OVERPARAMETRIZATION IN ADAPTIVE FILTERS

Albertus C. den Brinker

Faculteit der Elektrotechniek
Technische Universiteit Eindhoven, Eindhoven, Netherlands
A.C.d.Brinker@ele.tue.nl

ABSTRACT

Adaptive filters can be made fault tolerant by overparametrization. Conditions are derived such that no deterioration is caused by the redundancy under fault-free operation and that the deterioration caused by weight failures is minimized.

1. INTRODUCTION

Adaptive systems adjust their own parameters in order to reduce a certain specified error criterion. Hardware failures usually hamper the ability to minimize the error to the largest possible extent. By using a redundant number of degrees of freedom (the weights), the adaptive system can be made tolerant for hardware failures in the weights. The system then has the desirable property that it automatically compensates for this type of errors [1],[2],[3].

In order to perform an analysis of the behaviour of an adaptive system in the case of overparametrization, the adaptive filter problem and the overparametrization problem are separated. It is assumed that an adaptive filter has been designed with no redundant degrees of freedom such that a specified or intended level of performance in terms of steady-state accuracy and speed of convergence is reached. In all adaptive filters there is a fundamental trade-off between the steady-state accuracy and the speed of convergence. In principle, one cannot improve on this trade-off by overparametrization; the knowledge required to improve this trade-off by redundancy would immediately yield a better performance for the non-redundant case as well. The best we can do is to construct an overparametrized adaptive filter such that there is no deterioration in the performance caused by the redundancy itself. Since it is assumed that any weight can get stuck and that all weights have equal probability of failure it is also of interest to consider which condition on the overparametrization has to be imposed in order that the deterioration is independent of which weight gets stuck.

In our analysis we consider adaptive filters operating with an LMS-algorithm. The reason for this is that it is by far the most popular adaptive mechanism by virtue of its simplicity and its minor computational complexity. Furthermore, the LMS-algorithm can be used without alteration in the case of a redundant number of adaptive parameters even though there is no unique Wiener solution to the problem.

2. SCHEME AND DEFINITIONS

Consider the (overparametrized) adaptive filter shown in Fig. 1. The input signal is called $x(k)$, the output $y(k)$, the reference signal $d(k)$ and the error signal $e(k)$. The output of the adaptive filter is generated as a linear combination of internal signals $u_i(k)$ according to

$$y(k) = \underline{w}^h(k) A \underline{u}(k) \qquad (1)$$

where ^{h} denotes Hermitian transposition, $\underline{w}(k) = (w_0(k), \cdots, w_{N-1}(k))^t$ the weight vector (^{t} denotes transposition), $\underline{u}(k) = (u_0(k), \cdots, u_{M-1}(k))^t$ the vector containing the internal signals, and A is an $N \times M$ matrix ($N \geq M$). The internal signals are generated from the signals x by linear filters having impulse responses $f_m(k)$ and it is assumed that the impulse responses are linearly independent. If $A = I_M$ (the $M \times M$ identity matrix) we have the usual adaptive filtering case without overparametrization. The arbitrariness of the transfers $x \rightarrow u_n$ allows us to unify the analysis for all linear regression models including the tapped-delay-line, its extensions Laguerre and Kautz filters, frequency domain adaptive filters or the more general transform domain adaptive filters.
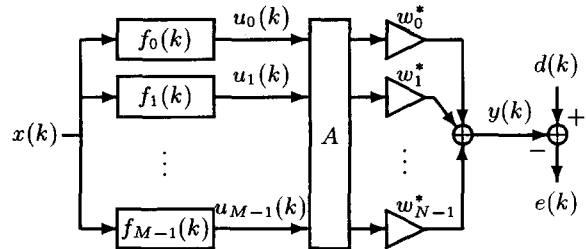

Figure 1: The overparametrized adaptive filter: subsystems and signals.

For the purpose of analysis we consider the adaptive filter working in a system identification setting and in a stationary environment. The reference signal $d(k)$ is assumed to consist of two additive parts: $d(k) = \tilde{\underline{w}}^h \underline{u}(k) + n(k)$, where n is a noise signal with zero mean and the signals x and n are statistically independent. This definition of $d(k)$ implies that we make the simplifying assumption that the unknown system is within the model set. The vector $\tilde{\underline{w}}$ is the reference weight vector which will be referred to as the Wiener solution.

The weights are driven by the LMS-algorithm:

$$\underline{w}(k+1) = \underline{w}(k) + \mu A\underline{u}(k)e^*(k), \qquad (2)$$

where * denotes conjugation.

3. ANALYSIS OF THE FAULT-FREE SITUATION

The analysis separates the adaptive filter problem from the overparametrization problem: we compare the overparametrized adaptive filter with the situation where $A = I_M$.

Consider the error signal. We have

$$e(k) = \{\tilde{w}^h - \underline{w}(k)^h A\}\underline{u}(k) + n(k). \qquad (3)$$

In view of this expression we define the weight-error vector $\underline{v}(k)$ according to

$$\underline{v}(k) = A^h\underline{w}(k) - \tilde{\underline{w}}. \qquad (4)$$

Note that the weight-error vector is only an M-dimensional vector. Variations in the overparametrized $N-M$ directions of vector $\underline{w}$ are not meaningful.

Combining (4) and (2) leads to the equation of motion for the weight-error vector

$$\begin{aligned}
\underline{v}(k+1) &= \underline{v}(k) + \mu A^h A\underline{u}(k)e^*(k) \\
&= (I_M - \mu A^h A\underline{u}(k)\underline{u}^h(k))\underline{v}(k) \\
&\quad + \mu A^h A\underline{u}(k)n^*(k). \qquad (5)
\end{aligned}$$

Clearly, if $A^h A = I_M$ the same equation of motion for the weight-error vector exists in the overparametrized case as for the case where $A = I_M$. Since the weight-error vector defines the error signal, the overparametrization in that case has no effect on the behaviour of the filter: the speed of adaptation and the accuracy of the final solution are in no way affected. The condition $A^h A = I_M$ means that the columns of A are orthonormal or, otherwise stated, that the rows of A constitute a frame in an M-dimensional Euclidian vector space with frame bounds equal to 1.

Taking $A^h A = I_M$ we obtain the same first-order approximations [4] for the weight-error correlation matrix as in the non-overparametrized case, i.e.,

$$RV + VR = \mu \sum_k r_{nn}(k)E[\underline{u}(n)\underline{u}^h(n-k)] \qquad (6)$$

and for the excess-mean-squared error

$$\begin{aligned}
\xi &= \mu E[|n(k)|^2]\left\{\sum_{m=0}^{M-1} E[|u_m(k)|^2]\right\} \\
&\quad - \frac{\mu}{2}\sum_k r_{nn}(k)E[\underline{u}^h(l)\underline{u}(l-k)]
\end{aligned}$$

where V is the weight-error correlation matrix, R is the covariance matrix of the internal signals u_m, and r_{nn} is the autocorrelation function of the signal $n(k)$.

Simple examples of matrices that adhere to $A^h A = cI_M$ ($c > 0$) are the following

- $A^h = [I_M U]$ with U an arbitrary unitary $M \times M$ matrix (e.g. the Fourier matrix [1], cosine matrix, Hadamard, Walsh, etc.);
- A is derived from an $N \times N$ unitary matrix by deleting some columns (e.g. the Fourier matrix [2]).

4. WEIGHT FAILURES

We now consider the case where one or more weights get stuck: by hardware failures they cannot change their value anymore. A weight stuck at zero can simply be accounted for, in terms of the previous analysis, by setting the appropriate row of A equal to zero or, equivalently, by considering a smaller overparametrization matrix where the appropriate row is deleted. A weight stuck at some nonzero value can be treated in the same way since the stuck (fixed) weight can be incorporated (with a minus sign) in the reference weight vector $\tilde{\underline{w}}$. The case of partially stuck weights is not considered here.

The previous analysis has shown that no deterioration of the adaptive system occurs if $A^h A = I_M$. Another desirable feature would be that the least possible deterioration in performance occurs if one or more weights get stuck. The extra conditions that have to be imposed on the matrix A to attain this are considered in this section.

We start from the notion that all weights have equal probability to fail. Furthermore, it is assumed that the covariance matrix $E\{\underline{u}(k)\underline{u}^h(k)\} = cI_M$ where c a positive real number. This is done not only for the sake of ease of analysis but also in view of a strict separation of the adaptive filtering problem from the overparametrization problem. We return to this issue later.

As can be seen from (5) the overparametrization matrix determines the behaviour of the weights. If the jth weight gets stuck, we can still use the analysis as before but have to replace the matrix A by a matrix B_j where B_j is identical to A but for the deletion of th jth row. Similarly, if more weights get stuck, we have to replace A by a matrix where the appropriate rows are deleted. In order to obtain a performance as close as possible to the failure-free situation, we require that the condition number of the matrix $B^h B$ is as close as possible to unity. The condition number of a matrix is defined as the ratio of its smallest and its largest eigenvalue, i.e., $\lambda_{min}/\lambda_{max}$.

As is to be expected, there is not a single matrix A which is optimal in the previous sense but rather there is a whole class. This can be seen as follows. Consider the case where $E\{\underline{u}(k)\underline{u}^h(k)\} = cI_M$. Using a unitary transformation $\hat{\underline{u}}(k) = U\underline{u}(k)$ leads to a new set of internal signals with identical properties in terms of the covariance matrix. This implies that the matrix A can only be determined save for a unitary transformation. Secondly, multiplying the signals just before the weights with an arbitrary complex number with unity absolute value will not influence the output and error signal (assuming that we initialize the weights by zero). This remark holds even in the case of stuck weights. Lastly, in view of the fact that an arbitrary weight can get stuck, row permutations in A are permitted. This leads to a class of optimal matrices

$$A = PDA_0 U \qquad (7)$$

where P is a permutation matrix, D a diagonal matrix with unity norm entries on the diagonal, U a unitary matrix, and A_0 is in some sense a fundamental solution to the problem.

4.1. Single weight failures

If a single weight gets stuck, then the adaptation proceeds with one degree of freedom less. We consider the eigenvalues of $B_j^h B_j$ and want to minimize the eigenvalue spread.

Theorem. The maximum over j of the minimal eigenvalue of $B_j^h B_j$ is given by $1 - M/N$. In that case $B_j^h B_j$ has $M - 1$ eigenvalues equal to 1 and one eigenvalue equal to $1 - M/N$, independent of j. Furthermore, all the rows of A have equal norm $\sqrt{M/N}$.

Proof. With the definition $A^h = (\underline{a}_0, \cdots, \underline{a}_{N-1})$ we have $B_j^h B_j = I_M - \underline{a}_j \underline{a}_j^h$ and thus it is clear that $B_j^h B_j$ has one (minimal) eigenvalue $1 - \underline{a}_j^h \underline{a}_j$ with eigenvector $\underline{a}_j$ and $M - 1$ eigenvalues equal to 1. Furthermore we have the property $M = \mathrm{tr}\{A^h A\} = \mathrm{tr}\{AA^h\} = \sum_j ||\underline{a}_j||_2^2$. Thus the maximum over j of the minimal eigenvalue occurs if all frame vectors $\underline{a}_j$ have equal norm $||\underline{a}_j||_2 = \sqrt{M/N}$.

Note that determining factor for the deterioration in the case of a single weight failure is not the number of redundant parameters but rather the relative degree of overparametrization $(N - M)/M$.

4.2. Several weight failures

If we have two stuck weights, say the j and kth weight, we introduce the matrix B_{jk} as identical to A except for the deletion of the j and kth row $(k \neq j)$. For the eigenvalues of $B_{jk}^h B_{jk}$ we find in a similar way as before that there are $M - 2$ eigenvalues equal to 1 and the remaining two eigenvalues are equal to $1 - M/N \pm |\underline{a}_k^h \underline{a}_j|$. In order to attain the largest possible condition number, we require that the rows of the matrix A are as close as possible to orthogonality. (Note that the result that the absolute value of the inner product of the rows determines the eigenvalues nicely agrees with the freedom in the additional matrix D.) The geometrical interpretation of minimizing the maximum of $|\underline{a}_j^h \underline{a}_k|$ is that the hyperplanes perpendicular to the vectors defined by the rows of A neatly divide the M-dimensional plane in N equal compartments.

Theorem. If the matrix A is chosen such that $A^H A = I_M$ and $||\underline{a}_j||_2 = \sqrt{M/N}$ and that $\max_{j,k} |\underline{a}_j^h \underline{a}_k|$ is minimized $(j \neq k)$, then for each j there are at least two vectors $\underline{a}_k$ which attain this value.

Proof. The solution to the minimization problem gives an equilibrium: in order that any vector $\underline{a}_j$ cannot be given an infinitesimally small change such that this minimum decreases, this change has to lead to an increase in the absolute value of the inner product with some other vector which immediately has to become larger than the obtained minimum.

The inner products $\underline{a}_k^h \underline{a}_j$ are all contained in the matrix AA^h. Since we are aiming at a 'neat' division of the the M-dimensional space, this requirement will be reflected in a certain structure in the matrix AA^h. In view of the expected symmetry of the solution and motivated by the freedom given by the permutation matrix P and the matrix D, it is conjectured that the fundamental 'neat' division

is reflected in the matrix $C = A_0 A_0^h$ in the form of C being a circulant matrix. In view of the previous analysis we have as additional constraints on C

- the matrix C is Hermitian;
- the matrix C has M eigenvalues equal to 1 and $N - M$ eigenvalues equal to 0;
- the diagonal of C contains entries equal to M/N;
- all off-diagonal entries are smaller than M/N in absolute value;
- the largest absolute value of the off-diagonal entries has to be minimized.

The requirement that $|\underline{a}_j^h \underline{a}_k|$ $(k \neq j)$ reaches its maximum value for at least two values of k for each j is automatically met in view of the circulant and Hermitian property of C.

5. CIRCULANT MATRICES

In this section we give the well-known relation between circulant matrices and Fourier matrices and apply this to our problem.

Definition. The Fourier matrix $F^{(N)}$ is defined as the $N \times N$ matrix having entries $\{F^{(N)}\}_{kl} = \exp\{-2\pi \jmath kl/N\}$ where $\jmath = \sqrt{-1}$ and $k, l = 0, 1, 2, \cdots, N - 1$.

Theorem. Given an $N \times N$ circulant matrix C then this matrix can be written as $C = F^{(N)} \Lambda F^{(N)h}/N$, where Λ is a diagonal matrix containing the eigenvalues of C.

Proof. Consider the N-points DFT of each of the rows of C. These transforms are stacked into a matrix and this is thus equal to $CF^{(N)}$. Since C contains circularly shifted versions of the first row, this can be written as $CF^{(N)} = F^{(N)} \Lambda$ where Λ is a diagonal matrix containing the N-points DFT of the first row of C. From this it follows that $C = F^{(N)} \Lambda F^{(N)h}/N$. Furthermore, we have the well-known result that the first row of a circulant matrix and the eigenvalues form a DFT-pair.

Corollary. Consider a circulant matrix $C = AA^h$ with M eigenvalues equal to 1 and $N - M$ eigenvalues equal to 0. Then A can be written as $A = A_0 U$ where A_0 is an $N \times M$ matrix consisting of M arbitrary rows of the scaled Fourier matrix, i.e., $F^{(N)}/\sqrt{N}$, and U is an arbitrary $M \times M$ unitary matrix.

Proof. From the previous theorem we have

$$AA^h = F^{(N)} \Lambda F^{(N)h}/N = \{F^{(N)} \Lambda/\sqrt{N}\}\{F^{(N)} \Lambda/\sqrt{N}\}^h$$

in view of the fact that $\Lambda^2 = \Lambda$. The matrix $F^{(N)} \Lambda/\sqrt{N}$ is an $N \times N$ matrix containing $(N - M)$ columns identical to zero. Eliminating these columns we have $AA^h = A_0 A_0^h$ which can be simply extended to $AA^h = A_0 U U^h A_0^h$.

It was shown that a selection of columns from a Fourier matrix for A_0 automatically meets the first three requirements from the previous section. However, the selection of the columns must still be made in order to meet the last two requirements.

Many choices of columns from the Fourier matrix yield the same matrix C in terms of absolute values of the entries. This can easily be seen from the following two examples.
1. The choice for a set of columns $i_1, i_2, \cdots, i_M$ gives the same entries in C in absolute value as $i_1 + k, i_2 + k, \cdots, i_M + k$ with k an arbitrary integer. (The addition of k has to be interpreted in the modulo N sense.) Such a shift in the

choice of the columns can be absorbed in the matrix D.

2. The choice for a set of columns $i_1, i_2, \cdots, i_M$ gives the same entries in C in absolute values as the choice $N - i_1, N - i_2, \cdots, N - i_M$ in view of the fact that these two sets of columns are each others conjugate. (For clarity, the columns in the Fourier matrix are indexed $0, 1, \cdots, N - 1$).

6. FULLY FAULT-TOLERANT DESIGNS

Another desirable feature of the overparametrization matrix A would be that is fully fault-tolerant by which we mean that $N - M$ weights can get stuck without loss of the its full adaptive capabilities. This implies that deleting an arbitrary $N - M$ rows of the matrix A must give a regular matrix. This issue is only partially covered by the following two Lemmas.

Lemma. Constructing the matrix A_0 in the above given way by selection of M consecutive columns from the Fourier matrix gives a full fault-tolerant solution.

Proof. Suppose we select the first M columns from an $N \times N$ Fourier matrix to construct A_0. The matrix is then a Vandermonde matrix with no identical rows and where the Vandermonde matrix is defined as

$$
\begin{pmatrix}
1 & \rho_0 & \rho_0^2 & \cdots & \rho_0^{M-1} \\
1 & \rho_1 & \rho_1^2 & \cdots & \rho_1^{M-1} \\
\vdots & & & & \\
1 & \rho_{N-1} & \rho_{N-1}^2 & \cdots & \rho_{N-1}^{M-1}
\end{pmatrix} .
$$

After deletion of $N - M$ columns, we still have a Vandermonde matrix which is regular. Since we may use an arbitrary matrix D (see (7)) we can extend this result to the selection of M consecutive columns from the Fourier matrix.

The previous statement can be extended. The Vandermonde character of the matrix is retained if we do not choose M consecutive columns, but also if we select M columns with an equal spacing κ from the Fourier matrix. (It is assumed that $M\kappa < N$.) However, in that case we do not always have a fully fault-tolerant solution as stated in the following.

Lemma. If N is a multiple of κ then the choice of columns from the Fourier matrix by an equal spacing of κ does not yield a fully fault-tolerant solution.

Proof. Define $N = \kappa L$ (L is an integer $L \geq 2$). We select our the columns by an equidistant spacing equal to κ, i.e., the columns $l\kappa + l_0$ with $l = 0, 1, \cdots, M - 1$ and l_0 an arbitrary fixed integer with $0 \leq l_0 < \kappa$. For convenience we take $l_0 = 0$. The entries in the matrix A_0 now become $\{A_0\}_{kl} = \exp\{-2\pi jkl/L\}$ ($k = 0, \cdots, N - 1, l = 0, \cdots, M - 1$). It is clear that the rows k and $k + L$ are identical. Thus if $N - M$ weights fail but two weights with spacing L are still functioning, the resulting $M \times M$ Vandermonde matrix is singular.

Note that this last statement is rather week: since two rows from the matrix A_0 are identical, it does not adhere to the last two requirements of Section 4.2. In fact, it might even be possible that any outcome of the minimization process which leads to A_0 automatically generates fully fault-tolerant solutions.

7. DISCUSSION

By separating the adaptive filtering problem from the overparametrization problem, we have found conditions relating the overparametrized case to the non-overparametrized one such that

i. there is no deterioration in performance in the fault-free situation;

ii. the deterioration caused by weight failures is minimized. In principle, this latter remark only applies in a first-order analysis of the LMS-behaviour if R is a scaled identity matrix since, essentially, one has to consider the eigenvalues of the matrices $B_j^h B_j R$. Having a priori knowledge of R, the filters f_i can be redesigned such that the covariance matrix becomes the scaled identity matrix. Also, adaptive orthogonalization procedures can be used to obtain orthogonal internal signals (e.g. lattice filters, Laguerre lattice filters [5]). Without such procedure and without knowledge of R the previous analysis holds as a worst-case scenario for any R since the worst possible condition number is the product of the condition numbers of the separate matrices.

Construction of the overparametrization matrix A was considered. This was based on the assumption that the fundamental solution gives rise to a circulant matrix for $A_0 A_0^h$. The fundamental solution is then a part of a Fourier matrix. However, which part has to be chosen to obtain the required optimal solution has not yet been answered. Nonetheless, our results suggest that the choice made in [2] is an interesting one. This is the more so since, if we aim at fault-tolerancy for the whole system and not just for the weights only, any additional processing to obtain fault-tolerancy must not be error-prone itself. Fortunately, there exist efficient and fault-tolerant implementations of the Fourier matrix [6],[7].

8. REFERENCES

[1] B.A. Schnaufer and W.K. Jenkins, "A fault tolerant FIR adaptive filter," in *Proc. 1992 ISCAS, San Diego, CA USA*, May 1992, pp. 2192–2195.

[2] B.A. Schnaufer and W.K. Jenkins, "An FFT-based fault tolerant FIR adaptive filter," in *Proc. 1994 ISCAS, London, UK*, 1994, pp. 261–264.

[3] W.K. Jenkins e.a., *Advanced Concepts in Adaptive Signal Processing*, Kluwer, Dordrecht, Netherlands, 1996.

[4] H.J. Butterweck, "A steady-state analysis of the LMS adaptive algorithm without use of the independence assumption," in *Proc. 1995 ICASSP, Detroit, MI USA*, 1995, pp. 1404–1407.

[5] T. Oliveira e Silva, "Optimality conditions for Laguerre lattice filters," *IEEE Signal Process. Letters*, vol. 2, pp. 97–98, 1995.

[6] Y.-H. Choi and M. Malek, "A fault-tolerant FFT processor," *IEEE Trans. Comput.*, vol. C-37, pp. 617–621, 1988.

[7] J.Y. Jou and J.A. Abraham, "Fault-tolerant FFT networks," *IEEE Trans. Comput.*, vol. C-37, pp. 548–561, 1988.