1 Deterministic Signal Processing

1.1 Transforms

$$F(j\omega) = \int_{-\infty}^{\infty} f(t)e^{-j\omega t}dt$$

$$f(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} F(j\omega)e^{j\omega t}d\omega$$

$$F(e^{j\Omega}) = \sum_{k=-\infty}^{\infty} f[k]e^{-j\Omega k}$$

$$f[n] = \frac{1}{2\pi}\int_{<2\pi>} F(e^{j\Omega})e^{j\Omega k}d\Omega$$

1.2 Systems

$$(f * g)(t) = \int_{-\infty}^{\infty} f(\tau)g(t-\tau)d\tau$$

$$(f * g)[n] = \sum_{k=-\infty}^{\infty} f[k]g[n-k]$$

1.3 Parseval's Theorem

$$\sum_k v^2[k] = \frac{1}{2\pi}\int_{<2\pi>} |V(e^{j\Omega})|^2 d\Omega$$

$$\int_{-\infty}^{\infty} v^2(t)dt = \frac{1}{2\pi}\int_{-\infty}^{\infty} |V(e^{j\omega})|^2 d\omega$$

1.4 Signal Classes

1. L1: finite action; absolutely summable (or integrable); $\sum_n |x[n]| < \infty$.
2. L2: finite energy; square summable (or integrable); $\sum_n |x[n]|^2 < \infty$.
3. L3: signals of slow growth.

1.5 Energy Spectral Density

$$ESD = |X(e^{j\Omega})|^2$$

For a signal passed through an ideal passband filter h with a passband P,

$$\epsilon_p = \frac{1}{2\pi}\int_{<2\pi>} |X(e^{j\Omega})H(e^{j\Omega})|^2 d\Omega = \frac{1}{2\pi}\int_P |X(e^{j\Omega})|^2 d\Omega$$

1.6 Sampling

$y[n] \to$ [PAM (T, p(t))] $\to y(t)$

Ideal Pulse: $p(t) = \frac{sin(\frac{\pi t}{T})}{\frac{\pi t}{T}}$, $P(j\omega) = \frac{p(0)}{2\pi}u(\frac{\pi}{T} - |\omega|)$

$$Y(j\omega) = Y_d(e^{j\Omega})|_{\Omega=\omega T}P(j\omega)$$

$x_c(t) \to$ [C/D (T)] $\to x_d[n]$

$X_d(e^{j\Omega}) = \frac{1}{T}X_c(j\omega)|_{\omega=\frac{\Omega}{T}}$ for $|\Omega| < \pi$

2 State-Space Modeling

$$\frac{d}{dt}\vec{q} = A\vec{q}(t) + \vec{b}x(t)$$

$$y(t) = \vec{c}^T\vec{q}(t) + dx(t)$$

$$\vec{Q}(s) = (sI-A)^{-1}\vec{b}X(s) + (sI-A)^{-1}\vec{q}(0)$$

$Y(s) = \vec{\gamma}^T(sI-\Lambda)^{-1}\vec{\beta}X(s) + \vec{\gamma}^T(sI-\Lambda)^{-1}\vec{r}(0) + dX(s)$, where V is the eigenvector matrix, Λ is the associated eigenvalue matrix, $\vec{\gamma}^T = \vec{c}^TV$, $\beta = V^{-1}\vec{b}$, and $\vec{r}(t) = V^{-1}\vec{q}(t)$.

$$y(t) = L^{-1}[\sum_{k=1}^{L}\frac{\gamma_k\beta_k}{s-\lambda_k}] * x(t) + \sum_{k=1}^{L}\gamma_k r_k(0)e^{\lambda_k t} + dx(t)$$

$$\vec{q}[n+1] = A\vec{q}[n]] + \vec{b}x[n]$$

$$y[n] = \vec{c}^T\vec{q}[n] + dx[n]$$

$$Q(z) = (zI-A)^{-1}\vec{b}X(z) + z(zI-A)\vec{q}[0]$$

$$y[n] = Z^{-1}[\sum_{k=1}^{L}\frac{\gamma_k\beta_k}{z-\lambda_k}] * x[n] + \sum_{k=1}^{L}\gamma_k\lambda_k^n r_k[0] + dx[n]$$

2.1 Observability & Reachability

Observability: When all γ_k are nonzero, the system is said to observable, ie, every state can be observed at the output, given appropriate initial conditions or input.
Reachability: When all β_k are nonzero, the system is said to be reachable, ie, every state can be reached given an appropriate input.

2.2 State-Space Representation of Nonlinear Systems

Given $\vec{q} = \vec{f}(\vec{q}, x)$ and $y = g(\vec{q}, x)$.
Steady state can be obtained by setting $\frac{d}{dt}\vec{q} = 0$ or $\vec{q}[n+1] = \bar{q}$.
Linearization is achieved by partially deriving each f_i with respect to each state, which yields A, partially deriving each f_i with respect to x, which yields $\vec{b}$, partially deriving g with respect to each state, which yields $\vec{c}^T$, and partially deriving g with respect to x, yielding d.

2.3 State-Space Feedback

In general, the plant of the system will be nonlinear and require a continuum of states to represent it fully. Our model attempts to capture the essential lower order behavior of the system. Even if our model were perfect, disturbances, noise, and a lack of knowledge of the initial state of the system would yield inaccurate estimates of the system's actual behavior. Therefore, feedback must be used to compensate for these discrepancies. Suppose a plant is described by the following equations:

$$\vec{q}[n+1] = A\vec{q}[n] + \vec{b}x[n] + \vec{w}[n]$$

$$y[n] = \vec{c}^T\vec{q}[n] + dx[n] + \varsigma[n]$$

$\vec{w}[n]$ represents the unknown disturbances of the system; $varsigma[n]$ represents the unknown output noise.
If we were to model our system using $\hat{q}[n+1] = A\hat{q}[n] + \vec{b}x[n]$ and $\hat{y}[n] = \vec{c}^T\hat{q}[n] + dx[n]$, then the error of the system, $\tilde{q}[n+1] = A\tilde{q}[n] + w[n]$, the error is directly determined by the disturbance $w[n]$. Thus, using a real-time estimator without feedback is undesirable.

2.3.1 State Observer

(NOTE: This is only used to model the plant, not to control it!) Now we express $\hat{q}[n+1]$ as $A\hat{q}[n] + \vec{b}x[n] - \vec{l}(y[n] - \hat{y}[n])$, where $\vec{l}$ is termed the observer gain.
$\tilde{q}[n+1] = (A + \vec{l}\vec{c}^T)\tilde{q}[n] + \vec{w}[n] + l\varsigma[n]$. The eigenvalues of the error are now those of $A + \vec{l}\vec{c}^T$.
The observer gain $\vec{l}$ can be used to set the observable eigenvalues of the system to any set of of self-conjugate points. Unobservable modes remain unobservable because no information about them "reaches" the output. A large gain $\vec{l}$, however, causes the output noise to become a factor in the equation. There is a tradeoff between the two.

2.3.2 State Feedback Control

Assume that the plant is modeled by $\vec{q}[n+1] = A\vec{q}[n] + \vec{b}x[n]$ and $y[n] = \vec{c}^T\vec{q}[n]$. Assume that to this system, we feed back $\vec{q}[n]$ and express $x[n] = p[n] + \vec{g}^T\vec{q}[n]$. With this choice of feedback, $q[n+1] = (A + \vec{b}\vec{g}^T)q[n] + \vec{b}p[n]$.
By varying $\vec{g}^T$, we can vary all reachable eigenvalues of the system; too high of a gain, however, can cause disturbances and output noise to be too much of a problem.

2.3.3 Observer-Based Feedback Control

Because we do not usually have access to the state of a system, we model the system using the observer and use its state to control the input: $x[n] = p[n] + \vec{g}^T\hat{q}[n] = p[n] + \vec{g}^T(\vec{q}[n] - \tilde{q}[n])$.
We apply observer feedback as in the state observer case: $\hat{q}[n+1] = A\hat{q}[n] + \vec{b}x[n] - \vec{l}(y[n] - haty[n])$. The resulting state-evolution equations for the system are summarized below:

$$\begin{bmatrix}\vec{q}[n+1]\\ \tilde{q}[n+1]\end{bmatrix} = \begin{bmatrix}A + \vec{b}\vec{g}^T & -bg^T\\ 0 & A + \vec{l}\vec{c}^T\end{bmatrix}\begin{bmatrix}\vec{q}[n]\\ \tilde{q}[n]\end{bmatrix} + \begin{bmatrix}\vec{b}\\ 0\end{bmatrix}p[n] + \begin{bmatrix}I\\ I\end{bmatrix}\vec{w}[n] + \begin{bmatrix}0\\ \vec{l}\end{bmatrix}\varsigma[n]$$

3 Probability

3.1 Helpful Definitions & Identities

$$E[X] \equiv \int_{-\infty}^{\infty} xf_X(x)dx$$

$$E[X+Y] = E[X] + E[Y]$$

$$E[aX+b] = aE[X] + b$$

$$\sigma_{xy} = cov(X,Y) = E[XY] - \mu_x\mu_y$$

$$\sigma_x^2 = cov(X,X) = E[X^2] - \mu_x^2$$

$$\text{var}(aX+b) = a^2\text{var}(X)$$

$$R_{XY} = E[XY]$$

$$S = \sum_{k=1}^{N} X_k \to f_S(s) = (f_{X_1} * \ldots * f_{X_N})(s)$$

$$E[E[Y|X]] = E[Y]$$

3.2 Estimation

3.2.1 Orthogonality

Given an MMSE estimate, the following conditions arise:

$$E_{Y,X}[\{Y - \hat{y}(X)\}h(X)] = 0$$

$$E_{Y,X}[Y - \hat{y}(X)] = 0 \text{ (Unbiased)}$$

3.2.2 MMSE

Known: $f_Y(y) \to \hat{Y} = E[Y]$
Known: $\vec{X} = \{X_1, \ldots, X_L\} \to \hat{Y} = \hat{Y}(\vec{X}) = E[Y|\vec{X}]$

3.2.3 LMMSE

$$\hat{Y}_l = \hat{y}_l(\vec{X}) = \mu_y + \sum_{j=1}^{L} a_j(X_j - \mu_{x_j}) = \mu_y + \sum_{j=1}^{L} a\tilde{X}_j$$

$$E[(Y - \hat{Y}_l)\tilde{X}_i] = 0\forall i$$

$$\begin{bmatrix}\sigma_{x_1x_1} & \cdots & \sigma_{x_1x_L}\\ \vdots & \ddots & \vdots\\ \sigma_{x_Lx_1} & \cdots & \sigma_{x_Lx_L}\end{bmatrix}\begin{bmatrix}a_1\\ \vdots\\ a_L\end{bmatrix} = \begin{bmatrix}\sigma_{x_1y}\\ \vdots\\ \sigma_{x_Ly}\end{bmatrix}$$

$$\vec{a} = \bar{\bar{C}}_{xx}^{-1}\bar{C}_{xy}$$

$$\epsilon = \sigma_y^2 - \bar{C}_{xy}^T(\bar{\bar{C}}_{xx})^{-1}\bar{C}_{xy}$$

4 Random Processes

The outcome of an experiment is a random signal. The signal itself can be thought of as an intersection of random variables at all times, ie:

$$X(t_1) \cap X(t_2) \cap \ldots \cap X(t_n), t_n - t_{n-1} = \Delta t, \Delta t \to 0$$

$X(t_1)$ denotes a possible outcome at time t_1 and has its own PDF $f_{X(t_1)}(x)$.

4.1 Definitions

Mean: $\mu_x(t_i) = E[X(t_i)]$
Auto-correlation: $R_{xx}(t_i, t_j) = E[X(t_i)X(t_j)]$
Auto-covariance: $C_{xx}(t_i, t_j) = R_{xx}(t_i, t_j) - \mu_x(t_i)\mu_x(t_j)$
Cross-correlation: $R_{xy}(t_i, t_j) = E[X(t_i)Y(t_j)]$
Cross-covariance: $C_{xy}(t_i, t_j) = R_{xy}(t_i, t_j) - \mu_x(t_i)\mu_y(t_j)$
Uncorrelated:
$C_{xy}(t_i, t_j) = 0 \to E[X(t_i)Y(t_j)] = E[X(t_i)]E[Y(t_j)]$

4.2 Strict and Wide Sense Stationary Processes

4.2.1 Definitions

Strict Sense Stationary: The statistics depend only on the relative times at which the samples are taken, ie,
$f_{x(t_1),\ldots,x(t_n)}(x_1, \ldots, x_n) =$
$f_{x(t_1+\tau),\ldots,x(t_n+\tau)}(x_1, \ldots, x_n)$
Independent and Identically Distributed: $f_{x(t_i)}(x) = f_x(x)\forall i$ and all pairs $X(t_i), X(t_j)$ are independent.
I.I.D $\to$ S.S.S.
$f_{x(t_1)\ldots,x(t_n)}(x_1, \ldots, x_n) = f_x(x_1)\ldots f_x(x_n)$
Wide Sense Stationary:

$$\mu_x(t) = \mu_x\forall t, R_{xx}(t_1, t_2) = R_{xx}(t_1 - t_2, 0) = R_{xx}(\tau)$$

4.2.2 Properties of WSS Processes

$$R_{xx}(\tau) = R_{xx}(-\tau), R_{xy}(\tau) = R_{yx}(-\tau)$$

$$C_{xx}(\tau) = C_{xx}(-\tau), C_{xy}(\tau) = C_{yx}(-\tau)$$

$$|C_{xx}(\tau)| \leq C_{xx}(0), R_{xx}(\tau) \geq 0\forall\tau$$

4.2.3 Power Spectral Density

$S_{xx}(j\omega) = FR_{xx}(\tau)$, $D_{xx}(j\omega) = FC_{xx}(\tau)$
S_{xx} and D_{xx} must be real, even, and positive.

4.3 Ergodicity

A process is termed ergodic when the characteristic of an ensemble of signals can be characterized by the time-dependent properties of one signal. A signal is ergodic in the mean when $\lim_{T\to\infty}\int_{-T}^{T} x(t)dt = \mu_x$. WSS processes with finite variance at each t and a $C_{xx} \to 0$ for $t \to \infty$ is ergodic in the mean. Second-order ergodicity implies that the process is ergodic in the mean and that $R_{xx}(\tau) = \lim_{T\to\infty}\frac{1}{2T}\int_{-\infty}^{\infty} x(t)dt$.

4.4 Estimation (based on other signals or past values)

$$\hat{x}[n_0 + m] = ax[n_0] + b$$

$$E\{(x[n_0+m] - \hat{[}n_0+m])x[n_0]\} = 0$$

$$E\{x[n_0+m] - \hat{x}[n_0+m]\} = 0$$

For a WSS process, $\hat{x}[n_0+m] = \mu_x + \frac{C_{xx}[m]}{C_{xx}[0]}(x[n_0] - \mu_x)$.

4.5 LTI Systems (for WSS processes)

$x(t) \to [h(t)\ (\text{L1})] \to y(t)$

$$E[y(t)] = H(0)\mu_x = \mu_y$$

$$R_{yx}(\tau) = h(\tau) * R_{xx}(\tau)$$

$$R_{yy}(\tau) = h(\tau)h(-\tau) * R_{xx}(\tau) = R_{xx}(\tau) * \bar{R}_{hh}(\tau)$$

$$S_{xy}(j\omega) = S_{xx}(j\omega)H(j\omega)$$

$$S_{yy}(j\omega) = S_{xx}(j\omega)|H(j\omega)|^2$$

4.6 Einstein-Wiener-Kinchin Theorem

$$x_T = w_Tx(t), w_T(t) = (1 - u(|t| - T))$$

$E[\bar{S}_{xx}] = \frac{1}{2T}R_{xx}(\tau)\int_{-\infty}^{\infty} w_T(\alpha)w_T(\alpha-\tau)d\alpha =$
$R_{xx}(\tau)\Lambda_T(\tau) \Leftrightarrow \frac{1}{2T}E[|X_T(j\omega)|^2]$
$\lim_{T\to\infty} R_{xx}(\tau)\Lambda_T(\tau) = R_{xx}(\tau)$
$\to S_{xx}(j\omega) = \lim_{T\to\infty}\frac{1}{2T}E[|X_T(j\omega)|^2]$

Implication: One can estimate $S_{xx}(j\omega)$ by averaging $|X_T(j\omega)|^2$ over many trials and dividing by $2T$. More iterations $\to$ less noise in the estimate. Longer $T \to$ more resolution.

4.7 Types of Noise

White: Flat power spectrum over all frequencies (delta in the τ/m domain). Colored: The opposite of white.

4.8 Estimation of $H(e^{j\Omega})$, S_{yx}, and $E\{x[n]\}$

$$H(e^{j\Omega}) = \frac{S_{yx}(e^{j\Omega})}{S_{xx}(e^{j\Omega})}$$

$$|H(e^{j\Omega})|^2 = \frac{S_{yy}(e^{j\Omega})}{S_{xx}(e^{j\Omega})}$$

Assume ergodicity:

$$E\{x[n]\} = \lim_{N\to\infty} \sum_{k=-N}^{N} \frac{x[k]}{2N+1}$$

$$S_{yx}[m] = E\{y[n+m]x[n]\} = \lim_{N\to\infty} \sum_{k=-N}^{N} \frac{y[k+m]x[k]}{2N+1}$$

4.9 Wiener Filtering

Given $x[n]$; would like to approximate $y[n]$ with $\hat{y}[n]$. $e[n] \equiv \hat{y}[n] - y[n]$. Minimizing $\epsilon = E\{e^2[n]\}$ for a given $h[\cdot]$ yields the best-case estimator.

The minimization is accomplished by setting $\frac{\partial \epsilon}{\partial h[m]} = 0 \forall h[m] \neq 0$.

Causal DT FIR: $R_{ex}[m] = E\{e[n]x[n-m]\} = 0 \to R_{\hat{y}x}[m] = R_{yx}[m] \to R_{\hat{y}x} = \sum_k h[k]R_{xx}[m-k] = R_{yx}[m]$, which yields N equations in N unknowns.

Non-causal DT IIR: $H(z) = \frac{S_{yx}(z)}{S_{xx}(z)}$. MMSE $= R_{ee}[0] = \frac{1}{2\pi}\int_{-\pi}^{\pi} S_{ee}(e^{j\Omega})d\Omega = \frac{1}{2\pi}\int_{-\pi}^{\pi}(S_{yy} - HS_{xy})d\Omega$ $= \frac{1}{2\pi}\int_{-\pi}^{\pi} S_{yy}(1 - \frac{S_{yx}S_{xy}}{S_{yy}S_{xx}}d\Omega = \frac{1}{2\pi}\int_{-\pi}^{\pi} S_{yy}(1 - \rho_{yx}\rho_{yx}^*)$.

Coherence function: $\rho_{yx}(e^{j\Omega}) = \frac{S_{yx}(e^{j\Omega})}{\sqrt{S_{yy}(e^{j\Omega})S_{xx}(e^{j\Omega})}}$.

Noncausal CT: $H(j\omega) = \frac{S_{yx}(j\omega)}{S_{xx}(j\omega)}$. Same expression for error as in the DT case, with bounds from $-\infty$ to ∞.

Causal CT & DT: Causal predictions involve converting a colored noise process to a zero-mean colored noise process, then to a white one through the use of an inverse filter, obtaining the best-case estimator for the white noise case, and adding back the mean.

Appropriate choices for filters are minimum-phase modeling filters. In the CT case, the poles and zeros of the filter must be in the left-half plane. In the DT case, the poles and zeros of the filter must be within the unit circle.

In the case of the one step estimator, $\tilde{x}[n+1] = \zeta[0]w[n+1] + \zeta[1]w[n] + \ldots$ The data point $w[n+1]$ is orthogonal to all of the previous data; therefore, the estimator $\zeta[1]w[n] + \ldots$ is the ideal linear estimator. The error is equal to $\zeta[0]w[n+1]$. The overall transfer function for this system is $z(M - \zeta[0])/M$.

4.10 Hypothesis Testing

Consider a signal $r(t) = h(t) * x(t) + v(t)$, where $x(t) = \sum a[n]p(t - nT)$, $p(t)$ is some pulse, and $v(t)$ is noise.

Focusing on a single symbol, $r(0) = a[0](p * h)(0) + v(0) = a[0]s(0) + v(0)$. Because the weight of $s(0)$ is irrelevant, $r(0)$ can be expressed as $r = a + v$.

We model R, A, and V as outcomes of random variables: $R = A + V$. Usually, A takes one of two values: a_0 and a_1.

On-off: $a_0 = 0$, $a_1 \neq 0$. Antipodal: $a_0 = -a_1 \neq 0$.

In order to determine which signal was sent, we formulate hypotheses:

$$H_0 : R = a_0 + V,\; H_1 : R = a_1 + V$$

4.10.1 Binary Hypothesis Testing

$\hat{H}, 'H_0', 'H_1'$ denote a decision.

$P(H_0 \text{ is true}) = P(H = H_0) = P(H_0) = p_0$

$P(H_1 \text{ is true}) = P(H = H_1) = P(H_1) = p_1$

Assuming we know $f_{R|H}(r|H_0)$, $f_{R|H}(r|H_1)$, p_0, and p_1, if V is independent of A:

$$f_{R|H}(r|H_0) = f_V(r - a_0),\; f_{R|H}(r|H_1) = f_V(r - a_1)$$

4.10.2 Hypothesis Error

$$P(\text{Error}) = P(H_0, 'H_1') + P(H_1, 'H_0')$$

$P('H_1'|H_0)$: probability of a False Alarm, or P_{FA}. $P('H_0'|H_1)$: probability of a Miss, or P_M. $P('H_1'|H_1)$: probability of Detection, or P_D.

Decision space: regions of the real line denoted by D_i, such that if $r \in D_i$, $'H_i'$ is chosen.

$$P_{FA} = \int_{D_1} f_{r|H}(r|H_0)dr,\; P_M = \int_{D_0} f_{r|H}(r|H_1)dr$$

Likelihood ratio: $\Lambda(r) = \frac{f_{r|H}(r|H_1)}{f_{r|H}(r|H_0)}$

4.10.3 Deciding with Minimum Probability of Error (MAP)

MAP: Maximum a posteriori rule.

Knowing nothing of R, $P(\text{Error}|'H_0') = 1 - p_0$ and $P(\text{Error}|'H_1') = 1 - p_1$. Therefore, you should choose the H that yields the lowest error.

Knowing $R = r$, we must choose the hypothesis with maximal conditional probability: $'H_1'$ if $P(H_1|R = r) > P(H_0|R = r)$ or vice versa.

The adopted notational convention is the following:

$$P(H_1|R = r) \underset{'H_0'}{\overset{'H_1'}{\gtrless}} P(H_0|R = r)$$

$P(\text{Error}|R = r) = min\{1 - P(H_0|R = r), 1 - P(H_1|R = r)\}$

$P(\text{Error}) = \int_{\{r|r\in R\}} P(Error|R = r)f_R(r)dr$

Using Bayes' Rule:

$$P(H_1|R = r) \underset{'H_0'}{\overset{'H_1'}{\gtrless}} P(H_0|R = r) \Leftrightarrow \frac{f_{R|H}(r|H_1)}{f_{R|H}(r|H_0)} \underset{'H_0'}{\overset{'H_1'}{\gtrless}} \frac{p_0}{p_1}$$

4.10.4 Neyman-Pearson Detection

p_0 and p_1 are unknown. Strategy: Maximize P_D while keeping P_{FA} below some threshold.

$$\Lambda(r) \underset{'H_0'}{\overset{'H_1'}{\gtrless}} \eta$$

4.11 Signal Detection

$r(t), s(t), w(t)$: received signal; deterministic, sent signal; white, additive noise.

Difference between standard hypothesis testing and signal detection: $R = r$ / $\vec{R} = \{R_0, R_1, \ldots, R_L\}$. This set of R is modeled as a random process $R[n]$ of finite length L.

$H_0 : R[n] = W[n]$, $H_1 : R[n] = s[n] + W[n]$

The "algorithm" (MAP):

$$P(H_1|\vec{r}) \underset{'H_0'}{\overset{'H_1'}{\gtrless}} P(H_0|\vec{R})$$

$$\frac{f_{R|H}(\vec{r}|H_1)}{f_{R|H}(\vec{r}|H_0)} \underset{'H_0'}{\overset{'H_1'}{\gtrless}} \frac{P(H_0)}{P(H_1)}$$

Because $W[n]$ is white and Gaussian:

$$f(\vec{r}|H_0) = \prod_{i=1}^{L} \frac{1}{\sqrt{2\pi\sigma^2}} \exp -\frac{r^2[i]}{2\sigma^2} = \frac{1}{(2\pi\sigma^2)^{L/2}} \exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{L} r^2[i]\right)$$

Similarly,

$$f(\vec{r}|H_1) = \frac{1}{(2\pi\sigma^2)^{L/2}} \exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{L}(r[i] - s[i])^2\right)$$

Given these PDFs, the Hypothesis condition simplifies to $g \underset{'H_0'}{\overset{'H_1'}{\gtrless}} \gamma$, where $g = \bar{R}_{rs}[0] = \sum_{i=1}^{L} r[i]s[i]$, $\epsilon = \sum_i s^2[i]$, and

$$\gamma = \sigma^2 \log\left(\frac{P(H_0)}{P(H_1)}\right) + \frac{\epsilon}{2}.$$

A random variable $G = \sum_{n=1}^{L} W[n]s[n]$ can be used to denote the sum in the Hypothesis condition. We will be dealing with this random variable G from now on.

$H_0 \to \sigma_G^2 = \sigma^2\epsilon$, $\mu_G = 0$, $H_1 \to \sigma_G^2 = \sigma^2\epsilon$, $\mu_G = \epsilon\sigma^2$. This is where the fabled $\frac{\epsilon}{\sigma^2}$ noise comes from.

4.11.1 Matched Filtering

The sum g can be calculated by passing r through an LTI filter $h(\cdot)$ and sampling at time 0. $h[n] = s[-n]$ leads to matched filtering.

4.11.2 Signal Classification (Multiple $s(t)$)s

M distinct hypotheses: $H_i : R[n] = S_i[n] + W[n]$. There are (M-1) non-zero s_i; the task is to pick H given $\vec{r}$. The H with greated $g_i + \frac{\epsilon_i}{2} + \sigma^2 \log(P(H_i))$ is picked under MAP.

4.11.3 General Detector Structure

Suppose we receive a signal $r[n] = s[n] + w[n]$ and we would like to pick a filter h that outputs $g[n]$, such that $g[0]$ minimizes the probability of error given some threshold.

$H_1 : g[n] = s[n] * h[n] + w[n] * h[n]$ / $H_0 : g[n] = w[n] * h[n]$

$H_1 : E\{g[n]\} = \sum_{n=-\infty}^{\infty} h[n]s[-n] \equiv \mu$ / $H_0 : E\{g[n]\} = 0$

For convenience, we normalize $h[n]$, ie, $\sum h^2[n] = 1$.

$v = w * h \to R_{vv}[m] = R_{ww}[m] * \bar{R}_{HH}[m] \to \sigma_v^2 = \sigma_w^2 \sum h^2[n] = \sigma_w^2$

$f_{G|H}(g|H_1) = N(\mu, \sigma_w^2)$ / $f_{G|H}(g|H_1) = N(0, \sigma_w^2)$

$$N = \frac{1}{2\pi\sigma^2} \exp\left(-\frac{(n-\mu)^2}{2\sigma^2}\right)$$

P_{FA} remains unaffected by our choice of $h[n]$.

$P_D = \int_\gamma^\infty f_{G|H}(g[0]|H_1)dg$ is affected by μ, which is determined by $h[n]$.

The resulting condition, $|\sum_n h[n]s[-n]|^2 \leq (\sum_n h^2[n])(\sum_n s^2[-n])$, by the Cauchy-Schwarz inequality.

The only way to achieve equality is set $h[n] = c_0 s[-n]$. The resulting optimal filter is $h[n] = \frac{1}{\sqrt{\epsilon_s}} s[-n]$.

4.11.4 Maximizing SNR

SNR $= \frac{E\{g[0]|H_1\}^2}{2}$. Maximizing the SNR tries to separate the two distributions as much as possible. Generally, maximizing SNR does not correspond to minimizing error.

4.11.5 Pulse Detection with Colored Noise

$H_1 : r[n] = s[n] + v[n]$ / $H_0 : r[n] = v[n]$, where $v[n]$ is a colored noise process such that $S_{vv}(e^{j\Omega}) > 0\forall\Omega$, then a whitening filter $h_w[n]$ can be applied to the input, yielding a white noise process and a new pulse $p[n] = s[n] * h_w[n]$. Optimization can be carried out for the additive white noise, yielding $h_f[n]$. The ideal filter, then, is simply $h[n] = h_f[n] * h_w[n]$. This process maximizes $\epsilon_P = \frac{1}{2\pi}\int_{<2\pi>} |H_w(e^{j\Omega})|^2|S(e^{j\Omega})|^2 d\Omega = \frac{1}{2\pi}\int_{<2\pi>} \frac{|S(e^{j\Omega})|^2}{S_{vv}(e^{j\Omega})} d\Omega$.

5 Pulse Amplitude Modulation

$x(t) = \sigma_n a[n]p(t - nT)$, where $a[n]$ corresponds to pulse amplitudes, T corresponds to pulse repetition interval, $\frac{1}{T}$ corresponds to the baud rate, and $p(t) = Au(1 - |\frac{t}{\Delta}|)$. Polar or antipodal systems broadcast {1,-1}; bipolar systems broadcast {1,0,-1}. Return-to-zero systems have a $\Delta < T$; non-return-to-zero systems have a $\Delta = T$.

Consider an input signal that is fed through a channel with transfer function $h(t)$, to which noise $\eta(t)$ is added, resulting in $r(t)$. $r(t)$ is then filtered, resulting in a signal $b(t)$, which is then sampled every T, in order to recover $x[n]$.

$X(j\omega) = A(e^{j\Omega})|_{\Omega=\omega T} P(j\omega)$. In the absence of noise, $R(j\omega) = H(j\omega)X(j\omega)$ and $B(j\omega) = F(j\omega)H(j\omega)X(j\omega)$. Note that the information of $a[n]$ will, in general, populate $|\Omega| < \pi$. Therefore, knowledge of $A(e^{j\Omega})$ for a range $|\Omega| < \Omega_a < \pi$ will be insufficient. Thus, if $P(j\omega)F(j\omega)H(j\omega) \neq 0$ for $|\omega| \leq \frac{\pi}{T}$, then all of the information is of $a[n]$ is preserved. Note that this implies that $P(j\omega) \neq 0$, $F(j\omega) \neq 0$, $H(j\omega) \neq 0$.

5.1 Intersymbol Interference

$b(t) = f(t) * h(t) * x(t) = \sum_n a[n]g(t - nT)$, where $g(t) = f(t) * h(t) * p(t)$.

The requirement for no ISI to occur is thus $g(0) = c \neq 0$; $g(nT) = 0 \forall n \neq 0$.

5.2 Nyquist Condition

Consider sampling $g(t)$ with deltas: $\hat{g}(t) = g(t)\sum_{n=-\infty}^{\infty} \delta(t - nT)$. By the proposed no-ISI condition, $\hat{g}(t) = c\delta(t)$, and $\hat{G}(j\omega) = \frac{1}{T}\sum_{m=-\infty}^{\infty} G(j\omega - jm\frac{2\pi}{T}) = c$, which implies that the sum of displaced $G(j\omega)$ must add up to a constant. Sinc pulses satisfy this condition; however, because of the slow roll-off of sincs and their non-causality, unwanted coupled-ISI propagates too much. Thus, pulses with roll-off that varies as $\frac{1}{t^2}$ are preferred.

Smoother transitions can be obtained by using the following formula:

$$f(t) * h(t) * p(t) = \frac{\sin\left(\frac{\pi}{T}t\right)}{\frac{\pi}{T}t} \frac{\cos\left(\beta\frac{\pi}{T}t\right)}{1-(2\beta t/T)^2}$$

5.2.1 Carrier Transmission

Passband PAM: $s(t) = \sum_n a[n]p(t - nT)\cos(\omega_c t + \theta_c)$

Frequency-Shift Keying (FSK): $s(t) = \sum_n a[n]p(t - nT)\cos((\omega_0 + \delta_n)t + \theta_c)$; information can be encoded in shifts in frequency.

Phase-Shift Keying (PSK): $s(t) = \sum_n a[n]p(t - nT)\cos(\omega_c t + \theta_n)$; $a[n] = a_0$; $\theta_n = \frac{2\pi b_n}{M}$, where M is the number of symbols.

$s(t) = \sum_n \Re\{a_0 e^{j\theta_n} p(t - nT)e^{j\omega_c t}\} = I(t)\cos(\omega_c t) - Q(t)\sin(\omega_c t)$, where $I(t) = \sum_n a_i[n]p(t - nT)$ and $Q(t) = \sum_n a_q[n]p(t - nT)$, and $a_i[n] = a\cos(\theta_n)$ and $a_q[n] = a\sin(\theta_n)$.

Quadrature Amplitude Modulation (QAM): $a_i, a_q \in \pm a, \pm 3a$.

$r_i(t) = \frac{1}{2}I(t) - \frac{1}{2}I(t)\cos(2\omega_c t) - \frac{1}{2}Q(t)\sin(2\omega_c t)$ and $r_q(t) = \frac{1}{2}I(t)\sin(2\omega_c t) + \frac{1}{2}Q(t) - \frac{1}{2}Q(t)\cos(2\omega_c t)$. Low-pass filtering recovers $r_i(t)$ and $r_q(t)$.