A Novel Method for Blind Identification of a (*n*, *n*-1, *m*) Convolutional Code

Shu Nan Han and Min Zhang

***Abstract*—The existing methods for identification of a $(n,n-1,m)$ convolutional code are not applicable in the cases of high bit error rates or need a large amount of computation. To overcome the limitations, a novel blind identification method is proposed. First, based on the parity check equation set, the parity check vector of a convolutional code is estimated by using the proposed recursive algorithm. Second, due to the orthogonality between the parity check matrix and the generator matrix, a set of polynomial generator bases are obtained. Finally, the canonical generator matrix is reconstructed by using the polynomial generator bases. Experimental results show the method is effective. The method has high robustness to bit errors. It does not need to search for the parity check vector exhaustively, and therefore its computational complexity is much lower than that of the existing method.**

***Index Terms*—Convolutional code, blind identification, parity check matrix, robustness.**

I. Introduction

In order to enhance the reliability of communication, convolutional codes are widely used in communication systems [1]. A $(n,n-1,m)$ convolutional code has the high code rate and is frequently used in practice [2]. In the fields of information interception and cognitive radio, the blind identification of a $(n,n-1,m)$ convolutional code is a key technology for non-corporative sides.

Filiol firstly proposes an algebraic identification method. The method is only applicable in the situation without noise [3]. Marazin proposes an identification method based on matrix analysis [4]. In this method, the parity check vector is estimated by converting the received bit matrix into a lower triangular matrix, and the parity check matrix is identified according to the estimation result. However, only estimating the parity check matrix is not sufficient for the identification of a $(n,n-1,m)$ convolutional code. The reconstruction of the generator matrix of a $(n,n-1,m)$ convolutional code is not discussed in [4]. Additionally, this method has low robustness to bit errors and is not useful in the non-corporative situation. In [5], matrix analysis and Walsh-Hadamard transform (WHT) are combined. The codeword length and constraint length are estimated by using matrix analysis first. Then the parity check matrix and generator matrix are reconstructed by WHT. Since the robustness of this method is limited by that of matrix analysis, it also cannot be applied under high error bit rates. Moreover, the generator matrix obtained by the method is not optimal. To satisfy the identification request in the cases of high bit error rates, a method based on exhaustively search is proposed in [6]. Since the parity check vector and the codeword length need to be searched for exhaustively. The computational complexity of the method increases exponentially along with the increase of the codeword and constraint lengths.

As above analysis, the existing identification methods are not suitable in the situations of high error bit rates or need a large amount of computation. In order to solve these limitations, a novel identification method is proposed in the paper. First, the parity check vector of a $(n,n-1,m)$ convolutional code is estimated by the proposed recursive algorithm and consequently the parity check matrix is obtained. Then due to the orthogonality between the parity check matrix and the generator matrix, a linear equation set is established. The polynomial generator bases are estimated by solving the equation set recursively. Finally, according to the property of a canonical generator matrix, some polynomial generator bases are selected to reconstruct the generator matrix.

The rest of this paper is organized as follows. In Section II, the mathematical model for identification is elaborated. In Section III, a recursive algorithm for estimating the parity check vector is proposed. In Section IV, the polynomial generator bases are estimated and the canonical generator matrix is reconstructed. The computational complexity is analyzed and simulation results are shown in Section V. Conclusions are given in Section VI.

II. Mathematical Model

A $(n,n-1,m)$ convolutional encoder can be described by a $(n-1)\times n$ polynomial generator matrix $\boldsymbol{G}(D)$. Correspondingly, there exists a $1\times n$ parity check matrix $\boldsymbol{H}(D)$ which is orthogonal to $\boldsymbol{G}(D)$ [7], i.e.,

$$\boldsymbol{G}(D)\cdot\boldsymbol{H}^{\mathrm{T}}(D)=0. \tag{1}$$

Suppose the information sequence is $m(D)$, the encoded sequence $c(D)$ can be expressed as

$$c(D)=m(D)\cdot\boldsymbol{G}(D). \tag{2}$$

Combining (1) and (2), the following equation is obtained.

$$c(D)\cdot\boldsymbol{H}^{\mathrm{T}}(D)=0 \tag{3}$$

Manuscript received March 7, 2018; revised May 23, 2018. This work was supported by the National Natural Science Foundation of China (61602491).

The authors are with the National University of Defense Technology, Hefei, China (e-mail: hsnnong@163.com).

doi: 10.18178/ijfcc.2018.7.2.519

According to (3), a parity check equation set can be established. The solution of the equation set is the parity check vector $\boldsymbol{h}$ which is consisted by coefficients of polynomials in $\boldsymbol{H}(D)$. Therefore, $\boldsymbol{H}(D)$ can be estimated by solving the parity check equation set. Furthermore, $\boldsymbol{G}(D)$ can be reconstructed using the orthogonality between $\boldsymbol{G}(D)$ and $\boldsymbol{H}(D)$.

III. Recursive Algorithm for Estimation of the Parity Check Vector

A. Principle of Recursive Algorithm

The parity check equation set derived from (3) is shown as (4).

$$\begin{bmatrix} c_1 & c_2 & \cdots & c_n & c_{n+1} & \cdots & c_{n(m+1)} \\ c_{n+1} & c_{n+2} & \cdots & c_{2n} & c_{2n+1} & \cdots & c_{n(m+2)} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ c_{(N-1)n+1} & c_{(N-1)n+2} & \cdots & c_{Nn} & c_{Nn+1} & \cdots & c_{n(N+m)} \end{bmatrix} \bullet \begin{bmatrix} h_1 \\ h_1 \\ \vdots \\ h_n \\ h_{n+1} \\ \vdots \\ h_{n(m+1)} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \tag{4}$$

The solution of the equation set is the parity check vector. In this subsection, we propose a recursive algorithm for the estimation of the parity check vector.

Let $\hat{\boldsymbol{H}}$ denote the matrix consisting of the values of elements having been estimated in the parity check vector and $\boldsymbol{h}'$ denote the vector of undetermined elements. The length of the parity check vector is $n(m+1)$. The steps of the recursive algorithm is as follows.

(1) Initialize $\hat{\boldsymbol{H}} = \phi$ and $\boldsymbol{h}' = [h_1, h_2, \cdots, h_{n(m+1)}]^{\mathrm{T}}$, where ϕ denotes a blank matrix.

(2) For simplicity, only the general j-th recursion of the algorithm is elaborated in this step. Assume the matrix $\boldsymbol{A}$ is composed of the columns of the received bit matrix corresponding to the elements having been estimated in $\boldsymbol{h}$, and the matrix $\boldsymbol{B}$ is composed of the columns of the received bit matrix corresponding to the unknown elements in $\boldsymbol{h}$; then the parity check equation set shown in (4) can be expressed as

$$\boldsymbol{B}\boldsymbol{h}' \oplus \boldsymbol{A}\hat{\boldsymbol{H}} = \boldsymbol{0}. \tag{5}$$

Find the sparsest row of $\boldsymbol{B}$ in which the number of 1 elements is the smallest. If the i-th row is the sparsest row and the 1 elements are the coefficients of the unknowns $h_{k_1}, h_{k_2}, \cdots, h_{k_p}$ respectively, then according to the i-th equation, the modulo 2 summation value of the elements $h_{k_1}, h_{k_2}, \cdots, h_{k_p}$ is obtained, i.e.,

$$h_{k_1} \oplus h_{k_2} \oplus \cdots \oplus h_{k_p} = \boldsymbol{a}_i \cdot \hat{\boldsymbol{H}}. \tag{6}$$

where $\boldsymbol{a}_i$ is the i-th row of $\boldsymbol{A}$. From (6), we can get all the possible values of the vector $[\hat{h}_{k_1}, \hat{h}_{k_2}, \cdots, \hat{h}_{k_p}]^{\mathrm{T}}$. The matrix $\hat{\boldsymbol{H}}$ is expanded according to these vectors, and the elements $h_{k_1}, h_{k_2}, \cdots, h_{k_p}$ in $\boldsymbol{h}'$ are deleted. The j-th recursion is finished.

If there are errors in the received bit sequence, it is probable to obtain an incorrect summation value of $h_{k_1}, h_{k_2}, \cdots, h_{k_p}$ by only one parity check equation. Since the number of correct equations is larger than that of incorrect equations in practice, we use several parity check equations to determine the summation value in each recursion. Assume there are N_{eq} equations which have the same form as (6); the numbers of equations by which the summation value is estimated to be 1 or 0 are N_{eq}^1 and N_{eq}^0 respectively. Then, the summation value is determined according to the following rule.

When $N_{\mathrm{eq}} \geq th_{\mathrm{eq}}$,

$$\hat{h}_{k_1} \oplus \hat{h}_{k_2} \oplus \cdots \oplus \hat{h}_{k_p} = \begin{cases} 1 & , N_{\mathrm{eq}}^1 > N_{\mathrm{eq}}^0 \\ 0 & , N_{\mathrm{eq}}^1 < N_{\mathrm{eq}}^0 \\ \{0,1\} & , N_{\mathrm{eq}}^1 = N_{\mathrm{eq}}^0 \end{cases}. \tag{7}$$

When $N_{\mathrm{eq}} < th_{\mathrm{eq}}$,

$$\hat{h}_{k_1} \oplus \hat{h}_{k_2} \oplus \cdots \oplus \hat{h}_{k_p} = \{0,1\}. \tag{8}$$

where th_{eq} is the smallest number of parity check equations required for determining the summation value and (8) denotes that the summation is assigned two possible values, 0 and 1.

The maximum probability $p_{\max}$ of incorrect determination for the summation value under the threshold th_{eq} is

$$p_{\max} = \sum_{i=\lceil th_{\mathrm{eq}}/2 \rceil + 1}^{2\lceil th_{\mathrm{eq}}/2 \rceil} C_{2\lceil th_{\mathrm{eq}}/2 \rceil}^{i} p_{\mathrm{e}}^{i} \left(1 - p_{\mathrm{e}}\right)^{2\lceil th_{\mathrm{eq}}/2 \rceil - i}. \tag{9}$$

$$p_{\mathrm{e}} = \sum_{i=1}^{\lceil w/2 \rceil} C_w^{2i-1} \eta^{2i-1} \left(1 - \eta\right)^{w-2i+1} \tag{10}$$

Equation (10) denotes the error probability of a parity check equation, where η is the bit error rate and w is the weight of the parity check vector. According to (9), we can calculate the threshold th_{eq} from the assumed $p_{\max}$. The threshold th_{eq} increases with the decrease of $p_{\max}$.

The recursion in step (2) is carried out again until all elements of $\boldsymbol{h}$ are estimated.

With the increase of the number of elements having been estimated, there are more and more rows which contain only one 1 element in the matrix $\boldsymbol{B}$. Therefore, the unknown elements of $\boldsymbol{h}$ can be estimated one by one in this case.

B. Correctness Verification of the Parity Check Vector Estimations

Since there is only one parity check vector for the $(n, n-1, m)$ convolutional code, we should choose the correct one from all the estimations. The received bit matrix is multiplied by an estimation $\hat{\boldsymbol{h}}$ of the parity check vector. 0 elements in the output vector represent that the corresponding equations hold and 1 elements represent that the equations do not hold. Define variables $\xi_i,(i=1,2,\cdots,N)$, where N is the number of parity check equations. If the i th equation hold, $\xi_i=1$; otherwise, $\xi_i=-1$. The testing statistic is defined as $\sum_{i=1}^{N}\xi_i$. Let the hypothesis H_1 denote the estimation of the parity check vector is correct and the hypothesis H_0 denote the estimation is incorrect. The probability distributions of $\sum_{i=1}^{N}\xi_i$ in the cases of hypotheses H_1 and H_0 are shown as follows [8].

$$\begin{aligned} H_1&:\sum_{i=1}^{N}\xi_i \sim N\left(N\left(1-2p_{\mathrm{e}}\right),4Np_{\mathrm{e}}(1-p_{\mathrm{e}})\right) \\ H_0&:\sum_{i=1}^{N}\xi_i \sim N\left(0,N\right) \end{aligned} \tag{11}$$

Assume the false-alarm probability is p_f. Based on the constant false-alarm criterion, it is straightforward to obtain the detecting threshold $th=\sqrt{N}\Phi^{-1}\left(1-p_f\right)$, where $\Phi(x)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x}e^{-t^2/2}dt$. The estimation whose testing statistic is larger than the detecting threshold and the statistics of other estimations is recognized as the parity check vector.

Establishing the parity check equation set require that the codeword length n and the constraint length m are known. If we do not have the prior knowledge of these parameters, we can assume the constraint length to be 16 which is large enough [6] and search for the codeword length from the smallest value. In practice, the intervals of the codeword length is $2\le n\le 9$ [9]. When the estimated codeword length is correct, there are solutions, the statistic values of which are larger than the detecting threshold. In this situation, there may be not only one solution, and the solution which corresponds to the parity check matrix with the smallest degree is identified as the parity check vector.

IV. Reconstruction of the Canonical Generator Matrix

The parity check matrix $\boldsymbol{H}(D)$ can be obtained from the parity check vector $\boldsymbol{h}$. Then according to (1), a linear equation set is established. Polynomial generator bases of the $(n,n-1,m)$ convolutional code can be estimated by solving the equation set. If the degree of $\boldsymbol{H}(D)$ is d, i.e., $\deg(\boldsymbol{H}(D))=d$, the degree of the i th row of $\boldsymbol{G}(D)$ is $e_i=\lfloor(d+i-1)/(n-1)\rfloor$ [10]. Therefore, we can establish $(d+e_i+1)$ linear equations with $(m+1)(e_i+1)$ unknowns, which are the coefficients of a polynomial generator basis. Because the number of unknowns is larger than the number of equations, the linear equation set cannot be solved by conventional Gaussian elimination algorithm.

Similarly, the equation set can be solved by our proposed recursive algorithm. But there are some differences in the recursion from that in Subsection A. First, due to all equations are correct, the summation value of unknowns can be determined by only one equation in each recursion. Second, if there are several sparsest rows of $\boldsymbol{B}$ in which the positions of 1 elements are identical, to make all equations hold, the dot products of the corresponding rows of $\boldsymbol{A}$ with any column of $\hat{\boldsymbol{H}}$ should be equal. Suppose the indices of these rows of $\boldsymbol{B}$ are $i_1,i_2,\cdots,i_q$, then

$$\boldsymbol{a}_{i_1}\cdot\hat{\boldsymbol{H}}=\boldsymbol{a}_{i_2}\cdot\hat{\boldsymbol{H}}=\cdots=\boldsymbol{a}_{i_q}\cdot\hat{\boldsymbol{H}}. \tag{12}$$

If a column of $\hat{\boldsymbol{H}}$ does not satisfy (12), the column is deleted from $\hat{\boldsymbol{H}}$. The recursion is carried out until all unknowns are estimated.

The optimal generator matrix should be canonical. After all polynomial generator bases having been estimated, some of them are selected to reconstruct the canonical generator matrix [6]. The definition and properties of a canonical generator matrix are elaborated in detail in [11]. However, there may exist not only one canonical generator matrix, and these canonical generator matrixes have the same error correction capability and encoding and decoding efficiency. As far as we know, there is no way to distinguish these canonical generator matrixes now.

V. Computational Complexity Analysis and Simulation Experiment

A. Computational Complexity Analysis

The computational complexity of the proposed method is intensive in estimating the parity check vector and verifying the correctness of the estimations. Define one operation is addition or multiplication between two elements in GF(2). If the codeword and constraint length is n and m respectively, the number of parity check equations is N and the number of the parity check vector estimations is n', the upper bound of the amount of computation for the estimation of the parity check vector is $N(2nm+2n-3)n'$ operations. Actually, the upper bound is very relax. Verifying the correctness of the estimations needs $N(2nm+2n-1)n'$ operations. Combining the amount of computation of these two parts, we drive the computational complexity of the proposed method is $O(Nnmn)$. Under the same condition, the computational complexity of Cluzeau's method is $O(Nnm2^{n(m+1)})$. Since $n'\ll 2^{n(m+1)}$, the computational complexity of our method is much lower than that of Cluzeau's method.

B. Verification of the Effectiveness of the Proposed Method

The identification of the $(4,3,2)$ convolutional code with the generator matrix

$$\boldsymbol{G}_{\mathrm{p}}(D)=\begin{bmatrix} 1+D+D^2 & 1+D & 0 & 1 \\ D & 1+D^2 & 1+D+D^2 & 1 \\ D & D & 1+D & 1+D+D^2 \end{bmatrix}$$

is considered in this experiment. The number of parity check equations is 4000. The false-alarm probability for detecting the parity check vector is 0.001. The differences between the statistic values of the estimations of the parity check vector and the detecting threshold are shown in Fig. 1.

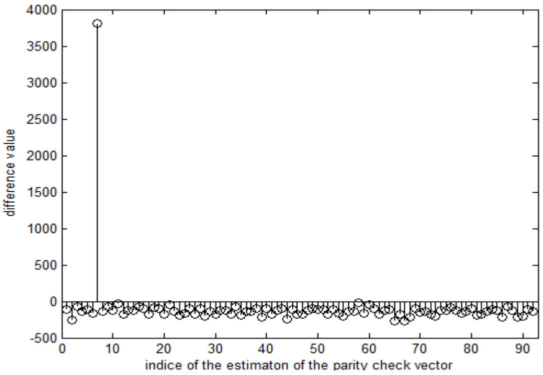

Fig. 1. Testing results of estimations of the parity check vector.

As shown in Fig. 1, only one statistic value is larger than the detecting threshold. Therefore, the corresponding vector $[0111110101110100101101011110]^{\mathrm{T}}$ is regarded as the identification result of the parity check vector.

According to the parity check matrix $\hat{\boldsymbol{H}}(D)=\begin{bmatrix}1+D^2+D^5\\1+D+D^3+D^4+D^5+D^6\\1+D^2+D^4+D^6\\D+D^2+D^4+D^5+D^6\end{bmatrix}^{\mathrm{T}}$ is obtained. Furthermore, all the polynomial generator bases is derived using $\hat{\boldsymbol{H}}(D)$, and they are $\lfloor D,D,D+D^2,1+D+D^2\rfloor$, $\lfloor D,1+D^2,1+D+D^2,1\rfloor$, $\lfloor 1+D+D^2,1+D,0,1\rfloor$, $\lfloor 1+D^2,1,D+D^2,D+D^2\rfloor$, $\lfloor 0,1+D+D^2,1,D+D^2\rfloor$, $\lfloor 1+D+D^2,D^2,1,1+D+D^2\rfloor$, $\lfloor 1+D^2,D+D^2,1+D+D^2,0\rfloor$ respectively. Any three polynomial generator bases can reconstruct a canonical generator matrix.

C. Robustness Analysis of the Proposed Method

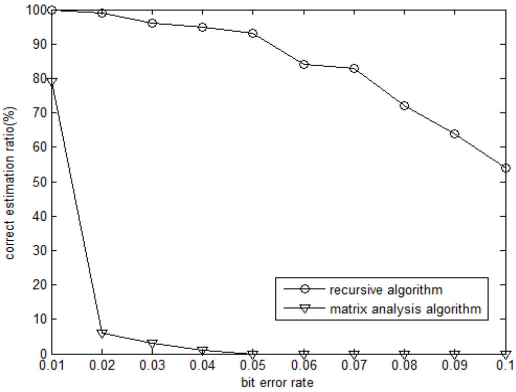

Fig. 2. Correct estimation ratio for the parity check vector of the (2,3,2) convolutional code.

The estimations of the parity check vectors of the (2,3,2) and (3,4,2) convolutional codes are considered respectively. The number of parity check equations is 4000. Assume the threshold $th_{\mathrm{eq}}=4$ in the recursive algorithm. Under different bit error rates, the correct estimation ratios of our recursive algorithm and the matrix analysis algorithm [9] for the parity check vectors of the two convolutional codes are shown in Fig. 2 and Fig. 3.

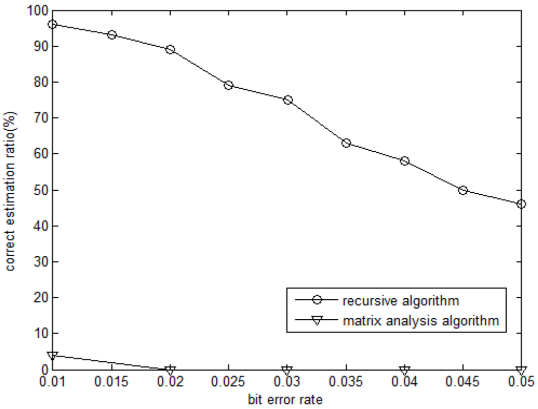

Fig. 3. Correct estimation ratio for the parity check vector of the (3,4,2) convolutional code.

As illustrated in Fig. 2 and Fig. 3, the robustness of our recursive algorithm to bit errors is much better than that of the matrix analysis algorithm. Additionally, under the same bit error rate, the correct estimation ratio for the parity check vector of the (2,3,2) convolutional code is larger than that of the (3,4,2) convolutional code. The reason is that the length of the parity check vector of the (3,4,2) convolutional code is larger and so there are more unknowns to be estimated.

VI. Conclusions

A novel method for the blind identification of a $(n,n-1,m)$ convolutional code is proposed. First, the parity check vector of a $(n,n-1,m)$ convolutional code is estimated by the proposed recursive algorithm. Then, due to the orthogonality between the generator matrix and the parity check matrix, a set of polynomial generator bases are derived. Finally, the canonical generator matrix is reconstructed by using these generator bases. The method is applicable in the cases of high bit error rates. Since the parity check vector does not need to be searched for exhaustively, its computational complexity is much lower than that of Cluzeau's method.

References

[1] K. M. Todd, *Error Correction Coding, Mathematical Methods and Algorithms*, 1st ed. Hoboken, USA: John Wiley & Sons, Inc., 2005, pp. 452-458.

[2] J. B. Cain, G. C. Clark, and J. Geist, "Punctured convolutional codes of rate (n-1)/n and simplified maximum likelihood decoding," *IEEE Transactions on Information Theory*, vol. 25, pp. 97-100, January 1979.

[3] E. Filiol, "Reconstruction of convolutional encoders over GF(q)," in *Proc. the 6th IMA conference*, 1997, pp. 101-109.

[4] M. Marazin, R. Gautier, and G. Burel, "Dual code method for blind identification of convolutional encoder for cognitive radio receiver design," in *Proc. Globecom Workshops*, Hawaii, USA, 2009, pp. 1-6.

[5] L. Huang, W. G. Chen, and E. H. Chen, "Blind recognition of k/n rate convolutional encoders from noisy observation," *Journal of Systems Engineering and Electronics*, vol. 28, pp. 235-243, February 2017.

[6] M. Cote and N. Sendrier, "Reconstruction of convolutional codes from noisy observation," in *Proc. the IEEE International Symposium on Information Theory*, Seoul, South Korea, 2009, pp. 546-550.

[7] M. Marazin, R. Gautier, and G. Burel, "Some interesting dual code properties of convolutional encoder for standards self-recognition," *IET Communications*, vol.6, pp. 931-935, August 2012.

[8] S. J. Su, J. Zhou, Z. P. Huang, C. W. Liu, and Y. M. Zhang, "Blind identification of convolutional encoder parameters," *The Scientific World Journal*, vol. 2014, pp. 1-9, May 2014.

[9] X. B. Liu, S. N. Koh, C. C. Chui, and X. W. Wu, "Reconstructing a linear scrambler with improved detection capability and in the presence of noise," *IEEE Transactions on Information Forensics and Security*, vol. 7, pp. 208-218, January 2012.

[10] P. Z. Lu, L. Shen, Y. Zou, and X. Y. Luo, "Blind recognition of punctured convolutional codes," *Science in China Ser. E Information Sciences*, vol. 35, pp. 173-185, February 2005.

[11] *Handbook of Coding Theory*, North-Holland, 1998, pp. 1065-1138.

Shu Nan Han was born in 1989. He received his B.S. degree and M.S. degree from Electronic Engineering Institute, Hefei, China in 2012 and 2015, respectively. He is currently pursuing his Ph.D. degree in National University of Defense Technology. His research interest is blind identification of channel encoder.

Min Zhang was born in 1966. He received his Ph.D. degree from Anhui University, China. Now he is a professor in National University of Defense Technology. His research interests include communication signal processing and intelligent computation.