On Sharp Thresholds of Monotone Properties: Bourgain’s Proof Revisited.

Deepak Bal
Department of Mathematical Sciences,
Carnegie Mellon University,
Pittsburgh, PA 15213.

Abstract

The purpose of this expository note is to give the proof of a theorem of Bourgain with some additional details and updated notation. The theorem first appeared as an appendix to the breakthrough paper by Friedgut, Sharp Thresholds of graph properties and the k-SAT Problem [2]. Throughout, we use notation and definitions akin to those in O’Donnell’s book, Analysis of Boolean Functions [5].

1 Background and Theorem

Random structures often exhibit what is called a threshold phenomenon. That is, a relatively small change in a parameter can cause a swift change in the structure of the overall system. In the random graph $G(n, p)$, the probability space consisting of $n$ vertices and edge probability $p$, this phenomenon is a central object of study. In his 1999 paper, Sharp Thresholds of graph properties and the k-SAT Problem, Friedgut gave a simple characterization of monotone graph properties with coarse thresholds. The result is important because unlike results which preceded, it holds when $p = p(n) \to 0$ like $n^{-\Theta(1)}$ which is a range in which many thresholds occur. In the appendix to that paper, Bourgain gave a characterization of general monotone properties (as opposed to graph properties) which exhibit coarse thresholds. In this note, we explain the proof of this result with more details.

Let $(\Omega, \pi)$ be a finite probability space and for $n \in \mathbb{N}$, let $(\Omega^n, \pi^\otimes n)$ be the $n$ dimensional product probability space. We will write $x \sim \pi^\otimes n$ to indicate

*dbal@cmu.edu
that $x$ is drawn from $\Omega^n$ according to $\pi^{\otimes n}$. Bourgain’s result concerns the particular product space $\{0,1\}^n$, $\mu_p$ is the $p$-biased distribution on $\{0,1\}$. So $\mu_p(1) = p$, $\mu_p(0) = q := 1 - p$. We will use the notation $\{0,1\}_p^n$ for $\{(0,1)^n, \mu_p^{\otimes n}\}$.

Throughout, unless otherwise specified, we will write $P[\cdot]$ for $P_{x \sim \pi^{\otimes n}}[\cdot]$ and $E[\cdot]$ for $E_{x \sim \pi^{\otimes n}}[\cdot]$. If we are in the context of $\{0,1\}_p^n$, then the probability and expectations will be with respect to $\mu_p^{\otimes n}$.

In this note, we will consider $f : \Omega^n \to \{-1,1\}$. This will simplify some calculations from Bourgain’s proof where the range is taken to be $\{0,1\}$. We say $f : \{0,1\}^n \to \{-1,1\}$ is monotone (increasing) if $f(x) \leq f(y)$ whenever $x \leq y$ component-wise. For any subset $S \subseteq [n]$, we write $x_S$ to refer to the coordinates of $x$ from $S$. In an abuse of notation, sometimes this will refer to a vector of length $|S|$ and sometimes we will want $x_S$ to be a vector of length $n$. Also, for $S \subseteq [n]$, we write $1_S$ for the vector of length $n$ with 1s in the positions corresponding to $S$ and 0s elsewhere.

Let $f : \Omega^n \to \{-1,1\}$. The $i$th expectation operator, $E_i$, applied to $f$ takes the expectation with respect to variable $x_i$. So

$$E_i f(x) = E_{x_i \sim \pi} [f(x_1, \ldots, x_{i-1}, x_i, x_{i+1}, \ldots, x_n)]$$

is a function of $x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n$. We also define the $i$th directional Laplacian operator, $L_i$, by

$$L_i f = f - E_i f.$$  

The influence of coordinate $i$ on $f$ is defined as

$$\text{Inf}_i[f] = \langle f, L_i f \rangle = \langle L_i f, L_i f \rangle$$

where the inner product is defined by

$$\langle f, g \rangle = E_{x \sim \pi^{\otimes n}} [f(x)g(x)].$$

The total influence of $f$ is $\|f\| = \sum_{i=1}^n \text{Inf}_i[f]$.

Let $f = \sum_{S \subseteq [n]} f^=S$ be the generalized Walsh expansion or orthogonal decomposition of $f$. Recall, that the orthogonal decomposition of $f$ is the unique decomposition that satisfies the following two properties

1. For every $S \subseteq [n]$, $f^=S(x) = f^=S(x_1, \ldots, x_n)$ depends only on $x_i$ for which $i \in S$.

2. For every $S \subseteq [n]$, $E_i f^=S(x) = 0$ for all $i \in S$.  

2
In the case of $\{0,1\}_p^n$, for any $S \subseteq [n]$, we have that
\[
f^S(x) = \hat{f}(S) \prod_{i \in S} r(x_i)
\]
where $r(0) = -\sqrt{\frac{1-p}{p}}$ and $r(1) = \sqrt{\frac{1-p}{p}}$. We also have that $\hat{f}(S) = \mathbb{E}_{x \sim \mu_p^n} [f(x) \prod_{i \in S} r(x_i)]$.

If $S \subseteq [n]$ and $\bar{S} = [n] \setminus S$, we let $f^{\leq S}$ represent the function dependent on the coordinates of $S$ where we take the expectation of $f$ over the variables in $\bar{S}$. So if we think of $x$ as $(x_S, x_{\bar{S}})$, then
\[
f^{\leq S}(x) = f^{\leq S}(x_S) = \mathbb{E}_{x_{\bar{S}} \sim \pi^n_{\bar{S}}} [f(x_S, x_{\bar{S}})].
\]

$f^S$ and $f^{\leq S}$ are related by the following two formulas:
\[
f^S = \sum_{J \subseteq S} (-1)^{|S| - |J|} f^{\leq J} \tag{1.1}
\]
and
\[
f^{\leq S} = \sum_{J \subseteq S} f^{=J} \tag{1.2}
\]

Basic Fourier formulas, which hold for the orthogonal decomposition, give us that
\[
L_i f = \sum_{S \ni i} f^{=S}, \quad \text{Inf}_i[f] = \sum_{S \ni i} \|f^{=S}\|_2^2 = \sum_{S \ni i} \hat{f}(S)^2 \tag{1.3}
\]
and
\[
\Pi[f] = \sum_{i=1}^n \|L_i f\|_2^2 = \sum_{S \subseteq [n]} |S| \|f^{=S}\|_2^2 \tag{1.4}
\]
where the last equality in (1.3) holds in the case of $\{0,1\}_p^n$.

For products of general finite probability spaces, we have the following result

**Theorem 1.** For any $f : \Omega^n \to \{-1,1\}$ with $\mathbb{E} [f(x)] = 0$ and $\Pi[f] < C$, we have
\[
\mathbb{E} \left[ \max_{0 < |S| \leq 10C} |f^{\leq S}(x)| \right] > \delta \tag{1.5}
\]
where $\delta = 2^{-O(C^2)}$. 

3
This result is the main ingredient in Bourgain’s proof and it does not rely on the space being $p$-biased bits, so we will prove it here without such an assumption.

For a monotone boolean function $f : \{0,1\}^n \rightarrow \{-1,1\}$, Margulis [4] and Russo [6] proved the following relationship between the total influence and the sharpness of the threshold:

$$p(1-p) \frac{d}{dp} \mathbb{P} [f(x) = 1] = \mathbb{I}[f]$$  \hspace{1cm} (1.6)

where $\mathbb{P}$ and $\mathbb{I}$ are both with respect to $\mu_p^\otimes n$. In other words, the rate of transition of $f$ from $-1$ to $1$ with respect to the rate of increase of $p$ is determined by the total influence. Hence functions with large total influence should have “sharp” thresholds and functions with small total influence should have “coarse” thresholds.

Bourgain’s result in [2] is now given. This result basically states the following. Let $f : \{0,1\}^n \rightarrow \{-1,1\}$ be a monotone boolean function and let $p$ be the critical probability (which is allowed to approach 0 rapidly with $n$), when $f$ is equally likely to be $-1$ or 1. Then if $f$’s total influence is bounded, either (1) a non-negligible portion (according to $\mu_p^\otimes n$) of the $x$’s with $f(x) = 1$ have a small witness, or (2) there exists a small set of coordinates such that conditioning on these coordinates being 1 boosts the expected value of $f$ by a non-negligible amount. Keep in mind that in the following statement, $E$, $\mathbb{P}$ and $\mathbb{I}$ are with respect to $\mu_p^\otimes n$.

**Corollary 1.** Let $f : \{0,1\}^n \rightarrow \{-1,1\}$ be monotone (increasing) and suppose that $p = p(n)$ is such that $E[f] = 0$ and $\mathbb{I}[f] < C$. Then there exists some $\delta' = 2^{-O(C^2)}$ such that if $p < \delta' \frac{C}{20C}$ then at least one of the following two possibilities holds:

1. \[ \mathbb{P} [\exists S \subseteq [n], |S| \leq 10C, 1_S \leq x, f(1_S) = 1] > \delta'. \hspace{1cm} (1.7) \]

2. There exists $S' \subseteq [n]$ with $|S'| \leq 10C$ with $f(1_{S'}) = 0$ such that

\[ f \subseteq S' (1_{S'}) > \delta'. \hspace{1cm} (1.8) \]

**Proof of Corollary [7].** Let $\delta' = \delta/2$ where $\delta$ is given by Theorem [1]. Suppose that the first alternative of the theorem, (1.7), does not hold, i.e.,

\[ \mathbb{P} [\exists S \subseteq [n], |S| \leq 10C, 1_S \leq x, f(1_S) = 1] \leq \delta'. \hspace{1cm} (1.9) \]
Then applying Theorem 1 if $n$ is sufficiently large, there must exist $\bar{x} \in \{0,1\}^n$ and $S \subseteq [n], |S| \leq 10C$ such that for all $x' \leq \bar{x}$ with at most $10C$ 1’s, we have $f(x') = 0$ and

$$|f^{\subseteq S}(\bar{x})| > \delta'.$$  \hfill (1.10)

Now by monotonicity of $f$, we have for all $S, x_S$

$$f^{\subseteq S}(x_S) = \mathbb{E}_{x_S \sim \mu_p^S}[f(x_S, x_S)]$$

$$\geq \mathbb{E}_{x_S \sim \mu_p^S}[f(\bar{0}_S, x_S)]$$

$$\geq \mathbb{E} \left[ f(x) - \sum_{i \in S} 1\{x_i = 1\} \right]$$

$$= -p |S| > \frac{\delta'}{2}.$$  \hfill (1.11)

So (1.10) implies that

$$f^{\subseteq S}(\bar{x}) > \delta'$$

which implies the second alternative of the theorem, (1.8), by taking

$$S' = S \cap \{i : \bar{x}_i = 1\}.$$  \hfill \qed

The following easy corollary may be a useful statement.

**Corollary 2.** Let $f : \{0,1\}^n \to \{-1,1\}$ be monotone (increasing) and suppose that $p = p(n) < \frac{\delta}{100C}$ is such that $\mathbb{E}[f] = 0$ where $\delta = 2^{-O(C^2)}$. Furthermore, suppose that $\mathbb{P}[f] < C$. Then there exists a subset $S \subseteq [n]$ with $|S| \leq 10C$ such that

$$\mathbb{E}[f(x) \mid x_S = (1, \ldots, 1)] > \delta.$$  \hfill (1.11)

To derive this from Corollary 1 note that if the first alternative holds, then there exists a small $S$ which makes the expectation in (1.11) equal to 1. If the second alternative holds, note that (1.8) and (1.11) are equivalent.

As a corollary of his very general theorem, Hatami [3] proves that in fact the expectation in (1.11) can be made arbitrarily close to 1. The size of the guaranteed $S$ may have size exponential in $C^2$, but it is still independent of $n$. 

5
2 The Proof

Proof of Theorem \[\text{[1]}\] First observe that the facts \(\|f=0\|^2 = 0\) and \(\sum_{S \subseteq [n]} \|f=S\|^2 = 1\) and the assumption that \(\sum_{S \subseteq [n]} |S| \|f=S\|^2 < C\) imply that

\[
\frac{9}{10} \leq \sum_{0 < |S| \leq 10C} \|f=S\|^2 \tag{2.1}
\]

since

\[
\sum_{|S| > 10C} \|f=S\|^2 \leq \sum_{S \subseteq [n]} \frac{|S|}{10C} \|f=S\|^2 < \frac{C}{10C} = 1/10.
\]

Now, consider the following functions

\[
h_i(x) := \left( \sum_{S \ni i, |S| \leq 10C} |f=S(x)|^2 \right)^{1/2}
\]

and

\[
h(x) = \left( \sum_{|S| \leq 10C} |f=S(x)|^2 \right)^{1/2}
\]

By Prop. 6 of \[\text{[1]}\], we may say that for a fixed \(1 < q \leq 2\), we get

\[
\|h_i(x)\|_q^q \leq c_1 \|L_i f\|_q^q = c_1 \mathbb{E} \|L_i f\|_q^q \\
\leq c_2 \mathbb{E} \|L_i f\| \\
\leq C_1 \mathbb{E} \|L_i f\|^2 \\
= C_1 \text{Inf}_i[f]. \tag{2.2}
\]

with \(C_1 = C_1(q) = 2^{O(C)}\) and \(c_1, c_2\) are some constants which also depend only on \(q\). The reader should note that in the proof that follows, we will only apply the result of \[\text{[1]}\] with \(q = 4/3\). If \(q' = \frac{q}{q-1}\), then we also have

\[
\|h(x)\|_{q'} \leq C_1 \|f\|_{q'} = C_1. \tag{2.3}
\]
Hence we have
\[
\sum_{i=1}^{n} \|h_i(x)\|^q \leq C \sum_{i=1}^{n} \text{Inf}_i[f] \leq C \cdot C_1.
\] (2.4)

Let \(0 < \varepsilon < M < \infty\) be constants which are taken to be \(\varepsilon = 2^{-O(C)}\) and \(M = O\left(\frac{1}{\varepsilon}\right)\) and let
\[
\eta_i(x) = \mathbb{1}\{h_i(x) > \varepsilon\},
\]
\[
\xi(x) = \mathbb{1}\{\sum_i \eta_i(x) < M\}.
\] (2.5) (2.6)

Specific values for \(M\) and \(\varepsilon\) may be determined in terms of \(C\) and \(C_1\) by analyzing the inequalities that follow.

Now \(1 - \xi(x)\) is the indicator of the event that there are more than \(M\) coordinates \(i\), such that \(h_i(x) > \varepsilon\). Given relation (1.4) and the assumption that total influence is bounded, we should expect this event to have small probability. Hence we have, using Markov’s theorem twice, that
\[
\mathbb{E}\left[1 - \xi(x)\right] \leq \frac{1}{M} \mathbb{E}\left[\sum_{i=1}^{n} \eta_i(x)\right] \\
\leq \frac{1}{M \varepsilon^2} \mathbb{E}\left[\sum_{i=1}^{n} h_i(x)^2\right] \\
\leq \frac{1}{M \varepsilon^2} \mathbb{E}\left[\sum_{i=1}^{n} \sum_{S \ni i} f^S(x)^2\right] \\
\leq \frac{C}{M \varepsilon^2}.
\]

Now, inequality (2.1) tells us that
\[
\frac{9}{10} < \mathbb{E}\left[\sum_{0 < |S| \leq 10C} f^S(x)^2\right].
\]

Note that for any \(x\), either there exist \(\geq M\) many \(i\) such that \(h_i(x) > \varepsilon\), or there are \(< M\) such \(i\). In the latter case, there are two types of \(S \subseteq [n]\) with \(0 < |S| \leq 10C\): those which contain an \(i\) such that \(h_i(x) \leq \varepsilon\) and those containing only \(i\)’s such that \(h_i(x) > \varepsilon\).
Hence, using the indicator functions $\xi, \eta$, and recalling the definitions of $h(x)$ and $h_i(x)$, we may split up the following expectation as

$$
\mathbb{E} \left[ \sum_{0 < |S| \leq 10C} f_S(x)^2 \right] \leq \mathbb{E} \left[ h(x)^2 (1 - \xi(x)) \right] \quad (2.7)
$$

$$
+ \mathbb{E} \left[ \sum_{i=1}^{n} h_i(x)^2 (1 - \eta_i(x)) \right] \quad (2.8)
$$

$$
+ \mathbb{E} \left[ \sum_{0 < |S| \leq 10C} f_S(x)^2 \left( \prod_{i \in S} \eta_i(x) \right) \xi(x) \right]. \quad (2.9)
$$

We now bound each of these terms in turn.

For (2.7), we apply Cauchy-Schwarz and see that

$$
\mathbb{E} \left[ h(x)^2 (1 - \xi(x)) \right] \leq \mathbb{E} \left[ h(x)^4 \right]^{1/2} \mathbb{E} \left[ (1 - \xi(x))^2 \right]^{1/2}
$$

$$
\leq \|h(x)\|_4^2 \mathbb{E} \left[ (1 - \xi(x))^2 \right]^{1/2}
$$

$$
\leq C_i^2 \cdot \frac{C}{M^2} \quad (2.10)
$$

where we used (2.3) with $q' = 4$ (and hence $q = 4/3$) to go from (2.11) to (2.12).

For (2.8), we note that in this expectation, $h_i(x) \leq \varepsilon$ for any $x$ such that $\eta_i(x) = 0$. Also, since each $h_i$ is a positive function, we may write $h_i^2 = h_i^{2/3} h_i^{4/3}$. So

$$
\mathbb{E} \left[ \sum_{i=1}^{n} h_i(x)^2 (1 - \eta_i(x)) \right] = \sum_{i=1}^{n} \mathbb{E} \left[ h_i(x)^{2/3} h_i(x)^{4/3} (1 - \eta_i(x)) \right]
$$

$$
\leq \varepsilon^{2/3} \sum_{i=1}^{n} \mathbb{E} \left[ |h_i(x)|^{4/3} \right]
$$

$$
= \varepsilon^{2/3} \sum_{i=1}^{n} \|h_i(x)\|_{4/3}^{4/3}
$$

$$
\leq \varepsilon^{2/3} \cdot C \cdot C_1
$$

where we used (2.4) with $q = 4/3$ to get the last line.

Finally, for (2.9), we first observe that for any $x$, we have that

$$
\sum_{0 < |S| \leq 10C} \left( \prod_{i \in S} \eta_i(x) \right) \xi(x) < M^{10C}
$$
since if \( \xi(x) = 1 \), then \( \mathcal{M}_x = \{ i : \eta_i(x) = 1 \} \) has \( |\mathcal{M}_x| < M \). So the non-zero terms in the sum correspond to \( S \subseteq \mathcal{M}_x, 0 < |S| \leq 10C \). So we get

\[
\mathbb{E} \left[ \sum_{0 < |S| \leq 10C} f^S(x)^2 \left( \prod_{i \in S} \eta_i(x) \right) \xi(x) \right] \\
\leq \mathbb{E} \left[ \max_{0 < |S| \leq 10C} f^S(x)^2 \sum_{0 < |S| \leq 10C} \left( \prod_{i \in S} \eta_i(x) \right) \xi(x) \right] \\
\leq M^{10C} \mathbb{E} \left[ \max_{0 < |S| \leq 10C} f^S(x)^2 \right].
\]

Adding these three estimates gives

\[
\frac{9}{10} < C_1^2 \sqrt{\frac{C}{M \varepsilon^2}} + \varepsilon^{2/3} C_1 + M^{10C} \mathbb{E} \left[ \max_{0 < |S| \leq 10C} f^S(x)^2 \right]. \tag{2.13}
\]

Now, by taking \( \varepsilon = 2^{-O(C)} \) and \( M = O(1/\varepsilon) \), we easily have that

\[
\mathbb{E} \left[ \max_{0 < |S| \leq 10C} f^S(x)^2 \right] > 2^{-O(C^2)}.
\]

Now note that for any \( S \subseteq [n] \), by (1.11),

\[
|f^S| = \left| \sum_{J \subseteq S} (-1)^{|S|-|J|} f^J \right| \\
\leq \sum_{J \subseteq S} |f^J| \leq 2^{|S|} \max_{J \subseteq S} \{ |f^J| \} \tag{2.14}
\]

\[
\leq 2^{|S|}. \tag{2.15}
\]

since \( |f^J| \leq 1 \). So applying (2.14) and (2.15) and using the fact that
\( f \subseteq \emptyset = \mathbb{E}[f] = 0 \), we have

\[
2^{-O(C^2)} \leq \mathbb{E} \left[ \max_{0 < |S| \leq 10C} f^{S}(x)^2 \right] \\
= \mathbb{E} \left[ \max_{0 < |S| \leq 10C} |f^{S}(x)| |f^{S}(x)| \right] \\
\leq 2^{10C} \mathbb{E} \left[ \max_{0 < |S| \leq 10C} |f^{S}(x)| \right] \\
\leq 2^{20C} \mathbb{E} \left[ \max_{0 < |S| \leq 10C} \max_{J \subseteq S} |f^{\subseteq J}(x)| \right] \\
= 2^{20C} \mathbb{E} \left[ \max_{0 < |S| \leq 10C} |f^{\subseteq S}(x)| \right]
\]

which completes the proof.

\[ \square \]

References

[1] J. Bourgain, “Walsh subspaces of \( L^p \)-product spaces.” Séminaire Analyse fonctionnelle (dit ”Maurey-Schwartz”) (1979-1980): 1-14. [http://eudml.org/doc/109239](http://eudml.org/doc/109239)

[2] E. Friedgut, J. Bourgain, “Sharp Thresholds of Graph Properties, and the \( k \)-SAT Problem.” Journal of the American Mathematical Society (1999): Vol. 12, No. 4, 1017-1054.

[3] H. Hatami, “A structure theorem for Boolean functions with small total influences.” Annals of Mathematics, to appear.

[4] G. A. Margulis, “Probabilistic characteristics of graphs with large connectivity.” Problemy Perdachi Informacii (1974): Vol. 10, No. 2, 101-108.

[5] R. O’Donnell, “Analysis of Boolean Functions.” [http://www.analysisofbooleanfunctions.org](http://www.analysisofbooleanfunctions.org)

[6] L. Russo, “An approximate zero-one law.” Z. Wahrsch. Verw. Gabiete (1981): Vol. 61, No. 1, 129-139.