Maximum Likelihood BSC Parameter Estimation for the Slepian-Wolf Problem

Velotiaray Toto-Zarasoa, Aline Roumy, Christine Guillemot

▶ **To cite this version:**

Velotiaray Toto-Zarasoa, Aline Roumy, Christine Guillemot. Maximum Likelihood BSC Parameter Estimation for the Slepian-Wolf Problem. IEEE Communications Letters, 2011, 15 (2), 10.1109/LCOMM.2011.122810.102182 . inria-00628996v1

HAL Id: inria-00628996

https://inria.hal.science/inria-00628996v1

Submitted on 4 Oct 2011 (v1), last revised 18 Oct 2011 (v2)

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

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 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 the decoding, with increasing reliability for longer code sizes. We exploit the probability of observing *ones* in the sum of the syndromes of the correlated sources. We show that the method is a *Maximum Likelihood* estimator of p with respect to the available data. The method can be extended to the parameter estimation for channel coding of binary sources over the BSC.**

Index terms — Source coding, Channel coding, Binary Symmetric Channel, On-line parameter estimation.

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. [2] shows that linear block codes can achieve the SW bound, provided that they achieve the capacity of the underlying BSC. Yet, 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, [3], or [5], the BSC parameter is assumed to be available at the decoder. Nevertheless, in practice, it is necessary to estimate this parameter *on-line*. It is only lately that some authors have proposed solutions to perform on-line parameter estimation for DSC. [4] proposes to estimate p with an expectation-maximization algorithm. However, no initialization of the estimate is proposed. The authors in [6] use the Log-Likelihood Ratio, propagated during the Message-Passing of LDPC decoding, to observe a function of p; this method is only efficient for high correlation between the sources. The authors in [7] use particle filtering and LDPC codes 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], [6], [7] closely depends on the initialization of $\hat{p}$. For the channel coding problem, Simons *et. al* [8], [9] 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 sequel, we propose to estimate p by exploiting the information from the *syndrome* bits of X, from the *side-information* Y, and from the *matrix* representation of the code. The estimator uses the *probability* of occurrence of *ones* in the sum of the syndromes of X and Y, noted $\mathbf{s_x}$ and $\mathbf{s_y}$. We show that this probability is a bijective function of p, and we derive the *Maximum Likelihood* (ML) estimator for p. The estimation is performed prior to decoding, which makes it independent of the decoding, and it does not degrade the decoding. We propose a second estimator of p with respect to $\mathbf{s_x}$ and Y, which requires an Expectation-Maximization (EM) algorithm. This EM algorithm improves the ML estimator. The proposed methods can be extended to the BSC parameter estimation for the channel coding problem.

II. Asymmetric DSC using linear block codes

Let X, Y be two correlated binary sources, and $\mathbf{x}$, $\mathbf{y}$ be two vectors of their respective 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 $\mathbb{P}(X \neq Y) = \mathbb{P}(Z = 1) = p$. Let $\mathbf{H}$ be the $(N - K) \times N$ parity check matrix of an (N, K) linear block code, where H_m is the m-th row of $\mathbf{H}$. In *asymmetric* DSC [1], $\mathbf{x}$ is compressed at a rate greater than $H(X|Y)$ by sending the syndrome $\mathbf{s_x} = \mathbf{Hx}$ of length $(N - K)$. $\mathbf{y}$ is compressed at a rate greater than $H(Y)$, allowing the decoder to retrieve it error-free. The decoder finds the best estimate $\hat{\mathbf{x}}$ as the closest sequence to $\mathbf{y}$ with syndrome $\mathbf{s_x}$: $\hat{\mathbf{x}} = \arg\min_{x s.t. \mathbf{Hx}=\mathbf{s_x}} d(\mathbf{x}, \mathbf{y})$

from the knowledge of $\mathbf{H}$, $\mathbf{s_x}$, and $\mathbf{y}$. For example, this search is efficiently done using the *sum-product* algorithm [11] for LDPC codes.

III. ML estimator of p with respect to $\mathbf{s_x}$ and $\mathbf{s_y}$

A. ML estimator for regular block codes

Regular block codes have the same number of *ones*, noted d_s, in every row of $\mathbf{H}$. d_s is also called *"syndrome degree"*. We first enunciate a lemma that characterizes the Bernoulli nature of the syndrome of $\mathbf{z}$.

Lemma 1. *Let $\mathbf{H}$ be the matrix of a linear block code in which all the rows are linearly independent and contain the same number of ones (d_s). 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 a Bernoulli random variable S_Z of parameter q, s.t.

This research was partially supported by the French European Commission in the framework of the FP7 Network of Excellence in Wireless COMmunications NEWCOM++ (contract n.216715).

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

Proof: First, since the rows of $\mathbf{H}$ are linearly independent, the syndrome symbols are independent of one another.

Let $\mathbf{s_z} = (s_{zm})_{m=1}^{(N-K)}$. $\forall m \in [1, (N-K)]$, let q be the probability of s_{zm} being a *one*. s_{zm} equals *one* if there is an odd number of *ones* in the symbols of $\mathbf{z}$ the sum of which yields s_{zm}. Since $Z \sim \mathcal{B}(p)$, $q(p)$ is given by Equation (1) and is the same for all the symbols $(s_{zm})_{m=1}^{(N-K)}$. Then the syndrome symbols are identically distributed.

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, we enunciate a corollary to this lemma that will help estimate the Bernoulli parameter of S_Z.

Corollary 1. *The estimate $\hat{q}$ of the mean of the Bernoulli variable S_Z is the ML estimator of q with respect to $\mathbf{s_z}$.*

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

Proof: Since $\mathbf{s_z}$ is the realization of a Bernoulli variable $S \sim \mathcal{B}(q)$ (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. ∎

Now, we turn to the DSC problem, where $\mathbf{s_z}$ is not observable, but only $\mathbf{s_x}$ and $\mathbf{y}$; the following Theorem gives the ML estimator of p with respect to the available data.

Theorem 1. *Let $\mathbf{H}$ be the matrix of a 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 $\mathbf{s_y} = (s_{ym})_{m=1}^{(N-K)}$. The joint probability of $\mathbf{s_x}$ and $\mathbf{s_y}$ can be factored as:

$$\begin{aligned} \mathbb{P}(\mathbf{s_x}, \mathbf{s_y}) &= \mathbb{P}(\mathbf{s_x}) \cdot \mathbb{P}(\mathbf{s_x} \oplus \mathbf{s_y}) \\ &= \mathbb{P}(\mathbf{s_x}) \cdot q^{\left(\sum_{m=1}^{N-K} s_{xm} \oplus s_{ym}\right)} (1-q)^{\left((N-K) - \sum_{m=1}^{N-K} s_{xm} \oplus s_{ym}\right)} \end{aligned} \tag{4}$$

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 thus $\hat{q} = \frac{1}{N-K} \sum_{m=1}^{N-K} s_{xm} \oplus s_{ym}$. 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 $\hat{p} = f^{-1}(\hat{q})$. ∎

This ML estimator $\hat{p}$ does not depend on the distribution of X and Y since $\mathbf{s_x} \oplus \mathbf{s_y}$ only depends on the BSC modeling their correlation. There is no analytical expression of the inverse function f^{-1}; the inversion can be implemented numerically (with a correspondence table for example). Note that our estimator is biased since f is not a linear function.

B. ML estimator for irregular block codes

In the present Section, we derive the ML estimator of p when the syndrome symbols have different degrees. For example, LDPC codes are characterized by their *variable degree distribution* $\Lambda(x)$ which is the distribution of the number of *ones* in the columns of $\mathbf{H}$, and by their *syndrome degree distribution* $\Phi(x)$ which is the distribution of the number of *ones* in the rows of $\mathbf{H}$. $\forall m \in [1, (N-K)]$, let d_{sm} be the degree of s_{zm}. Let d^{max} be the maximum syndrome degree. Let $f_m(p) = \sum_{\substack{i=1 \\ i \ odd}}^{d_{sm}} p^i (1-p)^{d_{sm} - i} \binom{d_{sm}}{i}$.

We want to find the best estimate $\hat{p}$ that yields $\mathbf{s_x}$ and $\mathbf{s_y}$, which is the solution of the maximization problem:

$$\hat{p} = \arg\max_p \mathbb{P}\left(\mathbf{s_x}, \mathbf{s_y}\right)(p)$$

The joint probability of the syndromes symbols can be expressed in function of p, as:

$$\mathbb{P}\left(\mathbf{s_x}, \mathbf{s_y}\right)(p) = \prod_{m=1}^{N-K} f_m(p)^{s_{xm} \oplus s_{ym}} \left(1 - f_m(p)\right)^{1 - s_{xm} \oplus s_{ym}} \tag{5}$$

Since $\log(\cdot)$ is a strictly increasing function, the value of $\hat{p}$ that maximizes $\mathbb{P}\left(\mathbf{s_x}, \mathbf{s_y}\right)(p)$, in (5), also maximizes $\log\left(\mathbb{P}\left(\mathbf{s_x}, \mathbf{s_y}\right)(p)\right)$. In the expression (6), the sum on the syndrome symbols has been re-organized in order to group the syndrome symbols having the same degrees.

$$\frac{d}{dp}\left(\log\left(\mathbb{P}\left(S_1^N\right)(p)\right)\right) = \sum_{j=1}^{d^{max}} \sigma_j \frac{f_j'(p)}{f_j(p)} - (1 - \sigma_j) \frac{f_j'(p)}{1 - f_j(p)} \tag{6}$$

where, $\forall j \in [1, d^{max}]$, $\sigma_j = \frac{1}{N_j} \sum_{\substack{m=1 \\ d_m = j}}^{N-K} s_{zm}$, and $f_j'(p) = \frac{d}{dp} f_j(p)$, and N_j is the number of symbols with degree j.

That re-organization is motivated by the fact that the syndrome symbols having the same degree are realizations of a common Bernoulli variable (from Lemma 1). The ML estimator $\hat{p}$ is found by zeroing the derivative (6).

IV. Improved estimation of p using an 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, the convergence is acquired since it improves the estimator likelihood at each iteration [14]. 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:

$$p^{(l+1)} = \arg\max_p \left(\mathbb{E}_{\mathbf{X}|\mathbf{s_x}, \mathbf{Y}, p^l}\left[\log\left(\mathbb{P}_p(\mathbf{s_x}, \mathbf{y}, \mathbf{x})\right)\right]\right) \tag{7}$$

Since the graph of the channel code contains cycles, the probability (7) cannot be computed exactly. The EM procedure consists into updating p^l thanks to the current *a posteriori probabilities* (APP) of $\mathbf{x}^l$ produced by the decoder. More precisely, $\forall n \in [1, N]$ let $\mathbb{P}_n$ represent the APP $\mathbb{P}_{p^l}(X_n = 1|\mathbf{s_x}, \mathbf{y})$. Then:

$$p^{(l+1)} = \frac{1}{N}\sum_{n=1}^{N} |y_n - \mathbb{P}_n| \tag{8}$$

This update rule is the same as presented in [4, Equation 3]. The EM algorithm is initialized with $p^0 = f^{-1}(\hat{q})$, see (3).

V. Simulation Results

We implement a DSC system using an LDPC code of rate $\frac{1}{2}$, which has the *variable degree distribution* $\Lambda(x) = 0.483949x + 0.294428x^2 + 0.085134x^5 + 0.074055x^6 + 0.062433x^{19}$ and the *syndrome 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 ranging from 0.05 to 0.11. We show the means of the estimated $\hat{p}$ in Fig. 1. For the code of length 1000, we also show the estimated parameters from the EM algorithm. The parameter is updated according to (8) each 20 iterations of the decoding to observe convergence.

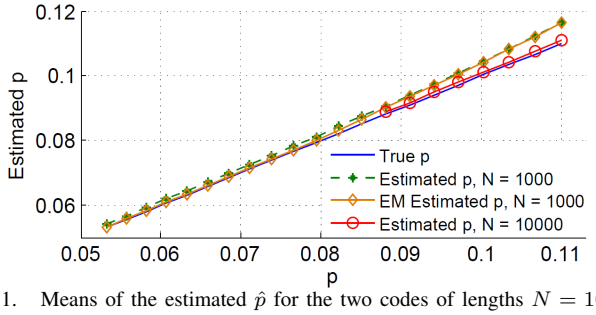

Fig. 1. Means of the estimated $\hat{p}$ for the two codes of lengths $N = 1000$ and $N = 10000$. Comparison with the true parameter p. The EM estimator improves the initial estimator.

The means of the estimated parameters are very close to the true values. The estimator performance improves with the code length; the gap to the true parameter goes under 10^{-3} for $N = 1000$, and it is under 10^{-4} for $N = 10000$. The EM always improves the performance of the estimator.

The performance of the decoder using the estimated parameter from the EM is compared to the performance of the genie-aided decoder using the true parameter. Both decoders iterate at most 100 times. The results presented in Fig. 2 indicate that no degradation at all is observable when using the estimated parameter or the true parameter.

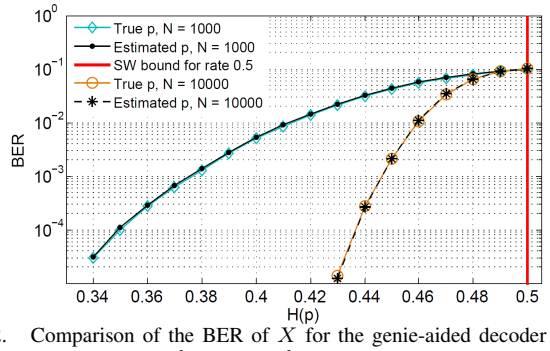

Fig. 2. Comparison of the BER of X for the genie-aided decoder and the proposed decoder, for N={1000,10000}. No rate loss is observable.

VI. On the particular case of channel coding

Let us consider the channel coding of binary sources. Let $\mathbf{x}$ be a valid *codeword* that is sent over the BSC represented by the signal $Z \sim \mathcal{B}(p)$, with noise realization $\mathbf{z}$. The output of the channel is $\mathbf{y} = \mathbf{x} \oplus \mathbf{z}$. The decoder uses the parity-check matrix $\mathbf{H}$, and $\mathbf{y}$, to estimate the original sequence $\mathbf{x}$, exploiting the fact that $\mathbf{Hx} = \mathbf{0}$. Note that $\mathbf{Hy} = \mathbf{H} \cdot (\mathbf{x} \oplus \mathbf{z}) = \mathbf{Hx} \oplus \mathbf{Hz} = \mathbf{Hz}$; this means that the received sequence $\mathbf{y}$ and the error pattern have the same syndrome.

Using the same argument as for the SW problem (section III), the estimator of the mean of the Bernoulli variable, $\hat{q}$, estimated from the syndrome of $\mathbf{y}$ can be found with the ML estimator, as in Equation (2); and this leads to the ML estimation of the parameter p, as in Equation (3). This estimation is a particular case of that proposed in Section III, when the code imposes $\mathbf{s_x} = 0$.

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, by using the ML estimator as an initialization. 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 Transactions on Information Theory*, vol. 19, no. 4, pp. 471–480, July 1973.

[2] A. Wyner, "Recent results in the Shannon theory," *IEEE Transactions on Information Theory*, vol. 20, 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 Communications Letters*, vol. 5, pp. 417–419, 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 Communications Letters*, vol. 6, no. 10, pp. 440–442, October 2002.

[6] Y. Fang and J. Jeong, "Correlation parameter estimation for ldpc-based slepian-wolf coding," *IEEE Communications Letters*, vol. 13, no. 1, pp. 37–39, 2009.

[7] S. Cheng, S. Wang, and L. Cui, "Adaptive slepian-wolf decoding using particle filtering based belief propagation," in *Proceedings of the 47th annual Allerton conference on Communication, control, and computing*, 2009, pp. 607–612.

[8] G. Simons, "Estimating distortion in a binary symmetric channel consistently," *IEEE transactions on information theory*, vol. 37, no. 5, pp. 1466–14 700, September 1991.

[9] K. Podgorski, G. Simons, and Y. Ma, "On estimation for a binary symmetric channel," *IEEE transactions on information theory*, vol. 44, no. 3, pp. 1261–1272, May 1998.

[10] L. R. Bahl, J. Cocke, F. Jelinek, and J. Raviv, "Optimal decoding of linear codes for minimizing symbol error rate," *IEEE Transactions on Information Theory*, pp. 284–287, March 1974.

[11] F. R. Kschischang, B. J. Frey, and H. Loeliger, "Factor graphs and the sum-product algorithm," *IEEE Transactions on Information Theory*, vol. 47, no. 2, pp. 498–519, 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," *Proceedings of the IEEE*, vol. 77, no. 2, pp. 257–286, February 1989.

[14] C. F. J. Wu, "On the convergence properties of the EM algorithm," *Annals of Statistics*, vol. 11, no. 1, p. 95103, 1983.