Objective of the course

Introduce how to use linear algebra techniques to solve practical problems in:

Image processing, Differential equations, Difference equations, Quantum Computing, Optimization, Deep Learning, etc.

A simple description of Deep Learning and Neural Network model.

- A simple example. Identify the images of $0, \ldots, 9$ using training data $x_1, \ldots, x_n$.
- Apply functions $F(x)$ so that it will correctly identify the outcome.
- It turns out that functions of the form $F(x) = L(R(L(R(\cdots(x)))))$, where $L(x) = Ax + b$ and $R(x) = (\max(0, x_1), \ldots, \max(0, x_n))^T$ for $x = (x_1, \ldots, x_n)^T$ work well.
- In the neural network setting, one uses the input v to adjust $L_k = A_k v_{k-1} + b_k$, to produce a new hidden layer.
- The composite function $F(v) = L_k(R(L_{k-1}(R \cdots (v))))$ adds depth to the network and leads to more successful model.

Basic notation and background

- $M_n(\mathbb{F})$, $M_{m,n}(\mathbb{F})$ are the set of $n \times n$ and $m \times n$ matrices over $\mathbb{F} = \mathbb{R}$ or $\mathbb{C}$ (or a general field).
- Basic operations of complex numbers are assumed.

  $z = x + iy = \rho e^{i\theta}$, $\rho = |z| = |\bar{z}z|^{1/2} = \sqrt{x^2 + y^2}$, $\bar{z} = x - iy$, $z_1 + z_2, z_1 z_2, z_1/z_2$ if $z_2 \neq 0$.
- $\mathbb{F}^n$ is the set of column vectors of length n with entries in $\mathbb{F}$.
- If $\mathbb{F}$ is clear, we use the notation $M_n, M_{m,n}$.
- Let $A \in M_{m,n}$. Then $A^T \in M_{n,m}$. For complex matrix A, we have $\bar{A}$ and $A^* = (\bar{A})^T = \overline{A^T}$.

I.1 A close look at $Ax = b$

- Linear equations, elementary row operations, solution sets.

  Recall. Let $A \in M_{m,n}(\mathbb{F}), b \in \mathbb{F}^n$, $Ax = b$.

  Find $E_r \cdots E_1[A|b] = [\tilde{A}|\tilde{b}]$ in row echelon form, $E_r \cdots E_1[A|I_n] = [I_n|A^{-1}]$.

- Matrices, column space, row space, null space, ranks.

  Recall. Let $A \in M_{m,n}(\mathbb{F})$, and $E_r \cdots E_1 A = [\tilde{A}]$ in echelon form.

  Then we can find the bases for column space, row space, and null space, and the rank of A.

Interpretation of $Ax = b$.

Example Let $A = [A_1 A_2] \in M_{3,2}, x = (x_1, x_2)^T, b \in \mathbb{F}^3$.

Then $Ax = b$ means $b = x_1 A_1 + x_2 A_2$.

All combination of A_1, A_2 form the column space.

The equation $Ax = b$ is solvable means that b is in the column space.

In general, if $A \in M_{3,n}$, the column space can be of dimensions $0, 1, 2, 3$.

All these comments hold for the general case: $Ax = b$.

For example, if $A = [A_1|\cdots|A_n] \in M_{m,n}(\mathbb{F})$ and $Ax = b \in \mathbb{F}^m$ then $b = x_1 A_1 + \cdots + x_n A_n$.

If $Ax = 0 \in \mathbb{F}^m$ has non-trivial solution then $\{A_1, \ldots, A_n\}$ is linearly independent, i.e., there is $x_1, \ldots, x_n$ not all zero such that $x_1 A_1 + \cdots + A_n = 0$.

The null space of $A \in M_{m,n}$ is the set/subspace $\{x \in \mathbb{F}^n : Ax = 0\}$.

Theorem rank + null space dimension = n.

Proposition If $A \in M_{m,n}$ has rank r, then $A = CR$, where $C \in M_{m,r}$ with independent columns forming a basis for the column space, and $R \in M_{r,n}$ has independent rows forming a basis for the row space. So, the row rank and column rank of A are the same.

Remark The result is useful for low rank factorization.

There will be better factorization, namely, the singular value decomposition.

I.2 Matrix Multiplication

- Note that $A \in M_{m,n}$ is rank one if and only if $A = uv^T$ for some nonzero vector $u \in \mathbb{F}^m, v \in \mathbb{F}^n$.
- In general if $A \in M_{m,n}$ has columns $u_1 \dots u_n$ and $B \in M_{n,p}$ has rows $b_1^T, \dots, b_n^T$, i.e., $A = [u_1 \cdots u_n], B^T = [b_1 \cdots b_n]$, then $AB = u_1v_1^T + \cdots + u_nv_n^T$.

  One can check the (i,j) entry of $C = AB$ and $\sum_{j=1}^m u_iv_j^T$.
- Furthermore, if $D = \operatorname{diag}(d_1, \dots, d_n)$, then $ADB = d_1u_1v_1^T + \cdots + d_nu_nv_n^T$.

Example

Proposition. Suppose $A \in M_{m,n}, B \in M_{n,p}, C \in M_{p,q}$. Show that $(AB)C = A(BC)$.

Case 1. Suppose $B = xy^T$ with $\mathbf{x} = (x_1, \dots, x_n)^t$, $\mathbf{y}^T = (y_1, \dots, y_p)$.
Let $A\mathbf{x} = (\hat{x}_1, \dots, \hat{x}_m)^T$ and $\hat{\mathbf{y}}^T = \mathbf{y}^T C = (\hat{y}_1, \dots, \hat{y}_q)$. Then

$$\begin{aligned}(AB)C &= (A[\mathbf{x}y_1|\cdots|\mathbf{x}y_p])C = [(A\mathbf{x})y_1|\cdots|(A\mathbf{x})y_p]C = ((A\mathbf{x})\mathbf{y}^T)C = (\hat{\mathbf{x}}\mathbf{y}^T)C \\ &= \begin{bmatrix}\hat{x}_1\mathbf{y}^T \\ \vdots \\ \hat{x}_m\mathbf{y}^T\end{bmatrix} C = \begin{bmatrix}\hat{x}_1\mathbf{y}^T C \\ \vdots \\ \hat{x}_m\mathbf{y}^T C\end{bmatrix} = \begin{bmatrix}\hat{x}_1\hat{\mathbf{y}}^T \\ \vdots \\ \hat{x}_m\hat{\mathbf{y}}^T\end{bmatrix} = [\hat{\mathbf{x}}\hat{y}_1|\cdots|\hat{\mathbf{x}}\hat{y}_q] \\ &= [(A\mathbf{x})\hat{y}_1|\dots|(A\mathbf{x})\hat{y}_q] = A\begin{bmatrix}x_1\hat{\mathbf{y}}^T \\ \cdots \\ x_n\mathbf{y}^T\end{bmatrix} = A(\mathbf{x}\mathbf{y}^T C) = A(BC).\end{aligned}$$

Case 2. If $B = \sum_{j=1}^r \mathbf{x}_j\mathbf{y}_j^T$, then

$$(A(\sum_{j=1}^r \mathbf{x}_j\mathbf{y}_j^T))C = (\sum_{j=1}^r A(\mathbf{x}_j\mathbf{y}_j^T))C = \sum_{j=1}^r A((\mathbf{x}_j\mathbf{y}_j^T)C)) = A(\sum_{j=1}^r \mathbf{x}_j\mathbf{y}_j^T C) = A(BC).$$

Matrix factorization.

$$A = LU, \quad A = QR, \quad A = RDR^{-1}S \quad A = Q\Lambda Q^T \text{ if } A \text{ is symmetric}, \quad A = U\Sigma V^T.$$

1. LU factorization for invertible $A \in M_n$.

   To solve $Ax = b$ for many different vectors $b \in \mathbb{F}^n$.

   Write $A = LU$ where L is lower triangular, and U as upper triangular.

   Then solve $Ax = LUx = b$ by solving $Ly = b$ and $Ux = y$.

   How to write $LU = A$? We will see this in Section 1.4

   One can do an exchange of rows by a permutation matrix P to get $PA = LU$ if A is invertible.

   Question What if A is not invertible or A is not a square matrix?

2. QR Decomposition for $A \in M_{m,n}$ with $m \geq n$.

   Recall the inner product on $\mathbb{F}^n$ defined by $\langle x, y\rangle = y^*x = \sum_{j=1}^n x_j\bar{y}_j$, and the inner product norm $\|x\| = \langle x, x\rangle^{1/2}$.

   We are interested in orthonormal set/basis $\{v_1, \ldots, v_n\}$ such that $V^*V = I_n$ if $V = [v_1|\cdots|v_n]$, i.e., $\langle v_i, v_j\rangle = \delta_{ij}$.

   To solve $Ax = b$ with b in the column space of b, we can solve $Rx = Q^*Ax = Q^*b$.

3. $A = RDR^{-1} = \sum_{j=1}^n \lambda_j x_j y_j^T$ where R has columns $x_1, \ldots, x_n$ and R^{-1} has rows $y_1^T, \ldots, y_n^T$. Then $A^k = \sum_{j=1}^n \lambda_j^k x_j y_j^T$.

   If $H = H^*$ is Hermitian, then $H = Q\Lambda Q^* = \sum_{j=1}^n \lambda_j v_j v_j^*$ with $\lambda_1, \ldots, \lambda_n \in \mathbb{R}$ so that $H^k = \sum_{j=1}^n \lambda_j^k v_j v_j^*$.

4. $A = U\Sigma V^*$. Then $A = \sum_{j=1}^r s_j u_j v_j^*$.

I.3 The Four Fundamental Subspaces

Let $A \in M_{m,n}$. We have
the column space $C(A)$ in $\mathbb{F}^m$ contains all combination of columns of A,
the null space $N(A)$ contains x in $\mathbb{F}^n$ such that $Ax = 0$.
the row space $C(A^T)$ contains all combination of columns of A^T,
the left null space $N(A^T)$ contains y in $\mathbb{F}^m$ such that $A^T y = 0$.

Proposition Let $A \in M_{m,n}$. Then $C(A)$ and $C(A^T)$ have the same dimension r; $N(A)$ has dimension $n - r$, $N(A^T)$ has dimension $m - r$.

Theorem Let $A \in M_{m,r}, B \in M_{r,n}$.

1. $\text{rank}\,(AB) \le \min\{\text{rank}\,(A), \text{rank}\,(B)\}$.
2. $\text{rank}\,(A + B) \le \text{rank}\,(A) + \text{rank}\,(B)$.
3. $\text{rank}\,(A^*A) = \text{rank}\,(AA^*) = \text{rank}\,(A) = \text{rank}\,(A^*)$.
4. $\text{rank}\,(AB) = r$ if $\text{rank}\,(A) = \text{rank}\,(B) = r$.

I.4 LU factorization

Recall that we want to write $A = LU$ for an invertible $A \in M_n$. How to do it?

- Use elementary operations by subtracting multiple of upper rows from lower rows to get an upper triangular matrix $E_r \cdots E_1 A = U$.
- Let $R = E_r \cdots E_1$. Then R is lower triangular, and $A = R^{-1}U$ so that $L = E_1^{-1} \cdots E_r^{-1}$.
- A better way: $A - u_1 v_1^T = [0] \oplus A_1$ if we let $u_1 = (a_{11}, \ldots, a_{n1})^T / a_{11}$ and $v_1^T = (a_{11}, \ldots, a_{1n})$.
- Then use induction on A_1 to get $A = \sum_{j=1}^n u_j v_j^T$.

Example

Remark We can apply the procedure as long as the $(1,1)$ entry of A_k is nonzero in each step.

Else, one can apply a permutation P to A so that $PA = LU$.

Just choose P so that the leading $k \times k$ submatrix is invertible for each $k = 1, \ldots, n$.

I.5 Orthogonality

- Inner product, norm, and orthogonal vectors.

- Pythagoras theorem. $\|x-y\|^2 = \|x\|^2 + \|y\|^2$.

- Cosine law. $\|x-y\|^2 = \|x\|^2 + \|y\|^2 - 2\|x\|\|y\|\cos\theta$, where $\langle x, y\rangle = \|x\|\|y\|\cos\theta$.

Gram-Schmidt process and QR factorization

- Let $\{u_1, \ldots, u_k\} \subseteq \mathbb{R}^n$ be a linearly independent set. We can an orthonormal set $\{v_1, \ldots, v_k\}$ by the Gram-Schmidt process as follows.

  Let $v_1 = u_1/\|u_1\|$. If $v_1, \ldots, v_\ell$ are constructed and $\ell < k$, set

  $$\tilde{v}_{\ell+1} = u_{\ell+1} - a_1 v_1 - \cdots a_\ell v_\ell \quad \text{and} \quad v_{\ell+1} = \tilde{v}_{\ell+1}/\|\tilde{v}_{\ell+1}\|$$

  with $a_j = \langle v_j, u_{\ell+1}\rangle$ for $j = 1, \ldots, \ell$.

- As a result, we can construct an orthonormal basis for a subspace.

- Orthogonal subspaces. For example, row space and null space are orthogonal subspaces.

- Also, we have the QR factorization.

  $$[u_1|\cdots|u_k] = [v_1|\cdots|v_k]R = QR$$

  for an upper triangular matrix $R \in M_k$.

Example

Special classes of matrices

- Matrix $Q \in M_{n,m}$ with orthonormal columns. $Q^*Q = I_m$. Then QQ^* is a projection.

- A matrix $Q \in M_n$ such that $Q^*Q = I_n$ is a unitary/orthogonal matrix.

  The set of unitary / orthogonal matrices form a group.

- Hadamard matrices.

  If $n = 4k$, then there is a Hadamard matrix in in $M_n(\mathbb{R})$.

  The smallest unknown case: 668.

- Hadamard matrices are useful in quantum computing.

- Householder reflection has the form $I - 2uu^*$ for a unit vector. The matrix has eigenvalues $-1, 1, \ldots, 1$.

I.6 Eigenvalues, eigenvectors

Recall. We solve $Ax = \lambda x$ by solving (1) $\det(\lambda I - A) = 0$ for λ, and (2) $(\lambda I - A)x = 0$.

Theorem There is an invertible S such that $S^{-1}AS = D = \operatorname{diag}(\lambda_1, \ldots, \lambda_n)$ if and only if A has n linearly independent eigenvectors (the columns of S).

Proof.

Example Find S such that $S^{-1}AS = \operatorname{diag}(d_1, d_2)$ for $A = \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix}$.

Applications

Recursive relation. Suppose $a_1, \ldots, a_k$ are given and $a_n = d_1 a_{n-1} + \cdots + d_k a_{n-k}$. We can find formula for a_n as follows.

Write $x_n = (a_n, \ldots, a_{n-k+1})^T$ and $A = \begin{pmatrix} d_1 & d_2 & \cdots & d_k \\ 1 & 0 & \cdots & 0 \\ & \ddots & \ddots & \vdots \\ & & 1 & 0 \end{pmatrix} \in M_k$.

Then $x_n = Ax_{n-1} = A^2 x_{n-2} = \cdots = A^{n-k} x_k$.

If $A = SDS^{-1} = \lambda_1 S(:,1)S^{-1}(1,:) + \lambda_k S(:,k)S^{-1}(k,:)$, where $D = \operatorname{diag}(\lambda_1, \ldots, \lambda_k)$, then

$$x_n = SD^k S^{-1} = \lambda_1 S(:,1)S^{-1}(1,:) + \lambda_k S(:,k)S^{-1}(k,:)x_k$$

so that $a_n = e_1^T SD^k S^{-1} x_k$.

Example Suppose $f_0 = 0, f_1 = 1$ and $f_n = f_{n-1} + f_{n-2}$. Then $x_n = A^{n-1}x_1$ with $x_1 = (1,0)^T$ and $A = \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix} = SDS^{-1}$ with $D = \operatorname{diag}(\lambda_1, \lambda_2) = \operatorname{diag}((1+\sqrt{5})/2, (1-\sqrt{5})/2)$ and

$$S = \begin{bmatrix} 1 & (\sqrt{5}-1)/2 \\ (\sqrt{5}-1)/2 & -1 \end{bmatrix} \quad \text{and} \quad S^{-1} = \frac{2}{5-\sqrt{5}} S.$$

Thus, $a_n = \frac{1}{\sqrt{5}}(\lambda_1^n + \lambda_2^n)$.

Exercise 1 Determine a formula for a_n if $f_0 = 1, f_1 = 2$ and $f_n = f_{n-1} + f_{n-2}$ for $n \geq 2$.

Differential equations. Consider the system $Ax' = x$ with $x(0) = x_0$.

For one variable, $x' = ax$ with $x(0) = x_0$, we have $x = e^{at}x_0$.

Suppose $S^{-1}AS = D = \operatorname{diag}(d_1, \ldots, d_n)$. Then

$$A = SDS^{-1} = d_1 S(:,1)S^{-1}(1,:) + \cdots + d_n S(:,n)S^{-1}(n,:).$$

For $y = Sx$, $Sx' = y'$ and $Dy' = y$ so that

$$y = e^{Dt}y_0 = \begin{bmatrix} e^{id_1t} & & \\ & \ddots & \\ & & e^{id_nt} \end{bmatrix} y_0 \quad \text{and} \quad x = S^{-1}y = S^{-1}e^{Dt}Sx_0.$$

That is,

$$x = (e^{d_1t}S(:,1)S^{-1}(1,:) + \cdots + e^{d_nt}S(:,n)S^{-1}(n,:))x_0.$$

Example: Solve $x' = Ax$ with $A = \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix}$ and $x(0) = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$. Then $A = \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix} = SDS^{-1}$ with $D = \operatorname{diag}(\lambda_1, \lambda_2) = \operatorname{diag}((1+\sqrt{5})/2, (1-\sqrt{5})/2)$ and

$$S = \begin{bmatrix} 1 & (\sqrt{5}-1)/2 \\ (\sqrt{5}-1)/2 & -1 \end{bmatrix} \quad \text{and} \quad S^{-1} = \frac{2}{5-\sqrt{5}}S.$$

So,

$$x = \frac{1}{\sqrt{5}} \begin{bmatrix} \lambda_1 e^{\lambda_1 t} - \lambda_2 e^{\lambda_2 t} \\ e^{\lambda_1 t} - e^{\lambda_2 t} \end{bmatrix}.$$

In fact, one can solve the equation $x'' = x' + x$ with $x(0) = (1,0)^T$.

Now, $\begin{pmatrix} x' \\ x \end{pmatrix}' = A \begin{pmatrix} x' \\ x \end{pmatrix}'$ so that $x = \frac{1}{\sqrt{5}} e^{\lambda_1 t} - e^{\lambda_2 t}$.

Exercise Solve the system $x' = Ax$ with the above A and $x_0 = (2,1)^T$.

1.7 Symmetric matrices, Hermitian matrices, positive definite matrices

Theorem (a) Every real symmetric matrix A can be written as QDQ^T for a real diagonal matrix $D = \operatorname{diag}(\lambda_1, \ldots, \lambda_n)$ and an orthogonal matrix Q.

(b) Every complex Hermitian matrix A can be written as UDU^* for a real diagonal matrix $D = \operatorname{diag}(\lambda_1, \ldots, \lambda_n)$ and a unitary matrix U.

Proof. If $A \in M_2(\mathbb{R})$ be symmetric, let $v \in= [v_1, v_2]^T \in \mathbb{R}^2$ be such that $v^T v = 1$ and

$$v^T A v = \lambda_1 = \max\{x^T A x : x \in \mathbb{R}^2, x^T x = 1\}.$$

Then $Q = \begin{pmatrix} v_1 & v_2 \\ v_2 & -v_1 \end{pmatrix}$ is orthogonal, and we claim that $Q^T A Q = \operatorname{diag}(\lambda_1, \lambda_2)$. If not, assume $Q^T A Q = \begin{pmatrix} \lambda_1 & b \\ b & c \end{pmatrix}$. Then

$$f(\theta) = [\cos\theta, \sin\theta] Q^T A Q [\cos\theta, \sin\theta]^T = \cos^2\theta\lambda_1 + \cos\theta\sin\theta b + \sin^2\theta c,$$

and $f'(0) = 2b \neq 0$. So, we can find θ near 0 such that $u^T A u > \lambda_1$ with $u = Q[\cos\theta, \sin\theta]^T$, which is a contradiction.

Now, for $A \in M_n(\mathbb{R})$. Let $v \in \mathbb{R}^n$ be a unit vector such that

$$v^T A v = \lambda_1 = \max\{x^T A x : x \in \mathbb{R}^n, x^T x = 1\}.$$

Suppose Q is an orthogonal matrix such that $Q^T A Q = A_1$. Then $A_1 = [\lambda_1] \oplus A_2$ ....

For a Hermitian matrix $A \in M_n(\mathbb{C})$, if $v \in \mathbb{C}^n$ and $\mu = v^* A V$, then $\bar{\mu} = (v^* A v)^* = v^* A^* v = v^* A v = \mu$. So, μ is real and we can consider a unit vector $v \in \mathbb{C}^n$ such that

$$v^* A v = \lambda_1 = \max\{u^* A u : u \in \mathbb{C}^n, u^* u = 1\}.$$

In practice, we can do the proof and diagonalization as follows. First, we can treat a real symmetric matrix as a complex Hermitian matrix A. Then there is a possible complex eigenvalue λ and a unit eigenvector x. We have $Ax = \lambda x$ and thus $x^* A x = x^* \lambda x = \lambda$. Now, $\bar{\lambda} = *x^* A x)^* = x^* A^* x x^* A x = \lambda$. So, λ is real.

Then we compute an eigenvalue λ and an unit eigenvector x, which is real in the real case, so that $Ax = \lambda x$, and let U_1 be a unitary matrix (an orthogonal matrix in the real case) with x as the first column. Then $U_1^* A U_1$ has the first column equal to $[\lambda, 0, \ldots, 0]^T$. But $U_1^* A U_1$ is also Hermitian (real symmetric). So, $U_1^* A U_1 = [\lambda] \oplus A_1$, where A_1 is also Hermitina (real symmetric). By induction, $U_2^T A_1 U_2 = D_2$, a real diagonal matrix. Thus, $U = U_1([1] \oplus U_2$ is unitary and $U^* A U = D$.

Positive semidefinite matrices.

Definition A real symmetric matrix $A \in M_n(\mathbb{R})$ is positive definite (semi-definite) if it has positive (nonnegative) eigenvalues

A complex Hermitian matrix $A \in M_n(\mathbb{C})$ is positive definite (semi-definite) if it has positive (nonnegative) eigenvalues

Theorem Let $A \in M_n$ be Hermitian. Then A is positive definite if and only if any one of the following holds.

1. $x^*Ax > 0$ for all nonzero vector x.

2. $A = B^*B$ for an invertible B. (In fact, B can be chosen to satisfy $B^* = B$.)

3. A has positive leading principal minors meaning ...

4. All pivots of A are positive. [In the Gaussian elimination process.] Then $A = LDL^*$ or $A = \tilde{L}\tilde{L}^*$ for some invertible lower triangular matrix $\tilde{L}$, the Cholesky factorization.

A matrix $A \in M_n$ is positive semi-definite if any of the following equivalent conditions hods.

1. $x^*Ax \geq 0$ for all nonzero vector x.

2. $S = A^*A$ for a matrix A.

3. A has nonnegative principal minors.

4. $A = \tilde{L}\tilde{L}^*$ for some lower triangular matrix L, the Cholesky factorization.

Example. $A = \begin{pmatrix} 4 & 3 \\ 3 & 4 \end{pmatrix}$.

Proof of Theorem. Suppose A is positive definite. Then $A = UDU^* = B^*B$ with $B = UD^{1/2}U^* = B^*$. Thus (2) holds.

If (2) holds, then for any nonzero vector x, $y = Bx$ is nonzero so that $x^*x = (x^*B^*)(Bx) = y^*y > 0$.

We prove (1). Note that $A = UDU^*$, where U is unitary and $D = \operatorname{diag}(\lambda_1, \ldots, \lambda_n)$ such that $\lambda_1, \ldots, \lambda_n > 0$. Then for any nonzero vector x, we can let $y = U^*x = (y_1, \ldots, y_n)^T$ so that $x^*Ax = x^*(UDU^*)x = y^*Dy = \lambda_1|y_1|^2 + \cdots + \lambda_n|y_n|^2 > 0$.

Suppose (1) holds. Then we show that (3) holds as follows. Assume that A_k is the $k\times k$ principal submatrix of A. If A_k has a non-positive eigenvalue μ, then there is a unit vector $y = (y_1, \ldots, y_k)^T \in \mathbb{C}^k$ such that $A_k y = \mu y$ so that $y^*A_k y = y^*\mu y = \mu \le 0$. Hence, for $x = (y_1, \ldots, y_k, 0, \ldots, 0)^T \in \mathbb{C}^n$, we have $x^*Ax = \mu \le 0$, contradicting (1). So, all eigenvalues of A_k is positive and $\det(A_k)$ is positive.

Suppose (3) holds. Then we show that (4) all pivots of A are positive, i.e., $A = LU$ for some LU factorization with a lower triangular L so that all the diagonal entries are 1, and an upper triangular matrix U. Now, the leading principal submatrix of $A_k = L_kU_k$, where L_k, U_k are the leading principal submatrices of L and $U = (u_{ij})$. We see that the diagonal entries of U are positive. So, we can write $U = DU$, where $D = \operatorname{diag}(1/u_{11}, \ldots, 1/u_{nn})$ is a positive matrix. Now, we have $A = LD\tilde{U} = d_1x_1y_1^* + \cdots + d_nx_ny_n^*$, where $L = [x_1 \cdots x_n]$, and $U^* = [y_1, \ldots, y_n]$ is in lower triangular form. Note that the first row and first column of A are the same as that of $d_1x_1y_1^*$. So, $x_1 = y_1$. Now, $A - d_1x_1y_1^* = d_2x_2y_2^* + \cdots + d_nx_ny_n^*$ is Hermitian. We see that $x_2 = y_2$. Inductively, we see that $x_k = y_k$ for all k so that $U = L^*$.

Suppose (4) holds. Then for any nonzero vector x, $y = L^*x = (y_1, \ldots, y_n)^T$ is nonzero as $\det L = 1 \neq 0$. So, $x^*Ax = x^*LDL^*x = d_1|y_1|^2 + \cdots + d_n|y_n|^2 > 0$. Thus (1) holds.

Hence, we have show that (1) $\Rightarrow$ (3) $\Rightarrow$ (4) $\Rightarrow$ (1).

Now, suppose A is positive definite, i.e., $A = UDU^*$ for some unitary and diagonal matrix D with positive diagonal. Then $A = B^*B$ with $B^* = B = UD^{1/2}U^*$. So, (2) holds.

Suppose (2) holds. Then for any nonzero vector x, the vector $y = Bx$ is nonzero. So, $x^*Ax = y^*Dy > 0$. So, (1) holds.

Suppose (1) holds. If A has a non-positive eigenvalue μ, then there is a unit eigenvector x for μ such that $x^*Ax = \mu \le 0$, which contradicts (1). So, we A has positive eigenvalues and is possitive definite.

So, we see that A is positive definite $\Rightarrow$ (2) $\Rightarrow$ (1) $\Rightarrow$ A is positive definite. □

Applications

1. Maximum and minimum of real valued function $f(x) = f(x_1, \ldots, x_n)$**.**

Then $f(x_1 + t_1, \ldots, x_n + t_n)$

$$= f(x_1, \ldots, x_n) + (f_{x_1}(x), \ldots, f_{x_n}(x))(t_1, \ldots, t_n)^T + (t_1, \ldots, t_n)J_f(x)(t_1, \ldots, t_n)^T + O(t^3),$$

where

$$J_f(x) = (f_{x_i,x_j}(x)) = \left(\frac{\partial^2 f(x)}{\partial x_i \partial x_j}\right).$$

Thus, $f(x)$ is minimum if $f_{x_j}(x) = 0$ for all $j = 1, \ldots, n$, and $J_f(x)$ is positive definite.

2. Major and minor axes of elliptical disk/ellipsoid.

Suppose the ellipse equation $1 = 5x^2 + 8xy + 5y^2$ is written as $1 = (x, y)A(x, y)^T$ with $A = \begin{pmatrix} 5 & 4 \\ 4 & 5 \end{pmatrix}$. Then A is positive definite, and $A = QDQ^T$ with $D = \text{diag}\,(9, 1)$ and $Q = \frac{1}{2}\begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix}$. Then the ellipse equation becomes $1 = 9X^2 + Y^2$ with $X = (x + y)\sqrt{2}$ and $Y = (-x + y)/\sqrt{2}$. Geometrically, we apply a rotation of $-\pi/4$, we get a vertical ellipse.

Application. Rotation of the ellipse to the upright position for facial recognition.

1.8 Singular value decomposition, polar decomposition.

Theorem Let $A \in M_{m,n}$. There exists unitary $U \in M_{m,k}$ and $V \in M_{n,k}$ each with orthonormal columns such that $U^*U = I_k = V^*V$ and $\Sigma = \text{diag}\,(s_1, \dots, s_k)$ with $s_1 \geq \cdots \geq s_k > 0$ such that

$$A = U\Sigma V^* = \sum_{j=1}^{k} s_j u_j v_j^*.$$

Proof. If $\tilde{V}^* A^* A \tilde{V} = \text{diag}\,(s_1^2, \dots, s_n^2)$, then the first k columns of $\tilde{V}$ are orthogonal columns of lengths $s_1, \dots, s_k$ corresponding to the nonzero eigenvalues $s_1^2, \dots, s_k^2$. So, there is a unitary $\tilde{U} \in M_m$ such that $A\tilde{V} = \tilde{U}\tilde{D}$, where $\tilde{D} \in M_{m,n}$ with (j,j) entry equal to s_j for $j = 1, \dots, k$, and all other entries zero. So, $A = \tilde{U}\tilde{D}\tilde{V}^* = \sum_{j=1}^{k} s_j u_j v_j^*$, where $u_1, \dots, u_k$ are the first k columns of $\tilde{U}$ and $v_1, \dots, v_k$ are the first k columns of V. □

Definition The values $s_1 \geq \cdots \geq s_k$ are known as the (nonzero) singular values of A. The produce $U\Sigma V^*$ is called the singular decomposition of A.

Note that AA^* and A^*A have eigenvalues $s_1^2 \geq \cdots \geq s_k^*$ and zoeros.

Also, the Wielandt matrix $S = \begin{bmatrix} 0 & A \\ A^* & 0 \end{bmatrix}$ has eigenvalues $s_1, \dots, s_k, -s_k, \dots, -s_1$ and zeros.

Theorem Let $A \in M_n$. Then $A = UP = QV$ for some unitary U, V and positive semidefinite P, Q with eigenvalues $s_1 \geq \cdots \geq s_n$.

Example If A is rank one, then ...

Let $A = \begin{pmatrix} 3 & 0 \\ 4 & 5 \end{pmatrix}$. Then $A^T A = \begin{pmatrix} 25 & 20 \\ 20 & 25 \end{pmatrix}$ has eigenvalues $45, 5$ and eigenvectors $(1,1)^T$ and $(1,-1)^T$. So, $AV = U\text{diag}\,(45, 5)$ with $u_1 = \sqrt{45}(1,3)^T/\sqrt{10}$ and $u_2 = (-3,1)^T/\sqrt{10}$.

The polar decomposition is:

Applications of SVD

Transforming a circle to a ellipse Every $A : \mathbb{R}^2 \to \mathbb{R}^2$ will transform the unit circle to an ellipse.

Finding the maximum of $\|Ax\|$ among all unit vectors x.

In the Hermitian/symmetric case:

Variation principle Finding $\max \|Ax\|/\|x\|$ subject to v_1^* where v_1 is a unit eigenvector for the eigenvalue s_1^2 of A^*A

Approximation of a matrix $A = \sum_{j=1}^{k} s_j u_j v_j^*$ by $\tilde{A} = \sum_{j=1}^{\ell} s_j u_j v_j^*$.

Application in function spaces We can find a basis of differentiable periodic functions on $[0, 2\pi]$ consisting of $\{\cos kt : k = 0, 1, 2, \} \cup \{\sin kt : k = 1, 2, \dots\}$. The inner product to two functions f, g equals $\int_0^{2\pi} f(t)g(t)dt$. Then we can approximate $f(t)$ by $\sum_{j=0}^{N}(a_j \cos jt + b_j \sin jt)$.

Finite difference In the study of the discrete form of a derivative, we consider $f(x) - f(x - \Delta x)$. Let $D = \begin{pmatrix} 1 & & \\ -1 & 1 & \\ & -1 & 1 \\ & & -1 \end{pmatrix}$ with $D^T = \begin{pmatrix} 1 & -1 & & \\ & 1 & -1 & \\ & & 1 & -1 \end{pmatrix}$. In computation, it is useful important find DD^T and $D^T D$; see p.66

I.9 Principal components and the best low rank approximation

Definition A norm $\|\cdot\|$ on a vector space V is a function such that

(1) $\|v\| \ge 0$ for all $v \in V$, where equality holds if and only if $v = 0$.

(2) $\|u + v\| \le \|u\| + \|v\|$ for all $u, v \in V$.

(3) $\|\mu v\| = |\mu|\|v\|$ for any scalar μ and $v \in V$.

Example For $x = (x_1, \dots, x_n)^T \in \mathbb{R}^n$, $\|x\| = (x^T x)^{1/2}$, $\|x\|_{\sup} = \max\{|x_j| : 1 \le j \le n\}$, $\|x\|_{\ell_1} = |x_1| + \cdots + |x_n|$ are norm functions.

Definition A norm on $M_{m,n}$ is unitarily invariant if a norm on $M_{m,n}$ satisfies $\|UAV\| = \|A\|$ for any $A \in M_{m,n}$, unitary $U \in M_m$ and $V \in M_n$. In other words, the norm depend on the singular values of A only.

Examples Let $A = (a_{ij}) = \sum_{j=1}^r s_j u_j v_j^* \in M_{m,n}$

Spectral norm: $\|A\|_2 = \max_{x \ne 0} \|Ax\|/\|x\| = s_1(A)$.

Frobenius norm: $\|A\|_F = (\sum_{i,j} |a_{ij}|^2)^{1/2} = (\mathrm{tr}\, A^*A)^{1/2} = (\sum_{j=1}^j s_1^2 + \cdots + s_r^2)^{1/2}$.

Nuclear norm or trace norm: $\|A\|_N = \sum_{j=1}^r s_j(A)$.

Theorem Suppose $\|\cdot\|$ is a unitarily invariant norm on $M_{m,n}$. Let $A = \sum_{j=1}^r s_j u_j v_j^*$. If $B = \sum_{j=1}^k s_j u_j v_j^*$, where $s_j = 0$ if $j > r$, then $\|A - B\| \le \|A - X\|$ for any $X \in M_{m,n}$ of rank at most k.

Proof for the spectral norm. We may focus on the case when $k < r$. Else, we can let $B = A$ be the optimizer. Suppose $B \in M_{m,n}$ has rank at most k. Then there is a unit vector $x = (a_1, \dots, a_k)^t$ such that $B[v_1|\cdots|v_{k+1}]x = 0$ because $B[v_1|\dots|v_{k+1}]$ has rank at most k. Then $\|x\| = 1$, $\|y\| = 1$ for $y = [v_1|\cdots|v_{k+1}]x$ and

$$\begin{aligned} \|A - B\|_2 &\ge \|(A - B)y\| = \|A(a_1 v_1, \dots, a_{k+1} v_{k+1})\| \\ &= \|\sum_{j=1}^{k+1} s_j a_j u_j\| = \{\sum_{j=1}^{k+1} (s_j^2 |a_j|^2)\}^{1/2} \\ &\ge s_{j+1} = \|A - \sum_{j=1}^k s_j u_j v_j^*\|_s. \end{aligned}$$

Here, we get $\|A - B\| = s_{k+1}$ if $B = \sum_{j=1}^k s_j u_j v_j^*$. $\square$

Proof for the Frobenius norm. Suppose B is the optimizer and has singular value decomposition $U_B D_B V_B^*$, where D_B has nonnegative entries $d_1, \dots, d_k$ in the $(1,1), \dots, (k,k)$ positions. Then

$$\|A - B\|_F = \|U_B^*(U_B A V_B^* - D_B)V_B^*\|_F.$$

If $\hat{A} = U_B A V_B^* = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}$, then

$$\|A - B\|_F^2 = \|\hat{A} - D_B\|_F^2 = \|A_{11} - \operatorname{diag}(d_1, \dots d_k)\|_F^2 + \|A_{12}\|_F^2 + \|A_{21}|_F^2 + \|A_{22}\|_F^2.$$

As a result, we must have $\hat{A}_{11} = \operatorname{diag}(d_1, \dots, d_n)$, $A_{21} = 0$, and $A_{21} = 0$. Otherwise,

$$B_1 = U_B^* \begin{pmatrix} A_{11} & A_{12} \\ 0 & 0 \end{pmatrix} \qquad \text{or } B_2 = U_B^* \begin{pmatrix} A_{11} & 0 \\ A_{21} & 0 \end{pmatrix}$$

will be better approximation because B_1, B_2 have rank at most k,

$$\|A - B_1\|_F^2 = \|A_{21}\|_F^2 + \|A_{22}\|^2 \quad \text{and} \quad \|A - B_2\|_F^2 = \|A_{12}\|_F^2 + \|A_{22}\|^2.$$

Thus, $A_{11} = \operatorname{diag}(d_1, \dots, d_k)$ and $d_1, \dots, d_k$ are the singular values of A, and

$$\|A - B\|_F^2 = \sum_{j=1}^{r} s_j^2 - \sum_{j=1}^{k} d_k^2,$$

which will be minimum if we choose $d_1, \dots, d_k$ to be the k largest singular values of A. □

Principal Component Analysis

- Let A be $m \times n$, each column is a sample with m measurements.
- Normalize the means to 0 for all measurement. So, sum of columns is the zero vector in $\mathbb{R}^m$.
- The variances are the diagonal entries of $AA^T \in M_m$.
- The co-variances are the off-diagonal entries of $AA^T \in M_n$.
- The sample co-variance matrix is $S = AA^T/(n-1)$.
- For example, when $m = 2$, the best rank one approximation of S is $s_1 u_1 v_1^t$. So, the slop of the line is the ratio of the second entry of u_1 to the first entry of u_1.
- Note that the line is different from finding the best fit $y = ax + b$.
- The total variance is $\mathrm{t}r\, A^T A/n-1) = \mathrm{t}r\, S/(n-1)$.