On the Sensitivity of a Suboptimal Precoding Scheme for Frequency-Selective Block-Based Channels With Respect to Channel Inaccuracies

Athanasios P. Liavas

***Abstract*—Water-filling is the precoding scheme that achieves capacity in Gaussian parallel or block-based channels. A suboptimal precoding scheme, which has been observed to perform quite close to water-filling, in some cases of great interest, is when all "active" eigenvectors of the input data covariance matrix receive the same power. Both techniques require perfect channel knowledge at the transmitter. In this work, we consider the sensitivity of the suboptimal scheme when a channel estimate is used at the transmitter as if it were the true channel. Using tools from matrix perturbation theory, we derive closed-form expressions and bounds relating the channel estimation error covariance matrix with the mean mutual information decrease. We thus uncover the factors that determine the behavior of the suboptimal precoding scheme under channel uncertainties. Simulations are in agreement with our theoretical results.**

***Index Terms*—Channel uncertainties, invariant subspaces, perturbation expansions, precoding, water-filling.**

I. Introduction

Block-based transmission is common in communications, with examples including packet-based transmission over frequency-selective channels, orthogonal frequency-division multiplexing (OFDM), discrete multitone (DMT), and flat-fading multiple-input multiple-output (MIMO) wireless communications.

If the channel is known at the transmitter, due to, e.g., feedback, then it is possible to maximize the information rate by precoding the channel input. This problem has been considered for frequency-selective channels in [1] and [2] and for MIMO channels in [3], which compute the input data covariance matrix that maximizes the mutual information between the channel input and output in the white Gaussian noise case. It turns out [3], that the eigenvectors of the optimal input covariance matrix are the right singular vectors of the channel matrix, while its eigenvalues are computed through water-filling using the (squared) singular values of the channel matrix. A significant observation of [1] is that a suboptimal input data covariance matrix, having all its nonzero eigenvalues equal, leads, in some important cases, e.g., digital subscriber line (DSL), to a mutual information that almost coincides with that obtained through water-filling.

A question that is directly related with the practical success of the aforementioned precoding schemes concerns their sensitivity with respect to channel and noise statistics inaccuracies. In this work, we consider the behavior of the suboptimal scheme when we use at the transmitter a channel estimate as if it were the true channel. This problem is of importance because the suboptimal scheme is computationally simpler than the optimal, while its performance has been observed to be very close to the optimal (practically, the same) in many cases of great interest (i.e., DSL, high signal-to-noise ratio (SNR)).

Manuscript received September 15, 2004; revised May 21, 2005. This work was supported in part by the EU under U-BROAD STREP #506790. The material in this correspondence was presented in part at the IEEE International Conference on Acoustics, Speech, and Signal Processing, Philadelphia, PA, March 2005.

The author is with the Department of Electronic and Computer Engineering, Technical University of Crete, 73100 Kounoupidiana, Chania, Greece (e-mail: liavas@telecom.tuc.gr).

Communicated by M. Médard, Associate Editor on Communications.
Digital Object Identifier 10.1109/TIT.2005.853328

Communication under channel uncertainty is a vast area, nicely summarized in [4] (see also references therein). More recent studies treating the impact of imperfect channel knowledge at the receiver include [5], [6]. Studies pursuing optimal training in frequency-selective block-based fading channels appear in [7], [8].

Using tools from matrix perturbation theory, we derive closed-form expressions and bounds relating the mean mutual information decrease with the channel estimation error covariance matrix. We thus uncover the factors that determine the sensitivity of the suboptimal precoding scheme under channel uncertainties. We observe that, for sufficiently good channel estimation, the mutual information degradation is insignificant.

Notation: Superscripts $(\cdot)^T, \bar{(\cdot)}, (\cdot)^H$ and $(\cdot)^\sharp$ denote, respectively, transpose, element-wise conjugate, Hermitian transpose, and pseudoinverse. I_i denotes the i-dimensional identity matrix, e_i denotes the ith canonical vector, that is, the vector with a 1 at the ith position and zeros elsewhere, 1_i denotes the i-dimensional vector composed of ones and $0_{m\times n}$ denotes the $m \times n$ zero matrix (when its dimensions are clear from the context, they are omitted). $\|\cdot\|$ and $\|\cdot\|_F$ denote, respectively, the matrix or vector 2-norm and the matrix Frobenious norm. We remind that the 2-norm of a positive semidefinite matrix equals its largest eigenvalue. $\mathcal{E}[\cdot]$ denotes expectation, while $\mathcal{R}(\cdot), |\cdot|$, and $\mathrm{Tr}(\cdot)$ denote, respectively, the column space, the determinant, and the trace of the matrix argument. For compatible matrices A and B, it holds that $\mathrm{Tr}(AB) = \mathrm{Tr}(BA)$. Symbol $\otimes$ denotes the Kronecker product. For compatible matrices A, B, and C, it holds that

$$\mathrm{vec}(ABC) = (C^T \otimes A)\,\mathrm{vec}(B)$$

where $\mathrm{vec}(\cdot)$ denotes the vectorization operator. We remind that the eigenvalues of $A\otimes B$ equal the products of the eigenvalues of A and B.

We define matrix $\mathcal{M}$ as

$$\mathcal{M} \triangleq \begin{bmatrix} 1_{N-1} \otimes \begin{bmatrix} I_{\nu+1} \\ 0_{N\times(\nu+1)} \end{bmatrix} \\ I_{\nu+1} \end{bmatrix}. \tag{1}$$

Finally, we define implicitly the commutation matrix $\mathcal{K}$ as [12, p. 115]

$$\mathrm{vec}(H) = \mathcal{K}\mathrm{vec}(H^T).$$

Structure: In Section II, we present the frequency-selective block-based channel model. In Section III, we consider the mutual information achieved by a suboptimal precoding scheme in the ideal and nonideal cases, i.e., perfect and imperfect channel state information (CSI) at the transmitter, respectively. In Section IV, we derive a second-order approximation, with respect to channel estimation errors, to the mutual information decrease. In Section V, we develop expressions and bounds relating the mean mutual information degradation with the channel estimation error covariance matrix. In Section VI, we check our theoretical results with numerical simulations and in Section VII we present some conclusions.

II. The Channel Model

We consider the baseband-equivalent discrete-time frequency-selective noisy communication channel modeled by the νth order linear time-invariant system with input–output relation [1]

$$y_l = \sum_{i=0}^{\nu} h_i x_{l-i} + w_l, \qquad l = k, \dots, k+N-1 \tag{2}$$

where $h \triangleq [h_0 \ \cdots \ h_\nu]^T$ is the channel impulse response vector and x_l, w_l, and y_l denote, respectively, the samples of the channel input, noise, and output. Defining the vectors

$$y \triangleq [y_{k+N-1} \ \cdots \ y_k]^T$$
$$x \triangleq [x_{k+N-1} \ \cdots \ x_{k-\nu}]^T$$

and

$$w \triangleq [w_{k+N-1} \ \cdots \ w_k]^T$$

we may rewrite (2) in matrix form as

$$y = Hx + w$$

where H is the $N \times (N+\nu)$ filtering matrix defined as

$$H \triangleq \begin{bmatrix} h_0 & \cdots & \cdots & h_\nu & & \\ & \ddots & & & \ddots & \\ & & h_0 & \cdots & \cdots & h_\nu \end{bmatrix}. \tag{3}$$

The noise vector is composed of independent and identically distributed complex-valued zero-mean circularly symmetric Gaussian random variables with covariance matrix

$$R_w \triangleq \mathcal{E}[ww^H] = \sigma_w^2 I_N.$$

The input symbols comprising x are complex-valued zero-mean circularly symmetric Gaussian (in order to achieve capacity), independent of the noise, with covariance matrix

$$R_x \triangleq \mathcal{E}[xx^H].$$

III. Capacity Analysis

A. Ideal Case: Perfect CSI at the Transmitter

A problem that has been considered in [1]–[3], is the computation of the input covariance matrix that maximizes the mutual information $I(X;Y)$ between the input vector X and the output vector Y of the above block-based channel.

Toward this end, the following singular value factorizations will prove useful:

$$H = V[\Sigma^{1/2} \quad 0]U^H \tag{4}$$

$$A \triangleq H^H H = U \begin{bmatrix} \Sigma & 0 \\ 0 & 0 \end{bmatrix} U^H \tag{5}$$

where V and U are, respectively, $N \times N$ and $(N+\nu) \times (N+\nu)$ unitary matrices consisted of the singular vectors of H and $\Sigma \triangleq \mathrm{diag}(\sigma_1, \ldots, \sigma_N)$, with σ_i being the ith largest eigenvalue of A. For later use, we define the ith column of U and V as u_i and v_i, respectively, and the matrices composed of the first n columns of U and V as U_n and V_n, respectively. Matrices U and V are partitioned as $U = [U_n \ \ U_n^c]$ and $V = [V_n \ \ V_n^c]$, with the definitions of U_n^c and V_n^c being obvious. Finally, we define $\Sigma_n \triangleq \mathrm{diag}(\sigma_1, \ldots, \sigma_n)$ and

$$\Sigma_n^c \triangleq \mathrm{diag}(\sigma_{n+1}, \ldots, \sigma_N, \underbrace{0, \ldots, 0}_{\nu}).$$

It turns out that the optimal input data covariance matrix is [3]

$$R_x^{\mathrm{opt}} = U \, \mathrm{diag}(\delta_1, \ldots, \delta_N, \underbrace{0, \ldots, 0}_{\nu}) \, U^H$$

where terms $\delta_i, i = 1, \ldots, N$ are computed through water-filling using the eigenvalues of A. A suboptimal approach, which has been observed to perform very close to the optimal in some cases of great interest [1], is to assume that all the nonzero eigenvalues of R_x^{opt} are equal, i.e.,

$$\delta = \delta_1 = \delta_2 = \cdots = \delta_n = \frac{E}{n}$$

where E is the total input power and $n \leq N$. In this case, the input data covariance matrix is $R_x = \delta \, U_n U_n^H$.

The mutual information per input sample between the channel input and output in the suboptimal case is given by [3]

$$\begin{aligned} I(X;Y) &= \frac{1}{N+\nu} \log_2 \left| I_N + \frac{1}{\sigma_w^2} H R_x H^H \right| \\ &= \frac{1}{N+\nu} \log_2 \left| I_{N+\nu} + \frac{1}{\sigma_w^2} H^H H R_x \right| \\ &= \frac{1}{N+\nu} \log_2 \left| I_{N+\nu} + \frac{\delta}{\sigma_w^2} U_N \Sigma_N U_N^H U_n U_n^H \right| \\ &= \frac{1}{N+\nu} \log_2 \left| I_n + \frac{\delta}{\sigma_w^2} U_n^H U_N \Sigma_N U_N^H U_n \right| \\ &= \frac{1}{N+\nu} \log_2 \left| I_n + \frac{\delta}{\sigma_w^2} \Sigma_n \right| \\ &= \frac{1}{N+\nu} \log_2 \left| \frac{\delta}{\sigma_w^2} \left(\Sigma_n + \frac{\sigma_w^2}{\delta} I_n \right) \right|. \end{aligned} \tag{6}$$

If we define

$$\mathcal{A} \triangleq \Sigma_n + \frac{\sigma_w^2}{\delta} I_n \tag{7}$$

and λ_i its ith eigenvalue (note that $\lambda_i = \sigma_i + \frac{\sigma_w^2}{\delta}$, then

$$\begin{aligned} I(X;Y) &= \frac{1}{N+\nu} \log_2 \left(\prod_{i=1}^{n} \frac{\delta}{\sigma_w^2} \lambda_i \right) \\ &= \underbrace{\frac{n}{N+\nu} \log_2 \left(\frac{\delta}{\sigma_w^2} \right)}_{c} + \frac{1}{N+\nu} \sum_{i=1}^{n} \log_2 \lambda_i. \end{aligned} \tag{8}$$

B. Nonideal Case: Imperfect CSI at the Transmitter

In practice, we do not know the true channel h but, instead, its estimate $\hat{h}$. Using at the transmitter $\hat{h}$ as if it were the true channel, we compute

$$\hat{A} \triangleq \hat{H}^H \hat{H} = \hat{U} \begin{bmatrix} \hat{\Sigma} & 0 \\ 0 & 0 \end{bmatrix} \hat{U}^H = \hat{U}_N \hat{\Sigma}_N \hat{U}_N^H.$$

In this case, we consider as "optimal" the input covariance matrix $\hat{R}_x = \delta \, \hat{U}_n \hat{U}_n^H$. The corresponding input and output random vectors are denoted as $\hat{X}$ and $\hat{Y}$, in order to be distinguished from the ideal-case quantities X and Y.

We define the errors in $\hat{h}, \hat{H}$, and $\hat{U}_n$ and the first-order error in $\hat{A}$ as follows:

$$\Delta h \triangleq \hat{h} - h, \quad \Delta H \triangleq \hat{H} - H, \quad \Delta U_n \triangleq \hat{U}_n - U_n \tag{9}$$

$$\Delta A \triangleq \hat{A} - A = H^H \Delta H + \Delta H^H H + O(\|\Delta h\|^2). \tag{10}$$

The mutual information between $\hat{X}$ and $\hat{Y}$ is

$$\begin{aligned} I(\hat{X};\hat{Y}) &= \frac{1}{N+\nu} \log_2 \left| I_N + \frac{1}{\sigma_w^2} H \hat{R}_x H^H \right| \\ &= \frac{1}{N+\nu} \log_2 \left| I_{N+\nu} + \frac{1}{\sigma_w^2} H^H H \hat{R}_x \right| \\ &= \frac{1}{N+\nu} \log_2 \left| I_{N+\nu} + \frac{\delta}{\sigma_w^2} U_N \Sigma_N U_N^H \hat{U}_n \hat{U}_n^H \right| \\ &= \frac{1}{N+\nu} \log_2 \left| I_n + \frac{\delta}{\sigma_w^2} \hat{U}_n^H U_N \Sigma_N U_N^H \hat{U}_n \right| \\ &= \frac{1}{N+\nu} \log_2 \left| \frac{\delta}{\sigma_w^2} \hat{\mathcal{A}} \right| \end{aligned} \tag{11}$$

where

$$\hat{\mathcal{A}} = \mathcal{A} + \Delta \mathcal{A} \tag{12}$$

with

$$\begin{aligned}\Delta\mathcal{A} &\triangleq \Delta U_n^H U_N \Sigma_N U_N^H U_n + U_n^H U_N \Sigma_N U_N^H \Delta U_n \\ &\quad + \Delta U_n^H U_N \Sigma_N U_N^H \Delta U_n \\ &= \Delta U_n^H U_n \Sigma_n + \Sigma_n U_n^H \Delta U_n + \Delta U_n^H U_N \Sigma_N U_N^H \Delta U_n. \quad (13)\end{aligned}$$

Thus,

$$I(\hat{X};\hat{Y}) = c + \frac{1}{N+\nu}\sum_{i=1}^{n} \log_2 \hat{\lambda}_i \quad (14)$$

where $\hat{\lambda}_i$ denotes the ith eigenvalue of $\hat{\mathcal{A}}$. The eigenvalue perturbation is defined as

$$\Delta\lambda_i \triangleq \hat{\lambda}_i - \lambda_i. \quad (15)$$

In order to determine the influence of the channel estimation errors on the performance of the suboptimal precoding scheme, we must assess the difference between the mutual information achieved in the ideal and the nonideal cases, defined as

$$\Delta I \triangleq I(X;Y) - I(\hat{X};\hat{Y}). \quad (16)$$

IV. Perturbation Analysis

A. Perturbations of Invariant Subspaces and Eigenvalues

In order to compute ΔI, we must relate the ideal and perturbed eigenvalues λ_i and $\hat{\lambda}_i$, which in turn needs relating matrices U_n and $\hat{U}_n$. Toward this end, we shall use tools from matrix perturbation theory.

We start by considering perturbations of invariant subspaces. Theorem 2.7 of [9, p. 236] gives the conditions under which if the columns of U_n and U_n^c form orthonormal bases for simple invariant subspaces of A (as it happens in our case for $n \leq N$ (see (5)), then there is a unique $(N+\nu-n)\times n$ matrix P such that the columns of

$$\hat{U}_n = (U_n + U_n^c P)\left(I_n + P^H P\right)^{-1/2} \quad (17)$$

and

$$\hat{U}_n^c = \left(U_n^c - U_n P^H\right)\left(I_{N+\nu-n} + PP^H\right)^{-1/2} \quad (18)$$

form orthonormal bases for simple orthogonal invariant subspaces of $\hat{A} = A + \Delta A$.

Loosely speaking, we can find such a P if the perturbation ΔA is sufficiently smaller than $\lambda_n - \lambda_{n+1}$, that is, the gap between the smallest eigenvalue associated with U_n and the largest eigenvalue associated with U_n^c. Since in this work we are mainly interested in perturbation expansions, we will assume that the perturbations are sufficiently small, guaranteeing the existence of P, and we will derive perturbation expansions, whose accuracy will be tested by simulations. Under the sufficiently small perturbation assumption, P can be computed by solving a *nonlinear* matrix equation. More specifically, since $\mathcal{R}(\hat{U}_n)$ and $\mathcal{R}(\hat{U}_n^c)$ are orthogonal invariant subspaces of $\hat{A}$, we have [9, p. 220]

$$\hat{U}_n^{cH}\hat{A}\,\hat{U}_n = 0.$$

Using (17) and (18), we obtain

$$\left(U_n^{cH} - PU_n^H\right)(A+\Delta A)(U_n + U_n^c P) = 0 \quad (19)$$

which is nonlinear in P. This implies that it is very difficult, if not impossible, to find a closed-form expression for P. It holds that $P = O(\|\Delta A\|)$ [9, p. 236], which, using (10), gives $P = O(\|\Delta h\|)$. By performing calculations in (19), ignoring higher order terms, that is, terms involving products of P and ΔA, and using (5), we construct the *linearized* version of the above equation for the first-order approximation $\tilde{P}$ of P (i.e., $\tilde{P} = P + O(\|\Delta h\|^2)$), as follows (see also [10]):

$$\Sigma_n^c \tilde{P} - \tilde{P}\Sigma_n = -U_n^{cH}\Delta A U_n. \quad (20)$$

Applying the vectorization operator at both sides of the previous equation, we obtain the closed-form expression for $\tilde{P}$

$$\left(I_n \otimes \Sigma_n^c - \Sigma_n \otimes I_{N+\nu-n}\right)\operatorname{vec}(\tilde{P}) = -\operatorname{vec}\left(U_n^{cH}\Delta A U_n\right). \quad (21)$$

Now, we turn to perturbations of eigenvalues.

Result: For Hermitian matrices $\mathcal{A}$ and $\hat{\mathcal{A}} = \mathcal{A} + \Delta\mathcal{A}$, if λ is a simple eigenvalue of $\mathcal{A}$ with associated eigenvector e, then there exists $\hat{\lambda}$ unique eigenvalue of $\hat{\mathcal{A}}$ such that

$$\hat{\lambda} = \lambda + e^H \Delta\mathcal{A} e + O(\|\Delta\mathcal{A}\|^2). \quad (22)$$

Proof: This result can be proved using Theorem 2.3 of [9, p. 183], which holds in the general non-Hermitian case, by noting that, in the Hermitian case, the left and right eigenvectors coincide.

B. Second-Order Approximation to ΔI

In this subsection, we derive a second-order approximation to ΔI. We start by providing second-order approximations to $\Delta\mathcal{A}$ and $\delta\lambda_i$.

Lemma 1: Let $\Delta\mathcal{A}$ be the perturbation to $\mathcal{A}$ defined in (13) and $\tilde{P}$ the first-order approximation to P defined in (20). Then

$$\Delta\mathcal{A} = -\frac{1}{2}\Sigma_n \tilde{P}^H\tilde{P} - \frac{1}{2}\tilde{P}^H\tilde{P}\Sigma_n + \tilde{P}^H\Sigma_n^c\tilde{P} + O(\|\Delta h\|^3). \quad (23)$$

Proof: The proof is provided in Appendix I. □

Notice that $\Delta\mathcal{A} = O(\|\tilde{P}\|^2) = O(\|\Delta h\|^2)$.

Assuming that the eigenvalues of the diagonal matrix $\mathcal{A}$ are simple, we obtain that its eigenvectors are the canonical vectors e_i. This leads to the following second-order approximation to the eigenvalue perturbation.

Lemma 2: Let $\Delta\lambda_i$ be the perturbation on λ_i defined in (15). Then

$$\Delta\lambda_i = -e_i^H\tilde{P}^H\underbrace{\left(\sigma_i I_{N+\nu-n} - \Sigma_n^c\right)}_{\mathcal{S}_i}\tilde{P}e_i + O(\|\Delta h\|^3). \quad (24)$$

Proof: Using (22) and (23) and the fact that $\Sigma_n e_i = \sigma_i e_i$, we obtain

$$\begin{aligned}\Delta\lambda_i &= -\frac{1}{2}e_i^H\Sigma_n\tilde{P}^H\tilde{P}e_i - \frac{1}{2}e_i^H\tilde{P}^H\tilde{P}\Sigma_n e_i \\ &\quad + e_i^H\tilde{P}^H\Sigma_n^c\tilde{P}e_i + O(\|\Delta h\|^3) \\ &= -\sigma_i e_i^H\tilde{P}^H\tilde{P}e_i + e_i^H\tilde{P}^H\Sigma_n^c\tilde{P}e_i + O(\|\Delta h\|^3)\end{aligned}$$

which is (24) and proves the lemma. □

Notice that $|\Delta\lambda_i| = O(\|\Delta h\|^2)$.

We can now provide a second-order approximation to the mutual information decrease due to channel estimation errors.

Proposition 1: Let ΔI be the difference between the mutual information achieved in the ideal and the nonideal cases defined in (16). Then

$$\Delta I = \frac{\log_2 e}{N+\nu}\sum_{i=1}^{n}\frac{1}{\lambda_i}e_i^H\tilde{P}^H\mathcal{S}_i\tilde{P}e_i + O(\|\Delta h\|^3) \quad (25)$$

where $\mathcal{S}_i$ is defined in (24).

Proof: Using (14) and (15), we obtain

$$\begin{aligned} I(\hat{X};\hat{Y}) &= c + \frac{1}{N+\nu}\sum_{i=1}^{n}\log_2\hat{\lambda}_i \\ &= c + \frac{1}{N+\nu}\sum_{i=1}^{n}\log_2(\lambda_i+\Delta\lambda_i) \\ &= c + \frac{1}{N+\nu}\sum_{i=1}^{n}\log_2\left(\lambda_i\left(1+\frac{\Delta\lambda_i}{\lambda_i}\right)\right) \\ &\overset{\text{(a)}}{=} I(X;Y) + \frac{1}{N+\nu}\sum_{i=1}^{n}\log_2\left(1+\frac{\Delta\lambda_i}{\lambda_i}\right) \\ &\overset{\text{(b)}}{=} I(X;Y) + \frac{\log_2 e}{N+\nu}\sum_{i=1}^{n}\frac{\Delta\lambda_i}{\lambda_i} + O(|\Delta\lambda_i|^2) \\ &\overset{\text{(c)}}{=} I(X;Y) - \frac{\log_2 e}{N+\nu}\sum_{i=1}^{n}\frac{1}{\lambda_i}e_i^H\tilde{P}^H\mathcal{S}_i\tilde{P}e_i + O(\|\Delta h\|^3) \end{aligned}$$

where at (a) we used (8), at (b) we used the first-order expansion $\ln(1+\Delta x) = \Delta x + O(|\Delta x|^2)$, and at (c) we used (24). Using (16), we obtain (25) to prove the proposition. □

We observe that if the perturbations are sufficiently small, then we will always have decrease of the mutual information because the sum in (25) is nonnegative.

In the sequel, in order to simplify notation, we shall omit the $O(\cdot)$ terms, because they are obvious from the above analysis.

V. Computation of the Mean Mutual Information Decrease

In this section, we assume that the channel estimation error Δh is zero-mean, circular, i.e., $\mathcal{E}\{\Delta h\,\Delta h^T\} = 0_{(\nu+1)\times(\nu+1)}$, with covariance matrix $R_{\Delta h} \triangleq \mathcal{E}\{\Delta h\,\Delta h^H\}$, and we derive second-order approximations to the mean degradation of the mutual information due to channel inaccuracies as

$$\mathcal{E}\{\Delta I\} = \frac{\log_2 e}{N+\nu}\sum_{i=1}^{n}\frac{1}{\lambda_i}\mathcal{E}\left\{\mathrm{Tr}\left(e_i^H\tilde{P}^H\mathcal{S}_i\tilde{P}e_i\right)\right\} \tag{26}$$

where the expectation is with respect to the channel errors.

A. Case $n = N$: Exact Expression

The case $n = N$ seems the most natural since we distribute equally the power to all available degrees of freedom. Furthermore, it leads to simple expressions. For these reasons, we consider it first.

Proposition 2: The mean mutual information degradation of the suboptimal scheme due to channel estimation errors with covariance matrix $R_{\Delta h}$ for $n = N$ is

$$\mathcal{E}\{\Delta I\} = \frac{\log_2 e}{N+\nu}\mathrm{Tr}\left(R_{\Delta h}^T\mathcal{M}^H\left(\bar{V}_N\mathcal{A}^{-1}V_N^T\otimes U_N^cU_N^{cH}\right)\mathcal{M}\right). \tag{27}$$

Proof: The proof is provided in Appendix II. □

Expression (27) relates the channel estimation error covariance matrix with the mean mutual information decrease. Matrix $\mathcal{M}$ depends only on the channel order ν and the block size N, while matrices $\bar{V}_N$ and U_N^c have orthonormal columns. Thus, for fixed ν and N, the term that mainly determines the magnification of the channel estimation errors is $\mathcal{A}^{-1}$, whose norm is large if σ_N and $((\sigma_w^2)/\delta)$ are small. However, it seems difficult to quantify precisely how large the magnification may be. A related bound is derived in the next subsection.

Any reasonable channel estimation procedure results in a channel estimation error covariance matrix that tends to zero as the noise variance tends to zero. On the other hand, as the noise variance tends to zero, $\|\mathcal{A}^{-1}\|$ tends to $\frac{1}{\sigma_N}$, which is finite. Thus, we obtain the intuitively satisfying fact that the mean mutual information degradation tends to zero as the noise variance tends to zero.

B. Case $n = N$: Bound

A case that is commonly encountered in practice is $R_{\Delta h} = \sigma_{\Delta h}^2 I_{\nu+1}$ [11, p. 786], giving $\mathcal{E}\{\|\Delta h\|^2\} = (\nu+1)\sigma_{\Delta h}^2$. Using the inequality [12, p. 44]

$$\mathrm{Tr}(X^HYX) \le \lambda_{\max}(Y)\,\mathrm{Tr}(X^HX) \tag{28}$$

for a positive semidefinite matrix Y and the relations

$$\left\|\bar{V}_N\mathcal{A}^{-1}V_N^T\otimes U_N^cU_N^{cH}\right\| = \|\mathcal{A}^{-1}\| \tag{29}$$

$$\mathrm{Tr}(\mathcal{M}^H\mathcal{M}) = \|\mathcal{M}\|_F^2 = N(\nu+1) \tag{30}$$

we obtain an upper bound for the mean mutual information decrease as

$$\begin{aligned} \mathcal{E}\{\Delta I\} &= \frac{\log_2 e\,\sigma_{\Delta h}^2}{N+\nu}\mathrm{Tr}\left(\mathcal{M}^H\left(\bar{V}_N\mathcal{A}^{-1}V_N^T\otimes U_N^cU_N^{cH}\right)\mathcal{M}\right) \\ &\le \frac{N(\nu+1)\log_2 e\sigma_{\Delta h}^2}{N+\nu}\|\mathcal{A}^{-1}\| \\ &= \frac{N(\nu+1)\log_2 e\,\sigma_{\Delta h}^2}{(N+\nu)\left(\sigma_N+\frac{\sigma_w^2}{\delta}\right)} \\ &\le \frac{\log_2 e}{\sigma_N+\frac{\sigma_w^2}{\delta}}\mathcal{E}\{\|\Delta h\|^2\}. \end{aligned} \tag{31}$$

This bound implies that the channel estimation error may be significantly magnified if the smallest nonzero singular value of H^HH, σ_N, and term $\frac{\sigma_w^2}{\delta}$ are small.

We may go one step further and relate $\sigma_{\Delta h}^2$ with the noise variance σ_w^2. For example, if we use the maximum-likelihood channel estimation procedure described in [11, p. 784], then

$$\sigma_{\Delta h}^2 = \frac{\sigma_w^2}{P_x(N_{\mathrm{tr}}-\nu)}$$

where P_x is the power per input sample, i.e., $(N+\nu)P_x = N\delta$, and N_{tr} is the number of training symbols used for the channel estimation. Then, substituting the value of $\sigma_{\Delta h}^2$ in the third line of (31), we obtain

$$\begin{aligned} \mathcal{E}\{\Delta I\} &\le \underbrace{\frac{\log_2 e\,(\nu+1)}{(N_{\mathrm{tr}}-\nu)}}_{d}\frac{\sigma_w^2}{\delta\sigma_N+\sigma_w^2} \\ &= d\,\frac{1}{1+\frac{\delta\sigma_N}{\sigma_w^2}} \end{aligned} \tag{32}$$

where the term d depends only on the channel order and the number of training symbols. Thus, the mean mutual information degradation bound depends mainly on the ratio $\frac{\delta\sigma_N}{\sigma_w^2}$ (note that $\delta\sigma_N$ is the power of the noiseless channel output corresponding to the Nth eigenvector u_N). If this ratio is large, then the mutual information degradation will be small, otherwise the degradation may be significant.

C. Case $n \le N$: Exact Expression

Proposition 3: The mean mutual information degradation of the suboptimal scheme due to channel estimation errors with covariance matrix $R_{\Delta h}$ for $n \le N$ is

$$\mathcal{E}\{\Delta I\} = \frac{\log_2 e}{N+\nu}\mathrm{Tr}\left(R_{\Delta h}^T\mathcal{B} + R_{\Delta_h}\mathcal{C}\right) \tag{33}$$

where

$$\mathcal{B} \triangleq \mathcal{M}^H\left[\sum_{i=1}^{n}\frac{\sigma_i}{\lambda_i}\left(\bar{v}_iv_i^T\otimes U_n^c\mathcal{S}_i^{-1}U_n^{cH}\right)\right]\mathcal{M} \tag{34}$$

$$\mathcal{C} \triangleq \mathcal{M}^H \mathcal{K}^H \left[\sum_{i=1}^{n} \frac{1}{\lambda_i} \left(\bar{u}_i u_i^T \otimes V_n^c \Sigma_{*n}^{c\,1/2\,H} \mathcal{S}_i^{-1} \Sigma_{*n}^{c\,1/2} V_n^{cH} \right) \right] \mathcal{K}\mathcal{M} \tag{35}$$

and $\Sigma_{*n}^{c1/2}$ is the $(N+\nu-n) \times (N-n)$ matrix defined as

$$\Sigma_{*n}^{c1/2} \triangleq \begin{bmatrix} \sigma_{n+1}^{1/2} & & \\ & \ddots & \\ & & \sigma_N^{1/2} \\ & 0_{\nu \times (N-n)} & \end{bmatrix}$$

Proof: The proof is given in Appendix III. □

Of course, for $n = N$, expression (33) coincides with (27), because $\Sigma_{*n}^{c1/2}$ and, thus, $\mathcal{C}$ become zero. The magnification of the channel estimation errors is determined by matrices $\mathcal{B}$ and $\mathcal{C}$. In the next subsection, we obtain related bounds.

D. Case $n \leq N$: Bound

For $R_{\Delta h} = \sigma_{\Delta h}^2 I_{\nu+1}$ we have

$$\mathcal{E}\{\Delta I\} = \frac{\log_2 e \sigma_{\Delta h}^2}{N+\nu} (\mathrm{Tr}(\mathcal{B}) + \mathrm{Tr}(\mathcal{C})).$$

From (28) and (30), we obtain

$$\mathrm{Tr}(\mathcal{B}) \leq \left\| \sum_{i=1}^{n} \frac{\sigma_i}{\lambda_i} \left(\bar{v}_i v_i^T \otimes U_n^c \mathcal{S}_i^{-1} U_n^{cH} \right) \right\| N(\nu+1).$$

It can be seen that, due to the orthogonality of v_i, the sum inside the norm is of the form

$$\sum_{i=1}^{n} \mu_i M_i$$

with $\mu_i > 0$, M_i positive semidefinite and $M_i^H M_j = 0$, for $j \neq i$. Thus, its 2-norm equals

$$\begin{aligned} \max_i (\mu_i \|M_i\|) &= \max_i \frac{\sigma_i}{\lambda_i} \left\| \mathcal{S}_i^{-1} \right\| = \max_i \frac{\sigma_i}{\lambda_i(\sigma_i - \sigma_{n+1})} \\ &= \max_i \frac{\sigma_i}{\left(\sigma_i + \frac{\sigma_w^2}{\delta} \right) (\sigma_i - \sigma_{n+1})}. \end{aligned}$$

By computing the derivative, we obtain that the function

$$f(\sigma) \triangleq \frac{\sigma}{(\sigma + \frac{\sigma_w^2}{\delta})(\sigma - \sigma_{n+1})}$$

increases for decreasing σ, for $\sigma \geq \sigma_{n+1}$, and thus the above maximum is

$$\frac{\sigma_n}{\lambda_n(\sigma_n - \sigma_{n+1})}.$$

Using similar arguments, it can be shown that

$$\mathrm{Tr}(\mathcal{C}) \leq \frac{N(\nu+1)}{\lambda_n} \left\| \Sigma_{*n}^{c1/2H} \mathcal{S}_n^{-1} \Sigma_{*n}^{c1/2} \right\| = \frac{N(\nu+1)\sigma_{n+1}}{\lambda_n(\sigma_n - \sigma_{n+1})}.$$

Thus, a bound for the mean mutual information degradation is as follows:

$$\mathcal{E}\{\Delta I\} \leq \frac{2 \log_2 e N(\nu+1) \sigma_n \sigma_{\Delta h}^2}{(N+\nu)\lambda_n(\sigma_n - \sigma_{n+1})}.$$

If we assume, as in Subsection V-B, that $\sigma_{\Delta h}^2 = \frac{\sigma_w^2}{P_x(N_{\mathrm{tr}} - \nu)}$, with $P_x = \frac{n\delta}{N+\nu}$, we obtain

$$\mathcal{E}\{\Delta I\} \leq \frac{2 \log_2 e(\nu+1)N}{n(N_{\mathrm{tr}} - \nu)} \frac{1}{1 - \frac{\sigma_{n+1}}{\sigma_n}} \frac{1}{1 + \frac{\delta\sigma_n}{\sigma_w^2}}. \tag{36}$$

This bound is small, implying that the degradation will be small, if $\sigma_{n+1} \ll \sigma_n$ and $\delta\sigma_n \gg \sigma_w^2$. It is large, implying that the degradation may be significant if $\sigma_n \approx \sigma_{n+1}$ and/or $\delta\sigma_n \approx \sigma_w^2$.

VI. Simulations

In this section, we use numerical simulations to illustrate our theoretical results. We assume that the power per input data sample is $P_x = 1$ and the total input power is $E = P_x(N+\nu)$, with block length $N = 50$. The power of the additive white Gaussian noise is σ_w^2 and the SNR is defined as $\mathrm{SNR} \triangleq \frac{P_x}{\sigma_w^2}$. The true channel is generated randomly and then normalized to unit 2-norm. It has order $\nu = 5$ and the realization we consider is

$$h = [0.3871 - 0.6087 - 0.1006 - 0.4675\; 0.1611 - 0.4713]^T.$$

In Fig. 1, we plot the mutual information $I(X;Y)$ that results from water-filling (implemented as "power loading algorithm #1" of [2]) and the suboptimal approach for $n = 50$ and $n = 30$, by assuming perfect channel knowledge. We observe the following.

1) For SNR higher than 4 dB, the water-filling and the suboptimal scheme for $n = N$ practically coincide, supporting the observations of [1].
2) For SNR higher than 5 dB, the suboptimal scheme for $n = N$ is superior than the one with $n = 30$, because it exploits more efficiently the degrees of freedom of the block channel. The difference increases for increasing the SNR. For very low SNR, the suboptimal scheme with $n = 30$ is slightly better than the one with $n = N$.

In order to check the accuracy of our approximation for $n = N$, we assume that the channel is estimated using $N_{\mathrm{tr}} = 12$ data samples of an ideal training sequence, giving that $R_{\Delta h} = \frac{\sigma_w^2}{P_x(N_{\mathrm{tr}} - \nu)} I_{\nu+1}$ [11, p. 786]. In Fig. 2, we plot the experimentally computed (over 10^4 independent noise realizations) mean mutual information degradation, the corresponding second-order approximation (27), and bound (32) for varying the SNR. We observe that for SNR higher than 15 dB, the experimentally computed degradation and the approximation practically coincide, showing the usefulness of our results. We note that for SNR = 15 dB, the noise variance is $\sigma_w^2 = 0.0316$, giving $R_{\Delta h} = 0.0045 I_{\nu+1}$. This yields

$$\mathcal{E}(\|\Delta h\|^2) = 0.0045(\nu+1) = 0.0270$$

(recall that $\|h\|^2 = 1$). Thus, we observe that the second-order approximation becomes accurate for

$$10 \log_{10} \left(\frac{\|h\|^2}{\mathcal{E}(\|\Delta h\|^2)} \right) \approx 15.7 \text{ dB}.$$

Furthermore, we observe that for SNR higher than 15 dB, the mean mutual information degradation is of the order of 10^{-2} (and smaller), implying that, for sufficiently high SNR, the degradation becomes negligible.

The bound (32) follows the general changes of the mean mutual information degradation but it is not tight, in general.

Analogous results are obtained in the cases where $n < N$.

VII. Conclusion

Water-filling is the optimal precoding scheme for Gaussian parallel or block-based channels. A suboptimal scheme, which has been observed to perform quite close to water-filling, is when all "active" eigenvectors of the input data covariance matrix receive the same power. Both techniques require perfect channel knowledge at the

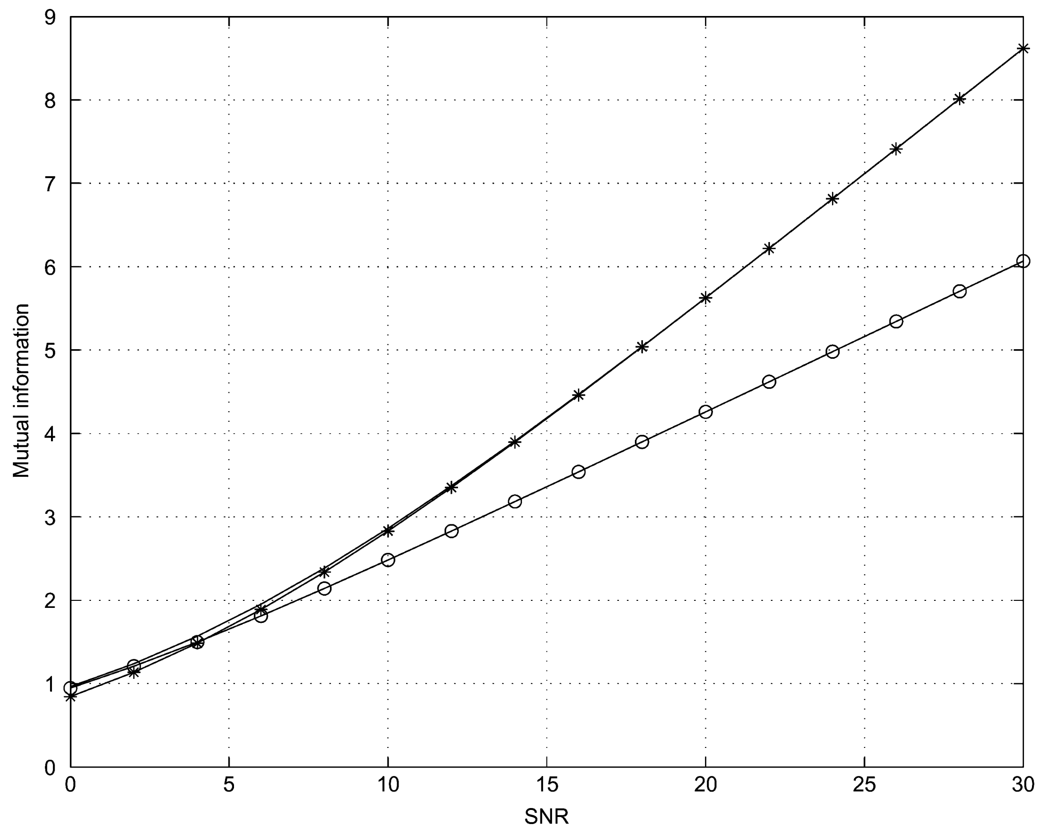

Fig. 1. Mutual information per sample versus SNR: water-filling (solid line), suboptimal scheme $n = N$ ("-*-"), $n = 30$ ("-o-").

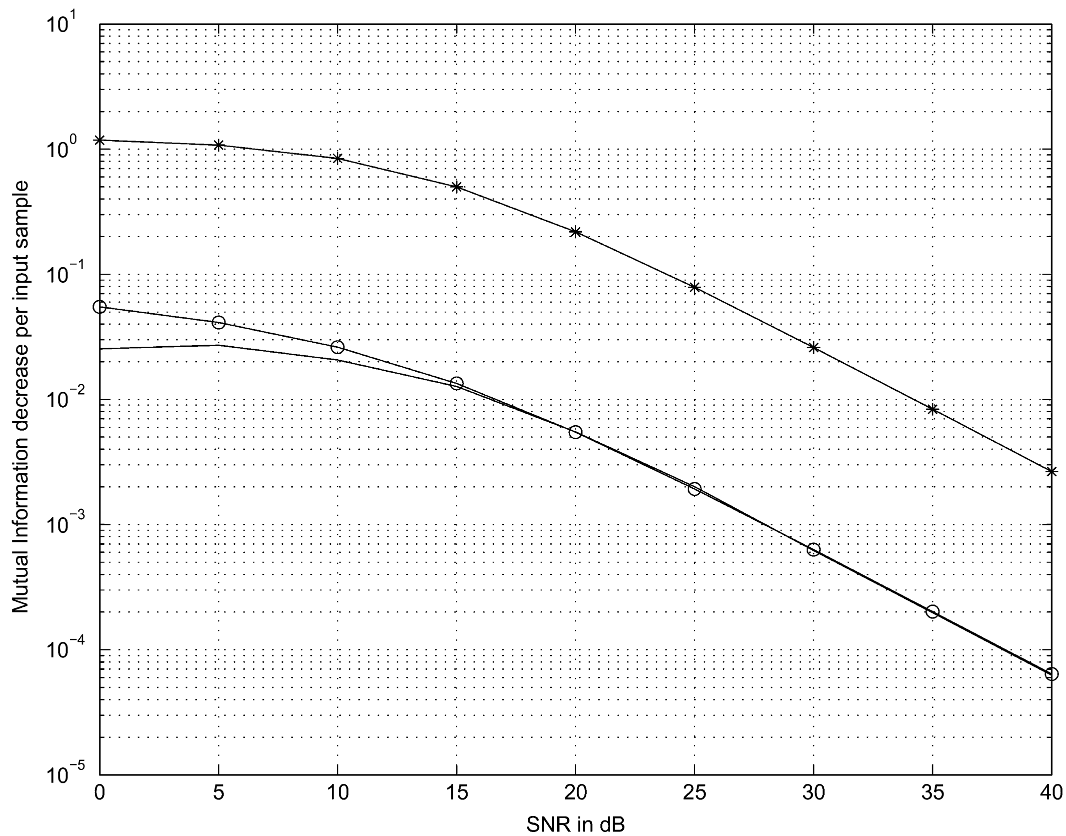

Fig. 2. Experimental mutual information decrease (solid line), second-order approximation ("-o-"), and bound (32) ("-*-") versus SNR.

transmitter. In this work, we considered the sensitivity of the suboptimal scheme when a channel estimate is used at the transmitter as if it were the true channel. Using tools from matrix perturbation theory, we derived closed-form expressions and bounds relating the channel estimation error covariance matrix with the mean mutual information decrease. We observed that for sufficiently good channel estimation the mutual information degradation is insignificant. Simulations are in agreement with our theoretical results.

Appendix I

Proof of Lemma 1: We start with an approximation to $(I_n + P^H P)^{-1/2}$. Let the singular value decomposition (SVD) of P be $P = U' \Sigma' V'^H$ (Notice that $\|\Sigma'\| = \|P\| = O(\|\Delta h\|)$.) Then

$$I_n + P^H P = I_n + V' \Sigma'^H \Sigma' V'^H = V'(I_n + \Sigma'^H \Sigma') V'^H.$$

Using the Taylor expansion

$$(1+x^2)^{-1/2} = 1 - \frac{1}{2}x^2 + O(x^4)$$

we obtain

$$\begin{aligned}(I_n + P^H P)^{-1/2} &= V'(I_n + \Sigma'^H \Sigma')^{-1/2} V'^H \\ &= V' \left(I_n - \frac{1}{2}\Sigma'^H \Sigma' + O(\|\Delta h\|^4)\right) V'^H \\ &= I_n - \frac{1}{2} P^H P + O(\|\Delta h\|^4). \qquad (37)\end{aligned}$$

Using (17), (37), and the fact that $P = O(\|\Delta h\|)$, we obtain

$$\hat{U}_n = U_n - \frac{1}{2} U_n P^H P + U_n^c P + O(\|\Delta h\|^3)$$

leading to

$$\Delta U_n = -\frac{1}{2} U_n P^H P + U_n^c P + O(\|\Delta h\|^3).$$

The error terms appearing in $\Delta \mathcal{A}$ become (see (13))

$$\Delta U_n^H U_n = -\frac{1}{2} P^H P + O(\|\Delta h\|^3)$$

and

$$\begin{aligned}\Delta U_n^H U_N \Sigma_N U_N^H \Delta U_n &= P^H U_n^{cH} U_N \Sigma_N U_N^H U_n^c P + O(\|\Delta h\|^3) \\ &= P^H \Sigma_n^c P + O(\|\Delta h\|^3)\end{aligned}$$

leading to

$$\Delta \mathcal{A} = -\frac{1}{2}\Sigma_n P^H P - \frac{1}{2} P^H P \Sigma_n + P^H \Sigma_n^c P + O(\|\Delta h\|^3).$$

This expression is difficult to compute because P is the solution of the nonlinear matrix (19). Using the facts that

$$P = O(\|\Delta h\|) \quad \text{and} \quad \tilde{P} = P + O(\|\Delta h\|^2)$$

we obtain

$$\begin{aligned}\tilde{P}^H \tilde{P} &= P^H P + O(\|\Delta h\|^3) \\ \tilde{P}^H \Sigma_n^c \tilde{P} &= P^H \Sigma_n^c P + O(\|\Delta h\|^3)\end{aligned}$$

and thus, we may approximate $\Delta \mathcal{A}$ as

$$\Delta \mathcal{A} = -\frac{1}{2}\Sigma_n \tilde{P}^H \tilde{P} - \frac{1}{2} \tilde{P}^H \tilde{P} \Sigma_n + \tilde{P}^H \Sigma_n^c \tilde{P} + O(\|\Delta h\|^3)$$

to prove the Lemma. □

Appendix II

Proof of Proposition 2: In the case $n = N, \Sigma_n^c = 0_{\nu\times\nu}$, giving (see (24))

$$\mathcal{S}_i = \sigma_i I_\nu.$$

Furthermore, from (20), we obtain

$$\tilde{P} = U_N^{cH} \Delta A U_N \Sigma_N^{-1}$$

which, using (10) and (4), gives

$$\tilde{P} = U_N^{cH} \Delta H^H V_N \Sigma_N^{-1/2}.$$

Thus,

$$\tilde{P} e_i = \sigma_i^{-1/2} U_N^{cH} \Delta H^H v_i$$

giving

$$\operatorname{vec}(\tilde{P} e_i) = \sigma_i^{-1/2} \underbrace{\left(v_i^T \otimes U_N^{cH}\right)}_{\alpha_i} \operatorname{vec}(\Delta H^H). \qquad (38)$$

It can be easily verified that

$$\operatorname{vec}(\Delta H^H) = \mathcal{M} \bar{\Delta h}. \qquad (39)$$

Thus,

$$\begin{aligned}\operatorname{Tr}\left(e_i^H \tilde{P}^H \mathcal{S}_i \tilde{P} e_i\right) &= \sigma_i \operatorname{Tr}\left(\operatorname{vec}(\tilde{P} e_i) \operatorname{vec}(\tilde{P} e_i)^H\right) \\ &= \operatorname{Tr}\left(\alpha_i \mathcal{M} \bar{\Delta h} \Delta h^T \mathcal{M}^H \alpha_i^H\right).\end{aligned}$$

Using (26), we obtain

$$\begin{aligned}\mathcal{E}\{\Delta I\} &= \frac{\log_2 e}{N+\nu} \sum_{i=1}^{N} \frac{1}{\lambda_i} \operatorname{Tr}\left(\alpha_i \mathcal{M}\, \mathcal{E}\left\{\bar{\Delta h} \Delta h^T\right\} \mathcal{M}^H \alpha_i^H\right) \\ &= \frac{\log_2 e}{N+\nu} \sum_{i=1}^{N} \frac{1}{\lambda_i} \operatorname{Tr}\left(R_{\Delta h}^T \mathcal{M}^H \alpha_i^H \alpha_i \mathcal{M}\right) \\ &= \frac{\log_2 e}{N+\nu} \operatorname{Tr}\left(R_{\Delta h}^T \mathcal{M}^H \left(\sum_{i=1}^{N} \frac{1}{\lambda_i} \alpha_i^H \alpha_i\right) \mathcal{M}\right).\end{aligned}$$

Using the relation (see (38))

$$\alpha_i^H \alpha_i = \bar{v}_i v_i^T \otimes U_N^c U_N^{cH}$$

and the fact that $\mathcal{A}$ is the diagonal matrix with elements λ_i (see (7)), we obtain

$$\begin{aligned}\sum_{i=1}^{N} \frac{1}{\lambda_i} \alpha_i^H \alpha_i &= \left(\sum_{i=1}^{N} \frac{1}{\lambda_i} \bar{v}_i v_i^T\right) \otimes U_N^c U_N^{cH} \\ &= \bar{V}_N \mathcal{A}^{-1} V_N^T \otimes U_N^c U_N^{cH}\end{aligned}$$

which proves the proposition. □

Appendix III

Proof of Proposition 3: We recall that in the general case (see (15) and (20))

$$\mathcal{S}_i = \sigma_i I_{N+\nu-n} - \Sigma_n^c$$

and

$$\Sigma_n^c \tilde{P} - \tilde{P} \Sigma_n = -U_n^{cH} \Delta A U_n.$$

Postmultiplying the above expression by e_i and using (4) and (10), we obtain

$$\mathcal{S}_i^{1/2} \tilde{P} e_i = \mathcal{S}_i^{-1/2} U_n^{cH} (\Delta H^H H + H^H \Delta H) u_i.$$

The terms of the right-hand side can be expressed as

$$\begin{aligned}\mathcal{S}_i^{-1/2} U_n^{cH} \Delta H^H H u_i &= \sigma_i^{1/2} \mathcal{S}_i^{-1/2} U_n^{cH} \Delta H^H v_i \\ \mathcal{S}_i^{-1/2} U_n^{cH} H^H \Delta H u_i &= \mathcal{S}_i^{-1/2} \Sigma_{*n}^{c\,1/2} V_n^{cH} \Delta H u_i.\end{aligned}$$

Furthermore

$$\mathrm{vec}(\Delta H) = \mathcal{K}\mathrm{vec}(\Delta H^T) = \mathcal{K}\mathcal{M}\Delta h.$$

Thus,

$$\mathrm{vec}(\mathcal{S}_i^{1/2}\tilde{P}e_i) = \sigma_i^{1/2}\left(v_i^T \otimes \mathcal{S}_i^{-1/2}U_n^{cH}\right)\mathcal{M}\bar{\Delta}h + \left(u_i^T \otimes \mathcal{S}_i^{-1/2}\Sigma_{*n}^{c\,1/2}V_n^{cH}\right)\mathcal{K}\mathcal{M}\Delta h.$$

From (26), we obtain

$$\mathcal{E}\{\Delta I\} = \frac{\log_2 e}{N+\nu}\sum_{i=1}^{n}\frac{1}{\lambda_i}\mathrm{Tr}\left(\mathrm{vec}(\mathcal{S}^{1/2}\tilde{P}e_i)\mathrm{vec}(\mathcal{S}^{1/2}\tilde{P}e_i)^H\right).$$

Working as in Appendix II and using the relations

$$\mathcal{E}\{\bar{\Delta}h\Delta h^H\} = \mathcal{E}\{\Delta h \Delta h^T\} = 0_{(\nu+1)\times(\nu+1)}$$

we obtain (33), to prove the proposition.

ACKNOWLEDGMENT

The author would like to thank the anonymous reviewers for their useful comments.

REFERENCES

[1] N. Al-Dhahir and J. M. Cioffi, "Block transmission over dispersive channels: Transmit filter optimization and realization, and MMSE-DFE receiver performance," *IEEE Trans. Inf. Theory*, vol. 42, no. 1, pp. 137–160, Jan. 1996.

[2] A. Scaglione, S. Barbarossa, and G. B. Giannakis, "Filterbank transceivers optimizing information rate in block transmissions over dispersive channels," *IEEE Trans. Inf. Theory*, vol. 45, no. 3, pp. 1019–1032, Apr. 1999.

[3] İ.E. Telatar, "Capacity of multi-antenna Gaussian channels," *Europ. Trans. Commun.*, vol. 10, no. 6, pp. 585–595, 1999.

[4] A. Lapidoth and P. Narayan, "Reliable communication under channel uncertainty," *IEEE Trans. Inf. Theory*, vol. 44, no. 6, pp. 2148–2177, Oct. 1998.

[5] M. Médard, "The effect upon channel capacity in wireless communications of perfect and imperfect knowledge of the channel," *IEEE Trans. Inf. Theory*, vol. 46, no. 3, pp. 933–946, May 2000.

[6] A. Lapidoth and S. Shamai (Shitz), "Fading channels: How perfect need "perfect side informatin" be?," *IEEE Trans. Inf. Theory*, vol. 48, no. 5, pp. 1118–1134, May 2002.

[7] S. Adireddy, L. Tong, and H. Viswanathan, "Optimal placement of training for frequency-selective block-fading channels," *IEEE Trans. Inf. Theory*, vol. 48, no. 8, pp. 2338–2353, Aug. 2002.

[8] H. Vikalo, B. Hassibi, B. Hochwald, and T. Kailath, "On the capacity of frequency-selective channels in training-based transmission schemes," *IEEE Trans. Signal Process.*, vol. 52, no. 9, pp. 2572–2583, Sep. 2004.

[9] G. Stewart and J. G. Sun, *Matrix Perturbation Theory*. New York: Academic, 1990.

[10] R. J. Vaccaro, "Weighted Subspace Fitting Using Subspace Perturbation Expansions," Tech. Rep. ESAT-SISTA/TR 1997-45.

[11] H. Meyr, M. Moeneclaey, and S. A. Fechtel, *Digital Communication Receivers*. New York: Wiley, 1998.

[12] H. Luetkepohl, *The Handbook of Matrices*. New York: Wiley,, 1996.

Addendum to "On Universal Simulation of Information Sources Using Training Data"

Neri Merhav, *Fellow, IEEE,* and
Marcelo J. Weinberger, *Senior Member, IEEE*

***Abstract*—In a recent paper [1], we studied the problem of universal simulation of an unknown information source of a certain parametric family, given a training sequence from that source and given a limited budget of purely random bits. The goal was to generate another random sequence (of the same length or shorter), whose probability law is identical to that of the given training sequence, but with minimum statistical dependency (minimum mutual information) between the input training sequence and the output sequence. In this addendum, we point out a concrete optimal simulation scheme that is easy to implement, as opposed to the nonconstructive existence result in that paper, and we make a number of additional observations on the universal simulation problem.**

***Index Terms*—Enumeration, mutual information, random process simulation, random number generators, typical sequences.**

A recent paper [1] studied the following universal simulation problem: An unknown source P, which is assumed to belong to a certain parametric family $\mathcal{P}$ (like the family of finite-alphabet memoryless sources, Markov sources, finite-state sources, parametric subsets of these families, etc.), is to be simulated. We are given a training sequence $X^m = (X_1, \ldots, X_m)$ that has emerged from this unknown source, as well as a string of k purely random bits $U^k = (U_1, \ldots, U_k)$, that are independent of X^m, and our goal is to generate an output sequence $Y^n = (Y_1, \ldots, Y_n)$, $n \le m$, corresponding to the simulated process, that satisfies the following three conditions.

C1. The mechanism by which Y^n is generated can be represented by a deterministic function $Y^n = \phi(X^m, U^k)$, where ϕ does not depend on the unknown source P.

C2. The probability distribution of Y^n is *exactly* the n-dimensional marginal of the probability law P corresponding to X^m for all $P \in \mathcal{P}$.

C3. The mutual information $I(X^m; Y^n)$ is as small as possible, simultaneously for all $P \in \mathcal{P}$.

In [1, Sec. IV-B], we referred to the case where $n < m$ and the key rate, $R \stackrel{\triangle}{=} k/n$, is finite. Unlike the other cases, for which we were able to demonstrate concrete simulation schemes that satisfy all three conditions, C1–C3, in this case, we only presented a nonconstructive existence result in a very large ensemble of schemes [1, Theorem 3].

The primary purpose of this addendum is to suggest a simple simulation scheme that satisfies the above conditions in the case where $n < m$ as well. In the sequel, lower case notation such as x^m, y^n, and u^k, will denote specific realizations of the random vectors X^m, Y^n, and U^k, respectively. For a given training sequence x^m, let

$$\phi(x^m, u^k) = \left[J_m^{-1}(J_m(x^m) \oplus \lceil f(u^k) \cdot |T_{x^m}|/2^k \rceil\right]_1^n \quad (1)$$

Manuscript received September 8, 2004; revised May 17, 2005. This work was done while N. Merhav was visiting Hewlett-Packard Laboratories, Palo Alto, CA, USA.

N. Merhav is with the Electrical Engineering Department, Technion–Israel Institute of Technology, Technion City, Haifa 32000, Israel (e-mail: merhav@ee.technion.ac.il).

M. J. Weinberger is with Hewlett-Packard Laboratories, Palo Alto, CA 94304 USA (e-mail: marcelo@hpl.hp.com).

Communicated by A. Lapidoth, Associate Editor for Shannon Theory.
Digital Object Identifier 10.1109/TIT.2005.853324