4.13. *Remark.*
(1) In case of $k = 1$ (one-step prediction) we shall simplify the notation omitting the superscript (1).
(2) Observe that dividing (4.2) by $\gamma(0) > 0$ leads to an equivalent SLE $R_n \mathbf{\Phi}_n^{(k)} = \boldsymbol{\rho}_n^{(k)}$ expressed in terms of autocorrelation function $\rho(\cdot)$ instead of the autocovariance function $\gamma(\cdot)$.
(3) Neither Γ_n nor $\boldsymbol{\gamma}_n^{(k)}$ depends on t (by stationarity) and thus hereafter we can assume $t = n$ without the loss of generality.

4.14. **Theorem** (Durbin-Levinson algorithm for $\mathbf{\Phi}_n$).
Let $X = \{X_t \,|\, t \in \mathbb{Z}\}$ be a stationary time series with $\mu_X = 0$ and autocovariance function $\gamma_X(h) \to 0$ for $h \to \infty$. If $\widehat{X}_{n+1} = \Phi_{n,1}X_n + \cdots + \Phi_{n,n}X_1$ is the best linear prediction as of Theorem 4.12 then coefficients $\Phi_{n,j}$ and the mean square error $v_n = \mathrm{E}|X_{n+1} - \widehat{X}_{n+1}|^2$ may be recursively computed as follows

Initial step with $n = 0$:

$$v_0 = \gamma_X(0). \tag{4.4a}$$

Recursive step with $n > 0$:

$$\Phi_{n,n} = \Big[\gamma_X(n) - \overbrace{\sum_{j=1}^{n-1} \Phi_{n-1,j}\, \gamma_X(n-j)}^{=0 \text{ for } n=1}\Big] v_{n-1}^{-1}, \tag{4.4b}$$

$$v_n = v_{n-1}(1 - |\Phi_{n,n}|^2), \tag{4.4c}$$

$$\begin{bmatrix} \Phi_{n,1} \\ \Phi_{n,2} \\ \vdots \\ \Phi_{n,n-1} \end{bmatrix} = \begin{bmatrix} \Phi_{n-1,1} \\ \Phi_{n-1,2} \\ \vdots \\ \Phi_{n-1,n-1} \end{bmatrix} - \Phi_{n,n} \begin{bmatrix} \Phi_{n-1,n-1} \\ \Phi_{n-1,n-2} \\ \vdots \\ \Phi_{n-1,1} \end{bmatrix} \quad \textit{for } n > 1. \tag{4.4d}$$

Proof. See [BD93, §5.2]. □

4.15. **Definition.** Let $\widehat{X} = P_{\mathcal{L}(1,X_1,\ldots,X_n)}X$ and $\widehat{Y} = P_{\mathcal{L}(1,X_1,\ldots,X_n)}Y$ where $X, Y, X_1, \ldots, X_n \in L_2$ then $\rho(X, Y \mid X_1, \ldots, X_n) := \rho(X - \widehat{X}, Y - \widehat{Y})$ is called **partial correlation coefficient of random variables X and Y given $X_1, \ldots, X_n$.**
Interpretation: partial correlation between X and Y given $X_1, \ldots, X_n$ is a correlation between X and Y cleaned of its part transmitted via the influence of random variables $X_1, \ldots, X_n$.

4.16. **Definition.** Let $X = \{X_t \mid t \in \mathbb{Z}\}$ be a stationary time series with autocorrelation function $\rho_X(\cdot)$. Then the **partial autocorrelation function (pacf) $\alpha_X(\cdot)$ of X** is defined as follows:

$$\alpha_X(0) = \rho_X(0) = 1,$$
$$\alpha_X(1) = \rho_X(1) = \rho(X_{t+1}, X_t)$$
$$\alpha_X(h) = \rho(X_{t+h}, X_t \mid X_{t+1}, \ldots, X_{t+h-1}) \text{ for } h \geq 2.$$

4.17. **Theorem.** *If $X = \{X_t \mid t \in \mathbb{Z}\}$ is a stationary time series with $\mu_X = 0$ and partial autocorrelation function $\alpha_X(\cdot)$ then $\alpha_X(n) = \Phi_{n,n}$ for $n \geq 1$ where $\mathbf{\Phi}_n = [\Phi_{n,1}, \ldots, \Phi_{n,n}]^T$ is the solution to the 1-step best linear prediction problem as of eq. (4.2), i.e. $\widehat{X}_{n+1} = \sum_{j=1}^{n} \Phi_{n,j} X_{n+1-j}$.*

Proof. See [BD93, §5.2]. □

Clearly, $\alpha_X(h)$ may be computed recursively using the intermediate result (4.4b) of the Durbin-Levinson algorithm 4.14. Another procedure is based on Cramer's rule according to the next corollary. Unfortunately, that method is computationally not much efficient for large n.

4.18. **Corollary.** *If the matrix Γ_n of eq. (4.2) is nonsingular, then*

$$\alpha_X(n) = \Phi_{n,n} = \frac{\det\Gamma_n^*}{\det\Gamma_n},$$

where $\Gamma_n^* := [\Gamma(:,1), \ldots, \Gamma(:,n-1), \boldsymbol{\gamma}_n]$.

4.19. **Definition** (ARMA process).
Stochastic process $X = \{X_t \,|\, t \in \mathbb{Z}\}$ is called **ARMA process of order** p, q $(0 \le p, q < \infty)$, we write $X \sim ARMA(p,q)$, if

$$\begin{aligned} X_t = & \underbrace{\Phi_1 X_{t-1} + \Phi_2 X_{t-2} + \cdots + \Phi_p X_{t-p}}_{\text{Autoregression component (AR)}} + \\ & + \underbrace{Z_t + \Theta_1 Z_{t-1} + \Theta_2 Z_{t-2} + \cdots + \Theta_q Z_{t-q}}_{\text{Moving average component (MA=Moving Average)}}, \end{aligned} \tag{4.5a}$$

where $Z := \{Z_t \,|\, t \in \mathbb{Z}\} \sim WN(0, \sigma^2)$ and $\Phi_p \neq 0,\ \Theta_q \neq 0,\ \sigma \neq 0$.
Rewriting (4.5a) into an equivalent form

$$\begin{aligned} X_t - \Phi_1 X_{t-1} - \Phi_2 X_{t-2} - \cdots - \Phi_p X_{t-p} = \\ = Z_t + \Theta_1 Z_{t-1} + \Theta_2 Z_{t-2} + \cdots + \Theta_q Z_{t-q}, \end{aligned} \tag{4.5b}$$

a short form may be used

$$\Phi(B) X_t = \Theta(B) Z_t \quad \text{or} \quad \Phi(z) X(z) = \Theta(z) Z(z), \tag{4.5c}$$

giving with $\Phi_0 = \Theta_0 = 1$

$$\begin{aligned} \Phi(z) &= 1 - \Phi_1 z - \cdots - \Phi_p z^p, \\ \Theta(z) &= 1 + \Theta_1 z + \cdots + \Theta_q z^q,\ z \in \mathbb{C}. \end{aligned}$$

4.20. *Remark*. In the preceding definition we assumed $\Theta_0 = 1$ without the loss of generality, because otherwise it would be sufficient to replace the original white noise by a modified one $\{\Theta_0 Z_t\} \sim WN(0, (\Theta_0 \sigma)^2)$, and the original Θ_i by $\frac{\Theta_i}{\Theta_0}$ for $i = 1, \ldots, q$.

4.21. <u>*Remark*</u> (Special cases).

a) Autoregressive process (AR process):
$X \sim AR(p) = ARMA(p,0): \quad \Phi(B)X_t = Z_t$
because $\Theta(z) \equiv 1$. Then

$$X_t = \Phi_1 X_{t-1} + \Phi_2 X_{t-2} + \cdots + \Phi_p X_{t-p} + Z_t. \tag{4.5d}$$

We admit $p = \infty$ provided that $\Phi := \{\Phi_j\}_{j=1}^{\infty} \in \ell_1$.

b) Moving average process (MA process):
$X \sim MA(q) = ARMA(0,q): \quad X_t = \Theta(B)Z_t$
because $\Phi(z) \equiv 1$. Then

$$X_t = Z_t + \Theta_1 Z_{t-1} + \Theta_2 Z_{t-2} + \cdots + \Theta_q Z_{t-q}. \tag{4.5e}$$

We admit $q = \infty$ provided that $\Theta := \{\Theta_j\}_{j=1}^{\infty} \in \ell_1$.

c) White noise:
White noise is the only process which is both AR and MA process:
$X \sim ARMA(0,0) = AR(0) = MA(0) = WN(0,\sigma^2)$: $X_t = Z_t$.

d) General ARMA process:
$X \sim ARMA(p,q),\ 0 < p,q < \infty$: True mixture of autoregressive and moving average components.

4.22. **Definition.**
$X = \{X_t \,|\, t \in \mathbb{Z}\}$, $X \sim ARMA(p,q)$ is called **causal ARMA process** if there exists $\psi = \{\psi_j\}_{j=0}^{\infty}$, $\sum_{j=0}^{\infty}|\psi_j| < \infty$ (i.e. $\psi \in \ell_1$) such that

$$X_t = \sum_{j=0}^{\infty} \psi_j Z_{t-j} \quad \text{(or in short form } X_t = \psi(B)Z_t\text{)},\ t \in \mathbb{Z}. \tag{4.6a}$$

X is called **invertible ARMA process** if there exists $\pi = \{\pi_j\}_{j=0}^{\infty}$, $\sum_{j=0}^{\infty}|\psi_j| < \infty$ (i.e. $\pi \in \ell_1$) such that

$$\sum_{j=0}^{\infty} \pi_j X_{t-j} = Z_t \quad \text{(or in short form } \pi(B)X_t = Z_t), \ t \in \mathbb{Z}. \quad (4.6b)$$

4.23. <u>*Remark*</u>.
Consequently, **causality** in this context says that ARMA process X is also a time series generated by white noise $\{Z_t\}$ in the sense of Remark 4.5(1), or $X \sim MA(\infty)$ in our notation.
On the other hand, **invertibility** means that the white noise $\{Z_t\}$ itself may be generated by the given ARMA process X, which is equivalent with $X \sim AR(\infty)$.
Above we assumed $\psi_0 = \pi_0 = 1$ again, which will be confirmed in section 4.32 later on. There the main issue will be the computation of the causal and invertible representation of an ARMA process.

4.24. **Theorem** (Autocovariance function of an MA process).
$\{X_t\} \sim MA(q)$, $q \leq \infty$ is a stationary process having zero mean $\mu_X = 0$ and autocovariance function

$$\gamma_X(h) = \sigma^2 \sum_{k=0}^{q} \Theta_{h+k}\Theta_k \text{ for } h \geq 0. \quad (4.7a)$$

Hence for $q < \infty$ we get

$$\gamma_X(h) = \begin{cases} \sigma^2 \sum_{k=0}^{q-h} \Theta_{h+k}\Theta_k & \text{for } 0 \leq h \leq q \\ 0 & \text{for } h > q \end{cases} \quad (4.7b)$$

and in particular $\gamma_X(q) = \sigma^2\Theta_q \neq 0$ in view of $\Theta_0 = 1$.

Proof. By Corollary 4.4 is $\{X_t\}$ stationary and it holds

$$\mu_X = \mu_Z \sum_{j=0}^{q} \Theta_j = 0, \text{ because } \mu_Z = 0.$$

$$\gamma_X(h) = \sum_{j,k=0}^{q} \Theta_j \Theta_k \gamma_Z(h-j+k), \text{ where } \gamma_Z(h) = \begin{cases} \sigma^2 & \text{for } h=0 \\ 0 & \text{for } h \neq 0 \end{cases}.$$

The only nonzero terms in the sum are with $h - j + k = 0$, i.e. with $j = h + k$, which yields (4.7a)

$$\gamma_X(h) = \sum_{k=0}^{q} \Theta_{h+k} \Theta_k \underbrace{\gamma_Z(0)}_{\sigma^2}.$$

With $q < \infty$ we have $\Theta_{h+k} = 0$ for $h + k > q$, or equivalently for $k > q - h$, which allows us to rewrite (4.7a) as (4.7b). □

4.25. **Corollary.**

$$\sigma_X^2 = \gamma_X(0) = \sigma^2 \sum_{k=0}^{q} |\Theta_k|^2. \tag{4.8a}$$

$$\sigma_X^2 = \sigma^2(1 + |\Theta_1|^2 + |\Theta_2|^2 + \cdots + + |\Theta_q|^2) \textit{ for } q < \infty.$$

$$\rho_X(h) = \frac{\sum_{k=0}^{q-h} \Theta_{h+k} \Theta_k}{\sum_{k=0}^{q} |\Theta_k|^2} \textit{ for } h \geq 0. \tag{4.8b}$$

4.26. **Theorem** (Pacf of a causal AR process)**.**
Let $\{X_t\} \sim AR(p)$, $p < \infty$ be a causal AR process. Then $\{X_t\}$ is stationary with zero mean $\mu_X = 0$ and partial autocorrelation function α_X satisfying $\alpha_X(p) = \Phi_p \neq 0$ and $\alpha_X(h) = 0$ for $h > p$. Moreover $\widehat{X}_t = \Phi_1 X_{t-1} + \cdots + \Phi_p X_{t-p} = P_{\mathcal{L}(X_{t-1},\ldots,X_{t-p})} X_t$ where $\Phi_1, \ldots, \Phi_p$ are precisely the 1-step best linear prediction coefficients.

Proof. In view of 4.23 and due to causality $\{X_t\} \sim MA(\infty)$ is zero-mean stationary by 4.24. By (4.6a) is $X_t = \sum_{j=0}^{\infty} \psi_j Z_{t-j}$ and consequently $X_t \in \mathcal{L}(Z_t, Z_{t-1}, \dots) =: \mathcal{L}_t$ (closure of a linear subspace in $L_2(\Omega, \mathcal{A}, P)$ spanned by random variables $Z_u,\ u \leq t$). Putting $\Phi_j = 0$ for $j > p$, we can write for each $n \geq p$ in view of (4.5d):

$$X_t = \underbrace{\sum_{j=1}^{n} \Phi_j X_{t-j}}_{\in \mathcal{L}_{t-1}} + Z_t.$$

Random variables Z_t are uncorrelated: $Z_t \perp Z_u$ for $t \neq u$. In particular $Z_t \perp Z_u$ for $u < t$ and thus $Z_t \perp \mathcal{L}_{t-1}$ by the continuity and bi-linearity of inner-product in $L_2(\Omega, \mathcal{A}, P)$. Applying the orthogonal projection theorem we get $\widehat{X}_t = \sum_{j=1}^{n} \Phi_j X_{t-j}$ as a unique best linear prediction X_t in terms of $X_{t-1}, \dots, X_{t-n}$ for every $n \geq p$. By the Theorem 4.17 it holds $\alpha(p) = \Phi_p$ and $\alpha(n) = \Phi_n = 0$ for $n > p$. □

Figures 4.1, 4.2 and 4.3 show typical behaviour of estimated autocorrelation and partial autocorrelation functions of simulated processes $AR(2)$, $MA(2)$ and $ARMA(2,2)$, respectively. Dashed band stands for the appropriate point confidence interval containing zero with probability 0.95. We see that processes $AR(2)$ on Fig. 4.1, or $MA(2)$ on Fig. 4.2, exhibit $\alpha_X(h) \approx 0$, or $\rho_X(h) \approx 0$ for $h > 2$ in accordance with theorems 4.26 and 4.24, respectively. Otherwise the envelope of $\rho_X(h)$, or $\alpha_X(h)$ (with $ARMA(2,2)$ on Fig. 4.3 both of them) exhibits exponential decay, eventually combined with oscillatory behaviour. It is because one can show that both ρ_X and α_X may be expressed in such cases as a linear combination of decreasing geometrical sequences and/or cosine waves with geometrically decreasing amplitudes.

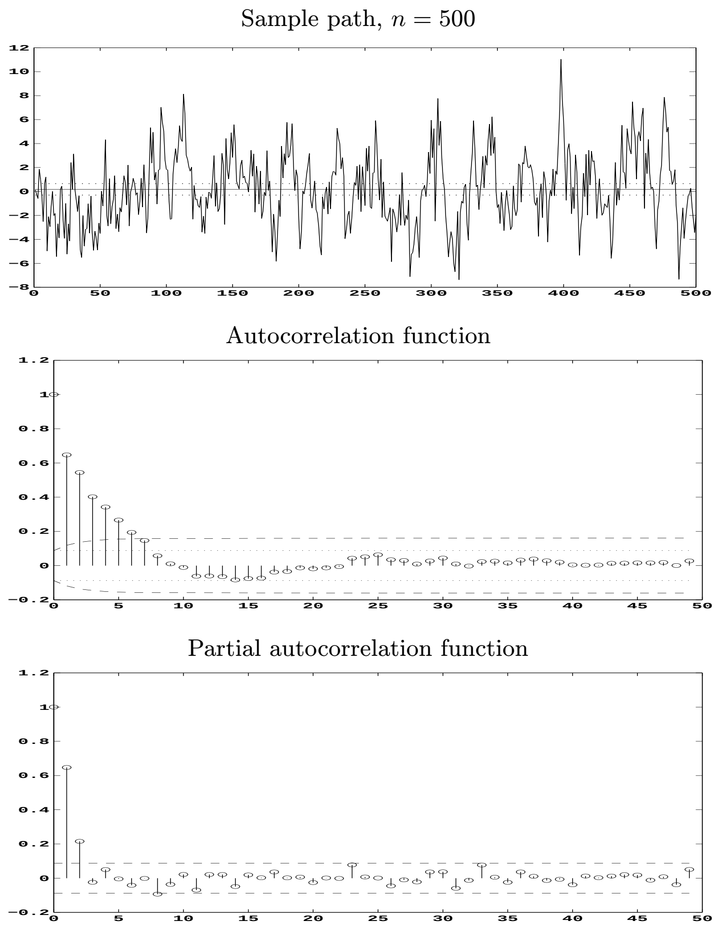

FIGURE 4.1. $AR(2):\ \mathbf{\Phi} = [0.5, 0.2],\ \sigma^2 = 2.25.$

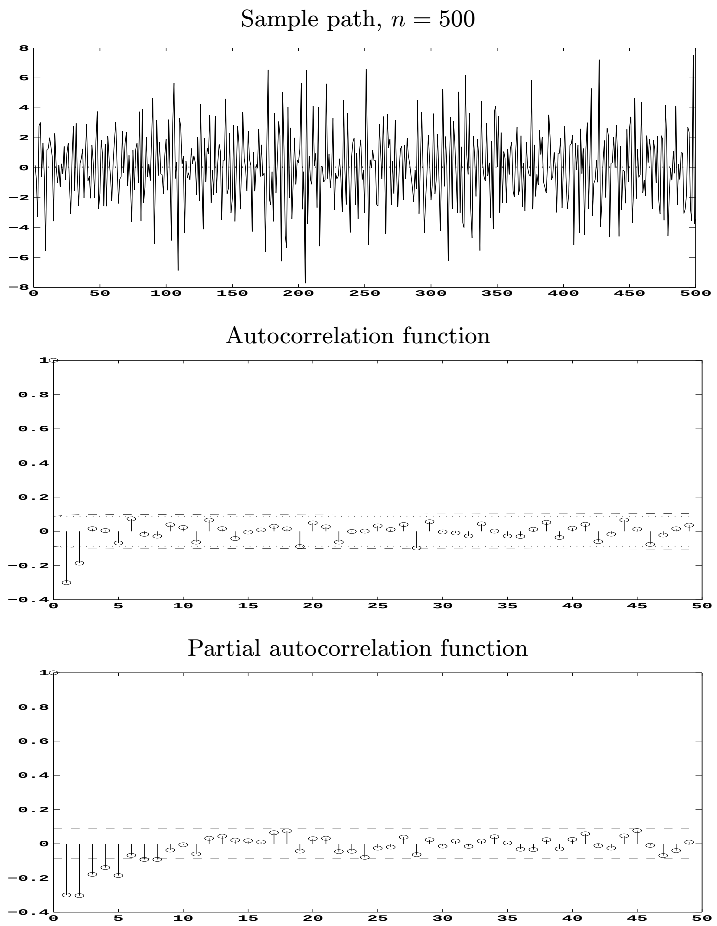

FIGURE 4.2. $MA(2): \ \mathbf{\Theta} = [-0.5, -0.2], \ \sigma^2 = 2.25.$

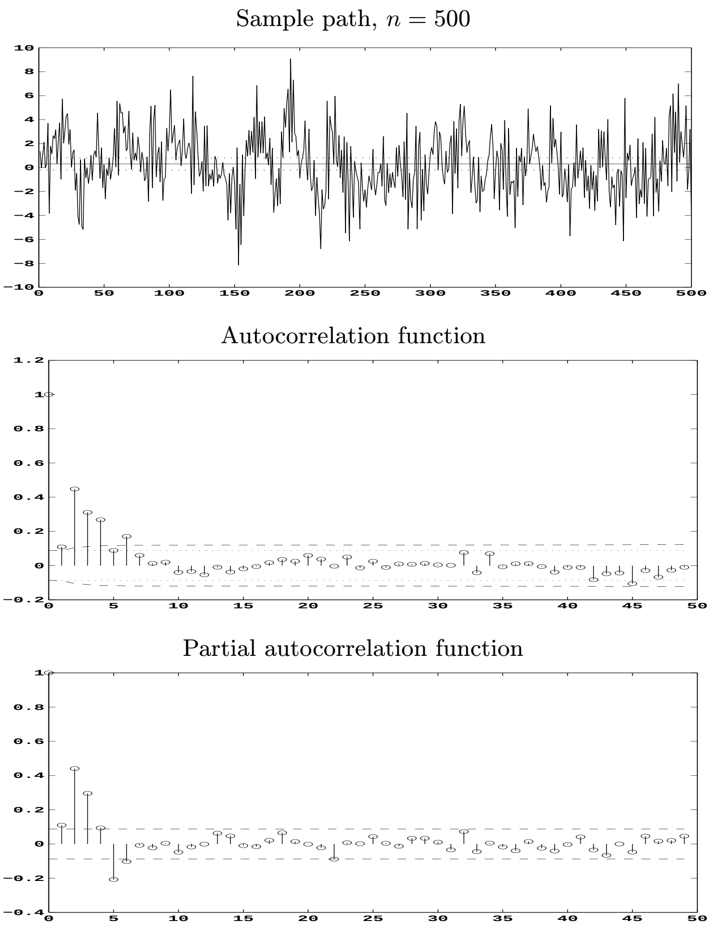

FIGURE 4.3. $ARMA(2,2)$: $\mathbf{\Phi} = [0.5, 0.2]$, $\mathbf{\Theta} = [-0.6, 0.3]$, $\sigma^2 = 2.25$.

References

[BD93] Peter J. Brockwell and Richard A. Davis, *Time series: Theory and methods*, 2-nd ed., Springer-Verlag, New York, 1991 (corrected 2-nd printing 1993).