

Some Notes

I. INFORMATION THEORY

A. Entropy

- For a discrete random variable X with alphabet \mathcal{X} and distributed according to the probability mass function $p(x)$, the entropy is defined as

$$\begin{aligned} H(X) &= \sum_{x \in \mathcal{X}} p(x) \cdot \log_2 \frac{1}{p(x)} \\ &= - \sum_{x \in \mathcal{X}} p(x) \cdot \log_2 p(x) \\ &= \mathcal{E} \left\{ \log_2 \frac{1}{p(x)} \right\}. \end{aligned} \tag{1}$$

- The entropy of a random variable is a measure of the uncertainty of the random variable. It is a measure of the amount of information required on the average to describe the random variable.
- With the base 2 logarithm, entropy is measured in bits.

B. Mutual Information and Channel Capacity

- $\mathcal{I}(X; Y)$ denotes the mutual information between X and Y and it is a measure of the amount of information that one random variable contains about another random variable.
- Representing the input and output of a memoryless wireless channel with the random variables X and Y respectively, the channel capacity is defined as

$$C = \max_{p(x)} \mathcal{I}(X; Y). \tag{2}$$

- According to definition in (2), the mutual information is maximized with respect to all possible transmitter statistical distributions $p(x)$.
- The mutual information between X and Y can also be written as

$$\begin{aligned} \mathcal{I}(X; Y) &= H(Y) - H(Y|X) \\ &= H(X) - H(X|Y). \end{aligned} \tag{3}$$

- From (3) above, it can be seen that mutual information can be described as the reduction in the uncertainty of one random variable due to knowledge of the other random variable.
- As an immediate consequence of (3), we have

1) *Capacity with a Transmit Power Constraint:*

- With an average transmit power constraint P_T , the channel capacity is defined as

$$C = \max_{p(x): P \leq P_T} \mathcal{I}(X; Y). \quad (4)$$

- If each symbol per channel use at the transmitter is denoted by x , the average power constraint can be expressed as $P = \mathcal{E} \{|x|^2\} \leq P_T$.
- Here the capacity of the channel is defined as the maximum of the mutual information between the input random variable X and the output random variable Y over all statistical distribution on the input that satisfy the power constraint.

2) *Capacity of a SISO Channel:*

- For a SISO flat fading wireless channel model, the input/output relations can be modelled by

$$y = hx + n \quad (5)$$

where y represent a single realization of the random variable Y (per channel use), h represent the complex channel between the transmitter and the receiver, x represents the transmitted complex symbol and n represents complex additive white gaussian noise (AWGN).

• **Assumptions:**

- 1) Perfect channel knowledge at the receiver.
- 2) X is independent of N , i.e., $p(N|X) = p(N)$ and $p(X|N) = p(X)$.
- 3) $N \sim \mathcal{N}(0, \sigma_n^2)$, i.e., $\mathcal{E} \{N\} = 0$ and $\mathcal{E} \{N^2\} = \sigma_n^2$

- **Mutual Information:** With $d_h(\cdot)$ denoting the differential entropy (entropy of a continuous random variable), the mutual information may be expressed as

$$\mathcal{I}(X; Y) = d_h(Y) - d_h(Y|X) \quad (6)$$

$$= d_h(Y) - d_h(hX + N|X) \quad (7)$$

$$= d_h(Y) - d_h(N|X) \quad (8)$$

$$= d_h(Y) - d_h(N). \quad (9)$$

- (8) follows from the fact that since h is assumed perfectly known by the receiver, there is no uncertainty in hX conditioned on X . (9) follows from the fact that N is assumed independent of X .

- **Noise differential entropy:** Since N is assumed to be complex gaussian random variable, the noise PDF is given by

$$f_N(n) = \frac{1}{\pi\sigma_n^2} e^{-\frac{n^2}{\sigma_n^2}} \quad (10)$$

and the noise differential entropy $d_h(N)$ is given by

$$\begin{aligned} d_h(N) &= - \int f_N(n) \log_2 f_N(n) dn \\ &= \log_2(\pi e \sigma_n^2) \end{aligned} \quad (11)$$

- Since $d_h(N)$ is given, the mutual information $\mathcal{I}(X; Y) = d_h(Y) - d_h(N)$ is maximized by maximizing $d_h(Y)$.
- Since normal distribution maximizes the entropy over all distributions with the same covariance, $\mathcal{I}(X; Y)$ is maximized when Y is assumed gaussian, i.e. $d_h(Y) = \log_2(\pi e \sigma_y^2)$, where $\mathcal{E}\{Y^2\} = \sigma_y^2$.
- Assuming the optimal gaussian distribution for X , the received average signal power σ_y^2 may be expressed as

$$\begin{aligned}\mathcal{E}\{Y^2\} &= \mathcal{E}\{(hX + N)(h^*X^* + N^*)\} \\ &= \sigma_x^2|h|^2 + \sigma_n^2,\end{aligned}\quad (12)$$

where $\mathcal{E}\{X^2\} = \sigma_x^2$ and $\mathcal{E}\{XN^*\} = \mathcal{E}\{NX^*\} = 0$.

- **SISO fading channel capacity:**

$$\begin{aligned}C &= d_h(Y) - d_h(N) \\ &= \log_2(\pi e(\sigma_x^2|h|^2 + \sigma_n^2)) - \log_2(\pi e\sigma_n^2) \\ &= \log_2\left(1 + \frac{P_T}{\sigma_n^2}|h|^2\right),\end{aligned}\quad (13)$$

where it is assumed that $\sigma_x^2 = P_T$.

- Denoting the total received signal-to-noise ratio (SNR) $\gamma_t = \frac{P_T}{\sigma_n^2}|h|^2$, the SISO fading channel capacity is given by $C = \log_2(1 + \gamma_t)$. Note that since γ_t is a random variable, the capacity also becomes a random variable.

C. Mutual Information and Entropy: Formulae

- Independence bound on Entropy: Let X_1, X_2, \dots, X_n be drawn according to $p(x_1, x_2, \dots, x_n)$, then

$$H(X_1, X_2, \dots, X_n) \leq \sum_{i=1}^n H(X_i),$$

with equality iff X_i are independent.

- Conditional mutual information:

$$\begin{aligned}I(\mathbf{X}, \mathbf{Y}; \mathbf{Z}) &= H(\mathbf{X}|\mathbf{Z}) - H(\mathbf{X}|\mathbf{Y}, \mathbf{Z}) \\ &= \mathcal{E}\left\{\log\left(\frac{p(x, y|z)}{p(x|z)p(y|z)}\right)\right\}\end{aligned}$$

- $I(\mathbf{X}, \mathbf{Y}; \mathbf{Z}) = I(\mathbf{Y}; \mathbf{Z}|\mathbf{X}) + I(\mathbf{X}; \mathbf{Z})$

- By the chain rule for mutual information

$$I(\mathbf{X}^T; \mathbf{Y}^T) = \sum_{t=1}^T I(\mathbf{X}^T; Y_t|\mathbf{Y}^{t-1})$$

- Directed mutual information is defined as

$$I(\mathbf{X}^T \rightarrow \mathbf{Y}^T) \triangleq \sum_{t=1}^T I(\mathbf{X}^t; Y_t | \mathbf{Y}^{t-1})$$

- $I(\mathbf{X}^T \rightarrow \mathbf{Y}^T) \neq I(\mathbf{Y}^T \rightarrow \mathbf{X}^T)$

II. TRIGONOMETRIC IDENTITIES

Reciprocal Identities:

$$\sin \theta = \frac{1}{\csc \theta} \quad \cos \theta = \frac{1}{\sec \theta} \quad \tan \theta = \frac{1}{\cot \theta}$$

Pythagorean Identities:

$$\begin{aligned} \sin^2 \theta + \cos^2 \theta &= 1 \\ \tan^2 \theta + 1 &= \sec^2 \theta \\ 1 + \cot^2 \theta &= \csc^2 \theta \end{aligned}$$

(14)

Quotient Identity:

$$\tan \theta = \frac{\sin \theta}{\cos \theta}$$

(15)

Co-Function Identities:

$$\begin{aligned} \sin\left(\frac{\pi}{2} - \theta\right) &= \cos \theta & \cos\left(\frac{\pi}{2} - \theta\right) &= \sin \theta & \tan\left(\frac{\pi}{2} - \theta\right) &= \cot \theta \\ \cos\left(\theta - \frac{\pi}{2}\right) &= \sin \theta & \sin\left(\theta - \frac{\pi}{2}\right) &= -\cos \theta \end{aligned}$$

Even-Odd Identities:

$$\sin(-\theta) = -\sin(\theta) \quad \cos(-\theta) = \cos(\theta) \quad \tan(-\theta) = -\tan(\theta)$$

Sum-Difference Formulas:

$$\begin{aligned} \sin(u \pm v) &= \sin u \cos v \pm \cos u \sin v \\ \cos(u \pm v) &= \cos u \cos v \mp \sin u \sin v \\ \tan(u \pm v) &= \frac{\tan u \pm \tan v}{1 \mp \tan u \tan v} \end{aligned}$$

Double Angle Formulas

III. PROBABILITY AND STATISTICS

A. Continuous Distributions

Uniform Distribution:

Probability density function:

$$f(y) = \frac{1}{\theta_2 - \theta_1}; \quad \theta_1 \leq y \leq \theta_2 \quad (16)$$

Mean:

$$\frac{\theta_1 + \theta_2}{2}$$

Variance:

$$\frac{(\theta_1 - \theta_2)^2}{12}$$

Moment Generating function:

$$\frac{e^{t\theta_2} - e^{t\theta_1}}{t(\theta_2 - \theta_1)}$$

Normal Distribution:

Probability density function:

$$f(y) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2\sigma^2}(y - \mu)^2\right]; \quad -\infty < y < +\infty \quad (17)$$

Mean:

$$\mu$$

Variance:

$$\sigma^2$$

Moment Generating function:

$$\exp\left(\mu t + \frac{t^2\sigma^2}{2}\right)$$

Exponential Distribution:

Probability density function:

$$f(y) = \frac{1}{\beta} e^{-y/\beta}; \quad 0 < y < \infty \quad (18)$$

Mean:

$$\beta$$

Variance:

$$\beta^2$$

Moment Generating function:

$$(1 - \beta t)^{-1}$$

Gamma Distribution:

Probability density function:

$$f(y) = \left[\frac{1}{\Gamma(\alpha)\beta^\alpha} \right] y^{\alpha-1} e^{-y/\beta}; \quad 0 < y < \infty \quad (19)$$

Mean:

$$\alpha\beta$$

Variance:

$$\alpha\beta^2$$

Moment Generating function:

$$(1 - \beta t)^{-\alpha}$$

When $\alpha = 1$, the resulting distribution is an exponential distribution with parameter β

Chi-square Distribution:

Probability density function:

$$f(y) = \frac{(y)^{\nu/2-1} e^{-y/2}}{2^{\nu/2} \Gamma(\nu/2)}; \quad y^2 > 0 \quad (20)$$

Mean:

$$\nu$$

Variance:

$$2\nu$$

Moment Generating function:

$$(1 - 2t)^{-\nu/2}$$

A random variable Y is said to have a *chi-square distribution with ν degree of freedom* if and only if Y is a gamma-distributed random variable with parameters $\alpha = \nu/2$ and $\beta = 2$, where ν is a positive integer.

Beta Distribution:

Probability density function:

$$f(y) = \left[\frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} \right] y^{\alpha-1} (1 - y)^{\beta-1}; \quad 0 < y < 1 \quad (21)$$

Mean:

$$\frac{\alpha}{\alpha + \beta}$$

Variance:

$$\frac{\alpha\beta}{(\alpha + \beta)^2(\alpha + \beta + 1)}$$

Moment Generating function: does not exist in close form.

Weibull Distribution:

Probability density function:

$$f(y) = \frac{m y^{m-1}}{\alpha} e^{-y^m/\alpha}; \quad 0 \leq y < \infty; \quad m, \alpha > 0 \quad (22)$$

Mean:

....

Variance:

....

Moment Generating function:

....

Rayleigh Distribution:

Probability density function:

$$f(y) = \frac{2y}{\Omega} \exp\left[-\frac{y^2}{\Omega}\right]; \quad y \geq 0 \quad (23)$$

$$E[y^2] = \Omega \quad (24)$$

Mean:

....

Variance:

....

Moment Generating function:

....

When $U = Y^2$, U has an exponential distribution with parameter $\beta = \Omega$.

Nakagami-m Distribution:

Probability density function:

$$f(y) = \frac{2m^m y^{2m-1}}{\Gamma(m)\Omega^m} \exp\left[-\frac{my^2}{\Omega}\right]; \quad y \geq 0; \quad m \geq \frac{1}{2} \quad (25)$$

$$E[y^2] = \Omega \quad (26)$$

Mean:

....

Variance:

....

Moment Generating function:

....

When $m = 1$, the Nakagami distribution becomes the Rayleigh distribution, when $m = 1/2$ it becomes a one-sided Gaussian distribution, and when $m \rightarrow \infty$ the distribution becomes an impulse. When $U = Y^2$, U has a Gamma distribution with pdf given by

$$f(u) = \left(\frac{m}{\Omega}\right)^m \frac{u^{m-1}}{\Gamma(m)} \exp\left[-\frac{mu}{\Omega}\right]$$

Multivariate Complex-Gaussian Distribution:

Probability density function [1]:

$$f(\mathbf{y}) = \frac{1}{\pi^s \det R} \exp [-(\mathbf{y} - \boldsymbol{\mu}_y)^\dagger \mathbf{R}^{-1}(\mathbf{y} - \boldsymbol{\mu}_y)] \quad (27)$$

where s is the length of \mathbf{y}

Mean:

$$\boldsymbol{\mu}_y$$

Covariance Matrix:

$$\mathbf{R}$$

Moment Generating function:

$$\dots$$

B. The law of Total Probability and Bayes's Rule

Conditional Probability: Given a probability space (Ω, F, P) and two events $A, B \in F$ with $P(B) > 0$, the conditional probability A given B is defined by

$$P(A|B) = \frac{P(A, B)}{P(B)}.$$

For two independent events A and B , $P(A|B) = P(A)$, $P(B|A) = P(B)$ and $P(A, B) = P(A)P(B)$.

Identities:

$$\begin{aligned} P(A, B|C) &= P(A|B, C)P(B|C), \\ P(A, B, C|D) &= P(A|B, C, D)P(B|C, D)P(C|D) \end{aligned}$$

Total Probability: Assume that $S = B_1 \cup B_2 \cup \dots \cup B_k$ where $P(B_i) > 0$, for $i = 1, 2, 3, \dots, k$ and $B_i \cup B_j = \emptyset$ for $i \neq j$. Then for any event A

$$P(A) = \sum_{i=1}^k P(B_i)P(A/B_i) \quad (28)$$

Bayes's Rule: Assume that $S = B_1 \cup B_2 \cup \dots \cup B_k$ where $P(B_i) > 0$, for $i = 1, 2, 3, \dots, k$ and $B_i \cup B_j = \emptyset$ for $i \neq j$. Then

$$P(B_j/A) = \frac{P(B_j)P(A/B_j)}{\sum_{i=1}^k P(B_i)P(A/B_i)} \quad (29)$$

C. The Multiplication rule

Suppose that (A_1, A_2, \dots, A_n) is a sequence of events in a random experiment whose intersection has positive probability. Then the multiplication rule of probability is given by

$$\begin{aligned} P(A_1, A_2, \dots, A_n) &= P(A_1)P(A_2/A_1)P(A_3/A_1, A_2) \cdots \\ &P(A_n/A_1 A_2 \cdots A_{n-1}) \end{aligned} \quad (30)$$

D. Correlation

Let X and Y be two random variables defined on the same sample space Ω . The *correlation* between X and Y is defined as $\mathbf{E}(XY)$. The *covariance* of X and Y is defined as:

$$COV(X, Y) = \mathbf{E}[(X - m_x)(Y - m_y)] \quad (31)$$

$$= \mathbf{E}(XY) - m_x m_y \quad (32)$$

The normalized version of the correlation, called the *correlation coefficient*, is denoted by $\rho_{x,y}$ and is defined by:

$$\rho_{x,y} = \frac{COV(X, Y)}{\sigma_x \sigma_y} \quad (33)$$

E. Transformation of Random variables

Transformations in one dimension: Consider a transformation $y = g(x)$ that maps a region R of x -values into a region S of y -values. Then the transformation is

$$f_Y(y) = f_X(g^{-1}(y)) \left| \frac{dx}{dy} \right|. \quad (34)$$

Suppose that x_1, x_2, \dots, x_n are the roots of y , then the transformation is

$$f_Y(y) = \sum_n \frac{f_X(x_n)}{|g'(x_n)|}. \quad (35)$$

Transformations in two dimension:

IV. BESSEL'S EQUATION

A. Bessel Functions of the First Kind

Bessel's Differential Equation:

$$x^2 y'' + xy' + (x^2 - v^2)y = 0 \quad (36)$$

The solution of (36) is known as the Bessel function of the first kind of order v and the solution:

$$J_v(x) = x^v \sum_{m=0}^{\infty} \frac{(-1)^m x^{2m}}{2^{2m+v} m! \Gamma(v + m + 1)} \quad (37)$$

$$J_{-v}(x) = x^{-v} \sum_{m=0}^{\infty} \frac{(-1)^m x^{2m}}{2^{2m-v} m! \Gamma(m - v + 1)} \quad (38)$$

If v is not an integer, a general solution of Bessel's equation (36) for all $x \neq 0$ is

$$y(x) = a_1 J_v(x) + a_2 J_{-v}(x) \quad (39)$$

**Integer values of v frequently denoted by n .
Thus, for $n \geq 0$**

$$J_n(x) = x^n \sum_{m=0}^{\infty} \frac{(-1)^m x^{2m}}{2^{2m+n} m! (n+m)!} \quad (40)$$

For integer $v = n$ the Bessel functions $J_n(x)$ and $J_{-n}(x)$ are linearly dependent because

$$J_{-n}(x) = (-1)^n J_n(x), \quad \text{for } n = 1, 2, 3, \dots \quad (41)$$

B. Bessel Functions of the Second Kind

For integer $v = n$ the Bessel function $J_n(x)$ and $J_{-n}(x)$ are linearly dependent, so that they do not form a basis of solutions. This poses the problem of obtaining a second linearly independent solution when v is an integer n . The standard second solution $Y_v(x)$ defined for all v by the formula

$$Y_v(x) = \frac{1}{\sin v\pi} [J_v(x) \cos v\pi - J_{-v}(x)] \quad (42)$$

$$Y_n(x) = \lim_{v \rightarrow n} Y_v(x) \quad (43)$$

The function (42) is known as the Bessel function of the second kind of order v or Neumann's function of order v . The series development of $Y_n(x)$ can be obtained by substituting J_v and J_{-v} into (42) and let v approaches n .

$$Y_n(x) = \frac{2}{\pi} J_n(x) \left(\ln \frac{x}{2} + \gamma \right) + \frac{x^n}{\pi} \sum_{m=0}^{\infty} \frac{(-1)^{m-1} (h_m + h_{m+n})}{2^{2m+n} m! (m+n)!} x^{2m} - \frac{x^{-n}}{\pi} \sum_{m=0}^{n-1} \frac{(n-m-1)!}{2^{2m-n} m!} x^{2m} \quad (44)$$

where $x > 0$, $n = 0, 1, 2, \dots$ and

$$h_0 = 0 \quad (45)$$

$$h_s = 1 + \frac{1}{2} + \frac{1}{3} + \dots + \frac{1}{s}, \quad \text{for } s = 1, 2, 3, \dots \quad (46)$$

where the number $\gamma = 0.57721566490\dots$ is the so called Euler Constant.

V. FOURIER SERIES

A. Euler Formulas

Assume that $f(x)$ is a periodic function of period T which can be represented by a trigonometric series

$$f(t) = a_0 + \sum_{n=1}^{\infty} \left(a_n \cos \frac{2n\pi}{T} t + b_n \sin \frac{2n\pi}{T} t \right) \quad (47)$$

Fourier coefficients of the periodic function $f(t)$ of period T are given by the Euler formulas

$$a_0 = \frac{1}{T} \int_{-T/2}^{T/2} f(t) dt \quad (48)$$

$$a_n = \frac{2}{T} \int_{-T/2}^{T/2} f(t) \cos \frac{2n\pi t}{T} dt \quad (49)$$

$$b_n = \frac{2}{T} \int_{-T/2}^{T/2} f(t) \sin \frac{2n\pi t}{T} dt \quad (50)$$

$n = 1, 2, 3, 4, \dots$ The interval of integration in (50) may be replaced by any interval of length T , for example, by the interval $0 \leq t \leq T$

B. *Fourier series of even and odd functions*

The Fourier series of an even function $f(t)$ of period T is a "Fourier cosine series".

$$f(t) = a_0 + \sum_{n=1}^{\infty} a_n \cos \frac{2n\pi}{T} t \quad (51)$$

with coefficients for $n = 1, 2, 3, 4, \dots$

$$a_0 = \frac{2}{T} \int_0^{T/2} f(t) dt \quad (52)$$

$$a_n = \frac{4}{T} \int_0^{T/2} f(t) \cos \frac{2n\pi t}{T} dt \quad (53)$$

The Fourier series of an odd function $f(t)$ of period T is a "Fourier sine series".

$$f(t) = \sum_{n=1}^{\infty} b_n \sin \frac{2n\pi}{T} t \quad (54)$$

with coefficient for $n = 1, 2, 3, 4, \dots$

$$b_n = \frac{4}{T} \int_0^{T/2} f(t) \sin \frac{2n\pi t}{T} dt \quad (55)$$

VI. SOME FORMULAS

A. *Power series*

Taylor Series: Let $f(z)$ be analytic in a domain D and let $z = a$ be any point in D . Then there exists precisely one power series with center at a which represents $f(z)$. This series is of the form:

$$f(z) = \sum_{n=0}^{\infty} \frac{f^{(n)}(a)}{n!} (z - a)^n, \quad n = 0, 1, 2, 3, \dots \quad (56)$$

The particular case where $a = 0$ is called the *Maclaurin series* of $f(z)$.

B. Examples of power series

$$\frac{1}{1-x} = \sum_{m=0}^{\infty} x^m = 1 + x + x^2 + \dots \quad (57)$$

$$e^x = \sum_{m=0}^{\infty} \frac{x^m}{m!} = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \dots \quad (58)$$

$$e^{x^2} = \sum_{m=0}^{\infty} \frac{x^{2m}}{m!} = 1 + x^2 + \frac{x^4}{2!} + \frac{x^6}{3!} + \dots \quad (59)$$

$$\cos x = \sum_{m=0}^{\infty} \frac{(-1)^m x^{2m}}{(2m)!} = 1 - \frac{x^2}{2!} + \frac{x^4}{4!} - + \dots \quad (60)$$

$$\sin x = \sum_{m=0}^{\infty} \frac{(-1)^m x^{2m+1}}{(2m+1)!} = x - \frac{x^3}{3!} + \frac{x^5}{5!} - + \dots \quad (61)$$

$$(62)$$

C. Logarithm, Natural Logarithm and dB relationships

$$\ln(xy) = \ln(x) + \ln(y) \quad (63)$$

$$\ln(x/y) = \ln(x) - \ln(y) \quad (64)$$

$$\ln(x^a) = a \ln(x) \quad (65)$$

1 Watt = 1000 milli-Watts

$$dBm = 10 \log_{10}(W) + 30 \quad (66)$$

$$W = 10^{\frac{(dBm-30)}{10}} \quad (67)$$

$$mW = 10^{\frac{dBm}{10}} \quad (68)$$

VII. INTEGRALS

- **For positive real matrix Λ ($\Lambda = \Lambda^* > 0$) and real $n \times 1$ vector x**
 $\int dx e^{-x^t \Lambda x} = \pi^{n/2} (\det \Lambda)^{-1/2}$.

VIII. CONVEX OPTIMIZATION

A. The Lagrangian Problem single constraint

$$\min f(x_1, x_2, \dots, x_n)$$

$$\text{subject to: } h(x_1, x_2, \dots, x_n) = 0$$

Form Lagrangian function:

$$\mathcal{L}(x_1, x_2, \dots, x_n, \mu) = f(x_1, x_2, \dots, x_n) - \mu h(x_1, x_2, \dots, x_n),$$

where μ is the Lagrangian multiplier.

Necessary conditions:

$$\frac{\partial \mathcal{L}(x_1, x_2, \dots, x_n, \mu)}{\partial x_i} = 0, \quad i = 1, 2, \dots, n \quad (69)$$

$$\frac{\partial \mathcal{L}(x_1, x_2, \dots, x_n, \mu)}{\partial \mu} = 0 \quad (70)$$

solve (69) and (70) for $\mu, x_1, x_2, \dots, x_n$.

IX. SOME DEFINITIONS

A

Analytic function: A function $f(x)$ is said to be analytic at a point $x = a$ if it can be represented by a power series in powers of $(x - a)$ with radius of convergence $R > 0$.

C

Convolutional Code - Tree, Trellis and State Diagrams: A rate k/n , constraint length K (K shift registers), convolutional code is characterized by 2^k branches emanating from each node of the tree diagram. The trellis and state diagram each have $2^{k(K-1)}$ possible states, where k is the length of each register in bits. There are 2^k branches entering each state and 2^k branches leaving each state (this is true after the initial transient). [?, page 475]

Commutation Matrix: Commutation matrix $K(m, n)$ is the unique $mn \times mn$ matrix such that $K(m, n) \text{vec}(A) = \text{vec}(A^H)$ for all $m \times n$ matrices A .

D

Diffuse Field: An acoustical environment with a high number of reflections – that is, highly reflective as in a reverberant chamber. At any given point in the diffuse field, sound will arrive from all angles in a uniform manner.

E

Euclidean Norm: $\|\cdot\|_2$ denotes the Euclidean Norm of vector v ,

$$\|v\|_2 = (v^\dagger v)^{1/2},$$

where v is a column vector.

F

Free Probability Theory: A tool to deal with infinite random matrices.

Frobenius Norm of a Matrix: The notation $\|\cdot\|$ denotes the Frobenius norm of a matrix:

$$\|X_{P \times Q}\|^2 = \sum_{i=1}^P \sum_{j=1}^Q |a_{ij}|^2$$

$$\begin{aligned}
\|X\|^2 &= \mathbf{trace}\{XX^\dagger\}, \\
\|XY\|^2 &= \mathbf{trace}\{XY Y^\dagger X^\dagger\}, \\
&= \sum_{\ell=1}^P x_\ell Y Y^\dagger x_\ell^\dagger, \\
&= x^\dagger \widehat{Y} \widehat{Y}^\dagger x,
\end{aligned}$$

where X is $P \times Q$, Y is $Q \times M$, x_ℓ is the ℓ -th row of X , $x = \text{vec}(X)$ is a column vector, $\widehat{Y} = Y \otimes I_P$ and \otimes denotes the Matrix Kronecker product.

G

Gamma Function $\Gamma(v + 1)$:

$$\Gamma(\alpha) = \int_0^\infty e^{-t} t^{\alpha-1} dt \quad (71)$$

$\Gamma(\alpha + 1) = \alpha \Gamma(\alpha)$ and $\Gamma(k + 1) = k!$

For integer $v = n$

$$\Gamma\left(n + \frac{1}{2}\right) = \frac{\sqrt{\pi}}{2^n} (2n - 1)!! \quad (72)$$

where $(2n - 1)!! = 1.3.5 \dots (2n - 1)$

H

Harmonic function and Harmonic oscillation: Simple harmonic motion or *harmonic oscillation* refers to oscillations with a sinusoidal waveform. Such functions satisfy the differential equation $\frac{d^2x}{dt^2} + \omega^2 x = 0$, which has solution $x(t) = A \cos(\omega t + \phi_1) + B \sin(\omega t + \phi_2)$. The word *harmonic analysis* is used to describe Fourier analysis, which breaks an arbitrary function into a superposition of sinusoids. In complex analysis, a *harmonic function* refers to a real-valued function $f(x, y)$ which satisfies Laplace's equation $\nabla^2 f(x, y) = 0$ where ∇^2 is the Laplacian. Although this definition is similar to the *harmonic oscillation*, it omits the second term in the differential equation. The Helmholtz differential equation is obtained if it is added back in, $\nabla^2 f(x, y) + k^2 f(x, y) = 0$.

Homogeneous function of order n: A function which satisfies $f(tx, ty) = t^n f(x, y)$ for a fixed n .

K

Kronecker product or Matrix direct product or tensor product, denoted by \otimes . The Kronecker product of the 2x2 matrix A and the 3x3 matrix B is given by the following matrix.

$$A \otimes B = \begin{pmatrix} a_{11}B & a_{12}B \\ a_{21}B & a_{22}B \end{pmatrix} \quad (73)$$

$$\mathbf{A} \otimes \mathbf{B} = \begin{pmatrix} a_{11}b_{11} & a_{11}b_{12} & a_{12}b_{11} & a_{12}b_{12} \\ a_{11}b_{21} & a_{11}b_{22} & a_{12}b_{21} & a_{12}b_{22} \\ a_{11}b_{31} & a_{11}b_{32} & a_{12}b_{31} & a_{12}b_{32} \\ a_{21}b_{11} & a_{21}b_{12} & a_{22}b_{11} & a_{22}b_{12} \\ a_{21}b_{21} & a_{21}b_{22} & a_{22}b_{21} & a_{22}b_{22} \\ a_{21}b_{31} & a_{21}b_{32} & a_{22}b_{31} & a_{22}b_{32} \end{pmatrix}$$

Some properties of the Kronecker product:

- 1) $\mathbf{A} \otimes (\mathbf{B} \otimes \mathbf{C}) = (\mathbf{A} \otimes \mathbf{B}) \otimes \mathbf{C}$ **associativity.**
- 2) $\mathbf{A} \otimes (\mathbf{B} + \mathbf{C}) = (\mathbf{A} \otimes \mathbf{B}) + (\mathbf{A} \otimes \mathbf{C})$
 $(\mathbf{A} + \mathbf{B}) \otimes \mathbf{C} = (\mathbf{A} \otimes \mathbf{C}) + (\mathbf{B} \otimes \mathbf{C})$, **distributive**
- 3) **For scalar a , $a \otimes \mathbf{A} = \mathbf{A} \otimes a = a\mathbf{A}$**
- 4) **For scalar a and b , $a\mathbf{A} \otimes b\mathbf{B} = ab\mathbf{A} \otimes \mathbf{B}$**
- 5) **For conforming matrices, $(\mathbf{A} \otimes \mathbf{B})(\mathbf{C} \otimes \mathbf{D}) = \mathbf{AC} \otimes \mathbf{BD}$**
- 6) $(\mathbf{A} \otimes \mathbf{B})^T = \mathbf{A}^T \otimes \mathbf{B}^T$,
 $(\mathbf{A} \otimes \mathbf{B})^H = \mathbf{A}^H \otimes \mathbf{B}^H$
- 7) **For square non-singular matrices \mathbf{A} and \mathbf{B} ,**
 $(\mathbf{A} \otimes \mathbf{B})^{-1} = \mathbf{A}^{-1} \otimes \mathbf{B}^{-1}$
- 8) **For $m \times m$ matrix \mathbf{A} and $n \times n$ matrix \mathbf{B} :**
 $|\mathbf{A} \otimes \mathbf{B}| = |\mathbf{A}|^n |\mathbf{B}|^m$
- 9) $\text{Tr}(\mathbf{A} \otimes \mathbf{B}) = \text{Tr}(\mathbf{A})\text{Tr}(\mathbf{B})$
- 10) $\text{rank}(\mathbf{A} \otimes \mathbf{B}) = \text{rank}(\mathbf{A})\text{rank}(\mathbf{B})$

L

Linear Time Invariant System: A Linear Time Invariant System is one that:

- 1) **Is unaffected by time. That is, if you perform an experiment on Monday to find the systems response to a sine wave, you will get the same result if you do the experiment again on Wednesday.**
- 2) **Is Linear. Given two input signals (ax, cx) and that they produce two output signals (by, dy) , the system is linear if, and only if, the input signal $ax + cx$ produces the output signal $by + dy$.**

More formally, in a Time Invariant system:

$$\begin{aligned} \text{If } x(t) &\implies y(t), \\ \text{then } x(t - t_0) &\implies y(t - t_0). \end{aligned}$$

In a linear system

$$\begin{aligned} \text{If } ax &\implies by \text{ and} \\ \text{If } cx &\implies dy \text{ and} \\ \text{If } ax + cx &\implies by + dy, \end{aligned}$$

then the system is linear.

L'Hopital's Rules:

i **Suppose** $\lim_{x \rightarrow a} f(x) = \lim_{x \rightarrow a} g(x) = 0$. **Then** $\lim_{x \rightarrow a} \frac{f(x)}{g(x)} = \lim_{x \rightarrow a} \frac{f'(x)}{g'(x)}$.

i **Suppose** $\lim_{x \rightarrow a} f(x) = \lim_{x \rightarrow a} g(x) = \infty$. **Then** $\lim_{x \rightarrow a} \frac{f(x)}{g(x)} = \lim_{x \rightarrow a} \frac{f'(x)}{g'(x)}$.

Remark: This rule only applies for the limit of a quotient, with $\frac{0}{0}$ and $\frac{\infty}{\infty}$.

M

Matrix Determinant Identities

- $|I + AB| = |I + BA|$
- $\det A = \prod_{m=1}^M \lambda_m$, where λ_m are the m eigenvalues of $M \times M$ matrix A .
- If a is a constant and A an $n \times n$ square matrix, then $|aA| = a^n |A|$
- Determinants are distributive, $|AB| = |A||B|$
- $|A||A^{-1}| = 1$ and $|BAB^{-1}| = |A|$
- $\det(I_m + aX) = 1 + a\text{Tr}(X) + \dots + a^m \det(X)$

Matrix Inverse: If A is a square non-singular matrix, then the inverse of A is

$$A^{-1} = \frac{1}{\det A} \begin{pmatrix} A_{11} & \dots & A_{1n} \\ \vdots & \ddots & \vdots \\ A_{n1} & \dots & A_{nn} \end{pmatrix} \quad (74)$$

where A_{ij} is the cofactor of a_{ij} [that is, $(-1)^{i+j}$ times the determinant of order $n - 1$ obtained by striking from A its i th row and j th column].

Matrix Inverse Identities

$$(A + BCD)^{-1} = A^{-1} - A^{-1}B(C^{-1} + DA^{-1}B)^{-1}DA^{-1}$$

Matrix Multiplication

- If A is a $m \times n$ matrix and B is a $n \times r$ matrix, then the product C of two matrices A and B is given by

$$c_{ij} = \sum_{k=1}^n a_{ik}b_{kj} \quad (75)$$

where c_{ij} is an entry in C and C is a $m \times r$ matrix.

- If A is a $m \times p$ matrix, B is a $q \times r$ matrix and H is a $p \times q$ matrix, then the product $C = AHB$ is given by

$$c_{ij} = \sum_{x=1}^p \sum_{y=1}^q a_{ix} h_{xy} b_{yj} \quad (76)$$

where c_{ij} is an entry in C and C is a $m \times r$ matrix.

Similarly,

$$C = \sum_{x=1}^p \sum_{y=1}^q \mathbf{a}_x h_{xy} \mathbf{b}_y \quad (77)$$

where \mathbf{a}_x are the column vectors of A and \mathbf{b}_y are the row vectors of B .

- If A is a $m \times n$ matrix, then the product $C = AA^\dagger$ is given by

$$c_{ij} = \sum_{k=1}^n a_{ik} a_{jk}^* \quad (78)$$

where c_{ij} is an entry in C and C is a $m \times m$ matrix.

- If A is a $m \times n$ matrix with $A = [\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_n]$, where \mathbf{a}_i is a $m \times 1$ vector, then

$$AA^\dagger = \sum_{i=1}^n \mathbf{a}_i \mathbf{a}_i^\dagger \quad (79)$$

Matrix Principal Minors: A principal minor is a sub-matrix in which certain columns together with rows with the same numbers have been deleted from the original matrix, e.g. the first and fourth rows and first and fourth columns.

Matrix Trace Identities

- $\text{Tr}(A) = \sum_{m=1}^M \lambda_m$, where λ_m are the m eigenvalues of matrix A .
- If x is $n \times 1$ column vector, then $\text{Tr}(xx^\dagger) = x^\dagger x$.
- If U is a unitary matrix (i.e., $UU^\dagger = U^\dagger U = I$) and A is non-negative definite then $\text{Tr}(A) = \text{Tr}(U^\dagger AU)$.

P

Positive Definite Matrix: An $n \times n$ complex matrix A is called positive definite if

$$x^\dagger Ax > 0,$$

for all non-zero vectors $x \in \mathbb{R}^n$. Matrix A is positive definite if all its eigenvalues are real and positive (no zero eigenvalues). The positive definite matrix is nonsingular, therefore A^{-1} exists.

Positive Semi-Definite Matrix (non-negative definite): An $n \times n$ complex matrix A is called positive definite if

$$x^\dagger Ax \geq 0,$$

for all non-zero vectors $x \in \mathbb{R}^n$. Matrix A is positive semi-definite if all its eigenvalues are real and non negative and at least one of them equals 0. The positive semi-definite matrix is singular

Q

Qasistatic fading: Fading is constant over a long period of time.

S

Scattering: The process in which a wave or beam of particles is diffused or deflected by collisions with particles of the medium which it traverses.

Singular value decomposition: Any matrix $H \in C^{r \times t}$ can be written as

$$H = UDV^\dagger$$

where $U \in C^{r \times r}$ and $V \in C^{t \times t}$ are unitary and $D \in C^{r \times t}$ is non-negative and diagonal. The diagonal entries of D are the non-negative square roots of the eigenvalues of HH^\dagger , the columns of U are the eigenvectors of HH^\dagger and the column vectors of V are the eigenvectors of $H^\dagger H$.

Singular Value Decomposition (SVD): Definition[2, page 70] If A is a real $m \times n$ matrix, then there exist orthogonal matrices

$$U = [\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_m] \in \mathbb{R}^{m \times m}$$

$$V = [\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n] \in \mathbb{R}^{n \times n}$$

such that

$$U^T AV = \text{diag}(\sigma_1, \sigma_2, \dots, \sigma_p), \quad p = \min\{m, n\}$$

where $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_p \geq 0$. The SVD reveals a great deal about the structure of a matrix. If the SVD of A is given by the above definition, and we define r by

$$\sigma_1 \geq \dots \geq \sigma_r \geq \sigma_{r+1} = \dots = \sigma_p = 0,$$

then

$$\text{rank}(\mathbf{A}) = r,$$

$$\text{null}(\mathbf{A}) = \text{span}\{\mathbf{v}_{r+1}, \dots, \mathbf{v}_n\},$$

$$\text{rang}(\mathbf{A}) = \text{span}\{\mathbf{u}_1, \dots, \mathbf{u}_r\},$$

and we have the SVD expansion

$$\mathbf{A} = \sum_{i=1}^r \sigma_i \mathbf{u}_i \mathbf{v}_i^T.$$

Stationarity:

- **First order stationary process:** A random process is classified as first-order stationary if its first-order probability function remains equal regardless of any shift in time to its time origin. Let x_{t_1} represents a given value at time t_1 , then a first-order stationary process satisfies

$$f_x(x_{t_1}) = f_x(x_{t_1+\tau}).$$

The physical significance of this equation is that density function $f_x(x_{t_1})$ is completely independent of t_1 and thus any time shift, τ .

The most important result of this statement, and the identifying characteristic of any first-order stationary process, is the fact that the mean is a constant, independent of any time shift. That is, for a random process, X with a discrete-time signal $x[n]$,

$$\begin{aligned}\bar{X} &= \mathcal{E} \{x[n]\}, \\ &= \text{constant} \quad (\text{independent of } n).\end{aligned}$$

- **Second order and Strict-sense stationary process:** A random process is classified as second-order stationary if its second-order probability density function does not vary over any time shift applied to both values. In other words, for values x_{t_1} and x_{t_2} , then for an arbitrary time shift τ , a second-order stationary process satisfies

$$f_x(x_{t_1}, x_{t_2}) = f_x(x_{t_1+\tau}, x_{t_2+\tau}).$$

These random processes are often referred to as strict sense stationary (SSS) when all of the distribution functions of the process are unchanged regardless of the time shift applied to them. For a second-order stationary process, we need to look at the autocorrelation function to see its most important property. Since the second-order stationary process depends only on the time difference, then all of these types of processes have the following property:

$$R_{xx}(t, t + \tau) = R_{xx}(\tau).$$

- **Wide-Sense Stationary Process:** In order to be WSS a random process only needs to meet the following two requirements.

$$\begin{aligned}\bar{X} &= \mathcal{E} \{x[n]\} = \text{constant}, \\ R_{xx}(t, t + \tau) &= R_{xx}(\tau),\end{aligned}$$

Note that a second-order stationary process will always be WSS; however, the reverse will not always hold true.

Sufficient statistic: Let $\{f_\theta\}$ be a statistical model with parameter θ . Let $X = (X_1, \dots, X_n)$ be a random vector of random variables representing n observations. A statistic $T = T(X)$ of X for the parameter θ is called a sufficient statistic, or a sufficient estimator, if the conditional probability distribution of X given $T(X) = t$ is not a function of θ (equivalently, does not depend on θ).

U

Unitary Matrix: A square matrix U is a unitary matrix if

$$U^* = U^{-1}$$

where U^* denotes the adjoint matrix and U^{-1} is the inverse matrix of U . The definition of the unitary matrix guarantees that $U^*U = I$.

V

Vector Majorization: A real vector $\alpha = [\alpha_i] \in \mathbb{R}^n$ majorizes another real vector $\beta = [\beta_i] \in \mathbb{R}^n$ if and only if the sum of the k smallest entries of α is greater than or equal to the sum of the k smallest entries of β for $k = 1, 2, \dots, n-1$ and the sums of the entries of α and β are equal (i.e., $\sum_{i=1}^n \alpha_i = \sum_{i=1}^n \beta_i$). Vector Majorization is a mathematical way to capture the vague notion that the components of a vector α are ‘less spread out’ or ‘more nearly equal’ than are the components of a vector β . Majorization is the precise relationship between the eigenvalues and the diagonal entries of a Hermitian matrix. That is, for a Hermitian matrix A the vector of diagonal entries $\{a_{ii}\}$ majorizes the vector of eigenvalues $\{\lambda_{ii}^A\}$ [2, Theorem 4.3.26].

W

Wronskian determinant: of function y_1, y_2 defined by

$$\begin{pmatrix} y_1 & y_2 \\ y_1' & y_2' \end{pmatrix} = y_1 y_2' - y_2 y_1'$$

In general, for functions y_1, y_2, \dots, y_n , **Wronskian** is the determinant of matrix

$$\begin{pmatrix} y_1 & y_2 & \dots & y_n \\ y_1' & y_2' & \dots & y_n' \\ y_1'' & y_2'' & \dots & y_n'' \\ \vdots & \vdots & \ddots & \vdots \\ y_1^{n-1} & y_2^{n-1} & \dots & y_n^{n-1} \end{pmatrix}$$

REFERENCES

- [1] N. R. Goodman, “Statistical analysis based on a certain multivariate complex gaussian distribution (an introduction),” *Ann. Math. Statist.*, vol. 34, pp. 152–177, 1963.
- [2] G. H. Golub and C. F. Van Loan, *Matrix Computations*, The Johns Hopkins University Press, Baltimore and London, third edition, 1996.