ADAPTIVE BITRATE MAXIMIZING TEQ DESIGN FOR DMT-BASED SYSTEMS

Koen Vanbleu, Geert Ysebaert, Gert Cuypers

Katholieke Universiteit Leuven
B-3001 Leuven, Belgium
{vanbleu,ysebaert,cuypers}@esat.kuleuven.ac.be

Geert Leus

Delft University of Technology
2628 CD Delft, The Netherlands
leus@cas.et.tudelft.nl

ABSTRACT

In a previous paper, we proposed a bitrate maximizing (BM) design criterion for the time-domain equalizer (TEQ) in a discrete multitone receiver. This BM-TEQ and the closely related BM per-group equalizers (PGEQ) get close to the performance of the so-called per-tone equalization (PTEQ). In this paper, we show that the BM-TEQ criterion, despite its nonlinear nature, is well suited for a recursive Levenberg-Marquardt (RLM) based design. This adaptive BM-TEQ also allows to track slow variations of the transmission channel and the noise. This RLM-based design uses the same second-order statistics (SOS) as the earlier presented recursive least-squares (RLS) based adaptive PTEQ and opens up a complete range of adaptive BM equalizers: from the computationally efficient RLS-based PTEQ with largest memory cost, over the RLM-based BM-PGEQ with intermediate memory cost, towards an RLM-based BM-TEQ with considerably smaller memory cost, but larger equalizer updating complexity.

1. INTRODUCTION

In a classical ADSL discrete multitone (DMT) receiver, a (real) T-tap channel shortening time domain equalizer (TEQ) is combined with 1-tap frequency domain equalizers (FEQ). Many TEQ design algorithms have been developed, but none of them truly optimizes bitrate (see [1] and references therein). In [2], an attractive alternative equalization scheme is proposed that always performs at least as well as - and usually better than - a TEQ based receiver while keeping complexity during data transmission at the same level. A complex *bitrate maximizing equalizer (BM-EQ)* is designed separately for each tone, hence the term *per-tone equalization* (PTEQ). The drawback of the PTEQ is its memory cost: N_aT complex equalizer taps (with N_a the number of active tones) need to be stored, instead of T taps in case of a TEQ.

In [1], we presented a nonlinear *bitrate maximizing (BM) TEQ* cost function based on an exact subchannel SNR model at the FEQ output. Instead, a *BM per-group equalization (PGEQ)* scheme can be devised [3]: the active tones are divided into N_g groups and each group is provided with a T-tap BM-EQ by solving the BM-TEQ cost function for that group. A BM-PGEQ with as few as 4 tone groups was found to perform close to the PTEQ in harsh environments with radio frequency interference (RFI) [3].

In ADSL, the TEQ is typically designed *offline*: the TEQ is computed during connection set-up and is then kept fixed during data transmission. However, a typical ADSL environment varies (slowly) with time. Moreover, the newest ADSL2 and ADSL2+ standards [4] support "seamless rate adaptation": the bitrate and bit allocation are adapted to varying line conditions. It is then desirable that the TEQ is *adaptive* and *tracks* changing conditions to keep the bitrate as large as possible. An adaptive equalizer can also be used for TEQ *design*: during connection set-up, a so-called medley signal of several seconds is transmitted for training.

So far, few adaptive *TEQ* designs have been presented [5, 6, 7]. A fast and reasonably cheap adaptive *PTEQ*, based on a recursive least-squares (RLS) algorithm, has been presented in [8]. It has a memory cost of N_aT complex equalizer taps and $\mathcal{O}(N_aT)$ second-order statistics (SOS) parameters at a computational load of $\mathcal{O}(N_aT)$ operations per update [8].

In this paper, we show that, despite its nonlinear and nonconvex nature, the BM-TEQ cost function also appears amenable to a *recursive* or *adaptive* design that is closely related to the RLS-based PTEQ. This adaptive BM-TEQ then opens up a complete range of *adaptive BM-EQs* (BM-TEQ, BM-PGEQ and PTEQ), all with the same SOS memory cost ($\mathcal{O}(N_aT)$), but each with a different number of equalizer taps and equalizer updating complexity. We refer to [3] for an extended version of this manuscript.

The authors would like to acknowledge Marc Moonen for his scientific contribution. They consider him to be a co-author of this paper, however he cannot be listed due to conference restrictions. This research work was carried out at the ESAT laboratory of the Katholieke Universiteit Leuven, in the frame of IUAP P5/22 and P5/11, GOA-MEFISTO-666, Research Project FWO nr.G.0196.02 and was partially sponsored by Alcatel-Bell. The scientific responsibility is assumed by its authors.

2. NOTATION AND KEY OBSERVATIONS

Here, we introduce the notation and some important basic equalities that will be applied furtheron. $\mathcal{S}_a$ is the set of N_a active tones; n is the tone index. N is the (I)FFT size; $\mathcal{F}_{\mathcal{S}_a}$ is a submatrix of the DFT matrix with the N_a active tone rows $\mathcal{S}_a$; the n-th DFT row is $\mathcal{F}_n$. $\mathbf{w}$ is the time-domain equalizer (TEQ, T taps); in the derivations, we assume a complex TEQ for reasons of conciseness. $\mathbf{D}$ is the $N_a \times 1$ vector of FEQs; D_n is the FEQ for tone n. $\boldsymbol{\theta} = \begin{bmatrix} \mathbf{w}^H & \mathbf{D}^H \end{bmatrix}^H$ and $\boldsymbol{\theta}_n = \begin{bmatrix} \mathbf{w}^H & D_n^* \end{bmatrix}^H$ are joint TEQ-FEQ parameter vectors. A tilde over a variable distinguishes frequency-domain symbols from time-domain symbols. k is the DMT symbol index. The k-th $N_a \times 1$ transmitted DMT symbol vector is $\tilde{\mathbf{x}}_k$; the symbol on tone n is $\tilde{x}_{k,n}$ and has a variance $\sigma^2_{n,\tilde{x}} = \mathcal{E}\left\{|\tilde{x}_{k,n}|^2\right\}$; $\hat{\tilde{x}}_{k,n}$ is the FEQ output. $\mathbf{I}_m$ is the $m \times m$ identity matrix. $\mathbf{a} \odot \mathbf{b}$ is the pointwise multiplication of $\mathbf{a}$ and $\mathbf{b}$. $\mathrm{diag}(\mathbf{a})$ is a diagonal matrix with $\mathbf{a}$ on the diagonal.

A *first key observation* exploits the associativity property in:

$$\hat{\tilde{\mathbf{x}}}_k = \mathbf{D} \odot \tilde{\mathbf{y}}_{k,\mathbf{w}} = \mathbf{D} \odot \mathcal{F}_{\mathcal{S}_a} \underbrace{(\mathbf{Y}_k \mathbf{w})}_{\mathbf{y}_{k,\mathbf{w}}} = \mathbf{D} \odot \underbrace{(\mathcal{F}_{\mathcal{S}_a} \mathbf{Y}_k)}_{\tilde{\mathbf{Y}}_k} \mathbf{w}$$

$$\hat{\tilde{x}}_{k,n} = D_n \tilde{y}_{k,n,\mathbf{w}} = D_n \mathcal{F}_n \underbrace{(\mathbf{Y}_k \mathbf{w})}_{\mathbf{y}_{k,\mathbf{w}}} = D_n \underbrace{(\mathcal{F}_n \mathbf{Y}_k)}_{\tilde{\mathbf{y}}_{k,n}} \mathbf{w} \qquad (1)$$

where $\mathbf{Y}_k$ is an $N \times T$ Toeplitz matrix of received samples with $\begin{bmatrix} y_{k,0} & \cdots & y_{k,-T+1} \end{bmatrix}$ on the first row and $\begin{bmatrix} y_{k,0} & \cdots & y_{k,N-1} \end{bmatrix}^T$

in the first column. It says (both for all tones and tone n) that the DFT of the convolution of the k-th DMT symbol and the TEQ, $\mathbf{Y}_k\mathbf{w}$, is equal to a linear combination $\mathbf{w}$ of the T outputs of a sliding DFT of the unequalized k-th DMT symbol, $\tilde{\mathbf{Y}}_k = \mathcal{F}_{\mathcal{S}_a}\mathbf{Y}_k$. A *second key observation* states that the sliding DFT output $\tilde{\mathbf{Y}}_k$ can be computed efficiently (based on 1 FFT $\tilde{y}_{k,n} = \tilde{\mathbf{y}}_{k,n}[1] = \mathcal{F}_n \left[\begin{array}{ccc} y_{k,0} & \cdots & y_{k,N-1} \end{array} \right]^T$, $T-1$ differences $\Delta\mathbf{y}_k[t] = y_{k,-T+t} - y_{k,-T+N+t}$, $1 \leq t \leq T-1$ and $N_a(T-1)$ recursions) using the recursion in

$$\tilde{\mathbf{y}}_{k,n} = \underbrace{\left[\begin{array}{cc} \Delta\mathbf{y}_k & \tilde{y}_{k,n} \end{array} \right]}_{\mathbf{z}_{k,n}} \underbrace{\left[\begin{array}{cccc} 0 & \cdots & 0 & 1 \\ \vdots & \iddots & \iddots & \alpha_n \\ 0 & 1 & \iddots & \vdots \\ 1 & \alpha_n & \cdots & \alpha_n^{T-1} \end{array} \right]}_{\mathbf{P}_n} \quad (2)$$

with $\alpha_n = \frac{1}{\sqrt{N}} e^{(-j2\pi(n-1)/N)}$. Both key observations underlie the RLS-based PTEQ [2, 8] and the here developed adaptive BM-TEQ. Based on the key observations, we obtain the following SOS definitions, used throughout the paper:

$$\mathcal{E}\left\{|\tilde{y}_{k,n,\mathbf{w}}|^2\right\} = \mathbf{w}^H \mathcal{E}\left\{\tilde{\mathbf{y}}_{k,n}^H \tilde{\mathbf{y}}_{k,n}\right\} \mathbf{w} = \mathbf{w}^H \Sigma^2_{n,\tilde{\mathbf{y}}} \mathbf{w} \quad (3)$$

$$\mathcal{E}\left\{\tilde{y}^*_{k,n,\mathbf{w}} \tilde{\mathbf{y}}_{k,n}\right\} = \mathbf{w}^H \mathcal{E}\left\{\tilde{\mathbf{y}}_{k,n}^H \tilde{\mathbf{y}}_{k,n}\right\} = \mathbf{w}^H \Sigma^2_{n,\tilde{\mathbf{y}}} \quad (4)$$

$$\mathcal{E}\left\{\tilde{x}^*_{k,n} \tilde{y}_{k,n,\mathbf{w}}\right\} = \mathcal{E}\left\{\tilde{x}^*_{k,n} \tilde{\mathbf{y}}_{k,n}\right\} \mathbf{w} = \Sigma_{n,\tilde{x}\tilde{\mathbf{y}}} \mathbf{w} \quad (5)$$

$$\Sigma^2_{n,\tilde{\mathbf{y}}} = \mathbf{P}_n^H \mathcal{E}\left\{\mathbf{z}_{k,n}^H \mathbf{z}_{k,n}\right\} \mathbf{P}_n = \mathbf{P}_n^H \Sigma^2_{n,\mathbf{z}} \mathbf{P}_n \quad (6)$$

3. A JOINT BM-TEQ-FEQ CRITERION

In [1], we introduced the nonlinear BM-TEQ criterion:

$$\arg\max_{\mathbf{w}} b_{DMT} = \arg\min_{\mathbf{w}} \sum_{n \in \mathcal{S}_a} \log\left(\frac{\mathbf{w}^H \mathbf{B}_n \mathbf{w}}{\mathbf{w}^H \mathbf{A}_n \mathbf{w}}\right) \quad (7)$$

where the tone-dependent matrices $\mathbf{A}_n$ and $\mathbf{B}_n$ are independent of $\mathbf{w}$ and depend on the SOS $\sigma^2_{n,\tilde{x}}$, $\Sigma^2_{n,\tilde{\mathbf{y}}}$ and $\Sigma_{n,\tilde{x}\tilde{\mathbf{y}}}$ defined in (3-5). This *BM-TEQ-only* criterion follows from a constrained nonlinear optimization problem in the joint parameter vector $\boldsymbol{\theta}$, which is the starting point for a new *joint BM-TEQ-FEQ* criterion:

$$\max_{\boldsymbol{\theta}} \sum_{n \in \mathcal{S}_a} \log_2\left(1 + \frac{\mathrm{SNR}_{n,\boldsymbol{\theta}_n}}{\Gamma_n}\right) \quad (8)$$

$$\text{with } \mathrm{SNR}_{n,\boldsymbol{\theta}_n} = \frac{\sigma^2_{n,\tilde{x}}}{\mathcal{E}\left\{|\tilde{e}_{k,n,\boldsymbol{\theta}_n}|^2\right\}} = \frac{\sigma^2_{n,\tilde{x}}}{\mathcal{E}\left\{|D_n \tilde{\mathbf{y}}_{k,n}\mathbf{w} - \tilde{x}_{k,n}|^2\right\}} \quad (9)$$

$$\text{subject to } D_n = \frac{\mathcal{E}\left\{|\tilde{x}_{k,n}|^2\right\}}{\mathcal{E}\left\{\tilde{x}^*_{k,n}\tilde{y}_{k,n,\mathbf{w}}\right\}} = \frac{\sigma^2_{n,\tilde{x}}}{\Sigma_{n,\tilde{x}\tilde{\mathbf{y}}}\mathbf{w}}, \forall n \in \mathcal{S}_a \quad (10)$$

i.e., maximizing (over $\boldsymbol{\theta}$) the number of bits per DMT symbol (8), subject to the use of unbiased MMSE FEQs (10), which render the subchannel SNR model in (9) exact. A constrained optimization criterion is typically restated as a Lagrangian cost minimization:

$$J_{\mathrm{L}}(\boldsymbol{\theta},\boldsymbol{\lambda}) = \sum_{n \in \mathcal{S}_a} \left[\log\left(1 + \frac{\mathrm{SNR}_{n,\boldsymbol{\theta}_n}}{\Gamma_n}\right)^{-1} - \Re\left\{\lambda_n^* \left(D_n \mathcal{E}\left\{\tilde{x}^*_{k,n}\tilde{y}_{k,n,\mathbf{w}}\right\} - \sigma^2_{n,\tilde{x}}\right)\right\}\right] \quad (11)$$

with λ_n the N_a Lagrange parameters and where the $\mathcal{R}$-operator ensures a real constraint term (which only matters for a real TEQ). Setting the gradient w.r.t. D_n, $\nabla_{D_n} J_{\mathrm{L}}$, to zero, i.e.,

$$\gamma_{n,\boldsymbol{\theta}_n} \mathcal{E}\left\{\tilde{y}^*_{k,n,\mathbf{w}} \tilde{e}_{k,n,\boldsymbol{\theta}_n}\right\} - \lambda_n \mathcal{E}\left\{\tilde{y}^*_{k,n,\mathbf{w}} \tilde{x}_{k,n}\right\} = 0 \quad (12)$$

where $\gamma_{n,\boldsymbol{\theta}_n} = \frac{\mathrm{SNR}^2_{n,\boldsymbol{\theta}_n}}{\sigma^2_{n,\tilde{x}}(\mathrm{SNR}_{n,\boldsymbol{\theta}_n} + \Gamma_n)}$ is a tone-dependent weight, results in $\lambda_n = \frac{\gamma_{n,\boldsymbol{\theta}_n}}{\mathrm{SNR}_{n,\boldsymbol{\theta}_n}}$ [3]. Combined with the gradient w.r.t. $\mathbf{w}$, $\nabla_{\mathbf{w}} J_{\mathrm{L}}$:

$$\sum_{n \in \mathcal{S}_a} \gamma_{n,\boldsymbol{\theta}_n} D_n^* \mathcal{E}\left\{\tilde{\mathbf{y}}_{k,n}^H \tilde{e}_{k,n,\boldsymbol{\theta}_n}\right\} - \lambda_n D_n^* \mathcal{E}\left\{\tilde{\mathbf{y}}_{k,n}^H \tilde{x}_{k,n}\right\} \quad (13)$$

the gradient of J_{L} w.r.t. $\boldsymbol{\theta}$ can be written compactly as:

$$\nabla_{\boldsymbol{\theta}} J_{\mathrm{L}} = \left[\begin{array}{c} \nabla_{\mathbf{w}} J_{\mathrm{L}} \\ \nabla_{\mathbf{D}} J_{\mathrm{L}} \end{array}\right] = \mathcal{E}\left\{\check{\mathbf{Y}}_{k,\boldsymbol{\theta}}^H \mathrm{diag}\left(\boldsymbol{\gamma}_{\boldsymbol{\theta}}\right) \check{\mathbf{e}}_{k,\boldsymbol{\theta}}\right\} \quad (14)$$

where the $N_a \times 1$ vector $\boldsymbol{\gamma}_{\boldsymbol{\theta}}$ has entries $\gamma_{n,\boldsymbol{\theta}_n}$, and where the $N_a \times (T + N_a)$ matrix $\check{\mathbf{Y}}_{k,\boldsymbol{\theta}}$ and $N_a \times 1$ vector $\check{\mathbf{e}}_{k,\boldsymbol{\theta}}$ are defined as:

$$\check{\mathbf{Y}}_{k,\boldsymbol{\theta}} = \left[\begin{array}{cc} \mathrm{diag}\left(\mathbf{D}\right)\tilde{\mathbf{Y}}_k & \mathrm{diag}(\tilde{\mathbf{Y}}_k\mathbf{w}) \end{array}\right] \quad (15)$$

$$\check{\mathbf{e}}_{k,\boldsymbol{\theta}} = \hat{\tilde{\mathbf{x}}}_k - \boldsymbol{\rho}_{\boldsymbol{\theta}}^{-2} \odot \tilde{\mathbf{x}}_k = \mathbf{D} \odot \left(\tilde{\mathbf{Y}}_k\mathbf{w}\right) - \boldsymbol{\rho}_{\boldsymbol{\theta}}^{-2} \odot \tilde{\mathbf{x}}_k \quad (16)$$

The vector $\boldsymbol{\rho}_{\boldsymbol{\theta}}^{-2}$ is short-hand notation for a vector with as entries the inverse of $\rho^2_{n,\boldsymbol{\theta}_n} = 1 + \frac{1}{\mathrm{SNR}_{n,\boldsymbol{\theta}_n}}$ (valued between 0 and 1, but typically close to 1 and accounting for the use of unbiased MMSE FEQs [1], [3]); $\check{\mathbf{e}}_{k,\boldsymbol{\theta}}$ in (16) is a vector of errors between the FEQ outputs $\hat{\tilde{x}}_{k,n}$ (1) and virtual symbols $\rho^{-2}_{n,\boldsymbol{\theta}_n}\tilde{x}_{k,n}$. Replacing (14) with an exponentially weighted time-average over K DMT symbols results in a generalized version (with weighting $\boldsymbol{\gamma}_{\boldsymbol{\theta}}$ and multiple-error vector $\check{\mathbf{e}}_{k,\boldsymbol{\theta}}$) of the orthogonality condition that is typically encountered when solving least-squares (LS) problems:

$$\nabla_{\boldsymbol{\theta}} J_{K,\mathrm{NL-WLS}} = \sum_{k=1}^{K} \lambda^{K-k} \check{\mathbf{Y}}_{k,\boldsymbol{\theta}}^H \mathrm{diag}\left(\boldsymbol{\gamma}_{\boldsymbol{\theta}}\right) \check{\mathbf{e}}_{k,\boldsymbol{\theta}} = \mathbf{0} \quad (17)$$

The weights $\boldsymbol{\gamma}_{\boldsymbol{\theta}}$ assume knowledge of the optimal $\mathrm{SNR}_{n,\boldsymbol{\theta}_n}$ (9) at time K, hence the optimal TEQ, FEQs and estimates of $\Sigma^2_{n,\tilde{\mathbf{y}}}$ and $\Sigma_{n,\tilde{x}\tilde{\mathbf{y}}}$. Adopting the ideas from [11, 12] to solve weighted LS problems with data and parameter dependent weights, $\boldsymbol{\gamma}_{\boldsymbol{\theta}_n}$ is replaced by an *instantaneous a priori* estimate, i.e., based on the *previous* parameter estimates $\mathbf{w}_{K-1}$ and $D_{K-1,n}$:

$$\hat{\gamma}_{K,n,\boldsymbol{\theta}_{K-1,n}} = \frac{\widehat{\mathrm{SNR}}^2_{K,n,\boldsymbol{\theta}_{K-1,n}}}{\sigma^2_{n,\tilde{x}}\left(\widehat{\mathrm{SNR}}_{K,n,\boldsymbol{\theta}_{K-1,n}} + \Gamma_n\right)} \quad (18)$$

$$\text{with } \widehat{\mathrm{SNR}}_{K,n,\boldsymbol{\theta}_{K-1,n}} = \frac{\sigma^2_{n,\tilde{x}}}{|D_{K-1,n}\tilde{\mathbf{y}}_{K,n}\mathbf{w}_{K-1} - \tilde{x}_{K,n}|^2}$$

$\widehat{\mathrm{SNR}}_{K,n,\boldsymbol{\theta}_{K-1,n}}$ is also used to approximate $\rho^{-2}_{n,\boldsymbol{\theta}_n}\tilde{x}_{k,n}$ in (16). With these approximations (18), the gradient (17) then also applies to the following *nonlinear weighted least-squares* (NL-WLS) problem with varying weights $\hat{\boldsymbol{\gamma}}_{K,\boldsymbol{\theta}_{K-1}}$:

$$J_{K,\mathrm{NL-WLS}}(\boldsymbol{\theta}) = \sum_{k=1}^{K} \lambda^{(K-k)} \left\| \sqrt{\hat{\boldsymbol{\gamma}}_{K,\boldsymbol{\theta}_{K-1}}} \odot \check{\mathbf{e}}_{k,\boldsymbol{\theta}} \right\|^2 \quad (19)$$

where $\check{\mathbf{e}}_{k,\boldsymbol{\theta}}$ and, hence, $J_{K,\mathrm{NL-WLS}}(\boldsymbol{\theta})$ depend nonlinearly on $\boldsymbol{\theta}$. The simulations of Section 5, using the adaptive algorithm developed in Section 4, confirm that the NL-WLS joint BM-TEQ-FEQ cost function (19) achieves the same performance as the BM-TEQ-only (7), despite the introduced approximations (18). The appearance of a least-squares (LS) problem does not come as a surprise, despite the nonlinear cost function (7): the denominator of SNR_n (9) is equal to the mean-square error (MSE) at the FEQ output; an MSE criterion is often adopted when designing an adaptive filter and naturally leads to a (linear or nonlinear) LS problem.

4. AN ADAPTIVE JOINT BM-TEQ-FEQ ALGORITHM

In this section, we solve the NL-WLS cost function (19) recursively (or adaptively) at each time k, based on a *recursive Levenberg-Marquardt* (RLM)[1] updating of the $(N_a + T) \times 1$ joint TEQ-FEQ parameter vector θ [9, 10]. From (19), one easily obtains the joint TEQ-FEQ updating rule:

$$\theta_k \quad \leftarrow \quad \theta_{k-1} - \mathbf{R}^{-1}_{k,\theta_{k-1},\delta_k} \mathbf{g}_{k,\theta_{k-1}} \tag{20}$$

where the gradient estimate $\mathbf{g}_{k,\theta}$ and the regularized approximate Hessian estimate $\mathbf{R}_{k,\theta,\delta}$ (both typical of an RLM algorithm) are[2]

$$\mathbf{g}_{k,\theta} = \check{\mathbf{Y}}^H_{k,\theta} \mathrm{diag}\left(\hat{\boldsymbol{\gamma}}_{k,\theta_{k-1}}\right) \check{\mathbf{e}}_{k,\theta} \tag{21}$$

$$\mathbf{R}_{k,\theta,\delta} = \sum_{\kappa=1}^{k} \check{\mathbf{Y}}^H_{\kappa,\theta} \mathrm{diag}\left(\hat{\boldsymbol{\gamma}}_{k,\theta_{k-1}}\right) \check{\mathbf{Y}}_{\kappa,\theta} + \delta \mathbf{I}_{T+N_a} \tag{22}$$

After each update, the TEQ is normalized (and the FEQs are scaled accordingly) to solve the parameter ambiguity in (19). $\mathbf{R}_{k,\theta,\delta}$ in (22) is a 2×2 block autocorrelation matrix estimate:

$$\mathbf{R}_{k,\theta,\delta} \quad = \quad \begin{bmatrix} \mathbf{E}_{k,\mathbf{D}} & \mathbf{F}_{k,\theta} \\ \mathbf{F}^H_{k,\theta} & \mathbf{G}_{k,\mathbf{w},\mathrm{diag}} \end{bmatrix} \tag{23}$$

The submatrices $\mathbf{E}_{k,\mathbf{D}}$ $(T \times T)$, $\mathbf{F}_{k,\theta}$ $(T \times N_a)$ and $\mathbf{G}_{k,\mathbf{w},\mathrm{diag}}$ ($N_a \times N_a$ and diagonal) follow from (3), (4)[3]:

$$\mathbf{E}_{k,\mathbf{D}} \quad = \quad \sum_{n \in \mathcal{S}_a} \hat{\gamma}_{k,n,\theta_{k-1,n}} |D_n|^2 \Sigma^2_{k,n,\tilde{\mathbf{y}}} + \delta \mathbf{I}_T \tag{24}$$

$$\mathbf{F}^H_{k,\theta} = \left[\begin{array}{c} \vdots \\ \left[\hat{\gamma}_{k,n,\theta_{k-1,n}} D_n \left(\mathbf{w}^H \Sigma^2_{k,n,\tilde{\mathbf{y}}}\right)\right] \\ \vdots \end{array} \right] \Big\updownarrow N_a \text{ rows} \tag{25}$$

$$\mathbf{G}_{k,\mathbf{w},\mathrm{diag}} = \mathrm{diag}\left(\left[\cdots \left[\hat{\gamma}_{k,n,\theta_{k-1,n}} \left(\mathbf{w}^H \Sigma^2_{k,n,\tilde{\mathbf{y}}} \mathbf{w}\right) + \delta\right] \cdots\right]\right) \tag{26}$$

Making use of (21), the block matrix inverse $\mathbf{R}^{-1}_{\theta}$:

$$\left[\begin{array}{c|c} \mathbf{Q}^{-1}_{\theta} & -\mathbf{Q}^{-1}_{\theta} \mathbf{F}_{\theta} \mathbf{G}^{-1}_{\mathbf{w},\mathrm{diag}} \\ -\mathbf{G}^{-1}_{\mathbf{w},\mathrm{diag}} \mathbf{F}^H_{\theta} \mathbf{Q}^{-1}_{\theta} & \mathbf{G}^{-1}_{\mathbf{w},\mathrm{diag}} \left(\mathbf{I}_{N_a} + \mathbf{F}^H_{\theta} \mathbf{Q}^{-1}_{\theta} \mathbf{F}_{\theta} \mathbf{G}^{-1}_{\mathbf{w},\mathrm{diag}}\right) \end{array} \right] \tag{27}$$

with

$$\mathbf{Q}_{\theta} = \mathbf{E}_{\mathbf{D}} - \mathbf{F}_{\theta} \mathbf{G}^{-1}_{\mathbf{w},\mathrm{diag}} \mathbf{F}^H_{\theta} \tag{28}$$

and the definition of $\check{\mathbf{Y}}_{k,\theta}$ (15), we obtain a stochastic-Newton-like updating equation for $\mathbf{w}$ in (20):

$$\mathbf{w}_k \leftarrow \mathbf{w}_{k-1} \underbrace{- \mathbf{Q}^{-1}_{k,\theta_{k-1}} \bar{\mathbf{Y}}^H_{k,\theta_{k-1}} \left(\hat{\boldsymbol{\gamma}}_{k,\theta_{k-1}} \odot \check{\mathbf{e}}_{k,\theta_{k-1}}\right)}_{\Delta \mathbf{w}_k} \tag{29}$$

with $\mathbf{Q}_{k,\theta}$ and $\check{\mathbf{e}}_{k,\theta_{k-1}}$ defined in (28) and (16), respectively, and

$$\bar{\mathbf{Y}}^H_{k,\theta} = \tilde{\mathbf{Y}}^H_k \mathrm{diag}\left(\mathbf{D}\right)^* - \mathbf{F}_{\theta} \mathbf{G}^{-1}_{\mathbf{w},\mathrm{diag}} \mathrm{diag}\left(\tilde{\mathbf{y}}_{k,\mathbf{w}}\right)^* \tag{30}$$

Thanks to the block structure and the diagonal submatrix $\mathbf{G}_{k,\mathbf{w},\mathrm{diag}}$, only the inverse of the full $T \times T$ matrix $\mathbf{Q}_{k,\theta}$ needs to be computed. The FEQ updating in (20) reduces to

$$\mathbf{D}_k \leftarrow \mathbf{D}_{k-1} - \mathbf{G}^{-1}_{k,\mathbf{w},\mathrm{diag}} \Big[\hat{\boldsymbol{\gamma}}_{k,\theta_{k-1}} \odot \tilde{\mathbf{y}}^*_{k,\mathbf{w}} \odot \check{\mathbf{e}}_{k,\theta_{k-1}} + \mathbf{F}^H_{k,\theta_{k-1}} \Delta \mathbf{w}_k\Big] \tag{31}$$

Despite the approximate Hessian $\mathbf{R}_{k,\theta_{k-1},\delta_k}$ having size $(T + N_a) \times (T + N_a)$, only the SOS estimates $\Sigma^2_{k,n,\tilde{\mathbf{y}}}$ are required to construct it. These are the exact same SOS as needed for the (square-root) RLS-based PTEQ [8]. Also here, as in [8], the SOS memory cost can be further reduced by exploiting the second key observation (2), which gives rise to (6): storing the upper-triangular Cholesky factor $\mathbf{L}_{k,n}$ of $\Sigma^2_{k,n,\mathbf{z}}$ instead of $\Sigma^2_{k,n,\tilde{\mathbf{y}}}$ reduces the total SOS memory to $\frac{T(T-1)}{2}$ real coefficients for the $T-1$ first, real, tone-independent columns of $\mathbf{L}_{k,n}$ (which should only be stored and updated once) plus $(2T-1)N_a$ coefficients for the tone-dependent complex last column of $\mathbf{L}_{k,n}$. In contrast to the RLS-based PTEQ, the SOS updating is computationally not the most demanding part of the adaptive joint BM-TEQ-FEQ. The complexity is rather dominated by $\mathcal{O}(N_a T^2)$ computations for the computation of $\mathbf{Q}_{k,\theta}$, which requires $\Sigma^2_{k,n,\tilde{\mathbf{y}}}$, rather than $\Sigma^2_{k,n,\mathbf{z}}$. Avoiding the transformation (6), which can be done efficiently with $\mathcal{O}(N_a T^2)$ computations [3], is an exclusive, computational advantage of the adaptive PTEQ over the BM-TEQ and BM-PGEQ: $\mathcal{O}(N_a T)$ computations suffice for the RLS-based PTEQ updating.

Table 1 compares the memory cost and computational complexity of the equalizer filtering and updating for the BM-TEQ, BM-PGEQ (with N_g tone groups) and PTEQ. It includes the dominant terms in memory cost and computational complexity of the equalizer filtering and updating in equivalent number of real coefficients and multiplications. Memory and complexity figure estimates are also included for $N_a = 224$, $T = 16 \rightarrow 32$ and $N_g = 4$. The SOS memory cost is the same and the equalizer filtering cost highly comparable for all BM-EQs. The PTEQ needs a large number of equalizer taps but has a computationally advantageous equalizer updating. We refer to [3] for a detailed discussion. In an application such as ADSL, equalizer updating (for design and tracking) can typically be done at a rate that is slower than the equalizer filtering rate of 4kHz: given that around 16000 DMT training symbols are available during connection set-up while convergence occurs within 200 to 300 symbols (see Section 5), then the updating speed can be decimated with a factor 50 to 80, resulting in 3.4 to 10.7 real multiplications per second for the PTEQ and 28.5 to 168 real multiplications per second for the BM-TEQ.

In [3], we discuss some further refinements to the algorithm.

- In case of an RLM algorithm, the choice of the exponential weighting factor for estimating the SOS does not only influence the tracking speed and estimation accuracy, but also the convergence speed [9]. Therefore, we increase λ from 0.9 (fast tracking during first 400 updates) over 0.95 (next 400 updates) to 0.99 (after 800 updates for high accuracy).
- The diagonal of $\mathbf{R}_{k,\theta_{k-1},\delta_k}$, especially $\mathbf{G}_{k,\mathbf{w}_{k-1},\mathrm{diag}}$, is not constant and can have a large dynamic range. This influences the condition of the approximate Hessian badly. We suggest a (cheap) energy normalization through a diagonal transformation of $\mathbf{R}_{k,\theta_{k-1},\delta_k}$ and $\mathbf{g}_{k,\theta_{k-1}}$. This reduces the diagonal elements of $\mathbf{R}_{k,\theta_{k-1},\delta_k}$ to $1 + \delta_k$.
- Both convergence speed and stability are affected by a suitable choice of the regularization parameter δ_k in (22): a too small δ_k could cause the RLM algorithm to go unstable, while a too large δ_k could induce slow convergence in directions of the parameter space that correspond to small eigenvalues. The parameter δ_k should be adapted, as the

[1] As with the PTEQ, stochastic gradient algorithms are found to converge too slowly.

[2] We omit the dependence on the update index of θ_{k-1} and δ_k for conciseness.

[3] The subscript k distinguishes SOS estimates $\Sigma^2_{k,n,\tilde{\mathbf{y}}}$ from the true SOS $\Sigma^2_{n,\tilde{\mathbf{y}}}$.

	Memory cost				Computational cost per update			
	EQ taps (K $= 10^3$)		SOS coeffs ($\times 10^3$)		Filtering (excl. FFT) ($\times 10^3$)		Updating ($\times 10^3$)	
real joint BM-TEQ-FEQ	$T + 2N_a$	$464 \rightarrow 480$	$2N_aT$	$7 \rightarrow 15$	$6N_aT$ or NT	$22 \rightarrow 43$ or $8 \rightarrow 16$	$8.5N_aT^2$	$570 \rightarrow 2100$
real joint BM-PGEQ-FEQ	$N_gT + 2N_a$	$512 \rightarrow 528$	$2N_aT$	$7 \rightarrow 15$	$6N_aT$	$22 \rightarrow 43$	$8.5N_aT^2$	$570 \rightarrow 2110$
complex PTEQ	$2N_aT$	$7\text{K} \rightarrow 14\text{K}$	$2N_aT$	$7 \rightarrow 15$	$2N_aT$	$8 \rightarrow 15$	$18N_aT$	$68 \rightarrow 134$

Table 1. Dominant terms in memory cost and computational complexity of equalizer filtering and updating in equivalent number of real coefficients and multiplications. Memory and complexity figure estimates for $N_a = 224$, $T = 16 \rightarrow 32$ and $N_g = 4$ tone groups.

condition of the first term in (22) depends on the estimates θ_{k-1} and hence changes during convergence. Based on the ideas in [10], we propose an adaptation rule for δ_k that is based on the ratio of instantaneous estimates of the actual and predicted cost reduction of $J_{k,\mathrm{NL-WLS}}$ (19).

5. SIMULATIONS

We include simulations for the downstream CSA4 loop (tones 33 to 256) with moderate front-end filtering.The noise is a superposition of AWG noise at -140dBm/Hz, residual echo and near-end crosstalk from 24 ADSL disturbers. We also include the harsh case of severe RFI (7 RFIs with powers between -30 and -50dBm) which can be treated effectively with the PTEQ. Further specifications and an extensive simulation section are included in [3].

In [1], the BM-TEQ was found to approach the PTEQ performance very closely, despite possible local minima of the non-convex BM-TEQ cost function. This result is confirmed here, when comparing the RLS-based PTEQ with the here presented RLM-based BM-TEQ. Figure 1 shows bitrate convergence curves (for $T = 32$) as a function of the update index. The PTEQ (with $\lambda = 0.999$) is compared with the BM-TEQ and a BM-PGEQ with 4 equally sized tone groups. If no RFI is present (thick lines), they all reach the same bitrate of 8.4Mbps; the PTEQ and BM-TEQ curve almost coincide, while the BM-PGEQ converges the fastest in around 100 updates. If RFI is present (thin lines), the BM-TEQ achieves 6.8Mbps, i.e., less than 300kbps (or only 4%) below the PTEQ bitrate; the BM-PGEQ fills the gap in convergence speed and bitrate between the PTEQ and the BM-TEQ. The convergence time can be decreased by initializing the RLM algorithm with a cheaply computed suboptimal TEQ (e.g., an MMSE-TEQ, see the thick dotted line), instead of $\mathbf{w}_0 = \begin{bmatrix} 1 & 0 & \cdots & 0 \end{bmatrix}^T$ elsewhere. The same thick dotted line also shows that the adaptive BM-TEQ is capable of tracking the disappearance of 2 RFIs at time instant 500.

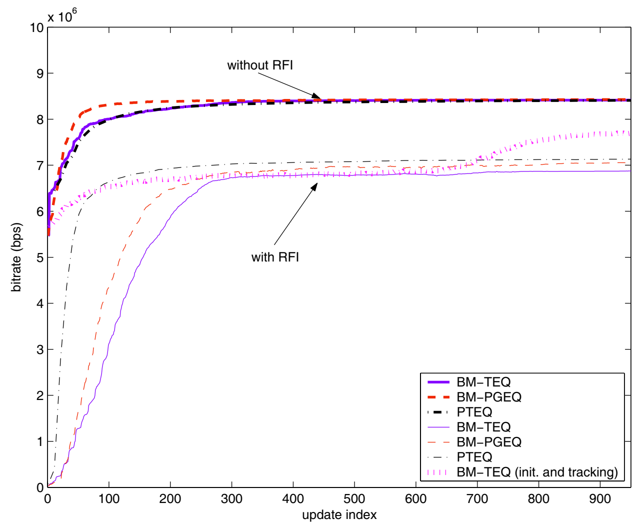

Fig. 1. Bitrate convergence of the BM-TEQ, BM-PGEQ (4 tone groups) and PTEQ, without (thin lines) and with RFI (thick lines). $T = 32$. The thick dotted line shows the adaptive BM-TEQ, initialized with an MMSE TEQ, tracking the disappearance of 2 RFIs at time instant 500.

6. REFERENCES

[1] K. Vanbleu, G. Ysebaert, G. Cuypers, M. Moonen, and K. Van Acker, "Bitrate maximizing time-domain equalizer design for DMT-based systems," accepted for publication in *IEEE Trans. on Comm.*

[2] K. Van Acker, G. Leus, M. Moonen, O. van de Wiel, and T. Pollet, "Per-tone equalization for DMT-based systems," *IEEE Trans. Comm.*, vol. 49, no. 1, pp. 109–119, 2001.

[3] K. Vanbleu, G. Ysebaert, G. Cuypers, M. Moonen, "Adaptive bitrate maximizing time-domain equalization design for DMT-based systems," Submitted. Available via ftp at ftp.esat.kuleuven.ac.be/pub/sista/vanbleu/reports/03-168.pdf.

[4] "ITU recommendation G.992.5. asymmetrical digital subscriber line (ADSL) transceivers - extended bandwidth ADSL2 (ADSL2+)," May 2003.

[5] J.S. Chow, J.M. Cioffi, and J.A.C. Bingham, "Equalizer training algorithms for multicarrier modulation systems," in *Proc. ICC*, 1993, vol. 2, pp. 761–765.

[6] R.K. Martin, J. Balakrishnan, W.A. Sethares, and C.R. Johnson Jr., "A blind, adaptive TEQ for multicarrier systems," *IEEE SPL*, vol. 9, no. 11, pp. 341–343, 2002.

[7] J. Balakrishnan, K. M. Martin, and C. R. Johnson Jr., "Blind, adaptive channel shortening by sum-squared auto-correlation minimization (SAM)," to appear in *IEEE Trans. SP*.

[8] K. Van Acker, G. Leus, M. Moonen, and T. Pollet, "RLS-based initialization for per-tone equalizers in DMT receivers," *IEEE Trans. Comm.*, vol. 51, no. 6, pp. 885–889, 2003.

[9] L. Ljung and T. Söderström, *Theory and Practice of Recursive Identification*, Cambridge, MA: MIT Press, 1983.

[10] L. S. H. Ngia and J. Sjöberg, "Efficient training of neural nets for nonlinear adaptive filtering using a recursive levenberg-marquardt algorithm," *IEEE Trans. SP*, vol. 48, no. 7, pp. 1915–1927, 2000.

[11] H. Dai and N. K. Sinha, "Robust recursive least-squares method with modified weights for bilinear system identification," *Proc. IEEE*, vol. 136, no. 3, pp. 122–126, 1989.

[12] A. T. Georgiadis and B. Mulgrew, "Adaptive bayesian decision feedback equalizer for alpha-stable noise environments," *Signal Processing*, vol. 81, no. 8, pp. 1603–1623, 2001.