Conjugate Gradient Based MMSE Decision Feedback Equalization

Guido Dietl*, Sabik Salam†, Wolfgang Utschick‡, and Josef A. Nossek°

*DOCOMO Communications Laboratories Europe GmbH, 80687 Munich, Germany
†Department of Electronics and Communication Engineering, Indian Institute of Technology, Kharagpur – 721302 WB, India
‡Associate Institute for Signal Processing, Technische Universität München, 80333 Munich, Germany
°Institute for Circuit Theory and Signal Processing, Technische Universität München, 80333 Munich, Germany

Abstract—**The performance of linear filters degrade drastically when applied to mitigate intersymbol interference caused by channels with frequent nulls in their spectral characteristics like, e.g., time-dispersive radio channels. In such cases, the well-known *Decision Feedback Equalizer* (DFE) is one possible nonlinear approach to improve the quality of the receiver. However, adapting the DFE filter coefficients to equalize time-varying channels is computationally intense, especially if the dimension of the observation vector is very high.**

In this paper, we apply the *Conjugate Gradient* (CG) *algorithm* to a conventional *Minimum Mean Square Error* (MMSE) DFE in order to reduce its computational complexity. Moreover, we compare its performance to the one of MMSE DFE versions which are based on the computationally efficient *Least-Mean-Square* (LMS) or *Recursive Least-Squares* (RLS) *algorithm*, respectively. The analysis additionally includes a detailed investigation of computational complexity with respect to the required number of *FLoating point OPerations* (FLOPs). Simulation results when applied to a digital communications system show the ability of the CG based MMSE DFE to outperform either the LMS and the RLS based MMSE DFE although its computational complexity is even smaller in most of the cases.

I. Introduction

The linear *Minimum Mean Square Error* (MMSE) *filter* estimates an unknown signal based on an observation by minimizing the *Mean Square Error* (MSE) between its output and the desired signal. Since the computation of the filter coefficients implies the solution of a system of linear equation—the so-called *Wiener-Hopf equation*—the computational complexity of the equalizer design increases with the dimension of the observation signal in cubic order. This can be especially computationally cumbersome if the filter needs to be periodically adjusted due to a time-varying channel.

In such a case, adaptive implementations of the MMSE filter like the *Least-Mean-Squares* (LMS) or the *Recursive Least-Squares* (RLS) *algorithm* [1] represent computationally efficient approximations of the optimal MMSE solution. Recently, the *Conjugate Gradient* (CG) *algorithm*—originally designed to solve iteratively systems of linear equations—has been applied to adaptive filtering [2], [3]. For example, Chowdhury et al. [4] used the CG algorithm as an adaptive linear MMSE equalizer for a multiuser multiple-antenna system and showed that the CG based equalizer has excellent convergence properties at a moderate computational cost. The CG algorithm neither requires the matrix manipulation as in the RLS nor has any instability problems that afflict some fast RLS methods [2].

All the above mentioned contributions are based on the application of the CG algorithm to linear adaptive filters in order to reduce their computational burden. However, if the spectral characteristic of the channel possesses frequent nulls which is the case for, e.g., time-dispersive radio channels, nonlinear processing like the *Decision Feedback Equalizer* (DFE) [5] is required to mitigate the effect of the dispersive channel. In order to reduce the computational complexity of the MMSE DFE, Zoltowski et al. [6] suggested recently a CG based implementation thereof and applied it to digital television which employs a 8-VSB modulation scheme with no memory.

The contribution of this paper is to apply the CG algorithm to the MMSE DFE where the statistics is either estimated via a *sample-mean* or *correlation procedure*, or computed based on the channel matrix which is estimated by using the *Least-Squares* (LS) *method* (e.g., [7]). The latter estimation approach has the advantage that it exploits special structures like, e.g., zero entries or Toeplitz structure, in the statistics, thus, reducing the number of parameters to be estimated and leading to a better estimate or performance if the length of the training sequence remains constant. Moreover, we present MMSE DFE versions which are based on the computationally efficient LMS or RLS algorithm and compare its performance to the one of the CG based MMSE DFE when applied to a digital communications system with parameters according to the *Enhanced Data rate for GSM Evolution* (EDGE) standard which suffers from severe intersymbol interference due to a special type of pulse shaping. Finally, we present a detailed investigation of the computational complexity for all proposed MMSE DFE implementations by counting the required number of *FLoating point OPerations* (FLOPs).

The next section briefly reviews the MMSE DFE. In Section III, the CG algorithm is introduced and its application to the MMSE DFE is explained. The LMS and RLS based MMSE DFE is derived in Section IV. Before applying the proposed algorithms to a digital communications system in

Section VI, we investigate the computational complexity of all algorithms in Section V.

II. MMSE Decision Feedback Equalization

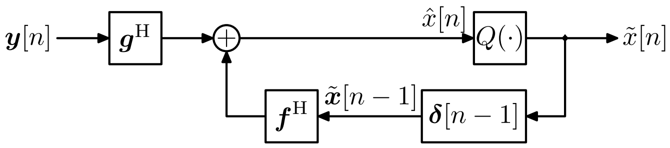

Fig. 1. MMSE decision feedback equalizer

The block diagram of the MMSE DFE which we consider in the following of this paper is depicted in Figure 1. The feedforward filter $\boldsymbol{g} \in \mathbb{C}^N$ and the feedback filter $\boldsymbol{f} \in \mathbb{C}^B$ estimate the desired signal $x[n-\nu]$ based on the observation vector $\boldsymbol{y}[n] \in \mathbb{C}^N$ which comprises the observed signal sequences at the multiple antennas of the receiver, as well as the already decided symbols

$$\begin{aligned}\tilde{\boldsymbol{x}}[n-1] &= \boldsymbol{\delta}[n-1] * \tilde{x}[n] \\ &= \begin{bmatrix}\delta[n-1]\\ \delta[n-2]\\ \vdots \\ \delta[n-B]\end{bmatrix} * \tilde{x}[n] = \begin{bmatrix}\tilde{x}[n-1]\\ \tilde{x}[n-2]\\ \vdots \\ \tilde{x}[n-B]\end{bmatrix} \in \mathbb{C}^B,\end{aligned} \tag{1}$$

respectively. Here, ν denotes the latency time of the DFE, the operation '$*$' convolution, and $\delta[n]$ the unit impulse. The decided symbols $\tilde{x}[n]$ are obtained from the estimated signal $\hat{x}[n]$ via quantization or hard decision denoted by $Q(\cdot)$.

The feedforward and feedback filter are designed to minimize the MSE

$$\xi(\boldsymbol{g}, \boldsymbol{f}) = \mathrm{E}\left\{ |x[n-\nu] - \hat{x}[n]|^2 \right\}, \tag{2}$$

i.e., the optimal filter coefficients compute as

$$(\boldsymbol{g}, \boldsymbol{f}) = \underset{(\boldsymbol{g}', \boldsymbol{f}')}{\operatorname{argmin}}\, \xi(\boldsymbol{g}', \boldsymbol{f}'). \tag{3}$$

If we assume that the statistics of the already decided symbols and the one of the desired symbols are the same which is true in case of no decision errors, the cross-correlation matrix between $\boldsymbol{y}[n]$ and $\tilde{\boldsymbol{x}}[n-1]$ can be written as

$$\begin{aligned}\boldsymbol{R}_{\boldsymbol{y}\tilde{\boldsymbol{x}}} &= \mathrm{E}\left\{\boldsymbol{y}[n]\tilde{\boldsymbol{x}}^{\mathrm{H}}[n-1]\right\} \\ &= \mathrm{E}\left\{\boldsymbol{y}[n]\boldsymbol{x}^{\mathrm{H}}[n-\nu-1]\right\} = \boldsymbol{R}_{\boldsymbol{y}\boldsymbol{x}},\end{aligned} \tag{4}$$

where we used the fact that the decided symbol $\tilde{x}[n] = Q(\hat{x}[n])$ is an estimation of the delayed symbol $x[n-\nu]$ due to the latency of the DFE (cf. also the cost function in Eq. 2). Note that throughout this paper, $(\cdot)^{\mathrm{T}}$ denotes transpose, $(\cdot)^*$ conjugate, and $(\cdot)^{\mathrm{H}}$ Hermitian, i.e., conjugate transpose. With the auto-correlation matrix $\boldsymbol{R}_{\boldsymbol{y}\boldsymbol{y}} = \mathrm{E}\{\boldsymbol{y}[n]\boldsymbol{y}^{\mathrm{H}}[n]\}$, the cross-correlation vector $\boldsymbol{r}_{\boldsymbol{y}x} = \mathrm{E}\{\boldsymbol{y}[n]x^*[n-\nu]\}$, and the variance $\sigma_x^2 = \mathrm{E}\{|x[n-\nu]|^2\}$, the MSE can be rewritten as

$$\begin{aligned}\xi(\boldsymbol{g}, \boldsymbol{f}) = \sigma_x^2 &- 2\,\mathrm{Re}\{\boldsymbol{g}^{\mathrm{H}}\boldsymbol{r}_{\boldsymbol{y}x}\} + \boldsymbol{g}^{\mathrm{H}}\boldsymbol{R}_{\boldsymbol{y}\boldsymbol{y}}\boldsymbol{g} \\ &+ 2\,\mathrm{Re}\{\boldsymbol{g}^{\mathrm{H}}\boldsymbol{R}_{\boldsymbol{y}\boldsymbol{x}}\boldsymbol{f}\} + \sigma_x^2\boldsymbol{f}^{\mathrm{H}}\boldsymbol{f},\end{aligned} \tag{5}$$

and the solution of the optimization in (3), i.e., the MMSE DFE filter coefficients finally compute as

$$\boldsymbol{g} = \left(\boldsymbol{R}_{\boldsymbol{y}\boldsymbol{y}} - \frac{1}{\sigma_x^2}\boldsymbol{R}_{\boldsymbol{y}\boldsymbol{x}}\boldsymbol{R}_{\boldsymbol{y}\boldsymbol{x}}^{\mathrm{H}}\right)^{-1}\boldsymbol{r}_{\boldsymbol{y}x}, \tag{6}$$

$$\boldsymbol{f} = -\frac{1}{\sigma_x^2}\boldsymbol{R}_{\boldsymbol{y}\boldsymbol{x}}^{\mathbf{H}}\boldsymbol{g}. \tag{7}$$

III. Conjugate Gradient Based MMSE DFE

The computation of the MSE optimal DFE filter coefficients according to (6) and (7) involves the solution of a system of N linear equations with N unknowns, i.e.,

$$\left(\boldsymbol{R}_{\boldsymbol{y}\boldsymbol{y}} - \frac{1}{\sigma_x^2}\boldsymbol{R}_{\boldsymbol{y}\boldsymbol{x}}\boldsymbol{R}_{\boldsymbol{y}\boldsymbol{x}}^{\mathrm{H}}\right)\boldsymbol{g} = \boldsymbol{r}_{\boldsymbol{y}x}, \tag{8}$$

each time when the statistics of the system is changing which is periodically the case if the wireless channel is time-varying. Adapting the filter coefficients to the channel state can be computationally cumbersome because the complexity of solving a system of linear equations is of cubic order.

One way of efficiently solving a system of linear equations is to apply the CG algorithm [8]. Here, we apply it to compute iteratively the coefficients $\boldsymbol{g}$ and $\boldsymbol{f}$ of the equalizer by solving (8). The CG method is based on *Gram-Schmidt orthogonalization* where the search directions $\boldsymbol{d}_n$ are constructed by conjugation of the residuals $\boldsymbol{r}_n = \boldsymbol{r}_{\boldsymbol{y}x} - \boldsymbol{R}_{\boldsymbol{y}\boldsymbol{y}}\boldsymbol{g}_n$ corresponding to the approximate solutions $\boldsymbol{g}_n$. The residuals are orthogonal to all previous search directions, and hence, it gives a new linearly independent search direction for each iteration, unless and until the residual is zero which is the case when the algorithm finds the optimum solution. Since the residuals are orthogonal to previous search directions, they are also orthogonal to previous residuals. Therefore, the new residual $\boldsymbol{r}_{n+1}$ is a linear combination of previous residuals and $\boldsymbol{R}_{\boldsymbol{y}\boldsymbol{y}}\boldsymbol{r}_n$. It can be shown that the $(n+1)$-dimensional subspace $\mathcal{D}_{n+1}$ spanned by the search directions $\boldsymbol{d}_0$, $\boldsymbol{d}_1$, $\ldots$, and $\boldsymbol{d}_n$, is formed by the sum of the previous subspace $\mathcal{D}_n$ and the subspace spanned by $\boldsymbol{R}_{\boldsymbol{y}\boldsymbol{y}}\mathcal{D}_n$, resulting in a *Krylov subspace* [9]. Algorithm 1 summarizes the CG method with D iterations to solve (8) in order to get finally the filter coefficients according to (6) and (7). Note that if $D < N$, the resulting filter coefficients are only approximations of the optimal MMSE DFE coefficients.

In order to perform Algorithm 1, the unknown statistics need to be estimated. This can be done by estimating the channel based on a LS approach and using the resulting channel estimate to compute the statistics like, e.g., presented in [7]. An alternative way is to estimate the statistics directly via a *sample-mean* or *correlation procedure*. From the received observation vector $\boldsymbol{y}[n]$ and the training symbols $x[n]$, considering the ergodicity theorem, we can estimate the auto-correlation matrix, the cross-correlation matrix, and the cross-

Algorithm 1 CG based MMSE DFE

$\boldsymbol{g}_0 \leftarrow \boldsymbol{0}_N$ (or the filter vector of the previous block)
$\boldsymbol{r}_0 \leftarrow \boldsymbol{r}_{\boldsymbol{yx}} - \boldsymbol{R}_{\boldsymbol{yy}}\boldsymbol{g}_0$
$\boldsymbol{d}_0 \leftarrow \boldsymbol{r}_0$
for $n = 0, 1, \ldots, D-1,\ D \leq N$ **do**
  $\alpha_n \leftarrow \boldsymbol{r}_n^{\mathrm{H}}\boldsymbol{r}_n / (\boldsymbol{d}_n^{\mathrm{H}}\boldsymbol{R}_{\boldsymbol{yy}}\boldsymbol{d}_n)$
  $\boldsymbol{g}_{n+1} \leftarrow \boldsymbol{g}_n + \alpha_n \boldsymbol{d}_n$
  $\boldsymbol{r}_{n+1} \leftarrow \boldsymbol{r}_n - \alpha_n \boldsymbol{R}_{\boldsymbol{yy}}\boldsymbol{d}_n$
  $\beta_{n+1} \leftarrow \boldsymbol{r}_{n+1}^{\mathrm{H}}\boldsymbol{r}_{n+1} / (\boldsymbol{r}_n^{\mathrm{H}}\boldsymbol{r}_n)$
  $\boldsymbol{d}_{n+1} \leftarrow \boldsymbol{r}_{n+1} + \beta_{n+1}\boldsymbol{d}_n$
end for
$\hat{\boldsymbol{g}} \leftarrow \boldsymbol{g}_D$
$\hat{\boldsymbol{f}} \leftarrow -\boldsymbol{R}_{\boldsymbol{yx}}^{\mathrm{H}}\boldsymbol{g}_D / \sigma_x^2$

correlation vector for an L_{p}-length training sequence via

$$\hat{\boldsymbol{R}}_{\boldsymbol{yy}} = \frac{1}{L_{\mathrm{p}}} \sum_{n=0}^{L_{\mathrm{p}}-1} \boldsymbol{y}[n]\boldsymbol{y}^{\mathrm{H}}[n], \tag{9}$$

$$\hat{\boldsymbol{R}}_{\boldsymbol{yx}} = \frac{1}{L_{\mathrm{p}}} \sum_{n=0}^{L_{\mathrm{p}}-1} \boldsymbol{y}[n]\boldsymbol{x}^{\mathrm{H}}[n-\nu-1], \tag{10}$$

$$\hat{\boldsymbol{r}}_{\boldsymbol{yx}} = \frac{1}{L_{\mathrm{p}}} \sum_{n=0}^{L_{\mathrm{p}}-1} \boldsymbol{y}[n]x^{*}[n-\nu], \tag{11}$$

respectively.

Note that the fundamental difference between both estimation approaches is the fact that the LS channel estimation based scheme is exploiting the structure of the correlation matrices, i.e., zero entries, Toeplitz structure, etc., whereas the correlation based method does not. We will see in the simulation results of Section VI that the exploitation of structures leads to a better performance at the same or even at a smaller computational cost. This is because the number of parameters to be estimated is reduced, hence, if the length of the training sequence is kept constant, the estimation of statistics is improved.

IV. LMS and RLS Based MMSE DFE

The LMS and RLS algorithms [1] are two well-known adaptive implementations of the MMSE equalizer. Here, we extend the LMS and RLS idea to adaptively compute the filter coefficients of the MMSE DFE. To do so, we reformulate (6) and (7) to get the following system of linear equations:

$$\begin{bmatrix} \boldsymbol{R}_{\boldsymbol{yy}} & \boldsymbol{R}_{\boldsymbol{yx}} \\ \boldsymbol{R}_{\boldsymbol{yx}}^{\mathrm{H}} & \sigma_x^2 \boldsymbol{1}_B \end{bmatrix} \begin{bmatrix} \boldsymbol{g} \\ \boldsymbol{f} \end{bmatrix} = \begin{bmatrix} \boldsymbol{r}_{\boldsymbol{yx}} \\ \boldsymbol{0}_B \end{bmatrix}. \tag{12}$$

Then, with the definitions

$$\boldsymbol{z}[n] = \begin{bmatrix} \boldsymbol{y}[n] \\ \tilde{\boldsymbol{x}}[n-1] \end{bmatrix} \in \mathbb{C}^{N+B}, \quad \boldsymbol{w} = \begin{bmatrix} \boldsymbol{g} \\ \boldsymbol{f} \end{bmatrix} \in \mathbb{C}^{N+B}, \tag{13}$$

the LMS and RLS algorithm can be applied to solve (12) as described by Algorithms 2 and 3. Compared to the CG based MMSE DFE which can be implemented by either exploiting the structure of statistics or not, the LMS and RLS based MMSE DFE cannot exploit the structure in any case. This is a major drawback of the LMS and RLS algorithm in the given context.

Algorithm 2 LMS based MMSE DFE

$\boldsymbol{w}_0 \leftarrow \boldsymbol{0}_{N+B}$ (or the filter vector of the previous block)
for $n = 0, 1, \ldots, L_{\mathrm{p}}-1$ **do**
  $\hat{x}[n] \leftarrow \boldsymbol{w}_n^{\mathrm{H}}\boldsymbol{z}[n]$
  $e[n] \leftarrow x[n-\nu] - \hat{x}[n]$
  $\boldsymbol{w}_{n+1} \leftarrow \boldsymbol{w}_n + \mu \boldsymbol{z}[n]e^{*}[n]$
end for
$\hat{\boldsymbol{w}} \leftarrow \boldsymbol{w}_{L_{\mathrm{p}}}$

Algorithm 3 RLS based MMSE DFE

$\boldsymbol{w}_0 \leftarrow \boldsymbol{0}_{N+B}$ (or the filter vector of the previous block)
$\boldsymbol{P}_0 \leftarrow \delta \boldsymbol{I}$
for $n = 0, 1, \ldots, L_{\mathrm{p}}-1$ **do**
  $\boldsymbol{k}_n \leftarrow \lambda^{-1}\boldsymbol{P}_n\boldsymbol{z}[n] / (1 + \lambda^{-1}\boldsymbol{z}^{\mathrm{H}}[n]\boldsymbol{P}_n\boldsymbol{z}[n])$
  $\xi[n] \leftarrow x[n-\nu] - \boldsymbol{w}_{n-1}^{\mathrm{H}}\boldsymbol{z}[n]$
  $\boldsymbol{w}_{n+1} \leftarrow \boldsymbol{w}_n + \boldsymbol{k}_n\xi^{*}[n]$
  $\boldsymbol{P}_{n+1} \leftarrow \lambda^{-1}\boldsymbol{P}_n - \lambda^{-1}\boldsymbol{k}_n\boldsymbol{z}^{\mathrm{H}}[n]\boldsymbol{P}_n$
end for
$\hat{\boldsymbol{w}} \leftarrow \boldsymbol{w}_{L_{\mathrm{p}}}$

V. Computational Complexity Analysis

In this section, we investigate the computational complexity of the different DFE implementations by counting the exact number of real-valued FLOPs involved when computing the equalizer coefficients. Here, one FLOP is defined as either one real multiplication or one real summation. Considering the Hermitian symmetry of the auto-correlation matrix, the CG based MMSE DFE requires

$$\begin{aligned} \mathcal{C}_{\text{CG-C}} = {} & \frac{4}{\kappa}\left(\kappa L_{\mathrm{p}} - \frac{N}{K} + 1\right)\left(N^2 + (2B+1)N + 1\right) \\ & + D\left(8N^2 + 26N - 2\right) + N^2(4B-2) \\ & + N\left(4B - \frac{2}{\kappa}(B+1) + 2\right) - 12 \text{ FLOPs}, \end{aligned} \tag{14}$$

if statistics is computed via correlation as given in (9), (10), and (11). If statistics is computed based on LS channel estimation, the CG based MMSE DFE requires

$$\begin{aligned} \mathcal{C}_{\text{CG-LS}} = {} & D\left(8N^2 + 26N - 2\right) + 4(2+B)N^2 \\ & + 8BN - 7N + 8KL(\kappa L_{\mathrm{p}} - L) + 6KL \\ & + \frac{2KL}{\kappa}\left(\frac{2KL}{\kappa} - 1\right) - 3 \text{ FLOPs}. \end{aligned} \tag{15}$$

Note that L is the length of the channel impulse response, K the number of receive antennas, and κ an oversampling factor. Here and in the simulation results of the next section, we assume that the received signal is sampled κ times per symbol duration before it is processed by the DFE (see, e.g., [10]).

Finally, the number of FLOPs required for the LMS and RLS based MMSE DFE computes as

$$\mathcal{C}_{\text{LMS}} = \frac{1}{\kappa}\left(\kappa L_{\text{p}} - \frac{N}{K} + 1\right)(16(N+B)+2), \tag{16}$$

$$\mathcal{C}_{\text{RLS}} = \frac{1}{\kappa}\left(\kappa L_{\text{p}} - \frac{N}{K} + 1\right)\left(16 + \frac{24}{N+B}\right) \tag{17}$$

$$\cdot (N+B)^2. \tag{18}$$

VI. Numerical Simulations

We consider a digital communications system with parameters according to the EDGE standard, i.e., with 8-PSK modulation and *Laurent pulse shaping* [11]. The Laurent impulse is a linearized *Gaussian Minimum-Shift Keying* (GMSK) impulse with a duration of five symbol times, thus, introducing severe intersymbol interference even without channel distortion. The symbol sequence is transmitted with a symbol time of $T = 3.69\,\mu$s and propagates over a time-varying Rayleigh fading channel with a maximum doppler frequency of about 83 Hz, corresponding to a velocity of 100 km/h at a carrier frequency of 900 MHz, and a delay spread of $\tau_{\max}$ equals three symbol times. Hence, the combination of pulse shaping and channel can be modeled as a $L = 6$ tap filter. We assume a constant channel during one burst with 148 symbols (excluding the guard symbols) where $L_{\text{p}} = 26$ symbols are used for training. At the receiver, the space-time observation vector $\boldsymbol{y}[n]$ comprises 15 time samples at each of the $K = 2$ antennas, i.e., its dimension computes as $N = 2 \cdot 15 = 30$, where $\kappa = 2$ samples have been taken per symbol duration. The feedback filter has $B = 6$ taps and the latency time $\nu = 17$. The following parameters have been chosen for Algorithms 2 and 3: $\mu = 0.03$, $\delta = 0.3$, and $\lambda = 1$.

TABLE I
Computational complexity of MMSE DFE

MMSE DFE type	FLOPs
LMS	9,056
RLS	345,312
CG-C, $D = 4$ it.	84,966
CG-C, $D = 6$ it.	100,922
CG-LS, $D = 4$ it.	66,559
CG-LS, $D = 6$ it.	82,515

Table I shows the computational complexity with respect to the number of FLOPs for the system parameters assumed in this section. We observe that the LMS based MMSE DFE is the computational cheapest solution whereas the RLS based MMSE DFE requires the highest number of FLOPs. Besides, the LS channel estimation based CG algorithm is computationally cheaper than the correlation based counterpart. In case of 4 iterations, the LS based CG implementation of the MMSE DFE requires less than a fifth of the number of FLOPs needed to perform the RLS based MMSE DFE.

Next, we investigate the performance of the proposed algorithms by evaluating the resulting uncoded *Bit Error Rate* (BER) for different *Signal-to-Noise Ratio* (SNR) values.

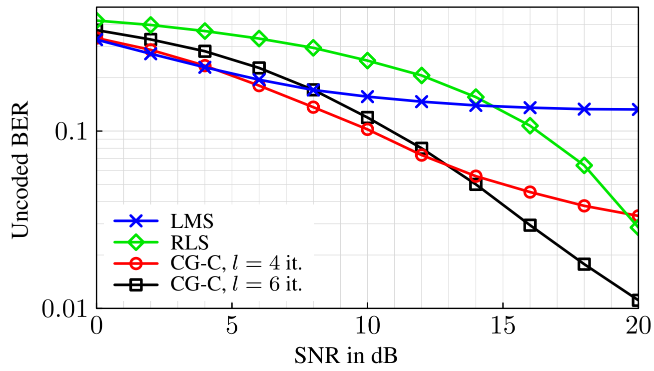

Fig. 2. BER comparison with correlation based CG method when starting algorithms with zero vector

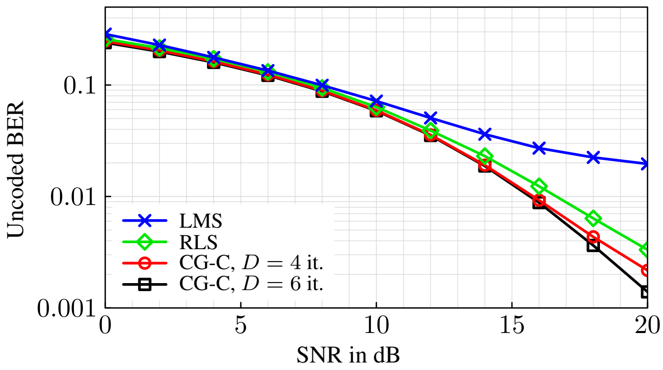

Fig. 3. BER comparison with correlation based CG method when starting algorithms with filter vector of the previous block

Figure 2 presents a BER comparison of the LMS, RLS, and CG based MMSE DFE where the latter estimates the statistics based on the correlation procedure. At the beginning of each block, all algorithms are initialized with a filter vector equals zero. The simulation results show that the CG based MMSE DFE with only 4 iterations performs significantly better than either the LMS and RLS based MMSE DFE at the given SNR range although its computational complexity is smaller. Only the LMS based MMSE DFE is computationally much more efficient than the CG based MMSE DFE, however, its performance is not acceptable. Note that the CG based MMSE DFE with 4 iterations outperforms the CG based MMSE DFE with 6 iterations at lower SNRs due to the regularizing effect [12] of the CG algorithm.

Figure 3 presents a BER comparison of the given algorithms if they are initialized with the filter coefficients of the previous block. Doing so improves the performance of all algorithms drastically. However, the CG based MMSE DFE still outperforms the remaining adaptive DFE implementations for all considered values of D. In case of the LMS based MMSE DFE, the initialization with the filter vector of the previous block results directly in a longer training sequence. Note that the performance of the RLS based MMSE DFE can be further improved by starting the correlation also with the

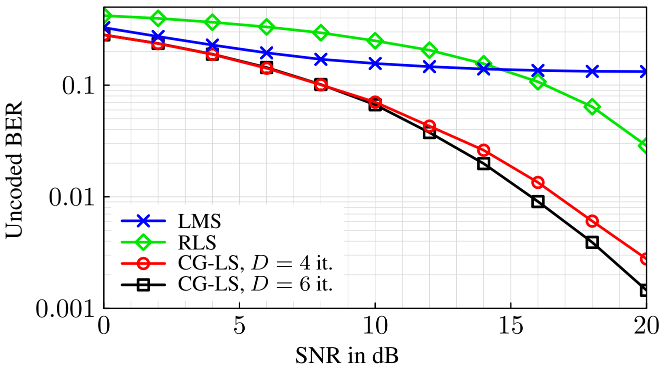

Fig. 4. BER comparison with LS based CG method when starting algorithms with zero vector

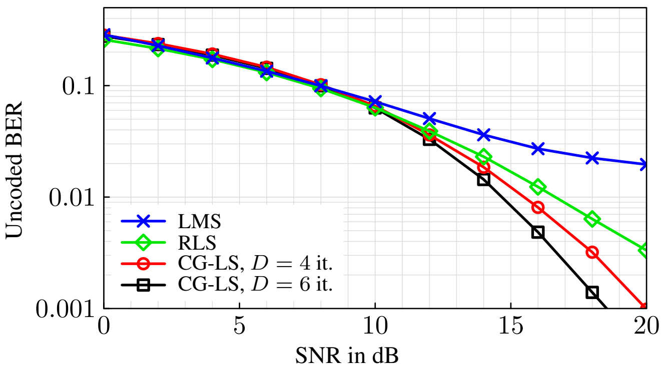

Fig. 5. BER comparison with LS based CG method when starting algorithms with filter vector of the previous block

final statistics estimate of the previous block, which can be also applied to the CG based MMSE DFE. However, this has not been considered in this paper.

Finally, Figures 4 and 5 present BER comparisons where the statistics of the CG based MMSE DFE is calculated based on a LS estimation of the channel. We see that the performance of the CG based MMSE DFE improves if the estimation of statistics exploits its structure appropriately. This is especially remarkable if we recall that the computational complexity is even smaller than in case of statistics estimation via correlation (cf. Table I).

VII. Conclusions

In this paper, we derived an MMSE DFE implementation based on the CG algorithm. The simulation results when applied to a digital communications system together with a detailed computational complexity analysis showed that the CG based MMSE DFE outperforms the computationally more expensive RLS based MMSE DFE both in terms of BER and in terms of computational complexity. Contrary to that, the LMS based MMSE is computationally most efficient among all algorithms, however, its performance is too poor in the considered scenario. Using the initial values of the coefficients as the values computed for the previous block improves the performance of all algorithms without any increase in computational complexity. It remains to mention that one should determine statistics via LS channel estimation if one decides to use the CG based MMSE DFE because this implementation achieves lower BERs than the correlation based version despite the decreased computational complexity.

References

[1] S. S. Haykin, *Adaptive Filter Theory*. Prentice Hall, 1986.

[2] G. K. Boray and M. D. Srinath, "Conjugate gradient techniques for adaptive filtering," *IEEE Transactions on Circuits and Systems–I: Fundamental Theory and Applications*, vol. 39, no. 1, pp. 1–10, January 1992.

[3] P. S. Chang and A. N. Willson, Jr., "Analysis of conjugate gradient algorithms for adaptive filtering," *IEEE Transactions on Signal Processing*, vol. 48, no. 2, pp. 409–418, February 2000.

[4] S. Chowdhury and M. D. Zoltowski, "Application of conjugate gradient methods in MMSE equalization for the forward link of DS-CDMA," in *Proceedings of the 2001 IEEE 54th Vehicular Technology Conference (VTC 2001-Fall)*, October 2001, pp. 2434–2438.

[5] P. Monsen, "Feedback equalization for fading dispersive channels," *IEEE Transactions on Information Theory*, vol. IT-17, no. 1, pp. 56–64, January 1971.

[6] M. D. Zoltowski, W. J. Hillery, S. Özen, and M. Fimoff, "Conjugate-gradient-based decision feedback equalization with structured channel estimation for digital television," in *Proceedings of SPIE – The International Society for Optical Engineering*, vol. 4740, April 2002.

[7] G. Dietl and W. Utschick, "Low-complexity equalizers—rank versus order reduction," in *Proceedings of the 7th IEEE Workshop on Signal Processing Advances in Wireless Communications (SPAWC 2006)*, July 2006, pp. 1–5.

[8] M. R. Hestenes and E. Stiefel, "Methods of conjugate gradients for solving linear systems," *Journal of Research of the National Bureau of Standards*, vol. 49, no. 6, pp. 409–436, December 1952.

[9] Y. Saad, *Numerical Methods for Large Eigenvalue Problems*. Halstead Press – *out of print*, 1992.

[10] M. D. Zoltowski and T. A. Thomas, "Novel zero-forcing, MMSE, and DFE equalizer structures employing oversampling and multiple receiver antennas," *Proceedings of the 36nd Conference on Signals, Systems, and Computers*, vol. 2, pp. 1111–1116, November 1998.

[11] European Telecommunications Standards Institute, "Digital cellular telecommunications system (phase 2+); modulation," EN 300 959 V7.1.1, ETSI, http://www.etsi.org, ETSI, June 2000.

[12] G. Dietl, M. Botsch, F. A. Dietrich, and W. Utschick, "Robust and reduced-rank matrix Wiener filter based on the block conjugate gradient algorithm," in *Proceedings of the 6th IEEE Workshop on Signal Processing Advances in Wireless Communications (SPAWC 2005)*, June 2005, pp. 555–559.