Boolean functions. In *Proceedings of the 29th Annual Symposium on Foundations of Computer Science (SFCS'88)* 68–80. IEEE Computer Society Press, Washington, DC, 1988. DOI: 10.1109/SFCS.1988.21923.

[11] N. Keller, E. Mossel and A. Sen. Geometric influences. *Ann. Probab.* **40** (3) (2012) 1135–1166. MR2962089

[12] M. A. Krasnosel'skiĭ and I. A. B. Rutitskiĭ. *Convex Functions and Orlicz Spaces*. P. Noordhoff, Groningen, 1961. MR0126722

[13] F. Martinelli and E. Olivieri. Approach to equilibrium of Glauber dynamics in the one phase region. I. The attractive case. *Comm. Math. Phys.* **161** (3) (1994) 447–486. MR1269387

[14] F. Martinelli and E. Olivieri. Approach to equilibrium of Glauber dynamics in the one phase region. II. The general case. *Comm. Math. Phys.* **161** (3) (1994) 487–514. MR1269388

[15] R. P. Stanley. *Enumerative Combinatorics*, **2**. Cambridge Univ. Press, Cambridge, 1999. MR1676282

[16] M. Talagrand. On Russo's approximate zero–one law. *Ann. Probab.* **22** (3) (1994) 1576–1587. MR1303654