

Optimal Training Sequences and Pilot Tones for OFDM Systems

Jonathan H. Manton, *Member, IEEE*

Abstract—Orthogonal frequency-division multiplex (OFDM) systems transmit data in blocks. The two simplest ways of identifying the channel in OFDM systems are to insert a training sequence between consecutive blocks or to insert pilot tones inside each block. This letter proves that both methods can achieve the same level of performance under certain conditions on the block length.

Index Terms—Channel estimation, orthogonal frequency-division multiplex, pilot tones, training sequence.

I. INTRODUCTION

AN UNKNOWN finite-impulse response channel can be identified in either the time domain, by using a training sequence, or in the frequency domain, by using pilot tones. This letter derives a necessary and sufficient condition for each method to be optimal, and proves that both methods can estimate the channel with the same accuracy.

Throughout, it is assumed that the known complex-valued training sequence $\mathbf{b} = [\mathbf{b}_1, \dots, \mathbf{b}_n]^T$ is sent through an unknown channel $\mathbf{h} = [\mathbf{h}_0, \dots, \mathbf{h}_{L-1}]^T$ whose length does not exceed some known constant L . The received symbols $\mathbf{y}_1, \dots, \mathbf{y}_{n-L+1}$ are given by

$$\mathbf{y}_i = \sum_{k=0}^{L-1} \mathbf{b}_{i-k+L-1} \mathbf{h}_k + \mathbf{w}_i \quad (1)$$

where the \mathbf{w}_i represent additive white Gaussian noise with zero mean and variance σ^2 . This model, which is appropriate for studying time domain channel estimation, will be shown in Section III to encompass frequency domain channel estimation as well. (This is not immediately apparent because the data symbols and pilot tones are interleaved in OFDM systems [1].)

It is convenient to rewrite (1) in matrix form as

$$\mathbf{y} = B\mathbf{h} + \mathbf{w}, \quad E[\mathbf{w}\mathbf{w}^H] = \sigma^2 I \quad (2)$$

where $B \in \mathbb{C}^{(n-L+1) \times L}$ is the Toeplitz matrix with ij th entry $B_{ij} = \mathbf{b}_{L+(i-j)}$. The superscript H denotes Hermitian transpose and I is the identity matrix.

Section II studies the channel estimation problem in the time domain, while Section III studies it in the frequency domain. Section IV explains why optimal training sequences should still

be used even if other information, such as second order statistics, is used to enhance the channel estimate.

Notation: The complex conjugate of x is denoted by \bar{x} . Dirac's delta function $\delta(i)$ is defined to be zero except when $i = 0$, in which case it equals one.

II. OPTIMAL TRAINING SEQUENCE

This section defines what it means for a training sequence to be optimal and then derives a necessary and sufficient condition for a training sequence to be optimal.

Based on the channel equation (2), a number of different "optimal" channel estimators can be constructed. For instance, and quite surprisingly, the James–Stein estimate of \mathbf{h} in the regression (2), as defined in [4], has a smaller mean-square error (MSE) $E[\|\hat{\mathbf{h}} - \mathbf{h}\|^2]$ than the traditional maximum-likelihood estimate (MLE)

$$\hat{\mathbf{h}} = (B^H B)^{-1} B^H \mathbf{y} \quad (3)$$

under certain conditions. However, the James–Stein estimator achieves a smaller MSE by introducing bias, and limited empirical evidence suggests that this bias results in poorer source estimates after channel inversion. Therefore, it is desirable to limit attention to unbiased channel estimates. It is a standard result [2] that the MLE $\hat{\mathbf{h}}$ in (3) is also the minimum variance unbiased estimate. Its variance is

$$E\left[\|\hat{\mathbf{h}} - \mathbf{h}\|^2\right] = \sigma^2 \text{tr}\{(B^H B)^{-1}\}, \quad (4)$$

Remark: A necessary condition for the channel to be identifiable is for $n \geq 2L - 1$, that is, B must have at least as many rows as columns.

Definition 1 (Optimal Training Sequence): The training sequence $\mathbf{b} \in \mathbb{C}^n$ is said to be optimal if no other training sequence of equal or less power $\mathbf{b}^H \mathbf{b}$, and possibly of different length n , results in a smaller variance $E[\|\hat{\mathbf{h}} - \mathbf{h}\|^2]$ of the minimum variance unbiased estimate $\hat{\mathbf{h}}$, defined in (3).

The following preliminary result is required in the proof of Theorem 1.

Lemma 1: Let $B \in \mathbb{C}^{m \times L}$ be such that $\text{tr}\{B^H B\} \leq \lambda L$ for some constant $\lambda > 0$. Then, provided the inverse exists, $\text{tr}\{(B^H B)^{-1}\} \geq \lambda^{-1} L$, with equality if and only if $B^H B = \lambda I$.

Proof: Let $\lambda_1, \dots, \lambda_L \in \mathbb{R}$ be the eigenvalues of $B^H B$. If $\text{tr}\{B^H B\} = \lambda_1 + \dots + \lambda_L = p$ then the Lagrange multiplier technique shows that $\text{tr}\{(B^H B)^{-1}\} = \lambda_1^{-1} + \dots + \lambda_L^{-1} \geq L^2/p$, with equality if and only if $\lambda_1 = \dots = \lambda_L = p/L$. If $p \leq \lambda L$ then $\text{tr}\{(B^H B)^{-1}\} \geq L^2/\lambda L = \lambda^{-1} L$. Moreover, if

Manuscript received November 13, 2000. The associate editor coordinating the review of this letter and approving it for publication was Dr. H. Sari. This work was supported by the Australian Research Council.

The author is with the ARC Special Research Centre for Ultra-Broadband Information Networks, Department of Electrical and Electronic Engineering, The University of Melbourne, Parkville, Vic. 3010, Australia (e-mail: j.manton@ee.mu.oz.au).

Publisher Item Identifier S 1089-7798(01)03699-7.

$\text{tr}\{(B^H B)^{-1}\} = \lambda^{-1}L$ then $\lambda_1 = \dots = \lambda_L = \lambda$, implying that $B^H B = \lambda I$. \square

Theorem 1 (Optimal Training Sequence): Let $\mathbf{b} \in \mathbb{C}^n$ be a training sequence and define L , σ^2 , B and $\hat{\mathbf{h}}$ as in equations (1)–(3).

- 1) The power constraint $\mathbf{b}^H \mathbf{b} = \lambda$ implies $\mathbf{E}\{|\hat{\mathbf{h}} - \mathbf{h}|^2\} \geq \sigma^2 \lambda^{-1} L$.
- 2) The training sequence \mathbf{b} is optimal if and only if $B^H B = (\mathbf{b}^H \mathbf{b})I$, in which case it achieves the lower bound $\mathbf{E}\{|\hat{\mathbf{h}} - \mathbf{h}|^2\} = \sigma^2 (\mathbf{b}^H \mathbf{b})^{-1} L$. A necessary condition for optimality is that $\mathbf{b}_k = 0$ for $k < L$ and $k > n - L + 1$.
- 3) If $2L - 1 \leq n \leq 3L - 2$ then \mathbf{b} is optimal if and only if it is of the form $\mathbf{b}_i = a\delta(i - c)$ for constants $a \in \mathbb{C}$ and $c \in \{L, \dots, n - L + 1\}$.

Proof: Since $\text{tr}\{B^H B\} = \sum_{k=1}^{L-1} k|\mathbf{b}_k|^2 + \sum_{k=L}^{n-L+1} L|\mathbf{b}_k|^2 + \sum_{k=1}^{L-1} k|\mathbf{b}_{n-k+1}|^2$, $\text{tr} B^H B \leq \lambda L$, with equality if and only if $\mathbf{b}_k = 0$ for $k < L$ and $k > n - L + 1$. Lemma 1 thus implies $\text{tr}\{(B^H B)^{-1}\} \geq \lambda^{-1}L$ with equality if and only if $B^H B = \lambda I$. This, together with (4), proves Parts 1 and 2. Finally, if $n \leq 3L - 2$ then expanding $B^H B = \lambda I$ in terms of the \mathbf{b}_i shows that at most one \mathbf{b}_i can be nonzero, proving Part 3. \square

A training sequence of the form $\mathbf{b}_i = a\delta(i - c)$ has a flat power spectrum; it excites all channel frequencies equally. If $n > 3L - 2$ then there exist optimal training sequences which do not excite all channel frequencies equally. Indeed, this is what opens the possibility of optimal pilot tones existing.

III. OPTIMAL PILOT TONES

This section derives necessary and sufficient conditions for pilot tones in OFDM systems to achieve the lower bound in Part 1 of Theorem 1.

Consider an OFDM system [1] with a cyclic prefix of length $L - 1$ and $n - L + 1$ virtual sub-channels. Define

$$C = \begin{bmatrix} 0_{(L-1) \times (n-2L+2)} & I_{L-1} \\ & I_{n-L+1} \end{bmatrix} \quad (5)$$

and let $D \in \mathbb{C}^{(n-L+1) \times (n-L+1)}$ be the normalized DFT matrix with ij th entry

$$D_{ij} = \frac{1}{\sqrt{n-L+1}} e^{-j2\pi((i-1)(j-1)/(n-L+1))}. \quad (6)$$

The data vector $\mathbf{s} \in \mathbb{C}^{n-L+1}$ is transmitted as $\mathbf{b} = CD^H \mathbf{s}$. The received signal \mathbf{y} , corrupted by additive white noise \mathbf{w} , is as given in (2).

If \mathbf{s} is known to the receiver then the minimum variance unbiased estimate of the channel is as given in (3), where B is defined in (2). However, it is not necessary for all the elements of \mathbf{s} to be known.

Lemma 2: Let $\mathbf{s} \in \mathbb{C}^{n-L+1}$ be an arbitrary vector with certain elements known to the receiver and define $\mathbf{b} = CD^H \mathbf{s}$. The variance of the minimum variance unbiased estimate of \mathbf{h} in (1) is $\sigma^2 \text{tr}\{(\tilde{B}^H \tilde{B})^{-1}\}$, where $\tilde{B} \in \mathbb{C}^{(n-L+1) \times L}$ is the Toeplitz matrix with ij th entry $\tilde{B}_{ij} = \tilde{\mathbf{b}}_{L+(i-j)}$, $\tilde{\mathbf{b}} = CD^H \mathbf{J} \mathbf{s}$ and \mathbf{J} is

the diagonal matrix whose i th diagonal element is one if \mathbf{s}_i is known and zero otherwise.

Proof: Rewrite (1) as $\mathbf{y} = H\mathbf{b} + \mathbf{w} = HCD^H \mathbf{s} + \mathbf{w}$ where H is the Toeplitz channel matrix. The unknown elements of \mathbf{s} can be removed by pre-multiplying \mathbf{y} by $D^H J D$. Indeed, since $DHCD^H$ is diagonal, $D^H J D \mathbf{y} = HCD^H \mathbf{J} \mathbf{s} + D^H J D \mathbf{w}$. Thus, $\tilde{\mathbf{y}} = \tilde{B}\tilde{\mathbf{h}} + \tilde{\mathbf{w}}$ where $\tilde{\mathbf{y}} = D^H J D \mathbf{y}$ and $\tilde{\mathbf{w}} = D^H J D \mathbf{w}$. Taking into account the singularity of $\mathbf{E}\{\tilde{\mathbf{w}}\tilde{\mathbf{w}}^H\}$, it can be shown [2, p. 271] that the variance of the minimum variance unbiased estimate of $\tilde{\mathbf{h}}$ is $\sigma^2 \text{tr}\{(\tilde{B}^H \tilde{B})^{-1}\}$.

Remark: If fewer than L elements of \mathbf{s} are known then $\tilde{B}^H \tilde{B}$ will be singular, proving that at least L elements must be known in order to identify the channel.

Comparing Lemma 2 and (4) shows that not knowing an element of \mathbf{s} results in the same performance as would be obtained if the element of \mathbf{s} was known to be zero. It therefore suffices to consider only pilot tone vectors, which are now defined.

Definition 3 (Pilot Tone Vector): The vector \mathbf{s} is called a pilot tone vector if all its nonzero elements are known to the receiver.

Since Lemma 2 and (4) showed that a pilot tone vector \mathbf{s} is equivalent to the training sequence $\mathbf{b} = CD^H \mathbf{s}$, it makes sense to define an optimal pilot tone vector as follows.

Definition 3 (Optimal Pilot Tones): The pilot tone vector \mathbf{s} is said to be optimal if the corresponding training sequence $\mathbf{b} = CD^H \mathbf{s}$ is optimal according to Definition 1.

Remark: There is a one-to-one correspondence between optimal pilot tones and optimal training sequences. Indeed, if \mathbf{b} is an optimal training sequence then, by Theorem 1, $\mathbf{b}_k = 0$ for $k < L$ and $k > n - L + 1$. Thus, there exists an \mathbf{s} for which $\mathbf{b} = CD^H \mathbf{s}$.

The following lemma is required in the proof of Theorem 2.

Lemma 3: Define the entries of $A \in \mathbb{C}^{n \times (n+1)}$ to be $A_{ij} = z_j^{i-1}$ where $z_j \neq z_k$ for $j \neq k$.

- 1) The null space of A is one dimensional.
- 2) If $A[1 \dots 1]^T = 0$ then $z_j \neq 0$ for $j = 0, \dots, n$, and, if $z_0 = 1$, then the z_j must satisfy $1 + z_j + \dots + z_j^n = 0$ for $j = 1, \dots, n$.

Proof: Part 1 follows from the fact that a square Vandermonde matrix has full rank if all its generating elements are nonzero and distinct. If $A[1 \dots 1]^T = 0$ then this also proves that $z_j \neq 0$ for $j = 0, \dots, n$. Define the polynomials $f_i^{(0)} = \sum_{k=0}^n z_k^i$ for $i = 1, \dots, n$. Recursively define $f_i^{(j)}$ for $j = 1, \dots, n - 1$ and $i = 1, \dots, n - j$ by the rule $f_i^{(j)} = f_{i+1}^{(j-1)} - z_j f_i^{(j-1)} + (1/j)(z_{j+1}^i + \dots + z_n^i) f_1^{(j-1)}$. Tedious algebraic manipulation shows that $f_i^{(j)}$ contains no terms in z_1, \dots, z_j , and moreover, $f_1^{(n-1)} = n \sum_{k=0}^n z_0^k z_n^{n-k}$. Therefore, if $A[1 \dots 1]^T = 0$ and $z_0 = 1$ then $1 + z_j + \dots + z_j^n = 0$ must hold for $j = n$, and, by symmetry, must therefore hold for $j = 1, \dots, n$. \square

Theorem 7 (Optimal Pilot Tones): Let $\mathbf{s} \in \mathbb{C}^{n-L+1}$ be a pilot tone vector and define L , C and D as in (1), (5) and (6).

- 1) The pilot tone vector \mathbf{s} is optimal if and only if it satisfies the following two conditions:

$$\sum_{k=1}^{n-L+1} \mathbf{s}_k e^{-j2\pi(i(k-1)/(n-L+1))} = 0, \quad \text{for } i = 1, \dots, L - 1. \quad (7)$$

$$\sum_{k=1}^{n-L+1} |\mathbf{s}_k|^2 e^{-j2\pi(i(k-1))/(n-L+1)} = 0, \quad \text{for } i = 1, \dots, L-1. \quad (8)$$

- 2) If $n-L+1$ is not a multiple of L then there does not exist an optimal pilot tone vector with L elements nonzero.
- 3) If there exists an integer p such that $n-L+1 = pL$, then \mathbf{s} is optimal with L elements nonzero if and only if it has the form $\mathbf{s}_k = a \sum_{i=1}^L \delta(k-ip+c)$ for constants $a \in \mathbb{C}$ and $c \in \{0, \dots, p-1\}$.

Proof: The i th element of \mathbf{b} is

$$\mathbf{b}_i = \frac{1}{\sqrt{n-L+1}} \sum_{k=1}^{n-L+1} \mathbf{s}_k e^{j2\pi((i-L)(k-1)/(n-L+1))}. \quad (9)$$

Thus, the condition $\mathbf{b}_k = 0$ for $k < L$ and $k > n-L+1$ in Theorem 1 can be expressed as (7). Under this condition, $B^H B$ is a Toeplitz matrix. Let α_i be such that $(B^H B)_{ij} = \alpha_{i-j}$. Then

$$\alpha_i = \sum_{k=1}^{n-L+1} |\mathbf{s}_k|^2 e^{j2\pi(i(k-1)/(n-L+1))}. \quad (10)$$

Thus, the condition $B^H B = \mathbf{b}^H \mathbf{b} \mathbf{I}$ in Theorem 1 can be expressed as $\alpha_i = (\mathbf{b}^H \mathbf{b}) \delta(i)$. For $i \neq 0$, this is equivalent to (8). Under condition (7), α_0 automatically equals $\mathbf{b}^H \mathbf{b}$. This proves Part 1.

Assume that the L elements $\mathbf{s}_{t_1}, \dots, \mathbf{s}_{t_L}$ are nonzero. Define $\tilde{\mathbf{s}} = [\mathbf{s}_{t_1}, \dots, \mathbf{s}_{t_L}]^T$ and let the unusual notation $|\tilde{\mathbf{s}}|^2$ denote the vector $|\tilde{\mathbf{s}}|^2 = [|\mathbf{s}_{t_1}|^2, \dots, |\mathbf{s}_{t_L}|^2]^T$. Then, the constraints (7) and (8) can be written in matrix form as $A\tilde{\mathbf{s}} = 0$ and $A|\tilde{\mathbf{s}}|^2 = 0$, where the ij th element of the matrix $A \in \mathbb{C}^{(L-1) \times L}$ is $A_{ij} = e^{j\theta_j}$ with $\theta_j = -2\pi((l_j-1)/(n-L+1))$. From Lemma 3, the dimension of the null space of A is one. Therefore, a necessary condition for $\tilde{\mathbf{s}}$ to satisfy both $A\tilde{\mathbf{s}} = 0$ and $A|\tilde{\mathbf{s}}|^2 = 0$ is for $\tilde{\mathbf{s}} = \lambda|\tilde{\mathbf{s}}|^2$ for some constant $\lambda \in \mathbb{C}$. Since no element of $\tilde{\mathbf{s}}$ is zero, the only way for $\tilde{\mathbf{s}} = \lambda|\tilde{\mathbf{s}}|^2$ to hold is if $\tilde{\mathbf{s}} = \lambda'[1 \dots 1]^T$ for some constant $\lambda' \in \mathbb{C}$. From Lemma 3, this implies the θ_j must be uniformly spaced, proving Parts 2 and 3. \square

Theorem 2 implies that there is no need to use a training sequence in OFDM systems. Indeed, the shortest optimal training sequence is of length $2L-1$ while the smallest number of optimally placed pilot tones is L . This also implies that the extra $L-1$ symbols required for the cyclic prefix in OFDM systems are not at all inefficient; $L + (L-1) = 2L-1$, the

smallest number of transmitted symbols required to identify the channel. The results of this section also imply that, somewhat counter-intuitively, the accuracy of the channel estimate based on optimally spaced pilot tones is not affected by channel spectral nulls.

IV. INCORPORATING OTHER INFORMATION

Blind estimation techniques can identify the channel based on known statistical or algebraic properties of the data symbols. (Algebraic properties arise from precoding the source symbols [3].) It is natural to ask whether or not the optimal training sequences and pilot tones proposed in this paper remain optimal if other information is used to enhance the channel estimate. This section briefly shows that, in certain cases at least, they do remain optimal.

Consider a training sequence followed by a data sequence. Let $\hat{\mathbf{h}}$ denote the channel estimate (3) obtained by using the training sequence. Assume that some property of the data sequence is used to obtain another channel estimate $\hat{\mathbf{h}}^{(2)}$. If the channel noise is white then $\hat{\mathbf{h}}$ and $\hat{\mathbf{h}}^{(2)}$ will be statistically independent. Therefore, there is no reason not to use an optimal training sequence.

A similar argument holds for an OFDM system. As the proof of Lemma 2 shows, the channel estimate $\hat{\mathbf{h}}$ based on the pilot tones is independent of the other data symbols sent. Thus, if the channel noise is white Gaussian, $\hat{\mathbf{h}}$ will be independent of any channel estimate based on the data symbols.

V. CONCLUSION

This letter derived necessary and sufficient conditions for training sequences and pilot tones to be optimal. It was proved that training sequences and pilot tones both can achieve the lower bound on the variance of the channel estimate.

REFERENCES

- [1] B. Le Floch, M. Alard, and C. Berrou, "Coded orthogonal frequency division multiplex," *Proc. IEEE*, vol. 83, pp. 982–996, June 1995.
- [2] J. R. Magnus and H. Neudecker, *Matrix Differential Calculus with Applications in Statistics and Econometrics*. New York: Wiley, 1994.
- [3] J. H. Manton and Y. Hua, "A frequency domain deterministic approach to channel identification," *IEEE Signal Processing Lett.*, vol. 6, pp. 323–326, Dec. 1999.
- [4] J. H. Manton, V. Krishnamurthy, and H. V. Poor, "James–Stein state filtering algorithms," *IEEE Trans. Signal Processing*, vol. 46, pp. 2431–2447, Sept. 1998.