

LINEAR IIR-MMSE PRECODING FOR FREQUENCY SELECTIVE MIMO CHANNELS

Sander Wahls and Holger Boche

Technische Universität München
 Lehrstuhl für Theoretische Informationstechnik
 Karlstraße 45-47/V, 80333 München, Germany
 {sander.wahls,boche}@tum.de

ABSTRACT

We consider the design of linear precoding filters with respect to the minimum mean square error (MMSE) criterion for systems that employ an additional scalar gain next to a fixed receive filter. The precoding filter and the scalar gain are to be jointly optimized. Currently, only the finite impulse response (FIR) solution to this problem is known. The goal of this paper is to derive the infinite impulse response (IIR) MMSE precoder both with and without causality constraint, i.e., finite and infinite latency time, respectively. We discuss the role of the scalar gain and its relationship to automatic gain control (AGC). We also show that causal precoding requires that the joint first arrival delay of channel and receive filter is not larger than the latency time, and that the IIR-MMSE precoder enjoys the same advantages over the FIR-MMSE precoder as the IIR-MMSE equalizer does over the FIR-MMSE equalizer, viz.: improved performance and no need for latency time optimization.

I. INTRODUCTION

Precoding is a well-researched method to migrate the computational burden in channel equalization from the receiver to the transmitter. Nevertheless, in contrast to the MMSE equalizer which is given by the classical Wiener filter, the MMSE precoder is a quite recent development (see, e.g., the discussion in [1, VI.A]). Various MMSE precoders have been derived for quite a few different system models. In this paper, we consider the system model depicted in Figure 1. The channel H and the receive filter G are considered to be fixed and known. The transmitter employs the precoding filter P . Additionally, we include the scalar gain α as an additional simple receive filter. The filters P and α are to be jointly optimized subject to a transmit power constraint. The idea to include a scalar gain at the receiver can be traced back at least to a paper of Karimi et. al. [2]. It offers two main advantages. The first advantage is of course an additional degree of freedom which usually will improve the overall system performance. The second advantage is that we can find closed-form solutions. The FIR-MMSE precoder for the model in Fig. 1 was derived by Choi and Murch [3] and, using a different technique, also by Joham et. al. [1]. However, the IIR counterpart is still unknown. In this paper, we are going to derive the IIR-MMSE precoder, both with and without causality constraint. We show that the IIR precoder has the same advantages compared to the FIR precoder as in the receive filter case. First of all, the IIR precoder is the optimal linear precoder and therefore can never be outperformed by the FIR precoder. Second, improving the latency

This work has been supported by the German Research Foundation (DFG) under grant BO 1734/5-2.

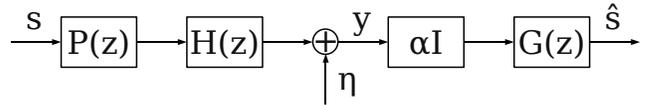


Fig. 1. System model.

time always improves the performance of an IIR precoder. This is in contrast to FIR precoding, where the optimal latency time is not obvious and has to be found by a costly exhaustive search or by suboptimal approaches [4]. IIR precoders do not require such optimization. Third, IIR precoders often outperform FIR precoders of similar complexity.

The paper is structured as follows. We start with a brief description of the system model in Section II and give an in-depth discussion of the role of the scalar gain at the receiver in Section III. Then, we give the optimal IIR-MMSE precoders with and without causality constraint in the Sections IV and V. We illustrate our results with numerical examples, and close the paper with a conclusion (Sections VI, VII).

Notation: The set of stable rational $m \times n$ matrices (i.e., no poles on $|z| = 1$) is denoted by $\mathcal{RL}_{\infty}^{m \times n}$ and the set of stable and causal rational $m \times n$ matrices (i.e., no poles in $|z| \geq 1$) by $\mathcal{RH}_{\infty}^{m \times n}$. The causal part of any rational matrix A is denoted by $\{A\}_+$ and the para-hermitian by $A^{\sim}(z) := A(\bar{z}^{-1})^*$. The two norm of any $B \in \mathcal{RL}_{\infty}^{m \times n}$ is defined as $\|B\|_2^2 := \langle B, B \rangle_2$, where $\langle C, D \rangle_2 := \oint_{|z|=1} \text{trace}(D^{\sim}(z)C(z)) \frac{dz}{2\pi i}$ for all $C, D \in \mathcal{RL}_{\infty}^{m \times n}$. Finally, with $G \in \mathcal{RH}_{\infty}^{r \times q}$ and $H \in \mathcal{RH}_{\infty}^{q \times p}$ denoting the equalizer and the channel in Figure 1, we introduce co-spectral factorizations $S_{\xi}^{\sim} S_{\xi} = H^{\sim} G^{\sim} G H + \xi I$ (i.e., $S_{\xi}, S_{\xi}^{-1} \in \mathcal{RH}_{\infty}^{p \times p}$ [5]) for all $\xi > 0$,¹ and define $\tilde{P}_{\xi} := S_{\xi}^{-1} \{z^{-L} S_{\xi}^{\sim} H^{\sim} G^{\sim}\}_+$.

II. SYSTEM MODEL

Our system model is depicted in Figure 1. In formulae, we have

$$\begin{aligned} y(z) &= H(z)P(z)s(z) + \eta(z) \\ \hat{s}(z) &= \alpha G(z)y(z) = \alpha G(z)H(z)P(z)s(z) + \alpha G(z)\eta(z), \end{aligned}$$

where $H \in \mathcal{RH}_{\infty}^{q \times p}$ is the channel, $G \in \mathcal{RH}_{\infty}^{r \times q}$ is a fixed receive filter, $P \in \mathcal{RH}_{\infty}^{p \times r}$ is the precoder, and $\alpha \geq 0$ is a scalar gain. The signals s, y, η denote transmitted and received signals, and noise. We assume that both s and η are mutually independent white

¹The spectral factorizations always exist but are non-unique. To ensure uniqueness, we further assume that $S_{\xi}(\infty)$ is upper triangular.

random sequences with zero mean and covariance matrices I and $\sigma_\eta^2 I > 0$, respectively. We want to minimize the mean square error²

$$\text{MSE}(P, \alpha) := \|z^{-L}I - \alpha GHP\|_2^2 + \alpha^2 \sigma_\eta^2 \|G\|_2^2, \quad (1)$$

where $L \in \mathbb{N}$ is the latency time, subject to the power constraint

$$\|P\|_2^2 \leq E_{tr}. \quad (2)$$

III. ROLE OF THE SCALAR GAIN AT THE RECEIVER

In many papers, it is claimed that the scalar gain α at the receiver can be implemented with a simple automatic gain control (AGC). See, e.g., [3, II] or [1, II]. An AGC rescales the input signal at the receiver such that its average power (or some other criterion like peak amplitude) is equal to some fixed reference value [6]. This reference value is chosen such that the input signal at the receiver fits the dynamic range of the analog-to-digital converter. The AGC is automatic in the sense that no other inputs next to the actual input signal are required. Hence, we can interpret the scaling by α as the action of an AGC if the scaled receiver input αy is normalized in some reasonable way. However, this is not the case in general. If all other system parameters are fixed, the optimal gain that minimizes

$$\begin{aligned} \text{MSE}(P, \alpha) &= \|z^{-L}I\|_2^2 - 2\alpha \Re(\langle z^{-L}I, GHP \rangle_2) \\ &\quad + \alpha^2 (\|GHP\|_2^2 + \sigma_\eta^2 \|G\|_2^2) \end{aligned}$$

can be found with a standard calculus argument,

$$\alpha_* = \frac{\Re(\langle z^{-L}I, GHP \rangle_2)}{\|GHP\|_2^2 + \sigma_\eta^2 \|G\|_2^2}.$$

We can assume that $\alpha_* \geq 0$ without loss of generality. Otherwise, we absorb the negative sign into the precoding filter P . The resulting average power of the scaled receiver input becomes

$$\begin{aligned} &\lim_{K \rightarrow \infty} \mathbb{E} \left[\frac{1}{K+1} \sum_{k=0}^K \|\alpha_* y_k\|_F^2 \right] \\ &= \alpha_*^2 (\|HP\|_2^2 + \sigma_\eta^2 \|I\|_2^2) \\ &= \frac{\Re(\langle z^{-L}I, GHP \rangle_2)^2}{(\|GHP\|_2^2 + \sigma_\eta^2 \|G\|_2^2)^2} (\|HP\|_2^2 + \sigma_\eta^2 \|I\|_2^2). \quad (3) \end{aligned}$$

If the noise power grows large, $\sigma_\eta^2 \rightarrow \infty$, we have

$$\begin{aligned} \lim_{K \rightarrow \infty} \mathbb{E} \left[\frac{1}{K+1} \sum_{k=0}^K \|\alpha_* y_k\|_F^2 \right] &\approx \frac{\Re(\langle z^{-L}I, GHP \rangle_2)^2 \|I\|_2^2}{\sigma_\eta^2 \|G\|_2^4} \\ &\leq \frac{E_{tr} \|GH\|_2^2 \|I\|_2^4}{\sigma_\eta^2 \|G\|_2^4} \quad (4) \end{aligned}$$

Note that the upper bound (4) is independent of the precoding filter P . Hence, the rescaled receiver input $\alpha_* y$ will be arbitrarily closed to zero if the noise power grows large, $\sigma_\eta^2 \rightarrow \infty$. This is not the action of an AGC. The situation is a little bit different if we consider the case of low noise power, $\sigma_\eta^2 \rightarrow 0$. Assume that the concatenation of the channel H and the receive filter G has a stable and causal right-inverse with latency time L , i.e., $GHX = z^{-L}I$ for some $X \in \mathcal{RH}_\infty^{p \times r}$ and that G is square. Then, the average power of the scaled receiver input converges to a value that is independent of the channel if optimal precoding filters and scalar gains are employed and $\sigma_\eta^2 \rightarrow 0$ (cf. Sec. IV-B). In that case,

²The stochastic interpretation of the 2-norm allows us write the MSE and the power constraint in (1) and (2) as system norms. See, e.g., [5, 10.3.1].

the scaling can indeed be interpreted as the action of an AGC. However, if no right inverse X exists, the average power of the scaled receiver input depends on the channel. Again, this is not what an AGC would do.

Hence, we can consider the scalar gain as an AGC only if the concatenation GH of channel and receive filter is stably and causally right-invertible with latency time L , G is square and the noise power σ_η^2 is not too small. Otherwise, we may e.g. use a preamble to let the receiver learn α_{opt} . Alternatively, we can employ a phase shift keying (PSK) modulation which is independent of the amplitude.

IV. OPTIMAL CAUSAL PRECODER

A well-known issue with IIR filters is that the optimal filters usually are non-causal and thus cannot be implemented in real-time. Therefore, a causality constraint in form of a finite latency time has to be incorporated explicitly. In this section, we derive and discuss the optimal precoder with causality constraint.

IV-A. Problem statement and main result

We start with a formal statement of the causal precoding problem.

Problem 1 (Optimal causal precoder). Find $P \in \mathcal{RH}_\infty^{p \times r}$ and $\alpha \geq 0$ such that the MSE (1) is minimized subject to (2).

The following theorem establishes a solution to Problem 1. The proof is given in the appendix.

Theorem 2 (Optimal causal precoder). Let $\{z^{-L}H \sim G\}_+ \neq 0$, and define $\xi_{opt} := E_{tr}^{-1} \sigma_\eta^2 \|G\|_2^2$. Then, with \tilde{P}_ξ as in Section I,

$$P_{opt} := \frac{1}{\alpha_{opt}} \tilde{P}_{\xi_{opt}} \text{ and } \alpha_{opt} := \frac{\|\tilde{P}_{\xi_{opt}}\|_2}{\sqrt{E_{tr}}} \quad (5)$$

are well-defined and solve Problem 1.

IV-B. Remarks and discussion

Let us give some remarks on Theorem 2.

The condition $\{z^{-L}H \sim G\}_+ \neq 0$: The condition $\{z^{-L}H \sim G\}_+ \neq 0$ in Theorem 2 has a very simple interpretation. Let $F(z) = \sum_{k=0}^{\infty} F_k z^{-k}$ denote the Laurent expansion of the concatenation of equalizer and channel, i.e., $F = GH$. Then, the condition is equivalent to that there is some $0 \leq k_0 \leq L$ such that $F_{k_0} \neq 0$. Intuitively, this means that the joint first arrival delay of channel and equalizer is not larger the latency time of the precoder.

Limiting cases $\xi \rightarrow 0$ and $\xi \rightarrow \infty$: The limiting cases are similar to the FIR case [1, VI.B]. Assume that the causal minimum-norm right-inverse of GH with latency time L , $P_{zf} := \text{argmin}\{\|P\|_2^2 : P \in \mathcal{RH}_\infty^{p \times q}, GH P = z^{-L}I\}$, is well-defined [7]. Then, the precoder P_{opt} in Theorem 2 converges towards $\sqrt{E_{tr}} \|P_{zf}\|_2^{-1} P_{zf}$ as $\xi \rightarrow 0$. On the other hand, P_{opt} converges towards $\sqrt{E_{tr}} \|M\|_2^{-1} M$, where $M := \{z^{-L}H \sim G\}_+$ is the causal matched filter, as $\xi \rightarrow \infty$.

Power of the scaled receiver input for $\sigma_\eta^2 \rightarrow 0$: Assume that the minimum-norm right-inverse P_{zf} from the previous subsection is well-defined, G is square and we employ P_{opt} and α_{opt} . Then, G is invertible with $G^{-1} = z^L H P_{zf}$. The average power (3) becomes

$$\lim_{K \rightarrow \infty} \mathbb{E} \left[\frac{1}{K+1} \sum_{k=0}^K \|\alpha_{opt} y_k\|_F^2 \right]$$

$$\begin{aligned}
&\approx \frac{\frac{E_{tr}}{\|P_{zf}\|_2^2} \Re\{z^{-L}I, z^{-L}I\}_2^2}{\left(\frac{E_{tr}}{\|P_{zf}\|_2^2} \|z^{-L}I\|_2^2\right)^2} \left(\frac{E_{tr}}{\|P_{zf}\|_2^2} \|HP_{zf}\|_2^2\right) \\
&= \|HP_{zf}\|_2^2 = \|z^{-L}G^{-1}\|_2^2
\end{aligned}$$

if $\sigma_\eta^2 \rightarrow 0$. Thus, the average power converges towards a fixed value that is independent of the channel H . Hence, we can consider the scaling by α_{opt} as the action of an AGC if σ_η^2 is small.

IIR-FIR structure: By definition of \tilde{P}_ξ , we see that the MMSE precoder (5) decomposes into two parts: an IIR part $\alpha^{-1}S_\xi^{-1}$ and a FIR part $\alpha^{-1}\{z^{-L}S_\xi^{-1}H\tilde{G}\}_+$. The filter length of FIR part equals the latency time L , while the IIR part is independent of L .

IV-C. Further results

In this subsection, we give some further results on the IIR-MMSE precoder. Note that the proofs have been moved to the Appendix.

Our first result addresses the question what happens if the assumption $\{z^{-L}H\tilde{G}\}_+ \neq 0$ in Theorem 2 fails. It is simple to see that the solution given in the Theorem then no longer is well-defined. Furthermore, the following Proposition shows that no transmission becomes an optimal strategy.

Proposition 3. *Suppose $\{z^{-L}H\tilde{G}\}_+ = 0$. Then, $(P, \alpha) = (0, 0)$ solves Problem 1.*

Next, we formalize the intuitive property that no FIR precoder can outperform the IIR-MMSE precoder (if their latency times are identical).

Proposition 4. *Let $P(z) = \sum_{n=0}^{N-1} P_n z^{-n}$, $\|P\|_2^2 \leq E_{tr}$, denote an arbitrary FIR precoder, and let $\alpha \geq 0$. Then, the MSE achieved by (P, α) is lower bounded by the MSE of (5).*

Finally, we consider the issue of latency time optimization. Finding the optimal latency time L that minimizes the MSE among the FIR-MMSE precoders of a fixed length is an intricate problem [4]. Basically, one has to test all possible latency times. Fortunately, this becomes a non-issue when IIR-MMSE precoders are used.

Proposition 5. *The MSE of (5) decreases monotonically with L .*

V. OPTIMAL NON-CAUSAL PRECODER

The optimal precoder without explicit causality constraint usually is non-causal and therefore cannot be implemented in real-time. Nevertheless, it still often is of theoretical interest because it constitutes the limit of the optimal causal precoder as the latency time goes to infinity. Formally, the optimal precoder without causality constraint is the solution to the following problem.

Problem 6 (Optimal non-causal precoder). Find $P \in \mathcal{LH}_\infty^{p \times r}$ and $\alpha \geq 0$ such that (1) is minimized for $L = 0$ subject to (2).

The following theorem, which can be obtained after simple modifications of the proof of Theorem 2, gives us the optimal precoder without causality constraint.

Theorem 7 (Optimal non-causal precoder). *Let $GH \neq 0$, and define $\xi_{opt} := E_{tr}^{-1} \sigma_\eta^2 \|G\|_2^2$ as well as $\tilde{P}_{\xi_{opt}} := (H\tilde{G}G^{-1}H + \xi_{opt}I)^{-1}H\tilde{G}$. Then,*

$$P_{opt} := \frac{1}{\alpha_{opt}} \tilde{P}_{\xi_{opt}} \text{ and } \alpha_{opt} := \frac{\|\tilde{P}_{\xi_{opt}}\|_2}{\sqrt{E_{tr}}}$$

are well-defined and solve Problem 6.

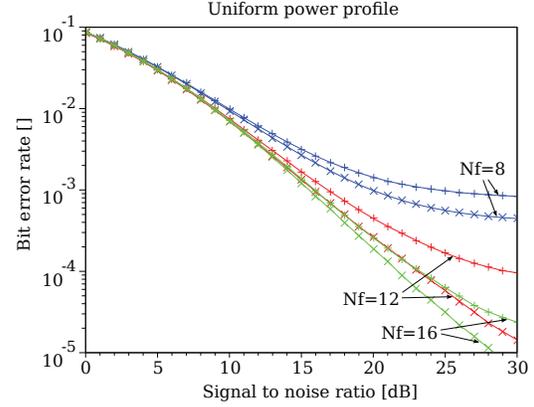


Fig. 2. First example. The “+” marks belong to the FIR-MMSE, the “x” marks to the IIR-MMSE.

VI. NUMERICAL EXAMPLES

VI-A. Simulation setup

The simulation setup is as follows. Per simulation run, we created 10.000 instances of the 2×2 FIR channel

$$H(z) = c_0 H_0 + c_1 H_1 z^{-1} + c_2 H_2 z^{-2},$$

where the real and imaginary parts of the entries of each H_k are random normally distributed variables with mean zero and variance one, and c_0, c_1, c_2 are fixed scalars. The equalizer used was $G(z) = I$. Per channel instance, we transmitted 1.000 binary phase shift keying (BPSK) modulated data signals, i.e., $s_k \in \{-1, 1\}^2$, at various signal to noise ratios $E_{tr}/\sigma_\eta = 1/\sigma_\eta$. We finally averaged the resulting uncoded bit error rates per signal to noise ratio.

VI-B. Precoder design

We compare the FIR-MMSE precoder for various filter lengths N_f with the IIR-MMSE precoder from Theorem 2. The latency time of the FIR-MMSE precoder was chosen such that the MSE is minimized [4]. In contrast, Proposition 5 shows us that we should make the latency time L of the IIR precoder as large as possible. In order to ensure a fair comparison, we chose L such that the number K_{IIR} of parameters necessary to describe the resulting IIR precoder is not larger than the number of parameters necessary to describe the FIR precoder, i.e., $K_{FIR} = N_f q p = 4N_f$. The IIR part of the precoder, i.e., $S_{\xi_{opt}}^{-1}$, has the same McMillan degree as the channel, i.e., $\kappa = 4$. Therefore, when we use a state-space realization in block controllable companion form [8, §5.4D] of $S_{\xi_{opt}}^{-1}$, $\kappa p^{-1} + \kappa q + q p = 14$ parameters are required to describe it. The FIR part of the IIR-MMSE precoder has length L . Therefore, the total number of parameters necessary in the IIR case is $K_{IIR} = 14 + L q p = 14 + 4L$. In our simulations, we chose $L = N_f - 4$, which gives $K_{IIR} = K_{FIR} - 2 < K_{FIR}$.

VI-C. First example: Uniform power profile

In this example, we chose an uniform power profile, i.e., $c_0 = c_1 = c_2 = 1$. We considered the filter lengths $N_f \in \{8, 12, 16\}$. Fig. 2 show the simulation results. We observe that the IIR-MMSE outperforms the FIR-MMSE for all three filter lengths with

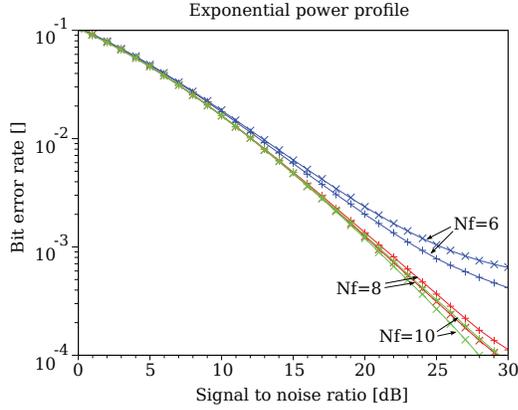


Fig. 3. Second example. The “+” marks belong to the FIR-MMSE, the “x” marks to the IIR-MMSE.

improvements up to an order of magnitude in the high SNR regime. Also note that there are error floors. The reason is that we consider square channels. These are likely to have unstable zeros, i.e., zeros outside the unit circle. Such channels cannot be causally inverted, which is why there are errors no matter how high the SNR.

VI-D. Second example: Exponential power profile

We consider an exponential power profile, i.e., $c_0 = 1.64$, $c_1 = 0.53$, and $c_2 = 0.16$, and filter lengths $N_f \in \{6, 8, 10\}$. Fig. 3 shows the simulation results. The IIR precoders outperform the FIR precoders for the filter lengths $N_f \in \{8, 10\}$, but they are outperformed by the FIR precoders for $N_f = 6$. (We point out that this does not contradict Proposition 4 because the FIR precoders may have a larger latency time than the IIR precoders.) However, the difference in performance between the FIR and IIR precoders in this example is small for all considered filter lengths.

VII. CONCLUSION

In this paper, we have derived the IIR-MMSE precoder with and without causality constraint. We clarified that the usual interpretation of the scalar gain at the receiver as an AGC is valid only under certain conditions. We also showed that the causal precoder requires a latency time equal or greater than the joint first arrival delay of channel and equalizer, while the non-causal precoder only requires that the concatenation of channel and equalizer is not identically zero. Another result was that the MSE decreases monotonously with the latency time. Numerical examples showed that IIR-MMSE precoders usually outperform FIR-MMSE precoders with a similar number of parameters even if the latency time of the FIR precoders has been optimized (which is a costly operation).

VIII. REFERENCES

- [1] M. Joham, W. Utschick, and J. A. Nossék, “Linear transmit processing in mimo communications systems,” *IEEE Trans. Signal Process.*, vol. 53, pp. 2700–2712, Aug. 2005.
- [2] H. R. Karimi, M. Sandell, and J. Salz, “Comparison between transmitter and receiver array processing to achieve interference nulling and diversity,” in *Proc. PIMRC*, (Osaka, Japan), pp. 997–1001, Sept. 1999.
- [3] R. L.-U. Choi and R. D. Murch, “New transmit schemes and simplified receivers for mimo wireless communication systems,” *IEEE Trans. Wireless Commun.*, vol. 2, pp. 1217–1230, Nov. 2003.
- [4] M. Joham, W. Utschick, and J. A. Nossék, “Latency time optimization for FIR and block transmit filters,” in *Proc. ISSPA*, (Paris, France), pp. 273–276, 2003.
- [5] B. Hassibi, A. H. Sayed, and T. Kailath, *Indefinite-Quadratic Estimation and Control*. Philadelphia: SIAM, 1999.
- [6] D. Whitlow, “Design and operation of automatic gain control loops for receivers in modern communication systems,” in *ADI Wireless Seminar*, 2006.
- [7] S. Wahls and H. Boche, “Novel System Inversion Algorithm with Application to Oversampled Perfect Reconstruction Filter Banks,” *IEEE Trans. Signal Process.*, vol. 58, no. 6, pp. 3008–3016, 2010.
- [8] P. J. Antsaklis and A. N. Michel, *Linear Systems*. Boston: Birkhäuser, 2006.

APPENDIX

Proof: (Proof of Theorem 2) We skip the well-definedness of P_{opt} and α_{opt} because of space limitations. Let us show optimality by modification of an argument in [3]. The MMSE can be lower bounded as follows,

$$\begin{aligned}
& \text{MMSE} \\
& := \inf_{\substack{\alpha \geq 0, P \in \mathcal{RH}_{\infty}^{p \times r} \\ \|P\|_2^2 \leq E_{tr}}} \text{MSE}(P, \alpha) \\
& = \inf_{\substack{\alpha \geq 0, P \in \mathcal{RH}_{\infty}^{p \times r} \\ \|P\|_2^2 \leq E_{tr}}} (\|z^{-L}I - \alpha GHP\|_2^2 + \alpha^2 \sigma_{\eta}^2 \|G\|_2^2) \\
& \geq \inf_{\substack{\alpha \geq 0, P \in \mathcal{RH}_{\infty}^{p \times r} \\ \|P\|_2^2 \leq E_{tr}}} (\|z^{-L}I - \alpha GHP\|_2^2 + \alpha^2 \sigma_{\eta}^2 \frac{\|P\|_2^2}{E_{tr}} \|G\|_2^2) \\
& = \inf_{\substack{\alpha \geq 0, P \in \mathcal{RH}_{\infty}^{p \times r} \\ \|P\|_2^2 \leq E_{tr}}} (\|z^{-L}I - GH(\alpha P)\|_2^2 + \frac{\sigma_{\eta}^2}{E_{tr}} \|G\|_2^2 \|\alpha P\|_2^2) \\
& = \inf_{X \in \mathcal{RH}_{\infty}^{p \times r}} (\|z^{-L}I - GHX\|_2^2 + \frac{\sigma_{\eta}^2}{E_{tr}} \|G\|_2^2 \|X\|_2^2).
\end{aligned}$$

Theorem 10.3.3 in [5] shows that $\tilde{P} = \alpha_{opt} P_{opt}$ is the unique minimizer of the lower bound, $\{\tilde{P}\} = \arg\min_{X \in \mathcal{RH}_{\infty}^{p \times r}} (\|z^{-L}I - GHX\|_2^2 + E_{tr}^{-1} \sigma_{\eta}^2 \|G\|_2^2 \|X\|_2^2)$. Since $\|P_{opt}\|_2^2 = E_{tr}$ and

$$\text{MSE}(P_{opt}, \alpha_{opt}) = \|z^{-L}I - GH\tilde{P}\|_2^2 + \frac{\sigma_{\eta}^2}{E_{tr}} \|G\|_2^2 \|\tilde{P}\|_2^2$$

by construction, we see that P_{opt} and α_{opt} solve Problem 1. ■

Proof: (of Proposition 3) Since $\{z^{-L}H \sim G \sim\}_+ = 0$ we can write $GH = z^{-(L+1)}\Psi$ for some $\Psi \in \mathcal{RH}_{\infty}^{p \times r}$. Then, by Parseval’s relation (see, e.g., [5, (10.3.4)]), the mean square error is lower bounded as follows, $\text{MSE} = \|z^{-L}I - \alpha GHP\|_2^2 + \alpha^2 \sigma_{\eta}^2 \|G\|_2^2 = \|z^{-L}I\|_2^2 + \alpha^2 \|\Psi P\|_2^2 + \alpha^2 \sigma_{\eta}^2 \|G\|_2^2 \geq \|z^{-L}I\|_2^2$. The choice $(P, \alpha) = (0, 0)$ achieves this lower bound. ■

Proof: (of Proposition 4). The proposition follows immediately from $P \in \mathcal{RH}_{\infty}^{p \times r}$ and Theorem 2. ■

Proof: (of Proposition 5) Let $P_1 \in \mathcal{RH}_{\infty}^{p \times r}$, $\|P_1\|_2^2 \leq E_{tr}$, $\alpha > 0$ denote any precoder with decision delay L . Then, with $P_2 := z^{-1}P_1$, (P_2, α) is a feasible precoder with decision delay $L + 1$, and the MSE of (P_2, α) equals the MSE of (P_1, α) . ■