Introduction to analysis on the discrete cube

Krzysztof Oleszkiewicz

University of Warsaw

Simons Institute for the Theory of Computing at UC Berkeley,
Real Analysis Boot Camp, September 9-13, 2013

Disclaimer

This is a revised version of a course given at the Kent State University in 2008, extended to include parts of other presentations.

It is an introduction to the subject, not a complete exposition of the theory, its history and recent developments. Its main purpose is to present various basic objects, notions and approaches. For a more detailed and systematic approach see Ryan O'Donnell's blog and book *Analysis of Boolean Functions*, in preparation.

Disclaimer

This is a revised version of a course given at the Kent State University in 2008, extended to include parts of other presentations.

It is an introduction to the subject, not a complete exposition of the theory, its history and recent developments. Its main purpose is to present various basic objects, notions and approaches. For a more detailed and systematic approach see Ryan O'Donnell's blog and book *Analysis of Boolean Functions*, in preparation.

Discrete cube

$[n] := \{1, 2, \ldots, n\}$

Discrete cube (hypercube) $C_n := \{-1, 1\}^n$, equipped with a normalized counting (uniform probability) measure $(\frac{1}{2}\delta_{-1} + \frac{1}{2}\delta_1)^{\otimes n}$

Disclaimer: There will be no "cheating" as long as the discrete cube C_n is considered, with $n < \infty$. Many results of the present talk can be extended to the case $n = \infty$ and more general product probability spaces. However, usually technical details become much more delicate then.

Hamming's metric: For $x, y \in C_n$ let

$$d(x, y) = |\{i \in [n] : x_i \neq y_i\}| = \frac{1}{2}\|x - y\|_1.$$

Expectation: For $f : C_n \longrightarrow \mathbf{R}$ we have

$$E[f] = 2^{-n} \sum_{x \in C_n} f(x).$$

Discrete cube

$[n] := \{1, 2, \ldots, n\}$

Discrete cube (hypercube) $C_n := \{-1, 1\}^n$, equipped with a normalized counting (uniform probability) measure $(\frac{1}{2}\delta_{-1} + \frac{1}{2}\delta_1)^{\otimes n}$

Disclaimer: There will be no "cheating" as long as the discrete cube C_n is considered, with $n < \infty$. Many results of the present talk can be extended to the case $n = \infty$ and more general product probability spaces. However, usually technical details become much more delicate then.

Hamming's metric: For $x, y \in C_n$ let

$$d(x, y) = |\{i \in [n] : x_i \neq y_i\}| = \frac{1}{2}\|x - y\|_1.$$

Expectation: For $f : C_n \longrightarrow \mathbf{R}$ we have

$$E[f] = 2^{-n} \sum_{x \in C_n} f(x).$$

Discrete cube

$[n] := \{1, 2, \ldots, n\}$

Discrete cube (hypercube) $C_n := \{-1, 1\}^n$, equipped with a normalized counting (uniform probability) measure $(\frac{1}{2}\delta_{-1} + \frac{1}{2}\delta_1)^{\otimes n}$

Disclaimer: There will be no "cheating" as long as the discrete cube C_n is considered, with $n < \infty$. Many results of the present talk can be extended to the case $n = \infty$ and more general product probability spaces. However, usually technical details become much more delicate then.

Hamming's metric: For $x, y \in C_n$ let

$$d(x, y) = |\{i \in [n] : x_i \neq y_i\}| = \frac{1}{2}\|x - y\|_1.$$

Expectation: For $f : C_n \longrightarrow \mathbf{R}$ we have

$$E[f] = 2^{-n} \sum_{x \in C_n} f(x).$$

Discrete cube

$[n] := \{1, 2, \ldots, n\}$

Discrete cube (hypercube) $C_n := \{-1, 1\}^n$, equipped with a normalized counting (uniform probability) measure $(\frac{1}{2}\delta_{-1} + \frac{1}{2}\delta_1)^{\otimes n}$

Disclaimer: There will be no "cheating" as long as the discrete cube C_n is considered, with $n < \infty$. Many results of the present talk can be extended to the case $n = \infty$ and more general product probability spaces. However, usually technical details become much more delicate then.

Hamming's metric: For $x, y \in C_n$ let

$$d(x, y) = |\{i \in [n] : x_i \neq y_i\}| = \frac{1}{2}\|x - y\|_1.$$

Expectation: For $f : C_n \longrightarrow \mathbf{R}$ we have

$$E[f] = 2^{-n} \sum_{x \in C_n} f(x).$$

Discrete cube

$[n] := \{1, 2, \ldots, n\}$

Discrete cube (hypercube) $C_n := \{-1, 1\}^n$, equipped with a normalized counting (uniform probability) measure $(\frac{1}{2}\delta_{-1} + \frac{1}{2}\delta_1)^{\otimes n}$

Disclaimer: There will be no "cheating" as long as the discrete cube C_n is considered, with $n < \infty$. Many results of the present talk can be extended to the case $n = \infty$ and more general product probability spaces. However, usually technical details become much more delicate then.

Hamming's metric: For $x, y \in C_n$ let

$$d(x, y) = |\{i \in [n] : x_i \neq y_i\}| = \frac{1}{2}\|x - y\|_1.$$

Expectation: For $f : C_n \longrightarrow \mathbf{R}$ we have

$$E[f] = 2^{-n} \sum_{x \in C_n} f(x).$$

L^2 structure

Scalar product: For $f, g : C_n \longrightarrow \mathbf{R}$ let

$$\langle f, g \rangle = E[f \cdot g] = 2^{-n} \cdot \sum_{x \in C_n} f(x)g(x).$$

We denote $\|f\|_p = (E[|f|^p])^{1/p}$ for $p > 0$ and $\|f\|_\infty = \max_{x \in C_n} |f(x)|$.

Note that $\langle f, f \rangle = \|f\|_2^2$.

Hilbert space:

$$\mathcal{H}_n := L^2(C_n, \mathbf{R}); \quad \dim \mathcal{H}_n = 2^n$$

L^2 structure

Scalar product: For $f, g : C_n \longrightarrow \mathbf{R}$ let

$$\langle f, g \rangle = E[f \cdot g] = 2^{-n} \cdot \sum_{x \in C_n} f(x)g(x).$$

We denote $\|f\|_p = (E[|f|^p])^{1/p}$ for $p > 0$ and $\|f\|_\infty = \max_{x \in C_n} |f(x)|$.

Note that $\langle f, f \rangle = \|f\|_2^2$.

Hilbert space:

$$\mathcal{H}_n := L^2(C_n, \mathbf{R}); \quad \dim \mathcal{H}_n = 2^n$$

L^2 structure

Scalar product: For $f, g : C_n \longrightarrow \mathbf{R}$ let

$$\langle f, g \rangle = E[f \cdot g] = 2^{-n} \cdot \sum_{x \in C_n} f(x)g(x).$$

We denote $\|f\|_p = (E[|f|^p])^{1/p}$ for $p > 0$ and $\|f\|_\infty = \max_{x \in C_n} |f(x)|$.

Note that $\langle f, f \rangle = \|f\|_2^2$.

Hilbert space:

$$\mathcal{H}_n := L^2(C_n, \mathbf{R}); \quad \dim \mathcal{H}_n = 2^n$$

Walsh system

Boolean function: $f : C_n \to \{-1, 1\}$

- theoretical computer science (bits)
- social choice theory (voting)
- combinatorics (family of subsets of $[n]$)

Walsh functions: For $x \in \{-1, 1\}^n$ and $S \subseteq [n]$ let

$$w_S(x) = \prod_{i \in S} x_i,$$

$$w_\emptyset \equiv 1$$

$r_i := w_i = w_{\{i\}}$ - i-th coordinate projection ($i \in [n]$)

$r_1, r_2, \ldots, r_n$ - a Rademacher sequence:
independent symmetric ± 1 Bernoulli random variables

Walsh system

Boolean function: $f : C_n \to \{-1, 1\}$

- theoretical computer science (bits)
- social choice theory (voting)
- combinatorics (family of subsets of $[n]$)

Walsh functions: For $x \in \{-1,1\}^n$ and $S \subseteq [n]$ let

$$w_S(x) = \prod_{i \in S} x_i,$$

$$w_\emptyset \equiv 1$$

$r_i := w_i = w_{\{i\}}$ - i-th coordinate projection ($i \in [n]$)

$r_1, r_2, \ldots, r_n$ - a Rademacher sequence:
independent symmetric ± 1 Bernoulli random variables

Walsh system

Boolean function: $f : C_n \to \{-1, 1\}$

- theoretical computer science (bits)
- social choice theory (voting)
- combinatorics (family of subsets of $[n]$)

Walsh functions: For $x \in \{-1, 1\}^n$ and $S \subseteq [n]$ let

$$w_S(x) = \prod_{i \in S} x_i,$$

$$w_\emptyset \equiv 1$$

$r_i := w_i = w_{\{i\}}$ - i-th coordinate projection ($i \in [n]$)

$r_1, r_2, \ldots, r_n$ - a Rademacher sequence:
independent symmetric ± 1 Bernoulli random variables

Walsh system

Boolean function: $f : C_n \to \{-1, 1\}$

- theoretical computer science (bits)
- social choice theory (voting)
- combinatorics (family of subsets of $[n]$)

Walsh functions: For $x \in \{-1, 1\}^n$ and $S \subseteq [n]$ let

$$w_S(x) = \prod_{i \in S} x_i,$$

$$w_\emptyset \equiv 1$$

$r_i := w_i = w_{\{i\}}$ - i-th coordinate projection ($i \in [n]$)

$r_1, r_2, \ldots, r_n$ - a Rademacher sequence:
independent symmetric ± 1 Bernoulli random variables

Walsh system

Boolean function: $f : C_n \to \{-1, 1\}$

- theoretical computer science (bits)
- social choice theory (voting)
- combinatorics (family of subsets of $[n]$)

Walsh functions: For $x \in \{-1, 1\}^n$ and $S \subseteq [n]$ let

$$w_S(x) = \prod_{i \in S} x_i,$$

$$w_\emptyset \equiv 1$$

$r_i := w_i = w_{\{i\}}$ - i-th coordinate projection ($i \in [n]$)

$r_1, r_2, \ldots, r_n$ - a Rademacher sequence:
independent symmetric ± 1 Bernoulli random variables

Orthonormality

$E[w_S] = 0$ for $S \neq \emptyset$ and $E[w_\emptyset] = 1$

Indeed, expectation of the product of independent random variables is equal to the product of their expectations (and they are all equal to zero).

Orthonormality: $w_S \cdot w_T = w_{S \Delta T}$ thus

$$\langle w_S, w_T \rangle = E[w_{S \Delta T}] = \delta_{S,T}$$

Here Δ denotes a symmetric set difference (XOR) while $\delta_{S,T} = 1$ if $S = T$ and $\delta_{S,T} = 0$ if $S \neq T$ (Kronecker's delta).

Example: $w_{\{1,2\}} \cdot w_{\{2,3\}} = r_1 r_2 \cdot r_2 r_3 = r_1 r_2^2 r_3 = r_1 r_3$.

We have proved that the Walsh system $(w_S)_{S \subseteq [n]}$ is orthonormal (and therefore linearly independent). Since it is of cardinality 2^n, which is equal to the linear dimension of $\mathcal{H}_n$, it spans the whole space and thus is complete.

Orthonormality

$E[w_S] = 0$ for $S \neq \emptyset$ and $E[w_\emptyset] = 1$

Indeed, expectation of the product of independent random variables is equal to the product of their expectations (and they are all equal to zero).

Orthonormality: $w_S \cdot w_T = w_{S\Delta T}$ thus

$$\langle w_S, w_T \rangle = E[w_{S\Delta T}] = \delta_{S,T}$$

Here Δ denotes a symmetric set difference (XOR) while $\delta_{S,T} = 1$ if $S = T$ and $\delta_{S,T} = 0$ if $S \neq T$ (Kronecker's delta).

Example: $w_{\{1,2\}} \cdot w_{\{2,3\}} = r_1 r_2 \cdot r_2 r_3 = r_1 r_2^2 r_3 = r_1 r_3$.

We have proved that the Walsh system $(w_S)_{S\subseteq[n]}$ is orthonormal (and therefore linearly independent). Since it is of cardinality 2^n, which is equal to the linear dimension of $\mathcal{H}_n$, it spans the whole space and thus is complete.

Orthonormality

$E[w_S] = 0$ for $S \neq \emptyset$ and $E[w_\emptyset] = 1$

Indeed, expectation of the product of independent random variables is equal to the product of their expectations (and they are all equal to zero).

Orthonormality: $w_S \cdot w_T = w_{S \Delta T}$ thus

$$\langle w_S, w_T \rangle = E[w_{S\Delta T}] = \delta_{S,T}$$

Here Δ denotes a symmetric set difference (XOR) while $\delta_{S,T} = 1$ if $S = T$ and $\delta_{S,T} = 0$ if $S \neq T$ (Kronecker's delta).

Example: $w_{\{1,2\}} \cdot w_{\{2,3\}} = r_1 r_2 \cdot r_2 r_3 = r_1 r_2^2 r_3 = r_1 r_3$.

We have proved that the Walsh system $(w_S)_{S \subseteq [n]}$ is orthonormal (and therefore linearly independent). Since it is of cardinality 2^n, which is equal to the linear dimension of $\mathcal{H}_n$, it spans the whole space and thus is complete.

Orthonormality

$E[w_S] = 0$ for $S \neq \emptyset$ and $E[w_\emptyset] = 1$

Indeed, expectation of the product of independent random variables is equal to the product of their expectations (and they are all equal to zero).

Orthonormality: $w_S \cdot w_T = w_{S \Delta T}$ thus

$$\langle w_S, w_T \rangle = E[w_{S \Delta T}] = \delta_{S,T}$$

Here Δ denotes a symmetric set difference (XOR) while $\delta_{S,T} = 1$ if $S = T$ and $\delta_{S,T} = 0$ if $S \neq T$ (Kronecker's delta).

Example: $w_{\{1,2\}} \cdot w_{\{2,3\}} = r_1 r_2 \cdot r_2 r_3 = r_1 r_2^2 r_3 = r_1 r_3$.

We have proved that the Walsh system $(w_S)_{S \subseteq [n]}$ is orthonormal (and therefore linearly independent). Since it is of cardinality 2^n, which is equal to the linear dimension of $\mathcal{H}_n$, it spans the whole space and thus is complete.

Elementary argument

There is also a straightforward way to see that every function from $\mathcal{H}_n$ is a linear combination of the Walsh functions. Indeed, for any $y \in C_n$ we have

$$1_y(x) = \prod_{i=1}^{n} \frac{1 + x_i y_i}{2} = 2^{-n} \sum_{S \subseteq [n]} w_S(y) w_S(x),$$

where 1_y denotes the indicator (the characteristic function) of $\{y\}$. Hence

$$f(x) = \sum_{y \in C_n} f(y) 1_y(x) = 2^{-n} \sum_{S \subseteq [n]} \Big(\sum_{y \in C_n} f(y) w_S(y) \Big) w_S(x) =$$

$$= \sum_{S \subseteq [n]} \langle f, w_S \rangle \cdot w_S(x).$$

Therefore every $f \in \mathcal{H}_n$ admits one and only one **Walsh-Fourier expansion:**

$$f = \sum \hat{f}(S) w_S.$$

Elementary argument

There is also a straightforward way to see that every function from $\mathcal{H}_n$ is a linear combination of the Walsh functions. Indeed, for any $y \in C_n$ we have

$$1_y(x) = \prod_{i=1}^{n} \frac{1 + x_i y_i}{2} = 2^{-n} \sum_{S \subseteq [n]} w_S(y) w_S(x),$$

where 1_y denotes the indicator (the characteristic function) of $\{y\}$. Hence

$$f(x) = \sum_{y \in C_n} f(y) 1_y(x) = 2^{-n} \sum_{S \subseteq [n]} \Big(\sum_{y \in C_n} f(y) w_S(y) \Big) w_S(x) =$$

$$= \sum_{S \subseteq [n]} \langle f, w_S \rangle \cdot w_S(x).$$

Therefore every $f \in \mathcal{H}_n$ admits one and only one **Walsh-Fourier expansion:**

$$f = \sum \hat{f}(S) w_S.$$

Simple consequences of the orthonormality

As we have seen above (it follows also from the orthonormality of the Walsh system):

$$\hat{f}(S) = \langle f, w_S \rangle = E[f \cdot w_S].$$

In particular, for every $f \in \mathcal{H}_n$ we have

$$E[f] = E[f \cdot \mathbf{1}] = E[f \cdot w_\emptyset] = \langle f, w_\emptyset \rangle = \hat{f}(\emptyset)$$

and

$$E[f^2] = E[f \cdot f] = \langle f, f \rangle = \langle \sum_{S \subseteq [n]} \hat{f}(S) w_S, \sum_{T \subseteq [n]} \hat{f}(T) w_T \rangle =$$

$$= \sum_{S,T \subseteq [n]} \hat{f}(S)\hat{f}(T)\langle w_S, w_T \rangle = \sum_{S \subseteq [n]} \hat{f}(S)^2 \text{ } \textbf{(Plancherel).}$$

Simple consequences of the orthonormality

As we have seen above (it follows also from the orthonormality of the Walsh system):

$$\hat{f}(S) = \langle f, w_S \rangle = E[f \cdot w_S].$$

In particular, for every $f \in \mathcal{H}_n$ we have

$$E[f] = E[f \cdot \mathbf{1}] = E[f \cdot w_\emptyset] = \langle f, w_\emptyset \rangle = \hat{f}(\emptyset)$$

and

$$E[f^2] = E[f \cdot f] = \langle f, f \rangle = \langle \sum_{S \subseteq [n]} \hat{f}(S) w_S, \sum_{T \subseteq [n]} \hat{f}(T) w_T \rangle =$$

$$= \sum_{S, T \subseteq [n]} \hat{f}(S)\hat{f}(T)\langle w_S, w_T \rangle = \sum_{S \subseteq [n]} \hat{f}(S)^2 \ \textbf{(Plancherel)}.$$

LCA setting

Remark: Note that $\{-1,1\}$ (with multiplication as a group action) is a locally compact (compact, in fact) abelian group and $C_n = \{-1,1\}^n$ (with coordinatewise multiplication as a group action) shares this property. The case of the Cantor group ($n = \infty$ with the natural product topology) is covered as well. The standard product probability measure on C_n is the Haar measure then and general harmonic analysis on LCA groups tools apply. It is easy to check that, for $n < \infty$, C_n is self-dual: the group of characters on C_n is just the Walsh system and it is isomorphic with C_n itself and the isomorphism is very natural - $S \subseteq [n]$ is identified with $x \in C_n$ such that $S = \{i \in [n] : x_i = -1\}$. Then the mapping $f \mapsto \hat{f}$, which sends a real function on C_n to its Walsh-Fourier coefficients collection, is just the classical Fourier transform (on LCA groups) up to some normalization. The transform applied twice returns the original function, up to a multiplicative factor. However, in what follows we will not take advantage (at least explicitly) of the group structure of C_n.

Computing the Walsh-Fourier transform

At first glance, it may seem that to compute the Walsh-Fourier transform of a function on C_n one needs $O(2^n \cdot 2^n)$, approximately quadratic in the data size, arithmetic operations. However, only $O(n \cdot 2^n)$ operations are needed.

Indeed, note that knowing the Walsh-Fourier transforms of the function restricted to two parallel $(n-1)$-dimensional faces of C_n one easily obtains the Walsh-Fourier transform of the function on the whole discrete cube, using only $O(2^n)$ operations – addition, subtraction and division by 2 suffice. Thus, if we denote by $\tau(n)$ the number of operations needed to compute the Walsh-Fourier transform on C_n then we have $\tau(n) \le 2\tau(n-1) + \kappa \cdot 2^n$, i.e.,

$$\frac{\tau(n)}{2^n} \le \frac{\tau(n-1)}{2^{n-1}} + \kappa,$$

so that $2^{-n}\tau(n) \le \tau_0 + \kappa n$.

Computing the Walsh-Fourier transform

At first glance, it may seem that to compute the Walsh-Fourier transform of a function on C_n one needs $O(2^n \cdot 2^n)$, approximately quadratic in the data size, arithmetic operations. However, only $O(n \cdot 2^n)$ operations are needed.

Indeed, note that knowing the Walsh-Fourier transforms of the function restricted to two parallel $(n-1)$-dimensional faces of C_n one easily obtains the Walsh-Fourier transform of the function on the whole discrete cube, using only $O(2^n)$ operations – addition, subtraction and division by 2 suffice. Thus, if we denote by $\tau(n)$ the number of operations needed to compute the Walsh-Fourier transform on C_n then we have $\tau(n) \leq 2\tau(n-1) + \kappa \cdot 2^n$, i.e.,

$$\frac{\tau(n)}{2^n} \leq \frac{\tau(n-1)}{2^{n-1}} + \kappa,$$

so that $2^{-n}\tau(n) \leq \tau_0 + \kappa n$.

Example: Discrete time symmetric random walk on C_n

Let $n \geq 2$ and let us consider a Markov chain with the state space C_n i.e. a sequence of C_n-valued random variables $(Y_t)_{t=0}^{\infty}$ satisfying the Markov condition and such that $Y_0 = (1, 1, \ldots, 1)$ a.s. and $\forall_t P(Y_{t+1} = x | Y_t = y) = 1/n$ whenever $d(x, y) = 1$. This models a random walk starting from $(1, 1, \ldots, 1)$ and moving in every step from a vertex it occupies to one of its neighbours, choosing each of them with equal probability. The starting point $(1, 1, \ldots, 1)$ is chosen for the sake of simplicity and it can be easily replaced by another vertex of the cube.

Let $f_t(x) = P(Y_t = x)$. Obviously,

$$f_0(x) = \prod_{i=1}^{n} \frac{1 + x_i}{2} = 2^{-n} \sum_{S \subseteq [n]} w_S(x).$$

Discrete time random walk

Example: Discrete time symmetric random walk on C_n

Let $n \geq 2$ and let us consider a Markov chain with the state space C_n i.e. a sequence of C_n-valued random variables $(Y_t)_{t=0}^{\infty}$ satisfying the Markov condition and such that $Y_0 = (1, 1, \ldots, 1)$ a.s. and $\forall_t P(Y_{t+1} = x | Y_t = y) = 1/n$ whenever $d(x, y) = 1$. This models a random walk starting from $(1, 1, \ldots, 1)$ and moving in every step from a vertex it occupies to one of its neighbours, choosing each of them with equal probability. The starting point $(1, 1, \ldots, 1)$ is chosen for the sake of simplicity and it can be easily replaced by another vertex of the cube.

Let $f_t(x) = P(Y_t = x)$. Obviously,

$$f_0(x) = \prod_{i=1}^{n} \frac{1 + x_i}{2} = 2^{-n} \sum_{S \subseteq [n]} w_S(x).$$

Discrete time walk - spectral properties

We have $f_{t+1} = Kf_t$, where $K : \mathcal{H}_n \longrightarrow \mathcal{H}_n$ is a linear operator defined by the following formula:

$$(Kf)(x) = \frac{1}{n} \cdot \sum_{y \in C_n:\, d(x,y)=1} f(y).$$

Hence $f_t = K^t f_0$.

Note that for $S \subseteq [n]$ we have

$$Kw_S = \frac{1}{n}\Big((n - |S|)w_S - |S|w_S\Big) = \Big(1 - 2\frac{|S|}{n}\Big)w_S,$$

which means that Walsh functions are eigenfunctions of the operator K (and therefore K is a **multiplier**). Indeed, for every $x \in C_n$ exactly $|S|$ out of n neighbours of x differ from x on a coordinate belonging to S (and w_S takes value $-w_S(x)$ on these vertices) whereas the remaining $n - |S|$ neighbour vertices have the same coordinates indexed by S as x and therefore w_S does not distinguish them from x (i.e. assigns the value $w_S(x)$ to them).

Discrete time walk - spectral properties

We have $f_{t+1} = Kf_t$, where $K : \mathcal{H}_n \longrightarrow \mathcal{H}_n$ is a linear operator defined by the following formula:

$$(Kf)(x) = \frac{1}{n} \cdot \sum_{y \in C_n:\, d(x,y)=1} f(y).$$

Hence $f_t = K^t f_0$.

Note that for $S \subseteq [n]$ we have

$$Kw_S = \frac{1}{n}\Big((n - |S|)w_S - |S|w_S\Big) = \Big(1 - 2\frac{|S|}{n}\Big)w_S,$$

which means that Walsh functions are eigenfunctions of the operator K (and therefore K is a **multiplier**). Indeed, for every $x \in C_n$ exactly $|S|$ out of n neighbours of x differ from x on a coordinate belonging to S (and w_S takes value $-w_S(x)$ on these vertices) whereas the remaining $n - |S|$ neighbour vertices have the same coordinates indexed by S as x and therefore w_S does not distinguish them from x (i.e. assigns the value $w_S(x)$ to them).

Discrete time walk - estimates

Consequently, we have

$$K^t w_S = \left(1 - 2\frac{|S|}{n}\right)^t w_S$$

and

$$f_t = K^t f_0 = 2^{-n} \sum_{S \subseteq [n]} \left(1 - 2\frac{|S|}{n}\right)^t \cdot w_S.$$

Since the Walsh functions are Boolean and $\left|1 - 2\frac{|S|}{n}\right| \leq 1 - \frac{2}{n}$ whenever $S \neq \emptyset$ and $S \neq [n]$, we deduce that

$$\left\| f_t - 2^{-n} w_\emptyset - (-1)^t 2^{-n} w_{[n]} \right\|_\infty \leq \left(1 - \frac{2}{n}\right)^t \leq e^{-2t/n}.$$

Recall that $w_\emptyset \equiv 1$ and $w_{[n]} = r_1 r_2 \ldots r_n$.

Discrete time walk - estimates

Consequently, we have

$$K^t w_S = \left(1 - 2\frac{|S|}{n}\right)^t w_S$$

and

$$f_t = K^t f_0 = 2^{-n} \sum_{S \subseteq [n]} \left(1 - 2\frac{|S|}{n}\right)^t \cdot w_S.$$

Since the Walsh functions are Boolean and $\left|1 - 2\frac{|S|}{n}\right| \leq 1 - \frac{2}{n}$ whenever $S \neq \emptyset$ and $S \neq [n]$, we deduce that

$$\left\| f_t - 2^{-n} w_\emptyset - (-1)^t 2^{-n} w_{[n]} \right\|_\infty \leq \left(1 - \frac{2}{n}\right)^t \leq e^{-2t/n}.$$

Recall that $w_\emptyset \equiv 1$ and $w_{[n]} = r_1 r_2 \ldots r_n$.

Discrete random walk - ergodicity

Hence

$$f_{2t} \longrightarrow 2^{-n}(1 + r_1 \dots r_n) = \frac{1}{2^{n-1}} \cdot 1_{y_1+\dots+y_n \equiv n \ (mod\ 2)}$$

and

$$f_{2t+1} \longrightarrow 2^{-n}(1 - r_1 \dots r_n) = \frac{1}{2^{n-1}} \cdot 1_{y_1+\dots+y_n \not\equiv n \ (mod\ 2)},$$

uniformly on C_n and with exponential speed, as $t \longrightarrow \infty$.

Clearly, it is just a precise form of the ergodic theorem for this Markov chain and the dependence on the parity of t is related to the fact that the chain is 2-periodic. It is so beacuse C_n is a bi-partite graph (we connect two vertices with an edge if and only if they are neighbours).

Discrete random walk - ergodicity

Hence

$$f_{2t} \longrightarrow 2^{-n}(1 + r_1 \dots r_n) = \frac{1}{2^{n-1}} \cdot 1_{y_1+\dots+y_n \equiv n\ (mod\ 2)}$$

and

$$f_{2t+1} \longrightarrow 2^{-n}(1 - r_1 \dots r_n) = \frac{1}{2^{n-1}} \cdot 1_{y_1+\dots+y_n \not\equiv n\ (mod\ 2)},$$

uniformly on C_n and with exponential speed, as $t \longrightarrow \infty$.

Clearly, it is just a precise form of the ergodic theorem for this Markov chain and the dependence on the parity of t is related to the fact that the chain is 2-periodic. It is so beacuse C_n is a bi-partite graph (we connect two vertices with an edge if and only if they are neighbours).

Lazy random walk

Let us modify the previous example a little bit. We choose $\lambda \in (0, 1/2]$ and define a new random walk $Z_t = Z_t^{v,\lambda}$, starting from $v \in C_n$. Now we set different transition probability rules: $\forall_t \ P(Z_{t+1} = x | Z_t = y) = \lambda/n$ whenever $d(x, y) = 1$, and $P(Z_{t+1} = x | Z_t = x) = 1 - \lambda$.

This random walk is "lazy" - sometimes it does not move (especially when λ is small). When it does move, it chooses the vertex to go to among the neighbours of its current position, each of them with the same probability (one can also describe $(Z_t)_{t=0}^{\infty}$ as a modification of $(Y_t)_{t=0}^{\infty}$ by some non-deterministic time change).

Let $f_t(x) = P(Z_t = x)$. Clearly,

$$f_0(x) = \prod_{i=1}^{n} \frac{1 + v_i x_i}{2} = 2^{-n} \sum_{S \subseteq [n]} w_S(v) w_S(x).$$

Lazy random walk

Let us modify the previous example a little bit. We choose $\lambda \in (0, 1/2]$ and define a new random walk $Z_t = Z_t^{v,\lambda}$, starting from $v \in C_n$. Now we set different transition probability rules: $\forall_t \ P(Z_{t+1} = x | Z_t = y) = \lambda/n$ whenever $d(x, y) = 1$, and $P(Z_{t+1} = x | Z_t = x) = 1 - \lambda$.

This random walk is "lazy" - sometimes it does not move (especially when λ is small). When it does move, it chooses the vertex to go to among the neighbours of its current position, each of them with the same probability (one can also describe $(Z_t)_{t=0}^{\infty}$ as a modification of $(Y_t)_{t=0}^{\infty}$ by some non-deterministic time change).

Let $f_t(x) = P(Z_t = x)$. Clearly,

$$f_0(x) = \prod_{i=1}^{n} \frac{1 + v_i x_i}{2} = 2^{-n} \sum_{S \subseteq [n]} w_S(v) w_S(x).$$

Let us modify the previous example a little bit. We choose $\lambda \in (0, 1/2]$ and define a new random walk $Z_t = Z_t^{v,\lambda}$, starting from $v \in C_n$. Now we set different transition probability rules: $\forall_t \; P(Z_{t+1} = x | Z_t = y) = \lambda/n$ whenever $d(x, y) = 1$, and $P(Z_{t+1} = x | Z_t = x) = 1 - \lambda$.

This random walk is "lazy" - sometimes it does not move (especially when λ is small). When it does move, it chooses the vertex to go to among the neighbours of its current position, each of them with the same probability (one can also describe $(Z_t)_{t=0}^{\infty}$ as a modification of $(Y_t)_{t=0}^{\infty}$ by some non-deterministic time change).

Let $f_t(x) = P(Z_t = x)$. Clearly,

$$f_0(x) = \prod_{i=1}^{n} \frac{1 + v_i x_i}{2} = 2^{-n} \sum_{S \subseteq [n]} w_S(v) w_S(x).$$

Lazy random walk - ergodicity

Now

$$f_{t+1} = K_\lambda f_t$$

where $(K_\lambda f)(x) = (1-\lambda)f(x) + \frac{\lambda}{n}\sum_{y\in C_n:\ d(x,y)=1} f(y)$,
i.e. $K_\lambda = (1-\lambda)Id + \lambda K$.

Hence

$$K_\lambda w_S = (1-\lambda)w_S + \lambda\cdot\left(1-\frac{2|S|}{n}\right)w_S = \left(1-\frac{2\lambda|S|}{n}\right)w_S$$

and, as before, we get

$$f_t = K_\lambda^t f_0 = 2^{-n}\sum_{S\subseteq[n]} w_S(v)\left(1-\frac{2\lambda|S|}{n}\right)^t w_S.$$

Now for every $S\neq\emptyset$ we have $|1-\frac{2\lambda|S|}{n}|\leq 1-\frac{2\lambda}{n}$, so that f_t converges uniformly on C_n and exponentially fast (but still possibly quite slow if λ is close to zero) to the constant function 2^{-n}, no matter where v was.

Lazy random walk - ergodicity

Now

$$f_{t+1} = K_\lambda f_t$$

where $(K_\lambda f)(x) = (1-\lambda)f(x) + \frac{\lambda}{n}\sum_{y\in C_n:\ d(x,y)=1} f(y)$,
i.e. $K_\lambda = (1-\lambda)Id + \lambda K$.

Hence

$$K_\lambda w_S = (1-\lambda)w_S + \lambda \cdot \Big(1 - \frac{2|S|}{n}\Big) w_S = \Big(1 - \frac{2\lambda|S|}{n}\Big) w_S$$

and, as before, we get

$$f_t = K_\lambda^t f_0 = 2^{-n} \sum_{S\subseteq[n]} w_S(v)\Big(1 - \frac{2\lambda|S|}{n}\Big)^t w_S.$$

Now for every $S \neq \emptyset$ we have $|1 - \frac{2\lambda|S|}{n}| \leq 1 - \frac{2\lambda}{n}$, so that f_t converges uniformly on C_n and exponentially fast (but still possibly quite slow if λ is close to zero) to the constant function 2^{-n}, no matter where v was.

Lazy random walk - ergodicity

Now

$$f_{t+1} = K_\lambda f_t$$

where $(K_\lambda f)(x) = (1-\lambda)f(x) + \frac{\lambda}{n}\sum_{y \in C_n:\ d(x,y)=1} f(y)$,
i.e. $K_\lambda = (1-\lambda)Id + \lambda K$.

Hence

$$K_\lambda w_S = (1-\lambda)w_S + \lambda \cdot \left(1 - \frac{2|S|}{n}\right) w_S = \left(1 - \frac{2\lambda|S|}{n}\right) w_S$$

and, as before, we get

$$f_t = K_\lambda^t f_0 = 2^{-n} \sum_{S \subseteq [n]} w_S(v) \left(1 - \frac{2\lambda|S|}{n}\right)^t w_S.$$

Now for every $S \neq \emptyset$ we have $|1 - \frac{2\lambda|S|}{n}| \leq 1 - \frac{2\lambda}{n}$, so that f_t converges uniformly on C_n and exponentially fast (but still possibly quite slow if λ is close to zero) to the constant function 2^{-n}, no matter where v was.

Lazy walk - limit behaviour

Obviously, this is the classical ergodic theorem again (the "laziness" destroyed the 2-periodicity which we observed in the previous example). What is more interesting is a time rescaling of the "lazy walk": since it really moves only in λ fraction of time steps, due to the Law of Large Numbers, it is natural to ask about $f_{\lceil nt/\lambda \rceil}$ for real $t > 0$. One easily arrives at

$$f_{\lceil nt/\lambda \rceil} \xrightarrow{\lambda \to 0^+} 2^{-n} \sum_{S \subseteq [n]} w_S(v) e^{-2t|S|} w_S.$$

If we start the process from some random point rather than from a fixed v, we have some non-negative function $f_0 = \sum_{S \subseteq [n]} a_S w_S$,

$$a_\emptyset = E[f_0] = 2^{-n} \sum_{x \in C_n} f_0(x) = 2^{-n}$$

and

$$f_{\lceil nt/\lambda \rceil} \xrightarrow{\lambda \to 0^+} \sum_{S \subseteq [n]} e^{-2t|S|} a_S w_S.$$

Lazy walk - limit behaviour

Obviously, this is the classical ergodic theorem again (the "laziness" destroyed the 2-periodicity which we observed in the previous example). What is more interesting is a time rescaling of the "lazy walk": since it really moves only in λ fraction of time steps, due to the Law of Large Numbers, it is natural to ask about $f_{\lceil nt/\lambda \rceil}$ for real $t > 0$. One easily arrives at

$$f_{\lceil nt/\lambda \rceil} \stackrel{\lambda \to 0^+}{\longrightarrow} 2^{-n} \sum_{S \subseteq [n]} w_S(v) e^{-2t|S|} w_S.$$

If we start the process from some random point rather than from a fixed v, we have some non-negative function $f_0 = \sum_{S \subseteq [n]} a_S w_S$,

$$a_\emptyset = E[f_0] = 2^{-n} \sum_{x \in C_n} f_0(x) = 2^{-n}$$

and

$$f_{\lceil nt/\lambda \rceil} \stackrel{\lambda \to 0^+}{\longrightarrow} \sum_{S \subseteq [n]} e^{-2t|S|} a_S w_S.$$

Poisson parity process

Let $(N(t))_{t\in[0,\infty)}$ be the standard Poisson process, i.e. an integer-valued Markov process with independent Poissonian increments:

$$N(0) = 0,\ \forall t > s \geq 0\ N(t) - N(s) \sim N(t-s) \sim Pois(t-s).$$

With probability one its trajectory $t \mapsto N(t)$ is a non-decreasing integer-valued function, and the time gaps between the trajectory's jumps (with probability one the function increases exactly by 1 at the point of jump) are i.i.d. exponential random variables (with expectation equal to 1).

Define $M(t) = (-1)^{N(t)}$. Although in general an image of a Markov process under some map does **not** have to be a Markov process, $(M(t))_{t\in[0,\infty)}$ **does** satisfy Markov's condition.

Poisson parity process

Let $(N(t))_{t\in[0,\infty)}$ be the standard Poisson process, i.e. an integer-valued Markov process with independent Poissonian increments:

$$N(0) = 0,\ \forall t > s \geq 0\ N(t) - N(s) \sim N(t-s) \sim Pois(t-s).$$

With probability one its trajectory $t \mapsto N(t)$ is a non-decreasing integer-valued function, and the time gaps between the trajectory's jumps (with probability one the function increases exactly by 1 at the point of jump) are i.i.d. exponential random variables (with expectation equal to 1).

Define $M(t) = (-1)^{N(t)}$. Although in general an image of a Markov process under some map does **not** have to be a Markov process, $(M(t))_{t\in[0,\infty)}$ **does** satisfy Markov's condition.

Poisson parity process

Let $(N(t))_{t\in[0,\infty)}$ be the standard Poisson process, i.e. an integer-valued Markov process with independent Poissonian increments:

$$N(0) = 0, \ \forall t > s \geq 0 \ N(t) - N(s) \sim N(t-s) \sim Pois(t-s).$$

With probability one its trajectory $t \mapsto N(t)$ is a non-decreasing integer-valued function, and the time gaps between the trajectory's jumps (with probability one the function increases exactly by 1 at the point of jump) are i.i.d. exponential random variables (with expectation equal to 1).

Define $M(t) = (-1)^{N(t)}$. Although in general an image of a Markov process under some map does **not** have to be a Markov process, $(M(t))_{t\in[0,\infty)}$ **does** satisfy Markov's condition.

Parity process - transition

The parity process defined above follows a simple transition rule:

$$P(M(t)=1|M(s)=1)=P(M(t)=-1|M(s)=-1)=(1+e^{-2(t-s)})/2,$$

$$P(M(t)=-1|M(s)=1)=P(M(t)=1|M(s)=-1)=(1-e^{-2(t-s)})/2,$$

for all $t>s\geq 0$.

Indeed,

$$P(M(t)=1|M(s)=1)=P(N(t)-N(s)\text{ is even}|N(s)\text{ is even})=$$

$$=P(N(t)-N(s)\text{ is even})=Pois(t-s)(\{0,2,4,\ldots\})=$$

$$e^{-(t-s)}\sum_{k=0}^{\infty}(t-s)^{2k}/(2k)!=e^{-(t-s)}\cdot\frac{e^{t-s}+e^{-(t-s)}}{2}=(1+e^{-2(t-s)})/2$$

and thus

$$P(M(t)=-1|M(s)=1)=1-P(M(t)=1|M(s)=1)=$$

$$=1-(1+e^{-2(t-s)})/2=(1-e^{-2(t-s)})/2.$$

Parity process - transition

The parity process defined above follows a simple transition rule:

$$P(M(t)=1|M(s)=1)=P(M(t)=-1|M(s)=-1)=(1+e^{-2(t-s)})/2,$$

$$P(M(t)=-1|M(s)=1)=P(M(t)=1|M(s)=-1)=(1-e^{-2(t-s)})/2,$$

for all $t>s\geq 0$.

Indeed,

$$P(M(t)=1|M(s)=1)=P(N(t)-N(s)\text{ is even}|N(s)\text{ is even})=$$

$$=P(N(t)-N(s)\text{ is even})=Pois(t-s)(\{0,2,4,\ldots\})=$$

$$e^{-(t-s)}\sum_{k=0}^{\infty}(t-s)^{2k}/(2k)!=e^{-(t-s)}\cdot\frac{e^{t-s}+e^{-(t-s)}}{2}=(1+e^{-2(t-s)})/2$$

and thus

$$P(M(t)=-1|M(s)=1)=1-P(M(t)=1|M(s)=1)=$$

$$=1-(1+e^{-2(t-s)})/2=(1-e^{-2(t-s)})/2.$$

Parity process - transition

The parity process defined above follows a simple transition rule:

$$P(M(t)=1|M(s)=1)=P(M(t)=-1|M(s)=-1)=(1+e^{-2(t-s)})/2,$$

$$P(M(t)=-1|M(s)=1)=P(M(t)=1|M(s)=-1)=(1-e^{-2(t-s)})/2,$$

for all $t > s \geq 0$.

Indeed,

$$P(M(t)=1|M(s)=1)=P(N(t)-N(s) \text{ is even}|N(s) \text{ is even})=$$

$$=P(N(t)-N(s) \text{ is even})=Pois(t-s)(\{0,2,4,\ldots\})=$$

$$e^{-(t-s)}\sum_{k=0}^{\infty}(t-s)^{2k}/(2k)! = e^{-(t-s)}\cdot\frac{e^{t-s}+e^{-(t-s)}}{2}=(1+e^{-2(t-s)})/2$$

and thus

$$P(M(t)=-1|M(s)=1)=1-P(M(t)=1|M(s)=1)=$$

$$=1-(1+e^{-2(t-s)})/2=(1-e^{-2(t-s)})/2.$$

Due to the properties of the Poisson process $(N(t))_{t\in[0,\infty)}$ the time gaps between the sign flips are again i.i.d. exponential random variables (with expectation equal to 1).

For notational simplicity we will consider the same process with time running two times slower, i.e. we define $X(t) = M(t/2)$ to obtain, for all $t > s \geq 0$,

$$P(X(t) = 1|X(s) = 1) = P(X(t) = -1|X(s) = -1) = (1+e^{-(t-s)})/2,$$

$$P(X(t) = -1|X(s) = 1) = P(X(t) = 1|X(s) = -1) = (1-e^{-(t-s)})/2.$$

Note that $(X(t))_{t\in[0,\infty)}$ is both time and space homogenous Markov process. In fact one may construct it out of scratch, at least as long as one cares only about the finite-dimensional distributions, forgetting about trajectories (which is our case) - one just needs to prove the consistency conditions which in this case amounts to checking whether the Chapman-Kolmogorov equations hold; then the Kolmogorov consistency (extension) theorem does the rest.

Parity procees - time rescaling

Due to the properties of the Poisson process $(N(t))_{t\in[0,\infty)}$ the time gaps between the sign flips are again i.i.d. exponential random variables (with expectation equal to 1).

For notational simplicity we will consider the same process with time running two times slower, i.e. we define $X(t) = M(t/2)$ to obtain, for all $t > s \geq 0$,

$$P(X(t) = 1|X(s) = 1) = P(X(t) = -1|X(s) = -1) = (1+e^{-(t-s)})/2,$$

$$P(X(t) = -1|X(s) = 1) = P(X(t) = 1|X(s) = -1) = (1-e^{-(t-s)})/2.$$

Note that $(X(t))_{t\in[0,\infty)}$ is both time and space homogenous Markov process. In fact one may construct it out of scratch, at least as long as one cares only about the finite-dimensional distributions, forgetting about trajectories (which is our case) - one just needs to prove the consistency conditions which in this case amounts to checking whether the Chapman-Kolmogorov equations hold; then the Kolmogorov consistency (extension) theorem does the rest.

Parity procees - time rescaling

Due to the properties of the Poisson process $(N(t))_{t\in[0,\infty)}$ the time gaps between the sign flips are again i.i.d. exponential random variables (with expectation equal to 1).

For notational simplicity we will consider the same process with time running two times slower, i.e. we define $X(t) = M(t/2)$ to obtain, for all $t > s \geq 0$,

$$P(X(t) = 1 | X(s) = 1) = P(X(t) = -1 | X(s) = -1) = (1 + e^{-(t-s)})/2,$$

$$P(X(t) = -1 | X(s) = 1) = P(X(t) = 1 | X(s) = -1) = (1 - e^{-(t-s)})/2.$$

Note that $(X(t))_{t\in[0,\infty)}$ is both time and space homogenous Markov process. In fact one may construct it out of scratch, at least as long as one cares only about the finite-dimensional distributions, forgetting about trajectories (which is our case) - one just needs to prove the consistency conditions which in this case amounts to checking whether the Chapman-Kolmogorov equations hold; then the Kolmogorov consistency (extension) theorem does the rest.

Parity process - consistency

Let $u > t > s \geq 0$ and $z, x \in \{-1, 1\}$. We need to prove that

$$p_{u-s}(x, z) = \sum_{y \in \{-1,1\}} p_{t-s}(x, y) p_{u-t}(y, z).$$

Well, if $x = z$ then we do have

$$(1 + e^{-(u-s)})/2 =$$

$$= (1+e^{-(t-s)})/2 \cdot (1+e^{-(u-t)})/2 + (1-e^{-(t-s)})/2 \cdot (1-e^{-(u-t)})/2,$$

and if $x = -z$ then we do have

$$(1 - e^{-(u-s)})/2 =$$

$$= (1+e^{-(t-s)})/2 \cdot (1-e^{-(u-t)})/2 + (1-e^{-(t-s)})/2 \cdot (1+e^{-(u-t)})/2,$$

so the proof is finished.

Parity process - consistency

Let $u > t > s \geq 0$ and $z, x \in \{-1, 1\}$. We need to prove that

$$p_{u-s}(x,z) = \sum_{y \in \{-1,1\}} p_{t-s}(x,y) p_{u-t}(y,z).$$

Well, if $x = z$ then we do have

$$(1 + e^{-(u-s)})/2 =$$

$$= (1+e^{-(t-s)})/2 \cdot (1+e^{-(u-t)})/2 + (1-e^{-(t-s)})/2 \cdot (1-e^{-(u-t)})/2,$$

and if $x = -z$ then we do have

$$(1 - e^{-(u-s)})/2 =$$

$$= (1+e^{-(t-s)})/2 \cdot (1-e^{-(u-t)})/2 + (1-e^{-(t-s)})/2 \cdot (1+e^{-(u-t)})/2,$$

so the proof is finished.

Parity process - consistency

Let $u > t > s \geq 0$ and $z, x \in \{-1, 1\}$. We need to prove that

$$p_{u-s}(x,z) = \sum_{y \in \{-1,1\}} p_{t-s}(x,y) p_{u-t}(y,z).$$

Well, if $x = z$ then we do have

$$(1 + e^{-(u-s)})/2 =$$

$$= (1+e^{-(t-s)})/2 \cdot (1+e^{-(u-t)})/2 + (1-e^{-(t-s)})/2 \cdot (1-e^{-(u-t)})/2,$$

and if $x = -z$ then we do have

$$(1 - e^{-(u-s)})/2 =$$

$$= (1+e^{-(t-s)})/2 \cdot (1-e^{-(u-t)})/2 + (1-e^{-(t-s)})/2 \cdot (1+e^{-(u-t)})/2,$$

so the proof is finished.

Continuous time random walk

Now we may construct a continuous time random walk on C_n. Let $(X_1(t))_{t\in[0,\infty)}$, $(X_2(t))_{t\in[0,\infty)}, \ldots$ $(X_n(t))_{t\in[0,\infty)}$ be i.i.d. copies of the process $(X(t))_{t\in[0,\infty)}$. Given $v \in C_n$ we define a C_n-valued Markov process $(\mathbf{X}^v(t))_{t\in[0,\infty)}$ by setting

$$\mathbf{X}^v(t) = \Big(v_1 \cdot X_1(t), v_2 \cdot X_2(t), \ldots, v_n \cdot X_n(t)\Big),$$

so that $\mathbf{X}^v(0) = v$. The process starts at v and jumps from a vertex to each of its n neighbours with probability $1/n$, the time gaps between jumps being i.i.d. exponential random variables with expectation $2/n$ (the factor 2 comes from the fact that we have slowed the time flow and the factor $1/n$ is related to the fact that now jumps may occur on n coordinates).

Semigroup

Given $t \geq 0$ and a function $f \in \mathcal{H}_n$ we define a new function $P_t f \in \mathcal{H}_n$:

$$(P_t f)(v) = E[f(\mathbf{X}^v(t))] = \sum_{x \in C_n} p_t(v, x) f(x)$$

for $v \in C_n$, where $p_t(v, x)$ denotes the transition probability from v to x for the process $(\mathbf{X}(t))_{t \in [0, \infty)}$ - here we understand the process as a set of transition rules, independent of the starting point.

Clearly, P_t is a linear operator with the following properties:

- $P_t \mathbf{1} = \mathbf{1}$ (this reads as the invariance of the product probability measure on C_n **because the semigroup is symmetric**),
- $f \geq 0$ a.s implies $P_t f \geq 0$ a.s. (positivity preserving); because of the linearity of P_t the second condition may be equvialently stated as
- $f \geq g$ a.s. implies $P_t f \geq P_t g$ a.s. (order preserving).

Semigroup

Given $t \geq 0$ and a function $f \in \mathcal{H}_n$ we define a new function $P_t f \in \mathcal{H}_n$:

$$(P_t f)(v) = E[f(\mathbf{X}^v(t))] = \sum_{x \in C_n} p_t(v,x) f(x)$$

for $v \in C_n$, where $p_t(v,x)$ denotes the transition probability from v to x for the process $(\mathbf{X}(t))_{t \in [0,\infty)}$ - here we understand the process as a set of transition rules, independent of the starting point.

Clearly, P_t is a linear operator with the following properties:

- $P_t \mathbf{1} = \mathbf{1}$ (this reads as the invariance of the product probability measure on C_n **because the semigroup is symmetric**),
- $f \geq 0$ a.s implies $P_t f \geq 0$ a.s. (positivity preserving); because of the linearity of P_t the second condition may be equvialently stated as
- $f \geq g$ a.s. implies $P_t f \geq P_t g$ a.s. (order preserving).

Semigroup property

Obviously, $P_0 = Id$, i.e. $P_0 f \equiv f$, and for $t, s \geq 0$ we have $P_t \circ P_s = P_{t+s}$ (the semigroup property). Indeed,

$$(P_t(P_s f))(v) = E[(P_s f)(\mathbf{X}^v(t))] = \sum_{x \in C_n} P(\mathbf{X}^v(t) = x) \cdot (P_s f)(x) =$$

$$= \sum_{x \in C_n} p_t(v,x) \cdot E[f(\mathbf{X}^x(s))] = \sum_{x \in C_n} (p_t(v,x) \sum_{y \in C_n} P(\mathbf{X}^x(s) = y)) \cdot f(y))$$

$$= \sum_{y \in C_n} \Big(\sum_{x \in C_n} p_t(v,x) \cdot P(\mathbf{X}^x(s) = y) \Big) f(y) =$$

$$= \sum_{y \in C_n} \Big(\sum_{x \in C_n} p_t(v,x) p_s(x,y) \Big) f(y) =$$

$$= \sum_{y \in C_n} p_{t+s}(v,y) f(y) = E[f(\mathbf{X}^v(t+s))] = (P_{t+s} f)(v),$$

where we have used the Chapman-Kolmogorov equation.

Markov semigroups

A semigroup, indexed by a time parameter $t \in [0, \infty)$, of linear operators on $L^2(\Omega, \mu)$ which preserve positivity and the constant function **1** is called **Markovian**.

We have proved that $(P_t)_{t\in[0,\infty)}$ is a Markov semigroup.

The Markovianity of a semigroup of linear operators and possibility of defining it via some time homogenous Markov process, as we did for $(P_t)_{t\in[0,\infty)}$, are strongly related.

We have already seen one way implication - the way in which we verified Markovian properties of $(P_t)_{t\in[0,\infty)}$ did not really use any specific property of C_n and thus can be easily generalized. Now we need to understand how to produce a homogenous Markov process given a Markov semigroup $(Q_t)_{t\in[0,\infty)}$. We will discuss it in the case of a finite Ω (all atoms with non-zero measure).

Markov semigroups

A semigroup, indexed by a time parameter $t \in [0, \infty)$, of linear operators on $L^2(\Omega, \mu)$ which preserve positivity and the constant function $\mathbf{1}$ is called **Markovian**.

We have proved that $(P_t)_{t\in[0,\infty)}$ is a Markov semigroup.

The Markovianity of a semigroup of linear operators and possibility of defining it via some time homogenous Markov process, as we did for $(P_t)_{t\in[0,\infty)}$, are strongly related.

We have already seen one way implication - the way in which we verified Markovian properties of $(P_t)_{t\in[0,\infty)}$ did not really use any specific property of C_n and thus can be easily generalized. Now we need to understand how to produce a homogenous Markov process given a Markov semigroup $(Q_t)_{t\in[0,\infty)}$. We will discuss it in the case of a finite Ω (all atoms with non-zero measure).

Markov semigroups

A semigroup, indexed by a time parameter $t \in [0, \infty)$, of linear operators on $L^2(\Omega, \mu)$ which preserve positivity and the constant function $\mathbf{1}$ is called **Markovian**.

We have proved that $(P_t)_{t\in[0,\infty)}$ is a Markov semigroup.

The Markovianity of a semigroup of linear operators and possibility of defining it via some time homogenous Markov process, as we did for $(P_t)_{t\in[0,\infty)}$, are strongly related.

We have already seen one way implication - the way in which we verified Markovian properties of $(P_t)_{t\in[0,\infty)}$ did not really use any specific property of C_n and thus can be easily generalized. Now we need to understand how to produce a homogenous Markov process given a Markov semigroup $(Q_t)_{t\in[0,\infty)}$. We will discuss it in the case of a finite Ω (all atoms with non-zero measure).

Markov semigroups - equivalent condition

We will describe the Markov process we look for by expressing its transition probabilities. For $x, y \in \Omega$ and $t \geq 0$ let

$$q_t(x, y) := (Q_t 1_y)(x);$$

let us recall that 1_x denotes the indicator function of $\{x\}$.

Certainly, $q_0(x, y) = (Q_0 1_y)(x) = 1_y(x) = \delta_{x,y}$. The fact that $(Q_t)_{t\in[0,\infty)}$ is positivity preserving ensures that $q_t(x, y) \geq 0$. We also see that

$$\sum_{y\in\Omega} q_t(x, y) = \sum_{y\in\Omega} (Q_t 1_y)(x) = Q_t(\sum_{y\in\Omega} 1_y)(x) = (Q_t \mathbf{1})(x) = \mathbf{1}(x) = 1.$$

Let $x, z \in \Omega$ and $t, s \geq 0$. For every $y \in \Omega$ we have $q_t(y, z) = (Q_t 1_z)(y)$ and therefore

$$Q_t 1_z = \sum_{y\in\Omega} q_t(y, z) \cdot 1_y.$$

Markov semigroups - equivalent condition

We will describe the Markov process we look for by expressing its transition probabilities. For $x, y \in \Omega$ and $t \geq 0$ let

$$q_t(x, y) := (Q_t 1_y)(x);$$

let us recall that 1_x denotes the indicator function of $\{x\}$.

Certainly, $q_0(x, y) = (Q_0 1_y)(x) = 1_y(x) = \delta_{x,y}$. The fact that $(Q_t)_{t \in [0,\infty)}$ is positivity preserving ensures that $q_t(x, y) \geq 0$. We also see that

$$\sum_{y \in \Omega} q_t(x, y) = \sum_{y \in \Omega} (Q_t 1_y)(x) = Q_t\Big(\sum_{y \in \Omega} 1_y\Big)(x) = (Q_t \mathbf{1})(x) = \mathbf{1}(x) = 1.$$

Let $x, z \in \Omega$ and $t, s \geq 0$. For every $y \in \Omega$ we have $q_t(y, z) = (Q_t 1_z)(y)$ and therefore

$$Q_t 1_z = \sum_{y \in \Omega} q_t(y, z) \cdot 1_y.$$

Markov semigroups - equivalent condition

We will describe the Markov process we look for by expressing its transition probabilities. For $x, y \in \Omega$ and $t \geq 0$ let

$$q_t(x, y) := (Q_t 1_y)(x);$$

let us recall that 1_x denotes the indicator function of $\{x\}$.

Certainly, $q_0(x, y) = (Q_0 1_y)(x) = 1_y(x) = \delta_{x,y}$. The fact that $(Q_t)_{t \in [0,\infty)}$ is positivity preserving ensures that $q_t(x, y) \geq 0$.
We also see that

$$\sum_{y \in \Omega} q_t(x, y) = \sum_{y \in \Omega} (Q_t 1_y)(x) = Q_t(\sum_{y \in \Omega} 1_y)(x) = (Q_t \mathbf{1})(x) = \mathbf{1}(x) = 1.$$

Let $x, z \in \Omega$ and $t, s \geq 0$. For every $y \in \Omega$ we have $q_t(y, z) = (Q_t 1_z)(y)$ and therefore

$$Q_t 1_z = \sum_{y \in \Omega} q_t(y, z) \cdot 1_y.$$

Markov semigroups - equivalence continued

Hence

$$q_{s+t}(x,z) = (Q_{s+t}1_z)(x) = \Big(Q_s(Q_t 1_z)\Big)(x) =$$

$$= \Big(Q_s(\sum_{y\in\Omega} q_t(y,z)\cdot 1_y)\Big)(x) =$$

$$= \sum_{y\in\Omega} q_t(y,z)(Q_s 1_y)(x) = \sum_{y\in\Omega} q_t(y,z) q_s(x,y).$$

We have verified the Chapman-Kolmogorov equation and thus finished the proof that $q_t(x,y)$ defined as above is a consistent family of transition probabilities.

Now it only remains to prove that the process defined by the above transition probabilities yields the same semigroup which we started with, i.e.

$$E[f(\mathbf{X}^v(t))] = (Q_t f)(v)$$

for every function $f : \Omega \longrightarrow \mathbf{R}$.

Hence

$$q_{s+t}(x,z) = (Q_{s+t}1_z)(x) = \Big(Q_s(Q_t 1_z)\Big)(x) =$$

$$= \Big(Q_s(\sum_{y\in\Omega} q_t(y,z)\cdot 1_y)\Big)(x) =$$

$$= \sum_{y\in\Omega} q_t(y,z)(Q_s 1_y)(x) = \sum_{y\in\Omega} q_t(y,z) q_s(x,y).$$

We have verified the Chapman-Kolmogorov equation and thus finished the proof that $q_t(x,y)$ defined as above is a consistent family of transition probabilities.

Now it only remains to prove that the process defined by the above transition probabilities yields the same semigroup which we started with, i.e.

$$E[f(\mathbf{X}^v(t))] = (Q_t f)(v)$$

for every function $f : \Omega \longrightarrow \mathbf{R}$.

Equivalence - this is the end...

$$E[f(\mathbf{X}^v(t))] \stackrel{?}{=} (Q_t f)(v)$$

If $f = 1_y$ for some $y \in \Omega$ then the above follows just from the very way in which we defined our process:

$$q_t(v, y) = (Q_t 1_y)(x).$$

By the linearity (with respect to f) the equation holds for every f as well, and the proof is finished.

Remark: In general setting, $q_t(v, A) = (Q_t 1_A)(v)$.

Equivalence - this is the end...

$$E[f(\mathbf{X}^v(t))] \stackrel{?}{=} (Q_t f)(v)$$

If $f = 1_y$ for some $y \in \Omega$ then the above follows just from the very way in which we defined our process:

$$q_t(v, y) = (Q_t 1_y)(x).$$

By the linearity (with respect to f) the equation holds for every f as well, and the proof is finished.

Remark: In general setting, $q_t(v, A) = (Q_t 1_A)(v)$.

Equivalence - this is the end...

$$E[f(\mathbf{X}^v(t))] \stackrel{?}{=} (Q_t f)(v)$$

If $f = 1_y$ for some $y \in \Omega$ then the above follows just from the very way in which we defined our process:

$$q_t(v, y) = (Q_t 1_y)(x).$$

By the linearity (with respect to f) the equation holds for every f as well, and the proof is finished.

Remark: In general setting, $q_t(v, A) = (Q_t 1_A)(v)$.

Invariant measure

We will say that a probability measure μ on Ω is an **invariant measure** for our semigroup, or a **stationary distribution** for our Markov process, if for every $y \in \Omega$ and $t > 0$ there is

$$\sum_{x \in \Omega} \mu(\{x\}) q_t(x, y) = \mu(\{y\}),$$

so that the total "immigration" to y balances "emigration" from y.

It amounts to $E[Q_t 1_y] = \sum_{x \in \Omega} \mu(\{x\})(Q_t 1_y)(x) = E[1_y]$, so that μ is an invariant measure for our semigroup if and only if Q_t's preserve expectation for all 1_y's, i.e., if and only if Q_t's preserve expectation for all functions.

As we will see soon, if the semigroup is symmetric and it preserves the constant function $\mathbf{1}$ then it also preserves expectation. Conversely, if the semigroup is symmetric and it preserves expectation then $Q_t \mathbf{1} = \mathbf{1}$.

Invariant measure

We will say that a probability measure μ on Ω is an **invariant measure** for our semigroup, or a **stationary distribution** for our Markov process, if for every $y \in \Omega$ and $t > 0$ there is

$$\sum_{x \in \Omega} \mu(\{x\}) q_t(x, y) = \mu(\{y\}),$$

so that the total "immigration" to y balances "emigration" from y.

It amounts to $E[Q_t 1_y] = \sum_{x \in \Omega} \mu(\{x\})(Q_t 1_y)(x) = E[1_y]$, so that μ is an invariant measure for our semigroup if and only if Q_t's preserve expectation for all 1_y's, i.e., if and only if Q_t's preserve expectation for all functions.

As we will see soon, if the semigroup is symmetric and it preserves the constant function $\mathbf{1}$ then it also preserves expectation. Conversely, if the semigroup is symmetric and it preserves expectation then $Q_t \mathbf{1} = \mathbf{1}$.

Invariant measure

We will say that a probability measure μ on Ω is an **invariant measure** for our semigroup, or a **stationary distribution** for our Markov process, if for every $y \in \Omega$ and $t > 0$ there is

$$\sum_{x \in \Omega} \mu(\{x\}) q_t(x, y) = \mu(\{y\}),$$

so that the total "immigration" to y balances "emigration" from y.

It amounts to $E[Q_t 1_y] = \sum_{x \in \Omega} \mu(\{x\})(Q_t 1_y)(x) = E[1_y]$, so that μ is an invariant measure for our semigroup if and only if Q_t's preserve expectation for all 1_y's, i.e., if and only if Q_t's preserve expectation for all functions.

As we will see soon, if the semigroup is symmetric and it preserves the constant function $\mathbf{1}$ then it also preserves expectation. Conversely, if the semigroup is symmetric and it preserves expectation then $Q_t \mathbf{1} = \mathbf{1}$.

Generator

Warning: A Markovian (in the sense described above) semigroup $(Q_t)_{t\in[0,\infty)}$ need **not** be continuous with respect to the parameter t. As an example one may consider $Q_0 f \equiv f$ and $Q_t f \equiv E[f]$ for $t > 0$ which is not time continuous unless f is constant a.s.

However, in many cases Markov semigroups are not only continuous but also differentiable with respect to time. A linear operator defined as $-\frac{d}{dt}Q_t\Big|_{t=0^+}$ is then called a **generator** of the semigroup $(Q_t)_{t\in[0,\infty)}$. Usually it cannot be defined on the whole L^2 function space but only on its dense linear subspace. There are quite many technical problems and extensive literature concerning relations between a Markov semigroup and its generator - we will discuss them very briefly below.

Generator

Warning: A Markovian (in the sense described above) semigroup $(Q_t)_{t\in[0,\infty)}$ need **not** be continuous with respect to the parameter t. As an example one may consider $Q_0 f \equiv f$ and $Q_t f \equiv E[f]$ for $t > 0$ which is not time continuous unless f is constant a.s.

However, in many cases Markov semigroups are not only continuous but also differentiable with respect to time. A linear operator defined as $-\frac{d}{dt} Q_t \Big|_{t=0^+}$ is then called a **generator** of the semigroup $(Q_t)_{t\in[0,\infty)}$. Usually it cannot be defined on the whole L^2 function space but only on its dense linear subspace. There are quite many technical problems and extensive literature concerning relations between a Markov semigroup and its generator - we will discuss them very briefly below.

Semigroup and generator - spectral properties

Let us see how the semigroup $(P_t)_{t\in[0,\infty)}$ acts on the Walsh functions. For $v \in C_n$ and $S \subseteq [n]$ we have

$$(P_t w_S)(v) = E[w_S(\mathbf{X}^v(t))] = E\Big[\prod_{i\in S} v_i X_i(t)\Big] =$$

$$= \Big(\prod_{i\in S} v_i\Big) \cdot \prod_{i\in S} E[X_i(t)] =$$

$$= w_S(v) \cdot \Big(\frac{1+e^{-t}}{2} \cdot 1 + \frac{1-e^{-t}}{2} \cdot (-1)\Big)^{|S|} = e^{-|S|t} w_S(v).$$

Hence $P_t w_S = e^{-|S|t} w_S$. If $f = \sum_{S\subseteq[n]} a_S w_S$ then

$$P_t f = \sum_{S\subseteq[n]} e^{-|S|t} a_S w_S$$

- compare to the formula for the limit of the lazy walk.

Semigroup - notation

Remark: One may also define a family of multipliers $T_\eta : \mathcal{H}_n \longrightarrow \mathcal{H}_n$ by

$$T_\eta\Big(\sum_{S\subseteq[n]} a_S w_S\Big) := \sum_{S\subseteq[n]} \eta^{|S|} a_S w_S.$$

This notation is better adapted for harmonic analysis use. Also, it is often used in theoretical computer science. However, it is less natural from the point of view of probability theory (note: "our" $(P_t)_{t\in[0,\infty)}$ is closely related to the Ornstein-Uhlenbeck semigroup on the Gaussian space).

Clearly, $T_{e^{-t}} \equiv P_t$ for $t \geq 0$ but sometimes it makes sense to consider also $|\eta| \leq 1$ or even η from some sector on the complex plane (holomorphic semigroups), and with vector coefficients a_S. Of course, $T_\eta \circ T_\rho = T_{\eta\rho}$.

Warning: The notation in literature does vary!

Semigroup - notation

Remark: One may also define a family of multipliers $T_\eta : \mathcal{H}_n \longrightarrow \mathcal{H}_n$ by

$$T_\eta\Big(\sum_{S\subseteq[n]} a_S w_S\Big) := \sum_{S\subseteq[n]} \eta^{|S|} a_S w_S.$$

This notation is better adapted for harmonic analysis use. Also, it is often used in theoretical computer science. However, it is less natural from the point of view of probability theory (note: "our" $(P_t)_{t\in[0,\infty)}$ is closely related to the Ornstein-Uhlenbeck semigroup on the Gaussian space).

Clearly, $T_{e^{-t}} \equiv P_t$ for $t \geq 0$ but sometimes it makes sense to consider also $|\eta| \leq 1$ or even η from some sector on the complex plane (holomorphic semigroups), and with vector coefficients a_S. Of course, $T_\eta \circ T_\rho = T_{\eta\rho}$.

Warning: The notation in literature does vary!

Semigroup - notation

Remark: One may also define a family of multipliers $T_\eta : \mathcal{H}_n \longrightarrow \mathcal{H}_n$ by

$$T_\eta\Big(\sum_{S\subseteq[n]} a_S w_S\Big) := \sum_{S\subseteq[n]} \eta^{|S|} a_S w_S.$$

This notation is better adapted for harmonic analysis use. Also, it is often used in theoretical computer science. However, it is less natural from the point of view of probability theory (note: "our" $(P_t)_{t\in[0,\infty)}$ is closely related to the Ornstein-Uhlenbeck semigroup on the Gaussian space).

Clearly, $T_{e^{-t}} \equiv P_t$ for $t \geq 0$ but sometimes it makes sense to consider also $|\eta| \leq 1$ or even η from some sector on the complex plane (holomorphic semigroups), and with vector coefficients a_S. Of course, $T_\eta \circ T_\rho = T_{\eta\rho}$.

Warning: The notation in literature does vary!

Time derivative

We know that

$$P_t w_S = e^{-|S|t} w_S.$$

Now it is easy to differentiate P_t :

$$\frac{d}{dt} P_t w_S = -|S| e^{-|S|t} w_S = -|S| P_t w_S.$$

Let $L : \mathcal{H}_n \longrightarrow \mathcal{H}_n$ be a linear operator defined by $L w_S := |S| w_S$, i.e.

$$L\Big(\sum_{S \subseteq [n]} a_S w_S\Big) = \sum_{S \subseteq [n]} |S| a_S w_S.$$

We have proved that $\frac{d}{dt} P_t f = -L P_t f = -P_t L f$ (the multipliers P_t and L obviously commute) and in particular $\frac{d}{dt} P_t \Big|_{t=0^+} = -L$.

Since for **every** Markovian semigroup $Q_t \mathbf{1} = \mathbf{1}$ for all $t \geq 0$,

$$L\mathbf{1} = -\frac{d}{dt} Q_t \mathbf{1}\Big|_{t=0^+} \equiv 0.$$

Time derivative

We know that

$$P_t w_S = e^{-|S|t} w_S.$$

Now it is easy to differentiate P_t :

$$\frac{d}{dt} P_t w_S = -|S| e^{-|S|t} w_S = -|S| P_t w_S.$$

Let $L : \mathcal{H}_n \longrightarrow \mathcal{H}_n$ be a linear operator defined by $L w_S := |S| w_S$, i.e.

$$L\Big(\sum_{S \subseteq [n]} a_S w_S \Big) = \sum_{S \subseteq [n]} |S| a_S w_S.$$

We have proved that $\frac{d}{dt} P_t f = -L P_t f = -P_t L f$ (the multipliers P_t and L obviously commute) and in particular $\frac{d}{dt} P_t \Big|_{t=0^+} = -L$.

Since for **every** Markovian semigroup $Q_t \mathbf{1} = \mathbf{1}$ for all $t \geq 0$,

$$L\mathbf{1} = -\frac{d}{dt} Q_t \mathbf{1} \Big|_{t=0^+} \equiv 0.$$

Time derivative

We know that

$$P_t w_S = e^{-|S|t} w_S.$$

Now it is easy to differentiate P_t :

$$\frac{d}{dt} P_t w_S = -|S| e^{-|S|t} w_S = -|S| P_t w_S.$$

Let $L : \mathcal{H}_n \longrightarrow \mathcal{H}_n$ be a linear operator defined by $L w_S := |S| w_S$, i.e.

$$L\Big(\sum_{S \subseteq [n]} a_S w_S \Big) = \sum_{S \subseteq [n]} |S| a_S w_S.$$

We have proved that $\frac{d}{dt} P_t f = -L P_t f = -P_t L f$ (the multipliers P_t and L obviously commute) and in particular $\frac{d}{dt} P_t \Big|_{t=0^+} = -L$.

Since for **every** Markovian semigroup $Q_t \mathbf{1} = \mathbf{1}$ for all $t \geq 0$,

$$L\mathbf{1} = -\frac{d}{dt} Q_t \mathbf{1} \Big|_{t=0^+} \equiv 0.$$

Retrieval of the semigroup from its generator

We can recover the semigroup from its generator by

$$P_t = e^{-tL} = Id + \sum_{k=1}^{\infty} (-t)^k L^k / k!.$$

Indeed,

$$e^{-tL} w_S = w_S + \sum_{k=1}^{\infty} (-1)^k t^k |S|^k w_S / k! = e^{-|S| t} w_S = P_t w_S.$$

However, this approach works well if L is bounded and thus cannot be easily generalized.

By writing a Markov semigroup $(Q_t)_{t \in [0,\infty)}$ in the form e^{-tL} one usually means that it is a solution of an operator differential equation $\frac{d}{dt} Q_t = -LQ_t$ $(t \geq 0)$ with a boundary condition $Q_0 = Id$.

Retrieval of the semigroup from its generator

We can recover the semigroup from its generator by

$$P_t = e^{-tL} = Id + \sum_{k=1}^{\infty} (-t)^k L^k / k!.$$

Indeed,

$$e^{-tL} w_S = w_S + \sum_{k=1}^{\infty} (-1)^k t^k |S|^k w_S / k! = e^{-|S|t} w_S = P_t w_S.$$

However, this approach works well if L is bounded and thus cannot be easily generalized.

By writing a Markov semigroup $(Q_t)_{t\in[0,\infty)}$ in the form e^{-tL} one usually means that it is a solution of an operator differential equation $\frac{d}{dt} Q_t = -LQ_t$ $(t \geq 0)$ with a boundary condition $Q_0 = Id$.

Explicit definition of the generator

There are more direct ways to define L. For any $f \in \mathcal{H}_n$ there is

$$(Lf)(x) = \frac{1}{2} \sum_{y \in C_n\colon d(x,y)=1} \Big(f(x) - f(y) \Big) = \frac{n}{2} f(x) - \frac{1}{2} \cdot \sum_{y \in C_n\colon d(x,y)=1} f(y),$$

i.e. $L = \frac{n}{2}(Id - K)$, where K is the operator related to the discrete time random walk on C_n. Indeed, it suffices to recall that

$$\sum_{y \in C_n\colon d(x,y)=1} w_S(y) = (n - |S|) \cdot w_S(x) + |S| \cdot (-w_S(x)) = (n - 2|S|) w_S(x),$$

so that

$$\frac{1}{2} \cdot \sum_{y \in C_n\colon d(x,y)=1} \Big(w_S(x) - w_S(y) \Big) = |S| w_S(x) = (L w_S)(x).$$

Now we can use the fact that the Walsh functions span $\mathcal{H}_n$.

Explicit definition of the generator

There are more direct ways to define L. For any $f \in \mathcal{H}_n$ there is

$$(Lf)(x) = \frac{1}{2} \sum_{y \in C_n \colon d(x,y)=1} \Big(f(x) - f(y) \Big) = \frac{n}{2} f(x) - \frac{1}{2} \cdot \sum_{y \in C_n \colon d(x,y)=1} f(y),$$

i.e. $L = \frac{n}{2}(Id - K)$, where K is the operator related to the discrete time random walk on C_n. Indeed, it suffices to recall that

$$\sum_{y \in C_n \colon d(x,y)=1} w_S(y) = (n - |S|) \cdot w_S(x) + |S| \cdot (-w_S(x)) = (n - 2|S|) w_S(x),$$

so that

$$\frac{1}{2} \cdot \sum_{y \in C_n \colon d(x,y)=1} \Big(w_S(x) - w_S(y) \Big) = |S| w_S(x) = (L w_S)(x).$$

Now we can use the fact that the Walsh functions span $\mathcal{H}_n$.

Symmetry

The operators P_t and L are symmetric (in fact, also bounded and thus self-adjoint). We can expand every $f, g \in \mathcal{H}_n$ as $f = \sum_{S\subseteq[n]} a_S w_S$ and $g = \sum_{S\subseteq[n]} b_S w_S$, and arrive at

$$E[f \cdot P_t g] = \langle f, P_t g\rangle = \sum_{S,T\subseteq[n]} a_S e^{-|T|t} b_T \langle w_S, w_T\rangle =$$

$$= \sum_{S\subseteq[n]} e^{-|S|t} a_S b_S = \langle P_t f, g\rangle = E[P_t f \cdot g]$$

and

$$E[f \cdot Lg] = \langle f, Lg\rangle = \sum_{S,T\subseteq[n]} a_S |T| b_T \langle w_S, w_T\rangle =$$

$$= \sum_{S\subseteq[n]} |S| a_S b_S = \langle Lf, g\rangle = E[Lf \cdot g].$$

Symmetry

The operators P_t and L are symmetric (in fact, also bounded and thus self-adjoint). We can expand every $f, g \in \mathcal{H}_n$ as $f = \sum_{S\subseteq[n]} a_S w_S$ and $g = \sum_{S\subseteq[n]} b_S w_S$, and arrive at

$$E[f \cdot P_t g] = \langle f, P_t g\rangle = \sum_{S,T\subseteq[n]} a_S e^{-|T|t} b_T \langle w_S, w_T\rangle =$$

$$= \sum_{S\subseteq[n]} e^{-|S|t} a_S b_S = \langle P_t f, g\rangle = E[P_t f \cdot g]$$

and

$$E[f \cdot Lg] = \langle f, Lg\rangle = \sum_{S,T\subseteq[n]} a_S |T| b_T \langle w_S, w_T\rangle =$$

$$= \sum_{S\subseteq[n]} |S| a_S b_S = \langle Lf, g\rangle = E[Lf \cdot g].$$

Warning: If a Markov semigroup $(Q_t)_{t\in[0,\infty)}$ is symmetric in the above sense it usually does **not** mean that $q_t(x,y) = q_t(y,x)$. Indeed, in the case of a finite probability space (Ω, μ) we have rather

$$\mu(\{x\}) \cdot q_t(x,y) = \mu(\{x\}) \cdot (Q_t 1_y)(x) = E[1_x Q_t 1_y] =$$

$$= E[1_y Q_t 1_x] = \mu(\{y\}) \cdot (Q_t 1_x)(y) = \mu(\{y\}) \cdot q_t(y,x).$$

The above concept of symmetry is equivalent to the symmetry of the transition matrix for C_n equipped with the uniform probability measure only because all atoms have equal measure in this case. In general, the symmetry meant here is the symmetry of operators with respect to the $L^2(\Omega, \mu)$ structure.

Symmetry - warning

Warning: If a Markov semigroup $(Q_t)_{t\in[0,\infty)}$ is symmetric in the above sense it usually does **not** mean that $q_t(x,y) = q_t(y,x)$. Indeed, in the case of a finite probability space (Ω,μ) we have rather

$$\mu(\{x\})\cdot q_t(x,y) = \mu(\{x\})\cdot (Q_t 1_y)(x) = E[1_x Q_t 1_y] =$$

$$= E[1_y Q_t 1_x] = \mu(\{y\})\cdot (Q_t 1_x)(y) = \mu(\{y\})\cdot q_t(y,x).$$

The above concept of symmetry is equivalent to the symmetry of the transition matrix for C_n equipped with the uniform probability measure only because all atoms have equal measure in this case. In general, the symmetry meant here is the symmetry of operators with respect to the $L^2(\Omega,\mu)$ structure.

Symmetry - a simple consequence

Let us note that $(P_t)_{t\in[0,\infty)}$ preserves expectation:

$$E[P_t f] = E[\mathbf{1}\cdot P_t f] = E[P_t\mathbf{1}\cdot f] = E[\mathbf{1}\cdot f] = E[f]$$

for every $t \geq 0$.

Obviously, the same holds for any **symmetric** Markov semigroup (as long as the expectation is taken with respect to the underlying invariant probability measure).

Symmetry - a simple consequence

Let us note that $(P_t)_{t\in[0,\infty)}$ preserves expectation:

$$E[P_t f] = E[\mathbf{1} \cdot P_t f] = E[P_t \mathbf{1} \cdot f] = E[\mathbf{1} \cdot f] = E[f]$$

for every $t \geq 0$.

Obviously, the same holds for any **symmetric** Markov semigroup (as long as the expectation is taken with respect to the underlying invariant probability measure).

Contractivity

The semigroup $(P_t)_{t\in[0,\infty)}$ is contractive in L^p for every $p \geq 1$, i.e.

$$\forall_{f\in\mathcal{H}_n} \; \|P_t f\|_p \leq \|f\|_p.$$

We will prove a more general fact (the above is just the case $\Phi(t) = |t|^p$):

For every convex function $\Phi : \mathbf{R} \longrightarrow \mathbf{R}$ there is

$$\forall_{t\geq 0,\, f\in\mathcal{H}_n} \; E[\Phi(P_t f)] \leq E[\Phi(f)].$$

Indeed, $\Phi(x) = \sup_\alpha (a_\alpha x + b_\alpha)$ - every convex function is a supremum of its supporting affine functions. For every α the pointwise inequalities $\Phi(f) \geq a_\alpha f + b_\alpha$ and, due to the order preserving property of $(P_t)_{t\in[0,\infty)}$, also

$$P_t(\Phi(f)) \geq P_t(a_\alpha f + b_\alpha) = a_\alpha P_t f + b_\alpha$$

hold.

Contractivity

The semigroup $(P_t)_{t\in[0,\infty)}$ is contractive in L^p for every $p \geq 1$, i.e.

$$\forall_{f\in\mathcal{H}_n} \ \|P_t f\|_p \leq \|f\|_p.$$

We will prove a more general fact (the above is just the case $\Phi(t) = |t|^p$):

For every convex function $\Phi : \mathbf{R} \longrightarrow \mathbf{R}$ there is

$$\forall_{t\geq 0,\, f\in\mathcal{H}_n} \ E[\Phi(P_t f)] \leq E[\Phi(f)].$$

Indeed, $\Phi(x) = \sup_\alpha (a_\alpha x + b_\alpha)$ - every convex function is a supremum of its supporting affine functions. For every α the pointwise inequalities $\Phi(f) \geq a_\alpha f + b_\alpha$ and, due to the order preserving property of $(P_t)_{t\in[0,\infty)}$, also

$$P_t(\Phi(f)) \geq P_t(a_\alpha f + b_\alpha) = a_\alpha P_t f + b_\alpha$$

hold.

Contractivity

The semigroup $(P_t)_{t\in[0,\infty)}$ is contractive in L^p for every $p \geq 1$, i.e.

$$\forall_{f\in\mathcal{H}_n} \;\; \|P_t f\|_p \leq \|f\|_p.$$

We will prove a more general fact (the above is just the case $\Phi(t) = |t|^p$):

For every convex function $\Phi : \mathbf{R} \longrightarrow \mathbf{R}$ there is

$$\forall_{t\geq 0,\, f\in\mathcal{H}_n} \;\; E[\Phi(P_t f)] \leq E[\Phi(f)].$$

Indeed, $\Phi(x) = \sup_\alpha(a_\alpha x + b_\alpha)$ - every convex function is a supremum of its supporting affine functions. For every α the pointwise inequalities $\Phi(f) \geq a_\alpha f + b_\alpha$ and, due to the order preserving property of $(P_t)_{t\in[0,\infty)}$, also

$$P_t(\Phi(f)) \geq P_t(a_\alpha f + b_\alpha) = a_\alpha P_t f + b_\alpha$$

hold.

Hence

$$P_t(\Phi(f)) \geq \sup_\alpha (a_\alpha P_t f + b_\alpha) = \Phi(P_t f)$$

pointwise and we infer $E[\Phi(f)] = E[P_t(\Phi(f))] \geq E[\Phi(P_t f)]$, where we have used the fact that $(P_t)_{t\in[0,\infty)}$ preserves expectation.

Thus the semigroup $(P_t)_{t\in[0,\infty)}$ is called a semigroup of **contractions.** Indeed, we have proved that $\|P_t\|_{L^p\to L^p} \leq 1$ for $p \in [1,\infty)$ and for $p = \infty$ this is a consequence of the fact that $(P_t)_{t\in[0,\infty)}$ preserves order: $-m \leq f \leq m$ a.s. implies $-m = P_t(-m) \leq P_t f \leq P_t m = m$ a.s.

Certainly, a similar reasoning works for every **symmetric** Markov semigroup.

Contractivity - continued

Hence

$$P_t(\Phi(f)) \geq \sup_\alpha (a_\alpha P_t f + b_\alpha) = \Phi(P_t f)$$

pointwise and we infer $E[\Phi(f)] = E[P_t(\Phi(f))] \geq E[\Phi(P_t f)]$, where we have used the fact that $(P_t)_{t\in[0,\infty)}$ preserves expectation.

Thus the semigroup $(P_t)_{t\in[0,\infty)}$ is called a semigroup of **contractions.** Indeed, we have proved that $\|P_t\|_{L^p\to L^p} \leq 1$ for $p \in [1,\infty)$ and for $p = \infty$ this is a consequence of the fact that $(P_t)_{t\in[0,\infty)}$ preserves order: $-m \leq f \leq m$ a.s. implies $-m = P_t(-m) \leq P_t f \leq P_t m = m$ a.s.

Certainly, a similar reasoning works for every **symmetric** Markov semigroup.

Contractivity - continued

Hence

$$P_t(\Phi(f)) \geq \sup_\alpha (a_\alpha P_t f + b_\alpha) = \Phi(P_t f)$$

pointwise and we infer $E[\Phi(f)] = E[P_t(\Phi(f))] \geq E[\Phi(P_t f)]$, where we have used the fact that $(P_t)_{t\in[0,\infty)}$ preserves expectation.

Thus the semigroup $(P_t)_{t\in[0,\infty)}$ is called a semigroup of **contractions.** Indeed, we have proved that $\|P_t\|_{L^p \to L^p} \leq 1$ for $p \in [1, \infty)$ and for $p = \infty$ this is a consequence of the fact that $(P_t)_{t\in[0,\infty)}$ preserves order: $-m \leq f \leq m$ a.s. implies $-m = P_t(-m) \leq P_t f \leq P_t m = m$ a.s.

Certainly, a similar reasoning works for every **symmetric** Markov semigroup.

Energy functional

Let us consider a bi-linear form $\mathcal{E} = \mathcal{E}_L : \mathcal{H}_n \times \mathcal{H}_n \longrightarrow \mathbf{R}$ defined by

$$\mathcal{E}(f, g) := E[f \cdot Lg].$$

Since $E[f \cdot Lg] = E[Lf \cdot g]$ the form is symmetric. It is also positive semi-definite.

Indeed, for $t \geq 0$ and $f \in \mathcal{H}_n$ let us set $\psi(t) = \|P_t f\|_2^2 = E[(P_t f)^2]$. Then

$$\psi'(t) = E\left[2P_t f \cdot \frac{d}{dt} P_t f\right] = E\left[2P_t f \cdot -LP_t f\right],$$

so that $\psi'(0^+) = -2E[f \cdot Lf] = -2\mathcal{E}(f, f)$. On the other hand, because of the contractivity of $(P_t)_{t\in[0,\infty)}$ we have $\psi(t) \leq \|f\|_2^2 = \|P_0 f\|_2^2 = \psi(0)$ for $t \geq 0$, so that $\psi'(0^+) \leq 0$. Thus $\mathcal{E}[f] := \mathcal{E}(f, f) \geq 0$.

The above proof works (up to quite many technical details) for a large subclass of **symmetric** Markov semigroups.

Energy functional

Let us consider a bi-linear form $\mathcal{E} = \mathcal{E}_L : \mathcal{H}_n \times \mathcal{H}_n \longrightarrow \mathbf{R}$ defined by

$$\mathcal{E}(f, g) := E[f \cdot Lg].$$

Since $E[f \cdot Lg] = E[Lf \cdot g]$ the form is symmetric. It is also positive semi-definite.

Indeed, for $t \geq 0$ and $f \in \mathcal{H}_n$ let us set
$\psi(t) = \|P_t f\|_2^2 = E[(P_t f)^2]$. Then

$$\psi'(t) = E\Big[2P_t f \cdot \frac{d}{dt} P_t f\Big] = E\Big[2P_t f \cdot -LP_t f\Big],$$

so that $\psi'(0^+) = -2E[f \cdot Lf] = -2\mathcal{E}(f, f)$. On the other hand, because of the contractivity of $(P_t)_{t \in [0,\infty)}$ we have
$\psi(t) \leq \|f\|_2^2 = \|P_0 f\|_2^2 = \psi(0)$ for $t \geq 0$, so that $\psi'(0^+) \leq 0$.
Thus $\mathcal{E}[f] := \mathcal{E}(f, f) \geq 0$.

The above proof works (up to quite many technical details) for a large subclass of **symmetric** Markov semigroups.

Energy functional

Let us consider a bi-linear form $\mathcal{E} = \mathcal{E}_L : \mathcal{H}_n \times \mathcal{H}_n \longrightarrow \mathbf{R}$ defined by

$$\mathcal{E}(f, g) := E[f \cdot Lg].$$

Since $E[f \cdot Lg] = E[Lf \cdot g]$ the form is symmetric. It is also positive semi-definite.

Indeed, for $t \geq 0$ and $f \in \mathcal{H}_n$ let us set
$\psi(t) = \|P_t f\|_2^2 = E[(P_t f)^2]$. Then

$$\psi'(t) = E\Big[2P_t f \cdot \frac{d}{dt} P_t f\Big] = E\Big[2P_t f \cdot -LP_t f\Big],$$

so that $\psi'(0^+) = -2E[f \cdot Lf] = -2\mathcal{E}(f, f)$. On the other hand, because of the contractivity of $(P_t)_{t \in [0,\infty)}$ we have
$\psi(t) \leq \|f\|_2^2 = \|P_0 f\|_2^2 = \psi(0)$ for $t \geq 0$, so that $\psi'(0^+) \leq 0$.
Thus $\mathcal{E}[f] := \mathcal{E}(f, f) \geq 0$.

The above proof works (up to quite many technical details) for a large subclass of **symmetric** Markov semigroups.

Energy functional - elementary approach

However, for the semigroup $(P_t)_{t\in[0,\infty)}$ the positive semi-definiteness of its generator is related to a more elementary observation.

Recall that

$$(Lf)(x) = \frac{1}{2} \cdot \sum_{y\in C_n:\, d(x,y)=1} \Big(f(x) - f(y)\Big).$$

Hence, for any $f, g \in \mathcal{H}_n$ we have

$$\mathcal{E}(f,g) = E[f \cdot Lg] = \frac{2^{-n}}{2} \sum_{x\in C_n} \sum_{y\in C_n:\, d(x,y)=1} f(x)\Big(g(x) - g(y)\Big) =$$

$$= 2^{-n-1} \sum_{x,y\in C_n:\, d(x,y)=1} \Big(f(x)g(x) - f(x)g(y)\Big) =$$

$$= 2^{-n-1} \sum_{x,y\in C_n:\, d(x,y)=1} \Big(f(y)g(y) - f(y)g(x)\Big).$$

Energy functional - elementary approach

However, for the semigroup $(P_t)_{t\in[0,\infty)}$ the positive semi-definiteness of its generator is related to a more elementary observation.

Recall that

$$(Lf)(x) = \frac{1}{2} \cdot \sum_{y\in C_n:\, d(x,y)=1} \Big(f(x) - f(y)\Big).$$

Hence, for any $f, g \in \mathcal{H}_n$ we have

$$\mathcal{E}(f,g) = E[f \cdot Lg] = \frac{2^{-n}}{2} \sum_{x\in C_n} \sum_{y\in C_n:\, d(x,y)=1} f(x)\Big(g(x) - g(y)\Big) =$$

$$= 2^{-n-1} \sum_{x,y\in C_n:\, d(x,y)=1} \Big(f(x)g(x) - f(x)g(y)\Big) =$$

$$= 2^{-n-1} \sum_{x,y\in C_n:\, d(x,y)=1} \Big(f(y)g(y) - f(y)g(x)\Big).$$

Discrete gradient

$$f(x)g(x) - f(x)g(y) + f(y)g(y) - f(y)g(x) = (f(x) - f(y))(g(x) - g(y)),$$

so

$$\mathcal{E}(f,g) = 2^{-n-2} \sum_{x,y \in C_n:\, d(x,y)=1} \Big(f(x) - f(y)\Big)\Big(g(x) - g(y)\Big),$$

in particular

$$E[f\,Lf] = \mathcal{E}[f] = \mathcal{E}(f,f) = 2^{-n} \sum_{x,y \in C_n:\, d(x,y)=1} \left(\frac{f(x) - f(y)}{2}\right)^2$$

is clearly nonnegative.

The last expression is a discrete counterpart of the averaged $|\nabla f|^2$ - the similarity to the physical kinetic energy notion explains the name given to this quadratic form. Quadratic forms of this type (under some additional conditions) are called Dirichlet forms and play important role in the theory of Markov semigroups.

process $\sim$ semigroup $\sim$ generator $\sim$ Dirichlet form

Discrete gradient

$f(x)g(x)-f(x)g(y)+f(y)g(y)-f(y)g(x) = (f(x)-f(y))(g(x)-g(y))$,
so

$$\mathcal{E}(f,g) = 2^{-n-2} \sum_{x,y \in C_n :\, d(x,y)=1} \Big(f(x)-f(y)\Big)\Big(g(x)-g(y)\Big),$$

in particular

$$E[f\,Lf] = \mathcal{E}[f] = \mathcal{E}(f,f) = 2^{-n} \sum_{x,y \in C_n :\, d(x,y)=1} \left(\frac{f(x)-f(y)}{2}\right)^2$$

is clearly nonnegative.

The last expression is a discrete counterpart of the averaged $|\nabla f|^2$ - the similarity to the physical kinetic energy notion explains the name given to this quadratic form. Quadratic forms of this type (under some additional conditions) are called Dirichlet forms and play important role in the theory of Markov semigroups.

process $\sim$ semigroup $\sim$ generator $\sim$ Dirichlet form

Energy under Lipschitz map

Let $\Psi : \mathbf{R} \longrightarrow \mathbf{R}$ be a Lipschitz map with constant C, i.e. $|\Psi(a) - \Psi(b)| \leq C|a-b|$.

Obviously, for any $f \in \mathcal{H}_n$ we have

$$\mathcal{E}[\Psi(f)] = 2^{-n} \sum_{x,y \in C_n:\, d(x,y)=1} \Big(\frac{\Psi(f(x)) - \Psi(f(y))}{2}\Big)^2 \leq$$

$$\leq 2^{-n} \sum_{x,y \in C_n:\, d(x,y)=1} C^2 \Big(\frac{f(x) - f(y)}{2}\Big)^2 = C^2 \mathcal{E}[f].$$

In particular, since $\Psi(a) = |a|$ is 1-Lipschitz we get

$$E[|f| L(|f|)] = \mathcal{E}[|f|] \leq \mathcal{E}[f] = E[f\, Lf].$$

Energy under Lipschitz map

Let $\Psi : \mathbf{R} \longrightarrow \mathbf{R}$ be a Lipschitz map with constant C, i.e. $|\Psi(a) - \Psi(b)| \leq C|a - b|$.

Obviously, for any $f \in \mathcal{H}_n$ we have

$$\mathcal{E}[\Psi(f)] = 2^{-n} \sum_{x,y \in C_n:\, d(x,y)=1} \left(\frac{\Psi(f(x)) - \Psi(f(y))}{2}\right)^2 \leq$$

$$\leq 2^{-n} \sum_{x,y \in C_n:\, d(x,y)=1} C^2 \left(\frac{f(x) - f(y)}{2}\right)^2 = C^2 \mathcal{E}[f].$$

In particular, since $\Psi(a) = |a|$ is 1-Lipschitz we get

$$E[|f| L(|f|)] = \mathcal{E}[|f|] \leq \mathcal{E}[f] = E[f\, Lf].$$

Energy under Lipschitz map

Let $\Psi : \mathbf{R} \longrightarrow \mathbf{R}$ be a Lipschitz map with constant C, i.e. $|\Psi(a) - \Psi(b)| \leq C|a - b|$.

Obviously, for any $f \in \mathcal{H}_n$ we have

$$\mathcal{E}[\Psi(f)] = 2^{-n} \sum_{x,y \in C_n \colon d(x,y)=1} \Big(\frac{\Psi(f(x)) - \Psi(f(y))}{2}\Big)^2 \leq$$

$$\leq 2^{-n} \sum_{x,y \in C_n \colon d(x,y)=1} C^2 \Big(\frac{f(x) - f(y)}{2}\Big)^2 = C^2 \mathcal{E}[f].$$

In particular, since $\Psi(a) = |a|$ is 1-Lipschitz we get

$$E[|f| L(|f|)] = \mathcal{E}[|f|] \leq \mathcal{E}[f] = E[f\, Lf].$$

Energy stability - general case

A similar phenomenon is observed for a larger class of **symmetric** Markov semigroups (again, up to technicalities).

Indeed, for all $a, b \in \mathbf{R}$ we have

$$(\Psi(a) - \Psi(b))^2 \leq C^2(a-b)^2.$$

For $t \geq 0$ and $x \in \Omega$ let us set $a = f(x)$ and $b = f(\mathbf{X}^x(t))$:

$$\Psi(f(x))^2 - 2\Psi(f(x))\Psi(f(\mathbf{X}^x(t))) + \Psi(f(\mathbf{X}^x(t)))^2 \leq$$

$$\leq C^2\Big(f(x)^2 - 2f(x)f(\mathbf{X}^x(t)) + f(\mathbf{X}^x(t))^2\Big).$$

By taking expectation (with respect to the Markov process $\mathbf{X}^x$) of both sides we obtain the following inequality:

$$\Psi(f)^2 - 2\Psi(f)Q_t(\Psi(f)) + Q_t(\Psi(f)^2) \leq C^2(f^2 - 2fQ_tf + Q_t(f^2)),$$

which holds pointwise (a.e. with respect to the invariant probability measure).

Energy stability - general case

A similar phenomenon is observed for a larger class of **symmetric** Markov semigroups (again, up to technicalities).
Indeed, for all $a, b \in \mathbf{R}$ we have

$$(\Psi(a) - \Psi(b))^2 \leq C^2(a-b)^2.$$

For $t \geq 0$ and $x \in \Omega$ let us set $a = f(x)$ and $b = f(\mathbf{X}^x(t))$:

$$\Psi(f(x))^2 - 2\Psi(f(x))\Psi(f(\mathbf{X}^x(t))) + \Psi(f(\mathbf{X}^x(t)))^2 \leq$$

$$\leq C^2\Big(f(x)^2 - 2f(x)f(\mathbf{X}^x(t)) + f(\mathbf{X}^x(t))^2\Big).$$

By taking expectation (with respect to the Markov process $\mathbf{X}^x$) of both sides we obtain the following inequality:

$$\Psi(f)^2 - 2\Psi(f)Q_t(\Psi(f)) + Q_t(\Psi(f)^2) \leq C^2(f^2 - 2fQ_tf + Q_t(f^2)),$$

which holds pointwise (a.e. with respect to the invariant probability measure).

Energy stability - general case

A similar phenomenon is observed for a larger class of **symmetric** Markov semigroups (again, up to technicalities).
Indeed, for all $a, b \in \mathbf{R}$ we have

$$(\Psi(a) - \Psi(b))^2 \leq C^2(a-b)^2.$$

For $t \geq 0$ and $x \in \Omega$ let us set $a = f(x)$ and $b = f(\mathbf{X}^x(t))$:

$$\Psi(f(x))^2 - 2\Psi(f(x))\Psi(f(\mathbf{X}^x(t))) + \Psi(f(\mathbf{X}^x(t)))^2 \leq$$

$$\leq C^2\Big(f(x)^2 - 2f(x)f(\mathbf{X}^x(t)) + f(\mathbf{X}^x(t))^2\Big).$$

By taking expectation (with respect to the Markov process $\mathbf{X}^x$) of both sides we obtain the following inequality:

$$\Psi(f)^2 - 2\Psi(f)Q_t(\Psi(f)) + Q_t(\Psi(f)^2) \leq C^2(f^2 - 2fQ_tf + Q_t(f^2)),$$

which holds pointwise (a.e. with respect to the invariant probability measure).

Energy stability - general case

A similar phenomenon is observed for a larger class of **symmetric** Markov semigroups (again, up to technicalities).
Indeed, for all $a, b \in \mathbf{R}$ we have

$$(\Psi(a) - \Psi(b))^2 \leq C^2(a-b)^2.$$

For $t \geq 0$ and $x \in \Omega$ let us set $a = f(x)$ and $b = f(\mathbf{X}^x(t))$:

$$\Psi(f(x))^2 - 2\Psi(f(x))\Psi(f(\mathbf{X}^x(t))) + \Psi(f(\mathbf{X}^x(t)))^2 \leq$$

$$\leq C^2\Big(f(x)^2 - 2f(x)f(\mathbf{X}^x(t)) + f(\mathbf{X}^x(t))^2\Big).$$

By taking expectation (with respect to the Markov process $\mathbf{X}^x$) of both sides we obtain the following inequality:

$$\Psi(f)^2 - 2\Psi(f)Q_t(\Psi(f)) + Q_t(\Psi(f)^2) \leq C^2(f^2 - 2fQ_tf + Q_t(f^2)),$$

which holds pointwise (a.e. with respect to the invariant probability measure).

Energy stability - end of the proof

Now we average over Ω (with respect to the invariant probability measure) and arrive at

$$0 \geq \alpha(t) = E[\Psi(f)^2] - 2E[\Psi(f)Q_t(\Psi(f))] + E[Q_t(\Psi(f)^2)] -$$

$$-C^2E[f^2] + 2C^2E[fQ_tf] - C^2E[Q_t(f^2)] =$$

$$= 2E[\Psi(f)^2] - 2E[\Psi(f)Q_t(\Psi(f))] - 2C^2E[f^2] + 2C^2E[fQ_tf],$$

where we have used the fact that symmetric Markov semigroups preserve expectation.

Since $Q_0 = Id$ we have $\alpha(0) = 0$. Thus $\alpha'(0^+) \leq 0$, i.e.

$$2E[\Psi(f)L(\Psi(f))] - 2C^2E[f\,Lf] \leq 0,$$

$$\mathcal{E}[\Psi(f)] \leq C^2\mathcal{E}[f].$$

Energy stability - end of the proof

Now we average over Ω (with respect to the invariant probability measure) and arrive at

$$0 \geq \alpha(t) = E[\Psi(f)^2] - 2E[\Psi(f)Q_t(\Psi(f))] + E[Q_t(\Psi(f)^2)] -$$

$$-C^2E[f^2] + 2C^2E[fQ_tf] - C^2E[Q_t(f^2)] =$$

$$= 2E[\Psi(f)^2] - 2E[\Psi(f)Q_t(\Psi(f))] - 2C^2E[f^2] + 2C^2E[fQ_tf],$$

where we have used the fact that symmetric Markov semigroups preserve expectation.

Since $Q_0 = Id$ we have $\alpha(0) = 0$. Thus $\alpha'(0^+) \leq 0$, i.e.

$$2E[\Psi(f)L(\Psi(f))] - 2C^2E[f\,Lf] \leq 0,$$

$$\mathcal{E}[\Psi(f)] \leq C^2\mathcal{E}[f].$$

Poincaré inequality

The classical Poincaré inequality comes from the partial differential equations area:

$$\int_D f^2 \le C_D \int_D |\nabla f|^2,$$

where $D \subset R^n$ is bounded, $f \in \mathcal{C}^1_c(D)$, and we integrate with respect to the Lebesgue measure.

We say that a probability Borel measure ν on $\mathbf{R}^n$ satisfies the Poincaré inequality with constant C if for every $\mathcal{C}^1$ function $f : \mathbf{R}^n \longrightarrow \mathbf{R}$ such that $\int_{\mathbf{R}^n} f\,d\nu < \infty$ there is

$$\int_{\mathbf{R}^n} f^2\,d\nu - \Big(\int_{\mathbf{R}^n} f\,d\nu\Big)^2 \le C \int_{\mathbf{R}^n} |\nabla f|^2\,d\nu.$$

On C_n the energy functional takes place of $\int |\nabla f|^2$.
We will prove that for every $f \in \mathcal{H}_n$

$$E[f^2] - (E[f])^2 \le E[f\,Lf].$$

Poincaré inequality

The classical Poincaré inequality comes from the partial differential equations area:

$$\int_D f^2 \leq C_D \int_D |\nabla f|^2,$$

where $D \subset R^n$ is bounded, $f \in \mathcal{C}_c^1(D)$, and we integrate with respect to the Lebesgue measure.

We say that a probability Borel measure ν on $\mathbf{R}^n$ satisfies the Poincaré inequality with constant C if for every $\mathcal{C}^1$ function $f : \mathbf{R}^n \longrightarrow \mathbf{R}$ such that $\int_{\mathbf{R}^n} f \, d\nu < \infty$ there is

$$\int_{\mathbf{R}^n} f^2 \, d\nu - \Big(\int_{\mathbf{R}^n} f \, d\nu \Big)^2 \leq C \int_{\mathbf{R}^n} |\nabla f|^2 \, d\nu.$$

On C_n the energy functional takes place of $\int |\nabla f|^2$.
We will prove that for every $f \in \mathcal{H}_n$

$$E[f^2] - (E[f])^2 \leq E[f \, Lf].$$

Poincaré inequality

The classical Poincaré inequality comes from the partial differential equations area:

$$\int_D f^2 \le C_D \int_D |\nabla f|^2,$$

where $D \subset R^n$ is bounded, $f \in \mathcal{C}^1_c(D)$, and we integrate with respect to the Lebesgue measure.

We say that a probability Borel measure ν on $\mathbf{R}^n$ satisfies the Poincaré inequality with constant C if for every $\mathcal{C}^1$ function $f : \mathbf{R}^n \longrightarrow \mathbf{R}$ such that $\int_{\mathbf{R}^n} f \, d\nu < \infty$ there is

$$\int_{\mathbf{R}^n} f^2 \, d\nu - \Big(\int_{\mathbf{R}^n} f \, d\nu \Big)^2 \le C \int_{\mathbf{R}^n} |\nabla f|^2 \, d\nu.$$

On C_n the energy functional takes place of $\int |\nabla f|^2$.
We will prove that for every $f \in \mathcal{H}_n$

$$E[f^2] - (E[f])^2 \le E[f \, Lf].$$

Poincaré inequality for even functions

Moreover, if f is even, i.e. $f(-x) = f(x)$ for all $x \in C_n$, then

$$E[f^2] - (E[f])^2 \leq \frac{1}{2} E[f\, Lf].$$

Indeed, let $f = \sum_{S \subseteq [n]} a_S w_S$. Recall that

$$E[f^2] = \sum_{S \subseteq [n]} a_S^2, \quad E[f] = a_\emptyset.$$

Thus

$$Var[f] = E[f^2] - (E[f])^2 = \sum_{S \subseteq [n]:\, S \neq \emptyset} a_S^2.$$

On the other hand

$$E[f\, Lf] = \sum_{S \subseteq [n]} |S| a_S^2 = \sum_{S \subseteq [n]:\, S \neq \emptyset} |S| a_S^2.$$

This ends the proof of the first assertion (and, by the way, it gives one more proof that L is positive semi-definite).

Poincaré inequality for even functions

Moreover, if f is even, i.e. $f(-x) = f(x)$ for all $x \in C_n$, then

$$E[f^2] - (E[f])^2 \leq \frac{1}{2} E[f\, Lf].$$

Indeed, let $f = \sum_{S \subseteq [n]} a_S w_S$. Recall that

$$E[f^2] = \sum_{S \subseteq [n]} a_S^2, \quad E[f] = a_\emptyset.$$

Thus

$$Var[f] = E[f^2] - (E[f])^2 = \sum_{S \subseteq [n]:\, S \neq \emptyset} a_S^2.$$

On the other hand

$$E[f\, Lf] = \sum_{S \subseteq [n]} |S| a_S^2 = \sum_{S \subseteq [n]:\, S \neq \emptyset} |S| a_S^2.$$

This ends the proof of the first assertion (and, by the way, it gives one more proof that L is positive semi-definite).

Poincaré inequality for even functions

Moreover, if f is even, i.e. $f(-x) = f(x)$ for all $x \in C_n$, then

$$E[f^2] - (E[f])^2 \leq \frac{1}{2} E[f\,Lf].$$

Indeed, let $f = \sum_{S \subseteq [n]} a_S w_S$. Recall that

$$E[f^2] = \sum_{S \subseteq [n]} a_S^2, \quad E[f] = a_\emptyset.$$

Thus

$$Var[f] = E[f^2] - (E[f])^2 = \sum_{S \subseteq [n]:\, S \neq \emptyset} a_S^2.$$

On the other hand

$$E[f\,Lf] = \sum_{S \subseteq [n]} |S| a_S^2 = \sum_{S \subseteq [n]:\, S \neq \emptyset} |S| a_S^2.$$

This ends the proof of the first assertion (and, by the way, it gives one more proof that L is positive semi-definite).

Poincaré inequality - proof of the second assertion

To prove the second assertion, note that if f is an even function then for all $S \subseteq [n]$ with $|S|$ odd we have

$$a_S = \langle f, w_S \rangle = E[f \cdot w_S] = 0.$$

Indeed, for $|S|$ odd, w_S is an odd function, so that $f \cdot w_S$ is odd as well and thus it has expectation zero.

Since all natural numbers strictly between 0 and 2 are odd, for every even $f \in \mathcal{H}_n$ we have

$$\frac{1}{2}E[f\,Lf] = \sum_{S\subseteq[n]:\,S\neq\emptyset} \frac{|S|}{2} a_S^2 = \sum_{S\subseteq[n]:\,|S|\geq 2} \frac{|S|}{2} a_S^2 \geq$$

$$\geq \sum_{S\subseteq[n]:\,|S|\geq 2} a_S^2 = \sum_{S\subseteq[n]:\,S\neq\emptyset} a_S^2 = E[f^2] - (E[f])^2.$$

Poincaré inequality - proof of the second assertion

To prove the second assertion, note that if f is an even function then for all $S \subseteq [n]$ with $|S|$ odd we have

$$a_S = \langle f, w_S \rangle = E[f \cdot w_S] = 0.$$

Indeed, for $|S|$ odd, w_S is an odd function, so that $f \cdot w_S$ is odd as well and thus it has expectation zero.

Since all natural numbers strictly between 0 and 2 are odd, for every even $f \in \mathcal{H}_n$ we have

$$\frac{1}{2}E[f\,Lf] = \sum_{S \subseteq [n]:\, S \neq \emptyset} \frac{|S|}{2} a_S^2 = \sum_{S \subseteq [n]:\, |S| \geq 2} \frac{|S|}{2} a_S^2 \geq$$

$$\geq \sum_{S \subseteq [n]:\, |S| \geq 2} a_S^2 = \sum_{S \subseteq [n]:\, S \neq \emptyset} a_S^2 = E[f^2] - (E[f])^2.$$

Spectral gap

The above **variance-energy inequalities** are also called **spectral gap inequalities** - as we have seen, they hold because there is a gap in the spectrum $\sigma(L)$ between eigenvalue 0, associated to the constant function $\mathbf{1}$, and $\sigma(L|_{f\in\mathcal{H}_n:\, E[f]=0})$.

For the proof of the Poincaré inequality for even functions we have used the existence of a gap between 0 and $\sigma(L|_{f\in\mathcal{H}_n:\, f \text{ even}, E[f]=0})$.

The existence of the spectral gap (of the first type) for a symmetric Markov semigroup $(Q_t)_{t\in[0,\infty)}$ implies

$$Q_t f \stackrel{t\to\infty}{\longrightarrow} E[f]$$

and the size of the gap is responsible for the speed of convergence. This is of uttermost importance in physics (and the Poincaré-type inequalities were considered in physics first, already in the middle of the nineteenth century).

Spectral gap

The above **variance-energy inequalities** are also called **spectral gap inequalities** - as we have seen, they hold because there is a gap in the spectrum $\sigma(L)$ between eigenvalue 0, associated to the constant function $\mathbf{1}$, and $\sigma(L|_{f\in\mathcal{H}_n:\, E[f]=0})$.

For the proof of the Poincaré inequality for even functions we have used the existence of a gap between 0 and $\sigma(L|_{f\in\mathcal{H}_n:\, f \text{ even}, E[f]=0})$.

The existence of the spectral gap (of the first type) for a symmetric Markov semigroup $(Q_t)_{t\in[0,\infty)}$ implies

$$Q_t f \overset{t\to\infty}{\longrightarrow} E[f]$$

and the size of the gap is responsible for the speed of convergence. This is of uttermost importance in physics (and the Poincaré-type inequalities were considered in physics first, already in the middle of the nineteenth century).

Spectral gap

The above **variance-energy inequalities** are also called **spectral gap inequalities** - as we have seen, they hold because there is a gap in the spectrum $\sigma(L)$ between eigenvalue 0, associated to the constant function $\mathbf{1}$, and $\sigma(L|_{f\in\mathcal{H}_n:\, E[f]=0})$.

For the proof of the Poincaré inequality for even functions we have used the existence of a gap between 0 and $\sigma(L|_{f\in\mathcal{H}_n:\, f \text{ even}, E[f]=0})$.

The existence of the spectral gap (of the first type) for a symmetric Markov semigroup $(Q_t)_{t\in[0,\infty)}$ implies

$$Q_t f \overset{t\to\infty}{\longrightarrow} E[f]$$

and the size of the gap is responsible for the speed of convergence. This is of uttermost importance in physics (and the Poincaré-type inequalities were considered in physics first, already in the middle of the nineteenth century).

Elementary inequality

For any $p > 1$ and $a, b \geq 0$ the following inequality holds:

$$(p-2)^2(a^p + b^p) - p^2(a^{p-1}b + ab^{p-1}) + 8(p-1)a^{p/2}b^{p/2} \geq 0.$$

Because of the homogeneity, it suffices to prove that for $t \geq 1$

$$u(t) = (p-2)^2 t^p - p^2 t^{p-1} + 8(p-1)t^{p/2} - p^2 t + (p-2)^2 \geq 0.$$

Indeed, $u(1) = 2(p^2 - 4p + 4) - 2p^2 + 8p - 8 = 0$, and

$$u'(t) = p(p-2)^2 t^{p-1} - p^2(p-1)t^{p-2} + 4p(p-1)t^{\frac{p}{2}-1} - p^2,$$

so that $u'(1) = (p^3 - 4p^2 + 4p) - (p^3 - p^2) + (4p^2 - 4p) - p^2 = 0$.
Now it suffices to note that

$$u''(t) = p(p-1)(p-2)^2 t^{p-2} - p^2(p-1)(p-2)t^{p-3} + 2p(p-1)(p-2)t^{\frac{p}{2}-2}$$

$$= p^2(p-1)(p-2)t^{p-2}\left(\frac{p-2}{p} + \frac{2}{p}t^{-p/2} - t^{-1}\right) =$$

$$= 2p(p-1)(2-p)t^{p-2}\left(\frac{2-p}{2} + \frac{p}{2}t^{-1} - t^{-p/2}\right).$$

Elementary inequality

For any $p > 1$ and $a, b \geq 0$ the following inequality holds:

$$(p-2)^2(a^p + b^p) - p^2(a^{p-1}b + ab^{p-1}) + 8(p-1)a^{p/2}b^{p/2} \geq 0.$$

Because of the homogeneity, it suffices to prove that for $t \geq 1$

$$u(t) = (p-2)^2 t^p - p^2 t^{p-1} + 8(p-1)t^{p/2} - p^2 t + (p-2)^2 \geq 0.$$

Indeed, $u(1) = 2(p^2 - 4p + 4) - 2p^2 + 8p - 8 = 0$, and

$$u'(t) = p(p-2)^2 t^{p-1} - p^2(p-1)t^{p-2} + 4p(p-1)t^{\frac{p}{2}-1} - p^2,$$

so that $u'(1) = (p^3 - 4p^2 + 4p) - (p^3 - p^2) + (4p^2 - 4p) - p^2 = 0.$
Now it suffices to note that

$$u''(t) = p(p-1)(p-2)^2 t^{p-2} - p^2(p-1)(p-2)t^{p-3} + 2p(p-1)(p-2)t^{\frac{p}{2}-2}$$

$$= p^2(p-1)(p-2)t^{p-2}\left(\frac{p-2}{p} + \frac{2}{p}t^{-p/2} - t^{-1}\right) =$$

$$= 2p(p-1)(2-p)t^{p-2}\left(\frac{2-p}{2} + \frac{p}{2}t^{-1} - t^{-p/2}\right).$$

Elementary inequality

For any $p > 1$ and $a, b \geq 0$ the following inequality holds:

$$(p-2)^2(a^p + b^p) - p^2(a^{p-1}b + ab^{p-1}) + 8(p-1)a^{p/2}b^{p/2} \geq 0.$$

Because of the homogeneity, it suffices to prove that for $t \geq 1$

$$u(t) = (p-2)^2 t^p - p^2 t^{p-1} + 8(p-1)t^{p/2} - p^2 t + (p-2)^2 \geq 0.$$

Indeed, $u(1) = 2(p^2 - 4p + 4) - 2p^2 + 8p - 8 = 0$, and

$$u'(t) = p(p-2)^2 t^{p-1} - p^2(p-1)t^{p-2} + 4p(p-1)t^{\frac{p}{2}-1} - p^2,$$

so that $u'(1) = (p^3 - 4p^2 + 4p) - (p^3 - p^2) + (4p^2 - 4p) - p^2 = 0$.

Now it suffices to note that

$$u''(t) = p(p-1)(p-2)^2 t^{p-2} - p^2(p-1)(p-2)t^{p-3} + 2p(p-1)(p-2)t^{\frac{p}{2}-2}$$

$$= p^2(p-1)(p-2)t^{p-2}\left(\frac{p-2}{p} + \frac{2}{p}t^{-p/2} - t^{-1}\right) =$$

$$= 2p(p-1)(2-p)t^{p-2}\left(\frac{2-p}{2} + \frac{p}{2}t^{-1} - t^{-p/2}\right).$$

Elementary inequality

For any $p > 1$ and $a, b \geq 0$ the following inequality holds:

$$(p-2)^2(a^p + b^p) - p^2(a^{p-1}b + ab^{p-1}) + 8(p-1)a^{p/2}b^{p/2} \geq 0.$$

Because of the homogeneity, it suffices to prove that for $t \geq 1$

$$u(t) = (p-2)^2 t^p - p^2 t^{p-1} + 8(p-1)t^{p/2} - p^2 t + (p-2)^2 \geq 0.$$

Indeed, $u(1) = 2(p^2 - 4p + 4) - 2p^2 + 8p - 8 = 0$, and

$$u'(t) = p(p-2)^2 t^{p-1} - p^2(p-1)t^{p-2} + 4p(p-1)t^{\frac{p}{2}-1} - p^2,$$

so that $u'(1) = (p^3 - 4p^2 + 4p) - (p^3 - p^2) + (4p^2 - 4p) - p^2 = 0.$
Now it suffices to note that

$$u''(t) = p(p-1)(p-2)^2 t^{p-2} - p^2(p-1)(p-2)t^{p-3} + 2p(p-1)(p-2)t^{\frac{p}{2}-2}$$

$$= p^2(p-1)(p-2)t^{p-2}\left(\frac{p-2}{p} + \frac{2}{p}t^{-p/2} - t^{-1}\right) =$$

$$= 2p(p-1)(2-p)t^{p-2}\left(\frac{2-p}{2} + \frac{p}{2}t^{-1} - t^{-p/2}\right).$$

Elementary inequality - end of the proof

Recall:

$$u''(t) = p^2(p-1)(p-2)t^{p-2}\Big(\frac{p-2}{p} + \frac{2}{p}t^{-p/2} - t^{-1}\Big) =$$

$$= 2p(p-1)(2-p)t^{p-2}\Big(\frac{2-p}{2} + \frac{p}{2}t^{-1} - t^{-p/2}\Big).$$

Since for $p \geq 2$ there is

$$\frac{p-2}{p} + \frac{2}{p}t^{-p/2} = \frac{p-2}{p}\cdot 1 + \frac{2}{p}\cdot t^{-p/2} \geq 1^{\frac{p-2}{p}}(t^{-p/2})^{\frac{2}{p}} = t^{-1},$$

while for $p \in (1,2]$ there is

$$\frac{2-p}{2} + \frac{p}{2}t^{-1} = \frac{2-p}{2}\cdot 1 + \frac{p}{2}\cdot t^{-1} \geq 1^{\frac{p-2}{2}}(t^{-1})^{\frac{p}{2}} = t^{-p/2},$$

we have $u''(t) \geq 0$ and the proof is finished.

For $p \in (0,1)$ and $a, b > 0$ the reverse inequality holds.

Elementary inequality - end of the proof

Recall:

$$u''(t) = p^2(p-1)(p-2)t^{p-2}\Big(\frac{p-2}{p} + \frac{2}{p}t^{-p/2} - t^{-1}\Big) =$$

$$= 2p(p-1)(2-p)t^{p-2}\Big(\frac{2-p}{2} + \frac{p}{2}t^{-1} - t^{-p/2}\Big).$$

Since for $p \geq 2$ there is

$$\frac{p-2}{p} + \frac{2}{p}t^{-p/2} = \frac{p-2}{p}\cdot 1 + \frac{2}{p}\cdot t^{-p/2} \geq 1^{\frac{p-2}{p}}(t^{-p/2})^{\frac{2}{p}} = t^{-1},$$

while for $p \in (1,2]$ there is

$$\frac{2-p}{2} + \frac{p}{2}t^{-1} = \frac{2-p}{2}\cdot 1 + \frac{p}{2}\cdot t^{-1} \geq 1^{\frac{p-2}{2}}(t^{-1})^{\frac{p}{2}} = t^{-p/2},$$

we have $u''(t) \geq 0$ and the proof is finished.

For $p \in (0,1)$ and $a, b > 0$ the reverse inequality holds.

Elementary inequality - end of the proof

Recall:

$$u''(t) = p^2(p-1)(p-2)t^{p-2}\Big(\frac{p-2}{p} + \frac{2}{p}t^{-p/2} - t^{-1}\Big) =$$

$$= 2p(p-1)(2-p)t^{p-2}\Big(\frac{2-p}{2} + \frac{p}{2}t^{-1} - t^{-p/2}\Big).$$

Since for $p \geq 2$ there is

$$\frac{p-2}{p} + \frac{2}{p}t^{-p/2} = \frac{p-2}{p} \cdot 1 + \frac{2}{p} \cdot t^{-p/2} \geq 1^{\frac{p-2}{p}}(t^{-p/2})^{\frac{2}{p}} = t^{-1},$$

while for $p \in (1,2]$ there is

$$\frac{2-p}{2} + \frac{p}{2}t^{-1} = \frac{2-p}{2} \cdot 1 + \frac{p}{2} \cdot t^{-1} \geq 1^{\frac{p-2}{2}}(t^{-1})^{\frac{p}{2}} = t^{-p/2},$$

we have $u''(t) \geq 0$ and the proof is finished.

For $p \in (0,1)$ and $a, b > 0$ the reverse inequality holds.

Elementary inequality - end of the proof

Recall:

$$u''(t) = p^2(p-1)(p-2)t^{p-2}\Big(\frac{p-2}{p} + \frac{2}{p}t^{-p/2} - t^{-1}\Big) =$$

$$= 2p(p-1)(2-p)t^{p-2}\Big(\frac{2-p}{2} + \frac{p}{2}t^{-1} - t^{-p/2}\Big).$$

Since for $p \geq 2$ there is

$$\frac{p-2}{p} + \frac{2}{p}t^{-p/2} = \frac{p-2}{p}\cdot 1 + \frac{2}{p}\cdot t^{-p/2} \geq 1^{\frac{p-2}{p}}(t^{-p/2})^{\frac{2}{p}} = t^{-1},$$

while for $p \in (1,2]$ there is

$$\frac{2-p}{2} + \frac{p}{2}t^{-1} = \frac{2-p}{2}\cdot 1 + \frac{p}{2}\cdot t^{-1} \geq 1^{\frac{p-2}{2}}(t^{-1})^{\frac{p}{2}} = t^{-p/2},$$

we have $u''(t) \geq 0$ and the proof is finished.

For $p \in (0,1)$ and $a, b > 0$ the reverse inequality holds.

Stroock-Varopoulos inequality (1984/85)

For any $p > 1$ and $f : C_n \longrightarrow [0, \infty)$ there is

$$E[f^{p/2}L(f^{p/2})] \leq \frac{p^2}{4(p-1)} E[f^{p-1}Lf].$$

The same inequality applies to any generator of a **symmetric** Markov semigroup (under some technical assumptions about f), with a proof similar to the one below.

Recall that for $a, b \geq 0$ there is

$$(p-2)^2(a^p + b^p) - p^2(a^{p-1}b + ab^{p-1}) + 8(p-1)a^{p/2}b^{p/2} \geq 0.$$

Hence, for any $a \geq 0$ we have

$$(p-2)^2(a^p + f^p) - p^2(a^{p-1}f + af^{p-1}) + 8(p-1)a^{p/2}f^{p/2} \geq 0$$

pointwise. P_t is linear and order preserving for any $t \geq 0$, so that

$$(p-2)^2(a^p + P_t(f^p)) - p^2(a^{p-1}P_t f + aP_t(f^{p-1})) + 8(p-1)a^{p/2}P_t(f^{p/2}) \geq 0.$$

Stroock-Varopoulos inequality (1984/85)

For any $p > 1$ and $f : C_n \longrightarrow [0, \infty)$ there is

$$E[f^{p/2}L(f^{p/2})] \leq \frac{p^2}{4(p-1)}E[f^{p-1}Lf].$$

The same inequality applies to any generator of a **symmetric** Markov semigroup (under some technical assumptions about f), with a proof similar to the one below.

Recall that for $a, b \geq 0$ there is

$$(p-2)^2(a^p + b^p) - p^2(a^{p-1}b + ab^{p-1}) + 8(p-1)a^{p/2}b^{p/2} \geq 0.$$

Hence, for any $a \geq 0$ we have

$$(p-2)^2(a^p + f^p) - p^2(a^{p-1}f + af^{p-1}) + 8(p-1)a^{p/2}f^{p/2} \geq 0$$

pointwise. P_t is linear and order preserving for any $t \geq 0$, so that

$$(p-2)^2(a^p + P_t(f^p)) - p^2(a^{p-1}P_tf + aP_t(f^{p-1})) + 8(p-1)a^{p/2}P_t(f^{p/2}) \geq 0.$$

Stroock-Varopoulos inequality (1984/85)

For any $p > 1$ and $f : C_n \longrightarrow [0, \infty)$ there is

$$E[f^{p/2}L(f^{p/2})] \leq \frac{p^2}{4(p-1)}E[f^{p-1}Lf].$$

The same inequality applies to any generator of a **symmetric** Markov semigroup (under some technical assumptions about f), with a proof similar to the one below.

Recall that for $a, b \geq 0$ there is

$$(p-2)^2(a^p + b^p) - p^2(a^{p-1}b + ab^{p-1}) + 8(p-1)a^{p/2}b^{p/2} \geq 0.$$

Hence, for any $a \geq 0$ we have

$$(p-2)^2(a^p + f^p) - p^2(a^{p-1}f + af^{p-1}) + 8(p-1)a^{p/2}f^{p/2} \geq 0$$

pointwise. P_t is linear and order preserving for any $t \geq 0$, so that

$$(p-2)^2(a^p+P_t(f^p))-p^2(a^{p-1}P_tf+aP_t(f^{p-1}))+8(p-1)a^{p/2}P_t(f^{p/2}) \geq 0.$$

Stroock-Varopoulos inequality (1984/85)

For any $p > 1$ and $f : C_n \longrightarrow [0, \infty)$ there is

$$E[f^{p/2}L(f^{p/2})] \leq \frac{p^2}{4(p-1)}E[f^{p-1}Lf].$$

The same inequality applies to any generator of a **symmetric** Markov semigroup (under some technical assumptions about f), with a proof similar to the one below.

Recall that for $a, b \geq 0$ there is

$$(p-2)^2(a^p + b^p) - p^2(a^{p-1}b + ab^{p-1}) + 8(p-1)a^{p/2}b^{p/2} \geq 0.$$

Hence, for any $a \geq 0$ we have

$$(p-2)^2(a^p + f^p) - p^2(a^{p-1}f + af^{p-1}) + 8(p-1)a^{p/2}f^{p/2} \geq 0$$

pointwise. P_t is linear and order preserving for any $t \geq 0$, so that

$$(p-2)^2(a^p + P_t(f^p)) - p^2(a^{p-1}P_tf + aP_t(f^{p-1})) + 8(p-1)a^{p/2}P_t(f^{p/2}) \geq 0.$$

Stroock-Varopoulos inequality - continued

$$(p-2)^2(a^p+P_t(f^p))-p^2(a^{p-1}P_tf+aP_t(f^{p-1}))+8(p-1)a^{p/2}P_t(f^{p/2})\geq 0$$

pointwise, for every $a \geq 0$. Thus also

$$(p-2)^2(f^p+P_t(f^p))-p^2(f^{p-1}P_tf+f\,P_t(f^{p-1}))+8(p-1)f^{p/2}P_t(f^{p/2})\geq 0$$

pointwise. By taking expectation we arrive at

$$(p-2)^2(E[f^p]+E[P_t(f^p)])-p^2(E[f^{p-1}P_tf]+E[f\,P_t(f^{p-1})])+$$
$$+8(p-1)E[f^{p/2}P_t(f^{p/2})]\geq 0.$$

Since P_t is symmetric and expectation preserving, we have

$$\beta(t)=2(p-2)^2E[f^p]-2p^2E[f^{p-1}P_tf]+8(p-1)E[f^{p/2}P_t(f^{p/2})]\geq 0$$

for $t \geq 0$. Since $P_0 = Id$ we have

$$\beta(0)=\Big(2(p-2)^2-2p^2+8(p-1)\Big)\cdot E[f^p]=0.$$

Stroock-Varopoulos inequality - continued

$$(p-2)^2(a^p+P_t(f^p))-p^2(a^{p-1}P_tf+aP_t(f^{p-1}))+8(p-1)a^{p/2}P_t(f^{p/2}) \geq 0$$

pointwise, for every $a \geq 0$. Thus also

$$(p-2)^2(f^p+P_t(f^p))-p^2(f^{p-1}P_tf+f\,P_t(f^{p-1}))+8(p-1)f^{p/2}P_t(f^{p/2}) \geq 0$$

pointwise. By taking expectation we arrive at

$$(p-2)^2(E[f^p]+E[P_t(f^p)])-p^2(E[f^{p-1}P_tf]+E[f\,P_t(f^{p-1})])+$$
$$+8(p-1)E[f^{p/2}P_t(f^{p/2})] \geq 0.$$

Since P_t is symmetric and expectation preserving, we have

$$\beta(t)=2(p-2)^2E[f^p]-2p^2E[f^{p-1}P_tf]+8(p-1)E[f^{p/2}P_t(f^{p/2})] \geq 0$$

for $t \geq 0$. Since $P_0 = Id$ we have

$$\beta(0)=\Big(2(p-2)^2-2p^2+8(p-1)\Big)\cdot E[f^p]=0.$$

Stroock-Varopoulos inequality - continued

$$(p-2)^2(a^p+P_t(f^p))-p^2(a^{p-1}P_tf+aP_t(f^{p-1}))+8(p-1)a^{p/2}P_t(f^{p/2}) \geq 0$$

pointwise, for every $a \geq 0$. Thus also

$$(p-2)^2(f^p+P_t(f^p))-p^2(f^{p-1}P_tf+f\,P_t(f^{p-1}))+8(p-1)f^{p/2}P_t(f^{p/2}) \geq 0$$

pointwise. By taking expectation we arrive at

$$(p-2)^2(E[f^p]+E[P_t(f^p)])-p^2(E[f^{p-1}P_tf]+E[f\,P_t(f^{p-1})])+$$
$$+8(p-1)E[f^{p/2}P_t(f^{p/2})] \geq 0.$$

Since P_t is symmetric and expectation preserving, we have

$$\beta(t)=2(p-2)^2E[f^p]-2p^2E[f^{p-1}P_tf]+8(p-1)E[f^{p/2}P_t(f^{p/2})] \geq 0$$

for $t \geq 0$. Since $P_0 = Id$ we have

$$\beta(0)=\Big(2(p-2)^2-2p^2+8(p-1)\Big)\cdot E[f^p]=0.$$

$$(p-2)^2(a^p+P_t(f^p))-p^2(a^{p-1}P_tf+aP_t(f^{p-1}))+8(p-1)a^{p/2}P_t(f^{p/2}) \geq 0$$

pointwise, for every $a \geq 0$. Thus also

$$(p-2)^2(f^p+P_t(f^p))-p^2(f^{p-1}P_tf+f\,P_t(f^{p-1}))+8(p-1)f^{p/2}P_t(f^{p/2}) \geq 0$$

pointwise. By taking expectation we arrive at

$$(p-2)^2(E[f^p]+E[P_t(f^p)])-p^2(E[f^{p-1}P_tf]+E[f\,P_t(f^{p-1})])+ \\ +8(p-1)E[f^{p/2}P_t(f^{p/2})] \geq 0.$$

Since P_t is symmetric and expectation preserving, we have

$$\beta(t)=2(p-2)^2E[f^p]-2p^2E[f^{p-1}P_tf]+8(p-1)E[f^{p/2}P_t(f^{p/2})] \geq 0$$

for $t \geq 0$. Since $P_0 = Id$ we have

$$\beta(0)=\Big(2(p-2)^2-2p^2+8(p-1)\Big)\cdot E[f^p]=0.$$

Stroock-Varopoulos inequality - continued

$$(p-2)^2(a^p+P_t(f^p))-p^2(a^{p-1}P_tf+aP_t(f^{p-1}))+8(p-1)a^{p/2}P_t(f^{p/2})\geq 0$$

pointwise, for every $a\geq 0$. Thus also

$$(p-2)^2(f^p+P_t(f^p))-p^2(f^{p-1}P_tf+f\,P_t(f^{p-1}))+8(p-1)f^{p/2}P_t(f^{p/2})\geq 0$$

pointwise. By taking expectation we arrive at

$$(p-2)^2(E[f^p]+E[P_t(f^p)])-p^2(E[f^{p-1}P_tf]+E[f\,P_t(f^{p-1})])+$$
$$+8(p-1)E[f^{p/2}P_t(f^{p/2})]\geq 0.$$

Since P_t is symmetric and expectation preserving, we have

$$\beta(t)=2(p-2)^2E[f^p]-2p^2E[f^{p-1}P_tf]+8(p-1)E[f^{p/2}P_t(f^{p/2})]\geq 0$$

for $t\geq 0$. Since $P_0=Id$ we have

$$\beta(0)=\Big(2(p-2)^2-2p^2+8(p-1)\Big)\cdot E[f^p]=0.$$

Stroock-Varopoulos inequality - end of the proof

Thus $\beta'(0^+) \geq 0$. On the other hand,

$$\beta'(0^+) = 2p^2 E[f^{p-1} Lf] - 8(p-1) E[f^{p/2} L(f^{p/2})],$$

so that

$$E[f^{p/2} L(f^{p/2})] \leq \frac{p^2}{4(p-1)} E[f^{p-1} Lf]$$

and the proof is finished.

Remark: In the case of the Ornstein-Uhlenbeck semigroup on $(\mathbf{R}^n, (2\pi)^{-n/2} e^{-|x|^2/2}\, dx)$ there is

$$(Lf)(x) = \langle x, \nabla f(x) \rangle - (\Delta f)(x),$$

$$E[f \cdot Lg] = (2\pi)^{-n/2} \int_{\mathbf{R}^n} \langle (\nabla f)(x), (\nabla g)(x) \rangle e^{-|x|^2/2}\, dx = E[\langle \nabla f, \nabla g \rangle]$$

(at least for $f, g \in \mathcal{C}_c^\infty$; strictly speaking, one must extend L from this dense subspace to a self-adjoint operator). It is easy to see that always there is equality in the Stroock-Varopoulos inequality in this case, at least if f is positive and smooth.

Stroock-Varopoulos inequality - end of the proof

Thus $\beta'(0^+) \geq 0$. On the other hand,

$$\beta'(0^+) = 2p^2 E[f^{p-1} Lf] - 8(p-1)E[f^{p/2}L(f^{p/2})],$$

so that

$$E[f^{p/2}L(f^{p/2})] \leq \frac{p^2}{4(p-1)} E[f^{p-1}Lf]$$

and the proof is finished.

Remark: In the case of the Ornstein-Uhlenbeck semigroup on $(\mathbf{R}^n, (2\pi)^{-n/2}e^{-|x|^2/2}\,dx)$ there is

$$(Lf)(x) = \langle x, \nabla f(x)\rangle - (\Delta f)(x),$$

$$E[f \cdot Lg] = (2\pi)^{-n/2} \int_{\mathbf{R}^n} \langle (\nabla f)(x), (\nabla g)(x)\rangle e^{-|x|^2/2}\,dx = E[\langle \nabla f, \nabla g\rangle]$$

(at least for $f, g \in \mathcal{C}_c^\infty$; strictly speaking, one must extend L from this dense subspace to a self-adjoint operator). It is easy to see that always there is equality in the Stroock-Varopoulos inequality in this case, at least if f is positive and smooth.

Khinchine-Kahane inequality

In his studies on the law of the iterated logarithm, A. Khinchine discovered that for every $p > q > 0$ there exists a positive constant $C_{p,q}$ such that for any natural n and arbitrary real numbers $a_1, a_2, \ldots, a_n$ the inequality

$$(E[|\sum_{i=1}^{n} a_i r_i|^p])^{1/p} \leq C_{p,q} \cdot (E[|\sum_{i=1}^{n} a_i r_i|^q])^{1/q}$$

holds, where $r_1, r_2, \ldots$ are independent symmetric ± 1 random variables.

J.-P. Kahane extended this result. He proved that for every $p > q > 0$ there exists a positive constant $K_{p,q}$ such that for any natural n, any normed linear space F and any collection of vectors $v_1, v_2, \ldots, v_n \in F$ there is

$$(E[\|\sum_{i=1}^{n} r_i v_i\|^p])^{1/p} \leq K_{p,q} \cdot (E[\|\sum_{i=1}^{n} r_i v_i\|^q])^{1/q}.$$

Khinchine-Kahane inequality

In his studies on the law of the iterated logarithm, A. Khinchine discovered that for every $p > q > 0$ there exists a positive constant $C_{p,q}$ such that for any natural n and arbitrary real numbers $a_1, a_2, \ldots, a_n$ the inequality

$$(E[|\sum_{i=1}^{n} a_i r_i|^p])^{1/p} \leq C_{p,q} \cdot (E[|\sum_{i=1}^{n} a_i r_i|^q])^{1/q}$$

holds, where $r_1, r_2, \ldots$ are independent symmetric ± 1 random variables.

J.-P. Kahane extended this result. He proved that for every $p > q > 0$ there exists a positive constant $K_{p,q}$ such that for any natural n, any normed linear space F and any collection of vectors $v_1,\ v_2, \ldots,\ v_n \in F$ there is

$$(E[\|\sum_{i=1}^{n} r_i v_i\|^p])^{1/p} \leq K_{p,q} \cdot (E[\|\sum_{i=1}^{n} r_i v_i\|^q])^{1/q}.$$

Kahane inequality - optimal constants

The optimal (least possible) constants in the Khinchine inequality were established for a large range of parameters p and q (Whittle, Szarek and others). U. Haagerup found the optimal $C_{p,2}$ for $p > 2$ and $C_{2,q}$ for $q \in (0, 2)$.

We will prove that the Kahane inequality holds with $K_{2,1} = \sqrt{2}$ and $K_{4,2} = \sqrt[4]{3}$ (R. Latała, S. Kwapień).

Both constants are optimal even for the Khinchine inequality (obviously, R is a special case of a normed linear space):

$$(E[|r_1 + r_2|^2])^{1/2}/E[|r_1 + r_2|] = \sqrt{2}$$

and in the $(4, 2)$ case there is asymptotic equality for $a_1 = \ldots = a_n = n^{-1/2}$ as $n \to \infty$ - by the CLT it is enough to check that

$$(E[G^4])^{1/4}/(E[G^2])^{1/2} = \sqrt[4]{3},$$

where $G \sim \mathcal{N}(0, 1)$.

Kahane inequality - optimal constants

The optimal (least possible) constants in the Khinchine inequality were established for a large range of parameters p and q (Whittle, Szarek and others). U. Haagerup found the optimal $C_{p,2}$ for $p > 2$ and $C_{2,q}$ for $q \in (0,2)$.

We will prove that the Kahane inequality holds with $K_{2,1} = \sqrt{2}$ and $K_{4,2} = \sqrt[4]{3}$ (R. Latała, S. Kwapień).

Both constants are optimal even for the Khinchine inequality (obviously, R is a special case of a normed linear space):

$$(E[|r_1 + r_2|^2])^{1/2} / E[|r_1 + r_2|] = \sqrt{2}$$

and in the $(4,2)$ case there is asymptotic equality for $a_1 = \ldots = a_n = n^{-1/2}$ as $n \to \infty$ - by the CLT it is enough to check that

$$(E[G^4])^{1/4} / (E[G^2])^{1/2} = \sqrt[4]{3},$$

where $G \sim \mathcal{N}(0,1)$.

Kahane inequality - optimal constants

The optimal (least possible) constants in the Khinchine inequality were established for a large range of parameters p and q (Whittle, Szarek and others). U. Haagerup found the optimal $C_{p,2}$ for $p > 2$ and $C_{2,q}$ for $q \in (0,2)$.

We will prove that the Kahane inequality holds with $K_{2,1} = \sqrt{2}$ and $K_{4,2} = \sqrt[4]{3}$ (R. Latała, S. Kwapień).

Both constants are optimal even for the Khinchine inequality (obviously, **R** is a special case of a normed linear space):

$$(E[|r_1 + r_2|^2])^{1/2} / E[|r_1 + r_2|] = \sqrt{2}$$

and in the $(4,2)$ case there is asymptotic equality for $a_1 = \ldots = a_n = n^{-1/2}$ as $n \to \infty$ - by the CLT it is enough to check that

$$(E[G^4])^{1/4} / (E[G^2])^{1/2} = \sqrt[4]{3},$$

where $G \sim \mathcal{N}(0,1)$.

Kahane inequality - optimal constants

The optimal (least possible) constants in the Khinchine inequality were established for a large range of parameters p and q (Whittle, Szarek and others). U. Haagerup found the optimal $C_{p,2}$ for $p > 2$ and $C_{2,q}$ for $q \in (0,2)$.

We will prove that the Kahane inequality holds with $K_{2,1} = \sqrt{2}$ and $K_{4,2} = \sqrt[4]{3}$ (R. Latała, S. Kwapień).

Both constants are optimal even for the Khinchine inequality (obviously, **R** is a special case of a normed linear space):

$$(E[|r_1 + r_2|^2])^{1/2} / E[|r_1 + r_2|] = \sqrt{2}$$

and in the $(4,2)$ case there is asymptotic equality for $a_1 = \ldots = a_n = n^{-1/2}$ as $n \to \infty$ - by the CLT it is enough to check that

$$(E[G^4])^{1/4} / (E[G^2])^{1/2} = \sqrt[4]{3},$$

where $G \sim \mathcal{N}(0,1)$.

Seminorm

For $n \geq 2$ let $v_1,\ v_2, \ldots,\ v_n$ be vectors of some linear space and let $\|\cdot\|$ be a seminorm on this space. For $x = (x_1, x_2, \ldots, x_n) \in \mathbf{R}^n$ let

$$H(x) = \|\sum_{i=1}^{n} x_i v_i\|.$$

Obviously, H is a seminorm on $\mathbf{R}^n$ and $h = H\Big|_{C_n}$ has the following properties:

- $h \geq 0$ (pointwise),
- h is even, i.e. $h(-x) = h(x)$ for all $x \in C_n$.

Now we will also prove that

- $Lh \leq h$ (pointwise).

Seminorm

For $n \geq 2$ let $v_1,\ v_2, \dots,\ v_n$ be vectors of some linear space and let $\|\cdot\|$ be a seminorm on this space. For $x = (x_1, x_2, \dots, x_n) \in \mathbf{R}^n$ let

$$H(x) = \|\sum_{i=1}^{n} x_i v_i\|.$$

Obviously, H is a seminorm on $\mathbf{R}^n$ and $h = H\Big|_{C_n}$ has the following properties:

- $h \geq 0$ (pointwise),
- h is even, i.e. $h(-x) = h(x)$ for all $x \in C_n$.

Now we will also prove that

- $Lh \leq h$ (pointwise).

Seminorm

For $n \geq 2$ let $v_1, v_2, \ldots, v_n$ be vectors of some linear space and let $\|\cdot\|$ be a seminorm on this space. For $x = (x_1, x_2, \ldots, x_n) \in \mathbf{R}^n$ let

$$H(x) = \|\sum_{i=1}^{n} x_i v_i\|.$$

Obviously, H is a seminorm on $\mathbf{R}^n$ and $h = H\Big|_{C_n}$ has the following properties:

- $h \geq 0$ (pointwise),
- h is even, i.e. $h(-x) = h(x)$ for all $x \in C_n$.

Now we will also prove that

- $Lh \leq h$ (pointwise).

Seminorm

For $n \geq 2$ let v_1, $v_2, \ldots,$ v_n be vectors of some linear space and let $\|\cdot\|$ be a seminorm on this space. For $x = (x_1, x_2, \ldots, x_n) \in \mathbf{R}^n$ let

$$H(x) = \|\sum_{i=1}^{n} x_i v_i\|.$$

Obviously, H is a seminorm on $\mathbf{R}^n$ and $h = H\Big|_{C_n}$ has the following properties:

- $h \geq 0$ (pointwise),
- h is even, i.e. $h(-x) = h(x)$ for all $x \in C_n$.

Now we will also prove that

- $Lh \leq h$ (pointwise).

Seminorm bound - first proof

Indeed, by the triangle inequality

$$(Lh)(x) = \frac{1}{2} \cdot \sum_{y \in C_n\colon d(x,y)=1} \Big(h(x) - h(y) \Big) = \frac{1}{2} \cdot \sum_{y \in C_n\colon d(x,y)=1} \Big(H(x) - H(y) \Big)$$

$$\leq \frac{n}{2} H(x) - \frac{1}{2} H\Big(\sum_{y \in C_n\colon d(x,y)=1} y \Big) = \frac{n}{2} H(x) - \frac{1}{2} H((n-2)x) =$$

$$= \frac{n}{2} H(x) - \frac{n-2}{2} H(x) = H(x) = h(x).$$

To understand why

$$\sum_{y \in C_n\colon d(x,y)=1} y = (n-2)x$$

(we add elements of C_n in $\mathbf{R}^n$) note that for each $i \in [n]$ exactly $n-1$ of the y's have the same i-th coordinate as x and exactly one of the y's has the same i-th coordinate as $-x$.

Seminorm bound - first proof

Indeed, by the triangle inequality

$$(Lh)(x) = \frac{1}{2} \cdot \sum_{y \in C_n :\, d(x,y)=1} \Big(h(x) - h(y) \Big) = \frac{1}{2} \cdot \sum_{y \in C_n :\, d(x,y)=1} \Big(H(x) - H(y) \Big)$$

$$\leq \frac{n}{2} H(x) - \frac{1}{2} H\Big(\sum_{y \in C_n :\, d(x,y)=1} y \Big) = \frac{n}{2} H(x) - \frac{1}{2} H((n-2)x) =$$

$$= \frac{n}{2} H(x) - \frac{n-2}{2} H(x) = H(x) = h(x).$$

To understand why

$$\sum_{y \in C_n :\, d(x,y)=1} y = (n-2)x$$

(we add elements of C_n in $\mathbf{R}^n$) note that for each $i \in [n]$ exactly $n-1$ of the y's have the same i-th coordinate as x and exactly one of the y's has the same i-th coordinate as $-x$.

Seminorm bound - second proof

Now let us follow a different reasoning. By the Hahn-Banach theorem the seminorm H can be expressed as a pointwise supremum of some family of linear functionals Φ:

$$\forall_{x \in \mathbf{R}^n} \quad H(x) = \sup_{\varphi \in \Phi} \varphi(x).$$

Obviously, each of these linear functionals, when restricted to C_n, is a linear combination of Rademacher functions and therefore $P_t \varphi = e^{-t}\varphi$ for $t \geq 0$. Since $H \geq \varphi$ pointwise and $(P_t)_{t \in [0,\infty)}$ is order preserving, we have

$$\forall_{\varphi \in \Phi} \quad P_t h \geq P_t \varphi = e^{-t}\varphi$$

and thus

$$P_t h \geq e^{-t} \sup_{\varphi \in \Phi} \varphi = e^{-t} h.$$

Hence

$$Lh = -\frac{d}{dt} P_t h \Big|_{t=0^+} = \lim_{t \to 0^+} \frac{P_0 h - P_t h}{t} \leq \lim_{t \to 0^+} \frac{h - e^{-t} h}{t} = h.$$

Seminorm bound - second proof

Now let us follow a different reasoning. By the Hahn-Banach theorem the seminorm H can be expressed as a pointwise supremum of some family of linear functionals Φ:

$$\forall_{x \in \mathbf{R}^n} \quad H(x) = \sup_{\varphi \in \Phi} \varphi(x).$$

Obviously, each of these linear functionals, when restricted to C_n, is a linear combination of Rademacher functions and therefore $P_t\varphi = e^{-t}\varphi$ for $t \geq 0$. Since $H \geq \varphi$ pointwise and $(P_t)_{t \in [0,\infty)}$ is order preserving, we have

$$\forall_{\varphi \in \Phi} \quad P_t h \geq P_t \varphi = e^{-t}\varphi$$

and thus

$$P_t h \geq e^{-t} \sup_{\varphi \in \Phi} \varphi = e^{-t} h.$$

Hence

$$Lh = -\frac{d}{dt} P_t h \Big|_{t=0^+} = \lim_{t \to 0^+} \frac{P_0 h - P_t h}{t} \leq \lim_{t \to 0^+} \frac{h - e^{-t} h}{t} = h.$$

Seminorm bound - second proof

Now let us follow a different reasoning. By the Hahn-Banach theorem the seminorm H can be expressed as a pointwise supremum of some family of linear functionals Φ:

$$\forall_{x \in \mathbf{R}^n} \quad H(x) = \sup_{\varphi \in \Phi} \varphi(x).$$

Obviously, each of these linear functionals, when restricted to C_n, is a linear combination of Rademacher functions and therefore $P_t \varphi = e^{-t} \varphi$ for $t \geq 0$. Since $H \geq \varphi$ pointwise and $(P_t)_{t \in [0,\infty)}$ is order preserving, we have

$$\forall_{\varphi \in \Phi} \quad P_t h \geq P_t \varphi = e^{-t} \varphi$$

and thus

$$P_t h \geq e^{-t} \sup_{\varphi \in \Phi} \varphi = e^{-t} h.$$

Hence

$$Lh = -\frac{d}{dt} P_t h \Big|_{t=0^+} = \lim_{t \to 0^+} \frac{P_0 h - P_t h}{t} \leq \lim_{t \to 0^+} \frac{h - e^{-t} h}{t} = h.$$

Seminorm bound - second proof

Now let us follow a different reasoning. By the Hahn-Banach theorem the seminorm H can be expressed as a pointwise supremum of some family of linear functionals Φ:

$$\forall_{x \in \mathbf{R}^n} \quad H(x) = \sup_{\varphi \in \Phi} \varphi(x).$$

Obviously, each of these linear functionals, when restricted to C_n, is a linear combination of Rademacher functions and therefore $P_t\varphi = e^{-t}\varphi$ for $t \geq 0$. Since $H \geq \varphi$ pointwise and $(P_t)_{t \in [0,\infty)}$ is order preserving, we have

$$\forall_{\varphi \in \Phi} \quad P_t h \geq P_t \varphi = e^{-t}\varphi$$

and thus

$$P_t h \geq e^{-t} \sup_{\varphi \in \Phi} \varphi = e^{-t} h.$$

Hence

$$Lh = -\frac{d}{dt} P_t h \Big|_{t=0^+} = \lim_{t \to 0^+} \frac{P_0 h - P_t h}{t} \leq \lim_{t \to 0^+} \frac{h - e^{-t} h}{t} = h.$$

Optimal constant $K_{2,1}$ - proof

We know that

- $h \geq 0$ (pointwise),
- h is even,
- $Lh \leq h$ (pointwise),

Thus, by the Poincaré inequality for even functions, we have

$$E[h^2] - (E[h])^2 \leq \frac{1}{2}E[hLh] \leq \frac{1}{2}E[h^2]$$

and therefore $(E[h^2])^{1/2} \leq \sqrt{2} \cdot E[h]$, i.e.

$$(E[\|\sum_{i=1}^{n} r_i v_i\|^2])^{1/2} \leq \sqrt{2} \cdot E[\|\sum_{i=1}^{n} r_i v_i\|].$$

Optimal constant $K_{2,1}$ - proof

We know that

- $h \geq 0$ (pointwise),
- h is even,
- $Lh \leq h$ (pointwise),

Thus, by the Poincaré inequality for even functions, we have

$$E[h^2] - (E[h])^2 \leq \frac{1}{2}E[hLh] \leq \frac{1}{2}E[h^2]$$

and therefore $(E[h^2])^{1/2} \leq \sqrt{2} \cdot E[h]$, i.e.

$$(E[\| \sum_{i=1}^{n} r_i v_i \|^2])^{1/2} \leq \sqrt{2} \cdot E[\| \sum_{i=1}^{n} r_i v_i \|].$$

Optimal constant $K_{4,2}$ - proof

Since h^2 is an even function as well, we have

$$E[h^4] - (E[h^2])^2 \leq \frac{1}{2} E[h^2 L(h^2)].$$

Recall the Stroock-Varopoulos inequality - for $p > 1$ and $f \geq 0$

$$E[f^{p/2} L(f^{p/2})] \leq \frac{p^2}{4(p-1)} E[f^{p-1} Lf],$$

in particular for $p = 4$ and $f = h$ we have

$$E[h^2 L(h^2)] \leq \frac{4}{3} E[h^3 Lh] \leq \frac{4}{3} E[h^4].$$

Hence

$$E[h^4] - (E[h^2])^2 \leq \frac{1}{2} \cdot \frac{4}{3} E[h^4]$$

and thus $(E[h^4])^{1/4} \leq \sqrt[4]{3} \cdot (E[h^2])^{1/2}$, i.e.

$$(E[\| \sum_{i=1}^{n} r_i v_i \|^4])^{1/4} \leq \sqrt[4]{3} \cdot (E[\| \sum_{i=1}^{n} r_i v_i \|^2])^{1/2}.$$

Optimal constant $K_{4,2}$ - proof

Since h^2 is an even function as well, we have

$$E[h^4] - (E[h^2])^2 \leq \frac{1}{2} E[h^2 L(h^2)].$$

Recall the Stroock-Varopoulos inequality - for $p > 1$ and $f \geq 0$

$$E[f^{p/2} L(f^{p/2})] \leq \frac{p^2}{4(p-1)} E[f^{p-1} L f],$$

in particular for $p = 4$ and $f = h$ we have

$$E[h^2 L(h^2)] \leq \frac{4}{3} E[h^3 L h] \leq \frac{4}{3} E[h^4].$$

Hence

$$E[h^4] - (E[h^2])^2 \leq \frac{1}{2} \cdot \frac{4}{3} E[h^4]$$

and thus $(E[h^4])^{1/4} \leq \sqrt[4]{3} \cdot (E[h^2])^{1/2}$, i.e.

$$(E[\| \sum_{i=1}^{n} r_i v_i \|^4])^{1/4} \leq \sqrt[4]{3} \cdot (E[\| \sum_{i=1}^{n} r_i v_i \|^2])^{1/2}.$$

Optimal constant $K_{4,2}$ - proof

Since h^2 is an even function as well, we have

$$E[h^4] - (E[h^2])^2 \leq \frac{1}{2} E[h^2 L(h^2)].$$

Recall the Stroock-Varopoulos inequality - for $p > 1$ and $f \geq 0$

$$E[f^{p/2} L(f^{p/2})] \leq \frac{p^2}{4(p-1)} E[f^{p-1} Lf],$$

in particular for $p = 4$ and $f = h$ we have

$$E[h^2 L(h^2)] \leq \frac{4}{3} E[h^3 Lh] \leq \frac{4}{3} E[h^4].$$

Hence

$$E[h^4] - (E[h^2])^2 \leq \frac{1}{2} \cdot \frac{4}{3} E[h^4]$$

and thus $(E[h^4])^{1/4} \leq \sqrt[4]{3} \cdot (E[h^2])^{1/2}$, i.e.

$$(E[\| \sum_{i=1}^{n} r_i v_i \|^4])^{1/4} \leq \sqrt[4]{3} \cdot (E[\| \sum_{i=1}^{n} r_i v_i \|^2])^{1/2}.$$

Optimal constant $K_{4,2}$ - proof

Since h^2 is an even function as well, we have

$$E[h^4] - (E[h^2])^2 \leq \frac{1}{2} E[h^2 L(h^2)].$$

Recall the Stroock-Varopoulos inequality - for $p > 1$ and $f \geq 0$

$$E[f^{p/2} L(f^{p/2})] \leq \frac{p^2}{4(p-1)} E[f^{p-1} Lf],$$

in particular for $p = 4$ and $f = h$ we have

$$E[h^2 L(h^2)] \leq \frac{4}{3} E[h^3 Lh] \leq \frac{4}{3} E[h^4].$$

Hence

$$E[h^4] - (E[h^2])^2 \leq \frac{1}{2} \cdot \frac{4}{3} E[h^4]$$

and thus $(E[h^4])^{1/4} \leq \sqrt[4]{3} \cdot (E[h^2])^{1/2}$, i.e.

$$(E[\| \sum_{i=1}^{n} r_i v_i \|^4])^{1/4} \leq \sqrt[4]{3} \cdot (E[\| \sum_{i=1}^{n} r_i v_i \|^2])^{1/2}.$$

Open problems

Question 1 (Kwapień): Is it true that for every $p > q > 0$ the optimal constants in the Khinchine and Khinchine-Kahane inequalities are equal, i.e., $C_{p,q} = K_{p,q}$?

It is known that this is the case for $p = 4, q = 2$, and for $q \in (0, 1]$, $p \in (q, 2)$. It is also known that $\sup_{q \in (1,p)} K_{p,q}/C_{p,q} \to 1$ as $p \to \infty$. Szarek showed that the Gaussian analog of this question, with ± 1 variables replaced by $\mathcal{N}(0, 1)$ random variables, follows from the S-inequality.

Question 2 (Pełczyński): Let $(v_{i,j})_{1 \le i < j \le n}$ belong to some normed linear space, and let $r_1, r_2, \ldots, r_n$ be independent symmetric ± 1 random variables. Let $S = \sum_{1 \le i < j \le n} r_i r_j v_{i,j}$. Does it follow that $E[\|S\|^2] \le 4(E[\|S\|])^2$?

Known to be true for $n \le 6$. In general, unknown even for real-valued coefficients. Its decoupled counterpart follows easily from the fact that $K_{2,1} = \sqrt{2}$.

Open problems

Question 1 (Kwapień): Is it true that for every $p > q > 0$ the optimal constants in the Khinchine and Khinchine-Kahane inequalities are equal, i.e., $C_{p,q} = K_{p,q}$?

It is known that this is the case for $p = 4, q = 2$, and for $q \in (0, 1]$, $p \in (q, 2)$. It is also known that $\sup_{q \in (1,p)} K_{p,q}/C_{p,q} \to 1$ as $p \to \infty$. Szarek showed that the Gaussian analog of this question, with ± 1 variables replaced by $\mathcal{N}(0, 1)$ random variables, follows from the S-inequality.

Question 2 (Pełczyński): Let $(v_{i,j})_{1 \le i < j \le n}$ belong to some normed linear space, and let $r_1, r_2, \ldots, r_n$ be independent symmetric ± 1 random variables. Let $S = \sum_{1 \le i < j \le n} r_i r_j v_{i,j}$. Does it follow that $E[\|S\|^2] \le 4(E[\|S\|])^2$?

Known to be true for $n \le 6$. In general, unknown even for real-valued coefficients. Its decoupled counterpart follows easily from the fact that $K_{2,1} = \sqrt{2}$.

Open problems

Question 1 (Kwapień): Is it true that for every $p > q > 0$ the optimal constants in the Khinchine and Khinchine-Kahane inequalities are equal, i.e., $C_{p,q} = K_{p,q}$?

It is known that this is the case for $p = 4, q = 2$, and for $q \in (0, 1]$, $p \in (q, 2)$. It is also known that $\sup_{q \in (1,p)} K_{p,q}/C_{p,q} \to 1$ as $p \to \infty$. Szarek showed that the Gaussian analog of this question, with ± 1 variables replaced by $\mathcal{N}(0, 1)$ random variables, follows from the S-inequality.

Question 2 (Pełczyński): Let $(v_{i,j})_{1 \leq i < j \leq n}$ belong to some normed linear space, and let $r_1, r_2, \ldots, r_n$ be independent symmetric ± 1 random variables. Let $S = \sum_{1 \leq i < j \leq n} r_i r_j v_{i,j}$. Does it follow that $E[\|S\|^2] \leq 4(E[\|S\|])^2$?

Known to be true for $n \leq 6$. In general, unknown even for real-valued coefficients. Its decoupled counterpart follows easily from the fact that $K_{2,1} = \sqrt{2}$.

Open problems

Question 1 (Kwapień): Is it true that for every $p > q > 0$ the optimal constants in the Khinchine and Khinchine-Kahane inequalities are equal, i.e., $C_{p,q} = K_{p,q}$?

It is known that this is the case for $p = 4, q = 2$, and for $q \in (0,1]$, $p \in (q,2)$. It is also known that $\sup_{q\in(1,p)} K_{p,q}/C_{p,q} \to 1$ as $p \to \infty$. Szarek showed that the Gaussian analog of this question, with ± 1 variables replaced by $\mathcal{N}(0,1)$ random variables, follows from the S-inequality.

Question 2 (Pełczyński): Let $(v_{i,j})_{1\le i<j\le n}$ belong to some normed linear space, and let $r_1, r_2, \ldots, r_n$ be independent symmetric ± 1 random variables. Let $S = \sum_{1\le i<j\le n} r_i r_j v_{i,j}$. Does it follow that $E[\|S\|^2] \le 4(E[\|S\|])^2$?

Known to be true for $n \le 6$. In general, unknown even for real-valued coefficients. Its decoupled counterpart follows easily from the fact that $K_{2,1} = \sqrt{2}$.

Chaos

We will say that a polynomial $V \in \mathbf{R}[x_1, x_2, \ldots, x_n]$ is **multilinear (polylinear)** or **of chaos type** if for every $i \in [n]$ there is

$$\partial_{ii} V = \frac{\partial^2 V}{\partial x_i^2} \equiv 0,$$

i.e. no variable appears squared or in higher power (with non-zero coefficient). This is obviously equivalent to the fact that V belongs to the linear span of the constant function $\mathbf{1}$ and multilinear monomials $x_1, x_2, \ldots, x_n, x_1x_2, x_1x_3, \ldots, x_{n-1}x_n, x_1x_2x_3, \ldots, x_1x_2\ldots x_n$.

If $Z_1, Z_2, \ldots, Z_n$ are independent random variables and $V \in \mathbf{R}[x_1, x_2, \ldots, x_n]$ is multilinear then $Z = V(Z_1, Z_2, \ldots, Z_n)$ is called a (tetrahedral) **chaos**.

Note that **any** real function on the discrete cube is a Rademacher chaos (due to the Walsh-Fourier expansion).

Chaos

We will say that a polynomial $V \in \mathbf{R}[x_1, x_2, \ldots, x_n]$ is **multilinear (polylinear)** or **of chaos type** if for every $i \in [n]$ there is

$$\partial_{ii} V = \frac{\partial^2 V}{\partial x_i^2} \equiv 0,$$

i.e. no variable appears squared or in higher power (with non-zero coefficient). This is obviously equivalent to the fact that V belongs to the linear span of the constant function $\mathbf{1}$ and multilinear monomials $x_1, x_2, \ldots, x_n, x_1x_2, x_1x_3, \ldots, x_{n-1}x_n, x_1x_2x_3, \ldots,$ $x_1x_2\ldots x_n$.

If $Z_1, Z_2, \ldots, Z_n$ are independent random variables and $V \in \mathbf{R}[x_1, x_2, \ldots, x_n]$ is multilinear then $Z = V(Z_1, Z_2, \ldots, Z_n)$ is called a (tetrahedral) **chaos.**

Note that **any** real function on the discrete cube is a Rademacher chaos (due to the Walsh-Fourier expansion).

Chaos

We will say that a polynomial $V \in \mathbf{R}[x_1, x_2, \ldots, x_n]$ is **multilinear (polylinear)** or **of chaos type** if for every $i \in [n]$ there is

$$\partial_{ii} V = \frac{\partial^2 V}{\partial x_i^2} \equiv 0,$$

i.e. no variable appears squared or in higher power (with non-zero coefficient). This is obviously equivalent to the fact that V belongs to the linear span of the constant function $\mathbf{1}$ and multilinear monomials $x_1, x_2, \ldots, x_n, x_1x_2, x_1x_3, \ldots, x_{n-1}x_n, x_1x_2x_3, \ldots, x_1x_2\ldots x_n$.

If $Z_1, Z_2, \ldots, Z_n$ are independent random variables and $V \in \mathbf{R}[x_1, x_2, \ldots, x_n]$ is multilinear then $Z = V(Z_1, Z_2, \ldots, Z_n)$ is called a (tetrahedral) **chaos.**

Note that **any** real function on the discrete cube is a Rademacher chaos (due to the Walsh-Fourier expansion).

Elementary moment comparison

Lemma Let $Z_1, Z_2, \ldots, Z_n$ be independent real random variables with $E[Z_i] = E[Z_i^3] = 0$, $E[Z_i^2] = 1$ and $E[Z_i^4] \leq 9$ for $i = 1, 2, \ldots, n$. Let $V \in R[x_1, x_2, \ldots, x_n]$ be of chaos type, $d = \deg V$, and let $Z = V(Z_1, Z_2, \ldots, Z_n)$.
Then $E[Z^4] \leq 9^d (E[Z^2])^2$.

The main example one can have in mind is $Z_i = r_i$ for $i \in [n]$ (comparison of moments for Rademacher chaos). The more general statement above was given just to underline those features of symmetric ± 1 random variables which will be used in the proof.

Proof: We will prove our assertion by induction on n. For $n = 1$ it is trivial. Assume $n > 1$. We can express V as
$V(x_1, x_2, \ldots, x_n) = P(x_1, x_2, \ldots, x_{n-1}) + x_n Q(x_1, x_2, \ldots, x_{n-1})$,
where P and Q are again chaos type polynomials, in at most $n-1$ variables, with $\deg P \leq d$ and $\deg Q \leq d - 1$.

Elementary moment comparison

Lemma Let $Z_1, Z_2, \ldots, Z_n$ be independent real random variables with $E[Z_i] = E[Z_i^3] = 0$, $E[Z_i^2] = 1$ and $E[Z_i^4] \leq 9$ for $i = 1, 2, \ldots, n$. Let $V \in R[x_1, x_2, \ldots, x_n]$ be of chaos type, $d = \deg V$, and let $Z = V(Z_1, Z_2, \ldots, Z_n)$.
Then $E[Z^4] \leq 9^d (E[Z^2])^2$.

The main example one can have in mind is $Z_i = r_i$ for $i \in [n]$ (comparison of moments for Rademacher chaos). The more general statement above was given just to underline those features of symmetric ± 1 random variables which will be used in the proof.

Proof: We will prove our assertion by induction on n. For $n = 1$ it is trivial. Assume $n > 1$. We can express V as $V(x_1, x_2, \ldots, x_n) = P(x_1, x_2, \ldots, x_{n-1}) + x_n Q(x_1, x_2, \ldots, x_{n-1})$, where P and Q are again chaos type polynomials, in at most $n - 1$ variables, with $\deg P \leq d$ and $\deg Q \leq d - 1$.

Lemma Let $Z_1, Z_2, \ldots, Z_n$ be independent real random variables with $E[Z_i] = E[Z_i^3] = 0$, $E[Z_i^2] = 1$ and $E[Z_i^4] \leq 9$ for $i = 1, 2, \ldots, n$. Let $V \in R[x_1, x_2, \ldots, x_n]$ be of chaos type, $d = \deg V$, and let $Z = V(Z_1, Z_2, \ldots, Z_n)$.
Then $E[Z^4] \leq 9^d (E[Z^2])^2$.

The main example one can have in mind is $Z_i = r_i$ for $i \in [n]$ (comparison of moments for Rademacher chaos). The more general statement above was given just to underline those features of symmetric ± 1 random variables which will be used in the proof.

Proof: We will prove our assertion by induction on n. For $n = 1$ it is trivial. Assume $n > 1$. We can express V as
$V(x_1, x_2, \ldots, x_n) = P(x_1, x_2, \ldots, x_{n-1}) + x_n Q(x_1, x_2, \ldots, x_{n-1})$,
where P and Q are again chaos type polynomials, in at most $n-1$ variables, with $\deg P \leq d$ and $\deg Q \leq d-1$.

Moment comparison - proof

Let

$$X = P(Z_1, Z_2, \ldots, Z_{n-1})$$

and

$$Y = Q(Z_1, Z_2, \ldots, Z_{n-1}).$$

Clearly, (X, Y) is independent of Z_n. We have

$$E[Z^4] = E[(X+Z_nY)^4] = E[X^4]+4E[X^3Y]\cdot E[Z_n]+6E[X^2Y^2]\cdot E[Z_n^2]+$$

$$+4E[XY^3]\cdot E[Z_n^3]+E[Y^4]\cdot E[Z_n^4] \leq E[X^4]+6E[X^2Y^2]+9E[Y^4] \leq$$

$$\leq 9^d(E[X^2])^2 + 6(E[X^4])^{1/2}(E[Y^4])^{1/2} + 9 \cdot 9^{d-1}(E[Y^2])^2 \leq$$

$$\leq 9^d(E[X^2])^2 + 6 \cdot 3^d E[X^2] \cdot 3^{d-1} E[Y^2] + 9^d(E[Y^2])^2 =$$

$$9^d(E[X^2]+E[Y^2])^2 = 9^d(E[X^2]+2E[XY]\cdot E[Z_n]+E[Y^2]\cdot E[Z_n^2])^2 =$$

$$= 9^d(E[X + Z_nY]^2)^2 = 9^d(E[Z^2])^2,$$

where the induction hypothesis was used for P and Q. The proof is finished.

Comparison: example

In fact, if Z_i's are just symmetric ± 1 random variables, the constant 9^d is not optimal, for example for $d = 1$ one can prove the above comparison of moments with factor 3 instead of 9. However, the following example indicates that the asymptotic behaviour of the constant is very close to optimal when $d \to \infty$ (even if we restrict our interest to Rademacher chaos only).

Denote by $\binom{[n]}{d}$ all subsets of $[n] = \{1, 2, \ldots, n\}$ with cardinality d. Let

$$V(x_1, x_2, \ldots, x_n) = \sum_{S \in \binom{[n]}{d}} \prod_{i \in S} x_i,$$

so that $deg(V) = d$, and let

$$Z = V(r_1, \ldots, r_n) = \sum_{S \in \binom{[n]}{d}} \prod_{i \in S} r_i.$$

Comparison: example

In fact, if Z_i's are just symmetric ± 1 random variables, the constant 9^d is not optimal, for example for $d = 1$ one can prove the above comparison of moments with factor 3 instead of 9. However, the following example indicates that the asymptotic behaviour of the constant is very close to optimal when $d \to \infty$ (even if we restrict our interest to Rademacher chaos only).

Denote by $\binom{[n]}{d}$ all subsets of $[n] = \{1, 2, \ldots, n\}$ with cardinality d. Let

$$V(x_1, x_2, \ldots, x_n) = \sum_{S \in \binom{[n]}{d}} \prod_{i \in S} x_i,$$

so that $deg(V) = d$, and let

$$Z = V(r_1, \ldots, r_n) = \sum_{S \in \binom{[n]}{d}} \prod_{i \in S} r_i.$$

Comparison example - computation

Let $\delta_i(A) = 1$ if $i \in A$, and $\delta_i(A) = 0$ if $i \notin A$, as usually. Then

$$E[Z^2] = E\Big[\sum_{S_1 \in \binom{[n]}{d}} \prod_{i \in S_1} r_i \cdot \sum_{S_2 \in \binom{[n]}{d}} \prod_{j \in S_2} r_j\Big] =$$

$$\sum_{S_1 \in \binom{[n]}{d}} \sum_{S_2 \in \binom{[n]}{d}} \prod_{i=1}^{n} E[r_i^{\delta_i(S_1)+\delta_i(S_2)}] = \sum_{S \in \binom{[n]}{d}} 1 = \binom{n}{d}.$$

Similarly, we have

$$E[Z^4] = \sum_{S_1,S_2,S_3,S_4 \in \binom{[n]}{d}} \prod_{i=1}^{n} E\left[r_i^{\delta_i(S_1)+\delta_i(S_2)+\delta_i(S_3)+\delta_i(S_4)}\right]$$

and all summands in the above sum are nonnegative.

Comparison example - computation

Let $\delta_i(A) = 1$ if $i \in A$, and $\delta_i(A) = 0$ if $i \notin A$, as usually. Then

$$E[Z^2] = E\Big[\sum_{S_1 \in \binom{[n]}{d}} \prod_{i \in S_1} r_i \cdot \sum_{S_2 \in \binom{[n]}{d}} \prod_{j \in S_2} r_j \Big] =$$

$$\sum_{S_1 \in \binom{[n]}{d}} \sum_{S_2 \in \binom{[n]}{d}} \prod_{i=1}^{n} E[r_i^{\delta_i(S_1)+\delta_i(S_2)}] = \sum_{S \in \binom{[n]}{d}} 1 = \binom{n}{d}.$$

Similarly, we have

$$E[Z^4] = \sum_{S_1, S_2, S_3, S_4 \in \binom{[n]}{d}} \prod_{i=1}^{n} E\Big[r_i^{\delta_i(S_1)+\delta_i(S_2)+\delta_i(S_3)+\delta_i(S_4)} \Big]$$

and all summands in the above sum are nonnegative.

Comparison example - computations

For simplicity assume that d is divisible by 3 and $n \geq 2d$.
Let $A_1, A_2, \ldots, A_6 \subseteq [n]$ be pairwise disjoint with cardinality $d/3$.
Let $S_1 = A_1 \cup A_4 \cup A_5$, $S_2 = A_1 \cup A_2 \cup A_6$, $S_3 = A_3 \cup A_4 \cup A_6$, $S_4 = A_2 \cup A_3 \cup A_5$ (so that $A_1 = S_1 \cap S_2$, $A_2 = S_2 \cap S_4$, $A_3 = S_3 \cap S_4$, $A_4 = S_1 \cap S_3$, $A_5 = S_1 \cap S_4$, $A_6 = S_2 \cap S_3$).

$$\text{Hence } E[Z^4] \geq \sum_{A_1,\ldots,A_6 \text{ as above}} \prod_{i=1}^{n} E\left[r_i^{\delta_i(A_1\cup A_4\cup A_5)+\delta_i(A_1\cup A_2\cup A_6)+\delta_i(A_3\cup A_4\cup A_6)+\delta_i(A_2\cup A_3\cup A_5)}\right]$$

$$= \sum_{A_1,\ldots,A_6} \prod_{i=1}^{n} E\left[r_i^{2\delta_i(A_1\cup A_2\cup A_3\cup A_4\cup A_5\cup A_6)}\right] =$$

$$= \sum_{A_1,\ldots,A_6} 1 = \binom{n}{d/3, d/3, d/3, d/3, d/3, d/3}.$$

Comparison example - computations

For simplicity assume that d is divisible by 3 and $n \geq 2d$.
Let $A_1, A_2, \ldots, A_6 \subseteq [n]$ be pairwise disjoint with cardinality $d/3$.
Let $S_1 = A_1 \cup A_4 \cup A_5$, $S_2 = A_1 \cup A_2 \cup A_6$, $S_3 = A_3 \cup A_4 \cup A_6$, $S_4 = A_2 \cup A_3 \cup A_5$ (so that $A_1 = S_1 \cap S_2$, $A_2 = S_2 \cap S_4$, $A_3 = S_3 \cap S_4$, $A_4 = S_1 \cap S_3$, $A_5 = S_1 \cap S_4$, $A_6 = S_2 \cap S_3$).

$$\text{Hence } E[Z^4] \geq \sum_{A_1,\ldots,A_6 \text{ as above}} \prod_{i=1}^{n}$$

$$E\left[r_i^{\delta_i(A_1\cup A_4\cup A_5)+\delta_i(A_1\cup A_2\cup A_6)+\delta_i(A_3\cup A_4\cup A_6)+\delta_i(A_2\cup A_3\cup A_5)}\right]$$

$$= \sum_{A_1,\ldots,A_6} \prod_{i=1}^{n} E\left[r_i^{2\delta_i(A_1\cup A_2\cup A_3\cup A_4\cup A_5\cup A_6)}\right] =$$

$$= \sum_{A_1,\ldots,A_6} 1 = \binom{n}{d/3, d/3, d/3, d/3, d/3, d/3}.$$

Comparison example - conclusion

Finally, we have

$$\left(E[Z^4]/(E[Z^2])^2\right)^{1/d} \geq \binom{n}{d/3, d/3, d/3, d/3, d/3, d/3}^{1/d} \Big/ \binom{n}{d}^{2/d}$$

$$\geq \left(\frac{n(n-1)\cdot \ldots (n-2d+1)}{((d/3)!)^6}\right)^{1/d} \Big/ \left(\frac{n^d}{d!}\right)^{2/d} \xrightarrow{n\to\infty}$$

$$\xrightarrow{n\to\infty} (d!)^{2/d}/((d/3)!)^{6/d} \xrightarrow{d\to\infty} 9,$$

by Stirling's formula.

Entropy

For an integrable nonnegative function g on a probability space we define its entropy as

$$Ent[g] := E[g \ln g] - E[g] \cdot \ln(E[g]),$$

where we adopt a natural convention, extending in a continuous way $\psi(s) = s \ln s$ from $(0, \infty)$ to $[0, \infty)$ by setting $\psi(0) = 0$.

Clearly, $Ent[g] < \infty$ if and only if $g \ln g$ is integrable.
Since ψ is strictly convex, always there is $Ent[g] \geq 0$,
and $Ent[g] = 0$ if and only if g is constant a.s.

For $\lambda > 0$ we have $Ent(\lambda g) = \lambda \cdot Ent(g)$.

For an integrable nonnegative function g on a probability space we define its entropy as

$$Ent[g] := E[g \ln g] - E[g] \cdot \ln(E[g]),$$

where we adopt a natural convention, extending in a continuous way $\psi(s) = s \ln s$ from $(0, \infty)$ to $[0, \infty)$ by setting $\psi(0) = 0$.

Clearly, $Ent[g] < \infty$ if and only if $g \ln g$ is integrable.
Since ψ is strictly convex, always there is $Ent[g] \geq 0$,
and $Ent[g] = 0$ if and only if g is constant a.s.

For $\lambda > 0$ we have $Ent(\lambda g) = \lambda \cdot Ent(g)$.

Entropy

For an integrable nonnegative function g on a probability space we define its entropy as

$$Ent[g] := E[g \ln g] - E[g] \cdot \ln(E[g]),$$

where we adopt a natural convention, extending in a continuous way $\psi(s) = s \ln s$ from $(0, \infty)$ to $[0, \infty)$ by setting $\psi(0) = 0$.

Clearly, $Ent[g] < \infty$ if and only if $g \ln g$ is integrable. Since ψ is strictly convex, always there is $Ent[g] \geq 0$, and $Ent[g] = 0$ if and only if g is constant a.s.

For $\lambda > 0$ we have $Ent(\lambda g) = \lambda \cdot Ent(g)$.

Logarithmic Sobolev inequality

The **logarithmic Sobolev inequality** (called also **entropy-energy inequality**) was introduced by L. Gross. It resembles the Poincaré inequality - the variance functional on the left hand side is replaced by entropy. However, both variance and energy functionals are quadratic forms while entropy is 1-homogenous. Therefore the inequality takes form:

$$Ent[f^2] \leq C \cdot \mathcal{E}[f].$$

Strictly speaking, a symmetric Markov semigroup $(Q_t)_{t\in[0,\infty)}$ on Ω, with an invariant measure μ and a self-adjoint (with respect to the $L^2(\Omega, \mu)$ structure) generator L, satisfies the logarithmic Sobolev inequality with constant $C > 0$ if for every function f belonging to the domain of L there is

$$E_\mu[f^2 \ln(f^2)] - E_\mu[f^2] \ln E_\mu[f^2] \leq C \cdot E_\mu[f\, Lf].$$

Logarithmic Sobolev inequality

The **logarithmic Sobolev inequality** (called also **entropy-energy inequality**) was introduced by L. Gross. It resembles the Poincaré inequality - the variance functional on the left hand side is replaced by entropy. However, both variance and energy functionals are quadratic forms while entropy is 1-homogenous. Therefore the inequality takes form:

$$Ent[f^2] \leq C \cdot \mathcal{E}[f].$$

Strictly speaking, a symmetric Markov semigroup $(Q_t)_{t \in [0,\infty)}$ on Ω, with an invariant measure μ and a self-adjoint (with respect to the $L^2(\Omega, \mu)$ structure) generator L, satisfies the logarithmic Sobolev inequality with constant $C > 0$ if for every function f belonging to the domain of L there is

$$E_\mu[f^2 \ln(f^2)] - E_\mu[f^2] \ln E_\mu[f^2] \leq C \cdot E_\mu[f\, Lf].$$

Logarithmic Sobolev inequalities - continued

We will prove that $(P_t)_{t\in[0,\infty)}$ satisfies the logarithmic Sobolev inequality with constant 2, i.e. for every $f \in \mathcal{H}_n$ we have

$$E[f^2 \ln(f^2)] - E[f^2] \ln E[f^2] \leq 2 \cdot E[f\, Lf].$$

To avoid technicalities we will concentrate on the case of $(P_t)_{t\in[0,\infty)}$ but most of our arguments, after some appropriate modifications, may be applied to a large class of **symmetric** Markov semigroups. Therefore we will first describe some equivalent formulations in which the constant C appears, and only then we will prove that in our discrete cube setting we can set $C = 2$.

Remark: Note that a linear change of time parameter t in $(P_t)_{t\in[0,\infty)}$ is reflected by an analogous rescaling of the semigroup's generator. Thus the optimal constants in the logarithmic Sobolev inequality for different symetric Markov semigroups may vary.

Logarithmic Sobolev inequalities - continued

We will prove that $(P_t)_{t\in[0,\infty)}$ satisfies the logarithmic Sobolev inequality with constant 2, i.e. for every $f \in \mathcal{H}_n$ we have

$$E[f^2 \ln(f^2)] - E[f^2] \ln E[f^2] \leq 2 \cdot E[f\, Lf].$$

To avoid technicalities we will concentrate on the case of $(P_t)_{t\in[0,\infty)}$ but most of our arguments, after some appropriate modifications, may be applied to a large class of **symmetric** Markov semigroups. Therefore we will first describe some equivalent formulations in which the constant C appears, and only then we will prove that in our discrete cube setting we can set $C = 2$.

Remark: Note that a linear change of time parameter t in $(P_t)_{t\in[0,\infty)}$ is reflected by an analogous rescaling of the semigroup's generator. Thus the optimal constants in the logarithmic Sobolev inequality for different symetric Markov semigroups may vary.

Logarithmic Sobolev inequalities - continued

We will prove that $(P_t)_{t\in[0,\infty)}$ satisfies the logarithmic Sobolev inequality with constant 2, i.e. for every $f \in \mathcal{H}_n$ we have

$$E[f^2 \ln(f^2)] - E[f^2] \ln E[f^2] \leq 2 \cdot E[f\, Lf].$$

To avoid technicalities we will concentrate on the case of $(P_t)_{t\in[0,\infty)}$ but most of our arguments, after some appropriate modifications, may be applied to a large class of **symmetric** Markov semigroups. Therefore we will first describe some equivalent formulations in which the constant C appears, and only then we will prove that in our discrete cube setting we can set $C = 2$.

Remark: Note that a linear change of time parameter t in $(P_t)_{t\in[0,\infty)}$ is reflected by an analogous rescaling of the semigroup's generator. Thus the optimal constants in the logarithmic Sobolev inequality for different symetric Markov semigroups may vary.

Logarithmic Sobolev inequality - equivalent versions

The following statements are equivalent:

- For every $f \in \mathcal{H}_n$

$$E[f^2 \ln(f^2)] - E[f^2] \ln E[f^2] \leq C \cdot E[f\, Lf].$$

- For every nonnegative $f \in \mathcal{H}_n$

$$E[f^2 \ln(f^2)] - E[f^2] \ln E[f^2] \leq C \cdot E[f\, Lf].$$

- For every nonnegative $f \in \mathcal{H}_n$ and **every** $p > 1$

$$E[f^p \ln(f^p)] - E[f^p] \ln E[f^p] \leq \frac{Cp^2}{4(p-1)} \cdot E[f^{p-1} Lf].$$

Logarithmic Sobolev inequalities - proof of the equivalence

The second statement (for nonnegative functions) trivially follows from the first one (for the whole $\mathcal{H}_n$). The reverse implication is also easy. Let $f \in \mathcal{H}_n$. We use the second statement for a nonnegative function $|f|$ and apply the inequality

$$\mathcal{E}[|f|] = E[|f|\, L|f|] \leq E[f\, Lf] = \mathcal{E}[f],$$

which we proved earlier.

The second statement is a special case of the third one (for $p = 2$). To prove the reverse implication we use the second statement for $f^{p/2}$ instead of f:

$$E[|f|^p \ln(f^p)] - E[f^p] \ln E[f^p] \leq C \cdot E[f^{p/2} L(f^{p/2})] \leq \frac{Cp^2}{4(p-1)} \cdot E[f^{p-1} Lf],$$

where we have used the Stroock-Varopoulos inequality.

Logarithmic Sobolev inequalities - proof of the equivalence

The second statement (for nonnegative functions) trivially follows from the first one (for the whole $\mathcal{H}_n$). The reverse implication is also easy. Let $f \in \mathcal{H}_n$. We use the second statement for a nonnegative function $|f|$ and apply the inequality

$$\mathcal{E}[|f|] = E[|f|\, L|f|] \leq E[f\, Lf] = \mathcal{E}[f],$$

which we proved earlier.

The second statement is a special case of the third one (for $p = 2$). To prove the reverse implication we use the second statement for $f^{p/2}$ instead of f:

$$E[|f|^p \ln(f^p)] - E[f^p] \ln E[f^p] \leq C \cdot E[f^{p/2} L(f^{p/2})] \leq \frac{Cp^2}{4(p-1)} \cdot E[f^{p-1} Lf],$$

where we have used the Stroock-Varopoulos inequality.

Logarithmic Sobolev inequality - semigroup application

For a nonnegative $f \in \mathcal{H}_n$ and $p > 1$ let us define $\phi_q : [q, \infty) \longrightarrow \mathbf{R}$ by

$$\phi_q(p) = \ln \|P_{t(p)} f\|_p = \frac{1}{p} \ln E[(P_{t(p)} f)^p],$$

where $t(p) = \frac{C}{4} \ln \frac{p-1}{q-1}$.

It is easy to see that $t(q) = 0$ and $t(p) \geq 0$ for $p \geq q$, so that $f_p := P_{t(p)} f \geq 0$. Note that $\phi_q(q) = \ln \|f\|_q$. An elementary computation shows that

$$\frac{d}{dp}\phi_q(p) = \frac{1}{p} \frac{E[\frac{d}{dp}(f_p^p)]}{E[f_p^p]} - \frac{1}{p^2} \ln E[f_p^p]$$

and

$$\frac{d}{dp}(f_p^p) = \frac{1}{p} f_p^p \ln(f_p^p) - \frac{Cp}{4(p-1)} f_p^{p-1} L f_p.$$

Logarithmic Sobolev inequality - semigroup application

For a nonnegative $f \in \mathcal{H}_n$ and $p > 1$ let us define $\phi_q : [q, \infty) \longrightarrow \mathbf{R}$ by

$$\phi_q(p) = \ln \|P_{t(p)} f\|_p = \frac{1}{p} \ln E[(P_{t(p)} f)^p],$$

where $t(p) = \frac{C}{4} \ln \frac{p-1}{q-1}$.

It is easy to see that $t(q) = 0$ and $t(p) \geq 0$ for $p \geq q$, so that $f_p := P_{t(p)} f \geq 0$. Note that $\phi_q(q) = \ln \|f\|_q$. An elementary computation shows that

$$\frac{d}{dp}\phi_q(p) = \frac{1}{p} \frac{E[\frac{d}{dp}(f_p^p)]}{E[f_p^p]} - \frac{1}{p^2} \ln E[f_p^p]$$

and

$$\frac{d}{dp}(f_p^p) = \frac{1}{p} f_p^p \ln(f_p^p) - \frac{Cp}{4(p-1)} f_p^{p-1} L f_p.$$

Semigroup application - continued

Thus $\frac{d}{dp}\phi_q(p) \leq 0$ if and only if

$$Ent(f_p^p) \leq \frac{Cp^2}{4(p-1)} \cdot E[f_p^{p-1} L f_p],$$

which, as we have seen, is the logarithmic Sobolev inequality with constant C applied to the function f_p.

Hence the logarithmic Sobolev inequality implies the fact that ϕ_q is decreasing. The **partial** converse follows from computing $\frac{d}{dp}\phi_q(p)\Big|_{p=q}$ and using the fact that it must be nonpositive.

Remark: In particular, it is just enough to know that ϕ_2 is nonincreasing to obtain

$$E[f^2 \ln(f^2)] - E[f^2] \ln E[f^2] \leq C \cdot E[f\, Lf],$$

$$\text{so } E[f^q \ln(f^q)] - E[f^q] \ln E[f^q] \leq \frac{Cq^2}{4(q-1)} \cdot E[f^{q-1} L f]$$

for $q > 1$, and thus also ϕ_q is nonincreasing for all $q > 1$.

Semigroup application - continued

Thus $\frac{d}{dp}\phi_q(p) \le 0$ if and only if

$$Ent(f_p^p) \le \frac{Cp^2}{4(p-1)} \cdot E[f_p^{p-1} L f_p],$$

which, as we have seen, is the logarithmic Sobolev inequality with constant C applied to the function f_p.

Hence the logarithmic Sobolev inequality implies the fact that ϕ_q is decreasing. The **partial** converse follows from computing $\frac{d}{dp}\phi_q(p)\Big|_{p=q}$ and using the fact that it must be nonpositive.

Remark: In particular, it is just enough to know that ϕ_2 is nonincreasing to obtain

$$E[f^2 \ln(f^2)] - E[f^2] \ln E[f^2] \le C \cdot E[f\, Lf],$$

$$\text{so } E[f^q \ln(f^q)] - E[f^q] \ln E[f^q] \le \frac{Cq^2}{4(q-1)} \cdot E[f^{q-1} Lf]$$

for $q > 1$, and thus also ϕ_q is nonincreasing for all $q > 1$.

Thus $\frac{d}{dp}\phi_q(p) \leq 0$ if and only if

$$Ent(f_p^p) \leq \frac{Cp^2}{4(p-1)} \cdot E[f_p^{p-1} L f_p],$$

which, as we have seen, is the logarithmic Sobolev inequality with constant C applied to the function f_p.
Hence the logarithmic Sobolev inequality implies the fact that ϕ_q is decreasing. The **partial** converse follows from computing $\frac{d}{dp}\phi_q(p)\Big|_{p=q}$ and using the fact that it must be nonpositive.
Remark: In particular, it is just enough to know that ϕ_2 is nonincreasing to obtain

$$E[f^2 \ln(f^2)] - E[f^2] \ln E[f^2] \leq C \cdot E[f\, Lf],$$

so $$E[f^q \ln(f^q)] - E[f^q] \ln E[f^q] \leq \frac{Cq^2}{4(q-1)} \cdot E[f^{q-1} Lf]$$

for $q > 1$, and thus also ϕ_q is nonincreasing for all $q > 1$.

Hypercontractivity

Let $C > 0$. The following statements are equivalent:

- For every nonnegative $f \in H_n$
$$E[f^2 \ln(f^2)] - E[f^2] \ln E[f^2] \leq C \cdot E[f\, Lf].$$
- For every $p > q > 1$ and every nonnegative $f \in \mathcal{H}_n$
$$\|P_t f\|_p \leq \|f\|_q$$
for $t = t(p,q) = \frac{C}{4} \ln \frac{p-1}{q-1}$.
- For every $p > q > 1$ and every nonnegative $f \in \mathcal{H}_n$
$$\|P_t f\|_p \leq \|f\|_q$$
for every $t \geq t(p,q) = \frac{C}{4} \ln \frac{p-1}{q-1}$.
- For every $p > q > 1$ and every $f \in \mathcal{H}_n$
$$\|P_t f\|_p \leq \|f\|_q$$
for every $t \geq t(p,q) = \frac{C}{4} \ln \frac{p-1}{q-1}$.

This property of the semigroup is called **hypercontractivity**.

Hypercontractivity

Let $C > 0$. The following statements are equivalent:

- For every nonnegative $f \in H_n$
$$E[f^2 \ln(f^2)] - E[f^2] \ln E[f^2] \leq C \cdot E[f\, Lf].$$
- For every $p > q > 1$ and every nonnegative $f \in \mathcal{H}_n$
$$\|P_t f\|_p \leq \|f\|_q$$
for $t = t(p,q) = \frac{C}{4} \ln \frac{p-1}{q-1}$.
- For every $p > q > 1$ and every nonnegative $f \in \mathcal{H}_n$
$$\|P_t f\|_p \leq \|f\|_q$$
for every $t \geq t(p,q) = \frac{C}{4} \ln \frac{p-1}{q-1}$.
- For every $p > q > 1$ and every $f \in \mathcal{H}_n$
$$\|P_t f\|_p \leq \|f\|_q$$
for every $t \geq t(p,q) = \frac{C}{4} \ln \frac{p-1}{q-1}$.

This property of the semigroup is called **hypercontractivity**.

Hypercontractivity - explanation

Many of the implications are obvious. Passing from $t = t(p,q)$ to general $t \geq t(p,q)$ follows from the fact that

$$P_t f = P_{t-t(p,q)}(P_{t(p,q)} f)$$

and from the contractivity of $P_{t-t(p,q)}$.

To pass from nonnegative to arbitrary $f \in \mathcal{H}_n$ we just note that $|f| \geq f \geq -|f|$ pointwise, and since P_t is order preserving we have

$$P_t|f| \geq P_t f \geq -P_t|f|,$$

i.e. $|P_t f| \leq P_t|f|$ pointwise, so that $\|P_t f\|_p \leq \|P_t|f|\,\|_p$. Thus we can apply the inequality for nonnegative functions to $|f|$ and deduce the general statement.

Hypercontractivity - explanation

Many of the implications are obvious. Passing from $t = t(p, q)$ to general $t \geq t(p, q)$ follows from the fact that

$$P_t f = P_{t-t(p,q)}(P_{t(p,q)} f)$$

and from the contractivity of $P_{t-t(p,q)}$.

To pass from nonnegative to arbitrary $f \in \mathcal{H}_n$ we just note that $|f| \geq f \geq -|f|$ pointwise, and since P_t is order preserving we have

$$P_t|f| \geq P_t f \geq -P_t|f|,$$

i.e. $|P_t f| \leq P_t|f|$ pointwise, so that $\|P_t f\|_p \leq \|P_t|f|\,\|_p$. Thus we can apply the inequality for nonnegative functions to $|f|$ and deduce the general statement.

Hypercontractivity - application

For every $p > q > 1$ and for any natural n, any normed linear space F and any collection of vectors $v_1, v_2, \ldots, v_n \in F$ there is

$$(E[\|\sum_{i=1}^{n} r_i v_i\|^p])^{1/p} \leq \left(\frac{p-1}{q-1}\right)^{C/4} \cdot (E[\|\sum_{i=1}^{n} r_i v_i\|^q])^{1/q},$$

where $C > 0$ is such that the logarithmic Sobolev inequality with constant C holds on C_n and $r_1, r_2, \ldots$ are independent symmetric ± 1 random variables.

Since the optimal $C = 2$, we get the Kahane inequality with constant $\sqrt{p-1}/\sqrt{q-1}$.

For every $p > q > 1$ and for any natural n, any normed linear space F and any collection of vectors $v_1, v_2, \ldots, v_n \in F$ there is

$$(E[\| \sum_{i=1}^{n} r_i v_i \|^p])^{1/p} \leq \left(\frac{p-1}{q-1}\right)^{C/4} \cdot (E[\| \sum_{i=1}^{n} r_i v_i \|^q])^{1/q},$$

where $C > 0$ is such that the logarithmic Sobolev inequality with constant C holds on C_n and $r_1, r_2, \ldots$ are independent symmetric ± 1 random variables.

Since the optimal $C = 2$, we get the Kahane inequality with constant $\sqrt{p-1}/\sqrt{q-1}$.

$K_{p,q} \leq \sqrt{p-1}/\sqrt{q-1}$ - proof

Again we consider the function $h(x) = \|\sum_{i=1}^{n} x_i v_i\|$. We have proved that $P_t h \geq e^{-t} h \geq 0$ pointwise for $t \geq 0$.

Therefore for $t = \frac{C}{4} \ln \frac{p-1}{q-1}$ we have

$$\left(\frac{q-1}{p-1}\right)^{C/4} \|h\|_p = e^{-t}\|h\|_p \leq \|P_t h\|_p \leq \|h\|_q,$$

so that

$$(E[h^p])^{1/p} \leq \left(\frac{p-1}{q-1}\right)^{C/4} \cdot (E[h^q])^{1/q}$$

and the proof is finished.

$K_{p,q} \leq \sqrt{p-1}/\sqrt{q-1}$ - proof

Again we consider the function $h(x) = \| \sum_{i=1}^{n} x_i v_i \|$. We have proved that $P_t h \geq e^{-t} h \geq 0$ pointwise for $t \geq 0$.

Therefore for $t = \frac{C}{4} \ln \frac{p-1}{q-1}$ we have

$$\left(\frac{q-1}{p-1}\right)^{C/4} \|h\|_p = e^{-t} \|h\|_p \leq \|P_t h\|_p \leq \|h\|_q,$$

so that

$$(E[h^p])^{1/p} \leq \left(\frac{p-1}{q-1}\right)^{C/4} \cdot (E[h^q])^{1/q}$$

and the proof is finished.

Log-Sobolev inequality for $(P_t)_{t\in[0,\infty)}$ - optimal constant

We will prove that the log-Sobolev inequality for the semigroup $(P_t)_{t\in[0,\infty)}$ holds with the constant $C = 2$.

This result is due to L. Gross but its equivalent versions were proved earlier by A. Bonami and W. Beckner. The main ideas and the very notion of hypercontractivity go back to the works of Nelson.

It is clear that the log-Sobolev inequality cannot hold with $C < 2$. Indeed, we know that the Khinchine-Kahane inequality holds with $C_{p,q} = (p-1)^{C/4}/(q-1)^{C/4}$ for any $p > q > 1$. For $p = 2k$, $q = 2$ we get by the CLT argument ($a_1 = \ldots = a_n = n^{-1/2}$, $n \longrightarrow \infty$)

$$\Big((2k-1)!!\Big)^{1/2k} = (E[G^{2k}])^{1/2k} \leq (2k-1)^{C/4}(E[G^2])^{1/2} = (2k-1)^{C/4}$$

where $G \sim \mathcal{N}(0,1)$. The left hand side grows like $\sqrt{k}$ thus $C/4 \geq 1/2$, so that $C \geq 2$.

We will prove that the log-Sobolev inequality for the semigroup $(P_t)_{t\in[0,\infty)}$ holds with the constant $C = 2$.

This result is due to L. Gross but its equivalent versions were proved earlier by A. Bonami and W. Beckner. The main ideas and the very notion of hypercontractivity go back to the works of Nelson.

It is clear that the log-Sobolev inequality cannot hold with $C < 2$. Indeed, we know that the Khinchine-Kahane inequality holds with $C_{p,q} = (p-1)^{C/4}/(q-1)^{C/4}$ for any $p > q > 1$. For $p = 2k$, $q = 2$ we get by the CLT argument $(a_1 = \ldots = a_n = n^{-1/2},\ n \longrightarrow \infty)$

$$\left((2k-1)!!\right)^{1/2k} = (E[G^{2k}])^{1/2k} \leq (2k-1)^{C/4}(E[G^2])^{1/2} = (2k-1)^{C/4}$$

where $G \sim \mathcal{N}(0,1)$. The left hand side grows like $\sqrt{k}$ thus $C/4 \geq 1/2$, so that $C \geq 2$.

We will prove that the log-Sobolev inequality for the semigroup $(P_t)_{t\in[0,\infty)}$ holds with the constant $C = 2$.

This result is due to L. Gross but its equivalent versions were proved earlier by A. Bonami and W. Beckner. The main ideas and the very notion of hypercontractivity go back to the works of Nelson.

It is clear that the log-Sobolev inequality cannot hold with $C < 2$. Indeed, we know that the Khinchine-Kahane inequality holds with $C_{p,q} = (p-1)^{C/4}/(q-1)^{C/4}$ for any $p > q > 1$. For $p = 2k$, $q = 2$ we get by the CLT argument ($a_1 = \ldots = a_n = n^{-1/2}$, $n \longrightarrow \infty$)

$$\Big((2k-1)!!\Big)^{1/2k} = (E[G^{2k}])^{1/2k} \le (2k-1)^{C/4}(E[G^2])^{1/2} = (2k-1)^{C/4}$$

where $G \sim \mathcal{N}(0,1)$. The left hand side grows like $\sqrt{k}$ thus $C/4 \ge 1/2$, so that $C \ge 2$.

Log-Sobolev inequality for $(P_t)_{t\in[0,\infty)}$ - optimal constant

We will prove that the log-Sobolev inequality for the semigroup $(P_t)_{t\in[0,\infty)}$ holds with the constant $C = 2$.

This result is due to L. Gross but its equivalent versions were proved earlier by A. Bonami and W. Beckner. The main ideas and the very notion of hypercontractivity go back to the works of Nelson.

It is clear that the log-Sobolev inequality cannot hold with $C < 2$. Indeed, we know that the Khinchine-Kahane inequality holds with $C_{p,q} = (p-1)^{C/4}/(q-1)^{C/4}$ for any $p > q > 1$. For $p = 2k$, $q = 2$ we get by the CLT argument ($a_1 = \ldots = a_n = n^{-1/2}$, $n \longrightarrow \infty$)

$$\Big((2k-1)!!\Big)^{1/2k} = (E[G^{2k}])^{1/2k} \leq (2k-1)^{C/4}(E[G^2])^{1/2} = (2k-1)^{C/4}$$

where $G \sim \mathcal{N}(0,1)$. The left hand side grows like $\sqrt{k}$ thus $C/4 \geq 1/2$, so that $C \geq 2$.

$L^p \to L^2$ hypercontractivity for $p \in (1, 2]$

Let $p \in (1, 2]$. We will prove that $P_{-\frac{1}{2}\ln(p-1)}$ is contractive as a linear operator from L^p to L^2, i.e.

$$\|P_{-\frac{1}{2}\ln(p-1)} f\|_2 \le \|f\|_p$$

for every $f \in \mathcal{H}_n$ (as we know, we may w.l.o.g. assume $f \ge 0$).

Clearly, the inequality turns into equality for $p = 2$. Therefore the above hypercontractive estimate implies

$$\frac{d}{dp}\|P_{-\frac{1}{2}\ln(p-1)} f\|_2 \Big|_{p=2^-} \ge \frac{d}{dp}\|f\|_p \Big|_{p=2^-}$$

which (after some elementary computation of the type we already know) takes form of

$$Ent[f^2] \le 2 \cdot E[f\, Lf].$$

Hence our task is reduced to proving that $\|P_{-\frac{1}{2}\ln(p-1)}\|_{L^p \to L^2} \le 1$.

$L^p \longrightarrow L^2$ hypercontractivity for $p \in (1, 2]$

Let $p \in (1, 2]$. We will prove that $P_{-\frac{1}{2}\ln(p-1)}$ is contractive as a linear operator from L^p to L^2, i.e.

$$\|P_{-\frac{1}{2}\ln(p-1)}f\|_2 \leq \|f\|_p$$

for every $f \in \mathcal{H}_n$ (as we know, we may w.l.o.g. assume $f \geq 0$).

Clearly, the inequality turns into equality for $p = 2$. Therefore the above hypercontractive estimate implies

$$\frac{d}{dp}\|P_{-\frac{1}{2}\ln(p-1)}f\|_2 \Big|_{p=2^-} \geq \frac{d}{dp}\|f\|_p \Big|_{p=2^-}$$

which (after some elementary computation of the type we already know) takes form of

$$Ent[f^2] \leq 2 \cdot E[f\, Lf].$$

Hence our task is reduced to proving that $\|P_{-\frac{1}{2}\ln(p-1)}\|_{L^p \to L^2} \leq 1$.

$L^p \longrightarrow L^2$ hypercontractivity for $p \in (1, 2]$

Let $p \in (1, 2]$. We will prove that $P_{-\frac{1}{2}\ln(p-1)}$ is contractive as a linear operator from L^p to L^2, i.e.

$$\|P_{-\frac{1}{2}\ln(p-1)} f\|_2 \leq \|f\|_p$$

for every $f \in \mathcal{H}_n$ (as we know, we may w.l.o.g. assume $f \geq 0$).

Clearly, the inequality turns into equality for $p = 2$. Therefore the above hypercontractive estimate implies

$$\frac{d}{dp}\|P_{-\frac{1}{2}\ln(p-1)} f\|_2 \Big|_{p=2^-} \geq \frac{d}{dp}\|f\|_p \Big|_{p=2^-}$$

which (after some elementary computation of the type we already know) takes form of

$$Ent[f^2] \leq 2 \cdot E[f\, Lf].$$

Hence our task is reduced to proving that $\|P_{-\frac{1}{2}\ln(p-1)}\|_{L^p \to L^2} \leq 1$.

Elementary inequalities

We need two easy elementary inequalities:

$$\forall_{a>-1,\, b>1} \quad (1+a)^b \geq 1 + ab$$

and

$$\forall_{a,b\in\mathbf{R}:\, a\geq|b|} \quad \left(\frac{(a+b)^p + (a-b)^p}{2}\right)^{2/p} \geq a^2 + (p-1)b^2.$$

The first one is well-known and trivial: $a \mapsto 1 + ab$ is a supporting (tangent) function of a convex function $a \mapsto (1+a)^b$ at $a = 0$.

To prove the second inequality let us consider a function $\gamma : [-1,1] \longrightarrow \mathrm{R}$ defined by

$$\gamma(u) = \frac{(1+u)^p + (1-u)^p}{2}.$$

Obviously, $\gamma(0) = 1$, $\gamma'(0) = 0$.

Elementary inequalities

We need two easy elementary inequalities:

$$\forall_{a>-1,\, b>1} \quad (1+a)^b \geq 1+ab$$

and

$$\forall_{a,b\in\mathbf{R}:\, a\geq|b|} \quad \left(\frac{(a+b)^p+(a-b)^p}{2}\right)^{2/p} \geq a^2+(p-1)b^2.$$

The first one is well-known and trivial: $a \mapsto 1+ab$ is a supporting (tangent) function of a convex function $a \mapsto (1+a)^b$ at $a=0$.

To prove the second inequality let us consider a function $\gamma : [-1,1] \longrightarrow \mathrm{R}$ defined by

$$\gamma(u) = \frac{(1+u)^p+(1-u)^p}{2}.$$

Obviously, $\gamma(0)=1$, $\gamma'(0)=0$.

Elementary inequalities

We need two easy elementary inequalities:

$$\forall_{a>-1,\, b>1} \quad (1+a)^b \geq 1+ab$$

and

$$\forall_{a,b\in\mathbf{R}:\, a\geq |b|} \quad \left(\frac{(a+b)^p+(a-b)^p}{2}\right)^{2/p} \geq a^2+(p-1)b^2.$$

The first one is well-known and trivial: $a \mapsto 1+ab$ is a supporting (tangent) function of a convex function $a \mapsto (1+a)^b$ at $a=0$.

To prove the second inequality let us consider a function $\gamma : [-1,1] \longrightarrow \mathbf{R}$ defined by

$$\gamma(u) = \frac{(1+u)^p+(1-u)^p}{2}.$$

Obviously, $\gamma(0)=1$, $\gamma'(0)=0$.

Elementary inequalities - continued

We have

$$\gamma''(u) = p(p-1)\frac{(1+u)^{p-2} + (1-u)^{p-2}}{2} \geq p(p-1)$$

because $s \mapsto s^{p-2}$ is convex on $[0,2]$.

Therefore $\gamma(u) \geq 1 + p(p-1)u^2/2$, so that

$$\Big(\frac{(1+u)^p + (1-u)^p}{2}\Big)^{2/p} = \gamma(u)^{2/p} \geq$$

$$\geq \Big(1 + \frac{p(p-1)}{2}u^2\Big)^{2/p} \geq 1 + \frac{2}{p} \cdot \frac{p(p-1)}{2}u^2 = 1 + (p-1)u^2,$$

where we have used the first inequality.

The homogeneity yields that for $|b| \leq a$ there is

$$\Big(\frac{(a+b)^p + (a-b)^p}{2}\Big)^{2/p} \geq a^2 + (p-1)b^2,$$

which is our second inequality.

Elementary inequalities - continued

We have

$$\gamma''(u) = p(p-1)\frac{(1+u)^{p-2} + (1-u)^{p-2}}{2} \geq p(p-1)$$

because $s \mapsto s^{p-2}$ is convex on $[0,2]$.

Therefore $\gamma(u) \geq 1 + p(p-1)u^2/2$, so that

$$\left(\frac{(1+u)^p + (1-u)^p}{2}\right)^{2/p} = \gamma(u)^{2/p} \geq$$

$$\geq \left(1 + \frac{p(p-1)}{2}u^2\right)^{2/p} \geq 1 + \frac{2}{p} \cdot \frac{p(p-1)}{2}u^2 = 1 + (p-1)u^2,$$

where we have used the first inequality.

The homogeneity yields that for $|b| \leq a$ there is

$$\left(\frac{(a+b)^p + (a-b)^p}{2}\right)^{2/p} \geq a^2 + (p-1)b^2,$$

which is our second inequality.

Elementary inequalities - continued

We have

$$\gamma''(u) = p(p-1)\frac{(1+u)^{p-2} + (1-u)^{p-2}}{2} \geq p(p-1)$$

because $s \mapsto s^{p-2}$ is convex on $[0,2]$.

Therefore $\gamma(u) \geq 1 + p(p-1)u^2/2$, so that

$$\left(\frac{(1+u)^p + (1-u)^p}{2}\right)^{2/p} = \gamma(u)^{2/p} \geq$$

$$\geq \left(1 + \frac{p(p-1)}{2}u^2\right)^{2/p} \geq 1 + \frac{2}{p} \cdot \frac{p(p-1)}{2}u^2 = 1 + (p-1)u^2,$$

where we have used the first inequality.

The homogeneity yields that for $|b| \leq a$ there is

$$\left(\frac{(a+b)^p + (a-b)^p}{2}\right)^{2/p} \geq a^2 + (p-1)b^2,$$

which is our second inequality.

$L^p \longrightarrow L^2$ hypercontractivity for $p \in (1,2]$ - case $n = 1$

We have proved that for $|b| \leq a$

$$\left(\frac{(a+b)^p + (a-b)^p}{2} \right)^{2/p} \geq a^2 + (p-1)b^2,$$

i.e.

$$\|P_{-\frac{1}{2}\ln(p-1)} f\|_2 = (a^2+(p-1)b^2)^{1/2} \leq \left(\frac{(a+b)^p + (a-b)^p}{2} \right)^{1/p} = \|f\|_p,$$

where $f : \{-1,1\} \longrightarrow \mathbf{R}$ is given by the formula $f(x_1) = a + bx_1$ or $f = a + br_1$, so that $P_{-\frac{1}{2}\ln(p-1)} f = a + (p-1)^{1/2} br_1$.

Since **every** nonnegative $f : \{-1,1\} \longrightarrow \mathbf{R}$ is of the above form with $|b| \leq a$, we have just proved the hypercontractive estimate on $C_1 = \{-1,1\}$.

We will use induction on n to tranfer this result to C_n.

$L^p \longrightarrow L^2$ hypercontractivity for $p \in (1,2]$ - case $n = 1$

We have proved that for $|b| \leq a$

$$\left(\frac{(a+b)^p + (a-b)^p}{2} \right)^{2/p} \geq a^2 + (p-1)b^2,$$

i.e.

$$\|P_{-\frac{1}{2}\ln(p-1)} f\|_2 = (a^2+(p-1)b^2)^{1/2} \leq \left(\frac{(a+b)^p + (a-b)^p}{2} \right)^{1/p} = \|f\|_p,$$

where $f : \{-1,1\} \longrightarrow \mathbf{R}$ is given by the formula $f(x_1) = a + bx_1$ or $f = a + br_1$, so that $P_{-\frac{1}{2}\ln(p-1)} f = a + (p-1)^{1/2} br_1$.

Since **every** nonnegative $f : \{-1,1\} \longrightarrow \mathbf{R}$ is of the above form with $|b| \leq a$, we have just proved the hypercontractive estimate on $C_1 = \{-1,1\}$.

We will use induction on n to tranfer this result to C_n.

$L^p \longrightarrow L^2$ hypercontractivity for $p \in (1, 2]$ - case $n = 1$

We have proved that for $|b| \leq a$

$$\left(\frac{(a+b)^p + (a-b)^p}{2} \right)^{2/p} \geq a^2 + (p-1)b^2,$$

i.e.

$$\|P_{-\frac{1}{2}\ln(p-1)} f\|_2 = (a^2+(p-1)b^2)^{1/2} \leq \left(\frac{(a+b)^p + (a-b)^p}{2} \right)^{1/p} = \|f\|_p,$$

where $f : \{-1, 1\} \longrightarrow \mathbf{R}$ is given by the formula $f(x_1) = a + bx_1$ or $f = a + br_1$, so that $P_{-\frac{1}{2}\ln(p-1)} f = a + (p-1)^{1/2} br_1$.

Since **every** nonnegative $f : \{-1, 1\} \longrightarrow \mathbf{R}$ is of the above form with $|b| \leq a$, we have just proved the hypercontractive estimate on $C_1 = \{-1, 1\}$.

We will use induction on n to tranfer this result to C_n.

Induction - notation

For $A \subseteq [n]$ we will denote by E_A expectation taken with respect to all coordinates indexed by A, so that for $f \in \mathcal{H}_n$ the expectation $E_A[f]$ is a **function** depending on coordinates indexed by $[n] \setminus A$.

For $A \subseteq [n]$ we will denote by P_t^A the semigroup action restricted to the coordinates indexed by A. Namely, for $S \subseteq [n]$ we set

$$P_t^A w_S = e^{-|S \cap A| t} w_S$$

and extend P_t^A to a linear operator on $\mathcal{H}_n$.

One can easily check that if $A \cup B = [n]$ and $A \cap B = \emptyset$ then

$$E_A\Big[E_B[f]\Big] = E[f] \text{ and } P_t^A(P_t^B f) = P_t f$$

for every $f \in \mathcal{H}_n$ and $t \geq 0$.

Induction - notation

For $A \subseteq [n]$ we will denote by E_A expectation taken with respect to all coordinates indexed by A, so that for $f \in \mathcal{H}_n$ the expectation $E_A[f]$ is a **function** depending on coordinates indexed by $[n] \setminus A$.

For $A \subseteq [n]$ we will denote by P_t^A the semigroup action restricted to the coordinates indexed by A. Namely, for $S \subseteq [n]$ we set

$$P_t^A w_S = e^{-|S \cap A| t} w_S$$

and extend P_t^A to a linear operator on $\mathcal{H}_n$.

One can easily check that if $A \cup B = [n]$ and $A \cap B = \emptyset$ then

$$E_A\Big[E_B[f]\Big] = E[f] \text{ and } P_t^A(P_t^B f) = P_t f$$

for every $f \in \mathcal{H}_n$ and $t \geq 0$.

Induction - notation

For $A \subseteq [n]$ we will denote by E_A expectation taken with respect to all coordinates indexed by A, so that for $f \in \mathcal{H}_n$ the expectation $E_A[f]$ is a **function** depending on coordinates indexed by $[n] \setminus A$.

For $A \subseteq [n]$ we will denote by P_t^A the semigroup action restricted to the coordinates indexed by A. Namely, for $S \subseteq [n]$ we set

$$P_t^A w_S = e^{-|S \cap A| t} w_S$$

and extend P_t^A to a linear operator on $\mathcal{H}_n$.

One can easily check that if $A \cup B = [n]$ and $A \cap B = \emptyset$ then

$$E_A\Big[E_B[f]\Big] = E[f] \text{ and } P_t^A(P_t^B f) = P_t f$$

for every $f \in \mathcal{H}_n$ and $t \geq 0$.

Induction step

Let $t = -\frac{1}{2}\ln(p-1)$.

$$\|P_t f\|_2 = (E[(P_t f)^2])^{1/2} = \left(E_A\left[E_B[(P_t^B(P_t^A f))^2]\right]\right)^{1/2} \le$$

$$\le \left(E_A\left[(E_B[(P_t^A f)^p])^{2/p}\right]\right)^{1/2} \overset{?}{\le} \left(E_B\left[(E_A[(P_t^A f)^2])^{p/2}\right]\right)^{1/p} \le$$

$$\le \left(E_B\left[E_A[f^p]\right]\right)^{1/p} = (E[f^p])^{1/p} = \|f\|_p,$$

where we have used the induction assumption $\|P_t^B\|_{L^p \to L^2} \le 1$ in the first inequality and the induction assumption $\|P_t^A\|_{L^p \to L^2} \le 1$ in the third inequality.

Now we only need to prove $\overset{?}{\le}$.

Induction step

Let $t = -\frac{1}{2}\ln(p-1)$.

$$\|P_t f\|_2 = (E[(P_t f)^2])^{1/2} = \left(E_A\left[E_B[(P_t^B(P_t^A f))^2]\right]\right)^{1/2} \leq$$

$$\leq \left(E_A\left[(E_B[(P_t^A f)^p])^{2/p}\right]\right)^{1/2} \overset{?}{\leq} \left(E_B\left[(E_A[(P_t^A f)^2])^{p/2}\right]\right)^{1/p} \leq$$

$$\leq \left(E_B\left[E_A[f^p]\right]\right)^{1/p} = (E[f^p])^{1/p} = \|f\|_p,$$

where we have used the induction assumption $\|P_t^B\|_{L^p \to L^2} \leq 1$ in the first inequality and the induction assumption $\|P_t^A\|_{L^p \to L^2} \leq 1$ in the third inequality.

Now we only need to prove $\overset{?}{\leq}$.

Hypercontractivity - main trick

We will prove that

$$\left(E_A\left[(E_B[(P_t^A f)^p])^{2/p}\right]\right)^{1/2} \leq \left(E_B\left[(E_A[(P_t^A f)^2])^{p/2}\right]\right)^{1/p}$$

for every nonnegative $f \in \mathcal{H}_n$.

Let $g = (P_t^A f)^p \geq 0$ and $s = 2/p \geq 1$. We need to prove that

$$\left(E_A\left[(E_B[g])^s\right]\right)^{1/s} \leq E_B\left[(E_A[g^s])^{1/s}\right],$$

which is just a form of the Minkowski inequality:

$$\Big\| E_B[g] \Big\|_{s,A} \leq E_B\left[\|g\|_{s,A}\right]$$

(an easy exercise).

Hypercontractivity - main trick

We will prove that

$$\left(E_A\left[(E_B[(P_t^A f)^p])^{2/p}\right]\right)^{1/2} \le \left(E_B\left[(E_A[(P_t^A f)^2])^{p/2}\right]\right)^{1/p}$$

for every nonnegative $f \in \mathcal{H}_n$.

Let $g = (P_t^A f)^p \ge 0$ and $s = 2/p \ge 1$. We need to prove that

$$\left(E_A\left[(E_B[g])^s\right]\right)^{1/s} \le E_B\left[(E_A[g^s])^{1/s}\right],$$

which is just a form of the Minkowski inequality:

$$\Big\| E_B[g] \Big\|_{s,A} \le E_B\left[\|g\|_{s,A}\right]$$

(an easy exercise).

Subadditivity

There is also a standard method of tensorizing the Poincaré and logarithmic Sobolev inequalities by using the subadditivity of the variance and entropy functionals.

Thus the hypercontractive estimates we have just proved can be also obtained by proving the logarithmic Sobolev inequality on $\{-1, 1\}$ and then deducing it on the discrete cube via subadditivity.

Hint (variational definition of entropy):
for every $f > 0$ we have

$$Ent(f) = \sup\{E[fg]; E[e^g] \leq 1\}.$$

Subadditivity

There is also a standard method of tensorizing the Poincaré and logarithmic Sobolev inequalities by using the subadditivity of the variance and entropy functionals.

Thus the hypercontractive estimates we have just proved can be also obtained by proving the logarithmic Sobolev inequality on $\{-1, 1\}$ and then deducing it on the discrete cube via subadditivity.

Hint (variational definition of entropy):
for every $f > 0$ we have

$$Ent(f) = \sup\{E[fg]; E[e^g] \leq 1\}.$$

Subadditivity

There is also a standard method of tensorizing the Poincaré and logarithmic Sobolev inequalities by using the subadditivity of the variance and entropy functionals.

Thus the hypercontractive estimates we have just proved can be also obtained by proving the logarithmic Sobolev inequality on $\{-1,1\}$ and then deducing it on the discrete cube via subadditivity.

Hint (variational definition of entropy):
for every $f > 0$ we have

$$Ent(f) = \sup\{E[fg]; E[e^g] \le 1\}.$$

Question (Talagrand):
Let $t > 0$ and let μ denote the normalized counting measure on $\{-1,1\}^n$. Does there exist a function $\psi_t : (1,\infty) \to (1,\infty)$ such that $\lim_{u\to\infty} \psi_t(u) = \infty$ and for any positive integer n, any $u > 1$, and every function $f : \{-1,1\}^n \to [0,\infty)$ there is

$$\mu\Big(\Big\{x \in \{-1,1\}^n : (P_t f)(x) > u \cdot E[f]\Big\}\Big) \leq \frac{1}{u\psi_t(u)}?$$

The bound with $\psi(u) \equiv 1$ follows trivially from the fact that $E[P_t f] = E[f]$, and from the Markov-Chebyshev inequality.
The problem is open, some partial affirmative answers have been obtained for its Ornstein-Uhlenbeck analog in the Gaussian setting.

Question (Talagrand):
Let $t > 0$ and let μ denote the normalized counting measure on $\{-1,1\}^n$. Does there exist a function $\psi_t : (1,\infty) \to (1,\infty)$ such that $\lim_{u\to\infty} \psi_t(u) = \infty$ and for any positive integer n, any $u > 1$, and every function $f : \{-1,1\}^n \to [0,\infty)$ there is

$$\mu\Big(\Big\{x \in \{-1,1\}^n : (P_t f)(x) > u \cdot E[f]\Big\}\Big) \leq \frac{1}{u\psi_t(u)}?$$

The bound with $\psi(u) \equiv 1$ follows trivially from the fact that $E[P_t f] = E[f]$, and from the Markov-Chebyshev inequality.
The problem is open, some partial affirmative answers have been obtained for its Ornstein-Uhlenbeck analog in the Gaussian setting.

Brief history of hypercontractivity

The following slides contain a sketch of the history of hypercontractivity. They were prepared as a part of a 2011 presentation, joint with Ryan O'Donnell and Elchanan Mossel.

Symmetric Markov semigroup setting

(Ω, μ) - probability measure space (with some reasonable σ-field)

$\mathcal{H} = L^2(\Omega, \mu)$ - Hilbert space

$L : \mathcal{H} \longrightarrow \mathcal{H}$ - positive semi-definite self-adjoint operator
(in fact, usually defined only on some dense subspace of $\mathcal{H}$),
L1=0; usually L provides a link to a geometric structure of Ω

For $f \in \mathcal{H}$ and $t \geq 0$ let $P_t f = e^{-tL} f$,
i.e. $P_0 f = f$ and $\frac{d}{dt} P_t f = -L P_t f$.

Semigroup property: $P_{t+s} = P_t \circ P_s$ for $t, s \geq 0$.

Assume, additionally, that $P_t : \mathcal{H} \longrightarrow \mathcal{H}$ is positivity preserving:
for $t > 0$, if $f \geq 0$ μ-a.e. then also $P_t f \geq 0$ μ-a.e.

Symmetric Markov semigroup setting

(Ω, μ) - probability measure space (with some reasonable σ-field)

$\mathcal{H} = L^2(\Omega, \mu)$ - Hilbert space

$L : \mathcal{H} \longrightarrow \mathcal{H}$ - positive semi-definite self-adjoint operator
(in fact, usually defined only on some dense subspace of $\mathcal{H}$),
L1=0; usually L provides a link to a geometric structure of Ω

For $f \in \mathcal{H}$ and $t \geq 0$ let $P_t f = e^{-tL} f$,
i.e. $P_0 f = f$ and $\frac{d}{dt} P_t f = -LP_t f$.

Semigroup property: $P_{t+s} = P_t \circ P_s$ for $t, s \geq 0$.

Assume, additionally, that $P_t : \mathcal{H} \longrightarrow \mathcal{H}$ is positivity preserving:
for $t > 0$, if $f \geq 0$ μ-a.e. then also $P_t f \geq 0$ μ-a.e.

Heat semigroups

$(\Omega, \mu) = \left(\{-1,1\}^n, (\frac{1}{2}\delta_{-1} + \frac{1}{2}\delta_1)^{\otimes n}\right)$

$(Lf)(x) = \frac{1}{2}\sum_{y\sim x}\left(f(x) - f(y)\right)$; we sum over neighbours of x, i.e. over y's that differ from x on exactly one coordinate.

Then $P_t w_S = e^{-|S|t} w_S$ for $w_S(x) = \prod_{i\in S} x_i$.

$(\Omega, \mu) = (\mathbf{R}^n, \gamma_n)$; $(Lf)(x) = \langle x, \nabla f(x)\rangle - \Delta f(x)$

Then $(P_t f)(x) = \mathbf{E} f(e^{-t}x + \sqrt{1-e^{-2t}}G)$, where $G \sim \mathcal{N}(0, Id_n)$, with Hermite polynomials as eigenfuntions (the Ornstein-Uhlenbeck semigroup, sort of a heat semigroup on $\mathbf{R}^n$ "compactified" by replacing the non-probabilistic Lebesgue measure λ_n with γ_n).

In both cases $(P_t f)(x) = \mathbf{E} f(X^x(t))$, where $(X^x(t))_{t\geq 0}$, $X^x(0) = x$ is either a symmetric random walk on $\{-1,1\}^n$ with occurences of jumps governed by a Poisson process, or the Ornstein-Uhlenbeck Gaussian process on $\mathbf{R}^n$.

Heat semigroups

$(\Omega, \mu) = \left(\{-1,1\}^n, (\frac{1}{2}\delta_{-1} + \frac{1}{2}\delta_1)^{\otimes n}\right)$
$(Lf)(x) = \frac{1}{2}\sum_{y\sim x}\left(f(x) - f(y)\right)$; we sum over neighbours of x, i.e. over y's that differ from x on exactly one coordinate.

Then $P_t w_S = e^{-|S|t} w_S$ for $w_S(x) = \prod_{i\in S} x_i$.

$(\Omega, \mu) = (\mathbf{R}^n, \gamma_n)$; $(Lf)(x) = \langle x, \nabla f(x)\rangle - \Delta f(x)$

Then $(P_t f)(x) = \mathbf{E} f(e^{-t}x + \sqrt{1-e^{-2t}}G)$, where $G \sim \mathcal{N}(0, Id_n)$, with Hermite polynomials as eigenfuntions (the Ornstein-Uhlenbeck semigroup, sort of a heat semigroup on $\mathbf{R}^n$ "compactified" by replacing the non-probabilistic Lebesgue measure λ_n with γ_n).

In both cases $(P_t f)(x) = \mathbf{E} f(X^x(t))$, where $(X^x(t))_{t\geq 0}$, $X^x(0) = x$ is either a symmetric random walk on $\{-1,1\}^n$ with occurences of jumps governed by a Poisson process, or the Ornstein-Uhlenbeck Gaussian process on $\mathbf{R}^n$.

Heat semigroups

$(\Omega, \mu) = \left(\{-1,1\}^n, (\frac{1}{2}\delta_{-1} + \frac{1}{2}\delta_1)^{\otimes n}\right)$
$(Lf)(x) = \frac{1}{2}\sum_{y\sim x}\left(f(x) - f(y)\right)$; we sum over neighbours of x, i.e. over y's that differ from x on exactly one coordinate.

Then $P_t w_S = e^{-|S|t} w_S$ for $w_S(x) = \prod_{i\in S} x_i$.

$(\Omega, \mu) = (\mathbf{R}^n, \gamma_n)$; $(Lf)(x) = \langle x, \nabla f(x)\rangle - \Delta f(x)$

Then $(P_t f)(x) = \mathbf{E} f(e^{-t}x + \sqrt{1 - e^{-2t}}G)$, where $G \sim \mathcal{N}(0, Id_n)$, with Hermite polynomials as eigenfuntions (the Ornstein-Uhlenbeck semigroup, sort of a heat semigroup on $\mathbf{R}^n$ "compactified" by replacing the non-probabilistic Lebesgue measure λ_n with γ_n).

In both cases $(P_t f)(x) = \mathbf{E} f(X^x(t))$, where $(X^x(t))_{t\geq 0}$, $X^x(0) = x$ is either a symmetric random walk on $\{-1,1\}^n$ with occurences of jumps governed by a Poisson process, or the Ornstein-Uhlenbeck Gaussian process on $\mathbf{R}^n$.

Hypercontractivity

Let $p > q > 1$. We will say that a symmetric Markov semigroup $(P_t)_{t \geq 0}$ is (p,q)-hypercontractive if there exists $t(p,q) > 0$ such that

$$\|P_{t(p,q)} f\|_p \leq \|f\|_q$$

for every $f \in (L^q \cap L^2)(\Omega, \mu)$.

Then the same inequality holds also for every $t \geq t(p,q)$.

Examples: $(\mathbf{R}^n, \gamma_n)$ and $\Big(\{-1,1\}^n, (\frac{1}{2}\delta_{-1} + \frac{1}{2}\delta_1)^{\otimes n}\Big)$ with $t(p,q) = \Big(\ln(p-1) - \ln(q-1)\Big)/2$ (for heat semigroups).

The first example follows from the second one via CLT.

Multiplier notation: $T_\rho = P_{-\ln \rho}$ $(0 < \rho \leq 1)$; then $T_{\rho\eta} = T_\rho \circ T_\eta$.

Hypercontractivity

Let $p > q > 1$. We will say that a symmetric Markov semigroup $(P_t)_{t\geq 0}$ is (p,q)-hypercontractive if there exists $t(p,q) > 0$ such that

$$\|P_{t(p,q)}f\|_p \leq \|f\|_q$$

for every $f \in (L^q \cap L^2)(\Omega, \mu)$.

Then the same inequality holds also for every $t \geq t(p,q)$.

Examples: $(\mathbf{R}^n, \gamma_n)$ and $\Big(\{-1,1\}^n, (\frac{1}{2}\delta_{-1} + \frac{1}{2}\delta_1)^{\otimes n}\Big)$ with $t(p,q) = \Big(\ln(p-1) - \ln(q-1)\Big)/2$ (for heat semigroups).

The first example follows from the second one via CLT.

Multiplier notation: $T_\rho = P_{-\ln\rho}$ $(0 < \rho \leq 1)$; then $T_{\rho\eta} = T_\rho \circ T_\eta$.

Hypercontractivity

Let $p > q > 1$. We will say that a symmetric Markov semigroup $(P_t)_{t \geq 0}$ is (p,q)-hypercontractive if there exists $t(p,q) > 0$ such that

$$\|P_{t(p,q)} f\|_p \leq \|f\|_q$$

for every $f \in (L^q \cap L^2)(\Omega, \mu)$.

Then the same inequality holds also for every $t \geq t(p,q)$.

Examples: $(\mathbf{R}^n, \gamma_n)$ and $\left(\{-1,1\}^n, (\frac{1}{2}\delta_{-1} + \frac{1}{2}\delta_1)^{\otimes n}\right)$ with $t(p,q) = \Big(\ln(p-1) - \ln(q-1)\Big)/2$ (for heat semigroups).

The first example follows from the second one via CLT.

Multiplier notation: $T_\rho = P_{-\ln \rho}$ $(0 < \rho \leq 1)$; then $T_{\rho\eta} = T_\rho \circ T_\eta$.

Hypercontractivity

Let $p > q > 1$. We will say that a symmetric Markov semigroup $(P_t)_{t \geq 0}$ is (p,q)-hypercontractive if there exists $t(p,q) > 0$ such that

$$\|P_{t(p,q)} f\|_p \leq \|f\|_q$$

for every $f \in (L^q \cap L^2)(\Omega, \mu)$.

Then the same inequality holds also for every $t \geq t(p,q)$.

Examples: $(\mathbf{R}^n, \gamma_n)$ and $\left(\{-1,1\}^n, (\frac{1}{2}\delta_{-1} + \frac{1}{2}\delta_1)^{\otimes n}\right)$ with $t(p,q) = \Big(\ln(p-1) - \ln(q-1)\Big)/2$ (for heat semigroups).

The first example follows from the second one via CLT.

Multiplier notation: $T_\rho = P_{-\ln \rho}$ $(0 < \rho \leq 1)$; then $T_{\rho\eta} = T_\rho \circ T_\eta$.

History of the hypercontractive bounds - early period

Nelson 1966, Glimm 1968, Federbush 1969, Segal 1970,
Nelson 1973 - quantum field theory (the Gaussian case)

Gross 1973 - logarithmic Sobolev inequality,
Gross 1975 - Nelson's result via LSI and CLT

Stam 1959 - a Euclidean variant of LSI (information theory)

Harmonic analysis:
Rudin 1960 - similar inequalities for $\mathbf{Z}_n$ instead of $\{-1,1\}^n$
Bonami 1968 - the (4,2)-hypercontractivity (on discrete cube)
Bonami 1970 - the Boolean setting result (general p and q)
Beckner 1975 - as above, for vector-valued functions;
applications to tight Fourier transform norm bounds

Kahn, Kalai, Linial 1988 - KKL theorem

History of the hypercontractive bounds - early period

Nelson 1966, Glimm 1968, Federbush 1969, Segal 1970,
Nelson 1973 - quantum field theory (the Gaussian case)

Gross 1973 - logarithmic Sobolev inequality,
Gross 1975 - Nelson's result via LSI and CLT

Stam 1959 - a Euclidean variant of LSI (information theory)

Harmonic analysis:
Rudin 1960 - similar inequalities for $\mathbf{Z}_n$ instead of $\{-1,1\}^n$
Bonami 1968 - the (4,2)-hypercontractivity (on discrete cube)
Bonami 1970 - the Boolean setting result (general p and q)
Beckner 1975 - as above, for vector-valued functions;
applications to tight Fourier transform norm bounds

Kahn, Kalai, Linial 1988 - KKL theorem

Logarithmic Sobolev inequality

We say that L (as before) satisfies the logaritmic Sobolev inequality with a constant $C > 0$ if

$$\int_\Omega f^2 \ln(f^2)\, d\mu - (\int_\Omega f^2\, d\mu) \ln(\int_\Omega f^2\, d\mu) \le C \cdot \int_\Omega f \cdot Lf\, d\mu$$

for every $f \in Dom(L)$.

Theorem (Gross): L satisfies the logarithmic Sobolev inequality with constant C if and only if for all $p > q > 1$ the semigroup $(P_t)_{t \ge 0}$ generated by L is (p, q)-hypercontractive with

$$t(p, q) = \frac{C}{4} \cdot \Big(\ln(p-1) - \ln(q-1) \Big).$$

Logarithmic Sobolev inequality

We say that L (as before) satisfies the logaritmic Sobolev inequality with a constant $C > 0$ if

$$\int_\Omega f^2 \ln(f^2)\, d\mu - (\int_\Omega f^2\, d\mu) \ln(\int_\Omega f^2\, d\mu) \leq C \cdot \int_\Omega f \cdot Lf\, d\mu$$

for every $f \in Dom(L)$.

Theorem (Gross): L satisfies the logarithmic Sobolev inequality with constant C if and only if for all $p > q > 1$ the semigroup $(P_t)_{t \geq 0}$ generated by L is (p, q)-hypercontractive with

$$t(p, q) = \frac{C}{4} \cdot \Big(\ln(p-1) - \ln(q-1) \Big).$$

Energy-entropy inequality

The left hand side of the LSI,

$$Ent_\mu(f^2) = \int_\Omega f^2 \ln(f^2)\, d\mu - \left(\int_\Omega f^2\, d\mu\right) \ln\left(\int_\Omega f^2\, d\mu\right) \leq C \cdot \int_\Omega f \cdot Lf\, d\mu,$$

is called entropy (here: entropy of f^2 with respect to μ).
It depends only on measure-theoretic properties of f,
i.e. distribution of f on (Ω, μ).

The non-negative shift-invariant quadratic form $\int_\Omega f \cdot Lf\, d\mu$ on the right hand side usually takes into account also the geometry of Ω (which is encoded in L). Expressions of this type are called energy functionals. Another example of energy: $\int_{\mathbf{R}^n} |\nabla f|^2\, d\mu$ (in the Ornstein-Uhlenbeck case it coincides with $\int_{\mathbf{R}^n} f\, Lf\, d\mu$).

Energy-entropy inequality

The left hand side of the LSI,

$$Ent_\mu(f^2) = \int_\Omega f^2 \ln(f^2)\, d\mu - \left(\int_\Omega f^2\, d\mu\right) \ln\left(\int_\Omega f^2\, d\mu\right) \leq C \cdot \int_\Omega f \cdot Lf\, d\mu,$$

is called entropy (here: entropy of f^2 with respect to μ).
It depends only on measure-theoretic properties of f,
i.e. distribution of f on (Ω, μ).

The non-negative shift-invariant quadratic form $\int_\Omega f \cdot Lf\, d\mu$ on the right hand side usually takes into account also the geometry of Ω (which is encoded in L). Expressions of this type are called energy functionals. Another example of energy: $\int_{\mathbf{R}^n} |\nabla f|^2\, d\mu$ (in the Ornstein-Uhlenbeck case it coincides with $\int_{\mathbf{R}^n} f\, Lf\, d\mu$).

Entropy

Lazare Carnot 1803, Sadi Carnot 1824
- first insights into the second law of thermodynamics

Second half of 19th century:
Clausius (thermodynamic definition),
Boltzmann (statistical definition), Gibbs, Maxwell

Carathéodory 1909 - links to irreversibility

Schrödinger, von Neumann, first half of 20th century
- in quantum mechanics

Information theory:
Shannon 1948 - information entropy,
Kullback and Leibler 1951 - relative entropy

Dynamical systems - Kolmogorov-Sinai and topological entropy
(middle of 20th century)

Entropy

Lazare Carnot 1803, Sadi Carnot 1824
- first insights into the second law of thermodynamics

Second half of 19th century:
Clausius (thermodynamic definition),
Boltzmann (statistical definition), Gibbs, Maxwell

Carathéodory 1909 - links to irreversibility

Schrödinger, von Neumann, first half of 20th century
- in quantum mechanics

Information theory:
Shannon 1948 - information entropy,
Kullback and Leibler 1951 - relative entropy

Dynamical systems - Kolmogorov-Sinai and topological entropy
(middle of 20th century)

Concentration of measure phenomenon

A metric space (Ω, ρ) equipped with a Borel probability measure μ enjoys concentration if:

Every 1-Lipschitz (i.e. $|f(x) - f(y)| \leq \rho(x, y)$ for all $x, y \in \Omega$) function $f : \Omega \longrightarrow \mathbf{R}$ is integrable, and it takes values far from its mean $\int_\Omega f \, d\mu$ only with a very small (uniformly with respect to choice of f) probability.

Or, equivalently, for $s > 0$ the concentration function

$$\alpha(s) = \sup_{A \subset \Omega : \mu(A) = 1/2} \mu\Big(\{x \in \Omega : dist_\rho(x, A) > s\}\Big)$$

decays quickly when s grows.

Lévy - sphere S^{n-1} with geodesic distance and uniform measure
Milman 1971 - convex geometry, proof of Dvoretzky theorem
Sudakov & Tsirelson, and Borell 1974/75 - Gaussian isoperimetry

Concentration of measure phenomenon

A metric space (Ω, ρ) equipped with a Borel probability measure μ enjoys concentration if:

Every 1-Lipschitz (i.e. $|f(x) - f(y)| \leq \rho(x, y)$ for all $x, y \in \Omega$) function $f : \Omega \longrightarrow \mathbf{R}$ is integrable, and it takes values far from its mean $\int_\Omega f \, d\mu$ only with a very small (uniformly with respect to choice of f) probability.

Or, equivalently, for $s > 0$ the concentration function

$$\alpha(s) = \sup_{A \subset \Omega : \mu(A) = 1/2} \mu\Big(\{x \in \Omega : dist_\rho(x, A) > s\}\Big)$$

decays quickly when s grows.

Lévy - sphere S^{n-1} with geodesic distance and uniform measure
Milman 1971 - convex geometry, proof of Dvoretzky theorem
Sudakov & Tsirelson, and Borell 1974/75 - Gaussian isoperimetry

Concentration of measure phenomenon

A metric space (Ω, ρ) equipped with a Borel probability measure μ enjoys concentration if:

Every 1-Lipschitz (i.e. $|f(x) - f(y)| \leq \rho(x, y)$ for all $x, y \in \Omega$) function $f : \Omega \longrightarrow \mathbf{R}$ is integrable, and it takes values far from its mean $\int_\Omega f \, d\mu$ only with a very small (uniformly with respect to choice of f) probability.

Or, equivalently, for $s > 0$ the concentration function

$$\alpha(s) = \sup_{A \subset \Omega : \mu(A) = 1/2} \mu\Big(\{x \in \Omega : dist_\rho(x, A) > s\}\Big)$$

decays quickly when s grows.

Lévy - sphere S^{n-1} with geodesic distance and uniform measure
Milman 1971 - convex geometry, proof of Dvoretzky theorem
Sudakov & Tsirelson, and Borell 1974/75 - Gaussian isoperimetry

Concentration via functional inequalities

Herbst (unpublished letter to Gross, mentioned in early 1980's):
LSI-type inequality $\forall_f \;\; Ent_\mu(f^2) \leq C \int_{\mathbf{R}^n} |\nabla f|^2 \, d\mu$
implies Gaussian concentration, $\alpha(s) \leq c_1 e^{-c_2 s^2}$ for ρ Euclidean

Gromov and Milman 1983:
Poincaré inequality $\forall_f \;\; \int_{\mathbf{R}^n} f^2 \, d\mu - \left(\int_{\mathbf{R}^n} f \, d\mu \right)^2 \leq C \int_{\mathbf{R}^n} |\nabla f|^2 \, d\mu$
implies exponential concentration, $\alpha(s) \leq c_1 e^{-c_2 s}$ for ρ Euclidean

Talagrand - concentration inequalities since late 1980s
Ledoux, Talagrand 1991 - *Probability in Banach spaces* book
Ledoux - since 1990s develops modern functional techniques
Bobkov 1997 - functional isoperimetry on the discrete cube
Beckner, Latała et al. 1990s - between Poincaré and log-Sobolev

Concentration via functional inequalities

Herbst (unpublished letter to Gross, mentioned in early 1980's):
LSI-type inequality $\forall_f \;\; Ent_\mu(f^2) \le C \int_{\mathbf{R}^n} |\nabla f|^2 \, d\mu$
implies Gaussian concentration, $\alpha(s) \le c_1 e^{-c_2 s^2}$ for ρ Euclidean

Gromov and Milman 1983:
Poincaré inequality $\forall_f \;\; \int_{\mathbf{R}^n} f^2 \, d\mu - \left(\int_{\mathbf{R}^n} f \, d\mu \right)^2 \le C \int_{\mathbf{R}^n} |\nabla f|^2 \, d\mu$
implies exponential concentration, $\alpha(s) \le c_1 e^{-c_2 s}$ for ρ Euclidean

Talagrand - concentration inequalities since late 1980s
Ledoux, Talagrand 1991 - *Probability in Banach spaces* book
Ledoux - since 1990s develops modern functional techniques
Bobkov 1997 - functional isoperimetry on the discrete cube
Beckner, Latała et al. 1990s - between Poincaré and log-Sobolev

Concentration via functional inequalities

Herbst (unpublished letter to Gross, mentioned in early 1980's):
LSI-type inequality $\forall_f \;\; Ent_\mu(f^2) \leq C \int_{\mathbf{R}^n} |\nabla f|^2 \, d\mu$
implies Gaussian concentration, $\alpha(s) \leq c_1 e^{-c_2 s^2}$ for ρ Euclidean

Gromov and Milman 1983:
Poincaré inequality $\forall_f \;\; \int_{\mathbf{R}^n} f^2 \, d\mu - \left(\int_{\mathbf{R}^n} f \, d\mu \right)^2 \leq C \int_{\mathbf{R}^n} |\nabla f|^2 \, d\mu$
implies exponential concentration, $\alpha(s) \leq c_1 e^{-c_2 s}$ for ρ Euclidean

Talagrand - concentration inequalities since late 1980s
Ledoux, Talagrand 1991 - *Probability in Banach spaces* book
Ledoux - since 1990s develops modern functional techniques
Bobkov 1997 - functional isoperimetry on the discrete cube
Beckner, Latała et al. 1990s - between Poincaré and log-Sobolev

Reverse hypercontractivity

Borell 1982 - reverse hypercontractivity for the discrete cube (Gaussian Ornstein-Uhlenbeck version follows via CLT):
For $q < p < 1$ and every $f : \{-1, 1\}^n \longrightarrow (0, \infty)$
there is

$$\|P_t f\|_q \geq \|f\|_p$$

for all $t \geq t(p, q) = \Big(\ln(1-q) - \ln(1-p)\Big)/2.$

In recent years many applications and new extensions:

- Boolean analysis, theoretical computer science
- mixing estimates for Markov processes

Reverse hypercontractivity

Borell 1982 - reverse hypercontractivity for the discrete cube (Gaussian Ornstein-Uhlenbeck version follows via CLT):
For $q < p < 1$ and every $f : \{-1,1\}^n \longrightarrow (0,\infty)$ there is

$$\|P_t f\|_q \geq \|f\|_p$$

for all $t \geq t(p,q) = \Big(\ln(1-q) - \ln(1-p)\Big)/2.$

In recent years many applications and new extensions:

- Boolean analysis, theoretical computer science
- mixing estimates for Markov processes

Borell, Krakowiak, Kwapień, Szulga, Woyczyński - since 1980s develop probabilistic Banach space version of hypercontractivity:

For $p > q > 1$ and $\sigma \in (0, 1)$ we say that a random vector X is (p, q, σ)-hypercontractive if

$$\left(\mathbf{E}\|x + \sigma \cdot X\|^p\right)^{1/p} \le \left(\mathbf{E}\|x + X\|^q\right)^{1/q}$$

for every vector x.

Among consequences: Khinchine-Kahane type inequalities for sums of independent random vectors and vector-valued chaoses, with important conseqences for stochastic integration theory

Combined with revival of Lindeberg's technique of proof of CLT yields universality principles (MOO 2005/10).

Probability in Banach spaces approach to hypercontractivity

Borell, Krakowiak, Kwapień, Szulga, Woyczyński - since 1980s develop probabilistic Banach space version of hypercontractivity:

For $p > q > 1$ and $\sigma \in (0, 1)$ we say that a random vector X is (p, q, σ)-hypercontractive if

$$\left(\mathbf{E}\|x + \sigma \cdot X\|^p\right)^{1/p} \leq \left(\mathbf{E}\|x + X\|^q\right)^{1/q}$$

for every vector x.

Among consequences: Khinchine-Kahane type inequalities for sums of independent random vectors and vector-valued chaoses, with important conseqences for stochastic integration theory

Combined with revival of Lindeberg's technique of proof of CLT yields universality principles (MOO 2005/10).

Probability in Banach spaces approach to hypercontractivity

Borell, Krakowiak, Kwapień, Szulga, Woyczyński - since 1980s develop probabilistic Banach space version of hypercontractivity:

For $p > q > 1$ and $\sigma \in (0,1)$ we say that a random vector X is (p,q,σ)-hypercontractive if

$$\left(\mathbf{E}\|x+\sigma\cdot X\|^p\right)^{1/p} \leq \left(\mathbf{E}\|x+X\|^q\right)^{1/q}$$

for every vector x.

Among consequences: Khinchine-Kahane type inequalities for sums of independent random vectors and vector-valued chaoses, with important conseqences for stochastic integration theory

Combined with revival of Lindeberg's technique of proof of CLT yields universality principles (MOO 2005/10).

MOO

In the slides that follow, there are sketched some basic ideas of a joint paper with Elchanan Mossel and Ryan O'Donnell (MOO), dealing with the noise stability and a related invariance principle.

Notation

Discrete cube (with a normalized counting measure): $\{-1,1\}^n$

Boolean function:

$$f : \{-1,1\}^n \to \{-1,1\}$$

Walsh functions: for $x \in \{-1,1\}^n$ and $S \subseteq [n]$,

$$w_S(x) = \prod_{i \in S} x_i$$

Fourier expansion: $f = \sum_{S \subseteq [n]} \hat{f}(S) w_S$

Influence of the i-th variable:

$$Inf_i(f) = E_x[Var_{x_i}[f(x)]] = \sum_{S \ni i} \hat{f}(S)^2$$

Noise stability

Noise stability of f: For $\rho \in [0,1]$, let

$$S_\rho(f) = \sum_{S \subseteq [n]} \rho^{|S|} \hat{f}(S)^2.$$

Let $x \in \{-1,1\}^n$ be chosen uniformly and let $\theta_1, \theta_2, \ldots, \theta_n$ be a sequence of independent random variables with

$$P[\theta_i = -1] = (1-\rho)/2, \;\; P[\theta_i = 1] = (1+\rho)/2,$$

independent of x. Let $y \in \{-1,1\}^n$ be given by $y_i = x_i \theta_i$. Then

$$E[f(x)f(y)] = S_\rho(f).$$

Thus: $\rho \simeq 0$ – great noise, $\rho \simeq 1$ – small noise.

Usually we assume $E[f] = 0$, $E[f^2] = 1$.
Then $S_0(f) = 0$ and $S_1(f) = 1$.

Majority (is Stablest)

Majority function: For n odd, let $Maj_n(x) = sgn(x_1 + \ldots + x_n)$. Then

$$\lim_{n \to \infty} S_\rho(Maj_n) = \frac{2}{\pi} \arcsin \rho.$$

Majority is Stablest conjecture:
For any $\rho \in [0,1]$ there is

$$\lim_{\tau \to 0^+} \sup_f S_\rho(f) = \frac{2}{\pi} \arcsin \rho,$$

where the supremum is taken over all Boolean f with $E[f] = 0$ and having all influences less than τ : $\max_i Inf_i(f) \leq \tau$ (and over all n).

Majority (is Stablest)

Majority function: For n odd, let $Maj_n(x) = sgn(x_1 + \ldots + x_n)$. Then

$$\lim_{n\to\infty} S_\rho(Maj_n) = \frac{2}{\pi}\arcsin\rho.$$

Majority is Stablest conjecture:
For any $\rho \in [0,1]$ there is

$$\lim_{\tau\to 0^+} \sup_f S_\rho(f) = \frac{2}{\pi}\arcsin\rho,$$

where the supremum is taken over all Boolean f with $E[f] = 0$ and having all influences less than τ : $\max_i Inf_i(f) \leq \tau$ (and over all n).

It Ain't Over Till It's Over conjecture [Kalai]:

Let $\rho \in [0,1)$. Let $x \in \{-1,1\}^n$ be chosen randomly with uniform measure. Each of its coordinates is revealed with probability ρ (independently for each i and independently of x). Then

$$\sup_f P\Big[E[f \,|\, rev.\, x_i's] > 1-\delta\Big] \xrightarrow{\delta,\tau\to 0^+} 0,$$

where the supremum is taken over all Boolean f with $E[f]=0$ and having all influences less than τ (and over all n).

For fixed δ, the limit (as $\tau \to 0$) is roughly $\delta^{(1-\rho)/\rho}$
– the asymptotics one gets for $f = Maj_n$ when n is large.

It Ain't Over Till It's Over conjecture [Kalai]:

Let $\rho \in [0, 1)$. Let $x \in \{-1, 1\}^n$ be chosen randomly with uniform measure. Each of its coordinates is revealed with probability ρ (independently for each i and independently of x). Then

$$\sup_f P\Big[E[f \,|\, rev.\, x_i's] > 1 - \delta\Big] \overset{\delta,\tau \to 0^+}{\longrightarrow} 0,$$

where the supremum is taken over all Boolean f with $E[f] = 0$ and having all influences less than τ (and over all n).

For fixed δ, the limit (as $\tau \to 0$) is roughly $\delta^{(1-\rho)/\rho}$
– the asymptotics one gets for $f = Maj_n$ when n is large.

Invariance Principle:
Given a multilinear polynomial (chaos) of bounded degree in independent random variables, one can replace them by independent $\mathcal{N}(0,1)$ Gaussians without changing the polynomial's distribution too much, under some **reasonable assumptions**.

Classical examples:
the Central Limit Theorem
the Berry-Esséen inequality

Related results:
Rotar' (1975, 1979)
Chatterjee (2004)

Invariance Principle:
Given a multilinear polynomial (chaos) of bounded degree in independent random variables, one can replace them by independent $\mathcal{N}(0,1)$ Gaussians without changing the polynomial's distribution too much, under some **reasonable assumptions**.

Classical examples:
the Central Limit Theorem
the Berry-Esséen inequality

Related results:
Rotar' (1975, 1979)
Chatterjee (2004)

Invariance Principle

Invariance Principle:
Given a multilinear polynomial (chaos) of bounded degree in independent random variables, one can replace them by independent $\mathcal{N}(0,1)$ Gaussians without changing the polynomial's distribution too much, under some **reasonable assumptions**.

Classical examples:
the Central Limit Theorem
the Berry-Esséen inequality

Related results:
Rotar' (1975, 1979)
Chatterjee (2004)

Reasonable assumptions:

$E[X_i] = 0$, $E[X_i^2] = 1$, small influences
– small coefficients do not suffice:

$$r_{n+1}(r_1 + r_2 + \ldots + r_n)/\sqrt{n} \xrightarrow{D} \mathcal{N}(0,1)$$

as $n \to \infty$ but

$$g_{n+1}(g_1 + \ldots + g_n)/\sqrt{n} \stackrel{D}{=} g_1 g_2 \stackrel{D}{\neq} \mathcal{N}(0,1).$$

Also: bounds on higher moments, e.g.,

$$\sup_i E[|X_i|^3] < \infty.$$

An obstacle: Dependence on degree.

What we want to measure

Closeness in distribution:

For cumulative distribution functions F and G, Lévy's metric is defined by

$$\rho_L(F,G) = \inf\{a > 0 : \forall_{t \in \mathbf{R}} F(t-a) - a \leq G(t) \leq F(t+a) + a\}.$$

Generalization of approach: sequences of independent orthonormal ensembles instead of independent random variables.

Motivation: Finite probability spaces

What we want to measure

Closeness in distribution:

For cumulative distribution functions F and G, Lévy's metric is defined by

$$\rho_L(F,G) = \inf\{a > 0 : \forall_{t\in\mathbf{R}} F(t-a) - a \leq G(t) \leq F(t+a) + a\}.$$

Generalization of approach: sequences of independent orthonormal ensembles instead of independent random variables.

Motivation: Finite probability spaces

Orthonormal ensemble

Orthonormal ensemble:
$\mathbf{X}_i = \{X_{i,0} \equiv 1, X_{i,1}, \ldots, X_{i,m_i}\}$

Examples:
Let (Ω, μ) be a finite probability space. Any orthonormal basis in $L^2(\Omega, \mu)$ to which the constant 1 belongs is OK. Also:
$\mathbf{G}_i = \{G_{i,0} \equiv 1, G_{i,1}, G_{i,2}, \ldots\}$, where $G_{i,j}$'s are i.i.d. $\mathcal{N}(0,1)$.

Sequence of independent ensembles:
$\mathbf{X} = (\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n)$

For any sequence of (measurable) functions $f_1, f_2, \ldots, f_n$ we want random variables $f_i(X_{i,1}, X_{i,2}, \ldots, X_{i,m_i})$ $(i = 1, 2, \ldots, n)$ to be independent.

Notation: $\|Z\|_p = (E[|Z|^p])^{1/p}$.

Hypercontractivity: Let $p > q > 1$. We will say that a real r.v. X is (p,q)-hypercontractive with constant $\eta \in (0,1)$ if
$\forall_{x,y \in \mathbf{R}}\ \|x + \eta y X\|_p \leq \|x + yX\|_q$ or, equivalently,
$\forall_{x \in \mathbf{R}}\ \|x + \eta X\|_p \leq \|x + X\|_q$.

We will say that an orthonormal ensemble $\mathbf{X_i} = \{X_{i,0} \equiv 1, X_{i,1}, \ldots, X_{i,m_i}\}$ is (p,q,η)-hypercontractive if for any sequence of reals $(a_1, \ldots, a_{m_i})$ a random variable $a_1 X_{i,1} + a_2 X_{i,2} + \ldots + a_{m_i} X_{i,m_i}$ is (p,q)-hypercontractive with constant η.

Hypercontractivity

Notation: $\|Z\|_p = (E[|Z|^p])^{1/p}$.

Hypercontractivity: Let $p > q > 1$. We will say that a real r.v. X is (p, q)-hypercontractive with constant $\eta \in (0,1)$ if
$\forall_{x,y \in \mathbf{R}}\ \|x + \eta y X\|_p \leq \|x + yX\|_q$ or, equivalently,
$\forall_{x \in \mathbf{R}}\ \|x + \eta X\|_p \leq \|x + X\|_q$.

We will say that an orthonormal ensemble $\mathbf{X_i} = \{X_{i,0} \equiv 1, X_{i,1}, \ldots, X_{i,m_i}\}$ is (p, q, η)-hypercontractive if for any sequence of reals $(a_1, \ldots, a_{m_i})$ a random variable $a_1 X_{i,1} + a_2 X_{i,2} + \ldots + a_{m_i} X_{i,m_i}$ is (p, q)-hypercontractive with constant η.

A sequence of independent orthonormal ensembles is called (p, q, η)-hypercontractive if all of the ensembles are (p, q, η)-hypercontractive. Hence any union of two independent (p, q, η)-hypercontractive sequences of independent orthonormal ensembles $\mathbf{X} \cup \mathbf{Y} = (\mathbf{X}_1, \ldots, \mathbf{X}_n, \mathbf{Y}_1, \ldots, \mathbf{Y}_m)$ is also a (p, q, η)-hypercontractive sequence of independent orthonormal ensembles.

Hypercontractive estimates

Proposition [Wolff]: Let (Ω, μ) be a finite probability space with

$$\alpha = \min_{x \in \Omega : \mu(x) > 0} \mu(x) \leq 1/2.$$

Then any orthonormal ensemble defined on (Ω, μ) is $(3, 2, \eta)$-hypercontractive, $\eta = ((\alpha^{-1} - 1)^{1/3} + (\alpha^{-1} - 1)^{-1/3})^{-1/2}$, i.e., $\eta \sim \alpha^{1/6}$ as $\alpha \to 0$. No better bound in terms of α is possible.

[Nelson/Bonami/Beckner/Gross] Theorem: An orthonormal Gaussian ensemble is $(p, q, \sqrt{q-1}/\sqrt{p-1})$-hypercontractive for any $p > q > 1$, as well as an orthonormal Rademacher ensemble (this way it is proved, then via CLT). The constant $\sqrt{q-1}/\sqrt{p-1}$ is optimal.

Hypercontractive estimates

Proposition [Wolff]: Let (Ω, μ) be a finite probability space with

$$\alpha = \min_{x \in \Omega: \mu(x) > 0} \mu(x) \le 1/2.$$

Then any orthonormal ensemble defined on (Ω, μ) is $(3, 2, \eta)$-hypercontractive, $\eta = ((\alpha^{-1} - 1)^{1/3} + (\alpha^{-1} - 1)^{-1/3})^{-1/2}$, i.e., $\eta \sim \alpha^{1/6}$ as $\alpha \to 0$. No better bound in terms of α is possible.

[Nelson/Bonami/Beckner/Gross] Theorem: An orthonormal Gaussian ensemble is $(p, q, \sqrt{q-1}/\sqrt{p-1})$-hypercontractive for any $p > q > 1$, as well as an orthonormal Rademacher ensemble (this way it is proved, then via CLT). The constant $\sqrt{q-1}/\sqrt{p-1}$ is optimal.

Auxiliary notation

Notation:
multi-index: $\sigma = (\sigma_1, \ldots, \sigma_n) \in \boldsymbol{N}^n$
degree: $|\sigma| = |\{i \in [n] : \sigma_i > 0\}|$
monomial: $x_\sigma = \prod_{i=1}^n x_{i,\sigma_i}$
multilinear polynomial: $Q(x) = \sum_\sigma c_\sigma x_\sigma$
$deg(Q) = \max_{\sigma : c_\sigma \neq 0} |\sigma|$

Replacing $x_{i,j}'$s by $X_{i,j}'s$ from a sequence of independent orthonormal ensembles $\mathbf{X} = (\mathbf{X}_1, \ldots, \mathbf{X}_n)$, we obtain a random variable $Q(\mathbf{X})$.

Contraction: $(T_\eta Q)(x) := \sum_\sigma \eta^{|\sigma|} c_\sigma x_\sigma$, so that $T_{\eta\xi} = T_\eta T_\xi$.
Hence $P_t := T_{e^{-t}}$ is a semigroup of contractions $(t \geq 0)$.

Auxiliary notation

Notation:
multi-index: $\sigma = (\sigma_1, \ldots, \sigma_n) \in \boldsymbol{N}^n$
degree: $|\sigma| = |\{i \in [n] : \sigma_i > 0\}|$
monomial: $x_\sigma = \prod_{i=1}^n x_{i,\sigma_i}$
multilinear polynomial: $Q(x) = \sum_\sigma c_\sigma x_\sigma$
$deg(Q) = \max_{\sigma : c_\sigma \neq 0} |\sigma|$

Replacing $x'_{i,j}$s by $X'_{i,j}s$ from a sequence of independent orthonormal ensembles $\mathbf{X} = (\mathbf{X}_1, \ldots, \mathbf{X}_n)$, we obtain a random variable $Q(\mathbf{X})$.

Contraction: $(T_\eta Q)(x) := \sum_\sigma \eta^{|\sigma|} c_\sigma x_\sigma$, so that $T_{\eta\xi} = T_\eta T_\xi$.
Hence $P_t := T_{e^{-t}}$ is a semigroup of contractions $(t \geq 0)$.

Auxiliary notation

Notation:
multi-index: $\sigma = (\sigma_1, \ldots, \sigma_n) \in \boldsymbol{N}^n$
degree: $|\sigma| = |\{i \in [n] : \sigma_i > 0\}|$
monomial: $x_\sigma = \prod_{i=1}^n x_{i,\sigma_i}$
multilinear polynomial: $Q(x) = \sum_\sigma c_\sigma x_\sigma$
$deg(Q) = \max_{\sigma : c_\sigma \neq 0} |\sigma|$

Replacing $x'_{i,j}$s by $X'_{i,j}s$ from a sequence of independent orthonormal ensembles $\mathbf{X} = (\mathbf{X}_1, \ldots, \mathbf{X}_n)$, we obtain a random variable $Q(\mathbf{X})$.

Contraction: $(T_\eta Q)(x) := \sum_\sigma \eta^{|\sigma|} c_\sigma x_\sigma$, so that $T_{\eta\xi} = T_\eta T_\xi$.
Hence $P_t := T_{e^{-t}}$ is a semigroup of contractions $(t \geq 0)$.

Hypercontractive lemma

Lemma: If a sequence of independent orthonormal ensembles $\mathbf{X}$ is (p,q,η)-hypercontractive then $\|(T_\eta Q)(\mathbf{X})\|_p \le \|Q(\mathbf{X})\|_q$.
(Proof: induction on the length of $\mathbf{X}$.)

Corollary: If a sequence of independent orthonormal ensembles $\mathbf{X}$ is $(p,2,\eta)$-hypercontractive then

$$\|Q(\mathbf{X})\|_p \le \eta^{-deg(Q)}\|Q(\mathbf{X})\|_2,$$

since summands in $R(\mathbf{X}) = \sum_\sigma \eta^{-|\sigma|} c_\sigma X_\sigma$ are orthogonal and $Q(x) = (T_\eta R)(x)$, so that we can use the above Lemma for R. Note that $deg(R) = deg(Q)$.

Hypercontractive lemma

Lemma: If a sequence of independent orthonormal ensembles $\mathbf{X}$ is (p,q,η)-hypercontractive then $\|(T_\eta Q)(\mathbf{X})\|_p \leq \|Q(\mathbf{X})\|_q$.
(Proof: induction on the length of $\mathbf{X}$.)

Corollary: If a sequence of independent orthonormal ensembles $\mathbf{X}$ is $(p,2,\eta)$-hypercontractive then

$$\|Q(\mathbf{X})\|_p \leq \eta^{-deg(Q)}\|Q(\mathbf{X})\|_2,$$

since summands in $R(\mathbf{X}) = \sum_\sigma \eta^{-|\sigma|} c_\sigma X_\sigma$ are orthogonal and $Q(x) = (T_\eta R)(x)$, so that we can use the above Lemma for R. Note that $deg(R) = deg(Q)$.

Proof of the Invariance Principle:

Choose $\Phi : \mathbf{R} \to \mathbf{R}$ with $|\Phi'''|$ uniformly bounded. Replace each of the $X'_{i,j}s$ ($i \in [n]$, $j \geq 1$), step by step, by i.i.d. $\mathcal{N}(0,1)$ r.v.'s $G_{i,j}$. Prove, by the Taylor theorem and comparison of moments, that the difference between $E[\Phi(Q(\mathbf{X}))]$ and $E[\Phi(Q(\mathbf{G}))]$ is small since the change is controlled in each step by $const(d,\eta) \cdot Inf_i(Q)^{3/2}$ and $\sum_i Inf_i(Q)^{3/2} \leq \sum_i Inf_i(Q) \cdot \sqrt{\max_i Inf_i(Q)} \leq d \cdot \sqrt{\max_i Inf_i(Q)}$. By an appropriate choice of Φ one can prove that distributions of $Q(\mathbf{X})$ and $Q(\mathbf{G})$ are close to each other (in Lévy's metric or some other sense).

Proof of Majority Is Stablest:

Note that $S_{\rho_1\rho_2}(f) = \| T_{\sqrt{\rho_1}}((T_{\sqrt{\rho_2}} f))\|_2^2$ for $\rho_1, \rho_2 > 0$, so that we can use part of ρ to kill high frequencies and obtain essentially a polynomial of bounded degree. Then we can transfer the problem to the Gaussian setting, where we can use an old theorem due to Borell to obtain the result on the Gauss space and then come back to the discrete cube setting (we use the fact that the heat semigroup on the cube and the Ornstein-Uhlenbeck semigroup modify $Q's$ coefficients in the same way).

Proof of It Ain't Over Till It's Over:

Let $\mathbf{X}$ denote a sequence of Rademacher ensembles. Given $\rho \in (0,1)$ let $V_1, \ldots, V_n$ be an i.i.d. sequence independent of $\mathbf{X}$; $P[V_i = 0] = 1 - \rho$, $P[V_i = 1] = \rho$. Then define a new sequence of orthonormal ensembles $\mathbf{X}^{(\rho)}$ by $X_{i,0} \equiv 1$ and $X_{i,1}^{(\rho)} = \rho^{-1/2} V_i X_{i,1}$ for $i \in [n]$. The i's for which $V_i = 1$ can be understood as revealed votes.

The key observation:
$E(Q(\mathbf{X}) \mid \textit{revealed } i's)$ has the same distribution as $(T_{\sqrt{\rho}} Q)(\mathbf{X}^{(\rho)})$, i.e., close to the distribution of $(T_{\sqrt{\rho}} Q)(\mathbf{G})$.

Proof of It Ain't Over Till It's Over:

Let $\mathbf{X}$ denote a sequence of Rademacher ensembles. Given $\rho \in (0,1)$ let $V_1, \ldots, V_n$ be an i.i.d. sequence independent of $\mathbf{X}$; $P[V_i = 0] = 1 - \rho$, $P[V_i = 1] = \rho$. Then define a new sequence of orthonormal ensembles $\mathbf{X}^{(\rho)}$ by $X_{i,0} \equiv 1$ and $X_{i,1}^{(\rho)} = \rho^{-1/2} V_i X_{i,1}$ for $i \in [n]$. The i's for which $V_i = 1$ can be understood as revealed votes.

The key observation:
$E(Q(\mathbf{X}) \mid \textit{revealed } i's)$ has the same distribution as $(T_{\sqrt{\rho}}Q)(\mathbf{X}^{(\rho)})$, i.e., close to the distribution of $(T_{\sqrt{\rho}}Q)(\mathbf{G})$.

Note that high frequencies in $T_{\sqrt{\rho}}Q$ are already "killed". Then we use the knowledge than in the Gauss space the contractions T_ρ applied to mean zero functions with values in $[-a, a]$ "push them away" from the ends of the interval, letting them stay close to these ends with a small probability only. This basically ends the proof.

Product placement

Theorem [Fund. Math. 1996] in search of applications:

Let $f : \{-1,1\}^n \to \mathbf{R}^k$ and assume that for every $x, y \in \{-1,1\}^n$ there is $\|f(x) - f(y)\| \leq d(x,y)$, where $\|\cdot\|$ is some norm on $\mathbf{R}^k$ and d denotes the Hamming metric on the discrete cube. Then there exists some $z \in \{-1,1\}^n$ such that

$$\|f(z) - f(-z)\| \leq \min(k, n).$$

Moreover, if the norm $\|\cdot\|$ is Euclidean then the bound $\min(k,n)$ may be strenghtened to $\min(\sqrt{k}, \sqrt{n})$.

Product placement

Theorem [Fund. Math. 1996] in search of applications:

Let $f : \{-1,1\}^n \to \mathbf{R}^k$ and assume that for every $x, y \in \{-1,1\}^n$ there is $\|f(x) - f(y)\| \leq d(x,y)$, where $\|\cdot\|$ is some norm on $\mathbf{R}^k$ and d denotes the Hamming metric on the discrete cube. Then there exists some $z \in \{-1,1\}^n$ such that

$$\|f(z) - f(-z)\| \leq \min(k,n).$$

Moreover, if the norm $\|\cdot\|$ is Euclidean then the bound $\min(k,n)$ may be strenghtened to $\min(\sqrt{k}, \sqrt{n})$.