## Linear Algebra Fundamentals

It can be argued that all of linear algebra can be understood using the four fundamental subspaces associated with a matrix. Because they form the foundation on which we later work, we want an explicit method for analyzing these subspaces- That method will be the Singular Value Decomposition (SVD). It is unfortunate that most first courses in linear algebra do not cover this material, so we do it here. Again, we cannot stress the importance of this decomposition enough- We will apply this technique throughout the rest of this text.

### 0.1 Representation, Basis and Dimension

Let us quickly review some notation and basic ideas from linear algebra:
Suppose that the matrix $V$ is composed of the columns $\boldsymbol{v}_{1}, \ldots, \boldsymbol{v}_{k}$, and that these columns form a basis basis for some subspace, $H$, in $\mathbb{R}^{n}$ (notice that this implies $k \leq n$ ). Then every data point in $H$ can be written as a linear combination of the basis vectors. In particular, if $\boldsymbol{x} \in H$, then we can write:

$$
\boldsymbol{x}=c_{1} \boldsymbol{v}_{1}+\ldots+c_{k} \boldsymbol{v}_{k} \doteq V \boldsymbol{c}
$$

so that every data point in our subset of $\mathbb{R}^{n}$ is identified with a point in $\mathbb{R}^{k}$ :

$$
\boldsymbol{x}=\left[\begin{array}{r}
x_{1} \\
x_{2} \\
\vdots \\
x_{n}
\end{array}\right] \longleftrightarrow\left[\begin{array}{r}
c_{1} \\
\vdots \\
c_{k}
\end{array}\right]=\boldsymbol{c}
$$

The vector $\boldsymbol{c}$, which contains the coordinates of $\boldsymbol{x}$, is the low dimensional representation of the point $\boldsymbol{x}$. That is, the data point $\boldsymbol{x}$ resides in $\mathbb{R}^{n}$, but $\boldsymbol{c}$ is in $\mathbb{R}^{k}$, where $k \leq n$.

Furthermore, we would say that the subspace $H$ (a subspace of $\mathbb{R}^{n}$ ) is isomorphic to $\mathbb{R}^{k}$. We'll recall the definition:

Definition 0.1.1. Any one-to-one (and onto) linear map is called an isomorphism. In particular, any change of coordinates is an isomorphism. Spaces that
are isomorphic have essentially the same algebraic structure- adding vectors in one space is corresponds to adding vectors in the second space, and scalar multiplication in one space is the same as scalar multiplication in the second.

Definition 0.1.2. Let $H$ be a subspace of vector space $X$. Then $H$ has dimension $k$ if a basis for $H$ requires $k$ vectors.

Given a linearly independent spanning set (the columns of $V$ ) to compute the coordinates of a data point with respect to that basis requires a matrix inversion (or more generally, Gaussian elimination) to solve the equation:

$$
\boldsymbol{x}=V \boldsymbol{c}
$$

In the case where we have $n$ basis vectors of $\mathbb{R}^{n}$, then $V$ is an invertible matrix, and we write:

$$
\boldsymbol{c}=V^{-1} \boldsymbol{x}
$$

If we have fewer than $n$ basis vectors, $V$ will not be square, and thus not invertible in the usual sense. However, if $\boldsymbol{x}$ is contained in the span of the basis, then we will be able to solve for the coordinates of $\boldsymbol{x}$.

Example 0.1.1. Let the subspace $H$ be formed by the span of the vectors $\boldsymbol{v}_{1}, \boldsymbol{v}_{2}$ given below. Given the point $\boldsymbol{x}_{1}, \boldsymbol{x}_{2}$ below, find which one belongs to $H$, and if it does, give its coordinates.

$$
\boldsymbol{v}_{1}=\left[\begin{array}{r}
1 \\
2 \\
-1
\end{array}\right] \quad \boldsymbol{v}_{2}=\left[\begin{array}{r}
2 \\
-1 \\
1
\end{array}\right] \quad \boldsymbol{x}_{1}=\left[\begin{array}{l}
7 \\
4 \\
0
\end{array}\right] \quad \boldsymbol{x}_{2}=\left[\begin{array}{r}
4 \\
3 \\
-1
\end{array}\right]
$$

SOLUTION: Rather than row-reduce twice, we'll do it once on the augmented matrix below.

$$
\left[\begin{array}{rr|rr}
1 & 2 & 7 & 4 \\
2 & -1 & 4 & 3 \\
-1 & 1 & 0 & -1
\end{array}\right] \rightarrow\left[\begin{array}{ll|ll}
1 & 0 & 3 & 2 \\
0 & 1 & 2 & 1 \\
0 & 0 & 1 & 0
\end{array}\right]
$$

How should this be interpreted? The second vector, $\boldsymbol{x}_{2}$ is in $H$, as it can be expressed as $2 \boldsymbol{v}_{1}+\boldsymbol{v}_{2}$. Its low dimensional representation (its coordinate vector) is thus $[2,1]^{T}$.

The first vector, $\boldsymbol{x}_{1}$, cannot be expressed as a linear combination of $\boldsymbol{v}_{1}$ and $\boldsymbol{v}_{2}$, so it does not belong to $H$.

If the basis is orthonormal, we do not need to perform any row reduction. Let us recall a few more definitions:

Definition 0.1.3. A real $n \times n$ matrix $\mathbb{Q}$ is said to be orthogonal if

$$
\mathbb{Q}^{T} \mathbb{Q}=I
$$

This is the property that makes an orthonormal basis nice to work with- it's inverse is its transpose. Thus, it is easy to compute the coordinates of a vector $\boldsymbol{x}$ with respect to this basis. That is, suppose that

$$
\boldsymbol{x}=c_{1} \boldsymbol{u}_{1}+\ldots+c_{k} \boldsymbol{u}_{k}
$$

Then the coordinate $c_{j}$ is just a dot product:

$$
\boldsymbol{x} \cdot \boldsymbol{u}_{j}=0+\ldots+0+c_{j} \boldsymbol{u}_{j} \cdot \boldsymbol{u}_{j}+0+\ldots 0 \Rightarrow c_{j}=\boldsymbol{x} \cdot \boldsymbol{u}_{j}
$$

We can also interpret each individual coordinate as the projection of $\boldsymbol{x}$ onto the appropriate basis vector. Recall that the orthogonal projection of $\boldsymbol{x}$ onto a vector $\boldsymbol{u}$ is the following:

$$
\operatorname{Proj}_{\boldsymbol{u}}(\boldsymbol{x})=\frac{\boldsymbol{u} \cdot \boldsymbol{x}}{\boldsymbol{u} \cdot \boldsymbol{u}} \boldsymbol{u}
$$

If $\mathbf{u}$ is unit length, the denominator is 1 and we have:

$$
\operatorname{Proj} \boldsymbol{u}(\boldsymbol{x})=\left(\mathbf{u}^{T} \mathbf{x}\right) \mathbf{u}=\left(\mathbf{u} \mathbf{u}^{T}\right) \mathbf{x}=\mathbf{u}\left(\mathbf{u}^{T} \mathbf{x}\right)
$$

Writing the coefficients in matrix form, with the columns of $U$ being the orthonormal vectors forming the basis, we have:

$$
\boldsymbol{c}=[\mathbf{x}]_{U}=U^{T} \boldsymbol{x}
$$

Additionally, the projection of $\boldsymbol{x}$ onto the subspace spanned by the (orthonormal) columns of a matrix $U$ is:

$$
\begin{equation*}
\operatorname{Proj}_{U}(\boldsymbol{x})=U \boldsymbol{c}=U U^{T} \boldsymbol{x} \tag{1}
\end{equation*}
$$

Example 0.1.2. We'll change our previous example slightly so that $\boldsymbol{u}_{1}$ and $\boldsymbol{u}_{2}$ are orthonormal. Find the coordinates of $\mathbf{x}_{1}$ with respect to this basis.

$$
\boldsymbol{u}_{1}=\frac{1}{\sqrt{5}}\left[\begin{array}{l}
1 \\
2 \\
0
\end{array}\right] \quad \boldsymbol{u}_{2}=\frac{1}{\sqrt{6}}\left[\begin{array}{r}
2 \\
-1 \\
1
\end{array}\right] \quad \boldsymbol{x}_{1}=\left[\begin{array}{r}
-1 \\
0 \\
2
\end{array}\right]
$$

SOLUTION:

$$
\boldsymbol{c}=U^{T} \boldsymbol{x} \quad \Rightarrow \quad \boldsymbol{c}=\left[\begin{array}{rrr}
1 / \sqrt{5} & 2 / \sqrt{5} & 0 \\
2 / \sqrt{6} & -1 / \sqrt{6} & 1 / \sqrt{6}
\end{array}\right]\left[\begin{array}{r}
-1 \\
8 \\
-2
\end{array}\right]=\left[\begin{array}{r}
3 \sqrt{5} \\
-2 \sqrt{6}
\end{array}\right]
$$

The reader should verify that this is accurate.
We summarize our discussion with the following theorem:
Representation Theorem. Suppose $H$ is a subspace of $\mathbb{R}^{n}$ with orthonormal basis vectors given by the $k$ columns of a matrix $U$ (so $U$ is $n \times k$ ). Then, given $\boldsymbol{x} \in H$,

- The low dimensional representation of $\boldsymbol{x}$ with respect to $U$ is the vector of coordinates, $\boldsymbol{c} \in \mathbb{R}^{k}$ :

$$
\boldsymbol{c}=U^{T} \boldsymbol{x}
$$

- The reconstruction of $\boldsymbol{x}$ as a vector in $\mathbb{R}^{n}$ is:

$$
\hat{\boldsymbol{x}}=U U^{T} \boldsymbol{x}
$$

where, if the subspace formed by $U$ contains $\boldsymbol{x}$, then $\boldsymbol{x}=\hat{\boldsymbol{x}}$ - Notice in this case, the projection of $\boldsymbol{x}$ into the columnspace of $U$ is the same as $\boldsymbol{x}$.

This last point may seem trivial since we started by saying that $\boldsymbol{x} \in U$, however, soon we'll be loosening that requirement.

Example 0.1.3. Let $\boldsymbol{x}=[3,2,3]^{T}$ and let the basis vectors be $\boldsymbol{u}_{1}=\frac{1}{\sqrt{2}}[1,0,1]^{T}$ and let $\boldsymbol{u}_{2}=[0,1,0]^{T}$. Compute the low dimensional representation of $\boldsymbol{x}$, and its reconstruction (to verify that $\boldsymbol{x}$ is in the right subspace).

SOLUTION: The low dimensional representation is given by:

$$
\boldsymbol{c}=U^{T} \boldsymbol{x}=\left[\begin{array}{ccc}
1 / \sqrt{2} & 0 & 1 / \sqrt{2} \\
0 & 1 & 0
\end{array}\right]\left[\begin{array}{l}
3 \\
2 \\
3
\end{array}\right]=\left[\begin{array}{c}
3 \sqrt{2} \\
2
\end{array}\right]
$$

And the reconstruction (verify the arithmetic) is:

$$
\hat{\boldsymbol{x}}=U U^{T} \mathbf{x}=\left[\begin{array}{rrr}
1 / 2 & 0 & 1 / 2 \\
0 & 1 & 0 \\
1 / 2 & 0 & 1 / 2
\end{array}\right]\left[\begin{array}{l}
3 \\
2 \\
3
\end{array}\right]=\left[\begin{array}{l}
3 \\
2 \\
3
\end{array}\right]
$$

For future reference, you might notice that $U U^{T}$ is not the identity, but $U^{T} U$ is the $2 \times 2$ identity:

$$
U^{T} U=\left[\begin{array}{ccc}
1 / \sqrt{2} & 0 & 1 / \sqrt{2} \\
0 & 1 & 0
\end{array}\right]\left[\begin{array}{rr}
1 / \sqrt{2} & 0 \\
0 & 1 \\
1 / \sqrt{2} & 0
\end{array}\right]=\left[\begin{array}{ll}
1 & 0 \\
0 & 1
\end{array}\right]
$$

Projections are important part of our work in modeling data- so much so that we'll spend a bit of time formalizing the ideas in the next section.

### 0.2 Special Mappings: The Projectors

In the previous section, we looked at projecting one vector onto a subspace by using Equation 1. In this section, we think about the projection as a function whose domain and range will be subspaces of $\mathbb{R}^{n}$.

The defining equation for such a function comes from the idea that if one projects a vector, then projecting it again will leave it unchanged.


Figure 1: Projections $P_{1}$ and $P_{2}$ in the first and second graphs (respectively). Asterisks denote the original data point, and circles represent their destination, the projection of the asterisk onto the vector $[1,1]^{T}$. The line segment follows the direction $P \boldsymbol{x}-\boldsymbol{x}$. Note that $P_{1}$ does not project in an orthogonal fashion, while the second matrix $P_{2}$ does.

Definition 0.2.1. A Projector is a square matrix $\mathbb{P}$ so that:

$$
\mathbb{P}^{2}=\mathbb{P}
$$

In particular, $\mathbb{P} \boldsymbol{x}$ is the projection of $\boldsymbol{x}$.
Example 0.2.1. The following are two projectors. Their matrix representations are given by:

$$
P_{1}=\left[\begin{array}{ll}
1 & 0 \\
1 & 0
\end{array}\right] \quad P_{2}=\frac{1}{2}\left[\begin{array}{ll}
1 & 1 \\
1 & 1
\end{array}\right]
$$

Some samples of the projections are given in Figure 1, where we see that both project to the subspace spanned by $[1,1]^{T}$.

Let's consider the action of these matrices on an arbitrary point:

$$
\begin{gathered}
P_{1} \boldsymbol{x}=\left[\begin{array}{ll}
1 & 0 \\
1 & 0
\end{array}\right]\left[\begin{array}{l}
x \\
y
\end{array}\right]=\left[\begin{array}{l}
x \\
x
\end{array}\right], P_{1}\left(P_{1} \boldsymbol{x}\right)=\left[\begin{array}{ll}
1 & 0 \\
1 & 0
\end{array}\right]\left[\begin{array}{l}
x \\
x
\end{array}\right]=\left[\begin{array}{l}
x \\
x
\end{array}\right] \\
P_{2} \boldsymbol{x}=\frac{1}{2}\left[\begin{array}{ll}
1 & 1 \\
1 & 1
\end{array}\right]\left[\begin{array}{l}
x \\
y
\end{array}\right]=\left[\begin{array}{l}
\frac{x+y}{2} \\
\frac{x+y}{2}
\end{array}\right]=\frac{x+y}{2}\left[\begin{array}{l}
1 \\
1
\end{array}\right]
\end{gathered}
$$

You should verify that $P_{2}^{2} \boldsymbol{x}=P_{2}\left(P_{2}(\boldsymbol{x})\right)=\boldsymbol{x}$.
You can deduce along which direction a point is projected by drawing a straight line from the point $\boldsymbol{x}$ to the point $\mathbb{P} \boldsymbol{x}$. In general, this direction will depend on the point. We denote this direction by the vector $\mathbb{P} \boldsymbol{x}-\boldsymbol{x}$.

From the previous examples, we see that $\mathbb{P} \boldsymbol{x}-\boldsymbol{x}$ is given by:

$$
P_{1} \boldsymbol{x}-\boldsymbol{x}=\left[\begin{array}{c}
0 \\
x-y
\end{array}\right], \text { and } P_{2} \boldsymbol{x}-\boldsymbol{x}=\left[\begin{array}{c}
\frac{-x+y}{2} \\
\frac{x-y}{2}
\end{array}\right]=\frac{x-y}{2}\left[\begin{array}{c}
-1 \\
1
\end{array}\right]
$$

You'll notice that in the case of $P_{2}, P_{2} \boldsymbol{x}-\boldsymbol{x}=\left(P_{2}-I\right) \boldsymbol{x}$ is orthogonal to $P_{2} \boldsymbol{x}$.

Definition 0.2.2. $\mathbb{P}$ is said to be an orthogonal projector if it is a projector, and the range of $\mathbb{P}$ is orthogonal to the range of $(I-\mathbb{P})$. We can show orthogonality by taking an arbitrary point in the range, $\mathbb{P} \boldsymbol{x}$ and an arbitrary point in $(I-\mathbb{P})$, $(I-\mathbb{P}) \boldsymbol{y}$, and show the dot product is 0 .

There is a property of real projectors that make them nice to work with: They are also symmetric matrices:

Theorem 0.2.1. The (real) projector $\mathbb{P}$ is an orthogonal projector iff $\mathbb{P}=\mathbb{P}^{T}$. For a proof, see for example, [?].

Caution: An orthogonal projector need not be an orthogonal matrix. Notice that the projector $P_{2}$ from Figure 1 was not an orthogonal matrix (that is, $\left.P_{2} P_{2}^{T} \neq I\right)$.

We have two primary sources for projectors:
Projecting to a vector: Let $\boldsymbol{a}$ be an arbitrary, real, non-zero vector. We show that

$$
\mathbb{P}_{\boldsymbol{a}}=\frac{\boldsymbol{a} \boldsymbol{a}^{T}}{\|\boldsymbol{a}\|^{2}}
$$

is a rank one orthogonal projector onto the span of $\boldsymbol{a}$ :

- The matrix $\boldsymbol{a} \boldsymbol{a}^{T}$ has rank one, since every column is a multiple of $\boldsymbol{a}$.
- The given matrix is a projector:

$$
\mathbb{P}^{2}=\frac{\boldsymbol{a} \boldsymbol{a}^{T}}{\|\boldsymbol{a}\|^{2}} \cdot \frac{\boldsymbol{a} \boldsymbol{a}^{T}}{\|\boldsymbol{a}\|^{2}}=\frac{1}{\|\boldsymbol{a}\|^{4}} \boldsymbol{a}\left(\boldsymbol{a}^{T} \boldsymbol{a}\right) \boldsymbol{a}^{T}=\frac{\boldsymbol{a} \boldsymbol{a}^{T}}{\|\boldsymbol{a}\|^{2}}=\mathbb{P}
$$

- The matrix is an orthogonal projector, since $\mathbb{P}^{T}=\mathbb{P}$.

Projecting to a Subspace: Let $Q=\left[\boldsymbol{q}_{1}, \boldsymbol{q}_{2}, \ldots, \boldsymbol{q}_{k}\right]$ be a matrix with orthonormal columns. Then

$$
\mathbb{P}=Q Q^{T}
$$

is an orthogonal projector to the column space of $Q$. This generalizes the result of the previous exercise. Note that if $Q$ was additionally a square matrix, $Q Q^{T}=I$.

Note that this is exactly the property that we discussed in the last example of the previous section.

## Exercises

1. Show that the plane $H$ defined by:

$$
H=\left\{\alpha_{1}\left[\begin{array}{l}
1 \\
1 \\
1
\end{array}\right]+\alpha_{2}\left[\begin{array}{r}
1 \\
-1 \\
0
\end{array}\right] \text { such that } \alpha_{1}, \alpha_{2} \in \mathbb{R}\right\}
$$

is isormorphic to $\mathbb{R}^{2}$.
2. Let the subspace $G$ be the plane defined below, and consider the vector $\boldsymbol{x}$, where:

$$
G=\left\{\alpha_{1}\left[\begin{array}{r}
1 \\
3 \\
-2
\end{array}\right]+\alpha_{2}\left[\begin{array}{r}
3 \\
-1 \\
0
\end{array}\right] \quad \text { such that } \alpha_{1}, \alpha_{2} \in \mathbb{R}\right\} \quad \boldsymbol{x}=\left[\begin{array}{l}
1 \\
0 \\
2
\end{array}\right]
$$

(a) Find the projector $P$ that takes an arbitrary vector and projects it (orthogonally) to the plane $G$.
(b) Find the orthogonal projection of the given $\boldsymbol{x}$ onto the plane $G$.
(c) Find the distance from the plane $G$ to the vector $\boldsymbol{x}$.
3. If the low dimensional representation of a vector $\boldsymbol{x}$ is $[9,-1]^{T}$ and the basis vectors are $[1,0,1]^{T}$ and $[3,1,1]^{T}$, then what was the original vector $\boldsymbol{x}$ ? (HINT: it is easy to compute it directly)
4. If the vector $\boldsymbol{x}=[10,4,2]^{T}$ and the basis vectors are $[1,0,1]^{T}$ and $[3,1,1]^{T}$, then what is the low dimensional representation for $\boldsymbol{x}$ ?
5. Let $\boldsymbol{a}=[-1,3]^{T}$. Find a square matrix $P$ so that $P \boldsymbol{x}$ is the orthogonal projection of $\boldsymbol{x}$ onto the span of $\boldsymbol{a}$.

### 0.3 The Four Fundamental Subspaces

Given any $m \times n$ matrix $A$, we consider the mapping $A: \mathbb{R}^{n} \rightarrow \mathbb{R}^{m}$ by:

$$
\boldsymbol{x} \rightarrow A \boldsymbol{x}=\boldsymbol{y}
$$

The four subspaces allow us to completely understand the domain and range of the mapping. We will first define them, then look at some examples.

## Definition 0.3.1. The Four Fundamental Subspaces

- The row space of $A$ is a subspace of $\mathbb{R}^{n}$ formed by taking all possible linear combinations of the rows of $A$. Formally,

$$
\operatorname{Row}(A)=\left\{\boldsymbol{x} \in \mathbb{R}^{n} \mid x=A^{T} \boldsymbol{y} \quad y \in \mathbb{R}^{m}\right\}
$$

- The null space of $A$ is a subspace of $\mathbb{R}^{n}$ formed by

$$
\operatorname{Null}(A)=\left\{\boldsymbol{x} \in \mathbb{R}^{n} \mid A \boldsymbol{x}=\mathbf{0}\right\}
$$

- The column space of $A$ is a subspace of $\mathbb{R}^{m}$ formed by taking all possible linear combinations of the columns of $A$.

$$
\operatorname{Col}(A)=\left\{\boldsymbol{y} \in \mathbb{R}^{m} \mid y=A \boldsymbol{x} \quad \in \mathbb{R}^{n}\right\}
$$

The column space is also the image of the mapping. Notice that $A \boldsymbol{x}$ is simply a linear combination of the columns of $A$ :

$$
A \boldsymbol{x}=x_{1} \boldsymbol{a}_{1}+x_{2} \boldsymbol{a}_{2}+\cdots+x_{n} \boldsymbol{a}_{n}
$$

- Finally, we define the null space of $A^{T}$ can be defined in the obvious way (see the Exercises).

The fundamental subspaces subdivide the domain and range of the mapping in a particularly nice way:

Theorem 0.3.1. Let $A$ be an $m \times n$ matrix. Then

- The nullspace of $A$ is orthogonal to the row space of $A$
- The nullspace of $A^{T}$ is orthogonal to the columnspace of $A$

Proof: See the Exercises.
Before going further, let us recall how to construct a basis for the column space, row space and nullspace of a matrix $A$. We'll do it with a particular matrix:

Example 0.3.1. Construct a basis for the column space, row space and nullspace of the matrix $A$ below:

$$
A=\left[\begin{array}{rrrr}
2 & 0 & -2 & 2 \\
-2 & 5 & 7 & 3 \\
3 & -5 & -8 & -2
\end{array}\right]
$$

SOLUTION: The row reduced form of $A$ is:

$$
\left[\begin{array}{rrrr}
1 & 0 & -1 & 1 \\
0 & 1 & 1 & 1 \\
0 & 0 & 0 & 0
\end{array}\right]
$$

The first two columns of the original matrix form a basis for the columnspace (which is a subspace of $\mathbb{R}^{3}$ ):

$$
\operatorname{Col}(A)=\operatorname{span}\left\{\left[\begin{array}{r}
2 \\
-2 \\
3
\end{array}\right],\left[\begin{array}{r}
2 \\
-2 \\
3
\end{array}\right]\right\}
$$

A basis for the row space is found by using the row reduced rows corresponding to the pivots (and is a subspace of $\mathbb{R}^{4}$ ). You should also verify that you can find a basis for the null space of $A$, given below (also a subspace of $\mathbb{R}^{4}$ ). If you're having any difficulties here, be sure to look it up in a linear algebra text:
$\operatorname{Row}(A)=\operatorname{span}\left\{\left[\begin{array}{r}1 \\ 0 \\ -1 \\ 1\end{array}\right],\left[\begin{array}{l}0 \\ 1 \\ 1 \\ 1\end{array}\right]\right\} \quad \operatorname{Null}(A)=\operatorname{span}\left\{\left[\begin{array}{r}1 \\ -1 \\ 1 \\ 0\end{array}\right],\left[\begin{array}{r}-1 \\ -1 \\ 0 \\ 1\end{array}\right]\right\}$
We will often refer to the dimensions of the four subspaces. We recall that there is a term for the dimension of the column space- That is, the rank.

Definition 0.3.2. The rank of a matrix $A$ is the number of independent columns of $A$.

In our previous example, the rank of $A$ is 2 . Also from our example, we see that the rank is the dimension of the column space, and that this is the same as the dimension of the row space (all three numbers correspond to the number of pivots in the row reduced form of $A$ ). Finally, a handy theorem for counting is the following.
The Rank Theorem. Let the $m \times n$ matrix $A$ have rank $r$. Then

$$
r+\operatorname{dim}(\operatorname{Null}(A))=n
$$

This theorem says that the number of pivot columns plus the other columns (which correspond to free variables) is equal to the total number of columns.

## Example 0.3.2. The Dimensions of the Subspaces.

Given a matrix $A$ that is $m \times n$ with rank $k$, then the dimensions of the four subspaces are shown below.

- $\operatorname{dim}(\operatorname{Row}(A))=k$
- $\operatorname{dim}(\operatorname{Col}(A))=k$
- $\operatorname{dim}(\operatorname{Null}(A))=n-k$
- $\operatorname{dim}\left(\operatorname{Null}\left(A^{T}\right)\right)=m-k$

There are some interesting implications of these theorems to matrices of data- For example, suppose $A$ is $m \times n$. With no other information, we do not know whether we should consider this matrix as $n$ points in $\mathbb{R}^{m}$, or $m$ points in $\mathbb{R}^{n}$. In one sense, it doesn't matter! The theorems we've discussed shows that the dimension of the columnspace is equal to the dimension of the rowspace. Later on, we'll find out that if we can find a basis for the columnspace, it is easy to find a basis for the rowspace. We'll need some more machinery first.

The Best Approximation Theorem If $W$ is a subspace of $\mathbb{R}^{n}$ and $\boldsymbol{x} \in$ $\mathbb{R}^{n}$, then the point closest to $\boldsymbol{x}$ in $W$ is the orthogonal projection of $\boldsymbol{x}$ into $W$. We prove this in the exercises below.

### 0.4 Exercises

1. Show that $\operatorname{Null}(A) \perp \operatorname{Row}(A)$.

Hint: One way to show this is to take an arbitrary $\boldsymbol{x}_{1} \neq \overrightarrow{0} \in \operatorname{Null}(A)$ and show that it is orthogonal to every row of $A$.
2. If $A$ is $m \times n$, how big can the rank of $A$ possibly be?
3. Show that multiplication by an orthogonal matrix preserves lengths: $\|\mathbb{Q} \boldsymbol{x}\|_{2}=$ $\|\boldsymbol{x}\|_{2}$ (Hint: Use properties of inner products). Conclude that multiplication by $\mathbb{Q}$ represents a rigid rotation.
4. Prove the Pythagorean Theorem by induction: Given a set of $n$ orthogonal vectors $\left\{\boldsymbol{x}_{i}\right\}$

$$
\left\|\sum_{i=1}^{n} \boldsymbol{x}_{i}\right\|_{2}^{2}=\sum_{i=1}^{n}\left\|\boldsymbol{x}_{i}\right\|_{2}^{2}
$$

5. Let $A$ be an $m \times n$ matrix where $m>n$, and let $A$ have $\operatorname{rank} n$. Let $\boldsymbol{x}, \boldsymbol{y} \in$ $\mathbb{R}^{m}$, such that $\boldsymbol{y}$ is the orthogonal projection of $\boldsymbol{x}$ onto the columnspace of $A$. We want a formula for the projector $\mathbb{P}: \mathbb{R}^{m} \rightarrow \mathbb{R}^{m}$ so that $\mathbb{P} \boldsymbol{x}=\boldsymbol{y}$.
(a) Why is the projector not $\mathbb{P}=A A^{T}$ ?
(b) Since $\boldsymbol{y}-\boldsymbol{x}$ is orthogonal to the range of $A$, show that

$$
\begin{equation*}
A^{T}(\boldsymbol{y}-\boldsymbol{x})=\mathbf{0} \tag{2}
\end{equation*}
$$

(c) Show that there exists $\boldsymbol{v}$ so that Equation (2) can be written as:

$$
\begin{equation*}
A^{T}(A \boldsymbol{v}-\boldsymbol{x})=0 \tag{3}
\end{equation*}
$$

(d) Argue that $A^{T} A$ is invertible, so that Equation (3) implies that

$$
\boldsymbol{v}=\left(A^{T} A\right)^{-1} A^{T} \boldsymbol{x}
$$

(e) Finally, show that this implies that

$$
\mathbb{P}=A\left(A^{T} A\right)^{-1} A^{T}
$$

Note: If $A$ has rank $k<m$, then we will need something different, since $A^{T} A$ will not be full rank. The missing piece is the singular value decomposition, to be discussed later.
6. The Orthogonal Decomposition Theorem: if $\boldsymbol{x} \in \mathbb{R}^{n}$ and $W$ is a (nonzero) subspace of $\mathbb{R}^{n}$, then $\boldsymbol{x}$ can be written uniquely as

$$
\boldsymbol{x}=\boldsymbol{w}+\boldsymbol{z}
$$

where $\boldsymbol{w} \in W$ and $\boldsymbol{z} \in W^{\perp}$.
To prove this, let $\left\{\boldsymbol{u}_{i}\right\}_{i=1}^{p}$ be an orthonormal basis for $W$, define $\boldsymbol{w}=$ $\left(\boldsymbol{x}, \boldsymbol{u}_{1}\right) \boldsymbol{u}_{1}+\ldots+\left(\boldsymbol{x}, \boldsymbol{u}_{p}\right) \boldsymbol{u}_{p}$, and define $\boldsymbol{z}=\boldsymbol{x}-\boldsymbol{w}$. Then:
(a) Show that $\boldsymbol{z} \in W^{\perp}$ by showing that it is orthogonal to every $\boldsymbol{u}_{i}$.
(b) To show that the decomposition is unique, suppose it is not. That is, there are two decompositions:

$$
\boldsymbol{x}=\boldsymbol{w}_{1}+\boldsymbol{z}_{1}, \quad \boldsymbol{x}=\boldsymbol{w}_{2}+\boldsymbol{z}_{2}
$$

Show this implies that $\boldsymbol{w}_{1}-\boldsymbol{w}_{2}=\boldsymbol{z}_{2}-\boldsymbol{z}_{1}$, and that this vector is in both $W$ and $W^{\perp}$. What can we conclude from this?
7. Use the previous exercises to prove the The Best Approximation Theorem If $W$ is a subspace of $\mathbb{R}^{n}$ and $\boldsymbol{x} \in \mathbb{R}^{n}$, then the point closest to $\boldsymbol{x}$ in $W$ is the orthogonal projection of $\boldsymbol{x}$ into $W$.

### 0.5 The Decomposition Theorems

### 0.5.1 The Eigenvector/Eigenvalue Decomposition

1. Definition: Let $A$ be an $n \times n$ matrix. Then an eigenvector-eigenvalue pair is a vector $\boldsymbol{v} \neq \mathbf{0}$, and a scalar $\lambda$ where

$$
\begin{equation*}
A \boldsymbol{v}=\lambda \boldsymbol{v} \Rightarrow(A-\lambda I) \boldsymbol{v}=\mathbf{0} \tag{4}
\end{equation*}
$$

2. Remark: If Equation (4) has a nontrivial solution, then

$$
\operatorname{det}(A-\lambda I)=0
$$

which leads to solving for the roots of a polynomial of degree $n$. This polynomial is called the characteristic polynomial.
3. Remark: We solve for the eigenvalues first, then solve for the nullspace of $(A-\lambda I)$ (which is called the eigenspace, $E_{\lambda}$ ), by solving

$$
(A-\lambda I) \boldsymbol{v}=\mathbf{0}
$$

4. Remark: Note that it is possible that one eigenvalue is repeated. This may or may not correspond with the same number of eigenvectors.
5. Definition: If eigenvalue $\lambda$ is repeated $k$ times as a root to the characteristic equation, then the algebraic multiplicity of $\lambda$ is $k$.
6. Definition: If the eigenspace $E_{\lambda}$ has dimension $k$, then $\lambda$ has geometric multiplicity $k$.
7. Example: Compute the eigenvalues and eigenvectors for: (i) the $2 \times 2$ identity matrix, (ii) The matrix (in Matlab notation): [1 2;0 1]
8. Theorem: If $a_{\lambda}$ is the algebraic multiplicity of $\lambda$ and $g_{\lambda}$ is the geometric multiplicity, then

$$
a_{\lambda} \geq g_{\lambda}
$$

We won't prove this here.
9. Definition: If, for some eigenvalue $\lambda$ of $A$, we have that $a_{\lambda}>g_{\lambda}, A$ is said to be defective.
10. Theorem: If $A$ is square, and $P$ is square and invertible, then $A$ and $P^{-1} A P$ have the same eigenvalues.
11. Exercise: Prove the previous theorem.
12. Remark: One method of characterizing eigenvalues in terms of the determinant and trace of a matrix:

$$
\operatorname{det}(A)=\prod_{i=1}^{n} \lambda_{i} \quad \operatorname{trace}(A)=\sum_{i=1}^{\infty} \lambda_{i}
$$

13. Remark: We will be especially interested in symmetric matrices. The rest of this section is devoted to them.
14. Definition: A matrix $A$ is orthogonally diagonalizeable if there is an orthogonal matrix $\mathbb{Q}$ and diagonal matrix $D$ so that so that $A=\mathbb{Q} D \mathbb{Q}^{T}$.
15. The Spectral Theorem: If $A$ is an $n \times n$ symmetric matrix, then:
(a) $A$ has $n$ real eigenvalues (counting multiplicity).
(b) For all $\lambda, a_{\lambda}=g_{\lambda}$.
(c) The eigenspaces for distinct eigenvalues are mutually orthogonal.
(d) $A$ is orthogonally diagonalizeable, with $D=\operatorname{diag}\left(\lambda_{1}, \lambda_{2}, \ldots, \lambda_{n}\right)$.

Some remarks about the Spectral Theorem:

- We assume that inside each eigenspace, we have an orthonormal basis of eigenvectors. This is not a restriction, since we can always construct such a basis using Gram-Schmidt.
- If a matrix is real and symmetric, the Spectral Theorem says that its eigenvectors form an orthonormal basis for $\mathbb{R}^{n}$.
- The full proof is beyond the scope of this course, but we can prove some parts (given below).

16. The following is a proof of part (c). Supply justification for each step: Let $\boldsymbol{v}_{1}, \boldsymbol{v}_{2}$ be eigenvectors from distinct eigenvalues, $\lambda_{1}, \lambda_{2}$. We show that $\boldsymbol{v}_{1} \cdot \boldsymbol{v}_{2}=0$ :

$$
\lambda_{1} \boldsymbol{v}_{1} \cdot \boldsymbol{v}_{2}=\left(A \boldsymbol{v}_{1}\right)^{T} \boldsymbol{v}_{2}=\boldsymbol{v}_{1}^{T} A^{T} \boldsymbol{v}_{2}=\boldsymbol{v}_{1}^{T} A \boldsymbol{v}_{2}=\lambda_{2} \boldsymbol{v}_{1} \cdot \boldsymbol{v}_{2}
$$

Now, $\left(\lambda_{1}-\lambda_{2}\right) \boldsymbol{v}_{1} \cdot \boldsymbol{v}_{2}=0$.
17. The Spectral Decomposition: Since $A$ is orthogonally diagonalizable, then

$$
A=\left[\begin{array}{llll}
\boldsymbol{q}_{1} & \boldsymbol{q}_{2} & \ldots & \boldsymbol{q}_{n}
\end{array}\right]\left[\begin{array} { c c c c } 
{ \lambda _ { 1 } } & { 0 } & { \ldots } & { 0 } \\
{ 0 } & { \lambda _ { 2 } } & { \ldots } & { 0 } \\
{ \vdots } & { \vdots } & { \ddots } & { \vdots } \\
{ 0 } & { 0 } & { \ldots } & { \lambda _ { n } }
\end{array} \left[\left[\begin{array}{c}
\boldsymbol{q}_{1}^{T} \\
\boldsymbol{q}_{2}^{T} \\
\vdots \\
\boldsymbol{q}_{n}^{T}
\end{array}\right]\right.\right.
$$

so that:

$$
A=\left[\begin{array}{lllll}
\lambda_{1} \boldsymbol{q}_{1} & \lambda_{2} \boldsymbol{q}_{2} & \ldots & \lambda_{n} \boldsymbol{q}_{n}
\end{array}\right]\left[\begin{array}{c}
\boldsymbol{q}_{1}^{T} \\
\boldsymbol{q}_{2}^{T} \\
\vdots \\
\boldsymbol{q}_{n}^{T}
\end{array}\right]
$$

so finally:

$$
A=\lambda_{1} \boldsymbol{q}_{1} \boldsymbol{q}_{1}^{T}+\lambda_{2} \boldsymbol{q}_{2} \boldsymbol{q}_{2}^{T}+\ldots+\lambda_{n} \boldsymbol{q}_{n} \boldsymbol{q}_{n}^{T}
$$

That is, $A$ is a sum of $n$ rank one matrices, each of which is a projection matrix.
18. Matlab Exercise: Verify the spectral decomposition for a symmetric matrix. Type the following into Matlab (the lines that begin with a \% denote comments that do not have to be typed in).

```
%Construct a random, symmetric, 6 x 6 matrix:
for i=1:6
    for j=1:i
                A(i,j)=rand;
        A(j,i)=A(i,j);
    end
end
%Compute the eigenvalues of A:
[Q,L]=eig(A); %NOTE: A = Q L Q'
                            %L is a diagonal matrix
%Now form the spectral sum
S=zeros(6,6); for i=1:6
    S=S+L(i,i)*Q(:,i)*Q(:,i)';
end
max(max(S-A))
```

Note that the maximum of $S-A$ should be a very small number! (By the spectral decomposition theorem).

### 0.5.2 The Singular Value Decomposition

There is a special matrix factorization that is extremely useful, both in applications and in proving theorems. This is mainly due to two facts, which we shall investigate in this section: (1) We can use this factorization on any matrix, (2) The factorization defines explicitly the rank of the matrix, and all four matrix subspaces.

In what follows, assume that $A$ is an $m \times n$ matrix (so $A$ maps $\mathbb{R}^{n}$ to $\mathbb{R}^{m}$ ).

1. Remark: Although $A$ itself is not symmetric, $A^{T} A$ is $n \times n$ and symmetric. Therefore, it is orthogonally diagonalizable. Let $\left\{\lambda_{i}\right\}_{i=1}^{n}$ and $V=\left[\boldsymbol{v}_{1}, \boldsymbol{v}_{2}, \ldots, \boldsymbol{v}_{n}\right]$ be the eigenvalues and orthonormal eigenvectors.
2. Exercise: Show that $\lambda_{i} \geq 0$ for $i=1$..n by showing that $\left\|A \boldsymbol{v}_{i}\right\|_{2}^{2}=\lambda_{i}$.
3. Definition: We define the singular values of $A$ by:

$$
\sigma_{i}=\sqrt{\lambda_{i}}
$$

where $\lambda_{i}$ is an eigenvalue of $A^{T} A$.
4. Remark: In the rest of the section, we will assume any list (or diagonal matrix) of eigenvalues of $A^{T} A$ (or singular values of $A$ ) will be ordered from highest to lowest: $\lambda_{1} \geq \lambda_{2} \geq \ldots \geq \lambda_{n}$.
5. Exercise: Let $\mathbf{v}_{i}, \mathbf{v}_{j}$ be non-zero vectors for $\lambda_{i}, \lambda_{j}$. We note that, if $\lambda_{i} \neq \lambda_{j}$, then the eigenvectors are orthogonal by the Spectral Theorem. If $\lambda_{i}=\lambda_{j}$, it is possible to choose the vectors so we have an orthonormal basis for $E_{\lambda}$.

Prove that, if $\boldsymbol{v}_{i}$ and $\boldsymbol{v}_{j}$ are distinct eigenvectors of $A^{T} A$, then their corresponding images, $A \boldsymbol{v}_{i}$ and $A \boldsymbol{v}_{j}$, are orthogonal.
SOLUTION:

$$
\left(A \mathbf{v}_{i}\right)^{T} A \mathbf{v}_{j}=\boldsymbol{v}_{i} A^{T} A \boldsymbol{v}_{j}=\lambda_{j} \boldsymbol{v}_{i}^{T} \boldsymbol{v}_{j}=0
$$

6. Exercise: Prove that, if $\boldsymbol{x}=\alpha_{1} \boldsymbol{v}_{1}+\ldots \alpha_{n} \boldsymbol{v}_{n}$, then

$$
\|A \boldsymbol{x}\|^{2}=\alpha_{1}^{2} \lambda_{1}+\ldots+\alpha_{n}^{2} \lambda_{n}
$$

where $\lambda_{i}, \mathbf{v}_{i}$ are the evals/evecs of $A^{T} A$.
SOLUTION: Just work it out using the dot product and orthogonality (or the Pythagorean Theorem), and Exercises 2 and 5.

$$
\|A \boldsymbol{x}\|^{2}=\left\|\alpha_{1} A \boldsymbol{v}_{1}+\ldots \alpha_{n} A \boldsymbol{v}_{n}\right\|^{2}=\alpha_{1}^{2}\left\|A \boldsymbol{v}_{1}\right\|^{2}+\ldots \alpha_{n}^{2}\left\|\boldsymbol{v}_{n}\right\|^{2}
$$

7. Exercise: Let $W$ be the subspace generated by the basis $\left\{\boldsymbol{v}_{j}\right\}_{j=k+1}^{n}$, where $\boldsymbol{v}_{j}$ are the eigenvectors associated with the zero eigenvalues of $A^{T} A$ (therefore, we are assuming that the first $k$ eigenvalues are NOT zero). Show that $W=\operatorname{Null}(A)$.
SOLUTION:
We show that $W$ is the nullspace of $A$ :
Let $\boldsymbol{x} \neq \mathbf{0} \in W$. Then

$$
\|A \boldsymbol{x}\|=x^{T} A^{T} A x=0
$$

since $A^{T} A \boldsymbol{x}=\mathbf{0}$. Therefore, $A \boldsymbol{x}=0$, and $\boldsymbol{x}$ is in the nullspace of $A$. Since $\boldsymbol{x}$ was chosen arbitrarily, $W$ is contained in nullspace of $A$.
Now, if we take a vector in the null space of $A$, is it possible that it is NOT contained in $W$ ?
If $\mathbf{x} \in$ null space, then $\|A \boldsymbol{x}\|=0$. Furthermore, we have the expansion

$$
\boldsymbol{x}=\alpha_{1} \boldsymbol{v}_{1}+\ldots+\alpha_{r} \boldsymbol{v}_{r}
$$

and therefore

$$
\|A \boldsymbol{x}\|^{2}=\alpha_{1}^{2} \lambda_{1}+\ldots+\alpha_{r}^{2} \lambda_{r}
$$

which is zero only if every $\alpha$ is zero.
Thus, $W$ is the nullspace of $A$.
8. Exercise: Prove that if the rank of $A^{T} A$ is $r$, then so is the rank of $A$. SOLUTION:
If the rank of $A^{T} A$ is $r$, the dimension of the nullspace of $A$ is $n-r$ by our previous exercise. By the rank theorem,

$$
\operatorname{rank}(A)+\operatorname{dim}(\operatorname{Null}(A))=n
$$

Thus, the rank of $A$ is also $r$.
9. Remark: Define

$$
\boldsymbol{u}_{i}=\frac{1}{\left\|A \boldsymbol{v}_{i}\right\|_{2}} A \boldsymbol{v}_{i}=\frac{1}{\sigma_{i}} A \boldsymbol{v}_{i}
$$

and let $U$ be the matrix whose $i^{\text {th }}$ column is $\boldsymbol{u}_{i}$.
10. Remark: This definition only makes sense sense for the first $r$ vectors $\boldsymbol{v}$ (otherwise, $A \boldsymbol{v}_{i}=\mathbf{0}$ ). Thus, we'll have to extend the basis to span all of $\mathbb{R}^{m}$.
11. Exercise: Sketch how you might do this.
12. Exercise: Show that $\boldsymbol{u}_{i}$ is an eigenvector of $A A^{T}$ whose eigenvalue is also $\lambda_{i}$.

## SOLUTION:

Show that $A A^{T} \boldsymbol{u}_{i}=\lambda_{i} \boldsymbol{u}_{i}$ :

$$
A A^{T} \boldsymbol{u}_{i}=A A^{T}\left(\frac{1}{\sigma_{i}} A \boldsymbol{v}_{i}\right)=\frac{1}{\sigma_{i}} A\left(A^{T} A \boldsymbol{v}_{i}\right)
$$

so that:

$$
A A^{T} \boldsymbol{u}_{i}=\frac{\lambda_{i}}{\sigma_{i}} A \boldsymbol{v}_{i}=\frac{\lambda_{i}}{\sigma_{i}} \sigma_{i} \boldsymbol{u}_{i}
$$

13. Exercise: Show that $A^{T} \boldsymbol{u}_{i}=\sigma_{i} \boldsymbol{v}_{i}$

SOLUTION:
We show this for non-zero $\sigma_{i}$ :

$$
A^{T} \boldsymbol{u}_{i}=A^{T}\left(\frac{1}{\sigma_{i}} A \boldsymbol{v}_{i}\right)=\frac{\lambda_{i}}{\sigma_{i}} \boldsymbol{v}_{i}=\sigma_{i} \boldsymbol{v}_{i}
$$

14. Remark: So far, we have shown how to construct two matrices, $U$ and $V$ given a matrix $A$. That is, the matrix $V$ is constructed by the eigenvectors of $A^{T} A$, and the matrix $U$ can be constructed using the $\boldsymbol{v}$ 's or by finding the eigenvectors of $A A^{T}$.
15. Exercise: Let $A$ be $m \times n$. Define the $m \times n$ matrix

$$
\Sigma=\operatorname{diag}\left(\sigma_{1}, \ldots, \sigma_{n}\right)
$$

where $\sigma_{i}$ is the $i^{\text {th }}$ singular value of the matrix $A$. Show that

$$
A V=U \Sigma
$$



Figure 2: The geometric meaning of the right and left singular vectors of the SVD decomposition. Note that $A \boldsymbol{v}_{i}=\sigma_{i} \boldsymbol{u}_{i}$. The mapping $x \rightarrow A x$ will map the unit circle on the left to the ellipse on the right.
16. The Singular Value Decomposition (SVD) Let $A$ be any $m \times n$ matrix of rank $r$. Then

$$
A=U \Sigma V^{T}
$$

where $U, \Sigma, V$ are the matrices defined in the previous exercises. That is, $U$ is an orthogonal $m \times m$ matrix, $\Sigma$ is a diagonal $m \times n$ matrix, and $V$ is an orthogonal $n \times n$ matrix. The $\boldsymbol{u}$ 's are called the left singular vectors and the $\boldsymbol{v}$ 's are called the right singular vectors.
17. Remark: Keep in mind the following relationship between the right and left singular vectors:

$$
\begin{aligned}
A \boldsymbol{v}_{i} & =\sigma_{i} \boldsymbol{u}_{i} \\
A^{T} \boldsymbol{u}_{i} & =\sigma_{i} \boldsymbol{v}_{i}
\end{aligned}
$$

18. Computing The Four Subspaces to a matrix $A$. Let $A=U \Sigma V^{T}$ be the SVD of $A$ which has rank $r$. Be sure that the singular values are ordered from highest to lowest. Then:
(a) A basis for the columnspace of $A, \operatorname{Col}(A)$ is $\left\{\boldsymbol{u}_{i}\right\}_{i=1}^{r}$
(b) A basis for nullspace of $A, \operatorname{Null}(A)$ is $\left\{\boldsymbol{v}_{i}\right\}_{i=r+1}^{n}$
(c) A basis for the rowspace of $A, \operatorname{Row}(A)$ is $\left\{\boldsymbol{v}_{i}\right\}_{i=1}^{r}$
(d) A basis for the nullspace of $A^{T}, \operatorname{Null}\left(A^{T}\right)$ is $\left\{\boldsymbol{u}_{i}\right\}_{i=r+1}^{m}$
19. We can also visualize the right and left singular values as in Figure 2. We think of the $\boldsymbol{v}_{i}$ as a special orthogonal basis in $R^{n}$ (Domain) that maps to the ellipse whose axes are defined by $\sigma_{i} \boldsymbol{u}_{i}$.
20. The SVD is one of the premier tools of linear algebra, because it allows us to completely reveal everything we need to know about a matrix mapping: The rank, the basis of the nullspace, a basis for the column space, the basis


Figure 3: The SVD of A ([U,S,V]=svd(A)) completely and explicitly describes the 4 fundamental subspaces associated with the matrix, as shown. We have a one to one correspondence between the rowspace and columnspace of $A$, the remaining $\boldsymbol{v}$ 's map to zero, and the remaining $\boldsymbol{u}$ 's map to zero (under $A^{T}$ ).
for the nullspace of $A^{T}$, and of the row space. This is depicted in Figure 3.
21. Lastly, the SVD provides a decomposition of any linear mapping into two "rotations" and a scaling. This will become important later when we try to deduce a mapping matrix from data (See the section on signal separation).
22. Exercise: Compute the SVD by hand of the following matrices:

$$
\left(\begin{array}{ll}
1 & 1 \\
0 & 0
\end{array}\right) \quad\left(\begin{array}{ll}
0 & 2 \\
0 & 0 \\
0 & 0
\end{array}\right)
$$

SOLUTION: For the first matrix,

$$
A A^{T}=\left[\begin{array}{ll}
2 & 0 \\
0 & 0
\end{array}\right] \quad \Rightarrow \quad U=\left[\begin{array}{ll}
1 & 0 \\
0 & 1
\end{array}\right] \quad D=\left[\begin{array}{ll}
2 & 0 \\
0 & 0
\end{array}\right]
$$

$$
A^{T} A=\left[\begin{array}{ll}
1 & 1 \\
1 & 1
\end{array}\right] \quad \Rightarrow \quad V=\frac{1}{\sqrt{2}}\left[\begin{array}{rr}
1 & -1 \\
1 & 1
\end{array}\right] \quad D=\left[\begin{array}{ll}
2 & 0 \\
0 & 0
\end{array}\right]
$$

Therefore,

$$
\left(\begin{array}{ll}
1 & 1 \\
0 & 0
\end{array}\right)=\left[\begin{array}{ll}
1 & 0 \\
0 & 1
\end{array}\right]\left[\begin{array}{rr}
\sqrt{2} & 0 \\
0 & 0
\end{array}\right] \frac{1}{\sqrt{2}}\left[\begin{array}{rr}
1 & -1 \\
1 & 1
\end{array}\right]
$$

23. Remark: If $m$ or $n$ is very large, it might not make sense to keep the full matrix $U$ and $V$.
24. The Reduced SVD Let $A$ be $m \times n$ with rank $r$. Then we can write:

$$
A=\tilde{U} \tilde{\Sigma} \tilde{V}^{T}
$$

where $\tilde{U}$ is an $m \times r$ matrix with orthogonal columns, $\tilde{\Sigma}$ is an $r \times r$ square matrix, and $\tilde{V}$ is an $n \times r$ matrix.
25. Theorem: (Actually, this is just another way to express the SVD). Let $A=U \Sigma V^{T}$ be the SVD of $A$, which has rank $r$. Then:

$$
A=\sum_{i=1}^{r} \sigma_{i} \boldsymbol{u}_{i} \boldsymbol{v}_{i}^{T}
$$

Therefore, we can approximate $A$ by the sum of rank one matrices.
26. Matlab and the SVD Matlab has the SVD built in. The function specifications are: $[\mathrm{U}, \mathrm{S}, \mathrm{V}]=\operatorname{svd}(\mathrm{A})$ and $[\mathrm{U}, \mathrm{S}, \mathrm{V}]=\operatorname{svd}(\mathrm{A}, 0)$ where the first function call returns the full SVD, and the second call returns a reduced SVD- see Matlab's help file for the details on the second call.
27. Matlab Exercise: Image Processing and the SVD. First, in Matlab, load the clown picture:
load clown
This loads a matrix $X$ and a colormap, map, into the workspace. To see the clown, type:

```
image(X); colormap(map)
```

We now perform a Singular Value Decomposition on the clown. Type in: $[\mathrm{U}, \mathrm{S}, \mathrm{V}]=\mathrm{svd}(\mathrm{X})$;

How many vectors are needed to retain a good picture of the clown? Try performing a $k$-dimensional reconstruction of the image by typing:
$H=U(:, 1: k) * S(1: k, 1: k) * V(:, 1: k) ' ; ~ i m a g e(H)$
for $k=5,10,20$ and 30 . What do you see?

## Generalized Inverses

Let a matrix $A$ be $m \times n$ with rank $r$. In the general case, $A$ does not have an inverse. Is there a way of restricting the domain and range of the mapping $\boldsymbol{y}=A \boldsymbol{x}$ so that the map is invertible?

We know that the columnspace and rowspace of $A$ have the same dimensions. Therefore, there exists a 1-1 and onto map between these spaces, and this is our restriction.

To "solve" $\boldsymbol{y}=A \boldsymbol{x}$, we replace $\boldsymbol{y}$ by its orthogonal projection to the columnspace of $A, \hat{\boldsymbol{y}}$. This gives the least squares solution, which makes the problem solvable. To get a unique solution, we replace $\boldsymbol{x}$ by its projection to the rowspace of $A$, $\hat{\boldsymbol{x}}$. The problem

$$
\hat{\boldsymbol{y}}=A \hat{\boldsymbol{x}}
$$

now has a solution, and that solution is unique. We can rewrite this problem now in terms of the reduced SVD of $A$ :

$$
\hat{\boldsymbol{x}}=V V^{T} \boldsymbol{x}, \quad \hat{\boldsymbol{y}}=U U^{T} \boldsymbol{y}
$$

Now we can write:

$$
U U^{T} \boldsymbol{y}=U \Sigma V^{T}\left(V V^{T} \boldsymbol{x}\right)
$$

so that

$$
V \Sigma^{-1} U^{T} \boldsymbol{y}=V V^{T} \boldsymbol{x}
$$

(Exercise: Verify that these computations are correct!)
Given an $m \times n$ matrix $A$, define its pseudoinverse, $A^{\dagger}$ by:

$$
A^{\dagger}=V \Sigma^{-1} U^{T}
$$

We have shown that the least squares solution to $\boldsymbol{y}=A \boldsymbol{x}$ is given by:

$$
\hat{\boldsymbol{x}}=A^{\dagger} \boldsymbol{y}
$$

where $\hat{\boldsymbol{x}}$ is in the rowspace of $A$, and its image, $A \hat{\boldsymbol{x}}$ is the projection of $\boldsymbol{y}$ into the columnspace of $A$.

Geometrically, we can understand these computations in terms of the four fundamental subspaces.


In this case, there is no value of $\boldsymbol{x} \in \mathbb{R}^{n}$ which will map onto $\boldsymbol{y}$, since $\boldsymbol{y}$ is outside the columnspace of $A$. To get a solution, we project $\boldsymbol{y}$ onto the columnspace of $A$ as shown below:



Now it is possible to find an $\boldsymbol{x}$ that will map onto $\hat{\boldsymbol{y}}$, but if the nullspace of $A$ is nontrivial, then all of the points on the dotted line will also map to $\hat{\boldsymbol{y}}$



Finally, we must choose a unique value of $\boldsymbol{x}$ for the mapping- We choose the $\boldsymbol{x}$ inside the rowspace of $A$.

This is a very useful idea, and it is one we will explore in more detail later. For now, notice that to get this solution, we analyzed our four fundamental subspaces in terms of the basis vectors given by the SVD.

## Exercises

1. Consider

$$
\left[\begin{array}{ccc}
2 & 1 & -1 \\
3 & 1 & 2
\end{array}\right]\left[\begin{array}{l}
x_{1} \\
x_{2} \\
x_{3}
\end{array}\right]=\left[\begin{array}{l}
5 \\
1
\end{array}\right]
$$

(a) Before solving this problem, what are the dimensions of the four fundamental subspaces?
(b) Use Matlab to compute the SVD of the matrix $A$, and solve the problem by computing the pseudoinverse of $A$ directly.
(c) Check your answer explicitly and verify that $\hat{\boldsymbol{x}}$ and $\hat{\boldsymbol{y}}$ are in the rowspace and columnspace. (Hint: If a vector $\boldsymbol{x}$ is already in the rowspace, what is $V V^{T} \boldsymbol{x}$ ?)
2. Consider

$$
\left[\begin{array}{rrrr}
2 & 1 & -1 & 3 \\
-1 & 0 & 1 & -2 \\
7 & 2 & -5 & 12 \\
-3 & -2 & 0 & -4 \\
4 & 1 & -3 & 7
\end{array}\right]\left[\begin{array}{l}
x_{1} \\
x_{2} \\
x_{3} \\
x 4
\end{array}\right]=\left[\begin{array}{r}
5 \\
1 \\
0 \\
-2 \\
6
\end{array}\right]
$$

(a) Find the dimensions of the four fundamental subspaces by using the SVD of $A$ (in Matlab).
(b) Solve the problem.
(c) Check your answer explicitly and verify that $\hat{\boldsymbol{x}}$ and $\hat{\boldsymbol{y}}$ are in the rowspace and columnspace.
3. Write the following in Matlab to reproduce Figure 2:

```
theta=linspace(0,2*pi,30);
z=exp(i*theta);
X=[real(z);imag(z)]; %The domain points
m=1/sqrt(2);
A=(m*[1,1;1,-1])*[1,0;0,3];
Y=A*X; %The image of the circle
t=linspace (0,1);
vec1=[0;0]*(1-t)+[0;1]*t; %The basis vectors v
vec2=[0;0]*(1-t)+[1;0]*t;
Avec1=A*vec1; Avec2=A*vec2; %Image of the basis vectors
figure(1) %The domain
plot(X(1,:),X(2,:),'k',vec1(1,:),vec1(2,:),'k',
    vec2(1,:),vec2(2,:),'k');
axis equal
figure(2) %The image
plot(Y(1,:),Y(2,:),'k',Avec1(1,:),Avec1(2,:),'k',
    Avec2(1,:),Avec2(2,:),'k');
axis equal
```

4. In the previous example, what was the matrix $A$ ? The vectors $\boldsymbol{v}$ ? The vectors $\boldsymbol{u}$ ? The singular values $\sigma_{1}, \sigma_{2}$ ?
Once you've written these down from the program, perform the SVD of $A$ in Matlab. Are the vectors the same that you wrote down?
NOTE: These show that the singular vectors are not unique- they vary by $\pm \boldsymbol{v}$, or $\pm \boldsymbol{u}$.

### 0.6 Interactions Between Subspaces and the SVD

Suppose that a matrix $A$ is $p \times n$ and $B$ is $q \times n$. Then we have four fundamental subspaces for each of $A, B$. In particular, the row spaces and null spaces of $A$ and $B$ are all in $\mathbb{R}^{n}$. Note that this interpretation is looking at the matrix $A$ as containing $p$ sample data points from $\mathbb{R}^{n}$, and the matrix $B$ as containing $q$ sample data points from $\mathbb{R}^{n}$.

It is natural to ask about the interaction of these subspaces of $\mathbb{R}^{n}$, which are enumerated below:

- The rowspace of $A$ separate from $B$. This is also the intersection of the rowspace of $A$ with the nullspace of $B$.
- The rowspace of $B$ separate from $A$. This is also the intersection of the rowspace of $B$ with the nullspace of $A$.
- The intersection of the rowspaces of $A$ and $B$.
- The intersection of the nullspaces of $A$ and $B$.

How can we obtain a basis for the intersection of nullspaces? This is in fact fairly easy as long as $p, q$ are relatively small. We construct a new matrix $Z$ that is $p+q \times n$ :

$$
Z=\left[\begin{array}{l}
A \\
B
\end{array}\right]
$$

and find the nullspace of this via the SVD of Z .
Exercise: Prove that, if $\mathbf{v} \in \mathbb{R}^{n}$, and $Z \mathbf{v}=\mathbf{0}$, then $\mathbf{v} \in \operatorname{Null}(A) \cap \operatorname{Null}(B)$.
Exercise: Prove directly that, if $\mathbf{v} \in \mathbb{R}^{n}$, and $Z \mathbf{v} \neq \mathbf{0}$, then $\mathbf{v} \in \operatorname{Row}(A) \cup$ Row(B).

Note that the first exercise proves the second in that:

$$
(\operatorname{Null}(A) \cap \operatorname{Null}(B))^{c}=\operatorname{Null}(A)^{c} \cup \operatorname{Null}(B)^{c}=\operatorname{Row}(A) \cup \operatorname{Row}(B)
$$

where $c$ is the complement of the set.

We can find bases for the $\operatorname{Row}(A)$ and $\operatorname{Row}(B)$ directly- Let us define the two reduced SVDs:

$$
A=U_{A} \Sigma_{A} V_{A}^{T}, \quad B=U_{B} \Sigma_{B} V_{B}^{T}
$$

Then the columns of $V_{A}, V_{B}$ form a basis for the rowspace of $A$ and the rowspace of $B$ respectively. Before continuing further in the discussion, lets consider the idea of distance and angles between two subspaces.

### 0.6.1 Angles Between Subspaces

As in Golub and Van Loan, let $F, G$ be two subspaces of $\mathbb{R}^{m}$. Without loss of generality, assume:

$$
p=\operatorname{dim}(F) \geq \operatorname{dim}(G)=q \geq 1
$$

The principal angles $\theta_{1}, \ldots, \theta_{q} \in\left[0, \frac{\pi}{2}\right]$ between $F$ and $G$ are defined recursively by:

$$
\cos \left(\theta_{k}\right)=\max _{u \in F} \max _{v \in G} u^{T} v=u_{k}^{T} v_{k}
$$

subject to the additional constraints that $u, v$ be unit length and they are orthogonal to the previously found vectors $u_{1}, \ldots, u_{k-1}$ and $v_{1}, \ldots v_{k-1}$. The vectors $u_{1}, \ldots u_{q}$ and $v_{1}, \ldots, v_{q}$ are called the principal vectors between subspaces $F$ and $G$.

Some remarks about this definition:

- Since $u, v$ are normalized, the maximum value of $u^{T} v$ is 1 , corresponding to $\theta=0$. Thus, principal vectors associated with this value of $\theta$ are the same. These vectors will also give the an orthonormal basis for the intersection of $F, G$.
- Since $\theta$ is restricted, the smallest value of $u^{T} v$ is zero, corresponding to $\theta=\pi / 2$. The corresponding vectors of $F$ and $G$ will be orthogonal.
- We can define the distance between subspaces $F$ and $G$ if $p=q$ by using the largest principal angle, $\theta_{p}$ :

$$
\operatorname{dist}(F, G)=\sqrt{1-\cos ^{2}\left(\theta_{p}\right)}=\sin \left(\theta_{p}\right)
$$

For example, if $F$ and $G$ are the same subspace, then $\theta_{k}=0$ for $k=$ $1, \ldots, p$, and the distance between them is 0 . On the other extreme, if $F$ and $G$ are orthogonal, then $\theta_{k}=\pi / 2$ for $k=1, \ldots, p$ and the distance between them is 1 .

### 0.6.2 Computing the principal angles and vectors.

Here we give an intuitive idea behind the algorithm; for a more details see Golub and Van Loan (Chapter 12).

Given two sets of orthonormal basis vectors for subspaces $F$ and $G$ (we'll use $V_{A}$ and $V_{B}$ found earlier), we can write:

$$
u^{T} v=y^{T} V_{A}^{T} V_{B} z
$$

so that $u=V_{A} y$ and $v=V_{B} z$. Thus, if $U, V$ are the matrices whose columns are the principal vectors, then

$$
U^{T} V=Y^{T}\left(V_{A}^{T} V_{B}\right) Z=\operatorname{diag}\left(\cos \left(\theta_{1}\right), \ldots, \cos \left(\theta_{q}\right)\right)=D
$$

Notice that this is the SVD of $V_{A}^{T} V_{B}$ :

$$
V_{A}^{T} V_{B}=Y D Z^{T}
$$

Note again that we have not proven anything- this was just an observation. For a proof, we would need additional facts about the SVD of a matrix that would take us too far afield.

Example: Let $F, G$ be subspaces of $\mathbb{R}^{3}$, where $F$ is the $x-y$ plane and $G$ is the $x-z$ plane. Clearly, there is a one-dimensional intersection. We show this using the computations in the previous discussion.

$$
F=\operatorname{span}\left\{\left[\begin{array}{l}
1 \\
0 \\
0
\end{array}\right],\left[\begin{array}{l}
0 \\
1 \\
0
\end{array}\right]\right\}, \quad G=\operatorname{span}\left\{\left[\begin{array}{l}
1 \\
0 \\
0
\end{array}\right],\left[\begin{array}{l}
0 \\
0 \\
1
\end{array}\right]\right\}
$$

We can use these basis vectors as $V_{A}, V_{B}$ respectively. Now,

$$
V_{A}^{T} V_{B}=\left[\begin{array}{lll}
1 & 0 & 0 \\
0 & 1 & 0
\end{array}\right]\left[\begin{array}{ll}
1 & 0 \\
0 & 0 \\
0 & 1
\end{array}\right]=\left[\begin{array}{ll}
1 & 0 \\
0 & 0
\end{array}\right]
$$

Since this matrix is already in diagonal form,

$$
\left[\begin{array}{ll}
1 & 0 \\
0 & 0
\end{array}\right]=I\left[\begin{array}{cc}
\cos \left(\theta_{1}\right) & 0 \\
0 & \cos \left(\theta_{2}\right)
\end{array}\right] I^{T}
$$

so that $Y=Z=I$ and

$$
U=V_{A} I=V_{A}, \quad V=V_{B} I=V_{B}
$$

Therefore, $u_{1}=v_{1}$ corresponding to the subspace intersection, and the distance between the subspaces is 1 . We also see that there is no nontrivial intersection between the nullspaces of $F, G$. We would see this if we took the SVD of $\left[\begin{array}{ll}V_{A}^{T} & V_{B}^{T}\end{array}\right]^{T}$ as suggested at the beginning of this section, since there would be no zero singular values.

Similarly, we can find the intersection of $F$ with $G^{\perp}$. In this case,

$$
V_{A}^{T} V_{B}=\left[\begin{array}{lll}
1 & 0 & 0 \\
0 & 1 & 0
\end{array}\right]\left[\begin{array}{l}
0 \\
1 \\
0
\end{array}\right]=\left[\begin{array}{l}
0 \\
1
\end{array}\right]=Y \cdot 1 \cdot 1=Y \cdot 1 \cdot Z^{T}
$$

Since this "matrix" is already in diagonal form,

$$
U=V_{A} Y=\left[\begin{array}{l}
0 \\
1 \\
0
\end{array}\right], \quad V=V_{B} \cdot 1=\left[\begin{array}{l}
0 \\
1 \\
0
\end{array}\right]
$$

