Maximum Likelihood BSC Parameter Estimation for the Slepian-Wolf Problem

Velotiaray Toto-Zarasoa, Aline Roumy, Christine Guillemot

▶ **To cite this version:**

HAL Id: inria-00628996

https://inria.hal.science/inria-00628996v2

Submitted on 18 Oct 2011

Maximum Likelihood BSC parameter estimation for the Slepian-Wolf problem

Velotiaray Toto-Zarasoa, *Student Member, IEEE,* Aline Roumy, *Member, IEEE*
and Christine Guillemot, *Senior Member, IEEE,*

Abstract—In the context of Distributed Source Coding, we propose a low complexity algorithm for the estimation of the cross-over probability p of the *Binary Symmetric Channel* (BSC) modeling the correlation between two binary sources. The coding is done with linear block codes. We propose a novel method to estimate p *prior* to decoding and show that it is the *Maximum Likelihood* estimator of p with respect to the syndromes of the correlated sources. The method can be utilized for the parameter estimation for channel coding of binary sources over the BSC.

Index terms — Source coding, Channel coding, Parameter estimation, Binary sequences.

I. INTRODUCTION

The Distributed Source Coding (DSC) problem was introduced by Slepian and Wolf (SW) [1]; the aim is to achieve lossless compression of correlated sources X and Y. They state [1] that no additional rate is needed when encoding the two sources *separately*, with respect to the *joint encoding* solution, provided that the decoding is performed jointly and that the compression rates are greater than $H(X|Y)$ and $H(Y|X)$ respectively. Here, $H(\cdot)$ stands for the entropy. We now consider the case of binary variables where the correlation between the sources is modeled as a virtual *Binary Symmetric Channel* (BSC) of *cross-over probability p*. It is shown in [2] that linear block codes can achieve the SW bound, provided that they achieve the capacity of the underlying BSC. Therefore, practical DSC solutions are based on channel codes, like Convolutional codes [3], Turbo codes [4] or Low-Density-Parity-Check (LDPC) codes [5].

In the literature, to perform the sum-product decoding in [3], or [5], the BSC parameter is assumed to be available at the decoder. However, in practice, it is necessary to estimate this parameter *on-line*. In channel coding, Simons *et. al* propose an estimator of the BSC parameter in [6], [7]. They observe the output of the BSC, and deduce p based on the assumption that certain finite sequences appear rarely in the input; this method is only efficient when the source distribution is known. In the DSC setup, [4], [8] propose to estimate p with an expectation-maximization (EM) algorithm. However, no initialization of the estimate is proposed, and the estimation accuracy depends on the quality of the initialization, especially for low correlation between the sources (large p). It is proposed in [9] to use the Log-Likelihood Ratio, propagated during the *sum-product* decoding of LDPC codes, to observe a function of p; this method is only efficient for high correlation between the sources. Particle filtering and LDPC codes are used in [10]

to iteratively update the estimate $\hat{p}$; the method can be used to pursue slow changes of p, but it needs a large number of iterations to converge. The performance of all these methods [4], [8], [9], [10] closely depends on the initialization of $\hat{p}$, and is efficient when the sources are highly correlated.

In the sequel, we propose a low complexity estimator that performs well for all value of p. It is performed *prior to decoding* and can serve as an initialization for the EM algorithm proposed in [4], [8]. The trick is that our solution is the *Maximum Likelihood* (ML) estimator for p with respect to a subset of the data available at the decoder. It therefore provides an efficient estimate of p that can be refined with the EM algorithm that exploits all the data available at the decoder (namely the side-information (SI) Y). More precisely, in the channel coding approach to solve the SW problem, the decoder is fed with the source Y and the syndrome of the source X. Our solution only uses the syndromes of both sources and estimates the distribution of the syndrome of the difference between the sources. From this, we deduce the distribution of the difference between the sources, that is p.

II. ASYMMETRIC DSC USING LINEAR BLOCK CODES

Let X, Y be two correlated binary sources, and $\mathbf{x}$, $\mathbf{y}$ their realizations of length N. $\mathbf{x}$ and $\mathbf{y}$ differ by some noise $\mathbf{z}$, which is the realization of a Bernoulli random variable Z of parameter $p \in [0, 0.5]$ (noted $Z \sim \mathcal{B}(p)$), s.t. $Y = X \oplus Z$, and $\mathbb{P}(X \neq Y) = \mathbb{P}(Z = 1) = p$. In *asymmetric* DSC [1], the source X is compressed at a rate greater than $H(X|Y)$, while the source Y is compressed at a rate greater than $H(Y)$ allowing the decoder to retrieve it error-free. A way to compress the source X consists in sending the syndrome $\mathbf{s_x} = \mathbf{Hx}$ of length $(N - K)$, where $\mathbf{H}$ is the $(N - K) \times N$ parity check matrix of an (N, K) linear block code. Wyner [2] shows that if the linear block code achieves the capacity of the underlying BSC, then it also achieves the SW bound $H(X|Y)$. The decoder finds the best estimate $\hat{\mathbf{x}}$ as the closest sequence to $\mathbf{y}$ with syndrome $\mathbf{s_x}$: $\hat{\mathbf{x}} = \arg\min_{\mathbf{x}\ s.t.\mathbf{Hx}=\mathbf{s_x}} d(\mathbf{x}, \mathbf{y}) = \arg\min_{\mathbf{x}\ s.t.\mathbf{Hx}=\mathbf{s_x}} \mathbb{P}(\mathbf{x}|\mathbf{y})$. For LDPC codes, this search is efficiently approximated by the *sum-product* algorithm [11], that requires knowledge of the parameter p.

III. ML ESTIMATOR OF p WITH RESPECT TO $\mathbf{s_x}$ AND $\mathbf{s_y}$

The *sum-product* decoder is fed by the syndrome $\mathbf{s_x}$ and the SI realization $\mathbf{y}$. In the following, we derive an estimate for p based on the knowledge of the syndromes $\mathbf{s_x}$ and $\mathbf{s_y}$. Without loss of generality, we consider *regular* block codes. Regular block codes have the same number of *ones*, noted d_c, in every

row of $\mathbf{H}$. In the following, d_c is called *"check degree"*. We first characterize the distribution of the syndrome of $\mathbf{z}$.

Lemma 1. *Let $\mathbf{H}$ be the matrix of a regular linear block code in which all the rows are linearly independent. Let $\mathbf{z}$ be the realization of $Z \sim \mathcal{B}(p), p \in [0, 0.5]$. Let $\mathbf{s_z} = \mathbf{Hz}$.*

The syndrome $\mathbf{s_z}$ can be seen as the realization of an i.i.d. Bernoulli random process S_Z of parameter q, s.t.

$$q(p) = \sum_{\substack{(i=1) \\ (i \ odd)}}^{d_c} p^i (1-p)^{d_c-i} \binom{d_c}{i} \tag{1}$$

Proof: First, since the rows of $\mathbf{H}$ are linearly independent, the syndrome symbols are independent. Now, let s_{zm} be the mth element of $\mathbf{s_z}$ and let q be the probability of s_{zm} being a *one*. Since Z is an i.i.d. process of law $\mathcal{B}(p)$, and since s_{zm} is the sum of d_c elements of Z, $q(p)$ is given by Equation (1) and is the same for all the symbols $(s_{zm})_{m=1}^{N-K}$. Therefore, the syndrome symbols are independent and identically distributed realizations of an (*iid*) binary source, i.e. a Bernoulli source, noted S_Z, of parameter q. ∎

Now, that we have characterized the process S_Z, we can derive an estimate of its Bernoulli parameter q. This is stated in the following corollary.

Corollary 1. *The ML estimator of q with respect to $\mathbf{s_z}$ is the estimate $\hat{q}$ of the mean of the i.i.d. Bernoulli process S_Z; i.e:*

$$\hat{q} = \frac{1}{N-K} \sum_{m=1}^{N-K} s_{zm} \tag{2}$$

Proof: Since $\mathbf{s_z}$ is the realization of an i.i.d. Bernoulli process $S \sim \mathcal{B}(q)$ (from Lemma 1), $\sum_{m=1}^{N-K} s_{zm}$ is a sufficient statistic for the estimation of q with respect to $\mathbf{s_z}$, and the mean of S_Z, in (2), is the ML estimator of q. ∎

Let us return to the DSC problem, where $\mathbf{s_x}$ and $\mathbf{y}$ are available at the decoder; the following Theorem gives the ML estimator of p with respect to $\mathbf{s_x}$ and $\mathbf{s_y} = \mathbf{Hy}$, i.e. parts of the available data.

Theorem 1. *Let $\mathbf{H}$ be the matrix of a regular linear block code in which all the rows are linearly independent and contain the same number of ones. Let $\mathbf{x}$ be the realization of the binary source X, and let $\mathbf{s_x} = \mathbf{Hx}$ be its syndrome. Let Y be another binary source which is correlated to X in the following manner: $\exists Z \sim \mathcal{B}(p)$ s.t. $p \in [0, 0.5]$ and $Y = X \oplus Z$. Let $\mathbf{y}$ be a realization of Y, and let $\mathbf{s_y}$ be its syndrome. Let $f : p \to q(p)$, where $q(p)$ is given in (1).*

The Maximum Likelihood estimator for p with respect to $(\mathbf{s_x}, \mathbf{s_y})$ is:

$$\hat{p} = f^{-1}(\hat{q}) \tag{3}$$

where $\hat{q}$ is given in (2), and f^{-1} is the inverse of f.

Proof: Let $\mathbf{s_x} = (s_{xm})_{m=1}^{N-K}$ and $\mathbf{s_y} = (s_{ym})_{m=1}^{N-K}$. From Lemma 1, the joint probability of $\mathbf{s_x}$ and $\mathbf{s_y}$ can be factored as $\mathbb{P}(\mathbf{s_x}, \mathbf{s_y}) = \mathbb{P}(\mathbf{s_x}) \cdot \mathbb{P}(\mathbf{s_x} \oplus \mathbf{s_y})$

$$= \mathbb{P}(\mathbf{s_x}) \cdot q^{\left(\sum_{m=1}^{N-K} s_{xm} \oplus s_{ym}\right)} (1-q)^{\left((N-K) - \sum_{m=1}^{N-K} s_{xm} \oplus s_{ym}\right)}$$

From [12, Theorem 5.1], $\sum_{m=1}^{N-K} s_{xm} \oplus s_{ym}$ is a sufficient statistic for the estimation of q. The ML estimator of q, with respect to $(\mathbf{s_x}, \mathbf{s_y})$, is $\hat{q} = \frac{1}{N-K} \sum_{m=1}^{N-K} s_{xm} \oplus s_{ym}$. We denote $f(p) = q(p)$ for clarity of notation.

f is a strictly increasing one-to-one function of p in $[0, 0.5]$, and we denote f^{-1} its inverse, s.t. $p = f^{-1}(q)$. It follows from [12, Theorem 7.2], that the ML estimator of p, with respect to $(\mathbf{s_x}, \mathbf{s_y})$, is $\hat{p} = f^{-1}(\hat{q})$. ∎

Note that this ML estimator $\hat{p}$ does not depend on the estimates of the source X as in the EM algorithm (see [4], [8] and equation (4)). It can therefore be performed *prior to decoding*. Moreover, it does not depend on the distribution of X and Y as in [6], [7] and is therefore more general. Finally, our estimate $\hat{p}$ is asymptotically unbiased and asymptotically efficient (since it is the ML estimator). However it is biased since $\hat{q}$ is unbiased and f is non-linear.

IV. Improved estimation with the EM algorithm

The ML estimator presented in Section III only uses the information from $\mathbf{s_x}$ and $\mathbf{s_y}$, which is not optimal since the information from $\mathbf{y}$ is not fully exploited. In this Section, an EM algorithm [13] is used to improve the estimator $\hat{p}$. The EM is an optimization procedure that updates $\hat{p}$ through the decoding iterations, convergence is acquired since it improves the estimator's likelihood at each iteration. Let l be the label of the current decoding iteration, and p^l be the current estimate. Then the next estimate is the solution to the maximization problem $p^{(l+1)} = \arg\max_p \left(\mathbb{E}_{\mathbf{X}|\mathbf{S_x}, \mathbf{Y}, p^l} \left[\log \left(\mathbb{P}_p(\mathbf{s_x}, \mathbf{y}, \mathbf{x}) \right) \right] \right)$ and is

$$p^{(l+1)} = \frac{1}{N} \sum_{n=1}^{N} |y_n - \mathbb{P}_n| \tag{4}$$

where $\mathbb{P}_n$ is the *a posteriori* probability $\mathbb{P}_{p^l}(X_n = 1 | \mathbf{s_x}, \mathbf{y})$ from the sum-product algorithm. This update rule is the same as presented in [4, Equation 3]. However, our EM algorithm differs from [4] since it is initialized with our efficient estimate $p^0 = f^{-1}(\hat{q})$, see (3).

V. Generalization of the estimator

The method can be used for parameter estimation in *channel coding* over the BSC, since the channel decoder is a particular case of the DSC decoder, with $\mathbf{s_x} = \mathbf{0}$. Moreover, for unsupervised learning of motion vectors in Distributed Video Coding [14], our estimator can be used to initialize the motion field M relating the unknown frame X to its SI Y, knowing its syndrome S. More precisely, for each sub-block B in X, the syndromes of the candidates in Y can be computed and used to estimate their respective correlation with B. Then, the most correlated one can be used as SI for B. The motion estimation can be performed in pixel or transform domain (see [14]).

VI. Simulation results

A. Precision of the estimator

We implement two DSC systems using LDPC codes of rates $R = 0.5$ and $R = 0.7$, which have the *variable degree distribution* $\Lambda(x) = 0.483949x + 0.294428x^2 + 0.085134x^5 + 0.074055x^6 + 0.062433x^{19}$ and the *check degree distribution* $\Phi(x) = 0.741935x^7 + 0.258065x^8$. Each row of $\mathbf{H}$ is linearly independent of one another. We consider two codes of respective lengths $N = 1000$ and $N = 10000$. The sources are uniform Bernoulli, and we test the estimation for p greater than 0.05. We show the means of the estimated $\hat{p}$ in Fig. 1 for $N = 1000$ and $N = 10000$. On the same Figure, we also show the estimated parameters from the EM algorithm.
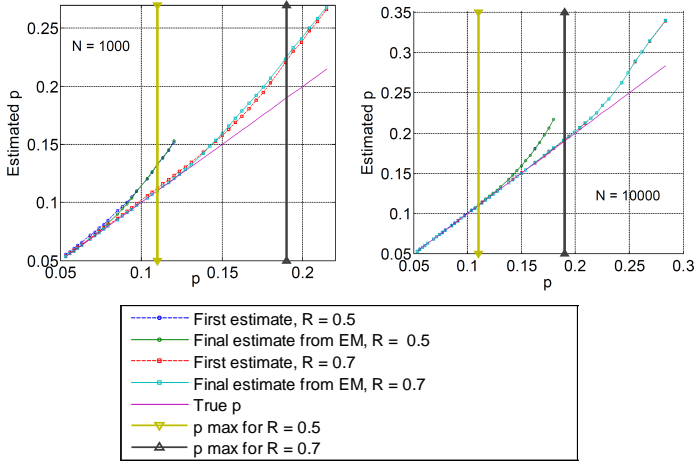
Fig. 1. Means of the estimated $\hat{p}$ in function of N and R

It can be seen in Fig. 1 that the precision of the estimator improves with the code length and with the code rate. The improvement brought by the EM is only minimal, since the first estimate is already very good. Note that we could assess the performance of the estimator for the whole range of $p \in [0, 0.5]$, but we limit the experiment to $p < 0.11$ for code rate 0.5 (resp. to $p < 0.19$ for code rate 0.7), because the estimated $\hat{X}$ must be sufficiently reliable. Actually, for the SW setup, the maximum value of p that yields reliable $\hat{X}$ verifies $H(p) = 0.5$ (resp. $H(p) = 0.7$), i.e. $p = 0.11$ (resp. $p = 0.19$) for the code of rate 0.5 (resp. 0.7). For efficient estimation of higher values of p, a code allowing higher compression rates must be used.

B. Convergence speed to the Cramer-Rao lower bound

For our estimation problem, a bound of the Cramer-Rao Lower Bound (CRLB) can be derived by considering the Minimum Variance Unbiased (MVU) estimator of p knowing the realization $\mathbf{z}$ of length N. This MVU estimator is $\hat{p} = \frac{1}{N} \sum_{n=1}^{N} z_n$ and its Mean Square Error (MSE), which corresponds to the bound of the CRLB, is $\frac{p(1-p)}{N}$. For the code of length $N = 10000$ and rate $R = 0.5$, we have assessed the number of iterations needed by the EM to reach the CRLB in function of p using our estimation method. The results are presented in Fig. 2.
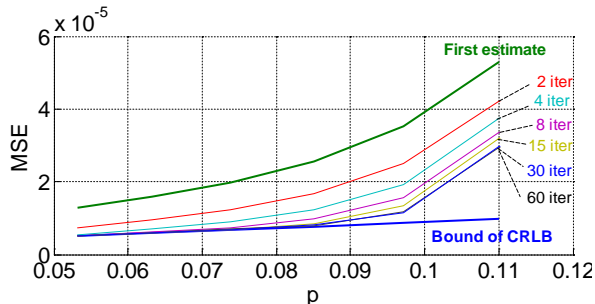

Fig. 2. Comparison of the MSE of our estimator with the CRLB.

First, note that the MSE of the first estimate is already very small ($< 10^{-4}$) before applying the EM algorithm. 4 iterations of the EM are sufficient to reach the CRLB for $p < 0.055$, and 30 iterations are sufficient to reach the CRLB for $p < 0.085$.

C. Comparison with genie-aided SW decoder

We have implemented a SW codec where $\hat{p}$ is initialized with the ML estimator (3). In Fig. 3, we observe the BERs of

X for decoding iteration numbers $2, 4, 8, 15, 30$ and 60 (*solid lines*); they are compared to the BERs yielded by the SW codec in which the initialization of the EM is done with the fixed value $p^0 = 0.2$ (*dashed lines*), and with the *genie-aided* SW codec (which knows the true p, in *dotted lines*). The same update rule (4) is used for the two systems that need to estimate the parameter p.
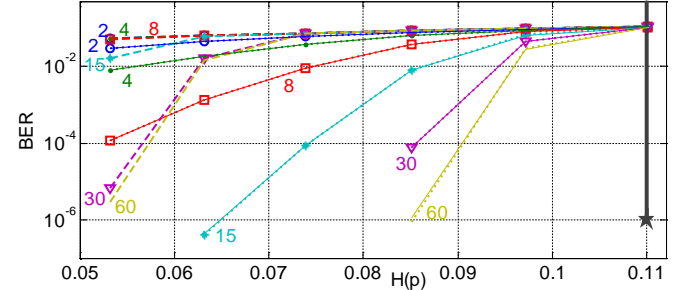

Fig. 3. Comparison of the BER of X for different iterations of the EM.

We see from Fig. 3 that the decoder initialized with our ML estimator is far more efficient than the one initialized with the fixed value $p^0 = 0.2$, beside it performs as well as the genie-aided decoder (the solid lines and the dotted lines are almost merged together, for all the iterations).

VII. CONCLUSION

We have proposed a novel and simple ML estimator of the cross-over parameter of the BSC modeling the correlation between two arbitrary binary sources. The estimation is performed prior to decoding, and only depends on the degree distribution of the matrix representing the DSC code. We also presented an EM algorithm that improves the estimation variance. The method can be extended to channel coding of binary sources over the BSC.

REFERENCES

[1] D. Slepian and J. K. Wolf, "Noiseless coding of correlated information sources," *IEEE Trans. Inf. Th.*, vol. 19, no. 4, pp. 471–480, July 1973.
[2] A. Wyner, "Recent results in the Shannon theory," *IEEE Trans. on Inf. Th.*, vol. 20, no. 1, pp. 2–10, January 1974.
[3] J. Li and H. Alqamzi, "An optimal distributed and adaptive source coding strategy using rate-compatible punctured convolutional codes," *ICASSP*, vol. 3, pp. 685–688, March 2005.
[4] J. Garcia-Frias and Y. Zhao, "Compression of correlated binary sources using Turbo codes," *IEEE Comm. Let.*, vol. 5, October 2001.
[5] A. D. Liveris, Z. Xiong, , and C. N. Georghiades, "Compression of binary sources with side information at the decoder using LDPC codes," *IEEE Comm. Let.*, October 2002.
[6] G. Simons, "Estimating distortion in a binary symmetric channel consistently," *IEEE Trans. Inf. Th.*, September 1991.
[7] K. Podgorski, G. Simons, and Y. Ma, "On estimation for a binary symmetric channel," *IEEE Trans. Inf. Th.*, May 1998.
[8] A. Zia, J. P. Reilly, and S. Shirani, "Distributed parameter estimation with side information: A factor graph approach," in *ISIT*, June 2007.
[9] Y. Fang and J. Jeong, "Correlation parameter estimation for LDPC-based Slepian-Wolf coding," *IEEE Comm. Let.*, 2009.
[10] S. Cheng, S. Wang, and L. Cui, "Adaptive Slepian-Wolf decoding using particle filtering based belief propagation," in *Allerton conf. Comm. Cont. Comp.*, 2009, pp. 607–612.
[11] F. R. Kschischang, B. J. Frey, and H. Loeliger, "Factor graphs and the sum-product algorithm," *IEEE Trans. Inf. Th.*, February 2001.
[12] S. M. Kay, *Fundamentals of statistical signal processing: estimation theory*. Prentice-Hall, Inc., 1993.
[13] L. R. Rabiner, "A tutorial on hidden Markov models and selected applications in speech recognition," *Proc. IEEE*, February 1989.
[14] D. Varodayan, D. Chen, M. Flierl, and B. Girod, "Wyner-Ziv coding of video with unsupervised motion vector learning," *EURASIP Sig. Proc., Sp. Issue on DVC*, vol. 23, no. 5, pp. 369–378, June 2008.