HIGH DIMENSIONAL APPROXIMATION SEMINAR

CONCENTRATION PHENOMENA

WOLFGANG DAHMEN

Abstract. This document has been prepared by Miss J.L. Nelson from lecture notes and transcriptions of Prof. Dahmen's talks on concentration of measure phenomena the High Dimensional Approximation Seminar during the Spring 2008 semester at the University of South Carolina.

1. Concentration Phenomena

1.1. **Classical Approximation Estimates.** Before addressing specific problems of approximation in high dimensions, let us briefly review the basic approximation concepts for moderate spatial dimension d.

A rough overview of constructive approaches reads as follows:

(a) Spectral schemes, raising (trigonometric) polynomial degree
(b) splines, piecewise-defined functions based on spatial localization, for example by refining partitions of the domain
(c) kernel/meshless methods (implicitly based on localization)

Let us briefly review the type of error estimates for such methods.

Example 1.1. (a): (A simple typical estimate) Consider the space of square integrable functions over the d-dimensional torus $\mathbb{T}^d = \mathbb{R}^d/(2\pi\mathbb{Z})^d$

$$X := L_2(\mathbb{T}^d)$$

with inner product

$$(f,g) := (2\pi)^{-d}\int_{\mathbb{T}^d} f(x)\overline{g(x)}dx.$$

Elementary arguments lead to the following bounds:

$$\|f - S_n(f)\| \le (n+1)^{-m}\left(\sum_{j=1}^{d}\|\partial_i^m f\|_{L_i}^2\right)^{1/2}, \tag{1.1}$$

where

$$S_n(f) := \sum_{|k|\le n}(f,e_k)e_k$$

and

$$e_k(x) := e^{ikx} \text{ for } k\in\mathbb{Z} \text{ and } x\in\mathbb{R}^d.$$

Date: 18, 20, 25 February 2008 .

Defining $T_n :=$ span$\{e_k : |k|_\infty \le n\}$, note that the dimension of T_n is $(2n+1)^d$. Thus for

$$U_m = \{f : \|f\|^2 + \sum_{j=1}^d \|\partial_j^m f\|^2 \le 1\}$$

to ensure that this holds

$$\sup_{f\in U_m} \|f - S_n(f)\| \le \varepsilon$$

one needs $(n+1)^{-m} \le \varepsilon$, which means $n+1 \ge \varepsilon^{-1/m}$. In other words, the necessary number of degrees of freedom to achieve accuracy ε is given by

$$\dim T_n = (2n+1)^d \ge (n+1)^d \ge \varepsilon^{-d/m}.$$

Hence, we have exponential growth of complexity with respect to target accuracy.

Example 1.2. (b). Methods based on spatial localization usually use local polynomial approximation estimates. Typical steps include:

- Choose a standard reference domain, D, such as the unit cube, simplex, or ball and estimate
$$\inf_{P\in\mathbb{P}_k} \|f-P\|_{L_p(D)} \le C\ |f|_{W^{k+1}(L_p(D))}$$
where $C = C(D,k,d)$.
- Rescale $D_h = A_h D$ with diam$D_h = h$. A scaling argument then gives
$$\inf_{p\in\mathbb{P}_k} \|f-P\|_{L_p(\Omega)} \le C\ h^{k+1}\ |f|_{W^{k+1}(L_p(D_h))}. \tag{1.2}$$
- Denoting by S_h a space of splines / piecewise polynomials over partitions $\mathcal{P}_h$ of the domain D consisting of shape regular cells, D_h, of above type.
  Combining now (1.2) with quasi-interpolant techniques yields
$$\inf_{s\in S_h} \|f-S\|_{L_p(D)} \le C\ h^{k+1}\ |f|_{W^{k+1}(L_p(D))}, \tag{1.3}$$
where the constant $C = C(k,d,D,\mathcal{P})$, however, is usually hard to quantify and typically exhibits the dependencies listed above.

Thus, for U_n, the unit ball of $W^{k+1}(L^p(D))$ one has

$$\sup_{f\in U_n}\left(\inf_{S\in S_h} \|f-S\|_{L_p(D)} \le \varepsilon\right)$$

which requires that

$$h \le \left(\frac{\varepsilon}{C}\right)^{1/(k+1)}.$$

However, the fact that the diam(D) is of order one implies that $\dim(S_h) = N(h) \sim h^{-d}$. This would entail that

$$N^{-1/d} \le \left(\frac{\varepsilon}{C}\right)^{1/(k+1)}$$

which may be written as

$$N \ge \left(\frac{C}{\varepsilon}\right)^{d/(k+1)},$$

where again $C = C(d,k)$. This is, in essence, the same exponential dependence of complexity on target accuracy, another manifestation of the "curse of dimensionality."

There are two types of remedies–or attempts there at:

- Use different constructive concepts such as hyperbolic crosses, sparse grids, superposition, and so on.
- The above estimates typically contain constants that are hard to pin down exactly. Look for a different framework for error estimation that takes the specificities of high dimensional geometry into account.

1.2. **Some Simple Observations.** Let the d-dimensional unit ball be given by

$$B^d(x;r) = \{y \in \mathbb{R}^d : |x-y|_2 \le r\}$$

then

$$|B^d(0,r)| = \frac{2}{d}\frac{r^d \pi^{d/2}}{\Gamma\left(\frac{d}{2}\right)} = \frac{2}{d}\frac{(r^2\pi)^{d/2}}{\Gamma\left(\frac{d}{2}\right)} = \frac{2}{d}\frac{|B^2(r)|^{d/2}}{\Gamma\left(\frac{d}{2}\right)}.$$

An immediate consequences of this is

- Volume concentrates near vertices of a cube as d grows:

$$\frac{|B^d(0,r)|}{|[-r,r]^d|} = \frac{2}{d}\frac{\left(\frac{\pi}{4}\right)^{d/2}}{\Gamma\left(\frac{d}{2}\right)} \to 0.$$

- Volume concentrates near the boundary of $B^d(0,r)$:

$$\frac{|B^d(0,r) \setminus B^d(0,r-\varepsilon)|}{|B^d(0,r)|} = \frac{r^d-(r-\varepsilon)^d}{r^d} = 1-\left(1-\frac{\varepsilon}{r}\right)^d \to 1 \text{ as } d \to \infty.$$

Systematic studies and theoretical foundations have been laid by P. Levi, V. Milman, M. Gromov, M. Talagrand, and M. Ledoux [3].

Such phenomena arise in different ways in various areas of mathematics:

- They play a key role in V. Milman's asymptotic geometric analysis [4].
- Another description is the "observable diameter" of S^d in Gromov's theory [2]. In particular, the observable diameter[1] of S^d, behaves like $\frac{1}{\sqrt{d}}$ as $d \to \infty$. (More will be said about this later.)
- In probability the *concentration of measure* is a property of a large number of variables. This is generally known as the law of large numbers:

$$\mathbb{P}\{|\mathbb{E}(x) - \frac{1}{d}\sum_{j=1}^d x_j| < \varepsilon\} \to 1 \text{ as } d \to \infty \text{ for any } \varepsilon > 0.$$

Borel's Geometric Interpretation of the Law of Large Numbers. Consider the d-dimensional hypercube, $[0,1]^d$. Let H be the hyperplane perpendicular to $e := (1,\ldots,1)$ which goes through the center, $(\frac{1}{2},\ldots,\frac{1}{2})$, and define its r-neighborhood

$$H_r := \{x \in [0,1] : \ \mathrm{dist}(x,H) < r\}.$$

Denoting by μ_d, the uniform measure on $[0,1]^d$, then for any $\varepsilon > 0$ we have

$$\mu_d(H_{\varepsilon\sqrt{d}}) \to 1 \text{ as } d \to \infty. \tag{1.4}$$

Next, identify the principal diagonal in $[0,1]^d$ connecting 0 and e with the interval $[0,\sqrt{d}]$ and note that the orthogonal projection of $x \in [0,1]^d$ to the diagonal is $\left(\frac{1}{\sqrt{d}}e, x\right)\frac{1}{\sqrt{d}}e$ which corresponds to the point $\frac{1}{d}(x_1+\ldots+x_d) \in [0,\sqrt{d}]$. Since the mean of independent, uniformly distributed $x_i \in [0,1]$ is $1/2$, (1.4) means that

[1] The transcripter's understanding is that the observable diameter is the distance between two the points of tangency of S^d to rays from the observer. Remind her to insert a little picture here.

orthogonally projecting the cube onto its principal diagonal puts almost all the mass onto segments of the form

$$\left[\frac{\sqrt{d}}{2} - \varepsilon\sqrt{d}, \frac{\sqrt{d}}{2} + \varepsilon\sqrt{d}\right].$$

In fact, the central limit theorem implies that $\varepsilon\sqrt{(d)}$ can be replaced by any $r_d \to \infty$ as $d \to \infty$.

Independent Bernoulli Variables. Let x_i be independent Bernoulli variables (that is, they take on the values $+1$ and -1 with equal probability). In that case, we have

$$\mathbb{P}\left\{\left|\frac{x_1 + \ldots + x_d}{d}\right| \geq r\right\} \leq 2e^{-dr^2/2} \text{ for } r > 0.$$

1.3. **The Classical Concentration Inequality.** Let us consider such concentration phenomena from a measure theoretic point of view. The classical example reads as follows.

Let $S^d \subset \mathbb{R}^d$ be the standard d-sphere, σ^d be the uniform measure on S^d, and $\rho(x, y)$ be the geodesic distance on S^d :

$$\begin{aligned} \rho(x, y) &= \theta(x, y) \\ \cos(\theta(x, y)) &= \frac{x \cdot y}{|x|_2 |y|_2} \text{ i.e. } \theta = \cos^{-1}(x \cdot y), \end{aligned}$$

where $|\cdot|_2$ is the Euclidean norm.

Now consider a Borel set $A \subset S^d$ and its r-neighborhood

$$A_r := \{x \in S^d : \rho(x, A) < r\}.$$

We wish to quantify how much mass concentrates near the perimeter of A.

Theorem 1.3. *For ever Borel set $A \subset S^d$ such that $\sigma^d(A) \geq \frac{1}{2}$ one has*

$$\sigma^d(A_r) \geq 1 - e^{-(d-1)r^2/2}. \tag{1.5}$$

One interpretation of this theorem is as follows: Almost all the points in S^d are within geodesic distance $\frac{1}{\sqrt{d}}$ from A.

The key to such an inequality is a so-called *isoperimetric inequality* which can be stated as follows:

Theorem 1.4. *(IPI) For any Borel set $A \subset S^d$ and any spherical cap $C \subset S^d$ with $\sigma^d(A) = \sigma^d(C)$ one has*

$$\sigma^d(A_r) \geq \sigma^d(C_r). \tag{1.6}$$

That is, spherical caps minimize the boundary measure at fixed volume.

Now, $\sigma^d(C_r)$ can be estimated quite accurately. In particular,

Proposition 1.5. *([3], p.25) Let $\sigma^d(C) = \frac{1}{2}$ (a geodesic ball of radius $\frac{\pi}{2}$) then*

$$1 - \sigma^d(C_r) \leq e^{-(d-1)r^2/2}. \tag{1.7}$$

(Actually, this isn't really the best estimate. A better estimate involves the constant $(\pi/8)^{1/2}$. For more information see [5] and [1].)

Combining (1.6) and (1.7) one obtains

$$\sup\left\{1-\sigma^d(A_r) : A \in \mathcal{B}(S^d),\ \sigma^d(A) \geq \frac{1}{2}\right\} \leq e^{-(d-1)r^2/2}. \tag{1.8}$$

If we define

$$\alpha_{(S^d,\rho,\sigma^d)}(r)$$

we have a prototype of a "concentration function."

Remark 1.6. Define

$$v(r) = \sigma^d(C)$$

where C is a spherical cap of radius r with $0 \leq r \leq \pi$ then

$$v(r) = s_d^{-1}\int_0^r \sin^{d-1}\theta d\theta \text{ where } s_d = \int_0^p i\sin^{d-1}\theta d\theta.$$

Rescaling gives analogous inequalities for the sphere S_R^d with radius $R:$

$$\alpha_{(S_R^d,\rho,\sigma_R^d)}(r) \leq e^{-(d-1)r^2/(2R)^2} \text{ for } r > 0. \tag{1.9}$$

2. The Concentration Function

A key tool for quantifying the above concentration phenomena is the "concentration function." The general setting involves the following ingredients:

- (X,ρ) a metric space,
- $\mathbf{B}(X)$ the Borel sets, the smallest σ-algebra of subsets of X containing all open sets, and
- μ a probability measure on $\mathbf{B}(X)$ (with $\mu(X) = 1$).

Finally this allows us to define the concentration function

$$\alpha_{(X,\rho,\mu)}(\eta) := \sup\{a - \mu(A_\eta) : A \in \mathbf{B}(X),\ \mu(A) \geq \frac{1}{2}\} \tag{2.1}$$

where

$$A_\eta := \{x \in X : \rho(X,A) < \eta\}$$

is an open η-neighborhood of the set A.

Remark 2.1. We may note the following about $\alpha(\eta) = \alpha_{(X,\rho,\mu)}(\eta)$

(1) $\alpha(r) \leq \frac{1}{2}$ and $\alpha(0) = \frac{1}{2}$,
(2) $\alpha(r)$ decreases as r increases
(3) $\lim_{r\to\infty}\alpha(r) = 0$, and
(4) if $\operatorname{diam}(X) := \sup\{\rho(x,y) : x,y \in X\} < \infty$ and $r >$ $\operatorname{diam}(X)$, then $\alpha(r) = 0$.

Proof. The proofs of (1) and (2) are clear from the definition of $\alpha(r)$.

However, (3) follows from fixing a given $0 < \varepsilon < \frac{1}{2}$ and choosing $x \in X$ and $r > 0$ such that $\mu(X \setminus B(x,r)) < \varepsilon$. For any $A \in \mathbf{B}(X)$ with $\mu(A) \geq \frac{1}{2}$ then, one has

$$A \cap B(x,r) \neq \emptyset$$

and so therefore $A_{2r} \supset B(x,r)$ which implies

$$\begin{aligned}\alpha(2r) &= \sup\{\mu(X \setminus A_{2r}) : A \in \mathbf{B}(X), \mu(A) \geq \frac{1}{2}\} \\ &\leq \mu(X\ B(x,r)) \leq \varepsilon.\end{aligned}$$

The proof of (4) is also obvious. □

Remark 2.2. A frequently use relation is as follows: For any $A \in \mathbf{B}(X)$ with $\mu A \geq \frac{1}{2}$ one has

$$\mu(A_r) \geq 1 - \alpha_\mu(r) \text{ or } 1 - \mu(A_r) \leq \alpha_\mu(r) \tag{2.2}$$

in words, the smaller $\alpha_m u(r)$ is, the more mass is concentrated in $A_r \setminus A$. We could also say that $\alpha_\mu(r)$ bounds $1 - \mu(A_r)$ so long as $\mu(A) \geq \frac{1}{2}$.

The constant $\frac{1}{2}$ may seem arbitrary, but it's significance is explained by the following lemma.

Lemma 2.3. *Let μ be a probability measure on the Borel sets $\mathbf{B}(X)$ of (X,ρ). If*

$$\mu(A) \geq \varepsilon > 0, \tag{2.3}$$

and $r_0 > 0$ is such that

$$\alpha_\mu(r_0) < \varepsilon \tag{2.4}$$

then we have

$$1 - \mu(A_{r_0+r}) \leq \alpha_\mu(r).$$

Proof. The idea of this proof is to show that $\mu(A_{r_0}) \geq \frac{1}{2}$. To this end, let $B = X \setminus A_{r_0}$ so that $A \subset X\ B_{r_0}$. First we claim that $\mu(B) < \frac{1}{2}$. If this were not the case, then $\mu(B) \geq \frac{1}{2}$ and (2.4) would imply that

$$\mu(A) \leq \mu(X \setminus B_{r_0}) = 1 - \mu(B_{r_0}) \leq \alpha_\mu(r_0) < \varepsilon$$

which is a contraction to the assumption made in (2.3). Thus, $\mu(A_{r_0}) \geq \frac{1}{2}$. Therefore, by (2.2) we have that

$$1 - \mu(A_{r_)+r}) = 1 - \mu((A_{r_0})_r) \leq \alpha_\mu(r).$$

□

Of particular interest are situation where $\alpha_\mu(r)$ decays rapidly, as in the sphere case.

Definition 2.4. μ is said to have normal concentration on (X,ρ)if there exist constants $C, c > 0$ such that

$$\alpha_\mu(r) \leq Ce^{-cr^2},\ r > 0.$$

In the previous example we have $c = \frac{d-1}{2}$ which is dependent on d, thus σ^d had a normal concentration on (S^d,ρ).

Example 2.5. A second classical example involves $X = \mathbb{R}^d$ and $\rho(x,y) = |x-y|_2$ with $\mu = \gamma_d$ and $\mathrm{d}\mu = \gamma_d \mathrm{d}x$. In this situation we have now

$$\gamma_d(x) = (2\pi)^{-d/2} e^{-|x|_2^2/2} \tag{2.5}$$

which is the standard Gaussian density.

From this fact we have the following theorem for the corresponding concentration function.

Theorem 2.6. *For $(\mathbb{R}^d, |\cdot|_2, \gamma_2)$ one has*

$$\alpha_{\gamma_d}(r) \le e^{-r^2/2} \tag{2.6}$$

that is, the normal concentration holds.

Before we prove Theorem 2.6, it is helpful to note its relationship with the concentration result for the sphere, which was

$$\alpha_{(S^d,\rho,\sigma^d)}(r) \le e^{-(d-1)r^2/2},\ r > 0$$

or, rescaled as in (1.9),

$$\alpha_{(S^d_R,\rho,\sigma^d_R)}(r) \le e^{-(d-1)r^2/(2R^2)},\ r > 0.$$

Choosing $R = \sqrt{d}$, the bound in (1.9) tends to $e^{-r^2/2}$ as the number of dimensions increases. Curiously, this is also the bound for $(\mathbb{R}^d, |\cdot|_2, \gamma_d)$. The mystery ceases, however, when it is noted that

$$\sigma^d_{\sqrt{d}} \to \gamma_\infty.$$

More precisely, if one allows $\sigma^{d-1}_{\sqrt{d}}$ to be the uniform measure on $S^{d-1}_{\sqrt{d}}$ and $\pi_{N,d} : S^{N-1}_{\sqrt{N}} \to \mathbb{R}^d$ to be the canonical projection, then we can define the measure

$$\mu_{N,d} = \pi_{N,d}\sigma^{N-1}_{\sqrt{N}}$$

by

$$\int_{\mathbb{R}^d} f(x)\mathrm{d}\mu_{N,d}(x) := \int_{S^{N-1}_{\sqrt{N}}} f(\pi_{N,d}y)d\sigma^{N-1}_{\sqrt{N}}(y).$$

We can now claim that

$$\int_{\mathbb{R}^d} f(x)\mathrm{d}\mu_{N,d}(x) \overset{N\to\infty}{\to} (2\pi)^{-d/2}\int_{\mathbb{R}^d} f(x)e^{-|x|^2/2}dx. \tag{2.7}$$

A sketch of the proof of (2.7) is as follows: Let $y_i \in \mathcal{N}(0,1)$ for $i = 1,\dots,N$ be standard Gaussian random variables so that

$$\mathbb{E}(y_i^2) = \mathbb{E}((y_i = \mathbb{E}(y_i)^2)) = 1 = (2\pi)^{-1/2}\int_{\mathbb{R}} y_i^2 e^{-|y_i|^2/2}dy_i.$$

The law of large numbers then says

$$\frac{1}{N}(y_1^2 + \ldots + y_N^2) \to 1 \tag{2.8}$$

which may be reconsidered as

$$\mathbb{P}\{y : |\frac{1}{N}(y_1^2 + \ldots + y_N^2) - 1| > \varepsilon\} \to 0,\ N \to \infty.$$

Now, for each N we have that the quantities

$$\frac{\sqrt{N}(y_1,\dots,y_N)}{(y_1^2+\ldots+y_N^2)^{1/2}}$$

are uniformly distributed on $S^{N-1}_{\sqrt{N}}$ with $\sigma^{N-1}_{\sqrt{N}}$ so by definition

$$\frac{\sqrt{N}(y_1,\dots,y_d)}{(y_1^2+\ldots+y_N^2)^{1/2}}$$

are uniformly distributed with $\pi_{N,d}\sigma^{N-1}_{\sqrt{N}}$. In fact, we have

$$\frac{\sqrt{N}(y_1,\dots,y_d)}{(y_1^2+\dots+y_N^2)^{1/2}} \stackrel{N\to\infty}{\to} (y_1,\dots,y_d)$$

by (2.8).

Proof. The proof of Theorem 2.6 can be broken into the following two parts:

(I) Prove the following isoperimetric inequality

Theorem 2.7. *Let $A \in \mathcal{B}(\mathbb{R}^d)$ and $H = \{x \in \mathbb{R}^d : x \cdot u < \lambda\}$ where $\lambda \in \mathbb{R}$ and $u \in \mathbb{R}^d$ are fixed such that*

$$\gamma_d(A) = \gamma_d(H).$$

Then

$$\gamma_d(A_r) \geq \gamma_d(H_r),\ r > 0 \tag{2.9}$$

(II) Evaluate $\gamma_d(H_r)$ using rotational invariance and product structure.[2]

For now, let us assume that (I) is true and proceed with part (II). Consider the Gaussian distribution function

$$\phi(t) = (2\pi)^{-1/2}\int_{-\infty}^{t} e^{-x^2/2}\mathrm{d}x,\ t \in \mathbb{R}.$$

It suffices to consider $u = e_1 = (1,0,\dots,0)$ so that $H = \{x \in \mathbb{R}^d : x_1 < \lambda\}$. In that case,

$$\begin{aligned} \gamma_d H &= (2\pi)^{-1/2}\int_{-\infty}^{\lambda} e^{-|x_1|^2/2}\mathrm{d}x_1 \underbrace{(2\pi)^{\frac{-(d-1)}{2}}\int_{\mathbb{R}^{d-1}} e^{\frac{-(x_2^2+\cdots+x_d^2)}{2}}\mathrm{d}x_2\dots\mathrm{d}x_d}_{=1} \\ &= \phi(\lambda). \end{aligned} \tag{2.10}$$

Next, note that since $\phi(0) = \frac{1}{2}$, the fact that $\gamma_d H \geq \frac{1}{2}$ means $\lambda \geq 0$. Thus, (2.9) implies

$$\begin{aligned} 1-\gamma_d(A_r) &\leq 1-\gamma_d(H_r) = 1-\phi(\lambda+r) \\ &\leq 1-\phi(r) \\ &= (2\pi)^{-1/2}\int_r^{\infty} e^{-x^2/2}\mathrm{d}x \\ &= (2\pi)^{-1/2}\int_0^{\infty} e^{-(r+t)^2/2}\mathrm{d}t \\ &= (2\pi)^{-1/2}e^{-t^2/2}\int_0^{\infty} e^{-(t)^2/2}e^{-tr}\mathrm{d}t \\ &\leq e^{-r^2/2} \end{aligned}$$

which is (2.6) and proves Theorem 2.6. □

It remains now to prove Theorem 2.7, the isoperimetric inequality.

Proof. The proof of Theorem 2.7 also proceeds in two parts:

(I) Verify

[2] γ_d is a product measure so that $\gamma_d(x) = \gamma_1(x_1)\cdots\gamma_1(x_d)$.

Remark 2.8. (2.9) is equivalent to

$$\phi^{-1}(\gamma_d(A_r)) \geq \phi^{-1}(\gamma_d(A)) + r \tag{2.11}$$

(II) Prove the validity of (2.11).

The proof of Remark 2.8 comes first. The first step is to show that (2.11) implies (2.9). We begin by noting that (2.11) implies that

$$\begin{aligned} \gamma_d(A_r) &\geq \phi(\phi^{-1}(\gamma_d(A)) + r) \\ &= \phi(\phi^{-1}(\gamma_d(H)) + r) \\ &\overset{(2.10)}{=} \phi(\phi^{-1}(\phi(\lambda)) + r) \\ &= \phi(\lambda + r) \overset{(2.10)}{=} \gamma_d(H_r) \end{aligned}$$

which is (2.9). Conversely, (2.9) implies that

$$\begin{aligned} \phi^{-1}(\gamma_d(A_r)) &\geq \phi^{-1}(\gamma_d(H_r)) \overset{(2.10)}{=} \lambda + r \\ &= \phi^{-1}(\gamma_d(H)) + r \\ &= \phi^{-1}(\gamma_d(A)) + r \end{aligned}$$

which is (2.11). □

It now remains to verify (2.11). In the interest of time, a sketch suffices.

Using then Poincaré limit for $N \geq d$ (the uniform measure on expanding spheres tends to the Gaussian measure), we have

$$\sigma^{N-1}_{\sqrt{N}}(\pi^{-1}_{N,d}(A)) \to \gamma_d(A), \ N \to \infty.$$

For $\gamma_d(A)$ let $a > -\infty$ such that $\phi(a) - \gamma_d(A) = \gamma_1(-\infty, a)$ and choose any $b < a$. Then, for N large enough,

$$\begin{aligned} \sigma^{N-1}_{\sqrt{N}}(\pi^{-}_{N,d}1(A)) &> \gamma_d(A) - \varepsilon_1 = \gamma_1(-\infty, a) \\ &> \gamma_1(-\infty, b) + \varepsilon_2 \\ &> \sigma^{N-1}_{\sqrt{N}}(\underbrace{\pi^{-1}_{N,1}(-\infty, b)}_{\text{cap}}) \end{aligned} \tag{2.12}$$

since $b < \gamma_d(A)$ is arbitrary, we have $\gamma(A_r) \geq \phi(\gamma(A)) + r$.

Note that

$$\pi^{-1}_{N,d}(A_r) \supset (\pi^{-1}_{N,d}(A))_r.$$

Now the isoperimetric inequality on spheres gives

$$\begin{aligned} \sigma^{N-1}_{\sqrt{N}}(\pi^{-1}_{N,d}(A_r)) &\geq \sigma^{N-1}_{\sqrt{N}}((\pi^{-1}_{N,d}(A))_r) \\ &\overset{(2.12)}{\geq} \sigma^{N-1}_{\sqrt{N}}((\pi^{-1}_{N,1}(-\infty, b])_r). \end{aligned}$$

Now

$$\pi^{-1}_{N,1}(-\infty, b])_r = \pi^{-1}(-\infty, b + r(N)]$$

as $r(N) \to r$ implies that as $N \to \infty$ we have

$$\gamma(A_r) > \gamma_1(-\infty, b + r) = \phi(b + r).$$

This completes the proof of Theorem 2.7. □

Situations where isoperimetric inequalities can be invoked are, however, rare. A possible tool to prove concentration estimates in other settings is the following simple observation

Proposition 2.9. *Suppose $\phi : (X, \rho) \to (Y, \delta)$ is Lipschitz, that is,*

$$\delta(\phi(x), \phi(x')) \leq \|\phi\|_{L_p} \rho(x, x'),\ x, x' \in X.$$

Let μ be probability measure on $\mathcal{B}(X)$ and consider the push-forward on Y defined by

$$\mu_\phi(B) = \mu(\phi^{-1}(b)),\ \text{for } B \in \mathcal{B}(Y).$$

μ_ϕ can be seen to be a probability measure on $\mathcal{B}(Y)$. Then one has

$$\alpha_{(Y,\delta,\mu_\phi)}(r) \leq \alpha_{(X,rho,\mu)}(r/(\|\phi\|_{\mathrm{Lip}})). \tag{2.13}$$

Proof. Let $B \in \mathcal{B}$ such that $\frac{1}{2} \leq \mu_\phi(B) = \mu(\phi^{-1}(B))$. Then

$$1 - \mu_\phi(B_r) = 1 - \mu(\phi^{-1}(B_r)).$$

Now observe that

$$\phi^{-1}(B_r) \supset (\phi^{-1}(B))_{r/\|\phi\|_{\mathrm{Lip}}}. \tag{2.14}$$

In fact, suppose $\rho(x, \phi^{-1}(B)) < r/\|\phi\|_{\mathrm{Lip}}$ then $\delta(\phi(x), B) \leq \|\phi\|_{\mathrm{Lip}} r/\|\phi\|_{\mathrm{Lip}} = r$
Thus,

$$\begin{aligned} 1 - \mu_\phi(B_r) &\leq 1 - \mu((\phi^{-1}(B))_{r/\|\phi\|_{\mathrm{Lip}}}) \\ &\leq \alpha_{(X,\rho,\mu)}(r/\|\phi\|_{\mathrm{Lip}}) \end{aligned}$$

□

Example 2.10. As a simple application, let us consider the unit cube, $X = [0, 1]^d$. With a slight abuse of notation we write $|\cdot|$ to denote the Euclidean metric $\rho(x, y) = |x - y|$. Let $\mu = \mu^d$, the uniform Lebesgue measure. Define $\varphi : x \to (\phi(x_1), \ldots, \phi(x_d))$, with

$$\varphi(t) = (2\pi)^{-1/2} \int_{-\infty}^{t} e^{-x^2/2} \mathrm{d}x$$

then

$$\begin{aligned} |\phi(x) - \phi(x')| &= |\phi'(\xi)(x - x')| = |(2\pi)^{-1/2} e^{-\xi^2/2}(x - x')| \\ &\leq (2\pi)^{-1/2} |x - x'| \end{aligned}$$

so that $\|phi\|_{\mathrm{Lip}} = (2\pi)^{-1/2}$ which by Proposition 2.9 becomes

$$\alpha_{([0,1]^d,|\cdot|,\mu_d)}(r) \leq e^{-((2\pi)^{1/2} r)^2/2} = e^{-\pi r^2}. \tag{2.15}$$

Thus one has a normal concentration on cubes with normalized Lebesque measure.

A much more general result comes from [3, Thm 3.1].

Theorem 2.11. *Let (X, g) be a compact Riemmanian manifold with normalized Riemmanian metric μ. Then*

$$\alpha_{(X,g,\mu)}(r) \leq e^{-r\sqrt{\lambda_1}/3}$$

where $\lambda_1 > 0$ is the first nontrivial eigenvalue of the Δ on X.

(For instance, when $X = S^d$, the unit sphere of dimension d then $\lambda_1 = d$, so the above estimate is not sharp.)

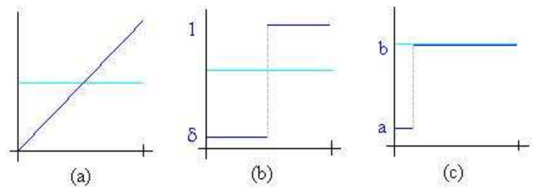

FIGURE 1. Various examples of functions and their medians.

3. Deviation Estimates

3.1. **The Role of the Median.** The key idea of this section is that well-behaved functions of many variables deviate significantly from their *median* on sets of small measure.

Given that the setting is (X, ρ, μ), m_f is called a *median* of f (measurable) with respect to μ, if

$$\mu(\{x : f(x) \leq m_f\}) \geq \frac{1}{2} \text{ and } \mu(\{x : f(x) \geq m_f\}) \geq \frac{1}{2}. \tag{3.1}$$

Of course in general, the median is different from the mean and need not be unique.

Example 3.1. If we let $X = [0, 1]$, $\mathrm{d}\mu(x) = \mathrm{d}x$, and $\rho(x, y) = |x - y|$ then

(a) for $f(x) = x$, $m_f = \frac{1}{2} =$ average (Figure 1(a));

(b) for

$$f(x) = \begin{cases} \delta, & x \in (0, \frac{1}{2}] \\ 1, & x \in (\frac{1}{2}, 1) \end{cases}$$

$m_f \in [\delta, 1]$ (Figure 1(b));

(c) for

$$f(x) = \begin{cases} a, & x \in (0, x_0] \\ b, & x \in (x_0, 1) \end{cases}$$

where $x_0 < \frac{1}{2}$, then $m_f = b$ (Figure 1(c)).

The modulus of continuity for $f \in C(X, \mathbb{R})$ is defined by

$$\omega(\eta, f) = \sup\{|f(x) - f(y)| : \rho(x, y) \leq \eta\},\ \eta > 0.$$

Now Levy's inequality can be stated as follows.

Lemma 3.2. *(Levy's inequality) For any median m_f of f one has*

$$\mu(\{x : |f(x) - m_f| \geq \omega(\eta, f)\}) \leq 2\alpha_{(X,\rho,\mu)}(\eta). \tag{3.2}$$

Proof. Defining

$$\underline{A}^f := \{x : f(x) \leq m_f\}$$

and

$$\overline{A}^f := \{x : f(x) \geq m_f\}$$

note that

$$\begin{cases} \underline{A}^f_\eta \subseteq \{x : f(x) \leq m_f + \omega(\eta, f)\} \\ \overline{A}^f_\eta \subseteq \{x : f(x) \geq m_f - \omega(\eta, f)\} \end{cases}. \tag{3.3}$$

Thus

$$\begin{aligned}\mu(\{x : f(x) > m_f + \omega(\eta, f)\}) &= 1 - \mu(\{x : f(x) \le m_f + \omega(\eta, f)\}) \\ &\le 1 - \mu(\underline{A}_\eta^f) \\ &\le \alpha_{(X,\rho,\mu)}(\eta).\end{aligned}$$

and

$$\begin{aligned}\mu(\{x : f(x) < m_f - \omega(\eta, f)\}) &= 1 - \mu(\{x : f(x) \ge m_f - \omega(\eta, f)\}) \\ &\le 1 - \mu(\overline{A}_\eta^f) \\ &\le \alpha_{(X,\rho,\mu)}(\eta)\end{aligned}$$

whence the assertion easily follows. □

3.2. **Lipschitz Functions.** Let us denote by $\|f\|_{\mathrm{Lip}}$ the smallest constant C such that

$$|f(x) - f(y)| \le C\rho(x, y),\ x, y \in X.$$

Theorem 3.3. *Let f be Lipschitz (X, ρ, μ) as before*

(a) *One has the deviation inequality*

$$\mu(\{x : |f(x) - m_f| \ge r\}) \le 2\alpha_{(X,\rho,\mu)}(\frac{r}{\|f\|_{\mathrm{Lip}}}) \tag{3.4}$$

(b) *If there exists some constant C such that*

$$\mu(\{x : f(x) \le C\}) \ge \varepsilon > 0 \tag{3.5}$$

and if $r_0 > 0$ is such that

$$\alpha_{X,\rho,\mu}(0) \le \frac{1}{2} \tag{3.6}$$

then

$$\mu(\{x : f(x) \ge C + r_0 + r\}) \le \alpha_{X,\rho,\mu}(\frac{r}{\|f\|_{\mathrm{Lip}}}) \tag{3.7}$$

(c) *(Converse result) Let $\|f\|_{\mathrm{Lip}} = 1$ (1-Lipschitz). If for some $\alpha : \mathbb{R}_+ \to \mathbb{R}_+$ the inequality*

$$\mu(\{x : f(x) \ge m_f + r\}) \le \alpha(r)$$

holds for all 1-Lipschitz functions f then

$$\alpha_{(X,\rho,\mu)}(r) \le \alpha_(r).$$

Proof.

(a) Use $\{x : |f(x) - m_f| \ge r\} \subseteq \{x : |f(x) - m_f| \ge \omega(\frac{r}{\|f\|_{\mathrm{Lip}}}, f)\}$ and Lemma 3.1.

(b) Let $A := \{x : f(x) \le C\}$ show as in Lemma 2.1 that $\mu(A_{r_0/\|f\|_{\mathrm{Lip}}}) \ge \frac{1}{2}$. In fact, otherwise $B := X \setminus A_{r_0/\|f\|_{\mathrm{Lip}}}$ would satisfy $\mu(B) > \frac{1}{2}$ and $B_{r_0/\|f\|_{\mathrm{Lip}}} \subseteq X \setminus A$, yet

$$\begin{aligned}\varepsilon &\le \mu(A) = 1 - \mu(X \setminus A) \\ &\le 1 - \mu(B_{r_0/\|f\|_{\mathrm{Lip}}}) \\ &\le \alpha_{(x,\rho,\mu)}(\frac{r_0}{\|f\|_{\mathrm{Lip}}}) < \varepsilon\end{aligned}$$

which is a contradiction. Thus, since, for $x \in A_{b/\|f\|_{\text{Lip}}}$ and $z(x) \in \partial A$ such that $\rho(x, z(x)) = \rho(x, A)$ gives

$$f(x) = f(z(x)) + f(x) - f(z(x)) \le C + \rho(x, z(x))\|f\|_{\text{Lip}} \le C + b.$$

Thus we conclude that

$$\begin{aligned} \mu(\{x : f(x) \ge C + r_0 + r\}) &= 1 - \mu(\{x : f(x) < C + r_0 + r\}) \\ &\le 1 - \mu((A_{r_0/\|f\|_{\text{Lip}}})_{r/\|f\|_{\text{Lip}}})) \\ &\le \alpha_{(X,\rho,\mu)}(\frac{r}{\|f\|_{\text{Lip}}}) \end{aligned}$$

(c) Let $\mu(A) \ge \frac{1}{2}$ and $f(x) := \rho(x, A)$ then $f(x)$ is 1-Lipschitz with median $m_f = 0$, and thus

$$\begin{aligned} \mu(\{x : f(x) \ge m_f + r\}) &= \mu(\{x : \rho(x, A) \ge r\}) \\ &= 1 - \mu(\{x : \rho(x, A) < r\}) \\ &= 1 - \mu(A_r). \end{aligned}$$

□

Several interesting corollaries ensue.

Corollary 3.4. *For $A, B \in \mathbf{B}(X)$ one has*

$$\mu(A)\mu(B) \le 4\alpha_{(X,\rho,\mu)}\left(\frac{\rho(A, B)}{2}\right)$$

where $\rho(A, B) := \inf\{\rho(x, y) : x \in A, y \in B\}$

Proof. Take $f(x) := \rho(x, B)$, which we know to be 1-Lipschitz. This implies that $f = 0$ on B and $f \ge \rho(A, B)$ on A. Now note that

$$\begin{aligned} \mu(A)\mu(B) &\le \mu \otimes \mu(\{(x, y) : |f(x) - f(y)| \ge 2r\}) \\ &\le 2\mu(\{|f - m_f| \ge r\}) \le 4\alpha_{(X,\rho,\mu)}(r) \end{aligned} \quad (3.8)$$

(3.9)

□

Corollary 3.5. *If f is 1-Lipschitz, then*

$$\mu \otimes \mu(\{(x, y) : |f(x) - f(y)| \ge r\}) \le 4\alpha_{(X,\rho,\mu)}(r/2).$$

Conversely, if

$$\mu \otimes \mu(\{(x, y) : |f(x) - f(y)| \ge r\}) \le \alpha(r)$$

for all 1-Lipschitz functions, then $\alpha_{(X,\rho,\mu)} \le 2\alpha$.

Proof. The first assertion follows from (3.8). The second assertion follows from taking $A \in \mathbf{B}(X)$, $\mu(A) \ge \frac{1}{2}$, and $B = X \setminus A_r$ such that $\rho(A, B) = r$ and using the first part of (3.8) to find that

$$\mu(A)(1 - \mu(A_r)) \le \mu \otimes \mu\{(x, y) : |f(x) - f(y)| > r\}$$

for $f(x) = \rho(x, A)$, which implies the conclusion. □

3.3. **Deviation from the Mean.** One can bootstrap the above arguments to estimate deviation from the mean.

Proposition 3.6. *Let f be measurable on $\mathcal{B}(X)$. Assume that for some $a_f \in \mathbb{R}$ and $\alpha : \mathbb{R}_+ \to \mathbb{R}_+$ we have $lim_{r\to\infty}\alpha(r) = 0$ and that*

$$\mu(\{x : |f(x) - a_f| \geq r\}) \leq \alpha(r). \tag{3.10}$$

Then

(a) *one can infer the following deviation estimates from medians*

$$\mu(\{x : |f(x) - m_f| \geq r + r_0\}) \leq \alpha(r) \tag{3.11}$$

where m_f is a median of f for μ and $r_0 > 0$ is sufficiently large to ensure that $\alpha(r_0)) < \frac{1}{2}$.

(b) *If*

$$\bar{\alpha} = \int_0^\infty \alpha(s)ds$$

then $f \in L_1(X, \mu)$ and $|a_f - \int f d\mu| \leq \bar{\alpha}$. Moreover,

$$\mu(\{x : |f(x) - \int f d\mu| \geq r + al\bar{p}ha\}) \leq \alpha(r). \tag{3.12}$$

(c) *If $\alpha(r) \leq Ce^{-cr^p}$, for some $0 < p < \infty$ then for $M \in \{m_f, \int_X f d\mu\}$,*

$$\mu(x : |f(x) - M| > r\}) \leq C'e^{-\kappa_p cr^p} \tag{3.13}$$

where C' depends on c and p, and κ_p depends only on p.

Proof.

(a) : The justification of (3.11) is based on the following remark.

Remark 3.7. (3.10) and the inequality $\alpha(r_0) < \frac{1}{2}$ imply that

$$|a_f - m_f| \leq r_0. \tag{3.14}$$

Proof of remark. Let $B = \{x : |f(x) - a_f| < r_0\}$. By assumption $\mu(X \setminus B) \leq \alpha(r_0) < \frac{1}{2}$ in other words, $\mu(B) > \frac{1}{2}$. Hence

$$B \cap \underline{A}^f \neq \emptyset \text{ and } B \cap \overline{A}^f \neq \emptyset. \tag{3.15}$$

Suppose that (3.14) was not true, so instead we would have

$$m_f - a_f > r_0 \text{ or } a_f - m_f > r_0.$$

In the first case, where $m_f > a_f$, we have $x \in \overline{A}^f$, which implies that $f(x) \geq m_f > a_f + r_0$ while $x \in B$ implies that $f(x) - a_f = |f(x) - a_f| < r_0$. These two statements contradict each other, so this case is impossible.

Likewise, assuming that $a_f - m_f > r_0$ turns out to contradict the assumption that $B \cap \underline{A}^f \neq \emptyset$. □

Now that we have that remark, observe

$$\begin{aligned} |f(x) - m_f| &\leq |f(x) - a_f| + |a_f - m_f| \\ &\leq |f(x) - a_f| + r_0, \end{aligned}$$

so that

$$\{x : |f(x) - m_f| \geq r + r_0\} \subseteq \{x : |f(x) - a_f| \geq r\}$$

hence

$$\mu(\{x : |f(x) - m_f| \geq r + r_0\}) \leq \mu(\{x : |f(x) - a_f| \geq r\}) \leq \alpha(r)$$

which is (3.11).

(b) Define $\|f\|_p^p := \int_X |f(x)|^p d\mu(x)$ and use the fact that for $p \geq 1$

$$\|f\|_p^p = p \int_0^\infty t^{p-1} \mu(\{x : |f(x) \geq t|\}) dt. \tag{3.16}$$

Especially for $p = 1$ this implies

$$\begin{aligned} \int_X |f(x) - a_f| d\mu(x) &= \int_0^\infty \mu(\{x : |f(x) - a_f| \geq r\}) dr \\ &\overset{(3.10)}{\leq} \int_0^\infty \alpha(r) dr = \bar{\alpha} < \infty. \end{aligned} \tag{3.17}$$

Thus,

$$\|f\|_1 - |a_f| = \int_X (|f(x)| - |a_f|) d\mu \leq \bar{\alpha}$$

shows that f is integrable. Moreover,

$$|\int_X f \mathrm{d}\mu - a_f| \leq \int_X |f(x) - a_f| \mathrm{d}\mu \leq \bar{\alpha}.$$

To see (3.12) note that

$$\begin{aligned} |f(x) - \int f \mathrm{d}\mu| &\leq |f(x) - a_f| + (a_f - \int f \mathrm{d}\mu) \\ &\leq |f(x) - a_f| + \overline{\alpha} \end{aligned}$$

so that

$$\{a : |f(x) - \int f \mathrm{d}\mu| \geq r + \overline{\alpha}\} \subset \{x : |f(x) - a_f| \geq r\}$$

and finally,

$$\mu(\{x : |f(x) - \int f d\mu| \geq r + al\bar{p}ha\}) \leq \alpha(r)$$

which is what was to be shown.

(c) These are straightforward estimates.

□

References

1. M. Gromov, *Paul lévy's isoperimetric inequality*, Preprint, 1980.
2. ———, *Metric structures for Riemannian and non-Riemannian spaces*, Birkhausen, 1998.
3. M. Ledoux, *The concentration of measure phenomenon*, Math. Survey and Mono. **89** (2001).
4. V. Milman and G. Scheditman, *Asymptotic theory of finite dimensional normed spaces*, Lecture Notes in math., no. 1200, Springer.
5. E. Schmidt, *Die Brunn-Minkowskische Ungleichung un ihr Spiegelbild sowie die isoperimetrische Eigenschaft de Kugel in der Euklisch en und nicht-Euklidischen Geometrie*, Math. Nachr. **1** (1948), 81–157.

University of South Carolina
E-mail address: dahmen@igpm.rwth-aachen.de