Math 510 **Fall 2022**
Notes for Homework 3 **Drew Armstrong**

Contents

1 Subspaces Associated to Linear Functions

In the last section we discussed purely symbolic properties of matrix inversion. Recall: Let A be an $m \times n$ matrix. An $n \times m$ matrix B is called a *right inverse of* A when $AB = I_m$ and an $m \times n$ matrix C is called a *left inverse of* A when $CA = I_n$. If A has both a right inverse B and a left inverse C then the two must be equal because

$$B = I_nB = (CA)B = C(AB) = CI_m = C.$$

In this case we say that $A^{-1} = B = C$ is the unique *two-sided inverse of* A. Any matrix having a two-sided inverse is called *invertible*. We also proved the following basic facts: If A^{-1} exists then $(A^*)^{-1}$ exists and is equal to $(A^{-1})^*$. If A^{-1}, B^{-1} and AB exist then $(AB)^{-1}$ exists and is equal to $B^{-1}A^{-1}$.

Precisely when do inverse matrices exist? This question is surprisingly subtle. In order to answer it we must ascend to a higher level of abstraction. To each linear function between vector spaces $V \to W$ we associate certain subspaces of V and W.

1.1 Kernel and Image of a Linear Function.

Consider a linear function $f : V \to W$ between vector spaces.[1] We define the *kernel* and the *image* of f as follows:

$$ker(f) := \{\text{the set of } \mathbf{v} \in V \text{ such that } f(\mathbf{v}) = \mathbf{0}\},$$
$$im(f) := \{\text{the set of } \mathbf{w} \in W \text{ such that } \mathbf{w} = f(\mathbf{v}) \text{ for some } \mathbf{v} \in V\}.$$

Remark: The kernel and image of f are sometimes called the *nullspace* and *range*.[2]

We observe that $ker(f) \subseteq V$ **is a subspace**. Indeed, given vectors $\mathbf{v}_1, \ldots, \mathbf{v}_n \in ker(f)$ in the kernel and scalars $a_1, \ldots, a_n$, the linearity of f implies

$$f(a_1\mathbf{v}_1 + \cdots + a_n\mathbf{v}_n) = a_1 f(\mathbf{v}_1) + \cdots + a_n f(\mathbf{v}_n) = a_1\mathbf{0} + \cdots + a_n\mathbf{0} = \mathbf{0},$$

so the linear combination $a_1\mathbf{v}_1 + \cdots + a_n\mathbf{v}_n$ is also in the kernel. Furthermore, we observe that $im(f) \subseteq W$ **is a subspace**. Indeed, consider any vectors $\mathbf{w}_1, \ldots, \mathbf{w}_n \in im(f)$ in the image and any scalars $a_1, \ldots, a_n$. By definition we can write $\mathbf{w}_i = f(\mathbf{v}_i)$ for some vectors $\mathbf{v}_i$, hence from the linearity of f we have

$$a_1\mathbf{w}_1 + \cdots + a_n\mathbf{w}_n = a_1 f(\mathbf{v}_1) + \cdots + a_n f(\mathbf{v}_n) = f(a_1\mathbf{v}_1 + \cdots + a_n\mathbf{v}_n).$$

Since $a_1\mathbf{w}_1 + \cdots + a_n\mathbf{w}_n = f(\mathbf{v}')$ for some vector $\mathbf{v}'$ we conclude that the linear combination $a_1\mathbf{w}_1 + \cdots + a_n\mathbf{w}_n$ is also in the image.

The invertibility of a linear function is closely related to its kernel and image. The first observation is true by definition of the words *image* and *surjective*:[3]

$$f : V \to W \text{ is surjective if and only if } im(f) = W.$$

[1] Over $\mathbb{R}$ or $\mathbb{C}$; it doesn't matter. Indeed, the same theory applies to vector spaces over arbitrary fields.

[2] Kernel and image are standard terminology in abstract algebra. Nullspace and range are more common in applied linear algebra. For matrices, the image/range is often called the *column space*. (Too many words; I know.) See the next section.

[3] The words *surjective* and *injective* were introduced by Bourbaki in the 1940s. The older equivalent terms are *onto* and *one-to-one*.

The next observation requires a short proof:

$$f : V \to W \text{ is injective if and only if } ker(f) = \{\mathbf{0}\}.$$

Proof. Recall that any linear function satisfies $f(\mathbf{0}) = \mathbf{0}$. If f is injective then $f(\mathbf{v}) = \mathbf{0} = f(\mathbf{0})$ implies $\mathbf{v} = \mathbf{0}$, and hence $ker(f) = \{\mathbf{0}\}$. Conversely, suppose that $ker(f) = \{\mathbf{0}\}$. To show that f is injective, let $f(\mathbf{v}_1) = f(\mathbf{v}_2)$ for some vectors $\mathbf{v}_1, \mathbf{v}_2$. Then we have

$$\begin{aligned} f(\mathbf{v}_1) &= f(\mathbf{v}_2) && \\ f(\mathbf{v}_1) - f(\mathbf{v}_2) &= \mathbf{0} && \\ f(\mathbf{v}_1 - \mathbf{v}_2) &= \mathbf{0} && \text{linearity of } f \\ \mathbf{v}_1 - \mathbf{v}_2 &= \mathbf{0} && ker(f) = \{\mathbf{0}\} \\ \mathbf{v}_1 &= \mathbf{v}_2. && \end{aligned}$$

Hence f is injective. □

1.2 Isomorphism of Vector Spaces.

Let $f : V \to W$ be a function between vector spaces. We say that f is an *isomorphism*[4] when the following properties are satisfied:

(a) f is linear,

(b) f is surjective,

(c) f is injective.

Properties (b) and (c) say that f is a *bijection*,[5] which is equivalent to being invertible. Furthermore, one can check that the inverse function $f^{-1} : W \to V$ is also linear. If there exists an isomorphism between vector spaces V and W then we will write

$$V \cong W.$$

When V and W are finite dimensional we have the following important fact:

Isomorphism of Finite Dimensional Vector Spaces.

$$\boxed{V \cong W \quad \iff \quad \dim(V) = \dim(W).}$$

Proof. $\Longrightarrow$: Suppose that $V \cong W$ and let $f : V \to W$ be a specific isomorphism. Suppose that $\dim(V) = n$ and let $\{\mathbf{b}_1, \ldots, \mathbf{b}_n\}$ be a basis for V. Then I claim that $\{f(\mathbf{b}_1), \ldots, f(\mathbf{b}_n)\}$ is a basis for W, from which it will follow that $\dim(W) = n$. There are two things to show:

[4] Also called a linear isomorphism, or an isomorphism of vector spaces.

[5] Another Bourbaki term. The older word is *one-to-one correspondence*.

- **Independent.** Suppose that $a_1 f(\mathbf{b}_1) + \cdots + a_n f(\mathbf{b}_n) = \mathbf{0}$ for some scalars $a_1, \ldots, a_n$. Linearity of f implies that

$$\mathbf{0} = a_1 f(\mathbf{b}_1) + \cdots + a_n f(\mathbf{b}_n) = f(a_1 \mathbf{b}_1 + \cdots + a_n \mathbf{b}_n),$$

  and then the fact that f is injective implies that

$$\mathbf{0} = a_1 \mathbf{b}_1 + \cdots + a_n \mathbf{b}_n.$$

  Finally, the fact that $\{\mathbf{b}_1, \ldots, \mathbf{b}_n\}$ is independent implies that $a_1 = \cdots = a_n = 0$.

- **Spanning.** Consider any vector $\mathbf{w} \in W$. Since f is surjective we have $\mathbf{w} = f(\mathbf{v})$ for some $\mathbf{v} \in V$, and since $\{\mathbf{b}_1, \ldots, \mathbf{b}_n\}$ spans $\mathbf{v}$ we can write

$$\mathbf{v} = a_1 \mathbf{b}_1 + \cdots + a_n \mathbf{b}_n$$

  for some scalars $a_1, \ldots, a_n$. Finally, by linearity of f we have

$$\mathbf{w} = f(\mathbf{v}) = f(a_1 \mathbf{b}_1 + \cdots + a_n \mathbf{b}_n) = a_1 f(\mathbf{b}_1) + \cdots + a_n f(\mathbf{b}_n),$$

  which shows that $\{f(\mathbf{b}_1), \ldots, f(\mathbf{b}_n)\}$ spans W.

$\Longleftarrow$: Suppose that $\dim(V) = \dim(W) = n$. Choose bases $\mathbf{v}_1, \ldots, \mathbf{v}_n \in V$ and $\mathbf{w}_1, \ldots, \mathbf{w}_n \in W$ and define a linear function $f : V \to W$ by sending $\mathbf{v}_i \mapsto \mathbf{w}_i$ for all i. Then for any vector $\mathbf{v} = a_1 \mathbf{v}_1 + \cdots + a_n \mathbf{v}_n \in V$ we have

$$f(a_1 \mathbf{v}_1 + \cdots + a_n \mathbf{v}_n) = a_1 f(\mathbf{v}_1) + \cdots + a_n f(\mathbf{v}_n) = a_1 \mathbf{w}_1 + \cdots + a_n \mathbf{w}_n.$$

Furthermore, the function $f^{:-1}: W \to V$ defined by sending $\mathbf{w}_i \mapsto \mathbf{v}_i$ is the inverse of f:

$$f^{-1}(a_1 \mathbf{w}_1 + \cdots + a_n \mathbf{w}_n) = a_1 f^{-1}(\mathbf{w}_1) + \cdots + a_n f^{-1}(\mathbf{w}_n) = a_1 \mathbf{v}_1 + \cdots + a_n \mathbf{v}_n.$$

□

As a consequence of this theorem, any n-dimensional vector space over $\mathbb{R}$ is isomorphic to $\mathbb{R}^n$. Indeed, let $\mathbf{v}_1, \ldots, \mathbf{v}_n \in V$ be a basis. Then the following function $V \to \mathbb{R}^n$ is an isomorphism:

$$a_1 \mathbf{v}_1 + \cdots + a_n \mathbf{v}_n \mapsto \begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix}.$$

We will apply these ideas in the next section.

2 Subspaces Associated to Matrices

Recall that an $m \times n$ matrix over $\mathbb{R}$ is the same thing as a linear function $\mathbb{R}^n \to \mathbb{R}^m$.[6] In this case the kernel and image have a special interpretation.

[6] When I write $\mathbb{R}^n$ I always assume that we are working with the standard basis.

2.1 The Nullspace of a Matrix.

Given an $m \times n$ matrix A we define the *nullspace*:

$$\mathcal{N}(A) := \{\text{the set of } \mathbf{x} \in \mathbb{R}^n \text{ such that } A\mathbf{x} = \mathbf{0}\}.$$

It is easy to check that $\mathcal{N}(A) \subseteq \mathbb{R}^m$ is a subspace. Indeed, $\mathcal{N}(A)$ is just the kernel of the linear function $A : \mathbb{R}^n \to \mathbb{R}^m$. More interestingly, we can use the concept of the nullspace to express the fact that a given vector is simultaneously orthogonal to a given set of vectors:

$$\boxed{\mathbf{x} \in \mathcal{N}(A) \quad \Longleftrightarrow \quad A\mathbf{x} = \mathbf{0} \quad \Longleftrightarrow \quad \mathbf{x} \text{ is orthogonal to every row of } A.}$$

Indeed, let $\mathbf{a}_i^T$ be the ith row vector of A. If $A\mathbf{x} = \mathbf{0}$ then we have

$$\begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix} = \mathbf{0} = A\mathbf{x} = \begin{pmatrix} - & \mathbf{a}_1^T & - \\ & \vdots & \\ - & \mathbf{a}_m^T & - \end{pmatrix} \mathbf{x} = \begin{pmatrix} \mathbf{a}_1^T\mathbf{x} \\ \vdots \\ \mathbf{a}_m^T\mathbf{x} \end{pmatrix}.$$

Comparing entries on the left and right gives $\mathbf{a}_i^T\mathbf{x} = 0$ for all i. In other words, the vector $\mathbf{x}$ is orthogonal to each row vector of A. Equivalently, we have

$$\boxed{A^T\mathbf{x} = \mathbf{0} \quad \Longleftrightarrow \quad \mathbf{x} \text{ is orthogonal to every column of } A.}$$

It is important to get comfortable with this idea because it is the foundation of least squares.[7]

2.2 The Column Space of a Matrix.

We can think of an $m \times n$ matrix A as a linear function $A : \mathbb{R}^n \to \mathbb{R}^m$. In this case the image of A is called the *column space*:

$$\mathcal{C}(A) = \{\text{the set of } A\mathbf{x} \in \mathbb{R}^m \text{ for all } \mathbf{x} \in \mathbb{R}^n\}.$$

But *why* is it called the column space? Let $\mathbf{a}_1, \ldots, \mathbf{a}_n \in \mathbb{R}^m$ be the column vectors of A. Then for any vector $\mathbf{x} = (x_1, \ldots, x_n) \in \mathbb{R}^n$ we have

$$A\mathbf{x} = \left(\begin{array}{c|c|c} \mathbf{a}_1 & \cdots & \mathbf{a}_n \end{array} \right) \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = x_1\mathbf{a}_1 + \cdots + x_n\mathbf{a}_n,$$

which is a linear combination of the columns of A. So we can also write

$$\mathcal{C}(A) = \{\text{all linear combinations of the columns of } A\}.$$

[7] For the impatient: Let $P\mathbf{x}$ be the orthogonal projection of a point $\mathbf{x}$ onto the column space of a matrix A. Then the vector $P\mathbf{x} - \mathbf{x}$ must be orthogonal to every column of A, hence $A^T(P\mathbf{x} - \mathbf{x}) = \mathbf{0}$.

Similarly, we can define the *row space* of A:

$$\mathcal{R}(A) := \mathcal{C}(A^T) = \{\text{all linear combinations of the rows of } A\}.$$

Note that $\mathcal{C}(A)$ is a subspace of $\mathbb{R}^m$ because each column of an $m \times n$ matrix lives in $\mathbb{R}^m$, while $\mathcal{R}(A)$ is a subspace of $\mathbb{R}^n$. So the row space and column space cannot be directly compared.

2.3 Orthogonality of the Subspaces.

We observed above that $\mathbf{x} \in \mathcal{N}(A)$ if and only if $\mathbf{x}$ is orthogonal to every row of A. We can express this as follows:

$$\boxed{\mathcal{N}(A) = \mathcal{R}(A)^{\perp}.}$$

In general, given a subspace $U \subseteq V$ of an inner product space V we let $U^{\perp} \subseteq V$ denote the set of vectors that are orthogonal to every vector in U:[8]

$$U^{\perp} = \{\text{the set of } \mathbf{v} \in V \text{ such that } \langle \mathbf{u}, \mathbf{v} \rangle = 0 \text{ for all } \mathbf{u} \in U\}.$$

You will check on the homework that $U^{\perp} \subseteq V$ is also a subspace. Furthermore, if V finite dimensional then you will prove the following dimension formula:

$$\dim U + \dim U^{\perp} = \dim V.$$

In the case of the rowspace and nullspace of a matrix A we obtain the following theorem.

The Rank-Nullity Theorem. For any matrix A we have

$$\boxed{\dim \mathcal{R}(A) + \dim \mathcal{N}(A) = \text{the number of columns of } A.}$$

Indeed, if A is $m \times n$ then $\mathcal{R}(A)$ and $\mathcal{N}(A)$ are orthogonal subspaces of $\mathbb{R}^n$, so that

$$\dim \mathcal{R}(A) + \dim \mathcal{N}(A) = n.$$

This is often called the *rank-nullity theorem* because $\dim \mathcal{R}(A)$ is called the *rank* and $\dim \mathcal{N}(A)$ is called the *nullity* of the matrix A.[9] By replacing A with A^T we obtain the equivalent formula

$$\dim \mathcal{C}(A) + \dim \mathcal{N}(A^T) = m,$$

which does not have a nice name.

[8] We read $U^{\perp}$ as "U perp".

[9] The dimension of $\mathcal{C}(A)$ is also called the rank of A. The fact that $\mathcal{R}(A)$ and $\mathcal{C}(A)$ have the same dimension is a deep fact called the Fundamental Theorem. See the next section.

3 The Fundamental Theorem

In this section we will prove the most important theorem about matrices. Following Gilbert Strang, I will call this "The Fundamental Theorem".

The Fundamental Theorem of Linear Algebra. For any $m \times n$ matrix A we have

$$\dim \mathcal{R}(A) = \dim \mathcal{C}(A).$$

This common dimension is called the *rank* of A, sometimes written $\operatorname{rank}(A)$.

This result is a bit surprising because the row space $\mathcal{R}(A)$ lives in $\mathbb{R}^n$, while the column space $\mathcal{C}(A)$ lives in $\mathbb{R}^m$, so there is no direct way to compare them. Evidently there is some subtle form of communication between the rows and columns of a matrix. We will see in the next section that the Fundamental Theorem implies the following facts:

- Invertible matrices are square.
- If A and B are square of the same size, then $AB = I$ if and only if $BA = I$.
- If A is square then A has orthonormal columns if and only if it has orthonormal rows.

The proof is more difficult than you might expect, but it is worth going through the details because the ideas in the proof quite useful. There are two main steps:

(1) Let E and F be any matrices such that E has a left inverse $E'E = I$ and F has a right inverse $FF' = I$.[10] Then we will show that

$$\dim \mathcal{R}(EAF) = \dim \mathcal{R}(A) \quad \text{and} \quad \dim \mathcal{C}(EAF) = \dim \mathcal{C}(A).$$

(2) For any matrix A, we will find matrices E and F, as in (1), so that EAF has the following simple form:

$$EAF = \left(\begin{array}{c|c} I_r & O_{r,n-r} \\ \hline O_{m-r,r} & O_{m-r,n-r} \end{array}\right),$$

where I_r is the square $r \times r$ identity matrix. Since the matrix on the right clearly has row space and column space of dimension r,[11] it will follow that

$$\dim \mathcal{R}(A) = \dim \mathcal{R}(EAF) = r = \dim \mathcal{C}(EAF) = \dim \mathcal{C}(A).$$

Aside from these two main steps, we will further organize the proof into substeps, labeled by (a), (b), etc., since there are many details.

[10] These one-sided inverses need not be unique.

[11] The first r rows are a basis for the row space, while the first r columns are a basis for the column space.

Proof of Step (1).

(a) For any matrix E such that EA exists, we have

$$\mathcal{R}(EA) \subseteq \mathcal{R}(A).$$

Indeed, I claim that each row of EA is a linear combination of the rows of A. To see this, let E have ith row $(e_{i1}, \ldots, e_{im})$ and let A have ith row $\mathbf{a}_i^T$. Then

$$\begin{aligned}
(i\text{th row of } EA) &= (i\text{th row of } E)A \\
&= \begin{pmatrix} e_{i1} & \cdots & e_{im} \end{pmatrix} A \\
&= \left(\begin{array}{c|c|c} e_{i1} & \cdots & e_{im} \end{array} \right) \left(\begin{array}{c} \mathbf{a}_1^T \\ \hline \vdots \\ \hline \mathbf{a}_m^T \end{array} \right) \\
&= e_{i1}\mathbf{a}_1^T + \cdots + e_{im}\mathbf{a}_m^T.
\end{aligned}$$

In the last step we used block multiplication. Since every row of EA is in the rowspace $\mathcal{R}(A)$ it follows that any linear combination of rows of EA is in $\mathcal{R}(A)$. In other words, $\mathcal{R}(EA) \subseteq \mathcal{R}(A)$.

(b) If E has a left inverse $E'E = I$ then we also have

$$\mathcal{R}(A) \subseteq \mathcal{R}(EA).$$

Indeed, applying step (a) to the matricx $B = EA$ and E' shows that

$$\mathcal{R}(A) = \mathcal{R}(E'EA) = \mathcal{R}(E'B) \subseteq \mathcal{R}(B) = \mathcal{R}(EA).$$

Then combining (a) and (b) shows that $\mathcal{R}(EA) = \mathcal{R}(A)$, hence

$$\dim \mathcal{R}(EA) = \dim \mathcal{R}(A).$$

(c) For any matrix F such that AF exists, we have

$$\mathcal{C}(A) \subseteq \mathcal{C}(AF).$$

Indeed, I claim that any column of AF is a linear combination of the columns of A. The proof is similar to part (a). Let $(f_{1j}, \ldots, f_{nj})$ be the jth column of F and let $\mathbf{a}_j$ be the jth column of A. Then we have

$$\begin{aligned}
(j\text{th column of } AF) &= A(j\text{th column of } F) \\
&= \left(\begin{array}{c|c|c} \mathbf{a}_1 & \cdots & \mathbf{a}_n \end{array} \right) \begin{pmatrix} f_{1j} \\ \vdots \\ f_{nj} \end{pmatrix}
\end{aligned}$$

$$= f_{1j}\mathbf{a}_1 + \cdots + f_{nj}\mathbf{a}_n.$$

(d) If F has a right inverse $FF' = I$, then applying (c) to the matrix $B = AF$ and F' gives

$$\mathcal{C}(AF) = \mathcal{C}(B) \subseteq \mathcal{C}(BF') = \mathcal{C}(AFF') = \mathcal{C}(A),$$

hence $\mathcal{C}(AF) = \mathcal{C}(A)$. It follows that

$$\dim \mathcal{C}(AF) = \dim \mathcal{C}(A).$$

Next we will show that $\dim \mathcal{R}(A) = \dim \mathcal{R}(AF)$ and $\dim \mathcal{C}(EA) = \dim \mathcal{C}(A)$. This time the corresponding spaces are **not equal**, but they are still **isomorphic**.

(e) For any matrix A with rows $\mathbf{a}_i^T$ and any matrix F of appropriate shape, note that

$$(i\text{th row of } AF) = (i\text{th row of } A)F = \mathbf{a}_i^T F.$$

Consider the function $\varphi : \mathcal{R}(A) \to \mathcal{R}(AF)$ defined by multiplying on the right by F. That is, for any vector[12] $\mathbf{b}^T = b_1\mathbf{a}_1^T + \cdots + b_m\mathbf{a}_m^T \in \mathcal{R}(A)$ we define

$$\begin{aligned}\varphi(\mathbf{b}^T) &:= \mathbf{b}^T F \\ &= \varphi(b_1\mathbf{a}_1^T + \cdots + b_m\mathbf{a}_m^T)F \\ &= b_1(\mathbf{a}_1^T F) + \cdots + b_m(\mathbf{a}_m^T F) \in \mathcal{R}(AF).\end{aligned}$$

Matrix multiplication is linear, so φ is a linear function. Next, for any vector

$$\mathbf{c}^T := c_1(\mathbf{a}_1^T F) + \cdots + c_m(\mathbf{a}_m^T F) \in \mathcal{R}(AF)$$

we have

$$\mathbf{c}^T = \varphi(c_1\mathbf{a}_1^T + \cdots + c_n\mathbf{a}_m^T),$$

so that φ is surjective. Finally, since F has a right inverse $FF' = I$ we see that φ is injective. Indeed, if $\varphi(\mathbf{b}^T) = \varphi(\mathbf{c}^T)$ then

$$\begin{aligned}\varphi(\mathbf{b}^T) &= \varphi(\mathbf{c}^T) \\ \mathbf{b}^T F &= \mathbf{c}^T F \\ (\mathbf{b}^T F)F' &= (\mathbf{c}^T F)F' \\ \mathbf{b}^T(FF') &= \mathbf{c}^T(FF') \\ \mathbf{b}^T &= \mathbf{c}^T.\end{aligned}$$

Hence φ is an isomorphism $\mathcal{R}(A) \cong \mathcal{R}(AF)$, and it follows from the previous section that

$$\dim \mathcal{R}(AF) = \dim \mathcal{R}(A).$$

[12] Usually we think of $\mathcal{R}(A)$ as space of column vectors, but for the purpose of this proof it is more convenient to think of $\mathcal{R}(A)$ as space of row vectors.

(f) Similarly, if E has a left inverse $E'E = I$ then we will show that $\mathcal{C}(EA) \cong \mathcal{C}(A)$. To do this we consider the function $\psi : \mathcal{C}(A) \to \mathcal{C}(EA)$ defined by multiplying on the left by E. To be explicit, let $\mathbf{a}_j$ be the jth column of A,[13] so that

$$(j\text{th column of } EA) = E(j\text{th column of } A) = E\mathbf{a}_j.$$

Consider the function $\psi : \mathcal{C}(A) \to \mathcal{C}(EA)$ defined by multiplying on the left by E. That is, for any vector $\mathbf{b} = b_1\mathbf{a}_1 + \cdots + b_n\mathbf{a}_n \in \mathcal{C}(A)$ we define

$$\begin{aligned}\psi(\mathbf{b}) &:= E\mathbf{b} \\ &= E(b_1\mathbf{a}_1 + \cdots + b_n\mathbf{a}_n) \\ &= b_1(E\mathbf{a}_1) + \cdots + b_n(E\mathbf{a}_n) \in \mathcal{C}(EA).\end{aligned}$$

Following an argument similar to (e), we see that ψ is a vector space isomorphism, and hence

$$\dim \mathcal{C}(EA) = \dim \mathcal{C}(A).$$

Proof of Step (2). The proof of this step is an algorithm. For this purpose we introduce the important new idea of *elementary matrices*.

(g) Elementary Matrices. We define three families of **square** matrices.[14]

- For any index i and **nonzero** scalar λ we define

$$D_i(\lambda) = \begin{pmatrix} 1 & & & & \\ & 1 & & & \\ & & \lambda & & \\ & & & 1 & \\ & & & & 1 \end{pmatrix}.$$

  The main diagonal entries are 1, except of the ii entry, which is λ. The off-diagonal entries are all zero.

- For any indices $i \neq j$ and any scalar λ we define

$$L_{ij}(\lambda) = \begin{pmatrix} 1 & & & & \\ & 1 & \cdots & \lambda & \\ & & 1 & \vdots & \\ & & & 1 & \\ & & & & 1 \end{pmatrix}.$$

  The main diagonal entries are 1. The only other nonzero entry is λ in the ij position.

[13] In part (e) we used $\mathbf{a}_i^T$ for the ith row of A. Hopefully you don't mind that I'm recycling the notation $\mathbf{a}_j$ for a different purpose. Gilbert Strang uses $\mathbf{a}_i^*$ to denote rows of a matrix, but I don't like this because I use $*$ for conjugate transpose.

[14] It is always a struggle to find a notation for elementary matrices. Here I use the Wikipedia notation. I guess that D is for Diagonal, T is for Transposition and L is for Lower triangular, since many algorithms only use lower triangular $L_{ij}(\lambda)$ (i.e., with $i > j$). I prefer to think of D for Dilation and L for eLimination.

- For any indices $i \neq j$ we define

$$T_{ij} = \begin{pmatrix} 1 & & & & \\ & 0 & \cdots & 1 & \\ & \vdots & 1 & \vdots & \\ & 1 & \cdots & 0 & \\ & & & & 1 \end{pmatrix}.$$

  The main diagonal entries are 1 except for zeros in the ii and jj positions. The off-diagonal entries are zero except for 1 in the ij and ji positions.

We observe that each of these (square) *elementary matrices* is invertible. That is, we have

$$\begin{aligned} D_i(\lambda)^{-1} &= D_i(1/\lambda) \\ L_{ij}(\lambda)^{-1} &= L_{ij}(-\lambda) \\ T_{ij}^{-1} &= T_{ij}. \end{aligned}$$

But what are these matrices **for**?

(h) Row and Column Operations. Let A be an $m \times n$ matrix. For any matrix E we have seen that each row of EA is a linear combination of the rows of A. To be precise, if $(e_{i1}, \ldots, e_{im})$ is the ith row of E and $\mathbf{a}_i^T$ is the ith row of A, then

$$(i\text{th row of } EA) = e_{i1}\mathbf{a}_1^T + \cdots + e_{im}\mathbf{a}_m^T.$$

When E is an $m \times m$ elementary matrix then we have the following *elementary row operations*.

- The function $A \rightsquigarrow D_i(\lambda)A$ multiplies the ith row of A by λ.
- The function $A \rightsquigarrow L_{ij}(\lambda)A$ replaces the ith row of A by itself plus λ times the jth row of A. Indeed, when $k \neq i$, the kth row of $L_{ij}(\lambda)$ is just a standard basis vector, whereas the ith row of $L_{ij}(\lambda)$ is $(0, \ldots, 0, 1, 0, \ldots, 0, \lambda, 0, \ldots, 0)$ with 1 in the ith position and λ in the jth position. Hence the ith row of EA is

$$0\mathbf{a}_1^T + \cdots + 0\mathbf{a}_{i-1}^T + 1\mathbf{a}_i^T + 0\mathbf{a}_{i+1}^T + \cdots + 0\mathbf{a}_{j-1}^T + \lambda\mathbf{a}_j^T + 0\mathbf{a}_{j+1}^T + \cdots + 0\mathbf{a}_m^T.$$

- The function $A \rightsquigarrow T_{ij}A$ swaps the ith and jth rows of A.

Similarly, if F has jth column $(f_{1j}, \ldots, f_{nj})$ and A has jth column $\mathbf{a}_j$, then

$$(j\text{th column of } AF) = f_{1j}\mathbf{a}_1 + \cdots + f_{nj}\mathbf{a}_n.$$

When F is an elementary matrix then we have the following *elementary column operations*.

- The function $A \rightsquigarrow AD_i(\lambda)$ multiplies the ith column of A by λ.
- The function $A \rightsquigarrow AL_{ij}(\lambda)$ replaces the jth column of A by itself plus λ times the ith column of A. The proof is the same as for rows.

- The function $A \rightsquigarrow AT_{ij}$ swaps the ith and jth columns of A.

(i) The Algorithm. Finally, we can use elementary matrices to put the $m \times n$ matrix A into a particularly nice form. If $E_1, \ldots, E_k$ are elementary $m \times m$ matrices and if $F_1, \ldots, F_\ell$ are elementary $n \times n$ matrices then by performing row and column operations we will obtain

$$E_k \cdots E_1 E_1 A F_1 F_2 \cdots F_\ell = EAF.$$

Since elementary matrices are invertible, the products $E = E_k \cdots E_2 E_1$ and $F = F_1 F_2 \cdots F_\ell$ are also invertible. In particular, E has a left inverse and F has a right inverse, so we can apply the results from step (1).

Now we explain how to choose the operations.[15] If the top left entry of A is zero, swap rows or columns until it is not zero. Then scale the first row or column so the top left entry is equal to 1. Next apply elimination matrices $L_{ij}(\lambda)$ on both sides to eliminate the other entries in the first row and column. The result is a matrix of the form

$$\left(\begin{array}{c|ccc} 1 & 0 & \cdots & 0 \\ \hline 0 & & & \\ \vdots & & A' & \\ 0 & & & \end{array}\right),$$

where A' has size $(m-1) \times (n-1)$. If A' is the zero matrix then we are done. Otherwise we repeat the previous steps on the smaller matrix to obtain

$$\left(\begin{array}{c|c|ccc} 1 & 0 & 0 & \cdots & 0 \\ \hline 0 & 1 & 0 & \cdots & 0 \\ \hline 0 & 0 & & & \\ \vdots & \vdots & & A'' & \\ 0 & 0 & & & \end{array}\right),$$

where A'' has size $(m-2) \times (n-2)$. We repeat this process until the bottom right corner is a zero matrix. If the process terminates after r steps then the bottom right corner is the zero matrix of size $(m-r) \times (n-r)$. □

This completes our proof of the Fundamental Theorem. This is not the shortest proof, but it is the clearest proof that I know. And it has the added benefit of introducing important ideas (such as elementary matrices) that we will use in the future.

Remark: There is a variant of this algorithm that works over the integers. The difference when working over $\mathbb{Z}$ is that we cannot divide, so we cannot scale the top left entry to equal 1.

[15] We are concerned here with clarity of exposition, not with efficiency of implementation.

However, we can arrange that the top left entry is as small as possible, and that each diagonal entry divides the next. We omit the proof because it requires a bit of number theory.[16]

Theorem (Smith Normal Form). Let A be an $n \times m$ matrix of rank r with integer entries. Then there exist invertible matrices E and F with integer entries, whose inverses E^{-1} and F^{-1} and also have integer entries, such that

$$EAF = \left(\begin{array}{cccc|c} d_1 & & & & \\ & d_2 & & & \\ & & \ddots & & O_{r,n-r} \\ & & & d_r & \\ \hline & O_{m-r,r} & & & O_{m-r,n-r} \end{array}\right).$$

The diagonal integers $0 \leq d_1 \leq \ldots \leq d_r$ have the property that d_{i+1} is an integer multiple of d_i for all i. These diagonal entries are called the *elementary divisors* of the matrix A. The Smith Normal Form is useful in cryptography and in algebraic topology, but we will have no use for it in this course.

4 Existence of Inverse Matrices

As promised, we now apply the Fundamental Theorem to the existence of inverse matrices. Before doing so we make a basic observation. For any $m \times n$ matrix A and $n \times 1$ column $\mathbf{b}$,

$$\boxed{\text{the matrix equation } A\mathbf{x} = \mathbf{b} \text{ has a solution } \mathbf{x} \in \mathbb{R}^n \text{ if and only if } \mathbf{b} \in \mathcal{C}(A).}$$

Indeed, this is just a way of rephrasing the definition of the column space, since every linear combination of the columns of A has the form $A\mathbf{x}$ for some vector $\mathbf{x}$.

First we state conditions for the existence of left and right inverse matrices.

4.1 Existence of Right Inverses.

Given an $m \times n$ matrix A, recall that a right inverse is any $n \times m$ matrix X satisfying $AX = I_m$. In order to find such a matrix, let $\mathbf{x}_j \in \mathbb{R}^n$ be the unknown jth column of X. Then using block multiplication gives

$$\left(\begin{array}{c|c|c} A\mathbf{x}_1 & \cdots & A\mathbf{x}_m \end{array}\right) = A\left(\begin{array}{c|c|c} \mathbf{x}_1 & \cdots & \mathbf{x}_m \end{array}\right) = AX = I_m = \left(\begin{array}{c|c|c} \mathbf{e}_1 & \cdots & \mathbf{e}_m \end{array}\right).$$

In other words, we have $AX = I_m$ if and only if we have $A\mathbf{x}_j = \mathbf{e}_j$ for each column vector $\mathbf{x}_j$, where $\mathbf{e}_j$ is the jth column of the identity matrix I_m, i.e., the jth standard basis vector in $\mathbb{R}^m$. By the previous remark, such vectors $\mathbf{x}_j$ exist if and only if each basis vector $\mathbf{e}_j \in \mathbb{R}^m$ is

[16] In general the algorithms for linear algebra over $\mathbb{Z}$ are much more expensive than for linear algebra over a field such as $\mathbb{R}$ or $\mathbb{C}$. The complexity of the algorithms makes the subject useful for cryptography.

in the column space $\mathcal{C}(A)$. Finally, since $\mathcal{C}(A)$ is a subspace of $\mathbb{R}^m$, this happens if and only if $\mathcal{C}(A)$ fills up all of $\mathbb{R}^m$.[17] Here is a summary:

$$\begin{aligned}
A \text{ has a right inverse} &\iff AX = I_m \text{ for some matrix } X \\
&\iff A\mathbf{x}_j = \mathbf{e}_j \text{ for some vectors } \mathbf{x}_1, \ldots, \mathbf{x}_m \\
&\iff \mathbf{e}_j \in \mathcal{C}(A) \text{ for the standard basis vectors } \mathbf{e}_1, \ldots, \mathbf{e}_m \\
&\iff \mathcal{C}(A) = \mathbb{R}^m \\
&\iff \dim \mathcal{C}(A) = m.
\end{aligned}$$

Furthermore, the Rank-Nullity Theorem tells us that $\dim \mathcal{C}(A) + \dim \mathcal{N}(A^T) = m$, hence

$$\begin{aligned}
A \text{ has a right inverse} &\iff \dim \mathcal{C}(A) = m \\
&\iff \dim \mathcal{N}(A^T) = 0 \\
&\iff \mathcal{N}(A^T) = \{\mathbf{0}\} \\
&\iff A^T\mathbf{x} = \mathbf{0} \text{ implies } \mathbf{x} = \mathbf{0} \\
&\iff \text{the columns of } A^T \text{ are independent} \\
&\iff \text{the rows of } A \text{ are independent.}
\end{aligned}$$

4.2 Existence of Left Inverses.

We could do this from scratch, or we could observe that A has a right inverse if and only if A^T has a left inverse. Indeed, if X is a right inverse of A then $AX = I_m$ implies $X^T A^T = I_m$, so that X^T is a left inverse of A^T. Conversely, if Y is a left inverse of A^T then $YA^T = I_n$ implies $AY^T = I_m$, so that Y^T is a right inverse of A. Hence

$$\begin{aligned}
A \text{ has a left inverse} &\iff A^T \text{ has a right inverse} \\
&\iff \mathcal{C}(A^T) = \mathbb{R}^n \\
&\iff \mathcal{R}(A) = \mathbb{R}^n \\
&\iff \dim \mathcal{R}(A) = n \\
&\iff \dim \mathcal{N}(A) = 0 \qquad \text{Rank-Nullity} \\
&\iff \mathcal{N}(A) = \{\mathbf{0}\} \\
&\iff A\mathbf{x} = \mathbf{0} \text{ implies } \mathbf{x} = \mathbf{0} \\
&\iff \text{the columns of } A \text{ are independent.}
\end{aligned}$$

[17] Indeed, if $\mathcal{C}(A)$ contains every basis vector $\mathbf{e}_1, \ldots, \mathbf{e}_n$ then since $\mathcal{C}(A)$ is a subspace, it contains every linear combination of the basis vectors, i.e., it contains every vector in $\mathbb{R}^m$.

4.3 Existence of Two-Sided Inverses.

Now we will use the Fundamental Theorem, which says that $\dim \mathcal{R}(A) = \dim \mathcal{C}(A)$. First we observe that

$$A \text{ has a two-sided inverse} \iff A \text{ has a right inverse and a left inverse.}$$

Indeed, any two sided inverse is by definition a right inverse and a left inverse. Conversely, suppose that A has a right inverse $AB = I_m$ and a left inverse $CA = I_n$. Then (as we have seen before) we must have

$$B = I_n B = (CA)B = C(AB) = CI_m = C,$$

so that $A^{-1} = B = C$ is the unique two-sided inverse of A. Finally, let r be the rank of A so that $r = \dim \mathcal{R}(A) = \dim \mathcal{C}(A)$ and observe that[18]

$$\begin{aligned} A \text{ has a two-sided inverse} &\iff A \text{ has a right inverse and a left inverse} \\ &\iff \dim \mathcal{C}(A) = m \text{ and } \dim \mathcal{R}(A) = n \\ &\iff r = m \text{ and } r = n \\ &\iff m = n = r. \end{aligned}$$

In particular, A must be square.

These ideas lead to some subtle properties of square matrices. Apparently the columns know what the rows are doing, and vice versa.

4.4 Proof that $AB = I \iff BA = I$ for Square Matrices.

Let A and B be square matrices with $r = \text{rank}(A)$. Then

$$\begin{aligned} AB = I &\implies A \text{ has a right inverse} \\ &\implies r = \text{ the number of columns of } A \\ &\implies r = \text{ the number of rows of } A \\ &\implies A \text{ has a left inverse, say } CA = I. \end{aligned}$$

But then from the above computation we must have $B = C$, so $BA = I$. Switching the roles of A and B shows that $BA = I$ implies $AB = I$.

Here is an interesting application.

[18] There are many more equivalent conditions for invertibility. Wolfram MathWorld lists twenty three: `https://mathworld.wolfram.com/InvertibleMatrixTheorem.html`. Twenty of these follow easily from the results in this section. The remaining three refer to determinants, eigenvalues and singular values, which we haven't discussed yet.

4.5 For Square Matrices, Orthonormal Columns $\iff$ Orthonormal Rows.

Let A be a square matrix. Then we have

$$\begin{aligned} A \text{ has orthonormal columns} &\iff A^T A = I \\ &\iff A A^T = I \\ &\iff A \text{ has orthonormal rows.} \end{aligned}$$

I think this theorem is a small miracle.

Now we know when inverse matrices exist. In the next section we will describe methods to compute inverse matrices.

5 Linear Systems

I assume you have some familiarity with the solution of linear systems, which is the main topic of Linear Algebra I. In this section we will go deeper into the topic.

Recall that a system of m linear equations in n unknowns has the form

$$\begin{cases} a_{11}x_1 + \cdots + a_{1n}x_n = b_1, \\ \vdots \qquad\qquad \vdots \quad \vdots \quad \vdots \\ a_{m1}x_1 + \cdots + a_{mn}x_n = b_m, \end{cases}$$

which can be expressed as a single matrix equation:

$$\begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & \cdots & a_{mn} \end{pmatrix} \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} b_1 \\ \vdots \\ b_m \end{pmatrix}.$$

At a higher level of abstraction we just write $A\mathbf{x} = \mathbf{b}$. Given a matrix of coefficients $A \in \mathbb{R}^{m \times n}$ and a vector of constants $\mathbf{b} \in \mathbb{R}^m$, the goal is to solve for the vector of unknowns $\mathbf{x} \in \mathbb{R}^n$. Recall from the previous section that

the system $A\mathbf{x} = \mathbf{b}$ has a solution for $\mathbf{x}$ if and only if $\mathbf{b}$ is in the column space $\mathcal{C}(A)$.

If this is the case, then we can view the solution of $A\mathbf{x} = \mathbf{b}$ as an $(n - r)$-dimensional affine subspace of $\mathbb{R}^n$, which is parallel to the nullspace $\mathcal{N}(A)$. To be precise, we have the following.

5.1 Shape of the Solution

Let A be an $m \times n$ matrix and consider any vector $\mathbf{b} \in \mathcal{C}(A)$ in the column space. By definition this means we can write $\mathbf{b} = A\mathbf{x}'$ for some vector $\mathbf{x}' \in \mathbb{R}^m$, which might not be unique. Then every solution $A\mathbf{x} = \mathbf{b}$ has the form

$$\mathbf{x} = \mathbf{x}' + \mathbf{x}_0$$

for some *homogeneous solution* $A\mathbf{x}_0 = \mathbf{0}$, i.e., for some element of the nullspace $\mathbf{x}_0 \in \mathcal{N}(A)$. In more colloquial terms:

$$(\text{general solution}) = (\text{one particular solution}) + (\text{general } \textit{homogeneous} \text{ solution}).$$

Proof. Fix a particular solution $A\mathbf{x}' = \mathbf{b}$. Then for any $\mathbf{x}_0 \in \mathcal{N}(A)$ we have

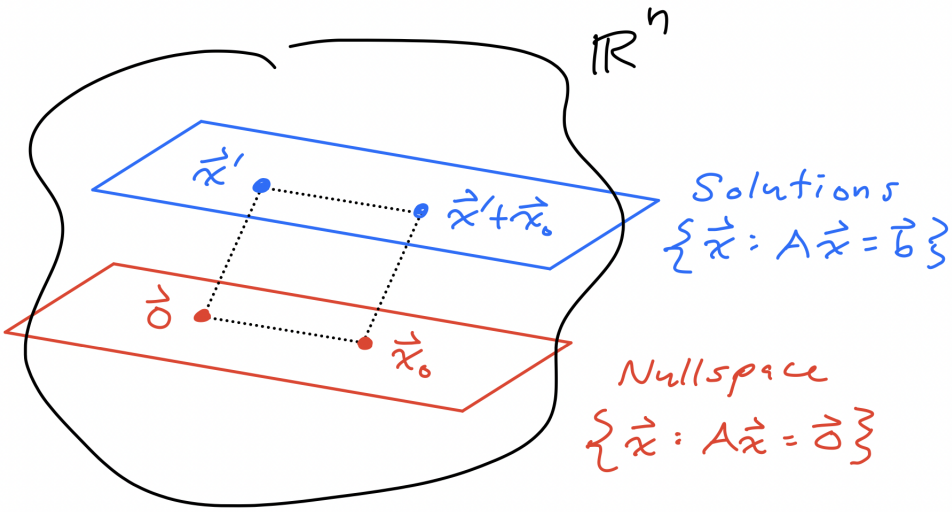
$$A(\mathbf{x}' + \mathbf{x}_0) = A\mathbf{x}' + A\mathbf{x}_0 = A\mathbf{x}' + \mathbf{0} = A\mathbf{x}' = \mathbf{b},$$

so that $\mathbf{x} = \mathbf{x}' + \mathbf{x}_0$ is also a solution. Conversely, let $\mathbf{x}$ be any solution $A\mathbf{x} = \mathbf{b}$. Then

$$\begin{aligned} \mathbf{b} &= \mathbf{b} \\ A\mathbf{x} &= A\mathbf{x}' \\ A\mathbf{x} - A\mathbf{x}' &= \mathbf{0} \\ A(\mathbf{x} - \mathbf{x}') &= \mathbf{0}, \end{aligned}$$

so that $\mathbf{x} - \mathbf{x}'$ is an element of the nullspace, say $\mathbf{x} - \mathbf{x}' = \mathbf{x}_0 \in \mathcal{N}(A)$. Hence every solution has the form $\mathbf{x} = \mathbf{x}' + \mathbf{x}_0$ for some $\mathbf{x}_0$. □

Here is a picture where the nullspace is a 2-dimensional plane living in $\mathbb{R}^n$, so the general solution is also a 2-dimensional plane:



5.2 Uniqueness of the Solution

Suppose that $\mathbf{b} \in \mathcal{C}(A)$, so the system $A\mathbf{x} = \mathbf{b}$ has a solution. In the previous section we saw that this solution has the same shape as the nullspace. Hence the solution is unique if and

only if the nullspace is a single point. If A has shape[19] $m \times n$ and rank r, recall from the Rank-Nullity theorem that $\dim \mathcal{N}(A) = n - \dim \mathcal{R}(A) = n - r$. Hence

$$\begin{aligned} \text{the solution to } A\mathbf{x} = \mathbf{b} \text{ is unique} &\iff \mathcal{N}(A) = \{\mathbf{0}\}, \\ &\iff \dim N(A) = 0 \\ &\iff r = n \\ &\iff A \text{ has independent rows} \\ &\iff A \text{ has a left inverse.} \end{aligned}$$

Indeed, suppose that $CA = I$ and $A\mathbf{x} = \mathbf{b}$. Then we must have

$$\begin{aligned} A\mathbf{x} &= \mathbf{b} \\ CA\mathbf{x} &= C\mathbf{b} \\ I\mathbf{x} &= C\mathbf{b} \\ \mathbf{x} &= C\mathbf{b}, \end{aligned}$$

so that $C\mathbf{b}$ is the **unique** solution.

5.3 How to Compute the Solution

Linear systems are solved using *row reduction*, also called *Gaussian elimination*. Gauss developed this method together with the method of least squares when he was 24, in order to determine the orbit of the dwarf planet Ceres. A similar method for solving linear systems was used in China since at least the 5th century AD.[20]

We will perform row reduction using elimination matrices, which were defined in the previous section. The goal is to put the system in a standardized simple form. Given a general matrix A, we first multiply on the left by **lower triangular** elimination matrices $L_{ij}(\lambda)$ (i.e., with $i > j$) until we obtain a matrix in "staircase form":

$$L_k \cdots L_2 L_1 A = \begin{pmatrix} * & \cdot & \cdot & \cdot & \cdot & \cdot \\ & & * & \cdot & \cdot & \cdot \\ & & & & * & \cdot \\ & & & & & \end{pmatrix}.$$

Here the blank entries are zero. The entries labeled $*$ are **nonzero**; these are called the *pivots*. And the entries marked $\cdot$ are arbitrary. Next we multiply by dilation matrices $D_i(\lambda)$ to turn the pivot entries into 1s:

$$D_\ell \cdots D_2 D_1 L_k \cdots L_2 L_1 A = \begin{pmatrix} 1 & \cdot & \cdot & \cdot & \cdot & \cdot \\ & & 1 & \cdot & \cdot & \cdot \\ & & & & 1 & \cdot \\ & & & & & \end{pmatrix}.$$

[19] Here I am using the word "shape" for matrices and for subspaces. Don't take it too literally in either case.

[20] The Chinese method was concerned with **integer solutions**, and is the precursor of the Chinese Remainder Theorem in abstract algebra.

Finally, we multiply by **upper triangular** elimination matrices $L_{ij}(\lambda)$ (i.e., with $i < j$) to eliminate the entries above the pivots:

$$U_m \cdots U_2 U_1 D_\ell \cdots D_2 D_1 L_k \cdots L_2 L_1 A = \begin{pmatrix} 1 & \cdot & 0 & \cdot & 0 & \cdot \\ & & 1 & \cdot & 0 & \cdot \\ & & & & 1 & \cdot \\ & & & & & \end{pmatrix}.$$

Finally, this is called the *reduced row echelon form* (or RREF) of A. It has the virtue of being **unique**, i.e., independent of the particular order of row operations.[21]

We can summarize this process as follows. Multiply the elementary matrices together to obtain

$$L := L_k \cdots L_2 L_1, \quad D := D_\ell \cdots D_2 D_1 \quad \text{and} \quad U := U_m \cdots U_2 U_1.$$

The names indicate that L is lower trianglular (i.e., has zeros above the diagonal), D is diagonal (i.e., has zeros away from the diagonal) and U is upper triangular (i.e., has zeros below the diagonal). Furthermore, let's define $E = UDL$, which is invertible because it is a product of invertible matrices. Let R denote the RREF of A, so that

$$EA = R.$$

Since E is invertible, it follows from the section on the Fundamental Theorem that R has the same row space and nullspace as A:

$$\mathcal{R}(R) = \mathcal{R}(A) \quad \text{and} \quad \mathcal{N}(R) = \mathcal{N}(A).$$

In other words, the **homogeneous** system equation $A\mathbf{x} = \mathbf{0}$ is equivalent to $R\mathbf{x} = \mathbf{0}$, and the solution of this second system is particularly easy to read off. To solve the **non-homogeneous** system $A\mathbf{x} = \mathbf{b}$ we simply multiply both sides on the left by E to obtain

$$\begin{aligned} A\mathbf{x} &= \mathbf{b} \\ EA\mathbf{x} &= E\mathbf{b} \\ R\mathbf{x} &= E\mathbf{b}, \end{aligned}$$

and the solution is again easy to read off.

Example. Solve the linear system

$$\begin{cases} x + 3y + 8z = 2, \\ x + 2y + 6z = 1, \\ 0 + y + 2z = 1, \end{cases}$$

which can be expressed in matrix notation as

$$\begin{pmatrix} 1 & 3 & 8 \\ 1 & 2 & 6 \\ 0 & 1 & 2 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 2 \\ 1 \\ 1 \end{pmatrix},$$

[21] We will not prove this uniqueness because it is a bit tricky, and we will never need it.

$$A\mathbf{x} = \mathbf{b}.$$

First we perform down elimination on A:

$$\begin{pmatrix} 1 & & \\ -1 & 1 & \\ & & 1 \end{pmatrix} \begin{pmatrix} 1 & 3 & 8 \\ 1 & 2 & 6 \\ 0 & 1 & 2 \end{pmatrix} = \begin{pmatrix} 1 & 3 & 8 \\ 0 & -1 & -2 \\ 0 & 1 & 2 \end{pmatrix}$$

$$\begin{pmatrix} 1 & & \\ & 1 & \\ & +1 & 1 \end{pmatrix} \begin{pmatrix} 1 & 3 & 8 \\ 0 & -1 & -2 \\ 0 & 1 & 2 \end{pmatrix} = \begin{pmatrix} 1 & 3 & 8 \\ 0 & -1 & -2 \\ 0 & 0 & 0 \end{pmatrix}$$

Next we scale the pivots:

$$\begin{pmatrix} 1 & & \\ & -1 & \\ & & 1 \end{pmatrix} \begin{pmatrix} 1 & 3 & 8 \\ 0 & -1 & -2 \\ 0 & 0 & 0 \end{pmatrix} = \begin{pmatrix} 1 & 3 & 8 \\ 0 & 1 & 2 \\ 0 & 0 & 0 \end{pmatrix}.$$

Then we perform up elimination:[22]

$$\begin{pmatrix} 1 & -3 & \\ & 1 & \\ & & 1 \end{pmatrix} \begin{pmatrix} 1 & 3 & 8 \\ 0 & 1 & 2 \\ 0 & 0 & 0 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 2 \\ 0 & 1 & 2 \\ 0 & 0 & 0 \end{pmatrix}.$$

The single matrix that performs the elimination is

$$\begin{aligned} E &= UDL \\ &= \left[\begin{pmatrix} 1 & -3 & \\ & 1 & \\ & & 1 \end{pmatrix} \right] \left[\begin{pmatrix} 1 & & \\ & -1 & \\ & & 1 \end{pmatrix} \right] \left[\begin{pmatrix} 1 & & \\ & 1 & \\ & +1 & 1 \end{pmatrix} \begin{pmatrix} 1 & & \\ -1 & 1 & \\ & & 1 \end{pmatrix} \right] \\ &= \begin{pmatrix} 1 & -3 & \\ & 1 & \\ & & 1 \end{pmatrix} \begin{pmatrix} 1 & & \\ & -1 & \\ & & 1 \end{pmatrix} \begin{pmatrix} 1 & & \\ -1 & 1 & \\ -1 & 1 & 1 \end{pmatrix} \\ &= \begin{pmatrix} -2 & 3 & 0 \\ 1 & -1 & 0 \\ -1 & 1 & 1 \end{pmatrix}. \end{aligned}$$

Check:

$$EA = R,$$

$$\begin{pmatrix} -2 & 3 & 0 \\ 1 & -1 & 0 \\ -1 & 1 & 1 \end{pmatrix} \begin{pmatrix} 1 & 3 & 8 \\ 1 & 2 & 6 \\ 0 & 1 & 2 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 2 \\ 0 & 1 & 2 \\ 0 & 0 & 0 \end{pmatrix}.$$

[22] In class I circled pivots and drew arrows, which is extremely difficult to do in LaTeX.

To solve the homogeneous system $A\mathbf{x} = \mathbf{0}$ we multiply both sides by E:

$$\begin{aligned} A\mathbf{x} &= \mathbf{0} \\ EA\mathbf{x} &= E\mathbf{0} \\ R\mathbf{x} &= \mathbf{0} \\ \begin{pmatrix} 1 & 0 & 2 \\ 0 & 1 & 2 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} &= \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}. \end{aligned}$$

This is equivalent to the linear system

$$\begin{cases} x + 0 + 2z = 0, \\ 0 + y + 2z = 0, \\ 0 + 0 + 0 = 0. \end{cases}$$

Note that the third equation is redundant, which shows that our original system of three equations really only contains two equations. The solution, which is also called the nullspace of A, is a line:

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} -2z \\ -2z \\ z \end{pmatrix} = z \begin{pmatrix} -2 \\ -2 \\ 1 \end{pmatrix}.$$

To solve the **non-homogeneous system** $A\mathbf{x} = \mathbf{b}$ we again multiply both sides by E:

$$\begin{aligned} A\mathbf{x} &= \mathbf{b} \\ EA\mathbf{x} &= E\mathbf{b} \\ R\mathbf{x} &= E\mathbf{b} \\ \begin{pmatrix} 1 & 0 & 2 \\ 0 & 1 & 2 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} &= \begin{pmatrix} -2 & 3 & 0 \\ 1 & -1 & 0 \\ -1 & 1 & 1 \end{pmatrix} \begin{pmatrix} 2 \\ 1 \\ 1 \end{pmatrix} \\ \begin{pmatrix} 1 & 0 & 2 \\ 0 & 1 & 2 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} &= \begin{pmatrix} -1 \\ 1 \\ 0 \end{pmatrix}. \end{aligned}$$

This is equivalent to the linear system

$$\begin{cases} x + 0 + 2z = -1, \\ 0 + y + 2z = 1, \\ 0 + 0 + 0 = 0, \end{cases}$$

whose solution is a line parallel to the null space:

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} -1 - 2z \\ 1 - 2z \\ z \end{pmatrix} = \begin{pmatrix} -1 \\ 1 \\ 0 \end{pmatrix} + z \begin{pmatrix} -2 \\ -2 \\ 1 \end{pmatrix}.$$

In the language of 5.1, $\mathbf{x}_0 = z(-2,-2,1)$ is the general homogeneous solution and $\mathbf{x}' = (-1,1,0)$ is one particular solution. Note that there are infinitely many equivalent ways to describe this solution. For example, we can also write

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 1 \\ 3 \\ -1 \end{pmatrix} + t \begin{pmatrix} 2 \\ 2 \\ 1 \end{pmatrix}.$$

On the other hand, the following system has **no solution** because $(1,0,0)$ is not in the column space of A:

$$\begin{pmatrix} 1 & 3 & 8 \\ 1 & 2 & 6 \\ 0 & 1 & 2 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}.$$

If we **try** to solve the system then we obtain

$$\begin{pmatrix} -2 & 3 & 0 \\ 1 & -1 & 0 \\ -1 & 1 & 1 \end{pmatrix} \begin{pmatrix} 1 & 3 & 8 \\ 1 & 2 & 6 \\ 0 & 1 & 2 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} -2 & 3 & 0 \\ 1 & -1 & 0 \\ -1 & 1 & 1 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$$

$$\begin{pmatrix} 1 & 0 & 2 \\ 0 & 1 & 2 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} -2 \\ 1 \\ -1 \end{pmatrix},$$

which is equivalent to the system

$$\begin{cases} x + 0y + 2z = -2, \\ 0x + y + 2z = 1, \\ 0x + 0y + 0z = -1. \end{cases}$$

This system has no solution because the third equation $0x + 0y + 0z = -1$ has no solution.

5.4 How to Compute the Inverse of a Square Matrix

We have seen a method for solving linear systems. Now we apply this method to compute the inverse of a square matrix. Let A be an invertible $n \times n$ matrix, and let E be the product of elementary matrices that puts A in reduced row echelon form: $EA = R$. Since A is invertible it has independent rows, and, since $\mathcal{R}(A) = \mathcal{R}(R)$, this implies that R has independent rows. In particular, R has no zero rows, which finally implies that R is the identity matrix. Summary:

The RREF of an invertible matrix A is the identity matrix I.

This idea gives an algorithm to compute the inverse. Begin with the *augmented matrix*

$$\left(\begin{array}{c|c} A & I \end{array} \right).$$

Then apply elementary matrices on the left to put A in RREF:

$$\left(\begin{array}{c|c} A & I \end{array} \right)$$

$$
\begin{aligned}
&\rightsquigarrow \left(\begin{array}{c|c} E_1A & E_1I \end{array}\right) \\
&\rightsquigarrow \left(\begin{array}{c|c} E_2E_1A & E_2E_1I \end{array}\right) \\
&\quad\vdots \\
&\rightsquigarrow \left(\begin{array}{c|c} E_k\cdots E_2E_1A & E_k\cdots E_2E_1I \end{array}\right) \\
&= \left(\begin{array}{c|c} EA & E \end{array}\right) \\
&= \left(\begin{array}{c|c} R & E \end{array}\right).
\end{aligned}
$$

If A is invertible, so that $R = I$ and $E = A^{-1}$ then the process gives

$$
\left(\begin{array}{c|c} A & I \end{array}\right) \overset{\text{RREF}}{\rightsquigarrow} \left(\begin{array}{c|c} I & A^{-1} \end{array}\right).
$$

We don't even need to keep track of the elementary matrices.

Example.

$$
\begin{aligned}
&\left(\begin{array}{c|c} A & I \end{array}\right) \\
&= \left(\begin{array}{ccc|ccc} 1 & 1 & 1 & 1 & 0 & 0 \\ 1 & 2 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 0 & 1 \end{array}\right) \\
&\rightsquigarrow \left(\begin{array}{ccc|ccc} 1 & 1 & 1 & 1 & 0 & 0 \\ 0 & 1 & -1 & -1 & 1 & 0 \\ 1 & 0 & 0 & 0 & 0 & 1 \end{array}\right) \\
&\rightsquigarrow \left(\begin{array}{ccc|ccc} 1 & 1 & 1 & 1 & 0 & 0 \\ 0 & 1 & -1 & -1 & 1 & 0 \\ 0 & -1 & -1 & -1 & 0 & 1 \end{array}\right) \\
&\rightsquigarrow \left(\begin{array}{ccc|ccc} 1 & 1 & 1 & 1 & 0 & 0 \\ 0 & 1 & -1 & -1 & 1 & 0 \\ 0 & 0 & -2 & -2 & 1 & 1 \end{array}\right) \\
&\rightsquigarrow \left(\begin{array}{ccc|ccc} 1 & 1 & 1 & 1 & 0 & 0 \\ 0 & 1 & -1 & -1 & 1 & 0 \\ 0 & 0 & 1 & 1 & -1/2 & -1/2 \end{array}\right) \\
&\rightsquigarrow \left(\begin{array}{ccc|ccc} 1 & 1 & 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1/2 & -1/2 \\ 0 & 0 & 1 & 1 & -1/2 & -1/2 \end{array}\right) \\
&\rightsquigarrow \left(\begin{array}{ccc|ccc} 1 & 1 & 0 & 0 & 1/2 & 1/2 \\ 0 & 1 & 0 & 0 & 1/2 & -1/2 \\ 0 & 0 & 1 & 1 & -1/2 & -1/2 \end{array}\right) \\
&\rightsquigarrow \left(\begin{array}{ccc|ccc} 1 & 0 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 1/2 & -1/2 \\ 0 & 0 & 1 & 1 & -1/2 & -1/2 \end{array}\right)
\end{aligned}
$$

$$= \left(\begin{array}{c|c} I & A^{-1} \end{array} \right).$$

Check:

$$\begin{pmatrix} 1 & 1 & 1 \\ 1 & 2 & 0 \\ 1 & 0 & 0 \end{pmatrix} \begin{pmatrix} 0 & 0 & 1 \\ 0 & 1/2 & -1/2 \\ 1 & -1/2 & -1/2 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

Recall that for square matrices A and B we have $AB = I$ if and only if $BA = I$ so we only need to check one.

What happens if we try to invert a non-invertible matrix? Consider the matrix A from Section 5.3. We perform elimination until we reach the RREF:

$$\begin{aligned} &\left(\begin{array}{ccc|ccc} 1 & 3 & 8 & 1 & 0 & 0 \\ 1 & 2 & 6 & 0 & 1 & 0 \\ 0 & 1 & 2 & 0 & 0 & 1 \end{array} \right) \\ &= \left(\begin{array}{c|c} A & I \end{array} \right) \\ &\rightsquigarrow \left(\begin{array}{c|c} EA & E \end{array} \right) \\ &= \left(\begin{array}{c|c} R & E \end{array} \right) \\ &= \left(\begin{array}{ccc|ccc} 1 & 0 & 2 & -2 & 3 & 0 \\ 0 & 1 & 2 & 1 & -1 & 0 \\ 0 & 0 & 0 & -1 & 1 & 1 \end{array} \right) \end{aligned}$$

And then we're stuck.

6 Least Squares Approximation

6.1 The Four Fundamental Subspaces

Let me summarize our results so far. To each $m \times n$ matrix A we associate four subspaces; two of $\mathbb{R}^m$ and two of $\mathbb{R}^n$:

$$\mathcal{R}(A), \mathcal{N}(A) \subseteq \mathbb{R}^n \quad \text{and} \quad \mathcal{C}(A), \mathcal{N}(A^T) \subseteq \mathbb{R}^m.$$

The subspaces $\mathcal{R}(A)$ and $\mathcal{N}(A)$ are orthogonal complements in $\mathbb{R}^n$, while $\mathcal{C}(A)$ and $\mathcal{N}(A^T)$ are orthogonal complements in $\mathbb{R}^m$.[23] It follows from the general theorem on dimensions of orthogonal complements[24] that

$$\dim \mathcal{R}(A) + \dim \mathcal{N}(A) = n \quad \text{and} \quad \dim \mathcal{C}(A) + \dim N(A^T) = m.$$

These results are called the Rank-Nullity Theorem. The Fundamental Theorem says that the rank of A is well-defined:

$$r = \operatorname{rank}(A) := \dim \mathcal{R}(A) = \dim \mathcal{C}(A).$$

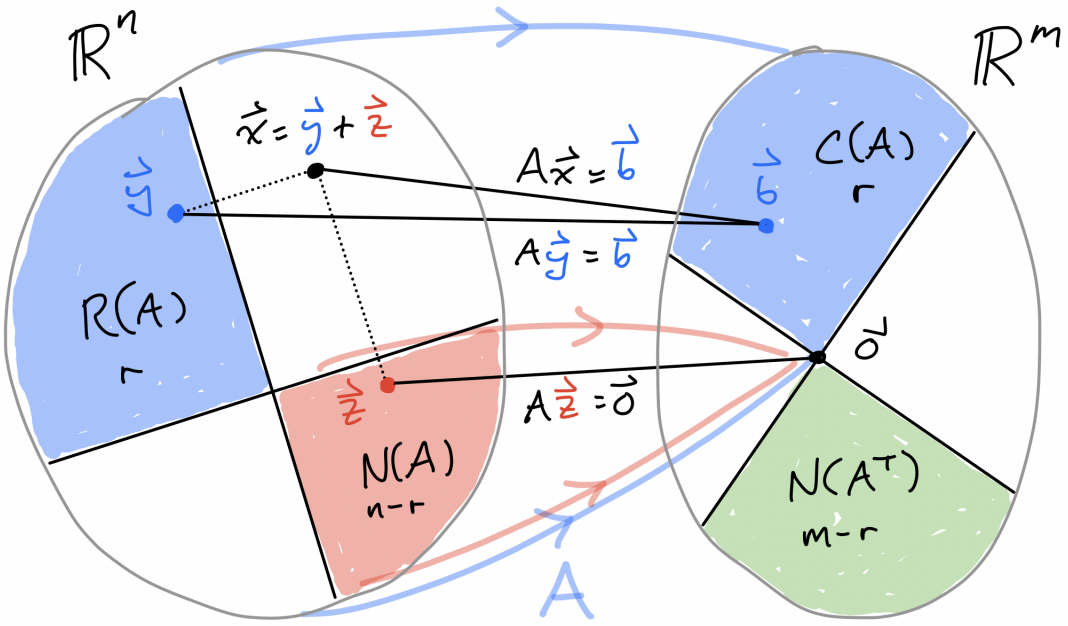
[23] Remind yourself right now why this is true.

[24] See the homework.

Hence we also have

$$\dim \mathcal{N}(A) = n - r \quad \text{and} \quad \dim \mathcal{N}(A^T) = m - r.$$

Here is "the big picture" in the style of Gilbert Strang:[25]



The matrix A maps the space $\mathbb{R}^n$ on the left to the space $\mathbb{R}^m$ on the right. Actually, A maps all of $\mathbb{R}^n$ onto the blue column space $\mathcal{C}(A)$. The red nullspace $\mathcal{N}(A)$ gets squashed onto the origin $\mathbf{0} \in \mathbb{R}^m$. Any vector $\mathbf{x} \in \mathbb{R}^n$ can be expressed uniquely as $\mathbf{x} = \mathbf{y} + \mathbf{z}$ with $\mathbf{y} \in \mathcal{R}(A)$ and $\mathbf{z} \in \mathcal{N}(A)$. If $A\mathbf{y} = \mathbf{b}$ then we also have $A\mathbf{x} = \mathbf{b}$ because

$$A\mathbf{x} = A(\mathbf{y} + \mathbf{z}) = A\mathbf{y} + A\mathbf{z} = \mathbf{b} + \mathbf{0} = \mathbf{b}.$$

This picture is rather impressionistic but it does a good job of showing a lot of information. One thing it doesn't show is the set of all solutions to the equation $A\mathbf{x} = \mathbf{b}$, which is an affine subspace of $\mathbb{R}^n$ that is parallel to $N(A)$ and passes through $\mathbf{x}$ and $\mathbf{y}$. I guess that would make the picture unreadable.

Next we work through an explicit example. Consider the rank 2 matrix

$$A = \begin{pmatrix} 1 & 3 & 8 \\ 1 & 2 & 6 \\ 0 & 1 & 2 \end{pmatrix}.$$

In Section 5.3 we already computed the nullspace:

$$\mathcal{N}(A) = \text{ the line in } \mathbb{R}^3 \text{ spanned by } (2, 2, -1).$$

[25] A similar picture appears on the cover of his 4th edition of *Introduction to Linear Algebra*.

The rowspace is the orthogonal complement of the nullspace, which is a plane:

$$\mathcal{R}(A) = \mathcal{N}(A)^\perp = \text{the plane in } \mathbb{R}^3 \text{ defined by } 2x + 2y - z = 0.$$

Since no two rows of A are parallel, any two rows will form a basis for $\mathcal{R}(A)$. More systematically, we can look at the RREF:

$$EA = R,$$
$$\begin{pmatrix} -2 & 3 & 0 \\ 1 & -1 & 0 \\ -1 & 1 & 1 \end{pmatrix} \begin{pmatrix} 1 & 3 & 8 \\ 1 & 2 & 6 \\ 0 & 1 & 2 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 2 \\ 0 & 1 & 2 \\ 0 & 0 & 0 \end{pmatrix}.$$

Since the product of elementary matrices E is invertible, we know that $\mathcal{R}(A) = \mathcal{R}(EA) = \mathcal{R}(R)$, and it is very easy to read a basis from R:

$$\mathcal{R}(A) = \mathcal{R}(R) = \text{span}\{(1,0,2),(0,1,2)\}.$$

To compute the column space $\mathcal{C}(A)$ and left nullspace $\mathcal{N}(A^T)$ we can apply the same methods to the transposed matrix A^T. That is, we should compute $RREF(A^T)$:[26]

$$\begin{pmatrix} 1 & 1 & 0 \\ 3 & 2 & 1 \\ 8 & 6 & 2 \end{pmatrix} \overset{\text{RREF}}{\rightsquigarrow} \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & -1 \\ 0 & 0 & 0 \end{pmatrix}.$$

From this we see that

$$\mathcal{C}(A) = \mathcal{R}(A^T) = \text{span}\{(1,0,1),(0,1,-1)\}.$$

Finally, the left nullspace is the solution to the homogeneous system $A^T\mathbf{x} = \mathbf{0}$, which from the RREF of A^T is equivalent to

$$\begin{cases} x + 0 + z = 0, \\ 0 + y - z = 0, \\ 0 + 0 + 0 = 0. \end{cases}$$

The solution is the line spanned by $(1,-1,-1)$:

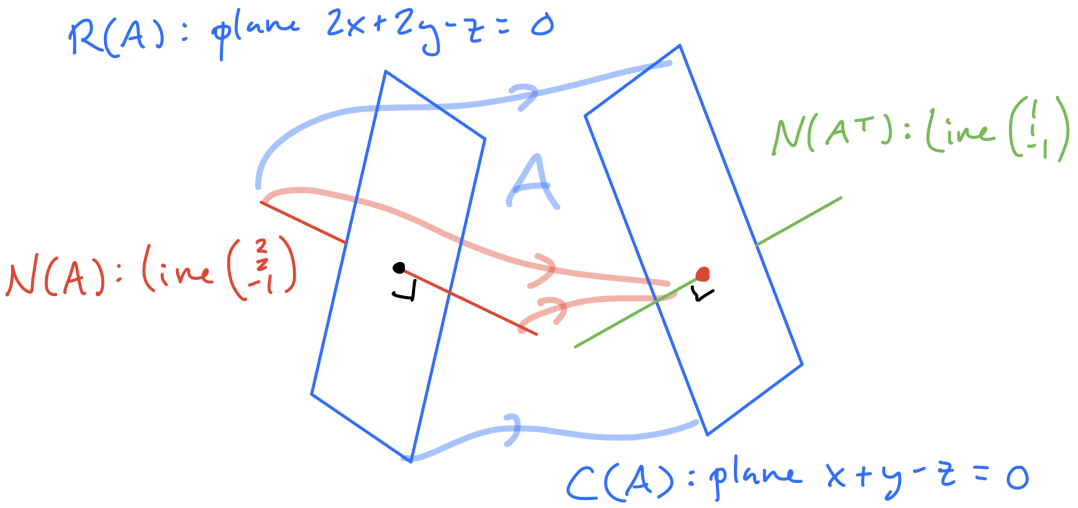
$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} -z \\ z \\ z \end{pmatrix} = z \begin{pmatrix} -1 \\ 1 \\ 1 \end{pmatrix} = \text{span}\{(1,-1,-1)\}.$$

As expected, this line is the orthogonal complement of the column space:

$$\text{span}\{(1,-1,-1)\}^\perp = (\text{plane } x - y - z = 0) = \text{span}\{(1,0,1),(0,1,-1)\} = \mathcal{C}(A).$$

Here is a picture:

[26] This is equivalent to applying elementary matrices on the right of A to compute *reduced column echelon form* (RCEF), but nobody uses this terminology.



Note that the line $\mathcal{N}(A)$ gets squashed onto the origin, while all of $\mathbb{R}^3$ gets squashed onto the plane $\mathcal{C}(A)$. Since this matrix is square, we could have drawn all four subspaces in the same copy of $\mathbb{R}^3$, but that would just be a mess.

In summary:

The four fundamental subspaces can be read off from RREF(A) and RREF(A^T).

6.2 The Matrices A^TA and AA^T

We have seen that a non-square matrix A cannot have an inverse. To fix this we sometimes consider the square matrices A^TA and AA^T. To be precise, suppose that A has shape $m \times n$, so that A^TA is square of shape $n \times n$ and AA^T is square of shape $m \times m$. We also observe that these matrices are *symmetric* because

$$(A^TA)^T = A^T(A^T)^T = A^TA$$

and

$$(AA^T)^T = (A^T)^TA^T = AA^T.$$

The matrices A^TA and AA^T show up surprisingly often in applied mathematics. We will see our first glimpse of this in the next section when we discuss least squares approximation. To prepare for this we develop some basic properties. The key observation is that A and A^TA have **the same nullspace**:

$$\boxed{\mathcal{N}(A^TA) = \mathcal{N}(A).}$$

This would be easy to prove if A^T had a left inverse. Indeed, if E is a matrix with a left inverse $E'E = I$ then we recall from Section 3 that

$$\mathcal{R}(EA) = \mathcal{R}(A),$$

and hence

$$\mathcal{N}(EA) = \mathcal{R}(EA)^\perp = \mathcal{R}(A)^\perp = \mathcal{N}(A).$$

But the matrix A^T might **not** have a left inverse, so we cannot use this fact. Instead we use a clever trick:[27]

$$\boxed{\text{For any } \mathbf{x} \in \mathbb{R}^n \text{ we have } \mathbf{x}^T A^T A\mathbf{x} = (A\mathbf{x})^T(A\mathbf{x}) = (A\mathbf{x}) \bullet (A\mathbf{x}) = \|A\mathbf{x}\|^2.}$$

Proof that $\mathcal{N}(A^TA) = \mathcal{N}(A)$. First we note that $\mathcal{N}(A) \subseteq \mathcal{N}(A^TA)$ because

$$A\mathbf{x} = \mathbf{0} \quad \Longrightarrow \quad (A^TA)\mathbf{x} = A^T(A\mathbf{x}) = A^T\mathbf{0} = \mathbf{0}.$$

On the other hand, suppose that $(A^TA)\mathbf{x} = \mathbf{0}$. Then from the trick we have

$$\|A\mathbf{x}\|^2 = \mathbf{x}^T A^T A\mathbf{x} = \mathbf{x}^T(A^TA\mathbf{x}) = \mathbf{x}^T\mathbf{0} = 0,$$

and hence $\|A\mathbf{x}\| = 0$. But recall that the standard norm $\| - \|$ satisfies $\|\mathbf{v}\| = 0$ if and only if $\mathbf{v} = \mathbf{0}$. Hence we must have $A\mathbf{x} = \mathbf{0}$ as desired. □

We obtain a similar identity by replacing A with A^T. To be precise, let $B = A^T$, so that

$$\boxed{\mathcal{N}(AA^T) = \mathcal{N}(B^TB) = \mathcal{N}(B) = \mathcal{N}(A^T).}$$

And it follows from these identities that

$$\boxed{\operatorname{rank}(A^TA) = \operatorname{rank}(A) = \operatorname{rank}(A^T) = \operatorname{rank}(AA^T).}$$

Indeed, the first and third equations follow by applying dimension to the identities $\mathcal{N}(A^TA) = \mathcal{N}(A)$ and $\mathcal{N}(AA^T) = \mathcal{N}(A^T)$, while the middle equation is just the Fundamental Theorem. This is quite interesting since the four matrices A, A^T, A^TA and AA^T have different shapes.

We combine these results to prove the main result of this section.

Theorem (Invertibility of A^TA and AA^T). For any matrix A, the matrices A^TA and AA^T are square, hence they might be invertible. I claim that

$$\begin{aligned} (A^TA)^{-1} \text{ exists} \quad &\Longleftrightarrow \quad A \text{ has independent columns,} \\ (AA^T)^{-1} \text{ exists} \quad &\Longleftrightarrow \quad A \text{ has independent rows.} \end{aligned}$$

Proof. Let A have shape $m \times n$ and rank r. To prove the first statement, note that A^TA has shape $n \times n$, hence

$$(A^TA)^{-1} \text{ exists} \quad \Longleftrightarrow \quad \operatorname{rank}(A^TA) = n$$

[27] The idea lurking in the background is that matrices of the form A^TA are related to inner products. See Problem 5 on Homework 3.

$$\begin{aligned} &\iff && \operatorname{rank}(A) = n && \text{previous result} \\ &\iff && \dim \mathcal{C}(A) = n \\ &\iff && A \text{ as independent columns.} \end{aligned}$$

Similarly, since AA^T has shape $m \times m$, we have

$$\begin{aligned} (AA^T)^{-1} \text{ exists} &\iff && \operatorname{rank}(AA^T) = m \\ &\iff && \operatorname{rank}(A) = m && \text{previous result} \\ &\iff && \dim \mathcal{R}(A) = m \\ &\iff && A \text{ as independent rows.} \end{aligned}$$

□

To end this section we give two theoretical applications.[28]

Explicit formulas for left and right inverses. For any matrix A we recall from 4.1 that

$$\begin{aligned} A \text{ has a left inverse} &\iff A \text{ has independent columns,} \\ A \text{ has a right inverse} &\iff A \text{ has independent rows.} \end{aligned}$$

Such left and right inverses are **not unique**, but we can use the previous theorem to give a formula for **specific** left and right inverse. If A has independent columns then $(A^TA)^{-1}$ exists and $(A^TA)^{-1}A^T$ is a left inverse:

$$[(A^TA)^{-1}A^T]A = (A^TA)^{-1}(A^TA) = I.$$

If A has independent rows then $(AA^T)^{-1}$ exists and $A^T(AA^T)^{-1}$ is a right inverse:

$$A[A^T(AA^T)^{-1}] = (AA^T)(AA^T)^{-1} = I.$$

CMR Factorization. Applied linear algebra is often expressed in terms of matrix factorizations. Here we will show that any $m \times n$ matrix A of rank r can be factored as $A = CMR$, where the matrices C, M and R have shapes $m \times r$, $r \times r$ and $r \times n$. The matrices C and R are defined as follows:

- Choose any r independent columns of A and let these be the columns of C.
- Choose any r independent rows of A and let these be the rows of R.

By construction, C has independent columns and R has independent rows, so the matrices $(C^TC)^{-1}$ and $(RR^T)^{-1}$ exist. In this case we will show that **there exists a unique** $r \times r$ **matrix** M **satisfying** $A = CMR$. This matrix is invertible and is determined by the formula

$$M = (C^TC)^{-1}(C^TAR^T)(RR^T)^{-1}.$$

It is difficult to see that the matrix defined by this formula has the desired properties, so we will proceed in two steps:

[28] The section on Least Squares below gives some practical applications.

(1) There exists an invertible matrix M satisfying $A = CMR$.

(2) The matrix from part (1) must satisfy the desired formula.

The proof of (1) is tricky and algorithmic.[29] Feel free to skip it.

(1): First let T be an invertible product of column transpositions so that the first r columns of AT are equal to C; let's say

$$AT = \left(\begin{array}{c|c} C & F \end{array}\right),$$

for some $m \times (n-r)$ matrix F. Next we consider the reduced row echelon form of AT. Let E be an invertible product of elementary row operations satisfying $EAT = \mathrm{RREF}(AT)$. Since the first r columns of AT (i.e., the columns of C) are independent, so will be the first r columns of the RREF, and it follows that

$$EAT = \mathrm{RREF}(AT) = \left(\begin{array}{c|c} I_r & G \\ \hline O_{m-r,r} & O_{m-r,n-r} \end{array}\right),$$

for some $r \times (n-r)$ matrix G. I claim that $AT = C\left(\begin{array}{c|c} I & G \end{array}\right)$. Indeed, if we write $E^{-1} = \left(\begin{array}{c|c} X & Y \end{array}\right)$ where X is $m \times r$ and Y is $m \times (m-r)$ then we find

$$\left(\begin{array}{c|c} C & F \end{array}\right) = AT = E^{-1}\left(\begin{array}{c|c} I & G \\ \hline O & O \end{array}\right) = \left(\begin{array}{c|c} X & Y \end{array}\right)\left(\begin{array}{c|c} I & G \\ \hline O & O \end{array}\right) = \left(\begin{array}{c|c} X & YG \end{array}\right),$$

which implies that $X = C$.[30] It follows that

$$AT = E^{-1}\left(\begin{array}{c|c} I & G \\ \hline O & O \end{array}\right) = \left(\begin{array}{c|c} C & Y \end{array}\right)\left(\begin{array}{c|c} I & G \\ \hline O & O \end{array}\right) = \left(\begin{array}{c|c} C & CG \end{array}\right) = C\left(\begin{array}{c|c} I & G \end{array}\right).$$

At this point we have

$$A = C\left(\begin{array}{c|c} I & G \end{array}\right)T^{-1} = CR',$$

where we have defined $R' := \left(\begin{array}{c|c} I & G \end{array}\right)T^{-1}$. Our final goal is to prove that $R' = MR$ for some invertible M. Since C has independent columns (and hence has a left inverse) we see from Section 3 that A and R' have the same row space:

$$\mathcal{R}(A) = \mathcal{R}(CR') = \mathcal{R}(R'),$$

Since this row space is r-dimensional, and since R' has r rows, it follows that the rows of R' are a basis for $\mathcal{R}(A)$. In particular, each row of R can be expressed as a linear combination of the rows of R', which gives a matrix equation $R = MR'$. Similarly, since the rows of R are a basis for $\mathcal{R}(A)$ we can write $R' = NR$ for some matrix N. Putting these together gives

[29] I apologize that I assigned this as homework; I didn't realize how tricky it is. Gilbert Strang fooled me. Maybe there is a more direct proof but I couldn't find it.

[30] We also have $YG = F$, but we don't care about this.

$R = MNR$. Finally, since R has a right inverse this implies $MN = I$, which implies that M is invertible.[31] □

(2): Once we know that M exists, it is not difficult to prove that satisfies the desired formula. Indeed, suppose that $A = CMR$. Then since $(C^TC)^{-1}$ and $(RR^T)^{-1}$ exist we must have

$$\begin{aligned} CMR &= A \\ C^T(CMR)R^T &= C^TAR^T \\ (C^TC)M(RR^T) &= C^TAR^T \\ M &= (C^TC)^{-1}C^TAR^T(RR^T)^{-1}. \end{aligned}$$

□

For example, let's consider our favorite matrix

$$A = \begin{pmatrix} 1 & 3 & 8 \\ 1 & 2 & 6 \\ 0 & 1 & 2 \end{pmatrix}.$$

This matrix has rank 2, so we should choose two independent columns and two independent rows. Choosing the first two columns and the first two rows gives

$$A = \begin{pmatrix} 1 & 3 \\ 1 & 2 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} -2 & 3 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} 1 & 3 & 8 \\ 1 & 2 & 6 \end{pmatrix}.$$

Choosing columns $1, 3$ and rows $2, 3$ gives

$$A = \begin{pmatrix} 1 & 8 \\ 1 & 6 \\ 0 & 2 \end{pmatrix} \begin{pmatrix} 1 & -3 \\ 0 & 1/2 \end{pmatrix} \begin{pmatrix} 1 & 2 & 6 \\ 0 & 1 & 2 \end{pmatrix}.$$

Remark: There is another interesting description of the matrix M. In the paper *LU and CR Elimination* by Strang and Moler,[32] they prove that M^{-1} is the matrix obtained from A by intersecting the columns of C with the rows of R. We observe that this is true for the two examples just given:

$$\begin{pmatrix} -2 & 3 \\ 1 & -1 \end{pmatrix}^{-1} = \begin{pmatrix} 1 & 3 \\ 1 & 2 \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} 1 & -3 \\ 0 & 1/2 \end{pmatrix}^{-1} = \begin{pmatrix} 1 & 6 \\ 0 & 2 \end{pmatrix}.$$

Pretty cool.

[31] Recall that $MN = I$ implies $NM = I$ for square matrices.

[32] I think there's a better proof in Hamm and Huang, `https://arxiv.org/abs/1907.12668`. I need to look into it.

6.3 Least Squares Approximation

We have seen that a linear system $A\mathbf{x} = \mathbf{b}$ has a solution for $\mathbf{x}$ if and only if $\mathbf{b}$ is in the column space of A. In fact, this statement is just the definition of the column space:

$$\begin{aligned}\mathcal{C}(A) &= \{\text{all linear combinations of the columns of } A\},\\ &= \{\text{all vectors of the form } A\mathbf{x} \text{ for some } \mathbf{x}\}.\end{aligned}$$

What happens when $\mathbf{b}$ is not in the column space?

The Problem of Least Squares. Given an $m \times n$ matrix A and an $m \times 1$ column vector $\mathbf{b}$, find an $n \times 1$ column vector $\mathbf{x}$ such that the distance $\|A\mathbf{x} - \mathbf{b}\|$ is minimized.

Obviously a true solution $A\mathbf{x} = \mathbf{b}$ makes $\|A\mathbf{x} - \mathbf{b}\| = 0$. When $\mathbf{b} \notin \mathcal{C}(A)$, the minimum value of $\|A\mathbf{x} - \mathbf{b}\|$ will be strictly positive. The problem is called *least squares approximation* since the length $\|A\mathbf{x} - \mathbf{b}\|$ is minimized if and only if the squared length $\|A\mathbf{x} - \mathbf{b}\|^2$ is minimized, and the squared length is a sum of squares:[33]

$$\|A\mathbf{x} - \mathbf{b}\|^2 = \left\| \begin{pmatrix} \mathbf{a}_1^T\mathbf{x} - b_1 \\ \vdots \\ \mathbf{a}_m^T\mathbf{x} - b_m \end{pmatrix} \right\|^2 = (\mathbf{a}_1^T\mathbf{x} - b_1)^2 + \cdots + (\mathbf{a}_m^T\mathbf{x} - b_m)^2,$$

where $\mathbf{a}_i^T$ is the ith row of A and $\mathbf{b} = (b_1, \ldots, b_m)$. There are two ways to solve this problem:

(1) Calculus

(2) Linear Algebra

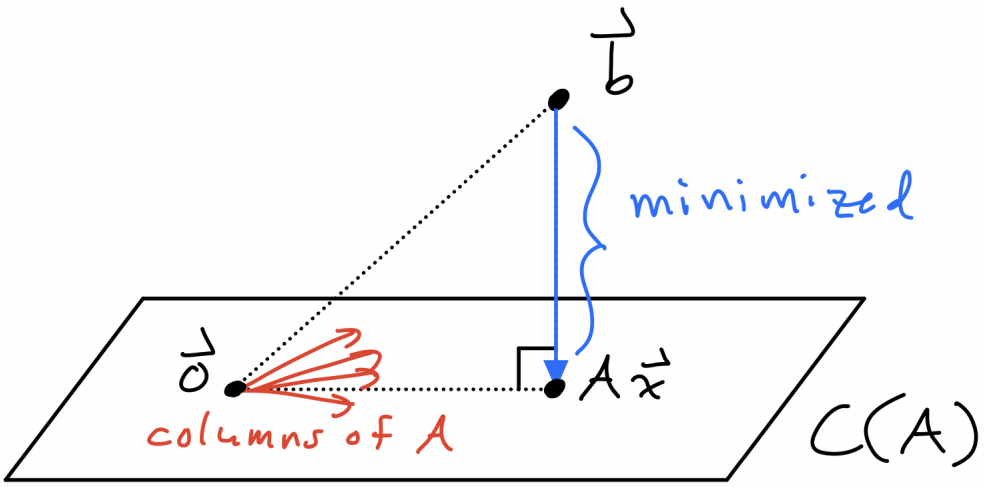
The calculus solution uses the typical method of Lagrange multipliers. This solution is more common in textbooks because every student knows calculus, whereas not every student knows linear algebra. However, the linear algebra solution is conceptually much simpler and is easier to generalize.

The key idea is to view $\|A\mathbf{x} - \mathbf{b}\|$ as the distance between two points in $\mathbb{R}^n$. The expression $A\mathbf{x}$ represents a general point of the column space, while $\mathbf{b}$ is a point that is not in the column space. Here is a picture:

[33] There are certainly other ways to define a "best approximate solution". For example, one could try to minimize the sum of absolute values:

$$|\mathbf{a}_1^T\mathbf{x} - b_1| + \cdots + |\mathbf{a}_m^T\mathbf{x} - b_m|.$$

This is a reasonable idea, but the mathematics is much more difficult. We will see some other methods of approximation after we discuss the singular value decomposition.



For geometric reasons[34] we see that the length of the blue vector $A\mathbf{x} - \mathbf{b}$ is minimized when it is perpendicular to the column space. Since the orthogonal complement of $\mathcal{C}(A)$ is $\mathcal{N}(A^T)$, this happens precisely when

$$\boxed{A^T(A\mathbf{x} - \mathbf{b}) = \mathbf{0}.} \qquad (*)$$

Since you might not remember the details, I will repeat them one more time.[35] Let $\mathbf{a}_i$ be the ith column of A, so that $\mathbf{a}_i^T$ is the ith row of A^T. To say that $A\mathbf{x} - \mathbf{b}$ is perpendicular to the column space means that $A\mathbf{x} - \mathbf{b}$ is perpendicular to each column. That is, we must have

$$\mathbf{a}_i^T(A\mathbf{x} - \mathbf{b}) = \mathbf{a}_i \bullet (A\mathbf{x} - \mathbf{b}) = 0 \quad \text{for all } i.$$

But this is equivalent to saying that $A^T(A\mathbf{x} - \mathbf{b}) = \mathbf{0}$ because

$$A^T(A\mathbf{x} - \mathbf{b}) = \begin{pmatrix} - & \mathbf{a}_1^T & - \\ & \vdots & \\ - & \mathbf{a}_m^T & - \end{pmatrix} (A\mathbf{x} - \mathbf{b}) = \begin{pmatrix} \mathbf{a}_1^T(A\mathbf{x} - \mathbf{b}) \\ \vdots \\ \mathbf{a}_m^T(A\mathbf{x} - \mathbf{b}) \end{pmatrix},$$

which is the zero vector if and only if each component $\mathbf{a}_i^T(A\mathbf{x} - \mathbf{b})$ is zero.

We may proceed to solve equation $(*)$ which is called the *normal equation*:[36]

$$\begin{aligned} A^T(A\mathbf{x} - \mathbf{b}) &= \mathbf{0} \\ A^TA\mathbf{x} - A^T\mathbf{b} &= \mathbf{0} \\ A^TA\mathbf{x} &= A^T\mathbf{b}. \end{aligned}$$

[34] Ultimately, this follows from the triangle inequality.

[35] David Hilbert said that every idea must be repeated five times before the students will remember it. See the very interesting biography of Hilbert by Constance Reid.

[36] The word normal here indicates that $A\mathbf{x} - \mathbf{b}$ is perpendicular to the columns of A.

Whereas the equation $A\mathbf{x} = \mathbf{b}$ did not have a solution, it is worth noting that the normal equation $A^TA\mathbf{x} = A^T\mathbf{b}$ always has a solution. To see this, we only need to check that $A^T\mathbf{b}$ is in the column space $\mathcal{C}(A^TA)$. In the previous section on the matrices A^TA and AA^T we proved the key fact that $\mathcal{N}(A^TA) = \mathcal{N}(A)$, which implies that

$$\mathcal{R}(A^TA) = \mathcal{N}(A^TA)^\perp = \mathcal{N}(A)^\perp = \mathcal{R}(A).$$

But then we must have

$$\mathcal{C}(A^TA) = \mathcal{R}((A^TA)^T) = \mathcal{R}(A^TA) = \mathcal{R}(A) = \mathcal{C}(A^T).$$

This implies that any vector in the column space of A^T, for example $A^T\mathbf{b}$, is in the column space of A^TA, so can be expressed in the form $A^TA\mathbf{x}$.

In general, suppose that A has shape $m \times n$ and rank r. Then the solution of the normal equation $A^TA\mathbf{x} = A^T\mathbf{b}$ is an affine subspace of $\mathbb{R}^n$ that is parallel to the nullspace $\mathcal{N}(A^TA) = \mathcal{N}(A)$, and so has dimension $n - r$. This solution will be unique if and only if $r = n$, i.e., if and only if A has independent columns. In this case we know from the previous section that $(A^TA)^{-1}$ exists, and hence the unique least squares solution has a symbolic form:

$$\begin{aligned} A^TA\mathbf{x} &= A^T\mathbf{b} \\ \mathbf{x} &= (A^TA)^{-1}A^T\mathbf{b}. \end{aligned}$$

Here is a summary:

- If $\mathbf{b} \in \mathcal{C}(A)$ then the system $A\mathbf{x} = \mathbf{b}$ has an exact solution.
- If $\mathbf{b} \notin \mathcal{C}(A)$ then the system $A\mathbf{x} = \mathbf{b}$ does not have an exact solution.
- The length $\|A\mathbf{x} - \mathbf{b}\|$ is minimized when $A\mathbf{x} - \mathbf{b}$ is perpendicular to $\mathcal{C}(A)$.
- This happens if and only if $A^T(A\mathbf{x} - \mathbf{b}) = \mathbf{0}$, or $A^TA\mathbf{x} = A^T\mathbf{b}$.
- The *normal equation* $A^TA\mathbf{x} = A^T\mathbf{b}$ always has a solution.
- If A has independent columns then A^TA is invertible, so the solution is unique:

$$\mathbf{x} = (A^TA)^{-1}A^T\mathbf{b}.$$

We often use a different notation such as $\hat{\mathbf{x}}$ to denote the least squares solution $A^TA\hat{\mathbf{x}} = A^T\mathbf{b}$, to distinguish it from an exact solution $A\mathbf{x} = \mathbf{b}$. However, if there exists an exact solution $A\mathbf{x} = \mathbf{b}$, then we note that $\hat{\mathbf{x}} = \mathbf{x}$ since multiplying both sides on the left gives

$$\begin{aligned} A\mathbf{x} &= \mathbf{b} \\ A^TA\mathbf{x} &= A^T\mathbf{b}. \end{aligned}$$

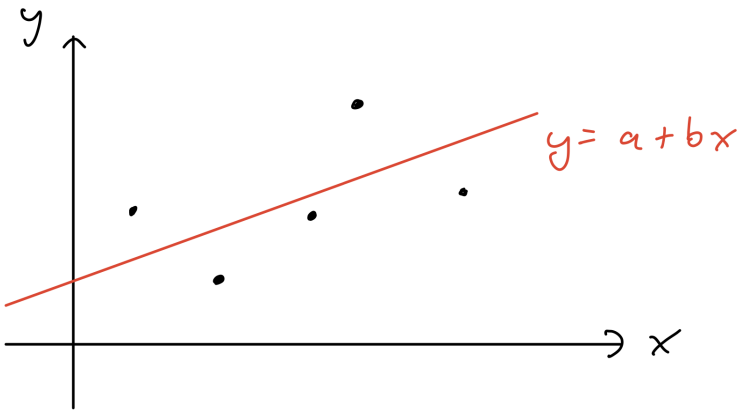
6.4 Examples of Least Squares

The classical application of least squares is to curve fitting. Indeed, this is the purpose for which Gauss invented the method.[37]

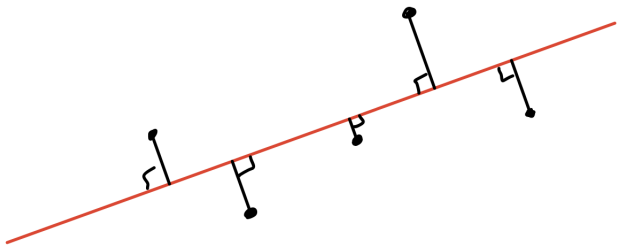
Curve Fitting. Suppose that we have a collection of n data points in the x, y-plane:

$$(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n).$$

We would like to find the line of the form $y = a + bx$ that is the "best fit" for these points:

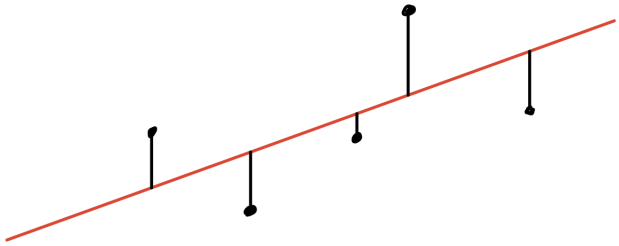


There are different ways one might interpret the word "best". The most obvious definition might be to minimize the orthogonal distances[38] from the points to the line:



This idea is called *total least squares*, or *orthogonal least squares*. It is a hard non-linear problem, which we will solve after discussing the singular value decomposition. In statistics this problem is called *principal component analysis*. It is much easier to minimize the sum of squares of the vertical distances:

[37] He used it to fit the elliptical orbit of the dwarf Planet Ceres to a collection of observed data points.
[38] Typically we want to minimize the sum of squared distances.



This problem is called *ordinary least squares*, or just least squares regression.

Here's how we solve it. We start by being optimistic and assuming that all of the data points fit perfectly on the line, which leads to a system of n linear equations in the two unknowns a and b:

$$\begin{cases} a + bx_1 = y_1, \\ a + bx_2 = y_2, \\ \vdots \\ a + bx_n = y_n, \end{cases}$$

It is an unfortunate feature of curve fitting problems that the roles of variables and constants get switched around, so instead of a system looking like $A\mathbf{x} = \mathbf{b}$ we get a system looking like $X\mathbf{a} = \mathbf{y}$. In our case we have

$$X\mathbf{a} = \mathbf{y}$$
$$\begin{pmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}.$$

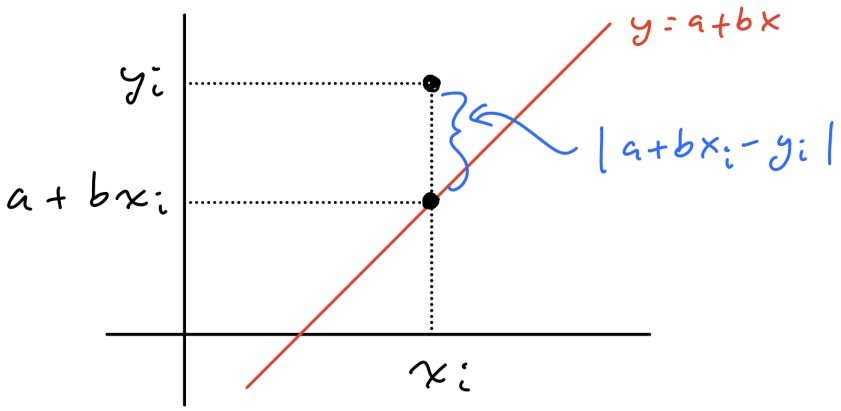
However, this system almost certainly does not have a solution, since any three or more points almost certainly do not fit perfectly on a straight line. Hence we will apply the method of least squares. If the data points do not all have the same x value, then the two columns of X are are independent and we get a unique solution:

$$\begin{aligned} X\mathbf{a} &= \mathbf{y} \\ X^T X\mathbf{a} &= X^T\mathbf{y} \\ \mathbf{a} &= (X^T X)^{-1} X\mathbf{y}. \end{aligned}$$

Recall that this "least squares solution" minimizes the length $\|X\mathbf{a}-\mathbf{y}\|$, hence it also minimizes the squared length $\|X\mathbf{a} - \mathbf{y}\|^2$. In terms of the data points, this becomes

$$\|X\mathbf{a} - \mathbf{y}\|^2 = \left\| \begin{pmatrix} a + bx_i - y_i \\ \vdots \\ a + bx_n - y_n \end{pmatrix} \right\|^2 = \sum (a + bx_i - y_i)^2,$$

which is, indeed, the sum of the squared vertical errors:



To be explicit, the normal equation has the following form, which you might recognize:

$$X^T X\mathbf{a} = X^T\mathbf{y}$$

$$\begin{pmatrix} 1 & \cdots & 1 \\ x_1 & \cdots & x_n \end{pmatrix} \begin{pmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} 1 & \cdots & 1 \\ x_1 & \cdots & x_n \end{pmatrix} \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}$$

$$\begin{pmatrix} n & \sum x_i \\ \sum x_i & \sum x_i^2 \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} \sum y_i \\ \sum x_i y_i \end{pmatrix},$$

which is equivalent to the linear system

$$\begin{cases} an & + & b\sum x_i & = & \sum y_i, \\ a\sum x_i & + & b\sum x_i^2 & = & \sum x_i y_i. \end{cases}$$

This is the form usually presented in introductory statistics courses, when the students don't know linear algebra.

However, the linear algebra formulation is much more powerful because it generalizes easily. For example, we can fit our data to polynomial curve of degree d:

$$y = a_0 + a_1 x + \cdots + \cdots + a_d x^d.$$

Assuming optimistically that all n data points lie on this curve gives a system of n linear equations in the $d+1$ unknown coefficients $a_0, \ldots, a_d$:

$$X\mathbf{a} = \mathbf{y}$$

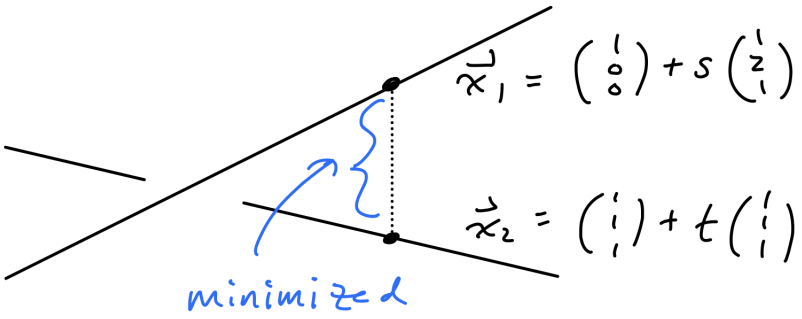
$$\begin{pmatrix} 1 & x_1 & x_1^2 & \cdots & x_1^d \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_n & x_n^2 & \cdots & x_n^d \end{pmatrix} \begin{pmatrix} a_0 \\ \vdots \\ a_d \end{pmatrix} = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}.$$

Then the least squares solution (which minimizes the sum of squares of the vertical errors) is given by the normal equation $X^T X\mathbf{a} = X^T\mathbf{y}$. This equation is much harder to obtain using calculus, and the explicit formulas for the entries of the matrix $X^T X$ are not so nice.

Distance Between Subspaces. Consider the following parametrized lines in $\mathbb{R}^3$:

$$\begin{aligned} L_1: &\quad (1,0,0) + s(1,2,1), \\ L_2: &\quad (1,1,1) + t(1,1,1). \end{aligned}$$

These lines (probably) do not intersect. We would like to find points $\mathbf{x}_1 \in L_1$ and $\mathbf{x}_2 \in L_2$ such that the distance $\|\mathbf{x}_1 - \mathbf{x}_2\|$ is minimized:



We could solve this problem from scratch, but instead we will apply the general theory of least squares. First we assume, optimistically, that the lines intersect, so that

$$\begin{gathered} \mathbf{x}_1 = \mathbf{x}_2 \\ \begin{pmatrix}1\\0\\0\end{pmatrix} + s\begin{pmatrix}1\\2\\1\end{pmatrix} = \begin{pmatrix}1\\1\\1\end{pmatrix} + t\begin{pmatrix}1\\1\\1\end{pmatrix} \\ s\begin{pmatrix}1\\2\\1\end{pmatrix} - t\begin{pmatrix}1\\1\\1\end{pmatrix} = \begin{pmatrix}1\\1\\1\end{pmatrix} - \begin{pmatrix}1\\0\\0\end{pmatrix} \\ \begin{pmatrix}1 & -1\\2 & -1\\1 & -1\end{pmatrix}\begin{pmatrix}s\\t\end{pmatrix} = \begin{pmatrix}0\\1\\1\end{pmatrix}. \end{gathered}$$

Whether this system has an exact solution or not,[39] we can proceed by multiplying on the left by the transpose of the coefficient matrix:

$$\begin{pmatrix}1 & -1\\2 & -1\\1 & -1\end{pmatrix}\begin{pmatrix}s\\t\end{pmatrix} = \begin{pmatrix}0\\1\\1\end{pmatrix}$$

[39] If the system did, unexpectedly, have an exact solution, we would see this at the end.

$$\begin{pmatrix} 1 & 2 & 1 \\ -1 & -1 & -1 \end{pmatrix} \begin{pmatrix} 1 & -1 \\ 2 & -1 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} s \\ t \end{pmatrix} = \begin{pmatrix} 1 & 2 & 1 \\ -1 & -1 & -1 \end{pmatrix} \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix}$$
$$\begin{pmatrix} 6 & -4 \\ -4 & 3 \end{pmatrix} \begin{pmatrix} s \\ t \end{pmatrix} = \begin{pmatrix} 3 \\ -2 \end{pmatrix}$$
$$\begin{pmatrix} s \\ t \end{pmatrix} = \begin{pmatrix} 6 & -4 \\ -4 & 3 \end{pmatrix}^{-1} \begin{pmatrix} 3 \\ -2 \end{pmatrix}$$
$$\begin{pmatrix} s \\ t \end{pmatrix} = \frac{1}{2} \begin{pmatrix} 3 & 4 \\ 4 & 6 \end{pmatrix} \begin{pmatrix} 3 \\ -2 \end{pmatrix}$$
$$= \frac{1}{2} \begin{pmatrix} 1 \\ 0 \end{pmatrix}$$
$$= \begin{pmatrix} 1/2 \\ 0 \end{pmatrix}.$$

The least squares solution $(s, t) = (1/2, 0)$ corresponds to the points

$$\mathbf{x}_1 = (1, 0, 0) + \frac{1}{2}(1, 2, 1) = (3/2, 1, 1/2) \quad \text{and} \quad \mathbf{x}_2 = (1, 1, 1) + 0(1, 1, 1) = (1, 1, 1).$$

But what exactly have we minimized here? Recall that the least squares solution of $A\mathbf{x} = \mathbf{b}$ minimizes the distance $\|A\mathbf{x} - \mathbf{b}\|$. In our case we have minimized the distance

$$\left\| \begin{pmatrix} 1 & -1 \\ 2 & -1 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} s \\ t \end{pmatrix} - \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix} \right\| = \left\| \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} + s \begin{pmatrix} 1 \\ 2 \\ 1 \end{pmatrix} - \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} - t \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \right\| = \|\mathbf{x}_1 - \mathbf{x}_2\|,$$

which is exactly what we wanted to do.

More generally, we can use this method to find the distance between any two affine subspaces living in $\mathbb{R}^n$. Recall that an *affine subspace* of $\mathbb{R}^n$ has the form

$$\mathbf{p} + U = \{\text{the set of points } \mathbf{p} + \mathbf{u} \text{ for all } \mathbf{u} \in U\},$$

where $\mathbf{p} \in \mathbb{R}^n$ is a point and $U \subseteq \mathbb{R}^n$ is a linear subspace (i.e., passing through $\mathbf{0}$). For the current discussion, it is convenient to represent a d-dimensional affine subspace as $\mathbf{p} + \mathcal{C}(A)$ for some $n \times d$ matrix A with independent columns. We can also express this as

$$\mathbf{p} + \mathcal{C}(A) = \{\text{the set of points } \mathbf{p} + A\mathbf{x} \text{ for all } \mathbf{x} \in \mathbb{R}^d\}.$$

Now let A and B be matrices of shapes $n \times d$ and $n \times e$, each with independent columns, and consider any two points $\mathbf{a}, \mathbf{b} \in \mathbb{R}^n$. We want to find the distance between the following two subspaces:

$$\mathbf{a} + \mathcal{C}(A) = \{\text{the set of } \mathbf{a} + A\mathbf{x} \text{ for } \mathbf{x} \in \mathbb{R}^d\},$$
$$\mathbf{b} + \mathcal{C}(B) = \{\text{the set of } \mathbf{b} + B\mathbf{y} \text{ for } \mathbf{y} \in \mathbb{R}^e\}.$$

We begin optimistically, by assuming that $\mathbf{a}+\mathcal{C}(A)$ and $\mathbf{b}+\mathcal{C}(B)$ share a common point:

$$\begin{aligned}
\mathbf{a}+A\mathbf{x} &= \mathbf{b}+B\mathbf{y}\\
A\mathbf{x}-B\mathbf{y} &= \mathbf{b}-\mathbf{a}\\
\left(\begin{array}{c|c} A & -B \end{array}\right)\left(\frac{\mathbf{x}}{\mathbf{y}}\right) &= \mathbf{b}-\mathbf{a}\\
C\mathbf{z} &= \mathbf{c},
\end{aligned}$$

where the matrices C, $\mathbf{z}$ and $\mathbf{c}$ have shapes $n\times(d+e)$, $(d+e)\times 1$ and $n\times 1$, respectively. Next we multiply on the left by C^T to obtain

$$\begin{aligned}
C\mathbf{z} &= \mathbf{c}\\
C^TC\mathbf{z} &= C^T\mathbf{c}\\
\left(\frac{A^T}{-B^T}\right)\left(\begin{array}{c|c} A & -B \end{array}\right)\mathbf{z} &= \left(\frac{A^T}{-B^T}\right)\mathbf{c}\\
\left(\begin{array}{c|c} A^TA & -A^TB \\ \hline -B^TA & B^TB \end{array}\right)\mathbf{z} &= \left(\frac{A^T\mathbf{c}}{-B^T\mathbf{c}}\right).
\end{aligned}$$

The matrix C need not have independent columns. However, if the column spaces $\mathcal{C}(A)$ and $\mathcal{C}(B)$ have trivial intersection (i.e., if $\mathcal{C}(A)\cap\mathcal{C}(B)=\{\mathbf{0}\}$), then C **will** have independent columns.[40] In this case the inverse $(C^TC)^{-1}$ exists and we have a unique least squares solution:

$$\left(\frac{\mathbf{x}}{\mathbf{y}}\right)=\left(\begin{array}{c|c} A^TA & -A^TB \\ \hline -B^TA & B^TB \end{array}\right)^{-1}\left(\frac{A^T(\mathbf{b}-\mathbf{a})}{-B^T(\mathbf{b}-\mathbf{a})}\right).$$

To check that this makes sense, we consider the case when

$$\mathbf{a}=\begin{pmatrix}1\\0\\0\end{pmatrix},\quad A=\begin{pmatrix}1\\2\\1\end{pmatrix},\quad \mathbf{b}=\begin{pmatrix}1\\1\\1\end{pmatrix},\quad B=\begin{pmatrix}1\\1\\1\end{pmatrix}.$$

This is just our previous example with $L_1=\mathbf{a}+\mathcal{C}(A)$ and $L_2=\mathbf{b}+\mathcal{C}(B)$. Then we have

$$\begin{aligned}
\left(\frac{\mathbf{x}}{\mathbf{y}}\right) &= \left(\begin{array}{c|c} A^TA & -A^TB \\ \hline -B^TA & B^TB \end{array}\right)^{-1}\left(\frac{A^T(\mathbf{b}-\mathbf{a})}{-B^T(\mathbf{b}-\mathbf{a})}\right)\\
&= \left(\begin{array}{c|c} \begin{pmatrix}1&2&1\end{pmatrix}\begin{pmatrix}1\\2\\1\end{pmatrix} & -\begin{pmatrix}1&2&1\end{pmatrix}\begin{pmatrix}1\\1\\1\end{pmatrix} \\ \hline -\begin{pmatrix}1&1&1\end{pmatrix}\begin{pmatrix}1\\2\\1\end{pmatrix} & \begin{pmatrix}1&1&1\end{pmatrix}\begin{pmatrix}1\\1\\1\end{pmatrix} \end{array}\right)^{-1}\left(\begin{array}{c} \begin{pmatrix}1&2&1\end{pmatrix}\begin{pmatrix}0\\1\\1\end{pmatrix} \\ \hline -\begin{pmatrix}1&1&1\end{pmatrix}\begin{pmatrix}0\\1\\1\end{pmatrix} \end{array}\right)
\end{aligned}$$

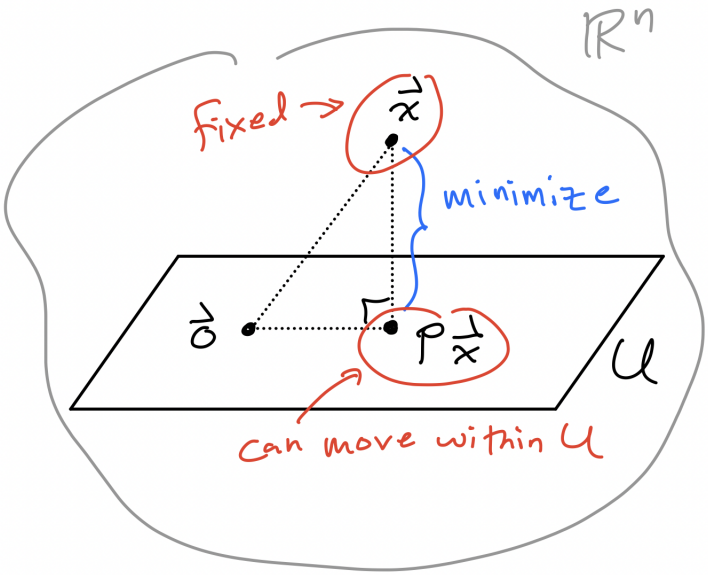
[40] This is a bit tricky so we omit the proof.

$$= \left(\begin{array}{c|c} 6 & -4 \\ \hline -4 & 3 \end{array}\right)^{-1} \left(\begin{array}{c} 3 \\ \hline -2 \end{array}\right),$$
$$= \left(\begin{array}{c} 1/2 \\ \hline 0 \end{array}\right),$$

which is exactly what we had before.

6.5 Projection Matrices

When solving the least squares problem we implicitly solved the problem of projecting onto a (linear) subspace. Given a linear subspace $U \subseteq \mathbb{R}^n$ and a point $\mathbf{x} \in \mathbb{R}^n$ we want to find the point $\mathbf{y} \in U$ that is closest to $\mathbf{x}$. We will denote the point by $\mathbf{y} = P(\mathbf{x})$ and call it the *projection of* $\mathbf{x}$ *onto* U. Here is a picture:



It is not immediately obvious, but we will see that $P : \mathbb{R}^n \to \mathbb{R}^n$ is a linear function, hence it corresponds to an $n \times n$ matrix. The easiest way to find this matrix is to represent U as a column space. Suppose that $\dim U = d$ and let $\mathbf{a}_1, \ldots, \mathbf{a}_d \in U$ be any basis. Then we can form the $n \times d$ matrix

$$A = \begin{pmatrix} | & & | \\ \mathbf{a}_1 & \cdots & \mathbf{a}_d \\ | & & | \end{pmatrix} \quad \text{so that} \quad U = \mathcal{C}(A).$$

From geometric considerations (the triangle inequality) we see that the distance $\|P(\mathbf{x}) - \mathbf{x}\|$ is minimized when the vector $P(\mathbf{x}) - \mathbf{x}$ is perpendicular to U. And since $U^\perp = \mathcal{C}(A)^\perp = \mathcal{N}(A^T)$, we see that[41]

$$P(\mathbf{x}) - \mathbf{x} \in U^\perp \quad \iff \quad A^T(P(\mathbf{x}) - \mathbf{x}) = \mathbf{0}.$$

[41] We already saw this argument in 6.3 so I went faster this time.

Furthermore, since $P(\mathbf{x}) \in U$ and since $U = \mathcal{C}(A)$ we can write $P(\mathbf{x}) = A\hat{\mathbf{x}}$ for some vector $\hat{\mathbf{x}} \in \mathbb{R}^d$.[42] Thus we have the following two facts about the projection:

- $A^T(P(\mathbf{x}) - \mathbf{x}) = \mathbf{0}$,
- $P(\mathbf{x}) = A\hat{\mathbf{x}}$.

Combining these facts gives

$$\begin{aligned} A^T(A\hat{\mathbf{x}} - \mathbf{x}) &= \mathbf{0} \\ A^TA\hat{\mathbf{x}} - A^T\mathbf{x} &= \mathbf{0} \\ A^TA\hat{\mathbf{x}} &= A^T\mathbf{x} \\ \hat{\mathbf{x}} &= (A^TA)^{-1}A^T\mathbf{x} && A \text{ has independent columns} \\ A\hat{\mathbf{x}} &= A(A^TA)^{-1}A^T\mathbf{x} \\ P(\mathbf{x}) &= A(A^TA)^{-1}A^T\mathbf{x}. \end{aligned}$$

Finally, since this equality holds for any vector $\mathbf{x} \in \mathbb{R}^n$ we conclude that P is linear and is represented by the $n \times n$ matrix $A(A^TA)^{-1}A^T$. We have thus proved the following theorem.

Theorem (Projection Onto a Subspace). Let A be an $n \times d$ matrix with independent columns, so the column space $U = \mathcal{C}(A)$ is a d-dimensional subspace of $\mathbb{R}^n$. The function $P : \mathbb{R}^n \to \mathbb{R}^n$ that projects onto U is linear and is represented by the following matrix:

$$\boxed{P = A(A^TA)^{-1}A^T.}$$

If A has **orthonormal columns** then the formula simplifies because $A^TA = I$:

$$\boxed{P = AA^T.}$$

A given subspace is represented by many matrices. For example, consider the 3×1 matrices

$$A = \begin{pmatrix} 1 \\ -1 \\ 1 \end{pmatrix} \quad \text{and} \quad B = \begin{pmatrix} -2 \\ 2 \\ -2 \end{pmatrix}.$$

The column spaces $\mathcal{C}(A)$ and $\mathcal{C}(B)$ are the same line in $\mathbb{R}^3$. Thus we expect that the matrices $A(A^TA)^{-1}A^T$ and $B(B^TB)^{-1}B^T$ are equal. Indeed, we have

$$\begin{aligned} A(A^TA)^{-1}A^T &= \begin{pmatrix} 1 \\ -1 \\ 1 \end{pmatrix} \left(\begin{pmatrix} 1 & -1 & 1 \end{pmatrix} \begin{pmatrix} 1 \\ -1 \\ 1 \end{pmatrix} \right)^{-1} \begin{pmatrix} 1 & -1 & 1 \end{pmatrix} \\ &= \begin{pmatrix} 1 \\ -1 \\ 1 \end{pmatrix} (3)^{-1} \begin{pmatrix} 1 & -1 & 1 \end{pmatrix} \end{aligned}$$

[42] In the least squares problem the vector $\hat{\mathbf{x}}$ is the main event. Here it is only a temporary convenience.

$$= \frac{1}{3}\begin{pmatrix} 1 \\ -1 \\ 1 \end{pmatrix}\begin{pmatrix} 1 & -1 & 1 \end{pmatrix}$$
$$= \frac{1}{3}\begin{pmatrix} 1 & -1 & 1 \\ -1 & 1 & -1 \\ 1 & -1 & 1 \end{pmatrix}$$

and

$$A(A^TA)^{-1}A^T = \begin{pmatrix} -2 \\ 2 \\ -2 \end{pmatrix}\left(\begin{pmatrix} -2 & 2 & -2 \end{pmatrix}\begin{pmatrix} -2 \\ 2 \\ -2 \end{pmatrix}\right)^{-1}\begin{pmatrix} -2 & 2 & -2 \end{pmatrix}$$
$$= \begin{pmatrix} -2 \\ 2 \\ -2 \end{pmatrix}(12)^{-1}\begin{pmatrix} -2 & 2 & -2 \end{pmatrix}$$
$$= \frac{1}{12}\begin{pmatrix} -2 \\ 2 \\ -2 \end{pmatrix}\begin{pmatrix} -2 & 2 & -2 \end{pmatrix}$$
$$= \frac{1}{12}\begin{pmatrix} 4 & -4 & 4 \\ -4 & 4 & -4 \\ 4 & -4 & 4 \end{pmatrix}$$
$$= \frac{1}{3}\begin{pmatrix} 1 & -1 & 1 \\ -1 & 1 & -1 \\ 1 & -1 & 1 \end{pmatrix}.$$

More generally, if A is $n \times d$ then for any invertible $d \times d$ matrix C we have

$$\mathcal{C}(AC) = \mathcal{C}(A).$$

If A has independent columns then AC also has independent columns, and we observe that

$$\begin{aligned}(AC)((AC)^T(AC))^{-1}(AC)^T &= AC(C^T(A^TA)C)^{-1}C^TA^T \\ &= ACC^{-1}(A^TA)^{-1}(C^T)^{-1}C^TA^T \\ &= AI(A^TA)^{-1}IA^T \\ &= A(A^TA)^{-1}A^T.\end{aligned}$$

So far we have discussed explicit properties of projection in Euclidean space. Next we discuss some abstract properties of projection that apply also to operators on infinite dimensional spaces.

Definition of Abstract Projection. Let V be a real inner product space and consider a linear function $P : V \to V$. If P satisfies certain mild conditions,[43] then there exists a unique linear function $P^T : V \to V$, called the *adjoint of* P, satisfying

$$\langle P\mathbf{u}, \mathbf{v}\rangle = \langle \mathbf{u}, P^T\mathbf{v}\rangle \quad \text{for all } \mathbf{u}, \mathbf{v} \in V.$$

[43]For example, this holds when V is complete and P is continuous.

We say that P is an *abstract projection operator* when

$$P^2 = P \quad \text{and} \quad P^T = P.$$

For example, if V is Euclidean space then the adjoint P^T is just the transpose matrix. In this case we observe that the matrix $P = A(A^TA)^{-1}A^T$ is an abstract projection because

$$\begin{aligned} P^2 &= [A(A^TA)^{-1}A^T][A(A^TA)^{-1}A^T] \\ &= A\cancel{(A^TA)^{-1}(A^TA)}(A^TA)^{-1}A^T \\ &= AI(A^TA)^{-1}A^T \\ &= P \end{aligned}$$

and

$$\begin{aligned} P^T &= [A(A^TA)^{-1}A^T]^T \\ &= (A^T)^T[(A^TA)^{-1}]^T(A)^T \\ &= A[(A^TA)^T]^{-1}A^T \\ &= A[A^T(A^T)^T]^{-1}A^T \\ &= A(A^TA)^{-1}A^T \\ &= P. \end{aligned}$$

Later we will see that any abstract projection matrix satisfying $P^2 = P$ and $P^T = P$ is a "real" (i.e., geometric) projection, hence it can be represented as $P = A(A^TA)^{-1}A^T$. To summarize: For any square matrix P we have

$$\boxed{P^2 = P \text{ and } P^T = P \quad \iff \quad P = A(A^TA)^{-1}A^T \text{ for some } A.}$$

I think that's pretty surprising. In fact, there is a more general version:[44] For any square matrix P we have

$$\boxed{P^2 = P \quad \iff \quad P = A(B^TA)^{-1}B^T \text{ for some } A \text{ and } B.}$$

If P has shape $n \times n$ and rank d then the matrices A and B both have shape $n \times d$ and independent columns. Geometrically, this is a "non-orthogonal projection". It projects all points onto the column space of A, but it does this at a strange angle that is perpendicular to the column space of B.

For example, suppose we want to project onto the line $t(1,1)$ in $\mathbb{R}^2$ in a direction that is perpendicular to $(3,1)$. Then we can take

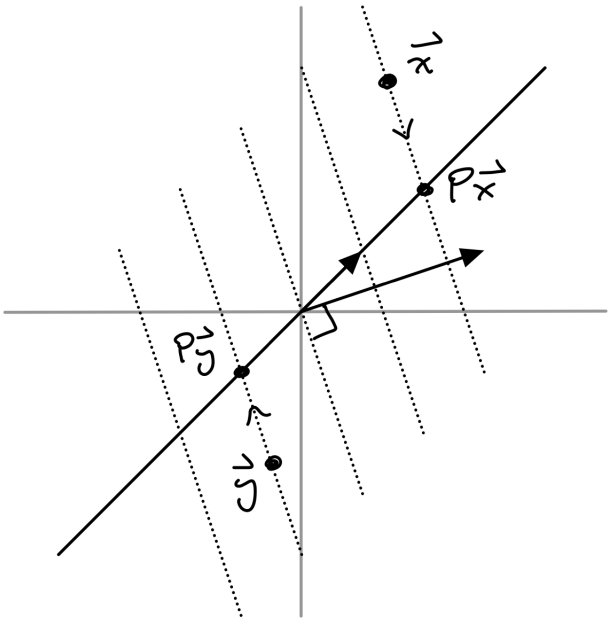
$$A = \begin{pmatrix} 1 \\ 1 \end{pmatrix} \quad \text{and} \quad B = \begin{pmatrix} 3 \\ 1 \end{pmatrix}$$

[44] Maybe we'll prove this later; maybe not. Here are some links:
`https://math.stackexchange.com/questions/600745/are-idempotent-matrices-always-diagonalizable`
`https://math.stackexchange.com/questions/2817221/decomposition-of-idempotent-matrix`

to get

$$P = A(B^TA)^{-1}B^T = \begin{pmatrix}1\\1\end{pmatrix}\left(\begin{pmatrix}3 & 1\end{pmatrix}\begin{pmatrix}1\\1\end{pmatrix}\right)^{-1}\begin{pmatrix}3 & 1\end{pmatrix}$$
$$= \begin{pmatrix}1\\1\end{pmatrix}(4)^{-1}\begin{pmatrix}3 & 1\end{pmatrix}$$
$$= \frac{1}{4}\begin{pmatrix}1\\1\end{pmatrix}\begin{pmatrix}3 & 1\end{pmatrix}$$
$$= \frac{1}{4}\begin{pmatrix}3 & 1\\3 & 1\end{pmatrix}.$$

Picture:



Projection Matrices Come in Pairs.[45] To end this section, I want to observe that projection matrices come in pairs. Let P be a projection matrix satisfying

$$P^2 = P \quad \text{and} \quad P^T = P.$$

Then the matrix $Q = I - P$ is also a projection since

$$Q^2 = (I-P)^2 = I^2 - 2P + P^2 = I - 2P + P = I - P = Q$$

and

$$Q^T = (I-P)^T = I^T - P^T = I - P = Q.$$

[45]This topic does not apply very well to infinite dimensional vector spaces, since one of the pair will have infinite rank.

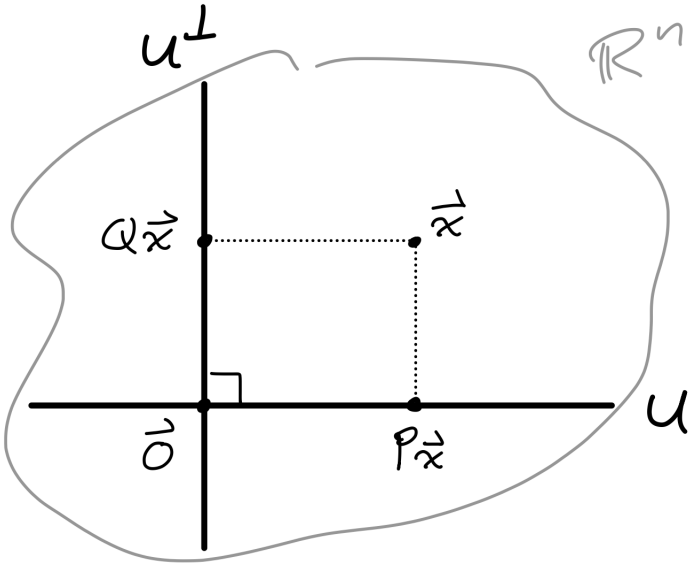
Furthermore, we observe that

$$PQ = QP = P^2 - P = P - P = O.$$

Thus we have the following situation:

- P and Q are projections,
- $P + Q = I$,
- $PQ = O$.

Suppose that P and Q have shape $n \times n$. If P is the projection onto a subspace $U \subseteq \mathbb{R}^n$ then Q is the projection onto the orthogonal complement $U^\perp \subseteq \mathbb{R}^n$ and vice versa. We can see this by looking at $\mathbb{R}^n$ "from the side":



For any point $\mathbf{x} \in \mathbb{R}^n$ we know that the four points $\mathbf{0}, \mathbf{x}, P\mathbf{x}, Q\mathbf{x}$ form a rectangle because

$$P\mathbf{x} + Q\mathbf{x} = (P + Q)\mathbf{x} = I\mathbf{x} = \mathbf{x}$$

and

$$(P\mathbf{x}) \bullet (Q\mathbf{x}) = (P\mathbf{x})^T(Q\mathbf{x}) = \mathbf{x}^T P^T Q\mathbf{x} = \mathbf{x}^T PQ\mathbf{x} = \mathbf{x}^T O\mathbf{x} = 0.$$

This pairing sometimes shortens calculations. For example, suppose that we want to find the 3×3 matrix P that projects onto the plane $x - 2y + z = 0$ in $\mathbb{R}^3$. Then the complementary matrix $Q = I - P$ projects onto the line generated by $(1, -2, 1)$, which is easier to calculate:

$$Q = \begin{pmatrix} 1 \\ -2 \\ 1 \end{pmatrix} \left(\begin{pmatrix} 1 \\ -2 \\ 1 \end{pmatrix} \begin{pmatrix} 1 & -2 & 1 \end{pmatrix} \right)^{-1} \begin{pmatrix} 1 & -2 & 1 \end{pmatrix}$$

$$= \begin{pmatrix} 1 \\ -2 \\ 1 \end{pmatrix} (6)^{-1} \begin{pmatrix} 1 & -2 & 1 \end{pmatrix}$$
$$= \frac{1}{6} \begin{pmatrix} 1 \\ -2 \\ 1 \end{pmatrix} \begin{pmatrix} 1 & -2 & 1 \end{pmatrix}$$
$$= \frac{1}{6} \begin{pmatrix} 1 & -2 & 1 \\ -2 & 4 & -2 \\ 1 & -2 & 1 \end{pmatrix}.$$

It follows that

$$\begin{aligned} P &= I - Q \\ &= \frac{1}{6} \begin{pmatrix} 6 & 0 & 0 \\ 0 & 6 & 0 \\ 0 & 0 & 6 \end{pmatrix} - \frac{1}{6} \begin{pmatrix} 1 & -2 & 1 \\ -2 & 4 & -2 \\ 1 & -2 & 1 \end{pmatrix} \\ &= \frac{1}{6} \begin{pmatrix} 5 & 2 & -1 \\ 2 & 2 & 2 \\ -1 & 2 & 5 \end{pmatrix}. \end{aligned}$$

Of course, we could also do this the long way, by first finding a basis for the plane $x-2y+z=0$. Let's take $(1,0,-1)$ and $(0,1,2)$ and form the matrix

$$A = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ -1 & 2 \end{pmatrix}.$$

Then with a bit of work, one can verify that

$$A(A^TA)^{-1}A^T = \frac{1}{6} \begin{pmatrix} 5 & 2 & -1 \\ 2 & 2 & 2 \\ -1 & 2 & 5 \end{pmatrix}.$$

It seems a bit surprising that these two methods give the same answer. To be more precise, consider any complementary subspaces U and $U^\perp$ in $\mathbb{R}^n$, and choose any matrices A and B with independent columns, such that $U = \mathcal{C}(A)$ and $U^\perp = \mathcal{C}(B)$. Then it must be true that

$$A(A^TA)^{-1}A^T + B(B^TB)^{-1}B^T = I,$$

but this seems mysterious. I'll end by giving an argument to make it feel more natural.

Suppose that $\dim U = d$ so that A has shape $n \times d$ and B has shape $n \times (n-d)$. Form the augmented matrix

$$C = \left(\begin{array}{c|c} A & B \end{array} \right),$$

which has shape $n \times n$. Since the columns of A are a basis for U and the columns of B are a basis for $U^\perp$, the columns of C are a basis for the whole space. In particular, C is invertible, which implies that

$$C(C^TC)^{-1}C^T = CC^{-1}(C^T)^{-1}C^T = I.$$

On the other hand, since every column of A is perpendicular to every column of B we know that $A^TB = O$ and $B^TA = O$, hence

$$C^TC = \left(\begin{array}{c} A^T \\ \hline B^T \end{array}\right) \left(\begin{array}{c|c} A & B \end{array}\right) = \left(\begin{array}{c|c} A^TA & A^TB \\ \hline B^TA & B^TB \end{array}\right) = \left(\begin{array}{c|c} A^TA & O \\ \hline O & B^TB \end{array}\right).$$

And since A and B each have independent columns, we know that A^TA and B^TB are invertible, hence

$$(C^TC)^{-1} = \left(\begin{array}{c|c} A^TA & O \\ \hline O & B^TB \end{array}\right)^{-1} = \left(\begin{array}{c|c} (A^TA)^{-1} & O \\ \hline O & (B^TB)^{-1} \end{array}\right).$$

Finally, we observe that

$$\begin{aligned} C(C^TC)^{-1}C^T &= \left(\begin{array}{c|c} A & B \end{array}\right) \left(\begin{array}{c|c} (A^TA)^{-1} & O \\ \hline O & (B^TB)^{-1} \end{array}\right) \left(\begin{array}{c} A^T \\ \hline B^T \end{array}\right) \\ &= \left(\begin{array}{c|c} A(A^TA)^{-1} & B(B^TB)^{-1} \end{array}\right) \left(\begin{array}{c} A^T \\ \hline B^T \end{array}\right) \\ &= A(A^TA)^{-1}A^T + B(B^TB)^{-1}B^T. \end{aligned}$$