Fundamentals and applications of linear algebra

- Diagonalisation and orthogonal diagonalisation of matrices
- Spectral decomposition
- Singular value decomposition
- Quadratic forms

Definition: Square matrix $\mathbf{A}$ is *diagonalisable* if $\mathbf{A} = \mathbf{PDP}^{-1}$ where $\mathbf{D}$ is a diagonal matrix.

Theorem: $n \times n$ matrix $\mathbf{A}$ is diagonalisable if and only if $\mathbf{A}$ has n linearly independent eigenvectors.

The columns of $\mathbf{P}$ are then the eigenvectors of $\mathbf{A}$.

The diagonal entries of $\mathbf{D}$ are the eigenvalues of $\mathbf{A}$ which correspond to the eigenvectors in $\mathbf{P}$.

Corollary: $\mathbf{A}$ is diagonalisable if and only if its eigenvectors form a basis of $\mathbb{R}^n$ (it is called the *eigenvector basis*).

Note: $n \times n$ matrix with n distinct eigenvalues is diagonalisable.

(caution: this is sufficient, but not necessary condition).

Example: for $\mathbf{A} = \begin{bmatrix} 4 & -1 & 6 \\ 2 & 1 & 6 \\ 2 & -1 & 8 \end{bmatrix}$, we know the eigenvalues

$\lambda_1 = 9$ and $\lambda_{2,3} = 2$, and three linearly independent eigenvectors

$$\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \quad (\text{for } \lambda_1) \qquad \text{and} \qquad \begin{bmatrix} 1/2 \\ 1 \\ 0 \end{bmatrix}, \quad \begin{bmatrix} -3 \\ 0 \\ 1 \end{bmatrix} \quad (\text{for } \lambda_{2,3})$$

Then $\mathbf{A} = \mathbf{P}\mathbf{D}\mathbf{P}^{-1}$ where

$$\mathbf{P} = \begin{bmatrix} 1 & 1/2 & -3 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix}, \quad \mathbf{D} = \begin{bmatrix} 9 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 2 \end{bmatrix}, \quad \mathbf{P}^{-1} = \frac{1}{7} \begin{bmatrix} 2 & -1 & 6 \\ -2 & 8 & -6 \\ -2 & 1 & 1 \end{bmatrix}$$

Definition: matrix $\mathbf{A}$ is called symmetric if $\mathbf{A}^\mathsf{T} = \mathbf{A}$.

Theorem: If $\mathbf{A}$ is symmetric, then any two eigenvectors from different eigenspaces (corresponding to distinct eigenvalues) are orthogonal.

Proof: Let $\mathbf{v}_1$ and $\mathbf{v}_2$ be eigenvectors that correspond to distinct eigenvalues $\lambda_1 \neq \lambda_2$. We must show that $\mathbf{v}_1 \cdot \mathbf{v}_2 = 0$.

$$\lambda_1 \mathbf{v}_1 \cdot \mathbf{v}_2 = (\lambda_1 \mathbf{v}_1)^\mathsf{T} \mathbf{v}_2 = (\mathbf{A}\mathbf{v}_1)^\mathsf{T} \mathbf{v}_2 = (\mathbf{v}_1^\mathsf{T} \mathbf{A}^\mathsf{T}) \mathbf{v}_2 = (\mathbf{v}_1^\mathsf{T} \mathbf{A}) \mathbf{v}_2$$

$$= \mathbf{v}_1^\mathsf{T} (\mathbf{A}\mathbf{v}_2) = \mathbf{v}_1^\mathsf{T} (\lambda_2 \mathbf{v}_2) = \lambda_2 \mathbf{v}_1^\mathsf{T} \mathbf{v}_2 = \lambda_2 \mathbf{v}_1 \cdot \mathbf{v}_2$$

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

But $\lambda_1 - \lambda_2 \neq 0$ hence $\mathbf{v}_1 \cdot \mathbf{v}_2 = 0$.

Definition: Matrix $\mathbf{A}$ is *orthogonally diagonalisable* if there are an orthogonal matrix $\mathbf{P}$ (that is $\mathbf{P}^{-1} = \mathbf{P}^{\mathsf{T}}$) and a diagonal matrix $\mathbf{D}$ such that

$$\mathbf{A} = \mathbf{P}\mathbf{D}\mathbf{P}^{\mathsf{T}}$$

Note: If $\mathbf{A}$ is orthogonally diagonalisable then

$$\mathbf{A}^{\mathsf{T}} = \left(\mathbf{P}\mathbf{D}\mathbf{P}^{\mathsf{T}}\right)^{\mathsf{T}} = \left(\mathbf{P}^{\mathsf{T}}\right)^{\mathsf{T}}\mathbf{D}^{\mathsf{T}}\mathbf{P}^{\mathsf{T}} = \mathbf{P}\mathbf{D}\mathbf{P}^{\mathsf{T}} = \mathbf{A}$$

and therefore $\mathbf{A}$ is symmetric. In fact (but not so quick to proove):

Theorem: An $n \times n$ matrix $\mathbf{A}$ is orthogonally diagonalisable if and only if $\mathbf{A}$ is a symmetric matrix.

Example: orthogonally diagonalise $\mathbf{A} = \begin{bmatrix} 3 & -2 & 4 \\ -2 & 6 & 2 \\ 4 & 2 & 3 \end{bmatrix}$

Solution: characteristic equation for this matrix is

$$0 = -\lambda^3 + 12\lambda^2 - 21\lambda - 98 = \cdots = -(\lambda - 7)^2(\lambda + 2)$$

Eigenvalues are $\lambda_{1,2} = 7$ (with multiplicity 2), and $\lambda_3 = -2$.

$$\mathbf{A} - 7\mathbf{I} = \begin{bmatrix} -4 & -2 & 4 \\ -2 & -1 & 2 \\ 4 & 2 & -4 \end{bmatrix} \to \begin{bmatrix} 1 & 1/2 & -1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \Rightarrow \left\{ \begin{bmatrix} -1 \\ 2 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} \right\}$$

$$\mathbf{A} + 2\mathbf{I} = \begin{bmatrix} 5 & -2 & 4 \\ -2 & 8 & 2 \\ 4 & 2 & 5 \end{bmatrix} \to \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 1/2 \\ 0 & 0 & 0 \end{bmatrix} \Rightarrow \left\{ \begin{bmatrix} -2 \\ -1 \\ 2 \end{bmatrix} \right\}$$

$\mathbf{v}_1 = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}$ and $\mathbf{v}_2 = \begin{bmatrix} -1 \\ 2 \\ 0 \end{bmatrix}$ are linearly independent but not orthogonal.

Use Gram-Schmidt process:

$$\mathbf{w}_2 = \mathbf{v}_2 - \frac{\mathbf{v}_2 \cdot \mathbf{v}_1}{\mathbf{v}_1 \cdot \mathbf{v}_1}\mathbf{v}_1 = \begin{bmatrix} -1 \\ 2 \\ 0 \end{bmatrix} - \frac{-1}{2}\begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} -1/2 \\ 2 \\ 1/2 \end{bmatrix}; \quad \mathbf{w}_2' = \begin{bmatrix} -1 \\ 4 \\ 1 \end{bmatrix}$$

and then normalise the vectors for an orthonormal set:

$$\mathbf{u}_1 = \frac{\mathbf{v}_1}{\|\mathbf{v}_1\|} = \begin{bmatrix} 1/\sqrt{2} \\ 0 \\ 1/\sqrt{2} \end{bmatrix}, \qquad \mathbf{u}_2 = \frac{\mathbf{w}_2}{\|\mathbf{w}_2\|} = \begin{bmatrix} -1/\sqrt{18} \\ 4/\sqrt{18} \\ 1/\sqrt{18} \end{bmatrix}$$

The third eigenvector (for λ_3) is automatically orthogonal to this pair (why?) and we only need to normalise it:

$$\mathbf{v}_3 = \begin{bmatrix} -2 \\ -1 \\ 2 \end{bmatrix} \quad \Rightarrow \quad \mathbf{u}_3 = \frac{\mathbf{v}_3}{\|\mathbf{v}_3\|} = \begin{bmatrix} -2/3 \\ -1/3 \\ 2/3 \end{bmatrix}$$

So the orthonormal set of eigenvectors is

$$\mathbf{u}_1 = \begin{bmatrix} 1/\sqrt{2} \\ 0 \\ 1/\sqrt{2} \end{bmatrix}, \quad \mathbf{u}_2 = \begin{bmatrix} -1/\sqrt{18} \\ 4/\sqrt{18} \\ 1/\sqrt{18} \end{bmatrix}, \quad \mathbf{u}_3 = \begin{bmatrix} -2/3 \\ -1/3 \\ 2/3 \end{bmatrix}$$

In this way, we have obtained $\mathbf{P}$ and $\mathbf{D}$:

$$\mathbf{P} = \begin{bmatrix} 1/\sqrt{2} & -1/\sqrt{18} & -2/3 \\ 0 & 4/\sqrt{18} & -1/3 \\ 1/\sqrt{2} & 1/\sqrt{18} & 2/3 \end{bmatrix}, \qquad \mathbf{D} = \begin{bmatrix} 7 & 0 & 0 \\ 0 & 7 & 0 \\ 0 & 0 & -2 \end{bmatrix}$$

and $\mathbf{A} = \mathbf{PDP}^{-1} = \mathbf{PDP}^{\mathsf{T}}$.

Note: the order of eigenvalues in $\mathbf{D}$ can be different, but then the order of columns in $\mathbf{P}$ must be changed accordingly.

The set of eigenvalues of matrix $\mathbf{A}$ is called the *spectrum* of $\mathbf{A}$.

Theorem:

An $n \times n$ symmetric matrix $\mathbf{A}$ has the following properties:

- $\mathbf{A}$ has n real eigenvalues (if counting multiplicities);
- The dimension of the eigenspace for each eigenvalue λ equals the multiplicity of λ as a root of the characteristic equation;
- The eigenspaces are mutually orthogonal (the eigenvectors corresponding to different eigenvalues are orthogonal);
- $\mathbf{A}$ is orthogonally diagonalisable: $\mathbf{A} = \mathbf{PDP}^{-1} = \mathbf{PDP}^{\mathsf{T}}$.

For an orthogonally diagonalisable matrix:

$$\mathbf{A} = \mathbf{P}\mathbf{D}\mathbf{P}^{\mathsf{T}} = \begin{bmatrix}\mathbf{u}_1 & \dots & \mathbf{u}_n\end{bmatrix} \begin{bmatrix}\lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n\end{bmatrix} \begin{bmatrix}\mathbf{u}_1^{\mathsf{T}} \\ \vdots \\ \mathbf{u}_n^{\mathsf{T}}\end{bmatrix}$$

$$= \lambda_1 \mathbf{u}_1 \mathbf{u}_1^{\mathsf{T}} + \lambda_2 \mathbf{u}_2 \mathbf{u}_2^{\mathsf{T}} + \dots + \lambda_n \mathbf{u}_n \mathbf{u}_n^{\mathsf{T}}.$$

This is called a *spectral decomposition* of $\mathbf{A}$.

- Each decomposition term is an $n \times n$ matrix with rank 1.
- Each matrix $\mathbf{u}_j \mathbf{u}_j^{\mathsf{T}}$ is a projection matrix:
  for $\mathbf{x} \in \mathbb{R}^n$, vector $\mathbf{u}_j \mathbf{u}_j^{\mathsf{T}} \mathbf{x}$ is the orthogonal projection of $\mathbf{x}$ onto the subspace spanned by $\mathbf{u}_j$.

Example: Spectral decomposition of matrix $\mathbf{A} = \begin{bmatrix} 7 & 2 \\ 2 & 4 \end{bmatrix}$:

Eigenvalues:

$$(7-\lambda)(4-\lambda) - 4 = \lambda^2 - 11\lambda + 24 = 0 \quad \Rightarrow \quad \lambda_1 = 8, \quad \lambda_2 = 3$$

Eigenvectors:

$$\begin{bmatrix} -1 & 2 \\ 2 & -4 \end{bmatrix} \to \begin{bmatrix} 1 & -2 \\ 0 & 0 \end{bmatrix} \quad \Rightarrow \quad \mathbf{v}_1 = \begin{bmatrix} 2 \\ 1 \end{bmatrix} \quad \Rightarrow \quad \mathbf{u}_1 = \begin{bmatrix} 2/\sqrt{5} \\ 1/\sqrt{5} \end{bmatrix}$$

$$\begin{bmatrix} 4 & 2 \\ 2 & 1 \end{bmatrix} \to \begin{bmatrix} 2 & 1 \\ 0 & 0 \end{bmatrix} \quad \Rightarrow \quad \mathbf{v}_2 = \begin{bmatrix} -1 \\ 2 \end{bmatrix} \quad \Rightarrow \quad \mathbf{u}_2 = \begin{bmatrix} -1/\sqrt{5} \\ 2/\sqrt{5} \end{bmatrix}$$

So then $\mathbf{P} = \begin{bmatrix} \mathbf{u}_1 & \mathbf{u}_2 \end{bmatrix}$ and

$$\mathbf{A} = \mathbf{P}\mathbf{D}\mathbf{P}^\mathsf{T} = \begin{bmatrix} 2/\sqrt{5} & -1/\sqrt{5} \\ 1/\sqrt{5} & 2/\sqrt{5} \end{bmatrix} \begin{bmatrix} 8 & 0 \\ 0 & 3 \end{bmatrix} \begin{bmatrix} 2/\sqrt{5} & 1/\sqrt{5} \\ -1/\sqrt{5} & 2/\sqrt{5} \end{bmatrix}$$

$$\mathbf{A} = \begin{bmatrix} 7 & 2 \\ 2 & 4 \end{bmatrix} = \mathbf{PDP}^\mathsf{T} = \begin{bmatrix} 2/\sqrt{5} & -1/\sqrt{5} \\ 1/\sqrt{5} & 2/\sqrt{5} \end{bmatrix} \begin{bmatrix} 8 & 0 \\ 0 & 3 \end{bmatrix} \begin{bmatrix} 2/\sqrt{5} & 1/\sqrt{5} \\ -1/\sqrt{5} & 2/\sqrt{5} \end{bmatrix}$$

Thus: $\quad \mathbf{A} = 8\mathbf{u}_1\mathbf{u}_1^\mathsf{T} + 3\mathbf{u}_2\mathbf{u}_2^\mathsf{T} = 8 \begin{bmatrix} 4/5 & 2/5 \\ 2/5 & 1/5 \end{bmatrix} + 3 \begin{bmatrix} 1/5 & -2/5 \\ -2/5 & 4/5 \end{bmatrix}$

Verifying this decomposition:

$$\mathbf{u}_1\mathbf{u}_1^\mathsf{T} = \begin{bmatrix} 2/\sqrt{5} \\ 1/\sqrt{5} \end{bmatrix} \begin{bmatrix} 2/\sqrt{5} & 1/\sqrt{5} \end{bmatrix} = \begin{bmatrix} 4/5 & 2/5 \\ 2/5 & 1/5 \end{bmatrix}$$

$$\mathbf{u}_2\mathbf{u}_2^\mathsf{T} = \begin{bmatrix} -1/\sqrt{5} \\ 2/\sqrt{5} \end{bmatrix} \begin{bmatrix} -1/\sqrt{5} & 2/\sqrt{5} \end{bmatrix} = \begin{bmatrix} 1/5 & -2/5 \\ -2/5 & 4/5 \end{bmatrix}$$

$$8\mathbf{u}_1\mathbf{u}_1^\mathsf{T} + 3\mathbf{u}_2\mathbf{u}_2^\mathsf{T} = \begin{bmatrix} 32/5 & 16/5 \\ 16/5 & 8/5 \end{bmatrix} + \begin{bmatrix} 3/5 & -6/5 \\ -6/5 & 12/5 \end{bmatrix} = \begin{bmatrix} 7 & 2 \\ 2 & 4 \end{bmatrix}$$

- Orthogonal diagonalisation is a very useful tool however only symmetric matrices can be decomposed as $\mathbf{A} = \mathbf{P}\mathbf{D}\mathbf{P}^{\mathsf{T}}$.
- There is a more general decomposition possible for a non-square matrix $\mathbf{A}$.

Note: $\mathbf{A}^{\mathsf{T}}\mathbf{A}$ is symmetric and can be orthogonally diagonalised.

Let $\{\mathbf{u}_1, \ldots \mathbf{u}_n\}$ be unit eigenvectors of $\mathbf{A}^{\mathsf{T}}\mathbf{A}$,
and $\lambda_1, \ldots \lambda_n$ be the corresponding eigenvalues. Then

$$\|\mathbf{A}\mathbf{u}_i\|^2 = (\mathbf{A}\mathbf{u}_i)^{\mathsf{T}}\mathbf{A}\mathbf{u}_i = \mathbf{u}_i^{\mathsf{T}}\mathbf{A}^{\mathsf{T}}\mathbf{A}\mathbf{u}_i = \mathbf{u}_i^{\mathsf{T}}(\lambda_i\mathbf{u}_i) = \lambda_i(\mathbf{u}_i^{\mathsf{T}}\mathbf{u}_i) = \lambda_i$$

therefore all the eigenvalues of $\mathbf{A}^{\mathsf{T}}\mathbf{A}$ are non-negative.

We can always rearrange them so that $\lambda_1 \geqslant \lambda_2 \geqslant \ldots \geqslant \lambda_n \geqslant 0$.

Definition: The *singular values* of $\mathbf{A}$ are the square roots of the eigenvalues $\lambda_1, \ldots \lambda_n$ of $\mathbf{A}^\mathsf{T}\mathbf{A}$, arranged in the descending order:

$$\sigma_i = \sqrt{\lambda_i} \qquad \sigma_i \geqslant \sigma_{i+1}$$

SVD: for an $m \times n$ matrix $\mathbf{A}$ with rank r, construct:

- $m \times n$ matrix Σ with the first r diagonal entries being the singular values of $\mathbf{A}$: $\sigma_1 \geqslant \sigma_2 \geqslant \ldots \geqslant \sigma_r > 0$; then zeros
- $n \times n$ orthogonal matrix $\mathbf{U} = \begin{bmatrix}\mathbf{u}_1 & \ldots & \mathbf{u}_n\end{bmatrix}$ where $\mathbf{u}_i$ are the normalised eigenvectors of $\mathbf{A}^\mathsf{T}\mathbf{A}$
- $m \times m$ orthogonal matrix $\mathbf{W}$: $\mathbf{w}_i = \dfrac{\mathbf{A}\mathbf{u}_i}{\sigma_i}$ for $1 \leqslant i \leqslant r$, extended to an orthonormal basis of $\mathbb{R}^m$ for $r < i \leqslant m$

Then $\mathbf{A} = \mathbf{W}\Sigma\mathbf{U}^\mathsf{T}$ is a *singular value decomposition* of $\mathbf{A}$.

- The decomposition of $\mathbf{A}$ involves an $m \times n$ "quasi-diagonal" matrix $\Sigma = \begin{bmatrix} \mathbf{D} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}$, where $\mathbf{D}$ is an $r \times r$ diagonal matrix $(r \leqslant m \;\&\; r \leqslant n)$.

  The second "row" in Σ contains $m - r$ rows.

  The second "column" in Σ contains $n - r$ columns.

- The matrices $\mathbf{U}$ and $\mathbf{W}$ in $\mathbf{A} = \mathbf{W}\Sigma\mathbf{U}^\mathsf{T}$ are not uniquely defined by $\mathbf{A}$ but the diagonal entries in Σ are uniquely determined (by the singular values of $\mathbf{A}$).

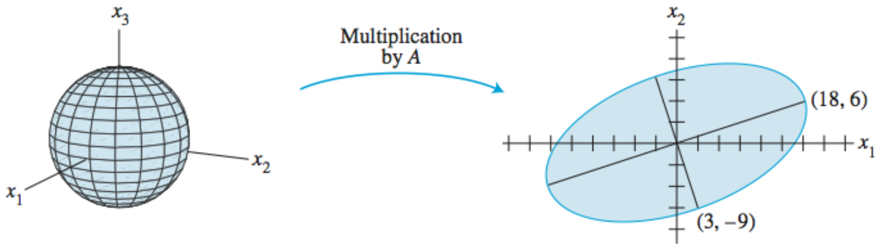
- The columns of $\mathbf{W}$ are called *left singular vectors* of $\mathbf{A}$ and the columns of $\mathbf{U}$ are called the *right singular vectors* of $\mathbf{A}$.

- The singular values of $\mathbf{A}$ are the lengths of $\mathbf{A}\mathbf{u}_i$ vectors.

Example: Construct a singular value decomposition of

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

Note: The transformation $\mathbf{x} \mapsto \mathbf{A}\mathbf{x}$ maps a unit sphere $\{\mathbf{x} : \|\mathbf{x}\| = 1\}$ in $\mathbb{R}^3$ onto an ellipse in $\mathbb{R}^2$.



Example: Construct a singular value decomposition of

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

Step 1: Construct an orthogonal diagonalisation of $\mathbf{A}^\mathsf{T}\mathbf{A}$.

$$\mathbf{A}^\mathsf{T}\mathbf{A} = \begin{bmatrix} 4 & 8 \\ 11 & 7 \\ 14 & -2 \end{bmatrix} \begin{bmatrix} 4 & 11 & 14 \\ 8 & 7 & -2 \end{bmatrix} = \begin{bmatrix} 80 & 100 & 40 \\ 100 & 170 & 140 \\ 40 & 140 & 200 \end{bmatrix}$$

The eigenvalues of this matrix are $\lambda_1 = 360$, $\lambda_2 = 90$, $\lambda_3 = 0$.
The corresponding unit eigenvectors are:

$$\mathbf{u}_1 = \begin{bmatrix} 1/3 \\ 2/3 \\ 2/3 \end{bmatrix}, \quad \mathbf{u}_2 = \begin{bmatrix} -2/3 \\ -1/3 \\ 2/3 \end{bmatrix}, \quad \mathbf{u}_3 = \begin{bmatrix} 2/3 \\ -2/3 \\ 1/3 \end{bmatrix}.$$

Then $$\mathbf{U} = \begin{bmatrix} \mathbf{u}_1 \ \mathbf{u}_2 \ \mathbf{u}_3 \end{bmatrix} = \begin{bmatrix} 1/3 & -2/3 & 2/3 \\ 2/3 & -1/3 & -2/3 \\ 2/3 & 2/3 & 1/3 \end{bmatrix}.$$

Step 2: Construct Σ using the singular values of $\mathbf{A}$.

Given the found eigenvalues of $\mathbf{A}$: $\lambda_1 = 360$, $\lambda_2 = 90$, $\lambda_3 = 0$ the singular values of $\mathbf{A}$ are:

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

The non-zero σ_i form the diagonal sub-matrix $\mathbf{D}$ within Σ:

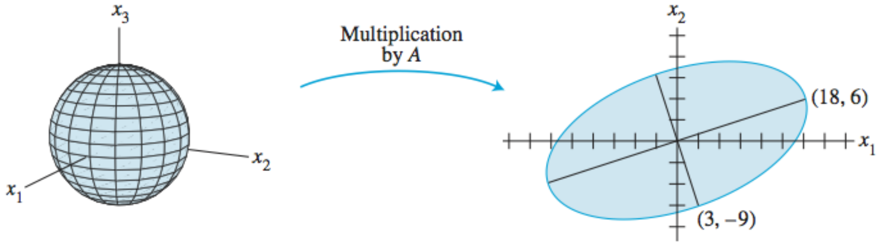
$$\mathbf{D} = \begin{bmatrix} 6\sqrt{10} & 0 \\ 0 & 3\sqrt{10} \end{bmatrix}$$

$$\Sigma = \begin{bmatrix} 6\sqrt{10} & 0 & 0 \\ 0 & 3\sqrt{10} & 0 \end{bmatrix}$$

The first singular value of $\mathbf{A}$ is the maximum of $\|\mathbf{Ax}\|$ for all $\mathbf{x}$ with $\|\mathbf{x}\| = 1$; this is obtained when $\mathbf{x} = \mathbf{u}_1$:

$$\mathbf{A}\mathbf{u}_1 = \begin{bmatrix} 4 & 11 & 14 \\ 8 & 7 & -2 \end{bmatrix} \begin{bmatrix} 1/3 \\ 2/3 \\ 2/3 \end{bmatrix} = \begin{bmatrix} 18 \\ 6 \end{bmatrix}.$$

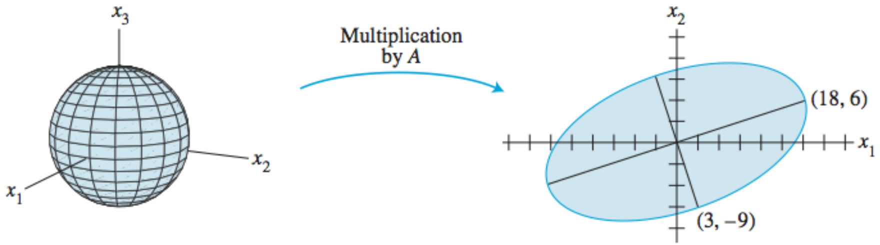
This is an ellipse point furthest from $\mathbf{0}$; the distance is $\sigma_1 = 6\sqrt{10}$:



The second singular value of $\mathbf{A}$ is the maximum of $\|\mathbf{Ax}\|$ over all unit vectors orthogonal to $\mathbf{u}_1$ and this is achieved at $\mathbf{x} = \mathbf{u}_2$:

$$\mathbf{A}\mathbf{u}_2 = \begin{bmatrix} 4 & 11 & 14 \\ 8 & 7 & -2 \end{bmatrix} \begin{bmatrix} -2/3 \\ -1/3 \\ 2/3 \end{bmatrix} = \begin{bmatrix} 3 \\ -9 \end{bmatrix}.$$

This is an ellipse point on the minor axis (distance $\sigma_2 = 3\sqrt{10}$):



Step 3: Construct $\mathbf{W}$. When $\mathbf{A}$ has rank r the first r columns of $\mathbf{W}$ are normalised vectors obtained from $\mathbf{Au}_1, \ldots \mathbf{Au}_r$.
Matrix $\mathbf{A}$ has two non-zero singular values so $\operatorname{rank}\mathbf{A} = 2$ and

$$\|\mathbf{Au}_1\| = \sigma_1, \quad \|\mathbf{Au}_2\| = \sigma_2.$$

Thus the columns of $\mathbf{W}$ are

$$\mathbf{w}_1 = \frac{\mathbf{Au}_1}{\sigma_1} = \frac{1}{6\sqrt{10}} \begin{bmatrix} 18 \\ 6 \end{bmatrix} = \begin{bmatrix} 3/\sqrt{10} \\ 1/\sqrt{10} \end{bmatrix}$$
$$\mathbf{w}_2 = \frac{\mathbf{Au}_2}{\sigma_2} = \frac{1}{3\sqrt{10}} \begin{bmatrix} 3 \\ -9 \end{bmatrix} = \begin{bmatrix} 1/\sqrt{10} \\ -3/\sqrt{10} \end{bmatrix}.$$

The set $\{\mathbf{w}_1, \mathbf{w}_2\}$ is already an orthonormal basis for $\mathbb{R}^2$, and so

$$\mathbf{W} = \begin{bmatrix} 3/\sqrt{10} & 1/\sqrt{10} \\ 1/\sqrt{10} & -3/\sqrt{10} \end{bmatrix}$$

.

Thus the singular value decomposition of

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

is $\mathbf{A} = \mathbf{W}\Sigma\mathbf{U}^{\mathsf{T}} =$

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

$\mathbf{A} = \mathbf{W}\Sigma\mathbf{U}^\mathsf{T}$ can be rewritten as

$$\mathbf{A} = \begin{bmatrix}\mathbf{w}_1 & \dots & \mathbf{w}_m\end{bmatrix} \begin{bmatrix} \sigma_1 & & 0 & 0 & \\ & \ddots & & 0 & \\ 0 & & \sigma_r & 0 & \dots \\ 0 & 0 & 0 & 0 & \\ & & \vdots & & \ddots \end{bmatrix} \begin{bmatrix}\mathbf{u}_1^\mathsf{T} \\ \vdots \\ \mathbf{u}_n^\mathsf{T}\end{bmatrix}$$

$$= \sigma_1 \mathbf{w}_1 \mathbf{u}_1^\mathsf{T} + \sigma_2 \mathbf{w}_2 \mathbf{u}_2^\mathsf{T} + \ldots + \sigma_r \mathbf{w}_r \mathbf{u}_r^\mathsf{T}.$$

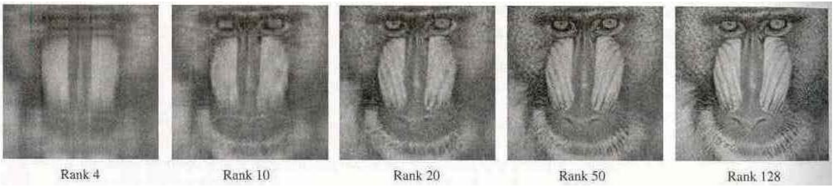
Original matrix $\mathbf{A}$ involves $m \times n$ values to be stored, whereas this expansion requires $(m \times r + n \times r + r) = r(m + n + 1)$.

Usually some of the singular values are very small so

$$\mathbf{A} \approx \mathbf{A}_k = \sigma_1 \mathbf{w}_1 \mathbf{u}_1^{\mathsf{T}} + \sigma_2 \mathbf{w}_2 \mathbf{u}_2^{\mathsf{T}} + \ldots + \sigma_k \mathbf{w}_k \mathbf{u}_k^{\mathsf{T}},$$

where $k < r$ is the *rank of approximation*; quite often $k \ll r$.

In that case, the storage size is reduced to $k(m + n + 1) \ll m \cdot n$.



In this way, for example, an SVD-based image compression works.

Definition: A *quadratic form* on $\mathbb{R}^n$ is a function Q defined as

$$Q(\mathbf{x}) = \mathbf{x}^\mathsf{T}\mathbf{A}\mathbf{x} \qquad : \qquad \mathbf{x} \in \mathbb{R}^n \qquad \mathbf{A} = \mathbf{A}^\mathsf{T}$$

where the $n \times n$ symmetric $\mathbf{A}$ is the *matrix of the quadratic form*.

Examples: (1) The simplest QF is: $\mathbf{x}^\mathsf{T}\,\mathbf{I}\,\mathbf{x} = \mathbf{x}^\mathsf{T}\mathbf{x} = \|\mathbf{x}\|^2$.

(2) Let $\mathbf{A} = \begin{bmatrix} 4 & 0 \\ 0 & 3 \end{bmatrix}$, $\quad \mathbf{B} = \begin{bmatrix} 3 & -2 \\ -2 & 7 \end{bmatrix}$, $\quad \mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$

$$\mathbf{x}^\mathsf{T}\mathbf{A}\mathbf{x} = \begin{bmatrix} x_1 & x_2 \end{bmatrix} \begin{bmatrix} 4 & 0 \\ 0 & 3 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} x_1 & x_2 \end{bmatrix} \begin{bmatrix} 4x_1 \\ 3x_2 \end{bmatrix} = 4x_1^2 + 3x_2^2$$

$$\mathbf{x}^\mathsf{T}\mathbf{B}\mathbf{x} = \begin{bmatrix} x_1 & x_2 \end{bmatrix} \begin{bmatrix} 3 & -2 \\ -2 & 7 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} x_1 & x_2 \end{bmatrix} \begin{bmatrix} 3x_1 - 2x_2 \\ -2x_1 + 7x_2 \end{bmatrix}$$

$$= x_1(3x_1 - 2x_2) + x_2(-2x_1 + 7x_2) = 3x_1^2 - 4x_1x_2 + 7x_2^2$$

Examples:

(3) $Q(\mathbf{x}) = 5x_1^2 - x_1x_2 + 3x_2^2 + 8x_2x_3 + 2x_3^2 \qquad \mathbf{x} \in \mathbb{R}^3$

Let us write this quadratic form as $\mathbf{x}^\intercal \mathbf{A}\mathbf{x}$:

The coefficients of x_1^2, x_2^2, x_3^2 provide the diagonal of $\mathbf{A}$.

Then, to make $\mathbf{A}$ symmetric we split the coefficients of x_ix_j between the i,j and j,i matrix elements:

$$Q(\mathbf{x}) = \mathbf{x}^\intercal \mathbf{A}\mathbf{x} = \begin{bmatrix} x_1 & x_2 & x_3 \end{bmatrix} \begin{bmatrix} 5 & -1/2 & 0 \\ -1/2 & 3 & 4 \\ 0 & 4 & 2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$$

Theorem (the principal axes theorem):

For a given quadratic form $\mathbf{x}^\top \mathbf{A}\mathbf{x}$ there is a change of coordinates (change of variable) $\mathbf{x} = \mathbf{P}\mathbf{y}$ that transforms it into a quadratic form $\mathbf{y}^\top \mathbf{D}\mathbf{y}$ with a diagonal matrix $\mathbf{D}$ (no cross-product terms).

Proof: $\mathbf{A} = \mathbf{A}^\mathsf{T}$ can be orthogonally diagonalised $\mathbf{A} = \mathbf{P}\mathbf{D}\mathbf{P}^\mathsf{T}$. Change variables as $\mathbf{x} = \mathbf{P}\mathbf{y}$, then $\mathbf{y} = \mathbf{P}^{-1}\mathbf{x} = \mathbf{P}^\mathsf{T}\mathbf{x}$. Then

$$\mathbf{x}^\top \mathbf{A}\mathbf{x} = (\mathbf{P}\mathbf{y})^\mathsf{T}\mathbf{A}(\mathbf{P}\mathbf{y}) = \mathbf{y}^\top \mathbf{P}^\mathsf{T}(\mathbf{P}\mathbf{D}\mathbf{P}^\mathsf{T})\mathbf{P}\mathbf{y} = \mathbf{y}^\top \mathbf{D}\mathbf{y}$$

and so the matrix in the quadratic form for $\mathbf{y}$ is diagonal.

Notes:

The columns of $\mathbf{P}$ are called the *principal axes* of the QF $\mathbf{x}^\top \mathbf{A}\mathbf{x}$.

Principal axes form an orthonormal basis for $\mathbb{R}^n$.

Example: $Q(\mathbf{x}) = x_1^2 - 8x_1x_2 - 5x_2^2 \quad \Leftrightarrow \quad \mathbf{A} = \begin{bmatrix} 1 & -4 \\ -4 & -5 \end{bmatrix}.$

Let us find the principal axes and eliminate the cross terms.

The eigenvalues are $\lambda_1 = 3$, $\lambda_2 = -7$, and the unit eigenvectors

$$\mathbf{u}_1 = \begin{bmatrix} 2/\sqrt{5} \\ -1/\sqrt{5} \end{bmatrix}, \qquad \mathbf{u}_2 = \begin{bmatrix} 1/\sqrt{5} \\ 2/\sqrt{5} \end{bmatrix}$$

are orthogonal because $\mathbf{A}$ is symmetric and $\lambda_1 \neq \lambda_2$.

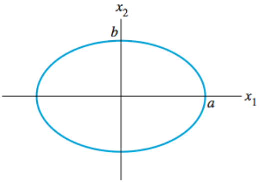
These vectors $\mathbf{u}_1$ and $\mathbf{u}_2$ are the principal axes of $Q(\mathbf{x})$.

$$\mathbf{P} = \begin{bmatrix} 2/\sqrt{5} & 1/\sqrt{5} \\ -1/\sqrt{5} & 2/\sqrt{5} \end{bmatrix}, \quad \mathbf{D} = \begin{bmatrix} 3 & 0 \\ 0 & -7 \end{bmatrix}.$$

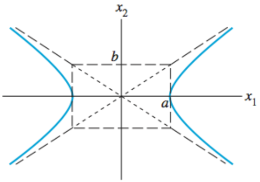
The change of variable is $\mathbf{y} = \mathbf{P}^{-1}\mathbf{x} = \mathbf{P}^{\mathsf{T}}\mathbf{x}$, and $\mathbf{A} = \mathbf{P}\mathbf{D}\mathbf{P}^{\mathsf{T}}$. So

$$x_1^2 - 8x_1x_2 - 5x_2^2 = \mathbf{x}^{\mathsf{T}}\mathbf{A}\mathbf{x} = \mathbf{y}^{\mathsf{T}}\mathbf{D}\mathbf{y} = 3y_1^2 - 7y_2^2$$

Example: $\mathbf{x}^{\mathsf{T}}\mathbf{A}\mathbf{x} = c \quad (\,\mathbf{x} \in \mathbb{R}^2,\ \ c \in \mathbb{R}\,)$ describes an ellipse, hyperbola, parabola, two lines, single point, or no point.

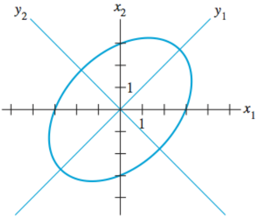


$\frac{x_1^2}{a^2} + \frac{x_2^2}{b^2} = 1,\ a > b > 0$

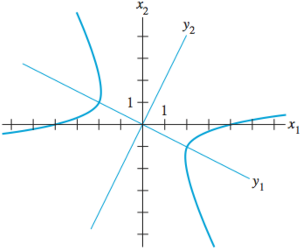


$\frac{x_1^2}{a^2} - \frac{x_2^2}{b^2} = 1,\ a > b > 0$

If $\mathbf{A}$ is diagonal then the graph is in the standard position.



(a) $5x_1^2 - 4x_1x_2 + 5x_2^2 = 48$

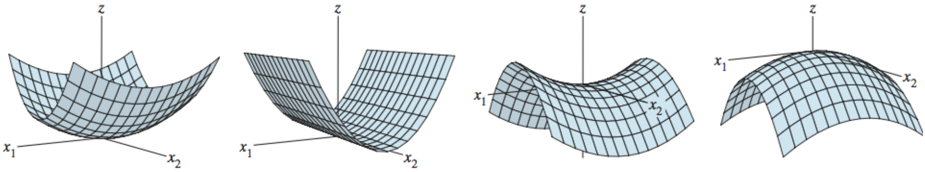


(b) $x_1^2 - 8x_1x_2 - 5x_2^2 = 16$

$$\mathbf{A} = \begin{bmatrix} 5 & -2 \\ -2 & 5 \end{bmatrix}$$

$$\mathbf{B} = \begin{bmatrix} 1 & -4 \\ -4 & -5 \end{bmatrix}$$

A few examples of $z = Q(\mathbf{x})$ for some typical cases ($\mathbf{x} \in \mathbb{R}^2$):



(a) $z = 3x_1^2 + 7x_2^2$ (b) $z = 3x_1^2$ (c) $z = 3x_1^2 - 7x_2^2$ (d) $z = -3x_1^2 - 7x_2^2$

Definition: A quadratic form Q is

(a) *positive definite*, if $Q(\mathbf{x}) > 0 \ \forall \, \mathbf{x} \neq 0$ all $\lambda > 0$

(b) *positive semidefinite*, if $Q(\mathbf{x}) \geqslant 0 \ \forall \, \mathbf{x}$

(c) *indefinite*, if $Q(\mathbf{x})$ takes positive and negative values

(d) *negative definite*, if $Q(\mathbf{x}) < 0 \ \forall \, \mathbf{x} \neq 0$ all $\lambda < 0$

(e) *negative semidefinite*, if $Q(\mathbf{x}) \leqslant 0 \ \forall \, \mathbf{x}$

This week: quick test 8

(fundamentals of orthogonality)

Next week:

Final class test (2 hours)

formally covers topics 8–10

but certainly implies the knowledge of previous topics