University of Warsaw
Faculty of Mathematics, Informatics and Mechanics

Bartłomiej Polaczyk

Concentration of measure and functional inequalities

PhD dissertation

Supervisor
dr hab. Radosław Adamczak
Institute of Mathematics
University of Warsaw

March 2023

Declaration of Authorship

Author's declaration:
I hereby declare that this dissertation is my own work.

....	....
Date	Signature

Supervisor's declaration:
The dissertation is ready to be reviewed

....	....
Date	Signature

Abstract

This thesis is devoted to the study of the concentration of measure phenomenon and its connections with functional inequalities. We focus on the relations between various types of inequalities and in the case of concentration estimates, we are mostly interested in discrete dependent random variables.

In particular, we prove that Beckner inequalities with constants separated from zero as $p \to 1^+$ are equivalent to the modified log Sobolev inequality. Further, we derive Sobolev type moment estimates which hold under these functional inequalities. We illustrate these results with applications to concentration of measure estimates for various stochastic models, including random permutations, zero-range processes, strong Rayleigh measures, exponential random graphs, and geometric functionals on the Poisson path space.

Then, we answer an open problem posed by Mossel–Oleszkiewicz–Sen regarding relations between p-log-Sobolev inequalities for $p \in (0, 1]$. We show that for any interval $I \subset (0, 1]$, there exist $q, p \in I$, $q < p$ and a measure μ for which the q-log-Sobolev inequality holds, while the p-log-Sobolev inequality is violated. As a tool we develop certain necessary and sufficient conditions characterizing those inequalities in the case of birth-death processes on N.

We also investigate concentration properties of functions of random vectors with values in the discrete cube, satisfying the stochastic covering property (SCP). Our result for SCP measures include subgaussian inequalities of bounded-difference type and their counterparts for matrix-valued setting. 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.

Finally, we prove a Bennett-type concentration bound for suprema of empirical processes based on sampling without replacement and a corresponding bound in the case of an arbitrary Hoeffding statistics.

2010 Mathematics Subject Classification: Primary: 60E15, Secondary: 60B20, 60C05, 60J27

Keywords: concentration of measure, log-Sobolev inequality, Beckner inequality, Poincaré inequality

Bartłomiej Polaczyk
March 2023

Streszczenie

Niniejsza rozprawa poświęcona jest badaniu zjawiska koncentracji miary i jego związków z nierównościami funkcyjnymi. Skupiamy się na związkach między różnymi typami nierówności, a w przypadku oszacowań koncentracyjnych interesują nas głównie dyskretne zmienne losowe zależne.

W szczególności udowadniamy, że nierówności Becknera ze stałymi oddzielonymi od zera wraz z $p \to 1^+$ są równoważne zmodyfikowanej nierówności log-Sobolewa. Ponadto, wyprowadzamy oszacowania momentów typu Sobolewa, które zachodzą przy tych nierównościach funkcyjnych. Powyższe wyniki ilustrujemy zastosowaniami do oszacowań koncentracji miary dla różnych modeli stochastycznych, w tym permutacji losowych, procesów zerowego zasięgu, silnych miar Rayleigha, wykładniczych grafów losowych i geometrycznych funkcji na przestrzeni Poissona.

Następnie, odpowiadamy na otwarty problem postawiony przez Mossela–Oleszkiewicza–Sena, dotyczący związków pomiędzy różnymi nierównościami p-log-Sobolewa dla $p \in (0,1]$. Pokazujemy, że dla dowolnego przedziału $I \subset (0,1]$ istnieje $q, p \in I$, $q < p$ i miara μ, dla której zachodzi nierówność q-log-Sobolewa, natomiast nie zachodzi nierówność p-log-Sobolewa. Jako narzędzie wyprowadzamy pewien warunek konieczny i powiązany warunek wystarczający dla powyższych nierówności w przypadku procesów narodzin i śmierci na $\mathbb{N}$.

Badamy również koncentrację funkcji wektorów losowych o wartościach w kostce dyskretnej, spełniających własność pokrycia stochastycznego. Uzyskane wyniki dla tych miar obejmują nierówności subgaussowskie oraz ich odpowiedniki w sytuacji macierzowej. Szczegółowo traktujemy także specjalny przypadek niezależnych zmiennych losowych Bernoulliego uwarunkowanych na ich sumę, dla których otrzymujemy wzmocnione oszacowania, wyprowadzając w szczególności zmodyfikowane nierówności log-Sobolewa, nierówność Talagranda dla odległości wypukłej oraz, w konsekwencji, wyniki koncentracji dla funkcji wypukłych i wielomianów, a także ulepszone oszacowania dla funkcji o wartościach macierzowych.

W ostatniej części pracy udowadniamy oszacowania koncentracyjne typu Bennetta w przypadku supremów procesów empirycznych opartych na próbkowaniu bez zwracania oraz analogiczne oszacowania dla dowolnych statystyk Hoeffdinga.

2010 Mathematics Subject Classification: 60E15, 60B20, 60C05, 60J27

Keywords: koncentracja miary, nierówność logarytmiczna Sobolewa, nierówność Becknera, nierówność Poincaré

Bartłomiej Polaczyk
Marzec, 2023

Acknowledgments

I thank my wife, for always believing in me, even when I could not.

I would also like to express immense gratitude towards my supervisor, Radosław Adamczak, for helping me grow as a mathematician as well as for his endless patience, guidance and constant support.

The research presented in this thesis was partially supported by the National Science Centre (Poland), grant Sonata Bis no. 2015/18/E/ST1/00214 and grant Preludium no. 2020/37/N/ST1/026.

Contents

Chapter 1

Introduction

1.1 Preliminaries

A classical observation lying at the foundations of the modern probability theory says that if we run a random experiment sufficiently many times, then the resulting average outcome is very close to the expected outcome of a single experiment. This observation was originally formalized in terms of various laws of large numbers, which are qualitative in nature. Soon later, some quantitative expressions regarding exponential deviation bounds for sums of independent random variables were obtained and, beginning with the work of Milman [157], these quantitative bounds were extended onto much more general situations and complicated functions.

The prevalence and strength of such bounds is a fascinating matter, being referred to as the concentration of measure phenomenon, which in its basic form has been neatly phrased by Michael Talagrand in his seminal work [192]:

> *a random variable that depends (in a "smooth" way) on the influence of many independent variables (but not too much on any of them) is essentially constant.*

The objective of this PhD thesis is to investigate concentration of measure inequalities and the interplay between various approaches that lead to them. In particular, we will be interested in the following two contexts:

1. Concentration for dependent random variables.

2. Functional inequalities.

The first goal of this PhD thesis is to study the concentration of measure phenomenon in the probabilistic and combinatorial setting. We will be analyzing specific, mostly discrete, models of dependent random variables that arise naturally in applications and deduce their concentration properties. We will also be interested in formulating some sufficient conditions for specific types of concentration (e.g., concentration for convex functions or polynomials) to hold and show they are satisfied in some important situations, improving previously known bounds.

The second goal of this PhD thesis is to investigate the abstract analytical setting of functional inequalities. These can be often seen as a bridge between the theory of convergence to stationarity of Markov processes and concentration properties of limiting distributions. In particular, we will be analyzing the dependence between various types of functional inequalities, deriving their characterization and investigating what concentration properties they yield.

1.2 Classical results

Since the 70s, the concentration of measure has been an active field of research, which, in spite of being thoroughly examined [138, 84, 51], still sparks a lot of attention, see, e.g., [170, 96, 40, 2, 101, 167]. Not only is it one of the main areas of research of the contemporary probability theory but also finds numerous applications in other branches of mathematics [60, 171, 49], as well as in other fields such as statistics, computer science, statistical physics and quantum mechanics to name only a few (cf. [198, 84, 159, 114]).

In principle, the investigation of concentration properties can be stated as a problem of finding the best possible $\alpha : [0, \infty) \to [0, 1]$ such that the deviation bound

$$\forall t \quad \mathbb{P}(f(X) - \mathbb{E}\, f(X) > t) \leq \alpha(t) \tag{1.2.1}$$

holds for all f belonging to some class $\mathcal{F}$ and a given random element X. We say informally that X has strong concentration properties if α decays rapidly to 0 as $t \to \infty$ and $\mathcal{F}$ is possibly large.

In the classical and well studied cases, $\mathcal{F}$ is a class of regular functions (e.g., 1-Lipschitz with respect to some metric) and X is a random vector satisfying strong probabilistic or geometric conditions (e.g., independence of the coordinates, uniform log-concavity). In such situations it is often easy to deduce strong concentration properties of X using classical (Hoeffding, McDiarmid, Bernstein etc.) bounds and criteria (e.g., Bakry-Emery approach). In applications however, these assumptions are often too restrictive. In particular, we often wish to find concentration bounds in cases when either X is a vector of dependent random variables (with no additional geometric assumptions), or the class $\mathcal{F}$ is wider than the class of 1-Lipschitz functions and whence the need to develop the theory beyond the classical scope. Sometimes, these objectives altogether may not allow for any applicable concentration bounds and we have to find some balance between them, e.g., imposing stronger assumptions on $\mathcal{F}$ allows to relax the assumptions put on X.

For example, it was firstly observed by Talagrand [192] that the restriction to the class of 1-Lipschitz and convex functions allows for weaker assumptions on the vector X. Namely, it suffices for X to have uniformly bounded and independent coordinates to imply the concentration bound (1.2.1) with $\alpha(t) = Ce^{-ct^2}$ for some C, c dependent on the upper bound on the coordinates of X but, most notably, independent on the dimension n. As noted in [191], such a result (with constant independent on the dimension) cannot hold for all 1-Lipschitz functions even for the simple case of the uniform distribution on the hypercube $\{-1, 1\}^n$. In the last years, much attention has been directed towards studying the dimension-free convex concentration, cf. [105, 185, 7, 149].

Obtaining concentration bounds (1.2.1) can be achieved via plethora of methods. It is beyond the scope of this outline to mention all the approaches, whence we sketch only some of them below.

Historically, the first proofs were geometric in nature. In [157] Milman gave a new proof of Dvoretzky's theorem which relies on Lévy's isoperimetric inequality. It is considered a starting point of the investigation of the concentration of measure phenomenon. Soon after, Tsirelson and Sudakov [186] and Borell [46] solved independently the isoperimetric problem for the Gaussian measure (which implies the concentration for 1-Lipschitz functions of independent Gaussian random variables).

Another metric approach is via the convex-hull approximation technique developed by Talagrand [192, 193, 189]. Some proofs of Talagrand were later simplified by Marton [146, 147] who developed the optimal transportation method. There's also a wide family of martingale methods, based on the Azuma-type inequality. These were developed firstly by Maurey [151] and then extended by Schechtman [182] and McDiarmid [152]. Together with the method of exchangeable pairs invented by Chatterjee [59], these techniques allow proving strong concentration bounds in many non-trivial situations, see e.g. [84, 61, 167].

Bounds (1.2.1) can be also approached more analytically, by means of functional inequalities such as, e.g., Poincaré's inequality [9], log-Sobolev inequality [108], Maurey's infimum convolution inequality [151] and the p-variance inequalities by Latała and Oleszkiewicz [136]. The approach via functional inequalities also displays many deep connections to the theory of optimal transport (cf. [165, 103, 7, 105, 102, 101, 104, 190]), making it a separate and dynamic area of research.

In the following sections, we briefly provide background and outline of the main results of this PhD thesis.

1.3 Functional inequalities

1.3.1 Background

A fruitful approach to concentration inequalities is the one relying on the semigroup techniques and functional inequalities. Due to the wide range of their applications, studying the relations between various forms of functional and transportation inequalities is an important area of research in the theory of concentration of measure.

Let L be a generator of some Markov process with invariant measure π, and $\mathcal{E}(f,g) = -\pi(fLg)$ be its Dirichlet form (here $\pi(f)$ denotes $\int f\,d\pi$). The functional inequalities we are interested in are of the two forms: the p-log-Sobolev inequalities considered in [158] generalizing the log-Sobolev inequality due to Gross [108],

$$\begin{cases} \Upsilon_p \operatorname{Ent}_\pi(f^p) \leq \frac{p^2}{p-1}\mathcal{E}(f^{p-1},f) & \text{for } p \in \mathbb{R}\setminus\{0,1\}, \\ \Upsilon_1 \operatorname{Ent}_\pi(f) \leq \mathcal{E}(\log f, f) & \text{for } p = 1, \\ \Upsilon_0 \operatorname{Var}(\log f) \leq \mathcal{E}(1/f, f) & \text{for } p = 0, \end{cases} \tag{1.3.1}$$

where $\Upsilon > 0$ and $\operatorname{Ent}_\pi(f) = \pi(f\log f) - \pi(f)\log\pi(f)$ is the entropy functional, and the family of Beckner inequalities introduced in [29, 42],

$$\alpha_p[\pi(f^p) - (\pi(f))^p] \leq \frac{p}{2}\mathcal{E}(f^{p-1},f) \quad \text{for } p \in (1,2]. \tag{1.3.2}$$

In this nomenclature, the standard log-Sobolev inequality corresponds to the 2-log-Sobolev inequality (1.3.1) and the usual Poincaré inequality corresponds to the Beckner inequality (1.3.2) with $p = 2$. Such inequalities in various forms describe appropriate qualities of the Markov process and concentration properties of the invariant distribution π and have been studied thoroughly in the recent years [42, 136, 109, 99, 158, 114].

In presence of the abundance of different functional inequalities, a natural question that arises is that of their hierarchy. For example, a classical observation [42, 109] is that if the Beckner inequality (1.3.2) is satisfied with $\alpha_p \geq \varepsilon$ for each $p \in (1,2]$ and some $\varepsilon > 0$, then dividing it by $p-1$ and taking the limit

$p \to 1^+$ yields the 1-log-Sobolev inequality (1.3.1) with $\Upsilon_1 \geq 2\varepsilon$. On the other hand, the reverse implication was established by Latała and Oleszkiewicz [136] for diffusions (or, more generally, whenever the associated carré-du-champ operator satisfies the chain rule). It is also natural to ask what happens in the case of general processes (i.e., in the absence of the chain rule). In such case the usual log-Sobolev inequality is much more restrictive than its modified version – on infinite spaces it is strictly stronger, while on finite spaces it often holds with much worse constants, which affects the concentration estimates.

Similarly, in [158] Mossel et al. have proven that if $0 \leq q \leq p \leq 2$ and the p-log-Sobolev is satisfied with $\Upsilon_p > 0$, then the q-log-Sobolev is also satisfied with $\Upsilon_q \geq \Upsilon_p > 0$, generalizing the previous results [22]. It is also known (cf. [158]) that this implication can be partially reversed. In particular, if $1+\varepsilon \leq q \leq p \leq 2$ and the q-log-Sobolev is satisfied with $\Upsilon_q > 0$ then the p-log-Sobolev is also satisfied with $\Upsilon_p \geq c(\varepsilon)\Upsilon_q$. This result cannot be extended to the whole interval $[1, 2]$, since there are known examples of measures that satisfy the 1-log-Sobolev and do not satisfy the p-log-Sobolev for any $p > 1$ (e.g., the Poisson distribution mentioned previously [38]). The importance of such results lies in the fact that they allow deducing reverse hypercontractive estimates, which is a main theme of [158]. In light of the above, Mossel et al. have posed the following open problem.

> *For which intervals I, there exist constant $c(I)$ such that for all $r, s \in I$, the r-log-Sobolev inequality with constant Υ_r implies the s-log-Sobolev inequality with constant $\Upsilon_s \geq c(I)\Upsilon_r$? A particular choice of interest are intervals $I \subset [0, 1]$.*

1.3.2 Our contribution – Beckner inequalities

In Chapter 2, based on a joint work [6], we show that if the 1-log-Sobolev (1.3.1) inequality is satisfied with $\Upsilon_1 > 0$, then each Beckner inequality (1.3.2) is satisfied with $\alpha_p \geq \Upsilon_1/6$. The importance of such a result stems from the following observations:

- there are many examples of measures that satisfy 1-log-Sobolev inequality (1.3.1) which, together with our result, allows for deducing also Beckner inequalities in these models (e.g., for zero-range processes, measures satisfying the SCP, exponential graphs etc. [115, 114, 100, 42]);

- Beckner inequalities (1.3.2) provide a convenient setting for deriving moment estimates in the spirit of [9] and [49], which we have exploited in the aforementioned work by deriving novel bounds. Such moment estimates were originally often deduced from the usual log-Sobolev inequalities, which in many cases of discrete distributions are satisfied with much worse constants (cf. random transposition model [42]) or do not hold at all (cf., e.g., birth and death process with invariant Poisson distribution [38]).

Building on that result, we obtain novel moment estimates in the following cases.

1. In the continuous case, for Cauchy-type measures.

2. In the case of jump processes, for stationary measures of Glauber dynamics, including the Ising model, exponential random graphs or hardcore model.

3. For the symmetric group as well as for empirical processes of sampling without replacement.

4. On the hypercube, for measures satisfying the stochastic covering property, strongly log-concave measures, and zero-range processes.

5. For the Poisson path space.

1.3.3 Our contribution – p-log-Sobolev inequalities

In Chapter 3, we answer the open problem of Mossel–Oleszkiewicz–Sen stated above, regarding the relations between p-log-Sobolev inequalities. In particular, we show that for any interval $I \subset (0,1]$, there exist $q,p \in I$, $q < p$, and a measure μ for which q-log Sobolev inequality holds, while p-log Sobolev inequality is violated. As a tool certain necessary and (distinct) sufficient conditions characterizing those inequalities in the case of birth-death processes on $\mathbb{N}$ are developed.

1.4 Dependent binary random variables

1.4.1 Background

Investigating families of dependent binary random variables is an important problem from the point of view of computer science and combinatorics. A wide and important class of such measures are the Rayleigh measures, i.e., measures satisfying the Rayleigh property, which is a way of measuring the negative association of the coordinates. The examples of Rayleigh measures are, e.g., vector of independent Bernoulli random variables conditioned on their sum, determinantal measures, point processes or measures obtained by running exclusion dynamics on the cube [170]. In the context of studying the measures satisfying the Rayleigh property, Pemantle and Peres [170] have introduced a more general notion of the stochastic covering property (abbrev. SCP).

Namely, for $x, y \in \Omega = \{0,1\}^n$, we say that x covers y ($x \rhd y$) if

$$x = y \quad \text{or} \quad \exists_{i \le n}\, x = y + e_i,$$

i.e., if $x \neq y$ then it can be obtained from y by increasing a single coordinate (e_i denotes the i-th unit vector of the standard basis in $\mathbb{R}^n$). For random measures μ, ν on Ω, we say that μ covers ν ($\mu \rhd \nu$) if there is a coupling of μ, ν supported on the set $\{(x,y) \in \Omega^2 \colon x \rhd y\}$. Let X be a random vector taking values in Ω. For a set $I \subseteq [n]$ we write $X_I = (x_i)_{i \in I}$. We say that X satisfies the SCP if

$$x \rhd y \quad \Rightarrow \quad \forall\, S \subset \Omega \quad \mathcal{L}(X_{S^c} \,|\, X_S = y_S) \rhd \mathcal{L}(X_{S^c} \,|\, X_S = x_S),$$

where $\mathcal{L}(X)$ denotes the distribution of the random variable X.

A measure on the hypercube is called k-homogenous if it is supported on the set $\{\, x : \sum x_i = k \,\}$. The main result of [170] says that any k-homogenous measure satisfying the SCP satisfies also the Gaussian concentration bound, i.e.,

$$\mathbb{P}(f > \mathbb{E}\, f + t) \le \exp(-ct^2/k) \tag{1.4.1}$$

for some universal constant c and all f being 1-Lipschitz with respect to the Hamming distance $d_H(x,y) = \sum_i \mathbf{1}_{x_i \neq y_i}$. This result was later extended by

Hermon and Salez [114] who (using semigroup techniques), apart from retrieving the original result by Pemantle and Peres, were able to remove the k-homogenity assumption with k replaced by $n/2$ in (1.4.1).

Note that (1.4.1) fails to be optimal for some choices of f even in the case of product measures. Namely, let ε_i be independent Rademacher random variables and $f(x_1, \dots, x_n) = \sum_i a_i x_i$ for some non-zero numbers a_i. It is easily verified that product measures satisfy the SCP and thus the result from [114] implies $\mathbb{P}(f > \mathbb{E} f + t) \leq \exp(-2ct^2/n\|a\|_\infty^2)$. On the other hand, the application of classical Hoeffding's inequality yields $\mathbb{P}(f > \mathbb{E} f + t) \leq \exp(-\tilde{c}t^2/\|a\|_2^2)$ which improves upon the previous bound whenever $\|a\|_2$ is much smaller than $\sqrt{n}\|a\|_\infty$.

1.4.2 Our contribution

In Chapter 4, based on a joint work [5], we develop two types of results.

The first series of results concerns general measures satisfying the SCP for which we refine the Azuma type martingale argument used by Pemantle and Peres [170] and generalize (1.4.1) 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. Next, we use the approach developed for the scalar case together with matrix bounded-difference inequality due to Tropp [197] to get an analogous concentration for matrix-valued functions, strengthening the results of Aoun et al. [16]. Under a stronger assumption of the SRP we are also able to extend the Bernstein-type inequality of Kyng and Song [131] from linear combinations with coefficients in nonnegative definite matrices to general functions satisfying a matrix bounded-difference type assumptions.

The second line of research concerns the functional approach to improved concentration inequalities. We develop an abstract condition 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. While we are not able to prove our condition holds for an arbitrary SCP measure, 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. In particular, we extend the results of Bobkov–Tetali [42] on uniform measures on the slices of the hypercube, which is partially motivated by applications to statistics and geometry.

1.5 Sampling without replacement and Hoeffding statistics

1.5.1 Background

Consider a set of vectors $\mathcal{X} \subset \mathbb{R}^n$. Let $I_1, \dots, I_n$ be a uniform sample without replacement and $J_1, \dots, J_n$ be a sample with replacement from the set $[n]$. For

$m \le n$, define

$$Z = \sup_{x \in \mathcal{X}} \sum_{k=1}^{m} x_{I_k}, \qquad Z' = \sup_{x \in \mathcal{X}} \sum_{k=1}^{m} x_{J_k}$$

so that Z' can be considered a supremum of the empirical process in independent random variables J_k. Tails of Z', in much greater generality, have been extensively studied beginning with the works of Talagrand [193] and Ledoux [137], see also, e.g., [30, 51].

When studying Z, it is often convenient to represent it as a supremum of Hoeffding statistics over a family of matrices, i.e., functions of the form

$$\sum_{k=1}^{n} a_{k\sigma(k)},$$

where $(a_{ij})_{i,j=1}^{n} \in \mathbb{R}^{n \times n}$ is some real matrix. Concentration properties of Hoeffding statists were studied in, e.g., [62, 30, 11].

1.5.2 Our contribution

In Chapter 5, based on the work [173], we prove a Bennett-type concentration bound for Z, improving (in some situations) upon a corresponding bound [195], where the authors consider the deviations of Z above $\mathbb{E}\, Z'$, which in some cases can be order of magnitudes greater that $\mathbb{E}\, Z$. A corresponding Bennett-type inequality in the case of arbitrary Hoeffding statistics is also derived in [173], providing the first result that captures both the subgaussian and Poisson behaviors of Hoeffding statistics. This in particular improves (up to the numerical constants in the exponent) upon bounds by Chatterjee [62], Bercu–Deylon–Rio [30] and Albert [11].

1.6 Dissertation structure

This PhD Thesis is organized as follows.

1. Chapter 2 is devoted to the modified log-Sobolev inequalities, Beckner inequalities and moment estimates outlined in Section 1.3.2. The results are based on a joint work [6].

2. Chapter 3 is devoted to relations between p-log-Sobolev inequalities and a solution to Mossel–Oleszkiewicz–Sen problem outlined in Section 1.3.3. The results are yet unpublished.

3. Chapter 4 is devoted to concentration inequalities for some negatively dependent binary random variables outlined in Section 1.4.2. The results are based on a joint work [5].

4. Chapter 5 is devoted to concentration bounds for sampling without replacement and Hoeffding statistics outlined in Section 1.5.2. The results are based on [173].

Chapter 2

Beckner inequalities and moment estimates

2.1 Introduction

2.1.1 Motivation and informal presentation

In the work [29] Beckner proposed a family of inequalities interpolating between the Poincaré and log-Sobolev inequalities and showed that they held true in the case of standard Gaussian measures. Their form was subsequently generalized by Latała and Oleszkiewicz [136] who used them to obtain intermediate concentration estimates between subexponential and subgaussian. See [51] for further developments.

While initially studied mostly in the analytic setting, for diffusions on $\mathbb{R}^n$ or on Riemannian manifolds, all the aforementioned inequalities have their counterparts for general Markov semigroups, including those of Markov processes on discrete spaces. They are however not unique, since due to the lack of the chain rule, two forms of a single inequality, which are equivalent in the continuous framework, may differ significantly in the general case.

In particular one distinguishes between the log-Sobolev inequality and a weaker modified log-Sobolev inequality. Also, Beckner's inequalities have two formulations, one of them stronger than the other one. Before stating our results in detail, let us briefly recall some of those inequalities. We will introduce the remaining ones in Section 2.1.4. For now, we will be working in the setting of Dirichlet forms and keep the presentation slightly informal. The general setting of this chapter will be described more precisely in Section 2.1.2. Below $(\mathcal{X}, \mathcal{B}, \mu)$ is a probability space and $\mathrm{Dom}(\mathcal{E}) \subseteq L_2(\mathcal{X}, \mu)$ is a linear subspace on which a Dirichlet form $\mathcal{E}$ is defined.

Recall that one says that μ and $\mathcal{E}$ satisfy the Poincaré inequality if there exists a constant $\lambda > 0$ such that

$$\lambda \operatorname{Var}_\mu(f) \le \mathcal{E}(f, f) \tag{2.1.1}$$

for any $f \in \mathrm{Dom}(\mathcal{E})$, where $\operatorname{Var}_\mu(f) = \mu(f^2) - \mu(f)^2$ is the variance of f treated as a random variable on the probability space $(\mathcal{X}, \mathcal{B}, \mu)$ (recall that we use the common notation $\mu(f) = \int_{\mathcal{X}} f \, d\mu$).

One says that the modified log-Sobolev inequality is satisfied if there exists a constant $\rho_0 > 0$ such that

$$\rho_0 \operatorname{Ent}_\mu(f) \le \mathcal{E}(f, \log f) \tag{2.1.2}$$

for any nonnegative $f \in \mathrm{Dom}(\mathcal{E})$ such that $\log f \in \mathrm{Dom}(\mathcal{E})$ (which corresponds to the 1-log-Sobolev inequality (1.3.1) from Chapter 1).

Finally, recall that the Beckner inequality with parameter $p \in (1, 2]$ holds if there exists a constant $\alpha_p > 0$ such that

$$\alpha_p(\mu(f^p) - \mu(f)^p) \leq \frac{p}{2}\mathcal{E}(f, f^{p-1}) \tag{2.1.3}$$

for any nonnegative $f \in \mathrm{Dom}(\mathcal{E})$ such that $f^{p-1} \in \mathrm{Dom}(\mathcal{E})$.

While for each individual p the Beckner inequality (2.1.3) is equivalent to the Poincaré inequality, from the point of view of the concentration of measure theory the full strength of (2.1.3) is captured in the behavior of the constants α_p as $p \to 1^+$.

In particular, it is a well known observation made by many authors that the Beckner inequality (2.1.3) with α_p separated from zero on the interval $(1, 2]$ implies the modified log-Sobolev inequality (2.1.2). Indeed, it is enough to divide both sides of (2.1.3) by $p - 1$ and take $\liminf$ as $p \to 1^+$ (see, e.g., [42, 109, 125]). However, somewhat surprisingly, the reverse implication is not present in the literature and in fact there are quite a few works where Beckner's inequalities are proved or discussed separately from the modified log-Sobolev inequality for the same models (see, e.g., [50, 47] and [42, 109, 125, 72, 201]). This is in contrast with the continuous case in which the equivalence has been obtained in [136] (for the Reader's convenience we describe all the connections between various inequalities in Section 2.1.4 below).

The main result of this chapter (the proof of which is presented in Section 2.2) can be summarized as follows.

Theorem 2.1.1. *The modified log-Sobolev inequality* (2.1.2) *holds with some constant* $\rho_0 > 0$ *if and only if the Beckner inequality* (2.1.3) *holds for every* $p \in (1, 2]$ *with some* α_p *bounded away from zero. Moreover, the optimal constants with which they hold satisfy* $\rho_0^{opt}(\mu) = 2\lim_{p\to1^+} \alpha_p^{opt}(\mu)$.

Apart from being important in its own right from the point of view of the abstract theory of functional inequalities for Markov semigroups, the above result is motivated by applications to the theory of concentration of measure. The classical Herbst's argument allows for deducing deviation estimates for Lipschitz (in an appropriate sense) functions from log-Sobolev inequalities. As proven by Aida and Stroock [9] (see also [36] for the discrete case), the usual log-Sobolev inequality due to Gross [108] (see Definition 2.1.3 below) implies certain moment estimates, which can be a starting point for obtaining concentration for more general functions, in particular polynomials or more generally functions with bounded derivatives of higher order. Such concentration results were obtained in [8, 39, 3] in the continuous setting and subsequently in [99, 4] in the discrete one (in particular for the Ising model). They have found numerous applications in signal processing, statistics and computer science, where they proved to be an important tool for deriving theoretical guarantees in particular for compressed sensing type algorithms or for learning Ising models (see, e.g., [199, 75, 160, 155]). They are also useful in random graph theory, allowing to obtain concentration inequalities for the subgraph count beyond the large deviation regime or for models with dependencies [8, 100].

In the discrete case however, the usual log-Sobolev inequality is much more restrictive than its modified version (2.1.2) – on infinite spaces it is strictly stronger, while on finite spaces it often holds with much worse constants, which affects the concentration estimates. At the same time Beckner inequalities for product distributions (treated as a special case of modified ϕ-Sobolev inequalities) were used by Boucheron, Bousquet, Lugosi, and Massart [47] in order to obtain moment estimates for functions of independent random variables which generalize the classical Efron–Stein inequality for the variance. It turns out that their argument can be adapted to the setting of general semigroups and beyond, providing moment estimates of the same nature as those by Aida–Stroock but under a weaker assumption of modified log-Sobolev inequality (2.1.2). This allows to treat a variety of models and obtain Sobolev type inequalities with various types of gradients. We remark that even though in certain situations it is possible to recover Beckner type inequalities by an appropriate modification of known arguments leading to modified log-Sobolev inequalities, there are cases in which such a modification of proofs does not seem straightforward at least at present. Examples include recent techniques based on approximate tensorization of entropy or entropic independence [149, 99, 178, 74, 71, 12, 34], discussed briefly in Sections 2.4.3 and 2.4.5.

Since the precise formulation of the general moment inequalities requires an introduction of some additional notation, we postpone it to Section 2.3 (see Propositions 2.3.1 and 2.3.3). Here let us just mention some of their applications, which we present in Sections 2.4 and 2.5. In the continuous case we derive L_r-Poincaré inequalities with optimal growth of constants as $r \to \infty$ for measures satisfying the Beckner–Latała–Oleszkiewicz inequality. We also obtain new inequalities for Cauchy-type measures (Section 2.4.1). In the case of jump processes, we obtain moment bounds with discrete gradients. In particular, we obtain estimates for stationary measures of Glauber dynamics, including the Ising model, exponential random graphs or hardcore model (Section 2.4.3). They can be used to derive higher order concentration inequalities, which when specialized to polynomials improve the results from [4, 100] (Section 2.5). For the symmetric group we generalize moment estimates obtained by Chatterjee for Hoeffding statistics [62] to general functions (Section 2.4.4). We apply them to empirical processes of sampling without replacement, improving recent results due to Tolstikhin, Zhivotovskiy, and Blanchard [195]. Building on recent work of Hermon and Salez [115, 114] and Cryan et al. [74] we also obtain Beckner inequalities and moment estimates for measures satisfying the stochastic covering property, strongly log-concave measures, and zero-range processes (Section 2.4.5). In Section 2.4.6 we obtain moment estimates for the Poisson path space. We remark that even though concentration of measure and functional inequalities for the Poisson space have been an object of intensive studies (to mention [15, 202, 57, 18, 19, 20, 175, 106]), to the best of our knowledge we are the first to prove moment estimates for the Poisson space providing subgaussian growth of moments[1]. Various inequalities with subexponential growth of moments were considered by Houdré and Privault in [119, 120].

2.1.2 General setting

Let $(\mathcal{X}, \mathcal{B}, \mu)$ be a probability space and consider a symmetric non-negative definite bilinear form $\mathcal{E}\colon \mathrm{Dom}(\mathcal{E}) \times \mathrm{Dom}(\mathcal{E}) \to \mathbb{R}$, where $\mathrm{Dom}(\mathcal{E})$ is a linear subspace of the space of $\mathcal{B}$-measurable functions $L_0(\mathcal{X}, \mu)$.

We will consider the following abstract assumption on $\mathcal{E}$.

[1]It is worth mentioning that soon after our results were announced, moment estimates analogous to ours were obtained by Gusakova–Sambale–Thäle [111].

Assumption 1. If $\varphi\colon \mathbb{R} \to \mathbb{R}$ is a contraction and $f \in \mathrm{Dom}(\mathcal{E})$, then also $\varphi \circ f \in \mathrm{Dom}(\mathcal{E})$. Moreover, for any $f_1, f_2, g_1, g_2 \in \mathrm{Dom}(\mathcal{E})$ if we have a pointwise inequality

$$(f_1(x) - f_1(y))(f_2(x) - f_2(y)) \le (g_1(x) - g_1(y))(g_2(x) - g_2(y)) \tag{2.1.4}$$

for all $x, y \in \mathcal{X}$, then

$$\mathcal{E}(f_1, f_2) \le \mathcal{E}(g_1, g_2). \tag{2.1.5}$$

Remark 2.1.2. Let us provide some basic consequences of Assumption 1, which we are going to use. First, if $f \in \mathrm{Dom}(\mathcal{E})$ and $\varphi\colon \mathbb{R} \to \mathbb{R}$ is a contraction, then

$$\mathcal{E}(\varphi(f), \varphi(f)) \le \mathcal{E}(f, f). \tag{2.1.6}$$

Another consequence is the equality

$$\mathcal{E}(f, c) = 0$$

for any constant $c \in \mathbb{R}$. Finally, if $\varphi\colon \mathbb{R} \to \mathbb{R}$ is non-decreasing and $f, \varphi \circ f \in \mathrm{Dom}(\mathcal{E})$, then

$$\mathcal{E}(f, \varphi \circ f) \ge 0.$$

Assumption 1 is satisfied in particular if $\mathcal{E}$ is a Dirichlet form of a reversible Markov semigroup (for completeness of the exposition we recall basic properties of Dirichlet forms in the Appendix A.3). This is one of the main motivations for our investigations, however we prefer not to restrict to this specific setting, since in applications to concentration of measure and moment inequalities one may encounter quadratic forms which do not correspond to Markov semigroups. For instance, it may happen that the functional inequalities of interest are in fact valid for a larger class of functions than the domain of the Dirichlet form associated with some μ-reversible Markov process or that the quadratic form appearing on the right-hand side does not correspond to a Dirichlet form, while it still satisfies Assumption 1 and the available functional inequalities are meaningful from the concentration of measure point of view. In addition, Assumption 1 will allow us to avoid unnecessary discussion of domains and help us state our main results in a more concise way.

In our examples with $\mathcal{E}$ one will often associate a subspace $\mathcal{A} \subseteq L_0(\mathcal{X}, \mu)$ and a symmetric bilinear function $\Gamma\colon \mathcal{A} \times \mathcal{A} \to L_0(\mathcal{X}, \mu)$ such that $\Gamma(f, f) \ge 0$ and for $f, g \in \mathcal{A} \cap \mathrm{Dom}(\mathcal{E})$,

$$\mathcal{E}(f, g) = \int_{\mathcal{X}} \Gamma(f, g) d\mu. \tag{2.1.7}$$

In what follows we will write $\Gamma(f)$ for $\Gamma(f, f)$.

In the Markovian setting Γ will be the *carré du champ* operator defined as

$$\Gamma(f, g) = \frac{1}{2}\Big(L(fg) - gLf - fLg\Big), \tag{2.1.8}$$

where L is a generator of a reversible Markov semigroup on $L_2(\mathcal{X}, \mu)$ with domain $\mathrm{Dom}(L)$. In this case Γ is first defined on a suitable algebra of functions $\mathcal{A}_0 \subseteq \mathrm{Dom}(L)$ and then extended to a larger algebra $\mathcal{A}$. We refer to the monograph [21] for a very detailed description of the relations between the domain of the infinitesimal generator, the domain of the Dirichlet form, and the algebra $\mathcal{A}$.

2.1.3 Examples

We will now provide several concrete examples covered by the setting described above. We remark that even though our setting is not the same as in [109] the exposition below parallels to some extent the one from this article.

As a first example let us take a diffusion $(X_t)_{t\geq 0}$ on $\mathcal{X} = \mathbb{R}^n$, with the infinitesimal generator L given by

$$Lf(x) = \sum_{i,j=1}^{n} a_{ij}(x)\frac{\partial^2 f(x)}{\partial x_i \partial x_j} + \sum_{i=1}^{n} b_i(x)\frac{\partial f(x)}{\partial x_i},$$

$a = \sigma\sigma^T$, where σ is a smooth, locally bounded function from $\mathbb{R}^n$ to the space of $n \times d$ matrices and $b\colon \mathbb{R}^n \to \mathbb{R}^n$ is a smooth function. In this case $\mathcal{A} = C^\infty(\mathbb{R}^d)$ is the set of all smooth functions and

$$\Gamma(f,g) = \sum_{i,j=1}^{n} a_{ij}\frac{\partial f}{\partial x_i}\frac{\partial g}{\partial x_j}.$$

In order to make this class of processes fit into our setting, we need to assume that $(X_t)_{t\geq 0}$ has an invariant probability measure μ, in which case one defines

$$\mathcal{E}(f,g) = \int_{\mathcal{X}} \sum_{i,j=1}^{n} a_{ij}\frac{\partial f}{\partial x_i}\frac{\partial g}{\partial x_j}\mu(dx)$$

for $f, g \in \mathcal{A}_0$ – the space of smooth compactly supported functions, and then extends this to an appropriate domain, which is the completion of $\mathcal{A}_0$ with respect to the norm $\|f\| = \sqrt{\mu(f^2) + \mathcal{E}(f,f)}$. The assumption concerning the existence of μ is satisfied, e.g., if a is the identity matrix and $b = -\nabla V$ for some function $V\colon \mathbb{R} \to \mathbb{R}$ such that e^{-V} is integrable. One can then show that the normalized measure $\mu(dx) = \frac{1}{Z}e^{-V(x)}dx$ is an invariant measure of the process. One can also consider more general diffusions on Riemannian manifolds. At this point we should stress that this class of examples satisfies the chain rule and as a consequence many functional inequalities become equivalent, even though in the general situation they are not. For this reason, this class will not be in our focus in the subsequent part of this chapter, even though we will state some Sobolev type estimates which to our best knowledge are new also in this setting (see Section 2.4.1).

Another particular case of the operator Γ, which will become for us an important source of examples, is given by

$$\Gamma(f,g)(x) = \frac{1}{2}\int_{\mathcal{X}} (f(y) - f(x))(g(y) - g(x))Q_x(dy), \tag{2.1.9}$$

where $x \mapsto Q_x$ is a map from $\mathcal{X}$ to the set of positive measures on $\mathcal{X}$ such that for all $A \in \mathcal{B}$, $x \mapsto Q_x(A)$ is measurable and Q_x, μ satisfy the following detailed balance condition:

$$Q_x(dy)\mu(dx) = Q_y(dx)\mu(dy).\text{[2]} \tag{2.1.10}$$

[2] We treat both hand sides of this condition as measures on the product $\mathcal{X} \times \mathcal{X}$. In particular, we assume that for $B \in \mathcal{B} \otimes \mathcal{B}$, the mapping $x \mapsto Q_x(B_x)$ (where B_x is the appropriate slice of B) is measurable. Such situation occurs in particular when $Q_x(\mathcal{X}) < \infty$ or when Q_x can be approximated by some kernels Q_x^n such that $Q_x^n(\mathcal{X}) < \infty$.

The bilinear form Γ is well-defined on $\mathcal{A} \times \mathcal{A}$, where

$$\mathcal{A} = \big\{ f \in L_0(\mathcal{X}, \mu) \colon \int_{\mathcal{X}} (f(y) - f(x))^2 Q_x(dy) < \infty \ \mu\text{-a.s.} \big\}.$$

In this case

$$\mathcal{E}(f, g) = \frac{1}{2} \int_{\mathcal{X}} \int_{\mathcal{X}} (f(y) - f(x))(g(y) - g(x)) Q_x(dy) \mu(dx) \tag{2.1.11}$$

with $\mathrm{Dom}(\mathcal{E}) = \{ f \in L_0(\mathcal{X}, \mu) \colon \int_{\mathcal{X}} \int_{\mathcal{X}} (f(y) - f(x))^2 Q_x(dy) \mu(dx) < \infty \}$. It is straightforward to check that in this case Assumption 1 is satisfied. Moreover, by the detailed balance condition (2.1.10) we can further write

$$\mathcal{E}(f, g) = \int_{\mathcal{X}} \int_{\mathcal{X}} (f(x) - f(y))_+ (g(x) - g(y)) Q_x(dy) \mu(dx) \tag{2.1.12}$$

and

$$\mathcal{E}(f, f) = \int_{\mathcal{X}} \int_{\mathcal{X}} (f(x) - f(y))_+^2 Q_x(dy) \mu(dx) = \int_{\mathcal{X}} \Gamma_+(f) d\mu, \tag{2.1.13}$$

where

$$\Gamma_+(f)(x) = \int_{\mathcal{X}} (f(x) - f(y))_+^2 Q_x(dy). \tag{2.1.14}$$

We remark that in many applications to concentration of measure, passing from Γ to Γ_+ is essential, since the latter can be often effectively bounded, especially under certain convexity or monotonicity assumptions on the function f.

The case when $\mathcal{X}$ is countable and $Q_x(\mathcal{X}) < \infty$ for all $x \in \mathcal{X}$, corresponds to the Markov jump process with generator

$$Lf(x) = \int_{\mathcal{X}} (f(y) - f(x)) Q_x(dy).$$

We will however see that examples of this nature appear also in spaces which are not necessarily discrete, e.g., on the Poisson space and for general product spaces endowed with Glauber type dynamics.

2.1.4 Functional inequalities

Let us now introduce more precisely the functional inequalities we will investigate. In addition to restating the definitions of Poincaré, modified log-Sobolev, and the Beckner inequalities in the abstract setting described in Section 2.1.2, we will introduce the usual log-Sobolev inequality and Beckner inequality in its original version from [29].

Definition 2.1.3. Let $\mathcal{E}$ be a symmetric, nonnegative definite bilinear form on $\mathrm{Dom}(\mathcal{E}) \times \mathrm{Dom}(\mathcal{E})$, where $\mathrm{Dom}(\mathcal{E})$ is a linear subspace of $L_0(\mathcal{X}, \mu)$. We will say that:

(i) the Poincaré inequality is satisfied if there exists a constant $\lambda > 0$ such that

$$\lambda \, \mathrm{Var}_\mu(f) \le \mathcal{E}(f, f) \tag{P}$$

for any $f \in \mathrm{Dom}(\mathcal{E})$;

(ii) the modified log-Sobolev inequality is satisfied if there exists a constant $\rho_0 > 0$ such that

$$\rho_0 \operatorname{Ent}_\mu(f) \le \mathcal{E}(f, \log f) \tag{mLSI}$$

for any nonnegative $f \in \operatorname{Dom}(\mathcal{E})$ such that $\log f \in \operatorname{Dom}(\mathcal{E})$;

(iii) Beckner's inequality (Bec-p) with parameter $p \in (1, 2]$ holds if there exists a constant $\alpha_p > 0$ such that

$$\alpha_p(\mu(f^p) - \mu(f)^p) \le \frac{p}{2}\mathcal{E}(f, f^{p-1}) \tag{Bec-p}$$

for any nonnegative $f \in \operatorname{Dom}(\mathcal{E})$ such that $f^{p-1} \in \operatorname{Dom}(\mathcal{E})$;

(iv) the log-Sobolev inequality is satisfied if there exists a constant $\rho_1 > 0$ such that

$$\rho_1 \operatorname{Ent}_\mu(g^2) \le \mathcal{E}(g, g) \tag{LSI}$$

for any $g \in \operatorname{Dom}(\mathcal{E})$;

(v) dual Beckner's inequality (Bec'-q) with parameter $q \in [1, 2)$ holds if there exists a constant $\beta_q > 0$ such that

$$\beta_q(\mu(g^2) - \mu(g^q)^{2/q}) \le (2 - q)\mathcal{E}(g, g) \tag{Bec'-q}$$

for any nonnegative $g \in \operatorname{Dom}(\mathcal{E})$.

Remark 2.1.4. Since we only assume that $\operatorname{Dom}(\mathcal{E}) \subseteq L_0(\mathcal{X}, \mu)$, the inequalities introduced above assert in particular that the left-hand sides are well-defined. In general estimates of the form $A \le B$ in this chapter should be understood as: *if $B < \infty$, then A is well-defined and the inequality holds.*

Remark 2.1.5. We note that Beckner [29] originally considered the inequality (Bec'-q) and that both (Bec-p) and (Bec'-q) are referred to in the literature as Beckner's inequalities. In this chapter we will be primarily interested in (Bec-p), however occasionally we will also refer to (Bec'-q). To avoid possible confusion we call the latter inequality "the dual Beckner inequality". We refer the Reader to [57] (see the remark after inequality (2.15) therein) for explanation on how in the smooth setting one can formally regard (Bec'-q) as a dual version of the Sobolev inequality.

If $\mathcal{E}$ is a Dirichlet form corresponding to a diffusion, then by a substitution $f = g^2$ and by the chain rule one can easily see that the modified log-Sobolev inequality (mLSI) and the log-Sobolev inequality (LSI) are equivalent. Similarly, Beckner's inequality (Bec-p) for given p is equivalent to dual Beckner's inequality (Bec'-q) for $q = 2/p$ (one substitutes $f^p = g^2$).

In general however there is no such equivalence. It remains true that the Poincaré inequality is implied by each of the other inequalities. Other known relations between them are presented in Figure 2.1. Below we briefly comment on each of the implications. Since usually they are proved in the literature in a particular context, not necessarily agreeing with our setting, in Appendix A.2 we also provide their proofs (being simple adjustments of the arguments known from the literature). In the next section we will prove the remaining implication, between (mLSI) and (Bec-p), in particular proving Theorem 2.1.1.

The implication (LSI) $\implies$ (mLSI) with $\rho_0 \ge 4\rho_1$ was obtained by Bobkov and Tetali in [42]. The reverse implication is not true in general: if $\mu = \operatorname{Poiss}(\lambda)$, and one considers

$$\mathcal{E}(f, g) = \sum_{n \ge 0} (f(n+1) - f(n))^2 \mu(\{n\})$$

$$\begin{tikzcd}[column sep=huge, row sep=huge]
\begin{array}{c}\text{Beckner (Bec-p)}\\ \text{with } \inf \alpha_p > 0\end{array} \arrow[r, Rightarrow, "\not"'] \arrow[d, Rightarrow, "\text{with } \rho_0 \geq 2\limsup_{p\to 1} \alpha_p"'] & \begin{array}{c}\text{dual Beckner (Bec'-q)}\\ \text{with } \inf \beta_q > 0\end{array} \arrow[l, Rightarrow, bend right=15, "\text{with } \alpha_p \geq \beta_q,\ p=2/q"'] \arrow[d, Rightarrow, shift right, "\text{with } \rho_1 \geq \frac{1}{2}\limsup_{q\to 2} \beta_q"'] \\
\text{modified log-Sobolev (mLSI)} \arrow[r, Rightarrow, "\not"'] & \text{log-Sobolev (LSI)} \arrow[l, Rightarrow, bend right=15, "\text{with } \rho_0 \geq 4\rho_1"'] \arrow[u, Rightarrow, shift right, "\text{with } \beta_q \geq q\rho_1"']
\end{tikzcd}$$

Figure 2.1: Arrows denote known implications.

corresponding to the birth and death Markov process with generator $Lf(n) = f(n+1) - f(n) + \lambda^{-1}n(f(n-1) - f(n))$, then the log-Sobolev inequality (LSI) does not hold (see [38]), while the modified log-Sobolev inequality (mLSI) is satisfied (see [77]).

As observed in the original article [29] by Beckner, if (Bec'-q) holds for every $q \in [1,2)$ with β_q bounded away from zero, then the log-Sobolev inequality (LSI) holds as well with $\rho_1 \geq \frac{1}{2}\limsup_{q\to 2^-} \beta_q$. The reverse implication can be found in [136].

The implication (Bec'-q) $\implies$ (Bec-p) with $\alpha_p \geq \beta_q$ (where $p = 2/q$) seems to be a part of folklore (we have not been able to find an explicit statement in the literature). It can be easily proved using arguments used of our knowledge for the first time in [81] (see Appendix A.2 for details). The reverse implication also holds, but in this case one gets $\beta_q \geq q(2-q)\alpha_p$, so the dependence on constants degenerates when $q \to 2$. Such a degeneration indeed takes place, as the Poisson measure satisfies (Bec-p) with α_p separated from zero (which can be easily proved by known results on the two point space [42] together with tensorization and Poisson limit theorem, similarly as it was done in [77] for the modified log-Sobolev inequality), whereas it cannot satisfy (Bec'-q) with β_q separated from zero, since this would imply (LSI), which as already mentioned fails for the Poisson measure.

The observation that if (Bec-p) holds for every $p \in (1,2]$ with α_p bounded away from zero, then the modified log-Sobolev inequality (mLSI) holds with $\rho_0 \geq 2\limsup_{p\to 1^+} \alpha_p$ can be found, e.g., in [42] or [109] (to see this divide both sides of (Bec-p) by $p-1$ and take $p \to 1^+$).

One can thus see that to complete the above diagram one should verify whether (mLSI) $\implies$ (Bec-p) with $\inf_{p\in(1,2]} \alpha_p > 0$. We will establish this implication in the following section.

2.2 From modified log-Sobolev to Beckner's inequalities

2.2.1 Main result

Throughout this section we assume that we are in the setting described in Section 2.1.2, in particular that Assumption 1 holds. The next theorem contains the precise statement of the result announced above, in Theorem 2.1.1.

Theorem 2.2.1. *Let μ be a probability measure which satisfies the modified log-Sobolev inequality* (mLSI) *with constant $\rho_0 > 0$. For $p \in (1,2]$ and $\theta \in (0,1)$*

denote

$$k(p,\theta) := \Big(1 - \frac{2\big((1+\theta)^p - 1\big)}{p(p-1)(1-\theta)^2}\Big) \cdot \frac{\theta^{p-1}}{e^{p-1}(1+\theta)^{p-1}},$$
$$K_p := \max\Big\{(1-1/p);\ \frac{p}{2} \cdot \sup_{\theta\in(0,1)} k(p,\theta)\Big\}.$$

Then, for any $p \in (1,2]$, μ *satisfies the Beckner inequality* (Bec-p) *with constant* $\alpha_p \geq K_p \rho_0$.

Moreover, $\lim_{p\to 1^+} K_p = \lim_{p\to 2^-} K_p = 1/2$ *and* $\inf_{p\in(1,2]} K_p \geq 0.17$. *In particular* $\alpha_p \geq \rho_0/6$.

Note that this result is sharp in the most interesting regime, $p \to 1^+$, since if the Beckner inequality (Bec-p) holds with some constants α_p, then the modified log-Sobolev inequality (mLSI) holds with $\rho_0 \geq 2 \limsup_{p\to 1^+} \alpha_p$. Combining this observation and the above theorem yields immediately the following corollary, which in particular implies Theorem 2.1.1 from Section 2.1.

Corollary 2.2.2. *The modified log-Sobolev inequality* (mLSI) *holds with some constant* $\rho_0 > 0$ *if and only if the Beckner inequality* (Bec-p) *holds for every* $p \in (1,2]$ *with some* α_p *bounded away from zero. Moreover, the optimal constants with which they hold satisfy* $\rho_0^{opt}(\mu) = 2\lim_{p\to 1^+} \alpha_p^{opt}(\mu)$.

2.2.2 Auxiliary lemmas

In this section we gather technical lemmas to be used in the proof of Theorem 2.2.1.

Since we work in the abstract setting described in Section 2.1.2, we need the following lemma which asserts that it suffices to check the validity of the inequality (Bec-p) for bounded functions only. Its proof, as well as proofs of some other auxiliary lemmas, is deferred to Appendix A.1.

Lemma 2.2.3. *If for some* $p \in (1,2]$ *the Beckner inequality* (Bec-p) *is satisfied (with some constant* $\alpha_p > 0$*) for all bounded nonnegative functions* f *such that* $f, f^{p-1} \in \mathrm{Dom}(\mathcal{E})$, *then it is satisfied with the same constant for all nonnegative functions* f *such that* $f, f^{p-1} \in \mathrm{Dom}(\mathcal{E})$. *In particular, for all such functions* $\mu(f^p) < \infty$.

We will also need the following two well-known lemmas.

Lemma 2.2.4 ([51, Lemma 14.4])**.** *For any nonnegative* $f \in L_p(\mathcal{X},\mu)$ *and* $p \in (1,2]$

$$\mu(f^p) - \mu(f)^p \leq \mathrm{Cov}_\mu(f, f^{p-1}).$$

Lemma 2.2.5 ([81, Lemma 2.6])**.** *For* $p \in (1,2]$, $a, b, > 0$,

$$(a-b)(a^{p-1} - b^{p-1}) \leq (a^{p/2} - b^{p/2})^2 \leq \frac{p^2}{4(p-1)}(a-b)(a^{p-1} - b^{p-1}).$$

Another point-wise inequality to be used in the proof of Theorem 2.2.1 is given in the next lemma.

Lemma 2.2.6. *If* $a, b \geq e$, *then for all* $p \geq 1$,

$$(a^p - b^p)(\log a - \log b) \leq (a-b)(a^{p-1}\log a - b^{p-1}\log b)\,.$$

Proof. The inequality is equivalent to $ab(a^{p-1} - b^{p-1})(\frac{\log a}{a} - \frac{\log b}{b}) \leq 0$, which follows since the function $\frac{\log x}{x}$ is decreasing for $x \geq e$. $\square$

The last lemma we need is a simple fact concerning differentiability of the bilinear form. To verify that it holds just under Assumption 1 we provide its complete proof in Appendix A.1.

Lemma 2.2.7. *Let $s \geq 1$. Assume that $f \in \mathrm{Dom}(\mathcal{E})$ is bounded and satisfies $0 < \inf f$. Then $v(s) = \mathcal{E}(f, f^{s-1})$ is well-defined, differentiable for $s \in (1, \infty)$, right-differentiable at $s = 1$ and its derivative is given by the (also well-defined) formula $v'(s) = \mathcal{E}(f, f^{s-1} \log f)$ for $s \in [1, \infty)$.*

2.2.3 Proof of Theorem 2.2.1

Let us start with a simple proposition, which allows to deduce Beckner's inequality (Bec-p) from the modified log-Sobolev inequality (mLSI) with constant α_p degenerating as $p \to 1^+$.

Proposition 2.2.8. *If μ satisfies the Poincaré (P) inequality, then for all $p \in (1, 2]$ it satisfies Beckner's inequality (Bec-p) with constants satisfying the relation*

$$\alpha_p \geq 2\frac{p-1}{p}\lambda.$$

In particular, if the modified log-Sobolev inequality (mLSI) holds, then $\alpha_p \geq \frac{p-1}{p}\rho_0$.

Proof of Proposition 2.2.8. Fix $p \in (1, 2]$ and take any nonnegative f such that $f, f^{p-1} \in \mathrm{Dom}(\mathcal{E})$. By Lemma 2.2.3 we may and do assume that f is bounded so that all the expressions below are well-defined. By Assumption 1, the Lipschitz property of the mapping $x^{p-1} \mapsto x^{p/2}$ on the set $[0, \sup f]$ implies that $f^{p/2} \in \mathrm{Dom}(\mathcal{E})$. We have $\mathrm{Cov}_\mu(f, g) = \frac{1}{2}\int_\mathcal{X}\int_\mathcal{X}(f(x) - f(y))(g(x) - g(y))\mu(dx)\mu(dy)$. Using Lemma 2.2.4, Lemma 2.2.5 and the Poincaré inequality (P) we see that

$$\lambda(\mu(f^p) - \mu(f)^p) \leq \lambda \,\mathrm{Cov}_\mu(f, f^{p-1}) \leq \lambda \,\mathrm{Cov}_\mu(f^{p/2}, f^{p/2}) \leq \mathcal{E}(f^{p/2}, f^{p/2}).$$

By Assumption 1 and another application of Lemma 2.2.5 we conclude that

$$\lambda(\mu(f^p) - \mu(f)^p) \leq \frac{p^2}{4(p-1)}\mathcal{E}(f, f^{p-1}),$$

which ends the proof. The second part follows from the fact that the modified log-Sobolev inequality (mLSI) implies the Poincaré inequality (P), see Proposition A.2.5 in the Appendix A.2. $\square$

To handle the case of $p \to 1^+$, we will need the following proposition.

Proposition 2.2.9. *Suppose that the modified log-Sobolev inequality (mLSI) holds with some constant $\rho_0 > 0$. Then for any $p \in (1, 2]$ and any bounded $f \in \mathrm{Dom}(\mathcal{E})$ such that $\inf f > 0$,*

$$\rho_0(\mu(f^p) - \mu(f)^p) \leq \left(e\frac{\mu(f)}{\inf f}\right)^{p-1} \mathcal{E}(f, f^{p-1}).$$

Proof. Fix any $p \in (1, 2]$ and any bounded $f \in \mathrm{Dom}(\mathcal{E})$ satisfying $\inf f > 0$. By homogeneity, we may and do assume that $\inf f = e$.

For $s \in [1, p]$, let $u(s) := \mu(f^s) - \mu(f)^s$ and $v(s) := \mathcal{E}(f, f^{s-1})$. For $s \in (1, p)$, Lemma 2.2.7 implies that $v(s)$ and $v'(s) = \mathcal{E}(f, f^{s-1} \log f)$ are well-defined. Recall the variational formula for the entropy

$$\mathrm{Ent}_\mu(g) = \sup_{h \in U} \mu(gh),$$

where U is the family of all measurable functions $h \colon \mathcal{X} \to \overline{\mathbb{R}}$, such that $\mu(e^h) = 1$ (see, e.g., [51, Theorem 4.13]).

Using this formula with $g = f^s$ and $h = \log \frac{f}{\mu(f)}$, we obtain

$$u'(s) = \mu(f^s \log \tfrac{f}{\mu(f)}) + \log(\mu(f))u(s) \le \mathrm{Ent}(f^s) + \log(\mu(f))u(s).$$

Hence, by the modified log-Sobolev inequality (mLSI) and Lemma 2.2.6 combined with Assumption 1,

$$\begin{aligned}\rho_0\big(u'(s) - \log(\mu(f))u(s)\big) \le \rho_0 \,\mathrm{Ent}(f^s) &\le s\mathcal{E}(f^s, \log f)\\ &\le s\mathcal{E}(f, f^{s-1}\log f) = sv'(s).\end{aligned}$$

Consequently, since $s\mu(f)^{1-s} \le se^{1-s} \le 1$, we arrive at

$$\rho_0\big(u(s)\mu(f)^{1-s}\big)' = \rho_0\mu(f)^{1-s}\big(u'(s) - \log(\mu(f))u(s)\big) \le s\mu(f)^{1-s}v'(s) \le v'(s).$$

Integrating both sides over the interval $[1, p]$ yields the result (recall that $\inf f = e$). $\square$

Having Propositions 2.2.8 and 2.2.9 we can turn to the proof of the main result.

Proof of Theorem 2.2.1. Fix any $p \in (1, 2]$ and take any bounded nonnegative function f such that $f, f^{p-1} \in \mathrm{Dom}(\mathcal{E})$. For $\theta \in (0, 1)$ denote $g = \max(f, \theta\mu(f))$ and $P_\theta = \mathbb{P}(f < \theta\mu(f))$. Then $\mu(f^p) \le \mu(g^p)$ and, since $\mu(g) \le (1 + \theta P_\theta)\mu(f)$ and $x \mapsto (1 + \theta x)^p$ is convex for $x \in [0, 1]$,

$$\begin{aligned}\mu(f^p) - \mu(f)^p &\le \mu(g^p) - \mu(g)^p + \mu(f)^p\big((1 + \theta P_\theta)^p - 1\big)\\ &\le \mu(g^p) - \mu(g)^p + \mu(f)^p P_\theta\big((1 + \theta)^p - 1\big).\end{aligned} \tag{2.2.1}$$

Clearly $\mu(g)/\inf g \le (1 + \theta)/\theta$, therefore Proposition 2.2.9 implies that

$$\rho_0(\mu(g^p) - \mu(g)^p) \le \big(e\tfrac{1+\theta}{\theta}\big)^{p-1}\mathcal{E}(g, g^{p-1}) \le \big(e\tfrac{1+\theta}{\theta}\big)^{p-1}\mathcal{E}(f, f^{p-1}), \tag{2.2.2}$$

where we also used the fact that $x \mapsto \max(x, a)$ is a contraction and Assumption 1. Combining (2.2.1) and (2.2.2) yields a defective Beckner inequality:

$$\rho_0(\mu(f^p) - \mu(f)^p) \le \rho_0\mu(f)^p P_\theta\big((1 + \theta)^p - 1\big) + \big(e\tfrac{1+\theta}{\theta}\big)^{p-1}\mathcal{E}(f, f^{p-1}). \tag{2.2.3}$$

It remains to deal with the first summand on the right-hand side

By Taylor's expansion around $\mu(f)$ with the integral form of the remainder (and since $p \le 2$)

$$\begin{aligned}\mu(f^p) - \mu(f)^p &= p(p-1)\mu\Big(\int_{\mu(f)}^{f} u^{p-2}(f - u)\,du\Big)\\ &\ge p(p-1)\mu\Big(\mathbf{1}_{\{f<\mu(f)\}}\int_{f}^{\mu(f)} u^{p-2}(u - f)\,du\Big)\\ &\ge \frac{p(p-1)}{2}\mu(f)^{p-2}\mu\big((\mu(f) - f)_+^2\big)\,,\end{aligned}$$

while by Chebyshev's inequality

$$P_\theta = \mathbb{P}\big((1-\theta)\mu(f) < (\mu(f)-f)_+\big) \le \frac{\mu\big((\mu(f)-f)_+^2\big)}{(1-\theta)^2(\mu(f))^2},$$

whence

$$P_\theta \mu(f)^p \le \frac{2(\mu(f^p)-\mu(f)^p)}{p(p-1)(1-\theta)^2}.$$

Plugging the above estimate into (2.2.3) and optimizing over $\theta \in (0,1)$ yields Beckner's inequality (Bec-p) with

$$\alpha_p \ge \rho_0 \cdot \frac{p}{2} \cdot \sup_{\theta\in(0,1)} k(p,\theta),$$

where we recall that

$$k(p,\theta) = \Big(1 - \frac{2\big((1+\theta)^p - 1\big)}{p(p-1)(1-\theta)^2}\Big) \cdot \frac{\theta^{p-1}}{e^{p-1}(1+\theta)^{p-1}}.$$

The extension to not necessarily bounded functions follows by Lemma 2.2.3.

Of course, for some values of $p \in (1,2]$ the bound $\alpha_p \ge (1-1/p)\rho_0$ provided by Proposition 2.2.8 may be better. We shall now compare both expressions to get some more explicit estimates on the multiplicative factor

$$K_p = \max\{1-1/p;\ p/2 \cdot \sup_{\theta\in(0,1)} k(p,\theta)\}.$$

It is easy to see that $\lim_{p\to 2^-} K_p = 1/2$. Since $\lim_{p\to 1^+} k(p,(p-1)^2) = 1$ and obviously $k(p,\theta) \le 1$, we conclude that $\lim_{p\to 1^+} K_p = 1/2$.

Moreover, one can check that

$$K_p \ge \max\{1-1/p;\ p/2 \cdot k(p, 0.25\cdot(1-p)^2)\} \ge 0.17. \tag{2.2.4}$$

This ends the proof of the theorem. □

Remark 2.2.10. Note that the numerical bound (2.2.4) cannot be substantially improved if we want it to hold for *all* $p \in (1,2]$. Indeed, we have

$$K_{6/5} \le 0.18.$$

Identification of the best constants K_p^{opt} such that $\alpha_p \ge K_p^{\mathrm{opt}}\rho_0$ seems to be an interesting open question.

2.3 Moment estimates

In this section we revisit the arguments by Boucheron et al. [51, Theorem 15.5] and present them in the context of general Beckner inequalities. We derive moment inequalities, which are valid in particular under the assumption of modified log-Sobolev inequality. These moment estimates will lie at the core of the applications presented in subsequent sections. In this section we still work with a probability space $(\mathcal{X},\mathcal{B},\mu)$. Accordingly, all the moments of functions/random variables are calculated with respect to the measure μ, i.e., for $g\colon \mathcal{X}\to\mathbb{R}$, we set $\|g\|_r = (\mu(|g|^r))^{1/r}$.

Proposition 2.3.1. *Assume that Γ_+ is defined, as in (2.1.14), via some kernel Q_x satisfying the detailed-balance condition (2.1.10). Let $\mathcal{E}$ be given by (2.1.11) and assume that for some $a > 0$, $s \geq 0$ and all $p \in (1, 2]$ the Beckner inequality (Bec-p) is satisfied with constant $\alpha_p \geq a(p-1)^s$. Then for every measurable $f\colon \mathcal{X} \to \mathbb{R}$ and $r \geq 2$,*

$$\|(f - \mu(f))_+\|_r^2 \leq (1 - 2^{-(s+1)})\frac{r^{s+1}}{a}\kappa(s)\|\Gamma_+(f)\|_{r/2}, \tag{2.3.1}$$

$$\|(\mu(f) - f)_+\|_r^2 \leq (1 - 2^{-(s+1)})\frac{r^{s+1}}{a}\kappa(s)\|\Gamma_+(-f)\|_{r/2}, \tag{2.3.2}$$

where $\kappa(s) = (1 - e^{-(s+1)/2})^{-1}$.

The case of $s = 0$ corresponds via Theorem 2.2.1 to the modified log-Sobolev inequality (mLSI), while the case $s = 1$ via Proposition 2.2.8 to the Poincaré inequality. In fact, if the inequality (Bec-p) holds for *some* $p \in (1, 2]$, then also the Poincaré inequality holds (see Proposition A.2.6), whence one can find $a > 0$, such that (Bec-p) holds for *all* $p \in (1, 2]$ with $\alpha_p \geq a(p - 1)$. Thus, the interesting range of the parameter s in the above proposition is $[0, 1]$.

In most applications that we have in mind, Γ will be indeed defined by some kernel. However, similar estimates may be derived also in a more abstract setting, encompassing in particular general reversible Markov semigroups. In Section 2.4.1 we will use such a statement to present certain weighted L_p Poincaré inequalities. In line with our general approach of writing the inequalities in an abstract form, under structural assumptions, we will formulate the next result in terms of the following additional assumption.

Assumption 2.

- For any bounded $f \in \mathcal{A}$, any $c \in \mathbb{R}$, and any $\gamma > 1$, $t \geq 1$,

$$\mathcal{E}(|f + c|^\gamma, |f + c|) \leq 2\gamma \big\| |f + c|^{\gamma - 1} \big\|_{\frac{t}{t-1}} \big\| \Gamma(f) \big\|_t. \tag{2.3.3}$$

- For any $f \in \mathcal{A}$ there exists a sequence f_n of bounded elements of $\mathcal{A}$, such that $f_n \to f$ and $\Gamma(f_n) \leq \Gamma(f)$ μ-a.s.

Remark 2.3.2. The first part of the above assumption is satisfied in particular if $\mathcal{A}$ is any algebra contained in the domain of the infinitesimal operator L of a Markov semigroup reversible with respect to μ. The second part may depend on the choice of $\mathcal{A}$, however in most cases in the theory of Dirichlet forms one chooses $\mathcal{A}$ which is stable under composition with smooth bounded Lipschitz functions, which allows for appropriate truncations, implying the second part (see, e.g., Definition 3.3.1 of the extended algebra $\mathcal{A}$ in the monograph [21]). We provide derivation of both parts of Assumption 2 in this standard Markovian setting in Proposition A.3.2 in the Appendix A.3.

Let us also note that Γ satisfying (2.1.7) is not unique even if $\mathcal{E}$ is a Dirichlet form corresponding to a Markov process, e.g., one may consider Γ which is not the carré du champ operator corresponding to a Markov process for which μ is the invariant measure, but for instance to a Markov process reversible with respect to some other measure of reference, while (2.1.7) still holds. Examples of this kind can be found, e.g., in [28], where the authors consider log-Sobolev inequalities for the Ising model as well as quenched log-Sobolev inequalities for the Sherrington–Kirkpatrick model and the Dirichlet form is written in terms of Γ being the carré du champ operator for the Glauber dynamics induced by the product measure on the cube (see [85] for a comparison of the Dirichlet form used in [28] with the one induced by the usual Glauber dynamics). In such situations in order to use Assumption 2 one needs to pass to the carré du champ operator related to μ (see also the discussion in [21, p. 121]).

Proposition 2.3.3. *Let $\mathcal{E}\colon \mathrm{Dom}(\mathcal{E}) \times \mathrm{Dom}(\mathcal{E}) \to \mathbb{R}$ be a nonnegative definite symmetric bilinear form and let $\Gamma\colon \mathcal{A} \times \mathcal{A} \to L_0(\mathcal{X}, \mu)$, where $\mathcal{A} \subseteq \mathrm{Dom}(\mathcal{E})$ is a linear subspace, be a bilinear form related to $\mathcal{E}$ by (2.1.7). If Assumptions 1 and 2 are satisfied and for all $p \in (1, 2]$ the Beckner inequality (Bec-p) holds with $\alpha_p \geq a(p-1)^s$ for some $a > 0$, $s \geq 0$, then for all $f \in \mathcal{A}$ and $r \geq 2$,*

$$\|f - \mu(f)\|_r^2 \leq \frac{r^{s+1}\kappa(s)}{a}\|\Gamma(f)\|_{r/2}, \tag{2.3.4}$$

where $\kappa(s)$ is as in Proposition 2.3.1.

Remark 2.3.4. We remark that the inequalities of Propositions 2.3.1 and 2.3.3 should be again understood in the following sense: *if the right-hand side is finite, then the left-hand side is well-defined and the inequality holds*. Let us also mention that the inequalities of Proposition 2.3.3 can be extended beyond the space $\mathcal{A}$, if one replaces the right-hand side via a family of norms extending the moments of $\sqrt{\Gamma(f)}$ and defined by appropriate duality. We will not pursue this direction and refer to the article [9] by Aida–Stroock where similar moment estimates were proved under the stronger assumption of the log-Sobolev inequality (LSI). The inequalities derived by Aida–Stroock from (LSI), in our setting read as

$$||f - \mu(f)||_r^2 \leq \rho_1^{-1}(r - 3/2)\|\Gamma(f)\|_{r/2}. \tag{2.3.5}$$

We note that the derivation of moment estimates from the log-Sobolev inequality by Aida and Stroock is based on computing the derivative of $\|f\|_t^2$ with respect to $t \in [2, r]$, and identification of a part corresponding to $\mathrm{Ent}\,|f|^t$, which can be estimated via (LSI). Such an estimation allows for convenient cancellations and yields a uniform bound on the derivative on the interval $[2, r]$. This approach has been subsequently used, e.g., in [41, 8, 3] in the context of weighted log-Sobolev inequalities or various modified log-Sobolev inequalities on $\mathbb{R}^n$ (of different nature than (mLSI)). It does not seem however that this approach can work with (mLSI). Theorem 2.2.1 allows passing from (mLSI) to (Bec-p) and use the argument introduced by Boucheron, Bousquet, Lugosi, and Massart for product measures. In particular, we derive the following improvement (up to numerical constant) over (2.3.5),

$$||f - \mu(f)||_r^2 \leq \rho_0^{-1} \cdot 6r\kappa(0)\|\Gamma(f)\|_{r/2},$$

where $\kappa(0) = (1 - e^{-1/2})^{-1}$.

It is also known (see, e.g., Proposition 2.5 in [156] for the proof in the case $\Gamma(f) = |\nabla f|^2$), that the Poincaré inequality implies moment estimates of the form

$$\|f - \mu(f)\|_r^2 \leq \frac{C}{\lambda}r^2\|\Gamma(f)\|_{r/2} \tag{2.3.6}$$

for $r \geq 2$. This corresponds to the case $s = 1$ in Proposition 2.3.3.

Remark 2.3.5. It is easy to check that Assumption 2 is verified in the setting of Proposition 2.3.1. In fact the moment inequality of Proposition 2.3.3 provides better constants than one would obtain by combining the two estimates of Proposition 2.3.1 and pointwise estimates $\Gamma_+(f), \Gamma_+(-f) \leq 2\Gamma(f)$.

Remark 2.3.6. An inspection of the proofs of Propositions 2.3.1 and 2.3.3 shows that if one assumes that the inequality (Bec-p) holds just for $p \in [p_0, 2]$ for some $p_0 > 1$, then the moment estimates will still hold, but for $2 \leq r \leq r_0 = \frac{p_0}{p_0-1}$. We will use this observation in Section 2.4.1.

Proof of Proposition 2.3.1. Let us start with the inequality (2.3.1) and consider the case of bounded functions $f \in \mathcal{A}$.

We will show by induction a slightly stronger statement, namely that for all positive integers k and $r \in (k, k+1]$

$$\|(f-\mu(f))_+\|_r^2 \leq c_r \|\Gamma_+(f)\|_{\max(r/2,1)},$$

where

$$c_r = \frac{1}{a}\max\big(\frac{\kappa_r(s) r^{s+1}}{\kappa_2(s)}; 1\big), \quad \kappa_r(s) = \left(1 - \left(\frac{r-1}{r}\right)^{(s+1)r/2}\right)^{-1} \nearrow \kappa(s) \tag{2.3.7}$$

as $r \to \infty$.

In what follows the parameter r will change while s will remain fixed, so to simplify the notation we will suppress the dependence of $\kappa_r(s)$ on s and write simply κ_r.

For $k = 1$ and any $r \in (1, 2]$,

$$\|(f-\mu(f))_+\|_r^2 \leq \|f-\mu(f)\|_r^2 \leq \|f-\mu(f)\|_2^2 \leq \frac{1}{a}\mathcal{E}(f,f) \leq c_r\|\Gamma_+(f)\|_1, \tag{2.3.8}$$

where in the second step we used Jensen's inequality, in the third one the Poincaré inequality (which holds if (Bec-p) holds, see Proposition A.2.6 in the Appendix), while the last one follows from (2.1.13) and $c_r \geq c_1 = 1/a$. This yields the induction basis.

Assume that the induction hypothesis holds for all integers smaller than some $k > 1$. Consider any $r \in (k, k+1]$ and a bounded function $f \colon \mathcal{X} \to \mathbb{R}$. Choose p such that $r = \frac{p}{p-1}$ and denote $\gamma = \frac{1}{p-1}$, $g = (f-\mu(f))_+$. Applying the Beckner inequality (Bec-p) to the function g^γ and using the form (2.1.12) of $\mathcal{E}$ (which is a consequence of the detailed balance condition (2.1.10)), together with the convexity of $x \mapsto x^\gamma$, we get

$$\begin{aligned}\alpha_p(\mu(g^{\gamma p}) - \mu(g^\gamma)^p) &\leq \frac{p}{2}\int (g^\gamma(x) - g^\gamma(y))_+ (g(x)-g(y))_+ \, Q_x(dy)\mu(dx)\\ &\leq \frac{\gamma p}{2}\int g^{\gamma-1}(x)\Gamma_+(g)\mu(dx).\end{aligned} \tag{2.3.9}$$

Since $(g(x)-g(y))_+ \leq (f(x)-f(y))_+$, we have $\Gamma_+(g) \leq \Gamma_+(f)$, and so by Hölder's inequality with exponents $\frac{\gamma p}{\gamma p - 2}$ and $\frac{\gamma p}{2}$ (recall that $\gamma + 1 = \gamma p = r$), we obtain

$$\alpha_p(\mu(g^{\gamma p}) - \mu(g^\gamma)^p) \leq \frac{\gamma p}{2}(\mu(g^{\gamma p}))^{\frac{\gamma p - 2}{\gamma p}} \|\Gamma_+(f)\|_{\gamma p/2}. \tag{2.3.10}$$

Denoting $l_r = \|g\|_r$ and observing that $\alpha_p \geq a/r^s$, the above inequality divided by α_p and restated in terms of r gives

$$l_r^r \leq l_{r-1}^r + l_r^{r-2}\frac{r^{s+1}}{2a}\|\Gamma_+(f)\|_{r/2}.$$

The induction hypothesis allows us to estimate l_{r-1}:

$$l_r^r \leq \big(c_{r-1}\|\Gamma_+(f)\|_{\max((r-1)/2,1)}\big)^{r/2} + l_r^{r-2}\frac{r^{s+1}}{2a}\|\Gamma_+(f)\|_{r/2}. \tag{2.3.11}$$

Note that we can assume that $\|\Gamma_+(f)\|_{r/2} > 0$, since otherwise (as $r > 2$) we obtain $\|\Gamma_+(f)\|_1 = 0$ and by the induction assumption $(f - \mu(f))_+ = 0$. Recall that $c_r = \max\big(\kappa_r r^{s+1}/\kappa_2; 1\big)/a$ and thus, by the monotonicity in u of κ_u and $\|\Gamma_+(f)\|_{\max(u/2,1)}$, and since $r > 2$,

$$\begin{aligned}\frac{c_{r-1}\|\Gamma_+(f)\|_{\max((r-1)/2,1)}}{c_r\|\Gamma_+(f)\|_{r/2}} &\leq \frac{c_{r-1}}{c_r} \\ &\leq \max\left(\frac{\kappa_{r-1}(r-1)^{s+1}}{\kappa_r r^{s+1}}, \frac{\kappa_2}{\kappa_r r^{s+1}}\right) \leq \left(\frac{r-1}{r}\right)^{s+1}.\end{aligned}$$

Consequently, dividing (2.3.11) by $(c_r\|\Gamma_+(f)\|_{r/2})^{r/2}$, leads to

$$\left(\frac{l_r^2}{c_r\|\Gamma_+(f)\|_{r/2}}\right)^{r/2} \leq \left(\frac{r-1}{r}\right)^{(s+1)r/2} + \frac{\kappa_2}{2\kappa_r}\left(\frac{l_r^2}{c_r\|\Gamma_+(f)\|_{r/2}}\right)^{(r-2)/2}. \tag{2.3.12}$$

The function

$$h(x) = \left(\frac{r-1}{r}\right)^{(1+s)r/2} + \frac{1}{\kappa_r}x^{1-2/r} - x$$

is strictly concave on $[0,\infty)$, positive at $x = 0$ and $h(1) = 0$ (by the definition of κ_r). As a consequence, $h(x) \geq 0$ implies $x \leq 1$, whence (note that $\kappa_2/2 \leq 1$)

$$l_r^2 \leq c_r\|\Gamma_+(f)\|_{r/2}$$

which proves the induction step and demonstrates (2.3.1) for bounded functions f.

Let us now remove the boundedness assumption. If $f\colon \mathcal{X} \to \mathbb{R}$ is an arbitrary function with $\|\Gamma_+(f)\|_r < \infty$, then $\mathcal{E}(f,f) = \|\Gamma_+(f)\|_1 < \infty$ and as a consequence, by the Poincaré inequality, we obtain $\mu(|f|) < \infty$ and

$$\mathrm{Var}_\mu(f) \leq \frac{1}{a}\mathcal{E}(f,f).$$

In particular, defining $f_M = \max(\min(f,M),-M)$ we obtain $f_M \to f$ pointwise and $\mu(f_M) \to \mu(f)$ as $M \to \infty$. Applying (2.3.1) to f_M we obtain

$$\|(f_M - \mu(f_M))_+\|_r^2 \leq (1 - 2^{-(s+1)})\frac{r^{s+1}}{a}\kappa(s)\|\Gamma_+(f_M)\|_{r/2}. \tag{2.3.13}$$

However,

$$\begin{aligned}\Gamma_+(f_M)(x) &= \int_{\mathcal{X}} (f_M(x) - f_M(y))_+^2 Q_x(dy) \\ &\leq \int_{\mathcal{X}} (f(x) - f(y))_+^2 Q_x(dy) = \Gamma_+(f)(x).\end{aligned}$$

Therefore, Fatou's lemma implies that (2.3.1) for f follows from (2.3.13) by letting $M \to \infty$.

The inequality (2.3.2) follows by (2.3.1) applied to $-f$. □

Proof of Proposition 2.3.3. The general scheme of the proof is analogous as in the case of Proposition 2.3.1, one just needs to appropriately replace the pointwise estimates with the kernel Q_x by the abstract assumptions. Therefore instead of writing the complete proof we will just explain how to modify the arguments leading to (2.3.1).

We again prove by induction that for all positive integers k, $r \in (k, k+1]$, and bounded $f \in \mathcal{A}$,

$$\|f - \mu(f)\|_r^2 \leq c_r \|\Gamma(f)\|_{\max(r/2,1)}$$

with $c_r = \kappa_r(s) r^{s+1}/a \geq 1/a$. The quantity $\kappa_r(s)$ is defined as in (2.3.7), note however the difference between the definition of c_r in this proof and therein.

For $k = 1$, this follows analogously as in (2.3.8), by ignoring the first inequality and using $\mathcal{E}(f,f) = \|\Gamma(f)\|_1$ in the last estimate (note that finiteness of $\|\Gamma(f)\|_1$ implies that $\mu(f)$ is well-defined).

As for the induction step, we consider $g = f - \mu(f)$ and $\gamma = \frac{1}{p-1}$ where $r = \frac{p}{p-1}$. By Assumption 1, $g, |g| \in \mathrm{Dom}(\mathcal{E})$.

Assume that $\|g\|_\infty = M$ and observe that for $a, b \in [-M, M]$,

$$\big| |a|^\gamma - |b|^\gamma \big| \leq \gamma M^{\gamma-1} |a - b|.$$

Therefore, again by Assumption 1, $|g|^\gamma \in \mathrm{Dom}(\mathcal{E})$.

Applying thus (Bec-p) with parameter p to $|g|^\gamma$ we obtain

$$\alpha_p \big(\mu(|g|^{\gamma p}) - \mu(|g|^\gamma)^p \big) \leq \frac{p}{2} \mathcal{E}(|g|^\gamma, |g|). \tag{2.3.14}$$

Now, by the first part of Assumption 2 applied with $t = \gamma p/2 = r/2$ together with the equality $\frac{t}{t-1} = \gamma p/(\gamma p - 2) = \gamma p/(\gamma - 1)$, we get

$$\alpha_p \big(\mu(|g|^{\gamma p}) - \mu(|g|^\gamma)^p \big) \leq \gamma p (\mu(|g|^{\gamma p}))^{\frac{\gamma p - 2}{\gamma p}} \|\Gamma(f)\|_{\gamma p/2}.$$

The last inequality is a direct analog of (2.3.10), the difference being just the lack of the factor $1/2$ on the right-hand side.

The rest of the induction step is the same as in the proof of Proposition 2.3.1, leading to

$$\|f - \mu(f)\|_r \leq c_r \|\Gamma(f)\|_{r/2}$$

for bounded $f \in \mathcal{A}$, the only difference being the lack of the factor $\kappa_2/2$ in the counterpart of (2.3.12), which in the proof of Proposition 2.3.1 was estimated from above by one.

The extension to general $f \in \mathcal{A}$ follows easily by approximation from the second part of Assumption 2. □

2.4 Applications

We will now present applications of our results to various stochastic models in which modified log-Sobolev inequalities or Beckner inequalities are proven. Our main goal is to obtain new moment inequalities and derive from them concentration estimates.

2.4.1 The continuous setting

As already mentioned in Section 2.1, in the diffusive case, when the chain rule is satisfied, there is equivalence between the modified log-Sobolev inequality (mLSI) and the usual log-Sobolev inequality (LSI) as well as between the two forms (Bec-p) and (Bec'-q) of Beckner's inequality. Therefore, as explained in Section 2.1.4 the equivalence between the log-Sobolev inequalities and Beckner inequalities has been known in this case. Nevertheless, the equivalence of (Bec-p) and (Bec'-q) as well as some known examples of measures satisfying (Bec'-q) allow us to obtain moment estimates in L_r with optimal rate of dependence on r as $r \to \infty$ in several situations of interest.

We will start with a result proved by Wang [200, Corollary 1.3] in the setting of Riemannian manifolds.

Proposition 2.4.1. *Let E be a d-dimensional non-compact connected complete Riemannian manifold with Ricci curvature bounded from below. Let $\rho(x)$ be the Riemannian distance between x and a fixed point o. Consider $\mu(dx) := Ze^V dx$, where V is a continuous function on E such that $V + \theta\rho^\gamma$ is bounded for some $\gamma \in (1, 2]$ and $\theta > 0$, dx stands for the Riemannian volume measure, and Z is the normalization. Let $\mathcal{E}(f, f) := \mu(|\nabla f|^2)$ with $\mathrm{Dom}(\mathcal{E}) = H^{1,2}(\mu)$. Then there exists $\beta > 0$ such that (Bec'-q) holds for all $q \in [1, 2)$ with $\beta_q \geq \beta(2 - q)^{2/\gamma - 1}$.*

As a consequence, by Proposition 2.3.3 applied to $\mathcal{A}$ being the class of smooth compactly supported functions, followed by standard approximation techniques, we obtain

Corollary 2.4.2. *In the setting of Proposition 2.4.1, there exists a constant C, depending only on β, such that for any smooth function $f : M \to \mathbb{R}$ and all $r \geq 2$,*

$$\|f - \mu(f)\|_r \leq C r^{1/\gamma} \|\nabla f\|_r, \tag{2.4.1}$$

where the norms are taken in $L_r(E, \mu)$.

We remark that the example of measures μ_γ ($\gamma \in [1, 2]$) on $\mathbb{R}$ with density $c_\gamma \exp(-|x|^\gamma)$ (as investigated by Latała and Oleszkiewicz in [136], who proved that in this case $\beta > 0$ can be taken to be a universal constant) shows that the exponent $1/\gamma$ in the above corollary is optimal for $d = 1$. More generally, by tensorization it follows that (Bec'-q) with universal $\beta > 0$ is satisfied by the measures on $\mathbb{R}^d$ with densities $c_\gamma^d \exp(-\sum_{i=1}^d |x_i|^\gamma)$. In particular, we obtain (2.4.1) with constant C independent of γ and d, yielding the optimality of the exponent $1/\gamma$ in the case of general d.

We also note that in the case of $\mathbb{R}^d$, moment estimates of the form (2.4.1) for $\gamma \in (1, 2)$ can be derived from a combination of recent result [27] and [8] (the case $\gamma = 2$ corresponds to results by Aida–Stroock, the case $\gamma = 1$ can be found in [156]). The former article establishes an implication between Beckner inequalities (Bec'-q) and certain log-Sobolev inequalities with modified energy form (introduced by Gentil et al. in [94]), which are shown in [8] to imply (2.4.1). However, in both of the said implications, additional dependence on γ is introduced, and the constants explode for $\gamma \to 1$. To the best of our knowledge, the above corollary is new even in the case of measures μ_γ, $\gamma \in (1, 2)$.

Another example we would like to discuss concerns weighted inequalities for heavy tailed measures. We will focus on the Cauchy measure, defined on $\mathbb{R}^n$ as

$$\nu_{n,b}(dx) = \frac{1}{Z(1 + |x|^2)^b}$$

for $b > n/2$ (where Z is a normalizing constant).

Being heavy-tailed, the measure $\nu_{n,b}$ cannot satisfy the usual functional inequalities of Definition 2.1.3 with $\mathcal{E}(f,g) = \mu(\langle \nabla f, \nabla g\rangle)$. Nevertheless, as shown in [183, 35, 44, 161], for $b \geq n+1$, $\nu_{n,b}$ satisfies the following weighted Poincaré inequality for smooth functions:

$$\mathrm{Var}_{\nu_{n,b}}(f) \leq \frac{1}{2(b-1)} \int_{\mathbb{R}^n} |\nabla f(x)|^2 (1+|x|^2)\nu_{n,b}(dx).$$

Moreover, the weight $1+|x|^2$ is of optimal growth at infinity and the constant $\frac{1}{2(b-1)}$ is optimal. We remark that the weighted Poincaré inequality (without optimal constant) was known in a larger range of parameters (see, e.g., [41]) and a recent result [123] provides the optimal constants for the whole range of parameters. In what follows we however restrict to $b \geq n+1$, as we are going to use recent Beckner inequalities obtained under this assumption.

In [56] the above inequality has been complemented by a weighted log-Sobolev inequality

$$\mathrm{Ent}_{\nu_{n,b}}(f^2) \leq C_{n,b} \int_{\mathbb{R}^n} |\nabla f(x)|^2 (1+|x|^2)\log(e+|x|^2)\nu_{n,b}(dx),$$

where $C_{n,b}$ is a constant, depending only on n, b. Again, the growth of the weight is optimal at infinity (the result was earlier proved with a weight of faster growth in [41]).

By known approaches to moment estimates, related to (2.3.6) and (2.3.5) the above results provide for $r \geq 2$ bounds of the form

$$\|f - \nu_{n,b}(f)\|_{L_r(\nu_{n,b})} \leq \frac{C}{\sqrt{b-1}} r \|\sqrt{\omega_1}\nabla f\|_{L_r(\nu_{n,b})}, \tag{2.4.2}$$

where $\omega_1(x) = 1+|x|^2$, and C is a universal constant, and

$$\|f - \nu_{n,b}(f)\|_{L_r(\nu_{n,b})} \leq \sqrt{C_{n,b}(r-3/2)} \|\sqrt{\omega_2}\nabla f\|_{L_r(\nu_{n,b})}, \tag{2.4.3}$$

where $\omega_2(x) = (1+|x|^2)\log(e+|x|^2)$. See [41, 56] where similar moment inequalities were considered for Lipschitz functions. It is easy to see that (2.4.2) and (2.4.3) are not comparable. The latter has better dependence on r, the former may perform better if the function is supported far from the origin.

Recently, Bakry, Gentil, and Scheffer [23] proved that for $q \in [1, 2 - \frac{2}{b-n+1}]$, the measure $\nu_{n,b}$ satisfies a weighted Beckner inequality

$$2(b-1)\Big(\nu_{n,b}(f^2) - \nu_{n,b}(f^q)^{2/q}\Big) \leq (2-q)\int_{\mathbb{R}^n} |\nabla f(x)|^2(1+|x|^2)\nu_{n,b}(dx). \tag{2.4.4}$$

Interpreting $\mathcal{E}(f,g) = \int_{\mathbb{R}^n} \langle \nabla f, \nabla g\rangle \omega_1 d\nu_{n,b}$ as a Dirichlet form related to the diffusion with generator $Lf = \omega_1 \Delta f + \langle \nabla \omega_1 - \omega_1 \nabla V, \nabla f\rangle$, where $V = -\log(\frac{\nu_{n,b}(dx)}{dx})$, and using the relation between the inequalities (Bec'-q) and (Bec-p) discussed in Section 2.1.4 we see that for all $p \in [1+\frac{1}{b-n}, 2]$,

$$2(b-1)(\nu_{n,b}(f^p) - \nu_{n,b}(f)^p) \leq \mathcal{E}(f, f^{p-1}).$$

Note that this inequality cannot be satisfied for all $p \in (1,2]$ with a uniform constant, since this would contradict the optimal growth of weight w_2 for the log-Sobolev inequality. Using thus Remark 2.3.6 with $\Gamma(f,g) = \langle \nabla f, \nabla g\rangle \omega_1$, we obtain the following

Corollary 2.4.3. *If $b \geq n+1$ then for any smooth function $f\colon \mathbb{R}^n \to \mathbb{R}$, and $r \in [2, b-n+1]$,*

$$\|f - \nu_{n,b}(f)\|_{L_r(\nu_{n,b})} \leq \frac{C}{\sqrt{(b-1)}} \sqrt{r} \|\sqrt{\omega_1} \nabla f\|_{L_r(\nu_{n,b})}, \tag{2.4.5}$$

where C is a universal constant.

The above corollary complements the inequalities (2.4.2) and (2.4.3), improving on some of their aspects in the situation when b is substantially larger than n and r is large, as it provides better dependence on r than (2.4.2) and at the same is based on the weight ω_1 which is smaller than ω_2 used in (2.4.3). However, in the case of fixed b the range of r for which the estimate holds is restricted. We remark that weighted Beckner inequalities for more general heavy tailed convex measures have been recently obtained in [162]. They have been also generalized to the manifold setting in [95]. In all these cases one can derive similar weighted moment inequalities, we chose the case of the Cauchy measure to simplify the exposition.

2.4.2 Product spaces

The Herbst argument, which is now the main tool for deriving concentration estimates from log-Sobolev type inequalities, appeared for the first time in the 1970s in an unpublished letter from I. Herbst to L. Gross. In the mid 1990s in the seminal paper [137] Ledoux demonstrated the strength of this argument in the context of concentration inequalities in product spaces, recovering many inequalities by Talagrand, obtained by a more difficult inductive approach based on appropriate notions of isoperimetry [193, 192]. Since then the method was further developed by many authors, most notably by Boucheron, Bousquet, Lugosi, and Massart. Massart [150] and Boucheron, Lugosi, and Massart [49, 50] developed many modified log-Sobolev inequalities for product spaces, which were applied to a variety of problems, ranging from information theory through combinatorics to statistics and probability in Banach spaces. In a subsequent paper with Bousquet [47] they also established moment estimates, which are a direct inspiration for our Proposition 2.3.1. For this purpose they developed Beckner inequalities of the form (Bec-p) in product spaces, by deriving first tensorization properties for ϕ-entropies (present also in the work [136] by Latała and Oleszkiewicz) and then establishing one dimensional cases of (Bec-p) (thus proceeding in a manner parallel to the derivation of the modified log-Sobolev inequality in product spaces, based on tensorization properties of the usual entropy functional).

Our contribution in the context of product measures is an observation that thanks to the equivalence of (Bec-p) and (mLSI) with a mild change in constants, the Beckner's inequalities obtained in [47] can be derived directly from the most basic modified log-Sobolev inequality for product distributions. We would like to stress again that the subsequent derivation of moment inequalities that we present in Proposition 2.3.1 relies very heavily on the approach from [47].

For Reader's information and for comparison with the more general case of Glauber dynamics, discussed in the next section, we will now state some moment inequalities presented in [47] (we remark that this paper provides also other moment inequalities obtained under additional assumptions on the random variables in question).

Theorem 2.4.4 ([47, Theorem 2]). *Let $(X_1, \ldots, X_n)$ be independent random variables with values in a measurable space E and let $Z = f(X_1, \ldots, X_n)$ for some measurable function f. Let moreover $X_1', \ldots, X_n'$ be independent copies of $X_1, \ldots, X_n$ and set $Z_i = f(X_1, \ldots, X_{i-1}, X_i', X_{i+1}, \ldots, X_n)$.*

Then for $r \geq 2$,

$$\|(Z - \mathbb{E}\, Z)_+\|_r \leq \sqrt{\kappa r}\|\sqrt{V_+}\|_r, \tag{2.4.6}$$

where

$$V_+ = \mathbb{E}\left(\sum_{i=1}^n (Z - Z_i)_+^2 \middle| X_1, \ldots, X_n\right)$$

and $\kappa = \frac{\sqrt{e}}{\sqrt{e}-1}$.

We remark that for $r = 2$ this result recovers (up to constants) the Efron–Stein inequality for the variance.

Let us now relate the above theorem to our Theorem 2.2.1 and Proposition 2.3.1 and explain how they imply a version of the estimate (2.4.6). Denote by μ the distribution of the sequence $(X_1, \ldots, X_n)$ and observe that the quantity V_+ coincides with our $\Gamma_+(f)$ (recall (2.1.14)) for Γ given by

$$\Gamma(f, g) = \frac{1}{2}\int_{E^n} (f(y) - f(x))(g(y) - g(x))Q_x(dy),$$

for the kernel

$$Q_x(A) = \int_{E^n} \sum_{i=1}^n \mathbf{1}_A(x_1, \ldots, x_{i-1}, y_i, x_{i+1}, \ldots, x_n)\mu(dy),$$

reversible with respect to μ (as already mentioned this can be seen as a special case of Glauber dynamics).

The modified log-Sobolev inequality (mLSI) holds in this case simply due to Jensen's inequality and tensorization (the idea present already in the paper [137] by Ledoux). Indeed, for any random variable X, denoting by X' its independent copy, we have

$$\begin{aligned}
\mathrm{Ent}(f(X)) &\leq \mathbb{E}\, f(X)\log f(X) - \mathbb{E}\, f(X)\,\mathbb{E}\log f(X) \\
&= \mathbb{E}\, f(X)(\log f(X) - \log f(X')) \\
&= \frac{1}{2}\,\mathbb{E}(f(X) - f(X'))(\log f(X) - \log f(X')),
\end{aligned}$$

which, when combined with the well known tensorization property of entropy (see, e.g., [51, Theorem 4.10]) gives

$$\mathrm{Ent}_\mu(f) \leq \int \sum_{i=1}^n \mathrm{Ent}_{\mu_i}(f)d\mu,$$

for $\mu = \mu_1 \otimes \cdots \otimes \mu_n$ gives (mLSI) with $\rho_0 = 1$ (here Ent_{μ_i} denotes the entropy computed on a product space just with respect to the i-th coordinate and the measure μ_i, with the other coordinates fixed). By Theorem 2.2.1 this gives Beckner's inequality (Bec-p) with $\alpha_p \geq \frac{1}{6}$. Now, Proposition 2.3.1 applied with $a = 1/6$ and $s = 0$ gives (2.4.6) with $\kappa = 3\frac{\sqrt{e}}{\sqrt{e}-1}$, which is worse than Theorem 2.4.4 just by a factor $\sqrt{3}$.

2.4.3 Glauber dynamics

Let us now consider $\mathcal{X} = E^I$, where I is a finite set and E is a Polish space endowed with the Borel σ-field. Let μ be a probability measure on $\mathcal{X}$. For $x \in \mathcal{X}$ and $J \subseteq I$, let $x_J = (x_i)_{i\in J}$. Let also $X = (X_i)_{i\in I}$ be an $\mathcal{X}$-valued random variable distributed according to μ. Finally, for $i \in I$ let X_i' be an E-valued random variable such that its (regular) conditional distribution given X satisfies

$$\mu_i(\cdot|x) := \mathbb{P}(X_i' \in \cdot\ |X = x) = \mathbb{P}(X_i \in \cdot\ |X_{\{i\}^c} = x_{\{i\}^c}).$$

In other words, given $X_1, \dots, X_{i-1}, X_{i+1}, \dots, X_n$, the random variables X_i' and X_i are conditionally i.i.d.

Denote $X^i = (Y_j)_{j\in I}$ where $Y_j = X_j$ for $j \neq i$ and $Y_i = X_i'$ (i.e., X^i is obtained from X by replacing X_i with X_i'). The Glauber dynamics (known also as the Gibbs sampler or heat bath) is given by a generator of the form

$$Lf(x) = \sum_{i\in I} \int_E (f(x_1, \dots, x_{i-1}, y, x_{i+1}, \dots, x_n) - f(x))\mu_i(dy|x)$$

and corresponds to a càdlàg Markov process $(X(t))_{t\geq 0}$ in which at rate $|I|$ a coordinate $i \in I$ is chosen uniformly and $X_i(t-)$ is replaced with a value drawn from the distribution $\mu_i(\cdot|X(t-))$, while the remaining coordinates are kept intact.

Let us note that if μ is a product measure, then $\mu_i(\cdot|x)$ equals to the i-th marginal of μ (in particular is independent of x) and the situation reduces to the case described in the previous section with $I = [n] := \{1, \dots, n\}$. In the general case the generator and the carré du champ operator are given by the kernel

$$Q_x(A) = \int_{\mathcal{X}} \sum_{i\in I} \mathbf{1}_A(x_1, \dots, x_{i-1}, y_i, x_{i+1}, \dots, x_n)\mu_i(dy|x). \tag{2.4.7}$$

Plugging this kernel into formulas (2.1.9) and (2.1.14), and using the definition of the variables X^i, we obtain

$$\Gamma(f) = \frac{1}{2}\sum_{i\in I} \mathbb{E}((f(X) - f(X^i))^2|X),$$
$$\Gamma_+(f) = \sum_{i\in I} \mathbb{E}((f(X) - f(X^i))_+^2|X).$$

Therefore, a combination of Theorem 2.2.1, Proposition 2.3.1 and Proposition 2.3.3 gives in this setting the following corollary.

Corollary 2.4.5. *In the setting described above, if the Glauber dynamics satisfies the modified log-Sobolev inequality (mLSI) with constant ρ_0, then for $p \in (1, 2]$ it satisfies (Bec-p) with $\alpha_p \geq \rho_0/6$. Moreover, for every function $f : \mathcal{X} \to \mathbb{R}$ and $r \geq 2$,*

$$\begin{aligned}\|(f(X) - \mathbb{E} f(X))_+\|_r &\leq K\sqrt{r}\Big\|\Big(\sum_{i\in I} \mathbb{E}((f(X) - f(X^i))_+^2|X)\Big)^{1/2}\Big\|_r \\ &\leq K\sqrt{r}\Big\|\Big(\sum_{i\in I}(f(X) - f(X^i))_+^2\Big)^{1/2}\Big\|_r\end{aligned}$$

and

$$\begin{aligned}\|f(X)-\mathbb{E}\,f(X)\|_r &\le K\sqrt{r}\Big\|\Big(\sum_{i\in I}\mathbb{E}((f(X)-f(X^i))^2|X)\Big)^{1/2}\Big\|_r\\ &\le K\sqrt{r}\Big\|\Big(\sum_{i\in I}(f(X)-f(X^i))^2\Big)^{1/2}\Big\|_r,\end{aligned}$$

where $K=\sqrt{\frac{3\sqrt{e}}{\rho_0(\sqrt{e}-1)}}$.

Inequalities of this type have been recently derived for measures on finite product spaces [99] using the Aida–Stroock approach, based on the usual log-Sobolev inequality (LSI). This results in the constant K being a multiple of $\rho_1^{-1/2}$. However, in many cases (even if μ is a product measure on a finite set) the constant ρ_1 is much smaller than ρ_0. Moreover, as shown in [100] in the general case $\rho_1>0$ for the Glauber dynamics only if μ is finitely supported, which is in contrast to ρ_0 which, as stated in the previous section, is positive, e.g., for all product measures. Using a Holley–Stroock type perturbation argument (cf. [118, 14]) one can also easily produce examples of non-product measures with infinite support and $\rho_0>0$.

Several examples satisfying the modified log-Sobolev inequality (mLSI) and the log-Sobolev inequality (LSI) have been recently presented by Sambale and Sinulis in [178]. They are based on a general theorem concerning approximate tensorization of entropy under a Dobrushin type condition due to Marton [149] (see also [99]). Recently [71, 34] have considered a more general notion of block factorization, allowing, e.g., to obtain modified log-Sobolev inequalities for the Glauber dynamics on q-colorings of graphs with constant maximum degree.

Let us now state the general result of [178].

Assume that E is finite and define the Dobrushin matrix $A=(A_{ij})_{i,j\in I}$ as

$$A_{ij}=\sup_{x,y\in\mathcal{X}:\,x_{\{j\}^c}=y_{\{j\}^c}}\|\mathcal{L}(X_i|X_{\{i\}^c}=x_{\{i\}^c})-\mathcal{L}(X_i|X_{\{i\}^c}=y_{\{i\}^c})\|_{TV}$$

for $i\neq j$ and $A_{ii}=0$ (where $\|\cdot\|_{TV}$ denotes the total variation norm). Let $\alpha=1-\|A\|_{\ell_2\to\ell_2}$, where $\|A\|_{\ell_2\to\ell_2}$ is the operator norm of the matrix A. Define also for $J\subsetneq I$ and $i\notin J$

$$\beta_{i,J}=\inf_{\substack{x_J\in E^J,y_{J^c}\in E^{J^c}\\ (x_J,y_{J^c})\in\operatorname{supp}(\mu)}}\mathbb{P}(X_i=(y_{J^c})_i|X_J=x_J)$$

(for $J=\emptyset$ we understand the above simply as $\inf_{y\in\operatorname{supp}(\mu)}\mathbb{P}(X_i=y_i)$).

Finally, set

$$\beta=\inf_{J\subsetneq I}\inf_{i\notin J}\beta_{i,J}.$$

Theorem 2.4.6 ([178, Theorem 4.1])**.** *If* $\alpha,\beta>0$, *then*

$$\rho_0\ge\alpha^2\beta,\quad \rho_1\ge\frac{\log(2)\alpha^2\beta}{2\log(\beta^{-1})}.$$

Recall that here ρ_0 is the constant on the left-hand side in the modified log-Sobolev (mLSI) and note that due to a different normalization of the Dirichlet form and a different convention concerning constants in [178] our parameter ρ_0 corresponds to $2|I|/\rho_0$ therein. Using Theorem 2.2.1 we immediately obtain

Corollary 2.4.7. *If $\alpha, \beta > 0$, then for $p \in (1,2]$ the inequality (Bec-p) holds with $\alpha_p \geq \frac{\alpha^2\beta}{6}$.*

Sambale and Sinulis apply Theorem 2.4.6 to several stochastic models, including exponential random graphs, random graph colorings, hardcore model. In an earlier paper [100] with Götze they also treat the Ising model. They are primarily interested in situations when for a family of models on sets I_n with $|I_n| \to \infty$ the constants ρ_0, ρ_1 are uniformly separated from zero. From this point of view the sufficient conditions that can be obtained from Theorem 2.4.6 are the same for both constants. If one is however interested in a more quantitative analysis, and looks at the dependence of the constants on the parameters of the model, then typically ρ_0 is of smaller order than ρ_1 (as β becomes small). In particular, a combination of Corollary 2.4.5 with estimates on ρ_0 given in Theorem 2.4.6 indeed gives better dependence of constants in moment inequalities than those derived from the Aida–Stroock approach based on ρ_1.

Below we discuss this in more detail for selected classical stochastic models.

Exponential random graphs

Let $I_n = \{(i,j) \in [n]^2 : i < j\}$ and identify elements of $\mathcal{G}_n = \{0,1\}^{I_n}$ with simple graphs on n-vertices in a natural way. For $\gamma = (\gamma_1, \ldots, \gamma_s) \in \mathbb{R}^s$ and simple connected graphs $G_i = (V_i, E_i)$, $i = 1, \ldots, s$, let μ_γ be a probability measure on $\mathcal{G}_n$ given by the weight of the form

$$\exp(-H_\gamma(x)), \quad x \in \mathcal{G}_n,$$

with

$$H_\gamma(x) = n^2 \sum_{i=1}^{s} \gamma_i \frac{N_{G_i}(x)}{n^{|V_i|}},$$

where for simple graphs $H = (V_H, E_H)$ and $G = (V_G, E_G)$, $N_H(G)$ is the number of homomorphisms from H to G, i.e., injective maps $i : V_H \to V_G$, which preserve edges. We assume (by convention) that G_1 is the complete graph on two vertices and that $|E_i| > 1$ for $i > 1$. Then, for $s = 1$, the measure μ_γ corresponds to the distribution of the Erdős–Rényi random graph $G(n,p)$ with $p = e^{\gamma_1}(1 + e^{\gamma_1})^{-1}$. The general model for arbitrary s and graphs G_i is known as the exponential random graph model and it has been an object of intensive studies in recent years, both from the theoretical point of view and in connection to applications, e.g., to modeling of social networks. We refer to [63] and [65] for a detailed presentation. Sambale and Sinulis provide a sufficient condition for the constants ρ_0 and ρ_1 to be separated from zero independently of the size n of the model. An inspection of their proof reveals that if

$$\delta := \frac{1}{2} \sum_{i=2}^{s} |\gamma_i||E_i|(|E_i| - 1) < 1,$$

then one can apply Theorem 2.4.6 with some $\alpha \geq 1 - \delta$ and $\beta \geq ce^{-2|\gamma_1|}$ for a universal constant $c > 0$. Thus, in this case the bounds on ρ_0 and ρ_1 differ by a factor of order $|\gamma_1|$ when $|\gamma_1| \to \infty$. It is an interesting question to verify if the constant ρ_0 indeed depends on γ_1, which corresponds to the Erdős–Rényi product-type behavior of the graph.

Ising model on finite sets

Let $I = \{1, \dots, n\}$ and consider the measure on $\mathcal{X} = \{+1, -1\}^n$ given by

$$\mu(\{\varepsilon\}) = \frac{1}{Z} \exp\Big(\frac{1}{2} \sum_{i,j=1}^{n} J_{ij}\varepsilon_i\varepsilon_j - \sum_{i=1}^{n} h_i\varepsilon_i\Big),$$

where $J = (J_{ij})_{i,j=1}^n$ is a symmetric matrix with vanishing diagonal and $h \in \mathbb{R}^n$. From the statistical physics point of view the matrix J correspond to interaction between spins, while h describes the external field. Concentration inequalities for the Ising model have been considered by many authors starting from the 1990s [148, 66], as it is arguably the most basic discrete model with dependencies. The interest in them has been recently revived in relation to algorithmic applications [79, 78, 96]. Estimates on ρ_1 given by Theorem 2.4.6 have been a starting point for inequalities obtained in [99, 4], with the Aida–Stroock approach playing a crucial role. Since each function of the discrete cube can be regarded as a polynomial, by considering its Fourier–Walsh expansion, it is natural to investigate concentration of measure in terms of characteristics related to the polynomial representation. In this case the dependence of the estimates from the said papers on the constant ρ_1 increases with the degree of the polynomial. Therefore, an application of Corollary 2.4.5 again allows to improve the behavior of inequalities in the asymptotic case. As for the parameters α, β of Theorem 2.4.6, an inspection of the calculations from [99] (cf. Lemma 3.1 therein) reveals that in this case they can be taken as

$$\alpha \geq 1 - \max_{i\leq n} \sum_{j\leq n} |J_{ij}|, \quad \beta \geq ce^{-\|h\|_\infty}. \tag{2.4.8}$$

Since the constants in the modified log-Sobolev inequalities do not depend on h in the product case $J = 0$, the same question as in the case of exponential random graphs seems natural also in this setting.

We note that in both cases it is not clear to us whether the above estimates on ρ_0 and ρ_1 can be improved in a general situation and what the true gap between the two constants is. Let us also point out that the gap in the estimates of Theorem 2.4.6 appears in the regime $\beta \to 0$ and is only logarithmic in $1/\beta$, while the dependence of the bounds for ρ_0 and ρ_1 on β is polynomial.

Let us remark that an approach to functional inequalities for the Ising model differing from the one based on the Dobrushin condition has been recently developed in a series of works by Bauerschmidt and Bodineau [28], Eldan et al. [85] and Anari et al. [12]. The conditions for functional inequalities are expressed in terms of the operator norm of the interaction matrix, without passing to absolute values of the coefficients, which allows treating cases which do not satisfy the Dobrushin condition, leading, e.g., to breakthroughs in the analysis of mixing times for the Sherrington–Kirkpatrick models of spin glasses (in particular [12] provides a modified log-Sobolev inequality). The conditions proposed in these articles are in general not comparable with the Dobrushin condition. Since our goal is just to illustrate potential application of our results we will not discuss these results in detail and only mention that they also can be combined with our estimates.

Hardcore model

We will conclude this section with another example of a classical stochastic model. In this case the model does demonstrate a gap between ρ_0 and ρ_1 and not just their known lower bounds. More specifically we will show a family of hardcore models on a growing sequence of graphs, for which ρ_0 remains separated from zero, while $\rho_1 \to 0$.

Let $G = (V, E)$ be a finite simple graph with maximum degree Δ and let $\eta > 0$ be a parameter. A binary function $\varepsilon = (\varepsilon_i)_{i\in V} \in \mathcal{X} := \{0,1\}^V$ will be called admissible if $\varepsilon_i\varepsilon_j = 0$ whenever $\{i,j\} \in E$. Thus, admissible functions describe allocations of particles on V in which one can have at most one particle per vertex and no two adjacent vertices can be occupied simultaneously. Let μ be a probability measure on $\mathcal{X}$ given by

$$\mu(\{\varepsilon\}) = \frac{1}{Z}\prod_{i\in V}\eta^{\varepsilon_i}\mathbf{1}_{\{\varepsilon \text{ is admissible}\}},$$

where Z is the normalization constant. Recently Conforti [72] obtained modified log-Sobolev inequalities and Beckner inequalities for this model. In particular, improving earlier estimates from [86], under the assumption $\eta\Delta < 1$ he proved that

$$\rho_0 := \rho_0(G,\eta) \geq \frac{1-\eta(\Delta-1)+2\min(\eta, 1-\eta\Delta)}{1+\eta} \tag{2.4.9}$$

(we remark that the Glauber dynamics considered by us is slowed down by a factor $1+\eta$ with respect to the one used in [72]). He obtained also general Φ-Sobolev inequalities, in particular (Bec-p). Estimates for ρ_0 and ρ_1 independent of $|V|$ have been also obtained in [178] by means of Theorem 2.4.6, under the assumption $\eta(\Delta-1) < 1$, however due to the dependence on the parameter β, they are of worse order.

Below we will provide a sequence of graphs $G_n = (V_n, E_n)$ with $|V_n| = n+1$, maximum degree $\Delta_n = n$, and such that $\rho_0 = \rho_0(G_n, 1/(2\Delta_n))$ is bounded away from zero, while $\rho_1 = \rho_1(G_n, 1/(2\Delta_n)) = O(\frac{1}{\log n})$. Let G_n be a star with center 0 and n rays, i.e., $V_n = \{0\}\cup[n]$, $E_n = \{\{0,i\}\colon i \in [n]\}$.

Note that the set of admissible ε's is composed of 2^n+1 elements: ε^* placing a single particle at zero and 2^n configurations with $\varepsilon_0 = 0$. Among them let us distinguish ε° such that $\varepsilon^\circ(0) = 0$, $\varepsilon^\circ(i) = 1$ for $i \in [n]$.

In particular, it follows from the above discussion that $Z = \eta + (1+\eta)^n$, $\mu(\{\varepsilon^*\}) = \frac{\eta}{Z}$, $\mu(\{\varepsilon^\circ\}) = \frac{\eta^n}{Z}$.

Let us test the inequality (LSI) with $f = \mathbf{1}_{\{\varepsilon^\circ\}}$. Denoting $p = \mu(\{\varepsilon^\circ\})$, we obtain

$$\operatorname{Ent} f = p\log(p^{-1}).$$

On the other hand

$$\mathcal{E}(\sqrt{f},\sqrt{f}) = \mathcal{E}(f,f) = \mathbb{E}\sum_{i=0}^{n}(f(X)-f(X^i))_+^2,$$

where X, X^i are defined at the beginning of this section.

Now, f is nonnegative and equal 0 on $\{\varepsilon^\circ\}^c$, therefore if $X \neq \varepsilon^\circ$, then

$$\sum_{i=0}^{n}(f(X)-f(X^i))_+^2 = 0.$$

On the other hand, if $X = \varepsilon^\circ$ then $X^0 = X$ and for $i \neq 0$, $X^i \neq X$ with conditional probability $\frac{1}{1+\eta}$. Thus,

$$\mathcal{E}(\sqrt{f},\sqrt{f}) = \frac{pn}{1+\eta}.$$

This shows that $\rho_1 \leq \frac{n}{(1+\eta)\log(p^{-1})}$. Since $p^{-1} = \frac{\eta+(1+\eta)^n}{\eta^n} \geq \frac{1}{\eta^n}$, we obtain

$$\rho_1 \leq \frac{1}{1+\eta} \frac{1}{\log(\eta^{-1})}.$$

In particular for $\eta = \frac{1}{2n}$, we get $\rho_1 = O\Big(\frac{1}{\log n}\Big)$, whereas by (2.4.9) $\rho_0 \geq c$ for some $c > 0$, independent of n.

2.4.4 The symmetric group

General moment estimates

Consider the symmetric group S_n of permutations of the set $[n]$ equipped with the uniform probability measure π_n. We will view this measure as the stationary distribution for the *interchange process*. Recall that this process describes the dynamics of n particles, labeled by the set $[n]$ which occupy n distinct sites (also labelled by $[n]$). At rate one a randomly chosen pair of particles exchange their positions. Let L be the infinitesimal operator for this process, i.e.,

$$\begin{aligned} Lf(\sigma) &= \frac{1}{n(n-1)} \sum_{i,j=1}^{n} \Big(f(\sigma \circ \tau_{ij}) - f(\sigma)\Big) \\ &= \frac{2}{n(n-1)} \sum_{1 \leq i < j \leq n} \Big(f(\sigma \circ \tau_{ij}) - f(\sigma)\Big), \end{aligned}$$

where τ_{ij} stands for the transposition of elements i and j. The corresponding Dirichlet form is

$$\begin{aligned} \mathcal{E}(f,g) &= \frac{1}{2n(n-1)n!} \sum_{\sigma \in S_n} \sum_{i,j=1}^{n} \Big(f(\sigma \circ \tau_{ij}) - f(\sigma)\Big)\Big(g(\sigma \circ \tau_{ij}) - g(\sigma)\Big) \\ &= \frac{1}{n(n-1)n!} \sum_{\sigma \in S_n} \sum_{1 \leq i < j \leq n} \Big(f(\sigma \circ \tau_{ij}) - f(\sigma)\Big)\Big(g(\sigma \circ \tau_{ij}) - g(\sigma)\Big). \end{aligned}$$

The modified log-Sobolev inequality for this process with $\rho_0 \geq \frac{1}{n-1}$ was obtained independently by Gao–Quastel [91] and Bobkov–Tetali [42], who also obtained the Beckner inequality (Bec-p) with $\alpha_p = \frac{p(n+2)}{2n(n-1)}$ (we note that the normalization of the generator L differs across various references, we provide here scaled constants matching our setting). The Poincaré constant was computed earlier by Diaconis and Shahshahani [82]. These results can be considered another example demonstrating that the behavior of constants in Poincaré, modified log-Sobolev or Beckner inequalities can be much better than of the constant in the classical log-Sobolev inequality, which was proved by Lee and Yau [140] to be of order $\frac{1}{n \log n}$. In a recent work Götze–Sambale–Sinulis [99] used the result from [140] in combination with the Aida–Stroock approach to obtain certain tail estimates on the symmetric group. However, the constants in their estimates explode as $n \to \infty$.

As a consequence of Beckner inequalities we obtain the following moment estimate for functions on the symmetric group. As explained below it generalizes to arbitrary functions the moment bound obtained by Chatterjee [62] for a special class of variables known as Hoeffding statistics.

Proposition 2.4.8. *Let σ be a uniform random permutation of the set $[n]$. For an arbitrary function $f\colon S_n \to \mathbb{R}$ and any $r \geq 2$,*

$$\|f(\sigma) - \mathbb{E}\, f(\sigma)\|_r \leq D_{2.4.8}\sqrt{r}\Big\|\Big(\frac{1}{n+2}\sum_{i,j=1}^{n}(f(\sigma) - f(\sigma\circ\tau_{ij}))^2\Big)^{1/2}\Big\|_r \tag{2.4.10}$$

and

$$\|(f(\sigma) - \mathbb{E}\, f(\sigma))_+\|_r \leq D_{2.4.8}\sqrt{r}\Big\|\Big(\frac{1}{n+2}\sum_{i,j=1}^{n}(f(\sigma) - f(\sigma\circ\tau_{ij}))_+^2\Big)^{1/2}\Big\|_r, \tag{2.4.11}$$

where $D_{2.4.8} = \sqrt{\frac{\sqrt{e}}{\sqrt{e}-1}}$.

Proof. We have

$$\Gamma(f)(\sigma) = \frac{1}{2}Lf^2(\sigma) - f(\sigma)Lf(\sigma) = \frac{1}{2n(n-1)}\sum_{i,j=1}^{n}\big(f(\sigma) - f(\sigma\circ\tau_{ij})\big)^2.$$

The assertion follows by the aforementioned result of Bobkov–Tetali and Propositions 2.3.1 and 2.3.3 (with $s = 0$ and $a = \inf_{p\in(1,2]}\frac{p(n+2)}{2n(n-1)} = \frac{(n+2)}{2n(n-1)}$). □

Hoeffding statistics

In the special case of Hoeffding statistics, i.e., functions of the form

$$f(\sigma) = \sum_{k=1}^{n} a_{k\sigma(k)}, \tag{2.4.12}$$

where $(a_{ij})_{i,j=1}^n$ is a real matrix, the inequality (2.4.10) was proved for integer r by Chatterjee [62] (with slightly different constants), who also obtained a Bernstein type inequality for matrices with positive bounded entries. Since then, concentration of measure for Hoeffding statistics has been investigated, e.g., by Albert [11] and Bercu–Delyon–Rio [30]. They obtained Bernstein type estimates for general bounded entries. The methods used in these references are quite diverse: while Chatterjee uses Stein's method, Albert relies on Talagrand's convex distance inequality on the symmetric group [192] and Bercu–Delyon–Rio on martingale methods (used for the first time in the context of random permutations by Maurey [151]).

Let us mention that Hoeffding statistics have been widely studied in the literature, starting from the article [116] of Hoeffding himself who obtained their asymptotic normality under certain assumptions (a result known as combinatorial CLT). They are important since they include many functions of interest in combinatorics or non-parametric statistics. In particular, it is easy to see that one can encode in the form (2.4.12) sums of functions of samples without replacement from a finite population.

Below we will use the second inequality of Proposition 2.4.8 together with an approach of Boucheron–Bousquet–Lugosi–Massart [47] to obtain an inequality for suprema of Hoeffding statistics. In the special case of sampling without replacement this inequality will improve certain aspects of an estimate obtained by Tolstikhin–Blanchard–Kloft [195]. Their main motivation were applications to transductive learning, we believe that bounds of this type may be also useful in the context of bootstrap for empirical processes. Note that in this section we focus on the moment estimates. We will revisit Hoeffding statistics in Chapter 5, where we prove Bennet-type estimates.

Proposition 2.4.9. *Let $\mathcal{A}$ be a collection of $n \times n$ matrices and let σ be a uniform random permutation of the set $[n]$. Define the random variable*

$$Z = \sup_{a \in \mathcal{A}} \sum_{k=1}^{n} a_{k\sigma(k)}.$$

Then for any $r \geq 2$,

$$\|(Z - \mathbb{E}\, Z)_+\|_r \leq 4D_{2.4.8}\sqrt{r}A + 10D_{2.4.8}^2 r B_r, \tag{2.4.13}$$

where $A = \mathbb{E}\sup_{a\in\mathcal{A}} \sqrt{\sum_{k=1}^{n} a_{k\sigma(k)}^2}$, $B_r = \big\| \max_{k\leq n} \sup_{a\in\mathcal{A}} |a_{k\sigma(k)}| \big\|_r$, and $D_{2.4.8}$ is the constant from Proposition 2.4.8. As a consequence, for any $r \geq 0$,

$$\mathbb{P}\left(Z \geq \mathbb{E}\, Z + 4eD_{2.4.8}\sqrt{r}A + 10eD_{2.4.8}^2 r B_r\right) \leq e^{2-r}. \tag{2.4.14}$$

Before we prove the above proposition, we will provide two examples of applications, comparing it with the results mentioned above.

Example 2.4.10. If $\mathcal{A}$ consists of a single element and one does not pay attention to universal constants, then inequality (2.4.14) is a strengthening of the results by Bercu–Delyon–Rio [30] and Albert [11]. Their results give

$$\mathbb{P}\left(Z \geq \mathbb{E}\, Z + K\Big(\sqrt{r}\Big(\frac{1}{n}\sum_{ij=1}^{n} a_{ij}^2\Big)^{1/2} + r\max_{i,j\leq n}|a_{ij}|\Big)\right) \leq 2e^{-r} \tag{2.4.15}$$

for a certain universal constant K. The parameters A and B_r of Proposition 2.4.9 read in this case as

$$A = \mathbb{E}\sqrt{\sum_{k=1}^{n} a_{k\sigma(k)}^2}, \quad B_r = \big\| \max_{k\leq n} |a_{k\sigma(k)}| \big\|_r.$$

Clearly $\max_{ij}|a_{ij}| \geq B_r$, moreover in certain situation $\max_{ij}|a_{ij}|$ may be significantly greater than the r-th moment B_r (this happens when there are few large elements in the matrix (a_{ij}) and r is not too large).

By Jensen's inequality we also have $(n^{-1}\sum_{ij=1}^{n} a_{ij}^2)^{1/2} \geq A$, but in fact the difference of these two quantities is at most of the order $\|\max_k |a_{k\sigma(k)}|\|_2$ (see the proof of Proposition 2.4.9 below), so it can be absorbed in the coefficient in front of r. Hence, (2.4.14) does not improve on the subgaussian coefficient of (2.4.15). This is not surprising, since (as observed in [30]) if one defines

$$d_{ij} = a_{ij} - n^{-1}\sum_{l=1}^{n} a_{il} - n^{-1}\sum_{l=1}^{n} a_{lj} + n^{-2}\sum_{l,m=1}^{n} a_{lm},$$

then $\sum_{k=1}^{n} d_{k\sigma(k)} = Z - \mathbb{E}\, Z$ and $n^{-1}\sum_{ij=1}^{n} d_{ij}^2 = \mathrm{Var}(Z)$.

To summarize, the main advantage of Proposition 2.4.9 over (2.4.15) is the fact that $\max_{i,j\leq n}|a_{ij}|$ can be replaced by a smaller parameter B_r.

Example 2.4.11. Let us now specialize to the setting of sampling without replacement and compare our result with Tolstikhin–Blanchard–Kloft [195]. To this end we will need to rephrase both results in the same notation. Let us consider a set of vectors $\mathcal{X} \subseteq \{x \in \mathbb{R}^n \colon x_1 + \ldots + x_n = 0\}$. For $m \leq n$ let

$I_1, \ldots, I_m$ be a uniform sample without replacement and $J_1, \ldots, J_m$ a sample with replacements from the set $[n]$. Define

$$Z = \sup_{x \in \mathcal{X}} \sum_{k=1}^{m} x_{I_k}, \qquad Z' = \sup_{x \in \mathcal{X}} \sum_{k=1}^{m} x_{J_k}.$$

Thus, Z' can be considered a supremum of the empirical process in independent random variables J_k. The tails of such suprema have been thoroughly studied, beginning with the seminal work by Talagrand [193], who obtained Bernstein and Bennett type inequalities. The authors of [195] combined optimal forms of such inequalities proved by Bousquet [52] with a stochastic domination between Z and Z' (due to Hoeffding) to derive a bound of the form

$$\mathbb{P}\Big(Z \geq \mathbb{E}\, Z' + \sqrt{2vr} + \frac{r}{3} \sup_{x \in \mathcal{X}} \|x\|_\infty\Big) \leq e^{-r} \tag{2.4.16}$$

for $r \geq 0$, where $v = m \sup_{x \in \mathcal{X}} \operatorname{Var}(x_{J_1}) + 2 \sup_{x \in \mathcal{X}} \|x\|_\infty \mathbb{E}\, Z'$.

One can easily see that the variable Z corresponds to the supremum of Hoeffding statistics over matrices given by $a^x_{ij} = x_j$ for $i \leq m$, and $a^x_{ij} = 0$ for $i > m$. Therefore, Proposition 2.4.9 yields

$$\mathbb{P}\Big(Z \geq \mathbb{E}\, Z + 4eD_{2.4.8}\sqrt{r}A + 10eD^2_{2.4.8} r B_r\Big) \leq e^{2-r} \tag{2.4.17}$$

with

$$A = \mathbb{E} \sup_{x \in \mathcal{X}} \Big(\sum_{k=1}^{m} x^2_{I_k}\Big)^{1/2}, \quad B_r = \Big\| \sup_{x \in \mathcal{X}} \max_{k \leq m} |x_{I_k}| \Big\|_r.$$

Again, in certain situations, especially for relatively small values of r, the quantity B_r may be of smaller order than $\sup_{x \in \mathcal{X}} \|x\|_\infty$ used in (2.4.16). However, the main difference between the two estimates is the fact that (2.4.17) provides deviation above $\mathbb{E}\, Z$, while (2.4.16) considers deviations above $\mathbb{E}\, Z'$, which always exceeds $\mathbb{E}\, Z$ (see the inequality (2.4.18) below) and in certain situation can be significantly larger. The authors of [195] provide a bound

$$\mathbb{E}\, Z' - \mathbb{E}\, Z \leq 2\frac{m^3}{n} \sup_{x \in \mathcal{X}} \|x\|_\infty.$$

Specializing to the case $\mathcal{X} \subseteq [-1, 1]^n$, it follows from the above estimate that if one is interested in a bound on $Z - \mathbb{E}\, Z$ which is of the order $\sqrt{m}$ (corresponding to the CLT type rates one would like to obtain in statistical applications), the inequality (2.4.16) is applicable for $m = O(n^{2/5})$. Note that $\mathbb{E}\, Z' \leq m$, so the quantity $\sqrt{v}$ is of the right order $\sqrt{m}$. On the other hand $\mathbb{E} \sup_{x \in \mathcal{X}} \Big(\sum_{k=1}^{m} x^2_{I_k}\Big)^{1/2}$ also equals at most $\sqrt{m}$, so (2.4.17) provides a bound on $Z - \mathbb{E}\, Z$ of the order $\sqrt{m}$ without any restrictions on m (we remark that the interesting case is $m \leq n/2$ since thanks to the mean zero assumption one can always pass from m to $n - m$).

Let us now discuss in more detail the subgaussian coefficients of the two inequalities. As pointed out in [145, 107] it follows from an argument due to Hoeffding [117] that if E is a normed space and $f \colon [n] \to E$, then for any convex function $\Psi \colon E \to \mathbb{R}$,

$$\mathbb{E}\, \Psi\Big(\sum_{k=1}^{m} f(I_k)\Big) \leq \mathbb{E}\, \Psi\Big(\sum_{k=1}^{m} f(J_k)\Big). \tag{2.4.18}$$

In particular this implies that

$$A^2 \le \mathbb{E} \sup_{x\in\mathcal{X}} \sum_{k=1}^m x_{I_k}^2 \le \mathbb{E} \sup_{x\in\mathcal{X}} \sum_{k=1}^m x_{J_k}^2 \le m \sup_{x\in\mathcal{X}} \mathrm{Var}(x_{J_1}) + 8 \sup_{x\in\mathcal{X}} \|x\|_\infty \,\mathbb{E}\, Z'',$$

with $Z'' = \sup_{x\in\mathcal{X}} \sum_{k=1}^m \varepsilon_k x_{J_k}$, where $\varepsilon_1, \ldots, \varepsilon_m$ are i.i.d. Rademacher's variables independent of $J_1, \ldots, J_m$. The last inequality is a classical result of the theory of empirical processes (see, e.g., [188, 150, 47]) based on symmetrization and Talagrand's contraction principle for Rademacher averages [139]. If the set $\mathcal{X}$ is symmetric with respect to the origin, one can further write $\mathbb{E}\, Z'' \le 2\,\mathbb{E}\, Z'$. Thus, in this case the subgaussian coefficient of (2.4.17) is up to absolute constants dominated by $\sqrt{v}$ used in (2.4.16). Let us note that using results from [47] one can also provide a similar bound on the subgaussian coefficient of (2.4.17) with $\|\max_{i\le m} \sup_{x\in\mathcal{X}} |x_{J_i}|\|_2$ in place of $\sup_{x\in\mathcal{X}} \|x\|_\infty$. Since our goal is rather to illustrate Proposition 2.4.8 than to provide the most general estimate, we skip the details.

The above discussion shows that our estimate (2.4.17) may give better bounds than (2.4.16). On the other hand (2.4.16) has better constants, in particular provides the optimal constant $\sqrt{2}$ in the subgaussian part. Let us remark that [195] contains also a more refined Bennett type inequality for the deviation of Z above $\mathbb{E}\, Z'$, which does not follow from the moment type bounds we consider here, however a similar improvement, giving concentration around $\mathbb{E}\, Z$ can be up to constants recovered from the modified log-Sobolev inequality on the symmetric group. We do not discuss it in detail in this chapter, since it is necessarily expressed in terms of v and $\sup_{x\in\mathcal{X}} \|x\|_\infty$ and here we are interested primarily in improvements one can obtain by looking at the p-th moments rather than the ℓ_∞-norm. We revisit this problem in depth in Chapter 5. In particular, we obtain therein the aforementioned Bennet-type deviation bound.

Let us now pass to the proof of Proposition 2.4.9.

Proof of Proposition 2.4.9. Without loss of generality we can assume that $\mathcal{A}$ is finite, the general case follows then by approximation. For $i, j \in [n]$ define $Z_{ij} = \sup_{a\in\mathcal{A}} \sum_{k=1}^n a_{k\sigma(\tau_{ij}(k))}$. Note that by the definition of Z and the triangle inequality in ℓ_2,

$$\begin{aligned}\Big(\sum_{i,j=1}^n (Z - Z_{ij})_+^2\Big)^{1/2} &\le \sup_{a\in\mathcal{A}} \Big(\sum_{i,j=1}^n (a_{i\sigma(i)} + a_{j\sigma(j)} - a_{i\sigma(j)} - a_{j\sigma(i)})^2\Big)^{1/2} \\ &\le 2\sqrt{n} \sup_{a\in\mathcal{A}} \Big(\sum_{i=1}^n a_{i\sigma(i)}^2\Big)^{1/2} + 2 \sup_{a\in\mathcal{A}} \Big(\sum_{i,j=1}^n a_{ij}^2\Big)^{1/2}.\end{aligned}$$

Therefore, by Proposition 2.4.8, we obtain

$$\begin{aligned}\|(Z - \mathbb{E}\, Z)_+\|_r \le 2D_{2.4.8}\sqrt{r} &\Big\| \sup_{a\in\mathcal{A}} \Big(\sum_{i=1}^n a_{i\sigma(i)}^2\Big)^{1/2} \Big\|_r \\ &+ 2D_{2.4.8}\sqrt{r}\Big(\sup_{a\in\mathcal{A}} \frac{1}{n} \sum_{i,j=1}^n a_{ij}^2\Big)^{1/2}.\end{aligned} \tag{2.4.19}$$

We start with estimating the first summand. Denote

$$S = \sup_{a\in\mathcal{A}} \Big(\sum_{k=1}^n a_{k\sigma(k)}^2\Big)^{1/2}, \quad S_{ij} = \sup_{a\in\mathcal{A}} \Big(\sum_{k=1}^n a_{k\sigma(\tau_{ij}(k))}^2\Big)^{1/2}.$$

By another application of Proposition 2.4.8, we get

$$\|(S - \mathbb{E}\, S)_+\|_r \le D_{2.4.8}\sqrt{r}\Big\|\Big(\frac{1}{n}\sum_{i,j=1}^{n}(S - S_{ij})_+^2\Big)^{1/2}\Big\|_r. \tag{2.4.20}$$

For a fixed value of σ let $a \in \mathcal{A}$ be such that

$$S = \Big(\sum_{i=1}^{n} a_{i\sigma(i)}^2\Big)^{1/2}.$$

Fix $i, j \in [n]$ and denote

$$s = \sqrt{\sum_{k \ne i,j} a_{k\sigma(k)}^2}, \qquad x = \sqrt{a_{i\sigma(i)}^2 + a_{j\sigma(j)}^2}, \qquad y = \sqrt{a_{i\sigma(j)}^2 + a_{j\sigma(i)}^2}.$$

Denote also by φ the function $t \mapsto \sqrt{s^2 + t^2}$. Then φ is convex and increasing on $\mathbb{R}_+$. Moreover, if $(S - S_{ij})_+$ is nonzero, then $x^2 > y^2$, in particular $x > 0$ and so φ is differentiable at x. As a consequence, by convexity of φ and monotonicity of the function $t \mapsto t_+^2$, we obtain

$$\begin{aligned}(S - S_{ij})_+^2 &\le (\varphi'(x)(x-y))_+^2 \le \varphi'(x)^2 x^2 = \frac{(a_{i\sigma(i)}^2 + a_{j\sigma(j)}^2)^2}{S^2} \\ &\le 2\frac{a_{i\sigma(i)}^2 + a_{j\sigma(j)}^2}{S^2}\max_{k \le n} a_{k\sigma(k)}^2.\end{aligned}$$

Summing over all $i, j \in [n]$ we obtain

$$\sum_{i,j=1}^{n}(S - S_{ij})_+^2 \le 4n \max_{k\le n}\sup_{a\in\mathcal{A}} a_{k\sigma(k)}^2,$$

which in combination with (2.4.20) gives

$$\|(S - \mathbb{E}\, S)_+\|_r \le 2D_{2.4.8}\sqrt{r}\Big\|\max_{k\le n}\sup_{a\in\mathcal{A}}|a_{k\sigma(k)}|\Big\|_r = 2D_{2.4.8}\sqrt{r}B_r. \tag{2.4.21}$$

Hence,

$$\|S\|_r \le \|\mathbb{E}\, S\|_r + \|(S - \mathbb{E}\, S)_+\|_r \le A + 2D_{2.4.8}\sqrt{r}B_r. \tag{2.4.22}$$

Let us note that

$$A = \mathbb{E}\, S \ge \sup_{a\in\mathcal{A}}\mathbb{E}\Big(\sum_{i=1}^{n} a_{i\sigma(i)}^2\Big)^{1/2}.$$

Applying (2.4.22) with $r = 2$ to the one element sets $\{a\}$ instead of $\mathcal{A}$, we obtain

$$\begin{aligned}\Big(\frac{1}{n}\sum_{i,j=1}^{n} a_{ij}^2\Big)^{1/2} = \Big(\mathbb{E}\sum_{i=1}^{n} a_{i\sigma(i)}^2\Big)^{1/2} &\le A + 2\sqrt{2}D_{2.4.8}\Big\|\max_{k\le n}|a_{k\sigma(k)}|\Big\|_2 \\ &\le A + 2\sqrt{2}D_{2.4.8}B_r.\end{aligned}$$

Combining the above inequality with (2.4.19) and (2.4.22) we obtain

$$\|(Z - \mathbb{E}\, Z)_+\|_r \le 4D_{2.4.8}\sqrt{r}A + 10D_{2.4.8}^2 rB_r,$$

which ends the proof of (2.4.13). The inequality (2.4.14) is now an easy consequence of Chebyshev's inequality in L_r (note that for $r \le 2$ the right-hand side exceeds one, so the inequality is trivial). $\square$

Multislices

Let us conclude this section with a remark concerning multislices. For a positive integers $n \geq l$ and a sequence $\kappa = (\kappa_1, \ldots, \kappa_l) \in \mathbb{N}_+^l$ such that $\kappa_1 + \ldots + \kappa_l = n$ consider

$$U_\kappa = \{x = (x_1, \ldots, x_n) \in [l]^n \colon \#\{j : x_j = i\} = \kappa_i \text{ for } i = 1, \ldots, l\}$$

– the *multislice* of $[l]^n$ consisting of all the sequences which for $i \leq l$ take the value i exactly κ_i times. If $l = 2$ then U_κ can be identified with a slice of the discrete cube $\{0,1\}^n$ by a hyperplane perpendicular to the vector $(1, \ldots, 1)$. The dynamics corresponding to switching a randomly chosen pair of coordinates of an element of U_κ is related to the Bernoulli–Laplace model of statistical mechanics (which can also be interpreted as an urn scheme). In [42] Bobkov and Tetali proved Beckner inequalities for this dynamics in the case of $l = 2$. From this result they inferred modified log-Sobolev inequalities, which were proven independently by Gao–Quastel [91]. Again the constant ρ_1 in the log-Sobolev inequality (LSI) degenerates as $n \to \infty$. It was first computed in [140] for $l = 2$. Subsequently, estimates on this constant for general n, l, κ were obtained in [88]. These estimates are optimal for l fixed but deteriorate as n tends to ∞. The authors of [88] put forward a conjecture regarding the sharp growth rate of ρ_1 for all values of l, which was recently proved by Salez [177].

We would like to point out that in the case of Beckner and modified log-Sobolev inequalities the results on the symmetric group cited in the previous section can be projected onto U_κ yielding inequalities with constants of a better order than ρ_1, which can be then used to conclude moment estimates. We skip the rather standard details.

2.4.5 Zero-range processes and negatively dependent binary variables

New examples of measures satisfying the modified log-Sobolev inequality have been recently obtained in the works of Hermon and Salez [114, 115] and Cryan et al. [74] – these regard some classes of negatively dependent random variables on the hypercube and zero-range processes. In this section we will provide just an outline of their results and briefly comment on what can be obtained by combining them with ours. A more in-depth investigation of the stochastic covering property (which is the main theme of [114]) and its relation to concentration is presented in Chapter 4, cf. also Section 1.4.

Stochastic covering property

The results in [114] concern measures on $\mathcal{X} \subseteq \{0,1\}^n$. Recall that for $x, y \in \mathcal{X}$ we will say that x covers y ($x \rhd y$) if

$$x = y \quad \text{or} \quad \exists_{i \leq n} \; x = y + e_i,$$

where e_i's are the standard basis vectors, i.e., if $x \neq y$ then x can be obtained from y by increasing a single coordinate. For probability measures ν_1, ν_2 on $\mathcal{X}$ we say that ν_1 covers ν_2 if there is coupling of ν_1, ν_2 supported on the set $\{(x,y) \in \mathcal{X}^2 \colon x \rhd y\}$.

Let μ be a probability measure on $\mathcal{X}$ and X a random vector with law μ. For a set $I \subseteq [n]$ we will write $X_I = (x_i)_{i \in I}$. We say that μ satisfies the SCP if for every $I \subseteq [n]$ and $x, y \in \{0,1\}^I$, such that $\mathbb{P}(X_I = x), \mathbb{P}(X_I = y) > 0$ and $x \rhd y$, one has

$$\mathbb{P}(X_{I^c} \in \cdot | X_I = y) \rhd \mathbb{P}(X_{I^c} \in \cdot | X_I = x).$$

Let us moreover introduce a relation $\sim$ on $\mathcal{X}$: $x \sim y$ if and only if x and y differ at a single coordinate or by a transposition of two coordinates.

Examples of measures satisfying the SCP are given, e.g., by laws of weighted random bases of balanced matroids [187], in particular the uniform measure on the set of all spanning trees of a given graph (we identify here the spanning tree with an element of $\{0,1\}^E$, where E is the set of edges). We refer to [170] for further examples.

The authors of [114] obtain modified log-Sobolev inequalities for measures with SCP.

Theorem 2.4.12. *Let X be a random vector with values in $\mathcal{X} \subseteq \{0,1\}^n$ and law μ, satisfying the SCP.*

(i) Let Q be any kernel, reversible with respect to μ. Then the measure μ satisfies (mLSI) with constant $\rho_0 \geq \min_{x,y \in \mathcal{X}, x \sim y} \max(Q_x(y), Q_y(x))$.

(ii) There exists a kernel Q such that for all $x \in \mathcal{X}$, $Q_x(\cdot)$ is supported on $\{y \in \mathcal{X} : y \sim x\}$, $\sum_{y \in \mathcal{X} \setminus \{x\}} Q_x(y) \leq 1$ and (mLSI) is satisfied with $\rho_0 \geq 1/n$.

(iii) If μ is supported on the set $\{x \in \mathcal{X} : \sum_{i=1}^n x_i = k\}$, then there exists a kernel supported on the set of $y \in \mathcal{X}$ such that x and y differ by a transposition of two coordinates, and such kernel verifies $\sum_{y \in \mathcal{X} \setminus \{x\}} Q_x(y) \leq 1$ and (mLSI) is satisfied with $\rho_0 \geq 1/(2k)$.

Combining the above theorem with Theorem 2.2.1 and Proposition 2.3.1 we immediately obtain the following corollary.

Corollary 2.4.13. *In the setting of Theorem 2.4.12, for $p \in (1,2]$, the measure μ satisfies (Bec-p) with $\alpha_p \geq \rho_0/6$. As a consequence, for any function $f : \mathcal{X} \to \mathbb{R}$ and $r \geq 2$,*

$$\|(f(X) - \mathbb{E}\, f(X))_+\|_r \leq K\sqrt{r} \Big\| \Big(\sum_{y \in \mathcal{X}} (f(y) - f(x))_-^2 Q_x(y) \Big)^{1/2} \Big\|_r, \qquad (2.4.23)$$

where $K = \sqrt{\frac{3\sqrt{e}}{\rho_0(\sqrt{e}-1)}}$.

To the best of our knowledge, Beckner inequalities for general measures satisfying the SCP have not been so far considered in the literature and this is the first result that establishes them in this setting.

The case (i) of Theorem 2.4.12 applies in particular to Q given by the Metropolis-Hastings kernel $Q_x(y) = \frac{1}{2kn} \min\left(\frac{\mu(x)}{\mu(y)}, 1\right)$ if $y \sim x$ and $Q_x(y) = 0$ otherwise. Note however that, e.g., in the case of the uniform distribution on slices of the cube application of this part of Theorem 2.4.12 gives a suboptimal bound (cf. the discussion in Section 2.4.4 and [42, 91]). Part (ii) gives the right order of ρ_0, however the existence of Q is obtained via an inductive procedure with respect to the dimension n and so in general it is not given by an explicit formula. It is worth noting that from the point of view of concentration of measure with respect to the Hamming distance, the construction of the kernel Q might play important role. In order to obtain the same concentration profile as in the product case, one has to construct an appropriate kernel Q – this idea is investigated further in Chapter 4 and is the main theme of Section 4.3.

Strongly log-concave measures

Another possible generalization of the strong Rayleigh property (apart from the SCP described in the previous section) is the *strong log-concavity* (abbrev. SLC) treated in detail in the context of homogeneous measures in [74]. Let us recall that a polynomial P in n real variables with nonnegative coefficients is log-concave at $x \in [0,\infty)^n$ if the Hessian of $\log P$ is negative semi-definite at x. A measure π on $\mathcal{X} \subset \{0,1\}^n$ is strongly log-concave if its generating polynomial is log-concave at all-ones vector $(1,\dots,1)$ after taking any sequence of pairwise distinct partial derivatives.

In [74] the authors verify that SCP and SLC are in general incomparable by constructing appropriate examples. It is also known, cf. [53], that any k-homogeneous SLC measure is supported on the set of bases of some matroid of rank k. Using this fact, and extending the previous results by Anari et al. [13] and Kaufman–Oppenheim [129], Cryan et al. [74] explicitly construct a *down-up random walk*, which has any given k-homogeneous strongly log-concave measure as a stationary distribution, and prove that it satisfies the modified log-Sobolev inequality (mLSI) with $\rho_0 = 1/k$.

Let us now revisit their construction. We refer the Reader to [166] for the basic definitions concerning matroids. For a given k-homogeneous SLC measure μ, let $\mathcal{M} \subset \{0,1\}^n$ be the associated matroid of rank k (therefore, if $\mathcal{M}_{k'} = \{x \in \mathcal{M} \colon \sum_{i=1}^n x_i = k'\}$ for $k' = 1,\dots,k$, then μ is supported on $\mathcal{M}_k$). For $x \in \mathcal{M}$, denote $\mathcal{M}^{\downarrow}_{k'}(x) = \{y \in \mathcal{M}_{k'} \colon x \geq y\}$ and $\mathcal{M}^{\uparrow}_{k'}(x) = \{y \in \mathcal{M}_{k'} \colon x \leq y\}$, where $x \geq y$ for $x, y \in \{0,1\}^n$ if for all $i = 1,\dots,n$, $x_i = 0$ implies that $y_i = 0$. Consider the kernel $Q = Q^{\uparrow} \circ Q^{\downarrow}$ acting on $\mathcal{M}_k$, where for $x \in \mathcal{M}_k$, $Q^{\downarrow}_x(y) = \mathbf{1}_{y \in \mathcal{M}^{\downarrow}_{k-1}(x)} / |\mathcal{M}^{\downarrow}_{k-1}(x)|$ is the probability kernel that samples uniformly from the set $\mathcal{M}^{\downarrow}_{k-1}(x)$ and for $y \in \mathcal{M}_{k-1}$, $Q^{\uparrow}_y(x) = \mu(x)\mathbf{1}_{x \in \mathcal{M}^{\uparrow}_k(y)} / \mu(\mathcal{M}^{\uparrow}_k(y))$ is the probability kernel sampling from the set $\mathcal{M}^{\uparrow}_k(y)$ with probability proportional to μ. The above construction asserts that Q is reversible with respect to μ.

The main result of [74] is the following theorem.

Theorem 2.4.14. *For any k-homogeneous SLC measure μ on $\{0,1\}^n$ and dynamics given by (2.1.9) with the kernel Q constructed above, the modified log-Sobolev inequality (mLSI) holds with $\rho_0 = 1/k$.*

Recently, extensions of Theorem 2.4.14 for more general classes of measures were obtained in [71, 12]. We do not however describe these results in detail for the sake of brevity and only mention that one can apply our results to their bounds on the modified log-Sobolev constants in the same manner as below in the case of [74]. Additionally, one notable feature of all mentioned works is that they do not rely on the approach developed by Lee–Yau [140] (which exploits the convexity of the function $(a,b) \mapsto (a-b)(\log a - \log b)$). Hence, the proofs do not seem to be easily adapted to general Phi-Sobolev inequalities (just by changing the underlying convex function). In particular, deriving the Beckner inequality (Bec-p) without using Theorem 2.2.1 is nontrivial in these cases.

Combining Theorem 2.4.14 with Theorem 2.2.1 and Proposition 2.3.1 we obtain the following corollary.

Corollary 2.4.15. *For any k-homogeneous SLC measure μ on $\{0,1\}^n$ and dynamics given by (2.1.9) with the kernel Q constructed above, the Beckner inequality (Bec-p) holds with $\alpha_p \geq \frac{1}{6k}$ for any $p \in (1,2]$.*

As a consequence, for any $f\colon \{0,1\}^n \to \mathbb{R}$,

$$\|(f-\mu(f))_+\|_r \le K \cdot \sqrt{rk}\, \left\|\sqrt{\Gamma_+(f)}\right\|_r,$$

where $K = \sqrt{\frac{3\sqrt{e}}{\sqrt{e}-1}}$ and Γ_+ is given by (2.1.14).

Zero-range processes

Another class of examples coming from the recent work of Hermon and Salez is described in [115] and concerns zero-range processes, i.e., stochastic systems in which a fixed number m of particles occupy n sites. The particles leave the present site, with rates λ_i depending on the number of particles they share it with, and choose the new site according to a prescribed probability measure p common for all the particles. More precisely, let m, n be positive integers and let $\mathcal{X} = \{x \in \mathbb{N}^n \colon \sum_{i=1}^n x_i = m\}$. Consider functions $\lambda_i\colon \{0\} \cup [m] \to [0,\infty)$, $i = 1,\ldots,n$, such that $\lambda_i(0) = 0$ and let $p = (p_1,\ldots,p_n)$ be a probability vector. The zero-range dynamics is given by a Markov generator of the form

$$Lf(x) = \sum_{i,j=1}^n (f(x+e_j-e_i) - f(x))\lambda_i(x_i)p_j, \tag{2.4.24}$$

where $e_1,\ldots,e_n$ is the standard basis in $\mathbb{R}^n$. This dynamics is reversible with respect to the probability measure μ on $\mathcal{X}$, defined by

$$\mu(\{x\}) = \frac{1}{Z}\prod_{i=1}^n \frac{p_i^{x_i}}{\lambda_i(1)\cdots\lambda_i(x_i)}. \tag{2.4.25}$$

Hermon and Salez obtained a modified log-Sobolev inequality for the case when the rates of escape are sandwiched between two linear functions, with constant ρ_0 depending only on the directional coefficients of the functions. In particular, this provides a solution to a conjecture posed by Caputo, Dai Pra, and Posta [55, 54]. Below we state their theorem and a corollary one can immediately obtain from it with our results.

We remark that Beckner's inequalities for zero-range processes were previously considered in [125] and very recently in [72] in the case of p being the uniform distribution and under a restriction on Δ, δ (for instance [72] assumes that $\Delta \le 2\delta$). See Remark 5.3 in [72] for a detailed discussion of the applicability of the Bakry-Émery approach used in these references. The equivalence with the modified log-Sobolev inequality allows going beyond this restriction and conclude Beckner inequalities directly from the result by Hermon and Salez.

Theorem 2.4.16. *Assume that for $l \in \{0\} \cup [m-1]$,*

$$\delta \le \lambda_i(l+1) - \lambda_i(l) \le \Delta, \tag{2.4.26}$$

where δ, Δ are positive constants. Then the zero range dynamics corresponding to the generator (2.4.24) satisfies the modified log-Sobolev inequality with $\rho_0 \ge \frac{\delta^2}{2\Delta}$.

Theorem 2.2.1 and Proposition 2.3.1 immediately yield the following

Corollary 2.4.17. *If the assumption (2.4.26) is satisfied, then the zero-range dynamics satisfies for any $p \in (1,2]$ the Beckner inequality (Bec-p) with constant $\alpha_p \ge \frac{\delta^2}{12\Delta}$.*

As a consequence, if $X = (X_1, \ldots, X_n)$ is a random vector with law μ given by (2.4.25), then for every function $f\colon \mathcal{X} \to \mathbb{R}$ and $r \geq 2$,

$$\|(f(X) - \mathbb{E}\, f(X))_+\|_r \leq K\frac{\sqrt{\Delta}}{\delta}\sqrt{r}\Big\|\Big(\sum_{i=1}^{n}\sum_{j=1}^{n}(f(X+e_j-e_i)-f(X))_-^2\lambda_i(X_i)p_j\Big)^{1/2}\Big\|_r,$$

where $K = \sqrt{\frac{6\sqrt{e}}{\sqrt{e}-1}}$.

2.4.6 The Poisson space

We will now present applications of our results to concentration of measure on the Poisson space. In literature there are quite a few results, providing functional inequalities and concentration estimates on path spaces of Poisson point processes. In particular Ané and Ledoux [15] obtained certain modified log-Sobolev inequalities (of a different form than (mLSI)), Wu [202] proved a modified log-Sobolev inequality implying in particular the one from [15] as well as (mLSI), Chafaï [57] considered general Φ-Sobolev inequalities (including ones of Beckner type), Reynaud-Bouret [176] obtained concentration estimates for suprema of compensated stochastic integrals (see also [121, 1] for multiple stochastic integrals). More recently Reitzner introduced a version of the convex distance inequality [175], Bachmann and Peccati [19] used modified log-Sobolev inequalities due to Wu to obtain concentration results with focus on geometric functionals, an approach which was subsequently extended by Bachmann [18], Bachmann and Reitzner [20]. Nourdin, Peccati, and Yang [163] proved restricted hypercontractive for certain classes of functions, whereas Gozlan, Herry, Peccati [106] obtained transportation type inequalities.

Our goal is to complement these results with moment estimates and derive from them concentration inequalities. Moment estimates of subexponential type, i.e., with linear growth of constants as $r \to \infty$, were obtained from the Poincaré and Cheeger inequalities by Houdré and Privault in [119, 120]. To the best of our knowledge the inequalities we present in Proposition 2.4.20 are the first subgaussian type moment estimates on the Poisson space. Moreover, as one can easily see using infinite divisibility and the Central Limit Theorem, the growth of constants in our estimate as $r \to \infty$ is optimal.

We will start by a necessarily brief introduction of the setting. We refer to [133, 132] for a detailed presentation of Poisson point processes and stochastic calculus on the Poisson space. Furthermore, we remark, that soon after our results were announced, the moment estimates were derived by Gusakova–Sambale–Thäle [111], also by means of the Boucheron–Bousquet–Lugosi–Massart approach, adapted to the case of the Poisson space. We refer the Reader to [111] for additional examples of geometric functionals which can be treated with general moment estimates

Let $(\mathcal{X}, \mathcal{B})$ be a measurable space endowed with a σ-finite measure λ. Let $\mathcal{N}$ be the space of all $\mathbb{N} \cup \{\infty\}$-valued measures on $(\mathcal{X}, \mathcal{B})$ which can be expressed as countable sums of $\mathbb{N}$-valued measures. The measurable structure on $\mathcal{N}$ that we consider is given by the smallest σ-field $\mathcal{G}$ such that for all $B \in \mathcal{B}$ the map $\mu \mapsto \mu(B)$ is $\mathcal{G}$-measurable. Recall that an $\mathcal{N}$-valued random variable η is a Poisson process with intensity λ if

(i) for every $B \in \mathcal{B}$, the random variable $\eta(B)$ has Poisson distribution with parameter $\lambda(B)$ (which we interpret as the Dirac mass at $\lambda(B)$ if $\lambda(B) \in \{0, \infty\}$),

(ii) for every positive integer m and all pairwise disjoint sets $B_1, \ldots, B_m \in \mathcal{B}$, the random variables $\eta(B_1), \ldots, \eta(B_m)$ are jointly independent.

A Poisson process η is called proper, if there exists a random variable $\kappa \leq \infty$ and a sequence of $\mathcal{X}$-valued random variables X_i such that

$$\eta = \sum_{i=1}^{\kappa} \delta_{X_i}, \tag{2.4.27}$$

where δ_x stands for Dirac's mass at x. Corollary 3.7 in [133] asserts that for every Poisson process there exists a proper Poisson process with the same distribution. We will use this fact together with σ-finiteness of λ to avoid certain measurability issues in the definition of quantities that we are about to consider.

More precisely, let $\mathcal{X}_n \in \mathcal{B}$ be a sequence of sets with $\bigcup_n \mathcal{X}_n = \mathcal{X}$, such that $\lambda(\mathcal{X}_n) < \infty$ for all n. We can and do assume that η is proper and consider η as a random variable with values in the space $\mathcal{M} \subset \mathcal{N}$ of measures of the form $\mu = \sum_{i=1}^{K} \delta_{x_i}$ where $K \leq \infty$ and $x_i \in \mathcal{X}$, such that for every n, $\mu(\mathcal{X}_n) < \infty$. We will again endow this space with the smallest σ-field $\mathcal{S}$ such that the maps $\mu \mapsto \mu(B)$ are measurable for all $B \in \mathcal{B}$.

Remark 2.4.18. It is not difficult to see that $\mathcal{S} = \{A \cap \mathcal{M} \colon A \in \mathcal{G}\}$. In particular $\mathcal{S}$-measurable functions on $\mathcal{M}$ are just restrictions of $\mathcal{G}$-measurable functions on $\mathcal{N}$. We stress that in what follows we will consider inequalities for functions defined on the path space $(\mathcal{N}, \mathcal{G})$, however one should remember that the underlying Poisson process takes values in $(\mathcal{M}, \mathcal{S})$, which makes the quantities we will deal with well-defined. In what follows, so as not to obscure the main ideas, we will not discuss in detail the standard but somewhat tedious measurability issues – we further comment on them in more detail in Appendix A.4.

For $F \colon \mathcal{N} \to \mathbb{R}$ and $x \in \mathcal{X}$ define

$$D_x^+ F(\eta) = F(\eta + \delta_x) - F(\eta)$$

and

$$D_x^- F(\eta) = F(\eta) - F(\eta - \delta_x)$$

if $\eta \geq \delta_x$ and $D_x^- F = 0$ otherwise.

In [202] Wu proved that for arbitrary positive integrable $F \colon \mathcal{N} \to [0, \infty)$,

$$\operatorname{Ent} F \leq \mathbb{E} \int_{\mathcal{X}} (D_x^+ \Phi(F) - \Phi'(F) D_x^+ F) \lambda(dx), \tag{2.4.28}$$

where $\Phi(t) = t \log t$ and the expectation is taken with respect to η. From this inequality it easily follows that η satisfies the modified log-Sobolev inequality

$$\operatorname{Ent} F \leq \mathcal{E}(F, \log F) \tag{2.4.29}$$

for nonnegative $F \colon \mathcal{N} \to [0, \infty)$, and

$$\mathcal{E}(F, G) = \int_{\mathcal{X}} \mathbb{E}(D_x^+ F)(D_x^+ G) \lambda(dx)$$

is a bilinear form with

$$\operatorname{Dom}(\mathcal{E}) = \Big\{ F \in L_0(\mathcal{N}, \mathcal{L}(\eta)) \colon \int_{\mathcal{X}} \mathbb{E}(D_x^+ F)^2 \lambda(dx) < \infty \Big\},$$

where $\mathcal{L}(\eta)$ denotes the law of η.

Recall now the Mecke formula (see, e.g., [133, Theorem 4.1]), which asserts that for every measurable function $H\colon \mathcal{N} \times \mathcal{X} \to [0,\infty)$,

$$\mathbb{E}\int_{\mathcal{X}} H(\eta,x)\eta(dx) = \int_{\mathcal{X}} \mathbb{E}\, H(\eta+\delta_x,x)\lambda(dx).$$

We will actually need a slightly different version of this formula, valid for proper Poisson processes, given in [133, Theorem 4.5]. Namely, for $\mu \in \mathcal{N}$ and $x \in \mathcal{X}$ define $\mu\backslash\delta_x$ as $\mu - \delta_x$ if $\mu \geq \delta_x$, and as μ otherwise. If η is a proper Poisson process, then for every H as above

$$\mathbb{E}\int_{\mathcal{X}} H(\eta\backslash\delta_x,x)\eta(dx) = \int_{\mathcal{X}} \mathbb{E}\, H(\eta,x)\lambda(dx). \tag{2.4.30}$$

We remark that the assumption that η is proper allows to interpret the left-hand side as

$$\mathbb{E}\sum_{i=1}^{\kappa} H\Big(\sum_{j=1}^{\kappa} \mathbf{1}_{\{j\neq i\}}\delta_{X_j}, X_i\Big).$$

Let us also note that clearly Mecke's formula holds also for measurable functions $H\colon \mathcal{N}\times\mathcal{X} \to \mathbb{R}$, provided that its left- or right-hand side with H replaced by $|H|$ is finite.

For $\rho = \sum_{k=1}^{K} \delta_{x_k} \in \mathcal{M}$ (with $K \leq \infty$) we define a measure Q_ρ on $\mathcal{M}$ as

$$\begin{aligned} Q_\rho(A) &= \int_{\mathcal{X}} \mathbf{1}_A(\rho+\delta_x)\lambda(dx) + \int_{\mathcal{X}} \mathbf{1}_A(\rho-\delta_x)\rho(dx) \\ &= \lambda(\{x\colon \rho+\delta_x \in A\}) + \sum_{k=1}^{\kappa} \mathbf{1}_A(\rho - \delta_{x_k}). \end{aligned} \tag{2.4.31}$$

Using Dynkin's π-λ theorem one can prove that the definition of Q does not depend on the representation of ρ as a sum of Dirac's deltas (note that we do not assume measurability of singletons, so such a representation of ρ as a measure on $\mathcal{B}$ may not be unique), moreover Q is a kernel on $\mathcal{M}$ (this is the main reason for which we introduce the space $\mathcal{M}$, cf. Appendix A.4 for more discussion).

By Mecke's formula (2.4.30) for every measurable $G\colon \mathcal{M}\times\mathcal{M} \to [0,\infty)$,

$$\begin{aligned} &\mathbb{E}\int_{\mathcal{M}} G(\sigma,\eta)Q_\eta(d\sigma) \\ &= \mathbb{E}\int_{\mathcal{X}} G(\eta+\delta_x,\eta)\lambda(dx) + \mathbb{E}\int_{\mathcal{X}} G(\eta-\delta_x,\eta)\eta(dx) \\ &= \mathbb{E}\int_{\mathcal{X}} G(\eta,\eta-\delta_x)\eta(dx) + \mathbb{E}\int_{\mathcal{X}} G(\eta,\eta+\delta_x)\lambda(dx) \\ &= \mathbb{E}\int_{\mathcal{M}} G(\eta,\sigma)Q_\eta(d\sigma). \end{aligned}$$

Thus, the kernel Q_σ is reversible with respect to the law of η.

Note that by another application of Mecke's formula (2.4.30), for any $F, G \in \mathrm{Dom}(\mathcal{E})$,

$$\begin{aligned} \mathcal{E}(F,G) = \frac{1}{2}\Big(&\mathbb{E}\int_{\mathcal{X}} (F(\eta)-F(\eta-\delta_x))(G(\eta)-G(\eta-\delta_x))\eta(dx) \\ &+ \mathbb{E}\int_{\mathcal{X}} (F(\eta+\delta_x)-F(\eta))(G(\eta+\delta_x)-G(\eta))\lambda(dx)\Big) \\ = \frac{1}{2}&\mathbb{E}\int_{\mathcal{M}} (F(\sigma)-F(\eta))(G(\sigma)-G(\eta))Q_\eta(d\sigma). \end{aligned} \tag{2.4.32}$$

We remark that the functions F, G above are defined on $\mathcal{N}$, but their restrictions to $\mathcal{M}$ are $\mathcal{S}$-measurable (cf. Remark 2.4.18), so the last expression in the above formula is well-defined. In particular one can see that the value of $\mathcal{E}(F, G)$ depends only on the behavior of F and G on $\mathcal{M}$.

Consider now the space

$$\mathcal{A} = \Big\{ F \in L_0(\mathcal{M}, \mathcal{L}(\eta)) \colon \int_{\mathcal{M}} (F(\eta) - F(\sigma))^2 Q_\eta(d\sigma) < \infty \text{ a.s.} \Big\}.$$

Observe also that the restriction from $\mathcal{N}$ to $\mathcal{M}$ gives a natural identification of $L_0(\mathcal{M}, \mathcal{L}(\eta))$ and $L_0(\mathcal{N}, \mathcal{L}(\eta))$ (cf. again Remark 2.4.18), therefore we can also consider $\mathcal{A}$ as a subspace of the latter space.

On $\mathcal{A} \times \mathcal{A}$ define

$$\begin{aligned}
\Gamma(F, G) &= \frac{1}{2} \int_{\mathcal{M}} (F(\eta) - F(\sigma))(G(\eta) - G(\sigma)) Q_\eta(d\sigma) \\
&= \frac{1}{2} \Big(\int_{\mathcal{X}} (F(\eta) - F(\eta - \delta_x))(G(\eta) - G(\eta - \delta_x)) \eta(dx) \\
&\qquad + \int_{\mathcal{X}} (F(\eta + \delta_x) - F(\eta))(G(\eta + \delta_x) - G(\eta)) \lambda(dx) \Big) \\
&= \frac{1}{2} \int_{\mathcal{X}} (D_x^- F(\eta))(D_x^- G(\eta)) \eta(dx) + \frac{1}{2} \int_{\mathcal{X}} (D_x^+ F(\eta))(D_x^+ G(\eta)) \lambda(dx)
\end{aligned}$$

and

$$\begin{aligned}
\Gamma_+(F) &= \int_{\mathcal{M}} (F(\eta) - F(\sigma))_+^2 Q_\eta(d\sigma) \qquad (2.4.33)\\
&= \int_{\mathcal{X}} (F(\eta) - F(\eta - \delta_x))_+^2 \eta(dx) + \int_{\mathcal{X}} (F(\eta + \delta_x) - F(\eta))_-^2 \lambda(dx) \\
&= \int_{\mathcal{X}} (D_x^- F(\eta))_+^2 \eta(dx) + \int_{\mathcal{X}} (D_x^+ F(\eta))_-^2 \lambda(dx).
\end{aligned}$$

By (2.4.32) we then have

$$\mathcal{E}(F, G) = \mathbb{E}\, \Gamma(F, G),$$

which shows that one can interpret Wu's inequality (2.4.29) in the setting of our main results (cf. (2.1.9), (2.1.11), (2.1.14)) and (2.4.29) becomes just (mLSI) with $\rho_0 = 1$. In particular, we also obtain Beckner's inequality (Bec-p) with $\alpha_p \geq 1/6$.

Remark 2.4.19. We remark that Γ is closely related to the carré du champ operator for the Ornstein–Uhlenbeck process on the Poisson space (see [132]). In [83, Proposition 2.6] it is shown that under suitable assumptions $\Gamma(F, G)$ actually coincides with the carré du champ operator. Similarly, as in [19] we find it however simpler to introduce Γ and Γ_+ via the Mecke formula (2.4.30), which gives greater generality and does not require a detailed discussion of domains.

Beckner's inequality (Bec-p) and Propositions 2.3.1 and 2.3.3 imply the following proposition providing Sobolev type inequalities on the Poisson space.

Proposition 2.4.20. *For any $F \colon \mathcal{N} \to \mathbb{R}$ and any $r \geq 2$,*

$$\begin{aligned}
\|F - \mathbb{E}\, F\|_r &\leq D_{2.4.20} \sqrt{r} \|\sqrt{\Gamma(F)}\|_r \\
&= D_{2.4.20} \sqrt{r} \Big\| \Big(\int_{\mathcal{X}} (D_x^+ F)^2 \lambda(dx) + \int_{\mathcal{X}} (D_x^- F)^2 \eta(dx) \Big)^{1/2} \Big\|_r
\end{aligned}$$

and

$$\begin{aligned}\|(F-\mathbb{E}\,F)_+\|_r &\le D_{2.4.20}\sqrt{r}\|\sqrt{\Gamma_+(F)}\|_r\\ &= D_{2.4.20}\sqrt{r}\Big\|\Big(\int_{\mathcal{X}}(D_x^+F)_-^2\lambda(dx)+\int_{\mathcal{X}}(D_x^-F)_+^2\eta(dx)\Big)^{1/2}\Big\|_r,\end{aligned}$$

where $D_{2.4.20}=\sqrt{3\frac{\sqrt{e}}{\sqrt{e}-1}}$.

Recently Bachmann and Peccati [19] used Wu's inequality (2.4.28) to derive concentration inequalities for Poisson functionals by various variants of the Herbst argument. They put special emphasis on increasing functionals, i.e., functionals F such that $D_x^+F\ge 0$ arguing that for them the second integral on the right-hand side of (2.4.33) vanishes, while the first integral can be often relatively easily estimated by appealing just to geometric properties of the functional, without taking into account the dependence on the intensity λ. Further applications of inequalities from [19] were presented in [18, 20]. The approach used in these papers relies on Herbst's argument, which usually requires either that $\Gamma_+(F)$ or $\Gamma(F)$ is uniformly bounded or that the function has some self-bounding properties (e.g., $\Gamma_+(F)\le\varphi(F)$ for some function φ). One aspect in which moment estimates of Proposition 2.4.20 complement this approach is that they can be easily used also if $\Gamma(F)$ or $\Gamma_+(F)$ have heavier tails, e.g., if they are not exponentially integrable.

Proposition 2.4.20 may also be an efficient tool in the self-bounded setting leading to inequalities which are (up to constants) comparable to those presented in said articles. We will illustrate it with use of the following proposition, which may be considered a counterpart of [19, Corrolary 3.5], which instead of moments concerns the Laplace transform.

Proposition 2.4.21. *Assume that* $F\colon\mathcal{N}\to[0,\infty)$ *is a measurable function which satisfies*

$$\Gamma_+(F)\le F^\alpha G \tag{2.4.34}$$

for some $\alpha\in[0,2)$ *and a function* $G\colon\mathcal{N}\to[0,\infty)$. *Then for* $r\ge 2$,

$$\begin{aligned}\|(F-\mathbb{E}\,F)_+\|_r \le\; &2D_{2.4.20}\sqrt{r}(\mathbb{E}\,F)^{\alpha/2}\|G^{1/(2-\alpha)}\|_r^{1-\alpha/2}\\ &+(2D_{2.4.20})^{2/(2-\alpha)}r^{1/(2-\alpha)}\|G^{1/(2-\alpha)}\|_r,\end{aligned}\tag{2.4.35}$$

where $D_{2.4.20}$ *is the constant from Proposition 2.4.20.*

Proof. Denote $A:=\|(F-\mathbb{E}\,F)_+\|_r$. We will first show that if $\mathbb{E}\,G^{r/(2-\alpha)}<\infty$, then $A<\infty$. Note that the inequality $a^{\alpha/2}(a^{1-\alpha/2}-b^{1-\alpha/2})_+\le(a-b)_+$ implies that

$$F^\alpha\Gamma_+(F^{1-\alpha/2})\le\Gamma_+(F)\le F^\alpha G.$$

As a consequence $\Gamma_+(F^{1-\alpha/2})\le G$ and thus if $\mathbb{E}\,G^{q/2}<\infty$, then by Proposition 2.4.20, $\mathbb{E}\,F^{q(1-\alpha/2)}<\infty$. Choosing $q=2r/(2-\alpha)$ we obtain that $A<\infty$.

Passing to the main part of the proof, we can assume that the right-hand side of (2.4.35) is finite and as a consequence $A<\infty$. By Proposition 2.4.20 and the assumption (2.4.34),

$$A^r\le D^r_{2.4.20}r^{r/2}\,\mathbb{E}\,F^{\alpha r/2}G^{r/2}\le D^r_{2.4.20}r^{r/2}(\mathbb{E}\,F^r)^{\alpha/2}(\mathbb{E}\,G^{r/(2-\alpha)})^{1-\alpha/2},$$

where in the second estimate we used Hölder's inequality. Using the triangle inequality in L_r, together with subadditivity of the function $t \mapsto t^{\alpha/2}$ we get

$$\begin{aligned} A &\le D_{2.4.20}\sqrt{r}\|F\|_r^{\alpha/2}\|G^{1/(2-\alpha)}\|_r^{1-\alpha/2} \\ &\le D_{2.4.20}\sqrt{r}A^{\alpha/2}\|G^{1/(2-\alpha)}\|_r^{1-\alpha/2} + D_{2.4.20}\sqrt{r}(\mathbb{E}\,F)^{\alpha/2}\|G^{1/(2-\alpha)}\|_r^{1-\alpha/2}, \end{aligned}$$

which easily implies that either

$$A \le 2D_{2.4.20}\sqrt{r}(\mathbb{E}\,F)^{\alpha/2}\|G^{1/(2-\alpha)}\|_r^{1-\alpha/2}$$

or

$$A \le (2D_{2.4.20})^{2/(2-\alpha)} r^{1/(2-\alpha)}\|G^{1/(2-\alpha)}\|_r,$$

proving the proposition. □

Let us illustrate Proposition 2.4.21 with two applications.

Suprema of Poisson stochastic integrals

Let $\mathcal{F}$ be a countable family of real valued functions on $\mathcal{X}$. Consider random variables of the form

$$Z = \sup_{f\in\mathcal{F}} \int_{\mathcal{X}} f(x)\eta(dx), \tag{2.4.36}$$

where all functions $f \in \mathcal{F}$ are nonnegative and $\mathcal{F} \subseteq L_1(\mathcal{X}, \lambda)$ and

$$S = \sup_{f\in\mathcal{F}} \int_{\mathcal{X}} f(x)(\eta - \lambda)(dx), \tag{2.4.37}$$

where $\mathcal{F} \subseteq L_2(\mathcal{X}, \lambda)$. Here the compensated integral is defined in the usual way, first directly on $L_1(\mathcal{X}, \lambda) \cap L_2(\mathcal{X}, \lambda)$, then extended by density – we refer to [133, Chapter 12] for background on the Wiener-Ito integral in the Poisson case. In the case when the functions in $\mathcal{F}$ are uniformly bounded and λ is finite, concentration inequalities for Z and S were obtained by Reynaud-Bouret in [176]. Here we will complement them with moment inequalities valid for not necessarily bounded classes or finite intensity measures.

The inequalities we obtain can be considered counterparts of results due to Giné–Latała–Zinn [97] for empirical processes in independent random variables. Originally they were derived from Talagrand's concentration inequality for empirical processes and the Hoffman–Jørgensen inequality; an alternate proof based on moment estimates of Theorem 2.4.4 was provided by Boucheron–Bousquet–Lugosi–Massart [47]. We remark that it should be possible to use this inequality together with infinite divisibility of Poisson processes similarly as in [176] to recover the estimates we present below (passing through finite intensity measures first), it seems however that this approach would require dealing with more technicalities in comparison with a direct application of general Poissonian moment estimates.

Let us start with the estimate on Z. Assume first that $\mathcal{F}$ is a finite class of functions. Note that if $f \ge 0$ for all $f \in \mathcal{F}$, then $Z(\eta) \le Z(\eta + \delta_x)$ and as a consequence, by (2.4.33),

$$\Gamma_+(Z) = \int_{\mathcal{X}} (D_y^- Z)_+^2 \eta(dy).$$

If $g_\eta \in \mathcal{F}$ is such that $Z(\eta) = \int_{\mathcal{X}} g_\eta(x)\eta(dx)$, then for all $y \in \mathrm{supp}(\eta) := \{X_1, X_2, \ldots\}$, where X_i are the random variables from the representation (2.4.27), we have

$$D_y^- Z \le \int_{\mathcal{X}} g_\eta(x)\eta(dx) - \int_{\mathcal{X}} g_\eta(x)(\eta - \delta_y)(dx) = g_\eta(y).$$

Thus,

$$\Gamma_+(Z) \le \int_{\mathcal{X}} g_\eta(y)^2\eta(dy) \le ZG,$$

where

$$G = \sup_{y\in\mathrm{supp}(\eta)} \sup_{f\in\mathcal{F}} f(y) \tag{2.4.38}$$

and as a consequence, an application of Proposition 2.4.21 with $\alpha = 1$ (followed by the monotone convergence theorem if $\mathcal{F}$ is infinite) gives

Corollary 2.4.22. *If Z is given by (2.4.36), and $\mathbb{E}\, Z < \infty$, then for a universal constant C, all $r \ge 2$ and any $\varepsilon > 0$,*

$$\begin{aligned}\|(Z - \mathbb{E}\, Z)_+\|_r &\le C(\sqrt{r}\sqrt{\mathbb{E}\, Z}\sqrt{\|G\|_r} + r\|G\|_r) \\ &\le C(\varepsilon\, \mathbb{E}\, Z + (1+\varepsilon^{-1})r\|G\|_r).\end{aligned}$$

Let us now pass to the variable S given by (2.4.37). By a limiting argument, we can again assume without loss of generality that $\mathcal{F}$ is finite. Further we can assume that $\mathcal{F} \subseteq L_1(\mathcal{X},\lambda) \cap L_2(\mathcal{X},\lambda)$, so that one can consider separately integration with respect to η and λ.

Let $g_\eta \in \mathcal{F}$ be such that $S = \int_{\mathcal{X}} g_\eta(x)\eta(dx) - \int_{\mathcal{X}} g_\eta(x)\lambda(dx)$. Arguing similarly as for the variable Z, we have

$$\begin{aligned}\Gamma_+(S) &\le \int_{\mathcal{X}} g_\eta(y)_-^2\lambda(dy) + \int_{\mathcal{X}} g_\eta(y)_+^2\eta(dy) \\ &\le \sup_{f\in\mathcal{F}} \int_{\mathcal{X}} f(x)^2\lambda(dx) + \sup_{f\in\mathcal{F}} \int_{\mathcal{X}} f(x)^2\eta(dx).\end{aligned}$$

Thus, by Proposition 2.4.20, the subadditivity of the function $x \mapsto x^{1/2}$ and the triangle inequality, we obtain

$$\|(S - \mathbb{E}\, S)_+\|_r \le C\sqrt{r}\Big(\Big(\sup_{f\in\mathcal{F}} \int_{\mathcal{X}} f(x)^2\lambda(dx)\Big)^{1/2} + \Big\|\sup_{f\in\mathcal{F}} \int_{\mathcal{X}} f(x)^2\eta(dx)\Big\|_{r/2}^{1/2}\Big).$$

The second term can be bounded from above by Corollary 2.4.22 applied to $\mathcal{F}' = \{f^2 \colon f \in \mathcal{F}\}$, which results in

Corollary 2.4.23. *If S given by (2.4.37) satisfies $\mathbb{E}\, S < \infty$, then for all $r \ge 4$,*

$$\|(S - \mathbb{E}\, S)_+\|_r \le C\Big(\sqrt{r}\Sigma + r\Big\| \sup_{x\in\mathrm{supp}\,\eta} \sup_{f\in\mathcal{F}} |f(x)|\Big\|_r\Big),$$

where

$$\Sigma^2 = \sup_{f\in\mathcal{F}} \int_{\mathcal{X}} f(x)^2\lambda(dx) + \mathbb{E} \sup_{f\in\mathcal{F}} \int_{\mathcal{X}} f^2(x)\eta(dx)$$

and C is a universal constant.

We remark that if the class $\mathcal{F}$ is uniformly bounded, then by Chebyshev's inequality the above corollary allows to recover (up to universal constants) the exponential upper tail estimates for S obtained in [176].

Non-negative U-statistics

Another application of Proposition 2.4.21 is related to geometric functionals of the Poisson process, specifically certain non-negative U-statistics, investigated recently by several authors [18, 19, 20, 106]. For a measurable kernel $h\colon \mathcal{X}^m \to [0,\infty)$, symmetric under permutation of arguments, let us define

$$U(\eta) = \sum_{i_1,\ldots,i_m}^{\neq} h(X_{i_1},\ldots,X_{i_m}),$$

where the X_i's are given by the representation (2.4.27) and the superscript $\neq$ indicates that the summation is taken over pairwise disjoint indices.

Let us note that for nonnegative h we have $D_X^+ h \geq 0$, therefore

$$\Gamma_+(U) = \int_{\mathcal{X}} (D_x^- U)_+^2 \eta(dx) = m^2 \sum_i \Big(\sum_{i_1,\ldots,i_{m-1}\colon i_j\neq i}^{\neq} h(X_{i_1},\ldots,X_{i_{m-1}},X_i)\Big)^2.$$

Therefore, using Proposition 2.4.21 and Chebyshev's inequality, we get the following corollary.

Corollary 2.4.24. *If U is an almost surely finite Poisson U-statistic based on a non-negative symmetric kernel h, and there exists $a \geq 0$ and $\alpha \in [0,2)$ such that*

$$\sum_i \Big(\sum_{i_1,\ldots,i_{m-1}\colon i_j\neq i}^{\neq} h(X_{i_1},\ldots,X_{i_{m-1}},X_i)\Big)^2 \leq aU^\alpha, \tag{2.4.39}$$

then for any $r \geq 2$,

$$\|(U - \mathbb{E}\,U)_+\|_r \leq C\sqrt{r}m\sqrt{a}(\mathbb{E}\,U)^{\alpha/2} + (Crm^2a)^{1/(2-\alpha)},$$

where C is some universal constant. As a consequence, for $t \geq 0$,

$$\mathbb{P}(U \geq \mathbb{E}\,U + t) \leq 2\exp\Big(-\min\Big(\frac{t^2}{C'm^2a(\mathbb{E}\,U)^\alpha}, \frac{t^{2-\alpha}}{C'm^2a}\Big)\Big),$$

where C' is some universal constant.

Let us remark that the references [19, 20, 106] provide also bounds on the left tail of U. It does not seem that such a bound can be easily recovered from the moment approach, since it relies heavily on another property of Poisson U-statistics with non-negative kernels, namely an appropriate notion of convexity, which allows for an application of certain correlation inequalities [19] or the Poisson convex distance inequality [20, 106]. It is an interesting question what moment estimates can be obtained under an additional convexity assumption. We remark that for the usual notion of convexity on $\mathbb{R}^n$, certain self-normalized moment estimates have been derived for all measures satisfying the convex concentration property [7].

The upper bounds on the upper tail of U, presented in the above references are

$$\mathbb{P}(U \geq \mathbb{E}\,U + t) \leq \exp\Big(-\frac{((\mathbb{E}\,U+t)^{1-\alpha/2} - (\mathbb{E}\,U)^{1-\alpha/2})^2}{2m^2a}\Big),$$

$$\mathbb{P}(U \geq \operatorname{Med} U + t) \leq 2\exp\Big(-\frac{t^2}{4m^2a(t+\operatorname{Med} U)^\alpha}\Big)\Big),$$

where $\operatorname{Med} U$ is any median of U. One can show that their behavior (disregarding the exact values of constants and using the fact that median and the mean of concentrated random variables are not far apart) is the same as of the upper bound of Corollary 2.4.24.

In [18] Bachmann and Reitzner verified the assumption (2.4.39) for a general class of U-statistics of Poisson processes on $\mathbb{R}^d$, with $\alpha = 2 - \frac{1}{m}$ and a depending on the dimension d and certain parameters of the kernel. In particular, they showed that this assumption is satisfied in the case when U is the number of copies of a given connected graph H on m vertices in the Gilbert graph based on η. They also proved that the above bounds are of the right order as $t \to \infty$ and agree with known limit theorems if one increases the intensity of the process.

This shows that the moment bounds of Proposition 2.4.20 may be an alternative for proving exponential inequalities for the upper tail of geometric functionals. It is an interesting question, however beyond the scope of this thesis, to verify whether Proposition 2.4.20 can give meaningful bounds in cases when functionals in question are known to have polynomial tails.

2.5 Remarks on higher order concentration

We will now describe applications of our results to higher order concentration inequalities, which provide estimates on probabilities of deviations from the mean for not necessarily Lipschitz functions, expressed in terms of higher order derivatives. Such estimates were obtained, e.g., in [8, 3, 39, 99, 4], both in the discrete and continuous settings. Since the latter case follows in a straightforward way from results in [8] we will focus here mainly on the discrete case. We will however start with an abstract statement, and only later specialize it to concrete examples.

2.5.1 Abstract inequality

Let $\mathcal{A}$ be a linear space of functions on $\mathcal{X}$ and $D_i \colon \mathcal{A} \to \mathbb{R}^{\mathcal{X}}$, $i = 1, \ldots, n$ be linear maps (we will think of them as directional derivatives). For positive integers $k, i_1, \ldots, i_k$ denote $D_{i_1 \ldots i_k} f = D_{i_1} \cdots D_{i_k} f$, $\mathbf{D}^k f = (D_{i_1 \ldots i_k} f)_{i_1, \ldots, i_k = 1}^n$. Thus, $\mathbf{D} = \mathbf{D}^1$ corresponds to the gradient and $\mathbf{D}^k$ for $k > 1$ to tensors of higher order derivatives, in particular $\mathbf{D}^k f(x) \in (\mathbb{R}^n)^{\otimes k}$. For $x^1, \ldots, x^k \in \mathbb{R}^n$, let $x^1 \otimes \cdots \otimes x^k = (x^1_{i_1} \cdots x^k_{i_k})_{i_1, \ldots, i_k = 1}^n \in (\mathbb{R}^n)^{\otimes k}$.

Let us also define the inner product on $(\mathbb{R}^n)^{\otimes k}$ with the formula

$$\langle x, y \rangle = \sum_{i_1, \ldots, i_k = 1}^n x_{i_1 \ldots i_k} y_{i_1 \ldots i_k}.$$

The following fact was proved in [8] in the case of the usual derivatives (see Proposition 3.2 therein). Since the easy proof is completely analogous to the one presented in [8] (it uses only linearity of D_i's) we will skip it. Below $|\cdot|$ denotes the euclidean norm on $\mathbb{R}^n$.

Proposition 2.5.1. *Let X be an $\mathcal{X}$-valued random variable. Assume that $r \geq 2$ and that there exists a constant K such that for all $f \in \mathcal{A}$,*

$$\|f(X) - \mathbb{E} f(X)\|_r \leq K \Big\| |\mathbf{D} f(X)| \Big\|_r.$$

Then for any integer d and any function $f\colon \mathcal{X} \to \mathbb{R}$ in the domain of $D_{i_1\ldots i_d}$, $i_1,\ldots,i_d \le n$ such that $D^d f(X) \in L_r$,

$$\begin{aligned}\|f(X) - \mathbb{E}\, f(X)\|_r \le & \frac{C^d K^d}{r^{d/2}} \|\langle \mathbf{D}^d f(X), G_1 \otimes \cdots \otimes G_d\rangle\|_r \\ & + \sum_{k=1}^{d-1} \frac{C^k K^k}{r^{k/2}} \|\langle \mathbb{E}_X \mathbf{D}^k f(X), G_1 \otimes \cdots \otimes G_k\rangle\|_r,\end{aligned}$$

where $G_1,\ldots,G_d$ are i.i.d. standard Gaussian vectors in $\mathbb{R}^n$, independent of X, and C is a universal constant.

If $\mathbf{D}^d f$ is uniformly bounded and X satisfies Beckner's inequality, then one can combine the above proposition with moment estimates of Theorem 2.3.1 and inequalities for multilinear forms in i.i.d. Gaussian vectors, obtained by Latała [135], which we will now recall.

Let us start by introducing the (rather involved) notation. For a multiindex $\mathbf{i} = (i_1,\ldots,i_d) \in [n]^d$ and $I \subseteq [d]$ we will write $\mathbf{i}_I = (i_k)_{k\in I}$. We will also denote $|\mathbf{i}| = \max_{j\le d} i_j$ and $|\mathbf{i}_I| = \max_{j\in I} i_j$. Let moreover P_d be 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

$$\|A\|_{\mathcal{I}} = \sup\Big\{ \sum_{\mathbf{i}\in[n]^d} a_{\mathbf{i}} \prod_{l=1}^k x^l_{\mathbf{i}_{I_l}} : \|(x^l_{\mathbf{i}_{I_l}})\|_2 \le 1, 1 \le l \le k\Big\}, \tag{2.5.1}$$

where $\|(x_{\mathbf{i}_{I_l}})\|_2 = \sqrt{\sum_{|\mathbf{i}_{I_l}|\le n} x^2_{\mathbf{i}_{I_l}}}$. Thus, e.g.,

$$\begin{aligned}\|(a_{ij})_{i,j\le n}\|_{\{1,2\}} &= \sup\Big\{ \sum_{i,j\le n} a_{ij}x_{ij} \colon \sum_{i,j\le n} x_{ij}^2 \le 1\Big\} = \sqrt{\sum_{i,j\le n} a_{ij}^2} \\ &= \|(a_{ij})_{i,j\le n}\|_{\mathrm{HS}}, \\ \|(a_{ij})_{i,j\le n}\|_{\{1\}\{2\}} &= \sup\Big\{ \sum_{i,j\le n} a_{ij}x_i y_j \colon \sum_{i\le n} x_i^2 \le 1, \sum_{j\le n} y_j^2 \le 1\Big\} \\ &= \|(a_{ij})_{i,j\le n}\|_{\ell_2^n\to\ell_2^n}, \\ \|(a_{ijk})_{i,j,k\le n}\|_{\{1,2\}\{3\}} &= \sup\Big\{ \sum_{i,j,k\le n} a_{ijk}x_{ij}y_k \colon \sum_{i,j\le n} x_{ij}^2 \le 1, \sum_{k\le n} y_k^2 \le 1\Big\},\end{aligned}$$

where for simplicity in the notation we skip the outer brackets and commas in the subscript and write, e.g., $\|\cdot\|_{\{1\}\{2\}}$ instead of $\|\cdot\|_{\{\{1\},\{2\}\}}$.

In the special case of $d = 2$, $\|\cdot\|_{\{1,2\}}$ and $\|\cdot\|_{\{1\}\{2\}}$ are just the Hilbert–Schmidt and operator norms of a matrix. We remark that for every d and $\mathcal{I} \in P_d$ we have $\|A\|_{\mathcal{I}} \le \|A\|_{\{[d]\}} = \sqrt{\sum_{|\mathbf{i}|\le n} a_{\mathbf{i}}^2}$. The norm $\|A\|_{\{[d]\}}$ can be considered a counterpart of the Hilbert–Schmidt norm for higher order tensors.

The result by Latała provides bounds on moments of multilinear forms in i.i.d. Gaussian variables in terms of the $\mathcal{I}$-norms of the corresponding matrix of coefficients.

Theorem 2.5.2 ([135]). *Let $G_1,\ldots,G_k$ be independent standard Gaussian vectors in $\mathbb{R}^n$ and let $A \in (\mathbb{R}^n)^{\otimes k}$. There exist constants C_k, depending only on k, such that for any $r \ge 2$,*

$$\frac{1}{C_k} \sum_{\mathcal{I}\in P_k} r^{|\mathcal{I}|/2} \|A\|_{\mathcal{I}} \le \|\langle A, G_1 \otimes \cdots \otimes G_k\rangle\|_r \le C_k \sum_{\mathcal{I}\in P_k} r^{|\mathcal{I}|/2} \|A\|_{\mathcal{I}},$$

Combining this result with Proposition 2.5.1, we obtain the following corollary.

Corollary 2.5.3. *Assume that there exist constants $M, \gamma > 0$ such that for all functions $f \in \mathcal{A}$, and all $r \geq 2$,*

$$\|f(X) - \mathbb{E}\, f(X)\|_r \leq M r^\gamma \Big\| |\mathbf{D} f(X)| \Big\|_r. \tag{2.5.2}$$

Then for every integer $d \geq 1$, any $r \geq 2$ and for every f in the domain of $\mathbf{D}^d$ such that $D^d f(X) \in L_r$,

$$\begin{aligned}\|f(X) - \mathbb{E}\, f(X)\|_r \leq & C_d\Big(M^d \sum_{\mathcal{J} \in P_d} r^{(\gamma - \frac{1}{2})d + \frac{|\mathcal{J}|}{2}} \Big\| \|\mathbf{D}^d f(X)\|_{\mathcal{J}} \Big\|_r \\ & + \sum_{k=1}^{d-1} M^k \sum_{\mathcal{J} \in P_k} r^{(\gamma - \frac{1}{2})k + \frac{|\mathcal{J}|}{2}} \|\, \mathbb{E}\, \mathbf{D}^k f(X)\|_{\mathcal{J}} \Big),\end{aligned}$$

where C_d depends only on d. Moreover, if $\mathbf{D}^d f(x)$ is uniformly bounded on $\mathcal{X}$, then for $t > 0$,

$$\mathbb{P}(|f(X) - \mathbb{E}\, f(X)| \geq t) \leq 2 \exp\Big(- \frac{1}{C'_d} \eta_f(t) \Big),$$

where C'_d is another constant depending only on d and

$$\eta_f(t) = \min(A, B)$$

with

$$\begin{aligned} A &= \min_{\mathcal{J} \in P_d} \Big(\frac{t}{M^d \sup_{x \in \mathcal{X}} \|\mathbf{D}^d f(x)\|_{\mathcal{J}}} \Big)^{\frac{2}{(2\gamma - 1)d + |\mathcal{J}|}}, \\ B &= \min_{1 \leq k \leq d-1} \min_{\mathcal{J} \in P_k} \Big(\frac{t}{M^k \|\, \mathbb{E}\, \mathbf{D}^k f(X)\|_{\mathcal{J}}} \Big)^{\frac{2}{(2\gamma - 1)k + |\mathcal{J}|}}. \end{aligned}$$

Proof. To obtain the moment estimate we combine Proposition 2.5.1 with $K = M r^\gamma$ and Theorem 2.5.2. The second part follows from the first one by an application of Chebyshev's inequality for the r-th moment and optimization in r. □

A typical application of the above corollary is the situation in which $\gamma = 1/2$ (such a subgaussian bound holds by Proposition 2.3.3, e.g., under the assumption of modified log-Sobolev inequalities) and f has bounded derivatives of second order. The tail bound one obtains is then

$$\mathbb{P}(|f(X) - \mathbb{E}\, f(X)| \geq t) \leq 2 \exp\Big(- c \min\Big(\frac{t^2}{\sup_{x \in \mathcal{X}} \|\mathbf{D}^2 f(x)\|_{\mathrm{HS}}^2 + |\, \mathbb{E}\, \mathbf{D} f(X)|^2}, \frac{t}{\sup_{x \in \mathcal{X}} \|\mathbf{D}^2 f(x)\|_{\ell_2 \to \ell_2}} \Big)\Big).$$

Estimates of this type are counterparts of the well known Hanson–Wright inequality for quadratic forms in independent subgaussian random variables (see [113, 134]).

Let us also mention that if $\mathcal{X} = \mathbb{R}^n$, D_i's correspond to the usual partial derivatives, X is a standard Gaussian vector and f is a polynomial of degree d

then the inequalities of Corollary 2.5.3 can be reversed (up to constants depending on d) [8]. The fact that in this case the assumptions of the corollary are satisfied was proved for the first time by Maurey and Pisier (see, e.g., [172]). Other continuous type examples for which this assumption is satisfied are described in [8, 3] by means of various types of modified log-Sobolev inequalities corresponding to tail behavior between exponential and Gaussian. As announced, we will not describe in details such examples. Instead, we will now focus on the discrete case and discuss a general situation, related to applications considered in previous sections, in which one can find gradients D_i satisfying (2.5.2).

2.5.2 Discussion on the choice of gradients

We will consider the following setting. Let G be a group with a set of generators $g_1, g_2, \ldots, g_m$, acting on a countable set $\mathcal{X}$ (we will denote the result of the action of $g \in G$ on $x \in \mathcal{X}$ simply by gx). Assume that $g_1, \ldots, g_m$ are pairwise distinct, distinct from the neutral element of G (denoted by e) and that no two distinct elements among the g_i's are reciprocal to each other.

Let μ be a probability measure on $\mathcal{X}$. Set $A = \{g_i, g_i^{-1} \colon i \leq m\}$ and let $\lambda \colon \mathcal{X} \times A \to [0, \infty)$ be a function satisfying the detailed balance condition

$$\lambda(x, g)\mu(x) = \lambda(gx, g^{-1})\mu(gx) \tag{2.5.3}$$

for $x \in \mathcal{X}$, $g \in A$.

Finally, consider the Markov process with the generator

$$Lf(x) = \sum_{g \in A} (f(gx) - f(x))\lambda(x, g)$$

and the corresponding Dirichlet form

$$\mathcal{E}(f, h) = \frac{1}{2} \sum_{x \in \mathcal{X}} \sum_{g \in A} (f(gx) - f(x))(h(gx) - h(x))\lambda(x, g)\mu(x) = \mathbb{E}\,\Gamma(f, h)$$

with

$$\begin{aligned}\Gamma(f, h)(x) &= \frac{1}{2} \sum_{g \in A} (f(gx) - f(x))(h(gx) - h(x))\lambda(x, g) \\ &= \frac{1}{2} \int_{\mathcal{X}} (f(y) - f(x))(h(y) - h(x))Q_x(dy),\end{aligned}$$

where

$$Q_x(y) = \sum_{g \in A \colon gx = y} \lambda(x, g)$$

Moreover the pair (Q, μ) satisfies the detailed balance condition. Therefore, if the form $\mathcal{E}$ satisfies Beckner's inequality with $\alpha_p \geq a(p-1)^s$, for $p \in (1, 2]$, then by Proposition 2.3.3 for all functions $f \colon \mathcal{X} \to \mathbb{R}$, and all $r \geq 2$

$$\|f - \mu(f)\|_r \leq K r^{\frac{1+s}{2}} \Big\| \Big(\sum_{g \in A} (f(gx) - f(x))^2 \lambda(x, g) \Big)^{1/2} \Big\|_r, \tag{2.5.4}$$

for $K = \sqrt{\kappa(s)/2a}$.

The above inequality allows for a direct use of Corollary 2.5.3 with gradients $D_g f(x) = (f(gx) - f(x))\sqrt{\lambda(x, g)}$. This choice however may have some disadvantages from the point of view of higher order concentration, especially when one deals with finite groups. To illustrate this let us focus on the situation when for some $g \in A$, $g^2 = e$ (in the sequel we will discuss natural examples when this is true for all elements of A). One then gets

$$\begin{aligned} D_g D_g f(x) &= D_g(f(gx) - f(x))\sqrt{\lambda(x,g)} \\ &= \Big((f(x) - f(gx))\sqrt{\lambda(gx,g)} - (f(gx) - f(x))\sqrt{\lambda(x,g)}\Big)\sqrt{\lambda(x,g)} \\ &= (f(x) - f(gx))(\sqrt{\lambda(x,g)} + \sqrt{\lambda(gx,g)})\sqrt{\lambda(x,g)}. \end{aligned}$$

In particular, if M is the constant from Corollary 2.5.3, it may happen that

$$M^2\|\mathbf{D}^2 f(x)\|_{\{1,2\}} \geq M\|\mathbf{D} f(x)\|_{\{1\}} = M|\mathbf{D} f(x)|$$

and so Corollary 2.5.3 applied with $d = 2$ is strictly weaker than its assumption corresponding to $d = 1$ (while the goal of introducing second order concentration is handling functions for which first order bounds are too conservative). Also, in certain situations, especially when dealing with a class of processes or when one does not have full knowledge about the transition rates, one may want to have a notion of gradient, which depends only on the function f and not on the rates $\lambda(x, g_i)$. For these reason one may want to replace the natural choice of the gradient with another one. We will now briefly discuss some possibilities.

Let us view the set $\mathcal{X}$ with the action G as a graph, i.e., define the set of edges $E = \{\{x,y\} \colon x, y \in \mathcal{X}, \exists_{i\leq m} y = g_i x \text{ or } x = g_i y\}$. Impose also an arbitrary orientation on the edges, by choosing functions $s, t \colon E \to \mathcal{X}$ such that for all $\{x,y\} \in E$, $\{s(x,y), t(x,y)\} = \{x,y\}$. Then one can define for $g \in A$,

$$D_g f(x) = \big(f(t(x,gx)) - f(s(x,gx))\big)\sqrt{\max(\lambda(x,g), \lambda(gx,g^{-1}))}.$$

Clearly, by (2.5.4), we then have

$$\|f(X) - \mathbb{E} f(X)\|_r \leq K r^{\frac{1+s}{2}} \Big\| |\mathbf{D} f(X)| \Big\|_r.$$

Moreover, for $g^2 = e$, $D_g f(x) = D_g f(gx)$ and so $D_g D_g f(x) = 0$. If $\lambda^* := \sup_{x\in\mathcal{X}} \max_{g\in A} \lambda(x,g) < \infty$ one can also take

$$\widetilde{D}_g f(x) = f(t(x,gx)) - f(s(x,gx)),$$

obtaining a gradient independent of the transition kernel, at the cost of changing the constant K by a factor $\sqrt{\lambda^*}$. Such situation may happen especially in the finite case, when the Markov semigroup is obtained by embedding in continuous time a discrete time Markov chain as, e.g., in the case of Glauber dynamics (cf. Section 2.4.3).

For instance if $\mathcal{X} = \{-1,1\}^m$ and g_i's act on $\mathcal{X}$ by flipping the i-th coordinate then all g_i's satisfy $g_i^2 = e$. In this case it is natural to choose $t(x,y) = \max(x,y)$, $s(x,y) = \min(x,y)$ where $\max, \min$ are taken with respect to the lexicographic order. It is then easy to see that in this case $\widetilde{D}_{g_i} f$ coincides, up to a factor of 2, with the usual partial derivative of the polynomial corresponding to the Fourier-Walsh representation of the function f. The article [4] uses the strategy described above to obtain counterparts of Latała's inequalities for Gaussian polynomials for polynomials in Ising models satisfying the Dobrushin condition discussed in Section 2.4.3. We will generalize the inequalities obtained therein in Corollary 2.5.4 below.

Another situation in which all the g_i's are of order two is related to transpositions, and the action of the symmetric group, corresponding to moment inequalities described in Propositions 2.4.8 and 2.4.12 (iii). In this case it is also natural to use the lexicographic order to define the functions t and s. Note that in this case we have $\binom{n}{2}$ directional derivatives along transpositions. For instance in the setting of Proposition 2.4.8, one has

$$D_{\tau_{ij}} f(x) = \frac{1}{n-1}\Big(f(\max(\sigma \circ \tau_{ij}, \sigma)) - f(\min(\sigma \circ \tau_{ij}, \sigma))\Big).$$

In the more general situation of Proposition 2.4.12 (ii) one can consider the action of the group generated by flips and swaps of coordinates on the discrete cube. Another types of estimates related to this example are presented in Chapter 4, where flip-swap random walks are examined. Their dynamics is expressed in terms of transpositions and changes on particular coordinates but the usual discrete gradient is considered. One may also consider actions of infinite groups, for instance $\mathbb{Z}$ or more generally $\mathbb{Z}^n$, as Beckner and log-Sobolev inequalities are well known for various measures on $\mathbb{Z}$, e.g., for stationary measures of various birth-and-death processes (see [58, 77, 54, 154, 25]). In particular, Chapter 3 is devoted to studying p-log-Sobolev inequalities on $\mathbb{N}$.

2.5.3 Applications to tetrahedral polynomials

Let us now provide an application of Corollary 2.5.3 to tetrahedral polynomials of random vectors with values in a cube, say $[-1,1]^n$, for which the corresponding Glauber dynamics satisfies (mLSI) (as discussed in Section 2.4.3). This will generalize results in [99, 4] concerning the Ising model and results from [100] concerning exponential random graphs (and by the discussion in Section 2.4.3 will allow also for a slight strengthening of the dependence of constants on the parameters in these models). Recall that a polynomial $f \colon \mathbb{R}^n \to \mathbb{R}$ is tetrahedral, if it is affine in every variable, i.e., it is of the form

$$f(x_1, \ldots, x_n) = \sum_{k=0}^{d} \sum_{I \subseteq [n], |I| = k} a_I \prod_{j \in I} x_i.$$

Assume that $X = (X_1, \ldots, X_n)$ is distributed according to a measure μ on $[-1,1]^n$, which satisfies (mLSI) with constant $\rho_0 > 0$ for the Glauber dynamics. Recall Corollary 2.4.5. Let $\mathcal{A}$ be the linear space of tetrahedral polynomials and note that for any $f \in \mathcal{A}$, the inequality $|X_i - X_i'| \le 2$ implies that

$$|f(X) - f(X^i)| \le 2\Big|\frac{\partial f}{\partial x_i}(X)\Big|.$$

As a consequence, by Corollary 2.4.5, the assumptions of Corollary 2.5.3 are satisfied with $\mathbf{D} = \nabla$, $\gamma = 1/2$ and $K = C\rho_0^{-1}$ for some universal constant C. Noting that partial derivatives of tetrahedral polynomials are tetrahedral, and for a polynomial f of degree d, $\nabla^d f$ is constant, we obtain the following result.

Corollary 2.5.4. *Assume that μ is a probability measure on $[-1,1]^n$, satisfying the inequality (mLSI) with $\rho_0 > 0$ for the Glauber dynamics. Let X be a random vector with law μ and let $f \colon \mathbb{R}^n \to \mathbb{R}$ be a tetrahedral polynomial of degree d. Then for any $t > 0$,*

$$\mathbb{P}\Big(\Big|f(X) - \mathbb{E}\, f(X)\Big| \ge t\Big) \le 2\exp\Big(-\frac{1}{C_d}\min_{1\le k\le d}\min_{\mathcal{J}\in P_k}\Big(\frac{\rho_0^{k/2} t}{\|\mathbb{E}\,\nabla^k f(X)\|_{\mathcal{J}}}\Big)^{2/|\mathcal{J}|}\Big). \tag{2.5.5}$$

In particular the above corollary applies to the Ising model, exponential random graphs and hardcore models under the assumptions presented in Section 2.4.3. Note that in these cases (or more generally for measures supported on $\{-1,0,1\}^n$) every polynomial can be reduced to a tetrahedral one.

Let us illustrate the above corollary with an application to triangle counts in random graphs. Consider a simple random graph $G = (V, E)$, where $|V| = n$. For distinct vertices $v, w \in V$, let $X_{v,w} = \mathbf{1}_{\{v,w\}\in E}$. Then, the number of triangles in G can be written as $T = \frac{1}{6}\sum_{u,v,w\in V}^{\neq} X_{u,v}X_{v,w}X_{w,u}$. The problem of tail behavior for subgraph counts in Erdős–Rényi random graphs has a long history, and a lot of progress has been made recently in the large deviation regime (see, e.g., [64, 144]). Here, we would like to illustrate how Corollary 2.5.4 may be used to obtain bounds for the non-product case.

Assume that G is exchangeable in the following sense: for any permutation $\sigma\colon V \to V$, the distribution of the random set $E_\sigma = \{\{\sigma(v), \sigma(w)\} \colon \{v, w\} \in E\}$ is the same as that of E (in other words, the adjacency matrix of G has distribution invariant under a simultaneous permutation of rows and columns). Assuming that $n \geq 3$ and $V = [n]$, let us define $A = \mathbb{E}\, X_{1,2}$, $B = \mathbb{E}\, X_{1,2}X_{2,3}$ (i.e., A is the probability of finding an edge and B is the probability of finding a *cherry* at a fixed place in G). If the random vector $(X_{v,w})_{1\leq v<w\leq n}$ satisfies (mLSI), then one can estimate the $\mathcal{I}$-norms appearing in Corollary 2.5.4 to find a tail estimate for T. Note that the expected derivatives of T are expressed in terms of A and B.

In the Erdős–Rényi case such estimates were obtained in [8] and combined with a slight refinement of (2.5.5), specific to the product situation which allowed to replace ρ_0 with an appropriate subgaussian norm. Surprisingly, the inequality obtained from such a general approach turned out to be optimal in the large deviation regime for edge probability $p \geq (n^{1/4}\sqrt{\log(n)})^{-1}$. In [100] the calculations from [8] were combined with (LSI) in the case of exponential random graphs (cf. Section 2.4.3). The dependence on ρ_1 is not specified there, but one can easily derive it from the proof. Corollary (2.5.4) allows to relax the dependence on ρ_1 to dependence on ρ_0, leading to the following estimate (we skip the detailed calculations, which are based on those from [8] and present just the final result):

$$
\begin{aligned}
&\mathbb{P}(|T - \mathbb{E}\, T| \geq t) \\
&\leq 2\exp\Big(-\frac{1}{C}\min\Big(\frac{t^2}{n^3(\rho_0^{-3} + \rho_0^{-2}A^2) + n^4\rho_0^{-1}B^2}, \frac{t}{\sqrt{n}\rho_0^{-3/2} + n\rho_0^{-1}A}, \frac{t^{2/3}}{\rho_0^{-1}}\Big)\Big)
\end{aligned}
$$

for $t \geq 0$ and a universal constant C.

Chapter 3

P-log-Sobolev inequalities

3.1 Background

Functional inequalities are one of the central objects of modern probability theory. They arise naturally when studying mixing times of Markov chains and are an important tool in proving concentration and hypercontractive estimates. Arguably, the most prominent examples are the Poincaré and log-Sobolev inequalities since most of the other functional inequalities known in the literature arise either as a modification of one of them (e.g., various modified log-Sobolev inequalities or inequalities with defects) or as a product of some sort of procedure that interpolates between the above two.

An important example of the latter type is a family of Beckner inequalities investigated in Chapter 2. Another particular example is a family of p-log-Sobolev inequalities introduced by Gross [108] for $p > 1$. His definition was then extended by Bakry [22] to any real p. Mossel et al. [158] studied p-log-Sobolev inequalities in the context of reverse hypercontraction.

A vibrant area of research is the study of the relations between various inequalities. E.g., it is by now classical that the log-Sobolev inequality implies the Poincaré inequality. Such sort of results can often have far-reaching consequences as demonstrated, e.g., in [158], where the authors prove that for $0 \leq q < p \leq 2$, the p-log-Sobolev inequality is stronger than the q-log-Sobolev inequality (cf. Theorem 3.1.6 below). They exploit this relation further to prove the reverse hypercontraction for measures satisfying the modified log-Sobolev inequality, which found various applications, cf., e.g., [174, 110]. A similar situation occurred in Chapter 2, where we have proved that the modified log-Sobolev inequality implies a particular family of Beckner inequalities, which allows deriving strong concentration and moment bounds from the modified log-Sobolev inequality. Both of the mentioned results are examples of positive results (i.e., claiming that one inequality implies the other) and follow from abstract arguments that involve direct comparison of Dirichlet forms. Similar results can be found in, e.g., the works of Diaconis–Saloff-Coste [81] or Bobkov–Tetali [42].

Another approach needs to be taken in the case of negative results, i.e., when showing that one inequality does not imply another. E.g., it is classical that the Poincaré inequality does not imply the log-Sobolev inequality. In that case, one counterexample is the exponential measure which satisfies the Poincaré inequality and does not satisfy the log-Sobolev inequality as the latter implies subgaussian concentration. To prove such statements, it is often useful to derive some characterization of the functional inequality in question. These characterizations are often of independent interest, as the conditions they are expressed in are usually much more accessible than a direct proof of the functional inequalities. Some of the results that provide such characterizations include: the work by Bobkov and Götze [37] who, by viewing the log-Sobolev inequality in the general framework of Orlicz spaces, characterize the log-Sobolev inequality on $\mathbb{R}$; the work by Miclo [154] who uses Hardy inequalities on $\mathbb{Z}$ to characterize the Poincaré inequality on trees or the work by Barthe–Roberto [26] who treat the case of the modified log-Sobolev inequalities on $\mathbb{R}$.

The aim of this chapter is to answer an open question from [158] on the relation between p-log-Sobolev inequalities for $p \in (0,1]$, cf. Problem 3.1.1 below. Our result falls into the category of negative examples described above. As a byproduct, we develop a sufficient condition and complement it with a closely related necessary condition for p-log-Sobolev inequalities, which are of independent interest. Below we describe our setting and results in more detail.

3.1.1 General setup

Let $(\Omega, \mathcal{B}, \mu)$ denote some discrete probability space. We assume that μ is fully supported on Ω. Let $P\colon [0,\infty) \times \Omega \times \mathcal{B} \to [0,1]$ be a homogeneous Markov transition function for which μ is an invariant measure. We assume that P is reversible with respect to μ and that it induces a strongly continuous semigroup $(P_t)_{t\geq 0}$ of operators on $L_2(\Omega,\mu)$, defined as

$$P_t f(x) = \int_\Omega f(y) P(t,x,dy).$$

It can be then shown that for each $f \in L_2(\Omega,\mu)$, the mapping

$$(0,\infty) \ni t \mapsto \frac{1}{2t}\int_\Omega\int_\Omega (f(x)-f(y))^2 P(t,x,dy)\mu(dx)$$

is non-increasing. Denote

$$\mathcal{H} = \left\{ f \in L_2(\Omega,\mu) : \sup_{t\geq 0} \frac{1}{2t}\int_\Omega\int_\Omega (f(x)-f(y))^2 P(t,x,dy)\mu(dx) < \infty \right\} \tag{3.1.1}$$

to be a domain of the Dirichlet form $\mathcal{E}$ associated with this semigroup, given by the formula

$$\mathcal{E}(f,g) = \lim_{t\to 0^+} \frac{1}{2t}\int_\Omega\int_\Omega (f(x)-f(y))(g(x)-g(y)) P(t,x,dy)\mu(dx) \tag{3.1.2}$$

for $f, g \in \mathcal{H}$. If L is the infinitesimal generator of the semigroup $(P_t)_{t\geq 0}$, defined via

$$Lf = \lim_{t\to 0^+} \frac{P_t f - f}{h}$$

with the convergence in the L_2 sense, then for f, g belonging to the domain $\mathcal{H}_L$ of L,

$$\mathcal{E}(f,g) = -\int fLg \, d\mu.$$

We refer the Reader to [67, 90, 142, 21, 9, 6] for a detailed treatment of Markov processes, generators, Dirichlet forms and associated domains.

Set $\mathcal{H}_+ = \mathcal{H} \cap \mathbb{R}_+^\Omega$, where $\mathbb{R}_+ = (0,\infty)$. Below we gather some notation and results from [158]. Note that while the results in [158] are stated for finite spaces, the extension to general discrete spaces is straightforward as discussed in Section 12.3 therein. Moreover, in order to be consistent with [158] and most of the relevant literature regarding characterization of functional inequalities [26, 154, 37], we change the convention from the remainder of this thesis, and throughout this chapter we keep the constants on the right-hand sides of functional inequalities in question.

Definition 3.1.1. For $p \in \mathbb{R} \setminus \{0,1\}$, the p-log-Sobolev inequality is satisfied with constant $C > 0$ if

$$\mathrm{Ent}_\mu(f^p) \leq \frac{Cp^2}{4(p-1)}\mathcal{E}(f^{p-1}, f), \tag{3.1.3}$$

for all $f \in \mathcal{H}_+$, such that $f^{p-1} \in \mathcal{H}_+$.

The 1-log-Sobolev inequality is satisfied with constant $C > 0$ if

$$\mathrm{Ent}_\mu(f) \leq \frac{C}{4}\mathcal{E}(f, \log f) \tag{3.1.4}$$

for all $f \in \mathcal{H}_+$, such that $\log f \in \mathcal{H}$.

The 0-log-Sobolev inequality is satisfied with constant $C > 0$ if

$$\mathrm{Var}_\mu(\log f) \leq \frac{C}{2}\mathcal{E}(f, -1/f) \tag{3.1.5}$$

for all $f \in \mathcal{H}_+$, such that $1/f \in \mathcal{H}_+$.

The Poincaré inequality is satisfied with constant $C > 0$ if

$$\mathrm{Var}_\mu(f) \leq \frac{C}{2}\mathcal{E}(f, f) \tag{3.1.6}$$

for all $f \in \mathcal{H}$.

We write p-LS(C) for short to denote the p-log-Sobolev inequality with constant C. We say that a pair $(\mu, \mathcal{E})$ satisfies the p-log-Sobolev inequality if p-LS(C) is satisfied with some finite $C > 0$. If the underlying Dirichlet form $\mathcal{E}$ is clear from context, we omit it and simply say the μ satisfies the p-log-Sobolev inequality. Note that the 1-log-Sobolev inequality is often referred to as the *modified* log-Sobolev (or entropic) inequality in the literature and that the 0-log-Sobolev and 1-log-Sobolev inequalities are limiting cases for p-log-Sobolev inequalities for $p \in \mathbb{R} \setminus \{0,1\}$.

Results below provide some basic relations between the inequalities introduced in Definition 3.1.1.

Proposition 3.1.2 ([158, Lemma 3.1])**.** *The Poincaré inequality* (3.1.6) *with constant $C > 0$ is equivalent to the 0-log-Sobolev inequality with the same constant C.*

We set $p' = p/(p-1)$ for the Hölder conjugate of $p \in \mathbb{R} \setminus \{1\}$.

Proposition 3.1.3 ([158, Lemma 3.2])**.** *For $p \in \mathbb{R} \setminus \{1\}$, p-LS(C) is equivalent to p'-LS(C).*

In particular, Proposition 3.1.3 implies that the study of relations between various p-log-Sobolev inequalities can be reduced to the case $p \in [0,2]$. Moreover, since in the continuous setting (in the presence of the chain rule) all p-log-Sobolev inequalities for $p \neq 0$ are equivalent to the usual log-Sobolev inequality[1] and since by Proposition 3.1.2, 0-log-Sobolev inequality is equivalent to the Poincaré inequality which is strictly weaker than the log-Sobolev inequality, the only interesting range of p is in fact $p \in (0, 2]$.

Denote

$$\mathcal{E}_p(f) = \begin{cases} pp'\mathcal{E}(f^{1/p}, f^{1/p'}) & \text{if } p \in (0,2] \setminus \{1\}, \\ \mathcal{E}(f, \log f) & \text{if } p = 1, \end{cases} \tag{3.1.7}$$

so that p-LS(C) for $p \in (0, 2]$ is equivalent to

$$\mathrm{Ent}_\mu(f) \leq \frac{C}{4}\mathcal{E}_p(f).$$

Proposition 3.1.4 ([158, Theorem 2.1]). *For any positive f, the mapping $(0, 2] \ni p \mapsto \mathcal{E}_p(f)$ is non-increasing.*

Remark 3.1.5. Note that [158, Theorem 2.1] is stated for $p \in (0, 2] \setminus \{1\}$ but a natural extension to the case $p = 1$ is straightforward as discussed in Remark 2.2 therein.

Proposition 3.1.4 serves as a tool for obtaining the following main result of [158].

Theorem 3.1.6 ([158, Theorem 1.7]). *For any $0 \leq q \leq p \leq 2$, p-LS(C) implies q-LS(C). Moreover, for any $1 < q \leq p \leq 2$, q-LS(C) implies p-LS($\tilde{C}$) with $\tilde{C} = Cqq'/pp'$.*

The main goal of this chapter is to provide an answer to the following question posed in [158, Section 12].

Problem. Is there any subset $I \subset (0, 1]$ with non-empty interior, such that for any $p, q \in I$, p-LS(C) implies q-LS($c(I)C$), where $c(I) > 0$ depends on I only?

3.2 Main result

Let μ be a measure on $\mathbb{N}$ with full support. We use the convention that $\mu_k = \mu(\{k\})$ and $\mu[k, \infty) = \mu([k, \infty))$. For any $f : \mathbb{N} \to \mathbb{R}$ and $k \in \mathbb{N}$, denote $Df(k) = f(k + 1) - f(k)$. Consider the following Dirichlet form

$$\mathcal{E}(f, g) = \sum_{k \geq 0} Df(k)Dg(k)\mu_k \tag{3.2.1}$$

defined for $f, g \in \{h \in L_2(\Omega, \mu) : \sum_{k\geq 0}(Dh(k))^2\mu_k < \infty\} =: \mathcal{H}$. The corresponding birth-death dynamics generator is given by the formula

$$Lf(k) = Df(k) - \mathbf{1}_{\{k>0\}}\frac{\mu_{k-1}}{\mu_k}Df(k - 1), \tag{3.2.2}$$

so that $\mathcal{E}(f, g) = -\int fLg\, d\mu$ for $f, g \in \mathcal{H}_L \subset \mathcal{H}$. A Markov process generated by L given by (3.2.2) and Dirichlet form given by (3.2.1) with stationary measure μ being a geometric measure is investigated in [77] as an example of a pair $(\mu, \mathcal{E})$ satisfying the Poincaré inequality (3.1.6) and violating the modified log-Sobolev inequality (3.1.4).

[1] This can be seen by substituting $f^2 \leftarrow f^p$ in the definition of the p-log-Sobolev inequality for $p \in \mathbb{R} \setminus \{0\}$.

Remark 3.2.1. It follows from the general theory of birth and death processes (see, e.g., [67, 142]), that $\mathcal{E}$ is indeed a Dirichlet form corresponding to a Markov process on $\mathbb{N}$. As a consequence, when proving a functional inequality in question it actually suffices to consider simple (i.e., having finitely many jumps) functions. Therefore, henceforth we will sometimes restrict from specifying particular domains and simply assume that the expressions we introduce are considered for functions for which they are well-defined as one can always restrict the attention to the class of simple functions.

For $x > 0$, set

$$H_p(x) = \begin{cases} pp'(x^{1/p} - 1)(x^{1/p'} - 1) & \text{if } p \in (0,1), \\ (x-1)\log(x) & \text{if } p = 1 \end{cases} \tag{3.2.3}$$

so that $\mathcal{E}_p$ defined in (3.1.7) is given by

$$\mathcal{E}_p(f) = \sum_{k\geq 0} f(k) H_p\Big(\frac{f(k+1)}{f(k)}\Big)\mu_k \tag{3.2.4}$$

and p-LS(C) for $p \in (0,1]$ is equivalent to

$$\mathrm{Ent}_\mu(f) \leq \frac{C}{4}\sum_{k\geq 0} f(k) H_p\Big(\frac{f(k+1)}{f(k)}\Big)\mu_k, \tag{3.2.5}$$

while 0-LS(C) (i.e., the Poincaré inequality) is equivalent to

$$\mathrm{Var}_\mu(f) \leq \frac{C}{2}\sum_{l=0}^{\infty}(Df(l))^2\mu_l. \tag{3.2.6}$$

Theorem below is our main result.

Theorem 3.2.2. *For any $p \in (0,1)$, there exists a measure μ on $\mathbb{N}$ that does not satisfy the p-log-Sobolev inequality but satisfies the q-log-Sobolev inequality for all $q \in (0,p)$. In particular, there are no intervals I that meet the conditions posed in Problem 3.1.1.*

We construct the required counterexample and verify that it satisfies the appropriate p-log-Sobolev inequalities with the use of the theorem below, which is of independent interest.

Theorem 3.2.3. *Choose any $p \in (0,1]$ and a measure μ on $\mathbb{N}$ with full support and with associated Dirichlet form $\mathcal{E}$, given by the birth-death process generator (3.2.1).*

If μ satisfies the Poincaré inequality (3.2.6) with some finite constant $C_P > 0$ and

$$\hat{C} := \sup_{n\geq 1}\Big\{\Big[H_p\Big(\frac{\mu[n-1,\infty)}{\mu[n,\infty)}\Big)\Big]^{-1}\cdot\log\Big(\frac{2}{\mu[n,\infty)}\Big)\Big\} < \infty, \tag{3.2.7}$$

then μ satisfies p-LS(C) (3.2.5) with some finite $C > 0$.

Contrarily, if there exists an increasing sequence $\tau_0 < \tau_1 < \dots$ such that

$$\lim_{n\to\infty}\Big\{\Big[H_p\Big(\frac{\mu[\tau_{n-1},\infty)}{\mu[\tau_n,\infty)}\Big)\Big]^{-1}\cdot\frac{\mu[\tau_{n-1},\infty)}{\mu[\tau_n-1,\infty)}\cdot\log\Big(\frac{2}{\mu[\tau_n,\infty)}\Big)\Big\} = \infty, \tag{3.2.8}$$

then μ does not satisfy p-LS(C) (3.2.5) with any finite $C > 0$.

Remark 3.2.4. The Poincaré inequality is implied by the p-log-Sobolev inequality for any $p \in (0, 2]$, cf. Theorem 3.1.6 and Proposition 3.1.2, therefore making it a part of the sufficient condition of Theorem 3.2.3 (alongside (3.2.7)) is non-restrictive.

Remark 3.2.5. The negation of (3.2.7) is equivalent to the existence of an increasing sequence $\tau_0 < \tau_1 < \dots$ such that

$$\lim_{n\to\infty} \Big\{ \Big[H_p\big(\frac{\mu[\tau_n - 1, \infty)}{\mu[\tau_n, \infty)}\big) \Big]^{-1} \cdot \log\big(\frac{2}{\mu[\tau_n, \infty)}\big) \Big\} = \infty.$$

If μ verifies the Poincaré inequality (3.2.6) with constant C_P, then by Lemma 3.3.4 (cf. also Remark 3.3.3) and by Lemma 3.3.10, (v) below, for any $\tau_{n-1} < \tau_n$ and some $c > 0$,

$$\Big[H_p\big(\frac{\mu[\tau_n - 1, \infty)}{\mu[\tau_n, \infty)}\big) \Big]^{-1} \geq \frac{1}{1+c} \Big[H_p\big(\frac{\mu[\tau_{n-1}, \infty)}{\mu[\tau_n, \infty)}\big) \Big]^{-1} \cdot \frac{\mu[\tau_{n-1}, \infty)}{\mu[\tau_n - 1, \infty)}.$$

Whence, condition (3.2.8) implies that $\hat{C} = \infty$ and thus (3.2.7) does not hold, but the reverse needs not to be true. Thus, providing a full characterization of p-log-Sobolev inequalities on $\mathbb{N}$ for $p \in (0, 1]$ remains open.

3.3 Auxiliary results

In this section we gather some lemmas and known results required for the proof of Theorem 3.2.3.

3.3.1 Hardy inequality

In the sequel we put $0 \cdot \infty = 0$ and $c/0 = \infty$ for any $c > 0$.

Definition 3.3.1. We say that a probability measure μ on $\mathbb{N}$ satisfies the Hardy inequality with constant C, if

$$\sum_{l=0}^{\infty} (f(l) - f(0))^2 \mu_l \leq C \sum_{l=0}^{\infty} (Df(l))^2 \mu_l = C\mathcal{E}(f, f) \tag{3.3.1}$$

for all $f \in \mathcal{H}$.

Define

$$C_\mu = \sup_{k \geq 1} \Big\{ \mu[k, \infty) \cdot \sum_{l=0}^{k-1} \frac{1}{\mu_l} \Big\}. \tag{3.3.2}$$

The following result states that C_μ characterizes the Hardy inequality (3.3.1).

Theorem 3.3.2 (Miclo [154])**.** *The best constant $\hat{C}_H$ in the Hardy inequality (3.3.1),*

$$\hat{C}_H = \sup_{f \in \mathcal{H}} \Big\{ \frac{\sum_{l=0}^{\infty} (f(l) - f(0))^2 \mu_l}{\mathcal{E}(f, f)} : \mathcal{E}(f, f) > 0 \Big\}, \tag{3.3.3}$$

verifies

$$C_\mu \leq \hat{C}_H \leq 4C_\mu.$$

Remark 3.3.3. It is easy to see that for fully supported measures, the Poincaré inequality (3.2.6) is satisfied if and only if the Hardy inequality (3.3.1) is satisfied. More precisely, the best constant $\hat{C}_P$ in the Poincaré inequality (3.2.6),

$$\hat{C}_P = 2 \sup_{f \in \mathcal{H}} \Big\{ \frac{\mathrm{Var}_\mu(f)}{\mathcal{E}(f,f)} : \mathcal{E}(f,f) > 0 \Big\}, \tag{3.3.4}$$

satisfies

$$2\mu_0 \hat{C}_H \le \hat{C}_P \le 2\hat{C}_H.$$

Indeed, $\hat{C}_P \le 2\hat{C}_H$ follows from the estimate $\mathrm{Var}_\mu(f) \le \int (f - f(0))^2 \, d\mu$. To see that $2\mu_0 \hat{C}_H \le \hat{C}_P$, let $f_\varepsilon \in \mathcal{H}$ for any $\varepsilon > 0$ be such that $f_\varepsilon(0) = 0$, $\mathcal{E}(f_\varepsilon, f_\varepsilon) > 0$ and $\int f_\varepsilon^2 \, d\mu \ge (\hat{C}_H - \varepsilon)\mathcal{E}(f_\varepsilon, f_\varepsilon)$. Then, by the Cauchy–Schwarz inequality, $(\int f_\varepsilon \, d\mu)^2 \le (1 - \mu_0) \int f_\varepsilon^2 \, d\mu$, whence

$$\frac{\hat{C}_P}{2} \mathcal{E}(f_\varepsilon, f_\varepsilon) \ge \mathrm{Var}_\mu(f_\varepsilon) \ge \mu_0 \int f_\varepsilon^2 \, d\mu \ge \mu_0 (\hat{C}_H - \varepsilon) \mathcal{E}(f_\varepsilon, f_\varepsilon),$$

and we conclude by taking $\varepsilon \to 0^+$.

The quantity C_μ is useful for controlling the tail behavior of μ as demonstrated in the lemmas below.

Lemma 3.3.4. *If μ is fully supported, then*

$$\sup_{k \ge 0} \frac{\mu[k, \infty)}{\mu_k} \le 1 + C_\mu \tag{3.3.5}$$

and

$$\inf_{k \ge 1} \Big\{ \frac{\mu[k-1, \infty)}{\mu[k, \infty)} \Big\} \ge 1 + \frac{1}{C_\mu}. \tag{3.3.6}$$

Proof. The estimate (3.3.5) follows from the definition (3.3.2) of C_μ. Using (3.3.5), we obtain for any $k \ge 1$,

$$\begin{aligned} \mu[k, \infty) &= \mu[k-1, \infty) - \mu_{k-1} \\ &\le \mu[k-1, \infty) - \frac{\mu[k-1, \infty)}{1 + C_\mu} = \Big(1 + \frac{1}{C_\mu}\Big)^{-1} \mu[k-1, \infty) \end{aligned}$$

and (3.3.6) follows. □

Lemma 3.3.5. *If μ is fully supported, then for $X \sim \mu$, $\mathbb{E}\, X \le 1 + C_\mu$.*

Proof. By Lemma 3.3.4, $\mathbb{E}\, X \le \sum_{k \in \mathbb{N}} \mu[k, \infty) \le (1 + C_\mu) \sum_{k \in \mathbb{N}} \mu_k = 1 + C_\mu$. □

3.3.2 Tail estimates

Proposition below states that the p-log-Sobolev inequality implies Poisson-type tail behavior for $p \in (0, 1]$. It is deduced by using a variant of Herbst's argument.

Proposition 3.3.6. *If μ is fully supported and satisfies p-LS(C) (3.2.5) for some $p \in (0, 1]$ and some $C < \infty$, then there exists $\varepsilon_p \colon (0, \infty) \to [0, 1]$, such that $\varepsilon_p(t) \to 0$ as $t \to \infty$ and*

$$\log\big(\mu[t, \infty)\big) \le -(1 - \varepsilon_p(t)) tp \log(t + 1) \tag{3.3.7}$$

for any $t \ge 0$.

For the proof of Proposition 3.3.6, we need the following lemma.

Lemma 3.3.7. *For any* $\lambda > 1$ *and* $p \in (0,1)$,

$$\int_1^\lambda \frac{e^{s/p}-1}{s^2}\,ds \le \frac{p}{1-p}\cdot\frac{e^{\lambda/p}}{\lambda} = -p'\cdot\frac{e^{\lambda/p}}{\lambda}. \tag{3.3.8}$$

Proof. For any $s > 1$,

$$\frac{1-p}{p}\cdot\frac{e^{s/p}-1}{s^2} \le \frac{1-p}{p}\cdot\frac{e^{s/p}}{s^2} \le \frac{s-p}{p}\cdot\frac{e^{s/p}}{s^2} = \frac{d}{ds}\Big[\frac{e^{s/p}}{s}\Big],$$

whence

$$\int_1^\lambda \frac{e^{s/p}-1}{s^2}\,ds \le \frac{p}{1-p}\Big(\frac{e^{\lambda/p}}{\lambda} - e^{1/p}\Big) \le \frac{p}{1-p}\frac{e^{\lambda/p}}{\lambda}$$

as desired. □

Proof of Proposition 3.3.6. Consider first the case $p \in (0,1)$. For $s,h > 0$, $x \in \mathbb{R}$, denote $\phi_h(x) = \min(x,h)$, $f_{s,h}(x) = \exp(s\phi_h(x)/p)$ and $g_{s,h}(x) = \exp(s\phi_h(x)/p')$. Let $X \sim \mu$ and for $h > 0$, set $X_h = \min(X,h)$. For any $s,h > 0$, $h \in \mathbb{N}$, applying p-LS(C) to $f_{s,h}$ and using the Dirichlet form formula (3.2.1), we obtain

$$\begin{aligned}\frac{d}{ds}\Big[\frac{\log \mathbb{E}\, e^{sX_h}}{s}\Big] &= \frac{\mathrm{Ent}(e^{sX_h})}{s^2\,\mathbb{E}\, e^{sX_h}}\\ &\le \frac{Cpp'}{4s^2\,\mathbb{E}\, e^{sX_h}}\mathcal{E}(f_{s,h}, g_{s,h})\\ &= \frac{Cpp'}{4s^2\,\mathbb{E}\, e^{sX_h}}\sum_{k=0}^{h-1}\big(e^{s(k+1)/p}-e^{sk/p}\big)\big(e^{s(k+1)/p'}-e^{sk/p'}\big)\mu_k\\ &= \frac{Cpp'}{4s^2}\big(e^{s/p}-1\big)\big(e^{s/p'}-1\big)\frac{\sum_{k=0}^{h-1} e^{sk/p+sk/p'}\mu_k}{\mathbb{E}\, e^{sX_h}}\\ &\le \frac{Cpp'}{4s^2}\big(e^{s/p}-1\big)\big(e^{s/p'}-1\big),\end{aligned} \tag{3.3.9}$$

where in the last inequality we have also used that $p'(e^{s/p'}-1) > 0$. By Theorem 3.1.6, μ satisfies the Poincaré inequality (3.2.6), whence by Lemma 3.3.5, $\mathbb{E}\, X < 1 + C_\mu < \infty$, cf. Remark 3.3.3. Consequently, for any $\lambda, h > 0$,

$$\begin{aligned}\frac{\log \mathbb{E}\, e^{\lambda X_h}}{\lambda} &= \mathbb{E}\, X_h + \int_0^\lambda \frac{d}{ds}\Big[\frac{\log \mathbb{E}\, e^{sX_h}}{s}\Big]\,ds\\ &\le \mathbb{E}\, X_h + \frac{Cpp'}{4}\int_0^\lambda \frac{\big(e^{s/p}-1\big)\big(e^{s/p'}-1\big)}{s^2}\,ds\\ &\le \mathbb{E}\, X + \frac{Cpp'}{4}\int_0^\lambda \frac{\big(e^{s/p}-1\big)\big(e^{s/p'}-1\big)}{s^2}\,ds < \infty.\end{aligned} \tag{3.3.10}$$

Taking $h \to +\infty$ and using monotone convergence theorem in (3.3.10), we obtain that for any $\lambda > 0$, $\mathbb{E}\, e^{\lambda X} < \infty$. Therefore, we can repeat reasoning from (3.3.9) to get that for any $s > 0$,

$$\frac{d}{ds}\Big[\frac{\log \mathbb{E}\, e^{sX}}{s}\Big] \le \frac{Cpp'}{4s^2}\big(e^{s/p}-1\big)\big(e^{s/p'}-1\big). \tag{3.3.11}$$

For any $\lambda > 1$, by (3.3.11), estimating $1 - e^{s/p'} \leq 1$ for $s > 1$ and using Lemma 3.3.7,

$$\int_1^\lambda \frac{d}{ds}\Big[\frac{\log \mathbb{E}\, e^{sX}}{s}\Big]\, ds \leq \frac{Cpp'}{4}\int_1^\lambda \frac{(e^{s/p}-1)(e^{s/p'}-1)}{s^2}\, ds$$
$$\leq \frac{Cp(-p')}{4}\int_1^\lambda \frac{e^{s/p}-1}{s^2}\, ds \leq \frac{Cpp'^2}{4\lambda} e^{\lambda/p},$$

whence, for $t > e^{1/p} - 1$, by the Chernoff bound and setting $\lambda = p\log(t+1) > 1$,

$$\begin{aligned}
\log\big(\mathbb{P}(X \geq t)\big) &\leq \log \mathbb{E}\exp(-\lambda t + \lambda X) \\
&= -\lambda t + \lambda \log \mathbb{E}\, e^X + \lambda \int_1^\lambda \frac{d}{ds}\Big[\frac{\log \mathbb{E}\, e^{sX}}{s}\Big]\, ds \\
&\leq -\lambda t + \lambda \log \mathbb{E}\, e^X + \frac{Cpp'^2}{4} e^{\lambda/p} \\
&= -pt\log(t+1) + p(\log \mathbb{E}\, e^X)\log(t+1) + \frac{Cpp'^2}{4}(t+1) \\
&= -\Big(1 - \frac{\log \mathbb{E}\, e^X}{t} - \frac{Cp'^2(t+1)}{4t\log(t+1)}\Big) tp\log(t+1).
\end{aligned}$$

Therefore, for any $p \in (0,1)$ and

$$\varepsilon_p(t) = \begin{cases} 1 & \text{if} \quad t \leq e^{1/p} - 1, \\ \min\big(1, \frac{\log \mathbb{E}\, e^X}{t} + \frac{C}{4}\cdot\frac{1}{(1/p-1)^2}\cdot\frac{t+1}{t\log(t+1)}\big) & \text{if} \quad t > e^{1/p} - 1, \end{cases} \tag{3.3.12}$$

the estimate (3.3.7) holds for any $t \geq 0$ as desired.

We turn to the case $p = 1$. Recall that by Theorem 3.1.6, for any $p \in (0,1)$ the 1-log-Sobolev inequality implies the p-log-Sobolev inequality with the same constant. Let ε_p be given by (3.3.12) and let $\varphi\colon (0,\infty) \to (0,1)$ be such that $\varphi(t) \to 1$ and $\varepsilon_{\varphi(t)}(t) \to 0$ as $t \to \infty$ (one can take, e.g., $\varphi(t) = 0.5$ for $t \leq e^e$ and $\varphi(t) = (1 + 1/\log\log t)^{-1}$ for $t > e^e$). For $t \in (0,\infty)$, set

$$\varepsilon_1(t) = 1 - \varphi(t)\big(1 - \varepsilon_{\varphi(t)}(t)\big).$$

Then $\varepsilon_1(t) \to 0$ as $t \to \infty$ and for any $t > 0$, we apply (3.3.7) with $p = \varphi(t)$ to obtain the conclusion. □

3.3.3 Reduction to increasing functions

The proposition below is an adaptation of [26, Proposition 3] to the discrete case. The proof goes along similar lines – we present it for completeness.

Proposition 3.3.8. *Recall the definition (3.3.2) of C_μ and assume $C_\mu < \infty$. For any $f\colon \mathbb{N} \to \mathbb{R}_+$ with a finite number of jumps and for any $\rho > 1$,*

$$\mathrm{Ent}_\mu(f) \leq 16 C_\mu (1 + \sqrt{\rho})^2 \mathcal{E}(\sqrt{f}, \sqrt{f}) + \sum_{k\colon g(k) \geq \rho\, \mathbb{E}_\mu g} g(k)\log\Big(\frac{g(k)}{\mathbb{E}_\mu g}\Big)\mu_k,$$

where $g(n) = f(0) + \sum_{k=0}^{n-1}(Df(k))_+$.

Proof. We use the assumption that f has a finite number of jumps to make sure that every quantity below is well-defined. Note that $g \geq f$ since $(Df(k))_+ \geq$

$Df(k)$ for all $k \in \mathbb{N}$. For $x, t > 0$, denote $\psi(t,x) = x \log \frac{x}{t} - (x-t)$, $\phi(x) = x \log x$ and set $\Theta = \{n \in \mathbb{N} \colon f(n) \geq \rho \mathbb{E}\, g\}$. Convexity of ϕ implies that $\phi(\mathbb{E}_\mu f) \geq \phi(\mathbb{E}_\mu g) + \phi'(\mathbb{E}_\mu g)(\mathbb{E}_\mu f - \mathbb{E}_\mu g)$, so that

$$\begin{aligned} \operatorname{Ent}(f) &= \mathbb{E}_\mu\big[\phi(f) - \phi(\mathbb{E}_\mu f)\big] \\ &\leq \mathbb{E}_\mu\big[\phi(f) - \phi(\mathbb{E}_\mu g) + \phi'(\mathbb{E}_\mu g)\cdot(\mathbb{E}_\mu g - \mathbb{E}_\mu f)\big] = \mathbb{E}_\mu \psi(\mathbb{E}_\mu g, f). \end{aligned} \tag{3.3.13}$$

Set $U_\rho = (1+\sqrt{\rho})^2$. Since for $x \in [0, \rho t]$,

$$\begin{aligned} \psi(t,x) \leq x\Big(\frac{x}{t} - 1\Big) - (x-t) = t\Big(\frac{x}{t} - 1\Big)^2 \\ = \Big(1 + \sqrt{\frac{x}{t}}\Big)^2(\sqrt{x} - \sqrt{t})^2 \leq U_\rho(\sqrt{x} - \sqrt{t})^2, \end{aligned}$$

we obtain

$$\begin{aligned} \mathbb{E}_\mu \psi(\mathbb{E}_\mu g, f)\mathbf{1}_{\Theta^c} &\leq U_\rho \mathbb{E}_\mu(\sqrt{f} - \sqrt{\mathbb{E}_\mu g})^2 \\ &\leq 2U_\rho\big(\mathbb{E}_\mu(\sqrt{f} - \sqrt{g})^2 + \mathbb{E}_\mu(\sqrt{g} - \sqrt{\mathbb{E}_\mu g})^2\big) \\ &\leq 2U_\rho\big(\mathbb{E}_\mu(\sqrt{f} - \sqrt{g})^2 + 2\operatorname{Var}_\mu(\sqrt{g})\big). \end{aligned} \tag{3.3.14}$$

Let $\hat{C}_H$ and $\hat{C}_P$ be the best constants in the Hardy inequality (3.3.1) and the Poincaré inequality (3.2.6) respectively, cf. the definitions (3.3.3) and (3.3.4). By Theorem 3.3.2 and as indicated in Remark 3.3.3, $\hat{C}_P \leq 2\hat{C}_H \leq 8C_\mu$. Recall that $\sqrt{f(0)} = \sqrt{g(0)}$, whence the RHS of (3.3.14) can be estimated from above by

$$\begin{aligned} 2U_\rho\big(4C_\mu \sum_{k=0}^{\infty}(D\sqrt{f} - D\sqrt{g})^2(k)\mu_k + 8C_\mu \sum_{k=0}^{\infty}(D\sqrt{g})^2(k)\mu_k\big) \\ \leq 16 C_\mu U_\rho \mathcal{E}(\sqrt{f}, \sqrt{f}), \end{aligned}$$

where we have used the fact that $4(D\sqrt{f} - D\sqrt{g})^2 + 8(D\sqrt{g})^2 \leq 8(D\sqrt{f})^2$, which follows from the pointwise estimate $0 \leq D\sqrt{g} \leq (D\sqrt{f})_+$ and the fact that for a fixed y, the convex mapping $x \mapsto 4(y-x)^2 + 8x^2$ on a closed interval admits a maximum at an end of this interval.

Turning to the remaining part of the RHS of (3.3.13), we get

$$\begin{aligned} \mathbb{E}_\mu \psi(\mathbb{E}\, g, f)\mathbf{1}_\Theta &= \mathbb{E}_\mu\big[f \log \frac{f}{\mathbb{E}_\mu g} - (f - \mathbb{E}_\mu g)\big]\mathbf{1}_\Theta \\ &\leq \mathbb{E}_\mu\big[g \log \frac{g}{\mathbb{E}_\mu g}\big]\mathbf{1}_\Theta \leq \sum_{k \colon g(k) \geq \rho \mathbb{E}_\mu g} g(k) \log \frac{g(k)}{\mathbb{E}_\mu g}\mu_k, \end{aligned}$$

where in both inequalities we have used the definition of Θ and the facts that $f \leq g$ and $\rho > 1$. Combining all the above estimates yields the conclusion. □

When proving that some p-log-Sobolev inequality is satisfied, Proposition 3.3.8 allows us to restrict our attention to a special subclass of functions from $\mathcal{H}_+$. This idea is formalized in the corollary below.

Corollary 3.3.9. *If $C_\mu < \infty$ and*

$$\sum_{k \colon g(k) \geq \rho \mathbb{E}_\mu g} g(k) \log\Big(\frac{g(k)}{\mathbb{E}_\mu g}\Big)\mu_k \leq C_\rho \mathcal{E}_p(g) \tag{3.3.15}$$

for some $\rho > 1$, $p \in (0,2]$, some constant $C_\rho > 0$ and any non-decreasing function $g\colon \mathbb{N} \to \mathbb{R}_+$ with a finite number of jumps, then μ satisfies the p-log-Sobolev inequality (3.2.5) with constant

$$C = 64C_\mu(1+\sqrt{\rho})^2 + 4C_\rho.$$

Proof. By applying Proposition 3.3.8 together with condition (3.3.15), we get that for any $f\colon \mathbb{N} \to \mathbb{R}_+$ with a finite number of jumps

$$\mathrm{Ent}_\mu(f) \le 16C_\mu(1+\sqrt{\rho})^2 \mathcal{E}(\sqrt{f}, \sqrt{f}) + C_\rho \mathcal{E}_p(g), \tag{3.3.16}$$

where $g(0) = f(0)$ and $g(n) = f(0) + \sum_{k=0}^{n-1} (Df(k))_+$ for $n \ge 1$.

Note that $f \le g$ by definition and that for any $x, \Delta > 0$, the mapping $x \mapsto xH_p(1 + \Delta/x)$ is non-increasing, which follows from the convexity of H_p on $[1, \infty)$, cf. Lemma 3.3.10, (i) below, and the fact that $H_p(1) = 0$. Using that $H_p \ge 0$ and monotonicity of $x \mapsto xH_p(1 + \Delta/x)$, we obtain

$$\begin{aligned}
\mathcal{E}_p(f) &= \sum_{k=0}^{\infty} f(k) H_p\Big(1 + \frac{Df(k)}{f(k)}\Big)\mu_k \\
&\ge \sum_{k=0}^{\infty} f(k) H_p\Big(1 + \frac{Df(k)}{f(k)}\Big)\mu_k \mathbf{1}_{\{Df(k)>0\}} \\
&\ge \sum_{k=0}^{\infty} g(k) H_p\Big(1 + \frac{Df(k)}{g(k)}\Big)\mu_k \mathbf{1}_{\{Df(k)>0\}} \\
&= \sum_{k=0}^{\infty} g(k) H_p\Big(1 + \frac{Dg(k)}{g(k)}\Big)\mu_k = \mathcal{E}_p(g).
\end{aligned} \tag{3.3.17}$$

By Proposition 3.1.4, $\mathcal{E}(\sqrt{f}, \sqrt{f}) \le \mathcal{E}_2(f) \le \mathcal{E}_p(f)$ for any $p \in (0,2]$. Therefore, combining (3.3.16) and (3.3.17), we obtain that for any $f\colon \mathbb{N} \to \mathbb{R}_+$ with a finite number of jumps,

$$\mathrm{Ent}_\mu(f) \le \big(16C_\mu(1+\sqrt{\rho})^2 + C_\rho\big)\mathcal{E}_p(f). \tag{3.3.18}$$

We conclude the result for arbitrary f by the monotone convergence theorem (applied to the RHS of (3.3.18)) and Fatou's lemma (applied to the LHS of (3.3.18)). $\square$

3.3.4 Technical lemmas

Recall the definition (3.2.3) of H_p and note that for $p \in (0,1)$ and $x > 0$, $H_p(x) = pp'(x - x^{1/p} - x^{1/p'} + 1)$.

Lemma 3.3.10. *For any $p \in (0,1]$, the following properties are true:*

(i) *H_p is increasing and convex on $[1, \infty)$;*

(ii) *$H_p(x) \ge (\log x)^2$ for $x \ge 1$.*

If additionally $x \ge \lambda$ for some $\lambda > 1$, then

(iii) *$H_p(x) \ge (x \log x) \cdot \min\{\frac{H_p(\lambda)}{\lambda \log \lambda}, \frac{H_p'(\lambda)}{1+\log \lambda}, 1\}$;*

(iv) *$H_p(x) \ge H_p(xc) \cdot \min\{\frac{H_p(\lambda)}{H_p(\lambda c)}, \frac{H_p'(\lambda)}{cH_p'(\lambda c)}, c^{-1/p}\}$ for any $c > 1$;*

(v) $\frac{H_p(y)}{y} \leq \frac{\lambda}{\lambda-1}\frac{H_p(x)}{x}$ *for any* $\lambda \leq y \leq x$.

Proof. We start with (i) and (ii).

For $p = 1$, $H_1(x) = (x-1)\log x$, $H_1(1) = H_1'(1) = 0$ and $H_1''(x) = \frac{1}{x^2}+\frac{1}{x} > 0$, yielding (i), while (ii) follows immediately as $\log x < x - 1$.

For $p \in (0,1)$, denote $h(x) = (\log x)^2$. Then, $H_p(1) = H_p'(1) = h(1) = h'(1) = 0$ and, using AM-GM inequality,

$$H_p''(x) = \frac{x^{1/p} + x^{1/p'}}{x^2} \geq \frac{2x^{1/2p+1/2p'}}{x^2} \geq \frac{2}{x^2} \geq 2\frac{1-\log x}{x^2} = h''(x), \tag{3.3.19}$$

whence (ii) follows. Moreover, (3.3.19) implies that $H_p''(x) \geq 0$, yielding also (i).

To see (iii), denote the function on its RHS by $\tilde{h}(x)$ and note that by the definition of $\tilde{h}$, $H_p(\lambda) \geq \tilde{h}(\lambda)$ and $H_p'(\lambda) \geq \tilde{h}'(\lambda)$. Since for any $x \geq 1$,

$$H_p''(x) = \frac{x^{1/p} + x^{1/p'}}{x^2} \geq x^{1/p-2} \geq \frac{1}{x} = (x\log x)'',$$

then also $H_p''(x) \geq \tilde{h}''(x)$ for any $x \geq \lambda$ and (iii) follows (note that the calculation above also covers the case $p = 1$, as then $1/p' = 0$).

Similarly, if $\tilde{h}(x)$ is the RHS of (iv), then $H_p(\lambda) \geq \tilde{h}(\lambda)$ and $H_p'(\lambda) \geq \tilde{h}'(\lambda)$ by the definition of $\tilde{h}$. As

$$H_p''(x) = \frac{x^{1/p} + x^{1/p'}}{x^2} \geq c^{-1/p} \cdot c^2\frac{(cx)^{1/p} + (cx)^{1/p'}}{(cx)^2} = \frac{d^2}{dx^2}\big(c^{-1/p}H_p(cx)\big),$$

it follows that $H_p''(x) \geq \tilde{h}''(x)$ for any $x \geq \lambda$, yielding (iv).

Finally, since $H_p(1) = 0$ and H_p is convex by (i), we have for any $\lambda \leq y \leq x$,

$$\frac{H_p(y)}{y} \leq \frac{H_p(y)}{y-1} \leq \frac{H_p(x)}{x-1} \leq \frac{\lambda}{\lambda-1}\frac{H_p(x)}{x}$$

yielding (v). □

For $1 < \rho \leq x$ and $k \geq 1$, denote

$$\alpha_{x,\rho}(k) = \inf\Big\{\sum_{s=0}^{k-1} H_p\big(\frac{g_{s+1}}{g_s}\big)\mu_s \ : \ g_0 = 1 \leq \ldots \leq g_{k-1} < \rho \leq x \leq g_k\Big\}. \tag{3.3.20}$$

This quantity plays a crucial role in providing sufficient condition for the p-log-Sobolev inequalities in Theorem 3.2.3. Its definition is partially inspired by an analogous quantity defined in [26] in the continuous setting.

Lemma 3.3.11. *For any* $1 < \rho \leq x$ *and* $k \geq 1$,

$$\alpha_{x,\rho}(k) \geq \Big[\sum_{s=0}^{k-1}\mu_s^{-1}\Big]^{-1} \cdot (\log x)^2 \tag{3.3.21}$$

and

$$\alpha_{x,\rho}(k) \geq H_p(x\rho^{-1})\mu_{k-1}. \tag{3.3.22}$$

Proof. We start with (3.3.21). By Lemma 3.3.10, (ii),

$$\begin{aligned}
\alpha_{x,\rho}(k) &\geq \inf\Big\{\sum_{s=0}^{k-1}\Big(\log\frac{g_{s+1}}{g_s}\Big)^2\mu_s \;:\; g_0=1\leq\ldots\leq g_{k-1}<\rho\leq x\leq g_k\Big\}\\
&\geq \inf\Big\{\sum_{s=0}^{k-1}\lambda_s^2\mu_s \;:\; \sum_{s=0}^{k-1}\lambda_s\geq\log x\Big\}\\
&\geq \inf\Big\{\Big[\sum_{s=0}^{k-1}\lambda_s\Big]^2\Big[\sum_{s=0}^{k-1}\mu_s^{-1}\Big]^{-1} \;:\; \sum_{s=0}^{k-1}\lambda_s\geq\log x\Big\}=\Big[\sum_{s=0}^{k-1}\mu_s^{-1}\Big]^{-1}\cdot(\log x)^2,
\end{aligned}$$

where in the last estimate we used the Cauchy-Schwarz inequality.

We turn to (3.3.22). By Lemma 3.3.10, H_p is non-negative and increasing on $[1,\infty)$, whence

$$\alpha_{x.\rho}(k)\geq\inf\Big\{\,H_p\big(\frac{g_k}{g_{k-1}}\big)\mu_{k-1} \;:\; 1\leq g_{k-1}<\rho\leq x\leq g_k\,\Big\}=H_p(x\rho^{-1})\mu_{k-1}$$

as desired. □

3.4 Proof of Theorem 3.2.3

3.4.1 Sufficient condition

Fix $p\in(0,1]$ and assume that μ satisfies the Poincaré inequality (3.2.6) with constant $C_P<\infty$ and that $\hat{C}<\infty$, i.e., the condition (3.2.7) holds. Recall the definition (3.3.2) of C_μ and that for fully supported measures, $C_\mu\leq C_P/2\mu_0<\infty$, cf. Theorem 3.3.2 and Remark 3.3.3. We show that μ satisfies the p-log-Sobolev inequality (3.2.5) with constant C bounded from above by a quantity depending on C_μ and $\hat{C}$ only.

Define ρ as

$$\rho=\min\Big(\big(\frac{1+C_\mu}{C_\mu}\big)^{1/3},2\Big). \tag{3.4.1}$$

By Corollary 3.3.9, it suffices to show that for any non-decreasing function $f\colon\mathbb{N}\to\mathbb{R}_+$ with a finite number of jumps,

$$\sum_{k\colon f(k)\geq\rho\,\mathbb{E}\,f} f(k)\log\big(\frac{f(k)}{\mathbb{E}_\mu f}\big)\mu_k\leq C_\rho\mathcal{E}_p(f)=C_\rho\sum_{k=0}^{\infty}f(k)H_p\big(\frac{f(k+1)}{f(k)}\big)\mu_k \tag{3.4.2}$$

for some constant $C_\rho>0$ independent of f. By the homogeneity of (3.4.2), we can assume that $f(0)=1$. Let us consider such f and denote

$$\tau_0=0,\qquad \tau_k=\inf\{\,l>\tau_{k-1}\colon f(l)\geq\rho f(\tau_{k-1})\,\}$$

for $k\geq1$. Since f has a finite number of jumps, then there exists $M\in\mathbb{N}\setminus\{0\}$ such that $\tau_{M-1}<\tau_M=\infty$. If $M=1$, then the LHS of (3.4.2) equals zero as $\mathbb{E}\,f\geq f(0)=1$ and (3.4.2) holds with any $C_\rho>0$. Assume therefore from now on that $M>1$. For $k\in\{1,\ldots,M-1\}$, denote

$$\gamma_k=\rho f(\tau_k)\log_+\big(\frac{\rho f(\tau_k)}{\mathbb{E}_\mu f}\big)\mu[\tau_k,\infty),$$

where $\log_+(x) := \max(\log x, 0)$ and let

$$\delta_k = f(\tau_{k-1}) \sum_{l=\tau_{k-1}}^{\tau_k - 1} H_p\big(\frac{f(l+1)}{f(l)}\big)\mu_l.$$

Since $\mathbb{E}\, f \geq f(0) = 1$ and as for $k \geq 1$ and $l \in [\tau_k, \tau_{k+1})$, $f(\tau_k) \leq f(l) < \rho f(\tau_k)$, we have

$$\begin{aligned}
\sum_{k:\ f(k) \geq \rho\, \mathbb{E}_\mu f} f(k) \log\big(\frac{f(k)}{\mathbb{E}_\mu f}\big)\mu_k &\leq \sum_{k:\ f(k) \geq \rho} f(k) \log_+\big(\frac{f(k)}{\mathbb{E}_\mu f}\big)\mu_k \\
&= \sum_{k=1}^{M-1} \sum_{l=\tau_k}^{\tau_{k+1}-1} f(l) \log_+\big(\frac{f(l)}{\mathbb{E}_\mu f}\big)\mu_l \\
&< \sum_{k=1}^{M-1} \rho f(\tau_k) \log_+\big(\frac{\rho f(\tau_k)}{\mathbb{E}_\mu f}\big)\mu[\tau_k, \tau_{k+1}) \leq \sum_{k=1}^{M-1} \gamma_k
\end{aligned}$$

and as $H_p \geq 0$ and by the monotonicity of f,

$$\sum_{k=0}^{\infty} f(k) H_p\big(\frac{f(k+1)}{f(k)}\big)\mu_k \geq \sum_{k=1}^{M-1} \delta_k.$$

Therefore, to prove (3.4.2), it suffices to show that

$$\sum_{k=1}^{M-1} \gamma_k \leq C_\rho \sum_{k=1}^{M-1} \delta_k. \tag{3.4.3}$$

Recall that $H_p(1) = 0$, $H_p \geq 0$ and recall the definition (3.3.20) of $\alpha_{x,\rho}$. Consider $g_l = 1$ for $l = 0, 1, \ldots, \tau_{k-1}$ and $g_l = f(l)/f(\tau_{k-1})$ for $l = \tau_{k-1} + 1, \ldots, \tau_k$. Then, by the definition of δ_k, we have for any $k \in \{1, \ldots, M-1\}$,

$$\alpha_{f(\tau_k)/f(\tau_{k-1}),\rho}(\tau_k) \leq \delta_k / f(\tau_{k-1}). \tag{3.4.4}$$

For $k = 1$, using the monotonicity of f, estimate (3.4.4) and the fact that $f(\tau_0) = 1 < \rho \leq f(\tau_1)$, we get

$$\begin{aligned}
\gamma_1 &\leq \rho f(\tau_1) \log_+\big(\frac{\rho f(\tau_1)}{\mu[0,\tau_1) + f(\tau_1)\mu[\tau_1,\infty)}\big)\mu[\tau_1,\infty) \cdot \frac{\delta_1}{\delta_1} \\
&\leq \rho f(\tau_1) \log_+\big(\frac{\rho f(\tau_1)}{\mu[0,\tau_1) + f(\tau_1)\mu[\tau_1,\infty)}\big)\mu[\tau_1,\infty)\big[\alpha_{f(\tau_1),\rho}(\tau_1)\big]^{-1} \cdot \delta_1 \\
&\leq \sup\Big\{\rho x \log_+\big(\frac{\rho x}{\mu[0,l) + x\mu[l,\infty)}\big)\mu[l,\infty)\big[\alpha_{x,\rho}(l)\big]^{-1}\ :\ x \geq \rho,\ l \geq 1\Big\} \cdot \delta_1 \\
&=: C_1 \delta_1.
\end{aligned} \tag{3.4.5}$$

For $2 \leq k \leq M-1$ (if such exist), we estimate each γ_k based on two cases. To that end, choose $\varepsilon = \rho^{-1}$ and note that by the definition (3.4.1) of ρ and by Lemma 3.3.4, for any $l \geq 1$,

$$\varepsilon < 1 < \rho < \rho^2 \leq \varepsilon \frac{\mu[l-1,\infty)}{\mu[l,\infty)}. \tag{3.4.6}$$

Consider the case $f(\tau_k)/f(\tau_{k-1}) > \varepsilon\mu[\tau_k-1,\infty)/\mu[\tau_k,\infty)$. Using monotonicity of f, Markov's inequality implies that $\frac{\rho f(\tau_k)}{\mathbb{E}_\mu f} \leq \frac{\rho}{\mu[\tau_k,\infty)}$. Using this estimate together with (3.4.4), we get that

$$\begin{aligned}
\gamma_k &\leq \rho f(\tau_k)\log\big(\frac{\rho}{\mu[\tau_k,\infty)}\big)\mu[\tau_k,\infty)\cdot\frac{\delta_k}{\delta_k}\\
&\leq \frac{\rho f(\tau_k)}{f(\tau_{k-1})}\log\big(\frac{\rho}{\mu[\tau_k,\infty)}\big)\mu[\tau_k,\infty)\big[\alpha_{f(\tau_k)/f(\tau_{k-1}),\rho}(\tau_k)\big]^{-1}\cdot\delta_k\\
&\leq \sup\Big\{\frac{\rho x\log\big(\frac{\rho}{\mu[l,\infty)}\big)\mu[l,\infty)}{\alpha_{x,\rho}(l)}\ :\ l\geq 1,\ x>\varepsilon\frac{\mu[l-1,\infty)}{\mu[l,\infty)}\Big\}\cdot\delta_k\\
&=: C_2\delta_k.
\end{aligned}\tag{3.4.7}$$

If $f(\tau_k)/f(\tau_{k-1}) \leq \varepsilon\mu[\tau_k-1,\infty)/\mu[\tau_k,\infty)$, then using the estimate $\log_+(xy)\leq \log_+(x)+\log_+(y)$, we split

$$\gamma_k \leq \underbrace{\rho f(\tau_k)\log_+\big(\frac{\rho f(\tau_{k-1})}{\mathbb{E} f}\big)\mu[\tau_k,\infty)}_{=:A} + \underbrace{\rho f(\tau_k)\log_+\big(\frac{f(\tau_k)}{f(\tau_{k-1})}\big)\mu[\tau_k,\infty)}_{=:B},$$

and we estimate A and B separately using the assumption as follows:

$$\begin{aligned}
A &\leq \varepsilon\rho f(\tau_{k-1})\log_+\big(\frac{\rho f(\tau_{k-1})}{\mathbb{E} f}\big)\mu[\tau_k-1,\infty)\\
&\leq \varepsilon\rho f(\tau_{k-1})\log_+\big(\frac{\rho f(\tau_{k-1})}{\mathbb{E} f}\big)\mu[\tau_{k-1},\infty) = \varepsilon\gamma_{k-1}
\end{aligned}\tag{3.4.8}$$

and, using monotonicity of f and (3.4.4),

$$\begin{aligned}
B &= \rho f(\tau_k)\log\big(\frac{f(\tau_k)}{f(\tau_{k-1})}\big)\mu[\tau_k,\infty)\cdot\frac{\delta_k}{\delta_k}\\
&\leq \frac{\rho f(\tau_k)}{f(\tau_{k-1})}\log\big(\frac{f(\tau_k)}{f(\tau_{k-1})}\big)\mu[\tau_k,\infty)\big[\alpha_{f(\tau_k)/f(\tau_{k-1}),\rho}(\tau_k)\big]^{-1}\cdot\delta_k\\
&\leq \sup\Big\{\frac{\rho x\log(x)}{\alpha_{x,\rho}(l)}\mu[l,\infty)\ :\ l\geq 1,\ \rho\leq x\leq\varepsilon\frac{\mu[l-1,\infty)}{\mu[l,\infty)}\Big\}\cdot\delta_k\\
&=: C_3\delta_k.
\end{aligned}\tag{3.4.9}$$

Combine estimates (3.4.7), (3.4.8) and (3.4.9) to get that for $k>1$,

$$\gamma_k \leq \varepsilon\gamma_{k-1} + (C_2+C_3)\delta_k.$$

By (3.4.5), $\gamma_1\leq C_1\delta_1$, whence

$$(1-\varepsilon)\sum_{k=1}^{M-1}\gamma_k \leq (C_1+C_2+C_3)\sum_{k=1}^{M-1}\delta_k. \tag{3.4.10}$$

It suffices therefore to estimate the terms C_1, C_2 and C_3.

Estimating C_3

Recall the definition (3.4.9) of C_3 and the relation (3.4.6) – we split the supremum based on whether $x\leq\rho^2$ or not. By Lemma 3.3.11, eq. (3.3.21), and by the definition (3.3.2) of C_μ,

$$\sup\Big\{\frac{\rho x\log(x)}{\alpha_{x,\rho}(l)}\mu[l,\infty)\ :\ l\geq 1,\ \rho\leq x\leq\rho^2\Big\}$$
$$\leq\sup\Big\{\frac{\rho x}{\log x}\big[\sum_{s=0}^{l-1}\mu_s^{-1}\big]\mu[l,\infty)\ :\ l\geq 1,\ \rho\leq x\leq\rho^2\Big\}\leq C_\mu\frac{\rho^3}{\log\rho}. \tag{3.4.11}$$

Similarly, by Lemma 3.3.11, eq. (3.3.22), and by the definition (3.3.2) of C_μ,

$$\sup\Big\{\frac{\rho x\log(x)}{\alpha_{x,\rho}(l)}\mu[l,\infty)\ :\ l\geq 1,\ \rho^2\leq x\leq\varepsilon\frac{\mu[l-1,\infty)}{\mu[l,\infty)}\Big\}$$
$$\leq\sup\Big\{\frac{\rho x\log(x)}{H_p(x\rho^{-1})}\frac{\mu[l,\infty)}{\mu_{l-1}}\ :\ l\geq 1,\ \rho^2\leq x\leq\varepsilon\frac{\mu[l-1,\infty)}{\mu[l,\infty)}\Big\}$$
$$\leq C_\mu\sup_{x\geq\rho^2}\frac{\rho x\log(x)}{H_p(x\rho^{-1})}. \tag{3.4.12}$$

Combining (3.4.11) and (3.4.12) and using (iii) of Lemma 3.3.10 with $\lambda=\rho$ and $x\rho^{-1}\geq\rho$ in place of x, we obtain

$$\begin{aligned}C_3&\leq C_\mu\Big(\frac{\rho^3}{\log\rho}+\max\Big\{1,\frac{\rho\log\rho}{H_p(\rho)},\frac{1+\log\rho}{H_p'(\rho)}\Big\}\cdot\sup_{x\geq\rho^2}\frac{\rho^2\log x}{\log(x\rho^{-1})}\Big)\\&=C_\mu\Big(\frac{\rho^3}{\log\rho}+2\max\Big\{1,\frac{\rho\log\rho}{H_p(\rho)},\frac{1+\log\rho}{H_p'(\rho)}\Big\}\cdot\rho^2\Big).\end{aligned} \tag{3.4.13}$$

Estimating C_2

Recall the definition (3.4.7) of C_2 and the relation (3.4.6). By Lemma 3.3.11, eq. (3.3.22) and using that $\rho\leq 2$ by definition, we get

$$\begin{aligned}C_2&\leq\sup\Big\{\frac{\rho x}{H_p(x\rho^{-1})}\cdot\log\big(\frac{\rho}{\mu[l,\infty)}\big)\cdot\frac{\mu[l,\infty)}{\mu_{l-1}}\ :\ l\geq 1,\ x>\varepsilon\frac{\mu[l-1,\infty)}{\mu[l,\infty)}\Big\}\\&\leq\sup\Big\{\frac{\rho x}{H_p(x\rho^{-1})}\cdot\log\big(\frac{2}{\mu[l,\infty)}\big)\cdot\frac{\mu[l,\infty)}{\mu_{l-1}}\ :\ l\geq 1,\ x>\varepsilon\frac{\mu[l-1,\infty)}{\mu[l,\infty)}\Big\}.\end{aligned} \tag{3.4.14}$$

For any $l\geq 1$ and $x>\varepsilon\frac{\mu[l-1,\infty)}{\mu[l,\infty)}$, applying (v) of Lemma 3.3.10 with $\rho^{-1}x\geq\rho$ in place of x, $\frac{\varepsilon}{\rho}\frac{\mu[l-1,\infty)}{\mu[l,\infty)}\geq\rho$ in place of y and $\lambda=\rho$ (recall (3.4.6)) we get that

$$\frac{\rho}{\varepsilon}\frac{\mu[l,\infty)}{\mu[l-1,\infty)}H_p\big(\frac{\varepsilon}{\rho}\frac{\mu[l-1,\infty)}{\mu[l,\infty)}\big)\leq\frac{\rho^2}{\rho-1}\frac{H_p(x\rho^{-1})}{x},$$

which after rearrangement (recall that $\varepsilon=\rho^{-1}$) is equivalent to

$$\frac{\rho x}{H_p(x\rho^{-1})}\cdot\mu[l,\infty)\leq\frac{\rho}{\rho-1}\Big[H_p\big(\frac{1}{\rho^2}\frac{\mu[l-1,\infty)}{\mu[l,\infty)}\big)\Big]^{-1}\mu[l-1,\infty),$$

which combined with Lemma 3.3.4 allows estimating further (3.4.14) as follows:

$$\begin{aligned}C_2&\leq\frac{\rho}{\rho-1}\sup_{l\geq 1}\Big\{\Big[H_p\big(\frac{1}{\rho^2}\frac{\mu[l-1,\infty)}{\mu[l,\infty)}\big)\Big]^{-1}\cdot\log\big(\frac{2}{\mu[l,\infty)}\big)\cdot\frac{\mu[l-1,\infty)}{\mu_{l-1}}\Big\}\\&\leq(1+C_\mu)\frac{\rho}{\rho-1}\sup_{l\geq 1}\Big\{\Big[H_p\big(\frac{1}{\rho^2}\frac{\mu[l-1,\infty)}{\mu[l,\infty)}\big)\Big]^{-1}\cdot\log\big(\frac{2}{\mu[l,\infty)}\big)\Big\}.\end{aligned} \tag{3.4.15}$$

For any $l \geq 1$, applying (iv) of Lemma 3.3.10 with $c = \rho^2$, $\lambda = \rho$ and $\frac{1}{\rho^2}\frac{\mu[l-1,\infty)}{\mu[l,\infty)} \geq \rho$ in place of x (recall (3.4.6)) gives

$$H_p\Big(\frac{1}{\rho^2}\frac{\mu[l-1,\infty)}{\mu[l,\infty)}\Big) \geq H_p\Big(\frac{\mu[l-1,\infty)}{\mu[l,\infty)}\Big)\min\Big\{\frac{H_p(\rho)}{H_p(\rho^3)}, \frac{H_p'(\rho)}{\rho^2 H_p'(\rho^3)}, \rho^{-2/p}\Big\},$$

which combined with (3.4.15) and assumption (3.2.7) results in

$$C_2 \leq (1+C_\mu)\frac{\rho}{\rho-1}\max\Big\{\frac{H_p(\rho^3)}{H_p(\rho)}, \frac{\rho^2 H_p'(\rho^3)}{H_p'(\rho)}, \rho^{2/p}\Big\}\cdot \hat{C} < \infty. \tag{3.4.16}$$

Estimating C_1.

Recall the definition (3.4.5) of C_1. We use the same ideas as above by considering two cases. For any $l \geq 1$, if $\rho \leq x \leq \varepsilon\mu[l-1,\infty)/\mu[l,\infty)$, then

$$\begin{aligned}\rho x \log\Big(\frac{\rho x}{\mu[0,l) + x\mu[l,\infty)}\Big)\mu[l,\infty) &\leq \rho x\log(\rho x)\mu[l,\infty) \\ &\leq 2\rho x\log(x)\mu[l,\infty) \leq 2C_3\cdot\alpha_{x,\rho}(l),\end{aligned} \tag{3.4.17}$$

where in the fist and second step we used that $x \geq \rho > 1$ and in the last step we used the definition (3.4.9) of C_3.

If $x > \varepsilon\mu[l-1,\infty)/\mu[l,\infty)$, then by the definition (3.4.7) of C_2,

$$\rho x\log\Big(\frac{\rho x}{\mu[0,l)+x\mu[l,\infty)}\Big)\mu[l,\infty) \leq \rho x\log\Big(\frac{\rho}{\mu[l,\infty)}\Big)\mu[l,\infty) \leq C_2\cdot\alpha_{x,\rho}(l) \tag{3.4.18}$$

and thus combining (3.4.17) and (3.4.18), we arrive at

$$C_1 \leq C_2 + 2C_3. \tag{3.4.19}$$

Final estimate

Combining (3.4.13) and (3.4.16) together with bounds (3.4.10) and (3.4.19) yields (3.4.3) with $C_\rho = \frac{2C_2+3C_3}{1-\varepsilon}$, which is bounded from above by

$$\begin{aligned}\frac{\rho}{\rho-1}\Big[&2(1+C_\mu)\hat{C}\frac{\rho}{\rho-1}\max\Big\{\frac{H_p(\rho^3)}{H_p(\rho)}, \frac{\rho^2H_p'(\rho^3)}{H_p'(\rho)}, \rho^{2/p}\Big\} \\ &+3C_\mu\Big(\frac{\rho^3}{\log\rho} + 2\rho^2\max\Big\{1, \frac{\rho\log\rho}{H_p(\rho)}, \frac{1+\log\rho}{H_p'(\rho)}\Big\}\Big)\Big] < \infty.\end{aligned}$$

Therefore, we obtain (3.4.2) for any non-decreasing $f\colon \mathbb{N}\to\mathbb{R}_+$ with a finite number of jumps and C_ρ as above. Thus, Corollary 3.3.9 implies that

$$\mathrm{Ent}_\mu(f) \leq [64C_\mu(1+\sqrt{\rho})^2 + 4C_\rho]\mathcal{E}_p(f)$$

for any $f\colon\mathbb{N}\to\mathbb{R}_+$, as desired.

3.4.2 Necessary condition

For the sake of contradiction, assume that there exists a sequence $0 = \tau_0 < \tau_1 < \ldots$ such that (3.2.8) holds, i.e.,

$$\beta_k := \Big[H_p\Big(\frac{\mu[\tau_{k-1},\infty)}{\mu[\tau_k,\infty)}\Big)\Big]^{-1}\cdot\frac{\mu[\tau_{k-1},\infty)}{\mu[\tau_k-1,\infty)}\cdot\log\Big(\frac{2}{\mu[\tau_k,\infty)}\Big) \to \infty$$

as $k \to \infty$, and that μ verifies p-LS(C) with some finite constant $C > 0$. By Theorem 3.1.6, μ satisfies the Poincaré inequality (3.1.6) with the same constant C and therefore $C_\mu < \infty$ (recall the definition (3.3.2) of C_μ and Remark 3.3.3). For $M \geq 1$, set

$$f_M = \Big[\sum_{k=0}^{M-1} \frac{\mathbf{1}_{[\tau_k,\tau_{k+1})}}{\mu[\tau_k,\infty)}\Big] + \frac{\mathbf{1}_{[\tau_M,\infty)}}{\mu[\tau_M,\infty)}.$$

Then $\mathbb{E}_\mu f_M \leq M+1$. Moreover, by Lemma 3.3.4, $\frac{\mu[\tau_k,\tau_{k+1})}{\mu[\tau_k,\infty)} \geq \frac{1}{1+C_\mu}$ for any $k \in \mathbb{N}$, whence

$$\mathrm{Ent}_\mu(f_M) \geq -(M+1)\log(M+1) + \frac{1}{1+C_\mu}\sum_{k=1}^{M} \log\big(\frac{1}{\mu[\tau_k,\infty)}\big).$$

Similarly (recall that $H_p(1) = 0$),

$$\begin{aligned}
\mathcal{E}_p(f_M) &= \sum_{k=1}^{M} \mu_{\tau_k-1} f(\tau_k - 1) H_p\big(\frac{f(\tau_k)}{f(\tau_k-1)}\big) \\
&\leq \sum_{k=1}^{M} \frac{\mu[\tau_k - 1,\infty)}{\mu[\tau_{k-1},\infty)} H_p\big(\frac{f(\tau_k)}{f(\tau_k-1)}\big) = \sum_{k=1}^{M} \beta_k^{-1}\log\big(\frac{2}{\mu[\tau_k,\infty)}\big)
\end{aligned}$$

and consequently, since $\mathrm{Ent}_\mu(f_M) \leq \frac{C}{4}\mathcal{E}_p(f_M)$ by assumption,

$$\begin{aligned}
&\sum_{k=1}^{M}\Big[\log\big(\frac{1}{\mu[\tau_k,\infty)}\big)\cdot\big(1 - \frac{C(1+C_\mu)}{4\beta_k}\big)\Big] \\
&\qquad \leq (1+C_\mu)\Big[\frac{C}{4}M(\log 2)\sup_k \beta_k^{-1} + (M+1)\log(M+1)\Big]
\end{aligned} \tag{3.4.20}$$

(recall that $\beta_k \to \infty$, whence $\sup_k \beta_k^{-1} < \infty$). Let k_0 be such that $2\beta_k \geq C(1+C_\mu)$ for every $k \geq k_0$. By Proposition 3.3.6, $\log\big(\frac{1}{\mu[\tau_k,\infty)}\big) \geq cp\tau_k \log \tau_k$ for some constant $c > 0$, whence by (3.4.20)

$$\begin{aligned}
\sum_{k=k_0}^{M} k\log k &\leq \sum_{k=k_0}^{M} \tau_k \log \tau_k \\
&\leq \frac{1}{cp}\sum_{k=k_0}^{M}\log\big(\frac{1}{\mu[\tau_k,\infty)}\big) \leq \tilde{c}\cdot(M+1)\log(M+1)
\end{aligned}$$

for M big enough and some constant $\tilde{c} > 0$ independent of M. We arrive at the desired contradiction by taking a limit as $M \to \infty$ and by noting that $\int_{k_0}^{M} x\log x\,dx \geq \hat{c}M^2\log M$ for M big enough and some constant $\hat{c} > 0$ independent of M.

3.5 Proof of Theorem 3.2.2

As a counterexample we will take the Conway–Maxwell–Poisson distribution with parameter $\nu > 0$, defined as $\mu_\nu(k) = \frac{1}{Z_\nu}(k!)^{-\nu}$ for $k \in \mathbb{N}$, where $Z_\nu = \sum_{k\geq 0}(k!)^{-\nu}$ is the normalizing constant, cf. [73].

For any $n \in \mathbb{N}$, using the estimate

$$b! \geq a!(b-a)! \tag{3.5.1}$$

valid for $b \geq a \geq 0$ we obtain

$$\begin{aligned}\mu_\nu(\{n\}) \leq \mu_\nu[n,\infty) &= \frac{1}{Z_\nu}\sum_{k\geq n}\frac{1}{(k!)^\nu} \\ &\leq \frac{1}{Z_\nu(n!)^\nu}\sum_{k\geq 0}\frac{1}{(k!)^\nu} = \frac{1}{(n!)^\nu} = Z_\nu\mu_\nu(\{n\}).\end{aligned} \tag{3.5.2}$$

Recall the definition (3.3.2) of C_μ. Using (3.5.2) and (3.5.1), we obtain for $n \in \mathbb{N}\setminus\{0\}$,

$$\mu_\nu[n,\infty)\sum_{k=0}^{n-1}\frac{1}{\mu_\nu(\{k\})} \leq Z_\nu\mu_\nu(\{n\})\sum_{k=0}^{n-1}\frac{1}{\mu_\nu(\{k\})} = Z_\nu\frac{\sum_{k=0}^{n-1}(k!)^\nu}{(n!)^\nu} \leq Z_\nu^2,$$

whence $C_{\mu_\nu} < \infty$ and thus μ_ν satisfies the Poincaré inequality (3.2.6) with constant $C_P = 8Z_\nu^2$, cf. Theorem 3.3.2 and Remark 3.3.3.

We first show that for any $\nu \in (0,1]$, μ_ν verifies the p-log-Sobolev inequality (3.2.5) for any $p < \nu$. Fix some $0 < p < \nu \leq 1$. By the first part of Theorem 3.2.3, it suffices to show that

$$\sup_{n\geq 1}\Big\{\Big[H_p\big(\frac{\mu_\nu[n-1,\infty)}{\mu_\nu[n,\infty)}\big)\Big]^{-1}\cdot\log\big(\frac{2}{\mu_\nu[n,\infty)}\big)\Big\} < \infty. \tag{3.5.3}$$

By (3.5.2), for any $n \geq 1$

$$\begin{aligned}\frac{n^\nu}{Z_\nu} = \frac{\mu_\nu(\{n-1\})}{\mu_\nu(\{n\})}\frac{1}{Z_\nu} &\leq \frac{\mu_\nu(\{n-1\})}{\mu_\nu[n,\infty)} \\ &\leq \frac{\mu_\nu[n-1,\infty)}{\mu_\nu[n,\infty)} \\ &\leq Z_\nu\frac{\mu_\nu(\{n-1\})}{\mu_\nu[n,\infty)} \leq Z_\nu\frac{\mu_\nu(\{n-1\})}{\mu_\nu(\{n\})} = Z_\nu n^\nu.\end{aligned} \tag{3.5.4}$$

This, together with Lemma 3.3.4 and the definition (3.2.3) of H_p implies that

$$\Big[H_p\big(\frac{\mu_\nu[n-1,\infty)}{\mu_\nu[n,\infty)}\big)\Big]^{-1} \leq \Big[H_p\big(\max\{\frac{1+C_{\mu_\nu}}{C_{\mu_\nu}},\frac{n^\nu}{Z_\nu}\}\big)\Big]^{-1} \leq \frac{C'}{n^{\nu/p}} \tag{3.5.5}$$

for any $n \geq 1$ and some big enough constant $C' > 0$ (independent on n but dependent on ν and p). By (3.5.2) and Stirling's formula

$$\lim_{n\to\infty}\frac{\log\big(\frac{2}{\mu_\nu[n,\infty)}\big)}{n\log n} = \nu. \tag{3.5.6}$$

Combining (3.5.5) with (3.5.6) and recalling that $p < \nu$, we obtain (3.5.3).

Finally, we show that for any $\nu \in (0,1)$, μ_ν does not satisfy the ν-log-Sobolev inequality. By Theorem 3.2.3, it suffices to show that there exists an increasing sequence $\tau_0 < \tau_1 < \ldots$ such that

$$\lim_{n\to\infty}\Big\{\Big[H_\nu\big(\frac{\mu[\tau_{n-1},\infty)}{\mu[\tau_n,\infty)}\big)\Big]^{-1}\cdot\frac{\mu[\tau_{n-1},\infty)}{\mu[\tau_n-1,\infty)}\cdot\log\big(\frac{2}{\mu[\tau_n,\infty)}\big)\Big\} = \infty. \tag{3.5.7}$$

We choose $\tau_n = n$ and (3.5.7) becomes

$$\lim_{n\to\infty}\Big\{\Big[H_\nu\big(\frac{\mu[n-1,\infty)}{\mu[n,\infty)}\big)\Big]^{-1}\cdot\log\big(\frac{2}{\mu[n,\infty)}\big)\Big\} = \infty. \tag{3.5.8}$$

Analogously as in (3.5.5), using (3.5.4) and the definition (3.2.3) of H_p, we get that

$$\Big[H_\nu\big(\frac{\mu_\nu[n-1,\infty)}{\mu_\nu[n,\infty)}\big)\Big]^{-1} \geq \Big[H_\nu\big(Z_\nu n^\nu\big)\Big]^{-1} \geq \frac{C''}{n} \tag{3.5.9}$$

for any $n \geq 1$ and some small enough constant $C'' > 0$ (independent on n but dependent on ν). We conclude (3.5.8) by combining (3.5.6) with (3.5.9).

Chapter 4

Stochastic Covering Property

4.1 Introduction

Investigating families of binary random variables with negatively dependent coordinates in the recent years has attracted considerable attention, see, e.g., [169, 184, 45, 170, 131, 92, 32, 13, 16, 126]. A wide and important class of such variables is constituted by those satisfying the *strong Rayleigh property* (abbrev. SRP) introduced by Borcea et al. [45]. More precisely, a probability measure π on the hypercube $\mathcal{B}_n := \{0,1\}^n$ satisfies the SRP if its generating polynomial

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

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, laws of point processes or measures obtained by running exclusion dynamics on the cube, cf. Pemantle and Peres [170].

The main purpose of this chapter 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 4.2.1) introduced by Pemantle and Peres [170], since this condition already turns out to provide a useful framework for proving concentration results [170, 114, 16, 131, 126, 6]. 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., by Lee–Yau, Bobkov–Tetali, Gao–Quastel [140, 42, 91] and more recently by Samson [181] and Sambale–Sinulis [179]. 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 [69, 68, 70, 194, 32] for properties and applications of this family of distributions.

4.1.1 State of the art

The landmark paper that initiated the study of concentration phenomenon implied by the SCP is due to Pemantle and Peres [170] 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 [114], building on the works [143, 124], retrieved this estimate by proving that the SCP implies the modified log-Sobolev inequality, cf. Section 2.4.5. 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., [13, 128, 74, 12] – 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 $\mathcal{B}_n$ and $f\colon \mathcal{B}_n \to \mathbb{R}$, we use the notation $\pi(f) := \int f\, d\pi$).

Theorem 4.1.1 ([170, 114])**.** *For a probability measure π on $\mathcal{B}_n$ satisfying the SCP and any $f\colon \mathcal{B}_n \to \mathbb{R}$ such that*

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

the following estimate holds for all $t > 0$

$$\pi\big(f > \pi(f) + t\big) \le \exp(-t^2/8n). \tag{4.1.1}$$

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 4.1.1 for matrix-valued functions has been shown by Aoun et al. [16], who 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 by Kyng–Song [131] and Kaufman, Kyng and Soldá [127] for functions of the form $f(x) = \sum_{i=1}^{n} x_i C_i$, where C_i are nonnegative definite matrices (see Theorem 4.2.8 and Remark 4.2.10 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–Tetali [42] and independently Gao–Quastel [91] 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 [192]. Samson [181] complemented this approach by proving corresponding transportation inequalities. Very recently Sambale and Sinulis [179], 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 (4.1.1) 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 [32] by Bertail and Clémençon in which the authors, motivated by applications to survey sampling, obtain precise Bernstein-type inequalities for linear functionals.

4.1.2 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 4.1.1. 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 (4.1.1) 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 [153] or the monograph [47]),
- 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 [191, 192], or bounds for polynomials which are especially important for the discrete cube due to their relation with the Fourier–Walsh expansion (see, e.g., [164, 130, 135, 8]),
- inequalities for general matrix-valued functions, see [197, 168] for a description of this rich theory in the independent setting (we note important results obtained for sums of linear combination of positive definite matrices [131, 127] as well as a subexponential bound obtained in [16] 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 and Hoeffding [17], Talagrand's powerful induction techniques [192] and functional inequalities brought forward by Ledoux [137] and developed by many authors (see [51] 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, cf. Section 2.4.3. 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 (4.1.1). This in our opinion is the main obstacle in obtaining counterparts of classical strong concentration inequalities in the negatively dependent setting. Our goal in this chapter 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 chapter.

The first series of results we obtain concerns general measures satisfying the SCP for which we refine the Azuma type martingale argument used by Pemantle and Peres [170] and generalize Theorem 4.1.1 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 4.2.3. Next, we use the approach developed for the scalar case together with matrix bounded-difference inequality due to Tropp [197] to get an analogous concentration for matrix-valued functions (Theorem 4.2.5), strengthening the results of Aoun et al. [16], in particular obtaining a subgaussian inequality in place of a subexponential one. We note that the proof in [16] 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 [131] from linear combinations with coefficients in nonnegative definite matrices to general functions satisfying a matrix bounded-difference type assumptions (Theorem 4.2.8).

The second line of research presented in thesis chapter concerns the functional approach to improved concentration inequalities. We develop an abstract condition (Definition 4.4.3) 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–Castel [91] and Bobkov–Tetali [42], as well as the convex distance inequality and polynomial concentration obtained recently by Sambale–Sinulis [179]. 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 [68] for a description of statistical applications of conditional Bernoulli distributions and to the monograph [194] for an algorithmic perspective. Let us 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., [31, 112]). 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.

4.1.3 Organization of this chapter

In Section 4.2 we present our results concerning concentration for general measures satisfying the SCP/SRP. In Section 4.3 we specialize our analysis to Bernoulli random variables conditioned on their sum being equal to some constant. Then, in Section 4.4 we formulate an abstract framework that allows to deduce the results of Section 4.3. Finally, all the proofs are presented in Sections 4.5, 4.6 and 4.7.

4.2 Concentration under the SCP and SRP

In this section we present our concentration results for general measures satisfying the SCP or SRP. Let us start with introducing some notation. For $x = (x_1, \ldots, x_n) \in \mathcal{B}_n := \{0,1\}^n$ 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 $x \in \mathcal{B}_n$ 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 $e_i \in \mathcal{B}_n$ 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 4.2.1 (Stochastic covering property)**.** For $x, y \in \mathcal{B}_n$, we say that x covers y, denoted $x \triangleright y$, if $x = y$ or $x = y + e_i$ for some $i \in [n]$.

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

Remark 4.2.2. As indicated in the introduction, the SCP is implied by the SRP, cf. [170]. The opposite however is not true, as is demonstrated, e.g., by Cryan et al. in [74, Appendix A], where the authors show that yet another possible generalization of the SRP, the strong log-concavity, is incomparable with the SCP. In particular, they construct a distribution that is supported on the bases of a matroid, and that satisfies the SCP and violates the log-concavity (and whence the SRP as well).

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

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 $x, y \in \mathcal{B}_n$ 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 chapter is the following generalization of Theorem 4.1.1.

Theorem 4.2.3. *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 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 4.2.4. Theorem 4.2.3 implies Theorem 4.1.1 (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 $|c|^2 \ll n|c|_\infty^2$ in the non-homogeneous or $\sum_{i=1}^k (c_i^{\downarrow})^2 \ll k|c|_\infty^2$ in the k-homogeneous case, one can see that Theorem 4.2.3 can give substantially better concentration estimates than Theorem 4.1.1. We remark that such general linear functional 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., [32]).

We now formulate the matrix analogue of Theorem 4.2.3. 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 4.2.5. *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*

$$\|f(x) - f(y)\| \le d_\alpha(x,y) \quad \forall\, x, y \in \mathcal{B}_n \tag{4.2.1}$$

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 4.2.6. Recently, Aoun et al. [16] 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

$$\pi\big(\lambda_{max}(f - \pi(f)) > t\big) \le d\exp\Big(-\frac{t^2}{8k + 2t\sqrt{2k}}\Big). \tag{4.2.2}$$

The exponent in (4.2.2) is proportional to $-t/2\sqrt{2k}$ for t big enough and whence for such t Theorem 4.2.5 applied with $\alpha = (1, \ldots, 1)$ improves on (4.2.2) (and on an analogous result from [126]) as it yields a sub-Gaussian estimate.

Remark 4.2.7. Using semigroup techniques together with matrix concentration results implied by the Poincaré inequality due to Aoun et al. [16], 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 4.2.5, cf. Remark 4.3.7.

When comparing the inequality of Theorem 4.2.5 or the results from [16] 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 (4.2.1), 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 by Kyng and Song in [131]. It turns out that their approach can be adapted to the setting of general functions, yielding the following theorem.

Theorem 4.2.8. *Let π be a k-homogeneous probability measure $\mathcal{B}_n$ satisfying the strong Rayleigh property and $f\colon \mathcal{B}_n \to \mathcal{H}_d$ be such that there exists a sequence $C_1, \ldots, C_n \in \mathcal{H}_d$ satisfying*

$$(f(x)-f(x^i))^2 \preccurlyeq C_i^2 \quad \forall\, x \in \mathcal{B}_n,\ i \in [n]. \tag{4.2.3}$$

Then for any $t > 0$,

$$\pi\big(\lambda_{max}(f - \pi(f)) > t\big) \le d \exp\Big(-\frac{t^2}{8\|\pi(\tilde{f})\| \log(ek) + \frac{4}{3}Kt}\Big), \tag{4.2.4}$$

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

Remark 4.2.9. In fact, the only place in the proof of Theorem 4.2.8 where we use the SRP in its full strength is to get that $\mathbb{P}(X_i = 1 \,|\, X_{i_1} = 1, \ldots, X_{i_l} = 1) \le \mathbb{P}(X_i = 1)$ for $X \sim \pi$ and any $i, k \in [n]$ and $i_1, \ldots, i_k \subset [n] \setminus \{i\}$. Therefore, in Theorem 4.2.8 it suffices to assume that π satisfies the SCP and negative association, which is implied by the SRP, cf. [170].

Remark 4.2.10. It is natural to expect that $\log(ek)$ in (4.2.4) is just an artefact of the proof. Very recently in Kaufman, Kyng and Soldá in [127] 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. [131].

Let us also point out that despite the logarithmic factor present in Theorem 4.2.8, when we specialize it to the linear function f as discussed above, it is not directly comparable with the result from [127], 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).

4.3 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 $i \in [n]$. Finally, 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 of the celebrated convex distance inequality, introduced for the first time by Talagrand [191] for product measures on the cube.

Theorem 4.3.1. *If $\pi = \pi(p,k)$ for some $p \in (0,1)^n$ and $k \in \{0, \dots, n\}$, then for any $A \subset \mathcal{B}_n$,*

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

where

$$d_T(x, A) = \sup_{\alpha\colon |\alpha| \le 1} d_\alpha(x, A) \quad \textit{for} \quad x \in \mathcal{B}_n,\ A \subset \mathcal{B}_n.$$

Let $\mathbb{M}_\pi f$ denote any median of f with respect to the measure π. A classical consequence of Theorem 4.3.1 is the following fact regarding the concentration around the median of convex functions [51]. 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 4.3.2. *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 $t > 0$,*

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

Remark 4.3.3. 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 [7].

Remark 4.3.4. If $f\colon \mathcal{B}_n \to \mathbb{R}$ 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 4.3.2 counterparts Theorem 4.2.3 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.

Our next result concerns concentration for matrix-valued functions under weaker assumptions than those in Theorem 4.2.5.

Theorem 4.3.5. *Let $\pi = \pi(p,k)$ for some $p \in (0,1)^n$ and $k \in \{0, \dots, n\}$. Assume that $f\colon \mathcal{B}_n \to \mathcal{H}_d$ is such that there is a sequence of positive semidefinite matrices $C_1, \dots, C_n$ satisfying*

$$(f(x) - f(x^i))^2 \preccurlyeq C_i^2, \quad \forall\, x \in \mathcal{B}_n,\ i \in [n], \tag{4.3.1}$$

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

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

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 4.3.6. Condition (4.3.1) 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 4.2.5, Theorem 4.3.5 may improve significantly on Theorem 4.2.5 through better parameters in the exponent.

Remark 4.3.7. By an adaptation of the proof of Theorem 4.3.5, one can obtain a similar result for general k-homogeneous measures satisfying the SCP condition with the variance proxy parameter

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

Finally, let us turn to the higher order concentration. Below, for Reader's convenience, we repeat some parts and notation of Section 2.5. Recall that by the Fourier–Walsh expansion (see e.g., [164]), every function $f \colon \mathcal{B}_n \to \mathbb{R}$ 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 $\mathcal{B}_n$ 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.

Let $|I|$ be the cardinality of a set I and for a multiindex $\mathbf{i} = (i_1, \ldots, i_d) \in [n]^d$ 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 $A = (a_{\mathbf{i}})_{\mathbf{i} \in [n]^d}$, recall the notation

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

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

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

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

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

where $\|\cdot\|_{HS}$ and $\|\cdot\|$ denote the Hilbert–Schmidt and the $\ell^2 \to \ell^2$ operator norm respectively.

Theorem 4.3.8. *If $\pi = \pi(p,k)$ for some $p \in (0,1)^n$ and $k \in \{0,\ldots,n\}$, then for any tetrahedral polynomial $f\colon \mathcal{B}_n \to \mathbb{R}$ of degree d,*

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

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 [135] 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 [8, 4, 178, 179, 6]). We remark that in the independent, subgaussian case and $d = 2$ they reduce to the well known Hanson–Wright inequality for quadratic forms, which has proved useful in non-asymptotic analysis of random matrices and in asymptotic geometric analysis (see, e.g., [198, Chapter 6]). It is worth mentioning that in the Gaussian case they may be reversed up to the value of the absolute constants, thus Theorem 4.3.8 shows that the measures $\pi(p,k)$ exhibit Gaussian type concentration for polynomials. As we have seen in Section 2.5.3, 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 [8, 178, 179]) or to statistical applications, e.g., in testing Ising models [76] and signal processing [199]. It is worth noting that Theorem 4.3.8 is not a direct consequence of Corollary 2.5.4 since it is based on a dynamics that allows to change two coordinates simultaneously.

4.4 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 4.3 and of Theorem 4.2.3. 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 4.4.3) and their adaptation to the setting of flip-swap random walks.

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 [142, 24, 42] for details.

4.4.1 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

$$\forall\, x,y \in M \quad \pi(x)L(x,y) = \pi(y)L(y,x), \tag{4.4.1}$$

which implies that π is a stationary measure for the Markov process and L is self-adjoint on $L^2(\pi)$. In this chapter 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

$$\Gamma(f,g)(x) = \frac{1}{2}\sum_{y\in M}(f(x)-f(y))(g(x)-g(y))L(x,y) \tag{4.4.2}$$

is the corresponding carré-du-champ operator. We use shorthand notation $\Gamma(f,f) =: \Gamma(f)$ and observe that by the detailed-balance condition (4.4.1) we have $\pi\big(\Gamma(f)\big) = \pi(\Gamma_+(f))$, where

$$\Gamma_+(f)(x) = \sum_{y\in M}(f(x)-f(y))_+^2 L(x,y). \tag{4.4.3}$$

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

$$\rho(L)\operatorname{Ent}_\pi(f) \leq \mathcal{E}(f,\log f) \tag{4.4.4}$$

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 [42] and [141, 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 [138] by Ledoux and [51] by Boucheron et al.), says that for any $f\colon M \to \mathbb{R}$,

$$\pi\big(f > \pi(f)+t\big) \leq \exp(-t^2\rho(L)/4\|\Gamma_+(f)\|_\infty), \tag{4.4.5}$$

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

4.4.2 Flip-swap random walks

Let us now recall the results due to Hermon and Salez, formulated already in Section 2.4.5 and rephrase them in the notation more convenient for the upcoming arguments. After Hermon and Salez [114], 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 [114] can be stated in the following way (cf. Theorem 2.4.12 for the extended formulation of their results).

Theorem 4.4.1 ([114]). *For any measure π on $\mathcal{B}_n$ satisfying the SCP, there exists a kernel L generating a reversible flip-swap random walk with stationary measure π such that $\rho(L) \geq 1$ and $\Delta(L) \leq n$. If π is also k-homogeneous, then $\Delta(L) \leq 2k$ as well.*

Theorem 4.4.1, by means of Herbst's argument (4.4.5), implies (up to an absolute constant in the exponent) the estimate from Theorem 4.1.1 after observing that for a flip-swap random walk and any $f\colon \mathcal{B}_n \to \mathbb{R}$ that is 1-Lipschitz with respect to the Hamming distance d_H,

$$\|\Gamma_+(f)\|_\infty \leq \Delta(L)\cdot \max_{x,y\in\mathcal{B}_n}\{\,(f(y)-f(x))_+^2\colon L(x,y)>0\,\} \leq 4\Delta(L). \tag{4.4.6}$$

Remark 4.4.2. 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., [42, 98, 178, 74]. 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 4.4.3) constant.

It turns out that for the proofs of all the statements of Section 4.3 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 (4.4.4) is known.

Definition 4.4.3 (Stability condition)**.** Let L be a generator of a flip-swap random walk on $\mathcal{B}_n$ 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 (4.4.4) and

$$\max_{x\in \operatorname{supp}\pi;\, i\in[n]} \sum_{y:\, y_i\neq x_i} L(x,y) \leq R\rho(L). \tag{4.4.7}$$

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 4.4.4. If π is not concentrated on a single point, then a random walk on $\mathcal{B}_n$ with a generator L that satisfies the modified log-Sobolev inequality (4.4.4) 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., [6, Proposition B.5]. Let $f(x) = \mathbf{1}_{\{x_i=1\}}$. By the stability condition (4.4.7) and reversibility of L we get that

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

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 $\mathcal{B}_n$ 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 [42, Example 3.7], note a different normalization of both the Dirichlet form and the constant in the modified log-Sobolev inequality), whereas for all $x \in \mathcal{B}_n$

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

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

Example 4.4.5 (Bernoulli–Laplace model). Let π be the uniform measure on the slice of $\mathcal{B}_n$ 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 [91] and independently in [42] 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 4.7.3 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 4.4.6. 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 $\mathcal{B}_n$ 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 $\mathcal{B}_n = \mathcal{B}_{n_1} \times \cdots \times \mathcal{B}_{n_m}$, 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 $P_j \colon \mathcal{B}_n \to \mathcal{B}_j$ is the projection onto the j-th factor in the product $\mathcal{B}_n = \mathcal{B}_{n_1} \times \cdots \times \mathcal{B}_{n_m}$ Thanks to the well known tensorization property of the entropy (see, e.g., [14, Chapter 3]) we have $\rho(L) = \rho$. Moreover, for $i \in (n_1 + \ldots + n_{j-1}, n_1 + \ldots + n_j]$,

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

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 od Section 4.3 to product of measures $\pi(n,k)$ allowing for more general conditioning of Bernoulli variables.

4.4.3 Concentration results

Finally, let us present the counterparts of the results of Section 4.2 and of Theorem 4.2.3 from Section 4.3 in the abstract language of the stability condition (4.4.7). We stress here that it is the sole property needed for their proofs, which are deferred to Section 4.7.

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

Proposition 4.4.7. *If a flip-swap random walk on $\mathcal{B}_n$ with stationary distribution π and generator L satisfies the stability condition* (4.4.7)*, then for any $f\colon \mathcal{B}_n \to \mathbb{R}$ and $\alpha \in \mathbb{R}^n_+$ such that*

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

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 4.4.8. 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 $\Delta(L)/R\rho(L) \leq k$.

Let us now pass to the matrix-valued case.

Proposition 4.4.9. *Let a flip-swap random walk on* $\mathcal{B}_n$ *with stationary distribution* π *and generator* L *satisfy the stability condition* (4.4.7). *Assume also that* $f\colon \mathcal{B}_n \to \mathcal{H}_d$ *is such that there is a sequence of positive semidefinite matrices* $C_1, \dots, C_n$ *satisfying*

$$(f(x) - f(x^i))^2 \preccurlyeq C_i^2 \quad \forall\, x \in \mathcal{B}_n,\ i \in [n], \tag{4.4.8}$$

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 4.4.10. *If a flip-swap random walk on* $\mathcal{B}_n$ *with some stationary distribution* π *and a generator* L *satisfies the stability condition* (4.4.7), *then for any set* $A \subset \mathcal{B}_n$

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

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

Proposition 4.4.11. *If a flip-swap random walk on* $\mathcal{B}_n$ *with some stationary distribution* π *and a generator* L *satisfies the stability condition* (4.4.7), *then for any tetrahedral polynomial* $f\colon \mathcal{B}_n \to \mathbb{R}$ *of degree* d

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

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

Remark 4.4.12. Although Proposition 4.4.7 gives a worse constant in the exponent than Theorem 4.2.3 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 4.4.13. 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 4.4.14. Recently, Cryan et al. [74] have shown some version of Theorem 4.4.1 in the case of 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 [74]. It is known, cf. Brändén and Huh [53], 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 by Anari et al. [13] and Kaufman and Oppenheim [128], Cryan et al. [74] explicitly construct a base-exchange random walk, which has any given strongly log-concave measure as a stationary distribution, and verify it satisfies the modified log-Sobolev inequality (4.4.4), cf. Section 2.4.5, where we recall their construction in detail.

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 (4.4.7), 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 $\pi(p, k) = \mathcal{L}(B \,|\, \sum_i B_i = k)$. 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 [74] is at best k-stable. Therefore, applying propositions of Section 4.4.3 to the base-exchange random walk gives much worse concentration constants than those of Section 4.3. On the other hand, the abstract construction of a flip-swap random walk proposed by Hermon and Salez [114], 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 4.3.

In view of the above, it is an interesting problem to analyze what other known kernels satisfy the stability condition (4.4.7) with good (dimension-independent) constant and to look for some other criteria that would allow to deduce this condition.

4.5 Proofs of the results of Section 4.2

In this section we provide proofs of Theorems 4.2.3, 4.2.5 and 4.2.8. Our approach relies on certain refinements of the Azuma type martingale argument originally used by Pemantle and Peres [170]. For Theorems 4.2.3, 4.2.5 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 $\mathcal{B}_n$ 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

$$\begin{aligned} \mathcal{L}(Y_1 \,|\, X) &= \operatorname{Unif}(\operatorname{supp} X \backslash \{1, \ldots, k\}) \quad \text{and} \\ \mathcal{L}(Y_l \,|\, X, Y_1, \ldots, Y_{l-1}) &= \operatorname{Unif}(\operatorname{supp} X \setminus \{1, \ldots, k, Y_1, \ldots, Y_{l-1}\}), \end{aligned} \tag{4.5.1}$$

for $l = 2, \ldots, k$, where $\operatorname{Unif}(A)$ stands for the uniform distribution on the set A, and for notational simplicity we set $\operatorname{Unif}(\emptyset)$ 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 4.2.3 and 4.2.5 will be based on the following two lemmas.

Lemma 4.5.1. *Let $\alpha \in \mathbb{R}_+^n$ be non-increasing and let $f\colon \mathcal{B}_n \to \mathcal{H}_d$ be 1-Lipschitz with respect to the distance d_α. Assume that X is a $\mathcal{B}_n$-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 $l \in [n]$,*

$$M_l^2 \preccurlyeq 4\alpha_l^2 I_d. \tag{4.5.2}$$

Lemma 4.5.2. *In the setting of Lemma 4.5.1, 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 $l \in [k]$,*

$$N_l^2 \preccurlyeq 4\alpha_l^2 I_d, \tag{4.5.3}$$

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

$$N_l^2 \preccurlyeq 4\alpha_k^2 I_d. \tag{4.5.4}$$

We postpone for now the proof of the above lemmas and firstly show how they imply Theorems 4.2.3 and 4.2.5. To this end let us recall the matrix version of the Azuma-Hoeffding inequality due to Tropp [197, Theorem 7.1], which asserts that if D_l, $l = 1, \ldots, n$ are $\mathcal{H}_d$-valued martingale differences and $D_l^2 \preccurlyeq C_l^2$ for some deterministic matrices $C_l \in \mathcal{H}_d$, 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., [84, Theorem 5.8]) allows to replace the constant $1/8$ by $1/2$.

Proof of Theorems 4.2.3 and 4.2.5. Since the SCP is invariant under permutations of coordinates of X, we may and do assume that $\alpha = \alpha^{\downarrow}$. By Lemma 4.5.1 the martingale differences M_l satisfy $M_l^2 \preccurlyeq C_l^2 := 4\alpha_l^2 I_d$. Clearly

$$\Big\| \sum_{l=1}^n C_l^2 \Big\| = 4|\alpha|^2. \tag{4.5.5}$$

If X is k-homogeneous, then by Lemma 4.5.2, $N_l^2 \preccurlyeq \widetilde{C}_l^2 := 4\alpha_{\min(l,k)}^2 I_d$. In this case

$$\Big\| \sum_{l=1}^{2k} \widetilde{C}_l^2 \Big\| = 4\Big[\Big(\sum_{l=1}^k \alpha_l^2\Big) + k\alpha_k^2\Big] \leq 8 \sum_{l=1}^k \alpha_l^2. \tag{4.5.6}$$

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 4.2.3 follows by applying estimates (4.5.5) and (4.5.6) for $d = 1$ together with the classical Azuma-Hoeffding inequality. Similarly, Theorem 4.2.5 follows from the matrix version of the Azuma-Hoeffding inequality. □

It remains to prove Lemmas 4.5.1 and 4.5.2.

Proof of Lemma 4.5.1. Let $A_l^x = \{X_1 = x_1, \ldots, X_l = x_l\}$ for $x = (x_1, \ldots, x_n) \in \mathcal{B}_n$ and $l = 0, \ldots, n$. Then, for $l = 1, \ldots, n$ and any $x \in \mathcal{B}_n$ such that $\mathbb{P}(A_l^x) > 0$,

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

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

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

which is equivalent to (4.5.2). □

Proof of Lemma 4.5.2. Note that for $l \leq k$, we have $\mathcal{G}_l = \mathcal{F}_l$. As a consequence $N_l = M_l$, where M_l are martingale increments defined in Lemma 4.5.1, which implies (4.5.3).

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 $\mathbb{P}(A_l^{x,v}) > 0$,

$$\mathbb{E}[f(X) \,|\, A_l^{x,v}] = \frac{\mathbb{E}[f(X)\mathbf{1}_{A_{l-1}^{x,v}}\mathbf{1}_{\{X_{v_r}=1\}}\mathbf{1}_{\{Y_r=v_r\}}]}{\mathbb{P}(A_{l-1}^{x,v}, X_{v_r} = 1, Y_r = v_r)}. \tag{4.5.7}$$

For $s \in [r]$ let $m_s = |\{i \in [k] \colon x_i = 1\}| + |\{j \in [s-1] \colon v_j \neq 0\}|$ be the number of ones sampled by the time $k + s - 1$. It follows from (4.5.1) that if $m_s < k$ then $\mathbb{P}(A_{k+s}^{x,v}) > 0$ implies that $v_s \neq 0$ 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 (4.5.7) 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

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

Note that the right-hand side may be non-zero only if $v_r \neq 0$. In this case using the inequality $\alpha_{v_s} \leq \alpha_k$ for $s \in [k]$ we can conclude as in the proof of Lemma 4.5.1.

□

Let us now pass to the proof of Theorem 4.2.8.

Proof of Theorem 4.2.8. Let X be a random vector with law π and define the random variables Y_l for $l \leq n$ as

$$\begin{aligned} \mathcal{L}(Y_1 \,|\, X) &= \mathrm{Unif}(\operatorname{supp} X) \quad \text{and} \qquad (4.5.8)\\ \mathcal{L}(Y_l \,|\, X, Y_1, \ldots, Y_{l-1}) &= \mathrm{Unif}(\operatorname{supp} X \setminus \{Y_1, \ldots, Y_{l-1}\}), \quad \text{for} \quad l = 2, \ldots, k, \end{aligned}$$

i.e., $Y_l's$ reveal in a uniformly random order the elements of $\operatorname{supp} X$. Set $\mathcal{H}_0 = \{\emptyset, \Omega\}$ 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 [196], which asserts (in a version specialized for our application) that if $\|D_l\| \leq a$ a.s. for all l, and $\| \sum_{l=1}^{k} \mathbb{E}[D_l^2 \,|\, \mathcal{H}_{l-1}]\| \leq \sigma^2$ a.s., then for any $t \geq 0$,

$$\mathbb{P}(\|f(X) - \mathbb{E}\, f(X)\| \geq t) \leq 2d \exp\Big(-\frac{t^2}{2\sigma^2 + 2at/3}\Big). \qquad (4.5.9)$$

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 4.5.2, 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

$$D_l \mathbf{1}_{A_l^v} = \mathbb{P}(X_{v_l} = 0 \,|\, B_{l-1}^v)\big(\mathbb{E}[f(X) \,|\, B_{l-1}^v, X_{v_l} = 1] - \mathbb{E}[f(X) | B_{l-1}^v, X_{v_l} = 0]\big)\mathbf{1}_{A_l^v}. \qquad (4.5.10)$$

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

$$\mathbb{E}[f(X) \,|\, B_{l-1}^v, X_{v_l} = 1] - \mathbb{E}[f(X) \,|\, B_{l-1}^v, X_{v_l} = 0] = \mathbb{E}[f(\tilde{Z}) - f(\hat{Z})], \qquad (4.5.11)$$

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

$$\begin{aligned} \Big(\mathbb{E}[f(X) \,|\, B_{l-1}^v, X_{v_l} = 1] - \mathbb{E}[f(X) \,|\, B_{l-1}^v, X_{v_l} = 0]\Big)^2 & \qquad (4.5.12)\\ &= \Big(\mathbb{E}[f(\tilde{Z}) - f(\hat{Z})]\Big)^2\\ &\preccurlyeq \mathbb{E}\Big[\big(f(\tilde{Z}) - f(\hat{Z})\big)^2\Big]\\ &= \mathbb{E}\Big[\big(f(\tilde{Z}) - f(\tilde{Z}^{v_l}) + f(\hat{Z}^{\tilde{Y}_l}) - f(\hat{Z})\big)^2\Big]\\ &\preccurlyeq 2\,\mathbb{E}\Big[\big(f(\tilde{Z}) - f(\tilde{Z}^{v_l})\big)^2\Big] + 2\,\mathbb{E}\Big[\big(f(\hat{Z}^{\tilde{Y}_l}) - f(\hat{Z})\big)^2\Big]\\ &\preccurlyeq 2C_{v_l}^2 + 2\,\mathbb{E}\, C_{\tilde{Y}_l}^2, \end{aligned}$$

where in the first and second inequality we used the operator convexity of the function $x \mapsto x^2$ (see [33, Example V.1.3]), and in the last inequality the assumption (4.2.3).

In particular, using (4.5.10), 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 (4.5.10) and (4.5.12) 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 as

$$D_l^2 \preccurlyeq 2 \sum_{v_l \in [n] \setminus \{v_1, \ldots, v_{l-1}\}} \big(C_{Y_l}^2 + \mathbb{E}[C_{\tilde{Y}_l}^2 \,|\, A_l^v]\big)\, \mathbb{P}(X_{v_l} = 0 \,|\, B_{l-1}^v)^2 \mathbf{1}_{A_l^v}. \tag{4.5.13}$$

Let us now go back to the equations (4.5.10) and (4.5.11) 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

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

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 (4.5.13) 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

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

where in the last inequality we used [131, Lemma 1.10], which asserts that $\mathbb{P}(X_{v_l} = 1) \geq \mathbb{P}(X_{v_l} = 1|B^v_{l-1})$ (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 (4.5.9) 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. □

4.6 Proofs of the results of Section 4.4

4.6.1 Propositions 4.4.7 and 4.4.9

The main idea behind the proof of Proposition 4.4.7 is to find an estimate on $\|\Gamma_+(f)\|_\infty$ in terms of α, refining (4.4.6), 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 4.4.9 as well.

Lemma 4.6.1. *Let $t = (t_1, \ldots, t_n)$ be a sequence of nonnegative numbers and let $D_1, \ldots, D_n \in \mathcal{H}_d$ be positive semidefinite matrices. Then for any $T_1 \geq |t|_1$ and $T_\infty \geq |t|_\infty$*

$$\Big\| \sum_{i=1}^{n} t_i D_i \Big\| \leq T_\infty \cdot \sup\Big\{ \Big\| \sum_{i \in \mathcal{I}} D_i \Big\| \colon \mathcal{I} \subset [n],\, |\mathcal{I}| \leq \lceil T_1/T_\infty \rceil \Big\}. \tag{4.6.1}$$

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 (4.6.1) equals to $\max\{\| \sum_{i=1}^n y_i D_i\| \colon y \in \mathcal{Y}\}$, while 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 4.4.7. We recall that for $x \in \mathcal{B}_n$ 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 $x \in \mathcal{B}_n$, using the definition (4.4.3) of Γ_+, Lipschitz property of f and inequality $(a+b)^2 \le 2(a^2+b^2)$ we get

$$\begin{aligned}\Gamma_+(f)(x) &= \sum_{i=1}^n (f(x)-f(x^i))_+^2 L(x,x^i) \\ &\qquad\qquad\qquad + \frac{1}{2}\sum_{i,j=1}^n (f(x)-f(x^{ij}))_+^2 L(x,x^{ij}) \\ &\le \sum_{i=1}^n \alpha_i^2 L(x,x^i) + \frac{1}{2}\sum_{i,j=1}^n (\alpha_i+\alpha_j)^2 L(x,x^{ij}) \mathbf{1}_{\{x\neq x^{ij}\}} \qquad (4.6.2)\\ &\le \sum_{i=1}^n \alpha_i^2 L(x,x^i) + 2\sum_{i=1}^n \alpha_i^2 \sum_{j=1}^n L(x,x^{ij})\mathbf{1}_{\{x\neq x^{ij}\}} \\ &\le 2\sum_{i=1}^n \alpha_i^2 \sum_{y\colon y_i\neq x_i} L(x,y).\end{aligned}$$

Whence, by the stability condition (4.4.7) we estimate $\|\Gamma_+(f)\|_\infty \le 2R\rho(L)|\alpha|^2$. Herbst's argument (4.4.5) allows to conclude the first part.

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

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

so by (4.6.2), Lemma 4.6.1 applied in the scalar setting $d=1$ with $t_i = 2\sum_{y\colon y_i\neq x_i} L(x,y)$, $D_i=\alpha_i^2$, $T_1 = 4\Delta(L)$ and $T_\infty = 4R\rho(L)$ we can estimate

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

and conclude again in virtue of Herbst's argument (4.4.5). □

The proof of Proposition 4.4.9 follows along similar lines to the proof of Proposition 4.4.7, the difference being that in the end, instead of Herbst's argument, we apply the concentration result of Aoun et al. [16], which asserts that if L satisfies the matrix Poincaré inequality with constant $C_P > 0$

$$\mathrm{Var}(f) \preccurlyeq -C_P\pi(fLf) \quad \forall f\colon \mathcal{B}_n \to \mathcal{H}_d, \qquad (4.6.3)$$

(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

$$\pi\big(\lambda_{max}(f-\pi(f)) > t\big) \le d\exp\left(\frac{-t^2}{2C_P v_f + t\sqrt{2C_P v_f}}\right), \qquad (4.6.4)$$

where $v_f = \sup_x \|\Gamma(f)(x)\|$ (where Γ is defined via (4.4.2), again with matrix multiplication, and $\|\cdot\|$ stands for the operator norm). Note that for $d=1$, (4.6.3) is just the usual scalar Poincaré inequality.

Proof of Proposition 4.4.9. For any $x\in\mathcal{B}_n$ and $i,j\in[n]$, using operator convexity of the function $x\mapsto x^2$ (see [33, Example V.1.3]) we get that

$$\begin{aligned}\big(f(x)-f(x^{ij})\big)^2 &= \big[\big(f(x)-f(x^i)\big)+\big(f(x^i)-f(x^{ij})\big)\big]^2 \\ &\preccurlyeq 2\big(f(x)-f(x^i)\big)^2 + 2\big(f(x^i)-f(x^{ij})\big)^2. \qquad (4.6.5)\end{aligned}$$

Whence, by the definition (4.4.2) of Γ, by the assumed Lipschitz property (4.4.8) of f and by (4.6.5), for any $x \in \mathcal{B}_n$,

$$\begin{aligned}\Gamma(f)(x) &= \frac{1}{2}\sum_{i=1}^{n}(f(x)-f(x^i))^2 L(x,x^i) + \frac{1}{4}\sum_{i,j=1}^{n}(f(x)-f(x^{ij}))^2 L(x,x^{ij}) \\ &\preccurlyeq \frac{1}{2}\sum_{i=1}^{n} C_i^2 L(x,x^i) + \frac{1}{2}\sum_{i,j=1}^{n}(C_i^2+C_j^2)L(x,x^{ij})\mathbf{1}_{\{x\neq x^{ij}\}} \\ &\preccurlyeq \sum_{i=1}^{n} C_i^2 \cdot \Big[\sum_{y\colon y_i\neq x_i} L(x,y)\Big].\end{aligned} \tag{4.6.6}$$

As both hand sides of (4.6.6) are positive semidefinite, their norms compare as well. Therefore, as in the proof of Proposition 4.4.7, by Lemma 4.6.1 with $t_i = \sum_{y\colon y_i\neq x_i} L(x,y)$, $T_1 = 2\Delta(L)$, $T_\infty = 2R\rho(L)$ and $D_i = C_i^2$

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

Since L satisfies the (scalar) modified log-Sobolev inequality (4.4.4), then it satisfies the (scalar) Poincaré inequality with constant $C_P = 2/\rho(L)$ (see, e.g., [42, p. 292], noting slightly different definitions of constants in functional inequalities used therein) and whence by [122, Proposition 2.2] or [93, Theorem 1.1] it satisfies the matrix Poincaré inequality (4.6.3) with the same constant, which yields the conclusion in virtue of (4.6.4). □

4.6.2 Proposition 4.4.10

The proof of Proposition 4.4.10 is based on the idea introduced by Boucheron et al. [48] for independent random variables and then developed by Paulin [167] for Glauber dynamics under the Dobrushin condition. We follow the exposition introduced in the recent works of Sambale and Sinulis [180, 179] in the context of sampling without replacement and adapt it to the more abstract setting involving the stability condition given in Definition 4.4.3.

We start with the following lemmas.

Lemma 4.6.2. *For any flip-swap random walk with generator L satisfying the stability condition (4.4.7) and for any $A\subset\mathcal{B}_n$*

$$\Gamma_+(d_T^2(\cdot,A))(x) \le 8R\rho(L)\cdot d_T^2(x,A). \tag{4.6.7}$$

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

$$d_T^2(x,A) - d_T^2(y,A) \le d_H(x,y). \tag{4.6.8}$$

Proof. For $x\in\mathcal{B}_n$, $\alpha\in\mathbb{R}^n$ and a probability measure μ on $\mathcal{B}_n$, let $h_x(\mu,\alpha) = \sum_i \alpha_i\mu(z\colon z_i\neq x_i)$. By Sion's minmax theorem, cf. [51, p. 227],

$$d_T(x,A) = \inf_{\mu\in\mathcal{M}(A)}\sup_{\alpha\in B_2^n} h_x(\mu,\alpha), \tag{4.6.9}$$

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 (4.4.7). We conclude (4.6.7) 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 (4.6.9) 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 $x,y\in\mathcal{B}_n$

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

as desired. $\square$

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

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

Therefore, the modified log-Sobolev inequality (4.4.4) implies the following inequality stated in Bobkov and Götze [37]:

$$\rho(L)\operatorname{Ent}_\pi(e^f) \le \pi\big(e^f\tilde\Gamma(f)^2\big) \tag{4.6.10}$$

with operator $\tilde\Gamma(f)=\sqrt{\Gamma_+(f)}$ (note that in [37] $\tilde\Gamma$ is denoted by Γ, we use $\tilde\Gamma$ to avoid a conflict of notation). As a consequence, the hypothesis of [37, Theorem 2.1] (formula (1.1)) therein holds under the assumption of the modified log-Sobolev inequality (4.4.4) (with $c=2/\rho(L)$). As a result, the following lemma follows directly by the derivation of [37, equation (2.4)] with a slight adjustment of constants (see also [10]).

Lemma 4.6.3. *If a measure π on $\mathcal{B}_n$ satisfies the modified log-Sobolev inequality (4.4.4) and $f\colon \mathcal{B}_n \to [0,\infty)$ is such that $\Gamma_+(f) \le Cf$ for some constant $C > 0$, then for all $t > C/\rho(L)$,*

$$\pi\big(\exp(f/t)\big) \le \exp\Big(\frac{\pi(f)}{t - C/\rho(L)}\Big). \tag{4.6.11}$$

We are finally in position to prove Proposition 4.4.10.

Proof of Proposition 4.4.10. To lighten notation, denote $f(x) = d_T^2(x, A)$ for $x \in \mathcal{B}_n$ and some fixed set $A \subset \mathcal{B}_n$. Denote also $h(z) = (e^z - 1)/z$ for $z \in [0,\infty)$ and $Df_y(x) = f(x) - f(y)$ for $x, y \in \mathcal{B}_n$, and note that h is an increasing function. Starting with the modified log-Sobolev inequality (4.4.4) and using the reversibility of L, we have for all $\lambda > 0$,

$$\begin{aligned}
\mathrm{Ent}_\pi(e^{-\lambda f}) &\le \lambda/\rho(L)\cdot \mathcal{E}(e^{-\lambda f}, -f)\\
&= \lambda/\rho(L)\sum_{x,y}\big(Df_y(x)\big)_+\big(e^{-\lambda f(y)} - e^{-\lambda f(x)}\big)L(x,y)\pi(x)\\
&= \lambda^2/\rho(L)\sum_x \pi(x)e^{-\lambda f(x)}\Big[\sum_y \big(Df_y(x)\big)_+^2 h\big(\lambda Df_y(x)\big)L(x,y)\Big]\\
&\overset{\text{by (4.6.8)}}{\le} \lambda^2 h(2\lambda)/\rho(L)\cdot \pi\big(e^{-\lambda f}\Gamma_+(f)\big)\\
&\le 8R\lambda^2 h(2\lambda)\cdot \pi\big(e^{-\lambda f} f\big)\\
&\overset{\text{by (4.6.7)}}{\le} 8R\lambda^2 h(2\lambda)\cdot \pi(e^{-\lambda f})\pi(f),
\end{aligned}$$

where the last inequality follows by convexity of the function $t \mapsto t\log t$. Therefore, using the entropy method (cf., e.g., [51, 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 $\log(1+x) \ge x/(x+1)$ for $x \ge 1$ gives

$$\pi(A) \le \exp\Big(-\frac{1}{4}\log\big(1 + \frac{1}{8R}\big)\pi(f)\Big) \le \exp\Big(-\frac{\pi(f)}{32R + 4}\Big). \tag{4.6.12}$$

We conclude by dividing (4.6.12) by its right-hand side and using Lemma 4.6.3 with $t = 4 + 40R$ and $C = 8R\rho(L)$ (in virtue of Lemma 4.6.2). ☐

4.6.3 Proposition 4.4.11

We prove Proposition 4.4.11 with help of Proposition 2.3.1 from Chapter 2, which we restate below in the current setting.

Proposition 4.6.4. *If a probability measure π on $\mathcal{B}_n$ satisfies the modified log-Sobolev inequality (4.4.4), then for any $p \geq 2$,*

$$\|(f-\pi(f))_+\|_p \leq C\sqrt{p/\rho(L)}\|\sqrt{\Gamma_+(f)}\|_p, \tag{4.6.13}$$

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

A general method of deriving estimates for polynomials from moment inequalities of the form (4.6.13) has been presented in [8] in the continuous case, and in [4, 6] in the context of Glauber dynamics, cf. Section 2.5. To obtain results for flip-swap random walks we will adapt a version of this method introduced recently by Sambale and Sinulis [179] for multislices. Note that we cannot apply Corollary 2.5.4 directly as a flip-swap kernel allows changing two coordinates simultaneously,

Proof of Proposition 4.4.11. 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 $f\colon \mathcal{B}_n \to \mathbb{R}$ be a tetrahedral polynomial. By ∂_i we denote the partial derivative with respect to the i-th coordinate. If $x, y \in \mathcal{B}_n$ 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

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

Thus,

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

where in the last inequality we used the stability condition (4.4.7). Note that since f is tetrahedral, $\partial_i\partial_i f(x) = 0$ for all i.

Combining the above equality with Proposition 4.6.4 we obtain that for every tetrahedral polynomial $f\colon \mathcal{B}_n \to \mathbb{R}$

$$\|f-\pi(f)\|_p \leq C\sqrt{p}\sqrt{R}\Big(\big\||\nabla f|\big\|_p + \big\|\|\nabla^2 f\|_{HS}\big\|_p\Big), \tag{4.6.14}$$

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 $\mathcal{J} = \{J_1, \ldots, J_l\} \in P_{d,\le 2}$ 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 (4.6.14) as

$$\|f(X) - \mathbb{E} f(X)\|_p \le C\sqrt{R}\Big(\|\langle \nabla f(X), G^{\{1\}}\rangle\|_p + \|\langle \nabla^2 f(X), G^{\{1,2\}}\rangle\|_p\Big), \tag{4.6.15}$$

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

The inequality (4.6.15) constitutes a basis for the induction argument leading to the following inequality valid for any $f \colon \mathcal{B}_n \to \mathbb{R}$, $d \ge 1$ and $p \ge 2$,

$$\begin{aligned}\|f(X) - \mathbb{E} f(X)\|_p \le & C_d\Big(\sum_{l=d}^{2d} \sum_{\mathcal{J}\in P_{l,\le 2}} R^{|\mathcal{J}|/2} \|\langle \nabla^l f(X), G_{\mathcal{J}}\rangle\|_p \\ & + \sum_{l=1}^{2d-2} \sum_{\mathcal{J}\in P_{l,\le 2}} R^{|\mathcal{J}|/2} \|\langle \mathbb{E}_X \nabla^l f(X), G_{\mathcal{J}}\rangle\|_p\Big).\end{aligned} \tag{4.6.16}$$

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, (4.6.16) 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 [135], which assert that for any l-tensor $A = (a_{\mathbf{i}})_{\mathbf{i}\in[n]^l}$ and $p \ge 2$,

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

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 4.4.4), which gives $R^{|\mathcal{J}|/2} \le C_d R^{l/2}$ for $\mathcal{J} \in P_{l,\le 2}$, we get

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

for $p \ge 2$. This implies the tail inequality of the proposition in the standard way by the use of Chebyshev's inequality $\mathbb{P}(|f(X) - \mathbb{E}\, f(X)| \ge e\|f(X) - \mathbb{E}\, f(X)\|_p) \le e^{-p}$ followed by an appropriate change of variables and adjustment of constants. We leave the details to the Reader and turn to the proof of (4.6.16).

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 (4.6.16) reads as

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

which is clearly weaker than (4.6.15). 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)

$$\begin{aligned}\|f(X) - \mathbb{E}\, f(X)\|_p \le & C_d\Big(\sum_{l=d-1}^{2d-2} \sum_{\mathcal{J}\in P_{l,\le 2}} R^{|\mathcal{J}|/2}\|\langle \nabla^l f(X), G_{\mathcal{J}}\rangle\|_p \\ & + \sum_{l=1}^{2d-4} \sum_{\mathcal{J}\in P_{l,\le 2}} R^{|\mathcal{J}|/2}\|\langle \mathbb{E}_X\, \nabla^l f(X), G_{\mathcal{J}}\rangle\|_p\Big) \\ \le & C_d\Big(\sum_{l=d-1}^{2d-2} \sum_{\mathcal{J}\in P_{l,\le 2}} R^{|\mathcal{J}|/2}\|\langle \nabla^l f(X), G_{\mathcal{J}}\rangle - \langle \mathbb{E}_X\, \nabla^l f(X), G_{\mathcal{J}}\rangle\|_p \\ & + \sum_{l=1}^{2d-2} \sum_{\mathcal{J}\in P_{l,\le 2}} 2R^{|\mathcal{J}|/2}\|\langle \mathbb{E}_X\, \nabla^l f(X), G_{\mathcal{J}}\rangle\|_p\Big).\end{aligned} \tag{4.6.17}$$

An application of inequality (4.6.15) 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 $\mathcal{J} \in P_{l,\le 2}$ (note that $h_{l,\mathcal{J}}$'s are tetrahedral polynomials), followed by the Fubini theorem, gives

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

which combined with (4.6.17) concludes the induction step, proving (4.6.16). □

4.7 Proofs of the results of Section 4.3

By virtue of the abstract results of Section 4.4, all the results of Section 4.3 will follow if one proves that there exist a flip-swap random walks on $\mathcal{B}_n$ with stationary measure $\pi = \pi(p,k)$ which satisfy the stability condition (4.4.7) with constant $R = 2$ for all $p \in (0,1)^n$ and $k = 0,\ldots,n$ (cf. Theorem 4.7.3). 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 [114]. 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 (4.1.1), 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 4.7.1. *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 $x_r = 0$,*

$$\mathbb{P}(Z = x + e_r \,|\, Z' = x) = \mathbb{E}\left[\frac{\mathbf{1}_{\{Z_r=1\}}}{\sum_{l=1}^n \mathbf{1}_{\{Z_l=1\}}\mathbf{1}_{\{x_l=0\}}}\right] \tag{4.7.1}$$

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

$$\mathbb{P}(Z' = x - e_r \,|\, Z = x) = \mathbb{E}\left[\frac{\mathbf{1}_{\{Z'_r=0\}}}{\sum_{l=1}^n \mathbf{1}_{\{Z'_l=0\}}\mathbf{1}_{\{x_l=1\}}}\right]. \tag{4.7.2}$$

Let us now recall the inductive construction of Hermon and Salez [114] in the k-homogeneous case. It works for any k-homogeneous probability measure π on $\mathcal{B}_n$, 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 $x \in \mathcal{B}_n$ and $l \in [n]$ sometimes as elements of $\{0,1\}^{[n]\setminus\{l\}}$ (this is how they were defined at the beginning of Section 4.2) 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

$$Q^{(l)}(x,y) = \begin{cases} \mathbb{P}\left(U = y_{\neq l} \,|\, V = x_{\neq l}\right)\mathbb{P}\left(X_l \neq x_l\right) & \text{if} \quad x_l \neq y_l, \\ Q^{(l)}_{x_l}(x_{\neq l}, y_{\neq l}) & \text{else,} \end{cases} \tag{4.7.3}$$

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[2] 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

[2] We note a small typo in the cited arxiv version of [114] – for Q to be self-adjoint we need to have $\pi(x)$ in the denominator of each expression in [114, equation (90)]. This change is also consistent with the subsequent part of the proof in [114].

$\rho(Q^{(l)}_{x_l}) \geq 1$ 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

$$Q^* = \frac{1}{n}\sum_{l=1}^{n} Q^{(l)}. \tag{4.7.4}$$

Then by (the proof of) [114, Theorem 2], 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 4.7.1 applied in dimension $n-1$ with $p_{\neq l}$ instead of p (note that since the right-hand side of (4.7.1) 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

$$L_\pi = \frac{1}{n}\sum_{l=1}^{n} L^{(l)} \tag{4.7.5}$$

with

$$L^{(l)}(x,y) = \begin{cases} \mathbb{P}\left(U = y_{\neq l} \,|\, V = x_{\neq l}\right)\mathbb{P}\left(X_l \neq x_l\right) & \text{if} \quad x_l \neq y_l, \\ L_{\pi_l}(x_{\neq l}, y_{\neq l}) & \text{else}, \end{cases} \tag{4.7.6}$$

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 4.7.1, 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 4.7.2. *The generator L_π constructed according to (4.7.5) 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 4.3 is

Theorem 4.7.3. *The generator L_π constructed according to (4.7.5) with stationary measure π satisfies the stability condition (4.4.7) with $R = 2$.*

Proof of Theorem 4.7.3. 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 (4.4.7) vanishes. Thus, the stability condition (4.4.7) 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 $k \in \{1,\dots,n-1\}$ as otherwise L_π trivializes.

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

$$\max_{x \in \operatorname{supp}\pi;\, i\in[n]} \sum_{y\colon y_i \neq x_i} L_\pi(x,y) \leq 2. \tag{4.7.7}$$

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

If $x_i = 0$, then by (4.7.5)

$$\begin{aligned}\sum_{y\colon y_i\neq x_i} L_\pi(x,y) &= \sum_{j\colon x_j=1} \frac{1}{n}\sum_{l=1}^n L^{(l)}(x,x^{ij})\\ &= \frac{1}{n}\sum_{j\colon x_j=1}\sum_{l\in[n]\setminus\{i,j\}} L^{(l)}(x,x^{ij}) + \frac{1}{n}\sum_{j\colon x_j=1} L^{(i)}(x,x^{ij})\\ &\qquad + \frac{1}{n}\sum_{j\colon x_j=1} L^{(j)}(x,x^{ij}),\end{aligned} \tag{4.7.8}$$

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 (4.7.6), $L^{(l)}(x,y) = L_{\pi_l}(x_{\neq l}, y_{\neq l})$, where $\pi_l = \pi(p_{\neq l}, k - x_l)$. Thus, denoting $r_l = \zeta_l(i)$, we get

$$\begin{aligned}\frac{1}{n}\sum_{j\colon x_j=1}\sum_{l\in[n]\setminus\{i,j\}} L^{(l)}(x,x^{ij}) &= \frac{1}{n}\sum_{l\in[n]\setminus\{i\}}\sum_{j\neq l\colon x_j=1} L_{\pi_l}(x_{\neq l},(x^{ij})_{\neq l})\\ &= \frac{1}{n}\sum_{l\in[n]\setminus\{i\}}\sum_{y\in\mathcal{B}_{n-1}\colon y_{r_l}\neq(x_{\neq l})_{r_l}} L_{\pi_l}(x_{\neq l},y) \leq \frac{n-1}{n}\cdot 2,\end{aligned} \tag{4.7.9}$$

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

The second term of (4.7.8) is estimated again using the definition (4.7.6). 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

$$\begin{aligned}\frac{1}{n}\sum_{j\colon x_j=1} L^{(i)}(x,x^{ij}) &= \frac{1}{n}\sum_{j\colon x_j=1}\mathbb{P}\left(U = (x^{ij})_{\neq i}\,|\,V = x_{\neq i}\right)\mathbb{P}\left(X_i\neq x_i\right)\\ &= \frac{1}{n}\mathbb{P}\left(X_i = 1\right) \leq \frac{1}{n}.\end{aligned} \tag{4.7.10}$$

Let us pass to the last term of (4.7.8). 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 (4.7.1) from Lemma 4.7.1 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 (4.7.6) is the coupling between the laws $\pi(p_{\neq j}, k)$ and $\pi(p_{\neq j}, k-1)$ given by Lemma 4.7.1 (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

$$\sum_{l=1}^n \mathbf{1}_{\{X_l=0\}}\mathbf{1}_{\{x_l=1\}} = \sum_{l=1}^n \mathbf{1}_{\{X_l=1\}}\mathbf{1}_{\{x_l=0\}}. \tag{4.7.11}$$

Putting all the above observations together and using Lemma 4.7.1 together with (4.7.6) in the first step, we obtain

$$\begin{aligned}
\frac{1}{n}\sum_{j:\,x_j=1} L^{(j)}(x,x^{ij}) &= \frac{1}{n}\sum_{j:\,x_j=1}\mathbb{E}\Big[\frac{\mathbf{1}_{\{Z_i^{(j)}=1\}}}{\sum_{l\neq j}\mathbf{1}_{\{Z_l^{(j)}=1\}}\mathbf{1}_{\{x_l=0\}}}\Big]\,\mathbb{P}\big(X_j=0\big)\\
&= \frac{1}{n}\sum_{j:\,x_j=1}\mathbb{E}\Big[\mathbf{1}_{\{X_i=1\}}\frac{\mathbf{1}_{\{X_j=0\}}}{\sum_{l\neq j}\mathbf{1}_{\{X_l=1\}}\mathbf{1}_{\{x_l=0\}}}\Big]\\
&\overset{x_j\,=\,1}{=} \frac{1}{n}\sum_{j:\,x_j=1}\mathbb{E}\Big[\mathbf{1}_{\{X_i=1\}}\frac{\mathbf{1}_{\{X_j=0\}}}{\sum_{l}\mathbf{1}_{\{X_l=1\}}\mathbf{1}_{\{x_l=0\}}}\Big]\\
&\overset{(4.7.11)}{=} \frac{1}{n}\sum_{j:\,x_j=1}\mathbb{E}\Big[\mathbf{1}_{\{X_i=1\}}\frac{\mathbf{1}_{\{X_j=0\}}}{\sum_{l}\mathbf{1}_{\{X_l=0\}}\mathbf{1}_{\{x_l=1\}}}\Big]\\
&= \frac{1}{n}\,\mathbb{P}\big(X_i=1\big)\le\frac{1}{n}.
\end{aligned}\tag{4.7.12}$$

Combining the estimates (4.7.9), (4.7.10) and (4.7.12) with (4.7.8) yields inequality (4.7.7) and thus the stability condition (4.4.7) with $R=2$ in the case $x_i=0$. The case $x_i=1$ is analogous, the main difference being that in (4.7.12) we use (4.7.2) in place of (4.7.1) from Lemma 4.7.1.

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

Let us conclude this section with the proof of Lemma 4.7.1.

Proof of Lemma 4.7.1. For $x\in\mathcal{B}_n$, let $\kappa(x)=\sum_i x_i$ and let B be a vector of independent Bernoulli random variables with probabilities of success given by $p=(p_1,\ldots,p_n)$. Consider three $\mathcal{B}_n$-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 $x,y\in\mathcal{B}_n$,

$$\mathbb{P}(Z=y|Z'=x)=h(y,x),\tag{4.7.13}$$

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 $x\in\mathcal{B}_n$ 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

$$\sum_{y\in\mathcal{B}_n}h(y,x)=\sum_{r:\,x_r=0}h(x+e_r,x)=1,$$

which guarantees the existence of the couple (Z,Z') satisfying (4.7.13). Thus, to prove (4.7.1) it is enough to show that $Z\sim\widehat{Z}$, i.e., that $\sum_{x\in\mathcal{B}_n}h(y,x)\,\mathbb{P}(Z'=x)=\mathbb{P}(\widehat{Z}=y)$ for any $y\in\mathcal{B}_n$ such that $\kappa(y)=k$.

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

$$\begin{aligned}
\frac{\mathbb{P}\left(Z'=x\right)}{\mathbb{P}(\widehat{Z}=x+e_r)} &= \frac{\mathbb{P}\left(B=x\right)}{\mathbb{P}\left(B=x+e_r\right)}\frac{\mathbb{P}\left(\kappa(B)=k\right)}{\mathbb{P}\left(\kappa(B)=k-1\right)}\\
&= \frac{1-p_r}{p_r}\frac{\mathbb{P}\left(\kappa(B)=k\right)}{\mathbb{P}\left(\kappa(B)=k-1\right)}.
\end{aligned}\tag{4.7.14}$$

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

$$\mathbb{E}\left[f(B)\mathbf{1}_{\{B_r=1\}}\right]\frac{1-p_r}{p_r} = \mathbb{E}\left[f(B+e_r)\mathbf{1}_{\{B_r=0\}}\right]. \tag{4.7.15}$$

We use (4.7.14) and (4.7.15) to get that for any such y and any $r \in [n]$ such that $y_r = 1$ and $\kappa(y) = k$,

$$\begin{aligned}
h(y, y-e_r)&\frac{\mathbb{P}(Z' = y-e_r)}{\mathbb{P}(\widehat{Z} = y)} \\
&\overset{(4.7.14)}{=} \quad h(y,y-e_r)\frac{1-p_r}{p_r}\frac{\mathbb{P}\left(\kappa(B)=k\right)}{\mathbb{P}\left(\kappa(B)=k-1\right)} \\
&= \quad \mathbb{E}\left[\frac{\mathbf{1}_{\{B_r=1\}}\mathbf{1}_{\{\kappa(B)=k\}}}{\mathbf{1}_{\{B_r=1\}}+\sum_{l\neq r}\mathbf{1}_{\{B_l=1\}}\mathbf{1}_{\{y_l=0\}}}\right]\frac{(1-p_r)/p_r}{\mathbb{P}\left(\kappa(B)=k-1\right)} \\
&\overset{(4.7.15)}{=} \quad \mathbb{E}\left[\frac{\mathbf{1}_{\{B_r=0\}}\mathbf{1}_{\{\kappa(B)=k-1\}}}{\mathbf{1}_{\{B_r=0\}}+\sum_{l\neq r}\mathbf{1}_{\{B_l=1\}}\mathbf{1}_{\{y_l=0\}}}\right]\frac{1}{\mathbb{P}\left(\kappa(B)=k-1\right)} \\
&= \quad \mathbb{E}\left[\frac{\mathbf{1}_{\{B_r=0\}}\mathbf{1}_{\{\kappa(B)=k-1\}}}{\mathbf{1}_{\{B_r=0\}}+\sum_{l\neq r}\mathbf{1}_{\{B_l=0\}}\mathbf{1}_{\{y_l=1\}}}\right]\frac{1}{\mathbb{P}\left(\kappa(B)=k-1\right)} \\
&= \quad \mathbb{E}\left[\frac{\mathbf{1}_{\{Z'_r=0\}}}{\sum_l \mathbf{1}_{\{Z'_l=0\}}\mathbf{1}_{\{y_l=1\}}}\right],
\end{aligned} \tag{4.7.16}$$

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 $\mathcal{B}_n$ 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 (4.7.16)

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

which completes the proof of (4.7.1). The equality (4.7.2) follows again by using (4.7.16). $\square$

Chapter 5

Sampling without replacement and Hoeffding statistics

5.1 Preliminaries

In this chapter we investigate concentration properties of particular functionals of uniform random permutations, complementing on the results from Section 2.4.4. Namely, we focus on the suprema of empirical processes when sampling without replacement. Such processes can be seen as Hoeffding statistics for matrices of a special form with repeated rows. We also obtain corresponding bounds for a single Hoeffding statistic for general underlying matrix. Such bounds were considered extensively in the literature, cf., e.g, [30, 51, 193], and they play an important role in various applications, e.g., in transductive learning [195], or statistical testing [11].

5.1.1 Organization of this chapter

In the rest of this section we introduce some core notation. In Section 5.2 we present our results concerning concentration for suprema of empirical processes when sampling without replacement. In Section 5.3 we present analogous results for a single Hoeffding statistic. We provide remaining proofs of our concentration estimates in Section 5.4. Proofs of auxiliary facts are moved to Appendix.

5.1.2 Basic notation

Let us recall the notation of Section 2.4.4. For $n \in \mathbb{N}$, consider the symmetric group S_n of permutations of the set $[n] := \{1, \ldots, n\}$ equipped with the uniform probability measure π_n. It is the stationary distribution of the *interchange process* defined via its generator L given by the formula

$$
\begin{aligned}
Lf(\sigma) &= \frac{1}{n(n-1)} \sum_{i,j=1}^{n} \big(f(\sigma \circ \tau_{ij}) - f(\sigma)\big) \\
&= \frac{2}{n(n-1)} \sum_{1 \leq i < j \leq n} \big(f(\sigma \circ \tau_{ij}) - f(\sigma)\big),
\end{aligned}
$$

where τ_{ij} stands for the transposition of elements i and j. By $\mathbb{E}$, we denote the expectation w.r.t. π_n. Moreover, for a function $f \colon S_n \to \mathbb{R}$, denote $f_{ij}(\cdot) =$

$f(\cdot \circ \tau_{ij})$ for short. The corresponding Dirichlet form is then expressed as

$$\begin{aligned}\mathcal{E}(f,g) &= \frac{1}{2n(n-1)}\,\mathbb{E}\sum_{i,j=1}^{n}(f_{ij}-f)(g_{ij}-g)\\ &= \frac{1}{n(n-1)}\,\mathbb{E}\sum_{1\le i<j\le n}(g_{ij}-g)(f_{ij}-f).\end{aligned}$$

If f and g have the same monotonicity, then by the reversibility of L we also have

$$\mathcal{E}(f,g) = \frac{1}{n(n-1)}\,\mathbb{E}\sum_{i,j=1}^{n}(g_{ij}-g)_+(f_{ij}-f)_+.$$

Recall that the *modified log-Sobolev* inequality is satisfied with constant $\rho_0 > 0$ if

$$\rho_0 \operatorname{Ent}_\mu(f) \le \mathcal{E}(f, \log f) \tag{5.1.1}$$

for all positive functions f. For this process, $\rho_0 \ge \frac{1}{n-1}$ was obtained independently by Gao–Quastel [91] and Bobkov–Tetali [42] (note that the normalization of the generator L differs across various references – we provide here scaled constants matching our setting).

5.2 Sampling without replacement – concentration for suprema

Consider a set of vectors $\mathcal{X} \subset \mathbb{R}^n$. As in Section 2.4.4, let $I_1, \ldots, I_n$ be a uniform sample without replacement and $J_1, \ldots, J_n$ be a sample with replacement from the set $[n]$. For $m \le n$, define

$$Z = \sup_{x\in\mathcal{X}}\sum_{k=1}^{m} x_{I_k}, \qquad Z' = \sup_{x\in\mathcal{X}}\sum_{k=1}^{m} x_{J_k} \tag{5.2.1}$$

so that Z' can be considered a supremum of the empirical process in independent random variables J_k.

To analyze the tails of Z, it is often convenient to represent it as a supremum of Hoeffding statistics over a family of matrices. Namely, for $x \in \mathcal{X}$, denote $a^x \in \mathbb{R}^{n\times n}$ to be such that the first m rows of a consist of copies of vector x and the remaining rows have zero entries only, i.e., $a_{ij} = x_j$ for $i \le m$, $j \in [n]$ and $a_{ij} = 0$ for $i > m$, $j \in [n]$. Then

$$Z = \sup_{x\in\mathcal{X}}\sum_{k=1}^{n} a^x_{k\sigma_k},$$

where $\sigma = (I_1, I_2, \ldots, I_n) \sim \pi_n$. Moreover, denote $\sigma_{ij} = \sigma \circ \tau_{ij}$ for any $i, j \in [n]$ and

$$Z_{ij} = \sup_{x\in\mathcal{X}}\sum_{k=1}^{n} a^x_{k\sigma_{ij}(k)},$$

so that the modified log-Sobolev inequality (5.1.1) applied to the Laplace transform of Z reads

$$\operatorname{Ent}(e^{\lambda Z}) \le \frac{\lambda}{n}\,\mathbb{E}\, e^{\lambda Z}\sum_{ij}(1-e^{-\lambda(Z-Z_{ij})})_+(Z-Z_{ij})_+.$$

In the sequel, we express our concentration results for Z using expectations of the following random variables

$$\Sigma^2 = \sup_{x\in\mathcal{X}} \sum_{k=1}^m x_{I_k}^2, \qquad \widetilde{\Sigma}^2 = \sup_{x\in\mathcal{X}} \sum_{k=1}^m x_{J_k}^2.$$

As pointed out in [107], it follows from an argument due to Hoeffding [117] (cf. also [145]) that if E is a normed space and $g\colon [n] \to E$, then for any convex function $\Psi\colon E \to \mathbb{R}$,

$$\mathbb{E}\,\Psi\Big(\sum_{k=1}^m g(I_k)\Big) \le \mathbb{E}\,\Psi\Big(\sum_{k=1}^m g(J_k)\Big). \tag{5.2.2}$$

The meaning of (5.2.2) in terms of Z and Z' and related quantities is explained in the following lemma, which in particular implies that $\mathbb{E}\,Z \le \mathbb{E}\,Z'$ and $\mathbb{E}\,\Sigma^2 \le \mathbb{E}\,\widetilde{\Sigma}^2$. We provide its proof for completeness in Appendix B.1.

Lemma 5.2.1. *Let $\phi\colon \mathbb{R} \to \mathbb{R}$ be convex and increasing, and let Z, Z' be given by (5.2.1). Then*

$$\mathbb{E}\,\phi(Z) \le \mathbb{E}\,\phi(Z').$$

Our main result regarding concentration of Z is the theorem below providing a Bennett-type bound.

Theorem 5.2.2. *Let Z be given by (5.2.1) and assume $\mathcal{X} \subset [-1,1]^n$. Then, for some absolute constants $C_1, C_2 > 0$,*

$$\forall\, t \ge 0 \qquad \mathbb{P}(Z \ge \mathbb{E}\,Z + t) \le 2\exp\Big(-\frac{t}{C_1}\log\Big(1 + \frac{t}{C_2\,\mathbb{E}\,\widetilde{\Sigma}^2}\Big)\Big),$$

where $\widetilde{\Sigma}^2 = \sup_{x\in\mathcal{X}} \sum_{k=1}^m x_{J_k}^2$. One can take $C_1 = 36$, $C_2 = 46$.

Remark 5.2.3. Denote

$$v = m \sup_{x\in\mathcal{X}} \mathrm{Var}(x_{J_1}) + 2\,\mathbb{E}\,Z'.$$

If $\mathcal{X} \subset \{x \in [-1,1]^n\colon \sum_i x_i = 0\}$, then Tolstikhin–Blanchard–Kloft [195, Theorem 2] proved that

$$\forall\, t \ge 0 \qquad \mathbb{P}(Z \ge \mathbb{E}\,Z' + t) \le \exp\Big(-t\log\big(1 + \frac{t}{v}\big) + t - v\log\big(1 + \frac{t}{v}\big)\Big). \tag{5.2.3}$$

Recall that by Hoeffding's argument (5.2.2), cf. Lemma 5.2.1, $\mathbb{E}\,Z \le \mathbb{E}\,Z'$ and in many situations the latter quantity can be significantly larger. Using symmetrization and Talagrand's contraction principle for Rademacher averages, cf., e.g., [139], we can estimate

$$\mathbb{E}\,\widetilde{\Sigma}^2 \le m \sup_{x\in\mathcal{X}} \mathrm{Var}(x_{J_1}) + 8\,\mathbb{E} \sup_{x\in\mathcal{X}} \sum_{k=1}^m \varepsilon_k x_{J_k},$$

where $\varepsilon_1, \ldots, \varepsilon_m$ are i.i.d. Rademacher variables independent of $J_1, \ldots, J_m$. Thus, in the case when the set $\mathcal{X}$ is symmetric with respect to the origin we obtain that

$$\mathbb{E}\,\widetilde{\Sigma}^2 \le m \sup_{x\in\mathcal{X}} \mathrm{Var}(x_{J_1}) + 16\,\mathbb{E}\,Z' \le 8v$$

and consequently our estimate of Theorem 5.2.2, in contrast to (5.2.3), provides a bound on deviations around the "proper" mean, while having no worse scaling behavior in the exponent (up to numerical constants).

It is worth noting that authors of [195] provide a bound $\mathbb{E}\, Z' \le \mathbb{E}\, Z + 2\frac{m^3}{n}$ which shows that one can replace $\mathbb{E}\, Z'$ with $\mathbb{E}\, Z$ under the probability estimate without losing much for small values of m. However, even in such a general case of non-symmetric set $\mathcal{X}$, it may also happen that $\mathbb{E}\, Z$ is significantly smaller than $\mathbb{E}\, Z'$, in which situation our estimate still improves upon (5.2.3).

To prove the Bennett-type inequality of Theorem 5.2.2, we need the following estimate due to Ledoux [137]. We provide the proof for completeness in Appendix B.2.

Lemma 5.2.4 ([137, Proof of Theorem 2.4])**.** *Let Z' be given by (5.2.1) and assume $\mathcal{X} \subset [0,1]^n$. Then*

$$\forall\, \lambda \ge 1/4 \qquad \log \mathbb{E}\, e^{\lambda Z'} \le \frac{1}{16} e^{8\lambda}\, \mathbb{E}\, Z'.$$

We also need the following proposition providing the Bernstein inequality for Z. We defer its proof to Section 5.4.

Proposition 5.2.5. *Let Z be given by (5.2.1) and assume $\mathcal{X} \subset [-1,1]^n$. Then*

$$\forall\, t \ge 0 \qquad \mathbb{P}(Z \ge \mathbb{E}\, Z + t) \le \exp\Big(-\min\Big(\frac{t}{32}, \frac{t^2}{128\, \mathbb{E}\, \Sigma^2}\Big)\Big),$$

where $\Sigma^2 = \sup_{x\in\mathcal{X}} \sum_k x_{I_k}^2$.

Proof of Theorem 5.2.2. If $32t < C_1 C_2\, \mathbb{E}\, \widetilde{\Sigma}^2$, then we apply Proposition 5.2.5 and estimate $\log(1+x) \le x$ to get that as long as $128 \le C_1 C_2$,

$$\begin{aligned}
\mathbb{P}(Z \ge \mathbb{E}\, Z + t) &\le \exp\Big(-\min\Big(\frac{t}{32}, \frac{t^2}{128\, \mathbb{E}\, \widetilde{\Sigma}^2}\Big)\Big), \\
&\le \exp\Big(-\min\Big(\frac{t}{32}, \frac{t^2}{C_1 C_2\, \mathbb{E}\, \widetilde{\Sigma}^2}\Big)\Big) \\
&= \exp\Big(-\frac{t^2}{C_1 C_2\, \mathbb{E}\, \widetilde{\Sigma}^2}\Big) \le \exp\Big(-\frac{t}{C_1} \log\Big(1 + \frac{t}{C_2\, \mathbb{E}\, \widetilde{\Sigma}^2}\Big)\Big)
\end{aligned}$$

and the result follows in this case.

If $32t \ge C_1 C_2\, \mathbb{E}\, \widetilde{\Sigma}^2$, then set

$$\rho^{-1} = \alpha \log\big(1 + \beta \frac{t}{\mathbb{E}\, \widetilde{\Sigma}^2}\big)$$

for some $\alpha, \beta > 0$ (to be fixed later) and denote

$$Z^{\downarrow} = \sup_{x\in\mathcal{X}} \sum_{k=1}^{m} x_{I_k} \mathbf{1}_{\{|x_{I_k}| \le \rho\}}$$

and

$$Z^{\uparrow} = \sup_{x\in\mathcal{X}} \sum_{k=1}^{m} |x_{I_k}| \mathbf{1}_{\{|x_{I_k}| > \rho\}}$$

so that $Z \le Z^{\downarrow} + Z^{\uparrow}$. We estimate the tail probabilities for $Z^{\downarrow}$ and $Z^{\uparrow}$.

Recall that by Lemma 5.2.1, $\mathbb{E}\,\Sigma^2 \le \mathbb{E}\,\widetilde{\Sigma}^2$. By the estimate $\log(1+x) \le x$, by the definition of ρ and as long as $\alpha\beta \le 1/4$,

$$\frac{t}{32\rho} \le \alpha\beta \cdot \frac{t^2}{32\,\mathbb{E}\,\widetilde{\Sigma}^2} \le \frac{t^2}{128\,\mathbb{E}\,\widetilde{\Sigma}^2} \le \frac{t^2}{128\,\mathbb{E}\,\Sigma^2},$$

whence, by Proposition 5.2.5 applied to $Z^{\downarrow}/\rho$,

$$\begin{aligned}\mathbb{P}(Z^{\downarrow} \ge \mathbb{E}\,Z^{\downarrow} + t) &\le \exp\Big(-\min\Big(\frac{t}{32\rho}, \frac{t^2}{128\,\mathbb{E}\,\Sigma^2}\Big)\Big)\\ &= \exp\Big(-\frac{t}{32\rho}\Big) = \exp\Big(-\frac{\alpha t}{32}\log\big(1+\beta\frac{t}{\mathbb{E}\,\widetilde{\Sigma}^2}\big)\Big).\end{aligned} \tag{5.2.4}$$

We turn to the tails of $Z^{\uparrow}$. Denote

$$Z'_\rho = \sup_{x\in\mathcal{X}} \sum_{k=1}^{m} |x_{J_k}| \mathbf{1}_{\{|x_{J_k}|>\rho\}}.$$

Lemma 5.2.1 applied with $\{\, (|x_i|\mathbf{1}_{\{|x_i|>\rho\}})_{i=1}^n \;:\; x\in\mathcal{X}\,\}$ in place of $\mathcal{X}$ together with Lemma 5.2.4 applied to Z'_ρ yield

$$\log \mathbb{E}\, e^{\lambda Z^{\uparrow}} \le \log \mathbb{E}\, e^{\lambda Z'_\rho} \le \frac{1}{16} e^{8\lambda}\,\mathbb{E}\,Z'_\rho \tag{5.2.5}$$

for all $\lambda \ge 1/4$. Choose

$$\lambda^* = \frac{1}{8}\log\Big(1+\beta\frac{t}{\mathbb{E}\,\widetilde{\Sigma}^2}\Big).$$

Since $32t \ge C_1C_2\,\mathbb{E}\,\widetilde{\Sigma}^2$ by assumption, then $\lambda^* \ge \frac{1}{8}\log(1+\frac{\beta C_1C_2}{32}) \ge \frac{1}{4}$, as long as $\beta C_1 C_2 \ge 32(e^2-1)$. Moreover, note that

$$\mathbb{E}\,Z'_\rho \le \rho^{-1}\,\mathbb{E}\,\widetilde{\Sigma}^2 \le \frac{32t\rho^{-1}}{C_1C_1}.$$

Consequently, by the Chernoff bound combined with (5.2.5),

$$\begin{aligned}\mathbb{P}(Z^{\uparrow} \ge t) &\le \exp\Big(-t\lambda^* + \frac{e^{8\lambda^*}}{16}\,\mathbb{E}\,Z'_\rho\Big)\\ &= \exp\Big(-\frac{t}{8}\log\Big(1+\beta\frac{t}{\mathbb{E}\,\widetilde{\Sigma}^2}\Big) + \frac{1}{16}\Big(\mathbb{E}\,Z'_\rho + \frac{\mathbb{E}\,Z'_\rho}{\mathbb{E}\,\widetilde{\Sigma}^2}t\beta\Big)\Big)\\ &\le \exp\Big(-\frac{t}{8}\log\Big(1+\beta\frac{t}{\mathbb{E}\,\widetilde{\Sigma}^2}\Big) + \frac{1}{16}\Big(\frac{32t\rho^{-1}}{C_1C_1} + t\beta\rho^{-1}\Big)\Big)\\ &= \exp\Big(-t\log\Big(1+\beta\frac{t}{\mathbb{E}\,\widetilde{\Sigma}^2}\Big)\cdot\Big(\frac{1}{8} - \frac{2\alpha}{C_1C_2} - \frac{\alpha\beta}{16}\Big)\Big).\end{aligned} \tag{5.2.6}$$

Using the estimate $\log(1+x) \le x$ we obtain that

$$|\mathbb{E}\,Z^{\downarrow} - \mathbb{E}\,Z| \le \mathbb{E}\,Z^{\uparrow} \le \frac{\mathbb{E}\,\widetilde{\Sigma}^2}{\rho} \le \alpha\beta t. \tag{5.2.7}$$

Thus, combining (5.2.4), (5.2.6) and (5.2.7) and as long as $\alpha\beta \le 1/4$ and $\beta C_1 C_2 \ge 32(e^2-1)$, we arrive at

$$\begin{aligned}\mathbb{P}(Z \ge \mathbb{E}\, Z + 2t + \alpha\beta t) &\le \mathbb{P}(Z^{\uparrow} + Z^{\downarrow} \ge \mathbb{E}\, Z + 2t + \alpha\beta t)\\ &\le \mathbb{P}(Z^{\uparrow} + Z^{\downarrow} \ge \mathbb{E}\, Z^{\downarrow} - |\,\mathbb{E}\, Z - \mathbb{E}\, Z^{\downarrow}| + 2t + \alpha\beta t)\\ &\le \mathbb{P}(Z^{\uparrow} + Z^{\downarrow} \ge \mathbb{E}\, Z^{\downarrow} + 2t)\\ &\le \mathbb{P}(Z^{\uparrow} \ge t) + \mathbb{P}(Z^{\downarrow} \ge \mathbb{E}\, Z^{\downarrow} + t)\\ &\le 2\exp\Big(-\min\Big(\frac{\alpha}{32}, \frac{1}{8} - \frac{\alpha\beta}{16} - \frac{2\alpha}{C_1C_2}\Big)\\ &\qquad\qquad \times t\log\Big(1 + \beta\frac{t}{\mathbb{E}\,\widetilde{\Sigma}^2}\Big)\Big).\end{aligned}$$

Substituting $t \leftarrow (2+\alpha\beta)^{-1}t$ and estimating $\frac{1}{2+\alpha\beta} \ge \frac{4}{9}$ yields

$$\begin{aligned}\mathbb{P}(Z \ge \mathbb{E}\, Z + t) &\le 2\exp\Big(-\frac{1}{2+\alpha\beta}\min\Big(\frac{\alpha}{32}, \frac{1}{8} - \frac{\alpha\beta}{16} - \frac{2\alpha}{C_1C_2}\Big)\\ &\qquad\qquad \times t\log\Big(1 + \frac{\beta}{2+\alpha\beta}\frac{t}{\mathbb{E}\,\widetilde{\Sigma}^2}\Big)\Big)\\ &\le 2\exp\Big(-\frac{4}{9}\min\Big(\frac{\alpha}{32}, \frac{1}{8} - \frac{\alpha\beta}{16} - \frac{2\alpha}{C_1C_2}\Big)\cdot t\log\Big(1 + \frac{4\beta}{9}\frac{t}{\mathbb{E}\,\widetilde{\Sigma}^2}\Big)\Big)\\ &\le 2\exp\Big(-\frac{4}{9}\min\Big(\frac{\alpha}{32}, \frac{1}{8} - \frac{\alpha\beta}{16} - \frac{2\alpha}{C_1C_2}\Big)\cdot t\log\Big(1 + \frac{t}{C_2\,\mathbb{E}\,\widetilde{\Sigma}^2}\Big)\Big)\end{aligned}$$

as long as $\alpha\beta \le 1/4$, $\beta C_1 C_2 \ge 32(e^2-1)$ and $4\beta C_2 \ge 9$. Setting $\alpha = 2$ and $\beta = \frac{1}{8}$ yields the result with $C_1 = 36$ and $C_2 = 46$. □

5.3 Concentration for a single Hoeffding statistic

In this section, we provide concentration bounds for single Hoeffding statistics, extending the results of Chatterjee [62], Bercu–Deylon–Rio [30] and Albert [11]. In the sequel, f denotes some Hoeffding statistics, i.e.,

$$f(\sigma) = \sum_{k=1}^{n} a_{k\sigma(k)}, \tag{5.3.1}$$

where $(a_{ij})_{i,j=1}^n \in \mathbb{R}^{n\times n}$ is some real matrix. The main result of this section is the following theorem. To the best of our knowledge, this is the first result that captures both the subgaussian and Poisson behaviors of Hoeffding statistics.

Theorem 5.3.1. *Let f be given by (5.3.1). If $a_{ij} \in [-1,1]$ for all i,j and $\sum_{ij} a_{ij} = 0$, then for some absolute constants $C_1, C_2 > 0$,*

$$\forall\, t \ge 0 \qquad \mathbb{P}(f \ge t) \le 2\exp\Big(-\frac{t}{C_1}\log\Big(1 + \frac{t}{C_2\,\mathbb{E}\,\Sigma^2}\Big)\Big),$$

where $\Sigma^2 = \sum_k a_{k\sigma_k}^2$ so that $\mathbb{E}\,\Sigma^2 = \frac{1}{n}\sum_{ij} a_{ij}^2$. One can take $C_1 = C_2 = 36$

Remark 5.3.2. As in Bercu–Deylon–Rio [30], note that setting

$$d_{ij} = a_{ij} - \frac{1}{n}\sum_{k=1}^{n}(a_{ik} + a_{kj}) + \frac{1}{n^2}\sum_{k,l=1}^{n} a_{kl}$$

yields $\mathrm{Var}(f) = \frac{1}{n-1}\sum_{ij} d_{ij}^2$ and $f - \mathbb{E} f = \sum_{k=1}^n d_{k\sigma(k)}$. Therefore, an application of Theorem 5.3.1 to $(f - \mathbb{E} f)/2$ in place of f (note that $\sum_{ij} d_{ij} = 0$, while $a_{ij} \in [-1,1]$ are arbitrary) provides that

$$\forall\, t \geq 0 \qquad \mathbb{P}(f \geq \mathbb{E} f + t) \leq 2\exp\Big(-\frac{t}{2C_1}\log\Big(1 + \frac{t}{2C_2\,\mathrm{Var}(f)}\Big)\Big). \tag{5.3.2}$$

The results of [62, 30, 11] provide bounds of the form $\exp(-ct^2/(\mathrm{Var}(f)+t))$ for some numerical constants c, which is comparable with (5.3.2) for small t and worse for t big enough. Moreover, as shown by Hoeffding in [116] (cf. also Bolthausen [43] for a Stein method based approach), as soon as

$$\lim_{n\to\infty} \frac{\max_{i,j\in[n]} d_{ij}}{\mathrm{Var}(S_n)} = 0,$$

then f verifies the CLT, i.e.,

$$\frac{f - \mathbb{E} f}{\sqrt{\mathrm{Var}(f)}} \overset{n\to\infty}{\longrightarrow} \mathcal{N}(0,1)$$

in law. Clearly, the bound from (5.3.2) becomes subgaussian for small values of t and whence matches the CLT behavior described above (up to numerical constants). Similarly, if one chooses $a_{ij} = \mathbf{1}_{\{i=j\}}$, then f becomes the number of fixed points of a random permutation σ. The exact tail distribution of f in such case is well known, cf. [87, Section IV.4], and is of order $\exp(-Ct\log t)$ for t big and some $C > 0$, which agrees with the bound (5.3.2). This shows that Theorem 5.3.1 is optimal up to the numerical constants.

To prove the Bennett inequality of Theorem 5.3.1, we first derive it for non-negative statistics in the theorem below.

Theorem 5.3.3. *Let f be given by (5.3.1). If $a_{ij} \in [0,1]$ for all i,j, then*

$$\forall\, t \geq 0 \qquad \mathbb{P}(f > \mathbb{E} f + t) \leq \exp\Big(-\frac{t}{4}\log\Big(1 + \frac{t}{4\,\mathbb{E} f}\Big)\Big).$$

Remark 5.3.4. Theorem 5.3.3 already improves (up to numerical constants in the exponent) upon a Bernstein-type bound

$$\forall\, t \geq 0 \qquad \mathbb{P}(f > \mathbb{E} f + t) \leq \exp\Big(-\frac{t^2}{4\,\mathbb{E} f + 2t}\Big)$$

obtained by Chatterjee [62, Proposition 1.1].

Proof of Theorem 5.3.3. Since $a_{ij} \in [0,1]$, then for any i,j,

$$\begin{aligned}\sum_{ij}(f_{ij} - f)_+ &= \sum_{ij}(a_{i\sigma_j} + a_{j\sigma_i} - a_{i\sigma_i} - a_{j\sigma_j})_+ \\ &\leq \sum_{ij}(a_{i\sigma_j} + a_{j\sigma_i}) = 2\sum_{ij} a_{ij} = 2n\,\mathbb{E} f.\end{aligned} \tag{5.3.3}$$

By the modified log-Sobolev inequality, using (5.3.3) and convexity of $x \mapsto e^{2x}$, we arrive at

$$\begin{aligned}\mathrm{Ent}(e^{\lambda f}) &\leq \frac{\lambda}{n}\,\mathbb{E}\, e^{\lambda f}\sum_{ij}(e^{\lambda(f_{ij}-f)_+} - 1)(f_{ij} - f)_+ \\ &\leq \frac{\lambda}{n}(e^{2\lambda} - 1)\,\mathbb{E}\, e^{\lambda f}\sum_{ij}(f_{ij} - f)_+ \\ &\leq 2\lambda(e^{2\lambda} - 1)\,\mathbb{E} f\,\mathbb{E}\, e^{\lambda f} \leq 4\lambda^2 e^{2\lambda}\,\mathbb{E} f\,\mathbb{E}\, e^{\lambda f}\end{aligned}$$

for all $\lambda \geq 0$. Hence, using Proposition B.3.1 with $a = 4\,\mathbb{E}\, f$, $b = 2$ gives the conclusion. □

Finally, to prove Theorem 5.3.1, we need the following proposition. We defer its proof to Section 5.4.

Proposition 5.3.5. *Let f be given by (5.3.1). If $a_{ij} \in [-1, 1]$ for all i, j, then*

$$\forall\, t \geq 0 \qquad \mathbb{P}(f \geq \mathbb{E}\, f + t) \leq \exp\Big(-\min\Big(\frac{t}{32}, \frac{t^2}{128\,\mathbb{E}\,\Sigma^2}\Big)\Big),$$

where $\Sigma^2 = \sum_k a^2_{k\sigma_k}$ so that $\mathbb{E}\,\Sigma^2 = \frac{1}{n}\sum_{ij} a^2_{ij}$.

Proof of Theorem 5.3.1. For a fixed $t > 0$, set

$$\rho^{-1} = 2\log\big(1 + \frac{t}{16\,\mathbb{E}\,\Sigma^2}\big)$$

and denote

$$f^{\downarrow}(\sigma) = \sum_i a_{i\sigma_i} \mathbf{1}_{\{|a_{i\sigma_i}| \leq \rho\}}$$

and

$$f^{\uparrow}(\sigma) = \sum_i |a_{i\sigma_i}| \mathbf{1}_{\{|a_{i\sigma_i}| > \rho\}}$$

so that $f \leq f^{\downarrow} + f^{\uparrow}$. We estimate the tail probabilities for $f^{\downarrow}$ and $f^{\uparrow}$.

By the estimate $\log(1 + x) \leq x$ and by the definition of ρ,

$$\frac{t}{32\rho} \leq \frac{t^2}{256\,\mathbb{E}\,\Sigma^2} \leq \frac{t^2}{128\,\mathbb{E}\,\Sigma^2},$$

whence by Proposition 5.3.5 applied to $f^{\downarrow}/\rho$,

$$\begin{aligned} \mathbb{P}(f^{\downarrow} \geq \mathbb{E}\, f^{\downarrow} + t) &\leq \exp\Big(-\min\Big(\frac{t}{32\rho}, \frac{t^2}{128\,\mathbb{E}\,\Sigma^2}\Big)\Big) \\ &= \exp\Big(-\frac{t}{32\rho}\Big) = \exp\Big(-\frac{t}{16}\log\big(1 + \frac{t}{16\,\mathbb{E}\,\Sigma^2}\big)\Big). \end{aligned} \tag{5.3.4}$$

By the definitions of $f^{\uparrow}, \rho$ and estimate $\log(1 + x) \leq 2\log(1 + \sqrt{x}) \leq 2\sqrt{x}$,

$$\mathbb{E}\, f^{\uparrow} \leq \frac{\mathbb{E}\,\Sigma^2}{\rho} = 2(\mathbb{E}\,\Sigma^2)\log\Big(1 + \frac{t}{16\,\mathbb{E}\,\Sigma^2}\Big) \leq \sqrt{t\,\mathbb{E}\,\Sigma^2},$$

whence by Theorem 5.3.3 applied to $f^{\uparrow}$,

$$\begin{aligned} \mathbb{P}(f^{\uparrow} \geq \mathbb{E}\, f^{\uparrow} + t) &\leq \exp\Big(-\frac{t}{4}\log\Big(1 + \frac{t}{4\,\mathbb{E}\, f^{\uparrow}}\Big)\Big) \\ &\leq \exp\Big(-\frac{t}{4}\log\Big(1 + \frac{1}{4}\sqrt{\frac{t}{\mathbb{E}\,\Sigma^2}}\Big)\Big) \\ &\leq \exp\Big(-\frac{t}{8}\log\Big(1 + \frac{t}{16\,\mathbb{E}\,\Sigma^2}\Big)\Big), \end{aligned} \tag{5.3.5}$$

where in the last step we have used again the estimate $2\log(1+\sqrt{x}) \geq \log(1+x)$. Using the assumption $\mathbb{E}\, f = 0$, triangle inequality and estimating $\log(1+x) \leq x$, we obtain

$$|\mathbb{E}\, f^{\downarrow}| = |\mathbb{E}\, f^{\downarrow} - \mathbb{E}\, f| \leq \mathbb{E}\, f^{\uparrow} \leq \frac{\mathbb{E}\,\Sigma^2}{\rho} \leq \frac{1}{8}t. \tag{5.3.6}$$

By combining (5.3.4), (5.3.5) and (5.3.6) we arrive at

$$\begin{aligned}\mathbb{P}(f \geq 9t/4) &\leq \mathbb{P}(f^{\downarrow} \geq 9t/8) + \mathbb{P}(f^{\uparrow} \geq 9t/8)\\ &\leq \mathbb{P}(f^{\downarrow} \geq \mathbb{E}\, f^{\downarrow} + t) + \mathbb{P}(f^{\uparrow} \geq \mathbb{E}\, f^{\uparrow} + t)\\ &\leq 2\exp\Big(-\frac{t}{16}\log\Big(1+\frac{t}{16\,\mathbb{E}\,\Sigma^2}\Big)\Big).\end{aligned}$$

Substituting $t \leftarrow 4t/9$ yields the result. □

5.4 Proof of Propositions 5.2.5 and 5.3.5

Both propositions are special cases of a more general result for suprema of Hoeffding statistics which we provide below. Let $R \subset \mathbb{R}^{n\times n}$ be a set of real matrices. Denote

$$S = \sup_{r\in R} \sum_{k=1}^{n} r_{k\sigma_k}. \tag{5.4.1}$$

The main result of this section is the following estimate. It is worth mentioning that it can be also derived (up to numerical constants) from Proposition 2.4.9.

Proposition 5.4.1. *Let S be given by (5.4.1) and assume $R \subset [-1,1]^{n\times n}$. Then*

$$\forall\, t \geq 0 \qquad \mathbb{P}(S \geq \mathbb{E}\, S + t) \leq \exp\Big(-\min\Big(\frac{t}{32}, \frac{t^2}{128\,\mathbb{E}\,\Sigma_R^2}\Big)\Big),$$

where $\Sigma_R^2 = \sup_{r\in R}\sum_k r_{k\sigma_k}^2$.

Propositions 5.2.5 and 5.3.5 are special cases of Proposition 5.4.1 as illustrated below.

Proof of Proposition 5.2.5. Apply Proposition 5.4.1 with $R = \{\, a^x : x \in \mathcal{X}\,\}$ (recall the definition of the matrix a^x introduced at the beginning of Section 5.2). □

Proof of Propositoin 5.3.5. Apply Proposition 5.4.1 with $R = \{a\}$. □

To prove Proposition 5.4.1, let us first state the modified log-Sobolev inequality (5.1.1) for the Laplace transform of S. For any $i, j \in [n]$, denote

$$S_{ij} = \sup_{r\in R} \sum_{k=1}^{n} r_{k\sigma_{ij}(k)}.$$

Then, the modified log-Sobolev inequality (5.1.1) implies that

$$\mathrm{Ent}(e^{\lambda S}) \leq \frac{\lambda}{n}\,\mathbb{E}\Big[e^{\lambda S}\sum_{ij}(1-e^{-\lambda(S-S_{ij})})_+(S-S_{ij})_+\Big],$$

which after estimating $1 - e^{-x} \leq x$ can be further specialized to

$$\mathrm{Ent}(e^{\lambda S}) \leq \frac{\lambda}{n}\,\mathbb{E}\Big[e^{\lambda S}\sum_{ij}(S-S_{ij})_+^2\Big]. \tag{5.4.2}$$

We need also the following auxiliary fact.

Lemma 5.4.2. *Let S be given by (5.4.1) and assume $R \subset [0,1]^{n\times n}$. Then*

$$\forall\, \lambda \in [0, 1/4] \qquad \log \mathbb{E}\, e^{\lambda S} \leq 2\lambda\, \mathbb{E}\, S.$$

Proof. Assume w.l.o.g. that R is finite. Let $\hat{r}$ be a random matrix taking values in R such that $S = \sum_{k=1}^{n} \hat{r}_{k\sigma_k}$. We have

$$\begin{aligned}\sum_{ij} (S - S_{ij})_+^2 &\leq \sum_{ij} (\hat{r}_{i\sigma_i} + \hat{r}_{j\sigma_j} - \hat{r}_{i\sigma_j} - \hat{r}_{j\sigma_i})_+^2 \\ &\leq \sum_{ij} (\hat{r}_{i\sigma_i} + \hat{r}_{j\sigma_j})^2 \leq 2n \sum_{i} (\hat{r}_{i\sigma_i})^2 \leq 2nS,\end{aligned} \tag{5.4.3}$$

where in the last inequality we have used that $R \in [0,1]^{n\times n}$.

By the modified log-Sobolev inequality (5.4.2) combined with (5.4.3), we arrive at

$$\operatorname{Ent}(e^{\lambda S}) \leq \frac{\lambda^2}{n}\, \mathbb{E}\Big[e^{\lambda S} \sum_{ij} (S - S_{ij})_+^2 \Big] \leq 2\lambda^2\, \mathbb{E}[e^{\lambda S} S]$$

for all $\lambda \geq 0$. Applying Proposition B.3.2 with $a = 2$, $b = 0$ results in $(1 - 2\lambda) \log \mathbb{E}\, e^{\lambda S} \leq \lambda\, \mathbb{E}\, S$, for all $\lambda \geq 0$, which yields the conclusion. $\square$

We are in position to prove Proposition 5.4.1.

Proof of Proposition 5.4.1. Let $\hat{r}$ be a random matrix taking values in R such that $S = \sum_{k=1}^{n} \hat{r}_{k\sigma_k}$. By the triangle inequality in ℓ^2,

$$\begin{aligned}\sum_{ij} (S - S_{ij})_+^2 &\leq \sum_{ij} (\hat{r}_{i\sigma_i} + \hat{r}_{j\sigma_j} - \hat{r}_{i\sigma_j} - \hat{r}_{j\sigma_i})_+^2 \\ &\leq 8 \sum_{ij} \hat{r}_{i\sigma_i}^2 + 8 \sum_{ij} \hat{r}_{i\sigma_j}^2 \leq 8n\Sigma_R^2 + 8 \sum_{ij} \hat{r}_{i\sigma_j}^2 .\end{aligned} \tag{5.4.4}$$

Note that

$$\sum_{ij} \hat{r}_{i\sigma_j}^2 = \sum_{ij} \hat{r}_{ij}^2 = n\, \mathbb{E} \sum_{i} \hat{r}_{i\sigma_i}^2 \leq n\, \mathbb{E} \sup_{r\in R} \sum_{i} r_{i\sigma_i}^2 = n\, \mathbb{E}\, \Sigma_R^2,$$

whence (5.4.4) can be further specialized to

$$\sum_{ij} (S - S_{ij})_+^2 \leq 8n(\Sigma_R^2 + \mathbb{E}\, \Sigma_R^2). \tag{5.4.5}$$

By the modified log-Sobolev inequality (5.4.2) combined with (5.4.5), we arrive at

$$\operatorname{Ent}(e^{\lambda S}) \leq \frac{\lambda^2}{n}\, \mathbb{E}\Big[e^{\lambda S} \sum_{ij} (S - S_{ij})_+^2 \Big] \leq 8\lambda^2 \big((\mathbb{E}\, e^{\lambda S})(\mathbb{E}\, \Sigma_R^2) + \mathbb{E}[e^{\lambda S} \Sigma_R^2] \big). \tag{5.4.6}$$

Recall the variational formula for entropy $\operatorname{Ent}(h) = \sup\big\{\, \mathbb{E}\, hg \colon\ \mathbb{E}\, e^g \leq 1 \,\big\}$, from which it follows that for any h, g

$$\mathbb{E}\, hg \leq \operatorname{Ent}(h) + (\mathbb{E}\, h) \log(\mathbb{E}\, e^g). \tag{5.4.7}$$

Applying first (5.4.7) with $h = e^{\lambda S}$, $g = \Sigma_R^2/4$ and then Lemma 5.4.2 yields

$$\mathbb{E}\big[e^{\lambda S}\Sigma_R^2\big] \le 4\,\mathrm{Ent}(e^{\lambda S}) + 4\big(\mathbb{E}\,e^{\lambda S}\big)\big(\log \mathbb{E}\,e^{\Sigma_R^2/4}\big) \le 4\,\mathrm{Ent}(e^{\lambda S}) + 2\big(\mathbb{E}\,e^{\lambda S}\big)\big(\mathbb{E}\,\Sigma_R^2\big),$$

which combined with (5.4.6) results in

$$(1 - 32\lambda^2)\,\mathrm{Ent}(e^{\lambda S}) \le 24\lambda^2\big(\mathbb{E}\,\Sigma_R{}^2\big)\big(\mathbb{E}\,e^{\lambda S}\big)$$

for all $\lambda \ge 0$, so that

$$\mathrm{Ent}(e^{\lambda S}) \le \frac{192}{7}\lambda^2\big(\mathbb{E}\,\Sigma_R{}^2\big)\big(\mathbb{E}\,e^{\lambda S}\big) \le 32\lambda^2\big(\mathbb{E}\,\Sigma_R{}^2\big)\big(\mathbb{E}\,e^{\lambda S}\big)$$

for all $\lambda \in [0, 1/16]$. We conclude by applying Proposition B.3.3 with $\varepsilon = \frac{1}{16}$ and $b = 32\,\mathbb{E}\,\Sigma_R^2$. $\square$

Bibliography

[1] Radosław Adamczak, *Moment inequalities for U-statistics*, Ann. Probab. **34** (2006), no. 6, 2288–2314. MR 2294982 44

[2] ———, *A note on the Hanson-Wright inequality for random vectors with dependencies*, Electron. Commun. Probab. **20** (2015), no. 72, 13. MR 3407216 2

[3] Radosław Adamczak, Witold Bednorz, and Paweł Wolff, *Moment estimates implied by modified log-Sobolev inequalities*, ESAIM Probab. Stat. **21** (2017), 467–494. MR 3743923 9, 21, 52, 55

[4] Radosław Adamczak, Michał Kotowski, Bartłomiej Polaczyk, and Michał Strzelecki, *A note on concentration for polynomials in the Ising model*, Electron. J. Probab. **24** (2019), Paper No. 42, 22. MR 3949267 9, 10, 32, 52, 56, 57, 88, 103

[5] Radosław Adamczak and Bartłomiej Polaczyk, *Concentration inequalities for some negatively dependent binary random variables*, arXiv preprint arXiv:2108.12636 (2021). 6, 7

[6] Radosław Adamczak, Bartłomiej Polaczyk, and Michał Strzelecki, *Modified log-Sobolev inequalities, Beckner inequalities and moment estimates*, J. Funct. Anal. **282** (2022), no. 7, Paper No. 109349, 76. MR 4372142 4, 7, 60, 79, 88, 90, 103

[7] Radosław Adamczak and Michał Strzelecki, *On the convex Poincaré inequality and weak transportation inequalities*, Bernoulli **25** (2019), no. 1, 341–374. MR 3892322 2, 3, 51, 86

[8] Radosław Adamczak and Paweł Wolff, *Concentration inequalities for non-Lipschitz functions with bounded derivatives of higher order*, Probab. Theory Related Fields **162** (2015), no. 3-4, 531–586. MR 3383337 9, 21, 25, 52, 55, 58, 81, 88, 103

[9] S. Aida and D. Stroock, *Moment estimates derived from Poincaré and logarithmic Sobolev inequalities*, Math. Res. Lett. **1** (1994), no. 1, 75–86. MR 1258492 3, 4, 9, 21, 60

[10] Shigeki Aida, Takao Masuda, and Ichirō Shigekawa, *Logarithmic Sobolev inequalities and exponential integrability*, J. Funct. Anal. **126** (1994), no. 1, 83–101. MR 1305064 101

[11] Mélisande Albert, *Concentration inequalities for randomly permuted sums*, High dimensional probability VIII—the Oaxaca volume, Progr. Probab., vol. 74, Birkhäuser/Springer, Cham, [2019] ©2019, pp. 341–383. MR 4181373 7, 35, 36, 111, 116, 117

[12] Nima Anari, Vishesh Jain, Frederic Koehler, Huy Tuan Pham, and Thuy-Duong Vuong, *Entropic independence in high-dimensional expanders: Modified log-sobolev inequalities for fractionally log-concave polynomials and the ising model*, arXiv preprint arXiv:2106.04105 (2021). 10, 32, 42, 80

[13] Nima Anari, Kuikui Liu, Shayan Oveis Gharan, and Cynthia Vinzant, *Log-concave polynomials II: High-dimensional walks and an FPRAS for counting bases of a matroid*, STOC'19—Proceedings of the 51st Annual ACM SIGACT Symposium on Theory of Computing, ACM, New York, 2019, pp. 1–12. MR 4003314 42, 79, 80, 93

[14] Cécile Ané, Sébastien Blachère, Djalil Chafaï, Pierre Fougères, Ivan Gentil, Florent Malrieu, Cyril Roberto, and Grégory Scheffer, *Sur les inégalités de Sobolev logarithmiques*, Panoramas et Synthèses [Panoramas and Syntheses], vol. 10, Société Mathématique de France, Paris, 2000, With a preface by Dominique Bakry and Michel Ledoux. MR 1845806 30, 91

[15] Cécile Ané and Michel Ledoux, *On logarithmic Sobolev inequalities for continuous time random walks on graphs*, Probab. Theory Related Fields **116** (2000), no. 4, 573–602. MR 1757600 10, 44

[16] Richard Aoun, Marwa Banna, and Pierre Youssef, *Matrix Poincaré inequalities and concentration*, Adv. Math. **371** (2020), 107251, 33. MR 4108222 6, 79, 80, 81, 82, 84, 85, 99

[17] Kazuoki Azuma, *Weighted sums of certain dependent random variables*, Tohoku Math. J. (2) **19** (1967), 357–367. MR 221571 81

[18] Sascha Bachmann, *Concentration for Poisson functionals: component counts in random geometric graphs*, Stochastic Process. Appl. **126** (2016), no. 5, 1306–1330. MR 3473096 10, 44, 48, 51, 52

[19] Sascha Bachmann and Giovanni Peccati, *Concentration bounds for geometric Poisson functionals: logarithmic Sobolev inequalities revisited*, Electron. J. Probab. **21** (2016), Paper No. 6, 44. MR 3485348 10, 44, 47, 48, 51

[20] Sascha Bachmann and Matthias Reitzner, *Concentration for Poisson U-statistics: subgraph counts in random geometric graphs*, Stochastic Process. Appl. **128** (2018), no. 10, 3327–3352. MR 3849811 10, 44, 48, 51

[21] D. Bakry, I. Gentil, and M. Ledoux, *Analysis and geometry of Markov diffusion operators*, Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences], vol. 348, Springer, Cham, 2014. MR 3155209 11, 20, 21, 60, 142, 143, 144

[22] Dominique Bakry, *L'hypercontractivité et son utilisation en théorie des semigroupes*, Lectures on probability theory (Saint-Flour, 1992), Lecture Notes in Math., vol. 1581, Springer, Berlin, 1994, pp. 1–114. MR 1307413 4, 59

[23] Dominique Bakry, Ivan Gentil, and Grégory Scheffer, *Sharp Beckner-type inequalities for Cauchy and spherical distributions*, Studia Math. **251** (2020), no. 3, 219–245. MR 4048609 26

[24] Dominique Bakry, Michel Ledoux, and Laurent Saloff-Coste, *Markov semigroups at Saint-Flour*, Probability at Saint-Flour, Springer, Heidelberg, 2012. MR 3075390 88

[25] F. Barthe and C. Roberto, *Sobolev inequalities for probability measures on the real line*, vol. 159, 2003, Dedicated to Professor Aleksander Pełczyński on the occasion of his 70th birthday (Polish), pp. 481–497. MR 2052235 57

[26] ______, *Modified logarithmic Sobolev inequalities on* $\mathbb{R}$, Potential Anal. **29** (2008), no. 2, 167–193. MR 2430612 60, 61, 67, 70

[27] Franck Barthe and MichałStrzelecki, *Functional inequalities for two-level concentration*, Potential Anal. **56** (2022), no. 4, 669–696. MR 4396831 25

[28] Roland Bauerschmidt and Thierry Bodineau, *A very simple proof of the LSI for high temperature spin systems*, J. Funct. Anal. **276** (2019), no. 8, 2582–2588. MR 3926125 20, 21, 32

[29] W. Beckner, *A generalized Poincaré inequality for Gaussian measures*, Proc. Amer. Math. Soc. **105** (1989), no. 2, 397–400. MR 954373 3, 8, 13, 14, 15

[30] Bernard Bercu, Bernard Delyon, and Emmanuel Rio, *Concentration inequalities for sums and martingales*, SpringerBriefs in Mathematics, Springer, Cham, 2015. MR 3363542 7, 35, 36, 111, 116, 117

[31] Patrice Bertail, Emilie Chautru, and Stephan Clémençon, *Empirical processes in survey sampling with (conditional) Poisson designs*, Scand. J. Stat. **44** (2017), no. 1, 97–111. MR 3619696 83

[32] Patrice Bertail and Stephan Clémençon, *Bernstein-type exponential inequalities in survey sampling: conditional Poisson sampling schemes*, Bernoulli **25** (2019), no. 4B, 3527–3554. MR 4010964 79, 81, 84

[33] Rajendra Bhatia, *Matrix analysis*, Graduate Texts in Mathematics, vol. 169, Springer-Verlag, New York, 1997. MR 1477662 96, 99

[34] Antonio Blanca, Pietro Caputo, Zongchen Chen, Daniel Parisi, Daniel Štefankovič, and Eric Vigoda, *On mixing of Markov chains: coupling, spectral independence, and entropy factorization*, Proceedings of the 2022 Annual ACM-SIAM Symposium on Discrete Algorithms (SODA), [Society for Industrial and Applied Mathematics (SIAM)], Philadelphia, PA, 2022, pp. 3670–3692. MR 4415182 10, 30

[35] Adrien Blanchet, Matteo Bonforte, Jean Dolbeault, Gabriele Grillo, and Juan-Luis Vázquez, *Hardy-Poincaré inequalities and applications to nonlinear diffusions*, C. R. Math. Acad. Sci. Paris **344** (2007), no. 7, 431–436. MR 2320246 26

[36] S. G. Bobkov, *The growth of* L_p*-norms in presence of logarithmic Sobolev inequalities*, Vestnik Syktyvkar Univ. 11.2 (2010), 92–111. 9

[37] S. G. Bobkov and F. Götze, *Exponential integrability and transportation cost related to logarithmic Sobolev inequalities*, J. Funct. Anal. **163** (1999), no. 1, 1–28. MR 1682772 60, 61, 101

[38] S. G. Bobkov and M. Ledoux, *On modified logarithmic Sobolev inequalities for Bernoulli and Poisson measures*, J. Funct. Anal. **156** (1998), no. 2, 347–365. MR 1636948 4, 15

[39] Sergey G. Bobkov, Gennadiy P. Chistyakov, and Friedrich Götze, *Second-order concentration on the sphere*, Commun. Contemp. Math. **19** (2017), no. 5, 1650058, 20. MR 3670792 9, 52

[40] Sergey G. Bobkov, Friedrich Götze, and Holger Sambale, *Higher order concentration of measure*, Commun. Contemp. Math. **21** (2019), no. 3, 1850043, 36. MR 3947067 2

[41] Sergey G. Bobkov and Michel Ledoux, *Weighted Poincaré-type inequalities for Cauchy and other convex measures*, Ann. Probab. **37** (2009), no. 2, 403–427. MR 2510011 21, 26

[42] Sergey G. Bobkov and Prasad Tetali, *Modified logarithmic Sobolev inequalities in discrete settings*, J. Theoret. Probab. **19** (2006), no. 2, 289–336. MR 2283379 3, 4, 6, 9, 14, 15, 34, 40, 41, 59, 79, 80, 82, 88, 89, 90, 100, 112, 138

[43] E. Bolthausen, *An estimate of the remainder in a combinatorial central limit theorem*, Z. Wahrsch. Verw. Gebiete **66** (1984), no. 3, 379–386. MR 751577 117

[44] M. Bonforte, J. Dolbeault, G. Grillo, and J. L. Vázquez, *Sharp rates of decay of solutions to the nonlinear fast diffusion equation via functional inequalities*, Proc. Natl. Acad. Sci. USA **107** (2010), no. 38, 16459–16464. MR 2726546 26

[45] Julius Borcea, Petter Brändén, and Thomas M. Liggett, *Negative dependence and the geometry of polynomials*, J. Amer. Math. Soc. **22** (2009), no. 2, 521–567. MR 2476782 79

[46] C. Borell, *The Brunn-Minkowski inequality in Gauss space*, Invent. Math. **30** (1975), no. 2, 207–216. MR 399402 2

[47] Stéphane Boucheron, Olivier Bousquet, Gábor Lugosi, and Pascal Massart, *Moment inequalities for functions of independent random variables*, Ann. Probab. **33** (2005), no. 2, 514–560. MR 2123200 9, 10, 27, 28, 35, 38, 49, 81

[48] Stéphane Boucheron, Gábor Lugosi, and Pacal Massart, *On concentration of self-bounding functions*, Electron. J. Probab. **14** (2009), no. 64, 1884–1899. MR 2540852 100

[49] Stéphane Boucheron, Gábor Lugosi, and Pascal Massart, *A sharp concentration inequality with applications*, Random Structures Algorithms **16** (2000), no. 3, 277–292. MR 1749290 2, 4, 27

[50] ______, *Concentration inequalities using the entropy method*, Ann. Probab. **31** (2003), no. 3, 1583–1614. MR 1989444 9, 27

[51] ______, *Concentration inequalities. A nonasymptotic theory of independence*, Oxford University Press, Oxford, 2013. MR 3185193 2, 7, 8, 16, 18, 19, 28, 81, 86, 89, 100, 102, 111, 147

[52] Olivier Bousquet, *A Bennett concentration inequality and its application to suprema of empirical processes*, C. R. Math. Acad. Sci. Paris **334** (2002), no. 6, 495–500. MR 1890640 37

[53] Petter Brändén and June Huh, *Lorentzian polynomials*, Ann. of Math. (2) **192** (2020), no. 3, 821–891. MR 4172622 42, 92

[54] Pietro Caputo, Paolo Dai Pra, and Gustavo Posta, *Convex entropy decay via the Bochner-Bakry-Emery approach*, Ann. Inst. Henri Poincaré Probab. Stat. **45** (2009), no. 3, 734–753. MR 2548501 43, 57

[55] Pietro Caputo and Gustavo Posta, *Entropy dissipation estimates in a zero-range dynamics*, Probab. Theory Related Fields **139** (2007), no. 1-2, 65–87. MR 2322692 43

[56] Patrick Cattiaux, Arnaud Guillin, and Li-Ming Wu, *Some remarks on weighted logarithmic Sobolev inequality*, Indiana Univ. Math. J. **60** (2011), no. 6, 1885–1904. MR 3008255 26

[57] Djalil Chafaï, *Entropies, convexity, and functional inequalities: on Φ-entropies and Φ-Sobolev inequalities*, J. Math. Kyoto Univ. **44** (2004), no. 2, 325–363. MR 2081075 10, 14, 44

[58] Djalil Chafaï and Aldéric Joulin, *Intertwining and commutation relations for birth-death processes*, Bernoulli **19** (2013), no. 5A, 1855–1879. MR 3129037 57

[59] S. Chatterjee, *Concentration inequalities with exchangeable pairs*, ProQuest LLC, Ann Arbor, MI, 2005, Thesis (Ph.D.)–Stanford University. MR 2707160 3

[60] ———, *The missing log in large deviations for triangle counts*, Random Structures Algorithms **40** (2012), no. 4, 437–451. MR 2925306 2

[61] S. Chatterjee and P. S. Dey, *Applications of Stein's method for concentration inequalities*, Ann. Probab. **38** (2010), no. 6, 2443–2485. MR 2683635 3

[62] Sourav Chatterjee, *Stein's method for concentration inequalities*, Probab. Theory Related Fields **138** (2007), no. 1-2, 305–321. MR 2288072 7, 10, 34, 35, 116, 117

[63] ———, *An introduction to large deviations for random graphs*, Bull. Amer. Math. Soc. (N.S.) **53** (2016), no. 4, 617–642. MR 3544262 31

[64] Sourav Chatterjee and Amir Dembo, *Nonlinear large deviations*, Adv. Math. **299** (2016), 396–450. MR 3519474 58

[65] Sourav Chatterjee and Persi Diaconis, *Estimating and understanding exponential random graph models*, Ann. Statist. **41** (2013), no. 5, 2428–2461. MR 3127871 31

[66] J.-R. Chazottes, P. Collet, and F. Redig, *On concentration inequalities and their applications for Gibbs measures in lattice systems*, J. Stat. Phys. **169** (2017), no. 3, 504–546. MR 3711609 32

[67] Mu-Fa Chen, *From Markov chains to non-equilibrium particle systems*, second ed., World Scientific Publishing Co., Inc., River Edge, NJ, 2004. MR 2091955 60, 63

[68] Sean X. Chen, *General properties and estimation of conditional Bernoulli models*, J. Multivariate Anal. **74** (2000), no. 1, 69–87. MR 1790614 79, 82

[69] Sean X. Chen and Jun S. Liu, *Statistical applications of the Poisson-binomial and conditional Bernoulli distributions*, Statist. Sinica **7** (1997), no. 4, 875–892. MR 1488647 79

[70] Xiang-Hui Chen, Arthur P. Dempster, and Jun S. Liu, *Weighted finite population sampling to maximize entropy*, Biometrika **81** (1994), no. 3, 457–469. MR 1311090 79

[71] Zongchen Chen, Kuikui Liu, and Eric Vigoda, *Optimal mixing of glauber dynamics: Entropy factorization via high-dimensional expansion*, Proceedings of the 53rd Annual ACM SIGACT Symposium on Theory of Computing (New York, NY, USA), STOC 2021, Association for Computing Machinery, 2021, pp. 1537–1550. 10, 30, 42

[72] Giovanni Conforti, *A probabilistic approach to convex (ϕ)-entropy decay for Markov chains*, Ann. Appl. Probab. **32** (2022), no. 2, 932–973. MR 4414692 9, 33, 43

[73] Richard W Conway and William L Maxwell, *A queuing model with state dependent service rates*, Journal of Industrial Engineering **12** (1962), no. 2, 132–136. 76

[74] Mary Cryan, Heng Guo, and Giorgos Mousa, *Modified log-Sobolev inequalities for strongly log-concave distributions*, Ann. Probab. **49** (2021), no. 1, 506–525. MR 4203344 10, 40, 42, 80, 83, 89, 92, 93

[75] Yuval Dagan, Constantinos Daskalakis, Nishanth Dikkala, and Anthimos Vardis Kandiros, *Estimating ising models from one sample*, CoRR **abs/2004.09370** (2020). 9

[76] Yuval Dagan, Constantinos Daskalakis, Nishanth Dikkala, and Anthimos Vardis Kandiros, *Learning Ising models from one or multiple samples*, Proceedings of the 53rd Annual ACM SIGACT Symposium on Theory of Computing (New York, NY, USA), STOC 2021, Association for Computing Machinery, 2021, pp. 161–168. 88

[77] P. Dai Pra, A. M. Paganoni, and G. Posta, *Entropy inequalities for unbounded spin systems*, Ann. Probab. **30** (2002), no. 4, 1959–1976. MR 1944012 15, 57, 62

[78] C. Daskalakis, N. Dikkala, and G. Kamath, *Concentration of multilinear functions of the ising model with applications to network data*, Advances in Neural Information Processing Systems 30: Annual Conference on Neural Information Processing Systems 2017, 4-9 December 2017, Long Beach, CA, USA, 2017, pp. 12–22. 32

[79] Constantinos Daskalakis, Nishanth Dikkala, and Gautam Kamath, *Testing Ising models*, Proceedings of the Twenty-Ninth Annual ACM-SIAM Symposium on Discrete Algorithms, SIAM, Philadelphia, PA, 2018, pp. 1989–2007. MR 3775918 32

[80] Víctor H. de la Peña and Evarist Giné, *Decoupling*, Probability and its Applications (New York), Springer-Verlag, New York, 1999, From dependence to independence, Randomly stopped processes. U-statistics and processes. Martingales and beyond. MR 1666908 137

[81] P. Diaconis and L. Saloff-Coste, *Logarithmic Sobolev inequalities for finite Markov chains*, Ann. Appl. Probab. **6** (1996), no. 3, 695–750. MR 1410112 15, 16, 59

[82] Persi Diaconis and Mehrdad Shahshahani, *Generating a random permutation with random transpositions*, Z. Wahrsch. Verw. Gebiete **57** (1981), no. 2, 159–179. MR 626813 34

[83] Christian Döbler and Giovanni Peccati, *Fourth moment theorems on the Poisson space: analytic statements via product formulae*, Electron. Commun. Probab. **23** (2018), Paper No. 91, 12. MR 3896829 47

[84] Devdatt P. Dubhashi and Alessandro Panconesi, *Concentration of measure for the analysis of randomized algorithms*, Cambridge University Press, Cambridge, 2009. MR 2547432 2, 3, 94

[85] Ronen Eldan, Frederic Koehler, and Ofer Zeitouni, *A spectral condition for spectral gap: fast mixing in high-temperature ising models*, Probability Theory and Related Fields (2021), 1–17. 21, 32

[86] Matthias Erbar, Christopher Henderson, Georg Menz, and Prasad Tetali, *Ricci curvature bounds for weakly interacting Markov chains*, Electron. J. Probab. **22** (2017), Paper No. 40, 23. MR 3646066 33

[87] Willliam Feller, *An introduction to probability theory and its applications, vol 2*, John Wiley & Sons, 2008. 117

[88] Yuval Filmus, Ryan O'Donnell, and Xinyu Wu, *A log-Sobolev inequality for the multislice, with applications*, 10th Innovations in Theoretical Computer Science Conference, ITCS 2019, January 10-12, 2019, San Diego, California, USA (Avrim Blum, ed.), LIPIcs, vol. 124, Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2019, pp. 34:1–34:12. 40

[89] Masatoshi Fukushima, Yoichi Oshima, and Masayoshi Takeda, *Dirichlet forms and symmetric Markov processes*, De Gruyter Studies in Mathematics, vol. 19, Walter de Gruyter & Co., Berlin, 1994. MR 1303354 142

[90] ———, *Dirichlet forms and symmetric Markov processes*, extended ed., De Gruyter Studies in Mathematics, vol. 19, Walter de Gruyter & Co., Berlin, 2011. MR 2778606 60

[91] Fuqing Gao and Jeremy Quastel, *Exponential decay of entropy in the random transposition and Bernoulli-Laplace models*, Ann. Appl. Probab. **13** (2003), no. 4, 1591–1600. MR 2023890 34, 40, 41, 79, 80, 82, 90, 112

[92] Kevin Garbe and Jan Vondrak, *Concentration of Lipschitz functions of negatively dependent variables*, arXiv preprint arXiv:1804.10084 (2018). 79

[93] Ankit Garg, Tarun Kathuria, and Nikhil Srivastava, *Scalar Poincaré implies matrix Poincaré*, Electron. Commun. Probab. **26** (2021), Paper No. 20, 4. MR 4245747 100

[94] Ivan Gentil, Arnaud Guillin, and Laurent Miclo, *Modified logarithmic Sobolev inequalities and transportation inequalities*, Probab. Theory Related Fields **133** (2005), no. 3, 409–436. MR 2198019 25

[95] Ivan Gentil and Simon Zugmeyer, *A family of Beckner inequalities under various curvature-dimension conditions*, Bernoulli **27** (2021), no. 2, 751–771. MR 4255214 27

[96] Reza Gheissari, Eyal Lubetzky, and Yuval Peres, *Concentration inequalities for polynomials of contracting Ising models*, Electron. Commun. Probab. **23** (2018), Paper No. 76, 12. MR 3873783 2, 32

[97] Evarist Giné, Rafał Latała, and Joel Zinn, *Exponential and moment inequalities for U-statistics*, High dimensional probability, II (Seattle, WA, 1999), Progr. Probab., vol. 47, Birkhäuser Boston, Boston, MA, 2000, pp. 13–38. MR 1857312 49

[98] Sharad Goel, *Modified logarithmic Sobolev inequalities for some models of random walk*, Stochastic Process. Appl. **114** (2004), no. 1, 51–79. MR 2094147 89

[99] F. Götze, H. Sambale, and A. Sinulis, *Higher order concentration for functions of weakly dependent random variables*, Electron. J. Probab. **24** (2019), Paper No. 85, 19. MR 4003138 3, 9, 10, 30, 32, 34, 52, 57

[100] Friedrich Götze, Holger Sambale, and Arthur Sinulis, *Concentration inequalities for bounded functionals via log-Sobolev-type inequalities*, J. Theoret. Probab. **34** (2021), no. 3, 1623–1652. MR 4289897 4, 9, 10, 30, 31, 57, 58

[101] N. Gozlan, *A characterization of dimension free concentration in terms of transportation inequalities*, Ann. Probab. **37** (2009), no. 6, 2480–2498. MR 2573565 2, 3

[102] ______, *Transport-entropy inequalities on the line*, Electron. J. Probab. **17** (2012), no. 49, 18. MR 2946156 3

[103] N. Gozlan, C. Roberto, and P. M. Samson, *Hamilton Jacobi equations on metric spaces and transport entropy inequalities*, Rev. Mat. Iberoam. **30** (2014), no. 1, 133–163. MR 3186934 3

[104] N. Gozlan, C. Roberto, P. M. Samson, Y. Shu, and P. Tetali, *Characterization of a class of weak transport-entropy inequalities on the line*, Ann. Inst. Henri Poincaré Probab. Stat. **54** (2018), no. 3, 1667–1693. MR 3825894 3

[105] N. Gozlan, C. Roberto, P. M. Samson, and P. Tetali, *Kantorovich duality for general transport costs and applications*, J. Funct. Anal. **273** (2017), no. 11, 3327–3405. MR 3706606 2, 3

[106] Nathaël Gozlan, Ronan Herry, and Giovanni Peccati, *Transport inequalities for random point measures*, J. Funct. Anal. **281** (2021), no. 9, Paper No. 109141, 45. MR 4283247 10, 44, 51

[107] David Gross and Vincent Nesme, *Note on sampling without replacing from a finite collection of matrices*, 2010. 37, 113

[108] Leonard Gross, *Logarithmic Sobolev inequalities*, Amer. J. Math. **97** (1975), no. 4, 1061–1083. MR 420249 3, 9, 59

[109] A. Guillin and A. Joulin, *Measure concentration through non-Lipschitz observables and functional inequalities*, Electron. J. Probab. **18** (2013), no. 65, 26. MR 3078024 3, 9, 12, 15

[110] Venkatesan Guruswami and Euiwoong Lee, *Strong inapproximability results on balanced rainbow-colorable hypergraphs*, Combinatorica **38** (2018), no. 3, 547–599. MR 3876877 59

[111] Anna Gusakova, Holger Sambale, and Christoph Thäle, *Concentration on Poisson spaces via modified Φ-Sobolev inequalities*, Stochastic Process. Appl. **140** (2021), 216–235. MR 4282691 10, 44

[112] Qiyang Han and Jon A. Wellner, *Complex sampling designs: uniform limit theorems and applications*, Ann. Statist. **49** (2021), no. 1, 459–485. MR 4206686 83

[113] D. L. Hanson and F. T. Wright, *A bound on tail probabilities for quadratic forms in independent random variables*, Ann. Math. Statist. **42** (1971), 1079–1083. MR 0279864 54

[114] Jonathan Hermon and Justin Salez, *Modified log-Sobolev inequalities for strong-Rayleigh measures*, arXiv preprint arXiv:1902.02775 (2019). 2, 3, 4, 5, 6, 10, 40, 41, 79, 80, 89, 93, 106, 107

[115] ______, *Entropy dissipation estimates for inhomogeneous zero-range processes*, Ann. Appl. Probab. **31** (2021), no. 5, 2275–2283. MR 4332696 4, 10, 40, 43

[116] Wassily Hoeffding, *A combinatorial central limit theorem*, Ann. Math. Statistics **22** (1951), 558–566. MR 44058 35, 117

[117] ______, *Probability inequalities for sums of bounded random variables*, J. Amer. Statist. Assoc. **58** (1963), 13–30. MR 144363 37, 113

[118] Richard Holley and Daniel Stroock, *Logarithmic Sobolev inequalities and stochastic Ising models*, J. Statist. Phys. **46** (1987), no. 5-6, 1159–1194. MR 893137 30

[119] Christian Houdré and Nicolas Privault, *Isoperimetric and related bounds on configuration spaces*, Statist. Probab. Lett. **78** (2008), no. 14, 2154–2164. MR 2458025 10, 44

[120] Christian Houdré and Nicolas Privault, *Surface measures and related functional inequalities on configuration spaces*, arXiv preprint arXiv:2103.11546 (2021). 10, 44

[121] Christian Houdré and Patricia Reynaud-Bouret, *Exponential inequalities, with constants, for U-statistics of order two*, Stochastic inequalities and applications, Progr. Probab., vol. 56, Birkhäuser, Basel, 2003, pp. 55–69. MR 2073426 44

[122] De Huang and Joel A. Tropp, *Nonlinear matrix concentration via semigroup methods*, Electron. J. Probab. **26** (2021), Paper No. 8, 31. MR 4216521 100

[123] Baptiste Nicolas Huguet, *Poincaré inequalities and integrated curvature-dimension criterion for generalised cauchy and convex measures*, arXiv preprint arXiv:2302.02684 (2023). 26

[124] Mark Jerrum, Jung-Bae Son, Prasad Tetali, and Eric Vigoda, *Elementary bounds on Poincaré and log-Sobolev constants for decomposable Markov chains*, Ann. Appl. Probab. **14** (2004), no. 4, 1741–1765. MR 2099650 80

[125] Ansgar Jüngel and Wen Yue, *Discrete Beckner inequalities via the Bochner-Bakry-Emery approach for Markov chains*, Ann. Appl. Probab. **27** (2017), no. 4, 2238–2269. MR 3693525 9, 43

[126] Tarun Kathuria, *A Matrix Bernstein Inequality for Strong Rayleigh Distributions*, arXiv preprint arXiv:2011.13340 (2020). 79, 84

[127] Tali Kaufman, Rasmus Kyng, and Federico Soldá, *Scalar and matrix Chernoff bounds from ℓ_∞-independence*, pp. 3732–3753. 80, 81, 85

[128] Tali Kaufman and Izhar Oppenheim, *High order random walks: beyond spectral gap*, Approximation, randomization, and combinatorial optimization. Algorithms and techniques, LIPIcs. Leibniz Int. Proc. Inform., vol. 116, Schloss Dagstuhl. Leibniz-Zent. Inform., Wadern, 2018, pp. Art. No. 47, 17. MR 3857285 80, 93

[129] ———, *High order random walks: beyond spectral gap*, Combinatorica **40** (2020), no. 2, 245–281. MR 4085988 42

[130] Jeong Han Kim and Van H. Vu, *Concentration of multivariate polynomials and its applications*, Combinatorica **20** (2000), no. 3, 417–434. MR 1774845 81

[131] Rasmus Kyng and Zhao Song, *A matrix Chernoff bound for strongly Rayleigh distributions and spectral sparsifiers from a few random spanning trees*, 59th Annual IEEE Symposium on Foundations of Computer Science—FOCS 2018, IEEE Computer Soc., Los Alamitos, CA, 2018, pp. 373–384. MR 3899605 6, 79, 80, 81, 82, 85, 97

[132] Günter Last, *Stochastic analysis for Poisson processes*, Stochastic analysis for Poisson point processes, Bocconi Springer Ser., vol. 7, Bocconi Univ. Press, [place of publication not identified], 2016, pp. 1–36. MR 3585396 44, 47

[133] Günter Last and Mathew Penrose, *Lectures on the Poisson process*, Institute of Mathematical Statistics Textbooks, vol. 7, Cambridge University Press, Cambridge, 2018. MR 3791470 44, 45, 46, 49, 145

[134] R. Latała, *Tail and moment estimates for some types of chaos*, Studia Math. **135** (1999), no. 1, 39–53. MR 1686370 54

[135] ———, *Estimates of moments and tails of Gaussian chaoses*, Ann. Probab. **34** (2006), no. 6, 2315–2331. MR 2294983 53, 81, 88, 104

[136] R. Latała and K. Oleszkiewicz, *Between Sobolev and Poincaré*, Geometric aspects of functional analysis, Lecture Notes in Math., vol. 1745, Springer, Berlin, 2000, pp. 147–168. MR 1796718 3, 4, 8, 9, 15, 25, 27, 139, 142

[137] Michel Ledoux, *On Talagrand's deviation inequalities for product measures*, ESAIM Probab. Statist. **1** (1995/97), 63–87. MR 1399224 7, 27, 28, 81, 114

[138] ———, *The concentration of measure phenomenon*, Mathematical Surveys and Monographs, vol. 89, American Mathematical Society, Providence, RI, 2001. MR 1849347 2, 89

[139] Michel Ledoux and Michel Talagrand, *Probability in Banach spaces*, Ergebnisse der Mathematik und ihrer Grenzgebiete (3) [Results in Mathematics and Related Areas (3)], vol. 23, Springer-Verlag, Berlin, 1991, Isoperimetry and processes. MR 1102015 38, 113

[140] Tzong-Yow Lee and Horng-Tzer Yau, *Logarithmic Sobolev inequality for some models of random walks*, Ann. Probab. **26** (1998), no. 4, 1855–1873. MR 1675008 34, 40, 42, 79

[141] David A. Levin and Yuval Peres, *Markov chains and mixing times*, American Mathematical Society, Providence, RI, 2017, Second edition of [MR2466937], With contributions by Elizabeth L. Wilmer, With a chapter on "Coupling from the past" by James G. Propp and David B. Wilson. MR 3726904 89

[142] Thomas M. Liggett, *Continuous time Markov processes*, Graduate Studies in Mathematics, vol. 113, American Mathematical Society, Providence, RI, 2010, An introduction. MR 2574430 60, 63, 88

[143] Sheng Lin Lu and Horng-Tzer Yau, *Spectral gap and logarithmic Sobolev inequality for Kawasaki and Glauber dynamics*, Comm. Math. Phys. **156** (1993), no. 2, 399–433. MR 1233852 80

[144] Eyal Lubetzky and Yufei Zhao, *On the variational problem for upper tails in sparse random graphs*, Random Structures Algorithms **50** (2017), no. 3, 420–436. MR 3632418 58

[145] Kyle Luh and Nicholas Pippenger, *Large-deviation bounds for sampling without replacement*, Amer. Math. Monthly **121** (2014), no. 5, 449–454. MR 3193733 37, 113

[146] K. Marton, *Bounding $\overline{d}$-distance by informational divergence: a method to prove measure concentration*, Ann. Probab. **24** (1996), no. 2, 857–866. MR 1404531 3

[147] ______, *A measure concentration inequality for contracting Markov chains*, Geom. Funct. Anal. **6** (1996), no. 3, 556–571. MR 1392329 3

[148] ______, *Measure concentration and strong mixing*, Studia Sci. Math. Hungar. **40** (2003), no. 1-2, 95–113. MR 2002993 32

[149] Katalin Marton, *Logarithmic Sobolev inequalities in discrete product spaces*, Combin. Probab. Comput. **28** (2019), no. 6, 919–935. MR 4015662 2, 10, 30

[150] Pascal Massart, *About the constants in Talagrand's concentration inequalities for empirical processes*, Ann. Probab. **28** (2000), no. 2, 863–884. MR 1782276 27, 38

[151] Bernard Maurey, *Construction de suites symétriques*, C. R. Acad. Sci. Paris Sér. A-B **288** (1979), no. 14, A679–A681. MR 533901 3, 35

[152] C. McDiarmid, *On the method of bounded differences*, Surveys in combinatorics, 1989 (Norwich, 1989), London Math. Soc. Lecture Note Ser., vol. 141, Cambridge Univ. Press, Cambridge, 1989, pp. 148–188. MR 1036755 3

[153] Colin McDiarmid, *Concentration*, Probabilistic methods for algorithmic discrete mathematics, Algorithms Combin., vol. 16, Springer, Berlin, 1998, pp. 195–248. MR 1678578 81

[154] L. Miclo, *An example of application of discrete Hardy's inequalities*, Markov Process. Related Fields **5** (1999), no. 3, 319–330. MR 1710983 57, 60, 61, 64

[155] Cassandra Milbradt and Martin Wahl, *High-probability bounds for the reconstruction error of PCA*, Statist. Probab. Lett. **161** (2020), 108741, 6. MR 4073556 9

[156] Emanuel Milman, *On the role of convexity in isoperimetry, spectral gap and concentration*, Invent. Math. **177** (2009), no. 1, 1–43. MR 2507637 21, 25

[157] V. D. Milman, *A new proof of A. Dvoretzky's theorem on cross-sections of convex bodies*, Funkcional. Anal. i Priložen. **5** (1971), no. 4, 28–37. MR 0293374 1, 2

[158] E. Mossel, K. Oleszkiewicz, and A. Sen, *On reverse hypercontractivity*, Geom. Funct. Anal. **23** (2013), no. 3, 1062–1097. MR 3061780 3, 4, 59, 60, 61, 62

[159] M. P. Müller, D. Gross, and J. Eisert, *Concentration of measure for quantum states with a fixed expectation value*, Comm. Math. Phys. **303** (2011), no. 3, 785–824. MR 2786217 2

[160] Matey Neykov, *Isotonic regression meets LASSO*, Electron. J. Stat. **13** (2019), no. 1, 710–746. MR 3914933 9

[161] Van Hoang Nguyen, *Dimensional variance inequalities of Brascamp-Lieb type and a local approach to dimensional Prékopa's theorem*, J. Funct. Anal. **266** (2014), no. 2, 931–955. MR 3132733 26

[162] ———, *Φ-entropy inequalities and asymmetric covariance estimates for convex measures*, Bernoulli **25** (2019), no. 4A, 3090–3108. MR 4003575 27

[163] Ivan Nourdin, Giovanni Peccati, and Xiaochuan Yang, *Restricted hypercontractivity on the Poisson space*, Proc. Amer. Math. Soc. **148** (2020), no. 8, 3617–3632. MR 4108865 44

[164] Ryan O'Donnell, *Analysis of Boolean functions*, Cambridge University Press, New York, 2014. MR 3443800 81, 87

[165] F. Otto and C. Villani, *Generalization of an inequality by Talagrand and links with the logarithmic Sobolev inequality*, J. Funct. Anal. **173** (2000), no. 2, 361–400. MR 1760620 3

[166] James Oxley, *Matroid theory*, second ed., Oxford Graduate Texts in Mathematics, vol. 21, Oxford University Press, Oxford, 2011. MR 2849819 42

[167] Daniel Paulin, *The convex distance inequality for dependent random variables, with applications to the stochastic travelling salesman and other problems*, Electron. J. Probab. **19** (2014), no. 68, 34. MR 3248197 2, 3, 100

[168] Daniel Paulin, Lester Mackey, and Joel A. Tropp, *Efron-Stein inequalities for random matrices*, Ann. Probab. **44** (2016), no. 5, 3431–3473. MR 3551202 81

[169] Robin Pemantle, *Towards a theory of negative dependence*, vol. 41, 2000, Probabilistic techniques in equilibrium and nonequilibrium statistical physics, pp. 1371–1390. MR 1757964 79

[170] Robin Pemantle and Yuval Peres, *Concentration of Lipschitz functionals of determinantal and other strong Rayleigh measures*, Combin. Probab. Comput. **23** (2014), no. 1, 140–160. MR 3197973 2, 5, 6, 41, 79, 80, 82, 83, 85, 93

[171] V. Pestov, *Dynamics of infinite-dimensional groups*, University Lecture Series, vol. 40, American Mathematical Society, Providence, RI, 2006. MR 2277969 2

[172] Gilles Pisier, *The volume of convex bodies and Banach space geometry*, Cambridge Tracts in Mathematics, vol. 94, Cambridge University Press, Cambridge, 1989. MR 1036275 55

[173] Bartłomiej Polaczyk, *Concentration bounds for sampling without replacement and hoeffding statistics*, arXiv preprint arXiv:2301.03096 (2023). 7

[174] Maxim Raginsky, *Strong data processing inequalities and Φ-sobolev inequalities for discrete channels*, IEEE Trans. Inf. Theory **62** (2016), no. 6, 3355–3389. 59

[175] Matthias Reitzner, *Poisson point processes: large deviation inequalities for the convex distance*, Electron. Commun. Probab. **18** (2013), no. 96, 7. MR 3151752 10, 44

[176] Patricia Reynaud-Bouret, *Adaptive estimation of the intensity of inhomogeneous Poisson processes via concentration inequalities*, Probab. Theory Related Fields **126** (2003), no. 1, 103–153. MR 1981635 44, 49, 50

[177] Justin Salez, *A sharp log-Sobolev inequality for the multislice*, Ann. H. Lebesgue **4** (2021), 1143–1161. MR 4353960 40

[178] Holger Sambale and Arthur Sinulis, *Logarithmic Sobolev inequalities for finite spin systems and applications*, Bernoulli **26** (2020), no. 3, 1863–1890. MR 4091094 10, 30, 33, 88, 89

[179] ______, *Concentration inequalities on the multislice and for sampling without replacement*, Journal of Theoretical Probability (2021), 1–26. 79, 80, 82, 88, 100, 103

[180] ______, *Modified log-Sobolev inequalities and two-level concentration*, ALEA Lat. Am. J. Probab. Math. Stat. **18** (2021), no. 1, 855–885. MR 4243518 100

[181] Paul-Marie Samson, *Transport-entropy inequalities on locally acting groups of permutations*, Electron. J. Probab. **22** (2017), Paper No. 62, 33. MR 3690287 79, 80

[182] G. Schechtman, *Lévy type inequality for a class of finite metric spaces*, Martingale theory in harmonic analysis and Banach spaces (Cleveland, Ohio, 1981), Lecture Notes in Math., vol. 939, Springer, Berlin-New York, 1982, pp. 211–215. MR 668548 3

[183] G. Scheffer, *Inégalités fonctionnelles, géométrie conforme et noyaux markoviens*, Ph.D. thesis, Univ. Toulouse 3, 2001. 26

[184] Qi-Man Shao, *A comparison theorem on moment inequalities between negatively associated and independent random variables*, J. Theoret. Probab. **13** (2000), no. 2, 343–356. MR 1777538 79

[185] Y. Shu and M. Strzelecki, *A characterization of a class of convex log-Sobolev inequalities on the real line*, Ann. Inst. Henri Poincaré Probab. Stat. **54** (2018), no. 4, 2075–2091. MR 3865667 2

[186] V. N. Sudakov and B. S. Cirelson, *Extremal properties of half-spaces for spherically invariant measures*, Zap. Naučn. Sem. Leningrad. Otdel. Mat. Inst. Steklov. (LOMI) **41** (1974), 14–24, 165, Problems in the theory of probability distributions, II. MR 0365680 2

[187] Feder T. and Mihail M., *Balanced matroids*, In Proceedings of the Twenty-fourth Annual ACM Symposium on Theory of Computing. STOC '92 (ACM, ed.), 1992, pp. 26–38. 41

[188] M. Talagrand, *Sharper bounds for Gaussian and empirical processes*, Ann. Probab. **22** (1994), no. 1, 28–76. MR 1258865 38

[189] ______, *A new look at independence*, Ann. Probab. **24** (1996), no. 1, 1–34. MR 1387624 3

[190] ______, *Transportation cost for Gaussian and other product measures*, Geom. Funct. Anal. **6** (1996), no. 3, 587–600. MR 1392331 3

[191] Michel Talagrand, *An isoperimetric theorem on the cube and the Kintchine-Kahane inequalities*, Proc. Amer. Math. Soc. **104** (1988), no. 3, 905–909. MR 964871 2, 81, 86

[192] ______, *Concentration of measure and isoperimetric inequalities in product spaces*, Inst. Hautes Études Sci. Publ. Math. (1995), no. 81, 73–205. MR 1361756 1, 2, 3, 27, 35, 80, 81

[193] ______, *New concentration inequalities in product spaces*, Invent. Math. **126** (1996), no. 3, 505–563. MR 1419006 3, 7, 27, 37, 111

[194] Yves Tillé, *Sampling algorithms*, Springer Series in Statistics, Springer, New York, 2006. MR 2225036 79, 82

[195] Ilya Tolstikhin, Nikita Zhivotovskiy, and Gilles Blanchard, *Permutational Rademacher complexity. A new complexity measure for transductive learning*, Algorithmic learning theory, Lecture Notes in Comput. Sci., vol. 9355, Springer, Cham, 2015, pp. 209–223. MR 3480745 7, 10, 35, 36, 37, 38, 111, 113, 114

[196] Joel A. Tropp, *Freedman's inequality for matrix martingales*, Electron. Commun. Probab. **16** (2011), 262–270. MR 2802042 96

[197] ______, *User-friendly tail bounds for sums of random matrices*, Found. Comput. Math. **12** (2012), no. 4, 389–434. MR 2946459 6, 81, 82, 94

[198] Roman Vershynin, *High-dimensional probability*, Cambridge Series in Statistical and Probabilistic Mathematics, vol. 47, Cambridge University Press, Cambridge, 2018, An introduction with applications in data science, With a foreword by Sara van de Geer. MR 3837109 2, 88

[199] Nicolas Verzelen and Elisabeth Gassiat, *Adaptive estimation of high-dimensional signal-to-noise ratios*, Bernoulli **24** (2018), no. 4B, 3683–3710. MR 3788186 9, 88

[200] Feng-Yu Wang, *A generalization of Poincaré and log-Sobolev inequalities*, Potential Anal. **22** (2005), no. 1, 1–15. MR 2127729 25

[201] Frederic Weber and Rico Zacher, *The entropy method under curvature-dimension conditions in the spirit of Bakry-Émery in the discrete setting of Markov chains*, J. Funct. Anal. **281** (2021), no. 5, Paper No. 109061, 81. MR 4253929 9

[202] Liming Wu, *A new modified logarithmic Sobolev inequality for Poisson point processes and several applications*, Probab. Theory Related Fields **118** (2000), no. 3, 427–438. MR 1800540 10, 44, 45

Appendix A

Beckner inequalities and moment estimates

Throughout this appendix, we use the notation from Chapter 2.

A.1 Auxiliary lemmas

In this section we provide proofs of technical lemmas used in Section 2.2.

Proof of Lemma 2.2.3. Consider a nonnegative f such that $f, f^{p-1} \in \mathrm{Dom}(\mathcal{E})$. For $t > 0$, let f_t denote $\min(f, t)$. The function $x \mapsto \min(x, t)$ is a contraction, whence $f_t, f_t^{p-1} \in \mathrm{Dom}(\mathcal{E})$. Moreover, for any $t > 0$ and any non-decreasing function φ

$$(f_t(x) - f_t(y))(\varphi(f_t(x)) - \varphi(f_t(y))) \leq (f(x) - f(y))(\varphi(f(x)) - \varphi(f(y))),$$

so by Assumption 1,

$$\sup_{t>0} \mathcal{E}(f_t, f_t^{p-1}) \leq \mathcal{E}(f, f^{p-1}).$$

It remains to show that $\mu(f^p) - \mu(f)^p$ is well-defined and is the limit of $\mu(f_t^p) - \mu(f_t)^p$ as $t \to \infty$. By the Lebesgue monotone convergence theorem, it suffices to show that $\mu(f) < \infty$. To this end it is enough to show that $\sup_t \mu(f_t) < \infty$. This inequality is trivially satisfied if $\sup_t \mu(f_t)^p \leq \mathcal{E}(f, f^{p-1})$. Assume thus that $\mu(f_t)^p > \mathcal{E}(f, f^{p-1})$ for some t. Applying Beckner's inequality (Bec-p) to the function f_t, we get that $\frac{1+\alpha_p}{\alpha_p}\mu(f_t)^p \geq \mu(f_t^p)$ (recall that $p/2 \leq 1$). Whence, by the Paley–Zygmund inequality (see, e.g., [80, Corollary 3.3.2])

$$\begin{aligned} \mu\big(f > \frac{1}{2}\mu(f_t)\big) &\geq \mu\big(f_t > \frac{1}{2}\mu(f_t)\big) \\ &\geq \Big(\frac{1}{2^p}\frac{\mu(f_t)^p}{\mu(f_t^p)}\Big)^{\frac{1}{p-1}} \geq 2^{-\frac{p}{p-1}}\Big(\frac{\alpha_p}{1+\alpha_p}\Big)^{\frac{1}{p-1}}. \end{aligned}$$

Thus, $\frac{1}{2}\mu(f_t)$ must be bounded by an appropriate quantile of f, whence we proved that $\sup_{t>0} \mu(f_t)$ is finite as desired. □

Proof of Lemma 2.2.7. Fix any $s \geq 1$. The Lipschitz property of the appropriate maps on the interval $[\inf f, \sup f]$ and Assumption 1 imply that for any $u \in \mathbb{R}$, $f^u, f^u \log f \in \mathrm{Dom}(\mathcal{E})$. Denote $h_\varepsilon(x) = \frac{x^\varepsilon - 1}{\varepsilon} - \log x$ for $\varepsilon \neq 0$. Then $f^{s-1}h_\varepsilon(f) \in \mathrm{Dom}(\mathcal{E})$ (since $\mathrm{Dom}(\mathcal{E})$ is a linear space) and by the bilinearity of $\mathcal{E}$,

$$\frac{v(s+\varepsilon) - v(s)}{\varepsilon} = \mathcal{E}(f, f^{s-1}\log f) + \mathcal{E}(f, f^{s-1}h_\varepsilon(f)).$$

We will show that

$$|x^{s-1}h_\varepsilon(x) - y^{s-1}h_\varepsilon(y)| \le C|\varepsilon| \cdot |x-y|$$

for all $x, y \in [\inf f, \sup f]$, $|\varepsilon|$ small enough, and some positive constant C dependent on s and f. By Assumption 1, this will allow to conclude that $|\mathcal{E}(f, f^{s-1}h_\varepsilon(f))| \le C|\varepsilon| \cdot \mathcal{E}(f,f) \to 0$ as $\varepsilon \to 0$.

By the Taylor expansion of x^ε in ε with the integral form of the remainder

$$h_\varepsilon(x) = \frac{1}{\varepsilon}\int_0^\varepsilon (\varepsilon - u)x^u \log^2 x\, du,$$

so

$$x^{s-1}h_\varepsilon(x) - y^{s-1}h_\varepsilon(y) = \frac{1}{\varepsilon}\int_0^\varepsilon (g_u(x) - g_u(y))(\varepsilon - u)\, du,$$

where $g_u(x) = x^{u+s-1}\log^2(x)$. We can and do assume without loss of generality that $\varepsilon \in (1-s, 1)$. Recalling that f is bounded and separated from zero, it follows that

$$\sup\Big\{\frac{1}{2}|g_u'(x)| \colon u \in (1-s,1),\, x \in [\inf f, \sup f]\Big\} =: C < \infty.$$

The proof is concluded by taking the absolute value, passing with it under the integral and estimating the increments of g_u.

□

A.2 Known implications between functional inequalities

In this section we provide sketches of proofs of previously known implications between functional inequalities discussed in Section 2.1.4. Although in the literature the results are commonly stated in the Markov kernel setting (sometimes only in the finite state space or continuous case), their proofs are mostly based on pointwise inequalities which imply comparison of Dirichlet forms. Hence, by virtue of Assumption 1 they pass directly to our setting at the cost of adding just a few technical details (needed mostly in order to make sure that all quantities are well-defined). We remark that a section containing implications between the Poincaré, log-Sobolev, and modified log-Sobolev inequality in an abstract setting, but under a somewhat different set of assumptions than ours, can be found in the article [42] by Bobkov and Tetali. In all the statements below we assume that Assumption 1 is satisfied.

Let us start with the implication between (LSI) and (mLSI).

Proposition A.2.1. *If the log-Sobolev inequality (LSI) holds with some constant ρ_1, then the modified log-Sobolev inequality (mLSI) holds with $\rho_0 \ge 4\rho_1$.*

Proof. The argument is based on the following pointwise inequality introduced by Bobkov and Tetali [42]:

$$4\big(\sqrt{a} - \sqrt{b}\big)^2 \le (a-b)(\log(a) - \log(b)), \quad a, b > 0. \tag{A.2.1}$$

Assuming that (LSI) holds for all $g \in \mathrm{Dom}(\mathcal{E})$, let us consider f, such that $f, \log f \in \mathrm{Dom}(\mathcal{E})$. Denote $f_\varepsilon = \max(f, \varepsilon)$. By Assumption 1, the fact that $x \mapsto \max(x, \varepsilon)$ is a contraction and monotonicity of the log, one can easily see that $\mathcal{E}(f_\varepsilon, \log(f_\varepsilon)) \leq \mathcal{E}(f, \log(f))$. Taking into account that $\mathrm{Ent}_\mu(f) = \lim_{\varepsilon\to 0} \mathrm{Ent}_\mu(f_\varepsilon)$, one can thus assume that f is separated from zero. Thus, again by Assumption 1, and Lipschitz property of the square root away from zero, $g = \sqrt{f} \in \mathrm{Dom}(\mathcal{E})$, and $\rho_1 \mathrm{Ent}_\mu(f) \leq \mathcal{E}(\sqrt{f}, \sqrt{f}) \leq \frac{1}{4}\mathcal{E}(f, \log(f))$, where in the last inequality we used (A.2.1) and Assumption 1. □

Let us now pass to the relation between (Bec'-q) and (LSI).

Proposition A.2.2. *If dual Beckner's inequality* (Bec'-q) *holds for every* $q \in [1, 2)$ *with* β_q *bounded away from zero, then the log-Sobolev inequality* (LSI) *holds as well with* $\rho_1 \geq \frac{1}{2} \limsup_{q\to 2^-} \beta_q$*. Conversely if* (LSI) *holds, then so does* (Bec'-q) *for every* $q \in [1, 2)$*, with* $\beta_q \geq q\rho_1$*.*

Proof. To obtain the first part of the proposition, it is enough to apply (Bec'-q) to $|g|$ (note that $|g| \in \mathrm{Dom}(\mathcal{E})$ by Assumption 1), divide by $2 - q$, and pass to the limit $q \to 2$, observing that

$$\frac{\mu(g^2) - \mu(g^q)^{2/q}}{2-q} \longrightarrow \frac{1}{2}\mathrm{Ent}_\mu(g^2) \quad \text{as} \quad q \to 2^-.$$

One obtains $(2^{-1} \limsup_{q\to 2-} \beta_q) \mathrm{Ent}_\mu(g^2) \leq \mathcal{E}(|g|, |g|) \leq \mathcal{E}(g, g)$, where the last inequality follows from another application of Assumption 1.

The second part follows from a lemma proved in [136], which asserts that if $g \in L_2(\mathcal{X}, \mu)$, then the function

$$[1, 2) \ni q \longmapsto \frac{\mu(g^2) - \mu(g^q)^{2/q}}{1/q - 1/2}$$

is increasing. Note that in our setting square integrability of g for $g \in \mathrm{Dom}(\mathcal{E})$ is a part of the assumption (LSI). □

Proposition A.2.3. *If the Beckner inequality* (Bec-p) *holds for every* $p \in (1, 2]$ *with* α_p *bounded away from zero, then the modified log-Sobolev inequality* (mLSI) *holds as well with* $\rho_0 \geq 2 \limsup_{p\to 1^+} \alpha_p$*.*

Proof. Let us consider f such that $f, \log(f) \in \mathrm{Dom}(\mathcal{E})$. Additionally, let us assume that f is bounded and separated from zero. In particular $f^p \in \mathrm{Dom}(\mathcal{E})$ for $p \geq 1$. Taking the right derivative of the function $p \mapsto \mu(f^p) - \mu(f)^p$ at $p = 1$ and using (Bec-p) together with Lemma 2.2.7, we obtain (mLSI) with $\rho_0 \geq 2 \limsup_{p\to 1+} \alpha_p$.

It remains to remove the additional assumptions on f. The assumption that f is separated from zero can be removed in the same way as in the proof of Proposition A.2.1. Let us therefore focus on the boundedness assumption. The argument is a variation of the one used in Lemma 2.2.3. Setting $f_t = \min(f, t)$, by Assumption 1 we have $\mathcal{E}(f_t, \log(f_t)) \leq \mathcal{E}(f, \log(f)) < \infty$, so it remains to show that $\sup_{t>0} \mu(f_t) < \infty$, as it will prove integrability of f, which will allow to pass to the limit with $t \to \infty$ in (mLSI) for f_t.

Let $\varphi\colon [0, \infty) \to [0, \infty)$ be any increasing convex function, $\varphi(0) = 0$, such that for large x, $\varphi(x) = x \log x$. Set $g_t = f_t/\mu(f_t)$ for $t > t_0 = \inf\{t\colon \mu(f_t) > 1\}$ – we assume that $t_0 < \infty$ as otherwise $\sup_{t>0} \mu(f_t) < \infty$ as desired. Since $x \log x$ is bounded from below and $\mathrm{Ent}_\mu(f_t)/\mu(f_t) = \mu(g_t \log(g_t))$, we obtain $\sup_{t>t_0} \mu(\varphi(g_t)) < \infty$. Thus, by convexity and monotonicity of φ,

$$\lim_{\delta\to 0} \sup_{t>t_0} \mu(\varphi(\delta g_t)) = 0.$$

Let $\delta < 1/4$ be such that for all $t > t_0$, $\mu(\varphi(\delta g_t)) < 1/4$. Denoting by φ^* the Legendre transform of φ, given by $\varphi^*(y) = \sup_{x\geq 0}(xy - \varphi(x))$, we have

$$1 = \mu(g_t) \leq \delta + \mu(g_t \mathbf{1}_{\{g_t>\delta\}}) \leq \delta + \mu(\varphi(\delta g_t)) + \varphi^*(\delta^{-1})\mu(g_t \geq \delta),$$

which gives

$$\mu\big(f \geq \delta\mu(f_t)\big) \geq \mu\big(f_t \geq \delta\mu(f_t)\big) = \mu(g_t \geq \delta) \geq \frac{1}{2\varphi^*(\delta^{-1})} > 0.$$

This shows that $\sup_{t>0}\mu(f_t) = \sup_{t>t_0}\mu(f_t)$ is dominated by an appropriate quantile of f and is therefore finite, ending the proof. □

Let us now pass to the relation between Beckner's inequalities (Bec-p) and (Bec'-q).

Proposition A.2.4. *Let $p \in (1,2]$ and $q = 2/p \in [1,2)$. If dual Beckner's inequality (Bec'-q) holds with constant β_q then Beckner's inequality (Bec-p) holds with constant $\alpha_p \geq \beta_q$. Conversely, if (Bec-p) holds with constant α_p, then (Bec'-q) holds with constant $\beta_q \geq q(2-q)\alpha_p$.*

Proof. Assume first (Bec'-q). By Lemma 2.2.3 in order to prove (Bec-p) it is enough to show that it holds for bounded f. Assume thus that f is bounded and $f, f^{p-1} \in \mathrm{Dom}(\mathcal{E})$. Set $g = f^{p/2}$. Since $p/2 \geq p-1$, $g \in \mathrm{Dom}(\mathcal{E})$ and thus

$$\begin{aligned}\beta_q(\mu(f^p) - \mu(f)^p) &= \beta_q(\mu(g^2) - \mu(g^q)^{2/q}) \\ &\leq (2-q)\mathcal{E}(g,g) = \frac{2(p-1)}{p}\mathcal{E}(g,g)\end{aligned}$$

By Lemma 2.2.5, $\mathcal{E}(g,g) \leq \frac{p^2}{4(p-1)}\mathcal{E}(f, f^{p-1})$, which implies (Bec-p).

The second part of the proposition follows by the first inequality from Lemma 2.2.5 for functions g separated from zero and infinity (the assumption is needed in order to assure that for $f = g^{2/p}$ we have $f, f^{p-1} \in \mathrm{Dom}(\mathcal{E})$. An extension to general functions can be obtained by appropriate truncations analogously as in the other implications we have considered so far. Since we do not use this implication in any part of this paper, we skip the details. □

Finally, let us show that the Poincaré inequality is implied both by the modified log-Sobolev inequality (mLSI) and by Beckner's inequality (Bec-p) (with fixed p)

Proposition A.2.5. *If the modified log-Sobolev inequality (mLSI) holds with constant ρ_0, then the Poincaré inequality (P) holds with $\lambda \geq \rho_0/2$.*

Proof. Again the argument is well known and one just needs to adjust it to our setting. If $f \in \mathrm{Dom}(\mathcal{E})$ is bounded (say $\|f\|_\infty = M$), then set $g_\varepsilon = e^{\varepsilon f}$. Using Taylor's expansion we get

$$\mathrm{Ent}_\mu(g_\varepsilon) = \frac{1}{2}\varepsilon^2 \mathrm{Var}_\mu(f) + o(\varepsilon^2).$$

On the other hand, using the inequality

$$(e^a - e^b)(a-b) \leq e^{\varepsilon M}(a-b)^2$$

valid for $a, b \in [-\varepsilon M, \varepsilon M]$, together with Assumption 1, we obtain

$$\mathcal{E}(g_\varepsilon, \log g_\varepsilon) \leq \varepsilon^2 e^{\varepsilon M}\mathcal{E}(f,f).$$

To obtain (P) for bounded functions it is thus enough to apply (mLSI) to g_ε and let $\varepsilon \to 0^+$. To extend this to $\mathrm{Dom}(\mathcal{E})$ note that the Poincaré inequality for bounded functions implies (in fact is equivalent to) (Bec-p) with $p = 2$ for bounded functions. Thus by Assumption 1 and Lemma 2.2.3 if $f \in \mathrm{Dom}(\mathcal{E})$, then f is square integrable. It is thus enough to set $f_t = \max(-t, \min(f, t))$ for $t > 0$, apply (P) to f_t and pass with $t \to \infty$, using the fact that by Assumption 1, $\mathcal{E}(f_t, f_t) \le \mathcal{E}(f, f)$. □

Proposition A.2.6. *Let $p \in (1, 2]$. If Beckner's inequality (Bec-p) holds with constant α_p, then the Poincaré inequality (P) holds with constant $\lambda \ge \alpha_p$.*

Proof. As in the proof of Proposition A.2.5, it is enough to prove (P) for bounded functions. Assume thus that $f \in \mathrm{Dom}(\mathcal{E})$ is bounded. Then, for sufficiently small ε, $(1 + \varepsilon f)^{p-1} \in \mathrm{Dom}(\mathcal{E})$. Thus

$$\begin{aligned}\alpha_p\Big(\mu((1+\varepsilon f)^p) - (\mu(1+\varepsilon f))^p\Big) &\le \frac{p}{2}\mathcal{E}(1+\varepsilon f, (1+\varepsilon f)^{p-1})\\ &= \frac{\varepsilon p}{2}\mathcal{E}(f, (1+\varepsilon f)^{p-1}).\end{aligned}$$

The Taylor expansion reveals that for $\varepsilon \to 0$,

$$\mu((1+\varepsilon f)^p) - (\mu(1+\varepsilon f))^p = \frac{1}{2}p(p-1)\varepsilon^2 \operatorname{Var}_\mu(f) + o(\varepsilon^2).$$

On the other hand

$$\varepsilon\mathcal{E}(f, (1+\varepsilon f)^{p-1}) = (p-1)\varepsilon^2\mathcal{E}(f, f) + \varepsilon\mathcal{E}(f, (1+\varepsilon f)^{p-1} - 1 - (p-1)\varepsilon f).$$

To finish the proof it is thus enough to show that

$$\mathcal{E}(f, (1+\varepsilon f)^{p-1} - 1 - (p-1)\varepsilon f) = o(\varepsilon)$$

for $\varepsilon \to 0+$. Denote $M = \|f\|_\infty$ and denote $g(x) = (1+x)^{p-1} - 1 - (p-1)x$. For $a, b \in [-M, M]$ we have

$$|a-b||g(\varepsilon a) - g(\varepsilon b)| \le \varepsilon(a-b)^2 A_\varepsilon.$$

where $A_\varepsilon = \sup_{t\in[-\varepsilon M, \varepsilon M]} |g'(t)|$. Thus, by Assumption 1, $\mathcal{E}(f, (1+\varepsilon f)^{p-1} - 1 - (p-1)\varepsilon f) \le \varepsilon A_\varepsilon \mathcal{E}(f, f)$ and it remains to show that $\lim_{\varepsilon\to 0+} A_\varepsilon = 0$. This is however true, since g is continuously differentiable in the neighborhood of 0 and $g'(0) = 0$. □

Finally we address the question of the monotonicity of the constants in Beckner's inequalities.

Lemma A.2.7. *For $1 < p - \varepsilon < p \le 2$ and $a, b, > 0$,*

$$\begin{aligned}&\big(a^{(p-\varepsilon)/p} - b^{(p-\varepsilon)/p}\big)\big(a^{(p-\varepsilon)(p-1)/p} - b^{(p-\varepsilon)(p-1)/p}\big)\\ &\qquad\qquad \le \frac{(p-\varepsilon)^2(p-1)}{p^2(p-\varepsilon-1)}(a-b)(a^{p-\varepsilon-1} - b^{p-\varepsilon-1}).\end{aligned}$$

Proof. By the integral version of Chebyshev's sum inequality,

$$\begin{aligned}&\frac{a^{(p-\varepsilon)/p} - b^{(p-\varepsilon)/p}}{a-b} \cdot \frac{a^{(p-\varepsilon)(p-1)/p} - b^{(p-\varepsilon)(p-1)/p}}{a-b}\\ &= \frac{(p-\varepsilon)/p}{a-b}\int_a^b s^{(p-\varepsilon)/p-1} ds \cdot \frac{(p-\varepsilon)(p-1)/p}{a-b}\int_a^b s^{(p-\varepsilon)(p-1)/p-1} ds\\ &\le \frac{(p-\varepsilon)^2(p-1)}{p^2}\frac{1}{a-b}\int_a^b s^{p-\varepsilon-2} ds\\ &= \frac{(p-\varepsilon)^2(p-1)}{p^2(p-\varepsilon-1)}\frac{a^{p-\varepsilon-1} - b^{p-\varepsilon-1}}{a-b}\end{aligned}$$

(note that $s \mapsto s^{p-\varepsilon-2}$ and $s \mapsto s^{(p-\varepsilon)(p-1)/p-1}$ are both decreasing). □

Proposition A.2.8. *Suppose that for some $p \in (1,2]$ Beckner's inequality (Bec-p) holds with constant $\alpha_p > 0$. Let $0 < \varepsilon < p-1$. Then, Beckner's inequality (Bec-p) holds for $p-\varepsilon$ (in place of p) with constant*

$$\alpha_{p-\varepsilon} \geq \frac{(p-\varepsilon-1)p}{(p-\varepsilon)(p-1)}\alpha_p.$$

In particular, if Beckner's inequality (Bec-p) holds for all (equivalently: for some) $p \in (1,2]$, then the function $p \mapsto \frac{p}{p-1}\alpha_p^{opt}$, $p \in (1,2]$, is non-increasing.

Proof. By Jensen's inequality, (Bec-p), and Lemma A.2.7,

$$\begin{aligned}\alpha_p\big(\mu(f^{p-\varepsilon}) - \mu(f)^{p-\varepsilon}\big) &\leq \alpha_p\big(\mu(f^{p-\varepsilon}) - \mu(f^{(p-\varepsilon)/p})^p\big)\\ &\leq \frac{p}{2}\mathcal{E}(f^{(p-\varepsilon)/p}, f^{(p-\varepsilon)(p-1)/p})\\ &\leq \frac{p-\varepsilon}{2}\cdot\frac{(p-\varepsilon)(p-1)}{(p-\varepsilon-1)p}\mathcal{E}(f, f^{p-\varepsilon-1}).\end{aligned}$$

This finishes the proof. □

Remark A.2.9. In the case of dual Beckner's inequality (Bec'-q) we have:

- $q \mapsto \frac{1}{2-q}\beta_q^{\mathrm{opt}}$, $q \in [1,2)$ is non-decreasing (just by the Jensen inequality),
- $q \mapsto \frac{1}{q}\beta_q^{\mathrm{opt}}$, $q \in [1,2)$ is non-increasing (by a lemma proved in [136] concerning the monotonicity of the function $q \mapsto \frac{\mu(g^2)-\mu(g^q)^{2/q}}{1/q-1/2}$, $q \in [1,2)$).

A.3 Connections with Dirichlet forms

In this section we will provide a link between our assumptions and the usual theory of Dirichlet forms associated with Markov semigroups, in particular showing that our main assumptions are satisfied in this setting. As reference, we suggest the monographs [89, 21]. Recall that we work on a probability space $(\mathcal{X}, \mathcal{B}, \mu)$.

Let $P\colon [0,\infty)\times\mathcal{X}\times\mathcal{B} \to [0,1]$ be a homogeneous Markov transition function for which μ is an invariant measure. We will assume that P is reversible with respect to μ. We will often write $P_t(x,B)$ for $P(t,x,B)$, and we will also denote by $(P_t)_{t\geq 0}$ the semigroup of operators on $L_2(\mathcal{X},\mu)$ related to the transition function and defined as

$$P_t f(x) = \int_{\mathcal{X}} f(y)P_t(x,dy).$$

We will assume that this semigroup is strongly continuous.

It can be shown that for each $f \in L_2(\mathcal{X},\mu)$ the function

$$t \mapsto \frac{1}{2t}\int_{\mathcal{X}}\int_{\mathcal{X}}(f(y)-f(x))^2 P_t(x,dy)\mu(dx)$$

is non-increasing. Denoting

$$\mathrm{Dom}(\mathcal{E}) = \Big\{f \in L_2(\mathcal{X},\mu)\colon \sup_{t\geq 0}\frac{1}{2t}\int_{\mathcal{X}}\int_{\mathcal{X}}(f(y)-f(x))^2 P_t(x,dy)\mu(dx) < \infty\Big\}$$

and defining for $f, g \in \mathcal{E}$,

$$\mathcal{E}(f,g) = \lim_{t\to 0}\frac{1}{2t}\int_{\mathcal{X}}\int_{\mathcal{X}}(f(y)-f(x))(g(y)-g(x))P_t(x,dy)\mu(dx)$$

we obtain a nonnegative definite symmetric quadratic form.

In particular, for $f \in \mathrm{Dom}(\mathcal{E})$ we obtain

$$\mathcal{E}(f,f) = \lim_{t\to 0}\frac{1}{2t}\int_{\mathcal{X}}\int_{\mathcal{X}}(f(y)-f(x))^2P_t(x,dy)\mu(dx),$$

and it is straightforward to check that the Assumption 1 is satisfied in this case.

Let us now discuss the Assumption 2. If L is the infinitesimal operator of the semigroup $(P_t)_{t\geq 0}$, defined as

$$Lf = \lim_{h\to 0^+}\frac{P_hf - f}{h}$$

with the convergence in the L_2 sense, and $f, g \in \mathrm{Dom}(L)$, then

$$\mathcal{E}(f,g) = -\int_{\mathcal{X}} fLgd\mu.$$

If also $fg \in \mathrm{Dom}(L)$, then one obtains further the equality (2.1.7) where the carré du champ operator Γ is given by (2.1.8). One shows that $\Gamma(f,f) \geq 0$. In most applications the operator Γ is first defined on a suitable algebra of functions $\mathcal{A}_0 \subseteq \mathrm{Dom}(L)$ and then extended to some larger class $\mathcal{A}$. This is the situation described, e.g., in Chapter 1.14 of [21]. In the case of diffusions on Riemannian manifolds one usually takes $\mathcal{A}_0$ to be the algebra of smooth compactly supported functions and $\mathcal{A}$ – the algebra of all smooth functions. However, in the abstract setting there is no canonical choice of $\mathcal{A}$, so we will stick here to the basic case of $\mathcal{A} \subseteq \mathrm{Dom}(L)$ and refer to Chapter 3 of [21] for the axiomatic approach, which allows to introduce a general framework for an abstract counterpart of the theory of diffusions in $\mathbb{R}^n$.

The following proposition shows that in our basic setting the first part of Assumption 2 is satisfied for every algebra $\mathcal{A} \subseteq \mathrm{Dom}(L)$.

Proposition A.3.1. *Assume that $f\colon \mathcal{X} \to \mathbb{R}$ is a bounded function and $f, f^2 \in \mathrm{Dom}(L)$. Let $t, t' \geq 1$ satisfy $\frac{1}{t} + \frac{1}{t'} = 1$. Then for every differentiable convex, non-decreasing function $\varphi\colon [0,\infty) \to \mathbb{R}$ and every $c \in \mathbb{R}$*

$$\mathcal{E}(\varphi(|f+c|), |f+c|) \leq 2\int_{\mathcal{X}}\varphi'(|f+c|)\Gamma(f)d\mu \leq 2\Big\|\varphi'(|f+c|)\Big\|_{t'}\Big\|\Gamma(f)\Big\|_{t}.$$

Proof. Note that if $f^2, f \in \mathrm{Dom}(L)$ and f is bounded then

$$\begin{aligned}&\frac{1}{u}\int_{\mathcal{X}}(f(x)-f(y))^2P_u(x,dy)\\ &\quad= \frac{1}{u}\int_{\mathcal{X}}(f^2(y)-f^2(x))P_u(x,dy) - 2f(x)\frac{1}{u}\int_{\mathcal{X}}(f(y)-f(x))P_u(x,dy)\\ &\qquad\qquad\overset{u\to 0+}{\to} 2\Gamma(f) \qquad \text{(A.3.1)}\end{aligned}$$

in L_2.

By boundedness of f, the fact that convex functions are locally Lipschitz, and Assumption 1 (which we know to be satisfied in the Markov case), $\varphi(|f+c|), |f+c| \in \mathrm{Dom}(\mathcal{E})$. Moreover, denoting $g = f + c$,

$$
\begin{aligned}
&\mathcal{E}(\varphi(|g|), |g|)\\
&= \lim_{u\to 0+} \frac{1}{2u} \int_{\mathcal{X}} \int_{\mathcal{X}} (\varphi(|g(x)|) - \varphi(|g(y)|))(|g(x)| - |g(y)|) P_u(x, dy)\mu(dx)\\
&= \lim_{u\to 0+} \frac{1}{u} \int_{\mathcal{X}} \int_{\mathcal{X}} (\varphi(|g(x)|) - \varphi(|g(y)|))_+(|g(x)| - |g(y)|)_+ P_u(x, dy)\mu(dx)\\
&\le \liminf_{u\to 0+} \int_{\mathcal{X}} \varphi'(|g(x)|)\frac{1}{u} \int_{\mathcal{X}} (f(x) - f(y))_+^2 P_u(x, dy)\mu(dx)\\
&\le \liminf_{u\to 0+} \int_{\mathcal{X}} \varphi'(|g(x)|)\frac{1}{u} \int_{\mathcal{X}} (f(x) - f(y))^2 P_u(x, dy)\mu(dx)\\
&= 2\int_{\mathcal{X}} \varphi'(|f(x) + c|)\Gamma(f)(x)d\mu(dx),
\end{aligned}
$$

where in the second equality we used reversibility of the semigroup together with monotonicity of φ, and in the first inequality – convexity of φ. The last equality follows by boundedness of f and (A.3.1).

This proves the first inequality of the proposition. The second one follows by Hölder's inequality. □

As for the second part of Assumption 2, it is satisfied, e.g., if $\mathcal{A} \subseteq \mathrm{Dom}(L)$ is an algebra stable under compositions with smooth bounded Lipschitz functions vanishing at zero, which is a common assumption in this context (see Chapter 1.13 of [21]). Indeed, in this case one can define an appropriate sequence of smooth bounded contractions $\psi_n \colon \mathbb{R} \to \mathbb{R}$ with $\psi_n(0) = 0$, converging to $\psi(x) = x$ pointwise and take $f_n = \psi_n(f)$. Then clearly $f_n \to f$ and $|f_n| \le |f|$ μ-a.s. Moreover, for general f such that $f, f^2 \in \mathrm{Dom}(L)$ one still has (A.3.1) but this time in L_1. Thus using the contraction property of ψ_n one can easily show that

$$\Gamma(f_n) = \Gamma(\psi_n(f)) \le \Gamma(f) \quad \mu\text{-a.s.}$$

Combining this observation with Proposition A.3.1 one obtains

Proposition A.3.2. *If $\mathcal{A} \subseteq \mathrm{Dom}(L)$ is an algebra stable under composition with smooth bounded Lipschitz functions vanishing at zero, then the Assumption 2 is satisfied.*

This gives the basic setting for applying moment estimates of Proposition 2.3.3 in the Markovian case. Clearly, in concrete applications the moment inequalities can be extended to larger classes of functions – the details of such an extension and the choice of the class of functions may however depend on the particular case. We again refer to Chapter 3 of [21] for an extensive discussion of this issue.

A.4 Measurability on the Poisson space

In this section we would like to address the measurability issues indicated in Section 2.4.6, cf. Remark 2.4.18 therein.

Let us briefly recall the notation. The pair $(\mathcal{X}, \mathcal{B})$ is a measurable space endowed with a σ-finite measure λ, while $\mathcal{N}$ is the space of all $\mathbb{N} \cup \{\infty\}$-valued measures on $(\mathcal{X}, \mathcal{B})$ which can be expressed as countable sums of $\mathbb{N}$-valued measures. The measurable structure on $\mathcal{N}$ is given by the smallest σ-field $\mathcal{G}$ such that for all $B \in \mathcal{B}$ the map $\mu \mapsto \mu(B)$ is $\mathcal{G}$-measurable. An $\mathcal{N}$-valued random variable η is a Poisson process with intensity λ.

Recall also that $\mathcal{X}_n \in \mathcal{B}$ is a sequence of sets with $\bigcup_n \mathcal{X}_n = \mathcal{X}$, such that $\lambda(\mathcal{X}_n) < \infty$ for all n. We consider the space $\mathcal{M} \subset \mathcal{N}$ of measures of the form $\mu = \sum_{i=1}^{K} \delta_{x_i}$ where $K \leq \infty$ and $x_i \in \mathcal{X}$, such that for every n, $\mu(\mathcal{X}_n) < \infty$. We endow this space with the smallest σ-field $\mathcal{S}$ such that the maps $\mu \mapsto \mu(B)$ are measurable for all $B \in \mathcal{B}$. Finally, for $\rho = \sum_{k=1}^{K} \delta_{x_k} \in \mathcal{M}$, $A \in \mathcal{S}$, recall the definition

$$Q_\rho(A) = \lambda(\{\, x \colon \rho + \delta_x \in A \,\}) + \sum_{k=1}^{K} \mathbf{1}_A(\rho - \delta_{x_k}). \tag{A.4.1}$$

One of the reasons why we restrict to the space $\mathcal{M}$ is the possibility of interpreting the integral with respect to ρ as a sum. For general integer-valued measures, when a representation with Dirac deltas may not exist (see [133]) one may encounter problems with measurability and interpretation of $\rho - \delta_x$. For instance, the example below shows that even when ρ itself is a Dirac's delta, the set $\{\, x \colon \rho \geq \delta_x \,\}$ may not be measurable. Below we wil show that the definition (A.4.1) of Q does not depend on the representation of ρ as a sum of Dirac masses, and Q defined by the above formula is indeed a kernel on $\mathcal{M}$. Moreover, it can be approximated with kernels admitting finite values only.

Example A.4.1. Let $\mathcal{X} = [0,1]$ and $\mathcal{B} = \sigma(\{\, A \subset [0,1) : A \text{ is countable} \,\})$ so that

$$\begin{aligned}\mathcal{B} = \{\, C \subset [0,1) :\ & C \text{ is countable} \,\} \\ & \cup \{\, C \subset [0,1] :\ [0,1] \setminus C \text{ is countable and } 1 \in C \,\}.\end{aligned}$$

Let $\rho = \delta_1$ be the Dirac mass at point 1. Then $\{\, x \colon \rho \geq \delta_x \,\} = \{1\}$ but $\{1\} \notin \mathcal{B}$.

Let us now discuss the definition (A.4.1) and show that $Q_\rho(A)$ is well-defined, i.e., it is independent on the representation of ρ as a sum and that the mapping $\mathcal{M} \ni \rho \mapsto Q_\rho(A)$ is $\mathcal{S}$-measurable.

Let us begin with showing the measurability of the first integral in the definition (A.4.2). By Fubini's theorem, it suffices to show that $(x, \rho) \mapsto \rho + \delta_x$ is $\mathcal{B} \otimes \mathcal{S}$-measurable. Note that $\mathcal{S}$ is generated by the sets $D_{B,C} = \{\, \rho \colon \rho(B) \in C \,\}$ for $B \in \mathcal{B}$, $C \in \mathcal{B}(\mathbb{R})$. Thus, for arbitrary $B \in \mathcal{B}$, $C \in \mathcal{B}(\mathbb{R})$,

$$\begin{aligned}\{\, (x,\rho) \colon \rho + \delta_x \in D_{B,C} \,\} &= \{\, (x,\rho) \colon (\rho + \delta_x)(B) \in C \,\} \\ &= \big(B \times \{\, \rho \colon \rho(B) \in C - 1 \,\}\big) \cup \big(B^c \times \{\, \rho \colon \rho(B) \in C \,\}\big) \in \mathcal{B} \otimes \mathcal{S}\end{aligned}$$

as desired.

Set

$$Q^n_\rho(A) = \int_{\mathcal{X}_n} \mathbf{1}_A(\rho + \delta_x)\, \lambda(dx) + \sum_{k=1}^{\kappa} \mathbf{1}_{\mathcal{X}_n}(x_k) \mathbf{1}_A(\rho - \delta_{x_k}). \tag{A.4.2}$$

Note that for any $A \in \mathcal{S}$ and $\rho \in \mathcal{M}$, $Q^n_\rho(A) \nearrow Q_\rho(A)$ as $n \to \infty$ and that $Q^n_\rho(\mathcal{M}) = \lambda(\mathcal{X}_n) + \rho(\mathcal{X}_n) < \infty$.

Denote $\mathcal{A}$ to be those elements $A \in \mathcal{S}$ such that for any $n \in \mathbb{N}$ and $\rho \in \mathcal{M}$, $Q^n_\rho(A)$ is independent on the representation of ρ as a sum and that $\rho \mapsto Q^n_\rho(A)$ is $\mathcal{S}$-measurable

We verify that $\mathcal{A}$ is a λ-system. Firstly, $\emptyset \in \mathcal{A}$. Secondly, for any pairwise disjoint sets $A_1, A_2, \ldots \subset \mathcal{A}$ and any $n \in \mathbb{N}$, $\rho \mapsto Q^n_\rho(\bigcup_i A_i) = \sum_i Q^n_\rho(A_i)$ is well-defined and $\mathcal{S}$-measurable as a countable sum of well-defined and measurable mappings. Finally, for any $A \in \mathcal{A}$ and $n \in \mathbb{N}$, $\rho \mapsto Q^n_\rho(A^c) = \lambda(\mathcal{X}_n) + \rho(\mathcal{X}_n) - Q^n_\rho(A)$ is also well-defined and $\mathcal{S}$-measurable as a combination of well-defined and measurable mappings (here we use the fact that $\lambda(\mathcal{X}_n) + \rho(\mathcal{X}_n) < \infty$ by the definition of $\mathcal{X}_n$). Thus, $\mathcal{A}$ is a λ-system.

Consider the following π-system,

$$\mathcal{A}' = \Big\{ \{\mu \in \mathcal{M} \colon \mu(B_1) = a_1\} \cap \ldots \cap \{\mu \in \mathcal{M} \colon \mu(B_k) = a_k\} : \\ k \in \mathbb{N}, B_i \in \mathcal{B}, a_i \in \mathbb{N} \text{ for all } i \in \{1, \ldots, k\} \Big\}$$

We verify that for each $A \in \mathcal{A}'$, A is $\mathcal{S}$-measurable. For $B \in \mathcal{B}$, denote $B^1 = B$ and $B^{-1} = B^c$. For $\rho \in \mathcal{M}$, let $\rho = \sum_{i=1}^K \delta_{x_i}$ for some $K \le \infty$ and $\mathcal{X}$-valued random variables $x_1, x_2, \ldots$ so that for any $A \in \mathcal{A}'$,

$$\begin{aligned}
\sum_{i=1}^K \mathbf{1}_{\mathcal{X}_n}(x_i)\mathbf{1}_A(\rho - \delta_{x_i}) &= \sum_{i=1}^K \mathbf{1}_{\{x_i \in \mathcal{X}_n\}} \prod_{j=1}^k \big(\mathbf{1}_{\{x_i \in B_j,\, \rho(B_j) = a_j + 1\}} + \mathbf{1}_{\{x_i \in B_j^c,\, \rho(B_j) = a_j\}}\big) \\
&= \sum_{i=1}^K \sum_{\varepsilon_1, \ldots, \varepsilon_k = \pm 1} \mathbf{1}_{\{x_i \in \mathcal{X}_n \cap \bigcap_{j=1}^k B_j^{\varepsilon_j}\}} \prod_{j=1}^k \mathbf{1}_{\{\rho(B_j) = a_j + \frac{\varepsilon_j + 1}{2}\}} \\
&= \sum_{\varepsilon_1, \ldots, \varepsilon_k = \pm 1} \prod_{j=1}^k \mathbf{1}_{\{\rho(B_j) = a_j + \frac{\varepsilon_j + 1}{2}\}} \sum_{i=1}^K \mathbf{1}_{\{x_i \in \mathcal{X}_n \cap \bigcap_{j=1}^k B_j^{\varepsilon_j}\}} \\
&= \sum_{\varepsilon_1, \ldots, \varepsilon_k = \pm 1} \prod_{j=1}^k \mathbf{1}_{\{\rho(B_j) = a_j + \frac{\varepsilon_j + 1}{2}\}} \cdot \rho\big(\mathcal{X}_n \cap \bigcap_{j=1}^k B_j^{\varepsilon_j}\big),
\end{aligned} \tag{A.4.3}$$

which is $\mathcal{S}$-measurable as a finite sum of finite products of measurable mappings. Let us note that the right-hand side of (A.4.3) is also independent on the representation of ρ as a sum of Dirac deltas. Therefore, by Sierpiński–Dynkin's π-λ theorem applied to systems $\mathcal{A}$ and $\mathcal{A}'$, Q is well-defined and the mapping $\mathcal{M} \ni \rho \mapsto Q_\rho(A)$ is $\mathcal{S}$-measurable for any $A \in \mathcal{S}$.

Appendix B

Sampling without replacement and Hoeffding statistics

Throughout this appendix, we use notation from Chapter 5.

B.1 Proof of Lemma 5.2.1

Set $E = \mathbb{R}^n$ and $g(i) = e_i$, where $e_i \in \mathbb{R}^n$ is a vector with 1 on the i-th coordinate and 0's elsewhere. Moreover, let for any $v \in \mathbb{R}^n$,

$$\Psi(v) = \phi\Big(\sup_{x\in\mathcal{X}}\langle x, v\rangle\Big),$$

where $\langle\cdot,\cdot\rangle$ is the standard dot product. Then,

$$\phi(Z) = \phi\Big(\sup_{x\in\mathcal{X}}\langle x, \sum_{k=1}^m e_{I_k}\rangle\Big) = \Psi\Big(\sum_{k=1}^m g(I_k)\Big)$$

and identically $\phi(Z') = \Psi(\sum_{k=1}^m g(J_k))$. Finally, for any $v, w \in \mathbb{R}^n$ and $t \in [0,1]$,

$$\begin{aligned}
\Psi(tw + (1-t)v) &= \phi\Big(\sup_{x\in\mathcal{X}}\langle x, tw + (1-t)v\rangle\Big) \\
&\le \phi\Big(t\sup_{x\in\mathcal{X}}\langle x, w\rangle + (1-t)\sup_{x\in\mathcal{X}}\langle x, v\rangle\Big) \\
&\le t\Psi(w) + (1-t)\Psi(v),
\end{aligned}$$

where in the first inequality we have used that ϕ is increasing, and in the second inequality we have used that ϕ is convex. We conclude by applying Hoeffding's argument (5.2.2) to the pair (g, Ψ).

B.2 Proof of Lemma 5.2.4

Let us recall some facts regarding entropy. For any random variable Y measurable w.r.t. $\sigma(J_1, \dots, J_m)$ and any $k \in [m]$, let $\mathbb{E}^{(k)}$ denote the expectation w.r.t. J_k only, i.e.,

$$\mathbb{E}^{(k)}[Y] = \mathbb{E}\big[Y \mid J_1, \dots, J_{k-1}, J_{k+1}, \dots, J_m\big].$$

For such positive Y, recall the tensorization of entropy formula (cf., e.g., [51, Theorem 4.10])

$$\mathrm{Ent}(Y) \le \mathbb{E}\sum_{k=1}^m \mathrm{Ent}^{(k)}(Y), \tag{B.2.1}$$

where

$$\mathrm{Ent}^{(k)}(Y) = \mathbb{E}^{(k)}\big[Y\log Y\big] - \mathbb{E}^{(k)}\big[Y\big]\log\mathbb{E}^{(k)}\big[Y\big]$$

is the entropy functional corresponding to $\mathbb{E}^{(k)}$. Moreover, recall the following variational formula for the entropy

$$\mathrm{Ent}(Y) = \inf_{c>0}\mathbb{E}\Big[Y(\log Y - \log c) - (Y - c)\Big]. \tag{B.2.2}$$

Proof of Lemma 5.2.4. For $k \in [m]$, let

$$Z'_k = \sup_{x\in\mathcal{X}} \sum_{l=1, l\neq k}^{m} x_{J_l}$$

(if $m = 1$, then we put $u_1 = 0$). By the tensorization of entropy (B.2.1) and by (B.2.2),

$$\begin{aligned}
\mathrm{Ent}(e^{\lambda Z'}) &\leq \mathbb{E}\sum_{k=1}^{m}\mathrm{Ent}^{(k)}(e^{\lambda Z'})\\
&= \mathbb{E}\sum_{k=1}^{m}\inf_{c_k>0}\mathbb{E}^{(k)}\Big[e^{\lambda Z'}(\lambda Z' - \log c_k) - (e^{\lambda Z'} - c_k)\Big]\\
&\leq \mathbb{E}\sum_{k=1}^{m}\mathbb{E}^{(k)}\Big[e^{\lambda Z'}(\lambda Z' - \lambda Z'_k) - (e^{\lambda Z'} - e^{\lambda Z'_k})\Big]\\
&\leq \mathbb{E}\Big[e^{\lambda Z'}\sum_{k=1}^{m}\phi(-\lambda(Z' - Z'_k))\Big],
\end{aligned} \tag{B.2.3}$$

where $\phi(z) = e^z - z - 1$.

Note that

$$\sum_{k=1}^{m}(Z' - Z'_k) \leq Z'$$

and that for any $z \in [0,1]$ and $\lambda \geq 1/4$, by the convexity of the function $z \mapsto e^{-z/4} - 1$

$$\begin{aligned}
\phi(-\lambda z) = e^{-\lambda z} - 1 + \lambda z &\leq e^{-z/4} - 1 + \lambda z\\
&\leq -\frac{z}{4}e^{-1/4} + \lambda z \leq \Big(\lambda - \frac{1}{8}\Big)z.
\end{aligned}$$

Since $\mathcal{X} \subset [0,1]^n$ by assumption, therefore $0 \leq Z' - Z'_k \leq 1$ and whence we can estimate (B.2.3) further for any $\lambda \geq 1/4$ as follows,

$$\mathrm{Ent}(e^{\lambda Z'}) \leq \Big(\lambda - \frac{1}{8}\Big)\mathbb{E}\Big[e^{\lambda Z'}\sum_{k=1}^{m}(Z' - Z'_k)\Big] \leq \Big(\lambda - \frac{1}{8}\Big)\mathbb{E}\big[e^{\lambda Z'}Z'\big],$$

which after rearrangement yields

$$\mathbb{E}\big[e^{\lambda Z'}Z'\big] \leq 8\,\mathbb{E}\,e^{\lambda Z'}\log\mathbb{E}\,e^{\lambda Z'},$$

which in turn is equivalent to

$$\frac{d}{d\lambda}\big(\log\mathbb{E}\,e^{\lambda Z'}\big) \leq 8\log\mathbb{E}\,e^{\lambda Z'}$$

for any $\lambda \geq 1/4$. Integrating w.r.t. λ yields that

$$\log \mathbb{E}\, e^{\lambda Z'} \leq e^{8\lambda - 2} \log \mathbb{E}\, e^{Z'/4}. \tag{B.2.4}$$

We turn to estimating the term $\log \mathbb{E}\, e^{Z'/4}$. Using again that $0 \leq Z' - Z'_k \leq 1$, we obtain that

$$\sum_{k=1}^{m} (Z' - Z'_k)^2 \leq Z'.$$

Moreover, by comparing the derivatives, we get that for any $z \geq 0$,

$$\phi(-z) \leq \frac{z^2}{2}$$

and thus we can also estimate further (B.2.3) as

$$\mathrm{Ent}(e^{\lambda Z'}) \leq \frac{\lambda^2}{2} \mathbb{E}\Big[e^{\lambda Z'} \sum_{k=1}^{m} (Z' - Z'_k)^2 \Big] \leq \frac{\lambda^2}{2} \mathbb{E}\big[e^{\lambda Z'} Z' \big].$$

Applying Proposition B.3.2 with $a = \frac{1}{2}$ and $b = 0$ yields that

$$\forall\, \lambda \geq 0 \qquad \Big(1 - \frac{\lambda}{2}\Big) \log \mathbb{E}\, e^{\lambda Z'} \leq \lambda\, \mathbb{E}\, Z'$$

so that

$$\forall\, \lambda \in [0, 1/4] \qquad \log \mathbb{E}\, e^{\lambda Z'} \leq \frac{8}{7} \lambda\, \mathbb{E}\, Z',$$

which combined with (B.2.4) yields

$$\log \mathbb{E}\, e^{\lambda Z'} \leq \frac{2}{7e^2} e^{8\lambda}\, \mathbb{E}\, Z' \leq \frac{1}{16} e^{8\lambda}\, \mathbb{E}\, Z'$$

as desired. □

B.3 Variants of the Herbst argument

Throughout this section, X is a random variable such that its Laplace transform F is well defined on $[0, \infty)$. In that case, recall that

$$\mathrm{Ent}(e^{\lambda X}) = \lambda F'(\lambda) - F(\lambda) \log F(\lambda)$$

for all $\lambda \geq 0$. Below we gather some variants of the celebrated Herbst argument.

Proposition B.3.1. *If for any $\lambda \geq 0$,*

$$\lambda F'(\lambda) - F(\lambda) \log F(\lambda) \leq a \lambda^2 e^{b\lambda} F(\lambda) \tag{B.3.1}$$

for some $a, b > 0$, then

$$\forall\, \lambda \geq 0 \qquad \log \mathbb{E}\, e^{\lambda (X - \mathbb{E}\, X)} \leq \frac{a}{b} \lambda (e^{b\lambda} - 1) \tag{B.3.2}$$

and in particular

$$\forall\, t \geq 0 \qquad \mathbb{P}\big(X \geq \mathbb{E}\, X + t\big) \leq \exp\Big(-\frac{t}{2b} \log\Big(1 + \frac{b}{2a} t\Big)\Big). \tag{B.3.3}$$

Proof. Set $H(\lambda) = \frac{\log F(\lambda)}{\lambda}$ for $\lambda > 0$. Then, (B.3.1) implies $H'(\lambda) \le ae^{b\lambda}$. Since $H(0^+) = \mathbb{E}\, X$, then for any $\lambda > 0$,

$$H(\lambda) \le \mathbb{E}\, X + \frac{a}{b}(e^{b\lambda} - 1),$$

which translates to (B.3.2) and consequently, by the Chernoff bound

$$\mathbb{P}\big(X \ge \mathbb{E}\, X + t\big) \le \inf_{\lambda>0} \exp\Big(-\lambda t + \frac{a}{b}\lambda(e^{b\lambda} - 1)\Big)$$

for all $t \ge 0$. Choosing $\lambda = \frac{1}{b}\log(1 + \frac{b}{2a}t)$ yields (B.3.3). □

Proposition B.3.2. *Assume that for all $\lambda \ge 0$,*

$$\lambda F'(\lambda) - F(\lambda)\log F(\lambda) \le \lambda^2\big(aF'(\lambda) + bF(\lambda)\big) \tag{B.3.4}$$

for some $a, b \in \mathbb{R}$. Then

$$\forall\, \lambda \ge 0 \qquad (1 - a\lambda)\log \mathbb{E}\, e^{\lambda X} \le \lambda\, \mathbb{E}\, X + b\lambda^2. \tag{B.3.5}$$

If additionally $a > 0$ and X is not constant, then $a\,\mathbb{E}\, X + b > 0$ and

$$\forall\, t \ge 0 \qquad \mathbb{P}\big(X \ge \mathbb{E}\, X + t\big) \le \exp\Big(-\min\Big(\frac{t}{4a}, \frac{t^2}{8(a\,\mathbb{E}\, X + b)}\Big)\Big). \tag{B.3.6}$$

Proof. Set $H(\lambda) = \frac{\log F(\lambda)}{\lambda}$ for $\lambda > 0$. Then, (B.3.4) implies

$$H'(\lambda) \le a\frac{F'(\lambda)}{F(\lambda)} + b = \frac{d}{d\lambda}\big(a\log F(\lambda) + b\lambda\big).$$

Consequently, for any $\lambda > 0$,

$$H(\lambda) \le H(0^+) + a\log F(\lambda) + b\lambda,$$

which is equivalent to (B.3.5) since $H(0^+) = \mathbb{E}\, X$. Subtracting $(1 - a\lambda)\lambda\,\mathbb{E}\, X$ from both sides gives

$$(1 - a\lambda)\log \mathbb{E}\, e^{\lambda(X - \mathbb{E}\, X)} \le \lambda^2(a\,\mathbb{E}\, X + b). \tag{B.3.7}$$

By Jensen's inequality and the fact that X is not constant, $\log \mathbb{E}\, e^{\lambda(X-\mathbb{E}\, X)} > 0$. If $\lambda \le 1/2a$, then $1/2 \le 1 - a\lambda$, whence (B.3.7) implies

$$\forall\, \lambda \in [0, 1/2a] \qquad 0 < \log \mathbb{E}\, e^{\lambda(X-\mathbb{E}\, X)} \le 2\lambda^2(a\,\mathbb{E}\, X + b).$$

Therefore, by the Chernoff bound

$$\mathbb{P}\big(X \ge \mathbb{E}\, X + t\big) \le \inf_{0\le\lambda\le 1/2a} \exp\Big(-\lambda t + 2\lambda^2(a\,\mathbb{E}\, X + b)\Big)$$

for all $t \ge 0$. Choosing $\lambda = \frac{t}{4(a\,\mathbb{E}\, X+b)}$ if $t \le \frac{2(a\,\mathbb{E}\, X+b)}{a}$ and $\lambda = \frac{1}{2a}$ otherwise yields (B.3.6). □

Proposition B.3.3. *Assume that for some $\varepsilon, b > 0$ and all $\lambda \in [0, \varepsilon]$,*

$$\lambda F'(\lambda) - F(\lambda)\log F(\lambda) \le b\lambda^2 F(\lambda). \tag{B.3.8}$$

Then

$$\forall\, t \ge 0 \qquad \mathbb{P}\big(X \ge \mathbb{E}\, X + t\big) \le \exp\Big(-\min\Big(\frac{\varepsilon t}{2}, \frac{t^2}{4b}\Big)\Big). \tag{B.3.9}$$

Proof. Dividing (B.3.8) by $\lambda^2 F(\lambda)$ and integrating w.r.t. λ yields

$$\frac{\log \mathbb{E}\, e^{\lambda X}}{\lambda} \le \mathbb{E}\, X + \lambda b$$

for all $\lambda \in [0, \varepsilon]$. Therefore, by the Chernoff bound

$$\mathbb{P}\big(X \ge \mathbb{E}\, X + t\big) \le \inf_{0\le\lambda\le\varepsilon} \exp\big(-\lambda t + b\lambda^2\big)$$

for all $t \ge 0$. Choosing $\lambda = \frac{t}{2b}$ if $t \le 2b\varepsilon$ and $\lambda = \varepsilon$ otherwise yields (B.3.9). □