F2E5216/TS1002 Adaptive Filtering and Change Detection

Bo Wahlberg (KTH) and Fredrik Gustafsson (LiTH)

Lecture 1

The goal of the course is to get an understanding for the theory of model based filtering and change detection, with particular attention to applications.

Course Organization

- **Lectures and compendium:** Theory, Algorithms, Applications, Evaluation
- **Toolbox and manual:** Algorithms, Evaluation $\leftrightarrow$ Implementations
- **Projects:** Application $\leftrightarrow$ Evaluation of Algorithms
- **Exercises:** To each part available from homepage
- **Examination, 6 credits:** Three-day take-home exam.

The Books

Steven Kay: Selected parts of Fundamentals of statistical signal processing: Volume II, detection theory. Prentice Hall, 1998.
Only Lecture 1

Fredrik Gustafsson: Adaptive filtering and change detection. John Wiley & Sons, 2001.

Lecture plan

Lecture 1: Detection theory from Kay's book.
Lecture 2: Applications. Detection of mean changes: likelihood principles; Chapters 1, 2, 3 .
Lecture 3: Detection of mean changes: The parallel filter concept. Chapters 3,4, Appendix 4.
Lecture 4: Parameter estimation: recursive least squares. Residual based detection. Chapter 5
Lecture 5: Parameter estimation: parallel filters and filter bank change detection . Chapter 6-7
Lecture 6: State estimation: Kalman, EKF and particle filters. Residual based detection. Chapter 8
Lecture 7: State estimation: parallel filters and filter bank change detection. Chapters 9, 10. Given by Fredrik
Lecture 8: State estimation: the parity space approach. Chapter 11
Lecture 9: State estimation: parity space diagnosability and PCA.

Lecture 1

- Basic Probability Theory
- Basic Detection Theory
- Neyman -Pearson's theorem
- Basic Hypothesis tests
- Generalized Likelihood Ratio Tests

Pages 1- 245 in Kay

Probability Theory

Stochastic variable: X

Probability density function (PDF): $p(x)$

Mean: $E(X) = \int_{-\infty}^{\infty} xp(x)dx$

Variance: $Var(X) = \int_{-\infty}^{\infty} (x - E(X))^2 p(x)dx$

Cummulative distribution function: $\Phi(x) = P(X < x) = \int_{-\infty}^{x} p(x)dx$

Right tail probability: $Q(x) = P(X > x) = \int_{x}^{\infty} p(x)dx$

Gaussian or Normal

$$\mathrm{N}(\mu, \sigma^2)$$

$p(x)$	$\frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$
$\mathrm{E}(x)$	μ
$\mathrm{Var}(x)$	σ^2
$Q(x)$	$\approx \frac{1}{\sqrt{2\pi}(x-\mu)/\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$
$\mathrm{E}(x^n)$	$1 \cdot 3 \cdot 5 \ldots (n-1)\sigma^2$
Matlab	`Q(x)=0.5*erfc((x-mu)/sigma/sqrt(2))`

Exponential

$$\mathrm{Exp}(\mu)$$

$p(x)$	$\frac{1}{\mu} e^{-\frac{x}{\mu}},\ x \geq 0$
$\mathrm{E}(x)$	μ
$\mathrm{Var}(x)$	μ^2
$Q(x)$	$e^{-\frac{x}{\mu}}$

Chi-squared with ν degrees

$$\chi^2_\nu$$

$p(x)$	$\frac{1}{2^{\nu/2}\Gamma(\nu/2)} x^{-\nu/2-1} e^{-x/2}, x > 0$
$\Gamma(u)$	$\int_0^\infty t^{u-1} e^{-t} dt, \Gamma(n) = (n-1)!$
	$\Gamma(u) = (u-1)\Gamma(u-1), \Gamma(0.5) = \sqrt{\pi}$
$\mathrm{E}(x)$	ν
$\mathrm{Var}(x)$	2ν
$Q(x)$	$e^{-x/2} \sum_{k=0}^{\nu/2-1} \frac{(x/2)^k}{k!}, \nu \geq 2$ and even
Comments	$\chi^2_2 = Exp(2)$.
	If $x = \sum_{i=1}^{\nu} x_i^2, x_i \in \mathrm{N}(0,1)$, then $x \in \chi^2_\nu$
	approximately Gaussian for large ν

Non-central chi-squared with ν degrees

$$\chi^2_\nu(\lambda)$$

$p(x)$	$\frac{1}{2}\left(\frac{x}{\lambda}\right)^{\frac{\nu-2}{4}} e^{-(x+\lambda)/2} I_{\nu/2-1}(\sqrt{\lambda x}), x > 0$
$I_r(u)$	$\sum_{k=0}^{\infty} \frac{(u/2)^{2k+r}}{k!\Gamma(r+k+1)}$ Bessel function
$\mathrm{E}(x)$	$\nu + \lambda$
$\mathrm{Var}(x)$	$2\nu + 4\lambda$
Comments	If $x = \sum_{i=1}^{\nu} x_i^2, x_i \in \mathrm{N}(\mu_i, 1)$, then $x \in \chi^2_\nu(\sum_i \mu_i)$

Detection Principles

Given: Data $x[n]$ with known distribution under the *null hypothesis* H_0 and the *alternative hypothesis* H_1, respectively.

General test: Form a test statistic $T(x)$.

$$\begin{array}{lll} \text{Decide } H_0 & \text{if} & T(x) < \gamma \textit{ (threshold)} \\ \text{Decide } H_1 & \text{if} & T(x) > \gamma \end{array}$$

Basic design parameters:

- *Probability of false alarm* (Type I error) $P_{FA} = P(H_1|H_0) = \int_{x:T(x)>\gamma} p(x|H_0)dx$
- *Probability of detection* (Power) $P_D = P(H_1|H_1) = \int_{x:T(x)>\gamma} p(x|H_1)dx$
- Miss $P_M = P(H_0|H_1) = 1 - P_D$ (Type II error)

The only 'hard' result in detection theory

Neyman-Pearson's theorem: To maximize the power P_D for a given probability of false alarm $P_{FA} = \alpha$, take

$$T(x) = \frac{p(x|H_1)}{p(x|H_0)}.$$

The test is called **LRT**, the **Likelihood Ratio Test**.

The threshold is found from

$$P_{FA} = \int_{x:T(x)>\gamma} p(x|H_0)dx = \alpha$$

Proof:

Define the LRT test $\mathcal{T} = \begin{cases} 1, & \text{if, } p(x|H_1) > \gamma p(x|H_0) \\ 0, & \text{if, } p(x|H_1) < \gamma p(x|H_0) \end{cases}$

Let $0 \leq \mathcal{T}' \leq 1$ denote any test with less or equal P'_{FA}.

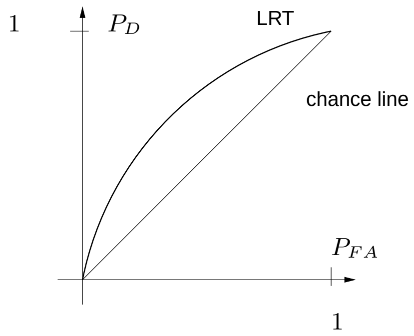
Key inequality (Check the signs!)

$$\int_{-\infty}^{\infty} (\mathcal{T} - \mathcal{T}')(p(x|H_1) - \gamma p(x|H_0))dx \geq 0$$

Re-arrange the integral: $(P_D - P'_D) \geq \gamma(P_{FA} - P'_{FA}) \geq 0$, where the last equality follows from the false alarm assumption.

This shows $\mathcal{T}$ optimize the power! One has to be a bit more careful if $p(x|H_1) = \gamma p(x|H_0)$ has a non-zero probability

Receiver Operating Characteristics



The slope of the LRT ROC curve is γ!
The curve is concave.
The LRT maximizes any performance criterion that values low false alarm probabilities and/or high detection probabilities.

Proof of Slope

$$E(T(x)^n|H_1) = \int \left[\frac{p(x|H_1)}{p(x|H_0)}\right]^n p(x|H_1)dx =$$

$$\int \left[\frac{p(x|H_1)}{p(x|H_0)}\right]^{n+1} p(x|H_0)dx = E(T(x)^{n+1}|H_0)$$

which can be written $\int t^n g(t|H_1)dt = \int t^{n+1} g(t|H_0)dt, \; \forall n$

where $g(t|H_i)$ is the PDF for the likelihood ratio under H_i. This means that $g(t|H_1) = tg(t|H_0)$. Now

$$P_D = \int_{\gamma}^{\infty} g(t|H_1)dt, \quad P_{FA} = \int_{\gamma}^{\infty} g(t|H_0)dt$$

and

$$\frac{dP_D}{dP_{FA}} = \frac{dP_D}{d\gamma} / \frac{dP_{FA}}{d\gamma} = \frac{g(t|H_1)}{g(t|H_0)} = \gamma$$

Observe that $\gamma \geq 0$

Sufficient Statistics

Replace x with the simpler sufficient statistics t,

$$T(x) = \frac{p(x|H_1)}{p(x|H_0)} = \frac{g(t|H_1)}{g(t|H_0)}$$

Example: $x = (x_0 \ldots x_{N-1})$, $x_i \in \mathrm{N}(m, \sigma^2)$, take

$$t = \frac{1}{N} \sum_{i=0}^{N-1} x_i$$

Basic Hypothesis Test

Is there a known mean A in the observed signal?

$$H_0 : x[n] = w[n]$$
$$H_1 : x[n] = A + w[n]$$

The more general problem of detecting a given signal

$$H_0 : x[n] = w[n]$$
$$H_1 : x[n] = s[n] + w[n]$$

can be rewritten to unknown mean.

Example 1: Communication

The signal $s(t) = 2A\sin(\omega t)$ is transmitted on the carrier ω and $y(t) = s(t) + w(t)$ is received. Let

$$\begin{aligned} x &= \frac{1}{T}\int_0^T y(t)\sin(\omega t) \\ &= A + \frac{1}{T}\int_0^T w(t)\sin(\omega t) \\ &= A + w, \quad \text{under } H_1 \end{aligned}$$

where $w \sim \mathrm{N}(0, \frac{\sigma^2}{2T})$.

Example 2: Matched filter

The signal $s[n]$ with energy E may be contained in $y[n]$. Let

$$\begin{aligned} x &= \frac{1}{N}\sum_{n=0}^{N-1} y[n]s[n] \\ &= E + w \quad \text{under } H_1 \end{aligned}$$

where $w \sim \mathrm{N}(0, \frac{E\sigma^2}{N})$.

Neyman-Pearson

$$T(x) = \frac{\frac{1}{(2\pi\sigma^2)^{N/2}} e^{-\frac{\sum_{n=0}^{N-1}(x[n]-A)^2}{2\sigma^2}}}{\frac{1}{(2\pi\sigma^2)^{N/2}} e^{-\frac{\sum_{n=0}^{N-1}(x[n])^2}{2\sigma^2}}} = e^{-\frac{NA^2 - \sum_{n=0}^{N-1} 2Ax[n]}{2\sigma^2}} > \gamma.$$

Taking logarithm and simplifying give

$$\bar{T}(x) = \frac{1}{N}\sum_{n=0}^{N-1} x[n] > \frac{\sigma^2}{NA}\log(\gamma) + \frac{A}{2} = \bar{\gamma}.$$

Now $\bar{T}(x) \in \begin{cases} \mathrm{N}\left(0, \frac{\sigma^2}{N}\right) & \text{under } H_0 \\ \mathrm{N}\left(A, \frac{\sigma^2}{N}\right) & \text{under } H_1 \end{cases}$

so $P_{FA} = P(\bar{T}(x) > \bar{\gamma}|H_0) = Q\left(\frac{\bar{\gamma}}{\sqrt{\sigma^2/N}}\right)$

$$\bar{\gamma} = \sqrt{\frac{\sigma^2}{N}} Q^{-1}(P_{FA})$$

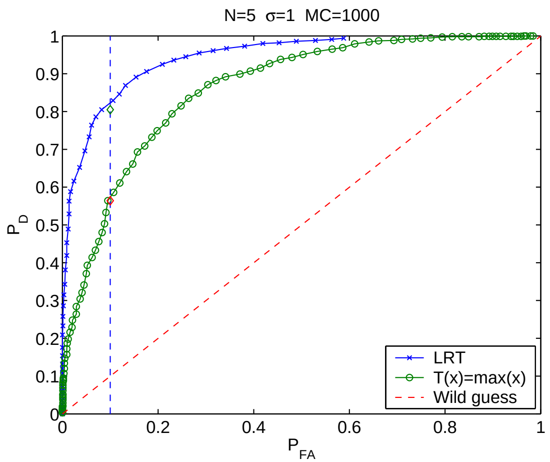
$$P_D = P(\bar{T}(x) > \bar{\gamma}|H_1) = Q\left(\frac{\bar{\gamma} - A}{\sqrt{\sigma^2/N}}\right) = Q\left(Q^{-1}(P_{FA}) - \sqrt{\frac{NA^2}{\sigma^2}}\right)$$

The last expression shows that P_D is a function of the desired P_{FA} and SNR, NA^2/σ^2, only.

Standard plots:

- **ROC:** Receiver Operating Characteristics plots P_D versus P_{FA} with $\bar{\gamma}$ being the parameter. Just guessing H_1 with probability P_{FA} gives $P_D = P_{FA}$. All clever tests give a curve above this straight line. Neyman-Pearson gives the upper bound curve for all tests.
- **Detection performance:** P_D versus SNR for a given P_{FA}. Again this performance is maximized for the Neyman-Pearson test.
- **BER:** Bit-Error Rate. In communication, both hypotheses are equal, and the design is to get $P_D = P_{FA}$. A BER-plot shows $P_D = P_{FA}$ versus SNR.

ROC



Typical Matlab code for ROC curves:

```
gamma=[-0.1:0.1:4]; MC=1000; N=5; A=1; sigma=1;
x=sigma*randn(N,MC); T1=mean(x); T2=max(x);
for k=1:length(gamma)
    PD1(k)=mean(T1+A>gamma(k));
    PD2(k)=mean(T2+A>gamma(k));
    PFA1(k)=mean(T1>gamma(k));
    PFA2(k)=mean(T2>gamma(k));
end
```

Simple line search to design γ and calculate P_D given P_{FA}:

```
PFAdesign=0.1; ind=find(PFA1<PFAdesign);
gamma1=gamma(ind(1)),
PD1design=PD1(ind(1)), ind=find(PFA2<PFAdesign);
gamma2=gamma(ind(1)), PD2design=PD2(ind(1))
```

`interp1` better to use.

Maximum Likelihood Estimation

Probability density function for the observation x parameterized by the unknown parameter θ:

$$p(x,\theta) \quad \Rightarrow \hat{\theta}_{ML} = \arg\max_{\theta} p(x_{observed},\theta)$$

Example: If $x = (x_0 \dots x_{N-1})$, $x_i \in \mathrm{N}(\theta, \sigma^2)$, then

$$\hat{\theta}_{ML} = \frac{1}{N}\sum_{i=0}^{N-1} x_i$$

Some Properties

Asymptotic distribution:

$$\sqrt{N}(\hat{\theta}_{ML} - \theta_0) \sim N(0, [I(\theta_0)]^{-1})$$

where the *Fisher information matrix* equals

$$I(\theta) = E\left[\frac{d}{d\theta}\log p(x,\theta)\right]\left[\frac{d}{d\theta}\log p(x,\theta)\right]^T$$

Asymptotically the best possible estimator
(reach the Cramér-Rao bound)

The GLRT for unknown A

$$\begin{aligned} H_0 &: x \in p_0(x|\theta_0) \\ H_1 &: x \in p_1(x|\theta_1) \end{aligned}$$

General principle: Plug in the maximum likelihood estimate of A under H_1, and let

$$L(x) = \frac{\max_{\theta_1} p_1(x|\theta_1)}{\max_{\theta_0} p_0(x|\theta_0)} = \frac{p_1(x|\hat{\theta}_1^{ML})}{p_0(x|\hat{\theta}_0^{ML})}$$

Generalized Likelihood Ratio Test (GLRT)

$$\begin{aligned} &\text{Decide } H_0 \quad \text{if} \quad L(x) < \gamma \\ &\text{Decide } H_1 \quad \text{if} \quad L(x) > \gamma \end{aligned}$$

Some more common special case:

$$\begin{aligned} H_0 &: \quad \theta = \theta_0 \\ H_1 &: \quad \theta \neq \theta_0 \end{aligned}$$

GLRT becomes $L(x) = \dfrac{\max_{\theta_1} p(x|\theta_1)}{p(x|\theta_0)} = \dfrac{p(x|\hat{\theta}_1^{ML})}{p(x|\theta_0)}$

General result for asymptotic distribution:

$$2\log L(x) \sim \begin{cases} \chi_n^2 \text{ under } H_0 \\ \chi_n^2(\lambda) \text{ under } H_1. \end{cases}$$

$$\lambda = (\theta_1 - \theta_0)^T I(\theta_0)(\theta_1 - \theta_0)$$

where $I(\theta)$ is Fisher's information matrix

Example:

Consider

$$H_0: \quad x[n] = w[n]$$
$$H_1: \quad x[n] = A + w[n]$$

with unknown A and Gaussian noise with known variance σ^2. Then

$$2 \log L(x) = \frac{N\hat{A}^2}{\sigma^2}$$
$$\hat{A} = \frac{1}{N}\sum_{n=0}^{N-1} x[n] \in \begin{cases} \mathrm{N}\left(0, \frac{\sigma^2}{N}\right) & \text{under } H_0 \\ \mathrm{N}\left(A, \frac{\sigma^2}{N}\right) & \text{under } H_1 \end{cases}$$
$$\Rightarrow 2 \log L(x) \in \begin{cases} \chi_1^2 \text{ under } H_0 \\ \chi_1^2(\frac{NA^2}{\sigma^2}) \text{ under } H_1. \end{cases}$$

The MLRT for unknown A

$$H_0: \quad x \in p_0(x|\theta_0)$$
$$H_1: \quad x \in p_1(x|\theta_1)$$

General principle: Use a prior on θ and integrate out its influence in the LRT. This is the *Bayesian approach*.

$$L(x) = \frac{\int_{\theta_1} p_1(x|\theta_1)p(\theta_1)d\theta_1}{\int_{\theta_0} p_0(x|\theta_0)p(\theta_0)d\theta_0}$$

This is the **Marginalized Likelihood Ratio Test (MLRT)**.

Mixed M/G LRT approach

Test:

$$H_0 : \theta = \theta_0, \eta$$
$$H_1 : \theta \neq \theta_0, \eta$$

Here η is an unknown *nuisance parameter*. General principle using the Bayesian approach: marginalize nuisance parameters, estimate the other parameters.

The LRT becomes

$$L(x) = \frac{\max_\theta \int_\eta p(x|\theta,\eta)d\eta}{\int_\eta p(x|\theta_0,\eta)d\eta} = \frac{\int_\eta p(x|\hat{\theta}^{ML},\eta)d\eta}{\int_\eta p(x|\theta_0,\eta)d\eta}$$

Nuisance parameters: classical approaches

Non-Bayesian approaches for the test

$$H_0: \quad \theta = \theta_0, \eta$$
$$H_1: \quad \theta \neq \theta_0, \eta$$

GLRT ML estimate of nuisance parameters:

$$L(x) = \frac{p(x|\hat{\theta}^{ML}, \hat{\eta}^{ML})}{p(x|\theta_0, \hat{\eta}^{ML})}$$

General result for asymptotic distribution:

$$2\log L(x) \sim \begin{cases} \chi^2_{\dim(\theta)} \text{ under } H_0 \\ \chi^2_{\dim(\theta)}(\lambda) \text{ under } H_1. \end{cases}$$

$$\lambda = (\theta_1 - \theta_0)^T \underbrace{\left(I_{\theta,\theta} - I_{\theta,\eta} I^{-1}_{\eta,\eta} I_{\eta,\theta}\right)}_{[I^{-1}]_{\theta,\theta}} (\theta_1 - \theta_0)$$

where $I(\theta)$ is Fisher's information matrix partitioned as

$$I = \begin{pmatrix} I_{\theta,\theta} & I_{\theta,\eta} \\ I_{\eta,\theta} & I_{\eta,\eta} \end{pmatrix}$$

evaluated at the true θ_0, η_0 (also θ_1 denotes the true value under H_1). Note that the underbraced expression $[I^{-1}]_{\theta,\theta}$ is the Schur complement to compute the upper left corner of the matrix inverse.

Wald test

Use

$$L(x) = (\hat{\theta} - \theta_0)^T [I^{-1}(\widehat{\theta,\eta}|H_1)]_{\theta,\theta} (\hat{\theta} - \theta_0)$$

where $\widehat{\theta,\eta}|H_1$ is the ML estimate under H_1.

Same performance as GLRT.

Rao test

Use

$$L(x) = \left.\frac{d\log(p(x|\theta,\eta))}{d\theta}\right|^T_{\theta_0,\hat{\eta}|H_0} [I^{-1}(\theta_0, \hat{\eta}|H_0)]_{\theta,\theta} \left.\frac{d\log(p(x|\theta,\eta))}{d\theta}\right|_{\theta_0,\hat{\eta}|H_0}$$

where $\hat{\eta}|H_0$ is the ML estimate under H_0.

Same performance as GLRT.

Note that θ does not have to be estimated here, as it has in the Wald test.

Locally Most Powerful (LMP) test

The Locally Most Powerful (LMP) test is a special case of the Rao test for the case of scalar parameter, no nuisance parameter and one-sided test:

$$H_0 : \theta = \theta_0$$
$$H_1 : \theta > \theta_0$$

Take the test statistic

$$L(x) = \frac{\left.\frac{d\log(p(x|\theta))}{d\theta}\right|_{\theta_0}}{\sqrt{I(\theta_0)}} \sim \begin{cases} \mathrm{N}(0,1) \text{ under } H_0 \\ \mathrm{N}(\sqrt{I(\theta_0)}(\theta_1 - \theta_0), 1) \text{ under } H_1 \end{cases}$$

for small changes $\theta_1 - \theta_0 > 0$. The derivation is based on local analysis by Taylor expansion.

Bayesian extensions

Priors: Assume priors $p(H_0)$ and $p(H_1)$, respectively. Bayesian decision rule, compare the *a posteriori* distribution of each hypothesis,

$$
\begin{aligned}
p(H_1|x) &\underset{H_1}{\overset{H_0}{\lessgtr}} p(H_0|x) \Leftrightarrow \\
\frac{p(x|H_1)p(H_1)}{p(x)} &\underset{H_1}{\overset{H_0}{\lessgtr}} \frac{p(x|H_0)p(H_0)}{p(x)}) \quad \Rightarrow \\
\frac{p(x|H_1)}{p(x|H_0)} &\underset{H_1}{\overset{H_0}{\lessgtr}} \frac{p(H_1)}{p(H_0)} = \gamma
\end{aligned}
$$

which shows that Bayes' rule provides a way to design the threshold in the LRT test.

Multiple tests

The use of prior enables multiple hypotheses tests:

$$\hat{k} = \arg\max_k p(H_k|x) = \arg\max_k p(x|H_k)p(H_k)$$

Bayes Risk

Define a cost to each decision

$$C_{ij} = P(H_i|H_j)$$

that is, C_{10} is the false alarm cost and $C_{0,1}$ is the missed detection cost.

Then

$$\frac{p(x|H_1)}{p(x|H_0)} \underset{H_1}{\overset{H_0}{\lessgtr}} \frac{(C_{10} - C_{00})p(H_1)}{(C_{01} - C_{11})p(H_0)} = \gamma$$

minimizes the Bayes risk or expected cost

$$\mathcal{R} = \mathrm{E}(C) = \sum_{i=0}^{1}\sum_{j=0}^{1} C_{ij}P(H_i|H_j)p(H_j)$$

Linear regressions

Model: $Y = \Phi^T\theta + E$

where $\mathrm{Cov}(E) = C$ and

$$p(Y|\theta) = \frac{1}{(2\pi)^{N/2}\sqrt{\det(C)}} e^{-\frac{1}{2}(Y-\Phi^T\theta)^T C^{-1}(Y-\Phi^T\theta)}.$$

Known parameter

Test: $\begin{cases} H_0: & \theta = 0 \\ H_1: & \theta = \theta_1 \end{cases}$

Neyman-Pearson gives LRT: $\log\left(\frac{p(Y|\theta_1)}{P(Y|\theta=0)}\right)$

$$= -0.5(Y - \Phi^T\theta_1)^T C^{-1}(Y - \Phi^T\theta_1) + 0.5Y^T C^{-1} Y$$
$$= Y^T C^{-1}\Phi^T\theta_1 - 0.5\theta_1^T \Phi C^{-1}\Phi^T\theta_1$$

The last term in independent of data, so take

$$L(Y) = Y^T C^{-1}\Phi^T\theta_1 \quad \in \mathrm{N}(\theta^T\Phi C^{-1}\Phi^T\theta_1, \theta_1^T\Phi C^{-1}\Phi^T\theta_1),$$

Constant in noise: $P_D = Q\left(Q^{-1}(P_{FA}) - \sqrt{\theta_1^T\Phi C^{-1}\Phi^T\theta_1}\right)$

Correlation detector or Matched Filter

Step 1 Pre-whiten the data:

$$\bar{Y} = C^{-1/2}Y, \quad \bar{\Phi} = C^{-1/2}\Phi \;\Rightarrow$$
$$\bar{Y} = \bar{\Phi}^T\theta + \bar{E} \qquad \mathrm{Cov}(\bar{E}) = I$$

Step 2: $L = \bar{Y}^T M$, where $M = \bar{\Phi}^T\theta_1$

Correlation detector: $L = \sum_{i=1}^{N} \bar{y}[i]m[i]$

Matched filter: Convolution $l(t) = \sum_{i=0}^{t-1} h[i]\bar{y}[t-i]$
Impulse response $h[i] = m[N-i]$, (matched to the signal)
Test statistics: $L = l(N)$

Unknown parameter

Test

$$H_0: \quad \theta = 0$$
$$H_1: \quad \theta \neq 0$$

ML estimate

$$\hat{\theta} = (\Phi\Phi^T)^{-1}\Phi Y$$

GLRT

$$L(x) = 2\log\left(\frac{p(Y|\hat{\theta})}{P(Y|\theta=0)}\right)$$
$$= -(Y - \Phi^T\hat{\theta})^T C^{-1}(Y - \Phi^T\hat{\theta}) + Y^T C^{-1} Y$$
$$= 2Y^T C^{-1}\Phi^T\hat{\theta} - \hat{\theta}^T\Phi C^{-1}\Phi^T\hat{\theta}$$
$$= ... = Y^T C^{-1}\Phi^T(\Phi C^{-1}\Phi^T)^{-1}\Phi C^{-1} Y$$
$$\in \chi^2_{\dim(\theta)}$$

See Chapter 6.1 in Gustafsson for more calculations on the linear model.

Exercises

- Check the lecture notes and report errors to me
- Groups of at most five persons
- I want one TEX version of problems + solution per group
- Kay: 1.1, 1.2, 1.4, 2.2, 3.1, 3.8, 3.13