Available online at www.sciencedirect.com
SCIENCE DIRECT®
Systems & Control Letters 49 (2003) 131 – 140
SYSTEMS & CONTROL LETTERS
www.elsevier.com/locate/sysconle

Design and analysis of linear precoders under a mean square error criterion, Part II: MMSE designs and conclusions

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

Part I of this two-part paper formulated the precoder design problem as an optimisation problem and solved it with respect to a worst case criterion. Part II studies a similar optimisation problem but with respect to a criterion measuring average performance. A stochastic optimisation algorithm is proposed for solving this problem. For special cases, closed form solutions are also given. These results indicate linear precoders reduce the effects of frequency distortion caused by multipath channels but are powerless to counteract additive white Gaussian noise. The conclusion is linear precoders should introduce only a small amount of redundancy and be used in conjunction with an error correcting code capable of combatting additive noise.

Keywords: Channel coding; Linear precoders; Multipath channels; Fading channels; Orthogonal frequency division multiplex (OFDM) systems; Zero padded systems; Wireless communications

1. Introduction

Redundant linear precoders [4,5,9,17,20,24] sometimes in conjunction with error correcting codes (ECC), have been proposed for transmitting digital data over multipath channels. This paper takes a step closer to resolving two open issues: how best to design the linear precoder and what sort of ECC should be used in conjunction with the linear precoder?

Part I [12] argued the trace of a certain Cramer–Rao Bound matrix, denoted $\operatorname{tr}\{R(\boldsymbol{h})\}$ since it depends on the channel $\boldsymbol{h}$, measures the *intrinsic* performance of a precoder, meaning it refers to the performance under the idealistic assumptions that each block of data is processed in isolation and a sufficiently powerful ECC coupled with an optimal maximum likelihood detector with perfect knowledge of the channel is used. This paper optimises $\boldsymbol{E}[\operatorname{tr}\{R(\boldsymbol{h})\}]$ where expectation is with respect to the random channel $\boldsymbol{h}$. It is shown that optimal solutions, called minimum mean square error (MMSE) precoders, reduce the effects of frequency distortion caused by multipath propagation. In other words, on the per-block basis considered here, a MMSE linear precoder makes the multipath channel resemble an additive white Gaussian noise (AWGN) channel. It is therefore appropriate to use an ECC designed for AWGN channels in conjunction with a MMSE linear precoder, as illustrated in Fig. 1. It is candidly stated though that Fig. 1 is not an optimal coding solution but a compromise between design simplicity, computational complexity and performance. It

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

0167-6911/03/$ - see front matter
PII: S0167-6911(02)00317-1

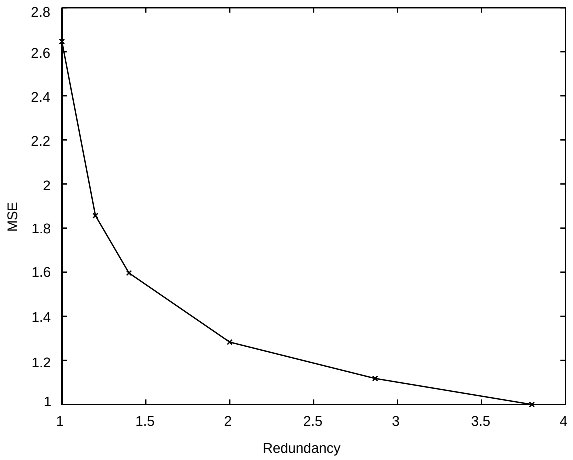

Fig. 1. The proposed two-stage coding scheme for communications over multipath channels. The MMSE linear precoder reduces the effects of multipath by spreading the symbols in the frequency domain. The outer ECC protects against additive noise.

is also emphasised that the use of a MMSE precoder in Fig. 1 may not be suitable if the ECC codes over multiple blocks; see [12, Section 2.1].

The inclusion of the scheme in Fig. 1 is merely to illustrate several theoretical findings. Rather, the main contributions of this paper are:

- The derivation of a stochastic optimisation algorithm for computing MMSE linear precoders.
- The derivation of closed form expressions for MMSE linear precoders in special cases.
- The conclusion that linear precoders cannot combat AWGN and should therefore introduce only a small amount of redundancy and be used in conjunction with an AWGN combatting ECC.

To put these contributions in perspective, the current status of the linear precoder design problem for multipath channels is summarised. Zero padded (ZP) systems [21] (such as trailing zero orthogonal frequency division multiplex systems and ZP only systems), which append a sequence of zeros to each transmitted block, have been shown to work better than simple ECC [22]. This suggests performance can be enhanced further by using a linear precoder introducing more redundancy. Although several design criteria for such redundant linear precoders were proposed in [17,18,22], no satisfactory algorithms for finding optimal precoders under these criteria were given. This may have been the motivation for randomly generated precoders being considered in [4]. The stochastic optimisation algorithm in Section 3.3 is believed to be the first algorithm capable of finding optimal precoders for random channels. Moreover, optimal precoders differ substantially from the randomly chosen ones in [4]. Fig. 4 is a histogram of the MSE of 1000 randomly generated isometric precoders [4] of size 21×15, showing a mean of 1.76 and a standard deviation of 0.02. By comparison, the optimal isometric precoder found by the algorithm in Section 3.3 has a MSE of 1.60, which is 8 standard deviations away from the mean. As the redundancy increases, the relative difference between random and optimal precoders appears to increase; for isometric precoders of size 43×15, the minimum MSE was 150 standard deviations away from the mean. This suggests good precoders cannot be constructed at random.

Other related work includes [9] and [19] where optimal linear precoders for *known* channels were proposed. This is a special case of the problem considered here because a random channel reduces to a known channel if it is constant with probability one. Section 4 compares the results in [9,19] with the results here.

Other approaches to the channel coding problem are taken in [2,8,23] to name just a few. Since these approaches exploit the ergodic nature of the random channel by spreading the source symbols over time, they are likely to perform better than the scheme in Fig. 1. It is therefore mentioned again that the present paper focuses on the case when the source symbols are broken into short blocks (relative to the channel coherence time) and each block is separately encoded and transmitted; see [12, Section 2.1].

The outline of this paper is as follows. Section 2 elaborates on the precoder design problem formulated in Part I, placing more emphasis on ZP systems because the distinction is more important here than in Part I. Section 3 derives a lower bound on the average performance of a linear precoder operating over a random channel. Closed form expressions for precoders meeting this lower bound are given. An algorithm for finding MMSE precoders in situations where the lower bound cannot be met is also presented. For completeness, Section 4 compares MMSE precoders with other optimal precoders when the channel is known. The implications of the theoretical results in Parts I and II are discussed in Section 5 and compared with existing results in the literature. Section 6 concludes the paper.

2. The linear precoder design problem

This section complements the formulation of the linear precoder design problem in Section 2 of Part I [12]. Minor notational changes are made to accommodate the greater emphasis placed here on ZP systems.

2.1. Block transmission systems

Two linear block transmission systems have received considerable attention in the literature; orthogonal frequency division multiplex (OFDM) systems and ZP systems [21]. This paper considers only these systems because a cyclic prefix or zero padding is required if the system is to have a reasonable MSE over almost any channel [10,14,12]. Fortuitously, using a cyclic prefix or zero padding also results in a mathematical simplification of the precoder design problem because it leads to a natural convex geometry [11,12].

A *cyclic prefixed system* breaks the source symbols $\{s_k\}_{k=-\infty}^{\infty}$ into blocks $\boldsymbol{s}^{(i)} = [s_{ip}, s_{ip+1}, \dots, s_{ip+p-1}]^{\mathrm{T}} \in \mathbb{C}^p$ of length p. A linear precoder matrix $P \in \mathbb{C}^{n\times p}$ (where $n \geqslant p$) then encodes each block. Finally, assuming the system operates over a finite impulse response (FIR) channel $\boldsymbol{h} = [h_0, \dots, h_{L-1}]^{\mathrm{T}} \in \mathbb{C}^L$ of length at most L, a cyclic prefix of length $L-1$ is added to each encoded block $Ps^{(i)}$. The ith received block is thus

$$\boldsymbol{y}^{(i)} = \mathscr{H}_{(\mathrm{cp})} P \boldsymbol{s}^{(i)} + \boldsymbol{n}^{(i)} \in \mathbb{C}^n, \tag{1}$$

where $\mathscr{H}_{(\mathrm{cp})}$ is the $n \times n$ circulant matrix whose first row is $[h_0, 0, \dots, 0, h_{L-1}, \dots, h_1]$ and $\boldsymbol{n}^{(i)} \in \mathbb{C}^n$ is AWGN.

A *ZP system* is defined analogously to a cyclic prefixed system except a sequence of $L-1$ zeros is appended to each transmitted block instead of a cyclic prefix. The ith received block is thus

$$\boldsymbol{y}^{(i)} = \mathscr{H}_{(\mathrm{zp})} P \boldsymbol{s}^{(i)} + \boldsymbol{n}^{(i)} \in \mathbb{C}^{n+L-1}, \tag{2}$$

where $\mathscr{H}_{(\mathrm{zp})}$ is the $(n+L-1)\times n$ lower triangular Toeplitz matrix with first column $[h_0, h_1, \dots, h_{L-1}, 0, \dots, 0]^{\mathrm{T}}$ and $\boldsymbol{n}^{(i)} \in \mathbb{C}^{(n+L-1)}$ is AWGN. Observe the received vector $\boldsymbol{y}^{(i)}$ is longer than the transmitted vector $P\boldsymbol{s}^{(i)}$, a consequence of the receiver not having to discard the guard interval as in a cyclic prefixed system [13,21].

To obtain a simple yet meaningful measure of the accuracy with which $\boldsymbol{s}$ can be recovered from $\boldsymbol{y}$, it is assumed throughout that the receiver (but not the transmitter) has perfect knowledge of the channel $\boldsymbol{h}$. In this case, and dropping the block index i, the maximum likelihood estimate (MLE) of $\boldsymbol{s}$ given $\boldsymbol{y}$ in (1) is

$$\hat{\boldsymbol{s}} = (P^{\mathrm{H}} \mathscr{H}_{(\mathrm{cp})}^{\mathrm{H}} \mathscr{H}_{(\mathrm{cp})} P)^{-1} P^{\mathrm{H}} \mathscr{H}_{(\mathrm{cp})}^{\mathrm{H}} \boldsymbol{y}. \tag{3}$$

This is also the minimum variance unbiased estimate of $\boldsymbol{s}$, and moreover, its error covariance matrix achieves the Cramer–Rao Bound. Indeed, the error covariance matrix is

$$R(\boldsymbol{h}) = \boldsymbol{E}[(\hat{\boldsymbol{s}} - \boldsymbol{s})(\hat{\boldsymbol{s}} - \boldsymbol{s})^{\mathrm{H}}] \tag{4}$$

$$= (P^{\mathrm{H}} \mathscr{H}_{(\mathrm{cp})}^{\mathrm{H}} \mathscr{H}_{(\mathrm{cp})} P)^{-1}. \tag{5}$$

The ith diagonal element of $R(\boldsymbol{h})$ is the lowest possible variance with which the ith source symbol can be estimated without any bias.

For reasons given in Part I, this paper uses the Cramer–Rao Bound ($P^{\mathrm{H}} \Lambda P^{-1}$) to measure the intrinsic performance of the linear precoder P where $\Lambda = \mathscr{H}_{(\mathrm{cp})}^{\mathrm{H}} \mathscr{H}_{(\mathrm{cp})}$ for a cyclic prefixed system and $\Lambda = \mathscr{H}_{(\mathrm{zp})}^{\mathrm{H}} \mathscr{H}_{(\mathrm{zp})}$ for a ZP system. Throughout, $\boldsymbol{h}$ and Λ are both referred to as the channel.

A ZP system is best studied via its associated cyclic prefixed system [13]; for any ZP system using the $n\times p$ precoder P, there exists a cyclic prefixed system using the $(n+L-1)\times p$ precoder P such that both systems have identical performance, that is,

$$(P^{\mathrm{H}} \mathscr{H}_{(\mathrm{zp})}^{\mathrm{H}} \mathscr{H}_{(\mathrm{zp})} P)^{-1} = (P^{\mathrm{H}} \mathscr{H}_{(\mathrm{cp})}^{\mathrm{H}} \mathscr{H}_{(\mathrm{cp})} P)^{-1} \tag{6}$$

holds for all channels $\boldsymbol{h}$.

3. MMSE linear precoders for random channels

The per-symbol performance of a linear precoder $P \in \mathbb{C}^{n\times p}$ operating over a random channel Λ is defined as

$$f(P) = \frac{1}{p} \boldsymbol{E}[\mathrm{tr}\{(P^{\mathrm{H}} \Lambda P)^{-1}\}]. \tag{7}$$

Part I [12, Section 2] explains why the trace is used in (7) and $1/p$ normalises the total MSE to a per-symbol MSE, thus allowing comparisons between precoders of different sizes. The probability distribution of $\boldsymbol{h}$, and thus of Λ, is arbitrary unless otherwise stated. Two examples are $\boldsymbol{h} \sim N(0, I)$ if nothing is known about the channel and $\boldsymbol{h} \sim N(\bar{\boldsymbol{h}}, \sigma^2 I)$ if an estimate $\bar{\boldsymbol{h}}$ of the channel $\boldsymbol{h}$, accurate to within a variance of σ^2, is somehow available.

Since $\mathrm{tr}\{(P^{\mathrm{H}}\Lambda P)^{-1}\}$ scales as $\|\boldsymbol{h}\|^{-2}$ and the distribution $\boldsymbol{h} \sim N(0, I)$ is symmetric, evaluating $f(P)$ when $\boldsymbol{h} \sim N(0, I)$ is equivalent to evaluating it when $\boldsymbol{h}$ is uniformly distributed on the unit ball $\|\boldsymbol{h}\|^2 = 1$. Therefore, rather than studying a Rayleigh fading channel $\boldsymbol{h} \sim N(0, I)$, it suffices to study a *normalised Rayleigh fading channel* obtained by first sampling from a $N(0, I)$ distribution and then normalising the result ($\boldsymbol{h} := \boldsymbol{h}/\|\boldsymbol{h}\|$). Moreover, for any reasonable probability distribution, there exists a distribution on the unit ball such that evaluating $f(P)$ with respect to either distribution gives the same answer. This motivates the first assumption below; the second is required to rule out channels for which one or more elements of $\boldsymbol{h}$ are constant with probability one.

A1 The random channel $\boldsymbol{h}$ is such that $\|\boldsymbol{h}\| = 1$ almost surely.

A2 The support of $\Lambda = \mathscr{H}^{\mathrm{H}}_{(\mathrm{cp})}\mathscr{H}_{(\mathrm{cp})}$, defined to be the set of all Λ having non-zero probability of occurring, contains a convex set of the largest possible dimension. (For a length L channel having unit norm, the largest dimension is $2(L-1)$; this is proved in [11] but can be deduced also from the proof of Theorem 6 in Part I.)

It is re-iterated (A1) is not restrictive in any sense.

3.1. Performance bounds

This section uses the convexity of $f(P)$ in (7) to derive bounds on the performance of a MMSE linear precoder.

Theorem 1. *The function $f(P)$ in (7) satisfies the lower bound*

$$f(P) \geqslant \frac{1}{p}\mathrm{tr}\{(P^{\mathrm{H}}\bar{\Lambda}P)^{-1}\}, \quad \bar{\Lambda} = \boldsymbol{E}[\Lambda] \tag{8}$$

in general. Furthermore, if (A1) and (A2) hold, equality holds in (8) if and only if $\mathrm{tr}\{(P^{\mathrm{H}}\Lambda P)^{-1}\}$ is constant with probability one.

Proof. Since $g(\Lambda) = \mathrm{tr}\{(P^{\mathrm{H}}\Lambda P)^{-1}\}$ is a convex function [12], $\boldsymbol{E}[g(\Lambda)] \geqslant g(\boldsymbol{E}[\Lambda])$. Furthermore, equality holds if and only if $g(\Lambda)$ is an affine function almost surely. The proof of Theorem 7 in Part I shows $g(\Lambda)$ is strictly convex unless it is constant. The proof now follows since (A1) and (A2) hold. □

Recall from Part I that P is called an isometric precoder if $P^{\mathrm{H}}P = I$.

Corollary 2. *Over a normalised Rayleigh fading channel, no isometric precoder has a per-symbol MSE better than unity, that is, $f(P) \geqslant 1$.*

Proof. Since $P^{\mathrm{H}}P = I$ and $\bar{\Lambda} = I$, $f(P) \geqslant 1$ in (8). □

Theorem 3. *Assume the length L random channel $\boldsymbol{h}$ satisfies (A1) and (A2). A necessary condition for $f(P)$ to achieve the lower bound in (8) is for $n \geqslant pL$ in a cyclic prefixed system, or for $n \geqslant pL - L + 1$ in a ZP system. Another necessary condition is $P^{\mathrm{H}}\Lambda P^{-1} = I$ for any channel $\boldsymbol{h}$ having unit norm.*

Proof. Consider first a cyclic prefixed system. If equality holds in (8) then Theorem 1 states $\mathrm{tr}\{(P^{\mathrm{H}}\Lambda P)^{-1}\}$ is constant with probability one. Precoders satisfying this property are called uniform precoders in Part I and it was proved the necessary conditions for a precoder to be uniform are as given here. For ZP systems, if P achieves the lower bound in (8) then the associated cyclic prefixed system $\tilde{P}$ in (6) must also achieve the lower bound. Since $\tilde{P}$ has $L-1$ more rows than P, the theorem follows. □

Since $(P^{\mathrm{H}}\Lambda P)^{-1}$ is the error covariance matrix at the output of the linear equaliser (3) depicted in Fig. 1, if $(P^{\mathrm{H}}\Lambda P)^{-1} = I$ for all channels $\boldsymbol{h}$ having unit norm then the precoder has eliminated the frequency distortion caused by the memory of the channel. Part I calls such precoders strictly uniform.

It is reasonable to expect the MSE of a MMSE linear precoder to decrease as the amount of redundancy increases. For $n \times p$ MMSE linear precoders, it is straightforward to prove the MSE cannot increase if

p decreases while n stays fixed. However, if p is held fixed and n increases (such as in Fig. 5), the size and hence the geometry of the channel Λ changes. It is therefore difficult to prove (and may not be true) that increasing n by one causes the MSE to decrease. However, since the next section shows the lower bound in (8) is achievable if n is sufficiently large, it is clear the MSE must decrease if n is increased sufficiently.

3.2. Precoders achieving the lower bound

Theorem 4 is analogous to Theorem 11 in Part I. Its trivial proof is omitted.

Theorem 4. *For any cyclic prefixed system operating over a length L channel, precoders of the form*

$$P = I_p \otimes \boldsymbol{e}_1, \tag{9}$$

where p is an arbitrary positive integer, I_p is the $p \times p$ identity matrix, $\otimes$ is Kronecker's product and *e_1 is the length L column vector $[1, 0, \ldots, 0]^{\mathrm{T}}$, satisfy $(1/p)\mathrm{tr}\{(P^{\mathrm{H}}\Lambda P)^{-1}\} = (1/n)\mathrm{tr}\{\Lambda\}$. Thus (9) achieves the lower bound in (8) regardless of the probability distribution of the channel. Furthermore, the amount of redundancy meets the lower bound in Theorem* 3. *For a ZP system, any precoder obtained by omitting the last $L-1$ rows of a precoder in (9) also meets both bounds with equality.*

Precoders of the form (9) transmit $L-1$ zeros after each symbol and are therefore unattractive coding schemes. The importance of Theorem 4 is it gives the limiting form of MMSE linear precoders as $n/p \to L$.

3.3. An algorithm for computing MMSE precoders

This section derives an algorithm for computing MMSE linear precoders. The algorithm converges to a local minimum of $f(P)$ in (7) subject to the peak power constraint $P^{\mathrm{H}}P = I$. (The algorithm is readily modified to use the average power constraint $\{PP^{\mathrm{H}}\} = p$ instead.) Although the theory only guarantees a local minimum is found, empirical evidence suggests this is not a problem in practice.

The difficulty in minimising $\boldsymbol{E}[\mathrm{tr}\{(P^{\mathrm{H}}\Lambda P\}]$ is the expectation does not have a closed form expression in general. It was proposed in [17] to approximate the expectation by a finite summation over a set of test channels and apply the optimisation technique in [15]. This section a more natural approach based on the theory of stochastic optimisation [1].

Stochastic gradient algorithms replace the gradient of the cost function by a stochastic approximation. Under mild conditions [1], a cost function of the form $\boldsymbol{E}[g(P;\Lambda)]$, where expectation is with respect to Λ, is minimised by the iteration

$$P^{(k+1)} = P^{(k)} - \frac{1}{k}\left.\frac{\partial g(P;\Lambda^{(k)})}{\partial P}\right|_{P=P^{(k)}}, \quad k = 1, 2, \ldots, \tag{10}$$

where $\Lambda^{(k)}$ is a sequence of independently generated random realisations of the channel Λ and $1/k$ is interpreted as a decreasing step size. The decreasing step size means the trajectory of (10) eventually "follows closely" the trajectory of the associated ODE [1]

$$\frac{\mathrm{d}P}{\mathrm{d}t} = \boldsymbol{E}\left[-\frac{\partial g(P;\Lambda)}{\partial P}\right]. \tag{11}$$

Swapping the expectation and partial derivative operators shows (11) is a gradient flow converging to a local minimum of $\boldsymbol{E}[g(P;\Lambda)]$, as required. This justifies interpreting the partial derivative term in (10) as an approximation of the true gradient of the cost function $\boldsymbol{E}[g(P;\Lambda)]$.

Choosing $g(P;\Lambda) = \mathrm{tr}\{P^{\mathrm{H}}\Lambda P\}$ in (10) leads to an algorithm for constructing MMSE linear precoders if the constraint $P^{\mathrm{H}}P = I$ is enforced. This is done [1] by constraining the gradient to lie in the tangent space of the surface $P^{\mathrm{H}}P = I$ and projecting P back onto this surface at each iteration. As shown in [15], the projection is accomplished by replacing P with its "Q-Factor" $\mathrm{qf}\{P\}$ defined as follows. If P has p columns and its QR decomposition [6] is $P = QR$ then $\mathrm{qf}\{P\}$ is the matrix formed from the first p columns of the unitary matrix Q. (This is actually a projection onto a Grassmann manifold but since $f(P)$ satisfies $f(PU) = f(P)$ for any unitary matrix U, it is the appropriate projection to use [15].)

Algorithm 5 (MMSE linear precoder). The following algorithm produces a sequence of isometric precoders $P^{(k)}$ converging to a local minimum of $\boldsymbol{E}[\mathrm{tr}\{(P^{\mathrm{H}}\Lambda P\}]$.

1. Initialise $P^{(1)}$ to a randomly chosen isometric matrix and set $k := 1$.

2. Generate a channel Λ at random. (See Section 2 for the definition of Λ.)
3. Calculate the stochastic steepest descent direction $G := \Lambda P^{(k)}((P^{(k)})^{\mathrm{H}}\Lambda P^{(k)})^{-2}$.
4. Perform the projection $G := G - P^{(k)}(P^{(k)})^{\mathrm{H}}G$.
5. Take a stochastic descent step $P^{(k+1)} := P^{(k)} + (1/k)G$.
6. Renormalise the columns of $P^{(k+1)}$ by setting $P^{(k+1)} := \mathrm{qf}\{P^{(k+1)}\}$ (see above paragraph).
7. Set $k := k + 1$ and go to step 2.

Explanatory notes: The matrix G in Step 3 is the gradient of $-\mathrm{tr}\{(P^{\mathrm{H}}\Lambda P)^{-1}\}$. The projection in Step 4 ensures the gradient lies in the tangent plane of the Grassmann manifold. Step 6 performs a numerically stable Gram–Schmidt orthogonalisation on the columns of P and is equivalent to projecting P onto the Grassmann manifold. See [15] for details.

Remark. The convergence rate of stochastic gradient algorithms can be speeded up by stochastic averaging techniques [7, Section 11]. This is not considered here because MMSE precoders are usually constructed offline and a fast convergence rate is not essential.

In practice, it is suggested Algorithm 5 is first run for ten thousand or so iterations with $P^{(1)}$ equal to the pseudo-identity matrix and the MSE $\boldsymbol{E}[\mathrm{tr}\{P^{\mathrm{H}}\Lambda P\}]$ of the resulting precoder evaluated by Monte Carlo simulation. Then Algorithm 5 can be run several more times, each time with $P^{(1)}$ randomly chosen. The precoder having the smallest MSE is the one ultimately chosen. Choosing a channel distribution satisfying (A1) improves convergence.

Empirical evidence suggests the problem of local minima is not severe. Choosing different initial values for $P^{(1)}$ usually leads to precoders having similar MSEs. Further evidence is given in Section 5.1. The only explanation offered for the good performance of Algorithm 5 is that by generating a different random channel at each iteration the precoder is moulded to perform well over a wide range of channels. *In fact, Algorithm 5 can be interpreted as a learning algorithm whereby the precoder learns the characteristics of the random channel.*

Extensions: The cost functions in [9,19] when the channel is known extend to cost functions when the channel is random by taking their expected values. Algorithm 5 is readily modified to compute optimal precoders with respect to these cost functions. Moreover, Algorithm 5 can be used as an adaptive algorithm for updating the precoder when more information about the channel (obtained via a reverse link) becomes available to the transmitter.

4. MMSE linear precoders for known channels

Theorem 6 obtains an expression for MMSE linear precoders operating over known channels. The MMSE linear precoder is compared with other optimal designs proposed in [9,19]. (Further comparisons are made in Section 5.3.)

The constrained minimum of $\mathrm{tr}\{(P^{\mathrm{H}}\Lambda P)^{-1}\}$ has an intuitively pleasing solution; Theorem 6 proves the columns of P must span the principal subspace of Λ.

Theorem 6. *Let P be a matrix with p columns and at least p rows. Let Λ be a Hermitian matrix with eigenvalue decomposition $\Lambda = V^{\mathrm{H}}\Sigma V$ where Σ is a diagonal matrix with elements arranged in descending order. Then $\mathrm{tr}\{P^{\mathrm{H}}\Lambda P)^{-1}\}$ achieves its minimum subject to $\lambda_{\max}\{P\} \leqslant 1$ when P is the sub-matrix formed by the first p columns of V.*

Proof. Theorem 1 in Part I proves it suffices to consider isometric precoders. The Lagrange multiplier technique shows a necessary condition for P to minimise $\mathrm{tr}\{(P^{\mathrm{H}}\Lambda P)^{-1}\}$ subject to $P^{\mathrm{H}}P = I$ is for the columns of P to span an invariant subspace of Λ. It is then straightforward to show the minimum occurs when the columns of P span the principal subspace of Λ. □

In [9] the optimisation problem differs from the one here in two ways; the MSE of the precoder is not measured by $\mathrm{tr}\{(P^{\mathrm{H}}\Lambda P)^{-1}\}$ but by the MSE at the output of a biased MMSE equaliser, and furthermore, the average power constraint $\mathrm{tr}\{PP^{\mathrm{H}}\} \leqslant p$ rather than the peak power constraint $\lambda_{\max}\{P\} \leqslant 1$ is used. The solution though also uses the principle subspace but the less restrictive power constraint provides scope for additional power allocation strategies within the principal subspace. Note that using the MMSE precoder in Theorem 6 does not preclude such power allocation strategies because the optimal precoder in [9] factors as PA where P is a MMSE precoder and A is a square matrix performing the power allocation. (This is consistent with the formulation in Part I that the ECC is designed after the precoder is designed.) Therefore, despite the different criteria used, it can be said MMSE precoders are equivalent to the precoders in [9] if the channel is known.

In [19] the precoder is designed to optimise the mutual information between the source symbols and the channel output. Section 5.3 explains why this criterion is not appropriate when coding over multiple blocks is prohibited. Interestingly though, the optimal precoders in [19] also use the principal subspace of the channel and are thus equivalent to MMSE precoders if the channel is known.

5. Discussion

Designing a good codeword constellation for a fixed [1] but unknown channel with memory is the underlying problem in packet networks and time-critical block transmissions described in Section 2.1 of Part I yet is an open problem. It is shown in [22] that using a linear precoder to shape the constellation yields better performance than using a simple ECC. However, the asymptotic analysis in [4] suggests there is no need for a linear precoder; standard ECCs are adequate. This section offers a simple explanation for these observations in addition to an interpretation of the theoretical results in this paper.

5.1. Numerical examples

Part I argued the trace of the Cramer–Rao Bound matrix $\mathrm{tr}\{R(\boldsymbol{h})\}$ measures the intrinsic performance of a linear precoder. Figs. 2 and 3 corroborate this assertion because they both show the BER decreases as $\mathrm{tr}\{R(\boldsymbol{h})\}$ decreases. (That $\mathrm{tr}\{R(\boldsymbol{h})\}$ decreases as n increases is seen from Fig. 5 and is discussed shortly.) Moreover, as anticipated in Part I, the performance of a linear precoder used in conjunction with a good ECC (Fig. 3) is significantly better than when no ECC is used (Fig. 2).

[1] Although the channel is quasi-stationary and changes from block to block, because coding across multiple blocks is prohibited in the problem formulation, the channel is effectively fixed but unknown for the duration of a single block.

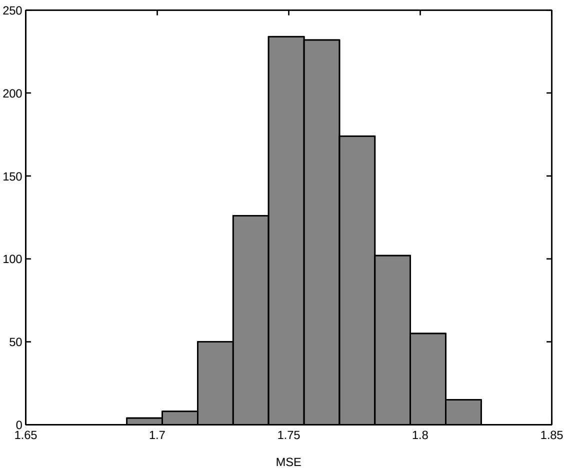

Fig. 2. Graphs of BER versus E_b/N_0 (or SNR) for BPSK transmissions over a normalised Rayleigh fading channel of length 4 using MMSE linear precoders of various sizes at the transmitter and a minimum variance unbiased equaliser at the receiver followed by quantisation. The size of each precoder is n by 15, where n ranges from 15 (top line) to 57 (bottom line).

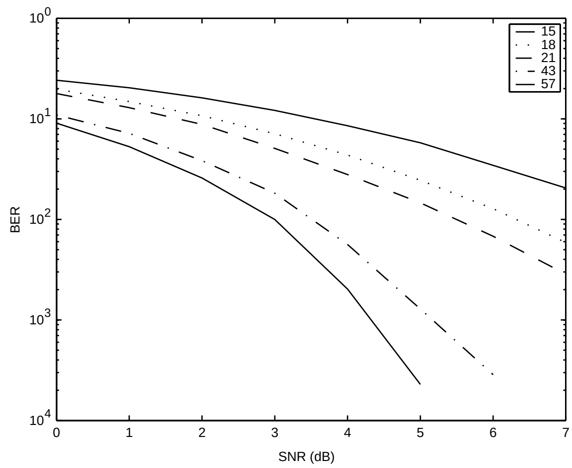

Fig. 3. Graphs of BER versus E_b/N_0 (or SNR) for BPSK transmissions over a normalised Rayleigh fading channel of length 4 using a BCH (15,7) code followed by a MMSE linear precoder at the transmitter and a minimum variance unbiased equaliser followed by a nearest neighbour decoder at the receiver. The size of each linear precoder is n by 15, where n ranges from 15 (top line) to 57 (bottom line).

Section 1 referred to the histogram in Fig. 4 to illustrate Algorithm 5 finds precoders with a significantly smaller MSE than those generated at random; this not

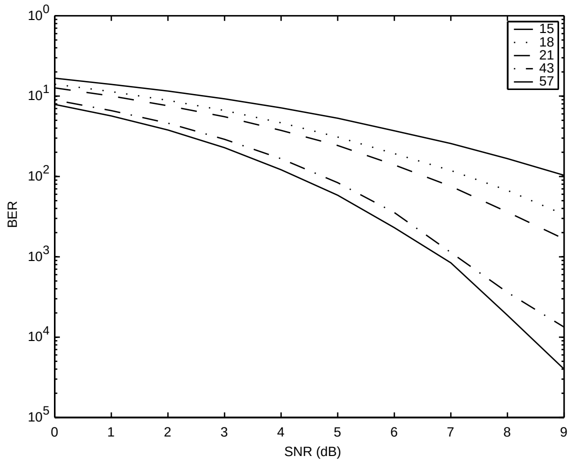

Fig. 4. Histogram showing the distribution of the MSEs of 1000 randomly generated isometric precoders of size 21 × 15 as measured over a normalised Rayleigh fading channel.

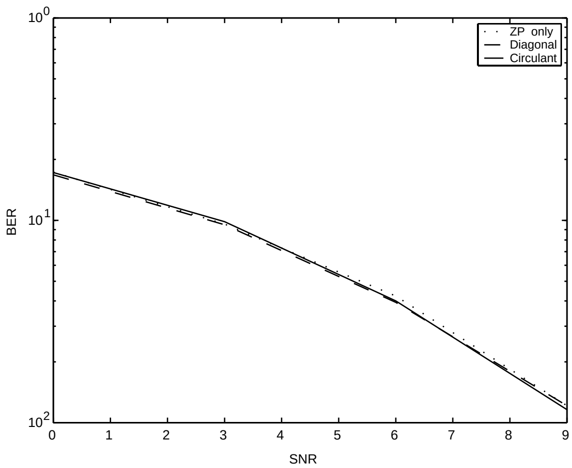

Fig. 5. Graph of MSE versus redundancy of MMSE linear precoders operating over a normalised Rayleigh fading channel of length 4. Here, the redundancy is defined as $n/15$ where each linear precoder is of size n by 15.

only demonstrates the correct functioning of Algorithm 5 but raises the question of whether or not results based on randomly generated precoders [4] hold for "optimal" precoders.

Fig. 5 also demonstrates the correct functioning of Algorithm 5 by showing the MSE of precoders found with the algorithm decreases as the redundancy increases, as to be expected. Furthermore, the convergence to the lower bound of unity is in perfect agreement with the theory in Section 3.1.

5.2. Interpretation of results

From both a robust and a MMSE perspective, this paper proved the best a linear precoder can do is convert the multipath channel into an AWGN channel. The performance cannot be reduced beyond this point. That is to say, *a linear precoder cannot offer any protection against additive white Gaussian noise*. This observation is consistent with, but does not follow from, the readily verifiable fact that a redundant linear precoder reduces the capacity of an AWGN channel; since E_b/N_0 is constant, the capacity per symbol of the linearly precoded channel is the same but the time it takes to transmit each symbol is longer [3]. (See Section 5.3 for a discussion on the relevance of channel capacity though.) Therefore, *a linear precoder should be used in conjunction with an ECC capable of combatting AWGN*. The substantial performance gains in Fig. 3 over Fig. 2 corroborate this recommendation.

The following intuitive explanation is offered. In the frequency domain, linear precoders spread the spectrum in an orderly way [16]. This spreading reduces the effects of frequency distortion caused by the memory of the channel. On the other hand, Shannon has shown ECCs designed for AWGN channels must spread the symbols in the time domain. Therefore, the best performance results when both types of coding are present. (The best is if a codeword constellation having both properties can be constructed; the choice of a standard ECC followed by a linear precoder should be viewed as either an interim or a low complexity solution to this coding problem.)

The results also suggest *the linear precoder should introduce only a small amount of redundancy*, that is, $(n-p)/L$ should be close to zero where $n \times p$ is the size of the precoder matrix and the normalising constant L is the length of the channel. (Since a length L channel has at most $L-1$ spectral nulls, L gives an indication of the amount of frequency distortion the linear precoder must overcome.) This is deduced from Fig. 5 which shows the MSE initially decreasing sharply as n increases but then tapering off. Specifically, because redundancy must be shared between the linear precoder and the ECC, the best is to allocate just enough redundancy to the linear precoder to let it combat the frequency distortion and to allocate the remaining redundancy to the ECC for combatting AWGN. The initial steep slope of Fig. 5 suggests precoders with small values of $(n-p)/L$ are a good compromise between combatting frequency distortion and leaving redundancy available for the ECC.

5.3. Comparison with other approaches

5.3.1. Comparison of design criteria

In [4,9], the performance of a precoder is measured by the MSE at the output of a biased MMSE equaliser rather than at the output of the minimum variance unbiased equaliser (3). This is because MMSE equalisers generally have a better BER when followed by per-symbol quantisation. However, this paper assumes an optimal detector is used, in which case trading bias for MSE reduction cannot improve the BER. Indeed, Eq. (5) in Part I shows the covariance matrix $R(\boldsymbol{h})$ of (3), and not the covariance matrix of the MMSE equaliser, directly influences the output of the optimal detector.

Another performance measure [19] is the mutual information between a block of source symbols and the corresponding channel output. However, mutual information is only meaningful if the ECC is permitted to code over multiple blocks, thereby exploiting time diversity [3]. Since coding over multiple blocks is forbidden in the problem formulation here, mutual information is not an appropriate performance measure.

5.3.2. Comparison of conclusions

The results in this paper do not conflict with those in [4,22]. The observation in [22] that linear precoders offer advantages over simple ECCs can be explained by postulating it is more important to combat frequency distortion using a linear precoder than it is to combat AWGN with a simple ECC under the conditions in [22]. The asymptotic analysis in [4] suggests linear precoders are not needed. However, the analysis assumes both n and p go to infinity with n/p fixed, where $n \times p$ is the precoder's size. This means $n-p$ goes to infinity too, and in particular, the argument in Section 5.2 that $n-p$ should be small if the precoder is to have any benefit is not contradicted. It is also mentioned that [4] considers random precoders whereas Section 1 gives a numerical example showing MMSE precoders differ substantially from randomly generated precoders.

6. Conclusion

This two-part paper studied the linear precoder design problem from a novel perspective. An intrinsic performance measure was introduced and used to formulate the precoder design problem as a constrained optimisation. Optimal solutions were found and used to gain a greater understanding of the strengths and limitations of linear precoders. A stochastic optimisation algorithm for constructing optimal linear precoders operating over random channels was also presented. The main finding is linear precoders should introduce a small amount of redundancy and be used in conjunction with ECC capable of correcting errors caused by additive white Gaussian noise.

Acknowledgements

The author gratefully acknowledges useful discussions with Professors Zhi Ding, Kawas Kaleh, Phillipe Loubaton and Lang White on the precoder design problem.

References

[1] A. Benveniste, M. Métivier, P. Priouret, Adaptive Algorithms and Stochastic Approximations, Springer, Berlin, 1990.

[2] J. Boutros, E. Viterbo, C. Rastello, J.-C. Belfiore, Good lattice constellations for both Rayleigh fading and Gaussian channels, IEEE Trans. Inform. Theory 42 (2) (1996) 502–518.

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

[4] 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

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

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

[7] H.J. Kushner, G.G. Yin, Stochastic Approximation Algorithms and Applications, Springer, Berlin, 1997.

[8] C. Lamy, J. Boutros, On random rotations diversity and minimum MSE decoding of lattices, IEEE Trans. Inform. Theory 46 (4) (2000) 1584–1589.
[9] 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.
[10] J.H. Manton, Dissecting OFDM: the independent roles of the cyclic prefix and the IDFT operation, IEEE Comm. Lett. 5 (12) (2001) 474–476.
[11] 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.
[12] J.H. Manton, Design and analysis of linear precoders under a mean square error criterion, Part I: foundations and worst case designs, Systems Control Lett. XRef. in this issue.
[13] J.H. Manton, An OFDM interpretation of zero padded block transmissions, Systems Control Lett. 47 (5) (2002) 393–399.
[14] J.H. Manton, On the importance of using a cyclic prefix or null guard in block transmissions over FIR channels, in: Information, Decision and Control, Adelaide, Australia, February, 2002.
[15] J.H. Manton, Optimisation algorithms exploiting unitary constraints, IEEE Trans. Signal Process. 50 (3) (2002) 635–650.
[16] J.H. Manton, Y. Hua, A frequency domain deterministic approach to channel identification, IEEE Signal Process. Lett. 6 (12) (1999) 323–326.
[17] J.H. Manton, Y. Hua, Robust frequency domain precoders, IEEE Comm. Lett. 5 (2) (2001) 40–42.
[18] S. Ohno, G.B. Giannakis, Universal redundant precoders for guaranteed recovery of block transmission through unknown FIR channels, in: IEEE Workshop on Signal Processing Advances in Wireless Communications, Taiwan, March, 2001.
[19] 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.
[20] 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.
[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] G.W. Wornell, Spread-response precoding for communication over fading channels, IEEE Trans. Inform. Theory 42 (2) (1996) 488–501.
[24] 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.