

**SIMON FRASER
UNIVERSITY**
School of Engineering Science

ENSC 810-3 Signal Processing

Solutions to Assignment 2

Semester 00-1

1. Maximum Likelihood and Cramer-Rao

(a) Define $\mathbf{x}$ and $\mathbf{n}$ as the N -vectors of measurements and noise, respectively. Then the correlation matrix is

$$\mathbf{R}_{\mathbf{x}} = E(\mathbf{x} \cdot \mathbf{x}^T) = \theta^2 \cdot \sigma_n^2 \cdot \mathbf{I} \qquad \mathbf{R}_{\mathbf{x}}^{-1} = \theta^{-2} \cdot \sigma_n^{-2} \cdot \mathbf{I}$$

The likelihood is just the conditional probability. From Section 1.2.3 of the notes

$$p(\mathbf{x} | \theta) = \frac{1}{\pi^N \cdot \theta^{2 \cdot N} \cdot \sigma_n^{2 \cdot N}} \cdot \exp \left[\frac{-(\|\mathbf{x}\|)^2}{\theta^2 \cdot \sigma_n^2} \right]$$

and the log-likelihood is

$$L(\mathbf{x}, \theta) = \frac{-(\|\mathbf{x}\|)^2}{\theta^2 \cdot \sigma_n^2} - 2 \cdot N \cdot \ln(\theta) - 2 \cdot N \cdot \ln(\sigma_n) \quad \text{after dropping the } 2\pi \text{ factor}$$

(b) The score is

$$V(\mathbf{x}, \theta) = \frac{d}{d\theta} L(\mathbf{x}, \theta) = \frac{2(\|\mathbf{x}\|)^2}{\theta^3 \cdot \sigma_n^2} - \frac{2 \cdot N}{\theta}$$

The mean value is taken over the $\mathbf{x}$ ensemble (equivalently the $\mathbf{n}$ ensemble).

$$E(V(\mathbf{x}, \theta)) = E \left[\frac{2(\|\mathbf{x}\|)^2}{\theta^3 \cdot \sigma_n^2} - \frac{2 \cdot N}{\theta} \right] = \frac{2 \cdot N \cdot \theta^2 \cdot \sigma_n^2}{\theta^3 \cdot \sigma_n^2} - \frac{2 \cdot N}{\theta} = 0$$

It equals zero, as required.

(c) The score is real, with zero mean, so we can write its variance as

$$E(V(\mathbf{x}, \theta)^2) = E \left[\left[\frac{2(\|\mathbf{x}\|)^2}{\theta^3 \cdot \sigma_n^2} - \frac{2 \cdot N}{\theta} \right]^2 \right]$$

$$J = \frac{4}{\theta^6 \cdot \sigma_n^4} \cdot E[(\|\mathbf{x}\|)^4] - \frac{8 \cdot N \cdot E[(\|\mathbf{x}\|)^2]}{\theta^4 \cdot \sigma_n^2} + \frac{4 \cdot N^2}{\theta^2} = J$$

The expectation in the middle term is easy - it's just $N \cdot \theta^2 \cdot \sigma_n^2$

The expectation in the first term can be written

$$E[(\|\mathbf{x}\|)^4] = E \left[\sum_{i=1}^N \sum_{j=1}^N (\|x_i\|)^2 \cdot (\|x_j\|)^2 \right]$$

Recall that the measurements are independent. Therefore the $N(N-1)$ terms in which $i \neq j$ are all

$$E[(\|x_i\|)^2] \cdot E[(\|x_j\|)^2] = \theta^4 \cdot \sigma_n^4$$

and the N terms in which $i=j$ are obtained from the question sheet appendix as

$$E[(\|x_i\|)^4] = 2 \cdot \theta^4 \cdot \sigma_n^4$$

Substitution into the score variance expression above gives

$$E(V(\mathbf{x}, \theta)^2) = \frac{4}{\theta^6 \cdot \sigma_n^4} \left[N \cdot (N-1) \cdot \theta^4 \cdot \sigma_n^4 + N \cdot 2 \cdot \theta^4 \cdot \sigma_n^4 \right] - \frac{8 \cdot N}{\theta^4 \cdot \sigma_n^2} \cdot N \cdot \theta^2 \cdot \sigma_n^2 + \frac{4 \cdot N}{\theta^2}$$

which simplifies to

$$J = \frac{4 \cdot N}{\theta^2}$$

(d) The maximum likelihood estimator is the $\theta(\mathbf{x})$ that makes the score equal zero. From

$$V(\mathbf{x}, \theta) = \frac{2(\|\mathbf{x}\|)^2}{\theta_{\text{hat}}^3 \cdot \sigma_n^2} - \frac{2 \cdot N}{\theta} = 0 \quad \text{we have} \quad \theta_{\text{hat}}^2 = \frac{(\|\mathbf{x}\|)^2}{N \cdot \sigma_n^2}$$

and since θ is non-negative real

$$\theta_{\text{hat}} = \frac{|\mathbf{x}|}{\sqrt{N \cdot \sigma_n^2}} = \frac{\sqrt{\sum_{i=1}^N (|x_i|)^2}}{\sqrt{N \cdot \sigma_n^2}}$$

Is it unbiased? Well, the expectation of θ_{hat}^2 is θ^2 , so we have an unbiased estimator for θ^2 . That should give us a sinking feeling about the estimator for θ itself, which we now examine. For $N > 1$, it's a bit tricky (see below), but the appendix of the question sheet gives the relations we need for $N=1$. The estimate is the magnitude of a complex Gaussian variate with variance $\sigma^2 = \theta^2$. From the appendix, the mean value of the magnitude is

$$\sqrt{\pi} \cdot \frac{\sigma}{2} = \frac{\sqrt{\pi}}{2} \cdot \theta \quad \text{so it is biased for } N=1, \text{ which is enough to disqualify it as an unbiased estimator}$$

Out of curiosity, let's see what we get for larger values of N . First, break each x_i into real and imaginary parts. This gives $2N$ real Gaussian variates that are independent with identical statistics. Then

$$\theta_{\text{hat}} = \sqrt{\sum_{k=1}^{2N} (u_k)^2} \quad \text{where all the real } u_k^2 \text{ have variance } \sigma^2 = \theta^2 / 2N$$

Clearly, θ_{hat}^2 has a χ^2 pdf with $2N$ degrees of freedom. If we denote it by y , then

$$p_y(y) = \frac{1}{2^N \cdot \sigma^{2 \cdot N} \cdot \Gamma(N)} \cdot y^N \cdot \exp\left(\frac{-y}{2 \cdot \sigma^2}\right) \quad (\chi^2, 2N \text{ d.f.})$$

Straightforward transformation to $\theta_{\text{hat}} = \sqrt{y}$ gives the generalized Rayleigh pdf

$$p_{\theta'}(\theta') = \frac{\theta'^{2 \cdot N - 1}}{2^{N-1} \cdot \sigma^{2 \cdot N} \cdot \Gamma(N)} \cdot \exp\left(\frac{-\theta'^2}{2 \cdot \sigma^2}\right) \quad \text{where } \theta' \text{ is } \theta_{\text{hat}} \text{ to save space}$$

The mean value can be obtained by repeated integration by parts, giving

$$\mu_{\theta'} = \sqrt{2 \cdot \sigma^2} \cdot \frac{\Gamma\left(N + \frac{1}{2}\right)}{(N-1)!} = \theta \cdot \frac{\Gamma\left(N + \frac{1}{2}\right)}{\sqrt{N} \cdot (N-1)!}$$

Evidently, the mean of θ_{hat} does not equal θ , so the estimator is biased.

(e) The Cramer-Rao bound states that the lowest error variance an unbiased estimator can achieve is just J^{-1} . where J was obtained in part (c) above. Consequently,

$$\sigma_e^2 \geq \frac{\theta^2}{4 \cdot N} \quad \text{In proportional terms (noise to signal ratio)} \quad \frac{\sigma_e^2}{\theta^2} \geq \frac{1}{4 \cdot N}$$

2. Estimate Array Response Vectors

(a) I'll do both formulations of the problem.

First, the vector version. Define the 6x1 extended gain vector $\mathbf{a}_e = \begin{bmatrix} \mathbf{a}_1 \\ \mathbf{a}_2 \end{bmatrix}$

and $\mathbf{r}(k) = \mathbf{S}(k) \cdot \mathbf{a}_e + \mathbf{n}(k)$

where the 3x6 signal matrix is $\mathbf{S}(k) = \begin{bmatrix} s_1(k) \cdot \mathbf{I}_3 & s_2(k) \cdot \mathbf{I}_3 \end{bmatrix}$ and $\mathbf{I}_3$ is the 3x3 identity matrix

Also stack the vectors by time, so that

$$\mathbf{r}_e = \begin{bmatrix} \mathbf{r}(1) \\ \mathbf{r}(2) \\ \vdots \\ \mathbf{r}(9) \end{bmatrix} \quad (27 \times 1) \quad \mathbf{n}_e = \begin{bmatrix} \mathbf{n}(1) \\ \mathbf{n}(2) \\ \vdots \\ \mathbf{n}(9) \end{bmatrix} \quad (27 \times 1) \quad \mathbf{S}_e = \begin{bmatrix} \mathbf{S}(1) \\ \mathbf{S}(2) \\ \vdots \\ \mathbf{S}(9) \end{bmatrix} \quad (27 \times 6)$$

and the model is summarized by $\mathbf{r}_e = \mathbf{S}_e \cdot \mathbf{a}_e + \mathbf{n}_e$

Second, the matrix version. Define $\mathbf{s}(k) = \begin{bmatrix} s_1(k) \\ s_2(k) \end{bmatrix}$ so that $\mathbf{r}(k) = \mathbf{A} \cdot \mathbf{s}(k) + \mathbf{n}(k)$

where the 3x2 matrix of unknown gains is $\mathbf{A} = \begin{bmatrix} \mathbf{a}_1 & \mathbf{a}_2 \end{bmatrix}$

Pack the received, transmitted and noise vectors into arrays $\mathbf{R}$, $\mathbf{S}$, $\mathbf{N}$ with 9 columns

$$\mathbf{R}^{<k>} = \mathbf{r}(k) \quad \mathbf{S}^{<k>} = \mathbf{s}(k) \quad \mathbf{N}^{<k>} = \mathbf{n}(k) \quad k=1..9$$

Then the model is summarized by $\mathbf{R} = \mathbf{A} \cdot \mathbf{S} + \mathbf{N}$

Read in the given data (only first 7 columns shown - scroll to see the rest):

$\mathbf{R} :=$

-2.48	1.77	1.83	-1.81	-2.52	2.53	-1.77
1.37	-0.44	-0.32	0.32	1.41	-1.38	0.36
1.1	0.4	0.48	-0.49	1.09	-1.06	-0.33

$\mathbf{S} :=$

1	-1	-1	1	1	-1	1
-1	-1	-1	1	-1	1	1

(b) Following Section 2.6.3 in the notes, we identify the $\mathbf{r}$ s with $\mathbf{x}$, the $\mathbf{a}$ s with θ and the $\mathbf{s}$ s with $\mathbf{A}$.

To do least squares with the first formulation, we extend the arrays as follows:

$$\mathbf{I}_3 := \text{identity}(3)$$

$$\mathbf{r}_e := \left| \begin{array}{l} \mathbf{r} \leftarrow \mathbf{R}^{<0>} \\ \text{for } k \in 1..8 \\ \quad \mathbf{r} \leftarrow \text{stack}(\mathbf{r}, \mathbf{R}^{<k>}) \\ \mathbf{r} \end{array} \right| \quad \mathbf{S}_e := \left| \begin{array}{l} \mathbf{S}_e \leftarrow \text{augment}(\mathbf{S}_{0,0} \cdot \mathbf{I}_3, \mathbf{S}_{1,0} \cdot \mathbf{I}_3) \\ \text{for } k \in 1..8 \\ \quad \mathbf{S}_e \leftarrow \text{stack}(\mathbf{S}_e, \text{augment}(\mathbf{S}_{0,k} \cdot \mathbf{I}_3, \mathbf{S}_{1,k} \cdot \mathbf{I}_3)) \\ \mathbf{S}_e \end{array} \right|$$

which gives the arrays shown below (only first 15 rows shown - scroll to see the rest)

$$\mathbf{r}_e = \begin{bmatrix} -2.48 \\ 1.37 \\ 1.103 \\ 1.767 \\ -0.437 \\ 0.403 \\ 1.833 \\ -0.319 \\ 0.484 \\ -1.815 \\ 0.319 \\ -0.486 \\ -2.523 \\ 1.413 \\ 1.092 \\ 2.526 \end{bmatrix} \quad \mathbf{S}_e = \begin{bmatrix} 1 & 0 & 0 & -1 & 0 & 0 \\ 0 & 1 & 0 & 0 & -1 & 0 \\ 0 & 0 & 1 & 0 & 0 & -1 \\ -1 & 0 & 0 & -1 & 0 & 0 \\ 0 & -1 & 0 & 0 & -1 & 0 \\ 0 & 0 & -1 & 0 & 0 & -1 \\ -1 & 0 & 0 & -1 & 0 & 0 \\ 0 & -1 & 0 & 0 & -1 & 0 \\ 0 & 0 & -1 & 0 & 0 & -1 \\ 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & -1 & 0 & 0 \\ 0 & 1 & 0 & 0 & -1 & 0 \\ 0 & 0 & 1 & 0 & 0 & -1 \\ -1 & 0 & 0 & 1 & 0 & 0 \end{bmatrix}$$

The solution from the notes can be adopted directly, and

$$\mathbf{a}_{\text{ehat}} := (\mathbf{S}_e^T \cdot \mathbf{S}_e)^{-1} \cdot \mathbf{S}_e^T \cdot \mathbf{r}_e$$

$$\mathbf{a}_{1\text{hat}} := \text{submatrix}(\mathbf{a}_{\text{ehat}}, 0, 2, 0, 0)$$

$$\mathbf{a}_{2\text{hat}} := \text{submatrix}(\mathbf{a}_{\text{ehat}}, 3, 5, 0, 0)$$

$$\mathbf{A}_{\text{hat}} := \text{augment}(\mathbf{a}_{1\text{hat}}, \mathbf{a}_{2\text{hat}})$$

$$\mathbf{A}_{\text{hat}} = \begin{bmatrix} -2.149 & 0.352 \\ 0.861 & -0.504 \\ 0.33 & -0.755 \end{bmatrix}$$

To use LS in the second formulation, note that each row of $\mathbf{A}$ can be estimated independently of the others, since they don't interact. Thus

$$\text{row}_k(\mathbf{R}) = \text{row}_k(\mathbf{A}) \cdot \mathbf{S} + \text{row}_k(\mathbf{N})$$

to which we apply LS row wise (in a transposed fashion), to get

$$\text{row}_k(\mathbf{A}_{\text{hat}}) = \text{row}_k(\mathbf{R}) \cdot \mathbf{S}^T \cdot (\mathbf{S} \cdot \mathbf{S}^T)^{-1}$$

or

$$\mathbf{A}_{\text{hat}} := \mathbf{R} \cdot \mathbf{S}^T \cdot (\mathbf{S} \cdot \mathbf{S}^T)^{-1} \quad \mathbf{A}_{\text{hat}} = \begin{bmatrix} -2.149 & 0.352 \\ 0.861 & -0.504 \\ 0.33 & -0.755 \end{bmatrix}$$

(c) Substituting the estimates of $\mathbf{A}$ back in the original equations, we have the projections

$$\mathbf{X}_{\text{hat}} := \mathbf{A}_{\text{hat}} \cdot \mathbf{S} \quad \mathbf{N}_{\text{hat}} := \mathbf{R} - \mathbf{X}_{\text{hat}}$$

$$\mathbf{X}_{\text{hat}} = \begin{bmatrix} -2.501 & 1.797 & 1.797 & -1.797 & -2.501 & 2.501 & -1.797 & 2.501 & -2.501 \\ 1.365 & -0.358 & -0.358 & 0.358 & 1.365 & -1.365 & 0.358 & -1.365 & 1.365 \\ 1.085 & 0.425 & 0.425 & -0.425 & 1.085 & -1.085 & -0.425 & -1.085 & 1.085 \end{bmatrix}$$

$$\mathbf{N}_{\text{hat}} = \begin{bmatrix} 0.021 & -0.03 & 0.036 & -0.018 & -0.022 & 0.025 & 0.022 & -0.017 & 7.4 \cdot 10^{-3} \\ 5.2 \cdot 10^{-3} & -0.079 & 0.039 & -0.039 & 0.048 & -0.015 & -1.75 \cdot 10^{-3} & 0.055 & -0.014 \\ 0.018 & -0.021 & 0.059 & -0.062 & 6.8 \cdot 10^{-3} & 0.021 & 0.099 & -9.8 \cdot 10^{-3} & -0.013 \end{bmatrix}$$

3. ML and Bayes Estimation

(a) This is just additive Gaussian noise, so

$$p_{\mathbf{n}}(\mathbf{n}) = \frac{1}{\pi \cdot \sigma_{\mathbf{n}}^2} \cdot \exp \left[\frac{-(\|\mathbf{n}\|)^2}{\sigma_{\mathbf{n}}^2} \right] \quad p(\mathbf{x} | \mathbf{n}) = \frac{1}{\pi \cdot \sigma_{\mathbf{n}}^2} \cdot \exp \left[\frac{-(\|\mathbf{x} - \theta\|)^2}{\sigma_{\mathbf{n}}^2} \right]$$

The conditional probability is maximized by $\theta_{\text{hat}} = \mathbf{x}$

(b) Is the ML estimate biased? We have $\theta_{\text{hat}} = \mathbf{x} = \theta + \mathbf{n}$

There is only one ensemble, and with zero mean noise, we have $E(\theta_{\text{hat}}) = \theta$

so it is unbiased.

(c) If θ is Gaussian, too, then we get an MMSE estimate $\hat{\theta} = \bar{w} \cdot x$. The scalar "normal equation" is

$$\sigma_x^2 \cdot \bar{w} = \sigma_{x\theta}^2 \quad \text{or} \quad \bar{w} = \frac{\sigma_{x\theta}^2}{\sigma_x^2} \quad \text{so} \quad \hat{\theta} = \bar{w} \cdot x = \frac{\sigma_{x\theta}^2}{\sigma_x^2} \cdot x = \frac{\sigma_\theta^2}{\sigma_\theta^2 + \sigma_n^2} \cdot x$$

(d) We have two different estimators, each claiming to be unbiased. We resolve the apparent contradiction by noting that we have different statistical characterizations: one ensemble for part (b), two ensembles for part (c).