Tomi Johnson[1]

Keble College - Michaelmas 2014

CP3&4: Mathematical methods I&II

Tutorial 5 - Vectors and matrices II

Prepare full solutions to the 'problems' with a self assessment of your progress on a cover page.
Leave these at Keble lodge by 5pm on Monday of 4th week.
Look at the 'class problems' in preparation for the tutorial session.
Suggested reading: RHB 8 and the lecturer's problem sets.

Goals

- Learn the basics types of matrices and appreciate their relationship to operators.
- Practice matrix calculations like multiplication, inversion, trace and determinant.
- Practice solving matrix problems, like simultaneous linear equations and eigenvalue problems, and appreciate their interpretation.

Problems

As with the problem set on vectors, here I will try and display to you some of the the formal mathematics underpinning matrices. It might seem overwhelming at first, and this is expected — don't panic. Detailed solutions are provided for the more abstract problems, so they shouldn't prevent you answering the more standard problems, on which you should focus if having difficulty.

In the previous problem set we introduced the concept of a vector $|v\rangle$ *and the representation of vectors in a basis by a column vector* $\mathbf{v}$. *In this problem set we will discuss linear operators* $\hat{O}$ *(often the hat is not written) that take a vector to another vector, i.e.* $|v\rangle = \hat{O}|u\rangle$ *(note, we follow standard convention and always consider things to act from right to left). As will be shown, these operators can be represented in some bases by a matrix* $\mathbf{O}$ *(as with column vectors, usually in longhand an underlined* $\underline{O}$ *is used instead of a bold symbol).*

A column vector $\mathbf{v}$ *stores the components of a vector* $|v\rangle$ *in a basis. Similarly, a matrix* $\mathbf{O}$ *stores how the components of vector* $|u\rangle$ *in one basis transform to the components of* $|v\rangle = \hat{O}|u\rangle$ *in another basis. To continue we therefore need to define these two bases* $\mathcal{B}_u = \{|u_i\rangle\}$ *and* $\mathcal{B}_v = \{|v_i\rangle\}$ *that span the input and output vector spaces, of dimensions denoted* N *and* M*, respectively. The action of the operator on an input basis vector must be some linear combination of the output basis vectors* $\hat{O}|u_j\rangle = \sum_i O_{ij}|v_i\rangle$*, where we have written the components of the output as* O_{ij}*. These* MN *components are usually written out as an* $M \times N$ *matrix*

$$\mathbf{O} = \begin{pmatrix} O_{11} & O_{12} & \cdots & O_{1N} \\ O_{21} & O_{22} & \cdots & O_{2N} \\ \vdots & & \ddots & \vdots \\ O_{M1} & O_{M2} & \cdots & O_{MN} \end{pmatrix}.$$

It is clear that a linear superposition $\hat{O} = z_1\hat{O}^{[1]} + z_2\hat{O}^{[2]}$ *of operators would be represented by* $O_{ij} = z_1 O^{[1]}_{ij} + z_2 O^{[2]}_{ij}$*, which defines matrix addition* $\mathbf{O} = z_1\mathbf{O}^{[1]} + z_2\mathbf{O}^{[2]}$.

---

[1]These problems were compiled by Prof. D. Jaksch based on problem sets by Prof. N. Harnew and past Oxford Prelims exam questions.

The components, now elements of matrix $\mathbf{O}$*, tell us everything we need to know about the action of the operator* $\hat{O}$ *in the bases being used. For example, let's consider the action of the operator on an arbitrary vector* $|u\rangle = \sum_i b_i |u_i\rangle$*, represented in basis* $\mathcal{B}_u$ *by column vector* $\mathbf{u} = (b_1, \ldots, b_N)^T$*, to give the vector* $|v\rangle = \sum_i c_i |v_i\rangle$*, represented in basis* $\mathcal{B}_v$ *by column vector* $\mathbf{v} = (c_1, \ldots, c_M)^T$*. Using definition* $\hat{O}|u_j\rangle = \sum_i O_{ij} |v_i\rangle$ *and linearity we can expand* $|v\rangle = \hat{O}|u\rangle$ *to arrive at* $\sum_i c_i |v_i\rangle = \sum_i \sum_j O_{ij} b_j |v_i\rangle$*. As the basis vectors in* $\mathcal{B}_v$ *are assumed to be linearly independent this means* $c_i = \sum_j O_{ij} b_j$*, and thus the components representing the output vector can be found from those representing the input vector and the operator. This defines the concept of multiplication* $\mathbf{v} = \mathbf{O}\mathbf{u}$ *of a column vector* $\mathbf{u}$ *by a matrix* $\mathbf{O}$*.*

We can continue in this fashion, considering a further operation on $|v\rangle$ *by* $\hat{P}$ *to give* $|w\rangle = \hat{P}|v\rangle = \hat{P}\hat{O}|u\rangle = \hat{Q}|u\rangle$ *(note again the right to left ordering) and describe the vector spaces of* $|u\rangle$*,* $|v\rangle$ *and* $|w\rangle$ *using bases* $\mathcal{B}_u$*,* $\mathcal{B}_v$ *and* $\mathcal{B}_w$*. We define* $\mathbf{P}$ *by components* P_{ij} *appearing in* $\hat{P}|v_j\rangle = \sum_i P_{ij} |w_i\rangle$ *and* $\mathbf{w} = (d_1, \ldots, d_L)^T$*. The operations are captured by the equations* $\mathbf{w} = \mathbf{P}\mathbf{v} = \mathbf{P}\mathbf{O}\mathbf{u} = \mathbf{Q}\mathbf{u}$ *where if we introduce the concept of multiplication* $\mathbf{Q} = \mathbf{P}\mathbf{O}$ *of a matrix by another matrix, defined in terms of components as* $Q_{ij} = \sum_k P_{ik} O_{kj}$*. Further examination reveals matrix multiplication, like operator multiplication, to be distributive, associative, but not commutative (as is revealed by some examples – think of two rotations in 3D space).*

Having defined multiplication and addition we can (similar to what we did for complex numbers) extend the domain/range of any analytic function $F(x)$*, e.g.* e^x*, via its Taylor series. For example* $F(x) = \sum_n F_n x^n$ *leads to the operator (matrix) function* $\hat{F}(\hat{O}) = \sum_n F_n \hat{O}^n$ *(*$\mathbf{F}(\mathbf{O}) = \sum_n F_n \mathbf{O}^n$*). If a matrix is diagonal* $O_{ij} = \delta_{ij} O_{ii}$ *then one finds* $\mathbf{F}(\mathbf{O})$ *is also diagonal with* $F_{ii} = F(O_{ii})$*. Thus it is vastly simpler to calculate functions of operators using a basis in which it is diagonal. Note, the expressions above are nonsensical unless the input and output vector spaces of* $\hat{O}$ *are the same, so* $\mathcal{B}_v = \mathcal{B}_u = \mathcal{B}$*, and* $\mathbf{O}$ *is a square* $N \times N$ *matrix. Also note that properties of* $F(x)$ *that assume commutativity do not generalise, e.g. generally* $\mathrm{e}^{\mathbf{O}+\mathbf{P}} \neq \mathrm{e}^{\mathbf{O}}\mathrm{e}^{\mathbf{P}}$*.*

There are a few types of operator (and thus matrix) that deserve a special mention. They arise in the case that the input and output vector spaces are the same and $\mathcal{B}_v = \mathcal{B}_u = \mathcal{B}$*. Firstly, it is possible to define an operator that does nothing, the identity operator* $\mathbb{1}$*, taking each vector to itself. We can infer from its desired action on the basis vectors that in any basis it must always be represented by the identity matrix*

$$\mathbf{I} = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{pmatrix}.$$

Secondly, we can think of an inverse $\hat{O}^{-1}$ *of an operator* $\hat{O}$ *as being one which, when applied in succession with* $\hat{O}$*, does nothing i.e.* $\hat{O}^{-1}\hat{O} = \hat{O}\hat{O}^{-1} = \mathbb{1}$*. It also follows that* $\hat{Q}^{-1} = \hat{O}^{-1}\hat{P}^{-1}$ *for* $\hat{Q} = \hat{P}\hat{O}$*. In matrix notation,* $\mathbf{O}^{-1}\mathbf{O} = \mathbf{O}\mathbf{O}^{-1} = \mathbf{I}$ *and* $\mathbf{Q}^{-1} = \mathbf{O}^{-1}\mathbf{P}^{-1}$ *for* $\mathbf{Q} = \mathbf{P}\mathbf{O}$*.*

We've gone to some effort to stress the basis dependence of column vectors and matrices. By introducing operators and matrices, and their inverses, we now have the tools to properly analyse this basis dependence. We'll do it for the case that the input and output vector spaces of an operator $\hat{O}$ *are the same and we use the same basis for both* $\mathcal{B}_v = \mathcal{B}_u = \mathcal{B} = \{|v_i\rangle\}$*, which is also the same basis we use for some vector* $|v\rangle$*. We wish to relate the column vector* $\mathbf{v}$ *and matrix* $\mathbf{O}$ *representing* $|v\rangle$ *and* $\hat{O}$ *in basis* $\mathcal{B}$ *to the column vector* $\mathbf{v}'$ *and matrix* $\mathbf{O}'$ *representing* $|v\rangle$ *and* $\hat{O}$ *in another basis* $\mathcal{B}' = \{|v'_i\rangle\}$*. Specifically, we now have the terminology to say that the original and transformed bases* $\mathcal{B}$ *and* $\mathcal{B}'$ *are related by some operator* $\hat{P}$ *with inverse* $\hat{P}^{-1}$ *according to* $|v'_i\rangle = \hat{P}|v_i\rangle$*, where operator* $\hat{P}$ *in the original basis* $\mathcal{B}$ *is represented by matrix* $\mathbf{P}$*.*

1. Show that $\mathbf{v}' = \mathbf{P}^{-1}\mathbf{v}$ and $\mathbf{O}' = \mathbf{P}^{-1}\mathbf{O}\mathbf{P}$ or, equivalently, $\mathbf{v} = \mathbf{P}\mathbf{v}'$ and $\mathbf{O} = \mathbf{P}\mathbf{O}'\mathbf{P}^{-1}$.

   Solution: Start by writing $|v\rangle = \sum_i c_i |v_i\rangle = \sum_i c_i' |v_i'\rangle = \sum_i c_i' \hat{P} |v_i\rangle$. Use $\hat{P} |v_i\rangle = \sum_j P_{ji} |v_j\rangle$ to get $\sum_i c_i |i\rangle = \sum_i \sum_j P_{ji} c_i' |v_i\rangle$. Assuming the vectors of basis $\mathcal{B}$ to be linearly independent this implies $c_i = \sum_j P_{ji} c_i'$, which is exactly what is represented by $\mathbf{v} = \mathbf{P}\mathbf{v}'$. A similar but more cumbersome procedure works for the operator transformation.

The key thing to note here is that $\mathbf{v}' = \mathbf{P}^{-1}\mathbf{v}$ is the representation, in the original basis $\mathcal{B}$, of the vector $\hat{P}^{-1} |v\rangle$. In other words, instead of $\mathbf{v}'$ representing the same vector in a different basis transformed by $\hat{P}$, it could equivalently represent a vector transformed by $\hat{P}^{-1}$ in the same basis (think of the components of a vector in real space along some axis, and imagine the effect on the components while e.g. rotating the axis and rotating the vector oppositely). Further $\mathbf{O}' = \mathbf{P}^{-1}\mathbf{O}\mathbf{P}$ is the representation, in the original basis $\mathcal{B}$, of an operator $\hat{O}' = \hat{P}^{-1}\hat{O}\hat{P}$ that first transforms a vector by $\hat{P}$, does the original operation $\hat{O}$ and transforms back again by $\hat{P}^{-1}$. This operator can be thought (defined) as the transformation of the operator $\hat{O}$ by $\hat{P}^{-1}$ since $\hat{O}'$ maps vectors $\hat{P}^{-1} |v\rangle$ exactly as $\hat{O}$ did for $|v\rangle$. Then instead of $\mathbf{O}'$ representing the same operation in a different basis transformed by $\hat{P}$, it could equivalently represent an operation transformed by $\hat{P}^{-1}$ in the same basis. Thus, as far as the column vector and matrix representation is concerned, there is no difference between a passive transformation in which the basis is transformed according to $\hat{P}$ and an active transformation in which the vectors and operators themselves are transformed according to $\hat{P}^{-1}$. This is why I have begun by introducing the real vectors and operators first, and gone to great lengths to make these distinct from their basis dependent representations as column vectors and matrices. Even if you spend the majority of your time working with the later, it can be essential for understanding to relate these back to the former. Transformations are a good example of this.

So far, we have only introduced one type of operator (matrix), the operator (matrix) inverse, and used it to understand transformations. What else can we do with it? The main application is the problem of finding $|r\rangle$ ($\mathbf{r}$) satisfying some linear equation $\hat{O} |r\rangle = |v\rangle$ ($\mathbf{O}\mathbf{r} = \mathbf{v}$), given $\hat{O}$ ($\mathbf{O}$) and $|v\rangle$ ($\mathbf{v}$). For simplicity we choose to discuss the problem in the column vector and matrix representation. It is a very important type of problem because, as can be seen by writing $\mathbf{r} = (x, y, z, \dots)^T$ and expanding the matrix equation, the problem is equivalent to finding the solutions to an arbitrary set of M linear equations in N variables x, y, $z\dots$ Thus identifying the number of solutions and finding them is an important procedure to learn. The relation to the inverse is the following: In the case that $\mathbf{O}$ is an invertible matrix, in which case it must also be square, there must always be a single solution $\mathbf{r} = \mathbf{O}^{-1}\mathbf{v}$. Equally if there is always exactly one solution to $\mathbf{O}\mathbf{r} = \mathbf{v}$ then $\mathbf{r} = \mathbf{O}^{-1}\mathbf{v}$ can be used to unambiguously define an inverse $\mathbf{O}^{-1}$.

There are many resources that go into great depth on this topic, so we only attempt to summarise. The most important concept is the rank $\mathrm{r}\{\mathbf{O}\}$ of the matrix $\mathbf{O}$, equal to the number of linearly independent rows (alternatively, columns — both types of rank are always the same), which can be calculated by row reduction. Obviously, we must have $\mathrm{r}\{\mathbf{O}\} \le \min(M, N)$. When may a solution exist? The equation $\mathbf{O}\mathbf{r} = \mathbf{v}$ can be interpreted as saying there is a linear combination of the N columns $\mathbf{o}_i$ of $\mathbf{O} = (\mathbf{o}_1 \dots \mathbf{o}_N)$ that equals $\mathbf{v}$, i.e. $\mathbf{v}$ is linearly dependent on the columns $\mathbf{o}_i$. Thus adding a column $\mathbf{v}$ to $\mathbf{O}$ to form an augmented matrix $(\mathbf{O}\ \mathbf{v}) = (\mathbf{o}_1 \dots \mathbf{o}_N\ \mathbf{v})$ should not increase the number of linearly independent columns and thus $\mathrm{r}\{\mathbf{O}\} = \mathrm{r}\{(\mathbf{O}\ \mathbf{v})\}$. It follows that if $\mathrm{r}\{\mathbf{O}\} < \mathrm{r}\{(\mathbf{O}\ \mathbf{v})\}$ then there may be no solutions, but there may be some if $\mathrm{r}\{\mathbf{O}\} = \mathrm{r}\{(\mathbf{O}\ \mathbf{v})\}$ (it is not possible that $\mathrm{r}\{\mathbf{O}\} > \mathrm{r}\{(\mathbf{O}\ \mathbf{v})\}$). How many solutions? If $\mathrm{r}\{\mathbf{O}\} = N$, only possible if $M \ge N$, then we have N independent linear equations in N independent variables. There is either zero or one solution. If $\mathrm{r}\{\mathbf{O}\} < N$, then we have too few ($< N$) independent linear equations to provide a unique solution for the N independent variables. So there is either zero or an infinite number of solutions.

In summary: the number of solutions is determined by whether $\mathrm{r}\{\mathbf{O}\}$ *is less than or equal to* $\mathrm{r}\{(\mathbf{O}\ \mathbf{v})\}$ *(none or some solutions) and, if the latter, whether* $\mathrm{r}\{\mathbf{O}\}$ *is less than or equal to* N *(infinite or a single solution). There is a nice way to visualise this. If we write the i-th row of* $\mathbf{O} = (\mathbf{n}_1 \ldots \mathbf{n}_M)^T$ *as* $\mathbf{n}_i$ *and the i-th component of* $\mathbf{v} = (d_1||\mathbf{n}_1||, \ldots, d_M||\mathbf{n}_M||)^T$ *as* $d_i||\mathbf{n}_i||$ *then the set of equations can be written* $\hat{\mathbf{n}}_i \cdot \mathbf{r} = d_i$*, which represents* M *planes in* N*-dimensional space. If* $\mathrm{r}\{\mathbf{O}\} < \mathrm{r}\{(\mathbf{O}\ \mathbf{v})\}$ *then all planes never intersect at the same point, while if* $\mathrm{r}\{\mathbf{O}\} = \mathrm{r}\{(\mathbf{O}\ \mathbf{v})\}$ *then all planes do intersect at a point or an infinite number of points. If* $\mathrm{r}\{\mathbf{O}\} < N$ *then any intersection must be along an infinite set of points (e.g. a line or a plane) while if* $\mathrm{r}\{\mathbf{O}\} = N$ *any intersection must occur at a single points.*

2. Draw the various possibilities, for the case of $M = 3$ linear equations in $N = 3$ variables, when the solution (a) is unique, (b) does not exist, (c) is a line, and (d) is a plane.

Let's return to the rank discussion, in which we found the number of solutions is determined by whether $\mathrm{r}\{\mathbf{O}\}$ *is less than or equal to* $\mathrm{r}\{(\mathbf{O}\ \mathbf{v})\}$ *(none or some solutions) and, if the latter, whether* $\mathrm{r}\{\mathbf{O}\}$ *is less than or equal to* N *(infinite or a single solution). In the case of a square* $N \times N$ *matrix* $\mathbf{O}$ *these two conditions are related. If* $\mathrm{r}\{\mathbf{O}\} = N$ *and the columns of* $\mathbf{O}$ *thus represent* N *linearly independent vectors, then these must span the whole* N*-dimensional vectors space and thus surely there is a linear combination of them that gives* $\mathbf{v}$*, thus* $\mathrm{r}\{\mathbf{O}\} = \mathrm{r}\{(\mathbf{O}\ \mathbf{v})\}$*. It follows that if we find* $\mathrm{r}\{\mathbf{O}\} = N$ *then there is guaranteed to be exactly one solution to the equation* $\mathbf{O}\mathbf{r} = \mathbf{v}$*, else there are zero or infinity solutions. Thus (see above) full rank* $\mathrm{r}\{\mathbf{O}\} = N$ *is equivalent to an inverse* $\mathbf{O}^{-1}$ *existing, and rank deficient* $\mathrm{r}\{\mathbf{O}\} < N$ *to* $\mathbf{O}$ *not being invertible. We have found the necessary and sufficient condition for the existence of an inverse. In fact this makes perfect sense if we think about it in another way. The columns* $\mathbf{o}_i$ *of* $\mathbf{O} = (\mathbf{o}_1 \ldots \mathbf{o}_N)$ *by definition are the transformation by* $\mathbf{O}$ *of column vectors* $\mathbf{v}_i = (0 \ldots 010 \ldots 0)^T$ *with a single non-zero i-th element. If these columns* $\mathbf{o}_i$ *are linearly dependent (and* $\mathbf{O}$ *is rank deficient) then it means that a non-zero vector* $\mathbf{v}$ *could be found that, as well as the zero vector* $\mathbf{0}$ *itself, goes to* $\mathbf{0}$ *when operated on by* $\mathbf{O}$*. Thus* $\mathbf{O}$ *can not possibly have an inverse. Note, there are other names used for the case of* $\mathbf{O}$ *not being invertible, e.g.* $\mathbf{O}$ *is often said to be singular.*

3. By inspection, state whether

$$\mathbf{O} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 2 \\ 0 & 5 & 0 \end{pmatrix},$$

has an inverse.

There is a nice quantity that tells us whether $\mathbf{O}$ *is invertible or not (only relevant and valid for square matrices* $M = N$*). This quantity, which is important for many other reasons also, is the determinant* $\det\{\mathbf{O}\} = |\mathbf{O}|$ *(not to be confused with vector norms). It is equal to zero if* $\mathbf{O}$ *is rank deficient and non-zero if* $\mathbf{O}$ *has full rank. Again, there are many books that detail the definitions and properties of determinants (with proofs), so I do not give them here, but you may wish to look them up and use them (reproduce them) in your working.*

4. Calculate the determinants of the matrices

$$\mathbf{O} = \begin{pmatrix} 0 & -\mathrm{i} & \mathrm{i} \\ \mathrm{i} & 0 & -\mathrm{i} \\ -\mathrm{i} & \mathrm{i} & 0 \end{pmatrix}, \quad \text{and} \quad \mathbf{P} = \frac{1}{\sqrt{8}} \begin{pmatrix} \sqrt{3} & -\sqrt{2} & -\sqrt{3} \\ 1 & \sqrt{6} & -1 \\ 2 & 0 & 2 \end{pmatrix}.$$

5. $\mathbf{O}$ is a non-singular 3×3 matrix and $\mathbf{P} = 2\mathbf{O}^{-1}$. Calculate $|\mathbf{O}||\mathbf{P}|$.

6. For which values of α is the matrix

$$\mathbf{O} = \begin{pmatrix} 1 & -3 & 2 \\ -\alpha & -1 & 2 \\ 3 & \alpha & -4 \end{pmatrix},$$

not invertible?

Solution: $\alpha = 4$ and $\alpha = 1$.

We now have everything we need to know in order to find the number of solutions to $\mathbf{Or} = \mathbf{v}$, but what about finding the solutions themselves? Always check first whether any solutions exist by calculating the determinant. Once it is verified that a solution (solutions) exist, the sure-fire method to find it (them) is to eliminate variables via substitution (Gaussian elimination), which can be performed in a structured and minimally cumbersome way by doing row reduction with the augmented matrix. Doing this will always work, regardless of the number of solutions and without knowing whether there will be zero, one or infinitely many. For the case of a square and invertible matrix $\mathbf{O}$, there are two alternative methods to calculate the single solution. We can calculate the matrix inverse $\mathbf{O}^{-1}$ and then calculate $\mathbf{O}^{-1}\mathbf{v}$ or we can use a method credited to Cramer. Both procedures are covered in detail elsewhere, but we note that there is a nice way to arrive at Cramer's method for the $N = 3$-dimensional case. If we write the columns of $\mathbf{O} = (\mathbf{v}_1\ \mathbf{v}_2\ \mathbf{v}_3)$ as $\mathbf{v}_1$, $\mathbf{v}_2$ and $\mathbf{v}_3$ then $\mathbf{Or} = \mathbf{v}$ may be written $\mathbf{v}_1 x + \mathbf{v}_2 y + \mathbf{v}_3 z = \mathbf{v}$. Taking the scalar product with $\mathbf{v}_2 \times \mathbf{v}_3$ and rearranging, we obtain $x = \{\mathbf{v}\mathbf{v}_2\mathbf{v}_3\}/\{\mathbf{v}_1\mathbf{v}_2\mathbf{v}_3\}$. Note that the triple scalar product $\{\mathbf{v}_1\mathbf{v}_2\mathbf{v}_3\} = |\mathbf{O}|$ is nothing but the determinant. Similarly, we find $y = \{\mathbf{v}\mathbf{v}_3\mathbf{v}_1\}/|\mathbf{O}|$ and $z = \{\mathbf{v}\mathbf{v}_1\mathbf{v}_2\}/|\mathbf{O}|$.

7. Whilst you might find this question slightly repetitive, it allows you to gain familiarity with the different methods that can be used to solve simultaneous equations. Solve the following equations for x, y and z by:

a) Calculating the matrix inverse.
b) Using Cramer's method.
c) Using row reduction.

$$x + 2y + 3z = 2,$$
$$3x + 4y + 5z = 4,$$
$$x + 3y + 4z = 6\,.$$

Solution: $x = -3$, $y = 7$, $z = -3$.

8. Show that the following equations have solutions only if $\eta = 1$ or 2, and find the solutions in these cases:

$$x + y + z = 1,$$
$$x + 2y + 4z = \eta,$$
$$x + 4y + 10z = \eta^2\,.$$

Solution: $\eta = 1$: $y = -3z = 3(1-x)/2$. $\eta = 2$: $x = 2z = 2(1-y)/3$.

Having discussed the solution of $\hat{O}\,|r\rangle = |v\rangle$ ($\mathbf{Or} = \mathbf{v}$) we are now in a position to discuss the solution of an even more important vector equation. The eigenvalue problem for an $N \times N$ square operator (matrix) $\hat{O}$ ($\mathbf{O}$) is to find non-trivial pairs of scalars λ, called eigenvalues, and (column) vectors $|v\rangle$ ($\mathbf{v}$) that satisfy the equation $\hat{O}\,|v\rangle = \lambda\,|v\rangle$ ($\mathbf{Ov} = \lambda\mathbf{v}$). To solve it (again, continuing in column vector and matrix representation), we rearrange the eigenvalue equation to get $(\mathbf{O}-\lambda\mathbf{I})\mathbf{v} = \mathbf{0}$. We know from above that this only has non-trivial solution if $\det\{\mathbf{O}-\lambda\mathbf{I}\} = 0$, *thus by solving this equation containing the determinant we obtain the possible eigenvalues λ. Since this equation is a polynomial equation of order N, we obtain N eigenvalues (roots[a] of what is called the characteristic polynomial of $\mathbf{O}$), though they are not necessarily distinct (i.e. repeated roots). The eigenvector(s) $\mathbf{v}$ corresponding to each possible value of λ are then found from $(\mathbf{O}-\lambda\mathbf{I})\mathbf{v} = \mathbf{0}$. There may be more than one eigenvector satisfying this equation for a given λ, and this eigenvalue would then be called degenerate. The eigenvectors are defined only up to a constant of proportionality. Get used to solving this type of equation, it's hugely important. The most fundamental equation of quantum mechanics, the Schrödinger equation is an eigenvalue equation.*

[a]Note that the characteristic polynomial is basis dependent as it depends on the basis $\mathcal{B}$ used in defining $\mathbf{O}$, but its roots are not as these are properties of the real vector $\hat{O}$.

9. Find the eigenvectors and eigenvalues of the matrices

$$\mathbf{O} = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}, \quad \text{and} \quad \mathbf{P} = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}.$$

Solution: The auxiliary equation $\det\{\mathbf{O} - \lambda\mathbf{I}\} = 0$ becomes $\lambda^2 = -1$ and thus we have the maximum of two eigenvalues $\lambda_\pm = \pm\mathrm{i}$. Writing the corresponding eigenvalues as $\mathbf{v}_\pm = (a_\pm, b_\pm)^T$, the eigenvector equation $(\mathbf{O} - \lambda_\pm\mathbf{I})\mathbf{v}_\pm = \mathbf{0}$ becomes

$$\begin{pmatrix} \mp\mathrm{i} & -1 \\ 1 & \mp\mathrm{i} \end{pmatrix}\begin{pmatrix} a_\pm \\ b_\pm \end{pmatrix} = \mathbf{0}.$$

Looking at the top line or bottom line gives $\mp\mathrm{i}a_\pm + b_\pm = 0$, which rearranges to $b_\pm = \pm\mathrm{i}a_\pm$ and so

$$\mathbf{v}_\pm = \begin{pmatrix} a_\pm \\ \pm\mathrm{i}a_\pm \end{pmatrix}.$$

So the eigenvectors up to a constant of proportionality are

$$\mathbf{v}_\pm = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ \pm\mathrm{i} \end{pmatrix}.$$

The second auxiliary equation $\det\{\mathbf{P} - \lambda\mathbf{I}\} = 0$ becomes $\lambda^2 = 0$ and thus we have only one eigenvalue $\lambda = 0$. Writing the corresponding eigenvalue (though being aware of the possibility of finding more than one) as $\mathbf{v} = (a, b)^T$. The eigenvector equation $(\mathbf{P} - \lambda\mathbf{I})\mathbf{v} = \mathbf{0}$ becomes

$$\begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}\begin{pmatrix} a \\ b \end{pmatrix} = \mathbf{0}.$$

Looking at the top line gives $b = 0$ and the bottom line $0 = 0$ yields no information, and so

$$\mathbf{v}_\pm = \begin{pmatrix} a \\ 0 \end{pmatrix}.$$

So the eigenvector up to a constant of proportionality is

$$\mathbf{v} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}.$$

We haven't yet used the scalar product. We haven't made use of any notions of length or angle. And we haven't thought about the effect this might have on operators. In the second half of this problem set we look at the implications of the scalar product on operators.

Recall from a previous tutorial the inner product on a complex vector space. It is a scalar function $f(|u\rangle, |v\rangle)$*, abbreviated* $\langle v|u\rangle$*, that is linear in its first argument and satisfies* $f(|u\rangle, |v\rangle) = (f(|v\rangle, |u\rangle))^*$*, or* $\langle v|u\rangle = (\langle u|v\rangle)^*$*. What does this mean for operators? Let's consider applying an operator* $\hat{O}$ *to the vector appearing in the first argument* $f(\hat{O}|u\rangle, |v\rangle)$ *and abbreviate this by* $\langle v|\hat{O}|u\rangle$*. It is tempting to think that the conjugate* $(f(\hat{O}|u\rangle, |v\rangle))^* = f(|v\rangle, \hat{O}|u\rangle)$ *should be abbreviated* $\langle u|\hat{O}|v\rangle$ *and is thus equal to* $f(|u\rangle, \hat{O}|v\rangle)$*. However, in general, there is no reason for this to be the case (in fact, in general, the last two equations are nonsensical as the input vector space to the operator doesn't necessarily equal the vector space of the object on which it is acting). Instead, in general, the operator appearing in the last two equations will not be* $\hat{O}$*, but will be some other operator that we write* $\hat{O}^\dagger$ *and define by this property* $(\langle v|\hat{O}|u\rangle)^* = (f(\hat{O}|u\rangle, |v\rangle))^* = f(|u\rangle, \hat{O}^\dagger|v\rangle) = \langle u|\hat{O}^\dagger|v\rangle$*. We call this operator* $\hat{O}^\dagger$ *the Hermitian conjugate of operator* $\hat{O}$*, and it plays as an important role in linear algebra as the operator inverse. Note that it follows that* $\hat{Q}^\dagger = \hat{O}^\dagger\hat{P}^\dagger$ *for* $\hat{Q} = \hat{P}\hat{O}$*.*

The other effect of introducing an inner product is that it gives us the concept of orthonormal bases in which we can now represent our operators. This turns out to be immensely useful, as it was for vectors. Consider orthonormal input and output bases $\mathcal{B}_u = \{|u_i\rangle\}$ *and* $\mathcal{B}_v = \{|v_i\rangle\}$ *satisfying* $\langle u_i|u_j\rangle = \langle v_i|v_j\rangle = \delta_{ij}$*. We have that* $\mathbf{O}$ *is defined by* $\hat{O}|u_j\rangle = \sum_i O_{ij}|v_i\rangle$*. By taking the inner product of this with* $|v_j\rangle$ *we find that, due to orthonormality* $O_{ij} = \langle v_j|\hat{O}|u_j\rangle$*, an expression that is so commonly associated with the idea of a matrix element that it is widely forgotten that this is only the case for orthonormal bases.*

These two pieces given to us by the scalar product can be combined to give two things. Firstly, we can find an expression for the matrix representing a Hermitian conjugate $\hat{O}^\dagger$ *in terms of the matrix representing* $\hat{O}$*, given an orthonormal basis. We use that from the definition of the Hermitian conjugate* $(\langle v_i|\hat{O}|u_j\rangle)^* = \langle u_j|\hat{O}^\dagger|v_i\rangle$*. The definitions of matrix elements in an orthonormal basis then immediately gives* $(O_{ij})^* = O^\dagger_{ji}$*. So the matrix* $\mathbf{O}^\dagger$ *representing the Hermitian conjugate* $\hat{O}^\dagger$ *in orthonormal bases is called the conjugate transpose* $\mathbf{O}$ *since it is obtained by swapping columns with rows (transposition) and taking the conjugate of each element. Secondly in an orthonormal basis with* $\mathbf{u} = (b_1 \cdots b_N)^T$ *and* $\mathbf{v} = (c_1 \cdots c_N)^T$ *we can show* $\langle v|u\rangle = \sum_i c_i^* b_i = \mathbf{v}^\dagger\mathbf{u}$*. People forget that these things only hold for orthonormal bases.*

So the inner product has given us the Hermitian conjugate and its representation in orthogonal bases, let's now discuss the amazing number of properties that follow from this. While the first half of this problem set (after defining an operator and a matrix) could be thought of as discussion surrounding the matrix inverse, the rest of this problem set can be thought of as examining the repercussions of introducing an inner product and thus Hermitian conjugate.

The introduction of the Hermitian conjugate allows us to define several extremely important types of operators. We'll also at the same time define several extremely important types of matrices (in the same way that defining a Hermitian conjugate operator defined a conjugate transpose matrix, using an orthonormal basis). For example, defining normal operators will define what we mean by normal matrices, though the fact that a normal operator is represented by a normal matrix is only true for orthonormal bases etc. To avoid the tedious cumbersome language we would otherwise have to use to ensure correctness, we thus from here on assume our bases always to be orthonormal so normal operators are represented by normal matrices etc. and the Hermitian conjugate corresponds to the conjugate transpose. We'll start by listing the operator and matrix types, then go on to discuss their properties. From their definitions it is clear that the input and output vector spaces for the operators must be the same and the matrices are clearly square.

The first are Hermitian operators (matrices), usually denoted $\hat{H}$ ($\mathbf{H}$), that satisfy $\hat{H} = \hat{H}^\dagger$ ($\mathbf{H} = \mathbf{H}^\dagger$). Secondly, unitary operators (matrices) are those with their Hermitian conjugate equal to their inverse $\hat{U}^{-1} = \hat{U}^\dagger$ ($\mathbf{U}^{-1} = \mathbf{U}^\dagger$). Thirdly, an operator (matrix) is normal if it satisfies $\hat{N}\hat{N}^\dagger = \hat{N}^\dagger\hat{N}$ ($\mathbf{N}\mathbf{N}^\dagger = \mathbf{N}^\dagger\mathbf{N}$). Both Hermitian and unitary operators have twins in the case of a real (rather than complex) vector space. In this case, pretty much everything remains the same but without the conjugates. The conjugate transpose becomes merely a transpose and we call real matrices satisfying $\mathbf{S} = \mathbf{S}^T$ symmetric, analogous to Hermitian matrices. Similarly, the analog of a unitary matrix is one whose inverse is its transpose $\mathbf{R}^{-1} = \mathbf{R}^T$ and is called an orthogonal matrix (the use of the letter R here is common because these types of matrices are associated with reflections and rotations). There are a few other related matrix types, the definitions of which you can look up. Let's get familiar with these matrix types.

10. State whether the matrices **O** and **P** in Question **4** are (i) real, (ii) diagonal, (iii) symmetric, (iv) antisymmetric, (v) singular, (vi) orthogonal, (vii) Hermitian, (viii) anti-Hermitian, (ix) unitary, (x) normal.

11. Prove that

(a) $(\mathbf{O}^\dagger)^\dagger = \mathbf{O}$ and $(\mathbf{O}^T)^T = \mathbf{O}$.

(b) $(\mathbf{O}^{-1})^{-1} = \mathbf{O}$.

(c) $(\mathbf{O}^\dagger)^{-1} = (\mathbf{O}^{-1})^\dagger$ and $(\mathbf{O}^T)^{-1} = (\mathbf{O}^{-1})^T$.

(d) **O** is Hermitian iff $\mathbf{O}^\dagger$ is Hermitian.

(e) **O** is unitary iff $\mathbf{O}^{-1}$ is unitary.

(f) **O** is Hermitian iff $\mathbf{O}^{-1}$ is Hermitian.

(g) **O** is unitary iff $\mathbf{O}^\dagger$ is unitary.

(h) If **O** has columns that are orthonormal column vectors then it is unitary.

Now let's begin to delve into the properties of these operator (matrix) types, beginning with Hermitian operators (matrices). This is a very important type and plays a central role in quantum mechanics. One important fact is that you can use the definition of a Hermitian operator (matrix) to show that its eigenvalues are real and eigenvectors can form a complete orthonormal basis for the vectors space, as we will now demonstrate.

12. Show (a) that Hermitian operators (matrices) must have real eigenvalues, and (b) that the eigenvectors of Hermitian operators (matrices) corresponding to different eigenvalues are orthogonal.

Solution: Let $|v_i\rangle$ be non-zero eigenvectors of a Hermitian operator $\hat{H}$ with eigenvalues λ_i. By the definition of Hermitivity $\langle v_i|\hat{H}|v_j\rangle = (\langle v_j|\hat{H}|v_i\rangle)^\dagger$. Using the eigenvalue equation, a property of the inner product, and rearranging, we are left with $\langle v_i|v_j\rangle(\lambda_j - \lambda_i^*) = 0$. If $i = j$ then $\langle v_i|v_j\rangle = ||\,|v_i\rangle\,||^2$ is non-zero and so $\lambda_i = \lambda_i^*$, proving (a) that λ_i must be real. If $i \neq j$ then $\lambda_j - \lambda_i^* = \lambda_j - \lambda_i$ is non-zero, assuming no degeneracy, and so $\langle v_i|v_j\rangle = 0$, proving (b) that the two eigenvectors must be orthogonal. Like all proofs we'll use here, this can be replaced by a matrix proof by representing the equations in some orthonormal basis $\mathcal{B}$, giving $\mathbf{v}_i^\dagger\mathbf{H}\mathbf{v}_j = (\mathbf{v}_j^\dagger\mathbf{H}\mathbf{v}_i)^\dagger$ and $\mathbf{v}_i^\dagger\mathbf{v}_j(\lambda_j - \lambda_i^*) = 0$ and so on.

This doesn't quite prove that there are N eigenvectors, all orthogonal. We haven't shown that there are N eigenvectors and we still need to prove that all eigenvectors corresponding to the same eigenvalue can be chosen to be orthogonal. The interested reader can find beyond-syllabus proofs of these things by looking up the 'spectral theorem', e.g. on Wikipedia, or reading material introducing mathematics for physicists, e.g., the first chapter of Shankar's 'Principles of quantum mechanics'. Essentially, the proofs use the following logic: (i) It is known there must be at least one eigenvalue and thus eigenvector. (ii) Selecting one, it is possible to think of a subspace, a vector space formed by all of the vectors orthogonal to that eigenvector. (iii) The Hermitian operator, due to its Hermitivity, transforms all vectors in this subspace to others in this subspace and we are in the same position with respect to this subspace as we were with the full vector space. Thus we can go back to (i) and repeat, selecting another eigenvector in this space that is automatically orthogonal to those previous, and so on until N are selected.

To find the eigenvalues and eigenvectors we simply use the usual procedure rather than the approach above, but the above allows us to know in advance of finding them the existence of N (possible not distinct) real eigenvalues λ_i corresponding to orthonormal eigenvectors $|v_i\rangle$ ($\mathbf{v}_i$) of Hermitian operator (matrix) $\hat{H}$ ($\mathbf{H}$). Such a set of vectors must form a complete orthonormal basis $\mathcal{B}_H = \{|v_i\rangle\}$ for the N-dimensional vector space. Exactly the same thing holds more generally, with a more complicated proof, for normal matrices (which include Hermitian matrices), except in that case the eigenvalues are not necessarily real. This allows us to define the operator itself by its eigenvalues and eigenvectors. Let us examine the consequence of this in the column vector and matrix representation.

13. Show that any normal (which includes Hermitian) matrix $\mathbf{N}$ (equivalently any matrix representing a normal operator $\hat{N}$ in an arbitrary orthonormal basis $\mathcal{B}$) (a) has the form $\mathbf{N} = \mathbf{V}\mathbf{D}\mathbf{V}^{-1}$ (equivalently $\hat{N} = \hat{V}\hat{D}\hat{V}^{-1}$), where (b) $\mathbf{D}$ is diagonal i.e. $D_{ij} = \delta_{ij}D_{ij}$ and (c) has diagonal elements $D_{ii} = \lambda_i$ equal to the eigenvalues of $\mathbf{N}$ (equivalently $\hat{N}$), and (d) the columns of $\mathbf{V} = (\mathbf{v}_1 \ldots \mathbf{v}_N)$ are the eigenvectors of $\mathbf{v}_i$ of $\mathbf{N}$ (equivalently, the column vectors $\mathbf{v}_i$ representing eigenvectors $|v_i\rangle$ in basis $\mathcal{B}$). Interpret $\mathbf{N} = \mathbf{V}\mathbf{D}\mathbf{V}^{-1}$.

Solution: The simple approach is to start with the answer. Write $\mathbf{N} = \mathbf{V}\mathbf{D}\mathbf{V}^{-1}$ with $\mathbf{V} = (\mathbf{v}_1 \ldots \mathbf{v}_N)$. The matrix $\mathbf{V}$ takes trivial vectors $\mathbf{v}'_i$ to $\mathbf{v}_i$, where $\mathbf{v}'_i$ is the vector with all elements zero except for the i-th element, which is unity. Thus $\mathbf{V}^{-1}$ takes $\mathbf{v}_i$ to $\mathbf{v}'_i$. Use this to show that $\mathbf{V}\mathbf{D}\mathbf{V}^{-1}$ acts as $\mathbf{N}$ should on its eigenvectors $\mathbf{v}_i$, and thus it must indeed be equal to $\mathbf{N}$. We can interpret the action of $\mathbf{N} = \mathbf{V}\mathbf{D}\mathbf{V}^{-1}$ on a vector as transforming via $\mathbf{V}^{-1}$ into the basis of its eigenvectors, in which it must act diagonally, performing this action, and then transforming the result back via $\mathbf{V}$. In fact this is how you would derive the expression from first principles.

14. Using the previous question, show that if $\mathbf{N}$ ($\hat{N}$) is a normal matrix (operator) then a matrix (an operator) function simplifies as $\mathbf{F}(\mathbf{N}) = \mathbf{V}\mathbf{F}(\mathbf{D})\mathbf{V}^{-1}$ ($\hat{F}(\hat{N}) = \hat{V}\hat{F}(\hat{D})\hat{V}^{-1}$), where the function $F(x)$ is analytic.

Solution: Choosing to work in the column vector and matrix representation, and writing $F(x) = \sum_n F_n x^n$ (this can be generalised easily to functions defined by a Taylor series rather than a Maclaurin series), we have $\mathbf{F}(\mathbf{N}) = \sum_n F_n \mathbf{N}^n$. Since $\mathbf{N} = \mathbf{V}\mathbf{D}\mathbf{V}^{-1}$ the matrices $\mathbf{V}$ and $\mathbf{V}^{-1}$ cancel each other such that $\mathbf{N}^n = \mathbf{V}\mathbf{D}^n\mathbf{V}^{-1}$, thus by linearity $\mathbf{F}(\mathbf{N}) = \mathbf{V}\mathbf{F}(\mathbf{D})\mathbf{V}^{-1}$.

15. Using the previous questions, calculate $\mathrm{e}^{\mathrm{i}\theta\mathbf{H}}$ where matrix $\mathbf{H}$ is that which is called $\mathbf{O}$ in question **9**.

Solution: By quick inspection $\mathbf{H}$ is Hermitian and thus normal. So all of the results above hold. Specifically we know $\mathbf{F}(\mathbf{H}) = \mathbf{V}\mathbf{F}(\mathbf{D})\mathbf{V}^{-1}$ with $\mathbf{H} = \mathbf{V}\mathbf{D}\mathbf{V}^{-1}$ and how to construct $\mathbf{D}$ and $\mathbf{V}$ from the eigenvalues and eigenvectors. Since we have already calculated these we can immediately write

$$\mathbf{D} = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \quad \text{and} \quad \mathbf{V} = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 & 1 \\ \mathrm{i} & -\mathrm{i} \end{pmatrix}.$$

Thus

$$\mathrm{e}^{\mathrm{i}\theta\mathbf{H}} = \frac{1}{2}\begin{pmatrix} 1 & 1 \\ \mathrm{i} & -\mathrm{i} \end{pmatrix}\begin{pmatrix} \mathrm{e}^{\mathrm{i}\theta} & 0 \\ 0 & \mathrm{e}^{-\mathrm{i}\theta} \end{pmatrix}\begin{pmatrix} 1 & -\mathrm{i} \\ 1 & -\mathrm{i} \end{pmatrix} = \begin{pmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{pmatrix}.$$

You may recognise this as the matrix for a rotation by angle θ in 2D space. You've just seen your first example of a generator of rotations! The Hamiltonian in quantum mechanics generates system dynamics in the same way.

Next let us move on to properties of unitary operators (matrices). The most important property is that the inner product of two arbitrary vectors is invariant under the action of an operator iff it is unitary. Thus unitary operators do not alter the norms of vectors and keep orthogonal vectors orthogonal, meaning they represent generalisations (to complex vectors spaces) of rotations and reflections.

16. Show that the inner product of two vectors is invariant under the action of a unitary operator.

Solution: By the definition of a Hermitian conjugate the inner product of $\hat{U}\,|u\rangle$ with $\hat{U}\,|v\rangle$ is $\langle v|\,\hat{U}^\dagger\hat{U}\,|u\rangle$. If $\hat{U}$ is unitary then by this property's definition we obtain $\langle v|\,\hat{U}^\dagger\hat{U}\,|u\rangle = \langle v|u\rangle$, equal to the inner product of the two vectors prior to their operation by $\hat{U}$. In column vector and matrix representation this would proceed in exactly the same way $(\mathbf{Uv})^\dagger\mathbf{Uu} = \mathbf{v}^\dagger\mathbf{U}^\dagger\mathbf{Uu} = \mathbf{v}^\dagger\mathbf{u}$.

The implication of this is that if $\mathcal{B}$ is an orthonormal basis then the basis $\mathcal{B}'$ related to it by operator $\hat{P}$ is orthonormal iff $\hat{P} = \hat{U}$ is unitary. A unitary operator can be used to take any orthonormal basis to any other. This means that a normal matrix $\mathbf{N}$ (equivalently the representation of a normal operator $\hat{N}$ in an orthonormal basis $\mathcal{B}$ can be written $\mathbf{N} = \mathbf{UDU}^{-1}$ because orthonormal bases $\mathcal{B}$ and $\mathcal{B}'$ must be related by a unitary operator. Further, since the columns $\mathbf{u}_i$ of $\mathbf{U} = (\mathbf{u}_1 \ldots \mathbf{u}_N)$ are column vectors representing the orthonormal eigenvectors $|v_i\rangle$ of $\hat{N}$ in a now orthonormal basis $\mathcal{B}'$, this means that $\mathbf{u}_i^\dagger\mathbf{u}_j = \delta_{ij}$, which shows that the columns of a matrix are orthonormal column vectors if it is unitary, which complements a proof asked for in an earlier questions that a matrix with columns that are orthonormal column vectors is unitary. If you ever forget which way around the $\mathbf{U}$ and $\mathbf{U}^{-1}$ go above, write out the columns, using $\mathbf{U}^{-1} = \mathbf{U}^\dagger$, as $\mathbf{N} = (\mathbf{u}_1 \ldots \mathbf{u}_N)\mathbf{D}(\mathbf{u}_1 \ldots \mathbf{u}_N)^\dagger$ and check that what happens when it acts on some $\mathbf{u}_i$ is as expected, using the orthonormal properties.

All the algebraic tools you need are above but there are a few extra objects, built from things we have already introduced, that ease the manipulation of equations. The most useful object is the outer product of two vectors $|v\rangle\langle u|$ defined as the operator that takes the inner product of the object to the right with $\langle u|$ and places this along $|v\rangle$. For an orthogonal basis $\mathcal{B} = \{|v_i\rangle\}$ we can use this to represent the operator $\hat{O}$ nicely and explicitly in terms of the matrix representing it in this basis $\hat{O} = \sum_{ij} O_{ij}\,|v_i\rangle\langle v_j|$ (think about why this is correct). Also, in some orthogonal basis $|v\rangle\langle u|$ must be, because of its action, represented by the matrix $\mathbf{vu}^\dagger$. A nice example is that it lets you immediately write $\hat{N} = \sum_i \lambda_i\,|v_i\rangle\langle v_i|$ for a normal operator with eigenvectors $|v_i\rangle$ and eigenvalues λ_i, which is clearly represented by a diagonal matrix $\mathbf{D}$ in the basis $\mathcal{B} = \{|v_i\rangle\}$ and thus by $\mathbf{UDU}^{-1}$ in any other orthonormal basis.

Linear algebra is a big topic, much too big for how long we get to spend on it in this year. The hope is that we can cover most of the important results, focusing initially on the uses and properties of column vectors and matrices and gradually then on the fact that column vectors and matrices are not the most fundamental objects, but a representation of them. Over time you will become familiar working in column vector and matrix notation, bra and ket notation, and performing all of the types of manipulations possible. Don't be discouraged at first and don't forget to learn by answering questions and using vectors and matrices to gain the intuition that will back up the more formal language you are acquiring.

Class problems

17. The Mathematica output in Figure 1 shows the application of a linear operation to a 2D vector. Interpret each step in these calculations in terms of rotations in the 2D plane.

We define a 2D vector of length r and angle α with the x-axis

```
In[142]:= rv = r {Cos[α], Sin[α]};
          Print["rv = " , MatrixForm[rv]]
```

$$\text{rv} = \begin{pmatrix} \text{r Cos}[\alpha] \\ \text{r Sin}[\alpha] \end{pmatrix}$$

The matrix R defines a linear operation on this vector

```
In[144]:= R = {{Cos[θ], -Sin[θ]}, {Sin[θ], Cos[θ]}};
          Print["R = " , MatrixForm[R]]
```

$$\text{R} = \begin{pmatrix} \text{Cos}[\theta] & -\text{Sin}[\theta] \\ \text{Sin}[\theta] & \text{Cos}[\theta] \end{pmatrix}$$

Applying this matrix to rv gives

```
In[147]:= Print["R.rv = ", R.rv // FullSimplify // MatrixForm]
```

$$\text{R.rv} = \begin{pmatrix} \text{r Cos}[\alpha+\theta] \\ \text{r Sin}[\alpha+\theta] \end{pmatrix}$$

The inverse operation is

```
In[148]:= Print["R^-1 = ", Inverse[R] // Simplify // MatrixForm]
```

$$\text{R\^{}-1} = \begin{pmatrix} \text{Cos}[\theta] & \text{Sin}[\theta] \\ -\text{Sin}[\theta] & \text{Cos}[\theta] \end{pmatrix}$$

Figure 1: A linear matrix operation in 2D.

18. Use row reduction to find the rank of the matrix

$$\mathbf{O} = \begin{pmatrix} 1 & 1 & 0 & -1 \\ 0 & 2 & -2 & 1 \\ 3 & 2 & 0 & -4 \\ 1 & -2 & x & 0 \end{pmatrix},$$

where x is a real parameter (note the rank depends on the value of x) and thus the number of possible solutions to $\mathbf{Or} = \mathbf{v}$.

19. Find the determinant of the matrix

$$\mathbf{O} = \begin{pmatrix} 0 & 4 & 0 & -3 \\ 1 & 1 & 5 & 2 \\ 1 & -2 & 0 & 6 \\ 3 & 0 & 0 & 1 \end{pmatrix}.$$

Solution: -250.

20. A linear homogeneous transformation that moves points $\mathbf{r} = (x, y, z)$ to new points $\mathbf{r}' = (x', y', z')$ can be represented by a matrix $\mathbf{O}$ such that $\mathbf{O}\mathbf{r} = \mathbf{r}'$. Determine the 3×3 matrices for the following situations:

(a) A counter-clockwise rotation by an angle α about the z-axis.

(b) A counter-clockwise rotation by an angle β about the x-axis.

(c) Counter-clockwise rotations, first by an angle β about the x-axis and then by an angle α about the z-axis.

21. An ellipse, defined by the equation $x^2 + 3y^2 - 2xy = 1$, may be rewritten in matrix form as $\mathbf{x}^T\mathbf{S}\mathbf{x} = 1$ where $\mathbf{x} = (x, y)^T$ and $\mathbf{S}$ is a symmetric matrix. Determine $\mathbf{S}$.

22. Using column vector/matrix methods, show that the line of intersection of the planes $x + 2y + 3z = 0$ and $3x + 2y + z = 0$ is equally inclined to the x- and z-axes and makes an angle $\arccos(-2/\sqrt{6})$ with the y-axis.

23. Find the 2×2 (two-dimensional) matrix $\mathbf{O}$ which has the properties:

$$\mathbf{O}\begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \quad \text{and} \quad \mathbf{O}^{-1}\begin{pmatrix} 1 \\ 2 \end{pmatrix} = \begin{pmatrix} 1 \\ -1 \end{pmatrix}.$$

Solution:

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

24. Let $\mathbf{O}$ be a square matrix.

(a) If $\mathbf{O}^2 = 0$, show that $\mathbf{I} - \mathbf{O}$ is invertible.

(b) If $\mathbf{O}^3 = 0$, show that $\mathbf{I} - \mathbf{O}$ is invertible.

(c) If $\mathbf{O}^n = 0$ for some positive integer n, show that $\mathbf{I} - \mathbf{O}$ is invertible.

(d) If $\mathbf{O}^2 + 2\mathbf{O} + \mathbf{I} = 0$, show that $\mathbf{O}$ is invertible.

25. For an operator, the trace is defined as $\mathrm{tr}\{\hat{O}\} = \sum_i \langle v_i | \hat{O} | v_i \rangle$ for any orthonormal basis $\mathcal{B} = \{|v_i\rangle\}$. For a matrix $\mathbf{O}$, the trace is defined as the sum of its diagonal elements, so $\mathrm{tr}\{\mathbf{O}\} = \sum_i O_{ii}$. If $\mathbf{O}$ represents $\hat{O}$ in some basis, these two traces are only equivalent if that basis is orthonormal.

(a) Show that $\mathrm{tr}\{\mathbf{OP}\} = \mathrm{tr}\{\mathbf{PO}\}$.

(b) Show that $\mathrm{tr}\{\mathbf{P}^{-1}\mathbf{OP}\} = \mathrm{tr}\{\mathbf{O}\}$. Discuss this result in relation to basis change.

(c) Show that $\mathrm{tr}\{\mathbf{U}^{-1}\mathbf{N}^n\mathbf{U}\} = \sum_i \lambda_i^n$ for a normal matrix $\mathbf{N}$ and unitary matrix $\mathbf{U}$, where λ_i are the eigenvalues of $\mathbf{N}$.