Background: Linear Algebra

EE4C03 Statistical Digital Signal Processing and Modeling

Delft University of Technology, The Netherlands

Outline

- Vectors and matrices
- Linear independence, vector spaces, and basis vectors
- Linear equations
- Eigenvalue decomposition
- Optimization theory

Vectors

- An N-dimensional vector is assumed to be a column vector:

$$\boldsymbol{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_N \end{bmatrix}$$

- Complex conjugate (Hermitian) transpose

$$\boldsymbol{x}^{\mathrm{H}} = (\boldsymbol{x}^{\mathrm{T}})^* = (\boldsymbol{x}^*)^{\mathrm{T}} = \begin{bmatrix} x_1^* & x_2^* & \ldots, x_N^* \end{bmatrix}$$

- For a discrete-time signal $x(n)$, we use the following vectors

$$\boldsymbol{x} = \begin{bmatrix} x(0) \\ x(1) \\ \vdots \\ x(N-1) \end{bmatrix} \qquad \boldsymbol{x}(n) = \begin{bmatrix} x(n) \\ x(n-1) \\ \vdots \\ x(n-N+1) \end{bmatrix}$$

Vectors

- Vector norms:

$$\text{Euclidean (2-norm):} \|\boldsymbol{x}\| = \|\boldsymbol{x}\|_2 = \left(\sum_{i=1}^{N} |x_i|^2\right)^{1/2} = \left(\sum_{i=1}^{N} x_i^* x_i\right)^{1/2} = (\boldsymbol{x}^{\mathrm{H}}\boldsymbol{x})^{1/2}$$

$$\text{1-norm:} \|\boldsymbol{x}\|_1 = \sum_{i=1}^{N} |x_i|$$

$$\infty\text{-norm:} \|\boldsymbol{x}\|_\infty = \max_i |x_i|$$

- The inner product is defined as

$$\langle \boldsymbol{a}, \boldsymbol{b} \rangle = \boldsymbol{a}^{\mathrm{H}}\boldsymbol{b} = \sum_{i=1}^{N} a_i^* b_i$$

  - Two vectors are *orthogonal* if $\langle \boldsymbol{b}, \boldsymbol{b} \rangle = 0$; if they are unit norm they are *orthonormal*
  - Properties of inner product:

$$|\langle \boldsymbol{a}, \boldsymbol{b} \rangle| \leq \|\boldsymbol{a}\| \|\boldsymbol{b}\| \text{(Cauchy-Schwarz)}$$

$$2|\langle \boldsymbol{a}, \boldsymbol{b} \rangle| \leq \|\boldsymbol{a}\|^2 + \|\boldsymbol{b}\|^2$$

Linear independence, vector spaces, and basis vectors

- A collection of N vectors $\boldsymbol{v}_1, \boldsymbol{v}_2, \dots, \boldsymbol{v}_N$ is called *linearly independent* if

$$\alpha_1 \boldsymbol{v}_1 + \alpha_2 \boldsymbol{v}_2 + \cdots + \alpha_N \boldsymbol{v}_N = 0 \quad \Leftrightarrow \quad \alpha_1 = \alpha_2 = \cdots = \alpha_N = 0$$

- The space $\mathcal{V}$ spanned by a collection of vectors $\boldsymbol{v}_1, \boldsymbol{v}_2, \dots, \boldsymbol{v}_N$ is called a *vector space*

$$\mathcal{V} = \left\{ \alpha_1 \boldsymbol{v}_1 + \alpha_2 \boldsymbol{v}_2 + \cdots + \alpha_N \boldsymbol{x}_N \;\middle|\; \alpha_i \in C \,, \;\forall i \right\}$$

  - If the vectors are linearly independent they are called a *basis* for that vector space
  - The number of basis vectors is called the *dimension* of the vector space
  - If the vectors are orthogonal → *orthogonal basis*
  - If the vectors are orthonormal → *orthonormal basis*

Matrices

- An $n \times m$ matrix has n rows and m columns:

$$\boldsymbol{A} = [a_{ij}] = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1m} \\ a_{21} & a_{22} & \cdots & a_{2m} \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nm} \end{bmatrix}$$

- Complex conjugate (Hermitian) transpose

$$\boldsymbol{A}^{\mathrm{H}} = (\boldsymbol{A}^{\mathrm{T}})^* = (\boldsymbol{A}^*)^{\mathrm{T}}$$

- Hermitian matrix

$$\boldsymbol{A} = \boldsymbol{A}^{\mathrm{H}}$$

Matrix inverse

- The rank of $\boldsymbol{A}$, denoted $\rho(\boldsymbol{A})$, is the number of independent columns or rows of $\boldsymbol{A}$

$$\text{Prototype rank-1 matrix:} \qquad \boldsymbol{A} = \boldsymbol{a}\boldsymbol{b}^{\mathrm{H}}$$

$$\text{Prototype rank-2 matrix:} \qquad \boldsymbol{A} = \boldsymbol{a}\boldsymbol{b}^{\mathrm{H}} + \boldsymbol{c}\boldsymbol{d}^{\mathrm{H}}$$

- The ranks of $\boldsymbol{A}$, $\boldsymbol{A}^{\mathrm{H}}$, $\boldsymbol{A}\boldsymbol{A}^{\mathrm{H}}$, and $\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A}$ are the same

$$\rho(\boldsymbol{A}) = \rho(\boldsymbol{A}^{\mathrm{H}}) = \rho(\boldsymbol{A}\boldsymbol{A}^{\mathrm{H}}) = \rho(\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A})$$

- If $\boldsymbol{A}$ is square and full rank ($\rho(\boldsymbol{A}) = n$), there is a unique inverse $\boldsymbol{A}^{-1}$ such that

$$\boldsymbol{A}\boldsymbol{A}^{-1} = \boldsymbol{A}^{-1}\boldsymbol{A} = \boldsymbol{I} = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix}$$

- If $\boldsymbol{A}$ is square and $\rho(\boldsymbol{A}) = n$, $\boldsymbol{A}$ is *invertible* or *nonsingular*
- If $\boldsymbol{A}$ is square and $\rho(\boldsymbol{A}) < n$, $\boldsymbol{A}$ is *noninvertible* or *singular*

Matrix inverse

- Properties

$$(\boldsymbol{AB})^{-1} = \boldsymbol{B}^{-1}\boldsymbol{A}^{-1}$$
$$(\boldsymbol{A}^{\mathrm{H}})^{-1} = (\boldsymbol{A}^{-1})^{\mathrm{H}}$$

- Matrix Inversion Lemma:

$$(\boldsymbol{A} + \boldsymbol{BCD})^{-1} = \boldsymbol{A}^{-1} - \boldsymbol{A}^{-1}\boldsymbol{B}(\boldsymbol{C}^{-1} + \boldsymbol{DA}^{-1}\boldsymbol{B})^{-1}\boldsymbol{DA}^{-1}$$

- Woodbury's Identity (special case of Matrix Inversion Lemma):

$$(\boldsymbol{A} + \boldsymbol{uv}^{\mathrm{H}})^{-1} = \boldsymbol{A}^{-1} - \frac{\boldsymbol{A}^{-1}\boldsymbol{uv}^{\mathrm{H}}\boldsymbol{A}^{-1}}{1 + \boldsymbol{v}^{\mathrm{H}}\boldsymbol{A}^{-1}\boldsymbol{u}}$$

Determinant and trace

- The determinant of an $n \times n$ matrix $\boldsymbol{A}$ is defined recursively by

$$\det(\boldsymbol{A}) = \sum_{i=1}^{n} (-1)^{i+j} a_{ij} \det(\boldsymbol{A}_{ij})$$

  where $\boldsymbol{A}_{ij}$ is the matrix obtained by removing the ith row and jth column from $\boldsymbol{A}$
- An $n \times n$ matrix $\boldsymbol{A}$ is invertible or nonsingular $\Leftrightarrow$ $\det(\boldsymbol{A}) \neq 0$
- Properties:

$$\begin{aligned} \det(\boldsymbol{AB}) &= \det(\boldsymbol{A}) \det(\boldsymbol{B}) \\ \det(\alpha \boldsymbol{A}) &= \alpha^n \det(\boldsymbol{A}) \\ \det(\boldsymbol{A}^{-1}) &= \frac{1}{\det(\boldsymbol{A})} \end{aligned}$$

- The trace of an $n \times n$ matrix $\boldsymbol{A}$ is given by

$$\operatorname{tr}(\boldsymbol{A}) = \sum_{i=1}^{n} a_{ii}$$

Linear equations

- A set of n linear equations in m unknowns (stacked in the vector $\boldsymbol{x}$), can be written as

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

  where $\boldsymbol{A}$ is an $m \times n$ matrix and $\boldsymbol{B}$ is an $m \times 1$ vector
- If $m = n$ and $\rho(\boldsymbol{A}) = n$ (square invertible), then $\boldsymbol{x} = \boldsymbol{A}^{-1}\boldsymbol{b}$
- If $m = n$ and $\rho(\boldsymbol{A}) < n$ (singular), then there is either no solution or many solutions;
  If $\boldsymbol{x}_0$ is a solution, then

$$\boldsymbol{x} = \boldsymbol{x}_0 + \alpha_1 \boldsymbol{z}_1 + \cdots + \alpha_k \boldsymbol{z}_k$$

  is also a solution with $\{\boldsymbol{z}_i, i = 1, 2, \ldots, k\}$ is a set of $k = n - \rho(\boldsymbol{A})$ linearly independent solutions of $\boldsymbol{A}\boldsymbol{z} = \boldsymbol{0}$

Linear equations

- If $n < m$, there are many solutions (the problem is *underdetermined*)
  We often take the solution with the minimal norm

$$\min_{\boldsymbol{x}} \|\boldsymbol{x}\| \quad \text{such that} \quad \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$$

  The solution is given by $\boldsymbol{x}_0 = \boldsymbol{A}^{\mathrm{H}}(\boldsymbol{A}\boldsymbol{A}^{\mathrm{H}})^{-1}\boldsymbol{b}$, where $\boldsymbol{A}^{+} = \boldsymbol{A}^{\mathrm{H}}(\boldsymbol{A}\boldsymbol{A}^{\mathrm{H}})^{-1}$ is the pseudo-inverse of $\boldsymbol{A}$ for the underdetermined problem ($\rho(\boldsymbol{A}) = n$)
- If $n > m$, there is generally no solution (the problem is *overdetermined*). We often take the least squares solution

$$\min_{\boldsymbol{x}} \|\boldsymbol{b} - \boldsymbol{A}\boldsymbol{x}\|$$

  The solution is given by $\boldsymbol{x}_0 = (\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A})^{-1}\boldsymbol{A}^{\mathrm{H}}\boldsymbol{b}$, where $\boldsymbol{A}^{+} = (\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A})^{-1}\boldsymbol{A}^{\mathrm{H}}$ is the pseudo-inverse of $\boldsymbol{A}$ for the overdetermined problem ($\rho(\boldsymbol{A}) = m$)

Special matrix forms

- Diagonal and block diagonal matrix:

$$\boldsymbol{A} = \begin{bmatrix} a_{11} & 0 & \cdots & 0 \\ 0 & a_{22} & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & a_{nn} \end{bmatrix}, \qquad \boldsymbol{A} = \begin{bmatrix} \boldsymbol{A}_{11} & 0 & \cdots & 0 \\ 0 & \boldsymbol{A}_{22} & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & \boldsymbol{A}_{kk} \end{bmatrix}$$

- Toeplitz and Hankel matrix (constant along (anti-)diagonal):

$$\boldsymbol{A} = \begin{bmatrix} 1 & 3 & 5 & 7 \\ 2 & 1 & 3 & 5 \\ 4 & 2 & 1 & 3 \\ 6 & 4 & 2 & 1 \end{bmatrix}, \qquad \boldsymbol{A} = \begin{bmatrix} 1 & 3 & 5 & 7 \\ 3 & 5 & 7 & 4 \\ 5 & 7 & 4 & 2 \\ 7 & 4 & 2 & 1 \end{bmatrix}$$

- A square matrix $\boldsymbol{A}$ is called *unitary* if $\boldsymbol{A}\boldsymbol{A}^{\mathrm{H}} = \boldsymbol{I}$ and $\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A} = \boldsymbol{I}$, or in other words, the columns and rows of $\boldsymbol{A}$ are orthonormal

Hermitian forms

- The quadratic form of an $n \times n$ Hermitian matrix $\boldsymbol{A}$ is

$$Q_A(\boldsymbol{x}) = \boldsymbol{x}^{\mathrm{H}} \boldsymbol{A} \boldsymbol{x} = \sum_{i=1}^{n} \sum_{j=1}^{n} x_i^* a_{ij} x_j$$

  where $\boldsymbol{x} = [x_1, x_2, \ldots, x_n]^{\mathrm{T}}$

- The matrix $\boldsymbol{A}$ is

$$\begin{array}{lll} \textit{positive definite}, & \boldsymbol{A} > 0, & \text{if } Q_A(\boldsymbol{x}) > 0,\ \forall \boldsymbol{x} \neq \mathbf{0}, \\ \textit{positive semidefinite}, & \boldsymbol{A} \geq 0, & \text{if } Q_A(\boldsymbol{x}) \geq 0,\ \forall \boldsymbol{x} \neq \mathbf{0} \\ & & \\ \textit{negative definite}, & \boldsymbol{A} < 0, & \text{if } Q_A(\boldsymbol{x}) < 0,\ \forall \boldsymbol{x} \neq \mathbf{0}, \\ \textit{negative semidefinite}, & \boldsymbol{A} \leq 0, & \text{if } Q_A(\boldsymbol{x}) \leq 0,\ \forall \boldsymbol{x} \neq \mathbf{0} \end{array}$$

- For any $n \times n$ matrix $\boldsymbol{A}$ and any $n \times m$ ($m \leq n$) matrix $\boldsymbol{B}$ with full rank m, the definiteness of $\boldsymbol{A}$ and $\boldsymbol{B}^{\mathrm{H}} \boldsymbol{A} \boldsymbol{B}$ are the same

Eigenvalues and eigenvectors

- For an $n \times n$ matrix $\boldsymbol{A}$ there are n *eigenvalues* λ_i and n *eigenvectors* $\boldsymbol{v}_i$ satisfying

$$\boldsymbol{A}\boldsymbol{v}_i = \lambda_i \boldsymbol{v}_i$$

- The eigenvalues are the roots of the *characteristic polynomial*

$$p(\lambda) = \det(\boldsymbol{A} - \lambda \boldsymbol{I})$$

- The eigenvectors have a scaling ambiguity and are often normalized, $\|\boldsymbol{v}_i\| = 1$
- The eigenvectors corresponding to distinct eigenvalues are linearly independent
- If $\boldsymbol{A}$ has rank $\rho(\boldsymbol{A})$, then $\boldsymbol{A}$ has $\rho(\boldsymbol{A})$ nonzero eigenvalues and $n - \rho(\boldsymbol{A})$ zero eigenvalues
- For a Hermitian matrix,
  - the eigenvalues are real
  - the eigenvectors are orthonormal
  - matrix positive (negative) definite $\Leftrightarrow$ all eigenvalues positive (negative)

Eigenvalue decomposition

- For an $n \times n$ matrix $\boldsymbol{A}$ with a set of n linearly independent eigenvectors we can perform an *eigenvalue decomposition* of $\boldsymbol{A}$

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

where $\boldsymbol{v}$ contains the eigenvectors and $\boldsymbol{\Lambda}$ is a diagonal matrix holding the eigenvalues

- Since for a Hermitian matrix there always exists a set of n orthonormal eigenvectors, the eigenvalue decomposition can be written as

$$\boldsymbol{A} = \boldsymbol{v}\boldsymbol{\Lambda}\boldsymbol{v}^{\mathrm{H}} = \lambda_1 \boldsymbol{v}_1 \boldsymbol{v}_1^{\mathrm{H}} + \lambda_2 \boldsymbol{v}_2 \boldsymbol{v}_2^{\mathrm{H}} + \cdots + \lambda_n \boldsymbol{v}_n \boldsymbol{v}_n^{\mathrm{H}}$$

where λ_i are the eigenvalues and $\boldsymbol{v}_i$ is a set of orthonormal eigenvectors

Optimization theory

- The local and global minima of an objective function $f(x)$, with x real, satisfy

$$\frac{df(x)}{dx} = 0 \qquad \frac{d^2 f(x)}{dx} > 0$$

  If $f(x)$ is convex, there is only one minimum, which is the global one.

- For an objective function $f(z)$, with z complex,
  - we rewrite $f(z)$ as $f(z, z^*)$ and treat z and z^* as two independent variables
  - minimize $f(z, z^*)$ w.r.t. z and z^*
  - the stationary points of $f(z, z^*)$ are found by setting the derivative of $f(z, z^*)$ w.r.t. to z or z^* to zero
  - but, the direction of the maximum rate of change is the gradient w.r.t. $\boldsymbol{z}^*$

Optimization theory

- For an objective function in two or more real variables, $f(x_1, x_2, \ldots, x_n) = f(\boldsymbol{x})$, the first-order derivative (gradient) and second-order derivative (Hessian) are required

$$\{\nabla_x f(\boldsymbol{x})\}_i = \frac{\partial f(\boldsymbol{x})}{\partial x_i} \qquad \{\boldsymbol{H}_x\}_{ij} = \frac{\partial^2 f(\boldsymbol{x})}{\partial x_i \partial x_j}$$

- The local and global minima of an objective function $f(\boldsymbol{x})$, with $\boldsymbol{x}$ real, satisfy

$$\nabla_x f(\boldsymbol{x}) = \boldsymbol{0} \qquad \boldsymbol{H}_x > 0$$