EE 4TM4: Digital Communications II

Matrix Factorization

Eigenvalue Decomposition: Every real symmetric matrix Σ can be written as $\Sigma = U\Lambda U^T$, where U is a unitary matrix (i.e., $UU^T = U^TU = I$) and Λ is a real diagonal matrix.

Note that U contains the eigenvectors of Σ while Λ contains the eigenvalues of Σ. Also note that the unitary transform preserves the L_2 norm, i.e., $\|Ua\|^2 = \|a\|^2$.

Singular Value Decomposition: Every matrix $A \in \mathbb{R}^{m\times n}$ can be written as

$$A = U \begin{pmatrix} S & 0 \\ 0 & 0 \end{pmatrix} V^T,$$

where U and V are respectively $m \times m$ and $n \times n$ unitary matrices, and $S = \text{diag}(\sigma_1, \cdots, \sigma_r)$. The positive numbers $\sigma_1 \geq \cdots \geq \sigma_r > 0$ are called the singular values of A, where r is the rank of A.

All the random objects are assumed to be of mean zero unless specified otherwise. For random vectors $X = (X_1, \cdots, X_k)^T$ and $Y = (Y_1, \cdots, Y_m)^T$, we define the correlation matrix between X and Y as $\Sigma_{X,Y} = E[XY^T]$; the covariance matrices of X and Y are defined as $\Sigma_X = E[XX^T]$ and $\Sigma_Y = E[YY^T]$, respectively. Note that the covariance matrix is symmetric.

A matrix Σ is said to be positive semidefinite if $a^T\Sigma a \geq 0$ for all a. Note that the covariance matrix is always positive semidefinite. This is because $a^T\Sigma_X a = E[(a^TX)^2] \geq 0$. When Σ is positive semidefinite, the diagonal entries of Λ is non-negative. The eigenvalue decomposition theorem implies that one can obtain a random vector with covariance matrix Σ by applying a linear transform U to a random vector with covariance matrix Λ. This also implies that every positive semidefinite matrix can be viewed as the covariance matrix of a random vector. Note that the determinant of a positive semidefinite matrix is non-negative.

Linear MMSE: For two jointly distributed random variables X and Y, find a to minimize $E[(X - aY)^2]$. By taking derivative with respect to a, it can be shown that $a = E[XY](E[Y^2])^{-1}$. A geometric interpretation: $E[(X - aY)Y] = 0$ implies $a = E[XY](E[Y^2])^{-1}$.

More generally, for two random vectors X and Y, we would like to find the optimal linear estimate from Y to X. Let $E[(X - AY)Y^T] = 0$. We have $A = \Sigma_{X,Y}\Sigma_Y^{-1}$. Note that for any B, we have $E[(X - AY)(X - AY)^T] \preceq E[(X - BY)(X - BY)^T]$, where $M_1 \preceq M_2$ means $M_2 - M_1$ is positive semidefinite. Now we proceed to prove this,

$$\begin{aligned}E[(X-BY)(X-BY)^T] &= E[(X-AY+AY-BY)(X-AY+AY-BY)^T] \\ &= E[(X-AY)(X-AY)^T] + E[(X-AY)(Y-BY)^T] \\ &\quad + E[(AY-BY)(X-AY)^T] + E[(AY-BY)(AY-BY)^T] \\ &= E[(X-AY)(X-AY)^T] + E[(AY-BY)(AY-BY)^T] \\ &\succeq E[(X-AY)(X-AY)^T].\end{aligned}$$

Note that $X - AY$ is uncorrelated with Y. In particular, if X and Y are jointly Gaussian, then $X - AY$ is independent of Y; moreover, $X - AY - Y$ form a Markov chain, i.e., X and Y are independent conditioned on AY (we say AY is a sufficient statistic of Y for estimating X).

Gram-Schmidt orthogonalization:

$$\begin{aligned}&I_1 = X_1, \\ &I_2 = X_2 - a_{2,1}I_1, \\ &I_3 = X_3 - a_{3,1}I_1 - a_{3,2}I_2, \\ &\vdots \\ &I_m = X_m - a_{m,1}I_1 - \cdots - a_{m,m-1}I_{m-1},\end{aligned}$$

where $I_1, \cdots, I_m$ are uncorrelated (they are sometimes referred to as innovation process). Therefore, we have

$$\begin{aligned}&X_1 = I_1, \\ &X_2 = a_{2,1}I_1 + I_2, \\ &X_3 = a_{3,1}I_1 + a_{3,2}I_2 + I_3, \\ &\vdots \\ &X_m = a_{m_1}I_1 + \cdots + a_{m_1,m-1}I_{m-1} + I_m.\end{aligned}$$

This yields the LDL^T decomposition of the covariance matrix of $X = (X_1, \cdots, X_m)$. Note that the diagonal entries of D is non-negative, and the diagonal entries of L are all equal to 1. One can further write the covariance matrix of X as the product of a lower triangular matrix and its transpose by grouping L and $D^{\frac{1}{2}}$.

Multivariate Normal Distribution: The density of a zero-mean Gaussian random vector $X = (X_1, \cdots, X_k)^T$ with covariance matrix Σ is given by

$$f_X(x_1, \cdots, x_k) = \frac{1}{(2\pi)^{k/2}|\Sigma|^{1/2}} \exp\left(-\frac{1}{2}x^T\Sigma^{-1}x\right).$$

Equivalently, we say X is a zero-mean Gaussian random vector with covariance matrix Σ if its moment generating function is given by $\exp(\frac{1}{2}t^T\Sigma t)$. It is easy to show that two jointly Gaussian random vectors are uncorrelated if and only if they are independent. Note that if X is a Gaussian random vector, then AX is also a Gaussian random vector. This is because the moment generating function of AX is

$$\mathbb{E}[\exp(t^T AX)] = \mathbb{E}[\exp((A^T t)^T X)] = \exp(\frac{1}{2}(A^T t)^T \Sigma A^T t) = \exp(\frac{1}{2} t^T (A\Sigma A^T) t),$$

which is exactly the moment generating function of a zero-mean Gaussian random vector with covariance matrix $A\Sigma A^T$.