ALEA, Lat. Am. J. Probab. Math. Stat. 20, 1283–1314 (2023)
DOI: 10.30757/ALEA.v20-48

Concentration inequalities for some negatively dependent binary random variables

Radosław Adamczak and Bartłomiej Polaczyk

Institute of Mathematics, University of Warsaw, Poland
E-mail address: R.Adamczak, B.Polaczyk @mimuw.edu.pl

Abstract. We investigate concentration properties of functions of random vectors with values in the discrete cube, satisfying the stochastic covering property (SCP) or the strong Rayleigh property (SRP).

Our results for SCP measures include subgaussian inequalities of bounded-difference type extending classical results by Pemantle and Peres and their counterparts for matrix-valued setting strengthening recent estimates by Aoun, Banna and Youssef. Under the stronger assumption of the SRP we obtain Bernstein-type inequalities for matrix-valued functions, generalizing recent bounds for linear combinations of positive definite matrices due to Kyng and Song.

We also treat in detail the special case of independent Bernoulli random variables conditioned on their sum for which we obtain strengthened estimates, deriving in particular modified log-Sobolev inequalities, Talagrand's convex distance inequality and, as corollaries, concentration results for convex functions and polynomials, as well as improved estimates for matrix-valued functions. These results generalize inequalities for the uniform measure on slices of the discrete cube, studied extensively by many authors. This case is based on a new, abstract condition implying strong concentration inequalities on the discrete cube (which is of independent interest) and recent results by Hermon and Salez.

1. Introduction

Investigating families of binary random variables with negatively dependent coordinates is an important problem from the point of view of computer science, statistics and combinatorics, which in the recent years has attracted considerable attention, see, e.g., Pemantle (2000); Shao (2000); Borcea et al. (2009); Pemantle and Peres (2014); Kyng and Song (2018); Garbe and Vondrak (2018); Bertail and Clémençon (2019); Anari et al. (2019); Aoun et al. (2020); Kathuria (2020). A wide and important class of such variables is constituted by those satisfying the *strong Rayleigh property* (abbrev. SRP) introduced in Borcea et al. (2009). More precisely, a probability measure π on the

Received by the editors July 22nd, 2022; accepted July 18th, 2023.

2010 *Mathematics Subject Classification.* 60E15, 60B20, 60J28.

Key words and phrases. concentration of measure, strong Rayleigh measure, stochastic covering property.

Research partially supported by the National Science Centre, Poland, via the Sonata Bis grant no. 2015/18/E/ST1/00214 (RA) and the Preludium grant no. 2020/37/N/ST1/02667 (BP).

hypercube $\mathcal{B}_{\mathrm{n}} := \{0,1\}^n$ satisfies the SRP if its generating polynomial

$$\mathbb{C}^n \ni z \mapsto \sum_{x\in\mathcal{B}_{\mathrm{n}}} \pi(x) \prod_{i=1}^{n} z_i^{x_i}$$

has no roots z whose all coordinates lie in the (strict) upper half-plane. The examples of such measures are, e.g., the law of independent Bernoulli random variables conditioned on their sum, determinantal measures, uniform measure on the bases of balanced matroids and laws of point processes or measures obtained by running exclusion dynamics on the cube, cf. Pemantle and Peres (2014).

The main purpose of this article is to deepen the understanding of the concentration of measure phenomenon in the context of strong Rayleigh distributions and related classes of probability measures on the discrete cube. In some of our considerations we will exploit only a more general notion of the *stochastic covering property* (abbrev. SCP, cf. Definition 2.1) introduced in Pemantle and Peres (2014), since this condition already turns out to provide a useful framework for proving concentration results, cf. Pemantle and Peres (2014); Hermon and Salez (2023); Aoun et al. (2020); Kyng and Song (2018); Kathuria (2020); Adamczak et al. (2022). On the other hand for some more specialized inequalities we will restrict our attention to independent Bernoulli variables conditioned on their sum taking some fixed value. Distributions of this type generalize the uniform measure on slices of the discrete cube, related to the Bernoulli–Laplace model, which has been extensively studied, e.g., in Lee and Yau (1998); Bobkov and Tetali (2006); Gao and Quastel (2003) and more recently in Samson (2017) and Sambale and Sinulis (2022). The non-uniform distribution given by conditioned Bernoulli variables has found applications, e.g., in survey sampling being a model of sampling without replacement from a finite population, with prescribed inclusion probabilities, which maximizes the entropy (often referred to as *conditional Poisson sampling*). We refer to Chen and Liu (1997); Chen (2000); Chen et al. (1994); Tillé (2006); Bertail and Clémençon (2019) for properties and applications of this family of distributions.

1.1. *State of the art.* The landmark paper that initiated the study of concentration phenomenon implied by the SCP is due to Pemantle and Peres (2014) who, using the martingale method, proved a sub-Gaussian concentration bound for measures satisfying the SCP and functions that are Lipschitz with respect to the Hamming distance $d_H(x,y) = \sum_i \mathbf{1}_{x_i \neq y_i}$. Recently, Hermon and Salez (2023) building on the works Lu and Yau (1993); Jerrum et al. (2004), retrieved this estimate by proving that the SCP implies the modified log-Sobolev inequality. Their result is one of many recent breakthrough achievements relating various types of negative dependence for binary random variables to logarithmic Sobolev inequalities, see, e.g., Anari et al. (2019); Kaufman and Oppenheim (2018); Cryan et al. (2021); Anari et al. (2021) – we provide a more detailed description of these developments in subsequent sections.

These findings in terms of concentration of measure can be summarized as follows (for a probability measure π on $\mathcal{B}_{\mathrm{n}}$ and $f\colon \mathcal{B}_{\mathrm{n}} \to \mathbb{R}$, we use the notation $\pi(f) := \int f\, d\pi$).

Theorem 1.1 (Pemantle and Peres (2014))**.** *For a probability measure π on $\mathcal{B}_{\mathrm{n}}$ satisfying the SCP and any $f\colon \mathcal{B}_{\mathrm{n}} \to \mathbb{R}$ such that*

$$|f(x) - f(y)| \le d_H(x,y) \quad \forall\, x,y \in \mathcal{B}_{\mathrm{n}}$$

the following estimate holds for all $t > 0$:

$$\pi\big(f > \pi(f) + t\big) \le \exp(-t^2/8n). \tag{1.1}$$

If π is k-homogeneous (i.e., it is supported on the set of binary vectors with exactly k coefficients equal to one), then n in the above expression can be replaced with k.

Recently, a sub-exponential version of Theorem 1.1 for matrix-valued functions has been shown in Aoun et al. (2020), where the authors develop a general framework for deducing concentration bounds for matrix-valued functions from the Poincaré inequality. A Bernstein-type bound for measures with the SRP, which in certain situations may give stronger concentration, has been also developed in Kyng and Song (2018) and Kaufman et al. (2022) for functions of the form $f(x) = \sum_{i=1}^{n} x_i C_i$, where C_i are nonnegative definite matrices (see Theorem 2.8 and Remark 2.10 below).

While concentration estimates and functional inequalities for general SCP measures are relatively recent, investigation of uniform measures on slices of the discrete cube in this context has much longer history. Such measures are of interest in relation to the Bernoulli–Laplace models of statistical physics and to uniform sampling without replacement. In particular Lee and Yau studied the Poincaré and log-Sobolev inequalities for such measures, whereas Bobkov and Tetali (2006) and independently Gao and Quastel (2003) investigated modified log-Sobolev inequalities relevant for concentration estimates. Strong concentration results for this case can be also obtained by projection from Talagrand's convex distance inequality for uniform measure on the symmetric group Talagrand (1995). Samson (2017) complemented this approach by proving corresponding transportation inequalities. Very recently Sambale and Sinulis (2022), investigating general multislices, recovered convex distance inequalities by means of functional inequalities and also obtained concentration for polynomials. One should stress that concentration results for slices of the cube provided by the above references are substantially stronger than those in the spirit of (1.1) coming from more general inequalities for SCP or SRP measures.

The uniform measure on slices of the cube can be seen as a special case of the distribution of independent Bernoulli random variables conditioned on their sum, when all the variables have the same probability of success. Such general distributions are known to be strong Rayleigh. To our best knowledge there has not been much work concerning refined concentration inequalities for general measures of this type. The only exception we are aware of is a recent article Bertail and Clémençon (2019), in which the authors, motivated by applications to survey sampling, obtain precise Bernstein-type inequalities for linear functionals.

1.2. *Overview of main results.* As mentioned in the prequel, various breakthrough results concerning negatively dependent measures on the discrete cube have been recently obtained, in particular in the context of functional inequalities. They have lead to optimal rates for the speed of convergence of the associated Markov chains allowing for improved sampling algorithms. Many of them also yield concentration results in the spirit of Theorem 1.1. Despite these important developments, the theory of concentration of measure for negatively dependent measures has not yet reached the level of completeness comparable to its counterpart in the independent setting. This concerns among others

- generalization of (1.1) to weighted Hamming distances, which would lead to a counterpart of the bounded difference inequality and allow to treat many functionals naturally arising in combinatorics or high dimensional geometry (see, e.g., the survey article McDiarmid (1998) or the monograph Boucheron et al. (2013)),
- improved bounds for special classes of functions, e.g., subgaussian bounds for convex Lipschitz functions, which in the independent setting were obtained first by Talagrand from his celebrated convex distance inequality (cf. Talagrand (1988, 1995)), or bounds for polynomials which are especially important for the discrete cube due to their relation with the Fourier–Walsh expansion (see, e.g., O'Donnell (2014); Kim and Vu (2000); Latała (2006); Adamczak and Wolff (2015)),
- inequalities for general matrix-valued functions, see Tropp (2012); Paulin et al. (2016) for a description of this rich theory in the independent setting (we note important results obtained for linear combinations with coefficients in positive definite matrices due to Kyng and Song

(2018); Kaufman et al. (2022) as well as a subexponential bound obtained in Aoun et al. (2020) by means of the matrix Poincaré inequality).

In the case of independent random variables, the results mentioned above have been obtained over the years by a mixture of many techniques, most importantly, the martingale approach, going back to Azuma (1967), Talagrand's powerful induction techniques (Talagrand (1995)) and functional inequalities brought forward in Ledoux (1995/97) and developed by many authors (see Boucheron et al. (2013) for a detailed list of references). The functional inequalities involved in the proofs in the independent setting from a broader perspective correspond to a special case of Glauber dynamics and involve changing just one coordinate in a product space at a time. This is in contrast with the dynamics considered in the aforementioned papers on negatively-dependent variables, especially in the homogeneous case. It turns out that the functional inequalities which are sufficient for proving strong results on the speed of convergence of the associated Markov processes may not lead directly to concentration results beyond (1.1). This in our opinion is the main obstacle in obtaining counterparts of classical strong concentration inequalities in the negatively depending setting. Our goal in this article is to explore such stronger concentration results, by adapting both the martingale and functional approach. Below we present informally our main results, referring for the details to the subsequent parts of the article.

The first series of results we obtain concerns general measures satisfying the SCP for which we refine the Azuma type martingale argument used in Pemantle and Peres (2014) and generalize Theorem 1.1 to Lipschitz functions with respect to more general weighted Hamming distances $d_\alpha(x, y) = \sum \alpha_i \mathbf{1}_{x_i \neq y_i}$ obtaining a bounded-difference type inequality (which corresponds to the first item on the list above). This is the content of Theorem 2.3. Next, we use the approach developed for the scalar case together with matrix bounded-difference inequality due to Tropp (2012) to get an analogous concentration for matrix-valued functions (Theorem 2.5), strengthening the results of Aoun et al. (2020), in particular obtaining a subgaussian inequality in place of a subexponential one. We note that the proof in Aoun et al. (2020) is based on the matrix Poincaré inequality, whereas our approach relies on matrix martingale inequalities. Under a stronger assumption of the SRP we are also able to extend the Bernstein-type inequality of Kyng and Song (2018) from linear combinations with coefficients in nonnegative definite matrices to general functions satisfying a matrix bounded-difference type assumptions (Theorem 2.8).

The second line of research presented in the article concerns the functional approach to improved concentration inequalities. We develop an abstract condition (Definition 4.3) based on a relation between the constant in the modified log-Sobolev inequality and some quantities related to the generator of the associated Markov process and show that this condition implies not only the bounded-difference type inequality but also Talagrand's convex distance inequality, matrix-Bernstein inequality and higher order concentration for tetrahedral polynomials.

It is natural to conjecture that our condition holds for an arbitrary SCP measure and an appropriately chosen Markov generator. While we are not able to prove it in such generality we show that this is the case for the distribution of Bernoulli random variables conditioned on their sum being equal to some constant, obtaining in particular all the aforementioned concentration results. This extends various previous works that treated uniform measures on slices of the hypercube to the case of non-uniform measures obtained by the above conditioning procedure. In particular, we extend the results on the modified log-Sobolev inequality for the Bernoulli–Laplace model due to Gao and Quastel (2003) and Bobkov and Tetali (2006), as well as the convex distance inequality and polynomial concentration obtained recently in Sambale and Sinulis (2022). We remark that conditioned Bernoulli distribution is a very natural generalization of the uniform measure on slices of the discrete cube, due to its relevance in survey sampling as well as information theoretic properties (as mentioned in the introduction, it is a measure with maximal entropy among all probability measures with prescribed inclusion probabilities). We refer to the survey article Chen (2000) for a description of statistical applications of conditional Bernoulli distributions and to the monograph Tillé

(2006) for an algorithmic perspective. We also mention that in recent years considerable attention in statistics has been devoted to Donsker type CLTs for empirical processes of sampling schemes, in particular for the conditional Poisson sampling (rejective sampling) relying on conditioned Bernoulli distribution (see, e.g., Bertail et al. (2017); Han and Wellner (2021)). We expect that improved concentration inequalities for this measure should lead to strengthened non-asymptotic estimates for such processes, as it was the case in the theory of empirical processes in independent random variables.

1.3. *Organization of the article.* In Section 2 we present our results concerning concentration for general measures satisfying the SCP/SRP. In Section 3 we specialize our analysis to Bernoulli random variables conditioned on their sum being equal to some constant. Then, in Section 4 we formulate an abstract framework that allows to deduce the results of Section 3. Finally, all the proofs are presented in Sections 5, 6 and 7.

2. Concentration under the SCP and SRP

In this section we present our concentration results for general measures satisfying the SCP or SRP. We start with introducing some notation. For $x = (x_1, \ldots, x_n) \in \mathcal{B}_{\mathrm{n}} := \{0,1\}^n$ and any $S \subset \{1, 2, \ldots, n\} =: [n]$ we use the shorthand notation $x_S = (x_i)_{i \in S}$. For any $r \in [n]$ we denote $x_{>r} = (x_i)_{i>r}$ (and analogously with relations other than $>$)[1]. We also write x^i for the vector obtained from $x \in \mathcal{B}_{\mathrm{n}}$ by flipping its i-th coordinate and x^{ij} for the vector obtained by swapping the i-th and j-th coordinate, i.e., $x^i = x \pm e_i$ and if $x_i \neq x_j$ then $x^{ij} = x \pm e_i \mp e_j$ for $i, j \in [n]$, where $e_i \in \mathcal{B}_{\mathrm{n}}$ is the vector with one on the i-th and zeros on the remaining coordinates; whereas $x^{ij} = x$ if $x_i = x_j$. We remark that the notation x^{ij} should not be confused with $(x^i)^j$. The law of a random variable X is denoted by $\mathcal{L}(X)$, whereas $\mathcal{L}(X|A)$ stands for the conditional law of X given an event A (with an analogous convention for conditioning with respect to σ-fields or other random variables).

Below we recall the definition of the SCP.

Definition 2.1 (Stochastic covering property)**.** For $x, y \in \mathcal{B}_{\mathrm{n}}$, we say that x covers y, denoted $x \rhd y$, if $x = y$ or $x = y + e_i$ for some $i \in [n]$.

A random variable X taking values in $\mathcal{B}_{\mathrm{n}}$ satisfies the SCP if for any $S \subset [n]$ and any $x, y \in \mathcal{B}_{\mathrm{n}}$ such that $\mathbb{P}(X_S = x_S), \mathbb{P}(X_S = y_S) > 0$ and $x_S \rhd y_S$, there exists a coupling (U, V) between the conditional distributions $\mathcal{L}(X_{S^c} \mid X_S = y_S)$ and $\mathcal{L}(X_{S^c} \mid X_S = x_S)$ such that $U \rhd V$. A measure π satisfies the SCP if X with law π does so.

Remark 2.2. As indicated in the introduction, the SCP is implied by the SRP, cf. Pemantle and Peres (2014). The opposite however is not true, as is demonstrated in, e.g., Cryan et al. (2021, Appendix A), where the authors study yet another possible generalization of the SRP, the strong log-concavity. In particular, they construct a distribution that is supported on the bases of a matroid, and that satisfies the SCP but violates the log-concavity (and hence violates also the SRP).

For a finite sequence x, we denote by $x^{\downarrow}$ the non-increasing rearrangement of the elements of x and for $\alpha \in [0,\infty)^n =: \mathbb{R}_+^n$ and $x, y \in \mathcal{B}_{\mathrm{n}}$ we define the α-weighted Hamming distance $d_\alpha(x,y) = \sum_i \alpha_i \mathbf{1}_{x_i \neq y_i}$. Finally, for $p \in [1,\infty]$, $|\cdot|_p$ is the ℓ_p norm on $\mathbb{R}^n$ and $|\cdot| := |\cdot|_2$ denotes the Euclidean norm.

The first main result of this paper is the following generalization of Theorem 1.1.

[1] We adopt the convention that if $x \in \mathcal{B}_{\mathrm{n}}$ then $x_{>n} = \emptyset$ and as a consequence, e.g., $\mathbb{P}(\cdot \mid X_{>n} = \emptyset) = \mathbb{P}(\cdot)$.

Theorem 2.3. *For a probability measure π on $\mathcal{B}_n$ satisfying the SCP, any $f\colon \mathcal{B}_n \to \mathbb{R}$ and $\alpha \in \mathbb{R}^n_+$ such that*
$$|f(x) - f(y)| \le d_\alpha(x,y) \quad \forall\, x,y \in \mathcal{B}_n$$
the following estimate holds for all $t > 0$:
$$\pi\big(f > \pi(f) + t\big) \le \exp(-t^2/8|\alpha|^2).$$
If the measure π is k-homogeneous, then $8|\alpha|^2$ in the above estimate can be replaced with $16\sum_{i=1}^{k}(\alpha_i^{\downarrow})^2$.

Remark 2.4. Theorem 2.3 implies Theorem 1.1 (up to an absolute constant in the exponent) by taking $\alpha = (1,1,\ldots,1)$. Moreover, by considering functions of the form $f(x) = \sum_i c_i x_i$ with $|c|^2 \ll n|c|^2_\infty$ in the non-homogeneous or $\sum_{i=1}^k (c_i^{\downarrow})^2 \ll k|c|^2_\infty$ in the k-homogeneous case, one can see that Theorem 2.3 can give substantially better concentration estimates than Theorem 1.1. We remark that such general linear functionals are important both from the abstract geometric perspective on high dimensional probability, but also from the statistical point of view. An important example is the Horvitz–Thompson estimator build over a sampling scheme defined by a k-homogeneous measure on the discrete cube (see, e.g., Bertail and Clémençon (2019)).

We now formulate the matrix analogue of Theorem 2.3. To this end, let us denote the space of d-dimensional Hermitian matrices by $\mathcal{H}_d$, the identity matrix in $\mathcal{H}_d$ by I_d, the maximal eigenvalue of $H \in \mathcal{H}_d$ by $\lambda_{max}(H)$ and the operator norm of H by $\|H\|$.

Theorem 2.5. *For a probability measure π on $\mathcal{B}_n$ satisfying the SCP, any $f\colon \mathcal{B}_n \to \mathcal{H}_d$ and $\alpha \in \mathbb{R}^n_+$ such that*
$$\|f(x) - f(y)\| \le d_\alpha(x,y) \quad \forall\, x,y \in \mathcal{B}_n \tag{2.1}$$
the following estimate holds for all $t > 0$:
$$\pi\big(\lambda_{max}(f - \pi(f)) > t\big) \le d\exp(-t^2/32|\alpha|^2).$$
If π is k-homogeneous then $32|\alpha|^2$ in the above estimate can be replaced with $64\sum_{i=1}^k(\alpha_i^{\downarrow})^2$.

Remark 2.6. Recently, Aoun et al. (2020) showed that for any k-homogeneous probability measure π on $\mathcal{B}_n$ satisfying the SCP and any $f\colon \mathcal{B}_n \to \mathcal{H}_d$ such that
$$\|f(x) - f(y)\| \le d_H(x,y) \quad \forall\, x,y \in \mathcal{B}_n$$
the following estimate applies:
$$\pi\big(\lambda_{max}(f - \pi(f)) > t\big) \le d\exp\Big(-\frac{t^2}{8k + 2t\sqrt{2k}}\Big). \tag{2.2}$$
The exponent in (2.2) is proportional to $-t/2\sqrt{2k}$ for t big enough and whence for such t Theorem 2.5 applied with $\alpha = (1,\ldots,1)$ strengthens on (2.2) (and on an analogous result from Kathuria (2020)) as it yields a sub-Gaussian estimate.

Remark 2.7. Using semigroup techniques together with matrix concentration results implied by the Poincaré inequality due to Aoun et al. (2020), we are also able to derive a sub-exponential concentration inequality for general measures satisfying the SCP under weaker assumptions on f than those of Theorem 2.5, cf. Remark 3.8.

When comparing the inequality of Theorem 2.5 or the results from Aoun et al. (2020) with results for matrix-valued functions of independent random variables, one can ask if it is possible to weaken the assumptions on the function f and instead of the Lipschitz constant with respect to d_α use some weaker parameter, involving bounds on the increments of the function in terms of the positive semidefinite order. In many situations one encounters functions for which $(f(x)-f(x^i))^2 \preccurlyeq C_i^2$ where

C_i are some positive semidefinite matrices and $\preccurlyeq$ stands for the positive semidefinite order (note that considering arbitrary matrices C_i is a generalization of the condition (2.1), which corresponds to the special case $C_i^2 = \alpha_i^2 I_d$). The simplest, yet important situation of this type is given by $f(x) = \sum_{i=1}^n x_i C_i$. Inequalities for such functions together with algorithmic applications were considered in Kyng and Song (2018). It turns out that their approach can be adapted to the setting of general functions, yielding the following theorem.

Theorem 2.8. *Let π be a k-homogeneous probability measure $\mathcal{B}_\mathrm{n}$ satisfying the strong Rayleigh property and $f\colon \mathcal{B}_\mathrm{n} \to \mathcal{H}_\mathrm{d}$ be such that there exists a sequence $C_1, \dots, C_n \in \mathcal{H}_\mathrm{d}$ satisfying*

$$(f(x) - f(x^i))^2 \preccurlyeq C_i^2 \quad \forall\, x \in \mathcal{B}_\mathrm{n},\ i \in [n]. \tag{2.3}$$

Then for any $t > 0$,

$$\pi\big(\lambda_{max}(f - \pi(f)) > t\big) \leq d \exp\Big(- \frac{t^2}{8\|\pi(\tilde{f})\| \log(ek) + \frac{4}{3}Kt}\Big), \tag{2.4}$$

where $\tilde{f}(x) = \sum_{i=1}^n x_i C_i^2$ and $K = \max_{i \leq n} \|C_i\|$.

Remark 2.9. In fact, the only place in the proof of Theorem 2.8 where we use the SRP in its full strength is to get that $\mathbb{P}(X_i = 1 \,|\, X_{i_1} = 1, \dots, X_{i_l} = 1) \leq \mathbb{P}(X_i = 1)$ for $X \sim \pi$ and any $i, l \in [n]$ and $\{i_1, \dots, i_l\} \subset [n] \setminus \{i\}$. Therefore, in Theorem 2.8 it suffices to assume that π satisfies the SCP and negative association, which is implied by the SRP, cf. Pemantle and Peres (2014).

Remark 2.10. It is natural to expect that $\log(ek)$ in (2.4) is just an artefact of the proof. Very recently, in Kaufman et al. (2022) the authors have obtained a Chernoff type inequality for functions of the form $f(x) = \sum_{i=1}^n x_i C_i$ for positive semidefinite matrices C_i, not containing this logarithmic factor, which improved certain algorithmic constructions related to graph sparsifiers constructed via random spanning trees, cf. Kyng and Song (2018).

Let us also point out that despite the logarithmic factor present in Theorem 2.8, when we specialize it to the linear function f as discussed above, it is not directly comparable with the result from Kaufman et al. (2022), which instead of $\|\pi(\tilde{f})\|$ uses a larger quantity $K\|\pi(\hat{f})\|$ with $\hat{f}(x) = \sum_{i=1}^n x_i C_i$ (recall that C_i's are nonnegative definite).

3. Concentration for conditional Bernoullis

In this section, we present our concentration results concerning Bernoulli random variables conditioned on their sum being constant. These include Talagrand's convex distance inequality, matrix-Bernstein inequality and concentration for polynomials.

We start with introducing the notation. For a sequence $p = (p_1, \dots, p_n) \in (0,1)^n$, let $B = (B_1, \dots, B_n)$ be a sequence of independent Bernoulli random variables with probabilities of success p_i, i.e., $\mathbb{P}(B_i = 1) = 1 - \mathbb{P}(B_i = 0) = p_i$ for $i \in [n]$. We set $X = (X_1, \dots, X_n) \sim \mathcal{L}\big(B \,|\, \sum_i B_i = k\big)$ for some $k \in \{0, \dots, n\}$ and denote the distribution of X by $\pi(p, k)$.

Our first contribution is a counterpart to the celebrated convex distance inequality, introduced for the first time in Talagrand (1988) for product measures on the cube.

Theorem 3.1. *If $\pi = \pi(p,k)$ for some $p \in (0,1)^n$ and $k \in \{0, \dots, n\}$, then for any $A \subset \mathcal{B}_\mathrm{n}$,*

$$\pi(A)\pi\big(d_T^2(\cdot, A)/84\big) \leq 1,$$

where

$$d_T(x, A) = \sup_{\alpha\colon |\alpha| \leq 1} d_\alpha(x, A) \quad \text{for} \quad x \in \mathcal{B}_\mathrm{n},\ A \subset \mathcal{B}_\mathrm{n}\,.$$

Let $\mathbb{M}_\pi f$ denote any median of f with respect to the measure π. A classical consequence of Theorem 3.1 is the following fact regarding the concentration around the median of convex functions (cf. Boucheron et al. (2013)). Let us recall the classical observation that subgaussian concentration around median and mean for convex Lipschitz functions are equivalent up to the change of constants by a universal factor.

Corollary 3.2. *If $\pi = \pi(p,k)$ for some $p \in (0,1)^n$ and $k \in \{0,\dots,n\}$, then for any convex function $f\colon \mathbb{R}^n \to \mathbb{R}$ that is L-Lipschitz with respect to the standard Euclidean distance on $\mathbb{R}^n$ and any $t > 0$,*

$$\pi\big(|f - \mathbb{M}_\pi f| > t\big) \leq 4\exp\big(-t^2/84L^2\big).$$

Remark 3.3. If one is interested just in the lower tail of a convex function, then one can in fact replace the Lipschitz constant L by $\pi\big(|\nabla f|\big)$ or even certain quantiles of $|\nabla f|$. We do not pursue this direction here and refer the reader to Adamczak and Strzelecki (2019).

Remark 3.4. If $f\colon \mathcal{B}_\mathrm{n} \to \mathbb{R}$ is d_α 1-Lipschitz, then it can be extended to a function on $\mathbb{R}^n$ which is $|\alpha|$-Lipschitz with respect to the standard Euclidean distance. Therefore, Corollary 3.2 counterparts Theorem 2.3 in the sense that it yields the same concentration profile while allowing for a weaker Lipschitz condition on f at the cost of assuming convexity. The standard Euclidean norm itself is an example of a convex 1-Lipschitz function which is not 1-Lipschitz with respect to any d_α with $|\alpha|^2 < n$. In this case Corollary 3.2 gives a much better estimate than Theorem 2.3.

Remark 3.5. Among typical applications of Corollary 3.2 one can mention estimates for functions of the form

$$f(x) = \Big\|\sum_{i=1}^n x_i v_i\Big\|, \tag{3.1}$$

where v_i's are vectors in a Banach space $(F, \|\cdot\|)$. In this case the Lipschitz constant can be expressed as

$$L = \sup_{w \in B_{F^*}} \Big(\sum_{i=1}^n w(v_i)^2\Big)^{1/2},$$

where B_{F^*} stands for the unit ball in the dual F^* of F. One of the consequences of Corollary 3.2 are, e.g., moment estimates of the form

$$\Big(\pi(\|f - \pi f\|^p)\Big)^{1/p} \leq C\sqrt{p}L$$

for a universal constant C. In the case of Rademacher variables they were proved in Talagrand (1988) and are a strengthening of classical Khintchine–Kahane inequalities. Note that the class of functions of the form (3.1) covers the case of suprema of empirical processes related to sampling schemes which are of interest in statistics. We also remark that beyond geometric application, through Corollary 3.2, Theorem 3.1 in the independent case allows for corollaries of combinatorial nature, concerning, e.g., longest common subsequences of random words and other extremal functionals (see Talagrand (1995); McDiarmid (1998)). Some of them may be of interest also in the dependent case, however we do not pursue this direction here.

Our next result concerns concentration for matrix-valued functions under weaker assumptions than those in Theorem 2.5.

Theorem 3.6. *Let $\pi = \pi(p,k)$ for some $p \in (0,1)^n$ and $k \in \{0,\dots,n\}$. Assume that $f\colon \mathcal{B}_\mathrm{n} \to \mathcal{H}_\mathrm{d}$ is such that there is a sequence of positive semidefinite matrices $C_1,\dots,C_n$ satisfying*

$$(f(x) - f(x^i))^2 \preccurlyeq C_i^2, \quad \forall\, x \in \mathcal{B}_\mathrm{n},\ i \in [n], \tag{3.2}$$

where $\preccurlyeq$ denotes the partial ordering of the set of positive semidefinite matrices. Define the variance proxy

$$\sigma^2 = 16 \sup \Big\{ \| \sum_{i \in \mathcal{I}} C_i^2 \| : \ |\mathcal{I}| = k, \, \mathcal{I} \subset [n] \Big\}.$$

Then for any $t > 0$,

$$\pi\big(\lambda_{max}(f - \pi(f)) > t\big) \leq d \exp\big(- t^2/(\sigma^2 + \sigma t)\big).$$

Remark 3.7. Condition (3.2) implies that f is 1-Lipschitz with respect to the distance d_α with $\alpha_i = \|C_i\|$. On the other hand, for many choices of matrices $C_1, \ldots, C_n$ it happens that $\sigma^2 \ll \sum_{i=1}^k \big(\|C_i\|^2\big)^{\downarrow}$ as $n, k \to \infty$. Therefore, while yielding only sub-exponential concentration as opposed to the sub-Gaussian one given by Theorem 2.5, Theorem 3.6 may improve significantly on Theorem 2.5 through better parameters in the exponent.

Remark 3.8. By an adaptation of the proof of Theorem 3.6, one can obtain a similar result for general k-homogeneous measures satisfying the SCP condition with the variance proxy parameter

$$\sigma^2 = 8 \sup \Big\{ \|\sum_{i \in \mathcal{I}} C_i^2 \| + k \max_{i \notin \mathcal{I}} \|C_i^2\| : |\mathcal{I}| \leq k, \, \mathcal{I} \subset [n] \Big\}.$$

Finally, we turn to the higher order concentration. By the Fourier–Walsh expansion (see, e.g., O'Donnell (2014)), every function $f \colon \mathcal{B}_{\mathrm{n}} \to \mathbb{R}$ can be written in a unique way as a tetrahedral polynomial, i.e., a polynomial which is affine with respect to every variable (in particular the degree of the polynomial is at most n). Therefore in what follows we restrict our attention to this representation. In particular, when we speak about the gradient $\nabla f = (\partial_1 f, \ldots, \partial_n f)$ or higher order derivatives $\nabla^k f$, we always think of the usual derivatives of the polynomial function on $\mathbb{R}^n$ given by the tetrahedral representation of f (sometimes referred to as the harmonic extension of f). We remark that the directional derivatives $\partial_i f$ coincide on $\mathcal{B}_{\mathrm{n}}$ with the discrete derivatives of f given by $D_i f(x) = f(\max(x, x^i)) - f(\min(x, x^i))$, where the maximum and minimum are taken coordinatewise.

In order to formulate concentration of measure estimates for tetrahedral polynomials, we need to introduce a family of injective tensor product norms on d-index matrices (d-tensors). Let us recall the notation introduced in Latała (2006).

Let $|I|$ be the cardinality of a set I and for $\mathbf{i} = (i_1, \ldots, i_d) \in [n]^d$ let $|\mathbf{i}| = \max_{j \leq d} i_j$ and $|\mathbf{i}_I| = \max_{j \in I} i_j$. Denote by P_d the set of partitions of $[d]$ into nonempty, pairwise disjoint sets. For a partition $\mathcal{I} = \{I_1, \ldots, I_k\} \in P_d$, and a d-indexed matrix $A = (a_{\mathbf{i}})_{\mathbf{i} \in [n]^d}$, define

$$\|A\|_{\mathcal{I}} = \sup \Big\{ \sum_{\mathbf{i} \in [n]^d} a_{\mathbf{i}} \prod_{l=1}^k x^{(l)}_{\mathbf{i}_{I_l}} : |(x^{(l)}_{\mathbf{i}_{I_l}})| \leq 1, 1 \leq l \leq k \Big\},$$

where $|(x_{\mathbf{i}_{I_l}})| = \sqrt{\sum_{|\mathbf{i}_{I_l}| \leq n} x^2_{\mathbf{i}_{I_l}}}$. Therefore, for example,

$$\|(a_{ij})_{i,j \leq n}\|_{\{1,2\}} = \sup \Big\{ \sum_{i,j \leq n} a_{ij} x_{ij} : \sum_{i,j \leq n} x_{ij}^2 \leq 1 \Big\} = \|(a_{ij})_{i,j \leq n}\|_{HS},$$

$$\|(a_{ij})_{i,j \leq n}\|_{\{1\}\{2\}} = \sup \Big\{ \sum_{i,j \leq n} a_{ij} x_i y_j : \sum_{i \leq n} x_i^2 \leq 1, \sum_{j \leq n} y_j^2 \leq 1 \Big\} = \|(a_{ij})_{i,j \leq n}\|,$$

$$\|(a_{ijk})_{i,j,k \leq n}\|_{\{1,2\}\{3\}} = \sup \Big\{ \sum_{i,j,k \leq n} a_{ijk} x_{ij} y_k : \sum_{i,j \leq n} x_{ij}^2 \leq 1, \sum_{k \leq n} y_k^2 \leq 1 \Big\},$$

where $\| \cdot \|_{HS}$ and $\| \cdot \|$ denote the Hilbert–Schmidt and the operator norm respectively.

Theorem 3.9. *If $\pi = \pi(p, k)$ for some $p \in (0,1)^n$ and $k \in \{0, \dots, n\}$, then for any tetrahedral polynomial $f\colon \mathcal{B}_{\mathrm{n}} \to \mathbb{R}$ of degree d,*

$$\pi\big(\big|f - \pi(f)\big| \geq t\big) \leq 2\exp\Big(-\frac{1}{C_d}\min_{1\leq r\leq d}\min_{\mathcal{J}\in P_r}\Big(\frac{t}{\|\pi(\nabla^r f)\|_{\mathcal{J}}}\Big)^{2/|\mathcal{J}|}\Big),$$

where C_d is a constant depending only on the degree d of f.

Inequalities of this type for polynomials of arbitrary degree were introduced for the first time by Latała (2006) for tetrahedral polynomials in i.i.d. standard Gaussian variables. Subsequently they were extended to general polynomials in independent subgaussian random variables and to certain dependent situations related to Glauber dynamics (see Adamczak and Wolff (2015); Adamczak et al. (2019); Sambale and Sinulis (2020, 2022); Adamczak et al. (2022)). We remark that in the independent, subgaussian case and $d = 2$ they reduce to the well known Hanson–Wright inequality for quadratic forms (see, e.g., Vershynin (2018, Chapter 6)),

$$\mathbb{P}\Big(\Big|\sum_{i,j=1}^n a_{ij}X_iX_j - \mathbb{E}\sum_{i,j=1}^n a_{ij}X_iX_j\Big| \geq t\Big) \leq 2\exp\Big(-c\min\Big(\frac{t^2}{\|A\|_{HS}^2 + |\,\mathbb{E}\,AX|^2}, \frac{t}{\|A\|}\Big)\Big).$$

This inequality has proved useful in non-asymptotic analysis of random matrices and in asymptotic geometric analysis. It is worth mentioning that in the Gaussian case the inequalities in question may be reversed up to the value of the absolute constants, thus Theorem 3.8 shows that the measures $\pi(p, k)$ exhibit Gaussian type concentration for polynomials. While calculating the norms $\|\cdot\|_{\mathcal{J}}$ is usually difficult, estimating them is sometimes possible, leading to applications involving subgraph counts (in the Erdős-Rényi case or for some models of random graphs with dependencies Adamczak and Wolff (2015); Sambale and Sinulis (2020, 2022)) or to statistical applications, e.g., in testing Ising models Dagan et al. (2021) and signal processing Verzelen and Gassiat (2018).

4. Abstract formulations

In this section we recall some notions from the theory of Markov semigroups and formulate the abstract counterparts of the results of Section 3 and of Theorem 2.3. We believe that the results presented in this section might be of separate interest as they provide a general framework for proving concentration on the hypercube. We stress that most of the proof techniques that we exploit were known previously – our main contribution is the abstract formulation of these results by means of the novel stability condition (cf. Definition 4.3).

Throughout this section we will rely on the usual notions from the theory of Markov processes and Dirichlet forms specialized to finite state space. We will briefly recall them and refer to Liggett (2010); Bakry et al. (2012); Bobkov and Tetali (2006) for details.

4.1. *Modified log-Sobolev inequalities.* Let L be the generator of a jump Markov process on some finite probability space (M, π). In what follows we will sometimes treat L as a linear operator on $\mathbb{R}^M$ and sometimes identify it with the corresponding matrix, indexed by the elements of M.

Assume that L satisfies the detailed-balance condition

$$\forall\, x, y \in M \quad \pi(x)L(x,y) = \pi(y)L(y,x), \tag{4.1}$$

which implies that π is a stationary measure for the Markov process and L is self-adjoint on $L^2(\pi)$. In this article we consider only Markov processes satisfying the above condition, which may not be stated explicitly in all the results.

For a given L, we define $\Delta(L) := \max_x -L(x,x) = \max_x \sum_{y\colon y\neq x} L(x,y)$ and write $\mathcal{E}(f,g) = -\pi(fLg)$ for the Dirichlet form associated with L. In particular $\mathcal{E}(f,g) = \pi\big(\Gamma(f,g)\big)$, where $\Gamma\colon \mathbb{R}^M \times$

$\mathbb{R}^M \to \mathbb{R}^M$ given by

$$\Gamma(f,g)(x) = \frac{1}{2}\sum_{y\in M}(f(x)-f(y))(g(x)-g(y))L(x,y) \tag{4.2}$$

is the corresponding carré-du-champ operator. We use shorthand notation $\Gamma(f,f) =: \Gamma(f)$ and observe that by the detailed-balance condition (4.1) we have $\pi\big(\Gamma(f)\big) = \pi(\Gamma_+(f))$, where

$$\Gamma_+(f)(x) = \sum_{y\in M}(f(x)-f(y))_+^2 L(x,y). \tag{4.3}$$

Finally, we denote by $\rho(L)$ the best (the greatest) constant such that the following modified log-Sobolev inequality is satisfied

$$\rho(L)\operatorname{Ent}_\pi(f) \le \mathcal{E}(f,\log f) \tag{4.4}$$

for all functions $f\colon M \to (0,\infty)$, where $\operatorname{Ent}_\pi(f) = \pi(f\log f) - \pi(f)\log\pi(f)$ is the entropy functional. We remark that $\rho(L)$ is positive iff L is irreducible on the support of π (see the discussion in Bobkov and Tetali (2006) and Levin and Peres (2017, Chapter 12)). In what follows we will restrict our attention to this situation, without mentioning this assumption explicitly in each statement.

A classical observation, often referred to as Herbst's argument (cf. the monographs Ledoux (2001) and Boucheron et al. (2013)), says that for any $f\colon M \to \mathbb{R}$,

$$\pi\big(f > \pi(f) + t\big) \le \exp(-t^2\rho(L)/4\|\Gamma_+(f)\|_\infty), \tag{4.5}$$

where $\|\cdot\|_\infty$ stands for the norm in $L^\infty(\pi)$.

4.2. *Flip-swap random walks.* After Hermon and Salez (2023), we say that a kernel L generates a flip-swap random walk if $L(x,y) > 0$ implies that $x = y^i$ for some $i \in [n]$ (i.e., x and y differ by a flip) or $x = y^{ij}$ for some $i \neq j$, $i,j \in [n]$ (i.e., x and y differ by a swap). The main contribution of Hermon and Salez (2023) can be stated in the following way.

Theorem 4.1 (Hermon and Salez (2023))**.** *For any measure π on $\mathcal{B}_{\mathrm{n}}$ satisfying the SCP, there exists a kernel L generating a reversible flip-swap random walk with stationary measure π such that $\rho(L) \ge 1$ and $\Delta(L) \le n$. If π is also k-homogeneous, then $\Delta(L) \le 2k$ as well.*

Theorem 4.1, by means of Herbst's argument (4.5), implies (up to an absolute constant in the exponent) the estimate from Theorem 1.1 after observing that for a flip-swap random walk and any $f\colon \mathcal{B}_{\mathrm{n}} \to \mathbb{R}$ that is 1-Lipschitz with respect to the Hamming distance d_H,

$$\|\Gamma_+(f)\|_\infty \le \Delta(L)\cdot \max_{x,y\in\mathcal{B}_{\mathrm{n}}}\{\,(f(y)-f(x))_+^2 \colon L(x,y) > 0\,\} \le 4\Delta(L). \tag{4.6}$$

Remark 4.2. There are many examples of flip-swap random walks on the hypercube in the literature, including, e.g., the Bernoulli–Laplace model, Glauber dynamics or base exchange random walk on matroids, cf. e.g., Bobkov and Tetali (2006); Goel (2004); Sambale and Sinulis (2020); Cryan et al. (2021). We note that the results of this section apply to any flip-swap random walk as long as we have control of its stability (cf. Definition 4.3) constant.

It turns out that for the proofs of all the statements of Section 3 it suffices to demonstrate that the following condition is true for some reversible generator L with stationary measure $\pi(p,k)$ for which the modified log-Sobolev inequality (4.4) is known.

Definition 4.3 (Stability condition)**.** Let L be a generator of a flip-swap random walk on $\mathcal{B}_{\mathrm{n}}$ with invariant probability distribution π. We say that the pair (L,π) meets the stability condition with constant $R \geq 0$ (i.e., is R-stable) if it satisfies the modified log-Sobolev inequality (4.4) and

$$\max_{x \in \operatorname{supp} \pi;\, i \in [n]} \sum_{y\colon y_i \neq x_i} L(x,y) \leq R\rho(L). \tag{4.7}$$

If it is clear from the context which measure π is associated with L, we will often omit it in the discussion and simply say that L is R-stable.

Remark 4.4. If π is not concentrated on a single point, then a random walk on $\mathcal{B}_\mathrm{n}$ with a generator L that satisfies the modified log-Sobolev inequality (4.4) may be at best 0.25-stable (i.e., $R \geq 0.25$). Indeed, in this case there exists i such that $\pi(\{x_i = 1\}), \pi(\{x_i = 0\}) > 0$. If L satisfies the modified log-Sobolev inequality, then it also satisfies the Poincaré inequality $\frac{1}{2}\rho(L)\operatorname{Var}_\pi(f) \leq \mathcal{E}(f,f)$, see e.g., Adamczak et al. (2022, Proposition B.5). Let $f(x) = \mathbf{1}_{\{x_i=1\}}$. By the stability condition (4.7) and reversibility of L we get that

$$\begin{aligned}
R\rho(L)\pi(\{x_i = 1\}) &\geq \sum_{x\colon x_i=1} \sum_{y\colon y_i=0} L(x,y)\pi(x) \\
&= \sum_{x,y} (x_i - y_i)_+^2 L(x,y)\pi(x) \\
&= \mathcal{E}(f,f) \\
&\geq \frac{1}{2}\rho(L)\operatorname{Var}_\pi(f) = \frac{1}{2}\rho(L)\pi(\{x_i = 1\})\pi(\{x_i = 0\}),
\end{aligned}$$

which gives $R \geq 0.5 \cdot \pi(\{x_i = 0\})$. Similarly, by considering $f(x) = \mathbf{1}_{\{x_i=0\}}$ we get that $R \geq 0.5 \cdot \pi(\{x_i = 1\})$ as well, yielding $R \geq 0.25$.

This bound is optimal, as can be seen for π being the uniform measure on $\mathcal{B}_\mathrm{n}$ and $L(x,y) = 1$ if there exists i such that $y = x^i$, $L(x,y) = -n$ if $y = x$ and $L(x,y) = 0$ otherwise (this corresponds to the special case of Glauber dynamics, in which at rate n, a random coordinate is flipped). In this case $\rho(L) = 4$ (see Bobkov and Tetali (2006, Example 3.7), note a different normalization of both the Dirichlet form and the constant in the modified log-Sobolev inequality), whereas for all $x \in \mathcal{B}_\mathrm{n}$

$$\max_i \sum_{y\colon y_i \neq x_i} L(x,y) = L(x,x^i) = 1 = 0.25 \cdot \rho(L).$$

Let us illustrate the notion of R-stability with another classical example.

Example 4.5 (Bernoulli–Laplace model). Let π be the uniform measure on the slice of $\mathcal{B}_\mathrm{n}$ consisting of elements with exactly k ones and let L be given by $Lf(x) = \frac{1}{n}\sum_{i<j}(f(x^{ij}) - f(x))$ (thus the corresponding Markov process at rate $(n-1)/2$ swaps a uniformly chosen pair of coordinates). In the matrix form this corresponds to $L(x,y) = \frac{1}{n}$ if $x \neq y$ and $y = x^{ij}$, $L(x,x) = -k(n-k)/n$ and $L(x,y) = 0$ otherwise. It has been proved in Gao and Quastel (2003) and independently in Bobkov and Tetali (2006) that $\rho_0(L) \geq 1/2$. At the same time $\sum_{y\colon y_i \neq x_i} L(x,y)$ equals to $(n-k)/n$ if $x_i = 1$ and to k/n otherwise. This shows that L is 2-stable, independently of n and k. As mentioned in the introduction, the uniform measure on the slice of the discrete cube can be interpreted as the distribution of i.i.d. Bernoulli variables conditioned on their sum being equal to k. In Theorem 7.3 we generalize the above observation on stability and show that if μ is the law of general independent Bernoulli variables conditioned on their sum being equal to a fixed constant, there exists a 2-stable generator of a random walk reversible with respect to μ.

Remark 4.6. Observe that the notion of stability is invariant under scaling of L (change of time), i.e., if L is R-stable then so is aL for any $a > 0$. This leads to a tensorization property for measures admitting an R-stable generator. More precisely, let $\pi_1, \ldots, \pi_m$ be measures on $\mathcal{B}_{n_1}, \ldots, \mathcal{B}_{n_m}$, for which there exist reversible flip-swap random walks with R-stable generators $L_1, \ldots, L_m$. By changing time, we can assume without loss of generality that $\rho(L_i) = \rho$ for all $i \leq m$. Let $n =$

$n_1 + \ldots + n_m$ and consider the product measure $\pi = \pi_1 \otimes \cdots \otimes \pi_m$ on $\mathcal{B}_{\mathrm{n}}$ together with the generator $L = L_1 + \ldots + L_m$, where we think of L_i as acting only on the i-th block of coordinates on $\mathcal{B}_{\mathrm{n}} = \mathcal{B}_{n_1} \times \cdots \times \mathcal{B}_{n_m}$, i.e., we identify L_i with its tensor product with identity on $\otimes_{j\neq i}\mathbb{R}^{\mathcal{B}_{n_j}}$. In the matrix form we have the representation

$$L(x,y) = \sum_{i=1}^{m} L_i(P_i x, P_i y) \prod_{j\neq i} \mathbf{1}_{\{P_j x = P_j y\}},$$

where $P_j \colon \mathcal{B}_{\mathrm{n}} \to \mathcal{B}_j$ is the projection onto the j-th factor in the product $\mathcal{B}_{\mathrm{n}} = \mathcal{B}_{n_1} \times \cdots \times \mathcal{B}_{n_m}$ Thanks to the well known tensorization property of the entropy (see, e.g., Ané et al. (2000, Chapter 3)) we have $\rho(L) = \rho$. Moreover, for $i \in (n_1 + \ldots + n_{j-1}, n_1 + \ldots + n_j]$,

$$\sum_{y\in\mathcal{B}_{\mathrm{n}}\colon y_i\neq x_i} L(x,y) = \sum_{y\in\mathcal{B}_{n_j}\colon y_l\neq (P_j x)_l} L_j((P_j x), y) \leq R\rho,$$

where $l = i - (n_1 + \ldots + n_{j-1})$. Thus, L is indeed R-stable.

This observation allows in particular to extend all the theorems of Section 3 to product of measures $\pi(n,k)$ allowing for more general conditioning of Bernoulli variables.

4.3. *Concentration results.* Finally, we present the counterparts of the results of Section 3 and of Theorem 2.3 from Section 2 in the abstract language of the stability condition (4.7). We stress here that it is the sole property needed for their proofs, which are deferred to Section 7.

We start with a bounded-difference type inequality for real valued functions.

Proposition 4.7. *If a flip-swap random walk on $\mathcal{B}_{\mathrm{n}}$ with stationary distribution π and generator L satisfies the stability condition (4.7), then for any $f\colon \mathcal{B}_{\mathrm{n}} \to \mathbb{R}$ and $\alpha \in \mathbb{R}^n_+$ such that*

$$|f(x) - f(y)| \leq d_\alpha(x,y) \quad \forall\, x,y \in \mathcal{B}_{\mathrm{n}}$$

the following estimate holds for all $t > 0$

$$\pi\big(f > \pi(f) + t\big) \leq \exp\Big(-\frac{t^2}{8R|\alpha|^2}\Big).$$

In the above estimate one can also replace $8|\alpha|^2$ with $16\sum_{i=1}^{\lceil \Delta(L)/R\rho(L)\rceil} (\alpha_i^{\downarrow})^2$.

Remark 4.8. Using the definitions of R-stability and of $\Delta(L)$ one can see that $\Delta(L)/R\rho(L) \leq n$ and if π is k-homogeneous, then $\Delta(L)/R\rho(L) \leq k$.

We now pass to the matrix-valued case.

Proposition 4.9. *Let a flip-swap random walk on $\mathcal{B}_{\mathrm{n}}$ with stationary distribution π and generator L satisfy the stability condition (4.7). Assume also that $f\colon \mathcal{B}_{\mathrm{n}} \to \mathcal{H}_{\mathrm{d}}$ is such that there is a sequence of positive semidefinite matrices $C_1, \ldots, C_n$ satisfying*

$$(f(x) - f(x^i))^2 \preccurlyeq C_i^2 \quad \forall\, x \in \mathcal{B}_{\mathrm{n}},\ i \in [n], \tag{4.8}$$

where $\preccurlyeq$ denotes the positive semidefinite order on the set of symmetric matrices. Set the variance proxy

$$\sigma^2 = 8R \cdot \sup\Big\{ \big\| \sum_{i\in\mathcal{I}} C_i^2 \big\| : \ |\mathcal{I}| = \lceil \Delta(L)/R\rho(L)\rceil,\ \mathcal{I} \subset [n] \Big\}.$$

Then for any $t > 0$,

$$\pi\big(\lambda_{max}(f - \pi(f)) > t\big) \leq d\exp\big(-t^2/(\sigma^2 + \sigma t)\big).$$

Our next proposition is the convex distance inequality under R-stability.

Proposition 4.10. *If a flip-swap random walk on $\mathcal{B}_{\mathrm{n}}$ with some stationary distribution π and a generator L satisfies the stability condition (4.7), then for any set $A \subset \mathcal{B}_{\mathrm{n}}$*

$$\pi(A)\pi\Big(\exp\Big(\frac{1}{40R+4} \cdot d_T^2(\cdot, A)\Big)\Big) \le 1.$$

Finally, we state the concentration result for polynomials in an abstract version.

Proposition 4.11. *If a flip-swap random walk on $\mathcal{B}_{\mathrm{n}}$ with some stationary distribution π and a generator L satisfies the stability condition (4.7), then for any tetrahedral polynomial $f\colon \mathcal{B}_{\mathrm{n}} \to \mathbb{R}$ of degree d*

$$\pi\big(\big|f - \pi(f)\big| \ge t\big) \le 2\exp\Big(-\frac{1}{C_d} \min_{1\le r\le d} \min_{\mathcal{J}\in P_r} \Big(\frac{t}{R^{r/2}\|\pi(\nabla^r f)\|_{\mathcal{J}}}\Big)^{2/|\mathcal{J}|}\Big),$$

where C_d is a constant depending only on the degree d of f.

Remark 4.12. Although Proposition 4.7 gives a worse constant in the exponent than Theorem 2.3 even in the case of conditional Bernoulli distributions $\pi(p,k)$, we state it here as in principle it does not assume that π satisfies the SCP and thus potentially can be applied in other settings.

Remark 4.13. The above propositions can be transferred to more general random walks that change at each step at most a fixed number of coordinates N (with $N = 2$ in the case of flip-swap random walks). We do not pursue this direction though and do not write all the theorems in full generality for the sake of readability.

Remark 4.14. Recently, in Cryan et al. (2021) the authors have shown a version of Theorem 4.1 for k-homogeneous strongly log-concave measures. Strong log-concavity is yet another possible generalization of the SRP, which is in general incomparable with the SCP, cf. Cryan et al. (2021). It is known, cf. Brändén and Huh (2020), that any k-homogeneous strongly log-concave measure is supported on the set of bases of some matroid of rank k. Using this fact, and extending the previous results for uniform measures on the bases of matroids from Anari et al. (2019) and Kaufman and Oppenheim (2018), in Cryan et al. (2021), the authors explicitly construct a base-exchange random walk, which has any given strongly log-concave measure as a stationary distribution, and verify that it satisfies the modified log-Sobolev inequality (4.4).

Since the base-exchange random walk proposed therein is a particular instance of a flip-swap random walk, a natural question is whether it satisfies the stability condition (4.7), which would allow deducing concentration results presented in this section. Unfortunately, the answer seems to be negative in full generality as can be seen already in the case of independent Bernoulli random variables $B = (B_1, \ldots, B_n)$ with different probabilities of success $\mathbb{P}(B_i = 1) = p_i$ conditioned on their sum being k, i.e., for the distribution $\pi(p,k) = \mathcal{L}(B \mid \sum_i B_i = k)$. If one chooses $p_1 \to 1^-$ and $p_j = c$ for $j > 1$ and some $c \in (0,1)$, then it is straightforward to verify that the base-exchange random walk of Cryan et al. (2021) is at best k-stable. Therefore, applying propositions of Section 4.3 to the base-exchange random walk gives much worse concentration constants than those of Section 3. On the other hand, as we will show in Section 7, the abstract construction of a flip-swap random walk proposed in Hermon and Salez (2023), when specialized to $\pi(p,k)$ and implemented with a proper choice of couplings, gives 2-stability. The appropriate selection of couplings is the main ingredient in the proofs of results of Section 3.

In view of the above, it is an interesting problem to analyze what other known kernels satisfy the stability condition (4.7) with good (dimension-independent) constant and to look for some other criteria that would allow to deduce this condition.

5. Proofs of the results of Section 2

In this section we provide proofs of Theorems 2.3, 2.5 and 2.8. Our approach relies on certain refinements of the Azuma type martingale argument originally used in Pemantle and Peres (2014).

For Theorems 2.3, 2.5 it is based on an appropriate choice of the filtration, adapted to the structure of the function f, as described below.

Let $X \sim \pi$ be a random variable with values in $\mathcal{B}_{\mathrm{n}}$ satisfying the SCP and denote $\operatorname{supp} X = \{ i \in \{1, \ldots, n\} : X_i = 1 \}$. In the non-homogenous case define a filtration $\mathcal{F} = (\mathcal{F}_l)_{l=0}^n$ by letting simply $\mathcal{F}_0 = \{\emptyset, \Omega\}$ and $\mathcal{F}_l = \sigma(X_1, \ldots, X_l)$ for $l = 1, \ldots, n$. In the k-homogenous case introduce a family of random variables $Y_1, \ldots, Y_k$ given by the conditions

$$\begin{aligned} \mathcal{L}(Y_1 \,|\, X) &= \mathrm{Unif}(\operatorname{supp} X \backslash \{1, \ldots, k\}) \quad \text{and} \\ \mathcal{L}(Y_l \,|\, X, Y_1, \ldots, Y_{l-1}) &= \mathrm{Unif}(\operatorname{supp} X \setminus \{1, \ldots, k, Y_1, \ldots, Y_{l-1}\}), \quad \text{for} \quad l = 2, \ldots, k, \end{aligned} \tag{5.1}$$

where $\mathrm{Unif}(A)$ stands for the uniform distribution on the set A, and for notational simplicity we set $\mathrm{Unif}(\emptyset)$ to be the Dirac mass at 0 and $X_0 \equiv 1$ (i.e., we add to X an additional coordinate providing no information and if the above sampling scheme yields all elements from $\operatorname{supp} X$ before sampling some Y_l, we set Y_i to zero for all $i \geq l$). Finally, define a filtration $\mathcal{G} = (\mathcal{G}_l)_{l=0}^{2k}$ setting $\mathcal{G}_0 = \{\emptyset, \Omega\}$ and $\mathcal{G}_l = \sigma(X_1, \ldots, X_l)$ for $l \in [k]$, $\mathcal{G}_{k+r} = \sigma(X_1, \ldots, X_k, Y_1, \ldots, Y_r\}$ for $r \in [k]$.

In other words in the first k-steps the subsequent values of X at the first k coordinates are revealed, while in the last k steps one reveals in a uniformly random order the remaining coordinates at which X takes the value 1. Note that if α is non-increasing (which we will assume without loss of generality) and f is 1-Lipschitz with respect to d_α then the first part of this sampling scheme promotes the coordinates which may have the greatest impact on the value of $f(X)$. The construction can be thought of as a modification of the sampling scheme proposed by Pemantle and Peres in which one immediately starts revealing in a random order the coordinates at which X takes the value 1, which does not allow to capture the most sensitive coordinates.

The proof of Theorems 2.3 and 2.5 will be based on the following two lemmas.

Lemma 5.1. *Let $\alpha \in \mathbb{R}_+^n$ be non-increasing and let $f\colon \mathcal{B}_{\mathrm{n}} \to \mathcal{H}_{\mathrm{d}}$ be 1-Lipschitz with respect to the distance d_α. Assume that X is a $\mathcal{B}_{\mathrm{n}}$-valued random vector satisfying the SCP. Let $M_l = \mathbb{E}[f(X) \,|\, \mathcal{F}_l] - \mathbb{E}[f(X) \,|\, \mathcal{F}_{l-1}]$ for $l \in [n]$. Then for every $l \in [n]$,*

$$M_l^2 \preccurlyeq 4\alpha_l^2 I_d. \tag{5.2}$$

Lemma 5.2. *In the setting of Lemma 5.1, let us assume additionally that X is k-homogeneous. For $l \in [2k]$ define $N_l = \mathbb{E}[f(X) \,|\, \mathcal{G}_l] - \mathbb{E}[f(X) \,|\, \mathcal{G}_{l-1}]$. Then for $l \in [k]$,*

$$N_l^2 \preccurlyeq 4\alpha_l^2 I_d, \tag{5.3}$$

while for $l = k+1, \ldots, 2k$,

$$N_l^2 \preccurlyeq 4\alpha_k^2 I_d. \tag{5.4}$$

We postpone for now the proof of the above lemmas and firstly show how they imply Theorems 2.3 and 2.5. To this end let us recall the matrix version of the Azuma–Hoeffding inequality due to Tropp (2012, Theorem 7.1), which asserts that if D_l, $l = 1, \ldots, n$ are $\mathcal{H}_{\mathrm{d}}$-valued martingale differences and $D_l^2 \preccurlyeq C_l^2$ for some deterministic matrices $C_l \in \mathcal{H}_{\mathrm{d}}$, then for all $t \geq 0$,

$$\mathbb{P}\left(\lambda_{max}\Big(\sum_{l=1}^n D_l\Big) \geq t\right) \leq d e^{-t^2/8\sigma^2},$$

where $\sigma^2 = \|\sum_{l=1}^n C_l^2\|$. Note also that for $d = 1$ the classical Azuma–Hoeffding inequality (see, e.g., Dubhashi and Panconesi (2009, Theorem 5.8)) allows to replace the constant 1/8 by 1/2.

Proof of Theorems 2.3 and 2.5: Since the SCP is invariant under permutations of coordinates of X, we may and do assume that $\alpha = \alpha^{\downarrow}$. By Lemma 5.1 the martingale differences M_l satisfy $M_l^2 \preccurlyeq C_l^2 := 4\alpha_l^2 I_d$. Clearly

$$\Big\|\sum_{l=1}^n C_l^2\Big\| = 4|\alpha|^2. \tag{5.5}$$

If X is k-homogeneous, then by Lemma 5.2, $N_l^2 \preccurlyeq \widetilde{C}_l^2 := 4\alpha^2_{\min(l,k)} I_d$. In this case

$$\Big\| \sum_{l=1}^{2k} \widetilde{C}_l^2 \Big\| = 4\Big[\Big(\sum_{l=1}^{k} \alpha_l^2\Big) + k\alpha_k^2\Big] \le 8 \sum_{l=1}^{k} \alpha_l^2. \tag{5.6}$$

We have $f(X) = \sum_{l=1}^n M_l$, whereas in the k-homogeneous case $f(X) = \sum_{l=1}^{2k} N_l$ (observe that after $2k$-steps of the sampling procedure all the nonzero coordinates of X are revealed and so X is $\mathcal{G}_{2k}$-measurable). Thus, the conclusion of Theorem 2.3 follows by applying estimates (5.5) and (5.6) for $d = 1$ together with the classical Azuma–Hoeffding inequality. Similarly, Theorem 2.5 follows from the matrix version of the Azuma–Hoeffding inequality. □

It remains to prove Lemmas 5.1 and 5.2.

Proof of Lemma 5.1: Let $A_l^x = \{X_1 = x_1, \ldots, X_l = x_l\}$ for $x = (x_1, \ldots, x_n) \in \mathcal{B}_{\mathrm{n}}$ and $l = 0, \ldots, n$. Then, for $l = 1, \ldots, n$ and any $x \in \mathcal{B}_{\mathrm{n}}$ such that $\mathbb{P}(A_l^x) > 0$,

$$\begin{aligned}\mathbb{E}[f(X) \,|\, A_l^x] - \mathbb{E}[f(X) \,|\, A_{l-1}^x] &= \mathbb{E}[f(X) \,|\, A_{l-1}^x, X_l = x_l] - \mathbb{E}[f(X) \,|\, A_{l-1}^x] \\ &= \mathbb{P}(X_l \ne x_l \,|\, A_{l-1}^x)\big(\mathbb{E}[f(X) \,|\, A_{l-1}^x, X_l = x_l] - \mathbb{E}[f(X) \,|\, A_{l-1}^x, X_l \ne x_l]\big).\end{aligned}$$

If $\mathbb{P}(X_l \ne x_l \,|\, A_{l-1}^x) \ne 0$, then by the SCP there exists a coupling $(\hat{X}, \hat{Y})$ between the distributions $\mathcal{L}(X \,|\, A_{l-1}^x, X_l = x_l)$ and $\mathcal{L}(X \,|\, A_{l-1}^x, X_l \ne x_l)$ that is supported on the set $\{\,(y, z) \in \mathcal{B}_n^2 :\ d_H((y_i)_{i>l}, (z_i)_{i>l}) \le 1\,\}$. Using this coupling, the Lipschitz property of f, Jensen's inequality and the fact that $\alpha_i \le \alpha_l$ for any $i > l$, we get that

$$\begin{aligned}&\|\mathbb{E}[f(X) \,|\, A_l^x] - \mathbb{E}[f(X) \,|\, A_{l-1}^x]\| \\ &\qquad\le \mathbb{P}(X_l \ne x_l \,|\, A_{l-1}^x)\, \mathbb{E}\, \|f((x_i)_{i\le l}, \hat{X}_{i>l}) - f((x_i)_{i<l}, 1 - x_l, \hat{Y}_{i>l})\| \\ &\qquad\le \mathbb{P}(X_l \ne x_l \,|\, A_{l-1}^x) \cdot 2\alpha_l \le 2\alpha_l,\end{aligned}$$

which is equivalent to (5.2). □

Proof of Lemma 5.2: Note that for $l \le k$, we have $\mathcal{G}_l = \mathcal{F}_l$. As a consequence $N_l = M_l$, where M_l are martingale increments defined in Lemma 5.1, which implies (5.3).

Consider now $l > k$ of the form $l = k + r$ and for $x = (x_1, \ldots, x_k) \in \mathcal{B}_k$ and $v = (v_1, \ldots, v_k) \in (\{0\} \cup \{k+1, \ldots, n\})^k$ set $A_l^{x,v} = \{X_1 = x_1, \ldots, X_k = x_k, Y_1 = v_1, \ldots, Y_r = v_r\}$. Then $\mathcal{F}_l$ is generated by the sets $A_l^{x,v}$. By the definition of the variables Y_r, we have $\{Y_r = i\} \subseteq \{X_i = 1\}$ and so for any x, v such that $\mathbb{P}(A_l^{x,v}) > 0$,

$$\mathbb{E}[f(X) \,|\, A_l^{x,v}] = \frac{\mathbb{E}[f(X)\mathbf{1}_{A_{l-1}^{x,v}} \mathbf{1}_{\{X_{v_r}=1\}} \mathbf{1}_{\{Y_r = v_r\}}]}{\mathbb{P}(A_{l-1}^{x,v}, X_{v_r} = 1, Y_r = v_r)}. \tag{5.7}$$

For $s \in [r]$ let $m_s = |\{i \in [k] \colon x_i = 1\}| + |\{j \in [s-1] \colon v_j \ne 0\}|$ be the number of ones sampled by the time $k + s - 1$. It follows from (5.1) that if $m_s < k$ then $\mathbb{P}(A_{k+s}^{x,v}) > 0$ implies that $v_s \ne 0$ and $\mathbb{P}(Y_s = v_s | X, Y_1, \ldots, Y_{s-1}) = \frac{1}{k - m_s}$ on $A_{k+s-1}^{x,v} \cap \{X_{v_s} = 1\}$, whereas if $m_s = k$, then $\mathbb{P}(A_{k+s}^{x,v}) > 0$ implies that $v_s = 0$ and $\mathbb{P}(Y_s = v_s \,|\, X, Y_1, \ldots, Y_{s-1}) = 1$ on $A_{k+s-1}^{x,v} \cap \{X_{v_s} = 1\} = A_{k+s-1}^{x,v}$. Going back to (5.7) and using this observation for $s = r, \ldots, 1$, we obtain that

$$\mathbb{E}[f(X) \,|\, A_l^{x,v}] = \mathbb{E}[f(X) \,|\, B_l^{x,v}],$$

where $B_l^{x,v} = \{X_1 = x_1, \ldots, X_k = x_k, X_{v_1} = \ldots = X_{v_{l-k}} = 1\}$. We thus obtain

$$\begin{aligned}&\mathbb{E}[f(X) \,|\, A_l^{x,v}] - \mathbb{E}[f(X) \,|\, A_{l-1}^{x,v}] \\ &\qquad= \mathbb{P}(X_{v_r} \ne 1 \,|\, B_{l-1}^{x,v})(\mathbb{E}[f(X) \,|\, B_{l-1}^{x,v}, X_{v_r} = 1] - \mathbb{E}[f(X) \,|\, B_{l-1}^{x,v}, X_{v_r} \ne 1]).\end{aligned}$$

Note that the right-hand side may be non-zero only if $v_r \ne 0$. In this case using the inequality $\alpha_{v_s} \le \alpha_k$ for $s \in [k]$ we can conclude as in the proof of Lemma 5.1. □

We now pass to the proof of Theorem 2.8.

Proof of Theorem 2.8: Let X be a random vector with law π and define the random variables Y_l for $l \le n$ as

$$\begin{aligned} \mathcal{L}(Y_1 \,|\, X) &= \mathrm{Unif}(\operatorname{supp} X) \quad \text{and} \\ \mathcal{L}(Y_l \,|\, X, Y_1, \ldots, Y_{l-1}) &= \mathrm{Unif}(\operatorname{supp} X \setminus \{Y_1, \ldots, Y_{l-1}\}), \quad \text{for} \quad l = 2, \ldots, k, \end{aligned} \tag{5.8}$$

i.e., $Y_l's$ reveal in a uniformly random order the elements of $\operatorname{supp} X$. Let $\mathcal{H}_0 = \{\emptyset, \Omega\}$ and $\mathcal{H}_l = \sigma(Y_1, \ldots, Y_l)$ for $l = 1, \ldots, k$. Then

$$f(X) - \mathbb{E}\, f(X) = \sum_{l=1}^{k} \mathbb{E}[f(X) \,|\, \mathcal{H}_l] - \mathbb{E}[f(X) \,|\, \mathcal{H}_{l-1}] =: \sum_{l=1}^{k} D_l.$$

We will use the matrix version of Freedman's inequality due to Tropp (2011), which asserts (in a version specialized for our application) that if $\|D_l\| \le a$ a.s. for all l, and $\|\sum_{l=1}^{k} \mathbb{E}[D_l^2 \,|\, \mathcal{H}_{l-1}]\| \le \sigma^2$ a.s., then for any $t \ge 0$,

$$\mathbb{P}\left(\lambda_{max}(f(X) - \mathbb{E}\, f(X)) \ge t\right) \le d \exp\Big(-\frac{t^2}{2\sigma^2 + 2at/3}\Big). \tag{5.9}$$

Consider thus a sequence of pairwise distinct $v_1, \ldots, v_k \in [n]$ and denote $A_l^v = \{Y_1 = v_1, \ldots, Y_l = v_l\}$. Similarly, as in the proof of Lemma 5.2, if $\mathbb{P}(A_l^v) > 0$, then we have

$$\mathbb{E}[f(X) \,|\, A_l^v] = \mathbb{E}[f(X) \,|\, B_l^v],$$

where $B_l^v = \{X_{v_1} = \ldots = X_{v_l} = 1\}$. Therefore, we have

$$D_l \mathbf{1}_{A_l^v} = \mathbb{P}(X_{v_l} = 0 \,|\, B_{l-1}^v)\big(\mathbb{E}[f(X) \,|\, B_{l-1}^v, X_{v_l} = 1] - \mathbb{E}[f(X)|B_{l-1}^v, X_{v_l} = 0]\big)\mathbf{1}_{A_l^v}. \tag{5.10}$$

Since the SRP implies the SCP, there exists a coupling $(\tilde{Z}, \hat{Z})$ between the distributions $\mathcal{L}(X \,|\, B_l^v)$ and $\mathcal{L}(X \,|\, B_{l-1}^v, X_{v_l} = 0)$ such that $\tilde{Z}$ and $\hat{Z}$ differ just at coordinate v_l and one additional coordinate (at which by k-homogeneity $\hat{Z}$ necessarily takes the value one). Let $\tilde{Y}_l$ be this coordinate. We have

$$\mathbb{E}[f(X) \,|\, B_{l-1}^v, X_{v_l} = 1] - \mathbb{E}[f(X) \,|\, B_{l-1}^v, X_{v_l} = 0] = \mathbb{E}[f(\tilde{Z}) - f(\hat{Z})], \tag{5.11}$$

Since $\hat{Z}^{\tilde{Y}_l} = \tilde{Z}^{v_l}$, we have

$$\begin{aligned} \Big(\mathbb{E}[f(X) \,|\, B_{l-1}^v, X_{v_l} = 1] - \mathbb{E}[f(X) \,|\, B_{l-1}^v, X_{v_l} = 0]\Big)^2 &= \Big(\mathbb{E}[f(\tilde{Z}) - f(\hat{Z})]\Big)^2 \\ \preccurlyeq \mathbb{E}\Big[\big(f(\tilde{Z}) - f(\hat{Z})\big)^2\Big] &= \mathbb{E}\Big[\big(f(\tilde{Z}) - f(\tilde{Z}^{v_l}) + f(\hat{Z}^{\tilde{Y}_l}) - f(\hat{Z})\big)^2\Big] \\ \preccurlyeq 2\,\mathbb{E}\Big[\big(f(\tilde{Z}) - f(\tilde{Z}^{v_l})\big)^2\Big] + 2\,\mathbb{E}\Big[\big(f(\hat{Z}^{\tilde{Y}_l}) - f(\hat{Z})\big)^2\Big] &\preccurlyeq 2C_{v_l}^2 + 2\,\mathbb{E}\, C_{\tilde{Y}_l}^2, \end{aligned} \tag{5.12}$$

where in the first and second inequality we used the fact that the function $x \mapsto x^2$ is operator convex (i.e., for any $A, B \in \mathcal{H}_{\mathrm{d}}$ and $\lambda \in [0,1]$, $((1-\lambda)A + \lambda B)^2 \preccurlyeq (1-\lambda)A^2 + \lambda B^2$, see Bhatia (1997, Example V.1.3)), and in the last inequality the assumption (2.3).

In particular, using (5.10), we obtain $\|D_l^2\| \le 4 \max_i \|C_i^2\|$, so $\|D_l\| \le 2K$. Moreover, as on A_l^v we have $Y_l = v_l$, by (5.10) and (5.12) we get that

$$D_l^2 \mathbf{1}_{A_l^v} \preccurlyeq 2(C_{Y_l}^2 + \mathbb{E}\, C_{\tilde{Y}_l}^2)\, \mathbb{P}(X_{v_l} = 0 \,|\, B_{l-1}^v)^2 \mathbf{1}_{A_l^v}.$$

Let us now slightly change our notation and think of $\tilde{Y}_l$ as of random variable defined on the same probability space as X, with conditional distribution with respect to the σ-field $\mathcal{H}_l$ given on each of its atoms A_l^v by the above construction, using the corresponding coupling (which depends on $v_1, \ldots, v_l$). Then the above inequality can be written on A_{l-1}^v as

$$D_l^2 \preccurlyeq 2 \sum_{v_l \in [n] \setminus \{v_1, \ldots, v_{l-1}\}} \big(C_{Y_l}^2 + \mathbb{E}[C_{\tilde{Y}_l}^2 \,|\, A_l^v]\big) \, \mathbb{P}(X_{v_l} = 0 \,|\, B_{l-1}^v)^2 \mathbf{1}_{A_l^v}. \tag{5.13}$$

We now go back to the equations (5.10) and (5.11) and let us apply them in the special case of the function $\tilde{f}(x) = \sum_{i=1}^n x_i C_i^2$, denoting the corresponding martingale increment by $\tilde{D}_l$. We obtain that

$$\tilde{D}_l \mathbf{1}_{A_l^v} = \mathbb{P}(X_{v_l} = 0 \,|\, B_{l-1}^v) \big(C_{Y_l}^2 - \mathbb{E}[C_{\tilde{Y}_l}^2 \,|\, A_l^v]\big) \mathbf{1}_{A_l^v}.$$

Thus, we get that

$$0 = \mathbb{E}[\tilde{D}_l \,|\, A_{l-1}^v] = \sum_{v_l \in [n] \setminus \{v_1, \ldots, v_{l-1}\}} \mathbb{E}\Big[\mathbb{P}(X_{v_l} = 0 \,|\, B_{l-1}^v) \mathbf{1}_{A_l^v} \big(C_{Y_l}^2 - \mathbb{E}[C_{\tilde{Y}_l}^2 \,|\, A_l^v]\big) \,\Big|\, A_{l-1}^v \Big],$$

i.e.,

$$\begin{aligned} &\sum_{v_l \in [n] \setminus \{v_1, \ldots, v_{l-1}\}} \mathbb{E}\Big[\mathbb{P}(X_{v_l} = 0 \,|\, B_{l-1}^v) \mathbf{1}_{A_l^v} C_{Y_l}^2 \,\Big|\, A_{l-1}^v \Big] \\ &\qquad = \sum_{v_l \in [n] \setminus \{v_1, \ldots, v_{l-1}\}} \mathbb{E}\Big[\mathbb{P}(X_{v_l} = 0 \,|\, B_{l-1}^v) \mathbf{1}_{A_l^v} \, \mathbb{E}(C_{\tilde{Y}_l}^2 \,|\, A_l^v) \,\Big|\, A_{l-1}^v \Big], \end{aligned}$$

which combined with the estimate (5.13) on D_l^2 (replacing $\mathbb{P}(X_{v_l} = 0 \,|\, B_{l-1}^v)^2$ by $\mathbb{P}(X_{v_l} = 0 \,|\, B_{l-1}^v)$) gives

$$\begin{aligned} \mathbb{E}[D_l^2 \,|\, A_{l-1}^v] &\preccurlyeq 2 \sum_{v_l \in [n] \setminus \{v_1, \ldots, v_{l-1}\}} \mathbb{E}\Big[\big(C_{Y_l}^2 + \mathbb{E}[C_{\tilde{Y}_l}^2 \,|\, A_l^v]\big) \, \mathbb{P}(X_{v_l} = 0 \,|\, B_{l-1}^v) \mathbf{1}_{A_l^v} \,\Big|\, A_{l-1}^v \Big] \\ &= 4 \sum_{v_l \in [n] \setminus \{v_1, \ldots, v_{l-1}\}} \mathbb{E}\Big[C_{Y_l}^2 \, \mathbb{P}(X_{v_l} = 0 \,|\, B_{l-1}^v) \mathbf{1}_{A_l^v} \,\Big|\, A_{l-1}^v \Big] \\ &\preccurlyeq 4 \sum_{v_l \in [n] \setminus \{v_1, \ldots, v_{l-1}\}} C_{v_l}^2 \, \mathbb{P}(A_l^v \,|\, A_{l-1}^v) \\ &= 4 \sum_{v_l \in [n] \setminus \{v_1, \ldots, v_{l-1}\}} C_{v_l}^2 \frac{1}{k - l + 1} \, \mathbb{P}(X_{v_l} = 1 \,|\, B_{l-1}^v) \\ &\preccurlyeq 4 \sum_{v_l \in [n] \setminus \{v_1, \ldots, v_{l-1}\}} C_{v_l}^2 \frac{1}{k - l + 1} \, \mathbb{P}(X_{v_l} = 1), \end{aligned}$$

where in the last inequality we used Kyng and Song (2018, Lemma 1.10), which asserts that $\mathbb{P}(X_{v_l} = 1) \geq \mathbb{P}(X_{v_l} = 1 | B_{l-1}^v)$ (we remark that this is the only place in the proof in which we use the full strength of the strong Rayleigh property).

Extending the summation to $[n]$, we thus obtain

$$\mathbb{E}[D_l^2 \,|\, \mathcal{H}_{l-1}] \preccurlyeq 4 \sum_{v=1}^n C_v^2 \, \mathbb{P}(X_v = 1) \frac{1}{k - l + 1},$$

whence

$$\sum_{l=1}^k \mathbb{E}[D_l^2 | \mathcal{H}_{l-1}] \preccurlyeq 4 \sum_{v=1}^n C_v^2 \, \mathbb{P}(X_v = 1) \log(ek) = 4 \log(ek) \cdot \mathbb{E}\Big[\sum_{v=1}^n X_v C_v^2 \Big].$$

Combining this with the already obtained bound $\|D_l\| \leq 2K$ allows us to apply (5.9) with $a = 2K$ and $\sigma^2 = 4 \| \mathbb{E} \sum_{v=1}^n X_v C_v^2 \| \log(ek)$, which ends the proof of the theorem. $\square$

6. Proofs of the results of Section 4

6.1. *Propositions 4.7 and 4.9.* The main idea behind the proof of Proposition 4.7 is to find an estimate on $\|\Gamma_+(f)\|_\infty$ in terms of α, refining (4.6), and then to use the Herbst argument. We will need the following lemma which we state in the matrix setting as it will be useful for the proof of Proposition 4.9 as well.

Lemma 6.1. *Let $t = (t_1, \ldots, t_n)$ be a sequence of nonnegative numbers and let $D_1, \ldots, D_n \in \mathcal{H}_{\mathrm{d}}$ be positive semidefinite matrices. Then for any $T_1 \geq |t|_1$ and $T_\infty \geq |t|_\infty$*

$$\Big\| \sum_{i=1}^n t_i D_i \Big\| \leq T_\infty \cdot \sup\Big\{ \Big\| \sum_{i\in\mathcal{I}} D_i \Big\| \colon \mathcal{I} \subset [n],\ |\mathcal{I}| \leq \lceil T_1/T_\infty \rceil \Big\}. \tag{6.1}$$

Proof: By homogeneity, we may assume without loss of generality that $T_\infty = 1$. We may also assume that T_1 is a positive integer. Let

$$\mathcal{X} = \Big\{ x \in [0,1]^n \colon \sum_{i=1}^n x_i \leq T_1 \Big\},\ \mathcal{Y} = \Big\{ y \in \{0,1\}^n \colon \sum_{i=1}^n y_i \leq T_1 \Big\}.$$

Since the right-hand side of (6.1) equals to $\max\{\| \sum_{i=1}^n y_i D_i\| \colon y \in \mathcal{Y}\}$, whereas the left-hand side is a convex function of t, the lemma will follow once we prove that $\mathcal{X} \subset \operatorname{conv}\mathcal{Y}$. To this end, by the Krein–Milman theorem, it is enough to show that $\mathcal{Y}$ is the set of all extreme points of the closed convex set $\mathcal{X}$. Consider any $x \in \mathcal{X}\setminus\mathcal{Y}$. Let $i_0 \in [n]$ be such that $x_{i_0} \in (0,1)$. If $\sum_i x_i < T_1$ then for ε sufficiently close to zero, $x + \varepsilon e_{i_0}, x - \varepsilon e_{i_0} \in \mathcal{X}$ and so $x = \frac{1}{2}(x+\varepsilon e_{i_0}) + \frac{1}{2}(x - \varepsilon e_{i_0})$ is not an extreme point of $\mathcal{X}$. If $\sum_i x_i = T_1$, then since T_1 is an integer, there exists $i_1 \neq i_0$ such that $x_{i_1} \in (0,1)$. Then $x = \frac{1}{2}u + \frac{1}{2}v$, where $u = x + \varepsilon e_{i_0} - \varepsilon e_{i_1}$, $v = x - \varepsilon e_{i_0} + \varepsilon e_{i_1}$. For ε close to zero $u, v \in \mathcal{X}$, thus again, x is not an extreme point. □

Proof of Proposition 4.7: We recall that for $x \in \mathcal{B}_{\mathrm{n}}$ and $i, j \in [n]$, x^i and x^{ij} denote the vectors obtained from x by flipping the i-th and swapping the i-th and j-th coordinates respectively. For any $x \in \mathcal{B}_{\mathrm{n}}$, using the definition (4.3) of Γ_+, Lipschitz property of f and inequality $(a+b)^2 \leq 2(a^2+b^2)$ we get

$$\begin{aligned}
\Gamma_+(f)(x) &= \sum_{i=1}^n (f(x) - f(x^i))_+^2 L(x,x^i) + \frac{1}{2}\sum_{i,j=1}^n (f(x) - f(x^{ij}))_+^2 L(x,x^{ij}) \\
&\leq \sum_{i=1}^n \alpha_i^2 L(x,x^i) + \frac{1}{2}\sum_{i,j=1}^n (\alpha_i + \alpha_j)^2 L(x,x^{ij}) \mathbf{1}_{\{x\neq x^{ij}\}} \\
&\leq \sum_{i=1}^n \alpha_i^2 L(x,x^i) + 2\sum_{i=1}^n \alpha_i^2 \sum_{j=1}^n L(x,x^{ij})\mathbf{1}_{\{x\neq x^{ij}\}} \\
&\leq 2\sum_{i=1}^n \alpha_i^2 \sum_{y\colon y_i\neq x_i} L(x,y).
\end{aligned} \tag{6.2}$$

Therefore, by the stability condition (4.7) we estimate $\|\Gamma_+(f)\|_\infty \leq 2R\rho(L)|\alpha|^2$. Herbst's argument (4.5) allows to conclude the first part.

The second part of the proposition follows by observing that for a flip-swap random walk

$$\sum_{i=1}^n \sum_{y\colon y_i \neq x_i} L(x,y) \leq 2\cdot\Delta(L)$$

so by (6.2), Lemma 6.1 applied in the scalar setting $d = 1$ with $t_i = 2\sum_{y:\, y_i \neq x_i} L(x,y)$, $D_i = \alpha_i^2$, $T_1 = 4\Delta(L)$ and $T_\infty = 4R\rho(L)$ we can estimate

$$\|\Gamma_+(f)\|_\infty \leq 4R\rho(L) \sum_{i=1}^{\lceil \Delta(L)/R\rho(L) \rceil} (\alpha_i^{\downarrow})^2$$

and conclude again in virtue of Herbst's argument (4.5). □

The proof of Proposition 4.9 follows along similar lines to the proof of Proposition 4.7, the difference being that in the end, instead of Herbst's argument, we apply the concentration result of Aoun et al. (2020), which asserts that if L satisfies the matrix Poincaré inequality with constant $C_P > 0$

$$\mathrm{Var}(f) \preccurlyeq -C_P \pi(fLf) \quad \forall\, f\colon \mathcal{B}_\mathrm{n} \to \mathcal{H}_\mathrm{d}, \tag{6.3}$$

(where L acts on the matrix-valued function f element-wise and fLf is the matrix product), then it satisfies the exponential concentration bound of the form

$$\pi\big(\lambda_{max}(f - \pi(f)) > t\big) \leq d \exp\Big(\frac{-t^2}{2C_P v_f + t\sqrt{2C_P v_f}}\Big), \tag{6.4}$$

where $v_f = \sup_x \|\Gamma(f)(x)\|$ (where Γ is defined via (4.2), again with matrix multiplication, and $\|\cdot\|$ stands for the operator norm). Note that for $d = 1$, (6.3) is just the usual scalar Poincaré inequality.

Proof of Proposition 4.9: For any $x \in \mathcal{B}_\mathrm{n}$ and $i,j \in [n]$, using operator convexity of the function $x \mapsto x^2$, i.e., the fact that $((1-\lambda)A + \lambda B)^2 \preccurlyeq (1-\lambda)A^2 + \lambda B^2$ for any $A, B \in \mathcal{H}_\mathrm{d}$ and $\lambda \in [0,1]$ (see Bhatia (1997, Example V.1.3)), we get that

$$\begin{aligned}\big(f(x) - f(x^{ij})\big)^2 &= \big[\big(f(x) - f(x^i)\big) + \big(f(x^i) - f(x^{ij})\big)\big]^2 \\ &\preccurlyeq 2\big(f(x) - f(x^i)\big)^2 + 2\big(f(x^i) - f(x^{ij})\big)^2.\end{aligned} \tag{6.5}$$

Therefore, by the definition (4.2) of Γ, by the assumed Lipschitz property (4.8) of f and by (6.5), for any $x \in \mathcal{B}_\mathrm{n}$,

$$\begin{aligned}\Gamma(f)(x) &= \frac{1}{2}\sum_{i=1}^n (f(x) - f(x^i))^2 L(x,x^i) + \frac{1}{4}\sum_{i,j=1}^n (f(x) - f(x^{ij}))^2 L(x,x^{ij}) \\ &\preccurlyeq \frac{1}{2}\sum_{i=1}^n C_i^2 L(x,x^i) + \frac{1}{2}\sum_{i,j=1}^n (C_i^2 + C_j^2) L(x,x^{ij}) \mathbf{1}_{\{x \neq x^{ij}\}} \\ &\preccurlyeq \sum_{i=1}^n C_i^2 \cdot \big[\sum_{y:\, y_i \neq x_i} L(x,y)\big].\end{aligned} \tag{6.6}$$

As both hand sides of (6.6) are positive semidefinite, their norms compare as well. Therefore, as in the proof of Proposition 4.7, by Lemma 6.1 with $t_i = \sum_{y:\, y_i \neq x_i} L(x,y)$, $T_1 = 2\Delta(L)$, $T_\infty = 2R\rho(L)$ and $D_i = C_i^2$

$$\sup_{x \in \mathcal{B}_\mathrm{n}} \|\Gamma(f)(x)\| \leq 2R\rho(L) \cdot \sup\Big\{\big\|\sum_{i \in \mathcal{I}} C_i^2\big\| : \mathcal{I} \subset [n],\, |\mathcal{I}| \leq \lceil \Delta(L)/R\rho(L) \rceil\Big\}.$$

Since L satisfies the (scalar) modified log-Sobolev inequality (4.4), then it satisfies the (scalar) Poincaré inequality with constant $C_P = 2/\rho(L)$ (see, e.g., Bobkov and Tetali (2006, p. 292), noting slightly different definitions of constants in functional inequalities used therein) and whence by Huang and Tropp (2021, Proposition 2.2) or Garg et al. (2021, Theorem 1.1) it satisfies the matrix Poincaré inequality (6.3) with the same constant, which yields the conclusion in virtue of (6.4). □

6.2. *Proposition 4.10.* The proof of Proposition 4.10 is based on the idea introduced in Boucheron et al. (2009) for independent random variables and then developed in Paulin (2014) for Glauber dynamics under the Dobrushin condition. We follow the exposition introduced in the recent works of Sambale and Sinulis (2021, 2022) in the context of sampling without replacement and adapt it to the more abstract setting involving the stability condition given in Definition 4.3.

We start with the following lemmas.

Lemma 6.2. *For any flip-swap random walk with generator L satisfying the stability condition (4.7) and for any $A \subset \mathcal{B}_{\mathrm{n}}$,*

$$\Gamma_+(d_T^2(\cdot, A))(x) \le 8R\rho(L) \cdot d_T^2(x, A). \tag{6.7}$$

Moreover, for any $x, y \in \mathcal{B}_{\mathrm{n}}$ and any set $A \subset \mathcal{B}_{\mathrm{n}}$,

$$d_T^2(x, A) - d_T^2(y, A) \le d_H(x, y). \tag{6.8}$$

Proof: For $x \in \mathcal{B}_{\mathrm{n}}$, $\alpha \in \mathbb{R}^n$ and a probability measure μ on $\mathcal{B}_{\mathrm{n}}$, let $h_x(\mu, \alpha) = \sum_i \alpha_i \mu(z\colon z_i \neq x_i)$. By Sion's minmax theorem, cf. Boucheron et al. (2013, p. 227),

$$d_T(x, A) = \inf_{\mu \in \mathcal{M}(A)} \sup_{\alpha \in B_2^n} h_x(\mu, \alpha), \tag{6.9}$$

where $\mathcal{M}(A)$ is the set of probability measures on A and $B_2^n = \{\, x \in \mathbb{R}^n\colon |x| \le 1 \,\}$ is the unit ball in $\mathbb{R}^n$. Let $\alpha^* \in \mathbb{R}_+^n \cap B_2^n$, $\mu^* \in \mathcal{M}(A)$ be such that $d_T(x, A) = h_x(\mu^*, \alpha^*)$ and set $\nu_y = \operatorname{argmin}_{\nu \in \mathcal{M}(A)} h_y(\nu, \alpha^*)$. Then

$$\begin{aligned}
\Gamma_+\big(d_T(\cdot, A)\big)(x) &= \sum_y \Big[h_x(\mu^*, \alpha^*) - \inf_{\nu \in \mathcal{M}(A)} \sup_{\alpha \in B_2^n} h_y(\nu, \alpha)\Big]_+^2 L(x, y) \\
&\le \sum_y \big[h_x(\mu^*, \alpha^*) - h_y(\nu_y, \alpha^*)\big]_+^2 L(x, y) \\
&\le \sum_y \big[h_x(\nu_y, \alpha^*) - h_y(\nu_y, \alpha^*)\big]_+^2 L(x, y) \\
&= \sum_y \Big[\sum_i \alpha_i^* \big(\nu_y(z\colon z_i \neq x_i) - \nu_y(z\colon z_i \neq y_i)\big)\Big]_+^2 L(x, y) \\
&\le \sum_y \Big[\sum_i \alpha_i^* \mathbf{1}_{\{x_i \neq y_i\}}\Big]^2 L(x, y) \\
&\le 2 \sum_i (\alpha_i^*)^2 \sum_{y\colon y_i \neq x_i} L(x, y) \le 2R\rho(L),
\end{aligned}$$

where the penultimate inequality follows since L is a flip-swap random walk and therefore $L(x, y) > 0$ implies that $d_H(x, y) \le 2$ and so at most two elements of the sum $\sum_i \alpha_i^* \mathbf{1}_{\{x_i \neq y_i\}}$ are non-zero, whence we may apply the inequality $(a + b)^2 \le 2(a^2 + b^2)$. The last inequality is a consequence of the condition $\alpha^* \in B_2^n$ and the stability condition (4.7). We conclude (6.7) using the definition of Γ_+ and estimating $(a - b)_+^2 (a + b)_+^2 \le 4a^2 (a - b)_+^2$.

To show the second part, note that (6.9) together with the Cauchy–Schwarz inequality imply that

$$d_T^2(x, A) = \inf_{\mu \in \mathcal{M}(A)} \sum_i \big(\mu(z\colon z_i \neq x_i)\big)^2 = \sum_i \big(\mu_x^*(z\colon z_i \neq x_i)\big)^2$$

for some $\mu_x^* \in \mathcal{M}(A)$. Therefore, for any $x, y \in \mathcal{B}_{\mathrm{n}}$,

$$\begin{aligned}
d_T^2(x, A) - d_T^2(y, A) &\le \sum_i \Big[\big(\mu_x^*(z\colon z_i \neq x_i)\big)^2 - \big(\mu_x^*(z\colon z_i \neq y_i)\big)^2\Big] \\
&\le \sum_i \mathbf{1}_{\{x_i \neq y_i\}},
\end{aligned}$$

as desired. □

Using the inequality $1 - e^{-z} \le z$ we observe that for any $f\colon \mathcal{B}_{\mathrm{n}} \to \mathbb{R}$,

$$\begin{aligned}\mathcal{E}(e^f, f) &= \sum_x \pi(x) e^{f(x)} \Big[\sum_y (f(x) - f(y))_+ (1 - e^{f(y)-f(x)}) L(x,y) \Big] \\ &\le \pi\big(e^f \Gamma_+(f)\big).\end{aligned}$$

Therefore, the modified log-Sobolev inequality (4.4) implies the following inequality stated in Bobkov and Götze (1999):

$$\rho(L)\,\mathrm{Ent}_\pi(e^f) \le \pi\big(e^f \tilde{\Gamma}(f)^2\big) \tag{6.10}$$

with operator $\tilde{\Gamma}(f) = \sqrt{\Gamma_+(f)}$ (note that in Bobkov and Götze (1999) $\tilde{\Gamma}$ is denoted by Γ, we use $\tilde{\Gamma}$ to avoid a conflict of notation). As a consequence, the hypothesis of Bobkov and Götze (1999, Theorem 2.1) (formula (1.1)) therein holds under the assumption of the modified log-Sobolev inequality (4.4) (with $c = 2/\rho(L)$). As a result, the following lemma follows directly by the derivation of Bobkov and Götze (1999, equation (2.4)) with a slight adjustment of constants (see also Aida et al. (1994)).

Lemma 6.3. *If a measure π on $\mathcal{B}_{\mathrm{n}}$ satisfies the modified log-Sobolev inequality (4.4) and $f\colon \mathcal{B}_{\mathrm{n}} \to [0,\infty)$ is such that $\Gamma_+(f) \le Cf$ for some constant $C > 0$, then for all $t > C/\rho(L)$,*

$$\pi\big(\exp(f/t)\big) \le \exp\Big(\frac{\pi(f)}{t - C/\rho(L)}\Big). \tag{6.11}$$

We are finally in position to prove Proposition 4.10.

Proof of Proposition 4.10: To lighten notation, denote $f(x) = d_T^2(x, A)$ for $x \in \mathcal{B}_{\mathrm{n}}$ and some fixed set $A \subset \mathcal{B}_{\mathrm{n}}$. Denote also $h(z) = (e^z - 1)/z$ for $z \in [0, \infty)$ and $Df_y(x) = f(x) - f(y)$ for $x, y \in \mathcal{B}_{\mathrm{n}}$, and note that h is an increasing function. Starting with the modified log-Sobolev inequality (4.4) we have for all $\lambda > 0$,

$$\begin{aligned}\mathrm{Ent}_\pi(e^{-\lambda f}) &\le \lambda/\rho(L) \cdot \mathcal{E}(e^{-\lambda f}, -f) \\ &= \lambda/\rho(L) \sum_{x,y} \big(Df_y(x)\big)_+ \big(e^{-\lambda f(y)} - e^{-\lambda f(x)}\big) L(x,y)\pi(x) && \text{(by reversibility of } L) \\ &= \lambda^2/\rho(L) \sum_x \pi(x) e^{-\lambda f(x)} \Big[\sum_y \big(Df_y(x)\big)_+^2 h\big(\lambda Df_y(x)\big) L(x,y) \Big] \\ &\le \lambda^2 h(2\lambda)/\rho(L) \cdot \pi\big(e^{-\lambda f}\Gamma_+(f)\big) && \text{(by (6.8))} \\ &\le 8R\lambda^2 h(2\lambda) \cdot \pi\big(e^{-\lambda f} f\big) && \text{(by (6.7))} \\ &\le 8R\lambda^2 h(2\lambda) \cdot \pi(e^{-\lambda f})\pi(f),\end{aligned}$$

where the last inequality follows from non-positive correlation between the functions f and $e^{-\lambda f}$. Therefore, using the entropy method (cf., e.g., Boucheron et al. (2013, Chapter 6)) and monotonicity of h, we have for every $\lambda > 0$,

$$\begin{aligned}\pi\big(\exp(\lambda(\pi(f) - f)\big) &= \exp\Big(\lambda \int_0^\lambda \frac{d}{ds}\Big[\frac{1}{s}\log \pi(e^{-sf})\Big]\, ds\Big) \\ &= \exp\Big(\lambda \int_0^\lambda \frac{\mathrm{Ent}_\pi(e^{-sf})}{s^2 \pi(e^{-sf})}\, ds\Big) \\ &\le \exp\Big(\lambda \cdot 8R\pi(f) \int_0^\lambda h(2s)\, ds\Big) \le \exp\big(4R\lambda(e^{2\lambda} - 1)\pi(f)\big).\end{aligned}$$

By Chebyshev's exponential inequality

$$\pi(A) = \pi\big(\pi(f) - f \ge \pi(f)\big) \le \exp\Big(\lambda\big(4R(e^{2\lambda} - 1) - 1\big)\pi(f)\Big).$$

Taking $\lambda = \frac{1}{2}\log(1+\frac{1}{8R})$ and estimating $\log(1+x) \geq x/(x+1)$ for $x \geq 0$ gives

$$\pi(A) \leq \exp\Big(-\frac{1}{4}\log\big(1+\frac{1}{8R}\big)\pi(f)\Big) \leq \exp\Big(-\frac{\pi(f)}{32R+4}\Big). \tag{6.12}$$

We conclude by dividing (6.12) by its right hand side and using Lemma 6.3 with $t = 4+40R$ and $C = 8R\rho(L)$ (in virtue of Lemma 6.2). □

6.3. *Proposition 4.11.* Before we move to the proof of Proposition 4.11, we comment a bit on a background result. Using the equivalence between the modified log-Sobolev inequality (4.4) and the family of Beckner inequalities together with the approach developed in Boucheron et al. (2005), it was shown in Adamczak et al. (2022, Proposition 3.1) that the following moment estimate is implied by the modified log-Sobolev inequality. Below we will denote by $\|\cdot\|_p$ the norm in $L^p(\pi)$.

Proposition 6.4. *If a probability measure π on $\mathcal{B}_{\mathrm{n}}$ satisfies the modified log-Sobolev inequality (4.4), then for any $p \geq 2$,*

$$\|(f-\pi(f))_+\|_p \leq C\sqrt{p/\rho(L)}\|\sqrt{\Gamma_+(f)}\|_p, \tag{6.13}$$

where $C = \sqrt{3\sqrt{e}/(\sqrt{e}-1)}$.

A general method of deriving estimates for polynomials from moment inequalities of the form (6.13) has been presented in Adamczak and Wolff (2015) in the continuous case, and in Adamczak et al. (2019, 2022) in the context of Glauber dynamics. To obtain results for flip-swap random walks we will adapt a version of this method introduced recently by Sambale and Sinulis (2022) for multislices.

Proof of Proposition 4.11: Below we write C to denote universal constants and C_a to denote constants depending only on the parameter a. In both cases the constants may change values between occurrences. Let $f\colon \mathcal{B}_{\mathrm{n}} \to \mathbb{R}$ be a tetrahedral polynomial. By ∂_i we denote the partial derivative with respect to the i-th coordinate. If $x, y \in \mathcal{B}_{\mathrm{n}}$ differ at the i-th coordinate only, then by the fact that f is linear in each coordinate

$$|f(x)-f(y)| = \big|\partial_i f(x)\big|.$$

Similarly, if x and y differ only by a swap of the i-th and j-th coordinate, we have

$$\begin{aligned}|f(x)-f(y)| &= |\partial_i f(x)(y_i-x_i) + \partial_j f(x)(y_j-x_j) + \partial_i\partial_j f(x)(y_i-x_i)(y_j-x_j)| \\ &\leq |\partial_i f(x)| + |\partial_j f(x)| + |\partial_i\partial_j f(x)|.\end{aligned}$$

Thus,

$$\begin{aligned}\Gamma(f)(x) &= \frac{1}{2}\sum_{i=1}^{n}(f(x)-f(x^i))^2 L(x,x^i) + \frac{1}{2}\sum_{1\leq i<j\leq n}(f(x)-f(x^{ij}))^2 L(x,x^{ij}) \\ &\leq \frac{1}{2}\sum_{i=1}^{n}|\partial_i f(x)|^2 L(x,x^i) + \frac{3}{2}\sum_{\substack{1\leq i<j\leq n \\ x^{ij}\neq x}}(|\partial_i f(x)|^2 + |\partial_j f(x)|^2 + |\partial_i\partial_j f(x)|^2)L(x,x^{ij}) \\ &\leq R\rho(L)\Big(3.5\sum_{i=1}^{n}|\partial_i f(x)|^2 + 0.75\sum_{i,j=1}^{n}|\partial_i\partial_j f(x)|^2\Big),\end{aligned}$$

where in the last inequality we used the stability condition (4.7). Note that since f is tetrahedral, $\partial_i\partial_i f(x) = 0$ for all i.

Combining the above equality with Proposition 6.4 we obtain that for every tetrahedral polynomial $f\colon \mathcal{B}_{\mathrm{n}} \to \mathbb{R}$

$$\|f - \pi(f)\|_p \le C\sqrt{p}\sqrt{R}\Big(\big\||\nabla f|\big\|_p + \big\|\|\nabla^2 f\|_{HS}\big\|_p\Big), \tag{6.14}$$

where C is a universal constant.

In the subsequent part of the proof we are going to need some auxiliary notation. For d-tensors $A = (a_{\mathbf{i}})_{\mathbf{i}\in[n]^d}$, $B = (b_{\mathbf{i}})_{\mathbf{i}\in[n]^d}$ define

$$\langle A, B\rangle = \sum_{\mathbf{i}\in[n]^d} a_{\mathbf{i}} b_{\mathbf{i}}.$$

Let us now consider a family of stochastically independent random tensors $\{G^I\colon I \subseteq \mathbb{N}, |I| \in \{1,2\}\}$, given by $G^{\{m\}} = (g_i^{\{m\}})_{i\in[n]}$, $G^{\{l,k\}} = (g_{i,j}^{\{l,k\}})_{i,j\in[n]}$, with coefficients being i.i.d. standard Gaussian variables. Denote by $P_{d,\le 2}$ the family of all partitions of the set $[d]$ into non-empty subsets of cardinality at most 2. Finally, for any positive integers d and l and $\mathcal{J} = \{J_1,\dots,J_l\} \in P_{d,\le 2}$ define the random d-tensor $G_{\mathcal{J}} = (\prod_{j=1}^l g_{\mathbf{i}_{J_j}}^{J_j})_{\mathbf{i}\in[n]^d}$. For instance $G_{\{\{1,3\},\{2\}\}} = (g_{i_1 i_3}^{\{1,3\}} g_{i_2}^{\{2\}})_{i_1,i_2,i_3\in[n]}$.

Using the fact that the p-th moment of a mean zero Gaussian variable with variance σ^2 is for $p \ge 2$ comparable to $\sqrt{p}\sigma$ up to universal constants, we can rewrite (6.14) as

$$\|f(X) - \mathbb{E}\, f(X)\|_p \le C\sqrt{R}\Big(\|\langle \nabla f(X), G^{\{1\}}\rangle\|_p + \|\langle \nabla^2 f(X), G^{\{1,2\}}\rangle\|_p\Big), \tag{6.15}$$

where X is a random vector with law π, independent of the family $\{G^I\}$.

The inequality (6.15) constitutes a basis for the induction argument leading to the following inequality valid for any $f\colon \mathcal{B}_{\mathrm{n}} \to \mathbb{R}$, $d \ge 1$ and $p \ge 2$,

$$\begin{aligned}\|f(X) - \mathbb{E}\, f(X)\|_p \le & C_d\Big(\sum_{l=d}^{2d}\sum_{\mathcal{J}\in P_{l,\le 2}} R^{|\mathcal{J}|/2}\|\langle \nabla^l f(X), G_{\mathcal{J}}\rangle\|_p \\ & + \sum_{l=1}^{2d-2}\sum_{\mathcal{J}\in P_{l,\le 2}} R^{|\mathcal{J}|/2}\|\langle \mathbb{E}_X \nabla^l f(X), G_{\mathcal{J}}\rangle\|_p\Big).\end{aligned} \tag{6.16}$$

Before we prove the above estimate, let us show how it implies the statement of the proposition. If f is a tetrahedral polynomial of degree d, then $\nabla^l f = 0$ for $l > d$, moreover $\nabla^d f$ is constant and so $\nabla^d f(X) = \mathbb{E}\,\nabla^d f(X)$. Thus, (6.16) reduces to

$$\|f(X) - \mathbb{E}\, f(X)\|_p \le C_d \sum_{l=1}^{d}\sum_{\mathcal{J}\in P_{l,\le 2}} R^{|\mathcal{J}|/2}\|\langle \mathbb{E}_X \nabla^l f(X), G_{\mathcal{J}}\rangle\|_p.$$

We can now use moment estimates for tetrahedral homogeneous polynomials in i.i.d. standard Gaussian variables due to Latała (2006), which assert that for any l-tensor $A = (a_{\mathbf{i}})_{\mathbf{i}\in[n]^l}$ and $p \ge 2$,

$$\|\langle A, G_{\{\{1\},\dots,\{l\}\}}\rangle\|_p \le C_l \sum_{\mathcal{J}\in P_l} p^{|\mathcal{J}|/2}\|A\|_{\mathcal{J}}.$$

Applying this inequality to $\langle \mathbb{E}_X \nabla^l f(X), G_{\mathcal{J}}\rangle$ (we treat here $\mathbb{E}_X \nabla^l f(X)$ as a $|\mathcal{J}|$-tensor by merging the indices according to the partition $\mathcal{J}$), we obtain

$$\|f(X) - \mathbb{E}\, f(X)\|_p \le C_d \sum_{l=1}^{d}\sum_{\mathcal{J}\in P_{l,\le 2}} R^{|\mathcal{J}|/2}\sum_{\mathcal{I}\in P_l\colon \mathcal{I}\succ\mathcal{J}} p^{|\mathcal{I}|/2}\|\mathbb{E}\,\nabla^l f(X)\|_{\mathcal{I}},$$

where $\mathcal{I} \succ \mathcal{J}$ if every element of $\mathcal{I}$ is a union of certain elements of $\mathcal{J}$. Rearranging the terms and taking into account that in a non-trivial case R is bounded away from zero by an absolute constant (see Remark 4.4), which gives $R^{|\mathcal{J}|/2} \leq C_d R^{l/2}$ for $\mathcal{J} \in P_{l,\leq 2}$, we get

$$\|f(X) - \mathbb{E} f(X)\|_p \leq C_d \sum_{l=1}^{d} \sum_{\mathcal{I} \in P_l} R^{l/2} p^{|\mathcal{I}|/2} \| \mathbb{E} \nabla^l f(X)\|_{\mathcal{I}}$$

for $p \geq 2$. This implies the tail inequality of the proposition in the standard way by the use of Chebyshev's inequality $\mathbb{P}(|f(X) - \mathbb{E} f(X)| \geq e\|f(X) - \mathbb{E} f(X)\|_p) \leq e^{-p}$ followed by an appropriate change of variables and adjustment of constants. We leave the details to the reader and turn to the proof of (6.16).

We will proceed by induction on d. For $d = 1$, using the definitions of $G_{\{1\}}$ and $G_{\{\{1,2\}\}}$ one can easily see that (6.16) reads as

$$\begin{aligned}\|f(X) - \mathbb{E} f(X)\|_p \leq & C\Big(\sqrt{R}\|\langle \nabla f(X), G^{\{1\}}\rangle\|_p + \sqrt{R}\|\langle \nabla^2 f(X), G^{\{1,2\}}\rangle\|_p \\ & + R\|\langle \nabla^2 f(X), G_{\{\{1\},\{2\}\}}\rangle\|_p\Big),\end{aligned}$$

which is clearly weaker than (6.15). Let us thus assume that the inequality holds for all positive integers smaller than d. Applying the inequality with $d-1$ and combining it with the triangle inequality in L_p we get (recall that the value of C_d may change between occurrences)

$$\begin{aligned}\|f(X) - \mathbb{E} f(X)\|_p \leq & C_d\Big(\sum_{l=d-1}^{2d-2} \sum_{\mathcal{J} \in P_{l,\leq 2}} R^{|\mathcal{J}|/2}\|\langle \nabla^l f(X), G_{\mathcal{J}}\rangle\|_p \\ & + \sum_{l=1}^{2d-4} \sum_{\mathcal{J} \in P_{l,\leq 2}} R^{|\mathcal{J}|/2}\|\langle \mathbb{E}_X \nabla^l f(X), G_{\mathcal{J}}\rangle\|_p\Big) \\ \leq & C_d\Big(\sum_{l=d-1}^{2d-2} \sum_{\mathcal{J} \in P_{l,\leq 2}} R^{|\mathcal{J}|/2}\|\langle \nabla^l f(X), G_{\mathcal{J}}\rangle - \langle \mathbb{E}_X \nabla^l f(X), G_{\mathcal{J}}\rangle\|_p \\ & + \sum_{l=1}^{2d-2} \sum_{\mathcal{J} \in P_{l,\leq 2}} 2R^{|\mathcal{J}|/2}\|\langle \mathbb{E}_X \nabla^l f(X), G_{\mathcal{J}}\rangle\|_p\Big).\end{aligned} \tag{6.17}$$

An application of inequality (6.15) conditionally on $G_{\mathcal{J}}$ to the functions $h_{l,\mathcal{J}}(x) = \langle \nabla^l f(x), G_{\mathcal{J}}\rangle$ for $l = d-1, \ldots, 2d-2$ and $\mathcal{J} \in P_{l,\leq 2}$ (note that $h_{l,\mathcal{J}}$'s are tetrahedral polynomials), followed by the Fubini theorem, gives

$$\begin{aligned}\|\langle \nabla^l f(X), G_{\mathcal{J}}\rangle - \langle \mathbb{E}_X \nabla^l f(X), G_{\mathcal{J}}\rangle\|_p \leq C\sqrt{R}\Big(&\|\langle \nabla^{l+1} f(X), G_{\mathcal{J} \cup \{\{l+1\}\}}\rangle\|_p \\ & + \|\langle \nabla^{l+2} f(X), G_{\mathcal{J} \cup \{\{l+1,l+2\}\}}\rangle\|_p\Big),\end{aligned}$$

which combined with (6.17) concludes the induction step, thus proving (6.16). □

7. Proofs of the results of Section 3

By virtue of the abstract results of Section 4, all the results of Section 3 will follow if one proves that there exists a flip-swap random walk on $\mathcal{B}_n$ with stationary measure $\pi = \pi(p, k)$ which satisfies the stability condition (4.7) with constant $R = 2$ for all $p \in (0, 1)^n$ and $k = 0, \ldots, n$ (cf. Theorem 7.3). The rest of this section is devoted to proving this theorem.

Before we proceed with the proof, let us present its outline. Our approach to defining an R-stable generator L_π will be based on the inductive construction of Hermon and Salez (2023). The construction is quite abstract and at each induction step it uses the coupling resulting from the definition of the stochastic covering property. For obtaining the modified log-Sobolev inequality sufficient to investigate the speed of convergence of the Markov chain or concentration inequality as in (1.1), the form of the coupling is not relevant, as long as it satisfies the SCP. In turn, in order to establish the stability condition, one needs to control additional properties of the couplings used at various steps of the construction. The main technical challenge is to choose them in an appropriate, *balanced* way. For conditioned Bernoulli distributions it is obtained by an explicit construction of the coupling, given in the following lemma, the proof of which is postponed until the end of this section.

Lemma 7.1. *For every $n \in \mathbb{N}$, $p \in (0,1)^n$ and $k \in [n]$, there exists a coupling (Z, Z') of measures $\pi(p,k)$ and $\pi(p,k-1)$ such that for all $x \in \operatorname{supp} \pi(p,k-1)$, and $r \in [n]$ such that $x_r = 0$,*

$$\mathbb{P}(Z = x + e_r \,|\, Z' = x) = \mathbb{E}\Big[\frac{\mathbf{1}_{\{Z_r=1\}}}{\sum_{l=1}^n \mathbf{1}_{\{Z_l=1\}}\mathbf{1}_{\{x_l=0\}}}\Big] \tag{7.1}$$

and for all $x \in \operatorname{supp}\pi(p,k)$, and $r \in [n]$ such that $x_r = 1$,

$$\mathbb{P}(Z' = x - e_r \,|\, Z = x) = \mathbb{E}\Big[\frac{\mathbf{1}_{\{Z'_r=0\}}}{\sum_{l=1}^n \mathbf{1}_{\{Z'_l=0\}}\mathbf{1}_{\{x_l=1\}}}\Big]. \tag{7.2}$$

Let us now recall the inductive construction of Hermon and Salez (2023) in the k-homogeneous case. It works for any k-homogeneous probability measure π on $\mathcal{B}_{\mathrm{n}}$, satisfying the SCP and produces a generator of a π-reversible flip-swap random walk Q^* such that $\rho(Q^*) \geq 1$ and $\Delta(Q^*) \leq 2k$.

To simplify the notation, we are going to treat vectors $x_{\neq l}$ for $x \in \mathcal{B}_{\mathrm{n}}$ and $l \in [n]$ sometimes as elements of $\{0,1\}^{[n]\setminus\{l\}}$ (this is how they were defined at the beginning of Section 2) and sometimes as elements of $\mathcal{B}_{n-1}$ (with the natural identification, i.e., preserving the order of coordinates). The exact meaning will be clear from the context. The same convention will apply to random vectors, e.g., to $X_{\neq l}$.

In the case $n = 1$, we let Q be the zero matrix, which restricted to the support of π gives the trivial generator on the one-point space. Clearly then $\rho(Q) = \infty$ and $\Delta(Q) = 0$.

For $n > 1$, $l \in [n]$ and $x, y \in \operatorname{supp}\pi$, $x \neq y$, we set

$$Q^{(l)}(x,y) = \begin{cases} \mathbb{P}\left(U = y_{\neq l} \,|\, V = x_{\neq l}\right)\mathbb{P}\left(X_l \neq x_l\right) & \text{if} \quad x_l \neq y_l, \\ Q^{(l)}_{x_l}(x_{\neq l}, y_{\neq l}) & \text{else,} \end{cases} \tag{7.3}$$

where X is a random vector with law π and (U,V) is any coupling between $\mathcal{L}(X_{\neq l} \,|\, X_l = y_l)$ and $\mathcal{L}(X_{\neq l} \,|\, X_l = x_l)$ given by the SCP and $Q^{(l)}_{x_l}$ is any flip-swap generator on $\mathcal{B}_{n-1}$ with stationary distribution $\mathcal{L}(X_{\neq l} \,|\, X_l = x_l)$ such that $\rho(Q^{(l)}_{x_l}) \geq 1$ and $\Delta(Q^{(l)}_{x_l}) \leq 2(k - x_i)$, the existence of which is provided by the induction scheme. We define the diagonal elements of $Q^{(l)}$ so that the row sums vanish. Finally, put

$$Q^* = \frac{1}{n}\sum_{l=1}^n Q^{(l)}. \tag{7.4}$$

Then by (the proof of) Hermon and Salez (2023, Theorem 2), Q is π-reversible and we have $\rho(Q^*) \geq 1$, $\Delta(Q^*) \leq 2k$.

Now we are in position to construct the generator L_π. Let $X \sim \pi = \pi(p,k)$ for some $p \in (0,1)^n$ and $k \in \{0,\dots,n\}$. Observe that for any $y_l \in \{0,1\}$, we have $\mathcal{L}(X_{\neq l} \,|\, X_l = y_l) = \pi(p_{\neq l}, k - y_l)$, in particular in the above recursive construction we can restrict our attention to the class of conditional Bernoulli distributions and use as $Q^{(l)}_{x_l}$ the generators defined for such measures in dimension $n-1$. Moreover, for (U,V) we can take the coupling (Z,Z') (if $y_l = 0$) or (Z',Z) (if $y_l = 1$) given by Lemma 7.1 applied in dimension $n-1$ with $p_{\neq l}$ instead of p (note that since the right-hand side of (7.1) summed over r such that $x_r = 0$ gives one, we indeed have $Z \triangleright Z'$, which makes this coupling a legitimate choice in the Hermon–Salez construction). Let us define L_π as the outcome of the Hermon–Salez construction with the above choices of $Q^{(l)}_{x_l}$ and (U, V). Thus, formally for $n = 1$ we let L_π be the trivial generator and for $n > 1$ and $l \in [n]$ we set

$$L_\pi = \frac{1}{n}\sum_{l=1}^{n} L^{(l)} \tag{7.5}$$

with

$$L^{(l)}(x,y) = \begin{cases} \mathbb{P}\left(U = y_{\neq l} \,|\, V = x_{\neq l}\right)\mathbb{P}\left(X_l \neq x_l\right) & \text{if} \quad x_l \neq y_l, \\ L_{\pi_l}(x_{\neq l}, y_{\neq l}) & \text{else,} \end{cases} \tag{7.6}$$

for $x \neq y$, where (U, V) is the coupling of $\pi(p_{\neq l}, k - y_l)$ and $\pi(p_{\neq l}, k - x_l)$ given by Lemma 7.1, and $\pi_l = \pi(p_{\neq l}, k - y_l)$ (again the diagonal elements are adjusted so that the row sums vanish).

Then, the results by Hermon and Salez, specialized to L_π give

Proposition 7.2. *The generator L_π constructed according to (7.5) generates a reversible flip-swap random walk with stationary measure π such that $\rho(L_\pi) \geq 1$ and $\Delta(L_\pi) \leq 2k$.*

Our main result concerning conditional Bernoulli distributions, underlying all the results from Section 3 is

Theorem 7.3. *The generator L_π constructed according to (7.5) with stationary measure π satisfies the stability condition (4.7) with $R = 2$.*

Proof of Theorem 7.3: We proceed by induction in the dimension n.

For $n = 1$ the only possibilities are $k = 0$ and $k = 1$ and in both cases the left-hand side of (4.7) vanishes. Thus, the stability condition (4.7) is satisfied with any nonnegative R.

Assume the induction hypothesis holds for $n - 1$ and fix $x \in \operatorname{supp}\pi$ and $i \in [n]$. We may and do assume that $k \in \{1, \ldots, n-1\}$ as otherwise L_π trivializes.

Since $\rho(L_\pi) \geq 1$, it is enough to show that

$$\max_{x\in\operatorname{supp}\pi;\, i\in[n]} \sum_{y\colon y_i\neq x_i} L_\pi(x,y) \leq 2. \tag{7.7}$$

As in the definition of L_π we will denote by X a random variable with distribution π.

If $x_i = 0$, then by (7.5),

$$\begin{aligned} \sum_{y\colon y_i\neq x_i} L_\pi(x,y) &= \sum_{j\colon x_j=1} \frac{1}{n}\sum_{l=1}^{n} L^{(l)}(x, x^{ij}) \\ &= \frac{1}{n}\sum_{j\colon x_j=1}\sum_{l\in[n]\setminus\{i,j\}} L^{(l)}(x,x^{ij}) + \frac{1}{n}\sum_{j\colon x_j=1} L^{(i)}(x,x^{ij}) \\ &\qquad + \frac{1}{n}\sum_{j\colon x_j=1} L^{(j)}(x,x^{ij}), \end{aligned} \tag{7.8}$$

where we recall that $x^{ij} = x + e_i - e_j$. We estimate each term on the right hand side separately.

For $l \in [n]$ let ζ_l be the unique increasing bijection between $[n] \setminus \{l\}$ and $[n-1]$. If $l \neq i, j$, then for $y = x^{ij}$ we have $y_l = x_l$ and so, by (7.6), $L^{(l)}(x, y) = L_{\pi_l}(x_{\neq l}, y_{\neq l})$, where $\pi_l = \pi(p_{\neq l}, k - x_l)$.

Thus, denoting $r_l = \zeta_l(i)$, we get

$$\begin{aligned}\frac{1}{n}\sum_{j\colon x_j=1}\sum_{l\in[n]\setminus\{i,j\}} L^{(l)}(x,x^{ij}) &= \frac{1}{n}\sum_{l\in[n]\setminus\{i\}}\sum_{j\neq l\colon x_j=1} L_{\pi_l}(x_{\neq l},(x^{ij})_{\neq l})\\ &= \frac{1}{n}\sum_{l\in[n]\setminus\{i\}}\sum_{y\in\mathcal{B}_{n-1}\colon y_{r_l}\neq (x_{\neq l})_{r_l}} L_{\pi_l}(x_{\neq l},y) \leq \frac{n-1}{n}\cdot 2, \end{aligned}\tag{7.9}$$

where the last inequality follows from the induction assumption applied to π_l.

The second term of (7.8) is estimated again using the definition (7.6). Indeed, if $x_j = 1$, then for $y = x^{ij}$ we have $x_i \neq y_i$. Thus, recalling that (U,V) is a coupling between the laws $\mathcal{L}(X_{\neq i}\,|\,X_i=1)$ and $\mathcal{L}(X_{\neq i}\,|\,X_i=0)$ such that $V \rhd U$, we obtain

$$\begin{aligned}\frac{1}{n}\sum_{j\colon x_j=1} L^{(i)}(x,x^{ij}) &= \frac{1}{n}\sum_{j\colon x_j=1}\mathbb{P}\left(U=(x^{ij})_{\neq i}\,|\,V=x_{\neq i}\right)\mathbb{P}\left(X_i\neq x_i\right)\\ &= \frac{1}{n}\,\mathbb{P}\left(X_i=1\right)\leq\frac{1}{n}.\end{aligned}\tag{7.10}$$

Let us pass to the last term of (7.8). We stress that this is the crucial part of the proof, the only one in which we use the specific form of the coupling (U,V) used in the construction of L_π.

To estimate this last term we use (7.1) from Lemma 7.1 combined with the fact that if $x_i = 0$ and $x_j = 1$, then for $y = x^{ij}$, $y_j = 0 \neq x_j$ and so (U,V) from (7.6) is the coupling between the laws $\pi(p_{\neq j},k)$ and $\pi(p_{\neq j},k-1)$ given by Lemma 7.1 (in dimension $n-1$). For $j\in[n]$ consider a $\mathcal{B}_{n-1}$-valued random vector $Z^{(j)}\sim\mathcal{L}(X_{\neq j}\,|\,X_j=0)=\pi(p_{\neq j},k)$. Note also that since X, x have the same number of ones, we have

$$\sum_{l=1}^{n}\mathbf{1}_{\{X_l=0\}}\mathbf{1}_{\{x_l=1\}}=\sum_{l=1}^{n}\mathbf{1}_{\{X_l=1\}}\mathbf{1}_{\{x_l=0\}}.\tag{7.11}$$

Putting all the above observations together and using Lemma 7.1 together with (7.6) in the first step, we obtain

$$\begin{aligned}\frac{1}{n}\sum_{j\colon x_j=1}L^{(j)}(x,x^{ij}) &= \frac{1}{n}\sum_{j\colon x_j=1}\mathbb{E}\,\Big[\frac{\mathbf{1}_{\{Z_i^{(j)}=1\}}}{\sum_{l\neq j}\mathbf{1}_{\{Z_l^{(j)}=1\}}\mathbf{1}_{\{x_l=0\}}}\Big]\,\mathbb{P}\left(X_j=0\right)\\ &= \frac{1}{n}\sum_{j\colon x_j=1}\mathbb{E}\,\Big[\mathbf{1}_{\{X_i=1\}}\frac{\mathbf{1}_{\{X_j=0\}}}{\sum_{l\neq j}\mathbf{1}_{\{X_l=1\}}\mathbf{1}_{\{x_l=0\}}}\Big]\\ &\overset{x_j=1}{=} \frac{1}{n}\sum_{j\colon x_j=1}\mathbb{E}\,\Big[\mathbf{1}_{\{X_i=1\}}\frac{\mathbf{1}_{\{X_j=0\}}}{\sum_{l}\mathbf{1}_{\{X_l=1\}}\mathbf{1}_{\{x_l=0\}}}\Big]\\ &\overset{(7.11)}{=} \frac{1}{n}\sum_{j\colon x_j=1}\mathbb{E}\,\Big[\mathbf{1}_{\{X_i=1\}}\frac{\mathbf{1}_{\{X_j=0\}}}{\sum_{l}\mathbf{1}_{\{X_l=0\}}\mathbf{1}_{\{x_l=1\}}}\Big]\\ &= \frac{1}{n}\,\mathbb{P}\left(X_i=1\right)\leq\frac{1}{n}.\end{aligned}\tag{7.12}$$

Combining the estimates (7.9), (7.10) and (7.12) with (7.8) yields (7.7) and thus the stability condition (4.7) with $R = 2$ in the case $x_i = 0$. The case $x_i = 1$ is analogous, the main difference being that in (7.12) we use (7.2) in place of (7.1) from Lemma 7.1.

Together the two cases give the induction step and conclude the proof of the theorem. □

We conclude this section with the proof of Lemma 7.1.

Proof of Lemma 7.1: For $x \in \mathcal{B}_{\mathrm{n}}$, let $\kappa(x) = \sum_i x_i$ and let B be a vector of independent Bernoulli random variables with probabilities of success given by $p = (p_1, \dots, p_n)$. Consider three $\mathcal{B}_{\mathrm{n}}$-valued random variables: $\widehat{Z} \sim \mathcal{L}(B \,|\, \kappa(B) = k)$, $Z' \sim \mathcal{L}(B \,|\, \kappa(B) = k-1)$ and Z such that for all $x, y \in \mathcal{B}_{\mathrm{n}}$,

$$\mathbb{P}(Z = y | Z' = x) = h(y, x), \tag{7.13}$$

where

$$h(y,x) = \mathbb{E}\Big[\frac{\mathbf{1}_{\{\widehat{Z}_r=1\}}}{\sum_l \mathbf{1}_{\{\widehat{Z}_l=1\}}\mathbf{1}_{\{x_l=0\}}}\Big]$$

if $y = x + e_r$ for some $r \in [n]$ and $\kappa(x) = k-1$, and $h(y,x) = 0$ otherwise. Note that for $x \in \mathcal{B}_{\mathrm{n}}$ such that $\kappa(x) = k-1$, $\sum_l \mathbf{1}_{\{\widehat{Z}_l=1\}}\mathbf{1}_{\{x_l=0\}} > 0$ with probability one, so $h(y,x)$ is well-defined. Moreover, for such x,

$$\sum_{y \in \mathcal{B}_{\mathrm{n}}} h(y,x) = \sum_{r\colon x_r=0} h(x+e_r, x) = 1,$$

which guarantees the existence of the couple (Z, Z') satisfying (7.13). Thus, to prove (7.1) it is enough to show that $Z \sim \widehat{Z}$, i.e., that $\sum_{x\in\mathcal{B}_{\mathrm{n}}} h(y,x)\,\mathbb{P}(Z' = x) = \mathbb{P}(\widehat{Z} = y)$ for any $y \in \mathcal{B}_{\mathrm{n}}$ such that $\kappa(y) = k$.

Observe that for any $r \in [n]$ such that $x_r = 0$ and $\kappa(x) = k-1$,

$$\frac{\mathbb{P}\left(Z' = x\right)}{\mathbb{P}(\widehat{Z} = x + e_r)} = \frac{\mathbb{P}\left(B = x\right)}{\mathbb{P}\left(B = x+e_r\right)}\frac{\mathbb{P}\left(\kappa(B) = k\right)}{\mathbb{P}\left(\kappa(B) = k-1\right)} = \frac{1-p_r}{p_r}\frac{\mathbb{P}\left(\kappa(B)=k\right)}{\mathbb{P}\left(\kappa(B)=k-1\right)}. \tag{7.14}$$

Moreover, for any $f\colon \mathcal{B}_{\mathrm{n}} \to \mathbb{R}$ and $r \in [n]$,

$$\mathbb{E}\left[f(B)\mathbf{1}_{\{B_r=1\}}\right]\frac{1-p_r}{p_r} = \mathbb{E}\left[f(B+e_r)\mathbf{1}_{\{B_r=0\}}\right]. \tag{7.15}$$

We use (7.14) and (7.15) to get that for any such y and any $r \in [n]$ such that $y_r = 1$ and $\kappa(y) = k$,

$$\begin{aligned}
\frac{h(y, y-e_r)\,\mathbb{P}(Z' = y-e_r)}{\mathbb{P}(\widehat{Z} = y)} &\overset{(7.14)}{=} h(y,y-e_r)\frac{1-p_r}{p_r}\frac{\mathbb{P}\left(\kappa(B)=k\right)}{\mathbb{P}\left(\kappa(B)=k-1\right)}\\
&= \mathbb{E}\Big[\frac{\mathbf{1}_{\{B_r=1\}}\mathbf{1}_{\{\kappa(B)=k\}}}{\mathbf{1}_{\{B_r=1\}} + \sum_{l\neq r}\mathbf{1}_{\{B_l=1\}}\mathbf{1}_{\{y_l=0\}}}\Big]\frac{(1-p_r)/p_r}{\mathbb{P}\left(\kappa(B)=k-1\right)}\\
&\overset{(7.15)}{=} \mathbb{E}\Big[\frac{\mathbf{1}_{\{B_r=0\}}\mathbf{1}_{\{\kappa(B)=k-1\}}}{\mathbf{1}_{\{B_r=0\}} + \sum_{l\neq r}\mathbf{1}_{\{B_l=1\}}\mathbf{1}_{\{y_l=0\}}}\Big]\frac{1}{\mathbb{P}\left(\kappa(B)=k-1\right)}\\
&= \mathbb{E}\Big[\frac{\mathbf{1}_{\{B_r=0\}}\mathbf{1}_{\{\kappa(B)=k-1\}}}{\mathbf{1}_{\{B_r=0\}} + \sum_{l\neq r}\mathbf{1}_{\{B_l=0\}}\mathbf{1}_{\{y_l=1\}}}\Big]\frac{1}{\mathbb{P}\left(\kappa(B)=k-1\right)}\\
&= \mathbb{E}\Big[\frac{\mathbf{1}_{\{Z'_r=0\}}}{\sum_l \mathbf{1}_{\{Z'_l=0\}}\mathbf{1}_{\{y_l=1\}}}\Big],
\end{aligned} \tag{7.16}$$

where the penultimate step comes from the fact that for any u, v such that $\kappa(u) = \kappa(v)$ one has $\sum \mathbf{1}_{\{u=0\}}\mathbf{1}_{\{v=1\}} = \sum \mathbf{1}_{\{u=1\}}\mathbf{1}_{\{v=0\}}$ applied to $u = \xi_r(B)$, $v = \xi_r(y)$, where ξ_r is the projection from $\mathcal{B}_{\mathrm{n}}$ to $\mathcal{B}_{n-1}$ obtained by skipping the r-th coordinate (note that if $B_r = 0$ and $\kappa(B) = k-1$ then $\kappa(u) = \kappa(v) = k-1$). Therefore, by (7.16),

$$\frac{\sum_x h(y,x)\,\mathbb{P}(Z' = x)}{\mathbb{P}(\widehat{Z} = y)} = \frac{\sum_{r\colon y_r=1} h(y,y-e_r)\,\mathbb{P}(Z' = y-e_r)}{\mathbb{P}(\widehat{Z} = y)} = 1,$$

which completes the proof of (7.1). The equality (7.2) follows again by (7.16). $\square$

References

Adamczak, R., Kotowski, M., Polaczyk, B., and Strzelecki, M. A note on concentration for polynomials in the Ising model. *Electron. J. Probab.*, **24**, Paper No. 42, 22 (2019). MR3949267.

Adamczak, R., Polaczyk, B., and Strzelecki, M. Modified log-Sobolev inequalities, Beckner inequalities and moment estimates. *J. Funct. Anal.*, **282** (7), Paper No. 109349, 76 (2022). MR4372142.

Adamczak, R. and Strzelecki, M. On the convex Poincaré inequality and weak transportation inequalities. *Bernoulli*, **25** (1), 341–374 (2019). MR3892322.

Adamczak, R. and Wolff, P. Concentration inequalities for non-Lipschitz functions with bounded derivatives of higher order. *Probab. Theory Related Fields*, **162** (3-4), 531–586 (2015). MR3383337.

Aida, S., Masuda, T., and Shigekawa, I. Logarithmic Sobolev inequalities and exponential integrability. *J. Funct. Anal.*, **126** (1), 83–101 (1994). MR1305064.

Anari, N., Jain, V., Koehler, F., Pham, H. T., and Vuong, T.-D. Entropic Independence in High-Dimensional Expanders: Modified Log-Sobolev Inequalities for Fractionally Log-Concave Polynomials and the Ising Model. *ArXiv Mathematics e-prints* (2021). arXiv: 2106.04105.

Anari, N., Liu, K., Gharan, S. O., and Vinzant, C. Log-concave polynomials II: High-dimensional walks and an FPRAS for counting bases of a matroid. In *STOC'19—Proceedings of the 51st Annual ACM SIGACT Symposium on Theory of Computing*, pp. 1–12. ACM, New York (2019). MR4003314.

Ané, C., Blachère, S., Chafaï, D., et al. *Sur les inégalités de Sobolev logarithmiques*, volume 10 of *Panoramas et Synthèses*. Société Mathématique de France, Paris (2000). ISBN 2-85629-105-8. MR1845806.

Aoun, R., Banna, M., and Youssef, P. Matrix Poincaré inequalities and concentration. *Adv. Math.*, **371**, 107251, 33 (2020). MR4108222.

Azuma, K. Weighted sums of certain dependent random variables. *Tohoku Math. J. (2)*, **19**, 357–367 (1967). MR0221571.

Bakry, D., Ledoux, M., and Saloff-Coste, L. *Markov semigroups at Saint-Flour*. Probability at Saint-Flour. Springer, Heidelberg (2012). ISBN 978-3-642-25937-1. MR3075390.

Bertail, P., Chautru, E., and Clémençon, S. Empirical processes in survey sampling with (conditional) Poisson designs. *Scand. J. Stat.*, **44** (1), 97–111 (2017). MR3619696.

Bertail, P. and Clémençon, S. Bernstein-type exponential inequalities in survey sampling: conditional Poisson sampling schemes. *Bernoulli*, **25** (4B), 3527–3554 (2019). MR4010964.

Bhatia, R. *Matrix analysis*, volume 169 of *Graduate Texts in Mathematics*. Springer-Verlag, New York (1997). MR1477662.

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

Bobkov, S. G. and Tetali, P. Modified logarithmic Sobolev inequalities in discrete settings. *J. Theoret. Probab.*, **19** (2), 289–336 (2006). MR2283379.

Borcea, J., Brändén, P., and Liggett, T. M. Negative dependence and the geometry of polynomials. *J. Amer. Math. Soc.*, **22** (2), 521–567 (2009). MR2476782.

Boucheron, S., Bousquet, O., Lugosi, G., and Massart, P. Moment inequalities for functions of independent random variables. *Ann. Probab.*, **33** (2), 514–560 (2005). MR2123200.

Boucheron, S., Lugosi, G., and Massart, P. On concentration of self-bounding functions. *Electron. J. Probab.*, **14**, no. 64, 1884–1899 (2009). MR2540852.

Boucheron, S., Lugosi, G., and Massart, P. *Concentration inequalities. A nonasymptotic theory of independence*. Oxford University Press, Oxford (2013). ISBN 978-0-19-953525-5. MR3185193.

Brändén, P. and Huh, J. Lorentzian polynomials. *Ann. of Math. (2)*, **192** (3), 821–891 (2020). MR4172622.

Chen, S. X. General properties and estimation of conditional Bernoulli models. *J. Multivariate Anal.*, **74** (1), 69–87 (2000). MR1790614.

Chen, S. X. and Liu, J. S. Statistical applications of the Poisson-binomial and conditional Bernoulli distributions. *Statist. Sinica*, **7** (4), 875–892 (1997). MR1488647.

Chen, X.-H., Dempster, A. P., and Liu, J. S. Weighted finite population sampling to maximize entropy. *Biometrika*, **81** (3), 457–469 (1994). MR1311090.

Cryan, M., Guo, H., and Mousa, G. Modified log-Sobolev inequalities for strongly log-concave distributions. *Ann. Probab.*, **49** (1), 506–525 (2021). MR4203344.

Dagan, Y., Daskalakis, C., Dikkala, N., and Kandiros, A. V. Learning Ising models from one or multiple samples. In *STOC '21—Proceedings of the 53rd Annual ACM SIGACT Symposium on Theory of Computing*, pp. 161–168. ACM, New York (2021). MR4398832.

Dubhashi, D. P. and Panconesi, A. *Concentration of measure for the analysis of randomized algorithms*. Cambridge University Press, Cambridge (2009). ISBN 978-0-521-88427-3. MR2547432.

Gao, F. and Quastel, J. Exponential decay of entropy in the random transposition and Bernoulli-Laplace models. *Ann. Appl. Probab.*, **13** (4), 1591–1600 (2003). MR2023890.

Garbe, K. and Vondrak, J. Concentration of Lipschitz functions of negatively dependent variables. *ArXiv Mathematics e-prints* (2018). arXiv: 1804.10084.

Garg, A., Kathuria, T., and Srivastava, N. Scalar Poincaré implies matrix Poincaré. *Electron. Commun. Probab.*, **26**, Paper No. 20, 4 (2021). MR4245747.

Goel, S. Modified logarithmic Sobolev inequalities for some models of random walk. *Stochastic Process. Appl.*, **114** (1), 51–79 (2004). MR2094147.

Han, Q. and Wellner, J. A. Complex sampling designs: uniform limit theorems and applications. *Ann. Statist.*, **49** (1), 459–485 (2021). MR4206686.

Hermon, J. and Salez, J. Modified log-Sobolev inequalities for strong-Rayleigh measures. *Ann. Appl. Probab.*, **33** (2), 1301–1314 (2023). MR4564432.

Huang, D. and Tropp, J. A. Nonlinear matrix concentration via semigroup methods. *Electron. J. Probab.*, **26**, Paper No. 8, 31 (2021). MR4216521.

Jerrum, M., Son, J.-B., Tetali, P., and Vigoda, E. Elementary bounds on Poincaré and log-Sobolev constants for decomposable Markov chains. *Ann. Appl. Probab.*, **14** (4), 1741–1765 (2004). MR2099650.

Kathuria, T. A Matrix Bernstein Inequality for Strong Rayleigh Distributions. *ArXiv Mathematics e-prints* (2020). arXiv: 2011.13340.

Kaufman, T., Kyng, R., and Soldá, F. Scalar and matrix Chernoff bounds from ℓ_∞-independence. In *Proceedings of the 2022 Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pp. 3732–3753 (2022). MR4415184.

Kaufman, T. and Oppenheim, I. High order random walks: beyond spectral gap. In *Approximation, randomization, and combinatorial optimization. Algorithms and techniques*, volume 116 of *LIPIcs. Leibniz Int. Proc. Inform.*, pp. Art. No. 47, 17. Schloss Dagstuhl. Leibniz-Zent. Inform., Wadern (2018). MR3857285.

Kim, J. H. and Vu, V. H. Concentration of multivariate polynomials and its applications. *Combinatorica*, **20** (3), 417–434 (2000). MR1774845.

Kyng, R. and Song, Z. A matrix Chernoff bound for strongly Rayleigh distributions and spectral sparsifiers from a few random spanning trees. In *59th Annual IEEE Symposium on Foundations of Computer Science—FOCS 2018*, pp. 373–384. IEEE Computer Soc., Los Alamitos, CA (2018). MR3899605.

Latała, R. Estimates of moments and tails of Gaussian chaoses. *Ann. Probab.*, **34** (6), 2315–2331 (2006). MR2294983.

Ledoux, M. On Talagrand's deviation inequalities for product measures. *ESAIM Probab. Statist.*, **1**, 63–87 (1995/97). MR1399224.

Ledoux, M. *The concentration of measure phenomenon*, volume 89 of *Mathematical Surveys and Monographs*. American Mathematical Society, Providence, RI (2001). ISBN 0-8218-2864-9. MR1849347.

Lee, T.-Y. and Yau, H.-T. Logarithmic Sobolev inequality for some models of random walks. *Ann. Probab.*, **26** (4), 1855–1873 (1998). MR1675008.

Levin, D. A. and Peres, Y. *Markov chains and mixing times.* American Mathematical Society, Providence, RI, second edition (2017). ISBN 978-1-4704-2962-1. MR3726904.

Liggett, T. M. *Continuous time Markov processes. An introduction*, volume 113 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI (2010). ISBN 978-0-8218-4949-1. MR2574430.

Lu, S. L. and Yau, H.-T. Spectral gap and logarithmic Sobolev inequality for Kawasaki and Glauber dynamics. *Comm. Math. Phys.*, **156** (2), 399–433 (1993). MR1233852.

McDiarmid, C. Concentration. In *Probabilistic methods for algorithmic discrete mathematics*, volume 16 of *Algorithms Combin.*, pp. 195–248. Springer, Berlin (1998). MR1678578.

O'Donnell, R. *Analysis of Boolean functions.* Cambridge University Press, New York (2014). MR3443800.

Paulin, D. The convex distance inequality for dependent random variables, with applications to the stochastic travelling salesman and other problems. *Electron. J. Probab.*, **19**, no. 68, 34 (2014). MR3248197.

Paulin, D., Mackey, L., and Tropp, J. A. Efron-Stein inequalities for random matrices. *Ann. Probab.*, **44** (5), 3431–3473 (2016). MR3551202.

Pemantle, R. Towards a theory of negative dependence. *J. Math. Phys.*, **41** (3), 1371–1390 (2000). MR1757964.

Pemantle, R. and Peres, Y. Concentration of Lipschitz functionals of determinantal and other strong Rayleigh measures. *Combin. Probab. Comput.*, **23** (1), 140–160 (2014). MR3197973.

Sambale, H. and Sinulis, A. Logarithmic Sobolev inequalities for finite spin systems and applications. *Bernoulli*, **26** (3), 1863–1890 (2020). MR4091094.

Sambale, H. and Sinulis, A. Modified log-Sobolev inequalities and two-level concentration. *ALEA Lat. Am. J. Probab. Math. Stat.*, **18** (1), 855–885 (2021). MR4243518.

Sambale, H. and Sinulis, A. Concentration inequalities on the multislice and for sampling without replacement. *J. Theoret. Probab.*, **35** (4), 2712–2737 (2022). MR4509083.

Samson, P.-M. Transport-entropy inequalities on locally acting groups of permutations. *Electron. J. Probab.*, **22**, Paper No. 62, 33 (2017). MR3690287.

Shao, Q.-M. A comparison theorem on moment inequalities between negatively associated and independent random variables. *J. Theoret. Probab.*, **13** (2), 343–356 (2000). MR1777538.

Talagrand, M. An isoperimetric theorem on the cube and the Kintchine-Kahane inequalities. *Proc. Amer. Math. Soc.*, **104** (3), 905–909 (1988). MR964871.

Talagrand, M. Concentration of measure and isoperimetric inequalities in product spaces. *Inst. Hautes Études Sci. Publ. Math.*, **81**, 73–205 (1995). MR1361756.

Tillé, Y. *Sampling algorithms.* Springer Series in Statistics. Springer, New York (2006). MR2225036.

Tropp, J. A. Freedman's inequality for matrix martingales. *Electron. Commun. Probab.*, **16**, 262–270 (2011). MR2802042.

Tropp, J. A. User-friendly tail bounds for sums of random matrices. *Found. Comput. Math.*, **12** (4), 389–434 (2012). MR2946459.

Vershynin, R. *High-dimensional probability. An introduction with applications in data science*, volume 47 of *Cambridge Series in Statistical and Probabilistic Mathematics*. Cambridge University Press, Cambridge (2018). ISBN 978-1-108-41519-4. MR3837109.

Verzelen, N. and Gassiat, E. Adaptive estimation of high-dimensional signal-to-noise ratios. *Bernoulli*, **24** (4B), 3683–3710 (2018). MR3788186.