Course Notes

Part VI

Probabilistic Combinatorics

and

Algorithms

J. A. Verstraete
Department of Mathematics
University of California San Diego
9500 Gilman Drive
La Jolla California 92037-0112

jacques@ucsd.edu
Talagrand’s Inequality

An isoperimetric inequality is an inequality which relates the volume of a set to its surface area. In this section, we are interested in isoperimetric inequalities which state that if $A$ is a subset of a metric space, then the set of points at distance at least $t$ from $A$ is large relative to $A$. There is now a very large body of research on isoperimetric inequalities and their applications to combinatorics, probability, geometric functional analysis, and so forth. We therefore concentrate our attention on the discrete cube, $Q_n$. One of the fundamental isoperimetric inequalities was established by Harper (1966):

**Theorem 1** Let $A \subset Q_n$, and let $A(s) = \{|x \in Q_n : d_H(x, A) \leq s\}$ and $B_t = \{|x \in Q_n : d_H(x, \emptyset) \leq t\}$. If $|A| \geq |B_t|$ then $|A(s)| \geq |B_t(s)|$.

Note that $B_t(s) = B_{t+s}$ and $|B_t| = \sum_{k \leq t} \binom{n}{k}$, and $d_H$ is the Hamming metric on $Q_n$, namely $d_H(x, y) = |x \triangle y|$. Accordingly, the Hamming distance between sets $A, B \subset Q_n$ is

$$d_H(A, B) = \min_{x \in A, y \in B} |x \triangle y|.$$

Harper’s Inequality can be read as saying that either $A$ is small or $|A(s)|$ is small, where $A(s)$ is the set of elements of $Q_n$ at distance more than $s$ from $A$. If we choose each coordinate in $Q_n$ independently with probability $\frac{1}{2}$, then we get a probabilistic version of Harper’s Inequality, involving the binomial expression for $|B_{t+s}|$ and $|B_t|$. Rather than stating this inequality, we derive a cleaner statement of probability using the Lipschitz Inequality for martingales. The proof relies on the observation that $d_H(x, A)$ is 1-Lipschitz, and since that is all that is needed, we can think later of more general metrics that $d_H(x, A)$.

**Theorem 2** If $A \subset Q_n$ and $B = \overline{A(t)}$, then

$$\frac{|A|}{2^n} \cdot \frac{|B|}{2^n} \leq e^{-\frac{t^2}{n}}.$$

**Proof** The expression on the left is a product of two probabilities. Now the function $f(x) = d_H(x, A)$ is 1-Lipschitz. By the Lipschitz Inequality of Chapter 6, with $c_i = 1$ for $i \in [n]$, $\lambda > 0$, and $\mu = \mathbb{E}(f)$, $\mathbb{P}(f - \mu \geq \lambda) \leq \exp(-\frac{2\lambda^2}{n})$. Now $\mathbb{P}(A) = \mathbb{P}(f = 0)$ and $\mathbb{P}(B) = \mathbb{P}(f \geq t)$ so

$$\mathbb{P}(A) \mathbb{P}(B) = \mathbb{P}(f = 0) \mathbb{P}(f \geq t)$$

$$\leq \mathbb{P}(f - \mu \leq -\mu) \mathbb{P}(f - \mu \geq t - \mu)$$

$$\leq e^{-\frac{2\mu^2}{n}} \cdot \frac{2(t - \mu)^2}{n}$$

$$\leq e^{-\frac{t^2}{n}}.$$

In the last line we note that $\mu = \frac{1}{2}t$ gives a maximum. This completes the proof.
In general, isoperimetric inequalities, such as Theorem 2, also give concentration inequalities. For example, if $X$ is a 1-Lipschitz random variable on $Q_n$ and each co-ordinate is zero or one with probability $\frac{1}{2}$, then for any number $s$ and any $t > 0$ we have

$$\mathbb{P}(X \leq s)\mathbb{P}(X \geq s + t) \leq e^{-\frac{t^2}{n}}.$$ 

Indeed, we just let $A$ be the event $X \leq s$ and let $B$ be the event $X \geq s + t$. Then $B \subset \overline{A(t)}$, and an application of Theorem 2 gives the above inequality.

Talagrand’s inequality (1995) is an isoperimetric equality on a product space but with a different but very powerful notion of distance. Let $(Ω_i, F_i, P_i)$ be probability spaces and let $Ω = \prod_{i=1}^{n} Ω_i$ be endowed with the product measure, $P$. The Talagrand distance from $x \in Ω$ to a set $A \subset Ω$ is

$$d_T(x, A) = \sup_{\|α\| = 1} \min_{y \in A} \sum_{i: x_i \neq y_i} α_i.$$ 

Here the supremum is over all unit vectors $α \in \mathbb{R}^n$ on the unit sphere $S(1)$, with the Euclidean norm. One of the most useful facts is that in the Talagrand definition of distance, the distance from $x$ to $A$ can be checked by allowing $α$ to depend on $x$ – we can choose our favourite $α$ for each $x$ to get bounds on the distance. The following further important facts are straightforward to prove:

- For all $x \in Ω$ and for all $A \subset Ω$,
  $$d_T(x, A) \leq d_H(x, A) \leq \sqrt{n}d_T(x, A).$$

- If $Ω = Q_n$, then the Talagrand distance from $x$ to $A$ is the Euclidean distance from $x$ to the convex hull of $A$ in $\mathbb{R}^n$.

So from the first statement, $d_T(x, A) \leq \ell$ implies there exists $y \in A$ such that $x$ and $y$ differ in at most $\ell\sqrt{n}$ co-ordinates.

The original form of Talagrand’s inequality, which is now one of many, is the following. Let $A_T(s) = \{x \in Q_n : d_T(x, A) \leq s\}$ – this is the analogue of $A(s)$ for the Hamming metric. By definition, $A_T(0) = A$.

**Theorem 3** Let $t$ be a positive real number and let $Ω$ be a product probability space with product measure $P$. Let $A \subset Ω$ and let $B = \overline{A_T(t)}$. Then

$$\mathbb{P}(A) \cdot \mathbb{P}(B) \leq e^{-t^2/4}.$$ 

An equivalent geometric definition of Talagrand’s Inequality is as follows: let $P(y, x)$ denote the set of zero-one vectors $z$ of length $n$ which are one on those coordinates $i$ for which $x_i \neq y_i$. For a set $A$ and a vector $x \in Ω$, let $P(A, x)$ be the union of all $P(y, x)$ with $y \in A$. Here is an equivalent definition of $d_T(x, A)$. Denote by $\text{conv} S$ the convex hull of a set $S \subset \mathbb{R}^n$. 

\[2\]
Lemma 4 Let $A \subset \Omega$ and let $x \in \Omega$. Then
\[ d_T(x, A) = \min \{ \| z \| : z \in \text{conv} P(A, x) \}, \]
where $\text{conv} X$ denotes the convex hull of a set $X \subset \mathbb{R}^n$.

Proof $\triangleright$ For the first statement, let $z^\star$ achieve the minimum. Recall
\[ d_T(x, A) = \sup_{\| \alpha \| = 1} \min_{z \in P(A, x)} \langle \alpha, z \rangle. \]
First we claim that there is a unit vector $\alpha$ such that $\langle \alpha, z \rangle \geq \| z^\star \|$ for all $z \in P(A, x)$. This shows that $\| z^\star \|$ is a lower bound for the Talagrand distance. By definition of $z^\star$, the hyperplane perpendicular to $z^\star$ separates $\text{conv} P(A, x)$ from the origin, so $\langle z, z^\star \rangle \geq \langle z^\star, z^\star \rangle$ for any $z \in \text{conv} P(A, x)$. Taking $\alpha = z^\star/\| z^\star \|$, we have $\| \alpha \| = 1$ and this proves the claim. To prove the lemma, we show that $\| z^\star \|$ is an upper bound for the Talagrand distance. To see this, note that there are positive real numbers $\lambda_i$ summing to one such that $z^\star = \sum_{i=1}^{n} \lambda_i z_i$ where $z_i \in P(A, x)$. Taking the inner product with any $\alpha$: $\| \alpha \| = 1$ we get
\[ \| z^\star \| \geq \langle \alpha, z^\star \rangle = \sum_{i=1}^{n} \lambda_i \langle \alpha, z_i \rangle \]
which means that one of the inner products in the sum is at most $\| z^\star \|$. \hfill \blacksquare

Let $x\omega$ denote the concatenation of a vector $x \in \prod_{i=1}^{n-1} \Omega_i$ with $\omega \in \Omega_n$.

Lemma 5 Let $\Omega = \prod_{i=1}^{n} \Omega_i$ be a product probability space, and $x \in \Omega$. Define $A_{\omega} = \{ x : x\omega \in A \}$ where $\omega \in \Omega_n$, and let $B = \bigcup \{ A_z : z \in \Omega_n \}$. Then for all $\lambda \in [0, 1]$,
\[ d_T(x\omega, A)^2 \leq (1 - \lambda)^2 + \lambda d_T(x, A_{\omega})^2 + (1 - \lambda) d_T(x, B)^2. \]

Proof $\triangleright$ Since $y \in P(B, x)$ implies $y1 \in P(A, x\omega)$ and $z \in P(A_{\omega}, x)$ implies $z0 \in P(A, x\omega)$, the same applies when we take convex hulls. In other words, for any $y \in \text{conv} P(B, x)$ and $z \in \text{conv} P(A_{\omega}, x)$, we have $((1 - \lambda)y + \lambda z, 1 - \lambda) \in \text{conv} P(A, x\omega)$. Consequently,
\[ d_T(x\omega, A)^2 \leq (1 - \lambda)^2 + \|(1 - \lambda)y + \lambda z\|^2 \leq (1 - \lambda)^2 + (1 - \lambda)\|y\|^2 + \lambda\|z\|^2. \]
We now take $y$ and $z$ which achieve equality in Lemma 4. \hfill \blacksquare

We combine the above two lemmas to prove Talagrand’s Inequality.
1.1 Proof of Talagrand’s Inequality

Let \( X(x) = d_T(x, A) \). We show, by induction on \( n \) that \( \mathbb{P}(A) \cdot \mathbb{E}(e^{X^2/4}) \leq 1 \). This is sufficient to prove Talagrand’s inequality, since we can apply Markov’s inequality to obtain

\[
\mathbb{P}(A) \cdot \mathbb{P}(\bar{A}_T(t)) = \mathbb{P}(A) \cdot \mathbb{P}(X > t) \leq \mathbb{P}(A) \cdot \mathbb{E}(e^{X^2/4})e^{-t^2/4} \leq e^{-t^2/4}.
\]

For \( n = 1 \), the result follows easily. Suppose \( n > 1 \). By Lemma 5,

\[
\mathbb{E}(e^{X^2/4}) = \int e^{\frac{1}{2n}d_T^2(x, \omega)}dx \omega
\]

\[
\leq \min_{\lambda \in [0,1]} e^{(1-\lambda)^2/4} \int [e^{\frac{1}{2n}d_T(x, \omega)}]^\lambda \cdot [e^{\frac{1}{2n}d_T(x, B)}]^{1-\lambda}
\]

\[
\leq \min_{\lambda \in [0,1]} e^{(1-\lambda)^2/4} \left[ \int e^{\frac{1}{2n}d_T(x, \omega)}^\lambda \right] \cdot \left[ \int e^{\frac{1}{2n}d_T(x, B)}^{1-\lambda} \right] \text{ Hölder’s Inequality}
\]

\[
\leq \min_{\lambda \in [0,1]} e^{(1-\lambda)^2/4} \mathbb{P}(A_\omega)^{-\lambda} \mathbb{P}(B)^{-1-\lambda} \text{ by induction.}
\]

\[
= \frac{1}{\mathbb{P}(B)} \cdot \min_{\lambda \in [0,1]} e^{(1-\lambda)^2/4} \left( \frac{\mathbb{P}(A_\omega)}{\mathbb{P}(B)} \right)^{-\lambda}.
\]

Let \( \lambda = 1 + 2 \log \frac{\mathbb{P}(A_\omega)}{\mathbb{P}(B)} \) so that the above expression is (after a calculation) at most

\[
\frac{\mathbb{P}(A_\omega)}{\mathbb{P}(B)} \left( 2 - \frac{\mathbb{P}(A_\omega)}{\mathbb{P}(B)} \right).
\]

Since \( \mathbb{P}(A_\omega) \leq \mathbb{P}(B) \) and \( \mathbb{P}(A) \leq \mathbb{P}(B) \), this expression is at most \( 1/\mathbb{P}(A) \), as required. \( \blacksquare \)

2 Certifiable functions

Let \( \Omega = \prod_{i=1}^n \Omega_i \) be a product space, let \( X \) be a random variable on \( \Omega \) and let \( f : \mathbb{N} \rightarrow \mathbb{N} \). Then \( X \) is \( f \)-certifiable if the event \( X(\omega) \geq s \) can be verified by checking at most \( f(s) \) co-ordinates of \( \omega \). In other words, there is a set of \( f(s) \) co-ordinates such that any \( \omega' \) equal to \( \omega \) on these co-ordinates also has \( X(\omega') \geq s \).

**Theorem 6** For \( s, t \in \mathbb{R} \), and any \( f \)-certifiable \( k \)-Lipschitz random variable \( X \) on \( \Omega = \prod_{i=1}^n \Omega_i \),

\[
\mathbb{P}(X \leq s - kt \sqrt{f(s)}) \cdot \mathbb{P}(X \geq s) \leq e^{-t^2/4}.
\]

**Proof** Let \( A \) be the event \( X < s - kt \sqrt{f(s)} \). By Talagrand’s inequality, we only need to prove that the event \( X \geq s \) is contained in \( A_T(t) \). Suppose, for a contradiction, that there exists \( \omega \in A_T(t) \) such that \( X(\omega) \geq s \). Since \( X \) is \( f \)-certifiable, there is a set \( I \) of indices of size \( m \leq f(s) \) which certifies \( X \geq s \). By taking \( a_i = 1/\sqrt{m} \) for \( i \in I \) in the definition of \( d_T(x, A) \), we deduce that there exists an \( \omega' \in A \) differing from \( \omega \) in at most \( t \sqrt{f(s)} \) co-ordinates of \( I \). Now suppose \( \omega'' \) agrees with \( \omega \) on \( I \) and agrees with \( \omega' \) outside of \( I \), then \( X(\omega'') \geq s \) because \( X \) is \( f \)-certifiable.
Now $\omega''$ and $\omega'$ differ in at most $t \sqrt{f(s)}$ co-ordinates, all of which are in $I$. Since $X$ is $k$-Lipschitz, this means
\[ X(\omega') \geq X(\omega'') - kt \sqrt{f(s)} \geq s - kt \sqrt{f(s)}. \]
However, this means $\omega' \not\in A$, which is a contradiction. The continuity of $e^{-t^2/4}$ ensures that $X < s - kt \sqrt{f(s)}$ can be replaced by $X \leq s - kt \sqrt{f(s)}$.

A similar inequality is valid if we replace $s - kt \sqrt{f(s)}$ with $s + kt \sqrt{f(s)}$ in the theorem. In general, the parameter $s$ in the theorem is taken to be the median of $X$, so that Talagrand’s inequality gives an upper bound for the probability of deviation below the median of a random variable. Practically speaking, the median is more difficult to determine than the mean, but the following lemma says that the median $\mathbb{M}(X)$ of an $f$-certifiable $k$-Lipschitz random variable $X$ and its mean are close together.

**Lemma 7** Let $X$ be a random variable, and let $\Omega$ be a product space with product probability measure $\mathbb{P}$. Suppose $X$ is $k$-Lipschitz and $f(s) = ds$ certifiable. Then there is a constant $\gamma > 0$ such that
\[ |\mathbb{E}(X) - \mathbb{M}(X)| \leq \gamma k \sqrt{d \mathbb{E}(X)}. \]

The proof of this lemma is left as an exercise.

### 2.1 Monotone subsequence problem

Let $\sigma$ be a sequence (permutation) of distinct letters in $[n]$ of length $n$. A natural question is the length of a longest increasing or decreasing subsequence of $\sigma$. Such a sequence is called monotone. Erdős and Szekeres proved that the length of a longest monotone subsequence in any sequence $\sigma$ has length at least $\sqrt{n}$ (we leave this as an exercise). Now suppose we take a permutation of $[n]$ uniformly and randomly, and let $X_n$ denote the length of a longest monotone subsequence. We show that $X_n$ is concentrated around its mean. In fact, this follows from Talagrand’s inequality. First, we generate our random permutation by taking $n$ real numbers $y_1, y_2, \ldots, y_n$ uniformly from the interval $[0,1]$, and then order them in increasing order. This gives a permutation of $[n]$ according to the order of the subscripts in the increasing order of the $y_i$s, because the chance that two of real numbers are equal is zero. To apply Theorem 6, note that $X_n \geq s$ if and only if there is a monotone increasing subsequence of $\sigma$ of length $s$, which can be certified by looking at the $s$ real numbers corresponding to that subsequence. Also, upon changing one of the real numbers, the value of $X_n$ changes by at most one, so $X_n$ is 1-Lipschitz. Therefore, by Theorem 6, if $\omega(n) \to \infty$, then by Lemma 7
\[ \mathbb{P}(|X_n - \mathbb{E}(X_n)| > \omega(n) \sqrt{\mathbb{E}(X)}) \to 0. \]

What is the expected length of a longest monotone subsequence? It turns out that $\mathbb{E}(X_n) \sim 2\sqrt{n}$, as shown by Vershik and Kerov, but no short proof of this fact is known. Much more precise statements are now known about the distribution of $X_n$: the limiting distribution was found by Baik, Deift and Johansson. A survey of this and related problems is given by Stanley.
2.2 Disjoint Triangles

For a vertex $v$ in a graph $G$, let $\triangle(G)$ denote the maximum number of edge-disjoint triangles in $G$. In the random graph $G$ with $n$ vertices and edges appearing independently with probability $p = \lambda/n$, $\triangle(G)$ is a 1-Lipschitz function and is $f(s) = 3s$ certifiable: revealing the $3k$ edges of $k$ edge-disjoint triangles certifies $\triangle(G) \geq k$, and any graph agreeing with $G$ on these edges also has $\triangle(G) \geq k$. By Talagrand’s Inequality, if $m$ is the median of $\triangle(G)$,

$$P(\triangle(G) \leq m - t \sqrt{3m}) \leq 2e^{-\frac{1}{4}t^2}.$$ 

It turns out that if $\lambda \rightarrow \infty$, then $E(\triangle(G)) \sim \frac{1}{6}pn^2$. So using Lemma 7, $G$ almost surely has a collection of edge-disjoint triangles covering most of its edges.

2.3 Colouring Graph Powers

Brook’s Theorem (1941) states that the (vertex) chromatic number of a graph $G$ is at most its maximum degree, unless $G$ is an odd cycle or a complete graph. This leads to the following question: what if $G$ is far from complete and the maximum degree $\Delta$ of $G$ is large? Can we say that $G$ has chromatic number much less than $\Delta$? The following theorem of Alon, Krivelevich and Sudakov (1999) shows that this is so if the neighbourhoods of vertices are not too dense:

**Theorem 8** Let $G$ be a graph of maximum degree $\Delta$, and suppose that the number of triangles on each vertex of $G$ is less than $\Delta^2/t$, where $2 \leq t \leq \Delta^2$. Then $G$ has chromatic number at most $c\Delta/\log t$ for some absolute constant $c$.

The $k$th power of a graph $G$ is the graph obtained by joining all pairs of vertices of $G$ at distance at most $k$. If $G$ has maximum degree $\Delta$, then $G^k$ potentially has maximum degree

$$\sum_{i=1}^{k} \Delta(\Delta - 1)^{i-1} \approx \Delta^k.$$ 

In the next theorem, we show that if $G$ contains no short cycles, then we can improve Brooke’s Theorem on $G^k$:

**Theorem 9** Let $G$ be a graph of girth at least $3k + 1$ and maximum degree $\Delta$, and let $G^k$ be the $k$th power of $G$. Then the

$$\chi(G^k) \ll \frac{\Delta^k}{\log \Delta}.$$ 

**Proof** One verifies that the maximum possible chromatic number of $G$ cannot be more than the given upper bound, by Theorem 8. Now we need a graph $G$ of maximum degree $\Delta$ such that the chromatic number of $G^k$ is at least $c\Delta^k/\log \Delta$ for some constant $c > 0$. In fact, all we need is that $G^k$ contains no independent set of size more than $n \log \Delta/c\Delta^k$. Such a graph is constructed using the probabilistic method.

Let $G_{n,p}$ denote the model of random graphs with edge-probability $p = \frac{d}{2n}$, where $d$ is fixed. For a set $S$ of $r$ vertices of $G_{n,p}$, let $X(S)$ denote the maximum number of internally disjoint paths
of length $k$ with endpoints in $S$, and with no other vertices in $S$. Note that $S$ is an independent set in $G^k$ if and only if no path of length at most $k$ has both endpoints in $S$. Now the expected number of paths of length $k$ with endpoints in $S$ is exactly

$$P = \binom{r}{2} (n - r)(n - r - 1) \cdots (n - r - k + 2)p^k.$$ 

Put $r = c_0 n \log d / d^k$, where $c_0$ is to be chosen later. Then the expected number of pairs of paths which share at least one common internal vertex is much smaller than $P^2$, so we can delete the shared internal vertices of paths to obtain a collection of at least

$$\frac{P}{2} \geq c_0^2 n (\log d)^2 / 2^{k+2} d^k$$

internally disjoint paths of length $k$ with both endpoints in $S$, and no other vertices in common. Thus for all sets $S$ of $r$ vertices, $E(X(S)) \geq \frac{P}{2}$.

Now $X = X(S)$ changes by at most one upon deleting or adding an edge to $G$, so $X$ is 1-Lipschitz. Furthermore, $X$ is $f$-certifiable, with $f(s) = ks$. Therefore, by Theorem 6, for any $s$ and $t > 0$,

$$\mathbb{P}(X \leq s - t\sqrt{f(s)}) \cdot \mathbb{P}(X \geq s) \leq e^{-t^2/4}.$$ 

By Lemma 7, there is an absolute constant $\gamma > 0$ such that

$$|\mathbb{E}(X) - \mathbb{M}(X)| \leq \gamma k \mathbb{E}(X).$$

Let $s = \mathbb{M}(X)$. It follows from the above inequality that there is a constant $c_1$ such that

$$\mathbb{P}(X \leq \frac{P}{8}) \leq 2e^{-c_1 n (\log d)^2 / d^k}.$$ 

Here $c_1$ depends only on $k$. It is not hard to check that

$$2e^{-c_1 n (\log d)^2 / d^k} \left( \frac{n}{r} \right) \to 0$$

as $n$ tends to infinity. So the probability that there exists a set $S$ of $r$ vertices which does not support at least $\frac{P}{8}$ internally disjoint paths tends to zero. In other words, a.a.s every set $S$ of size $r$ has at least $\frac{P}{8}$ internally disjoint paths of length $k$ with two vertices in $S$ and no other vertices in common.

Now we alter $G_{n,p}$: delete one vertex from each cycle of length at most $3k$ in $G_{n,p}$ and each vertex of degree greater than $d$ in $G_{n,p}$. The expected number of vertices deleted is at most $\frac{1}{3}(n2^{-d/16} + d^{3k})$, using the Chernoff bound for the probability that a fixed vertex has degree greater than $d$, so by Markov’s inequality we may delete $m = n2^{-d/16} + d^{3k}$ vertices to obtain a graph $G$ of girth at least $3k + 1$ and maximum degree $d$ with probability at least $3/4$. A.a.s, every set $S$ of $r$ vertices has $\frac{P}{8}$ internally disjoint paths in $G$ of length $k$ with endpoints in $S$. Since $\frac{P}{8} > m$, this means that at least one path survives the deletion for any set $S$, and therefore these two vertices of $S$ are joined in $G^k$. In particular, $G^k$ contains no independent set of size $r$, as required.
The girth condition in this theorem cannot be dropped. For example, it is known that there are $(q+1)$-regular bipartite graphs $P_q$ of girth six, with $q^2 + q + 1$ vertices in each part such that each pair of vertices in one part has a common neighbour, whenever $q$ is a prime power. Then $P^2_q$ is the union of two complete graphs of order $q^2 + q + 1$ on each part of $P_q$ together with $P_q$ itself. So the chromatic number of $G$ is at least $q^2 + q + 1$, whereas the maximum degree in $G$ is exactly $(q + 1)^2$. So the theorem required girth at least seven in the case $k = 2$. Alon and Mohar (2002) proved that if the girth of $G$ is at most $3k + 1$, then $G^k$ might have chromatic number at least $\Delta^k - \Delta + 1$. 