# Lecture 3

Math Foundations Team

BITS Pilani
Pilani | Dubai | Goa | Hyderabad

# Analytic geometry

- We have studied vector spaces in the previous lecture.
- Now we would like to provide some geometric interpretation to these concepts.
- We shall take a close look at geometric vectors and the concepts of lengths of vectors and angles between vectors.
- But first we need to add the concept of an inner product to our vector space.

# Norms

- A norm on a vector space is a function $\|.\| : V \to \mathbb{R}$, $\boldsymbol{x} \to \|\boldsymbol{x}\|$ which assigns to each vector $\boldsymbol{x}$ a length $\|\boldsymbol{x}\|$ such that for all $\lambda \in \mathbb{R}$ and $\boldsymbol{x}, \boldsymbol{y} \in V$ the following properties hold:
  - Absolutely homogeneous: $\|\lambda\boldsymbol{x}\| = |\lambda|\|\boldsymbol{x}\|$
  - Triangle inequality: $\|\boldsymbol{x}+\boldsymbol{y}\| \leq \|\boldsymbol{x}\| + \|\boldsymbol{y}\|$
  - Positive definite: $\|\boldsymbol{x}\| \geq 0$ and $\|\boldsymbol{x}\| = 0 \implies \boldsymbol{x} = 0$
- Manhattan norm : $\|\boldsymbol{x}\| = \sum_{i=1}^{i=n} |x_i|$
- Euclidean norm : $\|\boldsymbol{x}\|_2 = \sqrt{\sum_{i=1}^{i=n} x_i^2}$.

# Inner products

- Dot product in $\mathbb{R}^n$ is given by $\boldsymbol{x}^T\boldsymbol{y} = \sum_{i=1}^{i=n} x_i y_i$
- A bilinear mapping $\Omega$ is a mapping with two arguments and is linear in both arguments: Let $V$ be a vector space such that $\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z} \in V$, and let $\lambda, \psi \in \mathbb{R}$. Then we have $\Omega(\lambda\boldsymbol{x} + \psi\boldsymbol{y}, \boldsymbol{z}) = \lambda\Omega(\boldsymbol{x}, \boldsymbol{z}) + \psi\Omega(\boldsymbol{y}, \boldsymbol{z})$, and $\Omega(\boldsymbol{x}, \lambda\boldsymbol{y} + \psi\boldsymbol{z}) = \lambda\Omega(\boldsymbol{x}, \boldsymbol{y}) + \psi\Omega(\boldsymbol{x}, \boldsymbol{z})$.
- Let $V$ be a vector space and $\Omega : V \times V \to \mathbb{R}$ be a bilinear mapping that takes two vectors as arguments and returns a real number. Then $\Omega$ is called symmetric if $\Omega(\boldsymbol{x}, \boldsymbol{y}) = \Omega(\boldsymbol{y}, \boldsymbol{x})$. Also $\Omega$ is called positive-definite if $\forall x \in V \setminus \{0\},\ \ \Omega(\boldsymbol{x}, \boldsymbol{x}) > 0$ and $\Omega(\boldsymbol{0}, \boldsymbol{0}) = 0$.

# Inner products

- A positive-definite, symmetric bilinear mapping $\Omega : V \times V \to \mathbb{R}$ is called an inner product. To denote an inner product on $V$ we generally write $\langle \boldsymbol{x}, \boldsymbol{y} \rangle$.
- The pair $(V, \langle ., . \rangle)$ is called an inner product space.
- Next we introduce the concept of symmetric, positive-definite matrices and show we can express an inner product using such matrices.
- We recall that in a vector space $V$ any vector $\boldsymbol{x}$ can be written as linear combination of the basis vectors. We use this to express an inner product in terms of a matrix.

# Symmetric, positive-definite matrices

**Theorem**: For a real-valued, finite-dimensional vector space $V$ and an ordered basis $B$ of $V$, it holds that $\langle ., . \rangle : V \times V \to \mathbb{R}$ is an inner product if and only if there exists a symmetric, positive definite matrix $\boldsymbol{A} \in \mathbb{R}^{n \times n}$ with $\langle \boldsymbol{x}, \boldsymbol{y} \rangle = \hat{\boldsymbol{x}}^T \boldsymbol{A} \hat{\boldsymbol{y}}$.

**Proof**: One direction $\rightarrow$: $\langle ., . \rangle$ is an inner product $\implies$ $\boldsymbol{A}$ is symmetric, positive-definite such that $\langle \boldsymbol{x}, \boldsymbol{y} \rangle = \hat{\boldsymbol{x}}^T \boldsymbol{A} \hat{\boldsymbol{y}}$.

Other direction $\leftarrow$: $\boldsymbol{A}$ is symmetric, positive definite such that the operation $\langle \boldsymbol{x}, \boldsymbol{y} \rangle$ is defined as $\langle \boldsymbol{x}, \boldsymbol{y} \rangle = \hat{\boldsymbol{x}}^T \boldsymbol{A} \hat{\boldsymbol{y}}$ $\implies$ the operation defined is an inner product.

# Symmetric, positive-definite matrices

- We prove the $\rightarrow$ direction.
- Let $\langle \boldsymbol{x}, \boldsymbol{y} \rangle$ be the inner product between the vectors $\boldsymbol{x}, \boldsymbol{y}$ in $V$. We can write $\boldsymbol{x}$ in terms of say $n$ basis vectors as $\boldsymbol{x} = \sum_{i=1}^{i=n} \psi_i \boldsymbol{b}_i$. Similarly $\boldsymbol{y} = \sum_{i=1}^{i=n} \lambda_i \boldsymbol{b}_i$.
- Since the inner product is bilinear we can write $\langle \boldsymbol{x}, \boldsymbol{y} \rangle = \langle \sum_{i=1}^{i=n} \psi_i \boldsymbol{b}_i, \sum_{i=1}^{i=n} \lambda_i \boldsymbol{b}_i \rangle = \sum_{i=1}^{i=n} \sum_{j=1}^{j=n} \psi_i \langle \boldsymbol{b}_i, \boldsymbol{b}_j \rangle \lambda_j = \hat{\boldsymbol{x}}^T \boldsymbol{A} \hat{\boldsymbol{y}}$ where $A_{ij} = \langle \boldsymbol{b}_i, \boldsymbol{b}_j \rangle$.
- Here $\hat{\boldsymbol{x}}, \hat{\boldsymbol{y}}$ are vectors which represent the coordinates of the original vectors $\boldsymbol{x}, \boldsymbol{y}$ with respect to the basis vectors.

# Symmetric, positive-definite matrices

- This means that the inner product is entirely determined through the matrix $\boldsymbol{A}$. The symmetry of the inner product means that $A_{ij} = \langle \boldsymbol{b}_i, \boldsymbol{b}_j \rangle = A_{ji} = \langle \boldsymbol{b}_j, \boldsymbol{b}_i \rangle$. Thus $\boldsymbol{A}$ is symmetric.
- The positive-definiteness of the inner product means that $\forall \boldsymbol{x} \in V \backslash \{0\}, \boldsymbol{x}^T \boldsymbol{A} \boldsymbol{x} > 0$.

# Symmetric, positive-definite matrices

- Now let us consider an operation op such that $\boldsymbol{x} op \boldsymbol{y} = \hat{\boldsymbol{x}}^T A \hat{\boldsymbol{y}}$ where $\boldsymbol{A}$ is a symmetric, positive definite matrix.
- We shall show that "op" is an inner product by showing that it has all the properties of an inner product:
  - "op" has symmetry because $\boldsymbol{x} op \boldsymbol{y} = \hat{\boldsymbol{x}}^T A \hat{\boldsymbol{y}}$ and $\boldsymbol{y} op \boldsymbol{x} = \hat{\boldsymbol{y}}^T A \hat{\boldsymbol{x}} = \hat{\boldsymbol{y}}^T (A \hat{\boldsymbol{x}})$. By a property of the dot product we can write $\hat{\boldsymbol{y}}^T (A \hat{\boldsymbol{x}}) = (A \hat{\boldsymbol{x}})^T \hat{\boldsymbol{y}} = \hat{\boldsymbol{x}}^T \boldsymbol{A}^T \hat{\boldsymbol{y}} = \hat{\boldsymbol{x}}^T \boldsymbol{A} \hat{\boldsymbol{y}}$ where the last equality in the chain is possible since $\boldsymbol{A}$ is symmetric.
  - "op" also has bilinearity since we see that for $r \in R, (r\boldsymbol{x}) op \boldsymbol{y} = (r\hat{\boldsymbol{x}})^T \boldsymbol{A} \hat{\boldsymbol{y}} = r\hat{\boldsymbol{x}}^T \boldsymbol{A} \hat{\boldsymbol{y}} = r\boldsymbol{x} op \boldsymbol{y}$.
  - $(\boldsymbol{x} + \boldsymbol{y}) op \boldsymbol{z} = (\hat{\boldsymbol{x}} + \hat{\boldsymbol{y}})^T \boldsymbol{A} \hat{\boldsymbol{z}} = \hat{\boldsymbol{x}}^T \boldsymbol{A} \hat{\boldsymbol{z}} + \hat{\boldsymbol{y}}^T \boldsymbol{A} \hat{\boldsymbol{z}} = \boldsymbol{x} op \boldsymbol{z} + \boldsymbol{y} op \boldsymbol{z}$.
  - Finally if $\boldsymbol{x}$ is a non-zero vector then $\hat{\boldsymbol{x}}$ is also a non-zero vector, $\boldsymbol{x} op \boldsymbol{x} = \hat{\boldsymbol{x}}^T \boldsymbol{A} \hat{\boldsymbol{x}} > 0$ since we are given that $\boldsymbol{A}$ is positive-definite.

# Symmetric, positive definite matrices

- Can a symmetric, positive-definite matrix have less than full rank? We have $\boldsymbol{x}^T\boldsymbol{A}\boldsymbol{x} > 0$ for all non-zero $\boldsymbol{x}$. Thus $\boldsymbol{x} = \boldsymbol{0}$ is the only vector allowed in the nullspace. The nullspace is 0-dimensional so $\boldsymbol{A}$ has full rank.
- What can be said about the diagonal elements of a positive-definite matrix? From $(\boldsymbol{e}_i)^T\boldsymbol{A}\boldsymbol{e}_i > 0$ where $\boldsymbol{e}_i$ is the $i$th canonical basis vector, we see that $A_{ii} > 0$. Thus the diagonal entries are all strictly positive.

# Lengths and distances

- Inner products and norms are closely related in the sense that any inner product induces a norm: $\|\boldsymbol{x}\| = \sqrt{\langle \boldsymbol{x}, \boldsymbol{x} \rangle}$
- Not every norm is induced by an inner product, for example the Manhattan norm.
- For an inner product vector space $(V, \langle ., . \rangle)$, the induced norm $\|.\|$ satisfies the Cauchy-Schwarz inequality: $|\langle \boldsymbol{x}, \boldsymbol{y} \rangle| \leq \|\boldsymbol{x}\| \|\boldsymbol{y}\|$. Why is this true?

# Cauchy-Schwarz inequality

- Let $\boldsymbol{u}$ and $\boldsymbol{v}$ be two vectors and let us consider the length of the vector $\boldsymbol{u} - \alpha \boldsymbol{v}$ where $\alpha$ is a constant.
- The length of the vector $\boldsymbol{u} - \alpha \boldsymbol{v}$ is greater than or equal to zero. The length of the vector $\boldsymbol{u} - \alpha \boldsymbol{v}$ is $\|\boldsymbol{u} - \alpha \boldsymbol{v}\|^2 = \langle \boldsymbol{u} - \alpha \boldsymbol{v}, \boldsymbol{u} - \alpha \boldsymbol{v} \rangle = (\boldsymbol{u} - \alpha \boldsymbol{v})^T (\boldsymbol{u} - \alpha \boldsymbol{v})$.
- We can expand the dot product $(\boldsymbol{u} - \alpha \boldsymbol{v})^T (\boldsymbol{u} - \alpha \boldsymbol{v}) = \boldsymbol{u}^T \boldsymbol{u} - \alpha \boldsymbol{u}^T \boldsymbol{v} - \alpha \boldsymbol{v}^T \boldsymbol{u} + \alpha^2 \boldsymbol{v}^T \boldsymbol{v} \geq 0$
- Now set $\alpha = \frac{\boldsymbol{u}^T \boldsymbol{v}}{\boldsymbol{v}^T \boldsymbol{v}}$ to get $\boldsymbol{u}^T \boldsymbol{u} - \frac{(\boldsymbol{u}^T \boldsymbol{v})^2}{\boldsymbol{v}^T \boldsymbol{v}} \geq 0$ which leads us to $(\boldsymbol{u}^T \boldsymbol{u})(\boldsymbol{v}^T \boldsymbol{v}) \geq (\boldsymbol{u}^T \boldsymbol{v})^2$ which is Cauchy-Schwarz inequality.
- Note that although this proof was developed using $\langle \boldsymbol{u}, \boldsymbol{v} \rangle = \boldsymbol{u}^T \boldsymbol{v}$, it works for any definition of the inner product.

# Metric space

- Consider an inner product space $(V, \langle ., . \rangle)$. Define $d(\boldsymbol{x}, \boldsymbol{y})$ the distance between two vectors $\boldsymbol{x}$ and $\boldsymbol{y}$ to be $d(\boldsymbol{x}, \boldsymbol{y}) = \|\boldsymbol{x} - \boldsymbol{y}\| = \sqrt{\langle \boldsymbol{x} - \boldsymbol{y}, \boldsymbol{x} - \boldsymbol{y} \rangle}$.
- If we use the dot product as the inner product, then the distance is called the Euclidean distance.
- The mapping $d : V \times V \to \mathbb{R}$ is called a metric.

# Properties of a metric space

A metric $d$ has the following properties:

- $d$ is positive-definite which means $d(\boldsymbol{x}, \boldsymbol{y}) \geq 0 \;\; \forall \boldsymbol{x}, \boldsymbol{y} \in V$. $d(\boldsymbol{x}, \boldsymbol{y}) = 0 \implies \boldsymbol{x} = \boldsymbol{y}$.
- $d$ is symmetric which means $d(\boldsymbol{x}, \boldsymbol{y}) = d(\boldsymbol{y}, \boldsymbol{x}) \;\; \forall \boldsymbol{x}, \boldsymbol{y} \in V$.
- $d$ obeys the triangle inequality as follows: $d(\boldsymbol{x}, \boldsymbol{z}) \leq d(\boldsymbol{x}, \boldsymbol{y}) + d(\boldsymbol{y}, \boldsymbol{z}) \;\; \forall \boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z} \in V$

Inner products and metrics seem to be very similar in terms of their properties - however there is one important difference. When $\boldsymbol{x}$ and $\boldsymbol{y}$ are close to each other the inner product is large but the distance metric is small. On the other hand when $\boldsymbol{x}$ and $\boldsymbol{y}$ are far apart, then the inner product is small but the distance metric is large.

# Angles and orthogonality

- In addition to being able to capture the lengths of vectors and the distance between vectors, inner products can also capture the **angle** $\omega$ between two vectors and can thus capture the geometry of a vector space.
- The key to using the inner product to characterize the angle between two vectors is the Cauchy-Schwarz inequality.
- Assume that $\boldsymbol{x}$ and $\boldsymbol{y}$ are not the $\mathbf{0}$ vector. Then the Cauchy-Schwarz inequality tells us that

$$-1 \leq \frac{\boldsymbol{x}^T \boldsymbol{y}}{\|\boldsymbol{x}\| \|\boldsymbol{y}\|} \leq 1 \tag{1}$$

# Angles and orthogonality

- Since the Cauchy-Schwarz ratio lies between -1 and 1 we can set it equal to the cosine of a unique angle $\omega \in [0, \pi]$ such that

$$\cos(\omega) = \frac{\langle \boldsymbol{x}, \boldsymbol{y} \rangle}{\|\boldsymbol{x}\| \|\boldsymbol{y}\|} \tag{2}$$

- The angle $\omega$ is the angle between two vectors. What does it capture?
- The notion of angle captures similarity of orientation between two vectors. When the dot product is close to zero, the vectors are more or less pointing in orthogonal directions and $\omega \approx \pi/2$.

# Angles and orthogonality

- Food for thought: Suppose we choose vectors $\boldsymbol{x}$ and $\boldsymbol{y}$ uniformly at random in high dimensions. What happens to the dot product between the vectors and hence the angle between them?
- To choose a vector uniformly at random over a sphere let every component in the vector be an independent Gaussian random variable of mean 0 and unit variance.
- Write a small program to see what happens ...

# Angles and orthogonality

- A key feature of the inner product is that we can use it to characterize vectors that are orthogonal.
- Two vectors $\boldsymbol{x}$ and $\boldsymbol{y}$ are orthogonal if and only if the inner product between them is 0. For an orthogonal pair of vectors $\boldsymbol{x}$, $\boldsymbol{y}$ we can write $\boldsymbol{x} \perp \boldsymbol{y}$.
- By the above definition the $\mathbf{0}$-vector is orthogonal to all vectors.
- Vectors which are orthogonal with respect to one inner product need not be orthogonal with respect to another inner product.

# Example - angles and orthogonality

- Consider the vectors $\boldsymbol{x} = [1, 1]^T$ and $\boldsymbol{y} = [-1, 1]^T$
- With respect to the inner product defined as a dot product we see that $\langle \boldsymbol{x}, \boldsymbol{y} \rangle = \boldsymbol{x}^T \boldsymbol{y} = 1 * -1 + 1 * 1 = 0$.
- With respect to the inner product $\boldsymbol{x}^T \begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix} \boldsymbol{y}$ , the angle between the two vectors $\boldsymbol{x}$ and $\boldsymbol{y}$ becomes

$$\cos(\omega) = \frac{\langle \boldsymbol{x}, \boldsymbol{y} \rangle}{\|\boldsymbol{x}\| \|\boldsymbol{y}\|}$$

# Example - angles and orthogonality

- Continuing with our example we have

$$
\begin{aligned}
\cos(\omega) &= \frac{\boldsymbol{x}^T\boldsymbol{A}\boldsymbol{y}}{\sqrt{\boldsymbol{x}^T\boldsymbol{A}\boldsymbol{x}}\ \sqrt{\boldsymbol{y}^T\boldsymbol{A}\boldsymbol{y}}} \\
&= \frac{2x_1y_1 + x_2y_2}{\sqrt{(2x_1^2 + x_2^2)(2y_1^2 + y_2^2)}} \\
&= \frac{-1}{3}
\end{aligned}
$$

where $\boldsymbol{A} = \begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix}$.

- Thus, with respect to the new definition of inner product the vectors $\boldsymbol{x}$ and $\boldsymbol{y}$ are no longer orthogonal.

# Orthonormal matrix

- A square matrix $\boldsymbol{A} \in \mathbb{R}^{n \times n}$ is an orthogonal matrix if and only if its columns are orthonormal:

$$\boldsymbol{A}^T \boldsymbol{A} = \boldsymbol{I} = \boldsymbol{A}\boldsymbol{A}^T$$
$$\boldsymbol{A}^T = \boldsymbol{A}^{-1}$$

- If the columns of a matrix are orthonormal, why are its rows orthonormal too? This follows from the fact that the left-inverse of a square matrix is the same as the right-inverse. Let $\boldsymbol{A}$ be a square matrix with $\boldsymbol{B}$ and $\boldsymbol{C}$ the left and right inverses of $\boldsymbol{A}$: $\boldsymbol{BA} = I = \boldsymbol{AC} \implies \boldsymbol{B} = \boldsymbol{C}$. Why is this true?

# Orthonormal matrix

- Transformations by an orthonormal matrix preserve lengths. This can be seen as follows, using the dot product as the definition of the inner product:
  $\|\boldsymbol{Ax}\|^2 = (\boldsymbol{Ax})^T\boldsymbol{Ax} = \boldsymbol{x}^T\boldsymbol{A}^T\boldsymbol{Ax} = \boldsymbol{x}^T\boldsymbol{I}\boldsymbol{x} = \boldsymbol{x}^T\boldsymbol{x}$.
- An example of an orthonormal matrix is the 2D-rotation matrix which can be expressed as $\begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$ where $\theta$ is the angle of rotation.

# Orthonormal matrix

- Also the angle between two vectors $\boldsymbol{x}$ and $\boldsymbol{y}$ does not change after transformation by an orthonormal matrix. This can be seen as follows:

$$\begin{aligned}\cos(\omega) &= \frac{(\boldsymbol{Ax})^T\boldsymbol{Ay}}{\|\boldsymbol{Ax}\|\ \|\boldsymbol{Ay}\|}\\ &= \frac{\boldsymbol{x}^T\boldsymbol{A}^T\boldsymbol{Ay}}{\|\boldsymbol{x}\|\ \|\boldsymbol{y}\|}\\ &= \frac{\boldsymbol{x}^T\boldsymbol{y}}{\|\boldsymbol{x}\|\ \|\boldsymbol{y}\|}\end{aligned}$$

# Orthonormal basis

- We already looked at the concept of a basis of a vector space, and found that for the vector space $\mathbb{R}^n$ we need $n$ basis vectors.
- Our basis vectors needed only to be linearly independent - we can ensure linear independence by ensuring that our basis vectors point in different directions, so that a linear combination of $n-1$ basis vectors cannot cancel out the $n$th basis vector.
- Now we will look at a special case of a basis where the vectors are all mutually orthogonal in the sense of the inner product, and each vector is of unit length. We call such a basis an orthonormal basis.

# Orthonormal basis

- Question: Can you immediately think of an orthonormal basis for $\mathbb{R}^n$? Is an orthonormal basis for a vector space unique?
- Formal definition of an orthonormal basis: Consider an $n$-dimensional vector space $V$ and $n$ basis vectors $\{\boldsymbol{b}_1, \boldsymbol{b}_2, \dots \boldsymbol{b}_n\}$. If it is true that $\forall i, j = 1, \dots n, i \neq j \ \langle \boldsymbol{b}_i, \boldsymbol{b}_j \rangle = 0$ and $\langle \boldsymbol{b}_i, \boldsymbol{b}_i \rangle = 1$, then the basis is called an orthonormal basis.
- If the basis vectors are only mutually orthogonal but not of length unity, then we have an orthogonal basis.

# Gram-Schmidt process

- Given a set of basis vectors for a vector space, can we convert the given basis into an orthogonal basis? Yes, we shall use Gaussian elimination to construct such a basis.
- Let us start with an example: Consider $\mathbb{R}^2$ and two basis vectors $\boldsymbol{v}_1 = (3,1)^T$ and $\boldsymbol{v}_2 = (2,2)^T$. Put these vectors into columns of a matrix $\boldsymbol{A}$ such that $\boldsymbol{A} = \begin{bmatrix} 3 & 2 \\ 1 & 2 \end{bmatrix}$.
- The next step is to perform Gaussian elimination on the following augmented matrix: $[\boldsymbol{A}^T\boldsymbol{A}|\boldsymbol{A}^T] = \left[\begin{array}{cc|cc} 10 & 8 & 3 & 1 \\ 8 & 8 & 2 & 2 \end{array}\right]$
- On performing Gaussian elimination of this augmented matrix we end up with $\left[\begin{array}{cc|cc} 1 & 0.8 & 0.3 & 0.1 \\ 0 & 1 & -0.25 & 0.75 \end{array}\right]$

# Gram-Schmidt process

- Note that after the completion of Gaussian elimination the two rows on the right hand side are orthogonal. They form a basis for $\mathbb{R}^2$. We can normalize the vectors to get an orthonormal basis. Let us justify this technique.
- First we see that when the $m \times n$ matrix $\boldsymbol{A}$ has full column rank, then the matrix $\boldsymbol{A}^T\boldsymbol{A}$ is positive definite. To see this note that any solution $\boldsymbol{x}$ to $\boldsymbol{Ax} = \mathbf{0}$ is also a solution to $\boldsymbol{A}^T\boldsymbol{Ax} = \mathbf{0}$ and vice-versa. Why is this the case?
- When $\boldsymbol{A}$ has linearly independent columns, there are no non-trivial solutions to $\boldsymbol{Ax} = \mathbf{0}$. Thus the fact that there are no non-trivial solutions to $\boldsymbol{Ax} = \mathbf{0}$ means that $\forall \boldsymbol{x} \in \mathbb{R}^n, \boldsymbol{x} \neq \mathbf{0}, \boldsymbol{x}^T\boldsymbol{A}^T\boldsymbol{Ax} > \mathbf{0}$. Note that since $\boldsymbol{A}^T\boldsymbol{A}$ is positive-definite, Gaussian elimination can be carried out on $\boldsymbol{A}^T\boldsymbol{A}$ without row exchanges.

# Elementary transformations

- One of the steps of Gaussian elimination is the the subtraction of a multiple of a given row from a row below it. This step can be achieved by pre-multiplication of the given matrix by an elementary matrix. An elementary matrix is like an identity matrix except that one of the entries below the diagonal is allowed to be non-zero.
- To show how the process of elimination works using an elementary matrix consider the matrix $A = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}$ and assume that we want to subtract two times the first row from the second row.

# Elementary transformations

This can be accomplished by the following elementary matrix

$\boldsymbol{E} = \begin{bmatrix} 1 & 0 & 0 \\ -2 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$ so that the product

$$
\begin{aligned}
EA &= \begin{bmatrix} 1 & 0 & 0 \\ -2 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \\
&= \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} - 2a_{11} & a_{22} - 2a_{12} & a_{23} - 2a_{13} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}
\end{aligned}
$$

# Product of elementary transformations

- A series of Gaussian elimination steps can be represented as a product of elementary transformations acting on $\boldsymbol{A}$: $\boldsymbol{E}_m \boldsymbol{E}_{m-1} \ldots \boldsymbol{E}_1 \boldsymbol{A}$.
- The product of lower triangular matrices can be seen to be lower triangular, and the inverse of a lower triangular matrix can also be seen as a lower triangular matrix.
- Thus the action of Gaussian elimination operations can be seen in the following terms $\boldsymbol{L}^{-1}\boldsymbol{A} = \boldsymbol{U}$ where the product of the elementary transformations is represented as the inverse of a lower triangular matrix for notational convenience, and the right hand side $U$ is an upper triangular matrix. Thus we have $\boldsymbol{A} = \boldsymbol{L}\boldsymbol{U}$.

# Final argument

- Returning to our problem we are performing Gaussian elimination on the matrix $\boldsymbol{A}^T\boldsymbol{A}$ where $\boldsymbol{A}$ contains the basis vectors as its columns. Upon Gaussian elimination on the augmented matrix we reduce $[\boldsymbol{A}^T\boldsymbol{A}|\boldsymbol{A}^T]$ to get $[\boldsymbol{U}|\boldsymbol{L}^{-1}\boldsymbol{A}^T$ where $\boldsymbol{A}^T\boldsymbol{A} = \boldsymbol{LU}$.
- Now we shall show that $\boldsymbol{Q}^T = \boldsymbol{L}^{-1}\boldsymbol{A}^T$ is an orthogonal matrix whose rows are orthogonal.
- Consider $\boldsymbol{Q}^T\boldsymbol{Q} = \boldsymbol{L}^{-1}\boldsymbol{A}^T\boldsymbol{A}(\boldsymbol{L}^{-1})^T = \boldsymbol{U}(\boldsymbol{L}^{-1})^T =$ some upper triangular matrix
- But $\boldsymbol{Q}^T\boldsymbol{Q}$ is a symmetric matrix and can only be upper triangular if it is diagonal. Therefore $\boldsymbol{Q}$ is an orthogonal matrix whose columns are orthogonal. They can be normalized to obtain an orthonormal basis.