EECS 564 CRAMER-RAO BOUNDS

WINTER 1999

I. FOR ANY UNBIASED ESTIMATOR $\hat{a}(r)$ OF A PARAMETER (UNKNOWN CONSTANT) A,

$E[(\hat{a}(r)-A)^2] \geq 1/I_r(A)$ WHERE $I_r(A) = E[(\frac{\partial}{\partial A}\log p_{r|a}(R|A))^2] = -E[\frac{\partial^2}{\partial A^2}\log p_{r|a}(R|A)]$ (SEE BOTTOM).

MEAN SQUARE ERROR ≥ 1/FISHER INFO. FISHER INFO. IN r ABOUT A WILL GENERALLY BE A FUNCTION OF A

PROOF: LET $z_1 = \hat{a}(r) - A$ = ERROR BE RANDOM VARIABLES. $E[z_1] = 0$ (UNBIASED). $E[z_1^2] = E[(\hat{a}(r)-A)^2]$.

$z_2 = \frac{\partial}{\partial A}\log p_{r|a}(R|A)$. $E[z_2]=0$: SEE BELOW. $E[z_2^2] = I_r(A)$.

$$E[z_2] = \int\left(\frac{\partial}{\partial A}\log p_{r|a}(R|A)\right)p_{r|a}(R|A)\,dR = \int \frac{1}{p_{r|a}(R|A)}\left(\frac{\partial}{\partial A}p_{r|a}(R|A)\right)p_{r|a}(R|A)\,dR = \frac{\partial}{\partial A}\int p_{r|a}(R|A)\,dR = \frac{\partial}{\partial A}(1) = 0.$$

$$E[z_1z_2] = \int \underbrace{(\hat{a}(R)-A)}_{z_1}\underbrace{\left(\frac{\partial}{\partial A}\log p_{r|a}(R|A)\right)}_{z_2} p_{r|a}(R|A)\,dR \underset{\text{USING } E[z_2] \text{ ALGEBRA}}{=} \int(\hat{a}(R)-A)\frac{\partial}{\partial A}p_{r|a}(R|A)\,dR$$

$\boxed{\text{USE } u\,dv = d(uv) - v\,du}$ INTEGRATION BY PARTS

$$= \int\frac{\partial}{\partial A}[(\hat{a}(R)-A)p_{r|a}(R|A)]\,dR - \int p_{r|a}(R|A)(-1)\,dR = \frac{\partial}{\partial A}E[\underbrace{\cancel{\hat{a}(r)-A}}_{\text{UNBIASED}}]^{0} - (-1) = 1.$$

$\boxed{\text{PLUG INTO CAUCHY-SCHWARZ} \neq}$: $(E[z_1z_2])^2 \leq E[z_1^2]E[z_2^2] \rightarrow 1 \leq E[(\hat{a}(r)-A)^2]I_r(A) \rightarrow$ CRAMER RAO BOUND. (SINCE z_1, z_2 0-MEAN)

NOTE $\left\{\begin{array}{l}\frac{\partial^2}{\partial A^2}\log p_{r|a}(R|A)\\ \frac{\partial}{\partial A}\log p_{r|a}(R|A)\end{array}\right\}$ MUST EXIST AND BE ABSOLUTELY INTEGRABLE (USED ABOVE).

II. FOR ANY ESTIMATOR $\hat{a}(r)$ OF A PARAMETER A, LET BIAS $B(A) = E[\hat{a}(r)] - A$. THEN

$$E[(\hat{a}(r)-A-B(A))^2] = Var[\hat{a}(r)] \geq \left(1+\frac{dB(A)}{dA}\right)^2 / I_r(A).$$

1. NOTE THIS MAY BE LESS THAN PREVIOUS BOUND — REMOVING CONSTRAINT (UNBIASED) ON SET OF ESTIMATORS CONSIDERED.
2. IF $B(A)=0 \leftrightarrow$ UNBIASED, SAME AS ABOVE.

PROOF: NOW SET $z_1 = \hat{a}(r) - A - B(A)$, $E[z_1]=0$. $E[z_1^2] = Var(\hat{a}(r))$.

$z_2 = \frac{\partial}{\partial A}\log p_{r|a}(R|A)$ $E[z_2]$, $E[z_2^2]$ SAME AS ABOVE.

$$E[z_1z_2] = \int\underbrace{(\hat{a}(R)-A-B(A))}_{z_1}\underbrace{\left(\frac{\partial}{\partial A}\log p_{r|a}(R|A)\right)}_{z_2}p_{r|a}(R|A)\,dR \underset{\text{AS ABOVE}}{=} \int(\hat{a}(R)-A-B(A))\frac{\partial}{\partial A}p_{r|a}(R|A)\,dR$$

$\boxed{\text{AGAIN USE } u\,dv = d(uv) - v\,du}$

$$= \int\frac{\partial}{\partial A}[(\hat{a}(R)-A-B(A))p_{r|a}(R|A)]\,dR - \int\left(-1-\frac{dB}{dA}\right)p_{r|a}(R|A)\,dR = \frac{\partial}{\partial A}E[\cancel{\hat{a}(r)-A-B(A)}]^{0} + \left(1+\frac{dB}{dA}\right).$$

PLUG INTO CAUCHY-SCHWARZ ≠: $\left(1+\frac{dB}{dA}\right)^2 \leq Var[\hat{a}(r)]\,I_r(A) \rightarrow$ CRAMER RAO BOUND.

III. FOR ANY ESTIMATOR $\hat{a}(r)$ OF A RANDOM VARIABLE a, LET $I = E[(\frac{\partial}{\partial A}\log p_{r,a}(R,A))^2] = -E[\frac{\partial^2}{\partial A^2}\log p_{r,a}(R,A)]$

$Var[\hat{a}(r) - a] \geq 1/I$ NOTE E OVER BOTH r AND a. DON'T NEED "UNBIASED". WHY? $\frac{\partial B(A)}{\partial A} \neq 0$ BUT $\partial E(e)/\partial A = 0$.

PROOF: LET $z_1 = e - \underbrace{E[e]}_{\text{NUMBER}}$ WHERE $e = \hat{a}(r) - a$. $E[z_1]=0$. $E[z_1^2] = Var[e]$.

$z_2 = \frac{\partial}{\partial A}\log p_{r,a}(R,A)$ = RANDOM VARIABLE IN r AND a. $E[z_2]=0$ SAME ARGUMENT AS ABOVE. $E[z_2^2] = I$ = NUMBER.

NOTE: 1. NOW $p_{r,a}(R,A)$ JOINT pdf 2. E OVER BOTH r AND a, SO 3. I IS ALWAYS A NUMBER.

$$E[z_1z_2] = \iint(e-E(e))\left(\frac{\partial}{\partial A}\log p_{r,a}(R,A)\right)p_{r,a}(R,A)\,dR\,dA \underset{\text{AS ABOVE}}{=} \iint(e-E(e))\frac{\partial}{\partial A}p_{r,a}(R,A)\,dR\,dA$$

$\boxed{\text{AGAIN USE } u\,dv = d(uv) - v\,du}$

$$= \iint\frac{\partial}{\partial A}[(e-E(e))p_{r,a}(R,A)]\,dR\,dA - \iint(-1)p_{r,a}(R,A)\,dR\,dA = \left[\int(e-E(e))p_{r,a}(R,A)\,dR\right]\Big|_{A=-\infty}^{A=\infty} + 1 \text{ (INTEGRATE BY PARTS)}$$

REMEMBER $E[e]$ IS A NUMBER, NOT A FUNCTION OF A AS IN II

$$= E[e-E(e)|a=A]\,p_a(A)\Big|_{A=-\infty}^{A=\infty} + 1 = 1.$$

PLUG INTO CAUCHY-SCHWARZ ≠. (NEED $\lim_{|A|\to\infty} p_a(A) = 0$)

$\boxed{p_{r,a}(R,A) = p_{r|a}(R|A)\,p_a(A)}$

NOTE: $I_r(A) = \int\left(\frac{\partial}{\partial A}\log p_{r|a}(R|A)\right)^2 p_{r|a}(R|A)\,dR = \frac{\partial}{\partial A}\log p_{r|a}(R|A)\,p_{r|a}(R|A)\Big|_{-\infty}^{\infty} - \int\frac{\partial^2}{\partial A^2}\log p_{r|a}(R|A)\,p_{r|a}(R|A)$

INTEGRATION BY PARTS: $u = \frac{\partial}{\partial A}\log p_{r|a}(R|A)$, $dv = \left(\frac{\partial}{\partial A}\log p_{r|a}(R|A)\right)p_{r|a}(R|A) = \frac{\partial}{\partial A}p_{r|a}(R|A)$: $u\,dv = uv - \dots$

DEF: AN ESTIMATOR WHICH ATTAINS THE CRAMER-RAO BOUND (IN A PROBLEM) IS EFFICIENT.

Q: DOES AN EFFICIENT ESTIMATOR EXIST FOR A GIVEN PROBLEM?

A: SOMETIMES. THE ONLY ESTIMATOR THAT CAN BE EFFICIENT IS THE MLE: (UNLESS ANOTHER ESTIMATOR = MLE)

EQUALITY IN CAUCHY SCHWARZ: $z_2 = k(A) z_1 \rightarrow \frac{\partial}{\partial A} \log p_{r|a}(R|A) = k(A)(\hat{a}(R) - A)$ FOR ALL A. SET $A = \hat{a}(R)$! QED.

TO CHECK WHETHER $\hat{a}_{MLE}(R)$ IS EFFICIENT: 2 METHODS:

(1) COMPUTE $E[(\hat{a}_{MLE} - A)^2]$, $I_r(A)$, AND SEE IF $E[(\hat{a}_{MLE} - A)^2] = 1/I_r(A)$.

(2) COMPUTE $k(A) \overset{?}{=} \dfrac{\frac{\partial}{\partial A} \log p_{r|a}(R|A)}{\hat{a}_{MLE}(R) - A}$ ← NEED FOR $\hat{a}_{MLE}$ ANYWAY AND SEE IF IT'S INDEPENDENT OF R. [USUALLY EASIER THAN (1)].

IF SO, $\hat{a}_{MLE}(R)$ IS EFFICIENT (RATIO IS IN FACT $\pm I_r(A)$). IF NOT, THERE IS NO EFFICIENT ESTIMATOR.

Q: WHAT IF a IS A RANDOM VARIABLE? $z_2 = k z_1 \rightarrow \frac{\partial}{\partial A} \log p_{r,a}(R,A) = k(\hat{a}(R) - A - E(e)) \rightarrow$

A: AN EFFICIENT ESTIMATOR EXISTS IFF $p_{a|r}(A|R) = \dfrac{p_{r|a}(R|A)\, p_a(A)}{\int p_{r|a}(R|A')\, p_a(A')\, dA'}$ IS GAUSSIAN. $p_{r,a}(R,A) \sim e^{-k\int A\, dA} = e^{-kA^2/2}$

THEN $\hat{a}_{MAP}(R) = \hat{a}_{LEAST SQUARES}(R)$ IS EFFICIENT.
(EASIER TO COMPUTE)

CONSIDER THE PROBLEM $r = s(A) + n$ WHERE $n \sim N(0, \sigma^2)$ AND $s(A)$ = NONLINEAR FUNCTION OF A.

IN GENERAL, $\hat{a}_{MLE}(R)$ IS NOT EFFICIENT, BUT IF σ^2 IS SMALL $\hat{a}_{MLE}(R)$ ATTAINS C-R BOUND!!

WHY? 2 DERIVATIONS:

I. $r = s(A) + n$ AND $n \sim N(0, \sigma^2) \rightarrow r \sim N(s(A), \sigma^2) \rightarrow p_{r|a}(R|A) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(R - s(A))^2/2\sigma^2}$

$\rightarrow I_r(A) = E\left[\left(\frac{\partial}{\partial A} \log p_{r|a}(R|A)\right)^2\right] = E\left[\left(-\frac{-2}{2\sigma^2} \frac{ds(A)}{dA}(r - s(A))\right)^2\right] = \frac{1}{\sigma^4}\left(\frac{ds(A)}{dA}\right)^2 E\left[(r - s(A))^2\right]$

$\hat{a}_{MLE}(R)$ IS SUCH THAT $R = s(\hat{a}_{MLE}(R))$, i.e., $\hat{a}_{MLE}(R) = s^{-1}(R)$. (INVERSE FUNCTION) $= \left(\frac{ds}{dA}\right)^2 / \sigma^2$.

KEY POINT: σ^2 SMALL $\rightarrow \hat{a}_{MLE} \approx A \rightarrow r = s(A) + n \doteq s(\hat{a}_{MLE}) + \frac{ds}{dA}\Big|_{\hat{a}_{MLE}} (A - \hat{a}_{MLE}) + n$

$\boxed{\text{ASYMPTOTIC EFFICIENCY OF } \hat{a}_{MLE}(R)}$

$- \quad r = s(\hat{a}_{MLE})$

$0 = \frac{ds}{dA}\Big|_{\hat{a}_{MLE}} (A - \hat{a}_{MLE}) + n \rightarrow (\hat{a}_{MLE} - A) = \dfrac{n}{\frac{ds}{dA}\Big|_{\hat{a}_{MLE}}}$

$\rightarrow E[(\hat{a}_{MLE} - A)^2] = \dfrac{E[n^2]}{\left(\frac{ds}{dA}\right)^2_{\hat{a}_{MLE}}} = \dfrac{\sigma^2}{\left(\frac{ds}{dA}\right)^2_{\hat{a}_{MLE}}} \geq 1/I_r(A) = \dfrac{\sigma^2}{\left(\frac{ds}{dA}\right)^2}$ $\therefore$ EQUALITY IF $\frac{ds}{dA} \approx \frac{ds}{dA}\Big|_{\hat{a}_{MLE}}$. $(A \approx \hat{a}_{MLE})$

II. σ^2 SMALL $\rightarrow$ KNOW A APPROXIMATELY, e.g. FROM $s^{-1}(R)$. LET $A \approx A_{APPROX}$ = KNOWN.

THEN $r = s(A) + n \doteq s(A_{APPROX}) + \frac{ds}{dA}\Big|_{A_{APPROX}} (A - A_{APPROX}) + n$. LET $\underbrace{r'}_{\text{RANDOM VARIABLES}} = r - \underbrace{s(A_{APPROX}) + \frac{ds}{dA} A_{APPROX}}_{\text{KNOWN}}$ (NOT OBSERVATION R - YET)

$\rightarrow r' = \underbrace{\left(\frac{ds}{dA}\Big|_{A_{APPROX}}\right)}_{\text{KNOWN CONSTANT}} A + n$ LINEAR PROBLEM.

EASY TO SHOW $\hat{a}_{MLE}(R') = R' / \left(\frac{ds}{dA}\right)\Big|_{A_{APPROX}}$ IS EFFICIENT

THM: MLE COMMUTES WITH NONLINEAR OPS $\leftrightarrow [f(\hat{A})]_{MLE}(R) = f(\hat{A}_{MLE}(R)) \leftrightarrow$ IF $B = f(A)$, THEN $\hat{B} = f(\hat{A})$.

PROOF: $\hat{B}_{MLE}(R) = \underset{B}{\mathrm{ARGMAX}}\, p_{r|b}(R|B) = \underset{B}{\mathrm{ARGMAX}}\, p_{r|f(a)}(R|B)$. ALSO, $p_{r|a}(R|A) = p_{r|f(a)}(R|f(A))$ [GIVEN a=A!]

$\therefore$ IF $\hat{A}$ MAXIMIZES $p_{r|a}(R|A)$, $f(\hat{A})$ MAXIMIZES $p_{r|f(a)}(R|f(A))$.

EECS 564 DERIVATION OF THE MULTICHANNEL (VECTOR) CRAMER-RAO BOUND WINTER 1999

DEF: IF A IS A SYMMETRIC SQUARE MATRIX, "$A \geq 0$" $\leftrightarrow$ $\underline{x}^T A \underline{x} \geq 0$ FOR ANY VECTOR $\underline{x}$

$\leftrightarrow$ EIGENVALUES OF A ARE ALL ≥ 0.

DEF: "$A \geq B$" MEANS $A - B \geq 0$.

DEF: "$A^{1/2}$" MEANS THE MATRIX SUCH THAT $(A^{1/2})(A^{1/2})^T = A$ (UMOVE TO MULT. BY UNITARY MATRIX).

NOTE: $A \geq 0 \rightarrow BAB^T \geq 0$ FOR ANY MATRIX B: $\underline{x}^T(BAB^T)\underline{x} = (B^T\underline{x})^T A (B^T\underline{x}) \geq 0$. $A^{1/2}$ HAS EIGVAL $\sqrt{\text{EIGVAL}(A)}$.

MATRIX CAUCHY-SCHWARZ INEQUALITY: $E[AB^T]\,E[BB^T]^{-1}\,E[BA^T] \leq E[AA^T]$

IF A, B ARE RANDOM MATRICES (OR VECTORS) SUCH THAT $E[A] = E[B] = \underline{0}$.

PROOF: CONSIDER NORMALIZED RANDOM MATRICES $\tilde{A} = \underbrace{E[AA^T]^{-1/2}}_{\text{CONSTANT MATRIX}} A$, $\tilde{B} = \underbrace{E[BB^T]^{-1/2}}_{\text{CONSTANT MATRIX}} B$.

THEN
$$0 \underset{\text{(0-MEAN)}}{\leq} E\left[\begin{bmatrix}\tilde{A}\\ \tilde{B}\end{bmatrix}\begin{bmatrix}\tilde{A}^T & \tilde{B}^T\end{bmatrix}\right] = \begin{bmatrix} I & E(\tilde{A}\tilde{B}^T) \\ E(\tilde{A}\tilde{B}^T)^T & I\end{bmatrix} = \begin{bmatrix} I & \underline{\rho} \\ \underline{\rho}^T & I\end{bmatrix}$$
WHERE $\underline{\rho} = E(\tilde{A}\tilde{B}^T)$ $\left(= E(AA^T)^{-1/2} E(AB^T)\, E(BB^T)^{-1/2}\right)$

$$= \underset{\text{(TRY IT!)}}{\begin{bmatrix} I & \underline{\rho} \\ 0 & I\end{bmatrix}\begin{bmatrix} I - \underline{\rho}\underline{\rho}^T & 0 \\ 0 & I\end{bmatrix}\begin{bmatrix} I & \underline{\rho} \\ 0 & I\end{bmatrix}^T} \rightarrow I - \underline{\rho}\underline{\rho}^T \geq 0 \rightarrow \underline{\rho}\underline{\rho}^T \leq I.$$
INSERT $\uparrow$, PRE- AND POST-MULT. BY $E(AA^T)^{1/2}$.

USE THIS TO DERIVE VECTOR CRAMER-RAO BOUND:

LET $\underline{z}_1 = \underline{\hat{a}}(\underline{R}) - \underline{A}$ AND $\underline{z}_2 = \underset{\text{MAKES VECTOR}}{\nabla_A} \underbrace{\log p_{\underline{r}|\underline{a}}(\underline{R}|\underline{A})}_{\text{SCALAR}}$ WHERE $\nabla_A f(A) = \begin{bmatrix} \frac{\partial f}{\partial A_1} \\ \vdots \\ \frac{\partial f}{\partial A_n}\end{bmatrix}$ IF $A = \begin{bmatrix} A_1 \\ \vdots \\ A_n \end{bmatrix}$

NOTE $\nabla_A \log p_{\underline{r}|\underline{a}}(\underline{R}|\underline{A}) = \dfrac{1}{p_{\underline{r}|\underline{a}}(\underline{R}|\underline{A})} \nabla_A p_{\underline{r}|\underline{a}}(\underline{R}|\underline{A})$ (*).

PLUG INTO MATRIX CAUCHY-SCHWARZ INEQUALITY:

MEANS: $E(\underline{z}_1) = E(\underline{\hat{a}}(\underline{R})) - \underline{A} = 0$ (UNBIASED BY ASSUMPTION).

$$E(\underline{z}_2) = \int \underline{z}_2\, p_{\underline{r}|\underline{a}}(\underline{R}|\underline{A})\,d\underline{R} = \int \nabla_A \log p_{\underline{r}|\underline{a}}(\underline{R}|\underline{A})\, p_{\underline{r}|\underline{a}}(\underline{R}|\underline{A})\,d\underline{R} \overset{\text{USE (*)}}{=} \int \frac{p_{\underline{r}|\underline{a}}(\underline{R}|\underline{A})}{p_{\underline{r}|\underline{a}}(\underline{R}|\underline{A})} \nabla_A p_{\underline{r}|\underline{a}}(\underline{R}|\underline{A})\,d\underline{R}$$
$$\underset{\text{LEIBNIZ'S RULE}}{=} \nabla_A \int p_{\underline{r}|\underline{a}}(\underline{R}|\underline{A})\,d\underline{R} = \nabla_A(1) = 0.$$

SECOND MOMENTS: $E(\underline{z}_1\underline{z}_1^T) = E(\underline{\varepsilon}\underline{\varepsilon}^T)$ SINCE $\underline{\varepsilon} = \underline{\hat{a}}(\underline{R}) - \underline{A} = \underline{z}_1$.

$E(\underline{z}_2\underline{z}_2^T) = E(LL^T) \underset{\text{(FISHER INFO. MATRIX)}}{=} J$ SINCE $L = \nabla_A \log p_{\underline{r}|\underline{a}}(\underline{R}|\underline{A}) = \underline{z}_2$.

MIXED MOMENT: $E(\underline{z}_1\underline{z}_2^T) = E\left((\underline{\hat{a}}(\underline{R}) - \underline{A})\nabla_A^T \log p_{\underline{r}|\underline{a}}(\underline{R}|\underline{A})\right) = \int \underbrace{(\underline{\hat{a}}(\underline{R}) - \underline{A})}_{\underline{z}_1} \underbrace{\nabla_A^T \log p_{\underline{r}|\underline{a}}(\underline{R}|\underline{A})}_{\underline{z}_2}\, p_{\underline{r}|\underline{a}}(\underline{R}|\underline{A})\,d\underline{R}$

$$\underset{\text{USE (*)}}{=} \int \underbrace{(\underline{\hat{a}}(\underline{R}) - \underline{A})}_{f} \nabla_A^T \underbrace{p_{\underline{r}|\underline{a}}(\underline{R}|\underline{A})}_{g}\,d\underline{R} = \int \nabla_A^T\Big[\underbrace{(\underline{\hat{a}}(\underline{R}) - \underline{A})}_{f}\underbrace{p_{\underline{r}|\underline{a}}(\underline{R}|\underline{A})}_{g}\Big]d\underline{R} - \int \underbrace{\Big(\nabla_A^T(\underline{\hat{a}}(\underline{R}) - \underline{A})\Big)}_{f} \underbrace{p_{\underline{r}|\underline{a}}(\underline{R}|\underline{A})}_{g}\,d\underline{R}$$
USE $\nabla_A(fg) = (\nabla_A f)g + f(\nabla_A g)$

$$= \nabla_A^T E[\underline{\hat{a}}(\underline{R}) - \underline{A}] - \int (-I)\, p_{\underline{r}|\underline{a}}(\underline{R}|\underline{A})\,d\underline{R} \underset{\text{UNBIASED}}{=} 0 + I \cdot 1 = I. \quad E[\underline{z}_2\underline{z}_1^T] = I^T = I.$$
$\nabla_A^T(\underline{A}) = I$ (TRY IT!)

PLUG IN: $E[\underline{z}_1\underline{z}_2^T]\,E[\underline{z}_2\underline{z}_2^T]^{-1}\,E[\underline{z}_2\underline{z}_1^T] \leq E(\underline{z}_1\underline{z}_1^T) \rightarrow I J^{-1} I \leq E[\underline{\varepsilon}\underline{\varepsilon}^T]$

$\rightarrow J^{-1} \leq E[\underline{\varepsilon}\underline{\varepsilon}^T]$ QED. IN PARTICULAR: PRE- AND POST-MULT. BY $[0 \cdots 0 1 0 \cdots 0]$ TO GET VAN TREES RESULT.

HANDOUT
EXAMPLE OF VECTOR C-R BOUND

EXAMPLE: OBSERVE $\underline{Y} = H\underline{X} + \underline{V}$. $\underline{V} \sim N(\underline{0}, R)$. $\rightarrow Y \sim N(H\underline{X}, R)$.
COMPUTE $\hat{X}_{MLE}(\underline{Y})$ AND SHOW EFFICIENT (C-R BOUND).

LOG-LIKELIHOOD FUNC $= \log p_{\underline{Y}|\underline{X}}(\underline{Y}|\underline{X}) = \log\left[\frac{1}{(2\pi)^{n/2}|R|^{1/2}} e^{-(\underline{Y}-H\underline{X})^T R^{-1}(\underline{Y}-H\underline{X})/2}\right]$

$= -\frac{n}{2}\log 2\pi - \frac{1}{2}\log|R| - \underbrace{(\underline{Y}-H\underline{X})^T R^{-1}(\underline{Y}-H\underline{X})/2}_{\text{MAXIMIZE}}$. (VECTOR CALCULUS: STARK-WOODS p. 289)

$\nabla_{\underline{X}}\left[(\underline{Y}-H\underline{X})^T R^{-1}(\underline{Y}-H\underline{X})\right]/2 = H^T R^{-1}(\underline{Y}-H\underline{X}) = 0$

$\rightarrow \underline{\hat{X} = (H^T R^{-1} H)^{-1}(H^T R^{-1})\underline{Y}}$ PSEUDOINVERSE.

C-R BOUND: $J_{\underline{Y}}(\underline{a}) = E(LL^T]$, $L = \nabla_{\underline{X}} \log p_{\underline{Y}|\underline{X}}(\underline{Y}|\underline{X}) = H^T R^{-1}(\underline{Y}-H\underline{X})$
$= H^T R^{-1}\underline{V}$.

$J_{\underline{Y}}(a) = E(LL^T) = E\left[H^T R^{-1}\underline{V}\,\underline{V}^T R^{-1} H\right] = H^T R^{-1}(R)R^{-1}H = \underline{H^T R^{-1} H}$.

EFFICIENT?

$E(\hat{\underline{X}}] = (H^T R^{-1} H)^{-1}(H^T R^{-1})E(\underline{Y}) = (H^T R^{-1} H)^{-1}(H^T R^{-1})(H\underline{X}) = \underline{X}$. UNBIASED.

$\underline{e} = \underline{X} - \hat{\underline{X}} = \underline{X} - (H^T R^{-1} H)^{-1}(H^T R^{-1})(H\underline{X}+\underline{V}) = \underline{X} - I\underline{X} - (H^T R^{-1} H)^{-1}(H^T R^{-1})\underline{V}$.

$E[(\underline{X}-\hat{\underline{X}})(\underline{X}-\hat{\underline{X}})^T] = E[(H^T R^{-1} H)^{-1}(H^T R^{-1})\underline{V}\,\underline{V}^T(R^{-1}H)(H^T R^{-1} H)^{-1}]$

$= (H^T R^{-1} H)^{-1}(H^T R^{-1})R(R^{-1}H)(H^T R^{-1} H)^{-1} = \underline{(H^T R^{-1} H)^{-1}}$.

$\therefore E[\underline{e}\,\underline{e}^T] = J_{\underline{Y}}(a)^{-1}$ AND MLE IS EFFICIENT. NOT SURPRISING: LINEAR PROBLEM.

NOTE THAT $\underline{e} = -(H^T R^{-1} H)^{-1}(H^T R^{-1})\underline{V}$ IS GAUSSIAN.

$\therefore$ CAN CONSIDER QUESTIONS SUCH AS, "WHAT IS $Pr[\|\underline{e}\| > d]$"?

SOL'N: $Pr(\|\underline{e}\| > d] = Pr(\|\underline{e}\|^2 > d^2]$, AND $\|e\|^2$ IS SUM OF SQUARES OF 2 GAUSSIANS = EXPONENTIAL.

CONTOURS OF EQUAL PROBABILITY OF GIVEN VALUE OF $\underline{e}$ ARE ELLIPSES.

EECS 564 K-L EXPANSION PROPERTIES WINTER 1999

K-L EXPANSION: LET $X(t)$ BE A CONT.-TIME RANDOM PROCESS WITH 0 MEAN AND FINITE VARIANCE. $E(X(t))=0$ $\sigma^2_{X(t)}<\infty$
THEN, OVER THE FINITE INTERVAL $[T_i,T_f]$, $X(t)$ CAN BE EXPANDED AS:

$X(t) = \sum_{i=1}^{\infty} x_i \phi_i(t)$, $T_i \le t \le T_f$, WHERE $x_i = \int_{T_i}^{T_f} X(t)\phi_i(t)dt = \boxed{\text{PROJECTION}}$ OF $X(t)$ ON $\phi_i(t)$.

$X(t)$: RANDOM PROCESS; x_i: UNCORRELATED RANDOM VARIABLES; $\phi_i(t)$: ORTHOGONAL DETERMINISTIC BASIS FUNCTIONS; x_i: RANDOM VARIABLE; $X(t)$: RANDOM PROCESS. WILL HEAR THIS WORD A LOT!

PROPERTIES OF K-L EXPANSION: | ANALOG IN FINITE-DIMENSIONS

(1) $\{\phi_i(t)\}$ ORTHONORMAL: $\int_{T_i}^{T_f}\phi_i(t)\phi_j(t)dt = \delta_{ij}$ | $Q^TQ = I$

(2) $\{\phi_i(t)\}$ COMPLETE ORTHON. SET: $\sum_{i=1}^{N} x_i\phi_i(t) \xrightarrow[\text{L.I.M.}]{\text{MEAN SQ}} X(t)$

$\lim_{N\to\infty} E\left[\left(X(t)-\sum_{i=1}^{N}x_i\phi_i(t)\right)^2\right]=0$, $T_i\le t\le T_f$. | DOES NOT ARISE (MAJOR DIFF.)

(3) $\{x_i\}$ UNCORRELATED (OR WHY BOTHER?):

$E[x_ix_j] = \int_{T_i}^{T_f}\int_{T_i}^{T_f}\phi_i(t_1)K_X(t_1,t_2)\phi_j(t_2)dt_1dt_2 = \lambda_i\delta_{ij}$

x_i NEED NOT HAVE SAME VARIANCES

WANT K_Y DIAGONAL

WHERE CAN WE FIND SUCH A $\{\phi_i(t)\}$?

SOLVE FREDHOLM INTEGRAL EQN

$$\boxed{\int_{T_i}^{T_f} K_X(t_1,t_2)\phi_j(t_2)dt_2 = \lambda_j\phi_j(t_1),\quad T_i<t_1<T_f.}$$

$K_X\underline{\phi}_i = \lambda_i\underline{\phi}_i$ FIND EIGVAL λ_i EIGVEC ϕ_i OF K_X MATRIX

PROOFS: WHY THIS INTEGRAL EQN?

I **NECESSARY**: IF $X(t)=\sum_{i=1}^{\infty}x_i\phi_i(t)$ AND $x_i = \int_{T_i}^{T_f}X(t)\phi_i(t)dt$, $\{x_i\}$ UNCORRELATED,
THEN NECESSARILY $\int_{T_i}^{T_f}K(t_1,t_2)\phi_i(t_2) = \lambda_i\phi_i(t_1)$, $T_i<t<T_f$.

PROOF: $X(t)=\sum_{i=1}^{\infty}x_i\phi_i(t) \xrightarrow[\text{TAKE } E(\cdot)]{\text{MULT BY } x_j} E[X(t)x_j] = \sum_{i=1}^{\infty}E[x_ix_j]\phi_i(t) = \lambda_j\phi_j(t)$. WE WILL USE THIS RESULT ON OTHER SIDE OF HANDOUT.

$\lambda_j\delta_{ij}$ BY DEFINITION

ALSO, $E[X(t)x_j] = E\left[X(t)\int_{T_i}^{T_f}X(s)\phi_j(s)ds\right] = \int_{T_i}^{T_f}\underbrace{E[X(t)X(s)]}_{K_X(t,s)}\phi_j(s)ds$ MUST BE $\therefore$ = QED THIS IS INTEGRAL EQN.

(USE THIS LINE OF REASONING ON OTHER SIDE OF HANDOUT)

II **SUFFICIENT**: IF SOLVE INTEGRAL EQN, THEN:

(1) IF EIGVAL $\{\lambda_i\}$ DISTINCT, $\{\phi_i(t)\}$ ORTHONORMAL: LET $j=1$ IN INTEGRAL EQN.

MULT. BY $\phi_2(t_1)$ AND $\int_{T_i}^{T_f}dt_1$: $\int_{T_i}^{T_f}\int_{T_i}^{T_f}\phi_2(t_1)K_X(t_1,t_2)\phi_1(t_2)dt_1dt_2 = \int_{T_i}^{T_f}\lambda_1\phi_1(t_1)\phi_2(t_1)$

REARRANGE LEFT SIDE: $\int_{T_i}^{T_f}\phi_1(t_2)dt_2\underbrace{\int_{T_i}^{T_f}K_X(t_1,t_2)\phi_2(t_1)dt_1}_{\lambda_2\phi_2(t_2)} = \int_{T_i}^{T_f}\lambda_2\phi_1(t_2)\phi_2(t_2)dt_2$

SUBTRACT: $(\lambda_1-\lambda_2)\int_{T_i}^{T_f}\phi_1(t)\phi_2(t)dt = 0.$

$\therefore \lambda_1\ne\lambda_2 \to \int_{T_i}^{T_f}\phi_1(t)\phi_2(t)dt=0 \to \phi_1(t),\phi_2(t)$ ORTHOGONAL. QED.

HINT: CONSIDER PROOF FOR MATRIX: $\lambda_i\ne\lambda_j \to \underline{\phi}_i^T\underline{\phi}_j=0$. $\boxed{QQ^T=I}$

(2) $\{x_i\}$ UNCORRELATED: $x_i = \int_{T_i}^{T_f}X(t)\phi_i(t)dt$

$\boxed{K_Y \text{ DIAGONAL}}$

$E[x_ix_j] = \int_{T_i}^{T_f}\int_{T_i}^{T_f}\phi_i(t_1)\underbrace{K_X(t_1,t_2)\phi_j(t_2)dt_2}_{\text{USE INTEG EQN}}dt_1 = \int_{T_i}^{T_f}\phi_i(t_1)\lambda_j\phi_j(t_1)dt_1 = \lambda_j\delta_{ij}$. QED

(3) $\{\phi_i(t)\}$ IS A **COMPLETE** ORTHONORMAL SET, MEANING:

$$\underset{N\to\infty}{\text{L.I.M.}} \sum_{i=1}^{N} x_i\phi_i(t) = x(t) \Longleftrightarrow \lim_{N\to\infty} E[e_N(t)^2]=0 \text{ FOR ALL } T_i \le t \le T_f, \text{ WHERE}$$

$e_N(t) = x(t) - \sum_{i=1}^{N} x_i\phi_i(t)$ = **ERROR** RANDOM PROCESS IN REPRESENTING $x(t)$ BY THE **TRUNCATED** EXPANSION (ONLY N TERMS).

$\boxed{\text{SIGNIFICANCE OF "COMPLETENESS": BY INCLUDING ENOUGH TERMS IN } \sum_{i=1}^{N} x_i\phi_i(t) \text{, CAN APPROXIMATE } x(t) \text{ ARBITRARILY CLOSELY, IN THE MEAN-SQUARE SENSE.}}$

TO SHOW $\{\phi_i(t)\}$ IS A **COMPLETE ORTHONORMAL SET** (IF SOLVE INTEGRAL EQN):

LEMMA 1: $K_{e_N}(t,s) = K_x(t,s) - \sum_{i=1}^{N} \lambda_i\phi_i(t)\phi_i(s) \underset{psd}{\geq 0}$ (SINCE $K_{e_N}(t,s)$ IS A COVARIANCE FUNCTION).

PROOF: $K_{e_N}(t,s) = E[e_N(t)e_N(s)] = E\left[\left(x(t)-\sum_{i=1}^{N}x_i\phi_i(t)\right)\left(x(s)-\sum_{j=1}^{N}x_j\phi_j(s)\right)\right]$ EXPAND INTO 4 TERMS:

$$= \underbrace{E[x(t)x(s)]}_{K_x(t,s)} - \sum_{i=1}^{N}\phi_i(t)\underbrace{E[x_i x(s)]}_{\lambda_i\phi_i(s)} - \sum_{j=1}^{N}\phi_j(s)\underbrace{E[x_j x(t)]}_{\lambda_j\phi_j(t)} + \sum_{i=1}^{N}\sum_{j=1}^{N}\underbrace{E[x_i x_j]}_{\lambda_i\delta_{ij}}\phi_i(t)\phi_j(s) = K_x(t,s) - \sum_{i=1}^{N}\lambda_i\phi_i(t)\phi_i(s)$$

SEE OTHER SIDE OF HANDOUT; PROVED IN (2) ABOVE THAT $E(x_i x_j)=\lambda_i\delta_{ij}$ [2 MINUS, 1 PLUS → 1 MINUS TOTAL]

NOW SET $s=t$: $0 \le \sigma^2_{e_N(t)} = K_{e_N}(t,t) \underset{\text{LEMMA 1}}{=} K_x(t,t) - \sum_{i=1}^{N}\lambda_i\phi_i(t)^2 = \underbrace{\sigma^2_{x(t)}}_{\text{FINITE BY ASSUMPTION}} - \sum_{i=1}^{N}\underbrace{\lambda_i\phi_i(t)^2}_{\geq 0 \text{ SINCE } \lambda_i=\sigma^2_{x_i}=E(x_i^2)} \geq 0$

(**ALTERNATIVE DERIVATION:** $x(t) = \underbrace{e_N(t)}_{\sum_{i=N+1}^{\infty}x_i\phi_i(t)} + \sum_{i=1}^{N}x_i\phi_i(t)$. SINCE THE $\{x_i\}$ ARE UNCORRELATED AND THE VARIANCE OF A SUM OF UNCORRELATED RVS IS THE SUM OF THE VARIANCES $\rightarrow \sigma^2_{x(t)} = \sigma^2_{e_N(t)} + \sum_{i=1}^{N}\underbrace{\sigma^2_{x(i)}}_{\lambda_i}\underbrace{\phi_i(t)^2}_{\text{SCALE FACTOR SQUARED}}$)

[THE SEQUENCE OF PARTIAL SUMS $\left\{\sum_{i=1}^{N}\lambda_i\phi_i(t)^2\right\}$ IS **INCREASING** (SINCE $\lambda_i\phi_i(t)^2 \geq 0$) AND **BOUNDED** (BY $\sigma^2_{x(t)} < \infty$ **BY ASSUMPTION** (SEE FIRST LINE OF THIS HANDOUT)). ∴ IT **CONVERGES**.
∴ NECESSARY (NOT SUFFICIENT) CONDITION FOR CONVERGENCE IS $\lim_{i\to\infty}\lambda_i = 0$.]

EIGENVALUES/EIGENFUNCTIONS OF $K_{e_N}(t,s)$ ARE $\{\lambda_{N+1}, \lambda_{N+2}, \ldots\}$ / $\{\phi_{N+1}(t), \phi_{N+2}(t), \ldots\}$:

$$\int_{T_i}^{T_f} K_{e_N}(t,s)\phi_{N+k}(s)\,ds = \int_{T_i}^{T_f} K_x(t,s)\phi_{N+k}(s)\,ds - \int_{T_i}^{T_f}\sum_{i=1}^{N}\lambda_i\phi_i(t)\underbrace{\phi_i(s)\phi_{N+k}(s)}_{\text{NOTE } 1\le i\le N}ds = \lambda_{N+k}\phi_{N+k}(t) - 0.$$

(SEE 2ND LINE OF THIS SIDE OF HANDOUT)

LET $\lim_{N\to\infty} K_{e_N}(t,s) = K_{e_\infty}(t,s) < \sigma^2_{x(t)} < \infty$ SO $K_{e_\infty}(t,s)$ EXISTS.
BUT EIGENVALUES OF $K_{e_\infty}(t,s)$ ARE ALL 0, SINCE $\lim_{N\to\infty}\lambda_N = 0$!
ALL EIGENVALUES OF $K_{e_\infty}(t,s)$ ARE $0 \rightarrow K_{e_\infty}(t,s) = 0 \rightarrow \sigma^2_{e_\infty(t)} = K_{e_\infty}(t,t) = \lim_{N\to\infty}\sigma^2_{e_N(t)} = \lim_{N\to\infty}E[e_N(t)^2] = 0$.
(EXISTENCE THM)

[**ANOTHER WAY TO LOOK AT IT:** $x(t) = \sum_{i=1}^{\infty}x_i\phi_i(t) \rightarrow \sigma^2_{x(t)} = E[x(t)^2] \overset{\text{(A BIT IFFY HERE)}}{=} \sum_{i=1}^{\infty}\sigma^2_{x_i}\phi_i(t)^2 = \sum_{i=1}^{\infty}\lambda_i\phi_i(t)^2$ IS FINITE.
MEAN SQUARE ERROR $E[e_N(t)^2]$ OF TRUNCATED SERIES IS $\sum_{i=N+1}^{\infty}\lambda_i\phi_i(t)^2 \rightarrow 0$ AS $N\to\infty$. (FINITE SINCE $\int\phi_i(t)^2dt = 1$)]

MERCER'S THM: $\lim_{N\to\infty} K_{e_N}(t,s) = K_x(t,s) - \sum_{i=1}^{\infty}\lambda_i\phi_i(t)\phi_i(s) = 0 \longrightarrow K_x(t,s) = \sum_{i=1}^{\infty}\lambda_i\phi_i(t)\phi_i(s)$.

(1) THIS IS A FAMOUS RESULT IN THEORY OF INTEGRAL EQNS. USUALLY (e.g., VAN TREES) USE MERCER'S THM TO PROVE COMPLETENESS. P. 182

(2) WE WILL USE MERCER'S THM FOR DETECTION AND ESTIMATION IN COLORED (NONWHITE) GAUSSIAN NOISE.

(3) SET $s=t$ AND $\int_{T_i}^{T_f}dt$: EXPECTED TOTAL ENERGY IN (T_i, T_f) $= \int_{T_i}^{T_f}E[x(t)^2]dt = E\left[\int_{T_i}^{T_f}x(t)^2dt\right] = \int_{T_i}^{T_f}K_x(t,t)dt = \sum_{i=1}^{\infty}\lambda_i$ = TRACE OF $K_x(t,s)$