ON BLIND CHANNEL IDENTIFICATION AND EQUALIZATION OVER GALOIS FIELDS

Arie Yeredor

School of Electrical Engineering, Tel-Aviv University
arie@eng.tau.ac.il

ABSTRACT

We consider the problem of blind identification and equalization of a Linear, Time-Invariant (LTI) system, where the input and output signals, as well as the linear operations, all reside in a finite (Galois) field. We point out some fundamental differences from the classical version of this problem. We show that if the input process is a sequence of independent, identically distributed random variables, the system is identifiable if and only if the (marginal) distribution of the input is non-uniform. For an autoregressive (AR) channel a finite impulse response equalizer can be found by minimizing the marginal entropy of its output signal. However, an exhaustive search for the minimizing equalizer, although theoretically possible, is not necessary: Based on somewhat surprising properties of the AR channel's output (not shared by the classical case), we show that the equalizer can be found directly from the empirical characteristic tensor of this output. We demonstrate the success rate of the proposed methods in simulation.

Index Terms— Blind Equalization; System Identification; Moving Average; Autoregressive; Finite Fields; Galois Fields.

1. INTRODUCTION

Blind channel estimation and equalization over the real or complex valued fields is a well-studied topic with applications in diverse fields, such as communications, speech dereverberation or seismology (see, e.g., [1, 2] and references therein), to name just a few. In the basic classical model, an unobserved source signal, modeled as a random process with independent, identically distributed (iid) samples, is presented at the input of a linear, time-invariant (LTI) system (channel). It is desired to blindly estimate the channel's parameters (up to an acceptable scaling ambiguity), usually for the purpose of equalization. The blindness implies that no additional information is available on the channel or on the source signal, except for its iid time structure. This problem is closely related to the problem of Independent Component Analysis (ICA), in which a linear mixture of independent sources is observed, and it is desired to blindly estimate the mixing matrix (usually for separating the sources).

In recent years some interest has been taken in considering the ICA problem in other algebraic fields than the real or complex fields. The idea was first considered in 2007 [3] in a boolean algebraic framework, where binary sources were mixed by an "exclusive OR" (XOR) operation. This new paradigm was later expanded to general finite (Galois) fields of prime order P (denoted $\mathbb{GF}(P)$) in [4] and of general (prime power) orders in [5, 6]. While no immediate practical applications were associated with this problem at first, theoretical applications have been suggested in the context of eavesdropping on a Tomlinson-Harashima coded MIMO channel [4, 6] and in the context of Network Coding [7]. Some additional contributions to this emerging topic were recently published in [8, 9, 10, 11, 12].

An additional natural extension of ICA over finite fields can be considered in the context of blind channel identification and equalization over finite fields. Admittedly, like in its ICA counterpart, the current span of prospective applications is rather limited. The main reason is that LTI systems over finite fields, unlike LTI system over real or complex fields, are rarely met in nature. Nevertheless, there are possible contexts in which man-made LTI operations over finite fields are applied - such as in convolution coding or in network coding over networks with a cyclic topology. In addition, there are interesting theoretical aspects to this problem, and as we shall show in this paper, some of the basic principles are essentially different from their classical counterparts in the real or complex fields on one hand, and from principles of ICA over finite fields on the other hand.

Some rudimentary treatment of the problem of filtering over Galois fields has recently appeared in [13], and also in [14] as extensions of ICA to convolutive mixtures over such fields. However, the aspects of blind channel identification or equalization in these papers was limited to heuristic ideas and to empirical testing, mainly using exhaustive search procedures. In this paper we take a closer theoretical look at to blind identification and equalization over finite fields. We provide some fundamental theoretical results regarding the statistical distribution of the channel's output for rational (moving average (MA) / autoregressive (AR)) channels and discuss identifiability conditions. While classical tools of second or higher order statistics are irrelevant in the context of finite fields, one of the remaining keys for identification, separation and equalization is the (marginal) entropy of the signals. As we explain later on, the entropy (which is easy to estimate in a finite field) of an equalizer's output signal can serve as a criterion for the success of the attempted equalization. Since the number of possible equalizers of a given length is finite, an exhaustive search for the optimal equalizer is theoretically possible (and, indeed, is advocated in [13, 14]). However, we show that at least for the equalization of an AR channel, such an exhaustive search is not necessary, since it is possible to take advantage of specific properties of LTI AR filtering so as to obtain a direct estimate of the equalizer.

For simplicity of the exposition, we shall limit the discussion to finite fields of prime order P, where all the arithmetics in the field are applied modulu P. When referring to such arithmetics, we shall denote addition, subtraction and multiplication by $\oplus$, $\ominus$ and $\otimes$, respectively; The symbol $\sum^{\circ}$ will be used to denote summation; Vector and matrix multiplications will be denoted by $\circ$, e.g., $\boldsymbol{A} \circ \boldsymbol{b}$.

2. STATISTICAL PRELIMINARIES: RANDOM VARIABLES AND RANDOM VECTORS IN $\mathbb{GF}(P)$

A random variable (RV) u in $\mathbb{GF}(P)$ is characterized by a discrete probability distribution, fully described by a *probability vector* $\boldsymbol{p}_u = [p_u(0)\ p_u(1)\ \cdots\ p_u(P-1)]^T \in \mathbb{R}^P$, whose elements are the prob-

978-1-4799-2893-4/14/$31.00 ©2014 IEEE

abilities of u taking the respective values. An RV is called *uniform* if it takes all values with equal probability $\frac{1}{P}$, and *degenerate* if it deterministic, namely, if it takes a particular value with probability 1. The *entropy* of u is given by $H(u) = -\sum_{m=0}^{P-1} p_u(m) \log p_u(m)$. By maximizing the entropy with respect to $\boldsymbol{p}_u$, it is easy to show that among all random variables in $\mathbb{GF}(P)$, the uniform random variable has the largest entropy, given by $\log P$.

The *characteristic vector* of u is denoted $\tilde{\boldsymbol{p}}_u = [\tilde{p}_u(0) \ \tilde{p}_u(1) \ \cdots \ \tilde{p}_u(P-1)]^T \in \mathbb{C}^P$, and its elements are given by the discrete Fourier transform (DFT) of the elements of $\boldsymbol{p}$:

$$\tilde{p}_u(n) = E[W_P^{nu}] = \sum_{m=0}^{P-1} p_u(m) W_P^{mn} \quad n = 0, \ldots, P-1, \quad (1)$$

where the "twiddle factor" W_P is defined as $W_P = e^{-j2\pi/P}$ (note that the modulu-P operation is inherently present in the exponential part, so W_P^{mn} is equivalent to $W_P^{m \otimes n}$). Like the probability vector $\boldsymbol{p}_u$, the characteristic vector $\tilde{\boldsymbol{p}}_u$ provides full statistical characterization of the random variable u, since $\boldsymbol{p}_u$ can be directly obtained from $\tilde{\boldsymbol{p}}_u$ using the inverse DFT.

The following basic properties of $\tilde{\boldsymbol{p}}_u$ can be easily shown:

- P1. $\tilde{p}_u(0) = 1$;
- P2. Since $\boldsymbol{p}_u$ is real-valued, $\tilde{p}_u(n) = \tilde{p}_u^*(P-n)$ (where the superscript * denotes the complex-conjugate);
- P3. u is uniform $\Leftrightarrow \tilde{p}_u(n) = 0 \ \forall n \neq 0$;
- P4. u is degenerate $\Leftrightarrow \tilde{p}_u(n) = W_P^{nM} \ \forall n$;
- P5. $|\tilde{p}_u(n)| \leq 1 \ \forall n$, where for $n \neq 0$ equality holds if and only if (iff) u is degenerate.

The characteristic vector of the sum of two statistically independent RVs is given by the element-wise product of their characteristic vectors, since if $w = u \oplus v$, then for $n = 0, \ldots, P-1$

$$\tilde{p}_w(n) = E[W_P^{n(u+v)}] = E[W_P^{nu}] E[W_P^{nv}] = \tilde{p}_u(n) \tilde{p}_v(n). \quad (2)$$

For a $K \times 1$ random vector (RVec) $\boldsymbol{u}$ whose elements $u_1, \ldots, u_K$ are RVs in $\mathbb{GF}(P)$, the joint statistics are fully characterized by the K-dimensional *probability tensor* (matrix for $K = 2$) $\boldsymbol{\mathcal{P}}_{\boldsymbol{u}} \in \mathbb{R}^{P^{(\times K)}}$, whose elements are the probabilities $\mathcal{P}_{\boldsymbol{u}}(m_1, \ldots, m_K) = \Pr\{u_1 = m_1, \ldots, u_K = m_K\}$, $m_1, \ldots, m_K \in \{0, \ldots, P-1\}$. Using vector-index notations, where $\boldsymbol{m} = [m_1, \cdots, m_K]^T$, we may also express this relation more compactly as $\mathcal{P}_{\boldsymbol{u}}(\boldsymbol{m}) = \Pr\{\boldsymbol{u} = \boldsymbol{m}\}$. The *characteristic tensor* $\widetilde{\boldsymbol{\mathcal{P}}}_{\boldsymbol{u}} \in \mathbb{C}^{P^{(\times K)}}$ is given by the K-dimensional DFT of $\boldsymbol{\mathcal{P}}_{\boldsymbol{u}}$, which, using a similar index-vector notation, is given by

$$\widetilde{\mathcal{P}}_{\boldsymbol{u}}(\boldsymbol{n}) = E[W_P^{\boldsymbol{n}^T \boldsymbol{u}}] = \sum_{\boldsymbol{m}} \mathcal{P}_{\boldsymbol{u}}(\boldsymbol{m}) W_P^{\boldsymbol{n}^T \boldsymbol{m}}, \quad (3)$$

summing over all possible P^K indices combinations in $\boldsymbol{m}$.

The elements of an RVec are statistically independent iff its probability tensor is the outer product of their probability vectors, namely $\mathcal{P}_{\boldsymbol{u}}(m_1, \ldots, m_K) = p_{u_1}(m_1) \cdot p_{u_2}(m_2) \cdots p_{u_K}(m_K)$. Equivalently, the characteristic tensor is the outer product of the elements' caracteristic vectors. Accordingly, the characteristic tensor of a vector of iid uniform RVs is all-zeros, except for its $\boldsymbol{0}$-th element, which equals 1.

Similarly to the sum of independent RVs, the characteristic tensor of the sum of statistically independent RVecs is given by the element-wise product of their characteristic tensors.

If an RVec $\boldsymbol{u} = \boldsymbol{B} \circ \boldsymbol{v}$ is a linear transformation of another RVec $\boldsymbol{v}$, their characteristic tensors are related by $\widetilde{\mathcal{P}}_{\boldsymbol{u}}(\boldsymbol{n}) = E[W_P^{\boldsymbol{n}^T \boldsymbol{B} \boldsymbol{v}}] = \widetilde{\mathcal{P}}_{\boldsymbol{v}}(\boldsymbol{B}^T \circ \boldsymbol{n})$.

3. RATIONAL (MA AND AR) CHANNELS

Let $s[t] \in \mathbb{GF}(P)$ denote a discrete-time random process (the "source signal") at the input of a general rational LTI (linear over the field) channel. Such a channel's output is given by

$$x[t] = \sum_{m=0}^{M}{}^{\circ} b_m \otimes s[t-m] \ominus \sum_{r=1}^{R}{}^{\circ} a_r \otimes x[t-r], \quad (4)$$

where $b_0, ..., b_M \in \mathbb{GF}(P)$ are the channel's MA coefficients, and $a_1, ..., a_R \in \mathbb{GF}(P)$ are its AR coefficients, with M and R denoting the respective orders. We assume that these orders are "true", namely that b_0, b_M and a_R are non-zeros (other coefficients may or may not be zeros). In addition, to avoid scale ambiguities we shall assume $b_0 = 1$ as a scaling convention.

When the channel is a pure MA channel ($R = 0$), (4) amounts to a Finite Impulse Response (FIR) filtering relation, otherwise (for $R > 0$) the channel generally has an Infinite Impulse Response (IIR). While this is quite similar to the familiar real-valued (or complex-valued) counterparts, a fundamental difference in the IIR case in $\mathbb{GF}(P)$ is the absence of any stability issues on one hand, and the non-decaying nature of the filter's impulse response on the other hand. In other words, since the values are taken over a finite field, the channel's output is always "bounded", regardless of the specific AR coefficients, so such a channel is always "stable" and its impulse-response never "explodes"; however, it never decays, either: while a stable linear IIR channel in the real- or complex-valued case can always be approximated (to arbitrary precision) by a "sufficiently long" FIR filter, such an approximation is impossible in the finite-field framework, because there is no notion of "small" (or "large") numbers in a finite field, and hence the impulse response cannot be truncated while maintaining any tolerable approximation.

Nevertheless, such a channel can always be inverted by switching the roles of the MA and AR coefficients, since the same equation can also be written as

$$s[t] = \sum_{r=0}^{R}{}^{\circ} a_r \otimes x[t-r] \ominus \sum_{m=1}^{M}{}^{\circ} b_m \otimes s[t-m], \quad (5)$$

with $a_0 = 1$, so with $x[t]$ available for $t = 0, 1, ...$, and with known (e.g., zero) initial conditions for both $x[-1], x[-2], ..., x[-R]$ and $s[-1], s[-2], ..., s[-M]$, the source $s[t]$ can be fully recovered if the AR and MA coefficients are known.

Thus, when the MA and AR coefficients are unknown, the ultimate goal is to estimate these coefficients from observation of the output $x[t]$ for $t = 0, ..., T-1$ with some "sufficiently large" observation length T. The estimated coefficients may in turn be used to recover the source signal $s[t]$.

Naturally, without any prior knowledge regarding $s[t]$, such an estimation problem is ill-posed. To enable a solution, we shall adopt the common practice of classical blind channel estimation, and assume only that $s[t]$ is a sequence of iid random variables (but their specific probability distribution is not assumed to be known to the estimator). One possible key to estimating the coefficients is the observation that if the random variables $s[t]$ are non-degenerate and non-uniform, the entropy of $x[t]$ must be larger than that of $s[t]$. This is a direct consequence of the fact that the entropy of the sum (over $\mathbb{GF}(P)$) of any two independent random variables u and v is always larger than or equal to the entropy of each, namely $H(u \oplus v) \geq H(u), H(u)$, so that $H(u \oplus v) \geq \max\{H(u), H(v)\}$, where equality holds if and only if at least one of these variables is uniform or degenerate (see [4], Lemma 3). Since multiplication

by a nonzero constant over the field is bijective and therefore does not change the entropy, it follows that if $s[t]$ is iid, non-degenerate and non-uniform, the entropy of any linear combination of two or more samples of $s[t]$ is larger than the entropy of $s[t]$. Thus, when using any estimates of the coefficients $\hat{a}_1, ..., \hat{a}_R, \hat{b}_1, ..., \hat{b}_M$ to form an equalizer's output $y[n]$ as

$$y[t] = \sum_{r=0}^{R}{}^{\circ}\, \hat{a}_r \otimes x[t-r] \ominus \sum_{m=1}^{M}{}^{\circ}\, \hat{b}_m \otimes y[t-m] \qquad (6)$$

(with $\hat{a}_0 = 1$), the resulting entropy of $y[t]$ would always be larger than or equal to that of $s[t]$, and equality will hold if and only if the estimated parameters equal the true parameters. This is because the substitution of any other parameters in (6) implies a residual filter relating the equalizer's output $y[t]$ to the source signal $s[t]$.

Given M and R, there is a finite set of P^{MR} possible different combinations for the MA and AR parameters (all of which lead to a "legitimate" equalizer, since there are no stability issues), so a theoretically plausible approach (partly advocated in [13, 14]) is to run an exhaustive search over all possible equalizers and pick the one which yields the minimal entropy. In fact, the theoretical possibility of such a strategy leads to an identifiability condition:

Theorem 1. *Let $s[t]$ be a sequence of iid, non-degenerate random variables, and let $x[t]$ be given by* (4) *(with $b_0 = 1$). The parameters $a_1, ..., a_R$, $b_1, ..., b_R$ can be identified from exact knowledge of the joint statistics of the process $x[t]$ iff $s[t]$ is non-uniform.*

It can be shown that when an LTI channel's input is an iid, uniform sequence, the output is also iid and uniform, regardless of the channel coefficients (we omit the proof in here, due to space limitations). This means that if $s[t]$ is uniform, $x[t]$ is also iid and uniform, and therefore might as well be the source signal (and so might any LTI filtered version thereof), so without additional knowledge the channel cannot be identified. Conversely, when $s[t]$ is non-uniform, the channel's coefficient can be found by looking for the equalizer which minimizes the output's entropy (which can be calculated, for any equalizer, from the full statistical description of its input $x[t]$).

However, the direct implementation of an exhaustive search approach (as proposed in [13]) requires to apply each tested equalizer to the entire data sequence for estimation of the resulting output's entropy. This can be prohibitive even for relatively small values of P, M and R, if T is very large (as it should usually be for reliable estimation). It would be much more efficient to keep some estimated statistics of the observed data and to estimate the resulting output entropy of the prospective equalizers' outputs directly from these statistics, thereby relieving the estimator from the need to apply each tested equalizer to the entire signal. As we shall see, a more efficient strategy exists for AR channels (when looking for MA equalizers).

In order to equalize an AR channel of order R, an MA equalizer of the same order is required. Let the coefficients of a prospective equalizer be given by $\hat{\boldsymbol{a}} \triangleq [1\ \hat{a}_1\ \cdots\ \hat{a}_R]^T$, such that its output $y[t]$ is given by (6) with $M = 0$. Given the probability tensor $\boldsymbol{\mathcal{P}}_{\boldsymbol{x}}$ of a vector of $R+1$ consecutive samples $\boldsymbol{x} \triangleq [x[t]\ x[t-1]\ \cdots\ x[t-R]]^T$, the probability vector of $y[t]$ can be found as follows. Noting the relation $y[t] = \hat{\boldsymbol{a}}^T \circ \boldsymbol{x}$, the elements of the characteristic vector of $y[t]$ are given (for $n = 0, ..., P-1$) by

$$\tilde{p}_y(n) = E\left[W_P^{ny[t]}\right] = E\left[W_P^{n\hat{\boldsymbol{a}}^T\boldsymbol{x}}\right] = \widetilde{\mathcal{P}}_{\boldsymbol{x}}(n \otimes \hat{\boldsymbol{a}}) \qquad (7)$$

The characteristic tensor $\widetilde{\boldsymbol{\mathcal{P}}}_{\boldsymbol{x}}$ can be efficiently computed by applying an $(R+1)$-dimensional FFT to the probability tensor $\boldsymbol{\mathcal{P}}_{\boldsymbol{x}}$.

Evidently, the probability tensor $\boldsymbol{\mathcal{P}}_{\boldsymbol{x}}$ is unknown, but can be empirically estimated from the observations $x[0], ..., x[T-1]$ by looking at all $(R+1)$-tuples, counting their occurrences and dividing by their total number (which is $T-K+1$). The resulting estimate of the characteristic vector (hence of the probability vector) of $y[t]$ for a prospective equalizer can then be obtained from (7) with $\widetilde{\mathcal{P}}_{\boldsymbol{x}}$ substituted by its estimate. The estimated probability vector $\hat{\tilde{\boldsymbol{p}}}_y$ can in turn be used for estimating the entropy of the equalizer's output and for selecting the equalizer which minimizes this entropy. Therefore, an exhaustive search among all P^R possible different equalizers can be accomplished without actually applying each prospective equalizer to the channel's output. The resulting performance (in terms of the probability of selecting the correct equalizer) depends on the quality of the estimated $\boldsymbol{\mathcal{P}}_{\boldsymbol{x}}$, as well as on the entropy of $s[t]$, which is the minimum attainable entropy of $y[t]$. If the entropy of $s[t]$ is close to $\log P$ (namely, if $s[t]$ is nearly uniform), errors in estimating $\mathcal{P}_{\boldsymbol{x}}$ are more likely to cause a false minimum to be selected.

Note that unfortunately this strategy cannot be applied to finding the AR equalizer to an MA channel, since the relation $y[t] = \hat{\boldsymbol{a}}^T \circ \boldsymbol{x}$ is replaced with $x[t] = \hat{\boldsymbol{b}}^T \circ \boldsymbol{y}$, and there is no direct access to the characteristic vector of $y[t]$. Nevertheless, it is possible to obtain partial knowledge on the characteristic tensor of the vector $\boldsymbol{y}$, and to look for the best symmetric rank-1 approximation to this tensor, since upon successful equalization the elements of $\boldsymbol{y}$ should be independent and hence their characteristic tensor should admit a symmetric rank-1 decomposition. Unfortunately, we do not have sufficient room in here to further elaborate on this approach.

We now propose an alternative option for finding the MA equalizer to an AR cannel, which does not involve an exhaustive search. The proposal is based on the following properties of the output of an AR channel of order R with a non-degenerate iid input $s[n]$:

Theorem 2. *The asymptotic ($t \to \infty$) distribution of a vector of R consecutive samples $\bar{\boldsymbol{x}}_t \triangleq [x[t]\ x[t-1]\ \cdots\ x[t-R+1]]^T$ is uniform and iid.*

Proof. See the Appendix. □

This is a somewhat surprising and counter-intuitive result, which is in fierce contrast to the classical case of real- (or complex) valued signals: While for a classical AR process all samples are correlated (and hence statistically dependent), in the case of a finite field every R consecutive samples are mutually independent. However, every $R+1$ consecutive samples are generally mutually dependent, as implied by the following Theorem:

Theorem 3. *Denote the coefficients vector of the AR process as $\boldsymbol{a} \triangleq [1\ a_1\ \cdots\ a_R]^T$. The elements of the characteristic tensor $\widetilde{\boldsymbol{\mathcal{P}}}_{\boldsymbol{x}}$ of the asymptotic ($t \to \infty$) distribution of a vector of $R+1$ consecutive samples $\boldsymbol{x}_t \triangleq [x[t]\ x[t-1]\ \cdots\ x[t-R]]^T$ are given by:*

$$\widetilde{\mathcal{P}}_{\boldsymbol{x}}(\boldsymbol{n}) = \begin{cases} \tilde{p}_s(m) & \exists m | \boldsymbol{n} = m \otimes \boldsymbol{a} \\ 0 & \text{otherwise} \end{cases} \qquad (8)$$

Proof. See the Appendix. □

In other words, at most $P-1$ elements (out of P^R) of the probability tensor of $\boldsymbol{x}$ are nonzeros: their indices are multiples of the AR coefficients vector $\boldsymbol{a}$, and their values are the respective values of the characteristic vector $s[t]$. Note that if the $s[t]$ is uniform, all elements of its characteristic vector (except for the first, zero-indexed) are zeros, implying that $\widetilde{\boldsymbol{\mathcal{P}}}_{\boldsymbol{x}}$ is all-zeros as well (except for its $\boldsymbol{0}$-th

element), thus all the elements of $\boldsymbol{x}$ are iid and the AR coefficients are non-identifiable without additional information. However, if $s[t]$ is non-uniform, $\widetilde{\boldsymbol{\mathcal{P}}}_{\boldsymbol{x}}$ has at least one nonzero element. A convenient approach for estimating $\boldsymbol{a}$ is to look for the element with the largest absolute value in $\widetilde{\boldsymbol{\mathcal{P}}}_{\boldsymbol{x}}$ and to get its index-vector, denoted $\boldsymbol{m}$. Dividing $\boldsymbol{m}$ (over $\mathbb{GF}(P)$) by its first element $m(0)$, the coefficients vector $\boldsymbol{a}$ would be readily obtained. Naturally, since $\widetilde{\boldsymbol{\mathcal{P}}}_{\boldsymbol{x}}$ is unknown, its empirical estimate can be used instead, as described earlier.

4. SIMULATION RESULTS

We tested the two proposed methods for MA equalization of AR channels: An exhaustive entropy-minimization search and direct estimation of the equalizer, both based on the empirical characteristic tensor of $\boldsymbol{x}$ from $T = 40{,}000$ samples. Each trial is either "successful" if the correct equalizer is obtained, or "failed" otherwise. The methods were tested with different field orders (P), AR orders (R) and source distributions. The source distributions assigned a constant probability p_1 to all $P-1$ nonzero elements and probability $p_0 = 1-(P-1)p_1$ to $s[t]=0$, with a ratio $\eta \stackrel{\triangle}{=} \frac{p_0}{p_1}$, so as η approaches 1, $s[t]$ becomes "closer to uniform". Results are presented in Table 1 in terms of failures per 1000 independent trials. The channel's coefficient of order R were randomly drawn in each trial. Evidently, the results with the smaller values of P and R and with η sufficiently far from 1 are perfect for both methods. As expected, with $\eta = 1$ the channel is unidentifiable (cases of success are coincindencial). In the more difficult cases the exhaustive search (on the left-hand side in each column) performs better than the direct estimate (on the right-hand side).

Table 1. MA equalization of an AR channel: Failures in 1,000 trials

P	R	$\eta=0.8$	$\eta=0.9$	$\eta=1.0$	$\eta=1.1$	$\eta=1.2$
2	8	0|0	0|0	995|995	0|0	0|0
2	17	0|0	0|0	1000|1000	0|0	0|0
3	5	0|0	0|0	995|995	0|0	0|0
3	10	0|0	0|0	1000|1000	0|0	0|0
5	4	0|0	0|0	1000|999	0|4	0|0
5	7	0|0	1|46	1000|1000	4|75	0|0
7	3	0|0	3|61	997|995	5|94	0|0
7	6	0|0	50|565	1000|1000	101|603	0|0

5. APPENDIX

Proof of Theorem 2: Note that $\bar{\boldsymbol{x}}_t$ satisfies the recursion

$$\underbrace{\begin{bmatrix} x[t] \\ x[t-1] \\ \vdots \\ x[t-R+1] \end{bmatrix}}_{\bar{\boldsymbol{x}}_t} = \underbrace{\begin{bmatrix} a_1 & a_2 & \cdots & a_R \\ 1 & 0 & \cdots & 0 \\ & \ddots & \cdots & \\ 0 & \cdots & 1 & 0 \end{bmatrix}}_{\boldsymbol{A}} \circ \underbrace{\begin{bmatrix} x[t-1] \\ x[t-2] \\ \vdots \\ x[t-R] \end{bmatrix}}_{\bar{\boldsymbol{x}}_{t-1}} \oplus \underbrace{\begin{bmatrix} s[t] \\ 0 \\ \vdots \\ 0 \end{bmatrix}}_{\boldsymbol{s}_t}, \tag{9}$$

which can be expanded as

$$\bar{\boldsymbol{x}}_t = \boldsymbol{A} \circ \bar{\boldsymbol{x}}_{t-1} \oplus \boldsymbol{s}_t = \boldsymbol{A}^{\circ 2} \circ \bar{\boldsymbol{x}}_{t-2} \oplus \boldsymbol{A} \circ \boldsymbol{s}_{t-1} \oplus \boldsymbol{s}_{t-2} = \cdots$$
$$\cdots = \boldsymbol{A}^{\circ K} \circ \boldsymbol{x}_{t-K} \oplus \sum_{k=0}^{K-1}{}^{\circ} \boldsymbol{A}^{\circ k} \circ \boldsymbol{s}_{t-k} \tag{10}$$

(where $\boldsymbol{A}^{\circ k}$ denotes the k-th power of $\boldsymbol{A}$ over the field) with any $K < t$. Since all the RVecs on the last expression in (10) are statistically independent, the characteristic tensor of $\bar{\boldsymbol{x}}_t$ is given by the element-wise product of the characteristic tensors of these vectors. The elements of the characteristic tensor of $\boldsymbol{s}_t$ are evidently given by

$$\widetilde{\mathcal{P}}_{\boldsymbol{s}}(\boldsymbol{n}) = E\left[W_P^{\boldsymbol{n}^T \boldsymbol{s}_t}\right] = E\left[W_P^{n_1 s[t]}\right] = \tilde{p}_s(n_1). \tag{11}$$

Elements of the characteristic tensor of any linear transformation $\boldsymbol{w} = \boldsymbol{B} \circ \boldsymbol{s}_t$ of $\boldsymbol{s}_t$ are given by $\widetilde{\mathcal{P}}_{\boldsymbol{w}}(\boldsymbol{n}) = \widetilde{\mathcal{P}}_{\boldsymbol{s}}(\boldsymbol{B}^T \circ \boldsymbol{n})$, and if $\boldsymbol{B}$ is non-singular, the elements of $\widetilde{\boldsymbol{\mathcal{P}}}_{\boldsymbol{w}}$ are just a permutation of the elements of $\widetilde{\boldsymbol{\mathcal{P}}}_{\boldsymbol{s}}$. Now, since R is the "true" AR order, $a_R \neq 0$, and therefore the matrix $\boldsymbol{A}$ and all its powers are non-singular. This means that each non-$\boldsymbol{0}$-th element of the characteristic tensor of the sum in the last expression in (10) is a product of K non-$\boldsymbol{0}$-th elements of the characteristic tensor $\widetilde{\boldsymbol{\mathcal{P}}}_{\boldsymbol{s}}$, which in turn are simply non-0-th elements of the characteristic vector of $s[t]$. Since $s[t]$ is non-degenerate, the absolute values of all of its non-0-th elements are smaller than 1. Denoting the largest absolute value of a non-0-th element of $\tilde{\boldsymbol{p}}_s$ by $\lambda < 1$, we conclude that the absolute values of all non-$\boldsymbol{0}$-th elements of the characteristic tensor of $\bar{\boldsymbol{x}}_t$ are smaller than λ^K (regardless of the characteristic tensor of $\bar{\boldsymbol{x}}_{t-K}$, which are also all smaller or equal to 1 in absolute value). Since asymptotically K can be made arbitrarily large, all non-$\boldsymbol{0}$-th elements of $\widetilde{\boldsymbol{\mathcal{P}}}_{\boldsymbol{x}}$ vanish, so that $\widetilde{\mathcal{P}}_{\boldsymbol{x}}(\boldsymbol{0}) = 1$, and $\widetilde{\mathcal{P}}_{\boldsymbol{x}}(\boldsymbol{n}) \to 0$ for all $\boldsymbol{n} \neq \boldsymbol{0}$, and this is exactly the characteristic tensor of a uniform iid RVec.

However, when looking at $R+1$ consecutive samples, the augmented version of the matrix $\boldsymbol{A}$ in (9) would be singular, and this proof would not hold, since the product of powers of this matrix with some index-vectors $\boldsymbol{n}$ would yield the $\boldsymbol{0}$ index vector. Indeed, in this case we have the result of Theorem 3.

Proof of Theorem 3: Observe the following, alternative relation:

$$\underbrace{\begin{bmatrix} 1 & a_1 & \cdots & a_R \\ 0 & 1 & \cdots & a_{R-1} \\ & & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix}}_{\boldsymbol{H}} \circ \underbrace{\begin{bmatrix} x[t] \\ x[t-1] \\ \vdots \\ x[t-R] \end{bmatrix}}_{\boldsymbol{x}_t} = \underbrace{\begin{bmatrix} s[t] \\ s[t-1] \\ \vdots \\ s[t-R] \end{bmatrix}}_{\boldsymbol{s}_t}$$
$$\oplus \underbrace{\begin{bmatrix} 0 & 0 & \cdots & 0 \\ a_R & \ddots & \ddots & 0 \\ \vdots & \ddots & 0 & \vdots \\ a_1 & \cdots & a_R & 0 \end{bmatrix}}_{\boldsymbol{H}_c} \circ \underbrace{\begin{bmatrix} x[t-R-1] \\ x[t-R-2] \\ \vdots \\ x[t-2R-1] \end{bmatrix}}_{\boldsymbol{x}_{t-R-1}}. \tag{12}$$

Accordingly, due to the statistical independence between the RVecs $\boldsymbol{s}_t$ and $\boldsymbol{x}_{t-R-1}$ and due to the asymptotic stationarity of $\boldsymbol{x}_t$, the elements of the characteristic tensor of $\boldsymbol{x}$ satisfy

$$\widetilde{\mathcal{P}}_{\boldsymbol{x}}(\boldsymbol{H}^T \circ \boldsymbol{n}) = \widetilde{\mathcal{P}}_{\boldsymbol{s}}(\boldsymbol{n}) \widetilde{\mathcal{P}}_{\boldsymbol{x}}(\boldsymbol{H}_c^T \circ \boldsymbol{n}). \tag{13}$$

Observe that for the $P-1$ index vectors of the form $\boldsymbol{n} = [n_1 \ 0 \ \cdots \ 0]^T$ with $n_1 \neq 0$ we have $\widetilde{\mathcal{P}}_{\boldsymbol{s}}(\boldsymbol{n}) = \tilde{p}_s(n_1)$ and $\widetilde{\mathcal{P}}_{\boldsymbol{x}}(\boldsymbol{H}_c^T \circ \boldsymbol{n}) = \widetilde{\mathcal{P}}_{\boldsymbol{x}}(\boldsymbol{0}) = 1$. Since for such vectors $\boldsymbol{n}$ we also have $\boldsymbol{H}^T \circ \boldsymbol{n} = n_1 \otimes \boldsymbol{a}$, we conclude that $\widetilde{\mathcal{P}}_{\boldsymbol{x}}(n_1 \otimes \boldsymbol{a}) = \tilde{p}_s(n_1)$ for all $n_1 = 0, ..., P-1$. Next, we observe that (since $a_R \neq 0$) for any non-$\boldsymbol{0}$ index-vectors $\boldsymbol{n}$ of a different form, the product $\boldsymbol{H}_c \circ \boldsymbol{n}$ yields a non-$\boldsymbol{0}$ vector whose last element is zero. The characteristic tensor of $\boldsymbol{x}$ at any such index-vector of the form $\boldsymbol{n} = [\boldsymbol{m}^T 0]^T$ equals the characteristic tensor of $\bar{\boldsymbol{x}}$ (the length-R vector) at $\boldsymbol{m}$, which according to Theorem 2 is zero for all $\boldsymbol{m} \neq \boldsymbol{0}$. Consequently, $\widetilde{\mathcal{P}}_{\boldsymbol{x}}(\boldsymbol{H}^T \circ \boldsymbol{n})$ vanishes at any non-$\boldsymbol{0}$ index-vector $\boldsymbol{n}$ which is not of the form $\boldsymbol{n} = [n_1 \ 0 \ \cdots \ 0]^T$ - which concludes the proof.

6. REFERENCES

[1] Chong-Yung Chi, Ching-Yung Chen, Chil-Horng Chen, and Chih-Chun Feng, "Batch processing algorithms for blind equalization using higher-order statistics," *Signal Processing Magazine, IEEE*, vol. 20, no. 1, pp. 25–49, 2003.

[2] A.K. Takahata, E.Z. Nadalin, R. Ferrari, L.T. Duarte, R. Suyama, R.R. Lopes, J. M T Romano, and M. Tygel, "Unsupervised processing of geophysical signals: A review of some key aspects of blind deconvolution and blind source separation," *Signal Processing Magazine, IEEE*, vol. 29, no. 4, pp. 27–35, 2012.

[3] A. Yeredor, "ICA in boolean XOR mixtures," *Proc., ICA 2007, Lecture Notes in Computer Science (LNCS 4666)*, vol. 45, no. 1, pp. 17–24, 2007.

[4] A. Yeredor, "Independent component analysis over galois fields of prime order," *Information Theory, IEEE Transactions on*, vol. 57, no. 8, pp. 5342–5359, 2011.

[5] H. Gutch, P. Gruber, and F. J. Theis, "ICA over finite fields," *Proc., LVA/ICA 2010, Lecture Notes in Computer Science*, pp. 645–652, 2010.

[6] H. Gutch, P. Gruber, A. Yeredor, and F. J. Theis, "ICA over finite fieldsseparability and algorithms," *Signal Processing*, vol. 92, no. 8, pp. 1796 – 1808, 2012.

[7] I. Nemoianu, C. Greco, M. Castella, B. Pesquet-Popescu, and M. Cagnazzo, "Ona practical approach to source separation over finite fields for network coding applications," *Proc., IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP 2013)*, pp. 1335 – 1339, 2013.

[8] D.G. Silva, R. Attux, E.Z. Nadalin, L.T. Duarte, and R. Suyama, "An immune-inspired information-theoretic approach to the problem of ica over a galois field," in *Information Theory Workshop (ITW), 2011 IEEE*, 2011, pp. 618–622.

[9] D.G. Silva, E.Z. Nadalin, R. Attux, and J. Montalvao, "A modified version of the mexico algorithm for performing ica over galois fields," in *Machine Learning for Signal Processing (MLSP), 2012 IEEE International Workshop on*, 2012, pp. 1–6.

[10] D.G. Silva, E.Z. Nadalin, J. Montalvao, and R. Attux, "The modified MEXICO for ICA over finite fields," *Signal Processing*, vol. 93, no. 9, pp. 2525 – 2528, 2013.

[11] Huy Nguyen and Rong Zheng, "Binary independent component analysis with or mixtures," *Signal Processing, IEEE Transactions on*, vol. 59, no. 7, pp. 3168–3181, 2011.

[12] Huy Nguyen and Rong Zheng, "A binary independent component analysis approach to tree topology inference," *Signal Processing, IEEE Transactions on*, vol. 61, no. 12, pp. 3071–3080, 2013.

[13] D. Fantinato, D. G. Silva, R. Attux, R. Ferrari, T. L. Duarte, R. Suyama, J. M. Filho, A. Neves, and J. M. T. Romano, "Optimal linear filtering over Galois field: equalization and prediction," in *V Encontro dos Alunos e Docentes do DCA*, 2012.

[14] D. Fantinato, D. G. Silva, E. Z. Nadalin, R. Attux, J. M. T. Romano, A. Neves, and J. Montalvao, "Blind separation of convolutive mixture over Galois field," in *Machine Learning for Signal Processing (MLSP), 2013 IEEE International Workshop on*, 2013.