FUNDAMENTALS OF LEARNING AND INFORMATION PROCESSING

SESSION 02: LINEAR ALGEBRA (II)

Dr Gang Li

Deakin University, Geelong, Australia

2018-11-06

Table of Content

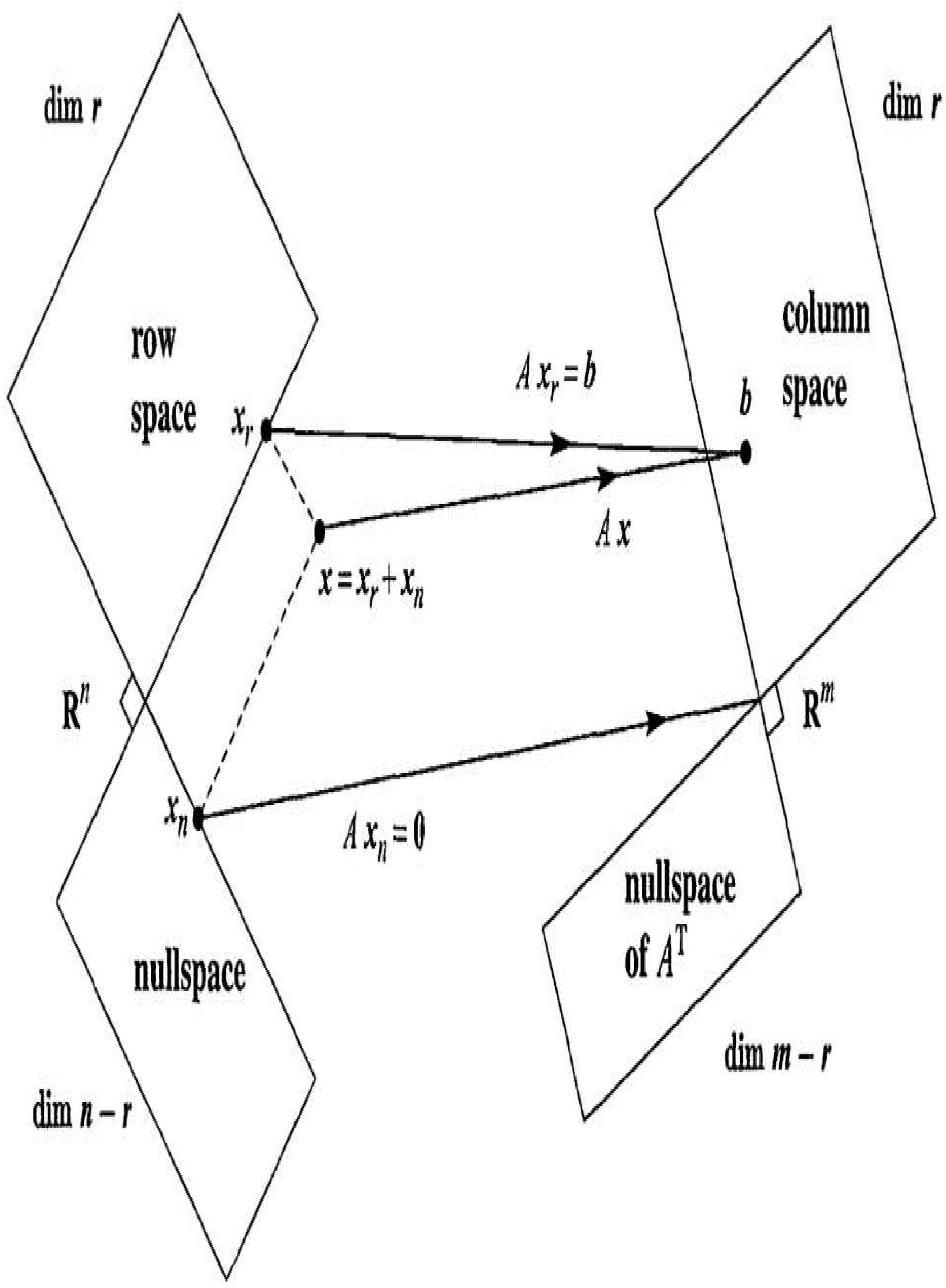
Orthogonality

Defn
- Vectors x and y are orthogonal, if and only if their *inner product* $x \cdot y = x^T y = 0$;
- Subspace S is orthogonal to subspace T, if and only if every vector in S is orthogonal to every vector in T.
- If S contains all vectors orthogonal to T, then S is orthogonal complement of V, denoted by $V^{\perp}$.

- Let q_j, $j = 1, \ldots, n$ be orthogonal, i.e., $q_i^T q_j = 0$ when $i \neq j$. Then they are linearly independent. (please prove)
- For a matrix A, its *row space* $\mathscr{R}(A^T)$ is orthogonal to its *nullspace* $\mathscr{N}(A)$, because $Ax = 0$, namely $\begin{pmatrix} r_1 \\ r_2 \\ \cdots \\ r_m \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \cdots \\ x_m \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ \cdots \\ 0 \end{pmatrix}$
  - A *nullspace* $\mathscr{N}(A)$ contains all vectors perpendicular to ($\perp$) *row space* $\mathscr{R}(A^T)$.
- Similarly, the *left nullspace* $\mathscr{N}(A^T)$ contains all vectors perpendicular to ($\perp$) *column space* $\mathscr{R}(A)$.

Orthogonality in Subspaces

Thm The *nullspace* is the *orthogonal complement* of the *row space* in $\mathscr{R}^n$. The *left nullspace* is the *orthogonal complement* of the *column space* in $\mathscr{R}^m$.



- The *row space* and the *column space* share the same dimension r (the rank)
- The nullspace component goes to zero: $Ax_n = 0$
- The row space component goes to the column space: $Ax_r = Ax$
- For a general vector $x = x_r + x_n$, it has a *row space* component x_r and a *nullspace* component x_n. $Ax = Ax_n + Ax_r = Ax$

Orthogonality

Defn Let the set of orthogonal vectors q_j, $j = 1, \ldots, m$ in $\mathbb{R}^m$ be normalized, $\| q \| = 1$. Then they are orthonomal, and constitute an orthonomal basis in $\mathbb{R}^m$.

- A matrix $Q = [q_1, q_2, \ldots, q_m] \in \mathbb{R}^{m\times m}$ with orthonormal columns is called an orthogonal matrix, which has a rank m.
- Properties:
  - The inverse of an orthogonal matrix Q is $Q^{-1} = Q^T$
  - The Euclidean distance of a vector is *invariant* under an orthogonal transformation Q: $\| Qx \|^2 = (Qx)^T(Qx) = x^T x = \| x \|^2$.
  - The product of two orthogonal matrices Q and P is orthogonal: $X^T X = (PQ)^T PQ = Q^T P^T PQ = Q^T Q = I$

Orthogonality

Exercises:

- Let $A = \begin{bmatrix} 1 & 0 & 2 \\ 1 & 1 & 4 \end{bmatrix}$
  - Find a basis for the *nullspace* $\mathcal{N}(A)$ and verify that it is orthogonal to the row space.
  - Give $x = (3,3,3)$, split into a *row space* component x_r, and a *nullspace* component x_n.

Projection

Defn Given a point b, find a point p on a subspace S of $\mathbb{R}^n$ that is closest to b, this point p is the projection of b onto the subspace S.

- The line from b to the closest point $p = \varpi a$ is perpendicular to the vector a:

$$(b - \varpi a) \perp a, \quad \text{or} \quad a^T(b - \varpi a) = 0 \quad \text{so} \quad \varpi = \frac{a^T b}{a^T a}$$

- Here ϖ is a scale, so the projection can be written with a slight twist:

$$p = \varpi a = a\varpi = a\frac{a^T b}{a^T a} = \frac{aa^T}{a^T a} b = Pb$$

  where $P = \frac{aa^T}{a^T a}$ is the projection matrix.
  - P is the matrix that multiplies b and produces p on a.
  - P is *symmetric*, with rank $r = 1$
  - $P^2 = P$

Least Squares Approximations

LSA For $Ax = b$, when b is not on the *column space* $\mathscr{C}(A)$ of A, we can project b to the *column space*. Let $p = A\bar{x}$ be the projection of b on to the *column space*. The *error vector* $b - A\bar{x}$ must be perpendicular to the column space $\mathscr{C}(A)$.

- The *error vector* must be perpendicular to every column of A:

$$\left\{\begin{array}{ll} a_1^T(b-A\bar{x}) & =0 \\ a_2^T(b-A\bar{x}) & =0 \\ \cdots & \\ a_n^T(b-A\bar{x}) & =0 \end{array}\right. \quad \text{or} \quad \begin{bmatrix} a_1^T \\ \vdots \\ a_n^T \end{bmatrix} \begin{bmatrix} b - A\bar{x} \end{bmatrix} = 0 \quad \text{or} \quad A^T(b - A\bar{x}) = 0$$

- Thus we have the least squares form:

$$A^T A\bar{x} = A^T b$$

Least Squares Approximations

LSA For the least squares form: $A^T A\bar{x} = A^T b$, if $A^T A$ is invertible, we have the least squares approximation $\bar{x} = (A^T A)^{-1}A^T b$. The projection of b to the column space of A is therefore $p = A\bar{x} = A(A^T A)^{-1}A^T b = Pb$, where $P = A(A^T A)^{-1}A^T$ is the *projection matrix*.

- $A^T A$ has the same *nullspace* as A.
  - If $Ax = 0$ then $A^T Ax = 0$, namely vectors x in the *nullspace* of A are also in the *nullspace* of $A^T A$;
  - Suppose $A^T Ax = 0$ and take the inner product with x:

    $$x^T A^T Ax = 0, \quad \text{or} \quad \| Ax \|^2 = 0 \quad \text{or} \quad Ax = 0$$

    Thus x is in the *nullspace* of A.
- If A has linearly independent columns, then $A^T A$ is *square*, *symmetric* and *invertible*.
  - Suppose $A^T Ax = 0$ and take the inner product with x:

    $$x^T A^T Ax = 0, \quad \text{or} \quad \| Ax \|^2 = 0 \quad \text{or} \quad Ax = 0$$

    As A has linearly independent columns, $x = 0$. So $A^T A$ is *invertible*.
- The projection matrix $P = A(A^T A)^{-1}A^T$ has two basic properties:
  - It equals its square: $P^2 = P$
  - It equals its transpose: $P^T = P$
  - Any symmetric matrix with $P^2 = P$ represents a projection:
    - Like any other matrix, P takes every vector b into its column space: Pb is a weighted combination of the columns.
    - On the other hand, the error vector $b - Pb$ is *orthogonal* to the space: $(b - Pb)^T Pc = b^T(I - P)^T Pc = b^T(P - P^2)c = 0$

Least Squares Approximations

Exercises:

- Let $Ax = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 1 \end{bmatrix}$ $x = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}$
  - Solve $Ax = b$ by L.S.A.
  - Find $p = A\bar{x}$, and verify that the error $b - p$ is perpendicular to the columns of A.
- Let $A = \begin{bmatrix} 1 & 1 \\ 2 & -1 \\ -2 & 4 \end{bmatrix}$, and $b = \begin{bmatrix} 1 \\ 2 \\ 7 \end{bmatrix}$
  - Find the projection of b onto the column space of A.
  - Split b into $p + q$, with p in the column space and q perpendicular to that space.
  - Which of the four subspaces contains q?

Solution of $Qx = b$

$Qx = b$

For $Qx = b$, with Q be a matrix with orthonormal columns. Two cases exist:
- Q is an orthogonal matrix, which is *square*;
- Q is a rectangular matrix with orthonormal columns;

- For the *orthogonal matrix* Q, we have $Q^{-1} = Q^T$, hence $x = Q^T b$.
  - If we have an orthogonal basis $q1, \cdots, q_n$, for a given vector b, it can be combined by $b = x_1 q_1 + \cdots + x_n q_n$, namely $b = Qx$. We have $x = Q^T b$
  - Any permutation matrix P is an orthogonal matrix. It is unit, and 1 appears in different place in each column. $P^{-1} = P^T$.
- For the m by n rectangular matrix Q with *orthogonal* columns, we still have $Q^T Q = I$ (*left inverse*). LSA gives us $\bar{x} = (Q^T Q)^{-1} Q^T b = Q^T b$
  - The *projection matrix* is then $P = Q(QQ^T)^{-1} Q^T = QQ^T$, which is a m by m matrix.

Gram-Schmidt Orthogonalization

G.S.O.

The G.S.O. process starts with indepdendent vectors $a_1, \cdots, a_n$ and ends with orthonormal vectors $q_1, \cdots, q_n$. At step j it subtracts from a_j its components in the directions that are already settled: $a'_j = a_j - (q_1^T a_j)q_1 - \cdots - (q_{j-1}^T a_j)q_{j-1}$. Then q_j is the unit vector $a'_j / \| a'_j \|$.

Exercise:

- Apply the G.S.O. process to

$$a = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \qquad b = \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix} \qquad c = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$$

  and write the result in the form $A = QR$, where Q is the same size m by n matrix as A, and R is a square matrix n by n.

QR Factorization

$A = QR$

In the G.S.O. process, both the A and the Q are m by n, when the vectors are in the m-dimensional space, there is an *upper triangular* matrix R that connects them:

$$\begin{bmatrix} a & b & c \end{bmatrix} = \begin{bmatrix} q_1 & q_2 & q_3 \end{bmatrix} \begin{bmatrix} q_1^T a & q_1^T b & q_1^T c \\ & q_2^T b & q_2^T c \\ & & q_3^T c \end{bmatrix}$$

- Every m by n matrix A with linearly independent columns can be factored into $A = QR$, where Q is with orthonormal columns, R is upper triangular and invertible.
- $Ax = b$ then becomes $QRx = b$, as $A^T A = R^T Q^T Q R = R^T R$, the L.S.A. formula is then $R^T R \bar{x} = R^T Q^T b$, or $R\bar{x} = Q^T b$, where R is upper triangular.

QR Factorization

Exercises:

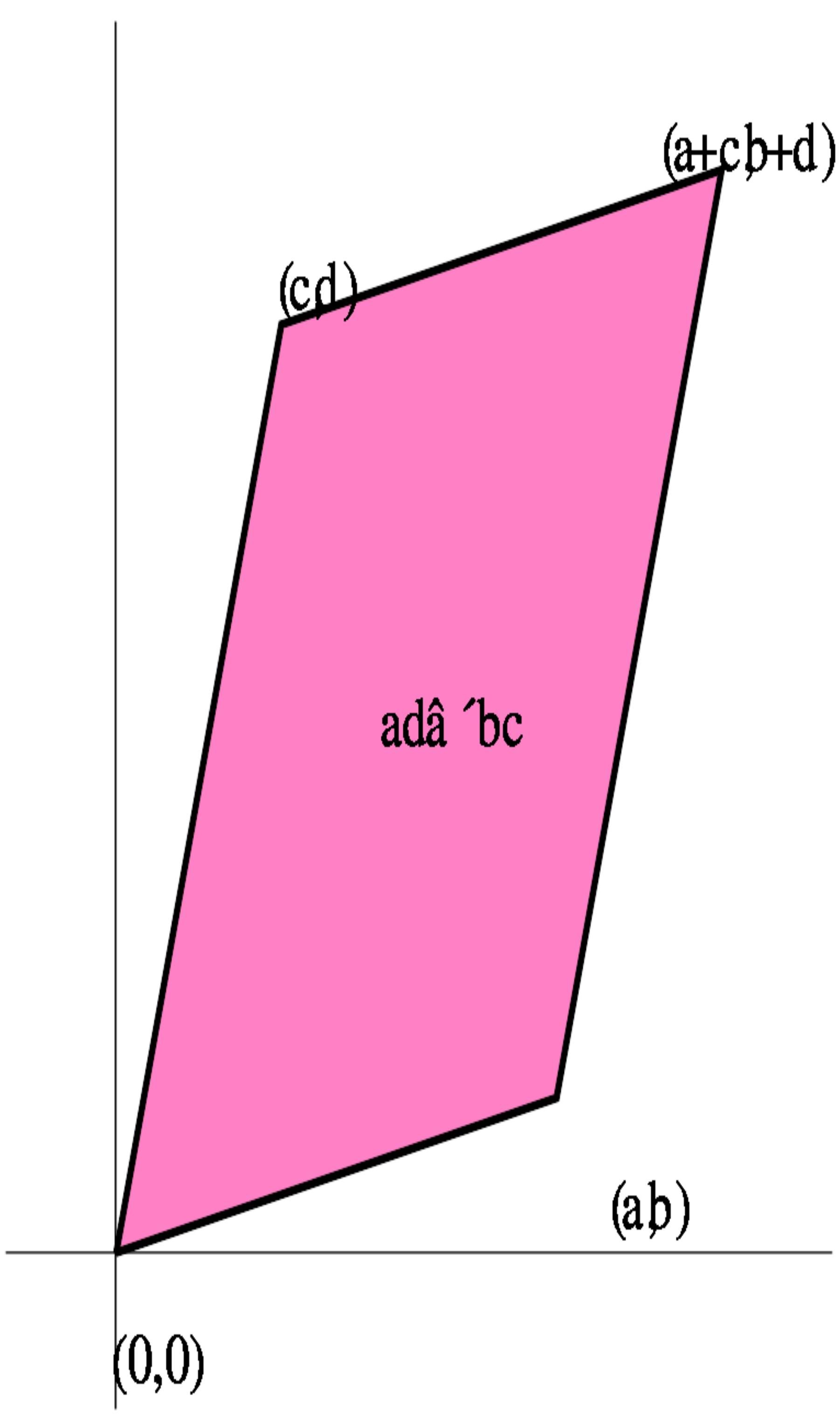
- Find an orthonormal set q_1, q_2 and q_3 for which q_1 and q_2 span the column space of

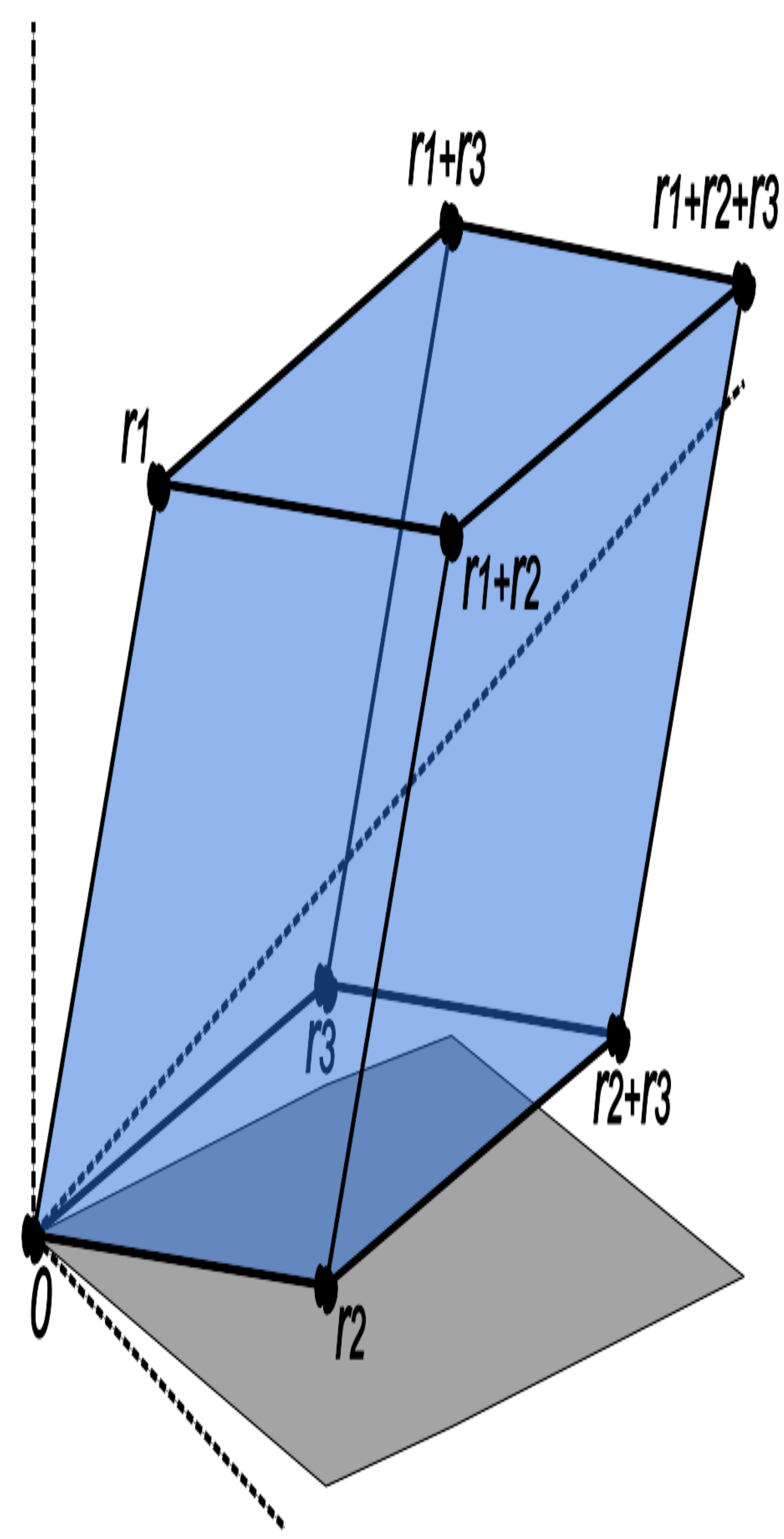
$$A = \begin{bmatrix} 1 & 1 \\ 2 & -1 \\ -2 & 4 \end{bmatrix}$$

  - Which fundamental subspace contains q_3?
  - What is the L.S.A. solution of $Ax = b$ if $b = [1, 2, 7]^T$?

Determinant

Defn *Determinant* is a value associated with a *square* matrix. A geometric interpretation can be given to the value of the determinant of a square matrix with real entries: the *absolute value* of the determinant gives the scale factor by which area or volume (or a higher dimensional analogue) is multiplied under the associated linear transformation, while its sign indicates whether the transformation preserves orientation.





Determinant

Rule 1 $det(I) = 1$ the unit box with a volume 1
Rule 2 Exchanging rows reverses the sign of $det(A)$.
Rule 3

Rule $3a$: $\begin{vmatrix} ta & tb \\ c & d \end{vmatrix} = t\begin{vmatrix} a & b \\ c & d \end{vmatrix}$ Rule $3b$: $\begin{vmatrix} a+a' & b+b' \\ c & d \end{vmatrix} = \begin{vmatrix} a & b \\ c & d \end{vmatrix} + \begin{vmatrix} a' & b' \\ c & d \end{vmatrix}$

Derived Rules

- Matrix with equal rows, the determinate is 0: $\begin{vmatrix} a & b \\ a & b \end{vmatrix} = 0$
- Subtract $l \times row_j$ from row_k, the determinant doesn't change: $\begin{vmatrix} a & b \\ c-la & d-lb \end{vmatrix} = \begin{vmatrix} a & b \\ c & d \end{vmatrix} + l\begin{vmatrix} a & b \\ -a & -b \end{vmatrix} = \begin{vmatrix} a & b \\ c & d \end{vmatrix}$
- The determinate for a Matrix with a row of 0 is 0.
- The determinant for an upper triangular matrix: $\begin{vmatrix} d_1 & * & * & \cdots \\ 0 & d_2 & * & \cdots \\ 0 & 0 & \cdots & \cdots \\ 0 & 0 & \cdots & d_n \end{vmatrix} = d_1 \times d_2 \times \cdots d_n$
- The determinant of a singular matrix is 0; if the determinate is not 0, then the matrix is invertible.
- $det(AB) = det(A)det(B)$, $det(A^{-1}) = \frac{1}{det(A)}$
- $det(A^T) = det(A)$

Determinant

Exercises:

- Use the row operations to verify

$$det\begin{vmatrix} 1 & a & a^2 \\ 1 & b & b^2 \\ 1 & c & c^2 \end{vmatrix} = (b-a)(c-a)(c-b)$$

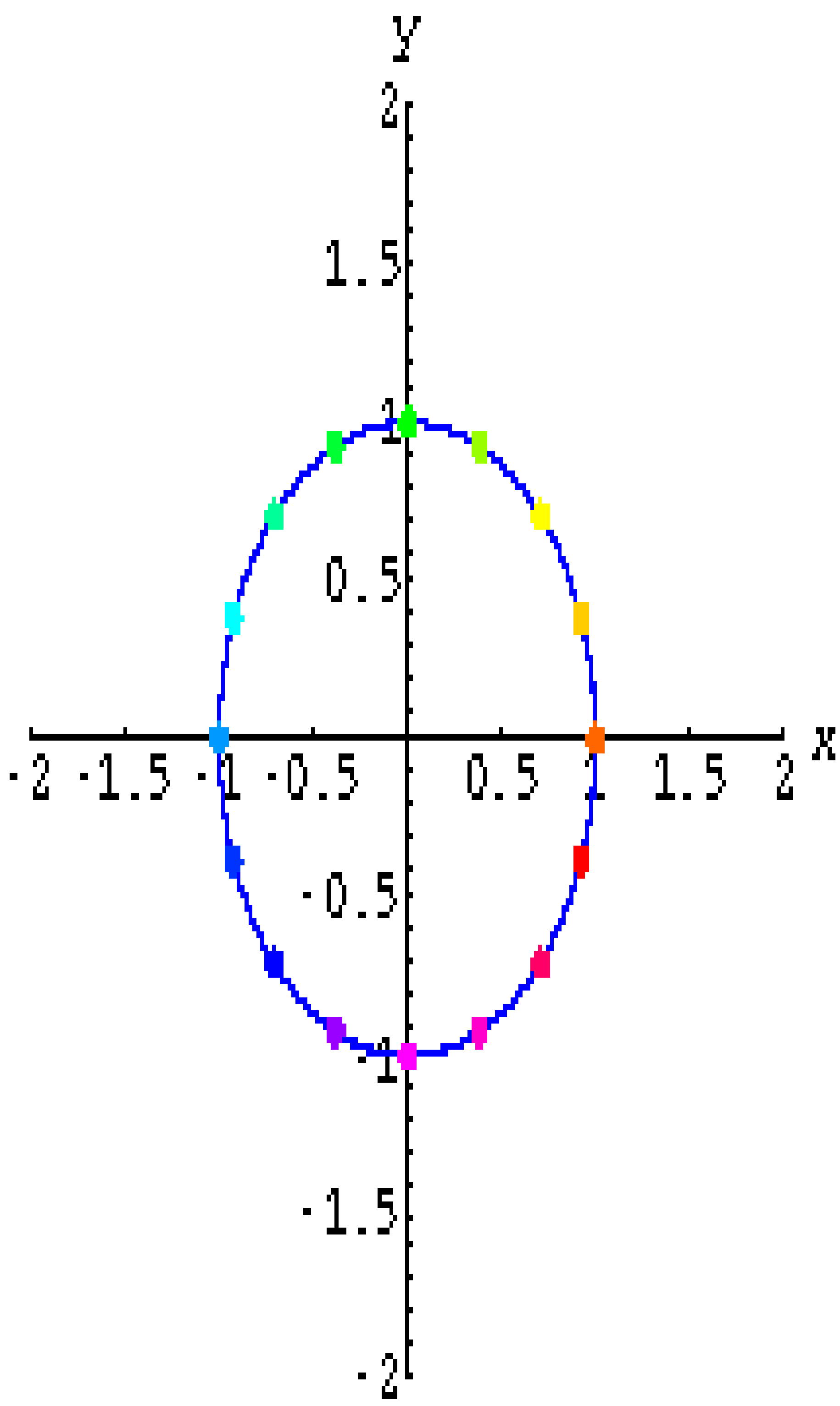
- Find the determinants by *Gaussian Elimination*:

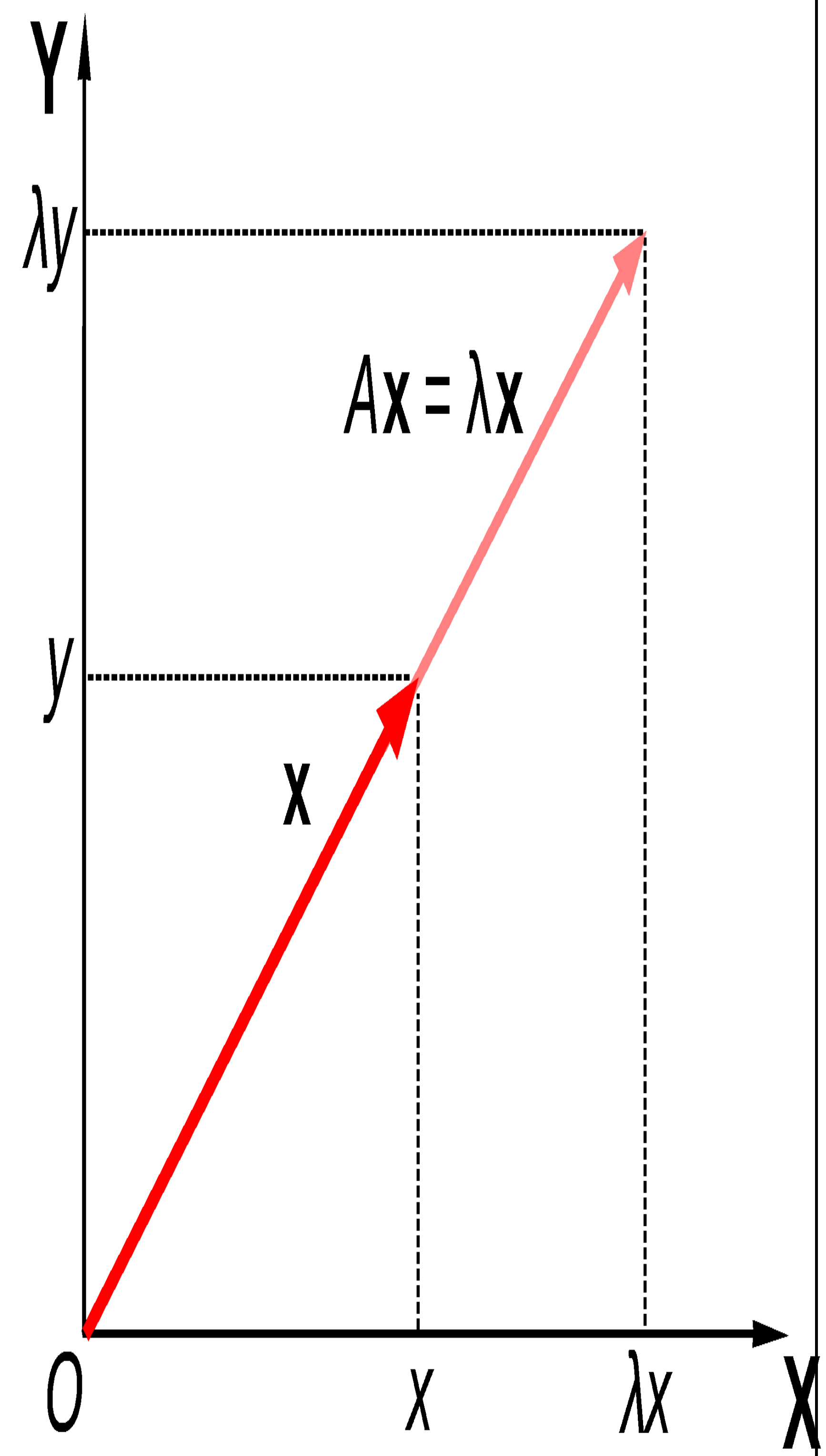
$$\begin{vmatrix} 11 & 12 & 13 & 14 \\ 21 & 22 & 23 & 24 \\ 31 & 32 & 33 & 34 \\ 41 & 42 & 43 & 44 \end{vmatrix} \quad \text{and} \quad \begin{vmatrix} 1 & t & t^2 & t^3 \\ t & 1 & t & t^2 \\ t^2 & t & 1 & t \\ t^3 & t^2 & t & 1 \end{vmatrix}$$

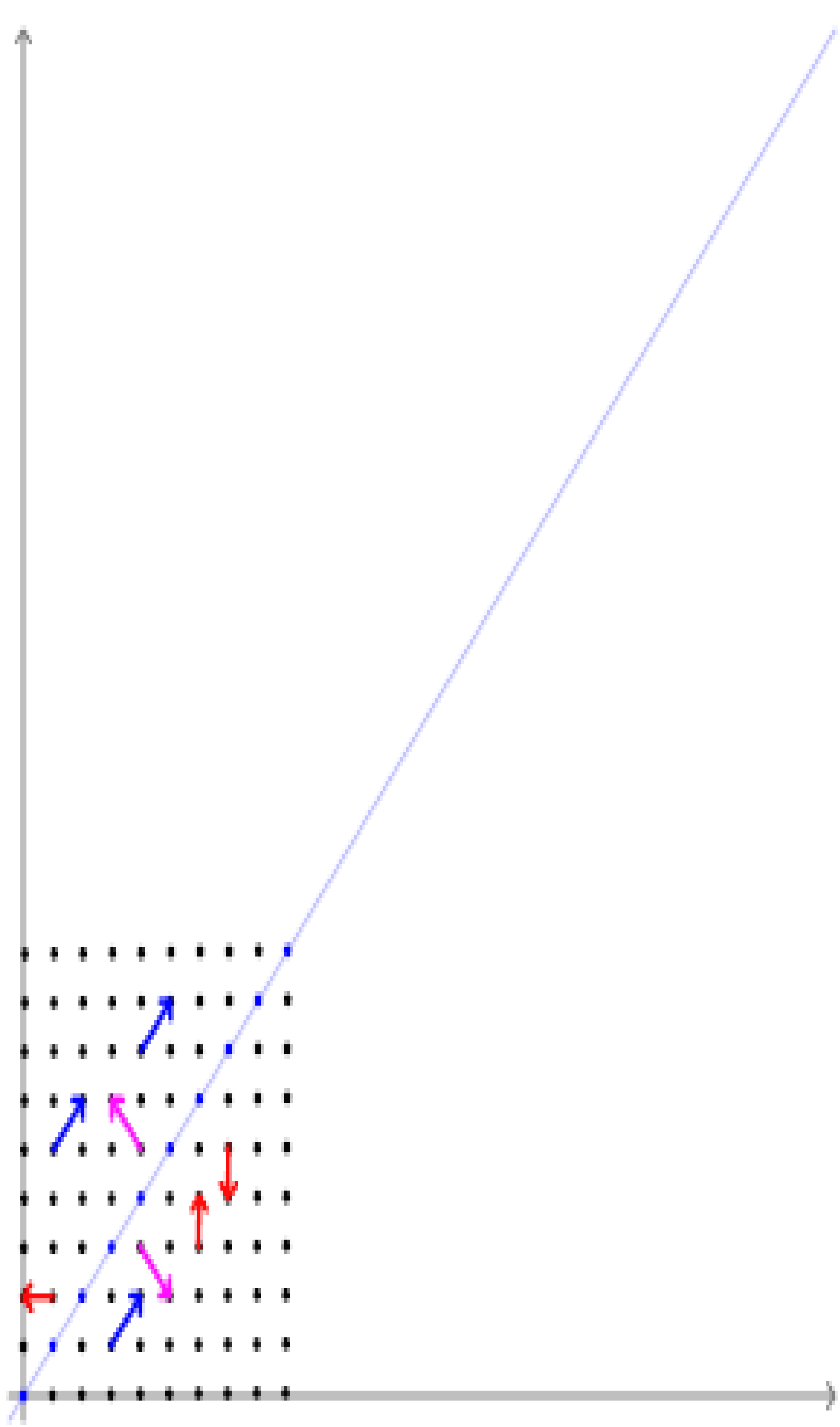
Eigenvectors and Eigenvalues

Defn: When we hit a vector x with a matrix A, but the matrix acts by stretching the vector x, not changing its direction, namely $Ax = \lambda x$, then x is an *eigenvector* of A, the stretching factor λ is the *eigenvalue*.

- When you *hit* a vector $x \in \mathbb{R}^m$ with a matrix $A \in \mathbb{R}^{m \times m}$, you get another vector Ax.







- Example:

$$A = \begin{bmatrix} 1 & 2 \\ 0 & -2 \end{bmatrix} \qquad x = \begin{bmatrix} -2 \\ 3 \end{bmatrix} \qquad y = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

If x and y are hit by matrix A,

Eigenvectors and Eigenvalues

Defn

In order to find the *eigenvalues* and *eigenvectors*, we can find the solutions to equation $Ax = \lambda x$:

- The vector x is in the nullspace of $A - \lambda I$
- The number λ is chosen so that $A - \lambda I$ has a nullspace

- The number λ is an eigenvalue of A if and only if

$$det(A - \lambda I) = 0$$

- The *sum* of the n *eigenvalues* equals the *trace* of A, namely the sum of the n diagonal entries:

$$\sum_{i=1}^{n} \lambda_i = \lambda_1 + \lambda_2 + \cdots + \lambda_n = a_{11} + \cdots + a_{nn}$$

- The *product* of the n *eigenvalues* equals the *determinant* of A:

$$\prod_{i=1}^{n} \lambda_i = \lambda_1 \times \cdots \times \lambda_n = det(A)$$

EVD: Eigenvectors Decomposition

Defn

Suppose matrix A is a n by n matrix with n linearly independent eigenvectors. Then if those vectors are chosen to be the columns of a matrix S, it follows that $S^{-1}AS$ is a diagonal matrix Λ, with the eigenvalues of A along its diagonal:

$$AS = S\Lambda \qquad \text{or} \qquad S^{-1}AS = \Lambda = \begin{bmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n \end{bmatrix} \qquad \text{or} \quad A = S\Lambda S^{-1}$$

- Proof: put the eigenvectors x_i in the columns of S, and compute the product AS one column at a time:

$$AS = A\begin{bmatrix} \uparrow & \uparrow & & \uparrow \\ x_1 & x_2 & \cdots & x_n \\ \downarrow & \downarrow & & \downarrow \end{bmatrix} = \begin{bmatrix} \uparrow & \uparrow & & \uparrow \\ \lambda_1 x_1 & \lambda_2 x_2 & \cdots & \lambda_n x_n \\ \downarrow & \downarrow & & \downarrow \end{bmatrix} = \begin{bmatrix} \uparrow & \uparrow & & \uparrow \\ x_1 & x_2 & \cdots & x_n \\ \downarrow & \downarrow & & \downarrow \end{bmatrix}\begin{bmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n \end{bmatrix}$$

Application of Eigenvectors Decomposition

Prob

Fibonacci sequence $F_{k+2} = F_{k+1} + F_k$:

$$0, 1, 1, 2, 3, 5, 8, 13, \cdots$$

How could we find the 1000th Fibonacci number?

- If $u_k = \begin{bmatrix} F_{k+1} \\ F_k \end{bmatrix}$, then $u_{k+1} = \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix} u_k$. Then, what is $u_{1000} = A^{1000}u_0$?
- If A can be diagonalized, $A = S\Lambda S^{-1}$, then

$$u_k = A^k u_0 = (S\Lambda S^{-1})(S\Lambda S^{-1})\cdots(S\Lambda S^{-1})u_0 = S\Lambda^k S^{-1}u_0$$

- Let $c = S^{-1}u_0$, as the columns of S are the eigenvectors of A, the solutions becomes

$$u_k = S\Lambda^k c = \begin{bmatrix} \uparrow & & \uparrow \\ x_1 & \cdots & x_n \\ \downarrow & & \downarrow \end{bmatrix}\begin{bmatrix} \lambda_1^k & & \\ & \ddots & \\ & & \lambda_n^k \end{bmatrix}\begin{bmatrix} c_1 \\ \vdots \\ c_n \end{bmatrix} = c_1\lambda_1^k x_1 + \cdots + c_n\lambda_n^k x_n$$

Application of Eigenvectors Decomposition

Prob *Fibonacci sequence* $F_{k+2} = F_{k+1} + F_k$:

$$0, 1, 1, 2, 3, 5, 8, 13, \cdots$$

How could we find the 1000th Fibonacci number?

- For $A = \left[\begin{smallmatrix} 1 & 1 \\ 1 & 0 \end{smallmatrix}\right]$, the determinant is $\lambda^2 - \lambda - 1$, and two eigenvalues are $\lambda_1 = \frac{1+\sqrt{5}}{2}$ and $\lambda_2 = \frac{1-\sqrt{5}}{2}$, corresponding to eigenvectors: $x_1 = \left[\begin{smallmatrix} \lambda_1 \\ 1 \end{smallmatrix}\right]$ and $x_2 = \left[\begin{smallmatrix} \lambda_2 \\ 1 \end{smallmatrix}\right]$
- As $F_0 = 0$ and $F_1 = 1$,

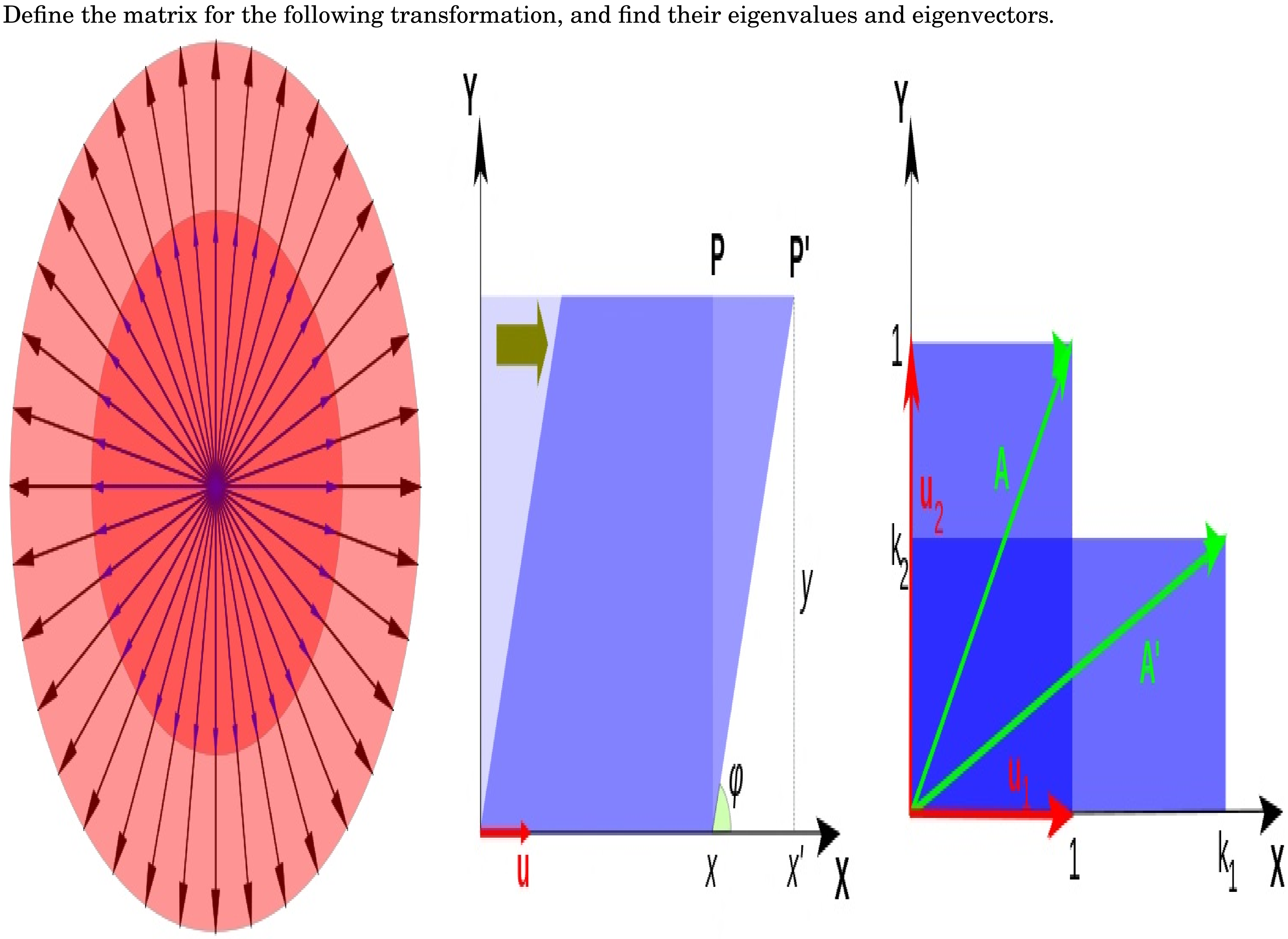
$$c = S^{-1} u_0 = \begin{bmatrix} \lambda_1 & \lambda_2 \\ 1 & 1 \end{bmatrix}^{-1} \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} \frac{1}{\lambda_1 - \lambda_2} \\ -\frac{1}{\lambda_1 - \lambda_2} \end{bmatrix}$$

  Hence we have $u_k = c_1 \lambda_1^k x_1 + c_2 \lambda_2^k x_2$.
- F_k is the second component of u_k, $F_k = c_1 \lambda_1^k + c_2 \lambda_2^k$. As λ_2 is less than 1, so F_k is dominated by the first term $\frac{1}{\sqrt{5}}(\frac{1+\sqrt{5}}{2})^{1000}$.

Application of Eigenvectors Decomposition

Exercises:

- Define the matrix for the following transformation, and find their eigenvalues and eigenvectors.



- Suppose we shift the matrix $A = \left[\begin{smallmatrix} 1 & -1 \\ 2 & 4 \end{smallmatrix}\right]$ by subtracting $7I$: $B = \left[\begin{smallmatrix} -6 & -1 \\ 2 & -3 \end{smallmatrix}\right]$. What are the eigenvalues of eigenvectors of B, and how are they related to those of A?

SED: Symmetric Eigenvectors Decomposition

Defn For any symmetric n by n matrix A, it can be diagonalized by an orthogonal matrix Q, whose columns are chosen to be the eigenvectors of A:

$$A = Q\Lambda Q^T \qquad \text{with} \qquad Q^T A Q = \Lambda = \begin{bmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n \end{bmatrix}$$

- For a symmetric matrix A with $A = A^T$, its eigenvalues are *real* values.

  proof $Ax = \lambda x \Rightarrow \bar{A}\bar{x} = \bar{\lambda}\bar{x} \Rightarrow \bar{x}^T A^T = \bar{x}^T \bar{\lambda} \Rightarrow \bar{x}^T A x = \bar{x}^T A^T x = \bar{x}^T \bar{\lambda} x$
  also we have $\bar{x}^T A x = \bar{x}^T \lambda x = \lambda \bar{x}^T x$,
  so we have $\lambda = \bar{\lambda}$
- For a symmetric matrix A with $A = A^T$, its eigenvectors from different eigenvalues are orthogonal to each other.
- Strictly speaking, this SED has been proven only when the eigenvalues of A are distinct. Nevertheless it is true that *even with repeated eigenvalues, a symmetric matrix still has a complete set of orthonormal eigenvectors*.

Spectral Theorem

Defn Every real symmetric A can be diagonalized by an orthogonal matrix Q, whose columns contain a complete set of *orthonormal* eigenvectors. If we multiply columns by rows, the matrix A becomes a combination of one-dimensional projections, which are the special matrix xx^T of rank one:

$$A = Q\Lambda Q^T = \begin{bmatrix} \uparrow & \uparrow & & \uparrow \\ x_1 & x_2 & \cdots & x_n \\ \downarrow & \downarrow & & \downarrow \end{bmatrix} \begin{bmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n \end{bmatrix} \begin{bmatrix} \leftarrow & x_1^T & \rightarrow \\ \leftarrow & x_2^T & \rightarrow \\ \leftarrow & \cdots & \rightarrow \\ \leftarrow & x_n^T & \rightarrow \end{bmatrix} = \lambda_1 x_1 x_1^T + \cdots \lambda_n x_n x_n^T$$

- As the projection matrix $P = \frac{aa^T}{a^Ta}$ projects a vector b to a vector a, $x_i x_i^T$ represents the projections onto the vector x_i. Hence *matrix hit* Ab can be represented by $Ab = \lambda_1 x_1 x_1^T b + \cdots + \lambda_n x_n x_n^T b$, where $x_i x_i^T b$ is the projection of b on x_i.
- A symmetric matrix can be completely represented by:

  eigenvectors determine the direction on which the matrix transformation doesn't change;
  eigenvalues control the weight of contribution along its corresponding eigenvectors.

Applications of Spectral Theorem

Prob Given a vector $b = \left[\begin{smallmatrix}1\\1\end{smallmatrix}\right]$, what is the vector after hitting it by a matrix $A = \left[\begin{smallmatrix}1&3\\3&1\end{smallmatrix}\right]$

- It can be calculated directly as $Ab = \left[\begin{smallmatrix}4\\4\end{smallmatrix}\right]$. However, we will try the spectral theorem.
- For the matrix A, the eigenvalues are $\lambda_1 = 4$ and $\lambda_2 = -2$, the eigenvectors are $x_1 = \begin{bmatrix}\frac{1}{\sqrt{2}}\\ \frac{1}{\sqrt{2}}\end{bmatrix}$ and $x_2 = \begin{bmatrix}\frac{1}{\sqrt{2}}\\ -\frac{1}{\sqrt{2}}\end{bmatrix}$. The spectral theorem tells us that

$$Ab = \lambda_1 x_1 x_1^T b + \lambda_2 x_2 x_2^T b = 4\begin{bmatrix}\frac{1}{\sqrt{2}}\\ \frac{1}{\sqrt{2}}\end{bmatrix}\left[\begin{smallmatrix}\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}}\end{smallmatrix}\right]\left[\begin{smallmatrix}1\\1\end{smallmatrix}\right] - 2\begin{bmatrix}\frac{1}{\sqrt{2}}\\ -\frac{1}{\sqrt{2}}\end{bmatrix}\left[\begin{smallmatrix}\frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}}\end{smallmatrix}\right]\left[\begin{smallmatrix}1\\1\end{smallmatrix}\right] = \left[\begin{smallmatrix}4\\4\end{smallmatrix}\right]$$

  - This is also understandable, as b is along the direction of x_1, hence Ab will not change its direction.
  - b is perpendicular to x_2, hence, the component in the 2nd part is 0.
- For a vector $b = \left[\begin{smallmatrix}1\\2\end{smallmatrix}\right]$ The spectral theorem tells us that

$$Ab = \lambda_1 x_1 x_1^T b + \lambda_2 x_2 x_2^T b = 4\begin{bmatrix}\frac{1}{\sqrt{2}}\\ \frac{1}{\sqrt{2}}\end{bmatrix}\left[\begin{smallmatrix}\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}}\end{smallmatrix}\right]\left[\begin{smallmatrix}1\\2\end{smallmatrix}\right] - 2\begin{bmatrix}\frac{1}{\sqrt{2}}\\ -\frac{1}{\sqrt{2}}\end{bmatrix}\left[\begin{smallmatrix}\frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}}\end{smallmatrix}\right]\left[\begin{smallmatrix}1\\2\end{smallmatrix}\right] = \left[\begin{smallmatrix}7\\5\end{smallmatrix}\right]$$

Applications of Spectral Theorem

Prob Every *quadratic* function in the n variables $x_1, x_2, \cdots, x_n$ can be expressed in the form $f(x) = x \cdot Hx = \sum_{i=1}^n \sum_{j=1}^n H_{ij} x_i x_j$, which involves n^2 terms, and the variables are typically coupled.

- when H is a diagonal matrix, the function can be simplified: $f(x) = \sum_{i=1}^n H_{ii} x_i^2$
- when H is a general matrix, the function can also be simplified by selecting a correct coordination system.

- Any vector $\begin{bmatrix}x_1\\x_2\\ \vdots\\ x_n\end{bmatrix}$ is implicitly in the standard basis $e^{(1)} = \begin{bmatrix}1\\0\\ \vdots\\0\end{bmatrix}$, $e^{(2)} = \begin{bmatrix}0\\1\\ \vdots\\0\end{bmatrix}$ etc.
- Any orthonormal set $u^{(1)}, \cdots, u^{(n)}$ forms an alternate basis, every vector x can then expressed as $x = \sum_{i=1}^n \alpha_i u^{(i)}$, with coefficients α_i as

$$\begin{bmatrix}\alpha_1\\ \vdots\\ \alpha_n\end{bmatrix} = \begin{bmatrix}u^{(1)} \cdot x\\ \vdots\\ u^{(n)} \cdot x\end{bmatrix} = U^T x$$

The key point is that $\{\alpha_1, \cdots, \alpha_n\}$ can be thought of as new variables representing the vector x. Specifically, $\{x_1, \cdots, x_n\}$ represent x in the standard basis $\{e^{(1)}, \cdots, e(n)\}$, while $\{\alpha_1, \cdots, \alpha_n\}$ represent x in the alternate basis $\{u^{(1)}, \cdots, u^{(n)}\}$.

Applications of Spectral Theorem

Prob

Every *quadratic* function in the n variables x_1, x_2, $\cdots$, x_n can be expressed in the form $f(x) = x \cdot Hx = \sum_{i=1}^n \sum_{j=1}^n H_{ij} x_i x_j$, which involves n^2 terms, and the variables are typically coupled.

- when H is a diagonal matrix, the function can be simplified: $f(x) = \sum_{i=1}^n H_{ii} x_i^2$
- when H is a general matrix, the function can also be simplified by selecting a correct coordination system.

- If $A \in R^{m \times n}$, then $y \cdot Ax = y^T Ax = (A^T y)^T x = (A^T y) \cdot x$ for all $x \in R^n$ and $y \in R^m$.
- Assuming $H \in R^{n \times n}$ is symmetric, it has a spectral decomposition $H = UDU^T$. Therefore, $x \cdot Hx = x \cdot UDU^T x = (U^T x) \cdot D(U^T x) = \sum_{i=1}^n \lambda_i \alpha_i^2$, where I have applied the change of variables $\alpha = U^T x$.
- Hence, the *quadratic* $f(x) = x \cdot Hx$ is a simple decoupled quadratic when expressed in terms of the alternate basis $\{u^{(1)}, \cdots, u^{(n)}\}$.
- Since every symmetric matrix has a spectral decomposition, this means that every quadratic function$f(x) = x \cdot Hx$ can be expressed as a simple decoupled quadratic, provided the correct coordinate system is chosen.

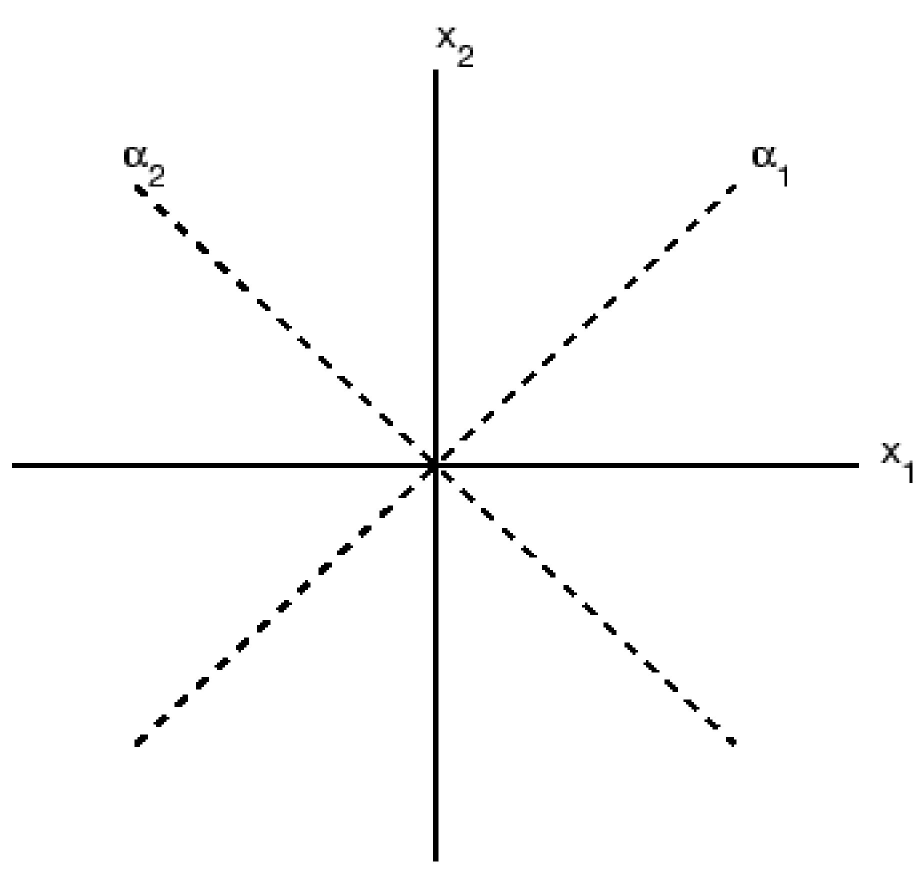
Applications of Spectral Theorem

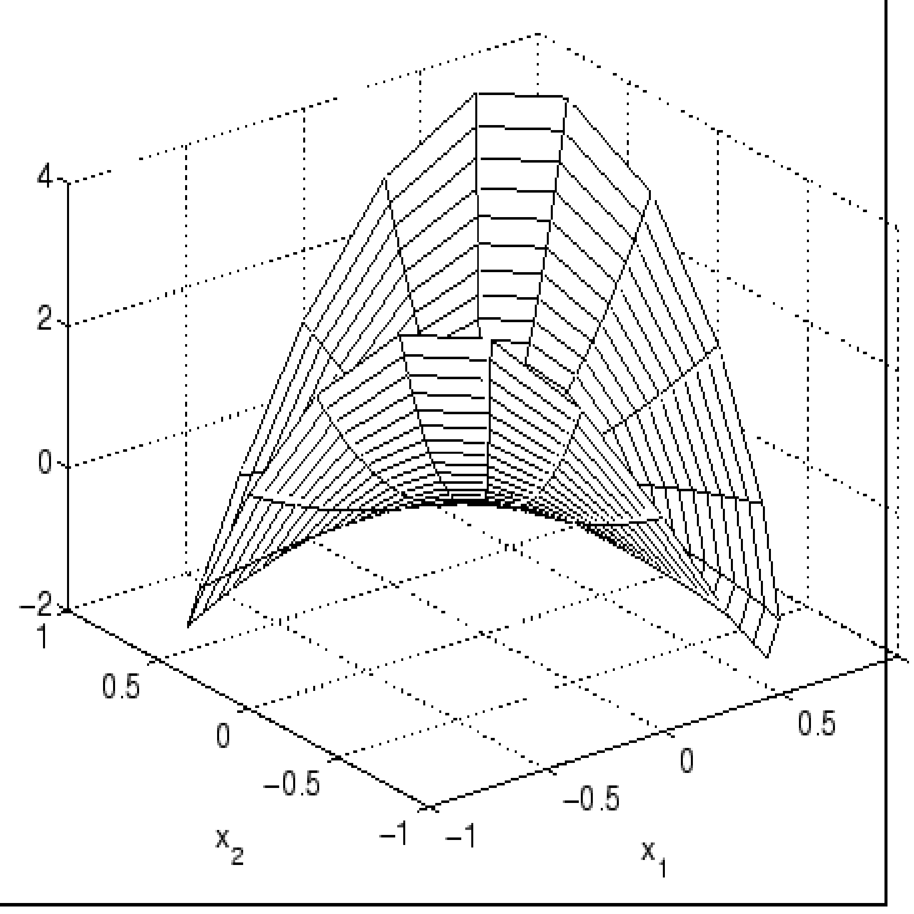
Prob

Every *quadratic* function in the n variables x_1, x_2, $\cdots$, x_n can be expressed in the form $f(x) = x \cdot Hx = \sum_{i=1}^n \sum_{j=1}^n H_{ij} x_i x_j$, which involves n^2 terms, and the variables are typically coupled.

- when H is a diagonal matrix, the function can be simplified: $f(x) = \sum_{i=1}^n H_{ii} x_i^2$
- when H is a general matrix, the function can also be simplified by selecting a correct coordination system.

- A quadratic function $f(x) = x_1^2 + 6x_1x_2 + x_2^2 = x \cdot Hx$, where $H = \left[\begin{smallmatrix} 1 & 3 \\ 3 & 1 \end{smallmatrix}\right]$ can be diagonalized $H = UDU^T$, with $U = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{bmatrix}$ and $D = \left[\begin{smallmatrix} 4 & 0 \\ 0 & -2 \end{smallmatrix}\right]$
- The new coordinator system is defined by $u^{(1)} = \left[\begin{smallmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{smallmatrix}\right]$ and $u^{(2)} = \left[\begin{smallmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{smallmatrix}\right]$. The function curves up in the direction of $u^{(1)}$ and down in the direction of $u^{(2)}$.





Applications of Spectral Theorem

Exercises:

- Find the eigenvalues and eigenvectors and the diagonalizing matrix S for $\begin{bmatrix} 1 & 0 \\ 2 & 3 \end{bmatrix}$
- If A has eigenvalues of 0 and 1, corresponding to the eigenvectors $\begin{bmatrix} 1 \\ 2 \end{bmatrix}$ and $\begin{bmatrix} 2 \\ -1 \end{bmatrix}$, how can you tell in advance that A is symmetric? What are its trace and determinant? What is A?
- Write the following matrix in the form $\lambda_1 x_1 x_1^T + \lambda_2 x_2 x_2^T$ of the spectral theorem: $\begin{bmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{bmatrix}$, and $\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$

Positive Definite Symmetric Matrix

Prob For a real symmetric matrix A to be positive definite, it needs satisfy any of the following:

- $x^T A x > 0$ for all non-zero vectors x
- All the eigenvalues of A satisfy $\lambda_i > 0$
- All the upper left submatrices A_k have positive determinants
- All the pivots (without row exchanges) satisfy $d_i > 0$

- In Gaussian elimination of a symmetric matrix A, the upper triangular U is the transpose of the lower triangular L. Then $A = LDU$ becomes $A = LDL^T$.

$$x^T A x = (x^T L)(D)(L^T x) = d_1 (L^T x)_1^2 + d_2 (L^T x)_2^2 + \cdots + d_n (L^T x)_n^2$$

- For example:

$$A = \begin{bmatrix} 2 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 2 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ -\frac{1}{2} & 1 & 0 \\ 0 & -\frac{2}{3} & 1 \end{bmatrix} \begin{bmatrix} 2 & & \\ & \frac{3}{2} & \\ & & \frac{4}{3} \end{bmatrix} \begin{bmatrix} 1 & -\frac{1}{2} & 0 \\ 0 & 1 & -\frac{2}{3} \\ 0 & 0 & 1 \end{bmatrix} = LDL^T$$

$$x^T A x = 2(x_1 - \frac{1}{2} x_2)^2 + \frac{3}{2}(x_2 - \frac{2}{3} x_3)^2 + \frac{4}{3}(x_3)^2$$

Positive Definite Symmetric Matrix

Prob

For a real symmetric matrix A to be positive definite, it needs satisfy any of the following:

- $x^T Ax > 0$ for all non-zero vectors x
- All the eigenvalues of A satisfy $\lambda_i > 0$
- All the upper left submatrices A_k have positive determinants
- All the pivots (without row exchanges) satisfy $d_i > 0$
- There is a matrix R with independent columns such that $A = R^T R$.

- Assume a rectangular matrix R and a least square problem $Rx = b$. The least square choice $\bar{x}$ is the solution of $R^T R\bar{x} = R^T b$. Provided that the columns of R are linearly independent, the matrix $R^T R$ is positive definite symmetric: $x^T R^T Rx = \|Rx\|^2$, which can not be negative or zero.
- When A is positive definite, we have two choices:
  - From SED: $A = Q\Lambda Q^T = (Q\sqrt{\Lambda})(\sqrt{\Lambda}Q^T) = R^T R$.
  - From Gaussian Elimination: $A = LDL^T = (L\sqrt{D})(\sqrt{D}L^T) = R^T R$.

Positive Semi-Definite Symmetric Matrix

Prob

For a real symmetric matrix A to be positive semidefinite, it needs satisfy any of the following:

- $x^T Ax \geq 0$ for all non-zero vectors x
- All the eigenvalues of A satisfy $\lambda_i \geq 0$
- All the upper left submatrices A_k have non-negative determinants
- No pivots are negative
- There is a matrix R, possibly with with dependent columns, such that $A = R^T R$.

- Example: $A = \begin{bmatrix} 2 & -1 & -1 \\ -1 & 2 & -1 \\ -1 & -1 & 2 \end{bmatrix}$ is semidefinite because:
  - $x^T Ax = (x_1 - x_2)^2 + (x_1 - x_3)^2 + (x_2 - x_3)^2 \geq 0$
  - The eigenvalues are $\lambda_1 = 0$, $\lambda_2 = \lambda_3 = 3$.
  - The submatrices determinants are 2, 3 and 0 respectively.
  - $\begin{bmatrix} 2 & -1 & -1 \\ -1 & 2 & -1 \\ -1 & -1 & 2 \end{bmatrix} \to \begin{bmatrix} 2 & 0 & 0 \\ 0 & \frac{3}{2} & -\frac{3}{2} \\ 0 & -\frac{3}{2} & \frac{3}{2} \end{bmatrix} \to \begin{bmatrix} 2 & 0 & 0 \\ 0 & \frac{3}{2} & 0 \\ 0 & 0 & 0 \end{bmatrix}$

Questions?

Contact Information

Associate Professor **Gang Li**
School of Information Technology
Deakin University, Australia

GANGLI@TULIP.ORG.AU
TEAM FOR UNIVERSAL LEARNING AND INTELLIGENT PROCESSING