Optimal precoder for block transmission over frequency-selective fading channels

J.-K. Zhang, T.N. Davidson and K.M. Wong

Abstract: The authors consider the design of a precoder for block transmission over a frequency-selective fading channel that minimises the worst-case averaged pairwise error probability (PEP) of the maximum likelihood detector. In applications in which the transmitter does not know the channel, the scaled identity matrix is shown to be an optimal precoder for the general uncorrelated frequency-selective Rayleigh fading channel. Such precoded communication systems automatically guarantee that the maximum likelihood detector extracts full diversity and that the optimal coding gain is achieved. A comparison of the error performance of the optimal precoded system with that of other systems with unitary precoders shows that the optimal system obtains a significant SNR gain (2–4 dB).

1 Introduction

We consider wireless communication systems with a single transmitting antenna and a single receiving antenna which transmit data over a frequency-selective fading channel. The systems which we consider mitigate the intersymbol interference generated by the channel by transmitting the data stream in consecutive equal-size blocks, which are subsequently processed at the receiver on a block-by-block basis, see, e.g. [1–6]. In order to remove inter-block interference, some redundancy is added to each block before transmission. There are several ways to add redundancy (e.g. [1, 6]), but in this paper we will focus on linearly precoded block-by-block communication systems with zero-padding redundancy; e.g. [4–6]. To describe our systems of interest in more detail, we assume that the channel is of length at most L (i.e. that L is an upper bound on the delay spread). The systems operate as follows: First, a length K data symbol vector $\boldsymbol{s}$ is linearly precoded by a $K\times K$ matrix $\boldsymbol{F}$ to form the vector $\boldsymbol{x}=\boldsymbol{F}\boldsymbol{s}$. Then, $L-1$ zeros are append to $\boldsymbol{x}$ to form $\boldsymbol{x}'$ which is of length $P=K+L-1$. The elements of $\boldsymbol{x}'$ are then serially transmitted through the channel. The impulse response of the channel is denoted by $\boldsymbol{h}=[h_0,h_1,\ldots,h_{L-1}]^T$ and is assumed to be constant over the transmission of a block. The length P received signal vector $\boldsymbol{r}$ can be written as

$$\boldsymbol{r}=\boldsymbol{H}\boldsymbol{F}\boldsymbol{s}+\boldsymbol{\xi} \tag{1}$$

where $\boldsymbol{\xi}$ denotes the vector of noise samples at the receiver and $\boldsymbol{H}$ denotes the $P\times K$ Toeplitz matrix [4–7]

$$\boldsymbol{H}=\sum_{k=0}^{L-1} h_k\boldsymbol{T}_k \tag{2}$$

where, for $0\le k\le L-1$, $1\le i\le P$ and $1\le j\le K$, the (i, j)th element of the matrix $\boldsymbol{T}_k$ is

$$[\boldsymbol{T}_k]_{i,j}=\begin{cases}1, & \text{if } i=j+k\\ 0, & \text{otherwise}\end{cases} \tag{3}$$

For applications in which the transmitter knows the channel impulse response, there exist solutions [3] to a large number of precoder design problems for systems of the form in (1), including maximisation of information rate [4], maximisation of a measure of the signal-to-noise ratio (SNR) [5], minimisation of the mean squared error [5] and minimisation of the bit error probability for zero-forcing equalisation [7]. However, in wireless communication systems, it is often difficult to provide sufficiently timely and accurate feedback to the transmitter for such designs to be practically viable. Many proposed systems for such scenarios consist of a transmitter designed without knowledge of the channel, and a receiver which possesses perfect knowledge of the channel and employs maximum likelihood (ML) detection. As the pairwise error probability (PEP), see e.g. [8], is a convenient measure of the performance of the ML detector at high SNRs, a natural precoder design question is:

Question 1: If the transmitter does not have knowledge of the channel and the receiver employs maximum likelihood detection (with precise channel knowledge), which precoder will minimise the pairwise error probability (PEP) of the system?

In the following Sections we will show that under a common channel model, the (scaled) identity precoder $\boldsymbol{F}\propto\boldsymbol{I}_K$ is an optimal precoder. Such systems correspond to simple serial transmission of the block of data with guard times between the blocks. More specifically, we will show that for an independent (but not necessarily identically distributed) frequency-selective Rayleigh fading channel model, the identity precoder achieves the minimum worst-case average pairwise error probability. This result complements an independent result [9] which is weaker, but applies to a broader class of channels. That result states that the identity precoder achieves full diversity and maximum coding gain, and hence that for a general correlated Rayleigh fading channel model, the averaged Chernoff bound on the error probability is minimised.

IEE Proceedings online no. 20059012
doi:10.1049/ip-com:20059012
Paper first received 29th September 2003 and in revised form 13th October 2004

The authors are with the Department of Electrical and Computer Engineering, McMaster University, 1280 Main Street West, Hamilton, Ontario, Canada, L8S 4K1
E-mail: davidson@mcmaster.ca

2 Precoder design

Throughout this paper, we adopt the following assumptions:

(i) Perfect channel estimates are available at the receiver to allow coherent detection.

(ii) The channel impulse response vector $\boldsymbol{h}$ is a sample of zero-mean circularly-symmetric complex Gaussian random vector with covariance matrix $\boldsymbol{\Lambda} = \mathrm{diag}\{\lambda_1, \lambda_2, \ldots, \lambda_L\}$. The matrix $\boldsymbol{\Lambda}$ is of rank $N \le L$; i.e. $L-N$ diagonal elements are zeros. For notational convenience we define $\boldsymbol{Q}$ to be the $L \times N$ tall matrix containing the N columns of $\boldsymbol{I}_L$ corresponding to the nonzero values of λ_k. We also define the $N \times N$ full rank matrix $\tilde{\boldsymbol{\Lambda}} = \boldsymbol{Q}^T \boldsymbol{\Lambda} Q$, which is the covariance matrix of $\tilde{\boldsymbol{h}} = \boldsymbol{Q}^T \boldsymbol{h}$.

(iii) The elements of $\boldsymbol{s}$ are uncoded independent identically distributed (i.i.d) equally likely signaling points from the same constellation $\mathcal{S}$, normalised so that $\mathbf{E}\{\mathbf{s}\ \mathbf{s}^H\} = \boldsymbol{I}_K$.

(iv) The noise vector $\boldsymbol{\xi}$ is zero-mean circularly-symmetric complex white Gaussian noise with covariance $N_0 \boldsymbol{I}_p$.

Given a channel realisation $\boldsymbol{h}$, the maximum likelihood (ML) detector and two vectors $\boldsymbol{s}$ and $\boldsymbol{s}'$, the pairwise error probability (PEP), $P(\boldsymbol{s} \to \boldsymbol{s}' | \boldsymbol{h})$, is the probability of transmitting $\boldsymbol{s}$ and deciding in favor of $\boldsymbol{s}' \neq \boldsymbol{s}$ at the decoder. Under the above assumptions, the PEP can be written as [8]

$$P(\boldsymbol{s} \to \boldsymbol{s}' | \boldsymbol{h}) = Q\left(\frac{d_r(\boldsymbol{s}, \boldsymbol{s}')}{\sqrt{2N_0}}\right) \tag{4}$$

where $d_r(\boldsymbol{s}, \boldsymbol{s}')$ is the Euclidean distance between the received code words $\boldsymbol{HFs}$ and $\boldsymbol{HFs}'$,

$$d_r^2(\boldsymbol{s}, \boldsymbol{s}') = (\boldsymbol{s} - \boldsymbol{s}')^H \boldsymbol{F}^H \boldsymbol{H}^H \boldsymbol{H} \boldsymbol{F} (\boldsymbol{s} - \boldsymbol{s}') \tag{5}$$

and

$$Q(t) = \left(1/\sqrt{2\pi}\right) \int_t^{\infty} e^{-t^2/2} dt$$

We will find it convenient to use the following alternative expression for the Q function [10]:

$$Q(t) = \frac{1}{\pi} \int_0^{\pi/2} \exp\left(-\frac{t^2}{2\sin^2\theta}\right) d\theta \tag{6}$$

At high SNRs, the union bound (e.g. [8]) can be used to bound the block error probability P_{ble} in terms of the pairwise error probability

$$P_{\mathrm{ble}} \le \sum_{\boldsymbol{s} \neq \boldsymbol{s}'} P(\boldsymbol{s}) P(\boldsymbol{s} \to \boldsymbol{s}' | \boldsymbol{h}) = \sum_{\boldsymbol{s} \neq \boldsymbol{s}'} P(\boldsymbol{s}) Q\left(\frac{d_r(\boldsymbol{s}, \boldsymbol{s}')}{\sqrt{2N_0}}\right) \tag{7}$$

An observation which assists our analysis is that by using the Toeplitz structure of $\boldsymbol{H}$, the received signal vector (1) can be rewritten as

$$\boldsymbol{r} = \boldsymbol{X_F}(\boldsymbol{s}) \boldsymbol{h} + \xi \tag{8}$$

where

$$\boldsymbol{X_F}(\boldsymbol{s}) = [\boldsymbol{T}_0 \boldsymbol{Fs}, \boldsymbol{T}_1 \boldsymbol{Fs}, \ldots, \boldsymbol{T}_{L-1} \boldsymbol{Fs}] \tag{9}$$

Using (9) we have that

$$d_r^2(\boldsymbol{s}, \boldsymbol{s}') = \boldsymbol{h}^H \boldsymbol{X}_{\boldsymbol{F}}^H(\boldsymbol{e}) \boldsymbol{X_F}(\boldsymbol{e}) \boldsymbol{h} \tag{10}$$

where $\boldsymbol{e} = \boldsymbol{s} - \boldsymbol{s}'$. By taking the average of (4) over the random vector $\boldsymbol{h}$, whose statistics are given in assumption (ii), the average pairwise error probability can be written as

$$P_{\boldsymbol{F}}(\boldsymbol{s} \to \boldsymbol{s}') = \frac{1}{\pi} \int_0^{\pi/2} \frac{d\theta}{\det\left(\boldsymbol{I}_N + \left(4N_0 \sin^2\theta\right)^{-1} \tilde{\boldsymbol{X}}_{\boldsymbol{F}}^H(\boldsymbol{e}) \tilde{\boldsymbol{X}}_{\boldsymbol{F}}(\boldsymbol{e}) \tilde{\boldsymbol{\Lambda}}\right)} \tag{11}$$

where $\tilde{\boldsymbol{X}}_{\boldsymbol{F}}(\boldsymbol{e})$ is defined by

$$\tilde{\boldsymbol{X}}_{\boldsymbol{F}}(\boldsymbol{e}) = \boldsymbol{X_F}(\boldsymbol{e}) \boldsymbol{Q} \tag{12}$$

and we have made the dependence of the PEP on $\boldsymbol{F}$ explicit. From the union bound (7) we see (e.g. [8]) that when the SNR is high (and the symbols are equally likely), the average performance of the ML detector is dominated by the worst-case averaged PEP. Therefore, our design problem in question 1 can now be formally stated as:

Problem 1: Let $p > 0$ be fixed. Find a matrix $\boldsymbol{F}$ that minimises the worst-case average pairwise error probability $P_{\boldsymbol{F}}(\boldsymbol{s} \to \boldsymbol{s}')$, subject to the power constraint $\mathrm{tr}\,(\boldsymbol{F}^H \boldsymbol{F}) \le p$. That is, find

$$\boldsymbol{F}^* = \arg \min_{\mathrm{tr}\left(\boldsymbol{F}^H \boldsymbol{F}\right) \le p} \max_{\substack{\boldsymbol{s}, \boldsymbol{s}' \in \mathcal{S}^K \\ \boldsymbol{s} \neq \boldsymbol{s}'}} P_{\boldsymbol{F}}(\boldsymbol{s} \to \boldsymbol{s}')$$

where $\mathcal{S}^K = \mathcal{S} \times \mathcal{S} \times \cdots \times \mathcal{S}$.

To assist with our derivation of a solution to problem 1, we let s be an element of the vector $\boldsymbol{s}$ and define the minimum distance of the constellation $\mathcal{S}$ as

$$d_{\min} = \min_{s, s' \in \mathcal{S}, s \neq s'} |s - s'| \tag{13}$$

Note that since $\boldsymbol{s} \in \mathcal{S}^K$, the minimum distance between the vectors $\boldsymbol{s}$ and $\boldsymbol{s}'$ is simply $d_{\min}$: i.e. $\min_{\boldsymbol{s}, \boldsymbol{s}' \in \mathcal{S}^K, \boldsymbol{s} \neq \boldsymbol{s}'} \| \boldsymbol{s} - \boldsymbol{s}' \| = d_{\min}$. The proof of our main result (theorem 1, below) will exploit the following two lemmas, which are proved in the Appendix (Sections 6.1 and 6.2, respectively). These lemmas generate lower and upper bounds, respectively, on the worst-case average PEP.

Lemma 1: Let $J_N(a)$ denote the integral,

$$J_N(a) = \frac{1}{\pi} \int_0^{\pi/2} \prod_{k=1}^{N} \left(1 + \frac{a\lambda_k}{\sin^2\theta}\right)^{-1} d\theta \quad \text{for } a > 0 \tag{14}$$

Then we have that

$$\max_{\boldsymbol{s}, \boldsymbol{s}' \in \mathcal{S}^K, \boldsymbol{s} \neq \boldsymbol{s}'} P_{\boldsymbol{F}}(\boldsymbol{s} \to \boldsymbol{s}') \ge J_N\left(\frac{d_{\min}^2 p}{4N_0 K}\right)$$

Lemma 2: Let $\boldsymbol{G} = \mathrm{diag}\{g_1, g_2, \ldots, g_N\}$ with $g_n \ge 0$. Then, for any nonzero vector $\boldsymbol{e}$, the following inequality holds:

$$\det\left(\boldsymbol{G} + \tilde{\boldsymbol{X}}_{\boldsymbol{I}_K}^H(\boldsymbol{e}) \tilde{\boldsymbol{X}}_{\boldsymbol{I}_K}(\boldsymbol{e})\right) \ge \det\left(\boldsymbol{G} + d_{\min}^2 \boldsymbol{I}_N\right) = \prod_{k=1}^{N} \left(g_k + d_{\min}^2\right) \tag{15}$$

where equality in (15) holds if and only if $\| \boldsymbol{e} \| = d_{\min}$.

We now formally state our main result.

Theorem 1: We have the following three statements:

(i) The precoder $\boldsymbol{F}^* = \sqrt{p/K}\, \boldsymbol{I}_K$ is an optimal solution for problem 1.

(ii) The minimal value of the objective of problem 1 is

$$\max_{\boldsymbol{s}, \boldsymbol{s}' \in \mathcal{S}^K, \boldsymbol{s} \neq \boldsymbol{s}'} P_{\boldsymbol{F}^*}(\boldsymbol{s} \to \boldsymbol{s}') = J_N\left(\frac{d_{\min}^2 p}{4N_0 K}\right) \tag{16}$$

(iii) In addition, $P_{\boldsymbol{F}^*}(s \to s') = J_N\left(d_{\min}^2 p / (4N_0 K)\right)$ if and only if $\| \boldsymbol{s} - \boldsymbol{s}' \| = d_{\min}$.

Proof: First we notice that when $\boldsymbol{F} = \boldsymbol{F}^* = \sqrt{(p/K)}\boldsymbol{I}_K$, $\boldsymbol{X}_{\boldsymbol{F}}(\boldsymbol{e}) = \sqrt{(p/K)}\boldsymbol{X}_{\boldsymbol{I}_K}(\boldsymbol{e})$. In this case, we have

$$\det\left(\boldsymbol{I}_N + \frac{1}{4N_0 \sin^2\theta}\tilde{\boldsymbol{X}}_{\boldsymbol{F}^*}^H(\boldsymbol{e})\tilde{\boldsymbol{X}}_{\boldsymbol{F}^*}(\boldsymbol{e})\tilde{\boldsymbol{\Lambda}}\right) = \left(\frac{p}{4N_0 K \sin^2\theta}\right)^N \det(\tilde{\boldsymbol{\Lambda}}) \det\left(\boldsymbol{G} + \tilde{\boldsymbol{X}}_{\boldsymbol{I}_K}^{\boldsymbol{H}}(\boldsymbol{e})\tilde{\boldsymbol{X}}_{\boldsymbol{I}_K}(\boldsymbol{e})\right) \tag{17}$$

where

$$\boldsymbol{G} = \frac{4N_0 K \sin^2\theta}{p}\tilde{\boldsymbol{\Lambda}}^{-1}$$

Using lemma 2, we obtain that for any nonzero vector $\boldsymbol{e}$ and nonzero θ in the interval $[0, \pi/2]$

$$\det\left(\boldsymbol{G} + \tilde{\boldsymbol{X}}_{\boldsymbol{I}_K}^H(\boldsymbol{e})\tilde{\boldsymbol{X}}_{\boldsymbol{I}_K}(\boldsymbol{e})\right) \geq \prod_{k=1}^{N}\left(\frac{4N_0\boldsymbol{K}\sin^2\theta}{p\lambda_k} + d_{\min}^2\right) \tag{18}$$

Here, the inequality holds with equality if and only if $\boldsymbol{s}$ and $\boldsymbol{s}'$ are neighbour points, i.e. $\| \boldsymbol{s} - \boldsymbol{s}' \| = d_{\min}$. Therefore, combining (17) with (18) yields

$$\det\left(\boldsymbol{I}_N + \frac{1}{4N_0 \sin^2\theta}\tilde{\boldsymbol{X}}_{\boldsymbol{F}^*}^H(\boldsymbol{e})\tilde{\boldsymbol{X}}_{\boldsymbol{F}^*}(\boldsymbol{e})\tilde{\boldsymbol{\Lambda}}\right) \geq \prod_{k=1}^{N}\left(1 + \frac{d_{\min}^2 p\lambda_k}{4N_0 K \sin^2\theta}\right)$$

This results in

$$\max_{\boldsymbol{s},\boldsymbol{s}' \in \mathcal{S}^K, \boldsymbol{s} \neq \boldsymbol{s}'} P_{\boldsymbol{F}^*}(\boldsymbol{s} \to \boldsymbol{s}') \leq J_N\left(\frac{d_{\min}^2 p}{4N_0 K}\right) \tag{19}$$

where the inequality holds with equality if and only if $\| \boldsymbol{s} - \boldsymbol{s}' \| = d_{\min}$. Combining (19) with lemma 1 yields

$$\min_{\mathrm{tr}(\boldsymbol{F}^H\boldsymbol{F}) \leq p} \max_{\substack{\boldsymbol{s},\boldsymbol{s}' \in \mathcal{S}^K \\ \boldsymbol{s}' \neq \boldsymbol{s}}} P_{\boldsymbol{F}}(\boldsymbol{s} \to \boldsymbol{s}') = J_N\left(\frac{d_{\min}^2 p}{4N_0 K}\right)$$

and hence statements (i), (ii) and (iii) of the theorem. □

We now present a sequence of observations regarding theorem 1.

(i) Since $|\sin\theta| \leq 1$, we obtain from (11) that $P_{\boldsymbol{F}}(\boldsymbol{s} \to \boldsymbol{s}') \leq \frac{1}{2}\det(\boldsymbol{I}_N + (4N_0)^{-1}\tilde{\boldsymbol{X}}_{\boldsymbol{F}}^H(\boldsymbol{e})\tilde{\boldsymbol{X}}_{\boldsymbol{F}}(\boldsymbol{e})\tilde{\boldsymbol{\Lambda}})^{-1}$, which is the bound one would obtain by applying the Chernoff bound to (7) and then taking the average over $\boldsymbol{h}$. In some related work on multiple antenna transmission and reception over flat fading channels, certain 'rank' and 'determinant' criteria were derived [11, 12] in order to design 'space–time' codes which render the Chernoff bound 'small'. Applying these criteria to the precoder $\boldsymbol{F}$, we find that the choice $\boldsymbol{F}^*$, enables the ML detector to extract full diversity and provides the optimal coding gain. However, theorem 1 tells us that the identity precoder not only extracts full diversity and achieves the optimal coding gain, it actually minimises the worst-case average pairwise error probability (for the case of zero-padded block-by-block single antenna transmission and reception over an independent frequency-selective Rayleigh fading channel).

(ii) Theorem 1 also tells us that the optimal performance is obtained by simply serially transmitting the data symbols and then adding a 'guard time' by padding the appropriate number of zeros. There is an interesting coincidence that our optimal precoder for problem 1 is also an optimal precoder for cyclic-prefix-based block transmission schemes with linear zero-forcing or minimum mean squared error equalisation [13] or an 'iterated decision' detector [14]. That said, the diversity of our scheme is L whereas that of the cyclic-prefix-based scheme [13] is only one.

(iii) Since the channel coeffcients are modelled as being uncorrelated, one might suspect that any unitary precoder would provide equally good performance. This would indeed be true if the channel matrix $\boldsymbol{H}$ in (2) was derived from an uncorrelated frequency-flat Rayleigh fading multiple antenna channel [11, 12]. However, we consider the case of single antenna frequency-selective channels and the channel matrix has the Toeplitz structure illustrated in (2). Hence it cannot absorb the unitary precoder without changing the distribution of the channel coefficients. That said, the matrix $\boldsymbol{F} = \sqrt{(p/K)}\boldsymbol{P}\,\mathrm{diag}(e^{j\theta_1}, e^{j\theta_2}, \ldots, e^{j\theta_K})$, where $\boldsymbol{P}$ is a $K \times K$ permutation matrix, is also an optimal precoder in the sense of theorem 1. While the assumption that the elements of $\boldsymbol{h}$ are uncorrelated is an idealisation that facilitates some of the analysis herein, it is approximated in practice when the channel path gains are uncorrelated (as is often the case) and the spectral shaping is mild.

(iv) A desirable property of the optimally precoded system in theorem 1 is that it preserves the Toeplitz structure of the channel matrix, i.e. $\boldsymbol{HF}$ is also Toeplitz. This allows us to take advantage of the Viterbi algorithm [15] to efficiently implement the ML detector when the channel memory is short and the constellation size is not too large. In addition, this Toeplitz structure exposes the potential for blind equalisation techniques based on second-order statistics.

3 Simulations

To verify our analysis, we demonstrate the error performance using two simple examples. For simplicity, we assume that the channel vector $\boldsymbol{h}$ is a sample of zero-mean circularly-symmetric complex white Gaussian random vector with covariance matrix $\boldsymbol{\Lambda} = \boldsymbol{I}_L$ and that the elements of $\boldsymbol{s}$ are uncoded and independent identically-distributed equally-likely ± 1 s.

Example 1: In this example, we compare the error performance of optimally precoded channels with different memory. We consider systems for which the data symbol block size is $K = 8$. We consider scenarios in which the channel memory is $L = 2$, 3, 4 or 5. We define the SNR to be the ratio of the transmitted signal energy per symbol to the receiver noise variance per sample, i.e.

$$\frac{E\{(\boldsymbol{Fs})^H\boldsymbol{Fs}\}/K}{N_0} = \frac{p}{KN_0}$$

The average block error rates of our Monte Carlo simulations are shown by the dashed lines in Fig. 1. To check how close the dominant term in the union bound is to the simulated block error rates for the optimally precoded system, we also indicate the theoretic average block error probability of the ML detector with the solid line in Fig. 1. To be precise, the solid line is $\tau_K(\mathcal{S})J_L(d_{\min}^2 p/(4N_0K))$, where $\tau_K(\mathcal{S})$ denotes the 'kissing number', e.g. [8]. In our system $\tau_K(\mathcal{S}) = 8$. At high SNRs this is a good approximation of the true block error rate. The fact that our precoded scheme achieves the full diversity, L, provided by the channel is evident from the slopes of the curves in Fig. 1 at high SNR, which are proportional to $-L$.

Example 2: In this example, we compare our optimal precoder with the following precoders: Zero-padded transmission precoded by the normalised inverse discrete Fourier transform (IDFT) matrix, i.e. 'zero-padded OFDM' [9]; and zero-padded transmission precoded by the Hadamard matrix. For all three precoders, we employ

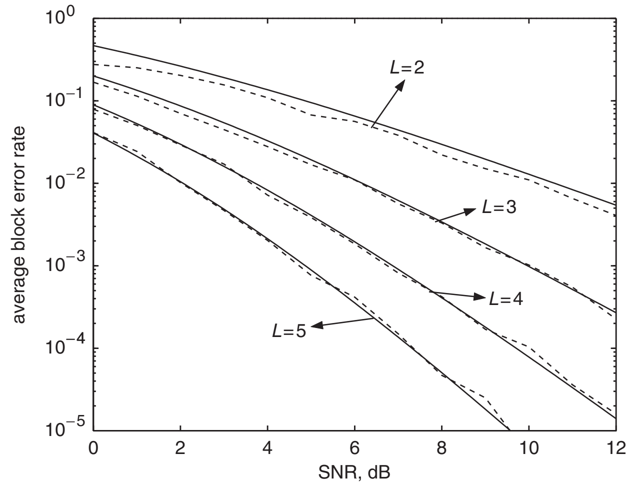

Fig. 1 *Simulated average block error rate (dashed) and the theoretic approximation to the average block error probability (solid) for optimally precoded schemes for frequency-selective fading channels of lengths $L = 2$, 3, 4 and 5*

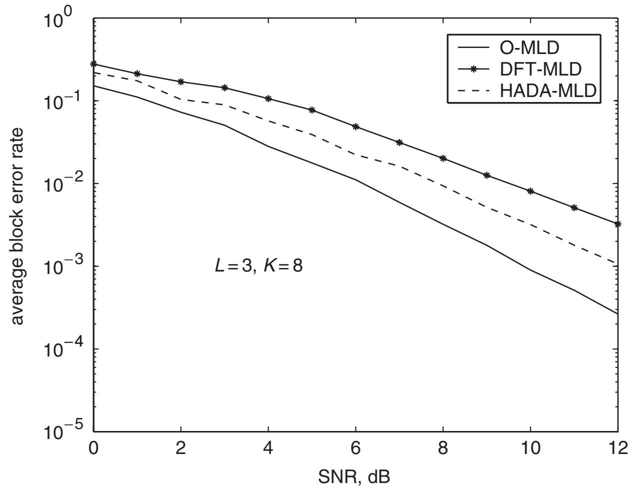

Fig. 2 *Block error rate performance comparison of the optimally precoded system (O-MLD) with systems which use the IDFT precoder (DFT-MLD) or Hadamard precoder (HADA-MLD) for the length $L = 3$ frequency-selective fading channel in example 1*

ML detection. The scenario corresponds to the case of the frequency-selective fading channel of length $L = 3$ from example 1. It is clear from Fig. 2 that the identity precoder can provide substantial SNR gain (up to 4 dB).

4 Conclusions

We have shown that a scaled identity matrix is an optimal zero-padded block precoder for the uncorrelated Gaussian frequency-selective fading channel. The precoder is optimal in the sense that it minimises the worst-case average pairwise error probability of the maximum likelihood detector in the scenario in which no channel state information is available at the transmitter (except an upper bound on the delay spread). Here, the average is taken over the distributions of the channel coefficients, which are assumed to be independent, but not necessarily identical, Gaussian distributions. The worst case is taken over the set of all pairs of data vectors. In addition to minimising the worst-case pairwise error probability, the identity precoder also guarantees that the maximum likelihood detector extracts full diversity from channel and that the optimal coding gain is achieved. Finally, our main result also identifies the vector symbol pairs which achieve the worst-case pairwise error probability. That information may be useful in the design of an outer code for our systems of interest. By choosing to minimise a measure of the pairwise error probability, we have implicitly focused on high SNR performance. However, our simulation results indicated that the scaled identity precoder continues to perform well at lower SNRs.

5 References

1 Al-Dhahir, N., and Diggavi, N.: 'Guard sequence optimization for block transmission over linear frequency-selective channels', *IEEE Trans. Commun.*, 2002, **50**, pp. 938–946

2 Muquet, B., Wang, Z., Giannakis, G.B., de Courville, M., and Duhamel, P.: 'Zero padding for wireless multicarrier transmissions', *IEEE Trans. Commun.*, 2002, **50**, pp. 2136–2148

3 Palomar, D.P., Cioffi, J.M., and Lagunas, M.A.: 'Joint Tx-Rx beamforming design for multicarrier MIMO channels: A unified framework for convex optimization', *IEEE Trans. Signal Process.*, 2003, **51**, pp. 2381–2401

4 Scaglione, A., Giannakis, G.B., and Barbarossa, S.: 'Filterbank transceivers optimizing information rate in block transmissions over dispersive channels', *IEEE Trans. Inf. Theory*, 1999, **45**, pp. 1019–1032

5 Scaglione, A., Giannakis, G.B., and Barbarossa, S.: 'Redundant filterbank precoders and equalizers Part I: Unification and optimal designs', *IEEE Trans. Signal Process.*, 1999, **47**, pp. 1988–2005

6 Wang, Z., and Giannakis, G.B.: 'Wireless multicarrier communications: Where Fourier meets Shannon', *IEEE Signal Process. Mag.*, 2000, **47**, pp. 29–48

7 Ding, Y.-W., Davidson, T.N., Luo, Z.-Q., and Wong, K.M.: 'Minimum BER block precoders for zero-forcing equalization', *IEEE Trans. Signal Process.*, 2003, **51**, pp. 2410–2423

8 Forney, G.D. Jr., and Ungerboeck, G.U.: 'Modulation and coding for linear Gaussian channel', *IEEE Trans. Inf. Theory*, 1998, **44**, pp. 2384–2415

9 Wang, Z., Ma, X., and Giannakis, G.B.: 'Optimality of single-carrier zero-padded block transmissions'. Proc. IEEE Wireless Commun. Networking Conf., Orlando, FL, USA, Mar. 2002, pp. 660–664

10 Simon, M.K., and Alouini, M.-S.: 'A unified approach to the performance analysis of digital communication over generalized fading channels', *Proc. IEEE.*, 1998, **86**, pp. 1860–1877

11 Guey, J.-C., Fitz, M.P., Bell, M.R., and Kuo, W.-Y.: 'Signal design for high data rate wireless communication systems over Rayleigh fading channels'. Proc. IEEE Vehicular Technology Conf., 1996, pp. 136–140

12 Tarokh, V., Jafarkhani, H., and Calderbank, A.R.: 'Space-time codes for high date rate wireless communication: performance criterion and code construction', *IEEE Trans. Inf. Theory*, 1998, **44**, pp. 744–765

13 Lin, Y.-P., and Phoong, S.-M.: 'BER optimized OFDM systems with channel independent precoders', *IEEE Trans. Signal Process.*, 2003, **51**, pp. 2369–2380

14 Chan, A.M., and Wornell, G.W.: 'Approaching the matched-filter bound using iterated-decision equalization with frequency-interleaved encoding'. Proc. Global Telecommun. Conf., Taipei, Taiwan, Nov. 2002, pp. 297–301

15 Forney, G.D. Jr.: 'Maximum-likelihood sequence estimation of digital sequences in the presence of intersymbol interference', *IEEE Trans. Inf. Theory*, 1972, **18**, pp. 363–378

16 Horn, R.A., and Johnson, C.R.: 'Matrix analysis' (Cambridge University Press, Cambridge, UK, 1985)

6 Appendix

6.1 Proof of lemma 1

From the definition (12) of the signal matrix $\tilde{\boldsymbol{X}}_{\boldsymbol{F}}(\boldsymbol{s})$, we observe that all diagonal entries of $\tilde{\boldsymbol{X}}_{\boldsymbol{F}}^H(\boldsymbol{e})\tilde{\boldsymbol{X}}_{\boldsymbol{F}}(\boldsymbol{e})$ are equal each other and that

$$\left[\tilde{\boldsymbol{X}}_{\boldsymbol{F}}^H(\boldsymbol{e})\tilde{\boldsymbol{X}}_{\boldsymbol{F}}(\boldsymbol{e})\right]_{k,k} = \boldsymbol{e}^H\boldsymbol{F}^H\boldsymbol{F}\boldsymbol{e}$$

for $k = 1, 2, \ldots, N$, where the notation $[\boldsymbol{A}]_{k,k}$ denotes the kth diagonal entry. Therefore,

$$\left[\boldsymbol{I}_N + \left(4N_0\sin^2\theta\right)^{-1}\tilde{\boldsymbol{X}}_{\boldsymbol{F}}^H(\boldsymbol{e})\tilde{\boldsymbol{X}}_{\boldsymbol{F}}(\boldsymbol{e})\tilde{\boldsymbol{\Lambda}}\right]_{k,k}$$
$$= 1 + \left(4N_0\sin^2\theta\right)^{-1}\lambda_k\boldsymbol{e}^H\boldsymbol{F}^H\boldsymbol{F}\boldsymbol{e}$$

for $k = 1, 2, \ldots, N$. Using Hadamard's inequality [16], we have

$$\det\left(\boldsymbol{I}_N + \frac{\tilde{\boldsymbol{X}}_{\boldsymbol{F}}^H(\boldsymbol{e})\tilde{\boldsymbol{X}}_{\boldsymbol{F}}(\boldsymbol{e})\tilde{\boldsymbol{\Lambda}}}{4N_0 \sin^2\theta}\right) = \det\left(\boldsymbol{I}_N + \frac{\tilde{\boldsymbol{\Lambda}}^{1/2}\tilde{\boldsymbol{X}}_{\boldsymbol{F}}^H(\boldsymbol{e})\tilde{\boldsymbol{X}}_{\boldsymbol{F}}(\boldsymbol{e})\tilde{\boldsymbol{\Lambda}}^{1/2}}{4N_0 \sin^2\theta}\right) \le \prod_{k=1}^{N}\left(1 + \frac{\lambda_k \boldsymbol{e}^H \boldsymbol{F}^H \boldsymbol{F}\boldsymbol{e}}{4N_0 \sin^2\theta}\right) \quad (20)$$

Let $m = \arg\min_{1\le k\le K}[\boldsymbol{F}^H\boldsymbol{F}]_{k,k}$. Setting $|e_m| = d_{\min}$ and $e_k = 0$, $k = 1, 2, \ldots, K$, $k \ne m$ in (20) yields

$$\det\left(\boldsymbol{I}_N + \frac{\tilde{\boldsymbol{X}}_{\boldsymbol{F}}^H(\boldsymbol{e})\tilde{\boldsymbol{X}}_{\boldsymbol{F}}(\boldsymbol{e})\tilde{\boldsymbol{\Lambda}}}{4N_0 \sin^2\theta}\right) \le \prod_{k=1}^{N}\left(1 + \frac{d_{\min}^2 \lambda_k [\boldsymbol{F}^H\boldsymbol{F}]_{m,m}}{4N_0 \sin^2\theta}\right) \quad (21)$$

However, we know that

$$[\boldsymbol{F}^H\boldsymbol{F}]_{m,m} \le \mathrm{tr}(\boldsymbol{F}^H\boldsymbol{F})/K \le p/K \quad (22)$$

Combining (22) with (21) leads to

$$\det\left(\boldsymbol{I}_N + \frac{\tilde{\boldsymbol{X}}_{\boldsymbol{F}}^H(\boldsymbol{e})\tilde{\boldsymbol{X}}_{\boldsymbol{F}}(\boldsymbol{e})\tilde{\boldsymbol{\Lambda}}}{4N_0 \sin^2\theta}\right) \le \prod_{k=1}^{N}\left(1 + \frac{d_{\min}^2 p\lambda_k}{4N_0 K \sin^2\theta}\right) \quad (23)$$

Therefore, for any nonzero vector $\boldsymbol{e}$

$$\frac{1}{\pi}\int_0^{\pi/2} \frac{d\theta}{\det\left(\boldsymbol{I}_N + (4N_0\sin^2\theta)^{-1}\tilde{\boldsymbol{X}}_{\boldsymbol{F}}^H(\boldsymbol{e})\tilde{\boldsymbol{X}}_{\boldsymbol{F}}(\boldsymbol{e})\tilde{\boldsymbol{\Lambda}}\right)} \ge J_N\left(\frac{d_{\min}^2 p}{4N_0 K}\right)$$

This completes the proof of lemma 1. □

6.2 Proof of lemma 2

To simplify the proof of this lemma, we first state and prove an auxiliary result which relates a measure of the distance between matrices $\tilde{\boldsymbol{X}}_{\boldsymbol{I}_K}(\boldsymbol{s})$ and $\tilde{\boldsymbol{X}}_{\boldsymbol{I}_K}(\boldsymbol{s}')$ to $d_{\min}$. The statement in the following lemma is more general than is needed to prove lemma 2, but it may be of independent interest. For notational convenience, let $\tilde{\boldsymbol{X}}_{\boldsymbol{I}_K}(\boldsymbol{e}, \{i_1, i_2, \ldots, i_n\})$ denote the matrix that remains after the columns of $\tilde{\boldsymbol{X}}_{\boldsymbol{I}_K}(\boldsymbol{e})$ indexed by $\{i_1, i_2, \ldots, i_n\}$ have been removed.

Lemma 3: Let $\boldsymbol{s}, \boldsymbol{s}' \in \mathcal{S}^K$ and $\boldsymbol{e} = \boldsymbol{s} - \boldsymbol{s}'$. Then, for any nonzero vector $\boldsymbol{e}$, we have

$$\det\left(\tilde{\boldsymbol{X}}_{\boldsymbol{I}_K}^H(\boldsymbol{e}, \{i_1, i_2, \ldots, i_n\})\tilde{\boldsymbol{X}}_{\boldsymbol{I}_K}(\boldsymbol{e}, \{i_1, i_2, \ldots, i_n\})\right) \ge d_{\min}^{2(N-n)} \quad \text{for } n = 0, 1, \ldots, N-1 \quad (24)$$

where the inequality holds with equality if and only if $\boldsymbol{s}$ and $\boldsymbol{s}'$ are neighbours, *i.e.* if and only if $\|\boldsymbol{e}\| = d_{\min}$.

Proof: Without the loss of generality, we can always assume that $\boldsymbol{e}_1 \ne 0$, where $\boldsymbol{e}_1$ is the first element of $\boldsymbol{e}$. Otherwise, we can permute the rows and columns of $\tilde{\boldsymbol{X}}_{\boldsymbol{I}_K}(\boldsymbol{e})$ such that the first entry is nonzero. (Recall that row and column permutations of a matrix $\boldsymbol{X}$ do not change the determinant of $\boldsymbol{X}^H\boldsymbol{X}$.) In this case, $\tilde{\boldsymbol{X}}_{\boldsymbol{I}_K}(\boldsymbol{e})$ can be written as

$$\tilde{\boldsymbol{X}}_{\boldsymbol{I}_K}(\boldsymbol{e}) = \begin{pmatrix} e_1 & 0 & \ldots & 0 \\ e_2 & e_1 & \ldots & 0 \\ \vdots & e_2 & \ddots & \vdots \\ e_N & \ddots & \ddots & e_1 \\ \vdots & e_N & \ddots & e_2 \\ e_K & \ddots & \ddots & \vdots \\ 0 & \ddots & \ddots & e_N \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \ddots & 0 & e_K \end{pmatrix}_{P\times N} \quad (25)$$

(Although (25) has been written for the case where $K > N$, the analysis is valid for any $K \ge 1$.) An important property of $\tilde{\boldsymbol{X}}_{\boldsymbol{I}_K}(\boldsymbol{e})$ in (25) is that the submatrix consisting of the first N rows of $\tilde{\boldsymbol{X}}_{\boldsymbol{I}_K}(\boldsymbol{e})$ is a lower triangular matrix with nonzero diagonal entries. Now, if we eliminate any n columns $\{i_1, i_2, \ldots, i_n\}$ from $\tilde{\boldsymbol{X}}_{\boldsymbol{I}_K}(\boldsymbol{e})$, we can always find matrices that permute the rows and columns of the remaining matrix $\tilde{\boldsymbol{X}}_{\boldsymbol{I}_K}(\boldsymbol{e}, \{i_1, i_2, \ldots, i_n\})$ so that it has a structure analogous to that in (25). That is, we can find a $P \times P$ permutation matrix $\boldsymbol{P}_1$ and an $(N-n)\times(N-n)$ permutation matrix $\boldsymbol{P}_2$ such that

$$\boldsymbol{P}_1\tilde{\boldsymbol{X}}_{\boldsymbol{I}_K}(\boldsymbol{e}, \{i_1, i_2, \ldots, i_n\})\boldsymbol{P}_2 = \begin{pmatrix} \boldsymbol{A} \\ \boldsymbol{B} \end{pmatrix}$$

where $\boldsymbol{A}$ is an $(N-n)\times(N-n)$ lower triangular matrix with nonzero diagonal entries and $\boldsymbol{B}$ denotes the remaining submatrix of $\boldsymbol{P}_1\tilde{\boldsymbol{X}}_{\boldsymbol{I}_K}(\boldsymbol{e}, \{i_1, i_2, \ldots, i_n\})\boldsymbol{P}_2$. Therefore, using a standard result on the determinant of the sum of a positive definite and a positive semidefinite matrix (e.g. [16, p.484]), we have

$$\det\left(\tilde{\boldsymbol{X}}_{\boldsymbol{I}_K}^H(\boldsymbol{e}, \{i_1, i_2, \ldots, i_n\})\tilde{\boldsymbol{X}}_{\boldsymbol{I}_K}(\boldsymbol{e}, \{i_1, i_2, \ldots, i_n\})\right) \ge \det(\boldsymbol{A}^H\boldsymbol{A}) + \det(\boldsymbol{B}^H\boldsymbol{B}) \ge d_{\min}^{2(N-n)} \quad (26)$$

The first inequality in (26) holds with equality if and only if $\boldsymbol{B}$ is the zero matrix, and the second inequality holds with equality if and only if $\det(\boldsymbol{B}^H\boldsymbol{B}) = 0$ and $\boldsymbol{A}^H\boldsymbol{A}$ is diagonal with diagonal elements equal to $d_{\min}^2$. Therefore, both inequalities in (26) hold with equality if and only if $\boldsymbol{s}$ and $\boldsymbol{s}'$ are neighbouring points. This completes the proof of lemma 3. □

We now proceed with the proof of lemma 2. For simplicity, we introduce the following notation:

$$\mathcal{D}(g_1, g_2, \ldots, g_N; \tilde{\boldsymbol{X}}_{\boldsymbol{I}_K}(\boldsymbol{e})) = \det(\boldsymbol{G} + \boldsymbol{R}) \quad (27)$$

where $\boldsymbol{R} = (r_{i,j})_{1\le i,j\le N} = \tilde{\boldsymbol{X}}_{\boldsymbol{I}_K}^H(\boldsymbol{e})\tilde{\boldsymbol{X}}_{\boldsymbol{I}_K}(\boldsymbol{e})$. We will prove lemma 2 via induction on N.

When $N = 1$, the matrices in (15) collapse to scalars and the inequality is obtained directly. However, to simplify the proof of the inductive step, we now explicitly consider the case in which $N = 2$. In that case

$$\boldsymbol{R} = \begin{bmatrix} r_{11} & r_{12} \\ r_{12}^* & r_{22} \end{bmatrix} \quad (28)$$

and hence

$$\mathcal{D}(g_1, g_2; \tilde{\boldsymbol{X}}_{\boldsymbol{I}_K}(\boldsymbol{e})) = \det(\boldsymbol{G} + \boldsymbol{R}) = \begin{vmatrix} g_1 + r_{11} & r_{12} \\ r_{12}^* & g_2 + r_{22} \end{vmatrix} \quad (29)$$

Expanding the determinant in (29) yields

$$\begin{aligned}
&\mathcal{D}\big(g_1, g_2; \tilde{\boldsymbol{X}}_{\boldsymbol{I}_K}(\boldsymbol{e})\big) \\
&= \begin{vmatrix} g_1 & r_{12} \\ 0 & g_2 + r_{22} \end{vmatrix} + \begin{vmatrix} r_{11} & 0 \\ r_{12}^* & g_2 \end{vmatrix} + \begin{vmatrix} r_{11} & r_{12} \\ r_{12}^* & r_{22} \end{vmatrix} \\
&= g_1(g_2 + r_{11}) + g_2 r_{22} + \det\Big(\tilde{\boldsymbol{X}}_{\boldsymbol{I}_K}^H(\boldsymbol{e})\tilde{\boldsymbol{X}}_{I_K}(\boldsymbol{e})\Big) \\
&= g_1\Big(g_2 + \tilde{\boldsymbol{X}}_{I_K}^H(\boldsymbol{e}, \{1\})\tilde{\boldsymbol{X}}_{I_K}(\boldsymbol{e}, \{1\})\Big) \\
&\quad + g_2 \tilde{\boldsymbol{X}}_{\boldsymbol{I}_K}^H(\boldsymbol{e}, \{2\})\tilde{\boldsymbol{X}}_{\boldsymbol{I}_K}(\boldsymbol{e}, \{2\}) + \det\Big(\tilde{\boldsymbol{X}}_{I_K}^H(\boldsymbol{e})\tilde{\boldsymbol{X}}_{I_K}(\boldsymbol{e})\Big) \\
&= g_1 \mathcal{D}\big(g_2; \tilde{\boldsymbol{X}}_{\boldsymbol{I}_K}(\boldsymbol{e}, \{1\})\big) + g_2 \mathcal{D}\big(0; \tilde{\boldsymbol{X}}_{\boldsymbol{I}_K}(\boldsymbol{e}, \{2\})\big) \\
&\quad + \mathcal{D}\big(0, 0; \tilde{\boldsymbol{X}}_{I_K}(\boldsymbol{e})\big)
\end{aligned} \tag{30}$$

Using lemma 3 we have that

$$\begin{aligned}
\mathcal{D}\big(g_2; \tilde{\boldsymbol{X}}_{\boldsymbol{I}_K}(\boldsymbol{e}, \{1\})\big) &= g_2 + \tilde{\boldsymbol{X}}_{\boldsymbol{I}_K}^H(\boldsymbol{e}, \{1\})\tilde{\boldsymbol{X}}_{\boldsymbol{I}_K}(\boldsymbol{e}, \{1\}) \\
&\geq g_2 + d_{\min}^2
\end{aligned} \tag{31a}$$

$$\mathcal{D}\big(0; \tilde{\boldsymbol{X}}_{\boldsymbol{I}_K}(\boldsymbol{e}, \{2\})\big) = \tilde{\boldsymbol{X}}_{\boldsymbol{I}_K}^H(\boldsymbol{e}, \{2\})\tilde{\boldsymbol{X}}_{\boldsymbol{I}_K}(\boldsymbol{e}, \{2\}) \geq d_{\min}^2 \tag{31b}$$

$$\mathcal{D}\big(0, 0; \tilde{\boldsymbol{X}}_{\boldsymbol{I}_K}(\boldsymbol{e})\big) = \det\Big(\tilde{\boldsymbol{X}}_{\boldsymbol{I}_K}^H(\boldsymbol{e})\tilde{\boldsymbol{X}}_{\boldsymbol{I}_K}(\boldsymbol{e})\Big) \geq d_{\min}^4 \tag{31c}$$

where the inequalities in (31) hold with equality if and only if $\| \boldsymbol{e} \| = d_{\min}$. Combining (31) with (30) we have that

$$\begin{aligned}
\mathcal{D}\big(g_1, g_2; \tilde{X}_{\boldsymbol{I}_K}(\boldsymbol{e})\big) &\geq g_1\big(g_2 + d_{\min}^2\big) + g_2 d_{\min}^2 + d_{\min}^4 \\
&= \big(g_1 + d_{\min}^2\big)\big(g_2 + d_{\min}^2\big) \\
&= \det\big(\boldsymbol{G} + d_{\min}^2 \boldsymbol{I}_2\big) \\
&= \mathcal{D}(g_1, g_2; d_{\min}\boldsymbol{I}_2)
\end{aligned}$$

Thus, lemma 2 holds for $N = 2$.

Now to prove that lemma 2 holds for all positive integers N, we make the inductive hypothesis that lemma 2 holds for $N = M$ and show that this hypothesis implies that lemma 2 holds for $N = M+1$. To that end, we note that by following the case where $N = 2$ we have

$$\begin{aligned}
&\mathcal{D}\big(g_1, g_2, \ldots, g_M, g_{M+1}; \tilde{\boldsymbol{X}}_{\boldsymbol{I}_K}(\boldsymbol{e})\big) \\
&= g_1 \mathcal{D}\big(g_2, \ldots, g_M, g_{M+1}; \tilde{\boldsymbol{X}}_{\boldsymbol{I}_K}(\boldsymbol{e}, \{1\})\big) \\
&\quad + \mathcal{D}\big(0, g_2, \ldots, g_M, g_{M+1}; \tilde{\boldsymbol{X}}_{\boldsymbol{I}_K}(\boldsymbol{e})\big) \\
&= \sum_{k=1}^{M+1} g_k \mathcal{D}\left(\underbrace{0, \ldots, 0}_{k-1}, g_{k+1}, \ldots, g_M, g_{M+1}; \tilde{\boldsymbol{X}}_{\boldsymbol{I}_K}(\boldsymbol{e}, \{k\})\right) \\
&\quad + \mathcal{D}\left(\underbrace{0, 0, \ldots, 0, 0}_{M+1}; \tilde{\boldsymbol{X}}_{\boldsymbol{I}_K}(\boldsymbol{e})\right)
\end{aligned} \tag{32}$$

Now using lemma 3 and exploiting the inductive hypothesis, we have

$$\begin{aligned}
&\mathcal{D}\left(\underbrace{0, \ldots, 0}_{k-1}, g_{k+1}, \ldots, g_M, g_{M+1}; \tilde{\boldsymbol{X}}_{\boldsymbol{I}_K}(\boldsymbol{e}, \{k\})\right) \\
&\geq \mathcal{D}\left(\underbrace{0, \ldots, 0}_{k-1}, g_{k+1}, \ldots, g_M, g_{M+1}; d_{\min}\boldsymbol{I}_M\right)
\end{aligned} \tag{33}$$

$$\mathcal{D}\left(\underbrace{0, \ldots, 0}_{M+1}; \tilde{\boldsymbol{X}}_{\boldsymbol{I}_K}(\boldsymbol{e})\right) \geq \mathcal{D}\left(\underbrace{0, \ldots, 0}_{M+1}; d_{\min}\boldsymbol{I}_{M+1}\right) \tag{34}$$

where the inequalities hold with equality if and only if $\| \boldsymbol{e} \| = d_{\min}$. Therefore

$$\begin{aligned}
&\sum_{k=1}^{M+1} g_k \mathcal{D}\left(\underbrace{0, \ldots, 0}_{k-1}, g_{k+1}, \ldots, g_M, g_{M+1}; \tilde{\boldsymbol{X}}_{\boldsymbol{I}_K}(\boldsymbol{e}, \{k\})\right) \\
&\quad + \mathcal{D}\left(\underbrace{0, \ldots, 0}_{M+1}; \tilde{\boldsymbol{X}}_{I_K}(\boldsymbol{e})\right) \\
&\geq \sum_{k=1}^{M+1} g_k \mathcal{D}\left(\underbrace{0, \ldots, 0}_{k-1}, g_{k+1}, \ldots, g_M, g_{M+1}; d_{\min}\boldsymbol{I}_M\right) \\
&\quad + \mathcal{D}\left(\underbrace{0, \ldots, 0}_{M+1}; d_{\min}\boldsymbol{I}_{M+1}\right) \\
&= \mathcal{D}(g_1, g_2, \ldots, g_M, g_{M+1}; d_{\min}\boldsymbol{I}_{M+1})
\end{aligned} \tag{35}$$

Combining (33) with (35) we have

$$\begin{aligned}
&\mathcal{D}\big(g_1, g_2, \ldots, g_M, g_{M+1}; \tilde{\boldsymbol{X}}_{\boldsymbol{I}_K}(\boldsymbol{e})\big) \\
&\geq \mathcal{D}(g_1, g_2, \ldots, g_M, g_{M+1}; d_{\min}\boldsymbol{I}_{M+1})
\end{aligned}$$

where the inequality holds with equality if and only if $\| \boldsymbol{e} \| = d_{\min}$. Thus if lemma 2 holds for $N = M$, then it also holds for $N = M+1$. This completes the proof of lemma 2. □