INTERFERENCE CANCELING RECEIVERS WITH GLOBAL MMSE-ZF STRUCTURE AND LOCAL MMSE OPERATIONS

Ahmet Baştuğ, Dirk T.M. Slock

Eurecom Institute
2229 Route des Crêtes, B.P. 193, 06904, Sophia Antipolis, Cedex, FRANCE
ahmet.bastug@@eurecom.fr, dirk.slock@@eurecom.fr

ABSTRACT
In this paper we consider iterative techniques for multi-user detection. In the overdetermined case in which signal and noise subspaces exist, polynomial expansion can perhaps most conveniently be applied to a global MMSE ZF linear receiver. In the resulting iteration sections, we then propose to replace the various receiver components by a (possibly non-linear) MMSE variant, by interpreting the signal estimation functionality of each component. The resulting receiver sections, which can also be applied in the underdetermined case, have interpretations in terms of interference cancellation. We shall apply the proposed receiver strategy to the downlink in the FDD mode of UMTS, in which we consider a mobile terminal that also combats explicitly intercell interference.

1. BASEBAND DOWNLINK TRANSMISSION MODEL

Figure 1: Baseband downlink transmission model

The baseband downlink transmission model of a CDMA-based cellular system with an interfering neighboring cell, BS2, is shown in Fig. 1. In practice, however, this model on which we base our discussion can be generalized to multiple interfering base stations. For each base station BSj, s.t. j ∈ {1, 2}, the Kj linearly modulated user signals are transmitted over the same linear multipath channel hj(t)

\[
\mathbf{y}[l] = \mathbf{u}_1[l] + \mathbf{u}_2[l] + \mathbf{v}[l]
\]

\[
\mathbf{y}_j[l] = \sum_{k=1}^{K_j} \sum_{i=0}^{N-1} h_{j,k}[i] h_{j,k}[l - i] \quad j \in \{1, 2\}
\]
\[
\begin{align*}
y_j[t] &= [y_{j,1}[t] \ldots y_{j,mq}[t]]^T, \\
h_j[t] &= [h_{j,1}[t] \ldots h_{j,mq}[t]]^T, \\
v[t] &= [v_1[t] \ldots v_{mq}[t]]^T.
\end{align*}
\]

Here, with a slight abuse of notation, \(h_j[t]\) represents the vectorized samples (represented at chip rate) of the overall channels (assumed to have the same delay spread of \(N\) chips) including the pulse shape, the propagation channel and the anti-aliasing receiver filter.

When we model the scramblers as unknown, i.i.d., aperiodic sequences and the symbol sequences as i.i.d., stationary, white sequences, then the chip sequences \(b_{l_1,2}[t]\) are also stationary and white. Therefore, both the intracell and intercell contribution to \(y[t]\) are stationary vector processes the continuous-time counterparts of which are cyclostationary with chip period. Finally, as the remaining noise, \(v[t]\), is also assumed to be white and stationary, the sum of interference and noise is a stationary chip rate vector process.

### 2. POLYNOMIAL EXPANSION (PE) AND ITERATIVE INTERFERENCE CANCELLING (IC) STRUCTURES

In this section, we develop intra and intercell interference cancelling (IC) structures based on polynomial expansion (PE), see [1] for an introduction to PE.

Assuming without loss of generality that there is a single highly interfering BS, i.e. \(BS_2\), a vector of received signal over one symbol period can be written as

\[
Y[n] = [H_1(z)S_1[n]C_1 \quad H_2(z)S_2[n]C_2] [A_1[n] \quad A_2[n]] + V[n] \\
= \tilde{G}(n,z)A[n] + V[n]
\]

representing the system at the symbol rate.

![Figure 2: Channel impulse response of \(H_j(z)\).](image)

As shown in Fig. 2, \(H_j(z) = \sum_{i=0}^{M_j-1} H_j[i] z^{-i}\) is the symbol rate \(Lmq \times L\) channel transfer function, \(z^{-1}\) being the symbol period delay operator. The block coefficients \(H_j[i]\) are the \(M_j = [L + N_j + d_j-1]\) parts of the block Toeplitz matrix with \(mq \times 1\) sized blocks, \(h_j\) being the first column whose top entries might be zero for it comprises the transmission delay \(d_j\) between \(BS_j\) and the mobile terminal. The \(L \times L\) matrices \(S_j[n]\) are diagonal and contain the scrambler of \(BS_j\) for symbol period \(n\). The column vectors \(A_j[n]\) contain the \(K_j\) symbols of \(BS_j\). \(A[n]\) contains the total \(K = K_1 + K_2\) user symbols in two base stations, and \(C_j\) is the \(L \times K_j\) matrix of the \(K_j\) active codes for \(BS_j\). Based on the equivalency of the chip rate and the symbol rate representations as demonstrated in Fig. 3 which shows the conversion between the two representations via serial-to-parallel and parallel-to-serial converters by vectorizing and sample rate conversion by a factor \(L\), one can pass from the chip level models represented by (1), (2) and (3) to the symbol level model represented by (4) where \(\tilde{G}(n,z)\) is a \(Lmq \times K\) channel-plus-spreading symbol rate filter. Although it is possible to find an FIR left inverse filter for \(\tilde{G}(n,z)\) provided that \(Lmq \geq K\), this is not practical since \(\tilde{G}(n,z)\) is time-varying due to the aperiodicity of the scrambling. Therefore, we will introduce a less complex approximation to this inversion based on the polynomial expansion technique [1]. Instead of basing the receiver directly on the received signal, we shall first introduce a dimensionality reduction step (from \(Lmq \to K\)) by equalizing the channels with minimum mean square error zero forcing (MMSE-ZF) chip rate equalizers \(F_j(z)\) followed by a bank of correlators [3]. Let \(X[n]\) be the \(K \times 1\) correlator output, which would correspond to the Rake receiver outputs if channel matched filters were used instead of channel equalizers. Then,

\[
X[n] = \tilde{F}(n,z)Y[n] \\
= [C^H S_1[n]F_1(z) \quad C^H S_2[n]F_2(z)] [\tilde{G}(n,z)A[n] + V[n]] \\
= M(n,z)A[n] + \tilde{F}(n,z)V[n]
\]

where \(M(n,z) = \tilde{F}(n,z)\tilde{G}(n,z)\) and ZF equalization re-
sults in $F_j(z)H_j(z) = I$. Hence,

$$M(n, z) = \sum_{i=-\infty}^{\infty} M[n, i]z^{-i} = \begin{bmatrix} I & \ast \\ \ast & J \end{bmatrix}$$

due to proper normalization of the code energies.

In order to obtain the estimate of $A[n]$, we initially consider the processing of $X[n]$ by a decorrelator as

$$\hat{A}[n] = M(n, z)^{-1}X[n]$$

$$= (I + M(n, z)^{-1}X[n].$$

The correlation matrix $M(n, z)$ has a coefficient $M[n, 0]$ with a dominant unit diagonal in the sense that all other elements of the $M[n, i]$ are much smaller than one in magnitude. Hence, the polynomial expansion approach suggests to develop $(I + M(n, z)^{-1})^i = \sum_{i=0}^{\infty}(-M(n, z))^i$ up to some finite order, which leads to

$$\hat{A}^{-1}[-n] = 0$$

$$i \geq 0 \hat{A}^{(i)}[n] = X[n] - M(n, z)^{i} \hat{A}^{(i)}[n].$$

In practice, we stop at the first-order expansion, the quality of which depends on the degree of dominance of the diagonal of the static part of $M(n, z)$ with respect to its off-diagonal elements and the dynamic part.

At first order, the expression for the user of interest (user one) becomes:

$$\hat{a}_{1,1}[n] = e_1^H \hat{A}^{(1)}[n]$$

$$= e_1^H(X[n] - M(n, z)^0 \hat{A}^{(0)}[n])$$

$$= e_1^H(2X[n] - M(n, z)X[n])$$

$$= e_1^H \hat{F}(n, z)(2Y[n] - \hat{G}(n, z)X[n])$$

$$= e_1^H \hat{S}_1^H[n]F_1(z)(2\hat{Y}[n] - \hat{G}(n, z)\hat{F}(n, z)Y[n])$$

where $e_i$ (i.e. $\{1, 2, \ldots, K \}$) is a column vector having a 1 at the $i$th position and the rest filled with 0s. From this symbol rate equation, one can obtain the chip rate signal processing diagram in Fig. 4 by using the equivalencies in Fig. 3. Each branch in the IC block is formed in order by a linear filter, a downsampler, a descrambler, a serial to parallel converter, a despreader, a resparser, a parallel to serial converter, a scrambler, an up sampler and a re-channeling filter.

Now that we have obtained the first order polynomial expansion decorrelating structure, we look for an equivalent interference cancelling representation that excludes the user signal estimation operation in the first IC branch [4]. Let

$$I_{K \times K} = [e_1e_2 \ldots e_K] = [e_1e_1]^H$$

$$\hat{G}(n, z) = \hat{g}_1(n, z) \hat{G}_1(n, z)$$

$$\hat{F}(n, z) = \hat{f}_1(n, z)^T \hat{F}_1(n, z)^T$$

$$A[n] = [a_{1,1}[n] \hat{A}_1[n]^T]^T$$

$$\hat{f}_1(n, z)\hat{g}_1(n, z) = 1.$$

Figure 4: Chip rate first order polynomial expansion structure.

Then, dropping the indices and starting from (5), we obtain

$$\hat{a}_{1,1} = (2\hat{f}_1\hat{g}_1 - \hat{f}_1\hat{g}_1 - \hat{f}_1\hat{G}_1\hat{F}_1\hat{g}_1)\alpha_{1,1} +$$

$$(2\hat{f}_1\hat{G}_1 - \hat{f}_1\hat{G}_1 - \hat{f}_1\hat{G}_1\hat{F}_1\hat{G}_1)\alpha_{1,1}$$

$$+ (\hat{f}_1\hat{G}_1 - \hat{f}_1\hat{G}_1\hat{F}_1\hat{G}_1)\alpha_{1,1}$$

$$+ (\hat{f}_1\hat{G}_1 - \hat{f}_1\hat{G}_1\hat{F}_1\hat{G}_1)\alpha_{1,1}$$

$$= \hat{f}_1(I - \hat{G}_1\hat{F}_1)Y$$

$$= e_1^H \hat{F}(Y - \hat{G}_1\hat{F}_1)Y$$

$$= e_1^H \hat{F}(Y - \hat{G}_1\hat{F}_1)Y.$$
3. LOCAL MMSE OPERATIONS

One of the advantages of the MMSE-ZF approach w.r.t. the MMSE approach is that clear symbol and chip sequence estimates appear at various points in the receiver which can be improved locally by replacing whatever the global MMSE-ZF structure yields as estimates by improved estimates in the MMSE sense. Any local MMSE improvement should lead to global MMSE improvement. In an iterative PE approach, such modifications should also lead to smaller offdiagonal power and hence faster convergence of the iterations to an estimate that is closer to a MMSE estimate. The interpretations to be discussed though being applicable to any iteration except for a possible dependence of a number of quantities on the iteration index, we will concentrate only on the first iteration displayed in Fig. 5 where the role of filters \( f_1 \) and \( f_2 \) is to produce estimates of the chip sequences of BS1 and BS2. Those estimates can be improved by replacing the MMSE-ZF chip rate equalizers \( f_1 \) and \( f_2 \) considered so far by MMSE equalizers which, though they perturb the orthogonality of the codes, do not enhance as much the intercell interference plus noise [2]. The estimated chip sequence then gets descrambled and passes by correlators to produce symbol estimates for the intracell/intercell interferers. These symbol estimates can be improved in a variety of ways by static linear or nonlinear functions \( \zeta (\cdot) \) such as to exploit the symbol variance to introduce a LMMSE weighting factor or to exploit the symbol constellation by taking hard decisions which however may not improve the estimate. Hence, one may replace the hard decision by a variety of soft decisions. A locally optimal MMSE estimate is obtained by using a hyperbolic tangent function and the estimated symbols are then respread, scrambled and added to produce again an estimate of the chip sequence. The purpose of the rechanneling filters \( g_j \) (in the PE so far equivalent to the channel impulse response \( h_j \)) is to produce an intracell/intercell interference estimate at the level of the received signal, on the basis of the chip sequences estimates. The linear filter after the IC canceller, \( f_3 \), which is considered to be equal to \( f_1 \) in the polynomial expansion setup, can also be (in fact must be in case of local MMSE modifications on the IC branches) replaced by a Rake or a LMMSE filter (taking into account reduced interference level). The second one would perform better but it requires the estimation of structured residual interferences from the two base stations.

For the spreading and despreading with the Walsh-Hadamard codes of the active users, it is suggested to (de)spread with all WH codes simultaneously, which can be done with the Fast Walsh-Hadamard transform (FWHT). Presence or absence of codes can be detected by estimating the powers at the correlator outputs.

So far we have considered a unique spreading factor. The multirate case corresponding to variable spreading factors can be reformulated as a multicode case. The tree structure of the orthogonal variable spreading factors (OVSFs) allows to interpret the interfering users as equivalent low rate users (at least if the symbol constellation of the interferers will not get exploited, in which case only second order statistics count). Hence in that case one needs at most to consider the multicode case at low rate for the user of interest. If on the other hand one also desires to exploit the interferers’ symbol constellation, then one needs to detect and use the actual spreading factors of the interferers. The tree structure of the OVSFs can be exploited to progressively explore increasing spreading factors and investigate the finite alphabet hypothesis at the various tree branches. In this case one cannot use the FWHT as such but the ingredients leading to its derivation (OVSF in fact) can still be used for fast (de)spreading.

The implementation of the hyperbolic tangent and the hard decisions requires the estimation of the user symbol powers. Let \( z_{i,j} \) represent the chip-rate channel-MMSE filter cascade of \( h_j \) and \( f_j \), including the external (up/down) samplers. Then, the expected value of the received power after the \( k^{th} \) decorrelator on the first branch is equal to

\[
\beta_{1,k} = \left| \frac{z_{1,k}(0)}{\alpha_{1,k}} \right|^2 + \frac{1}{L} \left( \| z_{1,1} \|^2 - | z_{1,1}(0) |^2 \right) \\
 \sum_{i=1}^{K_2} | \alpha_{2,i} |^2 + \| z_{1,2} \|^2 \sum_{i=1}^{K_1} | \alpha_{2,i} |^2 + \sigma_e^2 \| f_1 \|^2 \\
= \left| \frac{z_{1,1}(0)}{\alpha_{1,k}} \right|^2 + \left( \| z_{1,1} \|^2 - | z_{1,1}(0) |^2 \right) \sigma_t^2 \\
+ \| z_{1,2} \|^2 \sigma_e^2 + \sigma_e^2 \| f_1 \|^2 \\
= \left| \frac{z_{1,1}(0)}{\alpha_{1,k}} \right|^2 + \sigma_e^2 \\
\]

since the expected value of the correlation coefficient between any two nonorthogonal codes or between two shifted versions of the same code is equal to \( 1/L ; \) zero-forced channel coefficient is \( z_{1,1}(0) \) (which is in fact 1 for unbiased MMSE case). We assume that we know the channel parameters, the total received power from both base stations and the noise variance but we do not know each single user’s power. Therefore, the estimate of \( \| z_{1,1} \|^2 \), i.e. \( | \alpha_{1,k} |^2 \), can be calculated by subtracting the effective interference \( \sigma_e^2 \) from the long term average power \( \beta_{1,k} \) and scaling by \( 1/ \left| z_{1,1}(0) \right|^2 \). If the result is negative we replace it by 0. The powers of users in the second cell are estimated similarly.

For the MMSE filter construction after the subtractor, estimation of total residual interference powers originating from the two base stations are necessary. Let \( \gamma_{int} \) and \( \gamma_{user} \) be the average chip rate measured powers before the rescrambler blocks in the two IC branches and the user chip rate power respectively. Then the residual powers are \( \sigma_{leak1}^2 = [ \sigma_t^2 - \gamma_{user} - \gamma_{int} ] \) and \( \sigma_{leak2}^2 = [ \sigma_t^2 - \gamma_{int} ] \).
In the sequel, for simulations we will assume that the symbols are from a QPSK constellation. Hence we give here the optimal hyperbolic tangent estimator and the hard estimator for received QPSK symbols. Let \( r_{j,k} = \chi_{j,k} + i \psi_{j,k} \) be the correlator output for code \( k \) from BS\(_j\), then
\[
\hat{\alpha}_{j,k}^{hyp} = \frac{\alpha_{j,k}}{\sqrt{2}} \tanh(\sqrt{2} | z_{1,1}(0) | \frac{\alpha_{j,k}}{\sigma_{n,j}} \chi_{j,k}) \\
+ i \frac{\alpha_{j,k}}{\sqrt{2}} \tanh(\sqrt{2} | z_{1,1}(0) | \frac{\alpha_{j,k}}{\sigma_{n,j}} \psi_{j,k}) \\
\hat{\alpha}_{j,k}^{hard} = \frac{\alpha_{j,k}}{\sqrt{2}} (\text{sign}(\chi) + i \text{sign}(\psi))
\]

Figure 6: Linear, hyperbolic tangent and hard limiter decisions for real/imaginary parts of QPSK symbols.

4. SIMULATIONS

The \( K \) users of base station \( j \) use the same spreading factor and pass through the same downlink channel \( h_j \) which is a FIR filter, being the convolution of a sparse Vehicular A UMTS channel and a root-raised cosine pulse shape with roll-off factor of 0.22. The channel length is \( N = 19 \) chips at the UMTS chip rate of 3.84 Mchips/sec. An oversampling factor of \( m = 2 \) and one receive antenna \( q = 1 \) are used. User symbols are from a QPSK constellation. The user of interest has 10dB less power than the average interferer power, which represents a near-far situation. Fig. 7 and Fig. 8 show the SINR performance of the proposed receivers and the classical reference Rake receiver. As expected, they outperform the Rake receiver with high margins. Hard and hyperbolic tangent non-linearities give better results than the projector, hyperbolic tangent performing slightly better [5]. As for the linear filter after the subtractor, the MMSE filter which is constructed based on the residual interference powers from the two IC branches outperforms the Rake receiver.

5. REFERENCES