An Analysis of the Bidirectional LMS Algorithm over Fast-Fading Channels

Yavuz Yapıcı and Ali Ö zgür Yılmaz

Abstract

A bidirectional LMS algorithm is considered for estimation of fast frequency-selective time-varying channels with a promise of near optimal tracking performance and robustness to parameter imperfections under various scenarios at a practical level of complexity. The performance of the algorithm is verified by the theoretical steady-state MSE analysis and experimental bit error rate (BER) results.

I. INTRODUCTION

The adaptive least mean square (LMS) algorithm is of interest with its simple first order update equation [1]. Unfortunately, tracking performance of the LMS algorithm deteriorates dramatically in fast time-varying environments [2]. A recent bidirectional estimation strategy, which is pioneered by [3] and further elaborated in [4] and [5], offers an improved tracking performance for fast-time varying channels, but this time, at the expense of a severe computational complexity.

In this paper, we consider a bidirectional LMS algorithm [6], [7] over fast frequency-selective time-varying channels with an increased but still practical level of complexity. The tracking performance of the proposed algorithm at the steady-state is very close to that of the optimal minimum mean-square error (MMSE) filter in some settings of practical interest in terms of communication systems and is remarkably better than that of the conventional LMS. Although there are various work present in the literature on other forms of bidirectional estimation [3], [4], [5], none of them provide a theoretical analysis on the mean-square error (MSE) behavior. Therefore, as a major contribution of this paper, we analyze the tracking performance of the bidirectional LMS algorithm by deriving a novel step-size dependent steady-state MSE and optimal step-size expressions over fast frequency-selective time-varying channels. This derivation is applicable to many communication scenarios in the sense that it does not depend on the channel characteristics and the modulation scheme in use. The numerical evaluations show a very
good match between the theoretical and the experimental results most of the time. The robustness of the algorithm to the imperfect initialization and noisy Doppler and signal-to-noise ratio (SNR) values is also verified with the associated mean square identification error (MSIE) statistics. Finally, the promised performance is also investigated through BER results in a coded scenario as a more realistic application.

II. SYSTEM MODEL AND THE BIDIRECTIONAL LMS ALGORITHM

We consider an unknown time-varying frequency-selective communication channel represented by an $L_c$-tap fading vector $\mathbf{f}_k = [f_{k,0} \ldots f_{k,L_c-1}]^T$ with uncorrelated entries and assume the following discrete-time complex baseband model at an epoch $k$ given as

$$y_k = \sum_{l=0}^{L_c-1} f_{k,l} a_{k-l} + n_k = \mathbf{f}_k^T \mathbf{a}_k + n_k$$

where $y_k$ is the observation symbol, $\mathbf{a}_k = [a_k \ldots a_{k-L_c+1}]^T$ is the vector of data symbols chosen from a finite alphabet $\mathcal{A}$ in an independent and identical fashion, and $n_k$ is a circularly symmetric complex white Gaussian noise with zero-mean and variance $N_0$.

The bidirectional LMS algorithm is basically an extension of the conventional unidirectional LMS that operates both in the forward and the backward directions along an observation block. Defining $\mathbf{f}_k^f$ and $\mathbf{f}_k^b$ to be the channel estimates in the forward and the backward directions, respectively, the algorithm is given as

$$\mathbf{f}_{k+1}^f = \mathbf{f}_k^f + 2\mu e^f_k \mathbf{a}_k$$

$$\mathbf{f}_{k-1}^b = \mathbf{f}_k^b + 2\mu e^b_k \mathbf{a}_k$$

where $\mu$ is the step-size, $e^f_k = y_k - (\mathbf{f}_k^f)^T \mathbf{a}_k$ and $e^b_k = y_k - (\mathbf{f}_k^b)^T \mathbf{a}_k$ are the forward and the backward errors, respectively. The arithmetic average operation is preferred among various choices as a simple yet efficient combining strategy to obtain the final coefficient estimates $\mathbf{f}_k$ as follows

$$\mathbf{f}_k = \frac{\mathbf{f}_k^f + \mathbf{f}_k^b}{2}.$$
that the MMSE filter also requires a matrix inversion of complexity $O(K^3)$ and a matrix multiplication of complexity $O(L_cK^2)$ to compute optimal filter coefficients. As a result, the overall complexity of the bidirectional LMS is approximately twice that of the conventional LMS and significantly lower compared to the optimal MMSE estimation.

III. TRACKING PERFORMANCE OF THE BIDIRECTIONAL LMS ALGORITHM

A. Steady-State MSE and Optimal Step-Size Expressions

In this section, we will evaluate the tracking performance of the bidirectional LMS algorithm over a frequency-selective time-varying channel by deriving a steady-state MSE expression together with the optimal step-size. Since we are dealing with tracking performance, $a_k$'s are assumed to be perfectly known along a block of length $L$. The associated error performance surface, or equivalently the MSE expression, is given as

$$ J_{\text{MSE}} = E\{|e_k|^2\} = E\{|y_k - \hat{f}_k^T a_k|^2\} = E\{|n_k|^2\} + E\{|a_k|^2\} E\{\|f_k - \hat{f}_k\|^2\} $$

(5)

where $e_k$ is the overall tracking error, $J_{\text{min}}$ is the minimum achievable MSE due to the presence of additive noise and is equal to $N_0$, $E_s$ is the average symbol energy and $J_{\text{MSIE}}$ is the MSIE [8]. For a time-varying channel, MSIE is decoupled as [9]

$$ J_{\text{MSIE}} = E\{\|f_k - E\{\hat{f}_k\}\|^2\} + E\{\|E\{\hat{f}_k\} - \hat{f}_k\|^2\} $$

(6)

As a result, MSIE of the bidirectional LMS in time-varying environments is the sum of two terms which are called the self-noise ($J_{\text{self}}$) and the lag ($J_{\text{lag}}$) [9], and will be derived separately in this section. It is also discussed in [9] that $J_{\text{self}}$ arises from the noisy gradient estimation of the error performance surface, and $J_{\text{lag}}$ is due to the time variation.

Because the source of the self-noise is the noisy gradient estimation, we ignore any time variation while deriving $J_{\text{self}}$, as in [9]. To this end, we first model the gradient estimates in the forward and the backward directions as

$$ \hat{\nabla}^f_k = \nabla^f_k + \epsilon^f_k \quad \text{and} \quad \hat{\nabla}^b_k = \nabla^b_k + \epsilon^b_k, $$

where $\nabla_k = 2E_s(f_k - f_k)$ and $\nabla^b_k = 2E_s(f_k - f_k)$ are the true gradients and $\epsilon^f_k$ and $\epsilon^b_k$ are the noise terms. Following steps similar to
those in [1], the adaptations given in (2)-(3) become

\[
\begin{align*}
\mathbf{v}_k^{f+1} &= (1 - 2 \mu E_s)\mathbf{v}_k^f - \mu \mathbf{e}_k^f \\
\mathbf{v}_k^{b-1} &= (1 - 2 \mu E_s)\mathbf{v}_k^b - \mu \mathbf{e}_k^b
\end{align*}
\]

(7)

(8)

where \( \mathbf{v}_k^f = \hat{\mathbf{f}}_k^f - \mathbf{f}_k \) and \( \mathbf{v}_k^b = \hat{\mathbf{f}}_k^b - \mathbf{f}_k \) are the tap-weight tracking errors in the forward and the backward directions, respectively.

Since we ignore any time-variation in this particular case, the self-noise given in (6) becomes \( J_{self} = E\{\|\mathbf{v}_k\|^2\} \) (see [10] for details) where \( \mathbf{v}_k = \mathbf{f}_k - \hat{\mathbf{f}}_k \) is the overall tap-weight tracking error and is given as

\[
\mathbf{v}_k = \hat{\mathbf{f}}_k - \mathbf{f}_k = \frac{\mathbf{f}_k^f - \mathbf{f}_k + \mathbf{f}_k^b - \mathbf{f}_k}{2} = \frac{\mathbf{v}_k^f + \mathbf{v}_k^b}{2}.
\]

(9)

The self-noise could then be evaluated as

\[
J_{self} = \frac{E\{\|\mathbf{v}_k^f\|^2\}}{4} + \frac{E\{\|\mathbf{v}_k^b\|^2\}}{4} + \Re\{E\{\mathbf{v}_k^f(\mathbf{v}_k^b)^H\}\}
\]

(10)

where the last term could be further elaborated by iterative employment of (7)-(8) as follows

\[
E\{\mathbf{v}_k^f(\mathbf{v}_k^b)^H\} = (1 - 2 \mu E_s)E\{\mathbf{v}_0^f(\mathbf{v}_0^b)^H\}
\]

(11)

since \( \mathbf{e}_k^f \) and \( \mathbf{e}_k^b \) are assumed to be zero-mean random variables which are mutually independent of each other and of \( \mathbf{v}_k^f \) and \( \mathbf{v}_k^b \) [1]. As a result, the last term in (10) could safely be ignored since \( (1 - 2 \mu E_s)L \ll 1 \) in (11) due to the fact that \( |1 - 2 \mu E_s| < 1 \) is the mean-convergence condition of the LMS algorithm.

In [11], [12], an iterative expression is given for the mean-square energy of the tap-weight tracking error for the conventional LMS which could be expressed at the steady-state as

\[
E\{\|\mathbf{v}_k^f\|^2\} = E\{\|\mathbf{v}_k^b\|^2\} = \frac{\mu L_c^2 E_s^2}{E_s - \mu [(L_c - 1)E_s^2 + E_4]} J_{min}
\]

(12)

where \( E_4 = E\{|a_k|^4\} \). Therefore, the self-noise given in (10) becomes

\[
J_{self} = E\{\|\mathbf{v}_k\|^2\} = \frac{\mu L_c^2 E_s^2}{2(E_s - \mu [(L_c - 1)E_s^2 + E_4])} J_{min}
\]

(13)

which is observed to depend on the step-size \( \mu \), the number of channel taps \( L_c \), average energies \( E_s \) and \( E_4 \) of the input symbols and the minimum achievable MSE which is \( N_0 \).
Since contribution of the noisy gradient estimation into the overall MSE is considered by the self-noise part, perfect gradient estimation is assumed in the same way as in [9] while analyzing the lag component, and the focus is on the time variation only. The resulting adaptive processes could be expressed as

\[
\hat{f}_{k+1}^f = \hat{f}_k^f - \mu \nabla_k^f = (1 - 2\mu E_s)\hat{f}_k^f + 2\mu E_s f_k
\]

(14)

\[
\hat{f}_{k-1}^b = \hat{f}_k^b - \mu \nabla_k^b = (1 - 2\mu E_s)\hat{f}_k^b + 2\mu E_s f_k
\]

(15)

In order to cope with time variation, z-transforms of (14)-(15) are computed and combined according to (4) as follows

\[
\hat{f}(z) = \frac{\hat{f}'(z) + \hat{f}''(z)}{2} = \frac{1}{2} \left( \frac{1 - \beta}{z - \beta} \hat{f}(z) + \frac{1 - \beta}{z-1 - \beta} \hat{f}(z) \right)
\]

(16)

where \(\beta = 1 - 2\mu E_s\) is called the geometric ratio of the adjustments in [1]. Using (16), z-transform of the estimation error becomes

\[
\hat{f}(z) - f(z) = H(z) f(z)
\]

(17)

where \(H(z)\) is the transfer function for the bidirectional LMS algorithm, which is independent of the channel characteristics to be estimated, and is given as

\[
H(z) = -\frac{1 + \beta}{2\beta} + \frac{1 - \beta}{2\beta} \left( \frac{1}{-\beta z - 1} - \frac{1}{1 - \frac{1}{\beta} z - 1} \right).
\]

(18)

Since the gradient is assumed to be estimated perfectly for this particular case, the lag component given in (6) becomes \(J_{lag} = E\{\|\hat{f}_k - f_k\|^2\}\) [10]. Therefore, \(J_{lag}\) is the mean-square energy of the estimation error, and could be evaluated in the frequency domain using (17) and (18) as follows [8]

\[
J_{lag} = \frac{L_c}{2\pi} \int_{-\pi}^{\pi} |H(e^{jw})|^2 S_f(w) dw
\]

(19)

where \(S_f(w)\) is the power spectrum of the fading process under consideration, and \(H(e^{jw}) = H(z)|_{z=e^{jw}}\) from (18) [8]. Using (5), (13) and (19), the final form of MSE expression becomes

\[
J_{MSE} = \left( 1 + \frac{\mu L_c^2 E_s^3}{2(E_s - \mu [(L_c - 1)E_s^2 + E_s])} \right) J_{min} + \frac{L_c E_s}{2\pi} \int_{-\pi}^{\pi} |H(e^{jw})|^2 S_f(w) dw.
\]

(20)

In order to derive the optimal step-size, \(\mu_{opt}\), theoretically, we first take the derivative of (20) with respect
to $\beta$ as follows

$$
\frac{\partial J_{MSE}}{\partial \beta} = -\frac{E_s^2 L_c^2}{(2E_s^2 - (1 - \beta) [(L_c - 1)E_s^2 + E_4])^2} J_{\min} + \frac{E_s L_c}{\pi} \int_{\pi}^{\pi} H(e^{jw}) \frac{\partial H(e^{jw})}{\partial \beta} S_f(w) \, dw \tag{21}
$$

where $\frac{\partial H(e^{jw})}{\partial \beta}$ is evaluated to be

$$
\frac{\partial H(e^{jw})}{\partial \beta} = -\frac{(1 - \cos w)(1 - \beta^2 - 2\beta + 2\cos w)}{(1 + \beta^2 - 2\beta \cos w)^2}. \tag{22}
$$

The optimal values $\beta_{opt}$ and $\mu_{opt}$ could then be evaluated numerically using (21) and (22) as $\frac{\partial J_{MSE}}{\partial \beta}|_{\beta = \beta_{opt}} = 0$ and $\mu_{opt} = \frac{1 - \beta_{opt}}{2E_s}$, respectively.

**B. Numerical Results**

Without any loss of generality, a frequency-selective channel with $L_c = \{2, 4\}$ taps is assumed with wide-sense stationary uncorrelated scattering (WSSUS) Rayleigh fading generated according to the well-known Jakes’ model [13]. The channel has a fast time variation with the maximum normalized Doppler frequency of $f_d T_s = \{0.01, 0.02\}$. In each trial, a set of $L = 100$ information symbols are chosen independently from the the BPSK alphabet $A = \{-1, +1\}$ so that $E_s = E_4 = 1$, and the observations are produced according to the system model given in (1). We accordingly assume constant average SNR and maximum normalized Doppler frequency over a data block due to the transmission of short data blocks.

In Fig. 1, we plot theoretical and experimental normalized MSIE, i.e., $J_{MSIE}/L_c$, results for the bidirectional LMS (BiLMS) algorithm by using (5) and (20) for varying $\mu$ and at $\gamma = E_s/N_0 = 10$ dB for $f_d T_s = 0.01$ and $\gamma = 4$ dB for $f_d T_s = 0.02$ where $\gamma$ denotes the average received SNR. The experimental normalized MSIE results associated with the conventional unidirectional LMS (UniLMS) and MMSE filter are also provided. We observe a very good match between the theoretical and the experimental results for the bidirectional LMS for any choice of $\mu$. The tracking ability of the bidirectional LMS is also verified by achieving a minimum MSIE which is close to that of the MMSE filter which is far beyond that of the conventional LMS. We should also report that no significant performance improvement is observed in MMSE filter when $K > 31$ for $f_d T_s = 0.01$ and $K > 15$ for $f_d T_s = 0.02$.

In Fig. 2, the MSIE results are presented for varying $\gamma$ with the optimal step-size values ($\mu_{opt}$) over a frequency-selective channel with $L_c = \{2, 4\}$ taps. The experimental and the theoretical MSIE results for the bidirectional LMS algorithm are again observed to exhibit a very good match for various $\gamma$
choices. The theoretical $\mu_{\text{opt}}$’s computed according to (21)-(22) in Table I together with the experimental values produced through exhaustive search with 0.01 increments which are observed to be very close to each other for various $f_d T_s$ choices. Because the large step-size values contribute to the self-noise part and the small ones amplify the lag part of the associated MSIE, the optimal step-size appears to be a compromise to obtain the best performance in accordance with the results of Fig. 1 and should be greater to track much faster channels.

IV. ROBUSTNESS OF THE BIDIRECTIONAL LMS ALGORITHM

We now consider the robustness of the bidirectional LMS algorithm to imperfect initialization and knowledge of the maximum Doppler frequency and SNR through MSIE results. It is assumed that the algorithm is run with $\mu_{\text{opt}}$ computed through (21) and (22) by using the estimated parameters under consideration. In order to estimate the unknown parameters of interest in this section, a sequence of $L_T$ pilot symbols chosen from $A$ in an independent and identical fashion is employed prior to each of the transmitted block of length $L$.

A. Effects of Imperfect Doppler and SNR Information

In order to estimate the unknown maximum Doppler frequency $f_d$, we modify the least-squares (LS) approach given in [14] which chooses an estimate $\hat{f}_d$ minimizing the following cost function

$$F(f_d) = \frac{1}{Q L_c} \sum_{q=1}^{Q} \sum_{m=0}^{L_c-1} \sum_{l=1}^{L_T-1} \left| \frac{\hat{K}_{q,m}(l)}{K_{q,m}(0)} - \frac{r(l; f_d)}{r(0; f_d)} \right|^2$$

(23)

assuming that $f_d$ does not change during $Q$ blocks. Note that the cost function in (23) is different from the one given [14] in that the minimization is also over the independent fading taps. In (23), $r(\cdot; f_d)$ is the true autocorrelation value and $\hat{K}_{q,m}(l)$ is given as

$$\hat{K}_{q,m}(l) = \frac{1}{L_T - l} \sum_{k=0}^{L_T-l} \hat{f}_{k,m}^q \hat{f}_{k+l,m}^q$$

(24)

where $\hat{f}_{k,m}^q$ is LS estimate of $f_{k,m}$ derived from the pilot symbols in the $q$-th block which does not need $f_d$ and the SNR.

We prefer the ML approach [15] to estimate the unknown SNR as follows

$$\hat{\gamma} = \arg\max_{\gamma} \ln p(y; \gamma | A) = \arg\max_{\gamma} \left\{ -\ln |R_y| - y^H R_y^{-1} y \right\}$$

(25)
where \( p(y; \gamma | A) \) is the probability density function of \( y = [y_0 \ldots y_{L-1}]^T \) and \( R_y \) is

\[
R_y = E\{yy^H\} = AR_fA^H + \frac{E_s}{\gamma} I
\]

(26)

where \( A = \text{diag}\{a_0, \ldots, a_{L-1}\} \) and \( R_f \) is the channel autocorrelation matrix formed by using the estimate of \( f_d \) associated with (23).

For numerical evaluation, we assume a block of \( L = 200 \) BPSK symbols over a 2-tap Rayleigh fading channel with Jakes’ spectrum, \( f_d = 100 \) Hz, \( T_s = 1 \) ms, \( Q = 1 \) and \( \hat{f}_d \in [0, 500] \) Hz. The estimates \( \hat{f}_d \) and \( \hat{\gamma} \) associated with (23) and (25), respectively, are used to calculate the 31-tap MMSE filter together with the optimal step-size values for the unidirectional and bidirectional LMS algorithms. The algorithms run with these findings and the corresponding MSIE performances for various training length choices are depicted in Fig. 3. We observe that the performance degradation in the bidirectional LMS due to the estimated Doppler and SNR is small especially for \( \gamma > 6 \) dB and \( L_T = 20 \). Since the statistics available through the independent channel taps are also incorporated into (23) as an improvement over [14], we do not need \( Q > 1 \) which achieves no significant performance gain.

B. Effect of Imperfect Initialization

In practical systems of interest, there is no perfect information on the fading vector at the beginning and end of the transmitted block. We therefore consider to initialize the bidirectional LMS algorithm either with the zero vector or the LS estimate employing only \( L_c \) pilot symbols, i.e., \( L_T = L_c \), in order to investigate the associated performance under these stringent conditions. The associated Monte Carlo results for various data block lengths over a 2-tap Rayleigh fading channel with the Jakes’ spectrum and \( f_d T_s = 0.01 \) are presented in Fig. 4. We observe that zero initialization is sufficient for relatively long but still practical data blocks and that the LS initialization with \( L_T = 2 \) achieves a satisfactory performance even for the short blocks especially around the optimal step-size, i.e., minimum point of the MSIE curves.

V. AN APPLICATION: ITERATIVE CHANNEL ESTIMATION WITH THE BIDIRECTIONAL LMS

As a more realistic application, we consider a coded system in which the unknown channel is estimated iteratively by employing the soft decisions on the coded symbols, as suggested in [16]. We propose to
employ the bidirectional LMS algorithm in order to achieve a BER performance similar to that with the MMSE filter with significantly less computational complexity.

At the transmitter, a set of $L_d$ binary symbols are first encoded by a channel code, interleaved, modulated and then multiplexed with a set of known pilot symbols which are inserted into the stream with a period of $M$ symbols [17]. At the receiver, an initial estimate of the unknown channel is obtained by an optimal MMSE filter employing the pilots only. This estimate is then refined through iterations by employing the soft estimates of the coded symbols provided by the soft decoder, as well as the pilot symbols, in all of the estimation algorithms under consideration (see [16] for details).

We depict the experimental BER results in Fig. 5 assuming $L_d = 98$, $M = \{11, 21\}$, BPSK modulation and a convolutional code with the generator $(1, 5/7)_8$ over an equal-power 2-tap Rayleigh fading channel with the Jakes’ spectrum and $f_d T_s = 0.01$. We choose the step-size values optimally on a trial and error basis, and the number of channel estimation iterations to be 3 for $M = 11$ and 5 for $M = 21$ both of which provide satisfactory convergence. We observe that the BER results are significantly improved through iterative estimation of the unknown channel for which the performance of the perfectly initialized bidirectional LMS algorithm is very close to that of the 21-tap MMSE filter and is much better than that of the unidirectional LMS algorithm. We also observe that the imperfect initialization of the bidirectional LMS algorithm with the estimates from the previous channel estimation iteration does not cause any significant performance degradation as compared to perfect initialization, similar to the results of Section IV-B. As a final remark, when we decrease the number of pilot symbols by choosing $M = 21$ instead of $M = 11$, although the BER performance of the MMSE filter employing the pilots only deteriorates by approximately 2 dB, the bidirectional LMS algorithm with imperfect initialization achieves almost the same BER performance at the expense of increased but still practical number of channel estimation iterations.

VI. SUMMARY

A bidirectional LMS algorithm is considered and analyzed over fast frequency-selective time-varying channels. The tracking performance of the bidirectional LMS is shown to be very close to that of the optimal MMSE filter in some settings of practical interest, and remarkably better than that of the conventional LMS algorithm. A step-size dependent steady-state MSE together with the optimal step-size expressions are derived in order to provide a theoretical analysis, and the corresponding theoretical
results show a good match to the experimental ones most of the time. The algorithm is also shown to be robust to imperfect initialization together with noisy Doppler and SNR information, and achieves BER results very close to that of the MMSE filter in various scenarios.

REFERENCES


Fig. 1. Theoretical and experimental normalized MSIE over a frequency-selective Rayleigh fading channel of length \( L = 100 \) with \( L_c = 2 \) tap at \( \gamma = 10 \) dB for \( f_dT_s = 0.01 \) and \( \gamma = 4 \) dB for \( f_dT_s = 0.02 \).

Fig. 2. Theoretical and experimental normalized MSIE associated with the optimal step-size over a frequency-selective Rayleigh fading channel of length \( L = 100 \) with \( L_c = \{2, 4\} \) tap and \( f_dT_s = 0.01 \).
<table>
<thead>
<tr>
<th>f_dT_s</th>
<th>Experiment</th>
<th>Theory</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.01</td>
<td>0.060 0.060 0.070 0.080 0.090 0.100</td>
<td>0.056 0.062 0.069 0.076 0.084 0.092</td>
</tr>
<tr>
<td>0.02</td>
<td>0.090 0.100 0.110 0.130 0.140 0.150</td>
<td>0.089 0.100 0.110 0.122 0.134 0.146</td>
</tr>
</tbody>
</table>

Fig. 3. MSIE for UniLMS, BiLMS and 3-tap MMSE for known and estimated Doppler and SNR over a 2-tap Rayleigh fading channel with Jakes’ spectrum where f_d = 100 Hz and T_s = 0.1 ms.
Fig. 4. MSIE for BiLMS with zero and LS initializations together with perfectly initialized UniLMS over 2-tap Rayleigh fading channel with Jakes’ spectrum and $f_d T_s = 0.01$ at $\gamma = 10$ dB.

Fig. 5. BER for BiLMS, UniLMS and MMSE with $M = \{11, 21\}$ over a 2-tap Rayleigh fading ISI channel with $f_d T_s = 0.01$. The number of channel estimation iterations for BiLMS and UniLMS is 3 for $M = 11$ and 5 for $M = 21$. 