Post-Quantum Succinct Arguments

Two major families of interactive proof systems

① Zero Knowledge (goal: privacy)

$P(x,w) \rightleftarrows V(x)$ — V verifies $x \in L$, but doesn't learn anything else

Ex: Blum's protocol

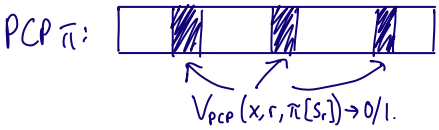
② Succinct Arguments (goal: efficient verification)

$P(x,w) \rightleftarrows V(x)$ — V verifies $x \in L$ and total communication $\ll w$.

Goal: Prove that classical succinct args are PQ secure.

Kilian's Protocol [K92]

PCP Theorem: Any NP statement x has a proof π that can be verified (w/ negl error) by checking $\text{polylog}(\lambda)$ locations.



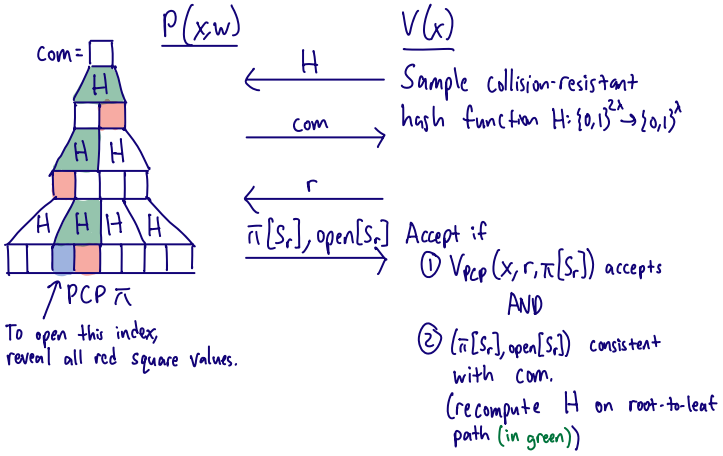
Notation: V_{PCP} samples random coins r, which specify a set S_r of PCP indices.

PCPs enable fast verification of a long proof $\bar{\pi}$.

Can we turn this into a succinct verification protocol?

Kilian's idea: P sends succinct "tree commitment" to $\bar{\pi}$.

Later send $\bar{\pi}[S_r]$ and local opening $\text{open}[S_r]$.



To open this index, reveal all red square values.

Classical Soundness: If x is false, then for any malicious PPT $\tilde{P}$,

$\Pr[V \text{ accepts } \tilde{P}] = \text{negl}$

Proof Intuition

Run $\tilde{P}$ on many random r, "stitch together" responses $\bar{\pi}[S_r]$ into a PCP $\bar{\pi}$

Eventually, get impossibly good $\bar{\pi}$ (violating PCP soundness)

OR inconsistent responses (breaking CRHF)

Formally, assume $\Pr[V \text{ accepts } \tilde{P}] = p$ (p is non-negl).

We'll extract an $\Omega(p)$-convincing PCP.

Step 0: Run $\tilde{P}$ on random H to get com.

$\tilde{P} \xleftarrow[\text{com}]{H}$ Fix (H, com).

(H, com) is p-good if $\Pr[V \text{ accepts } \tilde{P} \mid H, \text{com}] \geq p$.

Claim: $\Pr_{H,\text{com}}[(H,\text{com}) \text{ is } \frac{p}{2}\text{-good}] \geq \frac{p}{2}$

pf: If not, then $\Pr[V \text{ rejects } \tilde{P}] \geq (1-\frac{p}{2})(1-\frac{p}{2}) = 1 - p + \frac{p^2}{4}$

$\Rightarrow \Pr[V \text{ accepts } \tilde{P}] \leq p - \frac{p^2}{4}$, contradiction.

(We did this in Blum too!)

Main Step: Rewind $\tilde{P}$ many times for fixed (H, com).

1) Initialize "blank" PCP $\bar{\pi}$.

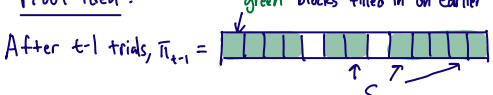
2) Repeat for $t = \text{poly}(\lambda)$ steps:

$\tilde{P} \xleftarrow{r} \xrightarrow{z}$ rewind (reset $\tilde{P}$ back to its state right after sending com)

(parse z as $\bar{\pi}[S_r], \text{open}[S_r]$)

If z is accepting, fill in $\bar{\pi}[S_r]$ at indices S_r (if collision, abort)

Proof idea:



Consider trial t: $\tilde{P} \xleftarrow{r} \xrightarrow{z}$ ($\geq p$ since all trials are independent) (renaming $\frac{p}{2}$ to p)

We can upper bound $\Pr[\tilde{P} \text{ wins on trial } t]$

$$\leq \Pr_r\left[V_{PCP} \text{ accepts } r, \bar{\pi}_{t-1}[S_r]\right] + \Pr_{z \leftarrow \tilde{P}(r)}\left[V(r,z) \text{ accepts and } S_r \text{ contains new index}\right] + \Pr_{z \leftarrow \tilde{P}(r)}\left[z \text{ inconsistent with } \bar{\pi}_{t-1}\right]$$

negl (collision resistance)

Suffices to show: On trial $t = \frac{2|\bar{\pi}|}{p}$, this is at most $\frac{p}{2}$.

(If we prove this, we're done since it implies $\Pr[V_{PCP} \text{ accepts } \bar{\pi}_{t-1}] \geq \frac{p}{2} - \text{negl}$)

Define event W_i = On trial i, $V(r,z)$ accepts and S_r contains a new index $\notin \pi_{i-1}$

① Only $|\overline{\pi}|$-many events W_i can occur.
So $\sum_{i=1}^{t} \Pr[W_i] = E\left[\sum_{i=1}^{t} W_i\right] \leq |\overline{\pi}|$.

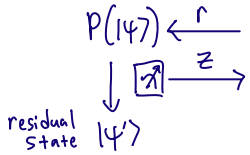
② $\Pr[W_i] \leq \Pr[W_{i-1}]$ (since trials are independent, but π can only have more indices filled at step i vs. $i-1$)

By ①+②: $|\overline{\pi}| \geq \sum_{i=1}^{t} \Pr[W_i] \geq t \Pr[W_t]$.

If $t = \frac{2|\overline{\pi}|}{p}$, then $\Pr[W_t] \leq \frac{p}{2}$.

What if $\widetilde{P}$ is quantum?

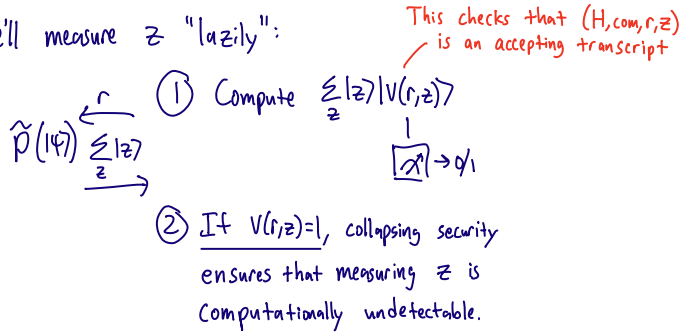
We still want to record many accepting z, but measuring z might disturb $\widetilde{P}$'s internal state $|\psi\rangle$.



Difficulty: Unclear whether $P(|\psi'\rangle)$ can answer further queries.

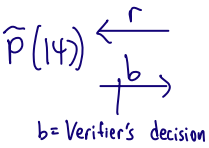
However, we can limit the disturbance to $|\psi\rangle$ if we use a collapsing hash function H in the protocol.

We'll measure z "lazily":



② If $V(r,z)=1$, collapsing security ensures that measuring z is computationally undetectable.

This reduces our task to measuring the verifier's 1-bit decision b.



Goal: Given only one copy of $|\psi\rangle$, make $\widetilde{P}$ succeed ($b=1$) on many random r.

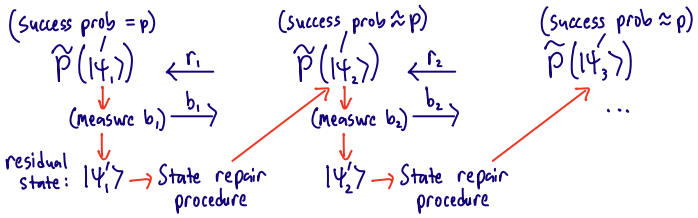
(In the actual reduction, we measure z whenever $b=1$.)

Aside: For Blum's protocol, we only needed to make $\widetilde{P}$ succeed twice.

High-level idea [CMSZZ21]

After each 1-bit measurement, perform a "Repair" procedure to restore success probability.

Initially: $\Pr_r[\widetilde{P}(|\psi\rangle, r) \text{ makes } V \text{ accept}] = p$
(success probability of $|\psi\rangle$)



Note: repair only guarantees $|\psi_1\rangle$ and $|\psi_2\rangle$ have (nearly) the same success prob. The states might be very different!

How is this possible?
To answer this, we'll start by giving a linear-algebraic characterization of success probability.

Understanding success probability

Success prob of $|\psi\rangle = \Pr_r[\widetilde{P}(|\psi\rangle, r) \text{ makes } V \text{ accept}]$

$= \frac{1}{R} \sum_r \left\| \overline{\Pi}_r |\psi\rangle \right\|^2$ (R = # of challenges r)

where $\{\overline{\Pi}_r, I - \overline{\Pi}_r\}$ is the projective measurement that tests if V accepts $\widetilde{P}(|\psi\rangle, r)$

(Formally: If $\widetilde{P}$ applies unitary U_r on challenge r, then $\overline{\Pi}_r := \sum_{z:V(r,z)=1} U_r^\dagger (|z\rangle\langle z| \otimes I) U_r$)

We can simplify this expression by considering a "purified" trial where $|r\rangle$ is in uniform superposition.

Define: $\cdot |t_R\rangle := \frac{1}{\sqrt{R}} \sum_r |r\rangle$

$\cdot \overline{\Pi}_A := \sum_r |r\rangle\langle r| \otimes \overline{\Pi}_r$. (A for "accept")

Then for all states of the form $|t_R\rangle|\psi\rangle$,

$\left\| \overline{\Pi}_A |t_R\rangle|\psi\rangle \right\|^2$ = Success prob of $|\psi\rangle$.

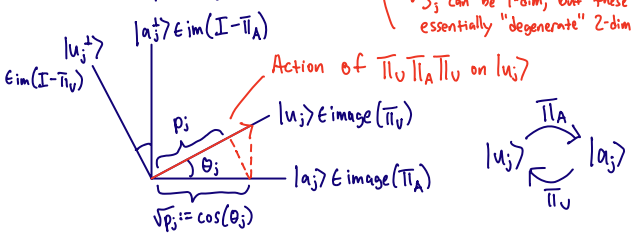
Let $\overline{\Pi}_U$ be the projector onto all such states:

$\overline{\Pi}_U := |t_R\rangle\langle t_R| \otimes I$ (U for "uniform")

Jordan's lemma guarantees existence of orthogonal 2-dim subspaces S_j where Π_A, Π_U are rank-1 in each S_j:

Notes:
- The $\{S_j\}$ span the entire space that Π_A, Π_U act on
- S_j can be 1-dim, but these are essentially "degenerate" 2-dim spaces

Subspace S_j:



Key Point: $|u_j\rangle = |t_R\rangle|\psi_j\rangle$ is a p_j-evector of $\Pi_U\Pi_A\Pi_U$ where $|\psi_j\rangle$ has success prob $\|\Pi_A|t_R\rangle|\psi_j\rangle\|^2 = p_j$.

Define $\Pi_p := \sum_{j:p_j=p} |\psi_j\rangle\langle\psi_j|$. $\text{im}(\Pi_p)$ is "p-success prob subspace".

- $|\psi\rangle \in \text{im}(\Pi_p) \Rightarrow |\psi\rangle$ has success prob p.
- In general, if $|\psi\rangle$ has success prob p', then $|\psi\rangle = \sum_p \alpha_p|\psi_p\rangle$ where $|\psi_p\rangle \in \text{im}(\Pi_p)$ and $p' = \sum |\alpha_p|^2 p$.

How does this relate to rewinding/state repair?

Suppose we can measure $\{\Pi_p\}$, i.e., on $|\psi\rangle = \sum_p \alpha_p|\psi_p\rangle$, this returns p with prob $|\alpha_p|^2$, and the state collapses to $|\psi_p\rangle \in \text{im}(\Pi_p)$

Claim: If $|\psi\rangle$ has initial success prob p', measuring $\{\Pi_p\}$ returns $p \geq \frac{p'}{2}$ with probability $\geq \frac{p'}{2}$.

Pf: $\sum_p |\alpha_p|^2 p = p'$ implies $\sum_{p \geq \frac{p'}{2}} |\alpha_p|^2 \geq \frac{p'}{2}$ by Markov.

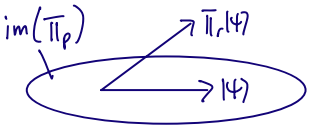
At this point, $|\psi\rangle \in \text{im}(\Pi_p)$ for some non-negl p. (and this occurs w/ non-negl prob)

Next, we run a trial:

$$P(|\psi\rangle) \xleftarrow{r} \xrightarrow{b}$$

Suppose $b=1$, so residual state is $\propto \Pi_r|\psi\rangle$.

Key point: to repair success prob, it suffices to "return" to $\text{im}(\Pi_p)$!



How can we return to $\text{im}(\Pi_p)$?

Turns out this is easy if we can measure $\{\Pi_p, I-\Pi_p\}$.

Note:
- $\{\Pi_p\}$ is a many-outcome measurement that returns some $p \in [0,1]$.
- For any fixed p, $\{\Pi_p, I-\Pi_p\}$ is a binary-outcome measurement that tests whether success prob is p.

Obvious first step: Just measure $\{\Pi_p, I-\Pi_p\}$.

If we get lucky and the outcome is Π_p, we're done! But what if we get $I-\Pi_p$?

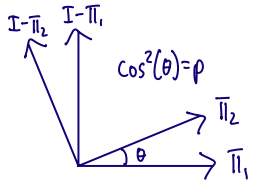
Idea: Use $\{\Pi_r, I-\Pi_r\}$ to get "unstuck" and then try again.

So we alternate Π_p, Π_r measurements until the $\{\Pi_p, I-\Pi_p\}$ measurement accepts (Π_p outcome).

In fact, this works for any pair of projectors:

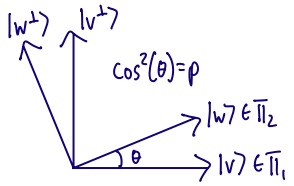
Return-to-subspace lemma: If $|\psi\rangle \in \text{im}(\Pi_1)$ and we measure with $\{\Pi_2, I-\Pi_2\}$, we can return to $\text{im}(\Pi_1)$ by alternating Π_1, Π_2 measurements in 2 expected steps.

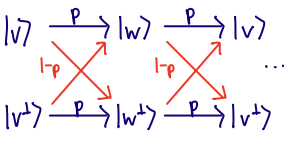
Intuition:



If initial outcome is Π_2, get back to Π_1 quickly.
If $I-\Pi_2$, get back to Π_1 slowly.
However, Π_2 is far more likely, so the effects cancel in expectation.

Pf: Consider Jordan subspaces of Π_1, Π_2





Sub-Claim: In each Jordan subspace, $E[T_{v\to v}] = 3$ where random var $T_{v\to v'}$:= # of measurements to go from $|v\rangle$ to $|v'\rangle$.

Let $q = 2p(1-p)$. Then $E[T_{v\to v}] = (1-q)\cdot 2 + q(2 + E[T_{v^\perp \to v}])$
$= 2 + qE[T_{v^\perp \to v}]$

Also: $E[T_{v^\perp\to v}] = 2q + (1-q)(2 + E[T_{v^\perp \to v}]) \Rightarrow E[T_{v^\perp \to v}] = \frac{1}{q}$.

Putting it all together: $E[T_{v\to v}] = 2 + q\cdot\frac{1}{q} = 3$.

In general, no guarantee that $|\psi\rangle$ lives in a Jordan subspace, i.e., $|\psi\rangle = \sum \alpha_j|v_j\rangle$.

Useful trick: Let Π_{S_j} be projection onto S_j.

Suppose we measure $\{\Pi_{S_j}\}$ (collapse to $|v_j\rangle$ w/ prob $|\alpha_j|^2$) before any Π_1, Π_2 measurements occur.

Key Point: $\{\Pi_{S_j}\}$ measurement commutes with Π_1, Π_2, so it doesn't affect Π_1, Π_2 measurement outcomes. Then we can apply 2-dim analysis to $|v_j\rangle$.

In $\Pi_1\Pi_2\Pi_1$ eigenbasis,

$\Pi_1 = \begin{pmatrix} 1 & & & \\ & 0 & & \\ & & 1 & \\ & & & 0 \\ & & & & \ddots \end{pmatrix}$ $\Pi_{S_1} = \begin{pmatrix} 1 & \\ & 1 \end{pmatrix}, \Pi_{S_2} = \begin{pmatrix} 0 & \\ & 0 \\ & & 1 \\ & & & 1 \end{pmatrix}, \ldots$

At this point, we can repair the state assuming $\{\Pi_p\}$ corresponds to an efficient measurement.

Big problem: No efficient way to measure $\{\Pi_p\}$.

However, it is possible to estimate p to any $\varepsilon = \frac{1}{\text{poly}}$ additive error. It turns out this is enough to repair the state!

High-level idea

Def: $|\psi\rangle$ is strongly p-successful if $|\psi\rangle$ is almost entirely supported on $\text{im}(\Pi_{p'})$ for $p' \in [p \pm \varepsilon]$

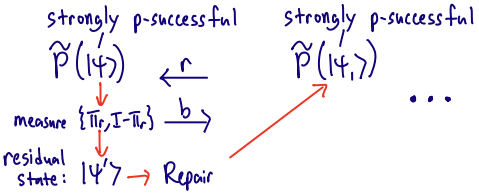
(Strongly p-successful implies success prob $\approx p$)

Using an estimator for p, we'll build a "repair" procedure that guarantees the state is strongly p-successful *before* each trial.

(This ensures that each trial succeeds w/ prob $\approx p$, which suffices for soundness of Kilian)

CMSZ Rewinding Template (after fixing H, com):

1) Estimate p.
2) Main Loop:



Goal: Implement a procedure Repair$(|\psi'\rangle)$ that outputs a strongly p-successful state.

Natural idea: Alternate Π_r w/ a measurement that tests whether the state is strongly p-successful.

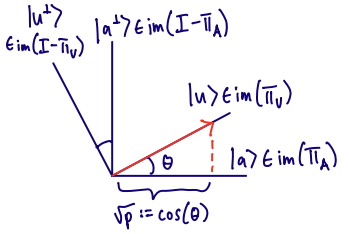
Rest of today:

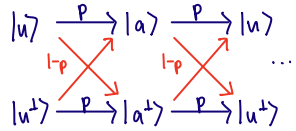
① Show how to estimate p.
(we'll use this to check if state is strongly p-successful)
② Use p-estimator to repair/recover a strongly p-successful state

Part 1: Estimating p

Given $|\psi\rangle$, want to output an "estimate" of p.

This is easy *if* $|u\rangle = |t_R\rangle|\psi\rangle$ is a p-evector of $\Pi_U\Pi_A\Pi_U$:





Est$(|\psi\rangle)$:

1) Initialize $|t_R\rangle|\psi\rangle$
2) Alternate Π_A, Π_U measurements, get outputs $b_1, b_2, b_3, \ldots, b_T$.
3) Output $p' = \frac{[\#\text{ of times } b_i = b_{i+1}]}{T-1}$.

How good is the estimate?

- If we run for $T = \text{poly}(\frac{1}{\varepsilon}, \log(\frac{1}{\delta}))$ steps,
$$\Pr[|p'-p| > \varepsilon] \leq \delta.$$
- So we can achieve $\varepsilon = \frac{1}{\text{poly}}$ additive error with failure probability $\delta = \frac{1}{2^{\text{poly}}}$ in $T = \text{poly}$ steps.

In general, $|t_R\rangle|\psi\rangle$ is *not* always an evector of $\Pi_U\Pi_A\Pi_U$, but a superposition of them:
$$|t_R\rangle|\psi\rangle = \sum_j \alpha_j |u_j\rangle = |t_R\rangle \sum_j \alpha_j |\psi_j\rangle$$
Each $|u_j\rangle$ has a different evalue p_j (where success prob of $|\psi\rangle$ is $\sum_j |\alpha_j|^2 p_j$)
What happens if we run Est $\to p'$?

Intuition: Consider $T = \infty$

- Est$(|\psi_j\rangle)$ always outputs the evalue p_j.
That is, Est *measures* in the $\{|\psi_j\rangle\}_j$ basis and outputs p_j.
- So Est$(\sum_j \alpha_j |\psi_j\rangle)$ outputs p_j with probability $|\alpha_j|^2$.

For large T, p' is an *approximate* measurement of the evalue p_j.

So Est$(\sum_j \alpha_j |\psi_j\rangle) \to p'$ where $p' \approx p_j$ w/ prob $|\alpha_j|^2$.

To "prove" this, we can use the same trick of introducing a *hypothetical* $\{\Pi_{S_j}\}$ measurement before running Est
- This measurement commutes w/ Π_A, Π_U measurements
- State collapses to a single 2-dim space

Two key properties of Est:

(i) $\mathbb{E}[p' \leftarrow \text{Est}(|\psi\rangle)]$ = success prob of $|\psi\rangle$.
(ii) State concentrates on $\approx p'$-evectors after running Est (i.e., strongly p'-successful)

Property (ii) implies that if we run Est *twice*, the two outcomes are ε-close w/ $1-\delta$ probability.

$|\psi\rangle$
↓
Est$(|\psi\rangle) \to p_1$
↓
Est$(|\psi'\rangle) \to p_2$

$$\Pr[|p_1 - p_2| \leq \varepsilon] \geq 1-\delta$$

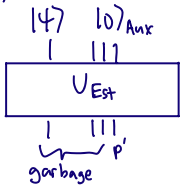
We'll say that Est is (ε, δ)-almost-projective.

Part 2: Using Est to repair

Challenge: Return-to-subspace lemma requires projections but Est is not projective.

Idea: Use Est to implement a projective measurement $\{\Pi_{\rho,\varepsilon}, I-\Pi_{\rho,\varepsilon}\}$ that "approximates" $\{\Pi_\rho, I-\Pi_\rho\}$, and use it to repair the state.

Formally, Est can be implemented as a unitary U_{Est}



Define $\Pi_{\rho,\varepsilon} = U_{Est}^{\dagger}\Big(I \otimes \underbrace{\sum_{\rho' \geq \rho-\varepsilon} |\rho'\rangle\langle\rho'|}_{\text{Check estimate} \geq \rho-\varepsilon}\Big) U_{Est}$ to replace the inefficient Π_ρ measurement.

Repair procedure:

Intuition: this is necessary because U_{Est} only implements Est if Aux = $|0\rangle$

Alternate $\Pi_{\rho,\varepsilon}$ and $\Pi_r \otimes |0\rangle\langle 0|_{Aux}$ until $\Pi_{\rho,\varepsilon}$ accepts.

Why does this work?

We'll show: For all subspaces S_j where $q_j \geq \frac{1}{\text{poly}}$, the state $\rho_j = Tr_{Aux}(|w_j\rangle\langle w_j|)$ is strongly $(\rho-\varepsilon)$-successful.

Suppose we run Est twice on $|\phi_j\rangle$:

1) Run Est$(|\phi_j\rangle) \to p_1$.
2) If $p_1 \geq \rho-\varepsilon$:
   Run Est on the residual state to get p_2.

Since Est is (ε,δ)-almost-projective:

$$\Pr[p_1 \geq \rho-\varepsilon] \cdot \Pr[p_2 < \rho-2\varepsilon \mid p_1 \geq \rho-\varepsilon] \leq \delta.$$

Rearrange: $\Pr[p_2 < \rho-2\varepsilon \mid p_1 \geq \rho-\varepsilon] \leq \dfrac{\delta}{\Pr[p_1 \geq \rho-\varepsilon]}$

These probabilities correspond to the following:

- $\Pr[p_1 \geq \rho-\varepsilon] = \left\| \Pi_{\rho,\varepsilon} |\phi_j\rangle|0\rangle_{Aux} \right\|^2$
- $\Pr[p_2 < \rho-2\varepsilon \mid p_1 \geq \rho-\varepsilon] = \Pr_{p_2 \leftarrow Est(\rho)}[p_2 < \rho-2\varepsilon]$

Together, this implies $\Pr_{p_2 \leftarrow Est(\rho)}[p_2 < \rho-2\varepsilon] \leq \dfrac{\delta}{q_j}$,

so ρ is strongly $(\rho-\varepsilon)$-successful (albeit w/ failure prob $\frac{\delta}{q_j}$)

Claim 1: $\Pi_{\rho,\varepsilon}$ accepts within ≈ 2 steps (in expectation)

Intuition: $|\psi\rangle|0\rangle_{Aux}$ is the state before $\Pi_r \otimes |0\rangle\langle 0|_{Aux}$ measurement.

If $|\psi\rangle|0\rangle_{Aux} \in \text{im}(\Pi_{\rho,\varepsilon})$, then the claim follows from return-to-subspace lemma.

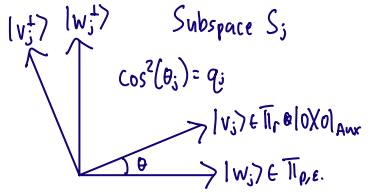
$|\psi\rangle|0\rangle_{Aux} \in \text{im}(\Pi_{\rho,\varepsilon})$ is **almost** true!

- Since $|\psi\rangle$ is strongly ρ-successful, if we run Est, the outcome is $\geq \rho-\varepsilon$ with prob ≈ 1 (technically, $\rho-2\varepsilon$)
- This implies $\left\| \Pi_{\rho,\varepsilon} |\psi\rangle|0\rangle_{Aux} \right\|^2 \approx 1$, so it's very close to a state that satisfies the conditions for the return-to-subspace lemma.

Claim 2: After repair, the state ρ (after discarding Aux) is strongly ρ-successful. (technically it's $\rho-\varepsilon$)

Pf Sketch:

Consider Jordan subspaces of $\Pi_{\rho,\varepsilon}$, $\Pi_r \otimes |0\rangle\langle 0|_{Aux}$.



$$|w_j\rangle = \frac{\Pi_{\rho,\varepsilon}|\phi_j\rangle|0\rangle_{Aux}}{\left\|\Pi_{\rho,\varepsilon}|\phi_j\rangle|0\rangle\right\|}$$

Within each 2-dim subspace, Repair maps $|v_j\rangle \to |w_j\rangle$ (and then discards Aux)

Remains to show: the repaired state is almost entirely supported on subspaces S_j where $q_j \geq \frac{1}{\text{poly}}$.

Rough sketch:

1) The alternating Π_A, Π_V measurements commute w/ $\{\Pi_{S_j}\}$, so it suffices to show that the state **before** the repair satisfies the same condition.
2) The state before repair is $(\Pi_r \otimes |0\rangle\langle 0|_{Aux})|\phi\rangle$ where $|\phi\rangle = |\psi\rangle|0\rangle_{Aux}$, and $|\phi\rangle$ is **almost** in $\Pi_{\rho,\varepsilon}$, i.e., $\left\|\Pi_{\rho,\varepsilon}|\phi\rangle\right\|^2 \geq 1-\delta$.
3) Let's pretend $|\phi\rangle \in \text{im}(\Pi_{\rho,\varepsilon})$ exactly. Then
   $|\phi\rangle = \sum_j \alpha_j |w_j\rangle$ for some $\{\alpha_j\}$.
   In terms of $\{\alpha_j\}$, the state **before** repair is (i.e., after $\Pi_r \otimes |0\rangle\langle 0|_{Aux}$ is applied)
   $$(\Pi_r \otimes |0\rangle\langle 0|)|\phi\rangle = \sum_j \alpha_j \sqrt{q_j}\, |v_j\rangle$$

Key point: The amplitude on S_j is re-weighted by a $\sqrt{q_j}$ factor, so we only have negl total weight on S_j where q_j = negl.

(Formalizing this requires a delicate argument, but the above "re-weighting" observation is the main step)