14. Singular value decomposition

- eigenvalues and diagonalization
- symmetric eigendecomposition
- singular value decomposition
- low-rank approximation
- least squares via SVD

Eigenvalues and eigenvectors

vector $x \neq 0$ is an *eigenvector* of the $n \times n$ matrix A associated with *eigenvalue* λ if

$$Ax = \lambda x$$

- together, (λ, x) is an *eigenpair*; set of all eigenvalues is called *spectrum* of A
- matrix expands/shrinks any vector lying in eigenvector direction by a scalar
- eigenvalues are useful in analyzing numerical methods
  - analysis of iterative methods for solving systems of equations and optimization problems
  - analysis of numerical methods for solving differential equations

Left eigenvector

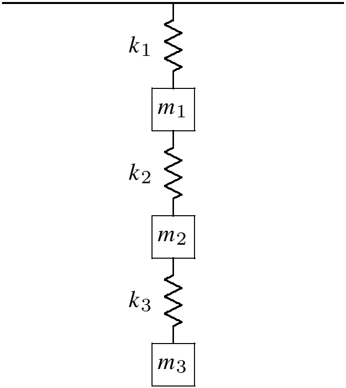
- w is a left *eigenvector*, associated with eigenvalue λ if $w^T A = \lambda w^T$
- a left eigenvector of A is a (right) eigenvector of A^T

Example: mass-spring system

$$M\frac{d^2y(t)}{dt^2} + Ky(t) = 0$$

$$M = \begin{bmatrix} m_1 & 0 & 0 \\ 0 & m_2 & 0 \\ 0 & 0 & m_3 \end{bmatrix}$$

$$K = \begin{bmatrix} k_1 + k_2 & -k_2 & 0 \\ -k_2 & k_2 + k_3 & -k_3 \\ 0 & -k_3 & k_3 \end{bmatrix}$$



- y_1, y_2, and y_3 are vertical displacements
- k_1, k_2, and k_3 are spring constants
- M is the mass matrix; K is the stiffness matrix

Example: mass-spring system

- system exhibits simple harmonic motion with natural frequency ω:

$$y_j(t) = x_j e^{\imath \omega t}, \quad j = 1, 2, 3$$

  where x_j is the amplitude or mode of vibration

- to determine the frequency ω and x_j, we note that

$$\frac{d^2 y_j(t)}{dt^2} = -\omega^2 x_j e^{\imath \omega t}$$

- substituting into the differential equation, we obtain the algebraic equation

$$Kx = \omega^2 Mx$$

- an eigenvalue problem $Ax = \lambda x$ with $A = M^{-1}K$ and $\lambda = \omega^2$

Characteristic polynomial

we can write the eigenvalue problem $Ax = \lambda x$ as a homogeneous linear system

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

- this means that $\lambda I - A$ must be singular and $x \neq 0$ is in the nullspace of $\lambda I - A$
- we can find λ by finding the roots of the *characteristic polynomial*:

$$p(\lambda) = \det(\lambda I - A) = \lambda^n + c_{n-1}\lambda^{n-1} + \cdots + c_1\lambda + (-1)^n \det(A) = 0$$

  which is a polynomial of degree n; $p(\lambda)$ has n roots counting multiplicities

- eigenvalues (and eigenvectors) can be complex even if A is real
  - complex eigenvalues of real A appear as conjugate pairs
- eigenvalues are typically computed using an iterative process
  - no closed-form formula exists for a polynomial of degree greater than or equal to 4

Cayley-Hamilton theorem: every $n \times n$ matrix A satisfies its characteristic equation

$$p(A) = A^n + c_{n-1}A^{n-1} + \cdots + c_1 A + c_0 I = 0$$

Eigenvalues of a 2×2 matrix

for a 2×2 matrix, the characteristic equation is

$$\begin{aligned} \det(\lambda I - A) &= \det \begin{bmatrix} \lambda - A_{11} & -A_{12} \\ -A_{21} & \lambda - A_{22} \end{bmatrix} \\ &= \lambda^2 - (A_{11} + A_{22})\,\lambda + (A_{11}A_{22} - A_{12}A_{21}) \end{aligned}$$

we therefore have to solve a quadratic equation of the form

$$\lambda^2 - b\lambda + c = 0$$

solving gives

$$\lambda_{1,2} = \frac{1}{2}(b \pm \sqrt{\Delta}) = \frac{1}{2}\left(A_{11} + A_{22} \pm \sqrt{\Delta}\right)$$

where $\Delta = b^2 - 4c = (A_{11} - A_{22})^2 + 4A_{12}A_{21}$

- if $\Delta > 0$, then there are two real eigenvalues
- if $\Delta = 0$, then there is the double real eigenvalue
- if $\Delta < 0$, then there are two complex eigenvalues

Example

- for the matrix

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

  - $\Delta = 3^2 + 4^2 = 25$
  - $\lambda_1 = (5 + 5)/2 = 5$
  - $\lambda_2 = (5 - 5)/2 = 0$

- for the matrix

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

  - $\Delta = -4$
  - $\lambda_1 = \imath$
  - $\lambda_2 = -\imath$

Some properties

let $A \in \mathbb{R}^{n \times n}$ with eigenvalues $\lambda_1, \ldots, \lambda_n$

- if x is an eigenvector, then for any scalar $\gamma \neq 0$, γx is also an eigenvector
- eigenvalues of $A + \alpha I$ are $\lambda_1 + \alpha, \ldots, \lambda_n + \alpha$
- eigenvalues of A^k are $\lambda_1^k, \ldots, \lambda_n^k$
- eigenvalues of A^{-1} are $1/\lambda_1, \ldots, 1/\lambda_n$
- eigenvalues of A^T are equal to the eigenvalues of A
- if A is a triangular matrix, then its eigenvalues are equal to its diagonal element
- if $x_1, \ldots, x_k$ are eigenvectors for k different eigenvalues:

  $$Ax_1 = \lambda_1 x_1, \quad \ldots, \quad Ax_k = \lambda_k x_k$$

  then $x_1, \ldots, x_k$ are linearly independent (converse is not true)
- $\rho(A) \leq \|A\|$ for any *induced norm*; $\rho(A) = \max_i |\lambda_i|$ is the *spectral radius*

Gershgorin's theorem

the eigenvalues of an $n \times n$ matrix A are all contained within the union of n disks:

$$D_i = \{\gamma \mid |\gamma - A_{ii}| \leq \sum_{j \neq i} |A_{ij}|\}, \quad i = 1, \ldots, n$$

- to see this, let λ be any eigenvalue with corresponding eigenvector x
- assume x is normalized so that $\|x\|_\infty = 1$ with $|x_i| = 1$
- because $Ax = \lambda x$, we have

$$(\lambda - A_{ii})\, x_i = \sum_{j \neq i} A_{ij} x_j$$

so that

$$|\lambda - A_{ii}| \leq \sum_{j \neq i} |A_{ij}|\, |x_j| \leq \sum_{j \neq i} |A_{ij}|$$

Similar matrices

two $n \times n$ matrices A and B are *similar* if there exists a nonsingular matrix T such that

$$T^{-1}AT = B$$

- the mapping $A \to T^{-1}AT$ is called a *similarity transformation*
- similar matrices have the same eigenvalues:

$$\det(\lambda I - B) = \det(\lambda I - T^{-1}AT) = \det(T^{-1}(\lambda I - A)T) = \det(\lambda I - A)$$

- if x is an eigenvector of A then $y = T^{-1}x$ is an eigenvector of B:

$$By = (T^{-1}AT)(T^{-1}x) = T^{-1}Ax = T^{-1}(\lambda x) = \lambda y$$

Diagonalizable matrices

if (λ_j, x_j) is an eigenpair, then

$$\begin{aligned} AX &= A\,[x_1 \;\; x_2 \;\; \cdots \;\; x_n] \\ &= [\lambda_1 x_1 \;\; \lambda_2 x_2 \;\; \cdots \;\; \lambda_n x_n] \\ &= X\Lambda \end{aligned}$$

where $\Lambda = \mathrm{diag}(\lambda_1, \lambda_2, \ldots, \lambda_n)$

Diagonalization: if the eigenvectors are linearly independent, then

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

- A is similar to a diagonal matrix $X^{-1}AX = \Lambda$
- this decomposition is the *eigendecomposition* of A
- not all square matrices are diagonalizable

Non-diagonalizable matrices

Algebraic multiplicity

- *algebraic multiplicity* of an eigenvalue is its multiplicity as a root of $\det(\lambda I - A)$
- the sum of the algebraic multiplicities of the eigenvalues of an $n \times n$ matrix is n

Geometric multiplicity

- *geometric multiplicity* is is the dimension of $\text{null}(\lambda I - A)$

  i.e., maximum number of linearly independent eigenvectors with eigenvalue λ
- sum is the maximum number of linearly independent eigenvectors of the matrix
- geometric multiplicity $\leq$ algebraic multiplicity

Non-diagonalizable matrices

- eigenvalue is *defective* if geometric multiplicity is less than algebraic multiplicity
- a matrix is *defective* if some of its eigenvalues are defective
- a defective matrix is not diagonalizable

Example

consider

$$A = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 2 \end{bmatrix}, \qquad \det(\lambda I - A) = (\lambda - 1)^2(\lambda - 2)^2$$

- there are two eigenvalues, $\lambda = 1$ and $\lambda = 2$, each with algebraic multiplicity two
- The eigenvalue $\lambda = 1$ is defective and has geometric multiplicity one:

$$\text{null}(I - A) = \text{span}\{(1, 0, 0, 0)\}$$

- eigenvalue $\lambda = 2$ is not defective and has geometric multiplicity two:

$$\text{null}(2I - A) = \text{span}\{(0, 0, 1, 0),\ (0, 0, 0, 1)\}$$

- the maximum number of linearly independent eigenvectors is three; for example,

$$(1, 0, 0, 0), \qquad (0, 0, 1, 0), \qquad (0, 0, 0, 1)$$

Outline

- eigenvalues and diagonalization
- **symmetric eigendecomposition**
- singular value decomposition
- low-rank approximation
- least squares via SVD

Symmetric matrix eigenvalues and eigenvectors

a symmetrix real matrix $A \in \mathbb{R}^{n \times n}$ satisfies the following properties

- all eigenvalues of A are real (and eigenvectors can assumed to be real)
- A has n linearly independent eigenvectors
- eigenvectors corresponding to different eigenvalues are orthogonal
- there is a set of n orthonormal eigenvectors of A

Symmetric eigendecomposition

let $A \in \mathbb{R}^{n\times n}$ be symmetric, then it can be factored as

$$A = Q\Lambda Q^T = \sum_{i=1}^{n} \lambda_i q_i q_i^T$$

- $Q \in \mathbb{R}^{n\times n}$ is orthogonal ($Q^TQ = QQ^T = I$)
- $\Lambda = \mathrm{diag}(\lambda_1, \ldots, \lambda_n)$ is diagonal with real diagonal elements
- A is diagonalizable by an orthogonal similarity transformation: $Q^TAQ = \Lambda$
- columns of Q forms an orthonormal set of eigenvectors of A
- this factorization is also called *spectral decomposition*
- not unique: some freedom exists in the choice of Λ and Q
  - diagonal elements of Λ and columns of Q can be permuted
  - k eigenvectors of eigenvalues with algebraic multiplicity k can be replaced by any orthonormal basis of their span

Range, nullspace, rank

for a symmetric matrix A, arrange the eigendecomposition so that

$$A = Q\Lambda Q^T = \begin{bmatrix} Q_1 & Q_2 \end{bmatrix} \begin{bmatrix} \Lambda_1 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} Q_1^T \\ Q_2^T \end{bmatrix} = Q_1 \Lambda_1 Q_1^T$$

where the diagonal entries of Λ_1 are the nonzero eigenvalues of A

- columns of Q_1 form an orthonormal basis for $\mathrm{range}(A)$
- columns of Q_2 form an orthonormal basis for $\mathrm{null}(A)$
- this is an example of a full-rank factorization:

$$A = BC, \qquad B = Q_1, \qquad C = \Lambda_1 Q_1^T$$

- the rank of A equals the number of nonzero eigenvalues, counted with multiplicity

Pseudo-inverse

we use the same notation as on the previous slide:

$$A = \begin{bmatrix} Q_1 & Q_2 \end{bmatrix} \begin{bmatrix} \Lambda_1 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} Q_1^T \\ Q_2^T \end{bmatrix} = Q_1 \Lambda_1 Q_1^T$$

where the diagonal entries of Λ_1 are the nonzero eigenvalues of A

- pseudo-inverse follows from full-rank factorization with $B = Q_1$ and $C = \Lambda_1 Q_1^T$
- the pseudo-inverse is $A^\dagger = C^\dagger B^\dagger = (Q_1 \Lambda_1^{-1}) Q_1^T$ or

$$A^\dagger = Q_1 \Lambda_1^{-1} Q_1^T = \begin{bmatrix} Q_1 & Q_2 \end{bmatrix} \begin{bmatrix} \Lambda_1^{-1} & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} Q_1^T \\ Q_2^T \end{bmatrix}$$

- the eigenvectors of $A^\dagger$ are the same as the eigenvectors of A
- the nonzero eigenvalues of $A^\dagger$ are the reciprocals of the nonzero eigenvalues of A
- the range, nullspace, and rank of $A^\dagger$ are the same as for A

Frobenius norm

recall the definition of the Frobenius norm of an $m \times n$ matrix B:

$$\|B\|_F = \left(\sum_{i=1}^{m} \sum_{j=1}^{n} B_{ij}^2 \right)^{1/2}$$

- this is a unitarily invariant norm: if U and V are orthogonal, then

$$\|UBV\|_F = \|B\|_F$$

- for a symmetric $n \times n$ matrix A with eigenvalues $\lambda_1, \ldots, \lambda_n$,

$$\|A\|_F = \|Q\Lambda Q^T\|_F = \|\Lambda\|_F = \left(\sum_{i=1}^{n} \lambda_i^2 \right)^{1/2}$$

Spectral norm (2-norm)

recall the definition of the 2-norm, or spectral norm, of an $m \times n$ matrix B:

$$\|B\|_2 = \max_{x \neq 0} \frac{\|Bx\|}{\|x\|}$$

- this norm is also unitarily invariant: if U and V are orthogonal, then

$$\|UBV\|_2 = \|B\|_2$$

- for a symmetric $n \times n$ matrix A with ordered eigenvalues $\lambda_1, \ldots, \lambda_n$,

$$\|A\|_2 = \|Q \Lambda Q^T\|_2 = \|\Lambda\|_2 = \max_{i=1,\ldots,n} |\lambda_i| = \max\{\lambda_1, -\lambda_n\}$$

Inequalities for quadratic forms

for symmetric matrix A, we have

$$\lambda_n \|x\|^2 \le x^T A x \le \lambda_1 \|x\|^2$$

where λ_1 is the largest eigenvalue and λ_n is the smallest

- to see this, consider the eigendecomposition

$$A = Q\Lambda Q^T, \qquad \Lambda = \mathrm{diag}(\lambda_1, \ldots, \lambda_n), \qquad \lambda_1 \ge \lambda_2 \ge \cdots \ge \lambda_n$$

  then

$$x^T A x = x^T Q \Lambda Q^T x = (Q^T x)^T \Lambda (Q^T x) = \sum_{i=1}^{n} \lambda_i (q_i^T x)^2$$

$$\le \lambda_1 \sum_{i=1}^{n} (q_i^T x)^2 = \lambda_1 \|x\|^2$$

- a similar argument shows that $x^T A x \ge \lambda_n \|x\|^2$

- the bounds are tight, since

$$q_1^T A q_1 = \lambda_1 \|q_1\|^2 \qquad q_n^T A q_n = \lambda_n \|q_n\|^2$$

Sign of eigenvalues and definiteness

- let $A \in \mathbb{R}^{n \times n}$ be symmetric
- let λ_1 and λ_n denote the largest and smallest eigenvalues:

$$\lambda_1 = \max_{i=1,\ldots,n} \lambda_i, \qquad \lambda_n = \min_{i=1,\ldots,n} \lambda_i$$

- sign pattern of its eigenvalues determines the properties:

property	condition on eigenvalues
positive definite	$\lambda_n > 0$
positive semidefinite	$\lambda_n \geq 0$
indefinite	$\lambda_n < 0$ and $\lambda_1 > 0$
negative semidefinite	$\lambda_1 \leq 0$
negative definite	$\lambda_1 < 0$

Outline

- eigenvalues and diagonalization
- symmetric eigendecomposition
- **singular value decomposition**
- low-rank approximation
- least squares via SVD

Singular value decomposition (SVD)

every $m \times n$ matrix A can be factored as

$$A = U\Sigma V^T$$

- U is $m \times m$ and orthogonal, V is $n \times n$ and orthogonal
- Σ is $m \times n$ and "diagonal":

$$\Sigma = \mathrm{diag}(\sigma_1, \ldots, \sigma_n) \qquad \text{if } m = n$$

$$\Sigma = \begin{bmatrix} \mathrm{diag}(\sigma_1, \ldots, \sigma_n) \\ 0_{(m-n)\times n} \end{bmatrix} \qquad \text{if } m > n$$

$$\Sigma = \begin{bmatrix} \mathrm{diag}(\sigma_1, \ldots, \sigma_m) & 0_{m\times(n-m)} \end{bmatrix} \qquad \text{if } m < n$$

- diagonal entries of Σ are nonnegative and ordered:

$$\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_{\min\{m,n\}} \geq 0$$

- in MATLAB, the command is `[U,Sigma,V] = svd(A)`

Example

$$A = \begin{bmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{bmatrix} = U\Sigma V^T$$

$$U = \begin{bmatrix} 0.229847696400071 & 0.883461017698525 & -0.408248290463864 \\ 0.524744818760294 & 0.240782492132547 & 0.816496580927726 \\ 0.819641941120516 & -0.401896033433433 & -0.408248290463863 \end{bmatrix}$$

$$\Sigma = \begin{bmatrix} 9.525518091565107 & 0 \\ 0 & 0.514300580658642 \\ 0 & 0 \end{bmatrix}$$

$$V = \begin{bmatrix} 0.619629483829340 & -0.784894453267053 \\ 0.784894453267053 & 0.619629483829340 \end{bmatrix}$$

Singular values and singular vectors

$$A = U\Sigma V^T$$

- columns of U are called *left singular vectors of* A
- columns of V are *right singular vectors of* A
- numbers σ_i are the *singular values of* A

if we write the factorization $A = U\Sigma V^T$ as

$$AV = U\Sigma, \quad A^T U = V\Sigma^T$$

and compare the ith columns on the left- and right-hand sides, we see that

$$Av_i = \sigma_i u_i \quad \text{and} \quad A^T u_i = \sigma_i v_i \quad \text{for } i = 1, \ldots, \min\{m, n\}$$

- if $m > n$ the additional $m - n$ vectors u_i satisfy $A^T u_i = 0$ for $i = n + 1, \ldots, m$
- if $n > m$ the additional $n - m$ vectors v_i satisfy $Av_i = 0$ for $i = m + 1, \ldots, n$

Reduced SVD

Tall matrix: if $m > n$, the last $m - n$ columns of U can be omitted to define

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

- U is $m \times n$ with orthonormal columns
- V is $n \times n$ and orthogonal
- Σ is $n \times n$ and diagonal with diagonal entries $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_n \geq 0$

Wide matrix: if $m < n$, the last $n - m$ columns of V can be omitted to define

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

- U is $m \times m$ and orthogonal
- V is $m \times n$ with orthonormal columns
- Σ is $m \times m$ and diagonal with diagonal entries $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_m \geq 0$

Eigendecomposition of Gram matrix

suppose A is an $m \times n$ matrix with (full) SVD

$$A = U\Sigma V^T$$

the SVD is related to the eigendecomposition of the Gram matrix A^TA:

$$A^TA = V\Sigma^T\Sigma V^T = \sum_{i=1}^{\min\{m,n\}} \sigma_i^2 v_i v_i^T$$

- V is an orthogonal $n \times n$ matrix
- $\Sigma^T\Sigma$ is a diagonal $n \times n$ matrix with (non-increasing) diagonal elements

$$\sigma_1^2,\ \sigma_2^2,\ \cdots,\ \sigma_{\min\{m,n\}}^2,\quad \underbrace{0,\ 0,\ \ldots,\ 0}_{n-\min\{m,n\} \text{ zeros}}$$

- the n diagonal elements of $\Sigma^T\Sigma$ are the eigenvalues of A^TA
- the right singular vectors (columns of V) are corresponding eigenvectors

Eigendecomposition of Gram matrix of transpose

the SVD also gives the eigendecomposition of AA^T:

$$AA^T = U\Sigma\Sigma^T U^T = \sum_{i=1}^{\min\{m,n\}} \sigma_i^2 u_i u_i^T$$

- U is an orthogonal $m \times m$ matrix
- $\Sigma\Sigma^T$ is a diagonal $m \times m$ matrix with (non-increasing) diagonal elements

$$\sigma_1^2,\ \sigma_2^2,\ \ldots,\ \sigma_{\min\{m,n\}}^2, \qquad \underbrace{0,\ 0,\ \cdots,\ 0}_{m-\min\{m,n\}\text{ zeros}}$$

- the m diagonal elements of $\Sigma\Sigma^T$ are the eigenvalues of AA^T
- the left singular vectors (columns of U) are corresponding eigenvectors

in particular, the first $\min\{m,n\}$ eigenvalues of A^TA and AA^T are the same:

$$\sigma_1^2, \quad \sigma_2^2, \quad \ldots, \quad \sigma_{\min\{m,n\}}^2$$

Rank and compact SVD

the rank of a matrix is equal to the number of positive singular values

Compact-form of SVD: partition the matrices in a full SVD of A as

$$
\begin{aligned}
A &= \begin{bmatrix} U_1 & U_2 \end{bmatrix} \left[\begin{array}{c|c} \Sigma_1 & 0_{r\times(n-r)} \\ \hline 0_{(m-r)\times r} & 0_{(m-r)\times(n-r)} \end{array}\right] \begin{bmatrix} V_1 & V_2 \end{bmatrix}^T \\
&= U_1 \Sigma_1 V_1^T = \sum_{i=1}^{r} \sigma_i u_i v_i^T
\end{aligned}
$$

- $\sigma_1, \ldots, \sigma_r$ are r strictly positive singular values:

$$\sigma_1 \geq \cdots \geq \sigma_r > 0 = \sigma_{r+1} = \cdots = \sigma_{\min\{m,n\}}$$

- $\Sigma_1 = \mathrm{diag}(\sigma_1, \ldots, \sigma_r)$ is $r \times r$ diagonal with the +ve singular values on diagonal
- U_1 is $m \times r$ and V_1 is $n \times r$; both have orthonormal columns
- hence $\mathrm{rank}(A) = r$ (see page 4.43)

Pseudo-inverse

consider the SVD of A with same notation in previous page:

$$A = \begin{bmatrix} U_1 & U_2 \end{bmatrix} \begin{bmatrix} \Sigma_1 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} V_1 & V_2 \end{bmatrix}^T = U_1 \Sigma_1 V_1^T$$

Pseudo-inverse

$$A^\dagger = V_1 \Sigma_1^{-1} U_1^T = \begin{bmatrix} V_1 & V_2 \end{bmatrix} \begin{bmatrix} \Sigma_1^{-1} & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} U_1^T \\ U_2^T \end{bmatrix}$$

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

- follows from page 4.44
- if A is square and nonsingular, this reduces to the inverse $A^{-1} = V\Sigma^{-1}U^T$
- if A is tall with linearly independent columns, $A^\dagger = (A^TA)^{-1}A^T$
- if A is wide with linearly independent rows, $A^\dagger = A^T(AA^T)^{-1}$

Four subspaces from the SVD

we continue with the same notation for the SVD of an $m \times n$ matrix A with rank r:

$$A = \begin{bmatrix} U_1 & U_2 \end{bmatrix} \begin{bmatrix} \Sigma_1 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} V_1 & V_2 \end{bmatrix}^T$$

where the diagonal entries of Σ_1 are the positive singular values of A

the SVD provides orthonormal bases for the four subspaces associated with A:

- columns of the $m \times r$ matrix U_1 form a basis for $\mathrm{range}(A)$
- columns of the $m \times (m-r)$ matrix U_2 form a basis for $\mathrm{range}(A)^\perp = \mathrm{null}(A^T)$
- columns of the $n \times r$ matrix V_1 form a basis for $\mathrm{range}(A^T)$
- columns of the $n \times (n-r)$ matrix V_2 form a basis for $\mathrm{null}(A)$

Frobenius norm and 2-norm

for an $m \times n$ matrix A with singular values σ_i :

$$\|A\|_F = \left(\sum_{i=1}^{\min\{m,n\}} \sigma_i^2 \right)^{1/2}, \quad \|A\|_2 = \sigma_1$$

this follows from the norm invariance properties

$$\|A\|_F = \left\| U \Sigma V^T \right\|_F = \|\Sigma\|_F = \left(\sum_{i=1}^{\min\{m,n\}} \sigma_i^2 \right)^{1/2}$$

and

$$\|A\|_2 = \left\| U \Sigma V^T \right\|_2 = \|\Sigma\|_2 = \sigma_1$$

Condition number

let A be an $m \times n$ matrix with $\mathrm{rank}(A) = n$ and full SVD

$$A = U \Sigma V^T$$

- the norm of A is $\|A\| = \|U \Sigma V^T\| = \|\Sigma\| = \sigma_1$
- the norm of $A^\dagger$ is $\|A^\dagger\| = 1/\sigma_n$
- the condition number of A is the ratio of largest to smallest singular values

$$\kappa(A) = \frac{\sigma_1}{\sigma_n}$$

- for an ill-conditioned matrix the smallest singular values are very small

Outline

- eigenvalues and diagonalization
- symmetric eigendecomposition
- singular value decomposition
- **low-rank approximation**
- least squares via SVD

Rank-r approximation

let A be an $m \times n$ matrix with $\mathrm{rank}(A) > r$ and full SVD

$$A = U\Sigma V^T = \sum_{i=1}^{\min\{m,n\}} \sigma_i u_i v_i^T, \quad \sigma_1 \geq \cdots \geq \sigma_{\min\{m,n\}} \geq 0, \quad \sigma_{r+1} > 0$$

the best rank-r approximation of A is the sum of the first r terms in the SVD:

$$B = \sum_{i=1}^{r} \sigma_i u_i v_i^T$$

- B is the best approximation for the Frobenius norm: for every C with rank r,

$$\|A - C\|_F \geq \|A - B\|_F = \left(\sum_{i=r+1}^{\min\{m,n\}} \sigma_i^2 \right)^{1/2}$$

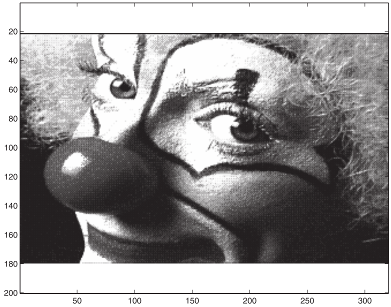
- B is also the best approximation for the 2-norm: for every C with rank r,

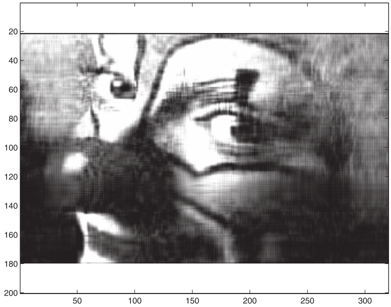
$$\|A - C\|_2 \geq \|A - B\|_2 = \sigma_{r+1}$$

Image compression

rank-r approximation makes it possible to devise a compression scheme:

- by storing the first r columns of U and V, as well as the first r singular values
- we obtain an approximation of A using only $r(m+n+1)$ locations instead of mn





- image of size $200 \times 320 = 64,000$
- rank-20 SVD approximation of size $20 \times (200 + 320 + 1) \approx 10,000$

Matlab code

```
colormap('gray')
load clown.mat;
figure(1)
image(X);
[U,S,V] = svd(X);
figure(2)
r = 20;
colormap('gray')
image(U(:,1:r)*S(1:r,1:r)*V(:,1:r)');
```

Document analysis

Task: identify documents that are relevant to a user's query

a collection of documents is represented by a *term–document matrix* $A \in \mathbb{R}^{n \times m}$

- n is the number of terms (words) and m is the number of documents
- each row corresponds to a word in a dictionary
- each column corresponds to a document
- entry i, j represents frequency of word i in document j, usually weighted as

$$A_{ij} = f_{ij} \log(m/m_i)$$

  - f_{ij} is frequency of term i in document j
  - m_i is number of documents that contain term i

Comparing documents and queries

Comparing documents: as a measure of document similarity, we can use

$$\frac{a_i^T a_j}{\|a_i\| \, \|a_j\|}$$

- a_i and a_j are the columns of A corresponding to documents i and j
- this is called the *cosine similarity*: the cosine of the angle between a_i and a_j

Query matching: we find the most relevant documents based on keywords in a query

- we treat the query as a simple document, represented by an n-vector x:

$$x_j = \begin{cases} 1, & \text{if term } j \text{ appears in the query} \\ 0, & \text{otherwise} \end{cases}$$

- we rank documents according to their cosine similarity with x:

$$\cos(\theta_i) = \frac{a_i^T x}{\|a_i\| \, \|x\|}, \qquad i = 1, \ldots, m$$

Example

two one-sentence documents:

"Numerical computations are fun."

"Numerical algorithms and numerical methods are interesting."

words:

1. "algorithms"
2. "computations"
3. "fun"
4. "interesting"
5. "methods"
6. "numerical"

term-document matrix

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

- we define 6-long vectors for representing queries; for example, the query "numerical methods" is represented as $x = (0, 0, 0, 0, 1, 1)$
- two documents deal with a similar issue
- but overlap between the two columns in the matrix A is relatively small

Reduced space

- truncated SVD $A \approx U_r \Sigma_r V_r^T$ with r small
- then cosine similarity in reduced space

$$\frac{(\Sigma_r V_r^T e_i)^T (U_r^T x)}{\|\Sigma_r V_r^T e_i\| \|x\|}$$

  where e_i is ith unit vector

- in the reduced space, the n-long query x is transformed into the r-long vector

$$\tilde{x} = U_r^T x$$

- called *latent semantic indexing* (LSI) (an early technique for search engines)

Outline

- eigenvalues and diagonalization
- symmetric eigendecomposition
- singular value decomposition
- low-rank approximation
- **least squares via SVD**

Least squares with full column rank

$$\text{minimize} \quad \|Ax - b\|^2, \quad A \in \mathbb{R}^{m\times n},\ b \in \mathbb{R}^m$$

suppose that A has full column rank with SVD

$$A = U\Sigma V^T = \begin{bmatrix} U_1 & U_2 \end{bmatrix} \begin{bmatrix} \Sigma_1 \\ 0_{(m-n)\times n} \end{bmatrix} \begin{bmatrix} V_1 & V_2 \end{bmatrix}^T$$

where $\Sigma_1 = \mathrm{diag}(\sigma_1, \ldots, \sigma_n)$; observe that

$$\begin{aligned} \|Ax - b\|^2 = \|U\Sigma V^T x - b\|^2 &= \|\Sigma V^T x - U^T b\|^2 \\ &= \left\| \begin{bmatrix} \Sigma_1 \\ 0 \end{bmatrix} V^T x - \begin{bmatrix} U_1^T b \\ U_2^T b \end{bmatrix} \right\|^2 \\ &= \left\| \begin{bmatrix} \Sigma_1 V^T x - U_1^T b \\ -U_2^T b \end{bmatrix} \right\|^2 \\ &= \|\Sigma_1 V^T x - U_1^T b\|^2 + \|U_2^T b\|^2 \end{aligned}$$

the above is minimized when

$$\hat{x} = V\Sigma_1^{-1} U_1^T b = A^\dagger b \quad \text{where } A^\dagger = V\Sigma_1^{-1} U_1^T = (A^T A)^{-1} A^T$$

Rank deficient least squares

$$\text{minimize} \quad \|Ax - b\| = \left\|U\Sigma V^T x - b\right\|$$

- A is $m \times n$ and rank deficient $\operatorname{rank}(A) = r < \min\{m, n\}$
- SVD of A is

$$A = \begin{bmatrix} U_1 & U_2 \end{bmatrix} \begin{bmatrix} \Sigma_1 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} V_1 & V_2 \end{bmatrix}^T$$

- the $m \times n$ matrix Σ has only r diagonal entries $\sigma_1, \ldots, \sigma_r$
- first we introduce the change of variables

$$z = V^T x \iff x = \begin{bmatrix} V_1 & V_2 \end{bmatrix} \begin{bmatrix} z_1 \\ z_2 \end{bmatrix}$$

where $z_1 \in \mathbb{R}^r$ and $z_2 \in \mathbb{R}^{n-r}$

Minimum norm solution

$$\begin{aligned}\|Ax-b\|^2 = \|U\Sigma V^T x - b\|^2 &= \|\Sigma V^T x - U^T b\|^2 \\ &= \left\| \begin{bmatrix} \Sigma_1 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} z_1 \\ z_2 \end{bmatrix} - \begin{bmatrix} U_1^T b \\ U_2^T b \end{bmatrix} \right\|^2 \\ &= \left\| \begin{bmatrix} \Sigma_1 z_1 - U_1^T b \\ -U_2^T b \end{bmatrix} \right\|^2 \\ &= \|\Sigma_1 z_1 - U_1^T b\|^2 + \|U_2^T b\|^2\end{aligned}$$

this is minimized by taking $z_1 = \Sigma_1^{-1} U_1^T b$; plugging back into x gives

$$x = \begin{bmatrix} V_1 & V_2 \end{bmatrix} \begin{bmatrix} z_1 \\ z_2 \end{bmatrix} = V_1 z_1 + V_2 z_2 = V_1 \Sigma_1^{-1} U_1^T b + V_2 z_2$$

- $z_2 \in \mathbb{R}^{n-r}$ is arbitrary and $V_2 z_2$ is in nullspace of A
- solution with least norm $\|x\|^2 = \|Vz\|^2 = \|z\|^2$ is obtained when $z_2 = 0$:

$$\hat{x} = V\Sigma^\dagger U^T b = V_1 \Sigma_1^{-1} U_1^T b = \sum_{i=1}^{r} \frac{1}{\sigma_i} v_i (u_i^T b) = A^\dagger b$$

Truncated SVD for least squares

---

given an $m \times n$ matrix A and threshold ϵ

1. form $A = U\Sigma V^T$
2. decide on truncation r such that $\sigma_k > \epsilon$ for $k = 1, \ldots, r$
3. compute $x = \sum_{i=1}^{r} \frac{1}{\sigma_i} v_i (u_i^T b)$

Complexity

- forming the SVD, costs approximately $2mn^2 + 11n^3$ flops
- for $m \gg n$ this is approximately the same cost as the QR-based approach $2mn^2$
- for $m \approx n$ or $n \geq m$ the SVD approach is more expensive

References and further readings

- L. Vandenberghe. *EE133B Lecture Notes*, University of California, Los Angeles.
- M. T. Heath. *Scientific Computing: An Introductory Survey* (revised second edition). Society for Industrial and Applied Mathematics, 2018.
- U. M. Ascher. *A First Course on Numerical Methods*. Society for Industrial and Applied Mathematics, 2011.