LINEAR PRECODING FOR TIME-VARYING MIMO CHANNELS WITH LOW-COMPLEXITY RECEIVERS

Jun Tong*, Peter J. Schreier‡, Steven R. Weller*, and Louis L. Scharf†∗

*School of Elec. Eng. & Comp. Sci., University of Newcastle, Australia, {jun.tong, steven.weller}@newcastle.edu.au
‡Signal and System Theory Group, Universität Paderborn, Germany, peter@peter-schreier.com
†Dept of Elec. & Comp. Eng., Colorado State University, scharf@engr.colostate.edu

ABSTRACT

This paper considers linear precoding for time-varying multiple-input multiple-output (MIMO) channels. We show that linear minimum mean-squared error (LMMSE) equalization based on the conjugate gradient (CG) method can result in significantly reduced complexity compared with conventional approaches. This reduction is achieved by incorporating a condition number constraint into the precoder optimization framework, which leads to clustered eigenvalues of the measurement covariance matrix. The cost is a small increase in MSE compared to the optimal precoder.

Index Terms—Condition number, conjugate gradient (CG), LMMSE, optimization, precoding.

1. INTRODUCTION

Linear precoding for multiple-input multiple-output (MIMO) channels has been studied extensively to improve power and spectral efficiency [1]-[4]. The potential gains may be realized by applying linear minimum mean-squared error (LMMSE) equalization at the receiver, which involves solving systems of linear equations. The direct implementation of LMMSE equalization has cubic complexity, which is computationally demanding when the dimension of the system is large. For time-invariant channels, the LMMSE equalizer can be reused for a number of transmission blocks, so the complexity is not a serious concern in this case. When the channel is time-varying, however, the equalizer must be recomputed frequently [3], and the complexity of equalization may thus grow significantly.

In this paper, we show that reduced-rank LMMSE equalization based on the conjugate gradient (CG) method [5] can reduce the complexity. By avoiding matrix-matrix products and matrix factorizations, each of which involves cubic complexity, the CG receiver can be built with only quadratic complexity. This is particularly useful for a time-varying environment where the equalizer must be updated in a timely fashion. We show in this paper that further complexity reduction can be achieved by improving the convergence rate of the CG algorithm. This is done by designing the precoder so that the eigenvalues of the measurement covariance matrix are clustered. The precoder is obtained by incorporating a condition number constraint into the precoding optimization framework. The condition number constraint is original to the literature of precoder and equalizer designs. The resulting optimization problem can be solved at low complexity by using majorization theory [4, 6] and convex optimization techniques [7]. The overall complexity of the system is thus significantly reduced compared with conventional approaches, at the price of a small increase in MSE.

2. LINEAR PRECODER DESIGN

2.1. Conventional Design

Consider a precoded block transmission system over a time-varying channel, where each frame consists of Q blocks precoded using the same precoder. Let q be the block index. The signal model is

\[ y_q = H_q S b_q + n_q, \]

where \( y_q \in \mathbb{C}^{N \times 1} \) is the received signal vector for block q; \( H_q \in \mathbb{C}^{N \times M} \) is the channel matrix; \( S \in \mathbb{C}^{M \times M} \) is the precoding matrix, which is shared by Q blocks; \( b_q \in \mathbb{C}^M \) contains the complex proper information symbols with \( E[b_q] = I_M \) and \( E[b_q^* b_q] = 0 \); \( \mathcal{X} \) is a rotationally symmetric constellation; \( (\cdot)^\dagger \) denotes conjugate transpose; and \( n_q \in \mathbb{C}^{N \times 1} \) is zero-mean complex proper noise with \( E[n_q n_q^*] = \sigma_n^2 I_N \) and \( E[n_q n_q^*] = 0 \). Without loss of generality, we assume that the number of transmit symbols in each block is no larger than the dimension of the received signal, i.e., \( M \leq N \), and the noise has unit average power, i.e., \( \sigma_n^2 = 1 \).

The generic linear model in (1) can be used to describe multiuser, multi-antenna systems over multipath channels with either single- or multi-carrier modulation [2, 4]. Define \( R_q = I_M + S^\dagger H_q^\dagger H_q S \). The LMMSE equalizer estimates \( b_q \) as

\[ \hat{b}_q = R_q^{-1} S^\dagger H_q^\dagger y_q = (I_M + S^\dagger H_q^\dagger H_q S)^{-1} S^\dagger H_q^\dagger y_q, \]

where we have used the Woodbury identity [7]. The MSEs for estimating the \( M \) symbols [1, 4] are given by the diagonal entries of

\[ E_q = R_q^{-1} = (I_M + S^\dagger H_q^\dagger H_q S)^{-1}. \]

The precoder \( S \) for a given frame is designed according to CSI \( H_0 \), which is assumed perfectly known at the last block of the previous frame. Let \( \xi_m \) be the mth diagonal entry of \( E_0 \). In [4], the minimization of a cost function, such as the sum MSE, under a total power constraint is considered:

\[ \min_S f(\{\xi_m\}), \text{ s.t. } \text{tr}(S^\dagger S) \leq P_0. \]

Employing majorization theory, this optimization problem can be cast in convex form and solved efficiently [4]. This precoding scheme assumes LMMSE equalization at the receiver. For time-invariant channels, the equalizer can be reused by a large number of blocks and so receiver complexity may not be a se-
Initialization: $\hat{b}_q^{(0)} = 0$, $d^{(1)} = r^{(1)} = y_q$.

Recursions: for $i = 1, 2, \ldots, D$

\[ \alpha^{(i)} = \frac{\langle r^{(i)} \rangle}{\langle dd^{\dagger} \rangle^{\frac{1}{2}}}, \]

\[ \beta^{(i+1)} = \beta^{(i)} + \frac{\langle d r \rangle^{\dagger}}{\langle d d^{\dagger} \rangle^{\frac{1}{2}}} \]

The CG method is an alternative to the Cholesky factorization method. It is equivalent to the multistage Wiener filter [8]-[10]. Let $y_q = S^H H_0^\dagger y_q$. The CG algorithm, adapted to the application at hand, is listed in Fig. 1. Note that the matrix-vector product $(I_M + S^H H_0^\dagger H_q S) d^{(i)}$ can be computed using four matrix-vector products as $d^{(i)} + (S^H (H_q^\dagger (H_q (S d^{(i)}))))$ without explicitly computing $R_q = I_M + S^H H_0^\dagger H_q S$. Since matrix-matrix products and Cholesky factorizations are avoided, the complexity of equalization can be reduced from cubic to quadratic. This can already result in a significant reduction in complexity. Furthermore, the CG method is straightforward to implement in parallel as a matrix-vector product is equivalent to multiple inner products, in contrast to the Cholesky factorization method, which involves recursive operations.

We now consider a further reduction of the CG complexity, which increases linearly with the number of iterations $D$. It is known that convergence of the CG algorithm is fast if the eigenvalues of $R_q$ are clustered [5, 11]. A useful parameter to quantify such clustering is the condition number

\[ \kappa(R_q) \triangleq \frac{\max\{\lambda(R_q)\}}{\min\{\lambda(R_q)\}}, \]

where $\{\lambda(R_q)\}$ denote the eigenvalues of $R_q$. Here, we extend the work in [1]-[4] by imposing a condition number constraint for the precoder optimization, such that the convergence of the CG method can be accelerated and the required $D$ be reduced. Following [3], we assume a fixed channel $H_0$ while designing the precoder. We consider Schur-convex cost functions $f(\{\xi_m\})$, covering a variety of problems (e.g., the minimization of the bit-error-rate (BER)) but the design can also be extended to Schur-concave cost functions. Let $V_H$ be the right singular matrix of $H_0$, i.e.,

\[ H_0 = U_H \text{diag}(\sqrt{\gamma}) V_H^\dagger, \]

where $\gamma = \{\gamma_m\}$ are the eigenvalues of $H_0^\dagger H_0$.

Lemma 1: Consider the optimization problem

\[ \min_{\xi} f(\{\xi_m\}), \quad \text{s.t. } \text{tr}(S^T S) \leq P_0, \kappa(R_0) \leq \kappa_0, \]

where $f(\cdot)$ is Schur-convex and increasing, and $\kappa_0 \geq 1$ is a desired condition number. The solution has the form

\[ S^* = Q \Pi, \]

where $\Pi$ is a unitary matrix which ensures that the MSE matrix $E_0$ of (3) has equal diagonal entries,

\[ Q = V_H \text{diag}(\sqrt{\beta^*}), \]

and $\beta^* = [p_1^*, p_2^*, \ldots, p_M^*]$ is the solution to

\[ \min_{\beta} \sum_{m=1}^M \frac{1}{1 + \gamma_m p_m}, \quad \text{s.t. } \max\{1 + \gamma_m p_m\} \leq \kappa_0, \]

\[ 1^T p \leq P_0, p_m \geq 0, m = 1, \ldots, M, \]

and 1 denotes the all-one vector of appropriate size.

Proof: The proof follows from majorization theory and [4, Section 3.4]. Let $d(E_0) = [\xi_1, \xi_2, \ldots, \xi_M]^T$ be the diagonal entries of $E_0$. From [4], $d(E_0)$ is majorized by $d(E_0)$. Since $f(\cdot)$ is Schur-convex, it follows that

\[ f(1(E_0)) \leq f(d(E_0)) \]

for any given precoder $\tilde{S}$. From [4, 6], there exists a unitary matrix $\Pi$ and a new precoder $\Pi S$ such that the MSE matrix $E_0$ has equal diagonal entries and $f(\cdot)$ is further reduced. Specifically, $\Pi$ can be selected as a Fourier or Hadamard matrix, or generated using Algorithm 2.2 of [4]! Note that the unitary transformation does not change the transmit power $\text{tr}(SS^T)$ nor the eigenvalues of $R_0$. Hence, the constructed precoder $\Pi S$ satisfies all the constraints of the original problem. Since $f(\cdot)$ is increasing, it is minimized when the trace of the MSE matrix $E_0$ is minimized, which can be achieved by solving

\[ \min_{\Pi} \text{tr}(E_0), \quad \text{s.t. } \text{tr}(SS^T) \leq P_0, \kappa(R_0) \leq \kappa_0. \]

Note that the sum-MSE is a Schur-convex and Schur-concave function of $\{\xi_m\}$ [4]. Since the eigenvalues of a Hermitian matrix majorize the diagonal entries, there is an optimal solution to (17) such that the MSE matrix $E_0$ is diagonal. With this property, the inverses of the diagonal entries of $E_0$ are the eigenvalues of $R_0 = I + S^H H_0^\dagger H_q S$. Therefore, any feasible choice of $S$ with given diagonal $E_0$ yields exactly the same eigenvalues of $R_0$ and so satisfies the condition number constraint. On the other hand, as shown in [4], among all such choices of $S$, $S = V_H \text{diag}(\sqrt{\beta})$ requires the minimum transmit power $\text{tr}(SS^T)$. In this case,

\[ E_0 = \text{diag} \left\{ \frac{1}{1 + \gamma_1 p_1}, \ldots, \frac{1}{1 + \gamma_M p_M} \right\}. \]
the eigenvalues of $R_0$ are given by
\[
\lambda_m(R_0) = 1 + \gamma_m p_m, \quad m = 1, 2, \ldots, M, \quad (19)
\]
and Lemma 1 is proved. □

Now the remaining problem is to solve (15), which can be reformulated by introducing a slack variable $u$ as
\[
\begin{align*}
\min_{p_1, \ldots, p_M, u} & \quad \sum_{m=1}^{M} \frac{1}{1 + \gamma_m p_m} \\
\text{s.t.} & \quad \sum_{m=1}^{M} p_m = P_0, \quad u \geq 1, \\
& \quad u \leq 1 + \gamma_m p_m \leq \kappa_0 u, \quad m = 1, 2, \ldots, M, \\
& \quad p_m \geq 0, \quad m = 1, 2, \ldots, M. \quad (20)
\end{align*}
\]
This problem is convex, involving only scalar variables. After examining the Karush-Kuhn-Tucker (KKT) conditions [7], we can characterize the solution by
\[
\begin{align*}
\sum_{m=1}^{M} p_m & = P_0, \\
p_m & = \left[ \frac{1}{\gamma_m} - 1 \right]^{-1} \frac{\kappa_0 u - 1}{\gamma_m - 1}, \quad m = 1, 2, \ldots, M, \\
\sum_{m=1}^{M} \left[ \frac{1}{\gamma_m u} - \frac{1}{\gamma_m \mu^2} \right] & = \kappa_0 \sum_{m=1}^{M} \left[ \frac{1}{\gamma_m u^2} - \frac{1}{\gamma_m \mu^2} \right], \quad u \geq 1, \quad \mu \geq 0, \quad (21)
\end{align*}
\]
where
\[
[x]_a^b = \begin{cases} 
  b, & x \geq b \\
  x, & a < x < b \\
  a, & x \leq a 
\end{cases}
\]
is a truncation function. Now the key is to solve these equations for $u$ and $\mu$. It can be verified that the optimal solution must satisfy $\gamma_m p_m \geq \gamma_k u$ for any $\gamma_m \geq \gamma_k$. Thus, truncation occurs only for large and small $\gamma_m$. We now search for a solution, analogous to what is done in the conventional waterfilling. We order the $\gamma_m$'s decreasingly. Let us say that the first $M_u$ powers $p_m$ and the last $M_t$ are truncated. This means $1 + \gamma_m u + 1 \leq \kappa_0 u$ and $1 + \gamma_{M-M_t} \mu \geq u$. For notational economy, let
\[
A_u \triangleq \sum_{m=1}^{M_u} \frac{1}{\gamma_m}, \quad A_t \triangleq \sum_{m=M-M_t+1}^{M} \frac{1}{\gamma_m}, \\
B \triangleq \sum_{m=M-M_t+1}^{M} \frac{1}{\gamma_m}, \quad \tilde{P} = P_0 + \sum_{m=1}^{M} \frac{1}{\gamma_m}. \quad (23)
\]
Then we need to check whether there exist $u$ and $\mu$ satisfying
\[
(k_0 A_u + A_t) u + B \mu - \tilde{P} = 0, \\
u - \sqrt{\frac{M_t + M_u}{A_t + k_0 A_u}} \mu = 0, \\
u \geq 1, \quad \mu \geq 0, \\
\mu \sqrt{\gamma_m u} - \kappa_0 u \leq 0, \\
\mu \sqrt{\gamma_{M-M_t}} - u \geq 0. \quad (24)
\]
This can be done with few real operations. (Note that we do not need to solve for $u$, $\mu$ unless all the inequalities in (24) are satisfied.) If the choice of $(M_u, M_t)$ satisfies the constraints, we keep this choice and substitute the values of $u$ and $\mu$ into (21) to find $p_m$; otherwise we try a different choice for $(M_u, M_t)$.

We remark that the conventional (optimal) precoder [4] is a special case of the proposed precoder with $\kappa_0 = \infty$. For $\kappa_0 < \infty$, the corresponding LMMSE equalization can converge faster than the equalizer with optimal precoding, but at the price of a slightly increased converged MSE. The tradeoff between the convergence rate and converged performance of the CG algorithm depends on $\kappa_0$. A smaller $\kappa_0$ should be chosen if faster convergence is required.

2.3. Computational Complexity

We first consider the complexity of the precoder optimization. We consider banded channel matrices $[H_q]$ with bandwidth $L$ [5], which is the case for multipath channels. The singular value decomposition (SVD) of $H_q$ can be implemented by a two-step algorithm [5], where the first step (i.e., bidiagonalization) can be realized using a fast algorithm specifically tailored to banded matrices [12]. The right singular vector matrix can be written as
\[
[ \Phi_H \simeq I - \beta_m \Phi_m \Phi_m^\dagger ]_q \frac{M}{M} \quad (25)
\]
where $\Phi_m = I - \beta_m \Phi_m \Phi_m^\dagger$ is a Householder matrix with $\phi_m$ a length-$M$ vector of up to $M - m - 1$ nonzero entries, and each $\Phi_m$ a Givens rotation matrix. From [5], a total of $\tau = M^2$ Givens rotations are sufficient to achieve good numerical accuracy. If the singular vector matrix $V_H$ is accumulated (i.e., computed explicitly), then the SVD requires about $\Theta_{svd} \approx 8LM^2 + 28/3M^3$ flops, where each flop represents a complex scalar addition, multiplication, division, or square root. However, it is not always advantageous to compute $V_H$ explicitly. If $V_H$ is stored in factored form as in (25), the complexity is $\Theta_{svd} \approx 8LM^2$ flops. For the power allocation stage of the precoder design, testing the feasibility of each choice for $(M_u, M_t)$ costs only a few real operations by exploiting the connections between $A_u, A_t$ and $B$ for different choices. The worst-case complexity for the power allocation is $O(M^2)$, since the number of choices is upper bounded by $M^2/2$. This is negligible compared with the complexity of the SVD and thus ignored.

We next consider the complexity of LMMSE equalization at the receiver. We assume that the precoder is computed at the receiver, so $V_H$ is perfectly known. By substituting $S = V_H \text{diag} (\sqrt{\beta}) \Pi$ into (2), the LMMSE equalization for block $q$ is achieved by solving
\[
\left( \begin{array}{c}
\Phi_q \\
\Psi_q \\
\end{array} \right) = \frac{1}{\tau} \left( \begin{array}{c}
PV_H H_q^\dagger H_q V_H \simeq I_M \quad (26)
\end{array} \right)
\]
where $P = \text{diag} (\sqrt{\beta})$. (Note that $\Psi_q$ and $R_q$ have the same eigenvalues.) If the conventional Cholesky factorization method is applied, the complexity of LMMSE equalization for $Q$ blocks is
\[
\Theta_{Chol} \approx Q(2NM + M^2 N + \frac{M^3}{3}) + 2M^2 + 2MN + \frac{3M}{2} \log_2 M \quad (27)
\]
flips, where the $2NM$ term arises from computing $F_q \triangleq H_q V_H$, $M^2 N$ from $F_q^T F_q$, $M^3/3$ from Cholesky factorization, and the last three terms in (27) are for backward/forward substitutions and matrix vector products (we assume $\Pi$ is the discrete Fourier transform (DFT) matrix). For the CG method, it is advantageous to use $V_H$ in factored form (without explicitly computing $V_H$). In this way, the

3094
complexity with $D$ iterations is
\[ \Theta_{CG} \approx Q \left( D(4NL + 16M^2) + 2NL + 8M^2 + 3/2M \log_2 M \right) \text{flops.} \]

3. EXAMPLE

We now present an example of the proposed precoder design applied to a single-carrier MIMO system over a time-varying multipath channel. We assume $n_t = 8$ transmit antennas and $n_r = 12$ receive antennas. The channel has $v = 4$ time-varying taps where each tap has average power $1/v$. The Doppler frequency is $f_d = 100 \text{ Hz}$. The bandwidth $W = 512 \text{ kHz}$. Each block spans $T = 16$ symbol intervals and each frame consists of $Q = 16$ blocks. The normalized Doppler frequency is 0.0036 for a block and 0.0576 for a frame. (Note that when $Q$ is increased, the complexity of the precoder optimization decreases but the performance also degrades.) In this case, $M = n_t T = 128$, $N = n_r (T + v - 1) = 228$ and $L = v n_t = 32$. The total transmit power is $P_t = 1280$. We compare the proposed precoder for $\kappa_0 = 2$ with the conventional precoder (i.e., $\kappa_0 = \infty$).

The simulation results are shown in Fig. 2, where the dashed curves marked by “fixed receiver” denote the scheme in which the LMMSE equalizer is designed assuming $H_q = H_0$ and reused for all blocks. The solid curves denote the solution in (2) using the updated channel matrix $H_q$ for each block. The “fixed receiver” leads to significant performance degradation. Thus, the receiver must be updated in a timely fashion to ensure satisfactory performance. In this case, the computational cost due to the calculations of the receiver coefficients becomes the bottleneck. Our precoders are designed to produce clustered eigenvalues of $R_0$. Over time, the clustering property decreases due to channel variations. This explains why the dashed curves converge faster than the solid curves. In all cases, however, our scheme converges faster than the conventional scheme.

With the Cholesky factorization-based method, the precoder design has complexity $\Theta_{\text{svd}} \approx 2.38 \times 10^7$ flops as $V_H$ must be explicitly computed. The complexity for equalizing $Q = 16$ blocks is about $\Theta_{\text{Chol}} \approx 1.02 \times 10^9$ flops. In contrast, when the CG method is used, the complexity of precoder design can be reduced by about 82% to $\Theta_{\text{svd}} \approx 4.19 \times 10^6$ flops because $V_H$ is stored in factored form. With $D = 3$, the complexity of equalization is significantly reduced by 85% to $\Theta_{CG} \approx 1.63 \times 10^7$ flops. Clearly, the CG method exhibits a large complexity advantage even accounting for the complexity of the precoder. Note that at $D = 3$, the best performance is achieved by the proposed precoder, which is slightly worse than the converged performance of the conventional precoder but still noticeably better than without precoding.

4. CONCLUSIONS

We have introduced a low-complexity precoding scheme for time-varying MIMO channels with iterative CG receivers. It allows trading off a small increase in MSE for faster convergence compared to the conventional (MMSE) precoder. Because the scheme is based on CG, it can easily be implemented in parallel.

5. REFERENCES