SOFT ITERATIVE DECODING FOR OVERLOADED CDMA

Gianmarco Romano[1], *Francesco Palmieri*[1], *Peter K. Willett*[2]

[1]Dipartimento di Ingegneria dell'Informazione, Seconda Università di Napoli, Aversa (Italy)
[2]Dept. of Electrical and Computer Engineering, The University of Connecticut, Storrs, CT (USA)

ABSTRACT

The probabilistic association algorithm (PDA) is proposed as a quasi-optimal solution for synchronous *overloaded* CDMA detection. Overloaded CDMA is useful in the applications in which the band is a limited resource and more users are to conveyed on the same channel. The PDA algorithm, already successfully applied to underloaded CDMA, is extended to the more difficult overloaded CDMA problem. PDA uses dynamic soft updates for the a posteriori probabilities and it is derived under the Gaussian assumption for the superposition of Multiuser Interference and Gaussian noise. A discussion on the separability of the two-class problem suggests how a Dynamic PDA (D-PDA) may solve the problem more effectively in comparison to a fixed-order PDA (F-PDA). Performance results are reported in this paper and show that D-PDA can provide good decoding with limited computational complexity with the bit error rate going to zero as the signal-to-noise ratio increases.

1. INTRODUCTION

Synchronous CDMA detection can be easily translated [1, 4] to the following linear regression problem: estimate $\mathbf{b}$ from

$$\mathbf{y} = SA\mathbf{b} + \mathbf{v} = F\mathbf{b} + \mathbf{v}, \tag{1}$$

where $\mathbf{y}$ is the N-dimensional observed column vector; $\mathbf{b} = (b_1, ..., b_M)^T = (\pm 1, \pm 1, ..., \pm 1)^T$ is the M-dimensional binary information vector and $\mathbf{v}$ is the N-dimensional noise. More specifically, each component of $\mathbf{b}$ represents the information bit emitted by one of the M users; the additive noise $\mathbf{v}$ is independent from the information bits, and it is Gaussian with zero mean and covariance matrix R_v. The matrix $A = diag(\alpha_1, ..., \alpha_M)$ controls the energy for each user (users may be at different distances from the base station) and matrix S contains in each column the "signature" for that user. The model allows for users' signatures to be real-valued and to have different energies. Therefore in the discussion that follows we will look at the second equivalent expression in which matrix F is the result of all the transformations such as: spreading, magnitude control, matched filtering, etc. For derivations of the standard model see, for example, [1, 4, 5].

For simplicity we assume in this paper that source symbols are binary and observations are real. The framework can be easily extended to complex symbols and it will not be discussed here. The relation between N and M and the nature of the matrix F and the noise $\mathbf{v}$ make the solution to the problem of estimating $\mathbf{b}$ from $\mathbf{y}$ of varying difficulty. If $N > M$ we have the so-called *underloaded* problem because the number of users is smaller than the spreading factor. The situation is similar in the case in which $N = M$ – the so-called *fully-loaded* problem. In such cases the usual approach is to multiply $\mathbf{y}$ by the pseudo-inverse of F. This amounts to computing $\mathbf{x} = F^{\#}\mathbf{y} = \mathbf{b} + \mathbf{n}$, where the users' source bits have been de-mixed and the contaminating noise $\mathbf{n}$ has correlation matrix $R_n = F^{\#} R_v F^{\#T}$. If R_v is diagonal and F has orthogonal columns, the problem is easily solved because also R_n is diagonal and detection for all the users can be performed component-by-component. Conversely, even if R_v is diagonal – but F has non-orthogonal columns – as happens in many well-known spreading codes [4], the problem is exponentially complex because the Maximum Likelihood (ML) detector requires in general comparison with all the 2^M binary configurations. Various sub-optimal solutions can be applied such as Sphere Decoding [6], Probabilistic Data Association [3, 4], Decision-Feedback (see references in [1]) and others [5].

The problem however becomes more difficult if $M > N$ (i.e., the *overloaded* case). Such a scenario is of great interest when N cannot be increased due to bandwidth constraints. Remember that bandwidth and bit-rate are determined by the chip frequency and the length N. If they are fixed, the only way to obtain an increase in the number of users on the channel is to increase power. The complexity of the decoding process however, remains crucial. In [1] Kapur and Varanasi point to group-detection strategies to solve the problem, since linear detectors (such as the linear MMSE decision-feedback receiver) are in general not able to separate the users [2].

In this paper we review the overloaded CDMA problem and with the aid of a two-dimensional example we show how a dynamic soft-decoding algorithm may be a good candidate solution with limited complexity.

2. OVERLOADED CDMA

To appreciate the nature of overloaded CDMA detection, let us look at the example shown in Figure 1, having $N = 2$ and $M = 5$ users. The $2^5 = 32$ points are generated with the equation $F\mathbf{b}$ with all the 32 binary configurations and a matrix chosen arbitrarily.

$$F = \begin{bmatrix} -1.90 & 1.75 & -1.40 & -0.52 & 0.22 \\ 1.61 & -0.92 & -1.02 & 0.02 & -0.17 \end{bmatrix} \quad (2)$$

The additive Gaussian noise disperses the observations around each signal point (typically in a spherical fashion), not shown in the picture. The five pictures show the partitions corresponding to the five bits $\{b_1, ..., b_5\}$. Note how user 3 is the easiest to decode, since a linear discriminant function can separate the two subsets and decode b_3. User 1 instead is not detectable via a linear classifier; but if b_3 were known, a linear classifier could detect b_1[1]. The nature of the problem suggests that for a linear receivers it is in general impossible to receive correctly all the users with an error probability that approaches zeros even if the signal-to-noise ratio becomes arbitrarily large (see [2]). The picture suggests that the receiver must be recursive in nature as the decisions about the various users must be taken progressively or more generally jointly.

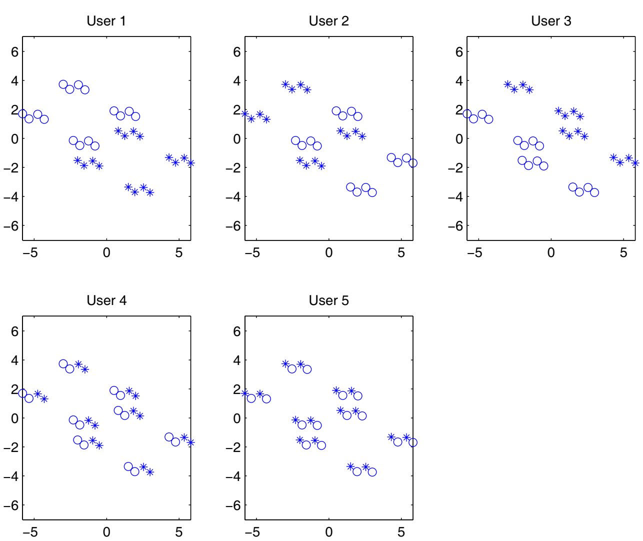

Fig. 1. Signal points in $N = 2$ dimensions and their partitions for an arbitrary F for $M = 5$ users.

The algorithms suggested in the following are built according to the idea of performing temporary estimates to be used iteratively: as the users become progressively separable the membership probabilities converge to a quasi-optimal solution.

[1]And so on, in an inductive fashion, with the other users.

3. PROBABILISTIC DATA ASSOCIATION

Let us rewrite the main equation as

$$\mathbf{y} = F\mathbf{b} + \mathbf{v} = \mathbf{f}_i b_i + F_i \mathbf{b}_i + \mathbf{v} = \mathbf{f}_i b_i + \mathbf{w}_i, \quad (3)$$

where we have separated the ith user's contribution. Vector $\mathbf{f}_i$ is the ith column of F and F_i is the $M \times (N-1)$ matrix F expurgated of its ith column. Vector $\mathbf{w}_i$ is the superposition of $N-1$ users' interference (MUI) and gaussian noise. Probabilistic Data Association (PDA) algorithms [3, 4] are derived assuming a Gaussian distribution for $\mathbf{w}_i$ (the Gaussian assumption on the MUI is often invoked in the CDMA literature; see for example [1]). The PDA is based on iterative computation of the a posteriori probabilities

$$P(i) = Pr\{b_i = 1 | \mathbf{y}, \{P(i)\}_{j \neq i}\}. \quad (4)$$

The approach is quite successful because it provides a means of taking into account partial knowledge about a user's bit to improve the decoding accuracy for the others. In the underloaded cases the PDA algorithm has been shown to provide excellent quasi-optimal solutions with very limited complexity with a number of users up to $M = 40$ [3, 4]. Various fixed orderings for the updates can be utilized based, for example, on the users' relative energies. The overloaded cases has shown various degrees of success [3] as performance seem to depend critically on the nature of matrix F and the update order. The likelihood ratio for the ith user can be written as

$$\begin{aligned} & Pr\{b_i = 1 | \mathbf{y}, \{P(i)\}_{j \neq i}\} / Pr\{b_i = -1 | \mathbf{y}, \{P(i)\}_{j \neq i}\} \\ = \; & f_{\mathbf{y}}\left(\mathbf{y} | b_i = 1, \{P(i)\}_{j \neq i}\right) / f_{\mathbf{y}}\left(\mathbf{y} | b_i = -1, \{P(i)\}_{j \neq i}\right) \\ = \; & \mathcal{N}\left(\mathbf{y}; \mathbf{f}_i + \mu_i, C_i\right) / \mathcal{N}\left(\mathbf{y}; -\mathbf{f}_i + \mu_i, C_i\right) \\ = \; & exp\left[2\left(\mathbf{y} - \mu_i\right)^T C_i^{-1} \mathbf{f}_i\right], \end{aligned}$$

where $\mu_i = E[\mathbf{w}_i]$ and $C = cov[\mathbf{w}_i]$. Since $(p/1-p) = e^x$, we have $p = 1/(1+e^{-x}) = \ell(x)$, where $\ell(x)$ is the logistic function. The a posteriori probability for the ith user is therefore written as

$$P(i) = \ell\left(2\left(\mathbf{y} - \mu_i\right)^T C_i^{-1} \mathbf{f}_i\right). \quad (5)$$

Temporary knowledge about the users is carried over in the iteration through dynamic updates of means and covariances:

$$\begin{aligned} \mu_i &= E[\mathbf{w}_i] = E[F_i \mathbf{b}_i + \mathbf{v}] = F_i E[\mathbf{b}_i] \\ &= F_i \begin{bmatrix} 2P(1) - 1 \\ . \\ 2P(i-1) - 1 \\ 2P(i+1) - 1 \\ . \\ 2P(M) - 1 \end{bmatrix} \end{aligned} \quad (6)$$

$$
\begin{aligned}
C_i &= cov[\mathbf{w}_i] \\
&= F_i E[\mathbf{b}_i \mathbf{b}_i^T] F_i^T + R_v - F_i E[\mathbf{b}_i] E[\mathbf{b}_i^T] F_i^T \\
&= F_i cov[\mathbf{b}_i] F_i^T + R_v \\
&= F_i diag\,(\,var[b_1], ..., var[b_{i-1}],\ var[b_{i+1}], ... \\
&\quad ,\ var[b_M])\, F_i^T + R_v, \qquad (7)
\end{aligned}
$$

where $var[b_i] = 4P(i)(1-P(i))$. Note that no previous knowledge is used about the ith user on its own updates. A *fixed-order PDA algorithm* (F-PDA) is as follows:

(1) Initialize $P(i) = 1/2,\ i = 1, ..., M$;

(2) Compute μ_{l_1} and C_{l_1} for a chosen l_1 and compute $P(l_1)$;

(3) with the available values of $P(i), i = 1, ..., M$ compute μ_{l_2} and C_{l_2} and $P(l_2)$;

(4) after having updated all the users, iterate.

The PDA strategy works very well in the underloaded case. Various order-updates are possible and the computational complexity remains limited and after only a few cycles the algorithm converges [3, 4]. Also, more efficient algorithms can be derived with the help of the matrix inversion lemma to avoid the inversion of matrix C_i [3].

4. THE SEPARATION CRITERION

The nature of the overloaded problem suggests that the PDA updates should follow a criterion that depends on how the various users are currently discriminated. More specifically, when a point bfy is observed, in order to discriminate a user's bit with a two-class decision, we could look at the user that would be well decoded with a linear classifier.

To see more clearly recall that in a two-class problem (classes 1 and 2) with conditional densities

$$
f(\mathbf{y}|1) = \mathcal{N}(\mathbf{y}; \mu_1, \Sigma)\,; \quad f(\mathbf{y}|2) = \mathcal{N}(\mathbf{y}; \mu_2, \Sigma)\,, \qquad (8)
$$

optimal ML decision is linear with an hyperplane dividing in two the N-dimensional space, i.e. $\hat{s}_{ML} = 1$ if

$$
\mu_1^T \Sigma^{-1}(\mu_1 - 2\mathbf{y}) < \mu_2^T \Sigma^{-1}(\mu_2 - 2\mathbf{y}), \qquad (9)
$$

else $\hat{s}_{ML} = 2$. The error probability is

$$
P_e = Q\left(\frac{1}{2}\sqrt{(\mu_1 - \mu_2)^T \Sigma^{-1}(\mu_1 - \mu_2)}\right). \qquad (10)
$$

The divergence parameter

$$
D = (\mu_1 - \mu_2)^T \Sigma^{-1}(\mu_1 - \mu_2), \qquad (11)
$$

can be used as a measure of performance: the larger the divergence is, the smaller the error probability becomes. Note that the divergence D reflects non only the mean distance between the means, but also the degree of separability of the two classes. If we apply this idea within the context of probabilistic data association, the decision on the ith user between the two classes

$$
\mathcal{N}(\mathbf{y}; \mathbf{f}_i + \mu_i, C_i)\,; \quad \mathcal{N}(\mathbf{y}; -\mathbf{f}_i + \mu_i, C_i)\,, \qquad (12)
$$

has divergence

$$
D_i = 4\ \mathbf{f_i}^T cov[\mathbf{w}_i]^{-1} \mathbf{f}_i. \qquad (13)
$$

Therefore the degree of solvability of the two-class problem can be related to the parameter $\mathbf{f_i}^T cov[\mathbf{w}_i]^{-1} \mathbf{f}_i$. Note that during iterative probabilistic association $cov[\mathbf{w}_i]$ depends on the probabilities estimated on the other users. At the beginning of the PDA iteration, the two clusters corresponding to a user can have a very small value of D_i and be highly superimposed. See for example user 4 in Figure 1. The situation may change as the iterations progress: the degree of separability of the various users becomes **dynamic**!

The above considerations suggest the following *Dynamic PDA Algoritm* (D-PDA):

(1) Inizialize $P(i) = 1/2,\ i = 1, ..., M$;

(2) compute $1/4D_i,\ i = 1, ..., M$;

(3) update P for the user with the largest D_i;

(4) recompute D_i for the other users excluding the one already considered;

(5) update P for the user with the largest D_i in the remaining group;

(6) Recompute D_i for the remaining users; etc.

The process spans all the users in M iterations and at each iteration all the means and covariances must be updated. The computations complexity of the D-PDA is larger, but the convergence is faster and much more accurate than fixed-order PDA.

5. SIMULATIONS

We have performed a good number of simulations for various values of N and M and various matrices F. The PDA approach is a sub-optimal but gives performance that are very close to the optimal ML detector. Reported here are the results for the example of Figure 1 with $N = 2$ and $M = 5$. The typical convergence behavior of the probabilities $P(i), i = 1, ..., M$ is shown in Figure 2. The fixed-order algorithm is consistently slower and sometimes tends to oscillate. The D-PDA instead is quite fast and in our simulations we have never observed oscillating behaviors. Dynamic PDA is more accurate in providing the a posteriori probabilities for each observation. We would like to point out that the soft-decoding approach can be very useful when a decision on the users' bits is postponed to another decoding stage at a higher hierarchical level such as that used in coded-CDMA [7].

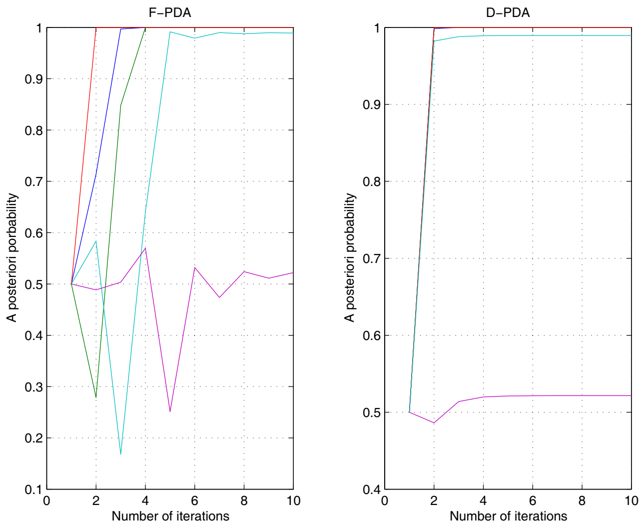

Fig. 2. Typical probability convergence patterns for the fixed-order PDA and the Dynamic PDA.

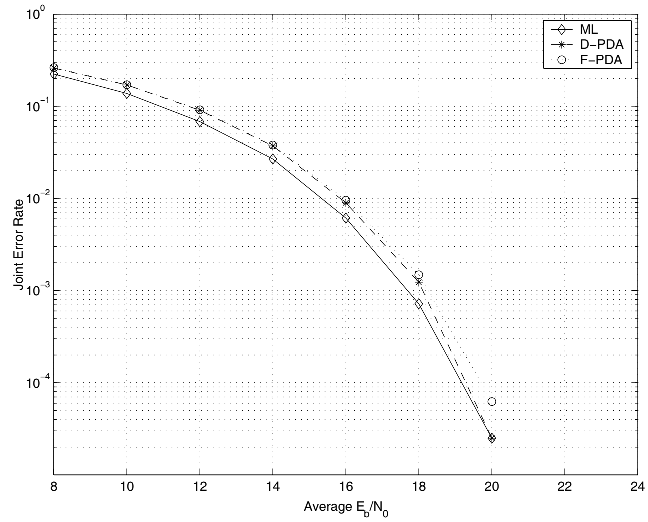

Fig. 3. BER for ML detector, F-PDA and D-PDA for various signal-to-noise ratios.

Figure 3 shows the probability of error for the two algorithms in comparison to the ML solution on the example $N = 2$, $M = 5$ discussed before. Final decisions are taken with a threshold at 1/2 on the final probabilities after 10 iterations. The dynamic updates are very consistent in providing an asymptotically null BER as the signal-to-noise ratio is increased. The error probabilities are the probabilities that at least one user is incorrectly decoded. The signal-to-noise ratio is computed on the average received energy. Note that the users have different energies and some are more critical than others due to the topology of the total constellation. We have found consistently that the D-PDA is generally better than the F-PDA as this one sometimes tends to floor for critical signal matrices F.

6. CONCLUSIONS

The PDA criterion is applied to synchronous overloaded CDMA. The problem has been reduced to a constrained linear regression problem with an unknown binary vector. The soft-decoding iterative strategy has been applied to the difficult problem of demixing users when their number is larger than the number of dimensions of the observation space. The detection in the overloaded problem cannot be performed with linear classifiers and the complexity of a general maximum likelihood receive calls for efficient sub-optimal algorithms. The PDA approach, especially in its dynamic version, can provide a solution with very limited computational complexity. We are currently pursuing further investigations on the conditions on the mixing matrix F that condition the different efficacies of the fixed-order algorithm in contrast to the dynamic ones.

7. REFERENCES

[1] A. Kapur and M.K. Varanasi, "Multiuser Detection for Overloaded CDMA Systems," *IEEE Trans. on Information Theory*, Vol. 49, n. 7, pp. 1728-1742, July 2003.

[2] A. Kapur, M.K. Varanasi and C.T. Mullis, "On the Limitations of Generalized Welch Bound Equality Signals under Linear MMSE Detection," Proc. of *2004 Conference on Information Sciences and Systems*, Princeton, pp. 999-1003, March 17-19 2004

[3] J. Luo, *Improved Multiuser Detection in Code-Division Multiple Access Communications*, Ph.D. Thesis, The University of Connecticut, 2002.

[4] J. Luo, K.R. Pattipati, P. Willett, and F. Hasegawa, "Near-optimal multiuser detection in synchronous cdma using probabilistic data association," *IEEE Transactions on Communications*, Vol. 5, n. 9, pp. 361-363, Sept 2001.

[5] S. Verdú, *Multiuser Detection*, Cambrige University Press, 1998.

[6] E. Viterbo and J. Boutros, "A Universal Lattice Code Decoder for Fading Channels," *IEEE Transactions on Information Theory*, Vol. 45, n. 5, pp. 1639-164, July 1999.

[7] X. Wang and H.V. Poor, "Iterative (Turbo) Soft Interference cancellation decoding for coded CDMA," *IEEE Trans. on Communications*, Vol. 47, n. 7, pp. 1046-1061, July 1999.