ELEG 5633 Detection and Estimation
Review of Linear Algebra

Jingxian Wu

Department of Electrical Engineering
University of Arkansas

January 30, 2017

Review of Linear Algebra

- Linear Vector Space
- Bases and Representations
- Orthogonal Projections
- Singular Value Decomposition and Eigen-decomposition

Definition

A linear vector space $\mathcal{X}$ is a collection of elements (called vectors) satisfying the following properties:

addition: $\forall x, y, z \in \mathcal{X}$,

- $x + y \in \mathcal{X}$
- $x + y = y + x$
- $(x + y) + z = x + (y + z)$
- $\exists 0 \in \mathcal{X}$, such that $x + 0 = x$
- $\forall x \in \mathcal{X}$, $\exists -x \in \mathcal{X}$ such that $x + (-x) = 0$

multiplication: $\forall x, y \in \mathcal{X}$ and $a, b \in \mathbb{R}$,

- $a \cdot x \in \mathcal{X}$
- $a \cdot (b \cdot x) = (a \cdot b) \cdot x$
- $(a + b) \cdot x = a \cdot x + b \cdot x$
- $a \cdot (x + y) = a \cdot x + a \cdot y$
- $1 \cdot x = x, \quad 0 \cdot x = 0 \ (0, 1 \in \mathbb{R})$

Example

Verify that the d-dimensional Euclidean space $\mathbb{R}^d$ is a linear vector space

Example

The space of finite energy signals/functions supported on the interval $[0, T]$

$$L_2([0,T]) := \left\{ x(t) : \int_0^T x^2(t)dt < +\infty \right\}$$

Verify that $L_2([0,T])$ is a linear vector space.

More examples of linear vector space

- The space of finite energy sequences

$$\ell_2(\mathbb{Z}) := \left\{ x[n] : \sum_{n=-\infty}^{+\infty} x^2[n] < +\infty \right\}$$

  - Example: $[0.5^n]_{n=0}^{\infty} \in \ell_2(\mathbb{Z})$, $[2^n]_{n=0}^{\infty} \notin \ell_2(\mathbb{Z})$.

- The space of random variables with finite variances

$$L_2(\Omega) := \left\{ X : \mathbb{E}[X^2] < +\infty \right\}$$

  - Example: $X \sim \mathcal{N}(0,1)$, $X \in L_2(\Omega)$.

Definition

A subset $\mathcal{M} \subset \mathcal{X}$ is subspace if $x, y \in \mathcal{M} \Rightarrow ax + by \in \mathcal{M}$, $\forall$ scalars $a, b \in \mathbb{R}$.

- $\mathcal{M}$ is a subspace $\Rightarrow 0 \in \mathcal{M}$

Definition

An inner product is a mapping from $\mathcal{X} \times \mathcal{X}$ to $\mathbb{R}$. The inner product between any $x, y \in \mathcal{X}$ is denoted by $\langle x, y \rangle$ and it satisfies the following properties for all $x, y, z \in \mathcal{X}$:

- $\langle x, y \rangle = \langle y, x \rangle$
- $\langle ax, y \rangle = a\langle x, y \rangle$, $\forall$ scalar a
- $\langle x + y, z \rangle = \langle x, z \rangle + \langle y, z \rangle$
- $\langle x, x \rangle \geq 0$
- $\langle x, x \rangle = 0 \Leftrightarrow x = 0$

Definition

A space $\mathcal{X}$ equipped with an inner product is called an inner product space.

Definition

An inner product space that contains all its limits is called a Hilbert Space, denoted by $\mathcal{H}$; i.e., if $x_1, x_2, \ldots$ are in $\mathcal{H}$ and $\lim_{n\to\infty} x_n$ exists, then the limit is also in $\mathcal{H}$.

Example

- Let $\mathcal{X} = \mathbb{R}^n$. Then $\langle x, y \rangle := x^T y = \sum_{i=1}^n x_i y_i$.
- Let $\mathcal{X} = \ell_2(\mathbb{Z})$. Then $\langle x, y \rangle := \sum_{n=-\infty}^{\infty} x[n]y[n]$
- Let $\mathcal{X} = L_2([0, T])$. Then $\langle x, y \rangle := \int_0^T x(t)y(t)dt$.
- Let $\mathcal{X} = L_2(\Omega)$. Then $\langle x, y \rangle := \mathbb{E}[XY]$.

Definition
The inner product induces a norm defined as

$$\|x\| := \sqrt{\langle x, x\rangle}$$

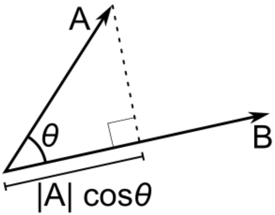
.

Cauchy-Schwarz Inequality

$$|\langle x, y\rangle| \leq \|x\|\|y\|$$

- with "=" iff $x = ay$, $\forall a \in \mathbb{R}$
- $\langle x, y\rangle = \|x\|\|y\| \cos(\theta)$ where $\theta = \arccos\left(\frac{\langle x,y\rangle}{\|x\|\|y\|}\right)$

A Geometric Interpretation



- The norm measures the length/size of x
- θ: the angle between two vectors.
- $|\langle x, y\rangle| = \|x\|\|y\|$ iff x and y are "parallel"; i.e., $\theta = 0, 180^\circ$
- $x \perp y$: Two vectors x, y are orthogonal if $\langle x, y\rangle = 0$; i.e., $\theta = +/-90^\circ$.

Example

Let $\mathcal{X} = \mathbb{R}^2$, then $x = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$ and $y = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$ are orthogonal.

Example

Let $\mathcal{X} = \mathbb{R}^2$, $x = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$ and $y = \begin{bmatrix} 2 \\ 0 \end{bmatrix}$. Find the angle θ between the two vectors.

Bases and Representations

Definition

A collection of vectors $\{x_1, \ldots, x_k\}$ are said to be linearly independent if none of them can be represented as a linear combination of the others. That is, for any x_i and every set of scalar weights $\{\theta_j\}$ we have $x_i \neq \sum_{j:j \neq i} \theta_j x_j$.

Definition

The set of all vectors that can be generated by taking linear combinations of $\{x_1, \ldots, x_k\}$ have the form

$$v = \sum_{i=1}^{k} \theta_i x_i$$

is called the span of $\{x_1, \ldots, x_k\}$, denoted $\text{span}(x_1, \ldots, x_k)$.

Definition

A set of linearly independent vectors $\{\phi_i\}_i$ is a basis for $\mathcal{H}$ if every $x \in \mathcal{H}$ can be represented as a unique linear combination of $\{\phi_i\}$. That is, every $x \in \mathcal{H}$ can be expressed as $x = \sum_i \theta_i \phi_i$ for a certain unique set of scalar weights $\{\theta_i\}$.

- $\{\phi_i\}_i$ is a basis for $\mathcal{H}$ if it spans $\mathcal{H}$ and $\{\phi_i\}_i$ are linearly independent.

Example

Let $\mathcal{H} = \mathbb{R}^2$. Then $\begin{bmatrix} 1 \\ 0 \end{bmatrix}$ and $\begin{bmatrix} 0 \\ 1 \end{bmatrix}$ are a basis (since they are orthogonal). Also, $\begin{bmatrix} 1 \\ 0 \end{bmatrix}$ and $\begin{bmatrix} 1 \\ 1 \end{bmatrix}$ are a basis because they are linearly independent (although not orthogonal).

Example

Let $\{x_i\}_{i=1}^n$ be a basis of $\mathbb{R}^n$. Then $X = [x_1, \cdots, x_n]$ is an $n \times n$ matrix with rank n. Let $y \in \mathbb{R}^n$ and assum $y = \sum_{i=1}^n \theta_i x_i$. Find the vector $\theta = [\theta_1, \cdots, \theta_n]^*$.

Orthonormal Basis

Definition

An orthonormal basis (orthobasis) is a basis $\{\phi_i\}_i$ satisfying

$$\langle \phi_i, \phi_j \rangle = \delta_{ij} := \begin{cases} 1, & i = j \\ 0, & i \neq j \end{cases}$$

Every $x \in \mathcal{H}$ can be represented in terms of an orthobasis $\{\phi_i\}_i$ as

$$x = \sum_i a_i \phi_i, \quad a_i = \langle x, \phi_i \rangle$$

Definition

Parsaval's Relation

Let $\{\phi_i\}_i$ be an orthonormal basis of $\mathcal{H}$, and

$$x = \sum_i a_i \phi_i, \quad a_i = \langle x, \phi_i \rangle$$

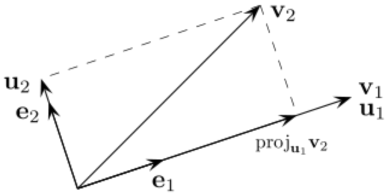
$$\begin{aligned} x &\longleftrightarrow a_i = \langle x, \phi_i \rangle \\ y &\longleftrightarrow b_i = \langle y, \phi_i \rangle \\ \langle x, y \rangle &= \sum_i a_i b_i \\ \|x\| &= \sum_i a_i^2 \end{aligned}$$

- An element x (function, random variable, vector) in a linear vector space can be equivalently represented as a vector $[a_1, \cdots, a_n] \in \mathbb{R}^n$ or $\{a[n]\}_n \in \ell_2(\mathbb{Z})$

Gram-Schmidt Orthogonalization

Any basis can be converted into an orthonormal basis using Gram-Schmidt Orthogonalization. Let $\{v_i\}$ be a basis for a vector space $\mathcal{X}$. An orthobasis $\{e_i\}$ for $\mathcal{X}$ can be constructed as follows.

$$
\begin{aligned}
u_1 &:= v_1; \quad e_1 = u_1/\|u_1\| \\
u_2 &:= v_2 - \langle v_2, e_1\rangle e_1; \quad e_2 = u_2/\|u_2\| \\
&\vdots \\
u_k &:= v_k - \sum_{i=1}^{k-1}\langle v_k, e_i\rangle e_i; \quad e_k = u_k/\|u_k\|
\end{aligned}
$$



Example

Perform Gram-Schmidt orthogonalization for the following basis of $\mathbb{R}^2$:
$v_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$ and $v_2 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$

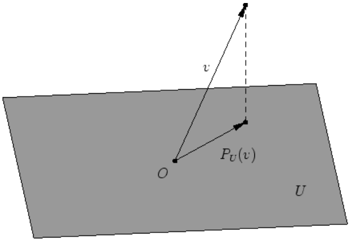
Orthogonal Projections

- Let $\mathcal{H}$ be a Hilbert space, and $\mathcal{M} \subset \mathcal{H}$ be a subspace. Let $x \in \mathcal{H}$, then

$$\hat{x} := \arg\min_{v \in \mathcal{M}} \|x - v\|$$

  is called the projection of x onto $\mathcal{M}$.
- $\hat{x}$ is the optimal approximation to x in terms of vectors in $\mathcal{M}$.
- The "error" is orthogonal to the subspace $\mathcal{M}$:

$$\begin{aligned} &e = x - \hat{x} \\ &e \perp \mathcal{M}, \quad i.e., \langle e, v \rangle = 0 \quad \forall v \in \mathcal{M} \end{aligned}$$



Orthogonal Projections

- let $\{\phi_i\}_{i=1}^r$ be an orthobasis for $\mathcal{M}$, i.e., $\mathcal{M}$ is spanned by $\{\phi_i\}_{i=1}^r$.
- $\mathcal{M}$ is an r-dimensional subspace of $\mathcal{H}$. For any $x \in \mathcal{H}$, the projection of x onto $\mathcal{M}$ is given by

$$y = \sum_{i=1}^{r} \langle \phi_i, x \rangle \phi_i$$

Example

Let $\mathcal{H} = \mathbb{R}^2$. Consider the canonical coordinate system $\phi_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$ and $\phi_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$. Consider the subspace spanned by ϕ_1. The projection of any $x = [x_1, x_2]^T \in \mathbb{R}_2$ onto this subspace is

$$P_1 x = \langle x, \begin{bmatrix} 1 \\ 0 \end{bmatrix} \rangle \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} x_1 \\ 0 \end{bmatrix}$$

The projection operator P_1 is just a matrix and it is given by

$$P_1 := \phi_1 \phi_1^T = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \begin{bmatrix} 1 & 0 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$$

If $\phi_1 = \begin{bmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{bmatrix}$ and $\phi_2 = \begin{bmatrix} 1/\sqrt{2} \\ -1/\sqrt{2} \end{bmatrix}$, what is the projection operator onto the span of ϕ_1 in this case?

Projection Matrix

- More generally suppose we are considering $\mathbb{R}^n$ and we have an orthobasis $\{\phi_i\}_{i=1}^r$ for some r-dimensional, $r < n$, subspace $\mathcal{M}$ of $\mathbb{R}^n$. Then the projection matrix is given by

$$P_{\mathcal{M}} = \sum_{i=1}^{r} \phi_i \phi_i^T = \Phi\Phi^T$$

where $\Phi = [\phi_1, \ldots, \phi_r]$, a matrix whose columns are the basis vectors.

- Moreover, if $\{\phi_i\}_{i=1}^r$ is a basis for $\mathcal{M}$, but not necessarily orthonormal, then

$$P_{\mathcal{M}} = \Phi(\Phi^T\Phi)^{-1}\Phi^T$$

Singular Value Decomposition (SVD)

A is an $m \times n$ matrix with entries from a field (e.g., $\mathbb{R}$ or $\mathbb{C}$). Then there exists a factorization of the form $A = UDV^*$:

- $U = [u_1 \dots u_m]$ is $m \times m$ with orthonormal columns, i.e., $U^*U = I_m$
- $V = [v_1 \dots v_n]$ is $n \times n$ with orthonormal columns, i.e., $V^*V = I_n$
- D is $m \times n$ and has the form

$$\begin{bmatrix} \sigma_1 & 0 & 0 & 0 & \dots & 0 \\ 0 & \sigma_2 & 0 & 0 & \dots & 0 \\ \vdots & 0 & \ddots & 0 & \dots & 0 \\ 0 & \dots & 0 & \sigma_m & 0 & \dots \end{bmatrix}$$

- The values $\sigma_1, \dots, \sigma_m$ are called the singular values of A.
- The factorization is called the singular value decomposition (SVD).
- Because of the orthonormality of the columns of U and V we have

$$Av_i = \sigma_i u_i, \quad A^* u_i = \sigma_i v_i, \quad i = 1, \dots, m$$

Eigenvalue Decomposition (EVD)

Letting the matrix A be square, the non-zero vector v is said to be an eigenvector of $\mathbf{A}$ if it satisfies

$$Av = \lambda v$$

- λ: scalar termed the eigenvalue associated with v.

Real symmetric matrices always have real eigenvalues and have an eigendecomposition of the form $A = UDU^*$,

- columns of U the orthonormal eigenvectors of A
- $D = \text{diag}(\lambda_1, \ldots, \lambda_n)$, and diagonal entries are the eigenvalues.
- This is just a special case of the SVD.

A symmetric pos-semidefinite matrix satisfies $v^T Av \geq 0$ for all v.
Thus, eigenvalues of symmetric pos-semidefinite matrices are non-negative.

SVD and EVD

- A is an $m \times n$ matrix with $n \geq m$. Then there exists a factorization of the form $A = UDV^*$, where $U^*U = I_m$ and $V^*V = I_n$. Denote the singular values as $\{\sigma_1, \cdots, \sigma_m\}$

- Define $B_1 = AA^*$ ($m \times m$ matrix), then

$$B_1 = UDV^*VD^*U^* = U\Lambda_1U^*$$

where $\Lambda_1 = \text{diag}([\lambda_1, \cdots, \lambda_m])$, with $\lambda_i = |\sigma_i|^2$, for $i = 1, \cdots, m$.

- Define $B_2 = A^*A$ ($n \times n$ matrix), then

$$B_1 = V^*D^*U^*UDV = V^*\Lambda_2V$$

where $\Lambda_2 = \text{diag}([\lambda_1, \cdots, \lambda_m, 0, \cdots, 0])$, with $\lambda_i = |\sigma_i|^2$, for $i = 1, \cdots, m$.

The eigenvalues of AA^* and A^*A are the amplitude squred singular values of A .

Example

Let X be a random vector taking values in $\mathbb{R}^n$ and recall the definition of the covariance matrix:

$$\Sigma := \mathbb{E}[(X - \mu)(X - \mu)T]$$

It is easy to see that $v^T \Sigma v \geq 0$, and Σ is symmetric. Therefore, every covariance matrix has an eigendecomposition of the form $\Sigma = UDU^*$, where $D = \mathrm{diag}(\lambda_1, \ldots, \lambda_n)$ and $\lambda_i \geq 0$ for $i = 1, \ldots, n$.

Example

Perform eigenvalue decomposition of the matrix

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