Inflated Lattice Precoding, Bias Compensation, and the Uplink/Downlink Duality: The Connection

Robert F.H. Fischer, Christian Siegl

Abstract—We review the uplink/downlink duality between decision-feedback equalization (DFE) and (nonlinear) precoding at the transmitter side. Thereby the effect of bias compensation in DFE is addressed, and its connection to the concept of “inflated lattice precoding” (ILP)—a tool which has been proven to be the key to achieve the capacity of channels with interference known at the transmitter—is elucidated. Concentrating on signal-to-interference plus noise ratios and hence error rates, pairs of schemes dual to each other are identified. It is shown that bias compensation in DFE and the scaling in ILP are dual operations and cancel each other.

I. INTRODUCTION

Over the last years, the theory of uplink/downlink duality has been developed, e.g., [13], [11]. Briefly, it states that the performance of the users in an uplink scenario (multipoint-to-point transmission) can be made exactly the same as in the corresponding downlink scenario (point-to-multipoint transmission) when the same sum power constraint applies. Based on this, transmission schemes in some sense dual to each other, for both scenarios can be identified.

In this letter, we review the uplink/downlink duality between decision-feedback equalization (DFE) and precoding, in particular the theoretically optimal concept of “Costa precoding” [3] and Tomlinson-Harashima precoding (THP) [6]. We concentrate on the effect of bias compensation in DFE and elucidate its connection to the concept of “inflated lattice precoding (ILP)” [4], [5]. Based on the results in [13] it is shown that bias compensation and the scaling in ILP are dual operations and cancel each other. Thereby, we concentrate on signal-to-interference plus noise ratios (SINR) and hence error rates. Pairs of schemes, dual to each other are identified, and the effect of appropriate scaling of known interference is generalized over the exposition in [8], [4], [9] to uplink/downlink scenarios.

In Section II, uplink transmission and DFE are reviewed. Precoding and its duality to DFE are addressed in Section III. Section IV shown some numerical results and gives a brief discussion.

II. UPLINK TRANSMISSION AND DECISION-FEEDBACK EQUALIZATION

First we consider transmission from, say K, non-cooperating users, distributed over some service area, with a central receiver, equipped with Nc receive antennas. The channel matrix $H_u$ (assuming flat fading and given in the equivalent complex baseband) is of dimension $N_c \times K$.

The authors are with Lehrstuhl für Informationsübertragung, Friedrich-Alexander-Universität Erlangen–Nürnberg, Cauerstrasse 7/LIT, 91058 Erlangen, Germany, Email: {fischer, siegl}@lnt.de

The data symbols $a_{u,k}$ (combined into the vector $a_u = [a_{u,1}, \ldots, a_{u,K}]^T$) are independently taken from QAM constellations with variance normalized to $\sigma_a^2$. Distribution of the sum transmit power is accomplished by the diagonal scaling matrix $A_u = \text{diag}(\lambda_{u,1}, \ldots, \lambda_{u,K})$ with real positive elements and $\sum_k \lambda_{u,k} = K$, and channel symbols $x_u = A_u a_u$ are obtained. At the receiver side, (spatially) white Gaussian noise $n_u$ with variance $\sigma_n^2$ per component is present.

In this typical uplink transmission scenario, joint processing is only possible at the receiver side. For separating the users, decision-feedback equalization, Fig. 1 (left), is a very popular strategy, because of its simplicity and optimality, as it is an implementation of the chain rule of information theory. In wireless applications, DFE is often called successive interference cancellation [12]; in case of MIMO transmission known as generalized DFE (GDFE) [2], and it is the basic principle behind the BLAST schemes [10].

Noteworthy, in (G)DFE decision order significantly influences performance. Here, we always assume a given ordering which is already included in the channel matrix $H_u$ by reordering its column accordingly. Decision is then done in the obvious order from user 1 up to user K.

Optimizing the DFE according to the minimum mean-squared error (MMSE) criterion, the required matrices are obtained by calculating a Cholesky factorization [6]

$$A_u H_u^H H_u A_u + \frac{\sigma_n^2}{\sigma_a^2} I = B_u^H \Sigma_u B_u$$

with real diagonal matrix $\Sigma_u = \text{diag}(\varsigma_{u,1}, \ldots, \varsigma_{u,K})$. The lower triangular matrix with unit main diagonal $B_u$ immediately constitutes the feedback matrix, whereas the feedforward matrix calculates to $F_u = \Sigma_u^{-1} B_u^H A_u H_u^H$. Ignoring error propagation in the DFE, the K users expire a SINR which calculates to

$$\text{SINR}_{u,k}^\text{II} = \frac{\sigma_a^2}{\sigma_n^2} \varsigma_{u,k} = \frac{1}{1 - K_u,k}, \quad k = 1, \ldots, K.$$  

The last equation is obtained after straightforward calculations using the definition of the end-to-end cascade $K_u = [K_u,k,l] = F_u H_u A_u$, and is a general principle in MMSE estimation, cf. [8], Footnote 5, Figures 4 (b), (c).

A close look at $K_u$ reveals that the main diagonal elements $K_u,k,k$ are smaller than one, i.e., a bias is present as some portion of the useful signal is falsely accounted to the noise, cf. [1]. Compensating for this bias by scaling with the real diagonal matrix $C_u = [c_{u,k},l] = \text{diag}(K_u^{-1}_{1,1}, \ldots, K_u^{-1}_{K,K})$ at

$^1$Notation: $A^T$ is the transpose of matrix $A$, $A^H$ the Hermitian (i.e., conjugate) transpose and $A^{-H}$ the inverse of the Hermitian transpose (for square matrix $A$), $I$: identity matrix, diag($\cdot$): diagonal matrix with given entries.

1 Notation: $A^T$ is the transpose of matrix $A$, $A^H$ the Hermitian (i.e., conjugate) transpose and $A^{-H}$ the inverse of the Hermitian transpose (for square matrix $A$), $I$: identity matrix, diag($\cdot$): diagonal matrix with given entries.
the input of the decision device lead to the unbiased MMSE DFE with SINR reduced by one, i.e.,

\[ \text{SINR}_{\text{UnB}}^{\text{UnB}} = \text{SINR}_{\text{UnB}}^{\text{B}} - 1 = \frac{K_{u,k,k}}{1 - K_{u,k,k}}. \]

(3)

However, this compensation of the inherent bias in the MMSE solution leads to an increased performance [1], as no longer a systematic error is present.

III. DOWNLINK TRANSMISSION AND PRECODING

An important result in information theory of the last years is the uplink/downlink duality, e.g., [13], [11]. Considering communication of a central transmitter with Nc transmit antennas with K non-cooperating users over the hermitian channel of the uplink scenario, i.e., \( H_d = H_u^H \), and independent Gaussian noise with variance \( \sigma^2_n \) at each receiver, the same (unbiased) SINRs of the users as in the uplink, i.e., \( \text{SINR}_{d,k}^{\text{UnB}} = \text{SINR}_{u,k}^{\text{UnB}}, \forall k \), can be achieved with the same sum transmit power.

The corresponding transmission scheme is obtained by “dualizing” the DFE structure, cf. Fig. 2: At the transmitter, precoding is performed using a feedback loop with feedback matrix \( B_d = B_u^H \) and reversed processing order [13], followed by a feedforward matrix \( F_d = F_u^H \). In theory, optimal precoding, often denoted as "Costa precoding" (named after the fundamental result in [3]), has to be used. Here, we consider its most simple realization, namely THP [6]. In THP, an one-dimensional modulo reduction is used, which—to complete the duality—is followed by scaling with \( C_d = C_u \). Going to higher-dimensional quantizations/modulo reductions the shaping loss (increased transmit power compared to Gaussian signaling) vanishes, e.g., [7].

In order to achieve the same SINR point, sum power has to be distributed among the users suitably. Following [13], the real diagonal scaling matrix \( \Lambda_d \) is obtained as

\[ \Lambda_d^2 = C_u^{-2} \frac{\sigma^2_n}{\sigma^2_u} \text{diag} \left( \Theta^{-1} \right), \]

(4)

where \( \Theta \) is a upper triangular matrix with entries \(-|K_{u,k,l}|^2, 1 \leq k < l \leq K\), and \( K_{u,k,l} = |K_{u,k,l}|^2, k = 1, \ldots, K, \) and 1 is the all-one vector.

At the receiver side, after compensating for the scaling (\( g_k = \lambda_k^{-1} \)), a threshold device, taking the modulo congruence of the signal point into account [6], generates estimates of the data symbols \( d_{u,k} \).

In [4], [5] it has been shown that precoding can be improved (significantly at low SNRs), if the so-called concept of “inflated lattice precoding” is applied. Here, basically an MMSE scaling of the interference term, which is presubtracted via the feedback loop, is performed (matrix \( A = \text{diag}(\alpha_1, \ldots, \alpha_K) \), shown in gray in Fig. 2). The same scaling also applies at the input of the decision device. Given \( \text{SINR}_{d,k}^{\text{UnB}} \), the scaling factor for the \( k \)th user is

\[ \alpha_k = \frac{1}{1 + 1/\text{SINR}_{d,k}^{\text{UnB}}} = K_{u,k,k}. \]

(5)

Since \( \text{SINR}_{d,k}^{\text{UnB}} = \text{SINR}_{u,k}^{\text{UnB}} \), and considering the right hand side of (3), the last equality in (5) is valid. However, since \( c_{u,k} = K_{u,k,k}^{-1} \), the scaling used in ILP is nothing else than the inverse of the bias compensation in DFE. Hence, the biased MMSE solution is immediately used in precoding, a fact already observed in [8] and in a related setting in [9].

Moving the matrix \( C_d \) in the precoder right in front of the modulo device, it cancels with the matrix \( A \) in the feedback loop, and the combined gain matrix \( \tilde{\Lambda}_d = A \Lambda_d \) remains in front of the loop. The modulo operation has then to be adjusted to the boundary region of the constellation scaled by \( \tilde{\lambda}_{d,k} \). At the receiver, scaling by \( g_k = \lambda_k^{-1} \) has to be performed.

The gain due to inflated lattice precoding over “unbiased precoding” emerges from the fact that the (possibly scaled) modulo output is independent of the data symbols. If applicable, this independency can be forced by utilizing a dither sequence, [8], [5]. As a result, an MMSE exchange between channel noise \( n_{d,k} \) and residual interference caused by \( x_{d,k} \) is possible without generating any bias, cf. [4], [8], [7].

In DFE this is not immediately possible. However, utilizing also a dither \( u_{d,k} \), uniformly distributed over the boundary region of the signal constellation, the same MMSE scaling is applicable and the biased MMSE DFE would outperform the unbiased version. The modifications are shown in Fig. 2. Yet, as in conventional THP, this is paid with an (slightly) increased average transmit power (precoding loss). The matrix calculation (1) should account for this increase by replacing \( \sigma^2_u \) by the actual transmit power.

\[ \text{Fig. 2.} \] Modifications for DFE with dither. The decision device takes the modulo congruence into account.

Using dither, the uplink/downlink duality is also perfect for biased MMSE DFE and inflated lattice precoding, i.e., “biased MMSE THP”. Figure 3 sketches the relations and correspondences.
IV. SIMULATION RESULTS AND CONCLUSIONS

In order to visualize the correspondences and differences of DFE/THP on the one hand and the effect of bias compensation/MMSE scaling on the other hand, numerical simulations were conducted.

The number of transmit antennas at the central unit is chosen to be $N_c = 4$ and $K = 4$ scattered, mobile users are assumed. The coefficients of the hence $4 \times 4$ channel matrix $H$ are selected randomly and unit-variance complex Gaussian. The results are averaged over a sufficiently large number of channel realizations and displayed over the ratio of average transmitted sum energy per infomation bit $E_{b,\text{TX}}$ and one-sided noise power spectral density $N_0$. In the sequel, we restrict to the obvious processing order 1 through $K$, and arbitrarily concentrate only on the third user. In DFE, perfect feedback is expected ("genie aided DFE"). 16QAM modulation with Gray labeling is assumed, and in the uplink power is distributed uniformly over the users, i.e., $A_u = I$.

SINR and (bit) error rates only have a defined one-to-one relation for fixed (and known) density of the sum interference. Since the central limit theorem does not hold for the $4 \times 4$ situation at hand, and we are interested only in the performance of one particular user, the transmit signals of all other users $(\epsilon_{u,k}$ and $\tilde{x}_d,k)$ are chosen Gaussian with the same variance as the respective transmission strategy. This guarantes that interference plus noise is Gaussian and SINR is related to error rates via the standard $Q$ function.

In Fig. 4 the bit error rates of DFE and THP are shown. On the left hand side, the curves for biased and unbiased MMSE DFE and for unbiased MMSE THP are plotted. For comparision, the THP curve is repeated but shifted (i) to the left by the precoding loss, i.e., the increase in average transmit power is neglected, and (ii) shifted down by a factor $4/3$, i.e., the increase in the average number of nearest neighbors is neglected. It is clearly visible that biased MMSE DFE performs slightly worse than the unbiased version, and, neglecting the losses of THP compared to the theoretical "Costa precoding", DFE and precoding perform exactly the same, as predicted by the uplink downlink duality. Noteworthy, this equivalence holds for each channel realization, not only for the average.

On the right hand side of Fig. 4, the curves for biased MMSE DFE employing dither, inflated lattice precoding (biased MMSE THP), and for reference, unbiased MMSE DFE, are given. As predicted, the biased DFE and THP versions with dither perform exactly the same. Shifting the curves, to compensate for the loss of THP as above (using dither DFE and THP have the same precoding loss and the same increase in the number of nearest neighbors), reveals that (using optimal "Costa precoding") a slightly better performance than unbiased MMSE DFE is possible.

To summarize, the numerical simulations clearly verify the correspondences and difference between the respective schemes. The main ingredient of inflated lattice precoding is—more evidently as so far in literature—identified as the dual to bias compensation in MMSE DFE. This point of view gives more complete picture of the relations concerning uplink/downlink duality.

REFERENCES