Applied Linear Algebra Refresher Course

Karianne Bergen
kbergen@stanford.edu

Institute for Computational and Mathematical Engineering,
Stanford University

September 16-19, 2013

Course Topics

- Matrix and vector properties
- Matrix algebra
- Linear transformations
- Solving linear systems
- Orthogonality
- Eigenvectors and eigenvalues
- Matrix decompositions

Vectors and Matrices

Scalars, vectors and matrices *

- *Scalar*: a single quantity or measurement that is invariant under coordinate rotations (e.g. mass, volume, temperature).

$$\alpha,\ \beta,\ \gamma \in \mathbb{R} \quad \text{(Greek letters)}$$

- *Vector*: an ordered collection of scalars (e.g. displacement, acceleration, force).

$$x = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}, \quad x_i \in \mathbb{R} \quad \forall i = 1, \ldots n$$

$$x,\ y,\ z \in \mathbb{R}^n \quad \text{(lower case) is a vector with } n \text{ entries.}$$

All vectors are column vectors unless otherwise noted.
Row vectors are denoted by $x^T = [\, x_1\ x_2\ \cdots\ x_n \,]$.

* For simplicity, we will restrict ourselves to real numbers. The generalization to the complex case can be found in any linear algebra text.

... Scalars, vectors and matrices

- *Matrix*: a two-dimensional collection of scalars.

$$A = \begin{bmatrix} a_1 & a_2 & \cdots & a_n \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}, \quad A^T = \begin{bmatrix} a_1^T \\ a_2^T \\ \\ a_m^T \end{bmatrix}$$

$A \in \mathbb{R}^{m \times n}$ is a matrix with m rows and n columns.

$$A,\ B,\ \Sigma,\ \Lambda \in \mathbb{R}^{m \times n} \quad \text{(upper case)}$$

Operations on vectors

- *Scalar multiplication:*

$$(\alpha x)_i = \alpha x_i, \qquad \alpha x = \alpha \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} \alpha x_1 \\ \alpha x_2 \\ \vdots \\ \alpha x_n \end{bmatrix}$$

- *Vector addition/subtraction:*

$$(x + y)_i = x_i + y_i, \qquad x + y = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} + \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} x_1 + y_1 \\ x_2 + y_2 \\ \vdots \\ x_n + y_n \end{bmatrix}$$

- *Inner Product (dot product, scalar product):*

$$x^T y \;=\; \sum_{i=1}^{n} x_i y_i \;=\; x_1 y_1 + x_2 y_2 + \cdots + x_n y_n$$

Other common notation for the inner product: $\langle x, y \rangle$, $\quad x \cdot y$

Vector norms

How do we measure the length or size of a vector?

Vector norm is a function $f : \mathbb{R}^n \to \mathbb{R}_+$ that satisfies the following properties:

1. Scale invariance: $f(\alpha v) = |\alpha| f(v), \quad \forall \alpha \in \mathbb{R}$ and $v \in \mathbb{R}^n$
2. Triangle inequality: $f(u + v) \leq f(u) + f(v), \quad \forall u,\ v \in \mathbb{R}^n$
3. Positivity: $f(x) = 0 \iff x \equiv 0$

Commonly used vector norms

- *1-norm ("taxi-cab" norm)*:

$$\|v\|_1 = \sum_{i=1}^{n} |v_i|$$

- *2-norm (Euclidean norm)*:

$$\|v\|_2 = \left(\sum_{i=1}^{n} v_i^2\right)^{\frac{1}{2}} = \sqrt{v^T v}$$

- *infinity-norm (max norm)*:

$$\|v\|_\infty = \max_i |v_i|$$

- These are examples of the *p-norm*, for $p = 1$, 2, and ∞:

$$\|v\|_p = \left(\sum_{i=1}^{n} |v_i|^p\right)^{\frac{1}{p}}, \quad p \geq 1$$

Questions?

Mini-Quiz 1:

Compute the 1-norm, 2-norm, infinity-norm of the following vector:

$$v = \begin{bmatrix} 3 \\ 6 \\ 2 \end{bmatrix}$$

Solution (Q1):

Solutions:

- 1-norm:

$$\|v\|_1 = \sum_{i=1}^{3} v_i = 3 + 6 + 2 = 11$$

- 2-norm:

$$\|v\|_2 = \left(\sum_{i=1}^{3} v_i^2 \right)^{\frac{1}{2}} = \sqrt{3^2 + 6^2 + 2^2} = \sqrt{9 + 36 + 4} = \sqrt{49} = 7$$

- ∞-norm

$$\|v\|_\infty = \max_i |v_i| = \max\{3, 6, 2\} = 6$$

Cauchy-Schwartz Inequality

- The inner product satisfies the *Cauchy-Schwartz inequality*:

  $$|x^T y| \leq \|x\|_2 \cdot \|y\|_2 \quad \text{for } x, y \in \mathbb{R}^n$$

  with equality when y is a scalar multiple of x.

  This inequality is arises in many applications, including
  - Geometry: triangle inequality
  - Physics: Heisenberg uncertainty principle
  - Statistics: Cramer-Rao lower bound

Inner Product and Orthogonality

- *Orthogonal vectors*: vectors u and v are orthogonal ($u \perp v$) if their inner product is zero.

$$u^T v = 0 \quad \Longleftrightarrow \quad u \perp v$$

- *Orthonormal vectors*: vectors u and v are orthonormal if they are orthogonal and both are normalized to length 1.

$$u \perp v \quad \text{and} \quad \|u\| = 1, \; \|v\| = 1$$

- The *orthogonal complement* of a set of vectors V, denoted $V^{\perp}$, is the set of vectors x such that $x^T v = 0$, $\forall v \in V$.

Inner Product and Projections

- *Projection*: the inner product gives us a way to compute the component of a vector v along any direction u:

$$\text{proj}_u v = \frac{u^T v}{\|u\|} \frac{u}{\|u\|}$$

When u is normalized to unit length, $\|u\| = 1$, this becomes

$$\text{proj}_u v = \left(u^T v\right) u$$

Questions?

Mini-Quiz 2:

Give an algebraic proof for the *triangle inequality*

$$\|x + y\|_2 \leq \|x\|_2 + \|y\|_2$$

using the Cauchy-Schwartz inequality. (Hint: expand $\|x + y\|^2$)

Solution (Q2):

Give an algebraic proof for the *triangle inequality*

$$\|x+y\|_2 \le \|x\|_2 + \|y\|_2$$

using the Cauchy-Schwartz inequality. (Hint: expand $\|x+y\|^2$)

Solution:

$$\begin{aligned}
\|x+y\|^2 &= (x+y)^T(x+y) \\
&= x^Tx + 2x^Ty + y^Ty \\
&= \|x\|^2 + \|y\|^2 + 2\left(x^Ty\right) \\
&\le \|x\|^2 + \|y\|^2 + 2|x^Ty| \\
&\le \|x\|^2 + \|y\|^2 + 2\|x\|\|y\| \quad \text{(by Cauchy-Schwartz)} \\
&= (\|x\| + \|y\|)^2
\end{aligned}$$

Operations with Matrices

- *Scalar multiplication*:

$$(\gamma A)_{ij} = \gamma A_{ij}, \qquad \gamma \begin{bmatrix} a_{11} & & a_{1n} \\ & \ddots & \\ a_{m1} & & a_{mn} \end{bmatrix} = \begin{bmatrix} \gamma a_{11} & & \gamma a_{1n} \\ & \ddots & \\ \gamma a_{m1} & & \gamma a_{mn} \end{bmatrix}$$

- *Matrix addition*:

$$(A + B)_{ij} = A_{ij} + B_{ij}$$

$$A + B = \begin{bmatrix} a_{11} + b_{11} & & a_{1n} + b_{1n} \\ & \ddots & \\ a_{m1} + b_{m1} & & a_{mn} + b_{mn} \end{bmatrix}, \quad A, B \in \mathbb{R}^{m \times n}$$

Matrix-vector multiplication

- *Matrix-vector multiplication*:

$$(Ax)_i = \sum_{j=1}^{n} A_{ij} x_j, \qquad A \in \mathbb{R}^{m \times n}, \quad x \in \mathbb{R}^n$$

- Matrix-vector product as linear combination of matrix columns

$$Ax = \begin{bmatrix} a_1 & a_2 & \cdots & a_n \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = x_1 \begin{bmatrix} a_1 \end{bmatrix} + x_2 \begin{bmatrix} a_2 \end{bmatrix} + \cdots + x_n \begin{bmatrix} a_n \end{bmatrix}$$

- Matrix-vector product in terms as a vector of scalar products

$$Ax = \begin{bmatrix} \bar{a}_1^T \\ \bar{a}_2^T \\ \vdots \\ \bar{a}_m^T \end{bmatrix} \begin{bmatrix} x \end{bmatrix} = \begin{bmatrix} \bar{a}_1^T x \\ \bar{a}_2^T x \\ \vdots \\ \bar{a}_m^T x \end{bmatrix}$$

Matrix-matrix multiplication

- *Matrix product*:

$$(AB)_{ij} = \sum_{k=1}^{p} A_{ik} B_{kj}, \qquad A \in \mathbb{R}^{m \times p}, \quad B \in \mathbb{R}^{p \times n}$$

- Matrix products in terms of inner products:

$$(AB)_{ij} = \bar{a}_i^T b_j, \qquad A = \begin{bmatrix} \bar{a}_1^T \\ \bar{a}_2^T \\ \vdots \\ \bar{a}_m^T \end{bmatrix}, \quad B = \begin{bmatrix} b_1 & b_2 & \cdots & b_n \end{bmatrix}$$

Since inner products are only defined for vectors of the same length, the matrix product requires that $\bar{a}_i$, $b_j \in \mathbb{R}^p$ for all i, j.

... Matrix-matrix multiplication

- Matrix products in terms of Matrix-vector products:

$$AB = A\begin{bmatrix} b_1 & b_2 & \cdots & b_n \end{bmatrix} = \begin{bmatrix} Ab_1 & Ab_2 & \cdots & Ab_n \end{bmatrix}$$

$$AB = \begin{bmatrix} \bar{a}_1^T \\ \bar{a}_2^T \\ \vdots \\ \bar{a}_m^T \end{bmatrix} B = \begin{bmatrix} \bar{a}_1^T B \\ \bar{a}_2^T B \\ \vdots \\ \bar{a}_m^T B \end{bmatrix}$$

  Matrix-matrix multiplication can be expressed in terms of the concatenation of matrix-vector products.

- For $A \in \mathbb{R}^{k \times l}$ and $B \in \mathbb{R}^{m \times n}$ the product AB is only defined if $l = m$, and the product BA is only defined if $k = n$.

Properties of Matrix-matrix multiplication

- Matrix multiplication is *associative*:

$$A(BC) = (AB)C$$

- Matrix multiplication is *distributive*:

$$A(B + C) = AB + AC$$

- Matrix multiplication is, in general, **not** *commutative*:

$$AB \neq BA$$

For example, let $A = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$ and $B = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}$, then

$$AB = \begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix} \neq \begin{bmatrix} 1 & 1 \\ 1 & 2 \end{bmatrix} = BA.$$

Vector Outer Product

- Using the definition of matrix multiplication, we can define the vector *outer product*:

$$xy^T = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_m \end{bmatrix} \begin{bmatrix} y_1 & y_2 & \cdots & y_n \end{bmatrix} = \begin{bmatrix} x_1y_1 & x_1y_2 & \cdots & x_1y_n \\ x_2y_1 & x_2y_2 & & x_2y_n \\ \vdots & & \ddots & \vdots \\ x_my_1 & x_my_2 & \cdots & x_my_n \end{bmatrix}$$

$$\left(xy^T\right)_{ij} = x_iy_j$$

$$x \in \mathbb{R}^m,\ y \in \mathbb{R}^n, \quad xy^T \in \mathbb{R}^{m\times n}$$

The outer product xy^T of vectors x and y is a matrix, and is defined even if the vectors are not the same length.

Questions?

Mini-Quiz 3:

Which of the following vector and matrix operations are well-defined?

1. αx
2. αA
3. $x - y$
4. $x^T y$
5. $x^T z$
6. $z x^T$
7. Az
8. $z^T A y$

$$\alpha = 5, \quad x = \begin{bmatrix} 1 \\ 5 \\ -2 \end{bmatrix}, \quad y = \begin{bmatrix} -4 \\ 0 \\ 1 \end{bmatrix}, \quad z = \begin{bmatrix} 3 \\ 0 \end{bmatrix}, \quad A = \begin{bmatrix} 4 & 0 & 2 \\ 1 & 5 & 3 \end{bmatrix}$$

Solution (Q3):

Which of the following vector and matrix operations are well-defined?

1. αx Yes
2. αA Yes
3. $x - y$ Yes
4. $x^T y$ Yes
5. $x^T z$ No
6. $z x^T$ Yes
7. Az No
8. $z^T A y$ Yes

$$\alpha = 5, \quad x = \begin{bmatrix} 1 \\ 5 \\ -2 \end{bmatrix}, \quad y = \begin{bmatrix} -4 \\ 0 \\ 1 \end{bmatrix}, \quad z = \begin{bmatrix} 3 \\ 0 \end{bmatrix}, \quad A = \begin{bmatrix} 4 & 0 & 2 \\ 1 & 5 & 3 \end{bmatrix}$$

Transpose of a Matrix

- *Transpose*: the transpose of matrix $A \in \mathbb{R}^{m \times n}$ is the matrix $A^T \in \mathbb{R}^{n \times m}$ formed by exchanging the rows and columns of A:

$$(A^T)_{ij} = A_{ji}$$

- Properties:

$$\begin{aligned} (A^T)^T &= A \\ (A+B)^T &= A^T + B^T \\ (AB)^T &= B^T A^T \\ (\alpha B)^T &= \alpha(B^T) \end{aligned}$$

Trace of a Matrix

- *Trace*: the trace of a square matrix $A \in \mathbb{R}^{n \times n}$ is the sum of its diagonal entries:

$$\mathrm{tr}(A) = \sum_{i=1}^{n} a_{ii}$$

- The trace is a linear map and has the following properties:

$$\begin{aligned} \mathrm{tr}(A+B) &= \mathrm{tr}(A) + \mathrm{tr}(B) \\ \mathrm{tr}(\gamma A) &= \gamma \mathrm{tr}(A), \quad \gamma \in \mathbb{R} \\ \mathrm{tr}(A) &= \mathrm{tr}(A^T) \\ \mathrm{tr}(AB) &= \mathrm{tr}(BA) \end{aligned}$$

- The trace of a matrix product functions similarly to the inner product of vectors:

$$\mathrm{tr}(A^T B) = \sum_{i,j} A_{ij} B_{ij}, \quad A, B \in \mathbb{R}^{m \times n}$$

Matrix norms

- *Induced norm (operator norm)* corresponding to vector norm $\|\cdot\|$:

$$\|A\| \;=\; \max_{\|x\|=1} \|Ax\|$$

  - $\|A\|_1$: maximum absolute column sum
  - $\|A\|_\infty$: maximum absolute row sum
  - $\|A\|_2 = \sqrt{\lambda_{max}(A^T A)} = \sigma_{max}(A)$ (*spectral norm*)

- *Frobenius norm*:

$$\|A\|_F = \sqrt{\sum_{i=1}^{m} \sum_{j=1}^{n} |a_{ij}|^2} = \sqrt{\text{trace}(A^T A)}$$

The Frobenius norm is equivalent to $\|\text{vec(A)}\|_2$.

Geometric Interpretation of Matrix Norms [1]

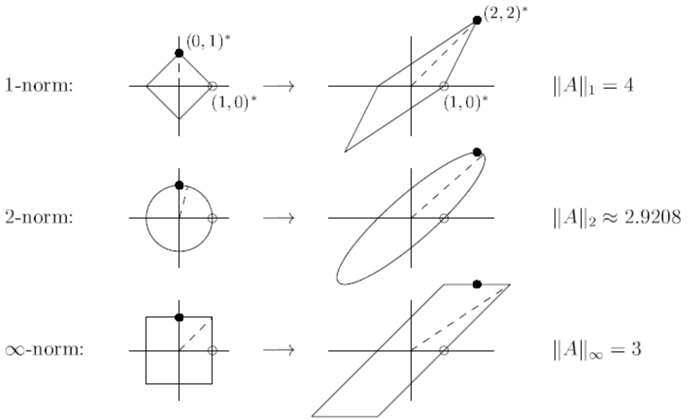


Figure 3.1. *On the left, the unit balls of* $\mathbb{R}^2$ *with respect to* $\|\cdot\|_1$, $\|\cdot\|_2$, *and* $\|\cdot\|_\infty$. *On the right, their images under the matrix* A *of* (3.7). *Dashed lines mark the vectors that are amplified most by* A *in each norm.*

[1] diagram from Trefethen and Bau.

Properties of Matrix Norms

- Matrix norms satisfy the same properties as vector norms:
  1. Scale invariance: $\|\gamma A\| = |\gamma| \|A\|$
  2. Triangle inequality: $\|A + B\| \leq \|A\| + \|B\|$
  3. Positivity: $\|A\| \geq 0$ and $\|A\| = 0$ only if $A \equiv 0$
- Some matrix norms, including induced norms and the Frobenius norm, also satisfy the *submultiplicative* property:

$$\|Ax\| \leq \|A\| \|x\|, \quad \text{and} \quad \|AB\| \leq \|A\| \|B\|$$

A few words on: Vector and Matrix Norms

- Vector norms is to give us a way to measure how "big" a vector is.
- 2-norm (Euclidean norm) is most common: it has nice relationship with the inner product $\|v\|_2^2 = v^T v$, and fits with our usual geometric notions of length and distance.
- The choice of norm may depend on the application:

  Let $y \in \mathbb{R}^n$ be a set of n measurements, and let $\hat{y} \in \mathbb{R}^n$ be the estimates for y from a model. Let $r = y - \hat{y}$ be the difference between the measured value and the model estimate. We want to optimize our model to make r as "small" as possible, but how should we measure "small"?
  - $\|r\|_\infty$: use this if no error should exceed a prescribed value δ.
  - $\|r\|_2$: use this if large errors are bad, but small errors are ok.
  - $\|r\|_1$: use this norm if you want to your measurement and estimate to match exactly for as many samples as possible.
- Matrix norms are an extension of vector norms to matrices.

Questions?

Mini-Quiz 4:

Determine whether each statement given below is True or False:

- There exist matrices $A, B \neq 0$ such that $AB = 0$.
- The trace is invariant under cyclic permutations:

$$\mathrm{tr}(ABC) = \mathrm{tr}(CAB) = \mathrm{tr}(BCA).$$

The definition of the induced norm allows us to use different vector norms on the numerator and denominator

$$\|A\|_{\alpha,\beta} = \max_{\|x\|_\beta = 1} \|Ax\|_\alpha.$$

- $\|A\|_{\infty,1}$ is equal to the largest element of A in absolute value .

Solutions (Q4):

Determine whether each statement given below is True or False:

- There exist matrices $A, B \neq 0$ such that $AB = 0$.

  TRUE : For example, when $A = \begin{bmatrix} 1 & 2 \\ 2 & 4 \end{bmatrix}$, and $B = \begin{bmatrix} -2 & 4 \\ 1 & -2 \end{bmatrix}$.

- The trace is invariant under cyclic permutations:

$$\mathrm{tr}(ABC) = \mathrm{tr}(CAB) = \mathrm{tr}(BCA).$$

TRUE

Use the property that $\mathrm{tr}(XY) = \mathrm{tr}(YX)$:
To show that $\mathrm{tr}(ABC) = \mathrm{tr}(CAB)$, let $X = AB$, and $Y = C$.
To show that $\mathrm{tr}(ABC) = \mathrm{tr}(BCA)$, let $X = A$ and $Y = BC$.

Solutions (Q4):

Determine whether each statement given below is True or False:

The definition of the induced norm allows us to use different vector norms on the numerator and denominator

$$\|A\|_{\alpha,\beta} = \max_{\|x\|_\beta = 1} \|Ax\|_\alpha.$$

- $\|A\|_{\infty,1}$ is equal to the largest element of A in absolute value .

TRUE

$$\|A\|_{\infty,1} = \max_{\|x\|_1 = 1} \|Ax\|_\infty = \max_{\|x\|_1 = 1} \left(\max_i (\bar{a}_i^T x_i) \right) = \max_{i,j} |a_{ij}|.$$

Special Matrices: Diagonal, Identity

- *Diagonal Matrix*: a matrix is diagonal if the only non-zero entries are on the matrix diagonal.

$$D_{ij} = \begin{cases} d_i & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases}, \qquad D = \begin{bmatrix} d_1 & 0 & \cdots & 0 \\ 0 & d_2 & & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \cdots & d_n \end{bmatrix}$$

- *Identity Matrix*: the identity $I_n \in \mathbb{R}^{n \times n}$ is a diagonal matrix with 1s along the diagonal (denoted by I when size is unambiguous).

$$I_{ij} = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases}, \qquad I = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix}$$

$$AI_n = I_m A = A \qquad \text{for rectangular } A \in R^{m \times n}$$

$$AI = IA = A \qquad \text{for square } A \in R^{n \times n}$$

Special Matrices: Triangular

- *Triangular Matrix*: a matrix is lower/upper triangular if all of the entries above/below the diagonal are zero:

$$L = \begin{bmatrix} l_{11} & & & \\ l_{21} & l_{22} & & \\ \vdots & \vdots & \ddots & \\ l_{m1} & l_{m2} & \cdots & l_{mn} \end{bmatrix}, \qquad U = \begin{bmatrix} u_{11} & u_{12} & \cdots & u_{1n} \\ & u_{22} & \cdots & u_{2n} \\ & & \ddots & \vdots \\ & & & u_{mn} \end{bmatrix}$$

  - The sum of two upper triangular matrices is upper triangular.
  - The product of two upper triangular matrices is upper triangular.
  - The inverse of an invertible upper triangular matrix is upper triangular.

Special Matrices: Symmetric, Orthogonal

- *Symmetric Matrix*: a square matrix is symmetric if it is equal to its own transpose.

$$A = A^T \in \mathbb{R}^{n \times n}, \qquad A = \begin{bmatrix} \alpha_1 & \beta & \cdots & \gamma \\ \beta & \alpha_2 & & \omega \\ \vdots & & \ddots & \vdots \\ \gamma & \omega & \cdots & \alpha_n \end{bmatrix}$$

- *Orthogonal (Unitary) Matrix*: a square matrix $U \in \mathbb{R}^{n \times n}$ is unitary if it satisfies

$$U^T U = U U^T = I.$$

The columns of a orthogonal matrix U are orthonormal.

Questions?

Mini-Quiz 5:

Determine whether each statement given below is True or False:

- For any $A \in \mathbb{R}^{n \times n}$, $A^T A$ is symmetric.
- For any $A \in \mathbb{R}^{n \times n}$, $A - A^T$ is symmetric.
- *Invariance under Unitary Multiplication.*
  For any matrix $x \in \mathbb{R}^m$ and unitary $Q \in \mathbb{R}^{m \times m}$, we have
  $$\|Qx\|_2 = \|x\|_2$$
- Let $D \in \mathbb{R}^{n \times n} = \operatorname{diag}(d_1, \cdots, d_n)$ be any diagonal matrix, then $DA = AD$ for any matrix $A \in \mathbb{R}^{n \times n}$.

Solution (Q5):

Determine whether each statement given below is True or False:

- For any $A \in \mathbb{R}^{n \times n}$, $A^T A$ is symmetric.

  $$\text{TRUE} : (A^T A)^T = (A)^T (A^T)^T = A^T A$$

- For any $A \in \mathbb{R}^{n \times n}$, $A - A^T$ is symmetric.

  FALSE : For example when $A = \begin{bmatrix} 1 & -1 \\ 1 & 0 \end{bmatrix}, \quad A - A^T = \begin{bmatrix} 0 & -2 \\ 2 & 0 \end{bmatrix}.$

Solution (Q5):

- *Invariance under Unitary Multiplication.*
  For any matrix $x \in \mathbb{R}^m$ and unitary $Q \in \mathbb{R}^{m \times m}$, we have

$$\|Qx\|_2 = \|x\|_2$$

TRUE : $\|Qx\|_2^2 = (Qx)^T(Qx) = x^T Q^T Q x = x^T x = \|x\|_2^2$

Similarly, for the matrix 2-norm and Frobenius norm, we have:

$$\begin{aligned} \|QA\|_2^2 &= \lambda_{max}\left((QA)^T(QA)\right) \\ &= \lambda_{max}\left(A^T Q^T Q A\right) \\ &= \lambda_{max}\left(A^T A)\right) \\ &= \|A\|_2^2 \end{aligned}$$

$$\text{and} \quad \begin{aligned} \|QA\|_F^2 &= \operatorname{tr}\left((QA)^T(QA)\right) \\ &= \operatorname{tr}\left(A^T A\right) \\ &= \|A\|_F^2 \end{aligned}$$

Solution (Q5):

Determine whether each statement given below is True or False:

- Let $D \in \mathbb{R}^{n\times n} = \text{diag}(d_1, \cdots, d_n)$ be any diagonal matrix, then $DA = AD$ for any matrix $A \in \mathbb{R}^{n\times n}$.

FALSE

DA rescales the rows of A while AD rescales the columns of A:

$$DA = \begin{bmatrix} d_1 & 0 & \cdots & 0 \\ 0 & d_2 & & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \cdots & d_n \end{bmatrix} \begin{bmatrix} \bar{a}_1^T \\ \bar{a}_2^T \\ \vdots \\ \bar{a}_n^T \end{bmatrix} = \begin{bmatrix} d_1\bar{a}_1^T \\ d_2\bar{a}_2^T \\ \vdots \\ d_n\bar{a}_n^T \end{bmatrix} \neq$$

$$AD = \begin{bmatrix} a_1 & \cdots & a_n \end{bmatrix} \begin{bmatrix} d_1 & 0 & \cdots & 0 \\ 0 & d_2 & & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \cdots & d_n \end{bmatrix} = \begin{bmatrix} d_1 a_1 & \cdots & d_n a_n \end{bmatrix}$$

Vector Spaces

Linear Transformations

- *Linear transformation*: $T : \mathbb{R}^n \to \mathbb{R}^m$ is a linear transformation if for all x, $y \in \mathbb{R}^n$ the following properties hold:
  - $T(x + y) = T(x) + T(y)$
  - $T(\alpha x) = \alpha T(x)$
- **Theorem:** Let $A \in \mathbb{R}^{m \times n}$ be a matrix and define the function $T_A(x) = Ax$. Then $T_A : \mathbb{R}^n \to \mathbb{R}^m$ is a linear transformation.
  - Matrix multiplication is a linear function.
- **Theorem:** Let $T : \mathbb{R}^n \to \mathbb{R}^m$, then there exists a matrix A such that $T(x) = Ax$.
  - There is a matrix for every linear function.

Linear Combinations

- Let $v_1, v_2, \cdots, v_r \in \mathbb{R}^m$ be a set of vectors. Then any vector v which can be written in the form

$$v = \alpha_1 v_1 + \alpha_2 v_2 + \cdots + \alpha_r v_r, \qquad \alpha_i \in \mathbb{R} \quad i = 1, \ldots, r$$

  is a linear combination of vectors $v_1, v_2, \cdots, v_r$.

- *Span*: the set of all linear combination of vectors $v_1, \cdots, v_r \in \mathbb{R}^m$ is called the span of $\{v_1, \cdots, v_r\}$
  - $span\{v_1, \cdots, v_r\}$, is always a subspace of $\mathbb{R}^m$.
  - If $S = span\{v_1, \cdots, v_r\}$, then S is *spanned by* $v_1, v_2, \cdots, v_r$.

- A set S is called a *subspace* if it is closed under vector addition and scalar multiplication:
  - if $x, y \in S$ and $\alpha \in \mathbb{R}$, then $x + y \in S$ and $\alpha x \in S$

Linear Independence

- *Linear dependence*: a set of vectors v_1, v_2, $\cdots$, v_r is linearly dependent if there exists a set of scalars α_1, α_2, $\cdots$, $\alpha_r \in \mathbb{R}$ with at least one $\alpha_i \neq 0$ such that

$$\alpha_1 v_1 + \alpha_2 v_2 + \cdots + \alpha_r v_r = 0$$

  That is, a set of vectors is linearly dependent if one of the vectors in the set can be written as a linear combination of one or more other vectors in the set.

- *Linear independence*: a set of vectors v_1, v_2, $\cdots$, v_r is linearly independent if it is *not* linearly dependent. That is

$$\alpha_1 v_1 + \alpha_2 v_2 + \cdots + \alpha_r v_r = 0 \quad \Longleftrightarrow \quad \alpha_1 = \cdots = \alpha_r = 0$$

Questions?

Mini-Quiz 6:

Are the following sets of vectors linearly independent or linearly dependent?

1. $\{a, b, c\}$
2. $\{b, c, d\}$
3. $\{a, b, d\}$
4. $\{a, b, c, d\}$

$$a = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \quad b = \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}, \quad c = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}, \quad d = \begin{bmatrix} 0 \\ -1 \\ 1 \end{bmatrix}$$

Solution (Q6):

Are the following sets of vectors linearly independent or linearly dependent?

1. $\{a, b, c\}$ independent
2. $\{b, c, d\}$ dependent: $-b + 2c = d$
3. $\{a, b, d\}$ independent
4. $\{a, b, c, d\}$ dependent: at most 3 linearly independent vectors in $\mathbb{R}^3$

$$a = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \quad b = \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}, \quad c = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}, \quad d = \begin{bmatrix} 0 \\ -1 \\ 1 \end{bmatrix}$$

Basis and Dimension

- A *basis* for a subspace S is a linearly independent set of vectors that span S
  - The basis for a subspace S is **not** unique, but all bases for the subspace contain the same number of vectors.
  - For example, consider the case where $S = \mathbb{R}^2$:

$$\left\{ \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \end{bmatrix} \right\} \quad \text{and} \quad \left\{ \begin{bmatrix} 1 \\ -1 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \end{bmatrix} \right\} \quad \text{are both bases that span } \mathbb{R}^2.$$

- The *dimension* of the subspace S, $dim(S)$, is the number of linearly independent vectors in the basis for S (in the example above, $dim(S) = 2$).

- **Theorem (unique representation)**: if vectors $v_1, \ldots, v_n$ are a basis for subspace S, then every vector in S can be uniquely represented as a linear combination of these basis vectors.

Range and Nullspace

- The *range (column space, image)* of a matrix $A \in \mathbb{R}^{m \times n}$, denoted by $\mathcal{R}(A)$, is the set of all linear combinations of the columns of A:

$$\mathcal{R}(A) := \{Ax \mid x \in \mathbb{R}^n\}, \quad \mathcal{R}(A) \subseteq \mathbb{R}^m$$

- The *nullspace (kernel)* of a matrix $A \in \mathbb{R}^{m \times n}$, denoted by $\mathcal{N}(A)$, is the set of vectors z such that $Az = 0$:

$$\mathcal{N}(A) := \{z \in \mathbb{R}^n \mid Az = 0\}, \quad \mathcal{N}(A) \subseteq \mathbb{R}^n$$

- The range and nullspace of A^T are called the *row space* and *left nullspace* of A.
  - These four subspaces of A are intrinsic to A and do not depend on the choice of basis.

Rank of a Matrix

- The *column rank* of A, denoted $\text{rank}(A)$, is dimension of $\mathcal{R}(A)$.
- The *row rank* of A is the dimension of $\mathcal{R}(A^T)$.
- *Rank of a matrix:* the column rank of a matrix is always equal to the row rank, and therefore we refer to this value as the rank of A.
- Matrix $A \in \mathbb{R}^{m \times n}$ is *full rank* if $rank(A) = \min\{m, n\}$.

Fundamental Theorem of Linear Algebra

- **Fundamental theorem of linear algebra**:
  1. The nullspace of A is the orthogonal complement of the row space.

$$\mathcal{N}(A) = \left(\mathcal{R}(A^T)\right)^{\perp}$$

  2. The left nullspace of A is the orthogonal complement of the column space.

$$\mathcal{N}(A^T) = \left(\mathcal{R}(A)\right)^{\perp}$$

- **Corollary (Rank-nullity)**:

$$dim\left(\mathcal{R}(A)\right) + dim\left(\mathcal{N}(A)\right) = n, \quad A \in \mathbb{R}^{m \times n}$$

Questions?

Mini-Quiz 7:

Give the range $\mathcal{R}(\cdot)$ and nullspace $\mathcal{N}(\cdot)$ for the matrices

$$A = \begin{bmatrix} 1 & 0 \\ 0 & 2 \\ 0 & 0 \end{bmatrix} \in \mathbb{R}^{3\times 2}, \quad \text{and} \quad B = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & -1 \end{bmatrix} \in \mathbb{R}^{2\times 3}$$

What are their dimensions ($dim\,(\mathcal{R}(\cdot))$ and $dim\,(\mathcal{N}(\cdot))$)?

Solution (Q7):

Give the range $\mathcal{R}(\cdot)$ and nullspace $\mathcal{N}(\cdot)$ for the matrices

$$A = \begin{bmatrix} 1 & 0 \\ 0 & 2 \\ 0 & 0 \end{bmatrix} \in \mathbb{R}^{3\times 2}, \quad \text{and} \quad B = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & -1 \end{bmatrix} \in \mathbb{R}^{2\times 3}$$

What are their dimensions ($dim\,(\mathcal{R}(\cdot))$ and $dim\,(\mathcal{N}(\cdot))$)?

Solution:

$$\mathcal{R}(A) = \left\{ \alpha \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} + \beta \begin{bmatrix} 0 \\ 2 \\ 0 \end{bmatrix},\ \alpha,\ \beta \in \mathbb{R} \right\}, \qquad \mathcal{N}(A) = \emptyset$$

$$\mathcal{R}(B) = \left\{ \alpha \begin{bmatrix} 1 \\ 0 \end{bmatrix} + \beta \begin{bmatrix} 0 \\ 1 \end{bmatrix},\ \alpha,\ \beta \in \mathbb{R} \right\}, \qquad \mathcal{N}(B) = \left\{ \gamma \begin{bmatrix} -1 \\ 1 \\ 1 \end{bmatrix},\ \gamma \in \mathbb{R} \right\}$$

$$dim\,(\mathcal{R}(A)) = 2 \qquad dim\,(\mathcal{N}(A)) = 0$$

$$dim\,(\mathcal{R}(B)) = 2 \qquad dim\,(\mathcal{N}(B)) = 1$$

Solving Linear Systems

Nonsingular Matrix

- **Theorem:** a matrix $A \in \mathbb{R}^{m \times n}$ with $m \geq n$ has full rank if and only if it maps no two distinct vectors to the same vector, and

$$Ax = Ay \implies x = y.$$

- *Singular matrix*: a square matrix $A \in \mathbb{R}^{n \times n}$ that is not full rank.
- *Nonsingular matrix*: a square matrix $A \in \mathbb{R}^{n \times n}$ of full rank.

  If A is nonsingular, we can uniquely express any vector $y \in \mathbb{R}^n$ as $y = Ax$ for some unique $x \in \mathbb{R}^n$. If we chose vectors $e_i = Ax_i$, $i = 1, \ldots, n$, we can write

$$AX = A \begin{bmatrix} x_1 & \cdots & x_n \end{bmatrix} = \begin{bmatrix} Ax_1 & \cdots & Ax_n \end{bmatrix} = \begin{bmatrix} e_1 & \cdots & e_n \end{bmatrix} = I$$

Matrix Inverse

- *Matrix inverse*: the inverse of a square matrix $A \in \mathbb{R}^{n\times n}$ is the matrix $B \in \mathbb{R}^{n\times n}$ such that

$$AB = BA = I \quad \text{where } I \text{ is the identity matrix}$$

  - The inverse of A is denoted by A^{-1}.
  - Any square, nonsingular matrix A has a unique inverse A^{-1} satisfying $AA^{-1} = A^{-1}A = I$.
  - $(AB)^{-1} = B^{-1}A^{-1}$ (assuming both inverses exist).
  - $(A^T)^{-1} = (A^{-1})^T$

- If $Ax = b$, then $A^{-1}b$ gives the vector of coefficients in the linear combination of the columns of A that yields b:

$$b = Ax = x_1a_1 + \cdots + x_na_n, \qquad x = A^{-1}b = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$$

Invertible Matrix Theorem

- Let $A \in \mathbb{R}^{n \times n}$ be a square matrix, then the following statements are equivalent:
    - A is invertible, A is nonsingular.
    - There exists a matrix A^{-1} such that $AA^{-1} = I_n = A^{-1}A$.
    - $Ax = b$ has exactly one solution for each $b \in \mathbb{R}^n$.
    - $Az = 0$ has only the trivial solution $z = 0$, i.e. $\mathcal{N}(A) = \emptyset$
    - The columns of A are linearly independent.
    - A is full rank, i.e. $rank(A) = n$.
    - $\det(A) \neq 0$.
    - 0 is not an eigenvalue of A.

Questions?

Mini-Quiz 8:

If A and $B \in \mathbb{R}^{n \times n}$ are two invertible matrices, which of the following formulas are necessarily true?

- A + B is invertible and $(A+B)^{-1} = A^{-1} + B^{-1}$.
- $(A+B)^2 = A^2 + 2AB + B^2$.
- $(ABA^{-1})^3 = AB^3A^{-1}$.
- $A^{-1}B$ is invertible and $\left(A^{-1}B\right)^{-1} = B^{-1}A$.

Solutions (Q8):

If A and $B \in \mathbb{R}^{n \times n}$ are two invertible matrices, which of the following formulas are necessarily true?

- A + B is invertible and $(A+B)^{-1} = A^{-1} + B^{-1}$.

FALSE : $A = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}, \quad B = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}, \quad A + B = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$

$$A^{-1} = \begin{bmatrix} 1 & -1 \\ 0 & 1 \end{bmatrix}, \quad B^{-1} = \begin{bmatrix} 1 & 0 \\ -1 & 1 \end{bmatrix}, \quad (A+B)^{-1} = \begin{bmatrix} \frac{2}{3} & -\frac{1}{3} \\ -\frac{1}{3} & \frac{2}{3} \end{bmatrix}$$

- $(A+B)^2 = A^2 + 2AB + B^2$.

FALSE : $(A+B)^2 = A^2 + AB + BA + B^2$ and $AB \neq BA$ for all invertible matrix pairs A, B.

Solutions (Q8):

If A and $B \in \mathbb{R}^{n\times n}$ are two invertible matrices, which of the following formulas are necessarily true?

- $(ABA^{-1})^3 = AB^3A^{-1}$ TRUE

$$\begin{aligned}(ABA^{-1})^3 &= (ABA^{-1})(ABA^{-1})(ABA^{-1} \\ &= AB(A^{-1}A)B(A^{-1}A)BA^{-1}) \\ &= AB^3A^{-1}\end{aligned}$$

- $A^{-1}B$ is invertible and $\left(A^{-1}B\right)^{-1} = B^{-1}A$ TRUE

$$\begin{aligned}(A^{-1}B)(A^{-1}B)^{-1} &= (A^{-1}B)(B^{-1}A) \\ &= A^{-1}(BB^{-1})A \\ &= A^{-1}A \\ &= I_n\end{aligned}$$

Systems of Linear Equations

Example: Find values x_1, $x_2 \in \mathbb{R}$ that satisfy

$$\begin{aligned} 2x_1 + x_2 &= 7 \\ \text{and} \quad x_1 - 3x_2 &= -7 \end{aligned}$$

(for example, if we want to find the intersection point of two lines).

This can be written as a matrix equation:

$$\begin{bmatrix} 2 & 1 \\ 1 & -3 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 7 \\ -7 \end{bmatrix} \qquad \text{or} \quad Ax = b$$

$$\text{where } A = \begin{bmatrix} 2 & 1 \\ 1 & -3 \end{bmatrix}, \quad x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}, \text{ and } b = \begin{bmatrix} 7 \\ -7 \end{bmatrix}$$

Linear Systems

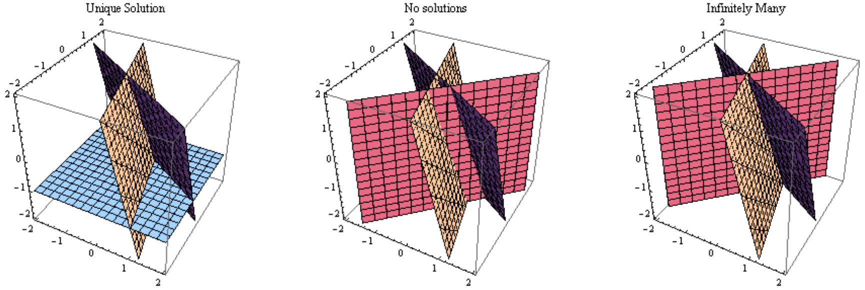
- One of the fundamental problems in linear algebra:
  Given $A \in \mathbb{R}^{m \times n}$ and $b \in \mathbb{R}^m$, find $x \in \mathbb{R}^n$ such that

$$Ax = b.$$

- The set of solutions of a linear system can behave in three ways:
  1. The system has one unique solution.
  2. The system has infinitely many solutions.
  3. The system has no solution.

- For a solution x to exist, we must have $b \in \mathcal{R}(A)$.
  - If $b \in \mathcal{R}(A)$ and $rank(A) = n$, then there is a unique solution (usually square systems: $m = n$).
  - If $b \in \mathcal{R}(A)$ and $rank(A) < n$ ($\mathcal{N}(A) \neq \emptyset$), then there are infinitely many solutions (usually underdetermined systems: $m < n$).
  - If $b \notin \mathcal{R}(A)$, then there is no solution (usually overdetermined systems: $m > n$).

Solution Set of Linear Systems [2]



[2] diagram from `http://oak.ucc.nau.edu/jws8/3equations3unknowns.html`

Gaussian Elimination (row reduction)

How do we solve $Ax = b$, $A \in \mathbb{R}^{n \times n}$?

- For some types of matrices, $Ax = b$ is easy to solve.
    - Diagonal matrices: equations independent, so $x_i = b_i / a_{ii}, \forall i$
    - Upper/Lower triangular matrices: back/forward substitution.
- *Gaussian Elimination*: Transform the system $Ax = b$ into an upper triangular system $Ux = y$ using elementary row operations.
- The following are elementary row operations that can be applied to the augmented system $\left[A \mid b\right]$ to introduce zeros below the diagonal without changing the solution set x:
    1. Multiply row by a scalar.
    2. Add scalar multiples of one row to another.
    3. Permute rows.

- *Step 1*: For $A \in \mathbb{R}^{3\times 3}$, $b \in \mathbb{R}^3$, form the augmented system.

$$\left[\begin{array}{c|c} A & b \end{array}\right] = \left[\begin{array}{ccc|c} \boxed{\times} & \times & \times & \times \\ \times & \times & \times & \times \\ \times & \times & \times & \times \end{array}\right]$$

- *Step 2*: Compute the row echelon form using row operations.

$$\left[\begin{array}{ccc|c} \boxed{\times} & \times & \times & \times \\ \times & \times & \times & \times \\ \times & \times & \times & \times \end{array}\right] \xrightarrow{L_1} \left[\begin{array}{ccc|c} \times & \times & \times & \times \\ 0 & + & + & + \\ 0 & + & + & + \end{array}\right]$$

$$\left[\begin{array}{ccc|c} \times & \times & \times & \times \\ & \boxed{\times} & \times & \times \\ & \times & \times & \times \end{array}\right] \xrightarrow{L_2} \left[\begin{array}{ccc|c} \times & \times & \times & \times \\ & + & + & + \\ & 0 & + & + \end{array}\right]$$

- *Step 3*: Solve the triangular system by back substitution.

$$\left[\begin{array}{ccc|c} \times & \times & \times & \times \\ & + & + & + \\ & & + & + \end{array}\right] = \left[\begin{array}{c|c} U & y \end{array}\right], \quad Ux = y \xrightarrow{\text{back substitution}} x = U^{-1}y$$

Questions?

Mini-Quiz 9:

Determine the lower triangular matrices L_i that correspond to the following elementary row operations:

- L_1 such that $L_1 A = A_1$:

$$L_1 \begin{bmatrix} 2 & 8 & 4 \\ 2 & 5 & 1 \\ 4 & 10 & -1 \end{bmatrix} = \begin{bmatrix} 1 & 4 & 2 \\ 2 & 5 & 1 \\ 4 & 10 & -1 \end{bmatrix}$$

- L_2 such that $L_2 A_1 = A_2$:

$$L_2 \begin{bmatrix} 1 & 4 & 2 \\ 2 & 5 & 1 \\ 4 & 10 & -1 \end{bmatrix} = \begin{bmatrix} 1 & 4 & 2 \\ 0 & -3 & -3 \\ 4 & 10 & -1 \end{bmatrix}$$

- L_3 such that $L_3 A_2 = A_3$:

$$L_3 \begin{bmatrix} 1 & 4 & 2 \\ 0 & -3 & -3 \\ 4 & 10 & -1 \end{bmatrix} = \begin{bmatrix} 1 & 4 & 2 \\ 0 & -3 & -3 \\ 0 & -6 & -9 \end{bmatrix}$$

Solution (Q9):

- $L_1 A = A_1$: $\begin{bmatrix} \frac{1}{2} & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 2 & 8 & 4 \\ 2 & 5 & 1 \\ 4 & 10 & -1 \end{bmatrix} = \begin{bmatrix} 1 & 4 & 2 \\ 2 & 5 & 1 \\ 4 & 10 & -1 \end{bmatrix}$

- $L_2 A_1 = A_2$: $\begin{bmatrix} 1 & 0 & 0 \\ -2 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 4 & 2 \\ 2 & 5 & 1 \\ 4 & 10 & -1 \end{bmatrix} = \begin{bmatrix} 1 & 4 & 2 \\ 0 & -3 & -3 \\ 4 & 10 & 1 \end{bmatrix}$

- $L_2 A_3 = A_3$: $\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -4 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 4 & 2 \\ 0 & -3 & -3 \\ 4 & 10 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 4 & 2 \\ 0 & -3 & -3 \\ 0 & -6 & -9 \end{bmatrix}$

- $L_3 L_2 L_1$ such that $L_3 L_2 L_1 A = A_3$:

$$\begin{bmatrix} \frac{1}{2} & 0 & 0 \\ -1 & 1 & 0 \\ -2 & 0 & 1 \end{bmatrix} \begin{bmatrix} 2 & 8 & 4 \\ 2 & 5 & 1 \\ 4 & 10 & -1 \end{bmatrix} = \begin{bmatrix} 1 & 4 & 2 \\ 0 & -3 & -3 \\ 0 & -6 & -9 \end{bmatrix}$$

Gaussian Elimination and LU Decomposition

- We can think of the elementary row operations applied to A as linear transformations, which can be represented as a series of lower triangular matrices, $L_1, \cdots, L_{n-1}$.

$$L_{n-1} \cdots L_1 A = U \qquad \text{or} \qquad L^{-1} A = U, \quad L^{-1} = L_1^{-1} \cdots L_{n-1}^{-1}$$

rearranging yields $A = LU$.

- If we can factor A as $A = LU$, then we can solve the system $Ax = b$ by solving two triangular systems:
  - Solve $Ly = b$ using forward substitution,
  - Solve $Ux = y$ using backward substitution.
- Gaussian elimination implicitly computes the LU factorization.

Partial Pivoting

- Pure Gaussian elimination can not be used to solve general linear systems because we may encounter a zero on the diagonal during the process, e.g.
$$A = \begin{bmatrix} 0 & 1 \\ 1 & 2 \end{bmatrix}$$
- In practical implementation of the algorithm, the rows of the matrix must be permuted to avoid division by zero.
- *Partial pivoting* swaps rows if an entry below the diagonal of the current column is larger in absolute value than the current diagonal entry (*pivot element*).

Permutations

- The reordering, or *permutation*, of the columns in partial pivoting can be represented as a linear transformation.

- *Permutation matrix*: a square, binary matrix $P \in \mathbb{R}^{n \times n}$ that has exactly one entry 1 in each row and each column.

    - Left-multiplication of a matrix A by a permuation matrix reorders the rows of A, while right-multiplication reorders the columns of A.

- For example, let $P = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix}, \quad A = \begin{bmatrix} 1 & 2 & 3 \\ 10 & 20 & 30 \\ 100 & 200 & 300 \end{bmatrix},$

  then the row and column permutations are, respectively:

$$PA = \begin{bmatrix} 10 & 20 & 30 \\ 100 & 200 & 300 \\ 1 & 2 & 3 \end{bmatrix}, \quad \text{and} \quad AP = \begin{bmatrix} 3 & 2 & 1 \\ 30 & 20 & 10 \\ 300 & 200 & 100 \end{bmatrix}.$$

LU Decomposition

- Let $A \in \mathbb{R}^{n \times n}$ be a square matrix, then the *LU decomposition* factors A into the product of a lower triangular matrix $L \in \mathbb{R}^{n \times n}$ and an upper triangular matrix $U \in \mathbb{R}^{n \times n}$

$$A = LU.$$

- When *partial pivoting* is used to permute the rows, we obtain a decomposition for the form $L_{n-1}P_{n-1} \cdots L_1 P_1 A = U$, which can be written[3] in form: $L^{-1}PA = U$.

- In general, any square matrix $A \in \mathbb{R}^{n \times n}$ (singular or nonsingular) has a factorization

$$PA = LU, \quad \text{where } P \text{ is a permutation matrix.}$$

[3] see *Trefethen and Bau: Chapter 21* for details.

Cholesky and Positive Definiteness

- *Positive definite matrix*: a matrix $A \in \mathbb{R}^{n \times n}$ is positive definite if

$$z^T A z > 0 \quad \text{for every } z \in \mathbb{R}^n, \quad \text{often denoted by } A \succ 0.$$

Example:

$$A = \begin{bmatrix} 4 & -2 \\ -2 & 10 \end{bmatrix}$$

$$\begin{aligned} x^T A x &= \begin{bmatrix} x_1 & x_2 \end{bmatrix} \begin{bmatrix} 4 & -2 \\ -2 & 10 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \\ &= 4x_1^2 - 4x_1 x_2 + 10x_2^2 \\ &= (2x_1 - x_2)^2 + 9x_2^2 > 0 \quad \forall x \neq 0 \end{aligned}$$

  - *Positive semi-definite matrix*: $z^T A z \geq 0$ for every $z \in \mathbb{R}^n$.
  - *Negative definite matrix*: $z^T A z < 0$ for every $z \in \mathbb{R}^n$.

... Cholesky and Positive Definiteness

- A symmetric matrix A is positive definite ($A \succ 0$) if and only if there exists a unique lower triangular matrix L such that

$$A = LL^T.$$

  This factorization is called the *Cholesky decomposition* of A.

- The *Cholesky algorithm*, a modified version of Gauss elimination, is used to compute the factor L.
  - *Basic idea*: In Gauss-elimination we use row operations to introduce zeros below the diagonal in each column. To maintain symmetry, we can apply the same operations to the columns of the matrix to introduce zeros in the first row.

$$A = \begin{bmatrix} \times & \times & \times \\ \times & \times & \times \\ \times & \times & \times \end{bmatrix} \xrightarrow{L_1 A} \begin{bmatrix} \times & \times & \times \\ 0 & + & + \\ 0 & + & + \end{bmatrix} = L_1 A = A_1$$

$$A_1 = \begin{bmatrix} \times & \times & \times \\ 0 & \times & \times \\ 0 & \times & \times \end{bmatrix} \xrightarrow{A_1 L_1^T} \begin{bmatrix} \times & 0 & 0 \\ 0 & + & + \\ 0 & + & + \end{bmatrix} = A_2 = A_1 L_1^T = L_1 A L_1^T$$

Questions?

Mini-Quiz 10:

- *Quadratic forms:* a function $q(x) : \mathbb{R}^n \to \mathbb{R}$ if it is a linear combination of functions of the form $x_i x_j$. A quadratic form can be written as

$$q(x) = x^T A x, \quad \text{where } A \in \mathbb{R}^{n \times n} \text{ is symmetric.}$$

  Let $B \in \mathbb{R}^{m \times n}$. Show that $q(x) = \|Bx\|^2$ is a quadratic form, find A such that $q(x) = x^T A x$, and determine the definiteness of A.

- Given A, B, find the permutation matrix P such that $PBP = A$:

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix}, \quad B = \begin{bmatrix} 1 & 3 & 2 \\ 7 & 9 & 8 \\ 4 & 6 & 5 \end{bmatrix}$$

Solution (Q10):

- Let $B \in \mathbb{R}^{m \times n}$. Show that $q(x) = \|Bx\|^2$ is a quadratic form, find A such that $q(x) = x^T Ax$, and determine the definiteness of A.

  Solution:

$$q(x) = (Ax)^T(Ax) = x^T A^T Ax = x^T Bx, \quad \text{where } B = A^T A.$$

  $q(x)$ is *positive semi-definite*, since $q(x) = \|Ax\|^2 \geq 0,\ \forall x \in \mathbb{R}^m$.

- Given A, B, find the permutation matrix P such that $PBP = A$:

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix}, \quad B = \begin{bmatrix} 1 & 3 & 2 \\ 7 & 9 & 8 \\ 4 & 6 & 5 \end{bmatrix}, \text{ and } P = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}$$

Orthogonalization

Orthogonal (Unitary) Matrices

- Recall that two vectors x, $y \in \mathbb{R}^n$ are *orthogonal* if $x^T y = 0$, and a set of vectors $\{u_1, \cdots, u_n\}$ is *orthonormal* if

$$u_i^T u_j = \begin{cases} 1 & \text{for } i = j \\ 0 & \text{for } i \neq j \end{cases}, \quad \|u_i\| = 1, \ \forall i$$

- *Orthogonal (Unitary) Matrix*: a matrix $Q \in \mathbb{R}^{n \times n}$ is orthogonal if its columns $q_1, \cdots, q_n \in \mathbb{R}^n$ form an orthonormal set:

$$Q^T Q = \begin{bmatrix} q_1^T \\ \vdots \\ q_n^T \end{bmatrix} \begin{bmatrix} q_1 & \cdots & q_n \end{bmatrix} = I_n$$

  - If Q is orthogonal, then $Q^{-1} = Q^T$ (inverse is easy to compute!)
  - Orthogonal transformations preserve lengths and angles and represent rotations or reflections.

$$\|Qx\|_2 = \sqrt{x^T Q^T Q x} = \sqrt{x^T I x} = \|x\|_2$$

Projectors

- Recall our notation for the *projection* of vector v onto u:

$$\mathrm{proj}_u v = \frac{u^T v}{\|u\|} \frac{u}{\|u\|}$$

- We can generalize the idea to projections onto subspaces of $\mathbb{R}^n$.
- *Projector:* a square matrix $P \in \mathbb{R}^{n \times n}$ is a projector if $P = P^2$.
  - For $v \in \mathbb{R}^n$, Pv represents the projection of v into the range of P.
  - If $v \in \mathcal{R}(P)$ $(\exists x : v = Px)$, then $Pv = P^2 x = Px = v$.

Orthogonal Projectors

- If $x \in \mathbb{R}^m$ and $A \in \mathbb{R}^{m \times n}$ is a matrix whose columns form a subspace of $\mathbb{R}^m$, then x can be decomposed as:

$$x = x^{\parallel} + x^{\perp}, \quad \text{where } x^{\parallel} \in \mathcal{R}(A) \text{ and } x^{\perp} \in \mathcal{R}(A)^{\perp}$$

- $x^{\parallel}$ is the *orthogonal projection* of x onto $\mathcal{R}(A)$.
  - $x^{\parallel}$ is the vector in $\mathcal{R}(A)$ "closest" to x, in the sense that

$$\|x - x^{\parallel}\| < \|x - v\|, \quad \forall v \in \mathcal{R}(A) \neq x^{\parallel}.$$

- An *orthogonal projector* is a projector for which the range and nullspace are orthogonal complements.
  - An orthogonal projector satisfies $P^T = P$.
  - If $u \in \mathbb{R}^n$ is a unit vector, then $P_u = uu^T$ is the orthogonal projection onto the line containing u.
  - If $u_1, \cdots, u_k$ are an orthonormal basis for subspace V and $A = [u_1 \cdots u_n]$, then the projection onto V is

$$P_A = AA^T$$

$$\text{or equivalently: } \operatorname{proj}_A x = P_A x \quad = \quad (u_1^T x)u_1 + \cdots + (u_n^T x)u_n$$
$$= \quad \operatorname{proj}_{u_1} x + \cdots + \operatorname{proj}_{u_n} x$$

Questions?

Mini-Quiz 11:

- Show that if P is an orthogonal projector, then $I - 2P$ is orthogonal.

  Hint: $P^T = P$ for orthogonal projectors and $P^2 = P$ for all projectors. Q is orthogonal if $Q^T Q = I$.

- Find the matrix for the orthogonal projector P_A onto $\mathcal{R}(A)$:

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

  Hint: The columns of A are orthogonal, but not orthonormal because the first column does not have length 1. The projection onto $\mathcal{R}(Q)$ is $P_Q = QQ^T$ if the columns of Q are orthonormal.

Solution (11):

- Show that if P is an orthogonal projector, then $I-2P$ is orthogonal. Solution:

$$\begin{aligned}(I-2P)^T(I-2P) &= I^2 - 2P^TI - 2IP + 4P^TP \\ &= 4P^2 - 4P + I \quad (P^T = P) \\ &= 4P - 4P + I \quad (P^2 = P) \\ &= I\end{aligned}$$

- Find the orthogonal projector P_A onto $\mathcal{R}(A)$: $A = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 0 \end{bmatrix}$.

$$\mathcal{R}(A) = \mathcal{R}\left(Q = \begin{bmatrix} \frac{1}{\sqrt{2}} & 0 \\ 0 & 1 \\ \frac{1}{\sqrt{2}} & 0 \end{bmatrix} \right)$$

$$P_A = P_Q = QQ^T = \begin{bmatrix} \frac{1}{2} & 0 & \frac{1}{2} \\ 0 & 1 & 0 \\ \frac{1}{2} & 0 & \frac{1}{2} \end{bmatrix}$$

Gram-Schmidt Orthogonalization

- *Gram-Schmidt algorithm*: a method for computing an orthonormal basis $\{q_1, \cdots, q_m\}$ for a set of vectors $A = [a_1, \cdots, a_n] \in \mathbb{R}^{m \times n}$.

- Each vector is orthogonalized with respect to the previous vectors and then normalized:

$$
\begin{aligned}
v_1 &= a_1 & q_1 &= \tfrac{v_1}{\|v_1\|} \\
v_2 &= a_2 - \mathrm{proj}_{q_1} a_2 & q_2 &= \tfrac{v_2}{\|v_2\|} \\
&\vdots & &\vdots \\
v_k &= a_k - \sum_{i=1}^{k-1} \mathrm{proj}_{q_i} a_k & q_k &= \tfrac{v_k}{\|v_k\|}
\end{aligned}
$$

In matrix form: $A = QR$, where

$$
Q = \big[q_1, \cdots, q_m\big] \text{ is orthogonal,} \quad \text{and} \quad R_{ij} = \begin{cases} q_i^T a_j & \text{if } i \le j \\ 0 & \text{if } i > j \end{cases}
$$

QR Decomposition

- In the process of orthogonalizing the columms of matrix A, the Gram-Schmidt algorithm computes the *QR decomposition* of A.

- *QR decomposition*: any matrix $A \in \mathbb{R}^{m\times n}$ can be decomposed as

$$A = QR$$

where $Q \in \mathbb{R}^{m\times m}$ is an orthogonal matrix, and $R \in \mathbb{R}^{m\times n}$ is an upper triangular matrix.

- For rectangular $A \in \mathbb{R}^{m\times n}$, $m \geq n$

$$A = QR = \begin{bmatrix} Q_1 & Q_2 \end{bmatrix} \begin{bmatrix} R_1 \\ 0 \end{bmatrix} = Q_1 R_1$$

where $R_1 \in \mathbb{R}^{n\times n}$ is upper triangular, and $Q_1 \in \mathbb{R}^{m\times n}$ and $Q_2 \in \mathbb{R}^{m\times(m-n)}$ both have orthogonal columns.

QR Decomposition for Solving Linear Systems

- Consider the linear system $Ax = b$, where $A \in \mathbb{R}^{n \times n}$ is a square, nonsingular matrix and $b \in \mathbb{R}^n$.
  - The QR decomposition expresses the matrix $A = QR$ as the product of an orthogonal matrix Q ($QQ^T = I_n$) and upper triangular R:

    $$Ax = b \implies QRx = b \implies Rx = Q^T b$$

  - $y = Q^T b$ can be easily computed, and $Rx = y$ is a triangular system that can be solved by back substitution.

Questions?

Mini-Quiz 12:

- Use Gram-Schmidt to compute an orthonormal basis for the columns of A :

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

  Hint: All three columns are normalized to length 1. The first and second columns are orthogonal, and the second and third columns are orthogonal. The projection operator for normalized vectors is $\mathrm{proj}_u v = (u^T v)u$.

Solution (Q12):

- Use Gram-Schmidt to compute an orthonormal basis for A :

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

$$\begin{aligned} q_1 &= a_1, \qquad q_2 = a_2 \\ v_3 &= a_3 - \mathrm{proj}_{q_1} a_3 - \mathrm{proj}_{q_2} a_3 = a_3 - (a_1^T a_3) a_1 \\ &= \begin{bmatrix} \frac{2}{3} \\ \frac{2}{3} \\ \frac{1}{3} \end{bmatrix} - \frac{5}{3\sqrt{3}} \begin{bmatrix} \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{3}} \end{bmatrix} = \begin{bmatrix} \frac{1}{9} \\ \frac{1}{9} \\ -\frac{2}{9} \end{bmatrix} \\ q_3 &= \frac{v_3}{\|v_3\|} = \begin{bmatrix} \frac{1}{\sqrt{6}} \\ \frac{1}{\sqrt{6}} \\ -\frac{2}{\sqrt{6}} \end{bmatrix} \end{aligned}$$

Least Squares Problems

Least Squares

- We want to solve $Ax = b$, but what do we do if $b \notin \mathcal{R}(A)$ (i.e. there does not exist a vector x such that $Ax = b$)?
  - For example, let $A \in \mathbb{R}^{m \times n}$, $m > n$ be a tall-and-skinny (*overdetermined*) matrix. Then for most $b \in \mathbb{R}^m$, there is no solution $x \in \mathbb{R}^n$ such that $Ax = b$.
- *Least Squares problem:* Define *residual* $r = b - Ax$ and find the vector x that minimizes

$$\|r\|_2^2 = \|b - Ax\|_2^2.$$

... Least Squares

- We can decompose any vector $b \in \mathbb{R}^m$ into components $b = b_1 + b_2$, with $b_1 \in \mathcal{R}(A)$ and $b_2 \in \mathcal{N}(A^T)$.

- Since b_2 is in the orthogonal complement of the $\mathcal{R}(A)$, we obtain the following expression for the residual norm:

$$\|r\|_2^2 = \|b_1 - Ax + b_2\|_2^2 = \|b_1 - Ax\|_2^2 + \|b_2\|_2^2$$

  which is minimized when $Ax = b_1$ and $r = b_2 \in \mathcal{N}(A^T)$.

Normal Equations

- The least squares solution x occurs when $r = b_2 \in \mathcal{N}(A^T)$ or equivalently $A^T r = A^T(b - Ax) = 0$.
- Rearranging this expression gives the normal equations:

$$A^T A x = A^T b$$

  - If A has full column rank, then $\tilde{A} = A^T A$ is invertible. Thus $\tilde{A}x = \tilde{b}$ (where $\tilde{b} = A^T b$) has a unique solution x.

Least Squares and Orthogonal Projections

- If $x^\star \in \mathbb{R}^n$ is the least squares solution to $Ax = b$, then $Ax^\star$ is the orthogonal projection of b onto $\mathcal{R}(A)$.

$$
\begin{aligned}
x^\star = \arg\min_x \|b - Ax\| \quad & \implies \quad \|b - Ax^\star\| \leq \|b - Ax\|, \quad \forall x \in \mathbb{R}^n \\
& \implies \quad Ax^\star = P_A b \\
& \qquad\quad (P_A \text{ is the projection onto } \mathcal{R}(A)) \\
& \implies \quad b - Ax^\star \in \mathcal{R}(A)^\perp = \mathcal{N}(A^T) \\
& \implies \quad A^T(b - Ax^\star) = 0 \\
& \implies \quad A^T A x^\star = A^T b
\end{aligned}
$$

- Combining the solution $x^\star = (A^T A)^{-1} A^T b$ and $Ax^\star = P_A b$ gives the matrix for the orthogonal projection onto $\mathcal{R}(A)$:

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

Questions?

Mini-Quiz 13

Find the least squares solution of

$$\begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix}.$$

by solving $Ax = b_1$ or $A^T A x = A^T b$

Hints:

$$\mathcal{R}(A) = \left\{ \alpha \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} + \beta \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix},\ \alpha, \beta \in \mathbb{R} \right\}, \quad \mathcal{N}(A^T) = \left\{ \gamma \begin{bmatrix} 0 \\ 1 \\ -1 \end{bmatrix},\ \gamma \in \mathbb{R} \right\},$$

$$b = 1 \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} + 1.5 \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix} + 0.5 \begin{bmatrix} 0 \\ 1 \\ -1 \end{bmatrix}$$

Mini-Quiz 13

Find the least squares solution of

$$\begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix}.$$

Solution:

$$b = b_1 + b_2 + \begin{bmatrix} 1 \\ 1.5 \\ 1.5 \end{bmatrix} + \begin{bmatrix} 0 \\ 0.5 \\ -0.5 \end{bmatrix}, \quad A^T A = \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix}, \quad A^T b = \begin{bmatrix} 1 \\ 3 \end{bmatrix}$$

$$Ax = b_1 \implies \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 1 \\ 1.5 \\ 1.5 \end{bmatrix} \implies \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 1 \\ 1.5 \end{bmatrix}$$

$$A^T A x = A^T b \implies \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 1 \\ 3 \end{bmatrix} \implies \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 1 \\ 1.5 \end{bmatrix}$$

Least Squares via QR

- The normal equation $A^T A x = A^T b$ can be rewritten using the QR factorization $A = QR$:

$$\begin{aligned} A^T A x &= A^T b \\ R^T Q^T Q R x &= R^T Q^T b \\ R^T R x &= R^T Q^T b \\ R x &= Q^T b \end{aligned}$$

- Solution using QR decomposition:
  - compute QR factorization of $A: \ A = QR$
  - form $y = Q^T b$
  - solve $Rx = y$ by back substitution

Least Squares via Cholesky

- $M = A^T A$ is symmetric and positive definite, and can be rewritten in terms the Cholesky factorization $M = LL^T$
- Solution using Cholesky factorization:
  - form $M = A^T A$
  - compute the Cholesky factorization of M: $M = LL^T$
  - form $y = A^T b$
  - solve $Lz = y$ by forward substitution
  - solve $L^T x = z$ by back substitution

Matrix Calculus

- Vector-by-scalar and Scalar-by-vector derivatives:

$$\alpha \in \mathbb{R},\ x \in \mathbb{R}^n, \quad \frac{\partial \alpha}{\partial x} = \begin{bmatrix} \frac{\partial \alpha}{\partial x_1} \\ \frac{\partial \alpha}{\partial x_2} \\ \vdots \\ \frac{\partial \alpha}{\partial x_n} \end{bmatrix}, \quad \frac{\partial x}{\partial \alpha} = \begin{bmatrix} \frac{\partial x_1}{\partial \alpha} & \frac{\partial x_2}{\partial \alpha} & \cdots & \frac{\partial x_n}{\partial \alpha} \end{bmatrix}$$

- Vector-by-vector derivatives:

$$y \in \mathbb{R}^m,\ x \in \mathbb{R}^n, \quad \frac{\partial y}{\partial x} = \begin{bmatrix} \frac{\partial y_1}{\partial x_1} & \frac{\partial y_2}{\partial x_1} & \cdots & \frac{\partial y_m}{\partial x_1} \\ \frac{\partial y_1}{\partial x_2} & \frac{\partial y_2}{\partial x_2} & & \frac{\partial y_m}{\partial x_2} \\ \vdots & & \ddots & \vdots \\ \frac{\partial y_1}{\partial x_n} & \frac{\partial y_2}{\partial x_n} & \cdots & \frac{\partial y_m}{\partial x_n} \end{bmatrix}$$

… Matrix Calculus[4]

- Vector-by-vector identities:

$$\frac{\partial(x^T x)}{\partial x} = 2x, \qquad \frac{\partial(b^T x)}{\partial x} = b, \qquad \frac{\partial(Ax)}{\partial x} = A^T$$

$$\frac{\partial(x^T A x)}{\partial x} = \left(A + A^T\right) x \quad \left(= 2Ax, \text{ if } A = A^T\right)$$

- Derivative of a quadratic:

$$\tfrac{\partial}{\partial x}\|b - Ax\|_2^2 = \tfrac{\partial}{\partial x}\left(x^T A^T A x - 2b^T A x + b^T b\right) = 2\left(A^T A x - A^T b\right)$$

  - Setting this expression equal to zero gives the normal equations:

$$\tfrac{\partial}{\partial x}\|b - Ax\|_2^2 = 0 \implies A^T A x = A^T b$$

[4] see `http://en.wikipedia.org/wiki/Matrix_calculus` for more properties.

Eigenvalues and Eigenvectors

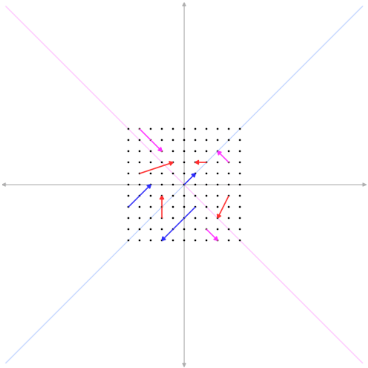
Eigenvalues and Eigenvectors

- For any square matrix $A \in \mathbb{R}^{n \times n}$, there is at least one scalar λ and a corresponding vector $v \neq 0$ such that:

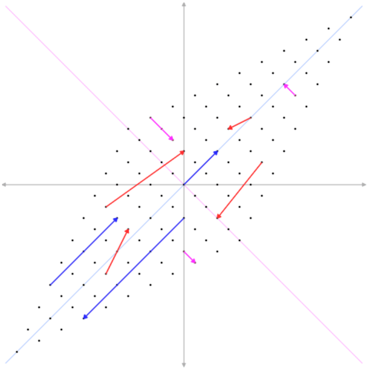
$$Av = \lambda v \qquad \text{or equivalently} \qquad (A - \lambda I)v = 0.$$

This scalar λ is an *eigenvalue* of A and v is an *eigenvector* corresponding to eigenvalue λ.

Geometric interpretation of Eigenvectors [5]



(a)



(b)

Figure: Under the transformation matrix $A = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$, the directions of vectors parallel to $v_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$ (blue) and $v_2 = \begin{bmatrix} 1 \\ -1 \end{bmatrix}$ (purple) are preserved.

[5] diagram from `wikipedia.org/wiki/Eigenvalues_and_eigenvectors`

- The set of distinct eigenvalues of A is called the *spectrum* of A and is denoted $\lambda(A)$:

$$\lambda(A) := \{\lambda \in \mathbb{R} \mid A - \lambda I \text{ is singular.}\}$$

- The magnitude of the largest eigenvalue in absolute value is called the *spectral radius* of A,

$$\rho(A) = \max_{\lambda_i \in \lambda(A)} |\lambda_i|.$$

- The set of all eigenvectors associated with an eigenvalue λ form the *eigenspace*, $E_\lambda = \mathcal{N}(A - \lambda I)$, associated with λ.

Similar Matrices

- Two matrices A and B are *similar* is they share the same eigenvalues.
  - Similar matrices represent the same linear transformation under two different bases.
  - If P is a nonsingular matrix, then $B = P^{-1}AP$ is similar to A. P is called a *similarity transformation* or *change of basis* matrix.
  - A and B do not in general have the same eigenvectors. If v is an eigenvector of A, then $P^{-1}v$ is an eigenvector of B.
- Given a square matrix $A \in \mathcal{R}^{n \times n}$, we wish to to reduce it to its simplest form by means of a similarity transformation.

Diagonalizable Matrices

- *Diagonalizable matrix*: a matrix $A \in \mathcal{R}^{n \times n}$ that is similar to a diagonal matrix, i.e. there exists an invertible matrix P such that

$$P^{-1}AP = D, \quad \text{where } D \text{ is diagonal.}$$

- A matrix is diagonalizable if and only if it has n linearly independent eigenvectors.
  - Matrices with n distinct eigenvalues are diagonalizable.
- Real symmetric matrices are diagonalizable by unitary matrices.
  - Matrices are diagonalizable by unitary matrices if and only if they are *normal* (a matrix is normal if it satisfies $A^T A = AA^T$).

Eigendecomposition

- Let $A \in \mathbb{R}^{n \times n}$ be a square, diagonalizable matrix, then A can be factorized as

$$A = V \Lambda V^{-1}$$

where $\Lambda = \text{diag}(\lambda_1, \cdots, \lambda_n)$ is a diagonal matrix whose elements are eigenvalues of A, and V is a square matrix whose ith column v_i is the eigenvector corresponding to eigenvalue $\Lambda_{ii} = \lambda_i$.

  - Note: this decomposition does not exist for every square matrix A. e.g. $A = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$ can not be diagonalized.

Spectral Theorem

- Any symmetric matrix $A = A^T \in \mathbb{R}^{n \times n}$, has n (not necessarily distinct) eigenvalues, and A can be decomposed with the *symmetric eigenvalue decomposition*:

$$A = \sum_{i=1}^{n} \lambda_i u_i u_i^T = U \Lambda U^T,$$

where U is orthogonal and $\Lambda = \mathrm{diag}(\lambda_1, \cdots, \lambda_n)$ is diagonal.

Properties of Eigenvalues

- The trace of $A \in \mathbb{R}^{n \times n}$ is equal to the sum of its eigenvalues:

$$\operatorname{tr}(A) = \sum_{i=1}^{n} \lambda_i$$

- The rank of $A \in \mathbb{R}^{n \times n}$ is equal to the number of non-zero eigenvalues of A.

- If $A \in \mathbb{R}^{n \times n}$ is non-singular with eigenvalue $\lambda_i \neq 0$, then $\frac{1}{\lambda_i}$ is an eigenvalue of A^{-1}.

- If $A \in \mathbb{R}^{n \times n}$ is symmetric positive definite ($A = A^T$ and $z^T A z > 0,\ \forall z \in \mathbb{R}^n$), then all of its eigenvalues are positive.

Questions?

Mini-Quiz 14

Show that if λ is an eigenvalue of a nonsingular matrix A with corresponding eigenvector v, then v is also an eigenvector of A^{-1} with eigenvalue $\frac{1}{\lambda}$.

Hint: we know that $Av = \lambda v$ and we want to show that $A^{-1}v = \frac{1}{\lambda}v$.

Solution (Q14)

Show that if λ is an eigenvalue of a nonsingular matrix A with corresponding eigenvector v, then v is also an eigenvector of A^{-1} with eigenvalue $\frac{1}{\lambda}$.

$$\begin{aligned} Av &= \lambda v \\ A^{-1}(Av) &= A^{-1}(\lambda v) \\ v &= \lambda(A^{-1}v) \\ \tfrac{1}{\lambda}v &= A^{-1}v \end{aligned}$$

Solving Eigenvalue Problems

- Solving eigenvalue problems $Av = \lambda v$ is a fundamentally more difficult problem than solving a linear system $Ax = b$.
- Solving the eigenvalue problem for an $n \times n$ matrix can be reduced to solving for the roots of an nth degree polynomial. For $n \geq 5$, no equation exists for the roots of an arbitrary nth degree polynomial given its coefficients.
- All algorithms to solve eigenvalue problems for matrices of arbitrary size are iterative. This contrasts with linear systems which have algorithms that are guaranteed to produce the solution in a finite number of steps.

Solving Small Eigenvalue Problems

- Recall that if $Av = \lambda v$, then $(A - \lambda I)v = 0$.
  - Since $(A - \lambda I)$ has a vector, v, in its nullspace, it is not singular (non-invertible).
  - Therefore finding the eigenvalues of A is equivalent to finding the values λ for which $(A - \lambda I)$ is singular.
- For a 2×2 matrix,

$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}, \text{ the matrix inverse is } A^{-1} = \tfrac{1}{ad-bc} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}.$$

  The quantity $ad - bc$ is called the *determinant* of A, $\det(A)$, and the inverse exists if and only if $\det(A) \neq 0$.
- Therefore, for our 2×2 eigenvalue problem, we want to find the eigenvalues λ such that $\det(A - \lambda I) = 0$:

$$(A-\lambda I) = \begin{bmatrix} (a - \lambda) & b \\ c & (d - \lambda) \end{bmatrix}, \quad \det(A-\lambda I) = (a-\lambda)(d-\lambda)-bc = 0$$

- Example:

$$A = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}, \quad (A - \lambda I) = \begin{bmatrix} (2-\lambda) & 1 \\ 1 & (2-\lambda) \end{bmatrix}$$

$$\begin{aligned} \det(A - \lambda I) &= (2-\lambda)(2-\lambda) - 1 \cdot 1 \\ &= \lambda^2 - 4\lambda + 3 \end{aligned}$$

This is a 2nd degree polynomial (i.e. quadratic). We want to find the roots of this polynomial, λ_1, and λ_2, which will satisfy $\det(A - \lambda_i I) = 0$.

$$\begin{aligned} \lambda^2 - 4\lambda + 3 &= 0 \\ (\lambda - 3)(\lambda - 1) &= 0 \end{aligned}$$

$$\implies \lambda_1 = 3, \quad \lambda_2 = 1$$

Power Method for Finding Eigenvectors

- One method for finding the eigenvector, v_1, corresponding to the largest eigenvalue, λ_1, of $A \in \mathbb{R}^{n\times n}$ is called the *power method*:
- Pick any vector $z \in \mathbb{R}^n$, and compute the sequence:

$$\left\{ Az,\, A^2z,\, A^3z,\, A^4z,\, \cdots,\, A^kz \right\}, \quad \text{then } A^kz \to v_1 \text{ as } k \to \infty.$$

z can be written as a linear combination of the (linearly independent) eigenvectors of A:

$$z = \alpha_1 v_1 + \cdots + \alpha_n v_n$$

$$\begin{aligned} \text{Therefore} \quad Az &= A(\alpha_1 v_1 + \cdots + \alpha_n v_n) \\ &= \alpha_1 A v_1 + \cdots + \alpha_n A v_n \\ &= \alpha_1 \lambda_1 v_1 + \cdots + \alpha_n \lambda_n v_n \end{aligned}$$

$$\text{and iterating gives} \quad A^kz = \alpha_1 \lambda_1^k v_1 + \cdots + \alpha_n \lambda_n^k v_n$$

since λ_1 is the largest eigenvalue, λ_1^k will dominate the sum, so A^kz will converge toward the direction of v_1.

The Singular Value Decomposition

Singular Value Decomposition

- **Theorem (Singular Value Decomposition)**: every matrix $A \in \mathbb{R}^{m \times n}$ has a decomposition, called the *singular value decomposition (SVD)* of the form

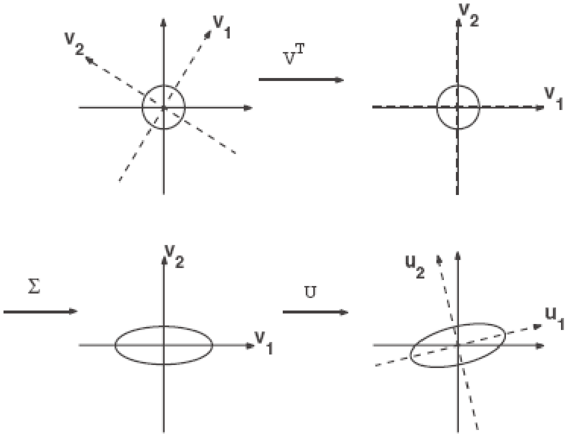
$$A = U \Sigma V^T = \sum_{i=1}^{r} \sigma_i u_i v_i^T,$$

  where $U \in \mathbb{R}^{m \times m}, V \in \mathbb{R}^{n \times n}$ are unitary matrices and $\Sigma = \begin{bmatrix} \Sigma_r & 0 \\ 0 & 0 \end{bmatrix} \in \mathbb{R}^{m \times n}$ is a diagonal matrix with $\Sigma_r = \mathrm{diag}(\sigma_1, \cdots, \sigma_r)$ and $r \leq \min(m, n) = rank(A)$.

- The positive numbers $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r > 0$ are called the *singular values* of A and are uniquely determined.

Geometric Interpretation of the SVD [6]

- The image of the unit sphere under a matrix is a hyper-ellipse.



[6] diagram: `http://people.sc.fsu.edu/~jburkardt/latex/fsu_2006/svd.png`

Singular Values

- The number of non-zero singular values r is the rank of the matrix.
- The singular values are also related to a class of matrix norms:
- *Spectral norm* (induced by the Euclidean vector norm):

$$\|A\|_2 = \sqrt{\lambda_{max}(A^T A)} = \sigma_{max}(A).$$

- *Frobenius norm*:

$$\|A\|_F = \sqrt{\sum_{i=1}^{m} \sum_{j=1}^{n} |a_{ij}|^2} = \sqrt{\text{trace}(A^T A)} = \sqrt{\sum_{i=1}^{r} \sigma_i^2}$$

Singular Values and Eigenvalues

- The singular value decomposition of $A = U\Sigma V^T \in \mathbb{R}^{m\times n}$ is related to the eigenvalue decomposition of symmetric matrix $A^TA \in \mathbb{R}^{n\times n}$:

$$A^TA = (V\Sigma U^T)(U\Sigma V^T) = V\Sigma^2 V^T = V\Lambda V^T$$

$$\text{and similarly} \quad AA^T = U\Sigma U^T = U\bar{\Lambda}U^T.$$

$$\implies \sigma_i(A) = \sqrt{\lambda_i(A^TA)} = \sqrt{\bar{\lambda}_i(AA^T)}, \quad i = 1, \cdots, r$$

- The singular values of A are the square roots of the eigenvalues of both A^TA and AA^T.
- The right singular vectors, V, of A are the eigenvectors of A^TA
- The left singular vectors, U, of A are the eigenvectors of AA^T

Singular Vectors

- The columns of $U = [u_1, \cdots, u_m]$ and $V = [v_1, \cdots, v_n]$ are called the *left* and *right singular vectors* of A, respectively.
- The columns of U and V form an orthonormal set of vectors, which can be regarded as orthonormal bases.
- The singular vectors provide a convenient way to represent the bases of the fundamental subspaces of matrix A.

  If $A \in \mathbb{R}^{m \times n}$ is a matrix of rank r, then

$$\begin{aligned} \mathcal{R}(A) &= \operatorname{span}\{u_1, \cdots, u_r\} \\ \mathcal{N}(A^T) &= \operatorname{span}\{u_{r+1}, \cdots, u_m\} \\ \mathcal{R}(A^T) &= \operatorname{span}\{v_1, \cdots, v_r\} \\ \mathcal{N}(A) &= \operatorname{span}\{v_{r+1}, \cdots, v_n\} \end{aligned}$$

Low-rank Approximation via the SVD

- From the definition of the SVD, any matrix $A \in \mathbb{R}^{m \times n}$ of rank r can be written as a sum of *rank-one matrices*:

$$A = U\Sigma V^T = \sum_{i=1}^{r} \sigma_i u_i v_i^T,$$

- **Theorem (low-rank approximation)**: Let $A \in \mathbb{R}^{m \times n}$ and $0 \leq k \leq r$, then

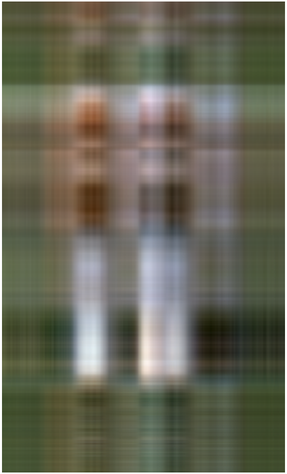
$$\min_{B\,:\,rank(B)=k} \|A - B\|_2 = \sigma_{k+1}$$

where the minimum is attained by $B^\star = A_k = \sum_{i=1}^{k} \sigma_i u_i v_i^T$.

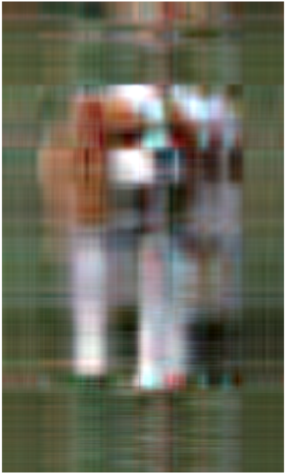
  - Application: image compression

Image Compression using SVD

(A) rank-1 approximation (18,456 bytes ≈ 0.6%)

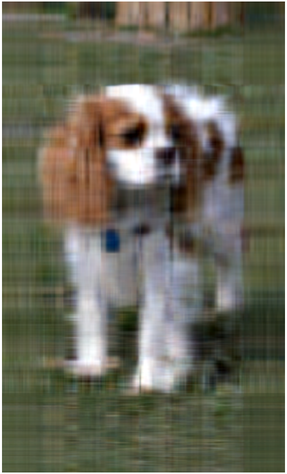


(B) rank-2 approximation (36,912 bytes ≈ 1.1%)

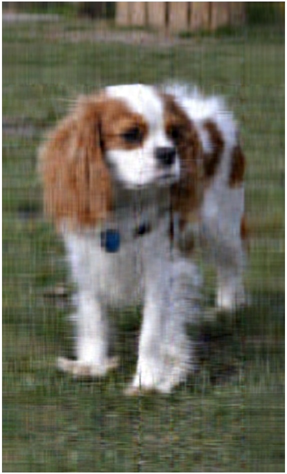


(C) rank-5 approximation (92,280 bytes ≈ 2.8%)

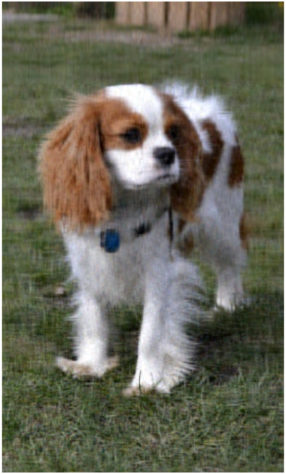
Image Compression using SVD



(A) rank-10 approximation (184,560 bytes ≈ 5.6%)



(B) rank-20 approximation (369,120 bytes ≈ 11.1%)



(C) rank-40 approximation (738,240 bytes ≈ 22.3%)

Image Compression using SVD

(a) original image
(3,317,760 bytes)

Questions?

Mini-Quiz 15

The SVD is useful in many linear algebra proofs, since it decomposes a matrix into orthogonal and diagonal factors which have nice properties.

- Let $A \in \mathbb{R}^{m \times n} = U\Sigma V^T$. Show that

$$\|A\|_2 = \sigma_{max}(A) = \max_{\|x\|=1,\ \|y\|=1} y^T A x.$$

  Hint: Use the SVD and unitary invariance of the Euclidean vector norm ($\|Ux\| = \|x\|$).

- *Polar decomposition*:
  Show how any square matrix $A \in \mathbb{R}^{n \times n} = U\Sigma V^T$ can be written as

$$A = QS, \quad Q \text{ orthogonal, } S \text{ symmetric positive semi-definite.}$$

  Hint: $V\Sigma V^T$ is symmetric positive semi-definite.

Solution (Q15)

- Let $A = U\Sigma V^T$. Show that $\|A\|_2 = \sigma_{max}(A) = \max\limits_{\|x\|=1,\|y\|=1} y^T Ax$.

  Solution:

$$\begin{aligned}
\max_{\|x\|=1,\|y\|=1} y^T Ax &= \max_{\|x\|=1,\ \|y\|=1} y^T (U\Sigma V^T)x \\
&= \max_{\|V^T x\|=1,\ \|U^T y\|=1} (y^T U)\Sigma(V^T x) \\
&= \max_{\|\bar{x}\|=1,\ \|\bar{y}\|=1} \bar{y}^T \Sigma \bar{x} \\
&= \max_i \Sigma_{ii} \\
&= \sigma_{max}(A)
\end{aligned}$$

- Show how any square matrix $A = U\Sigma V^T$ can be written as

  $A = QS$, Q orthogonal, S symmetric positive semi-definite.

  Solution:

$$A = U\Sigma V^T = UI_n\Sigma V^T = U(V^T V)\Sigma V^T = (UV^T)(V\Sigma V^T)$$

$$A = QS, \quad \text{where } Q = UV^T, \text{ and } S = V\Sigma V^T.$$

The End!

References and Acknowledgements

1. *Numerical Linear Algebra* by Lloyd N. Trefethen & David Bau III.
2. *Matrix Computations* by Gene H. Golub & Charles F. Van Load.
3. *Linear Algebra and its Applications* by Gilbert Strang.
4. *Linear Algebra with Applications.* by Otto Bretscher.
5. `www.Wikipedia.org` and `Mathworld.Wolfram.com`

- Thanks to former ICME Refresher Course instructors Yuekai Sun, Nicole Taheri, and Milinda Lakkam whose slides served as useful references in creating this presentation.
- Thanks to Carlos Sing-Long and Anil Damle for their feedback on this presentation.