2: Review of Linear Algebra

ECE 830, Spring 2014

Signal vectors

Linear vector space

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

- **addition:** $\forall\, x, y, z \in \mathcal{X}$
  1. $x + y$
  2. $x + y =$
  3. $(x + y) + z =$
  4. $\exists\ 0 \in \mathcal{X},$ such that $x + 0 =$
  5. $\forall x \in \mathcal{X}, \exists\ -x \in \mathcal{X}$ such that $x + (-x) =$
- **multiplication:** $\forall\, x, y \in \mathcal{X}$ and $a, b \in \mathbb{R}$
  1. ax
  2. $a(bx) =$
  3. $1x = x$, $0x = 0$
  4. $a(x + y) =$

Example: $\mathbb{R}^n$

$\mathbb{R}^n$, the n-dimensional Euclidean space, is a linear vector space.

Example: $\mathbb{C}^n$

$\mathbb{C}^n$, the n-dimensional complex space, is a linear vector space.

Example: $L_2([0,T])$

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

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

is a linear vector space.

Inner Products

Definition: Inner product

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

1. $\langle x, y \rangle =$
2. $\langle ax, y \rangle =$ for all scalars a
3. $\langle x + y, z \rangle =$
4. $\langle x, x \rangle \geq$ and $\langle x, x \rangle = 0 \implies$

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

Roughly speaking, $\langle x, y \rangle$ measures the alignment or similarity of x and y.

Definition: Orthogonal vectors

x and y are orthogonal vectors if

Example: Euclidean space

Let $\mathcal{X} = \mathbb{R}^n$. Then $\langle x, y \rangle :=$.

Example: Complex space

Let $\mathcal{X} = \mathcal{C}^n$. Then $\langle x, y \rangle :=$.

Example: L_2

Let $\mathcal{X} = L_2([0,1])$. Then $\langle x, y \rangle :=$.

Norms

Definition: Norm

The inner product induces a norm defined as $\|x\| := \sqrt{\langle x, x\rangle}$.

The norm measures the length/size of x. The inner product $\langle x, y\rangle = \|x\| \, \|y\| \cos(\theta)$, where θ is the angle between x and y.

Cauchy-Schwarz Inequality

For every $x, y \in \mathcal{X}$ we have

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

with equality if and only if x and y are linearly dependent or "parallel"; i.e., $\theta = 0$.

Triangle inequality

$$\|x - y\| \leq$$

Homogeneity

For $x \in \mathcal{X}$ and a a scalar

$$\|ax\| = \qquad\qquad .$$

Pythagorean theorem

If $x, y \in \mathcal{X}$ and $\langle x, y \rangle = 0$, then

$$\|x\|^2 + \|y\|^2 =$$

Parallelogram law

If $x, y \in \mathcal{X}$,

$$\|x + y\|^2 + \|x - y\|^2 =$$

Hilbert spaces

Definition: Hilbert space

An inner product space that contains all its limits is called a **Hilbert Space** and in this case we often denote the space 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 $\mathcal{H}$.

It is easy to verify that $\mathbb{R}^n$, $L_2([0,T])$, and $\ell_2(\mathbb{Z})$, the set of all finite energy sequences (e.g., discrete-time signals), are all Hilbert spaces.

Linear Independence

Definition: Linear independence

Consider a set of vectors

$$x_1, x_2, \ldots, x_p \in \mathcal{X}$$

If there exists a set of numbers $a_1, a_2, \ldots, a_p \in \mathbb{R}$ such that not all are zero and

$$\tag{1}$$

then we say that the vectors are **linearly dependent**. If Eq. 1 only holds for the case $a_1 = a_2 = \cdots = a_p = 0$, then the vectors are said to be **linearly independent**.

Note that if Eq. 1 holds and $a_k \neq 0$ then

$$x_k =$$

and x_k can be expressed as a linear combination of the other vectors (hence the term *dependent*).

Bases

Definition: Basis

A set of vectors $\phi_1, \ldots, \phi_n$ is a **basis** for $\mathcal{X}$ if every vector $x \in \mathcal{X}$ can be represented as a linear combination of $\{\phi_k\}_{k=1}^n$. That is, there exist numbers $\theta_1, \ldots, \theta_n$ so that

$$x =$$

Definition: Orthonormal basis

A basis $\{\phi_i\}$ is orthonormal if

$$\phi_i^\top \phi_j =$$

Orthobasis of Hilbert space

Every $x \in \mathcal{H}$ can be represented in terms of an orthonormal basis $\{\phi_i\}_{i\geq 1}$ (or 'orthobasis' for short) according to:

$$x \;=\; \sum_{i\geq 1}\langle x, \phi_i\rangle \; \phi_i$$

This is easy to see as follows. Suppose x has a representation $\sum_i \theta_i \phi_i$. Then

Example: Sample space orthonormal basis

$$\phi_k = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \qquad \text{i.e.} \qquad \phi_{k,i} = \begin{cases} 0 & i \neq k \\ 1 & i = k \end{cases} \tag{2}$$

Then $\{\phi_k\}_k$ is a basis for $\mathbb{R}^n$.

Example: Orthobasis of $L_2([0,1])$

For $i = 1, 2, \ldots$

$$\begin{aligned} \phi_{2i-1}(t) &:= \sqrt{2}\cos(2\pi(i-1)t) \\ \phi_{2i}(t) &:= \sqrt{2}\sin(2\pi i t) \end{aligned}$$

Gram-Schmidt Orthogonalization

Any basis can be converted into an orthonormal basis using *Gram-Schmidt Orthogonalization*.

Start with an arbitrary (non-orthogonal) basis $\{\phi_i\}$.

1. $$\psi_1 := \phi_1 / \|\phi_1\|$$

2. $$\begin{aligned} \widetilde{\psi}_2 &:= \phi_2 - \langle \psi_1, \phi_2 \rangle \psi_1 \\ \psi_2 &:= \widetilde{\psi}_2 / \|\widetilde{\psi}_2\| \end{aligned}$$

3. for $k = 3, \ldots, n$,

$$\begin{aligned} \widetilde{\psi}_k &:= \phi_k - \sum_{i=1}^{k-1} \langle \psi_i, \phi_k \rangle \psi_i \\ \psi_k &:= \widetilde{\psi}_k / \|\widetilde{\psi}_k\| \end{aligned}$$

Subspaces

Consider a set of vectors $x_1, x_2, \ldots, x_p \in \mathcal{X}$. The span of these vectors is the set of all vectors $x \in \mathcal{X}$ that can be generated from linear combinations of the set

$$\operatorname{span}\left(\{x_i\}_{i=1}^{p}\right) := \left\{ x : x = \sum_{i=1}^{p} a_i x_i, \;\; a_1, \ldots, a_p \in \mathbb{R} \right\}$$

This set is also called a signal subspace of $\mathcal{X}$.

Definition: subspace

A subset $\mathcal{M} \subset \mathcal{X}$ is a subspace if

$$x, y \in \mathcal{M} \Longrightarrow$$

If $\phi_1, \ldots, \phi_p$ is an orthonormal basis for $\mathcal{M} \subset \mathbb{R}^n$, then every $x \in \mathcal{M}$ can be written as

$$x = \sum_{i=1}^{p} \theta_i \phi_i.$$

Hence, even though the signal x is a length-n vector, the fact that it lies in the subspace $\mathcal{M}$ means that it is actually a function of only $p \leq n$ free parameters or "degrees of freedom".

We say that $\mathcal{M}$ is a p-dimensional subspace of $\mathbb{R}^n$ (and it is isometric to $\mathbb{R}^r$).

Example: $n = 3$

$$\phi_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \quad \phi_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$$

$$\mathcal{M} = \operatorname{span}(\phi_1, \phi_2) = \left\{ \begin{bmatrix} \\ \\ \\ \end{bmatrix} : a, b \in \mathbb{R} \right\}$$

Example: $n = 3$

$$\phi_1 = \begin{bmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \\ 0 \end{bmatrix} \quad \phi_2 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$$

$$\mathcal{M} = \operatorname{span}(\phi_1, \phi_2) = \left\{ \begin{bmatrix} \\ \\ \\ \end{bmatrix} : a, b \in \mathbb{R} \right\}$$

Orthogonal Projections

Let $\mathcal{H}$ be a Hilbert space and let $\mathcal{M} \subset \mathcal{H}$ be a subspace. Every $x \in \mathcal{H}$ can be written as

$$x = y + z$$

where $y \in \mathcal{M}$ and $z \perp \mathcal{M}$, which is shorthand for z orthogonal to $\mathcal{M}$; that is

$$\forall v \in \mathcal{M}, \qquad \langle v, z \rangle = 0.$$

The vector y is the optimal approximation to x in terms of vectors in $\mathcal{M}$ in the following sense:

$$\|x - y\| \;=\; \min_{v \in \mathcal{M}} \|x - v\|$$

The vector y is called the projection of x onto $\mathcal{M}$.

Orthogonal subspace projection

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

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

and this projection can be viewed as a sort of filter that removes all components of the signal x that are orthogonal to $\mathcal{M}$.

Example:

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

$$P_1 x \; =$$

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

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

It is also easy to check that $\phi_1 = \begin{bmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{bmatrix}$ and $\phi_2 = \begin{bmatrix} 1/\sqrt{2} \\ -1/\sqrt{2} \end{bmatrix}$ is an orthobasis for $\mathbb{R}^2$. What is the projection operator onto the span of ϕ_1 in this case?

Orthogonal projections in Euclidean subspaces

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

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

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

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

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

Example:

Let $\mathcal{H} = L_2([0,1])$ and let $\mathcal{M} = \{\text{linear functions on } [0,1]\}$. Since all linear functions have the form $y(t) = at + b$, for $t \in [0,1]$, here is a basis for $\mathcal{M}$:

$$\phi_1(t) =$$
$$\phi_2(t) =$$

Note that this means that $\mathcal{M}$ is two-dimensional. That makes sense since every line is defined by its slope and intercept (two real numbers). Using the Gram-Schmidt procedure we can construct the orthobasis $\psi_1(t) = 1$, $\psi_2(t) = (t - 1/2)/\sqrt{12}$. Now, consider any signal/function $x \in L_2([0,1])$. The projection of x onto $\mathcal{M}$ is

$$\begin{aligned} [P_{\mathcal{M}}\, x](t) &= \langle x, 1\rangle + \langle x, t - 1/2\rangle (t - 1/2) \\ &= \end{aligned}$$

Eigendecomposition of a Symmetric Matrix

Let C be a real, symmetric matrix ($C^\top = C$). $v \in \mathbb{R}^n$ is an eigenvector of C if

$$Cv = \underbrace{\lambda}_{\text{eigenvalue}} v$$

If C is $n \times n$, then there are n orthonormal eigenvectors $\{v_1, \ldots, v_n\}$ such that

$$\langle v_i, v_j \rangle = \delta_{i,j}.$$

Let $V = [v_1, \ldots, v_n]$. Then

$$C = V \Lambda V^\top$$

where

$$\Lambda = \begin{bmatrix} \lambda_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \lambda_n \end{bmatrix}$$

Singular Value Decomposition

The SVD of an $n \times p$ matrix H is written as

$$H = \underbrace{U}_{n\times p} \underbrace{\Sigma}_{p\times p} \underbrace{V^\top}_{p\times p}$$

- $U = [u_1 \cdots u_p]$ and $\{u_i\}_{i=1}^p$ are $n \times 1$ vectors called the left singular vectors of H. $U^\top U = I$
- $\Sigma = \begin{bmatrix} \sigma_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \sigma_p \end{bmatrix}$ (a diagonal $p \times p$ matrix) and $\{\sigma_i\}_{i=1}^p$ are the singular values of H, sorted so $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_p \geq 0$.
- $V = [v_1 \cdots v_p]$ and $\{v_i\}_{i=1}^p$ are $p \times 1$ vectors called the right singular vectors of H. $V^\top V = I$

Also note that

$$\begin{aligned} HH^\top &= \\ &= \\ H^\top H &= \\ &= \end{aligned}$$

So, $\{\sigma_1^2, \ldots, \sigma_p^2\}$ are the eigenvalues of $H^\top H$ and $\{v_1, \ldots, v_p\}$ are the corresponding eigenvectors. Also, $\{\sigma_1^2, \ldots, \sigma_p^2\}$ are the first p eigenvalues of $HH^\top$ ($n - p$ remaining eigenvalues are identically zero) and $\{u_1, \ldots, u_p\}$ are the associated eigenvectors.

Application of SVD

Suppose we want to solve the following overdetermined set of linear equations:

$$\underbrace{x}_{n\times 1} = \underbrace{H}_{n\times p}\underbrace{\theta}_{p\times 1}, \quad p < n$$

where x is the observation, H is a known matrix, and θ is unknown.
If H were square (and non-singular), then $\theta = H^{-1}x$, but here H is not square. Notice that

$$x = U\Sigma V^\top \theta$$

If the p columns of H are linearly independent, then $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_p > 0$. So, we can proceed in the following manner:

$$\begin{aligned} U^\top x &= \qquad\qquad\qquad\qquad, \text{ since } U^\top U = I_{p\times p} \\ \Sigma^{-1}U^\top x &= \qquad\qquad\qquad\qquad, \text{ since } \sigma_i > 0,\ i = 1, \ldots, p \\ V\Sigma^{-1}U^\top x &= \qquad\qquad\qquad\qquad, \text{ since } VV^\top = I_{p\times p} \end{aligned}$$

Definition: pseudoinverse

$$H^{\#} \equiv V\Sigma^{-1}U^{\top} = (H^{\top}H)^{-1}H^{\top}$$

is called the pseudoinverse of H.

proof: First, recall that $(ABC)^{-1} = C^{-1}B^{-1}A^{-1}$ and $V^{-1} = V^{\top}$.

$$(H^{\top}H)^{-1}H^{\top} =$$

$$=$$

$$=$$

$$=$$

$$=$$

Also, notice

$$
\begin{aligned}
x &= H\theta \\
H^\top x &= H^\top H\theta \\
\theta &= \underbrace{(H^\top H)^{-1}H^\top}x \\
\\
H\theta &= \underbrace{H(H^\top H)^{-1}H^\top}x
\end{aligned}
$$

where $H^\top H$ is $p \times p$ and invertible when columns of H are linearly independent.