1. (a) The ARMA(p,q) process $x(n)$ is described by the difference equation

$$x(n)+\sum_{l=1}^{p}a(l)x(n-l)=\sum_{l=0}^{q}b(l)v(n-l).$$

Expressing the process as a vectorial AR(1) process is based on the state space representation. The idea is to define a suitable vector $\mathbf{y}(n)$ that depends linearly on the previous state vector $\mathbf{y}(n-1)$ according to an equation of the AR(1) form:

$$\mathbf{y}(n)=\mathbf{A}\mathbf{y}(n-1)+\mathbf{b}v(n).$$

Here $\mathbf{y}(n)$ is a vector, $\mathbf{A}$ is a matrix and $\mathbf{b}$ is a vector. Let's choose $\mathbf{y}(n)$ so that it contains all the information that is needed to calculate the value $x(n+1)$ of the original process:

$$\mathbf{y}(n)=[x(n),\ldots,x(n-p+1),v(n),\ldots,v(n-q+1)]^T$$

except the noise term $v(n+1)$, that is obtained through the vector $\mathbf{b}$:

$$\mathbf{b}=[b(0),0,\ldots,0,1,0,\ldots,0]^T$$

where the element with value 1 is in the same position as the element $v(n)$ in the vector $\mathbf{y}(n)$.
The matrix $\mathbf{A}$ is constructed so that its first row gives the difference equation of the ARMA process:

$$A(1,1:n)=[-a(1),\ldots,-a(p),b(1),\ldots,b(q)]$$

The rows $2,3,\ldots,p$ have zero elements, except value 1 for elements multiplying the suitable elements of the vector $\mathbf{y}(n-1)$ to result in $x(n-k)=x(n-k)$.
The $p+1$:th row of the matrix contains just zeros as the corresponding element of the vector $\mathbf{y}(n)$ is $v(n)$ that is not included in the vector $\mathbf{y}(n-1)$ (instead, it is taken care of through the $\mathbf{b}$ term).
Rest of the rows contain zeros, except values 1 in positions resulting in equations $v(n-k)=v(n-k)$.

$$\underbrace{\begin{bmatrix} x(n)\\ x(n-1)\\ x(n-2)\\ \vdots\\ x(n-p+2)\\ x(n-p+1)\\ v(n)\\ v(n-1)\\ \vdots\\ v(n-q+2)\\ v(n-q+1)\end{bmatrix}}_{\mathbf{y}(n)}=\underbrace{\begin{bmatrix} -a(1) & -a(2) & \ldots & -a(p-1) & -a(p) & b(1) & \ldots & b(q-1) & b(q)\\ 1 & 0 & \ldots & 0 & 0 & 0 & \ldots & 0 & 0\\ 0 & 1 & \ldots & 0 & 0 & 0 & \ldots & 0 & 0\\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & & \vdots & \vdots\\ 0 & 0 & \ldots & 0 & 0 & 0 & \ldots & 0 & 0\\ 0 & 0 & \ldots & 1 & 0 & 0 & \ldots & 0 & 0\\ 0 & 0 & \ldots & 0 & 0 & 0 & \ldots & 0 & 0\\ 0 & 0 & \ldots & 0 & 0 & 1 & \ldots & 0 & 0\\ \vdots & \vdots & & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots\\ 0 & 0 & \ldots & 0 & 0 & 0 & \ldots & 0 & 0\\ 0 & 0 & \ldots & 0 & 0 & 0 & \ldots & 1 & 0\end{bmatrix}}_{\mathbf{A}}\underbrace{\begin{bmatrix} x(n-1)\\ x(n-2)\\ x(n-3)\\ \vdots\\ x(n-p+1)\\ x(n-p)\\ v(n-1)\\ v(n-2)\\ \vdots\\ v(n-q+1)\\ v(n-q)\end{bmatrix}}_{\mathbf{y}(n-1)}+\underbrace{\begin{bmatrix} b(0)\\ 0\\ 0\\ \vdots\\ 0\\ 0\\ 1\\ 0\\ \vdots\\ 0\\ 0\end{bmatrix}}_{\mathbf{b}}v(n)$$

(b) Let's calculate the conditional expectation and variance for the process

$$\mathbf{y}(n) = \sum_{k=0}^{\infty} \mathbf{A}^k \mathbf{b} v(n-k).$$

In this context, the variance of the vector $\mathbf{y}(n)$ is understood to mean the covariance matrix

$$\mathrm{E}[(\mathbf{y}(n) - m)\,(\mathbf{y}(n) - m)^H]$$

$$= \mathrm{E}\begin{bmatrix} (y_1(n) - m_1)(y_1(n) - m_1)^* & \dots & (y_1(n) - m_1)(y_{p+q}(n) - m_{p+q})^* \\ \vdots & \ddots & \vdots \\ (y_{p+q}(n) - m_{p+q})(y_1(n) - m_1)^* & \dots & (y_{p+q}(n) - m_{p+q})(y_{p+q}(n) - m_{p+q})^* \end{bmatrix}$$

where $\mathbf{m} = [m_1\, m_2\, \cdots\, m_{p+q}]^T = \mathrm{E}[\mathbf{y}(n)]$.

The expected value:

$$\begin{aligned}
\mathrm{E}(\mathbf{y}(n+k) \mid \mathbf{y}(n), \dots) &= \mathrm{E}\left(\sum_{l=0}^{\infty} \mathbf{A}^l \mathbf{b} v(n+k-l) \,\middle|\, \mathbf{y}(n), \dots \right) \\
&= \mathrm{E}\left(\sum_{l=0}^{k-1} \mathbf{A}^l \mathbf{b} v(n+k-l) \right) + \mathrm{E}\left(\sum_{l=k}^{\infty} \mathbf{A}^l \mathbf{b} v(n+k-l) \,\middle|\, \mathbf{y}(n), \dots \right) \\
&= 0 + \sum_{l=0}^{\infty} \mathbf{A}^{k+l} \mathbf{b} v(n-l) \\
&= \mathbf{A}^k \mathbf{y}(n)
\end{aligned}$$

The variance:

$$\begin{aligned}
\mathrm{Var}(\mathbf{y}(n+k) \mid \mathbf{y}(n), \dots) &= \mathrm{Var}\left(\sum_{l=0}^{\infty} \mathbf{A}^l \mathbf{b} v(n+k-l) \,\middle|\, \mathbf{y}(n), \dots \right) \\
&= \mathrm{Var}\left(\sum_{l=0}^{k-1} \mathbf{A}^l \mathbf{b} v(n+k-l) + \sum_{l=k}^{\infty} \mathbf{A}^l \mathbf{b} v(n+k-l) \,\middle|\, \mathbf{y}(n), \dots \right) \\
&= \mathrm{Var}\left(\sum_{l=0}^{k-1} \mathbf{A}^l \mathbf{b} v(n+k-l) \right) \\
\text{(uncorrelated terms)} \qquad &= \sum_{l=0}^{k-1} \mathrm{Var}(\mathbf{A}^l \mathbf{b} v(n+k-l)) \\
\text{(the mean is 0)} \qquad &= \sum_{l=0}^{k-1} \mathrm{E}\left[\mathbf{A}^l \mathbf{b} v(n+k-l) v^H(n+k-l) \mathbf{b}^H (\mathbf{A}^H)^l \right] \\
&= \sum_{l=0}^{k-1} \mathbf{A}^l \mathbf{b}\, \mathrm{E}\left[v(n+k-l) v^H(n+k-l) \right] \mathbf{b}^H (\mathbf{A}^H)^l \\
&= \sum_{l=0}^{k-1} \mathbf{A}^l \mathbf{b} \mathbf{b}^H (\mathbf{A}^H)^l
\end{aligned}$$

2. The parameter a_1 of the AR part is given by the latter of the Yule-Walker equations of the process

$$\begin{bmatrix} r_x(0) & r_x(1) \\ r_x(1) & r_x(0) \end{bmatrix} \begin{bmatrix} 1 \\ a_1 \end{bmatrix} = \begin{bmatrix} |b_0|^2 \\ 0 \end{bmatrix}.$$

The autocorrelation matrix estimated with the covariance method

$$\hat{\mathbf{R}}_{\mathbf{x}} = \begin{bmatrix} r_x(0,0) & r_x(0,1) \\ r_x(1,0) & r_x(1,1) \end{bmatrix}$$

can be inserted in the place of the autocorrelation matrix on the left hand side of the Yule-Walker equations. With this insertion, the latter of the equations yields

$$r_x(1,1)a_1 = -r_x(1,0).$$

The autocorrelation estimates

$$r_x(1,1) = \sum_{n=1}^{N-1} x(n-1)x^*(n-1) = \sum_{n=0}^{N-2} |x(n)|^2 = \sum_{n=0}^{N-2} |\rho|^{2n} = \frac{1-|\rho|^{2(N-1)}}{1-|\rho|^2}$$

$$r_x(1,0) = \sum_{n=1}^{N-1} x(n)x^*(n-1) = \sum_{n=0}^{N-2} \rho|x(n)|^2 = \rho\frac{1-|\rho|^{2(N-1)}}{1-|\rho|^2}.$$

Above we have used the sum of the geometric series $\alpha(1-q^N)/(1-q)$, where q is the ratio of consecutive terms, N is the number of terms and α is the first term. This results in the value

$$a_1 = -\frac{r_x(1,0)}{r_x(1,1)} = -\rho$$

for the parameter of the AR model.

The transfer function of the model $x(n) + a_1x(n-1) = b_0v(n)$ is

$$H(z) = \frac{X(z)}{V(z)} = \frac{b_0}{1+a_1z^{-1}}$$

which results in the transfer function having one pole at the point $z = -a_1 = \rho$. The system defined by the model is stable when $|\rho| < 1$. Then the pole is inside the unit circle. The squared error of the model

$$\sum_{n=p}^{N-1} |e(n)|^2 = \sum_{n=1}^{N-1} |x(n) + a_p(1)x(n-1)|^2 = \sum_{n=1}^{N-1} |\rho^n - \rho\rho^{n-1}|^2 = 0$$

3. Just as in the previous exercise problem, the autocorrelation matrix in the Yule-Walker equations is replaced with an estimate, the estimate given by the covariance method.

$$\begin{bmatrix} r_x(0,0) & r_x(0,1) & r_x(0,2) \\ r_x(1,0) & r_x(1,1) & r_x(1,2) \\ r_x(2,0) & r_x(2,1) & r_x(2,2) \end{bmatrix} \begin{bmatrix} 1 \\ a_p(1) \\ a_p(2) \end{bmatrix} = \begin{bmatrix} |b(0)|^2 \\ 0 \\ 0 \end{bmatrix}$$

where $r_x(k,l) = \sum_{n=p}^{N} x(n-l)x^*(n-k)$. An alternative way is to form the correlation matrix is by $\hat{\mathbf{R}}_x = \mathbf{X}^H \mathbf{X}$ where

$$\mathbf{X} = \begin{bmatrix} x(2) & x(1) & x(0) \\ x(3) & x(2) & x(1) \\ x(4) & x(3) & x(2) \\ x(5) & x(4) & x(3) \end{bmatrix} = \begin{bmatrix} 3 & 2 & -1 \\ 4 & 3 & 2 \\ 5 & 4 & 3 \\ 6 & 5 & 4 \end{bmatrix},$$

resulting in

$$\hat{\mathbf{R}}_x = \begin{bmatrix} r_x(0,0) & \cdots & r_x(0,p) \\ \vdots & \ddots & \vdots \\ r_x(p,0) & \cdots & r_x(p,p) \end{bmatrix} = \begin{bmatrix} 86 & 68 & 44 \\ 68 & 54 & 36 \\ 44 & 36 & 30 \end{bmatrix}$$

We notice that the matrix is not Toeplitz. Once again, let's take those of the equations that have zero on the right hand side, and move the leftmost column of the autocorrelation matrix to the right hand side. One selects the required elements from $\hat{\mathbf{R}}_x$ ands solves for the parameters a. This yields

$$\begin{bmatrix} a_p(1) \\ a_p(2) \end{bmatrix} = - \begin{bmatrix} 54 & 36 \\ 36 & 30 \end{bmatrix}^{-1} \begin{bmatrix} 68 \\ 44 \end{bmatrix} = \begin{bmatrix} -1.407 \\ 0.222 \end{bmatrix}.$$

Now the error is $e(n) = x(n) + a_p(1)x(n-1) + a_p(2)x(n-2)$. Let's form an error vector $\mathbf{e}$ out of the error values $e(p), \ldots, e(N)$:

$$\mathbf{e} = \mathbf{X} \begin{bmatrix} 1 \\ a_p(1) \\ a_p(2) \end{bmatrix}.$$

With this, $\sum_{n=p}^{N} |e(n)|^2 = \mathbf{e}^T \mathbf{e} = 0.0741$

Alternatively, the error could be found by solving $|b(0)|^2$ from the first row of the Yule-Walker equations:

$$|b(0)|^2 = r_x(0,0) + r_x(0,1)a_p(1) + r_x(0,2)a_p(2) = 0.0741$$

The time inverted signal is $(x(0), \ldots, x(N)) = (6, 5, 4, 3, 2, -1)$, so that

$$\mathbf{X} = \begin{bmatrix} 4 & 5 & 6 \\ 3 & 4 & 5 \\ 2 & 3 & 4 \\ -1 & 2 & 3 \end{bmatrix},$$

which once again results in

$$\hat{\mathbf{R}}_x = \mathbf{X}^H \mathbf{X} = \begin{bmatrix} 30 & 36 & 44 \\ 36 & 54 & 68 \\ 44 & 68 & 86 \end{bmatrix}$$

When one selects the required elements from this matrix, one gets

$$\begin{bmatrix} a_p(1) \\ a_p(2) \end{bmatrix} = - \begin{bmatrix} 54 & 68 \\ 68 & 86 \end{bmatrix}^{-1} \begin{bmatrix} 36 \\ 44 \end{bmatrix} = \begin{bmatrix} -5.20 \\ 3.60 \end{bmatrix}.$$

The error calculated using the time inverted x is $\sum_{n=p}^{N} |e(n)|^2 = 1.20$.

In the case of time inversion, one speaks about a backwards predicting filter. In this case, the forward predicting filter gave a better result. This doesn't mean that it would be the better alternative in general. In the case of perfect estimates (e.g. with unlimited amount of data), both forward and backward filters converge to give the same estimates as the true autocorrelations of a real valued ARMA process are invariant to time inversion.

4. (a) Part of the AR-signal remain uncaptured by the model. The prediction error contains some AR signal, it is not just white noise. In this case, the order of the predictor is too low. The linear predictor is not able to model the dependency between the values $x(n)$ and $x(n-p)$ when $M < p$. In the general case, AR(p) models include such dependencies. Usually the optimal predictor coefficients differ from the low-order coefficients of the true AR model. (One can confirm this e.g. by looking at the Levinson-Durbin recursion formulas.)

   (b) In principle one now estimates the true parameters of the AR(p) process and the normal equations yield just these parameter values. In the ideal case, the filter output is white noise. In practice the parameters of the filter do not exactly agree with the true AR parameters as they have to be estimated from limited data samples.

   (c) Now the filter order $M > p$ is too large, i.e. the filter has more coefficients than the AR model. The prediction result (error) does not improve from (b) because the extra filter coefficients try to predict white noise. (As white noise is an uncorrelated process, its linear prediction does not succeed better than random guess.) The normal equations have an unique solution. If one denotes the parameters of the modelling process with $a_M(k)$ and the parameters of the true process with $a_p(k)$, the normal equations have the solution
   $a_M(1) = a_p(1), \ldots, a_M(p) = a_p(p), a_M(p+1) = 0, \ldots, a_M(M) = 0.$

5. (a) Insert the autocorrelation-method estimate of the autocorrelation matrix into the Yule-Walker equations:

$$\hat{\mathbf{R}}_x \begin{pmatrix} 1 \\ a(1) \end{pmatrix} = \begin{pmatrix} \hat{r}_x(0) & \hat{r}_x(1) \\ \hat{r}_x(1) & \hat{r}_x(0) \end{pmatrix} \begin{pmatrix} 1 \\ a(1) \end{pmatrix} = \begin{pmatrix} |b(0)|^2 \\ 0 \end{pmatrix},$$

   where

$$\hat{r}_x(k) = \sum_{n=k}^{N} x(n)x^*(n-k).$$

   The required estimates for the autocorrelations:

$$\hat{r}_x(0) = x^2(0) + x^2(1)$$
$$\hat{r}_x(1) = x(0)x(1)$$

   From the lower equation we can solve

$$a(1) = -\frac{\hat{r}_x(1)}{\hat{r}_x(0)} = -\frac{x(0)x(1)}{x^2(0) + x^2(1)}.$$

This is enough to determine the stability of the model. As

$$(x(0) \pm x(1))^2 = x^2(0) + x^2(1) \pm 2x(0)x(1) \geq 0$$

then $x^2(0) + x^2(1) \geq 2|x(0)x(1)| \geq |x(0)x(1)|$.
Thus $|a(1)| < 1$ (equality is impossible, as $2|x(0)x(1)| > |x(0)x(1)|$, if $x(0)x(1) \neq 0$. And if $x(0)x(1) = 0$, then $a(1) = 0 < 1$). Thus the AR(1)-process will always be stable.
Having solved $a(1)$, we can get the magnitude of $b(0)$ from the first equation

$$|b(0)|^2 = \hat{r}_x(0)+a(1)\hat{r}_x(1) = \hat{r}_x(0)-\frac{\hat{r}_x(1)^2}{\hat{r}_x(0)} = \frac{\hat{r}_x(0)^2 - \hat{r}_x(1)^2}{\hat{r}_x(0)} = \frac{x^4(0) + x^2(0)x^2(1) + x^4(1)}{x^2(0) + x^2(1)}.$$

$b(0)$ cannot be uniquely defined, but it can be chosen to be real and positive, or to have an arbitrary phase.

(b) Insert the covariance-method estimate of the autocorrelation matrix into the Yule-Walker equations:

$$\hat{\mathbf{R}}_x \begin{pmatrix} 1 \\ a(1) \end{pmatrix} = \begin{pmatrix} \hat{r}_x(0,0) & \hat{r}_x(0,1) \\ \hat{r}_x(1,0) & \hat{r}_x(1,1) \end{pmatrix} \begin{pmatrix} 1 \\ a(1) \end{pmatrix} = \begin{pmatrix} |b(0)|^2 \\ 0 \end{pmatrix},$$

where

$$\hat{r}_x(k,l) = \sum_{n=p}^{N} x(n-l)x^*(n-k).$$

The elements of the matrix can be written in terms of the observations:

$$\begin{aligned} r_x(0,0) &= x^2(1) \\ r_x(1,1) &= x^2(0) \\ r_x(0,1) &= r_x(1,0) = x(0)x(1) \end{aligned}$$

Now

$$a(1) = -\frac{\hat{r}_x(1,0)}{\hat{r}_x(1,1)} = -\frac{x(0)x(1)}{x^2(0)} = -\frac{x(1)}{x(0)}.$$

Hence $a(1) = -1$ if $x(0) = x(1)$. The model will be unstable whenever $|x(1)| \geq |x(0)|$.
The magnitude of $b(0)$ can again be solved from the first equation

$$|b(0)|^2 = \hat{r}_x(0,0) + a(1)\hat{r}_x(0,1) = x^2(1) - \frac{x(1)}{x(0)}x(0)x(1) = 0$$

The result $b(0) = 0$ here means that the covariance method expects to predict all values exactly without error (this is unreasonable, but happens because the only value to predict was $x(1)$, and the model was optimised to estimate it exactly), hence in the corresponding ARMA-process there would be no contribution from noise. From the other perspective, considering the transfer function of the ARMA model:

$$H(z) = \frac{b(0)}{1 + a(1)z^{-1}}$$

having $b(0) = 0$ means $H(z) = 0$ everwhere, and there is no model. If the process should be studied from that point of view, some other value $b(0)$ would have to be chosen (e.g., in Hayes' book, Example 4.6.2, it is arbitrarily stated that $b(0) = 1$).