Chapter 5

Linear Algebra

The first step in processing data will involve using invertible linear transformations- since they are invertible and relatively easy, we can always reconstruct the original data set. Furthermore, we'll see how the matrix factorization formulas from linear algebra can be used to our advantage in solving "real world" problems.

It can be argued that all of linear algebra can be understood using the *Four Fundamental Subspaces* associated with a matrix. Because these ideas are the foundation, we need an explicit method for analyzing these subspaces- That method will be the *Singular Value Decomposition*. It is unfortunate that most first courses in linear algebra do not cover this material, so we do it here. Again, we cannot stress the importance of this decomposition enough- We will apply this technique throughout the rest of this text.

5.1 Representation, Basis and Dimension

In Chapter 2, we looked at representing data points using different sets of bases. We saw that, if $\mathcal{B} = \left\{\boldsymbol{v}^{(1)}, \ldots, \boldsymbol{v}^{(k)}\right\}$ is a basis for the data, then every data point in our set can be written as:

$$\boldsymbol{x}^{(j)} = \sum_{i=1}^{k} \alpha_i^{(j)} \boldsymbol{v}^{(i)}$$

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

$$\boldsymbol{x}^{(j)} = (x_1^{(j)}, \ldots, x_n^{(j)}) \longrightarrow (\alpha_1^{(j)}, \ldots, \alpha_j^{(j)}) \doteq [\boldsymbol{x}^{(j)}]_{\mathcal{B}}$$

where we call $[\boldsymbol{x}]_{\mathcal{B}}$ the *coordinates of $\boldsymbol{x}$ with respect to the basis $\mathcal{B}$*. Furthermore, if this is a basis for the data, then it is *isomorphic* to $\mathbb{R}^k$. We'll recall the definition:

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

Using this idea, we can define the **dimension of a subspace**: The dimension of a subspace is the number of basis vectors it requires to represent it. This implies that no matter what basis we choose, it always requires the same number (otherwise, the dimension would depend on the basis, and that wouldn't be good!).

Generically, given a linearly independent spanning set $\mathcal{B}$, to compute the coordinates of a data point with respect to $\mathcal{B}$ requires a matrix inversion (or more generally, Gaussian elimination) to solve the equation:

$$\boldsymbol{x} = \alpha_1 \boldsymbol{v}_1 + \alpha_2 \boldsymbol{v}_2 + \ldots + \alpha_n \boldsymbol{v}_n$$

for the coordinates α_i. We can write this in matrix form as:

$$\boldsymbol{x} = [\boldsymbol{v}_1 \ \ldots \ \boldsymbol{v}_n][\boldsymbol{x}]_{\mathcal{B}} = V[\boldsymbol{x}]_{\mathcal{B}}$$

so that, in the general case where we have n basis vectors of $\mathbb{R}^n$,

$$[\boldsymbol{x}]_{\mathcal{B}} = V^{-1}\boldsymbol{x}$$

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

It gets easier if the basis is orthonormal- before continuing, we'll want to recall the following:

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

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

This is the property that makes an orthonormal basis nice to work with- it's inverse is its transpose.

In an orthonormal basis, it is easy to compute the coordinates of a vector $\boldsymbol{x}$ with respect to this basis. Note that, if

$$\boldsymbol{x} = \alpha_1 \boldsymbol{u}_1 + \ldots + \alpha_k \boldsymbol{u}_k$$

Then the coordinate α_j is just a dot product:

$$\boldsymbol{x} \cdot \boldsymbol{u}_j = 0 + \ldots + 0 + \alpha_j \boldsymbol{u}_j \cdot \boldsymbol{u}_j + 0 + \ldots 0$$

We can also interpret each individual coordinate as the projection of $\boldsymbol{x}$ onto the appropriate basis vector:

$$\alpha_j = \mathrm{Proj}_{\boldsymbol{u}_j}(\boldsymbol{x}) = \boldsymbol{u}_j \cdot \boldsymbol{x}$$

Writing this in matrix form, let $\boldsymbol{c} = [\alpha_1, \ldots, \alpha_k]^T$ and $U = [\boldsymbol{u}_1, \ldots, \boldsymbol{u}_k]$, then

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

We summarize our discussion with the following important, but easily proved theorem (the proof is simply to multiply out the expressions and use the properties of an orthonormal basis).

Change of Basis Theorem: Let $\{\boldsymbol{u}_i\}_{i=1}^k$ be a new orthonormal basis for the subspace containing our data (so that this is a subspace of $\mathbb{R}^n$), and let U be the $n \times k$ matrix whose columns are formed from the basis vectors. Then

- *The Coordinates of $\boldsymbol{x}$ with respect to U:*

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

  which says that a change of coordinates can be performed via matrix multiplication.

- *The Reconstruction of $\boldsymbol{x}$ in $\mathbb{R}^n$ is given by:*

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

  where, if the subspace formed by U contains $\boldsymbol{x}$, then $\boldsymbol{x} = \hat{\boldsymbol{x}}$.

Example Let $\boldsymbol{x} = [3, 2, 3]^T$ and let the basis vectors be $\boldsymbol{u}_1 = \frac{1}{\sqrt{2}}[1, 0, 1]^T$ and let $\boldsymbol{u}_2 = [0, 1, 0]^T$. Then

$$[\boldsymbol{x}]_U = U^T \boldsymbol{x} = \begin{bmatrix} 1/\sqrt{2} & 0 & 1/\sqrt{2} \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} 3 \\ 2 \\ 3 \end{bmatrix} = \begin{bmatrix} 3\sqrt{2} \\ 2 \end{bmatrix}$$

which is the two dimensional representation of the (three dimensional) vector $\boldsymbol{x}$. To get $\boldsymbol{x}$ back into $\mathbb{R}^3$, we simply multiply the coordinates by U. (You should verify this).

5.2 The Four Fundamental Subspaces

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

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

This concept creates four fundamental subspaces:

Symbol	Definition	Name	Where
$\mathcal{R}(A)$	$\{\boldsymbol{y} \in \mathbb{R}^m \mid \boldsymbol{y} = A\boldsymbol{x}, \boldsymbol{x} \in \mathbb{R}^n\}$	Columnspace of A	$\mathbb{R}^m$
$\mathcal{N}(A)$	$\{\boldsymbol{x} \in \mathbb{R}^n \mid A\boldsymbol{x} = \boldsymbol{0}\}$	Nullspace of A	$\mathbb{R}^n$
$\mathcal{R}(A^T)$	$\{\boldsymbol{x} \in \mathbb{R}^n \mid \boldsymbol{x} = A^T\boldsymbol{y}, \boldsymbol{y} \in \mathbb{R}^m\}$	Row space of A	$\mathbb{R}^n$
$\mathcal{N}(A^T)$	$\{\boldsymbol{y} \in \mathbb{R}^m \mid A^T\boldsymbol{y} = \boldsymbol{0}\}$	Nullspace of A^T	$\mathbb{R}^m$

The column space is also the image of the function- This is clear if we write: $A = [\boldsymbol{a}_1, \boldsymbol{a}_2, \ldots, \boldsymbol{a}_n]$ so that $A\boldsymbol{x} = x_1\boldsymbol{a}_1 + x_2\boldsymbol{a}_2 + \cdots + x_n\boldsymbol{a}_n$ and we see that Ax is a linear combination of the columns of A.

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

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

$$\begin{array}{rcl} \mathcal{N}(A) & \perp & \mathcal{R}(A^T) \\ \mathcal{N}(A^T) & \perp & \mathcal{R}(A) \end{array}$$

Proof: See the exercises.

Definition: The *rank* of a matrix A is the number of independent columns of A..

Rank Theorem: Let the $m \times n$ matrix A have rank r. Then

$$r + \dim\left(\mathcal{N}(A)\right) = n$$

This is clear if we consider the matrix under Gaussian elimination. This theorem says that the number of pivot columns plus the other columns (which correspond to free variables) is equal to the total number of columns.

We also have that the dimension of the columnspace is equal to the rank. Now let's do a counting argument to get the dimensions of the rest of the spaces. If the dimension of the columnspace (the rank) is r, then the dimension of the nullspace of A is $n - r$ by the rank theorem. Furthermore, the dimension of the nullspace of A^T must be $m - r$, since there are m dimensions total in the range. Lastly, we also get that the dimension of the columnspace of A^T (or the rowspace) must also be equal to the rank, r. Thus we have shown that the dimension of the columnspace of A must be equal to the dimension of the rowspace. Furthermore, we have half of the theorem below:

Theorem: (We'll wait to show this until after Singular Value Decomposition):

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

We might ask what the implications are to a matrix of data- Let's consider some.

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

5.3 Special Mappings: The Projectors

It will be of benefit to understand certain theorems and algorithms as being projections. We have already seen what an orthogonal projection of one vector to another looks like. We present now a special matrix that performs projections.

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

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

Example: The following are two projectors. Their matrix representations are given by:

$$P_1 = \begin{bmatrix} 1 & 0 \\ 1 & 0 \end{bmatrix} \qquad P_2 = \frac{1}{2}\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$$

Some samples of the projections are given in Figure 5.1, where we see that both project to the subspace spanned by $(1,1)^T$.

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

$$P_1\boldsymbol{x} = \begin{bmatrix} 1 & 0 \\ 1 & 0 \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} x \\ x \end{bmatrix}, P_1(P_1\boldsymbol{x}) = \begin{bmatrix} 1 & 0 \\ 1 & 0 \end{bmatrix}\begin{bmatrix} x \\ x \end{bmatrix} = \begin{bmatrix} x \\ x \end{bmatrix}$$

$$P_2\boldsymbol{x} = \frac{1}{2}\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} \frac{x+y}{2} \\ \frac{x+y}{2} \end{bmatrix} = \frac{x+y}{2}\begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

You should verify that $P_2^2\boldsymbol{x} = P_2(P_2(\boldsymbol{x})) = \boldsymbol{x}$.

You can deduce along which direction a point is projected by drawing a straight line from the point $\boldsymbol{x}$ to the point $\mathbb{P}\boldsymbol{x}$. In general, this direction will depend on the point. We denote this direction by the vector $\mathbb{P}\boldsymbol{x} - \boldsymbol{x}$.

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

$$P_1\boldsymbol{x} - \boldsymbol{x} = \begin{bmatrix} 0 \\ x-y \end{bmatrix}, \text{ and } P_2\boldsymbol{x} - \boldsymbol{x} = \begin{bmatrix} \frac{-x+y}{2} \\ \frac{x-y}{2} \end{bmatrix} = \frac{x-y}{2}\begin{bmatrix} -1 \\ 1 \end{bmatrix}$$

You'll notice that in the case of P_2, $P_2\boldsymbol{x} - \boldsymbol{x} = (P_2 - I)\boldsymbol{x}$ is orthogonal to $P_2\boldsymbol{x}$.

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

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

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

Caution: An orthogonal projector need not be an orthogonal matrix. Notice that the projector P_2 from Figure 5.1 was not an orthogonal matrix (that is, $P_2P_2^T \neq I$).

We have two primary sources for projectors:

Example: Let $\boldsymbol{a}$ be an arbitrary, real, non-zero vector. Show that

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

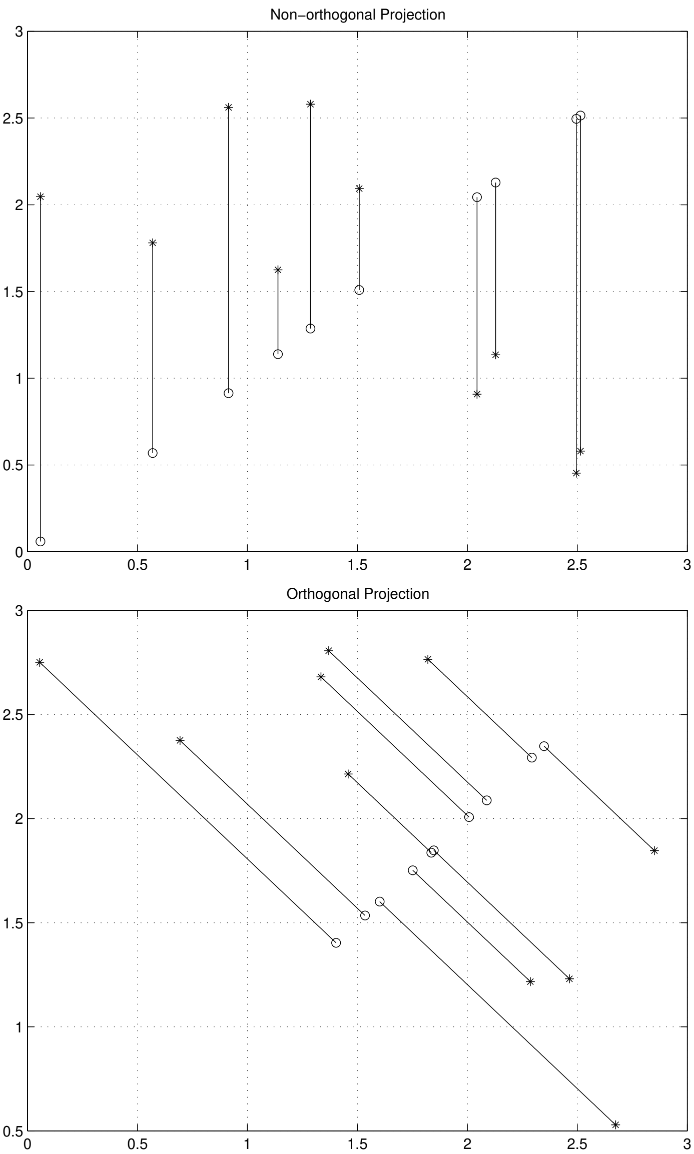


Figure 5.1: Projections P_1 (top picture) and P_2. Asterisks denote the original data point, and circles represent their destination. The line segment follows the direction $P\boldsymbol{x} - \boldsymbol{x}$. Note that P_1 does not project in an orthogonal fashion.

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

> The matrix $\boldsymbol{a}\boldsymbol{a}^T$ has rank one, since every column is a multiple of $\boldsymbol{a}$.
>
> The matrix is a projector:
>
> $$\mathbb{P}^2 = \frac{\boldsymbol{a}\boldsymbol{a}^T}{||\boldsymbol{a}||^2} \cdot \frac{\boldsymbol{a}\boldsymbol{a}^T}{||\boldsymbol{a}||^2} = \frac{1}{||\boldsymbol{a}||^4}\boldsymbol{a}(\boldsymbol{a}^T\boldsymbol{a})\boldsymbol{a}^T = \frac{\boldsymbol{a}\boldsymbol{a}^T}{||\boldsymbol{a}||^2} = \mathbb{P}$$
>
> The matrix is an orthogonal projector, since $\mathbb{P}^T = \mathbb{P}$.

Example: Let $Q = [\boldsymbol{q}_1, \boldsymbol{q}_2, \ldots, \boldsymbol{q}_k]$ be a matrix with orthonormal columns. Then

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

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

Note that this is exactly the property that we discussed earlier with $U = Q$, where the columns of U formed an orthonormal basis. That is, if $\boldsymbol{x}$ is in the columnspace of U, then

$$\boldsymbol{x} = UU^T\boldsymbol{x}$$

but if $\boldsymbol{x}$ is **not** in the columnspace of U:

$$UU^T\boldsymbol{x} \text{ is the projection of } \boldsymbol{x} \text{ into the columnspace of } U$$

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

5.4 Exercises

1. Show that $\mathcal{N}(A) \perp \mathcal{R}(A^T)$. You must show that, for arbitrary $\boldsymbol{x}_1 \in \mathcal{N}(A)$ and $\boldsymbol{x}_2 \in \mathcal{R}(A^T)$, we have $(\boldsymbol{x}_1, \boldsymbol{x}_2) = 0$. Hint: Write A in terms of its rows.

2. If A is $m \times n$, how big can the rank of A possibly be?

3. Show that multiplication by an orthogonal matrix preserves lengths: $||\mathbb{Q}\boldsymbol{x}||_2 = ||\boldsymbol{x}||_2$ (Hint: Use properties of inner products). Conclude that multiplication by $\mathbb{Q}$ represents a rigid rotation.

4. Prove the Pythagorean Theorem by induction: Given a set of n orthogonal vectors $\{\boldsymbol{x}_i\}$

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

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

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

   (b) Since $\boldsymbol{y} - \boldsymbol{x}$ is orthogonal to the range of A, show that
   $$A^T(\boldsymbol{y} - \boldsymbol{x}) = \mathbf{0} \tag{5.1}$$

   (c) Show that there exists $\boldsymbol{v}$ so that Equation (5.1) can be written as:
   $$A^T(A\boldsymbol{v} - \boldsymbol{x}) = 0 \tag{5.2}$$

   (d) Argue that $A^T A$ is invertible, so that Equation (5.2) implies that
   $$\boldsymbol{v} = \left(A^T A\right)^{-1} A^T \boldsymbol{x}$$

   (e) Finally, show that this implies that
   $$\mathbb{P} = A\left(A^T A\right)^{-1} A^T$$

   Note: If A has rank $k < m$, then we will need something different, since $A^T A$ will not be full rank. The missing piece is the singular value decomposition, to be discussed later.

6. The Orthogonal Decomposition Theorem: if $\boldsymbol{x} \in \mathbb{R}^n$ and W is a (non-zero) subspace of $\mathbb{R}^n$, then $\boldsymbol{x}$ can be written *uniquely* as
   $$\boldsymbol{x} = \boldsymbol{w} + \boldsymbol{z}$$
   where $\boldsymbol{w} \in W$ and $\boldsymbol{z} \in W^{\perp}$.

   To prove this, let $\{\boldsymbol{u}_i\}_{i=1}^p$ be an orthonormal basis for W, define $\boldsymbol{w} = (\boldsymbol{x}, \boldsymbol{u}_1)\boldsymbol{u}_1 + \ldots + (\boldsymbol{x}, \boldsymbol{u}_p)\boldsymbol{u}_p$, and define $\boldsymbol{z} = \boldsymbol{x} - \boldsymbol{w}$. Then:

   (a) Show that $\boldsymbol{z} \in W^{\perp}$ by showing that it is orthogonal to every $\boldsymbol{u}_i$.

   (b) To show that the decomposition is unique, suppose it is not. That is, there are two decompositions:
   $$\boldsymbol{x} = \boldsymbol{w}_1 + \boldsymbol{z}_1, \quad \boldsymbol{x} = \boldsymbol{w}_2 + \boldsymbol{z}_2$$
   Show this implies that $\boldsymbol{w}_1 - \boldsymbol{w}_2 = \boldsymbol{z}_2 - \boldsymbol{z}_1$, and that this vector is in both W and $W^{\perp}$. What can we conclude from this?

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

5.5 The Decomposition Theorems

5.5.1 The Eigenvector/Eigenvalue Decomposition

1. **Definition:** Let A be an $n \times n$ matrix. Then an eigenvector-eigenvalue pair is $\boldsymbol{v} \neq \mathbf{0}, \lambda$ where
$$A\boldsymbol{v} = \lambda \boldsymbol{v} \Rightarrow (A - \lambda I)\boldsymbol{v} = \mathbf{0} \tag{5.3}$$

2. **Remark:** If Equation (5.3) has a nontrivial solution, then
$$\det(A - \lambda I) = 0$$
which leads to solving for the roots of a polynomial of degree n. This polynomial is called the *characteristic* polynomial.

3. **Remark:** We solve for the eigenvalues first, then solve for the nullspace of $(A - \lambda_i I)$ by solving
$$(A - \lambda_i I)\boldsymbol{x} = \mathbf{0}$$

4. **Remark:** Note that it is possible that one eigenvalue is repeated. This may or may not correspond with the same number of eigenvectors.

5. **Definition:** If eigenvalue λ is repeated k times, then the *algebraic multiplicity* of λ is k.

6. **Definition:** If eigenvalue λ has k associated independent eigenvectors, λ has *geometric multiplicity* k.

7. **Example:** Compute the eigenvalues and eigenvectors for: (i) the 2×2 identity matrix, (ii) The matrix (in Matlab notation): `[1 2;0 1]`

8. **Theorem:** If a_λ is the algebraic multiplicity of λ and g_λ is the geometric multiplicity, then
$$a_\lambda \geq g_\lambda$$
We won't prove this here.

9. **Definition:** If, for some eigenvalue λ of A, we have that $a_\lambda > g_\lambda$, A is said to be defective.

10. **Definition:** The set of independent eigenvectors associated with an eigenvector λ, together with $\mathbf{0}$ forms a vector space. This space is called the *eigenspace*, and is denoted by E_λ.

11. **Theorem:** If X is square and invertible, then A and $X^{-1}AX$ have the same eigenvalues.

12. **Exercise:** Prove the previous theorem.

13. **Remark:** One method of characterizing eigenvalues in terms of the determinant and trace of a matrix:

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

14. **Remark:** We will be especially interested in symmetric matrices. The rest of this section is devoted to them.

15. **Definition:** A matrix A is *orthogonally diagonalizeable* if there is an orthogonal matrix $\mathbb{Q}$ and diagonal matrix D so that so that $A = \mathbb{Q}D\mathbb{Q}^T$.

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

    (a) A has n real eigenvalues (counting multiplicity).

    (b) For all λ, $a_\lambda = g_\lambda$.

    (c) The eigenspaces are mutually orthogonal.

    (d) A is orthogonally diagonalizeable, with $D = \operatorname{diag}(\lambda_1, \lambda_2, \ldots, \lambda_n)$.

    Some remarks about the Spectral Theorem:

    - We assume that inside each eigenspace, we have an orthonormal basis of eigenvectors. This is not a restriction, since we can always construct such a basis using Gram-Schmidt.
    - If a matrix is real and symmetric, the Spectral Theorem says that its eigenvectors form an orthonormal basis for $\mathbb{R}^n$.
    - The first part is somewhat difficult to prove in that we would have to bring in more machinery than we would like. If you would like to see a proof, it comes from the *Schur Decomposition*, which is given, for example, in "Matrix Computations" by Golub and Van Loan.

17. The following is a proof of part (c). Supply justification for each step: Let $\boldsymbol{v}_1$, $\boldsymbol{v}_2$ be eigenvectors from distinct eigenvalues, λ_1, λ_2. We show that $\boldsymbol{v}_1 \cdot \boldsymbol{v}_2 = 0$:

$$\lambda_1 \boldsymbol{v}_1 \cdot \boldsymbol{v}_2 = (A\boldsymbol{v}_1)^T \boldsymbol{v}_2 = \boldsymbol{v}_1^T A^T \boldsymbol{v}_2 = \boldsymbol{v}_1^T A \boldsymbol{v}_2 = \lambda_2 \boldsymbol{v}_1 \cdot \boldsymbol{v}_2$$

    Now, $(\lambda_1 - \lambda_2)\boldsymbol{v}_1 \cdot \boldsymbol{v}_2 = 0$.

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

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

so that:

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

so finally:

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

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

19. **Matlab Exercise:** Verify the spectral decomposition for a symmetric matrix. Type the following into Matlab (the lines that begin with a % denote comments that do not have to be typed in).

```
%Construct a random, symmetric, 6 x 6 matrix:
for i=1:6
  for j=1:i
    A(i,j)=rand;
    A(j,i)=A(i,j);
  end
end

%Compute the eigenvalues of A:
[Q,L]=eig(A);   %NOTE:  A = Q L Q'
                %L is a diagonal matrix

%Now form the spectral sum
S=zeros(6,6); for i=1:6
  S=S+L(i,i)*Q(:,i)*Q(:,i)';
end

max(max(S-A))
```

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

5.5.2 The Singular Value Decomposition

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

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

1. **Remark:** Although A itself is not symmetric, A^TA is $n \times n$ and symmetric. Therefore, it is orthogonally diagonalizable. Let $\{\lambda_i\}_{i=1}^n$ and $V = [\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_n]$ be the eigenvalues and orthonormal eigenvectors.

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

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

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

   where λ_i is an eigenvalue of A^TA.

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

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

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

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

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

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

9. **Remark:** Define

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

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

10. **Remark:** This definition only makes sense sense for the first r vectors $\boldsymbol{v}$ (otherwise, $A\boldsymbol{v}_i = \boldsymbol{0}$). Thus, we'll have to extend the basis to span all of $\mathbb{R}^m$.

11. **Exercise:** Sketch how you might do this.

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

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

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

15. **Exercise:** Let A be $m \times n$. Define the $m \times n$ matrix
$$\Sigma = \operatorname{diag}(\sigma_1, \ldots, \sigma_n)$$
where σ_i is the i^{th} singular value of the matrix A. Show that
$$AV = U\Sigma$$

16. **The Singular Value Decomposition (SVD)** Let A be any $m \times n$ matrix of rank r. Then
$$A = U\Sigma V^T$$
where U, Σ, V are the matrices defined in the previous exercises. That is, U is an orthogonal $m \times m$ matrix, Σ is a diagonal $m \times n$ matrix, and V is an orthogonal $n \times n$ matrix. The $\boldsymbol{u}$'s are called the *left singular vectors* and the $\boldsymbol{v}$'s are called the *right singular vectors*.

17. **Remark:** Keep in mind the following relationship between the right and left singular vectors:
$$\begin{aligned} A\boldsymbol{v}_i &= \sigma_i \boldsymbol{u}_i \\ A^T\boldsymbol{u}_i &= \sigma_i \boldsymbol{v}_i \end{aligned}$$

18. **Computing The Four Subspaces to a matrix** A. Let $A = U\Sigma V^T$ be the SVD of A which has rank r. Be sure that the singular values are ordered from highest to lowest. Then:
    (a) A basis for the columnspace of A, $\mathcal{R}(A)$ is $\{\boldsymbol{u}_i\}_{i=1}^r$
    (b) A basis for nullspace of A, $\mathcal{N}(A)$ is $\{\boldsymbol{v}_i\}_{i=r+1}^n$
    (c) A basis for the rowspace of A, $\mathcal{R}(A^T)$ is $\{\boldsymbol{v}_i\}_{i=1}^r$
    (d) A basis for the nullspace of A^T, $\mathcal{N}(A^T)$ is $\{\boldsymbol{u}_i\}_{i=r+1}^m$

19. We can also visualize the right and left singular values as in Figure 5.2. We think of the $\boldsymbol{v}_i$ as a special orthogonal basis in R^n (Domain) that maps to the ellipse whose axes are defined by $\sigma_i \boldsymbol{u}_i$.

20. The SVD is one of the premier tools of linear algebra, because it allows us to completely reveal everything we need to know about a matrix mapping: The rank, the basis of the nullspace, a basis for the column space, the basis for the nullspace of A^T, and of the row space. This is depicted in Figure 5.3.

21. Lastly, the SVD provides a decomposition of any linear mapping into two "rotations" and a scaling. This will become important later when we try to deduce a mapping matrix from data (See the section on *signal separation*).

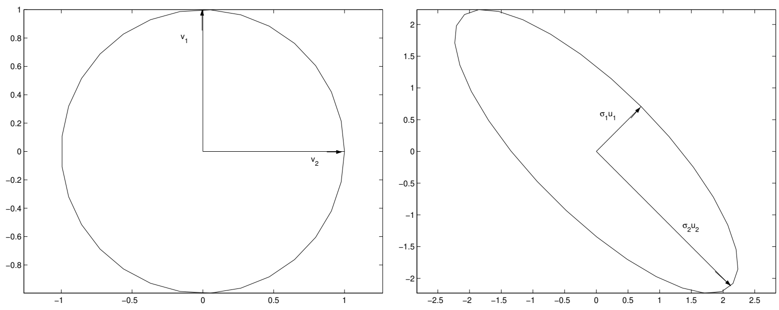


Figure 5.2: The geometric meaning of the right and left singular vectors of the SVD decomposition. Note that $A\boldsymbol{v}_i = \sigma_i \boldsymbol{u}_i$. The mapping $x \to Ax$ will map the unit circle on the left to the ellipse on the right.

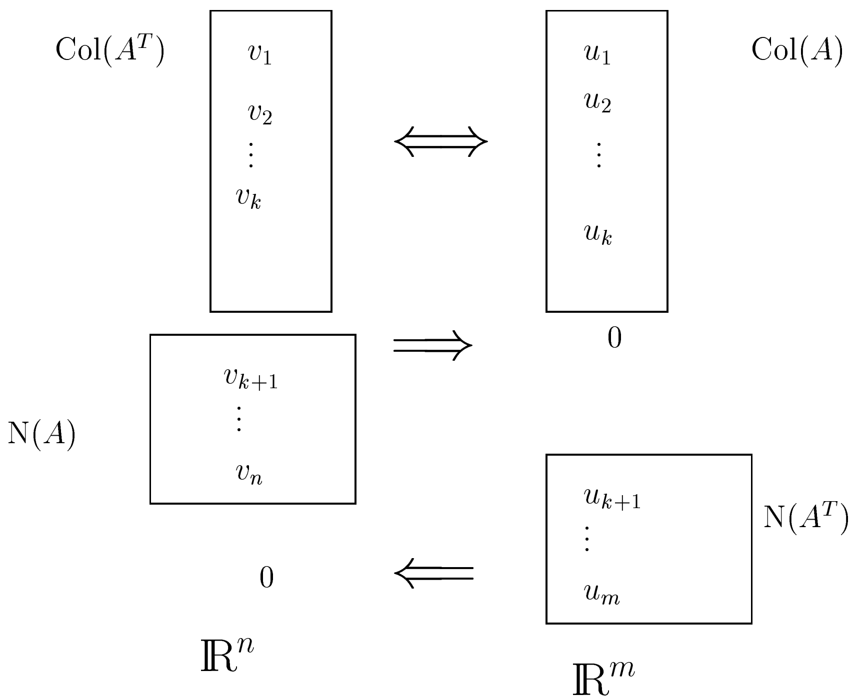


Figure 5.3: The SVD of A (`[U,S,V]=svd(A)`) completely and explicitly describes the 4 fundamental subspaces associated with the matrix, as shown. We have a one to one correspondence between the rowspace and columnspace of A, the remaining $\boldsymbol{v}$'s map to zero, and the remaining $\boldsymbol{u}$'s map to zero (under A^T).

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

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

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

24. **The Reduced SVD** Let A be $m \times n$ with rank r. Then we can write:

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

where $\tilde{U}$ is an $m \times r$ matrix with orthogonal columns, $\tilde{\Sigma}$ is an $r \times r$ *square* matrix, and $\tilde{V}$ is an $n \times r$ matrix.

25. **Theorem:** (Actually, this is just another way to express the SVD). Let $A = U\Sigma V^T$ be the SVD of A, which has rank r. Then:

$$A = \sum_{i=1}^{r} \sigma_i \boldsymbol{u}_i \boldsymbol{v}_i^T$$

Therefore, we can approximate A by the sum of rank one matrices.

26. **Matlab and the SVD** Matlab has the SVD built in. The function specifications are: `[U,S,V]=svd(A)` and `[U,S,V]=svd(A,0)` where the first function call returns the full SVD, and the second call returns a reduced SVD- see Matlab's help file for the details on the second call.

27. **Matlab Exercise:** Image Processing and the SVD. First, in Matlab, load the clown picture:

```
load clown
```

This loads a matrix X and a colormap, *map*, into the workspace. To see the clown, type:

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

We now perform a Singular Value Decomposition on the clown. Type in:

```
[U,S,V]=svd(X);
```

How many vectors are needed to retain a good picture of the clown? Try performing a $k-$dimensional reconstruction of the image by typing:

```
H=U(:,1:k)*S(1:k,1:k)*V(:,1:k)'; image(H)
```

for $k = 5, 10, 20$ and 30. What do you see?

Generalized Inverses

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

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

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

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

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

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

Now we can write:

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

so that

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

(Exercise: Verify that these computations are correct!)

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

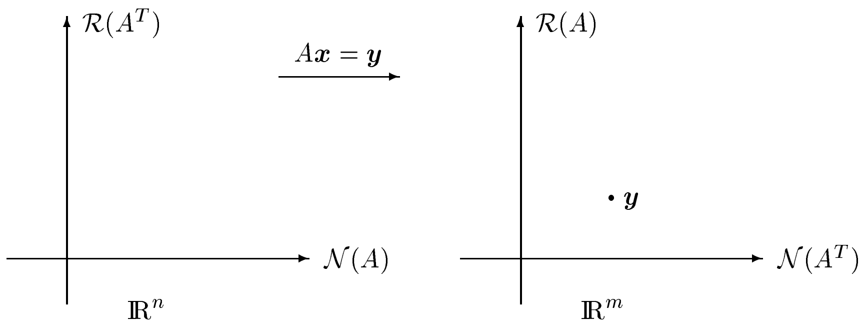
$$A^\dagger = V\Sigma^{-1}U^T$$

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

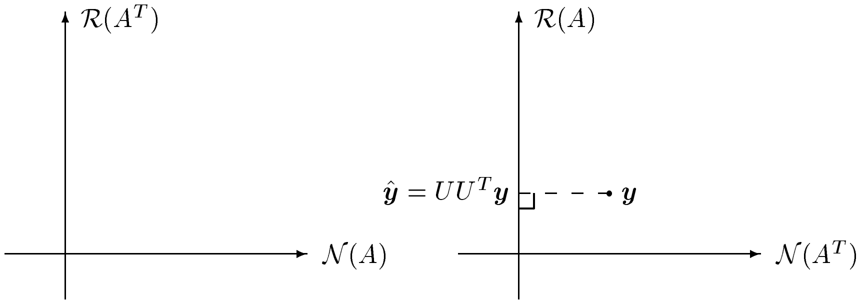
$$\hat{\boldsymbol{x}} = A^\dagger\boldsymbol{y}$$

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

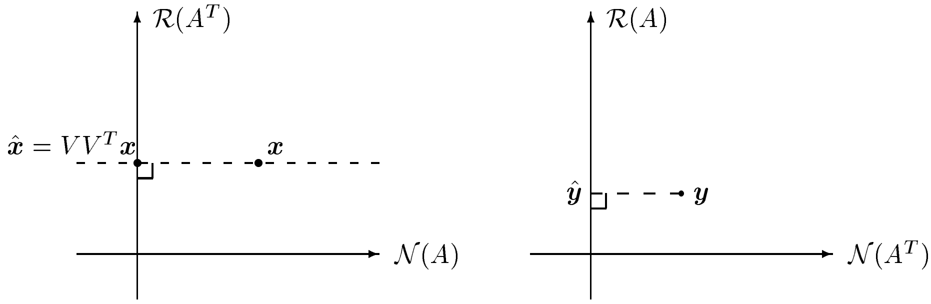
Geometrically, we can understand these computations in terms of the four fundamental subspaces.



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



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



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

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

Exercises

1. Consider

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

   (a) Before solving this problem, what are the dimensions of the four fundamental subspaces?

   (b) Use Matlab to compute the SVD of the matrix A, and solve the problem by computing the pseudoinverse of A directly.

(c) Check your answer explicitly and verify that $\hat{\boldsymbol{x}}$ and $\hat{\boldsymbol{y}}$ are in the rowspace and columnspace. (Hint: If a vector $\boldsymbol{x}$ is already in the rowspace, what is $VV^T\boldsymbol{x}$?)

2. Consider

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

(a) Find the dimensions of the four fundamental subspaces by using the SVD of A (in Matlab).

(b) Solve the problem.

(c) Check your answer explicitly and verify that $\hat{\boldsymbol{x}}$ and $\hat{\boldsymbol{y}}$ are in the rowspace and columnspace.

3. Write the following in Matlab to reproduce Figure 5.2:

```
theta=linspace(0,2*pi,30);
z=exp(i*theta);
X=[real(z);imag(z)]; %The domain points
m=1/sqrt(2);
A=(m*[1,1;1,-1])*[1,0;0,3];
Y=A*X;  %The image of the circle

t=linspace(0,1);
vec1=[0;0]*(1-t)+[0;1]*t;  %The basis vectors v
vec2=[0;0]*(1-t)+[1;0]*t;

Avec1=A*vec1; Avec2=A*vec2; %Image of the basis vectors

figure(1) %The domain
plot(X(1,:),X(2,:),'k',vec1(1,:),vec1(2,:),'k',
     vec2(1,:),vec2(2,:),'k');
axis equal
figure(2) %The image
plot(Y(1,:),Y(2,:),'k',Avec1(1,:),Avec1(2,:),'k',
     Avec2(1,:),Avec2(2,:),'k');
axis equal
```

4. In the previous example, what was the matrix A? The vectors $\boldsymbol{v}$? The vectors $\boldsymbol{u}$? The singular values σ_1, σ_2?

Once you've written these down from the program, perform the SVD of A in Matlab. Are the vectors the same that you wrote down?

NOTE: These show that the singular vectors are not unique- they vary by $\pm\boldsymbol{v}$*, or* $\pm\boldsymbol{u}$*.*

5.6 Interactions Between Subspaces and the SVD

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

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

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

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

$$Z = \begin{bmatrix} A \\ B \end{bmatrix}$$

and find the nullspace of this via the SVD of Z.

Exercise: Prove that, if $\mathbf{v} \in \mathbb{R}^n$, and $Z\mathbf{v} = \mathbf{0}$, then $\mathbf{v} \in \text{Null}(A) \cap \text{Null}(B)$.

Exercise: Prove directly that, if $\mathbf{v} \in \mathbb{R}^n$, and $Z\mathbf{v} \neq \mathbf{0}$, then $\mathbf{v} \in \text{Row}(A) \cup \text{Row}(B)$.

Note that the first exercise proves the second in that:

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

where c is the complement of the set.

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

$$A = U_A \Sigma_A V_A^T, \qquad B = U_B \Sigma_B V_B^T$$

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

5.6.1 Angles Between Subspaces

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

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

The *principal angles* $\theta_1, \ldots, \theta_q \in \left[0, \frac{\pi}{2}\right]$ between F and G are defined recursively by:

$$\cos(\theta_k) = \max_{u \in F} \max_{v \in G} u^T v = u_k^T v_k$$

subject to the additional constraints that u, v be unit length and they are orthogonal to the previously found vectors $u_1, \ldots, u_{k-1}$ and $v_1, \ldots v_{k-1}$. The vectors $u_1, \ldots u_q$ and $v_1, \ldots, v_q$ are called the principal vectors between subspaces F and G.

Some remarks about this definition:

- Since u, v are normalized, the maximum value of $u^T v$ is 1, corresponding to $\theta = 0$. Thus, principal vectors associated with this value of θ are the same. These vectors will also give the an orthonormal basis for the intersection of F, G.

- Since θ is restricted, the smallest value of $u^T v$ is zero, corresponding to $\theta = \pi/2$. The corresponding vectors of F and G will be orthogonal.

- We can define the *distance* between subspaces F and G if $p = q$ by using the largest principal angle, θ_p:

$$\mathrm{dist}(F, G) = \sqrt{1 - \cos^2(\theta_p)} = \sin(\theta_p)$$

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

5.6.2 Computing the principal angles and vectors.

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

Given two sets of orthonormal basis vectors for subspaces F and G (we'll use V_A and V_B found earlier), we can write:

$$u^T v = y^T V_A^T V_B z$$

so that $u = V_A y$ and $v = V_B z$. Thus, if U, V are the matrices whose columns are the principal vectors, then

$$U^T V = Y^T (V_A^T V_B) Z = \mathrm{diag}(\cos(\theta_1), \ldots, \cos(\theta_q)) = D$$

Notice that this is the SVD of $V_A^T V_B$:

$$V_A^T V_B = Y D Z^T$$

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

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

$$F = \text{span}\left\{\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}\right\}, \quad G = \text{span}\left\{\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}\right\}$$

We can use these basis vectors as V_A, V_B respectively. Now,

$$V_A^T V_B = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$$

Since this matrix is already in diagonal form,

$$\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} = I \begin{bmatrix} \cos(\theta_1) & 0 \\ 0 & \cos(\theta_2) \end{bmatrix} I^T$$

so that $Y = Z = I$ and

$$U = V_A I = V_A, \qquad V = V_B I = V_B$$

Therefore, $u_1 = v_1$ corresponding to the subspace intersection, and the distance between the subspaces is 1. We also see that there is no nontrivial intersection between the nullspaces of F, G. We would see this if we took the SVD of $[V_A^T \quad V_B^T]^T$ as suggested at the beginning of this section, since there would be no zero singular values.

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

$$V_A^T V_B = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \end{bmatrix} = Y \cdot 1 \cdot 1 = Y \cdot 1 \cdot Z^T$$

Since this "matrix" is already in diagonal form,

$$U = V_A Y = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \qquad V = V_B \cdot 1 = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$$