Blind Channel Identification for Direct-Sequence Spread-Spectrum Systems

Dennis L. Goeckel[1], Alfred O. Hero III[1], and Wayne E. Stark[1]
EECS Department
University of Michigan
Ann Arbor, MI 48109

Abstract

Channel identification for a binary phase-shift keyed (BPSK) direct-sequence spread-spectrum (DS/SS) system operating over a fading channel with sampling at the chip rate is considered in this work. The system is mapped to a discrete oversampled system, thereby allowing channel identification via second order statistics under a few nonrestrictive conditions. Using the method of subchannel response matching (SRM), the offline solution to this channel identification problem involves the determination of the eigenvector corresponding to the minimum eigenvalue of a matrix that depends on the correlation statistics of the samples of the received signal. A low complexity stochastic gradient method for finding this eigenvalue adaptively is derived and a convergence analysis under a few weak assumptions presented. For comparison, a method that utilizes trellis searching for joint data and channel identification when the system is not oversampled is extended in an obvious way to oversampled systems and a different adaptive algorithm developed than has been used in the past. Numerical results in the form of channel estimation error are obtained for the case when the spreading code is unknown but periodic with period equal to the symbol period.

1 Introduction

In this paper, channel identification for uncoded single-user direct-sequence spread spectrum (DS/SS) systems is considered. The systems will be characterized by symbol transmission period $[0, T_s]$ and employ a rectangular chip pulse restricted to $[0, T_c]$, $T_c \ll T_s$, where the processing gain is defined as $N = T_s/T_c$. Although it is not suboptimal for an ideal system on an additive white Gaussian noise (AWGN) channel with no intersymbol interference (ISI) to correlate with the full spreading waveform and sample only once per symbol period, in actual application the sampling is normally done at a much higher rate. In many cases, sampling is done at the chip rate and, after multiplication by the appropriate spreading sequence bit, the results summed to form the decision statistic for a given information bit, which then can be thresholded to determine an information estimate. For an ISI system that is oversampled, however, it is no longer possible to obtain a sufficient statistic for metric calculation of a given path of the Viterbi algorithm, the optimal sequence estimator for a system corrupted by ISI, by multiplying the samples within a symbol period by a transmitter spreading coefficient and then summing. The optimal combining of the samples in a given symbol period depends on the channel impulse response at each *sample* period, not just at the symbol period. This is the motivation for channel identification of oversampled systems at the sample period.

Since the DS/SS system is oversampled in continuous time, one predicts that it can be mapped to an oversampled system in discrete time. This is important as classical blind equalization results show that unique channel identification is not possible (unless the channel is known to have minumum phase) through the use of second order statistics for non-oversampled discrete systems when the input sequence is ergodic and wide-sense stationary [1]. Recent work [2], however, has shown that if the input symbols to the channel in a digital communications system are independent and identically distributed (IID) and the output of the channel is sampled at a rate that is a multiple of the input symbol rate, the oversampling of the cyclostationary input waveform makes the sampled process also cyclostationary, thus allowing for blind channel identification based on only second order statistics. Because estimation of second order statistics requires fewer samples than that of higher order statistics for a given level of accuracy, one expects algorithms based on second order statistics to exhibit faster convergence.

2 DS/SS Systems over ISI Channels

The input data bits, denoted by the sequence (s_k), are assumed to be IID. The spreading sequence will be denoted (a_l). Since binary phase-shift keying is being employed, the output of the transmitter can be expressed as (using complex baseband notation throughout) $x(t) = \sqrt{\frac{E_b}{NT_c}}s(t)a(t)e^{j\theta}$ where E_b is the energy per data bit, θ is the transmitter phase (which will be assumed to be worked into the channel response and suppressed from here forward), $s(t) = \sum_{k=-\infty}^{\infty} s_k p_{T_s}(t - kT_s)$, and $a(t) = \sum_{l=-\infty}^{\infty} a_l p_{T_c}(t - lT_c)$, where $p_T(t)$ is defined to be unit-amplitude pulse that is nonzero on $[0, T]$.

The time-nonselective, frequency-selective continuous channel response is denoted by $g(t)$. The received signal is given by $r(t) = g(t) * x(t) + n(t)$ where $n(t)$ is com-

[1]This work was supported by a National Science Foundation Graduate Fellowship and by the National Science Foundation under grants BCS-90243370 and NCR-9115969.

plex Gaussian channel noise. The receiver, using a chip-matched filter matched to the rectangular pulse, calculates the complex statistic

$$y_n(k) = y(kT_s + nT_c) = \int_{kT_s+nT_c}^{kT_s+(n+1)T_c} r(t)dt.$$

It is straightforward to show that the response g_m at sample time mT_c due to a unit amplitude pulse on $[0, T_c]$ is given by

$$g_m = \int_{-\infty}^{\infty} g(\tau) \int_0^{T_c} p_{T_c}(t + mT_c - \tau)dtd\tau$$

which implies by linearity that h_m^k, the response at sample time $kT_s + mT_c$ due to the spreading waveform $\hat{a}^k(t) = \sum_{p=0}^{N-1} \hat{a}_p^k p_{T_c}(t - pT_c - kT_s)$ on $[kT_s, (k+1)T_s]$, is given by

$$h_m^k = \sum_{p=0}^{N-1} \hat{a}_p^k g_{m-p} = \hat{a}_m^k * g_m$$

where one period of the spreading sequence $\{\hat{a}_p^k, p = 0, \ldots, N-1\}$ is defined in the obvious way.

Thus, it has been established that the overall channel response can be written as a convolution of the spreading sequence for a period and an effective channel response. Note that the index k that was carried through the calculation allows for the spreading sequence to vary over each symbol period. However, for the work considered in this paper, it is assumed that the spreading sequence is unknown and periodic in the symbol period. For this reason, the superscript k will be dropped from h in succeeding sections.

3 Adaptive Algorithms Using Subchannel Response Matching

3.1 Subchannel Response Matching

Since the SRM algorithm, an offline algorithm for channel identification in oversampled systems, is explained thoroughly elsewhere [3], the aim is only to set up the problem and present the final solution.

Let the impulse response $\{h_m, m = 0, 1, \ldots, LN-1\}$ of the overall channel be of duration LN samples. Consider the division of the oversampled output into N subchannels, one subchannel for each offset T_c from the input sample time. Each subchannel operates at sample rate T_s on the IID channel input symbols, and the output at time k of the n^{th} subchannel can be written as

$$y_n(k) = \sum_{l=0}^{L-1} h_n(l)s(k-l) + j_n(k)$$

where $h_n(l) = h_{lN+n}$ is the l^{th} sample of the n^{th} subchannel of the impulse response, and $j_n(k)$ is the zero-mean observation noise variance σ^2 on this sample.

The basic idea of the subchannel response matching algorithm is that since any two subchannels m and n have the same input, $y_m(k) * h_n(k) = y_n(k) * h_m(k)$. Thus, for a given pair, it is sought to match the output from the application of an estimate of channel n's impulse response on m's output to the application of an estimate of m's impulse response on n's output. If this is done in the mean squared sense for all distinct pairs of channels, one seeks to minimize

$$\mathcal{E}(\underline{\hat{h}}) = \sum_{m=0}^{N-2} \sum_{n=m+1}^{N-1} E[|\hat{h}_m(k) * y_n(k) - \hat{h}_n(k) * y_m(k)|^2]$$

subject to the constraint that $||\underline{\hat{h}}||^2 = 1$ where $\underline{\hat{h}} = [\hat{h}_0(0) \ldots \hat{h}_0(L-1)\ \hat{h}_1(0) \ldots \hat{h}_{N-1}(L-1)]^T$.

Define $\underline{y}_n(k) = [y_n(k)\ y_n(k-1) \ldots y_n(k-L+1)]^T$ and $R_{mn} = E[\underline{y}_m(k)\underline{y}_n^H(k)]$. The offline problem can be rewritten as: find the conjugate of the eigenvector corresponding to the minimum eigenvalue of the matrix S where

$$S = [I \otimes \sum_{n=0}^{N-1} R_{nn}] - V$$

$$V = \begin{bmatrix} R_{00} & R_{10} & \ldots & R_{(N-1)0} \\ R_{01} & R_{11} & & \vdots \\ \vdots & & \ddots & \\ R_{0(N-1)} & \ldots & & R_{(N-1)(N-1)} \end{bmatrix}.$$

Thus, S consists of the difference of two matrices, one that consists of L by L identical blocks on the diagonal and one that is seen to be *not* an outer product when considered carefully. Note that if the channel is identifiable via second order statistics [4], there will be a unique minimum eigenvalue and the conjugate of its eigenvector will be the unique solution. In the case of a noiseless system, S will be a positive semidefinite matrix, denoted S_Z, with a nullspace of dimension one. In a noisy system, $S = S_Z + (N-1)\sigma^2 I$ and thus will be positive definite.

3.2 Adaptive Algorithm

Although finding the eigenvector associated with the minimum eigenvalue of a matrix corresponds closely to Pisarenko's harmonic retrieval method [5], S is not simply an outer product of the observed vector unless $N = 2$, so Pisarenko's results cannot be applied. As an alternative, a low complexity stochastic gradient method that can be performed online is derived. The error signal considered is the Rayleigh quotient of S^*

$$e(\underline{\hat{h}}) = \frac{\underline{\hat{h}}^H S^* \underline{\hat{h}}}{||\underline{\hat{h}}||^2} = \frac{\underline{\hat{h}}^T S \underline{\hat{h}}^*}{||\underline{\hat{h}}||^2} \quad (1)$$

which is minimized at the conjugate of the miminum eigenvector of S.

Since the adaptive algorithm is implemented with sample averages as opposed to ensemble averages, the matrix S_k is introduced which is the estimate of S at time k based on the observed channel outputs at time $k, k-1, \ldots, k-L+1$ (i.e. S with the expectations removed). This definition suggests an empirical objective function at step k given by

$$J_k = \sum_{l=0}^{k} \lambda^{(k-l)} \frac{\underline{\hat{h}}_l^T S_l \underline{\hat{h}}_l^*}{||\underline{\hat{h}}_l||^2}$$

where $\lambda \in [0,1]$ is a forgetting factor. For the rest of this paper, it will be assumed that $\lambda = 0$, leading to the simpler objective function

$$J(\underline{\hat{h}}_k) = e(\underline{\hat{h}}_k) = \frac{\underline{\hat{h}}_k^T S_k \underline{\hat{h}}_k^*}{||\underline{\hat{h}}_k||^2}.$$

The standard update for the estimated channel, $\underline{\hat{h}}_k$, in the stochastic gradient algorithm is then given by $\underline{\hat{h}}_k = \underline{\hat{h}}_{k-1} - \mu \bigtriangledown e(\underline{\hat{h}}_{k-1})$ where μ is a user specified gain factor. Thus, the key is the efficient calculation of the gradient of $e(\underline{h})$. Direct calculation using the definition in equation (1) yields

$$\bigtriangledown e(\underline{h}) = 2\frac{(||\underline{h}||^2 S_k^* \underline{h} - \underline{h}^T S_k \underline{h}^* I \underline{h})}{||\underline{h}||^4}.$$

Due to the complicated form of S_k, "brute force" computation of the above requires $O(L^2N^2)$ operations because of the matrix-vector multiplication $S_k^*\underline{h}$. However, noting the conjugate symmetry of S, $\bigtriangledown(\underline{h}^T S_k \underline{h}^*) = 2S_k^*\underline{h}$, and it can be shown that

$$\begin{aligned}\frac{\partial(\underline{h}^T S_k \underline{h}^*)}{\partial h_m(l)} &= 2(\sum_{i=0}^{L-1} h_m(i) \sum_{n=0}^{N-1} y_n(k-i) y_n^*(k-l) \\ &\quad - \sum_{j=0}^{L-1} y_m(k-j) \sum_{n=0}^{N-1} h_n(j) y_n^*(k-l)).\end{aligned}$$

The analysis of the number of operations now goes as follows: the two sums over n must be done for each l and i (or j), thus leading to $O(L^2N)$ operations; given these sums, $\bigtriangledown(\underline{h}^T S_k \underline{h}^*)$ can be found in $O(LN)$ inner products of vectors of length L. Thus, the total number of operations required to obtain $S_k^*\underline{h}$ (and thus $\bigtriangledown e(\underline{h})$) is $O(L^2N)$, leading to a considerable savings if the processing gain is large.

3.3 Convergence Analysis

A gain factor μ must be specified for the stochastic gradient algorithm that guarantees convergence of $\underline{\hat{h}}_k$ to the desired solution $\underline{h}_{\text{opt}}$. Convergence of the stochastic gradient search for a particular sample path cannot be guaranteed; as a first approximation, convergence conditions are derived in this section for the gradient search assuming the matrix S is measured with no error. Even under this assumption, global convergence is still difficult to demonstrate due to the complexity of the error surface, including flat portions leading to algorithm stagnation at each of the eigenvectors. Therefore, a desireable goal is to choose μ such that the algorithm will converge to the correct solution when it is near the minimizing eigenvector.

The error at iteration k of the algorithm is given by

$$e_k = e(\underline{\hat{h}}_k) = \frac{\underline{\hat{h}}_k^T S \underline{\hat{h}}_k^*}{||\underline{\hat{h}}_k||^2} = \frac{\underline{\hat{h}}_k^T S_Z \underline{\hat{h}}_k^*}{||\underline{\hat{h}}_k||^2} + (N-1)\sigma^2 \quad (2)$$

In the following, the term involving σ^2 will be ignored as $(N-1)\sigma^2$ represents a constant residual error at the minimizing solution.

Assuming a noiseless system ($\sigma^2 = 0$), substituting in $\underline{\hat{h}}_k = \underline{\hat{h}}_{k-1} - \mu \bigtriangledown e(\underline{\hat{h}}_{k-1})$, and retaining only terms that are linear in e_{k-1} (since $e_{k-1} \ll 1$ near $\underline{h}_{\text{opt}}$ in a noiseless system), one obtains

$$e_k = \frac{\underline{\hat{h}}_{k-1}^T S_Z \underline{\hat{h}}_{k-1}^*}{||\underline{\hat{h}}_k||^2} - 2\mu \frac{\underline{\hat{h}}_{k-1}^T S_Z^2 \underline{\hat{h}}_{k-1}^*}{||\underline{\hat{h}}_k||^2 ||\underline{\hat{h}}_{k-1}||^2} + \mu^2 \frac{\underline{\hat{h}}_{k-1}^T S_Z^3 \underline{\hat{h}}_{k-1}^*}{||\underline{\hat{h}}_k||^2 ||\underline{\hat{h}}_{k-1}||^4}$$

Applying the modal decomposition $S_Z = \sum_{i=0}^{LN-1} \lambda_i \underline{v}_i \underline{v}_i^H$ where $\underline{v}_i$ is the i^{th} eigenvector of S_Z, convergence of e_k can be achieved by the convergence of each of the modes. For the i^{th} mode, one obtains

$$e_k^i = \left(\frac{\lambda_i ||\underline{\hat{h}}_{k-1}^T \underline{v}_i||^2}{||\underline{\hat{h}}_{k-1}||^2}\right) \left(\frac{||\underline{\hat{h}}_{k-1}||^2}{||\underline{\hat{h}}_k||^2}\right) \left(1 - 2\frac{\mu\lambda_i}{||\underline{\hat{h}}_{k-1}||^2} + \frac{\mu^2\lambda_i^2}{||\underline{\hat{h}}_{k-1}||^4}\right)$$

$$e_k^i \leq e_{k-1}^i \left(1 - \frac{\mu\lambda_i}{||\underline{\hat{h}}_{k-1}||^2}\right)^2 \quad (3)$$

where the last inequality comes from the fact that $||\underline{\hat{h}}_k||^2$ is nondecreasing as a function of k. This nondecreasing property can be derived as follows: note from the definition of $\bigtriangledown e(\underline{h})$ that $\underline{\hat{h}}_{k-1}^H(\underline{\hat{h}}_k - \underline{\hat{h}}_{k-1}) = \underline{\hat{h}}_{k-1}^H \bigtriangledown e(\underline{\hat{h}}_{k-1}) = 0$ which then implies

$$\begin{aligned}||\underline{\hat{h}}_k||^2 &= ||\underline{\hat{h}}_k - \underline{\hat{h}}_{k-1} + \underline{\hat{h}}_{k-1}||^2 \\ &= ||\underline{\hat{h}}_k - \underline{\hat{h}}_{k-1}||^2 + ||\underline{\hat{h}}_{k-1}||^2 \geq ||\underline{\hat{h}}_{k-1}||^2\end{aligned}$$

thus establishing the nondecreasing property.

If the initial guess is chosen on the unit circle (i.e. $||\underline{\hat{h}}_0^2|| = 1$) and is sufficiently close to the solution $\underline{h}_{\text{opt}}$, the nondecreasing property and (3) can be used to show that convergence occurs in all of the modes if $\mu < \frac{2}{\lambda_{\max}}$. Since $\lambda_{\max}$ is difficult to obtain, the conservative estimate $\mu < \frac{2}{tr(S_Z)}$ is used instead, where $tr(S_Z) = L(N-1)\sum_{i=0}^{N-1}(E[y_i(k)^2] - E[j_i^2(k)])$, which can be readily estimated if the signal-to-noise ratio (SNR) is known.

4 Trellis-Searching Algorithm

In [6], an algorithm for blind equalization is introduced. It extends the Viterbi algorithm, the optimal sequence estimator for channels with known ISI, to perform blind equalization by allowing M survivors per state and peforming channel estimation for each survivor at each step using an LMS algorithm. To extend this algorithm to oversampled systems, a given branch metric is now calculated as the total Euclidean distance between the samples in a symbol period and the samples predicted along that branch in the trellis.

The channel update algorithm is also altered. The minimum mean squared error solution for the channel update is given by $h_n^{\text{opt}}(l) = R_n(l) = E[y_n(k)s(k-l)]$ which is easily approximated. Thus, instead of using a gradient search algorithm, this ensemble average is approximated by a sample mean $\hat{R}_n(l) = \frac{1}{T}\sum_{k=0}^{T-1} y_n(k)\hat{s}(k-l)$ where $\hat{s}(k)$ is the estimate of the k^{th} data bit for the path under consideration.

5 Results and Conclusions

The complexity of the two algorithms must be considered in evaluating performance. The subchannel response matching (SRM) method, as mentioned earlier, involves $O(L^2N)$ operations per symbol to perform one iteration of the gradient search. The trellis searching algorithm discussed requires $O(LM2^LN)$ operations per symbol, significantly more than the SRM estimator, but also decodes the bits as it runs with no need for an additional sequence estimator.

For the case $N = 4, L = 5$, two channels labelled "good" and "bad" are considered with zeroes as specified below. One hundred trials, each with a random initial guess, were averaged over to obtain Figures 1-3 on the following page. The gain factor from section 3.3 for these channels is approximately $\mu = 0.07$. The normallized mean squared error (NMSE) is used as the figure of merit and is defined for a channel $\underline{\hat{h}}$ trying to identify $\underline{h}$ as $NMSE = \|\frac{\underline{\hat{h}}}{\|\underline{\hat{h}}\|^2} - \frac{\underline{h}}{\|\underline{h}\|^2}\|^2$.

Subchannel	*Good*	*Bad*
0	0.353 ± 0.353 j 0.259 ± 0.150 j	0.636 ±0.636j 0.692 ± 0.400j
1	-0.104 ± 0.590 j 1.691 ± 0.615 j	-0.121 ± 0.689j 1.122 ± 0.410j
2	0.069 ± 0.393 j -1.221 ± 1.455 j	0.148 ± 0.837j 1.034 ± 0.376j
3	0.752 ± 2.067 j -0.800 ± 1.385 j	0.393 ± 1.081j -0.525 ± 0.909j

One key item not displayed in Figures 1-3 for the SRM algorithm is that the plots are dominated by a few sample paths that take a long time to converge. Since the complexity of the SRM algorithm is low, this suggests using multiple starting points for one trial and choosing the one with the lowest e_k (calculated at each step to form the gradient) at the k^{th} iteration. Using a small number of starting points would greatly increase performance with complexity still less than the trellis-searching algorithm. An example of the improved performance (averaged over 30 sample paths) is shown in Figure 4.

6 Future Work

There are a few key areas still requiring work on this problem. The convergence analysis presented for the stochastic gradient algorithm must be updated to take into account variation of the sample data from the ensemble averages, and convergence in terms of mean and variance needs to be considered. Global convergence conditions must also be considered more closely. It would also be nice to have an algorithm with faster convergence properties; in particular, it would be desireable to find a fast method for computation of the Hessian so that a Newton-type algorithm could be efficiently implemented.

References

[1] S. Haykin, Ed., *Blind Deconvolution,* PTR Prentice-Hall, Englewood Cliffs, 1994.

[2] L. Tong, G. Xu, and T. Kailath, "Blind Identification and Equalization Based on Second-Order Statistics: A Time Domain Approach," *IEEE Transactions on Information Theory,* Vol.40, pp. 340-349, March 1994.

[3] S. Schell, D. Smith, and S. Roy, "Blind Channel Identification using Subchannel Response Matching," *Proceedings of the 1994 Conference on Information Science and Systems,* Vol. 2, pp. 858-862.

[4] L. Baccala and S. Roy, "A New Blind Time-domain Channel Identification Method Based on Cyclostationarity," *IEEE Signal Processing Letters,* pp. 89-91, June 1994.

[5] G. Mathew, S. Dasgupta, and V. U. Reddy, "Improved Newton-Type Algorithm for Adaptive Implementation of Pisarenko's Harmonic Retrieval Method and Its Convergence Analysis," *IEEE Transactions on Signal Processing,* Vol. 42, pp. 434-437, February 1994.

[6] N. Seshadri, "Joint Data and Channel Estimation Using Blind Trellis Search Techniques," *IEEE Transactions on Communications,* Vol. 42, pp. 1000-1011, February/March/April 1994.

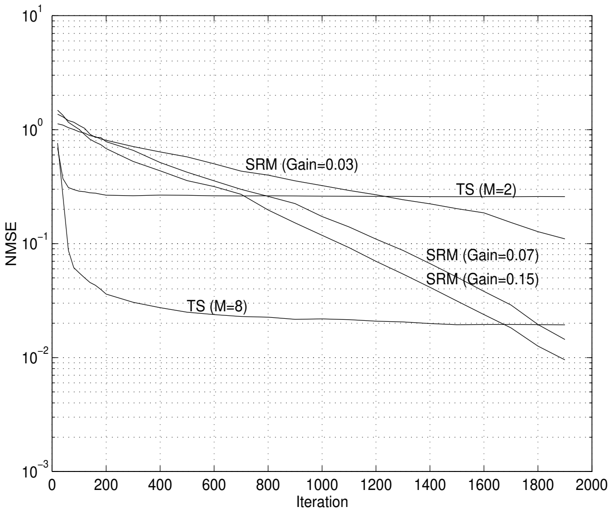

Figure 1: Normallized Mean Squared Error for the "Good" Channel, SNR = 18 dB: The gain factor of $\mu = 0.07$ appears too conservative.

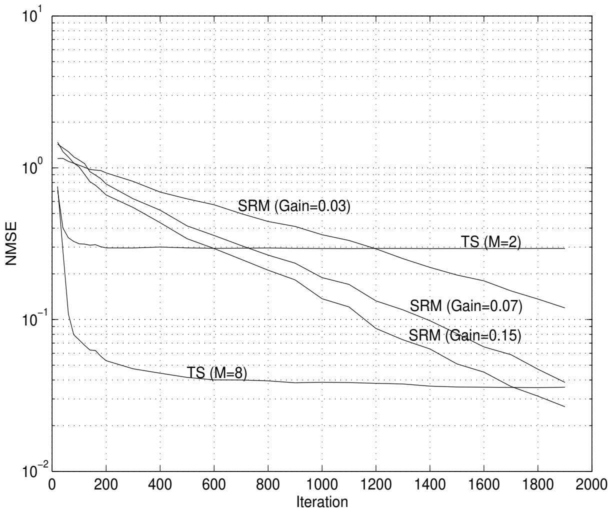

Figure 2: Normallized Mean Squared Error for the "Good" Channel, SNR = 8 dB: The drop in SNR of 10 dB affects both algorithms the same.

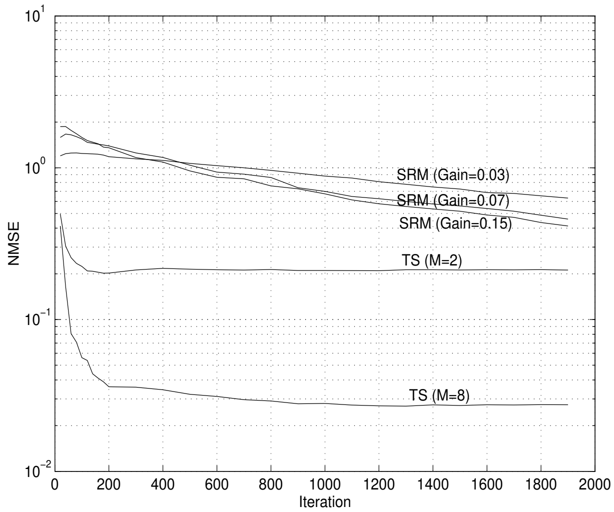

Figure 3: Normallized Mean Squared Error for the "Bad" Channel, SNR = 18 dB: The "bad" channel adversely affects SRM but not TS.

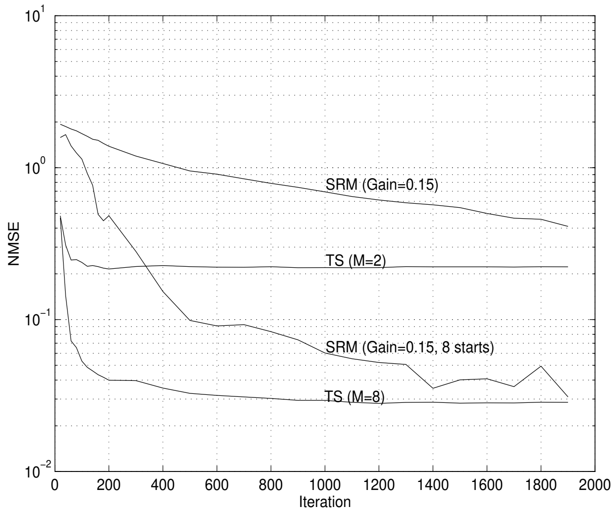

Figure 4: Normallized Mean Squared Error for the "Bad" Channel, SNR = 18 dB: The improved SRM algorithm shows a large gain vs. SRM on the "bad" channel.