Linear Precoding for Spatial Multiplexing MIMO Systems: Blind Channel Estimation Aspects

Abdelkader Medles, Dirk T.M. Slock
Eurecom Institute[1]
2229 route des Crêtes, B.P. 193
06904 Sophia Antipolis Cedex
FRANCE

***Abstract* - For the case of white uncorrelated inputs, most of the blind multichannel identification techniques are not very robust and only allow to estimate the channel up to a number of ambiguities, especially in the MIMO case. On the other hand, all current standardized communication systems employ some form of known inputs to allow channel estimation. The channel estimation performance in those cases can be optimized by a semiblind approach which exploits both training and blind information. When the inputs are colored and have sufficiently different spectra, the MIMO channel may become blindly identifiable up to one constant phase factor per input, and this under looser conditions on the channel. For the case of spatial multiplexing, possible cooperation between the channel inputs allows for more complex MIMO source prefiltering that may allow blind MIMO channel identification up to just one global constant phase factor. We introduce semiblind criteria that are motivated by the Gaussian ML approach. They combine a training based weighted least-squares criterion with a blind criterion based on linear prediction. A variety of blind criteria are considered for the various cases of source coloring.**

I. Introduction

The multichannel aspect has led to the development of a wealth of blind channel estimation techniques over the last decade. In this paper, blind identification shall mean channel identification on the basis of the second-order statistics of the received signal. Consider linear digital modulation over a linear channel with additive Gaussian noise. Assume that we have p transmitters and $m > p$ receiving channels (e.g. antennas in BLAST or SDMA). For more details on the used notation in this paper refer to [1].

II. Blind Identification for Colored Inputs

In this section we aim to improve the Second-Order Statistics (SOS) based blind channel identification by exploiting correlation in the inputs. In the context of digital communications, the inputs are symbol sequences which are typically uncorrelated. Correlation can be introduced by linear convolutive precoding, which corresponds to MIMO prefiltering of the actual vector sequence $\mathbf{b}_k$ of symbols to be transmitted with a MIMO prefilter $\mathbf{T}(z)$ such that the transmitted vector signal becomes $\mathbf{a}_k = \mathbf{T}(q)\,\mathbf{b}_k$. In this paper we consider full rate linear precoding so that $\mathbf{T}(z)$ is a $p \times p$ square matrix transfer function (in [2] an example of low rate precoding appears since the same symbol sequence gets distributed over all TX antennas). We get for the transmitted signal spectrum $S_{\mathbf{aa}}(z) = \mathbf{T}(z)\,S_{\mathbf{bb}}(z)\,\mathbf{T}^\dagger(z) = \sigma_b^2\,\mathbf{T}(z)\,\mathbf{T}^\dagger(z)$ and for the received signal spectrum $S_{\mathbf{yy}}(z) = \mathbf{H}(z)\,S_{\mathbf{aa}}(z)\,\mathbf{H}^\dagger(z) + S_{\mathbf{vv}}(z) = \sigma_b^2\,\mathbf{H}(z)\,\mathbf{T}(z)\,\mathbf{T}^\dagger(z)\,\mathbf{H}^\dagger(z) + \sigma_v^2 I_m$. The choice of appropriate prefiltering, as we shall see below, may reduce the non-identifiability to a phase factor per source or even to a global phase factor. In the context of wireless communications, two scenarios may be distinguished:

Noncooperative scenario: this scenario corresponds to the multi-user case (on the transmitter side) without cooperation between users. We shall consider the simple case in which the users transmit through only one antenna. This noncooperative scenario can also arise in other source separation applications since natural sources tend to have different spectra. In this scenario, $\mathbf{H}(z)$ has no structure, other than possibly being FIR, and $\mathbf{T}(z)$ and $S_{\mathbf{aa}}(z)$ are diagonal. This scenario has been considered in [3],[4].

Cooperative scenario: this is the single-user spatial multiplexing case. In this case, since transmit antennas are near each other and also receive antennas are near each other, all (FIR) entries in $\mathbf{H}(z)$ have the same delay spread and hence are polynomial of the same order. $S_{\mathbf{aa}}(z)$ is allowed to be nondiagonal.

In the noncooperative case the channel will tend to be irreducible, a characteristic we have assumed so far, due to the fact that the users tend to be spread out in space. In the spatial multiplexing case however, in which the TX antennas are essentially colocated, the irreducibility of the channel depends on the richness of the scattering environment. In general, we need to consider a reducible channel. Such a channel can be factored as $\mathbf{H}(z) = \mathbf{G}(z)\mathbf{C}(z)$ where $\mathbf{G}(z)$ is irreducible and column reduced with columns in order of e.g. non-increasing degree. If r is the (generic) rank of $\mathbf{H}(z)$, then $\mathbf{G}(z)$ is $m \times r$ whereas $\mathbf{C}(z)$ is $r \times p$. In the noncooperative scenario $\mathbf{C}(z)$ has no particular structure. In the cooperative scenario however, all entries in a particular row of $\mathbf{C}(z)$ have the same degree and the degrees of the rows are non-decreasing (the degree profile of the rows in $\mathbf{C}(z)$ is complementary to the degree profile of the columns in $\mathbf{G}(z)$).

If $r \leq m-1$, then σ_v^2 is blindly identifiable from $S_{\mathbf{yy}}(z)$ and $\mathbf{G}(z)$ is blindly identifiable from the signal/noise subspaces of $S_{\mathbf{yy}}(z)$ up to a postmultiplication factor $\mathbf{L}(z)$ that is block lower triangular with block sizes according to the multiplicities of the degrees of the columns of $\mathbf{G}(z)$ and $\mathbf{L}(z)$ is also polynomial with the degree of block (i,j) being the difference between the degrees of block i and block j of the columns of $\mathbf{G}(z)$ [5]. So in particular, the diagonal blocks of $\mathbf{L}(z)$ are constant. Also, $\mathbf{L}^{-1}(z)$ has the same polynomial structure as $\mathbf{L}(z)$. In the cooperative case, $\mathbf{L}^{-1}(z)\mathbf{C}(z)$ has the same polynomial structure as $\mathbf{C}(z)$. If $r \leq m-1$, there are essentially no restrictions on the number of inputs p for identifiability. If $r = m$, identifiability of σ_v^2 becomes an issue and there's no longer a point in considering a factorization of $\mathbf{H}(z)$ for its identifica-

[1] Eurecom's research is partially supported by its industrial partners: Ascom, Swisscom, France Télécom, La Fondation CEGETEL, Bouygues Télécom, Thales, ST Microelectronics, Motorola, Hitachi Europe and Texas Instruments. The work reported here was also partially supported by the French RNRT project ERMITAGES.

tion.

Finally, let us note that TX pulse shape filters can be incorporated in $\mathbf{T}(z)$ or $S_{\mathbf{aa}}(z)$ and that oversampling at the RX also leads to an increase in the number of RX channels. Also, the formulation of complex quantities as a superposition of real quantities may lead to an extra MIMO dimension. In the next two sections we investigate channel identifiability with diagonal or full prefiltering $\mathbf{T}(z)$.

III. Noncooperative/Diagonal Prefiltering

In general, we would like to handle the reducible channel case. The rank r can be identified from $S_{\mathbf{yy}}(z)$. If $r \leq m-1$, then we can denoise the SOS and identify the factor $\mathbf{G}(z)$ from the subspaces. $\mathbf{G}(z)$ is unique up to a factor $\mathbf{L}(z)$. For whichever $\mathbf{G}(z)$ in this equivalence class, it remains to identify $\mathbf{C}(z)$ in $\mathbf{H}(z) = \mathbf{G}(z)\mathbf{C}(z)$ from

$$S(z) = \mathbf{G}^{\#}(z)\,(S_{\mathbf{yy}}(z) - \sigma_v^2 I_m)\,\mathbf{G}^{\#\,\dagger}(z) = \mathbf{C}(z)\,S_{\mathbf{aa}}(z)\,\mathbf{C}^{\dagger}(z) \tag{1}$$

where $\mathbf{G}^{\#}(z)$ is a MMSE ZF equalizer for $\mathbf{G}(z)$: $\mathbf{G}^{\#}(z)\,\mathbf{G}(z) = I_r$. For $r = m$, the problem is similar to the one in (1) with $\mathbf{C}(z)$ replaced by $\mathbf{H}(z)$ (apart from the σ_v^2 identification issue which will be discussed below). The value of the rank $r \in \{1, 2, \ldots, \min(m,p)\}$ is unpredictible in general. For a certain rank r, subsets of $r-1$ columns of $\mathbf{C}(z)$ could be identified jointly from $S(z)$ using certain $S_{\mathbf{aa}}(z)$ and under certain conditions on $\mathbf{C}(z)$ (or subsets of r columns under more stringent conditions on $\mathbf{C}(z)$). So to be general, $S_{\mathbf{aa}}(z)$ should be such that it allows identifiability for the worst case of r, which is $r = 1$. In that case, each column of $\mathbf{C}(z)$ needs to be identified separately. On the other hand, since in the case $r = 1$ each column of $\mathbf{C}(z)$ is a scalar FIR transfer function, only its minimum-phase equivalent is identifiable. So a column would be truly identifiable only if it is minimum-phase. To avoid having zeros would require to impose $r \geq 2$. In any case, to be fully general, it is desirable to have $S_{\mathbf{aa}}(z)$ such that it allows identifiability of each column of $\mathbf{C}(z)$ separately. So the MIMO problem gets converted into a set of disconnected SIMO ($r > 1$) or SISO ($r = 1$) problems. This will allow identifiability of each column up to a constant phase factor of the form $e^{j\theta}$ if the column has no maximum-phase zeros (which is quite possible if $r \geq 2$ but highly unlikely for $r = 1$). Another issue is the degree of $\mathbf{C}_j(z)$, column j of $\mathbf{C}(z)$. We have $\mathbf{H}_j(z) = \mathbf{G}(z)\mathbf{C}_j(z)$. The degree of $\mathbf{C}_j(z)$ is unpredictible and can be up to $N_j - 1$, the degree of the corresponding column $\mathbf{H}_j(z)$ of $\mathbf{H}(z)$. For identifiability, we need to consider the worst case and hence we shall assume that the degree of $\mathbf{C}_j(z)$ is $N_j - 1$. Of course, the N_j themselves may be unpredictible and in practice need to be replaced by an upper bound. We now consider two approaches for identification, leading to two classes of solutions for $S_{\mathbf{aa}}(z)$.

Frequency domain approach: The idea here is to introduce zeros into the diagonal elements of $\mathbf{T}(z)$ or hence $S_{\mathbf{aa}}(z)$ such that all other elements other than diagonal element j share N_j zeros

$$\mathbf{T}_{jj}(z) = \prod_{i=1,\neq j}^{p} \prod_{k=1}^{N_i} (1 - z_{i,k} z^{-1}) \tag{2}$$

This allows identifiability of $\mathbf{C}_j(z)$ from $S(z)$ up to a phase since

$$S(z_{j,k}) = \mathbf{C}_j(z_{j,k}) S_{a_j a_j}(z_{j,k}) \mathbf{C}_j^{\dagger}(z_{j,k}) \;,\quad k = 1, \ldots, N_j \tag{3}$$

where $S_{a_j a_j}(z) = \sigma_b^2 T_{jj}(z) T_{jj}^{\dagger}(z)$. If all N_j are equal (to N_1), then we can choose equispaced zeros on the unit circle: $\prod_{k=1}^{N_i}(1 - z_{i,k} z^{-1}) = 1 - e^{j\theta_i} z^{-N_1}$. If we furthermore choose overall equispacing by taking $\theta_i = (i-1)2\pi/pN_1$, then $T_{ii}(z) = \frac{1 - z^{-pN_1}}{1 - e^{j\theta_i} z^{-N_1}}$. In [4] very similar work appears in which the degree of $S_{\mathbf{aa}}(z)$ (in the case of equal N_j) is at least pN_1 (compared to $(p-1)N_1$ here), but the discussion in [4] is limited to the case $m > p = 2$. Note that here we can easily allow $p > m$ even (more inputs than outputs!). Remark also that by introducing a number of zeros in $\mathbf{T}(z)$, we can furthermore identify an equal number of noise parameters (such as σ_v^2 for instance when $r = m$). Non-FIR $S_{\mathbf{aa}}(z)$ can be considered also. For instance we can consider the case in which the $S_{a_j a_j}(e^{j2\pi f})$ have at least N_j disjoint expansion coefficients in some orthogonal basis. An extreme example of this would be $S_{a_j a_j}(e^{j2\pi f})$ that are bandlimited with the p bands being non-overlapping.

Time domain approach: The idea here is to introduce delay in the prefilter so that the correlations of each $\mathbf{C}_j(z)$ appear separately in certain portions of the correlation sequence of $S(z)$. This can be obtained for instance with

$$\mathbf{T}_{jj}(z) = 1 - \alpha_j z^{-d_j} \quad , \quad d_j = \sum_{i=1}^{j-1} N_i \quad . \tag{4}$$

Identification can be done with a correlation sequence peeling approach that starts with the last column $\mathbf{C}_p(z)$ of which the (1-sided) correlation sequence appears in an isolated fashion in the last N_p correlations of $S(z)$. Identification of $\mathbf{C}_p(z)$ from its correlation sequence can be done up to a phase factor $e^{j\theta_p}$ (and up to the phase of zeros if $\mathbf{C}_p(z)$ has zeros). We can then subtract $S_{a_j a_j}(z)\,\mathbf{C}_p(z)\mathbf{C}_p^{\dagger}(z)$ (which does not require $\mathbf{C}_p(z)$ but only its correlation sequence) from $S(z)$ which will then reveal the correlation sequence of $\mathbf{C}_{p-1}(z)$ in its last N_{p-1} correlations, etc. The degree of $S_{\mathbf{aa}}(z)$ is in this case the degree d_p of $S_{a_p a_p}(z)$ which, in the case of all equal N_j, is again $(p-1)N_1$, which leads to a degree of $pN_1 - 1$ for $S(z)$ or hence pN_1 correlations. Such a degree for $S_{\mathbf{aa}}(z)$ is not only sufficient but also necessary since when $r = 1$, there are pN_1 parameters to be identified for which indeed at least pN_1 correlations are needed. Note that in the temporal approach, increasing all the delays d_j with an amount D allows furthermore the (straightforward) identification of MA($D-1$) noise (e.g. $D = 1$ for white noise with arbitrary spatial correlation).

In practice, with estimated correlations, the correlation peeling approach leads to increasing estimation errors as the columns of $\mathbf{C}(z)$ get processed. This error increase can be avoided by doubling the delay separation between sources, which may furthermore lead to simpler algorithms (e.g. SIMO subspace fitting with asymmetric covariance matrices). Time domain approaches also appear in [3].

IV. Cooperative/Spatial-Multiplexing Prefiltering

Non-cooperative approaches can of course also be applied in the cooperative scenario, so diagonal prefiltering can be used for spatial multiplexing. However, this leads to at least a unknown phase per TX antenna and hence requires either differential encoding or training symbols per TX antenna. By applying full prefiltering, such that $S_{\mathbf{aa}}(z)$ is not blockdiagonal

in which case it is said to be fully diverse, the channel may possibly be identified up to a global phase factor only. Since better identifiability results in this case, better estimation may possibly be another consequence. We consider here linear precoding by time-invariant MIMO prefiltering. In [6], a block precoding approach is considered.

We can work with the eigen or LDU decompositions of $S_{\mathbf{aa}}(z)$. To begin with, consider the eigendecomposition $\mathbf{S_{aa}}(z) = \mathbf{V}(z)\,\mathbf{D}(z)\,\mathbf{V}^\dagger(z)$ where $\mathbf{V}(z)$ is paraunitary (i.e. $\mathbf{V}^\dagger(\mathbf{z})\mathbf{V}(\mathbf{z}) = \mathbf{I}$) and contains the eigenvectors as columns, and $\mathbf{D}(z)$ is diagonal with the diagonal elements, the eigenvalues, being valid scalar spectra.

A paraunitary matrix $\mathbf{V}(z)$ is said to be full diverse, if $\boldsymbol{P}\mathbf{V}(z)\,\mathbf{V}^\dagger(1)\boldsymbol{P}^T$ cannot be made block diagonal for any permutation $\boldsymbol{P}$.

Theorem 1: : An irreducible FIR MIMO channel is blindly identifiable up to a phase factor per user if $S_{\mathbf{aa}}(z)$ has distinct eigen value functions, and up to one global phase factor if its eigenvector matrix is fully diverse.

It may perhaps be more practical to work with the LDU (Lower triangular-Diagonal-Upper triangular) decomposition $\mathbf{S_{aa}}(z) = \mathbf{L}(z)\,\mathbf{D}(\mathbf{z})\,\mathbf{L}^\dagger(\mathbf{z})$ where $\mathbf{L}(z)$ is lower triangular with unit diagonal and the non-zero off-diagonal elements being unconstrained transfer functions, and $\mathbf{D}(z)$ is diagonal with the diagonal elements being valid scalar spectra. The relation between the LDU decomposition and the prefilter $\mathbf{T}(z)$ is immediate if $\mathbf{T}(z)$ is of the form $\mathbf{T}(z) = \mathbf{L}(z)\Delta(z)$ where $\Delta(z)$ is diagonal. An example of such a $\mathbf{T}(z)$ that allows irreducible channel identification up to one global phase factor is $\Delta(z) = I_p$ and $\mathbf{T}(z) = \mathbf{L}(z) = I_p + \widetilde{D}\,z^{-1}$ where $\widetilde{D}$ has only non-zero elements on the first subdiagonal and those elements are all different constants.

Stationary precoding can be generalized to cyclostationary precoding via periodically timevariant prefiltering. By stacking q consecutive symbol period quantities $\mathbf{y}_k$, $\mathbf{v}_k$, $\mathbf{a}_k$, $\mathbf{b}_k$, we obtain $\mathbf{Y}_k$, $\mathbf{V}_k$, $\mathbf{A}_k$, $\mathbf{B}_k$. We can then introduce a $pq \times pq$ LTI MIMO prefilter $\mathbf{T}(z)$ such that

$$\mathbf{Y}_k - \mathbf{V}_k = (I_q \otimes \mathbf{H}(q))\,\mathbf{A}_k = (I_q \otimes \mathbf{H}(q))\,\mathbf{T}(q)\,\mathbf{B}_k \quad . \quad (5)$$

Cyclostationary prefiltering introduces more information, hence should allow improved estimation (and possibly avoid stationary noise).

V. Gaussian ML SemiBlind Channel Identification

For Gaussian ML, which will allow to exploit the SOS, we model the unknown symbols as uncorrelated Gaussian variables whereas the known symbols $\mathbf{b}_k$ lead to a non-zero mean. By neglecting the non-stationarity due to the known symbols, the Gaussian likelihood function can be written in the frequency domain:

$$\oint [M \ln det(S_{\mathbf{yy}}(z)) + (\mathbf{y}(z) - \mathbf{H}(z)\mathbf{T}(z)\mathbf{b}_K(z))^\dagger S_{\mathbf{yy}}^{-1}(z)\,(\mathbf{y}(z) - \mathbf{H}(z)\mathbf{T}(z)\mathbf{b}_K(z))] \quad (6)$$

where $\oint$ is short for $\frac{1}{2\pi j}\oint \frac{dz}{z}$ and $\mathbf{y}(z)$, $\mathbf{b}_K(z)$ denote the z transforms of the signal of M samples $\mathbf{y}_k$ and the known symbols $\mathbf{b}_k$. The gradient of this criterion is the same as the gradient of the following sum of two subcriteria. The first subcriterion is

$$\oint (\mathbf{y}(z) - \mathbf{H}(z)\mathbf{T}(z)\mathbf{b}_K(z))^\dagger S_{\mathbf{yy}}^{-1}(z)\,(\mathbf{y}(z) - \mathbf{H}(z)\mathbf{T}(z)\mathbf{b}_K(z)) \quad (7)$$

which is a weighted LS criterion (quadratic in $\mathbf{H}(z)$) with the training information. The second subcriterion is

$$\text{tr} \oint \{S_{\mathbf{yy}}^{-1}(z)\widetilde{S}_{\mathbf{yy}}(z)S_{\mathbf{yy}}^{-1}(z)\widetilde{S}_{\mathbf{yy}}(z)\} \quad (8)$$

where $\widetilde{S}_{\mathbf{yy}}(z) = S_{\mathbf{yy}}(z) - \widehat{S}_{\mathbf{yy}}(z)$, $\widehat{S}_{\mathbf{yy}}(z) = \frac{1}{M}\mathbf{y}(z)\mathbf{y}(z)^\dagger$ (periodogram), and the gradient is taken by considering $S_{\mathbf{yy}}^{-1}(z)$ as constant. This second criterion is one of weighted spectrum matching and expresses the blind information. By taking the gradient of the sum of the two subcriteria, we combine training and blind information in an optimal fashion (compare to the CRB expression for GML). Asymptotically we can replace $S_{\mathbf{yy}}^{-1}(z)$ by a consistent estimate $\widehat{S}_{\mathbf{yy}}^{-1}(z)$ such as the periodogram. Also, we can replace the periodogram by an AR model which matches the covariance sequence estimate appearing implicitely in the periodogram (or asymptotically by a consistent AR model) such that $\mathbf{P}(z)\widehat{S}_{\mathbf{yy}}(z)\mathbf{P}^\dagger(z) = I$ so that $\widehat{S}_{\mathbf{yy}}^{-1}(z) = \mathbf{P}^\dagger(z)\mathbf{P}(z)$, where $\mathbf{P}(z)$ is the MIMO prediction error filter in which a square-root of the prediction error covariance matrix inverse has been absorbed. The blind subcriterion then becomes

$$\oint \left\| \mathbf{P}(z) S_{\mathbf{yy}}(z) \mathbf{P}^\dagger(z) - I_m \right\|_F^2 \quad (9)$$

which is of fourth order in $\mathbf{H}(z)$. One solution consists of interpreting $\mathbf{H}_1(z)$ and $\mathbf{H}_2(z)$ in $S_{\mathbf{yy}}(z) = \mathbf{H}_1(z)S_{\mathbf{aa}}(z)\mathbf{H}_2(z) + \sigma_v^2\, I_m$ as different quantities and performing alternating optimizations between them (and σ_v^2).

VI. Blind GML Channel Identification for a Flat Channel

Here we shall focus on the blind identification part for a frequency flat channel $\mathbf{H} = \mathbf{G}\,\mathbf{C}$ with $r < m$. We can take $\mathbf{G} = V_{\mathcal{S}}$, an orthonormal matrix spanning the signal subspace. We shall estimate first $V_{\mathcal{S}}$ and then $\mathbf{C}$.

Identification of the signal subspace $V_{\mathcal{S}}$:

We can alternatively estimate the noise subspace $V_{\mathcal{N}}$. Ideally, $\mathbf{R}_L\,(I_L \otimes V_{\mathcal{N}}) = 0$ where $\mathbf{R}_L$ is the denoised covariance matrix of L symbol periods of $\mathbf{y}_k$. We shall estimate $V_{\mathcal{N}}$ using a weighted LS criterion

$$\min_{V_{\mathcal{N}}: V_{\mathcal{N}}^H V_{\mathcal{N}} = I_{m-r}} \{vect(\widehat{\mathbf{R}}_L\,(I_L \otimes V_{\mathcal{N}}))\}^H \mathbf{W} \{vect(\widehat{\mathbf{R}}_L\,(I_L \otimes V_{\mathcal{N}}))\} \quad (10)$$

where $\widehat{\mathbf{R}}_L$ is the sample covariance matrix, $\widehat{\mathbf{R}}_L = \mathbf{R}_L + \widetilde{\mathbf{R}}_L$. The optimal weighting is $\mathbf{W} = \text{E}\, vect(\widehat{\mathbf{R}}_L\,(I_L \otimes V_{\mathcal{N}}))\}\{vect(\widehat{\mathbf{R}}_L\,(I_L \otimes V_{\mathcal{N}}))\}^H$. With $\widehat{\mathbf{R}}_L$ based on M samples, we get $\text{E}\,\{vect(\widetilde{\mathbf{R}}_L)\}\{vect(\widetilde{\mathbf{R}}_L)\}^H = \frac{1}{M}\mathbf{R}_L^T \otimes \mathbf{R}_L$. This allows us to work out the WLS criterion (10) to become

$$\min_{V_{\mathcal{N}}: V_{\mathcal{N}}^H V_{\mathcal{N}} = I_{m-r}} tr\,\{V_{\mathcal{N}}^H\,\mathbf{r}_{\mathbf{yy}}(0)\,V_{\mathcal{N}}\} \quad (11)$$

where $\mathbf{r}_{\mathbf{yy}}(0)$ is the estimated correlation matrix of $\mathbf{y}_k$ at lag 0. The solution is clearly given by the noise subspace of the matrix in the middle, so that $V_{\mathcal{S}}$ becomes its signal subspace.

Identification of $\mathbf{C}$:

By equalizing $V_{\mathcal{S}}$, we get $\mathbf{r}(i) = V_S^H\,\mathbf{r}_{\mathbf{yy}}(i)\,V_S = \mathbf{C}\,\mathbf{r}_{\mathbf{aa}}(i)\,\mathbf{C}^H$. We introduce a normalization so that

$$\mathbf{r}^{-1/2}(0)\,\mathbf{C}\,\mathbf{r}_{\mathbf{aa}}^{1/2}(0)\,\overline{\mathbf{r}}_{\mathbf{aa}}(i)\,\mathbf{r}_{\mathbf{aa}}^{H/2}(0)\,\mathbf{C}^H\,\mathbf{r}^{-H/2}(0) = \overline{\mathbf{r}}(i) \quad (12)$$

for $i = 0, 1, \ldots, L$, and where $\overline{\mathbf{r}}(i) = \mathbf{r}^{-1/2}(0)\mathbf{r}(i)\mathbf{r}^{-H/2}(0)$ and $\overline{\mathbf{r}}_{\mathbf{aa}}(i) = \mathbf{r}_{\mathbf{aa}}^{-1/2}(0)\mathbf{r}_{\mathbf{aa}}(i)\mathbf{r}_{\mathbf{aa}}^{-H/2}(0)$. From $i = 0$, we observe that $\mathbf{r}^{-1/2}(0)\,\mathbf{C}\,\mathbf{r}_{\mathbf{aa}}^{1/2}(0) = Q$ for some matrix Q with orthonormal rows, or hence $\mathbf{C} = \mathbf{r}^{1/2}(0)\,Q\,\mathbf{r}_{\mathbf{aa}}^{-1/2}(0)$. To find estimate Q, we shall assume $r = p$ so that Q is square and unitary, and we shall solve (12), which becomes $Q\,\overline{\mathbf{r}}_{\mathbf{aa}}(i)\,Q^H = \overline{\mathbf{r}}(i)$, for $i = i, \ldots, L$ in a least-squares sense:

$$\min_{Q:\,\mathrm{tr}\{Q^HQ\}=r} \sum_{i=1}^{L} \|Q\,\overline{\mathbf{r}}_{\mathbf{aa}}(i) - \overline{\mathbf{r}}(i)\,Q\|_F^2 \tag{13}$$

The solution of this problem involves an eigendecomposition and is unique in general up to a phase factor.

VII. Precoder optimization

We focus here on the flat channel case, and we study the optimization of the precoder to maximize the available capacity of the system, which is also the mutual information when the channel is estimated $\mathbf{I}_{S_{\boldsymbol{aa}}}(\mathbf{y};\mathbf{b}|\widehat{\mathbf{H}})$.

The ergodic capacity of an AWGN channel, when the channel knowledge is absent at the transmitter and perfect at the receiver is given by:

$$\begin{aligned} C(S_{\boldsymbol{aa}}) &= \mathrm{E}\mathbf{I}_{S_{\boldsymbol{aa}}}(\mathbf{y};\mathbf{b}|\mathbf{H}) \\ &= \mathrm{E}\oint \ln\det(I + \tfrac{1}{\sigma_v^2}\,\mathbf{H}\,S_{\boldsymbol{aa}}(z)\,\mathbf{H}^H) \end{aligned} \tag{14}$$

Where $S_{\boldsymbol{aa}}(z) = \sum_i^L \mathbf{r}_{\mathbf{aa}}(i)\,z^{-i}$, the expectation E is here w.r.t. the distribution of the channel. As in [7], we assume the entries $\mathbf{H}_{i,j}$ of the channel to be mutually independent, identically distributed zero mean complex Gaussian variables (Rayleigh flat fading MIMO channel model). Telatar has shown [7] that for such a channel model, the optimization of the capacity subject to the TX power constraint $\oint \mathrm{tr}(S_{\boldsymbol{aa}}(z)) \le N_{tx}\sigma_b^2$ leads to the requirement of a white (and Gaussian) vector transmission signal $S_{\boldsymbol{aa}}(z) = \sigma_b^2\,I$.

For a block of length M (sufficiently large for the frequency domain expressions to be sufficiently accurate), and using a precoding $S_{\boldsymbol{aa}}(z) = \sigma_b^2(I + \beta\,\Delta S(z))$ ($\beta << 1$, $\Delta S(z) = \sum_i^L \Delta S_i\,z^{-i}$ and $\oint \mathrm{tr}(\Delta S(z)) = \mathrm{tr}(\Delta S_0) = 0$ for power constraint), the deviation of the available capacity for the whole block is:

$$\begin{aligned} \Delta =&\; MC(\sigma_b^2 I) - M\,\mathrm{E}\,\mathbf{I}_{S_{\boldsymbol{aa}}}(\mathbf{y};\mathbf{b}|\widehat{\mathbf{H}}) \\ =&\underbrace{M\,(C(\sigma_b^2 I) - C(S_{\boldsymbol{aa}}))}_{\Delta_1} + \underbrace{M\,(C(S_{\boldsymbol{aa}}) - \mathrm{E}\,\mathbf{I}_{S_{\boldsymbol{aa}}}(\mathbf{y};\mathbf{b}|\widehat{\mathbf{H}}))}_{\Delta_2} \end{aligned} \tag{15}$$

Δ_1 represents the deviation (decrease) of the capacity due to the precoding, whereas Δ_2 represents the deviation caused by channel estimation errors.

Let $\rho = \frac{\sigma_b^2}{\sigma_v^2}$ and $R = I + \rho\mathbf{H}\,\mathbf{H}^H$, then:

$$\begin{aligned} \Delta_1 =&\; -M\,\mathrm{E}\oint \ln\det(I + \beta\rho R^{-1}\mathbf{H}\,\Delta S(z)\,\mathbf{H}^H \\ =&\; -M\,\mathrm{E}\oint[\beta\rho\,tr\{R^{-1}\mathbf{H}\,\Delta S(z)\,\mathbf{H}^H\} \\ &-\beta^2\rho^2\,tr\{R^{-1}\mathbf{H}\,\Delta S(z)\,\mathbf{H}^H R^{-1}\mathbf{H}\,\Delta S(z)\,\mathbf{H}^H\} + O(\beta^3)] \\ \approx&\; -M\beta\rho\,tr\{\mathrm{E}\mathbf{H}^H R^{-1}\mathbf{H}\Delta S_0\} \\ &+M\beta^2\rho^2\,\mathrm{E}\,tr\{\oint \mathbf{H}^H R^{-1}\mathbf{H}\,\Delta S(z)\,\mathbf{H}^H R^{-1}\mathbf{H}\,\Delta S(z)\} \end{aligned} \tag{16}$$

We will show below that $\mathrm{E}\,\mathbf{H}^H R^{-1}\mathbf{H}$ is a multiple of the unity matrix I. In fact, for every permutation matrix P and unitary diagonal matrix $D = diag(e^{j\psi_1}, \ldots, e^{j\psi_p})$, $\mathbf{H}' = \mathbf{H}PD$ has the same distribution as $\mathbf{H}$, then $\mathrm{E}\,\mathbf{H}^H R^{-1}\mathbf{H} = D^*P^H\mathrm{E}\,\mathbf{H}^H R^{-1}\mathbf{H}PD$. By averaging over the set of permutations $\mathbf{P}$ and phases $[0, 2\pi]^p$, we get:

$$\begin{aligned} \mathrm{E}\,\mathbf{H}^H R^{-1}\mathbf{H} =& \tfrac{1}{(2\pi)^p}\int \frac{1}{p!}\sum_{P\in\mathbf{P}} D^*P^H\mathrm{E}\mathbf{H}^H R^{-1}\mathbf{H}PD\,d\psi_1\ldots d\psi_p \\ =&\; \tfrac{tr\mathrm{E}\,\mathbf{H}^H\mathrm{R}^{-1}\mathbf{H}}{p}\,I \end{aligned} \tag{17}$$

The first term in Δ_1 is $-M\beta\rho\frac{tr\mathrm{E}\mathbf{H}^H\mathrm{R}^{-1}\mathbf{H}}{p}\,tr\Delta S_0 = 0$.

$$\begin{aligned} \Delta_1 =&M\beta^2\rho^2\textstyle\sum_{i,j}\mathrm{E}\,tr\{\oint z^{-(i+j)}\mathbf{H}^H R^{-1}\mathbf{H}\Delta S_i\mathbf{H}^H R^{-1}\mathbf{H}\Delta S_j\} \\ =&M\,\beta^2\,\rho^2\textstyle\sum_i \mathrm{E}\,tr\{\oint \mathbf{H}^H R^{-1}\mathbf{H}\,\Delta S_i\,\mathbf{H}^H R^{-1}\mathbf{H}\,\Delta S_i^H\} \\ =&M\,\beta^2\textstyle\sum_i vect^H(\Delta S_i)\mathbf{W}vect(\Delta S_i). \end{aligned} \tag{18}$$

where $\mathbf{W} = \rho^2\mathrm{E}(\mathbf{H}^H R^{-1}\mathbf{H})^T \otimes (\mathbf{H}^H R^{-1}\mathbf{H})$, $\sum_i$ is done over all index values i for which $\Delta S_i \neq 0$ and $\rho\mathbf{H}^H R^{-1}\mathbf{H} = (I + (\rho\mathbf{H}^H\mathbf{H})^{-1})^{-1}$ then for high SNR i.e $\rho >> 1$: $\mathbf{W} \approx I$.

As shown in [8], the deviation of the capacity for a small channel estimation error is $\Delta_2 = M\rho E\left\|\widetilde{\mathbf{H}}\right\|^2$. To minimize Δ_2 the channel estimator should be the MMSE estimator. The MMSE estimator doesn't achieve in general the Cramer-Rao Bound (CRB) but a good measure of the deviation is $\Delta_2 = M\rho\,tr\,\mathrm{E}\,CRB = M\rho\,tr\,\mathrm{E}\,J_{HH}^{\#}$, where J_{HH} is the Fischer Information Matrix.

Subspace fitting gives an estimation of the channel $\widehat{\mathbf{H}}'$ from the estimated correlation at lag 0 (see Section VI.) $\mathbf{r}_{\mathbf{yy}}(0) = \rho\mathbf{H}S_0\mathbf{H}^H + I$, which leads asymptotically to $\mathbf{H}' = \mathbf{H}S_0^{1/2}\mathbf{Q}^H$. The identification of the unitary matrix $\mathbf{Q}$ is done from the remaining correlation lag, $\mathbf{H}'^{\#}\mathbf{r}_{\mathbf{yy}}(i)\mathbf{H}'^{\#\,H} = \mathbf{Q}S_0^{-1/2}S_iS_0^{-H/2}\mathbf{Q}^H$; as a result we can assume without loss of generality that $S_0 = I$ and hence $\Delta S_0 = 0$. The Mean Square Error of the subspace estimator $\mathbf{H}'$ is of order M^{-1}.

$$\mathbf{H} = \mathbf{H}'\mathbf{Q} = \underbrace{\widehat{\mathbf{H}}'\widehat{Q}}_{=\widehat{\mathbf{H}}} + \underbrace{\widehat{\mathbf{H}}'\widetilde{\mathbf{Q}}}_{\propto g(\beta)/\sqrt{M}} + \underbrace{\widetilde{\mathbf{H}}'\mathbf{Q}}_{\propto 1/\sqrt{M}} \tag{19}$$

where $g(\beta)$ is an increasing function of β^{-1}. For large M and small β we can neglect the last term, and assume that $\widehat{\mathbf{H}}' = \mathbf{H}'$. The source of the channel error is then caused by the estimation error on $\mathbf{Q}$. If $\mathbf{X} = (\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_{n_p})$ is a real parameterization of the unitary matrix, i.e $\mathbf{Q} = \mathbf{Q}(\mathbf{X})$, then:

$$\mathbf{J}_{HH} = \frac{\partial\mathbf{X}}{\partial\mathbf{h}}\,J\!\!J_{XX}\left(\frac{\partial\mathbf{X}}{\partial\mathbf{h}}\right)^T \tag{20}$$

where $\mathbf{h} = [Re(vect(\mathbf{H}))^T, Im(vect(\mathbf{H}))^T]^T$ and $J\!\!J_{XX} = -\mathrm{E}_{\mathbf{Y}|\mathbf{H}'}\left(\frac{\partial\ln p(\mathbf{Y}|\mathbf{X},\mathbf{H}')}{\partial\mathbf{X}}\right)\left(\frac{\partial\ln p(\mathbf{Y}|\mathbf{X},\mathbf{H}')}{\partial\mathbf{X}}\right)^T$. An asymptotic expression of $J\!\!J_{XX}$ for large M is then :

$$J\!\!J_{XX}(i,j) = M\oint tr\{S_{\mathbf{yy}}^{-1}(z)\frac{\partial S_{\mathbf{yy}}(z)}{\partial\mathbf{x}_i}S_{\mathbf{yy}}^{-1}(z)\frac{\partial S_{\mathbf{yy}}(z)}{\partial\mathbf{x}_j}\} \tag{21}$$

By proceeding as for Δ_1, we can show that:

$$JJ_{XX}(i,j) = M\{\beta^2 \sum_k vect^H(\Delta S_k) W^J_{i,j} vect(\Delta S_k) + O(\beta^3)\} \quad (22)$$

where $W^J_{i,j} = \left(\frac{\partial \mathbf{Q}^* \otimes \mathbf{Q}}{\partial \mathbf{x}_i}\right)^H R'^{-T} \otimes R'^{-1} \left(\frac{\partial \mathbf{Q}^* \otimes \mathbf{Q}}{\partial \mathbf{x}_j}\right)$ and $R' = I + (\rho \mathbf{H}'^H \mathbf{H}')^{-1}$. We can note that for high SNR i.e $\rho >> 1$: $R' \approx I$. The decrease is now $\Delta_2 = \beta^{-2}\rho \,\mathrm{E}\, tr\left(\frac{\mathbf{J}_{HH}}{M\beta^2}\right)^{\#} = \beta^{-2}\rho \,\mathrm{E}\, tr\left(\frac{JJ_{XX}}{M\beta^2}\right)^{-1}\left(\left(\frac{\partial \mathbf{X}}{\partial \mathbf{h}}\right)^T \frac{\partial \mathbf{X}}{\partial \mathbf{h}}\right)^{-1}$.

The minimization of the capacity lost Δ can be performed by evaluating the expectation, which is quite complex, and then optimizing the resulting cost function. An alternative approach is to consider a close lower bound of Δ that is easier to calculate and to minimize. From the solution for $\Delta S(z)$, we can then find the solution for the precoder $\mathbf{T}(z)$. For a precoder $\mathbf{T}(z) = (1-\beta^2)^{1/2} I + \beta \sum_k M_F^{-1} F_k \, z{-}k$, where $M_F = (\sum_i F_i F_i^H)^{1/2}$ verifies $M_F M_F^H = \sum_i F_i F_i^H$ and $F_0 = 0$, the power spectrum is:

$$S_{\boldsymbol{aa}}(z) = \sigma_b^2 \{I + \beta(1-\beta^2)^{1/2} \textstyle\sum_k (M_F^{-1} F_k + F_{-k}^H M_F^{-H}) z^{-k} + \beta^2 \left(M_F^{-1} F(z) F^\dagger(z) M_F^{-H} - I\right)\} \quad (23)$$

This spectrum verifies the power constraint as $\Delta S_0 = 0$. The first order approximation is:

$\Delta S(z) = \sum_k \left(M_F^{-1} F_k + F_{-k}^H M_F^{-H}\right) z^{-k}$.

Example of optimization for $p = 2$:

We introduce the following parameterization of $\mathbf{Q}$: $\mathbf{X} = (\theta, \varphi, \psi) \in [0, \pi] \times [0, 2\pi]^2$, that allows to determine $\mathbf{Q}$ up to a global phase factor:

$$\mathbf{Q}(\theta, \varphi, \psi) = \begin{bmatrix} e^{j\psi} \cos\theta & e^{j\phi} \sin\theta \\ -e^{-j\phi} \sin\theta & e^{-j\psi} \cos\theta \end{bmatrix} \quad (24)$$

We consider here the case of high SNR, i.e $\rho >> 1$, in which case $\Delta_1 = 2M\beta^2 \sum_1^L ||\Delta S_i||^2$ and $R'^{-T} \otimes R'^{-1} = I$. To optimize Δ, we fix Δ_1 by fixing $\sum_1^L ||\Delta S_i||^2 = 1$, we optimize then Δ_2 under this last constraint, and given the precoder solution, optimize finally the overall decrease Δ with respect to the scale factor β. We will skip below several details of the derivation due to lack of place. First we establish the following equalities: $\left(\frac{\partial \mathbf{X}}{\partial \mathbf{h}}\right)^{\#} = \frac{\partial \mathbf{h}^T}{\partial \mathbf{X}^T} = \frac{\partial \mathbf{Q}_v^T}{\partial \mathbf{X}^T} \frac{\partial \mathbf{h}^T}{\partial \mathbf{Q}_v}$, where $\mathbf{Q}_v = [Re(vect(\mathbf{Q}))^T, Im(vect(\mathbf{Q}))^T]^T$.

Considering now the independence between $\mathbf{H}'$and Q we can write:

$$\begin{aligned} \Delta_2 &= \beta^{-2}\rho \,\mathrm{E}_{\mathbf{X}} tr\left(\tfrac{JJ_{XX}}{M\beta^2}\right)^{-1} \tfrac{\partial \mathbf{Q}_v^T}{\partial \mathbf{X}^T} \,\mathrm{E}_{\mathbf{H}'}\left\{\tfrac{\partial \mathbf{h}^T}{\partial \mathbf{Q}_v}\left(\tfrac{\partial \mathbf{h}^T}{\partial \mathbf{Q}_v}\right)^T\right\}\left(\tfrac{\partial \mathbf{Q}_v^T}{\partial \mathbf{X}^T}\right)^T \\ &= \beta^{-2}\rho \tfrac{E||H||^2}{p} tr\,\mathrm{E}_{\mathbf{X}}\left\{\tfrac{JJ_{XX}}{M\beta^2}\left(\tfrac{\partial \mathbf{Q}_v^T}{\partial \mathbf{X}^T}\left(\tfrac{\partial \mathbf{Q}_v^T}{\partial \mathbf{X}^T}\right)^T\right)^{-1}\right\}^{-1} \\ &= \beta^{-2}\rho \tfrac{E||H||^2}{p} C(\Delta S) \end{aligned} \quad (25)$$

where $\mathrm{E}_{\mathbf{X}}$ (resp. $\mathrm{E}_{\mathbf{H}'}$) denotes expectation with respect to $\mathbf{X}$ (resp. $\mathbf{H}'$), and to obtain the second expression we used arguments similar to the ones used to prove that $\mathrm{E}\,\mathbf{H}^H R^{-1} \mathbf{H} = \frac{tr \mathrm{E}\,\mathbf{H}^H \mathrm{R}^{-1} \mathbf{H}}{p} I$ and applied here to $\mathbf{H}'$ (assumed to have the same distribution as $\mathbf{H}$) in order to prove that: $\mathrm{E}_{\mathbf{H}'} \frac{\partial \mathbf{h}^T}{\partial \mathbf{Q}_v}\left(\frac{\partial \mathbf{h}^T}{\partial \mathbf{Q}_v}\right)^T = \frac{E||H||^2}{p} I$. The main difficulty is to minimize $C(\Delta S)$, which is possible directly but at high numerical cost. To avoid it, let's first note that as result of the convexity character of the function $f(x) = 1/x$ over $\mathbf{R}^{*+}$: $C(\Delta S) \geq C'(\Delta S) = tr\left\{\mathrm{E}_{\mathbf{X}} \frac{JJ_{XX}}{M\beta^2}\left(\frac{\partial \mathbf{Q}_v^T}{\partial \mathbf{X}^T}\left(\frac{\partial \mathbf{Q}_v^T}{\partial \mathbf{X}^T}\right)^T\right)^{-1}\right\}^{-1}$, then for ΔS_{opt} optimum of C', $C(\Delta S_{opt}) \geq C_{opt} \geq C'_{opt}$. If $\frac{C(\Delta S_{opt}) - C'_{opt}}{C(\Delta S_{opt})} << 1$ then we can conclude that ΔS_{opt} minimizes Δ_2. In our case $C'(\Delta S) = \frac{1}{4}\frac{5 + \frac{m_1}{m_2}}{m_1 + m_2}$, where $m_1 = \sum_{i=1}^L ||\Delta S_i(1,1) - \Delta S_i(2,2)||^2$ and $m_2 = \sum_{i=1}^L ||\Delta S_i(1,2)||^2 + ||\Delta S_i(2,1)||^2$, the optimal solution ($C'_{opt} = 1.125$) verifies: $\Delta S_i(2,2) = -\Delta S_i(1,1), i = 1, \ldots, L, m_2 = \frac{2}{3}$, and $\sum_{i=1}^L ||\Delta S_i(1,1)||^2 + ||\Delta S_i(1,1)||^2 = \frac{1}{3}$. One choice of a causal precoder with the smallest order L=1 that verifies these optimality conditions is:

$$\mathbf{T}_{opt}(z) = (1-\beta^2)^{1/2} I + \beta \mathbf{A} z^{-1},\ \mathbf{A} = \frac{1}{\sqrt{3}} \begin{bmatrix} \frac{1}{\sqrt{2}} & 1 \\ 1 & -\frac{1}{\sqrt{2}} \end{bmatrix} \quad (26)$$

In the first order of β this leads to $\beta \Delta S_{opt}(z) = \beta(1-\beta^2)^{1/2} \mathbf{A}\left(z^{-1} + z\right) + \frac{-\beta^2}{2} I \approx \beta \mathbf{A}\left(z^{-1} + z\right)$. By numerical evaluation we get: $\frac{C(\Delta S_{opt}) - C'_{opt}}{C(\Delta S_{opt})} = \frac{1.375 - 1.125}{1.375} = 0.18$, hence ΔS_{opt} is close to optimal. The optimization of β gives $\beta = \left\{\frac{\rho(E||H||^2) C(\Delta S_{opt})}{2Mp}\right\}^{1/4}$, usually the normalization is $\frac{E||H||^2}{p} = 1$, then $\beta = 0.91\left\{\frac{M}{\rho}\right\}^{-1/4}$ and $\frac{\Delta}{MC(\sigma_b^2 I)} = \frac{1.66(\rho M)^{1/2}}{MC(\sigma_b^2 I)} \propto \frac{1}{\ln \rho}\sqrt{\frac{\rho}{M}}$. The same ratio in the case of the exclusive use of a training sequence gives $\frac{\Delta}{MC(\sigma_b^2 I)} \propto M_{TS}^{-1} \frac{pm}{\ln \rho}$. To achieve the same performance, the length of the training sequence needs to be of the order $\sqrt{\frac{M}{\rho}} m$ which can become very important for large MIMO systems.

REFERENCES

[1] A. Medles, D.T.M. Slock and E. De Carvalho "Linear prediction based semi-blind estimation of MIMO FIR channels". In Proc. *Third Workshop on Signal Processing Advances in Wireless Communications (SPAWC'01)*, pp. 58-61, Taipei, Taiwan, March 2001.

[2] G. Leus, P. Vandaele, and M. Moonen. "Deterministic Blind Modulation-Induced Source Separation for Digital Wireless Communications". *IEEE Trans. on Signal Processing*, SP-49(1):219–227, Jan. 2001.

[3] J. Xavier, V.A.N. Barroso and J.M.F Moura. "Closed-Form Correlative Coding(CFC_2)Blind Identification of MIMO Channels: Isometry Fitting to Second Order Statistics". *IEEE Trans. Signal Processing*, SP-49(5):1073–1086, May 2001.

[4] Y. Hua and Y. Xiang. "The BID for Blind Equalization of FIR MIMO Channels". In *Proc. ICASSP Conf.*, Salt Lake City, USA, May 2001.

[5] A. Gorokhov and Ph. Loubaton. "Subspace based techniques for second-order blind separation of convolutive mixtures with temporally correlated sources". *IEEE Trans. Circuits and Systems*, 44:813–820, Sept. 1997.

[6] H. Hatés and F. Hlawatsch. "Blind Equalization of MIMO Channels Using Deterministic Precoding". In *Proc. ICASSP Conf.*, Salt Lake City, USA, May 2001.

[7] I.E. Telatar. "Capacity of Multi-antenna Gaussian Channels". *Eur. Trans. Telecom.*, vol. 10, no. 6, pp. 585-595, Nov/Dec 1999.

[8] A. Medles and D.T.M. Slock. "Semiblind channel estimation for MIMO spatial multiplexing systems". In *Proc. 35th Asilomar Conference on Signals, Systems and Computers*, Pacific Grove, CA, USA, Nov. 2001.