Robust Turbo Equalization Under Channel Uncertainties

Nargiz Kalantarova, Suleyman Serdar Kozat, *Senior Member, IEEE*, and Alper T. Erdogan, *Member, IEEE*

***Abstract*—We investigate robust turbo equalization over frequency selective channels in the presence of channel uncertainties. The turbo equalization framework investigated here contains a linear equalizer to combat ISI and a trellis based decoder. However, instead of completely tuning the linear equalizer parameters to the available inaccurate channel information, a minimax scheme and a competitive scheme are studied, which incorporate the uncertainty in channel information to equalizer design in order to improve robustness. Approximate implementations of these methods are also presented with reduced computational complexity. The performance improvement obtained by the proposed algorithms are demonstrated through simulations under different scenarios.**

***Index Terms*—Channel uncertainties, competitive, linear turbo equalization, minimax.**

I. INTRODUCTION

We consider robust turbo equalization over communication channels with intersymbol interference (ISI) in the presence of channel uncertainties. Turbo equalization takes advantage of the concatenated code structure of the data path that consists of an error-correcting code (ECC) implemented at the transmitter and the convolutional structure of the communication channel perceived as a rate-1 convolutional code [1]. Turbo equalization mimics the classical turbo decoding procedure for the turbo codes, however, one of the intentional ECCs of the classical turbo coding framework is replaced by the "unintentional" convolutional channel [2]. Since the parameters of this unintentional code are medium dependent, and therefore random, they are to be estimated by the receiver. Consequently these "code parameters" are prone to estimation errors. The inaccuracies in the channel parameters may be either due to imperfect channel estimation caused by limited training data, high energy noise, or due to time variations of the channel parameters outside the training period which may be attributed to time variations in the channel (i.e., the violation of the quasi stationary assumption) or timing recovery problems [3]. Our goal in this article is to introduce novel turbo equalization approaches to achieve robustness against such potential uncertainties in the estimated channel parameters. In particular, we show that through use of equalization algorithms based on competitive [4] and minimax frameworks [5], [6], we can obtain turbo equalization methods whose performance are less sensitive against the channel estimation errors and better in terms of bit-error-rate (BER) over the plug-in methods in certain scenarios. We emphasize that the robust estimation setup here is *specific* to the turbo equalization framework, since unlike [5]–[7], due to the soft information generated from the soft input soft output (SISO) decoder, we require to estimate non zero mean random variables by observing non zero mean random variables through an *unknown* channel. Hence, this setup requires a particular equalizer formulation which is further explained in Section II. Moreover, due to the convolutive channel, robust formulations require constrained "affine" mapping yielding different formulation and optimization compared to [5]–[7].

In this paper, we study linear turbo equalization centered around the mean square error (MSE) minimization framework, however, we refrain from completely tuning the linear equalizer parameters to the available inaccurate channel information. The methods we introduce are based on minimizing certain MSE criterion which incorporate the channel inaccuracies in the problem formulation. In the first approach, we apply a minimax framework where the linear equalizer coefficients are selected by minimizing the MSE with respect to the worst possible channel around the inaccurate channel coefficients [6], [8], [9]. We then extend this framework and define a relative performance measure between the MSE of a linear equalizer and the MMSE of the linear MMSE equalizer calculated with the correct knowledge of the underlying channel [4], [5], [7]. This relative performance measure describes our "regret" using a linear equalizer that is not the correct linear MMSE equalizer (which is not available). We then seek for a linear equalizer that minimizes this regret with respect to the worst possible channel around the inaccurate channel coefficients. We demonstrate that obtaining the linear equalizers for both approaches can be formulated as a semi-definite programming problem (SDP), which can be efficiently solved [10] even for real-time applications [11]. For certain applications, as in the linear MMSE equalization setup, applying these approaches at each time instant may be computationally prohibitive [2]. Hence, we also provide approximate implementations of both methods with lower computational complexity.

The framework we investigate in this paper, where the MAP equalizer is replaced by a linear equalizer is initially studied in [12], where an LMS adaptive algorithm is used to train the linear equalizer parameters. Different extensions of this idea are further elaborated in [2], [13]. We emphasize that the introduced

Manuscript received July 28, 2010; revised March 14, 2011; accepted July 08, 2011. Date of publication July 22, 2011; date of current version December 16, 2011. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Subhrakanti Dey. This work was supported in part by TUBITAK Career Award, Contract No: 104E073, No: 108E195 and Turkish Academy of Sciences GEBIP Program.

The authors are with the Electrical Engineering Department, Koc University, Istanbul (e-mail: nkalantarova@ku.edu.tr; skozat@ku.edu.tr; alper@stanfordalumni.org).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TSP.2011.2162505

methods can be used in conjunction with such adaptive algorithms since usually the channel or system parameters cannot be learned perfectly by the adaptive algorithms and the uncertainty in learning can be incorporated in the equalizer design as in this paper. Along the lines of [13], in [2], authors replaced the MAP equalizer with a linear equalizer or a DFE, where the parameters of these filters are trained using the MMSE criteria. However, in [2], the parameters of the system are trained assuming perfect knowledge of the channel impulse response, hence the results cannot be applied here. However, we compare our results with the linear MMSE equalizer tuned to the inaccurate channel information and demonstrate that the introduced algorithms provide better BERs in certain scenarios in our simulations.

The robust minimax approach to equalization problems under channel uncertainties is studied in [5], [8], and [9]. In [9], the uncertainty in the channel information is represented in spectral domain as bounds on the phase and amplitude function of the unknown channel. However, unlike here where we explicitly provide the linear equalizer coefficients that are robust in a minimax sense, no expressions for a linear equalizer satisfying the functional forms of the phase and amplitude response are given in [9]. Although in [5] and [8] the minimax and competitive methods are used to incorporate the uncertainty in the model into the equalization problem, the framework, the application as well as the cost function definitions are different in this paper. Due to the turbo equalization framework, the problem we study here involves estimating a nonzero mean random variable using non-zero mean random observations under channel uncertainties, unlike in [5], [14], [15]. In [5], the data is assumed to be zero mean. However, in this paper, the received random variables cannot be normalized to have zero mean since the unknown transmitted sequence is observed through an unknown channel. Furthermore, in [5], the competitive minimax regret estimation framework has uncertainty only in the covariance matrix of the transmitted data and the channel is assumed to be *perfectly* known. For the minimax estimation framework (not for the regret case) in [5], the model has uncertainty either in the channel and/or in the covariance matrix of the transmitted signal. However, even in this preliminary case, the equalizer is optimized without any constraints, i.e., the channel matrix or the matrix generated by the linear equalizer are not required to have convolutive forms and the equalizer structure is different due to the bias terms. However, we use a similar SDP approach with [5], [7] to solve the convex constraint convex optimization problems. Although the well-known H^∞ equalization approach shares a similar minimax optimization setup as here, there are important differences. In the H^∞ framework, the cost criteria is the maximum energy gain from input signals to the output estimation errors, where the ratio of the error signal energy to the energy of the disturbances for all possible signals (with nonzero energy) is minimized [16]–[19]. The uncertainty in the H^∞ approach is specifically in the signals, where all signals and disturbances are considered as deterministic. However, in this paper, we specifically consider the case that the uncertainty is in the linear mapping, i.e., in the channel, the signals are taken as stochastic signals and naturally the cost function is different. The robust turbo equalization based on the minimax formulation is recently investigated in [20]. Note that our setup is different than [20] due to the specific structure of the channel matrix and norm based uncertainty ball, hence, requires different SDP formulations.

The competitive approach as an alternative to the minimax framework studied here has extensive roots in computational learning theory, information theory and signal processing [4], [21], [22]. The competitive approach studied here with a similar cost function is introduced in [5] and studied in [7] for linear estimation. However, the competitive framework of [5] concentrates on data estimation where the uncertainty is in the statistics of the desired and noise signals. In this paper, because of the nature of the communication problem, the uncertainty is in the communication channel; the statistics of the desired signal and the noise are assumed to be known.

The main contributions of this paper are as follows:

1) We focus on robust channel estimation where the inaccuracies in channel estimation are incorporated in the problem formulation through a minimax framework. Specific to the turbo equalization framework, this setup, unlike [5], [6], needs an adaptive bias term and, unlike [5], needs convolutive structure, where obtaining equalization parameters are formulated as an SDP problem.
2) We then study a competitive approach where the cost function is defined with respect to the performance of the best linear equalizer in MSE sense (which is unavailable). As in the minimax case, unlike [7] and [5], this competitive setup has a bias term and a convolutive structure that needs different formulation specific to the turbo equalization framework. Obtaining equalization parameters that optimizes this competitive setup is formulated as an SDP problem.

The organization of this article is as follows. In Section II, the basic setup for turbo equalization is described, along with the notation. We illustrate the proposed equalization approaches in Section III. We first study the linear MMSE equalization tuned to the inaccurate channel filter in order to introduce reduced complexity versions of the proposed algorithms. We then investigate the minimax approach and then the competitive approach, and demonstrate that both problems can be cast as SDP problems. Simulation results to illustrate the performance of the proposed algorithms are presented in Section IV. Finally, we conclude our paper with certain remarks in Section V.

II. Turbo Equalization System Description

Throughout this paper, bold lowercase letters will denote vectors and bold uppercase letters will denote matrices. All vectors are column vectors and l^2-norm of a vector $\mathbf{v}$ is defined as $\|\mathbf{v}\| = \sqrt{\mathbf{v}^H\mathbf{v}}$, where $(\cdot)^T$, $(\cdot)^H$ and $(\cdot)^+$ represent transpose, conjugate transpose and conjugation, respectively. The time index is shown in the subscripts. The operator E[.] denotes the expectation operator. For notational simplicity, the expected value of a random variable x is denoted as $\bar{x} = \mathrm{E}[x]$, and the expected value of a random vector $\mathbf{x}$ is $\bar{\mathbf{x}} = \mathrm{E}[\mathbf{x}]$. The matrix $\mathbf{I}$ denotes the identity matrix of appropriate dimensions. The vector (or matrix) $\mathbf{0}$ represents a vector (or matrix) of zeros, where the dimensions are understood from the context. Here, $\mathcal{N}(\mu, \sigma^2)$ denotes the Gaussian distribution with mean μ and variance σ^2. The operator "$*$" is the convolution operator.

In this section, we provide the basic description of the communication system studied in this paper, illustrated in Fig. 1. Here, $\{a_t\}$, $t = 1, \ldots, n_a$, $a_t \in \{0, 1\}$, is the transmitted

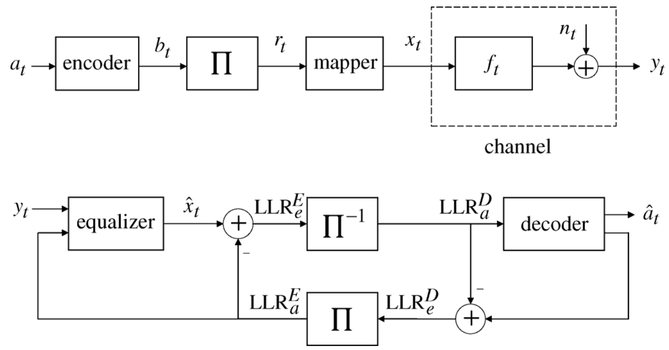


Fig. 1. A basic turbo equalization framework with the transmitter, the channel and the receiver. The receiver contains both the equalizer and the decoder parts.

signal. To incorporate redundancy in transmission, the input signal $\{a_t\}$ is encoded by a convolutional code to produce $\{b_t\}$, $t = 1, \ldots, n_b$. To further decrease the possible transmission errors, the encoded bits $\{b_t\}$ are interleaved using an S-random interleaver [23] to produce the interleaved and coded bits $\{r_t\}$. Finally, the interleaved bits are modulated to produce channel symbols $\{x_t\}$, e.g., if one uses BPSK signaling then $x_t = (-1)^{r_t+1}$. For notational simplicity, we assume BPSK signaling in the rest of the paper when we need to specify a particular modulation method. However, the formulations for the introduced equalization algorithms are given for complex modulated data. The modulated sequence, $\{x_t\}$, is transmitted through a baseband discrete-time channel with a finite-length impulse response $\{f_t\}$, $t = 0, 1, \ldots, M-1$, represented by $\mathbf{f} \triangleq [f_{M-1}, \ldots, f_0]^T$. Here, the transmitted signal $\{x_t\}$ is assumed to be uncorrelated due to the interleaver. The received signal y_t is given by

$$y_t \triangleq x_t * f_t + n_t = \left(\sum_{k=0}^{M-1} f_k x_{t-k}\right) + n_t$$

where $\{n_t\}$ is the additive complex white Gaussian noise with zero mean and circular symmetric variance σ_n^2.

Note that the underlying channel impulse response vector is not accurately known, however, an estimate of $\mathbf{f}$ is provided as $\tilde{\mathbf{f}}_t$ (which can be possibly time varying for certain adaptive methods [24]). The uncertainty in the channel impulse response vector is modeled by $\|\mathbf{f} - \tilde{\mathbf{f}}_t\| \le \delta, \delta \in \mathbb{R}^+, \delta < \infty$, where δ or a bound on δ is known. We emphasize that although the results we provide hold for time varying $\mathbf{f}_t$ and δ_t, we have dropped the time index from $\mathbf{f}$ and δ only for notational simplicity.

The received signal $\{y_t\}$ is then processed by a turbo equalization system comprised of an equalizer and a decoder as shown in Fig. 1. In this framework, the equalizer and decoder are considered as the inner decoder and outer decoder, respectively, and an iterative decoding scheme is used at the receiver of Fig. 1. The equalizer computes the a posteriori information using the received signal, transmitted signal estimate, channel convolution matrix (or an estimate of it) and *a priori* probability of the transmitted signals. After subtracting the *a priori* information LLR_a^E and de-interleaving the extrinsic information LLR_e^E, a SISO channel decoder computes the extrinsic information LLR_e^D on coded bits, which are fed back to the linear equalizer as *a priori* information LLR_a^E after interleaving. The *a priori* information from the decoder can be used to compute the mean and variance of the x_t as $\bar{x}_t \triangleq E[x_t : \{\mathrm{LLR}_{a,t}^E\}]$ and $q_t \triangleq E[x_t^2 : \{\mathrm{LLR}_{a,t}^E\}] - \bar{x}_t^2$, respectively, where $E[x : y]$ represents the expectation of x with respect to the distribution defined by y. As an example, for BPSK signaling, the mean and variance are given as $\bar{x}_t = \tanh\left(\frac{\mathrm{LLR}_{a,t}^E}{2}\right)$ and $q_t = 1 - |\bar{x}_t|^2$. Hence, the equalizer in turbo equalization system has access to second order statistics of $\{x_t\}$ in addition to $\{y_t\}$.

In this paper, a linear equalizer is used to combat the ISI. The estimate of the desired data x_t is constructed as

$$\hat{x}_t = \mathbf{c}_t^T \mathbf{y}_t + l_t + \bar{x}_t \tag{1}$$

where $\mathbf{c}_t = [c_{t,N_2}, \ldots, c_{t,-N_1}]^T$ is length $N = N_1 + N_2 + 1$ linear equalizer and $\mathbf{y}_t \triangleq [y_{t-N_2}, \ldots, y_{t+N_1}]^T$. We point that in (1), the equalizer is "affine", i.e., there is a bias term l_t since the received signal y_t is not zero mean and the mean sequence $\{\bar{y}_t\}$ is not known exactly due to uncertainty in the channel. Since the received data vector $\mathbf{y}_t$ is given by

$$\mathbf{y}_t = \mathbf{F}\mathbf{x}_t + \mathbf{n}_t$$

where $\mathbf{x}_t \triangleq [x_{t-M-N_2+1}, \ldots, x_{t+N_1}]^T$ and $\mathbf{F} \in \mathbb{C}^{N \times (N+M-1)}$

$$\mathbf{F} \triangleq \begin{bmatrix} f_{M-1} & f_{M-2} & \ldots & f_0 & 0 & \ldots & & 0 \\ 0 & f_{M-1} & f_{M-2} & \ldots & f_0 & 0 & \ldots & 0 \\ \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots \\ 0 & \ldots & & 0 & f_{M-1} & f_{M-2} & \ldots & f_0 \end{bmatrix}$$

is the convolution matrix corresponding to $\mathbf{f} = [f_{M-1}, \ldots, f_0]^T$, the estimate of x_t can be written as

$$\hat{x}_t = \mathbf{c}_t^T \mathbf{F}\mathbf{x}_t + \mathbf{c}_t^T \mathbf{n}_t + l_t + \bar{x}_t \tag{2}$$

or

$$\hat{x}_t = \mathbf{f}^T \mathbf{C}_t \mathbf{x}_t + \mathbf{c}_t^T \mathbf{n}_t + l_t + \bar{x}_t,$$

where $\mathbf{C}_t \in \mathbb{C}^{M \times (N+M-1)}$ is the convolution matrix corresponding to $\mathbf{c}_t$ and $\mathbf{c}_t^T \mathbf{F} = \mathbf{f}^T \mathbf{C}_t$.

As the linear equalizer, if one uses the linear MMSE equalizer, this yields

$$\begin{aligned} \mathbf{c}_t &= \left[\mathbf{v}^T \mathbf{Q}_t \mathbf{F}^H \left(\sigma_n^2 \mathbf{I} + \mathbf{F}\mathbf{Q}_t\mathbf{F}^H\right)^{-1}\right]^T \\ &= \left[\sigma_n^{-2} \mathbf{v}^T \left(\mathbf{Q}_t^{-1} + \sigma_n^{-2}\mathbf{F}^H\mathbf{F}\right)^{-1} \mathbf{F}^H\right]^T, \qquad (3) \\ l_t &= -\,\mathbf{c}_t^T \mathbf{F}\bar{\mathbf{x}}_t \qquad (4) \end{aligned}$$

where $\bar{\mathbf{x}}_t = [\bar{x}_{t-M-N_2+1}, \ldots, \bar{x}_{t+N_1}]^T$, $\mathbf{Q}_t \triangleq E[(\mathbf{x}_t - \bar{\mathbf{x}}_t)(\mathbf{x}_t - \bar{\mathbf{x}}_t)^H]$ is a diagonal matrix (due to uncorrelateness assumption on x_t) with diagonal entries $\mathbf{Q}_t = \mathrm{diag}([q_{t-M-N_2+1}, \ldots, q_t, \ldots, q_{t+N_1}])$ and $\mathbf{v} \in \mathbb{R}^{N+M-1}$ is a vector of all zeros except the $(M+N_2)$th entry is equal to 1. Then, the corresponding linear MMSE is given by

$$\begin{aligned} &\min_{\mathbf{c},l} E\left[\|x_t - \hat{x}_t\|^2\right] \\ &= \mathbf{v}^T \left[\mathbf{Q}_t - \mathbf{Q}_t\mathbf{F}^H \left(\sigma_n^2\mathbf{I} + \mathbf{F}\mathbf{Q}_t\mathbf{F}^H\right)^{-1} \mathbf{F}\mathbf{Q}_t\right] \mathbf{v} \\ &= \mathbf{v}^T \left(\mathbf{Q}_t^{-1} + \sigma_n^{-2}\mathbf{F}^H\mathbf{F}\right)^{-1} \mathbf{v}. \qquad (5) \end{aligned}$$

However to remove dependency of $\hat{x}_t$ to $\mathrm{LLR}_{a,t}^E$ due to using $\bar{x}_t$ and q_t in (3) and (2), we set $\mathrm{LLR}_{a,t}^E$ to 0 while computing

$\hat{x}_t$, yielding $\bar{x}_t = 0$ and $q_t = 1$ [2]. This changes the covariance matrix to $\mathbf{Q}'_t \triangleq \mathbf{Q} + (1-q_t)\mathbf{v}\mathbf{v}^T$ and the mean of $\mathbf{x}_t$ to $\bar{\mathbf{x}}_t - \bar{x}_t\mathbf{v}$, resulting in (2) and (3)

$$\mathbf{c}_t = \left[\sigma_n^{-2}\mathbf{v}^T\left(\mathbf{Q}_t'^{-1} + \sigma_n^{-2}\mathbf{F}^H\mathbf{F}\right)^{-1}\mathbf{F}^H\right]^T \quad \text{and}$$

$$l_t = -\,\mathbf{c}_t^T\mathbf{F}(\bar{\mathbf{x}}_t - \bar{x}_t\mathbf{v}) \tag{6}$$

$$\hat{x}_t = \mathbf{c}_t^T\mathbf{y}_t + l_t. \tag{7}$$

Since, the underlying channel vector $\mathbf{f}$ is not accurately known at the receiver, but an estimate $\tilde{\mathbf{f}}_t$, $\|\mathbf{f} - \tilde{\mathbf{f}}_t\| \leq \delta$, $\delta \in \mathbb{R}^+$, $\delta < \infty$ is provided, one cannot directly calculate (6) or (7). In the next section, we investigate three methods: a method using the inaccurate $\tilde{\mathbf{f}}_t$ to calculate the linear MMSE equalizer in (7); a minimax equalizer and a competitive equalizer that incorporate the uncertainty in the problem formulation to mitigate the effect of uncertainty on the equalization performance.

III. Equalization Methods

A. Linear MMSE Equalization

When the underlying communication channel $\mathbf{f}$ is not accurately known but estimated by $\tilde{\mathbf{f}}_t$, one may use a linear MMSE equalizer that is matched to the estimated channel vector $\tilde{\mathbf{f}}_t$ as

$$\begin{aligned}\tilde{\mathbf{c}}_t &= \left[\mathbf{v}^T\mathbf{Q}'_t\widetilde{\mathbf{F}}_t^H\left(\sigma_n^2\mathbf{I} + \widetilde{\mathbf{F}}_t\mathbf{Q}'_t\widetilde{\mathbf{F}}_t^H\right)^{-1}\right]^T \\ &= \left[\sigma_n^{-2}\mathbf{v}^T\left(\mathbf{Q}_t'^{-1} + \sigma_n^{-2}\widetilde{\mathbf{F}}_t^H\widetilde{\mathbf{F}}_t\right)^{-1}\widetilde{\mathbf{F}}_t^H\right]^T \\ \tilde{l}_t &= -\,\tilde{\mathbf{c}}_t^T\widetilde{\mathbf{F}}_t\bar{\mathbf{x}}_t\end{aligned} \tag{8}$$

where $\widetilde{\mathbf{F}}_t$ is the convolution matrix generated using $\tilde{\mathbf{f}}_t$. Calculating $\tilde{\mathbf{c}}_t$ using (8) at each time t may be computationally infeasible for certain applications since (8) requires $O(M^2 + N^2)$ operations (as shown in Fig. 2). Assuming that the channel estimate is time invariant, i.e., $\tilde{\mathbf{f}}_t = \tilde{\mathbf{f}}$, then one can reduce the computational complexity by approximating the covariance matrix $\mathbf{Q}'_t$. As an example, one can compute (8) using $\mathbf{Q}'_t = \mathbf{I}$, i.e., assuming unit variance and zero mean for each x_t, corresponding to a covariance matrix constructed without *a priori* information on x_t [2]. Then, (8) can be solved once and the resulting time invariant linear equalizer can be used over the whole block of received data in (3) and (4) [2], [25]. A better approximation can be achieved by computing (8) using $\mathbf{Q}_t = \frac{1}{n_x}(\sum_{i=1}^{n_x} q_i)\mathbf{I}$, i.e., using time averaged variances, yielding $\mathbf{Q}'_t = \frac{1}{n_x}(\sum_{i=1}^{n_x} q_i)\mathbf{I} - (1-q_t)\mathbf{v}\mathbf{v}^T$, where n_x is the size of the data block [25]. By this approximation (8) yields [25]

$$\begin{aligned}\tilde{\mathbf{c}}_t^{\mathrm{APP}} &= \left[\mathbf{v}^T\widetilde{\mathbf{F}}^H\left(\beta\widetilde{\mathbf{F}}\widetilde{\mathbf{F}}^H + \sigma_n^2\mathbf{I} + (1-q_t)\widetilde{\mathbf{F}}\mathbf{v}\mathbf{v}^T\widetilde{\mathbf{F}}^H\right)^{-1}\right]^T \\ &= \frac{\left[\mathbf{v}^T\widetilde{\mathbf{F}}^H\left(\beta\widetilde{\mathbf{F}}\widetilde{\mathbf{F}}^H + \sigma_n^2\mathbf{I}\right)^{-1}\right]^T}{1 + (1-q_t)\mathbf{v}^T\widetilde{\mathbf{F}}^H\left(\beta\widetilde{\mathbf{F}}\widetilde{\mathbf{F}}^H + \sigma_n^2\mathbf{I}\right)^{-1}\widetilde{\mathbf{F}}\mathbf{v}}\end{aligned} \tag{9}$$

where $\beta \triangleq \frac{1}{n_x}\sum_{i=1}^{n_x} q_i$, (9) follows from matrix inversion lemma and $\widetilde{\mathbf{F}}$ is the convolution matrix generated from the time invariant channel estimate $\tilde{\mathbf{f}}$. Note that to get a time invariant version of (9), one can use a time invariant $\mathbf{Q}'_t = \mathbf{I}$ assuming no *a priori* knowledge on x_t or $\mathbf{Q}'_t = \beta\mathbf{I}$, i.e., without $(1-q_t)\mathbf{v}\mathbf{v}^T$ term. The linear equalizer with time invariant approximation is given by

$$\begin{aligned}\tilde{\mathbf{c}}^{\mathrm{APP}} &= \left[\beta\mathbf{v}^T\widetilde{\mathbf{F}}^H\left(\sigma_n^2\mathbf{I} + \widetilde{\mathbf{F}}\widetilde{\mathbf{F}}^H\right)^{-1}\right]^T \\ &= \left[\sigma_n^{-2}\mathbf{v}^T\left(\beta\mathbf{I} + \sigma_n^{-2}\widetilde{\mathbf{F}}^H\widetilde{\mathbf{F}}\right)^{-1}\widetilde{\mathbf{F}}^H\right]^T.\end{aligned} \tag{10}$$

The required number of computations at each time t, per received symbol y_t, for $\tilde{\mathbf{c}}_t$ and $\tilde{\mathbf{c}}^{\mathrm{APP}}$ is given in Table I.

B. Linear Equalization With a Minimax Formulation

When the underlying channel $\mathbf{f}$ is unknown but estimated by $\tilde{\mathbf{f}}_t$, the uncertainty in the channel estimate can be incorporated in the equalizer design using a minimax framework in order to improve robustness over (8). In this minimax framework, one optimizes the MSE performance with respect to the worst possible communication channel around the channel estimate $\tilde{\mathbf{f}}_t$ and seeks for equalizer coefficients that minimize the worst case MSE, i.e.

$$\left\{\tilde{\mathbf{c}}_t^{\mathrm{MM}}, \tilde{l}_t^{\mathrm{MM}}\right\} = \arg\min_{\mathbf{c},l}\ \max_{\mathbf{f}=\tilde{\mathbf{f}}_t+\mathbf{df},\|\mathbf{df}\|\leq\delta}\left\{E\left[|x_t - \bar{x}_t - \mathbf{c}^T\mathbf{y}_t - l|^2\right]\right\}. \tag{11}$$

Note that we have

$$\begin{aligned}E&\left[|x_t - \bar{x}_t - \mathbf{c}^T\mathbf{y}_t - l|^2\right] \\ &= E\left[|x_t - \bar{x}_t - \mathbf{c}^T\mathbf{F}\mathbf{x}_t - \mathbf{c}^T\mathbf{n}_t - l + \mathbf{c}^T\mathbf{F}\bar{\mathbf{x}}_t - \mathbf{c}^T\mathbf{F}\bar{\mathbf{x}}_t|^2\right] \\ &= (\mathbf{v} - \mathbf{F}^T\mathbf{c})^H\mathbf{Q}'_t(\mathbf{v} - \mathbf{F}^T\mathbf{c}) + \sigma_n^2\mathbf{c}^H\mathbf{c} + |l + \mathbf{c}^T\mathbf{F}\bar{\mathbf{x}}_t|^2 \\ &= (\mathbf{v} - \mathbf{C}^T\mathbf{f})^H\mathbf{Q}'_t(\mathbf{v} - \mathbf{C}^T\mathbf{f}) + \sigma_n^2\mathbf{c}^H\mathbf{c} \\ &\quad + |l + \mathbf{f}^T\mathbf{C}\bar{\mathbf{x}}_t|^2\end{aligned} \tag{12}$$

where $\mathbf{C}$ is the convolution matrix constructed using $\mathbf{c}$, the second line follows since n_t is i.i.d. and $(\mathbf{x}_t - \bar{\mathbf{x}}_t)$ has zero mean. To find the minimax equalizer coefficients $\{\tilde{\mathbf{c}}_t^{\mathrm{MM}}, \tilde{l}_t^{\mathrm{MM}}\}$ satisfying (11), we formulate the corresponding problem in (11) as an SDP problem. We emphasize that SDP problems are convex constrained convex optimization problems, where efficient methods exist for their solutions [10]. The following theorem, whose proof is given in Section III-D, provides the corresponding robust linear equalizer while solving the corresponding SDP problem.

Theorem 1: Let $\{x_t\}$, $\{y_t\}$ and $\{n_t\}$ represent the transmitted, received and noise sequences in Fig. 1 such that $y_t = f_t * x_t + n_t$, where $\mathbf{f} = [f_{M-1}, \ldots, f_0]^T$ is the unknown channel impulse response vector and n_t is zero mean. At each time t, given an estimate $\tilde{\mathbf{f}}_t$ of the underlying communication channel response vector $\mathbf{f}$ satisfying $\mathbf{f} = \tilde{\mathbf{f}}_t + \mathbf{df}$, $\|\mathbf{df}\| \leq \delta$, then

$$\underset{\mathbf{c},l}{\text{minimize}}\ \underset{\mathbf{f}=\tilde{\mathbf{f}}_t+\mathbf{df},\|\mathbf{df}\|\leq\delta}{\text{maximize}}\ \Big[(\mathbf{v} - \mathbf{C}^T\mathbf{f})^H\mathbf{Q}'_t(\mathbf{v} - \mathbf{C}^T\mathbf{f}) + \sigma_n^2\mathbf{c}^H\mathbf{c} + |l + \mathbf{f}^T\mathbf{C}\bar{\mathbf{x}}_t|^2\Big] \tag{13}$$

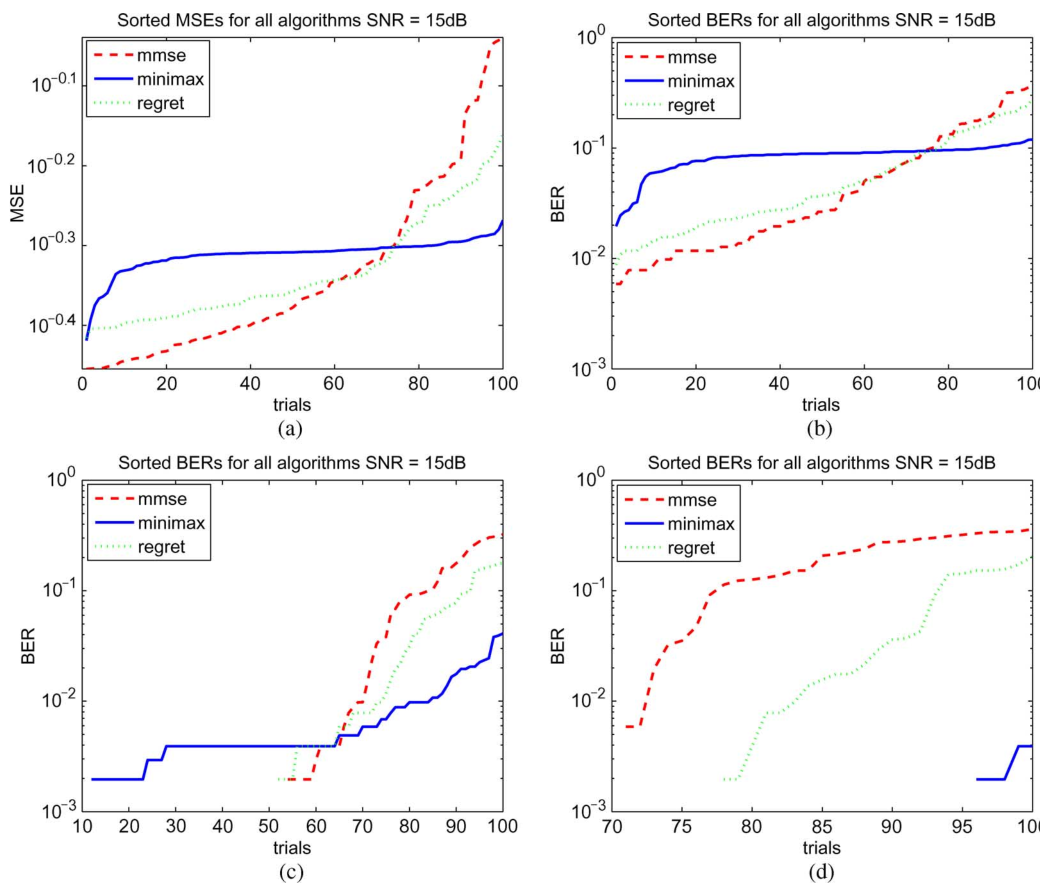


Fig. 2. Equalization results for the channel in (36) for 100 randomly introduced distortions with $|\mathbf{df}| \leq 0.3$. Here, $\mathrm{SNR} = 15$ dB. Included algorithms are $\tilde{\mathbf{c}}_t$ from (8) labeled "mmse," $\tilde{\mathbf{c}}_t^{\mathrm{MM}}$ from (11) labeled "minimax" and $\tilde{\mathbf{c}}_t^{\mathrm{CP}}$ from (21) labeled "regret." (a) Sorted MSEs for the 1st iteration. (b) Sorted BERs for the 1st iteration. (c) Sorted BERs for the 2nd iteration. (d) Sorted BERs for the 3rd iteration.

TABLE I
NUMBER OF REQUIRED COMPUTATIONS TO IMPLEMENT THE CORRESPONDING ALGORITHMS AT EACH TIME t PER RECEIVED SYMBOL y_t FOR LARGE PACKET SIZE. HERE, N IS THE EQUALIZER LENGTH, M IS THE CHANNEL LENGTH

	number of multiplications	number of additions
MMSE (time variant), $\tilde{\mathbf{c}}_t$	$O(N^2 + M^2)$	$O(N^2 + M^2)$
MMSE (time invariant), $\tilde{\mathbf{c}}^{\mathrm{APP}}$	$O(N + M)$	$O(N + M)$
Minimax (time variant), $\tilde{\mathbf{c}}_t^{\mathrm{MM}}$	$O(N\,(N + M)^{7/2})$	$O(N\,(N + M)^{7/2})$
Minimax (time invariant), $\tilde{\mathbf{c}}^{\mathrm{MM,APP}}$	$O(N + M)$	$O(N + M)$
Competitive (time variant), $\tilde{\mathbf{c}}_t^{\mathrm{CP}}$	$O(N\,(N + M)^{7/2})$	$O(N\,(N + M)^{7/2})$
Competitive (time invariant), $\tilde{\mathbf{c}}^{\mathrm{CP,APP}}$	$O(N + M)$	$O(N + M)$

where $\mathbf{c} = [c_{N_2}, \ldots, c_{-N_1}]^T$ and l are the coefficients of the linear equalizer, $\mathbf{C}$ is the convolution matrix generated from $\mathbf{c}$, $\mathbf{Q}_t' = \mathbf{Q}_t - (1 - q_t)\mathbf{v}\mathbf{v}^T$ and $E[\mathbf{n}_t\mathbf{n}_t^H] = \sigma_n^2\mathbf{I}$ are the covariance matrices of the transmitted and noise sequences, respectively, is equivalent to the SDP problem

$$\underset{\alpha,\mathbf{c},l,\tau}{\mathrm{minimize}}\,\alpha \tag{14}$$

such that [see (15) at the bottom of the next page]. The minimizer $\{\mathbf{c}, l\}$ in (14) yields the robust linear equalizer $\{\tilde{\mathbf{c}}_t^{\mathrm{MM}}, \tilde{l}_t^{\mathrm{MM}}\}$ in (11).

The proof of the theorem is provided in Section III-D. We note that in Theorem 1, for notational simplicity, we have dropped the time indices from $\mathbf{f}$ and δ. The same formulation equally applies to time varying δ.

To get the corresponding log-likelihood ratios, $\mathrm{LLR}_{e,t}^{E} = \ln \frac{p(\hat{x}_t|x_t=+1)}{p(\hat{x}_t|x_t=-1)}$, to fed into the decoder, we assume that $p(\hat{x}_t|x_t = x)$ is a Gaussian distribution with $\mathcal{N}(E[\hat{x}_t|x_t = x], \mathrm{Cov}(\hat{x}_t, \hat{x}_t|x_t = x))$ [2]. With the formulation $\hat{x}_t = \tilde{\mathbf{c}}_t^{\mathrm{MM},T}\mathbf{y}_t + \tilde{l}_t^{\mathrm{MM}}$, one calculates $E[\hat{x}_t|x_t = x]$ and $\mathrm{Cov}(\hat{x}_t, \hat{x}_t|x_t = x)$ as

$$\begin{aligned} &E[\hat{x}_t|x_t = x] \\ &\quad = E\left[\tilde{\mathbf{c}}_t^{\mathrm{MM},T}\mathbf{y}_t + \tilde{l}_t^{\mathrm{MM}}|x_t = x\right] \\ &\quad = \tilde{\mathbf{c}}_t^{\mathrm{MM},T}\mathbf{F}(\bar{\mathbf{x}}_t - \bar{x}_t\mathbf{v} + x\mathbf{v}) + \tilde{l}_t^{\mathrm{MM}}, \qquad (16) \\ &\mathrm{Cov}(\hat{x}_t, \hat{x}_t|x_t = x) \\ &\quad = \tilde{\mathbf{c}}_t^{\mathrm{MM},T}\left[\sigma_n^2\mathbf{I} + \mathbf{F}(\mathbf{Q}_t - q_t\mathbf{v}\mathbf{v}^T)\mathbf{F}^T\right]\tilde{\mathbf{c}}_t^{\mathrm{MM}}. \qquad (17) \end{aligned}$$

Since $\mathbf{F}$ is unknown, we use $\widetilde{\mathbf{F}}_t$ in (16) and (17) to calculate $\mathrm{LLR}_{e,t}^{E}$.

As in (8) of Section III-A, the SDP problem in (14) should be solved at each time t to compute $\tilde{\mathbf{c}}_t^{\mathrm{MM}}$ since $\mathbf{Q}_t'$ and (possibly) $\tilde{\mathbf{f}}_t$ are time dependent. Although there exist efficient methods to solve the corresponding SDP problem presented in Theorem 1, using these methods for each time instant t may be infeasible in certain applications since these calculations have $O\left(N(N+M)^{\frac{7}{2}}\right)$ computational complexity. To reduce computational complexity, assuming a time invariant channel estimate $\tilde{\mathbf{f}}_t = \tilde{\mathbf{f}}$, one can use a time invariant covariance matrix $\mathbf{Q}_t' = \mathbf{I}$ corresponding to a covariance matrix constructed without *a priori* information on x_t or $\mathbf{Q}_t' = \beta\mathbf{I}$ corresponding to time averaged variances. We note that unlike in Section III-A, we can not directly use $\mathbf{Q}_t' = \beta\mathbf{I} - (1-q_t)\mathbf{v}\mathbf{v}^T$, since this formulation is time dependent. Then, the corresponding SDP problem can be solved once to yield a time invariant equalizer, which can be used over the whole block, i.e.

$$\underset{\alpha,\mathbf{c},l,\tau}{\mathrm{minimize}}\ \alpha \qquad (18)$$

such that (see the equation at the bottom of the page).

The required number of computations at each time t, per received symbol y_t, for $\tilde{\mathbf{c}}_t^{\mathrm{MM}}$ and $\tilde{\mathbf{c}}^{\mathrm{MM,APP}}$ is given in Fig. 2.

C. Linear Equalization With Competitive Algorithm Formulation

We note that the minimax framework investigated in Section III-B to construct robust linear equalizers may produce overly conservative solutions in certain applications, since the linear equalizers are optimized to minimize the MSE corresponding to the worst possible channel. To improve the equalization performance, while trying to preserve robustness, a competitive approach may be used [4], [5], [7], [21]. In this competitive framework, instead of the usual MSE performance, the performance of a linear equalizer is defined with respect to the MMSE linear equalizer tuned to the underlying unknown channel, i.e., we compete against the linear equalizer that is constructed using the complete knowledge of the (unknown) underlying channel. For any affine equalizer coefficients $\{\mathbf{c}, l\}$, we define our regret for using the linear equalizer coefficients $\{\mathbf{c}, l\}$ instead of the linear MMSE equalizer tuned to $\mathbf{f}$ as

$$\begin{aligned} &E\left[|x_t - \bar{x}_t - \mathbf{c}^T\mathbf{y}_t - l|^2\right] \\ &\quad - \left(\min_{\mathbf{w},r} E\left[|x_t - \bar{x}_t - \mathbf{w}^T\mathbf{y}_t - r)|^2\right]\right) \\ &\quad = \left[(\mathbf{v} - \mathbf{C}^T\mathbf{f})^H\mathbf{Q}_t'(\mathbf{v} - \mathbf{C}^T\mathbf{f}) + \sigma_n^2\mathbf{c}^H\mathbf{c} + |l + \mathbf{f}^T\mathbf{C}\bar{\mathbf{x}}_t|^2\right] \\ &\quad - \left(\mathbf{v}^T\left[\mathbf{Q}_t'^{-1} + \sigma_n^{-2}\mathbf{F}^H\mathbf{F}\right]^{-1}\mathbf{v}\right), \qquad (19) \end{aligned}$$

where (12) and (5) are used in (19). However, to make the SDP problem formulation tractable, instead of directly using $(\mathbf{v}^T[\mathbf{Q}_t'^{-1} + \sigma_n^{-2}\mathbf{F}^H\mathbf{F}]^{-1}\mathbf{v})$ in the regret formulation of (19), one can use a first-order linear (Taylor) approximation around $\tilde{\mathbf{f}}_t$ [7], given in Appendix, as

$$\mathbf{v}^T\left[\mathbf{Q}_t'^{-1} + \sigma_n^{-2}\mathbf{F}^H\mathbf{F}\right]^{-1}\mathbf{v} = \eta_t + \mathbf{df}^H\mathbf{g}_t^{+} + \mathbf{g}_t^{T}\mathbf{df} + O\left(\|\mathbf{df}\|^2\right)$$

$$\begin{bmatrix} \alpha-\tau & \mathbf{c}^H & (\mathbf{v}-\mathbf{C}^T\tilde{\mathbf{f}}_t)^H & \left(l+\bar{\mathbf{x}}_t^T\mathbf{C}^T\tilde{\mathbf{f}}_t\right)^H & \mathbf{0} \\ \mathbf{c} & \sigma_n^{-2}\mathbf{I} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ (\mathbf{v}-\mathbf{C}^T\tilde{\mathbf{f}}_t) & \mathbf{0} & \mathbf{Q}_t'^{-1} & \mathbf{0} & -\delta\mathbf{C}^T \\ \left(l+\bar{\mathbf{x}}_t^T\mathbf{C}^T\tilde{\mathbf{f}}_t\right) & \mathbf{0} & \mathbf{0} & 1 & \delta\bar{\mathbf{x}}_t^T\mathbf{C}^T \\ \mathbf{0} & \mathbf{0} & -\delta\mathbf{C}^{+} & \delta\mathbf{C}^{+}\bar{\mathbf{x}}_t^{+} & \tau\mathbf{I} \end{bmatrix} \geq 0. \qquad (15)$$

$$\begin{bmatrix} \alpha-\tau & \mathbf{c}^H & (\mathbf{v}-\mathbf{C}^T\tilde{\mathbf{f}})^H & \left(l+\bar{\mathbf{x}}_t^T\mathbf{C}^T\tilde{\mathbf{f}}\right)^H & \mathbf{0} \\ \mathbf{c} & \sigma_n^{-2}\mathbf{I} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ (\mathbf{v}-\mathbf{C}^T\tilde{\mathbf{f}}) & \mathbf{0} & \beta\mathbf{I} & \mathbf{0} & -\delta\mathbf{C}^T \\ \left(l+\bar{\mathbf{x}}_t^T\mathbf{C}^T\tilde{\mathbf{f}}\right) & \mathbf{0} & \mathbf{0} & 1 & \delta\bar{\mathbf{x}}_t^T\mathbf{C}^T \\ \mathbf{0} & \mathbf{0} & -\delta\mathbf{C}^{+} & \delta\mathbf{C}^{+}\bar{\mathbf{x}}_t^{+} & \tau\mathbf{I} \end{bmatrix} \geq 0.$$

where $\eta_t \triangleq \mathbf{v}^T[\mathbf{Q}_t'^{-1} + \sigma_n^{-2}\widetilde{\mathbf{F}}_t^H\widetilde{\mathbf{F}}_t]^{-1}\mathbf{v}$ and $\mathbf{g}_t \triangleq -\widetilde{\mathbf{C}}_t(\mathbf{Q}_t' + \sigma_n^{-2}\widetilde{\mathbf{F}}_t^H\widetilde{\mathbf{F}}_t)^{-1}\mathbf{v}$ and $\widetilde{\mathbf{C}}_t$ is the convolution matrix constructed using $\tilde{\mathbf{c}}_t$ in (8). Using this in (19) yields the regret as

$$\left[(\mathbf{v}-\mathbf{C}^T\mathbf{f})^H\mathbf{Q}_t'(\mathbf{v}-\mathbf{C}^T\mathbf{f}) + \sigma_n^2\mathbf{c}^H\mathbf{c} + |l + \mathbf{f}^T\mathbf{C}\bar{\mathbf{x}}_t|^2\right] - \left(\eta_t + \mathbf{df}^H\mathbf{g}_t{}^+ + \mathbf{g}_t{}^T\mathbf{df}\right) \quad (20)$$

where the $O(\|\mathbf{df}\|^2)$ term is left out. Note that the first order Taylor approximation is introduced in order to make the solution of (19) in a minimax setting tractable. Clearly, the effect of this approximation diminishes as $\|\mathbf{df}\|$ gets smaller. For distortions with larger $\|\mathbf{df}\|$, one can use the higher order Taylor approximations instead, however, we have observed through our simulations that the solution using the first order approximation yields satisfactory results even for fairly large $\|\mathbf{df}\|$ (when compared to $\|\mathbf{f}\|$).

To get the competitive linear equalizer, we minimize this regret over all possible communication channels around the channel estimate $\tilde{\mathbf{f}}_t$, i.e.

$$\begin{aligned}\left\{\tilde{\mathbf{c}}_t^{\mathrm{CP}}, \tilde{l}_t^{\mathrm{CP}}\right\} &= \arg\min_{\mathbf{c},l} \max_{\mathbf{f}=\tilde{\mathbf{f}}_t+\mathbf{df}, \|\mathbf{df}\|\le\delta} \Big[(\mathbf{v}-\mathbf{C}^T\mathbf{f})^H\mathbf{Q}_t'(\mathbf{v}-\mathbf{C}^T\mathbf{f}) \\ &\quad + \sigma_n^2\mathbf{c}^H\mathbf{c} + |l + \mathbf{f}^T\mathbf{C}\bar{\mathbf{x}}_t|^2 \\ &\quad - \left(\eta_t + \mathbf{df}^H\mathbf{g}_t{}^+ + \mathbf{g}_t{}^T\mathbf{df}\right)\Big]. \quad (21)\end{aligned}$$

The problem in (21) that will yield the corresponding competitive linear equalizer can be formulated as an SDP problem as follows.

Theorem 2: Let $\{x_t\}$, $\{y_t\}$ and $\{n_t\}$ represent the transmitted, received and noise sequences in Fig. 1 such that $y_t = f_t * x_t + n_t$, where $\mathbf{f}$ is the unknown channel impulse response vector and n_t is zero mean. At each time t, given an estimate $\tilde{\mathbf{f}}_t$ of the underlying communication channel impulse response vector $\mathbf{f}$ satisfying $\mathbf{f} = \tilde{\mathbf{f}}_t + \mathbf{df}$, $\|\mathbf{df}\| \le \delta$, then

$$\begin{aligned}\underset{\mathbf{c},l}{\text{minimize}} \underset{\mathbf{f}=\tilde{\mathbf{f}}_t+\mathbf{df},\|\mathbf{df}\|\le\delta}{\text{maximize}} &\Big[(\mathbf{v}-\mathbf{C}^T\mathbf{f})^H\mathbf{Q}_t'(\mathbf{v}-\mathbf{C}^T\mathbf{f}) \\ &+ \sigma_n^2\mathbf{c}^H\mathbf{c} + |l + \mathbf{f}^T\mathbf{C}\bar{\mathbf{x}}_t|^2 - \left(\eta_t + \mathbf{df}^H\mathbf{g}_t{}^+ + \mathbf{g}_t{}^T\mathbf{df}\right)\Big] \quad (22)\end{aligned}$$

where $\mathbf{c} = [c_{N_2}, \ldots, c_{-N_1}]^T$ and l are the coefficients of the linear equalizer, $\mathbf{C}$ is the convolution matrix generated from $\mathbf{c}$, $\mathbf{Q}_t' = \mathbf{Q}_t - (1-q_t)\mathbf{v}\mathbf{v}^T$ and $E[\mathbf{n}_t\mathbf{n}_t^H] = \sigma_n^2\mathbf{I}$ are the covariance matrices of the transmitted and noise sequences, respectively, $\eta_t = \mathbf{v}^T[\mathbf{Q}_t'^{-1} + \sigma_n^{-2}\widetilde{\mathbf{F}}_t^H\widetilde{\mathbf{F}}_t]^{-1}\mathbf{v}$, $\mathbf{g}_t = -\widetilde{\mathbf{C}}_t(\mathbf{Q}_t' + \sigma_n^{-2}\widetilde{\mathbf{F}}_t^H\widetilde{\mathbf{F}}_t)^{-1}\mathbf{v}$, and >0, is equivalent to the SDP problem

$$\underset{\alpha,\mathbf{c},l,\tau}{\text{minimize}}\ \alpha \quad (23)$$

such that [see (24) at the bottom of the page]. The minimizer $\{\mathbf{c}, l\}$ in (23) yields the competitive linear equalizer coefficients $\{\tilde{\mathbf{c}}_t^{\mathrm{CP}}, \tilde{l}_t^{\mathrm{CP}}\}$ in (21).

The proof of the theorem is provided in Section III-D. We note that in Theorem 2, for notational simplicity, we have dropped the time indices from $\mathbf{f}$ and δ. The same formulation equally applies to time varying δ. Note that to get the corresponding $\mathrm{LLR}_{e,t}^E$, one needs to replace $\{\tilde{\mathbf{c}}_t^{\mathrm{MM}}, \tilde{l}_t^{\mathrm{MM}}\}$ with $\{\tilde{\mathbf{c}}_t^{\mathrm{CP}}, \tilde{l}_t^{\mathrm{CP}}\}$ in (16) and (17).

The proof of the theorem is given in Section III-D. As in Section III-C, instead of solving the SDP problem for all t, one can approximate the time varying correlation matrix $\mathbf{Q}_t'$ with a time invariant correlation matrix $\mathbf{Q}_t' = \mathbf{I}$ or $\mathbf{Q}_t' = \beta\mathbf{I}$. Then, assuming a time invariant channel estimate $\tilde{\mathbf{f}}$, the SDP problem in (23) can be solved only once. The linear equalizer $\tilde{\mathbf{c}}^{\mathrm{CP,APP}}$ calculated under this approximation can then be used over the whole block of received data with the corresponding SDP problem formulation

$$\underset{\alpha,\mathbf{c},l,\tau}{\text{minimize}}\ \alpha \quad (25)$$

such that (see the first equation at the bottom of the next page) where η and $\mathbf{g}$ are computed using time invariant $\tilde{\mathbf{f}}$ and $\mathbf{Q}'$.

The required number of computations at each time t, per received symbol y_t, for $\tilde{\mathbf{c}}_t^{\mathrm{CP}}$ and $\tilde{\mathbf{c}}^{\mathrm{CP,APP}}$ is given in Fig. 2.

D. Proofs of Theorem 1 and Theorem 2

Proof of Theorem 1: We first observe that the MSE expression in (13) can be written as

$$\min_{\mathbf{c},l} \max_{\mathbf{f}=\tilde{\mathbf{f}}_t+\mathbf{df},\|\mathbf{df}\|\le\delta} \Big[(\mathbf{v}-\mathbf{C}^T\mathbf{f})^H\mathbf{Q}_t'(\mathbf{v}-\mathbf{C}^T\mathbf{f}) + \sigma_n^2\mathbf{c}^H\mathbf{c} + |l + \mathbf{f}^T\mathbf{C}\bar{\mathbf{x}}_t|^2\Big] = \min_{\mathbf{c},l,\mathbf{df},\alpha} \alpha$$

such that

$$(\mathbf{v}-\mathbf{C}^T\mathbf{f})^H\mathbf{Q}_t'(\mathbf{v}-\mathbf{C}^T\mathbf{f}) + \sigma_n^2\mathbf{c}^H\mathbf{c} + |l + \mathbf{f}^T\mathbf{C}\bar{\mathbf{x}}_t|^2 \le \alpha \quad (26)$$

$$\begin{bmatrix} \alpha + \eta_t - \tau & \mathbf{c}^H & (\mathbf{v}-\mathbf{C}^T\tilde{\mathbf{f}}_t)^H & \left(l + \bar{\mathbf{x}}_t^T\mathbf{C}^T\tilde{\mathbf{f}}_t\right)^H & \delta\mathbf{g}_t{}^T \\ \mathbf{c} & \sigma_n^{-2}\mathbf{I} & \mathbf{0} & 0 & \mathbf{0} \\ (\mathbf{v}-\mathbf{C}^T\tilde{\mathbf{f}}_t) & \mathbf{0} & \mathbf{Q}_t'^{-1} & \mathbf{0} & -\delta\mathbf{C}^T \\ \left(l + \bar{\mathbf{x}}_t^T\mathbf{C}^T\tilde{\mathbf{f}}_t\right) & \mathbf{0} & \mathbf{0} & 1 & \delta\bar{\mathbf{x}}_t^T\mathbf{C}^T \\ \delta\mathbf{g}_t{}^+ & \mathbf{0} & -\delta\mathbf{C}^+ & \delta\mathbf{C}^+\bar{\mathbf{x}}_t^+ & \tau\mathbf{I} \end{bmatrix} \ge 0. \quad (24)$$

and $\|\mathbf{df}\| \leq \delta$. Using Lemma 2 from Appendix in (26) yields

$$\begin{bmatrix} \alpha-(\mathbf{v}-\mathbf{C}^T\mathbf{f})^H\mathbf{Q}'_t(\mathbf{v}-\mathbf{C}^T\mathbf{f})-|l+\mathbf{f}^T\mathbf{C}\bar{\mathbf{x}}_t|^2 & \mathbf{c}^H \\ \mathbf{c} & \sigma_n^{-2}\mathbf{I} \end{bmatrix} \geq 0. \tag{27}$$

Applying Lemma 2 from Appendix to (27) for the $(\mathbf{v}-\mathbf{C}^T\mathbf{f})^H\mathbf{Q}'_t(\mathbf{v}-\mathbf{C}^T\mathbf{f})$ term yields

$$\begin{bmatrix} \alpha-|l+\mathbf{f}^T\mathbf{C}\bar{\mathbf{x}}_t|^2 & \mathbf{c}^H & (\mathbf{v}-\mathbf{C}^T\mathbf{f})^H \\ \mathbf{c} & \sigma_n^{-2}\mathbf{I} & \mathbf{0} \\ (\mathbf{v}-\mathbf{C}^T\mathbf{f}) & \mathbf{0} & \mathbf{Q}'^{-1}_t \end{bmatrix} \geq 0. \tag{28}$$

Applying Lemma 2 to (28) for the term $|l+\mathbf{f}^T\mathbf{C}\bar{\mathbf{x}}_t|^2$ yields

$$\begin{bmatrix} \alpha & \mathbf{c}^H & (\mathbf{v}-\mathbf{C}^T\mathbf{f})^H & (l+\bar{\mathbf{x}}_t^T\mathbf{C}^T\mathbf{f})^H \\ \mathbf{c} & \sigma_n^{-2}\mathbf{I} & \mathbf{0} & \mathbf{0} \\ (\mathbf{v}-\mathbf{C}^T\mathbf{f}) & \mathbf{0} & \mathbf{Q}'^{-1}_t & \mathbf{0} \\ (l+\bar{\mathbf{x}}_t^T\mathbf{C}^T\mathbf{f}) & \mathbf{0} & \mathbf{0} & 1 \end{bmatrix} \geq 0. \tag{29}$$

However, (29) can be written as

$$\begin{bmatrix} \alpha & \mathbf{c}^H & (\mathbf{v}-\mathbf{C}^T\tilde{\mathbf{f}}_t)^H & \left(l+\bar{\mathbf{x}}_t^T\mathbf{C}^T\tilde{\mathbf{f}}_t\right)^H \\ \mathbf{c} & \sigma_n^{-2}\mathbf{I} & \mathbf{0} & \mathbf{0} \\ (\mathbf{v}-\mathbf{C}^T\tilde{\mathbf{f}}_t) & \mathbf{0} & \mathbf{Q}'^{-1}_t & \mathbf{0} \\ \left(l+\bar{\mathbf{x}}_t^T\mathbf{C}^T\tilde{\mathbf{f}}_t\right) & \mathbf{0} & \mathbf{0} & 1 \end{bmatrix} \geq \begin{bmatrix} \mathbf{0} \\ \mathbf{0} \\ \mathbf{C}^T \\ -\bar{\mathbf{x}}_t^T\mathbf{C}^T \end{bmatrix} \mathbf{df}[1 \quad 0 \quad 0 \quad 0] + \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} \mathbf{df}^H [\mathbf{0} \quad \mathbf{0} \quad \mathbf{C}^+ \quad -\mathbf{C}^+\bar{\mathbf{x}}_t^+]. \tag{30}$$

Applying Lemma 3 from Appendix to (30) yields (31), shown at the bottom of the page, with the constraint $\|\mathbf{df}\| \leq \delta$. Hence, using (31) in (26) results Theorem 1. This completes the proof of Theorem 1. □

Proof of Theorem 2: The proof of Theorem 2 follows the proof of Theorem 1. The MSE expression in (22) can be written as

$$\min_{\mathbf{c},l} \max_{\mathbf{f}=\tilde{\mathbf{f}}_t+\mathbf{df},\|\mathbf{df}\|\leq\delta} \Big[(\mathbf{v}-\mathbf{C}^T\mathbf{f})^H\mathbf{Q}'_t(\mathbf{v}-\mathbf{C}^T\mathbf{f})+\sigma_n^2\mathbf{c}^H\mathbf{c}+ |l+\mathbf{f}^T\mathbf{C}\bar{\mathbf{x}}_t|^2-\left(\eta_t+\mathbf{df}^H\mathbf{g}_t{}^+ +\mathbf{g}_t{}^T\mathbf{df}\right)\Big] = \min_{\mathbf{c},l,\mathbf{df},\alpha} \alpha$$

such that

$$(\mathbf{v}-\mathbf{C}^T\mathbf{f})^H\mathbf{Q}'_t(\mathbf{v}-\mathbf{C}^T\mathbf{f})+\sigma_n^2\mathbf{c}^H\mathbf{c}+|l+\mathbf{f}^T\mathbf{C}\bar{\mathbf{x}}_t|^2- \left(\eta_t+\mathbf{df}^H\mathbf{g}_t{}^+ +\mathbf{g}_t{}^T\mathbf{df}\right) \leq \alpha \tag{32}$$

and $\|\mathbf{df}\| \leq \delta$. Applying Lemma 2 to (32) for $(\mathbf{v}-\mathbf{C}^T\mathbf{f})^H\mathbf{Q}'_t(\mathbf{v} - \mathbf{C}^T\mathbf{f})$ and $|l+\mathbf{f}^T\mathbf{C}b\bar{f}x_t|^2$ successively two times yields the first equation at bottom of the next page, and (33), shown at the bottom of the next page. Since (33) can be written as

$$\begin{bmatrix} \alpha+\eta_t & \mathbf{c}^H & (\mathbf{v}-\mathbf{C}^T\tilde{\mathbf{f}}_t)^H & \left(l+\bar{\mathbf{x}}_t^T\mathbf{C}^T\tilde{\mathbf{f}}_t\right)^H \\ \mathbf{c} & \sigma_n^{-2}\mathbf{I} & \mathbf{0} & \mathbf{0} \\ (\mathbf{v}-\mathbf{C}^T\tilde{\mathbf{f}}_t) & \mathbf{0} & \mathbf{Q}'^{-1}_t & \mathbf{0} \\ \left(l+\bar{\mathbf{x}}_t^T\mathbf{C}^T\tilde{\mathbf{f}}_t\right) & \mathbf{0} & \mathbf{0} & 1 \end{bmatrix} \geq \begin{bmatrix} -\mathbf{g}_t{}^T \\ \mathbf{0} \\ \mathbf{C}^T \\ -\bar{\mathbf{x}}_t^T\mathbf{C}^T \end{bmatrix} \mathbf{df}[1 \quad 0 \quad 0 \quad 0] + \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} \mathbf{df}^H [-\mathbf{g}_t{}^+ \quad \mathbf{0} \quad \mathbf{C}^+ \quad -\mathbf{C}^+\bar{\mathbf{x}}_t^+] \tag{34}$$

$$\begin{bmatrix} \alpha+\eta-\tau & \mathbf{c}^H & (\mathbf{v}-\mathbf{C}^T\tilde{\mathbf{f}})^H & \left(l+\bar{\mathbf{x}}_t^T\mathbf{C}^T\tilde{\mathbf{f}}\right)^H & \delta\mathbf{g}^T \\ \mathbf{c} & \sigma_n^{-2}\mathbf{I} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ (\mathbf{v}-\mathbf{C}^T\tilde{\mathbf{f}}) & \mathbf{0} & \beta\mathbf{I} & \mathbf{0} & -\delta\mathbf{C}^T \\ \left(l+\bar{\mathbf{x}}_t^T\mathbf{C}^T\tilde{\mathbf{f}}\right) & \mathbf{0} & \mathbf{0} & 1 & \delta\bar{\mathbf{x}}_t^T\mathbf{C}^T \\ \delta\mathbf{g}^+ & \mathbf{0} & -\delta\mathbf{C}^+ & \delta\mathbf{C}^+\bar{\mathbf{x}}_t^+ & \tau\mathbf{I} \end{bmatrix} \geq 0.$$

$$\begin{bmatrix} \alpha-\tau & \mathbf{c}^H & (\mathbf{v}-\mathbf{C}^T\tilde{\mathbf{f}}_t)^H & \left(l+\bar{\mathbf{x}}_t^T\mathbf{C}^T\tilde{\mathbf{f}}_t\right)^H & \mathbf{0} \\ \mathbf{c} & \sigma_n^{-2}\mathbf{I} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ (\mathbf{v}-\mathbf{C}^T\tilde{\mathbf{f}}_t) & \mathbf{0} & \mathbf{Q}'^{-1}_t & \mathbf{0} & -\delta\mathbf{C}^T \\ \left(l+\bar{\mathbf{x}}_t^T\mathbf{C}^T\tilde{\mathbf{f}}_t\right) & \mathbf{0} & \mathbf{0} & 1 & \delta\bar{\mathbf{x}}_t^T\mathbf{C}^T \\ \mathbf{0} & \mathbf{0} & -\delta\mathbf{C}^+ & \delta\mathbf{C}^+\bar{\mathbf{x}}_t^+ & \tau\mathbf{I} \end{bmatrix} \geq 0 \tag{31}$$

with $\|\mathbf{df}\| \leq \delta$. Using Lemma 3 from Appendix in (35), shown at the bottom of the page. Using (35) in (32) yields Theorem 2. This completes the proof of Theorem 2. □

IV. Simulations

In this section, we illustrate the performance of the introduced algorithms under different settings. For all examples, we use the simulation setup from [26] with the channel examples from [3, Ch. 10]. Here, bits to be transmitted are encoded using a convolutional encoder with a generator matrix $G = [1\;0\;D^2; 1\;D\;D^2]$ [26]. An 8-random interleaver is used to shuffle the coded bits such that any consecutive bits will have a minimum distance of 8 bits after interleaving [23]. The coded bits are then BPSK modulated. We use linear equalizers introduced in the text and a MAP-based algorithm for decoding [3], [26].

In the first set of experiments, the modulated bits are transmitted through the ISI channel from [3, Ch. 10]

$$\mathbf{f} = [0.227 \quad 0.46 \quad 0.688 \quad 0.46 \quad 0.227]^T \tag{36}$$

with $\|\mathbf{f}\| = 1$, $M = 5$ and the noise variance σ_n^2 is determined by

$$\mathrm{SNR} = \frac{E\left[\|x_t\|^2\right]}{N_0} = \frac{1}{2\sigma_n^2}.$$

The channel estimates are constructed using $\tilde{\mathbf{f}} = \mathbf{f} + \mathbf{df}$, where the distortion $\mathbf{df}$ is randomly generated using a zero mean Gaussian distribution. In the first set of experiments, the norm of $\mathbf{df}$ is randomly scaled to give $|\mathbf{df}| \leq 0.3$ for each trial, the length of $\{x_t\}$ is selected as 2048 and the SNR is set to 15 dB. For all equalizers $N = 15$, $N_1 = 7$ and $N_2 = 7$. In Fig. 2(a), we plot the sorted MSEs, i.e., $E[\|x_t - \hat{x}_t\|^2]$, at the equalizer output for the first iteration of the turbo equalization with respect to 100 randomly selected $\mathbf{df}$'s. Here, we have $\tilde{\mathbf{c}}_t$ from (8) labeled "mmse", $\tilde{\mathbf{c}}^{\mathrm{MM}}$ from (11) labeled "minimax" and $\tilde{\mathbf{c}}^{\mathrm{CP}}$ from (21) labeled "regret." For the same algorithms, we also plot the sorted BERs at the decoder output with respect to randomly selected $\mathbf{df}$'s in Fig. 2(b). We observe that, as expected, the worst case MSE under channel distortion is minimized for the "minimax" algorithm. The same behavior is observed in BER plot in Fig. 2(b). However, although the "minimax" algorithm has the best worst case performance, its average performance over randomly selected channel distortions is worse than the "regret" and the "mmse" algorithms: the worst case and the average BERs for the "mmse" algorithm are 0.3542 and 0.0748, respectively; for the "minimax" algorithm are 0.1194 and 0.0847, respectively; for the "regret" algorithm are 0.2798 and 0.0682, respectively. Note that the worst case BER performance of the "regret" algorithm is worse than the "minimax" method but better than the plug-in MMSE. However, the average BER of the competitive approach is better than the "minimax" algorithm and the "mmse" algorithm. Hence, for these simulations, the competitive approach provides a trade-off between the worst case performance and the average case performance. We then plot the corresponding sorted BERs for the second and fourth iterations of turbo equalization. We observe similar results for the second and fourth iterations in Fig. 2(c) and Fig. 2(d), respectively, such that the robust methods outperform the plug-in "mmse" method for these simulations. We note that the performance improvement due to the robust methods becomes more noticeable as the turbo iteration count increases. We observe that since the "minimax" method is able to minimize the worst case performance over all random distortions (even in the first turbo iteration), it is able to further

$$\begin{bmatrix} \alpha - |l + \mathbf{f}^T\mathbf{C}\bar{\mathbf{x}}_t|^2 + \left(\eta_t + \mathbf{df}^H\mathbf{g}_t{}^+ + \mathbf{g}_t{}^T\mathbf{df}\right) & \mathbf{c}^H & (\mathbf{v} - \mathbf{C}^T\mathbf{f})^H \\ \mathbf{c} & \sigma_n^{-2}\mathbf{I} & \mathbf{0} \\ (\mathbf{v} - \mathbf{C}^T\mathbf{f}) & \mathbf{0} & \mathbf{Q}_t'^{-1} \end{bmatrix} \geq 0$$

$$\begin{bmatrix} \alpha + \left(\eta_t + \mathbf{df}^H\mathbf{g}_t{}^+ + \mathbf{g}_t{}^T\mathbf{df}\right) & \mathbf{c}^H & (\mathbf{v} - \mathbf{C}^T\mathbf{f})^H & (l + \mathbf{f}^T\mathbf{C}\bar{\mathbf{x}}_t)^H \\ \mathbf{c} & \sigma_n^{-2}\mathbf{I} & \mathbf{0} & \mathbf{0} \\ (\mathbf{v} - \mathbf{C}^T\mathbf{f}) & \mathbf{0} & \mathbf{Q}_t'^{-1} & \mathbf{0} \\ (l + \mathbf{f}^T\mathbf{C}\bar{\mathbf{x}}_t) & \mathbf{0} & \mathbf{0} & 1 \end{bmatrix} \geq 0. \tag{33}$$

$$\begin{bmatrix} \alpha + \eta_t - \tau & \mathbf{c}^H & (\mathbf{v} - \mathbf{C}^T\tilde{\mathbf{f}}_t)^H & \left(l + \bar{\mathbf{x}}_t^T\mathbf{C}^T\tilde{\mathbf{f}}_t\right)^H & \delta\mathbf{g}_t{}^T \\ \mathbf{c} & \sigma_n^{-2}\mathbf{I} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ (\mathbf{v} - \mathbf{C}^T\tilde{\mathbf{f}}_t) & \mathbf{0} & \mathbf{Q}_t'^{-1} & \mathbf{0} & -\delta\mathbf{C}^T \\ \left(l + \bar{\mathbf{x}}_t^T\mathbf{C}^T\tilde{\mathbf{f}}_t\right) & \mathbf{0} & \mathbf{0} & 1 & \delta\bar{\mathbf{x}}_t^T\mathbf{C}^T \\ \delta\mathbf{g}_t{}^+ & \mathbf{0} & -\delta\mathbf{C}^+ & \delta\mathbf{C}^+\bar{\mathbf{x}}_t^+ & \tau\mathbf{I} \end{bmatrix} \geq 0. \tag{35}$$

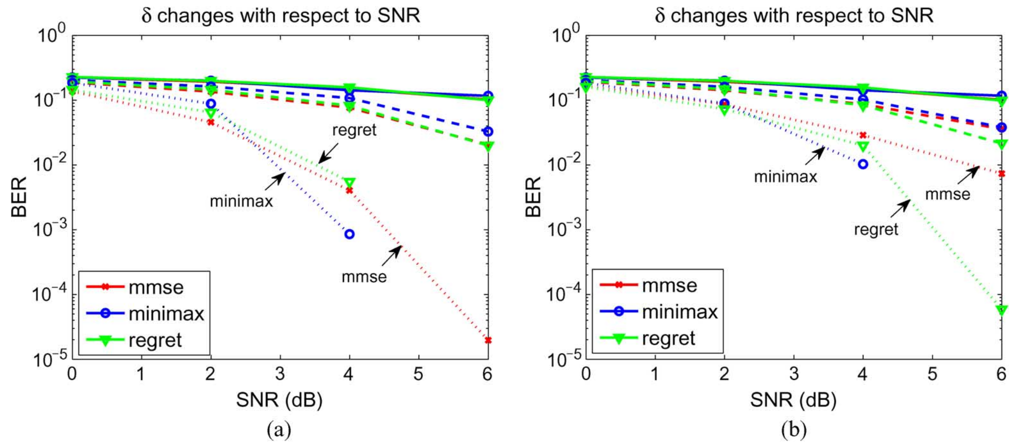


Fig. 3. Equalization results and average BERs for the channel in (36) under BPSK signaling and different SNRs. Here, the first iteration (the straight lines), the second iteration (dashed lines) and the fourth iteration (the dotted lines). (a) Included algorithms are $\tilde{\mathbf{c}}_t$ from (8) labeled "mmse," $\tilde{\mathbf{c}}_t^{\mathrm{MM}}$ from (11) labeled "minimax" and $\tilde{\mathbf{c}}_t^{\mathrm{CP}}$ from (21) labeled "regret." (b) Algorithms with low computational complexity: $\tilde{\mathbf{c}}^{\mathrm{APP}}$ from (10) labeled "mmse," $\tilde{\mathbf{c}}^{\mathrm{MM,APP}}$ from (18) labeled "minimax" and $\tilde{\mathbf{c}}^{\mathrm{CP,APP}}$ from (25) labeled "regret."

minimize the BERs (forcing them to zero) as turbo iteration count increases for all random distortions in these simulations.

In the next set of experiments, we simulate the performance of the introduced algorithms under different SNRs values over the channel in (36). However, since the channel estimates usually deteriorate with low SNR [27], we scale the bound for the norm of $\mathbf{df}$ inversely proportional to SNR to give $\|\mathbf{df}\| \leq 0.4$ for $\mathrm{SNR} = 0\ (\mathrm{dB})$ and $\|\mathbf{df}\| \leq 0.15$ for $\mathrm{SNR} = 6\ (\mathrm{dB})$, i.e., $\|\mathbf{df}\| \leq \frac{0.4-0.25\ \mathrm{SNR}}{6}$ (based on some empirical values). For these simulations, at each SNR, BERs are averaged over 200 random $\mathbf{df}$ and random $\{x_t\}$ with packet length 1024. Here, $N = 15$, $N_1 = 7$ and $N_2 = 7$. In Fig. 3(a), we present average BERs corresponding to the linear equalizers $\tilde{\mathbf{c}}_t$ from (8) labeled "mmse," $\tilde{\mathbf{c}}_t^{\mathrm{MM}}$ from (11) labeled "minimax" and $\tilde{\mathbf{c}}_t^{\mathrm{CP}}$ from (21) labeled "regret." We present BERs for the first iteration (the straight lines), the second iteration (dashed lines) and the fourth iteration (the dotted lines). We observe that although the robust algorithms are "optimized" with respect to the worst case MSE or to the worst case regret, their average performance is comparable and in certain SNRs much better than the plug-in method. We repeat the previous experiments with the same channels and the same system parameters to test the performance of the approximate implementations. In Fig. 3(b), we plot the BERs with respect to different SNRs over the channels from (36) for algorithms with low computational complexity: $\tilde{\mathbf{c}}^{\mathrm{APP}}$ from (10) labeled "mmse," $\tilde{\mathbf{c}}^{\mathrm{MM,APP}}$ from (18) labeled "minimax" and $\tilde{\mathbf{c}}^{\mathrm{CP,APP}}$ from (25) labeled "regret." For all equalizers, we use $\mathbf{Q}_t' = \mathbf{I}$. We present BERs corresponding to the first iteration (the straight lines), the second iteration (dashed lines) and the fourth iteration (the dotted lines). Although, as expected, the performance of the approximate implementations are inferior to exact implementations, we observe that the "regret" and "minimax" algorithms provide similar or better BERs with respect to the time invariant plug-in MMSE equalization algorithm for these simulations.

We next repeat the same set of experiments over a different channel from [3]

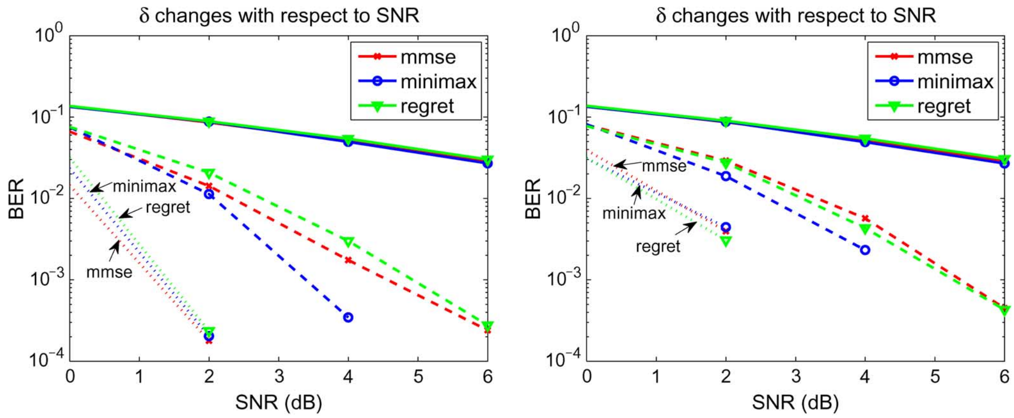
$$\mathbf{f} = [0.407 \quad 0.815 \quad 0.407]^T. \tag{37}$$

For these simulations, we run the experiments over 200 randomly selected channel distortions with packet length 4096. For this three tap channel, we choose $N = 7$, $N_1 = 3$, and $N_2 = 3$. The other system parameters are set to the same values as in the first set of experiments. As in the previous example, we scale the norm of randomly generated $\mathbf{df}$ inversely proportional to SNR such that $\|\mathbf{df}\| \leq 0.4$ for $\mathrm{SNR} = 0\ (\mathrm{dB})$ and $\|\mathbf{df}\| \leq 0.3$ for $\mathrm{SNR} = 6\ (\mathrm{dB})$, i.e., $\|\mathbf{df}\| \leq \frac{0.4-0.1\ \mathrm{SNR}}{6}$. The BERs with respect to different SNRs are plotted in Fig. 4(a). Note that since this channel introduces less severe ISI than (36), the BERs are better than the first channel. We observe similar behavior as in Fig. 3(a) such that the robust methods provide comparable or better BERs with respect to the plug-in method. In Fig. 4(b), we plot the BERs with respect to different SNRs over the channels from (37) for algorithms with low computational complexity and $\mathbf{Q}_t' = \mathbf{I}$ for all algorithms.

As the last set of examples, we perform robust turbo equalization in conjunction with the iterative channel estimation algorithm from [28]. We emphasize that our methods can be used along with such adaptive algorithms since usually the channel or system parameters cannot be learned perfectly by the adaptive algorithms and the uncertainty in learning can be incorporated in the equalizer design as done in this paper. In this setup, unlike the previous examples, each data block has a certain number of training samples known to both transmitter and receiver for channel estimation along with the test samples. At each turbo iteration, we perform channel estimation using the training data, the test data and the soft information provided by the SISO decoder. In particular, for channel estimation, we use a stochastic gradient update (which is suggested to be superior or comparable to the Newton based updates in [28] in the context of turbo equalization)

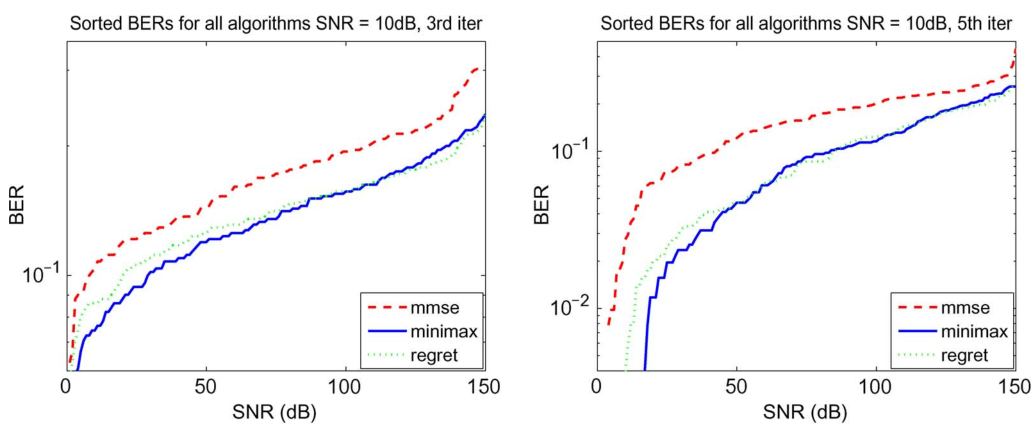
$$\tilde{\mathbf{f}}_{t+1} = \tilde{\mathbf{f}}_t + \mu e_t \mathbf{z}_t \tag{38}$$

where $\tilde{\mathbf{f}}_t$ is the length P estimated channel vector, for some P which may not be equal to the true channel length, $\mathbf{z}_t = [x_{t-P+1}, \ldots, x_t]^T$ in the training mode, $\mathbf{z}_t = [\tilde{x}_{t-P+1}, \ldots, \tilde{x}_t]^T$ in the first turbo iteration, where $\tilde{x}_t$ are the hard or quantized decisions (since no *a priori* statistics is available yet), i.e., the decision directed mode,



(a) (b)

Fig. 4. Equalization results and average BERs for the channel in (36) under BPSK signaling and different SNRs. Here, the first iteration (the straight lines), the second iteration (dashed lines) and the fourth iteration (the dotted lines). (a) Included algorithms are $\tilde{\mathbf{c}}_t$ from (8) labeled "mmse," $\tilde{\mathbf{c}}_t^{\mathrm{MM}}$ from (11) labeled "minimax" and $\tilde{\mathbf{c}}_t^{\mathrm{CP}}$ from (21) labeled "regret." (b) Algorithms with low computational complexity: $\tilde{\mathbf{c}}^{\mathrm{APP}}$ from (10) labeled "mmse," $\tilde{\mathbf{c}}^{\mathrm{MM,APP}}$ from (18) labeled "minimax" and $\tilde{\mathbf{c}}^{\mathrm{CP,APP}}$ from (25) labeled "regret."



(a) (b)

Fig. 5. Equalization results for the channel in (36) for 150 random independent realization of $\{x_t\}$. Here, $\mathrm{SNR} = 10\ \mathrm{dB}$. Included algorithms are $\tilde{\mathbf{c}}_t$ from (8) labeled "mmse," $\tilde{\mathbf{c}}_t^{\mathrm{MM}}$ from (11) labeled "minimax" and $\tilde{\mathbf{c}}_t^{\mathrm{CP}}$ from (21) labeled "regret." (a) Sorted BERs for the 3rd iteration. (b) Sorted BERs for the 5th iteration.

$\mathbf{z}_t = [\bar{x}_{t-P+1}, \ldots, \bar{x}_t]^T$ in the test data mode after the first turbo iteration, $e_t = y_t - \tilde{\mathbf{f}}_t^T \mathbf{z}_t$ and μ is the learning rate. We emphasize that our methods are generic with respect to how the channel estimation is performed. At each iteration after the final channel estimates are constructed for all methods, the estimated channels are used to construct the equalizers. Note that we get a different channel estimate for each equalization method, since the soft information from the SISO decoder are different. Naturally the channel estimates, $\tilde{\mathbf{f}}$'s, are not error free. To apply the introduced robust equalization algorithms, we define uncertainty in the channel around the final channel estimate based on the (conditional) mean square estimation error given by

$$E\left[\|\tilde{\mathbf{f}} - \mathbf{f}\|^2 : \{\mathrm{LLR}_{a,t}^E\}\right] = \mathrm{tr}\left\{E\left[(\tilde{\mathbf{f}} - \mathbf{f})(\hat{\mathbf{f}} - \mathbf{f})^H : \{\mathrm{LLR}_{a,t}^E\}\right]\right\} \quad (39)$$

around the final channel estimate. The Wiener solution and the corresponding MSE for the time varying (39), as well as the limiting values, are given in [29] for converged filters. Note that while performing equalization using the soft information, the received signal can be written as $y_t = \mathbf{f}^T \mathbf{x}_t + n_t = \mathbf{f}^T \bar{\mathbf{x}}_t + \mathbf{f}^T(\mathbf{x}_t - \bar{\mathbf{x}}_t) + n_t$. Under independence assumptions and assuming convergence, the corresponding mean square estimation error in (39) can be approximated for stationary data [30, Ch. 6] as

$$\mathrm{tr}\left\{E\left[(\tilde{\mathbf{f}} - \mathbf{f})(\tilde{\mathbf{f}} - \mathbf{f})^H : \{\mathrm{LLR}_{a,t}^E\}\right]\right\} \approx m\mu \frac{E\left[\left\|\mathbf{f}^T(\mathbf{x}_t - \bar{\mathbf{x}}_t) + n_t\right\|^2 : \{\mathrm{LLR}_{a,t}^E\}\right]}{2} \quad (40)$$

at convergence. However, since (40) is time varying, i.e., the underlying process is not stationary, we further approximate (40) with

$$\frac{\mu m}{2n} \sum_{t=1}^{n} E\left[\left\|\mathbf{f}^T(\mathbf{x}_t - \bar{\mathbf{x}}_t) + n_t\right\|^2 : \{\mathrm{LLR}_{a,t}^E\}\right] = \frac{\mu m}{2n} \sum_{t=1}^{n} \left(\mathbf{f}^T \mathbf{Q}_t \mathbf{f} + \sigma_n^2\right). \quad (41)$$

We define the "practical radius" of the channel uncertainty set, δ, in terms of the square root of the approximate MSE expression in (41). This scaling would also reflect the effect of using the estimated $\tilde{\mathbf{f}}$ and σ_w^2 in (41). We emphasize that at each iteration, $\tilde{\mathbf{f}}$ and (41) are updated using the most recently available soft information to construct the linear MMSE and robust estimators. For the experiments, we use the channel (36) with $N_1 = 7, N_2 = 7, N = 15$ and $P = M = 5$ for all methods and SNR is set to 10 dB. The learning parameter in (38) is set to 5×10^{-4} in the training mode and to 5×10^{-3} in the data mode to assure convergence. We choose the length of the training data 512 and the test data 1024 for each packet of $\{x_t\}$. Given the channel, we generate 150 independent realizations of $\{x_t\}$ and then perform estimation and equalization using the introduced algorithms. In Fig. 5, we plot the sorted BERs with respect to independent trials. The sorted BERs for the third and fifth iterations are displayed in Fig. 5(a) and in Fig. 5(b), respectively. In these simulations, as in the previous examples, we observe that the worst case BER performance of the regret algorithm is worse than the minimax method but better than the plug-in MMSE. Both robust methods provide better performance compared to the plug-in MMSE method by limiting the worst-case BER through the iterations.

V. Conclusion

In this paper, we investigate robust linear turbo equalization problem when the coefficients of the underlying discrete time communication channel are not accurately known. In order to incorporate the uncertainty in the channel coefficients in the equalizer design, we study a minimax approach and a competitive approach, which are centered around certain MSE optimality criteria. For both approaches, the problem of obtaining the linear equalizer coefficients is posed as an SDP problem. We observe through simulations that the introduced methods improve over the plug-in MMSE estimators for the examples from [3] under different distortions and SNRs. The performance gains of the introduced algorithms become more apparent as the iteration count increases.

Appendix

Lemma 1: The first order linear approximation of $\mathbf{v}^T[\mathbf{Q}_t'^{-1} + \sigma_n^{-2}\mathbf{F}^H\mathbf{F}]^{-1}\mathbf{v}$ around the channel estimate $\tilde{\mathbf{f}}_t$ is given by

$$\mathbf{v}^T\left[\mathbf{Q}_t'^{-1} + \sigma_n^{-2}\mathbf{F}^H\mathbf{F}\right]^{-1}\mathbf{v} = \eta_t + \mathbf{df}^H\mathbf{g}_t + \\ + \mathbf{g}_t{}^T\mathbf{df} + O\left(\|\mathbf{df}\|^2\right).$$

Proof: The derivation of Lemma 1 is similar to the derivation of [7, Lemma 1]. The first order linear approximation is given by

$$\mathbf{v}^T\left[\mathbf{Q}_t'^{-1} + \sigma_n^{-2}\mathbf{F}^H\mathbf{F}\right]^{-1}\mathbf{v} \approx \underbrace{\mathbf{v}^T\left[\mathbf{Q}_t'^{-1} + \sigma_n^{-2}\widetilde{\mathbf{F}}_t^H\widetilde{\mathbf{F}}_t\right]^{-1}\mathbf{v}}_{\eta_t} \\ + 2\Re e\left\{\mathrm{tr}\left[\mathbf{T}_t^H\mathbf{dF}\right]\right\} \quad (42)$$

where $\mathbf{T}_t \triangleq \nabla_{\mathbf{F}}(\mathbf{v}^T[\mathbf{Q}_t'^{-1} + \sigma_n^{-2}\mathbf{F}^H\mathbf{F}]^{-1}\mathbf{v})|_{\mathbf{F}=\tilde{\mathbf{F}}_t}$ and $\mathbf{dF}$ is the convolution matrix generated from $\mathbf{df}$. The gradient term $\mathbf{T}_t$ is derived in [7, Lemma 1] as

$$\mathbf{T}_t = -\sigma_n^{-2}\widetilde{\mathbf{F}}_t\left(\mathbf{Q}_t' + \sigma_n^{-2}\widetilde{\mathbf{F}}_t^H\widetilde{\mathbf{F}}_t\right)^{-1}\mathbf{v}\mathbf{v}^T \\ \times\left(\mathbf{Q}_t' + \sigma_n^{-2}\widetilde{\mathbf{F}}_t^H\widetilde{\mathbf{F}}_t\right)^{-1}. \quad (43)$$

Using (43), we have

$$\begin{aligned} &\mathrm{tr}\left[\mathbf{T}_t^H\mathbf{dF}\right] \\ &= \mathrm{tr}\left[-\sigma_n^{-2}\left(\mathbf{Q}_t' + \sigma_n^{-2}\widetilde{\mathbf{F}}_t^H\widetilde{\mathbf{F}}_t\right)^{-1}\mathbf{v}\mathbf{v}^T\right. \\ &\qquad \left.\times\left(\mathbf{Q}_t' + \sigma_n^{-2}\widetilde{\mathbf{F}}_t^H\widetilde{\mathbf{F}}_t\right)^{-1}\widetilde{\mathbf{F}}_t^H\mathbf{dF}\right] \\ &= \mathrm{tr}\left[-\sigma_n^{-2}\mathbf{v}^T\left(\mathbf{Q}_t' + \sigma_n^{-2}\widetilde{\mathbf{F}}_t^H\widetilde{\mathbf{F}}_t\right)^{-1}\widetilde{\mathbf{F}}_t^H\mathbf{dF}\right. \\ &\qquad \left.\times\left(\mathbf{Q}_t' + \sigma_n^{-2}\widetilde{\mathbf{F}}_t^H\widetilde{\mathbf{F}}_t\right)^{-1}\mathbf{v}\right] \\ &= \mathrm{tr}\left[-\tilde{\mathbf{c}}_t^T\mathbf{dF}\left(\mathbf{Q}_t' + \sigma_n^{-2}\widetilde{\mathbf{F}}_t^H\widetilde{\mathbf{F}}_t\right)^{-1}\mathbf{v}\right] \\ &= \left[\mathbf{df}^T\underbrace{\left(-\widetilde{\mathbf{C}}_t\left(\mathbf{Q}_t' + \sigma_n^{-2}\widetilde{\mathbf{F}}_t^H\widetilde{\mathbf{F}}_t\right)^{-1}\mathbf{v}\right)}_{\mathbf{g}_t}\right] \end{aligned} \quad (44)$$

where the second line is due to the properties of the trace operation, the third line follows from (8), the fourth line follows since $\mathbf{dF}$ is a convolution matrix and $\widetilde{\mathbf{C}}_t$ is the convolution matrix constructed using $\tilde{\mathbf{c}}_t$. Then, using (44) in (42) we get

$$\mathbf{v}^T\left[\mathbf{Q}_t'^{-1} + \sigma_n^{-2}\mathbf{F}^H\mathbf{F}\right]^{-1}\mathbf{v} \approx \underbrace{\mathbf{v}^T\left[\mathbf{Q}_t'^{-1} + \sigma_n^{-2}\widetilde{\mathbf{F}}_t^H\widetilde{\mathbf{F}}_t\right]^{-1}\mathbf{v}}_{\eta_t} \\ + 2\Re e[\mathbf{df}^T\mathbf{g}_t].$$

This completes the proof of Lemma 1. □

Lemma 2: The inequality

$$\begin{bmatrix} \mathbf{Q} & \mathbf{S} \\ \mathbf{S}^H & \mathbf{R} \end{bmatrix} \geq 0 \quad (45)$$

where $\mathbf{Q} = \mathbf{Q}^H$, $\mathbf{R} = \mathbf{R}^H$, and $\mathbf{R} > 0$ is equivalent to

$$\mathbf{R} > 0, \quad \mathbf{Q} - \mathbf{S}\mathbf{R}^{-1}\mathbf{S}^H \geq 0. \quad (46)$$

Proof of Lemma 2: Assume $\mathbf{R} > 0$ and $\mathbf{M}_1 = \begin{bmatrix} \mathbf{Q} & \mathbf{S} \\ \mathbf{S}^H & \mathbf{R} \end{bmatrix} \geq 0$. Then, using the nonsingular matrix $\mathbf{T} = \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ -\mathbf{R}^{-1}\mathbf{S}^H & \mathbf{I} \end{bmatrix}$, one can establish the congruence transformation

$$\mathbf{M}_2 = \mathbf{T}^H\mathbf{M}_1\mathbf{T} = \begin{bmatrix} \mathbf{Q} - \mathbf{S}\mathbf{R}^{-1}\mathbf{S}^H & \mathbf{0} \\ \mathbf{0} & \mathbf{R} \end{bmatrix}. \quad (47)$$

Assuming $\mathbf{M}_1 \geq 0$ yields $\mathbf{M}_1$ and $\mathbf{M}_2$ to have the same inertia. Since by assumption $\mathbf{R} > 0$, we conclude that $\mathbf{Q} - \mathbf{S}\mathbf{R}^{-1}\mathbf{S}^H \geq 0$. □

Lemma 3: Given matrices $\mathbf{P}$, $\mathbf{Q}$, and $\mathbf{A}$ with $\mathbf{A} = \mathbf{A}^H$

$$\mathbf{A} \geq \mathbf{P}^H\mathbf{Z}\mathbf{Q} + \mathbf{Q}^H\mathbf{Z}^H\mathbf{P}, \quad \forall\|\mathbf{Z}\| \leq \alpha$$

if and only if there exists a $\lambda \geq 0$ such that

$$\begin{bmatrix} \mathbf{A} - \lambda\mathbf{Q}^H\mathbf{Q} & -\alpha\mathbf{P}^H \\ -\alpha\mathbf{P} & \lambda\mathbf{I} \end{bmatrix} \geq 0.$$

This lemma is from [5, Prop. 2]. □

REFERENCES

[1] C. Berrou, A. Glavieux, and P. Thitimajshima, "Near Shannon limit error-correcting coding and decoding: Turbo codes," in *Proc. Int. Conf. Commun.*, Geneva, Switzerland, 1993, p. 1064–1070.

[2] M. Tuechler, R. Koetter, and A. C. Singer, "Turbo equalization: Principles and new results," *IEEE Trans. Commun.*, vol. 50, no. 5, pp. 754–767, May 2002.

[3] J. Proakis, *Digital Communications*. New York: McGraw-Hill, 1995.

[4] S. S. Kozat and A. C. Singer, "Universal switching linear least squares prediction," *IEEE Trans. Signal Process.*, vol. 56, pp. 189–204, Jan. 2008.

[5] Y. C. Eldar and N. Merhav, "A competitive minimax approach to robust estimation and random parameters," *IEEE Trans. Signal Process.*, vol. 52, no. 7, pp. 1931–1946, Jul. 2004.

[6] N. Vucic and H. Boche, "Robust minimax equalization of imperfectly known frequency selective MIMO channels," in *Proc. 41st Asilomar Conf. Signals, Syst. Comput.*, Pacific Grove, CA, Nov. 2007.

[7] S. S. Kozat and A. T. Erdogan, "Competitive linear estimation under model uncertainties," *IEEE Trans. Signal Process.*, vol. 58, no. 4, pp. 2388–2393, Apr. 2010.

[8] S. A. Kassam and H. V. Poor, "Robust signal processing for communication systems," *IEEE Commun. Mag.*, vol. 21, no. 1, pp. 20–28, 1983.

[9] G. Moustakides and S. A. Kassam, "Minimax robust equalization for random signals through uncertain channels," in *Proc. Allerton Conf. Commun., Contr., Comput.*, 1982, pp. 945–954.

[10] S. Boyd, L. El Ghaoui, E. Feron, and V. Balakrishnan, *Linear Matrix Inequalities in System and Control Theory*. Philadelphia, PA: SIAM, 1994.

[11] J. Mattingley and S. Boyd, "Automatic code generation for real-time convex optimization," in *Convex Optimization in Signal Processing and Communications*, Y. Eldar and D. Palomar, Eds. Cambridge, U.K.: Cambridge Univ. Press, 2010, pp. 1–41.

[12] A. Glavieux, C. Laot, and J. Labat, "Turbo equalization over a frequency selective channel," in *Proc. Int. Symp. Turbo Codes*, Brest, France, 1997, pp. 96–102.

[13] D. Raphaeli and A. Saguy, "Linear equalizers for turbo equalization: A new optimizaiton criterion for determining the equalizer taps," in *Proc. Int. Symp. Turbo Codes*, Brest, France, 2000, pp. 371–374.

[14] Y. C. Eldar, A. Ben-Tal, and A. Nemirovski, "Linear minimax regret estimation of deterministic parameters with bounded data uncertainties," *IEEE Trans. Signal Process.*, vol. 52, no. 8, pp. 2177–2188, 2004.

[15] Y. C. Eldar, A. Ben-Tal, and A. Nemirovski, "Robust mean-squared error estimation in the presence of model uncertainties," *IEEE Trans. Signal Process.*, vol. 53, no. 1, pp. 168–181, 2005.

[16] A. T. Erdogan, B. Hassibi, and T. Kailath, "MIMO decision feedback equalization from an H^∞ perspective," *IEEE Trans. Signal Process.*, vol. 52, no. 3, pp. 734–745, Mar. 2004.

[17] B. Hassibi, A. T. Erdogan, and T. Kailath, "On linear H^∞ equalization of communication channels," *IEEE Trans. Signal Process.*, vol. 48, no. 11, pp. 3227–3231, Nov. 2000.

[18] A. T. Erdogan, B. Hassibi, and T. Kailath, "FIR H^∞ equalization," *Signal Process.*, vol. 81, no. 5, pp. 907–917, May 2001.

[19] B. Hassibi, A. T. Erdogan, and T. Kailath, "MIMO linear equalization with an H^∞ criterion," *IEEE Trans. Signal Process.*, vol. 54, no. 2, pp. 499–511, Feb. 2006.

[20] M. D. Nisar and W. Utschick, "Robust turbo equalization—A minimax perspective," in *Proc. 6th Int. Symp. Turbo Codes and Iterative Inf. Process. (ISTC)*, 2010, pp. 379–383.

[21] S. S Kozat, A. C. Singer, and G. Zeitler, "Universal piecewise linear prediction via context trees," *IEEE Trans. Signal Process.*, vol. 55, no. 7, pp. 3730–3745, 2007.

[22] S. S. Kozat and A. C. Singer, "Competitive prediction under additive noise," *IEEE Trans. Signal Process.*, vol. 57, no. 9, pp. 3698–3703, 2009.

[23] K. Ganti, "Interleaver design for modified circular simplex turbo block coded modulator," Master's thesis, Russ College of Eng. Technol., Ohio Univ., Athens, Nov. 2004.

[24] T. Kailath, A. H. Sayed, and B. Hassibi, *Linear Estimation*. Englewood Cliffs, NJ: Prentice-Hall, 2000.

[25] M. Tuechler, "Iterative equalization using priors," Master's thesis, Univ. Illinois, Urbana-Champaign, IL, 2000.

[26] R. Koetter, A. C. Singer, and M. Tuechler, "Turbo equalization: An iterative equalization and decoding technique for coded data transmission," *IEEE Signal Process. Mag.*, vol. 21, pp. 67–80, Jan. 2004.

[27] S. Haykin, *Adaptive Filter Theory*. Englewood Cliffs, NJ: Prentice-Hall, 2002.

[28] R. Otnes and M. Tüchler, "Iterative channel estimation for turbo equalization of time-varying frequency-selective channels," *IEEE Trans. Wireless Commun.*, vol. 3, no. 6, pp. 1918–1923, 2004.

[29] S. Song, A. C. Singer, and K.-M. Sung, "Soft input channel estimation for turbo equalization," *IEEE Trans. Signal Process.*, vol. 52, no. 10, pp. 2885–2894, Oct. 2004.

[30] S. Haykin, *Adaptive Filter Theory*. Englewood Cliffs, NJ: Prentice-Hall, 1996.

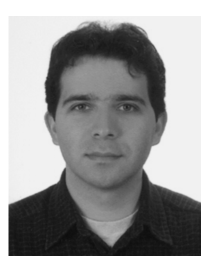
Nargiz Kalantarova was born on October 29, 1985, in Baku, Azerbaijan. She received the B.S. degree in electrical engineering and the B.S. degree in mathematics, both from Bogazici University, Istanbul, Turkey, in 2008.

Currently, she is a Research Assistant and is working toward the M.S. degree at the Electrical and Computer Engineering, Koc University, Istanbul. Her research interests include statistical signal processing, digital communications, and optimization.

Suleyman Serdar Kozat (A'10–M'11–SM'11) is a graduate of Ankara Fen Lisesi (Ankara Science High School). He received the B.S. degree with full scholarship and high honors from Bilkent University, Turkey. He received the M.S. and Ph.D. degrees in electrical and computer engineering from the Coordinated Science Laboratory, University of Illinois, Urbana-Champaign, in 2001 and 2004, respectively.

After graduation, he joined IBM Research T. J. Watson Research Center, Yorktown, NY, as a Research Staff Member with the Pervasive Speech Technologies Group, where he focused on problems related to statistical signal processing and machine learning. While doing pursuing the Ph.D. degree, he was also working as a Research Associate with the Cryptography and Anti-Piracy Group, Microsoft Research, Redmond, WA. He holds several patent applications for his works performed in IBM Research and Microsoft Research. Currently, he is an Assistant Professor with the Electrical and Electronics Engineering Department, Koc University, Istanbul, Turkey. He coauthored more than 40 papers in refereed high impact journals and conference proceedings and has several patent applications. Overall, his research interests include intelligent systems, adaptive filtering for smart data analytics, online learning, and machine learning algorithms for signal processing.

Dr. Kozat was awarded the Outstanding Young Researcher Award by the Turkish National Academy of Sciences in 2010, IBM Faculty Award by IBM Research in 2011, and holds Career Award by The Scientific Research Council of Turkey, 2009. He has been serving as an Associate Editor for the IEEE TRANSACTIONS ON SIGNAL PROCESSING and has been a reviewer for several IEEE TRANSACTIONS and Conferences. He has won several scholarships and medals in international and national science and math competitions.



Alper T. Erdogan (M'00) was born in Ankara, Turkey, in 1971. He received the B.S. degree from the Middle East Technical University, Ankara, Turkey, in 1993, and the M.S. and Ph.D. degrees from Stanford University, CA, in 1995 and 1999, respectively.

He was a Principal Research Engineer with the Globespan-Virata Corporation (formerly Excess Bandwidth and Virata Corporations) from September 1999 to November 2001. He joined the Electrical and Electronics Engineering Department, Koc University, Istanbul, Turkey, in January 2002, where he is currently an Associate Professor. His research interests include wireless, fiber and wireline communications, adaptive signal processing, optimization, system theory and control, and information theory.

Dr. Erdogan is the recipient of TUBITAK Career Award (2005), Werner Von Siemens Excellence Award (2007), and TUBA GEBIP Outstanding Young Scientist Award (2008). He currently serves as an Associate Editor for the IEEE TRANSACTIONS ON SIGNAL PROCESSING.