FOURIER TRANSFORM ON THE HYPERCUBE

JONAS SJÖSTRAND

Definitions

Let \( F \) be the real vector space of real functions on \( 2^{[n]} \) (the set of subsets of \( [n] = \{1, \ldots, n\} \)). Define an inner product on \( F \) by

\[
\langle f, g \rangle = \frac{1}{2^n} \sum_{A \subseteq [n]} f(A)g(A).
\]

For \( S \subseteq [n] \), let \( x^S \in F \) be the function \( x^S(A) = (-1)^{|A \cap S|} \). Now we compute

\[
\langle x^S, x^T \rangle = \frac{1}{2^n} \sum_{A \subseteq [n]} (-1)^{|A \cap S|}(-1)^{|A \cap T|} = \frac{1}{2^n} \sum_{A \subseteq [n]} (-1)^{|A \cap (S \triangle T)|} = \delta(S, T),
\]

where \( S \triangle T \) denotes the symmetric difference. There are as many \( x^S \)'s as there are \( \mathbb{R} \)-dimensions of \( F \), namely \( 2^n \). Thus, the \( x^S \)'s form an orthonormal basis for \( F \).

For \( f \in F \) we define the Fourier transform \( \hat{f} \in F \) of \( f \) as

\[
\hat{f}(S) = \langle f, x^S \rangle.
\]

The relation to the Fourier transform on \( \mathbb{R}^n \)

The usual inverse Fourier transform on \( \mathbb{R}^n \) is

\[
f(x) = \int_{\mathbb{R}^n} \hat{f}(s)e^{ix \cdot s} \, ds.
\]

What is the relation between the basis functions \( e^{ix \cdot s} \) and our basis functions \( x^S \)?

Identify the hypercube \( 2^{[n]} \) with the vector space \( \mathbb{Z}_2^n \) over \( \mathbb{Z}_2 \). A subset \( A \subseteq [n] \) corresponds to the element \( A = (a_1, \ldots, a_n) \in \mathbb{Z}_2^n \) where \( a_i = 1 \) if \( i \in A \) and \( a_i = 0 \) if \( i \notin A \). Define a scalar product on \( \mathbb{Z}_2^n \) by

\[
(a_1, \ldots, a_n) \cdot (b_1, \ldots, b_n) = a_1b_1 + \cdots + a_nb_n
\]

where the operations are performed in \( \mathbb{Z}_2 \). With this notation we get

\[
x^S(A) = (-1)^{|A \cap S|} = (-1)^A \cdot S = e^{i\pi(A \cdot S)}
\]

and the relation to \( e^{ix \cdot s} \) is evident.

Note that the functions \( A \mapsto A \cdot S \) for \( S \in \mathbb{Z}_2^n \) are precisely the linear functions from \( \mathbb{Z}_2^n \) to \( \mathbb{Z}_2 \).
WHY WE CALL \( x^S \chi_S \)

We can identify the hypercube \( 2^{[n]} \) with the set \( Q^n = \{-1, 1\}^n \). Here a subset \( A \subseteq [n] \) corresponds to \((x_1, \ldots, x_n) \in Q^n \) where \( x_i = -1 \) if \( i \in A \) and \( x_i = 1 \) if \( i \notin A \).

In this setting the function \( x^S \) is just evaluation of the monomial \( x_{s_1} \cdots x_{s_k} \) where \( S = \{s_1 < \cdots < s_k\} \). This explains why \( x^S \) is a natural notation.

We also see that any function \( f \in \mathcal{F} \) can be written uniquely as a real polynomial in the symbols \( x_1, \ldots, x_n \) where all monomials are squarefree. The coefficient of the monomial \( x^S = x_{s_1} \cdots x_{s_k} \) is \( \hat{f}(S) \). In fact, as an \( \mathbb{R} \)-algebra, \( \mathcal{F} \) (with the usual multiplication of functions) is isomorphic to

\[
\mathbb{R}[x_1, \ldots, x_n]/(x_1^2 - 1, \ldots, x_n^2 - 1)
\]

where \((x_1^2 - 1, \ldots, x_n^2 - 1)\) is the ideal generated by \( x_1^2 - 1, \ldots, x_n^2 - 1 \).

WHY THE GROUP PEOPLE CALL \( x^S \chi_S \)

From a group theoretical point of view we look at \( Q^n \) as a group under componentwise multiplication. Let \( \text{Irr}(Q^n) \) be the set of irreducible representations of \( Q^n \). Since \( Q^n \) is an Abelian group all irreducible representations are one-dimensional. Furthermore, every element in \( Q^n \) has order 1 or 2. This means that \( \text{Irr}(Q^n) \) is the set of group homomorphisms from \( Q^n \) to \( Q \).

Since the functions \( A \rightarrow A \cdot S \) for \( S \in \mathbb{Z}_2^n \) are precisely the linear functions from \( \mathbb{Z}_2^n \) to \( \mathbb{Z}_2 \), the homomorphisms from \( Q^n \) to \( Q \) are precisely the \( x^S \) for \( S \subseteq [n] \). For an Abelian group the irreducible characters are the same as the irreducible representations, so in group language \( x^S \) deserves being called \( \chi_S \).

THE TRINITY OF THE HYPERCUBE — A SUMMARY

<table>
<thead>
<tr>
<th>( 2^{[n]} )</th>
<th>( \mathbb{Z}_2^n )</th>
<th>( Q^n = {-1, 1}^n )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( A \subseteq [n] )</td>
<td>( A = (a_1, \ldots, a_n) ) where ( a_i = 1 \iff i \in A )</td>
<td>( A = (x_1, \ldots, x_n) ) where ( x_i = -1 \iff i \in A )</td>
</tr>
<tr>
<td>( A \triangle B ) (sym. diff.)</td>
<td>( A + B ) (c.w. add. mod 2)</td>
<td>( AB ) (c.w. mult.)</td>
</tr>
<tr>
<td>( x^S(A) = (-1)^{</td>
<td>A \cap S</td>
<td>} )</td>
</tr>
</tbody>
</table>

NICE PROPERTIES OF THE FOURIER TRANSFORM

In the following we will think of the hypercube as the group \( Q^n \).

For \( f, g \in \mathcal{F} \) let the convolution (sv. falling) \( f * g \in \mathcal{F} \) be defined by

\[
(f * g)(A) = \frac{1}{2^n} \sum_{A_1 A_2 = A} f(A_1) g(A_2).
\]

Observe that \( * \) is an associative operator. A very nice property of the Fourier transform is that

\[
\hat{f} * \hat{g} = \hat{f} \hat{g}.
\]

Parseval’s identity takes the following form on the hypercube:

\[
\frac{1}{2^n} \sum_{A \subseteq Q^n} |f(A)|^2 = \sum_{S \subseteq [n]} \hat{f}(S)^2.
\]
For boolean functions, that is functions \( f \in \mathcal{F} \) which attend only the values 1 and -1, we get
\[
\sum_{S \subseteq [n]} \hat{f}(S)^2 = 1.
\]

LINEARITY TESTING

A boolean function \( f \in \mathcal{F} \) is linear if \( f(AB) = f(A)f(B) \) for all \( A, B \in Q^n \). In other words, a function is linear if it is a group homomorphism from \( Q^n \) to \( Q \). Looking at \( f \) as a function from \( \mathbb{Z}_2^n \) to \( \mathbb{Z}_2 \), \( f \) is linear if and only if \( f(A + B) = f(A) + f(B) \) — a more common definition of linearity. We know that the only such functions are \( x^S \) for \( S \subseteq [n] \).

We introduce a metric on \( \mathcal{F} \):
\[
\operatorname{Dist}(f, g) = \operatorname{Prob}_{A \in Q^n} (f(A) \neq g(A)).
\]

To measure how linear a function is we define \( \operatorname{Dist}(f) \) to be the distance from \( f \) to the nearest linear function.

A linearity test on \( f \) is the following: Choose \( A \) and \( B \) independently in \( Q^n \) with uniform distribution. Accept if \( f(AB) = f(A)f(B) \).

We let \( \operatorname{Err}(f) \) be the probability that a linearity test on \( f \) does not accept.

The function \( \operatorname{Dist}(f) \) is hard to compute in practice, but \( \operatorname{Err}(f) \) can easily be approximated by performing the linearity test several times. Thus, we would like a relation between \( \operatorname{Dist}(f) \) and \( \operatorname{Err}(f) \).

**Theorem 1.** If \( f \in \mathcal{F} \) is a boolean function then \( \operatorname{Dist}(f) \leq \operatorname{Err}(f) \).

The proof needs two lemmas.

**Lemma 2.** Suppose \( f \in \mathcal{F} \) is a boolean function and \( S \subseteq [n] \). Then \( \hat{f}(S) \leq 1 - 2 \operatorname{Dist}(f) \).

**Proof.**
\[
\hat{f}(S) = \langle f, x^S \rangle = \frac{1}{2^n} \sum_{A \in Q^n} f(A)x^S(A) = \operatorname{Prob}_A(f(A) = x^S(A)) - \operatorname{Prob}_A(f(A) \neq x^S(A)) = 1 - 2 \operatorname{Dist}(f, x^S) \leq 1 - 2 \operatorname{Dist}(f).
\]

**Lemma 3.** If \( f \in \mathcal{F} \) is a boolean function, then
\[
\operatorname{Err}(f) = \frac{1}{2}(1 - (f * f * f)(1))
\]
where \( 1 = (1, \ldots, 1) \in Q^n \).

**Proof.** The linearity test chooses \( A, B \in Q^n \) and accepts if \( f(AB)f(A)f(B) = 1 \). Thus the expression \( \frac{1}{2}(1 - f(AB)f(A)f(B)) \) is an indicator for the rejection event in the linearity test. We get
\[
\operatorname{Err}(f) = \frac{1}{2^{2n}} \sum_{A, B \in Q^n} \frac{1}{2}(1 - f(AB)f(A)f(B)).
\]
From the definition of convolution it follows that
\[(f*f*f)(1) = \frac{1}{2^{2n}} \sum_{A,B \in Q^n} f(AB)f(A)f(B).\]

Now we are ready to prove Theorem 1.

**Proof of Theorem 1.** From Lemma 3 it suffices to analyze \((f*f*f)(1)\).

\[
(f*f*f)(1) = \sum_{S \subseteq [n]} (f*f*f)(S)x^S(1) = \sum_{S \subseteq [n]} f(S)^3 \\
\leq \left( \max_{S \subseteq [n]} \hat{f}(S) \right) \left( \sum_{S \subseteq [n]} \hat{f}(S)^2 \right) = \max_{S \subseteq [n]} \hat{f}(S) \quad \text{(Using Parseval’s identity)} \\
\leq 1 - 2 \text{Dist}(f) \quad \text{(Using Lemma 2)}
\]

Now using Lemma 3 we have

\[
\text{Err}(f) = \frac{1}{2} (1 - (f*f*f)(1)) \geq \frac{1}{2} (1 - (1 - 2 \text{Dist}(f))) = \text{Dist}(f).
\]

Influences

In this section we think of the hypercube as \(2^{[n]}\) and defines a function \(f \in \mathcal{F}\) to be **boolean** if it attains only the values 0 and 1.

For \(S \subseteq [n]\), the **influence of S over f**, denoted by \(I_f(S)\), is defined as

\[I_f(S) = \text{Prob}_{A \subseteq [n]}(\exists B \subseteq S : f(A \Delta B) \neq f(A)),\]

that is, the probability that the “variables” in \(S\) can affect the function value.

**Theorem 4.** Let \(f \in \mathcal{F}\) be a Boolean function which equals one with probability \(p \leq 1/2\). Then

\[\sum_{i=1}^{n} I_f(\{i\})^2 \geq Cp^2 \log^2 n/n\]

where \(C\) is an absolute constant (for example \(C = 1/16\) suffices for large \(n\)).

**Remark.** If \(p \geq 1/2\) the above theorem gives that

\[\sum_{i=1}^{n} I_f(\{i\})^2 \geq C(1-p)^2 \log^2 n/n.\]

**Proof.** First of all, let us write \(\beta_i := I_f(\{i\})\) for short, and introduce the convention that a summation sign \(\sum\) with nothing below it means summation over \(S \subseteq [n]\).

For \(1 \leq i \leq n\), let \(\Delta_i : \mathcal{F} \to \mathcal{F}\) be the linear functional defined by

\[(\Delta_i f)(A) = f(A) - f(A \Delta \{i\}).\]

The reason for introducing \(\Delta_i\) is that

\[\beta_i = \|\Delta_i f\|_2^2\]

which is fairly evident.
It is easy to see that \( \Delta_i x^S = 2x^S \) if \( i \in S \) and \( \Delta_i x^S = 0 \) otherwise. By Fourier expansion we get
\[
\Delta_i f = \Delta_i \sum \hat{f}(S)x^S = \sum \hat{f}(S)\Delta_i x^S = \sum_{i \in S \subseteq [n]} 2\hat{f}(S)x^S.
\]
Parseval’s identity gives the euclidean norm of \( \Delta_i f \):
\[
\beta_i = \| \Delta_i f \|^2 = 4 \sum_{i \in S \subseteq [n]} \hat{f}(S)^2.
\]
Summing this over all \( 1 \leq i \leq n \) we obtain
\[
\sum_{i=1}^{n} \beta_i = 4 \sum |S|\hat{f}(S)^2.
\]
We want to show that \( \sum_{i=1}^{n} \beta_i^2 \) is large, but this is approximately the same thing as showing that \( \sum_{i=1}^{n} \beta_i \) is large. Since we know that
\[
\sum \hat{f}(S)^2 = \| f \|^2_2 = p,
\]
in some sense we must show that the norm of \( f \) cannot be concentrated on those \( \hat{f}(S) \) with small \( |S| \). In other words we look for upper bounds on sums such as
\[
\sum_{|S| \leq b} \hat{f}(S)^2
\]
for some bound \( b \). Unfortunately, sums of this kind are not too convenient to work with. But we have the following lemma whose proof is found in the appendix.

**Lemma 5.** Let \( g \) be a function from \( 2^{[n]} \) to \( \{-1, 0, 1\} \). Let \( t \) be the probability that \( g \neq 0 \). Then
\[
t \geq \sum \delta^{|S|} \hat{g}(S)^2
\]
for every \( 0 < \delta < 1 \).

We apply this lemma with \( g = \Delta_i f \). The probability that \( \Delta_i f \neq 0 \) is exactly \( \beta_i \), so we obtain
\[
\beta_i^{1+\delta} \geq \sum \delta^{|S|} |\hat{S}f(S)|^2.
\]
Summing this over \( 1 \leq i \leq n \) we have
\[
\sum_{i=1}^{n} \beta_i^{1+\delta} \geq 4 \sum \delta^{|S|} |S| \hat{f}(S)^2.
\]
Now ignoring the portion of the sum contributed by the sets \( S \) of cardinality exceeding \( b \) (a parameter which we shortly select), we obtain
\[
\sum_{i=1}^{n} \beta_i^{1+\delta} \geq 4\delta^b \sum_{|S| \leq b} |S| \hat{f}(S)^2.
\]
We also keep in mind that
\[
p = \sum \hat{f}(S)^2 = \hat{f}(\emptyset).
\]
So also
\[ \sum_{i=1}^{n} \beta_i^{2+\frac{2}{\delta}} \geq 4\delta^b \left( -p^2 + \sum_{|S| \leq b} \hat{f}(S)^2 \right). \]

At the same time, since
\[ \sum_{i=1}^{n} \beta_i = 4 \sum_{i=1}^{n} |S| \hat{f}(S)^2 \]
we also have
\[ \sum_{i=1}^{n} \beta_i \geq 4b \sum_{|S| > b} \hat{f}(S)^2. \]

Now we combine these inequalities to obtain
\[
\delta - b \sum_{i=1}^{n} \beta_i^{2+\frac{2}{\delta}} + b^{-1} \sum_{i=1}^{n} \beta_i \geq 4 \left( -p^2 + \sum_{i=1}^{n} \hat{f}(S)^2 \right) = 4(-p^2 + p) \geq 2p
\]
where the last inequality comes from the assumption \( p \leq 1/2 \). Denote \( \sum_{i=1}^{n} \beta_i^2 \) by \( \lambda^2/n \). From Cauchy-Schwartz we have
\[ \sum_{i=1}^{n} \beta_i < \lambda. \]

Since \( \frac{2}{1+\delta} < 2 \) we can use the monotonicity of \( r \)-th power averages like this:
\[ \left( \frac{1}{n} \sum_{i=1}^{n} \beta_i^{2+\frac{2}{\delta}} \right)^{\frac{\delta}{2+\frac{2}{\delta}}} \leq \left( \frac{1}{n} \sum_{i=1}^{n} \beta_i^2 \right)^{\frac{\delta}{2}} = \frac{\lambda}{n}, \]
which yields
\[ \sum_{i=1}^{n} \beta_i^{2+\frac{2}{\delta}} \leq \lambda^{\frac{2}{\delta+1}} n^{\frac{1-\delta}{\delta+1}}. \]

If \( p = 0 \) the theorem is trivially true, so we assume \( p > 0 \). Choose \( b \) to be \( \lambda/p \). The second term in (1) cannot exceed \( p \) and so we remain with
\[ \delta - b \lambda \frac{2}{\delta+1} n^{\frac{1-\delta}{\delta+1}} \geq p. \]

Put \( \delta = 1/2 \) to get
\[ 2\lambda \lambda^{4/3} n^{-1/3} \geq p. \]

Define \( \mu \) so that \( \lambda = \mu p \log n \). We get
\[ n^{\mu - \frac{1}{3}} (\mu p \log n)^{4/3} \geq p. \]

If this should hold for large \( n \) clearly \( \mu \geq 1/4 \). Finally we get
\[ \sum_{i=1}^{n} \beta_i^2 = \frac{\lambda^2}{n} \geq \frac{1}{16} p^2 \log^2 n. \]
We can iterate the theorem about singleton influences to say something about influences of larger sets. For $S \subseteq [n]$ define the influence towards 1 of $S$ on $f$, denoted by $I^f_j(S)$, as

$$I^f_j(S) = \Pr_{A \subseteq [n]}(f(A) = 0 \land \exists B \subseteq S : f(A \triangle B) = 1),$$

that is, the probability that $f = 0$ but the “variables” in $S$ can make $f$ attain the value 1. Similarly, define the influence towards 0 of $S$ on $f$, denoted by $I^0_j(S)$, as

$$I^0_j(S) = \Pr_{A \subseteq [n]}(f(A) = 1 \land \exists B \subseteq S : f(A \triangle B) = 0).$$

Clearly

$$I^f_j(S) + I^0_j(S) = I_f(S) \tag{2}$$

and for singletons we also have

$$I^0_j(\{i\}) = I^f_j(\{i\}). \tag{3}$$

It is easy to see that $I^f_j(S) \leq 1 - p$. The following theorem tells us that there is a set $S$ of small cardinality that almost attains this maximum.

**Theorem 6.** Let $f : [n] \to \{0, 1\}$ be a boolean function, let $p = \Theta(1)$ be the probability that $f = 1$ and let $\omega = \omega(n)$ be any function tending to infinity with $n$. Then there is a set of cardinality $\leq \frac{1}{\log n} \omega(n) = o(n)$ whose influence towards one is $1 - p - o(1)$.

**Proof.** We will define a sequence of boolean functions $f_k : [k] \to \{0, 1\}$ for $k = n, n - 1, \ldots$ recursively. Let $p_k$ denote the probability that $f_k = 1$.

Start with $f_n = f$. Suppose we have already defined $f_k$ and now we are about to define $f_{k-1}$. By Theorem 4 there is some $1 \leq i \leq k$ with

$$I_{f_k}(\{i\}) \geq C p_k \log k / k$$

if $p_k \leq 1/2$. Without loss of generality we assume $i = k$. Now define the function $f_{k-1} : [k-1] \to \{0, 1\}$ by

$$f_{k-1}(A) := \begin{cases} 1 & \text{if } f_k(A) = 1 \text{ or } f_k(A \cup \{k\}) = 1, \\ 0 & \text{otherwise.} \end{cases}$$

It is easy to see that

$$p_{k-1} = p_k + I^f_j(\{k\}) = p_k + \frac{1}{2} I_{f_k}(\{k\})$$

where the last equality comes from (2) and (3). Thus

$$p_{k-1} \geq \min \left( \frac{1}{2}, p_k \left( 1 + C \frac{\log k}{2k} \right) \right).$$

Iterating $l$ steps yields $p_{n-l} \geq 1/2$ or

$$p_{n-l} \geq p \prod_{k=n-l+1}^n \left( 1 + C \frac{\log k}{2k} \right) \geq p \left( 1 + C \frac{\log n}{2n} \right)^l.$$ 

If we choose

$$l \approx \frac{2n}{C \log n} \log \frac{1}{p}$$
the right-hand side of (4) becomes greater that 1 and we get
\[ p_{n-l} \geq 1/2. \]

Now we continue defining the \( f_k \) for \( k < n-l \). We defined \( f_{n-l} \) before. Suppose we have already defined \( f_k \) with \( p_k \geq 1/2 \) and are about to define \( f_{k-1} \).

By the remark to Theorem 4 there is some \( 1 \leq i \leq k \) with
\[ I_{f_k}(\{i\}) \geq C(1-p_k)\log k/k. \]
Without loss of generality we assume \( i = k \). Now define the function \( f_{k-1} : [k-1] \rightarrow \{0,1\} \) by
\[
 f_{k-1}(A) := \begin{cases} 
 1 & \text{if } f_k(A) = 1 \text{ or } f_k(A \cup \{k\}) = 1, \\
 0 & \text{otherwise} 
\end{cases}
\]

exactly as before. We get
\[
 1 - p_{k-1} = 1 - (p_k + I_j^1(\{k\})) = (1 - p_k) - \frac{1}{2}I_{f_k}(\{k\}) 
\]
In particular \( p_{k-1} \geq 1/2 \) so that we are ready to define \( f_{k-2} \) in the next step.

Iterating \( j \) steps yields
\[
(5) \quad (1 - p_{n-l-j}) \leq (1 - p_{n-l}) \prod_{k=n-l-j+1}^{n-l} \left( 1 - C \frac{\log k}{2k} \right) \leq \frac{1}{2} \left( 1 - C \frac{\log n}{2n} \right)^j. 
\]
If we choose \( j \approx \frac{2n}{C \log n} u(n) \),
where \( u(n) \to \infty \) as \( n \to \infty \), then the right-hand side of (5) tends to zero as \( n \to \infty \). Thus we have showed that \( p_{n-(l+j)} = 1 - o(1) \) where
\[
l + j \approx \frac{2n}{C \log n} \left( u(n) + \log \frac{1}{p} \right) = \frac{n}{\log n} \omega(n) 
\]
if we choose
\[
u(n) = \frac{C}{2} \omega(n) - \log \frac{1}{p}. 
\]
Let \( S = \{n, n-1, \ldots, n-(l+j)\} \). It follows from the definition of \( f_{n-(l+j)} \) that
\[
p_{n-(l+j)} = \Pr_{A \subseteq [n]\backslash S} (\exists B \subseteq S : f(A \cup B) = 1) = \Pr_{A \subseteq [n]} (\exists B \subseteq S : f(A \Delta B) = 1) = p + I_j^1(S) 
\]
so we have showed that
\[
I_j^1(S) = 1 - p - o(1). \]
\[ \square \]
In this section we prove Lemma 5 which was used in the proof of Theorem 4.

For a finite set $X$, let $L^p(X)$ denote the metric space of all real functions on $X$ with norm

$$\|f\|_p = \left( \frac{1}{|X|} \sum_{x \in X} |f(x)|^p \right)^{1/p}.$$ 

Recall that $Q = \{-1, 1\}$. Let $0 < \varepsilon < 1$ be a real number. Define a functional $T : L^{1+\varepsilon^2}(Q) \to L^2(Q)$ by

$$(Tf)(x) = f(\varepsilon x).$$

**Lemma 7.** $\|T\| \leq 1$, that is, $\|Tf\|_2 \leq \|f\|_{1+\varepsilon^2}$ for every $f \in L^{1+\varepsilon^2}(Q)$.

**Proof.** Any real function $f$ on $Q$ can be written $f(x) = a + bx$ where $a$ and $b$ are real constants. Thus we need to show that

$$\left( \frac{(a - \varepsilon b)^2 + (a + \varepsilon b)^2}{2} \right)^{1/2} \leq \left( \frac{|a - b|^{1+\varepsilon^2} + |a + b|^{1+\varepsilon^2}}{2} \right)^{1/(1+\varepsilon^2)}.$$

By an appropriate scaling we can assume that $a = 1$, so we must show that

$$\sqrt{1 + (p - 1)b^2} \leq \left( \frac{|1 - b|^p + |1 + b|^p}{2} \right)^{1/p}$$

where $p := 1 + \varepsilon^2$. Observe that proving this inequality for $0 \leq b \leq 1$ will also imply the case of $b > 1$; just divide through by a factor of $b$. Also, for symmetry reasons, it suffices to consider $b \geq 0$. Consider the function

$$\varphi(b) = \frac{1}{p} \ln \frac{(1 - b)^p + (1 + b)^p}{2} - \frac{1}{2} \ln(1 + (p - 1)b^2).$$

We shall show that $\varphi(b) \geq 0$ for $0 \leq b \leq 1$ and $1 < p < 2$. We compute the derivative

$$\varphi'(b) = \left[ (1 - b)^p + (1 + b)^p \right]^{-1} [1 + (p - 1)b^2]^{-1} \theta(b)$$

where

$$\theta(b) = (1 + b)^{p-1} (1 - (p - 1)b) - (1 - b)^{p-1} [1 + (p - 1)b].$$

We also compute

$$\theta'(b) = p(p - 1)b[(1 - b)^{p-2} - (1 + b)^{p-2}].$$

For $0 \leq b \leq 1$ and $1 < p < 2$ we have $\theta'(b) \geq 0$ which implies $\varphi'(b) \geq 0$ which implies $\varphi(b) \geq 0$. □

The functional $T$ was designed for $Q$ but we are interested in the higher-dimensional hypercube $Q^n$. The following lemma is very useful when we increase the dimension.
Lemma 8. Let $p \leq q$ be positive real numbers. For $i = 1, 2$, let $X_i$ and $Y_i$ be finite sets and let $T_i : L^p(X_i) \to L^q(Y_i)$ be any two functionals. Let $T'_1$ and $T'_2$ be the functionals from $L^p(X_1 \times X_2)$ to $L^q(Y_1 \times Y_2)$ defined by

$$(T'_1 f)(x_1, x_2) = (T_1(* \mapsto f(*, x_2)))(x_1),$$

$$(T'_2 f)(x_1, x_2) = (T_2(* \mapsto f(x_1, *)))(x_2).$$

If $T_1$ and $T_2$ have norms at most 1 (i.e. $\|T_i f\|_q \leq \|f\|_p$ for every $f \in L^p(X_i)$), then the product

$$T'_1 T'_2 : L^p(X_1 \times X_2) \to L^q(Y_1 \times Y_2)$$

has norm at most 1 as well.

Proof. For any function $f \in L^p(X_1 \times X_2)$ the following holds.

$$\left( E_{X_1 \times X_2} |(T'_1 T'_2 f)(x_1, x_2)|^q \right)^{1/q} = \left( E_{X_1 \times X_2} |(T_1(* \mapsto (T'_2 f)(*, x_2))(x_1)|^q \right)^{1/q} \leq \left( E_{X_1} \left[ E_{Y_1} |(T'_2 f)(*, x_2))(y_1)|^p \right]^{q/p} \right)^{1/q} \left( \text{Since } |T_1| \leq 1 \right) \leq \left( E_{X_1} \left[ E_{Y_1} |T'_2f(x_2)|^q \right]^{p/q} \right)^{1/p} \left( \text{Minkowski's inequality} \right) = \left( E_{Y_1} \left[ E_{X_1} |T'_2f(x_2)|^q \right]^{p/q} \right)^{1/p} \left( \text{Since } |T_2| \leq 1 \right) = \left( E_{Y_1} E_{X_1} |f(y_1, x_2)|^p \right)^{1/p} \left( \text{Parseval} \right).$$

Here we have used Minkowski’s inequality, that is, for $r \geq 1$

$$\left( E_X |E_Y |F(x, y)|^r \right)^{1/r} \leq E_Y \left( E_X |F(x, y)|^r \right)^{1/r}.$$

In the computation above we take $r = q/p \geq 1$. □

Now we multiply $T$ by itself $n$ times in the sense of Lemma 8 to get

$$T_n := \underbrace{T \cdots T}_{n} \cdots T'$$

which is a functional from $L^{1+\varepsilon^2}(Q^n)$ to $L^2(Q^n)$. From Lemma 7 and 8 we know that $\|T_n\| \leq 1$. Since

$$(T_n f)(x_1, \ldots, x_n) = f(\varepsilon x_1, \ldots, \varepsilon x_n)$$

it is evident that the action of $T_n$ on a Fourier basis function is $T_n x^S = \varepsilon^{\hat{S}} x^S$. Putting $\varepsilon^2 := \delta$ we get

$$\|T_n g\|_2 = \sqrt{\sum \overline{(T_n g)(S)}^2} = \sqrt{\sum |\varepsilon^{\hat{S}} \hat{g}(S)|^2} = \sqrt{\sum |\hat{g}^{\hat{S}}\hat{g}(S)|^2} = \sqrt{\sum \overline{|\hat{g}^{\hat{S}}(S)|^2}} = \sqrt{\sum \overline{\hat{g}(S)|^2}}.$$

On the other hand

$$\|g\|_{1+\varepsilon^2} = \|g\|_{1+\delta} = \left( \frac{1}{2^n} \sum_{A \subseteq Q^n} |g(A)|^{1+\delta} \right)^{\frac{1}{1+\delta}} = \|g\|_{1+\delta}.$$
FOURIER TRANSFORM ON THE HYPERCUBE

11

TWO USEFUL INEQUALITIES

Here we prove two inequalities that we used earlier.

Theorem 9 (Minkowski’s inequality). Let $X$ and $Y$ be finite probability spaces. Let $r \geq 1$ be a real number and let $F(x, y)$ be a nonnegative real function on $X \times Y$. Then

$$(E_X[E_Y F(x, y)]^r)^{1/r} \leq E_Y (E_X F(x, y)^r)$$

Proof. Let $N(y) := (E_X F(x, y)^r)^{1/r}$. We have

$$\left( \frac{E Y F(x, y)^r}{E Y N(y)} \right)^r = \left( E_y \left( \frac{N(y)}{E Y N(y)} \right) \frac{F(x, y)^r}{N(y)^r} \right)^r \leq E_Y \left( \frac{N(y)}{E Y N(y)} \right) \frac{F(x, y)^r}{N(y)^r}$$

by the strong form of Jensen below, since $\ast \mapsto \ast^r$ is a convex function. Taking $E_X$ of this yields

$$E_X E_Y \left( \frac{N(y)}{E Y N(y)} \frac{F(x, y)^r}{N(y)^r} \right) = \frac{E Y N(y) E X F(x, y)^r}{E Y (E X F(x, y)^r)^{1/r}} = 1.$$

This means that

$$1 \geq E_X \left( \frac{E Y F(x, y)^r}{E Y N(y)} \right)^r = \frac{E X [E Y F(x, y)]^r}{(E Y (E X F(x, y)^r)^{1/r})^r}$$

which proves the theorem since $\ast \mapsto \ast^r$ is an increasing function. □

Theorem 10 (Inequality of $r$-th power averages). Let $x$ be a nonnegative real random variable and let $r \leq s$ be positive real numbers. Then

$$(E(x^r))^{1/r} \leq (E(x^s))^{1/s}.$$

Proof. With $y = x^r$ and $t = s/r$ the inequality can be written

$$(E y^t)^{1/t} \leq (E y^s)^{1/s}$$

which follows from Jensen since $t \geq 1$. □

For safety reasons we also state Jensen’s inequality, but without a proof.

Theorem 11 (Jensen’s inequality (weak version)). If $f$ is a convex function then

$$f \left( \sum \lambda_k x_k \right) \leq \sum \lambda_k f(x_k).$$

for all real $x_k$ and nonnegative $\lambda_k$ such that $\sum \lambda_k = 1$.

Theorem 12 (Jensen’s inequality (strong version)). Let $f$ be a convex function and let $x$ and $y$ be random variables. If $x \leq 0$ and $E(x) = 1$ then

$$f(E(xy)) \leq E(f(y)).$$

Proof. Let $\Omega$ be the underlying probability space (which we assume is finite) and let $F_x(\omega)$ and $F_y(\omega)$ be the density functions of $x$ and $y$ respectively. From the weak form of Jensen we have

$$f(E(xy)) = f \left( \frac{1}{|\Omega|} \sum_{\omega \in \Omega} F_x(\omega) F_y(\omega) \right) \leq \frac{1}{|\Omega|} \sum_{\omega \in \Omega} F_x(\omega) f(F_y(\omega)) = E(f(y)).$$

□