Concentration of Measure Inequalities in Information Theory, Communications, and Coding

Second Edition

Maxim Raginsky
Department of Electrical and Computer Engineering
Coordinated Science Laboratory
University of Illinois at Urbana-Champaign
Urbana, IL 61801, USA
maxim@illinois.edu

Igal Sason
Department of Electrical Engineering
Technion – Israel Institute of Technology
Haifa 32000, Israel
sason@ee.technion.ac.il

now
the essence of knowledge
Boston — Delft

Foundations and Trends® in Communications and Information Theory

Published, sold and distributed by:
now Publishers Inc.
PO Box 1024
Hanover, MA 02339
United States
Tel. +1-781-985-4510
www.nowpublishers.com
sales@nowpublishers.com

Outside North America:
now Publishers Inc.
PO Box 179
2600 AD Delft
The Netherlands
Tel. +31-6-51115274

The preferred citation for this publication is

M. Raginsky and I. Sason. *Concentration of Measure Inequalities in Information Theory, Communications, and Coding: Second Edition*. Foundations and Trends® in Communications and Information Theory, 2014.

This Foundations and Trends® issue was typeset in LaTeX using a class file designed by Neal Parikh. Printed on acid-free paper.

ISBN: 978-1-60198-907-9

Contents

Abstract

During the last two decades, concentration inequalities have been the subject of exciting developments in various areas, including convex geometry, functional analysis, statistical physics, high-dimensional statistics, pure and applied probability theory (e.g., concentration of measure phenomena in random graphs, random matrices, and percolation), information theory, theoretical computer science, and learning theory. This monograph focuses on some of the key modern mathematical tools that are used for the derivation of concentration inequalities, on their links to information theory, and on their various applications to communications and coding. In addition to being a survey, this monograph also includes various new recent results derived by the authors.

The first part of the monograph introduces classical concentration inequalities for martingales, as well as some recent refinements and extensions. The power and versatility of the martingale approach is exemplified in the context of codes defined on graphs and iterative decoding algorithms, as well as codes for wireless communication.

The second part of the monograph introduces the entropy method, an information-theoretic technique for deriving concentration inequalities. The basic ingredients of the entropy method are discussed first in the context of logarithmic Sobolev inequalities, which underlie the so-called functional approach to concentration of measure, and then from a complementary information-theoretic viewpoint based on transportation-cost inequalities and probability in metric spaces. Some representative results on concentration for dependent random variables are briefly summarized, with emphasis on their connections to the entropy method. Finally, we discuss several applications of the entropy method to problems in communications and coding, including strong converses, empirical distributions of good channel codes, and an information-theoretic converse for concentration of measure.

1

Introduction

1.1 An overview and a brief history

Concentration-of-measure inequalities provide bounds on the probability that a random variable X deviates from its mean, median or other typical value $\overline{x}$ by a given amount. These inequalities have been studied for several decades, with some fundamental and substantial contributions during the last two decades. Very roughly speaking, the concentration of measure phenomenon can be stated in the following simple way: "A random variable that depends in a smooth way on many independent random variables (but not too much on any of them) is essentially constant" [1]. The exact meaning of such a statement clearly needs to be clarified rigorously, but it often means that such a random variable X concentrates around $\overline{x}$ in a way that the probability of the event $\{|X - \overline{x}| \geq t\}$, for a given $t > 0$, decays exponentially in t. Detailed treatments of the concentration of measure phenomenon, including historical accounts, can be found, e.g., in [2, 3, 4, 5, 6, 7].

In recent years, concentration inequalities have been intensively studied and used as a powerful tool in various areas. These include convex geometry, functional analysis, statistical physics, dynamical systems, probability (random matrices, Markov processes, random graphs, percolation etc.), statistics, information theory, coding theory, learning theory, and theoretical computer science. Several techniques have been developed so far to prove concentration of measure inequalities. These include:

- The martingale approach (see, e.g., [6, 8, 9], [10, Chapter 7], [11, 12]), and its information-theoretic applications (see, e.g., [13] and references therein, [14]). This methodology will be covered in Chapter 2, which is focused on concentration inequalities for discrete-time martingales with bounded differences, as well as on some of their potential applications in information theory, coding and communications. A recent interesting avenue that follows from the martingale-based concentration inequalities which are introduced in Chapter 2 refers to their generalization to random matrices (see, e.g., [15, 16]).
- The entropy method and logarithmic Sobolev inequalities (see, e.g., [3, Chapter 5], [4] and references therein). This methodology and its many remarkable links to information theory will be considered in Chapter 3.
- Transportation-cost inequalities that originated from information theory (see, e.g., [3, Chapter 6], [17], and references therein). This methodology, which is closely related to the entropy method and log-Sobolev inequalities, will be considered in Chapter 3.
- Talagrand's inequalities for product measures (see, e.g., [1], [6, Chapter 4], [7] and [18, Chapter 6]) and their links to information theory [19]. These inequalities proved to be very useful in combinatorial applications (such as the study of common and/or increasing subsequences), in statistical physics, and in functional analysis. We do not discuss Talagrand's inequalities in detail.
- Stein's method (or the method of exchangeable pairs) was recently used to prove concentration inequalities (see, e.g., [20, 21, 22, 23, 24, 25, 26, 27, 28]).
- Concentration inequalities that follow from rigorous methods in statistical physics (see, e.g., [29, 30, 31, 32, 33, 34, 35, 36]).

- The so-called reverse Lyapunov inequalities were recently used to derive concentration inequalities for multi-dimensional log-concave distributions [37] (see also a related work in [38]). The concentration inequalities in [37] imply an extension of the Shannon–McMillan–Breiman strong ergodic theorem to the class of discrete-time processes with log-concave marginals.

The last three items are not addressed in this monograph.

We now give a synopsis of some of the main ideas underlying the martingale approach (Chapter 2) and the entropy method (Chapter 3).

The Azuma–Hoeffding inequality, as is introduced in Chapter 2, is by now a well-known tool to establish concentration results for *discrete-time bounded-difference martingales*. It is due to Hoeffding [9], who proved this inequality for a sum of independent and bounded random variables, and to Azuma [8], who later extended it to bounded-difference martingales. This inequality was introduced into the computer science literature by Shamir and Spencer [39], who used it to prove concentration of the chromatic number for random graphs around its expected value (the chromatic number of a graph is defined as the minimal number of colors required to color all the vertices of this graph such that no two adjacent vertices have the same color). Shamir and Spencer [39] established concentration of the chromatic number for the so-called *Erdös–Rényi* ensemble of random graphs, where an arbitrary pair of vertices is connected by an edge with probability $p \in (0, 1)$, independently of all other edges. Note that the concentration result in [39] was established without knowing the expected value of the chromatic number over this ensemble. This approach has been imported into coding theory in [40], [41] and [42], especially for exploring concentration of measure phenomena pertaining to codes defined on graphs and iterative message-passing decoding algorithms. The last decade has seen an ever-expanding use of the Azuma–Hoeffding inequality for proving concentration inequalities in coding theory (see, e.g., [13] and references therein). All these concentration inequalities serve in general to justify theoretically the ensemble approach to codes defined on graphs; nevertheless, much stronger concentration of measure phenomena are observed in practice.

Let $f \colon \mathbb{R}^n \to \mathbb{R}$ be a function that has *bounded differences*, i.e., the value of f changes by a bounded amount whenever any of its n input variables is changed arbitrarily while others are held fixed. A common method for proving concentration of such a function of n independent random variables around its expected value $\mathbb{E}[f]$ revolves around the so-called McDiarmid's inequality or the "independent bounded-differences inequality" [6]. This inequality, as is introduced in Chapter 2, was originally proved via the martingale approach [6]. Although the proof of McDiarmid's inequality has some similarity to the proof of the Azuma–Hoeffding inequality, the bounded-difference assumption on f that is used for the derivation of the former inequality yields an improvement in the exponent by a factor of 4. Nice applications of martingale-based concentration inequalities in discrete mathematics and random graphs, based on the Azuma–Hoeffding and McDiarmid inequalities, are exemplified in [6, Section 3], [10, Chapter 7], [13] and [18, Chapters 1, 2].

In spite of the large variety of problems where concentration of measure phenomena can be asserted via the martingale approach, as pointed out by Talagrand [1], "for all its qualities, the martingale method has a great drawback: it does not seem to yield results of optimal order in several key situations. In particular, it seems unable to obtain even a weak version of concentration of measure phenomenon in Gaussian space." In Chapter 3 of this monograph, we focus on another set of techniques, fundamentally rooted in information theory, that provide very strong concentration inequalities. These powerful techniques, commonly referred to as the *entropy method*, have originated in the work of Michel Ledoux [43], who found an alternative route to a class of concentration inequalities for product measures originally derived by Talagrand [7] using an ingenious inductive technique. Specifically, Ledoux noticed that the well-known Chernoff bounding technique, which bounds the deviation probability of the form $\mathbb{P}(|X - \bar{x}| > t)$, for an arbitrary $t > 0$, in terms of the moment-generating function (MGF) $\mathbb{E}[\exp(\lambda X)]$, can be combined with the so-called *logarithmic Sobolev inequalities*, which can be used to control the MGF in terms of the relative entropy.

Perhaps the best-known log-Sobolev inequality, first explicitly referred to as such by Leonard Gross [44], pertains to the standard Gaussian distribution in Euclidean space $\mathbb{R}^n$, and bounds the relative entropy $D(P\|G_n)$ between an arbitrary probability distribution P on $\mathbb{R}^n$ and the standard Gaussian measure G_n by an "energy-like" quantity related to the squared norm of the gradient of the density of P w.r.t. G_n. By a clever analytic argument which he attributed to an unpublished note by Ira Herbst, Gross has used his log-Sobolev inequality to show that the logarithmic MGF $\Lambda(\lambda) = \ln \mathbb{E}[\exp(\lambda U)]$ of $U = f(X^n)$, where $X^n \sim G_n$ and $f \colon \mathbb{R}^n \to \mathbb{R}$ is an arbitrary sufficiently smooth function with $\|\nabla f\| \le 1$, can be bounded as $\Lambda(\lambda) \le \lambda^2/2$. This bound then yields the optimal Gaussian concentration inequality $\mathbb{P}\left(|f(X^n) - \mathbb{E}[f(X^n)]| > t\right) \le 2\exp\left(-t^2/2\right)$ for $X^n \sim G_n$ and $t > 0$. (It should be pointed out that the Gaussian log-Sobolev inequality has a curious history, and it seems to have been discovered independently in various equivalent forms by several people, e.g., by Stam [45] in the context of information theory, and by Federbush [46] in the context of mathematical quantum field theory. Through the work of Stam [45], the Gaussian log-Sobolev inequality has been linked to several other information-theoretic notions, such as the concavity of entropy power [47, 48, 49, 50].)

In a nutshell, the entropy method takes this idea and applies it beyond the Gaussian case. In abstract terms, log-Sobolev inequalities are functional inequalities that relate the relative entropy between an arbitrary distribution Q w.r.t. the distribution P of interest to some "energy functional" of the density $f = \mathrm{d}Q/\mathrm{d}P$. If one is interested in studying concentration properties of some function $U = f(Z)$ with $Z \sim P$, the core of the entropy method consists in applying an appropriate log-Sobolev inequality to the *tilted distributions* $P^{(\lambda f)}$ with $\mathrm{d}P^{(\lambda f)}/\mathrm{d}P \propto \exp(\lambda f)$. Provided the function f is well-behaved in the sense of having bounded "energy," one can use the Herbst argument to pass from the log-Sobolev inequality to the bound $\ln \mathbb{E}[\exp(\lambda U)] \le c\lambda^2/(2C)$, where $c > 0$ depends only on the distribution P, while $C > 0$ is determined by the energy content of f. While there is no general technique for deriving log-Sobolev inequalities, there are nevertheless some underlying principles that can be exploited for that purpose. We discuss some of these principles in Chapter 3. More information on log-Sobolev inequalities can be found in several excellent monographs and lecture notes [3, 5, 51, 52, 53], as well as in recent papers [54, 55, 56, 57, 58] and references therein.

Around the same time that Michel Ledoux first introduced the entropy method [43], Katalin Marton showed in a breakthrough paper [59] that one can bypass functional inequalities and work directly on the level of probability measures (see also the survey paper [60], presented at the 2013 Shannon Award Lecture). More specifically, Marton has shown that Gaussian concentration bounds can be deduced from the so-called *transportation-cost inequalities*. These inequalities, discussed in detail in Section 3.4, relate information-theoretic quantities, such as the relative entropy, to a certain class of distances between probability measures on the metric space where the random variables of interest are defined. These so-called *Wasserstein distances* have been the subject of intense research activity that touches upon probability theory, functional analysis, dynamical systems, partial differential equations, statistical physics, and differential geometry. A great deal of information on this field of *optimal transportation* can be found in two books by Cédric Villani — [61] offers a concise and fairly elementary introduction, while a more recent monograph [62] is a lot more detailed and encyclopedic. Multiple connections between optimal transportation, concentration of measure, and information theory are also explored in [17, 19, 63, 64, 65, 66, 67]. Note that Wasserstein distances have been also used in information theory in the context of lossy source coding [68, 69, 70].

The first explicit invocation of concentration inequalities in an information-theoretic context appears in the work of Ahlswede *et al.* [71, 72]. These authors have shown that a certain delicate probabilistic inequality, which was referred to as the "blowing up lemma", and which we now (thanks to the contributions by Marton [59, 73]) recognize as a Gaussian concentration bound in the Hamming space, can be used to derive strong converses for a wide variety of information-theoretic problems, including multi-terminal scenarios. The importance of sharp concentration inequalities for characterizing fundamental limits of coding schemes in information theory is evident from the recent flurry of activity on *finite-blocklength* analysis of source and channel codes (see, e.g., [74, 75, 76, 77, 78, 79, 80, 81]). Thus, it is timely to revisit the use of concentration-of-measure ideas in information theory from a modern perspective. We hope that our treatment, which, above all, aims to distill the core information-theoretic ideas underlying the study of concentration of measure, will be helpful to researchers in information theory and related fields.

1.2 A reader's guide

This monograph is mainly focused on the interplay between concentration of measure and information theory, as well as applications to problems related to information theory, communications and coding. For this reason, it is primarily aimed at researchers and graduate students working in these fields. The necessary mathematical background is real analysis, elementary functional analysis, and a first graduate course in probability theory and stochastic processes. As a refresher textbook for this mathematical background, the reader is referred, e.g., to [82].

Chapter 2 on the martingale approach is structured as follows: Section 2.1 lists key definitions pertaining to discrete-time martingales, while Section 2.2 presents several basic inequalities (including the celebrated Azuma–Hoeffding and McDiarmid inequalities) that form the basis of the martingale approach to concentration of measure. Section 2.3 focuses on several refined versions of the Azuma–Hoeffding inequality under additional moment conditions. Section 2.4 discusses the connections of the concentration inequalities introduced in Section 2.3 to classical limit theorems of probability theory, including the central limit theorem for martingales, the moderate deviations principle for i.i.d. real-valued random variables, and the suitability of the concentration inequalities derived in Chapter 2 for some structured functions of discrete-time Markov chains. Section 2.5 forms the second part of Chapter 2, applying the concentration inequalities from Sections 2.2 and 2.3 to information theory and some related topics. Section 2.6 concludes with a summary of the chapter.

References

[1] M. Talagrand. A new look at independence. *Annals of Probability*, 24(1):1–34, January 1996.

[2] S. Boucheron, G. Lugosi, and P. Massart. *Concentration Inequalities - A Nonasymptotic Theory of Independence*. Oxford University Press, 2013.

[3] M. Ledoux. *The Concentration of Measure Phenomenon*, volume 89 of *Mathematical Surveys and Monographs*. American Mathematical Society, 2001.

[4] G. Lugosi. Concentration of measure inequalities - lecture notes, 2009.

[5] P. Massart. *The Concentration of Measure Phenomenon*, volume 1896 of *Lecture Notes in Mathematics*. Springer, 2007.

[6] C. McDiarmid. Concentration. In *Probabilistic Methods for Algorithmic Discrete Mathematics*, pages 195–248. Springer, 1998.

[7] M. Talagrand. Concentration of measure and isoperimteric inequalities in product space. *Publications Mathématiques de l'I.H.E.S*, 81:73–205, 1995.

[8] K. Azuma. Weighted sums of certain dependent random variables. *Tohoku Mathematical Journal*, 19:357–367, 1967.

[9] W. Hoeffding. Probability inequalities for sums of bounded random variables. *Journal of the American Statistical Association*, 58(301):13–30, March 1963.

[10] N. Alon and J. H. Spencer. *The Probabilistic Method*. Wiley Series in Discrete Mathematics and Optimization, third edition, 2008.

[11] F. Chung and L. Lu. *Complex Graphs and Networks*, volume 107 of *Regional Conference Series in Mathematics*. Wiley, 2006.

[12] F. Chung and L. Lu. Concentration inequalities and martingale inequalities: a survey. *Internet Mathematics*, 3(1):79–127, March 2006.

[13] T. J. Richardson and R. Urbanke. *Modern Coding Theory*. Cambridge University Press, 2008.

[14] Y. Seldin, F. Laviolette, N. Cesa-Bianchi, J. Shawe-Taylor, and P. Auer. PAC-Bayesian inequalities for martingales. *IEEE Trans. on Information Theory*, 58(12):7086–7093, December 2012.

[15] J. A. Tropp. User-friendly tail bounds for sums of random matrices. *Foundations of Computational Mathematics*, 12(4):389–434, August 2012.

[16] J. A. Tropp. Freedman's inequality for matrix martingales. *Electronic Communications in Probability*, 16:262–270, March 2011.

[17] N. Gozlan and C. Leonard. Transport inequalities: a survey. *Markov Processes and Related Fields*, 16(4):635–736, 2010.

[18] J. M. Steele. *Probability Theory and Combinatorial Optimization*, volume 69 of *CBMS–NSF Regional Conference Series in Applied Mathematics*. Siam, Philadelphia, PA, USA, 1997.

[19] A. Dembo. Information inequalities and concentration of measure. *Annals of Probability*, 25(2):927–939, 1997.

[20] S. Chatterjee. *Concentration Inequalities with Exchangeable Pairs*. PhD thesis, Stanford University, California, USA, February 2008.

[21] S. Chatterjee. Stein's method for concentration inequalities. *Probability Theory and Related Fields*, 138:305–321, 2007.

[22] S. Chatterjee and P. S. Dey. Applications of Stein's method for concentration inequalities. *Annals of Probability*, 38(6):2443–2485, June 2010.

[23] N. Ross. Fundamentals of Stein's method. *Probability Surveys*, 8:210–293, 2011.

[24] S. Ghosh and L. Goldstein. Concentration of measure via size-bias coupling. *Probability Theory and Related Fields*, 149:271–278, February 2011.

[25] S. Ghosh and L. Goldstein. Applications of size-biased couplings for concentration of measures. *Electronic Communications in Probability*, 16:70–83, January 2011.

[26] L. Goldstein and U. Işlak. Concentration inequalities via zero bias coupling. *Statistics and Probability Letters*, 86:17–23, January 2014.

[27] L. Mackey, M. I. Jordan, R. Y. Chen, B. Farrell, and J. A. Tropp. Matrix concentration inequalities via the method of exchangeable pairs. *Annals of Probability*, 10(2):906–945, 2014.

[28] D. Paulin. The convex distance inequality for dependent random variables, with applications to the stochastic travelling salesman and other problems. *Electronic Journal of Probability*, 19(68):1–34, August 2014.

[29] E. Abbe and A. Montanari. On the concentration of the number of solutions of random satisfiability formulas. *Random Structures and Algorithms*, 45(3):362–382, April 2013.

[30] S. B. Korada and N. Macris. On the concentration of the capacity for a code division multiple access system. In *Proceedings of the 2007 IEEE International Symposium on Information Theory*, pages 2801–2805, Nice, France, June 2007.

[31] S. B. Korada, S. Kudekar, and N. Macris. Concentration of magnetization for linear block codes. In *Proceedings of the 2008 IEEE International Symposium on Information Theory*, pages 1433–1437, Toronto, Canada, July 2008.

[32] S. Kudekar. *Statistical Physics Methods for Sparse Graph Codes*. PhD thesis, EPFL - Swiss Federal Institute of Technology, Lausanne, Switzeland, July 2009.

[33] S. Kudekar and N. Macris. Sharp bounds for optimal decoding of low-density parity-check codes. *IEEE Trans. on Information Theory*, 55(10):4635–4650, October 2009.

[34] S. B. Korada and N. Macris. Tight bounds on the capacity of binary input random CDMA systems. *IEEE Trans. on Information Theory*, 56(11):5590–5613, November 2010.

[35] A. Montanari. Tight bounds for LDPC and LDGM codes under MAP decoding. *IEEE Trans. on Information Theory*, 51(9):3247–3261, September 2005.

[36] M. Talagrand. *Mean Field Models for Spin Glasses*. Springer-Verlag, 2010.

[37] S. Bobkov and M. Madiman. Concentration of the information in data with log-concave distributions. *Annals of Probability*, 39(4):1528–1543, 2011.

[38] S. Bobkov and M. Madiman. The entropy per coordinate of a random vector is highly constrained under convexity conditions. *IEEE Trans. on Information Theory*, 57(8):4940–4954, August 2011.

[39] E. Shamir and J. Spencer. Sharp concentration of the chromatic number on random graphs. *Combinatorica*, 7(1):121–129, 1987.

[40] M. G. Luby, Mitzenmacher, M. A. Shokrollahi, and D. A. Spielmann. Efficient erasure-correcting codes. *IEEE Trans. on Information Theory*, 47(2):569–584, February 2001.

[41] T. J. Richardson and R. Urbanke. The capacity of low-density parity-check codes under message-passing decoding. *IEEE Trans. on Information Theory*, 47(2):599–618, February 2001.

[42] M. Sipser and D. A. Spielman. Expander codes. *IEEE Trans. on Information Theory*, 42(6):1710–1722, November 1996.

[43] M. Ledoux. On Talagrand's deviation inequalities for product measures. *ESAIM: Probability and Statistics*, 1:63–87, 1997.

[44] L. Gross. Logarithmic Sobolev inequalities. *American Journal of Mathematics*, 97(4):1061–1083, 1975.

[45] A. J. Stam. Some inequalities satisfied by the quantities of information of Fisher and Shannon. *Information and Control*, 2:101–112, 1959.

[46] P. Federbush. A partially alternate derivation of a result of Nelson. *Journal of Mathematical Physics*, 10(1):50–52, 1969.

[47] M. H. M. Costa. A new entropy power inequality. *IEEE Trans. on Information Theory*, 31(6):751–760, November 1985.

[48] A. Dembo, T. M. Cover, and J. A. Thomas. Information theoretic inequalities. *IEEE Trans. on Information Theory*, 37(6):1501–1518, November 1991.

[49] C. Villani. A short proof of the 'concavity of entropy power'. *IEEE Trans. on Information Theory*, 46(4):1695–1696, July 2000.

[50] G. Toscani. An information-theoretic proof of Nash's inequality. *Rendiconti Lincei: Matematica e Applicazioni*, 24(1):83–93, 2013.

[51] A. Guionnet and B. Zegarlinski. Lectures on logarithmic Sobolev inequalities. *Séminaire de probabilités (Strasbourg)*, 36:1–134, 2002.

[52] M. Ledoux. Concentration of measure and logarithmic Sobolev inequalities. In *Séminaire de Probabilités XXXIII*, volume 1709 of *Lecture Notes in Math.*, pages 120–216. Springer, 1999.

[53] G. Royer. *An Invitation to Logarithmic Sobolev Inequalities*, volume 14 of *SFM/AMS Texts and Monographs.* American Mathematical Society and Société Mathématiques de France, 2007.

[54] S. G. Bobkov and F. Götze. Exponential integrability and transportation cost related to logarithmic Sobolev inequalities. *Journal of Functional Analysis*, 163:1–28, 1999.

[55] S. G. Bobkov and M. Ledoux. On modified logarithmic Sobolev inequalities for Bernoulli and Poisson measures. *Journal of Functional Analysis*, 156(2):347–365, 1998.

[56] S. G. Bobkov and P. Tetali. Modified logarithmic Sobolev inequalities in discrete settings. *Journal of Theoretical Probability*, 19(2):289–336, 2006.

[57] D. Chafaï. Entropies, convexity, and functional inequalities: Φ-entropies and Φ-Sobolev inequalities. *J. Math. Kyoto University*, 44(2):325–363, 2004.

[58] C. P. Kitsos and N. K. Tavoularis. Logarithmic Sobolev inequalities for information measures. *IEEE Trans. on Information Theory*, 55(6):2554–2561, June 2009.

[59] K. Marton. Bounding $\bar{d}$-distance by informational divergence: a method to prove measure concentration. *Annals of Probability*, 24(2):857–866, 1996.

[60] K. Marton. Distance-divergence inequalities. *IEEE Information Theory Society Newsletter*, 64(1):9–13, March 2014.

[61] C. Villani. *Topics in Optimal Transportation.* American Mathematical Society, Providence, RI, 2003.

[62] C. Villani. *Optimal Transport: Old and New.* Springer, 2008.

[63] P. Cattiaux and A. Guillin. On quadratic transportation cost inequalities. *Journal de Matématiques Pures et Appliquées*, 86:342–361, 2006.

[64] A. Dembo and O. Zeitouni. Transportation approach to some concentration inequalities in product spaces. *Electronic Communications in Probability*, 1:83–90, 1996.

[65] H. Djellout, A. Guillin, and L. Wu. Transportation cost-information inequalities and applications to random dynamical systems and diffusions. *Annals of Probability*, 32(3B):2702–2732, 2004.

[66] N. Gozlan. A characterization of dimension free concentration in terms of transportation inequalities. *Annals of Probability*, 37(6):2480–2498, 2009.

[67] E. Milman. Properties of isoperimetric, functional and transport-entropy inequalities via concentration. *Probability Theory and Related Fields*, 152:475–507, 2012.

[68] R. M. Gray, D. L. Neuhoff, and P. C. Shields. A generalization of Ornstein's $\bar{d}$ distance with applications to information theory. *Annals of Probability*, 3(2):315–328, 1975.

[69] R. M. Gray, D. L. Neuhoff, and J. K. Omura. Process definitions of distortion-rate functions and source coding theorems. *IEEE Trans. on Information Theory*, 21(5):524–532, September 1975.

[70] Y. Steinberg and S. Verdú. Simulation of random processes and rate-distortion theory. *IEEE Trans. on Information Theory*, 42(1):63–86, January 1996.

[71] R. Ahlswede, P. Gács, and J. Körner. Bounds on conditional probabilities with applications in multi-user communication. *Z. Wahrscheinlichkeitstheorie verw. Gebiete*, 34:157–177, 1976. See correction in vol. 39, no. 4, pp. 353–354, 1977.

[72] R. Ahlswede and G. Dueck. Every bad code has a good subcode: a local converse to the coding theorem. *Z. Wahrscheinlichkeitstheorie verw. Gebiete*, 34:179–182, 1976.

[73] K. Marton. A simple proof of the blowing-up lemma. *IEEE Trans. on Information Theory*, 32(3):445–446, May 1986.

[74] Y. Altuğ and A. B. Wagner. Refinement of the sphere-packing bound: asymmetric channels. *IEEE Trans. on Information Theory*, 60(3):1592–1614, March 2014.

[75] A. Amraoui, A. Montanari, T. Richardson, and R. Urbanke. Finite-length scaling for iteratively decoded LDPC ensembles. *IEEE Trans. on Information Theory*, 55(2):473–498, February 2009.

[76] T. Nozaki, K. Kasai, and K. Sakaniwa. Analytical solution of covariance evolution for irregular LDPC codes. *IEEE Trans. on Information Theory*, 58(7):4770–4780, July 2012.

[77] Y. Kontoyiannis and S. Verdú. Optimal lossless data compression: non-asymptotics and asymptotics. *IEEE Trans. on Information Theory*, 60(2):777–795, February 2014.

[78] V. Kostina and S. Verdú. Fixed-length lossy compression in the finite blocklength regime. *IEEE Trans. on Information Theory*, 58(6):3309–3338, June 2012.

[79] W. Matthews. A linear program for the finite block length converse of Polyanskiy-Poor-Verdú via nonsignaling codes. *IEEE Trans. on Information Theory*, 59(12):7036–7044, December 2012.

[80] Y. Polyanskiy, H. V. Poor, and S. Verdú. Channel coding rate in finite blocklength regime. *IEEE Trans. on Information Theory*, 56(5):2307–2359, May 2010.

[81] G. Wiechman and I. Sason. An improved sphere-packing bound for finite-length codes on symmetric channels. *IEEE Trans. on Information Theory*, 54(5):1962–1990, 2008.

[82] J. S. Rosenthal. *A First Look at Rigorous Probability Theory*. World Scientific, second edition, 2006.

[83] A. Dembo and O. Zeitouni. *Large Deviations Techniques and Applications*. Springer, second edition, 1997.

[84] H. Chernoff. A measure of asymptotic efficiency of tests of a hypothesis based on the sum of observations. *Annals of Mathematical Statistics*, 23(4):493–507, 1952.

[85] S. N. Bernstein. *The Theory of Probability*. Gos. Izdat., Moscow/Leningrad, 1927. in Russian.

[86] S. Verdú. *Multiuser Detection*. Cambridge University Press, 1998.

[87] C. McDiarmid. Centering sequences with bounded differences. *Combinatorics, Probability and Computing*, 6(1):79–86, March 1997.

[88] C. McDiarmid. On the method of bounded differences. In *Surveys in Combinatorics*, volume 141, pages 148–188. Cambridge University Press, 1989.

[89] A. W. van der Vaart and J. A. Wellner. *Weak Convergence and Empirical Processes*. Springer, 1996.

[90] E. Rio. On Mcdiarmid's concentration inequality. *Electronic Communications in Probability*, 18(44):1–11, 2013.

[91] J. Dedecker and X. Fan. Deviation inequalities for separately Lipschitz functionals of iterated random variables. *Stochastic Processes and their Applications*, accepted in August 2014.

[92] M. J. Kearns and L. K. Saul. Large deviation methods for approximate probabilistic inference. In *Proceedings of the 14th Conference on Uncertaintly in Artifical Intelligence*, pages 311–319, San-Francisco, CA, USA, March 16-18 1998.

[93] D. Berend and A. Kontorovich. On the concentration of the missing mass. *Electronic Communications in Probability*, 18(3):1–7, January 2013.

[94] S. G. From and A. W. Swift. A refinement of Hoeffding's inequality. *Journal of Statistical Computation and Simulation*, pages 1–7, December 2011.

[95] X. Fan, I. Grama, and Q. Liu. Hoeffding's inequality for supermartingales. *Stochastic Processes and their Applications*, 122(10):3545–3559, October 2012.

[96] X. Fan, I. Grama, and Q. Liu. Large deviation exponential inequalities for supermartingales. *Electronic Communications in Probability*, 17(59):1–8, December 2012.

[97] P. Billingsley. *Probability and Measure*. Wiley Series in Probability and Mathematical Statistics, 3rd edition, 1995.

[98] G. Grimmett and D. Stirzaker. *Probability and Random Processes*. Oxford University Press, third edition, 2001.

[99] I. Kontoyiannis, L. A. Latras-Montano, and S. P. Meyn. Relative entropy and exponential deviation bounds for general Markov chains. In *Proceedings of the 2005 IEEE International Symposium on Information Theory*, pages 1563–1567, Adelaide, Australia, September 2005.

[100] A. Barg and G. D. Forney. Random codes: minimum distances and error exponents. *IEEE Trans. on Information Theory*, 48(9):2568–2573, September 2002.

[101] M. Breiling. A logarithmic upper bound on the minimum distance of turbo codes. *IEEE Trans. on Information Theory*, 50(8):1692–1710, August 2004.

[102] A. F. Molisch. *Wireless Communications*. John Wiley and Sons, 2005.

[103] G. Wunder, R. F. H. Fischer, H. Boche, S. Litsyn, and J. S. No. The PAPR problem in OFDM transmission: new directions for a long-lasting problem. *IEEE Signal Processing Magazine*, 30(6):130–144, November 2013.

[104] S. Litsyn and G. Wunder. Generalized bounds on the crest-factor istribution of OFDM signals with applications to code design. *IEEE Trans. on Information Theory*, 52(3):992–1006, March 2006.

[105] R. Salem and A. Zygmund. Some properties of trigonometric series whose terms have random signs. *Acta Mathematica*, 91(1):245–301, 1954.

[106] G. Wunder and H. Boche. New results on the statistical distribution of the crest-factor of OFDM signals. *IEEE Trans. on Information Theory*, 49(2):488–494, February 2003.

[107] I. Sason. On the concentration of the crest factor for OFDM signals. In *Proceedings of the 8th International Symposium on Wireless Communication Systems (ISWCS '11)*, pages 784–788, Aachen, Germany, November 2011.

[108] R. G. Gallager. *Low-Density Parity-Check Codes.* PhD thesis, MIT, Cambridge, MA, USA, 1963.

[109] T. Etzion, A. Trachtenberg, and A. Vardy. Which codes have cycle-free Tanner graphs? *IEEE Trans. on Information Theory*, 45(6):2173–2181, September 1999.

[110] I. Sason. On universal properties of capacity-approaching LDPC code ensembles. *IEEE Trans. on Information Theory*, 55(7):2956–2990, July 2009.

[111] I. Sason and R. Eshel. On concentration of measures for LDPC code ensembles. In *Proceedings of the 2011 IEEE International Symposium on Information Theory*, pages 1273–1277, Saint Petersburg, Russia, August 2011.

[112] M. G. Luby, Mitzenmacher, M. A. Shokrollahi, and D. A. Spielmann. Improved low-density parity-check codes using irregular graphs. *IEEE Trans. on Information Theory*, 47(2):585–598, February 2001.

[113] A. Kavčić, X. Ma, and M. Mitzenmacher. Binary intersymbol interference channels: Gallager bounds, density evolution, and code performance bounds. *IEEE Trans. on Information Theory*, 49(7):1636–1652, July 2003.

[114] R. Eshel. *Aspects of Convex Optimization and Concentration in Coding.* Technion - Israel Institute of Technology, Haifa, Israel, February 2012.

[115] J. Douillard, M. Jezequel, C. Berrou, A. Picart, P. Didier, and A. Glavieux. Iterative correction of intersymbol interference: turbo-equalization. *European Transactions on Telecommunications*, 6(1):507–511, September 1995.

[116] C. Méasson, A. Montanari, and R. Urbanke. Maxwell construction: the hidden bridge between iterative and maximum apposteriori decoding. *IEEE Trans. on Information Theory*, 54(12):5277–5307, December 2008.

[117] A. Shokrollahi. Capacity-achieving sequences. In *Volume in Mathematics and its Applications*, volume 123, pages 153–166, 2000.

[118] K. Xenoulis and N. Kalouptsidis. On the random coding exponent of nonlinear Gaussian channels. In *Proceedings of the 2009 IEEE International Workshop on Information Theory*, pages 32–36, Volos, Greece, June 2009.

[119] K. Xenoulis, N. Kalouptsidis, and I. Sason. New achievable rates for nonlinear Volterra channels via martingale inequalities. In *Proceedings of the 2012 IEEE International Workshop on Information Theory*, pages 1430–1434, MIT, Boston, MA, USA, July 2012.

[120] A. P. Godbole and P. Hitczenko. Beyond the method of bounded differences. In *DIMACS Series in Discrete Mathematics and Theoretical Computer Science*, volume 41, pages 43–58. American Mathematical Society, 1998.

[121] E. B. Davies and B. Simon. Ultracontractivity and the heat kernel for Schrödinger operators and Dirichlet Laplacians. *Journal of Functional Analysis*, 59(335-395), 1984.

[122] S. Verdú and T. Weissman. The information lost in erasures. *IEEE Trans. on Information Theory*, 54(11):5030–5058, November 2008.

[123] E. A. Carlen. Superadditivity of Fisher's information and logarithmic Sobolev inequalities. *Journal of Functional Analysis*, 101:194–211, 1991.

[124] R. A. Adams and F. H. Clarke. Gross's logarithmic Sobolev inequality: a simple proof. *American Journal of Mathematics*, 101(6):1265–1269, December 1979.

[125] G. Blower. *Random Matrices: High Dimensional Phenomena*. London Mathematical Society Lecture Notes. Cambridge University Press, Cambridge, U.K., 2009.

[126] O. Johnson. *Information Theory and the Central Limit Theorem*. Imperial College Press, London, 2004.

[127] E. H. Lieb and M. Loss. *Analysis*. American Mathematical Society, Providence, RI, 2nd edition, 2001.

[128] M. H. M. Costa and T. M. Cover. On the similarity of the entropy power inequality and the Brunn–Minkowski inequality. *IEEE Trans. on Information Theory*, 30(6):837–839, November 1984.

[129] P. J. Huber and E. M. Ronchetti. *Robust Statistics*. Wiley Series in Probability and Statistics, second edition, 2009.

[130] O. Johnson and A. Barron. Fisher information inequalities and the central limit theorem. *Probability Theory and Related Fields*, 129:391–409, 2004.

[131] S. Verdú. Mismatched estimation and relative entropy. *IEEE Trans. on Information Theory*, 56(8):3712–3720, August 2010.

[132] H. L. van Trees. *Detection, Estimation and Modulation Theory, Part I.* Wiley, 1968.

[133] L. C. Evans and R. F. Gariepy. *Measure Theory and Fine Properties of Functions.* CRC Press, 1992.

[134] M. C. Mackey. *Time's Arrow: The Origins of Thermodynamic Behavior.* Springer, New York, 1992.

[135] B. Øksendal. *Stochastic Differential Equations: An Introduction with Applications.* Springer, Berlin, 5 edition, 1998.

[136] I. Karatzas and S. Shreve. *Brownian Motion and Stochastic Calculus.* Springer, second edition, 1988.

[137] F. C. Klebaner. *Introduction to Stochastic Calculus with Applications.* Imperial College Press, second edition, 2005.

[138] T. van Erven and P. Harremoës. Rényi divergence and Kullback-Leibler divergence. *IEEE Trans. on Information Theory*, 60(7):3797–3820, July 2014.

[139] T. M. Cover and J. A. Thomas. *Elements of Information Theory.* John Wiley and Sons, second edition, 2006.

[140] A. Maurer. Thermodynamics and concentration. *Bernoulli*, 18(2):434–454, 2012.

[141] N. Merhav. *Statistical Physics and Information Theory*, volume 6 of Foundations and Trends in Communications and Information Theory. Now Publishers, Delft, the Netherlands, 2009.

[142] S. Boucheron, G. Lugosi, and P. Massart. Concentration inequalities using the entropy method. *Annals of Probability*, 31(3):1583–1614, 2003.

[143] I. Kontoyiannis and M. Madiman. Measure concentration for compound Poisson distributions. *Electronic Communications in Probability*, 11:45–57, 2006.

[144] B. Efron and C. Stein. The jackknife estimate of variance. *Annals of Statistics*, 9:586–596, 1981.

[145] J. M. Steele. An Efron–Stein inequality for nonsymmetric statistics. *Annals of Statistics*, 14:753–758, 1986.

[146] M. Gromov. *Metric Structures for Riemannian and Non-Riemannian Spaces.* Birkhäuser, 2001.

[147] S. Bobkov. A functional form of the isoperimetric inequality for the Gaussian measure. *Journal of Functional Analysis*, 135:39–49, 1996.

[148] L. V. Kantorovich. On the translocation of masses. *Journal of Mathematical Sciences*, 133(4):1381–1382, 2006.

[149] E. Ordentlich and M. Weinberger. A distribution dependent refinement of Pinsker's inequality. *IEEE Trans. on Information Theory*, 51(5):1836–1840, May 2005.

[150] T. Weissman, E. Ordentlich, G. Seroussi, S. Verdú, and M. J. Weinberger. Inequalities for the L_1 deviation of the empirical distribution. Technical Report HPL-2003-97 (R.1), Information Theory Research Group, HP Laboratories, Palo Alto, CA, June 2003.

[151] D. Berend, P. Harremoës, and A. Kontorovich. Minimum KL-divergence on complements of L_1 balls. *IEEE Trans. on Information Theory*, 60(6):3172–3177, June 2014.

[152] I. Sason. Improved lower bounds on the total variation distance and relative entropy for the Poisson approximation. In *Proceedings of the 2013 IEEE Information Theory and Applications (ITA) Workshop*, pages 1–4, San-Diego, California, USA, February 2013.

[153] I. Sason. Improved lower bounds on the total variation distance for the Poisson approximation. *Statistics and Probability Letters*, 83(10):2422–2431, October 2013.

[154] I. Kontoyiannis, P. Harremoës, and O. Johnson. Entropy and the law of small numbers. *IEEE Trans. on Information Theory*, 51(2):466–472, February 2005.

[155] I. Csiszár. Sanov property, generalized I-projection and a conditional limit theorem. *Annals of Probability*, 12(3):768–793, 1984.

[156] P. Dupuis and R. S. Ellis. *A Weak Convergence Approach to the Theory of Large Deviations.* Wiley Series in Probability and Statistics, New York, 1997.

[157] M. Talagrand. Transportation cost for Gaussian and other product measures. *Geometry and Functional Analysis*, 6(3):587–600, 1996.

[158] R. M. Dudley. *Real Analysis and Probability.* Cambridge University Press, 2004.

[159] F. Otto and C. Villani. Generalization of an inequality by Talagrand and links with the logarithmic Sobolev inequality. *Journal of Functional Analysis*, 173(2):361–400, June 2000.

[160] Y. Wu. A simple transportation-information inequality with applications to HWI inequalities, and predictive density estimation. Technical Report, September 2011.

[161] D. Cordero-Erausquin. Some applications of mass transport to Gaussian-type inequalities. *Archive for Rational Mechanics and Analysis*, 161(3):257–269, February 2002.

[162] D. Bakry and M. Emery. Diffusions hypercontractives. In *Séminaire de Probabilités XIX*, volume 1123 of *Lecture Notes in Mathematics*, pages 177–206. Springer, 1985.

[163] P.-M. Samson. Concentration of measure inequalities for Markov chains and ϕ-mixing processes. *Annals of Probability*, 28(1):416–461, 2000.

[164] K. Marton. A measure concentration inequality for contracting Markov chains. *Geometric and Functional Analysis*, 6:556–571, 1996. See also erratum in *Geometric and Functional Analysis*, vol. 7, pp. 609–613, 1997.

[165] K. Marton. Measure concentration for Euclidean distance in the case of dependent random variables. *Annals of Probability*, 32(3B):2526–2544, 2004.

[166] K. Marton. Correction to 'Measure concentration for Euclidean distance in the case of dependent random variables'. *Annals of Probability*, 38(1):439–442, 2010.

[167] R. L. Dobrushin and S. B. Shlosman. Completely analytical Gibbs fields. In *Statistical Physics and Dynamical Systems*, pages 371–403. Springer, 1985.

[168] K. Marton. Bounding relative entropy by the relative entropy of local specifications in product spaces, 2009.

[169] K. Marton. An inequality for relative entropy and logarithmic Sobolev inequalities in Euclidean spaces. *Journal of Functional Analysis*, 264(1):34–61, January 2013.

[170] I. Csiszár and J. Körner. *Information Theory: Coding Theorems for Discrete Memoryless Systems*. Cambridge University Press, 2nd edition, 2011.

[171] G. Margulis. Probabilistic characteristics of graphs with large connectivity. *Problems of Information Transmission*, 10(2):174–179, 1974.

[172] A. El Gamal and Y.-H. Kim. *Network Information Theory*. Cambridge University Press, 2011.

[173] G. Dueck. Maximal error capacity regions are smaller than average error capacity regions for multi-user channels. *Problems of Control and Information Theory*, 7(1):11–19, 1978.

[174] F. M. J. Willems. The maximal-error and average-error capacity regions of the broadcast channel are identical: a direct proof. *Problems of Control and Information Theory*, 19(4):339–347, 1990.

[175] T. M. Cover. Broadcast channels. *IEEE Trans. on Information Theory*, 18(1):2–14, January 1972.

[176] P. P. Bergmans. Random coding theorem for broadcast channels with degraded components. *IEEE Trans. on Information Theory*, 19(2):197–207, March 1973.

[177] A. D. Wyner. A theorem on the entropy of certain binary sequences and applications: Part II. *IEEE Trans. on Information Theory*, 19(6):772–777, March 1973.

[178] R. G. Gallager. Capacity and coding for degraded broadcast channels. *Problems of Information Transmission*, 10(3):3–14, July-September 1974.

[179] R. Ahlswede and J. Körner. Source coding with side information and a converse for degraded broadcast channels. *IEEE Trans. on Information Theory*, 21(6):629–637, November 1975.

[180] S. Shamai and S. Verdú. The empirical distribution of good codes. *IEEE Trans. on Information Theory*, 43(3):836–846, May 1997.

[181] T. S. Han and S. Verdú. Approximation theory of output statistics. *IEEE Trans. on Information Theory*, 39(3):752–772, May 1993.

[182] Y. Polyanskiy and S. Verdú. Empirical distribution of good channel codes with non-vanishing error probability. *IEEE Trans. on Information Theory*, 60(1):5–21, January 2014.

[183] M. Raginsky and I. Sason. Refined bounds on the empirical distribution of good channel codes via concentration inequalities. In *Proceedings of the 2013 IEEE International Workshop on Information Theory*, pages 221–225, Istanbul, Turkey, July 2013.

[184] F. Topsøe. An information theoretical identity and a problem involving capacity. *Studia Scientiarum Mathematicarum Hungarica*, 2:291–292, 1967.

[185] J. H. B. Kemperman. On the Shannon capacity of an arbitrary channel. *Indagationes Mathematicae*, 36:101–115, 1974.

[186] U. Augustin. Gedächtnisfreie Kanäle für diskrete Zeit. *Z. Wahrscheinlichkeitstheorie verw. Gebiete*, 6:10–61, 1966.

[187] R. Ahlswede. An elementary proof of the strong converse theorem for the multiple-access channel. *Journal of Combinatorics, Information and System Sciences*, 7(3):216–230, 1982.

[188] S. Shamai and I. Sason. Variations on the Gallager bounds, connections and applications. *IEEE Trans. on Information Theory*, 48(12):3029–3051, December 2001.

[189] Y. Kontoyiannis. Sphere-covering, measure concentration, and source coding. *IEEE Trans. on Information Theory*, 47(4):1544–1552, May 2001.

[190] Y. H. Kim, A. Sutivong, and T. M. Cover. State amplification. *IEEE Trans. on Information Theory*, 54(5):1850–1859, May 2008.