Closure of Linear Processes

Peter J. Bickel * and Peter Bühlmann †
University of California
Berkeley

September 1996

Abstract

We consider the sets of moving-average and autoregressive processes and study their closures under the Mallows metric and the total variation convergence on finite dimensional distributions. These closures are unexpectedly large, containing non-ergodic processes which are Poisson sums of i.i.d. copies from a stationary process. The presence of these non-ergodic Poisson sum processes has immediate implications. In particular, identifiability of the hypothesis of linearity of a process is in question.

A discussion of some of these issues for the set of moving-average processes has already been given without proof in Bickel and Bühlmann (1996). We establish here the precise mathematical arguments and present some additional extensions: results about the closure of autoregressive processes and natural sub-sets of moving-average and autoregressive processes which are closed.

Key words and phrases. AR process, Infinitely divisible law, MA process, distinction from non-linear process.

Short title: Closure of linear processes

*Research supported in part by grants NSA MDA 904-94-H-2020 and NSF DMS 95049555.
†Research supported in part by the Swiss National Science Foundation.

1 Introduction

We consider the characterization of closures of sets of stationary stochastic processes $(X_t)_{t\in\mathbb{Z}}$, $\mathbb{Z} = \{0, \pm 1, \pm 2, \ldots\}$, in order to obtain interesting implications about testing hypotheses, such as linearity of a process. A preliminary discussion of this issue has been given in Bickel and Bühlmann (1996), for additional motivation and interpretation we refer to that article.

A linear process $(X_t)_{t\in\mathbb{Z}}$ is most often referred to

$$X_t = \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j} \ (t \in \mathbb{Z}), \tag{1.1}$$

where $(\varepsilon_t)_{t\in\mathbb{Z}}$ is an i.i.d. sequence with $\mathbb{E}[\varepsilon_t] = 0$, $\mathbb{E}|\varepsilon_t|^2 < \infty$ and $\sum_{j=0}^{\infty} \psi_j^2 < \infty$. Such processes are also called moving-average (MA) processes. Here, we always assume existence of second moments. There is no loss of generality in assuming $\mathbb{E}[X_t] = 0$.
Under some circumstances, when the MA transfer function $\Psi(z) = \sum_{j=0}^{\infty} \psi_j z^j$ exists and has no zeros in $|z| \le 1$ ($z \in \mathbb{C}$), then such an MA process can be inverted and is also representable as an (invertible) autoregressive (AR) process

$$X_t = \sum_{j=1}^{\infty} \phi_j X_{t-j} + \varepsilon_t \ (t \in \mathbb{Z}), \tag{1.2}$$

where the coefficients $(\phi_j)_{j\in\mathrm{N}}$ are given by $1/\Psi(z) = 1 - \sum_{j=1}^{\infty} \phi_j z^j$ ($|z| \le 1,\ z \in \mathbb{C}$), cf. Hannan (1987).

We study here the closure of MA processes as given in (1.1) (MA closure) and of AR processes as given in (1.2) (AR closure). The notion of a closed set requires the specification of a topology. We work here with the Mallows metric (Mallows, 1972), also known as the Wasserstein metric, and with the variation metric. For details see section 2. We always identify real-valued stochastic processes, indexed by $\mathbb{Z}$, with their corresponding probability distributions; we then prefer to state our results in terms of stochastic processes.

Somewhat surprisingly, the set of MA processes as well as the set of invertible AR processes is not closed. We will show that the MA and AR closures are exhausted by three types of processes. The first type is the set of stationary Gaussian processes with mean zero, i.e.,

$$S_1 = \{(X_t)_{t\in\mathbb{Z}}; (X_t)_{t\in\mathbb{Z}} \text{ stationary Gaussian process with } \mathbb{E}[X_t] = 0\}.$$

The second type is the set of genuine MA processes, i.e.,

$$S_2 = \{(X_t)_{t\in\mathbb{Z}}; X_t \text{ as defined in (1.1)}\}.$$

The third type which arises is more surprising. We essentially can get Poisson sums of independent and identically distributed copies of stationary processes in the following sense. Denote by

$$(\xi_{t;1})_{t\in\mathbb{Z}},\ (\xi_{t;2})_{t\in\mathbb{Z}}, \ldots$$

a sequence of independent, real-valued, stationary processes with mean zero and finite second moments $\mathbb{E}|\xi_{t;1}|^2 = \sigma^2_{\xi;1}$, $\mathbb{E}|\xi_{t;2}|^2 = \sigma^2_{\xi;2}$, ... Moreover, we construct for every $i \in \mathbb{N} = \{1, 2, \ldots\}$ a sequence of independent copies of $(\xi_{t;i})_{t\in\mathbb{Z}}$, namely

$$(\xi_{t;i,1})_{t\in\mathbb{Z}},\ (\xi_{t;i,2})_{t\in\mathbb{Z}}, \ldots$$

Thus we have constructed a sequence of processes

$$\begin{aligned}&\{(\xi_{t;i,j})_{t\in\mathbb{Z}}\}_{i,j\in\mathbb{N}} \text{ independent processes over the index set } i,j \in \mathbb{N},\\ &(\xi_{t;i,1})_{t\in\mathbb{Z}},\ (\xi_{t;i,2})_{t\in\mathbb{Z}}, \ldots \text{ i.i.d., } \mathbb{E}[\xi_{t;i,j}] = 0 \text{ for all } j \in \mathbb{N}.\end{aligned} \tag{1.3}$$

Let

$$N_1,\ N_2, \ldots \text{ independent, } N_i \sim \text{ Poisson}(\lambda_i),\ \lambda_i \geq 0 \text{ for all } i \in \mathbb{N}. \tag{1.4}$$

Then the third type is given by the following set of processes,

$$\begin{aligned}S_3 = \{ \quad &(X_t)_{t\in\mathbb{Z}};\ X_t = \sum_{i=1}^{\infty}\sum_{j=1}^{N_i} \xi_{t;i,j},\ (\xi_{t;i,j})_{t\in\mathbb{Z}},\ N_i \text{ satisfying (1.3), (1.4)}\\ &\text{and } \sum_{i=1}^{\infty} \lambda_i \sigma^2_{\xi;i} < \infty\}.\end{aligned}$$

We make the convention that $\sum_{j=1}^{0} \xi_{t;i,j} = 0$. Elements of S_3, are typically non-ergodic processes whose finite dimensional distributions are infinitely divisible non-Gaussian.

The different sets $S_1,\ S_2,\ S_3$ are not disjoint and the representations are not unique. Also, to exhaust the MA and AR closures we need sums of processes of the different types. We introduce an adding operation for processes and define

$$\begin{aligned}&(X_t)_{t\in\mathbb{Z}} \oplus (Y_t)_{t\in\mathbb{Z}} \text{ is the process } (X_t + Y_t)_{t\in\mathbb{Z}}, \text{ where}\\ &\text{the processes } (X_t)_{t\in\mathbb{Z}} \text{ and } (Y_t)_{t\in\mathbb{Z}} \text{ are independent.}\end{aligned}$$

We then set

$$S_i \oplus S_j = \{(X_t)_{t\in\mathbb{Z}} \oplus (Y_t)_{t\in\mathbb{Z}};\ (X_t)_{t\in\mathbb{Z}} \in S_i,\ (Y_t)_{t\in\mathbb{Z}} \in S_j\},\ i,j \in \{1,2,3\},$$

and make the common convention that all S_i $(i = 1,2,3)$ also contain the null element $X_t \equiv 0$ for all $t \in \mathbb{Z}$.

We now summarize the discussion in Bickel and Bühlmann (1996) and some new results in a rather narrative way, without giving here the regularity assumptions we work with. The precise formulations and proofs are given in sections 2 – 5.

Fact 1.1 *The closure of the set of MA processes is characterized by*

$$MA\ closure = \{S_1 \oplus S_2\} \cup \{S_1 \oplus S_3\}.$$

Details are given in Theorem 3.1 and Theorem 3.2.

The limiting operation of sequences of MA processes converging to processes in S_3 is constructive.

Example 1.1 Consider the sequence of finite order MA processes,

$$X_t^{(n)} = \sum_{j=1}^{n} \xi_{j;1} U_{t-j;n} Z_{t-j;n} \ (t \in \mathbb{Z}),$$

with U_t i.i.d., $\mathbb{P}[U_t = 1] = 1 - \mathbb{P}[U_t = 0] = \lambda/n$ $(\lambda > 0)$, Z_t i.i.d. $\sim t_5$, Student's-t distribution with 5 degrees of freedom, and coefficients $(\xi_{j;1})_{j \in \mathrm{N}}$ which are a fixed realization of the Gaussian AR(1), $\xi_{j;1} = 0.9\xi_{j-1;1} + \eta_j$, η_j i.i.d. $\sim \mathcal{N}(0,1)$.
For every $n \in \mathbb{N}$, these are ergodic MA(n) processes. But for large n, they exhibit a behavior which can be interpreted as non-ergodic and non-stationary and which seems far from what one expects of a linear process. The reason is that they are then close to a non-ergodic member in S_3, see proof of Theorem 3.1 (ii), in particular formula (5.11).

We show in Figures 1.1 – 1.4 four long realizations of sample size 5000 of the process in Example 1.1 with $n = 5,\ 25,\ 50,\ 200$, always with the same realization $(\xi_{j;1})_{j \in \mathrm{N}}$. For small n, the realizations in Figures 1.1 and 1.2 look stationary. But for larger n, Figures 1.3 and 1.4 tell us that different stretches of the sequences exhibit very different behaviors, indicating non-stationarity and non-ergodicity. This is the typical pattern for a time series with innovation outliers, cf. Kleiner et al. (1979). Indeed, our model is an extreme case with innovations being either zero with probability $1 - \lambda/n$ or being a realization from a long-tailed distribution with probability λ/n. Note that outliers are with reference to the Gaussian distribution; it is the non-normality of innovations which can lead to MA processes being close to a process in S_3.

Fact 1.2 *Given any infinitely long realization $(\xi_t)_{t \in \mathbb{Z}}$ of a stationary process, there exists a non-ergodic, stationary process $(X_t)_{t \in \mathbb{Z}}$ in the MA closure, being an element of S_3, having with positive probability exactly the same sample path as $(\xi_t)_{t \in \mathbb{Z}}$. More precisely,*

$$\mathbb{P}[X_t = \xi_t \ \textit{for all}\ t \in \mathbb{Z} | (\xi_t)_{t \in \mathbb{Z}}] > 0.36 \ \textit{almost surely.}$$

Details are given in Theorem 3.3. This separation dilemma is evidently related to de Finetti's Theorem about the impossibility of distinguishing exchangeable from i.i.d. sequences, cf. Hartigan (1983, Ch. 4.6).

Fact 1.2 can be restated as,

Fact 1.3 *In testing the hypothesis H_0 about MA representation against any fixed one-point alternative H_A about a nonlinear, stationary process, there is no test with asymptotic significance level $\alpha < 0.36$ having limiting power one as the sample size tends to infinity.*

In some cases, there is a way out of the separation and testing dilemma.

Fact 1.4 *There exists a closed subset of MA processes with nice densities with respect to the Lebesgue measure for the innovations and with MA coefficients $(\psi_j)_{j \in \mathrm{N}_0}$ decaying as fast as $\sum_{j=0}^{\infty} j^{\beta} \psi_j^2 < \infty$ for some $\beta > 0$.*

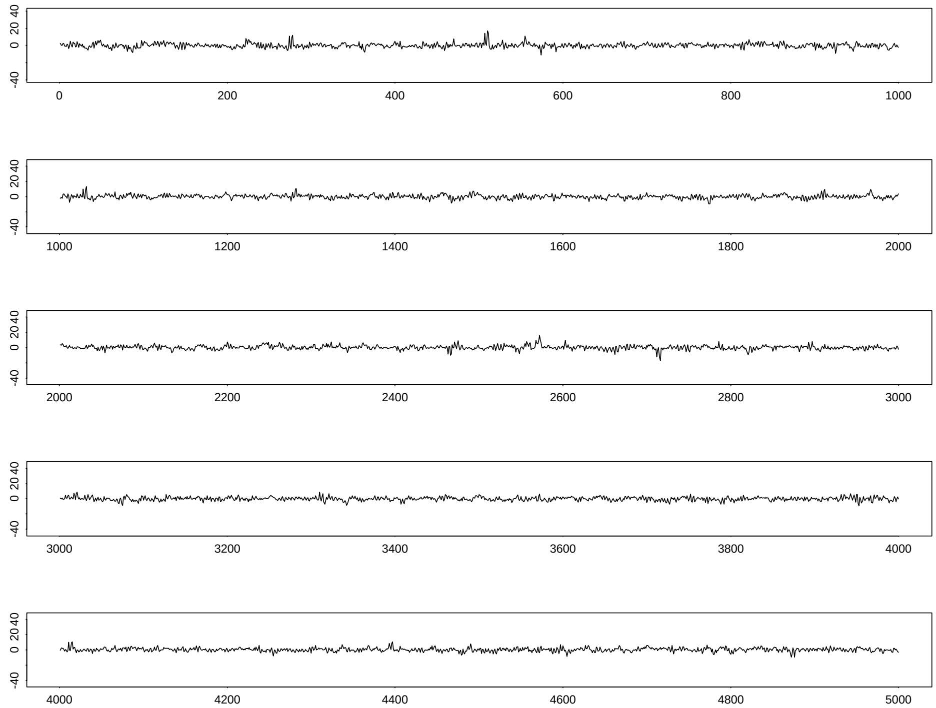

Figure 1.1: one long realization of Example 1.1 with $n = 5, \ \lambda = 5$

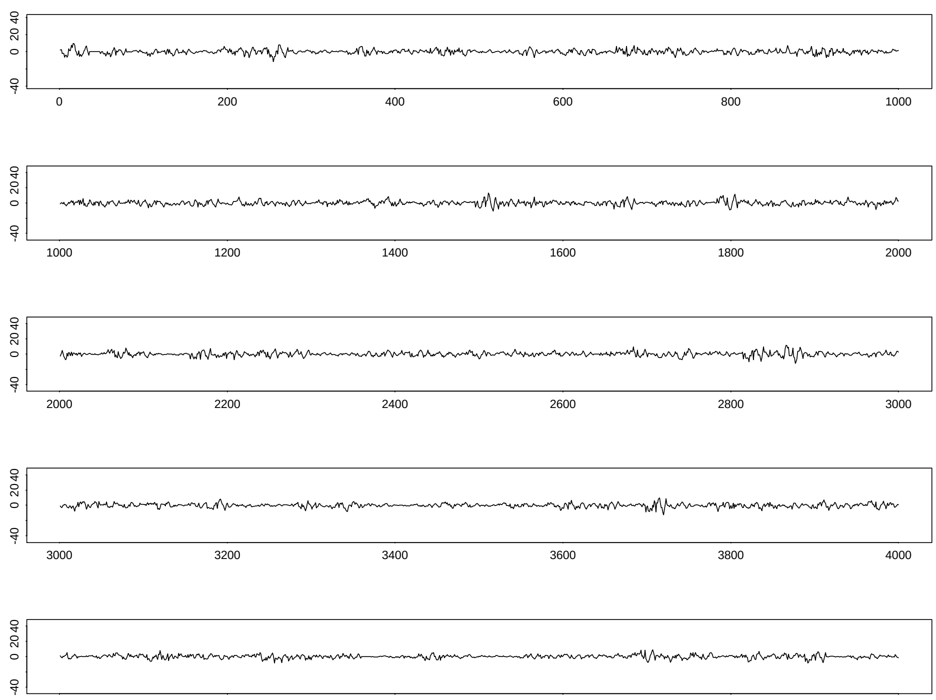

Figure 1.2: one long realization of Example 1.1 with $n = 25, \ \lambda = 5$

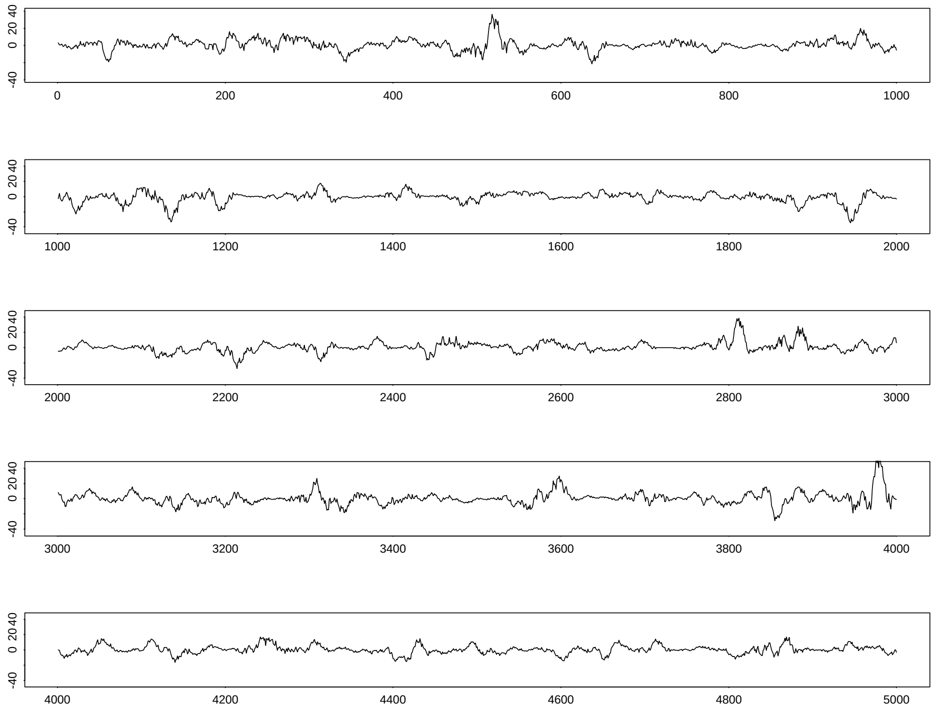

Figure 1.3: one long realization of Example 1.1 with $n = 50,\ \lambda = 5$

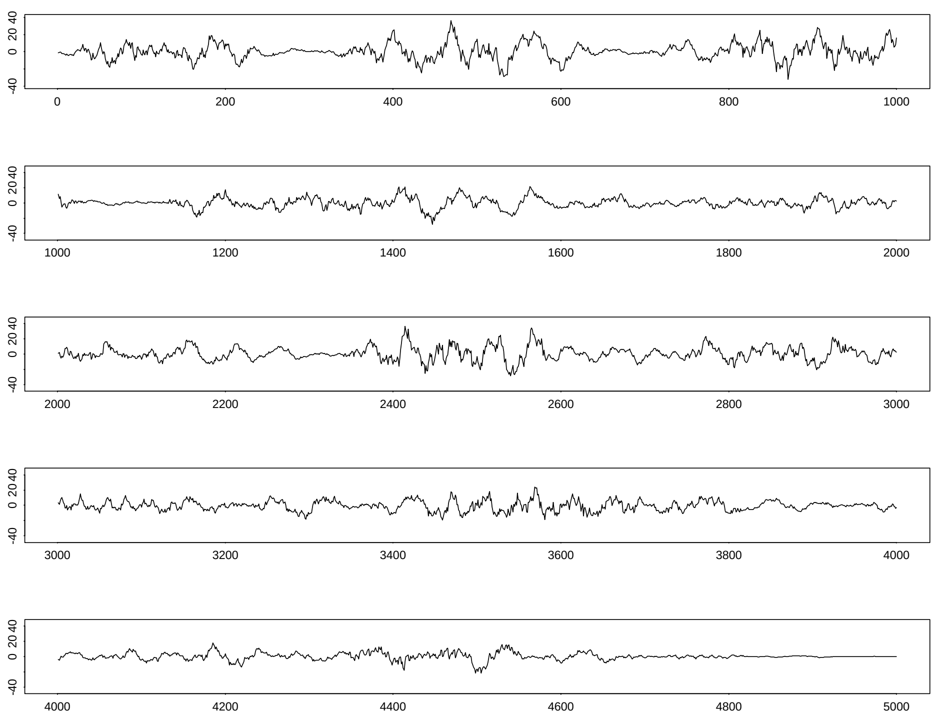

Figure 1.4: one long realization of Example 1.1 with $n = 200,\ \lambda = 5$

Details are given in Theorem 3.4.

For the closure of autoregressive processes we obtain the following,

Fact 1.5 *The closure of the set of invertible AR processes is described by*

$$AR\ closure \subset \{S_1 \oplus S_2\} \cup \{S_1 \oplus S_3\}.$$

Details are given in Theorem 4.1, Theorem 4.2 and Proposition 4.1. Similar to Fact 1.4 we have,

Fact 1.6 *There exists a closed subset of invertible (causal) AR processes with nice densities with respect to the Lebesgue measure for the innovations and with AR coefficients* $(\phi_j)_{j\in\mathbb{N}}$ *decaying as fast as* $\sum_{j=1}^{\infty} j^\beta|\phi_j| < \infty$ *for some* $\beta > 1$. *In particular, all the elements in such a closure are ergodic processes.*

Details are given in Theorem 4.3.

2 Probability space, metric and closure

Our framework is the following. We consider real-valued, stationary processes $(X_t)_{t\in\mathbb{Z}}$ with expectation zero and finite variances. The expectation zero assumption is not restrictive. Thus, an appropriate probability space is $(\mathbb{R}^{\mathbb{Z}}, \mathcal{B}, P)$, where $\mathcal{B}$ denotes the Borel σ-field on $\mathbb{R}^{\mathbb{Z}}$ and $\mathcal{P}$ a class of stationary probability measures on $(\mathbb{R}^{\mathbb{Z}}, \mathcal{B})$, such that for every $P \in \mathcal{P}$,

$$\mathbf{E}_P[X] = \int_{\mathbb{R}} x\, d(P \circ \pi_0^{-1})(x) = 0,\ \mathbf{E}_P|X|^2 = \int_{\mathbb{R}} x^2 d(P \circ \pi_0^{-1})(x) < \infty,$$

where $\pi_{t_1,\ldots,t_m} : \mathbb{R}^{\mathbb{Z}} \to \mathbb{R}^m,\ (x_t)_{t\in\mathbb{Z}} \mapsto (x_{t_1},\ldots,x_{t_m}),\ t_1,\ldots,t_m \in \mathbb{Z}$.
We always identify a probability measure $P \in \mathcal{P}$ with its corresponding real-valued stochastic process.

The space $\mathcal{P}$ can be equipped with a metric d, examples will be given in sections 2.1 and 2.2. We also use the notation for the corresponding processes on $\mathbb{R}^{\mathbb{Z}}$, $d((X_t)_{t\in\mathbb{Z}}, (Y_t)_{t\in\mathbb{Z}}) = d(P, Q)$, where $(X_t)_{t\in\mathbb{Z}} \sim P$, $(X_t)_{t\in\mathbb{Z}} \sim Q$. Such a metric d induces then the closure of sets in $\mathcal{P}$ in the usual topological sense.

Definition 2.1 *Let A be a set of real-valued, stationary processes, indexed by $\mathbb{Z}$, with corresponding probability measures in $\mathcal{P}$. The d closure $\bar{A}$ of A is defined as*

$$\begin{aligned}\bar{A} = \{ \quad & (X_t)_{t\in\mathbb{Z}};\ \exists \text{ a sequence } \{(X_{t;n})_{t\in\mathbb{Z}}\}_{n\in\mathbb{N}} \text{ with } (X_{t;n})_{t\in\mathbb{Z}} \in A \text{ for all } n \in \mathbb{N} \\ & \text{such that } d((X_{t;n})_{t\in\mathbb{Z}}, (X_t)_{t\in\mathbb{Z}}) \to 0\ (n \to \infty)\}.\end{aligned}$$

We are particularly interested in the d closures of moving average (MA) and autoregressive (AR) processes. Thus, we will consider sequences

$$\text{MA processes: } \{(X_{t;n} = \sum_{j=0}^{\infty} \psi_{j;n}\varepsilon_{t-j;n})_{t\in\mathbb{Z}}\}_{n\in\mathbb{N}}, \tag{2.1}$$

$$\text{AR processes: } \{(X_{t;n} = \sum_{j=1}^{\infty} \phi_{j;n}X_{t-j;n} + \varepsilon_{t;n})_{t\in\mathbb{Z}}\}_{n\in\mathbb{N}}. \tag{2.2}$$

We look here at MA and AR processes of infinite order. All our results are also true for sequences of finite (generally unbounded) order MA and AR processes, which are more common in statistical modeling. In the sequel we refer to the MA and AR closure with respect to the d metric the closure of sequences of MA and AR processes respectively, as given by Definition 2.1.

2.1 Mallows metric

The Mallows metric d_2 on $\mathcal{P}$, related to non-uniform weak convergence for finite dimensional distributions, is defined by

$$d_2(P_1, P_2) = \sum_{m=1}^{\infty} d_2^{(m)}(P_1 \circ \pi_{1,\ldots,m}^{-1}, P_2 \circ \pi_{1,\ldots,m}^{-1}) 2^{-m}, \ P_1,\ P_2 \in \mathcal{P},$$

where $d_2^{(m)}(P_1 \circ \pi_{1,\ldots,m}^{-1}, P_2 \circ \pi_{1,\ldots,m}^{-1}) = \inf\{(\mathbb{E}\|X - Y\|^2)^{1/2}\}$ when the infimum is taken over all jointly distributed $(X, Y) \in \mathbb{R}^{2m}$ having marginals $P_1 \circ \pi_{1,\ldots,m}^{-1}$ and $P_2 \circ \pi_{1,\ldots,m}^{-1}$; $\|.\|$ denotes the Euclidean norm in $\mathbb{R}^m$.

The following characterization is useful. Let $P_n,\ P \in \mathcal{P}$ and denote by $\Rightarrow$ weak convergence of probability measures. Then,

$$d_2(P_n, P) \to 0 \ (n \to \infty)$$

is equivalent to the following two statements

$$P_n \circ \pi_{t_1,\ldots,t_m}^{-1} \Rightarrow P \circ \pi_{t_1,\ldots,t_m}^{-1} \ (n \to \infty) \text{ for all } t_1, \ldots, t_m \in \mathbb{Z},\ m \in \mathbb{N},$$
$$\int_{\mathbf{R}} x^2 d(P_n \circ \pi_0^{-1})(x) \to \int_{\mathbf{R}} x^2 d(P \circ \pi_0^{-1})(x) \ (n \to \infty),$$

i.e., all finite dimensional distributions at $t_1, \ldots, t_m$ converge weakly and the variance of the one-dimensional marginal converges, see Bickel and Freedman (1981).

2.2 Variation metric

The question about distinguishing perfectly between two stationary processes requires a stronger metric than the Mallows d_2. The variation metric allows a precise formulation.

As before, let $P_1,\ P_2 \in \mathcal{P}$ and define the variation metric as

$$d_V(P_1, P_2) = \sum_{m=1}^{\infty} d_V^{(m)}(P_1 \circ \pi_{1,\ldots,m}^{-1}, P_2 \circ \pi_{1,\ldots,m}^{-1}) 2^{-m},$$

where $d_V^{(m)}(P_1 \circ \pi_{1,\ldots,m}^{-1}, P_2 \circ \pi_{1,\ldots,m}^{-1}) = \sup\{|P_1 \circ \pi_{1,\ldots,m}^{-1}[A] - P_2 \circ \pi_{1,\ldots,m}^{-1}[A]|;\ A \in \mathcal{B}(\mathbb{R}^m)\}$, $\mathcal{B}(\mathbb{R}^m)$ the Borel σ-field of $\mathbb{R}^m$. This definition reflects non-uniform convergence of finite dimensional distributions in the variation metric. Here we do not require convergence of second moments. Distinguishing perfectly is characterized as follows. Let $P_1,\ P_2$ be ergodic probability measures in $\mathcal{P}$. Then

$d_V(P_1, P_2) > 0$ if and only if
there exist test functions $\varphi_m : \mathbb{R}^m \to \mathbb{R},\ 0 \le \varphi_m \le 1,$ such that
$\mathbf{E}_{P_1}[\varphi_m(X_1, \ldots, X_m)] \to 0,\ \mathbf{E}_{P_2}[\varphi_m(X_1, \ldots, X_m)] \to 1 \ (m \to \infty).$

Note that such a sharp separation is only possible with the variation metric d_V but not with the Mallows metric d_2. Some of our results are in terms of the Mallows metric d_2. However, with regard to Facts 1.2 and 1.3 we will also use the stronger formulation in terms of the variation metric d_V.

3 Closure for MA processes

We consider first the Mallows d_2 closure for MA processes, i.e., we consider sequences as defined in (2.1). Without loss of generality we can scale the innovations and assume:

(**A.MA**): For every $n \in \mathbb{N}$, $(\varepsilon_{t;n})_{t\in\mathbb{Z}}$ is an i.i.d. sequence with

$$\mathbb{E}[\varepsilon_{t;n}] = 0,\ \mathbb{E}|\varepsilon_{t;n}|^2 = 1.$$

Under the assumption (A.MA) and assuming that $(X_{t;n})_{t\in\mathbb{Z}}$ converges in the d_2 metric, the behavior of the coefficients $(\psi_{j;n})_{j\in\mathbb{N}_0}$ is determined in the following way,

$$\mathbb{E}|X_{t;n}|^2 = \sum_{j=0}^{\infty} \psi_{j;n}^2 < \infty \text{ and } \lim_{n\to\infty} \mathbb{E}|X_{t;n}|^2 = \lim_{n\to\infty} \sum_{j=0}^{\infty} \psi_{j;n}^2 < \infty. \tag{3.1}$$

The following result describes the MA closure.

Theorem 3.1 *The closure of MA processes with respect to the Mallows metric d_2 is characterized as follows.*
(i) Consider a sequence of MA processes as defined in (2.1) converging in the d_2 metric, satisfying (A.MA) and one of the following:

(A1): $\lim_{n\to\infty} d_2^{(1)}(\varepsilon_{t;n}, \varepsilon_t) = 0$, *where $(\varepsilon_t)_{t\in\mathbb{Z}}$ is an i.i.d. sequence with $\mathbb{E}[\varepsilon_t] = 0$.*

(A2): $\lim_{n\to\infty} \max_{j\geq 0} |\psi_{j;n}| = 0$.

Then, the d_2 limit of such a sequence is in $\{S_1 \oplus S_2\} \cup \{S_1 \oplus S_3\}$.
(ii) Every element of $\{S_1 \oplus S_2\} \cup \{S_1 \oplus S_3\}$ can be obtained as a d_2 limit of a sequence of MA processes as defined in (2.1), satisfying (A.MA) and (A1) or (A2).

Remark 3.1: Assumptions (A1) and (A2) are not exclusive in that both of them can be true.

The proof of Theorem 3.1 is given in section 5.1. It will reveal a more precise characterization of the Mallows d_2 limits. To give the detailed characterization we take as a starting point the formula (3.1). We then know that for every $j \in \mathbb{N}_0$, the sequence $\{\psi_{j;n}\}_{n\in\mathbb{N}_0}$ is bounded. Thus, by the Theorem of Bolzano and Weierstrass, there exists a subsequence $\{n_{k(j)}\}_{k(j)\in\mathbb{N}} \subseteq \mathbb{N}$, possibly depending on j, such that

$$\psi_{j,n_{k(j)}} \to \psi_j\ (k \to \infty),$$

for some ψ_j.
Now by a 'diagonal argument' we can find a 'universal' subsequence $\{n_k\}_{k\in\mathbb{N}} \subseteq \mathbb{N}$, not depending on j, such that

$$\psi_{j,n_k} \to \psi_j\ (k \to \infty),\ j \in \mathbb{N}_0.$$

In the sequel we assume without loss of generality that

$$\psi_{j;n} \to \psi_j \ (n \to \infty), \ j \in \mathbb{N}_0. \tag{3.2}$$

By Fatou's Lemma we know that

$$\lim_{n\to\infty} \sum_{j=0}^{\infty} \psi_{j;n}^2 \geq \sum_{j=0}^{\infty} \psi_j^2.$$

We will see that the characterization of the Mallows d_2 limits will depend on whether one has equality or not in the above expression and whether assumption (A1) and/or (A2) hold in Theorem 3.1.

The next result describes the possible Mallows d_2 limits, i.e., the processes $(X_t)_{t\in\mathbb{Z}}$ which arise as a limit with respect to the d_2 metric of sequences of MA processes as defined in (2.1).

Theorem 3.2 *Assume that the sequence of MA processes as defined in (2.1) converges in the d_2 metric and satisfies (A.MA) and (3.2). Denote by (A1) and (A2) the same conditions as in Theorem 3.1.*
Then:

(i) If (A1) holds and $\lim_{n\to\infty}\sum_{j=0}^{\infty}\psi_{j;n}^2 = \sum_{j=0}^{\infty}\psi_j^2$, the Mallows d_2 limit is in S_2.

(ii) If (A1) holds and $\lim_{n\to\infty}\sum_{j=0}^{\infty}\psi_{j;n}^2 > \sum_{j=0}^{\infty}\psi_j^2$, the Mallows d_2 limit is in $S_1 \oplus S_2$.

(iii) If (A2) holds, then the Mallows d_2 limit is in $S_1 \oplus S_3$.

(iv) If (A1) and (A2) hold, then the Mallows d_2 limit is in S_1.

Remark 3.2: For assertion (i) it is sufficient to assume only $\varepsilon_{t;n} \Rightarrow \varepsilon_t$, where $(\varepsilon_t)_{t\in\mathbb{Z}}$ is an i.i.d. sequence with $\mathbb{E}[\varepsilon_t] = 0$, and $\lim_{n\to\infty}\sum_{j=0}^{\infty}\psi_{j;n}^2 = \sum_{j=0}^{\infty}\psi_j^2$; these two assumptions imply that $\mathbb{E}|\varepsilon_t|^2 = 1$. This can be seen in the proof of assertion (i) in section 5.1, which yields under these assumptions the d_2 limit $X_t = \sum_{j=0}^{\infty}\psi_j\varepsilon_{t-j}$. Thus by formula (3.1) we conclude that $d_2^{(1)}(\varepsilon_{t;n}, \varepsilon_t) \to 0$ $(n \to \infty)$.
Remark 3.3: It is impossible to get a Mallows d_2 limit which is in $S_2 \oplus S_3$.

The proof of Theorem 3.2 is given in section 5.1. Example 1.1 describes a sequence of MA processes with a d_2 limit in S_3.

This example can also be modified such that sequences of MA processes converge in the variation metric to a d_V limit in S_3. This corresponds to the following Theorem, describing that we can never distinguish perfectly between stationary processes and some finite moving average processes.

Theorem 3.3 *The MA closure with respect to the variation metric d_V has the following features.*

(i) Let $(\xi_t)_{t\in\mathbb{Z}}$ be any stationary process. Assume that the distributions of $(\xi_1, \ldots \xi_m)$ have densities with respect to the Lebesgue measure for all $m \in \mathbb{N}$.
Then, there exists a process $(X_t)_{t\in\mathbb{Z}} \in S_3$, which is a d_V limit of a sequence of MA processes satisfying (A.MA) and (A2), such that

$$\mathbb{P}[X_t = \xi_t \text{ for all } t \in \mathbb{Z} | (\xi_t)_{t\in\mathbb{Z}}] \geq exp(-1) > 0.36 \text{ almost surely.}$$

(ii) *There exist ergodic, stationary processes whose finite dimensional distributions have densities with respect to the Lebesgue measure and which are not representable as d_V limits of a sequence of MA processes, satisfying (A.MA).*

The proof of Theorem 3.3 (i) is given in section 5.1. For proving assertion (ii) it is sufficient to give an example.

Example 3.1 Consider the stationary binary Markov chain $(X_t)_{t\in\mathbb{Z}}$, given $\mathbb{P}[X_1 = 0] = \mathbb{P}[X_1 = 1] = 1/2$, $\mathbb{P}[X_1 = 0|X_0 = 0] = \mathbb{P}[X_1 = 0|X_0 = 1] = \pi$, $0 < \pi < 1/2$. Then $(X_t)_{t\in\mathbb{Z}}$ is ergodic. Moreover, the probability distribution of X_t is not decomposable, since the convolution of two non-degenerate distributions would place mass on at least three points, whereas X_t is only binary. Hence, $(X_t)_{t\in\mathbb{Z}}$ can not be approximated in the d_V metric by any MA process, saying that $(X_t)_{t\in\mathbb{Z}}$ can not be an element of the MA closure.

Example 3.2 Consider the Gaussian AR(1) process

$$Y_t = \phi Y_{t-1} + \varepsilon_t \ (t \in \mathbb{Z}),$$

where $0 < |\phi| < 1$ and $(\varepsilon_t)_{t\in\mathbb{Z}}$ an i.i.d sequence, $\varepsilon_t \sim \mathcal{N}(0, 1-\phi^2)$. The process $(Y_t)_{t\in\mathbb{Z}}$ is stationary and strong-mixing, cf. Gorodetskii (1977).
Let $F(x) = \Phi(x) - x\varphi(x)$ be the c.d.f. with density $f(x) = x^2\varphi(x)$, where $\Phi(.)$ and $\varphi(.)$ denote the c.d.f. and density of the standard normal distribution. This distribution is indecomposable, cf. Linnik (1964, Ch. 5.2). Construct, the process

$$X_t = F^{-1} \circ \Phi(Y_t) \ (t \in \mathbb{Z}).$$

The one-dimensional marginal distribution is $X_t \sim F$, being indecomposable. Thus, as in Example 3.1, $(X_t)_{t\in\mathbb{Z}}$ can not be approximated in the d_V metric by any MA process. Moreover, $(X_t)_{t\in\mathbb{Z}}$ is a stationary, strong-mixing, and hence ergodic process whose finite dimensional marginal distributions have densities with respect to the Lebesgue measure, as required in Theorem 3.3 (ii).

There are probably many ergodic, stationary processes, which are not elements of the MA closure. A possible candidate would be the bilinear process, given by

$$X_t = -0.4X_{t-1} + 0.4X_{t-1}\varepsilon_{t-1} + \varepsilon_t \ (t \in \mathbb{Z}),$$

where $(\varepsilon_t)_{t\in\mathbb{Z}}$ i.i.d $\sim \mathcal{N}(0,1)$, cf. Subba Rao and Gabr (1984, Figure 3.10).
This process is stationary and ergodic, cf. Akamanam et al. (1986). It is also immediate that the process is non-Gaussian. As argued in Subba Rao and Gabr (1984, Table 3.2 and Figure 3.3), this bilinear process is not representable as a moving average process. However, the MA closure also contains the class S_3 and it seems difficult to prove rigorously that the process is not an element of this class S_3.

It is possible to find a sub-set of MA processes which is closed with respect to the d_2 and to the d_V metric. Let

$$\begin{aligned} S_{MA;g(.),h(.),K,\beta,C} = \Big\{ & (X_t)_{t\in\mathbb{Z}}; X_t = \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j},\ (\varepsilon_t)_{t\in\mathbb{Z}} \text{ an i.i.d. sequence}, \\ & \varepsilon_t \sim f(x)dx,\ f \in \mathcal{F}_{g(.),h(.),K} \text{ and } (\psi_j)_{j\in\mathbb{N}_0} \in \mathcal{M}_{\beta,C}\Big\}, \end{aligned}$$

where

$$\mathcal{F}_{g(.),h(.),K} = \{ \quad f;\, f \geq 0,\ \int_{-\infty}^{\infty} f(x)dx = 1,\ \int_{-\infty}^{\infty} xf(x)dx = 0,\ \int_{-\infty}^{\infty} x^2 f(x)dx \leq K,$$
$$\int_{-\infty}^{\infty} |f(x) - f(x+c)|dx < \delta \text{ for all } |c| \leq g(\delta),$$
$$(\int_{-\infty}^{-A} + \int_{A}^{\infty})|f(x)|dx \leq \delta \text{ for all } A \geq h(\delta)\ (\delta > 0)\},$$

with $g : \mathbb{R}^+ \to \mathbb{R}^+,\ g \nearrow,\ g(0) = 0,\ h : \mathbb{R}^+ \to \mathbb{R}^+,\ h \searrow,\ h(0) = \infty$ and

$$\mathcal{M}_{\beta,C} = \{(\psi_j)_{j \in \mathbb{N}_0};\ \sum_{j=0}^{\infty} j^{\beta} \psi_j^2 \leq C\}.$$

Theorem 3.4 *Assume that $\beta > 0$, $K < \infty$ and $C < \infty$. Then the set $S_{MA;g(.),h(.),K,\beta,C}$ is closed under the d_2 and under the d_V metric.*

Remark 3.4: Under the additional assumption that $\beta > 3$ and $\sum_{j=0}^{\infty} \psi_j z^j \neq 0$ for $|z| \leq 1$ for every $(\psi_j)_{j \in \mathbb{N}_0} \in \mathcal{M}_{\beta,C}$, the processes in $S_{MA;g(.),h(.),K,\beta,C}$ are strong mixing and hence ergodic. This follows from the result in Gorodetskii (1977).

The proof of Theorem 3.4 is given in section 5.1.

4 Closure for causal AR processes

We consider now sequences of stationary, causal AR processes, as defined in (2.2). We always assume

(A.AR): For every $n \in \mathbb{N}$, $(\varepsilon_{t;n})_{t \in \mathbb{Z}}$ is an i.i.d. sequence with $\mathbb{E}[\varepsilon_{t;n}] = 0, \mathbb{E}|\varepsilon_{t;n}|^2 < \infty$.

Causal, or minimum phase, means that the autoregressive transfer function

$$\Phi_n(z) = 1 - \sum_{j=1}^{\infty} \phi_{j;n} z^j \neq 0 \text{ for } |z| \leq 1\ (z \in \mathbb{C}) \text{ and } \sum_{j=1}^{\infty} |\phi_{j;n}| < \infty. \tag{4.1}$$

The formula (4.1) implies that $X_{t;n}$ can be expressed as a function of the present and past of the innovation process $\varepsilon_{t;n},\ \varepsilon_{t-1;n}, \ldots$, namely

$$X_{t;n} = \sum_{j=0}^{\infty} \psi_{j;n} \varepsilon_{t-j;n}\ (t \in \mathbb{Z}), \tag{4.2}$$

where the coefficients $(\psi_{j;n})_{j=0}^{\infty}$ are given by the inverse of the AR polynomial, the so-called MA transfer function

$$1/\Phi_n(z) = \Psi_n(z) = \sum_{j=0}^{\infty} \psi_{j;n} z^j,\ |z| \leq 1.$$

Thus by (4.2), we see that the stationary, causal AR processes given by (2.2), satisfying (4.1) can always be represented as stationary MA processes.

We focus first on the Mallows d_2 closure. The characterization of the d_2 closure for such AR processes can now be derived via the characterization of the MA closure. However, we cannot force the innovation variances to be equal to one by a simple standardization as in the MA case. Moreover, as described below, for sequences of AR processes which converge in the d_2 metric, the innovations are also forced to converge in the d_2 metric, compare also with assumption (A1) in Theorem 3.1.

Lemma 4.1 *Assume that the sequence of AR processes as defined in (2.2) satisfies (A.AR), converges in the d_2 metric and $\varepsilon_{t;n} \Rightarrow \varepsilon_t$ $(n \to \infty)$, where $(\varepsilon_t)_{t\in\mathbb{Z}}$ denotes an i.i.d. sequence with $\mathbf{E}[\varepsilon_t] = 0$, $\mathbf{E}|\varepsilon_t|^2 < \infty$.*
Then, also

$$\mathbf{E}|\varepsilon_{t;n}|^2 \to \mathbf{E}|\varepsilon_t|^2 \ (n \to \infty),$$

i.e., $d_2^{(1)}(\varepsilon_{t;n}, \varepsilon_t) \to 0$ $(n \to \infty)$.

Proof: Denote by $U_{t;n} = \sum_{j=1}^{\infty} \phi_{j;n} X_{t-j;n}$. By the weak convergence of $X_{t;n} \Rightarrow X_t$ $(n \to \infty)$ and of $\varepsilon_{t;n} \Rightarrow \varepsilon_t$ $(n \to \infty)$ we know that $U_{t;n} \Rightarrow U_t = X_t - \varepsilon_t$ $(n \to \infty)$, where U_t is independent from ε_t. Thus,

$$\mathbf{E}|X_t|^2 = \mathbf{E}|U_t|^2 + \mathbf{E}|\varepsilon_t|^2,$$

and therefore

$$\mathbf{E}|X_t|^2 = \liminf_{n\to\infty} \mathbf{E}|X_{t;n}|^2 \geq \liminf_{n\to\infty} \mathbf{E}|U_{t;n}|^2 + \liminf_{n\to\infty} \mathbf{E}|\varepsilon_{t;n}|^2 \geq \mathbf{E}|U_t|^2 + \mathbf{E}|\varepsilon_t|^2 = \mathbf{E}|X_t|^2,$$

where the inequality cannot be strict. Hence equality holds in the above expression which completes the proof. □

The question of interest is then if the AR closure is smaller than the MA closure. The answer is yes although type S_3 can still arise, as described by the next theorem.

Theorem 4.1 *Assume that the sequence of AR processes as defined in (2.2) converges in the d_2 metric, satisfies (A.AR), (4.1) and one of the following:*

(A3): $\varepsilon_{t;n} \Rightarrow \varepsilon_t$ $(n \to \infty)$, where $(\varepsilon_t)_{t\in\mathbb{Z}}$ is an i.i.d. sequence with $\mathbf{E}[\varepsilon_t] = 0$.

(A4): $\sup_{n\in\mathbb{N}} \sup_{j\geq 0} |\psi_{j;n}| < \infty$, where $\sum_{j=0}^{\infty} \psi_{j;n} z^j = 1/\Phi_n(z)$ $(|z| \leq 1)$.

Then, the d_2 limit of such a sequence of AR processes is in $\{S_1 \oplus S_2\} \cup \{S_1 \oplus S_3\}$.

The proof of Theorem 4.1 is given in section 5.2. Also, the precise characterization of the Mallows d_2 limits is similar to the one in Theorem 3.2. We know from Lemma 4.1 that $\varepsilon_{t;n} \Rightarrow \varepsilon_t$ $(n \to \infty)$ implies $\sigma_n^2 = \mathbf{E}|\varepsilon_{t;n}|^2 \to \sigma^2 = \mathbf{E}|\varepsilon_t|^2 < \infty$ $(n \to \infty)$. We distinguish between two cases,
(I) $0 < \sigma^2 < \infty$,
(II) $\sigma^2 = 0$, i.e., the degenerate case $\varepsilon_t \equiv 0$ for all $t \in \mathbb{Z}$.
By assumption of Mallows d_2 convergence we know that

$$\mathbf{E}|X_{t;n}|^2 = \sum_{j=0}^{\infty} \psi_{j;n}^2 \sigma_n^2 \to c^2 \ (n \to \infty), \ 0 < c^2 < \infty.$$

(The case $c^2 = 0$ is degenerate and uninteresting, the d_2 limit $X_t \equiv 0$ is in S_2 with $\psi_j = 0$ for all j).
Thus, we consider the two cases
(I) $\sup_{n\in\mathbb{N}} \sum_{j=0}^{\infty} \psi_{j;n}^2 < \infty$,
(II) $\sup_{n\in\mathbb{N}} \sum_{j=0}^{\infty} \psi_{j;n}^2 = \infty$.
In case (I), we then know that for every $j \in \mathbb{N}_0$, the sequence $\{\psi_{j;n}\}_{n\in\mathbb{N}}$ is bounded. Thus, by the same argument as for formula (3.2),

$$\psi_{j;n} \to \psi_j \ (n \to \infty), \ j \in \mathbb{N}_0.$$

And again by Fatou's Lemma we know that

$$\lim_{n\to\infty} \sum_{j=0}^{\infty} \psi_{j;n}^2 \geq \sum_{j=0}^{\infty} \psi_j^2.$$

In case (II) the sum $\sum_{j=0}^{\infty} \psi_{j;n}^2$ is unbounded and no direct analysis as in case (I) applies.

Theorem 4.2 *Assume that the sequence of AR processes as defined in (2.2) converges in the d_2 metric, satisfies (A.AR) and (4.1). Denote by (A3) and (A4) the same conditions as in Theorem 4.1.*
Then:

(i) If (A3) holds, $\lim_{n\to\infty} \mathbb{E}|\varepsilon_{t;n}|^2 = \sigma^2 > 0$ and $\lim_{n\to\infty} \sum_{j=0}^{\infty} \psi_{j;n}^2 = \sum_{j=0}^{\infty} \psi_j^2$, the Mallows d_2 limit is in S_2.

(ii) If (A3) holds, $\lim_{n\to\infty} \mathbb{E}|\varepsilon_{t;n}|^2 = \sigma^2 > 0$ and $\lim_{n\to\infty} \sum_{j=0}^{\infty} \psi_{j;n}^2 > \sum_{j=0}^{\infty} \psi_j^2$, the Mallows d_2 limit is in $S_1 \oplus S_2$.

(iii) If (A4) holds, $\varepsilon_{t;n} = \sigma_n Z_t$ $(t \in \mathbb{Z})$ with $(Z_t)_{t\in\mathbb{Z}}$ i.i.d., $\mathbb{E}[Z_t] = 0$, $E|Z_t|^2 = 1$ and $\lim_{n\to\infty} \sigma_n = 0$, $\sigma_n \geq 0$, the Mallows d_2 limit is in S_1.

(iv) If (A4) holds and $\lim_{n\to\infty} \mathbb{E}|\varepsilon_{t;n}|^2 = 0$, the Mallows d_2 limit is in $S_1 \oplus S_3$.

The proof of Theorem 4.2 is given in section 5.2. To show that the statement (iv) does include non-zero elements of S_3 processes, we now give an example.

Example 4.1 Consider the sequence of AR(1) processes

$$X_{t;n} = \phi_n X_{t-1;n} + \varepsilon_{t;n} \ (t \in \mathbb{Z}),$$

where $0 < \phi_n < 1$, $\phi_n \to 1$ $(n \to \infty)$ and $(\varepsilon_{t;n})_{t\in\mathbb{Z}}$ i.i.d. with $\mathbb{P}[\varepsilon_{t;n} = 0] = \phi_n^2$, $\mathbb{P}[\varepsilon_{t;n} = \pm 1] = (1 - \phi_n^2)/2$.

Proposition 4.1 *In Example 4.1, $(X_{t;n})_{t\in\mathbb{Z}}$ converges in the d_2 metric to a process $(X_t)_{t\in\mathbb{Z}} \in S_3$ which has constant sample paths, i.e., $X_t = X_s$ for all $t, s \in \mathbb{Z}$.*

The proof of Proposition 4.1 is given in section 5.2. It is a difficult task to construct sequences of AR processes with d_2 limits in S_3 having non-constant sample paths. It is an open question to us if more complicated S_3 processes arise as d_2 limits of AR processes satisfying (A.AR) and (4.1).

In case (I), the condition $\sigma^2 > 0$ can be interpreted that the innovation variance $\mathbb{E}|\varepsilon_{t;n}|^2$ is of the same order as the process variance $\mathbb{E}|X_{t;n}|^2$. Then, only assertions (i) and (ii) of Theorem 4.2 apply and type S_3 does not arise. In that respect the AR closure is smaller and easier to understand than the MA closure.

Under additional assumptions we also can sharpen Theorem 4.1 for the case with non-vanishing innovation variance as $n \to \infty$. Denote by

$$\begin{aligned} S_{AR;g(.),h(.),K,\beta,C} = \{ & (X_t)_{t\in\mathbb{Z}}; X_t = \sum_{j=1}^{\infty} \phi_j X_{t-j} + \varepsilon_t,\ (\varepsilon_t)_{t\in\mathbb{Z}} \text{ an i.i.d. sequence,} \\ & \varepsilon_t \sim f(x)dx,\ f \in \mathcal{F}_{g(.),h(.),K} \text{ and } (\phi_j)_{j\in\mathbb{N}} \in \mathcal{A}_{\beta,C}\}, \end{aligned}$$

where $\mathcal{F}_{g(.),h(.),K}$ is defined as for $S_{MA;g(.),h(.),K,\beta,C}$ in section 3 and

$$\mathcal{A}_{\beta,C} = \{(\phi_j)_{j\in\mathbb{N}};\ \sum_{j=1}^{\infty} j^{\beta}|\phi_j| \le C \text{ and } 1 - \sum_{j=1}^{\infty} \phi_j z^j \ne 0 \text{ for } |z| \le 1\}.$$

Theorem 4.3 *Assume that $\beta > 1$, $K < \infty$ and $C < \infty$.*
Then, the set $S_{AR;g(.),h(.),K,\beta,C}$ is closed under the d_2 and under the d_V metric. Moreover, processes in $S_{AR;g(.),h(.),K,\beta,C}$ are strong-mixing and hence ergodic.

The proof of Theorem 4.3 is given in section 5.2.

One can also ask about the closure for ARMA processes. Most elegantly we represent sequences of ARMA processes of order (∞, ∞) by

$$\Phi_n(B)X_n = \Theta_n(B)\varepsilon_n,$$

where $\Phi_n(z) = 1 - \sum_{j=1}^{\infty} \phi_{j;n} z^j$, $\Theta_n(z) = \sum_{j=0}^{\infty} \theta_{j;n} z^j$ ($|z| \le 1$), B the back-shift operator, $X_n = (X_{t;n})_{t\in\mathbb{Z}}$ and $\varepsilon_n = (\varepsilon_{t;n})_{t\in\mathbb{Z}}$ an i.i.d. sequence with $\mathbb{E}[\varepsilon_{t;n}] = 0$, $\mathbb{E}|\varepsilon_{t;n}|^2 < \infty$.
If (4.1) holds for the autoregressive transfer function Φ_n and $\sum_{j=0}^{\infty} |\theta_{j;n}| < \infty$, we can represent such ARMA processes as $X_{t;n} = \sum_{j=0}^{\infty} \psi_{j;n}\varepsilon_{t-j;n}$, where $\Psi_n(z) = \sum_{j=0}^{\infty} \psi_{j;n} z^j = \Theta_n(z)/\Phi_n(z)$.
Then, under either the condition (A3) for the innovations $\{(\varepsilon_{t;n})_{t\in\mathbb{Z}}\}_{n\in\mathbb{N}}$, requiring additionally that $\lim_{n\to\infty} d_2^{(1)}(\varepsilon_{t;n}, \varepsilon_t) = 0$, or under condition (A4) for the variables $(\psi_{j;n})_{j\in\mathbb{N}_0}$ as defined above, the d_2 limits of such sequences of ARMA processes are in $\{S_1 \oplus S_2\} \cup \{S_1 \oplus S_3\}$.

5 Proofs

Lemma 5.1 *Let $(\xi_t)_{t\in\mathbb{Z}}$ be a stationary process with $\mathbb{E}[\xi_t] = 0$.*

(i) There exists a sequence of stationary, ergodic processes $\{(\xi_t^{(r)})_{t\in\mathbb{Z}}\}_{r\in\mathbb{N}}$ with $\mathbb{E}[\xi_t^{(r)}] = 0$ for every $r \in \mathbb{N}$, such that

$$\lim_{r\to\infty} d((\xi_t^{(r)})_{t\in\mathbb{Z}}, (\xi_t)_{t\in\mathbb{Z}}) = 0,\ d = d_2 \text{ or } d_V.$$

(ii) If the m-dimensional distributions of $(\xi_1, \dots, \xi_m)$ have densities with respect to the Lebesgue measure for all $m \in \mathbb{N}$, there exists a sequence of stationary, ergodic processes $\{(\xi_t^{(r)})_{t\in\mathbb{Z}}\}_{r\in\mathbb{N}}$ with m-dimensional distributions of $(\xi_1^{(r)}, \dots, \xi_m^{(r)})$ having densities with respect to the Lebesgue measure for all $m \in \mathbb{N}$ and $\mathbb{E}[\xi_t^{(r)}] = 0$ for every $r \in \mathbb{N}$, such that

$$\lim_{r\to\infty} d((\xi_t^{(r)})_{t\in\mathbb{Z}}, (\xi_t)_{t\in\mathbb{Z}}) = 0, \ d = d_2 \text{ or } d_V.$$

(iii) If the process $(\xi_t)_{t\in\mathbb{Z}}$ is Gaussian, there exists a sequence of stationary, ergodic Gaussian processes $\{(\xi_t^{(r)})_{t\in\mathbb{Z}}\}_{r\in\mathbb{N}}$, such that

$$\lim_{r\to\infty} d((\xi_t^{(r)})_{t\in\mathbb{Z}}, (\xi_t)_{t\in\mathbb{Z}}) = 0, \ d = d_2 \text{ or } d_V.$$

Proof: The statements (i) and (ii) follow by the technique of Durrett (1991, Ch. 6, Ex. 1.9 and Ex. 2.10).

Statement (iii) is rather standard: the spectral measure $F(.)$ of the Gaussian process $(\xi_t)_{t\in\mathbb{Z}}$ can be approximated by a sequence of spectral measures $\{F^{(r)}(.)\}_{r\in\mathbb{N}}$ (in the sense of complete convergence), having continuous spectral densities $\{f^{(r)}(.)\}_{r\in\mathbb{N}}$. These densities $f^{(r)}(.)$ can be approximated by polynomials $P^{(k_r)}(.)$ of order k_r, corresponding to Gaussian MA processes of order k_r, being now the approximating sequences $\{(\xi_t^{(k_r)})_{t\in\mathbb{Z}}\}_{r\in\mathbb{N}}$. For details, see Grenander and Szegö (1984, Ch. 1.9 and Ch. 1.12). □

5.1 Proofs of results in section 3

Proof of Theorem 3.1 (ii)

Step 1: Proof that $S_1 \oplus S_2$ is realizable with d_2 limits of MA sequences.
Let $(W_t \oplus Y_t)_{t\in\mathbb{Z}} \in S_1 \oplus S_2$ with $(W_t)_{t\in\mathbb{Z}} \in S_1$, $(Y_t)_{t\in\mathbb{Z}} \in S_2$. By Lemma 5.1 (iii) there exists a sequence of stationary, ergodic Gaussian processes $\{(W_{t;r})_{t\in\mathbb{Z}}\}_{r\in\mathbb{N}}$, such that

$$\lim_{r\to\infty} d_2((W_{t;r})_{t\in\mathbb{Z}}, (W_t)_{t\in\mathbb{Z}}) = 0. \tag{5.1}$$

We show now that $(W_{t;r} \oplus Y_t)_{t\in\mathbb{Z}}$ can be approximated by a sequence of MA processes. Assume that $Y_t = \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j}$ with $\mathbb{E}|\varepsilon_t|^2 = 1$. Let $M_n \to \infty$ $(n \to \infty)$ and define

$$X_{t;n,r} = \sum_{j=0}^{M_n} \psi_j \varepsilon_{t-j} + \sum_{j=M_n+1}^{\infty} \psi_{j,n;r} \varepsilon_{t-j}, \tag{5.2}$$

where $\psi_{j,n;r} = M_n^{-1/2} w_{j;r}$ $(M_n + 1 \le j \le 2M_n)$ and $\psi_{j,n;r} = 0$ $(j \ge 2M_n + 1)$, and $(w_{t;r})_{t\in\mathbb{Z}}$ is a fixed realization of the process $(W_{t;r})_{t\in\mathbb{Z}}$.
Clearly, for $\delta > 0$ there exists an $n_1 = n_1(\delta)$ such that

$$d_2((\sum_{j=0}^{M_n} \psi_j \varepsilon_{t-j})_{t\in\mathbb{Z}}, (Y_t)_{t\in\mathbb{Z}}) \le \delta \text{ for all } n \ge n_1. \tag{5.3}$$

Moreover, for $\delta > 0$ there exists an $n_{2,r} = n_2(r,\delta)$ such that

$$d_2((\sum_{j=M_n+c+1}^{\infty} \psi_{j,n;r}\varepsilon_{t-j})_{t\in\mathbb{Z}}, (W_{t;r})_{t\in\mathbb{Z}}) \le \delta \text{ for all } n \ge n_{2,r} \ (c \in \mathbb{N}_0). \tag{5.4}$$

(The constant $n_{2,r}$ might depend also on c).
To show (5.4) we first verify the Lindeberg condition; we first argue for the one-dimensional marginal distribution. Note that for $\gamma > 0$ there exists an $n_{3,r} = n_3(r,\gamma)$ such that

$$\max_{j\ge M_n+1} |\psi_{j,n;r}| \le \gamma \text{ for all } n \ge n_{3,r},$$

since $(w_{t;r})_{t\in\mathbb{Z}}$ is a realization of the stationary, ergodic Gaussian process $(W_{t;r})_{t\in\mathbb{Z}}$ and $\mathbb{P}[\max_{M_n+1\le j\le 2M_n} |W_{j;r}| > M_n^{1/4}] = o(1)$ $(n \to \infty)$, by the behavior of the Gaussian tail. Hence, for $\zeta > 0$, $\kappa > 0$ there exists an $n_{4,r} = n_4(r,\zeta,\kappa)$ such that

$$\sum_{j=M_n+1}^{\infty} \psi_{j,n;r}^2 \mathbb{E}[\varepsilon_t^2 1_{[|\varepsilon_t|>\kappa/|\psi_{j,n;r}|]}] < \zeta \text{ for all } n \ge n_{4,r}. \tag{5.5}$$

Here we used that $\sum_{j=M_n+1}^{\infty} \psi_{j,n;r}^2$ is bounded, see argumentation for (5.6).
Next, we verify that for $\zeta > 0$ there exists $n_{5,r} = n_5(r,\zeta)$ such that

$$|\sum_{j=M_n+c+1}^{\infty} \psi_{j,n;r}\psi_{j+k,n;r} - Cov(W_{0;r}, W_{k;r})| \le \zeta \text{ for all } n \ge n_{5,r} \ (k \in \mathbb{N}_0). \tag{5.6}$$

(The constant $n_{5,r}$ might depend on c and k). This states the convergence to the proper covariances.
Formula (5.6) follows immediately by the definition of $\psi_{j,n;r}$ and the ergodicity of $(W_{t;r})_{t\in\mathbb{Z}}$.
Thus, by (5.5) and (5.6) we have shown (5.4) for the one-dimensional marginals. The more general statement in (5.4) follows analogously by the Cramér-Wold device.
For any finite dimensional set $t_1 < \ldots < t_m$, $m \in \mathbb{N}$, we choose $c = t_m - t_1$ such that $\sum_{j=0}^{M_n} \psi_j\varepsilon_{t_i-j}$ and $\sum_{j=M_n+t_m-t_1+1}^{\infty} \psi_{j,n;r}\varepsilon_{t_k-j}$ are independent for all $i,k \in \{1,\ldots,m\}$. Moreover, $Var(\sum_{j=M_n+1}^{M_n+t_m-t_1} \psi_{j,n;r}\varepsilon_{t-j}) = o_P(1)$ $(n \to \infty)$, where the o_P-term depends on r. By splitting up

$$X_{t;n,r} = \sum_{j=0}^{M_n} \psi_j\varepsilon_{t-j} + \sum_{j=M_n+1}^{M_n+t_m-t_1} \psi_{j,n;r}\varepsilon_{t-j} + \sum_{M_n+t_m-t_1+1}^{\infty} \psi_{j,n;r}\varepsilon_{t-j}, \tag{5.7}$$

we see that the middle part $\sum_{j=M_n+1}^{M_n+t_m-t_1} \psi_{j,n;r}\varepsilon_{t-j}$ plays a negligible role as $n \to \infty$ and we can work with the remaining independent pieces. Using this independence and (5.3) and (5.4) we have: for $\delta > 0$ there exists an $n_{6,r} = n_6(r,\delta)$ such that

$$d_2((X_{t;n,r})_{t\in\mathbb{Z}}, (W_{t;r} + Y_t)_{t\in\mathbb{Z}}) \le \delta \text{ for all } n \ge n_{6,r},$$

with $(W_{t;r})_{t\in\mathbb{Z}}$ and $(Y_t)_{t\in\mathbb{Z}}$ independent for every $r \in \mathbb{N}$.
Restating,

$$\lim_{n\to\infty} d_2((X_{t;n,r})_{t\in\mathbb{Z}}, (W_{t;r} \oplus Y_t)_{t\in\mathbb{Z}}) = 0 \ (r \in \mathbb{N}). \tag{5.8}$$

Thus by (5.1) and (5.8), there exists a subsequence $\{n_r\}_{r\in\mathbb{N}} \subseteq \mathbb{N}$ such that

$$\lim_{n\to\infty} d_2((X_{t;n_r,r})_{t\in\mathbb{Z}}, (W_t \oplus Y_t)_{t\in\mathbb{Z}}) = 0.$$

Since $(X_{t;n_r,r})_{t\in\mathbb{Z}}$ is an MA process, satisfying (A.MA) and (A1) we complete the proof of step 1.

Step 2: Proof that $S_1 \oplus S_3$ is realizable with d_2 limits of MA processes.
Let $(X_t)_{t\in\mathbb{Z}} = (W_t \oplus Y_t)_{t\in\mathbb{Z}} \in S_1 \oplus S_3$ with $(W_t)_{t\in\mathbb{Z}} \in S_1$ and $(Y_t = \sum_{i=1}^{\infty}\sum_{j=1}^{N_i} \xi_{t;i,j})_{t\in\mathbb{Z}} \in S_3$. Approximate $(W_t)_{t\in\mathbb{Z}}$ by $(W_{t;r})_{t\in\mathbb{Z}}$ as in (5.1). Moreover, approximate $(Y_t)_{t\in\mathbb{Z}}$ by

$$Y_{t;k,u} = \sum_{i=1}^{k}\sum_{j=1}^{N_i} \xi_{t;i,j}^{(u)},\ \{(\xi_{t;i,j}^{(u)})_{t\in\mathbb{Z}}\}_{u\in\mathbb{N}} \text{ stationary, ergodic as in Lemma 5.1(i) } (k, u \in \mathbb{N}),$$
$$\lim_{u\to\infty} d_2((\xi_{t;i,j}^{(u)})_{t\in\mathbb{Z}}, (\xi_{t;i,j})_{t\in\mathbb{Z}}) = 0\ (i, j \in \mathbb{N}).$$

Then, by using $\sum_{i=1}^{\infty} \lambda_i \sigma_{\xi;i}^2 < \infty$ we get in a straightforward way,

$$\lim_{u\to\infty}\lim_{k\to\infty} d_2((Y_{t;k,u})_{t\in\mathbb{Z}}, (Y_t)_{t\in\mathbb{Z}}) = 0. \tag{5.9}$$

Denote by $(X_{t;r,k,u})_{t\in\mathbb{Z}} = (W_{t;r} + Y_{t;k,u})_{t\in\mathbb{Z}}$. Then, by (5.1) and (5.9),

$$\lim_{u\to\infty}\lim_{k\to\infty}\lim_{r\to\infty} d_2((X_{t;r,k,u})_{t\in\mathbb{Z}}, (X_t)_{t\in\mathbb{Z}}) = 0. \tag{5.10}$$

We show now that $(X_{t;r,k,u})_{t\in\mathbb{Z}}$ can be approximated by

$$X_{t;r,k,u}^{(n)} = \sum_{j=1}^{n} \bar{\xi}_{j;r,k,u} U_{t-j;n}\ (t \in \mathbb{Z},\ r, k, u \in \mathbb{N}), \tag{5.11}$$

where for every $n \in \mathbb{N}$, $(U_{t;n})_{t\in\mathbb{Z}}$ is an i.i.d. sequence, independent from $(\xi_{t;i,j}^{(u)})_{t\in\mathbb{Z}}$ for all $i, j, u = 1, 2, \ldots$ with $\mathbb{P}[U_{t;n} = 1] = 1 - \mathbb{P}[U_{t;n} = 0] = \lambda/n$ with $\lambda = 1 + \sum_{i=1}^{k} \lambda_i$ (if $(W_{t;r})_{t\in\mathbb{Z}} \equiv 0$, then $\lambda = \sum_{i=1}^{k} \lambda_i$), and $(\bar{\xi}_{j;r,k,u})_{j\in\mathbb{Z}}$ is a (fixed) realization of a process with distribution $F_{\bar{\xi}}$ on $(\mathbb{R}^{\mathbb{Z}}, \mathcal{B})$, given by $dF_{\bar{\xi}}(x) = \lambda^{-1}(dF_{W_r}(x) + \sum_{i=1}^{k} \lambda_i dF_i^{(u)}(x))$, $x \in \mathbb{R}^{\mathbb{Z}}$, where F_{W_r} and $F_i^{(u)}$ are the distributions of $(W_{t;r})_{t\in\mathbb{Z}}$ and $(\xi_{t;i,1}^{(u)})_{t\in\mathbb{Z}}$, respectively.
Note that the random quantities in the definition of $X_{t;r,k,u}^{(n)}$ are now only the $U_{t;n}$'s.
Then,

$$\begin{aligned}&(X_{t_1;r,k,u}^{(n)}, \ldots, X_{t_m;r,k,u}^{(n)}) \Rightarrow (X_{t_1;r,k,u}, \ldots, X_{t_m;r,k,u})\ (n \to \infty),\\ &\text{for all } t_1 < \ldots < t_m \in \mathbb{Z}, \text{ for all } m \in \mathbb{N}.\end{aligned} \tag{5.12}$$

To show (5.12) we first argue for the case $m = 1$. Denote by $\varphi_U(s) = \mathbb{E}[exp(isU_{1;n})]$, $s \in \mathbb{R}$. Then, we obtain for any $s \in \mathbb{R}$,

$$\begin{aligned}\varphi^{(n)}(s) &= \mathbb{E}[exp(isX_{t;r,k,u}^{(n)})] = \prod_{j=1}^{n} \varphi_U(\bar{\xi}_{j;r,k,u}s) = \prod_{j=1}^{n} (1 + \lambda/n(exp(is\bar{\xi}_{j;r,k,u}) - 1))\\ &\to exp\,(\lambda(\mathbb{E}[exp(is\bar{\xi}_{1;r,k,u})] - 1)) = \mathbb{E}[exp(isX_{t;r,k,u})]\ (n \to \infty).\end{aligned}$$

For the convergence to the limit with respect to n we use the ergodicity of $(\bar{\xi}_{j;r,k,u})_{j\in\mathbb{Z}}$ which implies $n^{-1}\sum_{j=1}^{n} exp(is\bar{\xi}_{j;r,k,u}) \to \mathbb{E}[exp(is\bar{\xi}_{1;r,k,u})]$ $(n \to \infty)$; note that $\bar{\xi}_{1;r,k,u}$ is now again a random variable. Moreover, the definition of $F_{\bar{\xi}}$ has justified the last equality in the formula above. This proves (5.12) for $m = 1$.
For $m > 1$ one argues similarly, for example for $m = 2$:

$$(X^{(n)}_{t_1;r,k,u}, X^{(n)}_{t_2;r,k,u})' = \sum_{j=1}^{n-1} (\bar{\xi}_{j;r,k,u}, \bar{\xi}_{j+t_2-t_1;r,k,u})' U_{t_1-j;n} + o_P(1)\ (n \to \infty)\ (t_1 < t_2),$$

and one uses now characteristic functions in $\mathbb{R}^2$.
On the other hand,

$$\begin{aligned} \mathbb{E}|X^{(n)}_{t;r,k,u}|^2 &= \mathbb{E}|U_{1;n}|^2 \sum_{j=1}^{n} \bar{\xi}^2_{j;r,k,u} = \lambda n^{-1} \sum_{j=1}^{n} \bar{\xi}^2_{j;r,k,u} \\ &\to \lambda \mathbb{E}|\bar{\xi}_{1;r,k,u}|^2 = \mathbb{E}|X_{t;r,k,u}|^2\ (n \to \infty). \end{aligned} \tag{5.13}$$

Therefore, by (5.12) and (5.13),

$$\lim_{n\to\infty} d_2((X^{(n)}_{t;r,k,u})_{t\in\mathbb{Z}}, (X_t)_{t\in\mathbb{Z}}) = 0. \tag{5.14}$$

Note that $\mathbb{E}[X^{(n)}_{t;r,k,u}] \neq 0$, but the formula (5.14) also holds for the mean corrected process $\tilde{X}^{(n)}_{t;r,k,u} = \sum_{j=1}^{n} \bar{\xi}_{j;r,k,u}(U_{t-j;n} - \lambda/n)$. Then, $(\tilde{X}^{(n)}_{t;r,k,u})_{t\in\mathbb{Z}}$ is an MA process of order n, satisfying (A.MA) and (A2), since $\mathbb{P}[\max_{1\le j\le n} |\bar{\xi}_{j;r,k,u}| \ge n^{3/4}] \to 0$ by the tail behavior, i.e., second moments of $\bar{\xi}_{j;r,k,u}$.
By (5.10) and (5.14),

$$\lim_{n\to\infty} \lim_{u\to\infty} \lim_{k\to\infty} \lim_{r\to\infty} d_2((\tilde{X}^{(n)}_{t;r,k,u})_{t\in\mathbb{Z}}, (X_t)_{t\in\mathbb{Z}}) = 0.$$

Thus, there exist subsequences $\{n_r\}_{r\in\mathbb{N}} \subseteq \mathbb{N}$, $\{u_r\}_{r\in\mathbb{N}} \subseteq \mathbb{N}$, $\{k_r\}_{r\in\mathbb{N}} \subseteq \mathbb{N}$ such that

$$\lim_{r\to\infty} d_2((\tilde{X}^{(n_r)}_{t;r,k_r,u_r})_{t\in\mathbb{Z}}, (X_t)_{t\in\mathbb{Z}}) = 0.$$

Therefore, $\{(\tilde{X}^{n_r}_{t;r,k_r,u_r})_{t\in\mathbb{Z}}\}_{r\in\mathbb{N}}$ serves as an approximating sequence of MA processes, which completes the proof for step 2.
By steps 1-2, the proof of Theorem 3.1 (ii) is complete.

Clearly, Theorem 3.2 implies Theorem 3.1 (i). □

Proof of Theorem 3.2

Proof of assertion (i):
We first argue for the one dimensional marginal $X_{t;n} = \sum_{j=0}^{\infty} \psi_{j;n}\varepsilon_{t-j;n}$ for a fixed $t \in \mathbb{Z}$. Let $\delta > 0$. Then there exists $k = k(\delta)$ such that

$$\sum_{j=k+1}^{\infty} \psi_j^2 \le \delta^2. \tag{5.15}$$

By (3.2) and $d_2^{(1)}(\varepsilon_{t;n}, \varepsilon_t) \to 0$ $(n \to \infty)$, there exists an $n_1 = n_1(\delta)$ such that

$$\begin{aligned} d_2(\sum_{j=0}^{k} \psi_{j;n}\varepsilon_{t-j;n}, \sum_{j=0}^{k} \psi_j \varepsilon_{t-j}) &\leq (\mathbb{E}|\sum_{j=0}^{k}(\psi_{j;n}\varepsilon_{t-j;n} - \psi_j\varepsilon_{t-j})|^2)^{1/2} \\ \leq \quad (\sum_{j=0}^{k}(\psi_{j;n} - \psi_j)^2)^{1/2} &+ (\sum_{j=0}^{k} \psi_j^2)^{1/2}(\mathbb{E}|\varepsilon_{t;n} - \varepsilon_t|^2)^{1/2} \leq \delta \text{ for } n \geq n_1. \end{aligned} \tag{5.16}$$

(Here we have used the Minkowski and Cauchy-Schwarz inequality).
Furthermore, again by the Cauchy-Schwarz inequality,

$$\begin{aligned} d_2(\sum_{j=0}^{k} \psi_j\varepsilon_{t-j}, \sum_{j=0}^{\infty} \psi_j\varepsilon_{t-j}) &\leq (\sum_{j=k+1}^{\infty} \psi_j^2)^{1/2} \leq \delta, \\ d_2(\sum_{j=0}^{k} \psi_{j;n}\varepsilon_{t-j;n}, \sum_{j=0}^{\infty} \psi_{j;n}\varepsilon_{t-j;n}) &\leq (\sum_{j=k+1}^{\infty} \psi_{j;n}^2)^{1/2} \end{aligned}$$

For the second inequality we bound in addition, by using $\sum_{j=0}^{\infty}(\psi_{j;n}^2 - \psi_j^2) \to 0$ $(n \to \infty)$: there exists a $n_2 = n_2(\delta)$ such that

$$\sum_{j=k+1}^{\infty} \psi_{j;n}^2 \leq \sum_{j=k+1}^{\infty} \psi_j^2 + |\sum_{j=k+1}^{\infty}(\psi_{j;n}^2 - \psi_j^2)| \leq (2\delta)^2 \text{ for all } n \geq n_2.$$

Thus,

$$d_2(\sum_{j=0}^{k} \psi_{j;n}\varepsilon_{t-j;n}, \sum_{j=0}^{\infty} \psi_{j;n}\varepsilon_{t-j;n}) \leq 2\delta \text{ for all } n \geq n_2.$$

By setting $n_0 = n_0(\delta) = max(n_1, n_2)$ we therefore obtain with (5.16) and its subsequent formulas,

$$d_2(\sum_{j=0}^{\infty} \psi_{j;n}\varepsilon_{t-j;n}, \sum_{j=0}^{\infty} \psi_j\varepsilon_{t-j}) \leq 4\delta \text{ for all } n \geq n_0,$$

implying

$$\lim_{n\to\infty} d_2^{(1)}(X_{t;n}, \sum_{j=0}^{\infty} \psi_j\varepsilon_{t-j}) = 0 \ (t \in \mathbb{Z}).$$

This proves assertion (i) for the one dimensional marginal case. The multidimensional case $(X_{t_1;n}, \ldots, X_{t_m;n})$ with $m > 1$ follows by the same argument and the Cramér-Wold device and hence

$$\lim_{n\to\infty} d_2((X_{t;n})_{t\in\mathbb{Z}}, (\sum_{j=0}^{\infty} \psi_j\varepsilon_{t-j})_{t\in\mathbb{Z}}) = 0.$$

Proof of assertion (ii):
We argue first for the one dimensional marginal case. Let $\{\delta_m\}_{m\in\mathbb{N}}$ be a monotonely

decreasing sequence with $\lim_{m\to\infty}\delta_m = 0$. Let $k_m = k(\delta_m)$ be as in (5.15). Then, there exists $n_{1,m} = n_1(\delta_m)$ and a constant $M_m \ge k_m$ such that

$$\max_{j\ge M_m+1} |\psi_{j;n}| \le \delta_m \text{ for } n \ge n_{1,m}. \tag{5.17}$$

To show formula (5.17) we assume the opposite. There exists $\delta_m > 0$ such that for all $n_{1,m}\in\mathbb{N}$ and for all $M_m \ge k_m = k(\delta_m)$

$$\max_{j\ge M_m+1} |\psi_{j;n}| > \delta_m \text{ for some } n \ge n_{1,m}.$$

This implies that for some $n \ge n_{1,m}$,

$$|\psi_{j;n}| > \delta_m \text{ for infinitely many } j \ge k_m + 1.$$

But this implies $\sum_{j=0}^{\infty}\psi_{j;n}^2 = \infty$ for some $n \ge n_{1,m}$ (for all $n_{1,m}\in\mathbb{N}$), which is a contradiction to $\lim_{n\to\infty}\mathbf{E}|X_{t;n}|^2 = \lim_{n\to\infty}\sum_{j=0}^{\infty}\psi_{j;n}^2 < \infty$.
As in (5.16) and its subsequent formula, there exists an $n_{2,m} = n_2(M_m,\delta_m)$ with

$$\begin{aligned} &d_2\Big(\sum_{j=0}^{M_m}\psi_{j;n}\varepsilon_{t-j;n}, \sum_{j=0}^{M_m}\psi_j\varepsilon_{t-j}\Big) \le \delta_m \text{ for all } n \ge n_{2,m} \\ &d_2\Big(\sum_{j=0}^{M_m}\psi_j\varepsilon_{t-j}, \sum_{j=0}^{\infty}\psi_j\varepsilon_{t-j}\Big) \le \delta_m. \end{aligned}$$

Therefore,

$$d_2\Big(\sum_{j=0}^{M_m}\psi_{j;n}\varepsilon_{t-j;n}, \sum_{j=0}^{\infty}\psi_j\varepsilon_{t-j}\Big) \le 2\delta_m \text{ for all } n \ge n_{2,m}.$$

For all $\delta > 0$ there exists an $m_1 = m_1(\delta)$ with $\delta_{m_1} \le \delta/2$ and hence

$$d_2\Big(\sum_{j=0}^{M_m}\psi_{j;n}\varepsilon_{t-j;n}, \sum_{j=0}^{\infty}\psi_j\varepsilon_{t-j}\Big) \le \delta, \text{ for all } m \ge m_1,\ n \ge n_{2,m}, \tag{5.18}$$

Let us consider now the tail part $\sum_{j=M_m+1}^{\infty}\psi_{j;n}\varepsilon_{t-j;n}$. Denote by $(W_t)_{t\in\mathbb{Z}}$ a stationary Gaussian process with $\mathbf{E}[W_t] = 0$ and $Cov(W_t, W_{t+k}) = \lim_{n\to\infty}\sum_{j=M_m+1}^{\infty}\psi_{j;n}\psi_{j+|k|;n}$ ($k\in\mathbb{Z}$). Hence, for the $d_2^{(1)}$ convergence of $\sum_{j=M_m+1}^{\infty}\psi_{j;n}\varepsilon_{t-j;n}$ to W_t, it will be sufficient to verify the Lindeberg condition.
For all $\zeta > 0$, for all $\kappa > 0$ there exists an $m_2 = m_2(\kappa,\zeta)$ and there exists $n_{3,m} = n_3(m,\kappa,\zeta)$ such that

$$\begin{aligned} &\sum_{j=M_m+1}^{\infty}\psi_{j;n}^2\mathbf{E}[\varepsilon_{t;n}^2 1_{[|\varepsilon_{t;n}|>\kappa/|\psi_{j;n}|]}] \le \sum_{j=M_m+1}^{\infty}\psi_{j;n}^2\mathbf{E}[\varepsilon_{t;n}^2 1_{[|\varepsilon_{t;n}|>\kappa/\max_{j\ge M_m+1}|\psi_{j;n}|]}] \\ \le\ &\sum_{j=M_m+1}^{\infty}\psi_{j;n}^2\mathbf{E}[\varepsilon_{t;n}^2 1_{[|\varepsilon_{t;n}|>\kappa/\delta_m]}] \le \zeta \text{ for all } m \ge m_2,\ n \ge n_{3,m}, \end{aligned}$$

here we have used the bound in (5.17) and the fact that $d_2^{(1)}(\varepsilon_{t;n}, \varepsilon_t) \to 0$ $(n \to \infty)$. Thus the Lindeberg condition holds. We now have: for all $\delta > 0$ there exists an $m_3 = m_3(\delta)$ and an $n_{4,m} = n_4(m, \delta)$ such that

$$d_2((\sum_{j=M_m+1}^{\infty} \psi_{j;n}\varepsilon_{t-j;n})_{t\in\mathbb{Z}}, (W_t)_{t\in\mathbb{Z}}) \le \delta \text{ for all } m \ge m_3,\ n \ge n_{4,m}. \tag{5.19}$$

Note that m_1 and m_3 in (5.18) and (5.19) are functions of $\delta > 0$. Thus, by defining $m_0 = m_0(\delta) = \max\{m_1, m_3\}$ and for $n_0 = n_0(\delta) = \max\{n_{2,m_0}, n_{4,m_0}\}$ we get by (5.18) and (5.19): for $n \ge n_0$

$$d_2(\sum_{j=0}^{M_{m_0}} \psi_{j;n}\varepsilon_{t-j;n}, \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j}) \le \delta,$$

$$d_2(\sum_{j=M_{m_0}+1}^{\infty} \psi_{j;n}\varepsilon_{t-j;n}, W_t) \le \delta.$$

By independence (for fixed t) of $\sum_{j=0}^{M_{m_0}} \psi_{j;n}\varepsilon_{t-j;n}$ and $\sum_{j=M_{m_0}+1}^{\infty} \psi_{j;n}\varepsilon_{t-j;n}$ and hence of $\sum_{j=0}^{\infty} \psi_j\varepsilon_{t-j}$ and W_t we get

$$\lim_{n\to\infty} d_2^{(1)}(X_{t;n}, \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j} \oplus W_t) = 0\ (t \in \mathbb{Z}),$$

(here, the $\oplus$ operation is for real valued random variables).
This completes the proof of assertion (ii) for the one dimensional case.
Similarly as in (5.7) we can argue that the pieces $(\sum_{j=0}^{M_{m_0}} \psi_{j;n}\varepsilon_{t-j;n})_{t\in\mathbb{Z}}$ and $(\sum_{j=M_{m_0}+1}^{\infty} \psi_{j;n}\varepsilon_{t-j;n})_{t\in\mathbb{Z}}$ become independent in the limit. By the same arguments as in the one dimensional case and using the Cramér-Wold device we then see that the Mallows d_2 limit of $\{(X_{t;n})_{t\in\mathbb{Z}}\}_{n\in\mathbb{N}}$ is in $S_1 \oplus S_2$.
Proof of assertion (iii):
We first characterize the Mallows d_2 limit $(X_t)_{t\in\mathbb{Z}}$ of the sequence $\{(X_{t;n})_{t\in\mathbb{Z}}; n \in \mathbb{N}\}$. Consider any finite dimensional vector

$$(X_{t_1;n}, \ldots, X_{t_m;n})' = \sum_{j=0}^{\infty} (\psi_{j;n}, \psi_{j+t_2-t_1;n}, \ldots, \psi_{j+t_m-t_1;n})' \varepsilon_{t-j;n} + o_P(1).$$

This is essentially a vector sum of independent variables, which are by assumption (A2) u.a.n. (uniformly asymptotically negligible). By Theorem 2 of Takano (1956) the characteristic function $\varphi_X(.)$ of the limiting distribution of $(X_{t_1;n}, \ldots, X_{t_m;n})'$ is thus characterized by the Khintchine-Lévy representation $\varphi_X(s) = exp(\psi_X(s))$ $(s \in \mathbb{R}^m)$ with

$$\psi_X(s) = -s'\Sigma s/2 + \int_{\mathbf{R}^m} (exp(is'x) - 1 - is'x) \frac{1}{\|x\|^2} dK(x),\ s \in \mathbb{R}^m, \tag{5.20}$$

where K is a measure with $K(\mathbb{R}^m) < \infty$, $K(\{0\}) = 0$ and Σ a non-negative definite $m \times m$ matrix.
The part $-s'\Sigma s/2$ in the $\psi_X(.)$ function corresponds to a Gaussian process $(W_t)_{t\in\mathbb{Z}}$ with

$\mathbb{E}[W_t] = 0$ and which satisfies $Cov(W(t_i), W(t_j)) = \Sigma_{i,j}$ $(i, j = 1, \ldots, m)$. Moreover, this Gaussian part is independent from the rest, given by the measure K, since it is an additive term in the $\psi_X(.)$ function.
We will show now that we can approximate the measure K, or the non-Gaussian part $\int_{\mathbb{R}^d}(exp(is'x) - 1 - is'x)\frac{1}{\|x\|^2}dK(x)$, by a quantity corresponding to a process in S_3. Decompose the non-Gaussian part of $\psi_X(s)$ as

$$\sum_{r=1}^{\infty}\int_{I_r}(exp(is'x) - 1 - is'x)\frac{1}{\|x\|^2}dK(x),\ s \in \mathbb{R}^d,$$

where $I_1 = \mathbb{R}^m \setminus [-1, 1]^m$, $I_r = [-(r-1)^{-1}, (r-1)^{-1}]^m \setminus [-r^{-1}, r^{-1}]^m$, $r = 2, 3, \ldots$. Note that the sets $I_1, I_s, \ldots$ are disjoint which will yield the independence of the Poisson variables $N_1, N_2, \ldots$ in the S_3-representation.
Consider

$$(Y_t)_{t\in\mathbb{Z}} \in S_3,\ Y_t = \sum_{i=1}^{\infty}\sum_{j=1}^{N_i}\xi_{t;i,j}$$

where $\xi_{t_1,\ldots,t_m;i,1} \sim F_i$ with $\lambda_i dF_i(x) = dK(x)/\|x\|^2 1_{[x\in I_i]}$, $i \in \mathbb{N}$ and $N_1, N_2, \ldots$ independent, $N_i \sim$ Poisson(λ_i), $\lambda_i = \int_{I_i} dK(x)/\|x\|^2$.
Then a straightforward calculation for the characteristic function $\varphi_Y(s) = \mathbb{E}[exp(is'(Y_{t_1}, \ldots, Y_{t_m})')]$ $(s \in \mathbb{R}^m)$ yields,

$$\varphi_Y(s) = \prod_{i=1}^{\infty} exp(\lambda_i(\varphi_i(s) - 1)),$$

where $\varphi_r(s) = \mathbb{E}[exp(is'(\xi_{t_1,r,1}, \ldots, \xi_{t_m,r,1})']$ $(r \in \mathbb{N})$.
Thus, in the Khintchine-Lévy representation

$$\begin{aligned}\psi_Y(s) &= log(\varphi_Y(s)) = \sum_{r=1}^{\infty}\lambda_r\int_{I_r}(exp(is'x) - 1 - is'x)dF_r(x) \\ &= \sum_{r=1}^{\infty}\int_{I_r}(exp(is'x) - 1 - is'x)\frac{1}{\|x\|^2}dK(x) = \psi_X(s). \qquad (5.21)\end{aligned}$$

Putting the pieces together, we consider the process

$$(W_t \oplus Y_t)_{t\in\mathbb{Z}} \ \in S_1 \oplus S_3$$

as a representation of $(X_t)_{t\in\mathbb{Z}}$.
Finally by (5.20),

$$\lim_{n\to\infty} d_2((X_{t;n})_{t\in\mathbb{Z}}, (W_t \oplus Y_t)_{t\in\mathbb{Z}}) = 0.$$

This completes the proof for assertion (iii).
For assertion (iv), observe that by (A2), $\psi_j = 0$ for all $j \in \mathbb{N}_0$. Thus, statement (iv) follows from (ii) and (iii). □

Proof of Theorem 3.3 (i)

Consider the process $(X_t)_{t\in\mathbb{Z}}$ in S_3, where

$$X_t = \sum_{j=1}^{N} \xi_{t;j} \ (t \in \mathbb{Z}),$$

and $(\xi_{t;1})_{t\in\mathbb{Z}} = (\xi_t)_{t\in\mathbb{Z}}$, $(\xi_{t;2})_{t\in\mathbb{Z}}, \ldots$ are independent copies of $(\xi_t)_{t\in\mathbb{Z}}$, $N \sim$ Poisson(1). By Lemma 5.1(ii), approximate $(\xi_t)_{t\in\mathbb{Z}}$ by a sequence of ergodic processes $\{(\xi_t^{(r)})_{t\in\mathbb{Z}}\}_{r\in\mathbb{N}}$, with densities for all marginal distributions, such that

$$\lim_{r\to\infty} d_V((\xi_t^{(r)})_{t\in\mathbb{Z}}, (\xi_t)_{t\in\mathbb{Z}}) = 0 \tag{5.22}$$

and $\mathbb{E}[\xi_t^{(r)}] = 0$ for every $r \in \mathbb{N}$. Without loss of generality we assume $\mathbb{E}|\xi_t^{(r)}|^2 < \infty$; this can be achieved by an additional truncation argument in the proof of Lemma 5.1(ii).
Then, consider the process

$$X_{t;r} = \sum_{j=1}^{N} \xi_{t;j}^{(r)} \ (t \in \mathbb{Z}),$$

where $(\xi_{t;1}^{(r)})_{t\in\mathbb{Z}}$, $(\xi_{t;2}^{(r)})_{t\in\mathbb{Z}}, \ldots$ are independent copies of $(\xi_t^{(r)})_{t\in\mathbb{Z}}$ $(r \in \mathbb{N})$.
Then, by straightforward arguments and using (5.22),

$$\lim_{r\to\infty} d_V((X_{t;r})_{t\in\mathbb{Z}}, (X_t)_{t\in\mathbb{Z}}) = 0. \tag{5.23}$$

Consider now the sequence of MA processes

$$X_{t;r}^{(n)*} = \sum_{j=1}^{n} \bar{\xi}_j^{(r)} U_{t-j;n,r}^* \ (t \in \mathbb{Z},\ r \in \mathbb{N}),$$

where $(\bar{\xi}_j^{(r)})_{j\in\mathbb{Z}}$ is now a fixed realization of the stationary, ergodic, mean zero process appearing in formula (5.22), and $U_{t;n,r}^* = U_{t;n} + Z_{t;n,r}$, $(U_{t;n})_{t\in\mathbb{Z}}$ an i.i.d. sequence with $\mathbb{P}[U_{t;n} = 1] = 1 - \mathbb{P}[U_{t;n} = 0] = n^{-1}$ and $(Z_{t;n,r})_{t\in\mathbb{Z}}$ an i.i.d. sequence with $Z_{t;n,r} \sim \mathcal{N}(0, \sigma_{n,r}^2)$, the processes being independent from $(X_{t;r})_{t\in\mathbb{Z}}$ and from each other. This is similar to the processes in Example 1.1. By choosing $\sigma_{n,r}^2 \to 0$ $(n \to \infty)$ appropriately, we will show that $(X_{t;r}^{(n)*})_{t\in\mathbb{Z}}$ converges in the variation metric to $(X_{t;r})_{t\in\mathbb{Z}}$. We will consider convergence for the one dimensional marginal distribution. Since $\xi_t^{(r)}$ has a density with respect to the Lebesgue measure (see Lemma 5.1(ii)), also $X_{t;r}$ has a density $f_r(.)$ for the corresponding c.d.f. $F_r(.)$.
We write

$$X_{t;r}^{(n)*} = X_{t;r}^{(n)} + W_{t;n,r},$$
$$X_{t;r}^{(n)} = \sum_{j=1}^{n} \bar{\xi}_j^{(r)} U_{t-j;n},\ W_{t;n,r} = \sum_{j=1}^{n} \bar{\xi}_j^{(r)} Z_{t-j;n,r},$$

so that $W_{t;n,r} \sim \mathcal{N}(0, \nu_{n,r}^2)$, $\nu_{n,r}^2 = \sigma_{n,r}^2 \sum_{j=1}^{n} (\bar{\xi}_j^{(r)})^2 \sim \sigma_{n,r}^2 n \mathbf{E}|\xi_1^{(r)}|^2$ $(n \to \infty)$.
Denoting the c.d.f. of $X_{t;r}^{(n)}$ by $F_{n,r}(.)$, $X_{t;r}^{(n)*}$ has the continuous density

$$f_{n,r}^*(u) = \int_{\mathbf{R}} \nu_{n,r}^{-1} \varphi(\frac{u-z}{\nu_{n,r}}) dF_{n,r}(z).$$

We can bound

$$d_V(X_{t;r}^{(n)*}, X_{t;r}) = 1/2\|X_{t;r}^{(n)*} - X_{t;r}\|_V = 1/2\int_{\mathbf{R}} |f_{n,r}^*(u) - f_r(u)|du. \tag{5.24}$$

Let $\tilde{f}_{n,r}^*(u) = \int_{\mathbf{R}} \nu_{n,r}^{-1}\varphi(\frac{u-z}{\nu_{n,r}})f_r(z)dz$. Clearly, as $\nu_{n,r}^2 \to 0$ $(n \to \infty)$,

$$\lim_{n\to\infty} |\tilde{f}_{n,r}^*(u) - f_r(u)| = 0 \ (u \in \mathbb{R}). \tag{5.25}$$

Moreover, rewrite by partial integration and change of variables,

$$\begin{aligned} |f_{n,r}^*(u) - \tilde{f}_{n,r}^*(u)| &= |\int_{\mathbf{R}} (F_{n,r}(u - v\nu_n) - F_r(u - v\nu_{n,r}))\nu_{n,r}^{-1}\varphi'(v)dv| \\ &\le \|F_{n,r} - F_r\|_\infty \nu_{n,r}^{-1} \int_{\mathbf{R}} |\varphi'(v)|dv. \end{aligned}$$

The distribution $F_{n,r}$ of $X_{t;r}^{(n)}$ converges weakly to the distribution F_r of $X_{t,r}$, this follows from the proof of formula (5.12). Since the distribution F_r is absolutely continuous, we get by Polya's Theorem $\|F_{n,r} - F_r\|_\infty \to 0$ $(n \to \infty)$. By choosing $\sigma_{n,r}^2 = \|F_{n,r} - F_r\|_\infty / n$ we get $\nu_{n,r}^2 \to 0$, $\|F_{n,r} - F_r\|_\infty \nu_{n,r}^{-1} \to 0$ $(n \to \infty)$ and therefore

$$\lim_{n\to\infty} |f_{n,r}^*(u) - \tilde{f}_{n,r}^*(u)| = 0 \ (u \in \mathbb{R}). \tag{5.26}$$

By (5.25), (5.26) and Scheffé's Theorem,

$$\lim_{n\to\infty} \int_{\mathbf{R}} |f_{n,r}^*(u) - f_r(u)|du = 0 \ (r \in \mathbb{N}).$$

This, together with (5.24) yields $d_V^{(1)}(X_{t;r}^{(n)*}, X_{t;r}) \to 0$ $(n \to \infty)$.
For the higher dimensional marginal distributions one can argue similarly. Therefore,

$$\lim_{n\to\infty} d_V((X_{t;r}^{(n)*})_{t\in\mathbb{Z}}, (X_{t;r})_{t\in\mathbb{Z}}) = 0 \ (r \in \mathbb{N}). \tag{5.27}$$

Note that $\mathbb{E}[X_{t;r}^{(n)*}] \neq 0$. But since $n^{-1}\sum_{j=1}^n \xi_j^{(r)} \to 0$ a.s $(n \to \infty)$, the formula (5.27) holds also for the mean corrected process $\tilde{X}_{t;r}^{(n)*} = \sum_{j=1}^n \xi_j^{(r)}(U_{t-j;n,r}^* - n^{-1})$. Also, the sequence $\{(\tilde{X}_{t;r}^{(n)*})_{t\in\mathbb{Z}}\}_{n,r\in\mathbb{N}}$ satisfies (A.MA) and (A2). By (5.23) and (5.27), there exists a subsequence $\{n_r\}_{r\in\mathbb{N}} \subseteq \mathbb{N}$, such that

$$\lim_{r\to\infty} d_V((X_{t;r}^{(n_r)*})_{t\in\mathbb{Z}}, (X_t)_{t\in\mathbb{Z}}) = 0.$$

Moreover,

$$\mathbb{P}[X_t = \xi_t \text{ for all } t \in \mathbb{Z} | (\xi_t)_{t\in\mathbb{Z}}] \ge \mathbb{P}[N = 1] = exp(-1) > 0.36.$$

This completes the proof. □

Proof of Theorem 3.4

With respect to the L^1-norm, the set $\mathcal{F}_{g(.),h(.),K}$ is uniformly bounded (as a set of probability densities) and therefore conditionally compact, cf. Dunford and Schwartz (1957, Th. 20, Ch. IV.8). In metric spaces, this is equivalent to sequential compactness. We now show that $\mathcal{F}_{g(.),h(.),K}$ is even closed in L^1 and hence compact. Let $\{f_n\}_{n\in\mathbb{N}}$ be a sequence in $\mathcal{F}_{g(.),h(.),K}$ with L^1 limit f, i.e., $\lim_{n\to\infty}\int_{\mathbf{R}}|f_n(x)-f(x)|dx=0$. By the L^1 convergence, we immediately get

$$f\geq 0,\ \int_{\mathbf{R}} f(x)dx=1,\ \int_{\mathbf{R}}|f(x)-f(x+c)|dx<\delta \text{ for all } |c|\leq g(\delta),$$
$$(\int_{-\infty}^{-A}+\int_{A}^{\infty})|f(x)|dx\leq\delta \text{ for all } A\geq h(\delta)\ (\delta>0).$$

Moreover, by Fatou's Lemma,

$$\int_{\mathbf{R}} x^2 f(x)dx\leq\liminf_{n\to\infty}\int_{\mathbf{R}} x^2 f_n(x)dx\leq K<\infty.$$

Finally, by uniform integrability, due to the bound $K<\infty$ for second moments,

$$\int_{\mathbf{R}} x f(x)dx=0.$$

This shows that $f\in\mathcal{F}_{g(.),h(.),K}$. Hence $\mathcal{F}_{g(.),h(.),K}$ is closed with respect to the L^1 norm.
Consider a sequence $\{(X_{t;n})_{t\in\mathbb{Z}}\}_{n\in\mathbb{N}}$, $X_{t;n}=\sum_{j=0}^{\infty}\psi_{j;n}\varepsilon_{t-j;n}$, of processes in $S_{MA;g(.),h(.),K,\beta,C}$ of which a suitable subsequence converges to a d_2 (or d_V) limit $(X_t)_{t\in\mathbb{Z}}$. The aim is to show that $(X_t)_{t\in\mathbb{Z}}\in S_{MA;g(.),h(.),K,\beta,C}$.
Denote by $f_n\in\mathcal{F}_{g(.),h(.),K}$ the innovation density, i.e., $\varepsilon_{t;n}\sim f_n(x)dx$. Since $\mathcal{F}_{g(.),h(.),K}$ is compact, the L^1 limit f is again in $\mathcal{F}_{g(.),h(.),K}$, where f is defined by

$$\lim_{n\to\infty}\int_{\mathbf{R}}|f_n(x)-f(x)|dx=0$$

(or take a suitable subsequence $\{n_k\}_{k\in\mathbb{N}}\subseteq\mathbb{N}$).
This again implies that

$$\lim_{n\to\infty} d_V^{(1)}(\varepsilon_{t;n},\varepsilon_t)=0\ (t\in\mathbb{Z}), \tag{5.28}$$

where $(\varepsilon_t)_{t\in\mathbb{Z}}$ is an i.i.d. sequence with $\varepsilon_t\sim f(x)dx$.
Without loss of generality we denote by $\psi_j=\lim_{n\to\infty}\psi_{j;n}$. Since $(\psi_{j;n})_{j\in\mathbb{N}_0}\in\mathcal{M}_{\beta,C}$ ($\beta>0$) for all $n\in\mathbb{N}$, we get

$$\lim_{n\to\infty}\sum_{j=0}^{\infty}\psi_{j;n}^2=\sum_{j=0}^{\infty}\psi_j^2.$$

We know from Theorem 3.2, see also Remark 3.2, that $\{(X_{t;n})_{t\in\mathbb{Z}}\}_{n\in\mathbb{N}}$ has a d_2 limit $(X_t)_{t\in\mathbb{Z}}$, where

$$X_t=\sum_{j=0}^{\infty}\psi_j\varepsilon_{t-j}\ (t\in\mathbb{Z}). \tag{5.29}$$

Since $\sup_{n\in\mathbb{N}}\sum_{j=0}^{\infty}j^{\beta}\psi_{j;n}^2\leq C$, we get $\sum_{j=0}^{\infty}j^{\beta}\psi_j^2\leq C$ and hence

$$(\psi_j)_{j\in\mathbb{N}_0}\in\mathcal{M}_{\beta,C}.$$

This together with (5.29) and the fact that $\varepsilon_t \sim f(x)dx,\ f \in \mathcal{F}_{g(.),h(.),K}$ implies $(X_t)_{t\in\mathbb{Z}} \in S_{MA;g(.),h(.),K,\beta,C}$.

From (5.28) and the fact that $\mathbb{E}|\varepsilon_t|^2 \leq K < \infty$, it follows from the proof of Theorem 3.2, see also Remark 3.2, that $(X_t)_{t\in\mathbb{Z}}$ as defined in (5.29) is also the weak convergence limit of $\{(X_{t;n})_{t\in\mathbb{Z}}\}_{n\in\mathbb{N}}$. Hence, the closure of $S_{MA;g(.),h(.),K,\beta,C}$ with respect to weak convergence on finite dimensional distributions is a subset (not strict) of $S_{MA;g(.),h(.),K,\beta,C}$. Since the closure of a set with respect to the d_V metric is always a subset (not strict) of the closure of the same set with respect to weak convergence on finite dimensional distributions, we conclude that the d_V closure of $S_{MA;g(.),h(.),K,\beta,C}$ is a subset (not strict) of $S_{MA;g(.),h(.),K,\beta,C}$. □

5.2 Proofs of results in section 4

Theorem 4.1 follows by Theorem 4.2.

Proof of Theorem 4.2

The proof follows in large parts the proofs in section 5.1. The key is to represent $(X_{t;n})_{t\in\mathbb{Z}}$ as an MA process, see formula (4.2). Then assertions (i) and (ii) follow as in the proof of Theorem 3.2. The cases in assertions (iii) and (iv) have in common that $\sigma_n^2 = \mathbb{E}|\varepsilon_{t;n}|^2 \to 0$ and hence $\sup_{n\in\mathbb{N}} \sum_{j=0}^{\infty} \psi_{j;n}^2 = \infty$. However, by (A4) the sequence of variables $\psi_{j;n}\varepsilon_{t-j;n}$ is still u.a.n., i.e.,

$$\max_{j\geq 0} \sigma_n|\psi_{j;n}| \to 0 \ (n \to \infty).$$

Now, assertion (iv) follows as in the proof of Theorem 3.2, assertion (iii) is a special case, where Lindeberg's condition holds, and therefore formula (5.20) consists only of the Gaussian part $-s'\Sigma s/2$. This completes the proof of Theorem 4.2. □

Proof of Proposition 4.1

The AR(1) process $(X_{t;n})_{t\in\mathbb{Z}}$ can be represented as

$$X_{t;n} = \sum_{j=0}^{\infty} \phi_n^j \varepsilon_{t-j;n} \ (t \in \mathbb{Z}).$$

Moreover, $\mathbb{E}|\varepsilon_{t;n}|^2 = 1 - \phi_n^2$ so that $\mathbb{E}|X_{t;n}|^2 = 1$ for all $n \in \mathbb{N}$. Next we will show that for fixed t,

$$\begin{aligned} &X_{t;n} \Rightarrow X_t \ (n \to \infty), \text{ with } \psi(s) = log(\mathbb{E}[exp(isX_t)]) \text{ given by} \\ &\psi(s) = \int_{\mathbf{R}} (exp(isx) - 1 - isx)/x^2 dK(x),\ dK(x) = x1_{[-1<x\leq 1]}dx. \end{aligned} \tag{5.30}$$

Note that X_t cannot be Gaussian since the Khintchine-Lévy measure $K(.)$ does not jump at zero. Moreover, formula (5.30) does imply d_2 convergence since $\mathbb{E}|X_t|^2 = \int_{\mathbf{R}} dK(x) = 1$, which equals $\lim_{n\to\infty} \mathbb{E}|X_{t;n}|^2$.
To show (5.30) it is sufficient to prove

$$K_n \text{ converges weakly to } K,\ K_n(y) = \sum_{j=0}^{\infty} \int_{-\infty}^{y} x^2 dF_{j;n}(x), \tag{5.31}$$

where $F_{j;n}$ is the distribution of $\phi_n^j \varepsilon_{t-j;n}$.
We consider

$$\sum_{j=0}^{\infty} \int_{\delta}^{1} dF_{j;n}(x) = (1-\phi_n^2)/2 \sum_{j=0}^{\infty} 1_{[\delta < \phi_n^j \leq 1]} \ (0 < \delta < 1).$$

Since $\phi_n^j > \delta$ is equivalent to $j < log(\delta)/log(\phi_n)$ and $-2log(\phi_n) \sim 1-\phi_n^2$ $(n \to \infty)$, we get

$$\sum_{j=0}^{\infty} \int_{\delta}^{1} dF_{j;n}(x) = (1-\phi_n^2)/2[log(\delta)/log(\phi_n)] \to -log(\delta) \ (n \to \infty).$$

By symmetry, the same is true for $-1 < \delta < 0$. By the Radon-Nikodym Theorem we then arrive at

$$\sum_{j=0}^{\infty} x^2 dF_{j;n}(x) \to x^2 x^{-1} 1_{[-1 < x \leq 1]} dx = x 1_{[-1 < x \leq 1]} dx \ (n \to \infty).$$

Therefore (5.31) follows and hence (5.30) holds.
Finally, since $\phi_n \to 1$ and $\varepsilon_{t;n} = o_P(1)$ we have $X_{t;n} - X_{t-1;n} = o_P(1)$ and therefore $X_t = X_{t-1}$ for all $t \in \mathbb{Z}$. □

Proof of Theorem 4.3

Consider a sequence $\{(X_{t;n})_{t\in\mathbb{Z}}\}_{n\in\mathbb{N}}$, $X_{t;n} = \sum_{j=1}^{\infty} \phi_{j;n} X_{t-j;n} + \varepsilon_{t;n}$, of processes in $S_{AR;g(.),h(.),K,\beta,C}$ of which a suitable subsequence converges to a d_2 (or d_V) limit $(X_t)_{t\in\mathbb{Z}}$. The aim is to show that $(X_t)_{t\in\mathbb{Z}} \in S_{AR;g(.),h(.),K,\beta,C}$.
We use the representation (4.2) and then get with the formula $\psi_{j;n} = \sum_{k=0}^{j-1} \psi_{k;n}\phi_{j-k;n}$, cf. Markushevich (1977, Vol. 1, p.438),

$$\sup_{n\in\mathbb{N}} \sum_{j=0}^{\infty} j^{\beta-1} |\psi_{j;n}| < \infty. \tag{5.32}$$

Without loss of generality, we denote by $\psi_j = \lim_{n\to\infty} \psi_{j;n}$. Note that now $\lim_{n\to\infty} \sum_{j=0}^{\infty} \psi_{j;n}^2 = \sum_{j=0}^{\infty} \psi_j^2$, since by (5.32), the tail sum $\sup_{n\in\mathbb{N}} \sum_{j=M+1}^{\infty} |\psi_{j;n}| \leq const. M^{-\beta+1} \to 0$ $(M \to \infty)$. Formula (5.32) also implies

$$\sum_{j=0}^{\infty} |\psi_j| < \infty. \tag{5.33}$$

Furthermore, since $\sup_{n\in\mathbb{N}} \sum_{j=1}^{\infty} |\phi_{j;n}| < \infty$, we have

$$\inf_{n\in\mathbb{N},\ |z|\leq 1} |\Psi_n(z)| > 0. \tag{5.34}$$

Since $\psi_{j;n} \to \psi_j$ $(n \to \infty)$ and by using (5.32)-(5.34) one can show in a straightforward way

$$\lim_{n\to\infty} \Psi_n(z) = \Psi(z) \text{ and } \Psi(z) = \sum_{j=0}^{\infty} \psi_j z^j \neq 0 \text{ for } |z| \leq 1. \tag{5.35}$$

By (5.33) and (5.35) we can invert and write $1/\Psi(z) = \Phi(z) = 1 - \sum_{j=1}^{\infty} \phi_j z^j \neq 0$ for $|z| \leq 1$.
Since $\Psi_n(z) \to \Psi(z)$ for $|z| \leq 1$ and by the invertibility of $\Psi_n(.)$ and $\Psi(.)$, we also have $\Phi_n(z) \to \Phi(z)$ for $|z| \leq 1$. By assumption, $\sup_{n\in\mathbb{N}} \sum_{j=1}^{\infty} j^{\beta}|\phi_{j;n}| \leq C < \infty$ and therefore $\sum_{j=1}^{\infty} j^{\beta}|\phi_j| \leq C$. Thus, $(\phi_j)_{j\in\mathbb{N}} \in \mathcal{A}_{\beta,C}$.
We know from the proof of Theorem 3.2 that $(X_{t;n})_{t\in\mathbb{Z}}$ has a d_2 limit $X_t = \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j}$, which can be represented by inversion as $X_t = \sum_{j=1}^{\infty} \phi_j X_{t-j} + \varepsilon_t$. By the proof of Theorem 3.4 in section 5.1, $\varepsilon_t \sim f(x)dx$ with $f \in \mathcal{F}_{g(.),h(.),K}$ and $\mathbf{E}|\varepsilon_t|^2 < \infty$. Moreover, since $(\phi_j)_{j\in\mathbb{N}} \in \mathcal{A}_{\beta,C}$, the process $(X_t)_{t\in\mathbb{Z}}$ is in $S_{AR;g(.),h(.),K,\beta,C}$.

As in the proof of Theorem 3.4 in section 5.1, the d_V closure of $S_{AR;g(.),h(.),K,\beta,C}$ is a subset (not strict) of $S_{AR;g(.),h(.),K,\beta,C}$.

Finally, the processes in $S_{AR;g(.),h(.),K,\beta,C}$ are strong-mixing, since their MA representations satisfy the conditions in Gorodetskii (1977). □

Acknowledgments: We thank David Freedman and David Aldous for helpful conversations.

References

[1] Akamanam, S.I., Rao, M. Bhaskara, Subramanyam, K. (1986). On the ergodicity of bilinear time series models. J. Time Series Anal. **7** 157- 163.

[2] Bickel, P.J. and Bühlmann, P. (1996). What is a linear process. To appear in Proc. National Academy Sciences of U.S.A.

[3] Bickel, P.J. and Freedman, D.A. (1981). Some asymptotic theory for the bootstrap. Ann. Statist. **9** 1196-1217.

[4] Doukhan, P. (1994). Mixing. Properties and Examples. Lect. Notes in Stat. **85**. Springer, New York.

[5] Dunford, N. and Schwartz, J.T. (1957). *Linear Operators. Part I: General Theory.* Interscience Publishers, New York.

[6] Durrett, R. (1991). *Probability: Theory and Examples.* Wadsworth & Brooks/Cole, Pacific Grove.

[7] Gorodetskii, V.V. (1977). On the strong mixing property for linear sequences. Theory Probab. Appl. **22** 411-413.

[8] Grenander, U. and Szegö, G. (1984). *Toeplitz Forms and their Applications.* 2nd ed. Chelsea, New York.

[9] Hannan, E.J. (1987). Rational transfer function approximation. Stat. Science **5** 105-138.

[10] Hartigan, J.A. (1983). *Bayes Theory.* Springer, New York.

[11] Kleiner, B., Martin, R.D. and Thomson, D.J. (1979). Robust estimation of power spectra. J. Roy. Statist. Soc. B **41** 313-351.

[12] Linnik, Y.V. (1964). *Decomposition of Probability Distributions.* Oliver & Boyd, Edinburgh.

[13] Mallows, C.L. (1972). A note on asymptotic joint normality. Ann. Math. Statist. **43** 508-515.

[14] Markushevich, A.I. (1977). *Theory of Functions of a Complex Variable.* 2nd ed. Translated by R.A. Silverman. Chelsea, New York.

[15] Subba Rao, T. and Gabr, M.M. (1984). An Introduction to Bispectral Analysis and Bilinear Time Series Models. Lect. Notes in Statist. **24**. Springer, Heidelberg.

[16] Takano, K. (1956). Multidimensional central limit criterion in the case of bounded variances. Ann. Inst. Statist. Math. **7** 81-93.

Department of Statistics
367 Evans Hall #3860
University of California
Berkeley, CA 94720-3860
USA