CSE 599d Quantum Computing Problem Set 1 Solutions

Author: Dave Bacon (*Department of Computer Science & Engineering, University of Washington*)

Due: January 20, 2006

Exercise 1: Majorization and Random Permutations

Let $x = (x_1, x_2, \dots, x_n)$ denote a vector of n real numbers, $x \in \mathbb{R}^n$. Define $x^\downarrow$ as the vector x sorted such that the components of the vector are in decreasing order, $x^\downarrow = (x_1^\downarrow, x_2^\downarrow, \dots, x_n^\downarrow)$ where $x_1^\downarrow \geq x_2^\downarrow \geq \cdots \geq x_n^\downarrow$. Thus, for example, $x_1^\downarrow$ is the largest component of x. We say that the vector x is majorized by the vector y if $\sum_{i=1}^k x_i^\downarrow \leq \sum_{i=1}^k y_i^\downarrow$ for all $k < n$ (i.e. $k = 1, 2, \dots, n-1$) and $\sum_{i=1}^n x_i^\downarrow = \sum_{i=1}^n y_i^\downarrow$. When x is majorized by y we write $x \prec y$.

(a) Suppose that $p \in \mathbb{R}^n$ is such that $p_i \geq 0$ and $\sum_{i=1}^n p_i = 1$ (we say that p is a vector of probabilities). There is a single vector of probabilities which is majorized by all other vectors of probabilities. What is this vector and prove that it is the only vector which has this property.

> The vector p which is majorized by all other vectors of probability is the vector with $p_i = \frac{1}{n}$ for all i. Let's show that this vector is indeed majorized by all other vectors of probability (note that I didn't ask for this, but it is nice to prove it.) Let x be an arbitrary probability vector which does not majorize p. We will prove by contradiction that such a vector cannot exist. For some k it must be that $\sum_{i=1}^k x_i^\downarrow < \sum_{i=1}^k \frac{1}{n}$ which implies that at least one of the x_l, $1 \leq l \leq k$ is less than $\frac{1}{n}$. But since the probability vectors sum to 1, we also have that $\sum_{i=k+1}^n x_i^\downarrow > \sum_{i=k+1}^n \frac{1}{n}$ which implies that one of these $x_l^\downarrow$, $k+1 \leq l \leq 1$ is less than $\frac{1}{n}$. But this contradicts the fact that the $x_i^\downarrow$ are sorted in decreasing order.
>
> Why is the uniform vector p the only probability vector which has the property of being majorized by all probability vectors (this is what I asked you to show)? Suppose that there was another such probability vector x which is not uniform but which is majorized by all probability vectors. Then we will show that it is not majorized by the uniform vector p. A probability vector which is not uniform must have at least one element $x_i > \frac{1}{n}$. But this implies that at least the first element of x must satisfy $x_1^\downarrow > \frac{1}{n} = p_1$. But this implies that there is a vector, namely the uniform vector, which x is not majorize by. This is a contradiction.

(b) An $n \times n$ matrix $A = (a_{ij})$ is called doubly stochastic if $a_{ij} \geq 0$ for all i and j, $\sum_{i=1}^n a_{ij} = 1$ for all j and $\sum_{j=1}^n a_{ij} = 1$ for all i. Show that every convex combination of a doubly stochastic matrices is a doubly stochastic matrix (recall that a convex combination of matrices $A_1, A_2, \dots, A_m$ is a sum of these matrices, $\sum_{j=1}^m q_j A_j$ with $q_j \geq 0$ and $\sum_{j=1}^m q_j = 1$.)

> Denote the matrix elements of A_k by $(a_k)_{ij}$. A convex combination of these elements is a matrix C with entries $c_{ij} = \sum_{k=1}^m q_k (a_k)_{ij}$. Since $q_k \geq 0$ (from definition of convex) and $(a_k)_{ij} \geq 0$ from definition of doubly stochastic, this implies that $c_{ij} \geq 0$. Next we take the column sums $\sum_{i=1}^n c_{ij} = \sum_{i=1}^n \sum_{k=1}^m q_k (a_k)_{ij} = \sum_{k=1}^m q_k \sum_{i=1}^n (a_k)_{ij} = \sum_{k=1}^m q_k = 1$, where we have used the fact that each A_k is doubly stochastic in the second to last equality and the convexity in the last equality. This holds for all j. A similar sum holds for each row, $\sum_{j=1}^n c_{ij} = \sum_{j=1}^n \sum_{k=1}^m q_k (a_k)_{ij} = \sum_{k=1}^m q_k \sum_{j=1}^n (a_k)_{ij} = \sum_{k=1}^m q_k = 1$ for all i. Thus we have shown that c_{ij} satisfies all of the properties of being doubly stochastic.

(c) Prove that if $Ax \prec x$ for all x then A must be doubly stochastic (hint consider the vector from part (a) as well as vectors like $(0, 0, 1, 0, \dots, 0)$.)

> If $Ax \prec x$ for all x then it must also do so for arbitrary vectors x and in particular for all probability vectors x. Consider the case where x has components $x_i = \delta_{ik}$, i.e. only one component, the kth is nonzero. Now the components of the vector Ax is $(Ax)_i = \sum_{j=1}^n A_{ij} x_i = A_{ik}$. The requirement that the sum over all the elements must be the same is then $\sum_{i=1}^n (Ax)_i = \sum_{i=1}^n A_{ik} = \sum_{i=1}^n x = 1$. This must hold for all k.
>
> $Ax \prec x$ must also hold for the uniform probability vector x with components $x_i = \frac{1}{n}$. But from above we know that the only probability vector which the uniform vector can majorize is itself. And in this case all our inequalities become strick equalities. $(Ax)_i = \sum_{j=1}^n A_{ij} x_j = \sum_{j=1}^n A_{ij} \frac{1}{n} = x_i = \frac{1}{n}$. Thus $\sum_{j=1}^n A_{ij} = 1$ for all i.

Finally we need to show that each $A_{ij} \geq 0$. To show this consider again the vector x with components $x_i = \delta_{ik}$. Then $(Ax)_i = \sum_{j=1}^n A_{ij}x_j = A_{ik}$. Now if we sort $(Ax)_i$ and x_i we will obtain the largest value of $(Ax)_i$ first. The first inequality which must be satisfied is therefore $(Ax)_1^\downarrow \leq 1$. The next inequality will be $(A_x)_2^\downarrow + (Ax)_1^\downarrow \leq 1$. The last inequality will be $\sum_{i=1}^{n-1}(Ax)_i^\downarrow \leq 1$. But the equality will be $\sum_{i=1}^n (Ax)_i^\downarrow = 1$. This implies that $\sum_{i=1}^{n-1}(Ax)_i^\downarrow = 1 - (Ax)_n^\downarrow$, so the last inequality becomes $1 - (Ax)_n^\downarrow \leq 1$ or $(Ax)_n \geq 0$ which is just $A_{nk} \geq 0$. Similarly we can turn the second to last inequality $\sum_{i=1}^{n-2}(Ax)_i^\downarrow \leq 1$ into $1 - (Ax)_{n-1} - (Ax)_n \geq 1$. Using these inequalities we obtain $A_{n-1,k} \geq 0$. It is easy to see that we can use induction to prove that $A_{i,k} \geq 0$ (the inequality $\sum_{i=1}^{n-k}(Ax)_i^\downarrow \leq 1$ can via the inductive hypothesis be turned into $1 - \sum_{i=n-k}^n (Ax)_i^\downarrow \leq 1$ and then using the previous inequalities we obtain the desired result.)

Thus we have have shown if $Ax \prec x$, then A must satisfy all of the properties of a doubly stochastic matrix.

(d) Prove that if A is doubly stochastic then $Ax \prec x$ for all vectors x.

For any vector x we can redefine A such that the columns of A are reordered so that x is a decreasing vector and the rows of A are reordered so that Ax is a decreasing vector. Without loss of generality we will use this A.

We will prove the kth inequality. Define $g_j = \sum_{i=1}^k A_{ij}$. Then because A is doubly stochastic, $0 \leq g_j \leq 1$. Notice also that $\sum_{j=1}^n g_j = k$. Define $y_i = \sum_{j=1}^n A_{ij}x_j$. We consider will $\sum_{i=1}^k (y_i - x_i)$. By definitions this is equal to

$$\sum_{i=1}^k (y_i - x_i) = \sum_{i=1}^k \sum_{j=1}^n A_{ij}x_j - \sum_{i=1}^k x_i \tag{1}$$

Next we add a term which is zero

$$\sum_{i=1}^k (y_i - x_i) = \sum_{i=1}^k \sum_{j=1}^n A_{ij}x_j - \sum_{i=1}^k x_i + (k - \sum_{j=1}^n g_j)x_k \tag{2}$$

Next rearrange the first sum

$$\sum_{i=1}^k (y_i - x_i) = \sum_{j=1}^n \sum_{i=1}^k A_{ij}x_j - \sum_{i=1}^k x_i + (k - \sum_{j=1}^n g_j)x_k \tag{3}$$

which we can rearrange as

$$\sum_{i=1}^k (y_i - x_i) = \sum_{j=1}^n g_j x_j - \sum_{i=1}^k x_i + (k - \sum_{j=1}^n g_j)x_k \tag{4}$$

and split up the sums

$$\sum_{i=1}^k (y_i - x_i) = \sum_{j=1}^k g_j x_j + \sum_{j=k+1}^n g_j x_j - \sum_{i=1}^k x_i + kx_k - \sum_{j=1}^k g_j x_k - \sum_{j=k+1}^n g_j x_k \tag{5}$$

Using $\sum_{i=1}^k 1 = k$ we can write this as

$$\sum_{i=1}^k (y_i - x_i) = \sum_{j=1}^k g_j x_j + \sum_{j=k+1}^n g_j x_j - \sum_{i=1}^k x_i + \sum_{i=1}^k x_k - \sum_{j=1}^k g_j x_k - \sum_{j=k+1}^n g_j x_k \tag{6}$$

which we then reexpress as

$$\sum_{i=1}^k (y_i - x_i) = \sum_{i=1}^k (g_i - 1)(x_i - x_k) + \sum_{j=k+1}^n g_j(x_j - x_k) \tag{7}$$

But notice that $x_i - x_k \geq 0$ for $i \leq k$ and $x_i - x_k \leq 0$ for $i \geq k$. Further since $0 \leq g_i \leq 1$, $g_i - 1 \leq 0$ and $g_j \geq 0$. Thus both of these terms are negative. Hence we have shown that

$$\sum_{i=1}^{k}(y_i - x_i) \leq 0 \tag{8}$$

Which is the required inequalities for $1 \leq k \leq n-1$. What about the equality? It is easy to show true by the properties of the doubly stochastic matrix: $\sum_{i=1}^{n}(Ax)_i = \sum_{i,j=1}^{n} A_{ij}x_j = \sum_{j=1}^{n} x_j$.

(e) Suppose that we have a machine with N configurations. One operation we can perform on such a system is to permute (map in a one-to-one manner) these configurations. Suppose that at any given time we apply one of $N!$ different permutations to the system with a fixed probability for each possible permutation. If p is a vector of probabilities describing our machine the evolution described by these random permutations is given by $q = Ap$ where q is the new description of our machine. Show that for process of applying random permutations, the matrix A is doubly stochastic. Can the vector of probabilities for a four state machine $[\frac{1}{12}\ \frac{1}{2}\ \frac{1}{12}\ \frac{1}{3}]^T$ ever evolve under one of these random permutations into the vector of probabilities $[\frac{1}{2}\ \frac{1}{6}\ \frac{1}{6}\ \frac{1}{6}]^T$? Why or why not?

A permutation is represented by a matrix A whose entries are all either 0 or 1 and each row and each column has exactly one 1 (because every configuration must be taken to only one other configuration and also this mapping is one to one.) Such matrices are doubly stochastic matrices. We represent a random permutation as a convex combination of these permutation matrices. This matrix, by part (b) must be doubly stochastic. From parts (c) and (d) we know that $Ax \prec x$ for all x iff A is doubly stochastic. Thus to see whether the above four state evolution is possible, we need to check whether $[\frac{1}{2}\ \frac{1}{6}\ \frac{1}{6}\ \frac{1}{6}]^T \prec [\frac{1}{12}\ \frac{1}{2}\ \frac{1}{12}\ \frac{1}{3}]^T$. Sorting in decreasing order the elements in these matrices we obtain $[\frac{1}{2}\ \frac{1}{6}\ \frac{1}{6}\ \frac{1}{6}]^T$ and $[\frac{1}{2}\ \frac{1}{3}\ \frac{1}{12}\ \frac{1}{12}]^T$. We obtain the following inequalities $\frac{1}{2} \leq \frac{1}{2}$, $\frac{1}{2} + \frac{1}{6} \leq \frac{1}{2} + \frac{1}{3}$, $\frac{1}{2} + \frac{1}{6} + \frac{1}{6} \leq \frac{1}{2} + \frac{1}{3} + \frac{1}{12}$, and the quality that both vectors sum to 1. Thus yes indeed such this evolution can occur. In fact it is easy to see that the following doubly stochastic matrix achieves this evolution

$$A = \begin{bmatrix} 0 & 1 & 0 & 0 \\ \frac{1}{3} & 0 & \frac{1}{3} & \frac{1}{3} \\ \frac{1}{3} & 0 & \frac{1}{3} & \frac{1}{3} \\ \frac{1}{3} & 0 & \frac{1}{3} & \frac{1}{3} \end{bmatrix} \tag{9}$$

The notion of majorization occurs naturally in various contexts. For example, in economics if $x_1, \ldots, x_n$ and $y_1, \ldots, y_n$ denote incomes of individuals $1, \ldots, n$, then $x \prec y$ would mean that there is a more equal distribution of incomes in the state x than in y. We will encounter majorization later when we talk about transformations on entangled quantum states.

Exercise 2: Paulis, Cliffords, and Toffolis

Recall that the single qubit Pauli operators are given by

$$I = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \quad X = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \quad Y = \begin{bmatrix} 0 & -i \\ i & 0 \end{bmatrix} \quad Z = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$$

Elements of the *Pauli group* are made up of tensor products of n of the Pauli matrices, along with a phase i^k, where $k \in \{0,1,2,3\}$. Elements of the Pauli group can be parameterized by two n bit strings, a and b along with $k = 0,1,2,3$ as

$$P(a,b,k) = i^k (X^{a_1} Z^{b_1}) \otimes (X^{a_2} Z^{b_2}) \otimes \cdots \otimes (X^{a_n} Z^{b_n})$$

(a) Show that all elements of the Pauli group on n qubits, $P(a,b,k)$, are unitary.

First we need to calculate $P(a,b,k)^\dagger$. Note that the adjoint of the tensor product of two matrices is the tensor product of the adjoint of these two matrices: $(A \otimes B)^\dagger = A^\dagger \otimes B^\dagger$. Further adjoint of a complex number times a matrix is the conjugate of that number times the adjoint of the matrix: $(cA)^\dagger = c^* A^\dagger$. Thus

$$P(a,b,k)^\dagger = (i^k (X^{a_1} Z^{b_1}) \otimes (X^{a_2} Z^{b_2}) \otimes \cdots \otimes (X^{a_n} Z^{b_n}))^\dagger = (i^k)^* (X^{a_1} Z^{b_1})^\dagger \otimes (X^{a_2} Z^{b_2})^\dagger \otimes \cdots \otimes (X^{a_n} Z^{b_n})^\dagger \tag{10}$$

Now use the fact that when you multiply tensor products they multiply tensor wise and a scalar multiplies separately $(cA \otimes B)(dC \otimes D) = cd((AC) \otimes (BC))$. This implies

$$\begin{aligned} P(a,b,k)P(a,b,k)^\dagger &= (i^k(X^{a_1}Z^{b_1}) \otimes (X^{a_2}Z^{b_2}) \otimes \cdots \otimes (X^{a_n}Z^{b_n})) \\ &\quad ((i^k)^*(X^{a_1}Z^{b_1})^\dagger \otimes (X^{a_2}Z^{b_2})^\dagger \otimes \cdots \otimes (X^{a_n}Z^{b_n})^\dagger) \\ &= i^k(i^k)^*(X^{a_1}Z^{b_1}(X^{a_1}Z^{b_1})^\dagger) \otimes (X^{a_2}Z^{b_2}(X^{a_2}Z^{b_2})^\dagger) \otimes \cdots \otimes (X^{a_n}Z^{b_n}(X^{a_n}Z^{b_n})^\dagger) \\ &= (X^{a_1}Z^{b_1}(X^{a_1}Z^{b_1})^\dagger) \otimes (X^{a_2}Z^{b_2}(X^{a_2}Z^{b_2})^\dagger) \otimes \cdots \otimes (X^{a_n}Z^{b_n}(X^{a_n}Z^{b_n})^\dagger) \qquad (11) \end{aligned}$$

Finally using $(AB)^\dagger = B^\dagger A^\dagger$, we can calculate hat $X^{a_i}Z^{b_i}(X^{a_i}Z^{b_i})^\dagger = X^{a_i}Z^{b_i}(Z^{b_i})^\dagger(X^{a_i})^\dagger$. But Z and X are hermitian, so this is just $X^{a_i}Z^{b_i}Z^{b_i}X^{b_i}$. Further $I^2 = I$, $Z^2 = I$, and $X^2 = I$, so this in turn is just $X^{a_i}Z^{b_i}(X^{a_i}Z^{b_i})^\dagger = I$. Thus we see that $P(a,b,k)P(a,b,k)^\dagger = I \otimes I \otimes \cdots \otimes I$ and thus $P(a,b,k)$ is unitary.

(b) Show that $P(a,b,k)P(c,d,l) = (-1)^m P(c,d,k)P(a,b,l)$ where $m = \sum_{i=1}^n a_i d_i + \sum_{i=1}^n b_i c_i \bmod 2$. Thus you will have shown that any two elements of the Pauli group either commute $m = 0$ or anti-commute $m = 1$.

First note that it is easy to calculate the following identity for single qubit Pauli matrices: $ZX = -XZ$, or $Z^uX^v = (-1)^{uv \bmod 2}X^vZ^u$ for $u, v \in \{0,1\}$. Consider $X^{a_i}Z^{b_i}X^{c_i}Z^{d_i}$. Then using our identity on the inner product we see that this is equal to $X^{a_i}Z^{b_i}X^{c_i}Z^{d_i} = (-1)^{b_ic_i \bmod 2}X^{a_i}X^{c_i}Z^{b_i}Z^{d_i}$. Since matrices commute with themselves and identity, this is equal to $X^{a_i}Z^{b_i}X^{c_i}Z^{d_i} = (-1)^{b_ic_i \bmod 2}X^{c_i}X^{a_i}Z^{d_i}Z^{b_i}$. Again using our identity we obtain $X^{a_i}Z^{b_i}X^{c_i}Z^{d_i} = (-1)^{b_ic_i \bmod 2}(-1)^{a_id_i \bmod 2}X^{c_i}Z^{d_i}X^{a_i}Z^{b_i}$ which we can express as $X^{a_i}Z^{b_i}X^{c_i}Z^{d_i} = (-1)^{b_ic_i + a_id_i \bmod 2}X^{c_i}Z^{d_i}X^{a_i}Z^{b_i}$. Now we need to apply this to $P(a,b,k)P(c,d,l)$.

$$\begin{aligned} P(a,b,k)P(c,d,l) &= \left(i^k(X^{a_1}Z^{b_1}) \otimes \cdots \otimes (X^{a_n}Z^{b_n})\right)\left(i^l(X^{c_1}Z^{d_1}) \otimes \cdots \otimes (X^{c_n}Z^{d_n})\right) \\ &= i^{k+l}(X^{a_1}Z^{b_1}X^{c_1}Z^{d_1}) \otimes \cdots \otimes (X^{a_n}Z^{b_n}X^{c_n}Z^{d_n}) \\ &= i^{k+l}((-1)^{b_1c_1 + a_1d_1 \bmod 2}X^{c_1}Z^{d_1}X^{a_1}Z^{b_1}) \otimes \cdots \otimes ((-1)^{b_nc_n + a_nd_n \bmod 2}X^{c_n}Z^{d_n}X^{a_n}Z^{b_n}) \\ &= i^{k+l}(-1)^{\sum_{i=1}^n b_ic_i + a_id_i \bmod 2}(X^{c_1}Z^{d_1}X^{a_1}Z^{b_1}) \otimes \cdots \otimes (X^{c_n}Z^{d_n}X^{a_n}Z^{b_n}) \\ &= (-1)^{\sum_{i=1}^n b_ic_i + a_id_i \bmod 2}(i^l(X^{c_1}Z^{d_1}) \otimes \cdots \otimes (X^{c_n}Z^{d_n}))(i^k(X^{a_1}Z^{b_1}) \otimes \cdots \otimes (X^{a_n}Z^{b_n})) \\ &= (-1)^{\sum_{i=1}^n b_ic_i + a_id_i \bmod 2}P(c,d,l)P(a,b,k) \qquad (12) \end{aligned}$$

Actually this is what I wanted you to show, but I wrote it incorrectly: there is a phase i^k and i^l which commute between both expressions. Thus it is easy to see that this is also equal to $P(a,b,k)P(c,d,l) = (-1)^{\sum_{i=1}^n b_ic_i + a_id_i \bmod 2}P(c,d,k)P(a,b,l)$

(c) Consider operators on n qubits of the form $R(P(a,b,k)) = \frac{1}{\sqrt{2}}(I + iP(a,b,k))$ where $P(a,b,k)$ is an element of the Pauli group. Show that if $P(a,b,k)$ is hermitian, then $R(P(a,b,k))$ is unitary. For the later parts of this problem, assume that $R(P(a,b,k))$ is indeed one of these unitary gates.

$R(P(a,b,k))^\dagger = \frac{1}{\sqrt{2}}(I + iP(a,b,k))^\dagger = \frac{1}{\sqrt{2}}(I^\dagger + i^*P(a,b,k)^\dagger)$. If $P(a,b,k)$ is hermitian, then this is equal to $R(P(a,b,ik))^\dagger = \frac{1}{\sqrt{2}}(I - iP(a,b,k))$. Next check it if is unitary: $R(P(a,b,k))R(P(a,b,k)^\dagger = \frac{1}{\sqrt{2}}(I + iP(a,b,k))\frac{1}{\sqrt{2}}(I - iP(a,b,k)) = \frac{1}{2}(I + iP(a,b,k) - iP(a,b,k) + i(-i)P(a,b,k)P(a,b,k)) = \frac{1}{2}(I + P(a,b,k)^2)$. But since $P(a,b,k)$ is hermitian, and we have shown that the $P(a,b,k)$ are unitary: $P(a,b,k)P(a,b,k)^\dagger = I$ and so $P(a,b,k)^2 = P(a,b,k)P(a,b,k)^\dagger = I$. Thus we see that $R(P(a,b,k))R(P(a,b,k)^\dagger = \frac{1}{2}(I + I) = I$, so $R(P(a,b,k))$ is indeed unitary.

(d) Show that $R(P(a,b,k))P(c,d,l)R(P(a,b,k))^\dagger = P(c,d,l)$ if $P(a,b,k)$ commutes with $P(c,d,l)$.

If $P(a,b,k)$ commutes with $P(c,d,l)$, then $R(P(a,b,k))$ commutes with $P(c,d,l)$ (because if a matrix commutes two matrices then it commutes with their sum, and $P(c,d,l)$ commutes with both I and $P(a,b,k)$.) Thus $R(P(a,b,k))P(c,d,l)R(P(a,b,k))^\dagger = P(c,d,l)R(P(a,b,k))R(P(a,b,k))^\dagger = P(c,d,l)I = P(c,d,l)$ where we have used the unitarity of $R(P(a,b,k))$.

(e) Show that $R(P(a,b,k))P(c,d,l)R(P(a,b,k))^\dagger = iP(a,b,k)P(c,d,l)$ if $P(a,b,k)$ anti-commutes with $P(c,d,l)$.

$R(P(a,b,k))P(c,d,l)R(P(a,b,k))^\dagger = \frac{1}{\sqrt{2}}\left(I + iP(a,b,k)\right)P(c,d,l)\frac{1}{\sqrt{2}}\left(I - iP(a,b,k)\right)$ We can expand this as

$$\begin{aligned} R(P(a,b,k))P(c,d,l)R(P(a,b,k))^\dagger = \frac{1}{2} \quad & (P(c,d,l) + P(a,b,k)P(c,d,l)P(a,b,k) \\ & + iP(a,b,k)P(c,d,l) - iP(c,d,l)P(a,b,k)) \qquad (13) \end{aligned}$$

. But since $P(a,b,k)P(c,d,l) = -P(c,d,l)P(a,b,k)$ and $P(a,b,k)^2 = I$ (see solution for part (c)), $P(a,b,k)P(c,d,l)P(a,b,k) = -P(c,d,l)$, so

$$\begin{aligned} R(P(a,b,k))P(c,d,l)R(P(a,b,k))^\dagger &= \frac{1}{2}\left(iP(a,b,k)P(c,d,l) - iP(c,d,l)P(a,b,k)\right) \\ &= \frac{1}{2}\left(iP(a,b,k)P(c,d,l) + iP(a,b,k)P(c,d,l)\right) = iP(a,b,k)P(c,d,l) \end{aligned} \tag{14}$$

where we have use the anti-commutivity of $P(a,b,k)$ and $P(c,d,l)$ once more.

(f) The Toffoli gate on 3 qubits is a controlled-controlled-NOT gate. In the computational basis, it acts as

$$CC_X = \begin{bmatrix} 1&0&0&0&0&0&0&0\\ 0&1&0&0&0&0&0&0\\ 0&0&1&0&0&0&0&0\\ 0&0&0&1&0&0&0&0\\ 0&0&0&0&1&0&0&0\\ 0&0&0&0&0&1&0&0\\ 0&0&0&0&0&0&0&1\\ 0&0&0&0&0&0&1&0 \end{bmatrix}$$

Show that no sequence of unitary operations $R(P(a,b,k))$ on 3 qubits can be used to reproduce the Toffoli gate (hint consider the effect of conjugating a Pauli operator by the Tofolli gate: $(CC_X)P(a,b,k)(CC_X)^\dagger$)

Consider $CC_X(X \otimes I \otimes I)CC_X^\dagger$. This is the monster matrix multiplication

$$\begin{bmatrix} 1&0&0&0&0&0&0&0\\ 0&1&0&0&0&0&0&0\\ 0&0&1&0&0&0&0&0\\ 0&0&0&1&0&0&0&0\\ 0&0&0&0&1&0&0&0\\ 0&0&0&0&0&1&0&0\\ 0&0&0&0&0&0&0&1\\ 0&0&0&0&0&0&1&0 \end{bmatrix} \begin{bmatrix} 0&0&0&0&1&0&0&0\\ 0&0&0&0&0&1&0&0\\ 0&0&0&0&0&0&1&0\\ 0&0&0&0&0&0&0&1\\ 1&0&0&0&0&0&0&0\\ 0&1&0&0&0&0&0&0\\ 0&0&1&0&0&0&0&0\\ 0&0&0&1&0&0&0&0 \end{bmatrix} \begin{bmatrix} 1&0&0&0&0&0&0&0\\ 0&1&0&0&0&0&0&0\\ 0&0&1&0&0&0&0&0\\ 0&0&0&1&0&0&0&0\\ 0&0&0&0&1&0&0&0\\ 0&0&0&0&0&1&0&0\\ 0&0&0&0&0&0&0&1\\ 0&0&0&0&0&0&1&0 \end{bmatrix} \tag{15}$$

which we can work out to be

$$\begin{bmatrix} 0&0&0&0&1&0&0&0\\ 0&0&0&0&0&1&0&0\\ 0&0&0&0&0&0&0&1\\ 0&0&0&0&0&0&1&0\\ 1&0&0&0&0&0&0&0\\ 0&1&0&0&0&0&0&0\\ 0&0&0&1&0&0&0&0\\ 0&0&1&0&0&0&0&0 \end{bmatrix} \tag{16}$$

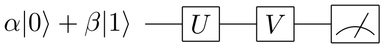
This is the gate $X \otimes C_X$ which acts as X on the first qubit and C_X on the second and third qubits. But C_X is not a member of the Pauli group. To see this note that the trace (sum of diagonal matrix elements) of all elements of Pauli group are zero or 2^n, but C_X has trace 2 and does not act on one qubit. Further $X \otimes C_X$ is also not a member of the Pauli group, since the Pauli group requires that all qubits should be acted upon by Pauli matrices. Thus we have seen that $CC_X(X \otimes I \otimes I)CC_X^\dagger$ produces an element not in the Pauli group. But by the previous two parts, we have seen that the $R(P(a,b,k))$ elements, when conjugated about a Pauli group member always produce a Pauli group member. So any sequence of such $R(P(a,b,k))$ elements must also, when conjugated about a Pauli group member always produce a Pauli group member. But the Toffoli gate doesn't do this. This is a contradiction with the idea that we can express the Toffoli as a product of $R(P(a,b,k))$ gates. Thus we cannot construct Toffoli from such a product of $R(P(a,b,k))$ gates.

What you've demonstrated in the last part of this problem is that there is no way to use the R gates alone to perform a Tofolli gate. In fact, we will learn when we get to quantum error correction that the R gates are what are called Clifford gates and that they do not form a universal set of quantum gates.

Exercise 3: Distinguishing Paulis

Suppose we are given access to a black box which implements one of the four Pauli operators I, X, Y, Z (given explicitly in Exercise 2.) We don't know which of these four Pauli operators the box implements and confoundingly the black box explodes after we use it, so that we only get to use it one time! We will call the black box U (i.e. U is from the set $\{I, X, Y, Z\}$.)

(a) Suppose that we are only allowed to use a single qubit pure state input into the black box, but that we can choose this single qubit state arbitrarily. After the black box U has acted, then we are allowed to make any single qubit unitary we wish and then measure in the computational ($|0\rangle$, $|1\rangle$) basis. In other words, we attempt to distinguish what U is by a circuit of the form



where V is an arbitrary unitary. Prove that it is impossible to choose a V and input $\alpha|0\rangle + \beta|1\rangle$ such that it is always possible to distinguish with perfect certainty which Pauli, I, X, Y, or Z, the black box U implements. This isn't as hard as it sounds.

There are four different Pauli's but our measurement outcome has only two outcomes! For each measurement outcome we need to guess one of the four Pauli's (if we want to succeed with certainty.) But there is no way to do this if we have only two measurement outcomes.

(b) Show that $(P \otimes I)\frac{1}{\sqrt{2}}(|00\rangle + |11\rangle)$, where P is one of the four single qubit Pauli matrices I, X, Y, or Z are orthogonal for the four different Ps.

Define $|\psi_P\rangle = (P \otimes I)\frac{1}{\sqrt{2}}(|00\rangle + |11\rangle)$. Then

$$\langle\psi_Q|\psi_P\rangle = \frac{1}{\sqrt{2}}(\langle 00| + \langle 11|)(Q^\dagger \otimes I)(P \otimes I)\frac{1}{\sqrt{2}}(|00\rangle + |11\rangle) = \frac{1}{2}(\langle 00| + \langle 11|)((QP) \otimes I)(|00\rangle + |11\rangle) \quad (17)$$

But this is equal to

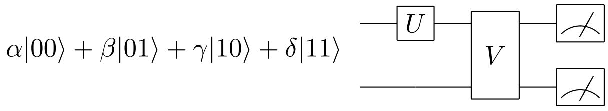
$$\langle\psi_Q|\psi_P\rangle = \frac{1}{2}(\langle 00|((QP) \otimes I)|00\rangle + \langle 00|((QP) \otimes I)|11\rangle + \langle 11|((QP) \otimes I)|00\rangle + \langle 11|((QP) \otimes I)|11\rangle) \quad (18)$$

But using the fact that $\langle 00|M \otimes I|11\rangle = \langle 0|M|1\rangle\langle 0|I|1\rangle = 0$ (and similar expressions for the other terms) we see that this is equal to

$$\langle\psi_Q|\psi_P\rangle = \frac{1}{2}(\langle 0|QP|0\rangle + \langle 1|QP|1\rangle) = \frac{1}{2}\mathrm{Tr}[QP] \quad (19)$$

where Tr is the trace (sum of diagonal elements of matrix.) Now it is easy to calculate that $\mathrm{Tr}[X] = \mathrm{Tr}[Y] = \mathrm{Tr}[Z] = 0$. Further since multiplying different elements from $\{I, X, Y, Z\}$ produces a matrix proportional to an element in $\{X, Y, Z\}$ this implies that $\langle\psi_Q|\psi_P\rangle = \frac{1}{2}\mathrm{Tr}[QP] = 0$ if $Q \neq P$.

(c) Now suppose that instead of the restricted circuit used in part (a), you are now allowed to input two-qubit states into the black box, then perform a two qubit gate after the evolution of the black box, and then perform a measurement in the computational basis. In other words the general circuit considered is now


$\alpha|00\rangle + \beta|01\rangle + \gamma|10\rangle + \delta|11\rangle$
U
V

where V is now a two-qubit unitary. Show that it is now possible to distinguish all for single qubit Paulis from each other with certainty by choosing the appropriate two qubit input state and two qubit unitary V.

Suppose we choose the two qubit input state to be $\frac{1}{\sqrt{2}}(|00\rangle + |11\rangle)$. Then the state after the Pauli will be one of the $|\Psi_Q\rangle$ states defined above, where $Q = U$. Since these states are orthogonal we should be able to choose a unitary V which rotates this basis into the computational basis. A measurement will then distinguish these Paulis with certainty. One way to choose this matrix is

$$V = \begin{bmatrix} \frac{1}{\sqrt{2}} & 0 & 0 & \frac{1}{\sqrt{2}} \\ 0 & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 \\ 0 & \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} & 0 \\ \frac{1}{\sqrt{2}} & 0 & 0 & -\frac{1}{\sqrt{2}} \end{bmatrix} \quad (20)$$

One can check that V is indeed unitary. Then if we apply V a measure, if $U = I$ we will obtain $|00\rangle$, if $U = X$ we will obtain $|01\rangle$, if $U = Y$ we obtain $|10\rangle$, and if $U = Z$ we obtain $|11\rangle$.

What you've just demonstrated is that it is possible to use entangled quantum states to help distinguish between different unknown unitary gates. This idea, generalized, is one way to think about how quantum algorithms work.