System Identification

1 First exercise

Exercise

Consider the data generation mechanism of the "real" system S:

$$S: \qquad y(t) = e(t) + \frac{1}{2} \cdot e(t-1), \qquad e(t) \sim WN(0, 1)$$

and the classes of models:

$$\begin{aligned}
&M_1(\vartheta): && y(t) = a \cdot y(t-1) + \eta(t), && \eta(t) \sim WN\left(0, \lambda_1^2\right) \\
&M_2(\vartheta): && y(t) = -a \cdot y(t-1) - b \cdot y(t-2) + \eta(t), && \eta(t) \sim WN\left(0, \lambda_2^2\right) \\
&M_3(\vartheta): && y(t) = \eta(t) + a \cdot \eta(t-1), && \eta(t) \sim WN\left(0, \lambda_3^2\right)
\end{aligned}$$

1. Compute the value ϑ^* of the vector of parameters ϑ which minimizes the loss function:

$$\begin{aligned}
\bar{V}(\vartheta) &= \mathbb{E}\left[(y(t) - \hat{y}(t|t-1;\vartheta))^2\right] \\
&= \mathbb{E}\left[\varepsilon(t;\vartheta)^2\right]
\end{aligned}$$

Solution

1) Compute the value ϑ^* of the vector of parameters ϑ which minimizes the loss function.

First of all notice that, since the data generation mechanism is given we can study the *asymptotical behaviour* of the identification method. So, when the data generation system is given we can assume we have collected an infinite amount of data:

$$N \rightarrow \infty$$

This means we study the asymptotic case:

$$\begin{aligned}
V_N(\vartheta) &= \frac{1}{N} \cdot \sum_{t=1}^{N} (y(t) - \hat{y}(t|t-1;\vartheta))^2 \\
&= \frac{1}{N} \cdot \sum_{t=1}^{N} \varepsilon(t;\vartheta)^2
\end{aligned}$$

$$V_N(\vartheta) \underset{N\rightarrow\infty}{\longrightarrow} \bar{V}(\vartheta)$$

Recall that, asymptotically, the estimated parameters can be retrieved solving the formula:

$$\vartheta^* = \arg\min_{\vartheta} \left\{\bar{V}(\vartheta)\right\}$$

1.1) Analysis of M_1

Notice that the system S is a $MA(1)$, while the system M_1 is an $AR(1)$. So the "real" system S is not included in the class of models $M_1(\vartheta)$:

$$S \notin M_1(\vartheta)$$

So we can state that, asymptotically, the identified model provided by the *PEM method* will be the “best” approximant of the system S in the class of models $M_1(\vartheta)$.

The optimal 1-step predictor $\hat{y}(t|t-1;\vartheta)$ of $M_1(\vartheta)$ (which is a function of the unknown parameter a) is:

$$\begin{aligned}
\hat{y}(t|t-1;\vartheta) &= \frac{C(z)-A(z)}{C(z)} \cdot y(t) \\
&= \frac{1-\left(1-a\cdot z^{-1}\right)}{1} \cdot y(t) \\
&= a \cdot z^{-1} \cdot y(t) \\
&= a \cdot y(t-1)
\end{aligned}$$

where $A(z) = 1 - a \cdot z^{-1}$, $C(z) = 1$ and $\vartheta = a$.

Let’s compute the value ϑ^* of ϑ which minimizes the loss function $\bar{V}(\vartheta)$:

$$\begin{aligned}
\vartheta^* &= \arg\min_{\vartheta} \left\{\bar{V}(\vartheta)\right\} \\
&= \arg\min_{\vartheta} \left\{\mathbb{E}\left[(y(t)-\hat{y}(t|t-1;\vartheta))^2\right]\right\} \\
&= \arg\min_{a} \left\{\mathbb{E}\left[(y(t)-a\cdot y(t-1))^2\right]\right\} \\
&= \arg\min_{a} \left\{\mathbb{E}\left[\left(\left(e(t)+\frac{1}{2}\cdot e(t-1)\right)-a\cdot\left(e(t-1)+\frac{1}{2}\cdot e(t-2)\right)\right)^2\right]\right\} \\
&= \arg\min_{a} \left\{\mathbb{E}\left[\left(e(t)+\frac{1}{2}\cdot e(t-1)-a\cdot e(t-1)-\frac{1}{2}\cdot a\cdot e(t-2)\right)^2\right]\right\} \\
&= \arg\min_{a} \left\{\mathbb{E}\left[\left(e(t)+\left(\frac{1}{2}-a\right)\cdot e(t-1)-\frac{1}{2}\cdot a\cdot e(t-2)\right)^2\right]\right\} \\
&= \arg\min_{a} \left\{\mathbb{E}\left[e(t)^2\right]+\left(\frac{1}{2}-a\right)^2\cdot\mathbb{E}\left[e(t-1)^2\right]+\frac{1}{4}\cdot a^2\cdot\mathbb{E}\left[e(t-2)^2\right]+\underset{e\sim WN}{\cancel{\cdots}}\right\} \\
&= \arg\min_{a} \left\{\lambda_e^2+\left(\frac{1}{2}-a\right)^2\cdot\lambda_e^2+\frac{1}{4}\cdot a^2\cdot\lambda_e^2\right\} \\
&= \arg\min_{a} \left\{1+\left(\frac{1}{4}+a^2-\cancel{2\cdot\frac{1}{2}}\cdot a\right)+\frac{1}{4}\cdot a^2\right\} = \arg\min_{a}\left\{\frac{5}{4}\cdot a^2-a+\frac{5}{4}\right\}
\end{aligned}$$

Notice that thet loss function to be minimized, the variance of the 1-step prediction error, is a *convex function* (convex parabola since it’s second derivative it’s positive):

$$\bar{V}(\vartheta) = \frac{5}{4}\cdot a^2 - a + \frac{5}{4}$$

Thus, this function has just one minimum point, which is equal to the vertex of the parabola. We can easily compute it:

$$\vartheta^* = \arg\min_{\vartheta}\left\{\bar{V}(\vartheta)\right\}$$

$$\begin{aligned}
\frac{\partial}{\partial \vartheta} \bar{V}(\vartheta)\bigg|_{\vartheta=\vartheta^*} &= 0 \\
\frac{\partial}{\partial \vartheta}\left\{\frac{5}{4}\cdot a^2 - a + \frac{5}{4}\right\}\bigg|_{\vartheta=\vartheta^*} &= 0 \\
\frac{5}{2}\cdot a^* - 1 &= 0
\end{aligned}$$

$$a^* = \frac{2}{5}$$

After computing the stationary point, check if the model $M_1(\vartheta^*)$ is in it's *canonical representation:*

$$M_1(a^*): \qquad y(t) = \frac{2}{5}\cdot y(t-1) + \eta(t), \qquad \eta(t) \sim WN\left(0, \lambda_1^2\right)$$

$$\begin{cases} A(z) = 1 - \frac{2}{5}\cdot z^{-1} \\ C(z) = 1 \end{cases}$$

The polynomials $A(z)$ and $C(z)$ satisfy the properties:

1. Same degree: ✓;
2. Coprime: ✓;
3. Monic: ✓;
4. Roots inside the unitary circle: ✓ since $|a^*| = \left|\frac{2}{5}\right| < 1$

Thus the model $M_1(\vartheta^*)$ is the best approximant of the system S in the class of $AR(1)$ models (notice that this is a WSS process).

Since $S \notin M_1(\vartheta)$, the prediction error $\varepsilon(t;\vartheta^*)$ is not a white noise, indeed:

$$\begin{aligned}
\varepsilon(t;\vartheta^*) &= y(t) - \hat{y}(t|t-1;\vartheta^*) \\
&= y(t) - a^*\cdot y(t-1) \\
&= \left(e(t) + \frac{1}{2}\cdot e(t-1)\right) - \frac{2}{5}\cdot\left(e(t-1) + \frac{1}{2}\cdot e(t-2)\right) \\
&= e(t) + \frac{1}{2}\cdot e(t-1) - \frac{2}{5}\cdot e(t-1) - \frac{1}{5}\cdot e(t-2) \\
&= e(t) + \frac{1}{10}\cdot e(t-1) - \frac{1}{5}\cdot e(t-2) \\
&= \eta(t)
\end{aligned}$$

which is an $MA(2)$ process (i.e. it's not a white noise). Notice that:

$$\begin{aligned}
\lambda_1^2 = \mathbb{E}\left[\eta(t)^2\right] = \mathbb{E}\left[\varepsilon(t;\vartheta^*)^2\right] &= \bar{V}(\vartheta^*) \\
&= \frac{5}{4}\cdot(a^*)^2 - a^* + \frac{5}{4} \\
&= \frac{5}{4}\cdot\frac{4}{25} - \frac{2}{5} + \frac{5}{4} \\
&= \frac{21}{20}
\end{aligned}$$

$$\lambda_1^2 = \frac{21}{20} > 1 = \lambda_e^2$$

1.2) Analysis of $M_2(\vartheta)$

$$M_2(\vartheta): \qquad y(t) = -a \cdot y(t-1) - b \cdot y(t-2) + \eta(t), \qquad \eta(t) \sim WN\left(0, \lambda_2^2\right)$$

Notice that the system S has a structure $MA(1)$, while the class of models $M_2(\vartheta)$ has a structure $AR(1)$. So the "real" system S is not included in the class of models $M_2(\vartheta)$:

$$S \notin M_2(\vartheta)$$

Again we can state that, asymptotically, the identified model provided by the *PEM method* will be the “best” approximant of the system S in the class of models $M_2(\vartheta)$.

The optimal 1-step predictor $\hat{y}(t|t-1;\vartheta)$ of $M_2(\vartheta)$ is:

$$\begin{aligned}
\hat{y}(t|t-1;\vartheta) &= \frac{C(z) - A(z)}{C(z)} \cdot y(t) \\
&= \frac{1 - \left(1 + a \cdot z^{-1} + b \cdot z^{-2}\right)}{1} \cdot y(t) \\
&= -a \cdot z^{-1} \cdot y(t) - b \cdot z^{-2} \cdot y(t) \\
&= -a \cdot y(t-1) - b \cdot y(t-2)
\end{aligned}$$

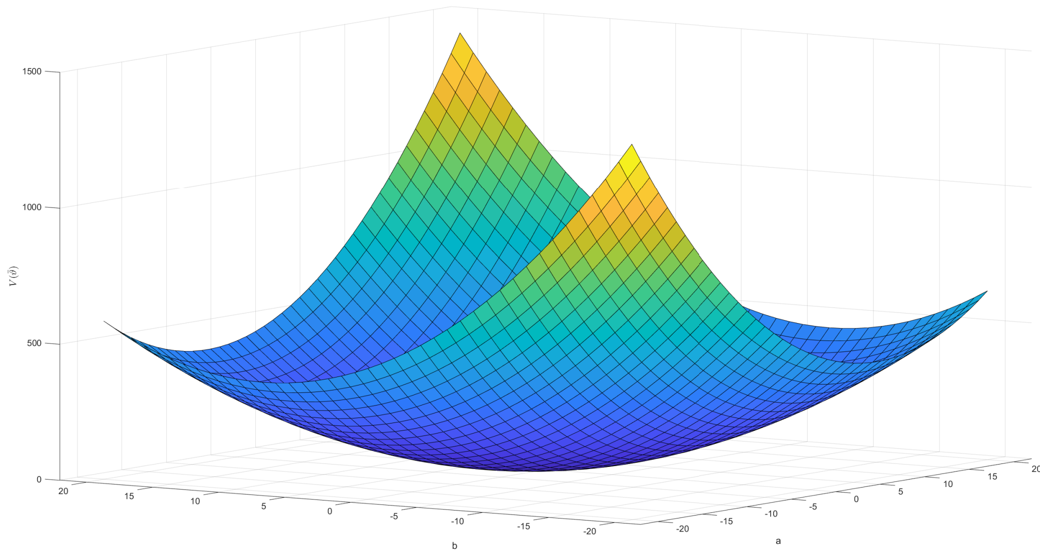
where $A(z) = 1 + a \cdot z^{-1} + b \cdot z^{-2}$, $C(z) = 1$ and $\vartheta = \begin{bmatrix} a \\ b \end{bmatrix}$. Notice that, in this case ϑ is a vector with dimensionality equal to 2.

Let’s compute the value ϑ^* of ϑ which minimizes the loss function $\bar{V}(\vartheta)$:

$$\begin{aligned}
\vartheta^* &= \arg\min_{\vartheta} \left\{\bar{V}(\vartheta)\right\} \\
&= \arg\min_{\vartheta} \left\{\mathbb{E}\left[(y(t) - \hat{y}(t|t-1;\vartheta))^2\right]\right\} \\
&= \arg\min_{a,b} \left\{\mathbb{E}\left[(y(t) + a \cdot y(t-1) + b \cdot y(t-2))^2\right]\right\} \\
&= \arg\min_{a,b} \left\{\mathbb{E}\left[\left(\left(e(t) + \frac{1}{2} \cdot e(t-1)\right) + a \cdot \left(e(t-1) + \frac{1}{2} \cdot e(t-2)\right) + b \cdot \left(e(t-2) + \frac{1}{2} \cdot e(t-3)\right)\right)^2\right]\right\} \\
&= \arg\min_{a,b} \left\{\mathbb{E}\left[\left(e(t) + \left(\frac{1}{2} + a\right) \cdot e(t-1) + \left(\frac{1}{2} \cdot a + b\right) \cdot e(t-2) + \frac{1}{2} \cdot b \cdot e(t-3)\right)^2\right]\right\} \\
&= \arg\min_{a,b} \left\{\mathbb{E}\left[e(t)^2\right] + \left(\frac{1}{2} + a\right)^2 \cdot \mathbb{E}\left[e(t-1)^2\right] + \left(\frac{1}{2} \cdot a + b\right)^2 \cdot \mathbb{E}\left[e(t-2)^2\right] + \left(\frac{1}{2} \cdot b\right)^2 \cdot \mathbb{E}\left[e(t-3)^2\right] + \underbrace{\cancel{\ldots}}_{e \sim WN}\right\} \\
&= \arg\min_{a,b} \left\{\lambda_e^2 + \left(\frac{1}{2} + a\right)^2 \cdot \lambda_e^2 + \left(\frac{1}{2} \cdot a + b\right)^2 \cdot \lambda_e^2 + \left(\frac{1}{2} \cdot b\right)^2 \cdot \lambda_e^2\right\} \\
&= \arg\min_{a,b} \left\{1 + \left(\frac{1}{4} + a^2 + \cancel{2} \cdot \frac{\cancel{1}}{\cancel{2}} \cdot a\right) + \left(\frac{1}{4} \cdot a^2 + b^2 + \cancel{2} \cdot \frac{\cancel{1}}{\cancel{2}} \cdot a \cdot b\right) + \left(\frac{1}{4} \cdot b^2\right)\right\} \\
&= \arg\min_{a,b} \left\{1 + \frac{1}{4} + a^2 + a + \frac{1}{4} \cdot a^2 + b^2 + a \cdot b + \frac{1}{4} \cdot b^2\right\} \\
&= \arg\min_{a,b} \left\{\frac{5}{4} \cdot a^2 + a \cdot b + \frac{5}{4} \cdot b^2 + a + \frac{5}{4}\right\}
\end{aligned}$$

Notice that the loss function to be minimized, the variance of the 1-step prediction error, is a *convex function* (convex paraboloid):

$$\bar{V}(\vartheta) = \frac{5}{4} \cdot a^2 + a \cdot b + \frac{5}{4} \cdot b^2 + a + \frac{5}{4}$$



This function has just one single minimum and can be easily computed. First of all we define the gradient of $\bar{V}(\vartheta)$ as:

$$\begin{aligned}\nabla\bar{V}(\vartheta) &= \nabla\bar{V}(a,b)\\ &= \begin{bmatrix}\frac{\partial}{\partial a}\bar{V}(a,b)\\ \frac{\partial}{\partial b}\bar{V}(a,b)\end{bmatrix}\\ &= \begin{bmatrix}\frac{5}{2}\cdot a+b+1\\ \frac{5}{2}\cdot b+a\end{bmatrix}\end{aligned}$$

The stationary point

$$\vartheta^* = \arg\min_{\vartheta}\left\{\bar{V}(\vartheta)\right\} = \begin{bmatrix}a^*\\ b^*\end{bmatrix}$$

can be computed as follow:

$$\nabla\bar{V}(\vartheta)\big|_{\vartheta=\vartheta^*} = 0$$

$$\begin{bmatrix}\frac{5}{2}\cdot a^*+b^*+1\\ \frac{5}{2}\cdot b^*+a^*\end{bmatrix} = \begin{bmatrix}0\\0\end{bmatrix}$$

Thus, we have to solve a system with 2 equations and 2 unknowns:

$$\begin{cases}\frac{5}{2}\cdot a^*+b^*+1=0\\ \frac{5}{2}\cdot b^*+a^*=0\end{cases}$$

$$\begin{cases}\frac{5}{2}\cdot a^*+b^*+1=0\\ a^*=-\frac{5}{2}\cdot b^*\end{cases}$$

$$\begin{cases}-\frac{25}{4}\cdot b^*+b^*+1=0\\ a^*=-\frac{5}{2}\cdot b^*\end{cases}$$

$$\begin{cases}b^*=\frac{4}{21}\\ a^*=-\frac{10}{21}\end{cases}$$

After computing the stationary point, check if the model $M_2(\vartheta^*)$ is in it's *canonical representation:*

$$M_2(a^*, b^*): \qquad y(t) = \frac{10}{21} \cdot y(t-1) - \frac{4}{21} \cdot y(t-2) + \eta(t), \qquad \eta(t) \sim WN\left(0, \lambda_2^2\right)$$

$$\begin{cases} A(z) = 1 - \frac{10}{21} \cdot z^{-1} + \frac{4}{21} \cdot z^{-2} \\ C(z) = 1 \end{cases}$$

The polynomials $A(z)$ and $C(z)$ satisfy the properties:

1. Same degree: ✓;
2. Coprime: ✓;
3. Monic: ✓;
4. Roots inside the unitary circle: ✓ since the roots of $A(z)$ are inside the unitary circle:

$$\begin{aligned} 1 - \frac{10}{21} \cdot z^{-1} + \frac{4}{21} \cdot z^{-2} &= 0 \\ z^2 - \frac{10}{21} \cdot z^1 + \frac{4}{21} &= 0 \end{aligned}$$

$$z_{1,2} = \frac{\frac{10}{21} \pm \sqrt{\frac{100}{441} - \frac{16}{21}}}{2} = \frac{5}{21} \pm j \cdot \sqrt{\frac{59}{441}}$$

We have two complex conjugate poles. So:

$$|z_1| = |z_2| = \sqrt{\frac{25}{441} + \frac{59}{441}} = \sqrt{\frac{84}{441}} < 1$$

So the poles are strictly inside the unitary circle, $M_2(\vartheta^*)$ is a WSS process and it is expressed in the canonical representation.

Again, notice that since $S \notin M_2(\vartheta)$, the prediction error $\varepsilon(t; \vartheta^*)$ is not a white noise:

$$\eta(t) = \varepsilon(t; \vartheta^*) = y(t) - \hat{y}(t|t-1; \vartheta^*)$$

Notice that:

$$\begin{aligned} \lambda_2^2 = \mathbb{E}\left[\eta(t)^2\right] = \mathbb{E}\left[\varepsilon(t; \vartheta^*)^2\right] &= \bar{V}(\vartheta^*) \\ &= \frac{5}{4} \cdot (a^*)^2 + a^* \cdot b^* + \frac{5}{4} \cdot (b^*)^2 + a^* + \frac{5}{4} \\ &= \frac{5}{4} \cdot \frac{100}{441} - \frac{10}{21} \cdot \frac{4}{21} + \frac{5}{4} \cdot \frac{16}{441} - \frac{10}{21} + \frac{5}{4} \\ &= \frac{125}{441} - \frac{40}{441} + \frac{20}{441} - \frac{210}{441} + \frac{5}{4} \\ &= \frac{-105 \cdot 4 + 5 \cdot 441}{441 \cdot 4} = \frac{1785}{1764} = \frac{85 \cdot 21}{84 \cdot 21} = \frac{85}{84} \end{aligned}$$

$$\lambda_e^2 = 1 < \lambda_2^2 = \frac{85}{84} < \lambda_1^2 = \frac{21}{20}$$

In the model $M_2(\vartheta^*)$ the variance of $\eta(t)$ is closer to the variance of $e(t)$ with respect to the model $M_1(\vartheta^*)$: this is due to the fact that an $MA(1)$ can be seen as an $AR(\infty)$; the greater is the m order of the AR model, the closer is the identified $AR(m)$ model to the system S.

1.3) Analysis of $M_3(\vartheta)$

$$M_3(\vartheta): \qquad y(t) = \eta(t) + a \cdot \eta(t-1), \qquad \eta(t) \sim WN\left(0, \lambda_3^2\right)$$

Notice that the system S has a structure $MA(1)$ and the class of models $M_3(\vartheta)$ has a structure $MA(1)$ too. So:

$$S \in M_3(\vartheta)$$

By recalling the theory, we can conclude that, asymptotically:

$$\hat{a}_N \underset{N\to\infty}{\longrightarrow} a^\circ = \frac{1}{2}$$

Thus the identified model is "asymptotically equivalent" to the system S. Thus, in this case, we have that:

$$a^* = \frac{1}{2}$$

However, if we do not recognize that $S \in M_3(\vartheta)$, we can apply the definition. The optimal 1-step predictor $\hat{y}(t|t-1;\vartheta)$ of $M_3(\vartheta)$ is:

$$\begin{aligned}
\hat{y}(t|t-1;\vartheta) &= \frac{C(z) - A(z)}{C(z)} \cdot y(t) \\
&= \frac{\left(1 + a \cdot z^{-1}\right) - 1}{\left(1 + a \cdot z^{-1}\right)} \cdot y(t) \\
&= \frac{a \cdot z^{-1}}{1 + a \cdot z^{-1}} \cdot y(t)
\end{aligned}$$

where $A(z) = 1$, $C(z) = 1 + a \cdot z^{-1}$ and $\vartheta = a$.

Let's compute the value ϑ^* of ϑ which minimizes the loss function $\bar{V}(\vartheta)$:

$$\begin{aligned}
\vartheta^* &= \arg\min_{\vartheta} \left\{\bar{V}(\vartheta)\right\} \\
&= \arg\min_{\vartheta} \left\{\mathbb{E}\left[\left(y(t) - \hat{y}(t|t-1;\vartheta)\right)^2\right]\right\} \\
&= \arg\min_{a} \left\{\mathbb{E}\left[\left(y(t) - \frac{a \cdot z^{-1}}{1 + a \cdot z^{-1}} \cdot y(t)\right)^2\right]\right\} \\
&= \arg\min_{a} \left\{\mathbb{E}\left[\left(\frac{\cancel{1 + a \cdot z^{-1} - a \cdot z^{-1}}}{1 + a \cdot z^{-1}} \cdot y(t)\right)^2\right]\right\} \\
&= \arg\min_{a} \left\{\mathbb{E}\left[\left(\frac{1}{1 + a \cdot z^{-1}} \cdot y(t)\right)^2\right]\right\} \\
&= \arg\min_{a} \left\{\mathbb{E}\left[\left(\frac{1 + \frac{1}{2} \cdot z^{-1}}{1 + a \cdot z^{-1}} \cdot e(t)\right)^2\right]\right\}
\end{aligned}$$

Notice that, under the condition $|a^*| < 1$ the minimum variance we can obtain is the variance of the noise $e(t)$: by inspection, we can see that this is possible if and only if $a = \frac{1}{2}$, which leads to:

$$a^* = \frac{1}{2}$$

This way, the digital transfer function turns out to be 1:

$$\frac{1 + \frac{1}{2} \cdot z^{-1}}{1 + a \cdot z^{-1}} \cdot e(t) = 1 \cdot e(t) = e(t)$$

However, we can always proceed in the classical way:

$$\begin{aligned}\varepsilon(t;\vartheta) &= y(t) - \hat{y}(t|t-1;\vartheta) \\ &= \frac{1+\frac{1}{2}\cdot z^{-1}}{1+a\cdot z^{-1}}\cdot e(t)\end{aligned}$$

$$\begin{aligned}\varepsilon(t;\vartheta) + a\cdot\varepsilon(t-1;\vartheta) &= e(t) + \frac{1}{2}\cdot e(t-1) \\ \varepsilon(t;\vartheta) &= -a\cdot\varepsilon(t-1;\vartheta) + e(t) + \frac{1}{2}\cdot e(t-1)\end{aligned}$$

So:

$$\begin{aligned}\bar{V}(\vartheta) &= \mathbb{E}\left[\varepsilon(t;\vartheta)^2\right] \\ &= \mathbb{E}\left[(y(t) - \hat{y}(t|t-1;\vartheta))^2\right] \\ &= \mathbb{E}\left[\left(-a\cdot\varepsilon(t-1;\vartheta) + e(t) + \frac{1}{2}\cdot e(t-1)\right)^2\right] \\ &= a^2\cdot\mathbb{E}\left[\varepsilon(t-1;\vartheta)^2\right] + \mathbb{E}\left[e(t)^2\right] + \frac{1}{4}\cdot\mathbb{E}\left[e(t-1)^2\right] - \\ &- 2\cdot a\cdot\cancel{\mathbb{E}[\varepsilon(t-1;\vartheta)\cdot e(t)]} - 2\cdot\frac{1}{2}\cdot a\cdot\mathbb{E}[\varepsilon(t-1;\vartheta)\cdot e(t-1)] + 2\cdot\frac{1}{2}\cdot\cancel{\mathbb{E}[e(t)\cdot e(t-1)]}\end{aligned}$$

The term $\mathbb{E}[\varepsilon(t-1;\vartheta)\cdot e(t)]$ can be calculated as follow:

$$\begin{aligned}\mathbb{E}[\varepsilon(t-1;\vartheta)\cdot e(t)] &= \mathbb{E}\left[\left(-a\cdot\varepsilon(t-2;\vartheta) + e(t-1) + \frac{1}{2}\cdot e(t-2)\right)\cdot e(t)\right] \\ &= -a\cdot\cancel{\mathbb{E}[\varepsilon(t-2;\vartheta)\cdot e(t)]} + \cancel{\mathbb{E}[e(t-1)\cdot e(t)]} + \frac{1}{2}\cdot\cancel{\mathbb{E}[e(t-2)\cdot e(t)]} \\ &= 0\end{aligned}$$

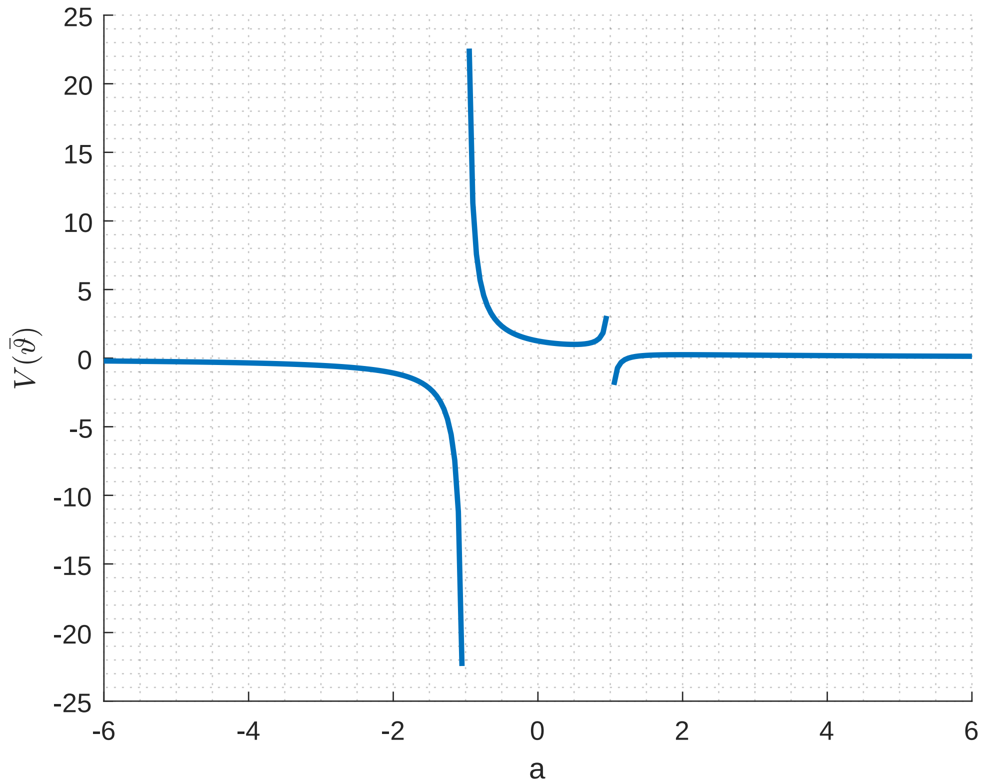
The term $\mathbb{E}[\varepsilon(t-1;\vartheta)\cdot e(t-1)]$ can be calculated as follow:

$$\begin{aligned}\mathbb{E}[\varepsilon(t-1;\vartheta)\cdot e(t-1)] &= \mathbb{E}\left[\left(-a\cdot\varepsilon(t-2;\vartheta) + e(t-1) + \frac{1}{2}\cdot e(t-2)\right)\cdot e(t-1)\right] \\ &= -a\cdot\cancel{\mathbb{E}[\varepsilon(t-2;\vartheta)\cdot e(t-1)]} + \mathbb{E}[e(t-1)\cdot e(t-1)] + \frac{1}{2}\cdot\cancel{\mathbb{E}[e(t-2)\cdot e(t-1)]} \\ &= \lambda_e^2\end{aligned}$$

Thus we can continue to compute the variance of the prediction error:

$$\begin{aligned}\bar{V}(\vartheta) &= \mathbb{E}\left[\varepsilon(t;\vartheta)^2\right] \\ \lambda_\varepsilon^2(a) &= a^2\cdot\lambda_\varepsilon^2(a) + \lambda_e^2 + \frac{1}{4}\cdot\lambda_e^2 - 2\cdot\frac{1}{2}\cdot a\cdot\mathbb{E}[\varepsilon(t-1;\vartheta)\cdot e(t-1)] \\ \left(1-a^2\right)\cdot\lambda_\varepsilon^2(a) &= \lambda_e^2 + \frac{1}{4}\cdot\lambda_e^2 - a\cdot\lambda_e^2\end{aligned}$$

$$\lambda_\varepsilon^2(a) = \frac{\frac{5}{4}-a}{1-a^2}$$



Thus the stationary point

$$\vartheta^* = \arg\min_{\vartheta} \left\{ \bar{V}(\vartheta) \right\} = a^* = \arg\min_{a} \left\{ \lambda_\varepsilon^2(a) \right\}$$

can be computed as follow:

$$\nabla \bar{V}(\vartheta)\Big|_{\vartheta=\vartheta^*} = 0$$

$$\frac{\partial}{\partial a} \lambda_\varepsilon^2(a)\Big|_{a=a^*} = 0$$

$$\frac{\partial}{\partial a}\left\{ \frac{\frac{5}{4} - a}{1 - a^2} \right\}\Bigg|_{a=a^*} = 0$$

$$-1 \cdot \frac{1}{(1-a^2)} + \left(\frac{5}{4} - a\right) \cdot \left(\frac{-(-2 \cdot a)}{(1-a^2)^2} \right)\Bigg|_{a=a^*} = 0$$

$$\frac{-(1-a^2) + 2 \cdot a \cdot \left(\frac{5}{4} - a\right)}{(1-a^2)^2}\Bigg|_{a=a^*} = 0$$

$$\frac{-1 + a^2 + \frac{5}{2} \cdot a - 2 \cdot a^2}{(1-a^2)^2}\Bigg|_{a=a^*} = 0$$

$$\frac{-a^2 + \frac{5}{2} \cdot a - 1}{(1-a^2)^2}\Bigg|_{a=a^*} = 0$$

$$\frac{-2 \cdot a^2 + 5 \cdot a - 2}{2 \cdot (1-a^2)^2}\Bigg|_{a=a^*} = 0$$

$$-2 \cdot a^2 + 5 \cdot a - 2\Big|_{a=a^*} = 0$$

$$-2 \cdot (a^*)^2 + 5 \cdot a^* - 2 = 0$$

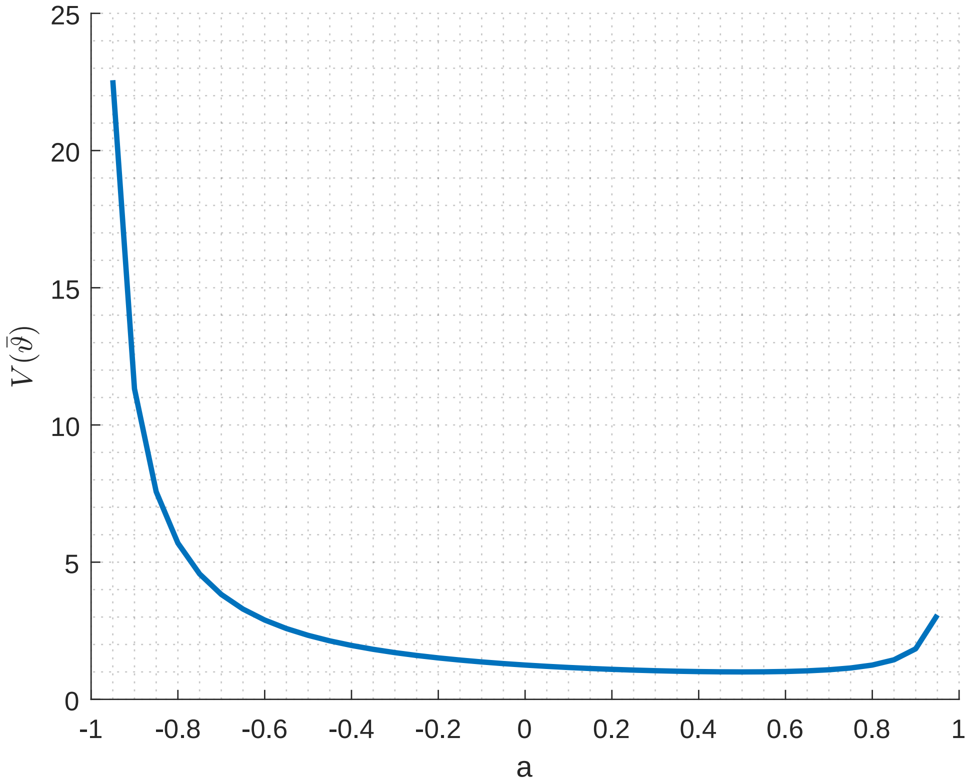
by exploiting the quotient rule.

Thus the stationary points are:

$$a^*_{1,2} = \frac{-5 \pm \sqrt{25-16}}{-4} = \frac{5 \pm 3}{4}$$

$$\begin{cases} a^*_1 = 2 & \text{maximum} \\ a^*_2 = \frac{1}{2} & \text{minimum} \end{cases}$$

Note that these are just local extrema: in fact, $\lambda^2_\varepsilon\left(a^* = a^*_1 = 2\right) = \frac{1}{4}$ and $\lambda^2_\varepsilon\left(a^* = a^*_2 = \frac{1}{2}\right) = 1$. Nevertheless, a^*_1 does not respect the condition on the canonical form of $M_3\left(\vartheta^*\right) = M_3\left(a^* = a^*_1\right)$. Thus, only $a^* = a^*_2$ is acceptable, and it represents the minimum point of $\lambda^2_\varepsilon\left(a\right)$ for $|a| < 1$.



Notice that since $S \in M_3\left(\vartheta\right)$, the prediction error $\varepsilon\left(t;\vartheta^*\right)$ is a white noise:

$$\begin{aligned} \eta\left(t\right) = \varepsilon\left(t;\vartheta^*\right) &= y\left(t\right) - \hat{y}\left(t|t-1;\vartheta^*\right) \\ &= \frac{1+\frac{1}{2}\cdot z^{-1}}{1+a^*_2\cdot z^{-1}}\cdot e\left(t\right) \\ &= e\left(t\right) \end{aligned}$$

Notice that:

$$\begin{aligned} \lambda^2_3 = \mathbb{E}\left[\eta\left(t\right)^2\right] = \mathbb{E}\left[\varepsilon\left(t;\vartheta^*\right)^2\right] &= \bar{V}\left(\vartheta^*\right) \\ &= \mathbb{E}\left[e\left(t\right)^2\right] \\ &= \lambda^2_e = 1 \end{aligned}$$

$$\lambda^2_3 = \lambda^2_e = 1 < \lambda^2_2 = \frac{85}{84} < \lambda^2_1 = \frac{21}{20}$$

2 Second exercise (December 2018)

Exercise

Consider the following dataset:

t	1	2	3	4	5
$u(t)$	1	0	-1	0	1
$y(t)$	1	-1	0	1	-1

and the following families of model:

$$\mathcal{M}_1 \ : \quad y(t) = b_1 \cdot u(t-1) + e_1(t) \qquad\qquad e_1(t) \sim WN\left(0, \lambda_1^2\right)$$
$$\mathcal{M}_2 \ : \quad y(t) = a_2 \cdot y(t-1) + b_2 \cdot u(t-1) + e_2(t) \qquad\qquad e_2(t) \sim WN\left(0, \lambda_2^2\right)$$

1. Using the dataset identifes the parameter b_1 of the family $\mathcal{M}_1$ according to the PEM criteria.
2. Using the dataset identifes the parameters a_2 and b_2 of the family $\mathcal{M}_2$ according to the PEM criteria.

Solution

First of all notice that, since the data generation mechanism is not given, we can not study the *asymptotical behaviour* of the identification method.

In general, when dealing with an $ARX(m, p+1)$ structure of models, it is more correct to write:

$$V_N(\vartheta) = \frac{1}{N-h} \cdot \sum_{t=h+1}^{N} \left(y(t) - \hat{y}(t|t-1;\vartheta)\right)^2$$

where h is:

$$h = \max\{m, p+1\}$$

in order to deal with the relative delay k between the output $y(t)$ and the exogenous input $u(t-k)$ using a finite set of data (values of $y(t)$ and $u(t)$ for $t \leq 0$ are unknown).

1) Analysis of $\mathcal{M}_1$

The family of model $\mathcal{M}_1$ has an *ARMAX* structure with the following polynomials:

$$\begin{cases} A(z) = 1 \\ B(z) = b_1 \\ C(z) = 1 \\ k = 1 \end{cases}$$

The optimal 1-step predictor $\hat{y}(t|t-1;\vartheta)$ of $\mathcal{M}_1$ is:

$$\begin{aligned} \hat{y}(t|t-r;\vartheta) &= \frac{B(z) \cdot Q_r(z)}{C(z)} \cdot u(t-k) + \frac{R_r(z)}{C(z)} \cdot y(t) \\ \hat{y}(t|t-1;\vartheta) &= \frac{B(z) \cdot Q_1(z)}{C(z)} \cdot u(t-k) + \frac{R_1(z)}{C(z)} \cdot y(t) \\ &= \frac{B(z) \cdot Q_1(z)}{C(z)} \cdot u(t-1) + \frac{R_1(z)}{C(z)} \cdot y(t) \end{aligned}$$

from the long division of the polynomials $A(z)$ and $B(z)$:

$$R_1(z) = 0$$
$$Q_1(z) = 1$$

Thus the predictor is:

$$\begin{aligned}\hat{y}(t|t-1;\vartheta) &= \frac{B(z) \cdot Q_1(z)}{C(z)} \cdot u(t-1) + \frac{R_1(z)}{C(z)} \cdot y(t) \\ &= \frac{b_1 \cdot 1}{1} \cdot u(t-1) + \frac{0}{1} \cdot y(t) \\ &= b_1 \cdot u(t-1)\end{aligned}$$

where $\vartheta = b_1$.

This predictor can also be found by noting that:

$$y(t) = \underbrace{b_1 \cdot u(t-1)}_{\text{known part}} + \underbrace{e_1(t)}_{\text{unknown part}}$$

and therefore the predictor is:

$$\hat{y}(t|t-1;\vartheta) = b_1 \cdot u(t-1)$$

Once the predictor is found, it's possible to write the loss function:

$$\begin{aligned}V_N(\vartheta) = V_N(b_1) &= \frac{1}{4} \cdot \sum_{t=2}^{5} \left(y(t) - \hat{y}(t|t-1;\vartheta)\right)^2 \\ &= \frac{1}{4} \cdot \sum_{t=2}^{5} \left(y(t) - b_1 \cdot u(t-1)\right)^2\end{aligned}$$

and its gradient:

$$\begin{aligned}\nabla V_N(\vartheta) = \nabla V_N(b_1) &= \frac{1}{4} \cdot \sum_{t=2}^{5} \frac{\partial}{\partial b_1} \left(y(t) - b_1 \cdot u(t-1)\right)^2 \\ &= \frac{1}{4} \cdot \sum_{t=2}^{5} 2 \cdot \left(y(t) - b_1 \cdot u(t-1)\right) \cdot \left(-u(t-1)\right) \\ &= \frac{1}{2} \cdot \sum_{t=2}^{5} \left[-y(t) \cdot u(t-1) + b_1 \cdot u(t-1)^2\right]\end{aligned}$$

The stationary points can be computed as:

$$\nabla \bar{V}(\vartheta)\big|_{\vartheta=\vartheta^*} = 0$$

Thus:

$$\nabla V_N\left(b_1\right) = 0$$

$$\cancel{\frac{1}{2}} \cdot \sum_{t=2}^{5} \left[-y\left(t\right) \cdot u\left(t-1\right) + b_1 \cdot u\left(t-1\right)^2\right] = 0$$

$$b_1 \cdot \sum_{t=2}^{5} u\left(t-1\right)^2 = \sum_{t=2}^{5} y\left(t\right) \cdot u\left(t-1\right)$$

$$b_1 \cdot \left(u\left(1\right)^2 + u\left(2\right)^2 + u\left(3\right)^2 + u\left(4\right)^2\right) = \left(y\left(2\right) \cdot u\left(1\right) + y\left(3\right) \cdot u\left(2\right) + y\left(4\right) \cdot u\left(3\right) + y\left(5\right) \cdot u\left(4\right)\right)$$

$$b_1 \cdot 2 = -2$$

$$b_1 = \frac{-2}{2} = -1$$

The only stationary point is then:

$$b_1^* = -1$$

1) Analysis of $\mathcal{M}_2$

The second family of models has the transfer function:

$$\begin{aligned} y\left(t\right) &= a_2 \cdot y\left(t-1\right) + b_2 \cdot u\left(t-1\right) + e_2\left(t\right) \\ y\left(t\right) &= a_2 \cdot z^{-1} \cdot y_2\left(t\right) + b_2 \cdot z^{-1} \cdot u\left(t\right) + e_2\left(t\right) \\ y_2\left(t\right) - a_2 \cdot z^{-1} \cdot y\left(t\right) &= b_2 \cdot z^{-1} \cdot u\left(t\right) + e_2\left(t\right) \\ \left(1 - a_2 \cdot z^{-1}\right) \cdot y\left(t\right) &= b_2 \cdot z^{-1} \cdot u\left(t\right) + e_2\left(t\right) \\ y\left(t\right) &= \frac{b_2 \cdot z^{-1}}{1 - a_2 \cdot z^{-1}} \cdot u\left(t\right) + \frac{1}{1 - a_2 \cdot z^{-1}} \cdot e_2\left(t\right) \\ y\left(t\right) &= \frac{b_2}{1 - a_2 \cdot z^{-1}} \cdot u\left(t-1\right) + \frac{1}{1 - a_2 \cdot z^{-1}} \cdot e_2\left(t\right) \end{aligned}$$

where the polynomials are:

$$\begin{aligned} A\left(z\right) &= 1 - a_2 \cdot z^{-1} \\ B\left(z\right) &= b_2 \\ C\left(z\right) &= 1 \\ k &= 1 \end{aligned}$$

from the long division we obtain:

$$\begin{aligned} R_1\left(z\right) &= a_2 \cdot z^{-1} \\ Q_1\left(z\right) &= 1 \end{aligned}$$

and the predictor:

$$\begin{aligned} \hat{y}\left(t|t-1; \vartheta\right) &= \frac{R_1\left(z\right)}{C\left(z\right)} \cdot y\left(t\right) + \frac{B\left(z\right) \cdot Q_1\left(z\right)}{C\left(z\right)} \cdot u\left(t-1\right) \\ &= a_2 \cdot z^{-1} \cdot y\left(t\right) + \frac{b_2 \cdot 1}{1} \cdot u\left(t-1\right) \\ &= a_2 \cdot y\left(t-1\right) + b_2 \cdot u\left(t-1\right) \end{aligned}$$

where $\vartheta = \begin{bmatrix} a_2 \\ b_2 \end{bmatrix}$.
This predictor can also be found by noting that:

$$y(t) = \underbrace{a_2 \cdot y(t-1) + b_2 \cdot u(t-1)}_{\text{known part}} + \underbrace{e_2(t)}_{\text{unknown part}}$$

and therefore the predictor is:

$$\hat{y}(t|t-1;\vartheta) = a_2 \cdot y(t-1) + b_2 \cdot u(t-1)$$

Once the predictor is found, it's possible to write the cost function:

$$\begin{aligned} V_N(\vartheta) = V_N(a_2, b_2) &= \frac{1}{4} \cdot \sum_{t=2}^{5} (y(t) - \hat{y}(t|t-1;\vartheta))^2 \\ &= \frac{1}{4} \cdot \sum_{t=2}^{5} (y(t) - a_2 \cdot y(t-1) - b_2 \cdot u(t-1))^2 \end{aligned}$$

and:

$$\begin{aligned} \frac{\partial}{\partial b_2} V_N(a_2, b_2) &= \frac{1}{4} \cdot \sum_{t=2}^{5} \frac{\partial}{\partial b_2} (y(t) - a_2 \cdot y(t-1) - b_2 \cdot u(t-1))^2 \\ &= \frac{1}{4} \cdot \sum_{t=2}^{5} 2 \cdot (y(t) - a_2 \cdot y(t-1) - b_2 \cdot u(t-1)) \cdot (-u(t-1)) \\ &= \frac{1}{2} \cdot \sum_{t=2}^{5} (y(t) - a_2 \cdot y(t-1) - b_2 \cdot u(t-1)) \cdot (-u(t-1)) \end{aligned}$$

$$\begin{aligned} \frac{\partial}{\partial a_2} V_N(a_2, b_2) &= \frac{1}{4} \cdot \sum_{t=2}^{5} \frac{\partial}{\partial a_2} (y(t) - a_2 \cdot y(t-1) - b_2 \cdot u(t-1))^2 \\ &= \frac{1}{4} \cdot \sum_{t=2}^{5} 2 \cdot (y(t) - a_2 \cdot y(t-1) - b_2 \cdot u(t-1)) \cdot (-y(t-1)) \\ &= \frac{1}{2} \cdot \sum_{t=2}^{5} (y(t) - a_2 \cdot y(t-1) - c_1 \cdot u(t-1)) \cdot (-y(t-1)) \end{aligned}$$

The stationary points can be computed as:

$$\nabla \bar{V}(\vartheta)\big|_{\vartheta=\vartheta^*} = 0$$

Thus:

$$\begin{aligned} \frac{\partial}{\partial b_2} V_N(a_2, b_2) &= 0 \\ \cancel{\frac{1}{2}} \cdot \sum_{t=2}^{5} (y(t) - a_2 \cdot y(t-1) - b_2 \cdot u(t-1)) \cdot (-u(t-1)) &= 0 \\ a_2 \cdot \sum_{t=2}^{5} y(t-1) \cdot u(t-1) + b_2 \cdot \sum_{t=2}^{5} u(t-1)^2 &= \sum_{t=2}^{5} y(t) \cdot u(t-1) \\ a_2 \cdot 1 + b_2 \cdot 2 &= -2 \\ a_2 + 2 \cdot b_2 &= -2 \end{aligned}$$

$$\frac{\partial}{\partial a_2} V_N\left(a_2, b_2\right) = 0$$

$$\cancel{\frac{1}{2}} \cdot \sum_{t=2}^{5} \left(y\left(t\right) - a_2 \cdot y\left(t-1\right) - b_2 \cdot u\left(t-1\right)\right) \cdot \left(-y\left(t-1\right)\right) = 0$$

$$a_2 \cdot \sum_{t=2}^{5} y\left(t-1\right)^2 + b_2 \cdot \sum_{t=2}^{5} u\left(t-1\right) \cdot y\left(t-1\right) = \sum_{t=2}^{5} y\left(t\right) \cdot y\left(t-1\right)$$

$$3 \cdot a_2 + 1 \cdot b_2 = -2$$

We have a system with 2 equations and 2 unknowns:

$$\begin{cases} a_2 + 2 \cdot b_2 = -2 \\ 3 \cdot a_2 + b_2 = -2 \end{cases}$$

$$\begin{cases} 3 \cdot \left(a_2 + 2 \cdot b_2\right) = 3 \cdot (-2) \\ 3 \cdot a_2 + b_2 = -2 \end{cases}$$

$$\begin{cases} 3 \cdot a_2 + 6 \cdot b_2 = -6 \\ 3 \cdot a_2 + b_2 = -2 \end{cases}$$

$$\begin{cases} 5 \cdot b_2 = -4 \\ 3 \cdot a_2 + b_2 = -2 \end{cases}$$

$$\begin{cases} b_2 = -\frac{4}{5} \\ 3 \cdot a_2 = -2 + \frac{4}{5} \end{cases}$$

$$\begin{cases} a_2 = -\frac{6}{5} \cdot \frac{1}{3} \\ b_2 = -\frac{4}{5} \end{cases}$$

$$\begin{cases} a_2 = -\frac{2}{5} \\ b_2 = -\frac{4}{5} \end{cases}$$

Thus the stationary points are:

$$\begin{cases} a_2^* = -\frac{2}{5} \\ b_2^* = -\frac{4}{5} \end{cases}$$