Linear systems of equations

system of linear equations:

$$\begin{array}{ccccccccc} y_1 & = & A_{11} & x_1 & + & \cdots & + & A_{1n} & x_n \\ & \vdots & & & & & & & \\ y_m & = & A_{m1} & x_1 & + & \cdots & + & A_{mn} & x_n \end{array}$$

matrix representation:

$$y = Ax$$

where

$$y = \begin{bmatrix} y_1 \\ \vdots \\ y_m \end{bmatrix}, \qquad A = \begin{bmatrix} A_{11} & \cdots & A_{1n} \\ \vdots & \ddots & \vdots \\ A_{m1} & \cdots & A_{mn} \end{bmatrix}, \qquad x = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$$

right side of system *defines* matrix-vector multiplication

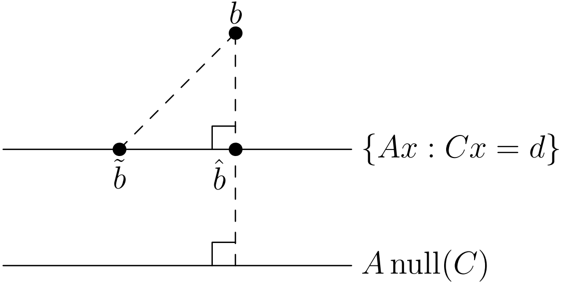
Linearly constrained least-squares problems

$$\begin{aligned} \underset{x \in \mathbb{R}^n}{\text{minimize}} &: \|Ax - b\| \\ \text{subject to} &: Cx = d \end{aligned}$$

normal equations:

$$\begin{bmatrix} A^\mathsf{T} A & C^\mathsf{T} \\ C & 0 \end{bmatrix} \begin{bmatrix} x \\ \lambda \end{bmatrix} = \begin{bmatrix} A^\mathsf{T} b \\ d \end{bmatrix}$$

The orthogonality principle



$$b - \hat{b} \perp A\,\mathrm{null}(C) \qquad \Rightarrow \qquad A^{\mathsf{T}}(b - \hat{b}) \in \mathrm{null}(C)^{\perp} = \mathrm{range}(C^{\mathsf{T}})$$

$$\begin{bmatrix} A^{\mathsf{T}}A & C^{\mathsf{T}} \\ C & 0 \end{bmatrix} \begin{bmatrix} x \\ \lambda \end{bmatrix} = \begin{bmatrix} A^{\mathsf{T}}b \\ d \end{bmatrix} \qquad \begin{array}{l} \text{orthogonality} \\ \text{feasibility} \end{array}$$

Linear dynamical systems

discrete-time linear dynamical system:

$$\begin{aligned} x(t+1) &= Ax(t) + Bu(t), \\ y(t) &= Cx(t) + Du(t) \end{aligned}$$

solutions of the state and measurement equations:

$$\begin{aligned} x(t) &= A^t x(0) + \sum_{t=0}^{t-1} A^{t-\tau-1} Bu(\tau), \\ y(t) &= CA^t x(0) + \sum_{t=0}^{t-1} CA^{t-\tau-1} Bu(\tau) + Du(t) \end{aligned}$$

Eigenvalues and eigenvectors

suppose $A \in \mathbb{R}^{n \times n}$

- $v \in \mathbb{C}^n$ is (right) eigenvector of A with eigenvalue $\lambda \in \mathbb{C}$ if

$$Av = \lambda v, \qquad v \neq 0$$

- in matrix form:

$$A \begin{bmatrix} v_1 & \cdots & v_k \end{bmatrix} = \begin{bmatrix} v_1 & \cdots & v_k \end{bmatrix} \begin{bmatrix} \lambda_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \lambda_k \end{bmatrix}$$

Eigenvalues and eigenvectors

- $w \in \mathbb{C}^n$ left eigenvector of A with eigenvalue $\lambda \in \mathbb{C}$ if

$$w^\mathsf{T} A = \lambda w^\mathsf{T}$$

- in matrix form:

$$\begin{bmatrix} w_1^\mathsf{T} \\ \vdots \\ w_k^\mathsf{T} \end{bmatrix} A = \begin{bmatrix} \lambda_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \lambda_n \end{bmatrix} \begin{bmatrix} w_1^\mathsf{T} \\ \vdots \\ w_n^\mathsf{T} \end{bmatrix}$$

Diagonalizable matrices

- A diagonalizable if there is linearly independent set of n eigenvectors:

$$A = \begin{bmatrix} v_1 & \cdots & v_n \end{bmatrix} \begin{bmatrix} \lambda_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \lambda_n \end{bmatrix} \begin{bmatrix} w_1^{\mathsf{T}} \\ \vdots \\ w_n^{\mathsf{T}} \end{bmatrix}$$

$$= \sum_{i=1}^{n} \lambda_i v_i w_i^{\mathsf{T}}$$

- where

$$\begin{bmatrix} w_1^{\mathsf{T}} \\ \vdots \\ w_n^{\mathsf{T}} \end{bmatrix} = \begin{bmatrix} v_1 & \cdots & v_n \end{bmatrix}^{-1}$$

Dominant-eigenvalue analysis

- order eigenvalues such that $|\lambda_1| \geq \cdots \geq |\lambda_n|$
- λ_1 is unique dominant eigenvalue if $|\lambda_1| > |\lambda_2|$
- then, for large t,
  $$A^t x \sim \lambda_1^t v_1 w_1^\mathsf{T} x$$
  assuming $w_1^\mathsf{T} x \neq 0$
- for example, if $x(t+1) = Ax(t)$, then
  $$\lim_{t \to \infty} \frac{\|x(t+1)\|}{\|x(t)\|} = |\lambda_1|$$
  assuming $w_1^\mathsf{T} x(0) \neq 0$
- if $|\lambda_1| > |\lambda_2| > |\lambda_3|$, then λ_2 is "vice" dominant eigenvalue

Quadratic forms

- quadratic form is function of the form

$$x^{\mathsf{T}}Ax = \sum_{i=1}^{n}\sum_{j=1}^{n} A_{ij}x_i x_j$$

- unique representation using symmetric matrix $A \in \mathbb{S}^n$

Symmetric matrices

suppose $A \in \mathbb{R}^{n \times n}$ is symmetric

- there is an orthonormal set of n eigenvectors:

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

- A called positive definite if $x^{\mathsf{T}} A x > 0$ for all $x \neq 0$
  - if and only if all $\lambda_i > 0$
- A called positive semidefinite if $x^{\mathsf{T}} A x \geq 0$ for all x
  - if and only if all $\lambda_i \geq 0$

Extremal-trace problems

$$\begin{aligned} \underset{Q \in \mathbb{R}^{n\times k}}{\text{maximize}} &: \operatorname{tr}(Q^{\mathsf{T}} A Q) \\ \text{subject to} &: Q^{\mathsf{T}} Q = I \end{aligned}$$

- equivalent formulation:

$$\begin{aligned} \underset{q_1, \dots, q_k \in \mathbb{R}^n}{\text{maximize}} &: \textstyle\sum_{i=1}^{k} q_i^{\mathsf{T}} A q_i \\ \text{subject to} &: q_i^{\mathsf{T}} q_j = \delta_{ij} \end{aligned}$$

- solution is matrix Q whose columns are eigenvectors of A corresponding to k largest eigenvalues
- to minimize, take matrix whose columns are eigenvectors corresponding to k smallest eigenvalues

The singular-value decomposition

- $A^\mathsf{T} A$ and AA^T are symmetric and positive semidefinite
- eigenvalue decompositions:

$$A^\mathsf{T} A = V\Sigma^2 V^\mathsf{T}, \qquad AA^\mathsf{T} = U\Sigma^2 U^\mathsf{T}$$

  - eigenvalues are nonnegative
  - eigenvalues are the same
- singular-value decomposition:

$$A = U\Sigma V^\mathsf{T}$$

The singular-value decomposition (matrix form)

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

- $U \in \mathbb{R}^{m\times r}$ has orthonormal columns
- $\Sigma \in \mathbb{R}^{r\times r}$ diagonal, nonsingular
- $V \in \mathbb{R}^{n\times r}$ has orthonormal columns

Ellipsoids

- generate members:

$$\{x_0 + Az : \|z\| \leq 1\}$$

- test membership:

$$\{x \in \mathbb{R}^n : (x - x_0)^{\mathsf{T}} S(x - x_0) \leq 1\}$$

- equivalent representations: $S = (AA^{\mathsf{T}})^{-1}$
- image of unit ball under A is ellipsoid with principal axes $\sigma_i u_i$

The singular-value decomposition (dyadic expansion)

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

- $\sigma_1 \geq \cdots \geq \sigma_r > 0$ are singular values
- $u_1, \ldots, u_r \in \mathbb{R}^m$ are output singular vectors
- $v_1, \ldots, v_r \in \mathbb{R}^n$ are input singular vectors
- $u_i v_i^\mathsf{T}$ is unit atom
  - $\mathrm{rank}(u_i v_i^\mathsf{T}) = 1$
  - $\|u_i v_i^\mathsf{T}\| = 1$
- SVD decomposes A into sum of unit atoms
  - singular values rank atoms in terms of importance

The pseudoinverse

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

- $A^\dagger b$ is least-norm least-squares vector

$$\begin{aligned} \underset{x \in \mathbb{R}^n}{\text{minimize}} &: \|x\| \\ \text{subject to} &: \|Ax - b\| = \min_z \|Az - b\| \end{aligned}$$

Extremal-trace problems (input)

$$\begin{aligned}\underset{Q \in \mathbb{R}^{n\times k}}{\text{maximize}} &: \operatorname{tr}(Q^\mathsf{T} A^\mathsf{T} A Q)\\ \text{subject to} &: Q^\mathsf{T} Q = I\end{aligned}$$

- equivalent formulation:

$$\begin{aligned}\underset{q_1,\ldots,q_k \in \mathbb{R}^{n}}{\text{maximize}} &: \textstyle\sum_{i=1}^{k} \|Aq_i\|^2\\ \text{subject to} &: q_i^\mathsf{T} q_j = \delta_{ij}\end{aligned}$$

- solution is matrix Q whose columns are eigenvectors of $A^\mathsf{T} A = V\Sigma^2 V^\mathsf{T}$ corresponding to k largest eigenvalues
- equivalently, solution is matrix Q whose columns are the first k input singular vectors of A

Extremal-trace problems (output)

$$\begin{array}{rl} \underset{Q \in \mathbb{R}^{m \times k}}{\text{maximize}} & : \operatorname{tr}(Q^\mathsf{T} A A^\mathsf{T} Q) \\ \text{subject to} & : Q^\mathsf{T} Q = I \end{array}$$

- equivalent formulation:

$$\begin{array}{rl} \underset{q_1, \ldots, q_k \in \mathbb{R}^m}{\text{maximize}} & : \sum_{i=1}^k \sum_{j=1}^n \| q_i q_i^\mathsf{T} A_{*j} \|^2 \\ \text{subject to} & : q_i^\mathsf{T} q_j = \delta_{ij} \end{array}$$

- solution is matrix Q whose columns are eigenvectors of $AA^\mathsf{T} = U\Sigma^2 U^\mathsf{T}$ corresponding to k largest eigenvalues
- equivalently, solution is matrix Q whose columns are the first k output singular vectors of A

Low-rank approximation

singular-value decomposition of $A \in \mathbb{R}^{m \times n}$:

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

best rank-p approximation of A:

$$\hat{A}_p = \sum_{i=1}^{p} \sigma_i u_i v_i^\mathsf{T}$$