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?

\[
\lim_{{n \to \infty}} \frac{\text{cost} \, (nT)}{n} = \text{cost}(T)
\]

Everywhere in complexity theory, CS, Math, Physics.
Shannon's source coding theorem

\[ \text{code}(M_i) \]

\[ M_i, \ldots, M_n \sim M \]

Theorem

One-way, amortized communication cost of sending a random message \( M \) is exactly the Shannon entropy \( H(M) \)
• Amortized Randomized communication
  also deeply studied.

$$f : x \times y \rightarrow \{0, 1\}$$

Alice

- $$x_i \in X$$
- $$x_n \in X$$

random bits

random bits

Bob

- $$y_i \in Y$$
- $$y_n \in Y$$

$$f(x_i, y_i)$$
- $$f(x_n, y_n)$$

Theorem [Braverman-Rao]

Amortized Randomized communication = information complexity
• Direct sum problems in disguise, ex: matrix multiplication

\[ \inf \]

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

• Known that \( 2 \leq w \leq 2.37 \) [Strassen, ..., Le Gall, Alman-Williams]

• No direct sum flavour ?!

... except, \( 2^w \) 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 $\mathbb{F}$.
- It is simple if it is the tensor product of $k$ vectors.
- Tensor rank $R(A) = \min \{ r : A \text{ is the sum of } r \text{ simple tensors} \}$.
- Asymptotic tensor rank $\tilde{R}(A) = \lim_{m \to \infty} R(A \otimes m)^{1/m}$.

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

Theorem [Gartenberg 85]

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

Strassen duality

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

For $\mathcal{E}$-tensors this preorder is defined analogously: $A = (U_1, U_2, \ldots, U_d) \cdot B$

Defn. [Strassen 86-88]

Let $X$ be the collection of all $\mu : \mathcal{E}$-tensors $\rightarrow \mathbb{R}_{\geq 0}$ so that

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

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

$\rightarrow$ General theory, applied to:
- Shannon capacity [$\mathbb{Z}19$]
- Sunflowers, cap sets, ...

$\rightarrow$ To understand matrix multiplication it suffices to understand $X$!
Boolean formulas

$L = \text{tree-like Boolean circuit}$

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

A formal complexity measure is a map

$\mu : \{\text{boolean functions}\} \rightarrow \mathbb{R}_{>0}$

such that

- $\mu$ is monotone wrt $\land, \lor$:
  
  $\mu(f \land g) \leq \mu(f) + \mu(g)$, $\mu(f \lor g) \leq \mu(f) + \mu(g)$

- $\mu$ is normalized on literals:
  
  $\mu(x_i), \mu(\overline{x_i}) \leq 1$

Theorem [Folklore] [For any $f$]

$C(f) = \max \mu(f)$

- $C$ is itself a formal complexity measure.
Strassen duality vs. Complexity measures?

\[ \tilde{R}(T) = \max_{\mu} m(T) \]

- tensor
- \( \leq_T \)-monotone
- normalized on diagonal tensors
- multipl., add.

\[ F(f) = \max_{\mu} m(f) \]

- 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

\[
\begin{align*}
&\begin{array}{ccc}
  x_1 & \rightarrow & x_2 \\
  & \rightarrow & x_3 \\
\end{array} &
\cdot f \\
\begin{array}{ccc}
  x_2 & \rightarrow & x_3 \\
\end{array} &
\cdot g
\end{align*}
\]

\[
G = \{ \text{or-gate (free)} \}
\]

\[
\{ \text{query gate}: f \mapsto (f \land x_i, f \land \overline{x_i}) \} \quad \text{(cost 2)}
\]

\[
f = 1 \text{ iff there is a path from some source to the sink for } f.
\]

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

**Ex.** $C_G(\{ f_1, f_2, f_3 \}) \leq 8$
First result: Duality theorem

Defn. \( \widehat{C}_G(f) = \text{amortized } G\text{-circuit complexity} = \lim_{m \to \infty} \frac{C_G(m^*\cdot f)}{m} \text{ multiset with } m \text{ copies of } f. \)

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: \( (f_1, \ldots, f_n) \to (g_1, \ldots, g_m) \) with cost \( c \), then \( \mu(g_1) + \cdots + \mu(g_m) \leq \mu(f_1) + \cdots + \mu(f_n) + c \)
- Normalized: \( \mu(f) \leq 1 \)

Theorem \( \widehat{C}_G(f) = \max_{\mu} \mu(f) \)
**Application:** submodular measures and comparator circuits

**Definition:**
\[ \mu : \mathcal{F}_{\text{bool. func.}} \to \mathbb{R}_{\geq 0} \]
- \( \mu(f \land g) + \mu(f \lor g) \leq \mu(f) + \mu(g) \)
- \( \mu(x_i), \mu(\bar{x}_i) \leq 1 \)

Introduced by Razborov

**Theorem [Raz'92]**
\[ \mu(f) \leq O(n) \]

\[ \text{bool. func. on } n \text{ vars.} \]

**Remark:** Also Potechin [Pot'17].

**Comparator Gate:**
\[ (f, g) \mapsto (f \land g, f \lor g) \]

\[ \begin{align*}
&x_1 \\
&x_2 \\
&x_3 \\
\end{align*} \]

- Comparator gates:
  - \( x_1 \land x_2 \land x_3 \)
  - \( \text{Maj}(x_1, x_2, x_3) \)
  - Amortized comp. circuit size is at most \( O(n) \).
Brief recap

**Formulas**

\[ \mu(f \land g) \leq \mu(f) + \mu(g) \]
\[ \mu(f \lor g) \leq \mu(f) + \mu(g) \]
\[ \mu(\ell) \leq 1 \]

**Comparator circuits**

\[ x_1 \quad x_2 \quad f_1 \]
\[ x_2 \quad x_1 \lor x_2 \quad f_2 \]
\[ x_3 \quad x_1 \land x_2 \quad f_3 \]

\[ x_1 \quad x_2 \quad x_3 \quad q \]

**Branching programs**

\[ \mu(f \land x_i) + \mu(f \land \overline{x_i}) \]
\[ \leq \mu(f) + 2 \]
\[ \mu(\ell) \leq 1 \]

\[ \max \mu(f) = \widehat{C_G}(f) \]
Second result: catalytic circuit complexity

\[
\mu(f) \leq \mu(g) \quad \iff \quad \mu(f) + \mu(h) \leq \mu(g) + \mu(h)
\]

\[\text{any bool. } f.\]

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.

Theorem

\[ \tilde{C}_G(f) \leq C_{G, \text{cat}}(f) \leq C_G(f) \]

optimal integral solution to some LP
Third result: Catalytic space

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.

\[ \text{TC}^1 \subseteq \text{catalytic log space}. \]

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.

Our duality does not characterize this.

Open problem: how related?
We can translate 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 $i$th start node, ends at the $i$th accept or reject node.

\[ f(x) = 1 \quad f(x) = 0 \]

---

**TM with catalytic tape**

- $x_1 \ldots x_n$, bp descr.  
- input tape  
- advice  
- work tape

\[ \alpha = \alpha_1, \alpha_2, \ldots \]

- $\alpha \in [m]$  
- catalytic tape  
- length $\log m$

---

Stronger than amortized BP.
Question [Gerard, Koucky, McKenzie]

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

Potechin [2017] \[
\text{Every } f \text{ has } m \text{-catalytic BP of size } O(mn) \quad (m = 2^n).
\]

Theorem

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

- Translation: a similar result that we prove using our duality.
- Exploits symmetry heavily.
Proof idea of duality

**Theorem**

\[
\begin{align*}
\hat{C}_G(f) = \max_{\mu} \mu(f)
\end{align*}
\]

gate set \mapsto \text{bool. func.}

Ex: Comparator circuits

\[
\begin{align*}
\max \mu(f) \\
\text{subject to } \mu(gv_h) + \mu(gr_h) \\
\leq \mu(g) + \mu(h) \quad \forall g, h \\
\mu(l) \leq 1 \quad \forall \text{literal } l
\end{align*}
\]

.G-gate monotone: \[ \forall f, g \ ; \ 
\begin{align*}
\sum_i \mu(f_i) \\
\Rightarrow \sum_i \mu(g_i)
\end{align*}
\]

normalized: \[ \mu(\text{literal}) \leq 1 \]

\[ \downarrow \text{ gate ineq.} \]

\[ \begin{align*}
\forall f, g \ ; \ 
\begin{align*}
\sum_i \text{cost}(r) y(r) \\
\text{subject to } \sum_{r \rightarrow g} y(r) \geq \sum_{g \rightarrow r} y(r) \\
\sum_{r \rightarrow f} y(r) \geq \sum_{f \rightarrow r} y(r) + 1 \\
y \geq 0
\end{align*}
\end{align*}
\]

\[ \rightarrow y \text{ encodes amortized circuit!} \]
\[
\sum_{r \in g} y(r) \geq \sum_{g \vdash r} y(r)
\]

\[
\sum_{r \in f} y(r) \geq \sum_{f \vdash r} y(r) + 1
\]

\( y \geq 0 \)

\[
\text{amortized circuit computing } f
\]

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

\[\Rightarrow\]

- \( y(r) \in \mathbb{Z}_{\geq 0} \)
- \( \exists n, \forall r, \; n \cdot y(r) \in \mathbb{N} \)
- build \underline{massively catalytic circuit}:
- boost catalytic to \underline{amortized}:

\[
\begin{array}{c}
\text{repeat} \\
\end{array}
\]

\[
\begin{array}{c}
x_1 \quad f \\
h_1 \quad f \\
\text{(x_1, h_1, f)} \\
\end{array}
\]

\[
\begin{array}{c}
x_2 \quad h_1 \\
\text{(x_2, h_1)} \\
\end{array}
\]

\[
\begin{array}{c}
\bar{x}_1 \quad h_2 \\
h_1 \quad f \\
\text{(\bar{x}_1, h_1, h_2, f)} \\
\end{array}
\]

\[
\begin{array}{c}
x_1 \quad f \\
h_1 \quad f \\
\text{(x_1, h_1, f)} \\
\end{array}
\]

\[
\begin{array}{c}
x_2 \quad h_1 \\
\text{(x_2, h_1)} \\
\end{array}
\]

\[
\begin{array}{c}
\bar{x}_1 \quad h_2 \\
h_1 \quad f \\
\text{(\bar{x}_1, h_1, h_2, f)} \\
\end{array}
\]

\[
\begin{array}{c}
x_1 \quad f \\
h_1 \quad f \\
\text{(x_1, h_1, f)} \\
\end{array}
\]

\[
\begin{array}{c}
x_2 \quad h_1 \\
\text{(x_2, h_1)} \\
\end{array}
\]

\[
\begin{array}{c}
\bar{x}_1 \quad h_2 \\
h_1 \quad f \\
\text{(\bar{x}_1, h_1, h_2, f)} \\
\end{array}
\]

\[
\begin{array}{c}
x_1 \quad f \\
h_1 \quad f \\
\text{(x_1, h_1, f)} \\
\end{array}
\]

\[
\begin{array}{c}
x_2 \quad h_1 \\
\text{(x_2, h_1)} \\
\end{array}
\]
Our duality

- semigroup \((S, +)\)
- good preorder
  - finitely generated
  - \(\leq\) gate preorder

\(G \leq F\) means \(G\) is "computable" from \(F\).

Strassen duality

- semiring \((S, +, \cdot)\)
- good preorder
  - not necess. fin. gen
  - \(\leq\) needed for graphs, tensors
Proof ideas for catalytic space

Theorem

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

Razborov \([92]\) \[ \mu(f) \leq O(n) \] \( \uparrow \) \( \subseteq \) bool. func. on \( n \) vars.

Proof relies on symmetry:

\[
\begin{align*}
\mu(f) &\leq 2 \cdot \text{Darg}(f) \\
\mu(\text{Orb}(f)) &\leq 2 \cdot |\text{Orb}(f)| \cdot \text{Darg}(f)
\end{align*}
\]

Ex: \( \text{Darg} \) (AND) = \( O(1) \)

\( \Rightarrow \) \text{average decision tree depth} \( \Rightarrow \) \text{technical}

à la Potechin

Lemma \[ \mu(f) = \mu(f^x) \] unif random on \( n \) vars

Symmetric BP measure: \( \mu(f) = \mu(f^x) \) unif random on \( n \) vars

Span \( \text{orb}(f) \) \( \Rightarrow \) 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?