Exercise Sheet 1: Basics of Quantum Information Theory

This exercise sheet tries to teach you some of the basics of quantum information theory.

Density Matrix Formalism

Introduction. There are multiple ways to approach a description of quantum mechanics from the analytical point of view. Quantum mechanics courses usually rely mostly on the formalism of quantum states, normally expressed as $|\psi\rangle$, and the Schrödinger equation. For our purposes, we will, however, mostly use the *density matrix formulation* of quantum mechanics, which allows for a simpler treatment of *probabilistic mixtures* of quantum states. These arise for example when a quantum system undergoes unwanted random interactions with an environment, introducing "noise" to a quantum state.

The density matrix formulation starts from the following (incomplete) set of postulates:

(I) Each physical system is associated with a Hilbert space $(\mathcal{H}, \langle\cdot|\cdot\rangle)$. The **(mixed) state** of a quantum system is described by a non-negative (also called positive semi-definite, this means that all eigenvalues of the matrix are non-negative), self-adjoint linear operator with unit trace, i.e. an element of[1]

$$\mathcal{D} := \{\rho \in L(\mathcal{H}) \mid \rho = \rho^\dagger,\ \rho \geq 0,\ \mathrm{Tr}[\rho] = 1\}. \tag{1}$$

Here, $\rho \geq 0$ is the notation we use to say that ρ is positive semi-definite.

(II) **Observables** are represented by Hermitian operators on $\mathcal{H}$. The expectation value of an observable A in the state ρ is given by $\langle A\rangle_\rho = \mathrm{Tr}[A\rho]$.

(III) The **time-evolution** of the state of a quantum system satisfies

$$\frac{\mathrm{d}\rho}{\mathrm{d}t} = -i[H, \rho],$$

where H is the *Hamiltonian*, the observable associated to the total energy of the system.

2 P. **Exercise 1** (Rank one projectors)**.** Show that the set

$$\mathcal{P} = \{\pi \in L(\mathcal{H}) \mid \pi = \pi^\dagger,\ \pi^2 = \pi,\ \mathrm{rank}\,\pi = 1\} \tag{2}$$

of orthogonal projectors onto one-dimensional subspaces of $\mathcal{H}$ is a subset of $\mathcal{D}$.

Solution

We need to show that any rank one orthogonal projectors π is positive semi-definite and has trace 1. Since $\pi^2 = \pi$, the only eigenvalues of π are 0 and 1, as a matter of fact if $|\lambda\rangle$ is an eigenvector then $\lambda|\lambda\rangle = \pi|\lambda\rangle = \pi^2|\lambda\rangle = \lambda^2|\lambda\rangle$ hence $\lambda^2 = \lambda$. This implies π is positive semi-definite. Since π has rank 1, there can only be one eigenvalue 1, hence $\mathrm{Tr}[\pi] = 1$.

Most probably, you have originally learned another definition for quantum states in your first quantum mechanics course. Namely, **pure quantum states** are rays of the Hilbert space $\mathcal{H}$. The rays of a Hilbert space are the equivalence classes of unit vector that only differ by

[1] In quantum information theory, it will be sufficient to consider finite-dimensional Hilbert spaces most of the time. A finite-dimensional Hilbert space is simply a vector space. In infinite dimension there are more subtleties, but these do not concern us.

a phase factor. In symbols, we have $\text{rays}(\mathcal{H}) = \{|\psi\rangle \in \mathcal{H} \mid \||\psi\rangle\|_2^2 = 1\}/\sim$. The equivalence relation $|\psi\rangle \sim |\phi\rangle$ captures the fact that if there exist $\alpha \in \mathbb{R}$ such that $|\psi\rangle = \mathrm{e}^{i\alpha}|\phi\rangle$, then $|\psi\rangle$ and $|\phi\rangle$ represent the same state. Often physicists tend to drop the equivalence relation and talk about unit vectors as quantum states instead of rays.

5 P. **Exercise 2** (Pure states)**.** In this exercise, we will show that the set $\mathcal{P}$ of Eq. (2) is equivalent to the set of pure quantum states and use this to derive the time-evolution in the density matrix formalism from the pure state Schrödinger equation.

2 P. (a) Show that the mapping

$$[|\psi\rangle] \mapsto |\psi\rangle\langle\psi| \tag{3}$$

is a bijection between the set $\mathcal{P}$ defined in Eq. (2) and $\text{rays}(\mathcal{H})$ irrespective of which representative of $[|\psi\rangle]$ is chosen.

Solution

Let $[|\psi\rangle] \in \text{rays}(\mathcal{H})$ and $|\psi\rangle$ be a representative of $[|\psi\rangle]$. We can then map it to $|\psi\rangle\langle\psi| \in \mathcal{P}$. This is well-defined since given another representative $|\psi'\rangle = \mathrm{e}^{i\alpha}|\psi\rangle$ with $\alpha \in \mathbb{R}$ we have $|\psi'\rangle\langle\psi'| = e^{i\alpha}|\psi\rangle\langle\psi|e^{-i\alpha} = |\psi\rangle\langle\psi|$. Let $\pi \in \mathcal{P}$. The mapping is inverted by choosing a normalized vector from the image of π. This choice is unique up to complex phase which does not matter because of the equivalence relation.

1 P. (b) Show that the mapping of Eq. (3) is the correct one as it reproduces the same expectation values for any observable A.

Solution

For any representative $|\psi\rangle$ of the ray $[|\psi\rangle]$, we have that

$$\langle A\rangle = \langle\psi|A|\psi\rangle.$$

Using the same representative under the mapping of Eq. (3), we have

$$\langle A\rangle = \mathrm{Tr}[A|\psi\rangle\langle\psi|] = \langle\psi|A|\psi\rangle,$$

proving the desired result.

2 P. (c) Starting from the Schrödinger equation for pure states, i.e.

$$\frac{\mathrm{d}}{\mathrm{d}t}|\psi\rangle = -iH|\psi\rangle \tag{4}$$

derive the corresponding evolution equation for density matrices

$$\frac{\mathrm{d}\rho}{\mathrm{d}t} = -i[H,\rho]. \tag{5}$$

Here $[A,B] = AB - BA$ is the commutator of two matrices. (Hint: start by proving this for $\rho = \pi$ a pure state, then use linearity.)

Solution

We start from $\rho = \pi$ a pure state. As shown in the previous point, $\pi = |\psi\rangle\langle\psi|$ for some time dependent normalized vector $|\psi\rangle(t)$. We have

$$\frac{\mathrm{d}\pi}{\mathrm{d}t} = \frac{\mathrm{d}}{\mathrm{d}t}(|\psi\rangle\langle\psi|) = \frac{\mathrm{d}}{\mathrm{d}t}(|\psi\rangle)\langle\psi| + |\psi\rangle\frac{\mathrm{d}}{\mathrm{d}t}(\langle\psi|).$$

If this looks strange, you can verify that this is true by writing $|\psi\rangle(t) = \sum_i \psi_i(t)|i\rangle$, where $|i\rangle$ are time independent basis vectors, then you need only apply the product rule to the coefficients ψ_i, which are just complex valued functions. Using the Schroedinger equation we get

$$\frac{\mathrm{d}\pi}{\mathrm{d}t} = -iH|\psi\rangle\langle\psi| + i|\psi\rangle\langle\psi|H = -i[H,\pi].$$

Any density matrix ρ can be decomposed as a sum of projectors by the spectral theorem (see also Exercise 4 below), $\rho = \sum_i \rho_i\pi_i$. Since the equation is linear, it holds for any ρ.

We have seen that density matrices describe both mixed and pure quantum states. Let us define the following function of the state

$$f : \mathcal{D} \to \mathbb{R}, \rho \mapsto \mathrm{Tr}[\rho^2]. \tag{6}$$

Before we come to the next exercise, we will also introduce the so-called *Hilbert-Schmidt inner product*, which endows the space of matrices with an inner product. It is defined as

$$\langle A, B\rangle := \mathrm{Tr}[A^\dagger B]. \tag{7}$$

As any proper inner product, it also obeys a Cauchy-Schwarz inequality:

$$\mathrm{Tr}[A^\dagger B]^2 \leq \mathrm{Tr}[A^\dagger A]\,\mathrm{Tr}[B^\dagger B]. \tag{8}$$

9 P. Exercise 3 (Purity)**.**

3 P. (a) Show that $\frac{1}{d} \leq f(\rho) \leq 1$, where d is the dimension of the Hilbert space $\mathcal{H}$. (Hint: Use the Cauchy-Schwarz inequality)

Solution

Let $\{\lambda_i\}_{i=1}^d$ be the eigenvalues of ρ. We then have that

$$\mathrm{Tr}[\rho^2] = \sum_{i=1}^d \lambda_i^2 \leq \sum_{i=1}^d \lambda_i = 1,$$

where we used that for all $0 \leq \lambda \leq 1$ we have that $\lambda^2 \leq \lambda$. This proves the upper bound. For the lower bound, we observe that we can insert an identity and use Cauchy-Schwarz to obtain

$$\begin{aligned} 1 &= \mathrm{Tr}[\rho]^2 \\ &= \mathrm{Tr}[\mathbb{I}\rho]^2 \\ &\leq \mathrm{Tr}[\mathbb{I}^2]\,\mathrm{Tr}[\rho^2] \\ &= \mathrm{Tr}[\mathbb{I}]f(\rho) \\ &= df(\rho), \end{aligned}$$

which after rearranging yields $f(\rho) \geq \frac{1}{d}$.

3 P. (b) Show that $f(\rho) = 1$ if and only if ρ is a pure state.

Solution

For $\rho = |\psi\rangle\langle\psi|$ a pure state, we have that

$$\begin{aligned} f(\rho) &= \mathrm{Tr}[|\psi\rangle\langle\psi|\psi\rangle\langle\psi|] \\ &= \mathrm{Tr}[|\psi\rangle\langle\psi|] \\ &= 1. \end{aligned}$$

We recall from the last exercise the for $\{\lambda_i\}_{i=1}^d$ the eigenvalues of ρ, we have that

$$\sum_{i=1}^d \lambda_i^2 \leq \sum_{i=1}^d \lambda_i = 1.$$

The only way equality can be achieved is if for all i we have $\lambda_i = \lambda_i^2$, which can only happen if $\lambda_i \in \{0, 1\}$. As there can only be a single 1 among the eigenvalues, we conclude that ρ has to be a projector, and hence a pure state, to fulfill $f(\rho) = 1$

3 P. (c) What state attains the lower bound $f(\rho) = \frac{1}{d}$? Argue that $f(\rho)$ can be seen as a measure of "purity" of the state ρ.

Solution

It is readily verified that $\rho = \frac{1}{d}\mathbb{I}$ is a valid quantum state that has $f(\rho) = \frac{1}{d}$. We can find this state by revisiting the argument from exercise (a), considering that the ρ that saturates (achieves the smallest value) in the Cauchy-Schwarz inequality must be proportional to the identity $\mathbb{I}$ in analogy to the Cauchy-Schwarz inequality for vectors.
This state is called the *maximally mixed state*, as it represents a mixture of all possible pure states of a system in any basis and is thus the least pure state we could have. Computing $f(\rho)$ thus determines how close a state is to being pure, where larger values signal larger purity and lower values a higher degree of mixedness.

5 P. Exercise 4 (Decompositions of mixed states)**.**

2 P. (a) Show that every mixed state of a finite-dimensional quantum system can be written as a convex decomposition of pure states.

Solution

Let $\rho \in L(\mathcal{H})$ for some finite-dimensional Hilbert space $\mathcal{H}$. Then ρ is in particular a self-adjoint linear operator, so by the spectral theorem, ρ has eigenvalues $\lambda_1, \ldots, \lambda_d \in \mathbb{R}$ with associated normalized eigenvectors $|\psi_1\rangle, \ldots, |\psi_d\rangle \in \mathcal{H}$. In Dirac bra-ket notation, this means

$$\rho = \sum_{i=1}^d \lambda_i |\psi_i\rangle\langle\psi_i| \,.$$

As $\rho \geq 0$, we have $\lambda_i \geq 0$ for all $1 \leq i \leq d$. Moreover, we have $\sum_{i=1}^d \lambda_i = \mathrm{Tr}[\rho] = 1$. Thus, the λ_i are non-negative and sum to 1, so the above decomposition is in fact a convex decomposition into pure states.

2 P. (b) Consider the following two (macroscopically different) preparation schemes of a large number of polarised photons:

Preparation A. For each photon we toss a fair coin. Depending on whether we get head or tail, we prepare the photon to have either vertical or horizontal *linear* polarisation.

Preparation B. For each photon we toss a fair coin. Depending on whether we get head or tail, we prepare the photon to have either left-handed or right-handed *circular* polarisation.

Note: You can simply think of the polarization of the light as a binary variable and of the polarization axis as a local basis. That is, the vertical and horizontal linear polarizations may be identified with the $|0\rangle$ and $|1\rangle$ eigenstates of the Z operator. Likewise you may interpret the left- and right handed circular polarizations as the $|+\rangle = \frac{|0\rangle+|1\rangle}{\sqrt{2}}$ and $|-\rangle = \frac{|0\rangle-|1\rangle}{\sqrt{2}}$ eigenstates of the X operator.

Write down the density matrices $\rho_A^{(m)}$ and $\rho_B^{(m)}$ describing the mixed quantum states obtained after m rounds of Preparations A and B, respectively.

Solution

A single round of Preparation A produces the mixed state

$$\rho_A = \frac{1}{2}|0\rangle\langle 0| + \frac{1}{2}|1\rangle\langle 1| = \frac{1}{2}\mathbb{1}_2 .$$

Similarly, a single round of Preparation B produces the mixed state

$$\rho_B = \frac{1}{2}|+\rangle\langle +| + \frac{1}{2}|-\rangle\langle -| = \frac{1}{2}\mathbb{1}_2 = \rho_A .$$

As the different rounds are independent (since the coin tosses are independent), the quantum states describing m rounds are given by

$$\rho_A^{(m)} = \left(\frac{1}{2}\mathbb{1}_2\right)^{\otimes m} = \frac{1}{2^m}\mathbb{1}_{2^m} = \rho_B^{(m)} .$$

1 P. (c) Use the result of (b) to argue that, having only access to the photons produced by the preparation procedures in (b), we cannot distinguish whether Preparation A or Preparation B was used.

Solution

According to our result in (b), no matter how large m is, we have $\rho_A^{(m)} = \rho_B^{(m)}$. Thus, the m photons coming from the preparation procedures are described by exactly the same mixed quantum state. Hence, there is no measurement that distinguishes the two preparation procedures when given only access to the photons.

4 P. Bonus Exercise 1. Argue that if it were possible to distinguish Preparation A from Preparation B (from the previous exercise) by measuring the photons, then this could be used to communicate a bit of information without actually sending any (classical or quantum) information carrier. Which fundamental physical principle would this violate? (Hint: What is the reduced state of the maximally entangled state $|\Omega\rangle = \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle) = \frac{1}{\sqrt{2}}(|{++}\rangle + |{--}\rangle)$?)

Solution

Protocol: EPR setting with Bell state $|\Omega\rangle = \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle) = \frac{1}{\sqrt{2}}(|{+}{+}\rangle + |{-}{-}\rangle)$
Bob chooses a measurement setting, X or Z, and measures his half of the EPR state without observing the measurement outcome. If Bob chooses to measure in the X-basis, the post-measurement state on Alice's side reads $\frac{1}{2}(|+\rangle\langle+| + |-\rangle\langle-|)$. If Bob chooses to measure in the Z-basis, the post-measurement state on Alice's side reads $\frac{1}{2}(|0\rangle\langle 0| + |1\rangle\langle 1|)$. If Alice had a way of distinguishing the two preparation procedures from a single photon, the two parties could have communicated a bit (encoded as $\{X, Z\}$) without Bob sending any message to Alice. This would violate locality. More generally, if Alice had a way of distinguishing the two preparation procedures from m photons, then Alice and Bob could start from m maximally entangled states to communicate a single bit without Bob sending any message to Alice, again violating locality.

Composite Quantum Systems

Next, we will see that the generalization to density matrices is a necessary one if we want to study subsystems. Consider a bipartite system AB with Hilbert space $\mathcal{H} = \mathbb{C}^{d_A} \otimes \mathbb{C}^{d_B}$ and an observable that is only supported on the subsystem A as $O_A \otimes \mathbb{I}_B$. We will see that the restriction to a subsystem is described by the *partial trace*: For a a linear operator (matrix) $M\colon \mathcal{H} \to \mathcal{H}$ on the composite system AB, the partial trace with respect to the system B is defined as

$$\mathrm{Tr}_B[M] = \sum_{j=1}^{d_B} (\mathbb{I}_A \otimes \langle j|_B) M (\mathbb{I}_A \otimes |j\rangle_B), \tag{9}$$

where $\{|j\rangle_B\}$ is an arbitrary orthonormal basis for $\mathbb{C}^{d_B}$ (as with the trace, this definition is independent of the particular choice of the basis). In quantum information theory, we usually say "we trace out the system B".

10 P. Exercise 5.

2 P. (a) As a technical prerequisite, prove that a self-adjoint operator is positive semi-definite, i.e. has only non-negative eigenvalues, if and only if $\langle v|M|v\rangle \geq 0$ for all $|v\rangle$.

Solution

If $\langle v|M|v\rangle \geq 0$ for all $|v\rangle$, this especially holds for the eigenvectors of M, and hence the eigenvalues are all non-negative, establishing one direction. The other direction is verified by direct computation by expanding both $M = \sum_{i=1}^{d} m_i |i\rangle\langle i|$ and $|v\rangle = \sum_{i=1}^{d} v_i |i\rangle$ in the eigenbasis of M. We then have

$$\begin{aligned}\langle v|M|v\rangle &= \left(\sum_{i=1}^{d} v_i^* \langle i|\right)\left(\sum_{j=1}^{d} m_j |j\rangle\langle j|\right)\left(\sum_{k=1}^{d} v_k |k\rangle\right)\\ &= \sum_{i=1}^{d} v_i^* v_i m_i\\ &= \sum_{i=1}^{d} |v_i|^2 m_i.\end{aligned}$$

As all $|v_i|^2 \geq 0$ and by assumption $m_i \geq 0$, the desired result follows.

3 P. (b) Show that the partial trace of a state with respect to the system B (density operator) is a valid state on the subsystem A.

Solution

We have to verify three properties: self-adjointness, unit trace and positive-semidefiniteness. We first observe that taking the adjoint is additive and hence:

$$\begin{aligned}
(\mathrm{Tr}_B(\rho))^\dagger &= \left(\sum_{j=1}^{d_B}(\mathbb{I}_A \otimes \langle j|)\rho(\mathbb{I}_A \otimes |j\rangle)\right)^\dagger \\
&= \sum_{j=1}^{d_B}((\mathbb{I}_A \otimes \langle j|)\rho(\mathbb{I}_A \otimes |j\rangle))^\dagger \\
&= \sum_{j=1}^{d_B}(\mathbb{I}_A \otimes \langle j|)\rho^\dagger(\mathbb{I}_A \otimes |j\rangle) \\
&= \sum_{j=1}^{d_B}(\mathbb{I}_A \otimes \langle j|)\rho(\mathbb{I}_A \otimes |j\rangle) \\
&= \mathrm{Tr}_B(\rho).
\end{aligned}$$

Next, we prove that the trace is preserved under the partial trace:

$$\begin{aligned}
\mathrm{Tr}(\mathrm{Tr}_B(\rho)) &= \mathrm{Tr}\left[\sum_{j=1}^{d_B}(\mathbb{I}_A \otimes \langle j|)\rho(\mathbb{I}_A \otimes |j\rangle)\right] \\
&= \sum_{i=1}^{d_A}\langle i|\sum_{j=1}^{d_B}(\mathbb{I}_A \otimes \langle j|)\rho(\mathbb{I}_A \otimes |j\rangle)|i\rangle \\
&= \sum_{i=1}^{d_A}\sum_{j=1}^{d_B}((\langle i| \otimes \langle j|)\rho(|i\rangle \otimes |j\rangle) \\
&= \mathrm{Tr}(\rho) = 1.
\end{aligned}$$

Here, the last line follows from observing that $|i\rangle\otimes|j\rangle$ is by definition an orthonormal basis for $\mathcal{H}$. For positivity, we use consider the expectation value

$$\langle\psi|_A \mathrm{Tr}_B(\rho)|\psi\rangle_A = \sum_j(\langle\psi|_A \otimes \langle j|_B)\rho(|\psi\rangle_A \otimes |j\rangle_B) \geq 0$$

since ρ is positive semi-definite by assumption.

2 P. (c) Prove that for any state ρ_{AB} we have

$$\mathrm{Tr}[\rho_{AB}(O_A \otimes \mathbb{I}_B)] = \mathrm{Tr}[\mathrm{Tr}_B[\rho_{AB}]O_A]. \tag{10}$$

for all observables O_A. That is, the partial trace is the *reduced state* on the subsystem A.

Solution

First notice that for any operator X_{AB}, $\mathrm{Tr}[X_{AB}] = \mathrm{Tr}_A[\mathrm{Tr}_B[X_{AB}]]$. Now, let $|i\rangle_B$ be a basis of $\mathcal{H}_B$, then

$$\begin{aligned}\mathrm{Tr}_B[\rho_{AB}(O_A \otimes \mathbb{I}_B)] &= \sum_{i=1}^{d_B} (\mathbb{I}_A \otimes \langle i|_B)\rho_{AB}(O_A \otimes \mathbb{I}_B)(\mathbb{I}_A \otimes |i\rangle_B) \\ &= \sum_{i=1}^{d_B} (\mathbb{I}_A \otimes \langle i|_B)\rho_{AB}(O_A \otimes |i\rangle_B) \\ &= \sum_{i=1}^{d_B} (\mathbb{I}_A \otimes \langle i|_B)\rho_{AB}(\mathbb{I}_A \otimes |i\rangle_B)O_A = \mathrm{Tr}_B(\rho_{AB})O_A\end{aligned}$$

where in the second to last equality we used $(O_A \otimes \mathbb{I}_B)(\mathbb{I}_A \otimes |i\rangle_B) = (\mathbb{I}_A \otimes |i\rangle_B)O_A$. Here, $(\mathbb{I}_A \otimes |i\rangle_B)$ can be seen as a map $\mathcal{H}_A \to \mathcal{H}_A \otimes \mathcal{H}_B$, $|\psi\rangle_A \mapsto |\psi\rangle_A \otimes |i\rangle_B$.

3 P. (d) Reduced states of pure states are not necessarily pure. Let $d_A = d_B = d$. Show that there is no pure state $|\psi_A\rangle\langle\psi_A|$ acting on A that satisfies

$$\mathrm{Tr}[\rho_{AB}(O_A \otimes \mathbb{I}_B)] = \mathrm{Tr}[|\psi_A\rangle\langle\psi_A|O_A] \tag{11}$$

for $\rho_{AB} = |\Omega_{AB}\rangle\langle\Omega_{AB}|$ and all observables O_A. Here,

$$|\Omega\rangle := d^{-\frac{1}{2}} \sum_{j=1}^{d} |j,j\rangle \tag{12}$$

is the *maximally entangled state.*

Solution

The statement we need to prove is the following: For any pure state $|\psi_A\rangle$ acting on A, there exists an observable O_A such that

$$\mathrm{Tr}[\rho_{AB}(O_A \otimes \mathbb{I}_B)] \neq \mathrm{Tr}[|\psi_A\rangle\langle\psi_A|O_A], \tag{13}$$

for ρ_{AB} the maximally mixed state introduced in the exercise.
We show it suffices to pick $O_A = |\psi_A\rangle\langle\psi_A|$.
The r.h.s. of Eq. (11) is easy to compute:

$$\mathrm{Tr}[|\psi_A\rangle\langle\psi_A|\psi_A\rangle\langle\psi_A|] = |\langle\psi_A|\psi_A\rangle|^2 = 1. \tag{14}$$

For the l.h.s, we use Eq. (10). For full generality, we allow $O_A = |\phi\rangle\langle\phi|$ be a projector to any pure state on A:

$$\begin{aligned}
\mathrm{Tr}[\rho_{AB}(O_A \otimes \mathbb{I}_B)] &= \mathrm{Tr}\left[\frac{1}{d}\sum_{j=1}^{d}\sum_{k=1}^{d}|j,j\rangle\langle k,k|(|\phi\rangle\langle\phi| \otimes \mathbb{I})\right] && (15)\\
&= \mathrm{Tr}\left[\frac{1}{d}\sum_{j=1}^{d}\sum_{k=1}^{d}(|j\rangle\langle k| \otimes |j\rangle\langle k|)\left(|\phi\rangle\langle\phi| \otimes \sum_{i=1}^{d}|i\rangle\langle i|\right)\right] && (16)\\
&= \sum_{i=1}^{d}\mathrm{Tr}\left[\frac{1}{d}\sum_{j=1}^{d}\sum_{k=1}^{d}|j\rangle\langle k|\phi\rangle\langle\phi| \otimes \langle i|j\rangle\langle k|i\rangle\right] && (17)\\
&= \sum_{i=1}^{d}\mathrm{Tr}\left[\frac{1}{d}|i\rangle\langle i|\phi\rangle\langle\phi|\right] && (18)\\
&= \frac{1}{d}\sum_{i=1}^{d}|\langle i|\phi\rangle|^2 && (19)\\
&= \frac{1}{d}, && (20)
\end{aligned}$$

where we exploited that for any orthonormal basis $\sum_{i=1}^{d}|\langle i|\phi\rangle|^2 = \||\phi\rangle\|_2^2 = 1$. So we see that the l.h.s. equals $1/d$ for any pure state projector. In particular, this is also true for our choice $O_A = |\psi_A\rangle\langle\psi_A|$.
In all, we saw that the l.h.s. is equal to $1/d$, whereas the r.h.s. is equal to 1, thus solving the exercise.

Total Points: 30 (+4)