Homework 7 Solutions

1. ***Matched filtering for sonar***

A sonar system operates by emitting a short pulse of sound, and listening for the reflection. The time between the emission and the reflection then determines the distance to the object. We will model the sonar in discrete-time, and use a sample period of 1 in appropriate units.

A commonly used signal shape for sonar is a *chirp*, which has the form of a sound with slowly rising frequency, given by

$$u(t) = \sin(c_1 t^2) \qquad \text{for } t \in [0, p-1]$$

Different frequency ranges are used depending on the desired range of distances at which objects are to be detected.

The emitted signal is then

$$u_{\text{emitted}} = u(t+p) \qquad \text{for } t \in [-p, -1]$$

For simplicity we assume that the received signal has the same shape as the emitted chirp. Of course, it is delayed by an amount proportional to the distance to the object, and corrupted by noise.

$$y(t) = \begin{cases} u(t-d) + w(t) & \text{if } d \le t \le p+d-1 \\ w(t) & \text{otherwise} \end{cases}$$

We listen for this reflection on the interval $[0, m-1]$. The noise $w(0), w(1), w(2), \ldots$ is IID and Gaussian, with

$$w(t) \sim \mathcal{N}(0, \sigma^2) \qquad \text{for all } t$$

We would now like to measure the signal $y(0), y(1), \ldots, y(m-1)$ and determine d. Our prior information regarding the random variable d is that it is uniformly distributed on the discrete interval $[0, m-p]$. To simplify notation, define

$$y = \begin{bmatrix} y(0) \\ y(1) \\ \vdots \\ y(m-1) \end{bmatrix} \qquad u = \begin{bmatrix} u(0) \\ u(1) \\ \vdots \\ u(p-1) \end{bmatrix} \qquad w = \begin{bmatrix} w(0) \\ w(1) \\ \vdots \\ w(m-1) \end{bmatrix}$$

Also define the shift matrices $S(k) \in \mathbb{R}^{m \times p}$ for $k = 0, 1, \ldots, m-p$ by

$$S(k) = \begin{bmatrix} 0 \\ I \\ 0 \end{bmatrix} = \begin{bmatrix} 0_{k \times p} \\ I_p \\ 0_{(m-p-k) \times p} \end{bmatrix}$$

where $O_{n \times m}$ denotes the $n \times m$ matrix of zeros, and I_p denotes the $p \times p$ identity matrix. Then we have

$$y = S(d)u + w$$

Then $\Sigma = \sigma^2 I_m$ is the covariance of w.

(a) Show that the conditional pdf of y given $d = k$ is

$$y \,|\, (d = k) \sim \mathcal{N}(S(k)u, \Sigma)$$

(b) We would like to construct an estimator of d given a measurement of y. Show that the estimator $f : R^m \to \{0, 1, \ldots, m-p\}$ which minimizes the probability of error is

$$f_{\text{est}}(y_{\text{meas}}) = \arg\max_k g_k(y_{\text{meas}})$$

where the functions g_k are defined by

$$g_k(y) = u^T S(k)^T y$$

(c) In fact there is a very efficient method of computing this. Let h be the vector u flipped upside down; that is

$$h(t) = u(p+1-t) \qquad \text{for } t = 0, \dots, p-1$$

Show that

$$g_k(y) = r(p+k-1)$$

where $r = h * y$ is the *convolution* of h with y, i.e.,

$$r(j) = \sum_{i=0}^{m} h(j-i)y(i)$$

with the usual convention that $h(i) = 0$ if i is too large or too small.

This means that, to find the estimate of d, all we do is construct the convolution, and look at the *peak* of the signal r. If this peak occurs at r_k, then the estimate for d is $k-p$. The nice thing about convolution is that it is simply a *filtering* operation, and the filter r is called the *matched* filter.

(d) We will use chirp parameters $c_1 = 0.003$ and $p = 100$, and listen for signals up to length $m = 500$.

Suppose the noise covariance is $\sigma^2 = 0.01$. Simulate the generation of y, and construct the signal r. Plot the signals r, y and u. It's worth using the same horizontal scale for all the plots, so you can see what the signal is you are looking for. Note that you can use the MATLAB commands `flipud` and `conv` to compute r.

(e) Now run the simulation a large number of times, say $50,000$. Use this to estimate the conditional probability of error

$$\text{Prob}(f_{\text{est}}(y) \neq k \,|\, d = k)$$

and also the conditional probability of an error of magnitude greater than 1

$$\text{Prob}(|f_{\text{est}}(y) - d| > 1 \,|\, d = k)$$

(f) Repeat parts (d) through (e) above with a noise covariance of $\sigma^2 = 1$. Notice how difficult it is to locate the signal simply by looking at the plot of y.

Solution.

(a) Let the conditional pdf of y conditioned on $d = k$ be the function q, so that

$$\text{Prob}(y \in R \,|\, d = k) = \int_R q(y,k)\, dk$$

for any subset $R \subset \mathbb{R}^n$. Hence we have

$$\begin{aligned} \int_R q(y,k)\, dk &= \text{Prob}(S(k)u + w \in R \,|\, d = k) \\ &= \text{Prob}(w \in R - S(k)u) \end{aligned}$$

where $R - S(k)u$ means

$$R - S(k)u = \Big\{ z \in \mathbb{R}^n \mid z = x - S(k)u \text{ for some } x \in R \Big\}$$

Hence

$$\int_R q(y,k)\, dk = \int_{R-S(k)u} p^w(w)\, dw$$

where p^w is the induced pdf of w. Changing variables gives

$$\int_R q(y,k)\,dk = \int_R p^w(y - S(k)u)\,dy$$

and since this holds for all R we have

$$q(y,k) = p^w(y - S(k)u)$$

Therefore q is the Gaussian pdf of w shifted by $S(k)u$, and it has mean $S(k)u$ and covariance Σ.

(b) We know that the classifier that minimizes the probability of error is the MAP classifier, i.e., given y, we pick $d = k$, where k maximizes the joint pdf

$$\text{Prob}(d = k)q(y,k)$$

where as above $q(y,k)$ is the conditional pdf of y given $d = k$. Since d is uniformly distributed, we have

$$\begin{aligned} f_{\text{est}}(y) &= \arg\max_k q(y,k) \\ &= \arg\max_k p^w(y - S(k)u) \\ &= \arg\max_k \exp\left(-\frac{1}{2}(y - S(k)u)^T\Sigma^{-1}(y - S(k)u)\right) \\ &= \arg\min_k \left(\big(y - S(k)u\big)^T\big(y - S(k)u\big)\right) \end{aligned}$$

where we have used the facts that the exponential is monotonic and $\Sigma = \sigma^2 I$. Now we can remove any terms on the right hand side that don't depend on k. We have

$$(y - S(k)u)^T(y - S(k)u) = y^Ty - 2u^TS(k)^Ty + u^TS(k)^TS(k)u$$

Clearly y^Ty is independent of k, and the last term is also since $S(k)^TS(k) = I$. Hence

$$f_{\text{est}}(y) = \arg\min_k -2u^TS(k)^Ty$$

which gives the desired result.

(c) We have $g_k(y) = u^TS(k)^Ty$ and hence

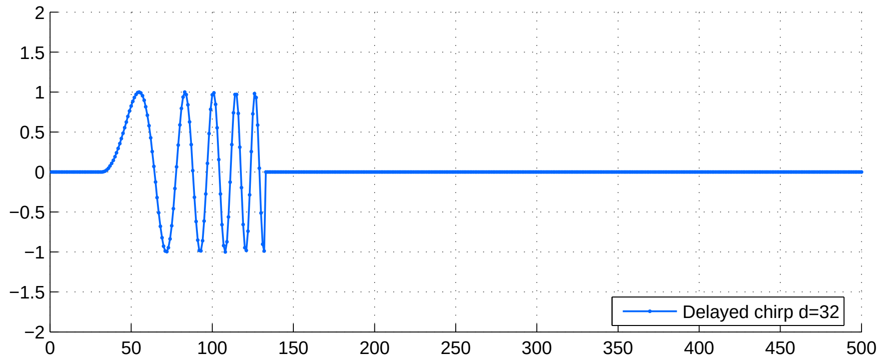
$$\begin{aligned} g_k(y) &= \sum_{t=0}^{p-1} u(t)y(t+k) \\ &= \sum_{i=k}^{k+p-1} u(i-k)y(i) \end{aligned}$$

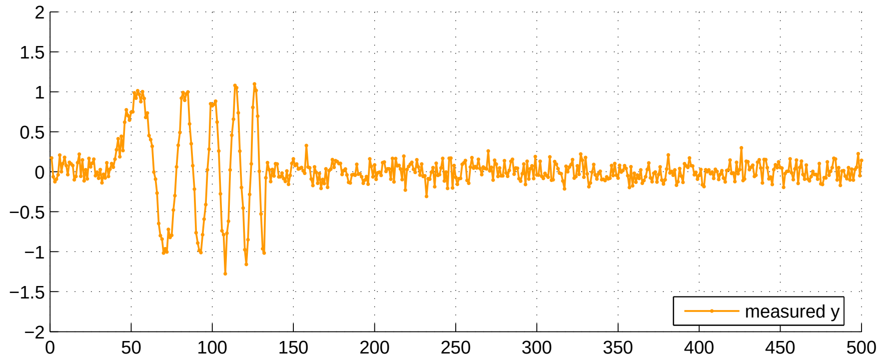
and since $u(t) = h(p-1-t)$ for all $t = 0, 1, \dots, p-1$ we have

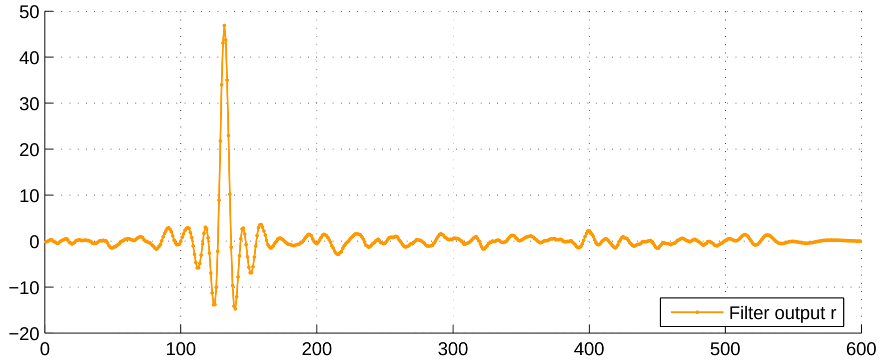
$$\begin{aligned} g_k(y) &= \sum_{i=k}^{k+p-1} h(p+k-1-i)y(i) \\ &= r(p+k-1) \end{aligned}$$

as desired.

(d) The plots are below.

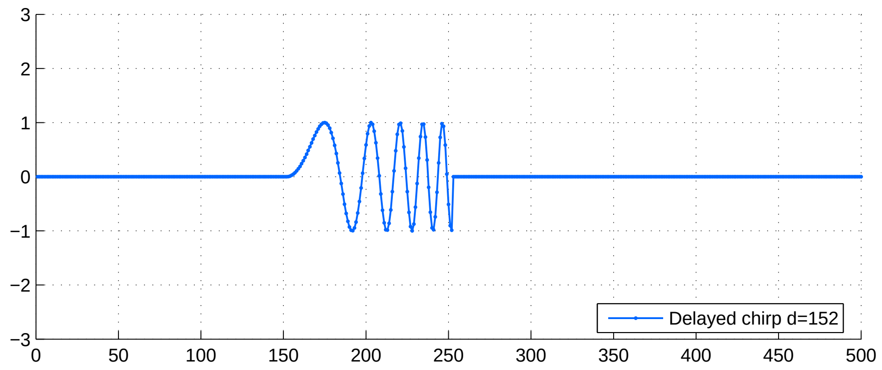


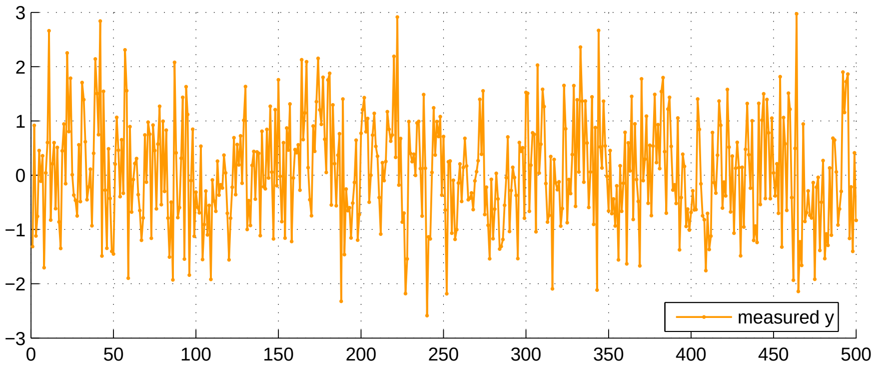


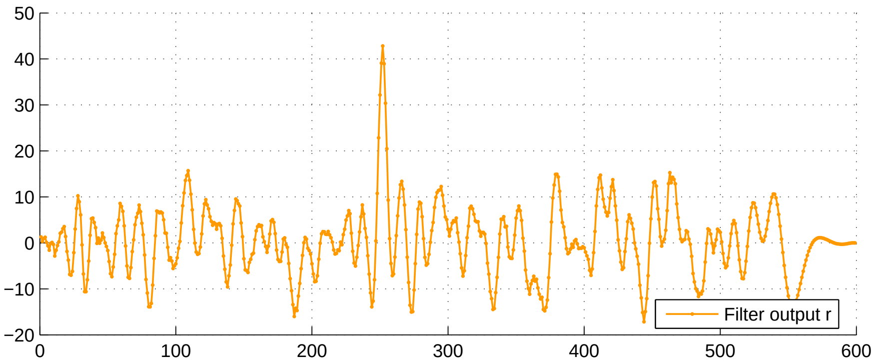


(e) For each simulation, pick d from a uniform distribution over the discrete set $[0, m - p]$. Then generate w from $\mathcal{N}(0, \sigma^2)$. In 50,000 trials, zero errors are made.

(f) The plots are below.







The estimates are

$$\text{Prob}(f_{\text{est}}(y) \neq k \mid d = k) \approx 0.196$$

and

$$\text{Prob}(|f_{\text{est}}(y) - d| > 1 \mid d = k) \approx 0.0012$$

2. ***Navigation***

Suppose a ship is located at position $x \in \mathbb{R}^2$. We measure distances to beacons located at

$$q_1 = \begin{bmatrix} 500 \\ 0 \end{bmatrix} \qquad q_2 = \begin{bmatrix} 0 \\ 500 \end{bmatrix} \qquad q_3 = \begin{bmatrix} 500 \\ -200 \end{bmatrix}$$

As usual we model this as $y = Ax + w$, where $y \in \mathbb{R}^3$ is a vector of measurements (with appropriate constants to make the relationship linear) and $w \in \mathbb{R}^3$ is Gaussian noise, $w \sim \mathcal{N}(0, \Sigma_w)$. We also have prior information $x \sim \mathcal{N}(\mu_x, \Sigma_x)$. Here

$$\mu_x = \begin{bmatrix} 2 \\ 3 \end{bmatrix} \qquad \Sigma_x = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \qquad \Sigma_w = \begin{bmatrix} 0.1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

(a) The MMSE estimate has the form

$$\phi_{\text{mmse}}(y) = \mu_x + L_{\text{mmse}}(y - A\mu_x)$$

Find the estimator gain L_{mmse}.

(b) Perform the experiment in Matlab; i.e., simulate x and w with the above distributions and compute y. Using this y, plot the 90% confidence ellipsoid for x conditioned on your measurement, centered on your estimate.

(c) Now collect data, consisting of 500 samples of $y = Ax + w$. For each measurement y, compute the corresponding estimate $\phi_{\text{mmse}}(y)$, and plot the error in $\mathbb{R}^2$ given by $x - \phi_{\text{mmse}}(y)$. Also plot the 90% confidence ellipsoid for the error (centered at the origin.)

(d) Suppose instead we were to use the least squares estimator

$$\phi_{\text{ls}}(y) = A^\dagger y$$

What is the mean square error

$$\mathrm{E}\|\phi_{\text{ls}}(y) - x\|^2$$

achieved by this estimator?

(e) Using the same data as above, plot the error in $\mathbb{R}^2$ given by $x - \phi_{\text{ls}}(y)$. Also plot the 90% confidence ellipsoid for the error, centered at the origin.

(f) Is it true that the ellipsoid of part (c) is always contained in the ellipsoid of part (e)? Prove or give a counterexample.

Solution.

(a) The MMSE estimator is

$$\phi_{\text{mmse}}(y) = \mu_x + L_{\text{mmse}}(y - A\mu_x)$$

where

$$L_{mmse} = \Sigma_{xy}\Sigma_y^{-1}$$

We have $y = Ax + w$ where A has the normalized beacon vectors as each row.

$$A = \begin{bmatrix} q_1^T/\|q_1\| \\ q_2^T/\|q_2\| \\ q_3^T/\|q_3\| \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0.9285 & -0.3714 \end{bmatrix}$$

We now have

$$\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} I & 0 \\ A & I \end{bmatrix} \begin{bmatrix} x \\ w \end{bmatrix}$$

and the joint covariance matrix is

$$\begin{bmatrix} \Sigma_x & \Sigma_{xy} \\ \Sigma_{yx} & \Sigma_y \end{bmatrix} = \begin{bmatrix} \Sigma_x & \Sigma_x A^T \\ A\Sigma_x & A\Sigma_x A^T + \Sigma_w \end{bmatrix}$$

Hence,

$$L_{\text{mmse}} = \Sigma_{xy}\Sigma_y^{-1} = \begin{bmatrix} 0.8470 & 0.0137 & 0.0736 \\ 0.1366 & 0.4699 & -0.1618 \end{bmatrix}$$

(b) We have

$$\mathrm{E}(x|y) = \phi_{\text{mmse}}(y)$$

and define

$$\Sigma_{x|y} = \mathrm{cov}(x|y)$$

then

$$\Sigma_{x|y} = \Sigma_x - \Sigma_x A^T(A\Sigma_x A^T + \Sigma_w)^{-1}A\Sigma_x$$

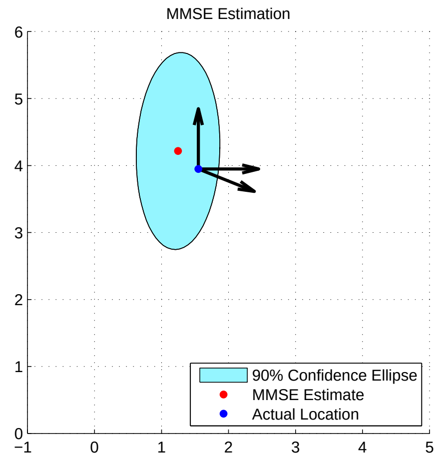
The 90% confidence ellipse for x conditioned on y is

$$\left\{ x \in \mathbb{R}^2 \mid (x - \phi_{\text{mmse}}(y))^T \Sigma_{x|y}^{-1}(x - \phi_{\text{mmse}}(y)) \le \alpha \right\}$$

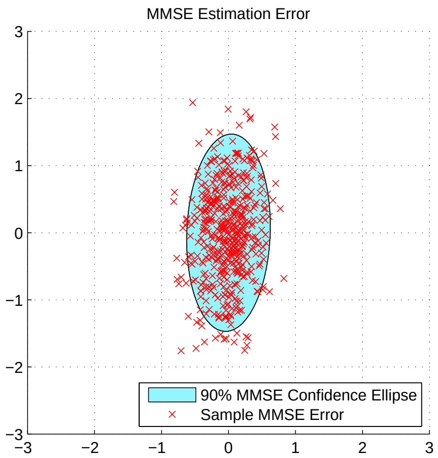
where

$$\alpha = F_{\chi_2^2}^{-1}(0.9) = 4.6052$$

The following figure shows the 90% confidence ellipse centered on the MMSE estimate for a particular simulation.



(c) Define the random variable $z = x - \phi_{mmse}(y)$. It has a Gaussian distribution with zero mean and covariance $\Sigma_{x|y}$. Thus, the 90% confidence ellipse for the error z looks the same as the ellipse in part b, translated to be centered on the origin, as seen in the following figure.



(d) Define the random variable $z_{\text{ls}} = \phi_{\text{ls}}(y) - x$. We have

$$\begin{aligned} z_{\text{ls}} &= A^\dagger y - x \\ &= A^\dagger (Ax + w) - x \\ &= A^\dagger w \end{aligned}$$

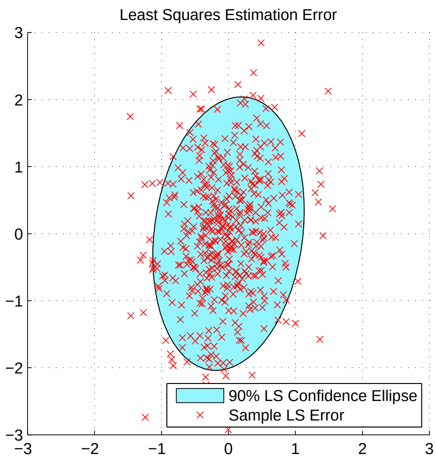
Therefore z_{ls} is Gaussian with zero mean. Let $\Sigma_{\text{ls}} = \text{cov}(z_{\text{ls}})$, then

$$\Sigma_{\text{ls}} = A^\dagger \Sigma_w A^{\dagger T}$$

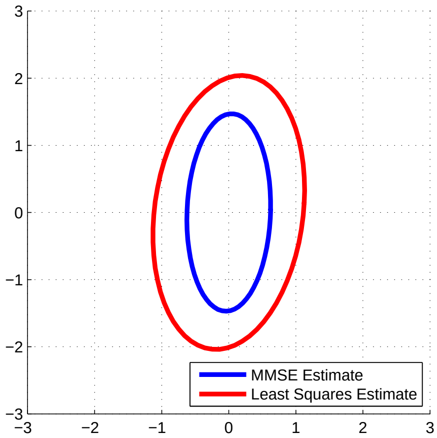
So, the mean square error achieved by this estimator is

$$\mathbf{E}\|\phi_{\text{ls}}(y) - x\|^2 = \text{trace}(\Sigma_{\text{ls}}) = 1.1819$$

(e) As in part (c), the 90% confidence ellipse corresponds to the covariance of z_{ls}, as shown in the following figure.



(f) We can visually verify this in the current example by plotting the two ellipses, as shown below.



In order to prove that the ellipse from part (c) is always contained within the ellipse from part (e), we need to show that

$$\Sigma_{x|y}^{-1} \geq \Sigma_{\text{ls}}^{-1}$$

We have

$$\begin{aligned}\Sigma_{x|y}^{-1} - \Sigma_{\text{ls}}^{-1} &= (\Sigma_x - \Sigma_x A^T (A\Sigma_x A^T + \Sigma_w)^{-1} A\Sigma_x)^{-1} - (A^\dagger \Sigma_w A^\dagger)^{-1} \\ &= \Sigma_x^{-1} + A^T \Sigma_w^{-1} A - A^T A (A^T \Sigma_w A)^{-1} A^T A\end{aligned}$$

by the matrix inversion lemma, and hence

$$\Sigma_{x|y}^{-1} - \Sigma_{\text{ls}}^{-1} = \Sigma_x^{-1} + A^T (\Sigma_w^{-1} - A(A^T \Sigma_w A)^{-1} A^T) A$$

To analyze the second term in the above expression, define the matrix

$$M = \begin{bmatrix} A^T \Sigma_w A & A^T \\ A & \Sigma_w^{-1} \end{bmatrix}$$

Then M satisfies

$$M = B^T \begin{bmatrix} \Sigma_w & 0 \\ 0 & 0 \end{bmatrix} B$$

where

$$B = \begin{bmatrix} A & \Sigma_w^{-1} \\ 0 & I \end{bmatrix}$$

and hence we have that $M \geq 0$. Since $M_{11} > 0$, we have by the completion of squares argument that the schur complement of M must be positive semidefinite also. That is,

$$M_{22} - M_{21} M_{11}^{-1} M_{12} \geq 0$$

and this is just

$$\Sigma_w^{-1} - A(A^T \Sigma_w A)^{-1} A^T \geq 0$$

As a result,

$$\Sigma_{x|y}^{-1} - \Sigma_{\text{ls}}^{-1} \geq 0$$

as desired.