Week 3 – Incomplete! Will update soon.

- A function $T : \mathbb{R}^k \rightarrow \mathbb{R}^n$ is called **injective**, or **one-to-one**, if each input gets a unique output. The rigorous definition is: for all $\vec{x}, \vec{y} \in \mathbb{R}^k$, we have that $[T(\vec{x}) = T(\vec{y})]$ implies $[\vec{x} = \vec{y}]$. In other words, the only time the outputs are the same is when in fact the inputs are the same as well.

- A function $T : \mathbb{R}^k \rightarrow \mathbb{R}^n$ is called **surjective**, or **onto**, if everything in the target space gets hit by $T$. The rigorous definition is: for all $\vec{b} \in \mathbb{R}^n$, there exists $\vec{x} \in \mathbb{R}^k$ such that $T(\vec{x}) = \vec{b}$. In other words, given anything in the co-domain/target space, we can find a vector in the domain that $T$ sends to that target vector.

- If $T : \mathbb{R}^k \rightarrow \mathbb{R}^n$ is linear, with matrix representation, say, $A$, then $T$ is 1-1 (short-hand for “one-to-one”) iff $\text{rref}(A)$ has a pivot in every column. If $\text{rref}(A)$ has a column without a pivot, there will be a free variable in the linear system $A\vec{x} = \vec{0}$, so $T(\vec{x}) = \vec{0}$ will have more than one solution, so $T$ won’t be 1-1. Conversely, if $\text{rref}(A)$ has a pivot in every column, then $T(\vec{x}) = T(\vec{y})$ means $A\vec{x} = A\vec{y}$, so $A(\vec{x} - \vec{y}) = 0$, and since $\text{rref}(A)$ has a pivot in each column, the only solution to $A\vec{u} = \vec{0}$ is $\vec{u} = \vec{0}$, which forces $\vec{x} - \vec{y} = \vec{0}$, and thus $\vec{x} = \vec{y}$, showing that $T$ is injective.

- For $T$ as above, $T$ is onto iff $\text{rref}(A)$ has a pivot in every row. If $\text{rref}(A)$ had a row without a pivot, then it’d have at least one row of zeros on the bottom. By setting the right hand side of constants of $\vec{c}_n$, and undoing each elementary row op taken to get $A$ into $\text{rref}(A)$, we can get a system $A\vec{x} = \vec{b}$ that’s inconsistent; in other words, $T(\vec{x}) = \vec{b}$ has no solutions $\vec{x}$ for our specially chosen $\vec{b}$, so $T$ is not onto. Conversely, if $\text{rref}(A)$ has a pivot in every row, then we will always be able to solve $A\vec{x} = \vec{b}$ for any $\vec{b} \in \mathbb{R}^n$, so $T$ is onto.

- We identify a matrix with its associated linear transform when we say that some matrix is onto or 1-1.

- some examples: $(\begin{smallmatrix} 1 & 0 & 0 \\ 0 & 1 & 2 \end{smallmatrix})$ is onto but not 1-1 since its $\text{rref}$ (which is conveniently itself...) has a pivot in every row, but not in every column. $(\begin{smallmatrix} 1 & 0 \\ 0 & 0 \end{smallmatrix})$ is 1-1 but not onto since it has a pivot in every column, but not in every row. $(\begin{smallmatrix} 1 & 1 & 0 \\ 0 & 0 & 0 \end{smallmatrix})$ is neither injective nor surjective, and
\[
\begin{pmatrix}
1 & 0 & 0 \\
0 & 1 & 0 \\
0 & 0 & 1
\end{pmatrix}
\] is both.

- If \(n \times k\) matrix \(A\) is 1-1, then it must be square or tall \((n \geq k)\), since each pivot gets a unique row. If \(A\) is onto, then it must be square or wide \((n \leq k)\), since each pivot gets a unique column.

- We say \(T : \mathbb{R}^k \rightarrow \mathbb{R}^n\) is bijective if \(T\) is both 1-1 and onto. In this case, by the above characterizations of onto and 1-1, if \(T\) is bijective, then its matrix representation must be square (combine the two inequalities \(n \leq k\) and \(n \geq k\) to get this). In fact, \(T\) is bijective iff its matrix representation \(A\) satisfies \(\text{rref}(A) = I_n\), since \(I_n\) is the only \(n \times n\) matrix in \(\text{rref}\) with a pivot in every row and column. We also say that a matrix \(A\) is bijective if its associated linear transform \((T(\vec{x}) = A\vec{x})\) is bijective.

- If matrix \(A\) is \(n \times n\) square and surjective, then its \(\text{rref}\) has \(n\) pivots (one for each row), which means also that \(\text{rref}(A)\)'s columns all have pivots as well, so \(A\) is injective, so \(A\) is bijective. Similarly, if \(A\) is \(n \times n\) square and injective, then its \(\text{rref}\) has \(n\) pivots (one for each column), which means also that \(\text{rref}(A)\)'s rows all have pivots as well, so \(A\) is surjective, so \(A\) is bijective.

- Bijective functions are special because they have inverses.

- We say \(T : \mathbb{R}^k \rightarrow \mathbb{R}^n\) has an inverse, which we denote by \(T^{-1}\), if \(T^{-1} : \mathbb{R}^n \rightarrow \mathbb{R}^k\), for all \(\vec{x} \in \mathbb{R}^k\) we have \(T^{-1}(T(\vec{x})) = \vec{x}\), and for all \(\vec{y} \in \mathbb{R}^n\) we have \(T(T^{-1}(\vec{y})) = \vec{y}\). This means that inverse functions switch inputs and outputs.

- The first condition requires \(T\) to be injective and \(T^{-1}\) to be surjective; the second condition requires \(T\) to be surjective and \(T^{-1}\) to be injective. So every invertible function is also bijective, and so is its inverse.

- Conversely, every bijective function has an inverse – namely, the function that takes elements of the target space back to the unique elements of the domain that the original function sends to those targets. For such a function to be defined, the original function must be onto (or else the inverse has some elements it can’t send anywhere) and 1-1 (or else the inverse has to send elements to multiple places).

- Long story short, a function is bijective iff it is invertible.

- Inverses are unique since where they send each element is determined entirely by the original function, and the inverse of the inverse is the original function. Both of these facts are consequences of the definition.

- If \(T\) is bijective, then we know its matrix representation is square. So it must be the case that \(n = k\), so \(T\) has the same domain and target space.
- if \( T: \mathbb{R}^k \rightarrow \mathbb{R}^n \) is invertible and linear, then \( T^{-1} \) is also linear, since for all \( \vec{x}, \vec{y} \in \mathbb{R}^n \) and for all \( c_1, c_2 \in \mathbb{R} \) we have \( T^{-1}(c_1 \vec{x} + c_2 \vec{y}) = T^{-1}(c_1 T(T^{-1}(\vec{x})) + c_2 T(T^{-1}(\vec{y}))) = T^{-1}(c_1 T^{-1}(\vec{x}) + c_2 T^{-1}(\vec{y})) = c_1 T^{-1}(\vec{x}) + c_2 T^{-1}(\vec{y}) \).

- We can treat matrices as linear transforms. Matrix \( A \) is invertible iff \( \text{tr}ix \) representation of \( T \) is invertible iff \( A \) is bijective iff \( \text{refr} \) of \( A \) is \( I_n \), and in this case, writing \( A^{-1} \) as the matrix representation of \( T^{-1} \), for all \( \vec{x} \in \mathbb{R}^n \) we have \( T^{-1}(T(\vec{x})) = A^{-1}A\vec{x} = \vec{x} \) and also \( T(T^{-1}(\vec{x})) = AA^{-1}\vec{x} = \vec{x} \) (since \( A \) is square, its domain and target space are identical, so we need not use two different variables).

- But in the above point, if we let \( \vec{x} = \vec{e}_i \) as \( i \) ranges from 1 to \( n \), this means that the \( i^{th} \) column of \( A^{-1}A \) is equal to \( \vec{e}_i \), and same for the \( i^{th} \) column of \( AA^{-1} \). So we may combine these columns to say \( AA^{-1}[\vec{e}_1 \cdots \vec{e}_n] = [\vec{e}_1 \cdots \vec{e}_n] \) and \( A^{-1}A[\vec{e}_1 \cdots \vec{e}_n] = [\vec{e}_1 \cdots \vec{e}_n] \), or simply: \( AA^{-1} = I_n \) and \( A^{-1}A = I_n \).

- Matrix inverses are unique because the inverses of the linear transforms they represent are unique. But we may easily double-check this by the following computation: if \( A \) is invertible, and \( B \) and \( C \) are two inverses of \( A \), then \( AB = I_n = AC \), so \( BAB = BAC \), and since \( BA = I_n \), we get that \( I_nB = I_nC \), or simply that \( B = C \), so in fact there was really only one inverse all along.

- If \( A \) and \( B \) are both \( n \times n \) square matrices and \( AB = I_n \), then \( B \) is 1-1: \( B\vec{x} = B\vec{y} \) implies \( AB\vec{x} = AB\vec{y} \), which implies \( I_n\vec{x} = I_n\vec{y} \), so \( \vec{x} = \vec{y} \). Also, \( A \) is onto: if \( \vec{y} \in \mathbb{R}^n \), then \( A(B\vec{y}) = (AB)\vec{y} = I_n\vec{y} = \vec{y} \). Since \( A \) and \( B \) are both square, we know by a comment above that both \( A \) and \( B \) are invertible. Then \( A = AB^{-1} = I_nB^{-1} = B^{-1} \) and \( B = A^{-1}AB = A^{-1}I_n = A^{-1} \), so \( A \) and \( B \) are inverses of each other. So two square matrices multiplying to the identity always means each of those matrices is invertible, and they are each other’s inverses!!

- If matrix \( A \) is \( n \times n \) invertible, then we already know that \( A \) is 1-1 and onto, so \( A\vec{x} = \vec{b} \) has exactly one solution for each \( \vec{b} \in \mathbb{R}^n \). We can get this solution by hitting both sides with \( A^{-1}: A\vec{x} = \vec{b} \implies A^{-1}A\vec{x} = A^{-1}\vec{b} \implies I_n\vec{x} = A^{-1}\vec{b} \implies \vec{x} = A^{-1}\vec{b} \), and we may check this solution by computing \( A(A^{-1}\vec{b}) = I_n\vec{b} = \vec{b} \).

- In order to find the inverse of an invertible matrix, we need to have a short digression on elementary matrices. This digression will also benefit us later when we discuss how to find a basis for the image of a matrix! Elementary matrices are best understood via example.
In the examples that follow, pay attention to the effect the left matrix has on the right matrix it’s multiplying!

\[
\begin{bmatrix}
1 & 0 & 0 \\
0 & 5 & 0 \\
0 & 0 & 1
\end{bmatrix}
\begin{bmatrix}
1 & 2 & 3 \\
4 & 5 & 6 \\
7 & 8 & 9
\end{bmatrix}
= 
\begin{bmatrix}
1 & 2 & 3 \\
20 & 25 & 30 \\
7 & 8 & 9
\end{bmatrix}
\]
scales second row by factor of 5

\[
\begin{bmatrix}
1 & 0 & 0 \\
0 & 0 & 1 \\
0 & 1 & 0
\end{bmatrix}
\begin{bmatrix}
1 & 2 & 3 \\
4 & 5 & 6 \\
7 & 8 & 9
\end{bmatrix}
= 
\begin{bmatrix}
1 & 2 & 3 \\
7 & 8 & 9 \\
4 & 5 & 6
\end{bmatrix}
\]
switches second and third rows

\[
\begin{bmatrix}
1 & 0 & 0 \\
0 & 1 & 0 \\
2 & 0 & 1
\end{bmatrix}
\begin{bmatrix}
1 & 2 & 3 \\
4 & 5 & 6 \\
7 & 8 & 9
\end{bmatrix}
= 
\begin{bmatrix}
1 & 2 & 3 \\
4 & 5 & 6 \\
9 & 12 & 15
\end{bmatrix}
\]
adds twice first row to third row

The matrices on the left are elementary matrices. An elementary matrix is just any matrix that results from performing one elementary row operation on the identity matrix. In the first case, \[
\begin{bmatrix}
1 & 0 & 0 \\
0 & 5 & 0 \\
0 & 0 & 1
\end{bmatrix}
\]
was obtained by multiplying the 2nd row of \(I_3\) by 5. \[
\begin{bmatrix}
1 & 0 & 0 \\
0 & 0 & 1 \\
0 & 1 & 0
\end{bmatrix}
\]
was obtained by switching the 2nd and 3rd rows of \(I_3\). \[
\begin{bmatrix}
1 & 0 & 0 \\
0 & 0 & 1 \\
2 & 0 & 1
\end{bmatrix}
\]
was obtained by adding twice the 1st row of \(I_3\) to its 3rd row. Now we see what’s going on – *an elementary matrix performs the SAME ROW OP on matrices it left-multiplies as the row op that was performed on the identity to obtain IT* (woahhhhhhh). This means these elementary matrices are little tools whose job is to perform one specific row op on matrices. The “input” matrices need not be square; they just have to have the same number of rows as the elementary matrix has columns.

Note that elementary matrices are always invertible – a consequence of the three types of elementary row ops being invertible. Specifically, \[
\begin{bmatrix}
1 & 0 & 0 \\
0 & 5 & 0 \\
0 & 0 & 1
\end{bmatrix}^{-1} = \begin{bmatrix}1 & 0 & 0 \\
0 & 1/5 & 0 \\
0 & 0 & 1
\end{bmatrix}, \quad \begin{bmatrix}1 & 0 & 0 \\
0 & 0 & 1 \\
0 & 1/0 & 0
\end{bmatrix}^{-1} = \begin{bmatrix}1 & 0 & 0 \\
0 & 0 & 1 \\
0 & 1 & 0
\end{bmatrix},
\]
and \[
\begin{bmatrix}
1 & 0 & 0 \\
0 & 0 & 1 \\
2 & 0 & 1
\end{bmatrix}^{-1} = \begin{bmatrix}1 & 0 & 0 \\
0 & 0 & 1 \\
-2 & 0 & 1
\end{bmatrix}.
\]

- For any matrix \(A\), since \(\text{rref}(A)\) is obtained through elementary row ops performed on \(A\), this means there is some sequence of elementary matrices that left-multiply \(A\) to get \(\text{rref}(A)\). In other words, we may write \(\text{rref}(A) = E_p \cdots E_1 A\) for some elementary matrices \(E_1, \ldots, E_p\).

- So, \(A\) is invertible \(\implies\) \(\text{rref}(A) = I_n \implies E_p \cdots E_1 A\) (for some elementary matrices \(E_1, \ldots, E_p\)). Then defining \(B = E_p \cdots E_1\), we get \(BA = I_n\), and since \(A\) and \(B\) are both square, this implies that \(B = A^{-1}\), by a comment above. So

\[
A^{-1} = B = E_p \cdots E_1 = E_p \cdots E_1 I_n.
\]

The last expression is the result of performing the same row ops on \(I_n\) that were used to get \(A\) into \(\text{rref}\). So the last string of equations means that we can always find the inverse of an invertible matrix by applying the same row ops on the identity as those that get \(A\) to \(\text{rref}(A)\).
SO – to find the inverse of a matrix $A$, you can do row reduction on $[A | I_n]$ until you get the identity on the left (if you don’t get the identity, then $\text{rref}(A) \neq I_n$, so $A$ wasn’t actually invertible), and then what remains on the right will be $A^{-1}$, since this is the result of performing the same row ops on the identity that were used to get $A$ into $\text{rref}(A)$.

Now we’ll discuss the image and kernel of a matrix. You can think of these (roughly) as measures of injectivity and surjectivity of a matrix – the larger the image is, the “closer” the matrix is to being surjective; the smaller the kernel is, the “closer” the matrix is to being injective. Following are the rigorous definitions of these terms.

- The image $\text{im}(A)$ of an $n \times k$ matrix $A$ is: $\text{im}(A) = \{ A\vec{x} : \vec{x} \in \mathbb{R}^k \}$. In other words, the image of a matrix is the set of all outputs of that matrix.

- The kernel $\text{ker}(A)$ of an $n \times k$ matrix $A$ is: $\text{ker}(A) = \{ \vec{x} \in \mathbb{R}^k : A\vec{x} = \vec{0} \}$. In other words, the kernel of a matrix is the set of all inputs that that matrix sends to the zero vector.

- VERY IMPORTANT: the image is a subset of the target space, whereas the kernel is a subset of the domain.

- Note that, by the definitions, the zero vectors (in $\mathbb{R}^n$ and $\mathbb{R}^k$, respectively) lie in $\text{im}(A)$ and $\text{ker}(A)$, since $A\vec{0} = \vec{0}$.

- If $\text{im}(A) = \mathbb{R}^n$, then that means (by definitions) $A$ is onto. If $\text{im}(A) = \{ \vec{0} \}$, then $A$ sends everything to the zero vector, so $A\vec{e}_i = \vec{0}$ for each applicable $i$, which means the $i^{th}$ column of $A$ is the zero vector for each $i$, which means $A$ is the zero matrix.

- If $\text{ker}(A) = \{ \vec{0} \}$, then the only input that gets sent to $\vec{0}$ is $\vec{0}$ (if $A$ is not square, then these zero vectors will of course have different size). So then $A\vec{x} = A\vec{y} \implies A\vec{x} - A\vec{y} = \vec{0} \implies A(\vec{x} - \vec{y}) = \vec{0} \implies \vec{x} - \vec{y} = \vec{0} \implies \vec{x} = \vec{y}$, so $A$ is 1-1. If $\text{ker}(A) = \mathbb{R}^k$, then $A$ sends everything to the zero vector, so again $A$ is the zero matrix.

- In the case that $\text{ker}(A) = \{ \vec{0} \}$, we say that $A$’s kernel is trivial, or that the kernel of $A$ is trivial.

- If $A$ is $n \times k$ and $\text{ker}(A)$ is trivial, then $A$ is 1-1, and so $\text{rref}(A)$ has a pivot in each column, which means that $n \geq k$ (since each pivot gets a unique row).

- If $A$ is $n \times k$ and $\text{im}(A)$ is all of $\mathbb{R}^n$, then $A$ is onto, so $\text{rref}(A)$ has a pivot in every row, which means that $n \leq k$, since each pivot gets its own column.

- To find the kernel of a matrix, simply solve for and parametrize the solutions to the system $A\vec{x} = \vec{0}$, usually done via Gaussian elimination.

- To find the image of a matrix, look at $A$’s columns, and take the “span” of those columns:
- If \( v_1, \ldots, v_q \in \mathbb{R}^k \), then the span of these vectors is defined as the set of all possible linear combinations of these vectors: \( \text{span}(\vec{v}_1, \ldots, \vec{v}_q) = \{a_1 \vec{v}_1 + \cdots + a_q \vec{v}_q : a_i \in \mathbb{R} \text{ for all } i\} \).

- The image of \( A \) is then just the span of its columns: say \( A = [\vec{v}_1 \cdots \vec{v}_q] \). Then \( \text{im}(A) = \{A\vec{x} : \vec{x} \in \mathbb{R}^k \} = \{x_1 \vec{v}_1 + \cdots + x_q \vec{v}_q : x_i \in \mathbb{R} \text{ for all } i\} = \text{span}(\vec{v}_1, \ldots, \vec{v}_q) \).

in progress- will complete these notes in a few days. To study for the midterm, also know: rank, nullity, rank-nullity theorem, subspaces, basis, dimension, how to find bases for the subspaces \( \text{im}(A) \) and \( \text{ker}(A) \), and linear independence/dependence. All these concepts are discussed in the textbook, in chapter 3. Solve all homework problems and as many problems nearby those as possible in order to cement understanding of these concepts! Good luck :)