Practical Coding Scheme for Universal Source Coding with Side Information at the Decoder

Elsa Dupraz, Aline Roumy, Michel Kieffer

Practical Coding Scheme for Universal Source Coding with Side Information at the Decoder

Elsa Dupraz[†], Aline Roumy[+] and Michel Kieffer[†,−]

[†] LSS - CNRS - SUPELEC - Univ Paris-Sud, 91192 Gif-sur-Yvette, France

[+] INRIA, Campus de Beaulieu, 35042 Rennes, France

[−] Institut universitaire de France

Abstract

This paper considers the problem of universal lossless source coding with side information at the decoder only. The correlation channel between the source and the side information is unknown and belongs to a class parametrized by some unknown parameter vector. A complete coding scheme is proposed that works well for any distribution in the class. At the encoder, the proposed scheme encompasses the determination of the coding rate and the design of the encoding process. Both contributions result from the information theoretical compression bounds of universal lossless source coding with side information. Then a novel decoder is proposed that takes into account the available information regarding the class. The proposed scheme avoids the use of a feedback channel or the transmission of a learning sequence, which both would result in a rate increase at finite length. Simulations are based on irregular and non-binary LDPC codes and show that, even at finite length, the proposed scheme performs close to the theoretical bound.

1 Introduction

The problem of lossless source coding with side information at the decoder has been well investigated when the correlation model between the source X and the side information (SI) Y is perfectly known. Several works, see, *e.g.*, [13, 19], propose practical coding schemes for the Slepian-Wolf (SW) problem. Most of them are based on channel codes [18], and particularly Low Density Parity Check (LDPC) codes [12]. This approach allows to leverage on many results on LDPC codes for the code construction and optimization [11, 14] even if there is a need to adapt the algorithms developed for channel coding to the case of SW coding [3].

Nonetheless, most of these works assume perfect knowledge of the correlation channel $Y|X$ between the source and the side information. This assumption is difficult to satisfy in practical situations such as video coding and sensor networks, due to the varying nature of the characteristics of the real signals. A usual solution to address

this problem is to use a feedback channel [1] or to allow interactions between the encoder and the decoder [20]. These interactions allow the encoder and the decoder to exchange information on the rate needed and on the correlation channel. These solutions are however difficult to implement in many practical situations such as sensor networks. Furthermore, solutions based on learning sequences [6] induce a rate increase at finite length.

Alternatively, universal coding schemes supposed to be able to decode the source whatever the correlation channel may be considered. Performance bounds for the universal setup are provided in [4]. We address the problem of constructing a practical universal coding scheme for the SW setup. At the encoder part, the rate has to be chosen and the coding process has to be designed. At the decoder part, the source has to be reconstructed despite the lack of knowledge on the correlation. When no feedback or learning sequence is allowed, several practical solutions based on LDPC codes and proposed for channel coding may be adapted to the SW problem. When hard decoding is performed, as suggested by [6] only symbol values are used, at the price of an important loss in performance. An alternative solution is the min-sum decoding algorithm proposed in [2, 15] for channel coding, respectively for binary and non-binary sources. The min-sum algorithm uses soft information for decoding, but does not require the knowledge of the correlation channel. The min-sum algorithm may be as efficient as the soft decoding algorithm, provided that a coefficient is chosen carefully. Unfortunately this choice depends on the unknown correlation channel.

In many applications, it is possible to restrict the correlation channel model to a given class (*e.g.*, binary symmetric, Gaussian etc.) due no the nature of the problem. Consequently in this paper, the universality is modeled by assuming that the correlation channel belongs to a given class and is parametrized by some unknown parameter vector $\boldsymbol{\theta}$. Hard and min-sum decoding are not able to exploit the knowledge of the structure of the class. The coding scheme we propose is based on non-binary LDPC codes. From an analysis of the performance bounds, we explain how to choose the coding rate and the LDPC coding matrix. Then, we propose a decoding algorithm that performs joint estimation of the parameter vector and of the source sequence with an Expectation Maximization (EM) algorithm. Furthermore, the main problem of the EM algorithm is its sensitivity to initialization. A method to produce a first raw estimate of the parameters is thus also provided.

The paper is organized as follows. Section 2 introduces the considered universal model. Section 3 presents an adaptation of the non-binary LDPC decoding algorithm for the SW problem. Section 4 describes the practical scheme we propose. To finish, Section 5 evaluates the performance of the considered scheme through simulations.

2 Model and performance

The source X to be compressed and the SI Y available at the decoder produce sequences of symbols $\{X_n\}_{n=1}^{+\infty}$ and $\{Y_n\}_{n=1}^{+\infty}$, respectively. $\mathcal{X}$ and $\mathcal{Y}$ denote the source and SI discrete alphabets. In this paper, we mainly consider the case where $\mathcal{X} = \mathcal{Y} =$ GF(q), the Galois Field of size q. Bold upper-case letters, e.g., $\mathbf{X}_1^N = \{X_n\}_{n=1}^N$, de-

note random vectors, whereas bold lower case letters, $\mathbf{x}_1^N = \{x_n\}_{n=1}^N$, represent their realizations. Moreover, when it is clear from the context that the distribution of a random variable X_n does not depend on n, the index n is omitted. Similarly, $\mathbf{X}_1^N$ is in general denoted $\mathbf{X}$.

In the universal setup we consider, the correlation channel is parametrized by an unknown vector $\boldsymbol{\theta}$. It is assumed fixed for a sequence $\{(X_n, Y_n)\}_{n=1}^{+\infty}$ but it is allowed to vary from sequence to sequence. Formally,

Definition 1. (WP-Source). *A* source (X, Y) Without Prior (WP-Source) *produces a sequence of independent symbols* $\{(X_n, Y_n)\}_{n=1}^{+\infty}$ *drawn from a distribution belonging to a family* $\{P(X, Y|\boldsymbol{\theta}) = P(X)P(Y|X, \boldsymbol{\theta})\}_{\boldsymbol{\theta}\in\mathcal{P}_{\boldsymbol{\theta}}}$ *parametrized by a vector* $\boldsymbol{\theta}$. *The vector* $\boldsymbol{\theta}$ *takes its value in a set* $\mathcal{P}_{\boldsymbol{\theta}}$ *that is either discrete or continuous. The source symbols* X *and* Y *take their values in the discrete sets* $\mathcal{X}$ *and* $\mathcal{Y}$*, respectively. Moreover, the parameter* $\boldsymbol{\theta}$ *is fixed for the sequence* $\{(X_n, Y_n)\}_{n=1}^{+\infty}$.

The WP-source, completely determined by $\mathcal{P}_{\boldsymbol{\theta}}$ and $\{P(X, Y|\boldsymbol{\theta})\}_{\boldsymbol{\theta}\in\mathcal{P}_{\boldsymbol{\theta}}}$, is stationary but non-ergodic [8, Section 3.5]. No distribution for $\boldsymbol{\theta}$ is specified, either because such a distribution is not known or because $\boldsymbol{\theta}$ cannot be modeled as a random variable.

For the WP-Source, the infimum of achievable rates in lossless SW coding is [4]

$$R_{X|Y}^{\mathrm{SW}} = \sup_{\boldsymbol{\theta}\in\mathcal{P}_{\boldsymbol{\theta}}} H(X|Y, \boldsymbol{\theta}) \ . \tag{1}$$

This result shows that the encoder (rate and coding matrix) has to be designed for the worst parameter case. However, since classical decoding algorithm require the knowledge of the true correlation channel, *i.e.*, $\boldsymbol{\theta}$, we propose a practical scheme capable of dealing with the lack of knowledge of the parameter at the decoder.

3 LDPC codes

LDPC codes are binary [7] or non-binary [5] linear correcting codes. In [12], they have been adapted to SW coding for binary sources with perfect model knowledge. This section generalizes the adaptation of LDPC codes to the SW non-binary case.

The SW encoding of a vector $\mathbf{x}$ of length N is performed by producing a vector $\mathbf{s}$ of length $M < N$ as $\mathbf{s} = H^T\mathbf{x}$. The matrix H is sparse, with non-zero coefficients uniformly distributed in $\mathrm{GF}(q)\backslash\{0\}$. In the following, $\oplus$, $\ominus$, $\otimes$, $\oslash$ are the usual operators in $\mathrm{GF}(q)$. In the bipartite graph representing the dependences between the random variables of $\mathbf{X}$ and $\mathbf{S}$, the entries of $\mathbf{X}$ are represented by Variable Nodes (VN) and the entries of $\mathbf{S}$ are represented by Check Nodes (CN). The set of CN connected to a VN n is denoted $\mathcal{N}(n)$ and the set of VN connected to a CN m is denoted $\mathcal{N}(m)$. The sparsity of H is determined by the VN degree distribution $\lambda(x) = \sum_{i\geq 2}\lambda_i x^{i-1}$ and the CN degree distribution $\rho(x) = \sum_{i\geq 2}\rho_i x^{i-1}$ with $\sum_{i\geq 2}\lambda_i = 1$ and $\sum_{i\geq 2}\rho_i = 1$. In SW coding, the rate $r(\lambda, \rho)$ of a code is given by $r(\lambda, \rho) = \frac{M}{N} = \frac{\sum_{i\geq 2}\rho_i/i}{\sum_{i\geq 2}\lambda_i/i}$.

The decoder performs a Maximum *A Posteriori* (MAP) estimation of $\mathbf{x}$ from the received codeword $\mathbf{s}$ and the observed side information $\mathbf{y}$ via a Message Passing (MP)

algorithm. The messages exchanged in the dependency graph are vectors of length q. The initial messages for each VN n are denoted $\mathbf{m}^{(0)}(n, y_n)$, with components

$$m_k^{(0)}(n, y_n) = \log \frac{P(X_n = 0|Y_n = y_n)}{P(X_n = k|Y_n = y_n)} \ . \tag{2}$$

The messages from CN to VN are computed with the help of a particular Fourier transform, denoted $\mathcal{F}(\mathbf{m})$. Its i-th component is given by [11]

$$\mathcal{F}_i(\mathbf{m}) = \frac{\sum_{j=0}^{q-1} r^{i \otimes j} e^{-m_j}}{\sum_{j=0}^{q-1} e^{-m_j}} \tag{3}$$

where r is the unit-root associated to GF(q). At iteration ℓ, the message $\mathbf{m}^{(\ell)}(m, n, s_m)$ from a CN m to a VN n is

$$\mathbf{m}^{(\ell)}(m, n, s_m) = \mathcal{A}[\bar{s}_m]\mathcal{F}^{-1}\left(\prod_{n' \in \mathcal{N}(m)\backslash n} \mathcal{F}\left(W\left[\overline{H}_{n'm}\right]\mathbf{m}^{(\ell-1)}(n', m, y_{n'})\right)\right) \tag{4}$$

where $\bar{s}_m = \ominus s_m \oslash H_{n,m}$, $\overline{H}_{n'm} = \ominus H_{n',m} \oslash H_{n,m}$ and $W[a]$ is a $q \times q$ matrix such that $W[a]_{k,n} = \delta(a \otimes n \ominus k), \forall 0 \leq k, n \leq q-1$. $\mathcal{A}[k]$ is a $q \times q$ matrix that maps a vector message $\mathbf{m}$ into a vector message $\mathbf{l} = \mathcal{A}[k]\mathbf{m}$ with $l_j = m_{j \oplus k} - m_k$. Note that the matrix $\mathcal{A}$ does not appear in the channel coding version of the algorithm and is specific to SW coding. At a VN n, a message $\mathbf{m}^{(\ell)}(n, m, y_i)$ is sent to the CN m and an *a posteriori* message $\tilde{\mathbf{m}}^{(\ell)}(n, y_n)$ is computed. The both satisfy:

$$\mathbf{m}^{(\ell)}(n, m, y_n) = \sum_{m' \in \mathcal{N}(n)\backslash m} \mathbf{m}^{(\ell)}(m', n, s_{m'}) + \mathbf{m}^{(0)}(n, y_n) \ , \tag{5}$$

$$\tilde{\mathbf{m}}^{(\ell)}(n, y_n) = \sum_{m' \in \mathcal{N}(n)} \mathbf{m}^{(\ell)}(m', n, s_{m'}) + \mathbf{m}^{(0)}(n, y_n) \ . \tag{6}$$

The algorithm ends if $\mathbf{s} = H^T\widehat{\mathbf{x}}^{(\ell)}$ or if the maximum number of iterations $L_{\max}$ is reached.

4 Practical Coding Scheme

When $\boldsymbol{\theta}$ is unknown, the LDPC decoding algorithm cannot be applied directly, since the initial messages (2) depend on $\boldsymbol{\theta}$. Therefore an EM algorithm jointly estimates the source vector $\mathbf{x}$ and the parameter vector $\boldsymbol{\theta}$. An efficient initialization is proposed for the EM, this being a known issue.

4.1 Joint estimation of $\boldsymbol{\theta}$ and $\mathbf{x}$

The joint estimation of the source vector $\mathbf{x}$ and the parameter $\boldsymbol{\theta}$ from the observed vectors $\mathbf{y}$ and $\mathbf{s}$ is performed via the EM algorithm [9]. Knowing some estimate $\boldsymbol{\theta}^{(\ell)}$

obtained at iteration ℓ, the EM algorithm maximizes, with respect to $\boldsymbol{\theta}$,

$$Q(\boldsymbol{\theta}, \boldsymbol{\theta}^{(\ell)}) = E_{\mathbf{X}|\mathbf{y},\mathbf{s},\boldsymbol{\theta}^{(\ell)}}\left[\log P(\mathbf{X}|\mathbf{y},\mathbf{s},\boldsymbol{\theta})\right] = \sum_{\mathbf{x}\in \mathrm{GF}(q)^n} P(\mathbf{x}|\mathbf{y},\mathbf{s},\boldsymbol{\theta}^{(\ell)}) \log P(\mathbf{y}|\mathbf{x},\mathbf{s},\boldsymbol{\theta}) \quad (7)$$

$$= \sum_{n=1}^{N}\sum_{k=0}^{q-1} P(X_n = k|y_n,\mathbf{s},\boldsymbol{\theta}^{(\ell)}) \log P(y_n|X_n = k,\boldsymbol{\theta}) \ .$$

Solving this maximization gives the update rules detailed in Lemma 1.

Lemma 1. *Let (X, Y) be a binary WP-Source. Let the correlation channel be a Binary Symmetric channel (BSC) with parameter $\theta = P(Y = 0|X = 1) = P(Y = 1|X = 0)$, $\theta \in [0, 1]$. The update equation for the EM algorithm is [17]*

$$\theta^{(\ell+1)} = \frac{1}{N}\sum_{n=1}^{N} |y_n - p_n^{(\ell)}| \quad (8)$$

where $p_n^{(\ell)} = P(X_n = 1|y_n, \mathbf{s}, \theta^{(\ell)})$.
Let (X, Y) be a WP-Source that generates symbols in GF(q). Let the correlation channel be such that $Y = X \oplus Z$, where Z is a random variable in GF(q), and $P(Z = k) = \theta_k$. The update equations for the EM algorithm are

$$\forall k \in GF(q),\ \theta_k^{(\ell+1)} = \frac{\sum_{n=1}^{N} P_{y_n\ominus k,n}^{(\ell)}}{\sum_{n=1}^{N}\sum_{k'=0}^{q-1} P_{y_n\ominus k',n}^{(\ell)}} \quad (9)$$

where $P_{k,n}^{(\ell)} = P(X_n = k|y_n, \mathbf{s}, \boldsymbol{\theta}^{(\ell)})$.

Proof. The binary case is provided by [17]. In the non-binary case, the updated estimate is obtained by maximizing (7) taking into account the constraints $0 \le \theta_k \le 1$ and $\sum_{k=0}^{q-1}\theta_k = 1$. The $P_{k,n}^{(\ell)} = P(X_n = k|y_n, \mathbf{s}, \boldsymbol{\theta}^{(\ell)})$ are obtained from the LDPC decoder considering that the true parameter is $\boldsymbol{\theta}^{(\ell)}$. □

4.2 Initialization of the EM algorithm

We now propose an efficient initialization of the EM algorithm valid for irregular codes and for sources X, Y taking values in $\mathrm{GF}(q)$. This generalizes the method proposed in [17] for regular and binary codes. The rationale is to derive a Maximum Likelihood (ML) estimate of a subpart $\mathbf{u} = H^T\mathbf{x} \oplus H^T\mathbf{y}$ of the observed data ($H^T\mathbf{x}$ and $\mathbf{y}$).

4.2.1 The BSC with irregular codes

Let Z denote the error of the BSC, *i.e.* $Z = X\oplus$, and let $\mathbf{U} = H^T\mathbf{Z}$, for a given matrix H.

In this case, each binary random variable U_m is the sum of random variables of $\mathbf{Z}$. Although each Z_n appears in several sums, we assume that each U_m is the sum of *i.i.d.* random variables $Z_j^{(m)}$. The validity of this assumption depends on the choice of the

matrix H and is not true in general. Although it produces a suboptimal solution, this choice may lead to a reasonable initialization for the EM algorithm. Furthermore, the number of terms in the sum for U_m depends on the degree of the CN m. One can use the CN degree distribution $\rho(x)$ as a probability distribution for these degrees, or decide to take into account the knowledge of the CN degrees. Both cases lead to a probability model for the U_m and enable to obtain an ML estimate for θ, as described in the two following lemmas.

Lemma 2. *Consider a vector* $\mathbf{U}$ *of* M *binary random variables such that each* U_m *is the sum of* J_m i.i.d. *binary random variables* $Z_j^{(m)}$ i.e. $U_m = \sum_{j=1}^{J_m} Z_j^{(m)}$. J_m *is an* i.i.d. *random variable taking its values in* $\{2, \dots, d_c\}$ *with known probability* $P(J = j) = \rho_j$. *Denote* $\theta = P(Z = 1)$, $\alpha = P(U = 1)$ *and assume that* θ *and* α *are unknown. Denote* $\widehat{\theta}$ *and* $\widehat{\alpha}$ *there respective ML estimates from an observed vector* $\mathbf{u}$. *Let* $f(\theta) = \frac{1}{2} - \frac{1}{2}\sum_{j=2}^{d_c} \rho_j (1-2\theta)^j$, $\forall \theta \in [0, \frac{1}{2}]$. *Then the random variables of* $\mathbf{U}$ *are* i.i.d., $\widehat{\alpha} = \frac{1}{M}\sum_{m=1}^{M} u_m$, $\alpha = f(\theta)$, f *is invertible, and* $\widehat{\theta} = f^{-1}(\widehat{\alpha})$.

Proof. The random variables U_m are independent because they are obtained from sums of independent variables. They are identically distributed because the J_m and the Z_j^m are identically distributed. are such that $\alpha = P(U = 1) = \sum_{j=2}^{d_c} \rho_j P(U = 1|J = j)$. Then, from [17], $P(U = 1|J = j) = \sum_{i=1, i \text{ odd}}^{j} \binom{j}{i} \theta^i (1-\theta)^{j-i}$ and from [7, Section 3.8], $P(U = 1|J = j) = \frac{1}{2} - \frac{1}{2}(1-2\theta)^j$. Thus $\alpha = f(\theta)$. A ML estimate $\widehat{\alpha}$ of α can be obtained as $\widehat{\alpha} = \frac{1}{M}\sum_{m=1}^{M} u_m$. Finally, as f is obviously invertible for $\left[0, \frac{1}{2}\right]$, then from[10, Theorem 7.2], the ML estimate of θ is given by $\widehat{\theta} = f^{-1}(\widehat{\alpha})$. □

Lemma 3. *Consider a vector* $\mathbf{U}$ *of* M *binary random variables such that each* U_m *is the sum of* j_m i.i.d. *binary random variables* $Z_j^{(m)}$, i.e., $U_m = \sum_{j=1}^{j_m} Z_j^{(m)}$. *The values of* j_m *are known and belong to* $\{2, \dots, d_c\}$. *Denote* $\theta = P(Z = 1)$ *and assume that* θ *is unknown. Then the entries of* $\mathbf{U}$ *are independent and the ML estimate* $\widehat{\theta}$ *from an observed vector* $\mathbf{u}$ *is the argument of the maximum of*

$$L(\theta) = \sum_{j=2}^{d_c} \mathbb{N}_{1,j}(\mathbf{u}) \log\left(\frac{1}{2} - \frac{1}{2}(1-2\theta)^j\right) + \sum_{j=2}^{d_c} \mathbb{N}_{0,j}(\mathbf{u}) \log\left(\frac{1}{2} + \frac{1}{2}(1-2\theta)^j\right) \quad (10)$$

where $\mathbb{N}_{1,j}(\mathbf{u})$ *and* $\mathbb{N}_{0,j}(\mathbf{u})$ *are the number of symbols in* $\mathbf{u}$ *obtained from the sum of* j *elements and respectively equal to* 1 *and* 0.

Proof. The random variables U_m are independent because they are obtained from sums of independent variables. We maximize the likelihood function

$$L(\theta) = \log P(\mathbf{u}|\theta) = \sum_{m=1}^{M} \log P(u_m|j_m, \theta) \quad (11)$$

with respect to θ. The second equality in (11) comes from the independence of the symbols. Then, (10) is obtained from the proof of Lemma 2. □

The method of Lemma 2 is simpler to implement but does not take into account the knowledge of the matrix H, at the price of a small loss in performance.

4.2.2 The non-binary discrete case

Only the case of a regular code is presented here, but the method can be easily generalized to irregular codes (see the previous section). Now, the probability mass function of Z is given by $\boldsymbol{\theta} = [\theta_0 \dots \theta_{q-1}]$ with $\theta_k = P(Z = k)\ \forall k \in \mathrm{GF}(q)$. As for the case of binary irregular codes, two signal models are first introduced through two lemmas. Now, each U_m is the sum of symbols of $\mathbf{Z}$, weighted by the coefficients contained in H. A first solution does not exploit the knowledge of these coefficients, but uses the fact that the non-zero coefficients of H are distributed uniformly in $\mathrm{GF}(q)\backslash\{0\}$ (Lemma 4). A second solution takes into account the knowledge of the coefficients (Lemma 5).

Lemma 4. *Consider a vector* $\mathbf{U}$ *of* M *random variables of GF(q) such that each* U_m *is the sum of* d_c i.i.d. *products of random variables,* i.e., $U_m = \sum_{j=1}^{d_c} H_j^{(m)} Z_j^{(m)}$. *The* $Z_j^{(m)}$ *and* $H_j^{(m)}$ *are independent and* i.i.d. *random variables.* H *is uniformly distributed in GF(q)\{0}. The of* $Z_j^{(m)}$ *take their values in GF(q). Denote* $\theta_k = P(Z = k)$, $\alpha_k = P(U = k)$ *and assume that* $\boldsymbol{\theta} = [\theta_0 \dots \theta_{q-1}]$ *and* $\boldsymbol{\alpha} = [\alpha_0 \dots \alpha_{q-1}]$ *are unknown. Denote* $\widehat{\boldsymbol{\theta}}$ *and* $\widehat{\boldsymbol{\alpha}}$ *there respective ML estimates from an observed vector* $\mathbf{u}$, *with* $\widehat{\alpha}_k = \frac{\mathbb{N}_k(\mathbf{u})}{M}$ *where* $\mathbb{N}_k(\mathbf{u})$ *is the number of occurrences of* k *in the vector* $\mathbf{u}$. *Let*

$$f(\boldsymbol{\theta}) = \sum_{n_0 \dots n_{q-1}} \binom{d_c}{n_0 \dots n_{q-1}} \left(\frac{1}{q}\right)^{d_c} \mathcal{F}^{-1}\left(\prod_{j=0}^{q-1} \left(\mathcal{F}\left(W[j]\boldsymbol{\theta}\right)\right)^{n_j}\right) \tag{12}$$

where the sum is on all the possible combinations of integers $n_0 \dots n_{q-1}$ *such that* $0 \leq n_k \leq d_c$ *and* $\sum_{k=0}^{q-1} n_k = d_c$. *Then the random variables of* $\mathbf{U}$ *are independent,* $\boldsymbol{\alpha} = f(\boldsymbol{\theta})$, *and if* f *is invertible,* $\widehat{\boldsymbol{\theta}} = f^{-1}(\widehat{\boldsymbol{\alpha}})$.

Proof. The random variables U_m are independent because they are obtained from sums of independent variables and $U_m = \sum_{j=1}^{d_c} H_j^{(m)} Z_j^{(m)}$. One has $\alpha_k = P(U = k) = \sum_{\{h_j\}_{j=1}^{d_c}} P(\{h_j\}_{j=1}^{d_c}) P(U = k | \{h_j\}_{j=1}^{d_c})$ in which the sum is on all the possible combinations of coefficients $\{h_j\}_{j=1}^{d_c}$. This can be simplified as $\alpha_k = \sum_{n_0 \dots n_{q-1}} P(N_0 = n_0 \dots N_{q-1} = n_{q-1}) P(U = k | n_0 \dots n_{q-1})$ where n_k is the number of occurences of k in a combination $\{h_j\}_{j=1}^{d_c}$. One has $P(N_0 = n_0 \dots N_{q-1} = n_{q-1}) = \binom{d_c}{n_0 \dots n_{q-1}} \left(\frac{1}{q}\right)^{d_c}$ Then, the vector denoted

$$P_{\mathbf{U}|n_0 \dots n_{q-1}} = [P(U = 0 | n_0 \dots n_{q-1}) \dots P(U = q-1 | n_0 \dots n_{q-1})] \tag{13}$$

can be expressed as $P_{\mathbf{U}|n_0 \dots n_{q-1}} = \mathcal{F}^{-1}\left(\prod_{j=0}^{q-1} \left(\mathcal{F}\left(W[j]\boldsymbol{\theta}\right)\right)^{n_j}\right)$. Therefore,

$$\boldsymbol{\alpha} = [\alpha_0 \dots \alpha_{q-1}] = \sum_{n_0 \dots n_{q-1}} \binom{d_c}{n_0 \dots n_{q-1}} \left(\frac{1}{q}\right)^{d_c} \mathcal{F}^{-1}\left(\prod_{j=0}^{q-1} \left(\mathcal{F}\left(W[j]\boldsymbol{\theta}\right)\right)^{n_j}\right) . \tag{14}$$

The ML estimates $\widehat{\alpha}_k$ of α_k are $\widehat{\alpha}_k = \frac{\mathbb{N}_k(\mathbf{u})}{M}$. Finally, if f is invertible, then from [10, Theorem 7.2] the ML estimate of $\boldsymbol{\theta}$ is given by $\widehat{\boldsymbol{\theta}} = f^{-1}(\widehat{\boldsymbol{\alpha}})$ □

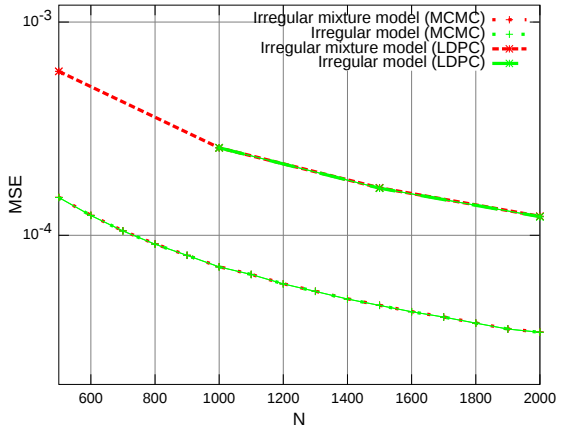

Figure 1: MSE for the binary irregular case

Lemma 5. *Consider a vector* $\mathbf{U}$ *of* M *random variables of* $GF(q)$ *such that each* U_m *is the sum of* d_c i.i.d. *random variables such that* $U_m = \sum_{j=1}^{d_c} h_j^{(m)} Z_j^{(m)}$. *The* $Z_j^{(m)}$ *are* i.i.d. *random variables taking their values in GF(q). The values of the coefficients* $h_j^{(m)}$ *are known and belong to* $GF(q)\backslash\{0\}$. *Denote* $\theta_k = P(Z = k)$, $\alpha_k = P(U = k)$ *and assume that* $\boldsymbol{\theta} = [\theta_0 \dots \theta_{q-1}]$ *and* $\boldsymbol{\alpha} = [\alpha_0 \dots \alpha_{q-1}]$ *are unknown. Then the random variables of* $\mathbf{U}$ *are independent and the ML estimate* $\widehat{\boldsymbol{\theta}}$ *from an observed vector* $\mathbf{u}$ *is the* $\boldsymbol{\theta}$ *that maximizes*

$$L(\boldsymbol{\theta}) = \sum_{m=1}^{M} \log \mathcal{F}_m^{-1} \left(\prod_{j=1}^{dc} \mathcal{F}(W[h_{s_{zm},j}]\boldsymbol{\theta}) \right) \tag{15}$$

and satisfies $0 \leq \theta_k \leq 1$ *and* $\sum_{k=0}^{q-1} \theta_k = 1$.

Proof. The random variables U_m are independent because they are obtained from sums of independent variables The ML estimate $\widehat{\boldsymbol{\theta}}$ is the value that maximizes the likelihood function given by

$$L(\boldsymbol{\theta}) = \log P(\mathbf{u}|\boldsymbol{\theta}, \{h_j^{(m)}\}_{j=1,m=1}^{d_c,M}) = \sum_{m=1}^{M} \log P(u_m|\boldsymbol{\theta}, \{h_j^{(m)}\}_{j=1}^{d_c}) \tag{16}$$

with respect to $0 \leq \theta_k \leq 1$ and $\sum_{k=0}^{q-1} \theta_k = 1$. The second equality comes from the independence assumption. Following the steps of Lemma 4, we show that (16) becomes $L(\boldsymbol{\theta}) = \sum_{m=1}^{M} \log \mathcal{F}_m^{-1} \left(\prod_{j=1}^{d_c} \mathcal{F}(W[h_j^{(m)}]\boldsymbol{\theta}) \right)$. □

5 Simulations

For the binary case, we consider a code $\lambda(x) = 0.4295x^3 + 0.2750x^4 + 0.0745x^{10} + 0.1150x^{11} + 0.0035x^{12} + 0.0930x^{16} + 0.0095x^{17}$, $\rho(x) = 0.2187x^7 + 0.7760x^8 + 0.0053x^9$

obtained from a code optimization realized with a differential evolution algorithm [16]. The rate of this code is $R = 0.75$ bit/symbol.

We first focus on the proposed initialization method. The ML performance of the models introduced in Lemmas 2 and 3 are first evaluated via Monte Carlo (MC) simulations [9]. More precisely, 50000 vectors $\mathbf{U}$ of length M are generated from the models defined in Lemmas 2 and 3, for $\theta = 0.1$. Then, the two proposed estimation methods are applied to each realization. We average the relative squared error $(\theta - \widehat{\theta})^2$ over the realizations, in order to obtain an approximate version of the Mean Squared Error (MSE) $E\left[(\theta - \widehat{\theta})^2\right]$. This gives the two stacked lower curves of Figure 1. They represents the approximate MSE for several values of $N = \frac{M}{R}$. The fact that the curves are stacked suggest that the ML performance is the same for the two models. Hence, one should choose the simpler model of Lemma 2 for the initialization of the EM algorithm. Then, 10000 vectors $\mathbf{Z}$ of length N are generated with respect to θ, and a matrix H of the defined code is applied to each vector. The two proposed estimation methods are applied to each realization. As before, an approximate version of the MSE is obtained. This gives the two stacked upper curves of Figure 1. We observe an important loss compared to the samples generated directly from the models. However, the performance seems sufficient for the initialization of the EM algorithm.

Second, we are interested in the EM algorithm. For a length $N = 10000$, we compare the Bit Error Rate (BER) of three setups over 100 realizations. The first setup is the genie-aided setup, *i.e*, the setup where the the true parameter θ is given to the decoder. The second setup corresponds to the EM algorithm initialized with the proposed method. The third setup corresponds to the EM algorithm initiliazed with a random θ. The results are presented in Figure **??**. The BER is plotted with respect to the number of iterations of the EM algorithm, for several values of θ. We see that the EM algorithm initialized properly converges faster than the one initialized at random, and that there is only a small loss compared to the genie aided setup.

6 Conclusion

This paper presents a universal Slepian-Wolf coding scheme based on non-binary LDPC codes. The proposed method allows to decode whatever the correlation channel in a given class by performing joint estimation of the source vector and of the parameter of the correlation channel. A method to initialize the EM algorithm realizing the joint estimation is also introduced.

The objective of future works is to provide tools based on density evolution methods to evaluate the performance of the proposed scheme. From such tools, one would be able to optimize the coding matrix both for the decoding of the source vector and for the estimation of the parameters.

References

[1] A. Aaron, R. Zhang, and B. Girod. Wyner-Ziv coding of motion video. In *Conference Record of the Thirty-Sixth Asilomar Conference on Signals, Systems and Computers, 2002.*, volume 1, pages 240–244, 2002.

[2] J. Chen and M.P.C. Fossorier. Near optimum universal belief propagation based decoding of low-density parity check codes. *IEEE Trans. on Comm.*, 50(3):406–414, 2002.

[3] J. Chen, D.K. He, and A. Jagmohan. The equivalence between Slepian-Wolf coding and channel coding under density evolution. *IEEE Trans. on Comm.*, 57(9):2534–2540, 2009.

[4] I. Csiszar. Linear codes for sources and source networks: Error exponents, universal coding. *IEEE Trans. on Inf. Th.*, 28(4):585–592, 1982.

[5] M.C. Davey and D.J.C. MacKay. Low Density Parity Check codes over GF (q). In *Information Theory Workshop*, pages 70–71. IEEE, 1998.

[6] E. Dupraz, A. Roumy, and M. Kieffer. Source coding with side information at the decoder : models with uncertainty, performance bounds, and practical coding schemes. 2012.

[7] R.G. Gallager. *Low-Density Parity Check Codes.* PhD thesis, Massachusetts Institute of Technology, 1963.

[8] R.G. Gallager. *Information theory and reliable communication.* Wiley, 1968.

[9] T. Hastie, R. Tibshirani, and J. Friedman. *The elements of statistical learning: data mining, inference and prediction.* Springer, 2009.

[10] S.M. Kay. *Fundamentals of Statistical Signal Processing, Estimation theory.* Prentice Hall PTR, 1993.

[11] G. Li, I.J. Fair, and W.A. Krzymien. Density evolution for nonbinary LDPC codes under Gaussian approximation. *IEEE Trans. on Inf. Th.*, 55(3):997–1015, 2009.

[12] A. Liveris, Z. Xiong, and C. Georghiades. Compression of binary sources with side information at the decoder using LDPC codes. *IEEE Comm. Letters*, 6:440–442, 2002.

[13] R. Puri and K. Ramchandran. PRISM: A new robust video coding architecture based on distributed compression principles. In *Proc. of the Annual Allerton Conf. on Comm. Cont. and Comp.*, volume 40, pages 586–595. Citeseer, 2002.

[14] T.J. Richardson and R.L. Urbanke. The capacity of Low-Density Parity-Check codes under message-passing decoding. *IEEE Trans. on Inf. Th.*, 47(2):599–618, 2001.

[15] V. Savin. Min-Max decoding for non binary LDPC codes. In *IEEE Int. Symp. on Inf. Th.*, pages 960–964. IEEE, 2008.

[16] Storn, R. and Price, K. Differential evolution– a simple and efficient heuristic for global optimization over continuous spaces. *Journal of global optimization*, 11(4):341–359, 1997.

[17] V. Toto-Zarasoa, A. Roumy, and C. Guillemot. Maximum Likelihood BSC Parameter Estimation for the Slepian-Wolf Problem. *IEEE Communications Letters*, 15(2), February 2011.

[18] V. Stankovic and A.D. Liveris and Z. Xiong and C.N. Georghiades. Design of Slepian-Wolf codes by channel code partitioning. In *Data Compression Conference, 2004. Proceedings. DCC 2004*, pages 302–311, march 2004.

[19] V. Stankovic and A.D.Liveris and Z. Xiong and C.N. Georghiades. On code design for the Slepian-Wolf problem and lossless multiterminal networks. *IEEE Trans. on Inf. Th.*, 52(4):1495–1507, april 2006.

[20] E.H. Yang and D.K. He. Interactive encoding and decoding for one way learning: Near lossless recovery with side information at the decoder. *IEEE Trans. on Inf. Th.*, 56(4):1808–1824, 2010.