Sequential Decoding of Lattice Codes[1]

Vahid Tarokh, Alexander Vardy, and Kenneth Zeger
Coordinated Science Laboratory, Dept. of Electrical Engineering, University of Illinois
1308 W. Main St., Urbana, IL 61801 USA.
Email: {v-tarokh, vardy, zeger}@shannon.csl.uiuc.edu

Abstract — **Viterbi decoding of lattices codes is optimal, but its computational complexity grows exponentially as a function of the coding gain. An alternative algorithm is sequential decoding. We demonstrate that sequential decoding is a computationally efficient near-optimal technique for decoding lattice codes. Furthermore, in contrast to convolutional codes, there is no possibility of buffer overflow for sequential decoding.**

I. Introduction

The computational complexity of the Viterbi decoding algorithm is known to increase exponentially as a fucntion of the coding gain [12]. Sequential decoding is known to have near-optimal performance. The main result of our paper is to demonstrate that sequential decoding of lattice codes is computationally efficient (Propositions 1 and 2).

Exponential growth of the trellis diagrams of lattices as a function of the coding gain was established in [12]. This result motivates development of efficient sub-optimal trellis decoding algorithms for lattice codes. Sequential decoding is perhaps the most well-known sub-optimal trellis decoding algorithm. Historically, it has been studied for decoding convolutional codes and has an older history than maximum likelihood decoding [5]. The success of sequential decoding for convolutional codes, both in theoretical and practical terms, and the growing applications of lattice codes in communications motivate a similar study for lattice codes. Such a study is the subject of this work.

This paper is organized as follows. In Section II, we use an argument of Massey to derive the Fano metric for sequential decoding of Euclidean codes. We then introduce lattice tree codes. These codes are easy to encode and are suitable for sequential decoding purposes. Moreover, they benefit from the structure of the defining lattice. We derive an appropriate sequential decoding metric for these codes. Section III is devoted to analyzing the running time of the sequential decoding algorithm for lattice tree codes. An analog of Pareto's distribution is obtained which upper bounds the expected running time of the algorithm. In Section IV, we study the implications of the results of Section III. We demonstrate that the label group complexity plays an important role in the sequential decoding of lattice codes. Since this is known to grow at most linearly in terms of the coding gain, sequential decoding is an attractive alternative to Viterbi decoding of lattice codes. This is similar to the case of convolutional codes. It is well known that Viterbi decoding of convolutional codes with large constraint lengths has large complexity. Nevertheless, sequential decoding has been successfully implemented in practice. In contrast to the case of convolutional codes, there is no possibility of buffer overflow with sequential decoding.

We start by reviewing the construction of trellis diagrams for lattices [6]. Let Γ be a lattice. and assume that

$$\{(0,\cdots,0)\} = V_0 \subseteq V_1 \subseteq \cdots \subseteq V_N = \mathbb{R}^N$$

is a nested sequence, or *chain*, of vector spaces in $\mathbb{R}^N$ with $\dim V_i = i$. Let $P_j : \mathbb{R}^N \to V_j$ denote the projection operator on V_j, and let $\Gamma_j = \Gamma \cap V_j$. Then Γ_j is a lattice and is called the *intersection lattice* at time j, and $P_j(\Gamma)$ is a lattice called the *projection lattice* at time j.

The *state space* at time $j = 0, 1, \ldots, N$ (with respect to the chain $V_0, V_1, \ldots, V_N$) is defined to be the abelian group $\Sigma_j(\Gamma) = P_j(\Gamma)/\Gamma_j$. For $x \in \Gamma$, define

$$\sigma_j(x) \;=\; P_j(x) + \Gamma_j \;\in\; \Sigma_j(\Gamma)$$

and $\boldsymbol{\sigma}(x) = (\sigma_0(x), \ldots, \sigma_N(x))$. We say that x *passes through* $\boldsymbol{\sigma}(x)$.

For every $j = 1, 2, \ldots, N$, let W_j be the orthogonal complement of V_{j-1} in V_j and let $P_{W_j} : \Gamma \to W_j$ denote the projection operator on W_j. Then, for $j = 1, 2, \ldots, N$, the *label group* $G_j(\Gamma)$ at time j is defined to be $P_{W_{j+1}}(\Gamma)/(\Gamma \cap W_{j+1})$. For $x \in \Gamma$, let

$$l_j(x) \;=\; P_{W_j}(x) + (\Gamma \cap W_j) \;\in\; G_{j-1}(\Gamma)$$

and let $\boldsymbol{l}(x) = (l_1(x), \ldots, l_N(x))$. The *trellis diagram* $\mathcal{T}$, of Γ, with respect to the chain $V_0, V_1, \ldots, V_N$, is defined to be a directed graph whose vertices are partitioned into a collection of sets called *levels*. The vertices of level j are the elements of $\Sigma_j(\Gamma)$. Edges from vertices of level j to those of level $j+1$ are labeled by the elements of the label group $G_j(\Gamma)$. Each $x \in \Gamma$ defines an *associated path* in the trellis diagram, corresponding to the state sequence $\boldsymbol{\sigma}(x)$ and the label sequence $\boldsymbol{l}(x)$. In other words, there is an edge between a state $\sigma_1 \in \Sigma_j(\Gamma)$ and a state $\sigma_2 \in \Sigma_{j+1}(\Gamma)$ if and only if there exists $x \in \Gamma$ such that $\sigma_j(x) = \sigma_1$ and $\sigma_{j+1}(x) = \sigma_2$. The graph of $\mathcal{T}$ without the edge labels is referred to as the *defining graph* of $\mathcal{T}$.

[1]This work was supported in part by the National Science Foundation under Grant NCR–9415860.

A lattice Γ has a *finite trellis* if there exists a trellis diagram for Γ, with respect to some chain $V_0, V_1, \ldots, V_N$, having a finite number of edges. Let $\mathcal{L}$ denote the collection of all lattices Γ with finite trellises. If $\Gamma \in \mathcal{L}$ then the elements of the label group at time j are translates of a one-dimensional lattice $v_j\mathbb{Z}$. The vectors $v_0, v_1, \ldots, v_{N-1} \in \Gamma$ defined in this way are orthogonal, that is $v_i \cdot v_j = 0$ for $i \neq j$. Conversely, as shown in [12], if there exist non-zero vectors $v_0, v_1, \ldots, v_{N-1} \in \Gamma$ such that $v_i \cdot v_j = 0$ whenever $i \neq j$, then $\Gamma \in \mathcal{L}$. In the present work, the collection $\mathcal{L}$ is of special interest. As any rational lattice is in $\mathcal{L}$ (cf. [12]), our attention is further restricted to the family of rational lattices.

II. The Fano metric

In 1963, Fano heuristically postulated a metric for the purpose of sequential decoding of convolutional codes [2]. Massey proved that this metric is in fact the correct criterion for this purpose [11]. We exploit the basic argument of [11] to derive the optimum metric for sequential decoding of lattice codes.

Definition: A variable length *Euclidean code* $\mathcal{C}$ is a finite subset of $\cup_{i=1}^{n}\mathbb{R}^i$ for some positive integer n. Each element of $\mathcal{C}$ is called a *codeword* and the vector dimension of a codeword is called its *length.*

Let $\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_M\}$ be a variable length Euclidean code whose codewords are respectively of lengths $\{n_1, n_2, \cdots, n_M\}$, and where $x_{m,i}$ denotes the i-th coordinate of $\mathbf{x}_m$, starting at $i = 0$. Let N be any positive integer larger than $\max_m(n_m)$. To each message $\mathbf{x}_m = [x_{m,0}x_{m,2}\cdots x_{m,n_m-1}]$ having probability P_m, a random *tail* $\mathbf{t}_m = [t_{m,n_m}\cdots t_{m,N-1}]$ is appended, where $t_{m,j} \in S_j$. The resulting word is denoted $\mathbf{z} = [z_0 z_1, \cdots z_{N-1}] = [x_{m,0}x_{m,1}\ldots x_{m,n_m-1}t_{m,n_m}\cdots t_{m,N-1}]$ and is sent over the channel. It is assumed that $\mathbf{x}_m$ and the $t_{m,j}$ are mutually independent random variables, for $n_m \leq j \leq N-1$, and we denote the probability mass function of $t_{m,j}$ by $p_j(\cdot)$.

We assume throughout this paper that the components of the channel noise are i.i.d. Gaussian random variables with variance σ^2 and joint density function $\mathbf{f}(\mathbf{y} \mid \mathbf{z}) = \prod_{i=0}^{N-1} f(y_i \mid z_i)$, and we let $\mathbf{y} = (y_0, y_1, y_2, \cdots, y_{N-1}) \in \mathbb{R}^N$ denote the word received from the channel. By independence, $Pr(\mathbf{t}_m \mid \mathbf{x}_m) = Pr(\mathbf{t}_m) = \prod_{k=n_m}^{N-1} p_k(t_{m,k})$, and the joint probability density of appending a tail $\mathbf{t}_m$ to a codeword $\mathbf{x}_m$ and receiving $\mathbf{y}$ is

$$\begin{aligned}\mathbf{f}(\mathbf{x}_m, \mathbf{t}_m, \mathbf{y}) &= P_m \Pr(\mathbf{t}_m)\mathbf{f}(\mathbf{y} \mid \mathbf{x}_m\mathbf{t}_m) \\ &= P_m \prod_{k=n_m}^{N-1} p_k(t_k) \cdot \prod_{j=0}^{n_m-1} f(y_j \mid x_{m,j}) \\ &\quad \cdot \prod_{i=n_m}^{N-1} f(y_i \mid t_{m,i}).\end{aligned}$$

Summing over all random tails gives

$$\mathbf{f}(\mathbf{x}_m, \mathbf{y}) = P_m \prod_{k=0}^{n_m-1} f(y_k \mid x_{m,k}) \cdot \prod_{i=n_m}^{N-1} f_i(y_i),$$

where

$$f_k(y_k) = \sum_{w \in S_k} f(y_k \mid w)p_k(w).$$

For a given received word $\mathbf{y}$, the maximum a posteriori decoding rule is to choose an $\mathbf{x}_m$ which maximizes $\Pr(\mathbf{x}_m|\mathbf{y})$. Equivalently, one can maximize

$$\mathbf{f}(\mathbf{x}_m, \mathbf{y}) / \prod_{i=0}^{N-1} f_i(y_i)$$

since the denominator is independent of $\mathbf{x}_m$. Taking logarithms, the final statistic to be maximized by the optimum decoder is

$$L(\mathbf{x}_m, \mathbf{y}) = \sum_{i=0}^{n_m-1}\left[\log\left(\frac{f(y_i \mid x_{m,i})}{f_i(y_i)}\right) + \frac{1}{n_m}\log(P_m)\right].$$

We refer to $L(\mathbf{x}_m, \mathbf{y})$ as the *Fano metric* for sequential decoding of Euclidean codes, and we apply it to the sequential decoding of lattice codes.

Definition: For a given finite tree whose edges are labeled with real numbers, a *tree code* T is a set of codewords of the form $\mathbf{x}_m = x_0x_1\cdots x_{n_m-1} \in \mathbb{R}^{n_m}$, a concatenated sequence of labels of a path of the tree starting from the root and ending in a leaf node of the tree.

A sequential decoder is one that computes a subset of all the paths in a trellis diagram in a sequential manner. At each stage, each new path is an extension of a path previously examined in the previous stage. The decision as to which path is extended at each stage is based only on already examined paths in the previous stages.

It is well known that the performance of a sequential decoder at high rates is essentially identical to that of maximum likelihood decoding. This was proven for convolutional codes by Forney and can be easily generalized to lattice codes. Simulation results also indicate that this is indeed the case when sequential decoding is applied to the trellis diagram of lattices.

One can view sequential decoding as a method for obtaining a good estimate of the path followed by the encoder of a tree code. The set of vertices at depth i from the root of the tree is referred to as *level* i and is denoted by $lev(i)$. The set $B(i)$ of edges going between elements of $lev(i)$ and $lev(i+1)$ is called the *branch space* at time i. Let r_i denote the outdegree of each vertex of level i and assume that no two paths in the tree correspond to the same codeword.

Suppose that $\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_M\}$ denotes the set of paths that have been explored up to the present by a sequential decoder. It is assumed that the decoder knows nothing about the labels in the unexplored part of the encoding tree except that they are selected independently according to the corresponding distributions $p_k(\cdot)$. However, by performing one computation, it can obtain the knowledge of the labels of the branches stemming from the terminal nodes on any already explored path. A sequential decoding algorithm can be thought of as a rule for deciding which of these paths to extend [11].

The decoding problem is to decide which of the explored paths is the initial portion of the path followed by the encoder. This is precisely the variable length decoding problem considered in [11].

Definition: Consider a tree with levels $0, 1, \cdots, N$, such that each vertex of $lev(i)$ points to r_i vertices in $lev(i+1)$. Let $B(i)$ be the set of edges from $lev(i)$ to $lev(i+1)$, and suppose each edge $e \in B(i)$ is labeled with a set label(e) of q_i real numbers. A *generalized tree code* **GTC** associated with the tree is the union of all sets $\{a_0 \cdots a_{N-1} : a_i \in \text{label}(e_i)\ \forall i\}$ of length-N codewords, taken over all paths $e_0 \cdots e_{N-1}$ in the tree. It is required that each codeword $a_0 \cdots a_{N-1}$ uniquely determines a path $e_0 e_1 \cdots e_{N-1}$ satisfying $a_i \in \text{label}(e_i)$.

Clearly **GTC** has $r_0 q_0 r_1 q_1 \cdots r_{N-1} q_{N-1}$ codewords. A **GTC** can be represented by a tree code T with levels $0, 1, 2, \cdots, N$ such that each vertex of level i points to $r_i q_i$ vertices in level $i+1$ for all i. We refer to T as the *equivalent tree code.*

We assume that the codewords of any code considered here are equally likely, so that

$$P_m = \frac{1}{r_0 q_0 r_1 q_1 \cdots r_{n_m-1} q_{n_m-1}}.$$

Applying the optimal decision statistic (1) to the equivalent tree code T gives

$$\begin{aligned} L(\mathbf{x}_m, \mathbf{y}) &= \sum_{i=0}^{n_m-1} \left[\log\left(\frac{f(y_i \mid x_{m,i})}{f_i(y_i)}\right) \right. \\ &\quad \left. - \frac{1}{n_m} \sum_{k=0}^{n_m-1} (\log r_k + \log q_k) \right] \\ &= \sum_{i=0}^{n_m-1} \left[\log\left(\frac{f(y_i \mid x_{m,i})}{f_i(y_i)}\right) - \log r_i - \log q_i \right] \end{aligned}$$

for sequential decoding of T and hence for **GTC**. To each branch having label $x_{m,i}$ in level i of the tree code **GTC**, we associate the *branch metric*

$$\log \frac{f(y_i \mid x_{m,i})}{f_i(y_i)} - \log r_i - \log q_i.$$

Let q be a positive odd integer and let $b + v\mathbb{Z}$ denote the translate of a one dimensional lattice $v\mathbb{Z}$ by a vector $b = cv$ where $v, b \in \mathbb{R}^n$, and $c \in (-1/2, 1/2]$. A *q-interval* based on $b + v\mathbb{Z}$ is defined as

$$I(q, b+v\mathbb{Z}) = \{b - \frac{q-1}{2}v, \cdots, b-v, b, b+v, \cdots, b+\frac{q-1}{2}v\}.$$

Let Γ be a rational lattice of dimension N and let $\mathcal{T}$ be a finite trellis diagram of Γ. Then the label of any edge $e_i \in B(i)$ in the trellis $\mathcal{T}$ is a translate of a one-dimensional lattice $v_i\mathbb{Z}$, and $v_i \cdot v_j = 0$ whenever $0 \leq i \neq j \leq N-1$ [5] [12]. Without loss generality, it is assumed that v_i points in the direction of the i-th standard coordinate vector.

Definition: A *lattice tree code* LTC based on Γ and $\mathcal{T}$ is a generalized tree code obtained by replacing each element $b + v_i\mathbb{Z}$ of the label group at time i with a q_i-interval $I(q_i, b + v_i\mathbb{Z})$ in $\mathcal{T}$ for all i. The *reduced trellis diagram* $\mathcal{RT}$ of LTC, given $\mathbf{y} = (y_0, y_1, \cdots, y_{N-1}) \in \mathbb{R}^N$ is the directed graph obtained by replacing label(e_i) by the element of label(e_i) closest to y_i. This is done for all i and for all $e_i \in B(i)$.

III. Running Time

Definition: The *running time* of a sequential decoding algorithm for a lattice tree code LTC is the expected number of decoding operations performed upon receiving the vector $\mathbf{y} \in \mathbb{R}^N$. Let LTC denote a lattice tree code based on a trellis diagram $\mathcal{T}$ of an N-dimensional lattice Γ. For $i = 0, 1, \ldots, N-1$, let $p_i(\cdot) = 1/|S_i|$ denote the probability mass function of a uniform random variable X_i taking values in the set S_i. Let RAND denote the ensemble of random codes generated by assigning q_i independent outcomes of X_i to branches between levels i and $i+1$ of the defining graph of $\mathcal{T}$. Elements of RAND are referred to as *random trellises* or *random codes.*

It is presently unknown how to compute the running time of the sequential decoding algorithm for LTC. Instead, we investigate the average running time of the algorithm taken over all the members of RAND.

Assume that $\mathbf{y} \in \mathbb{R}^N$ is received, and a random code $\mathcal{C} \in$ RAND is given. Let x be the closest element of $\mathcal{C}$ to $\mathbf{y}$. The path P associated with $x \in \mathcal{C}$ is referred to as the *correct path.* The time-j *incorrect subset* I_j is the set of those paths of $\mathcal{C}$ originating from the j-th vertex of the correct path P that are not tails of P_j. The *number of time-j incorrect computations* C_j is the number of paths in the time-j incorrect subset that are expanded by the sequential decoder.

The quantity C_0 is of particular interest and the following theorem sheds light on the average of C_0 over the ensemble RAND and over the set of all possible received words $\mathbf{y}$.

Proposition 1 *For every $\rho \in (0, 2]$, there exists $\theta \geq 0$, such that for all $M \in \mathbb{N}$*

$$\Pr\{C_0 \geq M\} \;\leq\; \theta M^{-\rho}. \qquad (1)$$

IV. Lattice Decoding

In this section, we study the application of previous results to decoding lattice codes and designing signal constellations. Let $\Gamma \in \mathcal{L}$ be an N-dimensional lattice with a finite trellis diagram $\mathcal{T}$. Let $d(\Gamma)$ denote the minimum distance of Γ and let LTC be a lattice tree code based on Γ and $\mathcal{T}$. We use the notations of the previous sections. In particular σ denotes the noise variance per dimension.

Any computation of real numbers performed by a machine has some degree of limited precision accuracy. We define the *decoder's finite precision error* as the error that the decoder allows when computing a branch metric while performing the sequential decoding algorithm.

Proposition 2 *Suppose that the decoder's finite precision error is at most* $\epsilon > 0$. *Then for fixed* $d(\Gamma)/\sigma$ *the branch metrics at time* i *can be computed with* $K \cdot |G_i|$ *operations and with an error of at most* 2ϵ, *where* K *is an appropriate constant.*

Clearly the above result leads to a significant decrease of the complexity of the sequential decoding algorithm.

Let G_i denote the label group at time i for $\mathcal{T}$. Suppose that that $d(\Gamma)/\sigma$ and the decoder's computation accuracy $\epsilon > 0$ are fixed. Let K be as in Proposition 2, then at most $K \sum_{j=1}^{N} |G_j|$ operations must be performed for computing the branch metrics of all branches of $\mathcal{T}$. The sum $\sum_{j=1}^{N} |G_j|$ thus plays an important role in determining whether LTC is attractive for sequential decoding purposes. In [12] $\mathcal{G}(\Gamma)$ was defined to be the minimum of the sums $\frac{1}{N} \sum_{j=1}^{N} |G_j|$ taken over all possible trellis diagrams of Γ. Moreover, using tensor products and Kitaoka's theorem, families of lattices $\{\Gamma_i, i = 1, 2, \cdots\}$ such that $\mathcal{G}(\Gamma_i)$, grows linearly as a function of the coding gain of Γ_i were constructed. Later Forney [7] observed that the Barnes-Wall family can be constructed from the same construction. Thus it seems that the Barnes-Wall family and trellises constructed from tensor products are attractive for sequential decoding purposes. The higher dimensional members of this family however, are not attractive for trellis-based maximum likelihood decoding purposes as their asymptotic coding gain grows without bound [12]. This is similar to the case of convolutional codes and TCM schemes. Trellises having too many states are not attractive for Viterbi decoding but have been successfully implemented in practice using sequential decoding.

One difference between sequential decoding of lattices and that of convolutional codes is that there is no possibility of buffer overflow in the former case. This is due to the finiteness of the trellis in this case and could make lattice tree codes attractive for some applications.

Acknowledgement. The authors would like to thank Dr. David Forney and Dr. Bert Hochwald for many useful suggestions.

References

[1] J.H. Conway and N.J.A. Sloane, *Sphere Packings, Lattices and Groups*, New York: Springer-Verlag, 1993.

[2] R.M. Fano, "A heuristic discussion of probabilistic decoding," *IEEE Trans. Inform. Theory*, vol. 9, pp. 64–73, 1963.

[3] G.D. Forney Jr. and M.D. Trott, " The dynamics of group codes: state spaces, trellis diagrams, and canonical encoders," *IEEE Trans. Inform. Theory*, vol. 39, pp. 1491–1513, 1993.

[4] G.D. Forney Jr., "Convolutional codes II: maximum likelihood decoding," *Inform. and Control*, vol. 25, pp. 222–266, 1974.

[5] G.D. Forney Jr., "Convolutional codes III: sequential decoding," *Inform. and Control*, vol. 25, pp. 267–297, 1974.

[6] G.D. Forney, "Density/length profiles and trellis complexity of lattices," *IEEE Trans. Inform. Theory*, Vol 40, pp. 1753-1772, 1995.

[7] G.D. Forney Jr., Private communication.

[8] R.G. Gallager, "A simple derivation of the coding theorem and some applications," *IEEE Trans. Inform. Theory*, vol. 11, pp. 3–18, 1965.

[9] R.G. Gallager, *Information Theory and Reliable Communication*, John Wiley & Sons, 1968.

[10] I.M. Jacobs and E.R. Berlekamp, "A lower bound to the distribution of computation for sequential decoding," *IEEE Trans. Inform. Theory*, vol. 18, pp. 196–198, 1972.

[11] J.L. Massey, "Variable-length codes and the Fano metric," *IEEE Trans. Inform. Theory*, vol. 18, pp. 196–198, 1972.

[12] V. Tarokh, *Trellis Complexity versus the Coding Gain of Lattice-based Communication Systems*, Ph.D. Thesis, University of Waterloo, Ontario, Canada, 1995.