CHAPTER 3

Tensor Networks

Quantum learning involves systems with many degrees of freedom (e.g. many qubits), which are described by tensor products of Hilbert spaces. In some circumstances, the standard notation describing operators and states in tensor product Hilbert spaces can be unwieldy, and obscures certain structural intuitions. Here we develop a standardized *diagrammatic* notation for manipulating tensors on tensor product Hilbert spaces which illuminates various kinds of proofs. We will use this notation on occasion in this book.

Before diving into our review, we begin with an anecdote. One of the earlier usages of tensor diagrams is by Roger Penrose, which is why in some communities such diagrams are called 'Penrose graphical notation'. Penrose relayed to one of the authors the following story. When Penrose was a PhD student at Cambridge under the direction of Hodge, he developed his graphical notation to help him better visualize certain proofs in algebraic geometry. One day when he met with Hodge to report his progress, Penrose used these diagrams on Hodge's blackboard; Hodge was puzzled since he had never seen such diagrams before. Penrose said that he would go write up a note explaining the notation to Hodge, and did so in the ensuing week. He gave Hodge a 50 page manuscript with many diagrams, and by Penrose's account, Hodge thought that Penrose must have lost his mind, given that he was claiming tensor algebra could be performed by manipulating a bunch of squiggles. Penrose was of course correct, and so we commence with the squiggles.

1. Review of tensor network diagrams

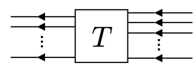
As promised, the so-called 'tensor network' diagrams will render the index contraction of higher-rank tensors more transparent than standard notations. Our discussion here is based off of [**CCHL22**], and we also refer the interested reader to [**Lan11**, **BC17**] for a more comprehensive overview of tensor networks.

Diagrams for individual tensors

Throughout, a rank (m, n) tensor will mean a multilinear map $T : \mathcal{H}^{*\,\otimes m} \otimes \mathcal{H}^{\otimes n} \to \mathbb{C}$. If $\{|i\rangle\}$ is an orthonormal basis of $\mathcal{H}$, then in bra–ket notation T admits the expansion

$$T = \sum_{\substack{i_1,\ldots,i_m \\ j_1,\ldots,j_n}} T^{i_1\cdots i_m}_{j_1\cdots j_n} \left(|i_1\rangle \otimes \cdots \otimes |i_m\rangle\right)\left(\langle j_1| \otimes \cdots \otimes \langle j_n|\right).$$

for some $T^{i_1\cdots i_m}_{j_1\cdots j_n} \in \mathbb{C}$. A quantum state $|\Psi\rangle$ on $\mathcal{H}$ is thus a rank $(1, 0)$ tensor (a map $\mathcal{H}^* \to \mathbb{C}$), and its dual $\langle\Psi|$ is rank $(0, 1)$. Moreover, a matrix $M = \sum_{ij} M^i_j\, |i\rangle\langle j|$ is a rank $(1,1)$ tensor. We will depict T diagrammatically as



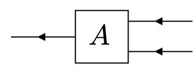
(15)

which carries m outgoing legs on the left and n incoming legs on the right. Each leg corresponds to one index of $T^{i_1\cdots i_m}_{j_1\cdots j_n}$. Our convention is that outgoing legs are ordered counter-clockwise, while incoming legs are ordered clockwise. Concretely, in (15) the upper left outgoing leg is i_1, the one below is i_2, etc.; symmetrically on the right, the top incoming leg is j_1, the next is j_2, and so forth.

Tensor contraction

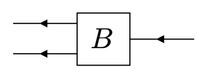
We now describe how to indicate tensor-network contractions. For illustration, consider a rank $(2,1)$ tensor

$$A = \sum_{ijk} A^i_{jk}\, |i\rangle\, (\langle j| \otimes \langle k|) \;=$$



and a rank $(1,2)$ tensor

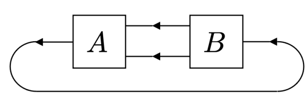
$$B = \sum_{\ell mn} B^{mn}_{\ell}\, (|m\rangle \otimes |n\rangle)\, \langle \ell| \;=$$



Suppose we wish to evaluate

$$\sum_{ijk} A^i_{jk} B^{jk}_i\,. \tag{16}$$

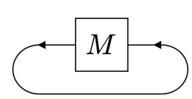
Here lower indices pair with upper indices, reflecting vector–covector contraction. The corresponding diagram is



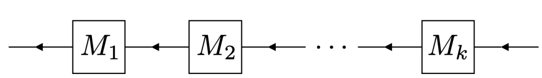
Reading this against (16), the contracted indices are precisely those whose incoming and outgoing legs are glued. Only legs with compatible orientations may be joined, encoding the rule that vectors contract with covectors.

As another instance, for a matrix $M = \sum_{ij} M^i_j\, |i\rangle\langle j|$, the trace is drawn as

$$\mathrm{tr}(M) = \sum_i M^i_i =$$



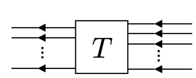
If $M_1, M_2, ..., M_k$ are matrices, then their product $M_1 M_2 \cdots M_k$ appears as



Multiplication by a scalar

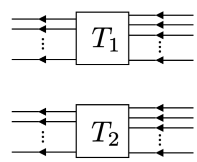
For a tensor T and scalar α, we notate αT. In diagrams we simply write

α

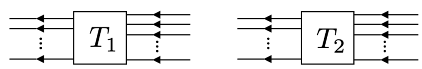


Tensor products

Given tensors T_1 and T_2, their tensor product $T_1 \otimes T_2$ is represented by



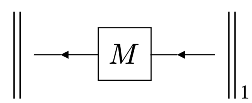
or equivalently by



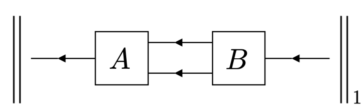
The ordering (e.g. $T_1 \otimes T_2$ versus $T_2 \otimes T_1$) will be evident from context.

Taking norms

Matrix norms are often conveniently expressed in this notation. If M is a matrix, its 1-norm $\|M\|_1$ is indicated by



Here the diagram for M acts as a placeholder inside $\|M\|_1$. This is especially useful when M itself arises from a contraction whose structure we wish to emphasize; for example, if $M = \sum_{ijk\ell} A^i_{k\ell} B^{k\ell}_j \, |i\rangle\langle j|$, then



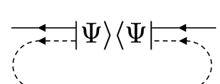
Tensors with legs of different dimensions

Thus far we have treated rank (m, n) tensors as maps $T : \mathcal{H}^{*\otimes m} \otimes \mathcal{H}^{\otimes n} \to \mathbb{C}$. More generally, consider

$$T : (\mathcal{H}_1^* \otimes \cdots \otimes \mathcal{H}_m^*) \otimes (\mathcal{H}_{m+1} \otimes \cdots \otimes \mathcal{H}_{m+n}) \to \mathbb{C},$$

where the Hilbert spaces need not be isomorphic. The same diagrammatic rules apply, with the additional restriction that two legs may be contracted only if they correspond to a Hilbert space and its dual of the same dimension.

As an example, take a state $|\Psi\rangle \in \mathbb{C}^2 \otimes \mathbb{C}^3$ and its density operator $|\Psi\rangle\langle\Psi|$. We will draw the $\mathbb{C}^2$ (qubit) legs as solid and the $\mathbb{C}^3$ (qutrit) legs as dotted. A partial trace over the qutrit subsystem reads



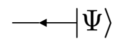
(17)

We return to partial traces in greater detail below.

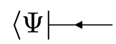
Identity operator

The identity on $\mathcal{H}$ is represented by a single oriented line:

Thus for a state in $\mathcal{H}$,

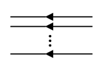


left-multiplication by the identity leaves the diagram (and therefore the state) unchanged. Likewise, for the dual state

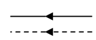


right-multiplying by the identity returns the same diagram.

For k copies, the identity on $\mathcal{H}^{\otimes k}$ is



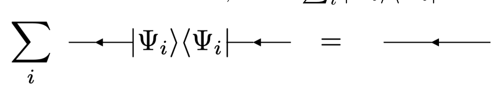
If instead the overall Hilbert space is $\mathcal{H} \otimes \mathcal{H}'$ with different factor dimensions, we take $\mathcal{H}$-legs to be solid and $\mathcal{H}'$-legs dotted; then



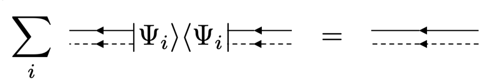
and the evident generalization covers more than two distinct factors.

Resolutions of the identity

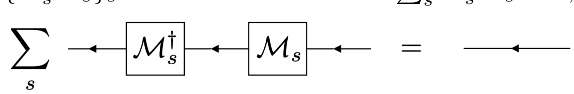
If $\{|\Psi_i\rangle\}_i$ is an orthonormal basis of $\mathcal{H}$, then $\sum_i |\Psi_i\rangle\langle\Psi_i| = \mathbb{1}$ is depicted by



If instead $\{|\Psi_i\rangle\}_i$ resolves the identity on $\mathcal{H} \otimes \mathcal{H}'$ with non-identical factor dimensions, we analogously draw



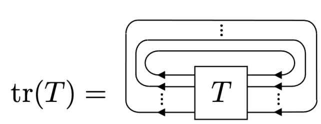
Similarly, if $\{M_s^\dagger M_s\}_s$ is a POVM on $\mathcal{H}$ with $\sum_s M_s^\dagger M_s = \mathbb{1}$, we write



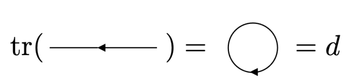
and the same idea extends to $\mathcal{H} \otimes \mathcal{H}'$ and larger tensor products.

Taking traces and partial traces

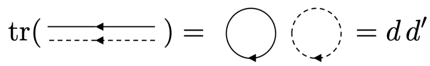
For a rank (n, n) tensor $T : \mathcal{H}^{*\otimes n} \otimes \mathcal{H}^{\otimes n} \to \mathbb{C}$, the trace is $\mathrm{tr}(T) = \sum_{i_1,\ldots,i_n} T^{i_1\cdots i_n}_{i_1\cdots i_n}$, drawn as



A particularly useful identity is the trace of $\mathbb{1} = \sum_i |i\rangle\langle i|$, viewed as a rank $(1, 1)$ tensor:



Thus a closed loop equals the dimension of the Hilbert space associated to that curve. For $\mathbb{1}_{d\times d}\otimes\mathbb{1}_{d'\times d'}$ on $\mathcal{H}\otimes\mathcal{H}'$, where $\dim(\mathcal{H})=d$ and $\dim(\mathcal{H}')=d'$, we have



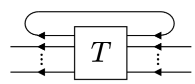
with solid denoting $\mathcal{H}$ and dotted denoting $\mathcal{H}'$.

Partial traces are handled analogously. Define the partial trace over the 'kth subsystem' by

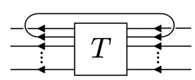
$$\mathrm{tr}_k(T)=\sum_{\substack{i_1,\dots,i_{k-1},i_{k+1},\dots,i_n\\ j_1,\dots,j_{k-1},j_{k+1},\dots,j_n}}\left(\sum_{i_k}T^{i_1\cdots i_n}_{j_1\cdots j_n}\right)|i_1\rangle\langle j_1|\otimes\cdots\otimes|i_{k-1}\rangle\langle j_{k-1}|\otimes|i_{k+1}\rangle\langle j_{k+1}|\otimes\cdots\otimes|i_n\rangle\langle j_n|\,.$$

Note that $\mathrm{tr}_\ell(\mathrm{tr}_k(T))=\mathrm{tr}_k(\mathrm{tr}_\ell(T))$, so we may write $\mathrm{tr}_{k,\ell}(T)$, and $\mathrm{tr}_{1,\dots,n}(T)=\mathrm{tr}(T)$.

Diagrammatically, the partial trace over the first subsystem is

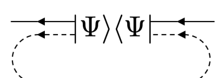
$\mathrm{tr}_1(T) =$ 

Over the second subsystem:

$\mathrm{tr}_2(T) =$ 

and so on.

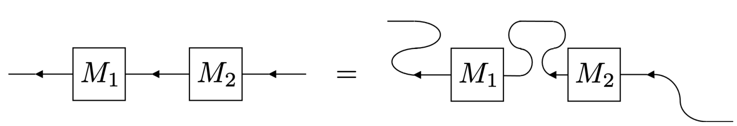
If the legs of a tensor correspond to Hilbert spaces of differing dimensions, traces and partial traces are still available whenever the paired spaces match. For example, if $T:(\mathcal{H}_1^*\otimes\cdots\otimes\mathcal{H}_n^*)\otimes(\mathcal{H}_1'\otimes\cdots\otimes\mathcal{H}_m')\to\mathbb{C}$ and $\mathcal{H}_k=\mathcal{H}_k'$, we may compute $\mathrm{tr}_k(T)$. As a simple instance, for $|\Psi\rangle\in\mathcal{H}\otimes\mathcal{H}'$, the density operator $|\Psi\rangle\langle\Psi|$ is a $(2,2)$ tensor $(\mathcal{H}^*\otimes\mathcal{H}'^{*})\otimes(\mathcal{H}\otimes\mathcal{H}')\to\mathbb{C}$, and

$\mathrm{tr}_2(|\Psi\rangle\langle\Psi|) =$ 

which matches the example in (17); a similar figure represents $\mathrm{tr}_1(|\Psi\rangle\langle\Psi|)$.

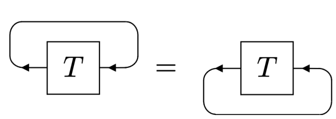
Isotopies

Tensor-network diagrams are interpreted up to isotopy of the legs: bending or smoothly deforming them does not change the meaning. For instance, for a product M_1M_2 we may equally draw

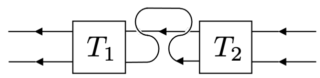


and the same holds in other cases.

Isotopies need not be planar; e.g.

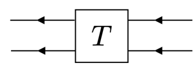


Leg crossings are also allowed:

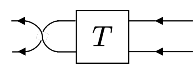


and we will not distinguish over- from under-crossings.

However, we keep fixed the relative ordering of the incoming and outgoing endpoints. Reordering them would permute the tensor factors on which the tensor acts. For example, let $T : (\mathcal{H}_1^* \otimes \mathcal{H}_2^*) \otimes (\mathcal{H}_1 \otimes \mathcal{H}_2) \to \mathbb{C}$ be drawn as



Then



corresponds to a tensor on $(\mathcal{H}_2^* \otimes \mathcal{H}_1^*) \otimes (\mathcal{H}_1 \otimes \mathcal{H}_2)$, where the dual factors have been swapped. See also the discussion of permutation operators below.

Permutation operators

Let S_k denote the permutation group on k elements, and let $\tau \in S_k$. Define $\mathrm{Perm}(\tau)$ acting on $\mathcal{H}^{\otimes k}$ by

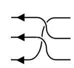
$$\mathrm{Perm}(\tau)|\psi_1\rangle \otimes |\psi_2\rangle \otimes \cdots \otimes |\psi_n\rangle = |\psi_{\tau^{-1}(1)}\rangle \otimes |\psi_{\tau^{-1}(2)}\rangle \otimes \cdots \otimes |\psi_{\tau^{-1}(n)}\rangle$$

and extend linearly. With this convention we have

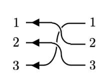
$$\mathrm{Perm}(\tau) \cdot \mathrm{Perm}(\sigma) = \mathrm{Perm}(\tau\sigma)$$

where $\tau\sigma$ denotes the group product (composition $\tau \circ \sigma$).

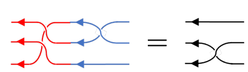
These representations admit an especially transparent diagrammatics. For S_3 and $\tau = (123)$, we draw



which becomes clear upon labeling the endpoints:



The group product is just as visible; e.g., $\mathrm{Perm}((123)) \cdot \mathrm{Perm}((12))$ is



with $\mathrm{Perm}((123))$ drawn in red and $\mathrm{Perm}((12))$ in blue for emphasis; allowable isotopies (without reordering endpoints) show the result is $\mathrm{Perm}((23))$. Note also that horizontally flipping the diagram for $\mathrm{Perm}(\tau)$ yields that for $\mathrm{Perm}(\tau^{-1})$.

As another example, acting with $\mathrm{Perm}((123))$ on a state $|\Psi\rangle \in \mathcal{H}^{\otimes 3}$ gives

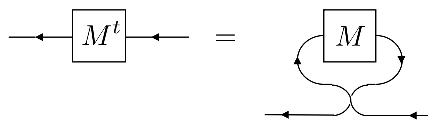
making it evident that the tensor factors are permuted according to $(123)^{-1} = (132)$.

In later arguments, when no confusion can arise, we will abbreviate $\mathrm{Perm}(\tau)$ simply as τ.

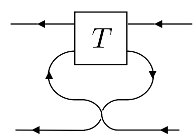
Transposes and partial transposes

Let $M = \sum_{i,j} M^i_j \, |i\rangle\langle j|$ be viewed as a rank $(1,1)$ tensor. Its transpose $M^t = \sum_{i,j} M^i_j \, |j\rangle\langle i|$ can be indicated diagrammatically as follows.

Here we dualize each leg by reversing its arrow, then use isotopy to reorient so the in-arrow enters from the right and the out-arrow exits to the left; this is done to match the orientation of the diagram on the left.



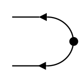
For a higher-rank tensor, e.g. a rank $(2,2)$ tensor $T = \sum_{ijk\ell} T^{ij}_{k\ell} \, |i\rangle\langle k| \otimes |j\rangle\langle \ell|$, we may transpose only one subsystem; the partial transpose on the second subsystem, $\sum_{ijk\ell} T^{ij}_{k\ell} \, |i\rangle\langle k| \otimes |\ell\rangle\langle j|$, is shown as

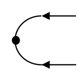


and the same notation extends in the obvious way to higher rank.

Maximally entangled state

The maximally entangled state is $|\Omega\rangle = \sum_i |i\rangle|i\rangle$ in the computational basis, taken unnormalized. We depict $|\Omega\rangle$ and its Hermitian conjugate by

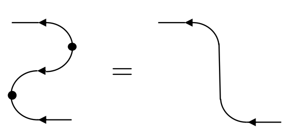
$|\Omega\rangle =$ 

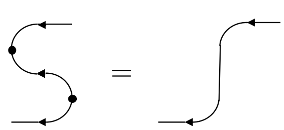
$\langle\Omega| =$ 

Let $\mathcal{H}_A \simeq \mathcal{H}_B \simeq \mathcal{H}_C$. Then

$$(\mathbb{1}_A \otimes \langle\Omega|_{BC})\,(|\Omega\rangle_{AB} \otimes \mathbb{1}_C) = \sum_i |i\rangle_A\langle i|_C$$
$$(\langle\Omega|_{AB} \otimes \mathbb{1}_C)\,(\mathbb{1}_A \otimes |\Omega\rangle_{BC}) = \sum_i |i\rangle_C\langle i|_A$$

which we draw as





One can view the black dot as implementing a transpose, since it flips the leg's orientation; two such dots cancel, reflecting that a double transpose is the identity.

2. Some applications

We give three initial applications of tensor network diagrams to illustrate how they illuminate certain kinds of mathematical relationships and proofs in multilinear algebra.

Example 1: A property of the maximally entangled state. First, consider a Hilbert space $\mathcal{H} \simeq \mathbb{C}^d$ with an orthonormal basis $\{|i\rangle\}_{i=1}^d$. As discussed before, the identity matrix can be written as $\mathbb{1} = \sum_{i=1}^d |i\rangle\langle i|$. Then we have the following definition:

Definition 45 (Maximally entangled state)**.** *The (normalized)* ***maximally entangled state*** *on* $\mathbb{C}^d \otimes \mathbb{C}^d$ *is given by*

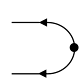
$$|\Phi^+\rangle := \frac{1}{\sqrt{d}} \sum_{i=1}^d |i\rangle|i\rangle\,.$$

This is related to our previous notation above by $|\Phi^+\rangle = \frac{1}{\sqrt{d}}\,|\Omega\rangle$.

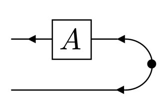
We observe that the maximally entangled state is proportional to the identity matrix if we take the 'transpose' of the subsystems, namely $|\Phi^+\rangle \propto \sum_{i=1}^d |i\rangle(\langle i|)^T = \sum_{i=1}^d |i\rangle|i\rangle$.

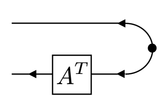
As an aside, the reason we use transpose here, and not Hermitian transposition, is as follows. Consider a more general state $\sum_{i,j=1}^d c_{ij}|i\rangle|j\rangle$ where the c_{ij} are complex. Upon transposing the second subsystem, we find the linear operator $\sum_{i,j=1}^d c_{ij}|i\rangle(|j\rangle)^T = \sum_{i,j=1}^d c_{ij}|i\rangle\langle j|$, and conversely we can go from an operator back to a state via a transposition. Note that the transpose is inert if we group the c_{ij}'s with the $|j\rangle$'s, namely $\sum_{i,j=1}^d |i\rangle(c_{ij}\,|j\rangle)^T = \sum_{i,j=1}^d c_{ij}|i\rangle\langle j|$. But if instead we considered Hermitian conjugation then, we would have $\sum_{i,j=1}^d c_{ij}|i\rangle(|j\rangle)^\dagger = \sum_{i,j=1}^d c_{ij}|i\rangle\langle j|$ and $\sum_{i,j=1}^d |i\rangle(c_{ij}\,|j\rangle)^\dagger = \sum_{i,j=1}^d c_{ij}^*|i\rangle\langle j|$, which are not equal. That is, it would matter if we 'grouped' the c_{ij}'s with the $|j\rangle$'s or not. Said a different way, taking the 'partial Hermitian conjugation' of a state, operator, or tensor *violates* multilinearity, whereas taking a 'partial transpose' maintains multilinearity. This is why partial transposition is a valid operation to do.

With the above considerations in mind, we can represent the maximally entangled state by the tensor

$$|\Phi^+\rangle = \frac{1}{\sqrt{d}}$$


which is proportional to the identity tensor with a transpose inserted in. Now consider an operator $A \otimes \mathbb{1}$ on $\mathbb{C}^d \otimes \mathbb{C}^d$. Then applying this operator to $|\Phi^+\rangle$ and applying elementary tensor network manipulations, we find

$$\frac{1}{\sqrt{d}}$$


$$= \quad \frac{1}{\sqrt{d}}$$


Thus we see that

$$(A \otimes \mathbb{1})|\Phi^+\rangle = (\mathbb{1} \otimes A^T)|\Phi^+\rangle\,,$$

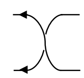
which is a useful property of the maximally entangled state.

Example 2: SWAP trick and SWAP test. Let us define that **swap operator** $\mathsf{SWAP} : \mathbb{C}^d \otimes \mathbb{C}^d \to \mathbb{C}^d \otimes \mathbb{C}^d$ by its action on basis states as

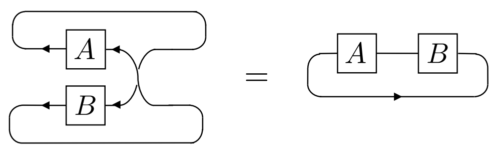
$$\mathsf{SWAP}|i\rangle|j\rangle = |j\rangle|i\rangle$$

for all i, j. The action of SWAP extends to other states by multilinearity, and is a permutation operator on two tensor factors. We observe that $\mathsf{SWAP}^\dagger = \mathsf{SWAP}$, and $\mathsf{SWAP}^2 = \mathbb{1}$, so it is both Hermitian and unitary. In line with our notation above, SWAP is expressed diagrammatically as

SWAP =



Now let A and B be linear operators acting on $\mathbb{C}^d$. We have



and so we have shown that $\mathrm{tr}(\mathsf{SWAP} \cdot A \otimes B) = \mathrm{tr}(AB)$.

The above identity is the foundation for the so-called **swap test**. Suppose we are given two states $|\Psi\rangle$ and $|\Phi\rangle$ and want to test if they are the same or not. Since the states in question are pure, their corresponding density matrices are $|\Psi\rangle\langle\Psi|$ and $|\Phi\rangle\langle\Phi|$. Since SWAP is a Hermitian operator, it is an observable, and so we are welcome to compute the 'observable' expectation value

$$\mathrm{tr}(\mathsf{SWAP}\, |\Psi\rangle\langle\Psi| \otimes |\Phi\rangle\langle\Phi|) = |\langle\Psi|\Phi\rangle|^2\,,$$

which gives the overlap of the two states. Thus, the overlap between two given states is observable; if it is close to one then states are close to being parallel; if it is close to being zero then the states are close to being orthogonal.

The swap test can also be used for the task of **purity testing**. We will return to this in more detail later, but informally, if we are given copies of a density matrix ρ, we would like to ascertain how close it is to being 'rank one' or 'pure'. Diagonalizing ρ as $\rho = \sum_{i=1}^d p_i |v_i\rangle\langle v_i|$ where $p_i \geq 0$ and $\sum_{i=1}^d p_i = 1$, we see that performing the swap test on two copies of ρ we obtain

$$\mathrm{tr}(\mathsf{SWAP}\, \rho \otimes \rho) = \mathrm{tr}(\rho^2) = \sum_{i=1}^d p_i^2\,.$$

If ρ is pure (so that one $p_i = 1$ and all the rest are zero), then the right-hand side of the above is one; if ρ is impure, then the right-hand side is less than one. Indeed, the smallness of $\sum_{i=1}^d p_i^2$ is a measure of the *impurity* of ρ.

As a special case of purity testing, consider a pure state density matrix $|\Psi\rangle\langle\Psi|$ and a maximally mixed density matrix $\frac{1}{d}\,\mathbb{1}$. Then if we perform the swap test on two copies of $|\Psi\rangle\langle\Psi|$, we find

$$\mathrm{tr}(\mathsf{SWAP}\, |\Psi\rangle\langle\Psi| \otimes |\Psi\rangle\langle\Psi|) = 1\,,$$

whereas if we perform the swap test on $\frac{1}{d}\mathbb{1}$ we find

$$\mathrm{tr}(\mathsf{SWAP}\,\frac{\mathbb{1}}{d}\otimes\frac{\mathbb{1}}{d}) = \frac{1}{d}\,.$$

If d is large, then the difference between the 'pure state' and 'maximally mixed state' swap tests is stark; the former is one, and the latter is $1/d$ which is close to zero.

Example 3: A completeness identity for orthonormal operator bases. For our final example, we derive a rather interesting (and useful) identity. First we require a definition.

Definition 46 (Hilbert-Schmidt inner product)**.** *Consider $\mathbb{C}^{d\times d}$ as a vector space of $d\times d$ matrices. We can turn it into a Hilbert space in its own right via the **Hilbert-Schmidt inner product***

$$\langle A, B\rangle_{\mathrm{HS}} := \mathrm{tr}(A^\dagger B) = \sum_{i,j=1}^{d} A_{ij}^* B_{ij}\,.$$

Let $\{M_i\}_{i=1}^{d^2}$ be a complete orthonormal basis of linear operators on $\mathbb{C}^d$. (We note that there must be d^2 such basis elements since the dimension of the space of $d\times d$ matrices is d^2.) Here we mean orthonormal with respect to the Hilbert-Schmidt inner product. Orthonormality means that

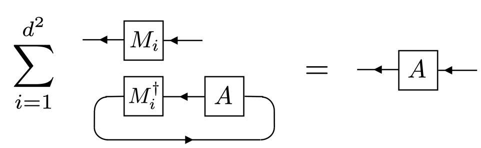
$$\langle M_i, M_j\rangle_{\mathrm{HS}} = \mathrm{tr}(M_i^\dagger M_j) = \delta_{ij}\,, \tag{18}$$

and completeness means that any operator A can be written as $A = \sum_{i=1}^{d^2} c_i\, M_i$ for some coefficients c_i. In fact, using (18) fixes the c_i's to be

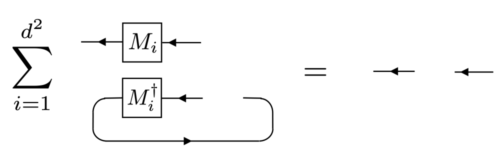
$$A = \sum_{i=1}^{d^2} \mathrm{tr}(M_i^\dagger A)\, M_i\,, \tag{19}$$

namely $c_i = \langle M_i, A\rangle_{\mathrm{HS}} = \mathrm{tr}(M_i^\dagger A)$.

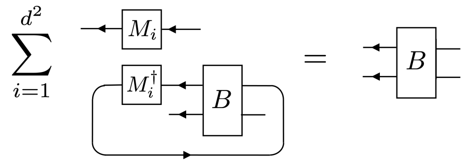
We can write (19) diagrammatically as



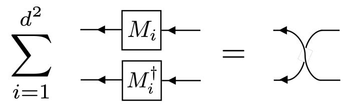
and since it holds for all A, we can remove the A to find the tensor identity



Since the above is a tensor identity, we are welcome to stick in an operator B on $\mathbb{C}^d \otimes \mathbb{C}^d$; this gives



Now if we happen to choose $B = \mathsf{SWAP}$, we find



where we have labeled the ends of the tensor legs with indices for clarity. The above can be written in an algebraic form as the identity

$$\sum_{i=1}^{d} M_i \otimes M_i^\dagger = \mathsf{SWAP}\,.$$

This is a striking identity. As an example if $d = 2^n$, we can let $\{M_i\}_{i=1}^{4^n}$ be the set of n-qubit normalized Pauli strings $\{\frac{1}{2^{n/2}}\,P_i\}_{i=1}^{4^n}$ which form an orthonormal basis of $\mathbb{C}^{2^n \otimes 2^n}$ with respect to the Hilbert-Schmidt inner product. Thus we find

$$\frac{1}{2^n}\sum_{i=1}^{4^n} P_i \otimes P_i = \mathsf{SWAP}\,,$$

where we have used that $P_i^\dagger = P_i$ for Pauli strings. Since the Paulis include the identity operator, sometimes the above is rearranged as

$$\sum_{P_i \neq \mathbb{1}} P_i \otimes P_i = 2^n\,\mathsf{SWAP} - \mathbb{1} \otimes \mathbb{1}\,.$$