Lecture 1. Poincaré, Cheeger, and some other inequalities.

Cheeger's inequality. Let μ be a probability measure on $\mathbb{R}^n$, or more genrally on some metric space (X,d) equipped with its Borel σ-field. The isoperimetric problem for μ asks the following questions: Among sets of given measure, which sets have minimal perimeter? There are several possible notions of perimeter. For our purposes, the most convenient one is the exterior Minkowski content, defined as follows: for every measurable subset A of the ambient space we let

$$\mu_+(A) = \liminf_{\varepsilon \to 0} \frac{\mu(A_\varepsilon \backslash A)}{\varepsilon}.$$

where A_ε is the ε-neighborhood of A, namely the set of points whose distance to A is at most ε. The exact answer to the isoperimetric problem is only known in a handful of very specific cases. For instance, for the uniform on the sphere equipped with the geodesic distance, spherical caps (i.e. geodesic balls) are the solution. This is usually attributed to P. Lévy (1922). The answer is also known on Gauss space, and this time affine halfspaces solve the isoperimetric problem. This was proved in 1975 by Sudakov and Tsirelson, and independently by Borell [5]. In general solving exactly the isoperimetric problem is hopeless and we content ourselves with a more modest task, such as finding lower bounds on the perimeter of a set A in terms of its measure. When this lower bound is linear, we say that μ satisfies Cheeger's inequality.

Definition 1. We say that μ satisfies Cheeger's inequality if there is a constant C such that

$$\min(\mu(A), 1 - \mu(A)) \leq C\mu_+(A), \tag{1}$$

for every measurable set A. The smallest C such that this holds true is called the Cheeger constant, and we denote it ψ_μ below.

Cheeger's inequality can be seen as an L^1-Poincaré inequality. Indeed we have the following result.

Lemma 2. *Inequality (1) is equivalent to the following:*

$$\min_{c \in \mathbb{R}} \int_X |f - c| \, d\mu \leq C \int_X |\nabla f| d\mu, \tag{2}$$

for every Lipschitz function f.

Remark. In the right-hand side the quantity $|\nabla f(x)|$ should be interpreted as the local Lipschitz constant of f, namely

$$|\nabla f(x)| = \limsup_{y \to x} \frac{|f(x) - f(y)|}{d(x,y)}.$$

This only make sense in a metric space with no isolated points. Actually we will only investigate the case $X = \mathbb{R}^n$ equipped with its usual Euclidean metric from now on.

Remark. It is well known that the infimum in the left-hand side is attained at any median for f, i.e. any real c such that both $\mu(f \leq c)$ and $\mu(f \geq c)$ are at least 1/2.

Proof. We only give a proof sketch, and refer to Bobkov and Houdré [4] (for instance) for more details. The derivation of (2) from (1) relies on the co-area formula: for any Lipschitz f we have

$$\int_X |\nabla f| \, d\mu \geq \int_{\mathbb{R}} \mu^+(f > t) \, dt.$$

In most cases this inequality is actually an equality, but we only need this inequality, which admits a soft proof, again see [4]. Applying Cheeger's inequality to the right-hand side then yields (2). For the converse implication, given a set A, we apply (2) to some suitable Lipschitz approximation of the indicator function of A. A bit more precisely, we pick $\varepsilon_n \to 0$ such that

$$\lim \frac{\mu(A_{\varepsilon_n} \backslash A)}{\varepsilon_n} \to \mu_+(A),$$

we pick another positive sequence (δ_n) tending to 0 (for instance $\delta_n = 1/n$) and we observe that the sequence (f_n) given by

$$f_n = \left(1 - \frac{1}{(1-\delta_n)\varepsilon_n} \cdot d(x, A_{\delta_n \varepsilon_n})\right)_+$$

satifies $0 \leq f_n \leq 1$ for every n, $f_n \to \mathbb{1}_{\overline{A}}$ pointwise, and $\limsup \int |\nabla f_n| \, d\mu \leq \mu_+(A)$. Applying (2) to f_n and letting n tend to $+\infty$ yields (1) after some computation. □

From this version of Cheeger's inequality it is relatively straightfoward to see that Cheeger's inequality is stronger than the Poincaré inequality. Recall from the first lecture of Bo'az that we say that μ satisfies Poincaré if there is a constant C such that

$$\operatorname{var}_\mu(f) \leq C \int_{\mathbb{R}^n} |\nabla f|^2 \, d\mu$$

for every Lipschitz function f. Also we let $C_P(\mu)$ be the best constant C such that this holds true.

Proposition 3 (Cheeger 1970)**.** *Let μ be a probability measure on $\mathbb{R}^n$ satisfying the Cheeger inequality. Then μ satisfies Poincaré, and we have*

$$C_P(\mu) \leq 4\psi_\mu^2.$$

Remark. Maybe it is unfortunate but our convention for the Cheeger constant and Poincaré constant do not have the same homogeneity. The Cheeger constant of a probability measure on $\mathbb{R}^n$ is 1-homogeneous, if we scale μ by a factor λ then the Cheeger constant is multiplied by λ. One the other hand the Poincaré constant is 2-homogeneous.

Proof. Assume f is Lipschitz and bounded, and has its median at 0. Applying (2) to f_+^2 we get

$$\int_{\mathbb{R}^n} f_+^2 \, d\mu \leq \psi_\mu \int_{\mathbb{R}^n} |\nabla f_+^2| \, d\mu = 2\psi_\mu \int_{\mathbb{R}^n} f_+ |\nabla f_+| \, d\mu$$

Applying Cauchy-Schwartz we get

$$\int_{\mathbb{R}^n} f_+^2 \, d\mu \leq 4\psi_\mu^2 \int_{\mathbb{R}^n} |\nabla f_+|^2 \, d\mu = 4\psi_\mu^2 \int_{\mathbb{R}^n} |\nabla f|^2 \mathbb{1}_{\{f>0\}} \, d\mu$$

We can do the same with f_- and adding up the two inequalities yields the result. □

The converse inequality is not true in general, one can cook up examples on the line. However it turns out that if we restrict to log-concave measures then the converse is true. This is a result of Buser from 1982, to which we will come back later on in this lecture.

Semigroup tools. Let μ be a probability measure on $\mathbb{R}^n$. We do not need log-concavity for now but let us assume that μ is supported on the whole space and has a smooth density ρ. Letting $V = -\log \rho$ be the potential of μ, the Laplace operator associated to μ is the differential operator given by

$$L_\mu = \Delta - \nabla V \cdot \nabla,$$

initially defined on the space of compactly supported smooth functions. For such functions, an integration by parts gives

$$\int_{\mathbb{R}^n} (L_\mu f) g \, d\mu = - \int_{\mathbb{R}^n} \nabla f \cdot \nabla g \, d\mu.$$

This shows in particular that L_μ is symmetric and that $-L_\mu$ is a monotone (unbounded) operator on $L^2(\mu)$. Moreover this operator is known to be essentially self-adjoint, in the sense that its minimal extension is self-adjoint. By a slight abuse of notation we still call L_μ this extension. A bit more explicitly, we call $\mathcal{D}$ the space of functions $f \in L^2(\mu)$ for which there exists a sequence (f_n) of smooth compactly supported functions such that $f_n \to f$ and $(L_\mu f_n)$ converges. The limit of $L_\mu f_n$ does not depend on the choice of the converging sequence (f_n) (this is an immediate consequence of the symmetry of L_μ) and we set $L_\mu f = \lim L_\mu f_n$. The fact that this new L_μ is self adjoint is not quite immediate, not every monotone operator is essentially self adjoint. This has to do with elliptic regularity, we refer to [1, Corollary 3.2.2] for the details. From the integration by parts above we can see that if (f_n) and $(L_\mu f_n)$ converge then also ∇f_n converges. This means that the domain $\mathcal{D}$ contains $H^1(\mu)$ and that the integration by parts $\langle L_\mu f, g \rangle = -\langle \nabla f, \nabla g \rangle$ remains valid for every f, g in the domain. Here the inner product is the one from $L^2(\mu)$, and when we apply it to tensors it has to be interpreted coordinate wise. Being self-adjoint and monotone (negative) the operator L_μ admits a spectral decomposition

$$L_\mu = - \int_0^\infty \lambda \, dE_\lambda. \tag{3}$$

The semigroup associated to L_μ is then defined as

$$P_t = \mathrm{e}^{tL_\mu} = \int_0^\infty \mathrm{e}^{-t\lambda} \, dE_\lambda.$$

For fixed t the operator P_t is a self-adjoint bounded operator in $L^2(\mu)$ and we have the semigroup property $P_t \circ P_s = P_{t+s}$. If f is a fixed function of $L^2(\mu)$ the function $F(t,x) = P_t f(x)$ is the solution to the parabolic equation

$$\begin{cases} F(0,\cdot) = f \\ \partial_t F = L_\mu F, \end{cases}$$

at least in a weak sense.

We now move on to the probabilistic representation of the semigroup (P_t). Consider the diffusion (X_t) given by

$$dX_t = \sqrt{2} \cdot dW_t - \nabla V(X_t) \, dt, \tag{4}$$

where (W_t) is standard Brownian motion. Then (X_t) is a Markov process, and (P_t) is the corresponding semigroup. Namely for every test function f we have

$$P_t f(x) = \mathbb{E}_x f(X_t)$$

where the subscript x next to the expectation denotes the starting point of (X_t). This allows to prove inequalities for the semigroup (P_t) using probabilistic techniques.

Lemma 4. *If μ is log-concave then Lipschitz functions are preserved along the semigroup, and moreover $\|P_t f\|_{\mathrm{Lip}} \leq \|f\|_{\mathrm{Lip}}$ for every f and every $t > 0$.*

Proof. Let $x, y \in \mathbb{R}^n$, and let (X_t^x) and (X_t^y) be two solutions of the SDE (4) using the same Brownian motion, but starting at two different points x and y. This is called parallel coupling. Then the process $(X_t^x - Y_t^x)$ is an absolutely continuous function of t (the Brownian part cancels out). Moreover, thanks to the convexity of V,

$$\frac{d}{dt}|X_t^x - X_t^y|^2 = -2(X_t^x - X_t^y) \cdot (\nabla V(X_t^x) - \nabla V(X_t^y)) \leq 0.$$

So the distance $|X_t^x - X_t^y|$ is almost surely decreasing. Therefore its expectation is also decreasing, and in particular

$$\mathbb{E}|X_t^x - X_t^y| \leq |x - y|.$$

Now suppose f is a Lipschitz function. Then from the previous inequality we get

$$|P_t f(x) - P_t f(y)| = |\mathbb{E} f(X_t^x) - \mathbb{E} f(X_t^y)| \leq \mathbb{E}|f(X_t^x) - f(X_t^y)| \leq \|f\|_{\mathrm{Lip}} \cdot |x - y|,$$

which is the result. □

I believe the next result is originally due to Varopoulos [15].

Proposition 5. *Suppose μ is log-concave. Then for every bounded function f and every $t > 0$ the function $P_t f$ is Lipschitz and moreover*

$$\|P_t f\|_{\mathrm{Lip}} \leq \frac{1}{\sqrt{t}} \cdot \|f\|_\infty.$$

Proof. Again we use a coupling argument. Suppose that f is a bounded function. Fix $x, y \in \mathbb{R}^n$, and let (X_t^x) and (X_t^y) be two processes solving the SDE (4) initiated at x and y respectively. Then

$$|P_t f(x) - P_t f(y)| \leq \mathbb{E}|f(X_t^x) - f(X_t^y)| \leq 2\|f\|_\infty \cdot \mathbb{P}(X_t^x \neq X_t^y). \tag{5}$$

It remains to choose a coupling for which the right-hand side is small. Parallel coupling is awful here, as it actually prevents X_t^x and X_t^y from meeting. Instead, we choose the Brownian increment for X_t^y to be the reflection of that of X_t^x with respect to the hyperplane $(X_t^x - X_t^y)^\perp$. If (W_t) is the Browian motion for X_t^x, the equation for X_t^y is thus

$$dX_t^y = \sqrt{2} \cdot \left(\mathrm{Id} - 2v_t^{\otimes 2}\right) dW_t - \nabla V(X_t^y)\, dt$$

where (v_t) is the unit vector $(X_t^x - X_t^y)/|X_t^x - X_t^y|$. Actually we do so until the first time (denoted τ) when the two processes meet. After time τ we just set $X_t^y = X_t^x$. We will not justify properly here why this is well defined, but this coupling technique, usually referred to as *mirror coupling*, is a relatively standard tool, see for instance [11]. Itô's formula shows that up to the coupling time τ the equation for the distance between the two processes is

$$d|X_t^x - X_t^y| = -2\sqrt{2} v_t \cdot dW_t - v_t \cdot (\nabla V(X_t^x) - \nabla V(X_t^y))\, dt.$$

Itô's term vanishes because the Brownian increment takes place in a direction where the Hessian matrix of the norm vanishes. Once again, in the log-concave case the second term from the right hand side is negative. Notice also that $B_t := \int_0^t v_s \cdot dW_s$ is a standard (one dimensional) Brownian motion. Therefore up to the coupling time τ we have

$$|X_t - Y_t| \leq |x-y| - 2\sqrt{2} B_t,$$

where (B_t) is some standard one diemnsional Brownian motion. Therefore

$$\mathbb{P}(X_t \neq Y_t) = \mathbb{P}(\tau > t) \leq \mathbb{P}\left(\forall s \leq t \colon B_s < \frac{|x-y|}{2\sqrt{2}} \right).$$

By the reflection principle for the Brownian motion

$$\mathbb{P}\left(\exists s \leq t \colon B_s \geq \frac{|x-y|}{2\sqrt{2}} \right) = 2 \cdot \mathbb{P}\left(B_t \geq \frac{|x-y|}{2\sqrt{2}} \right) = \mathbb{P}\left(|g| \geq \frac{|x-y|}{2\sqrt{2t}} \right)$$

where g is a standard Gaussian variable. Hence the inequality

$$\mathbb{P}(X_t \neq Y_t) \leq \Psi\left(\frac{|x-y|}{2\sqrt{2t}} \right),$$

where $\Psi(r) = (2/\pi)^{1/2} \int_0^r \mathrm{e}^{-u^2/2}\, du$ is the distribution function of $|g|$. Recalling (5) and taking the supremum over x, y gives

$$\|P_t f\|_{\mathrm{Lip}} \leq \frac{1}{\sqrt{2t}} \cdot \sup_{a>0} \left\{ \frac{\Psi(a)}{a} \right\} \cdot \|f\|_\infty .$$

The expression inside the sup is decreasing, so the sup equals the limit as a tends to 0, which is $(2/\pi)^{1/2}$. We thus get the desired inequality (even with a better constant than announced). □

The next corollary is taken from Ledoux [9].

Corollary 6. *If μ is log-concave, then for every locally Lipschitz function f we have*

$$\|f - P_t f\|_{L^1(\mu)} \leq 2\sqrt{t} \cdot \||\nabla f|\|_{L^1(\mu)}.$$

Also for every set measurable set A we have

$$\mu(A)(1-\mu(A)) = \mathrm{var}_\mu(\mathbb{1}_A) \leq \sqrt{2t} \cdot \mu^+(A) + \mathrm{var}_\mu(P_t \mathbb{1}_A).$$

Proof. Let f be a Lipschitz function and g be a smooth bounded function. Using the fact that the semigroup is self adjoint, and the integration by part formula, we get

$$\langle f - P_t f, g \rangle = \langle f, g - P_t g \rangle = - \int_0^t \langle f, L P_s g \rangle \, dt = \int_0^t \langle \nabla f, \nabla P_s g \rangle \, ds.$$

By the previous proposition,

$$\langle \nabla f, \nabla P_s g \rangle \leq \||\nabla f|\|_{L^1(\mu)} \cdot \|P_s g\|_{\mathrm{Lip}} \leq \frac{1}{\sqrt{s}} \||\nabla f|\|_{L^1(\mu)} \|g\|_\infty .$$

Integrating between 0 and t and plugging back in the previous display we get

$$\langle f - P_t f, g \rangle \leq 2\sqrt{t} \cdot \| |\nabla f| \|_{L^1(\mu)} \|g\|_\infty,$$

which is the result. For the second inequality, applying the first one to a suitable Lipschitz approximation of the indicator function of A, as in the proof of Lemma 2, we get

$$\| \mathbb{1}_A - P_t \mathbb{1}_A \|_1 \leq 2\sqrt{t} \cdot \mu^+(A).$$

Moreover, using reversibility, it is not hard to see that

$$\| \mathbb{1}_A - P_t \mathbb{1}_A \|_1 = 2 \left(\mathrm{var}_\mu(\mathbb{1}_A) - \mathrm{var}_\mu(P_{t/2} \mathbb{1}_A) \right).$$

Hence the result. □

A result of E. Milman. We said earlier that the inequality $C_P(\mu) \leq C\psi_\mu^2$ can be reversed in the log-concave case. Actually we will prove a much stronger statement, which is due to E. Milman.

Definition 7. If μ is a probability measure on $\mathbb{R}^n$, the function

$$I_\mu \colon r \in [0,1] \mapsto \inf\{ \mu_+(\partial S) \colon \mu(S) = r \}.$$

is called the isoperimetric profile of μ.

With this definition Cheeger's inequality can be rewritten

$$\psi_\mu \cdot I_\mu(r) \geq \min(r, 1-r).$$

The following is a deep result from geometric measure theory.

Theorem 8. *The isoperimetric profile of a log-concave measure is concave.*

We will use this as a blackbox, we refer to the appendix of [12] for an historical account and the relevant references. Another good reference for this is Bayle's Ph.D. thesis [2] (if you read french). This has important implications for us. Indeed, since the isoperimetric profile is non negative, its concavity implies that

$$I_\mu(t) \geq 2 \cdot I_\mu(1/2) \min(t, 1-t).$$

In particular the Cheeger constant of μ satisfies

$$\psi_\mu \leq \frac{1}{2 \cdot I_\mu(1/2)}. \tag{6}$$

Therefore, for a log-concave measure, in order to prove Cheeger's inequality, it is enough to look at the perimeter of sets of measure $1/2$. Combining this information with the results from the previous section we arrive at the following.

Theorem 9. *If μ is log-concave, then there exist a 1-Lischitz function f satifying*

$$\|f\|_\infty^2 \approx \mathrm{var}_\mu(f) \approx \psi_\mu^2 \approx C_P(\mu).$$

Here the symbol $\approx$ means that the ratio between the two quantities is comprised between two universal constants. The theorem asserts in particular that the Cheeger constant and the Poincaré constant are of the same order, which is the result of Buser that we mentioned earlier. This result is essentially due to E. Milman [12]. The proof we give is very much inspired by Ledoux's proof of Buser's inequality [9].

Proof. By (6) if A is a set of measure $1/2$ that has near minimal surface, say up to a factor 2, then

$$\mu_+(A) \leq \frac{1}{\psi_\mu}. \tag{7}$$

Let $t > 0$. By Corollary 6, and since $\mu(A) = 1/2$,

$$\frac{1}{4} \leq \sqrt{2t} \cdot \mu_+(A) + \mathrm{var}_\mu(P_t \mathbb{1}_A) \leq \frac{\sqrt{2t}}{\psi_\mu} + \mathrm{var}_\mu(P_t \mathbb{1}_A)$$

If t is a sufficiently small multiple of ψ_μ^2 we thus get $\mathrm{var}_\mu(P_t \mathbb{1}_A) \geq \frac{1}{8}$ (say). On the other hand, by Proposition 5,

$$\|P_t \mathbb{1}_A\|_{\mathrm{Lip}} \leq \frac{1}{\sqrt{t}} \leq \frac{C}{\psi_\mu},$$

for some constant C. Putting everything together we see that the function $f = (\psi_\mu / C) \cdot P_t \mathbb{1}_A$ is 1-Lipschitz and satisfies

$$\psi_\mu^2 \lesssim \mathrm{var}_\mu(f) \leq \|f\|_\infty^2 \lesssim \psi_\mu^2,$$

Hence $\psi_\mu^2 \approx \mathrm{var}_\mu(f) \approx \|f\|_\infty^2$. On the other hand since f is 1-Lipschitz, applying Poincaré to f yields $\mathrm{var}_\mu(f) \leq C_P(\mu)$. Since we always have $C_P(\mu) \leq 4\psi_\mu^2$ we indeed get the Buser inequality $\psi_\mu^2 \approx C_P(\mu)$. □

In the last part of the proof, we upper bounded $\|\nabla f\|_2$ by the Lipschitz constant of f, which is very wasteful. So the theorem actually yields a lot more. It implies that it is enough to bound the variance of Lipschitz functions to get Poincaré (or Cheeger). More precisely, we get the following.

Corollary 10 (E. Milman [12])**.** *For any log concave measure μ*

$$\psi_\mu^2 \approx C_P(\mu) \approx \sup \{ \mathrm{var}_\mu(f) \colon \|f\|_{\mathrm{Lip}} \leq 1 \}.$$

Bo'az will use this later on this weak, and also give another proof that avoids the concavity of the isoperimetric profile blackbox. Let us point out though that this corollary does not use the full strength of Theorem 9, it does not use the information about the L^∞ norm of f. So we actually have stronger form of the Corollary. Namely, in the log-concave case, to get Cheeger, or Poincaré, it is enough to bound the variance of a bounded Lipschitz function whose Lipschitz constant is 1, and whose L^∞-norm is of the same order as its standard deviation.

Related material that I didn't have time to cover.

Concentration of measure.

Definition 11. Let (X, d, μ) be a metric measured space. The concentration function of μ is defined by

$$\alpha_\mu \colon r \mapsto \sup \{1 - \mu(S_r) \colon \mu(S) = 1/2\}$$

where S_r is the r-neighborhood of the set S.

In some very specific models such as the uniform measure on the sphere or the Gaussian measure the exact value of the concentration function is known. In general it is hopeless to compute it exactly and we are happy with an upper bound for α_μ. The most interesting types of upper bound for us are the case of Gaussian concentration and of exponential concentration.

Definition 12. We say that μ satisfies Gaussian concentration if there exist constants C_0, C_1 such that

$$\alpha_\mu(r) \leq C_0 \cdot \exp\left(-\frac{r^2}{C_1}\right), \quad \forall r \geq 0.$$

We say that μ satisfies exponential concentration if there exist constants C_0, C_1 such that

$$\alpha_\mu(r) \leq C_0 \cdot \exp\left(-\frac{r}{C_1}\right), \quad \forall r \geq 0.$$

In the following the prefactor C_0 is always of order 1 and regarded as irrelevant. The constant that matters is the one inside the exponent. We call it the Gaussian or exponential concentration constant.

We are interested here in concentration properties of log-concave measures on $\mathbb{R}^n$. Gaussian concentration cannot be true in general (think of μ being the exponential measure) but there is no obstruction to having exponential concentration with a dimension free constant for isotropic log-concave measures, and this is in fact equivalent to the KLS conjecture that Bo'az introduced this morning. Indeed, it is well-known that the Poincaré inequality yields exponential concentration, and more precisely that for any probability measure μ on $\mathbb{R}^n$ satisfying the Poincaré inequality we have

$$\alpha_\mu(r) \leq C \cdot \exp\left(-\frac{r}{L \cdot \sqrt{C_P(\mu)}}\right), \quad \forall r \geq 0,$$

where C and L are universal constants. We will skip the derivation of this from Poincaré here, but this is not very hard, see for instance [1, section 4.4.2].
Once again, in the log-concave case this implication can be reversed. Indeed, by E. Milman's theorem (Corollary 10) the Poincaré constant of is a largest variance of a 1-Lipschitz function (up to a constant). If f is 1-Lipshitz, by definition of the concentration function we have

$$\mu(f - m \geq r) \leq \alpha_\mu(r),$$

for every $r > 0$, and where m is a median for f. From this we obtain easily

$$\mathrm{var}_\mu(f) \leq 4 \int_0^\infty r \alpha_\mu(r)\, dr.$$

Therefore, in the log-concave case

$$C_P(\mu) \lesssim \int_0^\infty r \cdot \alpha_\mu(r)\, dr. \tag{8}$$

This implies in particular that the Poincaré constant of μ and the exponential concentration constant squared are actually of the same order.

Log-Sobolev and Talagrand. We have seen earlier that Poincaré is weaker than Cheeger in general but equivalent to it within the class of log-concave measures. We'll see now that log-concavity also allows to reverse the hierarchy between the log-Sobolev inequality and the transportation inequality. A probability measure μ on $\mathbb{R}^n$ is said to satisfy the logarithmic Sobolev inequality if there exists a constant $C > 0$ such that

$$D(\nu \mid \mu) \leq \frac{C}{2} I(\nu \mid \nu)$$

for every probability measure ν, where $D(\nu \mid \mu)$ and $I(\nu \mid \mu)$ denote the relative entropy and Fisher information, respectively

$$D(\nu \mid \mu) = \int_{\mathbb{R}^n} \log(\frac{d\nu}{d\mu})\, d\nu \quad \text{and} \quad D(\nu \mid \mu) = \int_{\mathbb{R}^n} |\nabla \log(\frac{d\nu}{d\mu})|^2\, d\nu.$$

The best constant C is called the log-Sobolev constant, denoted $C_{LS}(\mu)$ below. The factor $1/2$ is just a matter of convention. With this convention the log-Sobolev constant of the standard Gaussian 1. This is a stronger inequality than Poincaré. More precisely we have $C_P(\mu) \leq C_{LS}(\mu)$ for any μ. This is easily seen by applying log-Sobolev to a probability measure whose density with respect to μ is $1 + \varepsilon f$ and letting ε tend to 0. Not every log-concave measure satisfy log-Sobolev, simply because log-Sobolev implies sub-Gaussian tails, so for instance the exponential measure (on $\mathbb{R}$) does not statisfy log-Sobolev. A bit more precisely, log-Sobolev implies Gaussian concentration: if μ satisfies log-Sobolev then for any set S we have

$$\mu(S)(1 - \mu(S_r)) \leq \exp\left(-c \cdot \frac{r^2}{C_{LS}(\mu)} \right).$$

Again see [1] for a proof.

Recall that if μ, ν are probability measures on $\mathbb{R}^n$, the quadratic transportation cost from μ to ν is defined as

$$T_2(\nu, \nu) = \inf \left\{ \int_{\mathbb{R}^n \times \mathbb{R}^n} |x - y|^2\, d\pi \right\},$$

where the infimum is taken over every coupling π of μ and ν, namely every probability measure on the product space whose marginals are μ and ν. Tomorrow Bo'az will speak about the Monge transport cost, which is the L^1 version of this.

Proposition 13 (Otto and Villani [14]). *If μ satisfies log-Sobolev then for every probability measure ν we have*

$$T_2(\nu, \mu) \leq 2C_{LS}(\mu) \cdot D(\nu \mid \mu).$$

This transportation/entropy inequality is sometimes called Talagrand's inequality, as it was first established by Talagrand for the Gaussian measure (before the work of Otto and Villani). Again in the log-concave case the implication log-Sobolev/Talagrand can be reversed. Indeed, we have the following, also due to Otto and Villani.

Proposition 14. *If μ is log-concave then*

$$D(\nu \mid \mu) \leq \sqrt{T_2(\nu,\mu)\cdot I(\nu \mid \mu)}.$$

This is only a particular case of the Otto-Villani result, there's also a version for semi-log-concave measures, namely measures for which we have a possibly negative lower bound on the Hessian of the potential. This inequality goes by the name HWI. The reason for this name is not apparent from our choice of notations, but relative entropy is often denoted H, and the transport cost T_2 can also be denoted W_2 or rather W_2^2 (for Wasserstein). From the HWI inequality we see that the implication between log-Sobolev and Talagrand can be reversed for log-concave measures: if we happen to know

$$T_2(\nu,\mu) \leq C_2 D(\nu \mid \mu)$$

for μ log-concave, then we get log-Sobolev for μ and $C_{LS}(\mu) \leq 2C_2$. We will not spell out the proofs of the Otto-Villani results here and we refer to [14] (see also [3]).

We've seen above that the equivalence between Cheeger and Poincaré can be considerably reinforced. This is also the case here, and this is yet again a result of E. Milman.

Theorem 15 (E. Milman [13])**.** *For a log-concave measure μ we have equivalence between Gaussian concentration and the log-Sobolev inequality, and moreover the log-Sobolev constant and the Gaussian concentration constant are within a universal factor of each other.*

Proof. There are several proofs of this result in the literature, see [13, 10]. The proof sketch that we give here is is taken from Gozlan, Roberto, Samson [8]. Before spelling it out, let us first explain why Talagrand's inequality implies Gaussian concentration. By some convex duality principle T_2 can be also expressed as a supremum, namely

$$T_2(\mu,\nu) = \sup_f \left\{ \int_{\mathbb{R}^n} Q_{1/2} f \, d\mu - \int_{\mathbb{R}^n} f \, d\nu \right\}$$

where the $Q_t f$ is the infimum convolution of f with some multiple of the distance squared:

$$Q_t f(x) = \inf_{y\in\mathbb{R}^n} \left\{ f(y) + \frac{1}{2t}|x-y|^2 \right\}.$$

It can also be shown that (Q_t) is a semigroup of operators, namely we have $Q_s Q_t = Q_{s+t}$. There's also some duality between the log-Laplace transform and the relative entropy:

$$\log \int_{\mathbb{R}^n} \mathrm{e}^f \, d\mu = \sup_\nu \{ \int_{\mathbb{R}^n} f \, d\nu - D(\nu \mid \mu) \},$$

where the supremum is taken over every probability measure ν. Using this, it is pretty easy to see that Talagrand's inequality

$$T_2(\nu,\mu) \leq 2C_T \cdot D(\nu \mid \mu), \quad \forall \nu$$

is equivalent to

$$\int_{\mathbb{R}^n} \exp(Q_{C_T} f) \, d\mu \leq \exp\left(\int_{\mathbb{R}^n} f \, d\mu \right), \quad \forall f.$$

Applying this to both $Q_{C_T} f$ and $-Q_{C_T} f$, using the fact that (Q_t) is a semigroup, and multiplying the two inequalities together we get

$$\int_{\mathbb{R}^n} \exp(Q_{C_T}(-Q_{C_t} f)\, d\mu \cdot \int_{\mathbb{R}^n} \exp(Q_{2C_T} f)\, d\mu \leq 1.$$

But clearly $Q_t(-Q_t f) \leq -f$ so we obtain

$$\int_{\mathbb{R}^n} \exp(-f)\, d\mu \cdot \int_{\mathbb{R}^n} \exp(Q_{2C_T} f)\, d\mu \leq 1.$$

Applying to $f = -\log \mathbb{1}_A$ we get

$$\int_{\mathbb{R}^n} \exp\left(\frac{d(x,A)^2}{2C_T}\right) dx \leq \frac{1}{\mu(A)},$$

for every set A. By Markov inequality this implies

$$\alpha_\mu(r) \leq 2 \cdot \exp\left(-\frac{r^2}{2C_T}\right).$$

So Talagrand implies Gaussian concentration, and moreover the Gaussian concentration constant is at most the constant in Talagrand, up to a factor 2. Now we want to reverse this, so we assume

$$\alpha_\mu(r) \lesssim \mathrm{e}^{-r^2/C_G}.$$

I'm using this notation to emphasize the fact that I will not keep track of the dependence on the prefactor in front of the exponential. It is easily seen to imply

$$\int_{\mathbb{R}^n} \exp(Q_{2C_G} f)\, d\mu \lesssim \exp(m_f).$$

for every f, where m_f is a median for f. Again, applying this $-Qf$ and Qf and multiplying we get

$$\int_{\mathbb{R}^n} \mathrm{e}^{-f}\, d\mu \cdot \int_{\mathbb{R}^n} \exp(Q_{2C_G} f)\, d\mu \lesssim 1,$$

hence by Jensen's inequality

$$\int_{\mathbb{R}^n} \exp(Q_{2C_G} f)\, d\mu \lesssim \exp\left(\int f\, d\mu\right).$$

In other words we get the dual version of Talagrand, but with some prefactor. In terms of transport and entropy this gives

$$T_2(\nu, \mu) \lesssim C_G(D(\nu \mid \mu) + 1).$$

So we have an additional additive constant in the right-hand side of Talagrand. So far we've not used log-concavity, this would be true for any measure satisfying Gaussian concentration. Now assuming log-concavity, we can plug this into HWI. We get

$$D(\nu \mid \mu) \lesssim C_G \cdot I(\nu \mid \mu) + 1.$$

Again, we get some weak form of log-Sobolev with an additional constant term in the right-hand side. This is sometimes called non-tight log-Sobolev inequality. To get rid of that constant, observe first that we clearly have from the first theorem of E. Milman (see equation (8))

$$C_P(\mu) \lesssim C_G.$$

Moreover, non-tight log-Sobolev can be reformulated as

$$\mathrm{ent}(f^2) \lesssim C_G \int_{\mathbb{R}^n} |\nabla f|^2 \, d\mu + \int f^2 \, d\mu,$$

where the entropy of a non negative function f is defined as

$$\mathrm{ent}_\mu(f) = \int_{\mathbb{R}^n} f \log f \, d\mu - \left(\int_{\mathbb{R}^n} f \, d\mu \right) \log \left(\int_{\mathbb{R}^n} f \, d\mu \right).$$

Now there is a nice inequality by Rothaus, which states that for any $f\colon \mathbb{R}^n \to \mathbb{R}$ and any constant c we have

$$\mathrm{ent}_\mu((f+c)^2) \le \mathrm{ent}_\mu(f^2) + 2 \int_{\mathbb{R}^n} f^2 \, d\mu,$$

(look for the keyword Rothaus lemma in [1]). Using this inequality it is easy to see that our non tight version of log-Sobolev and the bound that we have on $C_P(\mu)$ altogether imply

$$\mathrm{ent}_\mu(f) \lesssim C_G \int_{\mathbb{R}^n} |\nabla f|^2 \, d\mu,$$

which is a reformulation of the desired log-Sobolev inequality. □

References

[1] Bakry, D., Gentil, I., Ledoux, M., *Analysis and geometry of Markov diffusion operators.* Springer, 2014.

[2] Bayle, V., *Propriétés de concavité du profil isopérimétrique et applications*, Ph. D., Université Joseph Fourier (Grenoble), 2003.

[3] Bobkov, S.G., Gentil, I, Ledoux, M., *Hypercontractivity of Hamilton-Jacobi equations.* J. Math. Pures Appl. (9) 80, No. 7, (2001), 669-696.

[4] Bobkov. S., Houdré, C., *Isoperimetric constants for product probability measures.* Ann. Probab. 25 (1), 184-205 (1997).

[5] Borell, C., *The Brunn-Minkowski inequality in Gauss space*, Invent. Math. 30, 207-216 (1975).

[6] Buser, P., *A note on the isoperimetric constant.* Ann. Sci. École Norm. Sup., Vol. 15, (1982), 213—230.

[7] Cheeger, J., *A lower bound for the smallest eigenvalue of the Laplacian.* Problems in analysis, Princeton Univ. Press, (1970), 195—199.

[8] Gozlan, N.; Roberto, C.; Samson, P.-M. *From concentration to logarithmic Sobolev and Poincaré inequalities.* J. Funct. Anal. 260, No. 5, 1491-1522 (2011).

[9] Ledoux, M. *A simple analytic proof of an inequality by P. Buser.* Proc. Am. Math. Soc. 121, No. 3, 951-959 (1994).

[10] Ledoux, M. *From concentration to isoperimetry: semigroup proofs.* In Houdré, Christian (ed.) et al., Concentration, functional inequalities and isoperimetry. Contemporary Mathematics 545, 155-166 (2011).

[11] T. Lindvall, L.C.G. Rogers, *Coupling of Multidimensional Diffusions by Reflection.* Ann. Probab. 14, No. 3, 860-872 (1986).

[12] Milman, E., *On the role of convexity in isoperimetry, spectral gap and concentration.* Invent. Math. 177, No. 1, 1-43 (2009).

[13] , Milman, E. *Isoperimetric and concentration inequalities: equivalence under curvature lower bound.* Duke Math. J. 154, No. 2, 207-239 (2010).

[14] Otto, F.; Villani, C, *Generalization of an inequality by Talagrand and links with the logarithmic Sobolev inequality.* J. Funct. Anal. 173, No. 2, 361-400 (2000).

[15] Varopoulos, N. T., *Small time Gaussian estimates of heat diffusion kernels I: the semigroup technique.* Bull. Sci. Math. (2) 113 (1989), 253-277.