ECE 830 Fall 2010 Statistical Signal Processing

instructor: R. Nowak , **scribe:** R. Nowak

Lecture 3: Review of Linear Algebra

Very often in this course we will represent signals as vectors and operators (e.g., filters, transforms, etc) as matrices. This lecture reviews basic concepts from linear algebra that will be useful.

1 Linear Vector Space

Definition 1 *A linear vector space $\mathcal{X}$ is a collection of elements satisfying the following properties:*

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

a) $x + y \in \mathcal{X}$

b) $x + y = y + x$

c) $(x + y) + z = x + (y + z)$

d) $\exists\ 0 \in \mathcal{X},$ *such that* $x + 0 = x$

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

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

a) $a\, x \in \mathcal{X}$

b) $a(b\, x) = (ab)\, x$

c) $1x = x,\ 0x = 0$

d) $a(x + y) = ax + ay$

Example 1 *Here are two examples of linear vector spaces. The familiar d-dimensional Euclidean space $\mathbb{R}^d$ and the space of finite energy signals/functions supported on the interval $[0.T]$*

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

It is easy to verify the properties above for both examples.

Definition 2 *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.*

Definition 3 *An* inner product *is a mapping from $\mathcal{X} \times X$ to $\mathbb{R}$. The inner product between any $x, y \in \mathcal{X}$ is denoted by $\langle x, y \rangle$ and it must satisfy two properties for all $x, y, z \in \mathcal{X}$:*

a) $\langle x, y \rangle = \langle y, x \rangle$

b) $\langle ax, y \rangle = a\langle x, y \rangle$, $\forall$ *scalars* a

c) $\langle x + y, z \rangle = \langle x, z \rangle + \langle y, z \rangle$

d) $\langle x, x \rangle \geq 0$

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

Example 2 *Let* $\mathcal{X} = \mathbb{R}^n$. *Then* $\langle x, y\rangle := x^T y = \sum_{i=1}^n x_i y_i$.

Example 3 *Let* $\mathcal{X} = L_2([0,1])$. *Then* $\langle x, y\rangle := \int_0^1 x(t)y(t)\,dt$.

The inner product measures the alignment of the two vectors. 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. Thus, in general, 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., $\theta = 0$. This is called the *Cauchy-Schwarz Inequality*. Two vectors x, y are *orthogonal* if $\langle x, y\rangle = 0$.

Example 4 *Let* $\mathcal{X} = \mathbb{R}^2$, *then* $x = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$ *and* $y = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$ *are orthogonal, as are* $u = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$ *and* $v = \begin{bmatrix} 1 \\ -1 \end{bmatrix}$.

Definition 4 *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.

2 Bases and Representations

Definition 5 *A collection of vectors* $\{x_1, \ldots, x_k\}$ *are said to be* linearly independent *if none of them can be written as a linear combination of the others. That is, for any* x_i *and every set of scalar weights* $\{w_j\}$ *we have* $x_i \neq \sum_{j\neq i} w_j x_j$.

Definition 6 *A set of linearly independent vectors* $\{\phi_i\}_{i\geq 1}$ *is a* basis *for* $\mathcal{H}$ *if every* $x \in \mathcal{H}$ *can be represented as a linear combination of* $\{\phi_i\}$. *That is, every* $x \in \mathcal{H}$ *can be expressed as*

$$x \;=\; \sum_{i\geq 1} \theta_i \phi_i$$

for a certain unique set of scalar weights $\{\theta_i\}$.

Example 5 *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).*

Definition 7 *An* orthonormal basis *is one satisfying*

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

Any basis can be converted into an orthonormal basis using the *Gram-Schmidt Orthogonalization* procedure. Let $\{\phi_i\}$ be a basis. Generate an orthonormal basis according to the following steps.

1. $\psi_1 := \phi_1/\|\phi_1\|$
2. $\psi_2' := \phi_2 - \langle \psi_1, \phi_2\rangle\psi_1$; $\psi_2 = \psi_2'/\|\psi_2'\|$

$\vdots$

k. $\psi_k' = \phi_k - \sum_{i=1}^{k-1}\langle \psi_i, \phi_k\rangle\psi_i$; $\psi_k = \psi_k'/\|\psi_k'\|$

$\vdots$

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 $\langle x, \phi_j \rangle = \langle \sum_i \theta_i \phi_i, \phi_j \rangle = \sum_i \theta_i \langle \phi_i, \phi_j \rangle = \sum_i \theta_i \delta_{i,j} = \theta_j$.

Example 6 *Here is an orthobasis for $L_2([0,1])$: for $i = 1, 2, \ldots$*

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

Doesn't it look familiar?

3 Orthogonal Projections and Filters

One of the most important tools that we will use from linear algebra is the notion of an *orthogonal projection*. Most linear filters used in signal processing (e.g., bandpass filters, averaging filters, etc.) may all be interpreted as or related to orthogonal projections. Let $\mathcal{H}$ be a Hilbert space and let $\mathcal{M} \subset \mathcal{H}$ be a subspace. Every $x \in H$ 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 $\forall v \in \mathcal{M}$, $\langle v, z \rangle = 0$. 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 $\mathcal{M}$.

Here is an application of this concept. Let $\mathcal{M} \subset \mathcal{H}$ and let $\{\phi_i\}_{i=1}^r$ be an orthobasis for $\mathcal{M}$. We say that the subspace $\mathcal{M}$ is *spanned* by $\{\phi_i\}_{i=1}^r$. Note that this implies that $\mathcal{M}$ is an r-dimensional subspace of $\mathcal{H}$ (and it is isometric to $\mathbb{R}^r$). 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 that are orthogonal to $\mathcal{M}$.

Example 7 *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*

$$P_1 x \;=\; \langle x, \begin{bmatrix} 1 \\ 0 \end{bmatrix} \rangle \begin{bmatrix} 1 \\ 0 \end{bmatrix} \;=\; \left([x_1\, x_2] \begin{bmatrix} 1 \\ 0 \end{bmatrix} \right) \begin{bmatrix} 1 \\ 0 \end{bmatrix} \;=\; \begin{bmatrix} x_1 \\ 0 \end{bmatrix}$$

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

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

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?

More generally suppose we are considering $\mathbb{R}^n$ and we have a 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 $P_{\mathcal{M}} = \sum_{i=1}^r \phi_i \phi_i^T$. Moreover, if $\{\phi_i\}_{i=1}^r$ is a basis for $\mathcal{M}$, but not necessarily orthonormal, then $P_{\mathcal{M}} = \Phi(\Phi^T\Phi)^{-1}\Phi$, where $\Phi = [\phi_1, \ldots, \phi_r]$, a matrix whose columns are the basis vectors.

Example 8 *Let $\mathcal{H} = L_2([0,1])$ and let $\mathcal{M} = \{$linear functions on $[0,1]\}$. Since all linear functions have the form $at + b$, for $t \in [0,1]$, here is a basis for $\mathcal{M}$: $\phi_1(t) = 1$, $\phi_2(t) = t$. 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 $\psi_1(t) = 1$, $\psi_2(t) = t - 1/2$. Now, consider any signal/function $x \in L_2([0,1])$. The projection of x onto $\mathcal{M}$ is*

$$\begin{aligned} P_{\mathcal{M}}\, x &= \langle x, 1\rangle + \langle x, t - 1/2\rangle (t - 1/2) \\ &= \int_0^1 x(\tau)\, d\tau \;+\; (t - 1/2) \int_0^1 (\tau - 1/2) x(t)\, d\tau \end{aligned}$$

4 Eigenanalysis

Suppose A is an $m \times n$ matrix with entries from a field (e.g., $\mathbb{R}$ or $\mathbb{C}$, the latter being the complex numbers). Then there exists a factorization of the form

$$A \;=\; U\, D\, V^*$$

where $U = [u_1 \cdots u_m]$ is $m \times m$ with orthonormal columns, $V = [v_1 \cdots v_n]$ is $n \times n$ with orthonormal columns and the superscript * means transposition and conjugation (if complex-valued), and the matrix D is $m \times n$ and has the form

$$D \;=\; \begin{bmatrix} \sigma_1 & 0 & 0 & \cdots & 0 \\ 0 & \sigma_2 & 0 & \cdots & 0 \\ \vdots & 0 & \ddots & \cdots & \vdots \\ 0 & 0 & \cdots & \sigma_m & 0 \cdots \end{bmatrix}$$

The values $\sigma_1, \ldots, \sigma_m$ are called the singular values of A and the factorization is called the *singular value decomposition* (SVD). Because of the orthonormality of the columns of U and V we have $Av_i = \sigma_i u_i$ and $A^* u_i = \sigma_i v_i$, $i = 1, \ldots, m$.

A vector u is called an *eigenvector* of A if $A\, u = \lambda\, u$ for some scalar λ. The scalar λ is called the *eigenvalue* associated with u. Symmetric matrices (which are always square) always have real eigenvalues and have an *eigendecomposition* of the form $A = UDU^*$, where the columns of U the orthonormal eigenvectors of A, D is a diagonal matrix written $D = \mathrm{diag}(\lambda_1, \ldots, \lambda_n)$, and diagonal entries are the eigenvalues. This is just a special case of the SVD. A symmetric positive-semidefinite matrix satisfies the property $v^T A v \geq 0$ for all v. This implies that the eigenvalues of symmetric positive-semidefinite matrices are non-negative.

Example 9 *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 of course Σ is symmetric. Therefore, every covariance matrix has an eigendecomposition of the form $\Sigma = UDU^*$, where $D = diag(\lambda_1, \ldots, \lambda_n)$ and $\lambda_i \geq 0$ for $i = 1, \ldots, n$.

The *Karhunen-Lo'eve Transform* (KLT), which is also called *Principal Component Analysis* (PCA), is based on transforming a random vector X into the coordinate system associated with the eigendecomposition of the covariance of X. Let X be an n-dimensional random vector with covariance matrix $\Sigma = UDU^*$. Let $u_1, \ldots, u_n$ be the eigenvectors. Assume that the eigenvectors and eigenvalues are ordered such that $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$. The KLT or PCA representation of the random vector X is given by

$$X \;=\; \sum_{i=1}^n (u_i^T X) u_i$$

The coefficients in this representation can be arranged in a vector as $\theta = U^T X$, where U is as defined above. The vector θ is called the KLT or PCA of X. Using this representation we can define the approximation to X in the span of the first $r < n$ eigenvectors

$$X_r \;=\; \sum_{i=1}^{r} (u_i^T X) u_i$$

Note that this approximation involves only r scalar random variables $\{(u_i^T X)\}_{i=1}^r$ rather than n. In fact, it is easy to show that among all r-term linear approximations of X in terms of r random variables, X_r has the smallest mean square error; that is if we let $\mathcal{S}_r$ denote all r-term linear approximations to X, then

$$\mathbb{E}[\|X - X_r\|^2] \;=\; \min_{Y_r \in \mathcal{S}_r} \mathbb{E}[\|X - Y_r\|]$$