



**HAL**  
open science

# Autocovariance Varieties of Moving Average Random Fields

Carlos Amendola, Viet Son Pham

► **To cite this version:**

Carlos Amendola, Viet Son Pham. Autocovariance Varieties of Moving Average Random Fields. MEGA 2019 - International Conference on Effective Methods in Algebraic Geometry, Jun 2019, Madrid, Spain. hal-02912047

**HAL Id: hal-02912047**

**<https://inria.hal.science/hal-02912047>**

Submitted on 5 Aug 2020

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Autocovariance Varieties of Moving Average Random Fields

Carlos Améndola, Viet Son Pham

## Abstract

We study the autocovariance functions of moving average random fields over the integer lattice  $\mathbb{Z}^d$  from an algebraic perspective. These autocovariances are parametrized polynomially by the moving average coefficients, hence tracing out algebraic varieties. We derive dimension and degree of these varieties and we use their algebraic properties to obtain statistical consequences such as identifiability of model parameters. We connect the problem of parameter estimation to the algebraic invariants known as euclidean distance degree and maximum likelihood degree. Throughout, we illustrate the results with concrete examples. In our computations we use tools from commutative algebra and numerical algebraic geometry.

**MSC 2020:** 14Q15; 60G60; 62F10; 62M10; 62R01

## 1 Introduction

Moving average random fields indexed by the integer lattice  $\mathbb{Z}^d$  generalize the class of discrete-time moving average processes and constitute an important statistical spatial model. They are used to model texture images (cf. [11]), as well as in image segmentation and restoration (cf. [14]). Furthermore, they are connected to ARMA (autoregressive moving average) random fields (cf. [10] and the references therein) and the sampling problem of CARMA (continuous autoregressive moving average) random fields, in which the autocovariance functions of moving average random fields play a crucial role (cf. [16, Section 4.3]).

A *moving average random field*  $(Y_t)_{t \in \mathbb{Z}^d}$  of order  $q = (q_1, q_2, \dots, q_d) \in \mathbb{N}^d$  is defined by the equation

$$Y_t = \sum_{k_1=0}^{q_1} \cdots \sum_{k_d=0}^{q_d} a_k Z_{t-k}, \quad t \in \mathbb{Z}^d,$$

where  $k = (k_1, \dots, k_d)$ ,  $a_k$  are real coefficients and  $(Z_t)_{t \in \mathbb{Z}^d}$  is a real-valued zero-mean white noise (see Definition 2.1). The autocovariance function

$$\gamma(t) = \text{Cov}[Y_0, Y_t], \quad t \in \mathbb{Z}^d,$$

for this type of random field is compactly supported, i.e. only finitely many values are nonzero. More precisely, we have  $\gamma(t) = 0$  for every  $t = (t_1, \dots, t_d) \in \mathbb{Z}^d$  with entries satisfying  $|t_i| > q_i$  for at least one  $i \in \{1, \dots, d\}$ .

We study the autocovariance functions of moving average random fields from an algebraic perspective. Our motivation stems from the field of algebraic statistics [18]. Specifically, inspired by the concept of *moment varieties* [2], here we introduce *autocovariance varieties*. The *moving average variety*  $\mathcal{MA}_q \subseteq \mathbb{P}^N$  (see Definition 3.1) is parametrized by  $(q_1+1)\cdots(q_d+1)$  moving average coefficients  $a_k$  where the indices  $k$  satisfy  $0 \leq k_i \leq q_i$  for  $i = 1, \dots, d$ . These coefficients induce  $(2q_1+1)\cdots(2q_d+1)$  nonzero autocovariance values  $\gamma(t)$ . However, we only consider half of them since the relation  $\gamma(-v) = \gamma(v)$  holds for all  $v \in \mathbb{Z}^d$ .

**Example 1.1.** *Let  $d = 2$  and  $q = (1, 1)$ . Then the  $Q = (1+1)(1+1) = 4$  parameters  $a_{00}, a_{01}, a_{10}, a_{11}$  define the autocovariances*

$$\begin{aligned}\gamma(0, 0) &= a_{00}^2 + a_{10}^2 + a_{01}^2 + a_{11}^2, \\ \gamma(1, 0) &= a_{00}a_{10} + a_{11}a_{01}, \\ \gamma(0, 1) &= a_{00}a_{01} + a_{11}a_{10}, \\ \gamma(1, -1) &= a_{10}a_{01}, \\ \gamma(1, 1) &= a_{00}a_{11}.\end{aligned}$$

*The moving average variety  $\mathcal{MA}_{(1,1)} \subseteq \mathbb{P}^4$  is expected to be 3-dimensional. We characterize it in Theorem 3.4.*

This paper is organized as follows. In Section 2, we give the main definition of a moving average random field and its autocovariance function. We define our main object of study, namely autocovariance varieties, in Section 3. We contrast the properties between moving average processes (one-dimensional) from the higher dimensional moving average random fields. In Theorem 3.9 we establish the dimension and degree of these varieties. In Section 4 we investigate identifiability of the associated models and prove that they are algebraically identifiable. In contrast to the  $d = 1$  case where the degree of the fiber grows with  $q$ , we show that for  $d > 1$  there are generically only two sets of parameters that yield the same autocovariance function. Next, we study two different approaches to estimate model parameters from given samples in Section 5. First, we fit the empirical autocovariance function to the theoretical counterpart using a least squares method. Second, we consider maximum likelihood estimation. Both approaches connect nicely to concepts from algebraic statistics: respectively the ED degree and the ML degree. In Example 5.8, we conduct a simulation study comparing classical local optimization methods to numerical homotopy continuation, where we find that the numerical algebraic geometry (NAG) method performs slightly better.

We use the following notation and terminology in this paper: The components of a vector  $u \in \mathbb{R}^d$  are given by  $u_1, \dots, u_d$  if not stated otherwise. If  $u, v \in \mathbb{Z}^d$ , then we set  $[u, v] := \{s \in \mathbb{Z}^d \mid u_i \leq s_i \leq v_i, 1 \leq i \leq d\}$ , which may be an empty set. The symbol  $\leq$  denotes the lexicographic order and for  $x \in \mathbb{R}^d$  and  $t \in \mathbb{Z}^d$  we define  $x^t := x_1^{t_1} \cdots x_d^{t_d}$ . If  $g: A \rightarrow B$  is a mapping and  $y \in g(A)$ , then  $g^{-1}(y)$  is called fiber of  $y$  and each point inside the fiber is called preimage. A statement that holds *generically*, or for a *generic* point, holds outside an algebraic hypersurface in  $x \in \mathbb{C}^d$ , hence in particular for almost all  $x \in \mathbb{C}^d$  with respect to Lebesgue measure.

## 2 Moving Average Random Fields

Throughout this article, all stochastic objects are defined on a fixed complete probability space  $(\Omega, \mathcal{F}, \mathbb{P})$ .

**Definition 2.1.** (1) A random field  $(Y_t)_{t \in \mathbb{Z}^d}$  is called weakly stationary if  $Y_t \in L^2(\Omega, \mathcal{F}, \mathbb{P})$  for every  $t \in \mathbb{Z}^d$  and  $\gamma(t) := \text{Cov}[Y_0, Y_t] = \text{Cov}[Y_s, Y_{s+t}]$  for every  $t, s \in \mathbb{Z}^d$ . It is called a white noise if  $\gamma(0) > 0$  and  $\gamma(t) = 0$  for every  $0 \neq t \in \mathbb{Z}^d$ . In this case,  $\sigma^2 := \gamma(0)$  is called the white noise variance.

(2) Let  $q_1, \dots, q_d$  be positive integers and  $(Z_t)_{t \in \mathbb{Z}^d}$  be a real-valued zero-mean white noise on  $\mathbb{Z}^d$ . A random field  $(Y_t)_{t \in \mathbb{Z}^d}$  is called a moving average random field of order  $q = (q_1, \dots, q_d)$ , denoted  $MA(q)$ , if it satisfies the equation

$$Y_t = \sum_{k \in [0, q]} a_k Z_{t-k}, \quad t \in \mathbb{Z}^d, \quad (2.1)$$

where  $a_k \in \mathbb{R}$  such that for each  $i = 1, \dots, d$  there exist at least two index vectors  $l, m \in [0, q]$  satisfying  $a_l \neq 0$ ,  $a_m \neq 0$ ,  $l_i = q_i$  and  $m_i = 0$ .

The last condition on the two index vectors  $l, m$  guarantees that the  $MA(q)$  random field has indeed order  $q$  and not a smaller order. We associate to each  $MA(q)$  random field the moving average polynomial

$$\theta(x) = \sum_{k \in [0, q]} a_k x^k,$$

and further, we define the formal backshift operators  $B_1, \dots, B_d$  which act on any random field  $(X_t)_{t \in \mathbb{Z}^d}$  in the following way:

$$(B_i X)_t = X_{(t_1, \dots, t_{i-1}, t_i-1, t_{i+1}, \dots, t_d)}, \quad t \in \mathbb{Z}^d, \quad i = 1, \dots, d.$$

Since the operators  $B_i$  and  $B_j$  commute for any  $i, j$ , equation (2.1) can be written as

$$Y_t = \theta(B)Z_t, \quad t \in \mathbb{Z}^d,$$

where  $B = (B_1, \dots, B_d)$ .

The following proposition establishes the link between the moving average polynomial  $\theta$  and the autocovariance function  $\gamma$  of a  $MA(q)$  random field.

**Proposition 2.2.** Suppose that  $(Y_t)_{t \in \mathbb{Z}^d}$  is a  $MA(q)$  random field driven by a white noise  $(Z_t)_{t \in \mathbb{Z}^d}$  with variance  $\sigma^2$ . Then  $Y$  is weakly stationary, its autocovariance function  $\gamma$  is compactly supported and we have

$$\sigma^2 \theta(x) \theta(x^{-1}) = \sum_{t \in \mathbb{Z}^d} \gamma(t) x^t. \quad (2.2)$$

*Proof.* The facts that  $Y$  is weakly stationary and  $\gamma$  is compactly supported are straightforward. Let  $S = \{k \in \mathbb{Z}^d: a_k \neq 0\}$  denote the set of indices with non-vanishing coefficient  $a_k$ . Then we have that

$$\begin{aligned} \sigma^2 \theta(x) \theta(x^{-1}) &= \sigma^2 \left( \sum_{k \in [0, q]} a_k x^k \right) \left( \sum_{k \in [0, q]} a_k x^{-k} \right) = \sigma^2 \sum_{t \in \mathbb{Z}^d} \left( \sum_{k, k+t \in S} a_k a_{k+t} \right) x^t \\ &= \sum_{t \in \mathbb{Z}^d} \text{Cov} \left[ \sum_{k \in [0, q]} a_k Z_{0-k}, \sum_{k+t \in [0, q]} a_{k+t} Z_{0-k} \right] x^t \\ &= \sum_{t \in \mathbb{Z}^d} \text{Cov}[\theta(B)Z_0, \theta(B)Z_t] x^t = \sum_{t \in \mathbb{Z}^d} \text{Cov}[Y_0, Y_t] x^t = \sum_{t \in \mathbb{Z}^d} \gamma(t) x^t. \quad \square \end{aligned}$$

### 3 Autocovariance Varieties

We have seen that for  $q = (q_1, q_2, \dots, q_d) \in \mathbb{N}^d$ , the autocovariance function of a moving average random field is only dependent on the coefficients  $a_k$  of the moving average polynomial  $\theta(x) = \sum_{k \in [0, q]} a_k x^k$  and the white noise variance  $\sigma^2$ . In order to avoid redundancies in model specification, one can assume without loss of generality that  $\sigma^2 = 1$  and we will do so for the rest of this paper. There are  $Q + 1 := \prod_{i=1}^d (q_i + 1)$  coefficients  $a_k$  and  $2N + 1 := \prod_{i=1}^d (2q_i + 1)$  non-zero autocovariances  $\gamma(t)$  for  $t \in [-q, q]$ . Ordering them in two vectors  $a$  and  $\gamma_a$ , we can think of this correspondence as a polynomial map  $\Gamma_q: \mathbb{R}^{Q+1} \mapsto \mathbb{R}^{2N+1}$  given by  $a \mapsto \gamma_a$ .

Since  $\gamma(-t) = \gamma(t)$  for all  $t \in \mathbb{Z}^d$ , we can drop half of the autocovariances and only consider  $\gamma(t)$  with  $t \in [-q, q]$  and  $t \geq 0$ , where  $\geq$  denotes the lexicographic order. In this way, we have a map  $\mathbb{R}^{Q+1} \mapsto \mathbb{R}^{N+1}$  which we still denote by  $\Gamma_q$ .

The points in the image represent the set of autocovariance functions of moving average random fields. Geometrically, this is a semialgebraic set, defined by polynomial equalities and inequalities. The closure of the image of this parametrization will give a real affine algebraic variety. However, as is standard in algebraic statistics, we will first change the underlying field to be algebraically closed (so the map becomes  $\Gamma_q: \mathbb{C}^{Q+1} \mapsto \mathbb{C}^{N+1}$  over the complex numbers  $\mathbb{C}$ ) and then we pass to projective space arriving at  $\Gamma_q: \mathbb{P}^Q \rightarrow \mathbb{P}^N$ . This last step requires the polynomials  $\gamma(t)$  to be homogeneous, and indeed they are in our case.

**Definition 3.1.** Let  $q = (q_1, q_2, \dots, q_d) \in \mathbb{N}^d$  and define  $Q := \prod_{i=1}^d (q_i + 1) - 1$  and  $N := (\prod_{i=1}^d (2q_i + 1) - 1)/2$ . The autocovariance variety  $\mathcal{MA}_q$  is the image of the autocovariance map  $\Gamma_q: \mathbb{P}^Q \rightarrow \mathbb{P}^N$ .

#### 3.1 Moving average processes

If  $d = 1$ , moving average random fields are also called moving average processes. These processes are well-studied and belong to the important class of ARMA processes (cf. [6, Chapter 3]). Suppose that  $(Y_t)_{t \in \mathbb{Z}}$  is a  $MA(q)$  process given by the equation

$$Y_t = \sum_{k=0}^q a_k Z_{t-k}, \quad t \in \mathbb{Z}.$$

Then, the autocovariance function  $\gamma$  of  $Y$  has the simple expression

$$\gamma(t) = \begin{cases} \sum_{k=0}^{q-|t|} a_k a_{k+|t|}, & \text{if } |t| \leq q, \\ 0, & \text{if } |t| > q. \end{cases}$$

For the class of moving average processes we have that  $Q = N = q$ . Thus, in this special case the autocovariance map takes the form

$$\Gamma_q : \mathbb{P}^q \rightarrow \mathbb{P}^q.$$

In the next subsection we will see that the map is actually defined in all of  $\mathbb{P}^q$  (there are no *base points*), so we conclude the following.

**Proposition 3.2.** *If  $d = 1$ , then  $\mathcal{MA}_q = \mathbb{P}^q$ .*

While  $\mathcal{MA}_q$  is not particularly interesting when  $d = 1$ , the parametrization coming from  $\Gamma_q : \mathbb{P}^q \rightarrow \mathbb{P}^q$  has interesting fibers and computing them is important for statistical applications. This issue of identifiability will be explored in Section 4.

**Remark 3.3.** *Going back for a moment to the real picture (over  $\mathbb{R}$ ), the equality  $\mathcal{MA}_q = \mathbb{P}^q$  is analogous to the statement that when  $d = 1$ , any autocovariance function with support  $[-q, q]$  is an autocovariance function of a  $MA(q)$  process [6, Prop. 3.2.1].*

## 3.2 Moving average random fields

We start by carefully analyzing the case  $\mathcal{MA}_{(1,1)}$  mentioned in the introduction. We denote polynomial ring variables by  $g_t = \gamma(t)$ .

**Theorem 3.4.** *The autocovariance variety  $\mathcal{MA}_{(1,1)} \subseteq \mathbb{P}^4$  is a threefold of degree 4. It is the hypersurface defined by the quartic*

$$\begin{aligned} g_{10}^2 g_{01}^2 - g_{00} g_{10} g_{01} g_{11} + g_{10}^2 g_{11}^2 + g_{01}^2 g_{11}^2 - g_{00} g_{10} g_{01} g_{1-1} + g_{00}^2 g_{11} g_{1-1} - 2g_{10}^2 g_{11} g_{1-1} - 2g_{01}^2 g_{11} g_{1-1} \\ - 4g_{11}^3 g_{1-1} + g_{10}^2 g_{1-1}^2 + g_{01}^2 g_{1-1}^2 + 8g_{11}^2 g_{1-1}^2 - 4g_{11} g_{1-1}^3 = 0. \end{aligned} \quad (3.1)$$

*Its singular locus is a quadratic surface, which is the union of the three irreducible components corresponding to the prime ideals*

$$\langle g_{10} - g_{01}, g_{00} - 2g_{11} - 2g_{1-1}, 4g_{11}g_{1-1} - g_{01}^2 \rangle, \quad (3.2)$$

$$\langle g_{10} + g_{01}, g_{00} + 2g_{11} + 2g_{1-1}, 4g_{11}g_{1-1} - g_{01}^2 \rangle, \quad (3.3)$$

and

$$\langle g_{11} - g_{1-1}, g_{00}g_{1-1} - g_{10}g_{01} \rangle. \quad (3.4)$$

*Proof.* The proof is computational. One way to obtain the quartic (3.1) is through the following Macaulay2 [12] commands:

```

R = QQ[a00,a01,a10,a11]
S = QQ[g00,g01,gm11,g10,g11]
h = map(R,S,{ a00^2 + a10^2 + a01^2 + a11^2,      a00*a01 + a10*a11,
              a10*a01, a00*a10 + a01*a11,  a00*a11} )
I = kernel h

```

For the singular locus, we compute the radical ideal of the quartic along with its vanishing gradient, and then compute its prime decomposition.  $\square$

**Remark 3.5.** *Substituting the parametrization of Example 1.1 into (3.2) to (3.4), we find that the three irreducible components of the singular locus correspond to the three conditions*

$$a_{10} = a_{01} \quad \text{and} \quad a_{00} = a_{11}, \quad (3.5)$$

$$a_{10} = -a_{01} \quad \text{and} \quad a_{00} = -a_{11}, \quad (3.6)$$

and

$$a_{00}a_{11} = a_{01}a_{10}. \quad (3.7)$$

*These conditions represent submodels and we will analyze Equation (3.7) in more detail in Example 4.5.*

The complexity of  $\mathcal{MA}_q$  increases rapidly when  $d > 1$ . It is computationally challenging to obtain generators for its prime ideal even for small values of  $q$  and  $d$ . Beyond  $q = (1, 1)$ , we were also able to do this for  $q = (1, 2)$  and  $q = (1, 1, 1)$ .

**Proposition 3.6.** *The autocovariance variety  $\mathcal{MA}_{(1,2)} \subseteq \mathbb{P}^7$  is 5-dimensional of degree 16. Its prime ideal is cut out by 7 sextics. One of those is*

$$\begin{aligned}
& 4g_{12}^3g_{1-2}g_{11}^2 - 28g_{12}^2g_{1-2}g_{11}^2 + 28g_{12}g_{1-2}g_{11}^3 - 4g_{1-2}g_{11}^4 - 4g_{1-2}g_{11}^4 - g_{12}^2g_{11}^2g_{02}^2 + 6g_{12}g_{1-2}g_{11}^2g_{02}^2 \\
& \quad - 5g_{1-2}^2g_{11}^2g_{02}^2 - 32g_{12}^3g_{1-2}g_{11}g_{1-1} + 64g_{12}^2g_{1-2}g_{11}g_{1-1} - 32g_{12}g_{1-2}g_{11}^3g_{1-1} \\
& - 8g_{12}g_{1-2}g_{11}^3g_{1-1} + 8g_{1-2}^2g_{11}^3g_{1-1} + 6g_{12}^2g_{11}g_{02}^2g_{1-1} - 12g_{12}g_{1-2}g_{11}g_{02}^2g_{1-1} + 6g_{1-2}^2g_{11}g_{02}^2g_{1-1} - 4g_{12}^4g_{1-1}^2 \\
& \quad + 28g_{12}^3g_{1-2}g_{1-1}^2 - 28g_{12}^2g_{1-2}g_{1-1}^2 + 4g_{12}g_{1-2}g_{1-1}^3 - 4g_{12}^2g_{11}^2g_{1-1}^2 + 16g_{12}g_{1-2}g_{11}^2g_{1-1}^2 - 4g_{1-2}^2g_{11}^2g_{1-1}^2 \\
& - 5g_{12}^2g_{02}^2g_{1-1}^2 + 6g_{12}g_{1-2}g_{02}^2g_{1-1}^2 - g_{1-2}^2g_{02}^2g_{1-1}^2 + 8g_{12}^2g_{11}g_{1-1}^3 - 8g_{12}g_{1-2}g_{11}g_{1-1}^3 - 4g_{12}^2g_{1-1}^4 + 8g_{12}^2g_{1-2}g_{11}^2g_{10} \\
& \quad - 16g_{12}g_{1-2}g_{11}^2g_{10} + 8g_{1-2}^3g_{11}^2g_{10} - 2g_{12}g_{11}^2g_{02}^2g_{10} + 6g_{1-2}g_{11}^2g_{02}^2g_{10} + 4g_{12}g_{11}g_{02}^2g_{1-1}g_{10} \\
& \quad + 4g_{1-2}g_{11}g_{02}^2g_{1-1}g_{10} + 8g_{12}^3g_{1-1}^2g_{10} - 16g_{12}^2g_{1-2}g_{1-1}^2g_{10} + 8g_{12}g_{1-2}^2g_{1-1}^2g_{10} + 6g_{12}g_{02}^2g_{1-1}^2g_{10} \\
& - 2g_{1-2}g_{02}^2g_{1-1}^2g_{10} + 4g_{12}g_{1-2}g_{11}^2g_{10}^2 - 4g_{1-2}^2g_{11}^2g_{10}^2 - g_{11}^2g_{02}^2g_{10}^2 - 2g_{11}g_{02}^2g_{1-1}g_{10}^2 - 4g_{12}^2g_{1-1}^2g_{10}^2 \\
& \quad + 4g_{12}g_{1-2}g_{1-1}^2g_{10}^2 - g_{02}^2g_{1-1}^2g_{10}^2 + 12g_{12}^2g_{1-2}g_{11}g_{02}g_{01} - 16g_{12}g_{1-2}^2g_{11}g_{02}g_{01} + 4g_{1-2}^3g_{11}g_{02}g_{01} \\
& \quad + 4g_{1-2}g_{11}g_{02}g_{01} - 2g_{12}g_{11}g_{02}^3g_{01} + 2g_{1-2}g_{11}g_{02}^3g_{01} + 4g_{12}^3g_{02}g_{1-1}g_{01} \\
& \quad - 16g_{12}^2g_{1-2}g_{02}g_{1-1}g_{01} + 12g_{12}g_{1-2}^2g_{02}g_{1-1}g_{01} + 4g_{12}g_{11}^2g_{02}g_{1-1}g_{01} - 8g_{1-2}g_{11}^2g_{02}g_{1-1}g_{01} \\
& \quad + 2g_{12}g_{02}^3g_{1-1}g_{01} - 2g_{1-2}g_{02}^3g_{1-1}g_{01} - 8g_{12}g_{11}g_{02}g_{1-1}^2g_{01} + 4g_{1-2}g_{11}g_{02}g_{1-1}^2g_{01} + 4g_{12}g_{02}g_{1-1}^3g_{01} \\
& \quad - 8g_{1-2}^2g_{11}g_{02}g_{10}g_{01} - 2g_{11}g_{02}^3g_{10}g_{01} - 8g_{12}^2g_{02}g_{1-1}g_{10}g_{01} - 2g_{02}^3g_{1-1}g_{10}g_{01} + 4g_{1-2}g_{11}g_{02}g_{10}^2g_{01} \\
& + 4g_{12}g_{02}g_{1-1}g_{10}^2g_{01} - g_{12}^2g_{02}^2g_{01}^2 + 2g_{12}g_{1-2}g_{02}^2g_{01}^2 - g_{1-2}^2g_{02}^2g_{01}^2 - g_{11}^2g_{02}^2g_{01}^2 + 2g_{11}g_{02}^2g_{1-1}g_{01}^2 - g_{02}^2g_{1-1}^2g_{01}^2
\end{aligned}$$

$$\begin{aligned}
& + 2g_{12}g_{02}^2g_{10}g_{01}^2 + 2g_{1-2}g_{02}^2g_{10}g_{01}^2 - g_{02}^2g_{10}^2g_{01}^2 - 4g_{1-2}^2g_{11}^2g_{02}g_{00} - 8g_{12}g_{1-2}g_{11}g_{02}g_{1-1}g_{00} - 4g_{12}^2g_{02}g_{1-1}^2g_{00} \\
& - 4g_{12}^2g_{1-2}g_{11}g_{01}g_{00} + 12g_{12}g_{1-2}^2g_{11}g_{01}g_{00} + g_{12}g_{11}g_{02}^2g_{01}g_{00} + g_{1-2}g_{11}g_{02}^2g_{01}g_{00} + 12g_{12}^2g_{1-2}g_{1-1}g_{01}g_{00} \\
& - 4g_{12}g_{1-2}^2g_{1-1}g_{01}g_{00} + g_{12}g_{02}^2g_{1-1}g_{01}g_{00} + g_{1-2}g_{02}^2g_{1-1}g_{01}g_{00} - 4g_{12}g_{1-2}g_{11}g_{10}g_{01}g_{00} + g_{11}g_{02}^2g_{10}g_{01}g_{00} \\
& - 4g_{12}g_{1-2}g_{1-1}g_{10}g_{01}g_{00} + g_{02}^2g_{1-1}g_{10}g_{01}g_{00} - 4g_{12}g_{1-2}g_{02}g_{01}^2g_{00} + g_{1-2}^2g_{11}^2g_{00}^2 + 2g_{12}g_{1-2}g_{11}g_{1-1}g_{00}^2 \\
& \quad + g_{12}^2g_{1-1}^2g_{00}^2 - g_{1-2}g_{11}g_{02}g_{01}g_{00}^2 - g_{12}g_{02}g_{1-1}g_{01}g_{00}^2 + g_{12}g_{1-2}g_{01}^2g_{00}^2
\end{aligned}$$

The autocovariance variety  $\mathcal{MA}_{(1,1,1)} \subseteq \mathbb{P}^{13}$  is 7-dimensional of degree 64. Its prime ideal is cut out by 56 quartics, 90 quintics and 50 sextics.

Table 1 presents the basic properties of the first autocovariance varieties  $\mathcal{MA}_{(q_1, q_2)}$ , that is, for  $d = 2$ . The dimension appears to be the expected one, while the degree follows a clear pattern as a power of two. We will prove that this actually holds for any  $\mathcal{MA}_q$ . To that end, we use the next two lemmas.

$q_1$	$q_2$	$\dim(\mathcal{MA}_q)$	$N$	$\deg(\mathcal{MA}_q)$	generators
1	1	3	4	4	1 quartic
1	2	5	7	16	7 sextics
1	3	7	10	64	?
2	2	8	12	128	?

Table 1: Summary of first autocovariance varieties for  $d = 2$

**Lemma 3.7.** *The map  $\Gamma_q : \mathbb{P}^Q \dashrightarrow \mathbb{P}^N$  has no base points.*

*Proof.* Assume that  $\Gamma_q(a) = 0$ . We know from (2.2) that

$$\theta(x)\theta(x^{-1}) = \sum_{t \in \mathbb{Z}^d} \gamma(t)x^t = 0.$$

Multiplying both sides by the monomial  $x^q = x_1^{q_1} \cdots x_d^{q_d}$ , we obtain the product of two polynomials  $\theta(x) \cdot x^q \theta(x^{-1})$  that equals the zero polynomial. Since the polynomial ring  $K[x]$  is an integral domain when  $K$  is a field, we must have that either  $\theta(x) = 0$  or  $x^q \theta(x^{-1}) = 0$ . In particular, all the coefficients  $a_k = 0$ , that is,  $a = 0$  is the zero vector.  $\square$

**Lemma 3.8.** *The autocovariance variety  $\mathcal{MA}_q$  is a linear projection of the Veronese variety. Furthermore,  $\gamma(t)$  is the sum of exactly  $(q_1 - |t_1| + 1) \cdots (q_d - |t_d| + 1)$  quadratic monomials for every  $t \in [-q, q]$  with  $t \geq 0$ , and each monomial appears exactly once.*

*Proof.* The quadratic Veronese embedding precisely consists of all quadratic monomials. The parametrization of  $\mathcal{MA}_q$  consists of quadrics, each one is a sum of quadratic monomials. Moreover, Proposition 2.2 implies that

$$\gamma(t) = \sum_{k, k+t \in [0, q]} a_k a_{k+t}, \quad (3.8)$$

for every  $t \in [-q, q]$ , which shows the second part of the assertions.  $\square$

Now we state our main theorem concerning the varieties  $\mathcal{MA}_q$ .

**Theorem 3.9.** *Let  $q \in \mathbb{N}^d$ . Then*

$$\dim(\mathcal{MA}_q) = Q = \prod_{i=1}^d (q_i + 1) - 1$$

and if  $d > 1$ , then

$$\deg(\mathcal{MA}_q) = 2^{Q-1} = 2^{\prod_{i=1}^d (q_i+1)-2}.$$

*Proof.* Let  $D := \dim(\mathcal{MA}_q)$  denote the dimension of  $\mathcal{MA}_q$  and consider the regular map  $\Gamma_q : \mathbb{P}^Q \rightarrow \mathbb{P}^N$ . Since the domain is  $Q$ -dimensional, the inequality  $D \leq Q$  has to hold. However,  $\Gamma_q$  is not a constant map and has no base points by Lemma 3.7. As there are no nonconstant regular maps from a projective space to a variety of smaller dimension, we must have equality, that is,  $D = Q$ .

Now consider the map  $\Gamma_q$  as a composition of the quadratic Veronese embedding  $\nu_2$  of  $\mathbb{P}^Q$  followed by the linear projection to  $\mathbb{P}^N$ , as in Lemma 3.8. By [15, Proposition 5.5], we have the following equality of degrees:

$$\deg(\nu_2(\mathbb{P}^Q)) = \deg(\mathcal{MA}_q) \deg(\Gamma^{-1}(\gamma))$$

where  $\gamma \in \mathcal{MA}_q$  is generic. Here we know that the linear space we project from does not intersect the Veronese variety  $\mathcal{V}_{Q,2} = \nu_2(\mathbb{P}^Q)$ , by Lemma 3.7. Thus the left hand side equals the degree of the Veronese variety, which is  $2^Q$ . In addition, the identifiability Theorem 4.4 below proves that  $\deg(\Gamma^{-1}(\gamma)) = 2$ . Hence, we conclude that  $\deg(\mathcal{MA}_q) = 2^{Q-1}$ .  $\square$

## 4 Identifiability

We show that all models  $MA(q)$  with  $q \in \mathbb{N}^d$  are algebraically identifiable (in the sense of [3]). This means that the map from the model parameters to the autocovariances is generically finite to one.

### 4.1 Moving Average Processes

The following result shows that the map  $\Gamma_q$  is not injective and the moving average parameters  $a_i$  are not identifiable from a second order point of view if  $d = 1$ .

**Proposition 4.1.** *If  $d = 1$ , the fibers of a generic point  $\gamma \in \mathcal{MA}_q$  consist of  $2^q$  points.*

*Proof.* Let  $\alpha_1, \dots, \alpha_q \in \mathbb{C}$  be the  $q$  roots of the moving average polynomial

$$\theta(x) = \sum_{k=0}^q a_k x^k = a_q (x - \alpha_1) \cdots (x - \alpha_q), \quad x \in \mathbb{C}.$$

Using Proposition 2.2, we see that there are exactly  $2^{q+1}$  polynomials which generate  $\gamma$  as above, all of which have the form

$$\pm a_q (x - \alpha_1)_\pm \cdots (x - \alpha_q)_\pm,$$

where

$$(x - \alpha_i)_+ := (x - \alpha_i) \quad \text{and} \quad (x - \alpha_i)_- := (\alpha_i x - 1). \quad (4.1)$$

Hence, the fiber of any point in  $\mathcal{MA}_q$  under  $\Gamma_q : \mathbb{P}^q \rightarrow \mathbb{P}^q$  consists in general of  $2^q$  points.  $\square$

Proposition 4.1 shows that it is possible to deduce all preimage points from a single one by inverting the roots of  $\theta$  as suggested in (4.1).

In order to obtain injectivity of  $\Gamma_q$ , one usually imposes the condition that all roots  $\alpha_i$  of the polynomial  $\theta$  lie strictly outside the unit disk (and  $a_0 > 0$ ). It follows from the proof above that this is always possible as long as all roots don't lie on the boundary of the unit disk. Compare to [6, Proposition 3.5.1]. This property is also called *invertibility* in the literature [6, Definition 3.1.4] since it holds if and only if there exists coefficients  $\pi_0, \pi_1, \dots$  with  $\sum_{k=0}^{\infty} |\pi_k| < \infty$  such that the white noise sequence  $(Z_t)_{t \in \mathbb{Z}}$  can be expressed as

$$Z_t = \sum_{k=0}^{\infty} \pi_k Y_{t-k}, \quad t \in \mathbb{Z}.$$

**Example 4.2.** Let  $q = 1$ . This is the simplest moving average model  $MA(1)$ . We have that  $\theta(x) = a_0 + a_1 x$  and  $\Gamma_1 : \mathbb{P}^1 \rightarrow \mathbb{P}^1$  is given by

$$\Gamma_1(a_0, a_1) = (a_0^2 + a_1^2, a_0 a_1).$$

The fiber of a generic point  $\gamma = (\gamma_0, \gamma_1)$  consists of  $2 = 2^1$  points in  $\mathbb{P}^1$ . They are  $(\tilde{a}_0, \tilde{a}_1)$  and  $(\tilde{a}_1, \tilde{a}_0)$  where

$$\tilde{a}_0 = \sqrt{\frac{\gamma_0 + \sqrt{\gamma_0^2 - 4\gamma_1^2}}{2}}, \quad \tilde{a}_1 = \sqrt{\frac{\gamma_0 - \sqrt{\gamma_0^2 - 4\gamma_1^2}}{2}}. \quad (4.2)$$

The invertibility condition is equivalent to  $|a_0| > |a_1|$ .

The observed symmetry of the two points  $(\tilde{a}_0, \tilde{a}_1)$  and  $(\tilde{a}_1, \tilde{a}_0)$  above extends to higher  $q$ . In fact, it holds that

$$\Gamma_q(a_0, a_1, \dots, a_{q-1}, a_q) = \Gamma_q(a_q, a_{q-1}, \dots, a_1, a_0). \quad (4.3)$$

This can be seen from (2.2), where the reversal occurs by inverting all the roots in (4.1).

For general  $q > 1$ , there exist algorithms to numerically approximate the invertible solution with  $a_0 = 1$ . A basic one is the *innovations algorithm*, which recursively converges to the moving average parameters  $a_k$  given the autocovariance values  $\gamma(t)$  under the invertibility condition (we refer to Section 2 in [5] for details). Other approaches use spectral factorization methods [17]. While we do not pursue this in this paper, the fact remains that the desired parameters are solutions to a polynomial system of equations, so it would be interesting to compare these with state-of-the-art algorithms in numerical algebraic geometry. See Example 5.8 for an illustration of such techniques.

Furthermore, the symmetry in the polynomial system means that one does not necessarily need to find a root of a polynomial of degree  $2^q$  even when there are  $2^q$  solutions. We illustrate this with  $q = 2$ .

**Example 4.3.** For  $q = 2$  we have  $\theta(x) = a_0 + a_1x + a_2x^2$  and  $\Gamma_2 : \mathbb{P}^2 \longrightarrow \mathbb{P}^2$  given by

$$\Gamma_2(a_0, a_1, a_2) = (a_0^2 + a_1^2 + a_2^2, a_0a_1 + a_1a_2, a_0a_2).$$

The fiber of a generic point  $\gamma = (\gamma_0, \gamma_1, \gamma_2) \in \mathcal{MA}_2$  consists of  $2^2 = 4$  points. A Gröbner basis elimination from the system  $\Gamma_2(a_0, a_1, a_2) = (\gamma_0, \gamma_1, \gamma_2)$  with order  $a_2, a_0, a_1$  reveals a triangular system with a quadric in  $a_1^2$ :

$$\begin{aligned} a_1^4 - (\gamma_0 + 2\gamma_2)a_1^2 + \gamma_1^2 &= 0 \\ a_0^2a_1 - \gamma_1a_0 + a_1\gamma_2 &= 0 \\ a_2a_1 + a_0a_1 - \gamma_1 &= 0. \end{aligned}$$

And hence the solutions for  $(a_0, a_1, a_2)$  in terms of  $(\gamma_0, \gamma_1, \gamma_2)$  can be obtained as

$$a_1 = \sqrt{\frac{\gamma_0 + 2\gamma_2 \pm \sqrt{(\gamma_0 + 2\gamma_2)^2 - 4\gamma_1^2}}{2}}, \quad a_0 = \sqrt{\frac{\gamma_1 \pm \sqrt{\gamma_1^2 - 4a_1^2\gamma_2}}{2a_1}}, \quad a_2 = \frac{\gamma_1 - a_0a_1}{a_1}.$$

## 4.2 Moving Average Random Fields

The following result demonstrates a fundamental difference between  $d = 1$  and  $d > 1$  in terms of identifiability. On the other hand, it shows how the symmetry in (4.3) generalizes to higher dimensions.

**Theorem 4.4.** *Suppose that the moving average polynomial  $\theta$  is generic. Then for  $d > 1$ , the fibers of a point  $\gamma \in \mathcal{MA}_q$  are only two points  $a$  and  $a'$  in  $\mathbb{P}^Q$ . One is obtained from the other by  $a'_k = a_{q-k}$  for any  $k \in [0, q]$ .*

*Proof.* Let  $\gamma$  be the image of the coefficients  $a_k$  of a moving average polynomial  $\theta$  under the mapping  $\Gamma_q$  and assume that  $\theta'$  is another polynomial which also generates  $\gamma$  and has coefficients  $a'_k$ . Due to Proposition 2.2, the polynomial equation

$$\theta(x)(x^q\theta(x^{-1})) = \theta'(x)(x^q\theta'(x^{-1})) \tag{4.4}$$

has to hold. Since generically  $\theta$  is irreducible, we either have  $\theta' = \theta$  or  $\theta' = x^q\theta(x^{-1})$ , which proves the assertion.  $\square$

**Example 4.5.** *We consider again the autocovariance variety  $\mathcal{MA}_{(1,1)}$  and assume that  $\gamma \in \mathcal{MA}_{(1,1)}$  is generated by a generic moving average polynomial  $\theta$  as in the setting of Theorem 4.4, that is,  $\theta$  is irreducible. Using Macaulay2 to compute a Gröbner basis, where*

equation (3.1) is added to the list of polynomials, we obtain the fiber of  $\gamma$  explicitly as

$$\begin{aligned} a_{00} &= \sqrt{\frac{\gamma_{00}\gamma_{11} - \gamma_{01}\gamma_{10} - \sqrt{(\gamma_{01}\gamma_{10} - \gamma_{00}\gamma_{11})^2 - 4\gamma_{11}^2(\gamma_{11} - \gamma_{1-1})^2}}{2(\gamma_{11} - \gamma_{1-1})}}, \\ a_{10} &= \sqrt{\frac{-\gamma_{00}\gamma_{1-1} + \gamma_{01}\gamma_{10} - \sqrt{(\gamma_{01}\gamma_{10} - \gamma_{00}\gamma_{1-1})^2 - 4\gamma_{1-1}^2(\gamma_{11} - \gamma_{1-1})^2}}{2(\gamma_{11} - \gamma_{1-1})}}, \\ a_{01} &= \sqrt{\frac{-\gamma_{00}\gamma_{1-1} + \gamma_{01}\gamma_{10} + \sqrt{(\gamma_{01}\gamma_{10} - \gamma_{00}\gamma_{1-1})^2 - 4\gamma_{1-1}^2(\gamma_{11} - \gamma_{1-1})^2}}{2(\gamma_{11} - \gamma_{1-1})}}, \\ a_{11} &= \sqrt{\frac{\gamma_{00}\gamma_{11} - \gamma_{01}\gamma_{10} + \sqrt{(\gamma_{01}\gamma_{10} - \gamma_{00}\gamma_{11})^2 - 4\gamma_{11}^2(\gamma_{11} - \gamma_{1-1})^2}}{2(\gamma_{11} - \gamma_{1-1})}}, \end{aligned}$$

and  $a'_{00} = a_{11}$ ,  $a'_{01} = a_{10}$ ,  $a'_{10} = a_{01}$ ,  $a'_{11} = a_{00}$ . Substituting in the formulas from Example 1.1, we observe that the discriminants

$$\begin{aligned} &(\gamma_{01}\gamma_{10} - \gamma_{00}\gamma_{11})^2 - 4\gamma_{11}^2(\gamma_{11} - \gamma_{1-1})^2, \\ &(\gamma_{01}\gamma_{10} - \gamma_{00}\gamma_{1-1})^2 - 4\gamma_{1-1}^2(\gamma_{11} - \gamma_{1-1})^2, \end{aligned}$$

are equal to

$$\begin{aligned} &(a_{00}^2 - a_{11}^2)^2 (a_{01}a_{10} - a_{00}a_{11})^2, \\ &(a_{01}^2 - a_{10}^2)^2 (a_{01}a_{10} - a_{00}a_{11})^2, \end{aligned}$$

which are nonnegative in the real case (as they should be when the moving average parameters are real).

If however  $\theta$  is not irreducible, it has to be the product of two linear factors. Then identifiability from the above theorem fails and we have up to 4 preimages in the fiber of  $\gamma$ . We note that this explains the irreducible component (3.4) of the singular locus from Theorem 3.4, which is equivalent to (3.7). In order to see this, we first assume that Equation (3.7) holds. This implies that there exists a constant  $b^* \in \mathbb{R}$  such that we have  $a_{01} = b^*a_{00}$  and  $a_{11} = b^*a_{10}$ . Thus, the polynomial  $\theta$  satisfies

$$\theta(x) = (a_{00} + a_{10}x_1)(1 + b^*x_2)$$

and is therefore reducible. On the other hand, assuming that

$$\theta(x) = a_{00} + a_{10}x_1 + a_{01}x_2 + a_{11}x_1x_2 = (b_{10} + b_{11}x_1)(b_{20} + b_{21}x_2)$$

for some real-valued coefficients  $b_{10}, b_{11}, b_{20}, b_{21}$ , we can deduce (3.7). One of the four preimage points in the fiber of  $\gamma$  is given by the equations

$$a_{00} = \frac{1}{2} \sqrt{\gamma_{00} + \sqrt{\gamma_{00}^2 - 4(\gamma_{01}^2 + \gamma_{10}^2 - 4\gamma_{11}^2)} - \sqrt{2\gamma_{00}\sqrt{\gamma_{00}^2 - 4(\gamma_{01}^2 + \gamma_{10}^2 - 4\gamma_{11}^2)} + 2\gamma_{00}^2 - 4(\gamma_{01}^2 + \gamma_{10}^2)},$$

$$\begin{aligned}
a_{01} &= \frac{1}{2} \sqrt{\gamma_{00} - \sqrt{\gamma_{00}^2 - 4(\gamma_{01}^2 + \gamma_{10}^2 - 4\gamma_{11}^2)} - \sqrt{-2\gamma_{00}\sqrt{\gamma_{00}^2 - 4(\gamma_{01}^2 + \gamma_{10}^2 - 4\gamma_{11}^2)} + 2\gamma_{00}^2 - 4(\gamma_{01}^2 + \gamma_{10}^2)}, \\
a_{10} &= \frac{1}{2} \sqrt{\gamma_{00} - \sqrt{\gamma_{00}^2 - 4(\gamma_{01}^2 + \gamma_{10}^2 - 4\gamma_{11}^2)} + \sqrt{-2\gamma_{00}\sqrt{\gamma_{00}^2 - 4(\gamma_{01}^2 + \gamma_{10}^2 - 4\gamma_{11}^2)} + 2\gamma_{00}^2 - 4(\gamma_{01}^2 + \gamma_{10}^2)}, \\
a_{11} &= \frac{1}{2} \sqrt{\gamma_{00} + \sqrt{\gamma_{00}^2 - 4(\gamma_{01}^2 + \gamma_{10}^2 - 4\gamma_{11}^2)} + \sqrt{2\gamma_{00}\sqrt{\gamma_{00}^2 - 4(\gamma_{01}^2 + \gamma_{10}^2 - 4\gamma_{11}^2)} + 2\gamma_{00}^2 - 4(\gamma_{01}^2 + \gamma_{10}^2)}.
\end{aligned}$$

**Remark 4.6.** *If, in contrast to the setting of Theorem 4.4,  $\theta$  is not irreducible, then the fiber of  $\gamma$  under  $\Gamma_q$  consists of more than two preimages, as illustrated in the previous example. By (4.1), the maximum number of preimages is  $2^{q_1 + \dots + q_d}$  and occurs exactly when  $\theta$  is completely separable, that is, a product of linear forms.*

## 5 Parameter Estimation

In this section we go one step further and consider the problem of parameter estimation from observed sample points. We consider two methods: least squares estimation and maximum likelihood estimation. Both involve solving polynomial systems of equations. Algebraically, the computational complexity of the estimation problem is measured by the *ED degree* [9] of the associated variety in the first case and by the *ML degree* [1, 7] in the second.

### 5.1 Least squares estimation

Let  $(Y_t)_{t \in \mathbb{Z}^d}$  be a  $MA(q)$  random field, which by Definition 2.1 has mean zero. If we are given observations of  $Y$  on a lattice  $L = \{1, \dots, n\}^d$ , we can estimate the autocovariance function  $\gamma(t)$  by the empirical autocovariance estimator

$$\hat{\gamma}_n(t) := \frac{1}{|B_{n,t}|} \sum_{s \in B_{n,t}} Y(t+s)Y(s), \quad t \in [-q, q], \quad t \geq 0,$$

where

$$B_{n,t} := \{s \in \mathbb{Z}^d \mid s, s+t \in L\} \text{ and } |B_{n,t}| = \prod_{i=1}^d (n - |t_i|) \mathbb{1}_{\{|t_i| \leq n\}}.$$

If  $\hat{\gamma}_n(t)$  were exact values, we would be in the situation of the previous section. However, these are just numerical estimates which form a point  $\hat{\gamma}_n$  that almost surely lies outside the model  $\mathcal{MA}_q$ . One approach is to project the estimated vector  $\hat{\gamma}_n$  onto the autocovariance variety  $\mathcal{MA}_q$ , that is, obtaining  $\gamma_n^* \in \mathcal{MA}_q$  which has the smallest Euclidean distance to  $\hat{\gamma}_n$ :

$$\gamma_n^* := \operatorname{argmin}_{\gamma \in \mathcal{MA}_q} \|\gamma - \hat{\gamma}_n\|. \quad (5.1)$$

The number of critical points of this least squares optimization problem is counted by the Euclidean distance degree (*ED degree*).

**Proposition 5.1.** *The ED degree of  $\mathcal{MA}_q$  is 1 if  $d = 1$ . The ED degree of  $\mathcal{MA}_{(1,1)}$  is 16.*

*Proof.* The first part is a consequence of Proposition 3.2. In fact, for  $d = 1$  the unique critical point for (5.1) is  $\gamma_n^* = \hat{\gamma}_n$ . For the second we use the following M2 code:

```
R = QQ[g00,g01,gm11,g10,g11]
I = ideal(g01^2*gm11^2-g00*g01*gm11*g10+g01^2*g10^2+gm11^2*g10^2+
g00^2*gm11*g11-2*g01^2*gm11*g11-4*gm11^3*g11-g00*g01*g10*g11
-2*gm11*g10^2*g11+g01^2*g11^2+8*gm11^2*g11^2+g10^2*g11^2-4*gm11*g11^3)
sing = ideal singularLocus I
u = {5,7,13,11,3};
M = (matrix{apply(# gens R,i->(gens R)_i-u_i)}|(transpose(jacobian I)));
time J = saturate(I + minors(2,M), sing);
dim J, degree J
```

The vector  $u$  represents a generic choice of  $\gamma$  and the saturation is needed to remove the critical points that lie in the singular locus.  $\square$

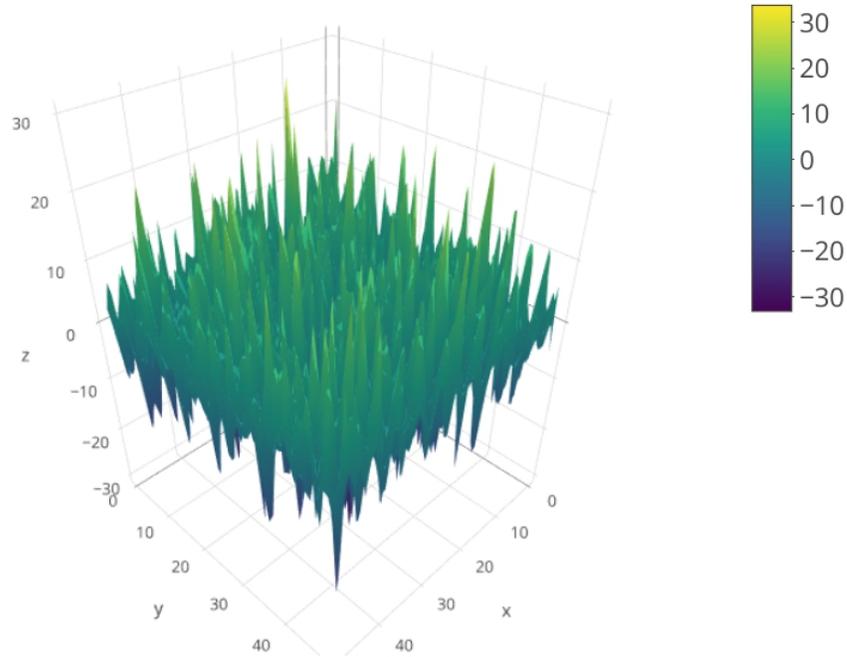


Figure 1:  $MA(1,1)$  random field in Example 5.2

We illustrate with an example:

**Example 5.2.** We simulate 2500 points of a  $MA(1,1)$  random field on a  $50 \times 50$  grid in  $R$  (see Figure 1). As white noise we take an *i.i.d.* standard Gaussian random field. The

moving average parameters are chosen as

$$a_{00} = 7, \quad a_{01} = -5, \quad a_{10} = 3, \quad a_{11} = 1$$

and the corresponding autocovariances values are

$$\gamma = (\gamma_{00}, \gamma_{01}, \gamma_{10}, \gamma_{11}, \gamma_{1-1}) = (84, 16, -32, 7, -15).$$

After centering the sample, we compute the empirical autocovariances

$$\hat{\gamma}_n = (\hat{\gamma}_n(0, 0), \hat{\gamma}_n(0, 1), \hat{\gamma}_n(1, 0), \hat{\gamma}_n(1, 1), \hat{\gamma}_n(1, -1)).$$

By Proposition 5.3, we expect 16 complex critical points, and we compute them numerically. Six of them are real

$$\begin{aligned} & (87.1147, 18.6511, -33.4739, 5.78808, -17.312), \\ & (80.8137, 30.7661, -23.1126, -3.96875, -28.7833), \\ & (61.9284, -24.7157, -16.0001, 1.76548, 19.994), \\ & (55.2165, 8.80716, 26.5528, 0.977029, 8.45708), \\ & (71.9207, -7.85594, -8.51067, 35.9693, 0.649541), \\ & (63.1632, -18.9463, -12.5151, 0.0189543, 24.6219). \end{aligned}$$

The first line has the lowest Euclidean distance to the estimated point

$$\hat{\gamma}_n = (86.6439, 19.1877, -34.2433, 6.6726, -17.3195),$$

and therefore

$$\gamma_n^* = (87.1147, 18.6511, -33.4739, 5.78808, -17.312).$$

Moreover, we have that

$$\|\gamma - \hat{\gamma}_n\| = 5.2604 \quad \text{and} \quad \|\gamma - \gamma_n^*\| = 5.0711,$$

so that projecting onto the autocovariance variety improves the empirical estimate.

$q$	ED degree
(1,1)	16
(1,2)	169
(1,3)	1600
(1,4)	14641

Table 2: ED degree of the variety  $\mathcal{MA}_q$  with  $q = (1, k)$ ,  $k = 1, 2, 3, 4$

The computation of the ED degree for  $\mathcal{MA}_{(1,k)}$  with  $k > 1$  is harder than for  $\mathcal{MA}_{(1,1)}$ . We therefore resort to numerical methods and obtain Table 2 above. The computations suggest the following pattern.

**Conjecture 5.3.** *The ED degree of  $\mathcal{MA}_{(1,k)}$  equals  $\frac{(3^{k+1}-1)^2}{4}$  for all  $k > 0$ .*

Note that the optimization problem (5.1) gives a point in  $\mathcal{MA}_q$  and not a corresponding  $a^* \in \mathbb{P}^Q$ . Theoretically, one could apply the identifiability results of the last section to obtain such  $a^*$  by  $a_n^* := \Gamma_q^{-1}(\gamma_n^*)$ . However, since  $\gamma_n^*$  will most often be a numerical approximation, this is not feasible in practice. Instead, one should solve the optimization problem in parametrized form:

$$a_n^* := \operatorname{argmin}_{a \in \Theta} \|\gamma_a - \hat{\gamma}_n\|,$$

where  $\Theta \subseteq \mathbb{R}^{Q+1}$  is a compact parameter space. We note then that the ED degree gets multiplied by the algebraic identifiability degree of the model parametrization.

## 5.2 Maximum likelihood estimation

Suppose as before that  $(Y_t)_{t \in \mathbb{Z}^d}$  is a  $MA(q)$  random field with mean zero. Furthermore, we assume that  $n$  observations  $Y(t^1), \dots, Y(t^n)$  are given, where the vectors  $t^1, \dots, t^n \in \mathbb{Z}^d$  are ordered according to the lexicographic order. If the driving white noise  $(Z_t)_{t \in \mathbb{Z}^d}$  is Gaussian, then the vector  $Y := (Y(t^1), \dots, Y(t^n))^T$  is Gaussian as well, and its likelihood is of the form

$$L(a) \propto |\Sigma|^{-1/2} \exp\left(-\frac{1}{2} Y^T \Sigma^{-1} Y\right),$$

where  $\Sigma = \Sigma(a)$  is the covariance matrix of  $Y$  and  $|\Sigma|$  its determinant. The *maximum likelihood estimator (MLE)* is then defined as the value which maximizes the log-likelihood:

$$\hat{a}_n := \operatorname{argmax}_{a \in \Theta} -\frac{1}{2} \log(|\Sigma|) - \frac{1}{2} Y^T \Sigma^{-1} Y, \quad (5.2)$$

where  $\Theta \subseteq \mathbb{R}^{Q+1}$  is a compact parameter space. If  $Z$  is not Gaussian, then the latter estimator is called the *quasi maximum likelihood estimator (QMLE)*.

**Remark 5.4.** *In [19] it was shown that under mild assumptions including an invertibility condition, the QMLE  $\hat{a}_n$  is consistent as  $n$  tends to infinity. Furthermore, a slightly modified version of  $\hat{a}_n$  (to account for the edge effect) is shown to be asymptotically normal in [19, Theorem 2].*

Conveniently, the optimization problem (5.2) is *still* algebraic, in the sense that the critical or score equations form a system of rational functions of  $a \in \Theta$ . The number of critical points of the log-likelihood is invariant under generic data  $Y$  and this is known as the maximum likelihood degree (*ML degree*). In terms of likelihood, this is the same definition as ML degree for a discrete finite model. However, we avoid saying ‘ML degree of the autocovariance variety’ to avoid confusion with the discrete case, where the model is identified by a set of vectors in probability coordinates and hence one can also talk of the ML degree of the associated algebraic variety, as in [13].

We analyze the first nontrivial case, when  $q = 1$  and  $n = 2$ . Even this simple model is interesting. It has been observed that the MLE can sometimes correspond to non-invertible models, which in this case is equivalent to  $|a_0| = |a_1|$ , and contrary to what was previously thought, this occurs with positive probability [8].

**Proposition 5.5.** Consider the MA(1) model with observed sample  $Y = (Y_1, Y_2)$ . The ML degree is 4, and these four critical points can be divided into three groups:

(1) The parameters  $a_0$  and  $a_1$  satisfy the two equations

$$a_0 a_1 = Y_1 Y_2 \quad \text{and} \quad a_0^2 + a_1^2 = \frac{Y_1^2 + Y_2^2}{2}.$$

(2)

$$a_0 = a_1 = \sqrt{\frac{Y_1^2 + Y_2^2 - Y_1 Y_2}{3}}$$

(3)

$$a_0 = -a_1 = \sqrt{\frac{Y_1^2 + Y_2^2 + Y_1 Y_2}{3}}$$

If  $Y_1 Y_2 = 0$ , then the MLE corresponds to a degenerate model ( $a_0 a_1 = 0$ ). Otherwise let  $W = \frac{Y_1^2 + Y_2^2}{2Y_1 Y_2}$  and the MLE is given as:

- the point in (3) if  $-2 < W < 0$
- the point in (2) if  $0 < W < 2$
- the points in (1) otherwise.

*Proof.* Since we have a MA(1) process and  $Y = (Y_1, Y_2)$ , we have  $\Sigma = \begin{pmatrix} \gamma(0) & \gamma(1) \\ \gamma(1) & \gamma(0) \end{pmatrix}$ , and the log-likelihood takes the form

$$\ell(a_0, a_1) = -\frac{1}{2} \log((a_0^2 + a_1^2)^2 - a_0^2 a_1^2) - \frac{1}{2} (Y_1, Y_2) \begin{pmatrix} a_0^2 + a_1^2 & a_0 a_1 \\ a_0 a_1 & a_0^2 + a_1^2 \end{pmatrix}^{-1} (Y_1, Y_2)^\top. \quad (5.3)$$

There are generically four solutions to the system  $\frac{\partial \ell}{\partial a_0} = \frac{\partial \ell}{\partial a_1} = 0$ . This means the ML degree is 4. The critical points can be divided into the three groups (1), (2) and (3) of the statement. In order to find the MLE depending on the values of  $Y_1, Y_2$ , we evaluate the likelihood function  $\ell$  at these 3 groups of points. In fact, substituting  $a_0$  and  $a_1$  from (1), (2) and (3) into the log-likelihood function, we obtain

- (i)  $-\frac{1}{2} \log((Y_1^2 - Y_2^2)^2) - 1 + \log(2)$ ,
- (ii)  $-\frac{1}{2} \log\left(\frac{1}{3}(Y_1^2 - Y_1 Y_2 + Y_2^2)^2\right) - 1$ ,
- (iii)  $-\frac{1}{2} \log\left(\frac{1}{3}(Y_1^2 + Y_1 Y_2 + Y_2^2)^2\right) - 1$ .

Computing (i) - (iii) gives the expression

$$\frac{1}{2} \left( \log \left( 4(Y_1^2 - Y_1 Y_2 + Y_2^2)^2 \right) - \log \left( 3(Y_1^2 - Y_2^2)^2 \right) \right),$$

which is always nonnegative since

$$4(Y_1^2 - Y_1 Y_2 + Y_2^2)^2 - 3(Y_1^2 - Y_2^2)^2 = (Y_1^2 - 4Y_1 Y_2 + Y_2^2)^2 \geq 0.$$

Analogously, (i) is greater than or equal to (ii) from  $(Y_1^2 + 4Y_1 Y_2 + Y_2^2)^2 \geq 0$ .

Hence, the first value (i) is always larger than or equal to the values (ii) and (iii), independently of  $Y_1$  and  $Y_2$ . We would conclude that the maximizers are always given by (1), but the points may not be real. Indeed, under (1), if

$$a_0 a_1 > 0 \text{ then } \frac{a_0^2 + a_1^2}{a_0 a_1} > 2 \text{ and thus } W = \frac{Y_1^2 + Y_2^2}{2Y_1 Y_2} > 2$$

while

$$a_0 a_1 < 0 \text{ implies } \frac{a_0^2 + a_1^2}{a_0 a_1} < -2 \text{ and hence } W < -2.$$

Direct inspection reveals that the likelihood for (2) is larger than the one for (3) if and only if  $W > 0$ . Note that when  $W = -2$  the points (1) and (3) coincide, while  $W = 2$  means that (1) and (2) coincide.  $\square$

Compare our conditions for  $W$  with the similar ones found by [8] in their effort of computing the distribution of the MLE in this  $q = 1, n = 2$  case (note the different parametrization in terms of  $\sigma, \theta$ ). Furthermore, it is gratifying to see that our computations provide a simple explanation for the ‘curious’ phenomenon that the MLE can belong to a non-invertible model. Algebraically, the points in (1) always maximize the likelihood, but for the specified region of  $Y_1, Y_2$  these points are strictly complex (even though evaluating at the likelihood yields real values!), which means then that (2) or (3) becomes the MLE.

In [20], standard numerical optimization routines were used to find the MLE in samples of  $MA(q)$  models with  $q = 1, 2, 3, 4$ . The simulations show the MLE can again lie on the non-invertible boundary.

**Example 5.6.** Consider a  $MA(1)$  process with  $n = 3$  sample points  $Y = (Y_1, Y_2, Y_3)$ . The ML degree is now 8. The expressions for the two non-invertible models  $|a_0| = |a_1|$  are:

$$a_0 = a_1 = \sqrt{\frac{3x_1^2 + 4x_2^2 + 3x_3^2 - 4x_1 x_2 + 2x_1 x_3 - 4x_2 x_3}{12}}$$

$$a_0 = -a_1 = \sqrt{\frac{3x_1^2 + 4x_2^2 + 3x_3^2 + 4x_1 x_2 + 2x_1 x_3 + 4x_2 x_3}{12}}$$

Obtaining closed form expressions for the other 6 critical points is also possible.

For  $n > 2$ , the matrix  $\Sigma$  is tridiagonal: it has  $\gamma(0)$  in the diagonal and  $\gamma(1)$  in the upper and lower diagonal. Our ML degree computations of  $MA(1)$  for  $n = 2, 3, \dots$  reveal the following pattern:

**Conjecture 5.7.** *The ML degree of  $MA(1)$  for  $n > 1$  sample points is equal to  $4(n - 1)$ .*

In contrast, the pattern for  $MA(2)$  is not as clear. The first values for  $n = 3, 4, \dots$  are recorded in Table 3.

$n$	ML degree
3	29
4	69
5	129
6	205

Table 3: ML degrees of the  $MA(2)$  model by number of sample points

Not unusually, Gröbner basis computations quickly become prohibitive. However, this does not mean that our algebraic approach is not useful. In applied algebraic geometry, this often means one needs to go into numerical techniques. Indeed, as far as we know, the algebraic nature of the ML problem has not been exploited yet, and a numerical algebraic geometry approach brings both a fresh perspective and efficient computational tools. Knowing the ML degree beforehand helps homotopy continuation and monodromy methods find all solutions to the critical equations and thus guarantee that the MLE will be found. In contrast, classical local search methods may only find a local maximum of the likelihood function. One way to compare these methods is to conduct simulation studies such as the one in the next example.

**Example 5.8.** *We simulate 500 independent paths of a  $MA(1)$  process with  $n = 8$  observations for each path. A sample path for this process is illustrated in Figure 2.*

*As moving average parameters we take*

$$a_0 = 1 \quad \text{and} \quad a_1 = 0.5.$$

*The  $MA(1)$  process is driven by i.i.d. standard Gaussian noise. Having simulated the  $MA(1)$  process, we proceed by estimating the model parameters with the MLE in (5.2) in two different ways.*

*For our first approach we use the standard R command `optim` for minimization of the objective function. As it is standard in time series analysis, we take the output of the innovations algorithm as the initial value for the optimization routine (cf. [5, Section 2]).*

*For our second approach we differentiate the likelihood function with respect to the moving average parameters and set the derivatives to zero. In order to compute the critical points of the likelihood function we solve the resulting polynomial system using homotopy continuation. This is implemented in the `julia` package **HomotopyContinuation** [4].*

*Finally, we evaluate the likelihood at every critical point and choose the maximal one. The summary of the estimation results are given in Tables 4 and 5 below.*

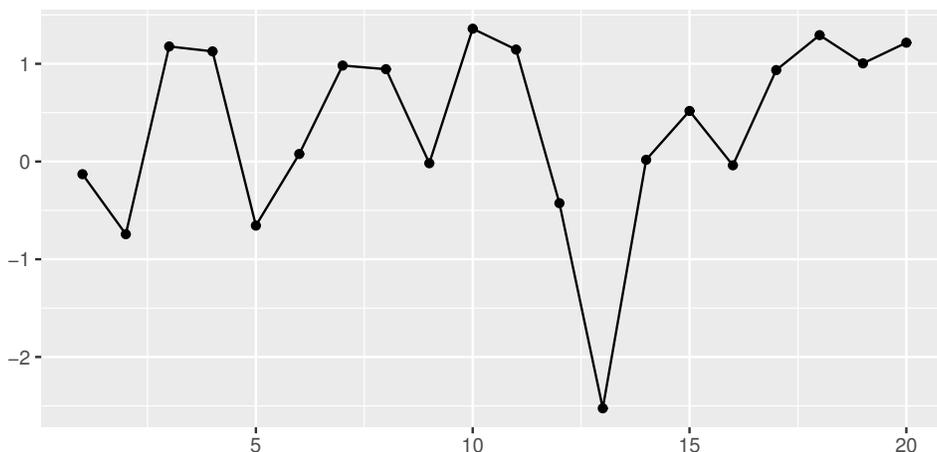


Figure 2: Sample path of a  $MA(1)$  process.

We observe that using homotopy continuation reduces the bias for both  $a_0$  and  $a_1$ , whereas it increases the standard deviation for  $a_1$  and decreases the standard deviation for  $a_0$ .

	True Value	Mean	Bias	Std
$a_0$	1.0000	0.8642	-0.1358	0.2459
$a_1$	0.5000	0.4503	-0.0497	0.4071

Table 4: Parameter estimation results for  $MA(1)$  with  $n = 8$  and R command `optim`.

	True Value	Mean	Bias	Std
$a_0$	1.0000	0.8818	-0.1182	0.2129
$a_1$	0.5000	0.4678	-0.0322	0.5094

Table 5: Parameter estimation results for  $MA(1)$  with  $n = 8$  and homotopy continuation.

Finally, we close this section by reporting the ML degree of  $MA(1, 1)$ :

**Proposition 5.9.** *Assume that  $n = 4$  sample points  $Y = (Y_{11}, Y_{12}, Y_{21}, Y_{22})$  over the lattice  $L = \{1, 2\}^2$  of a  $MA(1, 1)$  random field are given. The autocovariance matrix  $\Sigma$  of  $Y$  is*

$$\Sigma = \begin{pmatrix} \gamma_{00} & \gamma_{01} & \gamma_{10} & \gamma_{11} \\ \gamma_{01} & \gamma_{00} & \gamma_{1-1} & \gamma_{10} \\ \gamma_{10} & \gamma_{1-1} & \gamma_{00} & \gamma_{01} \\ \gamma_{11} & \gamma_{10} & \gamma_{01} & \gamma_{00} \end{pmatrix}.$$

The ML degree of the model is 192 over  $\mathbb{P}^3$ .

Further study of the autocovariance varieties of these models with their ED and ML degrees, as well as robust numerical (algebraic) estimation of their parameters is a natural direction for future work.

**Acknowledgments.** We are grateful to Claudia Klüppelberg for her support during this project. We thank Daniele Agostini for helpful conversations. Carlos Améndola was partially supported by the Deutsche Forschungsgemeinschaft (DFG) in the context of the Emmy Noether junior research group KR 4512/1-1. Viet Son Pham acknowledges support from the graduate program TopMath at the Technical University of Munich and the Studienstiftung des deutschen Volkes.

## References

- [1] C. Améndola, N. Bliss, I. Burke, C. R. Gibbons, M. Helmer, S. Hoşten, E. D. Nash, J. I. Rodríguez, and D. Smolkin. The maximum likelihood degree of toric varieties. *Journal of Symbolic Computation*, 92:222 – 242, 2019.
- [2] C. Améndola, J.C. Faugère, and B. Sturmfels. Moment varieties of Gaussian mixtures. *Journal of Algebraic Statistics*, 7(1):14–28, 2016.
- [3] C. Améndola, K. Ranestad, and B. Sturmfels. Algebraic identifiability of Gaussian mixtures. *International Mathematics Research Notices*, 2018(21):6556–6580, 2018.
- [4] P. Breiding and S. Timme. Homotopycontinuation. jl: A package for homotopy continuation in julia. In *International Congress on Mathematical Software*, pages 458–465. Springer, 2018.
- [5] P.J. Brockwell and R.A. Davis. Simple consistent estimation of the coefficients of a linear filter. *Stoch. Process. Appl.*, 28(1):47–59, 1988.
- [6] P.J. Brockwell and R.A. Davis. *Time Series: Theory and Methods*. Springer, New York, 2nd edition, 1991.
- [7] F. Catanese, S. Hoşten, A. Khetan, and B. Sturmfels. The maximum likelihood degree. *American Journal of Mathematics*, 128(3):671–697, 2006.
- [8] J.D. Cryer and J. Ledolter. Small-sample properties of the maximum likelihood estimator in the first-order moving average model. *Biometrika*, 68(3):691–694, 1981.
- [9] J. Draisma, E. Horobeţ, G. Ottaviani, B. Sturmfels, and R. Thomas. The euclidean distance degree of an algebraic variety. *Foundations of computational mathematics*, 16(1):99–149, 2016.
- [10] M. Drapatz. Strictly stationary solutions of spatial ARMA equations. *Ann. Inst. Stat. Math.*, 68(2):385–412, 2016.
- [11] J. M. Francos, A. Narasimhan, and J. W. Woods. Maximum likelihood parameter estimation of textures using a wold decomposition based model. *IEEE Trans. Image Processing*, 4:1655–1666, 1995.
- [12] D. R. Grayson and M. E. Stillman. Macaulay2, a software system for research in algebraic geometry. Available at <http://www.math.uiuc.edu/Macaulay2/>.

- [13] J. Huh. The maximum likelihood degree of a very affine variety. *Compositio Mathematica*, 149(8):1245–1266, 2013.
- [14] R. Krishnamurthy, J. W. Woods, and J. M. Francos. Adaptive restoration of textured images with mixed spectra using a generalized Wiener filter. *IEEE Trans. Image Processing*, 5:648–652, 1996.
- [15] D. Mumford. *Algebraic geometry I: complex projective varieties*. Springer Science & Business Media, 1995.
- [16] V.S. Pham. Lévy-driven causal CARMA random fields. 2018. Submitted. arXiv:1805.08807.
- [17] A. H. Sayed and T. Kailath. A survey of spectral factorization methods. *Numerical linear algebra with applications*, 8(6-7):467–496, 2001.
- [18] S. Sullivant. *Algebraic Statistics*. Graduate Studies in Mathematics. American Mathematical Society, 2018.
- [19] Q. Yao and P.J. Brockwell. Gaussian maximum likelihood estimation for ARMA models II: Spatial processes. *Bernoulli*, 12(3):403–429, 2006.
- [20] Y. Zhang and A.I. McLeod. Fitting MA(q) models in the closed invertible region. *Stat. Probab. Letters*, 76(13):1331–1334, 2016.

**Authors' addresses:**

Zentrum Mathematik, Technische Universität München  
 Boltzmannstrasse 3, 85748 Garching (b. München), Germany  
 carlos.amendola@tum.de , vietson.pham@tum.de