Iterative multiuser joint detection and parameter estimation: A factor-graph approach

Giuseppe Caire
Institut Eurécom
Sophia Antipolis, France
e-mail: giuseppe.caire@eurecom.fr

Antonia Tulino
Università del Sannio
Benevento, Italy
e-mail: tulino@unisannio.it

Ezio Biglieri
Politecnico di Torino
Torino, Italy
e-mail: biglieri@polito.it

Abstract — **We examine a multiple-access AWGN channel with synchronous DS-CDMA in which the channel amplitude and noise variance parameters are unknown a priori. We derive an iterative joint multiuser decoder and parameter estimator based on soft interference cancellation and on soft decision-driven least-squares estimation. Our derivation is obtained by applying the sum-product algorithm to the factor graph of the joint a posteriori probability measure of the information bits and of the unknown channel parameters.**

I. Introduction

In [1], several known iterative multiuser joint decoders for encoded DS-CDMA have been derived and analyzed in a unified way. Here we combine the approach of [1] with that of [2] to the purpose of deriving a conceptually simple iterative multiuser joint decoder and parameter estimator based on soft interference cancellation and on soft decision-directed least squares (LS) estimation. For simplicity's sake, our derivation is restricted to the case of synchronous CDMA and of constant (but unknown) frequency-flat channels. However, it can be extended without too much effort to more general channels. The following notations are used throughout:

- With $\mathbf{A}$ a matrix, $\mathbf{a}_n, \mathbf{a}^k$, and $a_{k,n}$ (or equivalently $[\mathbf{A}]_{k,n}$) denote the nth column, the kth row and the (k,n)th element of $\mathbf{A}$, respectively.
- $\mathbf{z} \sim \mathcal{N}_{\mathbb{C}}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ indicates that the random vector $\mathbf{z}$ is complex circularly-symmetric jointly Gaussian with mean $E[\mathbf{z}] = \boldsymbol{\mu}$ and covariance $E[(\mathbf{z}-\boldsymbol{\mu})(\mathbf{z}-\boldsymbol{\mu})^\dagger] = \boldsymbol{\Sigma}$.
- The superscript † indicates Hermitian transpose.
- The notation $A \propto B$ ($A \doteq B$) indicates that A and B differ by a multiplicative (additive) term.

II. System Model

Consider the uplink of a coded direct-sequence CDMA system with synchronous transmission over frequency-non-selective channels and Nyquist chip-shaping pulses. The system is frame-oriented, i.e., encoding and decoding is performed frame-by-frame and users are synchronous also at the frame level. In each frame, the complex baseband equivalent discrete-time signal originated by sampling at the chip rate the output of a chip-matched filter is given by

$$\begin{cases} \mathbf{Y} = \mathbf{S}\mathbf{W}\mathbf{X} + \mathbf{N} & \text{Data transmission phase} \\ \mathbf{Y}^{(t)} = \mathbf{S}\mathbf{W}\mathbf{X}^{(t)} + \mathbf{N}^{(t)} & \text{Training phase} \end{cases} \tag{1}$$

where:

- $\mathbf{Y} \in \mathbb{C}^{L\times N}$ and $\mathbf{Y}^{(t)} \in \mathbb{C}^{L\times T}$ are the arrays of received signal samples in the data and training phases, respectively.
- $\mathbf{N} \in \mathbb{C}^{L\times N}$ and $\mathbf{N}^{(t)} \in \mathbb{C}^{L\times T}$ are the corresponding arrays of noise samples, assumed complex circularly-symmetric Gaussian i.i.d. $\sim \mathcal{N}_{\mathbb{C}}(0, N_0)$.
- $\mathbf{S} \in \mathbb{C}^{L\times K}$ contains the user spreading sequences by columns.
- $\mathbf{W} = \text{diag}(\mathbf{w})$ where $\mathbf{w} = (w_1, \dots, w_K)$ contains the user complex amplitudes w_k.
- $\mathbf{X} \in \mathbb{C}^{K\times N}$ is the array of transmitted code symbols.
- $\mathbf{X}^{(t)} \in \mathbb{C}^{K\times T}$ is the array of transmitted training symbols (known at the receiver).
- N, T, L and K denote the code block length and the training sequence length (in symbols), the spreading factor (number of chips per symbol) and the number of users, respectively.

The total frame includes $N+T$ symbols. With reference to the above model and to our notation conventions, $\mathbf{s}_k, \mathbf{x}^k, \mathbf{y}_n$ and $\mathbf{x}_n$ denote the k-th user spreading sequence, the k-th user code word, the received signal vector in the n-th symbol interval and the transmitted symbol vector in the n-th symbol interval, respectively. The user spreading sequences are normalized to yield $|\mathbf{s}_k|^2 = 1$ for all k. Hence, the signal-to-noise ratio (SNR) of user k is $\text{SNR}_k = |w_k|^2/N_0$.

At each frame, each user encodes a sequence of information bits $\mathbf{b}_k$ into a code word $\mathbf{x}^k = \phi_k(\mathbf{b}_k) \in \mathcal{C}_k$, where ϕ_k denotes the encoding function and $\mathcal{C}_k$ is the code book of user k, defined over a given complex signal set (e.g., a PSK or QAM constellation). For the sake of simplicity, here we consider binary codes mapped onto BPSK, so that $x_{k,n} \in \{-1,+1\}$. Each code word is independently interleaved before transmission.

III. Factor-graph representation

The receiver we seek minimizes the per-user information-bit error probability while considering the user complex amplitudes w_k and the noise variance N_0 as unknown parameters. We write $\boldsymbol{\theta} \triangleq (\mathbf{w}, N_0)$, with an *a priori* density $p(\boldsymbol{\theta})$ uniform over a subset of *feasible* parameter realizations Θ. Let $\Pi \triangleq \Pr(\mathbf{b}_1, \dots, \mathbf{b}_K, \boldsymbol{\theta} \mid \mathbf{Y}, \mathbf{Y}^{(t)}, \mathbf{X}^{(t)})$ denote the joint *a posteriori* probability distribution of the users information bits and of the parameters given the observation $\mathbf{Y}, \mathbf{Y}^{(t)}$ and the known training symbols $\mathbf{X}^{(t)}$. The optimal joint decoder makes bit-decisions based on the marginal a posteriori probabilities (APP)

$$\mathrm{APP}_{k,\ell}(b) = \sum_{\mathbf{b}_1 \dots, \mathbf{b}_K : b_{k,\ell} = b} \int \Pi \, d\boldsymbol{\theta} \qquad (2)$$

Direct calculation of (2) is too complex in any practical situation. Hence, we seek a low-complexity iterative approximation of the a posteriori marginals (2) by applying the sum-product algorithm to the factor graph representing Π. Straightforward calculations yield

$$\Pi \propto \Xi \, T(\mathbf{w}, N_0) p(\boldsymbol{\theta}) \Pr(\mathbf{X}^{(t)}) \qquad (3)$$

where $\Xi \triangleq p(\mathbf{Y} \mid \mathbf{b}_1, \dots, \mathbf{b}_K, \boldsymbol{\theta}, \mathbf{Y}^{(t)}, \mathbf{X}^{(t)})$, and $T(\mathbf{w}, N_0) \triangleq p(\mathbf{Y}^{(t)} \mid \mathbf{X}^{(t)}, \boldsymbol{\theta})$ is the likelihood function of the parameters given the observation over the training phase only. Also, we have used the fact that $\mathbf{Y}^{(t)}$ is conditionally independent of $\mathbf{b}_1, \dots, \mathbf{b}_K$ given $\mathbf{X}^{(t)}$ and $\boldsymbol{\theta}$, that $\mathbf{b}_1, \dots, \mathbf{b}_K$ are uniformly distributed over all possible binary information sequences, and that they are independent of $\mathbf{X}^{(t)}$ and of $\boldsymbol{\theta}$.

Since $p(\boldsymbol{\theta})$ is uniform, we include it into the proportionality constant in (3) for $\boldsymbol{\theta} \in \Theta$. Moreover, since the training symbols are known and fixed, $\Pr(\mathbf{X}^{(t)}) = 1$ if $\mathbf{X}^{(t)}$ is the actual training sequence and zero elsewhere.

Now, we focus separately on the first and second factors of the RHS of (3). Since users encode their information bits independently and the synchronous CDMA channel (1) is memoryless, we have, after some algebra,

$$\Xi = \prod_{n=1}^{N} \frac{1}{\pi N_0} e^{-|\mathbf{y}_n - \mathbf{S}\mathbf{W}\mathbf{x}_n|^2/N_0} \prod_{k=1}^{K} 1\{\mathbf{x}^k = \phi_k(\mathbf{b}_k)\} \quad (4)$$

where we have used the fact that $\mathbf{Y}$ is independent of $\mathbf{Y}^{(t)}$ and $\mathbf{X}^{(t)}$ given $\boldsymbol{\theta}$ and $\mathbf{X}$, and the fact that $\mathbf{X}$ is independent of anything else given $\mathbf{b}_1, \dots, \mathbf{b}_K$, since $\mathbf{X}$ is a function of the information bits (via the encoding functions $\phi_1, \dots, \phi_K$).

Since $\mathbf{Y}^{(t)}$ is conditionally jointly Gaussian given $\boldsymbol{\theta}$ and $\mathbf{X}^{(t)}$, we have

$$T(\mathbf{w}, N_0) = \frac{1}{(\pi N_0)^{LT}} e^{-\sum_{\ell=1}^{T} \left| \mathbf{y}_\ell^{(t)} - \mathbf{S}\mathcal{X}_\ell^{(t)} \mathbf{w} \right|^2 / N_0} \qquad (5)$$

where we define $\mathcal{X}_\ell^{(t)} = \mathrm{diag}(\mathbf{x}_\ell^{(t)})$, a diagonal matrix with the training symbols $\mathbf{x}_\ell^{(t)}$ as diagonal elements.

By using (4) and (5) in (3), we can write the joint a posteriori probability distribution as proportional to a multivariate function of $\mathbf{b}_1, \dots, \mathbf{b}_K$, $\mathbf{w}, N_0$ and $\mathbf{X}$, where the training symbols $\mathbf{X}^{(t)}$ and the channel output observed signals $\mathbf{Y}$ and $\mathbf{Y}^{(t)}$ play the role of given parameters. We obtain

$$\Pi \;\propto\; \frac{1}{N_0^{LT}} e^{-\sum_{\ell=1}^{T} \left| \mathbf{y}_\ell^{(t)} - \mathbf{S}\mathcal{X}_\ell^{(t)} \mathbf{w} \right|^2 / N_0} \cdot \prod_{n=1}^{N} \frac{1}{N_0^{L}} e^{|\mathbf{y}_n - \mathbf{S}\mathbf{W}\mathbf{x}_n|^2/N_0} \prod_{k=1}^{K} 1\{\mathbf{x}^k = \phi_k(\mathbf{b}_k)\}$$

The factor-graph representation [3] of the above is a bipartite graph with *variable nodes* corresponding to the variables $\{b_{k,\ell}\}$, $\mathbf{w}, N_0$ and $\{x_{k,n}\}$ and *function nodes* corresponding to the terms in the factorization (6), and an edge between a variable node and a function node if the variable is an argument of the function.

The terms $\frac{1}{N_0^L} \exp\left(-\frac{1}{N_0} |\mathbf{y}_n - \mathbf{S}\mathbf{W}\mathbf{x}_n|^2 \right)$ are referred to as *channel transition functions*, since they are proportional to the channel transition pdf (for given parameters $\mathbf{w}, N_0$) and the terms $1\{\mathbf{x}^k = \phi_k(\mathbf{b}_k)\}$ are referred to as *code constraint functions*, since they define the constraint that $\mathbf{x}^k$ is the code word of $\mathcal{C}_k$ corresponding to the information sequence $\mathbf{b}_k$. The factor-graph corresponding to (6) is shown in Fig. 1. The variable and the function nodes are represented as circles and squares, respectively.

IV. Application of the sum-product algorithm

The sum-product algorithm applied to the graph of Fig. 1 yields the following local computation "steps": [3]

Multiuser detection (MUD) step. This computation is carried out at the n-th channel transition function node, for $n = 1, \dots, N$, with output edge towards $x_{k,n}$ and input edges from $\mathbf{w}, N_0$ and from $x_{j,n}$ for all $j \neq k$. The output message is the *extrinsic probability* (EXT) [1] for the symbol-by-symbol detection of symbol $x_{k,n}$ from the observation $\mathbf{y}_n$. Notice that a basic consequence of the sum-product algorithm computation rules is that only EXT pmfs are propagated at each step.

Soft-in Soft-out (SISO) decoding step. This computation is carried out at the k-th code constraint function node, for $k = 1, \dots, K$, with output edge towards $x_{k,n}$ and input edges from $x_{k,\ell}$ for all $\ell \neq n$ and from all $b_{k,j}$. The output message is the EXT pmf for the symbol-by-symbol detection of symbol $x_{k,n}$ given an observation in the form of an priori product pmf and the code constraint defining $\mathcal{C}_k$.

Local parameter estimation (LPE) step. This computation is carried out at the n-th channel transition function node, for $n = 1, \dots, N$, with output towards $\mathbf{w}, N_0$ and inputs from $x_{k,n}$ for all $k = 1, \dots, K$. The output message is the likelihood function for the estimation of $\mathbf{w}, N_0$ based on $\mathbf{y}_n$ (since only $\mathbf{y}_n$ rather than the whole observed signal $\mathbf{Y}$ is used here, we refer to this estimation as "local").

Global parameter estimation (GPE) step. This computation is carried out at the parameter variable node $\mathbf{w}, N_0$, with output towards the n-th channel transition function node, and input from all other channel transition function nodes $\ell \neq n$. and from the training-only likelihood function node $T(\mathbf{w}, N_0)$. The output message is the global likelihood function for the parameters $\mathbf{w}, N_0$ with observations extended over all $\{\mathbf{y}_\ell : \ell \neq n\}$ and over the whole training phase.

V. Simplifications: practical receiver structure

The sum-product algorithm summarized above is too complex for a practical implementation. Simplifying assumptions are necessary to obtain a practical algorithm for iterative joint data detection and parameter estimation.

Soft-decision driven LS estimation. Following [2], we replace $f_n(\mathbf{w}, N_0)$ by its single-mass point approximation $\delta(\mathbf{w} - \widehat{\mathbf{w}}_n)\delta(N_0 - \widehat{N_0}_n)$, where

$$(\widehat{\mathbf{w}}_n, \widehat{N_0}_n) = \arg \max_{\mathbf{w}, N_0} f_n(\mathbf{w}, N_0) \tag{6}$$

is the "global" ML parameter estimate obtained from the likelihood function $f_n(\mathbf{w}, N_0)$.

Next, we make the key assumption that the code symbols $x_{k,n}$ are independent and Gaussian distributed, i.e., we replace the discrete pmf $P_{k,n}(a)$ with the Gaussian pdf $p_{k,n}(a)$ having the same mean $\bar{x}_{k,n}$ and variance. By defining the total observed signal vector of length $L(N+T-1)$ as

$$\underline{\mathbf{y}}_n = (\underbrace{\mathbf{y}_1^T, \ldots, \mathbf{y}_N^T}_{\ell \neq n}, (\mathbf{y}_1^{(t)})^T, \ldots, (\mathbf{y}_1^{(t)})^T)^T$$

and the block matrix of size $L(N+T-1) \times K$

$$\underline{\mathbf{M}}_n = \left[\underbrace{(\mathbf{S}\bar{\mathcal{X}}_1)^T, \ldots, (\mathbf{S}\bar{\mathcal{X}}_N)^T}_{\ell \neq n}, (\mathbf{S}\mathcal{X}_1^{(t)})^T, \ldots, (\mathbf{S}\mathcal{X}_1^{(t)})^T \right]^T$$

(where $\bar{\mathcal{X}}_\ell \triangleq \mathrm{diag}(\bar{\mathbf{x}}_\ell)$) we estimate $\mathbf{w}$ in the form

$$\widehat{\mathbf{w}}_n = \left(\underline{\mathbf{M}}_n^H \underline{\mathbf{M}}_n \right)^{-1} \underline{\mathbf{M}}_n^H \underline{\mathbf{y}}_n \tag{7}$$

and N_0 in the form

$$\widehat{N_0}_n = \frac{1}{L(N+T-1)} \left(|\underline{\mathbf{y}}_n|^2 - |\widehat{\mathbf{w}}_n|^2 \right) \tag{8}$$

The parameter estimates (7) and (8) can be interpreted as follows. The first is the LS solution of the problem $\min_{\mathbf{w}} |\underline{\mathbf{y}}_n - \underline{\mathbf{M}}_n \mathbf{w}|^2$, originated by replacing the (unknown) data symbols by their soft estimates. For this reason, we refer to (7) as a soft-decision directed LS channel estimation. The second can be interpreted as the difference by the total received signal energy $|\underline{\mathbf{y}}_n|^2$ and the total estimated signal energy $|\widehat{\mathbf{w}}_n|^2$, divided by the number of noise samples $L(N+T-1)$, over the data and the training phase (the n-th symbol interval excluded).

An approximate analysis of the bias of the above estimators, for large block size N, shows that the method advocated here yields almost unbiased channel amplitude estimates when symbols are either almost unknown or known with high probability, corresponding to the beginning and to steady-state convergence of the iterative decoder.

Soft interference cancellation. In order to break the exponential complexity of the MUD step, we use again the Gaussian approximation of the symbol distribution. With this approximation, we obtain a cancellation structure that can be interpreted as the result of soft IC followed by conditional MMSE filtering, given the output log-ratios produced by the SISO decoders at the previous decoder iteration.

References

[1] J. Boutros and G. Caire, "Iterative multiuser decoding: unified framework and asymptotic performance analysis." submitted to IEEE Trans. on Inform. Theory, August 2000.

[2] A. Worthen and W. Stark, "Unified design of iterative receivers using factor graphs," *IEEE Trans. Inform. Theory*, vol. 47, pp. 843–849, February 2001.

[3] F. Kschischang, B. Frey, and H. Loeliger, "Factor graphs and the sum-product algorithm," *IEEE Trans. Inform. Theory*, vol. 47, pp. 498–519, February 2001.

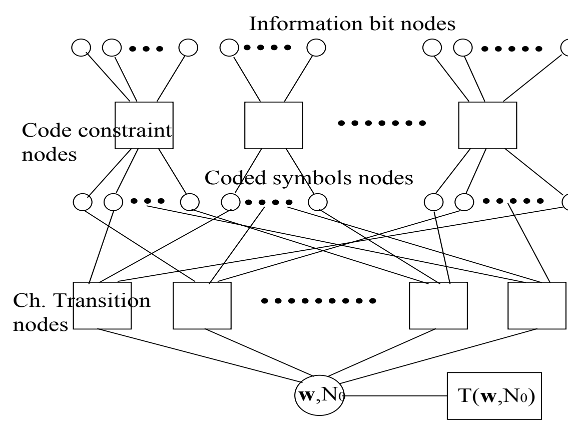

Figure 1: The edges connecting the coded symbols nodes with the channel transition nodes represent random interleaving.