11.5 MAP Estimator

Recall that the "hit-or-miss" cost function gave the MAP estimator… it maximizes the a posteriori PDF

Q: Given that the MMSE estimator is "the most natural" one… why would we consider the MAP estimator?

A: If $\mathbf{x}$ and θ are not jointly Gaussian, the form for MMSE estimate requires integration to find the conditional mean.

MAP avoids this Computational Problem!

Note: MAP doesn't require this integration

Trade "natural criterion" vs. "computational ease"

What else do you gain? More flexibility to choose the prior PDF

Notation and Form for MAP

Notation: $\hat{\theta}_{MAP}$ maximizes the posterior PDF

$$\hat{\theta}_{MAP} = \arg\max_{\theta} p(\theta \mid \mathbf{x})$$

"arg max" extracts the value of θ that causes the maximum

Equivalent Form (via Bayes' Rule):

$$\hat{\theta}_{MAP} = \arg\max_{\theta} [p(\mathbf{x} \mid \theta)\, p(\theta)]$$

Proof: Use $p(\theta \mid \mathbf{x}) = \dfrac{p(\mathbf{x} \mid \theta) p(\theta)}{p(\mathbf{x})}$

$$\hat{\theta}_{MAP} = \arg\max_{\theta} \left[\frac{p(\mathbf{x} \mid \theta) p(\theta)}{p(\mathbf{x})} \right] = \arg\max_{\theta} [p(\mathbf{x} \mid \theta)\, p(\theta)]$$

Does not depend on θ

Vector MAP

< Not as straight-forward as vector extension for MMSE >

The obvious extension leads to problems:

Choose $\hat{\theta}_i$ to minimize $\mathcal{R}(\hat{\theta}_i) = E\{C(\theta_i - \hat{\theta}_i)\}$

Exp. over $p(\mathbf{x}, \theta_i)$

$$\Rightarrow \quad \hat{\theta}_i = \arg\max_{\theta_i} p(\theta_i \mid \mathbf{x})$$

1-D marginal conditioned on $\mathbf{x}$

Need to integrate to get it!!

$$p(\theta_1 \mid \mathbf{x}) = \int \cdots \int p(\theta \mid \mathbf{x})\, d\theta_2 \cdots d\theta_p$$

Problem: The whole point of MAP was to avoid doing the integration needed in MMSE!!!

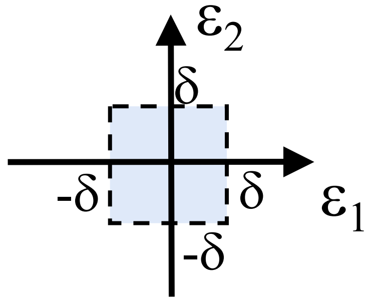
Is there a way around this?
Can we find an Integration-Free Vector MAP?

Circular Hit-or-Miss Cost Function

Not in Book

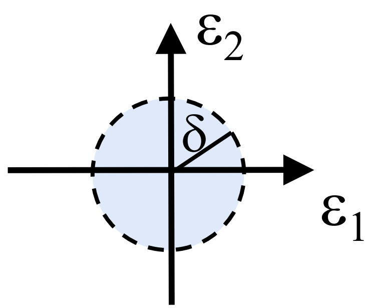
First look at the p-dimensional cost function for this "troubling" version of a vector map:

It consists of p individual applications of 1-D "Hit-or-Miss"



$$C(\varepsilon_1,\varepsilon_2) = \begin{cases} 0, & (\varepsilon_1,\varepsilon_2) \text{ in square} \\ 1, & (\varepsilon_1,\varepsilon_2) \text{ not in square} \end{cases}$$

The corners of the square "let too much in" $\Rightarrow$ use a circle!



$$C(\boldsymbol{\varepsilon}) = \begin{cases} 0, & \|\boldsymbol{\varepsilon}\| < \delta \\ 1, & \|\boldsymbol{\varepsilon}\| \geq \delta \end{cases}$$

This actually seems more natural than the "square" cost function!!!

MAP Estimate using Circular Hit-or-Miss

Back to Book

So… what vector Bayesian estimator comes from using this circular hit-or-miss cost function?

Can show that it is the following "Vector MAP"

$$\hat{\boldsymbol{\theta}}_{MAP} = \arg\max_{\boldsymbol{\theta}} \ p(\boldsymbol{\theta} \mid \mathbf{x})$$

Does Not Require Integration!!!

That is… find the maximum of the joint conditional PDF

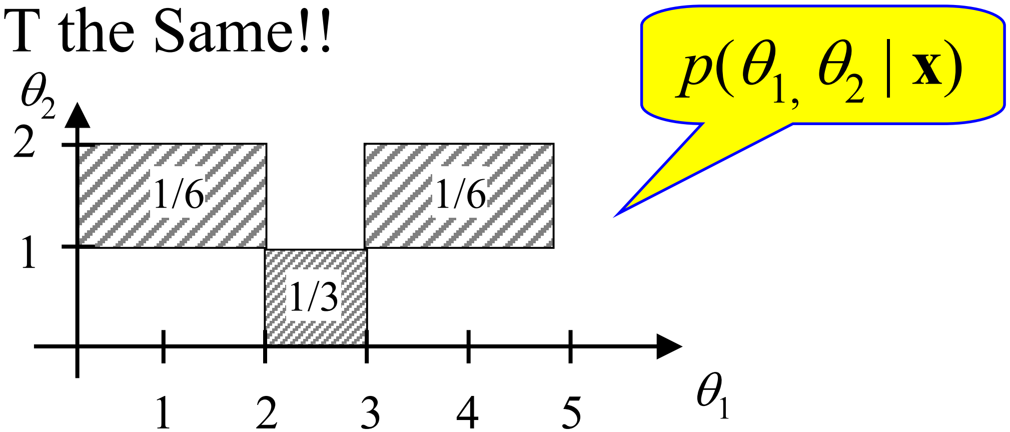
in all θ_i

conditioned on $\mathbf{x}$

How Do These Vector MAP Versions Compare

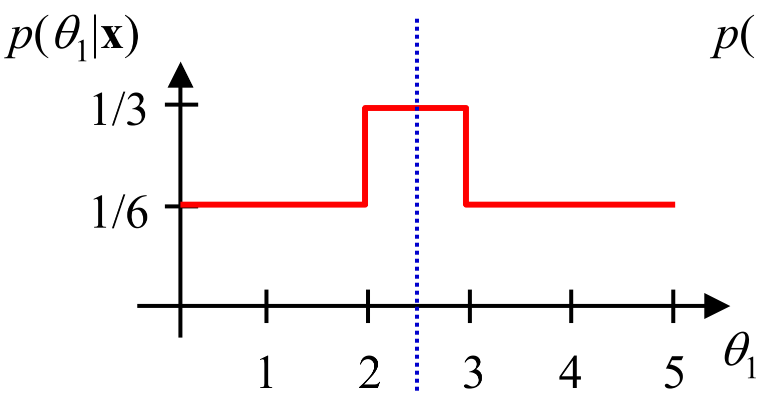
In general: They are NOT the Same!!

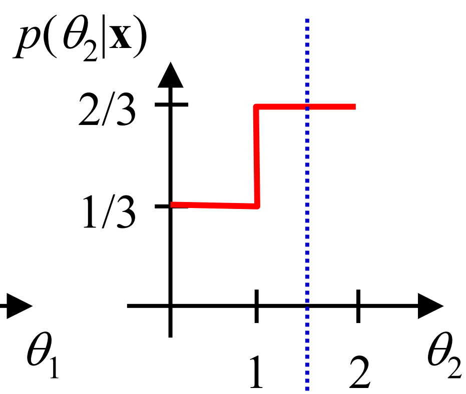
Example: $p = 2$



The vector MAP using Circular Hit-or-Miss is: $\hat{\boldsymbol{\theta}} = [2.5 \quad 0.5]^T$

To find the vector MAP using the element-wise maximization:





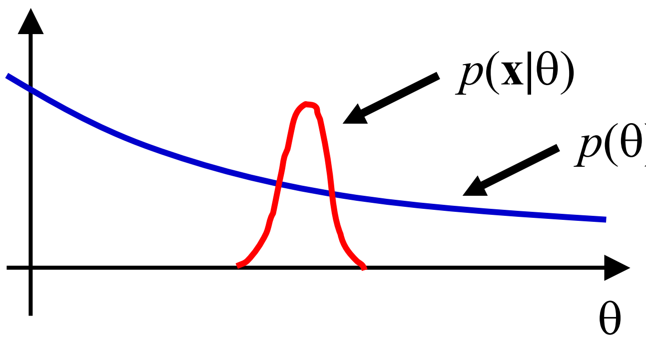
$\hat{\boldsymbol{\theta}} = [2.5 \quad 1.5]^T$

"Bayesian MLE"

Recall… As we keep getting good data, $p(\theta|\mathbf{x})$ becomes more concentrated as a function of θ. But… since:

$$\hat{\boldsymbol{\theta}}_{MAP} = \arg\max_{\boldsymbol{\theta}} p(\boldsymbol{\theta} \,|\, \mathbf{x}) = \arg\max_{\boldsymbol{\theta}} \left[p(\mathbf{x} \,|\, \boldsymbol{\theta})\; p(\boldsymbol{\theta}) \right]$$

… $p(\mathbf{x}|\theta)$ should also become more concentrated as a function of θ.



- Note that the prior PDF is nearly constant where $p(\mathbf{x}|\theta)$ is non-zero
- This becomes truer as $N \rightarrow \infty$, and $p(\mathbf{x}|\theta)$ gets more concentrated

$$\underbrace{\arg\max_{\boldsymbol{\theta}} \left[p(\mathbf{x} \,|\, \boldsymbol{\theta})\; p(\boldsymbol{\theta}) \right]}_{\text{MAP}} \approx \underbrace{\arg\max_{\boldsymbol{\theta}} p(\mathbf{x} \,|\, \boldsymbol{\theta})}_{\text{"Bayesian MLE"}}$$

Uses conditional PDF rather than the parameterized PDF

11.6 Performance Characterization

The performance of Bayesian estimators is characterized by looking at the estimation error:

$$\varepsilon = \theta - \hat{\theta}$$

Random (due to *a priori* PDF)

Random (due to **x**)

Performance characterized by error's PDF $p(\varepsilon)$

We'll focus on Mean and Variance

If ε is Gaussian then these tell the whole story

This will be the case for the Bayesian Linear Model (see Thm. 10.3)

We'll also concentrate on the MMSE Estimator

Performance of Scalar MMSE Estimator

The estimator is: $\hat{\theta} = E\{\theta \,|\, \mathbf{x}\}$

$$= \int \theta \, p(\theta \,|\, \mathbf{x}) d\theta$$

Function of **x**

So the estimation error is: $\varepsilon = \theta - E\{\theta \,|\, \mathbf{x}\} = f(\mathbf{x}, \theta)$

Function of two RV's

General Result for a function of two RVs: $Z = f(X, Y)$

$$E\{Z\} = \iint f(x,y) \, p_{XY}(x,y) \, dx \, dy$$

$$\text{var}\{Z\} = E\left\{(Z - E\{Z\})^2\right\} = \iint (f(x,y) - E\{Z\})^2 \; p_{XY}(x,y) \, dx \, dy$$

So… applying the mean result gives:

$$E\{\varepsilon\} = E_{\mathbf{x},\theta}\{\theta - E\{\theta \mid \mathbf{x}\}\}$$

$$= E_{\mathbf{x}}\left\{E_{\theta|\mathbf{x}}[\theta - E\{\theta \mid \mathbf{x}\}]\right\}$$

$$= E_{\mathbf{x}}\left\{ E_{\theta|\mathbf{x}}[\theta] - \underbrace{E_{\theta|\mathbf{x}}[E\{\theta \mid \mathbf{x}\}]}_{E\{\theta|\mathbf{x}\}} \right\}$$

$$= E_{\mathbf{x}}\left\{E_{\theta|\mathbf{x}}[\theta] - E\{\theta \mid \mathbf{x}\}\right\}$$

$$= E_{\mathbf{x}}\{0\} = 0$$

See Chart on "*Decomposing Joint Expectations*" in "Notes on 2 RVs"

Pass $E_{\theta|\mathbf{x}}$ through the terms

Evaluated as seen below

Two Notations for the same thing

$$E_{\theta|\mathbf{x}}[E\{\theta \mid \mathbf{x}\}] = \int_{\theta} \underbrace{E\{\theta \mid \mathbf{x}\}}_{\substack{\text{does not} \\ \text{depend on} \\ \theta}} p_{\theta|\mathbf{x}}(\theta \mid \mathbf{x})\, d\theta$$

$$= E\{\theta \mid \mathbf{x}\} \int_{\theta} p_{\theta|\mathbf{x}}(\theta \mid \mathbf{x})\, d\theta = E\{\theta \mid \mathbf{x}\}$$

$$\Rightarrow \quad E\{\varepsilon\} = 0$$

i.e., the Mean of the Estimation Error (over data & parm) is Zero!!!!

And… applying the variance result gives:

Use
$E\{\varepsilon\} = 0$

$$\operatorname{var}\{\varepsilon\} = E\left\{(\varepsilon - E\{\varepsilon\})^2\right\} = E\left\{\varepsilon^2\right\} = E\left\{(\theta - \hat{\theta})^2\right\}$$

$$= \iint (\theta - \hat{\theta})^2 \, p_{\mathbf{x},\theta}(\mathbf{x},\theta)\, d\mathbf{x}\, d\theta$$

$$= Bmse(\hat{\theta})$$

So… the MMSE estimation error has:

- mean = 0
- var = *Bmse*

So… when we minimize *Bmse* we are minimizing the variance of the estimate

If ε is Gaussian then $\varepsilon \sim N\left(0, Bmse(\hat{\theta})\right)$

Ex. 11.6: DC Level in WGN w/ Gaussian Prior

We saw that $$Bmse(\hat{A}) = \frac{1}{N/\sigma^2 + 1/\sigma_A^2}$$

with $$\hat{A} = \underbrace{\left(\frac{\sigma_A^2}{\sigma_A^2 + \sigma^2/N}\right)}_{\text{constant}} \bar{x} + \underbrace{\left(\frac{\sigma^2/N}{\sigma_A^2 + \sigma^2/N}\right)}_{\text{constant}} \mu_A$$

If X is Gaussian then
$Y = aX + b$
is also Gaussian

this is Gaussian because it is a linear combo of the jointly Gaussian data samples

So… $\hat{A}$ is Gaussian

$$\varepsilon \sim N\left(0, \frac{1}{N/\sigma^2 + 1/\sigma_A^2}\right)$$

Note: As N gets large this PDF collapses around 0.

This estimate is “consistent in the Bayesian sense”

Bayesian Consistency: For large N $\hat{A} \approx A$

(regardless of the realization of A!)

Performance of Vector MMSE Estimator

Vector estimation error: $\boldsymbol{\varepsilon} = \boldsymbol{\theta} - \hat{\boldsymbol{\theta}}$ The mean result is obvious.

Must extend the variance result:

$$\operatorname{cov}\{\boldsymbol{\varepsilon}\} = \mathbf{C}_{\boldsymbol{\varepsilon}} = E_{\mathbf{x},\boldsymbol{\theta}}\{\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^T\} \triangleq \mathbf{M}_{\hat{\boldsymbol{\theta}}}$$

Some New Notation… **"Bayesian Mean Square Error Matrix"**

Look some more at this:

$$\mathbf{M}_{\hat{\boldsymbol{\theta}}} = E_{\mathbf{x},\boldsymbol{\theta}}\{[\boldsymbol{\theta} - E\{\boldsymbol{\theta}|\mathbf{x}\}][\boldsymbol{\theta} - E\{\boldsymbol{\theta}|\mathbf{x}\}]^T\}$$

$$= E_{\mathbf{x}}\{\underbrace{E_{\boldsymbol{\theta}|\mathbf{x}}\{[\boldsymbol{\theta} - E\{\boldsymbol{\theta}|\mathbf{x}\}][\boldsymbol{\theta} - E\{\boldsymbol{\theta}|\mathbf{x}\}]^T\}}_{= \mathbf{C}_{\boldsymbol{\theta}|\mathbf{x}}}\}$$

$$= E_{\mathbf{x}}\{C_{\boldsymbol{\theta}|\mathbf{x}}\}$$

See Chart on "*Decomposing Joint Expectations*"

In general this is a function of **x**

***General* Vector Results**:

$$E\{\boldsymbol{\varepsilon}\} = \mathbf{0}$$

$$\mathbf{C}_{\boldsymbol{\varepsilon}} = \mathbf{M}_{\hat{\boldsymbol{\theta}}} = E_{\mathbf{x}}\{C_{\boldsymbol{\theta}|\mathbf{x}}\}$$

Why do we call the error covariance the "Bayesian MSE Matrix"?

The <u>Diagonal</u> Elements of $\mathbf{M}_{\hat{\boldsymbol{\theta}}}$ are Bmse's of the Estimates

To see this:

$$\left[E_{\mathbf{x},\boldsymbol{\theta}}\{\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^T\}\right]_{ii} = \int_{\mathbf{x}}\int_{\theta_1}\cdots\int_{\theta_p}[\theta_i - E\{\theta_i \mid \mathbf{x}\}]^2 \; p(\mathbf{x},\boldsymbol{\theta})\, d\mathbf{x}\, d\boldsymbol{\theta}$$

$$= \int_{\mathbf{x}}\int_{\theta_i}[\theta_i - E\{\theta_i \mid \mathbf{x}\}]^2 \; p(\mathbf{x},\theta_i)\, d\mathbf{x}\, d\theta_i$$

$$= Bmse(\theta_i)$$

Integrate over all the <u>other</u> parameters… "marginalizing" the PDF

Perf. of MMSE Est. for Jointly Gaussian Case

Let the data vector **x** and the parameter vector θ be jointly Gaussian.

Nothing new to say about the mean result: $E\{\boldsymbol{\varepsilon}\} = \mathbf{0}$

Now… look at the Error Covariance (i.e., Bayesian MSq Matrix):

Recall General Result: $\mathbf{C}_{\boldsymbol{\varepsilon}} = \mathbf{M}_{\hat{\boldsymbol{\theta}}} = E_{\mathbf{x}}\left\{C_{\boldsymbol{\theta}|\mathbf{x}}\right\}$

Thm 10.2 says that for Jointly Gaussian Vectors we get that…

$\mathbf{C}_{\theta|\mathbf{x}}$ does NOT depend on x

$$\Rightarrow \quad \mathbf{C}_{\boldsymbol{\varepsilon}} = \mathbf{M}_{\hat{\boldsymbol{\theta}}} = E_{\mathbf{x}}\left\{C_{\boldsymbol{\theta}|\mathbf{x}}\right\} = C_{\boldsymbol{\theta}|\mathbf{x}}$$

Thm 10.2 also gives the form as:

$$\begin{aligned}\mathbf{C}_{\boldsymbol{\varepsilon}} = \mathbf{M}_{\hat{\boldsymbol{\theta}}} &= C_{\boldsymbol{\theta}|\mathbf{x}} \\ &= \mathbf{C}_{\boldsymbol{\theta}} - \mathbf{C}_{\boldsymbol{\theta}\mathbf{x}}\mathbf{C}_{\mathbf{x}}^{-1}\mathbf{C}_{\mathbf{x}\boldsymbol{\theta}}\end{aligned}$$

Perf. of MMSE Est. for Bayesian Linear Model

Recall the model: $\mathbf{x} = \mathbf{H}\boldsymbol{\theta} + \mathbf{w}$

~$N(\boldsymbol{\mu}_\theta, \mathbf{C}_\theta)$

~$N(\mathbf{0}, \mathbf{C}_w)$

Nothing new to say about the mean result: $E\{\boldsymbol{\varepsilon}\} = \mathbf{0}$

Now… for the error covariance… this is nothing more than a special case of the jointly Gaussian case we just saw:

Results for Jointly Gaussian Case

$$\mathbf{C}_\varepsilon = \mathbf{M}_{\hat{\theta}} = C_{\theta|x} = \mathbf{C}_\theta - \mathbf{C}_{\theta x}\mathbf{C}_x^{-1}\mathbf{C}_{x\theta}$$

Evaluations for Bayesian Linear

$$\mathbf{C}_{\theta x} = \mathbf{C}_\theta \mathbf{H}^T$$

$$\mathbf{C}_x = \mathbf{H}\mathbf{C}_\theta\mathbf{H}^T + \mathbf{C}_w$$

$$\mathbf{C}_\varepsilon = \mathbf{M}_{\hat{\theta}} = C_{\theta|x} = \mathbf{C}_\theta - \mathbf{C}_\theta\mathbf{H}^T\left(\mathbf{H}\mathbf{C}_\theta\mathbf{H}^T + \mathbf{C}_w\right)^{-1}\mathbf{H}\mathbf{C}_\theta$$

$$= \left(\mathbf{C}_\theta^{-1} + \mathbf{H}^T\mathbf{C}_w^{-1}\mathbf{H}\right)^{-1}$$

Alternate Form … see (10.33)

Summary of MMSE Est. Error Results

1. For all cases: Est. Error is zero mean $E\{\boldsymbol{\varepsilon}\} = \mathbf{0}$

2. Error Covariance for three "Nested" Cases: $Bmse(\theta_i) = \left[\mathbf{M}_{\hat{\boldsymbol{\theta}}}\right]_{ii}$

General Case:

$$\mathbf{C}_{\boldsymbol{\varepsilon}} = \mathbf{M}_{\hat{\boldsymbol{\theta}}} = E_{\mathbf{x}}\left\{C_{\boldsymbol{\theta}|\mathbf{x}}\right\}$$

Jointly Gaussian:

$$\mathbf{C}_{\boldsymbol{\varepsilon}} = \mathbf{M}_{\hat{\boldsymbol{\theta}}} = C_{\boldsymbol{\theta}|\mathbf{x}} = \mathbf{C}_{\boldsymbol{\theta}} - \mathbf{C}_{\boldsymbol{\theta}\mathbf{x}}\mathbf{C}_{\mathbf{x}}^{-1}\mathbf{C}_{\mathbf{x}\boldsymbol{\theta}}$$

Bayesian Linear:

Jointly Gaussian & Linear Observation

$$\mathbf{C}_{\boldsymbol{\varepsilon}} = \mathbf{M}_{\hat{\boldsymbol{\theta}}} = C_{\boldsymbol{\theta}|\mathbf{x}} = \mathbf{C}_{\boldsymbol{\theta}} - \mathbf{C}_{\boldsymbol{\theta}}\mathbf{H}^T\left(\mathbf{H}\mathbf{C}_{\boldsymbol{\theta}}\mathbf{H}^T + \mathbf{C}_{\mathbf{w}}\right)^{-1}\mathbf{H}\mathbf{C}_{\boldsymbol{\theta}}$$

$$= \left(\mathbf{C}_{\boldsymbol{\theta}}^{-1} + \mathbf{H}^T\mathbf{C}_{\mathbf{w}}^{-1}\mathbf{H}\right)^{-1}$$

Main Bayesian Approaches

MMSE
"Squared" Cost Function
(In General: Nonlinear Estimate)

Estimate: $\hat{\boldsymbol{\theta}} = E\left\{\boldsymbol{\theta} \middle| \mathbf{x}\right\}$

Err. Cov.: $\mathbf{M}_{\hat{\boldsymbol{\theta}}} = E_{\mathbf{x}}\left\{\mathbf{C}_{\boldsymbol{\theta}|\mathbf{x}}\right\}$

Hard to Implement… numerical integration

Jointly Gaussian x and θ
(Yields Linear Estimate)

Estimate: $\hat{\boldsymbol{\theta}} = E\{\boldsymbol{\theta}\} + \mathbf{C}_{\boldsymbol{\theta}\mathbf{x}}\mathbf{C}_{\mathbf{xx}}^{-1}\left(\mathbf{x} - E\{\mathbf{x}\}\right)$

Err. Cov.: $\mathbf{M}_{\hat{\boldsymbol{\theta}}} = \mathbf{C}_{\boldsymbol{\theta\theta}} - \mathbf{C}_{\boldsymbol{\theta}\mathbf{x}}\mathbf{C}_{\mathbf{xx}}^{-1}\mathbf{C}_{\mathbf{x}\boldsymbol{\theta}}$

Easier to Implement… Determining $\mathbf{C}_{\boldsymbol{\theta}\mathbf{x}}$ can be hard to find

Bayesian Linear Model
(Yields Linear Estimate)

Estimate: $\hat{\boldsymbol{\theta}} = E\{\boldsymbol{\theta}\} + \mathbf{C}_{\boldsymbol{\theta}}\mathbf{H}^T\left(\mathbf{H}\mathbf{C}_{\boldsymbol{\theta}}\mathbf{H}^T + \mathbf{C}_w\right)^{-1}\left(\mathbf{x} - \mathbf{H}\boldsymbol{\mu}_{\boldsymbol{\theta}}\right)$

Err. Cov.: $\mathbf{M}_{\hat{\boldsymbol{\theta}}} = \mathbf{C}_{\boldsymbol{\theta}} - \mathbf{C}_{\boldsymbol{\theta}}\mathbf{H}^T\left(\mathbf{H}\mathbf{C}_{\boldsymbol{\theta}}\mathbf{H}^T + \mathbf{C}_w\right)^{-1}\mathbf{H}\mathbf{C}_{\boldsymbol{\theta}}$

"Easy" to Implement… Only need accurate model: $\mathbf{C}_{\boldsymbol{\theta}}$, $\mathbf{C}_{\mathbf{w}}$, $\mathbf{H}$

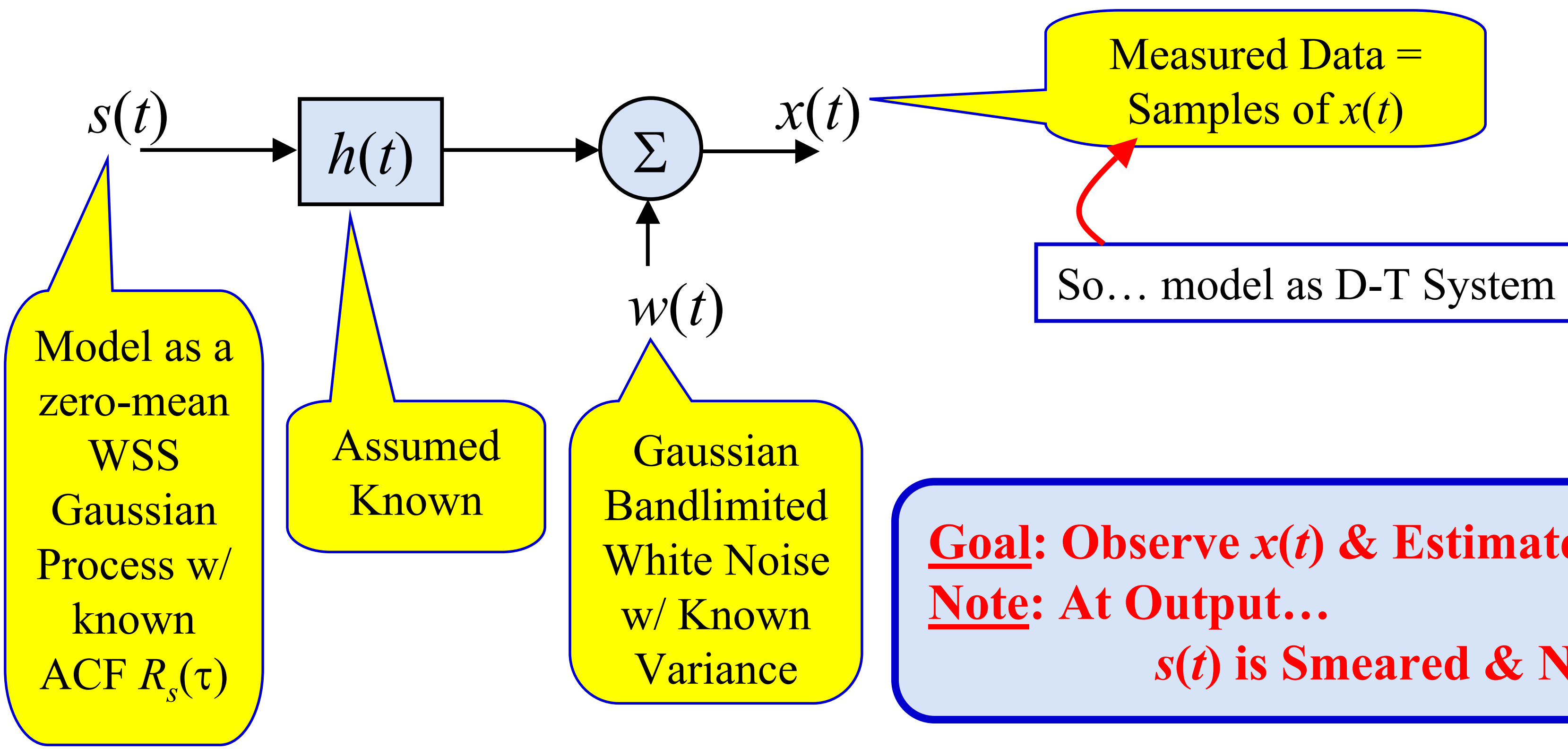
MAP
"Hit-or-Miss" Cost Function

Estimate: $\hat{\boldsymbol{\theta}} = \arg\max_{\boldsymbol{\theta}} \; p(\boldsymbol{\theta} \mid \mathbf{x})$

Easy to Implement… Performance Analysis is Challenging

11.7 Example: Bayesian Deconvolution

This example shows the power of Bayesian approaches over classical methods in signal estimation problems (i.e. estimating the signal rather than some parameters)



Goal: Observe $x(t)$ & Estimate $s(t)$

Note: At Output… $s(t)$ is Smeared & Noisy

Sampled-Data Formulation

$$\begin{bmatrix} x[0] \\ x[1] \\ \vdots \\ x[N-1] \end{bmatrix} = \begin{bmatrix} h[0] & 0 & 0 & \cdots & 0 \\ h[1] & h[0] & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ h[N-1] & h[N-1] & \cdots & \cdots & h[N-n_s] \end{bmatrix} \begin{bmatrix} s[0] \\ s[1] \\ \vdots \\ s[n_s-1] \end{bmatrix} + \begin{bmatrix} w[0] \\ w[1] \\ \vdots \\ w[N-1] \end{bmatrix}$$

Measured Data Vector **x**

Known Observation Matrix **H**

Signal Vector **s** to Estimate

AWGN: **w** $\mathbf{C_w} = \sigma^2\mathbf{I}$

We have modeled $s(t)$ as zero-mean WSS process with known ACF…
So… $s[n]$ is a D-T WSS process with known ACF $R_s[m]$…
So… vector **s** has a known covariance matrix (Toeplitz & Symmetric) given by:

$$\mathbf{C_s} = \begin{bmatrix} R_s[0] & R_s[1] & R_s[2] & \cdots & R_s[n_s-1] \\ R_s[1] & R_s[0] & R_s[1] & \ddots & \vdots \\ R_s[2] & R_s[1] & R_s[0] & \ddots & R_s[2] \\ \vdots & \ddots & \ddots & \ddots & R_s[1] \\ R_s[n_s-1] & \cdots & R_s[2] & R_s[1] & R_s[0] \end{bmatrix}$$

Model for Prior PDF is then $\mathbf{s} \sim N(\mathbf{0},\mathbf{C_s})$

s and **w** are independent

MMSE Solution for Deconvolution

We have the case of the Bayesian Linear Model… so:

$$\hat{\mathbf{s}} = \underbrace{\mathbf{C_s H}^T \left(\mathbf{H C_s H}^T + \sigma^2 \mathbf{I} \right)^{-1}}\mathbf{x}$$

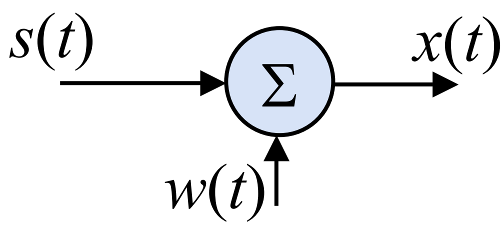
Note that this is a <u>linear</u> estimate
This matrix is called "The Weiner Filter"

The performance of the filter is characterized by:

$$\mathbf{C}_{\boldsymbol{\varepsilon}} = \mathbf{M}_{\hat{\mathbf{s}}} = \left(\mathbf{C}_{\mathbf{s}}^{-1} + \mathbf{H}^T \mathbf{H} / \sigma^2 \right)^{-1}$$

Sub-Example: No Inverse Filtering, Noise Only

Direct observation of $\mathbf{s}$ with $\mathbf{H} = \mathbf{I}$… $\mathbf{x} = \mathbf{s} + \mathbf{w}$



Goal: Observe $x(t)$ & "De-Noise" $s(t)$
Note: At Output… $s(t)$ with Noise

$$\hat{\mathbf{s}} = \mathbf{C_s}\left(\mathbf{C_s} + \sigma^2\mathbf{I}\right)^{-1}\mathbf{x}$$

$$\mathbf{C}_{\boldsymbol{\varepsilon}} = \mathbf{M}_{\hat{\mathbf{s}}} = \left(\mathbf{C_s}^{-1} + \mathbf{I}/\sigma^2\right)^{-1}$$

Note: Dimensionality Problem… **# of "parms" = # of observations**

Classical Methods Fail… $\hat{\mathbf{s}} = \mathbf{x}$ **Bayesian methods can solve it!!**

For insight… consider "single sample" case:

$$\hat{s}[0] = \frac{R_s[0]}{R_s[0] + \sigma^2} x[0] = \frac{\eta}{\eta + 1} x[0] \qquad \eta = \frac{R_s[0]}{\sigma^2} \quad (SNR)$$

High SNR	*Low SNR*
$\hat{s}[0] \approx x[0]$	$\hat{s}[0] \approx 0$
Data Driven	**Prior PDF Driven**

Sub-Sub-Example: Specific Signal Model

Direct observation of **s** with **H** = **I**… **x** = **s** + **w**

But here… the signal follows a specific random signal model

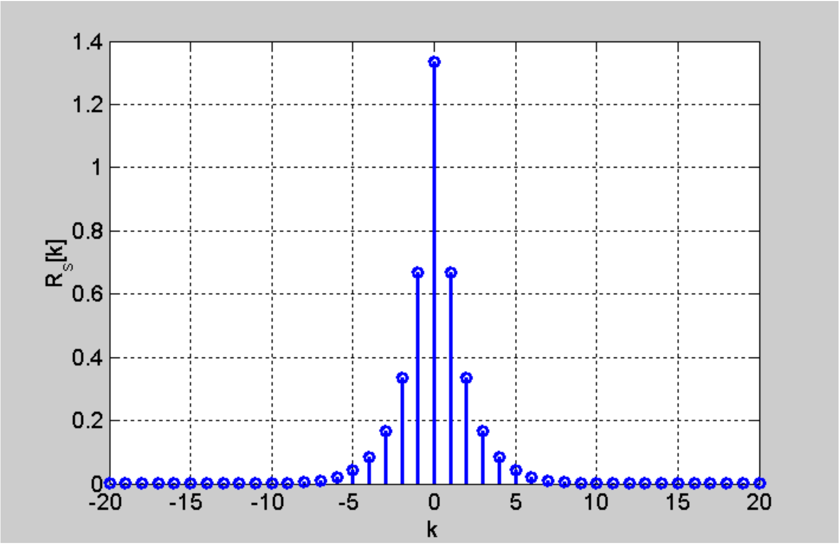
$$s[n] = -a_1 s[n-1] + u[n]$$

$u[n]$ is White Gaussian "Driving Process"

This is a 1st-order "auto-regressive" model: AR(1)

Such a random signal has an ACF & PSD of

$$R_s[k] = \left[\frac{\sigma_u^2}{1-a_1^2}\right](-a_1)^{|k|} \qquad P_s(f) = \frac{\sigma_u^2}{\left|1 + a_1\, e^{-j2\pi f}\right|^2}$$



See Figures 11.9 & 11.10 in the Textbook