Supplemental Material – Linear Independent Component Analysis over Finite Fields: Algorithms and Bounds

Amichai Painsky, *Member, IEEE,* Saharon Rosset and Meir Feder, *Fellow, IEEE*

I. Illustrations for Linear ICA over Finite Fields

In this experiment we examine our ability to recover d independent ternary sources that were mixed by an unknown matrix B. Let $\underline{S} \in \{0,1,2\}^d$ be a d-dimensional ternary vector. Assume that the components of $\underline{S}$ are i.i.d. and follow a marginal distribution $P(S_i) = [0.2, 0.3, 0.5]^T$. This means that the joint entropy of $\underline{S}$ is $d \cdot 1.485$. We draw $10,000$ i.i.d. samples from $\underline{S}$ and mix them with a binary matrix B. Then, we apply our binary ICA approach to recover the original samples and the mixing matrix B. Figure 1 demonstrates the results we achieve for different number of components d. We compare our suggested approach with three alternative methods: AMERICA, MEXICO and cobICA, as described in the main text.

We first notice that both GLICA and AMERICA successfully recover the mixing matrix B (up to permutation of the sources), as they achieve an empirical sum of marginal entropies, which equals to the entropy of the samples prior to the mixture (blue curve at the bottom). Second, we notice that our suggested lower bound is tight, as GLICA and AMERICA attain it. The green curve corresponds to MEXICO, which demonstrates inferior performance. Finally, the red curve with the circles is cobICA, which is less competitive as the dimension increases. It is important to emphasize that while AMERICA and MEXICO are designed under the assumption that a perfect decomposition exists, cobICA and GLICA do not assume a specific generative model. Nevertheless, GLICA shows to successfully recover the mixing matrix B.

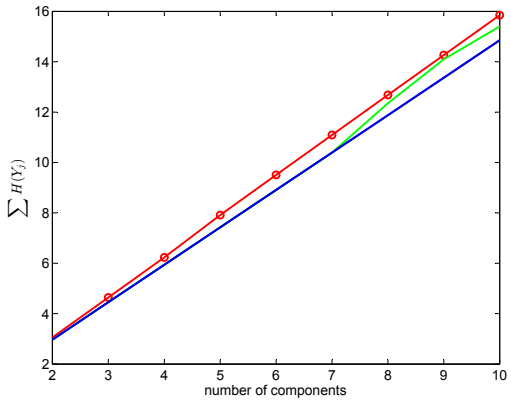

Fig. 1: Recovering independent sources experiment. The blue curve corresponds to GLICA and AMERICA, the green curve is MEXICO while red curve with the circles is cobICA

II. On the Flexibility of Linear Transformations

A. Non-linear binary ICA

1) Worst-case Analysis:

Theorem 1: For any random vector $\underline{X} \sim \underline{p}$, over an alphabet size $m = 2^d$ we have that

$$\max_{\underline{p}} C(\underline{p}, g_{opt}) = \Theta(d)$$

Proof We first notice that $\sum_{j=1}^{d} H(Y_j) = \sum_{j=1}^{d} h_b(P(Y_j = 0)) \leq d$. In addition, $H(\underline{X}) \geq 0$. Therefore, we have that $C(\underline{p}, g_{opt})$ is bounded from above by d. Let us also show that this bound is tight, in the sense that there exists a joint

A. Painsky is with the School of Computer Science and Engineering, The Hebrew University of Jerusalem, Israel. contact: amichai.painsky@huji.mail.ac.il
S. Rosset is with the Statistics Department, Tel Aviv University, Tel Aviv, Israel. contact: saharon@post.tau.ac.il
M. Feder is with the Department of Electrical Engineering, Tel Aviv University, Tel Aviv, Israel
The material in this paper was presented in part at the 2016 IEEE International Workshop on Machine Learning for Signal Processing

probability distribution $\tilde{\underline{p}}$ such that $C(\tilde{\underline{p}}, g_{opt})$ is linear in d. Let $\tilde{p}_1 = \tilde{p}_2 = \cdots = \tilde{p}_{m-1} = \frac{1}{3(m-1)}$ and $\tilde{p}_m = \frac{2}{3}$. Then, $\tilde{\underline{p}}$ is ordered and satisfies $P(Y_i = 0) = \frac{m}{6(m-1)}$.

In addition, we notice that assigning symbols in a decreasing order to $\tilde{\underline{p}}$ results with an optimal permutation. This is simply since $P(Y_j = 0) = \frac{m}{6(m-1)}$ is the minimal possible value of any $P(Y_j = 0)$ that can be achieved when summing any $\frac{m}{2}$ elements of $\tilde{p}_i$. Further we have that,

$$\begin{aligned} C(\tilde{\underline{p}}, g_{opt}) = &\sum_{j=1}^{d} H(Y_j) - H(\underline{X}) = \sum_{j=1}^{d} h_b(P(Y_j = 0)) - H(\underline{X}) = \\ &\log(m) \cdot h_b\left(\frac{m}{6(m-1)}\right) + \left((m-1)\frac{1}{3(m-1)} \log \frac{1}{3(m-1)} + \frac{2}{3}\log\frac{2}{3}\right) = \\ &\log(m) \cdot h_b\left(\frac{m}{6(m-1)}\right) - \frac{1}{3}\log(m-1) + \frac{1}{3}\log\frac{1}{3} + \frac{2}{3}\log\frac{2}{3} \underset{m\to\infty}{\longrightarrow} \\ &\log(m) \cdot \left(h_b\left(\frac{1}{6}\right) - \frac{1}{3}\right) - h_b\left(\frac{1}{3}\right). \end{aligned} \tag{1}$$

Therefore, $\max_{\underline{p}} C(\underline{p}, g_{opt}) = \Theta(\log(m)) = \Theta(d)$. □

2) Average-case analysis: In this section we show that the expected value of $C(\underline{p}, g_{opt})$ is bounded by a small constant, when averaging uniformly over all possible $\underline{p}$ over an alphabet size m.

To prove this, we recall that $C(\underline{p}, g_{opt}) \leq C(\underline{p}, g_{ord})$ for any given probability distribution $\underline{p}$. Therefore, we would like to find the expectation of $C(\underline{p}, g_{ord})$ where the random variables are $p_1, \ldots, p_m$, taking values over a uniform simplex.

Proposition 1: Let $\underline{X} \sim \underline{p}$ be a random vector of an alphabet size m and a joint probability distribution $\underline{p}$. The expected joint entropy of $\underline{X}$, where the expectation is over a uniform simplex of joint probability distributions $\underline{p}$ is

$$\mathbb{E}_{\underline{p}}\{H(\underline{X})\} = \frac{1}{\log_e 2}\left(\psi(m+1) - \psi(2)\right)$$

where ψ is the *digamma function*. □

Proof We first notice that a uniform distribution over a simplex of a size m is equivalent to a Direchlet distribution with parameters $\alpha_i = 1, i = 1, \ldots, m$. The Direchlet distribution can be generated through normalized independent random variables from a Gamma distribution. This means that for statistically independent $Z_i \sim \Gamma(k_i = 1, \theta_i = 1), i = 1, \ldots, m$ we have that

$$\frac{1}{\sum_{k=1}^{m} Z_k}\left(Z_1, \ldots Z_m\right) \sim Dir\left(\alpha_1 = 1, \ldots, \alpha_m = 1\right). \tag{2}$$

We are interested in the expected joint entropy of draws from (2),

$$\begin{aligned} \mathbb{E}_{\underline{p}}\{H(\underline{X})\} = &-\sum_{i=1}^{m} \mathbb{E}\left\{\frac{Z_i}{\sum_{k=1}^{m} Z_k} \log \frac{Z_i}{\sum_{k=1}^{m} Z_k}\right\} = \\ &- m\mathbb{E}\left\{\frac{Z_i}{\sum_{k=1}^{m} Z_k} \log \frac{Z_i}{\sum_{k=1}^{m} Z_k}\right\} \end{aligned} \tag{3}$$

It can be shown that for two independent Gamma distributed random variables $X_1 \sim \Gamma(\alpha_1, \theta)$ and $X_2 \sim \Gamma(\alpha_2, \theta)$, the ratio $\frac{X_1}{X_1+X_2}$ follows a Beta distribution with parameters (α_1, α_2). Let us denote $\tilde{Z}_i \triangleq \frac{Z_i}{\sum_{k=1}^{m} Z_k} = \frac{Z_i}{Z_i + \sum_{k\neq i} Z_k}$. Notice that $Z_i \sim \Gamma(1,1)$ and $\sum_{k\neq i} Z_i \sim \Gamma(m-1, 1)$ are mutually independent. Therefore,

$$f_{\tilde{Z}_i}(z) = Beta(1, m-1) = \frac{(1-z)^{(m-2)}}{B(1, m-1)}. \tag{4}$$

This means that

$$\begin{aligned} &\mathbb{E}\left\{\frac{Z_i}{\sum_{k=1}^{m} Z_k} \log \frac{Z_i}{\sum_{k=1}^{m} Z_k}\right\} = \mathbb{E}\left\{\tilde{Z}_i \log \tilde{Z}_i\right\} = \\ &\frac{1}{B(1, m-1)} \int_0^1 z \log(z)(1-z)^{(m-2)} dz = \\ &\frac{B(2, m-1)}{B(1, m-1)} \frac{1}{\log_e(2)} \frac{1}{B(2, m-1)} \int_0^1 \log_e(z) z (1-z)^{(m-2)} dz = \\ &\frac{1}{m \log_e(2)} \mathbb{E}\left(\log_e(U)\right) \end{aligned} \tag{5}$$

where U follows a Beta distribution with parameters $(2, m-1)$. The expected natural logarithm of a Beta distributed random variable, $V \sim Beta(\alpha_1, \alpha_2)$, follows $\mathbb{E}\left(\log_e(V)\right) = \psi(\alpha_1) - \psi(\alpha_1 + \alpha_2)$ where ψ is the *digamma function*. Putting this together with (3) and (5) we attain

$$\mathbb{E}_{\underline{p}}\left\{H(\underline{X})\right\} = -m\mathbb{E}\left\{\frac{Z_i}{\sum_{k=1}^{m} Z_k}\log\frac{Z_i}{\sum_{k=1}^{m} Z_k}\right\} = \frac{1}{\log_e(2)}\left(\psi(m+1) - \psi(2)\right) \tag{6}$$

□

We now turn to examine the expected sum of the marginal entropies, $\sum_{j=1}^{d} H(Y_j)$ under the order permutation. As described above, the order permutation suggests sorting the probability distribution $p_1, \ldots, p_m$ in an ascending order, followed by mapping of the i^{th} symbol (in a binary representation) the i^{th} smallest probability. Let us denote $p_{(1)} \leq \cdots \leq p_{(m)}$ the ascending ordered probabilities $p_1, \ldots, p_m$. [1] show that the expected value of $p_{(i)}$ is

$$\mathbb{E}\left\{p_{(i)}\right\} = \frac{1}{m}\sum_{k=m+1-i}^{m}\frac{1}{k} = \frac{1}{m}\left(K_m - K_{m-i}\right) \tag{7}$$

where $K_m = \sum_{k=1}^{m}\frac{1}{k}$ is the Harmonic number. Denote the ascending ordered binary representation of all possible symbols in a matrix form $A \in \{0,1\}^{(m\times d)}$. This means that entry A_{ij} corresponds to the j^{th} bit in the i^{th} symbol, when the symbols are given in an ascending order. Therefore, the expected sum of the marginal entropies of $\underline{Y}$, when the expectation is over a uniform simplex of joint probability distributions p, follows

$$\begin{aligned}\mathbb{E}_{\underline{p}}\left\{\sum_{j=1}^{d} H(Y_j)\right\} \underset{(a)}{\leq} \sum_{j=1}^{d} h_b(\mathbb{E}_{\underline{p}}\{Y_j\}) \underset{(b)}{=} \sum_{j=1}^{d} h_b\left(\frac{1}{m}\sum_{i=1}^{m} A_{ij}\left(K_m - K_{m-i}\right)\right) \underset{(c)}{=} \\ \sum_{j=1}^{d} h_b\left(\frac{1}{2}K_m - \frac{1}{m}\sum_{i=1}^{m} A_{ij}K_{m-i}\right)\end{aligned} \tag{8}$$

where (a) follows from Jensen's inequality, (b) follows from (7) and (c) follows since $\sum_{i=1}^{m} A_{ij} = \frac{1}{2}$ for all $j = 1, \ldots, d$.

We now turn to derive asymptotic bounds of the expected difference between the sum of $\underline{Y}$'s marginal entropies and the joint entropy of $\underline{X}$.

Theorem 2: Let $\underline{X} \sim \underline{p}$ be a random vector of an alphabet size m and joint probability distribution $\underline{p}$. Let $\underline{Y} = g_{ord}(\underline{X})$ be the order permutation. For $d \geq 10$, the expected value of $C(\underline{p}, g_{ord})$, over a uniform simplex of joint probability distributions $\underline{p}$, satisfies

$$\mathbb{E}_{\underline{p}} C(\underline{p}, g_{ord}) = \mathbb{E}_{\underline{p}}\left\{\sum_{j=1}^{d} H(Y_j) - H(\underline{X})\right\} < 0.0162 + O\left(\frac{1}{m}\right)$$

Proof Let us first derive the expected marginal entropy of the least significant bit, $j = 1$, according to (8).

$$\begin{aligned}\mathbb{E}_{\underline{p}}\left\{H(Y_1)\right\} \leq & h_b\left(\frac{1}{2}K_m - \frac{1}{m}\sum_{i=1}^{m/2} K_{m-i}\right) = \\ & h_b\left(\frac{1}{2}K_m - \frac{1}{m}\left(\sum_{i=1}^{m-1} K_i - \sum_{i=1}^{\frac{m}{2}-1} K_i\right)\right) \underset{(a)}{=} \\ & h_b\left(\frac{1}{2}K_m - \frac{1}{m}\left(mK_m - m - \frac{m}{2}K_{\frac{m}{2}} + \frac{m}{2}\right)\right) = \\ & h_b\left(\frac{1}{2}\left(K_{\frac{m}{2}} - K_m + 1\right)\right) \underset{(b)}{\leq} \\ & h_b\left(\frac{1}{2}\log_e\left(\frac{1}{2}\right) + \frac{1}{2} + O\left(\frac{1}{m}\right)\right) \underset{(c)}{\leq} \\ & h_b\left(\frac{1}{2}\log_e\left(\frac{1}{2}\right) + \frac{1}{2}\right) + O\left(\frac{1}{m}\right) h_b'\left(\frac{1}{2}\log_e\left(\frac{1}{2}\right) + \frac{1}{2}\right) = \\ & h_b\left(\frac{1}{2}\log_e\left(\frac{1}{2}\right) + \frac{1}{2}\right) + O\left(\frac{1}{m}\right)\end{aligned} \tag{9}$$

where (a) and (b) follow the harmonic number properties: (a)

1) $\sum_{i=1}^{m} K_i = (m+1)K_{m+1} - (m+1)$

2) $\frac{1}{2(m+1)} < K_m - \log_e(m) - \gamma < \frac{1}{2m}$, where γ is the Euler-Mascheroni constant [2]
and (c) results from the concavity of the binary entropy.
Repeating the same derivation for different values of j, we attain

$$\begin{aligned}\mathbb{E}_{\underline{p}}\left\{H(Y_j)\right\} \leq & h_b\left(\frac{1}{2}K_m - \frac{1}{m}\sum_{l=1}^{2^j-1}(-1)^{l+1}\sum_{i=1}^{l\frac{m}{2^j}}K_{m-i}\right) = \\ & h_b\left(\frac{1}{2}K_m - \frac{1}{m}\sum_{l=1}^{2^j}(-1)^{l}\sum_{i=1}^{l\frac{m}{2^j}-1}K_{i}\right) = \\ & h_b\left(\frac{1}{2}K_m - \frac{1}{m}\sum_{l=1}^{2^j}(-1)^{l}\left(l\frac{m}{2^j}K_{l\frac{m}{2^j}} - l\frac{m}{2^j}\right)\right) < \\ & h_b\left(\sum_{i=1}^{2^j-1}(-1)^{i+1}\frac{i}{2^j}\log_e\left(\frac{i}{2^j}\right) + \frac{1}{2}\right) + O\left(\frac{1}{m}\right) \qquad \forall j = 1,\ldots,d.\end{aligned} \tag{10}$$

We may now evaluate the sum of expected marginal entropies of $\underline{Y}$. For simplicity of derivation let us obtain $\mathbb{E}_{\underline{p}}\left\{H(Y_j)\right\}$ for $j = 1,\ldots,10$ according to (10) and upper bound $\mathbb{E}_{\underline{p}}\left\{H(Y_j)\right\}$ for $j > 10$ with $h_b\left(\frac{1}{2}\right) = 1$. This means that for $d \geq 10$ we have

$$\begin{aligned}\mathbb{E}_{\underline{p}}\left\{\sum_{j=1}^{d}H(Y_j)\right\} < & \sum_{j=1}^{10}\mathbb{E}_{\underline{p}}\left(H\left\{Y_j\right\}\right) + \sum_{j=11}^{d}h_b\left(\frac{1}{2}\right) < \\ & 9.4063 + (d-10) + O\left(\frac{1}{m}\right).\end{aligned} \tag{11}$$

The expected joint entropy may also be expressed in a more compact manner. In Proposition 1 it is shown than $\mathbb{E}_{\underline{p}}\left\{H(\underline{X})\right\} = \frac{1}{\log_e 2}\left(\psi(m+1) - \psi(2)\right)$. Following the inequality in [2], the Digamma function, $\psi(m+1)$, is bounded from below by $\psi(m+1) = H_m - \gamma > \log_e(m) + \frac{1}{2(m+1)}$. Therefore, we conclude that for $d \geq 10$ we have that

$$\begin{aligned}\mathbb{E}_{\underline{p}}\left\{\sum_{j=1}^{d}H(Y_j) - H(\underline{X})\right\} < & 9.4063 + (d-10) - \log(m) + \\ & \frac{\psi(2)}{\log_e 2} + O\left(\frac{1}{m}\right) = 0.0162 + O\left(\frac{1}{m}\right)\end{aligned} \tag{12}$$

□

In addition, we would like to evaluate the expected difference between the sum of marginal entropies and the joint entropy of $\underline{X}$, that is, without applying any permutation. This shall serve us as a reference to the upper bound we achieve in Theorem 2.

Theorem 3: Let $\underline{X} \sim \underline{p}$ be a random vector of an alphabet size m and joint probability distribution $\underline{p}$. The expected difference between the sum of marginal entropies and the joint entropy of $\underline{X}$, when the expectation is taken over a uniform simplex of joint probability distributions $\underline{p}$, satisfies

$$\mathbb{E}_{\underline{p}}\left\{\sum_{j=1}^{d}H(X_j) - H(\underline{X})\right\} < \frac{\psi(2)}{\log_e 2} = 0.6099$$

Proof We first notice that $P\left(X_j = 1\right)$ equals the sum of one half of the probabilities $p_i, i = 1,\ldots,m$ for every $j = 1\ldots d$. Assume p_i's are randomly (and uniformly) assigned to each of the m symbols. Then, $\mathbb{E}\{P\left(X_j = 1\right)\} = \frac{1}{2}$ for every $j = 1\ldots d$. Hence,

$$\begin{aligned}\mathbb{E}_{\underline{p}}\left\{\sum_{j=1}^{d}H(X_j) - H(\underline{X})\right\} = & \sum_{j=1}^{d}\mathbb{E}_{\underline{p}}\left\{H_b(X_j)\right\} - \mathbb{E}_{\underline{p}}\{H(\underline{X})\} < \\ & d - \log(m) + \frac{1}{\log_e 2}\left(\psi(2) - \frac{1}{2(m+1)}\right) < \frac{\psi(2)}{\log_e 2}\end{aligned}$$

□

To conclude, we show that for a random vector $\underline{X}$ over an alphabet size m, we have

$$\mathbb{E}_{\underline{p}} C(\underline{p}, g_{opt}) \leq \mathbb{E}_{\underline{p}} C(\underline{p}, g_{bst}) \leq \mathbb{E}_{\underline{p}} C(\underline{p}, g_{ord}) < 0.0162 + O\left(\frac{1}{m}\right)$$

for $d \geq 10$, where the expectation is over a uniform simplex of joint probability distributions $\underline{p}$. This means that when the alphabet size is large enough, even the simple order permutation achieves, on the average, a sum of marginal entropies which is only 0.0162 bits greater than the joint entropy, when all possible probability distributions $\underline{p}$ are equally likely to appear. Moreover, we show that the simple order permutation reduced the expected difference between the sum of the marginal entropies and the joint entropy of $\underline{X}$ by more than half a bit, for sufficiently large m.

REFERENCES

[1] I. Bairamov, A. Berred, and A. Stepanov, "Limit results for ordered uniform spacings," *Statistical Papers*, vol. 51, no. 1, pp. 227–240, 2010.
[2] R. M. Young, "75.9 euler's constant," *The Mathematical Gazette*, pp. 187–190, 1991.