APPPHYS225 - Tuesday 14 October 2008

Measurements on ensembles

Recall the notion of an ensemble of quantum states $\{p_k, \rho_k\}$. We imagine that someone has given us a quantum system that is guaranteed to have been prepared in one of the states ρ_k. We don't know which one for sure, but we know that the relative probabilities are p_k. Last time we considered a two-dimensional quantum system and the two-membered ensemble

$$p_0 = \frac{1}{2}, \quad \rho_0 = |\Psi_+(\theta)\rangle\langle\Psi_+(\theta)|,$$

$$p_1 = \frac{1}{2}, \quad \rho_1 = |\Psi_-(\theta)\rangle\langle\Psi_-(\theta)|,$$

where

$$|\Psi_\pm(\theta)\rangle = \cos\theta|0\rangle \pm \sin\theta|1\rangle.$$

By performing a measurement we can learn something about the identity of the state that was prepared. That is, we can gain some information about whether the initial preparation corresponded to ρ_0 or ρ_1.

How can we quantify this gain of information? One good way is by computing the expected reduction of Shannon entropy

$$H \equiv -\sum_k p_k \ln p_k.$$

For those not familiar with Shannon entropy from other contexts, it may suffice to note that the maximum value of H for a pair of probabilities $p,\ 1-p$ is obtained when $p = \frac{1}{2}$,

$$H = -\frac{1}{2}\ln\frac{1}{2} - \frac{1}{2}\ln\frac{1}{2} = -\ln\frac{1}{2} = \ln 2.$$

When $p \to 0$ or $p \to 1,$ we have $H \to \ln 1 = 0.$ Hence H is a good measure of how 'flat' a probability distribution is, and thus of just how ignorant we really are. Incidentally, in the classical probability setting we can define the Shannon entropy of a probability distribution function $m(\cdot)$ as

$$H = \langle L_m \rangle,$$

where $L_m(\cdot)$ is a random variable derived from the probability distribution function via

$$L_m(\omega_i) = -\ln m(\omega_i).$$

For an arbitrary random variable $X(\cdot) = \sum_j x_j \chi_{x_j}(\cdot)$ we can define

$$H(X) = -\sum_j \langle \chi_{x_j}\rangle \ln\langle \chi_{x_j}\rangle.$$

Getting back to our example, we have $H_{pre} = \ln 2$ before making any measurement since $p_0 = p_1 = 1/2.$ If we perform the projective measurement $\{\mathbf{E}_0 = |+\rangle\langle+|, \mathbf{E}_1 = |-\rangle\langle-|\},$ where

$$|\pm\rangle = \frac{1}{\sqrt{2}}(|0\rangle \pm |1\rangle),$$

we know that the outcome probabilities are:

$$k=0,\quad |\Psi_+(\theta)\rangle:\quad \Pr(i=0)=\frac{1}{2}(\cos\theta+\sin\theta)^2,\ \Pr(i=1)=\frac{1}{2}(\cos\theta-\sin\theta)^2,$$

$$k=1,\quad |\Psi_-(\theta)\rangle:\quad \Pr(i=0)=\frac{1}{2}(\cos\theta-\sin\theta)^2,\ \Pr(i=1)=\frac{1}{2}(\cos\theta+\sin\theta)^2.$$

To compute the post-measurement probability distrubtion, we can use Bayes' Rule

$$P(k|i)=\frac{P(i|k)P(k)}{P(i)}$$

on the classical probability distribution $\{p_0,p_1\}$. We first compute the conditioned probability distrubtion if the outcome $i=0$ is obained, then the distribution if $i=1$ is obtained:

$$P(k=0|i=0)=\frac{\frac{1}{2}(\cos\theta+\sin\theta)^2p_0}{\frac{1}{2}(\cos\theta+\sin\theta)^2p_0+\frac{1}{2}(\cos\theta-\sin\theta)^2p_1}$$

$$=\frac{(\cos\theta+\sin\theta)^2}{2}=\frac{1}{2}(\cos\theta+\sin\theta)^2,$$

$$P(k=1|i=0)=\frac{\frac{1}{2}(\cos\theta-\sin\theta)^2p_1}{\frac{1}{2}(\cos\theta-\sin\theta)^2p_0+\frac{1}{2}(\cos\theta+\sin\theta)^2p_1}$$

$$=\frac{1}{2}(\cos\theta-\sin\theta)^2,$$

$$P(k=0|i=1)=\frac{\frac{1}{2}(\cos\theta-\sin\theta)^2p_0}{\frac{1}{2}(\cos\theta-\sin\theta)^2p_0+\frac{1}{2}(\cos\theta+\sin\theta)^2p_1}$$

$$=\frac{1}{2}(\cos\theta-\sin\theta)^2,$$

$$P(k=1|i=1)=\frac{\frac{1}{2}(\cos\theta+\sin\theta)^2p_0}{\frac{1}{2}(\cos\theta+\sin\theta)^2p_0+\frac{1}{2}(\cos\theta-\sin\theta)^2p_1}$$

$$=\frac{1}{2}(\cos\theta+\sin\theta)^2.$$

In either case, the new Shannon entropy will be

$$H_{post}\to-\frac{1}{2}(\cos\theta+\sin\theta)^2\ln\left[\frac{1}{2}(\cos\theta+\sin\theta)^2\right]$$

$$-\frac{1}{2}(\cos\theta-\sin\theta)^2\ln\left[\frac{1}{2}(\cos\theta-\sin\theta)^2\right]$$

$$=-(\cos\theta+\sin\theta)^2\ln[\cos\theta+\sin\theta]+\frac{1}{2}(\cos\theta+\sin\theta)^2\ln 2$$

$$-(\cos\theta-\sin\theta)^2\ln[\cos\theta-\sin\theta]+\frac{1}{2}(\cos\theta-\sin\theta)^2\ln 2$$

$$=-(\cos\theta+\sin\theta)^2\ln[\cos\theta+\sin\theta]-(\cos\theta-\sin\theta)^2\ln[\cos\theta-\sin\theta]+\ln 2.$$

Hence the change in Shannon entropy is

$$\Delta H\equiv H_{post}-H_{pre}=-(\cos\theta+\sin\theta)^2\ln[\cos\theta+\sin\theta]-(\cos\theta-\sin\theta)^2\ln[\cos\theta-\sin\theta].$$

As $\theta\to 0,\ \Delta H\to 0$ since no information can be gained by any measurement. As $\theta\to\pi/4,$

$$\Delta H \to -\left(\frac{2}{\sqrt{2}}\right)^2 \ln\left[\frac{2}{\sqrt{2}}\right] = -2\ln[2^{1/2}] = -\ln 2,$$

which makes sense since we can determine the state with perfect certainty!

It is interesting to consider also the case of our 'optimal' tri-valued POVM,

$$\mathbf{E}_0 = \frac{1}{2}\cos^{-2}\theta\,\mathbf{\Pi}_{-\perp},$$
$$\mathbf{E}_1 = \frac{1}{2}\cos^{-2}\theta\,\mathbf{\Pi}_{+\perp},$$
$$\mathbf{E}_2 = (1-\tan^2\theta)\mathbf{\Pi}_0.$$

Last time we computed the table of probabilities $P(i|k)$,

	$k=0:\Psi_+(\theta)$	$k=1:\Psi_-(\theta)$
$i=0:+$	$2\sin^2\theta$	0
$i=1:-$	0	$2\sin^2\theta$
$i=2:?$	$1-2\sin^2\theta$	$1-2\sin^2\theta$

so we may again use Bayes' Rule to find

$$P(k=0|i=0) = \frac{2\sin^2\theta p_0}{2\sin^2\theta p_0 + 0} = 1,$$
$$P(k=1|i=0) = 0,$$
$$P(k=0|i=1) = 0,$$
$$P(k=1|i=1) = 1,$$
$$P(k=0|i=2) = \frac{(1-2\sin^2\theta)p_0}{(1-2\sin^2\theta)p_0 + (1-2\sin^2\theta)p_1} = \frac{1}{2},$$
$$P(k=1|i=2) = \frac{1}{2}.$$

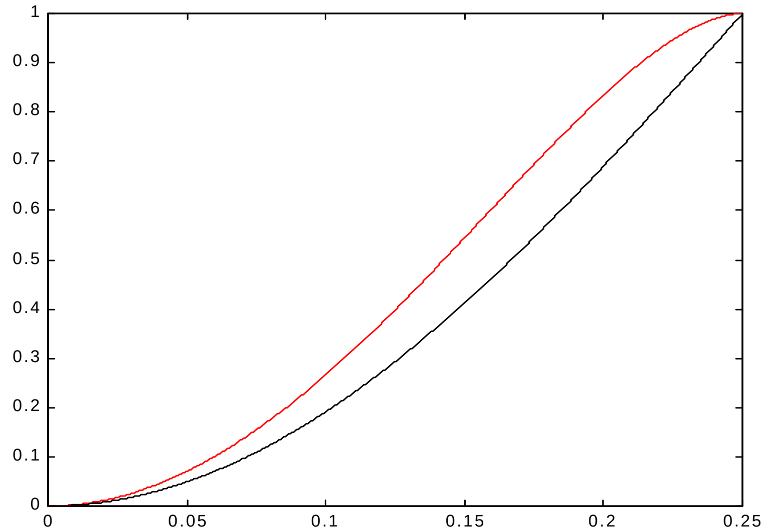
Hence we find that the post-measurement Shannon entropy depends on the result that is obtained. If $i=0$ or $i=1$, then $H_{post}=0$, but if $i=2$ then $H_{post}=\ln 2$. The changes in entropy are then

$$\Delta H_0 = -\ln 2, \quad \Delta H_1 = -\ln 2, \quad \Delta H_2 = 0.$$

Averaging these different cases with their respective probabilities,

$$\langle \Delta H\rangle = \Delta H_0 \Pr(i=0) + \Delta H_1 \Pr(i=1) + \Delta H_2 \Pr(i=2)$$
$$= -\ln 2\,\sin^2\theta - \ln 2\,\sin^2\theta$$
$$= -(2\ln 2)\sin^2\theta.$$

As $\theta \to 0$, $\langle \Delta H\rangle \to 0$ as it must, and as $\theta \to \pi/4$, $\langle \Delta H\rangle$ goes nicely to $-\ln 2$. Comparing now the simple projective measurement to our fancy POVM, in terms of $\langle \Delta H\rangle$, we get the following graph:



As usual, the x-axis corresponds to θ in units of π, and the vertical axis is now $-\langle \Delta H \rangle / H_{pre}$, which we will take as our definition of the information gained in the measurement (other definitions could be equally valid). For $\theta \in (0, \pi/4)$, we see that the projective measurement (in red) is actually better than the POVM (in black)! Hence it appears that although we did gain something by going to an indirect measurement procedure (no false positives), the tri-valued POVM actually gives us less 'information' on average, measured as ΔH.

Selective vs. non-selective evolution; disturbance

Consider an indirect measurement procedure, having N possible outcomes, defined by the operation elements $\{\mathbf{A}_1 \ldots \mathbf{A}_N\}$. For a given pre-measurement system state ρ the outcome probabilities will be

$$\Pr(i) = \mathrm{Tr}\left[\rho \mathbf{A}_i^\dagger \mathbf{A}_i\right],$$

and the post-measurement states are given by

$$\text{outcome } i: \quad \rho \mapsto \rho_i = \frac{\mathbf{A}_i \rho \mathbf{A}_i^\dagger}{\Pr(i)}.$$

Note that the division by $\Pr(i)$ guarentees normalization of the ρ_i.

The evolution rule $\rho \mapsto \rho_i$, which presumes that we have full knowledge of the outcome of the measurement, is termed selective or conditional evolution. What should our description of the post-measurement system state be, if we know that a measurement described by $\{\mathbf{A}_1 \ldots \mathbf{A}_N\}$ has been performed but we have no idea what result was obtained? Following basic probabilistic intuition,

$$\rho \mapsto \sum_{i=1}^{N} \Pr(i) \rho_i = \sum_{i=1}^{N} \mathbf{A}_i \rho \mathbf{A}_i^\dagger.$$

This rule is termed non-selective or unconditional evolution.

To see explicitly how different these two cases can be, let's look at a very simple example. Consider the measurement $\{\Pi_0, \Pi_1\}$ performed on a two-dimensional quantum system, where $\{|0\rangle, |1\rangle\}$ is an orthonormal basis for the system Hilbert space and

$$\Pi_0 = |0\rangle\langle 0|, \quad \Pi_1 = |1\rangle\langle 1|.$$

For the system pre-measurment state

$$|\Psi\rangle = \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle),$$

the outcome probabilities will clearly be

$$\Pr(0) = \Pr(1) = \frac{1}{2}.$$

The post-measurement states for selective evolution are likewise clearly

$$\rho_0 = |0\rangle\langle 0|,$$
$$\rho_1 = |1\rangle\langle 1|.$$

If the measurement $\{\Pi_0, \Pi_1\}$ is performed but for some reason we do not know the outcome, then the unconditional post-measurement state will be

$$\rho = \sum_{i=0,1} \Pr(i)\rho_i$$
$$= \frac{1}{2}|0\rangle\langle 0| + \frac{1}{2}|1\rangle\langle 1|$$
$$= \frac{1}{2}\mathbf{1}.$$

That is, instead of being left with a pure state after the measurement we end up with the maximally mixed state.

Note that in classical scenarios one can generally assume the existence of 'non-invasive' measurements, such that performing the measurement but ignoring the result is equivalent to never having performed the measurement at all. In the formal language of Bayesian statistics, the conditioning of a prior probability distribution $P(x)$ by a measurement result i is given by Bayes' Rule

$$P(x|i) = \frac{P(i|x)P(x)}{P(i)}.$$

This plays an analogous role to the quantum selective evolution rule

$$\rho_i = \frac{\mathbf{A}_i \rho \mathbf{A}_i^\dagger}{\Pr(i)}.$$

In the classical case, the non-selective evolution is

$$\sum_i P(x|i)P(i) = \sum_i P(i|x)P(x) = \sum_i P(i,x) = P(x),$$

where $P(i,x)$ is the joint probability of i and x. That is, we find a post-measurement probability distribution identical to the pre-measurement distribution. In the quantum case we should already have a sense that for essentially any measurement $\{\mathbf{A}_i\}$,

$$\rho \mapsto \sum_i \mathbf{A}_i \rho \mathbf{A}_i^\dagger \neq \rho$$

for general ρ. One exception to this is the class of measurements in which every $\mathbf{A}_i \propto \mathbf{1}$, but these are measurements that yield no information about the system state! A more subtle special case worth mentioning is one where the $\mathbf{A}_i$ are fine-tuned to correspond to projectors onto eigenspaces of the original ρ, but of course a measurement of this type is only 'non-disturbing' for a restricted class of density opertors.

Generally speaking, there are many possible ways to quantify the disturbance induced by a measurement procedure. If we restrict our attention to pure initial states, one good measure is the 'discrepancy rate'

$$D \equiv 1 - \left\langle \Psi_{pre} \middle| \rho_{post} \middle| \Psi_{pre} \right\rangle.$$

Let's apply this to our ensemble examples from the first section. For the projective measurement, the nonselective post-measurement states are

$$\begin{aligned}
\rho_0 &= |\Psi_+(\theta)\rangle\langle\Psi_+(\theta)| \\
&\mapsto \Pi_0|\Psi_+(\theta)\rangle\langle\Psi_+(\theta)|\Pi_0 + \Pi_1|\Psi_+(\theta)\rangle\langle\Psi_+(\theta)|\Pi_1, \\
\rho_1 &= |\Psi_-(\theta)\rangle\langle\Psi_-(\theta)| \\
&\mapsto \Pi_0|\Psi_-(\theta)\rangle\langle\Psi_-(\theta)|\Pi_0 + \Pi_1|\Psi_-(\theta)\rangle\langle\Psi_-(\theta)|\Pi_1.
\end{aligned}$$

The corresponding discrepancy rates are then

$$\begin{aligned}
D_0 &= 1 - \langle\Psi_+(\theta)|[\Pi_0|\Psi_+(\theta)\rangle\langle\Psi_+(\theta)|\Pi_0 + \Pi_1|\Psi_+(\theta)\rangle\langle\Psi_+(\theta)|\Pi_1]|\Psi_+(\theta)\rangle \\
&= 1 - \Pr(i=0|k=0)^2 - \Pr(i=1|k=0)^2 \\
&= 1 - \left[\frac{1}{2}(\cos\theta + \sin\theta)^2\right]^2 - \left[\frac{1}{2}(\cos\theta - \sin\theta)^2\right]^2 \\
&= 1 - \frac{1}{4}(1 + 2\cos\theta\sin\theta)^2 - \frac{1}{4}(1 - 2\cos\theta\sin\theta)^2 \\
&= \frac{1}{2} - 2\cos^2\theta\sin^2\theta, \\
D_1 &= 1 - \langle\Psi_-(\theta)|[\Pi_0|\Psi_-(\theta)\rangle\langle\Psi_-(\theta)|\Pi_0 + \Pi_1|\Psi_-(\theta)\rangle\langle\Psi_-(\theta)|\Pi_1]|\Psi_-(\theta)\rangle \\
&= 1 - \Pr(i=0|k=1)^2 - \Pr(i=1|k=1)^2 \\
&= 1 - \left[\frac{1}{2}(\cos\theta - \sin\theta)^2\right]^2 - \left[\frac{1}{2}(\cos\theta + \sin\theta)^2\right]^2 \\
&= \frac{1}{2} - 2\cos^2\theta\sin^2\theta.
\end{aligned}$$

Since the discrepancy rate is equivalent for either initial state, we find that the rate is

$$\langle D\rangle = \frac{1}{2} - 2\cos^2\theta\sin^2\theta$$

for the projective measurement scheme. As $\theta \to 0$, $\langle D\rangle \to \frac{1}{2}$, and as $\theta \to \pi/4$, $\langle D\rangle \to 0$.

What about our fancy POVM? The non-selective post-measurement states will be

$\rho_0 \mapsto \sqrt{\frac{1}{2}\cos^{-2}\theta}\,\Pi_{-\perp}\vert\Psi_+(\theta)\rangle\langle\Psi_+(\theta)\vert\sqrt{\frac{1}{2}\cos^{-2}\theta}\,\Pi_{-\perp} +$
$\sqrt{1-\tan^2\theta}\,\Pi_0\vert\Psi_+(\theta)\rangle\langle\Psi_+(\theta)\vert\sqrt{1-\tan^2\theta}\,\Pi_0,$
$\rho_1 \mapsto \sqrt{\frac{1}{2}\cos^{-2}\theta}\,\Pi_{+\perp}\vert\Psi_-(\theta)\rangle\langle\Psi_-(\theta)\vert\sqrt{\frac{1}{2}\cos^{-2}\theta}\,\Pi_{+\perp} +$
$\sqrt{1-\tan^2\theta}\,\Pi_0\vert\Psi_-(\theta)\rangle\langle\Psi_-(\theta)\vert\sqrt{1-\tan^2\theta}\,\Pi_0,$

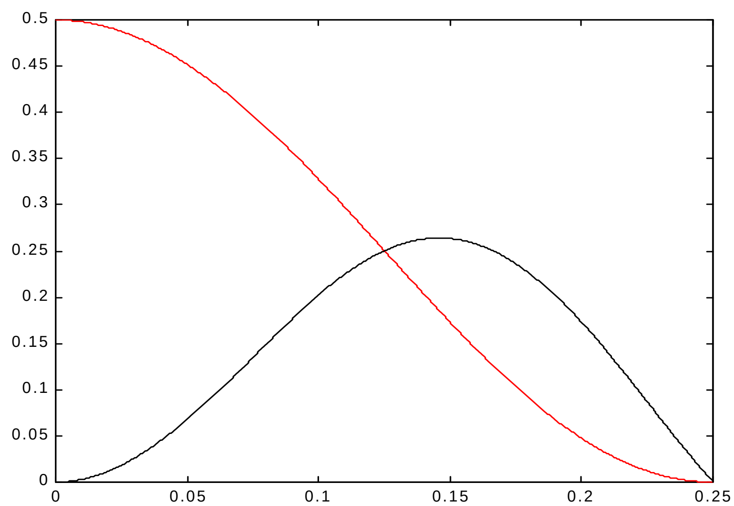
so

$D_0 = 1 - \frac{1}{2}\cos^{-2}\theta\langle\Psi_+(\theta)\vert\Pi_{-\perp}\vert\Psi_+(\theta)\rangle^2 - (1-\tan^2\theta)\langle\Psi_+(\theta)\vert\Pi_0\vert\Psi_+(\theta)\rangle^2$
$= 1 - 8\sin^4\theta\cos^2\theta - (1-\tan^2\theta)\cos^4\theta$
$= 1 - 8\sin^4\theta\cos^2\theta - \cos^4\theta + \sin^2\theta\cos^2\theta,$
$D_1 = 1 - 8\sin^4\theta\cos^2\theta - \cos^4\theta + \sin^2\theta\cos^2\theta,$

hence

$$\langle D\rangle = 1 - 8\sin^4\theta\cos^2\theta - \cos^4\theta + \sin^2\theta\cos^2\theta.$$

We can again check the limits, and this time find that $\langle D\rangle \to 0$ as $\theta \to 0$ and $\langle D\rangle \to 0$ as $\theta \to \pi/4.$ Plotting the two average disturbances, we find:



Again the x-axis is θ in units of $\pi,$ the red curve is $\langle D\rangle$ for the projective measurement, and the black curve is $\langle D\rangle$ for the POVM. It seems that the POVM does much better at minimizing disturbance for small $\theta,$ but much worse in the range of θ greater than $\sim \pi/8.$ Hence we find some quantitative meaning to the notion that generalized measurements allow one to play different inference-disturbance tradeoffs.

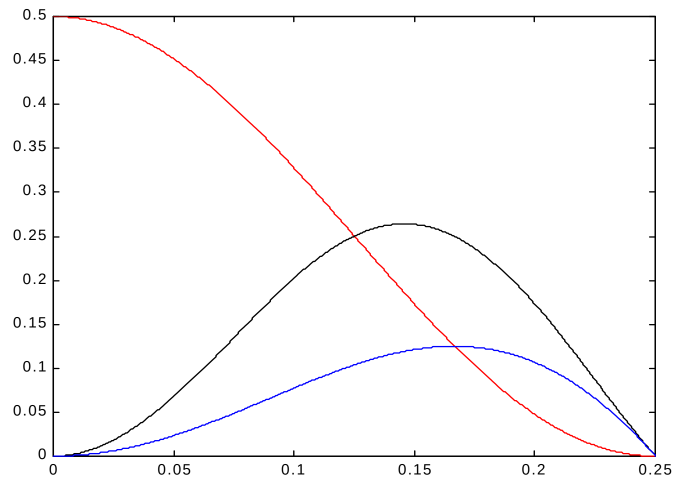
It is worth noting that we can 'patch up' our POVM in a sneaky way to improve its disturbance properties. In the case that we obtain results $i = 0$ (definitely not $\vert\Psi_-(\theta)\rangle$) or $i = 1$ (definitely not $\vert\Psi_+(\theta)\rangle$), we know the initial state with complete certainty. Thus, although we know that our indirect measurement procedure will leave the state as $\vert -_\perp\rangle$ or $\vert +_\perp\rangle$ respectively, we can apply a unitary rotation

$$i = 0 : \quad |-_{\perp}\rangle \mapsto |\Psi_{+}(\theta)\rangle,$$
$$i = 1 : \quad |+_{\perp}\rangle \mapsto |\Psi_{-}(\theta)\rangle,$$

before we declare ourselves 'done' with the measurement! Note carefully that we are talking about two different unitary operators, to be applied depending on whether $i = 0$ or $i = 1.$ Then in the cases where $i \neq 2,$ we see that the disturbances can be made exactly zero. Our residual average disturbance will then be

$$\begin{aligned}\langle D\rangle &= 1 - \Pr(i = k) - \Pr(i = 2)^2 \\ &= 1 - 2\sin^2\theta - (1 - \tan^2\theta)\cos^4\theta \\ &= 1 - 2\sin^2\theta - \cos^4\theta + \sin^2\theta\cos^2\theta.\end{aligned}$$

We note that this is precisely one-half the value for the straight POVM, and the comparison graph now looks like



Here red is the projective scheme, black is the original POVM, and blue is the modified POVM. The x-axis is still $\theta/\pi,$ and the vertical axis is $\langle D\rangle.$

C. A. Fuchs and A. Peres, "Quantum-state disturbance versus information gain: Uncertainty relations for quantum information," Phys. Rev. A **53**, 2038 (1996); C. A. Fuchs and K. Jacobs, "Information-tradeoff relations for finite-strength quantum measurements," Phys. Rev. A **63**, 062305 (2001).

Note that we have above considered two different 'realizations' of our optimal three-element POVM,

$$\mathbf{E}_0 = \frac{1}{2}\cos^{-2}\theta\,\mathbf{\Pi}_{-\perp}, \quad \mathbf{E}_1 = \frac{1}{2}\cos^{-2}\theta\,\mathbf{\Pi}_{+\perp}, \quad \mathbf{E}_2 = (1 - \tan^2\theta)\mathbf{\Pi}_0.$$

In the first we took the minimal assumption for the decompositions

$$\mathbf{E}_i = \sum_j \mathbf{A}_{ij}^{\dagger}\mathbf{A}_{ij},$$

by setting

$$\mathbf{E}_0 = \mathbf{A}_0^\dagger \mathbf{A}_0 \to \left(\frac{1}{\sqrt{2}\cos\theta}\mathbf{\Pi}_{-\perp}\right)^\dagger \left(\frac{1}{\sqrt{2}\cos\theta}\mathbf{\Pi}_{-\perp}\right),$$

$$\mathbf{E}_1 = \mathbf{A}_1^\dagger \mathbf{A}_1 \to \left(\frac{1}{\sqrt{2}\cos\theta}\mathbf{\Pi}_{+\perp}\right)^\dagger \left(\frac{1}{\sqrt{2}\cos\theta}\mathbf{\Pi}_{+\perp}\right),$$

$$\mathbf{E}_2 = \mathbf{A}_2^\dagger \mathbf{A}_2 \to \left(\frac{1}{\sqrt{1-\tan^2\theta}}\mathbf{\Pi}_0\right)^\dagger \left(\frac{1}{\sqrt{1-\tan^2\theta}}\mathbf{\Pi}_0\right).$$

In the second we used the alternative decomposition

$$\mathbf{A}_0^\dagger \mathbf{A}_0 \to \left(\mathbf{U}_+\frac{1}{\sqrt{2}\cos\theta}\mathbf{\Pi}_{-\perp}\right)^\dagger \left(\mathbf{U}_+\frac{1}{\sqrt{2}\cos\theta}\mathbf{\Pi}_{-\perp}\right),$$

$$\mathbf{A}_1^\dagger \mathbf{A}_1 \to \left(\mathbf{U}_-\frac{1}{\sqrt{2}\cos\theta}\mathbf{\Pi}_{+\perp}\right)^\dagger \left(\mathbf{U}_-\frac{1}{\sqrt{2}\cos\theta}\mathbf{\Pi}_{+\perp}\right),$$

where $\mathbf{U}_+$ is a unitary operator that rotates $|\Psi_{-\perp}\rangle$ to $|\Psi_+\rangle$ and $\mathbf{U}_-$ is a unitary operator that rotates $|\Psi_{+\perp}\rangle$ to $|\Psi_-\rangle$. It is easily seen that the unitary modifications to $\mathbf{A}_0$ and $\mathbf{A}_1$ do not effect $\mathbf{E}_0 = \mathbf{A}_0^\dagger\mathbf{A}_0$ or $\mathbf{E}_1 = \mathbf{A}_1^\dagger\mathbf{A}_1$, while they do have a significant effect on

$$\rho_0 = \mathbf{A}_0\rho\mathbf{A}_0^\dagger, \quad \rho_1 = \mathbf{A}_1\rho\mathbf{A}_1^\dagger.$$

This type of "feedback" modification, in which an additional operation is performed on the system in a manner that depends on the measurement result, thus generally changes the disturbance of the overall procedure without changing the information gain.

Measurement with feedback can be a useful tool for engineering evolutions that may be difficult to obtain via the Schrödinger Equation alone. For example, the 'initializing' state map

$$\rho \mapsto |\Psi_0\rangle\langle\Psi_0|,$$

which maps all possible initial states ρ to a single pure state $|\Psi_0\rangle$, obviously cannot be implemented as a unitary evolution

$$\rho \to \mathbf{U}\rho\mathbf{U}^\dagger.$$

On the other hand, consider the measurement-feedback procedure described by the operation elements

$$\mathbf{A}_0 = \mathbf{\Pi}_0, \quad \mathbf{A}_1 = \mathbf{U}\mathbf{\Pi}_1,$$

where

$$\mathbf{\Pi}_0 = |\Psi_0\rangle\langle\Psi_0|, \quad \mathbf{\Pi}_1 = |\Psi_1\rangle\langle\Psi_1|,$$

$$\mathbf{U} = |\Psi_0\rangle\langle\Psi_1| + |\Psi_1\rangle\langle\Psi_0|,$$

and $\{|\Psi_0\rangle, |\Psi_1\rangle\}$ span the system Hilbert space (we are considering a two-dimensional example). Here we can verify

$$
\begin{aligned}
\mathbf{U}^{\dagger}\mathbf{U} &= (|\Psi_0\rangle\langle\Psi_1| + |\Psi_1\rangle\langle\Psi_0|)(|\Psi_0\rangle\langle\Psi_1| + |\Psi_1\rangle\langle\Psi_0|) \\
&= |\Psi_0\rangle\langle\Psi_0| + |\Psi_1\rangle\langle\Psi_1| = \mathbf{1}, \\
\mathbf{A}_0^{\dagger}\mathbf{A}_0 + \mathbf{A}_1^{\dagger}\mathbf{A}_1 &= \mathbf{\Pi}_0 + \mathbf{\Pi}_1 = \mathbf{1}.
\end{aligned}
$$

Note that

$$
\mathbf{A}_1 = \mathbf{U}\mathbf{\Pi}_1 = (|\Psi_0\rangle\langle\Psi_1| + |\Psi_1\rangle\langle\Psi_0|)|\Psi_1\rangle\langle\Psi_1| = |\Psi_0\rangle\langle\Psi_1|.
$$

Thus for an arbitrary initial density matrix $\boldsymbol{\rho}$ we have the (overall) nonselective evolution

$$
\begin{aligned}
\boldsymbol{\rho} &\mapsto \mathbf{A}_0\boldsymbol{\rho}\mathbf{A}_0^{\dagger} + \mathbf{A}_1\boldsymbol{\rho}\mathbf{A}_1^{\dagger} \\
&= |\Psi_0\rangle\langle\Psi_0|\boldsymbol{\rho}|\Psi_0\rangle\langle\Psi_0| + |\Psi_0\rangle\langle\Psi_1|\boldsymbol{\rho}|\Psi_1\rangle\langle\Psi_0| \\
&= |\Psi_0\rangle\langle\Psi_0|(\langle\Psi_0|\boldsymbol{\rho}|\Psi_0\rangle + \langle\Psi_1|\boldsymbol{\rho}|\Psi_1\rangle).
\end{aligned}
$$

Since the quantity in parenthesis is the trace of $\boldsymbol{\rho}$ (equal to one), we see that the measurement-feedback procedure indeed realizes the desired initialization.