ECSE 6520: Estimation and Detection Theory

Linear Algebra Preliminaries, Signal Subspace Model and Deterministic Least Squares

Class Notes - 2

January 26, 2014

Contents

1 Notation **2**

2 Basic Definitions **2**

3 Eigen Decomposition of Hermitian Symmetric Matrices **4**

4 Quadratic Forms and Positive Definiteness **4**

5 Singular value Decomposition of a Matrix **5**

6 Vector Differentiation **5**

7 Signal Subspace Model **6**

8 Deterministic Least Squares **7**

1 Notation

- In general upper case letters, e.g. X, Y, Z, from the end of the alphabet denote random variables, i.e. functions on a sample space, and their lower case versions, e.g. x, denote realizations, i.e. evaluations of these functions at a sample point, of these random variables.
- We reserve lower case letters from the beginning of the alphabet, e.g. a, b, c, for constants and lower case letters in the middle of the alphabet, e.g. i, j, k, l, m, n, for integer variables.
- The letter f is reserved for a probability density function and p is reserved for a probability mass function. Finally in many cases we deal with functions of two or more variables, e.g. the density function $f(x;\theta)$ or $f_\theta(x)$ of a random variable X parameterized by a parameter θ.
- However, when dealing with multivariate densities for clarity we will prefer to explicitly subscript with the appropriate ordering of the random variables, e.g. $f_{X,Y}(x,y;\theta)$, $f_\theta(x,y)$ or $f_{X|Y}(x|y;\theta)$.
- We will define vectors as column vectors unless otherwise specified; and use T to denote its transpose, e.g. $x = [x_1, x_2, ..., x_N]^T$. We will use H to denote Hermitian transpose, i.e., $x^H = [x_1^*, x_2^*, ..., x_N^*]$ where x_i^* denotes complex conjugate.

2 Basic Definitions

- **2-Norm of a vector -** We will denote the length of a vector $x \in \mathbb{C}^n$ by $\|x\| = \sqrt{x^H x} = \sqrt{\sum_{i=1}^n |x_i|^2}$.
- **Distance between two vectors -** We will denote the "distance" between two vectors $x, y \in \mathbb{C}^n$ as $x^H y$. Note that xy^H is called the "outer product" of x and y.

- **Orthogonal vectors -** Two vectors $x \in \mathbb{C}^n$ and $y \in \mathbb{C}^n$ are said to be orthogonal if $x^H y = 0$. In addition, if $\|x\| = 1$ and $\|y\| = 1$, then x and y are said to be orthonormal.

- **Linear independence of vectors -** Let $x_1, ...x_n$ be a set of p dimensional (column) vectors. $x_1, ...x_n$ is said to be linearly independent if $c_1 x_1 + c_2 x_2 + ... + c_n x_n = 0$ implies that $c_i = 0$ for all $i = 1, ..., n$.

- **Linear span of a set of vectors -** Let $x_1, ...x_n$ be a set of p dimensional (column) vectors and construct the $p \times n$ matrix

$$X = [x_1, ..., x_n].$$

Let $a = [a_1, ...a_n]^T$ be a vector of coefficients. Then $y = \sum_{i=1}^{n} a_i x_i = X_a$ is another p dimensional vector that is a linear combination of the columns of X. The linear span of the vectors $x_1, ..., x_n$, equivalently, the column space or range of X, is defined as the subspace of $\mathbb{R}^p$ that contains all such linear combinations. In other words, when we allow a to sweep over its entire domain $\mathbb{R}^n$, y sweeps over the linear span of $x_1, ..., x_n$.

- **Rank of a matrix -** The (column) rank of a matrix A is equal to the number its columns which are linearly independent.

- **Orthogonal matrices -** A real square matrix A is said to be orthogonal if all of its columns are orthonormal, i.e.,

$$A^T A = I.$$

Thus if A is an orthogonal matrix, it is invertible and has a very simple inverse $A^{-1} = A^T$.

- **Unitary matrices -** The generalization of orthogonality to complex matrices A is the property of being unitary,

$$A^H A = I.$$

3 Eigen Decomposition of Hermitian Symmetric Matrices

If R is arbitrary $n \times n$ symmetric matrix, that is, $R^T = R$, then there exist a set of n orthonormal ${v_i}^T v_j = \delta(i-j)$ and a set of associated eigenvectors λ_i such that: $Rv_i = \lambda_i v_i$ for $i = 1, ..., n$. This result extends to Hermitian symmetric matrices, i.e., to matrices where $R^H = R$.

Theorem 1 *If $A \in \mathbb{C}^{N \times N}$ is Hermitian, then there exists a unitary matrix U and a diagonal matrix Λ such that*

$$A = U \Lambda U^H.$$

If $A \in \mathbb{R}^{N \times N}$ is symmetric, the same result holds where now U is orthogonal.

4 Quadratic Forms and Positive Definiteness

For a square symmetric (or Hermitian symmetric) matrix R and a compatible vector x, a quadratic form is the scalar defined by $x^H R x$. The matrix R is *non-negative definite* (nnd) if for any x

$$x^H R x \geq 0.$$

R is *positive definite* (pd) if $x^H R x > 0$. Examples of nnd (pd) matrices:

- $R = B^H B$ for arbitrary matrix B.
- R symmetric with only non-negative (positive) eigenvalues.

5 Singular value Decomposition of a Matrix

Theorem 2 *Let $A \in \mathbb{C}^{n\times m}$ arbitrary matrix. Then, there exist $m \times m$ and $n \times n$ unitary matrices U and V, and $\lambda_1, ...\lambda_p$, $p = min\{n, m\}$ positive constant such that*

$$A = U\Lambda V^H$$

where Λ is the diagonal matrix with its diagonal elements equal to $\lambda_1, ...\lambda_p$.

- Columns of the matrix U and V are called the left- and right-singular vectors of A, respectively. λ_i, $i = 1, ..., p$ are called the singular values of A.
- Left-singular vectors span the range space of A, whereas right-singular vectors spans the null space of A.
- Let $A \in \mathbb{C}^{n\times m}$ arbitrary matrix. Then, $A^H A$ is an $m \times m$ Hermitian symmetric, non-negative definite matrix. Similarly, AA^H is an $n\times n$ Hermitian symmetric, non-negative definite matrix. Thus, from the spectral decomposition theorem, there exists $m \times m$ and $n \times n$ unitary matrices U and V, such that $A^H A = U\Lambda_2 U^H \quad AA^H = V\Lambda_1 V^H$.

6 Vector Differentiation

Differentiation of functions of a vector variable often arise in signal processing and estimation theory. If $h = [h_1, ..., h_n]^T$ is an $n \times 1$ vector and $g(h)$ is a scalar function then the gradient of $g(h)$, denoted $\nabla g(h)$ or $\nabla_h g(h)$ when necessary for conciseness, is defined as the (column) vector of partials

$$\nabla g = [\frac{\partial g}{\partial h_1}, ..., \frac{\partial g}{\partial h_n}]^T.$$

- If c is a constant, $\nabla_h c = 0$.
- If $x = [x^1, ..., x^n]^T$, $\nabla_h(h^T x) = \nabla_h(x^T h) = x$.

- If B is an $n \times n$ matrix, $\nabla_h (h-x)^T B(h-x) = 2B(h-x)$.

- For a vector valued function $g(h) = [g_1(h), ..., g_m(h)]^T$, the gradient of $g(h)$ is an $m \times n$ matrix.

- In particular, for a scalar function $g(h)$, the two applications of the gradient $\nabla(\nabla(g))^T$ gives the $n \times n$ *Hessian* matrix of g, denoted as $\nabla^2 g$. This yields useful and natural identities such as: $\nabla^2 (h-x)^T B(h-x) = 2B$. For a more detailed discussion of vector differentiation, see Kay.

7 Signal Subspace Model

- Let $x_1, x_2, ..., x_p \in \mathbb{R}^n$ be linearly independent ($p \le N$), and consider the $N \times p$ matrix

$$X = [x_1, x_2, ..., x_p].$$

  Let $\mathcal{A}$ denotes the column space of X, then the dimension of $\mathcal{A}$ is p. Let $\mathcal{B}$ denote the space orthogonal to $\mathcal{A}$. Then

$$\mathbb{R}^n = \mathcal{A} \oplus \mathcal{B}.$$

- In the subspace model, we assume our observed signal $y \in \mathbb{R}^n$ has the form

$$y = s + \omega$$

  where $s \in \mathcal{A}$ is the signal of interest to be recovered and ω is noise. We say $\mathcal{A}$ is the *signal subspace* and $\mathcal{B}$ is the *noise subspace.*

- How do we estimate s from y knowing that s lies in the column space of X? Answer: Do an orthogonal projection onto $\mathcal{A}$.

- The orthogonal projection onto $\mathcal{A}$ is given by

$$X(X^T X)^{-1} X^T.$$

It turns out that the orthogonal projection is closely related to the least squares solution.

8 Deterministic Least Squares

Postulate an "input-output" relationship:

$$y = \sum_{i=1}^{p} x_i a_i + \omega$$

where $y = [y_1, ..., y_n]$ is the vector of observation/measurements; $x_i = [x_{i1}, x_{i2}, ..., x_{in}]^T, \quad i = 1, ..., p$ are the independent vectors; ω is the $1 \times n$ noise vector; and $a = [a_1, ..., a_p]$ is unknown $1 \times p$ coefficient vector to be estimated.

- Our objective is to estimate a from noisy measurements by means of least squares, i.e., by minimizing the squared error (SE) (SSE): The estimation criterion can be stated as

$$SE(a) = (y - Xa)^T (y - Xa).$$

- Least squared error solution of a:
  - Identify vector space containing y: $\mathbb{R}^n$. The inner product $< y, z >= y^T z$.
  - Identify the solution subspace containing Xa: $\mathcal{A} = \text{span}\{columns\ of\ X\}$ which contains vectors of the form $Xa = \sum_{k=1}^{p} a_k [x_{1k}, x_{2k}, ..., x_{nk}]^T$.
  - Differentiate $(y - Xa)^T (y - Xa)$ with respect to a and set it equal to zero.

$$0^T = (y - Xa)^T X = y^T X - a^T X^T X.$$

If X has full column rank of p, then $X^T X$ is invertable and

$$\hat{a} = (X^T X)^{-1} X^T y.$$

$(X^T X)^{-1} X^T$ is called the pseudo-inverse of X.

We next specify the projection operator form of predicted output response

$$\hat{y} = X\hat{a}$$

which using above can be represented as the orthogonal projection of y onto $\mathcal{A}$, the column space of X.

$$\hat{y} = X\hat{a} = \underbrace{X(X^TX)^{-1}X^T}_{P.} y$$

Properties of the orthogonal projection operator -

$$\Pi_X = X(X^TX)^{-1}X^T.$$

- Π_X projects vectors onto the column space of X. Define decomposition of y into components y_X in column space of X and $y_X^\perp$ orthogonal to the column space of X:

  $$y = y_X + y_X^\perp.$$

  Then for some vector $\alpha = [\alpha_1, ..., \alpha_p]^T$

  $$y_X = X\alpha, \quad X^T y_X^\perp = 0.$$

  $$\begin{aligned} \Pi_X y &= \Pi_X(y_X + y_X^\perp) \\ &= X \underbrace{(X^TX)^{-1}X^TX}_{I} \alpha + X(X^TX)^{-1}X^T y_X^\perp \end{aligned}$$

  Thus

  $$y = \Pi_X y + (I - \Pi_X)y$$

  and that $I - \Pi_X$ projects onto the subspace orthogonal to the column space of X.

- Π_X is symmetric and idempotent: $\Pi_X^T \Pi_X = \Pi_X$.

- $(I - \Pi_X)\Pi_X = 0$.