The Cross-Entropy Method for Blind Multiuser Detection

Zaifei Liu, Arnaud Doucet and Sumeetpal S. Singh *

December 1, 2003

Abstract

We consider the problem of blind multiuser detection. We adopt a Bayesian approach where unknown parameters are considered random and integrated out. Computing the *maximum a posteriori* estimate of the input data sequence requires solving a combinatorial optimization problem. We propose here to apply the Cross-Entropy method recently introduced by Rubinstein. The Cross-Entropy method is an efficient and principled adaptive importance sampling method which can be adapted to solve optimization problems. The performance of cross-entropy is compared to Markov chain Monte Carlo. For similar Bit Error Rate performance, we demonstrate that Cross-Entropy outperforms a generic Markov chain Monte Carlo method in terms of operation time.

Index Terms: multiuser detection, Bayesian inference, cross-entropy, Markov chain Monte Carlo.

1 Introduction

Blind multiuser detection in CDMA systems has numerous applications. It is relatively easy to come up formally with the expression of the optimal sequence estimates, say Maximum Likelihood (ML) or Maximum A Posteriori (MAP) estimates, but practically it is typically necessary to solve combinatorial optimization problems to obtain them. Consequently, optimal solutions cannot be implemented in practice because of their prohibitively high computational complexity. This has motivated the study of a number of low-complexity suboptimal techniques. Although these suboptimal methods have a reasonable complexity, their performance can be far below those achieved by the theoretically optimal ones.

The recently emerged Markov chain Monte Carlo (MCMC) methods have demonstrated a great potential for solving complex inference problems arising in wireless communications [1], [2], [10], [11]. These methods are iterative stochastic algorithms to generate random samples from complex probability distributions. In the Bayesian wireless communications context, one samples from the posterior distribution of the unknown symbols. In the CDMA context, the computational complexity only increases linearly with the user number K. However, MCMC are iterative algorithms and it is difficult to assess their convergence.

In this paper, we propose to apply the Cross-Entropy (CE) technique to address the problem of MAP sequence estimation. CE is a principled adaptive importance sampling where, at each iteration, the importance distribution is adapted so as to minimize the Kullback-Leibler distance, i.e. cross-entropy, with respect to the target distribution. It turns out that the minimization of the cross-entropy can be usually done analytically. Moreover, this procedure can be adapted so as to solve combinatorial optimization problems. This efficient method recently introduced by Rubinstein [8] has become very popular in operation research [5]. To the best of our knowledge, this paper is the first application of this methodology to wireless communications.

The rest of this paper is organized as follows. In Section 2, we detail the Bayesian model and a standard MCMC algorithm. In Section 3, we present a

*Signal Processing Group, Dept. of Engineering, Univ. of Cambridge, Cambridge CB2 1PZ, UK; zl217@eng.cam.ac.uk; ad2@eng.cam.ac.uk; ssss@ee.mu.oz.au

generic form of the CE method and show how it can be adapted to blind multiuser detection. Finally in Section 4, we demonstrate the performance of this new approach.

2 Bayesian Model and MCMC

Suppose in a CDMA system multiple users transmit symbol sequences through the same physical channel synchronously, each user employs a distinct signature waveform. Here we use the same model as in [10] yet with different priors. We then review very briefly a generic MCMC method.

2.1 Bayesian Model for Multiuser Detection

Suppose there are K users in the multiuser CDMA system. The k^{th} user employs a spreading waveform $\mathbf{s}_k$ of the form

$$\mathbf{s}_k = \frac{1}{\sqrt{P}}[s_{1,k}, s_{2,k}, ..., s_{P,k}]^{\mathrm{T}},\ s_{p,k} \in \{+1, -1\}$$

where P is the spreading factor. It is assumed that the receiver knows the spreading waveforms of all users in the system. The transmitted information symbols are $x_{k,1}, x_{k,2}, ..., x_{k,T},\ x_{k,t} \in \{+1, -1\}$, T is the number of observations. The received signal in this system is given by

$$\mathbf{y}(t) = \sum_{k=1}^{K} A_k x_{k,t}\mathbf{s}_k + \mathbf{n}(t),\ t = 0, ..., T-1$$

where A_k is the unknown amplitude of the k^{th} user; $\mathbf{n}(t) = [n_1(t), n_2(t), ..., n_P(t)]^{\mathrm{T}}$ is a zero-mean white Gaussian noise vector with variance σ^2. Our main goal is to estimate the input information sequence $X = \{x_{k,t}\}_{k=1,...,K}^{t=0,...,T-1}$, based on the received signals $Y = \{\mathbf{y}(t)\}_{t=0}^{T-1}$; the channel coefficients $\mathbf{a} = (A_1, ..., A_K)^T$ and σ^2 being unknown.

We follow a Bayesian approach where it is assumed that the parameters $\mathbf{a}$ and σ^2 are unknown and follow the following standard conjugate prior distributions [7]

$$\mathbf{a} \sim \mathcal{N}(\mathbf{a}_0, \sigma^2\Sigma_0),\ \sigma^2 \sim \mathcal{IG}(\frac{\alpha}{2}, \frac{\beta}{2})$$

where $\mathcal{N}$ (resp. $\mathcal{IG}$) denotes the normal (resp. inverse-Gamma) distribution. The symbols are assumed statistically independent. Formally each symbol follows a Bernoulli distribution

$$X \sim \mathcal{B}e(\mathbf{P}_0)$$

where

$$\begin{cases} \Pr(x_{k,t} = 1) = \mathbf{P}_{0,(k,t)}, \\ \Pr(x_{k,t} = -1) = 1 - \mathbf{P}_{0,(k,t)}. \end{cases}$$

We will consider here the case where $\mathbf{P}_{0,(k,t)}$=0.5 for any k, t. It follows that the introduction of the notation $\mathcal{B}e(\mathbf{P}_0)$ may appear tedious but it will prove useful when we discuss the CE algorithm in Section 3. The marginal likelihood for the observation Y given the input block of data X is given as

$$\begin{aligned} p(Y|\,X) &= \int p\left(Y|\,X, \mathbf{a}, \sigma^2\right) p\left(\mathbf{a}|\,\sigma^2\right) p(\sigma^2)\ d\mathbf{a}\ d\sigma^2 \\ &\propto |\Sigma|^{1/2} (V_M)^{-\frac{T+\alpha}{2}} \end{aligned} \quad (1)$$

where

$$\Sigma^{-1} = \sum_{t=0}^{T-1} B_{(t)} R B_{(t)} + \Sigma_0^{-1},$$

$$\mathbf{m} = \Sigma\ \left(\sum_{t=0}^{T-1} B_{(t)} S^T \mathbf{y}_{(t)} + \Sigma_0^{-1}\mathbf{a_0}\right),$$

$$V_M = \beta + \mathbf{a}_0^{\mathrm{T}}\Sigma_0^{-1}\mathbf{a_0} + \sum_{t=0}^{T-1} \mathbf{y}_{(t)}^{\mathrm{T}}\mathbf{y}_{(t)} - \mathbf{m}^{\mathrm{T}}\Sigma^{-1}\mathbf{m},$$

and $B(t)$, S, R are given as

$$B(t) = \mathrm{diag}(x_{1,t}, ..., x_{K,t}),$$

$$S = [\mathbf{s}_1, ..., \mathbf{s}_K],\ R = S^T S.$$

We are interested in maximizing the posterior distribution

$$p\left(X|\,Y\right) \propto p\left(Y|\,X\right) \quad (2)$$

as all sequences are a priori equiprobable. Clearly, there are 2^{KT} potential combinations for the sequence of symbols X so maximizing (1) cannot be performed exactly for realistic scenarios.

2.2 Markov Chain Monte Carlo

Recently MCMC have been introduced in wireless communications [1], [2], [10], [11]. These methods allow to sample from complex posterior distributions only known up to a normalizing constant. In the CDMA context presented here, we are interested in sampling from the posterior distribution (2). We turn to the Gibbs sampler, which iteratively samples from each of the full conditionals of a joint distribution, to explore the posterior. The draws from successive iterations are correlated. However the algorithm generates asymptotically samples whose distribution is the posterior distribution of interest. Thus with high probability the MAP sequence will be drawn. Formally, on each iteration n, we need to sample from

$$x_{k,t}^{(n)} \sim p(\cdot|\, X_{-k,t}^{(n)}, Y) \propto p(Y|\, X^{(n)})\; p(x_{k,t}^{(n)})$$

where $X_{-k,t}^{(n)} = \{x_{1,0}^{(n)}, \ldots, x_{1,T-1}^{(n)}, \ldots, x_{k-1,0}^{(n)}, \ldots, x_{k-1,T-1}^{(n)}, x_{k,0}^{(n)} ..., x_{k,t-1}^{(n)}, x_{k,t+1}^{(n-1)}, ..., x_{k,T-1}^{(n-1)}, x_{k+1,0}^{(n-1)}, \ldots\}$.

Note that if one wants to obtain an algorithm converging to a given sequence, then it is sufficient to replace the previous simulation step by

$$x_{k,t}^{(n)} \sim \overline{p}^{\gamma_n}(\cdot|\, X_{-k,t}^{(n)}, Y)$$

where

$$\overline{p}^{\gamma_n}(x|\, X_{-k,t}^{(n)}, Y) \propto p^{\gamma_n}(x|\, X_{-k,t}^{(n)}, Y)$$

and $\{\gamma_n\}$ is a positive increasing sequence such that $\gamma_n \to \infty$. In this case, one obtains a special case of the simulated annealing algorithm.

3 Cross-Entropy Method

3.1 Generic Cross-Entropy Method

Let χ be a finite set of states and S be the *Score Function* on χ. We are interested in finding an element, say x^*, of the set

$$\arg\max_{x \in \chi} \; S(x) \tag{3}$$

and we write $\gamma^* = S(x^*)$. CE is an adaptive important sampling which transforms the deterministic optimization problem (3) into a family of stochastic sampling problems. At each iteration j of the algorithm, a collection of random samples $\left(X_1^{(j)}, \ldots, X_N^{(j)}\right)$ is obtained from an importance distribution $f(\cdot; v^{(j-1)})$. Then a new value $v^{(j)}$ is obtained, this parameter is selected so as to minimize the approximate *cross entropy* between $f(\cdot; v)$ and the optimal importance distribution. The general adaptive CE algorithm is described in [8] and [4]. Assume a family of probability distributions $\{f(\cdot; v)\}$ has been selected and let ρ be a parameter that controls the number of samples drawn in each iteration.

The algorithm proceeds as follows.

Algorithm 1 [General Cross-entropy Algorithm]

1. Start with an initial value $v^{(0)}$. Set the iteration counter $j = 1$.

2. Generate for $n = 1, \ldots, N$ $X_n^{(j)} \sim f(\cdot; v^{(j-1)})$.

3. Calculate the score functions $\{S(X_n^{(j)})\}_{n=1}^N$ and order them from largest to smallest, $S_{(1)} \geq ... \geq S_{(N)}$. Set $\gamma^{(j)} = S_{([\rho N])}$ where $[\rho N]$ be the integer part of ρN.

4. Calculate the updated parameter

$$v^{(j)} = \arg\max_v \sum_{n=1}^{N} 1_{\left[\gamma^{(j)}, \infty\right)} \left(S\left(X_n^{(j)}\right)\right) \log f(X_n^{(j)}; v)$$

5. If convergence not reached then set $j \leftarrow j+1$ and go back to step 2.

In the algorithm, $1_A(x)$ is defined as

$$1_A(x) = \begin{cases} 1 \text{ if } x \in A, \\ 0 \text{ otherwise} \end{cases}$$

Step 3 can usually be solved analytically. From the above algorithm, we see that CE method updates the parameter v and the threshold γ simultaneously until convergence is reached which can be shown to happen in a finite number of iterations.

3.2 Cross-Entropy for Multiuser Detection

In the multiuser detection case, we are interested in maximizing the log-posterior (2) over the set of potential sequences so

$$S(X) = \log p\left(X \mid Y\right).$$

To apply the CE algorithm, we need to define a family of importance distributions. In this case, one can clearly adopted $\mathcal{B}e(\mathbf{P})$; i.e. a product of independent Bernoulli distributions such that

$$\Pr\left(x_{k,t}=1\right) = 1 - \Pr\left(x_{k,t}=-1\right) = \mathbf{P}_{(k,t)}.$$

Combined with the algorithm given in Section.1, the cross entropy algorithm for the multiuser detection proceeds as follows

Algorithm 2 [Cross-entropy Multiuser Detector]

1. Start with initial $\mathbf{P}^{(0)} = \mathbf{P}_0$. Set the iteration counter $j = 1$.

2. Generate N symbols sequences $X_n^{(j)} \sim \mathcal{B}e(\mathbf{P}^{(j-1)})$ $(n = 1, \ldots, N)$. $x_{n,k,t}^{(j)}$ denotes the (k,t) element of the sample n at iteration j.

3. Calculate the score functions $\{S(X_n^{(j)})\}_{n=1}^{N}$ and order them from largest to smallest, $S_{(1)} \geq ... \geq S_{(N)}$. Set $\gamma^{(j)} = S_{([\rho N])}$ where $[\rho N]$ be the integer part of ρN.

4. Calculate the updated parameters

$$\mathbf{P}_{k,t}^{(j)} = \frac{\sum_{n=1}^{N} 1_{\left[\gamma^{(j)},\infty\right)}\left(S(X_n^{(j)})\right) 1_{\{1\}}\left(x_{n,k,t}^{(j)}\right)}{\sum_{n=1}^{N} 1_{\left[\gamma^{(j)},\infty\right)}\left(S(X_n^{(j)})\right)}.$$

5. If convergence not reached then set $j \leftarrow j+1$ and go back to step 2.

This algorithm is a special instance of the generic CE procedure described in the previous subsection. We have explicated the parameter updating step.

According to [4], to prevent the situation that 0's or 1's occur in the parameter matrix $\mathbf{P}$ (which will usually remain so forever), we introduce a smoothing factor λ and change the updating equation to

$$\mathbf{P}^{(j)} \leftarrow \lambda \mathbf{P}^{(j)} + (1-\lambda)\,\mathbf{P}^{(j-1)}.$$

When $\lambda = 1$, one obviously recovers the previous formulation.

The main advantage of the CE entropy over standard MCMC methods such as the Gibbs sampler is that it does not based on a Markov chain approach and as such does not rely on iterative local perturbations of the current state. This makes this procedure much less sensitive to initialization.

4 Simulations

In the implementation of the CE and Gibbs sampling multiuser detectors, we consider a five-user $(K = 5)$ synchronous CDMA channel with processing gain $P = 10$. The channel coefficient is $\mathbf{a} = [\ 5.0176,\ 0.1073,\ -0.5685,\ 2.9487,\ -3.7755\ ]^T$. Other parameters are $\mathbf{a}_0 = 0_{K\times 1}$, $\Sigma_0 = 1000I$, $\alpha = 2$, $\beta = 0.2$ and $\mathbf{P} = \frac{1}{2} I_{K\times T}$. The data size of each user is $T = 50$. For each data block, the Gibbs sampling is performed for 1000 iterations, with the first 500 iterations as the 'burn-in' period. For cross-entropy, the sample block is $N = 500$, $\rho = 0.1$, the smoothing factor $\lambda = 0.8$ and the iteration number is $J = 50$. DPSK is employed as necessary to prevent the phase ambiguity in both algorithms. The computer used for simulations is a Pentium IV 1.8G.

The performance of cross-entropy is shown in Figure 1. CE (blind) performs better than the linear MMSE with known channel coefficients. Figure 2 demonstrates that CE is a robust algorithm with the number of samples $N \geq 500$ (SNR=20dB). From Figure 3, we find that adding a smoothing factor will improve the BER performance of cross-entropy. Figure 4 illustrates that being both optimization methods, cross-entropy outperforms Gibbs sampler under certain conditions (the iterations used in calculating Monte Carlo estimates are 500). With the performance shown in Figure 4, the operation time for Gibbs sampler (totally 1000 iterations) is around 2.6505e+003 seconds while it only takes cross-entropy

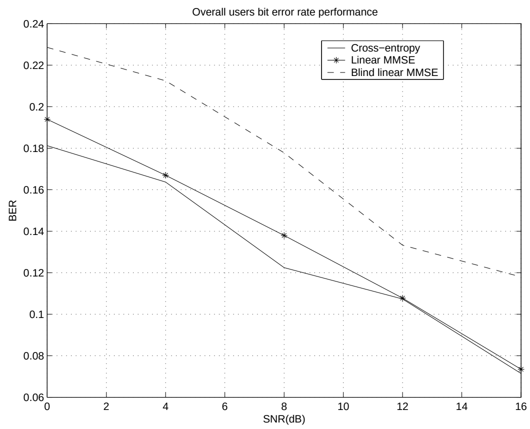

Figure 1: The BER Performance of Cross-entropy

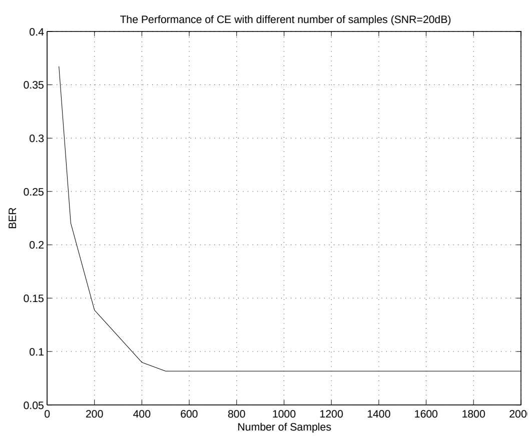

Figure 2: The Performance of CE with Different Number of Samples

336.6800 seconds to converge. As we increase the iteration number in Gibbs sampler (500 burn-ins and 1000 iterations for calculating the Monte Carlo estimate), it achieves a similar BER performance with CE, while its operation time is 3.9456e+003 seconds which is about 11.7 times of the operation time of cross-entropy method (the performance of Gibbs sampler does not make any notable improvement with further increase of iteration number).

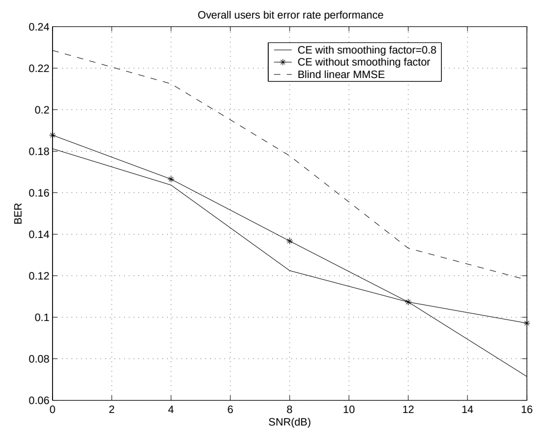

Figure 3: The Effect of Adding Smoothing Factor

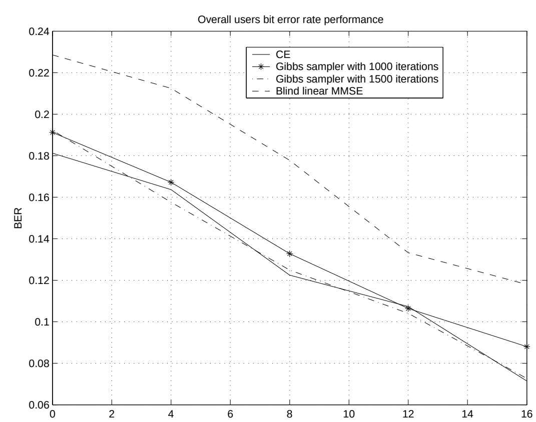

Figure 4: The BER Performance Comparison of CE and Gibbs Sampler

5 Conclusions and Future Work

We have presented an overview of the cross-entropy method and its application in multiuser detection. The cross-entropy method and Gibbs sampler are both powerful approaches to overcome the computational complexity problem encountered by optimal solutions. Compared with some traditional detection methods, i.e., linear MMSE, they both achieve a much better performance. With similar BER performance, CE outperforms the Gibbs sampler in terms of operation time.

Considering its ideal BER performance and relatively fast operation, the cross-entropy method shows a great potential in solving a variety of *optimal* signal processing problems encountered in wireless communications, such as mitigation of the interchannel interference in frequency-selective channels, user detection in MIMO etc. In the near future, we will implement CE in a more realistic model with error correction coding. Furthermore, we can try combining CE with sequential Monte Carlo methods [3] to obtain better proposals $f(\cdot; v)$.

References

[1] O. Cappé, A. Doucet, and E. Moulines, "Simulation-Based Methods for Blind Maximum-Likelihood Filter Identification", *Signal Processing*, vol. 73, pp. 3-25, 1999.

[2] R. Chen, J.S. Liu, and X. Wang, "Convergence Analyses and Comparisons of Markov Chain Monte Carlo Algorithms in Digital Communications," *IEEE Transactions on Signal Processing*, vol. 50, pp.255-270, 2002.

[3] A. Doucet, N. de Freitas, and N.J. Gordon, *Sequential Monte Carlo Methods in Practice,* New York: Springer-Verlag, 2001.

[4] P.T. De Boer, D.P. Kroese, S. Mannor, R.Y. Rubinstein,"A Tutorial on the Cross-entropy Method," http://www.cs.utwente.nl/ptdeboer/ce/, 2002.

[5] P.T. De Boer, D.P. Kroese, S. Mannor, R.Y. Rubinstein (editor), Special issue on Cross-entropy Algorithms for Combinatorial Optimisation, *Annals of Operations Research*, to appear.

[6] W.R. Gilks, S. Richardson, and D.J. Spiegelhalter (Eds.), *Markov Chain Monte Carlo in Practice*, London: Chapman

[7] C.P. Robert and G. Casella, *Monte Carlo Statistical Methods*, Springer-Verlag: New York, 1999.

[8] R.Y. Rubinstein, "The Cross-entropy Method for Combinatorial and Continuous Optimization," *Methodology and Computing in Applied Probability*, vol. 1, pp. 127-190, 1999.

[9] J.G. Proakis, *Digital Communications*, 3rd ed., New York: McGraw-Hill, 1995.

[10] X. Wang, R. Chen, J.S. Liu,"Monte Carlo Bayesian Signal Processing for Wireless Communications," *Journal of VLSI Signal Processing*, vol. 30, pp. 89-105, 2002.

[11] Z. Yang, B. Lu, and X. Wang, "Bayesian Monte Carlo Multiuser Receiver for Space-time Coded Mutli-Carrier CDMA Systems," *IEEE Journal of Selected Areas in Communications*, vol. 19, pp.1625-1637, 2001.