Bayesian Blind PARAFAC Receivers for DS-CDMA Systems.

Alexandre de Baynast*, Behnaam Aazhang
Rice University, MS-366
Department of
Electrical and Computer Eng.
6100 Main Street - Houston, Texas 77005

David Declercq, Lieven De Lathauwer†
Université de Cergy-Pontoise - ENSEA
Equipe Traitement des Images et du Signal
6 avenue du Ponceau
Cergy-Pontoise, France, 95014 Cedex

ABSTRACT
In this paper an original Bayesian approach for blind detection for Code Division Multiple Access (CDMA) Systems in presence of spatial diversity at the receiver is developed. In the noiseless context, the blind detection/identification problem relies on the canonical decomposition (also referred as Parallel Factor analysis [Sidiropoulos, IEEE SP’00], PARAFAC). The author in [Bro,INCINC’96] proposes a suboptimal solution in least-squares sense. However, poor performance are obtained in presence of high noise level. The recently emerged Markov chain Monte Carlo (MCMC) signal processing method provide a novel paradigm for tackling this problem. Simulation results are presented to demonstrate the effectiveness of this method.

1. INTRODUCTION
Let us start from the following memoryless baseband data model for $R$-user DS-CDMA depicted on Fig. 1:

$$y_{ijk} = \sum_{r=1}^{R} a_{ir} c_{jr} s_{kr} + n_{ijk} = x_{ijk} + n_{ijk}.$$  

Eq. (1) is the output of the $i$th antenna for chip $j$ and symbol $k$ ($1 \leq i \leq I$, $1 \leq j \leq J$, $1 \leq k \leq K$, with $I$ the number of antennas, $J$ the code length and $K$ the number of transmitted symbols). $a_{ir}$ is the fading / gain between user $r$ and antenna element $i$. $c_{jr}$ is the $j$th chip of the spreading sequence of user $r$ and $s_{kr}$ is the $k$th symbol transmitted by user $r$. If no power control is used in the transmission scheme, we can keep the same model provided we replace $c_{jr}$ by $h_{jr}$:

$$y_{ijk} = \sum_{r=1}^{R} a_{ir} h_{jr} s_{kr} + n_{ijk}.$$  

$h_{jr}$, for varying $j$ and fixed $r$, is the result of the product between the $j$th chip of the spreading sequence of user $r$ and the channel fading. As shown on Fig. 2, Eq. (2) can be written in a 3-way data array format as:

$$\mathcal{Y} = \sum_{r=1}^{R} \mathbf{a}_r \circ \mathbf{h}_r \circ \mathbf{s}_r + \mathcal{N},$$  

Eq. (3) has a number of inherent indeterminacies. First, the canonical decomposition (3) is unique in the noiseless case, apart for a set of Lebesgue measure 0.)

In the generic case, it was shown in [2, 1] that the canonical decomposition (3) is unique in the noiseless case, apart from the inherent trivial determinacies mentioned in the previous paragraph if the number of sources $R$ verifies:

$$R \leq \frac{1}{2} (\min(R, I) + \min(R, J) + \min(R, K)) - 1.$$  

(We call a property “generic” when it holds everywhere, except for a set of Lebesgue measure 0.)

Note that the number of users can be higher than the length of the spreading sequence and the spreading codes need not be orthogonal to guarantee the uniqueness of the canonical decomposition.
Fig. 1. Discrete-time baseband-equivalent scheme of CDMA transmission.

Fig. 2. Parallel Factors Model (PARAFAC) or canonical decomposition of the 3-way data array in rank-1 terms in the noiseless case.

2. DETECTION/IDENTIFICATION COMPUTATION

2.1. Trilinear Alternating Least Squares Regression

The joint estimation of the three sets (symbols, antenna gain and channel fading) can be computed by means of an Alternating Least Squares (ALS) algorithm [1]. Let consider Eq. (3). Compact vectorial representations of the model given by Eq. (3) are made possible by employing the Khatri-Rao matrix product (column-wise Kronecker product). It leads to the following equation sets:

\[
\begin{align*}
\mathbf{y}_i &= (H \odot S) \mathbf{a}_i + \mathbf{n}_i, \quad i = 1, \ldots, I, \\
\mathbf{y}_j &= (S \odot A) \mathbf{h}_j + \mathbf{n}_j, \quad j = 1, \ldots, J, \\
\mathbf{y}_k &= (A \odot H) \mathbf{s}_k + \mathbf{n}_k, \quad k = 1, \ldots, K,
\end{align*}
\]

where \( \mathbf{y}_i \in \mathbb{C}^{JK} \) results of the concatenation of the columns of the \( i \)th “slab” of the 3-way data array \( \mathcal{Y} \) following the first dimension and \( \mathbf{a}_i = [a_{i1}, a_{i2}, \ldots, a_{iR}]^T \). Vectors \( \mathbf{y}_j \in \mathbb{C}^{JK} \), \( \mathbf{x}_k \in \mathbb{C}^{JK} \) and \( \mathbf{h}_j = [h_{j1}, h_{j2}, \ldots, h_{jR}]^T \), \( \mathbf{s}_k = [s_{k1}, s_{k2}, \ldots, s_{kR}]^T \) are defined in the same manner. Vectors \( \mathbf{n}_i, \mathbf{n}_j \) and \( \mathbf{n}_k \) are defined accordingly.

The basic idea behind ALS is simple: each time update one matrix, using least squares conditioned on previously obtained estimates for the remaining matrices; proceed to update the other matrices; repeat until convergence of the least squares cost function. More precisely, ALS aims to minimize the cost function \( \mathcal{F} \) in a noisy context:

\[
\mathcal{F}_{A, H, S} = \left\| \sum_i \mathbf{y}_i - (H \odot S) \mathbf{a}_i \right\|^2
\]

Due to the complete symmetry of the trilinear model, the conditional least squares updates are:

\[
\begin{align*}
\mathbf{a}_i^{(l+1)} &= \left[ \hat{H}^{(l-1)} \odot \hat{S}^{(l-1)} \right]^\dagger \mathbf{y}_i, \quad i = 1, \ldots, I, \\
\mathbf{h}_j^{(l+1)} &= \left[ \hat{S}^{(l-1)} \odot \hat{A}^{(l)} \right]^\dagger \mathbf{y}_j, \quad j = 1, \ldots, J, \\
\mathbf{s}_k^{(l+1)} &= \left[ \hat{A}^{(l)} \odot \hat{H}^{(l)} \right]^\dagger \mathbf{y}_k, \quad k = 1, \ldots, K.
\end{align*}
\]

where \((\cdot)^\dagger\) denotes the pseudo-inverse.

The conditional update of any given matrix may either improve or maintain but cannot worsen the current fit. Global monotone convergence to at least a local minimum follows directly from this observation. For a high level noise, the performance suffer from spurious minima. To increase the chance of finding the global optimum, one may rerun the algorithm starting from different initial values if
convergence towards a spurious minimum occurs. In our case, we assume the sources are QPSK, i.e. \( s_{kr} \in \{ \pm 1 \pm i \} \) where \( i = \sqrt{-1} \). Spurious minima may be detected by considering the standard deviation between the estimated symbols \( \hat{s}_{kr} \) and its nearest QPSK symbol states given by \( \arg \min_{Z \in \{ \pm 1 \pm i \}} |\hat{s}_{kr} - Z| \). If this deviation is larger than a given threshold, then we assume that the algorithm converged towards a spurious minimum. For a high level noise, the number of reinitializations could become large and significantly increase the computational complexity.

Furthermore, the ALS algorithm does not take into account any statistical property of the sources. In particular, modulations with finite alphabet are used in most wireless communication systems. It results that the performance could be far from the ultimate optimal solution predicted by statistical theory.

In the sequel, we propose a solution based on the Bayesian estimation theory close to the maximum a posteriori solution as shown with the simulations.

### 2.2. Bayesian Blind Estimation/Detection

Markov Chain Monte Carlo is a class of algorithms that allow one to draw pseudo-random samples from a target probability distribution, the maximum a posteriori of all parameters given the observation samples in our case. The basic idea behind these algorithms is that one can achieve the sampling from the target distribution by running a Markov chain whose equilibrium distribution is exactly the target distribution. In this paper, we use the Gibbs sampler. The validity of this algorithm can be proved by the basic Markov Chain theory; for more details, see [3, 4]. The key idea behind this algorithm is simple: Suppose \( x = (x_1 \ldots x_n) \), where \( x_i \) is either a scalar or a vector. In the Gibbs sampler, one systematically choose a coordinate, say \( x_i \), and then update its value with a new sample \( x_i \) drawn from the conditional distribution \( p(x_i | x_{i-1},x_{i+1},\ldots,x_n) \).

Consider the signal model described by Eq. (3). The unknown quantities \( A, H, S \) and \( \sigma^2 \) are regarded as realizations of random variables with the following prior distributions. For the unknown antenna gains \( A \) and the spreading sequences \( H \), Gaussian prior distribution are assumed,

\[
p(a_i) \sim \mathcal{N}(0, I), \quad i = 1, \ldots, I, \quad (9)
p(h_j) \sim \mathcal{N}(0, I), \quad j = 1, \ldots, J \text{ respectively.} \quad (10)
\]

For the noise variance, we adopt an uninformative prior distribution by selecting Jeffrey’s prior distribution for \( \sigma^2 \), i.e. \( p(\sigma^2) \propto \sigma^{-2}, ([5]) \).

Finally since the symbols \( \{s_{kr}\}_{r,k} \) are assumed to be independent and identically distributed, the prior distribution \( p(S) \) can be expressed in terms of the prior symbol probabilities as,

\[
p(S) = \prod_{k,r} \frac{1}{4} (\delta_{s_{kr},-1-i} + \delta_{s_{kr},1+i})
\]

where \( \delta_{.,.} \) is the Kronecker symbol defined as \( \delta_{x,x'} = 1 \) if \( x = x' \) and 0 otherwise.

The blind Bayesian multiuser detector based on the Gibbs sampler is summarized as follows.

**Algorithm:** Given the initial values of the unknown quantities \( \{ H^{(0)}, S^{(0)}, \sigma^2(0) \} \) drawn from their prior distributions,

1. Sample \( a_{i}^{(l)}, \quad i = 1, \ldots, I \) from:

\[
p(a_{i}^{(l)} | H, S, \sigma^2) \sim \mathcal{N} ( \left( H \otimes S \right)^{t} \sum_{i} \left( H \otimes S \right)^{t} H \otimes S)^{t} \).
\]

2. Sample \( h_{j}^{(l)}, \quad j = 1, \ldots, J \) from:

\[
p(h_{j}^{(l)} | A, S, \sigma^2) \sim \mathcal{N} \left( (S \otimes A)^{t} \sum_{j} (S \otimes A)^{t} S \otimes A)^{t} \right).
\]

3. Sample \( s_{kr}^{(l)}, \quad r = 1, \ldots, R, \quad k = 1, \ldots, K \) from:

\[
p(s_{kr}^{(l)} = Z | s_{k1}, \ldots, s_{k-1,r}, s_{k+1,r}, \ldots, s_{KR}, A, H, \sigma^2) \sim \exp\left( -\frac{1}{2} \right),
\]

\[
\frac{1}{2} \left[ \sum_{i=1}^{I} \left| Y_{i} - \left( H \otimes S \right) a_{i} \right|^{2} \right].
\]

where \( Z \in \{ \pm 1 \pm i \} \) represents each of the four QPSK symbol states and \( s' = [\ldots s_{k,r-1} Z s_{k,r+1} \ldots]^{T} \).

4. Sample \( \sigma^{2(l)} \) from:

\[
p(\sigma^{2(l)} | S, A, H) \sim \mathcal{IG} \left( \frac{\alpha H + \beta}{2}, \frac{\sum_{i=1}^{I} \left| Y_{i} - \left( H \otimes S \right) a_{i} \right|^{2}}{2} \right)
\]

where the Inverse Gamma probability distribution is defined by \( \mathcal{IG}(\alpha, \beta) = \frac{\beta^{\alpha}}{\Gamma(\alpha)} x^{-(\alpha+1)} \exp(-\frac{\beta}{x} \), \( \Gamma(.) \) being the Gamma function ([4]).

To ensure convergence, the above procedure is usually carried out for \( N_0 + N \) iterations and samples from the last \( N \) iterations are used to calculate the Bayesian estimates of the unknown quantities. In particular, the a posteriori symbol probabilities are approximated as

\[
p(s_{kr} = Z | Y) \approx \frac{1}{N} \sum_{l=N_0+1}^{N} \delta_{s_{kr},Z}, \quad Z \in \{ \pm 1 \pm i \}.
\]
3. SIMULATION RESULTS

We now illustrate the performance of the blind Bayesian PARAFAC receiver. We consider a 3-user CDMA system with processing gain $S_F = 3$ (the load is equal to 1) and 2 receive antennas. The block size of the symbols is 10. The condition given by Eq. (4) is verified. The emitted signals are of the QPSK-type, taking values in $\{-1, 1\}$. The noise is zero-mean white (in all dimensions) Gaussian, with variance $\sigma^2$ for all antennas. The Signal-to-Noise Ratio (SNR) at the input of the multiuser receiver is defined by

$$\text{SNR} = 10 \log_{10}(2J/\sigma^2) [\text{dB}].$$

The channel impulse responses are not normalized to unit norm (no power control is assumed); this, along with the presence of fading, means that the effective signal power varies considerably from run to run and from user to user, which is a challenging setup. In computing the symbol probabilities, the Gibbs sampler is iterated 500 runs for each data block. The proposed method is tested across a range of SNR points and at each SNR point the experiment is repeated for 500 Monte-Carlo realizations. The results are averaged over all users and all runs.

We compare against the Maximum Likelihood Sequence Estimation (MLSE) which corresponds to the lowest error that can be obtained by an unbiased receiver ([6]). In contrast to our algorithm, the MLSE receiver assumes perfect knowledge of channel fading coefficients, spreading codes, antenna gains and noise variance.

Fig. 3 displays the average symbol error rate performance of the blind Bayesian PARAFAC receiver. The results of the blind Bayesian PARAFAC receiver are close to the MLSE results (less than 1 [dB] for all cases) and outperform the ALS results. Moreover, ALS frequently converges to spurious minima and needs to be rerun after reinitialization increasing considerably the computational complexity. Table 1 shows the percentage of the simulations for which at least one convergence towards a spurious minimum is detected according to the approach proposed in Section 2.1.

<table>
<thead>
<tr>
<th>SNR [dB]</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ill-convergence [%]</td>
<td>32</td>
<td>20</td>
<td>19</td>
<td>16</td>
<td>14</td>
<td>12</td>
</tr>
</tbody>
</table>

Table 1. Percentage of the simulations for which at least one convergence towards a spurious minimum occurs, requiring reinitialization and rerun of the ALS algorithm.

4. CONCLUSION

In this paper, we have proposed a new Bayesian blind detection/identification algorithm for CDMA system in presence of spatial diversity at the receiver. Such algorithm provides good performance even for short frame length (typically, 10 symbol periods) and for large system load. Furthermore, simulation results show that the proposed method is close to the MLSE algorithm and outperforms the ALS algorithm.

5. REFERENCES


