Optimal Weighting of Soft-Information in a SAGE-based Iterative Receiver for Coded CDMA

Alexander Kocian, Ingmar Land, and Bernard H. Fleury
Department of Communication Technology, Aalborg University, DK-9220 Aalborg, Denmark
Email: {ak,il,bl}@kom.aau.dk

Abstract—An iterative receiver for joint multiuser-decoding and channel-estimation of coded CDMA is derived by applying the noise-splitting approach within the Space Alternating Generalized Expectation-Maximization (SAGE) framework. We consider asynchronous single-rate DS/CDMA over flat Rayleigh fading channels. The resulting receiver structure comprises partial Successive Interference-Cancellation (SIC), channel estimation, and soft-input/hard-output Maximum Likelihood Sequence Decoding (MLSD) for each user. Additionally, one obtains a set of noise-weighting coefficients that can be freely chosen within weak constraints. These coefficients determine the amount of feedback from the decoder output to the decoder input in the subsequent iteration, and thus, “how extrinsic” the decoder output values are. The noise-weighting coefficients strongly influence the system performance. Their optimization within the SAGE framework leads to extrinsic output values in most of the cases. The proposed receiver is evaluated by Monte-Carlo simulations, and it shows two major advantages: high load is supported and convergence is guaranteed.

I. INTRODUCTION

Code-Division Multiple Access (CDMA) is used in many modern communication systems. To support a large number of users and to be resistant against the near-far problem, the receiver should employ multiuser (MU) detection in the case of uncoded CDMA or MU-decoding in the case of coded CDMA. Since the optimal MU-decoder [1] is infeasible, efficient suboptimal algorithms are required [2].

Initially applied to MU-detection only [3], the Expectation-Maximization (EM) algorithm and the Space-Alternating Generalized Expectation-maximization (SAGE) algorithm [4] have been extended to joint MU-detection and channel-estimation of uncoded DS/CDMA in [5], [6].

In the present paper, SAGE-based Joint MU-Decoding and Channel-Estimation (SAGE-JDCE) is further investigated. Resulting from the noise-splitting, the input sequence of each user’s decoder is a weighted superposition of a priori information from the previous iteration and extrinsic (EXT) information from the interference cancelation (IC) device when the channel is known to the receiver, and they are asymptotically a priori and EXT information when the channel is unknown but the observation length is large. Accordingly, the output values of the decoder may have asymptotically a priori or EXT character, depending on the values of the weighting coefficients. These coefficients strongly influence the system performance, and they may be optimized for each individual symbol. The optimal weighting coefficients are shown to lead to asymptotically EXT decoder outputs in almost all cases - even though finite code-lengths are considered.

The approaches in [5], [6] are extended in the present paper with respect to several aspects: A coded CDMA system is considered, and thus the receiver performs MU-decoding. The receiver is based on the SAGE algorithm, and so the analysis and optimization of the weighting coefficients greatly differs from that in [6] for the EM algorithm. Furthermore, the optimal weighting coefficients are in agreement with the SAGE framework and thus convergence of the algorithm is guaranteed; this is not the case for the best weighting coefficients found for the EM algorithm in [6].

The paper is organized as follows: Section II describes the system model and introduces the notation. In Section III, the SAGE-JDCE receiver is derived; in particular, the weighting coefficients resulting from the noise-splitting are analyzed and optimized. The performance is evaluated by means of Monte-Carlo simulations in Section IV.

II. SYSTEM DESCRIPTION

We consider an asynchronous single-rate convolutional-coded CDMA system over frequency-nonsselective channels with K active users.

Each user transmits \( \tilde{L}_b \) information bits per frame. The information bits of user k are written as the column vector \( b_k = \text{col}\{b_{k0}, \ldots, b_{kL_b - 1}\} \). For convenience, we write the information bits of all users as \( b = \text{col}\{b_1, \ldots, b_K\} \), and those of all users excluding user k as \( b_k = \text{col}\{b_1, \ldots, b_{k-1}, b_{k+1}, \ldots, b_K\} \).

For user k, the information bit sequence \( b_k \) is encoded by a user-independent encoder of rate R into the data sequence \( d_k \) and interleaved by an user-individual interleaver \( \Pi_k \). The overall frame comprises \( N \in \mathbb{Z}^+ \) blocks, and each block consists of \( L_d \) data symbols multiplexed with \( L_p \in \mathbb{Z}^+ \) random preamble symbols. Hence, the block length and the frame length are \( L = L_d + L_p \) and \( \hat{L} = LN \) code symbols (data symbols or preamble symbols), respectively. The average number of information bits per code symbol is thus \( \hat{L}_b / L = R(L - L_p)/L \).

One code symbol is transmitted in each signalling interval. During the ith signalling interval, user k is assigned a real-valued signature waveform \( s_k(t) \) of duration \( T_s \), spreading length \( N_c \), and chip duration \( T_c = T_s / N_c \). The waveforms \( s_1(t), \ldots, s_K(t) \) are normalized to have unit energy, i.e., \( \int_0^{T_k} |s_k(t)|^2 \, dt = 1, \quad k = 1, \ldots, K \).
The waveforms are transmitted over Rayleigh fading channels that are assumed to be quasi-static for each block. The relative transmission delay of user \( k \) is denoted by \( \tau_k \); without loss of generality, we assume \( 0 \leq \tau_1 \leq \ldots \leq \tau_K \leq \tau_0 \). The channel coefficient of user \( k \) during its \( l \)-th signalling interval is denoted by \( a_k[l] \), describing a complex circular-symmetric Gaussian random variable with zero variance \( \sigma_k^2 \). The resulting received complex baseband signal reads

\[
\begin{align*}
    r(t) &= \sum_{k=1}^{K} \sum_{l=0}^{\widetilde{L}-1} a_k[l] \, d_k[l] \, s_k(t - lT_s - \tau_k) + w(t),
\end{align*}
\]

where \( w(t) \) represents Additive complex-valued White Gaussian Noise (AWGN) with spectral height \( N_0 \).

At the receiver site, for each user \( k \), the output of a filter matched to the signature waveform \( s_k(t) \) is sampled at time instants \((l + 1)T_s + \tau_k \) with \( k = 1, \ldots, K \), and \( l = 0, \ldots, \widetilde{L} - 1 \). Let the column vector \( z[l] \triangleq \text{col}(z_1[l], \ldots, z_K[l]) \in \mathbb{C}^K \) contain the output samples of the \( K \) Matched Filters (MF) in bin \( l \). From (1), \( z[l] \) may then be written as

\[
\begin{align*}
\end{align*}
\]

(2) where \((\cdot)^T\) denotes transposition. The entries of the matrix \( R[l, 0] \in \mathbb{R}^{K \times K} \) are the zero-offset cross-correlations between the signature waveforms \( s_j(t) \) and \( s_k(t) \) in bin \( l \), i.e.,

\[
\begin{align*}
    [R[l, 0]]_{jk} &= \begin{cases} 
        \rho_{jk}[l, 0] & \text{for } j < k \\
        \rho_{kj}[l, 0] & \text{for } j \geq k
    \end{cases},
\end{align*}
\]

and for \( j < k \). The entries of the upper-triangular matrix \( R[l, 1] \in \mathbb{R}^{K \times K} \) are the one-offset cross-correlations between the signature waveforms \( s_j(t) \) and \( s_k(t) \) in bin \( l \), i.e.,

\[
\begin{align*}
    [R[l, 1]]_{jk} &= \begin{cases} 
        \rho_{jk}[l, 1] & \text{for } j < k \\
        0 & \text{for } j \geq k
    \end{cases},
\end{align*}
\]

for \( j < k \).

The symbol diagonal matrix \( D[l] \in \mathbb{R}^{K \times K} \) contains

\[
[D[l]]_{kk} = \begin{cases} 
    d_k[l] & \text{for } 0 \leq l \leq \widetilde{L} - 1 \\
    0 & \text{elsewhere}
\end{cases},
\]

as the diagonal elements. The channel column vector \( a[l] \in \mathbb{C}^K \) is defined by \( a[l] \triangleq \text{col}(a_1[l], \ldots, a_K[l]) \) with

\[
a_k[l] = \begin{cases} 
    a_k((l + 1)T_s + \tau_k) & \text{for } 0 \leq l \leq \widetilde{L} - 1 \\
    0 & \text{elsewhere}
\end{cases},
\]

With the assumptions made above the covariance matrix of \( a[l] \) is given by \( \Sigma_a[l] \triangleq \text{diag}(\sigma_1^2, \ldots, \sigma_K^2) \). Finally, the column vector \( n[l] \in \mathbb{C}^K \) contains circularly symmetric colored Gaussian noise \( \mathcal{C}_\mathcal{N} \).

To ease the derivation of the SAGE-JDCE scheme, \( z[l] \) is processed by a noise-whitening (anti-causal) filter with output-input relation \( z[l] = \sum_{\Delta l=0}^{L-1} F[l + \Delta l, \Delta l] g[l + \Delta l], y[l] \in \mathbb{C}^K, l = 0, \ldots, \widetilde{L} - 1 \) [7], where \( \Delta l \) denotes the time-offset. If the lower left triangular matrix \( F[l, 0] \in \mathbb{R}^{K \times K} \) and the upper right triangular matrix with zero diagonal \( F[l, 1] \in \mathbb{R}^{K \times K} \) are related to \( R[l, 0] \) and \( R[l, 1] \) according to

\[
\begin{align*}
    R[l, 0] &= F[l, 0]^T F[l, 0] + F[l + 1, 1]^T F[l + 1, 1] \\
    R[l, 1] &= F[l, 0]^T F[l, 1],
\end{align*}
\]

(3) and

\[
\begin{align*}
    y[l] &= \sum_{\Delta l=0}^{L-1} F[l, \Delta l] D[l - \Delta l] a[l - \Delta l] + w[l].
\end{align*}
\]

(5) The noise vector \( w[l] \) is white and has the covariance matrix \( N_0I \), where \( I \) denotes the \( K \times K \) identity matrix [7].

III. THE SAGE-JDCE RECEIVER

A. Receiver Design

To smartly apply the SAGE algorithm [4] to JDCE, we apply the noise splitting approach [8] to decompose the output signal vector of the whitened matched filter in (5) into the sum of two terms

\[
\begin{align*}
    y[l] &= x_k[l] + \bar{x}_k[l],
\end{align*}
\]

(6) The first term on the r.h.s of the equation represents the \( k \)-th user’s signal given by

\[
\begin{align*}
    x_k[l] &= \sum_{\Delta l=0}^{L-1} F_k[l, \Delta l] d_k[l - \Delta l] a_k[l - \Delta l] + w_k[l],
\end{align*}
\]

(7) \( k = 1 \ldots K \), with \( F_k[l, \Delta l] \) denoting the \( k \)-th column of \( F[l, \Delta l] \). The white Gaussian noise vector \( w_k[l] \) has variance \( \beta_k[l]N_0 \). The second term in (6) corresponds to the \( k \)-th user’s multiple access interference (MAI) with noise variance \( (1 - \beta_k[l])N_0 \). The real-valued adaptive weight coefficients \( \beta_k[l] \) satisfy

\[
0 \leq \beta_k[l] \leq 1, \quad l = 0, \ldots, \widetilde{L} - 1.
\]

(8) These weight coefficients are free parameters, and they will be used to optimize the performance of the SAGE-JDCE receiver in Section III-B. In order to get rid of the nuisance parameter \( a \), the admissible data for each \( b_k \) is selected as the decomposed signal along with knowledge about the channel i.e., \( \mathcal{X}_k = \{x_k, a\} \), where \( x_k \triangleq \text{col}(x_k[0], \ldots, x_k[\widetilde{L} - 1]) \).

In this case, the E-step of the SAGE algorithm yields

\[
Q\left(b_k|b^0\right) = E\left\{L(x_k[a]; b_k, b^0)\right\} | y, b^0\right\}.
\]

(9) Inserting (7) into (9), invoking the relations (3) and (4), and neglecting terms independent of \( b_k \), it follows after some straightforward algebraic manipulations
For each signalling interval $l$, a new fit of the $k^{th}$ user’s signal is computed using a fraction of its signal based on an estimate of MAI from the current iteration and a part of its tentative bit decision obtained from the previous iteration. The weight coefficients determine the balance. Looking at the same problem from a different point of view, it can be seen that the amount of noise split between user $k$ and its MAI $k$ is governed by $\beta_k[l]$ in an adaptive fashion. For $L \rightarrow \infty$, i.e. known channel, we observe the following: If the vector $\beta_k \triangleq \{\beta_k[0], \ldots, \beta_k[L-1]\} = 0$, the SAGE-JDCE scheme passes a priori values on the $k^{th}$ user’s code symbols from the channel estimator/IC device to the MLSD. In contrast if $\beta_k = 1$, EXT values on the $k^{th}$ user’s code symbols are forwarded. For unknown channel, however, the channel estimate in (11) is governed by all users’ symbol estimates. The influence of the $k^{th}$ user’s $l^{th}$ symbol, however, disappears as $L \rightarrow \infty$. Hence, we refer to these kinds of soft information as asymptotically a priori and EXT, respectively.

### B. Noise Splitting

The weight coefficient vector $\beta_k$ can be used to split noise between the $k^{th}$ user and its MAI such that the joint bit-error-probability is minimized as the number of iterations tend to infinity. This task, however, is difficult to meet due to the feedback and non-linear structure of the SAGE-JDCE scheme on the one hand, and the constraints on the code on the other hand. Hence, we rely on sub-optimal methods for optimizing the weight coefficients. In [6], the performance of an EM-based JDE scheme is enhanced by two optimization criteria. The first has been formulated within, the second outside the EM framework. In the following, the latter is adapted to the SAGE-JDCE scheme within the SAGE framework so that convergence is ensured. As a result we will get new insight into joint partial interference cancellation and data decoding.

We project different time-shifted versions of the estimation error $(a_k[l]^*x_k[l] - (a_k[l]^*x_k[l])^i)$ belonging to the $k^{th}$ user’s $l^{th}$ symbol on the corresponding versions of $F_{k}[:i]$, take the conditional expectation of the mean squared sum with respect to $b_k[i]$, and minimize the result according to

$$
\tilde{\beta}_k[l]^i = \arg\min_{\beta_k[l]^i} E \left( \sum_{\Delta l=0}^{1} F_{k}[l + \Delta l, \Delta l]^T (a_k[l] x_k[l] + \Delta l) - (a_k[l] x_k[l] + \Delta l)^i b_k[i] \right)^2.
$$

From (7), by invoking the relations (3) and (4), it follows for the projected true signal vector in (13)

$$
a_k[l]^* \sum_{\Delta l=0}^{1} F_{k}[l + \Delta l, \Delta l]^T x_k[l + \Delta l] = d_k[l] (a_k[l])^2 + a_k[l]^* \sum_{\Delta l=0}^{1} F_{k}[l + \Delta l, \Delta l]^T u_k[l + \Delta l].
$$
Analogously, the summands in (10) corresponds to its estimate. To make the derivation simpler, we only assume $L \rightarrow \infty$, i.e., the channel coefficients are assumed to be known to the receiver. Invoking the identity $\sqrt{\beta_k} n_k[l] = \sum_{a=0}^{N-1} F_k[l + \Delta l, \Delta l] w_k[l + \Delta l]$, making use of $E\{a_k[l]a_j[l]\} = 0$, $j \neq k$, and dropping terms independent of $\beta_k[l]$, the minimization problem in (13) is equivalent to

$$\beta_k[l]\hat{=} \arg \min_{\beta_k[l]} f(\beta_k[l]) \tag{14}$$

where

$$f(\beta_k[l]) = \left\{ \begin{array}{ll}
\beta_k[l] \sigma_k^2 & \text{if}\ n_k[l] n_k[l] (1 - \beta_k[l]) |b_k| \nonumber \\
-2\beta_k[l] \mathbb{E} \left\{ \left( I_k[l] - I_k[l] \right)^2 |b_k| \right\} \\
+ \beta_k[l]^2 \sum_{j=1}^{K} \mathbb{E} \left\{ \left( I_j[l] - I_j[l] \right)^2 |b_k| \right\} \end{array} \right.$$  

Notice that $\mathbb{E} \{ n_k[l] n_k[l] |b_k| \} = N_0$. Long codes can be modeled as random codes with Uniform distributed delay $\tau_k \in [0, T_s]$, $k = 1, \ldots, K$ if the period of the pseudo-noise waveforms is much larger than the symbol period [1]. Hence, $E\{\rho_j[k, 0, \rho_j[k, 1]\} = 0$. The remaining expectations in (14) can be computed as

$$E\left[ \{a_j[l]\}^4 \{d_j[l] - d_j[l]\} |b_k|^2 \right] = \left\{ \begin{array}{ll}
12\sigma_j^2 P_{d, j}[l]^{[i]} & j = k \\
0 & j \neq k \end{array} \right.$$  

with the bit-error-probability $P_{d, j}[l]^{[i]} = P\{d_j[l] | \neq d_j[l]\}$. Following this approach, the objective function $f(\beta_k[l])$ in (14) behaves as depicted in Fig. 2. It can be shown that only in the very low SNR-region the objective function $f(\beta_k[l])$ does not have a sole minimum at $\beta_k[l] = 1$ implying that the SAGE-JDCE scheme forwards EXT values on all user’s symbols in consecutive iterations. Notice that this behavior is independent of system load $K/N_c$ and interleaver size $L$ in contrast to the approach followed by Boutros and Caire [9].

Assuming a convolutional code with generator matrix $G = [5, 7]$ and infinite interleaver size, the bit-error-probability of the SAGE-JDCE scheme can be lower bounded by [10] $P_{d, j}[l] \geq 5 Q(\sqrt{2\gamma_d})$ with $Q(x)$ denoting the error function $Q(x) = 1/\sqrt{2\pi} \int_{x}^{\infty} e^{-x^2/2} dx$, $\gamma_d = \frac{\gamma_0 + \gamma_1}{N_0}$, and $\gamma_1$ is the deinterleaved version of $l$, i.e., $l = \Pi_{-1}[\{l\}]$.

C. Implementation Issues

Re-estimation of the channel coefficients strongly influences the computational complexity of the adaptive SAGE-JDCE scheme. The computation of (11) requires $O\{K^3 + K^2 L\}$ operations. Hence, the time complexity per bit of the SAGE-JDCE scheme is upper bounded by $O\{K^3 L^{-1} + K^2\}$.

The MMSE estimate of $a$ given the MF-output $z$ and the user specific (uncoded) pilot symbols $d_p$ yields the initial estimate $a^{[0]}$. Notice that the first and last bin in each preamble block have been omitted from joint channel estimation due to unknown MAI arising from the adjacent bins to the preamble. To keep the computational complexity low, the output signal of the $k$th user’s MF, weighted by $a^{[0]}$, $k = 1, \ldots, K$ forms the initial symbol estimate $d^{[0]}$.

IV. NUMERICAL EXAMPLES

The feedback structure of the EM-based JDE schemes makes a theoretical performance analysis cumbersome. We therefore evaluate the performance of the SAGE-JDCE scheme by means of Monte Carlo simulations. The system parameters are chosen as follows: All users employ a rate $R = 1/2$ convolutional code with generator matrix $G = [5, 7]$. Gold sequences of length $2^{25} - 1$ are assigned to $s_1(t), \ldots, s_{K;1}(t)$.

For comparison purpose, the performance of the initial MF and a windowed MMSE linear MU detector with window size 20 bins including 2 excess bins on each side followed by separate MLSD have also been included in subsequent plots. Both schemes are supplemented with a separate MMSE channel estimator as described in Section III-C. We refer to these decoding methods as, respectively, MF separate decoding and channel estimation (MF-SDE) and MMSE-SDCE.

The average bit-error-rate (BER) of the proposed receiver is plotted in Fig. 3 versus the average effective SNR $\frac{\gamma_d}{\gamma_b}$, $k = 1, \ldots, K$, for a scenario with $K = 8$ and 32 users, processing gain $N_c = 8$, $L_p = 4$ pilot symbols, block length $L = 40$ and frame length $L = 800$. The number of stages in the receiver has been limited to 20. Within one stage every user’s data sequence is updated once. Notice that averaging has been performed over both the realizations of the channel.
coefficients and the users since all users perform on average the same. It can be seen that the SAGE-JDCE scheme can support an effective load $K_{\text{max}} R/N_c = 2$ ($K_{\text{max}} = 32$). For $\text{BER} = 10^{-2}$, the performance loss of the SAGE-JDCE scheme w.r.t. to the SU-case in known channel is 1.1 dB and 2.8 dB, for $K = 8$ and $K = 32$ users, respectively. Notice that in the limit case $K = 32$, the SAGE-JDCE scheme has difficulties with handling MAI and hence, the BER curve floors out as the average SNR tends to infinity. For the same settings, the gap in performance in known channel of the MMSE-SDCE scheme w.r.t. to the SU-case, in contrast, is given by 5.6 dB and more than 10 dB, respectively. Of course, the MF-SDE scheme cannot cope with MAI. Fig. 4 shows the convergence behavior of the SAGE-JDCE scheme versus the average effective SNR ranging from 0 to 14 dB. It can be seen that for $K = 8$ and $K = 32$ users and low average effective SNR, the SAGE-JDCE roughly requires 14 stages. For $K = 8$, it can be seen that the SAGE-JDCE scheme converges the faster, the higher the average SNR is. This effect may be expected since the quality of the initial estimates also increases with increasing average SNR. At the other end of the scale, the SAGE-JDCE converges after around 1.5 and 3.5 stages for $K = 8$ and $K = 32$, respectively. For $K = 32$ users, the number of stages shows a maximum for about 6 dB. This maximum corresponds to the region of the BER curve, in which the iterative receiver starts to operate successfully. This effect is well-known for turbo codes operating at the decoding threshold [11]. Even if $KR/N_c = 0.5$ (low load), the convergence curve shows an inflection point around 4 dB.

V. SUMMARY

Based on the noise splitting approach, the SAGE algorithm has been used to derive an iterative receiver performing joint channel estimation, interference cancellation, and data decoding. The resulting receiver forwards soft-values from the interference cancellation device to the decoder but receives hard values on the coded symbols in return. The noise weighting coefficients has been used to control the quality of this soft-information. We have formulated one optimization criterion and found an analytical expression for it. The optimized receiver forwards asymptotically extrinsic values on the code symbols in most cases independent of the interleaver size. Monte Carlo simulations for asynchronous DS/CDMA over flat Rayleigh fading channels show that the SAGE based receiver can support an effective system load of $KR/N_c \leq 2$.

REFERENCES