# Chapter 5

# Linear Algebra

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.

## 5.1 Representation, Basis and Dimension

Let us quickly review some notation and basic ideas from linear algebra. First, a **basis** is a linearly independent spanning set, and the **dimension** of a subspace is the number of basis vectors it needs.

Suppose we have a subspace $H \subset \mathbb{R}^n$, and a basis for $H$ in the columns of matrix $V$.

By the definition of a basis, every vector in the subspace $H$ can be written as a linear combination of the basis elements. In particular, if $\boldsymbol{x} \in H$, then we can write:

$$\boldsymbol{x} = c_1\boldsymbol{v}_1 + \ldots + c_k\boldsymbol{v}_k \doteq V\boldsymbol{c} = V[\mathbf{x}]_V \tag{5.1}$$

where $\boldsymbol{c}$ is sometimes denoted as $[\boldsymbol{x}]_V$, and is referred to as the *coordinates of* $\boldsymbol{x}$ *with respect to the basis in* $V$.

Therefore, every vector in our subset of $\mathbb{R}^n$ can be identified with a point in $\mathbb{R}^k$, which gives us a function that we'll call the **coordinate mapping**:

$$\boldsymbol{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \in \mathbb{R}^n \longleftrightarrow \begin{bmatrix} c_1 \\ \vdots \\ c_k \end{bmatrix} = \boldsymbol{c} \in \mathbb{R}^k$$

If $k$ is small (with respect to $n$) we think of $\boldsymbol{c}$ is the **low dimensional representation** of the vector $\boldsymbol{x}$, and that $H$ is *isomorphic* to $\mathbb{R}^k$ (Isormorphic meaning one to one and onto linear map is the isormorphism)

**Example 5.1.1.** If $\boldsymbol{v}_i, \boldsymbol{v}_j$ are two linearly independent vectors, then the subspace created by their span is *isomorphic* to the plane $\mathbb{R}^2$ - but is not *equal* to the plane. The isomorphism is the coordinate mapping.

Generally, finding the coordinates of $\boldsymbol{x}$ with respect to an arbitrary basis (as columns of a matrix $V$) means that we have to solve Equation 5.1. However, if the columns of $V$ are orthogonal, then it is very simple to compute the coordinates.

Start with Equation 5.1:

$$\boldsymbol{x} = c_1\boldsymbol{v}_1 + \cdots + c_k\boldsymbol{v}_k$$

Now take the inner product of both sides with $\boldsymbol{v}_j$:

$$\boldsymbol{x} \cdot \boldsymbol{v}_j = c_1 \boldsymbol{v}_1 \cdot \boldsymbol{v}_j + \cdots + c_j \boldsymbol{v}_j \cdot \boldsymbol{v}_j + \cdots + c_k \boldsymbol{v}_k \cdot \boldsymbol{v}_j$$

All the dot products are 0 (due to orthogonality) except for the dot product with $\boldsymbol{x}$ and $\boldsymbol{v}_j$ leading to the formula:

$$c_j = \frac{\boldsymbol{x} \cdot \boldsymbol{v}_j}{\boldsymbol{v}_j \cdot \boldsymbol{v}_j}$$

And we see that this is the scalar projection of $\boldsymbol{x}$ onto $\boldsymbol{v}_j$. Recall the formula from Calc III:

$$\mathrm{Proj}_v(\mathbf{u}) = \frac{\mathbf{u} \cdot \mathbf{v}}{\mathbf{v} \cdot \mathbf{v}} \mathbf{v}$$

Therefore, we can think of the linear combination as the following, which simplifies if we use **orthonormal basis vectors**:

$$\begin{aligned} \mathbf{x} &= \mathrm{Proj}_{v_1}(\mathbf{x}) + \mathrm{Proj}_{v_2}(\mathbf{x}) + \cdots + \mathrm{Proj}_{v_k}(\mathbf{x}) \\ &= (\mathbf{x} \cdot \mathbf{v}_1)\mathbf{v}_1 + (\mathbf{x} \cdot \mathbf{v}_2)\mathbf{v}_2 + \cdots + (\mathbf{x} \cdot \mathbf{v}_k)\mathbf{v}_k \end{aligned} \tag{5.2}$$

IMPORTANT NOTE: In the event that $\mathbf{x}$ is NOT in $H$, then Equation 5.2 gives the (orthogonal) **projection of $\mathbf{x}$** into $H$.

## Projections

Consider the following example. If a matrix $U = [\mathbf{u}_1, \ldots, \mathbf{u}_k]$ has orthonormal columns (so if $U$ is $n \times k$, then that requires $k \leq n$), then

$$U^T U = \begin{bmatrix} \mathbf{u}_1^T \\ \vdots \\ \mathbf{u}_k \end{bmatrix} [\mathbf{u}_1, \ldots, \mathbf{u}_k] = \begin{bmatrix} \mathbf{u}_1^T \mathbf{u}_1 & \mathbf{u}_1^T \mathbf{u}_2 & \cdots & \mathbf{u}_1^T \mathbf{u}_k \\ \mathbf{u}_2^T \mathbf{u}_1 & \mathbf{u}_2^T \mathbf{u}_2 & \cdots & \mathbf{u}_2^T \mathbf{u}_k \\ \vdots & & & \vdots \\ \mathbf{u}_k^T \mathbf{u}_1 & \mathbf{u}_k^T \mathbf{u}_2 & \cdots & \mathbf{u}_k^T \mathbf{u}_k \end{bmatrix} = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & & & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix} = I_k$$

But $UU^T$ (which is $n \times n$) is NOT the identity if $k \neq n$ (If $k = n$, then the previous computation proves that the inverse is the transpose).

Here is a computation one might make for $UU^T$ (these are OUTER products):

$$UU^T = [\mathbf{u}_1, \ldots, \mathbf{u}_k] \begin{bmatrix} \mathbf{u}_1^T \\ \vdots \\ \mathbf{u}_k \end{bmatrix} = \mathbf{u}_1 \mathbf{u}_1^T + \mathbf{u}_2 \mathbf{u}_2^T + \cdots + \mathbf{u}_k \mathbf{u}_k^T$$

**Example 5.1.2.** Consider the following computations:

$$U = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \qquad U^T U = 1 \qquad UU^T = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$$

If $UU^T$ is not the identity, what is it? Consider the following computation:

$$\begin{aligned} UU^T \mathbf{x} &= \mathbf{u}_1 \mathbf{u}_1^T \mathbf{x} + \mathbf{u}_2 \mathbf{u}_2^T \mathbf{x} + \cdots + \mathbf{u}_k \mathbf{u}_k^T \mathbf{x} \\ &= \mathbf{u}_1 (\mathbf{u}_1^T \mathbf{x}) + \mathbf{u}_2 (\mathbf{u}_2^T \mathbf{x}) + \cdots + \mathbf{u}_k (\mathbf{u}_k^T \mathbf{x}) \end{aligned}$$

which we recognize as the projection of $\mathbf{x}$ into the space spanned by the orthonormal vectors of $U$. In summary, we can think of: the following matrix form for the coordinates:

$$[\boldsymbol{x}]_U = \boldsymbol{c} = U^T \boldsymbol{x}.$$

and the projection matrix $P$ that takes a vector $\mathbf{x}$ and produces the projection of $\mathbf{x}$ into the space spanned by the orthonormal columns of $U$ is

$$P = UU^T$$

## Exercises

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. (NOTE: The basis vectors are NOT orthonormal)

$$\boldsymbol{v}_1 = \begin{bmatrix} 1 \\ 2 \\ -1 \end{bmatrix} \quad \boldsymbol{v}_2 = \begin{bmatrix} 2 \\ -1 \\ 1 \end{bmatrix} \quad \boldsymbol{x}_1 = \begin{bmatrix} 7 \\ 4 \\ 0 \end{bmatrix} \quad \boldsymbol{x}_2 = \begin{bmatrix} 4 \\ 3 \\ -1 \end{bmatrix}$$

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

$$H = \left\{ \alpha_1 \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} + \alpha_2 \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix} \text{ such that } \alpha_1, \alpha_2 \in \mathbb{R} \right\}$$

is isormorphic to $\mathbb{R}^2$.

3. Let the subspace $G$ be the plane defined below, and consider the vector $\boldsymbol{x}$, where:

$$G = \left\{ \alpha_1 \begin{bmatrix} 1 \\ 3 \\ -2 \end{bmatrix} + \alpha_2 \begin{bmatrix} 3 \\ -1 \\ 0 \end{bmatrix} \text{ such that } \alpha_1, \alpha_2 \in \mathbb{R} \right\} \qquad \boldsymbol{x} = \begin{bmatrix} 1 \\ 0 \\ 2 \end{bmatrix}$$

   (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 $\boldsymbol{x}$.

4. 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)

5. 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 $\boldsymbol{x}$?

6. 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 $\boldsymbol{a}$.

7. To prove that we have an *orthogonal* projection, the vector $\text{Proj}_u(\boldsymbol{x}) - \boldsymbol{x}$ should be orthogonal to $\boldsymbol{u}$. Use this definition to show that our earlier formula was correct- that is,

$$\text{Proj}_u(\boldsymbol{x}) = \frac{\boldsymbol{x} \cdot \boldsymbol{u}}{\boldsymbol{u} \cdot \boldsymbol{u}} \boldsymbol{u}$$

is the orthogonal projection of $\mathbf{x}$ onto $\mathbf{u}$.

8. Continuing with the last exercise, show that $UU^T\mathbf{x}$ is the *orthogonal* projection of $\mathbf{x}$ into the space spanned by the columns of $U$ by showing that $(UU^T\boldsymbol{x} - \boldsymbol{x})$ is orthogonal to $\mathbf{u}_i$ for any $i = 1, 2, \cdots, k$.

## 5.2 The Four Fundamental Subspaces

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

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

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 5.2.1. 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,

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

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

$$\text{Null}(A) = \{ \boldsymbol{x} \in \mathbb{R}^n \,|\, A\boldsymbol{x} = \boldsymbol{0} \}$$

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

$$\text{Col}(A) = \{ \boldsymbol{y} \in \mathbb{R}^m \,|\, y = A\boldsymbol{x} \quad \in \mathbb{R}^n \}$$

  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 5.2.1.** *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:** We'll prove the first statement, the second statement is almost identical to the first. To prove the first statement, we have to show that if we take any vector $\mathbf{x}$ from nullspace of $A$ and any vector $\mathbf{y}$ from the row space of $A$, then $\mathbf{x} \cdot \mathbf{y} = 0$.

Alternatively, if we can show that $\mathbf{x}$ is orthogonal to each and every row of $A$, then we're done as well (since $\mathbf{y}$ is a linear combination of the rows of $A$).

In fact, now we see a strategy: Write out what it means for $\mathbf{x}$ to be in the nullspace using the rows of $A$. For ease of notation, let $\mathbf{a}_j$ denote the $j^{\text{th}}$ **row** of $A$, which will have size $1 \times n$. Then:

$$A\mathbf{x} = \mathbf{0} \quad \Rightarrow \quad \begin{bmatrix} \mathbf{a}_1 \\ \mathbf{a}_2 \\ \vdots \\ \mathbf{a}_m \end{bmatrix} \mathbf{x} = \begin{bmatrix} \mathbf{a}_1\mathbf{x} \\ \mathbf{a}_2\mathbf{x} \\ \vdots \\ \mathbf{a}_m\mathbf{x} \end{bmatrix} = \mathbf{0}$$

Therefore, the dot product between any row of $A$ and $\mathbf{x}$ is zero, so that $\mathbf{x}$ is orthogonal to every row of $A$. Therefore, $\mathbf{x}$ must be orthogonal to any linear combination of the rows of $A$, so that $\mathbf{x}$ is orthogonal to the row space of $A$. $\Box$

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 5.2.1.** Construct a basis for the column space, row space and nullspace of the matrix $A$ below that is row equivalent to the matrix beside it, RREF($A$):

$$A = \begin{bmatrix} 2 & 0 & -2 & 2 \\ -2 & 5 & 7 & 3 \\ 3 & -5 & -8 & -2 \end{bmatrix} \qquad \text{RREF}(A) = \begin{bmatrix} 1 & 0 & -1 & 1 \\ 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

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

$$\text{Col}(A) = \text{span}\left\{ \begin{bmatrix} 2 \\ -2 \\ 3 \end{bmatrix}, \begin{bmatrix} 2 \\ -2 \\ 3 \end{bmatrix} \right\}$$

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:

$$\text{Row}(A) = \text{span}\left\{ \begin{bmatrix} 1 \\ 0 \\ -1 \\ 1 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 1 \\ 1 \end{bmatrix} \right\} \qquad \text{Null}(A) = \text{span}\left\{ \begin{bmatrix} 1 \\ -1 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} -1 \\ -1 \\ 0 \\ 1 \end{bmatrix} \right\}$$

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 5.2.2.** 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

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

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 5.2.2. 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.

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

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.

## 5.3 Exercises

1. Show that $\text{Null}(A^T) \perp \text{Col}(A)$. Hint: You may use what we already proved.
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 $\{\boldsymbol{x}_i\}$

$$\|\sum_{i=1}^{n} \boldsymbol{x}_i\|_2^2 = \sum_{i=1}^{n} \|\boldsymbol{x}_i\|_2^2$$

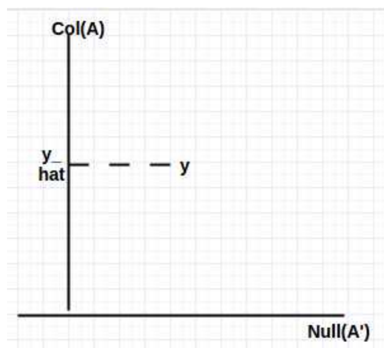
The case where $n = 1$ is trivial, so you might look at $n = 2$ first. Try starting with

$$\|\mathbf{x} + \mathbf{y}\|^2 = (\mathbf{x} + \mathbf{y})^T(\mathbf{x} + \mathbf{y}) = \cdots$$

and then simplify to get $\|\mathbf{x}\|^2 + \|\mathbf{y}\|^2$. Now try the induction step on your own.

5. Let $A$ be an $m \times n$ matrix where $m > n$, and let $A$ have rank $n$. Let $\boldsymbol{y}, \hat{\boldsymbol{y}} \in \mathbb{R}^m$, such that $\hat{\boldsymbol{y}}$ is the orthogonal projection of $\boldsymbol{y}$ onto the column space of $A$. We want a formula for the matrix $\mathbb{P} : \mathbb{R}^m \to \mathbb{R}^m$ so that $\mathbb{P}\boldsymbol{y} = \hat{\boldsymbol{y}}$.

The following image shows the relevant subspaces:




(a) Why is the projector not $\mathbb{P} = AA^T$?

(b) Since $\hat{\boldsymbol{y}} - \boldsymbol{y}$ is orthogonal to the column space of $A$, show that

$$A^T(\hat{\boldsymbol{y}} - \boldsymbol{y}) = \mathbf{0} \tag{5.3}$$

(c) Show that there exists $\boldsymbol{x} \in \mathbb{R}^n$ so that Equation (5.3) can be written as:

$$A^T(A\boldsymbol{x} - \boldsymbol{y}) = 0 \tag{5.4}$$

(d) Argue that $A^TA$ (which is $n \times n$) is invertible, so that Equation (5.4) implies that

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

(e) Finally, show that this implies that

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

Note: If $A$ has rank $k \neq n$, then we will need something different, since $A^TA$ 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 (non-zero) 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 $\{\boldsymbol{u}_i\}_{i=1}^{p}$ be an orthonormal basis for $W$, define $\boldsymbol{w} = (\boldsymbol{x} \cdot \boldsymbol{u}_1)\boldsymbol{u}_1 + \ldots + (\boldsymbol{x} \cdot \boldsymbol{u}_p)\boldsymbol{u}_p$, 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 $W$.

## 5.4 The Decomposition Theorems

The matrix factorization that arises from an eigenvector/eigenvalue decomposition is useful in many applications, so we'll briefly review it here and build from it until we get to our main goal, the Singular Value Decomposition.

### 5.4.1 The Eigenvector/Eigenvalue Decomposition

First we have a basic definition:

Let $A$ be an $n \times n$ matrix. If there exists a scalar $\lambda$ and non-zero vector $\boldsymbol{v}$ so that

$$A\boldsymbol{v} = \lambda \boldsymbol{v}$$

then we say that $\lambda$ is an eigenvalue and $\boldsymbol{v}$ is an associated eigenvector.

An equivalent formulation of the problem is to solve $A\mathbf{v} - \mathbf{v} = \mathbf{0}$, or, factoring $\mathbf{v}$ out,

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

This equation always has the zero solution (letting $\mathbf{v} = \mathbf{0}$), however, we need to have non-trivial solutions, and the only way that will happen is if $A - \lambda I$ is non-invertible, or equivalently,

$$\det(A - \lambda I) = 0$$

which, when multiplied out, is a polynomial equation in $\lambda$ that is called the **characteristic equation**.

Therefore, we find the eigenvalues first, then for each $\lambda$, there is an associated subspace- The null space of $A - \lambda I$, or the **eigenspace** associated with $\lambda$, denoted by $E_\lambda$.

The way to finish the problem is to give a "nice" basis for the eigenspace- If you're working by hand, try one with integer values. If you're on the computer, it is often convenient to make them unit vectors.

Some vocabulary associated with eigenvalues: Solving the characteristic equation will mean that we can have repeated solutions. The number of repetitions is the *algebraic multiplicity* of $\lambda$. On the other hand, for each $\lambda$, we find the eigenspace which will have a certain dimension- The dimension of the eigenspace is the *geometric multiplicity* of $\lambda$.

**Examples:**

1. Compute the eigenvalues and eigenvectors for the $2 \times 2$ identity matrix.

   SOLUTION: The eigenvalue is 1 (a double root), so the algebraic multiplicity of 1 is 2.

   On the other hand, if we take $A - \lambda I$, we simply get the zero matrix, which implies that every vector in $\mathbb{R}^2$ is an eigenvector. Therefore, we can take any basis of $\mathbb{R}^2$ is a basis for $E_1$, and the geometric multiplicity is 2.

2. Consider the matrix
$$\begin{bmatrix} 1 & 2 \\ 0 & 1 \end{bmatrix}$$
Again, the eigenvalue 1 is a double eigenvalue (so the algebraic multiplicity is 2), but solving $(A-\lambda I)\boldsymbol{v} = 0$ gives us:
$$2v_2 = 0 \quad \Rightarrow \quad v_2 = 0$$
That means $v_1$ is free, and the basis for $E_1$ is $[1, 0]^T$. Therefore, the algebraic multiplicity is 1.

**Definition:** A matrix for which the algebraic and geometric multiplicities are not equal is called *defective*.

There is a nice theorem relating eigenvalues:
**Theorem:** If $X$ is square and invertible, then $A$ and $X^{-1}AX$ have the same eigenvalues.
Sometimes this method of characterizing eigenvalues in terms of the determinant and trace of a matrix:
$$\det(A) = \Pi_{i=1}^{n} \lambda_i \quad \text{trace}(A) = \sum_{i=1}^{\infty} \lambda_i$$

## Symmetric Matrices and the Spectral Theorem

There are some difficulties working with eigenvalues and eigenvectors of a general matrix. For one thing, they are only defined for square matrices, and even when they are defined, we may get real or complex eigenvalues.

If a matrix is symmetric, beautiful things happen with the eigenvalues and eigenvectors, and it is summarized below in the Spectral Theorem.

**The Spectral Theorem:** If $A$ is an $n \times n$ symmetric matrix, then:

1. $A$ has $n$ real eigenvalues (counting multiplicity).
2. For each distinct $\lambda$, the algebraic and geometric multiplicities are the same.
3. The eigenspaces are mutually orthogonal- both for distinct eigenvalues, and we'll take each $E_\lambda$ to have an orthonormal basis.
4. $A$ is orthogonally diagonalizeable, with $D = \text{diag}(\lambda_1, \lambda_2, \ldots, \lambda_n)$. That is, if $V$ is the matrix whose columns are the (othornormal) eigenvectors of $A$, then
$$A = VDV^T$$

Some remarks about the Spectral Theorem:

- If a matrix is real and symmetric, the Spectral Theorem says that its eigenvectors form an orthonormal basis for $\mathbb{R}^n$.
- The first part is somewhat difficult to prove in that we would have to bring in more machinery than we would like. If you would like to see a proof, it comes from the *Schur Decomposition*, which is given, for example, in "Matrix Computations" by Golub and Van Loan.

The following is a proof of the third part. Supply justification for each step: Let $\boldsymbol{v}_1$, $\boldsymbol{v}_2$ be eigenvectors from distinct eigenvalues, $\lambda_1, \lambda_2$. We show that $\boldsymbol{v}_1 \cdot \boldsymbol{v}_2 = 0$:
$$\lambda_1 \boldsymbol{v}_1 \cdot \boldsymbol{v}_2 = (A\boldsymbol{v}_1)^T \boldsymbol{v}_2 = \boldsymbol{v}_1^T A^T \boldsymbol{v}_2 = \boldsymbol{v}_1^T A \boldsymbol{v}_2 = \lambda_2 \boldsymbol{v}_1 \cdot \boldsymbol{v}_2$$
Now, $(\lambda_1 - \lambda_2)\boldsymbol{v}_1 \cdot \boldsymbol{v}_2 = 0$.

**The Spectral Decomposition:** Since $A$ is orthogonally diagonalizable, then

$$A = (\boldsymbol{q}_1\ \boldsymbol{q}_2\ \dots\ \boldsymbol{q}_n) \begin{pmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \lambda_n \end{pmatrix} \begin{pmatrix} \boldsymbol{q}_1^T \\ \boldsymbol{q}_2^T \\ \vdots \\ \boldsymbol{q}_n^T \end{pmatrix}$$

so that:

$$A = (\lambda_1 \boldsymbol{q}_1\ \lambda_2 \boldsymbol{q}_2\ \dots\ \lambda_n \boldsymbol{q}_n) \begin{pmatrix} \boldsymbol{q}_1^T \\ \boldsymbol{q}_2^T \\ \vdots \\ \boldsymbol{q}_n^T \end{pmatrix}$$

so finally:

$$A = \lambda_1 \boldsymbol{q}_1 \boldsymbol{q}_1^T + \lambda_2 \boldsymbol{q}_2 \boldsymbol{q}_2^T + \dots + \lambda_n \boldsymbol{q}_n \boldsymbol{q}_n^T$$

That is, $A$ is a sum of $n$ rank one matrices, each of which is a projection matrix.

## Exercises

1. Prove that if $X$ is invertible and matrix $A$ is square, then $A$ and $X^{-1}AX$ have the same eigenvalues.
   **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).

### 5.4.2 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 gives orthonormal bases for all four matrix subspaces.

In what follows, assume that $A$ is an $m \times n$ matrix (so $A$ maps $\mathbb{R}^n$ to $\mathbb{R}^m$ and is not square).

1. Although $A$ itself is not symmetric, $A^TA$ is $n \times n$ and symmetric, so the Spectral Theorem applies.

   In particular, $A^TA$ is orthogonally diagonalizable. Let $\{\lambda_i\}_{i=1}^n$ and $V = [\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_n]$ be the eigenvalues and orthonormal eigenvectors so that

   $$A^TA = VDV^T$$

   and $D$ is the diagonal matrix with $\lambda_i$ along the diagonal.

2. **Exercise:** Show that $\lambda_i \geq 0$ for $i = 1..n$ by showing that $\|A\boldsymbol{v}_i\|_2^2 = \lambda_i$.

   As a starting point, you might rewrite

   $$\|A\boldsymbol{v}_i\|^2 = (A\boldsymbol{v})^T A\boldsymbol{v}$$

3. **Definition:** We define the singular values of $A$ by:

   $$\sigma_i = \sqrt{\lambda_i}$$

   where $\lambda_i$ is an eigenvalue of $A^TA$.

4. In the rest of the section, we will assume any list (or diagonal matrix) of eigenvals of $A^TA$ (or singular values of $A$) will be ordered from highest to lowest: $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_n$.

5. **Exercise:** Prove that, if $\boldsymbol{v}_i$ and $\boldsymbol{v}_j$ are distinct eigenvectors of $A^TA$, then their corresponding images, $A\boldsymbol{v}_i$ and $A\boldsymbol{v}_j$, are orthogonal.

6. **Exercise:** Prove that, if $\boldsymbol{x} = \alpha_1\boldsymbol{v}_1 + \ldots \alpha_n\boldsymbol{v}_n$, then

   $$\|A\boldsymbol{x}\|^2 = \alpha_1^2\lambda_1 + \ldots + \alpha_n^2\lambda_n$$

7. **Exercise:** Let $W$ be the subspace generated by the basis $\{\boldsymbol{v}_j\}_{j=k+1}^n$, where $\boldsymbol{v}_j$ are the eigenvectors associated with the *zero* eigenvalues of $A^TA$ (therefore, we are assuming that the first $k$ eigenvalues are NOT zero). Show that $W = \text{Null}(A)$.

   Hint: To show this, take an arbitrary vector $\boldsymbol{x}$ from $W$. Rather than showing directly that $A\boldsymbol{x} = \mathbf{0}$, instead show that the magnitude of $A\boldsymbol{x}$ is zero. We also need to show that if we take any vector from the nullspace of $A$, then it is also in $W$.

8. **Exercise:** Prove that if the rank of $A^TA$ is $r$, then so is the rank of $A$.

   Hint: How does the previous exercise help?

9. Define the columns of a matrix $U$ as:

   $$\boldsymbol{u}_i = \frac{1}{\|A\boldsymbol{v}_i\|_2} A\boldsymbol{v}_i = \frac{1}{\sigma_i} A\boldsymbol{v}_i$$

   and let $U$ be the matrix whose $i^{\text{th}}$ column is $\boldsymbol{u}_i$.

   We note that 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 $\mathbb{R}^m$, which can be done using Gram-Schmidt.

10. **Exercise:** Show that $\boldsymbol{u}_i$ is an eigenvector of $AA^T$ whose eigenvalue is also $\lambda_i$.

11. **Exercise:** Show that $A^T\boldsymbol{u}_i = \sigma_i\boldsymbol{v}_i$

12. 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 $A^TA$, and the matrix $U$ can be constructed using the $\boldsymbol{v}$'s or by finding the eigenvectors of $AA^T$.

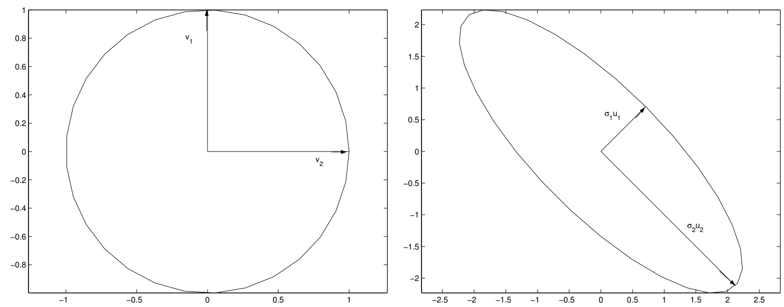



Figure 5.1: The geometric meaning of the right and left singular vectors of the SVD decomposition. Note that $A\boldsymbol{v}_i = \sigma_i \boldsymbol{u}_i$. The mapping $x \to Ax$ will map the unit circle on the left to the ellipse on the right.

13. **Exercise:** Let $A$ be $m \times n$. Define the $m \times n$ matrix

$$\Sigma = \operatorname{diag}(\sigma_1, \ldots, \sigma_n)$$

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

$$AV = U\Sigma$$

14. **Theorem: The Singular Value Decomposition (SVD)** Let $A$ be any $m \times n$ matrix of rank $r$. Then we can factor the matrix $A$ as the following product:

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

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*.

15. Keep in mind the following relationship between the right and left singular vectors:

$$\begin{aligned} A\boldsymbol{v}_i &= \sigma_i \boldsymbol{u}_i \\ A^T \boldsymbol{u}_i &= \sigma_i \boldsymbol{v}_i \end{aligned}$$

16. **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 $A$, Col($A$) is $\{\boldsymbol{u}_i\}_{i=1}^r$

    (b) A basis for nullspace of $A$, Null($A$) is $\{\boldsymbol{v}_i\}_{i=r+1}^n$

    (c) A basis for the rowspace of $A$, Row($A$) is $\{\boldsymbol{v}_i\}_{i=1}^r$

    (d) A basis for the nullspace of $A^T$, Null($A^T$) is $\{\boldsymbol{u}_i\}_{i=r+1}^m$

17. We can also visualize the right and left singular values as in Figure 5.1. 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 $\sigma_i \boldsymbol{u}_i$.

18. 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 for the nullspace of $A^T$, and of the row space. This is depicted in Figure 5.2.

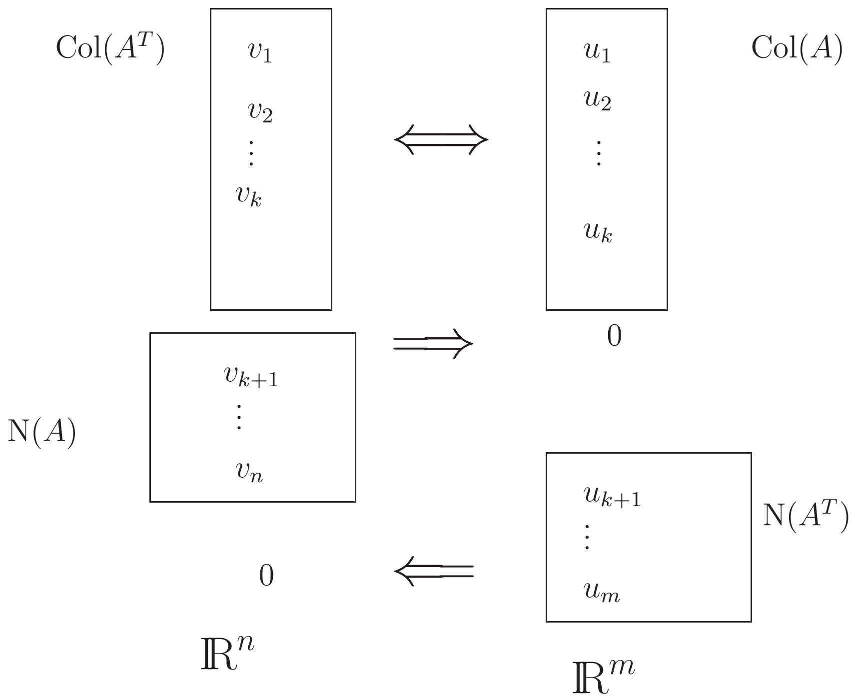



Figure 5.2: The SVD of A (`[U,S,V]=svd(A)`) completely and explicitly describes the 4 fundamental subspaces associated with the matrix, as shown. We have a one to one correspondence between the rowspace and columnspace of $A$, the remaining $\boldsymbol{v}$'s map to zero, and the remaining $\boldsymbol{u}$'s map to zero (under $A^T$).

19. 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*).

20. **Exercise:** Compute the SVD by hand of the following matrices:

$$\begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix} \qquad \begin{pmatrix} 0 & 2 \\ 0 & 0 \\ 0 & 0 \end{pmatrix}$$

21. **Remark:** If $m$ or $n$ is very large, it might not make sense to keep the full matrix $U$ and $V$.

22. **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.

23. **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.

24. **Matlab and the SVD** Matlab has the SVD built in. The function specifications are: `[U,S,V]=svd(A)` and `[U,S,V]=svd(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.

25. **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:

```
[U,S,V]=svd(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)'; image(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 $A$:

$$\hat{\boldsymbol{x}} = VV^T\boldsymbol{x}, \qquad \hat{\boldsymbol{y}} = UU^T\boldsymbol{y}$$

Now we can write:

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

so that

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

(Exercise: Verify that these computations are correct!)

Given an $m \times n$ matrix $A$, define its pseudoinverse, $A^\dagger$ by:

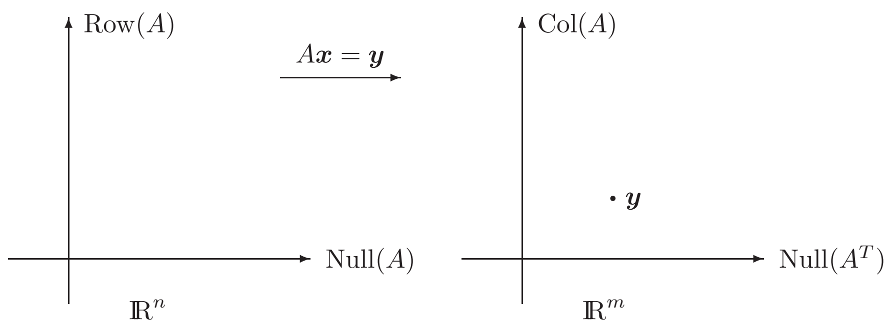
$$A^\dagger = V\Sigma^{-1}U^T$$

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

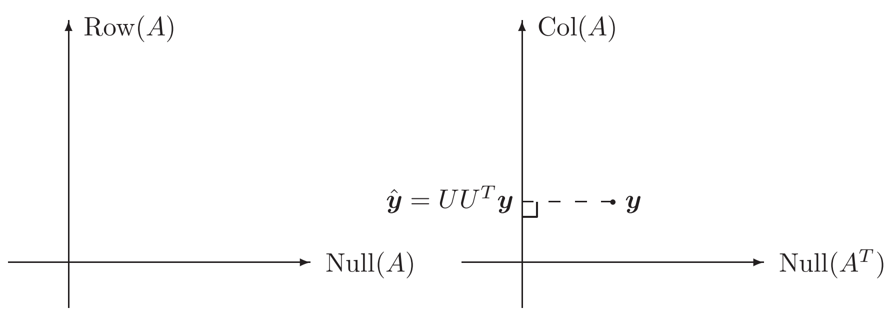
$$\hat{\boldsymbol{x}} = A^\dagger\boldsymbol{y}$$

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 $A$.

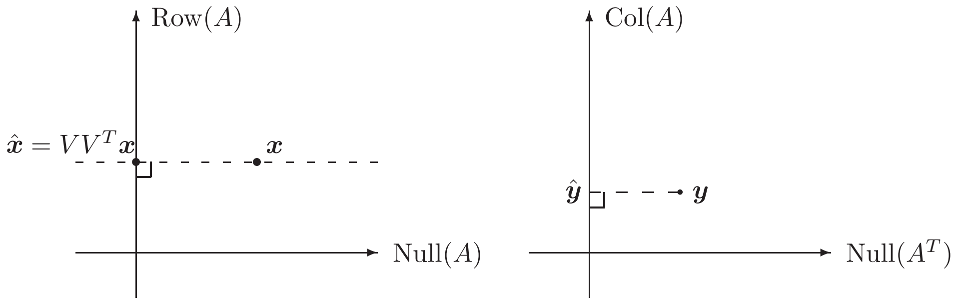
Geometrically, we can understand these computations in terms of the four fundamental subspaces.




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

$$\begin{bmatrix} 2 & 1 & -1 \\ 3 & 1 & 2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 5 \\ 1 \end{bmatrix}$$

   (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 $VV^T\boldsymbol{x}$?)

2. Consider

$$\begin{bmatrix} 2 & 1 & -1 & 3 \\ -1 & 0 & 1 & -2 \\ 7 & 2 & -5 & 12 \\ -3 & -2 & 0 & -4 \\ 4 & 1 & -3 & 7 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x4 \end{bmatrix} = \begin{bmatrix} 5 \\ 1 \\ 0 \\ -2 \\ 6 \end{bmatrix}$$

   (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 5.1:

```
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 $\pm\boldsymbol{u}$.*