Strong Direct Sum for Randomized Query Complexity

Joshua Brody
Swarthmore College

Eric Blais
University of Waterloo

Interactive Complexity Workshop
Simons Institute for the Theory of Computing
UC Berkeley 10/16/18

Direct Sum Theorems

Does computing f(x) on k copies scale with k?

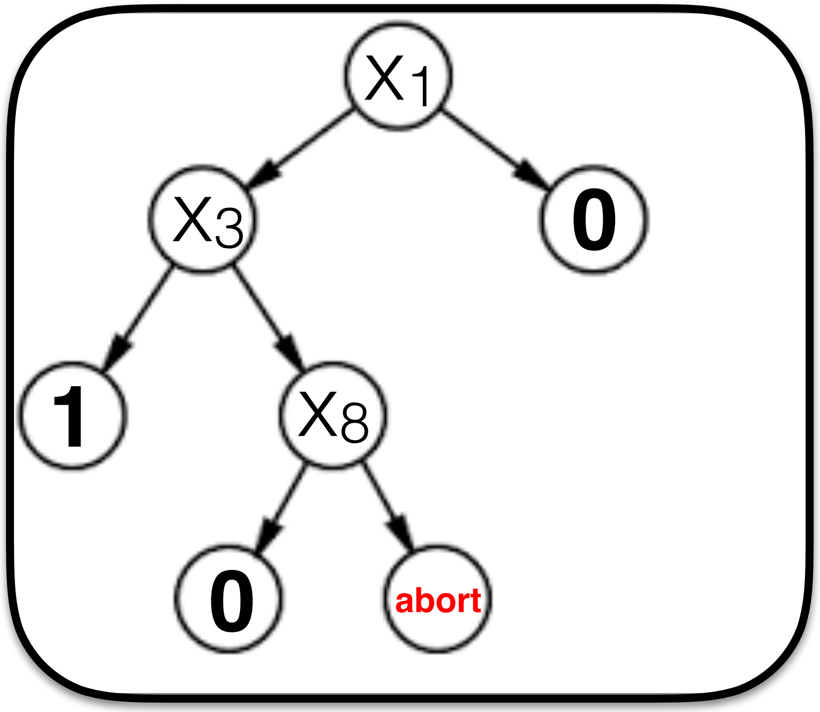
Direct Sum Theorem: Computing **k** copies of **f** requires **k** times the resources

Direct Product Theorem: Success prob. of computing **k** copies of **f** with **<<** **k** resources is $2^{-\Omega(k)}$

Strong Direct Sum: computing k copies of f(x) requires k*log(k) times the resources

Query Complexity

aka Decision Tree Complexity



Decision Tree for $f: \{0,1\}^n \to \{0,1\}$:

- internal nodes labeled w/input bits x_i
- leaves labeled w/output or **ABORT**
- **cost(T):** depth of **T** == worst-case #queries

Randomized DT:

- distribution *A* on decision trees
- **cost(*A*) = max$_T$ cost(T)**

Distributional QC $D^{\mu}_{\delta,\varepsilon}(f)$: min **cost(T)** s.t. **Pr[abort] ≤ δ** and **Pr[error] ≤ ε**

Randomized QC $R_{\delta,\varepsilon}(f)$: minimum cost of randomized algorithm s.t.

Pr[abort] ≤ δ and **Pr[error] ≤ ε**

(q, δ, ε)-algorithm: q queries, abort prob. δ, error prob. ε

Query Complexity w/aborts

Minimax Lemma: $D^{\mu}_{2\delta,2\varepsilon}(f) \le R_{\delta,\varepsilon}(f) \le D^{\mu}_{\delta/2,\varepsilon/2}(f)$

Error Reduction: $R_{O(1/t,\ O(1/t))}(f) \le O(\log(t) R_{1/2,\ 1/3}(f))$

Previous Work

[MWY13, MWY15]:

- strong direct sum for *information complexity w/aborts + error*
- applications for streaming/sketching algorithms

[Drucker12]:

- direct product theorems for randomized query complexity

[GPW15, ABBLSS17]:

- query complexity separations based on *pointer functions*
- polynomial separation $\mathbf{R_0(f)}$ vs $\mathbf{R_\varepsilon(f)}$

Theorem: Suppose any T-query algorithm computing f has success ≤ 1-*ε* under μ. Then, any (*ε*Tk)/2-query algorithm for computing f^k has success < $(1-\varepsilon/2)^k$ under μ^k **[Drucker 12]**

Our Results

Strong Direct Sum Theorem: $D^{\mu^k}_{0,\varepsilon}(f^k) = \Omega(kD^{\mu}_{1/5,40\varepsilon/k}(f))$

Scaling with ε: There is $f : \{0,1\}^N \rightarrow \{0,1\}$ such that for all $\varepsilon > 2^{-\log(N)^2}$, we have $R_{\delta,\varepsilon}(f) = \Theta(N'\log(1/\varepsilon))$

Corollary: There is f such that $R_\varepsilon(f^k) = \Omega(k\log(k)R_\varepsilon(f))$

Query-resistant codes: probabilistic encoding $G: \Sigma \rightarrow \{0,1\}^N$ such that $N/2$ bits of $G(x)$ needed to learn anything about x

Query Resistant Codes

Definition: a *δ**N**-query resistant code* of **Σ** is a set of distribs **{G(x)}**

- For each **x ϵ Σ**, **G(x)** is a distribution on $\{0,1\}^N$
- { ***support***(G(x)) : x ϵ Σ} partition $\{0,1\}^N$
- For all **S** ⊆ **[N]** with **|S| ≤ δN**, all **z ϵ** $\{0,1\}^{|S|}$ and all **x ≠ x'**,

  distributions **G(x), G(x')** conditioned on ***S-bits = z*** are equal
- *"decoding function"* **h(y) := x** iff **y ϵ support(G(x))**

Lemma: For any **Σ**, there is a **(N/2)**-query resistant code with **N = |Σ|**. Furthermore, conditional distributions

$G(x)|_{S=z}$ are uniform.

Query Resistance

For $f : \Sigma^n \to \{0,1\}$, define $F : \{0,1\}^{nN} \to \{0,1\}$ as:

$$F(y_1,\ldots,y_n) := f(h(y_1),\ldots, h(y_n))$$

Theorem: $R^{cell}_{\delta,\varepsilon}(f) \le (2/N)R_{\delta,\varepsilon}(F)$

Proof: Let A be a (q, δ, ε)-algorithm for F.

```
Algorithm B(x₁,…,xₙ) {
  emulate A(G(x₁),…,G(xₙ))
  when A queries G(xᵢ) for kth time:
    if k < N/2, sample G(xᵢ) cond. on prev. queries
    if k = N/2, sample xᵢ
    if k ≥ N/2, sample G(xᵢ) cond. on prev. history.
}
```

Functions

GapID: $\{0,1\}^n \rightarrow \{0,1\}$.

- GapID(x) = 1 if $x = 0^n$, 0 if $|x|=n/2$

Conclusion: $R_{\delta,\varepsilon}(\text{EncFcn}) \geq (N/2)\, R^{cell}_{\delta,\varepsilon}(\text{PtrFcn})$

$\geq (N/2)\, R_{\delta,2\varepsilon}(\text{BlueRed})$

$\geq (Nm/14)\, R_{\delta+0.1,2\varepsilon}(\text{GapID})$

$\geq \Omega(Nm\log(1/\varepsilon))$

- BlueRed(y) = 0 if half colored entries **Red**, half **Blue**
- theorem: $R_{\delta+0.1,\varepsilon}(\text{GapID}) \leq (7/m) * R_{\delta,\varepsilon}(\text{BlueRed})$

Theorem: $R_{0,\varepsilon}(\text{EncFcn}) = O(Nm\log(1/\varepsilon))$ [ABBLSS17]

- theorem: $R_{\delta,2\varepsilon}(\text{BlueRed}) \leq R^{cell}_{\delta,\varepsilon}(\text{PtrFcn})$

EncFcn: ***query resistant code+PtrFcn***

- theorem: $R^{cell}_{\delta,\varepsilon}(\text{PtrFcn}) \leq (2/N) R_{\delta,\varepsilon}(\text{EncFcn})$

GapID Lower Bound

Theorem: **$R_{\delta,\varepsilon}$(GapID) = $\Omega(\log(1/\varepsilon))$**

Hard Distribution **$X \sim \mu$**:

w/prob **$\alpha := \max(\delta, 1.001\varepsilon)$**, **$X=0^n$**

w/prob **$1-\alpha$**, **X** uniform on **$|x|=n/2$**

Fix (**$\log((1-\alpha)/\varepsilon)$**, δ,ε)-algorithm T for GapID

wlog output NO if **$X_i = 1$** queried

When all queries = 0:

- *abort:* **Pr[abort] > $\alpha \geq \delta$**
- *output 0:* **Pr[error] = $\alpha > \varepsilon$**
- *output 1:* **Pr[error] $\cong$ $(1-\alpha)2^{-q} > \varepsilon$**

In all cases, abort prob. > δ or error prob. > ε .

BlueRed Lower Bound

Theorem: $R_{\delta+0.1,\varepsilon}(\text{GapID}) \leq (7/m)^* R_{\delta,\varepsilon}(\text{BlueRed})$

Emulate (q,δ,ε)-algorithm A for **BlueRed**

- each colored entry in uniform row
- pick each $i_j \in [m]$ uniformly
- map **0 → Red**, **1 → Blue**
- abort if A queries > 7q/m colored entries

Claim: Pr[> 7q/m colored entries probed] ≤ 1/10.

BlueRed Lower Bound

Claim: Pr[> 7q/m colored entries probed] ≤ 1/10.

- For any column, **Pr[colored entry found on k-th query] = 1/m**
- For any leaf w/**t** colored entries found, **Pr[leaf] ≤ m^{-t}**
- there are **{q choose t}** leaves w/**t** colored entries found

$$\Pr[> 7q/m \text{ colored entries}] \le \Sigma_{t > 7q/m} \{q \text{ choose } t\} m^{-t}$$
$$\le \Sigma_{t > 7q/m} (qe/mt)^t$$
$$< \Sigma_{t > 7q/m} (e/7)^t$$
$$< 1/10.$$

PtrFcn Lower Bound

Theorem: $R_{\delta,2\varepsilon}(\text{BlueRed}) \leq R^{\text{cell}}_{\delta,\varepsilon}(\text{PtrFcn})$

Partially emulate **(q,δ,ε)**-algorithm ***A*** for **PtrFcn:**

- map **Black → [1,⊥,⊥...,⊥]**
- map **Red → [0,⊥,⊥...,⊥]**
- **Blue**: halt, output **NO**

Claim: Let **x ϵ BlueRed^{-1}(0)**. Then **Pr[no Blue entries queried] < 2ε**

Proof:

- Let **z ϵ PtrFcn^{-1}(1)** be *consistent* with **x.**
- **z' := z**, w/value of special entry = 0.
- ***A*(z)=*A*(z')** unless special entry queried.
- **Pr[no blue entries queried] ≤ Pr[special entry not queried] ≤ 2ε**

Strong Direct Sum Theorem: $D^{\mu^k}_{0,\varepsilon}(f^k) = (kD^{\mu}_{1/5,40\varepsilon/k}(f))$

Let A be an ε-**error** algorithm for f^k.

Let $y = (y_1, \ldots, y_k)$.

Embed(y,i,x) := **y**, w/i-th coord replaced by **x**.

```
Algorithm B(x) {
  carefully select y,i
  emulate A(EMBED(y,i,x))
  abort if problems found
}
```

Strong Direct Sum Theorem: $D^{\mu^k}_{0,\varepsilon}(f^k) = (kD^{\mu}_{1/5,40\varepsilon/k}(f))$

$$1\text{-}\varepsilon \le \Pr_{Y\sim\mu^k}[A(Y) = f^k(Y)] = \prod_{i=1}^{k} \Pr_{Y\sim\mu^k}[A(Y)_i = f^k(Y)_i \mid A(Y)_{<i} = f^k(Y)_{<i}]$$

- at least **2k/3 i** give $\Pr[A(Y)_i = f^k(Y)_i \mid A(Y)_{<i} = f^k(Y)_{<i}] \le 10\,\varepsilon/k$ **(1)**
- $q_i(Y)$: # queries of Y_i
- $q \ge \Sigma_i E_Y[q_i(Y)]$ $\Rightarrow$ $\ge$ **2k/3 i** have $E[q_i(Y)] \le 3q/k$ **(2)**

Fix **i*** to get **(1)** and **(2)**. **Y* := Embed(Y,i*,x).**

This **i*** satisfies:

1. $E_{Y\sim\mu k}[\ \Pr_{x\sim\mu}[A(Y^*)_{<i} \neq f^k(Y^*)_{<i}]\ ] \le \varepsilon$
2. $E_Y[\Pr_{x\sim\mu}[A(Y^*)_i \neq f^k(Y^*)_i \mid A(Y)_{<i} = f^k(Y)_{<i}] \le 10\,\varepsilon/k$
3. $E_Y[\ E_X\ [q_i(Y^*)]\ ] \le 3q/k$

Strong Direct Sum Theorem: $D^{\mu^k}_{0,\varepsilon}(f^k) = \Omega(kD^{\mu}_{1/5,40\varepsilon/k}(f))$

This **i*** satisfies:

1. $E_{Y\sim\mu^k}[\ Pr_{x\sim\mu}[A(Y^*)_{<i} \neq f^k(Y^*)_{<i}]\ ] \leq \varepsilon$
2. $E_Y[Pr_{x\sim\mu}[A(Y^*)_i \neq f^k(Y^*)_i \mid A(Y)_{<i} = f^k(Y)_{<i}] \leq 10\ \varepsilon/k$
3. $E_Y[\ E_X\ [q_i(Y^*)]\ ] \leq 3q/k$

Markov Inequality: there is **y*** such that

1. $Pr_{x\sim\mu}[A(Y^*)_{<i} \neq f^k(Y^*)_{<i}] \leq 4\varepsilon$
2. $Pr_{x\sim\mu}[A(Y^*)_i \neq f^k(Y^*)_i \mid A(Y)_{<i} = f^k(Y)_{<i}] \leq 40\ \varepsilon/k$
3. $E_X\ [q_i(Y^*)] \leq 12q/k$

```
Algorithm B(x) {
  z := EMBED(y*,i*,x)
  emulate A(z)
  abort if q_{i*}(z) > 120q/k
  abort if A(z)_{<i*} ≠ f^k(z)_{<i*}
}
```

abort probability: $1/10 + 4\varepsilon < 1/5$

error probability: $40\varepsilon/k$

Open Problems

1. Give a more efficient *query resistant code*
2. *Characterize* functions robust to **aborts**
3. **Strong Direct Sum** for Composed Functions
4. How does $\mathbf{R}_{\delta,\varepsilon}(\mathbf{f})$ compare to other QC measures?

Thanks!

slide of common stuff

$D^{\mu}_{\delta,\varepsilon}$ $R_{\delta,\varepsilon}$ [Y90], [HG91], [BT94]

$D^{\mu}_{\delta,\varepsilon}$ $R_{\delta,\varepsilon}$ $\delta\varepsilon\epsilon\mu\geq \rightarrow \leq\Sigma\Theta\Omega\Omega\Omega\perp\alpha$

Fact: If **f** $\in$ **ACC⁰** then **f** has NOF protocol with **poly(log n)** communication and **k = poly(log n)** players

$R_{\delta,\varepsilon}$(GapID)