LARGE DEVIATIONS VIA TRANSFERENCE PLANS *

Sergey G. Bobkov †

1 Introduction

Let P be a log-concave probability measure on $\mathbf{R}^n$. Equivalently, P is concentrated on some affine subspace $E \subset \mathbf{R}^n$ where it has a density p, with respect to Lebesgue measure on E, such that

$$p(tx + (1-t)y) \geq p(x)^t p(y)^{1-t}, \quad \text{for all } t \in (0,1) \text{ and } x, y \in E.$$

We refer the reader to the classical 1974 paper by C. Borell [Bor] for a general theory of such measures.

As is known, given a function f on $\mathbf{R}^n$, certain possible distributional properties of f with respect to P can be controled by the behavior of this function along lines. For example, when $f(x) = \|x\|$ is an arbitrary norm, we have an inequality for large deviations

$$P\{f > \lambda \mathbf{E} f\} \leq C e^{-c\lambda}, \quad \lambda \geq 0, \tag{1.1}$$

where C and c are positive numerical constants, and where we use probabilistic notations $\mathbf{E} f = \int f \, dP$ for the expectation with respect to P. In an equivalent form this fact first appeared in [Bor], cf. Lemma 3.1. If f is a polynomial in n real variables of degree d, we have a similar inequality

$$P\{|f| > \lambda \mathbf{E}|f|\} \leq C(d)\, e^{-c(d)\, \lambda^{r(d)}}, \quad \lambda \geq 0, \tag{1.2}$$

thus with the right hand side depending on d, but independent of the measure P. This observation, which gave an affirmative answer to a conjecture of V. D. Milman, is due to J. Bourgain [Bou] who considered for P the uniform distribution on an arbitrary convex body in $\mathbf{R}^n$.

Both (1.1) and (1.2) can be united by a more general scheme. With every continuous function f on $\mathbf{R}^n$ and $\varepsilon \in (0,1)$, we associate the quantity

$$\delta_f(\varepsilon) = \sup_{x_0, x_1 \in \mathbf{R}^n} \text{mes}\{t \in (0,1) : |\, f(tx_0 + (1-t)x_1)\,| < \varepsilon |f(x_0)|\}.$$

*Key words: Large deviations, Khinchine-type inequalities, transportation of mass

†Supported in part by an NSF grant

As turns out, the behavior of δ_f near zero is connected with large deviations of f, and moreover, the corresponding inequalities can be made independent of P. To study the polynomial case, J. Bourgain established the property

$$\delta_f(\varepsilon_0) \leq \delta_0 \tag{1.3}$$

with $\delta_0 = 1/2$ and with some $\varepsilon_0 \in (0,1)$ depending upon d. This was already enough to derive a very general statement on large deviations in the form (1.2). However, the altitude of $\delta_f(\varepsilon)$ for small ε's may contain an additional information on the strength of deviations. In this note, we refine and extend Bourgain's approach to arbitrary functions f and log-concave measures P, with resulting estimates depending upon δ_f, only. In particular, we prove:

Theorem 1.1. *Let P be a log-concave probability measure on $\mathbf{R}^n$, and let f be a continuous function on $\mathbf{R}^n$. Then, for all $\lambda > 2e$ such that $\delta_f(2e/\lambda) \leq 1/2$,*

$$P\{|f| > \lambda\, \mathbf{E}|f|\} \leq \exp\left\{-\frac{1}{2\delta_f(2e/\lambda)}\right\}. \tag{1.4}$$

Once $\delta_f(\varepsilon) \to 0$, as $\varepsilon \to 0$, the assumption $\delta_f(2e/\lambda) \leq 1/2$ is fulfilled for all λ large enough. In case $\delta_f(\varepsilon) \leq C\varepsilon^r$, for all $\varepsilon \in (0,1)$ and some $C \geq 1$, $r > 0$, we thus arrive at the estimate of the form

$$P\{|f| > \lambda\, \mathbf{E}|f|\} \leq c_1 e^{-c_2 \lambda^r}, \quad \lambda \geq 0.$$

As an example, we will observe that $\delta_f(\varepsilon) = \frac{2\varepsilon}{1+\varepsilon}$ for any norm $f(x) = \|x\|$ on $\mathbf{R}^n$, and we are thus lead to (1.1). In the polynomial case, $\delta_f(\varepsilon) = O(\varepsilon^{1/d})$ that leads to (1.2) with the correct power $r(d) = 1/d$.

The main Bourgain argument based on the existence of suitable measure-preserving maps is described in section 2. In the next section 3, we study large deviations under the condition 1.3. The latter turns out to be related to a property known as Markov's inequality for polynomials in one real variable. In section 4, we consider Theorem 1.1 itself and show how it can be applied to norms and polynomials, by computing or estimating the quantity $\delta_f(\varepsilon)$. In section 5, we apply Theorem 1.1 to study deviations of convex functions f from their mean $\mathbf{E}f$. In particular, we will consider the case of the euclidean norm for which it is possible to reach exponentially decreasing tails in terms of P-variance. At the end of this note, we put an appendix devoted to triangular measure-preserving maps.

2 Bourgain's argument

As a first basic step, we prove:

Theorem 2.1. *Let f be a continuous function on $\mathbf{R}^n$, and let P be a log-concave probability measure on $\mathbf{R}^n$. Let $\varepsilon \in (0,1)$ and $\delta = \delta_f(\varepsilon) < 1$. Then, for all $\lambda \geq 0$ and $\gamma \in (0, 1-\delta]$,*

$$P\{|f| > \lambda\varepsilon\} \geq \gamma P\{|f| > \lambda\}^{\delta/(1-\gamma)}. \tag{2.1}$$

Proof. Without loss of generality, we may assume that P is absolutely continuous with respect to Lebesgue measure on $\mathbf{R}^n$. Thus, P is concentrated on an open convex set $A_0 \subset \mathbf{R}^n$ where it has a positive, log-concave density $p(x)$. We may also assume that $0 < \lambda < \operatorname{ess\,sup} f$. For such values, introduce a family of non-empty open subsets of A_0,

$$A_\lambda = \{x \in A_0 : |f(x)| > \lambda\}.$$

Fix λ and take a regular subset A of A_λ (e.g., a finite union of open balls, cf. Appendix for details). In particular, $P(A) > 0$. We follow an argument of J. Bourgain [Bou]: There exists unique continuous bijective triangular map $T : A \to A_0$ which pushes forward the normalized restriction P_A of P to A to the measure P. Moreover, the components $T_i = T_i(x_1, \dots, x_i)$ of T, $i = 1, \dots, n$, are C^1-smooth with respect to x_i-coordinates and satisfy $\frac{\partial T_i}{\partial x_i} > 0$ so that the Jacobian

$$J(x) = \prod_{i=1}^n \frac{\partial T_i(x)}{\partial x_i}, \quad x \in A,$$

is continuous and positive on A. Since P_A has density $p_A(x) = \frac{p(x)}{P(A)}$, $x \in A$, the property that T pushes forward P_A to P is equivalent to saying that

$$\frac{p(x)}{P(A)} = p(T(x))\, J(x), \quad x \in A. \tag{2.2}$$

Now, for each $t \in (0,1)$, introduce another map,

$$T_t(x) = tx + (1-t)T(x), \quad x \in A,$$

which is also continuous, triangular, with components that are C^1-smooth with respect to x_i-coordinates. Moreover, its Jacobian J_t satisfies

$$J_t(x) = \prod_{i=1}^n \left(t + (1-t)\, \frac{\partial T_i(x)}{\partial x_i}\right) \geq J(x)^{1-t}, \quad x \in A. \tag{2.3}$$

Consider the set

$$B_t = \{x \in A : |f(tx + (1-t)T(x))| > \lambda\varepsilon\}$$

and its image $B'_t = T_t(B_t)$. Clearly, if $y \in B'_t$, then $y = tx + (1-t)T(x)$, for some $x \in B_t$, hence $|f(y)| > \lambda\varepsilon$, that is, $y \in A_{\lambda\varepsilon}$. This means that $B'_t \subset A_{\lambda\varepsilon}$, and therefore

$$P(B'_t) \leq P(A_{\lambda\varepsilon}). \tag{2.4}$$

On the other hand, using the log-concavity $p(tx+(1-t)x') \geq p(x)^t p(x')^{1-t}$ (which will be needed with $x' = T(x)$), and applying (2.2)-(2.3), we get

$$\begin{aligned} P(B_t') &= \int_{T_t(B_t)} p(y)\,dy = \int_{B_t} p(T_t(x))\,J_t(x)\,dx \\ &\geq \int_{B_t} p(T_t(x))\,J(x)^{1-t}\,dx \ \geq \ \int_{B_t} p(x)^t\,p(T(x))^{1-t}\,J(x)^{1-t}\,dx \\ &= P(A)^{t-1} \int_{B_t} p(x)\,dx \ = \ P(A)^{t-1}\,P(B_t). \end{aligned}$$

Together with (2.4), this yields

$$P(A_{\lambda\varepsilon}) \geq P(A)^{t-1}\,P(B_t). \tag{2.5}$$

Now, in order to further estimate from below the last term in (2.5), it is the time to involve the function δ_f. By the definition, for any $x \in A$,

$$\mathrm{mes}\{t \in (0,1) : |f(tx+(1-t)T(x)| < \varepsilon|f(x)|\} \leq \delta.$$

Since $A \subset A_\lambda$, we have $|f(x)| > \lambda$, so,

$$\mathrm{mes}\{t \in (0,1) : |f(tx+(1-t)T(x)| \leq \varepsilon\lambda\} \leq \delta,$$

or equivalently,

$$\int_0^1 1_{\{|f(tx+(1-t)T(x)|>\lambda\varepsilon\}}\,dt \geq 1-\delta.$$

Integrating this inequality over the measure P_A and interchanging the integrals, we get

$$\int_0^1 P_A(B_t)\,dt \geq 1-\delta. \tag{2.6}$$

Thus, the function $\psi(t) = P_A(B_t)$ being bounded by 1 satisfies $\int_0^1 \psi(t)\,dt \geq 1-\delta$. This actually implies that $\psi(t) \geq \gamma$, for some $t \in (0, t_0]$ where $t_0 = \frac{\delta}{1-\gamma} \in (0,1]$. Indeed, assuming that $\psi(t) < \gamma$, whenever $t \in (0, t_0]$, we would get

$$\int_0^1 \psi(t)\,dt = \int_0^{t_0} \psi(t)\,dt + \int_{t_0}^1 \psi(t)\,dt < \gamma t_0 + (1-t_0) = 1-(1-\gamma)t_0 = 1-\delta.$$

Thus, $\int_0^1 \psi(t)\,dt < 1-\delta$ that contradicts to (2.6). We can therefore conclude that

$$\frac{P(B_t)}{P(A)} = P_A(B_t) \geq \gamma, \quad \text{for some} \quad t \in (0, t_0].$$

Applying this in (2.5), we arrive at $P(A_{\lambda\varepsilon}) \geq \gamma P(A)^t$, and since $t \leq t_0$,

$$P(A_{\lambda\varepsilon}) \geq \gamma P(A)^{t_0}. \tag{2.7}$$

At last, approximating from below the set A_λ by regular subsets A so that $P(A) \uparrow P(A_\lambda)$, we get from (2.7) in the limit $P(A_{\lambda\varepsilon}) \geq \gamma P(A_\lambda)^{t_0}$, that is, exactly (2.1).

Theorem 2.1 is proved.

Remark 2.2. The above argument still works with many other measure preserving maps. For example, one may take for T the Brenier map, i.e., of the form $T = \nabla\varphi$, for some mod(P)-uniquely defined convex function φ, cf. [Bre] and [M]. In this case, the derivative $T'(x)$ represents a positively definite matrix, and the crucial inequality (2.3) should be replaced with

$$\det(t\,\mathrm{Id} + (1-t)\,T'(x)) \geq \det{}^{1-t}(T'(x))$$

which is a particular (and log-concave) case of the Brunn-Minkowski-type inequality for determinants $\det^{1/n}(A+B) \geq \det^{1/n}(A) + \det^{1/n}(B)$ in the class of all positively definite $n \times n$-matrices. However, to make the argument following (2.3) absolutely rigorous (the change of the variable formula), it is desirable to require that the map T_t be in a certain sense regular. C^1-smoothness seems too strong requirement, but specializing in triangular maps, it is enough to have C^1-smoothness of the components T_i of T with respect to i-th coordinates. We provide more details in appendix.

Remark 2.3. An attempt to choose an optimal γ in (2.1) complicates this inequality, but in essense does not give an improvement. For further applications, at the sake of some loss in constants, one may use Theorem 2.1 with $\gamma = \frac{1}{2}$, for example.

It is however interesting to know how sharp the inequality (2.1) is. The definition of δ_f reflects the behavior of the function f along all lines, so one may try to derive inequalities of this kind by appealing to the localization technique going back to the papers by M. Gromov and V. D. Milman, cf. [G-M], [A], and L. Lovász and M. Simonovits [L-S], cf. also [K-L-S]. The advantage of this approach is that it allows one to reduce many problems to dimension one where it is much easier to explore extremal situations. In a recent preprint [N-S-V], F. Nazarov, M. Sodin, and A. Volberg employ the localization ideas to prove the following remarkable statement which they call the geometric KLS lemma in the spirit of [K-L-S]: Given a compact convex set K in $\mathbf{R}^n$, its closed subset F, and a number $\alpha > 1$, define

$$F_\alpha = \left\{ x \in K : \text{for every interval } J \text{ such that } x \in J \subset K,\ \frac{|F \cap J|}{|J|} \geq \frac{\alpha - 1}{\alpha} \right\}.$$

Then, if $\mathrm{vol}_n(F) > 0$,

$$\frac{\mathrm{vol}_n(F_\alpha)}{\mathrm{vol}_n(K)} \leq \left(\frac{\mathrm{vol}_n(F)}{\mathrm{vol}_n(K)}\right)^\alpha.$$

It is noted in [N-S-V] that in the definition of F_α it is enough to consider only the intervals J that have x as one of their endpoints. Moreover, the above inequality extends to arbitrary log-concave probability measures P in the form

$$P(F_\alpha) \leq P(F)^\alpha. \tag{2.8}$$

To see a connection with (2.1), take $F = \{x : |f(x)| \geq \lambda\varepsilon\}$, $G = \{x : |f(x)| \geq \lambda\}$, and assume that $\delta = \delta_f(\varepsilon) < 1$. Then, by the very definition of δ_f, we have $G \subset F_\alpha$ for $\alpha = 1/\delta$, so (2.8) turns into

$$P\{|f| \geq \lambda\varepsilon\} \geq P\{|f| \geq \lambda\}^\delta. \tag{2.9}$$

This is an improved and more correct version of (2.1): the factor γ can thus be replaced with 1 while the power $\delta/(1-\gamma)$ can be replaced with δ. The inequality (2.9) is sharp already in some special sitiuations. For example, for an arbitrary norm $f(x) = \|x\|$ on $\mathbf{R}^n$, we have $\delta_f(\varepsilon) = \frac{2\varepsilon}{1+\varepsilon}$, $\varepsilon \in (0,1)$, so $P\{|f| \geq \lambda\varepsilon\} \geq P\{|f| \geq \lambda\}^{2\varepsilon/(1+\varepsilon)}$ which is another version of the ineqiality

$$1 - P\Big(\frac{1}{\varepsilon}A\Big) \leq (1 - P(A))^{2\varepsilon/(1+\varepsilon)} \tag{2.10}$$

for the class of all centrally symmetric convex sets A in $\mathbf{R}^n$. The latter was proved using a localization lemma by L. Lovász and M. Simonovits [L-S] for euclidean balls A and later extended by O. Guédon [G] to the general case. He also observed that equality in (2.10) is attained in dimension one for any interval $A = (-a, a)$, $a > 0$, at the (non-symmetric) exponential measure P with $P(x, +\infty) = e^{-(x+a)}$, $x > -a$. We do not know whether the argument based on the transference plans can appropriately be modified to reach the sharp forms (2.9)-(2.10).

Remark 2.4. It follows from (2.1) by letting $\lambda \downarrow 0$ that $P\{f = 0\} = 0$, if $f \neq 0$ mod(P) and $\delta_f(\varepsilon) \to 0$, as $\varepsilon \downarrow 0$. Note also that in Theorem 2.1 one may assume that f is defined on A_0 (rather than on the whole space), and restrict the points x_0 and x_1 in the definition of δ_f to the set A_0.

3 Iteration procedure. Markov's classes

In order to apply the inequality (2.1), the weakest assumption which should be required from f is the property $\delta_f(\varepsilon) \not\equiv 0$ (identically), that is,

$$\delta_f(\varepsilon_0) \leq \delta_0 \tag{3.1}$$

for some $\varepsilon_0 \in (0,1)$ and $\delta_0 \in (0,1)$. As shown in [Bou], already in this situation one can recover exponentially decreasing tails for f by iterating the inequality (2.1). Namely, we have:

Theorem 3.1. *Under the condition* (3.1), *there exist positive numbers C, c, r, depending on $(\varepsilon_0, \delta_0)$, only, such that, for all $\lambda \geq 0$,*

$$P\{|f| > \lambda \mathbf{E}|f|\} \leq C\, e^{-c\lambda^r}. \tag{3.2}$$

The power r appearing in (3.2) can be chosen as close to the number $r_0 = \frac{\log(1/\delta_0)}{\log(1/\varepsilon_0)}$, as we wish. In particular, f has finite moments $\mathbf{E}|f|^q$ of any order q, and, moreover, for all $0 < r < r_0$,

$$\mathbf{E} \exp\{|f|^r\} = \int \exp\{|f|^r\}\, dP < +\infty.$$

In addition, f satisfies Khinchine-type inequalities

$$(\mathbf{E}|f|^q)^{1/q} \leq C\, \mathbf{E}|f|, \qquad C = C(q, \varepsilon_0, \delta_0), \quad q \geq 1. \tag{3.3}$$

It would therefore be interesting to explore the class of all functions f possessing the property (3.1). One sufficient condition was suggested by Yu. V. Prokhorov in his study of Khinchine-type inequalities for polynomials over Gaussian and Γ-distributions on the real line, cf. [Pr1], [Pr2]. Prokhorov's proof of (3.3) is based on Markov's inequality,

$$\max_{0 \leq t \leq 1} |Q'(t)| \leq \kappa \max_{0 \leq t \leq 1} |Q(t)|, \tag{3.4}$$

which holds true for any polynomial Q in real variable t of degree d with (optimal) constant $\kappa = 2d^2$. Let us then say that a given function f on $\mathbf{R}^n$ belongs to the (Markov) class $M(\kappa)$ with constant $\kappa \geq 1$, if, for all vectors $x_0, x_1 \in \mathbf{R}^n$, the function $Q(t) = f(tx_0 + (1-t)x_1)$ is absolutely continuous on $\mathbf{R}$ and has a Radon-Nikodym derivative Q' satisfying the inequality (3.4). With this definition, we have:

Proposition 3.2. *Every function f in $M(\kappa)$ satisfies* (3.1) *with*

$$\varepsilon_0 = \frac{1}{2}, \quad \delta_0 = 1 - \frac{1}{2\kappa}.$$

Indeed, following an argument of [Pr1-2], let t_0 be a point of maximum of $|Q(t)|$ on [0,1], and assume for definiteness that $Q(t_0) > 0$. Then, by (3.4), for all $t \in [0, 1]$,

$$Q(t) \geq Q(t_0)(1 - \kappa|t - t_0|) \geq \frac{1}{2} Q(t_0),$$

where the second inequality holds true in a smaller subinterval $|t - t_0| \leq 1/(2\kappa)$, $0 \leq t \leq 1$. This interval has length at least $1/(2\kappa)$, so

$$\begin{aligned} \mathrm{mes}\{|\, f(tx_0 + (1-t)x_1)\,| < \varepsilon |f(x_0)|\} &= \mathrm{mes}\Big\{t \in (0,1) : |Q(t)| < \frac{1}{2}|Q(1)|\Big\} \\ &\leq \mathrm{mes}\Big\{t \in (0,1) : |Q(t)| < \frac{1}{2}|Q(t_0)|\Big\} \leq 1 - \frac{1}{2\kappa}. \end{aligned}$$

Thus, according to Theorem 3.1 and Proposition 3.2, any function f in $M(\kappa)$ shares the large deviation inequality (3.2) and the Khinchine-type inequality (3.3). Moreover, for large values of κ, the critical value r_0 in (3.2) is of order at most C/κ.

According to Markov's inequality, any polynomial f on $\mathbf{R}^n$ of degree d belongs to the class $M(2d^2)$. Another important example: any norm $f(x) = \|x\|$ belongs to the class $M(2)$. Indeed, the function $Q(t) = f(tx_0 + (1-t)x_1)$ is convex and satisfies, by the triangle inequality, $Q(t) \geq Q_0(t) \equiv |\, t\|x_0\| - (1-t)\|x_1\|\,|$. Since $Q_0(0) = Q(0)$, $Q_0(1) = Q(1)$, we conclude that

$$\max_{0 \leq t \leq 1} |Q'(t)| \leq \max_{0 \leq t \leq 1} |Q_0'(t)| = \|x_0\| + \|x_1\| \leq 2 \max_{0 \leq t \leq 1} |Q(t)|.$$

Proof of Theorem 3.1. Assume for definiteness that $f \neq 0 \bmod(P)$ and write the inequality (2.1) with $\varepsilon = \varepsilon_0$, $\delta = \delta_0$ and an arbitrary fixed $\gamma \in (0, 1-\delta_0)$ as

$$P\{|f| > \lambda\} \leq \alpha^\beta \, P\{|f| > \lambda\varepsilon_0\}^\beta, \quad \lambda \geq 0, \tag{3.5}$$

where $\alpha = \frac{1}{\gamma}$, $\beta = \frac{1-\gamma}{\delta_0}$. Thus, $\alpha > 1$ and $\beta > 1$. In case $\lambda = 0$, we get in particular that

$$P\{|f| > 0\} \geq \alpha^{-\frac{\beta}{\beta-1}} \tag{3.6}$$

Now, applying (3.5) to $\lambda\varepsilon_0$, we get $P\{|f| > \lambda\} \leq \alpha^{\beta+\beta^2} \, P\{|f| > \lambda\varepsilon_0^2\}^{\beta^2}$. Similarly, on the k-th step, we will have

$$P\{|f| > \lambda\} \leq \alpha^{\beta + \ldots + \beta^k} \, P\{|f| > \lambda\varepsilon_0^k\}^{\beta^k}.$$

Using $\beta + \ldots + \beta^k \leq \frac{\beta}{\beta-1} \, \beta^k$, we obtain a simpler estimate

$$P\{|f| > \lambda\} \leq \left(\alpha^{\frac{\beta}{\beta-1}} \, P\{|f| > \lambda\varepsilon_0^k\}\right)^{\beta^k}. \tag{3.7}$$

Now denote by m a quantile of $|f|$ of order $e^{-1} \, \alpha^{-\frac{\beta}{\beta-1}}$, that is, any number such that

$$P\{|f| > m\} \leq \frac{1}{e \, \alpha^{\frac{\beta}{\beta-1}}}, \quad P\{|f| < m\} \leq \frac{1}{e \, \alpha^{\frac{\beta}{\beta-1}}} \tag{3.8}$$

By (3.6), such a number m must be positive. Furthermore, the inequality (3.7) with $\lambda\varepsilon_0^k = m$ yields, for all $k = 1, 2 \ldots,$

$$P\left\{\frac{|f|}{m} > \varepsilon_0^{-k}\right\} \leq \exp\left\{-\beta^k\right\}. \tag{3.9}$$

Now take any $x \geq 1/\varepsilon_0$ and pick up a natural namber k such that $\varepsilon_0^{-k} \leq x < \varepsilon_0^{-(k+1)}$. Then, $k \geq \frac{\log x}{\log(1/\varepsilon_0)} - 1$, so, $\beta^k \geq \frac{1}{\beta} x^{\log\beta/\log(1/\varepsilon_0)}$. Since $P\left\{\frac{|f|}{m} > x\right\} \leq P\left\{\frac{|f|}{m} > \varepsilon_0^{-k}\right\}$, we derive from (3.9)

$$P\left\{\frac{|f|}{m} > x\right\} \leq \exp\left\{-\frac{1}{\beta} \, x^{\frac{\log\beta}{\log(1/\varepsilon_0)}}\right\}, \quad x \geq \frac{1}{\varepsilon_0}. \tag{3.10}$$

The power $r = \frac{\log \beta}{\log(1/\varepsilon_0)}$ in (3.10) is less than $r_0 = \frac{\log(1/\delta_0)}{\log(1/\varepsilon_0)}$ but it can be made close to this number by choosing small values of γ.

Now, in order to replace the quantile with the mean $\mathbf{E}|f|$, we may use the inequality $\mathbf{E}|f| \geq \frac{1}{e\,\alpha^{\frac{\beta}{\beta-1}}}\, m$, so (3.10) yields

$$P\left\{|f| > e\,\alpha^{\frac{\beta}{\beta-1}}\, x\,\mathbf{E}|f|\right\} \leq \exp\left\{-\frac{1}{\beta}\, x^{\frac{\log \beta}{\log(1/\varepsilon_0)}}\right\}, \quad x \geq \frac{1}{\varepsilon_0}.$$

Making the change $\lambda = e\,\alpha^{\frac{\beta}{\beta-1}}\, x$, we get the desired inequality $P\{|f| > \lambda\,\mathbf{E}|f|\} \leq e^{-c\,\lambda^r}$, $\lambda \geq \lambda_0$, with an arbitrarily chosen $\gamma \in (0, 1-\delta_0)$, and

$$r = \frac{\log \beta}{\log(1/\varepsilon_0)}, \quad c = \frac{1}{\beta\left(e\,\alpha^{\frac{\beta}{\beta-1}}\right)^{\frac{\log \beta}{\log(1/\varepsilon_0)}}}, \quad \lambda_0 = \frac{e\,\alpha^{\frac{\beta}{\beta-1}}}{\varepsilon_0}.$$

4 Theorem 1.1. Norms and polynomials

To derive the inequality of Theorem 1.1 from Theorem 2.1, assume f is normalized so that $\mathbf{E}|f| = 1$. By Chebyshev's inequaity, $P\{|f| \geq x\} \leq 1/x$, for all $x > 0$. If $\delta_f(\varepsilon) \leq 1/2$, we can take in (2.1) $\gamma = 1/2$ which leads to

$$P\{|f| > \lambda\} \leq (2\,P\{|f| > \lambda\varepsilon\})^{1/(2\delta_f(\varepsilon))} \leq \left(\frac{2}{\lambda\varepsilon}\right)^{1/(2\delta_f(\varepsilon))}, \quad \lambda \geq 0.$$

Choosing if possible $\varepsilon = 2e/\lambda$, we then arrive at the estimate (1.4), that is,

$$P\{|f| > \lambda\,\mathbf{E}|f|\} \leq \exp\left\{-\frac{1}{2\delta_f(2e/\lambda)}\right\}, \qquad \lambda > 2e,\ \delta_f(2e/\lambda) \leq 1/2. \tag{4.1}$$

The above inequality immediately implies:

Corollary 4.1. *If $\delta_f(\varepsilon) \leq C\varepsilon^r$, for all $\varepsilon \in (0,1)$ and some $C \geq 1$, $r > 0$, then*

$$P\left\{\frac{1}{2e}\,|f| \geq \lambda\,\mathbf{E}|f|\right\} \leq \exp\left\{-\frac{\lambda^r}{2C}\right\}, \quad \lambda^r \geq 2C. \tag{4.2}$$

As we see, the inequalities (4.1)-(4.2) may contain more precise information in comparison with the general Markov classes $M(\kappa)$. This concerns in particular such functions f as norms and polynomials for which it would be interesting to explore the behavior of δ_f near zero. We start with an arbitrary norm $f(x) = \|x\|$ on $\mathbf{R}^n$.

Proposition 4.2. *For any norm f, we always have $\delta_f(\varepsilon) = \frac{2\varepsilon}{1+\varepsilon}$, $\varepsilon \in (0,1)$.*

Proof. By the triangle inequality, for all $x_0, x_1 \in \mathbf{R}^n$, and $t \in (0,1)$,

$$\|tx_0 + (1-t)x_1\| \geq |\, t\|x_0\| - (1-t)\|x_1\|\,|.$$

Hence,

$$\begin{gathered}\mathrm{mes}\{t \in (0,1) : \|tx_0 + (1-t)x_1\| < \varepsilon\|x_0\|\} \leq \\ \mathrm{mes}\{t \in (0,1) : |\, t\|x_0\| - (1-t)\|x_1\|\,| < \varepsilon\|x_0\|\}.\end{gathered}$$

Putting $a = \|x_0\|$, $b = \|x_1\|$, $c = \frac{a}{a+b} \leq 1$, and assuming $a > 0$, we obtain

$$\begin{aligned}\mathrm{mes}\{t : |\, ta - (1-t)b\,| < \varepsilon a\} &= \mathrm{mes}\{t : -\varepsilon a < ta - (1-t)b < \varepsilon a\} \\ &= \mathrm{mes}\{t : (1-\varepsilon)\, a < t\,(a+b) < (1+\varepsilon)\, a\} \\ &= \mathrm{mes}\{t : (1-\varepsilon)\, c < t < (1+\varepsilon)\, c\} \\ &= \min\{1, (1+\varepsilon)\, c\} - (1-\varepsilon)\, c.\end{aligned}$$

The last quantity is maximized in $0 \leq c \leq 1$ at $c = \frac{1}{1+\varepsilon}$ which gives $\frac{2\varepsilon}{1+\varepsilon}$. The optimality of this upper bound can be seen in case $x_1 = -\varepsilon x_0$, $\|x_0\| = 1$. Proposition 4.2 is proved.

Thus, the assumption made in Corollary 4.1 is fulfilled for the norm-function with $r = 1$ and $C = 2$. Hence, by (4.2), for all $\lambda \geq 8e$,

$$P\,\{|f| > \lambda\, \mathbf{E}|f|\} \leq e^{-\lambda/(16e)}.$$

The numerical constants are certainly not optimal and can be improved by virtue of (2.10).

Now let f be an arbitrary polynomial of degree at most $d \geq 1$. In this case, the maximal possible value of $\delta_f(\varepsilon)$ is completely determined in dimension one, so assume $n = 1$. In [Bou] it was shown that, for some numerical $c_0 \in (0,1)$,

$$\mathrm{mes}\left\{t \in (0,1) : |f(t)| < c_0^d\, \|f\|_{L^\infty(0,1)}\right\} \leq \frac{1}{2}.$$

Thus, we always have $\delta_f(c_0^d) \leq \frac{1}{2}$ which complements Proposition 3.2 in the polynomial case, namely, $\delta_f(1/2) \leq 1 - \frac{1}{4d^2}$. As for small values of ε, we have:

Proposition 4.3. *For any polynomial f of degree at most $d \geq 1$, for all $\varepsilon \in (0,1)$,*

1) $\delta_f(\varepsilon) \leq 2d\,\varepsilon^{1/d}$; 2) $\delta_f(\varepsilon) \leq 2\,\varepsilon^{1/d} \log \frac{1}{\varepsilon^{1/d}}$.

Proof. Let $f(t) = \prod_{i=1}^d (t - z_i)$ with $z_i \in \mathbf{C}$, $1 \leq i \leq d$. Then, on the interval (0,1),

$$\mathrm{mes}\{t \in (0,1) : |f(t)| < \varepsilon\, |f(0)|\} \quad \leq \quad \mathrm{mes} \bigcup_{i=1}^d \{|t - z_i| < \varepsilon^{1/d}\, |z_i|\}$$

$$\leq \sum_{i=1}^{d} \mathrm{mes}\{|t - z_i| < \varepsilon^{1/d}\,|z_i|\}.$$

Since $|t - z_i| \geq |t - |z_i||$, the roots z_i may be assumed to be real non-negative numbers. But, for any $c \in (0,1)$ and $z > 0$, the quantity $\mathrm{mes}\{t \in (0,1) : |t - z| < cz\}$ is maximized at $z = \frac{1}{1+c}$ and is equal to $\frac{2c}{1+c}$. This gives the first inequality.

To get the second one, we follow an argument of [Bou]. Fix $\alpha \in (0,1)$ and put $u_i = 1/z_i$ ($z_i > 0$). By Chebyshev's and Hölder's inequalities,

$$\begin{aligned}
\mathrm{mes}\{t \in (0,1) : |f(t)| < \varepsilon\,|f(0)|\} &= \mathrm{mes}\left\{\prod_{i=1}^{d} |u_i t - 1|^{-\alpha/d} > \varepsilon^{-\alpha/d}\right\} \\
&\leq \varepsilon^{\alpha/d} \int_0^1 \prod_{i=1}^{d} |u_i t - 1|^{-\alpha/d}\,dt \\
&\leq \varepsilon^{\alpha/d} \left(\prod_{i=1}^{d} \int_0^1 |u_i t - 1|^{-\alpha}\,dt\right)^{1/d} \leq \frac{2\varepsilon^{\alpha/d}}{1-\alpha}
\end{aligned}$$

where we used a simple inequality $\int_0^1 |ut - 1|^{-\alpha}\,dt \leq \frac{2}{1-\alpha}$ ($u \geq 0$) on the last step. It remains to optimize over all $\alpha \in (0,1)$.

Thus, the condition of Corollary 4.1 is fulfilled with $r = 1/d$ and $C = 2d$. Hence:

Corollary 4.4. *For all* $\lambda \geq (4d)^d$,

$$P\left\{|f| > \lambda\,\mathbf{E}|f|\right\} \leq e^{-\lambda^{1/d}/(8ed)}.$$

The upper bound can further be sharpened with the help of the localization lemma of Lovász-Simonovits [L-S] which allows one to get in Khinchine-type inequalities for polynomials a correct order of constants as functions of degree d. As shown in [B1-2], for all $p \geq 1$,

$$(\mathbf{E}|f|^p)^{1/p} \leq (cp)^d\,\mathbf{E}|f|,$$

where $c > 1$ is a universal constant. Hence, by Chebyshev's inequality, for all $\lambda > 0$, $P\{|f| > \lambda\,\mathbf{E}|f|\} \leq \frac{(cp)^{pd}}{\lambda^p}$. Optimizing the right hand side over $p \geq 1$, we arrive at

$$P\{|f| > \lambda\,\mathbf{E}|f|\} \leq e^{-d\lambda^{1/d}/(ce)},$$

provided that $\lambda \geq (ce)^d$.

5 Deviations from the mean

Large deviations of f from the mean $\mathbf{E}f = \int f\,dP$ can be controled once we know how to estimate the quantity δ_{f-c} uniformly over all $c \in \mathbf{R}$. For example, since the class of polynomials f of degree d is closed under translations $f \to f + \mathrm{const}$, (1.2) implies the bound

$$P\{|f - \mathbf{E}f| > \lambda\sigma\} \leq C(d)\, \exp\left\{-c(d)\,\lambda^{1/d}\right\}, \quad \lambda \geq 0,$$

in terms of the variance $\sigma^2 = \mathbf{E}(f - \mathbf{E}f)^2$. One may therefore hope to reach similar dimension-free inequalities for other classes of functions. The question is stimulated by the observation (typical in concentration problems, cf. [M-S], [L]) that many interesting f's have very large expectations $\mathbf{E}f$, but relatively small variances σ^2. In this situation, bounds for $P\{|f - \mathbf{E}f| > \lambda\sigma\}$ are certainly more delicate and preferable in comparison with those for $P\{|f| > \lambda\mathbf{E}|f|\}$. However, if we wish to involve into consideration arbitrary norms, the desired extension of (1.1) to the larger class $f(x) = \|x\| - c$ is no longer valid, and some extra condition on the norm like the uniform convexity is required. To illustrate these ideas, we will consider here the example of the euclidean norm $f(x) = \|x\|_2$ on $\mathbf{R}^n$.

To start with, it might be reasonable to find an appropriate form of Theorem 1.1 for the case of devations from constants. To every continuous function f on $\mathbf{R}^n$ and $\varepsilon > 0$, we may associate another quantity $\Delta_f(\varepsilon)$ defined to be the least number $\Delta \in [0,1]$ such that, for all $x_0, x_1 \in \mathbf{R}^n$, the function $Q(t) = f(tx_0 + (1-t)x_1)$ satisfies

$$\mathrm{mes}\left\{t \in [0,1] : Q(t) - \min_{0 \leq s \leq 1} Q(s) < \varepsilon \left[\max_{0 \leq s \leq 1} Q(s) - \min_{0 \leq s \leq 1} Q(s)\right]\right\} \leq \Delta.$$

Theorem 5.1. *Let P be a log-concave probability measure on $\mathbf{R}^n$, and let f be a convex function on $\mathbf{R}^n$ with mean $\mathbf{E}f$ and variance σ^2. Then, for all $\lambda > 2e$ such that $\Delta_f(4e/(\lambda + 2e)) \leq 1/2$,*

$$P\{|f - \mathbf{E}f| > \lambda\,\sigma\} \leq \exp\left\{-\frac{1}{2\delta_f(4e/(\lambda + 2e))}\right\}.$$

The statement follows immediately from Theorem 1.1 and

Lemma 5.2. *For every convex f on $\mathbf{R}^n$, for all $\varepsilon \in (0,1)$,*

$$\sup_{c \in \mathbf{R}} \delta_{f-c}(\varepsilon) = \Delta_f\left(\frac{2\varepsilon}{1+\varepsilon}\right). \tag{5.1}$$

Proof. Fix $x_0, x_1 \in \mathbf{R}^n$ and the corresponding function $Q(t) = f(tx_0 + (1-t)x_1)$, $0 \le t \le 1$ (not identically a constant on [0,1]). Since Q is convex, it attains its maximum at $t = 0$ or $t = 1$. By homogeneity and translation invariance of (5.1), and replacing x_0 with x_1 if necessary, we may assume that $Q(1) = \max_{0 \le s \le 1} Q(s) = 1$ and $Q(t_0) = \min_{0 \le s \le 1} Q(s) = 0$, for some $t_0 \in [0,1)$. Consider the quantity

$$\varphi(c) = \mathrm{mes}\{t \in [0,1] : |Q(t) - c| < \varepsilon|1 - c|\}$$

appearing in the definition of $\delta_{f-c}(\varepsilon)$. In view of $\sup_{c \in \mathbf{R}} \delta_{f-c}(\varepsilon) = \sup_{x_0,x_1} \sup_{c \in \mathbf{R}} \varphi(c)$, we need to maximize the latter function over all c.

If $c > 1$, $|Q(t) - c| < \varepsilon|1 - c|$ implies $Q(t) > 1$ that is not possible. So, assume $c \le 1$ in which case the definition becomes

$$\varphi(c) = \mathrm{mes}\{t \in [0,1] : (1+\varepsilon)c - \varepsilon < Q(t) < (1-\varepsilon)c + \varepsilon\}. \tag{5.2}$$

In the range $\{c : (1+\varepsilon)c - \varepsilon < 0\} = (-\infty, \frac{\varepsilon}{1+\varepsilon})$, the first inequality in (5.2) is fulfilled automatically, while the upper bound $(1-\varepsilon)c + \varepsilon$ increases with c. Thus, we may also assume $c \ge c_0 \equiv \frac{\varepsilon}{1+\varepsilon}$.

As c varies in $[c_0, 1]$, the interval $((1+\varepsilon)c - \varepsilon, (1-\varepsilon)c + \varepsilon)$ moves to the right and its length $2\varepsilon(1-c)$ decreases from $2c_0$ to 0. By the convexity of Q, this implies that the length of the interval $\{t \in [t_0, 1] : (1+\varepsilon)c - \varepsilon < Q(t) < (1-\varepsilon)c + \varepsilon\}$ decreases as a function of c. Indeed, if Q is not a constant in any neighborhood of t_0, then it increases in $[t_0, 1]$, the inverse function $Q^{-1} : [0,1] \to [0, t_0]$ is concave, so, for any positive decreasing function $h = h(u)$, the function $Q^{-1}(u+h) - Q^{-1}(u)$ is decreasing in u, as well. A similar argument applies to Q restricted to the interval $[0, t_0]$. Therefore, $c = c_0$ is the point of minimum to φ. To involve a possible "degenerate" case, we should write

$$\sup_{c \in \mathbf{R}} \varphi(c) = \lim_{c \uparrow c_0} \varphi(c) = \mathrm{mes}\{t \in [0,1] : Q(t) < (1-\varepsilon)c_0 + \varepsilon\}.$$

It remains to note that $(1-\varepsilon)c_0 + \varepsilon = 2c_0 = \frac{2\varepsilon}{1+\varepsilon}$, and the lemma follows.

Now, let us turn to the particular case $f(x) = \|x\|_2$. The euclidean norm can be related to the polynomial f^2 of degree $d = 2$ via the following observation: For every convex $f \ge 0$ on $\mathbf{R}^n$, for all $\varepsilon > 0$ and $q \ge 1$, we have $\Delta_f(\varepsilon) \le \Delta_{f^q}(\varepsilon)$. The latter statement easily follows from the definition and a simple inequality $\frac{b^q - a^q}{c^q - a^q} \le \frac{b-a}{c-a}$, $0 \le a \le b \le c$ ($a \ne c$). Now, appropriate computations show that

$$\Delta_{\|x\|_2^2}(\varepsilon) = \frac{2\sqrt{\varepsilon}}{1+\sqrt{\varepsilon}}, \quad \varepsilon \in (0,1).$$

Hence, $\Delta_{\|x\|_2}(\varepsilon) \le \Delta_{\|x\|_2^2}(\varepsilon) \le 2\sqrt{\varepsilon}$. Thus, from Theorem 5.1, we obtain

Corollary 5.3. *Let $X = (X_1, \dots, X_n)$ be a random vector in $\mathbf{R}^n$ with a log-concave distribution. Let σ^2 be the variance of $\|X\|_2$. Then,*

$$\mathrm{Prob}\{|\,\|X\|_2 - \mathbf{E}\|X\|_2\,| > \lambda\sigma\} \le Ce^{-c\sqrt{\lambda}}, \quad \lambda \ge 0,$$

where C and c are positive numerical constants.

Since one can relate the strength of concentration of $\|X\|_2$ about its mean to the standard deviation σ, one may wonder how to bound the variance itself. For normalization, let the covariances of the components of X satisfy

$$\mathrm{cov}(X_i, X_j) \equiv \mathbf{E} X_i X_j - \mathbf{E} X_i \mathbf{E} X_j = \delta_{ij}, \quad 1 \leq i, j \leq n, \tag{5.3}$$

where δ_{ij} is the Kronecker symbol. Under this (isotropy) assumption, the question of whether or not $\sigma^2 = \mathrm{Var}(\|X\|_2)$ does not exceed a universal constant represents a week form of a conjecture of R. Kannan, L. Lovász and M. Simonovits, cf. [K-L-S]. One simple sufficient condition of dimension free boundedness of σ^2, namely, the property

$$\mathrm{cov}(X_i^2, X_j^2) \leq 0, \quad 1 \leq i < j \leq n, \tag{5.4}$$

was recently proposed by K. Ball and I. Perissinaki [B-P]. Indeed, for positive random variables ξ's, there is a general estimate $\mathrm{Var}(\xi) \leq \frac{\mathrm{Var}(\xi^2)}{\mathbf{E}\xi^2}$, which for $\xi = \|X\|_2$ in view of (5.3) becomes $\mathrm{Var}(\|X\|_2) \leq \frac{\mathrm{Var}(\|X\|_2^2)}{n}$. On the other hand,

$$\mathrm{Var}(\|X\|_2^2) = \sum_{i=1}^n \mathrm{Var}(X_i^2) + 2 \sum_{i<j} \mathrm{cov}(X_i^2, X_j^2) \leq \sum_{i=1}^n \mathbf{E} X_i^4 \leq Cn,$$

where we used (5.4) and Khinchine-type inequality $\mathbf{E} X_i^4 \leq C\,(\mathbf{E} X_i^2)^2 = C$.

In [B-P], a property implying (5.3) was verified for random vectors X uniformly distributed in ℓ_p^n balls in $\mathbf{R}^n$.

6 Appendix: Triangular maps

Here we recall some facts about triangular maps which are needed for the proof of Theorem 2.1. A map $T = (T_1, \dots, T_n) : G \rightarrow \mathbf{R}^n$ defined on an open non-empty set G in $\mathbf{R}^n$ is called triangular if its components are of the form

$$T_i = T_i(x_1, \dots, x_i), \qquad x \in G, \quad 1 \leq i \leq n.$$

The triangular map T will be called increasing if, for all $i \leq n$, the component T_i is a (strictly) increasing function with respect to x_i-coordinate while the rest coordinates are fixed (x_i may vary within an open interval which depends on the rest coordinates x_j, $j < i$).

Such maps were used by H. Knothe [Kn] to reach some generalizations of the Brunn-Minkowski inequality. The following statement is often refered to as the construction of the Knothe mapping [Kn].

Theorem 6.1. *Let A and B be open, bounded, non-empty convex sets in $\mathbf{R}^n$. There exists a continuous, bijective, triangular map $T : A \to B$ such that*

a) *the partial derivatives $\frac{\partial T_i}{\partial x_i}$ are continuous and positive on A ;*

b) *the Jacobian $J(x) = \mathrm{Det}(T'(x)) = \prod_{i=1}^n \frac{\partial T_i}{\partial x_i}$ is constant on A and satisfies*

$$J(x) = \frac{\mathrm{Vol}_n(B)}{\mathrm{Vol}_n(A)}, \qquad x \in A;$$

c) *the map T pushes forward the uniform distribution on A to the uniform distribution on B.*

Note that T is not required to be C^1-smooth, so the property b) first defines a function "Jacobian" and then pustulates that it is a constant.

To complete Bourgain's argument, we need an appropriate generalization of Theorem 6.1 for measures. In [Bou], Theorem 6.1 is stated without convexity assumption on A which might lead to singularity problems. Indeed, consider, for example, the sets $B = (0,1)\times(0,1)$ and $A = (0,1)\times(0,2)\cup(0,2)\times(0,1) \subset \mathbf{R}^2$. The set A is open, bounded and has Lebesgue measure $|A| = 3$. Let P be a probability measure which has density $p(x) = 1/3$, for $x \in A$, and $p = 0$ outside A. Then, the distrubution P_1 of x_1-coordinate under P is concentrated on the interval $A_1 = (0,2)$ and has there density

$$p_1(x_1) = \begin{cases} 2/3, & \text{if} \quad 0 < x_1 < 1 \\ 1/3, & \text{if} \quad 1 < x_1 < 2 \end{cases}$$

That is, P_1 does not have any continuous density on $(0,2)$. But the property that P_1 has a continuous density is necessary for smoothness of triangular maps which push forward P to the uniform measure Q on B.

Thus, to save the property a) in the general non-convex case, some extra condition is required. First note that, given random vectors $X = (X_1, \dots, X_n)$ and $Y = (Y_1, \dots, Y_n)$ with values in open sets A and B and distributed according to P and Q, respectively, the first i coordinates $(X_1, \dots, X_i)$ and $(Y_1, \dots, Y_i)$ $(1 \le i \le n)$ have distributions P_i and Q_i supported on the open sets

$$A_i = \{x \in \mathbf{R}^i : \exists\, t \in \mathbf{R}^{n-i} \ \ (x,t) \in A\},$$

$$B_i = \{x \in \mathbf{R}^i : \exists\, t \in \mathbf{R}^{n-i} \ \ (x,t) \in B\},$$

which are projections of A and B to $\mathbf{R}^i$. In particular, $A_n = A$, $B_n = B$.

We will say that P is regular if it has a (necessarily continuous) density p on A such that the following two conditions are satisfied. The first condition is that, for each $i \le n$, the measure P_i has a positive continuous density p_i on A_i. This is equivalent to saying that the integral

$$p_i(x_1, \dots, x_i) = \int_{\mathbf{R}^{n-i}} p(x_1, \dots, x_i, t_{i+1}, \dots, t_n)\, dt_{i+1} \dots dt_n$$

is finite for every $(x_1, \dots, x_i) \in A_i$ and represents a continuous function on A_i (when $i = n$, we just have $p_i = p$). The second condition is that, for each $i \leq n$, the conditional distribution function

$$\mathrm{Prob}\{X_i \leq x_i \,|\, X_j = x_j,\ j = 1, \dots, x_{i-1}\}$$

is continuous in $(x_1, \dots, x_i) \in A_i$. Under the first condition, this is equivalent to saying that the integral

$$\int_{-\infty}^{x_i} \int_{\mathbf{R}^{n-i}} p(x_1, \dots, x_{i-1}, t_i, \dots, t_n)\, dt_i \dots dt_n$$

defines a continuous function on A_i. We can now state a corresponding generalization of Theorem 6.1.

Theorem 6.2. *For all regular probability measures P and Q supported on an open set A and on an open convex set B, respectively, there exists unique increasing, continuous, triangular, bijective map $T : A \to B$ which pushes forward P to Q. Moreover, the components T_i are C^1-smooth with respect to x_i-coordinates and satisfy $\frac{\partial T_i}{\partial x_i} > 0$ on A_i.*

Some examples of regular measures will be described at the end of this section, and we turn to the study of triangular maps T themselves. Many properties of such maps are determined by behavior of functions T_i with respect to x_i-coordinates. We collect some properties in the two lemmas below.

Lemma 6.3. *If $T = (T_1, \dots, T_n) : G \to \mathbf{R}^n$ is a continuous, increasing, triangular map, then the image $T(G)$ is an open set, and T represents a homeomorphism between G and $T(G)$.*

Lemma 6.4. *Let $T = (T_1, \dots, T_n) : G \to \mathbf{R}^n$ be a continuous triangular map whose components T_i have continuous positive partial derivatives $\frac{\partial T_i}{\partial x_i}$ on G. Then, for every integrable function f on $\mathbf{R}^n$,*

$$\int_G f(T(x))\, J(x)\, dx = \int_{T(G)} f(y)\, dy, \tag{6.1}$$

where $J(x) = \prod_{i=1}^n \frac{\partial T_i(x)}{\partial x_i}$.

Note that, by Lemma 6.3, the map T from Lemma 6.4 is increasing, so T is a bijection from G to the open set $T(G)$. The topological Lemma 6.3 can easily be proved by virtue of Brauer's theorem, so we omit the proof.

Proof of Lemma 6.4. If T is C^1-smooth on G, i.e., T has a continuous derivative $T' = (\frac{\partial T_i}{\partial x_j})_{1 \leq i,j \leq n}$, the "true" Jacobian $J(x) = \mathrm{Det}(T'(x)) = \prod_{i=1}^n \frac{\partial T_i(x)}{\partial x_i}$

is well-defined and is everywhere positive on G. Hence, the equality (6.1) holds true by the well-known theorem on the change of the variable in the Lebesgue integral.

In general, we approximate T by smooth triangular maps T^ε. Recall that the domain of T_i is the open set

$$G_i = \{x = (x_1, \dots, x_i) \in \mathbf{R}^i : \exists y \in \mathbf{R}^{n-i}\ (x, y) \in \mathbf{R}^n\}.$$

Now, take C_0^∞-functions $K_i \geq 0$, $i = 1, \dots, n$, supported on the unit ball $D_i(0,1) \subset \mathbf{R}^i$ and such that $\int_{D_i(0,1)} K_i(y)\, dy = 1$, and introduce convolutions T_i^ε of T_i with $K_i^\varepsilon(x) = \frac{1}{\varepsilon^i} K_i(x/\varepsilon)$, $\varepsilon > 0$:

$$T_i^\varepsilon(x) = \int_{G_i} K_i^\varepsilon(x - z)\, T_i(z)\, dz, \quad x \in \mathbf{R}^i.$$

The above integral is well-defined and represents a C^∞-function on $\mathbf{R}^i$. Differentiating over x_i, we obtain a C^∞-function

$$\frac{\partial T_i^\varepsilon(x)}{\partial x_i} = \int_{G_i} \frac{\partial K_i^\varepsilon(x - z)}{\partial x_i}\, T_i(z)\, dz, \quad x \in \mathbf{R}^i.$$

The kernel K_i^ε is supported on $D_i(0, \varepsilon)$. Hence, this integral may be taken over the whole space $\mathbf{R}^i$ as soon as $D_i(x, \varepsilon)$ lies in G_i:

$$\frac{\partial T_i^\varepsilon(x)}{\partial x_i} = \int_{\mathbf{R}^i} \frac{\partial K_i^\varepsilon(x - z)}{\partial x_i}\, T_i(z)\, dz, \quad x \in \mathbf{R}^i.$$

Let $\varepsilon_i(x)$ be the supremum of such ε's. Integrating by parts, for $\varepsilon \in (0, \varepsilon_i(x)]$, $\varepsilon < +\infty$, we get

$$\frac{\partial T_i^\varepsilon(x)}{\partial x_i} = \int_{\mathbf{R}^i} K_i^\varepsilon(x - z)\, \frac{\partial T_i(z)}{\partial z_i}\, dz = \int_{G_i} K_i^\varepsilon(x - z)\, \frac{\partial T_i(z)}{\partial z_i}\, dz, \quad x \in \mathbf{R}^i.$$

The integral on the right is again well-defined and represents a C^∞-function as the convolution of $\frac{\partial T_i}{\partial z_i}$ with K_i^ε. Moreover, it is positive, by the assumptions on T_i and K_i.

Thus, the map $T^\varepsilon = (T_1^\varepsilon, \dots, T_n^\varepsilon)$ is triangular, C^∞-smooth, with positive Jacobian

$$J_\varepsilon(x) = \prod_{i=1}^n \frac{\partial T_i^\varepsilon(x_1, \dots, x_i)}{\partial x_i}$$

at every point $x \in G$ and for all $\varepsilon \in (0, +\infty)$ such that $0 < \varepsilon \leq \varepsilon_n(x)$ (note that the numbers $\varepsilon_i(x_1, \dots, x_i)$ decrease when i increases from 1 to n). Moreover, $J_\varepsilon(x) > 0$ on the set $G_\varepsilon = \{x \in G : D_n(x, \varepsilon) \subset G\}$, the open ε-interior of G. Therefore, we can apply (6.1) to any open set $A \subset G_\varepsilon$ and every intergrable function f on $\mathbf{R}^n$ to get

$$\int_A f(T^\varepsilon(x))\, J_\varepsilon(x)\, dx = \int_{T^\varepsilon(A)} f(y)\, dy. \tag{6.2}$$

Now, by the choice of K_i's, and since T and $\frac{\partial T_i}{\partial x_i}$ are continuous, $T^\varepsilon(x) \to T(x)$ and $\frac{\partial T_i^\varepsilon(x)}{\partial x_i} \to \frac{\partial T_i(x)}{\partial x_i}$, as $\varepsilon \downarrow 0$ uniformly over all $x \in A$, for every A whose closure $\mathrm{clos}(A)$ is compact and lies in G. Similarly, $J_\varepsilon(x) \to J(x)$.

Let A be open with compact closure $\mathrm{clos}(A) \subset G$. Since $G_\varepsilon \uparrow G$, as $\varepsilon \downarrow 0$, there is $\varepsilon_0 > 0$ such that $\mathrm{clos}(A) \subset G_{\varepsilon_0}$. Hence, we can apply the Lebesgue dominated convergence theorem: for every continuous function f on $\mathbf{R}^n$,

$$\int_A f(T^\varepsilon(x))\, J_\varepsilon(x)\, dx \to \int_A f(T(x))\, J(x)\, dx, \quad \text{as} \quad \varepsilon \downarrow 0. \tag{6.3}$$

To find the limit of the right hand side of (6.2), first note that

$$\limsup_{\varepsilon \downarrow 0} T^\varepsilon(A) \subset \mathrm{clos}(T(A)) = T(\mathrm{clos}(A)).$$

On the other hand, by a topological argument, we have $T(A) \subset \liminf_{\varepsilon \downarrow 0} T^\varepsilon(A)$, that is, whenever $a \in A$, the point $b = T(a)$ is contained in $T^\varepsilon(A)$, for all $\varepsilon > 0$ small enough. As a result, $1_{T(A)}(y) \leq \liminf_{\varepsilon \downarrow 0} 1_{T^\varepsilon(A)}(y) \leq \limsup_{\varepsilon \downarrow 0} 1_{T^\varepsilon(A)}(y) \leq 1_{T(\mathrm{clos}(A))}(y)$, for every $y \in \mathbf{R}^n$. Hence, for every non-negative bounded continuous function f on $\mathbf{R}^n$,

$$\int_{T(A)} f(y)\, dy \leq \liminf_{\varepsilon \downarrow 0} \int_{T^\varepsilon(A)} f(y)\, dy \leq \limsup_{\varepsilon \downarrow 0} \int_{T^\varepsilon(A)} f(y)\, dy \leq \int_{T(\mathrm{clos}(A))} f(y)\, dy.$$

Together with (6.2)-(6.3) we get

$$\int_{T(A)} f(y)\, dy \leq \int_A f(T(x))\, J(x)\, dx \leq \int_{T(\mathrm{clos}(A))} f(y)\, dy.$$

This already easily implies the equality (6.1). Lemma 6.4 follows.

In order to turn to the proof of Theorem 6.2, let us first emphasize what exactly we need to prove. Assume we have two absolutely continuous probability measures P and Q on $\mathbf{R}^n$ which are supported on some open sets A and B and have there densities $p(x)$ and $q(y)$, respectively. We wish to construct a continuous bijective map $T = (T_1, \dots, T_n) : A \to B$ which pushes forward P to Q. This property is denoted $Q = PT^{-1}$ or $Q = T(P)$ and can be defined via the equality $\int_B f\, dQ = \int_A f(T)\, dP$ or, in terms of densities, as

$$\int_B f(y) q(y)\, dy = \int_A f(T(x)) p(x)\, dx, \tag{6.4}$$

holding for every bounded measurable function f on B. If T is C^1-smooth and has at every point $x \in A$ an invertible matrix $T'(x) = (\frac{\partial T_i(x)}{\partial x_j})_{1 \leq i,j \leq n}$ of the first derivatives, one can make in the first integral the change of variable $y = T(x)$, and (6.4) becomes

$$\int_A f(T(x))\, q(T(x))\, |\mathrm{Det}(T'(x))|\, dx = \int_A f(T(x))\, p(x)\, dx.$$

Thus, if T is bijective, C^1-smooth, and the Jacobian $J(x) = \mathrm{Det}(T'(x))$ is everywhere positive, the necessary and sufficient condition for $Q = PT^{-1}$ is that, for almost all $x \in A$,

$$q(T(x))J(x) = p(x). \tag{6.5}$$

In the case where the map T is increasing and triangular, one can weaken the smoothness requirement and just assume that the components T_i have positive continuous derivatives $\frac{\partial T_i}{\partial x_i}$. Indeed, if $B = T(A)$, then, by Lemma 6.4, the equality

$$\int_B g(y)\,dy = \int_A g(T(x))\,J(x)\,dx$$

holds true for every integrable function g on B with $J(x) = \prod_{i=1}^n \frac{\partial T_i(x)}{\partial x_i}$. Applying this equality to $g(y) = f(y)q(y)$, we get

$$\int_B f(y)q(y)\,dy = \int_A f(T(x))q(T(x))\,J(x)\,dx.$$

Therefore, (6.4) would immediately follow from (6.5). Thus, we may conclude:

Lemma 6.5. *Let $T = (T_1, \dots, T_n) : A \to \mathbf{R}^n$ be a continuous triangular map whose components T_i have continuous positive partial derivatives $\frac{\partial T_i}{\partial x_i}$ on A. Let $B = T(A)$. If the equality* (6.5) *holds true for almost all $x \in A$, then the map T pushes forward P to Q.*

However, the existence of the triangular map T satisfying (6.5) requires more properties such as regularity of P and Q.

Proof of Theorem 6.2. We use induction over n, and prove at the same time that the components T_i, $1 \le i \le n$, satisfy, for all $(x_1, \dots, x_i) \in A_i$, the relation

$$\frac{\int_{-\infty}^{x_i} \int_{\mathbf{R}^{n-i}} p(x_1, \dots, x_{i-1}, t_i, \dots, t_n)\,dt_i \dots dt_n}{\int_{-\infty}^{+\infty} \int_{\mathbf{R}^{n-i}} p(x_1, \dots, x_{i-1}, t_i, \dots, t_n)\,dt_i \dots dt_n} =$$

$$\frac{\int_{-\infty}^{T_i} \int_{\mathbf{R}^{n-i}} q(T_1, \dots, T_{i-1}, t_i, \dots, t_n)\,dt_i \dots dt_n}{\int_{-\infty}^{+\infty} \int_{\mathbf{R}^{n-i}} q(T_1, \dots, T_{i-1}, t_i, \dots, t_n)\,dt_i \dots dt_n}\,, \tag{6.6}$$

where it is also claimed that all the integrals are finite and positive. For $i = 1$, the above formula becomes

$$\int_{-\infty}^{x_1} \int_{\mathbf{R}^{n-1}} p(t_1, \dots, t_n)\,dt_1 \dots dt_n = \int_{-\infty}^{T_1} \int_{\mathbf{R}^{n-1}} q(t_1, \dots, t_n)\,dt_1 \dots dt_n\,, \tag{6.7}$$

while for $i = n$, it reads as

$$\frac{\int_{-\infty}^{x_n} p(x_1, \dots, x_{n-1}, t_n)\,dt_n}{\int_{-\infty}^{+\infty} p(x_1, \dots, x_{n-1}, t_n)\,dt_n} = \frac{\int_{-\infty}^{T_n} q(T_1, \dots, T_{n-1}, t_n)\,dt_n}{\int_{-\infty}^{+\infty} q(T_1, \dots, T_{n-1}, t_n)\,dt_n}\,. \tag{6.8}$$

Note that the formulas (6.6)-(6.8) may be written in a more compact probabilistic form as

$$\mathrm{Prob}\{X_i \leq x_i \,|\, X_1 = x_1, \dots, X_{i-1} = x_{i-1}\} = \mathrm{Prob}\{Y_i \leq T_i \,|\, Y_1 = T_1, \dots, Y_{i-1} = T_{i-1}\},$$

where for $i \leq n$ we write for short $T_i = T_i(x_1, \dots, x_i)$.

The case $n = 1$ is obvious: the desired map $T = T_1(x_1)$ is unique and is determined by

$$\int_{-\infty}^{x_1} p(t_1)\, dt_1 = \int_{-\infty}^{T_1(x_1)} q(t_1)\, dt_1. \tag{6.9}$$

Clearly, T_1 is a C^1-smooth increasing bijection from A to B (B is an interval).

Now, to perfom the induction step, assume $n \geq 2$ and recall that P_i and Q_i denote the distrubution of the first i variables $(x_1, \dots, x_i)$ under P and Q, respectively. By the induction hypothesis, there is unique continuous, increasing, triangular bijective map $(T_1, \dots, T_{n-1}) : A_{n-1} \to B_{n-1}$ which transports P_{n-1} to Q_{n-1}, and moreover, the equality (6.6) holds true on A_i for all $i \leq n-1$.

According to (6.9) for the case $n = 1$, the equality (6.7) expresses the fact that the measure P_1 is transported to Q_1 by the map T_1. Similarly and more generally, the equality (6.6) expresses the fact that, given a vector $(x_1, \dots, x_{i-1}) \in A_{i-1}$, the function

$$x_i \to T_i(x_1, \dots, x_{i-1}, x_i)$$

transports the corresponding conditional measure $P_{x_1,\dots,x_{i-1}}$ of P on the line in $\mathbf{R}^i$ with these first $i-1$ coordinates to the conditional measure $Q_{T_1(x_1),\dots,T_{i-1}(x_1,\dots,x_{i-1})}$ of Q on the line with fixed coordinates $T_1(x_1), \dots, T_{i-1}(x_1, \dots, x_{i-1})$. In order to make the same to be valid when $i = n$, we just postulate equality (6.8) as the definition of T_n. Note that $P_{x_1,\dots,x_{n-1}}$ represents a probability measure which is supported on the open one dimensional set

$$A(x_1, \dots, x_{n-1}) = \{x \in \mathbf{R} : (x_1, \dots, x_{n-1}, x) \in A\},$$

while $Q_{T_1(x_1),\dots,T_{n-1}(x_1,\dots,x_{n-1})}$ is a probability measure supported (by convexity of B) on the open segment

$$B(x_1, \dots, x_{n-1}) = \{y \in \mathbf{R} : (T_1(x_1), \dots, T_{n-1}(x_1, \dots, x_{n-1}), y) \in B\}.$$

In addition, by the regularity assumption made on P and Q, these measures have positive continuous densities on $A(x_1, \dots, x_{n-1})$ and $B(x_1, \dots, x_{n-1})$, respectively. Hence, as well as in the case $n = 1$, for all $(x_1, \dots, x_{n-1}) \in A_{n-1}$, the function

$$x_n \to T_n(x_1, \dots, x_{n-1}, x_n)$$

represents a C^1-smooth increasing bijection from $A(x_1, \dots, x_{n-1})$ to $B(x_1, \dots, x_{n-1})$. This proves that $(T_1, \dots, T_{n-1}, T_n)$ is an increasing bijection from A to B together with the fact that all components T_i are C^1-smooth with respect to x_i-coordinates.

It should also be clear that, for each $i \le n$, the function T_i continuously depends on $(x_1, \dots, x_i)$. Indeed, the case $i = 1$ does not need to be verified, while for $i \ge 2$ we may argue using induction over i. Assuming that $T_1, \dots, T_{i-1}$ are continuous, introduce the function

$$\psi(x_1, \dots, x_{i-1}, y) = \int_{-\infty}^{y} \int_{\mathbf{R}^{n-i}} q(T_1, \dots, T_{i-1}, t_i, \dots, t_n)\, dt_i \dots dt_n$$

and write the equality (6.6) as

$$R(x_1, \dots, x_i) = \psi(x_1, \dots, x_{i-1}, T_i).$$

By the regularity assumption on P and Q and the induction hypothesis, both R and ψ are continuous functions defined respectively on the open sets A_i and

$$\{(x_1, \dots, x_{i-1}, y) : (x_1, \dots, x_{i-1}) \in A_{i-1},\ (T_1, \dots, T_{i-1}, y) \in B_i\}.$$

In particular, if $x'_j \to x_j$ for all $j = 1, \dots, i$, and $y = T_i(x_1, \dots, x_i)$, $y' = T_i(x'_1, \dots, x'_i)$, we get that

$$\psi(x'_1, \dots, x'_{i-1}, y') \to \psi(x_1, \dots, x_{i-1}, y).$$

The function ψ increases with respect to y. So, if y' does not converge to y, and for definiteness $y' \le y - \varepsilon$ for some $\varepsilon > 0$, then for some $\delta > 0$,

$$\begin{aligned} \psi(x'_1, \dots, x'_{i-1}, y') &\le \psi(x'_1, \dots, x'_{i-1}, y - \varepsilon) \to \psi(x_1, \dots, x_{i-1}, y - \varepsilon) \\ &< \psi(x_1, \dots, x_{i-1}, y) - \delta \end{aligned}$$

which is a contradiction. Hence, $y' \to y$, and thus T_i is continuous.

Now, differentiating (6.7) over x_1, (6.6) over x_i, where we assume that $2 \le i \le n - 1$, and (6.8) over x_n, we get respectively,

$$\int_{\mathbf{R}^{n-1}} p(x_1, t_2, \dots, t_n)\, dt_2 \dots dt_n = \int_{\mathbf{R}^{n-1}} q(T_1, t_2, \dots, t_n)\, dt_2 \dots dt_n \ \frac{\partial T_1}{\partial x_1}, \quad (6.10)$$

$$\frac{\int_{\mathbf{R}^{n-i}} p(x_1, \dots, x_i, t_{i+1}, \dots, t_n)\, dt_{i+1} \dots dt_n}{\int_{\mathbf{R}^{n-i+1}} p(x_1, \dots, x_{i-1}, t_i, \dots, t_n)\, dt_i \dots dt_n} =$$

$$\frac{\int_{\mathbf{R}^{n-i}} q(T_1, \dots, T_i, t_{i+1}, \dots, t_n)\, dt_{i+1} \dots dt_n}{\int_{\mathbf{R}^{n-i+1}} q(T_1, \dots, T_{i-1}, t_i, \dots, t_n)\, dt_i \dots dt_n} \ \frac{\partial T_i}{\partial x_i}, \quad (6.11)$$

$$\frac{p(x_1, \dots, x_n)}{\int_{-\infty}^{+\infty} p(x_1, \dots, x_{n-1}, t_n)\, dt_n} = \frac{q(T_1, \dots, T_n)}{\int_{-\infty}^{+\infty} q(T_1, \dots, T_{n-1}, t_n)\, dt_n} \ \frac{\partial T_n}{\partial x_n}. \quad (6.12)$$

Myltiplying (6.10)-(6.11)-(6.12) by each other, we arrive at

$$p(x_1,\dots,x_n) = q(T_1,\dots,T_n)\,\prod_{i=1}^n \frac{\partial T_i}{\partial x_i}$$

which is exactly (6.5). It remains to apply Lemma 6.5, and the existence part of Theorem 6.2 immediately follows.

The uniqueness follows from the requirement that, given a vector $(x_1,\dots,x_{i-1}) \in A_{i-1}$, the function $x_i \to T_i(x_1,\dots,x_{i-1},x_i)$ must transport the conditional measure $P_{x_1,\dots,x_{i-1}}$ to the conditional measure $Q_{T_1(x_1),\dots,T_{i-1}(x_1,\dots,x_{i-1})}$.

Theorem 6.2 is now proved.

To give some examples of regular measures (in the above sense), we need another definition. In addition to the projections A_i, with every set $A \subset \mathbf{R}^n$, we also associate its sections

$$A_{x_1,\dots,x_i} = \{t \in \mathbf{R}^{n-i} : (x_1,\dots,x_i,t) \in A\}, \qquad (x_1,\dots,x_i) \in \mathbf{R}^i,\ 1 \le i \le n-1.$$

We say that A is regular, if for all $i \le n-1$ and for all $(x_1,\dots,x_i) \in A_i$, the section $(\partial A)_{x_1,\dots,x_i}$ of the boundary of A has the $(n-i)$-dimensional Lebesgue measure zero.

For example, a finite union of balls represents a regular set. Another simple example is provided by an arbitrary open convex set in $\mathbf{R}^n$. As for regularity of measures, the following lemma covers most interesting cases.

Lemma 6.6. *Assume that a probability measure P is concentrated on an open set $B \subset \mathbf{R}^n$ where it has a positive continuous density p such that, for each $i \le n-1$,*

$$\int_{\mathbf{R}^{n-i}} \sup_{x \in B_i} p(x,t)\,dt < +\infty \tag{6.13}$$

(where it is assumed that $p = 0$ outside B). Then, the normalized restriction of P to any regular set $A \subset B$ is a regular measure.

The condition (6.13) is fulfilled, for example, if with some positive constants C and c, the density p satsfies an inequality

$$p(x) \le Ce^{-c|x|}, \quad x \in B. \tag{6.14}$$

Proof of Lemma 6.7. By the assumption, the function

$$p_i(x) = \int_{\mathbf{R}^{n-i}} p(x,t)\,1_A(x,t)\,dt$$

is finite for every $x \in B_i$, and moreover the function under the integral sign is bounded by an integrable function. We should show that p_i is continuous on A_i.

So, take a sequence $x^{(k)} \in A_i$ converging to a point $x \in A_i$, as $k \to \infty$. Then, for every $t \in \mathbf{R}^{n-1}$,

$$1_A(x,t) \leq \liminf_{k\to\infty} 1_A(x^{(k)},t) \leq \limsup_{k\to\infty} 1_A(x^{(k)},t) \leq 1_{\mathrm{clos}(A)}(x,t).$$

Since $p_i(x^{(k)},t) \to p(x,t)$, as $k \to \infty$, and since A is open, we may apply Lebesgue dominated converging theorem which gives

$$\begin{aligned} p_i(x) &\leq \liminf_{k\to\infty} \int_{\mathbf{R}^{n-i}} p(x^{(k)},t)\, 1_A(x^{(k)},t)\, dt \\ &\leq \limsup_{k\to\infty} \int_{\mathbf{R}^{n-i}} p(x^{(k)},t)\, 1_A(x^{(k)},t)\, dt \ \leq \ \int_{\mathbf{R}^{n-i}} p(x,t)\, 1_{\mathrm{clos}(A)}(x,t)\, dt. \end{aligned}$$

Now note that, by regularity of A, for any $x \in A_i$

$$1_{\mathrm{clos}(A)}(x,t) - 1_A(x,t) = 1_{\partial A}(x,t) = 0, \quad \text{for almost all } \ t \in \mathbf{R}^{n-i}$$

with respect to Lebesgue measure on $\mathbf{R}^{n-i}$. Hence, $\int_{\mathbf{R}^{n-i}} p(x,t)\, 1_{\mathrm{clos}(A)}(x,t)\, dt = p_i(x)$, and thus p_i is continuous. The first condition involved in the definition of regularity of a measure is therefore fulfilled. The second condition requires to verify that, for each $i \leq n$, the function

$$\begin{aligned} r_i(x,x_i) &= \int_{-\infty}^{x_i} \int_{\mathbf{R}^{n-i}} p(x,t_i,t)\, 1_A(x,t_i,t)\, dt_i\, dt \\ &= \int_{-\infty}^{+\infty} \int_{\mathbf{R}^{n-i}} p(x,t_i,t)\, 1_A(x,t_i,t)\, 1_{(-\infty,x_i]\times\mathbf{R}^{n-i}}(t_i,t)\, dt_i\, dt \end{aligned}$$

is continuous in $(x,x_i) \in A_i$, as well, where for short we write $x = (x_1,\dots,x_{i-1})$, $t = (t_{i+1}\ \dots,t_n)$. In case $i = 1$, the above expression depends on x_1, only,

$$r_1(x_1) = \int_{-\infty}^{+\infty} \int_{\mathbf{R}^{n-1}} p(t_1,t)\, 1_A(t_1,t)\, 1_{(-\infty,x_1]\times\mathbf{R}^{n-1}}(t_1,t)\, dt_1\, dt,$$

and is clearly continuous on A_1. In the case $i \geq 2$, we use the property that, for every $(t_i,t) \in \mathbf{R} \times \mathbf{R}^{n-i}$, the function $(x,x_i) \to p(x,t_i,t)\, 1_{(-\infty,x_i]\times\mathbf{R}^{n-i}}(t_i,t)$ is continuous on A_i, and then argue as before: for any $x \in A_{i-1}$,

$$1_{\mathrm{clos}(A)}(x,t_i,t) - 1_A(x,t_i,t) = 1_{\partial A}(x,t_i,t) = 0, \quad \text{for almost all } \ (t_i,t) \in \mathbf{R} \times \mathbf{R}^{n-i}$$

with respect to Lebesgue measure on $\mathbf{R}^{n-i+1}$, and therefore, once more by the Lebesgue dominated convergence theorem, r_i is continuous on A_i. Lemma 6.6 follows.

Corollary 6.7. *Uniform distrubution on a bounded regular set is a regular measure.*

This statement appears as a particular case of Lemma 6.6 with $A = B$ and $p = 1/|A|$ on A (where $|A|$ stands for the Lebesgue measure).

At last, since absolutely continuous log-concave measure on $\mathbf{R}^n$ are known to satisfy (6.14), we also obtain:

Corollary 6.8. *Every absolutely continuous log-concave measure P on $\mathbf{R}^n$ is regular. Moreover, the normalized restriction of P to an arbitrary regular set A of positive Lebesgue measure in the support of P represents a regular measure.*

REFERENCES

[A] Alesker, S. Localization technique on the sphere and the Gromov-Milman theorem on the concentration phenomenon on uniformly convex sphere. *Convex Geometric Analysis* (Berkeley, CA, 1996). *Math. Sci. Res. Inst. Publ.*, 34, Cambridge Univ. Press, Cambridge, 1999.

[B-P] Ball, K., Perissinaki, I. Subindependence of coordinate slabs in ℓ_p^n balls. *Israel J. of Math.*, 107 (1998), 289–299.

[B1] Bobkov, S.G. On the growth of L^p-norms of polynomials. *Geom. Aspects of Func. Anal., Lecture Notes in Math.*, 1745 (2000), 27–35.

[B2] Bobkov, S.G. Some generalizations of the results of Yu.V. Prokhorov on Khinchine-type inequalities for polynomials. *Probab. Theory and its Appl.*, 45 (2000), No. 4, 745–748.

[Bor] Borell, C. Convex measures on locally convex spaces. *Ark. Math.*, 12 (1974), 239–252.

[Bou] Bourgain, J. On the distribution of polynomials on high dimensional convex sets. Israel Seminar (GAFA) 1989–90, Lecture Notes in Math., 1469 (1991), 127–137.

[Bre] Brenier, Y. Polar factorization and monotone rearrangement of vector-valued functions. *Comm. Pure Appl. Math.*, 44 (1991), 375–417.

[G-M] Gromov, M., Milman, V.D. Generalization of the spherical isoperimetric inequality to uniformly convex Banach spaces. *Compositio Math.*, 62 (1987), 263–282.

[G] Guédon, O. Kahane-Khinchine type inequalities for negative exponents. *Mathematika*, 46 (1999), 165–173.

[K-L-S] Kannan, R., Lovász, L., Simonovits, M. Isoperimetric problems for convex bodies and a localization lemma. *Discrete and Comput. Geom.*, 13 (1995), 541–559.

[Kn] Knothe, H. Contributions to the theory of convex bodies. Michigan Math. J., 4 (1957), 39–52.

[L] Ledoux, M. The concetration of measure phenomenon. Math. Surveys and Monographs, AMS, vol. 89, 2001.

[L-S] Lovász, L., Simonovits, M. Random walks in a convex body and an improved volume algorithm. *Random Structures and Algorithms*, 4 (1993), No. 3, 359–412.

[M-S] Milman, V. D., Schechtman, G. Asymptotic theory of finite dimensional normed spaces. *Lecture Notes in Math.*, vol. 1200, 1986, Springer-Verlag.

[M] McCann, R. Existence and uniqueness of monotone measure-preserving maps. *Duke Math. J.*, 80 (1995), No. 2, 309–323.

[N-S-V] Nazarov, F., Sodin, M., Voberg, A. The geometric KLS lemma, dimension-free estimates for the distribution of values of polynomials, and distribution of zeroes of random analytic functions. Preprint (2001).

Sergey G. Bobkov
School of Mathematics
University of Minnesota
127 Vincent Hall, 206 Church St. S.E.
Minneapolis, MN 55455