Amortized circuit complexity

Formal complexity measures

and catalytic algorithms

Robert Robere
McGill

Jeroen Zuiddam
nyu

Amortized circuit complexity, formal complexity measures, and catalytic algorithms

Robert Robere & Jeroen Zuiddam

1. Direct sum problems
2. Strassen duality
3. Boolean formulas and formal compl. measures
4. Amortized circuit complexity
5. First result: duality
6. Second result: catalytic circuits
7. Third result: catalytic space
8. Proof ideas

Direct sum problems

Is the fastest way to solve n instances of some computational task T, to run the fastest algorithm for 1 instance n times?

Or, can we achieve economy of scale, and compute all n instances faster as a group?

$$\lim_{n \to \infty} \frac{\text{cost}(nT)}{n} \stackrel{?}{=} \text{cost}(T)$$

Everywhere in complexity theory, CS, Math, Physics.

- Shannon's source coding theorem

$$\text{Alice} \xrightarrow{\text{code}(M_i)} \text{Bob}$$

$$M_1, \_, M_n \sim \mathcal{M}$$

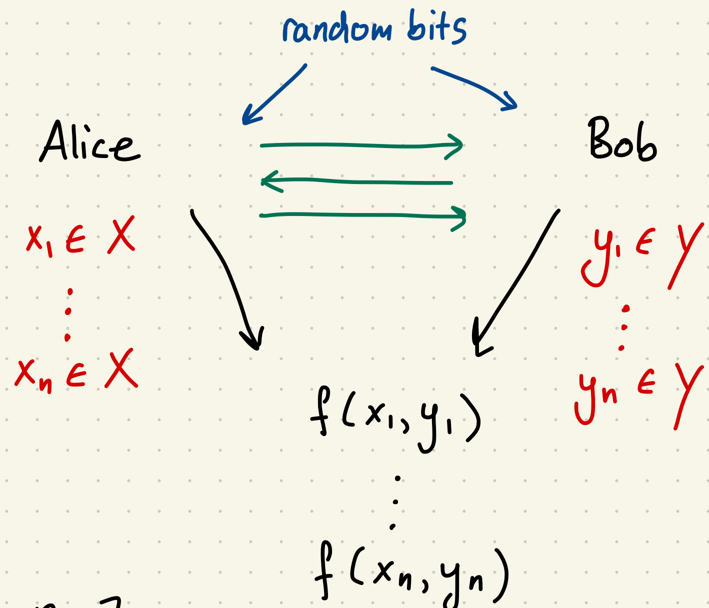
Theorem

$$\boxed{\text{One-way, amortized communication cost of sending a random message } \mathcal{M} \text{ is exactly the Shannon entropy } H(\mathcal{M})}$$

• Amortized Randomized communication

also deeply studied.

$f: X \times Y \to \{0,1\}$



Theorem [Braverman–Rao]

$\boxed{\text{Amortized Randomized communication} = \text{information complexity}}$

- Direct sum problems in disguise,

  ex: matrix multiplication

What is the smallest $\omega \in \mathbb{R}$ (infimum) such that two $n \times n$ matrices can be multiplied using $\mathcal{O}(n^{\omega})$ operations?

- Known that $2 \leq \omega \leq 2.37$ [Strassen (1969), ..., Le Gall (2014), Alman-Williams (2020)]

- No direct sum flavour ?!

- ... except, 2^{ω} is exactly the asymptotic tensor rank of a certain tensor!

→ Tensor rank crash course

- a k-tensor is a k-dimensional array over a field $\mathbb{F}$
- ↳ simple if it is the tensor product of k vectors

  ↑ outer product
- Tensor rank $R(A) = \min\{r \cdot A \text{ is the sum of } r \text{ simple tensors}\}$
- Asymptotic tensor rank $\tilde{R}(A) = \lim_{m\to\infty} R(A^{\otimes m})^{1/m}$

  ↑ m-fold kronecker product

  Matrices ↓
- 2-tensors : $R(A \otimes B) = R(A)\,R(B)$
- k-tensors with $k > 2$: only $\leq$

Theorem [Gartenberg 85]

$\boxed{\text{There is a 3-tensor } A \text{ such that } \tilde{R}(A) = 2^{\omega}}$

Strassen duality

For matrices A, B write $A \leq_T B$ if there are matrices U, V such that $A = UBV$.

For k-tensors this preorder is defined analogously: $A = (U_1, U_2, \ldots, U_k) \cdot B$

Defn. [Strassen 86-88]

Let X be the collection of all $\mu : \{\text{tensors}\} \to \mathbb{R}_{\geq 0}$ so that

- μ is $\otimes$-multiplicative and $\oplus$-additive
- μ is $\leq_T$-monotone
- μ is normalized to n on the diagonal tensor of size n

Theorem [S] $\boxed{\tilde{R}(T) = \max_{\mu \in X} \mu(T)}$

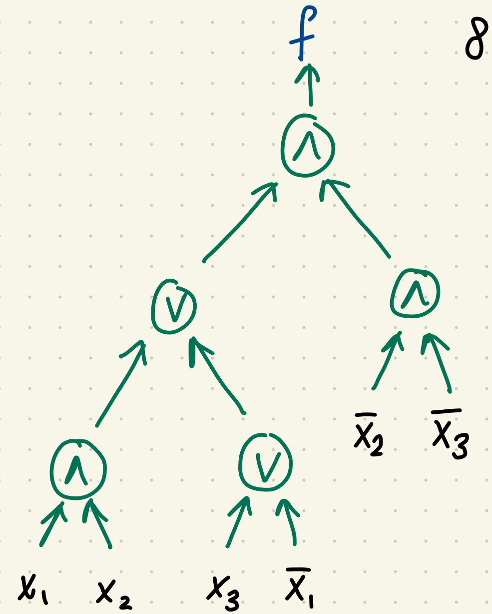
→ General theory, applied to:

- Shannon capacity [Z19]
- Sunflowers, cap sets, ...

→ To understand matrix multiplication it suffices to understand X!

Boolean formulas

$\hookrightarrow$ = tree-like Boolean circuit



Proving lower bounds on Boolean formula size $F(f)$ is a long-standing open problem

A formal complexity measure is a map

$$\mu : \{\text{boolean functions}\} \to \mathbb{R}_{\geq 0}$$

such that

- μ is monotone wrt $\wedge, \vee$:

$$\mu(f \wedge g) \leq \mu(f) + \mu(g), \quad \mu(f \vee g) \leq \mu(f) + \mu(g)$$

- μ is normalized on literals :

$$\mu(x_i), \mu(\bar{x}_i) \leq 1$$

Theorem [Folklore] $\boxed{\text{For any } f \quad C(f) = \max_{\mu} \mu(f)}$

Proof :

- $\mu(f) \leq C(f)$ by induction
- C is itself a formal compl. meas.

Strassen duality vs. Complexity measures?

[Strassen]

$$\boxed{\tilde{R}(T) = \max_{\mu} \mu(T)}$$

tensor

- $\leq_T$ - monotone
- normalized on diagonal tensors
- multipl., add.

[Folklore]

$$\boxed{F(f) = \max_{\mu} \mu(f)}$$

boolean function

- monotone wrt $\wedge, \vee$
- normalized on literals

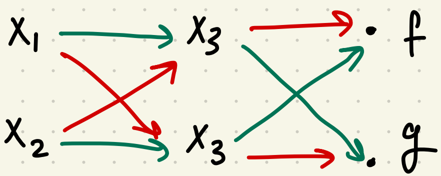
Coincidence? No!

Amortized circuit complexity

$G = \{\text{finite gate set}\}$

- Allow multiple inputs/outputs, different costs for different gates

Ex: Branching programs



$$G = \begin{cases} \bullet \text{ OR-gate} \quad (\text{free}) \\ \bullet \text{ query gate} : f \mapsto (f \wedge x_i, f \wedge \bar{x}_i) \quad (\text{cost } 2) \end{cases}$$

$f = 1$ iff there is a path from <u>some</u> source to the sink for f

multiset of boolean funcs.

Defn. $\boxed{C_G(F) = \text{minimum cost of } G\text{-circuit computing } F = \{\{f_1, \ldots, f_n\}\}}$

Ex. $C_G(\{\{f, g\}\}) \leq 8$

First result: Duality theorem

Defn. $\tilde{C}_G(f)$ = amortized G-circuit complexity $= \lim_{m \to \infty} \frac{C_G(m \cdot f)}{m}$

multiset with m copies of f.

Defn. A G-complexity measure is a function $\mu : \{\text{boolean functions}\} \to \mathbb{R}_{\geq 0}$ such that

- G-gate monotone: if there is a G-gate: $(f_1, \ldots, f_n) \mapsto (g_1, \ldots, g_m)$ with cost c, then $\mu(g_1) + \cdots + \mu(g_m) \leq \mu(f_1) + \cdots + \mu(f_n) + c$
- Normalized: $\mu(\ell) \leq 1$

literal.

Theorem

$$\boxed{\tilde{C}_G(f) = \max_{\mu} \mu(f)}$$

gate set — bool. func.

Application: submodular measures and comparator circuits

Def

$\mu : \{\text{bool. func.}\} \to \mathbb{R}_{\geq 0}$

- $\mu(f \wedge g) + \mu(f \vee g) \leq \mu(f) + \mu(g)$
- $\mu(x_i), \mu(\bar{x}_i) \leq 1$

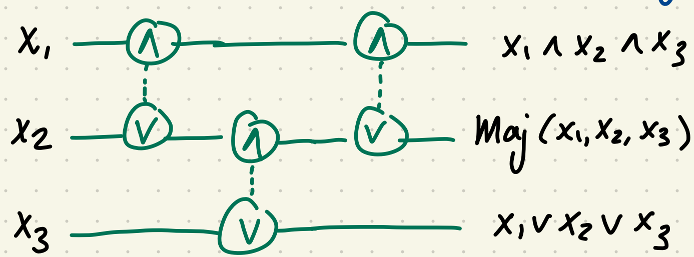
Introduced by Razborov

Theorem [Raz92] $\boxed{\mu(f) \leq O(n)}$

bool. func. on n vars.

Duality $\Longrightarrow$

comparator gate: $(f, g) \mapsto (f \wedge g, f \vee g)$

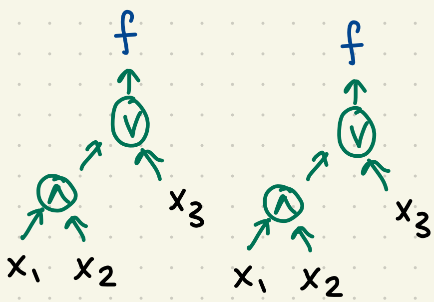


Amortized comp. circuit size is at most $O(n)$.

Rem. $\boxed{\text{Also Potechin [Pot17].}}$

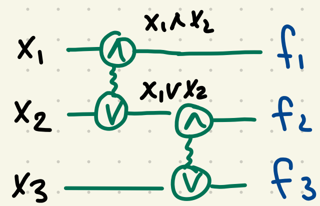
Brief recap

formulas



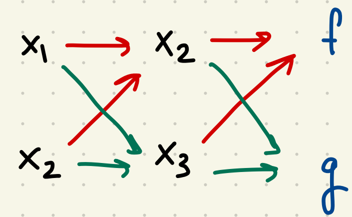
- $\mu(f \wedge g) \le \mu(f) + \mu(g)$
- $\mu(f \vee g) \le \mu(f) + \mu(g)$
- $\mu(\ell) \le 1$

comparator circuits



- $\mu(f \wedge g) + \mu(f \vee g) \le \mu(f) + \mu(g)$
- $\mu(\ell) \le 1$

branching programs



- $\mu(f \wedge x_i) + \mu(f \wedge \bar{x}_i) \le \mu(f) + 2$
- $\mu(\ell) \le 1$

$$\boxed{\max_{\mu} \mu(f) = \tilde{C}_G(f)}$$

Second result: catalytic circuit complexity

bool. func

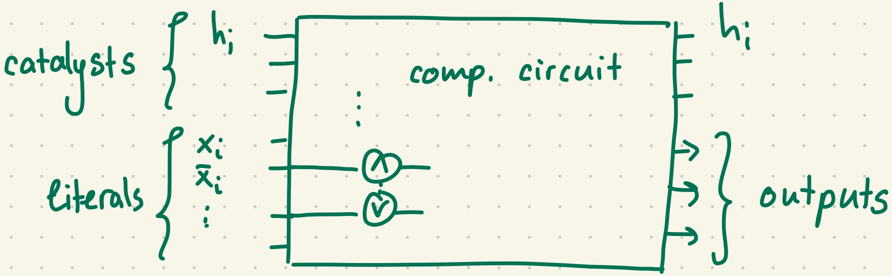
$$\mu(f) \leq \mu(g) \quad \Longleftrightarrow \quad \mu(f) + \mu(h) \leq \mu(g) + \mu(h)$$

submodular measure, say

any bool. f.

Def. A catalytic comparator circuit is a comparator circuit C which, besides literals, takes bool. functions h_i as input and produces another copy of h_i as outputs.

↳ catalyst



theorem

$$\boxed{\tilde{C}_G(f) \leq C_{G,cat}(f) \leq C_G(f)}$$

$$\parallel$$

optimal integral solution to some LP

Third result: $\underbrace{\text{Catalytic space}}$ Sounds similar ...

Def [Buhrman, Cleve, Koucký, Loff, Speelman 2014]

A catalytic space TM has an extra catalytic tape that starts with arbitrary content, and can be much longer than the work tape. At the end of the computation, the catalytic tape must be restored to the original content.

$TC^1 \subseteq$ catalytic log space.

Catalytic circuits

There exist catalysts h_i that can be used by the circuit, and the h_i must be reproduced.

Circuit can depend on the catalyst.

Catalytic space

For all catalytic tape contents, the TM computes the function with the catalytic tape restored.

TM cannot depend on content.

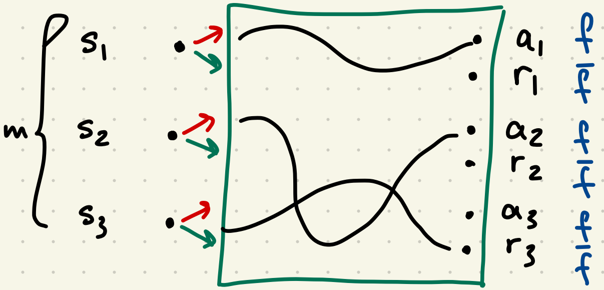
Open problem: how related

Our duality does not characterize this

We can traslate some new results proved with our new duality to catalytic space!

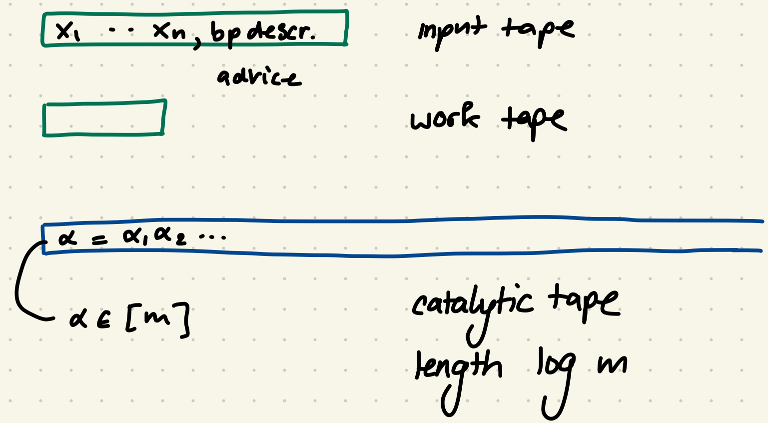
Def (non-uniform catalytic space) [Girard, Koucky, McKenzie 2015]

An m-catalytic branching program for f is a branching program with m start, accept, reject nodes, such that: for every $x \in \{0,1\}^n$ the computation path from the ith start node, ends at the ith $\underbrace{\text{accept}}_{f(x)=1}$ or $\underbrace{\text{reject}}_{f(x)=0}$ node.



Stronger than amortized BP.

TM with catalytic tape



Question [Gerard, Koucky, McKenzie]

For which boolean functions is m-catalytic branching program size smaller than branching program size (on average)

Potechin [2017] Every f has m-catalytic BP of size $O(mn)$

$(m = 2^{2^n})$

Theorem

For every f there is an m-catalytic BP computing f of size $O(mn)$ where $m = 2^{\binom{n}{\leq d}}$ and $d = \deg_2 f$.

- translate a similar result that we prove using our duality.
- Exploits symmetry heavily.

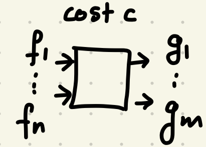
Proof idea of duality

Theorem

$$\boxed{\tilde{C}_G(f) = \max_{\mu} \mu(f)}$$

gate set — bool. func.

- G-gate monotone:



$\Rightarrow c + \sum_i \mu(f_i) \geq \sum_i \mu(g_i)$.

- normalized: $\mu(\text{literal}) \leq 1$

$\underbrace{\qquad\qquad\qquad}$

↓ gate ineq.

$\{f, g\} \geq_{cc} \{f \vee g, f \wedge g\}$, $\{1\} \geq_{cc} \{\ell\}$

notation: $r \vdash f$: r produces f

$f \vdash r$: r consumes f

Ex: Comparator circuits

$$\max \quad \mu(f)$$

subject to $\mu(g \vee h) + \mu(g \wedge h) \leq \mu(g) + \mu(h) \quad \forall g, h$

$\mu(\ell) \leq 1 \quad \forall$ literal ℓ

~ LP dual ~>

$$\min \quad \sum_r \text{cost}(r)\, y(r)$$

subject to

$$\sum_{r \vdash g} y(r) \geq \sum_{g \vdash r} y(r)$$

$$\sum_{r \vdash f} y(r) \geq \sum_{f \vdash r} y(r) + 1$$

$$y \geq 0$$

→ y encodes amortized circuit!

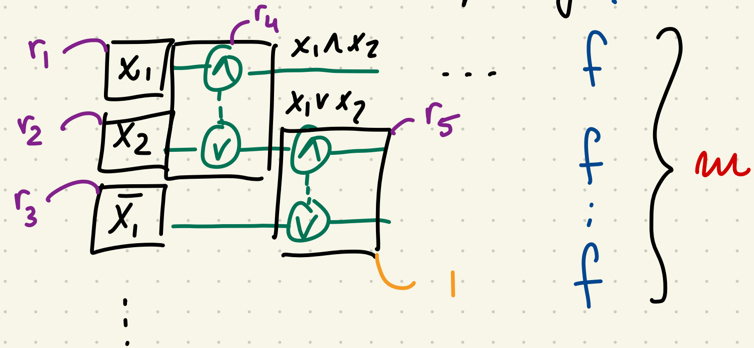
$$\sum_{r \vdash g} y(r) \geq \sum_{g \vdash r} y(r)$$

$$\sum_{r \vdash f} y(r) \geq \sum_{f \vdash r} y(r) + 1$$

$$y \geq 0$$

clear $\Longleftarrow$

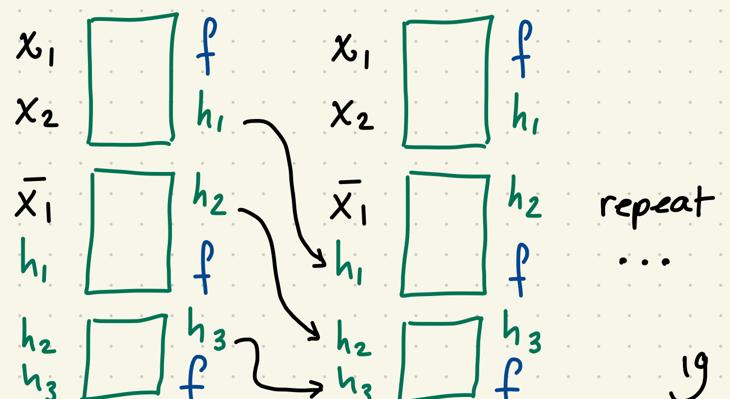
amortized circuit computing f



$$y(r) = \frac{\#\text{ occurrences of } r}{m}$$

$\Rightarrow$

- $y(r) \in \mathbb{Q}_{\geq 0}$
- $\exists n, \forall r,\ n\, y(r) \in \mathbb{N}$
- build <u>massively catalytic</u> circuit:
- <u>boost</u> catalytic to amortized:



Our duality

e.g. multisets of bool. f

- semigroup $(S, +)$
- good preorder
  - finitely generated
  - ..

gate preorder

$G \leq F$ means G is "computable" from F.

Strassen duality

- semiring $(S, +, \cdot)$
- good preorder
  - not necess. fin. gen

  $\rightarrow$ needed for graphs, tensors

Proof ideas for catalytic space

Theorem

$$\boxed{\text{For every } f \text{ there is an } m\text{-catalytic BP computing } f \text{ of size } O(m\,n) \text{ where } m = 2^{\binom{n}{\le d}} \text{ and } d = \deg_2 f.}$$

submodular

Razborov [92] $\left[\mu(f) \le O(n)\right]$

bool. func. on n vars.

same distr

Proof relies on symmetry: $\left[\, f \sim (f_0 \wedge x_n) \vee (f_1 \wedge \bar{x}_n) \sim (f_1 \wedge x_n) \vee (f_0 \wedge \bar{x}_n) \right]$

unif random on n vars

unif random on n−1 vars

symmetric BP measure: $\mu(f) = \mu(f^{\oplus i})$

Lemma $\left[\mu(f) \le 2 \cdot D_{avg}(f)\right]$

Ex: $D_{avg}(\mathrm{AND}) = O(1)$

average decision tree depth

à la Potechin

Theorem $\left[\mu(\mathrm{Orb}(f)) \le 2 \cdot |\mathrm{Orb}(f)| \cdot D_{avg}(f)\right]$ $\xrightarrow{\text{technical}}$

- $\mathrm{span}_{\mathbb{F}_2}(\mathrm{orb}(f))$
- implement as cat BP.

Conclusion

- What other direct sum problems can we express?
  - Information = Randomized Comm.?
  - Query compl?
  - ⋮
- Can the catalytic space bound be further improved?
- Relating different preorders?

⋮