Notes on Matrix Algebra and Inverse Matrices

Inverse of a linear transformation

<u>**Definition**</u>: We call a linear transformation $T:\mathbf{R}^n \to \mathbf{R}^n$ ***invertible*** (also called nonsingular) if it is both <u>one-to-one</u> (if $T(\mathbf{x}) = T(\mathbf{y})$ then necessarily $\mathbf{x} = \mathbf{y}$) and <u>onto</u> the codomain (for every vector $\mathbf{z} \in \mathbf{R}^n$ there is a (unique) $\mathbf{x} \in \mathbf{R}^n$ such that $T(\mathbf{x}) = \mathbf{z}$).

It's relatively easy to see why invertibility will only make sense for linear transformations $T:\mathbf{R}^n \to \mathbf{R}^n$ given by (square) $n\times n$ matrices, $T(\mathbf{x}) = \mathbf{Ax}$; and certainly not all such transformations will have inverses.

This is the same notion of invertibility we have for functions elsewhere. However, in the context of linear transformations given by $T(\mathbf{x}) = \mathbf{Ax}$ we have a simple algorithmic way of not only determining if this linear transformation is invertible, but also for determining the matrix of this inverse function (referred to as its **inverse matrix** $\mathbf{A}^{-1}$) if it exists. It all comes down to an enhanced view of row reduction and what invertibility means in terms of rank and the reduced row-echelon form of an associated $n\times 2n$ matrix.

> Some of you may already know about inverse matrices and may be tempted to use them to solve arbitrary systems of linear equations. **This is a very bad idea!** Linear systems can be inconsistent, and they can also have infinitely many solutions. If you restrict yourself to using inverse matrices for solving all linear systems, you will very soon come to regret this. Row reduction is universally valid.

Consider a simple example like $T(\mathbf{x}) = \mathbf{Ax}$ where $\mathbf{A} = \begin{bmatrix} 3 & 1 \\ 2 & -1 \end{bmatrix}$. Given any input vector $\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$, this transformation will give the output vector $T(\mathbf{x}) = \mathbf{Ax} = \begin{bmatrix} 3 & 1 \\ 2 & -1 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \mathbf{y}$. To be invertible, given any vector $\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix}$, we would have to be able to solve uniquely for $\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$ in terms of the components of $\mathbf{y}$.

What does this mean in terms of algebra?

$$T(\mathbf{x}) = \mathbf{Ax} = \begin{bmatrix} 3 & 1 \\ 2 & -1 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 3x_1 + x_2 \\ 2x_1 - x_2 \end{bmatrix} = \begin{bmatrix} y_1 & \\ & y_2 \end{bmatrix}.$$

I've staggered the right-hand-sides a bit to suggest the approach. All we have to do is augment the matrix a little more and represent these two equations by entering the coefficients on both the left-hand-side <u>and</u> the right-hand side. This gives $\left[\begin{array}{cc|cc} 3 & 1 & 1 & 0 \\ 2 & -1 & 0 & 1 \end{array}\right] = [\mathbf{A} \mid \mathbf{I}]$ where $\mathbf{I}$ is the appropriate Identity matrix. If it's possible to solve uniquely for $\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$, we'll discover this by row reduction. That is:

$$\left[\begin{array}{cc|cc} 3 & 1 & 1 & 0 \\ 2 & -1 & 0 & 1 \end{array}\right] \to \left[\begin{array}{cc|cc} 1 & 2 & 1 & -1 \\ 2 & -1 & 0 & 1 \end{array}\right] \to \left[\begin{array}{cc|cc} 1 & 2 & 1 & -1 \\ 0 & -5 & -2 & 3 \end{array}\right]$$

$$\to \left[\begin{array}{cc|cc} 1 & 2 & 1 & -1 \\ 0 & 1 & 2/5 & -3/5 \end{array}\right] \to \left[\begin{array}{cc|cc} 1 & 0 & 1/5 & 1/5 \\ 0 & 1 & 2/5 & -3/5 \end{array}\right] = [\mathbf{I} \mid \mathbf{A}^{-1}]$$

This last array reads: $\left\{\begin{array}{l} x_1 = \frac{1}{5}y_1 + \frac{1}{5}y_2 \\ x_2 = \frac{2}{5}y_1 - \frac{3}{5}y_2 \end{array}\right\}$.

We discover two things from this example:

(1) If the matrix $\mathbf{A}$ has full rank, then we will be able to solve uniquely for $\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$.

(2) If the matrix **A** has full rank, the matrix of its inverse will appear in the right half of

$\text{rref}\left[\mathbf{A} \mid \mathbf{I}_n\right] = \left[\mathbf{I}_n \mid \mathbf{A}^{-1}\right]$.

The situation in general is no different. If $T : \mathbf{R}^n \to \mathbf{R}^n$ is given by a (square) $n \times n$ matrix, $T(\mathbf{x}) = \mathbf{A}\mathbf{x}$, and if we write $T(\mathbf{x}) = \mathbf{A}\mathbf{x} = \mathbf{y}$, then we write the $n \times 2n$ matrix $\left[\mathbf{A} \mid \mathbf{I}_n\right]$ and carry out row reduction to determine whether this has full rank *n*. If it doesn't have full rank, then we can't solve uniquely for $\mathbf{x}$, and the transformation (and its matrix) is not invertible. However,

(1) If the matrix **A** has full rank *n*, then we will be able to solve uniquely for $\mathbf{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$ in terms of $\mathbf{y} = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}$.

(2) If the matrix **A** has full rank *n*, the matrix of its inverse will appear in the right half of

$\text{rref}\left[\mathbf{A} \mid \mathbf{I}_n\right] = \left[\mathbf{I}_n \mid \mathbf{A}^{-1}\right]$.

Note: This is generally the simplest way to find the inverse of a matrix by hand. There is a formulaic way of doing this using determinants (based on Cramer's Rule), but it's impractical for matrices larger than 3×3.

There is an ever-so-simple way to find the inverse of a 2×2 matrix $\mathbf{A} = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$. First, calculate its determinant $\det(\mathbf{A}) = ad - bc$. You can easily show using our row reduction method that if $\det(\mathbf{A}) = ad - bc = 0$, then the matrix **A** will not have full rank and will not be invertible. If $\det(\mathbf{A}) = ad - bc \neq 0$, then **A** will have full rank and will be invertible, and $\mathbf{A}^{-1} = \dfrac{1}{\det(\mathbf{A})}\begin{bmatrix} d & -b \\ -c & a \end{bmatrix} = \dfrac{1}{ad - bc}\begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$.

For example, if $\mathbf{A} = \begin{bmatrix} 3 & 1 \\ 2 & -1 \end{bmatrix}$, then $\det(\mathbf{A}) = (3)(-1) - (1)(2) = -5 \neq 0$ and $\mathbf{A} = -\frac{1}{5}\begin{bmatrix} -1 & -1 \\ -2 & 3 \end{bmatrix} = \begin{bmatrix} 1/5 & 1/5 \\ 2/5 & -3/5 \end{bmatrix}$.

The corresponding method for 3×3 matrices has similar elements to this, but involves far more calculation.

Matrix algebra

Definition: Given any scalar $k \in \mathbf{R}$ and an $m \times n$ matrix $\mathbf{A} = \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \cdots & \vdots \\ a_{m1} & \cdots & a_{mn} \end{bmatrix}$, we define the scalar multiple of the matrix as $k\mathbf{A} = \begin{bmatrix} ka_{11} & \cdots & ka_{1n} \\ \vdots & \cdots & \vdots \\ ka_{m1} & \cdots & ka_{mn} \end{bmatrix}$. In the case of an $m \times 1$ matrix (a column vector) or a $1 \times n$ matrix (a row vector), this is the same as the ordinary scaling of a vector.

Example: $3\begin{bmatrix} 2 & 1 \\ -1 & 5 \end{bmatrix} = \begin{bmatrix} 6 & 3 \\ -3 & 15 \end{bmatrix}$

Definition: Given two $m \times n$ matrices $\mathbf{A} = \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \cdots & \vdots \\ a_{m1} & \cdots & a_{mn} \end{bmatrix}$ and $\mathbf{B} = \begin{bmatrix} b_{11} & \cdots & b_{1n} \\ \vdots & \cdots & \vdots \\ b_{m1} & \cdots & b_{mn} \end{bmatrix}$, we define the sum of these two like matrices by adding their respective entries. That is:

$$\mathbf{A} + \mathbf{B} = \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \cdots & \vdots \\ a_{m1} & \cdots & a_{mn} \end{bmatrix} + \begin{bmatrix} b_{11} & \cdots & b_{1n} \\ \vdots & \cdots & \vdots \\ b_{m1} & \cdots & b_{mn} \end{bmatrix} = \begin{bmatrix} (a_{11} + b_{11}) & \cdots & (a_{1n} + b_{1n}) \\ \vdots & \cdots & \vdots \\ (a_{m1} + b_{m1}) & \cdots & (a_{mn} + b_{mn}) \end{bmatrix}.$$

Example: $3\begin{bmatrix} 2 & 1 \\ -1 & 5 \end{bmatrix} + 5\begin{bmatrix} 1 & 0 \\ 3 & -1 \end{bmatrix} = \begin{bmatrix} 6 & 3 \\ -3 & 15 \end{bmatrix} + \begin{bmatrix} 5 & 0 \\ 15 & -5 \end{bmatrix} = \begin{bmatrix} 11 & 3 \\ 12 & 10 \end{bmatrix}$

Proposition: For any $m \times n$ matrix $\mathbf{A}$, any scalar *k*, and any $1 \times n$ column vector **x**, $\boxed{(k\mathbf{A})\mathbf{x} = k(\mathbf{A}\mathbf{x})}$.

Proof: This is a straightforward calculation. Writing $\mathbf{A} = \begin{bmatrix} \uparrow & & \uparrow \\ \mathbf{v}_1 & \cdots & \mathbf{v}_n \\ \downarrow & & \downarrow \end{bmatrix}$ in terms of its column vectors, we have

$$(k\mathbf{A})\mathbf{x} = \left(k \begin{bmatrix} \uparrow & & \uparrow \\ \mathbf{v}_1 & \cdots & \mathbf{v}_n \\ \downarrow & & \downarrow \end{bmatrix} \right) \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} \uparrow & & \uparrow \\ k\mathbf{v}_1 & \cdots & k\mathbf{v}_n \\ \downarrow & & \downarrow \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} = x_1(k\mathbf{v}_1) + \cdots + x_n(k\mathbf{v}_n) = k(x_1\mathbf{v}_1 + \cdots + x_n\mathbf{v}_n)$$

$$= k\left(\begin{bmatrix} \uparrow & & \uparrow \\ \mathbf{v}_1 & \cdots & \mathbf{v}_n \\ \downarrow & & \downarrow \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} \right) = k(\mathbf{A}\mathbf{x}).$$

Proposition: For any $m \times n$ matrices **A** and **B** and any $1 \times n$ column vector **x**, $\boxed{(\mathbf{A}+\mathbf{B})\mathbf{x} = \mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{x}}$.

Proof: If we write $\mathbf{A} = \begin{bmatrix} \uparrow & & \uparrow \\ \mathbf{v}_1 & \cdots & \mathbf{v}_n \\ \downarrow & & \downarrow \end{bmatrix}$ and $\mathbf{B} = \begin{bmatrix} \uparrow & & \uparrow \\ \mathbf{w}_1 & \cdots & \mathbf{w}_n \\ \downarrow & & \downarrow \end{bmatrix}$, and $\mathbf{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$, then

$$(\mathbf{A}+\mathbf{B})\mathbf{x} = \left(\begin{bmatrix} \uparrow & & \uparrow \\ \mathbf{v}_1 & \cdots & \mathbf{v}_n \\ \downarrow & & \downarrow \end{bmatrix} + \begin{bmatrix} \uparrow & & \uparrow \\ \mathbf{w}_1 & \cdots & \mathbf{w}_n \\ \downarrow & & \downarrow \end{bmatrix} \right) \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} \uparrow & & \uparrow \\ (\mathbf{v}_1 + \mathbf{w}_1) & \cdots & (\mathbf{v}_n + \mathbf{w}_n) \\ \downarrow & & \downarrow \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$$

$$= x_1(\mathbf{v}_1 + \mathbf{w}_1) + \cdots + x_n(\mathbf{v}_n + \mathbf{w}_n) = x_1\mathbf{v}_1 + x_1\mathbf{w}_1 + \cdots + x_n\mathbf{v}_n + x_n\mathbf{w}_n$$

$$= x_1\mathbf{v}_1 + \cdots + x_n\mathbf{v}_n + x_1\mathbf{w}_1 + \cdots + x_n\mathbf{w}_n = \begin{bmatrix} \uparrow & & \uparrow \\ \mathbf{v}_1 & \cdots & \mathbf{v}_n \\ \downarrow & & \downarrow \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} + \begin{bmatrix} \uparrow & & \uparrow \\ \mathbf{w}_1 & \cdots & \mathbf{w}_n \\ \downarrow & & \downarrow \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} = \mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{x}.$$

Matrix products

Though it's possible to take a formulaic approach to the multiplication of matrices, it's much better to think of each matrix as representing a linear transformation and to define matrix product by considering the composition of these linear transformations.

Proposition: Where defined, the composition of linear transformations is a linear transformation.

Proof: Suppose **A** is an $m \times n$ matrix that corresponds to a linear transformation $T_{\mathbf{A}} : \mathbf{R}^n \to \mathbf{R}^m$, i.e. $T_{\mathbf{A}}(\mathbf{y}) = \mathbf{A}\mathbf{y}$. Also, suppose **B** is an $n \times p$ matrix that corresponds to a linear transformation $T_{\mathbf{B}} : \mathbf{R}^p \to \mathbf{R}^n$, i.e. $T_{\mathbf{B}}(\mathbf{x}) = \mathbf{B}\mathbf{x}$. We can then define the composition $T_{\mathbf{A}} \circ T_{\mathbf{B}} : \mathbf{R}^p \to \mathbf{R}^m$ by $(T_{\mathbf{A}} \circ T_{\mathbf{B}})(\mathbf{x}) = T_{\mathbf{A}}(T_{\mathbf{B}}(\mathbf{x}))$. Since both of these functions are linear, for any scalars c_1, c_2 and vectors $\mathbf{v}_1, \mathbf{v}_2 \in \mathbf{R}^p$, we have:

$$(T_{\mathbf{A}} \circ T_{\mathbf{B}})(c_1\mathbf{v}_1 + c_2\mathbf{v}_2) = T_{\mathbf{A}}(T_{\mathbf{B}}(c_1\mathbf{v}_1 + c_2\mathbf{v}_2)) = T_{\mathbf{A}}(c_1 T_{\mathbf{B}}(\mathbf{v}_1) + c_2 T_{\mathbf{B}}(\mathbf{v}_2))$$
$$= c_1 T_{\mathbf{A}}(T_{\mathbf{B}}(\mathbf{v}_1)) + c_2 T_{\mathbf{A}}(T_{\mathbf{B}}(\mathbf{v}_2)) = c_1(T_{\mathbf{A}} \circ T_{\mathbf{B}})\mathbf{v}_1 + c_2(T_{\mathbf{A}} \circ T_{\mathbf{B}})\mathbf{v}_2$$

So $T_{\mathbf{A}} \circ T_{\mathbf{B}} : \mathbf{R}^p \to \mathbf{R}^m$ is also linear and is represented by an $m \times p$ matrix. Call this matrix **AB**.

Corollary (really a restatement of the definition): For any vector $\mathbf{x} \in \mathbf{R}^p$, $\boxed{(\mathbf{AB})\mathbf{x} = \mathbf{A}(\mathbf{B}\mathbf{x})}$.

This statement look very much like an associative law for multiplication, but it's really just the statement that **AB** is defined to be the matrix of the composition.

Calculation of the matrix product

How do we actually calculate the matrix product **AB** (where defined)? Perhaps the simplest way to do this is to recall the meaning of the columns of any matrix. The columns tell us where the corresponding linear function takes the elementary vectors $\{\mathbf{e}_1,\cdots,\mathbf{e}_p\}$, so $\mathbf{AB}=\begin{bmatrix}\uparrow & & \uparrow \\ \mathbf{AB}(\mathbf{e}_1) & \cdots & \mathbf{AB}(\mathbf{e}_p) \\ \downarrow & & \downarrow\end{bmatrix}=\begin{bmatrix}\uparrow & & \uparrow \\ \mathbf{A}(\mathbf{Be}_1) & \cdots & \mathbf{A}(\mathbf{Be}_p) \\ \downarrow & & \downarrow\end{bmatrix}$.

But if we write $\mathbf{B}=\begin{bmatrix}\uparrow & & \uparrow \\ \mathbf{B}(\mathbf{e}_1) & \cdots & \mathbf{B}(\mathbf{e}_p) \\ \downarrow & & \downarrow\end{bmatrix}=\begin{bmatrix}\uparrow & & \uparrow \\ \mathbf{v}_1 & \cdots & \mathbf{v}_p \\ \downarrow & & \downarrow\end{bmatrix}$, we then see that $\mathbf{AB}=\begin{bmatrix}\uparrow & & \uparrow \\ \mathbf{A}(\mathbf{v}_1) & \cdots & \mathbf{A}(\mathbf{v}_p) \\ \downarrow & & \downarrow\end{bmatrix}$.

That is, $\mathbf{AB}=\mathbf{A}\begin{bmatrix}\uparrow & & \uparrow \\ \mathbf{v}_1 & \cdots & \mathbf{v}_p \\ \downarrow & & \downarrow\end{bmatrix}=\begin{bmatrix}\uparrow & & \uparrow \\ \mathbf{Av}_1 & \cdots & \mathbf{Av}_p \\ \downarrow & & \downarrow\end{bmatrix}$.

In other words, the matrix **A** simply individually multiplies each of the column vectors of **B**.

Example: If $\mathbf{A}=\begin{bmatrix}2 & -1 \\ 1 & 5\end{bmatrix}$ and $\mathbf{B}=\begin{bmatrix}1 & 0 & 4 \\ -1 & 2 & 1\end{bmatrix}$, then the product **AB** is defined (though **BA** is not), and

$$\mathbf{AB}=\begin{bmatrix}2 & -1 \\ 1 & 5\end{bmatrix}\begin{bmatrix}1 & 0 & 4 \\ -1 & 2 & 1\end{bmatrix}=\begin{bmatrix}3 & -2 & 7 \\ -4 & 10 & 9\end{bmatrix}.$$

It should be relatively easy to see that each entry is calculated as $(\mathbf{AB})_{ij}=(i\text{th row of }\mathbf{A})\cdot(j\text{th column of }\mathbf{B})$.

This dot product is only defined when the number of columns of **A** matches the number of rows of **B**.

Matrix multiplication (where defined) is not commutative: $\boxed{\mathbf{AB}\neq\mathbf{BA}}$

It's easy to understand why matrix multiplication cannot be commutative even in the case where both products are defined. Matrix product is just the composition of functions, and composing functions in reverse order does not generally give the same functions, i.e. $f\circ g\neq g\circ f$. This is most simply understood by thinking about it in less mathematical terms. For example, if you put on your socks and then put on your shoes, this is clearly different than first putting on your shoes and then putting on your socks. Sometimes you can get the same result, just as it is the case that there are some matrices **A** and **B** such that $\mathbf{AB}=\mathbf{BA}$, but this will not generally be the case.

Matrix multiplication (where defined) is associative: $\boxed{(\mathbf{AB})\mathbf{C}=\mathbf{A}(\mathbf{BC})}$

This follows from the corresponding fact about composition of functions, namely that $(f\circ g)\circ h=f\circ(g\circ h)$.

The Identity matrix acts as a multiplicative identity: For an $m\times n$ matrix **A**, $\boxed{\mathbf{I}_m\mathbf{A}=\mathbf{A}}$ and $\boxed{\mathbf{AI}_n=\mathbf{A}}$.

Though this is easy to see by calculation, it follows from the general fact about functions that $Id\circ f=f$ and $f\circ Id=f$, i.e. for any *x* in the domain of *f*, we have $(Id\circ f)(x)=Id(f(x))=f(x)$ and $(f\circ Id)(x)=f(Id(x))=f(x)$.

Proposition: If **A** is an invertible $n\times n$ matrix with inverse matrix $\mathbf{A}^{-1}$, then $\boxed{\mathbf{A}^{-1}\mathbf{A}=\mathbf{I}}$ and $\boxed{\mathbf{AA}^{-1}=\mathbf{I}}$.

Proof: These follow directly from the fact that matrix product represents the composition of linear functions and the fact that a function composed with its inverse (in either order) yields the identity function.

Proposition: If both **A** and **B** are invertible $n \times n$ matrices, then **AB** is also invertible and $\boxed{(\mathbf{AB})^{-1} = \mathbf{B}^{-1}\mathbf{A}^{-1}}$.

Proof: This also follows directly from the general fact about functions, i.e. $\left(f \circ g\right)^{-1} = g^{-1} \circ f^{-1}$. In nonmathematical terms, if you first put on your socks and then put on your shoes, the inverse of this is to first take off your shoes and then take off your socks.

More easy-to-prove matrix algebra facts: For any scalar k and appropriate sized matrices

$\mathbf{A}(\mathbf{C}+\mathbf{D}) = \mathbf{AC}+\mathbf{AD}$ (left-hand distributive law)
$(\mathbf{A}+\mathbf{B})\mathbf{C} = \mathbf{AC}+\mathbf{BC}$ (right-hand distributive law)
$\left(k\mathbf{A}\right)\mathbf{C} = k\left(\mathbf{AC}\right)$

These and the facts previously stated are not meant to be exhaustive. Except for the fact that matrix multiplication is not commutative, most of the familiar algebraic rules are also true for matrices.

An application to trigonometry: Sum of angle formulas for sine and cosine

We previously showed that counterclockwise rotation in $\mathbf{R}^2$ through an angle θ is a linear transformation represented by the rotation matrix $\mathbf{R}_\theta = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$. It is geometrically clear that the composition of rotation through angle ϕ and rotation through angle θ is just rotation through the angle $(\theta+\phi)$, so $\mathbf{R}_\theta\mathbf{R}_\phi = \mathbf{R}_{\theta+\phi}$. Therefore:

$$\begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}\begin{bmatrix} \cos\phi & -\sin\phi \\ \sin\phi & \cos\phi \end{bmatrix} = \begin{bmatrix} \cos\theta\cos\phi - \sin\theta\sin\phi & -(\sin\theta\cos\phi + \cos\theta\sin\phi) \\ \sin\theta\cos\phi + \cos\theta\sin\phi & \cos\theta\cos\phi - \sin\theta\sin\phi \end{bmatrix}$$

$$= \begin{bmatrix} \cos(\theta+\phi) & -\sin(\theta+\phi) \\ \sin(\theta+\phi) & \cos(\theta+\phi) \end{bmatrix} \quad \Rightarrow \quad \boxed{\begin{array}{l} \cos(\theta+\phi) = \cos\theta\cos\phi - \sin\theta\sin\phi \\ \sin(\theta+\phi) = \sin\theta\cos\phi + \cos\theta\sin\phi \end{array}}$$

Notes by Robert Winters