General Strong Polarization

Madhu Sudan
Harvard University

Joint work with Jaroslaw Blasiok (Harvard), Venkatesan Guruswami (CMU), Preetum Nakkiran (Harvard) and Atri Rudra (Buffalo)
This Talk: Martingales and Polarization

- [0,1]-Martingale: Sequence of r.v.s $X_0, X_1, X_2, \ldots$
  \[ X_t \in [0,1], \quad \mathbb{E}[X_{t+1}|X_0, \ldots, X_t] = X_t \]

- Well-studied in algorithms:
  - [MotwaniRaghavan, §4.4], [MitzenmacherUpfal, §12]
  - E.g., $X_t$ = Expected approx. factor of algorithm given first $t$ random choices.

- Usually study “concentration”
  - E.g. ... whp $X_{\text{final}} \in [X_0 \pm \epsilon] \iff \lim_{t \to \infty} X_t = X_0$

- Today: Polarization:
  - ... whp $X_{\text{final}} \notin (\epsilon, 1 - \epsilon) \iff \lim_{t \to \infty} X_t = \text{Bern}(X_0)$

- Why? – Polar codes ...
Main Result: Definition and Theorem

- **Strong Polarization:** (informally)
  \[
  \Pr[X_t \in (\tau, 1 - \tau)] \leq \epsilon \quad \text{if} \quad t \geq \max\{O(\log 1/\epsilon), o(\log 1/\tau)\}
  \]
  formally \quad \forall \gamma > 0 \exists \beta < 1, c \ s.t. \forall t \ \Pr[X_t \in (\gamma^t, 1 - \gamma^t)] \leq c \cdot \beta^t

- A non-polarizing example: \( X_0 = \frac{1}{2}; \ X_{t+1} = X_t \pm 2^{-t-2} \)

- **Local Polarization:**
  - **Variance in the middle:** \( X_t \in (\tau, 1 - \tau) \)  
    \( \forall \tau > 0 \ \exists \sigma > 0 \ \text{s.t.} \forall t, X_t \in (\tau, 1 - \tau) \Rightarrow V\text{ar}[X_{t+1} | X_t] \geq \sigma \)
  - **Suction at the ends:** \( X_t \not\in (\tau, 1 - \tau) \)  
    \( \exists \theta > 0, \forall c < \infty, \exists \tau > 0 \ \text{s.t.} X_t < \tau \Rightarrow \Pr\left[ X_{t+1} < \frac{X_t}{c} \right] \geq \theta \)

- **Theorem:** Local Polarization \( \Rightarrow \) Strong Polarization.

Both definitions qualitative!
Rest of the talk

- **Background/Motivation:**
  - The Shannon Capacity Challenge
  - Resolution by [Arikan’08], [Guruswami-Xia’13], [Hassani,Alishahi,Urbanke’13]: Polar Codes!

- **Key ideas:**
  - Polar Codes & Arikan-Martingale.
  - Implications of Weak/Strong Polarization.

- **Our Contributions: Simplification & Generalization**
  - Local Polarization $\Rightarrow$ Strong Polarization
  - Arikan-Martingale Locally Polarizes (generally)
Shannon and Channel Capacity

- Shannon ['48]: For every (noisy) channel of communication, $\exists$ capacity $C$ s.t. communication at rate $R < C$ is possible and $R > C$ is not.
  - Needs lot of formalization – omitted.

Example: $X \in \mathbb{F}_2$  

- Acts independently on bits
- Capacity $= 1 - H(p)$; $H(p)$ = binary entropy!
- $H(p) = p \cdot \log \frac{1}{p} + (1 - p) \cdot \log \frac{1}{1-p}$
“Achieving” Channel Capacity

- Commonly used phrase. Many mathematical interpretations.
- Some possibilities:
  - How small can $n$ be?
  - Get $R > C - \epsilon$ with polytime algorithms?
  - [Luby et al.’95] Make running time $\text{poly} \left( \frac{n}{\epsilon} \right)$?

- Shannon ‘48: $n = O(C - R)^{-2}$
- Forney ‘66: time $= \text{poly} \left( n, 2^\frac{1}{e^2} \right)$
- Open till 2008
- Arikan’08: Invented “Polar Codes” ...
- Final resolution: Guruswami+Xia’13, Hassani+Alishahi+Urbanke’13 – Strong analysis
Context for today’s talk

- [Arikan’08]: Martingale associated to general class of codes.
  - Martingale polarizes (weakly) implies code achieves capacity (weakly)
  - Martingale polarizes weakly for entire class.
- [GX’13, HAU’13]: Strong polarization for one specific code.
- Why so specific?
  - Bottleneck: Strong Polarization
- Rest of talk: Polar codes, Martingale and Strong Polarization
Lesson 0: Compression ⇒ Coding

- Linear Compression Scheme:
  - \((H, D)\) for compression scheme for \(\text{bern}(p)^n\) if
    - Linear map \(H: \mathbb{F}_2^n \rightarrow \mathbb{F}_2^m\)
    - \(\Pr_{Z \sim \text{bern}(p)^n} [D(H \cdot Z) \neq Z] = o(1)\)
    - Hopefully \(\frac{m}{n} \leq H(p) + \epsilon\)

- Compression ⇒ Coding
  - Let \(H^\perp\) be such that \(H \cdot H^\perp = 0\);
  - Encoder: \(X \mapsto H^\perp \cdot X\)
  - Error-Corrector: \(Y = H^\perp \cdot X + \eta \mapsto Y - D(H \cdot Y) = w.p.(1-o(1)) H^\perp \cdot X\)
Question: How to compress?

- Arikan’s key idea:
  - Start with $2 \times 2$ “Polarization Transform”:
    \[
    (U, V) \rightarrow (U + V, V)
    \]
  - Invertible – so does nothing?
  - If $U, V$ independent,
    - then $U + V$ “more random” than either
    - $V \mid U + V$ “less random” than either
  - Iterate (ignoring conditioning)
    - End with bits that are almost random, or almost determined (by others).
    - Output “totally random part” to get compression!
Information Theory Basics

- **Entropy:**
  - **Defn:** for random variable $X \in \Omega$,
    \[
    H(X) = \sum_{x \in \Omega} p_x \log \frac{1}{p_x}
    \]
    where $p_x = \Pr[X = x]$
  - **Bounds:** $0 \leq H(X) \leq \log|\Omega|$

- **Conditional Entropy**
  - **Defn:** for jointly distributed $(X,Y)$
    \[
    H(X|Y) = \sum_y p_y H(X|Y = y)
    \]
  - **Chain Rule:** $H(X,Y) = H(X) + H(Y|X)$
  - **Conditioning does not increase entropy:** $H(X|Y) \leq H(X)$
    - **Equality $\iff$ Independence:** $H(X|Y) = H(X) \iff X \perp Y$

April 4, 2018
Rutgers: General Strong Polarization
10 of 21
Iteration by Example:

- \((A, B, C, D) \rightarrow (A + B + C + D, B + D, C + D, D)\)
- \((E, F, G, H) \rightarrow (E + F + G + H, F + H, G + H, H)\)
The Polarization Butterfly

\[ P_n(U, V) = \left( \frac{P_n(U + V)}{2}, \frac{P_n(V)}{2} \right) \]

Polarization \( \Rightarrow \exists S \subseteq [n] \)
\[ H(W_{[n]-S} | W_S) \rightarrow 0 \]
\[ |S| \leq (H(p) + \epsilon) \cdot n \]

Compression \( E(Z) = P_n(Z)_S \)

Encoding time \( = O(n \log n) \)
(given \( S \))

To be shown:
1. Decoding = \( O(n \log n) \)
2. Polarization!
The Polarization Butterfly: Decoding

Decoding idea:
Given \( S, W_S = P_n(U, V) \)
Compute \( U + V \sim \text{Bern}(2p - 2p^2) \)
Determine \( V \sim ? \)

Decoding Algorithm:
Given \( S, W_S = P_n(U, V), p_1, \ldots, p_n \)
- Compute \( q_1, \ldots, q_n \sim p_1 \ldots p_n \)
- Determine \( U + V = D \left( W_S^+, q_1 \ldots q_n \right) \)
- Compute \( r_1 \ldots r_n \sim p_1 \ldots p_n; U + V \)
- Determine \( V = D \left( W_S^-, r_1 \ldots r_n \right) \)

Key idea: Stronger induction!
- non-identical product distribution
Polarization? Martingale?

Idea: Follow \( X_t = H(Z_{i,t} | Z_{<i,t}) \) for random \( i \)

Martingale?

\[
\begin{align*}
Z_1 & \rightarrow Z_1 + Z_{n/2+1} \\
Z_2 & \rightarrow Z_2 + Z_{n/2+2} \\
Z_{i,0} & \rightarrow Z_{i,1} \ldots Z_{i,t} \ldots \\
Z_{n/2} & \rightarrow Z_{n/2} + Z_n \\
Z_{n/2+1} & \rightarrow Z_{n/2+1} \\
Z_{n/2+2} & \rightarrow Z_{n/2+2} \\
Z_n & \rightarrow Z_n \\
W_1 & \\
W_2 & \\
W_n/2 & \\
W_n/2+1 & \\
W_n/2+2 & \\
W_n & \\
Z_{j,t} & \rightarrow Z_{j,t+1} \\
Z_{k,t} & \rightarrow Z_{k,t+1} \\
X_t \text{ can’t distinguish } i = j \text{ from } i = k \\
H(Z_{j,t}, Z_{k,t}) = H(Z_{j,t+1}, Z_{k,t+1}) \\
\text{Chain Rule ...} \\
\Rightarrow \mathbb{E}[X_{t+1} | X_t] = X_t
\end{align*}
\]
Remaining Tasks:

1. Prove that $X_t \overset{\text{def}}{=} H(Z_{i,t} \mid Z_{<i,t})$ polarizes locally

2. Prove that Local polarization $\Rightarrow$ Strong Polarization.

- Recall
  - **Strong Polarization:** (informally)
    \[
    \Pr[X_t \in (\tau, 1 - \tau)] \leq \epsilon \text{ if } t \geq \max\{O(\log 1/\epsilon), o(\log 1/\tau)\}
    \]
    formally \[ \forall \gamma > 0 \exists \beta < 1, c \text{ s.t. } \forall t \Pr[X_t \in (\gamma^t, 1 - \gamma^t)] \leq c \cdot \beta^t \]

  - **Local Polarization:**
    - **Variance in the middle:** $X_t \in (\tau, 1 - \tau)$
      \[ \forall \tau > 0 \exists \sigma > 0 \text{ s.t. } \forall t, X_t \in (\tau, 1 - \tau) \Rightarrow \text{Var}[X_{t+1} \mid X_t] \geq \sigma \]
    - **Suction at the ends:** $X_t \not\in (\tau, 1 - \tau)$
      \[ \exists \theta > 0, \forall c < \infty, \exists \tau > 0 \text{ s.t. } X_t < \tau \Rightarrow \Pr\left[X_{t+1} < \frac{X_t}{c}\right] \geq \theta \]
Local ⇒ (Global) Strong Polarization

- **Step 1: Medium Polarization:**
  - \( \exists \alpha, \beta < 1 \Pr[ X_t \in (\beta^t, 1 - \beta^t) ] \leq \alpha^t \)
  - **Proof:** Potential \( \Phi_t = \mathbb{E}[\min\{\sqrt{X_t}, \sqrt{1 - X_t}\}] \)
  - Variance+Suction ⇒ \( \exists \eta < 1 \) s.t. \( \Phi_t \leq \eta \cdot \Phi_{t-1} \)
  - (Aside: Why \( \sqrt{X_t} \)? Clearly \( X_t \) won’t work. And \( X_t^2 \) increases!)

  \[ \Rightarrow \Phi_t \leq \eta^t \Rightarrow \Pr[ X_t \in (\beta^t, 1 - \beta^t) ] \leq \left( \frac{\eta}{\beta} \right)^t \]
Local ⇒ (Global) Strong Polarization

- **Step 1: Medium Polarization:**
  - \( \exists \alpha, \beta < 1 \) \( \Pr[X_t \in (\beta^t, 1 - \beta^t)] \leq \alpha^t \)

- **Step 2: Medium Polarization + \( t \)-more steps ⇒ Strong Polarization:

- **Proof:** Assume \( X_0 \leq \alpha^t \); Want \( X_t \leq \gamma^t \) w.p. \( 1 - O(\alpha_1^t) \)
  - Pick \( c \) large & \( \tau \) s.t. \( \Pr[X_i < \frac{X_{i-1}}{c} | X_{i-1} \leq \tau] \geq \theta \)
  - 2.1: \( \Pr[\exists i \text{ s.t. } X_i \geq \tau] \leq \tau^{-1} \cdot \alpha^t \) (Doob’s Inequality)
  - 2.2: Let \( Y_i = \log X_i \); Assuming \( X_i \leq \tau \) for all \( i \)
    - \( Y_i - Y_{i-1} \leq -\log c \) w.p. \( \geq \theta \); \& \( \geq k \) w.p. \( \leq 2^{-k} \).
    - \( \mathbb{E}[Y_t] \leq -t \left( \frac{\log c}{\theta} + 1 \right) \leq 2t \cdot \log \gamma \)
    - **Concentration!** \( \Pr[Y_t \geq t \cdot \log \gamma] \leq \alpha_2^t \) QED!
Polarization of Arikan Martingale

- Ignoring conditioning:
  \[ X_t = H(p) \Rightarrow X_{t+1} = H(2p - 2p^2) \text{ w.p. } \frac{1}{2} \]
  \[ = 2H(p) - H(2p - 2p^2) \text{ w.p. } \frac{1}{2} \]

- Variance in the middle: Continuity of \( H(\cdot) \) and separation of \( 2p - 2p^2 \) from \( p \)

- Suction at high end:
  \[ X_t = 1 - \epsilon \Rightarrow X_{t+1} = 1 - O(\epsilon^2) \text{ w.p. } \frac{1}{2} \]

- Suction at low end:
  \[ X_t = H(p) \Rightarrow X_{t+1} = H\left(\frac{p}{\log p}\right) \text{ w.p. } \frac{1}{2} \]
General Strong Polarization?

- So far: \( P_n = H_2 \otimes \ell \) for \( H_2 = \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix} \)
- What about other matrices?
  - E.g., \( P_n = H_3 \otimes \ell \) for \( H_3 = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \).
- Open till our work!
- Necessary conditions: \( H \) invertible, not-lower-triangular
- Our work: Necessary conditions suffice (for local polarization)!
Conclusion

- Simple General Analysis of Strong Polarization
- But main reason to study general polarization:
  - Maybe some matrices polarize even faster!
  - This analysis does not get us there.
  - But hopefully following the (simpler) steps in specific cases can lead to improvements.
Thank You!