Quantum Coupon Collector

Ronald de Wolf

Joint with Srinivasan Arunachalam (IBM),
Aleksandrs Belovs (Riga), Andrew Childs (Maryland),
Robin Kothari (Microsoft), Ansis Rosmanis (Nagoya)
Collecting coupons

- Every year Albert Heijn hands out cards of soccer players (“voetbalplaatjes”). You get a random one for each 2.50€ you spend on groceries.

- There are 18 teams, 11 players per team: 198 different cards.

- Your nephew really wants to have a complete set. How much money do you need to spend to get each card?

- Obviously, at least $198 \times 2.5 = 495€$

- But it’s worse: if you already have a copy of most of the cards, then your next 2.5€ spending will likely give you a card that you already have.
Analysis of coupon collector

- Suppose you already have \( i \) out of \( k \) coupons, and you get another, uniformly random coupon.

\[
\Pr[\text{see new coupon}] = \frac{k - i}{k}
\]

- Total expected number of samples:

\[
\sum_{i=0}^{k-1} \frac{k}{k - i} = k \sum_{j=1}^{k} \frac{1}{j} = k \ln(k) + \Theta(k)
\]

- With \( k = 198 \), that’s 1162 samples. Need to spend 2905€!

- You’re unlikely to finish much earlier. Variance in number of samples is \( \Theta(k^2) \), so typically you need \( k \ln(k) \pm O(k) \) samples.
Can quantum help somehow?

- Suppose get “quantum samples” instead of random samples. You want to learn unknown set $S \subseteq [n]$ of size $k$ from states

\[ |S\rangle = \frac{1}{\sqrt{k}} \sum_{i \in S} |i\rangle \]

- If you measure, you get a uniformly random sample from $S$, and we know $k \ln(k)$ random samples needed to learn $S$

- Maybe there’s something smarter, using fewer copies of $|S\rangle$?
  - **yes** if the number of “missing items” ($m = n - k$) is small
  - **no** otherwise
Learning $S$ by sampling the missing elements

$|S\rangle = \sqrt{\frac{k}{n}} |[n]\rangle + \sqrt{\frac{m}{n}} |\psi\rangle$

where $|\psi\rangle = \sqrt{\frac{m}{n}} |S\rangle - \sqrt{\frac{k}{n}} |\bar{S}\rangle \approx -|\bar{S}\rangle$ if $m \ll n$

If we measure one copy of $|S\rangle$ with 2-outcome measurement $|[n]\rangle\langle[n]|$ vs $I - |[n]\rangle\langle[n]|$, then with prob $\frac{m}{n}$ we obtain $|\psi\rangle$

Thus we can convert an expected number of $\frac{n}{m}$ copies of $|S\rangle$ into one copy of $|\bar{S}\rangle$ (up to small error).

Now we can sample uniformly from the complement of $S$!

$\bar{S}$ has $m$ elements, so $O(m \log(m + 1))$ copies of $|\bar{S}\rangle$ suffice.

Hence $O(n \log(m + 1))$ copies of $|S\rangle$ suffice for a “quantum coupon collector”. For $m = O(1)$ and $k = n - O(1)$, this beats classical coupon collector by a log-factor.
Matching lower bound on number of copies of $|S\rangle$

- Claim: you need $T = \Omega(k \log(m + 1))$ copies of $|S\rangle$
  to learn the $k$-set $S \subseteq [n]$ (assume $m \leq n/2$)

- Approach: Use the adversary lower bound (without queries!)
  A learner should do state transformation: $|S\rangle^{\otimes T} \mapsto S$
  Consider Gram matrix $M_{SS'} = (\langle S|S'\rangle)^T$; and $F_{SS'} = 1 - \delta_{SS'}$
  State transformation problem can be solved iff $\gamma_2(M \circ F)$ is small

- Can witness $\gamma_2(M \circ F) \geq 1/2$ via an adversary matrix:
  $$\gamma_2(M \circ F) = \max_{\|\Gamma\| \leq 1} \| \Gamma \circ M \circ F \|$$
  How to construct such $\Gamma$? Note that $M, F$-entries only depend on $|S \cap S'|$, so we need math that respects this symmetry.
Using the Johnson association scheme

- Define Boolean matrices $A_0, \ldots, A_m$ of dimension $N = \binom{n}{k}$, with $(A_j)_{SS'} = 1$ iff $|S \cap S'| = k - j$

- $\exists$ pairwise-orthogonal projectors $E_0, \ldots, E_m$ spanning the same space: $A_i = \sum_{j=0}^{m} p_i(j) E_j$, $E_j = \frac{1}{N} \sum_{i=0}^{m} q_j(i) A_i$, $E_i \circ E_j = \frac{1}{N} \sum_{\ell=0}^{m} q_{i,j}(\ell) E_\ell$. These parameters are known.

- $M_{SS'}$ entries only depend on $|S \cap S'|$, so we can write $M$ as linear combination of $A_i$s and hence of $E_j$s

- Adversary matrix $\Gamma = \sum_{j=0}^{m} \gamma_j E_j$, with $\gamma_0 = \cdots = \gamma_{m-1} = 1$, $\gamma_m \in [-1, 0]$. Ensures $\| \Gamma \| \leq 1$, and $\text{diag}(\Gamma)=0$ (so $\Gamma \circ F = \Gamma$)

- Complicated calculation involving Krein parameters (similar to classical coupon!): if $T \ll k \log(m + 1)$ then $\| \Gamma \circ M \| \geq 1/2$
Relevance for proper vs improper PAC learning

- PAC learner $\mathcal{A}$ for a concept class $\mathcal{C} = \{f : [n] \rightarrow \{0, 1\}\}$: given samples $(x, f(x))$, $x \sim D$, for unknown target concept $f \in \mathcal{C}$, find hypothesis $h : [n] \rightarrow \{0, 1\}$ that is close to $f$:

$$\forall f \in \mathcal{C} \quad \forall D : \Pr_{x \sim D} [f(x) \neq h(x)] \leq \varepsilon \text{ w.h.p.}$$

- Fundamental Thm: Required # of samples is $\Theta(\text{VCdim}(\mathcal{C})/\varepsilon)$

- $\mathcal{A}$ is called a proper learner if $h \in \mathcal{C}$

- Requiring $\mathcal{A}$ to be proper can increase sample complexity: $\exists \mathcal{C}$ where proper learner needs $\Theta(\text{VCdim}(\mathcal{C}) \log(1/\varepsilon)/\varepsilon)$ examples.

  Related to coupon collector with $m = 1$, $\varepsilon = 1/n$:

  $\mathcal{C} = \{f_S : [n] \rightarrow \{0, 1\} \text{ is indicator of } S\}$

- $O(\text{VCdim}(\mathcal{C})/\varepsilon)$ quantum examples suffice for proper learner
What if you can also reflect through $|S\rangle$?

- If you can get copies of $|S\rangle$, then maybe you actually have a quantum machine to produce such copies? $U : |0\rangle \mapsto |S\rangle$

- Doing $U$ and $U^{-1}$ would allow you to reflect through $|S\rangle$!
  $R_S : |S\rangle \mapsto |S\rangle$, $R_S : |\psi\rangle \mapsto -|\psi\rangle$ whenever $\langle \psi | S \rangle = 0$

- **Finding $S$ more quickly**, using amplitude amplification:
  1. Start from $|[n]\rangle$, rotate to $|S\rangle$. Measure, get $i_1 \in S$
  2. Start from $|[n]\rangle$, rotate to $|S\setminus\{i_1\}\rangle$. Measure, get $i_2 \in S$

Cost of finding all elements of $S$: $\sum_{j=0}^{m-1} \sqrt{\frac{n-j}{m-j}} = O(\sqrt{km})$

- We show that this is tight for $m \leq n/2$, and also get tight bound $\Theta(k)$ for case $m \geq n/2$
Summary

- **Classical coupon collector:**
  learn $k$-set $S \subseteq [n]$ from $\Theta(k \log k)$ uniform samples

- **Quantum coupon collector:**
  learn $k$-set $S \subseteq [n]$ from $\Theta(k \log(m + 1))$ uniform superpositions ($m = n - k$ is number of missing items)

- We also gave tight bounds for learning $S$ from copies of $|S\rangle$ and reflections through $|S\rangle$

- Open problem: are the quantum sample complexities of proper and improper learning the same for all $\mathcal{C}$?