TD1: Parameter and Spectrum Estimation Solutions

1 Parameter Estimation

Problem 1. Poisson Process - Bayesian Parameter Setting

- **The conditional pdf $\boldsymbol{f(\theta|y)}$** is computed using Bayes' rule: $f(\theta|y) = \dfrac{f(y|\theta)f(\theta)}{\int_\Theta f(y|\theta)f(\theta)d\theta}$
  $f(y) = \lambda \int_0^{+\infty} \theta e^{-\theta(y+\lambda)} d\theta = \dfrac{\lambda}{(y+\lambda)^2}$ for $y \geq 0$ and $f(y) = 0$ for $y < 0$.
  $$f(\theta|y) = \begin{cases} (y+\lambda)^2 \theta e^{-\theta(y+\lambda)} & \theta \geq 0 \text{ and } y \geq 0 \\ 0 & \theta < 0 \text{ or } y < 0 \end{cases}$$

- $\boldsymbol{\widehat{\theta}_{MMSE} = E(\theta|y) = \int_\Theta \theta f(\theta|y) d\theta}$$= (y+\lambda)^2 \int_0^{+\infty} \theta^2 e^{-\theta(y+\lambda)} d\theta = (y+\lambda)^2 \dfrac{2!}{(y+\lambda)^3}$.
  Hence: $\widehat{\theta}_{MMSE} = \dfrac{2}{y+\lambda}$

- $\boldsymbol{\widehat{\theta}_{MAP}} =$ **argmax**$\boldsymbol{f(\theta|y)}$: $\frac{\partial \ln f(\theta|y)}{\partial \theta} = \frac{1}{\theta} - (y+\lambda) = 0 \Rightarrow \theta = \frac{1}{y+\lambda} > 0$ and $\frac{\partial^2 \ln f(y|\theta)}{\partial \theta^2} = -\frac{1}{\theta^2} < 0$. So the extremum is a maximum and: $\widehat{\theta}_{MAP} = \frac{1}{y+\lambda}$

- $\boldsymbol{\widehat{\theta}_{ABS}}$ **such that** $\boldsymbol{F(\theta|y) = \frac{1}{2}}$**:** $F(\theta|y) = \int_0^\theta f(\theta'|y)d\theta' = (y+\lambda)^2 \int_0^\theta \theta' e^{-\theta'(y+\lambda)} d\theta'$
  $F(\theta|y) = 1 - e^{-\theta(y+\lambda)} - (y+\lambda)\theta e^{-\theta(y+\lambda)}$, which is a function of $\theta(y+\lambda)$. An approximate solution of the transcendental equation $0.5 - e^{-x} - xe^{-x} = 0$ is $x = 1.68$, so: $\widehat{\theta}_{ABS} = \frac{1.68}{y+\lambda}$.
  (One can obtain this solution using the Matlab function fsolve).

- **Performance of the estimators:** $R_{\widetilde{\theta}_{MMSE}} = E_{y,\theta}\left(\theta - \widehat{\theta}_{MMSE}(y)\right)\left(\theta - \widehat{\theta}_{MMSE}(y)\right)^T$
  In this case: $R_{\widetilde{\theta}_{MMSE}} = \sigma^2_{\widetilde{\theta}_{MMSE}} = E_{y,\theta}\left(\frac{2}{y+\lambda} - \theta\right)^2$
  $$= \int_0^{+\infty} \int_0^{+\infty} \left(\frac{2 - \theta(y+\lambda)}{y+\lambda}\right)^2 \lambda \theta e^{-\theta(y+\lambda)} dy d\theta = \frac{2}{3\lambda^2}.$$
  $$\begin{cases} R_{\widetilde{\theta}_{MMSE}} & \simeq 0.67/\lambda^2 \\ R_{\widetilde{\theta}_{MAP}} & = 1/\lambda^2 \\ R_{\widetilde{\theta}_{ABS}} & \simeq 0.7/\lambda^2 \end{cases}$$
  $\widehat{\theta}_{MMSE}$ has the best performance as it minimizes the MSE, i. e. $E(\theta - \widehat{\theta})^2$. $\widehat{\theta}_{ABS}$ minimizes $E|\theta - \widehat{\theta}|^\alpha$ with $\alpha = 1$ while $\widehat{\theta}_{MAP}$ minimizes $E|\theta - \widehat{\theta}|^\alpha$ for $\alpha = 0$: $\alpha = 1$ being closer to 2 than 0, it is intuitively expected that $\widehat{\theta}_{ABS}$ is better than $\widehat{\theta}_{MAP}$.

Problem 2. MAP and MMSE Estimation - Hard and Soft Decisions for Gaussian Noise

(a) We obtain the marginal distribution from the joint distribution by summing:

$$f_y(y_0) = \sum_{\theta_0 \pm 1} f_{y|\theta}(y_0|\theta_0)\mathrm{Pr}_\theta(\theta = \theta_0). \tag{1}$$

Since $f_{y|\theta}(y_0|\theta_0) = f_v(y_0 - h\theta_0)$, (1) becomes

$$\begin{aligned} f_y(y_0) &= f_{y|\theta}(y_0|\theta_0 = +1)\mathrm{Pr}_\theta(\theta = +1) + f_{y|\theta}(y_0|\theta_0 = -1)\mathrm{Pr}_\theta(\theta = -1) \\ &= f_{y|\theta}(y_0|\theta_0 = 1)\tfrac{1}{2} + f_{y|\theta}(y_0|\theta_0 = -1)\tfrac{1}{2} \\ &= \tfrac{1}{2}f_v(y_0 - h) + \tfrac{1}{2}f_v(y_0 + h). \end{aligned}$$

The posterior distribution of θ given y is the ratio of joint to marginal distributions, hence

$$\begin{aligned} \mathrm{Pr}(\theta|y) &= \begin{cases} \dfrac{\frac{1}{2}f_v(y-h)}{\frac{1}{2}f_v(y-h) + \frac{1}{2}f_v(y+h)}, & \text{for } \theta = 1, \\ \dfrac{\frac{1}{2}f_v(y+h)}{\frac{1}{2}f_v(y-h) + \frac{1}{2}f_v(y+h)}, & \text{for } \theta = -1. \end{cases} \\ &= \begin{cases} \dfrac{f_v(y-h)}{f_v(y-h) + f_v(y+h)}, & \text{for } \theta = 1, \\ \dfrac{f_v(y+h)}{f_v(y-h) + f_v(y+h)}, & \text{for } \theta = -1. \end{cases} \end{aligned}$$

(b) The MAP estimate of θ given y verifies

$$\begin{aligned} & \begin{cases} \text{if } \mathrm{Pr}_{\theta|y}(1|y) > \mathrm{Pr}_{\theta|y}(-1|y) \Rightarrow \widehat{\theta}_{MAP} = 1, \\ \text{if } \mathrm{Pr}_{\theta|y}(1|y) < \mathrm{Pr}_{\theta|y}(-1|y) \Rightarrow \widehat{\theta}_{MAP} = -1. \end{cases} \\ \Leftrightarrow & \begin{cases} \text{if } \dfrac{f_v(y-h)}{f_v(y-h)+f_v(y+h)} > \dfrac{f_v(y+h)}{f_v(y-h)+f_v(y+h)} \Rightarrow \widehat{\theta}_{MAP} = 1, \\ \text{if } \dfrac{f_v(y-h)}{f_v(y-h)+f_v(y+h)} < \dfrac{f_v(y+h)}{f_v(y-h)+f_v(y+h)} \Rightarrow \widehat{\theta}_{MAP} = -1. \end{cases} \\ \Leftrightarrow & \begin{cases} \text{if } f_v(y-h) > f_v(y+h) \Rightarrow \widehat{\theta}_{MAP} = 1, \\ \text{if } f_v(y-h) < f_v(y+h) \Rightarrow \widehat{\theta}_{MAP} = -1. \end{cases} \\ \Leftrightarrow & \begin{cases} \text{if } \frac{1}{\sqrt{2\pi}\sigma_v}e^{-\frac{(y-h)^2}{2\sigma_v^2}} > \frac{1}{\sqrt{2\pi}\sigma_v}e^{-\frac{(y+h)^2}{2\sigma_v^2}} \Rightarrow \widehat{\theta}_{MAP} = 1, \\ \text{if } \frac{1}{\sqrt{2\pi}\sigma_v}e^{-\frac{(y-h)^2}{2\sigma_v^2}} < \frac{1}{\sqrt{2\pi}\sigma_v}e^{-\frac{(y+h)^2}{2\sigma_v^2}} \Rightarrow \widehat{\theta}_{MAP} = -1. \end{cases} \\ \Leftrightarrow & \begin{cases} \text{if } -(y-h)^2 > -(y+h)^2 \Rightarrow \widehat{\theta}_{MAP} = 1, \\ \text{if } -(y-h)^2 < -(y+h)^2 \Rightarrow \widehat{\theta}_{MAP} = -1. \end{cases} \\ \Leftrightarrow & \begin{cases} \text{if } -y^2 + 2yh - h^2 > -y^2 - 2yh - h^2 \Rightarrow \widehat{\theta}_{MAP} = 1, \\ \text{if } -y^2 + 2yh - h^2 < -y^2 - 2yh - h^2 \Rightarrow \widehat{\theta}_{MAP} = -1. \end{cases} \\ \Leftrightarrow & \begin{cases} \text{if } 4yh > 0 \Rightarrow \widehat{\theta}_{MAP} = 1, \\ \text{if } 4yh < 0 \Rightarrow \widehat{\theta}_{MAP} = -1. \end{cases} \Leftrightarrow \begin{cases} \text{if } y > 0 \Rightarrow \widehat{\theta}_{MAP} = 1, \\ \text{if } y < 0 \Rightarrow \widehat{\theta}_{MAP} = -1. \end{cases} \\ \Leftrightarrow & \ \widehat{\theta}_{MAP} = \mathrm{sign}(y). \end{aligned}$$

(c) The MMSE estimate of the symbol θ in terms of y is

$$\begin{aligned}
\widehat{\theta}_{MMSE} &= E(\theta|y) = \sum_{k=\pm 1} k \mathrm{Pr}_{\theta|y}(\theta = k|y) \\
&= (+1)\mathrm{Pr}_{\theta|y}(\theta = +1|y) + (-1)\mathrm{Pr}_{\theta|y}(\theta = -1|y) \\
&= \frac{f_v(y-h) - f_v(y+h)}{f_v(y-h) + f_v(y+h)} \\
&= \frac{\frac{1}{\sqrt{2\pi}\sigma_v} e^{-\frac{(y-h)^2}{2\sigma_v^2}} - \frac{1}{\sqrt{2\pi}\sigma_v} e^{-\frac{(y+h)^2}{2\sigma_v^2}}}{\frac{1}{\sqrt{2\pi}\sigma_v} e^{-\frac{(y-h)^2}{2\sigma_v^2}} + \frac{1}{\sqrt{2\pi}\sigma_v} e^{-\frac{(y+h)^2}{2\sigma_v^2}}} \\
&= \frac{e^{\frac{yh}{\sigma_v^2}} - e^{\frac{-yh}{\sigma_v^2}}}{e^{\frac{yh}{\sigma_v^2}} + e^{\frac{-yh}{\sigma_v^2}}} = \tanh\left(\frac{h\,y}{\sigma_v^2}\right).
\end{aligned}$$

Problem 3. Bayesian Fourier Analysis

Let

$$Y = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} \qquad H = \begin{bmatrix} \cos(2\pi f_o) & \sin(2\pi f_o) \\ \vdots & \vdots \\ \cos(2\pi f_o n) & \sin(2\pi f_o n) \end{bmatrix} \qquad \theta = \begin{bmatrix} a \\ b \end{bmatrix} \qquad V = \begin{bmatrix} v_1 \\ \vdots \\ v_n \end{bmatrix} \tag{2}$$

Then

$$Y = H\theta + V \quad . \tag{3}$$

As θ and V are Gaussian, the Bayesian linear model applies directly.

(i) $\widehat{\theta}_{LMMSE} = \left(H^T H + \frac{\sigma_v^2}{\sigma_\theta^2} I_2\right)^{-1} H^T Y$ $(= \widehat{\theta}_{AMMSE}$ as θ and Y are centered).

$$\begin{cases}
\left(H^T H\right)_{1,1} &= \sum_{k=1}^{n} \cos^2(2\pi f_o k) = \frac{1}{2}\sum_{k=1}^{n} (1 + \cos 4\pi f_o k) \\
\left(H^T H\right)_{2,2} &= \sum_{k=1}^{n} \sin^2(2\pi f_o k) = \frac{1}{2}\sum_{k=1}^{n} (1 - \sin 4\pi f_o k) \\
\left(H^T H\right)_{1,2} &= \sum_{k=1}^{n} \cos(2\pi f_o k)\sin(2\pi f_o k) = \frac{1}{2}\sum_{k=1}^{n} (\sin 4\pi f_o k) \\
\left(H^T H\right)_{2,1} &= \left(H^T H\right)_{1,2}
\end{cases} \tag{4}$$

$$\sum_{k=1}^{n} \cos 4\pi f_o k + j\sum_{k=1}^{n} \sin 4\pi f_o k = \sum_{k=1}^{n} e^{4\pi f_o k} = e^{4\pi f_o}\frac{1 - e^{4\pi f_o n}}{1 - e^{4\pi f_o}} = 0 \tag{5}$$

because f_o is a multiple of $\frac{1}{n}$, so:

$$H^T H = \frac{n}{2} I_2 \tag{6}$$

$$\widehat{\theta}_{LMMSE} = \left(\frac{n}{2} + \frac{\sigma_v^2}{\sigma_\theta^2}\right)^{-1} \begin{bmatrix} \sum_{k=1}^{n} y_k \cos(2\pi f_o k) \\ \sum_{k=1}^{n} y_k \sin(2\pi f_o k) \end{bmatrix} \tag{7}$$

(ii) $\widehat{\theta}_{MMSE} = \widehat{\theta}_{MAP} = \widehat{\theta}_{LMMSE} = \widehat{\theta}_{AMMSE}$ because θ and Y are jointly Gaussian with zero mean.

(iii) $\widehat{\theta}_{LMMSE}$ is efficient

$$\begin{aligned} J(\theta) &= \tfrac{1}{\sigma_v^2}\left(H^T H + \tfrac{\sigma_v^2}{\sigma_\theta^2} I_2\right) \\ CRB = R_{\widetilde{\theta}_{LMMSE}} &= \sigma_v^2 \left(H^T H + \frac{\sigma_v^2}{\sigma_\theta^2} I_2\right)^{-1} = \frac{\sigma_v^2}{\frac{n}{2} + \frac{\sigma_v^2}{\sigma_\theta^2}} I_2 \end{aligned} \tag{8}$$

Problem 4. Deconvolution

(a) $y_k = \theta_k - \theta_{k-2} + v_k$, $k = 0, 1, \ldots, n = m+2$ and $\theta_k = 0$ for $k \leq 0$ or $k > m$

$$\Rightarrow Y = H\,\theta + V, \quad H = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \ddots & \vdots \\ -1 & 0 & \ddots & 0 \\ 0 & -1 & \ddots & 1 \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & -1 \end{bmatrix} \quad \text{(of dimension } (m+2)\times m\text{)}$$

(b) $\widehat{\theta}_{zd} = FY = \underbrace{FH}_{=I}\theta + FV = \theta + FV$

$b_{\widehat{\theta}_{zd}}(\theta) = E_{Y|\theta}\widehat{\theta}(Y) - \theta = E_{V|\theta}FV = F\,\underbrace{E_{V|\theta}V}_{=0} = 0.$

Note: $b_{\widehat{\theta}_{zd}}(\theta) = 0 \Leftrightarrow FH = I$. $FH = I$ gives an underdetermined system of equations in F and F cannot be completely specified. $F = (H^T H)^{-1} H^T$ (Moore-Penrose pseudo-inverse of H) verifies $FH = I$, but also $F = (C^T H)^{-1} C^T$, where C is any matrix of dimension $(m+2) \times m$ and full-column rank.

(c) WLS linear model, course notes $\Rightarrow \widehat{\theta}_{WLS} = \underbrace{(H^T W H)^{-1} H^T W}_{F_{WLS}} Y$.

$F_{WLS} H = I$, so $\widehat{\theta}_{WLS}$ is a possible $\widehat{\theta}_{zd}$.

(d) $\widehat{\theta}_{LS} = (H^T H)^{-1} H^T Y \Rightarrow (H^T H)\widehat{\theta}_{LS} = H^T Y$.

Let's compute $H^T H$.

$$H = \begin{bmatrix} I_m \\ 0_{2\times m} \end{bmatrix} - \begin{bmatrix} 0_{2\times m} \\ I_m \end{bmatrix}$$

$$\begin{aligned} \Rightarrow H^T H \;=\; & \begin{bmatrix} I_m & 0_{m\times 2} \end{bmatrix} \begin{bmatrix} I_m \\ 0_{2\times m} \end{bmatrix} - \begin{bmatrix} I_m & 0_{m\times 2} \end{bmatrix} \begin{bmatrix} 0_{2\times m} \\ I_m \end{bmatrix} \\ & - \left(\begin{bmatrix} I_m & 0_{m\times 2} \end{bmatrix} \begin{bmatrix} 0_{2\times m} \\ I_m \end{bmatrix} \right)^T + \begin{bmatrix} 0_{m\times 2} & I_m \end{bmatrix} \begin{bmatrix} 0_{2\times m} \\ I_m \end{bmatrix} \end{aligned}$$

$$\Rightarrow H^T H \;=\; I_m - Z^2 - Z^{2T} + I_m = 2I_m - Z^2 - Z^{2T}, \text{ with } Z \text{ defined in question (i)}$$

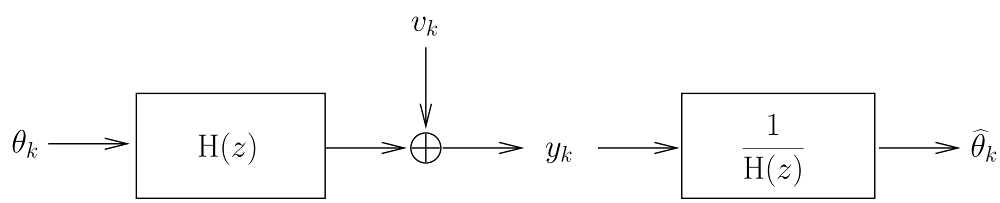
$$= \begin{bmatrix} 2 & 0 & -1 & 0 & \cdots & 0 \\ 0 & \ddots & \ddots & \ddots & \ddots & \vdots \\ -1 & \ddots & \ddots & \ddots & \ddots & 0 \\ 0 & \ddots & \ddots & \ddots & \ddots & -1 \\ \vdots & \ddots & \ddots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & -1 & 0 & 2 \end{bmatrix}$$

So the difference equation between the $\widehat{\theta}_k$'s and the y_k's is:

$$-\widehat{\theta}_{k-2} + 2\widehat{\theta}_k - \widehat{\theta}_{k+2} = y_k - y_{k+2}, \tag{9}$$

$k = 1, \cdots, m$ and $\widehat{\theta}_k = 0$ for $k \leq 0$ or $k > m$. If (9) is valid for any k, $\widehat{\theta}_k$ can be interpreted as the output of the filter $\mathrm{F}(z)$ taking y_k as input, with:

$$\mathrm{F}(z) = \frac{1-z^2}{-z^{-2}+2-z^2} = \frac{1-z^2}{(1-z^2)(1-z^{-2})} = \frac{1}{1-z^{-2}} = \frac{1}{\mathrm{H}(z)} \tag{10}$$



$\mathrm{H}(z)$ is not stable (it has zeros on the unit circle). Let $\widehat{\theta}(f)$, $\theta(f)$ and $V(f)$ be the Fourier transforms of $\{\widehat{\theta}_k\}_{k\in]-\infty,+\infty[}$, $\{\theta_k\}_{k\in]-\infty,+\infty[}$ and $\{v_k\}_{k\in]-\infty,+\infty[}$.

$$\widehat{\theta}(f) = \theta(f) + \frac{1}{H(f)} V(f) \tag{11}$$

At the frequencies where $H(f)$ is zero, the noise term is infinite: we say that there is noise amplification. The MSE is infinite.

(e) Linear Model, Course notes $\Rightarrow \widehat{\theta}_{ML} = \left(H^T H\right)^{-1} H^T Y$. $\widehat{\theta}_{ML}$ is unbiased.

(f) $m_\theta = 0$

$R_{\theta\theta} = C_{\theta\theta} = \sigma_\theta^2 I$: indeed, $C_{\theta\theta}(i,j) = 0$, $i \neq j$, because the θ_k's are independent, $C_{\theta\theta}(i,i) = \sigma_\theta^2$.

θ and Y centered $\Rightarrow C_{\theta Y} = R_{\theta\theta}$, $C_{YY} = R_{YY}$, then $\widehat{\theta}_{LMMSE} = C_{\theta Y} {C_{YY}}^{-1} Y$.

(g) $C_{\theta Y} = R_{\theta Y} = E_{Y,\theta}\theta Y^T = E\theta\theta^T H^T + E\theta V^T = (\sigma_\theta^2 I)\, H^T + (\underbrace{E\theta}_{=0})(\underbrace{EV}_{=0})^T = \sigma_\theta^2 H^T$

$C_{YY} = R_{YY} = H R_{\theta\theta} H^T + H R_{\theta V} + R_{V\theta} H^T + R_{VV} = \sigma_\theta^2 H H^T + \sigma_v^2 I$

Course notes $\Rightarrow$

$$\widehat{\theta}_{LMMSE} = C_{\theta Y} C_{YY}^{-1} Y = (\sigma_\theta^{-2} I + \sigma_v^{-2} H^T H)^{-1} H^T \sigma_v^{-2} Y = \left(H^T H + \frac{\sigma_v^2}{\sigma_\theta^2} I\right)^{-1} H^T Y$$

When $\dfrac{\sigma_v^2}{\sigma_\theta^2} \to 0$, $\widehat{\theta}_{LMMSE} \to \widehat{\theta}_{ML}$.

(h) Conditional bias:

$$\begin{aligned} b_{\widehat{\theta}_{LMMSE}}(\theta) &= E_{Y|\theta}\widehat{\theta}_{LMMSE} - \theta = \left(H^T H + \frac{\sigma_v^2}{\sigma_\theta^2} I\right)^{-1} H^T E_{Y|\theta} Y - \theta \\ &= \left(H^T H + \frac{\sigma_v^2}{\sigma_\theta^2} I\right)^{-1} H^T H \theta - \theta = -\frac{\sigma_v^2}{\sigma_\theta^2}\left(H^T H + \frac{\sigma_v^2}{\sigma_\theta^2} I\right)^{-1} \theta \neq 0 \end{aligned}$$

(i) Course notes: $R_{\widetilde{\theta}_{ML}\widetilde{\theta}_{ML}} = \sigma_v^2 \left(H^T H\right)^{-1}$, $R_{\widetilde{\theta}_{LMMSE}\widetilde{\theta}_{LMMSE}} = \sigma_v^2 \left(H^T H + \dfrac{\sigma_v^2}{\sigma_\theta^2}\right)^{-1}$

If we proceed now as in question (d), the difference equation between the $\widehat{\theta}_k$'s and y_k's is:

$$-\widehat{\theta}_{k-2} + (2 + \frac{\sigma_v^2}{\sigma_\theta^2})\widehat{\theta}_k - \widehat{\theta}_{k+2} = y_k - y_{k+2}, \tag{12}$$

$k = 1, \cdots, m$ and $\widehat{\theta}_k = 0$ for $k \leq 0$, $k > m$. If (12) is considered valid for any k, $\widehat{\theta}_k$ can be interpreted as the output of the filter $F(z)$ taking y_k as input, with:

$$\mathrm{F}(z) = \frac{1 - z^2}{-z^2 + 2 + \frac{\sigma_v^2}{\sigma_\theta^2} - z^{-2}} \tag{13}$$

F(z) is now stable (but non-causal) and there is no noise amplification. See the Wiener Filtering chapter for further clarifications.

Some comments:

This estimation problem corresponds to an equalization problem. The observations are of the form: $y_k =$ symbol of interest $+$ ISI terms $+$ noise terms. The symbol of interest (the one we want to estimate) at time k is here θ_k. The ISI (InterSymbol Interference) terms contain symbol contributions from time $i \neq k$, that are not of interest: here, ISI terms$=$ $-\theta_{k-2}$. The noise term is v_k. The purpose of equalization is to estimate the channel input signal θ_k by trying to reduce the ISI and noise terms.

$\mathrm{F}_{ZF}(z) = \dfrac{1}{\mathrm{H}(z)}$ is called the Zero-Forcing (ZF) equalizer: it eliminates all the ISI but without taking into account the noise contribution: $\widehat{\theta}_k = \theta_k + f_k * v_k$. As explained in question (d), we risk noise amplification.

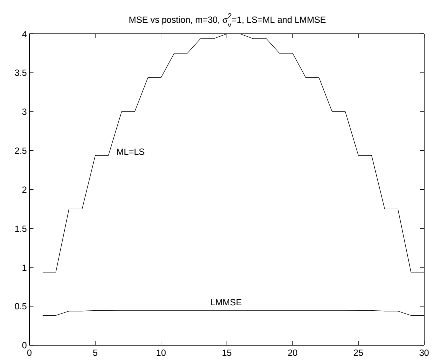
The LMMSE equalizer $\mathrm{F}_{LMMSE}(z)$ does not completely eliminate the ISI but makes a compromise between ISI and noise reduction and avoids the noise amplification phenomenon. Furthermore, the final MSE for the LMMSE equalizer is lower than for the ZF equalizer.

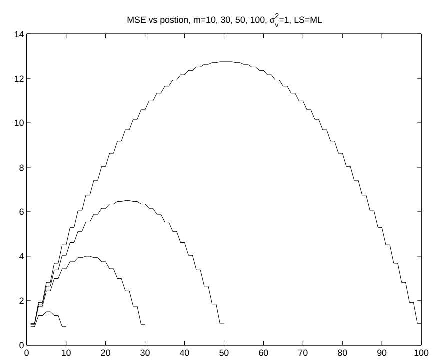
$\mathrm{F}_{ZF}(z)$ and $\mathrm{F}_{LMMSE}(z)$ are time-invariant equalizers when we have an infinite amount of data (from time $-\infty$ to $+\infty$): we will say that we are in a *continuous processing* mode. In this problem, the equalizers were developed in the case of a finite amount of data: we say that we are in a *burst processing* mode.

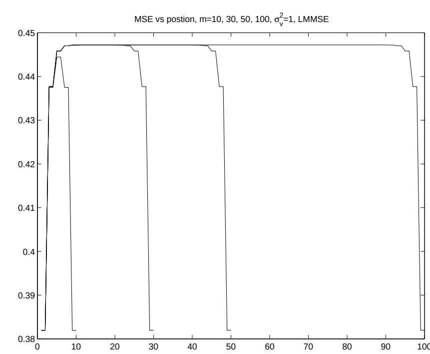
In burst mode, the ZF equalizer is $F_{ZF} = \left(H^T H\right)^{-1} H^T$. As stated in question (b), several ZF equalizers exist: the one considered here gives the lowest MSE. The LMMSE equalizer is $\mathrm{F}_{LMMSE} = \left(H^T H + \frac{\sigma_v^2}{\sigma_\theta^2} I\right)^{-1} H^T$. Line k of F_{ZF} or F_{LMMSE} is the equalizer filter of time k: $\widehat{\theta}_k = \sum_{i=1}^{m} F(k,i) y_i$. For each $\widehat{\theta}_k$, the equalizer filter varies (unlike in the continuous processing case, where it is time-invariant).

In the figures below, we plot the MSE for each $\widehat{\theta}_{k,ZF}$ $(= \widehat{\theta}_{k,ML} = \widehat{\theta}_{k,WLS})$ and $\widehat{\theta}_{k,LMMSE}$ versus the time index k. We observe:

- $\widehat{\theta}_{LMMSE}$ has a better performance than $\widehat{\theta}_{ML}$ (see first figure, on the left).
- The MSE is lower at the edges than in the middle of the burst of θ's. Indeed, the middle observations contain 2 symbols. The edge observations (k=1,2, m+1,m+2) contain only 1 symbol so that there is more information on those symbols, which are then better estimated, and because those edge symbols are better estimated, the symbols nearby get also better estimated.
- The MSE for $\widehat{\theta}_{ZF}$ grows unboundedly when m increases (see 2^{nd} figure, in the middle). For the middle symbols ($1 \ll k \ll m$) when m grows to infinity, you can consider that you have all the data available (from $-\infty$ to $+\infty$): we approach the continuous processing case and we have noise amplification.
- The MSE for $\widehat{\theta}_{LMMSE}$ remains bounded as m increases (see 3^{rd} figure, on the right). The MSE for the middle symbol tends to the MSE in the continuous case, which is bounded.







Problem 5. Poisson Process - Deterministic Parameter Setting

(a) $\{y_k\}_{k=1,\cdots,n}$ independent $\Rightarrow f(Y|\theta) = \prod_{i=1}^{n} f(y_k|\theta) = \theta^n e^{-n\theta\overline{y}}$.

(b) $\widehat{\theta}_{ML} = \arg\max_\theta f(Y|\theta) = \arg\max_\theta \ln f(Y|\theta)$

$\ln f(Y|\theta) = n \ln\theta - n\theta\overline{y}$

$$\frac{\partial \ln f(Y|\theta)}{\partial \theta} = n\frac{1}{\theta} - n\overline{y} = 0 \Rightarrow \widehat{\theta}_{ML} = \frac{1}{\overline{y}}$$
$$\frac{\partial^2 \ln f(Y|\theta)}{\partial \theta^2}_{|\theta=\widehat{\theta}_{ML}} = -n\frac{1}{\left(\widehat{\theta}_{ML}\right)^2} < 0 \Rightarrow \widehat{\theta}_{ML} \text{ is a maximum.}$$

(c) Different types of consistency can be defined. Let $\widehat{\theta}_n(Y)$ be the estimate of θ based on n observations Y.

- Mean-square consistency.
  $\widehat{\theta}_n(Y)$ is said mean-square consistent if: $\mathrm{tr} R_{\widetilde{\theta}_n\widetilde{\theta}_n} \overset{n\to\infty}{\longrightarrow} 0$.
- Strong consistency.
  $\widehat{\theta}_n(Y)$ is said strongly consistent if: $\widehat{\theta}_n(Y) \overset{n\to\infty}{\longrightarrow} \theta$ with probability 1 (w.p. 1).

In this question, you are asked to show strong consistency. (Mean square consistency is more difficult to show). The strong law of large numbers tells us that if the y_i's are independent random variables with a common mean Ey, then $\frac{1}{n}\sum_{i=1}^{n} y_i \overset{n\to\infty}{\longrightarrow} Ey$ w.p. 1.
Hence, $\overline{y} \to Ey_k = \frac{1}{\theta}$ w.p. 1. Or $\widehat{\theta}_{ML} \to \theta$ w.p. 1 : $\widehat{\theta}_{ML}$ is (strongly) consistent.

Problem 6. Variance Estimation of Zero-Mean Gaussian i.i.d. Variables

(a) $f(Y|\theta) = \frac{1}{\left(\sqrt{2\pi\theta}\right)^n} \exp\left[-\frac{1}{2\theta}\sum_{i=1}^{n} y_k^2\right].$

$$\ln f(Y|\theta) = -\frac{n}{2}\ln(2\pi) - \frac{n}{2}\left[\ln\theta + \frac{1}{\theta}\overline{y^2}\right] \qquad \left(\overline{y^2} = \frac{1}{n}\sum_{k=1}^{n} y_k^2\right)$$
$$\frac{\partial \ln f(Y|\theta)}{\partial \theta} = -\frac{n}{2}\left[\frac{1}{\theta} - \frac{1}{\theta^2}\overline{y^2}\right] = 0 \Rightarrow \widehat{\theta}_{ML} = \overline{y^2}$$
$$\left(\frac{\partial^2 \ln f(Y|\theta)}{\partial \theta^2} = -\frac{n}{2}\left[-\frac{1}{\theta^2} + \frac{2}{\theta^3}\overline{y^2}\right];\ \left.\frac{\partial^2 \ln f(Y|\theta)}{\partial \theta^2}\right|_{\theta=\widehat{\theta}_{ML}} = -\frac{n}{2\left(\overline{y^2}\right)^2} < 0 \Rightarrow \widehat{\theta}_{ML} \text{ maximum}\right).$$

(b) $b_{\widehat{\theta}_{ML}}(\theta) = E_{Y|\theta}\overline{y^2} - \theta = \overline{Ey_k^2} - \theta = 0$: $\widehat{\theta}_{ML}$ is unbiased.

(c) $E\widetilde{\theta}_{ML}^2 = \mathrm{var}\left(\widehat{\theta}_{ML}\right) = \mathrm{var}\left(\overline{y^2}\right) = \frac{1}{n}\mathrm{var}\left(y_k^2\right)$, because the y_k's are i.i.d.

$$E\widetilde{\theta}_{ML}^2 = \frac{1}{n}\left(Ey_k^4 - \left(Ey_k^2\right)^2\right) = \frac{3\theta^2 - \theta^2}{n} = \frac{2}{n}\theta^2.$$
$E\widetilde{\theta}_{ML}^2 \overset{n\to\infty}{\longrightarrow} 0$: $\widehat{\theta}_{ML}$ is (mean-square) consistent.

(d) $\widehat{\theta}_{LMMSE} = R_{\theta Y}R_{YY}^{-1}Y$

$$R_{\theta Y} = E_{Y,\theta}\theta Y^T = E_\theta E_{Y|\theta}\theta Y^T = E_\theta\theta \underbrace{E_{Y|\theta}Y^T}_{=0} = 0 \Rightarrow \widehat{\theta}_{LMMSE} = 0.$$

$$E\widetilde{\theta}^2_{LMMSE} = E\theta^2 = \lambda\int_0^\infty \theta^2 e^{-\lambda\theta}d\theta \quad = \lambda\underbrace{\left[-\frac{1}{\lambda}\theta^2 e^{-\lambda\theta}\right]_0^\infty}_{=0} + 2\int_0^\infty \theta e^{-\lambda\theta}d\theta$$

$$= \underbrace{\left[-\frac{2}{\lambda}\theta e^{-\lambda\theta}\right]_0^\infty}_{=0} + \frac{2}{\lambda}\int_0^\infty e^{-\lambda\theta}d\theta$$

(2 integrations by parts)

Hence $E\widetilde{\theta}^2_{LMMSE} = \dfrac{2}{\lambda^2}$.

(e) $\widehat{\theta}_{LMMSE} = R_{\theta X}R_{XX}^{-1}X$.

$R_{\theta x_k} = E\theta y_k^2 = E_\theta \theta \underbrace{E_{y_k|\theta}y_k^2}_{=\theta} = E_\theta\theta^2 \overset{(d)}{=} \dfrac{2}{\lambda^2}$

for $i \neq k$, $R_{x_i x_k} = Ey_i^2 y_k^2 = E_\theta E_{(y_i,y_k)|\theta}\left(y_i^2 y_k^2\right) = E_\theta \underbrace{\left(E_{y_i}y_i^2\right)}_{=\theta}\underbrace{\left(E_{y_k}y_k^2\right)}_{=\theta} = E_\theta\theta^2 = \dfrac{2}{\lambda^2}$

$R_{x_i x_i} = Ey_i^4 = E_\theta E_{y_i|\theta}y_i^4 = E_\theta 3\theta^2 = \dfrac{6}{\lambda^2}$

Hence, $R_{\theta X} = \dfrac{2}{\lambda^2}\mathbf{1}$, $R_{XX} = \dfrac{2}{\lambda^2}\left(2I + \mathbf{1}\mathbf{1}^T\right)$, where $\mathbf{1}$ is a vector of 1's.

$R_{\theta X}R_{XX}^{-1} = \mathbf{1}^T(2I+\mathbf{1}\mathbf{1}^T)^{-1} = \dfrac{1}{2}\mathbf{1}^T(I+\mathbf{1}\dfrac{1}{2}\mathbf{1}^T)^{-1} = \dfrac{1}{2}\mathbf{1}^T(I - \mathbf{1}(\mathbf{1}^T\mathbf{1}+2)^{-1}\mathbf{1}^T)$

(using the matrix inversion lemma.)

$R_{\theta X}R_{XX}^{-1} = \dfrac{1}{2}(2 - \dfrac{n}{n+2})\mathbf{1}^T = \dfrac{1}{n+2}\mathbf{1}^T$

$\widehat{\theta}_{LMMSE} = \dfrac{n}{n+2}\underbrace{\dfrac{1}{n}\mathbf{1}^T X}_{\overline{y^2}} = \dfrac{n}{n+2}\overline{y^2}$ (independent of λ because $Ey_i^4 \sim (Ey_i^2)^2$)

Note: $f(\theta)$ says that small values for θ are more likely, so it is not surprising that $\widehat{\theta}_{LMMSE} < \widehat{\theta}_{ML}$.

(f) Conditional mean of the estimator: $E_{Y|\theta}\widehat{\theta}_{LMMSE} = \dfrac{n}{n+2}E_{Y|\theta}\overline{y^2} = \dfrac{n}{n+2}\theta$.

Contiditional bias: $b_{\widehat{\theta}_{LMMSE}}(\theta) = \dfrac{n}{n+2}\theta - \theta = -\dfrac{2}{n+2}\theta$.

Average bias: $E_\theta b_{\widehat{\theta}_{LMMSE}}(\theta) = -\dfrac{2}{n+2}E_\theta\theta = -\dfrac{2}{n+2}\lambda\underbrace{\int_0^\infty \theta e^{-\lambda\theta}d\theta}_{1/\lambda^2} = -\dfrac{2}{n+2}\dfrac{1}{\lambda} \overset{n\to\infty}{\longrightarrow} 0.$

$\widehat{\theta}_{LMMSE}$ is asymptotically unbiased.

(g) MSE $= E_{Y,\theta}\widetilde{\theta}^2_{LMMSE} = E_{Y,\theta}\left(\widehat{\theta}-\theta\right)^2 = E_\theta E_{Y|\theta}\left(\widehat{\theta}-\theta\right)^2$ (we denote: $\widehat{\theta} = \widehat{\theta}_{LMMSE}$)

$$
\begin{aligned}
E_{Y|\theta}(\widehat{\theta}-\theta)^2 &= E_{Y|\theta}\left[\left(\widehat{\theta}-E_{Y|\theta}\widehat{\theta}\right)+\left(E_{Y|\theta}\widehat{\theta}-\theta\right)\right]^2 \\
&= E_{Y|\theta}(\widehat{\theta}-E_{Y|\theta}\widehat{\theta})^2+\left(E_{Y|\theta}\widehat{\theta}-\theta\right)^2+2\left(E_{Y|\theta}\widehat{\theta}-\theta\right)\underbrace{E_{Y|\theta}(\widehat{\theta}-E_{Y|\theta}\widehat{\theta})}_{=0} \\
&= E_{Y|\theta}\left(\frac{n}{n+2}\overline{y^2}-\frac{n}{n+2}\theta\right)^2+\left(\frac{n}{n+2}\theta-\theta\right)^2 \\
&= \left(\frac{n}{n+2}\right)^2 E_{Y|\theta}\left(\overline{y^2}-\theta\right)^2+\left(\frac{2}{n+2}\right)^2\theta^2 \\
&= \left(\frac{n}{n+2}\right)^2 \mathrm{var}_{Y|\theta}\left(\overline{y^2}\right)^2+\left(\frac{2}{n+2}\right)^2\theta^2 \\
&\overset{(c)}{=} \left(\frac{n}{n+2}\right)^2\frac{2\theta^2}{n}+\left(\frac{2}{n+2}\right)^2\theta^2 \\
&= \frac{1}{(n+2)^2}(2n+4)\theta^2 \\
&= \frac{2}{n+2}\theta^2 \\
&< \frac{2}{n}\theta^2 \text{ in (c) for ML!}
\end{aligned}
$$

$$
MSE = E\widetilde{\theta}^2_{LMMSE} = E_\theta\left(\frac{2}{n+2}\theta^2\right) = \frac{2}{n+2}E_\theta\theta^2 = \frac{2}{n+2}\frac{2}{\lambda^2}.
$$

Problem 7. Signal Amplitude Estimation in White Noise

You can use the results in the course notes to answer this problem. In this solution, we give proofs for certain questions, as a review, but you could omit them.

(a) $\widehat{\theta}_{ML} = (S^T C_{VV}^{-1} S)^{-1} S^T C_{VV}^{-1} Y$ (deterministic linear model)

(b) $b_{\widehat{\theta}_{ML}}(\theta) = 0$ since $EV = 0$ (deterministic linear model).

Proof. $Y = S\theta + V \Rightarrow$

$$
\begin{aligned}
\widehat{\theta}_{ML}-\theta &= \underbrace{\left(S^T C_{VV}^{-1} S\right)^{-1} S^T C_{VV}^{-1} S}_{=I}\,\theta+\left(S^T C_{VV}^{-1} S\right)^{-1} S^T C_{VV}^{-1} V-\theta \\
\widehat{\theta}_{ML}-\theta &= \left(S^T C_{VV}^{-1} S\right)^{-1} S^T C_{VV}^{-1} V \\
b_{\widehat{\theta}_{ML}}(\theta) &= E_{Y|\theta}\widehat{\theta}_{ML}-\theta = \left(S^T C_{VV}^{-1} S\right)^{-1} S^T C_{VV}^{-1} E_{V|\theta}V = 0, \text{ because } EV = 0.
\end{aligned}
$$

(c) $E\widetilde{\theta}^2_{ML} = \left(S^T C_{VV}^{-1} S\right)^{-1}$ (deterministic linear model)

Proof.

$$
\begin{aligned}
E_{Y|\theta}\widetilde{\theta}^2_{ML} &= E_{Y|\theta}\widetilde{\theta}_{ML}\widetilde{\theta}^T_{ML} \\
E_{Y|\theta}\widetilde{\theta}^2_{ML} &= \left(S^T C_{VV}^{-1} S\right)^{-1} S^T C_{VV}^{-1}\underbrace{EVV^T}_{C_{VV}} C_{VV}^{-1} S\left(S^T C_{VV}^{-1} S\right)^{-1} \\
E_{Y|\theta}\widetilde{\theta}^2_{ML} &= \underbrace{\left(S^T C_{VV}^{-1} S\right)^{-1} S^T C_{VV}^{-1} S}_{=I}\left(S^T C_{VV}^{-1} S\right)^{-1}
\end{aligned}
$$

(d) $\widehat{\theta}_{ML} = (S^TS)^{-1}S^TY$, $b_{\widehat{\theta}_{WLS}}(\theta) = 0$, $E\widetilde{\theta}^2_{ML} = \sigma_v^2 \left(S^TS\right)^{-1}$.

(e) $\widehat{\theta}_{ML}$ is (mean-square) consistent if $E\widetilde{\theta}^2_{ML} = \dfrac{\sigma_v^2}{S^TS} = \dfrac{\sigma_v^2}{\sum\limits_{k=1}^{n} s_k^2} \xrightarrow{n\to\infty} 0$.

$\widehat{\theta}_{ML}$ is consistent if $\sum\limits_{k=1}^{+\infty} s_k^2 = +\infty$.

(f) $\widehat{\theta}_{LMMSE} = \left(\dfrac{1}{\sigma_\theta^2} + \dfrac{1}{\sigma_v^2}S^TS\right)^{-1} \dfrac{1}{\sigma_v^2}S^TY = \left(S^TS + \dfrac{\sigma_v^2}{\sigma_\theta^2}\right)^{-1} S^TY$
(Bayesian linear model)

(g) $b_{\widehat{\theta}_{LMMSE}}(\theta) = \dfrac{S^T}{S^TS + \dfrac{\sigma_v^2}{\sigma_\theta^2}} \underbrace{E_{Y|\theta}Y}_{S\theta} - \theta = -\dfrac{\sigma_v^2}{\sigma_\theta^2} \dfrac{1}{S^TS + \dfrac{\sigma_v^2}{\sigma_\theta^2}} \theta \neq 0$: $\widehat{\theta}_{LMMSE}$ is conditionally unbiased.

$E_\theta b_{\widehat{\theta}_{LMMSE}}(\theta) = -\dfrac{\sigma_v^2}{\sigma_\theta^2} \dfrac{1}{S^TS + \dfrac{\sigma_v^2}{\sigma_\theta^2}} \underbrace{E_\theta \theta}_{=0} = 0$: $\widehat{\theta}_{LMMSE}$ is unbiased on the average.

(h) $E\widetilde{\theta}^2_{ML} = \dfrac{1}{\dfrac{1}{\sigma_\theta^2} + \dfrac{1}{\sigma_v^2}S^TS} = \dfrac{\sigma_v^2}{S^TS + \dfrac{\sigma_v^2}{\sigma_\theta^2}}$ (Bayesian linear model)

$S^TS + \dfrac{\sigma_v^2}{\sigma_\theta^2} > S^TS \Rightarrow MSE_{LMMSE} < MSE_{ML}$

2 Spectrum Estimation

2.1 Non-Parametric Spectrum Estimation

Problem 8. Periodogram spectral leakage in the case of a sinusoid

The periodogram of y_k is given by

$$\widehat{S}_{PER}(f) = \frac{1}{N}\left|\sum_{k=0}^{N-1} A\cos(2\pi f_0 k + \phi) e^{-j2\pi f k}\right|^2 ,$$

when evaluated at frequencies multiple of $\frac{1}{N}$, ($f_n = \frac{n}{N}$) it becomes

$$\widehat{S}_{PER}(f_n) = \frac{1}{N}\left|\sum_{k=0}^{N-1} A\cos(2\pi f_0 k + \phi) e^{-j\frac{2\pi}{N}nk}\right|^2 .$$

Since $f_0 = \frac{m}{N}$ with m being an integer, we have

$$A\cos(2\pi f_0 k + \phi) = \frac{A}{2}\left(e^{j\left(\frac{2\pi}{N}mk+\phi\right)} + e^{-j\left(\frac{2\pi}{N}mk+\phi\right)}\right) .$$

Hence the periodogram becomes

$$\begin{aligned}
\widehat{S}_{PER}(f_n) &= \frac{1}{N}\left|\frac{Ae^{j\phi}}{2}\sum_{k=0}^{N-1} e^{j\frac{2\pi}{N}(m-n)k} + \frac{Ae^{-j\phi}}{2}\sum_{k=0}^{N-1} e^{-j\frac{2\pi}{N}(m+n)k}\right|^2 \\
&= \frac{1}{N}\left|\frac{Ae^{j\phi}}{2}S_1 + \frac{Ae^{-j\phi}}{2}S_2\right|^2 . \qquad (14)
\end{aligned}$$

Now,

$$S_1 = \begin{cases} N & \text{if } n = m \\ \\ 0 & \text{if } n \neq m \end{cases} , \quad S_2 = \begin{cases} N & \text{if } n = N-m \ \text{ or } \ n = m = 0 \\ \\ 0 & \qquad\qquad \text{otherwise.} \end{cases} \qquad (15)$$

So clearly, when $f_0 = \frac{m}{N}$, there are two nonzero points of the periodogram if $m \neq 0$, namely $\widehat{S}_{PER}(\frac{m}{N})$ and $\widehat{S}_{PER}(\frac{N-m}{N})$, and the priodogram is nonzero in only one point for $m = 0$, namely $\widehat{S}_{PER}(0)$. When $f_0 \neq \frac{m}{N}$ for any m, there are non-zero contributions at all f_n.
We could obtain this result in an alternative fashion if we use the fact that the periodogram can be considered as the output at time $k = 0$ of an anticausal FIR filter consisting of a modulated rectangular window (see course notes).

$$H(f) = \frac{1}{N}\sum_{k=0}^{N-1} e^{j2\pi fk} .$$

In the present case, the input to the system consists of two complex exponentials and we know that such functions are eigenfunctions for linear time invariant systems. Hence, the output of the modulated filter $H(f - f_n)$ with input $\frac{Ae^{j\phi}}{2}e^{j2\pi f_0 k} + \frac{Ae^{-j\phi}}{2}e^{-j2\pi f_0 k}$ is

$$\frac{Ae^{j\phi}}{2}H(f_0 - f_n)e^{j2\pi f_0 k} + \frac{Ae^{-j\phi}}{2}H(f_0 + f_n)e^{-j2\pi f_0 k} .$$

Taking the output at time $k = 0$, we obtain the periodogram as

$$\widehat{S}_{PER}(f_n) = N\left|\frac{Ae^{j\phi}}{2}H(f_0 - f_n) + \frac{Ae^{-j\phi}}{2}H(f_0 + f_n)\right|^2 .$$

A plot of $|H(f - f_0)|$ is given in the course notes. We can see that if $f_0 = \frac{m}{N}$, frequency f_n will fall in the "holes" of $H(f)$ except for f_m and f_{N-m} and that if $f_0 \neq \frac{m}{N}$ then $H(f - f_0)$ and $H(f + f_0)$ get sampled at frequencies where they are non-zero.

Problem 9. Bias and Variance of the (averaged) Periodogram on White Noise

(a) $S_{yy}(f) = \sigma_y^2$.

(b) Consider $z_k = \frac{1}{\sigma_y\sqrt{N}}\sum_{k=0}^{N-1} y_k$. Since, z_k is proportional to a sum of independent Gaussian random variables, z_k is a Gaussian random variable.

$$\begin{aligned} Ez_k &= \tfrac{1}{\sigma_y\sqrt{N}} \sum_{k=0}^{N-1} Ey_k = 0 \\ \mathrm{var} z_k &= Ez_k^2 = \tfrac{1}{\sigma_y^2 N} \sum_{k=0}^{N-1} \mathrm{var} y_k = \frac{1}{\sigma_y^2 N} N\sigma_y^2 = 1 \ , \end{aligned}$$

which shows that $\frac{1}{\sigma_y\sqrt{N}} \sum_{k=0}^{N-1} y_k \sim \mathcal{N}(0,1)$. Note that $Ez_k^4 = 3$.

(c) $\widehat{S}_{PER}(0) = \frac{1}{N}\left(\sigma_y\sqrt{N}z_k\right)^2 = \sigma_y^2 z_k^2$, it follows

$$\begin{aligned} E\widehat{S}_{PER}(0) &= \sigma_y^2 E z_k^2 = \sigma_y^2 \\ \mathrm{var}\widehat{S}_{PER}(0) &= E\widehat{S}^2_{PER}(0) - \left(E\widehat{S}_{PER}(0)\right)^2 = E\sigma_y^4 z_k^4 - \sigma_y^4 = 2\sigma_y^4 \ , \end{aligned}$$

clearly, the variance does not go to 0 as $N \to \infty$.

(d) Consider $\widehat{S}_{AVPER}(0) = \frac{1}{N} \sum_{n=0}^{N-1} y_n^2$.

$$E\widehat{S}_{AVPER}(0) = \frac{1}{N} \sum_{n=0}^{N-1} Ey_n^2 = \sigma_y^2 \ ,$$

$$E\widehat{S}^2_{AVPER}(0) = E\frac{1}{N^2} \sum_{l=0}^{N-1} \sum_{n=0}^{N-1} y_l^2 y_n^2 = \frac{1}{N^2} \sum_{l=0}^{N-1} \sum_{l=0}^{N-1} Ey_l^2 y_n^2 \ .$$

Now, y_k being an i.i.d. sequence,

$$Ey_l^2 y_n^2 = \begin{cases} Ey_l^4 = 3\sigma_y^4 & \text{when } n = l \\ Ey_l^2 Ey_n^2 = \sigma_y^4 & \text{when } n \neq l \ . \end{cases}$$

So

$$E\widehat{S}^2_{AVPER}(0) = \frac{1}{N^2}\left(3N\sigma_y^4 + N(N-1)\sigma_y^4\right) = \frac{N+2}{N}\sigma_y^4 \ ,$$

and the variance is

$$\mathrm{var}\widehat{S}_{AVPER}(0) = E\widehat{S}^2_{AVPER}(0) - \left(E\widehat{S}_{AVPER}(0)\right)^2 = \frac{2}{N}\sigma_y^2 \ ,$$

and

$$\lim_{N\to\infty} \mathrm{var}\widehat{S}_{AVPER}(0) = 0 \ .$$

The averaged periodogram (considered here) is a consistent estimator while the periodogram is not consistent.

Problem 10. Periodogram Mean for Sinusoids in Colored Noise

(i) We show that y_k is a stationary process as a sum of three independent stationary processes: v_k and the two sinusoids. Since $v_k = \mathrm{H}(q)\, e_k \;=\; \sum_{i=0}^{4} h_i\, q^{-i}\, e_k$ is the stationary process e_k filtered by a linear time-invariant filter, v_k is a stationary process. Now, each one of the two sinusoids can be written as :

$$x_k = A\, \cos(2\pi f k + \phi)$$

where A and ϕ are f, and ϕ is random with a uniform distribution over $[0, 2\pi]$. Consider a shift in time k_0, then

$$\begin{aligned}
x_{k-k_0} &= A\, \cos(2\pi f(k-k_0) + \phi) \\
&= A\, \cos(2\pi f k + \phi - 2\pi f k_0) \\
&= A\, \cos(2\pi f k + \underbrace{(\phi - 2\pi f k_0) \mathrm{mod} 2\pi}_{\psi}) \\
&= A\, \cos(2\pi f k + \psi),
\end{aligned}$$

and ψ is uniformly distributed over $[0, 2\pi]$. Hence the distribution of the whole process x_k (not just one sample) remains unchanged after a shift in time, hence x_k is a stationary process.

So y_k is a stationary process as a sum of independent stationary processes.

(ii) The covariance sequence is

$$\begin{aligned}
r_{yy}(n) &= \frac{A_1^2}{2}\cos(2\pi f_1 n) + \frac{A_2^2}{2}\cos(2\pi f_2 n) + r_{vv}(n), \\
r_{vv}(n) &= h_n * h_{-n} = \begin{bmatrix} -1 & -4 & -4 & 4 & 10 & 4 & -4 & -4 & -1 \end{bmatrix}.
\end{aligned}$$

(iii) The power spectral density function is

$$S_{yy}(f) = \frac{A_1^2}{4}\delta(f - f_1) + \frac{A_1^2}{4}\delta(f + f_1) + \frac{A_2^2}{4}\delta(f - f_2) + \frac{A_2^2}{4}\delta(f + f_2) + |\mathrm{H}(f)|^2 .$$

(iv) Introduce $w'_{-n} = w_n e^{-j2\pi f n} \Rightarrow W'(-f_1) = W(f_1 + f) \Rightarrow W'(f_1) = W(f - f_1)$.

$$\begin{aligned}
\widehat{S}_{PER,w}(f) &= \frac{c_{N,w}}{N}\left|\sum_{n=0}^{N-1} w_n y_n e^{-j2\pi f n}\right|^2 \\
&= \frac{c_{N,w}}{N}\left|\sum_{n=0}^{N-1} w'_{-n} y_n\right|^2 = \frac{c_{N,w}}{N}\left|\left[\sum_{n=0}^{N-1} w'_{k-n} y_n\right]_{k=0}\right|^2 \\
&= \frac{c_{N,w}}{N}\left|\int_{-\frac{1}{2}}^{\frac{1}{2}} W'(f_1) Y(f_1) df_1\right|^2 \\
&= \frac{c_{N,w}}{N}\left(\int_{-\frac{1}{2}}^{\frac{1}{2}} W'(f_1) Y(f_1) df_1\right)\left(\int_{-\frac{1}{2}}^{\frac{1}{2}} W'^*(f_2) Y^*(f_2) df_2\right) \\
&= \frac{c_{N,w}}{N}\int_{-\frac{1}{2}}^{\frac{1}{2}} W'(f_1)\left(\int_{-\frac{1}{2}}^{\frac{1}{2}} W'^*(f_2) Y(f_1) Y^*(f_2) df_2\right) df_1
\end{aligned}$$

$$
\begin{aligned}
E\,\widehat{S}_{PER,w}(f) &= \frac{c_{N,w}}{N}\int_{-\frac{1}{2}}^{\frac{1}{2}} W'(f_1)\left(\int_{-\frac{1}{2}}^{\frac{1}{2}} W'^*(f_2)\; E\,Y(f_1)Y^*(f_2)df_2\right)df_1\\
&= \frac{c_{N,w}}{N}\int_{-\frac{1}{2}}^{\frac{1}{2}} W'(f_1)\left(\int_{-\frac{1}{2}}^{\frac{1}{2}} W'^*(f_2)S_{yy}(f_1)\delta_1(f_1-f_2)df_2\right)df_1\\
&= \frac{c_{N,w}}{N}\int_{-\frac{1}{2}}^{\frac{1}{2}} W'(f_1)S_{yy}(f_1)\underbrace{\left(\int_{-\frac{1}{2}}^{\frac{1}{2}} W'^*(f_2)\delta_1(f_1-f_2)df_2\right)}_{W'^*(f_1)}df_1\\
&= \frac{c_{N,w}}{N}\int_{-\frac{1}{2}}^{\frac{1}{2}} \left|W'(f_1)\right|^2 S_{yy}(f_1)df_1\\
&= \frac{c_{N,w}}{N}\int_{-\frac{1}{2}}^{\frac{1}{2}} \left|W(f-f_1)\right|^2 S_{yy}(f_1)df_1\\
&= \frac{c_{N,w}}{N}\left|W(f)\right|^2 * S_{yy}(f)\\
&= \frac{c_{N,w}}{N}\mathcal{F}\left\{r_{ww}(k)r_{yy}(k)\right\}
\end{aligned}
$$

where $r_{ww}(k) = w_k * w_{-k}$.

(v) $E\widehat{S}_{PER,w}(f) = \frac{c_{N,w}}{N}\mathcal{F}\left\{r_{ww}(k)r_{yy}(k)\right\} \overset{N\to\infty}{\longrightarrow} \frac{c_{N,w}}{N}r_{ww}(0)\underbrace{\mathcal{F}\left\{r_{yy}(k)\right\}}_{S_{yy}(f)}.$

For unbiasedness $\frac{c_{N,w}}{N}r_{ww}(0) = 1 \Rightarrow$ choose $c_{N,w} = \frac{N}{r_{ww}(0)} = \frac{N}{\|w\|^2} = \frac{N}{\sum_{k=0}^{N-1} w_k^2}$

(vi) We have

$$
\begin{aligned}
E\widehat{S}_{PER,w}(f) &= \tfrac{1}{\|w\|^2}\left\{\tfrac{A_1^2}{4}\left|W(f-f_1)\right|^2 + \tfrac{A_1^2}{4}\left|W(f+f_1)\right|^2 + \tfrac{A_2^2}{4}\left|W(f-f_2)\right|^2 + \tfrac{A_2^2}{4}\left|W(f+f_1)\right|^2\right\}\\
&+ \tfrac{1}{\|w\|^2}\,\mathcal{F}\{r_{ww}(k)\quad \underbrace{r_{vv}(k)}_{\text{only } \neq 0 \text{ for } |k|\leq 4}\quad\}.
\end{aligned}
$$

For $r_{ww}(k)$, work in the time domain and transfrom (if work in the freq-domain directly $\Rightarrow$ would need a convolution between two continuous functions of frequency: $|W(f)|^2$ and $S_{vv}(f) = |H(f)|^2$).

2.2 Parametric Spectrum Estimation

Problem 11. Autoregressive Processes

(a) First of all, we know $r_{-k} = r_k$, so we have to compute $\{r_k, 0 \leq k \leq 100\}$. For $0 \leq k \leq N$, we can use the Yule-Walker equation *i.e.*

$$
\begin{bmatrix} r_0 & r_1 & \cdots & r_N \\ r_1 & r_0 & \cdots & r_{N-1} \\ \vdots & \vdots & \ddots & \vdots \\ r_N & r_{N-1} & \cdots & r_0 \end{bmatrix}
\begin{bmatrix} 1 \\ a_1 \\ \vdots \\ a_N \end{bmatrix}
=
\begin{bmatrix} \sigma_e^2 \\ 0 \\ \vdots \\ 0 \end{bmatrix},
\tag{16}
$$

the r_i's being the unknowns, we have to rewrite the Yule-Walker equations in a more appropriate form, *viz.*

$$\left(\begin{bmatrix} 0 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ 0 & a_1 & 1 & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & 0 \\ 0 & a_{N-1} & \cdots & a_1 & 1 \end{bmatrix} + \begin{bmatrix} 1 & a_1 & \cdots & a_{N-1} & a_N \\ a_1 & a_2 & \cdots & a_N & 0 \\ a_2 & a_3 & \cdot^{\cdot^{\cdot}} & 0 & 0 \\ \vdots & \cdot^{\cdot^{\cdot}} & \cdot^{\cdot^{\cdot}} & & \vdots \\ a_N & 0 & \cdots & \cdots & 0 \end{bmatrix} \right) \begin{bmatrix} r_0 \\ r_1 \\ \vdots \\ r_N \end{bmatrix} = \begin{bmatrix} \sigma_e^2 \\ 0 \\ \vdots \\ 0 \end{bmatrix} , \tag{17}$$

this form allows the computation of the first $N+1$ correlation lags. In order to compute the rest, we can use the following relationship (Yule-Walker equations)

$$r_k + a_1 r_{k-1} + \ldots + a_N r_{k-N} = 0 \ , \quad k > N.$$

Finally, denoting by $\mathcal{A}$ the matrix appearing in the left hand side of (17), the procedure for computing correlation lags associated to $\{a_1, \ldots, a_N, \sigma_e^2\}$ is as follows

- Compute the matrix $\mathcal{A}$.
- Compute first column of its inverse: $\eta = \mathcal{A}^{-1} \left[10 \cdots 0\right]^T$.
- $\left[r_0 r_1 \cdots r_N\right]^T = \sigma_e^2 \eta$.
- For $k > N \quad r_k = -a_1 r_{k-1} - a_2 r_{k-2} - \ldots - a_N r_{k-N}$.
- $r_{-k} = r_k$.

(b) In the case of an AR(1) process, we have

$$\mathcal{A} = \begin{bmatrix} 1 & a_1 \\ a_1 & 1 \end{bmatrix} . \tag{18}$$

We have to solve

$$\begin{array}{rcl} r_0 + a_1 r_1 & = & \sigma_e^2 \\ a_1 r_o + r_1 & = & 0 \end{array} ,$$

and for $k > 1$

$$r_k = -a_1 r_{k-1} = (-a_1)^{k-1} r_1 \ .$$

The solution for $k \geq 0$ is $r_k = \dfrac{(-a_1)^k}{1 - a_1^2} \sigma_e^2$. Finally, the general solution is given by

$$r_k = \frac{(-a_1)^{|k|}}{1 - a_1^2} \sigma_e^2 \ . \tag{19}$$

Problem 12. Linear Prediction and Triangular Matrix Factorization

(i)

$$f_{i,i} = y_i + \sum_{l=1}^{i} A_{i,l} y_{i-l}$$

and we know from the orthogonality principle that $f_{i,i}$ is orthogonal to the space spanned by $\{y_{i-1} \cdots y_0\}$ *i.e.*

$$E f_{i,i} y_{i-l} = 0 \text{ for } l = 1, \ldots, i \ . \tag{20}$$

Now, consider the case where $i > j$, we have

$$E f_{i,i} f_{j,j} = E f_{i,i} \left(y_j + \sum_{l=1}^{j} A_{j,l} y_{j-l} \right) = E f_{i,i} y_j + \sum_{l=1}^{j} E f_{i,i} y_{j-l} \ ,$$

Since $i > j$: $E f_{i,i} y_{j-l} = 0$ for $l = 0, \ldots, j$ (see (20). When $i < j$ the same results hold and when $i = j$, $E f_{i,i} f_{j,j} = \sigma^2_{f,i}$. Finally

$$E(f_{i,i} f_{j,j}) \ = \ \sigma^2_{f,i} \delta_{ij} \tag{21}$$

Consider $LY = F$. we have: $EFF^T = ELYY^T L^T = LR_{n+1} L^T$. From (21) we know that EFF^T is a diagonal matrix: $EFF^T = diag\{\sigma^2_{f,0},\ \sigma^2_{f,1},\ \cdots, \sigma^2_{f,n}\}$. Hence

$$LR_{n+1} L^T = D = diag\{\sigma^2_{f,0},\ \sigma^2_{f,1},\ \cdots, \sigma^2_{f,n}\} \ . \tag{22}$$

(ii) By inverting (22), one has $L^{-T} R^{-1}_{n+1} L^{-1} = D^{-1}$. Premultiplying by L^T and postmultiplying by L, it follows that

$$R^{-1}_{n+1} = L^T D^{-1} L \ . \tag{23}$$

From (22), we have

$$\det L \det\ R_{n+1} \det L^T = \det\ R_{n+1} = \det D = \prod_{i=0}^{n} \sigma^2_{f,i} \ . \tag{24}$$

(iii) For a stationary process y_k satisfying $y_k + a\, y_{k-1} \ = \ e_k$ where e_k is white noise with variance σ^2, we have the following

$$\sigma^2_{f,0} = E(y_k^2) = E\left(e_k - a y_{k-1}\right)^2 = \sigma^2 + a^2 \sigma^2_{f,0} \ .$$

So

$$\sigma^2_{f,0} = \frac{\sigma^2}{1 - a^2} \ . \tag{25}$$

Note that $E e_k y_{k-1} = 0$ because y_{k-1} depends on $e_{k-1}, e_{k-2}, \ldots$ and e_k is a white noise.
For an AR(1) process with parameter a, the optimal forward prediction filter is $A_1 = [\ 1\ a\ ]^T$ and the optimal forward prediction filters of higher orders are equal to A_1 filled with the required number of zeros. We know also that the optimal forward prediction variance is $\sigma^2_{f,1} = \sigma^2$ and that $\sigma^2_{f,i} = \sigma^2$ for $i \geq 1$. From these two remarks, we have

$$L = \begin{bmatrix} 1 & 0 & \cdots & \cdots & 0 \\ a & 1 & 0 & & \vdots \\ 0 & a & 1 & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & 0 \\ 0 & 0 & \cdots & a & 1 \end{bmatrix} , \tag{26}$$

and

$$D = diag\{\frac{\sigma^2}{1-a^2}, \sigma^2, \dots, \sigma^2\} \, . \tag{27}$$

From (24), we have

$$\det R_{n+1} = \frac{\sigma^{2(n+1)}}{1-a^2} \, , \tag{28}$$

and the covariance matrix inverse is computed as in (23) and gives

$$R_{n+1}^{-1} = \sigma^{-2} \begin{bmatrix} 1 & a & 0 & \cdots & 0 \\ a & 1+a^2 & a & & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & a & 1+a^2 & a \\ 0 & \cdots & 0 & a & 1 \end{bmatrix} . \tag{29}$$

Note that the covariance matrix inverse is a banded matrix of bandwidth 3 (diagonals). This result can be generalized as follows: the inverse covariance matrix of an AR(N) process is a symmetric banded matrix of bandwidth $2N+1$.

Problem 13. Linear Prediction of a Constant Signal in White Noise

(i)

$$\begin{aligned} r_{yy}(i) &= E y_k y_{k-i} = E(s_k+v_k)(s_{k-i}+v_{k-i}) = E s_k s_{k-i} + E v_k v_{k-i} \\ &= \sigma_s^2 + \delta_{i,0}\, \sigma_v^2 \;=\; (1+\delta_{i,0}\,\rho)\sigma_s^2 \, , \end{aligned} \tag{30}$$

where $\delta_{i,0} = 1$ if $i=0$ and $\delta_{i,0} = 0$ if $i \neq 0$.
(ii) The covariance matrix is equal to

$$R_n = \sigma_s^2 \left(\rho I_n + \mathbf{1}_n \mathbf{1}_n^T\right) \, . \tag{31}$$

where $\mathbf{1}_n = [1 \cdots 1]^T$ of dimension $n \times 1$.
(iii) The optimal forward prediction filter of order n, A_n is the solution to the Yule-Walker equations:

$$R_{n+1} A_n = \begin{bmatrix} \sigma_{f,n}^2 \\ 0 \\ \vdots \\ 0 \end{bmatrix} ,$$

$A_n = [1 A_{n,1} \cdots A_{n,n}]^T$. Let's compute the coefficients $A_{n,i}$. They are the solution of the system

$$R_n \begin{bmatrix} A_{n,1} \\ \vdots \\ A_{n,n} \end{bmatrix} = - \begin{bmatrix} r_{yy}(1) \\ \vdots \\ r_{yy}(n) \end{bmatrix} . \tag{32}$$

Using the matrix inversion lemma, we have

$$\begin{aligned}
R_n^{-1} &= \sigma_s^{-2}\left(\rho I_n + \mathbf{1}_n\mathbf{1}_n^T\right)^{-1} \\
&= \sigma_s^{-2}\left(\rho^{-1}I_n - \rho^{-1}I_n\mathbf{1}_n\left(1+\mathbf{1}_n^T\rho^{-1}I_n\mathbf{1}_n\right)^{-1}\mathbf{1}_n\rho^{-1}I_n\right) \\
&= \sigma_s^{-2}\rho^{-1}\left(I_n - \frac{1}{n+\rho}\mathbf{1}_n\mathbf{1}_n^T\right) . \qquad (33)
\end{aligned}$$

Hence, the solution to (32) is

$$\begin{aligned}
\begin{bmatrix} A_{n,1} \\ \vdots \\ A_{n,n} \end{bmatrix} &= -\sigma_s^{-2}\rho^{-1}\left(I_n - \frac{1}{n+\rho}\mathbf{1}_n\mathbf{1}_n^T\right)\sigma_s^2\mathbf{1}_n = -\rho^{-1}\left(\mathbf{1}_n - \frac{n}{n+\rho}\mathbf{1}_n\right) \\
&= -\frac{1}{n+\rho}\mathbf{1}_n . \qquad (34)
\end{aligned}$$

(iv) The predicted value of y_k is

$$\widehat{y}_k = -\left[A_{n,1}\cdots A_{n,n}\right]\begin{bmatrix} y_{k-1} \\ \vdots \\ y_{k-n} \end{bmatrix} = \frac{1}{n+\rho}\sum_{i=1}^{n} y_{k-i} . \qquad (35)$$

When $\rho \to 0$ *i.e.* $\sigma_s^2 \gg \sigma_v^2$, the mean of the last N samples of the signal is taken in order to predict the value y_k. In this situation, the measure is nearly noiseless and the mean will be the constant C (realization of the random variable). In the opposite case, when $\rho \to \infty$ *i.e.* $\sigma_v^2 \gg \sigma_s^2$, the noise is too important and the best we can do is to take the mean of C which is 0. In this case, the measurements, which are the n past samples of the signal, are not considered.

(v) The prediction error variance is given by

$$\begin{aligned}
\sigma_{f,n}^2 &= r_{yy}(0) + \left[r_{yy}(1)\cdots r_{yy}(N)\right]\begin{bmatrix} A_{n,1} \\ \vdots \\ A_{n,n} \end{bmatrix} = \sigma_s^2(1+\rho) + \sigma_s^2\mathbf{1}_n^T\left(\frac{-1}{n+\rho}\mathbf{1}_n\right) \\
&= \sigma_s^2(1+\rho) - \frac{n\sigma_s^2}{n+\rho} \qquad (36) \\
&= \rho\frac{n+\rho+1}{n+\rho}\sigma_s^2 \;=\; \sigma_v^2(1+\frac{1}{n+\rho}) .
\end{aligned}$$

When $n \to \infty$ we have

$$\lim_{n\to\infty}\sigma_{f,n}^2 = \rho\sigma_s^2 = \sigma_v^2 .$$

The prediction error variance is strictly decreasing w.r.t. n. The minimum achievable error variance is the variance of the noise v_k.

(vi) Note: the “ hat” in $\widehat{S}_{yy}(f)$ does not refer to estimation but to approximation due to autoregressive modeling.

$$\widehat{S}_{yy}(0) = \frac{\sigma_{f,n}^2}{|A_n(0)|^2} ,$$

since $A_n(0) = \sum_{i=0}^{n} A_{n,i} = 1 - \frac{n}{n+\rho} = \frac{\rho}{n+\rho}$ using (36) we find

$$\widehat{S}_{yy}(0) = \frac{1}{\rho}(n+\rho+1)(n+\rho)\sigma_s^2 \ . \tag{37}$$

Clearly, $\widehat{S}_{yy}(0)$ is not a good estimator of σ_s^2 because it is increasing w.r.t. n (without bound).

(vii) $A_n(\frac{1}{2}) = \sum_{i=0}^{n} A_{n,i}(-1)^i = 1 - \frac{1}{n+\rho}\sum_{i=1}^{n}(-1)^i$.

It appears that when n is even $\widehat{S}_{yy}(\frac{1}{2}) = 1$ and when n is odd $A_n(\frac{1}{2}) = 1 - \frac{1}{n+\rho} = \frac{n+\rho+1}{n+\rho}$, then

$$\widehat{S}_{yy}(\frac{1}{2}) = \begin{cases} \dfrac{n+\rho+1}{n+\rho}\sigma_v^2 & \text{if } n \text{ even} \\ \\ \dfrac{n+\rho}{n+\rho+1}\sigma_v^2 & \text{if } n \text{ odd} \ . \end{cases} \tag{38}$$

So in any case:

$$\lim_{n\to\infty} \widehat{S}_{yy}(\frac{1}{2}) = \sigma_v^2.$$

Because $y_k = C + v_k$ we know that the spectral density function is $S_{yy}(f) = \sigma_s^2\delta(f) + \sigma_v^2$ or hence $S_{yy}(\frac{1}{2}) = \sigma_v^2$. From the previous results, we see that $\widehat{S}_{yy}(f)$ is a good estimator for $S_{yy}(f)$ only for f near $\frac{1}{2}$.