

On the Equivalence of Modes of Convergence for Log-Concave Measures

Elizabeth S. Meckes and Mark W. Meckes

Abstract An important theme in recent work in asymptotic geometric analysis is that many classical implications between different types of geometric or functional inequalities can be reversed in the presence of convexity assumptions. In this note, we explore the extent to which different notions of distance between probability measures are comparable for log-concave distributions. Our results imply that weak convergence of isotropic log-concave distributions is equivalent to convergence in total variation, and is further equivalent to convergence in relative entropy when the limit measure is Gaussian.

1 Introduction and Statements of Results

An important theme in recent work in asymptotic geometric analysis is that many classical implications between different types of geometric or functional inequalities can be reversed in the presence of convexity. A particularly striking recent example is the work of E. Milman [11–13], showing for example that, on a Riemannian manifold equipped with a probability measure satisfying a convexity assumption, the existence of a Cheeger inequality, a Poincaré inequality, and exponential concentration of Lipschitz functions are all equivalent. Important earlier examples of this theme are C. Borell’s 1974 proof of reverse Hölder inequalities for log-concave measures [4], and K. Ball’s 1991 proof of a reverse isoperimetric inequality for convex bodies [1].

In this note, we explore the extent to which different notions of distance between probability measures are comparable in the presence of a convexity assumption. Specifically, we consider **log-concave** probability measures; that is, Borel probability measures μ on \mathbb{R}^n such that for all nonempty compact sets $A, B \subseteq \mathbb{R}^n$ and every $\lambda \in (0, 1)$,

$$\mu(\lambda A + (1 - \lambda)B) \geq \mu(A)^\lambda \mu(B)^{1-\lambda}.$$

E.S. Meckes • M.W. Meckes (✉)

Department of Mathematics, Case Western Reserve University, 10900 Euclid Ave., Cleveland, OH 44106, USA

e-mail: elizabeth.meckes@case.edu; mark.meckes@case.edu

We moreover consider only those log-concave probability measures μ on \mathbb{R}^n which are **isotropic**, meaning that if $X \sim \mu$ then

$$\mathbb{E}X = 0 \quad \text{and} \quad \mathbb{E}XX^T = I_n.$$

The following distances between probability measures μ and ν on \mathbb{R}^n appear below.

1. The **total variation distance** is defined by

$$d_{TV}(\mu, \nu) := 2 \sup_{A \subseteq \mathbb{R}^n} |\mu(A) - \nu(A)|,$$

where the supremum is over Borel measurable sets.

2. The **bounded Lipschitz distance** is defined by

$$d_{BL}(\mu, \nu) := \sup_{\|g\|_{BL} \leq 1} \left| \int g \, d\mu - \int g \, d\nu \right|,$$

where the bounded-Lipschitz norm $\|g\|_{BL}$ of $g : \mathbb{R}^n \rightarrow \mathbb{R}$ is defined by

$$\|g\|_{BL} := \max \left\{ \|g\|_{\infty}, \sup_{x \neq y} \frac{|g(x) - g(y)|}{\|x - y\|} \right\}$$

and $\|\cdot\|$ denotes the standard Euclidean norm on \mathbb{R}^n . The bounded-Lipschitz distance is a metric for the weak topology on probability measures (see, e.g., [6, Theorem 11.3.3]).

3. The L_p **Wasserstein distance** for $p \geq 1$ is defined by

$$W_p(\mu, \nu) := \inf_{\pi} \left[\int \|x - y\|^p \, d\pi(x, y) \right]^{\frac{1}{p}},$$

where the infimum is over couplings π of μ and ν ; that is, probability measures π on \mathbb{R}^{2n} such that $\pi(A \times \mathbb{R}^n) = \mu(A)$ and $\pi(\mathbb{R}^n \times B) = \nu(B)$. The L_p Wasserstein distance is a metric for the topology of weak convergence plus convergence of moments of order p or less. (See [15, Sect. 6] for a proof of this fact, and a lengthy discussion of the many fine mathematicians after whom this distance could reasonably be named.)

4. If μ is absolutely continuous with respect to ν , the **relative entropy**, or **Kullback–Leibler divergence** is defined by

$$H(\mu \mid \nu) := \int \left(\frac{d\mu}{d\nu} \right) \log \left(\frac{d\mu}{d\nu} \right) \, d\nu = \int \log \left(\frac{d\mu}{d\nu} \right) \, d\mu.$$

It is a classical fact that for any probability measures μ and ν on \mathbb{R}^n ,

$$d_{BL}(\mu, \nu) \leq d_{TV}(\mu, \nu). \tag{1}$$

This follows from a dual formulation of total variation distance: the Riesz representation theorem implies that

$$d_{TV}(\mu, \nu) = \sup \left\{ \left| \int g \, d\mu - \int g \, d\nu \right| : g \in C(\mathbb{R}^n), \|g\|_\infty \leq 1 \right\}. \tag{2}$$

In the case that μ and ν are log-concave, there is the following complementary inequality.

Proposition 1. *Let μ and ν be log-concave isotropic probability measures on \mathbb{R}^n . Then*

$$d_{TV}(\mu, \nu) \leq C \sqrt{nd_{BL}(\mu, \nu)}.$$

In this result and below, C, c , etc. denote positive constants which are independent of n, μ , and ν , and whose values may change from one appearance to the next.

In the special case in which $n = 1$ and $\nu = \gamma_1$, Brehm, Hinow, Vogt and Voigt proved a similar comparison between total variation distance and Kolmogorov distance d_K .

Proposition 2 ([5, Theorem 3.3]). *Let μ be a log-concave measure on \mathbb{R} . Then*

$$d_{TV}(\mu, \gamma_1) \leq C \sqrt{\max \{1, \log(1/d_K(\mu, \gamma_1))\} d_K(\mu, \gamma_1)}.$$

Together with (1), Proposition 1 implies the following.

Corollary 3. *On the family of isotropic log-concave probability measures on \mathbb{R}^n , the topologies of weak convergence and of total variation coincide.*

Corollary 3 will probably be unsurprising to experts, but we have not seen it stated in the literature.

Proposition 1 and Corollary 3 are false without the assumption of isotropicity. For example, a sequence of nondegenerate Gaussian measures $\{\mu_k\}_{k \in \mathbb{N}}$ on \mathbb{R}^n may weakly approach a Gaussian measure μ supported on a lower-dimensional subspace, but $d_{TV}(\mu_k, \mu) = 2$ for every k . It may be possible to extend Corollary 3 to a class of log-concave probability measures with, say, a nontrivial uniform lower bound on the smallest eigenvalue of the covariance matrix, but we will not pursue this here.

The Kantorovitch duality theorem (see [15, Theorem 5.10]) gives a dual formulation of the L_1 Wasserstein distance similar to the formulation of total variation distance in (2):

$$W_1(\mu, \nu) = \sup_g \left| \int g \, d\mu - \int g \, d\nu \right|,$$

where the supremum is over 1-Lipschitz functions $g : \mathbb{R}^n \rightarrow \mathbb{R}$. An immediate consequence is that for any probability measures μ and ν ,

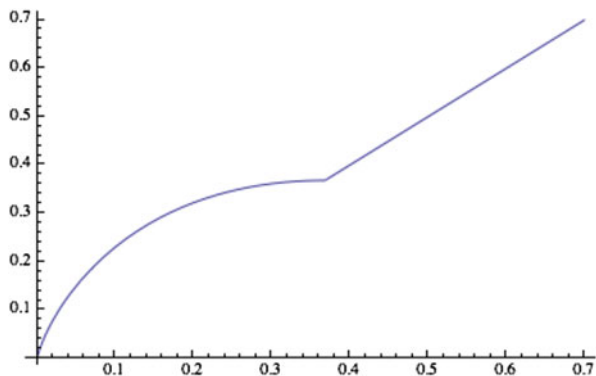
$$d_{BL}(\mu, \nu) \leq W_1(\mu, \nu).$$

The following complementary inequality holds in the log-concave case.

Proposition 4. *Let μ and ν be log-concave isotropic probability measures on \mathbb{R}^n . Then*

$$W_1(\mu, \nu) \leq C \max \left\{ \sqrt{n}, \log \left(\frac{\sqrt{n}}{d_{BL}(\mu, \nu)} \right) \right\} d_{BL}(\mu, \nu). \tag{3}$$

The following graph of $f(x) = \max \left\{ 1, \log \left(\frac{1}{x} \right) \right\} x$ may be helpful in visualizing the bounds in Proposition 4 and the results below.



In particular, when d_{BL} is moderate, we simply have $W_1 \leq C \sqrt{n} d_{BL}$. When d_{BL} is small, the right hand side of (3) is not quite linear in d_{BL} , but is $o(n^{\varepsilon/2} d_{BL}^{1-\varepsilon})$ for each $\varepsilon > 0$.

From Hölder’s inequality, it is immediate that if $p \leq q$, then $W_p(\mu, \nu) \leq W_q(\mu, \nu)$. In the log-concave case, we have the following.

Proposition 5. *Let μ and ν be isotropic log-concave probability measures on \mathbb{R}^n and let $1 \leq p < q$. Then*

$$W_q(\mu, \nu)^q \leq C \left(\max \left\{ \sqrt{n}, \log \left(\frac{(c \max\{q, \sqrt{n}\})^q}{W_p(\mu, \nu)^p} \right) \right\} \right)^{q-p} W_p(\mu, \nu)^p.$$

Because the bounded-Lipschitz distance metrizes the weak topology, and convergence in L_p Wasserstein distance implies convergence of moments of order smaller than p , Propositions 4 and 5 imply the following.

Corollary 6. *Let $\mu, \{\mu_k\}_{k \in \mathbb{N}}$ be isotropic log-concave probability measures on \mathbb{R}^n such that $\mu_k \rightarrow \mu$ weakly. Then all moments of the μ_k converge to the corresponding moments of μ .*

The following, known as the Csiszár–Kullback–Pinsker inequality, holds for any probability measures μ and ν :

$$d_{TV}(\mu, \nu) \leq \sqrt{2H(\mu \mid \nu)}. \tag{4}$$

(See [3] for a proof, generalizations, and original references.) Unlike the other notions of distance considered above, $H(\cdot \mid \cdot)$ is not a metric, and $H(\mu \mid \nu)$ can only be finite if μ is absolutely continuous with respect to ν . Nevertheless, it is frequently used to quantify convergence; (4) shows that convergence in relative entropy is stronger than convergence in total variation. Convergence in relative entropy is particularly useful in quantifying convergence to the Gaussian distribution, and it is in that setting that (4) can be essentially reversed under an assumption of log-concavity.

Proposition 7. *Let μ be an isotropic log-concave probability measure on \mathbb{R}^n , and let γ_n denote the standard Gaussian distribution on \mathbb{R}^n . Then*

$$H(\mu \mid \gamma_n) \leq C \max \left\{ \log^2 \left(\frac{n}{d_{TV}(\mu, \gamma_n)} \right), n \log(n + 1) \right\} d_{TV}(\mu, \gamma_n).$$

The proof of Proposition 7 uses a rough bound on the isotropic constant $L_f = \|f\|_\infty^{1/n}$ of the density f of μ . Better estimates are available but only result in a change in the absolute constants in our bound. In the case that the isotropic constant is bounded independent of n (e.g. if μ is the uniform measure on an unconditional convex body, or if the hyperplane conjecture is proved), then the bound above can be improved slightly to

$$H(\mu \mid \gamma_n) \leq C \max \left\{ \log^2 \left(\frac{n}{d_{TV}(\mu, \gamma_n)} \right), n \right\} d_{TV}(\mu, \gamma_n).$$

Corollary 8. *Let $\{\mu_k\}_{k \in \mathbb{N}}$ be isotropic log-concave probability measures on \mathbb{R}^n . The following are equivalent:*

1. $\mu_k \rightarrow \gamma_n$ weakly.
2. $\mu_k \rightarrow \gamma_n$ in total variation.
3. $H(\mu_k \mid \gamma_n) \rightarrow 0$.

It is worth noting that Proposition 7 implies that B. Klartag’s central limit theorem for convex bodies (proved in [8, 9] in total variation) also holds in the *a priori* stronger sense of entropy, with a polynomial rate of convergence.

2 Proofs of the Results

The proof of Proposition 1 uses the following deconvolution result of R. Eldan and B. Klartag.

Lemma 9 ([7, Proposition 10]). *Suppose that f is the density of an isotropic log-concave probability measure on \mathbb{R}^n , and for $t > 0$ define*

$$\varphi_t(x) = \frac{1}{(2\pi t^2)^{n/2}} e^{-\|x\|^2/2t^2}.$$

Then

$$\|f - f * \varphi_t\|_1 \leq cnt.$$

Proof (Proof of Proposition 1). Let $g \in C(\mathbb{R}^n)$ with $\|g\|_\infty \leq 1$. For $t > 0$, let $g_t = g * \varphi_t$, where φ_t is as in Lemma 9. It follows from Young's inequality that $\|g_t\|_\infty \leq 1$ and that g_t is $1/t$ -Lipschitz. We have

$$\left| \int g \, d\mu - \int g \, d\nu \right| \leq \left| \int (g - g_t) \, d\mu \right| + \left| \int g_t \, d\mu - \int g_t \, d\nu \right| + \left| \int (g_t - g) \, d\nu \right|.$$

It is a classical fact due to C. Borell [4] that a log-concave probability measures which is not supported on a proper affine subspace of \mathbb{R}^n has a density. If f is the density of μ , then by Lemma 9,

$$\left| \int (g - g_t) \, d\mu \right| = \left| \int g(f - f * \varphi_t) \right| \leq \|f - f * \varphi_t\|_1 \leq cnt,$$

and

$$\left| \int (g - g_t) \, d\nu \right| \leq cnt$$

similarly. Furthermore,

$$\left| \int g_t \, d\mu - \int g_t \, d\nu \right| \leq d_{BL}(\mu, \nu) \|g_t\|_{BL} \leq d_{BL}(\mu, \nu) \max\{1, 1/t\}.$$

Combining the above estimates and taking the supremum over g yields

$$d_{TV}(\mu, \nu) \leq d_{BL}(\mu, \nu) \max\{1, 1/t\} + cnt$$

for every $t > 0$. The proposition follows by picking $t = \sqrt{d_{BL}(\mu, \nu)/2n} \leq 1$.

The remaining propositions all depend in part on the following deep concentration result due to G. Paouris.

Proposition 10 ([14]). *Let X be an isotropic log-concave random vector in \mathbb{R}^n . Then*

$$\mathbb{P}[\|X\| \geq R] \leq e^{-cR}$$

for every $R \geq C\sqrt{n}$, and

$$(\mathbb{E}\|X\|^p)^{1/p} \leq C \max\{\sqrt{n}, p\}$$

for every $p \geq 1$.

The following simple optimization lemma will also be used in the remaining proofs.

Lemma 11. *Given $A, B, M, k > 0$,*

$$\inf_{t \geq M} (At^k + Be^{-t}) \leq A \left(1 + (\max\{M, \log(B/A)\})^k \right).$$

Proof. Set $t = \max\{M, \log(B/A)\}$.

Proof (Proof of Proposition 4). Let $g : \mathbb{R}^n \rightarrow \mathbb{R}$ be 1-Lipschitz and without loss of generality assume that $g(0) = 0$, so that $|g(x)| \leq \|x\|$. For $R > 0$ define

$$g_R(x) = \begin{cases} -R & \text{if } g(x) < -R, \\ g(x) & \text{if } -R \leq g(x) \leq R, \\ R & \text{if } g(x) > R, \end{cases}$$

and observe that $\|g_R\|_{BL} \leq \max\{1, R\}$. Let $X \sim \mu$ and $Y \sim \nu$. Then

$$\begin{aligned} |\mathbb{E}g(X) - \mathbb{E}g(Y)| &\leq \mathbb{E}|g_R(X) - g_R(Y)| + \mathbb{E}|g(X) - g_R(X)| + \mathbb{E}|g(Y) - g_R(Y)| \\ &\leq \max\{1, R\}d_{BL}(\mu, \nu) + \mathbb{E}\|X\| \mathbb{1}_{\|X\| \geq R} + \mathbb{E}\|Y\| \mathbb{1}_{\|Y\| \geq R}. \end{aligned}$$

By the Cauchy–Schwarz inequality and Proposition 10,

$$\mathbb{E}\|X\| \mathbb{1}_{\|X\| \geq R} \leq \sqrt{n\mathbb{P}[\|X\| \geq R]} \leq \sqrt{ne^{-cR}}$$

for $R \geq C\sqrt{n}$, and the last term is bounded similarly. Combining the above estimates and taking the supremum over g yields

$$W_1(\mu, \nu) \leq \max\{1, R\}d_{BL}(\mu, \nu) + 2\sqrt{ne^{-cR}}$$

for every $R \geq C\sqrt{n}$. The proposition follows using Lemma 11.

Proof (Proof of Proposition 5). Let (X, Y) be a coupling of μ and ν on $\mathbb{R}^n \times \mathbb{R}^n$. Then for each $R > 0$,

$$\mathbb{E} \|X - Y\|^q \leq R^{q-p} \mathbb{E} [\|X - Y\|^p \mathbb{1}_{\|X - Y\| \leq R}] + \sqrt{\mathbb{P} [\|X - Y\| \geq R] \mathbb{E} \|X - Y\|^{2q}}.$$

By Proposition 10,

$$\mathbb{P} [\|X - Y\| \geq R] \leq \mathbb{P} [\|X\| \geq R/2] + \mathbb{P} [\|Y\| \geq R/2] \leq e^{-cR}$$

when $R \geq C\sqrt{n}$, and

$$\left(\mathbb{E} \|X - Y\|^{2q}\right)^{1/2q} \leq \left(\mathbb{E} \|X\|^{2q}\right)^{1/2q} + \left(\mathbb{E} \|Y\|^{2q}\right)^{1/2q} \leq C \max\{q, \sqrt{n}\},$$

so that

$$\mathbb{E} \|X - Y\|^q \leq R^{q-p} \mathbb{E} \|X - Y\|^p + (C \max\{q, \sqrt{n}\})^q e^{-cR}$$

for every $R \geq C\sqrt{n}$. Taking the infimum over couplings and then applying Lemma 11 completes the proof.

The proof of Proposition 7 uses the following variance bound which follows from a more general concentration inequality due to Bobkov and Madiman.

Lemma 12 (See [2, Theorem 1.1]). *Suppose that μ is an isotropic log-concave probability measure on \mathbb{R}^n with density f , and let $Y \sim \mu$. Then*

$$\text{Var}(\log f(Y)) \leq Cn.$$

Proof (Proof of Proposition 7). Let f be the density of μ , and let $\varphi(x) = (2\pi)^{-n/2} e^{-\|x\|^2/2}$ be the density of γ_n . Let $Z \sim \gamma_n$, $Y \sim \mu$, $X = \frac{f(Z)}{\varphi(Z)}$, and $W = \frac{f(Y)}{\varphi(Y)}$. Then

$$H(\mu | \gamma_n) = \mathbb{E} X \log X.$$

In general, if μ and ν have densities f_μ and f_ν , it is an easy exercise to show that $d_{TV}(\mu, \nu) = \int |f_\mu - f_\nu|$; from this, it follows that

$$d_{TV}(\mu, \gamma_n) = \mathbb{E} |X - 1| = \frac{1}{2} \mathbb{E} (X - 1) \mathbb{1}_{X \geq 1}.$$

Let $h(x) = x \log x$. Since h is convex and $h(1) = 0$, we have that $h(x) \leq a(x - 1)$ for $1 \leq x \leq R$ as long as a is such that $h(R) \leq a(R - 1)$. Let $R \geq 2$, so that $\frac{R}{R-1} \leq 2$. Then

$$h(R) = R \log R \leq 2(R - 1) \log R = a(R - 1)$$

for $a = 2 \log R$. Thus

$$\begin{aligned} \mathbb{E}X \log X &\leq \mathbb{E}(X \log X) \mathbb{1}_{X \geq 1} \\ &\leq a \mathbb{E}(X - 1) \mathbb{1}_{X \geq 1} + \mathbb{E}(X \log X) \mathbb{1}_{X \geq R} \\ &= (\log R) d_{TV}(\mu, \gamma_n) + \mathbb{E}(X \log X) \mathbb{1}_{X \geq R}. \end{aligned}$$

The Cauchy–Schwarz inequality implies that

$$\mathbb{E}(X \log X) \mathbb{1}_{X \geq R} = \mathbb{E}(\log W) \mathbb{1}_{W \geq R} \leq \sqrt{\mathbb{E}(\log W)^2} \sqrt{\mathbb{P}[W \geq R]}.$$

By the L^2 triangle inequality, we have

$$\begin{aligned} \sqrt{\mathbb{E}(\log W)^2} &= \sqrt{\mathbb{E}|\log f(Y) - \log \varphi(Y)|^2} \\ &\leq \sqrt{\mathbb{E}|\log f(Y)|^2} + \sqrt{\mathbb{E}|\log \varphi(Y)|^2}, \end{aligned}$$

and by Proposition 10,

$$\mathbb{E}|\log \varphi(Y)|^2 = \mathbb{E}\left(\frac{n}{2} \log 2\pi + \frac{\|Y\|^2}{2}\right)^2 \leq Cn^2.$$

By Lemma 12,

$$\mathbb{E}|\log f(Y)|^2 \leq (\mathbb{E} \log f(Y))^2 + Cn.$$

Recall that the **entropy** of μ is

$$-\int f(y) \log f(y) dy = -\mathbb{E} \log f(Y) \geq 0,$$

and that γ_n is the maximum-entropy distribution with identity covariance, so that

$$(\mathbb{E} \log f(Y))^2 \leq (\mathbb{E} \log \varphi(Z))^2 = (n \log \sqrt{2\pi e})^2.$$

Thus

$$\sqrt{\mathbb{E}(\log W)^2} \leq Cn.$$

By [10, Theorem 5.14(e)], $\|f\|_\infty \leq 2^{8n} n^{n/2}$, and so

$$\begin{aligned} \mathbb{P}[W \geq R] &= \mathbb{P}\left[\frac{f(Y)}{\varphi(Y)} \geq R\right] \leq \mathbb{P}\left[e^{\|Y\|^2/2} \geq (2^{17}\pi n)^{-n/2} R\right] \\ &= \mathbb{P}\left[\|Y\| \geq \sqrt{2 \log((2^{17}\pi n)^{-n/2} R)}\right] \end{aligned}$$

for each $R \geq (2^{17} \pi n)^{n/2}$. Proposition 10 now implies that

$$\mathbb{P}[W \geq R] \leq e^{-c \sqrt{\log R - \frac{n}{2} \log(2^{17} \pi n)}} \leq e^{-c' \sqrt{\log R}}$$

for $\log R \geq Cn \log(n+1)$.

Substituting $S = c \sqrt{\log R}$, all together this shows that

$$H(\mu \mid \gamma_n) \leq C (S^2 d_{TV}(\mu, \gamma_n) + n e^{-S})$$

for every $S \geq c \sqrt{n \log(n+1)}$. The result follows using Lemma 11.

Acknowledgements This research was partially supported by grants from the Simons Foundation (#267058 to E.M. and #264103 to M.M.) and the U.S. National Science Foundation (DMS-1308725 to E.M.). This work was carried out while the authors were visiting the Institut de Mathématiques de Toulouse at the Université Paul Sabatier; the authors thank them for their generous hospitality.

References

1. K. Ball, Volume ratios and a reverse isoperimetric inequality. *J. Lond. Math. Soc.* (2) **44**(2), 351–359 (1991)
2. S. Bobkov, M. Madiman, Concentration of the information in data with log-concave distributions. *Ann. Probab.* **39**(4), 1528–1543 (2011)
3. F. Bolley, C. Villani, Weighted Csiszár-Kullback-Pinsker inequalities and applications to transportation inequalities. *Ann. Fac. Sci. Toulouse Math.* (6) **14**(3), 331–352 (2005)
4. C. Borell, Convex measures on locally convex spaces. *Ark. Mat.* **12**, 239–252 (1974)
5. U. Brehm, P. Hinow, H. Vogt, J. Voigt, Moment inequalities and central limit properties of isotropic convex bodies. *Math. Z.* **240**(1), 37–51 (2002)
6. R.M. Dudley, *Real Analysis and Probability*. Cambridge Studies in Advanced Mathematics, vol. 74 (Cambridge University Press, Cambridge, 2002). Revised reprint of the 1989 original
7. R. Eldan, B. Klartag, Pointwise estimates for marginals of convex bodies. *J. Funct. Anal.* **254**(8), 2275–2293 (2008)
8. B. Klartag, A central limit theorem for convex sets. *Invent. Math.* **168**(1), 91–131 (2007)
9. B. Klartag, Power-law estimates for the central limit theorem for convex sets. *J. Funct. Anal.* **245**(1), 284–310 (2007)
10. L. Lovász, S. Vempala, The geometry of logconcave functions and sampling algorithms. *Random Struct. Algorithms* **30**(3), 307–358 (2007)
11. E. Milman, On the role of convexity in functional and isoperimetric inequalities. *Proc. Lond. Math. Soc.* (3) **99**(1), 32–66 (2009)
12. E. Milman, On the role of convexity in isoperimetry, spectral gap and concentration. *Invent. Math.* **177**(1), 1–43 (2009)
13. E. Milman, Isoperimetric and concentration inequalities: equivalence under curvature lower bound. *Duke Math. J.* **154**(2), 207–239 (2010)
14. G. Paouris, Concentration of mass on convex bodies. *Geom. Funct. Anal.* **16**(5), 1021–1049 (2006)
15. C. Villani, *Optimal Transport: Old and New*. Grundlehren der Mathematischen Wissenschaften (Fundamental Principles of Mathematical Sciences), vol. 338 (Springer, Berlin, 2009)