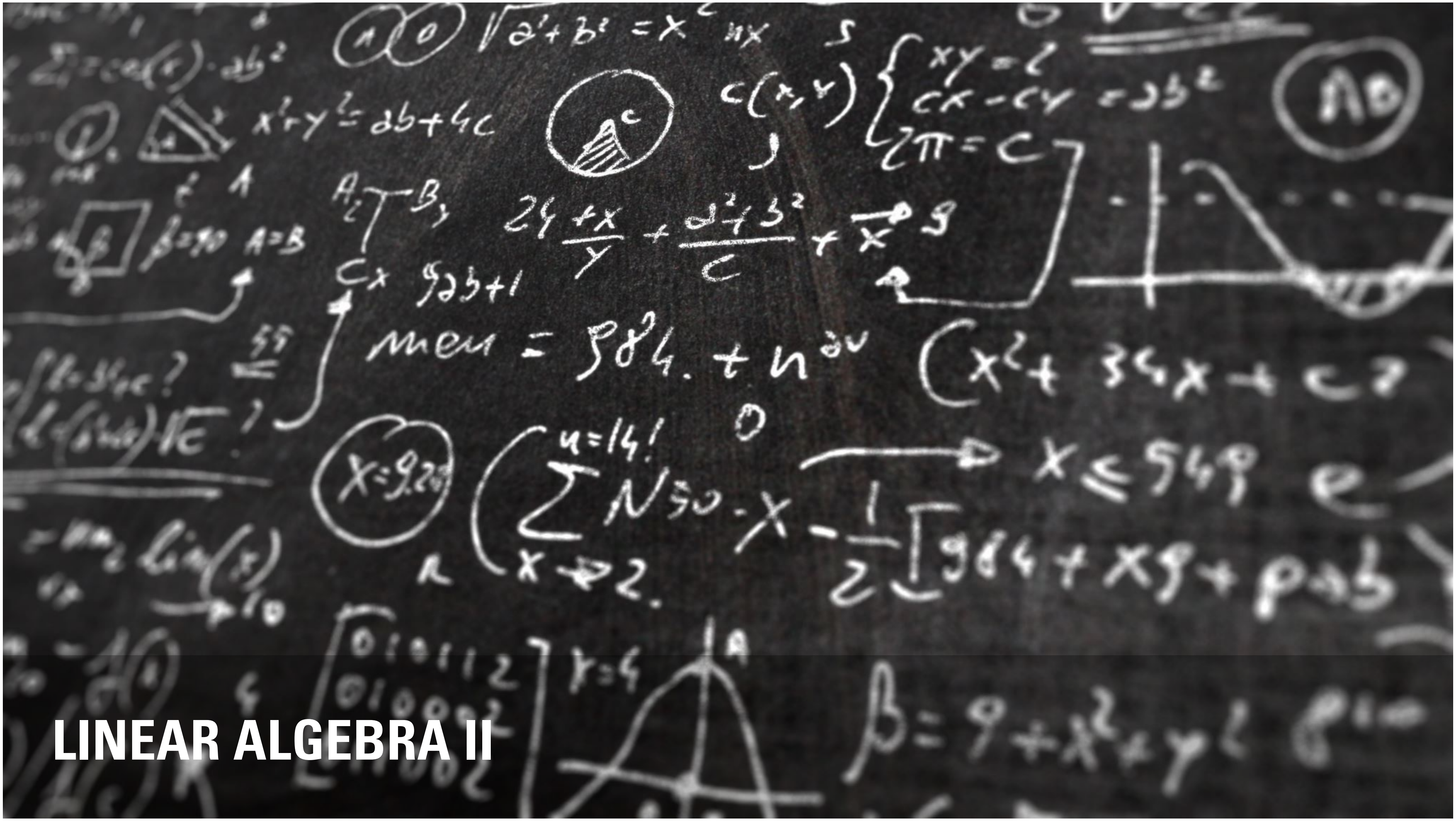


LINEAR ALGEBRA II

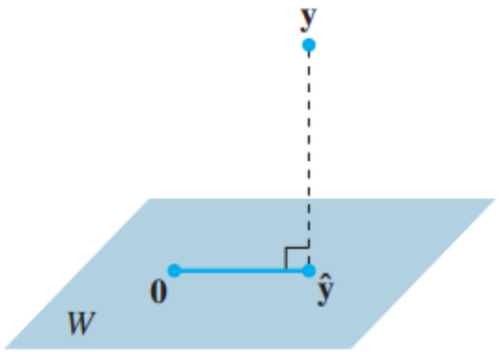
Introduction

The orthogonal projection of a point in $\mathbb{R}^2$ onto a line through the origin has an important analogue in $\mathbb{R}^n$.

Given a vector $\boldsymbol{y}$ and a subspace W in $\mathbb{R}^n$, there is a vector $\hat{\boldsymbol{y}}$ in W such that

(1) $\hat{\boldsymbol{y}}$ is the unique vector in W for which $\boldsymbol{y} - \hat{\boldsymbol{y}}$ is orthogonal to W, and

(2) $\hat{\boldsymbol{y}}$ is the unique vector in W closest to $\boldsymbol{y}$.



The Orthogonal Decomposition Theorem

Let W be a subspace of $\mathbb{R}^n$. Then each $\boldsymbol{y}$ in $\mathbb{R}^n$ can be written uniquely in the form

$$\boldsymbol{y} = \widehat{\boldsymbol{y}} + \boldsymbol{z}$$

where $\widehat{\boldsymbol{y}}$ is in W and $\boldsymbol{z}$ is in $W^{\perp}$.

In fact, if $\{\boldsymbol{u}_1, \dots, \boldsymbol{u}_p\}$ is any orthogonal basis of W, then

$$\widehat{\boldsymbol{y}} = \frac{\boldsymbol{y}.\boldsymbol{u}_1}{\boldsymbol{u}_1.\boldsymbol{u}_1}\boldsymbol{u}_1 + \cdots + \frac{\boldsymbol{y}.\boldsymbol{u}_p}{\boldsymbol{u}_p.\boldsymbol{u}_p}\boldsymbol{u}_p$$

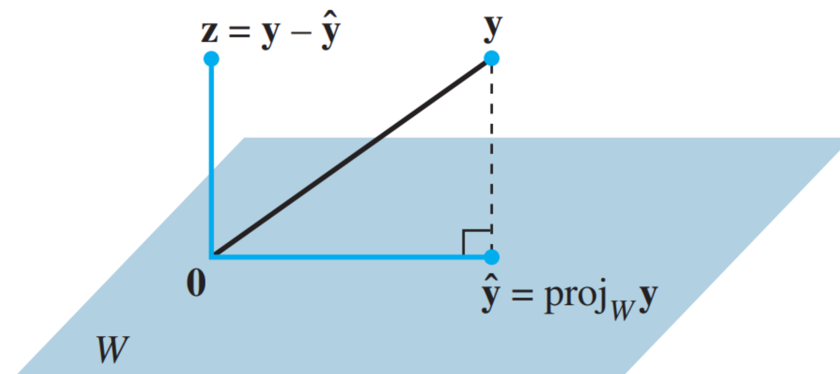
and $\boldsymbol{z} = \boldsymbol{y} - \widehat{\boldsymbol{y}}$.

The Orthogonal projection

Let W be a subspace of $\mathbb{R}^n$. Then each $\boldsymbol{y}$ in $\mathbb{R}^n$ can be written uniquely in the form

$$\boldsymbol{y} = \hat{\boldsymbol{y}} + \boldsymbol{z}$$

The vector $\hat{\boldsymbol{y}}$ is called the **orthogonal projection of $\boldsymbol{y}$ onto W** and often is written as $\mathrm{proj}_W\, \boldsymbol{y}$.



Example 2

Let $\boldsymbol{u}_1 = \begin{bmatrix} 2 \\ 5 \\ -1 \end{bmatrix}, \boldsymbol{u}_2 = \begin{bmatrix} -2 \\ 1 \\ 1 \end{bmatrix}, \boldsymbol{y} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$. Observe that $\{\boldsymbol{u}_1, \boldsymbol{u}_2\}$ is an orthogonal basis for $W = \text{Span}\{\boldsymbol{u}_1, \boldsymbol{u}_2\}$. Write $\boldsymbol{y}$ as the sum of a vector in W and a vector orthogonal to W.

Solution

The orthogonal projection of $\boldsymbol{y}$ onto W is

$$\widehat{\boldsymbol{y}} = \frac{\boldsymbol{y}.\boldsymbol{u}_1}{\boldsymbol{u}_1.\boldsymbol{u}_1}\boldsymbol{u}_1 + \frac{\boldsymbol{y}.\boldsymbol{u}_2}{\boldsymbol{u}_2.\boldsymbol{u}_2}\boldsymbol{u}_2$$

$$= \frac{9}{30}\begin{bmatrix} 2 \\ 5 \\ -1 \end{bmatrix} + \frac{3}{6}\begin{bmatrix} -2 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} -2/5 \\ 2 \\ 1/5 \end{bmatrix}$$

Also, $\boldsymbol{y} - \widehat{\boldsymbol{y}} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} - \begin{bmatrix} -2/5 \\ 2 \\ 1/5 \end{bmatrix} = \begin{bmatrix} 7/5 \\ 0 \\ 14/5 \end{bmatrix}$

Solution (continued)

The orthogonal decomposition theorem ensures that $\boldsymbol{y} - \hat{\boldsymbol{y}}$ is in $W^{\perp}$. To check the calculations, however, it is a good idea to verify that $\boldsymbol{y} - \hat{\boldsymbol{y}}$ is orthogonal to both $\boldsymbol{u}_1$ and $\boldsymbol{u}_2$ and hence to all of W. The desired decomposition of $\boldsymbol{y}$ is

$$\boldsymbol{y} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} = \begin{bmatrix} -2/5 \\ 2 \\ 1/5 \end{bmatrix} + \begin{bmatrix} 7/5 \\ 0 \\ 14/5 \end{bmatrix}$$

Properties of Orthogonal Projections

If $\boldsymbol{y}$ is in $W = \text{Span}\,\{\boldsymbol{u}_1, \dots, \boldsymbol{u}_p\}$, then $\text{proj}_W\, \boldsymbol{y} = \boldsymbol{y}$

The Best Approximation Theorem

Let W be a subspace of $\mathbb{R}^n$, let $\boldsymbol{y}$ be any vector in $\mathbb{R}^n$, and let $\widehat{\boldsymbol{y}}$ be the orthogonal projection of $\boldsymbol{y}$ onto W. Then $\widehat{\boldsymbol{y}}$ is the closest point in W to $\boldsymbol{y}$, in the sense that

$$\|\boldsymbol{y} - \widehat{\boldsymbol{y}}\| < \|\boldsymbol{y} - \boldsymbol{v}\|$$

for all $\boldsymbol{v}$ in W distinct from $\widehat{\boldsymbol{y}}$.

PROJECTION MATRIX

A **projection matrix** is a special type of square matrix that represents a linear transformation orthogonally **projecting** vectors onto a subspace of the vector space.

PROJECTION MATRIX

When we want to project onto the **column space of a matrix *A*** when columns of A are linearly independent. The **projection matrix** is:

$$A\left(\mathrm{A}^{\mathrm{T}}\mathrm{A}\right)^{-1}\mathrm{A}^{\mathrm{T}}$$

This is particularly useful when A is not necessarily a square matrix (e.g., A has more rows than columns).

EXAMPLE:

Let $A = \begin{bmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{bmatrix}$. Find the projection matrix P that projects vectors onto the column space of A.

$$P = \begin{bmatrix} \frac{5}{6} & \frac{1}{3} & -\frac{1}{6} \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ -\frac{1}{6} & \frac{1}{3} & \frac{5}{6} \end{bmatrix}$$

EXAMPLE:

For the projection matrix

$$P = A(A^T A)^{-1} A^T$$

Show that $P^2 = P$.

THE GRAM–SCHMIDT PROCESS

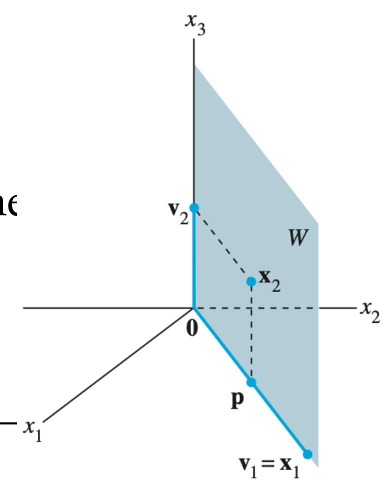
The Gram–Schmidt process is a simple algorithm for producing an orthogonal or orthonormal basis for any nonzero subspace of $\mathbb{R}^n$.

EXAMPLE 1:

Let $W = span\,\{\boldsymbol{x}_1, \boldsymbol{x}_2\}$, where $\boldsymbol{x}_1 = \begin{bmatrix} 3 \\ 6 \\ 0 \end{bmatrix}$ and $\boldsymbol{x}_2 = \begin{bmatrix} 1 \\ 2 \\ 2 \end{bmatrix}$. Construct an orthogonal basis $\{\boldsymbol{v}_1, \boldsymbol{v}_2\}$ for W.

SOLUTION The subspace W is shown in the figure, along with $\boldsymbol{x}_1, \boldsymbol{x}_2$, and the projection p of $\boldsymbol{x}_2$ onto $\boldsymbol{x}_1$. The component of $\boldsymbol{x}_2$ orthogonal to $\boldsymbol{x}_1$is $\boldsymbol{x}_2 - \boldsymbol{p}$, which is in W because it is formed from $\boldsymbol{x}_2$ and a multiple of $\boldsymbol{x}_1$. Let $\boldsymbol{v}_1 = \boldsymbol{x}_1$ and

$$\boldsymbol{v}_2 = \boldsymbol{x}_2 - \boldsymbol{p} = \boldsymbol{x}_2 - \left(\frac{\boldsymbol{x}_2 . \boldsymbol{x}_1}{\boldsymbol{x}_1 . \boldsymbol{x}_1}\right)\boldsymbol{x}_1 = \begin{bmatrix} 0 \\ 0 \\ 2 \end{bmatrix}$$



Then $\{\boldsymbol{v}_1, \boldsymbol{v}_2\}$ is an orthogonal set of nonzero vectors in W. Since, dim $W = 2$, the $\{\boldsymbol{v}_1, \boldsymbol{v}_2\}$ is a basis for W .

EXAMPLE 2:

Let $\boldsymbol{x}_1 = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}, \boldsymbol{x}_2 = \begin{bmatrix} 0 \\ 1 \\ 1 \\ 1 \end{bmatrix}$,and $\boldsymbol{x}_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 1 \end{bmatrix}$. Then $\{\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{x}_3\}$ is clearly linearly independent and thus is a basis for a subspace W of $\mathbb{R}^4$. Construct an orthogonal basis for W.

SOLUTION

Step 1. Let $\boldsymbol{v}_1 = \boldsymbol{x}_1$ and $W_1 = Span\{\boldsymbol{x}_1\} = Span\{\boldsymbol{v}_1\}$.

Step 2. Let $\boldsymbol{v}_2$ be the vector produced by subtracting from $\boldsymbol{x}_2$ its projection onto the subspace W_1. That is, let

$$\boldsymbol{v}_2 = \boldsymbol{x}_2 - proj_{W_1}\,\boldsymbol{x}_2 = \boldsymbol{x}_2 - \frac{\boldsymbol{x}_2.\boldsymbol{v}_1}{\boldsymbol{v}_1.\boldsymbol{v}_1}\boldsymbol{v}_1 = \begin{bmatrix} -3/4 \\ 1/4 \\ 1/4 \\ 1/4 \end{bmatrix}$$

As in Example 1, $\boldsymbol{v}_2$ is the component of $\boldsymbol{x}_2$ orthogonal to $\boldsymbol{x}_1$, and $\{\boldsymbol{v}_1, \boldsymbol{v}_2\}$ is an orthogonal basis for the subspace W_2 spanned by $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$.

SOLUTION

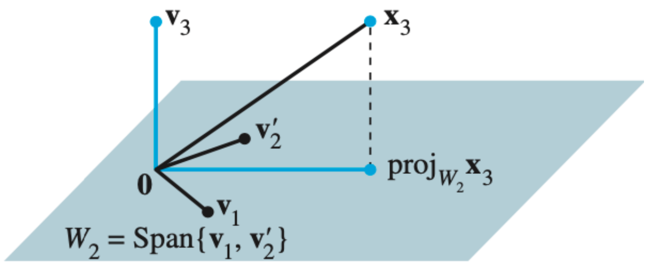
Step 2' (optional). If appropriate, scale $\boldsymbol{v}_2$ to simplify later computations. Since $\boldsymbol{v}_2$ has fractional entries, it is convenient to scale it by a factor of 4 and replace $\{\boldsymbol{v}_1, \boldsymbol{v}_2\}$ by the orthogonal basis

$$\boldsymbol{v}_1 = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}, \boldsymbol{v'}_2 = \begin{bmatrix} -3 \\ 1 \\ 1 \\ 1 \end{bmatrix}$$

Step 3. Let $\boldsymbol{v}_3$ be the vector produced by subtracting from $\boldsymbol{x}_3$ its projection onto the subspace W_2. Use the orthogonal basis $\{\boldsymbol{v}_1, \boldsymbol{v'}_2\}$ to compute this projection onto W_2:

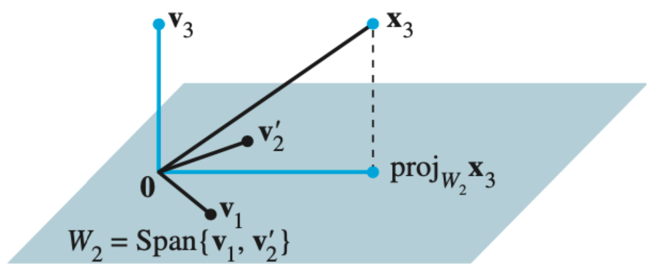
$$proj_{W_2}\,\boldsymbol{x_3} = \frac{\boldsymbol{x}_3.\boldsymbol{v}_1}{\boldsymbol{v}_1.\boldsymbol{v}_1}\boldsymbol{v}_1 + \frac{\boldsymbol{x}_3.\boldsymbol{v'}_2}{\boldsymbol{v'}_2.\boldsymbol{v'}_2}\boldsymbol{v'}_2 = \begin{bmatrix} 0 \\ 2/3 \\ 2/3 \\ 2/3 \end{bmatrix}$$

Then $\boldsymbol{v}_3$ is the component of $\boldsymbol{x}_3$ orthogonal to W_2, namely,



$$\boldsymbol{v_3} = \boldsymbol{x}_3 - proj_{W_2}\,\boldsymbol{x}_3 = \begin{bmatrix} 0 \\ -2/3 \\ 1/3 \\ 1/3 \end{bmatrix}$$

See the figure for a diagram of this construction. Observe that $\boldsymbol{v}_3$ is in W, because $\boldsymbol{x}_3$ and $\mathrm{proj}_{\mathrm{W}_2}\boldsymbol{x}_3$ are both in W. Thus $\{\boldsymbol{v}_1, \boldsymbol{v}'_2, \boldsymbol{v}_3\}$ is an orthogonal set of nonzero vectors and hence a linearly independent set in W. Note that W is three-dimensional since it was defined by a basis of three vectors. Hence, by the Basis Theorem in Section 4.5, $\{\boldsymbol{v}_1, \boldsymbol{v}'_2, \boldsymbol{v}_3\}$ is an orthogonal basis for W.



THEOREM: THE GRAM-SCHMIDT PROCESS

Given a basis $\{\boldsymbol{x}_1, \boldsymbol{x}_2, \dots, \boldsymbol{x}_p\}$ for a nonzero subspace W of $\mathbb{R}^n$, define

$$\boldsymbol{v}_1 = \boldsymbol{x}_1$$

$$\boldsymbol{v}_2 = \boldsymbol{x}_2 - \frac{\boldsymbol{x}_2.\boldsymbol{v}_1}{\boldsymbol{v}_1.\boldsymbol{v}_1}\,\boldsymbol{v}_1$$

$$\boldsymbol{v}_3 = \boldsymbol{x}_3 - \frac{\boldsymbol{x}_3.\boldsymbol{v}_1}{\boldsymbol{v}_1.\boldsymbol{v}_1}\,\boldsymbol{v}_1 - \frac{\boldsymbol{x}_3.\boldsymbol{v}_2}{\boldsymbol{v}_2.\boldsymbol{v}_2}\,\boldsymbol{v}_2$$

$$\vdots$$

$$\boldsymbol{v}_p = \boldsymbol{x}_p - \frac{\boldsymbol{x}_p.\boldsymbol{v}_1}{\boldsymbol{v}_1.\boldsymbol{v}_1}\,\boldsymbol{v}_1 - \frac{\boldsymbol{x}_p.\boldsymbol{v}_2}{\boldsymbol{v}_2.\boldsymbol{v}_2}\,\boldsymbol{v}_2 - \cdots - \frac{\boldsymbol{x}_p.\boldsymbol{v}_{p-1}}{\boldsymbol{v}_{p-1}.\boldsymbol{v}_{p-1}}\,\boldsymbol{v}_{p-1}$$

THEOREM: THE GRAM-SCHMIDT PROCESS

Then $\{\boldsymbol{v}_1, \dots, \boldsymbol{v}_p\}$ is an orthogonal basis for W. In addition

$$Span\{\boldsymbol{v}_1, \dots, \boldsymbol{v}_k\} = Span\{\boldsymbol{x}_1, \dots, \boldsymbol{x}_k\}\; for\; 1 \leq k \leq p \qquad (1)$$

ORTHONORMAL BASES

An orthonormal basis is constructed easily from an orthogonal basis $\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_p\}$: simply normalize (i.e., "scale") all the $\boldsymbol{v}_k$. When working problems by hand, this is easier than normalizing each $\boldsymbol{v}_k$ as soon as it is found (because it avoids unnecessary writing of square roots).

PROJECTION MATRIX

When we want to project onto the **column space of a matrix *A*** when columns of A are linearly independent. The **projection matrix** is:

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

This is particularly useful when A is not necessarily a square matrix (e.g., A has more rows than columns).

Let W be the column space of A, then $P\boldsymbol{y}$ will give us the projection of the vector $\boldsymbol{y}$ onto W.

$$proj_W\boldsymbol{y} = P\boldsymbol{y}.$$

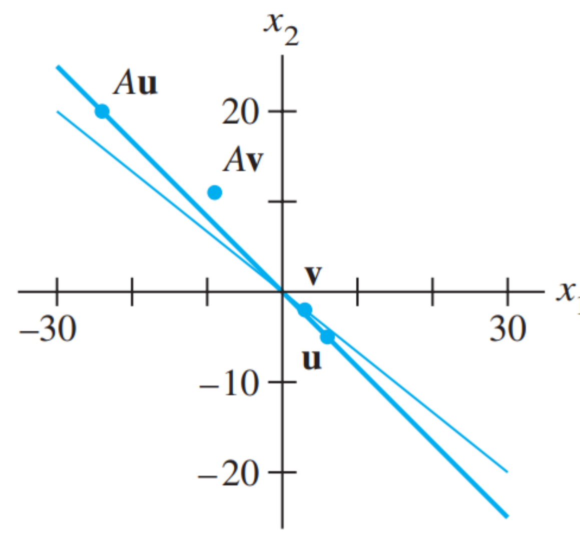
EIGENVALUES AND EIGENVECTORS

Example

Let $A = \begin{bmatrix} 1 & 6 \\ 5 & 2 \end{bmatrix}$, $\boldsymbol{u} = \begin{bmatrix} 6 \\ -5 \end{bmatrix}$ and $\boldsymbol{v} = \begin{bmatrix} 3 \\ -2 \end{bmatrix}$. Find $A\boldsymbol{u}$ and $A\boldsymbol{v}$, what do you observe?

Solution

$$A\boldsymbol{u} = \begin{bmatrix} 1 & 6 \\ 5 & 2 \end{bmatrix} \begin{bmatrix} 6 \\ -5 \end{bmatrix} = \begin{bmatrix} -24 \\ 20 \end{bmatrix} = -4 \begin{bmatrix} 6 \\ -5 \end{bmatrix}$$
$$= -4\boldsymbol{u}$$



$$A\boldsymbol{v} = \begin{bmatrix} 1 & 6 \\ 5 & 2 \end{bmatrix} \begin{bmatrix} 3 \\ -2 \end{bmatrix} = \begin{bmatrix} -9 \\ 11 \end{bmatrix} \neq \lambda \begin{bmatrix} 3 \\ -2 \end{bmatrix}$$

Thus, $\boldsymbol{u}$ is an eigenvector corresponding to an eigenvalue (-4), but $\boldsymbol{v}$ is not an eigenvector of A, because $A\boldsymbol{v}$ is not a multiple of $\boldsymbol{v}$.

Definition

- An **eigenvector** of an $n \times n$ matrix A is a nonzero vector $\boldsymbol{x}$ such that $A\boldsymbol{x} = \lambda\boldsymbol{x}$ for some scalar λ.
- A scalar λ is called an **eigenvalue** of A if there is a nontrivial solution $\boldsymbol{x}$ of
$$A\boldsymbol{x} = \lambda\boldsymbol{x};$$
such an $\boldsymbol{x}$ is called an eigenvector corresponding to λ.
- Note that
$$A\boldsymbol{x} = \lambda\boldsymbol{x} \Leftrightarrow (A - \lambda I)\boldsymbol{x} = \mathbf{0}$$
- In other words, λ is an **eigenvalue** of A if and only if
$$(A - \lambda I)\boldsymbol{x} = \mathbf{0}$$
has a non-trivial solution.

Example

Show that 7 is an eigenvalue of

$$A = \begin{bmatrix} 1 & 6 \\ 5 & 2 \end{bmatrix}$$

and find the corresponding eigenvectors.

Solution

The scalar 7 is an eigenvalue of A if and only if the equation

$$A\boldsymbol{x} = 7\boldsymbol{x} \qquad (1)$$

has a nontrivial solution. But (1) is equivalent to $A\boldsymbol{x} - 7\boldsymbol{x} = \mathbf{0}$, or

$$(A - 7I)\boldsymbol{x} = \mathbf{0} \qquad (2)$$

To solve this homogeneous equation, form the matrix

$$A - 7I = \begin{bmatrix} 1 & 6 \\ 5 & 2 \end{bmatrix} - \begin{bmatrix} 7 & 0 \\ 0 & 7 \end{bmatrix} = \begin{bmatrix} -6 & 6 \\ 5 & -5 \end{bmatrix}$$

Solution (continued)

The columns of $A - 7I$ are obviously linearly dependent, so (2) has nontrivial solutions. Thus 7 *is* an eigenvalue of A. To find the corresponding eigenvectors, use row operations:

$$\begin{bmatrix} -6 & 6 & 0 \\ 5 & -5 & 0 \end{bmatrix} \sim \begin{bmatrix} 1 & -1 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

The general solution has the form $x_2 \begin{bmatrix} 1 \\ 1 \end{bmatrix}$. Each vector of this form with $x_2 \neq 0$ is an eigenvector corresponding to $\lambda = 7$.

Eigen Space

For any $n \times n$ matrix A, λ is an eigenvalue of A if and only if the equation

$$(A - \lambda I)\boldsymbol{x} = \mathbf{0} \qquad (1)$$

has a nontrivial solution.

The set of all solutions of (1) is just the null space of the matrix $A - \lambda I$.

This set is a subspace of $\mathbb{R}^n$ and is called the **eigen space** of A corresponding to λ.

The eigen space consists of the zero vector and all the eigenvectors corresponding to λ.

Example

Let $A = \begin{bmatrix} 4 & -1 & 6 \\ 2 & 1 & 6 \\ 2 & -1 & 8 \end{bmatrix}$. An eigenvalue of A is 2. Find a basis for the corresponding eigenspace.

Solution

$$A - 2I = \begin{bmatrix} 4 & -1 & 6 \\ 2 & 1 & 6 \\ 2 & -1 & 8 \end{bmatrix} - \begin{bmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 2 \end{bmatrix} = \begin{bmatrix} 2 & -1 & 6 \\ 2 & -1 & 6 \\ 2 & -1 & 6 \end{bmatrix}$$

and row reduce the augmented matrix for $(A - 2I)\boldsymbol{x} = \mathbf{0}$:

$$\begin{bmatrix} 2 & -1 & 6 & 0 \\ 2 & -1 & 6 & 0 \\ 2 & -1 & 6 & 0 \end{bmatrix} \sim \begin{bmatrix} 2 & -1 & 6 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

Solution (continued)

At this point, it is clear that 2 is indeed an eigenvalue of A because the equation $(A - 2I)\boldsymbol{x} = \mathbf{0}$ has free variables. The general solution is

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = x_2 \begin{bmatrix} 1/2 \\ 1 \\ 0 \end{bmatrix} + x_3 \begin{bmatrix} -3 \\ 0 \\ 1 \end{bmatrix}; \; x_2 \text{ and } x_3 \text{ free}$$

The eigenspace is a two-dimensional subspace of $\mathbb{R}^3$. A basis is

$$\left\{ \begin{bmatrix} 1 \\ 2 \\ 0 \end{bmatrix}, \begin{bmatrix} -3 \\ 0 \\ 1 \end{bmatrix} \right\}$$

Theorem

The eigenvalues of a triangular matrix are the entries on its main diagonal.

Example

Find the eigenvalues of

$$A = \begin{bmatrix} 3 & 6 & -8 \\ 0 & 0 & 6 \\ 0 & 0 & 2 \end{bmatrix} \text{ and } B = \begin{bmatrix} 4 & 0 & 0 \\ -2 & 1 & 0 \\ 5 & 3 & 4 \end{bmatrix}.$$

Solution

The eigenvalues of A are 3, 0, and 2. The eigenvalues of B are 4 and 1.

Eigen value of 0

What does it mean for a matrix A to have an eigenvalue of 0.

This happens if and only if the equation

$$A\boldsymbol{x} = 0\boldsymbol{x}$$

has a nontrivial solution.

But $A\boldsymbol{x} = 0\boldsymbol{x}$ is equivalent to $A\boldsymbol{x} = \mathbf{0}$,

which has a nontrivial solution if and only if A is not invertible.

Thus,

$$0 \textit{ is an eigenvalue of } A \textit{ if and only if } A \textit{ is not invertible}.$$

Theorem

If $\boldsymbol{v_1}, \cdots, \boldsymbol{v_r}$ are eigenvectors that correspond to distinct eigenvalues $\lambda_1, \cdots, \lambda_r$ of an $n \times n$ matrix A, then the set $\{\boldsymbol{v_1}, \cdots, \boldsymbol{v_r}\}$ is linearly independent.

APPLICATIONS IN AI

Application Area	Role of Eigenvalues/Eigenvectors
PCA / Dimensionality Reduction	Directions of max variance (eigenvectors), importance (eigenvalues)
Face Recognition (Eigenfaces)	Feature extraction from images
Graph AI / Spectral Clustering	Partitioning via Laplacian eigenvectors
Optimization (Hessian)	Curvature → stability, convergence
Neural Nets (RNN stability)	Prevent exploding/vanishing gradients
Markov Chains / RL	Stationary distribution = dominant eigenvector
PageRank	Ranking via largest eigenvalue eigenvector
Kernel Methods (SVM, KPCA)	Nonlinear projections using kernel eigen-decomposition

DEFINITION

A matrix A is said to be diagonalizable if A is similar to a diagonal matrix, that is, if $A = PDP^{-1}$ for some invertible matrix P and some diagonal matrix D.

THEOREM: THE DIAGONALIZATION THEOREM

An $n \times n$ matrix A is diagonalizable if and only if A has n linearly independent eigenvectors.

In fact,

- $A = PDP^{-1}$, with D a diagonal matrix, if and only if the columns of P are n linearly independent eigenvectors of A.
- In this case, the diagonal entries of D are eigenvalues of A that correspond, respectively, to the eigenvectors in P .

EXAMPLE

Diagonalize the following matrix, if possible.

$$A = \begin{bmatrix} 1 & 3 & 3 \\ -3 & -5 & -3 \\ 3 & 3 & 1 \end{bmatrix}$$

- Find the eigenvalues of A.
- Find three linearly independent eigenvectors of A.
- Construct P, whose columns are eigenvectors.
- Construct D, whose diagonal entries are eigenvalues.

- Step 1

To find eigen values solve characteristic equation

$$\det\begin{bmatrix} 1-\lambda & 3 & 3 \\ -3 & -5-\lambda & -3 \\ 3 & 3 & 1-\lambda \end{bmatrix} = -\lambda^3 - 3\lambda^2 + 4$$

$$= -(\lambda - 1)(\lambda + 2)^2$$

The eigenvalues are $\lambda=1$ *and* $\lambda=-2$

- **Step 2**: Find n linearly independent eigenvectors of A

- For each eigenvalue λ, we find a basis for the solution space of $(A - \lambda I)\boldsymbol{x} = \mathbf{0}$.

- For $\lambda = 1$, we solve $(A - 1I)\boldsymbol{x} = \mathbf{0}$:
- The augmented matrix is $\begin{bmatrix} 0 & 3 & 3 & 0 \\ -3 & -6 & -3 & 0 \\ 3 & 3 & 0 & 0 \end{bmatrix}$
- This reduces to $\begin{bmatrix} 1 & 0 & -1 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$

- For $\lambda = -2$, we solve $(A - (-2)I)\boldsymbol{x} = \mathbf{0}$:
- The augmented matrix is $\begin{bmatrix} 3 & 3 & 3 & 0 \\ -3 & -3 & -3 & 0 \\ 3 & 3 & 3 & 0 \end{bmatrix}$
- This reduces to $\begin{bmatrix} 1 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$

BASIS

- The basis for $\lambda = 1$ is $\boldsymbol{v}_1 = \begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix}$

- The basis for $\lambda = -2$ is $\boldsymbol{v}_2 = \begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix}$ and $\boldsymbol{v}_3 = \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix}$

- Now we construct the matrices P and D

- The basis vectors we found form the columns of P:

$$P = \begin{array}{c} \begin{bmatrix} 1 & -1 & -1 \\ -1 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix} \\ \begin{array}{ccc} \uparrow & \uparrow & \uparrow \\ \vec{v}_1 & \vec{v}_2 & \vec{v}_3 \end{array} \end{array}$$

- The corresponding eigenvalues form the diagonal entries of D:

$$D = \begin{bmatrix} 1 & 0 & 0 \\ 0 & -2 & 0 \\ 0 & 0 & -2 \end{bmatrix}$$

Example 2

- Diagonalize $A = \begin{bmatrix} 2 & 4 & 3 \\ -4 & -6 & -3 \\ 3 & 3 & 1 \end{bmatrix}$, if possible.

- **Step 1**: Find the eigenvalues of A
- **Step 2**: Find n linearly independent eigenvectors of A
- **Step 3**: Construct P from the vectors found in Step 2
- **Step 4**: Construct D from the corresponding eigenvalues

EXAMPLE 2

- **Step 1**: Find the eigenvalues of A

$$\det\begin{bmatrix} 2-\lambda & 4 & 3 \\ -4 & -6-\lambda & -3 \\ 3 & 3 & 1-\lambda \end{bmatrix} = -\lambda^3 - 3\lambda^2 + 4$$

- This is the same characteristic equation as in Example 1!

- So the eigenvalues are $\lambda = 1$ and $\lambda = -2$

- **Step 2**: Find n linearly independent eigenvectors of A

- We first solve $(A - 1I)\boldsymbol{x} = \mathbf{0}$:

$$\begin{bmatrix} 1 & 4 & 3 & 0 \\ -4 & -7 & -3 & 0 \\ 3 & 3 & 0 & 0 \end{bmatrix} \rightarrow \begin{bmatrix} 1 & 0 & -1 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

- **Step 2**: Find n linearly independent eigenvectors of A

- Now we solve $(A - (-2)I)\boldsymbol{x} = \mathbf{0}$:

$$\begin{bmatrix} 4 & 4 & 3 & 0 \\ -4 & -4 & -3 & 0 \\ 3 & 3 & 3 & 0 \end{bmatrix} \rightarrow \begin{bmatrix} 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

- We have found two linearly independent eigenvectors:

$$\boldsymbol{v}_1 = \begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix} \text{ and } \boldsymbol{v}_2 = \begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix}$$

- There are no other eigenvalues, and every eigenvector of A is either a multiple of $\boldsymbol{v}_1$ or $\boldsymbol{v}_2$
- Thus it is not possible to find three linearly independent eigenvectors of A, and so A is not diagonalizable!

THEOREM

If A is symmetric, then any two eigenvectors from different eigenspaces are orthogonal.

DEFINITION

- An **orthogonal matrix** is a square invertible matrix U such that $U^{-1} = U^T$.
- Such a matrix has orthonormal columns.
- Any square matrix with orthonormal columns is an orthogonal matrix.
- Surprisingly, such a matrix must have orthonormal rows, too.

DEFINITION

An $n{\times}n$ matrix A is said to be **orthogonally diagonalizable** if there are an orthogonal matrix P (with $P^{-1} = P^T$) and a diagonal matrix D such that

$$A = PDP^T = PDP^{-1}$$

THE SINGULAR VALUE DECOMPOSITION

Introduction

- The diagonalization theorems play a part in many interesting applications.
- Unfortunately, as we know, not all matrices can be factored as $A = PDP^{-1}$ with D diagonal.
- However, a factorization $A = QDP^{-1}$ is possible for any $m \times n$ matrix A.
- A special factorization of this type, called the *singular value decomposition*, is one of the most useful matrix factorizations in applied linear algebra.

The singular value decomposition is based on the following property of the ordinary diagonalization that can be imitated for rectangular matrices:

- The absolute values of the eigenvalues of a symmetric matrix A measure the amounts that A stretches or shrinks certain vectors (the eigenvectors).
- If $A\boldsymbol{x} = \lambda\boldsymbol{x}$ and $\|\boldsymbol{x}\| = 1$, then
$$\|A\boldsymbol{x}\| = \|\lambda\boldsymbol{x}\| = |\lambda|\|\boldsymbol{x}\| = |\lambda| \tag{1}$$
- If λ_1 is the eigenvalue with the greatest magnitude, then a corresponding unit eigenvector $\boldsymbol{v}_1$ identifies a direction in which the stretching effect of A is greatest. That is, the length of $A\boldsymbol{x}$ is maximized when $\boldsymbol{x} = \boldsymbol{v}_1$, and $\|A\boldsymbol{v}_1\| = |\lambda_1|$, by (1).
- This description of $\boldsymbol{v}_1$ and $|\lambda_1|$ has an analogue for rectangular matrices that will lead to the singular value decomposition.

THE SINGULAR VALUES OF AN $m \times n$ MATRIX

Let A be an $m \times n$ matrix.

Then $A^{\mathrm{T}}A$ is symmetric and can be orthogonally diagonalized.

Let $\{\boldsymbol{v}_1, \dots, \boldsymbol{v}_n\}$ be an orthonormal basis for $\mathbb{R}^n$ consisting of eigenvectors of $A^T A$, and let $\lambda_1, \dots, \lambda_n$ be the associated eigenvalues of $A^T A$. Then, for $1 \leq i \leq n$,

$$\begin{aligned}
\|A\boldsymbol{v}_i\|^2 &= (A\boldsymbol{v}_i)^T A\boldsymbol{v}_i = \boldsymbol{v}_i^T A^T A\boldsymbol{v}_i \\
&= \boldsymbol{v}_i^T(\lambda_i \boldsymbol{v}_i) \qquad \textit{Since } \boldsymbol{v}_i \textit{ is an eigenvector of } A^T A \\
&= \lambda_i \qquad\qquad\quad \textit{Since } \boldsymbol{v}_i \textit{ is a unit vector}
\end{aligned}$$

So, the eigenvalues of $A^T A$ are all nonnegative.

THE SINGULAR VALUES OF AN $m \times n$ MATRIX

- The **singular values** of A are the square roots of the eigenvalues of $A^T A$, denoted by $\sigma_1, \dots \sigma_n$, and they are arranged in decreasing order.
- That is, $\sigma_i = \sqrt{\lambda_i}$ for $1 \leq i \leq n$. The **singular values of *A*** are the lengths of the vectors $A\boldsymbol{v}_1, \dots, A\boldsymbol{v}_n$.

EXAMPLE:

Let $A = \begin{bmatrix} 4 & 11 & 14 \\ 8 & 7 & -2 \end{bmatrix}$. Since the eigenvalues of $A^T A$ are 360, 90, and 0, the singular values of A are

$$\sigma_1 = \sqrt{360} = 6\sqrt{10}, \sigma_2 = \sqrt{90} = 3\sqrt{10}, \sigma_3 = 0$$

THEOREM

Suppose $\{\boldsymbol{v}_1, \dots, \boldsymbol{v}_n\}$ is an orthonormal basis of $\mathbb{R}^n$ consisting of eigenvectors of A^TA, arranged so that the corresponding eigenvalues of A^TA satisfy $\lambda_1 \geq \cdots \geq \lambda_n$, and suppose A has r nonzero singular values.

Then $\{A\boldsymbol{v}_1, \dots, A\boldsymbol{v}_r\}$ is an orthogonal basis for Col A, and rank $A = r$.

THE SINGULAR VALUE DECOMPOSITION

The decomposition of A involves an $m \times n$ "diagonal" matrix of the form

$$\Sigma = \begin{bmatrix} D & 0 \\ 0 & 0 \end{bmatrix} \qquad (3)$$

where D is an $r \times r$ diagonal matrix for some r not exceeding the smaller of m and n. (If r equals m or n or both, some or all of the zero matrices do not appear.)

THE SINGULAR VALUE DECOMPOSITION

Let A be an $m \times n$ matrix with rank r. Then there exists an $m \times n$ matrix as in (3) for which the diagonal entries in D are the first r singular values of A, $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r > 0$, and there exist an $m \times m$ orthogonal matrix U and an $n \times n$ orthogonal matrix V such that

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

Any factorization $A = U\Sigma V^T$, is called a **singular value decomposition** (or **SVD**) of A.

THE SINGULAR VALUE DECOMPOSITION

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

- Any factorization $A = U\Sigma V^T$, with U and V orthogonal, Σ as in (3), and positive diagonal entries in D, is called a **singular value decomposition** (or **SVD**) of A.
- The matrices U and V are not uniquely determined by A, but the diagonal entries of Σ are necessarily the singular values of A.
- The columns of U in such a decomposition are called **left singular vectors** of A, and
- the columns of V are called **right singular vectors** of A.

EXAMPLE: CONSTRUCT A SINGULAR VALUE DECOMPOSITION OF

$$A = \begin{bmatrix} 4 & 11 & 14 \\ 8 & 7 & -2 \end{bmatrix}$$

SOLUTION A construction can be divided into three steps.

Step 1. Find an orthogonal diagonalization of $\boldsymbol{A^T A}$. That is, find the eigenvalues of $A^T A$ and a corresponding orthonormal set of eigenvectors. If A had only two columns, the calculations could be done by hand. Larger matrices usually require a matrix program. The eigenvalues of $A^T A$ are $\lambda_1 = 360, \lambda_2 = 90$, and $\lambda_3 = 0$. Corresponding unit eigenvectors are, respectively,

$$\boldsymbol{v}_1 = \begin{bmatrix} 1/3 \\ 2/3 \\ 2/3 \end{bmatrix}, \boldsymbol{v}_2 = \begin{bmatrix} -2/3 \\ 1/3 \\ 2/3 \end{bmatrix}, \boldsymbol{v}_3 = \begin{bmatrix} 2/3 \\ -2/3 \\ 1/3 \end{bmatrix}$$

EXAMPLE: CONSTRUCT A SINGULAR VALUE DECOMPOSITION OF

$$A = \begin{bmatrix} 4 & 11 & 14 \\ 8 & 7 & -2 \end{bmatrix}$$

SOLUTION (Continued) **Step 2**. **Set up V and Σ.** Arrange the eigenvalues of A^TA in decreasing order. The eigenvalues are already listed in decreasing order: 360, 90, and 0. The corresponding unit eigenvectors, $\boldsymbol{v}_1, \boldsymbol{v}_2, and\ \boldsymbol{v}_3$, are the right singular vectors of A.

$$V = [\boldsymbol{v}_1 \quad \boldsymbol{v}_2 \quad \boldsymbol{v}_3] = \begin{bmatrix} 1/3 & -2/3 & 2/3 \\ 2/3 & -1/3 & -2/3 \\ 2/3 & 2/3 & 1/3 \end{bmatrix}$$

The square roots of the eigenvalues are the singular values:

$$\sigma_1 = 6\sqrt{10}, \sigma_2 = 3\sqrt{10}, \sigma_3 = 0$$

The nonzero singular values are the diagonal entries of D. The matrix Σ is the same size as A, with D in its upper left corner and with 0's elsewhere.

$$D = \begin{bmatrix} 6\sqrt{10} & 0 \\ 0 & 3\sqrt{10} \end{bmatrix}, \Sigma = [D \quad 0] = \begin{bmatrix} 6\sqrt{10} & 0 & 0 \\ 0 & 3\sqrt{10} & 0 \end{bmatrix}$$

EXAMPLE: CONSTRUCT A SINGULAR VALUE DECOMPOSITION OF

$$A = \begin{bmatrix} 4 & 11 & 14 \\ 8 & 7 & -2 \end{bmatrix}$$

SOLUTION (Continued) **Step** ***3. Construct U.*** When A has rank r, the first r columns of U are the normalized vectors obtained from $A\boldsymbol{v}_1, \dots, A\boldsymbol{v}_r$.

In this example, A has two nonzero singular values, so rank $A = 2$. Since, $\|A\boldsymbol{v}_1\| = \sigma_1\ and \|A\boldsymbol{v}_2\| = \sigma_2$. Thus

$$\boldsymbol{u}_1 = \frac{1}{\sigma_1} A\boldsymbol{v}_1 = \begin{bmatrix} 3/\sqrt{10} \\ 1/\sqrt{10} \end{bmatrix}, \boldsymbol{u}_2 = \frac{1}{\sigma_2} A\boldsymbol{v}_2 = \begin{bmatrix} 1/\sqrt{10} \\ -3/\sqrt{10} \end{bmatrix}$$

Note that $\{\boldsymbol{u}_1, \boldsymbol{u}_2\}$ is already a basis for $\mathbb{R}^2$. Thus, no additional vectors are needed for U , and $U = [\boldsymbol{u}_1 \ \ \boldsymbol{u}_2]$. The singular value decomposition of A is

$$A = \begin{bmatrix} 3/\sqrt{10} & 1/\sqrt{10} \\ 1/\sqrt{10} & -3/\sqrt{10} \end{bmatrix} \begin{bmatrix} 6\sqrt{10} & 0 & 0 \\ 0 & 3\sqrt{10} & 0 \end{bmatrix} \begin{bmatrix} 1/3 & 2/3 & 2/3 \\ -2/3 & -1/3 & 2/3 \\ 2/3 & -2/3 & 1/3 \end{bmatrix}$$

EXAMPLE: CONSTRUCT A SINGULAR VALUE DECOMPOSITION OF

$$A = \begin{bmatrix} 1 & -1 \\ -2 & 2 \\ 2 & -2 \end{bmatrix}$$

SOLUTION First, compute $A^T A = \begin{bmatrix} 9 & -9 \\ -9 & 9 \end{bmatrix}$. The eigenvalues of $A^T A$ are 18 and 0, with corresponding unit eigenvectors

$$\boldsymbol{v}_1 = \begin{bmatrix} 1/\sqrt{2} \\ -1/\sqrt{2} \end{bmatrix}, \boldsymbol{v}_2 = \begin{bmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{bmatrix}$$

These unit vectors form the columns of V :

$$V = [\boldsymbol{v}_1 \quad \boldsymbol{v}_2] = \begin{bmatrix} 1/\sqrt{2} & 1/\sqrt{2} \\ -1/\sqrt{2} & 1/\sqrt{2} \end{bmatrix}$$

EXAMPLE: CONSTRUCT A SINGULAR VALUE DECOMPOSITION OF

$$A = \begin{bmatrix} 1 & -1 \\ -2 & 2 \\ 2 & -2 \end{bmatrix}$$

SOLUTION The singular values are $\sigma_1 = \sqrt{18} = 3\sqrt{2}$ and $\sigma_2 = 0$. Since there is only one nonzero singular value, the "matrix" D may be written as a single number. That is, $D = 3\sqrt{2}$. The matrix Σ is the same size as A, with D in its upper left corner:

$$\Sigma = \begin{bmatrix} D & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} 3\sqrt{2} & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}$$

To construct U, first construct $A\boldsymbol{v}_1$ and $A\boldsymbol{v}_2$:

$$A\boldsymbol{v}_1 = \begin{bmatrix} 2/\sqrt{2} \\ -4/\sqrt{2} \\ 4/\sqrt{2} \end{bmatrix}, A\boldsymbol{v}_2 = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

EXAMPLE: CONSTRUCT A SINGULAR VALUE DECOMPOSITION OF

$$A = \begin{bmatrix} 1 & -1 \\ -2 & 2 \\ 2 & -2 \end{bmatrix}$$

SOLUTION The only column found for U so far is

$$\boldsymbol{u}_1 = \frac{1}{\sigma_1} A\boldsymbol{v}_1 = \begin{bmatrix} 2/3 \\ -4/3 \\ 4/3 \end{bmatrix}$$

The other columns of U are found by extending the set $\{\boldsymbol{u}_1\}$ to an orthonormal basis for $\mathbb{R}^3$. In this case, we need two orthogonal unit vectors $\boldsymbol{u}_2$ and $\boldsymbol{u}_3$ that are orthogonal to $\boldsymbol{u}_1$. Each vector must satisfy $\boldsymbol{u}_1^T\, \boldsymbol{x} = 0$, which is equivalent to the equation $x_1 - 2x_2 + 2x_3 = 0$. A basis for the solution set of this equation is

$$\boldsymbol{w}_1 = \begin{bmatrix} 2 \\ 1 \\ 0 \end{bmatrix}, \boldsymbol{w}_2 = \begin{bmatrix} -2 \\ 0 \\ 1 \end{bmatrix}$$

EXAMPLE: CONSTRUCT A SINGULAR VALUE DECOMPOSITION OF

$$A = \begin{bmatrix} 1 & -1 \\ -2 & 2 \\ 2 & -2 \end{bmatrix}$$

SOLUTION $\boldsymbol{w}_1 = \begin{bmatrix} 2 \\ 1 \\ 0 \end{bmatrix}, \boldsymbol{w}_2 = \begin{bmatrix} -2 \\ 0 \\ 1 \end{bmatrix}$

Apply the Gram–Schmidt process (with normalizations) to $\{\boldsymbol{w}_1, \boldsymbol{w}_2\}$, and obtain

$$\boldsymbol{u}_2 = \begin{bmatrix} 2/\sqrt{5} \\ 1/\sqrt{5} \\ 0 \end{bmatrix}, \boldsymbol{u}_3 = \begin{bmatrix} -2/\sqrt{45} \\ 4/\sqrt{45} \\ 5/\sqrt{45} \end{bmatrix}$$

$U = [\boldsymbol{u}_1 \;\; \boldsymbol{u}_2 \;\; \boldsymbol{u}_3]$. The singular value decomposition of A is

$$A = \begin{bmatrix} 1/3 & 2/\sqrt{5} & -2/\sqrt{45} \\ -2/3 & 1/\sqrt{5} & 4/\sqrt{45} \\ 2/3 & 0 & 5/\sqrt{45} \end{bmatrix} \begin{bmatrix} 3\sqrt{2} & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 1/\sqrt{2} & -1/\sqrt{2} \\ 1/\sqrt{2} & 1/\sqrt{2} \end{bmatrix}$$

NOTE:

- The matrix $\boldsymbol{AA^T}$ and $\boldsymbol{A^TA}$ are very special in linear algebra.
- Consider any m × n matrix $\boldsymbol{A}$, we can multiply it with A^T to form $\boldsymbol{AA^T}$ and $\boldsymbol{A^TA}$ separately. These matrices are
  - symmetrical,
  - square,
  - at least positive semidefinite (eigenvalues are zero or positive),
  - both matrices have the same positive eigenvalues, and
  - both have the same rank r as A.