BLIND EQUALIZATION OF MOVING AVERAGE CHANNELS OVER GALOIS FIELDS

Arie Yeredor

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

ABSTRACT

We consider the blind estimation / equalization of a Moving Average (MA) channel over a finite field. In this framework, the channel's input and output signals, as well as its coefficients, belong to a finite (Galois) field, and all summation and multiplication operations are calculated modulo the field's prime order. The input is assumed to be a sequence of independent, identically distributed (iid) samples with an unknown distribution, and the goal is to estimate the channel coefficients based on its observed output only. We derive two different estimation approaches: One is based on sequential identification of factors of the channel's associated polynomial; The other is based on an attempted factorization of the empirical characteristic function of the channel's output signal. We explain the trade-offs between the methods and demonstrate their performance by simulation.

Index Terms— Blind System Identification; Blind Equalization; Moving Average; Galois Fields; Polynomial Factorization.

1. INTRODUCTION

While classical signal processing is usually aimed at signals over the real- or complex-valued fields (denoted $\mathbb{R}$ and $\mathbb{C}$, resp.), during the past decade some interest has been taken in signal processing over finite fields (see, e.g., [1–13], to name a few). In particular, blind signal processing (BSP) tasks such as blind source separation, blind system identification or blind equalization have been considered in different scenarios over finite fields, where not only the inputs, but also the coefficients and the results, are all confined to the field. In finite Galois fields of prime order P, denoted $\mathbb{GF}(P)$, this is simply attained by restricting all elements to the set $\{0, 1, \ldots, P-1\}$ and then applying all additions and multiplications modulo P.

Naturally, such linear operations over finite fields are rarely encountered in nature. Nevertheless, they are sometimes present by design in man-made systems, e.g., in computer architecture and in digital communication, where coding schemes such as convolution coding (e.g., [14]), Tomlinson-Harashima channel precoding (e.g., [15], ch.5), Network Coding (e.g., [16]) or convolution Network Coding (e.g., [17, 18]) are involved. Therefore, although the range of viable applications of BSP over finite fields is still modest, the growing line of ongoing research in this field demonstrates its two-fold motivation: prospective applications as well as mere theoretical interest.

Blind channel estimation and equalization over $\mathbb{R}$ or $\mathbb{C}$ is a well-studied topic with applications in diverse fields, such as communications, speech dereverberation or seismology (see, e.g., [19, 20] 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), with no additional information on the channel or on the source signal. However, classical methods used in $\mathbb{R}$ or $\mathbb{C}$ are inapplicable over finite fields, since moments, covariance and correlation are useless over $\mathbb{GF}(P)$, due to the modulo operation.

In the framework of finite fields, some initial treatment of the problem by Fantinato *et al.* has appeared in [4], [5], and more recently in [12]. However, the scenarios considered in [4] and in [12] were not fully blind, as some access to the source signal itself (in [4]) or to its multivariate probability mass function (PMF) (in [12]) were assumed. In [5] a convolutive multi-channel blind separation problem was considered, focusing more on separation than on equalization. In addition, the derivations and empirical testing were limited to the binary case ($\mathbb{GF}(2)$), rather than to the more general $\mathbb{GF}(P)$.

In our previous work [9] on blind channel estimation over $\mathbb{GF}(P)$ we offered an explicit solution (consistent estimate) only for the case of an auto-regressing (AR) channel with an infinite impulse response (IIR), which can be equalized using a finite impulse response (FIR) filter (equalizer). As we shall explain in this paper, the converse case of a moving-average (MA) channel being blindly equalized by an AR equalizer is far more involved, both conceptually and computationally. Thus, our contribution here will consist of proposing (and demonstrating by simulation) two alternative strategies for blindly estimating / equalizing an MA channel over $\mathbb{GF}(P)$ based on its observed output only, under the common assumption that its input is a sequence of iid symbols over the same field (with an unknown probability distribution).

For simplicity, we limit our discussion in here to fields of prime order[1], P, such that all addition, subtraction and multiplication operations are calculated modulo P.

2. BACKGROUND: RANDOM VARIABLES IN $\mathbb{GF}(P)$

Any random variable (RV) u in $\mathbb{GF}(P)$ is fully characterized by a *probability vector* $\boldsymbol{p}_u = [p_u(0)\ p_u(1)\ \cdots\ p_u(P-1)]^T \in \mathbb{R}^P$, where $p_u(m) \triangleq \Pr\{u = m\}\ \forall m \in \mathbb{GF}(P)$. An RV u is called *uniform* if $p_u(m) = \frac{1}{P}\ \ \forall m \in \mathbb{GF}(P)$, and *degenerate* if it deterministic, namely, if for some $m_0 \in \mathbb{GF}(P)$, $p_u(m_0) = 1$.

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 $\boldsymbol{p}$:

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

where $W_P \triangleq e^{-j\frac{2\pi}{P}}$. Note that u is fully characterized by $\tilde{\boldsymbol{p}}_u$ as well, since $\boldsymbol{p}_u$ can be directly obtained from $\tilde{\boldsymbol{p}}_u$ via an inverse DFT.

This research was supported by the ISRAEL SCIENCE FOUNDATION (grant No. 2427/19).

[1] Rather than the more general case of a prime-power order, P^M.

Note also, that for any RV u, $\tilde{p}_u(0) = 1$, and $|\tilde{p}_u(n)| \leq 1 \; \forall n$, where for $n \neq 0$ equality holds iff u is degenerate. If u is uniform, $\tilde{p}_u(n) = 0 \; \forall n \neq 0$. The characteristic vector of the sum of two statistically independent RVs is given by the element-wise product of their characteristic vectors (e.g., [9]).

A K-dimensional random vector (RVec) $\boldsymbol{u}$ is characterized by the K-ways *probabilities tensor* (matrix for $K = 2$) $\boldsymbol{\mathcal{P}}_{\boldsymbol{u}} \in \mathbb{R}^{P^{(\times K)}}$, where $\mathcal{P}_{\boldsymbol{u}}(m_1, \ldots, m_K) = \Pr\{u_1 = m_1, ..., u_K = m_K\}$, $\forall m_1, \ldots, m_K \in \mathbb{GF}(P)$, expressed more compactly using the "index-vector" $\boldsymbol{m}$ 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 $\boldsymbol{\mathcal{P}}_{\boldsymbol{u}}$, namely

$$\widetilde{\mathcal{P}}_{\boldsymbol{u}}(\boldsymbol{n}) = E\left[W_P^{\boldsymbol{n}^T \boldsymbol{u}}\right] = \sum_{\boldsymbol{m}} \mathcal{P}_{\boldsymbol{u}}(\boldsymbol{m}) W_P^{\boldsymbol{n}^T \boldsymbol{m}}, \qquad (2)$$

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

The elements $u_1, ..., u_K$ of $\boldsymbol{u}$ are statistically independent iff $\mathcal{P}_{\boldsymbol{u}}$ 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, $\widetilde{\mathcal{P}}_{\boldsymbol{u}}$ is the outer product of the elements' caracteristic vectors.

If $\boldsymbol{u} = \boldsymbol{B}\boldsymbol{v}$ is a linear transformation of $\boldsymbol{v}$ (where $\boldsymbol{B}$ is a deterministic matrix with elements in $\mathbb{GF}(P)$), their characteristic tensors are related by $\widetilde{\mathcal{P}}_{\boldsymbol{u}}(\boldsymbol{n}) = \widetilde{\mathcal{P}}_{\boldsymbol{v}}(\boldsymbol{B}^T\boldsymbol{n})$. Note that for RVs as a particular case, if $u = bv$ is a scaled version of v (where $b \in \mathbb{GF}(P)$ is deterministic), then $\tilde{p}_u(n) = \tilde{p}_v(bn)$, resulting in a permutation of the characteristic vector of v.

An RV u is called *rich* if all elements of $\tilde{\boldsymbol{p}}_u$ are non-zeros. Obviously, a rich RV cannot be uniform. A linear combination of two independent rich RVs is also rich. Note that "most" random variables are rich, in the sense that the set of non-rich RVs has Lebesgue measure zero in the space of all RVs in $\mathbb{GF}(P)$. In $\mathbb{GF}(2)$ and in $\mathbb{GF}(3)$ any non-uniform random variable is rich.

3. PROBLEM FORMULATION AND DISCUSSION

Let $s[t] \in \mathbb{GF}(P)$ denote a discrete-time random process (the "source signal") at the input of a general rational LTI channel in $\mathbb{GF}(P)$. The output is given by

$$x[t] = \sum_{m=0}^{M} b_m s[t-m] - \sum_{r=1}^{R} a_r x[t-r], \qquad (3)$$

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. To avoid scale ambiguities, we assume $b_0 = 1$ as a scaling convention.

By switching the roles of the MA and AR coefficients, a similar equation representing the inverse channel can be written as

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

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.

In its general form, the blind identification / equalization problem consists of estimating the MA and AR coefficients from observation of the output $x[t]$ (only) for $t = 0, ..., T$–1. The only assumption made regarding the unobserved input is that $s[t]$ is a sequence of iid random variables (with an unknown probability distribution $\boldsymbol{p}_s$). The estimated coefficients may in turn be used (in (4)) to recover the source signal $s[t]$.

In previous work [9] we focused on the case of a pure AR channel ($M = 0$ in (3)). We showed that thanks to the FIR structure of the equalizer (obtained only in the case of an AR channel), the AR coefficients can be consistently estimated directly from the empirical characteristic tensor $\widetilde{\boldsymbol{\mathcal{P}}}_{\boldsymbol{x}}$ of $R + 1$ consecutive output samples.

Unfortunately, when the channel includes an MA part ($M > 0$), the implied equalizer no longer has a FIR structure, so there is no convenient linear transition from a finite-dimensional multivariate PMF of consecutive samples of $x[t]$ to the marginal PMF of an equalizer's output $y[t]$. An exhaustive search through all possible equalizers would therefore involve re-applying each tested equalizer to the entire T-long signal $x[t]$ in order to obtain the resulting $y[t]$ and to estimate its PMF directly.

In this paper we focus on the case of a pure MA channel ($M > 0$, $R = 0$ in (3)), and propose to circumvent this prohibitive limitation using two different approaches, derived in the following section.

4. THE PROPOSED ESTIMATION APPROACHES

We propose two blind estimation approaches for the case of an MA channel, assuming that $s[t]$ is iid, rich and non-degenerate, and that the MA order M is known. The first approach is based on successive equalizations by "short" equalizers, using an attempted factorization of the unknown MA polynomial (to be defined hereafter) into a product of smaller, irreducible polynomials. This approach essentially seeks the overall equalizer by "stripping" its shorter irreducible factors one by one, and is therefore called "MA STripping EstimatoR" (MASTER). This approach can potentially reduce the number of tested equalizers quite significantly relative to an exhaustive search, but still requires the application of each tested equalizer to a T-long signal.

The second approach is based on direct estimation of the MA coefficients from the attempted factorization of the empirical characteristic tensor of consecutive samples of $x[t]$, and is therefore called "CHaracteristic-tensor based Estimation by Factorization" (CHEF). This approach requires an exhaustive search through all P^M possible sets of possible MA coefficients, but avoids the need to re-apply the implied equalizer for each tested set.

4.1. MASTER

We begin with the following Theorem (proved in the Appendix):

Theorem 1. *Assume an order-M MA channel ($R = 0$, $M > 0$ in (3)). If $s[t]$ is iid, rich and non-degenerate, then $x[t]$ and $x[t-M]$ are statistically dependent (for all t).*

It is important to note, that unlike the classical real- or complex-valued fields, where this property is trivial, and holds for the output of any FIR channel regardless of the input distribution (as long as it is non-degenerate), in the case of a finite field this property may not hold if the source is not rich. For example, in $\mathbb{GF}(11)$, if the characteristic vector of the source is given by $\tilde{p}_s = [1 \; \theta \; 0 \; 0 \; 0 \; 0 \; 0 \; 0 \; 0 \; 0 \; \theta]^T$ with any $\theta \in (0, 0.5)$ (such a form always corresponds to a feasible probabilities vector), it can be shown that the output of a 2nd-order MA channel with $b_0 = 1, b_1 = 2$ is an iid sequence (of uniform random variables).

Note in addition, that for any MA channel of order M, driven by a non-degenerate, iid input $s[t]$, the samples $x[t]$ and $x[t-M']$

are statistically independent for all $M' > M$ (and for all t), because they are functions of disjoint sets of independent RVs.

To proceed, note that any MA channel of order M can be represented by an associated order-M polynomial (over the field)

$$Q(z) \triangleq b_0 + b_1 z^{-1} + b_2 z^{-2} + \cdots + b_M z^{-M} \quad (b_M \neq 0). \quad (5)$$

A polynomial is called *irreducible* over the field if it cannot be represented as a product of polynomials of smaller orders (larger or equal to 1) over the field (and is called *reducible* otherwise). For example, all polynomials over $\mathbb{C}$ of orders larger than 1 are reducible (over $\mathbb{C}$), but quadratic polynomials over $\mathbb{R}$ with no real-valued roots are irreducible (over $\mathbb{R}$). In $\mathbb{GF}(P)$ the number of irreducible polynomials of each order can be a small fraction of the total (finite) number of polynomials of the same order. Obviously, all $P-1$ monic polynomials (with $b_0 = 1$) of order 1 in $\mathbb{GF}(P)$ are irreducible. But, for example, considering polynomials of order 5, the number of irreducible polynomials is 6 out of 16 in $\mathbb{GF}(2)$; 48 out of 162 in $\mathbb{GF}(3)$; 624 out of 2,500 in $\mathbb{GF}(5)$; and 3,360 out of 14,406 in $\mathbb{GF}(7)$.

Therefore, if $Q(z)$ is reducible, the channel can be equalized by a series of equalizers corresponding to its factors. Even if $Q(z)$ is irreducible, searching only for irreducible factors entails trying out significantly fewer equalizers than an exhaustive search.

We propose a recursive search process, which proceeds as follows. Assume an MA channel of order M, whose associated polynomial $Q(z)$ can be factored as $Q(z) = (1+\beta z^{-1})\breve{Q}(z)$ with some $\beta \in \mathbb{GF}(P)$, where $\breve{Q}(z)$ has no 1st-order factors. According to Theorem 1, $x[t]$ and $x[t-M]$ are statistically dependent, whereas $x[t]$ and $x[t-M']$ are statistically independent for all $M' > M$. Suppose now that $x[t]$ is passed through an order-1 AR equalizer:

$$y[t] = -a \cdot y[t-1] + x[t] \quad , \quad t = 0, \ldots, T-1 \qquad (6)$$

with some $a \in \mathbb{GF}(P)$ and zero initial conditions. Then if (and only if) $a = \beta$, the output $y[t]$ will be an order-$(M-1)$ MA process (with the associated polynomial $\breve{Q}(z) = Q(z)/(1+\beta z^{-1})$), so that $y[t]$ and $y[t-M]$ (as well as $y[t]$ and $y[t-M']$ for all $M' > M$) would be independent. Otherwise (if $a \neq \beta$), $y[t]$ would be an ARMA$(1, M)$ process, which does not share this property [2].

We can therefore try equalizing $x[t]$ using each of the $P-1$ order-1 equalizers (with $\alpha = 1, 2, \ldots, P-1$) in (6), and select the equalizer which minimizes the empirical dependence between $y[t]$ and $y[t-M]$ (and possibly also between $y[t]$ and $y[t-M']$ for some values of $M' > M$). To quantify the dependence, we propose to use the Kullback-Leibler Divergence (KLD) between their empirical joint PMF and the product of their empirical marginal PMFs (which are the same). More specifically, for all $m, m_1, m_2 \in \mathbb{GF}(P)$ let

$$\hat{p}_y(m) \triangleq \frac{1}{T} \sum_{t=0}^{T-1} I\{y[t] = m\} \qquad (7)$$

$$\hat{P}_y(m_1, m_2) \triangleq \frac{1}{T-M} \sum_{t=M}^{T-1} I\{y[t] = m_1, y[t-M] = m_2\} \quad (8)$$

denote the empirical PMFs (where $I\{\cdot\}$ is the Indicator function, which equals 1 if the argument condition is met and 0 otherwise). Then the proposed measure of independence would be

$$D_M([1 \;\; a]) \triangleq -\sum_{m1,m2=0}^{P-1} \hat{p}_y(m_1)\hat{p}_y(m_2) \log \frac{\hat{P}_y(m_1, m_2)}{\hat{p}_y(m_1)\hat{p}_y(m_2)} \quad (9)$$

(which is also the mutual information between $y[n]$ and $y[n-M]$).

If, indeed, one of the factors of $Q(z)$ is a single 1st-order polynomial of the form $1 + \beta z^{-1}$, then $D_M([1 \;\; \beta])$ will be significantly smaller than $D_M([1 \;\; a])$ for all $a \neq \beta$. If $Q(z)$ has additional 1st-order factors, there may be additional values of a leading to similar minimal values of $D_M([1 \;\; a])$. Conversely, if $Q(z)$ does not have any 1st-order factors, then $D_M([1 \;\; a])$ would remain relatively high for all $a \in [1, P-1]$.

Asymptotically (in T), the empirical KLD tends to zero when a correct factor is used, and to some positive value when an incorrect factor is tried. For a finite T, a threshold should be determined first, to which the minimal obtained D_M would be compared in order to infer whether or not the minimizing equalizer is indeed a factor of $Q(z)$. We propose to obtain the threshold (before attempting any equalization) by averaging values of $D_{M'}$ obtained from the original signal $x[t]$ for several values of M', ranging from $M+1$ to some $M_{max} > M$. Since all pairs of $x[t]$ and $x[t-M']$ for $M' > M$ are independent, the associated $D_{M'}$ values can provide a reference for a "small" KLD for the specific signal and observation length T.

Once a minimizing 1st-order factor (producing D_M below the threshold) is found, that factor would be "stripped" from $Q(z)$, and the new generated signal $y[t]$ (which is now an MA$(M-1)$ process) would be subject to a similar search of equalizing factors, generating new equalized signals $y[t]$. The process is repeated until all factors have been "stripped-out" and $x[t]$ is fully equalized.

When all the attempted 1st-order factors yield "high" D_M values (above threshold), we move to look for 2nd-order factors. Recall that we only need to look for *irreducible* 2nd-order factors (fewer than all possible 2nd-order factors). When true 2nd-order factors are used, the current MA order M_c is expected to be reduced by 2, so we would compare D_{M_c} and $D_{M_c - 1}$ to the threshold, and proceed to "strip" such 2nd-order factors by re-equalizing $y[t]$ accordingly. Once these are exhausted, we would proceed to 3rd-order factors (comparing D_{M_c}, D_{M_c-1} and D_{M_c-2} to the threshold), and so forth.

4.2. CHEF

Our second approach avoids the need to re-apply a set of equalizers to T-long signals, but requires the empirical estimation of a multivariate PMF, and an exhaustive search through all $P^{M-1}(P-1)$ possible combinations of M coefficients (with $b_M \neq 0$).

Let L be a positive integer, and define the L-long vector $\boldsymbol{x}_t \triangleq [x[t] \;\; x[t-1] \;\; \cdots \;\; x[t-L+1]]^T$ and the $(M+L-1)$-long vector $\boldsymbol{s}_t \triangleq [s[t] \;\; s[t-1] \;\; \cdots \;\; s[t-M-L+1]]^T$.

For any presumed set of channel coefficients $\hat{b}_1, \hat{b}_2, ..., \hat{b}_M$, we can write $\boldsymbol{x}_t = \hat{\boldsymbol{B}}\boldsymbol{s}_t$, where $\hat{\boldsymbol{B}}$ is an $L \times (M+L)$ Toeplitz matrix,

$$\hat{\boldsymbol{B}} = \begin{bmatrix} 1 & \hat{b}_1 & \hat{b}_2 & \cdots & \hat{b}_M & 0 & \cdots & 0 \\ 0 & 1 & \hat{b}_1 & \hat{b}_2 & \cdots & \hat{b}_M & 0 & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & 1 & \hat{b}_1 & \hat{b}_2 & \cdots & \hat{b}_M \end{bmatrix}, \qquad (10)$$

which implies the relation $\widetilde{\mathcal{P}}_{\boldsymbol{x}}(\boldsymbol{n}) = \widetilde{\mathcal{P}}_{\boldsymbol{s}}(\hat{\boldsymbol{B}}^T\boldsymbol{n})$ between the characteristic tensors of $\boldsymbol{x}_t$ and $\boldsymbol{s}_t$. Although $\widetilde{\mathcal{P}}_{\boldsymbol{s}}$ is unknown, recalling the iid structure of $s[t]$ we can write, for any $(M+L)$-long index-vector $\boldsymbol{m} = [m_1, ..., m_{M+L}]^T$ in $\mathbb{GF}(P)$,

$$\widetilde{\mathcal{P}}_{\boldsymbol{s}}(\boldsymbol{m}) = \tilde{p}_s(m_1) \cdot \tilde{p}_s(m_2) \cdots \tilde{p}_s(m_{M+L}), \qquad (11)$$

where $\tilde{p}_s(1)$, $\tilde{p}_s(2)$,... $\tilde{p}_s(P-1)$ (the elements of the characteristic vector of $s[t]$) are unknown parameters (note that $\tilde{p}_s(0) = 1$ is

[2] A formal proof that under mild conditions an ARMA(R, M) process cannot share this property (independence of $y[t]$ and $y[t-M']$ for all $M' > M$) is omitted due to lack of space.

always known). Taking the log of $\widetilde{\mathcal{P}}_{\boldsymbol{x}}(\boldsymbol{n}) = \widetilde{\mathcal{P}}_{\boldsymbol{s}}(\hat{\boldsymbol{B}}^T \boldsymbol{n})$ we obtain, for any L-long index-vector $\boldsymbol{n}$ in $\mathbb{GF}(P)$,

$$\log(\widetilde{\mathcal{P}}_{\boldsymbol{x}}(\boldsymbol{n})) = \theta(\hat{\boldsymbol{b}}_1^T \boldsymbol{n}) + \theta(\hat{\boldsymbol{b}}_2^T \boldsymbol{n}) + \cdots + \theta(\hat{\boldsymbol{b}}_{M+L}^T \boldsymbol{n}), \qquad (12)$$

where $\boldsymbol{b}_m$ denotes the m-th column of $\boldsymbol{B}$, and where $\theta(n) \triangleq \log(\tilde{p}_s(n))$ for $n = 1, ..., P-1$ are $P-1$ unknown (complex-valued) parameters (note that $\theta(0) = \log(1) = 0$ is always known).

Thus, for any presumed set of coefficients $\hat{\boldsymbol{b}} \triangleq [\hat{b}_1, ..., \hat{b}_M]^T$, we construct the matrix $\hat{\boldsymbol{B}}$ (cf. (10)) and use its columns to construct P tensors, $\boldsymbol{\mathcal{H}}_0, \boldsymbol{\mathcal{H}}_1, ...\boldsymbol{\mathcal{H}}_{P-1} \in \mathbb{N}^{P(\times L)}$ as follows:

$$\boldsymbol{\mathcal{H}}_k(\boldsymbol{n}) = \sum_{m=1}^{M+L} I\{\hat{\boldsymbol{b}}_m^T \boldsymbol{n} = k\} \quad , \quad k = 0, 1, ...P-1. \qquad (13)$$

Using these tensors, we may rewrite (12) as

$$\log(\widetilde{\mathcal{P}}_{\boldsymbol{x}}(\boldsymbol{n})) = \sum_{k=0}^{P-1} \boldsymbol{\mathcal{H}}_k(\boldsymbol{n})\theta(k) = \sum_{k=1}^{P-1} \boldsymbol{\mathcal{H}}_k(\boldsymbol{n})\theta(k), \qquad (14)$$

(where the second equality is due to the fact that $\theta(0) = 0$).

Now, given the observed signal $x[t]$, we can estimate the PMF tensor $\boldsymbol{\mathcal{P}}_{\boldsymbol{x}} \in \mathbb{R}^{P(\times L)}$ of $\boldsymbol{x}_t$ using

$$\widehat{\mathcal{P}}_{\boldsymbol{x}}(\boldsymbol{m}) = \frac{1}{T-L+1} \sum_{t=L-1}^{T-1} I\{\boldsymbol{x}_t = \boldsymbol{m}\} \qquad (15)$$

for all L-long index vectors $\boldsymbol{m}$ in $\mathbb{GF}(P)$. Then, by applying an L-dimensional FFT to $\widehat{\boldsymbol{\mathcal{P}}}_{\boldsymbol{x}}$ we get the estimated characteristic tensor $\widehat{\widetilde{\mathcal{P}}}_{\boldsymbol{x}}$. Using (14) we can then test, for each presumed set of channel coefficients $\hat{\boldsymbol{b}}$, how well $\widehat{\widetilde{\mathcal{P}}}_{\boldsymbol{x}}$ can be "explained" as the characteristic tensor of the implied linear transformation of an $(M+L-1)$-long vector of iid samples $\boldsymbol{s}_t$ with some (unknown) PMF vector. In other words, each presumed set $\hat{\boldsymbol{b}}$ is given a "fitness score" as:

$$F(\hat{\boldsymbol{b}}) \triangleq \min_{\boldsymbol{\theta} \in \mathbb{C}^{P-1}} \sum_{\boldsymbol{n}} \left| \widehat{\widetilde{\mathcal{P}}}_{\boldsymbol{x}}(\boldsymbol{n}) - \sum_{k=1}^{P-1} \boldsymbol{\mathcal{H}}_k(\boldsymbol{n})\theta(k) \right|^2 \qquad (16)$$

(the dependence on $\hat{\boldsymbol{b}}$ is implicitly included in the tensors $\boldsymbol{\mathcal{H}}_k$ through (13)). Thanks to the conversion to the log form, the minimization in (16) amounts to a linear Least Squares problem (in $\boldsymbol{\theta}$), which admits a closed-form solution for each $\hat{\boldsymbol{b}}$. The selected estimate of the channel coefficients is the value of $\hat{\boldsymbol{b}}$ for which $\log(\widehat{\widetilde{\boldsymbol{\mathcal{P}}}}_{\boldsymbol{x}})$ is "best explained" by the implied channel model, namely the value of $\hat{\boldsymbol{b}}$ attaining the minimal $F(\hat{\boldsymbol{b}})$ in (16).

5. SIMULATION RESULTS

We demonstrate the performance of the two approaches (MASTER and CHEF) comparing running-times (with Matlab® R2020b on Windows 10, PC with i7-8700 CPU, 3.20GHz) and performance, vs. the observation length (T) for three different cases in $\mathbb{GF}(2)$, $\mathbb{GF}(3)$ and $\mathbb{GF}(5)$ with MA channels of orders $M = 9$, $M = 6$ and $M = 4$ (respectively). Note that due to the finite fields framework, success is binary: either the correct channel is fully identified (equalized) or not, as there are no "tolerable" errors, in general. Therefore the performance is shown in terms of the empirical success rate in 100 independent trials, and the running-times are averaged over the same trials. Both methods were applied to the same signals $x[t]$, generated by passing randomly drawn iid source signals $s[t]$ through the specified channels. The probability vectors of the source signals were of the form $p_s(1) = 0.9$, and $p_s(m) = 0.1/(P-1)$ for $m \neq 1$, where $P = 2, 3, 5$ is the field order. We used $M_{max} = L = M + 3$.

	MASTER			CHEF		
P M	$T = 10^4$	10^5	10^6	10^4	10^5	10^6
2 9	0.16[s]	2.6[s]	22[s]	0.23[s]	0.24[s]	0.35[s]
3 6	0.10[s]	1.0[s]	11[s]	1.8[s]	1.7[s]	1.8[s]
5 4	0.10[s]	1.0[s]	11[s]	6.7[s]	6.8[s]	6.8[s]
2 9	54%	39%	50%	100%	100%	100%
3 6	54%	59%	58%	100%	100%	100%
5 4	76%	99%	98%	100%	100%	100%

Table 1. Avg. Run Times (top); Empirical Success Rates (bottom).

6. CONCLUSION

We presented two alternative methods, MASTER and CHEF, for blind MA channel estimation over $\mathbb{GF}(P)$, based on the observed output (only), assuming that the input is an iid sequence (with an unknown distribution) and that the order M is known.

The MASTER approach spares the need for an exhaustive search by sequentially identifying and "stripping" irreducible factors of the channel's polynomial, but requires the application of each prospective equalizer to a T-long signal. The CHEF approach requires an exhaustive search through all possible combinations, but spares the need to apply each tested combination to the signal. For shorter observation times CHEF can be considerably more time-consuming than MASTER, but its accuracy is significantly better.

7. APPENDIX: PROOF OF THEOREM 1

Proof. Consider the two samples (RVs)

$$\alpha \triangleq x[t] = \sum_{m=0}^{M} b_m s[t-m] \triangleq u + b_M s[t-M]$$

$$\beta \triangleq x[t-M] = \sum_{m=0}^{M} b_m s[t-M-m] \triangleq s[t-M] + v,$$

where u and v are defined implicitly as the sums of the remaining terms. Since both u and v are linear combinations of rich RVs, they are both rich. In addition, being based on disjoint groups of independent RVs, both are independent of each other and of $s[t-M]$. Thus, the characteristic vectors $\tilde{\boldsymbol{p}}_\alpha$ and $\tilde{\boldsymbol{p}}_\beta$ of α and β (resp.) are given by $\tilde{p}_\alpha(n) = \tilde{p}_u(n)\tilde{p}_s(b_M n)$ and $\tilde{p}_\beta(n) = \tilde{p}_s(n)\tilde{p}_v(n)$ ($\forall n \in [0, P-1]$). Their characteristic tensor (matrix) $\widetilde{\boldsymbol{\mathcal{P}}}$ is

$$\begin{aligned}\widetilde{\mathcal{P}}(n_1, n_2) &= E\left[W_P^{n_1\alpha + n_2\beta}\right] = E\left[W_P^{n_1(u+b_M s)+n_2(s+v)}\right] \\ &= \tilde{p}_u(n_1)\tilde{p}_s(b_M n_1 + n_2)\tilde{p}_v(n_2)\end{aligned}$$

($\forall n_1, n_2 \in [0, P-1]$). $x[t]$ and $x[t$–$M]$ are independent iff $\forall n_1, n_2$: $\widetilde{\mathcal{P}}(n_1, n_2) = \tilde{p}_\alpha(n_1)\tilde{p}_\beta(n_2)$. We observe that since u and v are rich, this is satisfied iff $\tilde{p}_s(b_M n_1 + n_2) = \tilde{p}_s(b_M n_1)\tilde{p}_s(n_2)$ for all n_1, n_2. However, let $\check{n}_2$ denote the index of the element with the smallest absolute value in $\tilde{\boldsymbol{p}}_s$. Since $s[t]$ is non-degenerate, for all nonzero n_1 we have $|\tilde{p}_s(b_M n_1)| < 1$, and therefore the absolute value of the product $\tilde{p}_s(b_M n_1)\tilde{p}_s(\check{n}_2)$ is smaller than the absolute value of $\tilde{p}_s(\check{n}_2)$, so this product cannot equal another element $\tilde{p}_s(b_M n_1 + \check{n}_2)$ of $\tilde{p}_s$, and the independence condition cannot be satisfied. □

8. REFERENCES

[1] 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.

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

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

[4] 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.

[5] 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.

[6] 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.

[7] 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.

[8] I. Nemoianu, C. Greco, M. Castella, B. Pesquet-Popescu, and M. Cagnazzo, "On a 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.

[9] A. Yeredor, "On blind channel identification and equalization over Galois fields," *Proc., IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP 2014)*, pp. 4239–4243, 2014.

[10] S. Painsky, A. amd Rosset and M. Feder, "Binary independent component analysis: Theory, bounds and algorithms,," in *Machine Learning for Signal Processing (MLSP), 2016 IEEE International Workshop on*, 2016, pp. 1–6.

[11] S. Painsky, A. amd Rosset and M. Feder, "Large alphabet source coding using independent component analysis," *IEEE Transactions on Information Theory*, vol. 63, pp. 6514–6529, 2017.

[12] D. G. Fantinato, A. Neves, Silva D. G., and R. Attux, "Blind channel equalization of encoded data over Galois fields," in *Machine Learning for Signal Processing (MLSP), 2017 IEEE International Workshop on*, 2017.

[13] S. Painsky, A. amd Rosset and M. Feder, "Linear independent component analysis over finite fields: Algorithms and bounds," *IEEE Transactions on Signal Processing*, vol. 66, pp. 5875–5886, 2018.

[14] R. Johannesson and K.Sh. Zigangirov, *Fundamentals of Convolutional Coding, 2nd Ed.*, Wiley, IEEE Press, 2015.

[15] F. A. Dietrich, *Robust Signal Processing for Wireless Communications*, in Springer series on Foundations in Signal Processing, Communications and Networking. Springer-Verlag Berlin Heidelberg, 2008.

[16] M. Médard and A. Sprinston, *Network Coding: Fundamentals and Applications*, Academic Press, 2012.

[17] U. Erez and M. Feder, "Efficient network code design for cyclic networks," *IEEE Transactions on Information Theory*, vol. 56, pp. 3862–3878, 2010.

[18] M. Lvov and H. H. Permuter, "Initialization algorithms for convolutional network coding," *IEEE Transactions on Information Theory*, vol. 64, pp. 5277–5295, 2018.

[19] 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.

[20] 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.