# FAST MAXIMUM LIKELIHOOD SEQUENCE DETECTION OVER VECTOR INTERSYMBOL INTERFERENCE CHANNELS

*Jie Luo*

ECE Dept., Colorado State Univ., Fort Collins, CO 80523
rockey@engr.colostate.edu

## ABSTRACT

We consider the communication system that transmits a sequence of binary vector symbols over a vector intersymbol interference (ISI) channels subject to additive white Gaussian noise . Conventionally, maximum likelihood (ML) sequence is computed using the Viterbi Algorithm (VA), whose complexity scales exponentially in both the symbol vector length and the number of ISI channel taps. We show that, as the signal to noise ratio (SNR) goes to infinity, the ML sequence can be obtained with an asymptotic complexity scaling linearly in the number of channel taps and quadratically in the symbol vector length.

*Index Terms*— Maximum likelihood, Sequence detection, Statistical information, Viterbi algorithm

## 1. INTRODUCTION

Consider the scenario where a sequence of vector symbols, with each vector having $K$ binary elements, are sent from a transmitter to a receiver through a vector intersymbol interference (ISI) channel, whose number of taps is $L$, subject to additive Gaussian noise. Assume the source vector symbols are independently generated with all possible values being equal probable. If the receiver is willing to minimize the probability of sequence detection error, the optimal decision is given by the maximum likelihood (ML) sequence that maximizes the log likelihood function. Finding such sequence is known as the maximum likelihood sequence detection (MLSD) problem.

Conventionally, the ML sequence is computed using the well known Viterbi algorithm (VA) [1], whose complexity scales linearly in the sequence length, but exponentially in the source symbol vector length $K$, and exponentially in the number of ISI channel taps $L$. Such complexity can be prohibitive for systems with large $KL$ values. Throughout the past three decades, many attempts have been made to find sequence detectors performing about the same as the VA, but less complex in terms of the scaling law in the Markov states [3][4][5]. The main idea considered in these algorithms is to update only a selected number of routes upon the reception of each observation so that the worst case complexity of the algorithm is un-der control. However, a consequence of such limited search is that none of these complexity-reduction methods can guarantee the ML sequence, which is the sequence that maximizes the log likelihood function. On the other hand, if the length of the input vector sequence, $N$, is small, one can regard the MLSD problem as a maximum likelihood (ML) lattice decoding problem with an input symbol vector of length $NK$ [7]. Consequently, ML sequence can be obtained using various versions of the sphere decoding algorithm with low average complexity, under the assumption of high signal to noise ratio (SNR) [6][7]. Unfortunately, due to the difficulty of handling a lattice of infinite dimension, these algorithms cannot extend directly to the situation of stream input where the length of the source sequence is practically infinity. In summary, most existing complexity reduction methods for MLSD either cannot guarantee the ML sequence, or are not suitable for stream input.

Although the VA is computation efficient in the sense of exploiting the underlying Markov chain structure, it does not fully exploit the statistical information of the system. Particularly, the observations of the system are related to the Markov states through a statistic model, which is usually known to the receiver. If the observation perturbation is small, the observation sequence provides a strong inference about the underlying Markov states. Such information can be used to significantly reduce the number of routes one should visit in the VA. For the communication system studied in this paper, we develop an examination method which guarantees the truncated sequence passing the examination is indeed the truncated ML sequence. As SNR goes to infinity, the proposed examination method becomes highly efficient in the sense of passing the actual truncated source sequence with asymptotic probability one. Together with the help of an asymptotically efficient sequential detector whose probability of symbol detection error is asymptotically zero, the proposed algorithm obtains the ML sequence with an asymptotic complexity of $O(LK^2)$ per symbol. This complexity is, in the scaling order sense, no higher than any of the efficient sequence detectors, including suboptimal ones, that can achieve diminishing symbol detection error as SNR goes to infinity. In the situation of finite-length input sequence, the worst case complexity of the MLSD algorithm is in the same order of the VA.

The proposed MLSD algorithm is presented in a simple form in order to show clearly the insight of asymptotic complexity reduction. We make no effort in reducing the complexity further as long as the desired asymptotic scaling law is achieved. The proofs of the theorems are presented in [8].

## 2. SYSTEM MODEL

Let $\boldsymbol{x}(0), \boldsymbol{x}(1), \ldots, \boldsymbol{x}(n), \ldots, \boldsymbol{x}(N-1)$ be a sequence of source symbol vectors, each of length $K$, with binary-valued elements. Let $x_i(n)$ be the $i^{\text{th}}$ element of $\boldsymbol{x}(n)$. $x_i(n) \in \{-1, +1\}$, $\forall i, n$. Assume the source vectors are randomly and independently generated, with all possible values being equal probable. The sequence of vectors are then transmitted to the receiver through a vector ISI channel. The $n^{\text{th}}$ received symbol vector, also of length $K$, is given by,

$$\boldsymbol{y}(n) = \sum_{l=0}^{L-1} \boldsymbol{F}[l]\boldsymbol{x}(n-l) + \boldsymbol{v}(n) \tag{1}$$

Here $\boldsymbol{F}[l]$, $l = 0, \ldots, L-1$ are the channel matrices, each of size $K \times K$ with real-valued elements; $\boldsymbol{v}(n)$ is the $K$ length white Gaussian noise vector with zero mean and variance $\sigma^2 \boldsymbol{I}$. We assume all the channel matrices are known to the receiver and are time-invariant throughout the communication. We assume the input sequence length $N$ is large enough to be practically considered as infinity. We also assume $\boldsymbol{x}(n) = \boldsymbol{0}$ for $n < 0$ and $n \geq N$.

The following is a key assumption required for the derivation of the results in Section 3. A discussion about this assumption is given at the end of Section 3.

**Assumption 1** *We assume the channel matrix $\boldsymbol{F}[0]$ is lower triangular with all its diagonal elements being strictly positive. Define $\boldsymbol{F}(\omega)$ as the vector Fourier transform of the channel matrices,*

$$\boldsymbol{F}(\omega) = \sum_{l=0}^{L-1} \boldsymbol{F}[l]e^{-j\omega} \tag{2}$$

*Define $\lambda_{\min}(\boldsymbol{A})$ as the minimum eigenvalue of a Hermitian matrix $\boldsymbol{A}$. Let $\boldsymbol{F}(\omega)^H$ be the conjugate transpose of $\boldsymbol{F}(\omega)$. Define*

$$\lambda_0 = \min_{\omega \in [0, 2\pi)} \lambda_{\min}(\boldsymbol{F}(\omega)^H \boldsymbol{F}(\omega)) \tag{3}$$

*We assume $\lambda_0 > 0$.*

Define

$$\Psi_i(\{\boldsymbol{x}(n)\}) = -\sum_{n=0}^{i} \left\| \boldsymbol{y}(n) - \sum_{l=0}^{L-1} \boldsymbol{F}[l]\boldsymbol{x}(n-l) \right\|^2 \tag{4}$$

We use $\Psi(\{\boldsymbol{x}(n)\})$ to refer to the sum log likelihood function $\Psi_{N+L-2}(\{\boldsymbol{x}(n)\})$ corresponding to the sequence $\{\boldsymbol{x}(n)\}$.

Upon the reception of the observation sequence $\{\boldsymbol{y}(n)\}$, the optimal detection that minimized the probability of sequence detection error is given by the ML sequence.

$$\{\boldsymbol{x}^{\text{ML}}(n)\} = \arg \max_{\{\boldsymbol{x}(n)\}, \boldsymbol{x}(n) \in \{-1, +1\}^K} \Psi_{N+L-2}(\{\boldsymbol{x}(n)\}) \tag{5}$$

We are interested in the complexity of the MLSD per symbol, as $\sigma$ goes to zero.

## 3. SEQUENCE DETECTION

We first present in the following theorem the optimality examination method that can verify whether a particular vector in a finite-length truncated decision sequence is identical to the corresponding vector in the ML sequence.

**Theorem 1** *Given the observation sequence $\{\boldsymbol{y}(n)\}$, let $\{\hat{\boldsymbol{x}}(n)\}$ be a decision sequences. Define*

$$U_0 = \max_{\boldsymbol{x}(i) \in \{-2, 0, +2\}^K, \forall i} \sum_{i=0}^{L-2} \left\| \sum_{l=0}^{i} \boldsymbol{F}[l]\boldsymbol{x}(i-l) \right\|^2 \tag{6}$$

*Let $0 < \delta < \frac{\lambda_0}{L}$. Let $D$ be a positive integer satisfying*

$$D > \frac{2U_0 + \frac{L-1}{L}\lambda_0}{\delta} - \frac{3}{2}L + 2 \tag{7}$$

*For any $M \geq L+D-1$, if the following inequality is satisfied for all $M - D < n < M + D + L - 1$,*

$$\left\| \boldsymbol{y}(n) - \sum_{l=0}^{L-1} \boldsymbol{F}[l]\hat{\boldsymbol{x}}(n-l) \right\|^2 < \frac{\lambda_0}{L} - \delta \tag{8}$$

*then $\hat{\boldsymbol{x}}(M) = \boldsymbol{x}^{ML}(M)$ must be true.*

The optimality examination method presented in Theorem 1 is asymptotically efficient in the following sense. If $\{\hat{\boldsymbol{x}}(n)\}$ is indeed the actual source sequence, for any $M \geq L+D-1$, the probability of (8) being true for all $M - D < n < M + D + L - 1$ goes to one as $\sigma$ goes to zero.

Let $\boldsymbol{x}_L(n) = [\boldsymbol{x}(n-L+1)^T, \ldots, \boldsymbol{x}(n)^T]^T$ be the $n^{\text{th}}$ state vector for the underlying Markov chain. $\Psi(\{\boldsymbol{x}(n)\})$ can be computed via the following recursive algorithm.

$$\begin{aligned} \Psi_M(\{\boldsymbol{x}(n)\}) &= \Psi_M(\boldsymbol{x}_L(n \leq M)) \\ &= \Psi_{M-1}(\{\boldsymbol{x}(n)\}) - \left\| \boldsymbol{y}(M) - \sum_{l=0}^{L-1} \boldsymbol{F}[l]\boldsymbol{x}(M-l) \right\|^2 \\ &= \Psi_{M-1}(\boldsymbol{x}_L(n \leq M-1)) + \gamma_M(\boldsymbol{x}_L(M-1), \boldsymbol{x}_L(M)) \end{aligned} \tag{9}$$

Here $\gamma_M(\boldsymbol{x}_L(M-1), \boldsymbol{x}_L(M))$ is the value of the path connecting Markov states $\boldsymbol{x}_L(M-1)$ and $\boldsymbol{x}_L(M)$.

$$\gamma_M(\boldsymbol{x}_L(M-1), \boldsymbol{x}_L(M)) = - \left\| \boldsymbol{y}(M) - \sum_{l=0}^{L-1} \boldsymbol{F}[l]\boldsymbol{x}(M-l) \right\|^2 \tag{10}$$

We also define

$$\Gamma_M(\boldsymbol{x}_L(M)) = \max_{\boldsymbol{x}(n \le M-L) \in \{-1,+1\}^K} \Psi_M(\{\boldsymbol{x}(n)\}) \quad (11)$$

as the value of the Markov state $\boldsymbol{x}_L(M)$. Note that when a state value is obtained, we also obtain the corresponding route in the Markov graph that achieves this value.

**The Simple MLSD Algorithm**

Initialize $M = 0$. Use the decision feedback detector to obtain $\{\hat{\boldsymbol{x}}(n) | n \le D + L - 2\}$, where $\hat{x}_i(n)$ is given by

$$\text{sign}\left(y_i(n) - \sum_{l=1}^{L-1}\sum_{j=1}^{K} f_{ij}[l]\hat{x}_j(n-l) - \sum_{j=1}^{i-1} f_{ij}[0]\hat{x}_j(n)\right) \quad (12)$$

Initialize a Markov state set $\boldsymbol{X}_{-1}$, which contains only the zero state, i.e.,

$$\boldsymbol{X}_{-1} = \{\boldsymbol{x}_L(-1)\}, \qquad \boldsymbol{x}_L(-1) = [\boldsymbol{0}^T, \dots, \boldsymbol{0}^T]^T \quad (13)$$

Let $\hat{\boldsymbol{x}}_L(L-1) = [\hat{\boldsymbol{x}}(0)^T, \hat{\boldsymbol{x}}(1)^T, \dots, \hat{\boldsymbol{x}}(L-1)^T]^T$. Compute the values of the paths connecting $\boldsymbol{x}_L(-1)$ and $\hat{\boldsymbol{x}}_L(L-1)$; also compute all the values of the Markov states passed by this path.

The algorithm then performs the following three steps recursively for each $M$.

In step 1, we compute $\hat{\boldsymbol{x}}(M+D+L-1)$ using the decision feedback algorithm (12).

In step 2, we carry out the optimality examination for $\hat{\boldsymbol{x}}(M)$. The truncated observation sequence and the decision sequence involved in the examination are $\{\boldsymbol{y}(n) | M - D < n < M + D + L - 1\}$ and $\{\hat{\boldsymbol{x}}(n) | M - D - L + 1 < n < M + D + L - 1\}$, respectively. We say $\hat{\boldsymbol{x}}(M)$ passes the examination if and only if $M \ge L + D - 1$ and Inequality (8) holds for all $M - D < n < M + D + L - 1$.

If $\hat{\boldsymbol{x}}(M)$ does pass the examination, we construct a Markov state set $\boldsymbol{X}_M$ as follows: for all Markov states $\boldsymbol{x}_L(M-1) = [\boldsymbol{x}(M-L)^T, \boldsymbol{x}(M-L+1)^T, \dots, \boldsymbol{x}(M-1)^T]^T$ in $\boldsymbol{X}_{M-1}$, we let $\boldsymbol{x}_L(M) = [\boldsymbol{x}(M-L+1)^T, \dots, \boldsymbol{x}(M-1)^T, \hat{\boldsymbol{x}}(M)^T]^T$ be a Markov state in $\boldsymbol{X}_M$. We then compute the state value $\Gamma_{M+L}(\hat{\boldsymbol{x}}_L(M+L))$.

If $\hat{\boldsymbol{x}}(M)$ does not pass the examination, in step 3, we construct the Markov state set $\boldsymbol{X}_M$ as follows: for each Markov state $\boldsymbol{x}_L(M-1) = [\boldsymbol{x}(M-L)^T, \boldsymbol{x}(M-L+1)^T, \dots, \boldsymbol{x}(M-1)^T]^T$ in $\boldsymbol{X}_{M-1}$, we add the Markov states $\boldsymbol{x}_L(M) = [\boldsymbol{x}(M-L+1)^T, \dots, \boldsymbol{x}(M-1)^T, \boldsymbol{x}(M)^T]^T$ corresponding to all possible values of $\boldsymbol{x}(M)$ into the Markov state set $\boldsymbol{X}_M$. We then compute the values of all the paths connecting the states in $\boldsymbol{X}_{M-1}$ and the states in $\boldsymbol{X}_M$. We also compute values of all the paths connecting the states in $\boldsymbol{X}_M$ to the Markov state $\hat{\boldsymbol{x}}_L(M + L)$. The values for the states in $\boldsymbol{X}_M$ and the state value $\Gamma_{M+L}(\hat{\boldsymbol{x}}_L(M + L))$ are computed.

Let $M = M + 1$ and repeat the three steps for $M$. $\diamondsuit$

Since if $\hat{\boldsymbol{x}}(M)$ passes the optimality examination, we know $\boldsymbol{x}^{\text{ML}}(M) = \hat{\boldsymbol{x}}(M)$, the ML route must pass one of the Markov states in Markov state set $\boldsymbol{X}_M$ at time index $M$.

In step 3, when we compute the state value $\Gamma_{M+L}(\hat{\boldsymbol{x}}_L(M+L))$, all values of the Markov states along the pathes from the states in $\boldsymbol{X}_M$ to $\hat{\boldsymbol{x}}_L(M + L)$ are also obtained. Consequently, in step 2, we do not need extra computation to obtain the values of the states in $\boldsymbol{X}_M$ since they are all on the paths from states in $\boldsymbol{X}_{M-1}$ to $\hat{\boldsymbol{x}}_L(M + L - 1)$. Since the simple MLSD algorithm always compute the values of the Markov states in $\boldsymbol{X}_M$, the values of all the states on the ML path must have been enclosed.

The computation of the simple MLSD algorithm contains three parts.

The first part is the computation to obtain the suboptimal decision sequence $\{\hat{\boldsymbol{x}}(n)\}$. This complexity is in the order of $O(LK^2)$ per symbol.

The second part is to carry out the optimality examination for $\hat{\boldsymbol{x}}(n)$ (for each $n$) and update the path values and state values if $\hat{\boldsymbol{x}}(n)$ passes the examination. Since we do not perform the examination on any sequence other than $\{\hat{\boldsymbol{x}}(n)\}$, by sharing temporary results among successive examinations, the complexity of running the examination is in the order of $O(LK^2)$ per symbol. If $\hat{\boldsymbol{x}}(n)$ passes the examination, we need to compute the state value $\Gamma_{n+L}(\hat{\boldsymbol{x}}_L(n + L))$. Fortunately, since all paths connecting the states in $\boldsymbol{X}_n$ and $\hat{\boldsymbol{x}}_L(n+L)$ pass the Markov state $\hat{\boldsymbol{x}}_L(n + L - 1)$, whose value has already been obtained in previous steps corresponding to time index $n - 1$. To obtain the state value of $\hat{\boldsymbol{x}}_L(n + L)$, we only need to compute the value of the path connecting $\hat{\boldsymbol{x}}_L(n+L-1)$ and $\hat{\boldsymbol{x}}_L(n+L)$. The overall complexity of the second part is then in the order of $O(LK^2)$ per symbol vector.

The third part is the path values and state values updates when $\hat{\boldsymbol{x}}(n)$ does not pass the examination. This part of complexity is in the order of $O(2^{LK})$ per symbol.

**Theorem 2** *Assume the probability of symbol detection error of the decision feedback detector goes to zero as $\sigma \to 0$. If the input sequence is of infinite length, the complexity of the simple MLSD is in the order of $O(LK^2)$ per symbol as $\sigma \to 0$. If the input sequence has a finite length, then the worst case complexity per symbol of the simple MLSD algorithm is in the order of $O(2^{LK})$.*

We can conceptually think the MLSD algorithm contains three components. We have an asymptotically efficient suboptimal sequence detector, whose symbol detection error goes to zero as $\sigma \to 0$. We also developed an asymptotically efficient optimality examination method, that can check whether each vector of the obtained sequence is identical to its ML sequence correspondence. If we are unable to tell whether a particular vector is in the ML sequence or not, we use a backup search plan similar to the VA to make sure we do not miss the ML route.

Note that in the VA, a Markov route is eliminated only when it passes the same state at a particular time index with another Markov route, who achieves a lower state value. Therefore, a route comparison in the VA can only disprove, rather

than verify, the optimality of a Markov route. In other words, no matter how small the state value is, we cannot say a route indeed gives the truncated ML sequence unless all the survival routes merge naturally. The optimality examination used in the simple MLSD algorithm, on the other hand, can verify, rather than disprove, the optimality of a vector in a truncated decision sequence.
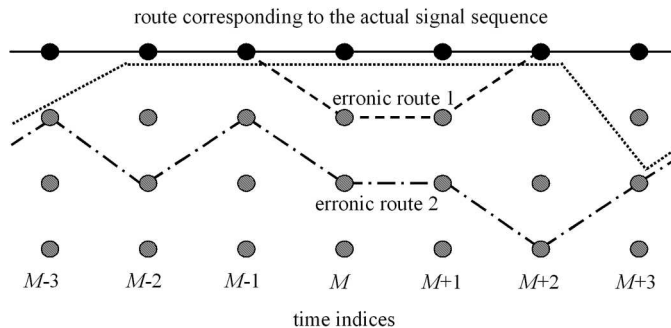


**Fig. 1**. Illustration of the optimality examination in the simple MLSD algorithm.

The idea of designing the optimality examination method can be explained using the Markov graph illustrated in Figure 1. Suppose the Markov states corresponding to the actual source sequence are given by the solid dots. We term these Markov states and the paths connecting them the actual states and the actual paths, respectively. We want to check wether the particular state at time index $M$ is optimal, i.e., whether the ML route passes the Markov state illustrated by the solid dot at time index $M$. Since when the noise power is small, the values of the actual paths are close to zero, while the values of other erronic paths are, in general, significantly lower than zero. If a route does not pass the actual state at time index $M$, we say it makes a decision error at time index $M$. Suppose the erronic route merges with the actual route after making several decision errors, as shown by "erronic route 1" in Figure 1. Due to the fact that all the values of the actual routes are close to zero, the sum log likelihood of the erronic route are usually lower than the sum log likelihood of the actual route. Now, suppose the erronic route does not merge with the actual route after making significant number of decision errors, as illustrated by "erronic route 2" in Figure 1. We can construct a new route that connects the actual paths to the erronic route, as illustrated by the dotted route in Figure 1. Although in making such connection, the constructed route also makes several decision errors and hence can have low sum log likelihood value, the value can still be larger than the sum log likelihood of erronic path 2 since the number of decision errors made by erronic route 2 is much more significant.

Note that as long as the probability of symbol detection error of the ML detector can go to zero as $\sigma \rightarrow 0$, without requiring Assumption 1, we can find asymptotically efficient optimality examination method, similar to the one presented in Theorem 1, with a complexity per symbol in the order of $O(LK^2)$. However, if Assumption 1 is not true, finding an asymptotically efficient suboptimal detector with a complexity of $O(LK^2)$ per symbol becomes non-trivial, if not impossible.

## 4. CONCLUSION

We considered the maximum likelihood sequence detection (MLSD) problem of transmitting a sequence of binary vector symbols over a vector intersymbol interference channel. We show that as the signal to noise ratio (SNR) goes to infinity, the ML sequence can be obtained with a complexity of $O(LK^2)$ per symbol, where $L$ is the number of channel taps and $K$ is the vector length of the source symbol, under certain conditions. Such a complexity is no higher in order than any of the efficient sequence detectors, including suboptimal ones, that can achieve diminishing symbol detection error as SNR goes to infinity.

## 5. REFERENCES

[1] G. Forney, "The Viterbi Algorithm", Proc. of The IEEE, Vol. 61, No. 3, pp. 268-278, Mar. 1973.

[2] A. Viterbi, "A Personal History of the Viterbi Algorithm", IEEE Sig. Proc. Magazine, Vol. 23, No. 4, pp. 120-142, July 2006.

[3] G. Foschini, "A Reduced State Variant of Maximum-likelihood Sequence Detection Attaining Optimum Performance for High Signal-to-Noise Ratios", IEEE Trans. Inform. Theory, Vol. IT-23, No. 5, pp. 605-609, Sep. 1977.

[4] J. Hayes, "The Viterbi Algorithm Applied to Digital Data Transmission", IEEE Commun. Magzine, Vol. 13, No. 2, pp. 26-32, May 2002.

[5] T. Aulin, "Breadth-First Maximum-Likelihood Sequence Detection: Geometry", IEEE Trans. Commun., Vol. 51, No. 6, pp. 2071-2080, June 2003.

[6] J. Luo, K. Pattipati, P. Willett and G. Levchuk, "Fast Optimal and Sub-optimal Any-Time Algorithms for CDMA Multiuser Detection based on Branch and Bound", IEEE Trans. Commun., Vol. 52, No. 4, pp. 632-642, Apr. 2004.

[7] B. Hassibi and H. Vikalo, "On The Sphere Decoding Algorithm I. Expected Complexity", IEEE Trans. Sig. Proc., Vol. 53, No. 8, pp. 2806-2818, Aug. 2005.

[8] J. Luo, "On Low Complexity Maximum Likelihood Sequence Detection under High SNR", submitted to IEEE Trans. Info. Theory, Sep. 2006.