15. Singular value decomposition

- eigenvalues and diagonalization
- singular value decomposition
- SVD and matrix properties
- least squares via SVD

Eigenvalues and eigenvectors

scalar λ is an *eigenvalue* of a square $n \times n$ matrix A if

$$Ax = \lambda x \quad \text{for} \quad x \neq 0$$

- x is an *eigenvector* associated with eigenvalue λ
- 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

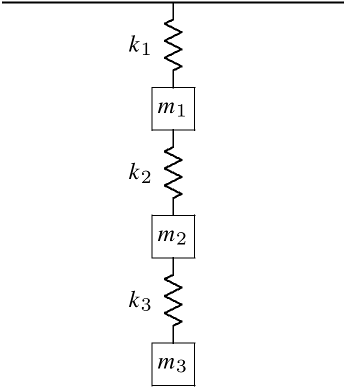
- w is a left *eigenvector*, associated with eigenvalue λ, if $w^T A = \lambda w^T$
- a left eigenvector of A is a (right) eigenvector of A^T

Example: mass-spring system

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

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

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



- k_1, k_2, and k_3 are spring constants
- y_1, y_2, and y_3 are vertical displacements
- 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

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

- 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

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

  since we want a nontrivial x, this means that $\lambda I - A$ must be singular

- we can find λ by finding the roots of the *characteristic polynomial*:

$$p(\lambda) = \det(\lambda I - A) = (\lambda - \lambda_1)(\lambda - \lambda_2)\cdots(\lambda - \lambda_n) = 0$$

  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:** A satisfies its own characteristic equation

$$p(A) = A^n + \alpha_{n-1}A^{n-1} + \cdots + \alpha_1 A + \alpha_0 I = 0$$

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 γx is an also an eigenvector for any scalar $\gamma \neq 0$
- 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 $v_1, \ldots, v_k$ are eigenvectors for k different eigenvalues:

$$Av_1 = \lambda_1 v_1, \quad \ldots, \quad Av_k = \lambda_k v_k$$

  then $v_1, \ldots, v_k$ are linearly independent
- $\rho(A) \leq \|A\|$ for any *induced* norm; $\rho(A) = \max_i |\lambda_i|$ is the *spectral radius*

Similar matrices

square matrices A and B are *similar* if there exists a nonsingular matrix T such that

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

- we call the transformation $A \rightarrow T^{-1}AT$ a *similarity transformation* of A
- 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

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

- this decomposition is the *spectral decomposition* of A
- not all 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 max no. of lin. indep. eigenvectors with eigenvalue λ
- sum is the maximum number of linearly independent eigenvectors of the matrix
- geometric multipicity never exceeds algebraic multiplicity (can be less)

Non-diagonalizable matrices

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

Example

consider the matrix

$$A = \begin{bmatrix} 4 & 1 \\ 0 & 4 \end{bmatrix}$$

- has the eigenvalue 4 with algebraic multiplicity 2
- eigenvector $(1, 0)$; no other linearly independent eigenvector
- the geometric multiplicity of the eigenvalue 4 is, then, only 1

consider

$$A = \begin{bmatrix} 4 & 0 \\ 0 & 4 \end{bmatrix}$$

- two linearly independent eigenvectors, are $(1, 0)$ and $(0, 1)$
- so the geometric multiplicity of the eigenvalue 4 equals 2

Example

take the first matrix from last example, and perturb its off-diagonal elements slightly

$$A = \begin{bmatrix} 4 & 1.01 \\ 0.01 & 4 \end{bmatrix}$$

- using the MATLAB function `eig` we find that the eigenvalues: 4.1005 and 3.8995
- perturbation of 0.01 produced a change of magnitude 0.1005 in eigenvalues!
- eigenvalue is ill-conditioned
- applying the same perturbation to the second, diagonal matrix from last example produces eigenvalues 4.01 and 3.99, so the eigenvalue 4 is well-conditioned here

Symmetric eigendecomposition

let A be a real symmetric matrix ($A = A^T \in \mathbb{R}^{n \times n}$), then

- all eigenvalues of A are real
- A has n linearly independent eigenvectors
- there is a set of n orthonormal eigenvectors of A

Symmetric eigendecomposition: let $A \in \mathbb{R}^{n \times n}$ be symmetric, then

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

- $Q \in \mathbb{R}^{n \times n}$ is *orthogonal* ($Q^T Q = I$)
- $\Lambda = \mathrm{diag}(\lambda_1, \ldots, \lambda_n)$
- columns of Q forms an orthonormal set of eigenvectors of Q
- this factorization is called *symmetric eigenvalue decomposition*

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
- corresponding eigenvector x is normalized so that $\|x\|_\infty = 1$
- suppose $|x_i| = 1$ (by definition of the ∞-norm)
- because $Ax = \lambda x$, we have

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

so that

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

Outline

- eigenvalues and diagonalization
- **singular value decomposition**
- SVD and matrix properties
- 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":

$$\begin{array}{ll} \Sigma = \mathrm{diag}(\sigma_1, \ldots, \sigma_n) & \text{if } m = n \\ \Sigma = \left[\begin{array}{c} \mathrm{diag}(\sigma_1, \ldots, \sigma_n) \\ 0_{(m-n)\times n} \end{array}\right] & \text{if } m > n \\ \Sigma = \left[\begin{array}{cc} \mathrm{diag}(\sigma_1, \ldots, \sigma_m) & 0_{m\times(n-m)} \end{array}\right] & \text{if } m < n \end{array}$$

- 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

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 $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_n \geq 0$

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 transpose of Gram matrix

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:

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

Outline

- eigenvalues and diagonalization
- singular value decomposition
- **SVD and matrix properties**
- least squares via SVD

Rank

- rank of a matrix is the maximum number of linearly independent columns
- number of positive singular values is the rank of a matrix

Compact-form of SVD: suppose there are r positive singular values:

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

partition the matrices in a full SVD of A as

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

- Σ_1 is $r \times r$ with the positive singular values $\sigma_1, \ldots, \sigma_r$ on the diagonal
- U_1 is $m \times r$ and V_1 is $n \times r$ have orthonormal columns

Pseudo-inverse

consider the SVD of A:

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

Pseudo-inverse

$$\begin{aligned} A^\dagger = V_1 \Sigma_1^{-1} U_1^T &= \begin{bmatrix} V_1 & V_2 \end{bmatrix} \begin{bmatrix} \Sigma_1^{-1} & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} U_1^T \\ U_2^T \end{bmatrix} \\ &= V \Sigma^\dagger U^T = \sum_{i=1}^{r} \frac{1}{\sigma_i} v_i u_i^T \end{aligned}$$

- 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, $A^\dagger = A^T(AA^T)^{-1}$

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

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,

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

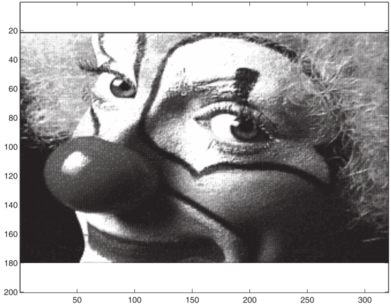
- B is also the best approximation for the 2-norm: for every C with rank r,

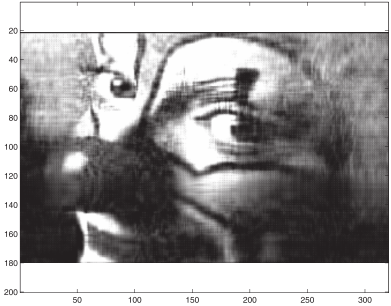
$$\|A - C\|_2 \geq \|A - B\|_2 = \sigma_{r+1}$$

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)');
```

Latent semantic analysis

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

a **term-document matrix** is an $n \times m$ matrix

- n is the number of terms (words)
- m is the number of documents
- entry i, j represents a function of the frequency of word i in document j
- simplest function is a simple count

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 $q = (0, 0, 0, 0, 1, 1)$
- two documents deal with a similar issue
- but overlap between the two columns in the matrix A is relatively small

Reduced space

- one popular measure to answer the question whether a query and a document have much in common is by using angles; for a given query q, we set

$$\cos(\theta_j) = \frac{(Ae_j)^T q}{\|Ae_j\| \|q\|}, \quad j = 1, \ldots, m$$

- if $A_r = U_r \Sigma_r V_r^T$ with r small, then

$$\cos(\theta_j) = \frac{e_j^T V_r \Sigma_r (U_r^T q)}{\|\Sigma_r V_r^T e_j\| \|q\|}$$

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

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

Outline

- eigenvalues and diagonalization
- singular value decomposition
- SVD and matrix properties
- **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 linearly independent column ($m \geq n$) 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 $\mathrm{rank}(A) = r < \min\{m, n\}$
- 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

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

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

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

- where $z_2 \in \mathbb{R}^{n-r}$ is arbitrary
- 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 expensive

References and further readings

- U. M. Ascher. *A First Course on Numerical Methods*. Society for Industrial and Applied Mathematics, 2011.
- M. T. Heath. *Scientific Computing: An Introductory Survey* (revised second edition). Society for Industrial and Applied Mathematics, 2018.
- L. Vandenberghe. *EE133B lecture notes,* Univ. of California, Los Angeles. (`https://www.seas.ucla.edu/~vandenbe/ece133b.html`)