OPTIMAL UPDATE PROFILING FOR STEEPEST DESCENT ALGORITHMS

Mile Milisavljević

Cicada Semiconductor Corporation
811 Barton Springs Road, Ste. 550
Austin, Texas 78704
mm@cicada-semi.com

ABSTRACT

In this paper, the methods for use of prior information about the operating environment, in improving adaptive filter convergence properties are discussed. More concretely, the gain selection, profiling and scheduling in steepest descent algorithms is treated in detail. Optimality criteria for steepest descent gains is derived, its impact on the residual mean square error is analyzed, and results are extended to affine filtering algorithms. Further, it is demonstrated that optimal update gains are closely connected to the Newton's adaptation algorithm. It is demonstrated that with no added complexity a substantial increase of convergence rate of steepest descent algorithms can be achieved.

1. INTRODUCTION

Due to their low implementational cost and good numerical properties, steepest descent techniques play important role in modern signal processing applications. They can be applied in adaptive elements such as feed-forward equalizer (FFE), decision-feedback equalizer (DFE), near end crosstalk (NEXT) cancellers (NC), and echo cancellers (EC).

A typical application environment for steepest descent techniques in NC and EC is given in Figure 1.

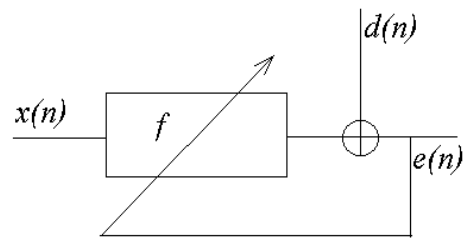

Figure 1: Common adaptive system adapting to minimize the residual error.

The convergence of the iterative algorithm is governed by the difference equation:

$$\begin{aligned} \mathbf{v}_{n+1} &= (\mathbf{I} - \mu \mathbf{R}_{xx})\mathbf{v}_n \\ \mathbf{v}_n &= \mathbf{f}_n - \mathbf{f}^* \end{aligned} \tag{1}$$

where $\mathbf{f}$ represents the filter coefficient vector, $\mathbf{v}$ represent sthe filter coefficient error, and $\mathbf{R}_{xx}$ denotes the autocorrelation matrix of the signal . Relationship (1) indicates that in order to ensure the stability of the algorithm one needs to choose the adaptation step to be within bounds: $0 < \mu < \frac{2}{\lambda_{max}(\mathbf{R}_x)}$. Larger values of the adaptation step lead to faster convergence, but increase the residual error and may potentially render the algorithm unstable. In practical applications, such problems are usually resolved with adaptive change of the adaptation step size [1], or with a conservative choice of fixed step size[2].

More concretely, the convergence of the steepest descent algorithm is governed by the smallest eigenvalue of the correlation matrix [7] via:

$$\tau_n \approx \frac{1}{\mu \lambda_n}, \tag{2}$$

where, τ_n is the time constant corresponding to the eigenvalue λ_n. Equation (2) indicates that while the adaptation step-size is limited by the reciprocal of the largest eigenvalue, the smallest eigenvalue is the one that governs the convergence of the slowest mode. The relationships (1) and (2) expose the fundamental problem of application of steepest descent procedures to problems with large eigenvalue disparity.

In this paper the problem of update profiling and scheduling of steepest descent algorithms is treated. It is shown that with some prior knowledge, and **without any added computational complexity**, performance of steepest descent algorithms can be substantially improved. The improvement is commensurate to the eigenvalue spread of the correlation matrix. Moreover, the results obtained for traditional steepest descent algorithms are straightforwardly extended to the class of affine update algorithms.

2. FORMULATION OF THE GRADED UPDATES PROBLEM

A practically more appealing version of steepest descent procedure is the least-mean-square (LMS) algorithm. In the LMS algorithm the gradient is substituted by its instantaneous estimate:

$$\nabla = \varepsilon_k \mathbf{x}_k = \mathbf{r}_{dx,\ k} - \mathbf{x}_k \mathbf{x}_k' \mathbf{f} \approx (d_k - \mathbf{x}_k' \mathbf{f}) \mathbf{x}_k \qquad (3)$$

Inclusion of the 3 in the gradient update yields:

$$\mathbf{f}_{n+1} = \mathbf{f}_n + \mu \varepsilon_n \mathbf{x_n}. \qquad (4)$$

Definition: *Graded Update Gains* (or Graded Updates) refers to the application of different gain to every coefficient (tap) in the vector implementation of the steepest descent based algorithm.

In graded update version therefore, every coefficient of the filter $\mathbf{f}$ has a different update rate μ_k. Thus (5) changes to:

$$\mathbf{f}_{n+1} = \mathbf{f}_n + \mathbf{M} \varepsilon_n \mathbf{x_n}. \qquad (5)$$

where $\mathbf{M}$ is the diagonal matrix:

$$\mathbf{M} = \begin{bmatrix} \mu_1 & 0 & \dots & 0 \\ 0 & \mu_2 & & 0 \\ & & \ddots & \\ 0 & 0 & \dots & \mu_N \end{bmatrix} \qquad (6)$$

and the error evolution equation (1) becomes:

$$\mathbf{v}_{n+1} = (\mathbf{I} - \mathbf{M}\mathbf{R}_{xx})\mathbf{v}_n \qquad (7)$$

The main problem treated in this paper is one of finding optimal diagonal matrix $\mathbf{M}$ such that the descent rate is maximized. The application of graded updates in LMS filtering have been long used and considered. However, the methods of determination of graded update gains have been mostly heuristic. Most of authors dealing with this problem [4, 5] concentrated on the expected value of filter coefficients and adapted coefficients of larger magnitude with larger update gains. In this paper it is shown that contrary to the popular belief, optimal graded update gains may **only coincidentally** be connected to the expected value of coefficients.

Two flavors of the problem are treated: in Section 3, the problem of determining $\mathbf{M}$ with partial statistical knowledge is treated. In Section 4, more complete problem of full statistical knowledge is solved. Section 5 treats the extension of the presented methods to affine filtering gain selection. Section 6 treats the excess Mean Square Error (excess MSE) computation, while Section 7 presents some numerical examples.

3. MAXIMUM ENTROPY APPROACH - USE OF PARTIAL STATISTICAL KNOWLEDGE

The basic idea behind the solution of the graded updates problem is minimization of the expected error variance at every iteration. This optimization procedure yields (possibly time varying) set of gains $\mathbf{M}$ which allow fastest descent down the expected quadratic bowl.

In case when initial error statistics is unknown, or disregarded, the designer can then adopt the maximum entropy approach and assume white statistics of initial error. Let $\mathbf{Q}$ be the matrix that diagonalizes (7):

$$\tilde{\mathbf{v}}_{n+1} = \mathbf{Q}\mathbf{v}_{n+1} = \mathbf{Q}\left(\mathbf{I} - \mathbf{M}\mathbf{R}_{xx}\right)\mathbf{Q}'\mathbf{Q}\mathbf{v}_n = \mathbf{\Lambda}\tilde{\mathbf{v}}_n \qquad (8)$$

Proposition 1: If $\mathbf{x}$ is white (i.e. $E\{\mathbf{x}\mathbf{x}'\} = \sigma^2\mathbf{I}$, and $\mathbf{Q}$ is an orthogonal matrix, then $\tilde{\mathbf{x}} = \mathbf{Q}\mathbf{x}$ is also white.
Proof: $E\{\tilde{\mathbf{x}}\tilde{\mathbf{x}}'\} = \mathbf{Q}E\{\mathbf{x}\mathbf{x}'\}\mathbf{Q}' = \sigma^2\mathbf{Q}\mathbf{Q}' = \sigma^2\mathbf{I}$, QED.

Via Proposition 1, if $\mathbf{v}$ is assumed to be white, then $\tilde{\mathbf{v}}$ is also white. Expected value of the norm of $\tilde{\mathbf{v}}_{n+1}$ is then minimized when sum of squared eigenvalues (on the diagonal of) $\mathbf{\Lambda}$ is minimized, since:

$$\begin{aligned} E\{\tilde{\mathbf{v}}_{n+1}'\tilde{\mathbf{v}}_{n+1}\} = \; & E\{\mathbf{v}'\mathbf{\Lambda}^2\mathbf{v}\} = \\ & \textstyle\sum_{i=1}^N \lambda_i^2 E\{\tilde{v}_n^2(i)\} = \\ & \sigma_{\tilde{\mathbf{v}}_n}^2 \textstyle\sum_{i=1}^N \lambda_i^2 \end{aligned} \qquad (9)$$

The solution to this problem corresponds to the optimal $\mathbf{M}$ which is the diagonal approximation to the inverse of the $\mathbf{R}_{xx}$ in Frobenious norm:

$$\mathbf{M} = \text{argmin} \|\mathbf{I} - \mathbf{M}\mathbf{R}_{xx}\|_F \qquad (10)$$

Due to its specific structure, the problem can be represented as a set of column-wise optimization problems:

$$\begin{aligned} & \min \|\mathbf{e}_1 - \mathbf{M}(1,1)\mathbf{r}_1\|_2 \\ & \min \|\mathbf{e}_2 - \mathbf{M}(2,2)\mathbf{r}_2\|_2 \\ & \vdots \\ & \min \|\mathbf{e}_N - \mathbf{M}(N,N)\mathbf{r}_N\|_2 \end{aligned} \qquad (11)$$

Where $\mathbf{r}_i$ is the i^{th} column of $\mathbf{R}_{xx}$ and $\mathbf{e}_i$ is the i^{th} column of $\mathbf{I}$. Each of these optimization problems can be solved easily to yield the general solution:

$$\mathbf{M}(i,i) = \frac{\mathbf{e}_i'\mathbf{r}_i}{||\mathbf{r}_i||^2} = \frac{\mathbf{R}_{xx}(i,i)}{||\mathbf{r}_i||^2} \tag{12}$$

4. USE OF COMPLETE STATISTICAL KNOWLEDGE

In physical systems, the statistics of the initial error is often non-white. A good example is a design of Feed-Forward Equalizer (FFE) linear filter. FFE taps often exhibit the alternating coefficient signs (due to its high-pass nature). Hence at least initial correlation of the error vector is going to be non white. In this section the possibility of use of initial information on error statistics is explored.

To derive the optimal graded update gains matrix $\mathbf{M}$, from (7) first compute the dynamics of the filter error norm $\mathbf{v}_n'\mathbf{v}_n$:

$$\mathbf{v}_{n+1}'\mathbf{v}_{n+1} = \mathbf{v}_{n+1}'\left(\mathbf{I} - \mu\mathbf{R}_{xx}\right)'\left(\mathbf{I} - \mu\mathbf{R}_{xx}\right)\mathbf{v}_{n+1} \tag{13}$$

Using the fact that the trace of a scalar is equal to the scalar and linearity of the expectation and trace operators derivation follows:

$$\begin{aligned}
\sigma_v^2[n+1] &= E\{\mathbf{v}_{n+1}'\mathbf{v}_{n+1}\} \\
&= E\left\{\mathrm{Tr}\left\{\mathbf{v}_n'\left(\mathbf{I} - \mu\mathbf{R}_{xx}\right)'\left(\mathbf{I} - \mu\mathbf{R}_{xx}\right)\mathbf{v}_n\right\}\right\} \\
&= E\left\{\mathrm{Tr}\left\{\left(\mathbf{I} - \mu\mathbf{R}_{xx}\right)'\left(\mathbf{I} - \mu\mathbf{R}_{xx}\right)\mathbf{v}_n\mathbf{v}_n'\right\}\right\} \\
&= \mathrm{Tr}\left\{\left(\mathbf{I} - \mu\mathbf{R}_{xx}\right)'\left(\mathbf{I} - \mu\mathbf{R}_{xx}\right)\mathbf{R}_{vv}[n]\right\} \\
&= \mathrm{Tr}\left\{\mathbf{R}_{vv}\right\} - \\
&\quad \mathrm{Tr}\left\{\left(2\mathbf{R}_{xx}'\mathbf{M} - \mathbf{R}_{xx}\mathbf{R}_{xx}'\mathbf{M}\mathbf{M}\right)\mathbf{R}_{vv}[n]\right\}
\end{aligned} \tag{14}$$

Noting that $\mathrm{Tr}\left\{\mathbf{R}_{vv}\right\} = E\{\mathbf{v}_{n+1}'\mathbf{v}_{n+1}\} = \sigma_v^2[n]$, we form the problem of finding $\mathbf{M}$ as an optimization problem:

$$\max_{\mathbf{M}\text{ diagonal}} ||\sigma_v^2[n] - \sigma_v^2[n+1]|| \tag{15}$$

Maximization of the difference between two subsequent errors ensures the maximization of the convergence rate of the algorithm. Cost function in (15) can be rewritten:

$$\min_{\mathbf{M}\text{ diagonal}} \mathrm{Tr}\left\{\mathbf{R}_{vv}[n]\mathbf{R}_{xx}\mathbf{R}_{xx}'\mathbf{M}\mathbf{M} - 2\mathbf{R}_{vv}[n]\mathbf{R}_{xx}'\mathbf{M}\right\} \tag{16}$$

subject to $\mathbf{M}$ diagonal.

Proposition 2: Stationary points of (16) correspond to the global maxima.
Proof of the Proposition 2 is straightforward and relies on the positive definite nature of correlation matrices $\mathbf{R}_{xx}$ and $\mathbf{R}_{vv}$, and their product.

Differentiating (16) optimality criteria for $\mathbf{M}$ becomes:

$$2\mathbf{R}_{vv}\mathbf{R}_{xx}\left(\mathbf{I} - \mathbf{M}\mathbf{R}_{xx}\right) = \mathbf{0} \tag{17}$$

While this form reveals that the direction of graded updates should be to approximate the inverse of $\mathbf{R}_{xx}$ by a diagonal matrix, it is not clear how to incorporate the diagonality constraint.

Since the optimization is performed over a quadratic bowl (it is easy to expand the problem in (15) to its full quadratic equivalent by adding a positive constant), Proposition 3 follows:

Proposition 3: Consider the following optimization strategy for problem in (15): first find the optimal solution $\mathbf{M} = \mathbf{R}_{xx}^{-1}$, and then find the matrix $\mathbf{M}$ closest to the $\mathbf{R}_{xx}^{-1}$ in Frobenious sense. Such matrix will be the optimal constrained solution of problem in (15).

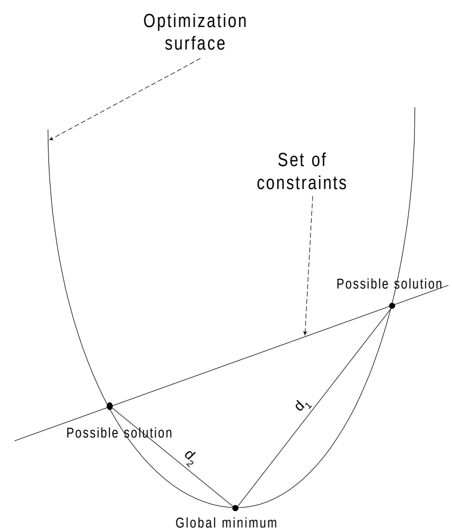

Figure 2: Illustration of optimization principle.

Proof: It is sufficient to consider one dimensional problem: $f(x) = ax^2 + bx + c$, for $a < 0$, and with minimum $x^* = -\frac{b}{2a}$. Now, consider two possible constrained solutions x_1 and x_2 with respective distances d_1 and d_2 from x^*, as in Figure 2. Assume also $|d_2| < |d_1|$. Then following holds:

$$\begin{aligned}
&f(x_1) - f(x_2) = f(x^* + d_1) - f(x^* + d_2) = \\
&= (d_1 - d_2)(2ax^* + a(d_1 + d_2) + b) = \\
&= d_1^2 - d_2^2 > 0
\end{aligned} \tag{18}$$

Hence, closer solution will always yield smaller performance measure. QED.

Some notes are due: First, the information about the statistics of the error was never used. Second, it can be trivially shown that optimization problems in (10) and in (17) yield different results. Subtle difference is that optimization $\mathbf{M} = \text{argmin}\|\mathbf{I} - \mathbf{M}\mathbf{R}_{xx}\|_F$, is different from optimization $\mathbf{M} = \text{argmin}\|\mathbf{M} - \mathbf{R}_{xx}^{-1}\|_F$. While first optimization rule minimizes the sum of squared eigenvalues (without regard to their respective sizes), therefore minimizing the expected descent rate, second one targets the largest eigenvalue (due to the optimization on a quadratic bowl) and reduces it as much as possible, therefore maximizing the actual descent rate and eigenvalue disparity. Obviously from the point of convergence rate second approach is superior.

Moreover, it is important to note, that solution of (17) is a close cousin of the Newton's algorithm. The descent strategies derived from Newton's algorithm have a general structure [2]:

$$\mathbf{f}_{n+1} = \mathbf{f}_n + \mu\mathbf{R}_{xx}^{-1}\varepsilon_n\mathbf{x}_n \tag{19}$$

Diagonal approximation of $\mathbf{R}_{xx}^{-1}$ and full Newton's algorithm are only two ends of the spectra of algorithms which span different approximations of $\mathbf{R}_{xx}^{-1}$. In-between there is number of approximations that can be applied to affine filtering.

5. EXTENSION TO OPTIMAL UPDATES IN AFFINE FILTERING ALGORITHMS

While straightforward use of graded updates for stationary signals yields performance advantage, with a bit of extra work, great deal of convergence speed improvements can be obtained.

The idea here is to use the local correlation in the signal (i.e. correlation of adjacent symbols) in order to accelerate the convergence of the algorithm. Reformulation of the algorithm to include this information is straightforward:

$$\mathbf{f}_{n+1} = \mathbf{f}_n + \mathbf{M}\varepsilon_n\mathbf{x_n} \tag{20}$$

and

$$\mathbf{M} = \begin{bmatrix} \mu_1 & \mu_{12} & 0 & \dots & 0 \\ \mu_{12} & \mu_2 & \mu_{23} & & 0 \\ & & & \ddots & \\ 0 & 0 & \dots & \mu_{N_1,N} & \mu_N \end{bmatrix} \tag{21}$$

As it was shown in the previous sections, the optimal solution for matrix $\mathbf{M}$ is the approximation of the inverse of the covariance matrix by a symmetric constrained matrix. Constraint on the approximation of $\mathbf{R}_{xx}^{-1}$ determines how many sub-diagonals are used for approximation. Construction of gain matrix $\mathbf{M}$ follows from previous section by retaining the main diagonal and desired number sub-diagonals.

6. ERROR DYNAMICS

As in ordinary LMS algorithms it is possible to investigate the dynamics of the error, and estimate the excess MSE (EMSE) due to the uncertainty in the gradient estimate.

Using the error evolution from Section 1, the error vector of the LMS update with graded coefficients obeys the following difference equation:

$$\mathbf{v}_{n+1} = (\mathbf{I} - \mathbf{M}\mathbf{R}_{xx})\mathbf{v}_n - \mathbf{M}\mathbf{n}_n \tag{22}$$

where $\mathbf{n}_n$ is the error contributed due to the instantaneous estimate of the steepest descent direction. If $\mathbf{Q}$ is a matrix that diagonalizes the product $\mathbf{MR}$, so that $\mathbf{Q^T MRQ} = \mathbf{\Lambda}$ then the error dynamics in rotated coordinates takes the form:

$$\mathbf{v'}_{n+1} = (\mathbf{I} - \mathbf{\Lambda}_{xx})\mathbf{v'}_n - \mathbf{n'}_n \tag{23}$$

where $\mathbf{v'} = \mathbf{vQ}$ and $\mathbf{n'} = \mathbf{MnQ}$ are the rotated versions of the coefficient vector and noise vector. Excess MSE contribution per coordinate is then:

$$\sigma_{excess}(j) = \frac{E\{\mathbf{n'}(j)^2\}}{2\lambda_j - \lambda_j^2} \tag{24}$$

and complete excess MSE expression takes form of:

$$\sigma_{excess} = \Sigma_{j=0}^{N-1}\frac{E\{\mathbf{n'}(j)^2\}}{2\lambda_j - \lambda_j^2} \tag{25}$$

Vector $\mathbf{n'}$ is zero mean, and with correlation matrix $\mathbf{R}_{nn} = \mathbf{MQ'\Lambda Q}$. Unfortunately, past this point, not much can be said in general about the behavior of the error without going into the specific cases of the correlation matrix.

7. EXAMPLE

Consider an example of correlation matrix:

$$\mathbf{R}_{xx} = \begin{bmatrix} 1.2446 & 0.7140 & 0.5580 & 0.4961 & 0.0720 \\ 0.7140 & 1.2446 & 0.7140 & 0.5580 & 0.4961 \\ 0.5580 & 0.7140 & 1.2446 & 0.7140 & 0.5580 \\ 0.4961 & 0.5580 & 0.7140 & 1.2446 & 0.7140 \\ 0.0720 & 0.4961 & 0.5580 & 0.7140 & 1.2446 \end{bmatrix} \tag{26}$$

corresponding to a baseband transmission channel. Eigenvalues of the correlation matrix are $\text{Spectrum}(\mathbf{R}_x x) =$

$\{0.6032, 0.5253, 0.3220, 1.2560, 3.5165, \}$ indicating that the maximum adaptation step size is $\mu_{max} = 0.2844$. The eigenvalue spread for this matrix is 10.92.

Inverse of the correlation matrix is computed to be:

$$\mathbf{R}_{xx}^{-1} = \begin{bmatrix} 1.48 & -0.73 & -0.23 & -0.45 & 0.57 \\ -0.73 & 1.62 & -0.44 & 0.07 & -0.45 \\ -0.23 & -0.44 & 1.52 & -0.44 & -0.23 \\ -0.45 & 0.07 & -0.44 & 1.62 & -0.73 \\ 0.57 & -0.45 & -0.23 & -0.73 & 1.48 \end{bmatrix} \tag{27}$$

yielding the graded update gains:

$$\mathbf{M} = \begin{bmatrix} 0.074 & 0 & 0 & 0 & 0 \\ 0 & 0.081 & 0 & 0 & 0 \\ 0 & 0 & 0.076 & 0 & 0 \\ 0 & 0 & 0 & 0.081 & 0 \\ 0 & 0 & 0 & 0 & 0.074 \end{bmatrix} \tag{28}$$

and update gain matrix for affine filtering:

$$\mathbf{M}_2 = \begin{bmatrix} 0.074 & -0.037 & 0 & 0 & 0 \\ -0.037 & 0.081 & -0.022 & 0 & 0 \\ 0 & -0.022 & 0.076 & -0.022 & 0 \\ 0 & 0 & -0.022 & 0.081 & -0.037 \\ 0 & 0 & 0 & -0.037 & 0.074 \end{bmatrix} \tag{29}$$

Optimization with white error constraint yields:

$$\mathbf{M}_{white} = \begin{bmatrix} 0.474 & 0 & 0 & 0 & 0 \\ 0 & 0.398 & 0 & 0 & 0 \\ 0 & 0 & 0.390 & 0 & 0 \\ 0 & 0 & 0 & 0.398 & 0 \\ 0 & 0 & 0 & 0 & 0.474 \end{bmatrix} \tag{30}$$

The error dynamics for four algorithms under consideration is shown in Figure 3. For a given example, dashed line in Figure 3 presents the error dynamics of the standard uniform gain steepest descent algorithm. Dash-dotted line with slightly higher descent rate presents the maximum entropy approach which minimizes the trace of $\mathbf{I} - \mathbf{M}\mathbf{R}_{xx}^{-1}$. Solid line with significantly higher descent rate shows the error dynamics of the graded update steepest descent algorithm with $\mathbf{M}$ set to the diagonal of the inverse of the correlation matrix. Dotted line with yet higher descent rate presents the error dynamics of the affine algorithm of order 3, using $\mathbf{M}_2$ grading matrix.

It is interesting to note that for a different correlation matrix:

$$\mathbf{R}_{xx} = \begin{bmatrix} 2.0245 & 0.0476 & 0.0450 & 0.0270 & 0.0046 \\ 0.0476 & 2.0245 & 0.0476 & 0.0450 & 0.0270 \\ 0.0450 & 0.0476 & 2.0245 & 0.0476 & 0.0450 \\ 0.0270 & 0.0450 & 0.0476 & 2.0245 & 0.0476 \\ 0.0046 & 0.0270 & 0.0450 & 0.0476 & 2.0245 \end{bmatrix} \tag{31}$$

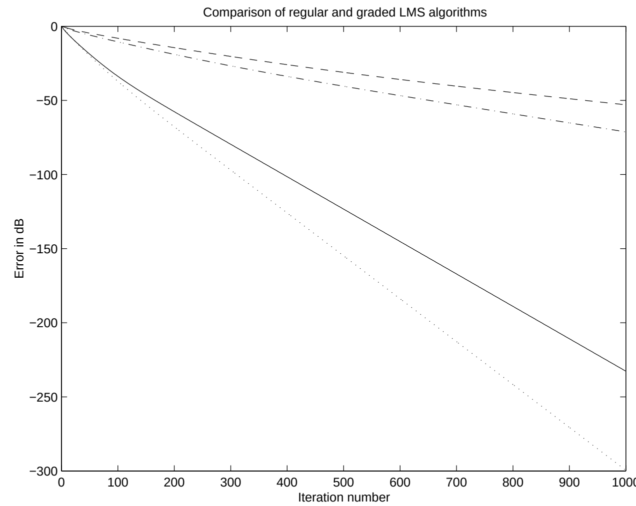

Figure 3: Illustration of the performance of four algorithms for a relatively large eigenvalue disparity case: LMS with uniform update rate (dashed), LMS with maximum entropy computed graded update (dash-dotted), LMS with inverse $\mathbf{R}_{xx}$ graded update (solid) and LMS with affine updates (dotted).

with a smaller eigenvalue spread of 1.1112, and gain matrices selected in same manner as in the previous example, the error dynamics looks quite different. Figure 4 shows error rate dynamics of four investigated algorithms. The convergence rates are more similar for variations of graded steepest descent algorithms, and they do not differ as much from ones obtained with the single update version. Reason for such behavior is the structure of the correlation matrix which is much closer to the diagonal matrix when eigenvalue spread is small. The inverse of the correlation matrix is then going to have a diagonal with nearly uniform elements yielding therefore nearly uniform optimal graded update gains.

8. CONCLUSIONS

In this paper the use of prior information to select the optimal update profile for vector steepest descent algorithms was considered. The optimal update profile (graded update gains) design procedure has been outlined for an arbitrary descent algorithm. Two flavors of graded updates have been discussed: first originating in maximum entropy principle and yielding the set of gains that reduce the sum of squared eigenvalues of $\mathbf{I} - \mathbf{M}\mathbf{R}_{xx}$, and second which reduces the largest eigenvalue of $\mathbf{I} - \mathbf{M}\mathbf{R}_{xx}$. The principle of graded updates

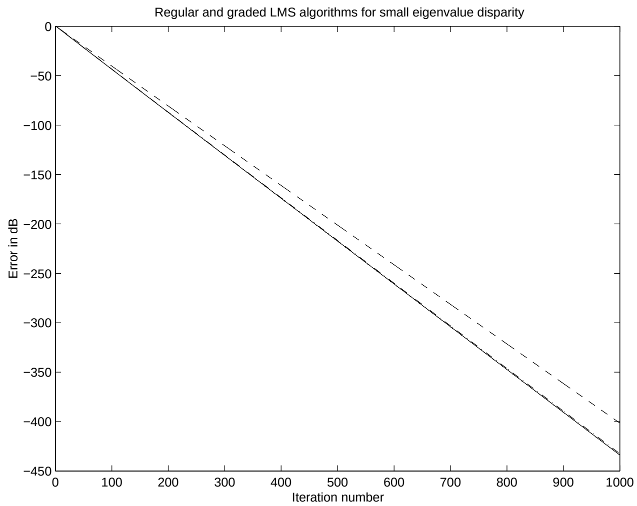

Figure 4: Illustration of the performance of four algorithms for small eigenvalue disparity: LMS with uniform update rate (dashed), LMS with maximum entropy computed graded update (dash-dotted), LMS with inverse $\mathbf{R}_{xx}$ graded update (solid) and LMS with affine updates (dotted).

was extended to affine filtering algorithms for which optimal update matrices have been designed. It was indicated that benefit from application of graded update gains increases with eigenvalue disparity of the correlation matrix $\mathbf{R}_{xx}$.

REFERENCES

[1] Bishop, C.M., "Neural Networks for Pattern Recognition", Clarendon Press - Oxford 1995.

[2] Clarkson, P. M., "Optimal and Adaptive Signal Processing", CRC Press 1993.

[3] Golub, G. and Van Loan, C., "Matrix Computations", Johns Hopkins 1996.

[4] Makino, S., Kaneda, Y., Koizumi, N., "Exponentially Weighted Stepsize NLMS Adaptive Filter Based on Statistics of a Room Impulse Response", IEEE Transactions on Speech and Audio Processing, Vol. 1, No. 1, January 1993.

[5] McCaslin, S. and Van Bavel, N., "Effects of Quasi-Periodic Training Signals on the Performance of Acoustic Echo Cancellers", Annales des Telecommunications, Vol. 49, No. 7-8, pp. 380-385, July 1994.

[6] Rupp, M. "Bursting the LMS Algorithm", IEEE Transactions on Signal Processing, Vol. 43, No. 10, October 1995.

[7] Widrow, B. and Stearns, S.D., "Adaptive Signal Processing", Engelwood Cliffs, N.J. Prentice Hall 1985.