CSC2429 / MAT1304: Circuit Complexity January 17, 2019

# Lecture 2: Gate elimination and formula lower bounds

*Instructor: Benjamin Rossman*

**Last time:**

- Boolean functions, DeMorgan circuits and formulas, $\mathcal{C}(f)$, $\mathcal{L}(f)$
- Discussion of uniform vs. concrete models of computation
- Turing machine time $t(n)$ implies circuit size $O(t(n)^2)$
- Formula balancing: Every formula of size $s$ is equivalent to a formula of depth $O(\log s)$
- Lupanov's upper bound: Every $n$-ary boolean function has circuit size $O(2^n/n)$ (with more careful analysis: $2^n/n + o(2^n/n)$).
- Shannon's lower bound: Almost every $n$-ary boolean function has circuit size $> 2^n/n$.

A corollary of the Lupanov and Shannon bounds is the following size hierarchy theorem (which we didn't have time for last week). For a function $s : \mathbb{N} \to \mathbb{N}$, let SIZE$[s]$ be the class of boolean functions $f : \{0,1\}^* \to \{0,1\}$ such that $\mathcal{C}(f_n) \le s(n)$ for all $n \in \mathbb{N}$.

**Theorem 1** (Circuit Size Hierarchy Theorem)**.** *If* $n \le s(n) \le 2^{n-2}/n$, *then* $\mathrm{SIZE}[s] \subsetneq \mathrm{SIZE}[4s]$.

*Proof.* Pick $m \le n$ such that $s(n) \le 2^m/m \le 2s(n)$. By Shannon, there exists $f : \{0,1\}^m \to \{0,1\}$ such that $\mathcal{C}(f_n) > 2^m/m = s(n)$. By Lupanov, $\mathcal{C}(f) \le 2 \cdot 2^m/m \le 4s(n)$. □

Remark: This result bears similarity to the Time Hierarchy Theorem (for Turing machines), which states that $\mathrm{DTIME}(o(t(n)/\log t(n)))$ is a proper subclass of $\mathrm{DTIME}(t(n))$ for every time-constructible function $t(n)$. The proof is a diagonalization argument (not counting).

**Today:**

- Khrapchenko's lower bound (1971): $\mathcal{L}(\mathrm{XOR}_n) \ge n^2$
- 1-bit restrictions and gate elimination: $\mathcal{C}(\mathrm{XOR}_n) \ge 3(n-1)$ (Schnorr 1974)
- The $p$-random restriction and shrinkage of formulas (Subbotovskaya 1961, Håstad 1998, Tal 2014)
- Composition of boolean functions

# 1 Khrapchenko / Koutsoupias lower bound

For brevity, I will write $\mathrm{XOR}_n$ for $\mathrm{PARITY}_n$ and $\overline{\mathrm{XOR}}_n$ for $1-\mathrm{PARITY}_n$. Last week we observed the following upper bounds on the DeMorgan circuit and formula size of $\mathrm{XOR}_n$:

$$\mathcal{C}(\mathrm{XOR}_n) \leq 3(n-1) \quad \text{and} \quad \mathcal{L}(\mathrm{XOR}_n) \leq O(n^2).$$

Start with a balanced binary tree of $n-1$ $\oplus$ gates computing $\mathrm{XOR}_n$. Replace each $x \oplus y$ with the depth-2 DeMorgan circuit $(x \wedge \neg y) \vee (\neg x \wedge y)$. Result is a DeMorgan circuit of size $3(n-1)$ and depth $2\lceil \log n \rceil$. This is equivalent to a DeMorgan formula of size at most $2^{2\lceil \log n \rceil} \leq 4n^2$ (we get $\leq n^2$ when $n$ is a power of 2). (In fact, Yablonskii (1954) showed that $\mathcal{L}(\mathrm{XOR}_n) \leq \frac{9}{8}n^2$.)

We will show a lower bound $\mathcal{L}(\mathrm{XOR}_n) \geq n^2$ using Krapchenko's method (1971). We present a slightly stronger version of the method due to Koutsoupias (1993).

**Notation 2.** Let $\lambda(P)$ denote the largest eigenvalue of a symmetric matrix $P$. We will use the elementary fact from linear algebra: $\lambda(P+Q) \leq \lambda(P) + \lambda(Q)$ for symmetric matrices $P, Q$ of the same dimension.

**Notation 3.** For nonempty sets $A, B \subseteq \{0,1\}^n$, let $M \in \{0,1\}^{A \times B}$ be the $A \times B$ matrix

$$M_{a,b} := \begin{cases} 1 & \text{if } a_i \neq b_i \text{ for a unique } i \in [n] \text{ (i.e. } a, b \text{ are neighbors in the Hamming cube),} \\ 0 & \text{otherwise.} \end{cases}$$

We have symmetric matrices $M^T M \in \{0,1\}^{B \times B}$ and $MM^T \in \{0,1\}^{A \times A}$. Another elementary fact from linear algebra: $M^T M$ and $MM^T$ have the same nonzero eigenvalues. In particular $\lambda(M^T M) = \lambda(MM^T)$.

**Theorem 4.** *For any $f : \{0,1\}^n \to \{0,1\}$ and nonempty sets $A \subseteq f^{-1}(0)$ and $B \subseteq f^{-1}(1)$,*

$$\mathcal{L}(f) \geq \lambda(M^T M).$$

*Proof.* Induction on $\mathcal{L}(f)$. In the base case $\mathcal{L}(f) = 1$, $f(x)$ is $x_i$ or $1 - x_i$. We have $M^T M = 1_B$ (the $B \times B$ identity matrix). Therefore, $\lambda(M^T M) = 1$.

For the induction step, let $F$ be a minimal formula for $f$ with leafsize $\mathcal{L}(f) \geq 2$. Consider the case that $F = F_1 \wedge F_2$ where $F_1$ and $F_2$ compute functions $f_1$ and $f_2$. Note that $\mathcal{L}(f) = \mathcal{L}(f_1) + \mathcal{L}(f_2)$.

Let $A_1 := F_1^{-1}(0)$ and $A_2 := A \setminus A_2$. Note that $A_2 \subseteq F_2^{-1}(0)$ and $B \subseteq F_1^{-1}(1) \cap F_2^{-1}(1)$.

Note that matrices $M_1 \in \{0,1\}^{A_1 \times B}$ and $M_2 \in \{0,1\}^{A_2 \times B}$ satisfy $M^T M = M_1^T M_1 + M_2^T M_2$. Therefore,

$$\begin{aligned} \mathcal{L}(f) &= \mathcal{L}(f_1) + \mathcal{L}(f_2) \\ &\geq \lambda(M_1^T M_1) + \lambda(M_2^T M_2) \quad \text{(by induction hypothesis)} \\ &\geq \lambda(M_1^T M_1 + M_2^T M_2) \\ &= \lambda(M^T M). \end{aligned}$$

The argument when $F = F_1 \wedge F_2$ is symmetric in $A$ and $B$, using the fact that $\lambda(M^T M) = \lambda(MM^T)$. $\square$

**Corollary 5** (Khrapchenko's bound). $\mathcal{L}(f) \geq \dfrac{(\sum_{a\in A}\sum_{b\in B} M_{a,b})^2}{|A|\cdot|B|}$

Obs: $\sum_{a\in A}\sum_{b\in B} M_{a,b} = |\{(a,b)\in A\times B : a, b \text{ are neighbors in the Hamming cube}\}|$.

*Proof.* We have

$$\begin{aligned}
\lambda(M^TM) &= \max_{z\in\mathbb{R}^B\setminus\{\vec{0}\}} \frac{z^TM^TMz}{z^Tz} \\
&\geq \frac{\sum_{b,b'\in B}(M^TM)_{b,b'}}{|B|} && \text{(letting } z \text{ be the all-1 vector)} \\
&= \frac{\sum_{b,b'\in B}\sum_{a\in A} M_{a,b}M_{a,b'}}{|B|} \\
&= \frac{\sum_{a\in A}(\sum_{b\in B} M_{a,b})^2}{|B|} \\
&\geq \frac{(\sum_{a\in A}\sum_{b\in B} M_{a,b})^2}{|A|\cdot|B|} && \text{(Cauchy-Schwarz).}
\end{aligned}$$

□

Remark: There is a direct proof of Khrapchenko's bound by a similar argument to Koutsoupias, but using Cauchy-Schwarz in a less elegant way. Koutsoupias's bound is stronger by a constant factor in some cases.

We can use Khrapchenko's bound to prove a lower bound on $\mathcal{L}(\mathrm{XOR}_n)$. Let $A = \{$all even weight strings$\}$ and $B = \{$all odd weight strings$\}$. Then

$$\mathcal{L}(\mathrm{XOR}_n) \geq \frac{(\sum_{a\in A}\sum_{b\in B} M_{a,b})^2}{|A|\cdot|B|} = \frac{(n2^{n-1})^2}{2^{n-1}\cdot 2^{n-1}} = n^2.$$

EXERCISE: (1) Show $\mathcal{L}(\mathrm{MAJ}_n) = \Omega(n^2)$ using Khrapchenko's bound. (2) Can you devise a polynomial upper bound on $\mathcal{L}(\mathrm{MAJ}_n)$? (Later on, we will see a polynomial upper bound on $\mathcal{L}_{\mathrm{mon}}(\mathrm{MAJ}_n)$.)

# 2 Gate elimination and random restrictions

## 2.1 1-bit restrictions

For $i \in [n]$ and $b \in \{0,1\}$, we consider the **1-bit restriction** "$x_i \leftarrow b$" which sets the $i$th variable $x_i$ to the constant $b$.

1-bit restrictions operate on boolean functions $f^{(x_i\leftarrow b)}$ as well as *syntactically* on DeMorgan circuits $C^{(x_i\leftarrow b)}$. (Since we measure size by the number of $\wedge$ and $\vee$ gates, we shall consider circuits with $\wedge$ and $\vee$ gates only, where negations appear on wires.)

- $f^{(x_i\leftarrow b)}$ is the $(n-1)$-ary formula defined by

$$f^{(x_i\leftarrow b)}(x_1,\ldots,x_{i-1},x_{i+1},\ldots,x_n) = f(x_1,\ldots,x_{i-1},b,x_{i+1},\ldots,x_n).$$

- $C^{(x_i \leftarrow b)}$ is the $(n-1)$-ary circuit obtained from $C$ as follows.
  - First, substitute $x_i \rightsquigarrow b$ for all inputs labeled by $x_i$.
  - Next, perform the following **constant simplifications** on subcircuits of $C$ whenever possible:

$$\begin{array}{lcllcllcl} 0 \wedge C' & \rightsquigarrow & 0, & 0 \vee C' & \rightsquigarrow & C', & \neg 0 & \rightsquigarrow & 1, \\ 1 \wedge C' & \rightsquigarrow & C', & 1 \vee C' & \rightsquigarrow & 1, & \neg 1 & \rightsquigarrow & 0. \end{array}$$

    Note that the order of applying these simplifications doesn't matter.

Obs 0: If $C$ computes $f$, then $C^{(x_i \leftarrow b)}$ computes $f^{(x_i \leftarrow b)}$.

Obs 1: If $x_i$ appears below a gate in $C$, then for both settings of $b \in \{0,1\}$,

$$\text{size}(C^{(x_i \leftarrow b)}) \leq \text{size}(C) - 1.$$

Obs 2: If $x_i$ appears below two gates in $C$, then for at least one setting of $b \in \{0,1\}$,

$$\text{size}(C^{(x_i \leftarrow b)}) \leq \text{size}(C) - 2.$$

For example, if $(x_i \wedge C') \vee C''$ is a subcircuit of $C$, then setting $x_i \leftarrow 0$ kills both gates in this subcircuit (whereas setting $x_i \leftarrow 1$ kills only the $\wedge$ gate).

## 2.2 The lower bound $\mathcal{C}(\text{XOR}_n) \geq 3(n-1)$ (Schnorr 1974)

**Lemma 6.** *In any circuit $C$ computing $\text{XOR}_n$ or $\overline{\text{XOR}}_n$ where $n \geq 2$, some 1-bit restriction eliminates 3 gates.*

*Proof.* Let $g$ be any bottom-level $\wedge$ or $\vee$ gate in $C$ (such that no $\wedge$ or $\vee$ appears below $g$). Without loss of generality, we may assume that the one of the wires feeding into $g$ computes $x_i$ or $\overline{x}_i$ and the other wire computes $x_j$ or $\overline{x}_j$ for distinct variables $x_i$ and $x_j$.

Claim 1: $x_i$ appears direct below another gate $h$ of $C$, which is distinct from $g$. (If not, then some 1-bit restriction $C^{(x_j \leftarrow b)}$ kills $g$, making $x_i$ irrelevant to the computation; but this cannot happen since $C$ computes $\text{XOR}_n$ or $\overline{\text{XOR}}_n$.)

Claim 2: $h$ is not the output of $C$. (If it were, then some 1-bit restriction $C^{(x_i \leftarrow b)}$ makes $C$ constant, which cannot happen since $C$ computes $\text{XOR}_n$ or $\overline{\text{XOR}}_n$.)

Let $h'$ be any gate which receives $h$ as input. Note that $g, h, h'$ are three distinct gates in $C$. (Obs: $g$ and $h'$ are distinct by minimality of $g$.) For both values of $b \in \{0,1\}$, gates $g$ and $h$ are both eliminated in the circuit $C^{(x_i \leftarrow b)}$. There exist $b \in \{0,1\}$, such that $h^{(x_i \leftarrow b)}$ is fixed to a constant; for this $b$, the gate $h'$ is also eliminated in $C^{(x_i \leftarrow b)}$. ☐

**Corollary 7.** $\mathcal{C}(\text{XOR}_n) \geq 3(n-1)$

*Proof.* By induction, using the fact that $\text{XOR}_n^{(x_i \leftarrow b)}$ is equivalent to $\text{XOR}_{n-1}$ or $\overline{\text{XOR}}_{n-1}$ for every 1-bit restriction (and we have $\mathcal{C}(\text{XOR}_{n-1}) = \mathcal{C}(\overline{\text{XOR}}_{n-1}) \geq 3(n-2)$ by the induction hypothesis and invariance of $\mathcal{C}(\cdot)$ under negations). ☐

More sophisticated versions of this gate elimination argument (with more general kinds of 1-bit restrictions) are used in best lower bounds on the DeMorgan circuit size of explicit functions, currently $5n - o(n)$. In the full binary basis, the best lower bound was recently improved from $3n - o(n)$ to $(3 + \frac{1}{86})n - o(n)$.

## 3 Subbotovskaya's Method (1961)

We say that a formula $F$ is **nice** for every subformula $x_i \wedge F'$ or $\overline{x}_i \wedge F'$ or $x_i \vee F'$ or $\overline{x}_i \vee F'$, the variable $x_i$ does not appear in $F'$. Note that formula is equivalent to a nice formula of the same (or lesser) leafsize: simply perform the following syntactic transformations:

$$\begin{aligned} x_i \wedge F &\quad\rightsquigarrow\quad x_i \wedge F^{(x_i \leftarrow 1)}, \\ \overline{x}_i \wedge F &\quad\rightsquigarrow\quad \overline{x}_i \wedge F^{(x_i \leftarrow 0)}, \\ x_i \vee F &\quad\rightsquigarrow\quad x_i \wedge F^{(x_i \leftarrow 0)}, \\ \overline{x}_i \wedge F &\quad\rightsquigarrow\quad \overline{x}_i \wedge F^{(x_i \leftarrow 1)}. \end{aligned}$$

Repeatedly applying these transformations to all subformulas of a given formula produces an equivalent nice formula. (Note: This statement is not true of circuits. Make sure you understand why!)

As a consequence, any minimal formula $F$ for a function $f$ (such that $F$ has leafsize $\mathcal{L}(f)$) is nice.

**Lemma 8.** *For every $n$-ary Boolean function $f$,*

$$\mathop{\mathbb{E}}_{i\in[n],\, b\in\{0,1\}}[\ \mathcal{L}(f^{(x_i\leftarrow b)})\ ] \leq \left(1 - \frac{1}{n}\right)^{1.5} \mathcal{L}(f).$$

*Proof.* Let $F$ be a minimum-size nice formula computing $f$. For $i \in [n]$, let $\ell_i$ be the number of leaves of $F$ labeled with $x_i$ or $\overline{x}_i$. So, $\mathcal{L}(f) = \text{leafsize}(F) = \sum_{i=1}^{n} \ell_i$.

We may assume that leafsize$(F) \geq 2$ (since the lemma is trivial if leafsize$(F) = 1$). Therefore, every leaf in $F$ has a sibling-subformula. That is, each leaf $\lambda$ belongs to a subformula gate$(\lambda, F')$ of $F$ (where gate $\in \{\wedge, \vee\}$); we call $F'$ the *sibling of* $\lambda$. For random $b \in \{0,1\}$, the 1-bit restriction $F^{(x_i\leftarrow b)}$ kills the leaf $\lambda$ with probability 1 and, in addition, kills all leaves of $F'$ with probability $\frac{1}{2}$. For random $b \in \{0,1\}$, at least 1.5 leaves of gate$(\lambda, F')$ are killed in expected under the 1-bit restriction $F^{(x_i\leftarrow b)}$.

For each $i \in [n]$, we have

$$\mathop{\mathbb{E}}_{b\in\{0,1\}}[\ \underbrace{\text{leafsize}(F) - \text{leafsize}(F^{(x_i\leftarrow b)})}_{\#\text{ of leaves killed by the 1-bit restriction}}\ ] \geq 1.5\ell_i.$$

(Here we rely on niceness of $F$ to ensure that we are not overcounting.) By linearity of expectations,

$$\mathop{\mathbb{E}}_{i\in[n],\, b\in\{0,1\}}[\ \text{leafsize}(F) - \text{leafsize}(F^{(x_i\leftarrow b)})\ ] \geq \frac{1}{n}\sum_{i=1}^{n} 1.5\ell_i = \frac{1.5}{n}\text{leafsize}(F).$$

Therefore,

$$\mathop{\mathbb{E}}_{i\in[n],\, b\in\{0,1\}}[\ \mathcal{L}(f^{(x_i\leftarrow b)})\ ] \leq \mathop{\mathbb{E}}_{i\in[n],\, b\in\{0,1\}}[\ \text{leafsize}(F^{(x_i\leftarrow b)})\ ]$$
$$\leq \Big(1-\frac{1.5}{n}\Big)\text{leafsize}(F)$$
$$\leq \Big(1-\frac{1}{n}\Big)^{1.5}\mathcal{L}(f). \qquad \square$$

Remark: This lemma implies $\mathcal{L}(\text{XOR}_n) \geq n^{1.5}$ (easy exercise). This is weaker than the $n^2$ lower bound of Khrapchenko's method.

**Definition 9.** A **restriction** $\rho$ is a function $[n] \to \{0,1,*\}$. We think of $\rho$ as a partial assignment of variables to 0 or 1, where $\rho(i) = *$ means that the $i$th variable is unrestricted. We say that $\rho$ is a $k$-**star restriction** if $|\rho^{-1}(*)| = k$. (A 1-bit restriction is an $(n-1)$-star restriction.)

For $f : \{0,1\}^n \to \{0,1\}$ and a restriction $\rho : [n] \to \{0,1,*\}$, we denote by $f{\restriction}\rho : \{0,1\}^{\rho^{-1}(*)} \to \{0,1\}$ the restricted boolean function (defined in the obvious way).

**Theorem 10** (Subbotovskaya's bound)**.** *Let $f : \{0,1\}^n \to \{0,1\}$ be any boolean function and let $\boldsymbol{\rho}$ is a uniform random $k$-star restriction. Then*

$$\mathbb{E}[\ \mathcal{L}(f{\restriction}\boldsymbol{\rho})\ ] \leq \Big(\frac{k}{n}\Big)^{1.5}\mathcal{L}(f).$$

*Proof.* Repeatedly applying Lemma 8, we have

$$\mathbb{E}[\ f{\restriction}\boldsymbol{\rho}\ ] \leq \Big(1-\frac{1}{n}\Big)^{1.5}\Big(1-\frac{1}{n-1}\Big)^{1.5}\cdots\Big(1-\frac{1}{k+1}\Big)^{1.5}\mathcal{L}(f) = \Big(\frac{k}{n}\Big)^{1.5}\mathcal{L}(f). \qquad \square$$

**Definition 11.** For $p \in [0,1]$, the $p$-**random restriction** is the random restriction $\mathbf{R}_p : [n] \to \{0,1,*\}$ such that $\mathbb{P}[\ \mathbf{R}_p(i) = *\ ] = p$ and $\mathbb{P}[\ \mathbf{R}_p(i) = 0\ ] = \mathbb{P}[\ \mathbf{R}_p(i) = 1\ ] = \frac{1-p}{2}$ independently for each $i \in [n]$.

Subbotovskaya's bound has the following corollary.

**Corollary 12.** $\mathbb{E}[\ \mathcal{L}(f{\restriction}\mathbf{R}_p)\ ] \leq O(p^{1.5}\mathcal{L}(f)+1)$

A stronger version of this result is known:

**Theorem 13** (Håstad 1998, Tal 2014)**.** *For every Boolean function $f$ and $p \in [0,1]$,*

$$\mathbb{E}[\ \mathcal{L}(f{\restriction}\mathbf{R}_p)\ ] \leq O(p^2\mathcal{L}(f)+1).$$

Tal proves a tight bound $O(p^2\mathcal{L}(f) + p\sqrt{\mathcal{L}(f)})$ (more about his proof in a moment). Theorem 13 implies a lower bound $\mathcal{L}(\text{XOR}_n) = \Omega(n^2)$ (weaker than Khrapchenko's bound by a constant factor).

**Remark 14.** The maximum constant $\Gamma$ such that $\mathcal{L}(f{\restriction}\mathbf{R}_p) \leq O(p^\gamma\mathcal{L}(f)+1)$ for every $\gamma < \Gamma$ is called **shrinkage exponent** of DeMorgan formulas. Theorem 10 establishes that $\Gamma \geq 1.5$. This was improved to 1.55 by Impagliazzo and Nisan (1993) and 1.63 by Paterson and Zwick (1993). Finally, Håstad (1998) showed that $\Gamma = 2$.

For the class of *read-once formulas* (in which each variables occurs at most once), Håstad, Razborov, Yao (1985) showed that that $\Gamma_{\text{read-once}}$ is exactly $1/\log(\sqrt{5}-1) \approx 3.27$ . It is an open problem to determine $\Gamma_{\text{monotone}}$, the shrinkage exponent of *monotone formulas*. Note that $\Gamma \leq \Gamma_{\text{monotone}} \leq \Gamma_{\text{read-once}}$. It is conjectured that $\Gamma_{\text{monotone}} = \Gamma_{\text{read-once}}$.

### 3.1 Outline of Tal's proof of Theorem 13 [mostly skipped in lecture]

The *approximate degree* $\widetilde{\deg}(f)$ of a boolean function $f : \{0,1\}^n \to \{0,1\}$ is the minimum degree of a real polynomial $p \in \mathbb{R}[x_1, \dots, x_n]$ such that $|f(x) - p(x)| \le 1/3$ for every $x \in \{0,1\}^n$.

Approximate degree has an upper in terms of formula size:

**Theorem 15.** $\widetilde{\deg}(f) \le O(\sqrt{L(f)})$

The proof of this theorem goes through *quantum query complexity*. For any boolean function $f$, it is known that $\widetilde{\deg}(f) \le O(Q_2(f))$ and $Q_2(f) \le O(\sqrt{L(f)})$. (Open problem: Give a direct proof of Theorem 15 that does not go through quantum query complexity.)

Using boolean analysis (which we'll discuss later in the course), Tal showed:

**Lemma 16.** $\mathbb{E}[\ \mathcal{L}(f{\upharpoonright}\mathbf{R}_{1/\widetilde{\deg}(f)})\ ] = O(1)$

An corollary of this lemma and Theorem 15 (exercise):

**Corollary 17.** *If* $p \le 1/\sqrt{\mathcal{L}(f)}$*, then* $\mathbb{E}[\ \mathcal{L}(f{\upharpoonright}\mathbf{R}_p)\ ] = O(p{\cdot}\sqrt{\mathcal{L}(f)})$.

For the remaining case where $p > 1/\sqrt{\mathcal{L}(f)}$, Tal uses the following decomposition:

**Lemma 18.** *Let $F$ be an formula of leafsize $\ell$ and let $\ell \in \mathbb{N}$. Then there exist $m = O(s/\ell)$ formulas $G_1, \dots, G_m$, each of size $\le \ell$, and a read-once formula $H$ with $m$ inputs such that $F = H(G_1, \dots, G_m)$.*

(The proof of this lemma is a formula-balancing argument, somewhat along the lines of Spira's theorem, which we saw in the last lecture.)

Still assuming $p > 1/\sqrt{\mathcal{L}(f)}$, we take $F$ to be a minimal formula for $f$ (of leafsize $\mathcal{L}(f)$). Applying the above decomposition with $\ell = 1/p^2$ and $m = p^2\mathcal{L}(f)$, we get

$$\begin{aligned}
\mathbb{E}[\ \mathcal{L}(f{\upharpoonright}\mathbf{R}_p)\ ] &\le \sum_{i=1}^{m} \mathbb{E}[\ \mathcal{L}(G_i{\upharpoonright}\mathbf{R}_p)\ ] && \text{(linearity of expectations)}\\
&\le m \cdot O(p\sqrt{\ell}) && \text{(since } p \le 1/\sqrt{\mathcal{L}(G_i)} \le 1/\ell \text{ for each } i \in [m]\\
&= O(p^2\mathcal{L}(f)).
\end{aligned}$$

Combining the two cases for $p$, we get $\mathbb{E}[\ \mathcal{L}(f{\upharpoonright}\mathbf{R}_p)\ ] = O(p^2\mathcal{L}(f) + p\sqrt{\mathcal{L}(f)})$.

## 4 Composition of boolean functions

Andreev showed how to get a better lower bound $\Omega(n^{\Gamma+1-o(1)})$ on the leafsize on an explicit $n$-variable function. We will see this in the next lecture. Andreev's function is a based on a composition of boolean functions.

**Definition 19.** For $f : \{0,1\}^k \to \{0,1\}$ and $g : \{0,1\}^m \to \{0,1\}$, the **composition** $f \otimes g : (\{0,1\}^m)^k \to \{0,1\}$ is defined by

$$(f \otimes g)(X_1, \dots, X_k) := f(g(X_1), \dots, g(X_k)).$$

(Properly speaking, $f \otimes g$ is the composition of $f$ with the $k$-output function $g^k : \{0,1\}^m \to \{0,1\}^k$.) Viewing $X \in \{0,1\}^{k \times m}$ as a matrix with rows $X_1, \ldots, X_k$, we first apply $g$ to each row and then $f$ to the resulting vector of $g$-values.

Note that $\mathcal{L}(f \otimes g) \leq \mathcal{L}(f) \cdot \mathcal{L}(g)$. For example, $\mathcal{L}(f \otimes \mathrm{XOR}_m) \leq \mathcal{L}(f) \cdot O(m^2)$. The next lemma gives the reverse inequality (with a polylog($k$) loss).

**Lemma 20.** *For all $k, m \geq 1$ and $f : \{0,1\}^k \to \{0,1\}$,*

$$\mathcal{L}(f \otimes \mathrm{XOR}_m) \geq \mathcal{L}(f) \cdot \Omega \left( \frac{m}{\log k} \right)^2 .$$

We will see the proof next time. We mention this lemma is a special case of a general conjecture on the leafsize of composed functions. This is known as the KRW Conjecture (after Karchmer, Raz and Wigderson), one version of which states that $\mathcal{L}(f \otimes g) = \widetilde{\Omega}(\mathcal{L}(f) \cdot \mathcal{L}(g))$ for all functions $f$ and $g$ (where $\widetilde{\Omega}(t(n)) = \Omega(t(n))/(\log t(n))^{O(1)}$ for any function $t(n)$).