Vectorization is great and you should all use it

These notes introduce a bunch of simple but helpful vectorization identities to try and emphasize why this formalism is useful. Essentially these notes are some things that I wish I'd noticed/been told sooner. The moto to take away is that: *Vectorization makes life easier because matrices (e.g. density operators) become vectors and linear maps (e.g. channels) become matrices.*

Contents

1 Vectorization

To recap. Let A be an operator on a Hilbert space with orthonormal basis $\{|i\rangle\}$, written as

$$A = \sum_{i,j} a_{ij} \; |i\rangle \langle j| \, .$$

We define its vectorized form as

$$|\mathrm{vec}(A)\rangle = \sum_{i,j} a_{ij} \; |i\,j\rangle = \begin{pmatrix} a_{11} \\ a_{12} \\ \vdots \\ a_{1d} \\ a_{21} \\ a_{22} \\ \vdots \\ a_{2d} \\ \vdots \\ a_{d1} \\ \vdots \\ a_{dd} \end{pmatrix}$$

where $|i\,j\rangle \equiv |i\rangle \otimes |j\rangle$ is the standard basis in the tensor product space.

Notice that the vectorization operation $\mathrm{vec}(\cdot)$ is a bijection. In particular, one can take the "de-vectorization" map $\text{de-vec}(\cdot) = \sum_{i,j} \langle i,j|\cdot\rangle |i\rangle\langle j|$, such that $\text{de-vec}(|\mathrm{vec}(A)\rangle) = A$.

2 The super useful Ricochet Identity

A super useful vectorization identity is that:

$$|\mathrm{vec}(AXB)\rangle = (A \otimes B^T)|\mathrm{vec}(X)\rangle \, . \tag{1}$$

You saw the proof in an earlier lecture but I'll just include it here for completeness:

Proof. Using the shorthand $X_{ij} = \langle i|\, X \,|j\rangle$, the vectorization of X is defined by

$$|\mathrm{vec}(X)\rangle = \sum_{i,j} X_{ij} \; |i\rangle \otimes |j\rangle \; .$$

Thus,

$$|\mathrm{vec}(AXB)\rangle = \sum_{i,j} (AXB)_{ij} \; |i\rangle \otimes |j\rangle \; ,$$

with

$$(AXB)_{ij} = \sum_{k,l} A_{ik} \, X_{kl} \, B_{lj} \, .$$

It follows that

$$|\mathrm{vec}(AXB)\rangle = \sum_{i,j,k,l} A_{ik}\, X_{kl}\, B_{lj}\ |i\rangle \otimes |j\rangle\ .$$

On the other hand, we have

$$(A \otimes B^T)\, |\mathrm{vec}(X)\rangle = (A \otimes B^T) \sum_{k,l} X_{kl}\ |k\rangle \otimes |l\rangle\ .$$

Using the definition of the Kronecker product,

$$(A \otimes B^T)(|k\rangle \otimes |l\rangle) = A\, |k\rangle \otimes B^T\, |l\rangle = \sum_{i,j} A_{ik}\, B^T_{jl}\ |i\rangle \otimes |j\rangle\ .$$

Thus,

$$(A \otimes B^T)\, |\mathrm{vec}(X)\rangle = \sum_{i,j,k,l} A_{ik}\, X_{kl}\, B^T_{jl}\ |i\rangle \otimes |j\rangle\ .$$

Noting that by definition $B^T_{jl} = B_{lj}$ and that the dummy indices are summed over, we can relabel the summation indices to see that this expression is identical to the one obtained for $|\mathrm{vec}(AXB)\rangle$. Hence,

$$|\mathrm{vec}(AXB)\rangle = \left(A \otimes B^T\right) |\mathrm{vec}(X)\rangle\ ,$$

which completes the proof. □

3 Trace as an inner product

For two operators

$$A = \sum_{i,j} a_{ij}\ |i\rangle\, \langle j| \quad \text{and} \quad B = \sum_{k,l} b_{kl}\ |k\rangle\, \langle l|\,,$$

their Hilbert–Schmidt inner product is defined as

$$\langle A, B\rangle_{\mathrm{HS}} := \mathrm{Tr}\left(A^\dagger B\right)\ .$$

In vectorized form this is equivalent to the (vector) inner product between the vectorized versions of A and B. That is,

$$\mathrm{Tr}\left(A^\dagger B\right) = \langle \mathrm{vec}(A)|\mathrm{vec}(B)\rangle \tag{2}$$

We can see this by writing out the two expressions explicitly and comparing. Namely, we have

$$\langle A, B\rangle_{\mathrm{HS}} := \mathrm{Tr}\left(A^\dagger B\right) = \mathrm{Tr}\left(\sum_{iji'j'} a^*_{ij} b_{i'j'} |j\rangle\langle i|i'\rangle\langle j'|\right) = \sum_{ijj'k} a^*_{ij} b_{ij'} \langle k|j\rangle\langle j'|k\rangle = \sum_{ik} a^*_{ik} b_{ik}\,.$$

Similarly, as

$$|\mathrm{vec}(A)\rangle = \sum_{i,j} a_{ij}\ |i\,j\rangle \quad \text{and} \quad |\mathrm{vec}(B)\rangle = \sum_{k,l} b_{kl}\ |k\,l\rangle\,,$$

the inner product in the vectorized space is

$$\langle \mathrm{vec}(A)|\mathrm{vec}(B)\rangle = \sum_{i,k=j} a^*_{ij}\, b_{ij} \equiv \sum_{i,k} a^*_{ik}\, b_{ik}\,.$$

Thus the two expressions are equivalent, as claimed.

4 Expectation Values

A very common thing to study in quantum information theory (well, all of quantum mechanics, actually) is an expectation value for an observable O after applying a channel $\mathcal{E}$ to a state ρ:

$$\mathrm{Tr}\big[O\,\mathcal{E}(\rho)\big]\,.$$

In vectorized form we see immediately from Eq. (2) that this takes the form:

$$\mathrm{Tr}\big[O\,\mathcal{E}(\rho)\big] = \langle \mathrm{vec}(O) | \mathrm{vec}(\mathcal{E}(\rho)) \rangle$$

That is, we can compute expectation values by computing $|\mathrm{vec}(\mathcal{E}(\rho))\rangle$ and taking its inner product with $|\mathrm{vec}(O)\rangle$. So how do we compute $|\mathrm{vec}(\mathcal{E}(\rho))\rangle$?

5 Unitary Evolutions

Consider a density operator ρ transformed by a unitary U,

$$\rho \longrightarrow U\rho U^\dagger$$

in vectorized form this becomes

$$\Big|\mathrm{vec}(U\rho U^\dagger)\Big\rangle = (U \otimes U^*)\, |\mathrm{vec}(\rho)\rangle\ .$$

We can see this immediately from Eq. (1) that

$$|\mathrm{vec}(AXB)\rangle = (A \otimes B^T)|\mathrm{vec}(X)\rangle \tag{3}$$

and letting $A = U$, $B = U^\dagger$ and $X = \rho$.

6 Quantum Channels in Kraus Form

Consider a quantum channel with Kraus operators $\{K\}$

$$\mathcal{E}(\rho) = \sum_K K\,\rho\,K^\dagger.$$

In vectorization form we have

$$|\mathrm{vec}(\mathcal{E}(\rho))\rangle = \sum_K \Big(K \otimes K^*\Big)\, |\mathrm{vec}(\rho)\rangle\,.$$

Thus the effect of the channel on the state ρ (a matrix) has been transformed into the action of a matrix $M_{\mathcal{E}} := \sum_K \Big(K \otimes K^*\Big)$ acting on the vector $|\mathrm{vec}(\rho)\rangle$.

7 Product of Unitaries

Suppose the applied unitary is a product of unitaries,

$$U = U_n U_{n-1} \cdots U_1.$$

Then the transformation

$$\rho \longrightarrow U \rho U^\dagger$$

becomes, using our earlier result for each unitary,

$$\left| \mathrm{vec}(U \rho U^\dagger) \right\rangle = (U \otimes U^*) \left| \mathrm{vec}(\rho) \right\rangle .$$

Exploiting the multiplicative property of the Kronecker product, we have:

$$\left| \mathrm{vec}(U \rho U^\dagger) \right\rangle = (U_n \otimes U_n^*) \left(U_{n-1} \otimes U_{n-1}^* \right) \cdots (U_1 \otimes U_1^*) \left| \mathrm{vec}(\rho) \right\rangle .$$

8 Concatenation of Channels

Consider two quantum channels $\mathcal{E}^{(1)}$ and $\mathcal{E}^{(2)}$ with Kraus representations:

$$\mathcal{E}^{(1)}(\rho) = \sum_j K_j^{(1)} \rho \left(K_j^{(1)} \right)^\dagger, \quad \mathcal{E}^{(2)}(\rho) = \sum_k K_k^{(2)} \rho \left(K_k^{(2)} \right)^\dagger .$$

Their concatenation $\mathcal{E} = \mathcal{E}^{(2)} \circ \mathcal{E}^{(1)}$ acts as

$$\mathcal{E}(\rho) = \mathcal{E}^{(2)} \Big(\mathcal{E}^{(1)}(\rho) \Big).$$

In vectorized form,

$$\left| \mathrm{vec}(\mathcal{E}^{(1)}(\rho)) \right\rangle = \sum_j \left(K_j^{(1)} \otimes \left(K_j^{(1)} \right)^* \right) \left| \mathrm{vec}(\rho) \right\rangle ,$$

and then

$$\left| \mathrm{vec}(\mathcal{E}(\rho)) \right\rangle = \sum_k \left(K_k^{(2)} \otimes \left(K_k^{(2)} \right)^* \right) \left| \mathrm{vec}(\mathcal{E}^{(1)}(\rho)) \right\rangle .$$

Thus, the overall mapping is given by:

$$\left| \mathrm{vec}(\mathcal{E}(\rho)) \right\rangle = \left[\sum_k \left(K_k^{(2)} \otimes \left(K_k^{(2)} \right)^* \right) \right] \left[\sum_j \left(K_j^{(1)} \otimes \left(K_j^{(1)} \right)^* \right) \right] \left| \mathrm{vec}(\rho) \right\rangle .$$

9 Averaging over Unitaries

Consider the averaged expectation value over an ensemble of unitaries:

$$E_{\boldsymbol{\theta}} \operatorname{Tr}\Big[U_{\boldsymbol{\theta}} \rho U_{\boldsymbol{\theta}}^\dagger O \Big] ,$$

where $U_{\boldsymbol{\theta}} \equiv U(\boldsymbol{\theta})$ is some unitary that depends on a real vector of parameters $\boldsymbol{\theta}$. Using the vectorized form, we have:

$$\mathrm{Tr}\Big[U_{\boldsymbol{\theta}}\rho U_{\boldsymbol{\theta}}^{\dagger}\,O\Big] = \langle \mathrm{vec}(O)|\,\Big(U_{\boldsymbol{\theta}} \otimes U_{\boldsymbol{\theta}}^{*}\Big)\,|\mathrm{vec}(\rho)\rangle\,.$$

Since the expectation value is linear, the averaging can be brought inside the sum:

$$E_{\boldsymbol{\theta}}\,\mathrm{Tr}\Big[U_{\boldsymbol{\theta}}\rho U_{\boldsymbol{\theta}}^{\dagger}\,O\Big] = \langle \mathrm{vec}(O)|\,\Big(E_{\boldsymbol{\theta}}\Big[U_{\boldsymbol{\theta}} \otimes U_{\boldsymbol{\theta}}^{*}\Big]\Big)\,|\mathrm{vec}(\rho)\rangle\,.$$

Defining

$$M \equiv E_{\boldsymbol{\theta}}\Big[U_{\boldsymbol{\theta}} \otimes U_{\boldsymbol{\theta}}^{*}\Big],$$

we can express the averaged expectation value succinctly as

$$E_{\boldsymbol{\theta}}\,\mathrm{Tr}\Big[U_{\boldsymbol{\theta}}\rho U_{\boldsymbol{\theta}}^{\dagger}\,O\Big] = \langle \mathrm{vec}(O)|\;M\;|\mathrm{vec}(\rho)\rangle\,.$$

Thus the action of evolving and averaging can be captured via a single matrix M which we sandwich between $|\mathrm{vec}(\rho)\rangle$ and $|\mathrm{vec}(O)\rangle$.

Example: Averaging over a Single-Qubit Z Rotation Consider the single-qubit unitary rotation about the Z-axis:

$$U_z(\theta) = e^{-i\theta Z/2} = \begin{pmatrix} e^{-i\theta/2} & 0 \\ 0 & e^{i\theta/2} \end{pmatrix},$$

where θ is a random variable uniformly distributed in $[0, 2\pi]$. The vectorized form for the average in case 3) is given by:

$$M = E_{\theta}\Big[U_z(\theta) \otimes U_z(\theta)^{*}\Big].$$

We first compute the tensor product:

$$U_z(\theta) \otimes U_z(\theta)^{*} = \begin{pmatrix} e^{-i\theta/2} & 0 \\ 0 & e^{i\theta/2} \end{pmatrix} \otimes \begin{pmatrix} e^{i\theta/2} & 0 \\ 0 & e^{-i\theta/2} \end{pmatrix}.$$

In the standard basis $\{|00\rangle, |01\rangle, |10\rangle, |11\rangle\}$, this tensor product is diagonal:

$$U_z(\theta) \otimes U_z(\theta)^{*} = \mathrm{diag}\Big(e^{-i\theta/2}e^{i\theta/2},\, e^{-i\theta/2}e^{-i\theta/2},\, e^{i\theta/2}e^{i\theta/2},\, e^{i\theta/2}e^{-i\theta/2}\Big) = \mathrm{diag}\Big(1,\, e^{-i\theta},\, e^{i\theta},\, 1\Big).$$

Taking the average over θ yields:

$$M = \frac{1}{2\pi}\int_0^{2\pi} \mathrm{diag}\Big(1,\, e^{-i\theta},\, e^{i\theta},\, 1\Big)\,d\theta.$$

Since

$$\frac{1}{2\pi}\int_0^{2\pi} e^{\pm i\theta}\,d\theta = 0,$$

we obtain:

$$M = \mathrm{diag}\Big(1,\, 0,\, 0,\, 1\Big).$$

Thus, we can write:

$$M = |00\rangle\langle 00| + |11\rangle\langle 11| .$$

This matrix M then captures the average effect of a Z rotation on the vectorized density matrix. Consequently, the averaged expectation value for an observable O is given by:

$$E_\theta \operatorname{Tr}\Big[U_z(\theta)\rho U_z(\theta)^\dagger O\Big] = \langle \mathrm{vec}(O)|M|\mathrm{vec}(\rho)\rangle,$$

with

$$M = |00\rangle\langle 00| + |11\rangle\langle 11| .$$

We note that this is just the vectorized version of the completely dephasing channel. This should make sense to you (particularly if you did my QP2 course) because we know that the effect averaging over random rotations around the Z axis (i.e., *twirling* a state with R_z rotations) kills of all coherences with respect to the Z eigenbasis (the computational basis).

10 The moment operator

More generally, higher order moments of the form

$$E_{\boldsymbol{\theta}}\left(\operatorname{Tr}\Big[U_{\boldsymbol{\theta}}\rho U_{\boldsymbol{\theta}}^\dagger O\Big]^k\right).$$

can be computed in vectorized form by computing an appropriate momentum operator as follows.

We first note that as

$$\operatorname{Tr}[X]^2 = \operatorname{Tr}[X \otimes X]$$

we can write

$$E_{\boldsymbol{\theta}}\left(\operatorname{Tr}\Big[U_{\boldsymbol{\theta}}\rho U_{\boldsymbol{\theta}}^\dagger O\Big]^k\right) = E_{\boldsymbol{\theta}}\left(\operatorname{Tr}\Big[(U_{\boldsymbol{\theta}}\rho U_{\boldsymbol{\theta}}^\dagger O)^{\otimes k}\Big]\right) = E_{\boldsymbol{\theta}}\left(\operatorname{Tr}\Big[(U_{\boldsymbol{\theta}}^{\otimes k}\rho^{\otimes k}(U_{\boldsymbol{\theta}}^\dagger)^{\otimes k})\, O^{\otimes k}\Big]\right).$$

Next we vectorize this as

$$E_{\boldsymbol{\theta}}\left(\langle \mathrm{vec}(O^{\otimes k})|\mathrm{vec}(U_{\boldsymbol{\theta}}^{\otimes k}\rho^{\otimes k}(U_{\boldsymbol{\theta}}^\dagger)^{\otimes k})\rangle\right) = E_{\boldsymbol{\theta}}\left(\langle \mathrm{vec}(O^{\otimes k})|U_{\boldsymbol{\theta}}^{\otimes k}\otimes U_{\boldsymbol{\theta}}^{*\otimes k}|\mathrm{vec}(\rho^{\otimes k})\rangle\right) = \langle \mathrm{vec}(O^{\otimes k})|M_k|\mathrm{vec}(\rho^{\otimes k})\rangle$$

where in the final line we have defined the moment operator

$$M_k = E_{\boldsymbol{\theta}}\left(U_{\boldsymbol{\theta}}^{\otimes k}\otimes U_{\boldsymbol{\theta}}^{*\otimes k}\right).$$

For example, the variance of an expectation value $f_{\boldsymbol{\theta}} = \operatorname{Tr}\Big[U_{\boldsymbol{\theta}}\rho U_{\boldsymbol{\theta}}^\dagger O\Big]$ over a circuit U takes the form

$$\begin{aligned}\mathrm{Var}_{\boldsymbol{\theta}}[f_{\boldsymbol{\theta}}] &= E_{\boldsymbol{\theta}}\left(\operatorname{Tr}\Big[U_{\boldsymbol{\theta}}\rho U_{\boldsymbol{\theta}}^\dagger O\Big]^2\right) - \left(E_{\boldsymbol{\theta}}\left(\operatorname{Tr}\Big[U_{\boldsymbol{\theta}}\rho U_{\boldsymbol{\theta}}^\dagger O\Big]\right)\right)^2 \\ &= \langle \mathrm{vec}(O^{\otimes 2})|M_2|\mathrm{vec}(\rho^{\otimes 2})\rangle - (\langle \mathrm{vec}(O)|M_1|\mathrm{vec}(\rho)\rangle)^2 .\end{aligned}$$

11 Tensor products of operators

In the note above we had terms of the form $|\mathrm{vec}(\rho^{\otimes k})\rangle$. Now if you do not think about things too carefully you might be tempted to assume that $|\mathrm{vec}(\rho^{\otimes k})\rangle = |\mathrm{vec}(\rho)\rangle^{\otimes k}$.... but NO. This is wrong.

Let's look more carefully as the action of the vectorization operation on tensor products. We assume

$$A = \sum_{i,j=1}^{n} a_{ij}\ |i\rangle\langle j|\,, \quad B = \sum_{k,l=1}^{n} b_{kl}\ |k\rangle\langle l|\,.$$

Then their vectorizations are given by

$$|\mathrm{vec}(A)\rangle = \sum_{i,j=1}^{n} a_{ij}\ |ij\rangle\,, \quad |\mathrm{vec}(B)\rangle = \sum_{k,l=1}^{n} b_{kl}\ |kl\rangle$$

Thus, the tensor product of these vectorizations is

$$|\mathrm{vec}(A)\rangle \otimes |\mathrm{vec}(B)\rangle = \sum_{i,j,k,l=1}^{n} a_{ij}\, b_{kl}\ |ij\rangle \otimes |kl\rangle = \sum_{i,j,k,l=1}^{n} a_{ij}\, b_{kl}\ |ijkl\rangle$$

On the other hand, the Kronecker product is defined as

$$A \otimes B = \sum_{i,j,k,l=1}^{n} a_{ij}\, b_{kl} \Big(|i\rangle\langle j|\Big) \otimes \Big(|k\rangle\langle l|\Big) = \sum_{i,j,k,l=1}^{n} a_{ij}\, b_{kl} \Big(|ik\rangle\langle jl|\Big),$$

and its vectorization yields

$$|\mathrm{vec}(A\otimes B)\rangle = \sum_{i,j,k,l=1}^{n} a_{ij}\, b_{kl}\ |ik\rangle \otimes |jl\rangle = \sum_{i,j,k,l=1}^{n} a_{ij}\, b_{kl}\ |ikjl\rangle\,.$$

Notice that the order of the basis vectors differs: the tensor product $|\mathrm{vec}(A)\rangle \otimes |\mathrm{vec}(B)\rangle$ arranges the indices as (i, j, k, l) while $|\mathrm{vec}(A\otimes B)\rangle$ arranges them as (i, k, j, l).

(Note, if you look back at the definition of the moment operator M_k in the previous section, it should now be extra clear why we had "$U_{\boldsymbol{\theta}}^{\otimes k} \otimes U_{\boldsymbol{\theta}}^{*\otimes k}$" rather than "$(U \otimes U^*)^{\otimes k}$".)

Remark: As the order of basis vectors of $|\mathrm{vec}(A\otimes B)\rangle$ and $|\mathrm{vec}(A)\rangle \otimes |\mathrm{vec}(B)\rangle$ is the only difference of their expressions, we can easily see that, for inner products, the following equality holds

$$\langle \mathrm{vec}(C\otimes D)|\mathrm{vec}(A\otimes B)\rangle = \sum_{i',j',k',l'=1}^{n} \sum_{i,j,k,l=1}^{n} c^*_{i'j'}\, d^*_{k'l'} a_{ij}\, b_{kl}\ \big\langle i'k'j'l' \big| ikjl \big\rangle \tag{4}$$

$$= \sum_{i',j',i,j=1}^{n} c^*_{i'j'} a_{ij}\ \big\langle i'j' \big| ij \big\rangle \sum_{k,l,k',l'=1}^{n} d^*_{k'l'}\, b_{kl}\ \big\langle k'l' \big| kl \big\rangle \tag{5}$$

$$= \langle \mathrm{vec}(C)|\mathrm{vec}(A)\rangle\, \langle \mathrm{vec}(D)|\mathrm{vec}(B)\rangle\ , \tag{6}$$

where C and D are also $n \times n$ matrices.

We write the inner products of the vectorizations back into traces of the corresponding matrices, and obtain $\mathrm{Tr}\big[(C\otimes D)^\dagger (A\otimes B)\big] = \mathrm{Tr}\big[C^\dagger A\big]\,\mathrm{Tr}\big[D^\dagger B\big]$, which verifies the equality for trace of tensor product of matrices.

12 Generic Map Defined by Basis Action

Consider a linear map $\mathcal{E}$ defined by its action on the basis operators:

$$\mathcal{E}\Big(|i\rangle\langle j|\Big) = \sum_{k,l} c^{ij}_{kl}\,|k\rangle\langle l|\,.$$

For a general operator

$$\rho = \sum_{i,j} \rho_{ij}\,|i\rangle\langle j|\,,$$

linearity of $\mathcal{E}$ gives:

$$\mathcal{E}(\rho) = \sum_{i,j} \rho_{ij}\,\mathcal{E}\Big(|i\rangle\langle j|\Big) = \sum_{i,j}\sum_{k,l} \rho_{ij}\,c^{ij}_{kl}\,|k\rangle\langle l|\,.$$

Taking the vectorized form yields:

$$|\mathrm{vec}(\mathcal{E}(\rho))\rangle = \sum_{i,j}\sum_{k,l} \rho_{ij}\,c^{ij}_{kl}\,|k\,l\rangle\,.$$

We now define a matrix M that encapsulates the action of $\mathcal{E}$ in the vectorized space:

$$M = \sum_{i,j,k,l} c^{ij}_{kl}\,|k\,l\rangle\langle i\,j|\,.$$

With this definition, the action of $\mathcal{E}$ in vectorized form is

$$|\mathrm{vec}(E(\rho))\rangle = M\,|\mathrm{vec}(\rho)\rangle\,,$$

and the expectation value of an observable O becomes

$$\mathrm{Tr}\Big[E(\rho)\,O\Big] = \langle\mathrm{vec}(O)|\,|\mathrm{vec}(E(\rho))\rangle = \langle\mathrm{vec}(O)|\,M\,|\mathrm{vec}(\rho)\rangle\,.$$

Example for Transpose Map Consider the linear map E defined by the transpose:

$$\mathcal{E}\Big(|i\rangle\langle j|\Big) = |j\rangle\langle i|\,.$$

For a general operator

$$\rho = \sum_{i,j} \rho_{ij}\,|i\rangle\langle j|\,,$$

its transpose is given by

$$\rho^T = \sum_{i,j} \rho_{ij}\,|j\rangle\langle i|\,.$$

Using our vectorization convention,

$$|\mathrm{vec}(\rho)\rangle = \sum_{i,j} \rho_{ij}\,|i\,j\rangle\,,$$

the vectorized form of the transposed operator is

$$\left|\text{vec}(\rho^T)\right\rangle = \sum_{i,j} \rho_{ij} \; |j\, i\rangle \, .$$

We now define a matrix M that captures the action of the transpose on the vectorized state. In bra-ket notation, this is given by

$$M = \sum_{i,j} |j\, i\rangle \, \langle i\, j| = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} = \text{SWAP}$$

in the standard $\{|00\rangle\,, |01\rangle\,, |10\rangle\,, |11\rangle\}$ basis. This permutation matrix swaps the indices as required by the transpose map,

$$\left|\text{vec}(\rho^T)\right\rangle = M \; |\text{vec}(\rho)\rangle \, .$$

13 Pauli Transfer Matrix Formalism

The choice of basis for representing operators is of course arbitrary. Above we worked in the standard computational basis: $\{|0\rangle = |00\rangle\,, |1\rangle = |01\rangle\,, |2\rangle = |10\rangle\,, |3\rangle = |11\rangle\}$. One may, however, choose to work in the Pauli basis. For a single qubit we define the basis as

$$I \equiv |0\rangle\,, \quad X \equiv |1\rangle\,, \quad Y \equiv |2\rangle\,, \quad Z \equiv |3\rangle\,.$$

A quantum operation (or channel) acting on an operator can be described by its action on the Pauli basis elements. Namely, if

$$\mathcal{E}(P_i) = \sum_j C^i_j \, P_j,$$

then the coefficients C^i_j form the *Pauli Transfer Matrix* (PTM),

$$M = \sum_{ij} C^i_j |j\rangle\langle i| \, , \tag{7}$$

and fully capture the action of the map.

Example: Hadamard Gate in the Pauli Transfer Matrix formalism. The Hadamard gate is given by

$$H = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}.$$

Its action on the Pauli matrices is well known:

$$HIH^\dagger = I, \quad HXH^\dagger = Z, \quad HYH^\dagger = -Y, \quad HZH^\dagger = X.$$

Hence, the PTM for the Hadamard gate is

$$M = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & -1 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix}.$$

We can then use the PTM to compute the action of the Hadamard gate on any operator written in the Pauli basis.

Let's now use this example to show that the PTM formalism is equivalent to the general vectorized expression for a linear map we saw in Section 12. Recall that we saw earlier that the action of a unitary channel $U(\rho)U^\dagger$ in vectorized form is given buy

$$|\text{vec}(E(\rho))\rangle = M\ |\text{vec}(\rho)\rangle\,,$$

where $M = U \otimes U^*$. In the computational basis we know that

$$H = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}$$

and thus we have that

$$M = H \otimes H^* = \frac{1}{2} \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & -1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 \end{pmatrix}.$$

This matrix describes the action of the Hadamard gate when the state is vectorized in the computational basis. Now, by choosing the Pauli basis we are simply working in a different operator basis. In particular, the set

$$\left\{ \frac{1}{\sqrt{2}} |\text{vec}(I)\rangle\,,\ \frac{1}{\sqrt{2}} |\text{vec}(X)\rangle\,,\ \frac{1}{\sqrt{2}} |\text{vec}(Y)\rangle\,,\ \frac{1}{\sqrt{2}} |\text{vec}(Z)\rangle \right\}$$

is equivalent (up to unimportant phase factors) to the standard Bell basis:

$$\begin{aligned} |\Phi^+\rangle &= \frac{1}{\sqrt{2}}\big(|00\rangle + |11\rangle\big), \\ |\Psi^+\rangle &= \frac{1}{\sqrt{2}}\big(|01\rangle + |10\rangle\big), \\ |\Phi^-\rangle &= \frac{1}{\sqrt{2}}\big(|00\rangle - |11\rangle\big), \\ |\Psi^-\rangle &= \frac{1}{\sqrt{2}}\big(|01\rangle - |10\rangle\big). \end{aligned}$$

Let S be the change-of-basis matrix that maps the computational (Bell) basis to the normalized vectorized Pauli basis:

$$\frac{1}{\sqrt{2}} |\text{vec}(P_i)\rangle = S\ |b_i\rangle\,,$$

where $\{|b_i\rangle\}$ is the Bell basis. Then the representation of the Hadamard map in the Pauli basis is given by

$$M = S\,(H \otimes H)\,S^{\dagger}.$$

One can show (by explicit calculation) that this change-of-basis yields exactly

$$M = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & -1 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix}.$$

Thus, the PTM in the Pauli basis is equivalent to the superoperator $H \otimes H$ in the computational basis, confirming that the description of the map is independent of the chosen basis. But the description is much more intuitive in the Pauli (or, equivalently, Bell) basis. Hence the advantage of working in the Pauli basis.