We shall state two theoretical results (without proof) confirming the appropriacy of causal invertible MA or AR models for stationary time series with decreasing autocorrelation function [see BD93, §3.2, §8.1].

4.27. Theorem.

Let $\{X_t\}$ be an arbitrary stationary time series with $\mu_X = 0$ and $\gamma_X(h) = \begin{cases} 0 & \text{for } h > q \\ \neq 0 & \text{for } h = q \end{cases}$. Then $\{X_t\} \sim MA(q)$, i.e. there exists a white noise $\{Z_t\} \sim WN(0, \sigma^2)$ with some variance σ^2 and parameters $\theta_1, \theta_2, \dots, \theta_q \in \mathbb{R}$, $\theta_q \neq 0$ such that:

$X_t = Z_t + \theta_1 Z_{t-1} + \dots + \theta_q Z_{t-q}$.

4.28. Theorem

Let $\{X_t\}$ be an arbitrary stationary time series with $\mu_X = 0$ and $\gamma_X(\cdot)$ such that $\gamma_X(h) \to 0$ for $h \to \infty$. Then for each $p \geq 0$ there exists a causal time series $Y = \{Y_t\} \sim AR(p)$ such that $\gamma_Y(h) = \gamma_X(h)$ for $h = 0, 1, \dots, p$.

4.29. Remark

(1) It can be shown that in 4.28 there need not exist for any $q \geq 0$ an $Y = \{Y_t\} \sim MA(q)$ with analogical property, namely that $\gamma_Y(h) = \gamma_X(h)$ for $h = 0, 1, \dots, q$.

(2) As invertible $MA(q)$ process from Theorem 4.27 may equivalently rephrased as a causal $AR(\infty)$-process, we can state that 4.27 is actually a special case of 4.28 which may be interpreted roughly as follows: every stationary time series with decreasing autocovariance (or autocorrelation) function may be well approximated by a causal $AR(p)$ process of sufficienty large order p, making the length of mismatching trailing parts of autocovariance functions arbitrarily small: they completely match at least at the limit $p = \infty$. As $AR(p), p \leq \infty$ is also trivially invertible, it is possible to show that its approximations by a causal and invertible $ARMA(p', q')$ model exist (not unique) reducing with $p' + q' < p$ the number of parameters.

4.30 Theorem [cf. Theorem 3.18]

Let $\{X_t\} \sim ARMA(p,q)$: $\phi(B)X_t = \theta(B)Z_t$ such that $\phi(z)$ and $\theta(z)$ do not have common roots. Then it holds

(i) $\{X_t\}$ is causal iff $\phi(z) \neq 0$ for all $z \in \mathbb{C}$, $|z| \leq 1$ (no roots inside unit disc)
In such a case ψ_j are given by $\psi(z) = \frac{\theta(z)}{\phi(z)} = \sum_{j=0}^{\infty} \psi_j \cdot z^j$, where $\psi := \{\psi_j\}_{j=0}^{\infty} \in l_1$ is unique

(ii) $\{X_t\}$ is invertible iff $\theta(z) \neq 0$ for all $z \in \mathbb{C}$, $|z| \leq 1$ (no roots inside unit d.)
In such a case π_j are given by $\pi(z) = \frac{\phi(z)}{\theta(z)} = \sum_{j=0}^{\infty} \pi_j z^j$, where $\pi := \{\pi_j\}_{j=0}^{\infty}$, $\pi \in l_1$, is unique.

Proof:

(i) $\overset{?}{\Leftarrow}$: $\phi(z) \neq 0$ at $|z| \leq 1 \Rightarrow \frac{1}{\phi(z)}$ has no poles in $|z| \leq 1 \Rightarrow$ exists Taylor expansion $\frac{1}{\phi(z)} = \sum_{j=0}^{\infty} \xi_j \cdot z^j =: \xi(z)$ absolutely convergent on $|z| \leq 1$

$\Rightarrow \sum_{j=0}^{\infty} |\xi_j| < \infty$.

Then $\phi(B)X_t = \theta(B)Z_t \overset{4.4}{\Rightarrow} \underbrace{\xi(B)\phi(B)}_{1} X_t = \underbrace{\xi(B)\theta(B)}_{=:\psi(B)} Z_t$, where

$\psi(z) = \xi(z) \cdot \theta(z) = \frac{\theta(z)}{\phi(z)}$ is absolutely convergent on $|z| \leq 1$ as well (by 3.14(6)), and $\psi \in l_1$ (or $\sum_{j=0}^{\infty} |\psi_j| < \infty$) by 3.14(3).

$\overset{?}{\Rightarrow}$: $\{X_t\}$ causal $\Rightarrow \exists \psi = \{\psi_j\} \in l_1$: $X_t = \sum_{j=0}^{\infty} \psi_j Z_{t-j} \overset{3.14(3)}{\Rightarrow} \psi(z)$ is absolutely convergent on $|z| \leq 1$. Then in view of $\{Z_t\} \sim WN(0, \sigma^2)$, $\sigma > 0$:

(*) $\underbrace{\theta(B) Z_t}_{\sum_{j=0}^{q} \theta_j Z_{t-j}} = \phi(B)X_t \overset{4.4}{=} \underbrace{\phi(B)\psi(B)}_{=:\eta(z)} Z_t = \sum_{j=0}^{\infty} \eta_j Z_{t-j}$ on $|z| \leq 1$ by 3.14(6) $\quad / \langle \cdot, Z_{t-k} \rangle \ \forall k$

$\Downarrow$

$\sum_{j=0}^{q} \theta_j \underbrace{\langle Z_{t-j}, Z_{t-k} \rangle}_{\sigma^2 \delta_{jk}} = \sum_{j=0}^{\infty} \eta_j \underbrace{\langle Z_{t-j}, Z_{t-k} \rangle}_{\sigma^2 \delta_{jk}} \Rightarrow \begin{cases} \eta_k = \theta_k & \text{for } k = 0, 1, \dots, q \\ \eta_k = 0 & \text{for } k > q \end{cases}$
in view of $Z_{t-j} \perp Z_{t-k}$ for $j \neq k$

and consequently $\theta(z) = \phi(z) \cdot \psi(z)$ for $|z| \leq 1$

As $|\psi(z)| < \infty$ for $|z| \leq 1$, it must be $\phi(z) \neq 0$ for $|z| \leq 1$ [if $\phi(z_0) = 0$ for some $|z_0| \leq 1$ then $\theta(z_0) = 0$ as well, and z_0 would be a common root of $\phi(z)$ and $\theta(z)$ – contradiction]. Thus we have $\psi(z) = \frac{\theta(z)}{\phi(z)}$ for $|z| \leq 1$.

Uniqueness: $\sum_{j=0}^{\infty} \psi_j Z_{t-j} = \sum_{j=0}^{\infty} \psi_j' Z_{t-j} \ / \langle \cdot, Z_{t-k} \rangle \Rightarrow \psi_k = \psi_k'$ as in (*).

(ii) By interchanging the role of $\phi(z)$ and $\theta(z)$. When proving the implication $\Rightarrow$ we utilize the identities:

$\phi(B)Z_t \overset{4.4}{=} \phi(B)\pi(B)X_t = \pi(B)\phi(B)X_t = \pi(B)\theta(B)Z_t \overset{\text{as of } (*)}{\Rightarrow} \phi(z) = \pi(z)\theta(z)$

4.31 Remark

If $\{X_t\} \sim ARMA(p,q)$: $\phi(B)X_t = \theta(B)Z_t$ where $\phi(z)$ and $\theta(z)$ have common roots violating the assumption of Theorem 4.30. then a more detailed analysis shows that:

a) if all common roots lie outside the unit circle $|z|=1$, then the problem is not critical: one can still apply Theorem 4.30 after cancelling the common roots when evaluating
$\psi(z) = \frac{\theta(z)}{\phi(z)}$ or $\pi(z) = \frac{\phi(z)}{\theta(z)}$

b) if one of the common roots lies on the unit circle $|z|=1$,
then Theorem 4.30 remains valid with one exception: there may exist more causal or invertible solutions $\{\psi_j\}$ or $\{\pi_j\}$ (uniqueness is lost).

4.32. Computation of ψ_j (or π_j) of a causal (or invertible) ARMA(p,q)

(i) causal representation ψ_j

By 4.30.(i): $\psi(z) = \frac{\theta(z)}{\phi(z)} \iff \psi(z)\phi(z) = \theta(z)$

ψ_j are obtained by comparing coefficients at z^j, $j=0,1,2,\ldots$ of the left-hand side power series product $(\psi_0+\psi_1 z+\psi_2 z^2 \ldots)(1-\phi_1 z-\ldots\phi_p z^p)$ with those on the right-hand side $\theta(z) = 1+\theta_1 z+\ldots+\theta_q z^q$.

At $z^0=1$: $\psi_0 \cdot 1 = 1 \Rightarrow \psi_0 = 1$

$z^1 = z$: $\psi_1 \cdot 1 - \psi_0 \cdot \phi_1 = \theta_1 \Rightarrow \psi_1 = \theta_1 + \underbrace{\psi_0}_{=1}\phi_1 = \theta_1 + \phi_1$

z^2 : $\psi_2 \cdot 1 - \psi_1\phi_1 - \psi_0\phi_2 = \theta_2 \Rightarrow \psi_2 = \theta_2 + (\theta_1+\phi_1)\phi_1 + \phi_2$

$\vdots$

where we set $\theta_j := 0$ for $j>q$ and $\phi_i = 0$ for $i>p$.

Equivalent systems of linear equations in matrix form:

$$(4.9a)\quad \begin{bmatrix} 1 & & & & \\ -\phi_1 & 1 & & 0 & \\ \vdots & & & & \\ -\phi_p & -\phi_{p-1} & \cdots & 1 & \\ 0 & -\phi_p & \cdots & & 1 \\ \vdots & \vdots & & & & \ddots \end{bmatrix} \underbrace{\begin{bmatrix} \psi_0 \\ \psi_1 \\ \psi_2 \\ \vdots \\ \vdots \end{bmatrix}}_{\psi} = \underbrace{\begin{bmatrix} 1 \\ \theta_1 \\ \theta_2 \\ \vdots \\ \theta_q \\ 0 \\ \vdots \end{bmatrix}}_{\theta}$$

columns: $\underline{a}_1$ $\underline{a}_2$, ...

(4.9a) is a SLE with lower triangular (infinite) matrix and the above described procedure is clearly the standard solution of such systems by the so-called forward-substitution technique. Solving $\psi_0 = 1$ from the first equation in the first step we obtain from $\underline{a}_1 \underbrace{\psi_0}_{1} + \underline{a}_2 \psi_1 + \dots = \underline{\theta}$ after moving the first term to the right-hand side:

$\underline{a}_2 \psi_1 + \underline{a}_3 \psi_2 + \dots = \underline{\theta} - \underline{a}_1$, which is another equivalent SLE:

$$(4.9b) \quad \begin{bmatrix} 1 & & & & \\ -\phi_1 & 1 & & \mathcal{O} & \\ \vdots & \vdots & \ddots & & \\ -\phi_p & -\phi_{p-1} & & 1 & \dots \\ 0 & -\phi_p & & \vdots & \ddots \\ \vdots & & & \vdots & \end{bmatrix} \begin{bmatrix} \psi_1 \\ \psi_2 \\ \vdots \\ \vdots \\ \vdots \end{bmatrix} = \begin{bmatrix} \theta_1 + \phi_1 \\ \theta_2 + \phi_2 \\ \vdots \\ \theta_{\max(p,q)} + \phi_{\max(p,q)} \\ 0 \\ \vdots \end{bmatrix}$$

(ii) invertible representation π_j:

Analogously by 4.30(ii): $\pi(z) = \dfrac{\phi(z)}{\theta(z)} \Leftrightarrow \pi(z) \cdot \theta(z) = \phi(z)$

leads again to SLE's:

$$(4.10a) \quad \begin{bmatrix} 1 & & & & \\ \theta_1 & 1 & & \mathcal{O} & \\ \vdots & & & & \\ \theta_q & \theta_{q-1} & \dots & 1 & \dots \\ 0 & \theta_q & \dots & \dots & 1 & \dots \\ \vdots & \vdots & & & & \ddots \end{bmatrix} \begin{bmatrix} \pi_0 \\ \pi_1 \\ \pi_2 \\ \vdots \\ \vdots \\ \vdots \end{bmatrix} = \begin{bmatrix} 1 \\ -\phi_1 \\ -\phi_2 \\ \vdots \\ -\phi_p \\ 0 \\ \vdots \end{bmatrix}$$

or after substituting $\pi_0 = 1$:

$$(4.10b) \quad \begin{bmatrix} 1 & & & & \\ \theta_1 & 1 & & \mathcal{O} & \\ \vdots & & & & \\ \theta_q & \theta_{q-1} & \dots & 1 & \\ 0 & \theta_q & \dots & \dots & 1 & \dots \\ \vdots & \vdots & & & & \end{bmatrix} \begin{bmatrix} \pi_1 \\ \pi_2 \\ \vdots \\ \vdots \\ \vdots \\ \vdots \end{bmatrix} = - \begin{bmatrix} \theta_1 + \phi_1 \\ \theta_2 + \phi_2 \\ \vdots \\ \theta_{\max(p,q)} + \phi_{\max(p,q)} \\ 0 \\ \vdots \end{bmatrix}$$

4.33. Computation of γ_X of a causal ARMA(p,q) process $X = \{X_t\}, \mu_X = 0$

A. Approximate procedure

1. Computing sufficiently many ψ_j from (4.9a) or (4.9b), $1 \le j \le J$
2. X causal $\Rightarrow X_t = \psi(B) Z_t$ by (4.6a) $\overset{(4.7a)}{\Rightarrow}$ $\gamma_X(h) \approx \sigma^2 \sum_{j=0}^{J} \psi_{j+h} \psi_j, \; h \ge 0$

B. Exact procedure

Using the theory of homogeneous linear difference equations with constant coefficients [BD93, §3.3, 3.6] one can derive the following formula for $\gamma_X(h)$:

$$\boxed{\gamma_X(h) = \sum_{i=1}^{k} \sum_{j=0}^{r_i - 1} \beta_{ij} h^j \xi_i^{-h}} \quad \text{for } h \ge \max(p, q+1) - p = \max(0, q-p+1) \tag{4.11}$$

where $\xi_i, i = 1, \dots, k$, are different roots of $\phi(z)$
r_i ... multiplicity of ξ_i, i.e.: $p = \sum_{i=1}^{k} r_i$
$\Big\} \; \phi(z) = -\phi_p \prod_{i=1}^{k} (z - \xi_i)^{r_i}$

↑ connection for the term at z^p to be $-\phi_p$.

β_{ij} ... unknown coefficients which may be obtained from finitely many values $\psi_0 = 1, \psi_1, \dots, \psi_q$ [see below]

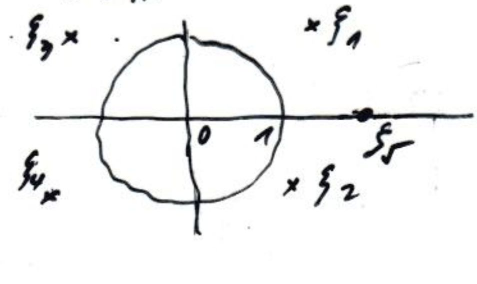
If $q - p + 1 \le 0$, then $h = 0$ is in the admissible range and we get from (4.11) a formula for variance σ_X^2:

$$\boxed{\sigma_X^2 = \gamma_X(0) = \sum_{i=1}^{k} \beta_{i0} \quad \text{for } p > q.} \tag{4.12}$$

4.33.1. Remark

By 4.30(i) we have $|\xi_i| > 1$ due to causality of X. Then $|\xi_i|^{-h} = \left(\frac{1}{|\xi_i|}\right)^h \to 0$ for $h \to \infty$ is faster than $h^j \to \infty \Rightarrow$ $\gamma_X(h) \to 0$ for $h \to \infty$. The rate of decrease is exponential and accelerates with growing distance of the roots ξ_i from the unit circle, i.e. with increasing value $\min_{i=1,\dots,k} |\xi_i|$.

Observe also that ϕ_j are real iff all complex roots form complex-conjugate pairs.



4.33.2 Computation of β_{ij} and $\gamma_X(h)$, $0 \le h < \max(p,q+1)-p$

$\{X_t\} \sim ARMA(p,q)$: $\phi(B)X_t = \Theta(B)Z_t$, or $X_t - \sum_{i=1}^{p}\phi_i X_{t-i} = \sum_{j=0}^{q}\Theta_j Z_{t-j}$

& causal : $X_t = \psi(B)Z_t$, or $X_t = \sum_{m=0}^{\infty}\psi_m Z_{t-m}$

$$\Downarrow$$

$$X_t - \sum_{i=1}^{p}\phi_i X_{t-i} = \sum_{j=0}^{q}\Theta_j Z_{t-j} \quad \Big/ \times X_{t-k} \text{ for } k=0,1,2,\dots$$

multiply

$$\Downarrow$$

$$X_t X_{t-k} - \sum_{i=1}^{p}\phi_i X_{t-i}X_{t-k} = \sum_{j=0}^{q}\Theta_j Z_{t-j}X_{t-k} \quad \Big/ E(\cdot)\dots$$

take expectation and subst. $X_{t-k} = \sum_{m=0}^{\infty}\psi_m Z_{t-k-m}$

$$\Downarrow$$

$$\underbrace{E X_t X_{t-k}}_{\gamma_X(k)} - \sum_{i=1}^{p}\phi_i \underbrace{E X_{t-i}X_{t-k}}_{\gamma_X(k-i)} = \sum_{j=0}^{q}\Theta_j E\Big[Z_{t-j}\sum_{m=0}^{\infty}\psi_m Z_{t-k-m}\Big]$$

$\Downarrow$ by continuity of $E(\cdot,\cdot\cdot) = \langle\cdot,\cdot\cdot\rangle$ with respect to the 2-nd argument ..

$$\underbrace{\gamma_X(k) - \sum_{i=1}^{p}\phi_i\gamma_X(k-i)}_{-\sum_{i=0}^{p}\phi_i\gamma_X(k-i)} = \sum_{j=0}^{q}\Theta_j\sum_{m=0}^{\infty}\psi_m \underbrace{E Z_{t-j}Z_{t-k-m}}_{\sigma^2\delta_{j,k+m}}$$

with $\phi_0 = -1$

by orthogonality $Z_{t-j} \perp Z_{t-k-m}$

$\Downarrow j = k+m \Leftrightarrow m = j-k$

$$= \sigma^2\sum_{j=0}^{q}\Theta_j\psi_{j-k}$$

$$= \sigma^2\sum_{j=k}^{q}\Theta_j\psi_{j-k} \quad \text{because } \psi_{j-k}=0 \text{ for } j-k<0$$

having $\sum_{j=k}^{q}\Theta_j\psi_{j-k} = 0$ for $k \ge q+1$, we arrive at the final form:

$$(4.13)\quad \boxed{\underbrace{\gamma_X(k) - \sum_{i=1}^{p}\phi_i\gamma_X(k-i)}_{LHS} = \begin{cases}\sigma^2\sum_{j=k}^{q}\Theta_j\psi_{j-k} & \text{for } 0 \le k < \max(p,q+1)\\ \underbrace{0}_{RHS} & \text{for } k \ge \max(p,q+1)\end{cases}}$$

(actualy even for $k \ge q+1$ if $p > q+1$)

RHS: $0 \le k < \max(p,q+1)$, $k \le j \le q \Rightarrow 0 \le j \le q \Rightarrow$ we substitute for $q+1$ values ψ_j, $j=0,\dots,q$ the solution of the first $q+1$ equations (4.9a) $\Rightarrow$ RHS values are functions of Θ_j and ϕ_i.

LHS: $0 \le k < \max(p,q+1)$, $0 \le i \le p \Rightarrow 0-p \le k-i \le \max(p,q+1)-1-0 < \max(p,q+1)$ $\Rightarrow 0 \le |k-i| \le \max(p,q+1)$. As $\gamma_X(-h) = \gamma_X(h)$, the LHS is linear combination of $\gamma_X(h)$, $h=|k-i|$, $0 \le h \le \max(p,q+1)$. Substituting from (4.11) for $\gamma_X(h)$, $\max(p,q+1)-p \le h \le \max(p,q+1)$ we obtain $\max(p,q+1)$ linear equations for p unknowns β_{ij} and remaining $\gamma_X(h)$, $0 \le h < \max(p,q+1)-p$, if any.

Let us express that SLE explicitly in matrix form:

Put $K := \max(p, q+1)$, $K' := \max(K, p+1)$, $L := K - p$.

RHS: $A := [a_1, a_2, \dots, a_{q+1}]$ ... $(q+1)$ columns of (4.9a).

Clearly A is nonsingular because $|A| = 1$ (determinant of A)

Solving (4.9a) with unknowns $\underline{\psi} = [\psi_0, \psi_1, \dots, \psi_q]^T$ and right-hand side $\underline{\theta} := [1, \theta_1, \dots, \theta_q]^T$ we obtain $\boxed{\underline{\psi} = A^{-1}\underline{\theta}}$

RHS of (4.13) is then

$$\underline{b} = \sigma^2 \underbrace{\begin{array}{r}\\ \\ \\ \\ K-(q+1)\left\{\begin{array}{c}\\ \\ \end{array}\right.\end{array}\begin{bmatrix} 1 & \theta_1 & \theta_2 & \cdots & \theta_{q-1} & \theta_q \\ \theta_1 & \theta_2 & \theta_3 & \cdots & \theta_q & 0 \\ & & & \iddots & & \vdots \\ \theta_q & 0 & \cdots & & & 0 \\ 0 & \cdots & & & & 0 \\ 0 & \cdots & & & & 0 \end{bmatrix}}_{Q \text{ of size } K\times(q+1)} \cdot \begin{bmatrix} \psi_0 \\ \psi_1 \\ \vdots \\ \psi_q \end{bmatrix} = \sigma^2 Q A^{-1} \underline{\theta}$$

LHS: k-th row:

$$\left[-\sum_{i=0}^{p} \phi_i \gamma_x(k-i)\right] = \underbrace{\underbrace{\sum_{i=0}^{k-1} (-\phi_i)\gamma_x(k-i)}_{=:F^-\underline{g}} + \underbrace{\sum_{i=k}^{p} (-\phi_i)\gamma_x(i-k)}_{=:F^+\underline{g}}}_{(F^- + F^+)\underline{g}} \quad \text{where}$$

$$F := \underbrace{\begin{bmatrix} 0 & 0 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & & & \vdots \\ 0 & -\phi_1 & 1 & & & \vdots \\ 0 & -\phi_2 & -\phi_1 & 1 & & \\ \vdots & & & & \ddots & \\ 0 & \cdots & & & & \end{bmatrix}}_{F^- \text{ of size } K\times K'} + \underbrace{\begin{bmatrix} 1 & -\phi_1 & -\phi_2 & \cdots \\ -\phi_1 & -\phi_2 & -\phi_3 & \cdots \\ -\phi_2 & -\phi_3 & -\phi_4 & \cdots \\ \vdots & & & \end{bmatrix}}_{F^+ \text{ of size } K\times K'} = \begin{bmatrix} 1 & -\phi_1 & -\phi_2 & \cdots \\ -\phi_1 & 1-\phi_2 & -\phi_3 & \cdots \\ -\phi_2 & -\phi_1-\phi_3 & 1-\phi_4 & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{bmatrix}$$

$$\underline{g} := \underset{K'\times 1}{\begin{bmatrix} \gamma_x(0) \\ \vdots \\ \gamma_x(L-1) \\ \gamma_x(L) \\ \gamma_x(L+1) \\ \vdots \\ \gamma_x(K'-1) \end{bmatrix}} = \underbrace{\begin{array}{c}\\ \\ \\ \\ h\rightarrow \\ \\ \\ \end{array}\left[\begin{array}{ccc|cccccc} 1 & & 0 & & & & & & \\ & \ddots & & & & 0 & & & \\ 0 & & 1 & & & & & & \\ \hline & 0 & & & & & & & \\ \cdots & \cdots & \cdots & \xi_0^{-h} & \xi_0^{-h}h & \xi_1^{-h} & \xi_1^{-h}h & \cdots & \\ & 0 & & & & \vdots & & & \\ & L & & & & p & & & \end{array}\right]}_{=:G \text{ of size } K'\times K} \underbrace{\begin{bmatrix} \gamma_x(0) \\ \vdots \\ \gamma_x(L-1) \\ B_{00} \\ B_{01} \\ B_{10} \\ B_{11} \\ \vdots \end{bmatrix}}_{\substack{=:\underline{\beta} \\ K\times 1}} \begin{array}{l} \left.\begin{array}{c}\\ \\ \\ \end{array}\right\} L \\ \left.\begin{array}{c}\\ \\ \\ \\ \end{array}\right\} p \end{array}$$

SLE to solve:

(4.14) $\boxed{F \cdot G\, \underline{\beta} = \sigma^2 Q A^{-1} \underline{\theta}}$

4.34. Computation of α_X of a causal ARMA(p,q) process $X=\{X_t\}$, $\mu_X=0$

By Theorem 4.17: $\alpha_X(n)=\phi_{nn}$ where $\phi_{nn}=\phi_{nn}^{(1)}$ is the last component of the solution to the 1-step best linear prediction problem (4.2)

If Γ_n is ~~regular~~ nonsingular then by (4.2):

$$\underline{\phi}_n=\begin{bmatrix}\phi_{n1}\\ \vdots\\ \phi_{n,n-1}\\ \boxed{\phi_{nn}}\end{bmatrix}=\underbrace{\begin{bmatrix}g_{11} & \cdots & g_{1n}\\ \vdots & & \\ g_{n-1,1} & \cdots & g_{n-1,n}\\ g_{n1} & \cdots & g_{nn}\end{bmatrix}}_{=:\Gamma_n^{-1}}\underbrace{\begin{bmatrix}\gamma_X(1)\\ \vdots\\ \gamma_X(n)\end{bmatrix}}_{\underline{\gamma}_n}\Rightarrow\boxed{\alpha_X(n)=\sum_{j=1}^{n}g_{nj}\,\gamma_X(j)}$$

4.35. Identification of ARMA type and orders based on a.c.f. and p.a.c.f. graphs.

By 4.33.1 each root ξ_i of $\phi(z)$ contributes to the sum (4.11) by r_i (multiplicity of ξ_i) geometrically (exponentially) decreasing terms $\underbrace{\left(\sum_{j=0}^{r_i-1}\beta_{ij}h^j\right)}_{\text{polynomial factor slowing down the decrease if } r_i>1}\xi_i^{-h}$ for $h\geq\max(0,q-p+1)$:

- Single real root $\Rightarrow$ fastest geometrical decrease by the term $\beta_{i0}\xi_i^{-h}$.
- Pair of comlex conjugate roots $\Rightarrow \xi_i^{-h}=|\xi_i|^{-h}(\cos(-h\alpha_i)+i\sin(-h\beta_i))\Rightarrow$ $\Rightarrow$ geometrically damped sinusoidal term

$\rightarrow$ all that mixed together $\Rightarrow$ resulting shape of $\gamma_X(h)$ for $h\geq\max(0,q-p+1)$, hereafter called U-shape

As $\alpha_X(h)$ is a partial autocorrelation to $\rho_X(h)=\frac{\gamma_X(h)}{\gamma_X(0)}$ a similar behaviour of $\alpha_X(h)$ is expected. Indeed, it can be shown that $\alpha_X(\cdot)$ is bounded by a U-shaped curve for $h\geq\max(0,p-q+1)$ — cf. (interchange)

Theorems 4.24 and 4.26 yield so-called identification point (IP):

$X\sim MA(q)\overset{4.24}{\Rightarrow} h_0=q$ is IP for $\gamma_X(h)$: $\gamma_X(h_0)\neq0$ and $\gamma_X(h)=0$ for $h>h_0$

$X\sim AR(p)\overset{4.26}{\Rightarrow} h_0=p$ is IP for $\alpha_X(h)$: $\alpha_X(h_0)\neq0$ and $\alpha_X(h)=0$ for $h>h_0$

Let us summarize all the information into the following table

4.35.1. Table

	AR(p), see Fig. 4.1	MA(q), see Fig. 4.2	ARMA(p,q), see Fig. 4.3
$\rho_X(h)$	IP h_0 deos not exist $\boxed{\text{Shape } \mathcal{U} \text{ for } h \geq 0}$	$\boxed{\text{IP } h_0 = q}$ (See (4.76) in Th. 4.24)	IP h_0 does not exist Shape $\mathcal{U}$ for $h \geq \max(0, q-p+1)$
$\alpha_X(h)$	$\boxed{\text{IP } h_0 = p}$ (see Theorem 4.26)	IP h_0 does not exist $\boxed{\text{Bounded by } \mathcal{U}\text{-shape for } h \geq 0}$	IP h_0 does not exist Bounded by $\mathcal{U}$-shape for $h \geq \max(0, p-q+1)$

4.36. Best linear prediction of zero-mean stationary time series using Innovations Algorithm. [BD 93, §5.2]

In contrast with the Durbin-Levinson algorithm 4.14, which is based on 4.11, we allow $\{X_t\}$ to be a possibly non-stationary process with mean $\mu_X = 0$ for the Innovations Algorithm (IA).

We assume again for simplicity that $t = n$ where n is the length of the prediction: $X_1, X_2, \dots, X_{n+1}$ ... time series up to the time $n+1$ where all $X_i \in L_2$.

Using notation analogical to Def. 4.7:

$\mathcal{L}_i := \mathcal{L}(X_1, \dots, X_i)$ ... linear space spanned by $X_1, \dots, X_i$

$\mathcal{L}_0 := \{0\}$ Clearly $\mathcal{L}_0 \subset \mathcal{L}_1 \subset \dots \subset \mathcal{L}_i$

$\hat{X}_{i+1} := P_{\mathcal{L}_i}(X_{i+1})$ ... best linear predictions for $i = 1, \dots, n$

Clearly $\hat{X}_1 = 0$ which is in accordance with $EX_1 = 0$ by 4.9. due to $\mu_X = 0$.

4.36.1. Theorem [Gram-Schmidt orthogonalization procedure]

$$\mathcal{L}_k = \mathcal{L}(X_1 - \hat{X}_1, \dots, X_k - \hat{X}_k) \text{ for } k = 1, \dots, n \text{ where } X_j - \hat{X}_j \perp X_k - \hat{X}_k \text{ for } j \neq k$$

Proof: I. $\hat{X}_j \in \mathcal{L}_{j-1} \subset \mathcal{L}_k$, $X_j \in \mathcal{L}_j \subseteq \mathcal{L}_k$ for $j = 1, \dots, k \Rightarrow X_j - \hat{X}_j \in \mathcal{L}_k$ for $j = 1, \dots, k \Rightarrow \mathcal{L}(X_1 - \hat{X}_1, \dots, X_k - \hat{X}_k) \subseteq \mathcal{L}_k$

II. Denote $\mathcal{L}_k' := \mathcal{L}(X_1-\hat X_1, \dots, X_k - \hat X_k)$. We prove by induction on k that $\mathcal{L}_k \subseteq \mathcal{L}_k'$ for $k=1,\dots,n$:

$\underline{k=1}$: $\mathcal{L}_1 = \mathcal{L}(X_1) = \mathcal{L}(X_1 - \underbrace{\hat X_1}_{0}) = \mathcal{L}_1'$

$\underline{k>1}$: $\mathcal{L}_{k-1} \subseteq \mathcal{L}_{k-1}' \subseteq \mathcal{L}_k'$ by induction hypothesis. Then

$X_1,\dots,X_{k-1} \in \mathcal{L}_{k-1} \subseteq \mathcal{L}_k'$ and $X_k = \underbrace{X_k - \hat X_k}_{\in \mathcal{L}_k'} + \underbrace{\hat X_k}_{\in \mathcal{L}_{k-1} \subseteq \mathcal{L}_k'} \in \mathcal{L}_k'$ as well $\Rightarrow$

$\Rightarrow X_1,\dots,X_k \in \mathcal{L}_k' \Rightarrow \mathcal{L}_k = \mathcal{L}(X_1,\dots,X_k) \subseteq \mathcal{L}_k'$.

III. Let $j \neq k$, e.g. $j<k$. Then $X_k - \hat X_k \perp \mathcal{L}_{k-1}$ by the orthogonal projection theorem. As $X_j - \hat X_j \in \mathcal{L}_j' = \mathcal{L}_j \subseteq \mathcal{L}_{k-1}$, we have $X_k - \hat X_k \perp X_j - \hat X_j$.

4.36.2. Corollary. It holds for $k=1,\dots,n$, and $v_k := \|X_{k+1} - \hat X_{k+1}\|^2 = E|X_{k+1} - \hat X_{k+1}|^2$

$$\hat X_{k+1} = \sum_{j=1}^{k} \theta_{kj} (X_{k+1-j} - \hat X_{k+1-j}) \text{ where } v_{k-j}\,\theta_{kj} = \langle \hat X_{k+1}, X_{k+1-j} - \hat X_{k+1-j}\rangle = \langle X_{k+1}, X_{k+1-j} - \hat X_{k+1-j}\rangle$$

Proof: $\hat X_{k+1} \in \mathcal{L}_k = \mathcal{L}(X_1 - \hat X_1, \dots, X_k - \hat X_k) \Rightarrow \hat X_{k+1}$ must be a linear combination of $X_1 - \hat X_1, \dots, X_k - \hat X_k$: $\exists\, \theta_{kj}$: $\hat X_{k+1} = \sum_{j=1}^{k} \theta_{kj} (X_{k+1-j} - \hat X_{k+1-j})$

By orthogonality of $e_i := X_{k+1-i} - \hat X_{k+1-i}$ for $i=0,1,\dots,k$:

$$\langle X_{k+1}, e_i\rangle = \langle \underbrace{X_{k+1} - \hat X_{k+1}}_{e_0} + \hat X_{k+1}, e_i\rangle = \underbrace{\langle e_0, e_i\rangle}_{=0} + \langle \hat X_{k+1}, e_i\rangle =$$

$$= \langle \sum_{j=1}^{k} \theta_{kj} e_j, e_i\rangle = \sum_{j=1}^{k} \theta_{kj} \underbrace{\langle e_j, e_i\rangle}_{v_{k-i}\,\delta_{ij}} = v_{k-i}\,\theta_{ki} \quad \text{for } i=1,\dots,k.$$

4.36.3. Theorem. Innovations algorithm

If $\{X_t\}$ has zero mean and $\varkappa(i,j) := E X_i X_j$ where the matrix $[\varkappa(i,j)]_{i,j=1}^{n}$ is non-singular for each $n=1,2,\dots$, then the one-step predictors $\hat X_{n+1}$, $n \geq 0$, and their mean squared errors v_n, $n \geq 1$, are given by

$$(4.15) \quad \hat X_{n+1} = \begin{cases} 0 & \text{if } n=0 \\ \sum_{j=1}^{n} \theta_{nj} (X_{n+1-j} - \hat X_{n+1-j}) & \text{if } n \geq 1 \end{cases}$$

and

$$\boxed{\begin{aligned} v_0 &= \varkappa(1,1) \\ \theta_{n,n-k} &= v_k^{-1} \Big[\varkappa(n+1,k+1) - \sum_{j=0}^{k-1} \theta_{k,k-j}\, \theta_{n,n-j}\, v_j \Big] \\ v_n &= \varkappa(n+1,n+1) - \sum_{j=0}^{n-1} \theta_{n,n-j}^2\, v_j . \end{aligned}}$$

for $k=0,1,\dots,n-1$ recursively in the order $v_0;\ \theta_{11}, v_1;\ \theta_{22}, \theta_{21}, v_2; \dots$

Proof: $v_0 = E|X_1 - \overset{\overset{0}{\approx}}{\hat{X}_1}|^2 = E|X_1|^2 = EX_1X_1 = \mathcal{K}(1,1)$.

By 4.36.2 (with n instead of k and n-k instead of j):

(*) $\langle X_{n+1}, X_{k+1} - \hat{X}_{k+1} \rangle = \theta_{n,n-k} v_k$ for $k = 0, 1, \dots, n-1$

Nonsingularity of $[\mathcal{K}(i,j)] \Rightarrow v_k \neq 0$ [otherwise $v_k = \|X_{k+1} - \hat{X}_{k+1}\|^2 = 0$ $\Rightarrow X_{k+1} = \hat{X}_{k+1} \in \mathcal{L}(X_1, \dots, X_k) \Rightarrow \exists c_m : X_{k+1} = \sum_{m=1}^{k} c_m X_m \Rightarrow \mathcal{K}(k+1, j) =$
$= E(\sum_{m=1}^{k} c_m X_m) X_j = \sum_{m=1}^{k} c_m E X_m X_j = \sum_{m=1}^{k} c_m \mathcal{K}(m,j) \quad \forall j \Rightarrow$
(k+1)-th row is linear combination of preceding rows, which contradicts nonsingularity].

Thus we can divide (*) by v_k to obtain:

$$\theta_{n,n-k} = v_k^{-1} \langle X_{n+1}, X_{k+1} - \hat{X}_{k+1} \rangle = v_k^{-1} \{ \langle X_{n+1}, X_{k+1} \rangle - \langle X_{n+1}, \hat{X}_{k+1} \rangle \} =$$

$$v_k^{-1} \{ \mathcal{K}(n+1, k+1) - \sum_{j=1}^{k} \theta_{k,j} \langle X_{n+1}, X_{k+1-j} - \hat{X}_{k+1-j} \rangle \} \overset{j \mapsto k-j}{=}$$

$$= v_k^{-1} \{ \mathcal{K}(n+1, k+1) - \sum_{j=0}^{k-1} \theta_{k,k-j} \underbrace{\langle X_{n+1}, X_{j+1} - \hat{X}_{j+1} \rangle}_{\theta_{n,n-j} v_j \text{ by } (*)} \}.$$

$$v_n = \|X_{n+1} - \hat{X}_{n+1}\|^2 = \langle X_{n+1} - \hat{X}_{n+1}, X_{n+1} - \hat{X}_{n+1} \rangle =$$

$$\langle X_{n+1}, X_{n+1} - \hat{X}_{n+1} \rangle - \underbrace{\langle \hat{X}_{n+1}, X_{n+1} - \hat{X}_{n+1} \rangle}_{=0} =$$

$$\langle X_{n+1}, X_{n+1} \rangle - \langle X_{n+1} - \hat{X}_{n+1} + \hat{X}_{n+1}, \hat{X}_{n+1} \rangle = \mathcal{K}(n+1, n+1) -$$

$$- \underbrace{\langle X_{n+1} - \hat{X}_{n+1}, \hat{X}_{n+1} \rangle}_{=0} - \langle \hat{X}_{n+1}, \hat{X}_{n+1} \rangle \overset{\mathcal{K}(n+1,n+1)}{=} - \langle \sum_{k=0}^{n-1} \theta_{n,n-k} (X_{k+1} - \hat{X}_{k+1}),$$

$$\sum_{j=0}^{n-1} \theta_{n,n-j} (X_{j+1} - \hat{X}_{j+1}) \rangle \overset{\mathcal{K}(n+1,n+1)}{=} - \sum_{k=0}^{n-1} \sum_{j=0}^{n-1} \theta_{n,n-k} \theta_{n,n-j} \underbrace{\langle X_{k+1} - \hat{X}_{k+1}, X_{j+1} - \hat{X}_{j+1} \rangle}_{v_k \delta_{jk}} =$$

$$\mathcal{K}(n+1, n+1) - \sum_{k=0}^{n-1} \theta_{n,n-k}^2 v_k.$$