BUT POSTERIOR PDF IS GAUSSIAN SINCE $\underline{x}, \underline{\theta}$ ARE JOINTLY GAUSSIAN.

LET $\underline{z} = \begin{bmatrix} \underline{x} \\ \underline{\theta} \end{bmatrix}$

$$\underline{z} = \begin{bmatrix} \underline{H}\underline{\theta} + \underline{w} \\ \underline{\theta} \end{bmatrix} = \begin{bmatrix} \underline{H} & \underline{I} \\ \underline{I} & \underline{0} \end{bmatrix} \begin{bmatrix} \underline{\theta} \\ \underline{w} \end{bmatrix}$$

$\underline{\theta}, \underline{w}$ ARE INDEPENDENT AND MARGINALLY GAUSSIAN $\Rightarrow$ JOINTLY GAUSSIAN

LET $\underline{x} \rightarrow \underline{H}\underline{\theta} + \underline{w}$

$\quad \underline{y} \rightarrow \underline{\theta}$

AND APPLY MULTIVARIATE GAUSSIAN RESULTS.

$$E(\underline{x}) = E(\underline{H}\underline{\theta} + \underline{w}) = \underline{H}E(\underline{\theta}) = \underline{H}\underline{\mu}_\theta$$

$$E(\underline{y}) = E(\underline{\theta}) = \underline{\mu}_\theta$$

$$\begin{aligned} C_{\underline{x}x} &= E\left[(\underline{x} - E(\underline{x}))(\underline{x} - E(\underline{x}))^T\right] \\ &= E\left[(\underline{H}\underline{\theta} + \underline{w} - \underline{H}\underline{\mu}_\theta)(\quad)^T\right] \\ &= E\left[(\underline{H}(\underline{\theta} - \underline{\mu}_\theta) + \underline{w})(\quad)^T\right] \\ &= \underline{H}E\left[(\underline{\theta} - \underline{\mu}_\theta)(\underline{\theta} - \underline{\mu}_\theta)^T\right]\underline{H}^T + E(\underline{w}\underline{w}^T) \\ &= \underline{H}C_\theta \underline{H}^T + C_w \end{aligned}$$

SINCE $\underline{\theta}, \underline{w}$ ARE INDEPENDENT

$$\underline{C}_{yx} = E[(\underline{y}-E(\underline{y}))(\underline{x}-E(\underline{x}))^T]$$
$$= E[(\underline{\theta}-\underline{\mu}_\theta)(\underline{H}(\underline{\theta}-\underline{\mu}_\theta)+\underline{w})^T]$$
$$= \underline{C}_\theta \underline{H}^T$$

$$\Rightarrow E(\underline{\theta}|\underline{x}) = \underline{\mu}_\theta + \underline{C}_\theta \underline{H}^T(\underline{H}\underline{C}_\theta\underline{H}^T+\underline{C}_w)^{-1}(\underline{x}-\underline{H}\underline{\mu}_\theta)$$

$$\underline{C}_{\theta|x} = \underline{C}_\theta - \underline{C}_\theta \underline{H}^T(\underline{H}\underline{C}_\theta\underline{H}^T+\underline{C}_w)^{-1}\underline{H}\underline{C}_\theta$$

<u>EXAMPLE</u> - DC LEVEL IN WGN - GAUSSIAN PRIOR

$$x[n] = \underset{\sim N(\mu_A, \sigma_A^2)}{\underset{\uparrow}{A}} + \underset{\text{WGN}}{\underset{\uparrow}{w[n]}} \qquad n = 0, 1, \ldots, N-1$$

A, w[n] ARE INDEPENDENT

$$\Rightarrow \underline{x} = \underline{1}A + \underline{w} \qquad \left(\begin{matrix} \mu_\theta = \mu_A & \underline{H} = \underline{1} \\ \underline{C}_\theta = \sigma_A^2 & \underline{C}_w = \sigma^2\underline{I} \end{matrix}\right)$$

$$\hat{A} = E(A|\underline{x}) = \mu_A + \sigma_A^2 \underline{1}^T(\underline{1}\sigma_A^2\underline{1}^T + \sigma^2\underline{I})^{-1} \cdot (\underline{x} - \underline{1}\mu_A)$$

USE WOODBURY'S IDENTITY TO INVERT MATRIX (SEE TEXT)

$$\hat{A} = \underset{\substack{\uparrow \\ \text{PRIOR} \\ \text{ESTIMATE}}}{\mu_A} + \underbrace{\underset{\substack{\uparrow \\ \text{GAIN}}}{\frac{\sigma_A^2}{\sigma_A^2 + \sigma^2/N}} \underbrace{(\bar{x} - \mu_A)}_{\text{"ERROR"}}}_{\text{DATA CORRECTION}} = \alpha\bar{x} + (1-\alpha)\mu_A$$

ALSO, $VAR(A|\underline{x}) = \sigma_A^2 - \sigma_A^2 \underline{1}^T (\underline{1}\sigma_A^2 \underline{1}^T + \sigma^2 \underline{I})^{-1} \cdot \underline{1}\sigma_A^2$

$$= \frac{\frac{\sigma^2}{N}\sigma_A^2}{\sigma_A^2 + \sigma^2/N}$$

CAN ALSO SHOW THAT

$$E(\underline{\theta}|\underline{x}) = \underline{\mu}_\theta + (\underline{C}_\theta^{-1} + \underline{H}^T \underline{C}_W^{-1} \underline{H})^{-1} \underline{H}^T \underline{C}_W^{-1} (\underline{x} - \underline{H}\underline{\mu}_\theta)$$

$$\underline{C}_{\theta|x} = (\underline{C}_\theta^{-1} + \underline{H}^T \underline{C}_W^{-1} \underline{H})^{-1}$$

USING STANDARD MATRIX IDENTITIES.

FOR NO PRIOR KNOWLEDGE, $\underline{C}_\theta^{-1} = \underline{0}$

$$E(\underline{\theta}|\underline{x}) = (\underline{H}^T \underline{C}_W^{-1} \underline{H})^{-1} \underline{H}^T \underline{C}_W^{-1} \underline{x}$$

SAME FORM AS EFFICIENT OR MVU ESTIMATOR FOR CLASSICAL LINEAR MODEL

NUISANCE PARAMETERS

CLASSICAL ESTIMATION – DC LEVEL IN WGN, A AND σ^2 ARE UNKNOWN BUT WE ARE

ONLY INTERESTED IN σ^2. MUST ALSO ESTIMATE A AS WELL. A TERMED <u>NUISANCE</u> PARAMETER.

BAYESIAN ESTIMATION - IF WE HAVE $p(A,\sigma^2|\underline{x})$, WE CAN "INTEGRATE OUT" A OR

$$p(\sigma^2|\underline{x}) = \int p(A,\sigma^2|\underline{x})\,dA$$

IN GENERAL, $\underline{\theta}$, $\underline{\alpha}$ ARE UNKNOWN. TO ESTIMATE $\underline{\theta}$,

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

OR ALTERNATIVELY SINCE

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

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

IF $\underline{\theta}$, $\underline{\alpha}$ ARE INDEPENDENT

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

$\Rightarrow$ FIND CONDITIONAL PDF $p(\underline{x}|\underline{\theta},\underline{\alpha})$, INTEGRATE OUT $\underline{\alpha}$ AND THEN USE

BAYES THEOREM AS USUAL

SEE EXAMPLE 10.3.

USING BAYESIAN ESTIMATORS FOR DETERMINISTIC PARAMETERS

SUPPOSE WE CAN'T FIND MVU ESTIMATOR. WITHIN BAYESIAN FRAMEWORK MMSE ESTIMATOR ALWAYS EXISTS ($E(\theta|\underline{x})$). AT LEAST WORKS WELL ON AVERAGE (DIFFERENT θ). MAY NOT WORK WELL FOR A GIVEN θ.

WHAT HAPPENS IF WE APPLY BAYESIAN ESTIMATOR TO CLASSICAL PROBLEM?

EXAMPLE - DC LEVEL IN WGN - CLASSICAL

BAYESIAN ESTIMATOR -

$$\hat{A} = \alpha \bar{x} + (1-\alpha)\mu_A \qquad 0 < \alpha < 1$$

$$\alpha = \frac{\sigma_A^2}{\sigma_A^2 + \sigma^2/N}$$

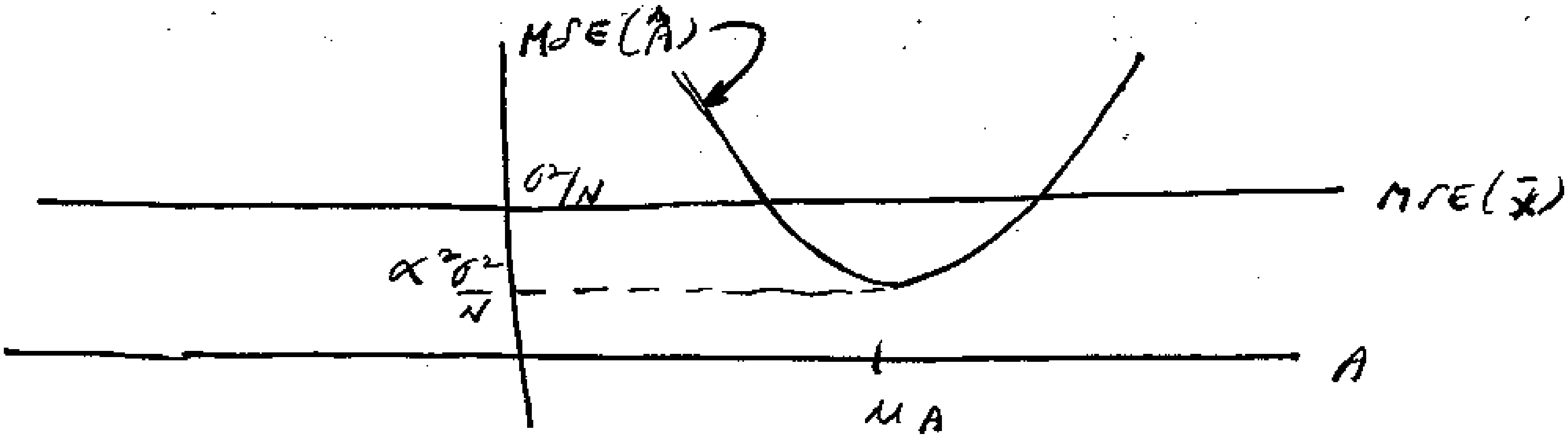
A IS NOW DETERMINISTIC. CLEARLY, $\hat{A}$ IS BIASED $\Rightarrow$ EVALUATE PERFORMANCE USING CLASSICAL MSE OR

$$mse(\hat{A}) = VAR(\hat{A}) + b^2(\hat{A})$$

WHERE $b(\hat{A}) = E(\hat{A}) - A = BIAS$

$$MSE(\hat{A}) = \underbrace{\alpha^2 VAR(\bar{x})}_{VAR(\hat{A})} + (\underbrace{\alpha A + (1-\alpha)\mu_A}_{E(\hat{A})} - A)^2$$

$$= \alpha^2 \sigma^2/N + (1-\alpha)^2 (A - \mu_A)^2$$



IF A IS CLOSE TO μ_A, WE DO BETTER THAN CLASSICAL, OTHERWISE NOT.
BAYESIAN ESTIMATOR IS NOT UNIFORMLY BETTER, BUT ONLY "ON AVERAGE"

NOW ASSUME A IS RANDOM, $A \sim N(\mu_A, \sigma_A^2)$
$\Rightarrow$ A WILL BE NEAR μ_A MOST OF TIME

$$BMSE(\hat{A}) = E_A(MSE(\hat{A}))$$
$$= \alpha^2\sigma^2/N + (1-\alpha)^2 E_A((A-\mu_A)^2)$$

$$= \alpha^2 \sigma^2/N + (1-\alpha)^2 \sigma_A^2$$

$$= \frac{\sigma^2}{N} \frac{\sigma_A^2}{\sigma_A^2 + \sigma^2/N} < \frac{\sigma^2}{N} = BMSE(\bar{x})$$

BAYESIAN ESTIMATOR USED WHEN A IS RANDOM TRADES OFF BIAS FOR VARIANCE. CAN REDUCE VARIANCE MORE THAN INCREASING $BIAS^2$ SO THAT OVERALL MSE REDUCED ON AVERAGE.

TO AVOID BIAS OF BAYESIAN ESTIMATOR

$$E(\hat{A}) = \frac{\sigma_A^2}{\sigma_A^2 + \sigma^2/N} A + \frac{\sigma^2/N}{\sigma_A^2 + \sigma^2/N} \mu_A$$

(BIAS TOWARD μ_A) LET $\sigma_A^2 \to \infty$
$\Rightarrow E(\hat{A}) \to A$. EQUIVALENT TO ASSUMING FLAT PRIOR PDF.

CALLED <u>NONINFORMATIVE</u> PRIOR PDF. WHEN NO PRIOR INFORMATION AVAILABLE BUT WE WANT TO USE BAYESIAN ESTIMATOR CHOOSE THIS PRIOR. SEE BOX-TIAO.