Available online at www.sciencedirect.com

SCIENCE DIRECT®

Systems & Control Letters 49 (2003) 121 – 130

SYSTEMS & CONTROL LETTERS

www.elsevier.com/locate/sysconle

Design and analysis of linear precoders under a mean square error criterion, Part I: foundations and worst case designs ☆

Jonathan H. Manton*

Department of Electrical and Electronic Engineering, ARC Special Research Centre for Ultra-Broadband Information Networks, The University of Melbourne, Victoria 3010, Australia

Received 1 December 2001; received in revised form 19 September 2002; accepted 7 November 2002

Abstract

Much debate surrounds the pros and cons of linear precoders in wireless communication systems. This two-part paper contributes to the debate by formulating the precoder design problem as an optimisation problem and studying the optimal solutions, thereby gaining a better understanding of how precoders work and what they can do. Part I builds a mathematical foundation for the study of linear precoders. It is shown that under a mean square error criterion, a natural convex geometry arises. This geometry facilitates the derivation of a necessary and sufficient condition for a linear precoder to be maximally robust, meaning the precoder's worst case performance is no worse than that of any other linear precoder. Part II studies precoders having the lowest average mean square error, deriving closed form solutions in special cases and developing a stochastic optimisation algorithm for computing optimal precoders in general.

Keywords: Orthogonal frequency division multiplex; Multipath channels; Wireless communications; Channel coding; Linear precoders; Strictly uniform precoders; Robust precoders; Minimax precoders

1. Introduction

The underlying problem in wireless communications is how best to encode a message so that the receiver can recover the message despite interference caused by multipath propagation and additive noise [17]. One proposed method for encoding the message is to use a linear precoder [3,19], either on its own or as the inner code in a two stage coding strategy, with the outer code being a standard error correcting code over a finite alphabet [2,5,10]. Although the design of linear precoders under the simplifying assumption that the channel is known has been considered in [6,18,19], the design of precoders for unknown channels has received relatively little attention to date [14,15]. Furthermore, it is not clear from the literature whether linear precoders can offer improved performance; in [22] it is claimed they do whereas Debbah et al. [2] suggests perhaps they do not. This two-part paper addresses both these issues.

Part I establishes a mathematical framework in which to study linear precoders. It converts the precoder design problem into a constrained optimisation problem. It goes on to use the recently discovered

☆ This work was performed in part while the author was a Tan Chin Tuan Exchange Fellow at Nanyang Technological University, Singapore.

* Tel.: +61383446791; fax: +61383446678.
E-mail address: jon@ee.mu.oz.au (J.H. Manton).

0167-6911/03/$ - see front matter
PII: S0167-6911(02)00316-X

convex geometry in linearly precoded systems [9] to derive closed form solutions to this optimisation problem under a worst case criterion. Part II [10] derives optimal solutions under a criterion measuring average performance rather than worst case performance. These results indicate linear precoders protect messages from frequency distortion caused by multipath propagation but not from additive white Gaussian noise (AWGN).

The motivation for studying robust linear precoders is the desire to design linear precoders which work well over any channel. Robust precoders are suitable for broadcast systems over fixed [1] channels and where the individual channels between the transmitter and each receiver differ substantially. Unfortunately though, one of the theoretical results of this paper is that the guaranteed performance of *maximally* robust linear precoders comes at the cost of requiring the number of transmitted symbols to be no less than L times the number of source symbols, where L is the channel length. Maximally robust linear precoders are therefore only interesting from a theoretical perspective; they establish a benchmark against which to compare all other robust precoders. Indeed, any other robust precoder must trade off worst case performance for better spectral utilisation (that is, less redundancy introduced by the precoder).

Related work in the literature is now summarised. Filter bank precoders, which are one way of implementing linear precoders, were explicitly introduced in [3]. Their purpose was not only to remove the need for oversampling the channel output in order to identify blindly the channel, but also to remove the conditions on the locations of the channel zeros for blind identification to be possible. It was soon realised (see the references in [19]) that filter bank precoders had other advantages, such as allowing the channel to be equalised perfectly using an FIR equaliser [25] and allowing the source symbols to be recovered regardless of the location of the channel zeros [19,21]. An explanation for these advantageous properties was given in [13] where it was shown that all redundant filter bank precoders spread the spectrum of the transmitted symbols in a predictable way. As previously mentioned though, although the optimal design of linear precoders for known channels has been studied in [6,18,19], little attention has been given to designing optimal linear precoders for unknown channels [14,15].

The organisation of this paper is as follows. The mathematical foundations of linear precoder design are laid down in Section 2. The main results are that the Cramer–Rao Bound (CRB) provides an intrinsic measure of performance of a linear precoder and that optimal precoders exist under quite general conditions. Section 3 defines three classes of robust linear precoders; maximally robust precoders, uniform precoders and strictly uniform precoders. Section 4 proves the somewhat surprising result that these three classes are identical. Section 5 derives necessary and sufficient conditions for a precoder to be maximally robust and presents explicit expressions for maximally robust precoders introducing the least possible redundancy. These minimally redundant maximally robust precoders are discussed further in Section 6. Section 7 summarises the results of this paper; a full conclusion appears at the end of Part II [10].

2. Foundations

This section formulates mathematically the optimal precoder design problem and proves the existence of optimal precoders.

2.1. Transmission model

The following two applications result in the same problem formulation considered in this paper.

Packet Networks: Consider a wireless packet network where short messages are sent sporadically between users. The wireless link is capable of transmitting a block $\boldsymbol{x} \in \mathbb{C}^m$ of m symbols but the received block will differ from the transmitted block due to multipath propagation and additive noise (explained later). The underlying problem is how best to encode each message as a codeword $\boldsymbol{x}$ in m-dimensional space. This paper considers a special case of this coding problem; it assumes the mapping from messages to codewords in $\mathbb{C}^m$ occurs in two steps. First the message is mapped to a codeword $\boldsymbol{s}$ in $\mathbb{C}^p$ (using a traditional error correcting code, for instance) and then a linear precoder matrix $\tilde{P} \in \mathbb{C}^{m \times p}$ is used to

[1] For time varying channels, maximising average performance over time is likely to be more appropriate [10].

obtain the codeword $\boldsymbol{x} = \tilde{P}\boldsymbol{s}$ in $\mathbb{C}^m$. The following issues are addressed (both here and in Part Two [10]): How should the dimension p be chosen? Given p, how should the precoder $\tilde{P}$ be designed? To what extent can the linear precoder take the place of more traditional error correcting codes?

Block transmissions: Consider transmitting a continuous stream of digitised voice data in real time. The real time requirement is met if the data is broken into blocks and each block transmitted separately. In other words, the encoding problem is equivalent to the one described above for packet networks. It is emphasised though that this equivalence does not hold if either the digitised voice data is allowed to be coded over multiple blocks (by using interleavers, for instance) or if the receiver is allowed to use previously received blocks to assist in the decoding of the current block. Indeed, if the channel is time varying and ergodic, it is possible to exploit the channel's ergodicity [24] by coding over multiple blocks to improve performance, and in particular, the design criteria in this paper would not be appropriate.

The transmission model is now formulated mathematically. The multipath propagation is modelled by a finite impulse response (FIR) channel $\boldsymbol{h} = [\boldsymbol{h}_0, \ldots, \boldsymbol{h}_{L-1}]^{\mathrm{T}} \in \mathbb{C}^L$ of length L, where T denotes transpose. For reasons given below, a mild restriction is imposed on the class of precoders $\tilde{P}$ considered in this paper. For given values of n and L, define the cyclic prefix matrix $C \in \mathbb{C}^{(n+L-1)\times n}$ to be

$$C = \begin{bmatrix} 0_{(L-1)\times(n-L+1)} \; I_{L-1} \\ I_n \end{bmatrix}, \tag{1}$$

where 0 and I are the zero and identity matrices of sizes as given by their subscripts. Define $D \in \mathbb{C}^{n\times n}$ to be the discrete Fourier transform matrix whose ijth element is given by $D_{ij} = \mathrm{e}^{-\jmath 2\pi(i-1)(j-1)/n}$. Only precoders of the form $\tilde{P} = CD^{\mathrm{H}}P$ for some arbitrary matrix $P \in \mathbb{C}^{n\times p}$, with $n \geqslant p$, are considered. The codeword $\boldsymbol{s} \in \mathbb{C}^p$ is thus mapped to $\boldsymbol{x} = CD^{\mathrm{H}}P\boldsymbol{s}$. The received vector $\boldsymbol{y} \in \mathbb{C}^n$ is the noisy convolution of $\boldsymbol{x}$ with $\boldsymbol{h}$, namely

$$\boldsymbol{y} = \mathscr{H}CD^{\mathrm{H}}P\boldsymbol{s} + \boldsymbol{n}, \qquad \boldsymbol{n} \sim N(0, I), \tag{2}$$

where $\mathscr{H}$ is the upper triangular $n\times(n+L-1)$ Toeplitz channel matrix with first row $[\boldsymbol{h}_{L-1}, \ldots, \boldsymbol{h}_0, 0, \ldots, 0]$, superscript H denotes Hermitian transpose and $\boldsymbol{n}$ denotes AWGN with unit variance ($\boldsymbol{E}[\boldsymbol{n}\boldsymbol{n}^{\mathrm{H}}] = I$); the true variance of the noise is unimportant. Henceforth, P and not $\tilde{P}$ is referred to as the precoder matrix.

Remarks. Although (2) can be interpreted as sending the precoded symbols $P\boldsymbol{s}$ through an OFDM system [23], it is only the cyclic prefix component of the OFDM system that must be present because P is able to cancel out the D^{H} operation in (2) if it so desires. Note too that a zero padded system is obtained by setting the last $L-1$ rows of P to zero; this is elaborated on in Part Two [10].

Justification for restricting attention to precoders of the form $\boldsymbol{x} = CD^{\mathrm{H}}P\boldsymbol{s}$ is now given. Since $\boldsymbol{x}$ passes through a channel with memory $L-1$, the first $L-1$ symbols of $\boldsymbol{x}$ must clear the memory of the channel while the last $L-1$ symbols of $\boldsymbol{x}$ must flush the memory of the channel so no data is lost. This appears to require $2L-2$ redundant symbols. However, the cyclic prefix matrix C introduces only $L-1$ redundant symbols yet achieves a similar effect [7,11]. A more impressive feature of the cyclic prefix is it allows channels with unstable inverses to be inverted accurately with a linear block equaliser [7,11]. Therefore, only linear precoders adding a cyclic prefix are considered here. The introduction of D^{H} is for notational convenience only; any precoder $\tilde{P} = CP_1$ is obtained by setting $P = DP_1$.

2.2. Figure of merit

Before a good linear precoder can be designed, it is necessary to decide what good means. Empirical evidence [2] suggests linear precoders on their own perform much worse than linear precoders used in conjunction with other error correcting codes (such as convolutional codes). In fact, Part II [10] explains why this is so. Consequently, this paper defines a good precoder as one having the ability to achieve a low bit error rate (BER) if it is used in conjunction with a suitably powerful error correcting code at the transmitter and an optimal maximum likelihood (ML) detector at the receiver. That is to say, this paper considers the *intrinsic* performance of a precoder, which is conceptually very different from extrinsic indicators such as the achievable BER with respect to a specific coding and decoding algorithm.

It is assumed throughout that the receiver has perfect knowledge of the channel $\boldsymbol{h}$. This is a reasonable simplification because it is easier to make the channel estimation error small rather than the symbol estimation error small; optimal training sequences or pilot tones [8] facilitate the estimation of the channel, for instance. Define the CRB matrix [7]

$$R(\boldsymbol{h}) = (P^{\mathrm{H}} D C^{\mathrm{H}} \mathscr{H}^{\mathrm{H}} \mathscr{H} C D^{\mathrm{H}} P)^{-1} \tag{3}$$

which is a function of the channel vector $\boldsymbol{h}$. It is well-known the error covariance matrix $\boldsymbol{E}[(\boldsymbol{s}-\hat{\boldsymbol{s}})(\boldsymbol{s}-\hat{\boldsymbol{s}})^{\mathrm{H}}]$ of any unbiased estimate $\hat{\boldsymbol{s}}$ of $\boldsymbol{s}$ in (2), given $\boldsymbol{y}$ and $\boldsymbol{h}$, is lower bounded by $R(\boldsymbol{h})$. In fact, this lower bound is met with equality if the ML estimator

$$\hat{\boldsymbol{s}} = (P^{\mathrm{H}} D C^{\mathrm{H}} \mathscr{H}^{\mathrm{H}} \mathscr{H} C D^{\mathrm{H}} P)^{-1} P^{\mathrm{H}} D C^{\mathrm{H}} \mathscr{H}^{\mathrm{H}} \boldsymbol{y} \tag{4}$$

is used.

It is proposed to use $\operatorname{tr}\{R(\boldsymbol{h})\}$, the trace of the CRB, as a figure of merit for the following reasons. Assume $\boldsymbol{s}$ comes from a finite set Ω of codewords. By definition, the optimal ML detector outputs the element $\hat{\boldsymbol{s}}^{\mathrm{q}} \in \Omega$ which minimises the norm of the error vector $\boldsymbol{n}$ in (2); here, q denotes quantisation. In fact, referring to (4), $\hat{\boldsymbol{s}}^{\mathrm{q}}$ is the point in Ω closest to $\hat{\boldsymbol{s}}$ as measured by the weighted distance

$$d(\hat{\boldsymbol{s}}^{\mathrm{q}}, \hat{\boldsymbol{s}}) = (\hat{\boldsymbol{s}}^{\mathrm{q}} - \hat{\boldsymbol{s}})^{\mathrm{H}} R^{-1}(\boldsymbol{h}) (\hat{\boldsymbol{s}}^{\mathrm{q}} - \hat{\boldsymbol{s}}), \tag{5}$$

where $R^{-1}(\boldsymbol{h})$ is the inverse of the CRB matrix. Quantisation errors occur frequently if two or more codewords in Ω are too close to each other. The definition of $d(\cdot,\cdot)$ shows that, loosely speaking, the "larger" $R(\boldsymbol{h})$ is, the closer codewords become in Ω. Therefore, $R(\boldsymbol{h})$ measures the *intrinsic* ability of a linear precoder to reduce the overall BER of the system; the "smaller" $R(\boldsymbol{h})$ is, the easier it should be to design codeword constellations and decoding algorithms having low BERs.

The matrix $R(\boldsymbol{h})$ itself is not suitable as a figure of merit because matrices are not well-ordered; given two CRBs, it is not always clear which is better. The following two observations suggest it is appropriate to use $\operatorname{tr}\{R(\boldsymbol{h})\}$ as a figure of merit. In (4), it can be shown $\hat{\boldsymbol{s}} = \boldsymbol{s} + \tilde{\boldsymbol{n}}$ where $\tilde{\boldsymbol{n}} \sim N(0, R(\boldsymbol{h}))$. If the transmitter does not know $R(\boldsymbol{h})$, as is the case here, the channel capacity [1] is a function of the SNR alone, or (the reciprocal of) $\operatorname{tr}\{R(\boldsymbol{h})\}$ in this instance. (Note though that channel capacity can only be achieved if the number of elements of $\boldsymbol{s}$ goes to infinity.) The second observation is if two codewords are generated at random then $\operatorname{tr}\{R(\boldsymbol{h})\}$ gives a reasonable indication of the distance, as measured by (5), between them.

2.3. Optimal linear precoders

The channel dependent figure of merit $\operatorname{tr}\{R(\boldsymbol{h})\}$ introduced in Section 2.2 can be used to form a channel independent figure of merit $f : \mathbb{C}^{n\times p} \to \mathbb{R}$ assigning an overall figure of merit $f(P)$ to a precoder P. Two examples are the worst case MSE

$$f(P) = \sup_{\substack{\boldsymbol{h}\in\mathbb{C}^L \\ \|\boldsymbol{h}\|=1}} \operatorname{tr}\{R(\boldsymbol{h})\} \tag{6}$$

and the average MSE

$$f(P) = \int \operatorname{tr}\{R(\boldsymbol{h})\}\, p(\boldsymbol{h})\, \mathrm{d}\boldsymbol{h} \tag{7}$$

for some probability density function $p(\boldsymbol{h})$ of $\boldsymbol{h}$. (Since $\operatorname{tr}\{R(\boldsymbol{h})\}$ scales as $\|\boldsymbol{h}\|^{-2}$, the energy constraint $\|\boldsymbol{h}\|=1$ in (6) is the natural one to use. Here, the Euclidean norm is used.)

The optimal precoder design problem is to find a precoder P minimising an appropriate overall figure of merit; for this to be meaningful though, either a peak or an average energy constraint must be imposed on P. For reasons given in the remark below, the peak energy constraint $\lambda_{\max}\{P^{\mathrm{H}}P\} \leqslant 1$ is used, where $\lambda_{\max}\{P^{\mathrm{H}}P\}$ denotes the largest eigenvalue of $P^{\mathrm{H}}P$. Under this constraint, Theorem 1 shows it suffices to consider isometric precoders, meaning $P^{\mathrm{H}}P = I$ where I is the identity matrix.

Theorem 1. *Let $P_1 \in \mathbb{C}^{n\times p}$ be a linear precoder satisfying the energy constraint $\lambda_{\max}\{P_1^{\mathrm{H}}P_1\} \leqslant 1$. Let $R_1(\boldsymbol{h})$ denote the CRB matrix associated with P_1, as defined in (3). There exists a precoder $P \in \mathbb{C}^{n\times p}$ satisfying $P^{\mathrm{H}}P = I$ and such that its associated CRB matrix R satisfies $\operatorname{tr}\{R(\boldsymbol{h})\} \leqslant \operatorname{tr}\{R_1(\boldsymbol{h})\}$ for all channel vectors $\boldsymbol{h}$.*

Proof. Use the thin SVD [4] to decompose P_1 as $P_1 = USV^{\mathrm{H}}$ where U is an $n\times p$ matrix and S is diagonal. Then $\lambda_{\max}\{P_1^{\mathrm{H}}P_1\} \leqslant 1$ implies all the diagonal elements of S lie between 0 and 1 inclusively. Choose $P = U$. It is clear $P^{\mathrm{H}}P = I$ by construction, and furthermore, it is straightforward to show

$$\begin{aligned}\mathrm{tr}\{R_1(\boldsymbol{h})\} &= \mathrm{tr}\{(VSU^{\mathrm{H}}DC^{\mathrm{H}}\mathscr{H}^{\mathrm{H}}\mathscr{H}CD^{\mathrm{H}}USV^{\mathrm{H}})^{-1}\}\\ &\geqslant \mathrm{tr}\{(U^{\mathrm{H}}DC^{\mathrm{H}}\mathscr{H}^{\mathrm{H}}\mathscr{H}CD^{\mathrm{H}}U)^{-1}\}\\ &= \mathrm{tr}\{R(\boldsymbol{h})\}. \qquad \Box \qquad (8)\end{aligned}$$

Remark. Since the linear precoder acts on codewords rather than the raw message, in some ways it is more natural for the precoder to preserve peak energy rather than average energy. Nevertheless, the average energy constraint $\mathrm{tr}\{PP^{\mathrm{H}}\} \leqslant p$ is also sensible. However, it introduces secondary effects by allowing the precoder to weight the power distribution across sub-channels (a form of water-filling) [10, Section 4], and in particular, it complicates the analysis and is not considered here. It is expected though that an average energy constraint would not change qualitatively the results in this paper [10, Section 4]. Moreover, it is remarked that restricting attention to isometric precoders (preserving peak energy) has been justified under information theoretic criteria in [2,16].

Theorem 2 shows optimal linear precoders always exist provided the figure of merit $f(P)$ is continuous in P.

Theorem 2. *Let $f : \mathbb{C}^{n\times p} \to \mathbb{R}$ be a continuous function and let*

$$c = \inf_{\substack{P\in\mathbb{C}^{n\times p}\\ P^{\mathrm{H}}P=I}} f(P). \qquad (9)$$

There exists a P satisfying $P^{\mathrm{H}}P = I$ and such that $f(P) = c$.

Proof. Let $P_k \in \mathbb{C}^{n\times p}$ be a sequence of matrices all satisfying $P_k^{\mathrm{H}}P_k = I$ and such that $f(P_k) \to c$. Since the set of all matrices satisfying $P^{\mathrm{H}}P=I$ is compact (in the usual topology), there exists a P satisfying $P^{\mathrm{H}}P=I$ and such that $P_{k'} \to P$ for some subsequence k' of k. By continuity of f, $P_{k'} \to P$ implies $f(P_{k'}) \to f(P)$. Thus, $f(P) = c$, as required. $\Box$

Remark. If $f(P)$ depends on P via $\mathrm{tr}\{R(\boldsymbol{h})\}$ only, such as in (6) and (7), then $f(PQ) = f(P)$ for any unitary matrix Q. Minimising a function with this property subject to the constraint $P^{\mathrm{H}}P = I$ is equivalent to minimising a cost function on a Grassmann manifold [12].

3. Robust performance criteria

If the channel can be modelled as a random process then a natural figure of merit is the average MSE (7), as considered in Part II [10]. However, if the channel characteristics are unknown, an indication of the performance of a linear precoder can be based on its best and worst case performances, defined to be

$$f_1(P) = \inf_{\substack{\boldsymbol{h}\in\mathbb{C}^L\\ \|\boldsymbol{h}\|=1}} \mathrm{tr}\{R(\boldsymbol{h})\} \qquad (10)$$

and

$$f_2(P) = \sup_{\substack{\boldsymbol{h}\in\mathbb{C}^L\\ \|\boldsymbol{h}\|=1}} \mathrm{tr}\{R(\boldsymbol{h})\}, \qquad (11)$$

respectively. Here, $R(\boldsymbol{h})$ is the CRB matrix (3) associated with the precoder P. The constraint $\|\boldsymbol{h}\|^2=\boldsymbol{h}^{\mathrm{H}}\boldsymbol{h}=1$ is an energy constraint; $\mathrm{tr}\{R(\boldsymbol{h})\}$ scales as $\|\boldsymbol{h}\|^{-2}$.

According to Theorem 1, it suffices to consider precoders $P \in \mathbb{C}^{n\times p}$ for which $P^{\mathrm{H}}P = I$. In this case, $f_1(P) \leqslant p$ and $f_2(P) \geqslant p$ because $R(\boldsymbol{h})$ is the identity matrix when $\boldsymbol{h} = [1\ 0\ \cdots\ 0]^{\mathrm{T}}$. This motivates the following definitions.

Note: The definitions are with respect to a given channel length L.

Definition 3 (Maximally robust). A maximally robust precoder $P \in \mathbb{C}^{n\times p}$ is one which satisfies both $P^{\mathrm{H}}P = I$ and $f_2(P) = p$. That is, no other precoder has better worst case performance.

Definition 4 (Uniform). A uniform precoder $P \in \mathbb{C}^{n\times p}$ is one which satisfies both $P^{\mathrm{H}}P = I$ and $f_1(P) = f_2(P)$. That is, its best and worst case performances are the same.

Definition 5 (Strictly uniform). A strictly uniform precoder $P \in \mathbb{C}^{n\times p}$ is one which satisfies both $P^{\mathrm{H}}P=I$ and $R(\boldsymbol{h}) = I$ for all channels $\boldsymbol{h}$ having unit norm, where $R(\boldsymbol{h})$ is defined in (3).

These three definitions are successively stronger interpretations of “robust”. Definition 3 is only concerned with the worst case performance whereas Definition 4 requires the performance to be independent of the channel. Definition 5 is the strongest interpretation; a strictly uniform precoder eliminates multipath effects completely because the ML estimator (4) will have the same covariance matrix regardless of the channel $\boldsymbol{h}$. Somewhat surprisingly though, Section 4 proves these definitions are equivalent.

4. The equivalence of robust performance criteria

This section proves the equivalence of Definitions 3–5. In doing so, it is proved the figure of merit $\operatorname{tr}\{R(\boldsymbol{h})\}$, defined in (3), is convex as a function of the power spectrum of the channel $\boldsymbol{h}$.

Referring to (3), define

$$\Lambda(\boldsymbol{h}) = DC^{\mathrm{H}}\mathscr{H}^{\mathrm{H}}\mathscr{H}CD^{\mathrm{H}} \tag{12}$$

where $\mathscr{H}$, C and D are defined in Section 2.2. It is straightforward to show $\Lambda(\boldsymbol{h})$ is a diagonal matrix with kth diagonal element, indexed from zero, equal to

$$\Lambda_{kk}(\boldsymbol{h}) = |\boldsymbol{H}_k|^2, \quad \boldsymbol{H}_k = \sum_{i=0}^{L-1} \boldsymbol{h}_i \mathrm{e}^{-\jmath 2\pi(ik/n)} \tag{13}$$

for $k=0,\ldots,n-1$. In OFDM systems, $\boldsymbol{H}_k$ is referred to as the attenuation in the kth sub-channel, and indeed, the diagonal elements of Λ correspond to uniformly spaced samples of the power spectrum

$$\Gamma(\omega) = \left|\sum_{i=0}^{L-1} \boldsymbol{h}_i \mathrm{e}^{-\jmath\omega i}\right|^2 \tag{14}$$

of the channel $\boldsymbol{h} = [\boldsymbol{h}_0,\ldots,\boldsymbol{h}_{L-1}]^{\mathrm{T}}$. Note the figure of merit $\operatorname{tr}\{R(\boldsymbol{h})\}$ can be written in terms of Λ as

$$R(\boldsymbol{h}) = \operatorname{tr}\{(P^{\mathrm{H}}\Lambda(\boldsymbol{h})P)^{-1}\}. \tag{15}$$

The reason for introducing Λ is the figure of merit (15) turns out to be convex in Λ.

Theorem 6. *Define the set*

$$C = \{\Lambda \in \mathbb{R}^{n\times n}\colon \Lambda = \Lambda(\boldsymbol{h}),\ \boldsymbol{h}\in\mathbb{C}^L,\ \|\boldsymbol{h}\|^2 = 1\}, \tag{16}$$

where $\Lambda(\boldsymbol{h})$ is defined in (13). Then C is a compact convex set. Furthermore, the function $R : C \to \mathbb{R}$ defined by $R(\Lambda) = \operatorname{tr}\{(P^{\mathrm{H}}\Lambda P)^{-1}\}$, is convex.

Proof (taken from Manton [9]). It is a direct consequence of the spectral representation theorem that a function $\Gamma(\omega)$ is the power spectrum of an FIR channel of length at most L if and only if it can be written as

$$\Gamma(\omega) = \alpha_0 + \sum_{i=1}^{L-1} \alpha_i \cos(\omega i) + \beta_i \sin(\omega i), \quad \alpha_i, \beta_i \in \mathbb{R} \tag{17}$$

and is everywhere non-negative ($\forall\omega,\ \Gamma(\omega) \geqslant 0$). Furthermore, it follows from Parseval's theorem that the channel $\boldsymbol{h}$ has unit norm if and only if its power spectrum (14), when written in the form (17), has $\alpha_0 = 1$. Let $\boldsymbol{h}_1$ and $\boldsymbol{h}_2$ be any two channel vectors with unit norm, and let Γ_1 and Γ_2 be the corresponding power spectrums (14). For any given $0 \leqslant \lambda \leqslant 1$, define $\Gamma = \lambda\Gamma_1 + (1-\lambda)\Gamma_2$. Then, from (17) and Parseval's theorem, it follows that there exists an $\boldsymbol{h}$ with unit norm and such that its power spectrum is Γ. Since $\Lambda_{kk}(\boldsymbol{h}) = \Gamma(2\pi k/n)$, it is clear $\Lambda(\boldsymbol{h}) = \lambda\Lambda(\boldsymbol{h}_1) + (1-\lambda)\Lambda(\boldsymbol{h}_2) \in C$, proving C is convex. Clearly, C is compact since the non-negativity of (17) implies C is a closed and bounded set.

To prove the convexity of $R(\Lambda)$, it must be proved that

$$\forall\lambda\in(0,1), \quad g(\lambda) \leqslant \lambda g(1) + (1-\lambda)g(0), \tag{18}$$

where $g(\lambda) = \operatorname{tr}\{(P^{\mathrm{H}}(\lambda\Lambda_1 + (1-\lambda)\Lambda_2)P)^{-1}\}$ for arbitrary $\Lambda_1, \Lambda_2 \in C$. If either $g(0)$ or $g(1)$ is infinite, (18) holds by convention [20, Section 2.1]. Define $Z = (P^{\mathrm{H}}(\lambda\Lambda_1 + (1-\lambda)\Lambda_2)P)^{-1}$, assuming for the moment that the inverse exists. Note that Z is positive definite and Hermitian for $0 \leqslant \lambda \leqslant 1$. Then

$$\frac{1}{2}\frac{\mathrm{d}^2 g(\lambda)}{\mathrm{d}\lambda^2} = \operatorname{tr}\{ZP^{\mathrm{H}}(\Lambda_1 - \Lambda_2)PZP^{\mathrm{H}}(\Lambda_1 - \Lambda_2)PZ\} \tag{19}$$

$$\begin{aligned} &= \operatorname{tr}\{ZP^{\mathrm{H}}\Lambda_1 PZP^{\mathrm{H}}\Lambda_1 PZ \\ &\quad -2ZP^{\mathrm{H}}\Lambda_1 PZP^{\mathrm{H}}\Lambda_2 PZ \\ &\quad +ZP^{\mathrm{H}}\Lambda_2 PZP^{\mathrm{H}}\Lambda_2 PZ\} \end{aligned} \tag{20}$$

$$= \operatorname{tr}\{((A-B)Z)^{\mathrm{H}}((A-B)Z)\} \tag{21}$$

$$\geqslant 0, \tag{22}$$

where $A = Z^{1/2}P^{\mathrm{H}}\Lambda_1 P$, $B = Z^{1/2}P^{\mathrm{H}}\Lambda_2 P$ and $Z^{1/2}$ is any matrix such that $(Z^{1/2})^{\mathrm{H}}Z^{1/2} = Z$. This proves not only that $g(0)$ and $g(1)$ being finite implies Z is well-defined for $0 \leqslant \lambda \leqslant 1$, but that (18) always holds. $\square$

In light of Theorem 6, the best and worst performances of a precoder P, defined in (10) and (11), can be rewritten as

$$f_1(P) = \min_{\Lambda \in C} \operatorname{tr}\{(P^{\mathrm{H}} \Lambda P)^{-1}\},$$

$$f_2(P) = \max_{\Lambda \in C} \operatorname{tr}\{(P^{\mathrm{H}} \Lambda P)^{-1}\}. \tag{23}$$

Theorem 7. *Referring to the definitions in Section 3, the following three statements are equivalent:*

(1) *The precoder P is maximally robust.*
(2) *The precoder P is uniform.*
(3) *The precoder P is strictly uniform.*

Proof. If $L = 1$ the theorem holds trivially. Assume $L \geqslant 2$. It follows from (17) that $\Lambda = I$ is an interior point of the convex set C. Thus, the only way for the convex function $\operatorname{tr}\{(P^{\mathrm{H}} \Lambda P)^{-1}\}$ to achieve its maximum at $\Lambda = I$ is if it is everywhere constant, proving a maximally robust precoder is uniform. Let P be a uniform precoder and consider the function $g(\lambda)$ defined just after (18), with Λ_1 chosen arbitrarily and $\Lambda_2 = I$. Then, since P is uniform, $\left.\frac{\mathrm{d}^2 g}{\mathrm{d}\lambda^2}\right|_{\lambda=0} = 0$. Thus, (21) implies $(A - B)Z = 0$, or equivalently, $P^{\mathrm{H}} \Lambda_1 P = P^{\mathrm{H}} P = I$, proving P is strictly uniform. If P is strictly uniform it is clearly maximally robust, completing the proof. $\square$

5. Maximally robust precoders

This section derives necessary and sufficient conditions for a precoder to be maximally robust. Since Theorem 7 implies these conditions also apply to uniform and strictly uniform precoders, the results below are stated for strictly uniform precoders for convenience.

The following lemma provides a simple test for maximal robustness, or equivalently, for strict uniformity.

Lemma 8. *A linear precoder $P \in \mathbb{C}^{n \times p}$ satisfying $P^{\mathrm{H}} P = I$ is strictly uniform over channels up to length L if and only if $P^{\mathrm{H}} \operatorname{diag}\{c_i\} P = P^{\mathrm{H}} \operatorname{diag}\{s_i\} P = 0$ for $i = 1, \ldots, L-1$, where* diag *converts a vector into a diagonal matrix and*

$$\boldsymbol{c}_i = [\cos(0) \quad \cos(2\pi i/n) \quad \cos(2\pi 2i/n) \cdots \cos(2\pi(n-1)i/n)]^{\mathrm{T}}, \tag{24}$$

$$\boldsymbol{s}_i = [\sin(0) \quad \sin(2\pi i/n) \quad \sin(2\pi 2i/n) \cdots \sin(2\pi(n-1)i/n)]^{\mathrm{T}}. \tag{25}$$

Proof. By definition, P is strictly uniform if and only if $P^{\mathrm{H}} \Lambda P = I$ for all $\Lambda \in C$, where C is defined in (16). The lemma follows from the fact that, as can be seen from (17) and Parseval's theorem in the proof of Theorem 6, the affine hull of C is the set

$$I + \operatorname{span}\{\operatorname{diag}\{\boldsymbol{c}_1\}, \ldots, \operatorname{diag}\{\boldsymbol{c}_{L-1}\}, \operatorname{diag}\{\boldsymbol{s}_1\}, \ldots, \operatorname{diag}\{\boldsymbol{s}_{L-1}\}\}. \quad \square \tag{26}$$

The next lemma requires the vector

$$\boldsymbol{H} = [\boldsymbol{H}_0, \ldots, \boldsymbol{H}_{n-1}]^{\mathrm{T}} \in \mathbb{C}^n \tag{27}$$

where $\boldsymbol{H}_0, \ldots, \boldsymbol{H}_{n-1}$ are defined in (13).

Lemma 9. *If the precoder $P = [\boldsymbol{u}_1, \ldots, \boldsymbol{u}_p] \in \mathbb{C}^{n \times p}$ is strictly uniform then its columns satisfy*

$$\boldsymbol{u}_i \perp \boldsymbol{u}_j, \quad V_i \perp V_j, \quad A|\boldsymbol{u}_i|^2 = 0 \tag{28}$$

for $i, j = 1, \ldots, p$ with $i \neq j$. Here, $\perp$ denotes orthogonality ($\boldsymbol{x} \perp \boldsymbol{y}$ iff $\boldsymbol{x}^{\mathrm{H}} \boldsymbol{y} = 0$), $|\boldsymbol{u}_i|^2$ denotes the vector obtained from $\boldsymbol{u}_i$ by taking the square of the magnitude of each element,

$$V_i = \{\boldsymbol{y} \in \mathbb{C}^n \colon \boldsymbol{y} = \operatorname{diag}\{\boldsymbol{H}\}\boldsymbol{u}_i, \ \boldsymbol{h} \in \mathbb{C}^L\} \tag{29}$$

where $\boldsymbol{H}$ is defined in (27) and

$$A = [\, \boldsymbol{c}_1 \quad \cdots \quad \boldsymbol{c}_{L-1} \quad \boldsymbol{s}_1 \quad \cdots \quad \boldsymbol{s}_{L-1} \,]^{\mathrm{T}}, \tag{30}$$

where $\boldsymbol{c}_i$ and $\boldsymbol{s}_i$ are defined in (24) and (25). Furthermore, the sets V_i are vector spaces.

Proof. If P is strictly uniform then, by definition, $P^{\mathrm{H}} \Lambda P = P^{\mathrm{H}} P = I$ for all $\Lambda \in C$, where C is defined in (16). Thus $\boldsymbol{u}_i \perp \boldsymbol{u}_j$ for $i \neq j$. Furthermore, $\boldsymbol{u}_i^{\mathrm{H}} \Lambda \boldsymbol{u}_i = 1$ and $\boldsymbol{u}_j^{\mathrm{H}} \Lambda \boldsymbol{u}_i = 0$ for all $\Lambda \in C$ and $i \neq j$. From (13), $\Lambda = \operatorname{diag}\{\boldsymbol{H}\}^{\mathrm{H}} \operatorname{diag}\{\boldsymbol{H}\}$. The condition $\boldsymbol{u}_j^{\mathrm{H}} \operatorname{diag}\{\boldsymbol{H}\}^{\mathrm{H}} \operatorname{diag}\{\boldsymbol{H}\} \boldsymbol{u}_i = 0$ for all $\boldsymbol{h}$ with unit norm is sufficient to imply that $V_i \perp V_j$, even though there is no norm constraint in (29). That the V_i are vector spaces follows from $\boldsymbol{H}$ being linear in $\boldsymbol{h}$; see (13).

Since $P^{\mathrm{H}}P=I$, the constraint $\boldsymbol{u}_i^{\mathrm{H}}\Lambda\boldsymbol{u}_i=1$ is equivalent to $\boldsymbol{u}_i^{\mathrm{H}}(\Lambda-I)\boldsymbol{u}_i=0$, and furthermore, this can be written as $\sum_{k=1}^{n}\lambda_k|(\boldsymbol{u}_i)_k|^2=0$, where λ_k is the kth diagonal element of $(\Lambda-I)$ and $(\boldsymbol{u}_i)_k$ is the kth element of $\boldsymbol{u}_i$. This, together with the fact that the affine hull of C is the set (26), proves that $\boldsymbol{u}_i^{\mathrm{H}}(\Lambda-I)\boldsymbol{u}_i=0$ for all $\Lambda\in C$ implies $A|\boldsymbol{u}_i|^2=0$. □

The usefulness of Lemma 9 is it makes precise the following three intuitive requirements for a precoder to be strictly uniform. Note first that since the symbols $\boldsymbol{s}$ are transmitted as $P\boldsymbol{s}$, the ith column $\boldsymbol{u}_i$ dictates how the ith symbol is sent. The total received energy must be independent of the shape of the channel spectrum; for the ith symbol, this requires $A|\boldsymbol{u}_i|^2=0$. The transmitted symbols must start off orthogonal; this requires $\boldsymbol{u}_i\perp\boldsymbol{u}_j$ for $i\neq j$. Lastly, the received symbols must remain orthogonal regardless of the channel; this requires $V_i\perp V_j$ for $i\neq j$.

Theorem 10. *A necessary condition for the linear precoder $P\in\mathbb{C}^{n\times p}$ to be either maximally robust or strictly uniform over channels up to length L is for $n\geqslant pL$.*

Proof. Define V_i as in (29). It will be proved that if P is strictly uniform then $\dim V_i=L$. Since the $V_i\subset\mathbb{C}^n$ for $i=1,\ldots,p$ are mutually orthogonal (see Lemma 9), their union will thus span pL dimensions. This is only possible if $n\geqslant pL$, proving the theorem.

First, it is proved that $\boldsymbol{u}_i$, the ith column of P, contains at least L non-zero elements. Assume to the contrary that it contains between 1 and $L-1$ non-zero elements. (It cannot be the zero vector because $P^{\mathrm{H}}P=I$.) Since $A|\boldsymbol{u}_i|^2=0$ in Lemma 9, this implies there exist $L-1$ columns of A which are linearly dependent. Assume these columns, indexed from zero, are numbered $j_1,\ldots,j_{L-1}$, and define $\lambda_k=\mathrm{e}^{\jmath 2\pi j_k/n}$. Then, by expressing the elements of A as the sums of exponentials, it can be shown that the dependence of $L-1$ columns of A implies the square Vandermonde matrix generated by $\lambda_1,\ldots,\lambda_{L-1}$ is rank deficient. However, since the λ_k are distinct, this is not possible. Thus, $\boldsymbol{u}_i$ must have at least L non-zero elements.

Finally, it is proved that $\boldsymbol{u}_i$ having at least L non-zero elements implies $\dim V_i=L$. Let $D\in\mathbb{C}^{n\times L}$ be the truncated discrete Fourier transform matrix, so that $\boldsymbol{H}=D\boldsymbol{h}$; see (13). Then V_i in (29) consists of all vectors of the form $\operatorname{diag}\{\boldsymbol{u}_i\}D\boldsymbol{h}$. It has dimension L unless there exists a non-zero $\boldsymbol{h}$ such that $\operatorname{diag}\{\boldsymbol{u}_i\}D\boldsymbol{h}=0$. Assume such an $\boldsymbol{h}$ exists. Then, since $\boldsymbol{u}_i$ has at least L non-zero elements, there exists an $L\times L$ sub-matrix $\tilde{D}$ of D such that $\tilde{D}\boldsymbol{h}=0$. However, $\tilde{D}$ is a square Vandermonde matrix with distinct generators and thus $\tilde{D}\boldsymbol{h}=0$ implies $\boldsymbol{h}=0$, a contradiction. □

Theorem 11 exhibits a class of maximally robust precoders introducing the least amount of redundancy. This class is discussed further in Section 6.

Theorem 11. *Precoders of the form $P=L^{-1/2}(\mathbf{1}_L\otimes I_p)$, where $\mathbf{1}_L$ is the column vector of L ones, I_p is the $p\times p$ identity matrix and $\otimes$ is Kronecker's product, are strictly uniform over channels up to length L. Moreover, they introduce the least amount of redundancy necessary to achieve either maximal robustness or strict uniformity.*

Proof. It is straightforward to verify the conditions in Lemma 8 are satisfied. Furthermore, the size of P satisfies the lower bound in Theorem 10. □

6. Minimally redundant maximally robust precoders

In [8] it was proved that sending L equally spaced pilot tones allows the receiver to identify the channel with the same accuracy regardless of the shape of the channel spectrum. This is because the total received energy over L equally spaced sub-channels, given by

$$\sum_{j=0}^{L-1}|\boldsymbol{H}_{k+jp}|^2=L\|\boldsymbol{h}\|^2,\qquad k=0,\ldots,p-1,\tag{31}$$

depends only on the magnitude $\|\boldsymbol{h}\|^2$ of the channel. (In (31), $\boldsymbol{H}$ is as defined in (27) with $n=pL$.)

It is apparent from (31) that if the same symbol is sent over L equally spaced sub-channels then, because the total received energy is independent of the shape of the channel spectrum, the receiver is able to estimate the symbol in AWGN with the same accuracy regardless of the shape of the channel spectrum. The precoder $P=L^{-1/2}(\mathbf{1}_L\otimes I_p)$ in Theorem 11 does just this; it transmits each symbol on L equally spaced sub-channels.

As well as having a simple frequency domain structure, precoders of the form $P = L^{-1/2}(\mathbf{1}_L \otimes I_p)$ have a simple time domain structure. It can be shown that $D^{\mathrm{H}}P\boldsymbol{s}$, the inverse discrete Fourier transform of $P\boldsymbol{s}$ (see Section 2.1), has $L-1$ consecutive zeros between each non-zero element. These zeros eliminate inter-symbol interference.

Unfortunately then, maximally redundant precoders are unattractive from a practical point of view. From a theoretical perspective though, Theorems 7 and 10 establish the important result that any precoder introducing less than $(L-1)p$ redundant symbols must trade off worst case performance for this decrease in redundancy.

7. Summary

This paper presented the first half of a theoretical study of the linear precoder design problem. Section 2 defined an intrinsic figure of merit for a linear precoder and introduced a new way of formulating mathematically the linear precoder design problem. Starting from Section 3, attention was restricted to robust linear precoders, where three successively stronger definitions of robust precoders were given. Somewhat surprisingly, it was shown in Section 4 that these definitions are equivalent. Robust precoders so defined were referred to as either maximally robust precoders or strictly uniform precoders, the former name referring to their best worst case performance and the latter to their ability to eliminate multipath effects completely. Necessary and sufficient conditions for a precoder to be maximally robust were given in Section 5 along with explicit expressions for a class of maximally robust precoders introducing the least amount of redundancy. Section 6 explained how maximally robust precoders achieve their robustness. Part II [10] continues this study and makes concluding remarks.

References

[1] T.M. Cover, J.A. Thomas, Elements of Information Theory, Wiley, New York, 1991.

[2] M. Debbah, W. Hachem, P. Loubaton, M. de Courville, Asymptotic performance analysis for redundant block precoded OFDM systems with MMSE equalization. in: IEEE Workshop on Statistical Signal Processing, Singapore, 2001, pp. 389–392.

[3] G.B. Giannakis, Filterbanks for blind channel identification and equalization, IEEE Signal Process. Lett. 4 (6) (1997) 184–187.

[4] G.H. Golub, C.F. Van Loan, Matrix Computations, 3rd Edition, The Johns Hopkins University Press, Baltimore, MD, 1996.

[5] B. Le Floch, M. Alard, C. Berrou, Coded orthogonal frequency division multiplex, Proc. IEEE 83 (6) (1995) 982–996.

[6] T. Li, Z. Ding, Joint transmitter–receiver optimization for partial response channels based on nonmaximally decimated filterbank precoding technique, IEEE Trans. Signal Process. 47 (9) (1999) 2407–2418.

[7] J.H. Manton, Dissecting OFDM: the independent roles of the cyclic prefix and the IDFT operation, IEEE Commun. Lett. 5 (12) (2001) 474–476.

[8] J.H. Manton, Optimal training sequences and pilot tones for OFDM systems, IEEE Commun. Lett. 5 (4) (2001) 151–153.

[9] J.H. Manton, The convex geometry of sub-channel attenuation coefficients in linearly precoded OFDM systems, IEEE Trans. Inform. Theory 48 (5) (2002) 1203–1206.

[10] J.H. Manton, Design and analysis of linear precoders under a mean square error criterion, Part II: MMSE designs and conclusions, Systems Control Lett. 49 (2003) 131–140, in this issue.

[11] J.H. Manton, On the importance of using a cyclic prefix or null guard in block transmissions over FIR channels, in: Proceedings of the Information, Decision and Control, Adelaide, Australia, February, 2002.

[12] J.H. Manton, Optimisation algorithms exploiting unitary constraints, IEEE Trans. Signal Process. 50 (3) (2002) 635–650.

[13] J.H. Manton, Y. Hua, A frequency domain deterministic approach to channel identification, IEEE Signal Process. Lett. 6 (12) (1999) 323–326.

[14] J.H. Manton, Y. Hua, Robust frequency domain precoders, IEEE Commun. Lett. 5 (2) (2001) 40–42.

[15] J.H. Manton, I.M.Y. Mareels, Y. Hua, Affine precoders for reliable communications, in: IEEE Conference on Acoustics, Speech and Signal Processing, Vol. V, Istanbul, Turkey, June, 2000, pp. 2749–2752.

[16] T.L. Marzetta, B.M. Hochwald, Capacity of a mobile multiple-antenna communication link in Rayleigh flat fading, IEEE Trans. Inform. Theory 45 (1) (1999) 139–157.

[17] J.G. Proakis, Digital Communications, 3rd Edition, McGraw-Hill, New York, 1995.

[18] A. Scaglione, S. Barbarossa, G.B. Giannakis, Filterbank transceivers optimizing information rate in block transmissions over dispersive channels, IEEE Trans. Inform. Theory 45 (3) (1999) 1019–1032.

[19] A. Scaglione, G.B. Giannakis, S. Barbarossa, Redundant filterbank precoders and equalizers, Part I: unification and optimal designs, IEEE Trans. Signal Process. 47 (1999) 1988–2006.

[20] H. Tuy, Convex Analysis and Global Optimization, Kluwer Academic Publishers, Dordrecht, 1998.

[21] Z. Wang, G.B. Giannakis, Wireless multicarrier communications: where Fourier meets Shannon, IEEE Signal Process. Mag. 17 (2000) 29–48.

[22] Z. Wang, G.B. Giannakis, Linearly precoded or coded OFDM against channel fades? in: IEEE Workshop on Signal Processing Advances in Wireless Communications, Taiwan, March, 2001.
[23] S.B. Weinstein, P.M. Ebert, Data transmission by frequency division multiplexing using the discrete Fourier transform, IEEE Trans. Commun. 19 (5) (1971) 628–634.
[24] G.W. Wornell, Spread-response precoding for communication over fading channels, IEEE Trans. Inform. Theory 42 (2) (1996) 488–501.
[25] X.-G. Xia, New precoding for intersymbol interference cancellation using nonmaximally decimated multirate filterbanks with ideal FIR equalizers, IEEE Trans. Signal Process. 45 (1997) 2431–2441.