STATISTICAL PREFILTER DESIGN FOR MIMO ZF AND MMSE RECEIVERS
BASED ON MAJORIZATION THEORY

Mario Kiessling\textsuperscript{1,2}, Joachim Speidel\textsuperscript{1}

\textsuperscript{1}Institute of Telecommunications, University of Stuttgart, Germany
\textsuperscript{2}Siemens AG, Information and Communication Mobile, Ulm, Germany
mario.kiessling@gmx.net, speidel@inue.uni-stuttgart.de

ABSTRACT
Recently, the authors have proposed statistical prefilters for MMSE and ZF receivers that minimize SER. In this paper, we give a general derivation of their structure based on majorization theory. For both receiver types it is shown that the optimal prefilter essentially transmits on the strongest long-term eigenmodes of the channel with proper power allocation. Moreover, simple closed-form power allocation schemes are presented for Rayleigh and Ricean fading environments. Interestingly, while the statistical prefilters exhibit the same basic mathematical structure as their short-term counterparts, they require only statistical information of the correlation properties and the Rice component of the channel. Monte-Carlo simulations show that the proposed filters can achieve a considerable performance gain. Specifically, it is demonstrated that they can completely counteract the SER degradation due to a Ricean channel component.

1. INTRODUCTION
With the availability of short-term (ST) channel state information (CSI) at the transmitter, it is well-known that the symbol error rate (SER) of wireless MIMO systems can be reduced by adequate linear prefilters. Filter designs are available for zero-forcing (ZF) \cite{1} and minimum mean squared error (MMSE) \cite{2,3,4} receivers. However, in many cases it is impossible to acquire ST CSI at the transmitter, e.g. with high user mobility or in frequency division duplex (FDD) systems. Therefore, in \cite{8,9} the authors have proposed long-term (LT) statistical prefilters that are based on the correlation properties of Rayleigh fading MIMO channels only. In this paper, we extend these results by providing exact proofs of the optimal prefilter structures based on majorization theory \cite{5}. It is shown that in both the ZF and MMSE case, the prefilter transmits independent data streams along the strongest LT eigenmodes of the channel with a suitable power allocation (PA). On the other hand, the optimum PA strategy in general is a complicated function of the prevailing fading statistics of the wireless channel and becomes hardly analytically tractable. One possible remedy to this problem is a costly numerical Monte-Carlo optimization. We avoid this by introducing low-complexity closed-form power allocation policies for Rayleigh fading environments based on bounding techniques, which are also effectively deployed after simple manipulations in the presence of Ricean fading. Interestingly, there is a duality between the prefilter structures in the ST and LT case, i.e. it turns out that the LT prefilters can be derived by simply replacing the ST eigenvectors and eigenvalues of the channel by their LT equivalents. Monte-Carlo simulations demonstrate the effectiveness of the proposed prefilter schemes and a significant improvement in the symbol error rate (SER).

2. SIGNAL AND CHANNEL MODEL
We consider a flat fading MIMO link modeled by

\begin{equation}
y = Hs + n,
\end{equation}

where \( s \) is the \( L \times 1 \) TX symbol vector, \( F \) is a \( M_{\text{TX}} \times L \) linear matrix transmit prefilter, \( H \) is the \( M_{\text{RX}} \times M_{\text{TX}} \) MIMO channel matrix with correlated fading elements, \( n \) is the \( M_{\text{RX}} \times 1 \) noise vector, and \( y \) is the \( M_{\text{RX}} \times 1 \) receive vector. By \( M_{\text{RX}} \geq L \) we denote the number of RX antennas, \( M_{\text{TX}} \) is the number of TX antennas and \( L \) is the number of independent subchannels. Note that \( L \) can in general be smaller than the number of transmit antennas. In the following we assume additive Gaussian noise with covariance matrix \( R_{\text{nn}} \).

Using a widely accepted channel model, the Rayleigh fading MIMO channel with TX correlation can be described by the matrix product

\begin{equation}
H_{\text{Ray}} = H_{\text{w}}B,
\end{equation}

where \( H_{\text{w}} \) is a \( M_{\text{RX}} \times M_{\text{TX}} \) matrix of complex i.i.d. Gaussian variables of unity variance and

\begin{equation}
BB^{H} = B^{H}B = R_{\text{TX}} = \frac{1}{M_{\text{TX}}}E[H_{\text{w}}^{H}H_{\text{Ray}}],
\end{equation}

whereas we assume without loss of generality that \( B \) is Hermitian and \( R_{\text{TX}} \) is the long-term stable (normalized) transmit correlation matrix, respectively. \( E \) denotes expectation with respect to \( x \). In the presence of a deterministic component, i.e. Ricean fading, the channel model in (2) can be extended to (see e.g. \cite{6})

\begin{equation}
H_{\text{Rice}} = \frac{1}{\sqrt{1 + K}}H_{\text{w}}B + \sqrt{K/(1 + K)}M,
\end{equation}

where \( M \) is the non-fading component, normalized such that \( \text{tr}(M^{H}M) = M_{\text{Rice}} \cdot M_{\text{TX}} \), and \( K \) is the Ricean factor ranging from 0 (purely stochastic) to infinity (purely deterministic).

3. ZF STATISTICAL PREFILTER DESIGN
Using majorization theory, we calculate a general solution of the prefilter optimization problem for Rayleigh fading channels given...
in [9]. Via a simple moment fit the result is then extended to Ricean fading.

3.1. Rayleigh fading

By considering the Chernoff bound (valid for the high SNR region) on the symbol error rate (SER), in [9] the authors have derived that the SER optimizing linear prefilter matrix \( F \) for a ZF receiver with transmit correlated Rayleigh fading channel and additive white Gaussian noise (AWGN) can be obtained by solving the following constrained optimization problem

\[
F_{opt} = \arg\min_F \frac{1}{L} \sum_{k=1}^{L} \text{tr}( (F^H R_{TX} F)^{-1} )_{kk} \quad \text{s.t. } \text{tr}(FF^H) = \rho \numberTag{5}
\]

with the diversity parameter \( N = M_R X L + 1 \) and transmit power restriction \( \rho \). Using a direct differentiation approach, in [9] a general solution for arbitrary system parameters could not be found. This gap is now closed by

Theorem 1. The optimum SER minimizing prefilter with ZF receiver in the sense of (5) is given by

\[
F_{opt} = \Phi \Phi^H D_L,
\]

where \( D_L \) is a \( L \times L \) discrete Fourier transform (DFT) matrix, \( \Phi \) is a matrix of the eigenvectors corresponding to the \( L \) strongest eigenvalues of the diagonalization decomposition (EVD) in

\[
R_{TX} = [\begin{bmatrix} \Lambda \end{bmatrix}^H \begin{bmatrix} \Lambda \end{bmatrix}] [\begin{bmatrix} \Lambda \end{bmatrix}^H \begin{bmatrix} \Lambda \end{bmatrix}],
\]

where the matrix \( \Lambda \) contains the \( L \) largest eigenvalues in increasing order. The diagonal power allocation matrix \( \Phi \) reads

\[
\Phi = \left( \frac{\rho}{\text{tr}(\Lambda^{-1/2})} \right)^{1/2} \Lambda^{1/4}.
\]

Proof: In this paper, we use majorization theory for solving problem (5). To this end, we first introduce the auxiliary \( L \times L \) matrix

\[
X = (F^H R_{TX} F)^{-1}
\]

with the vector of real diagonal elements

\[
x = \text{diag}(X) = [x_1(F), x_2(F), \ldots, x_L(F)].
\]

Without loss of generality, assume that the elements of \( x \) are arranged in decreasing order. Now we can reformulate the problem in (5) as

\[
F_{opt} = \arg\min_F \frac{1}{L} \sum_{k=1}^{L} x_k(F)^N \quad \text{s.t. } \text{tr}(FF^H) = \rho \numberTag{6}
\]

Note that \( F^H R_{TX} F \) is a positive definite matrix, i.e., all minors are positive definite and thus the diagonal of the inverse is positive. Now consider the objective function in (11). Obviously, it is symmetric in its arguments. Furthermore, it is convex in each of its arguments. It follows from Theorem 4 in the appendix that it is a Schur-convex function. Using Theorem 3, the objective function is minimized for equal elements in \( x \). From Theorem 6, we can find a real symmetric matrix \( Q \) such that \( Q^{1/2} X \) has identical diagonal elements, i.e., we use prefilter matrix

\[
F = \hat{F} Q.
\]

Note that this does not change the transmit power. Another straightforward choice for the matrix \( Q \) is a DFT matrix of size \( L \times L \). Using the transmit filter structure in (12), we find

\[
[(F^H R_{TX} F)^{-1}]_{kk} = \frac{1}{L} \text{tr}((F^H R_{TX} F)^{-1}.
\]

Obviously, with (13) the optimization problem in (5) can be reduced to (minimum trace minimizes each diagonal element)

\[
F_{opt} = \arg\min_F \text{tr}((F^H R_{TX} F)^{-1} \quad \text{s.t. } \text{tr}(FF^H) = \rho \numberTag{14}
\]

Now (14) is a Schur-concave function of the diagonal elements of \( (F^H R_{TX} F)^{-1} \). Applying Theorem 5 we can choose \( F \) by proper application of a rotation matrix such that \( (F^H R_{TX} F)^{-1} \) is diagonal with elements in decreasing order or equivalently such that the diagonal elements of \( F^H R_{TX} F \) are in increasing order. The diagonalizing rotation matrix consists of \( L \) eigenvectors of \( R_{TX} \).

It can then be shown by Lemma 2 that we have to chose the eigenvectors corresponding to the largest \( L \) eigenvalues of \( R_{TX} \), such that the optimum \( F \) has the structure in (6). Constrained Lagrange optimization of the problem in (14) then leads to (see also [9]) the power allocation matrix in (8). QED.

3.2. Ricean fading

The SER performance analysis that led to the optimum prefilter design problem in (5) becomes extremely complicated in the case of Ricean fading and to the author’s best knowledge there are no solutions available in literature. Therefore, we propose an approximation of the Ricean fading statistics (essentially, we approximate a non-central Wishart distribution via a central Wishart distribution [7]) by a moment fit

\[
E_{H_1} [H_{Ray}^H H_{Ray}] = E_{H_1} [H_{Wish}^H H_{Wish}],
\]

where \( H_{Ray} \) is the Rayleigh fading channel matrix approximating the Ricean statistics. Calculating the expected values in (15) and equating results in an approximating transmit covariance of \( H_{Ray} \)

\[
\tilde{R}_{TX} = \frac{1}{1 + K} R_{TX} + K \frac{1}{1 + K} \cdot M^H M.
\]

Using the approximation in (16), for the prefilter design in Ricean fading we can just use equations (6)(7)(8), with \( R_{TX} \) replaced by \( \tilde{R}_{TX} \). We note that the proposed scheme is applicable to arbitrary MIMO channel statistics.

4. MMSE STATISTICAL PREFILTER DESIGN

4.1. Rayleigh fading

The resulting average mean squared error (MSE) summed over all subchannels with the signal and channel models in (1) and (2) and MMSE receiver reads [8]

\[
\varepsilon(F) = E_H [\text{tr}((F^H B H^H H^H R_{TX} H^H H_{Ray} B F + I)^{-1})].
\]

Now we have the optimization problem for the prefilter (such that it minimizes the average MSE)

\[
F_{opt} = \arg\min_F \varepsilon(F) \quad \text{s.t. } \text{tr}(FF^H) = \rho \numberTag{18}
\]

The solution is given in

Theorem 2. The optimum prefilter with MMSE receiver in the sense that it minimizes the average MSE and forces equal SER on each subchannel in a Rayleigh fading environment is given by

\[
F_{opt} = \Phi \Phi^H D_L.
\]
with diagonal PA matrix $\Phi$, L×L DFT matrix $\mathbf{D}_t$, and the EVD in (7).

Proof: Introducing the singular value decomposition (SVD)

$$BF = YDZ^H,$$

where $Y$ contains the left singular vectors, $Z$ the right singular vectors and $D$ the singular values we get

$$e(D) = E_{H_1}[\text{tr}((DH_1^HR_0^H)^{-1}D + I)^{-1}].$$

with a $M_{RX} \times L$ matrix of complex Gaussian i.i.d. entries $H_1$. Obviously, the average MSE is a function of the singular values in $D$ only. Now noting that left multiplication of $F$ with unitary $Z$ does not change the objective function and the constraint in (18), we can introduce

$$\tilde{F} = FZ \Leftrightarrow \tilde{F}Z^H = F$$

and solve problem (18) with $F$ being replaced by $\tilde{F}$. We can arbitrarily chose the unitary matrix $Z$. Here, we let $Z^{H}=\mathbf{D}_b$ be a DFT matrix for minimizing the overall SER [3]. We find from (20) and (22)

$$\tilde{F}^H B^\dagger \tilde{F} = \tilde{F}^H R_{TX} \tilde{F} = D^2.$$  

Finally, we can apply Lemma 2 and arrive at (19). QED.

Theorem 2 describes only the general structure of the prefilter and we still have to define the PA matrix $\Phi$. To this end, let

$$E[H_{RX}^HR_{RX}^{-1}H_{RX}] = \text{tr}(R_{RX}^{-1}) \cdot R_{TX} = \left[ \begin{array}{c c} \mathbf{A}_b & \mathbf{A}_d \end{array} \right] \left[ \begin{array}{c c} \mathbf{A}_b^H & \mathbf{A}_d^H \end{array} \right]^H.$$ (24)

We note that in (24) obviously $\mathbf{V}_b \mathbf{V}_b^H = \mathbf{F} \tilde{F}^H$ defined in (7). Via a bound on the average MSE and Lagrange optimization, in [8] a low complexity long-term power allocation (PA) policy was proposed with

$$\Phi = (\mu^{-1/2} \mathbf{A}_b^{-1/2} - \mathbf{A}_d^{-1})^{1/2},$$  

where the constant $\mu$ is chosen according to the power constraint, resulting in

$$\mu^{1/2} = \frac{\text{tr}(\mathbf{A}_d^{1/2})}{\text{tr}(\mathbf{A}_b^{1/2}) + \rho}.$$ (26)

We have to assure $\rho > 0$ for all $l$, which is indicated by the plus sign in (25). This means that in certain situations the weakest LT eigenmodes are not used for transmission by setting the corresponding PA coefficient to 0.

4.2 Ricean fading

Similar to (24), we now calculate for the Ricean channel

$$E[H_{Rice}^HR_{Rice}^{-1}H_{Rice}] = \frac{\text{tr}(R_{RX}^{-1}) \cdot R_{TX} + K \cdot M^H R_{RX}^{-1} M}{1 + K}.$$ (27)

With the corresponding EVD of (27), we can use the prefilter structure in (19) together with the power allocation (25) also for a Ricean fading environment. We note that the resulting prefilter is optimum (i.e. it is no approximation) in the sense that it minimizes the bound on the average MSE given in [8] based on Jensen’s inequality.

5. SIMULATION RESULTS

In the following simulations, we assume a uniform linear array with 0.5 lambda element spacing at the transmitter and receiver. In the presence of transmit fading correlation, for the Rayleigh fading scattering component of the channel we consider the case of one main direction of departure (DOD) at 20 degrees with respect to the array perpendicular, and a Laplacian power distribution with 10 degrees angular spread (AS). The transmit correlation matrix is determined according to these assumptions via Monte-Carlo simulation, while uncorrelated fading is assumed at the receiver.

On the other hand, we consider the case of a rank-1 deterministic channel component $M$ corresponding to a single line-of-sight (LOS) component, whereas we assume a DOD at the transmitter and a DOA at the receiver of 45 degrees. We study an AWGN (i.e. no colored interference) scenario with the SNR given by

$$SNR = 10 \cdot \log_{10} \left( \frac{M_{RX} \cdot E_b}{N_0} \right) \quad (dB),$$

where $E_b$ is the transmit energy per information bit. Throughout our simulations we normalize the total transmitted energy to $\rho = M_{TX}$ and assume QPSK modulation.

In Fig. 1 we show BER results of a ZF receiver in an uncorrelated Ricean fading channel with $K=1$ (equal split between determinis-tic and scattering component) for a 4×4 system with L=2 independent data streams. For reference we have also depicted results for an i.i.d. Rayleigh fading channel ($M=0$) with blind transmission and results for a prefilter algorithm based on ST CSI [1].
blind transmission in a Ricean correlated channel compared to
blind transmission in an i.i.d. Rayleigh fading channel with
(note that the 4 independent data streams are transmitted on
the outer 4 transmit antennas). Obviously, via the LT CSI based pre-
filter we can successfully exploit the knowledge of the channel
statistics and one can see a considerable performance improve-
ment.

Definition 1. For any \( x \in \mathbb{R}^n \), let
\[
x_1 \geq \ldots \geq x_n \tag{29}
\]
denote the elements of \( x \) in decreasing order (also termed order
statistics).

Definition 2. [5, definition 1.A.1] Let \( x, y \in \mathbb{R}^n \). Vector \( x \) is
majorized by vector \( y \) (\( y \) majorizes \( x \)) if
\[
\sum_{i=1}^{k} x_i \leq \sum_{i=1}^{k} y_i \quad \forall 1 \leq k \leq n-1 \quad \wedge \quad \sum_{i=1}^{n} x_i = \sum_{i=1}^{n} y_i \tag{30}
\]
and is represented by \( x \prec y \).

Definition 3. [5, definition 3.A.1] A real-valued function \( \phi \)
defined on a set \( A \subseteq \mathbb{R}^n \) is said to be Schur-convex
on \( A \) if
\[
x \prec y \quad \Rightarrow \quad \phi(x) \leq \phi(y) \tag{31}
\]
Similarly, \( \phi \) is said to be Schur-concave on \( A \) if
\[
x \succ y \quad \Rightarrow \quad \phi(x) \leq \phi(y) \tag{32}
\]

Theorem 4. [5, proposition 3.C.2] If \( \phi \) is symmetric (i.e. symmetric
in its arguments) and convex, then \( \phi \) is Schur-convex.

Theorem 5. [5, theorem 9.B.1] Let \( R \) be an \( n \times n \) Hermitian matrix
with diagonal elements denoted by the vector \( d \) and eigenvalues
denoted by the vector \( \lambda \), then
\[
d \prec \lambda. \tag{34}
\]

Theorem 6. [5, theorem 9.B.2] If \( h \prec \lambda \) on \( \mathbb{R}^n \), then there exists a
real symmetric matrix \( H \) with diagonal elements \( h \) and characteristic
roots \( \lambda \).

Lemma 1. [5, lemma 9.H.1.h] If \( U \) and \( V \) are \( n \times n \) positive
semidefinite Hermitian matrices, then
\[
\text{tr}(UV) \geq \sum_{i=1}^{n} \lambda_i(U) \cdot \lambda_{n-i+1}(V), \tag{35}
\]

where the eigenvalues are sorted in decreasing order.

Lemma 2. [4, lemma 12 in appendix] Let \( R \) be a full rank positive
semidefinite Hermitian matrix with eigenvalue decomposition
\[
R = \begin{bmatrix} U & \tilde{U} \\ \tilde{L} & \tilde{L} \end{bmatrix}, \tag{36}
\]

where the matrix \( L \) contains the \( n \) largest eigenvalues in increasing
order and \( U \) the corresponding eigenvectors. Given a \( M_{TX} \times L \)
(\( M_{TX} > L \)) matrix \( B \) such that \( BB^H \) is diagonal with diagonal
elements in increasing order, it is always possible to find a matrix \( \tilde{B} \)
of the form \( B = US \) with diagonal matrix \( S \), such that it satisfies
\[
BB^H = BB^H \ewith \text{tr}(BB^H) \leq \text{tr}(BB^H). \tag{37}
\]

7. REFERENCES


process \( 51, no. 9, Sept. 2003 \)