ESE 524: Detection and Estimation Theory

Linear algebra review

Washington University in St. Louis

Welcome to The Matrix

- This review is intended to cover the important topics from matrix algebra that are needed for linear models and other lectures later in the semester.
- An $M \times N$ matrix $\boldsymbol{A}$ is indexed by M rows and N columns. The i^{th} row and j^{th} column are indexed as A_{ij}:

$$\boldsymbol{A} = \begin{bmatrix} A_{11} & A_{12} & \cdots & A_{1N} \\ A_{21} & A_{22} & \cdots & A_{2N} \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ A_{N1} & A_{N2} & \cdots & A_{NN} \end{bmatrix}$$

- The transpose of $\boldsymbol{A}$ is achieved by switching the rows and columns and denoted $\boldsymbol{A}^{\mathrm{T}}$.

Common Matrix Operations

- Elementwise addition and scalar multiplication: for two matrices with the same dimension, $c\boldsymbol{A} + d\boldsymbol{B} = [cA_{ij} + cB_{ij}]$, $\forall i \in \{1, 2, ..., M\}, j \in \{1, 2, ..., N\}, c, d \in \mathbb{R}$.
- Matrix multiplication: For $\boldsymbol{A} \in \mathbb{R}^{M_1 \times N_1}$ and $\boldsymbol{B} \in \boldsymbol{R}^{N_1 \times N_2}$, the product is a matrix with M_1 rows and N_2 columns: $\boldsymbol{AB} \in \boldsymbol{R}^{M \times N_2}$. Note that this is not symmetric - $\boldsymbol{AB} \neq \boldsymbol{BA}$ in general.
- Multiplication is associative $(\boldsymbol{AB})\boldsymbol{C} = \boldsymbol{A}(\boldsymbol{BC})$
- Multiplication is distributive $\boldsymbol{A}(\boldsymbol{B} + \boldsymbol{C}) = \boldsymbol{AB} + \boldsymbol{AC}$.
- $(\boldsymbol{AB})^{\mathrm{T}} = \boldsymbol{B}^{\mathrm{T}}\boldsymbol{A}^{\mathrm{T}}$

Different ways to express Matrix Products

- Given two matrices $A \in \mathbb{R}^{M\times N}$, and $\boldsymbol{B} \in \mathbb{R}^{N\times P}$, express $\boldsymbol{A}$ as a matrix of row vectors and $\boldsymbol{B}$ as a matrix of column Vectors:

$$\boldsymbol{C} = \boldsymbol{A}\boldsymbol{B} = \begin{bmatrix} \boldsymbol{a}_1^{\mathrm{T}} \\ \boldsymbol{a}_2^{\mathrm{T}} \\ \cdot \\ \cdot \\ \cdot \\ \boldsymbol{a}_M^{\mathrm{T}} \end{bmatrix} \begin{bmatrix} \boldsymbol{b}_1 & \boldsymbol{b}_2 & \cdots & \boldsymbol{b}_P \end{bmatrix} = \begin{bmatrix} \boldsymbol{a}_1^{\mathrm{T}}\boldsymbol{b}_1 & \boldsymbol{a}_1^{\mathrm{T}}\boldsymbol{b}_2 & \cdots & \boldsymbol{a}_1^{\mathrm{T}}\boldsymbol{b}_P \\ \boldsymbol{a}_2^{\mathrm{T}}\boldsymbol{b}_1 & \boldsymbol{a}_2^{\mathrm{T}}\boldsymbol{b}_2 & \cdots & \boldsymbol{a}_2^{\mathrm{T}}\boldsymbol{b}_P \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \boldsymbol{a}_M^{\mathrm{T}}\boldsymbol{b}_1 & \boldsymbol{a}_M^{\mathrm{T}}\boldsymbol{b}_2 & \cdots & \boldsymbol{a}_M^{\mathrm{T}}\boldsymbol{b}_P \end{bmatrix}$$

- Conversely, expressing $\boldsymbol{A}$ as column vectors and $\boldsymbol{B}$ as row vectors:

$$\boldsymbol{C} = \boldsymbol{A}\boldsymbol{B} = \begin{bmatrix} \boldsymbol{a}_1 & \boldsymbol{a}_2 & \cdot & \cdot & \cdot & \boldsymbol{a}_N \end{bmatrix} \begin{bmatrix} \boldsymbol{b}_1^{\mathrm{T}} \\ \boldsymbol{b}_2^{\mathrm{T}} \\ \cdot \\ \cdot \\ \cdot \\ \boldsymbol{b}_N^{\mathrm{T}} \end{bmatrix} = \sum_{i=1}^{N} \boldsymbol{a}_i \boldsymbol{b}_i^{\mathrm{T}}$$

which is a sum of outer products.

Identity Matrix

- The identity matrix $\boldsymbol{I}$ is defined as:

$$I_{ij} = \begin{cases} 1 & i = j \\ 0 & \text{otherwise} \end{cases}$$

- Any multiplication with the identity results in the original matrix: $\boldsymbol{IA} = \boldsymbol{B}$ or $\boldsymbol{BI} = \boldsymbol{B}$ results in the original matrix.
- This is a special case of a diagonal matrix - i.e., all the off-diagonal entries are zero.

Determinants

- Given a square matrix $\boldsymbol{A} \in \mathbb{R}^{N \times N}$, the determinant is a function:

$$\det(\boldsymbol{A}) : \mathbb{R}^{N \times N} \to \mathbb{R}$$

  which is also denoted $|\boldsymbol{A}|$.
- The determinant represents the area of an N-dimensional parrallelogram spanned by $\sum_{i=1}^{N} c_i \boldsymbol{a}_i$ for $c_i \in [0, 1]$.
- Some properties:
  - $|\boldsymbol{A}| = |\boldsymbol{A}^{\mathrm{T}}|$
  - $|\boldsymbol{A}\boldsymbol{B}| = |\boldsymbol{A}||\boldsymbol{B}|$
  - For $N = 2$, $|\boldsymbol{A}| = A_{11}A_{22} - A_{12}A_{21}$
  - For $N = 3$, $|\boldsymbol{A}| = A_{11}A_{22}A_{33} + A_{12}A_{23}A_{31} + A_{13}A21A_{32} - A_{11}A_{23}A_{32} - A_{12}A_{21}A_{33} - A_{13}A_{22}A_{31}$
  - For larger N there is a recursive formula relying on something called "expansion by minors".
  - If $|\boldsymbol{A}| = 0$ we call the matrix singular.

The Trace Operator

- For a square matrix $\boldsymbol{A} \in \mathbb{R}^{N \times N}$ the trace operator is the sum of all the diagonal elements:

$$\mathrm{tr}(\boldsymbol{A}) = \sum_{i=1}^{N} A_{ii}$$

- We have already seen the trace in lecture 2.
- Some properties:
  - $\mathrm{tr}(\boldsymbol{A} + \boldsymbol{B}) = \mathrm{tr}(\boldsymbol{A}) + \mathrm{tr}(\boldsymbol{B})$
  - $\mathrm{tr}(\boldsymbol{AB}) = \mathrm{tr}(\boldsymbol{BA})$, and $\mathrm{tr}(\boldsymbol{ABC}) = \mathrm{tr}(\boldsymbol{BCA}) = \mathrm{tr}(\boldsymbol{CBA})$, and so one for the product of more matrices.

Eigenvalues

- Given a vector $\boldsymbol{x} \in \mathbb{R}^N$ a constant λ and a square matrix $\boldsymbol{A} \in \mathbb{R}^{N \times N}$, $\boldsymbol{x}$ is an eigenvector if

$$\boldsymbol{A}\boldsymbol{x} = \lambda \boldsymbol{x}$$

- Eigenvalues can be found by computing the roots of characteristic polynomial, $|\lambda \boldsymbol{I} - \boldsymbol{A}| = 0$
- Interesting Properties:
  - $\mathrm{tr}(\boldsymbol{A}) = \sum_i \lambda_i$
  - $|\boldsymbol{A}| = \prod_i \lambda_i$ - singular matrices have at least one eigenvalue equal to zero.

Norm

- Given a vector $\boldsymbol{x} \in \mathbb{R}^N$, a norm is a function $||\cdot||$ that measure the "size" of $\boldsymbol{x}$ and satisfies:
  - $||\boldsymbol{x}|| \geq 0$
  - $||\boldsymbol{x}|| = 0 \iff \boldsymbol{x} = 0$
  - For $c \in \mathbb{R}$, $||c\boldsymbol{x}|| = |c| \cdot ||\boldsymbol{x}$.
  - For all x, y, $||\boldsymbol{x} + \boldsymbol{y}|| \leq ||\boldsymbol{x}|| + ||\boldsymbol{y}||$
- The most common norm we will use in this class is the 2-norm:

$$||\boldsymbol{x}||_2 = \sqrt{\sum_{i=1}^{N} x_i^2}$$

Inverses

- For a square matrix $\boldsymbol{A} \in \mathbb{R}^{N \times N}$, an inverse is a matrix $\boldsymbol{A}^{-1}$ such that $\boldsymbol{A}\boldsymbol{A}^{-1} = \boldsymbol{A}^{-1}\boldsymbol{A} = \boldsymbol{I}$.
  Note: Do not try to divide matrices by one another in your homework.
- Not all matrices have inverses. For example, non-square matrices.
- The determinant is $|A^{-1}| = \frac{1}{|\boldsymbol{A}|}$ - thus only non-singular matrices have inverses.
- If the $\{\lambda_1, \lambda_2, ..., \lambda_N\}$ are the eigenvalues of $\boldsymbol{A}$, then $\{1/\lambda_1, 1/\lambda_2, ...\}$ are the eigenvalues of $\boldsymbol{A}^{-1}$.
- For a 2×2 matrix, the inverse is:

$$\boldsymbol{A}^{-1} = \frac{1}{|\boldsymbol{A}|} \begin{bmatrix} A_{22} & -A_{12} \\ -A_{21} & A_{11} \end{bmatrix}$$

Orthogonal Matrices

- A square matrix $\boldsymbol{A}$ is orthonormal if $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A} = \boldsymbol{A}\boldsymbol{A}^{\mathrm{T}} = \boldsymbol{I}$.
- In this case each column is at a 90 degree angle from every other column.
- Orthonormal matrices preserve norms, i.e., $||\boldsymbol{Ax}||_2 = ||x||_2$.
- **Proof:**
  $||\boldsymbol{Ax}||_2 = \sqrt{(\boldsymbol{Ax})^{\mathbf{T}}(\boldsymbol{Ax})} = \sqrt{\boldsymbol{x}^{\mathbf{T}}(\boldsymbol{A}^{\mathbf{T}}\boldsymbol{A})\boldsymbol{x}} = \sqrt{\boldsymbol{x}^{\mathbf{T}}\boldsymbol{I}\boldsymbol{x}} = \sqrt{\boldsymbol{x}^{\mathbf{T}}\boldsymbol{x}} = ||x||_2$

Symmetric Matrices

- A matrix $\boldsymbol{A}$ is symmetric if $\boldsymbol{A} = \boldsymbol{A}^{\mathrm{T}}$.
- Symmetric Matrices have real eigenvalues, and the matrix containing all the eigenvectors $\boldsymbol{X}$ is orthonormal.
- This means that the matrix can be decomposed as: $\boldsymbol{A} = \boldsymbol{X}\boldsymbol{\Lambda}\boldsymbol{X}^{\mathrm{T}}$, where $\boldsymbol{\Lambda}$ is a diagonal matrix whose non-zero elements are the eigenvalues of $\boldsymbol{A}$.
- This means that

  $$\max_{\boldsymbol{x}} \boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x} = \boldsymbol{x}_{\max}$$

  is the eigenvector corresponding to the largest eigenvalue.
- Why is this important? Covariance matrices are symmetric and positive semi-definite (p.s.d.).

The Covariance matrix is p.s.d.

Proof. Let $x \in \mathbb{R}^n$ be a random vector. Then

$$\boldsymbol{C} \triangleq \mathbb{E}(\boldsymbol{x} - \overline{\boldsymbol{x}})(\boldsymbol{x} - \overline{\boldsymbol{x}})^{\mathrm{T}}$$

Let $\boldsymbol{u} \in \mathbb{R}^n$ be any deterministic vector.

$$\boldsymbol{u}^{\mathrm{T}}\boldsymbol{C}\boldsymbol{u} = \boldsymbol{u}^{\mathrm{T}}\mathbb{E}(\boldsymbol{x} - \overline{\boldsymbol{x}})(\boldsymbol{x} - \overline{\boldsymbol{x}})^{\mathrm{T}}\boldsymbol{u} = \mathbb{E}\boldsymbol{u}^{\mathrm{T}}(\boldsymbol{x} - \overline{\boldsymbol{x}})(\boldsymbol{x} - \overline{\boldsymbol{x}})^{\mathrm{T}}\boldsymbol{u}$$

Define $s = \boldsymbol{u}^{\mathrm{T}}\boldsymbol{x}$. s is a random variable with mean

$$\mathrm{E}(s) = \mathrm{E}\boldsymbol{u}^{\mathrm{T}}\boldsymbol{x} = \boldsymbol{u}^{\mathrm{T}}\mathrm{E}\boldsymbol{x} = \boldsymbol{u}^{\mathrm{T}}\overline{\boldsymbol{x}}$$

.

$$\Rightarrow \mathrm{E}\boldsymbol{u}^{\mathrm{T}}(\boldsymbol{x} - \overline{\boldsymbol{x}})(\boldsymbol{x} - \overline{\boldsymbol{x}})^{\mathrm{T}}\boldsymbol{u} = \mathrm{E}(s^{\mathrm{T}}s) = \mathrm{var}(s) \geq 0$$

□

Matrix Calculus: Real-Valued Functions

- Given a vector $\boldsymbol{x} = [x_1, x_2, ..., x_N]^{\mathrm{T}} \in \mathbb{R}^N$ and a function $f : \mathbb{R}^N \to \mathbb{R}$, the gradient is:

$$\frac{\partial f}{\partial \boldsymbol{x}^{\mathrm{T}}} = \nabla f(\boldsymbol{x}) = [\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}, \cdots, \frac{\partial f}{\partial x_N}]$$

- The "second derivative" is given by a the Hessian matrix:

$$\nabla^2 f(\boldsymbol{x}) = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_N} \\ \frac{\partial^2 f}{\partial x_2 \partial x_1} & \frac{\partial^2 f}{\partial x_2^2} & \cdots & \frac{\partial^2 f}{\partial x_2 \partial x_N} \\ . & . & . & . \\ . & . & . & . \\ \frac{\partial^2 f}{\partial x_N \partial x_1} & \frac{\partial^2 f}{\partial x_N \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_N^2} \end{bmatrix}$$

Matrix Calculus: Vector Valued Functions

- If $\boldsymbol{f} = [f_1(\boldsymbol{x}), f_2(\boldsymbol{x}), \cdots, f_M(\boldsymbol{x})]$ is a vector valued function, the Jacobian matrix is:

$$\frac{\partial \boldsymbol{f}}{\partial \boldsymbol{x}^{\mathrm{T}}} = \begin{bmatrix} \nabla f_1(\boldsymbol{x}) \\ \nabla f_2(\boldsymbol{x}) \\ \vdots \\ \nabla f_M(\boldsymbol{x}) \end{bmatrix} = \begin{bmatrix} \frac{\partial^2 f_1}{\partial x_1} & \frac{\partial f_1}{\partial x_2} & \cdots & \frac{\partial f_1}{\partial x_N} \\ \frac{\partial f_2}{\partial x_1} & \frac{\partial f_2}{\partial x_2} & \cdots & \frac{\partial f_2}{\partial x_M} \\ . & . & . & . \\ . & . & . & . \\ \frac{\partial f_M}{\partial x_1} & \frac{\partial f_M}{\partial x_2} & \cdots & \frac{\partial f_M}{\partial x_N} \end{bmatrix}$$

- Then the Hessian becomes a third order tensor, "stacking" the Hessian matrix of each f_i:

$$\nabla^2 \boldsymbol{f}(\boldsymbol{x}) = [\nabla^2 f_1(\boldsymbol{x}), \nabla^2 f_2(\boldsymbol{x}), \cdots, \nabla^2 f_M(\boldsymbol{x})]$$

Quadratic Form

- Given $\boldsymbol{x} \in \mathbb{R}^N$ and a symmetric matrix $\boldsymbol{A} \in \mathbb{R}^{N\times N}$, then a Quadratic Form is:

$$f(\boldsymbol{x}) = \boldsymbol{x}^{\mathrm{T}} \boldsymbol{A} \boldsymbol{x}$$

- $\boldsymbol{A}$ is positive definite if $f(\boldsymbol{x}) > 0 \, \forall \boldsymbol{x}$.
- $\boldsymbol{A}$ is positive semi-definite if $f \geq 0$.
- The eigenvalues of a positive (semi) definite matrix are all positive (non-negative).
- This is a multidimensional version of a quadratic equation, and the derivative follows a similar formula:

$$\frac{\partial f}{\partial \boldsymbol{x}} = 2\boldsymbol{A}\boldsymbol{x}$$

Matrix Decompositions

- It is often useful to express a matrix $\boldsymbol{A} \in \mathbb{R}^{M \times N}$ as a product of other matrices with useful properties.
- Eigen-Decomposition: If the eigenvectors of $\boldsymbol{A}$ are linearly independent:

$$\sum_{i=1} = 1^N a_i \boldsymbol{x}^i = 0 \iff a_i = 0 \, \forall i \in [1, ..., N]$$

then $\boldsymbol{A}$ is diagonalizable and can be expressed as

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

- QR Decomposition: $\boldsymbol{A} \in \mathbb{R}^{M \times N}$ can be decomposed into $\boldsymbol{QR}$ where $\boldsymbol{Q}$ is an orthonormal matrix and $\boldsymbol{R}$ is upper triangular.
- The QR Decomposition can be used to solve linear regression problems, find eigenvalues, and save computation time.