The Behavior of LMS and NLMS Algorithms in the Presence of Spherically Invariant Processes

Markus Rupp

Abstract—In this paper, the behavior of the least mean square (LMS) and normalized least mean square (NLMS) algorithms with spherically invariant random processes (SIRP's) as excitation is shown. SIRP's are of particular interest because many random processes fall under this category and SIRP's closely resemble speech signals. Consequently, it is of great interest to know the influence of SIRP's on the stability and convergence speed of stochastic gradient algorithms. The properties of SIRP's and the methods used in this paper are not restricted to LMS and NLMS algorithms but also allow a deeper insight into the behavior of some derivatives of the LMS algorithm.

I. INTRODUCTION

SINCE the derivation of the least mean square (LMS) algorithm by Widrow and Hoff in 1960, many authors have published the properties of this algorithm and its normalized version. However, all the authors have utilized simplifications like those used by Gardner [5] in 1984 which are appropriate only for several practical applications. In the last few years, more precise calculations of the behavior have been published. After Feuer and Weinstein [4] in 1985 published the calculation of the LMS algorithm behavior for Gaussian input sequences, Nitzberg [12], [13] showed in 1985 the behavior of the first moment of the weight-error vector in the normalized least mean square (NLMS) algorithm, and in 1986 Bershad [1] demonstrated the behavior of the second moments for the same excitation signal. Unfortunately, most analyses utilized either the properties of the Gaussian density function, making it such that the calculations could not be modified for other random processes, or the analyses used identically independent distributions in order to simplify the calculations.

It has not been clear until recently how the algorithms perform with a speech signal as input. Brehm [2], [3] and Stammler [14] have shown that the statistical properties of speech signals are well described by spherically invariant random processes (SIRP's). By incorporating SIRP's within the LMS/NLMS algorithm analyses, not only is the scope of the analyses broadened to include a larger class of random processes, but the results are germane to speech applications.

Since not every reader may have been exposed to SIRP's, the most important properties of these random

processes pertinent to this article are summarized in the following section. In the third section the LMS algorithm is introduced, and the first- and second-order moments of the weight-error vector $\epsilon(n) = \underline{w}_0 - \underline{w}(n)$ between the Wiener solution $\underline{w}_0$ and the estimated solution $\underline{w}(n)$ are shown. The behavior of the NLMS algorithm is derived in Sections IV and V in which the first- and second-order moments of the weight-error vector $\epsilon(n)$ are, respectively, calculated. After a verification of the results with the already known results for a white noise process and a colored Gaussian process as input sequences, some simulation results with a K_0-process are shown in Section VI. The article ends with a conclusion in Section VII.

II. SIRP's

A common problem in the field of signal processing is to describe the behavior of algorithms in the presence of speech signals. Gaussian properties do not adequately model speech signals but are often suitable for calculation purposes. The use of special one-dimensional probability density functions (pdf's) like K_0, Laplace, or gamma has led to a better description but has also suffered from being ambiguous with respect to the higher dimensional joint density functions. However, Brehm and Stammler [3] have shown that spherically invariant processes very suitably describe speech signals. This class of stationary random processes is characterized by having a multivariate density function that depends only on a radius r. However, a more characteristic property of these random processes is that all joint density functions are computable utilizing only the one-dimensional density function. Experiments have shown that speech signals do in fact exhibit these properties, especially the bivariate density function.

Because SIRP's are not well known, some important properties are listed here. A good tutorial is [3] where more references about SIRP's can be obtained and the generation of synthetical SIRP's with the desired density function and autocorrelation function (ACF) is also described. In speech signals the density functions are not spherically but elliptically invariant which can be represented by a correlated SIRP. The density function of a correlated SIRP $\underline{u}(n)$ of order M, can be presented as in the following equation:

$$f_{\underline{u}}(\underline{u}) = \frac{1}{\pi^{M/2} \det (R_{\underline{u}\underline{u}})^{1/2}} f(\underline{u}^T R_{\underline{u}\underline{u}}^{-1} \underline{u}, M). \quad (2.1)$$

Here $f(\cdot, M)$ is a function that describes an uncorrelated SIRP of order M, and R_{uu} is the autocorrelation matrix.

Thus, to model the speech signal of a certain speaker, it is only necessary to measure the ACF and the one-dimensional density function. The latter can generally be described by a G-function. These generalizations of hypergeometric functions, introduced in 1936 by Meijer [7], fit very well the measured density functions of speech. The G_{02}^{20} class of functions, which includes the Gaussian, K_0, Laplace, and gamma densities, was especially suitable. The Gaussian density function differs from the others in that it is the only ergodic spherically invariant process. The other density functions describe nonergodic processes (e.g., observed speech signals).

The first-order moments of the processes considered here are all assumed to be zero. Without any loss of generality, the correlated random process $u(n)$ can be considered as a linear transformation A of an uncorrelated process $x(n)$ of the same order M:

$$u(n) = Ax(n). \qquad (2.2)$$

Furthermore, it is assumed that the variance σ_x^2 of every element of the uncorrelated process $x(n)$ is the same and equals one (without any loss of generality). However, the elements of the uncorrelated process are not independent identically distributed (except for the Gaussian process) as is sometimes assumed in several derivations. The ACF matrix R_{uu} can then be obtained by $R_{uu} = AA^T$. Let Q be an orthogonal matrix[1] that diagonalizes $R_{uu} Q^T R_{uu} Q = \Lambda$. The elements of the diagonal matrix Λ are the eigenvalues of the ACF matrix. Q^T can also diagonalize A, i.e., $Q^T A = \Lambda^{1/2}$.

SIRP's can be represented as statistically independent with the usage of spherical coordinates. Hereby, the Cartesian coordinates $x^T = (x_1, \cdots, x_M)$ are substituted by spherical coordinates $\rho^T = (r, \phi^T)$, where $r^2 = x^T x$ and the angles have the following relations:

$$\phi^T = (\phi_1, \cdots, \phi_{M-1}) \qquad (2.3a)$$

$$x_1 = r \prod_{j=1}^{M-1} \sin \phi_j \qquad (2.3b)$$

$$x_k = r \cos (\phi_{M+1-k}) \prod_{j=1}^{M-k} \sin \phi_j;$$

$$k \in (2, M - 1) \qquad (2.3c)$$

$$x_M = r \cos \phi_1$$

$$\phi_1 \in [-\pi, \pi];$$

and for $k = 2 \cdots M - 1$: $\phi_k \in [0, \pi]$

$$(2.3d)$$

$$|\Delta_M| = r^{M-1} \prod_{j=1}^{M-2} (\sin \phi_j)^{M-1-j} = r^{M-1} \Delta(\phi)$$

$$(2.3e)$$

[1] In the case that the process is complex, Q is a unitary matrix.

$$d\underline{x} = r^{M-1} \Delta(\underline{\phi}) \, d\underline{\rho} = r^{M-1} \, dr \, \Delta(\underline{\phi}) \, d\underline{\phi}.$$

$$(2.3f)$$

Equation (2.3e) is the determinant of the Jacobian matrix. Because of the symmetry of the joint density functions of the uncorrelated process, it is easy to show that all moments that contain at least one uneven order equal zero, e.g., the first or third moments.

III. BEHAVIOR OF LMS ALGORITHM

Let $u(n)$ be a correlated input sequence and $w(n)$ the tap weights of a linear combiner. The LMS algorithm can then be stated as

$$e(n) = d(n) - w(n)^T u(n)$$

$$w(n + 1) = w(n) + \mu u(n) e(n). \qquad (3.1)$$

Hereby, $e(n)$ is the error of the desired output $d(n)$ and the estimated process $w(n)^T u(n)$, and μ is the step size. The following derivations apply for the general linear combiner case, a special case of which is the transversal filter (often used in speech applications). Building a weight-error vector $\varepsilon(n) = w_0 - w(n)$ and using the expectation operator and the independence theory [11],[2] the following vector equation results:

$$E[\varepsilon(n + 1)] = (I - \mu R_{uu}) E[\varepsilon(n)]. \qquad (3.2)$$

Expressing the ACF matrix R_{uu} in terms of uncorrelated $x(n)$ as in (2.2), and using the orthogonal matrix Q for diagonalization, (3.2) becomes finally $E[\varepsilon(n + 1)] = Q(I - \mu\Lambda) Q^T E[\varepsilon(n)]$. Stability is then guaranteed for every step-size μ that satisfies the inequality $0 < \mu < 2/\lambda_{max}$, where λ_{max} is the largest of the eigenvalues λ_i of the ACF matrix R_{uu}. This does not guarantee the stability of the complete algorithm, and therefore, convergence of the second-order moments of the error signal is desired. The desired $d(n)$ is describable by a linearly predictable component of order M and a second part $e_0(n)$ which is not linearly predictable for order M:

$$d(n) = w_0^T u(n) - e_0(n). \qquad (3.3)$$

The first-order moment of $e_0(n)$ is assumed to be zero, and the second moment shall be called the minimum mean-squared error of the Wiener solution:

$$E[e_0(n)] = 0 \qquad (3.4)$$

$$E[e_0^2(n)] = J_{min}. \qquad (3.5)$$

Solving for the second-order moment of the error signal, there is a relation between the expectation of the squared error and the covariance matrix of the weight-error vector:

$$E[e^2(n)] = \text{tr} \, (R_{uu} E[\varepsilon(n)\varepsilon^T(n)]) + J_{min}. \qquad (3.6)$$

[2] The independence theory [6] is commonly used in which the input vector sequence $u(n)$ is regarded as statistically independent. This can be true in the linear combiner case, whereas in the transversal filter case it is usually wrong. However, experiments often show good agreement with this assumption even in the case where the assumption fails.

The trace term is to be minimized in order to minimize $E[e^2(n)]$. Let $K(n) = E[\underline{\varepsilon}(n)\underline{\varepsilon}^T(n)]$ be the covariance matrix of the weight-error vector. $K(n)$ can then be expressed as a recursion:

$$K(n + 1) = E[(I - \mu\underline{u}(n)\underline{u}^T(n))K(n)(I - \mu\underline{u}(n)\underline{u}^T(n))]$$
$$+ \mu^2 E[\underline{u}(n)\underline{u}^T(n)e_0^2(n)]. \tag{3.7}$$

The terms $(I - \mu\underline{u}(n)\underline{u}^T(n))\underline{\varepsilon}(n)\underline{u}^T(n)$ $d(n)$ and $\underline{u}(n)$ $\underline{\varepsilon}^T(n)(I - \mu\underline{u}(n)\underline{u}^T(n))$ $d(n)$ are neglected since their expectation is zero. For the second term on the right-hand side in (3.7), it is assumed that $e_0(n)$ is uncorrelated with

$$\frac{m_x^{(4)}}{m_x^{(2,2)}} = \frac{\displaystyle\int_0^\pi \cos^4 \phi_1 \sin^{M-2} \phi_1 \, d\phi_1 \int_{-\pi}^\pi \sin^{M-3} \phi_2 \, d\phi_2}{\displaystyle\int_0^\pi \cos^2 \phi_1 \sin^M \phi_1 \, d\phi_1 \int_{-\pi}^\pi \cos^2 \phi_2 \sin^{M-3} \phi_2 \, d\phi_2}$$

the input process $\underline{u}(n)$, which is a correct assumption for sufficient order M. The second term of (3.7) is then simplified to $\mu^2 E[\underline{u}(n)\underline{u}^T(n)e_0^2(n)] = \mu^2 R_{\underline{uu}}J_{min}$. After also solving the first term, (3.7) can be expressed as

$$K(n + 1) = K(n) - \mu K(n)R_{\underline{uu}} - \mu R_{\underline{uu}}K(n)$$
$$+ \mu^2 E[\underline{u}(n)\underline{u}^T(n)K(n)\underline{u}(n)\underline{u}^T(n)]$$
$$+ \mu^2 R_{\underline{uu}}J_{min}. \tag{3.8}$$

Now consider the term with $K(n)$ that is proportional to μ^2. After replacing $\underline{u}(n)$ by the decorrelated process $\underline{x}(n)$ and using the abbreviation $Z(n) = A^T K(n)A$ with respect to spherically invariant processes, the following results:

$$E[\underline{x}(n)\underline{x}^T(n)Z(n)\underline{x}(n)\underline{x}^T(n)]_{ij} \tag{3.9a}$$

$$= \sum_k \sum_l E[\underline{x}_i(n)\underline{x}_k(n)Z_{kl}(n)\underline{x}_l(n)\underline{x}_j(n)] \tag{3.9b}$$

$$= \begin{cases} 2m_x^{(2,2)}Z_{ij}(n); & \text{for } i \neq j \\ m_x^{(4)}Z_{ii}(n) + \sum_{k \neq i} Z_{kk}(n)m_x^{(2,2)}; & \text{for } i = j \end{cases}$$
$$\tag{3.9c}$$

where[3]

$$m_x^{(2,2)} = E[\underline{x}_k^2(n)(\underline{x}_l^2(n)]; \quad \text{for } k \neq l \tag{3.10a}$$

$$m_x^{(4)} = E[\underline{x}_k^4(n)]. \tag{3.10b}$$

Defining γ and diag $(Z(n))$ as

$$\gamma = \frac{m_x^{(4)}}{m_x^{(2,2)}} \tag{3.11}$$

$$\text{diag } (Z(n)) = \begin{pmatrix} Z_{11}(n) & 0 & \cdots \\ & \ddots & \\ \cdots & 0 & Z_{MM}(n) \end{pmatrix} \tag{3.12}$$

[3]For the Gaussian case $m_x^{(2,2)}$ is one, but all other SIRP's have a second second moment greater than one.

$K(n)$ can finally be expressed as

$$K(n + 1) = K(n) - \mu R_{\underline{uu}}K(n) - \mu K(n)R_{\underline{uu}}$$
$$+ \mu^2 m_x^{(2,2)}(2R_{\underline{uu}}K(n)R_{\underline{uu}} + \text{tr } (Z(n))R_{\underline{uu}}$$
$$+ A \text{ diag } (Z(n))A^T(\gamma - 3))$$
$$+ \mu^2 R_{\underline{uu}}J_{min}. \tag{3.13}$$

For SIRP's γ always equals three, and the term $(\gamma - 3)$ can thus be omitted. To prove this, the moments are written in spherical integral form neglecting all parts that are equal:

$$= \frac{1 - 2\dfrac{M-1}{M} + \dfrac{M-1}{M}\dfrac{M+1}{M+2}}{\left(\dfrac{M-1}{M} - \dfrac{M-1}{M}\dfrac{M+1}{M+2}\right)\left(1 - \dfrac{M-2}{M-1}\right)} = 3.$$
$$\tag{3.14}$$

A very similar derivation shows that the quotient of the normalized fourth-order moment $m_x^{N(4)}$ and its joint fourth-order moment counterpart $m_x^{N(2,2)}$, both of which are used in the NLMS algorithm case, also equals three. Using $Q^T K(n)Q = C(n)$ and diagonalization with the matrix Q leads to

$$C(n + 1) = C(n) - \mu\Lambda C(n) - \mu C(n)\Lambda$$
$$+ \mu^2 m_x^{(2,2)}(2\Lambda C(n)\Lambda + \text{tr } (\Lambda C(n))\Lambda)$$
$$+ \mu^2\Lambda J_{min}. \tag{3.15}$$

Because all terms on the main diagonal of $C(n + 1)$ depend only on terms of the main diagonal of $C(n)$, these terms can be written in a vector $\underline{c}(n)$. Thus, (3.15) can be expressed using a vector $\underline{c}(n + 1)$, where

$$\underline{c}(n + 1) = B\underline{c}(n) + \mu^2\underline{\lambda}J_{min}. \tag{3.16}$$

The vector $\underline{\lambda}$ is filled with the eigenvalues from the diagonal matrix Λ. The entries of B are

$$B_{ij} = \begin{cases} \mu^2 m_x^{(2,2)}\lambda_i\lambda_j; & i \neq j \\ 1 - 2\mu\lambda_i + 3\mu^2 m_x^{(2,2)}\lambda_i^2; & i = j. \end{cases} \tag{3.17}$$

With the theorem of Gershgorin, it can be shown that the algorithm is stable when the step-size μ satisfies the following inequality:

$$\mu < \frac{2}{3 \text{ tr } (R_{\underline{uu}})}\frac{1}{m_x^{(2,2)}}. \tag{3.18}$$

For a Gaussian process the inequality is the same result as that from Feuer and Weinstein [4]. The limit in (3.18) should not mislead the reader; in the transversal filter case when the input sequence is strongly correlated, the step-size μ has to be much smaller because the independence assumption no longer holds.

To complete the calculations, the relative system mismatch $S_{rel}(n)$ and the average excess mean-squared error

is given:

$$S_{rel}(n) = \frac{E[\underline{\epsilon}(n)^T \underline{\epsilon}(n)]}{E[\underline{\epsilon}(0)^T \underline{\epsilon}(0)]} \qquad (3.19a)$$

$$= \frac{\| B^n \underline{c}(0) + \mu^2 J_{min}(I - B)^{-1}(I - B^n)\underline{\lambda}\|_{L_1}}{tr\ (C(0))}. \qquad (3.19b)$$

The final system mismatch can be calculated by using the matrix inversion lemma:

$$S_{rel}(\infty) = \frac{\mu J_{min}}{\underline{w}_o^T \underline{w}_o} \frac{\sum_{i=1}^{M} \frac{1}{1 - \mu m_x^{(2,2)} \lambda_i}}{2 - \mu m_x^{(2,2)} \sum_{i=1}^{M} \frac{\lambda_i}{1 - \mu m_x^{(2,2)} \lambda_i}}. \qquad (3.20)$$

In the same way the average excess mean-squared error is obtained:

$$E[e^2(\infty)] = tr\ (R_{\underline{u}\underline{u}} K(\infty)) + J_{min} \qquad (3.21a)$$

$$= \mu J_{min} \frac{\sum_{i=1}^{M} \frac{\lambda_i}{1 - \mu m_x^{(2,2)} \lambda_i}}{2 - \mu m_x^{(2,2)} \sum_{i=1}^{M} \frac{\lambda_i}{1 - \mu m_x^{(2,2)} \lambda_i}}$$

$$+ J_{min}. \qquad (3.21b)$$

For small step-sizes μ the term $1 - \mu m_x^{(2,2)} \lambda_i \approx 1$, and in the presence of a Gaussian process ($m_x^{(2,2)} = 1$) the average excess mean-squared error simplifies to

$$E[e^2(\infty)] \cong \mu J_{min} \frac{\sum_{i=1}^{M} \lambda_i}{2 - \mu \sum_{i=1}^{M} \lambda_i} + J_{min} \qquad (3.22)$$

which is also given in [8]. Equation (3.21b) leads to a more stringent bound for the step-size μ then (3.18). To limit the average excess mean-squared error the following two conditions for all $i = 1 \cdots M$ have to be satisfied:

$$0 < \mu < \frac{1}{m_x^{(2,2)} \lambda_i} \qquad (3.23a)$$

$$\mu m_x^{(2,2)} \sum_{i=1}^{M} \frac{\lambda_i}{1 - \mu m_x^{(2,2)} \lambda_i} < 2. \qquad (3.23b)$$

Since the largest eigenvalue $\lambda_{max}^{(B)}$ of the matrix B determines the convergence speed, it is of greatest interest. It is bounded by

$$1 - 2\mu\lambda_{max} + 2\mu^2 m_x^{(2,2)} \lambda_{max}^2 < \lambda_{max}^{(B)}. \qquad (3.24)$$

If only the linear terms in μ are considered, it is possible to give an estimate for $\lambda_{max}^{(B)}$ dependent on the eigenvalue ratio $\chi = \lambda_{max}/\lambda_{min}$.

$$1 - 2\frac{\mu M \sigma_u^2}{M - 1 + \chi} \leq \lambda_{max}^{(B)} \leq 1 - 2\frac{\mu M \sigma_u^2}{1 + (M - 1)\chi}. \qquad (3.25)$$

IV. Behavior of the First-Order Moments of NLMS Algorithm

The step-size μ in (3.1) is replaced with a step-size α and a normalizing term:

$$\mu = \frac{\alpha}{\underline{u}^T(n)\underline{u}(n)}. \qquad (4.1)$$

This leads to a new recursive equation for the expectation of the weight-error vector $\underline{\epsilon}(n)$:

$$E[\underline{\epsilon}(n + 1)] = E\left[I - \alpha \frac{\underline{u}(n)\underline{u}^T(n)}{\underline{u}^T(n)\underline{u}(n)} \right] E[\underline{\epsilon}(n)]$$

$$- \alpha E\left[\frac{\underline{u}(n)}{\underline{u}^T(n)\underline{u}(n)} (d(n) - \underline{w}_0^T \underline{u}(n)) \right]. \qquad (4.2)$$

Two terms are of particular interest: first, the normalized ACF term, and second, the error term at the end.

In the following investigations the usage of the matrices $A(\beta)$ and $L = L(\beta)$, which are dependent on a scalar β, proved to be useful. The two matrices are related as follows:

$$L(\beta) = A(\beta)A(\beta)^T \qquad (4.3a)$$

$$= (R_{\underline{u}\underline{u}}^{-1} + 2\beta I)^{-1} \qquad (4.3b)$$

where I is the identity matrix. As can be seen, for $\beta = 0$, (4.3b) reduces to the ACF matrix. It is desired to solve the expectation of the normalized matrix:

$$R_{\underline{u}\underline{u}}^N = E\left[\frac{\underline{u}(n)\underline{u}^T(n)}{\underline{u}^T(n)\underline{u}(n)} \right]. \qquad (4.4)$$

To facilitate this, a function $F_{ij}(\beta)$ is introduced:

$$F_{ij}(\beta) = \int_{\infty}^{\beta} \frac{\partial F_{ij}(x)}{\partial x}\ dx$$

$$= \frac{1}{\pi^{M/2}\ \sqrt{\det\ (R_{\underline{u}\underline{u}})}} \underbrace{\int \cdots \int}_{M\,times} \left(\frac{(\underline{u}\underline{u}^T)}{\underline{u}^T\underline{u}}\right)_{ij}$$

$$\cdot f(\underline{u}^T L^{-1} \underline{u},\ M)\ d\underline{u} \qquad (4.5)$$

where $f(\cdot, M)$ is a function that depends only on a scalar, the radius r of the SIRP. Thus, for $\beta = 0$ corresponding to $L(0) = R_{\underline{u}\underline{u}}$, $F_{ij}(0)$ is the i, jth element of the normalized ACF matrix.

The next step is to differentiate $F_{ij}(\beta)$ with respect to β and to describe the correlated process $\underline{u}(n)$ with an uncorrelated process $\underline{x}(n)$ and a matrix $A(\beta)$. In Appendix A it is shown that under the condition

$$\lim_{s \to \infty} s^{M/2} f(s, M) - s^{M/2} f(s, M)|_{s=0} = 0 \qquad (4.6)$$

$F_{ij}(\beta)$ can be written for spherically invariant processes as

$$\frac{dF_{ij}(\beta)}{d\beta} = -\sqrt{\frac{\det\ (L)}{\det\ (R_{\underline{u}\underline{u}})}}\ L_{ij}. \qquad (4.7)$$

Therefore, the matrix Q that diagonalizes the ACF matrix also diagonalizes the normalized ACF matrix

$$F(0) = Q \int_0^\infty \frac{(\Lambda^{-1} + 2\beta I)^{-1}}{\sqrt{\prod_{k=1}^M (1 + 2\beta\lambda_k)}} \, d\beta \, Q^T$$

$$= Q\Lambda^N Q^T \tag{4.8}$$

where Λ^N is the normalized eigenvalue matrix. This is the same result as Bershad's [1] for Gaussian processes. So it can be seen that every spherically invariant process leads to the same normalized ACF matrix that depends only on the eigenvalues and no longer on the density function.

Continuing with the concept of Appendix A and the methods of [1], it can be shown that the second term in (4.2) equals zero. Because all normalized eigenvalues are less than one, stability is guaranteed for $0 < \alpha < 2$. The integral in (4.8) is particularly significant and justifies close examination. Every eigenvalue λ_i is transformed into its normalized counterpart λ_i^N by

$$\lambda_i^N = \int_0^\infty \frac{\lambda_i \, d\beta}{(1 + 2\beta\lambda_i) \sqrt{\prod_{j=1}^M (1 + 2\beta\lambda_j)}}. \tag{4.9}$$

This elliptical integral is rather difficult to solve in the general case, but with the help of Schwarz' inequality an estimate is obtained (see Appendix B):

$$(\lambda_i^N)^2 \leq \lambda_i/4 \sum_{j=1}^M \frac{\lambda_j^{M-2}}{\prod_{k=1, k \neq j}^M (\lambda_j - \lambda_k)} \ln \lambda_j. \tag{4.10}$$

This inequality is very useful. When the eigenvalue spread is not too large (a typical value is less than 100), the solutions are very exact. But the most important concept is that the ordering sequence of the eigenvalues holds: i.e., if the eigenvalues are ordered and numbered according to increasing value ($\lambda_{min} = \lambda_1 < \lambda_2 < \cdots < \lambda_{M-1} < \lambda_M = \lambda_{max}$), after the transformation the normalized eigenvalues will also remain in increasing order ($\lambda_{min}^N = \lambda_1^N < \lambda_2^N < \cdots < \lambda_{M-1}^N < \lambda_M^N = \lambda_{max}^N$). This can be illustrated by the differences between two neighboring eigenvalues. The quotient of the difference Δ of the eigenvalues and Δ^N of the normalized eigenvalues is always positive. An estimation for the difference between the largest and the smallest normalized eigenvalue ($\lambda_{max}^N - \lambda_{min}^N$) is as follows:

$$\frac{1}{M+2}(\chi - 1) \geq \lambda_{max}^N - \lambda_{min}^N \geq \frac{1}{M+2} \frac{\chi - 1}{\chi} \tag{4.11}$$

where χ is the eigenvalue spread of the ACF matrix R_{uu}. Equation (4.11) suggests the superiority of the NLMS over the LMS algorithm. The normalized eigenvalue spread is approximately only the square root of the eigenvalue spread.

V. THE BEHAVIOR OF THE SECOND-ORDER MOMENTS OF THE NLMS ALGORITHM

The calculations for the second order moments of the weight-error vector $\underline{c}(n)$ proved to be very similar to the second-order case of the LMS algorithm. Using the same abbreviations as in Section III,

$$K(n+1) = E\left[\left(I - \alpha \frac{\underline{u}(n)\underline{u}^T(n)}{\underline{u}^T(n)\underline{u}(n)}\right) \right.$$

$$\left. \cdot K(n)\left(I - \alpha \frac{\underline{u}(n)\underline{u}^T(n)}{\underline{u}^T(n)\underline{u}(n)}\right)\right]$$

$$+ \alpha^2 E\left[\frac{\underline{u}(n)\underline{u}^T(n)}{\underline{u}^T(n)\underline{u}(n)\underline{u}^T(n)\underline{u}(n)}\right] J_{min}. \tag{5.1}$$

The last term describes the behavior of the minimal error term and can be solved, if the following conditions for the pdf are satisfied:

$$\lim_{s \to \infty} s^{M/2} \frac{\partial f(s, M)}{\partial s} - s^{M/2} \frac{\partial f(s, M)}{\partial s}\bigg|_{s=0} = 0 \tag{5.2a}$$

$$\lim_{s \to \infty} s^{(M/2)-1} f(s, M) - s^{(M/2)-1} f(s, M)\big|_{s=0} = 0. \tag{5.2b}$$

The calculations are similar to those in Appendix A. Using the zeroth order normalized moment

$$m_x^{N(0)} = E\left[\frac{1}{\underline{x}^T(n)\underline{x}(n)}\right] \tag{5.3}$$

the term reads as

$$E\left[\frac{\underline{u}(n)\underline{u}^T(n)}{\underline{u}^T(n)\underline{u}(n)\underline{u}^T(n)\underline{u}(n)}\right]$$

$$= (M - 2)m_x^{N(0)} Q$$

$$\cdot \int_0^\infty \int_{\beta_1 = \beta_2}^\infty \frac{(\Lambda^{-1} + 2\beta_1 I)^{-1} \, d\beta_1 \, d\beta_2}{\sqrt{\prod_{i=1}^M (1 + 2\beta_1\lambda_i)}} \, Q^T. \tag{5.4}$$

Solving for the first terms in (5.1),

$$E\left[\left(I - \alpha \frac{\underline{u}(n)\underline{u}^T(n)}{\underline{u}^T(n)\underline{u}(n)}\right) K(n) \left(I - \alpha \frac{\underline{u}(n)\underline{u}^T(n)}{\underline{u}^T(n)\underline{u}(n)}\right)\right]$$

$$= K(n) - \alpha R_{uu}^N K(n) - \alpha K(n) R_{uu}^N$$

$$+ \alpha^2 E\left[\frac{\underline{u}(n)\underline{u}^T(n)}{\underline{u}^T(n)\underline{u}(n)} K(n) \frac{\underline{u}(n)\underline{u}^T(n)}{\underline{u}^T(n)\underline{u}(n)}\right]. \tag{5.5}$$

Using $C(n) = Q^T K(n) Q$ again, the term in α^2 can be expressed in a form very similar to (5.4):

$$E\left[\frac{\underline{u}(n)\underline{u}^T(n)}{\underline{u}^T(n)\underline{u}(n)} K(n) \frac{\underline{u}(n)\underline{u}^T(n)}{\underline{u}^T(n)\underline{u}(n)}\right]$$

$$= Q \int_0^\infty \int_{\beta_1 = \beta_2}^\infty \frac{(\Lambda^{-1} + 2\beta_1 I)^{-1}}{\sqrt{\prod_{i=1}^M (1 + \beta_1\lambda_i)}}$$

$$\cdot (2C(n)(\Lambda^{-1} + 2\beta_1 I)^{-1}$$

$$+ \text{tr } (C(n)(\Lambda^{-1} + 2\beta_1 I)^{-1})I) \; d\beta_1 \; d\beta_2 Q^T.$$

$$(5.6)$$

To prove (5.6), the derivations from Section III and the methods from Appendix A must be used. It is then necessary to show that the normalized moments

$$m_x^{N(2,2)} = E\left[\frac{(\underline{x}(n)_i)^2(\underline{x}(n)_j)^2}{(\underline{x}^T(n)\underline{x}(n))^2}\right]; \quad \text{for } i \neq j$$

$$m_x^{N(4)} = E\left[\frac{(\underline{x}(n)_i)^4}{(\underline{x}^T(n)\underline{x}(n))^2}\right]$$

show the same dependency as their unnormalized counterparts, i.e., $m_x^{N(4)} = 3m_x^{N(2,2)}$, independent of the pdf. Furthermore, it can be shown that

$$m_x^{N(2,2)} = \frac{1}{M(M+2)} \quad (5.7)$$

which simplifies (5.1) to (5.6) under the conditions

$$\lim_{s \to \infty} s^{(M/2)+1} \frac{\partial f(s, M)}{\partial s} - s^{(M/2)+1} \frac{\partial f(s, M)}{\partial s}\Big|_{s=0} = 0$$

$$(5.8a)$$

$$\lim_{s \to \infty} s^{M/2} f(s, M) - s^{M/2} f(s, M)|_{s=0} = 0.$$

$$(5.8b)$$

All conditions (4.6), (5.2a), (5.2b), (5.8a) and (5.8b) are valid for the common density functions (Gaussian, Laplace, gamma, K_0). Composing a vector $\underline{c}(n)$ with diagonal terms of $C(n)$ leads to a vector equation similar to that of the LMS algorithm case:

$$\underline{c}(n+1) = B\underline{c}(n) + \alpha^2(M-2)m_x^{N(0)}\underline{\lambda}^H J_{\min}. \quad (5.9)$$

The vector $\underline{\lambda}^H$ is filled with the diagonal terms of the integral in (5.4), and B is a matrix with the following elements:

The elements of the matrix B determine its eigenvalues and also the stability bound of α. For the case that all eigenvalues are equal, it is easy to find a proof for stability for the bound $\alpha = 2$. A general proof is given in Appendix C where it is shown that for $0 < \alpha < 2$, stability can be guaranteed independent of the eigenvalues. As with the investigations in Section III, it is possible to estimate the largest eigenvalue $\lambda_{\max}^{(B)}$ of the matrix B. If the terms in α^2 are dropped and inequality (4.10) is used,

$$\lambda_{\max}^{(B)} \approx 1 - \frac{\alpha}{\sqrt{M\chi}}. \quad (5.11)$$

In contrast to the LMS case, here only the square root of the eigenvalue spread influences the convergence speed.

VI. EXAMPLES

In this section several examples will be presented, whereby results of simulations and calculations are compared. Two SIRP's were used as input sequences: first, a Gaussian SIRP for comparing already known results, and second, a K_0-SIRP as an approximation of speech signals. The variance of both processes was chosen to be one. The equivalent fourth-order moments were three ($m_x^{(2,2)} = 1$) and nine ($m_x^{(2,2)} = 3$) for the Gaussian and the K_0 case, respectively. The SIRP's were generated by a product process $\sigma(n)\eta(n)$ as described in [3]. The first factor $\sigma(n)$ of this product was a slowly varying Gaussian process that was optionally passed through a nonlinearity to obtain the desired probability density functions. The second factor $\eta(n)$ was derived from a white Gaussian process colored by a linear filter to obtain the desired ACF. Both SIRP's were provided in white and colored versions. The colored signals were produced by the following recursion:

$$\eta(n) = 0.994\eta(n-1) + 0.109w(n) \quad (6.1)$$

where $w(n)$ was a white Gaussian process. In Fig. 1 a single realization of the colored K_0-SIRP is presented. In the following examples, a system of order $M = 4$ was to be adapted. The eigenvalue ratio of the corresponding ACF matrix exceeded 1000, which is typical for speech signals. The output of the system was disturbed with a white noise ($J_{\min} = 10^{-10}$).

In the first example a white process with identical eigenvalues, i.e., $\lambda_i = \lambda$ (for $i = 1 \cdots M$), was used. The matrix A was therefore $\lambda^{1/2}I$, and the orthogonal matrix Q equaled the identity matrix I. First, the eigenvalues of the first-order moments of the LMS algorithm were in-

$$B_{ij} = \begin{cases} \alpha^2 \int_0^\infty \int_{\beta_2}^\infty \dfrac{\lambda_i\lambda_j \; d\beta_1 \; d\beta_2}{(1 + 2\beta_1\lambda_i)(1 + 2\beta_1\lambda_j)\sqrt{\prod_{k=1}^M (1 + 2\beta_1\lambda_k)}}; & \text{for } i \neq j \\[4ex] 1 - 2\alpha\lambda_i^N + 3\alpha^2 \int_0^\infty \int_{\beta_2}^\infty \dfrac{\lambda_i^2 \; d\beta_1 \; d\beta_2}{(1 + 2\beta_1\lambda_i)^2 \sqrt{\prod_{k=1}^M (1 + 2\beta_1\lambda_k)}}; & \text{for } i = j. \end{cases}$$
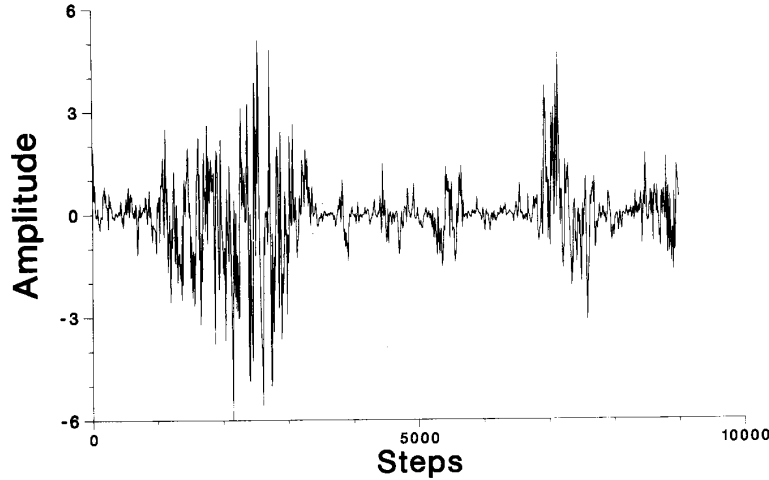
$$(5.10)$$

Fig. 1. Single realization of a SIRP with K_0 density function.

vestigated. These eigenvalues were equal to $(1 - \mu\lambda)$ which resulted in a first bound for the step-size μ: $0 < \mu < 2/\lambda$. The eigenvalues describing the convergence of the second-order moments were not all equal. The components of the matrix B were described as

$$B_{ij} = \begin{cases} \mu^2\lambda^2 m_x^{(2,2)}; & \text{for } i \neq j \\ 1 - 2\mu\lambda + 3\mu^2\lambda^2 m_x^{(2,2)}; & \text{for } i = j. \end{cases} \quad (6.2)$$

Defining E as the matrix with all components equal to one, the matrix B could be rewritten as

$$B = \mu^2\lambda^2 m_x^{(2,2)}E + (1 - 2\mu\lambda + 2\mu^2\lambda^2 m_x^{(2,2)})I. \quad (6.3)$$

Because $M - 1$ eigenvalues of E were zero and one was equal to the order M, each eigenvalue of B assumed one of two values:

$$1 - 2\mu\lambda + 2\mu^2 m_x^{(2,2)}\lambda^2$$
$$1 - 2\mu\lambda + (M + 2)\mu^2 m_x^{(2,2)}\lambda^2.$$

The second value was always the larger and therefore it determined the behavior of the algorithm. For stability, μ must meet the following condition[4]:

$$0 < \mu < \frac{2}{m_x^{(2,2)}(M + 2)\lambda}. \quad (6.4)$$

It was even possible to derive the optimal step size:

$$\mu_{\text{opt}} = \frac{1}{m_x^{(2,2)}(M + 2)\lambda}. \quad (6.5)$$

The curves shown in Fig. 2 were obtained by averaging 100 independent trials and depict the first 1000 steps of the relative system mismatch S_{rel}. The step sizes were

[4]The stability bound given here is an extension to that in [5]. For general density functions of white, ergodic processes, the author of [5] derived a bound depending only on the fourth-order moment. Thus, the results for the case of a Gaussian process coincide.

chosen to be optimal ($\mu = 1/6$ and $\mu = 1/18$ for the Gaussian and the K_0 cases, respectively). For the LMS algorithm with colored signals, an analytic calculation was more difficult, therefore, numerical results were used. The step sizes remained the same as in the previous simulation but were not optimal in this case. In Figure 3 the simulation results are depicted.

In the following, the performance of the NLMS algorithm is summarized. For the white processes the normalized eigenvalues λ^N were all equal. The normalized ACF matrix R_{uu}^N equaled $(1/M)I$, and therefore, the eigenvalues responsible for the first-order behavior were $1 - (\alpha/M)$. Stability was guaranteed for: $0 < \alpha < 2M$. This interval was surely much too large for stabilty of the second-order moments. Considering the second-order moments, the entries of B now simplified to

$$B_{ij} = \begin{cases} \dfrac{\alpha^2}{M(M + 2)}; & \text{for } i \neq j \\ 1 - 2\dfrac{\alpha}{M} + \dfrac{3\alpha^2}{M(M + 2)}; & \text{for } i = j. \end{cases} \quad (6.6)$$

As in the LMS algorithm case, two different eigenvalues resulted:

$$1 - \frac{\alpha}{M}\left(2 - \frac{2\alpha}{M + 2}\right)$$

$$1 - \frac{\alpha}{M}(2 - \alpha).$$

The second value was the larger one and again determined the algorithm behavior. It can be seen that in the undisturbed case the curve was independent of the density function (fulfilling some conditions as stated in the Appendix A), which remained true for colored sequences. However, in a disturbed environment the steady-state mismatch depends on the zeroth order normalized mo-
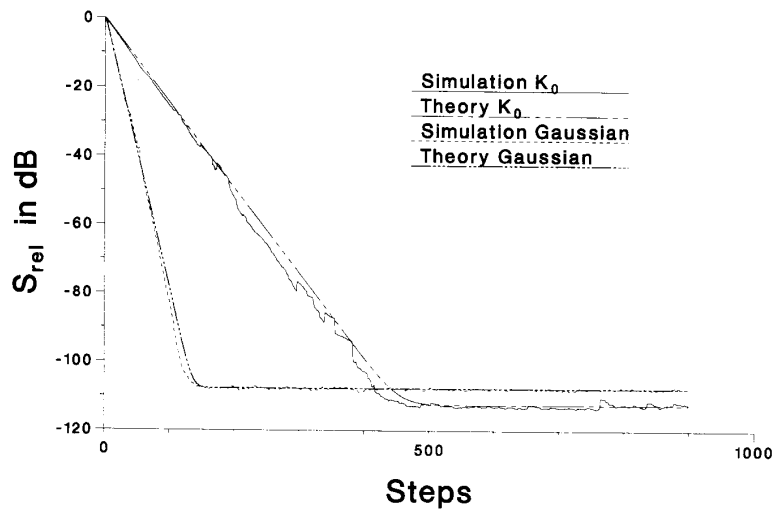
Fig. 2. Learning curves for LMS algorithm with a white Gaussian ($\mu = \mu_{opt} = 1/6$) and K_0 process ($\mu = \mu_{opt} = 1/18$).
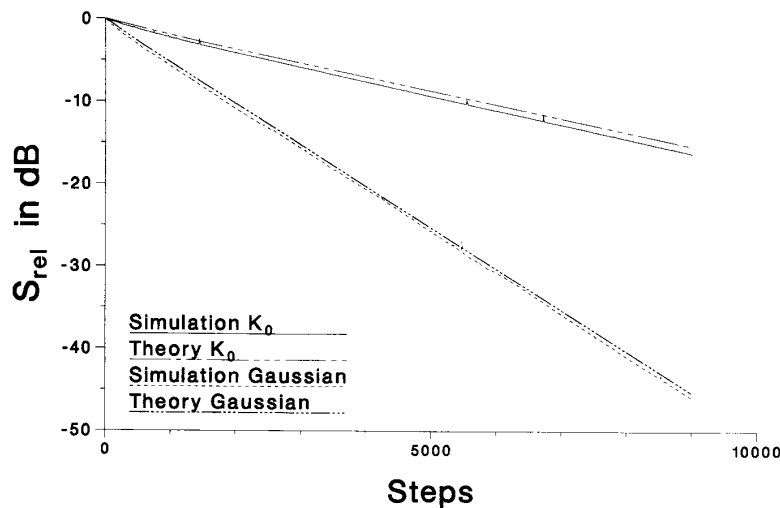

Fig. 3. Learning curves for LMS algorithm with a colored Gaussian ($\mu = 1/6$) and K_0 process ($\mu = 1/18$), $\chi > 1000$.

ment as can be seen in (5.4). The zeroth order normalized moment is in contrast to a Gaussian process not finite for a K_0 process. In simulations the value for this moment was obtained by averaging. The results are illustrated in Fig. 4. Stability was guaranteed for $0 < \alpha < 2$, which coincided exactly with the result known from observed data. In this case an optimal step-size α_{opt} could be calculated. From the largest eigenvalue $1 - 2(\alpha/M) + \alpha^2/M$, the optimal step size was $\alpha_{opt} = 1$.

In the fourth example the NLMS algorithm was excited by colored sequences. In comparison with the LMS algorithm (see Figs. 2 and 3) the difference between white and colored signals was not so large. The steady-state mismatch was also larger than in the white signal case.

The results of the simulations and calculations are shown in Fig. 5.

A comparison of the linear combiner case to the transversal filter has also been performed. Only the NLMS algorithm, which seemed to be superior, was used for adaptation. With the white process no great difference between transversal filter and linear combiner was found, whereas in the colored case a greater difference was noted. It can be inferred that in the colored case the independence assumption does not hold. Nevertheless, the differences to the linear combiner were not too large and stability held also for a step-size α less than two. Therefore, the concepts stated here can be applied as an approximation.
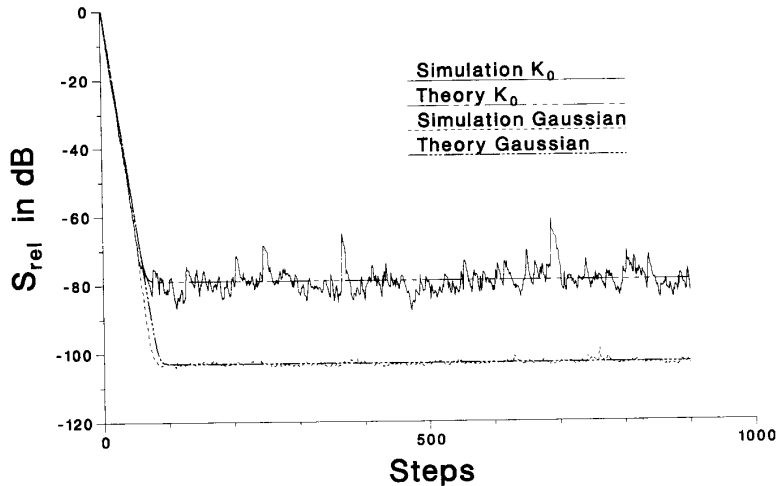
Fig. 4. Learning curves for NLMS algorithm with a white Gaussian and K_0 process ($\alpha = 1$).
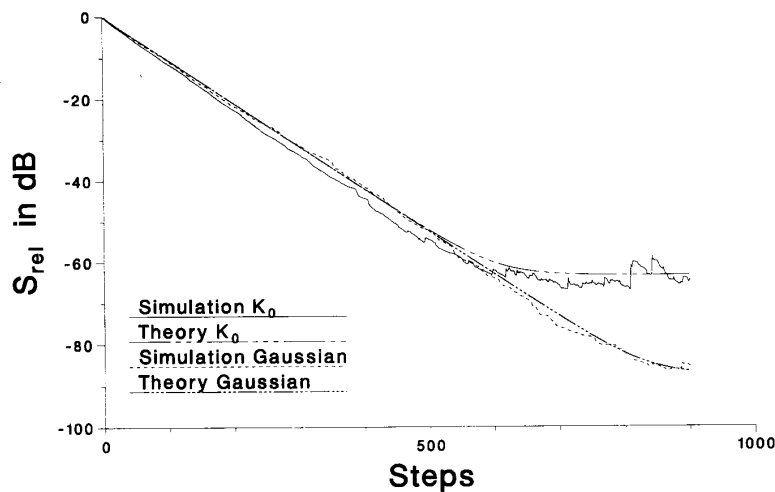

Fig. 5. Learning curves of NLMS algorithm with a colored Gaussian and K_0 process ($\alpha = 1$, $\chi > 1000$).

VII. CONCLUSIONS AND SUGGESTIONS

Spherically invariant processes (SIRP's) are characterized by their rich mathematical structures and their suitability for modeling speech signals. In the study of LMS and NLMS algorithms it is therefore of great interest to examine the behavior of these algorithms in the presence of SIRP's. Stability bounds have been derived and a matrix has been presented for which its eigenvalues determine the adaptation process. It turns out that the behavior of the LMS algorithm relies essentially on the fourth- and joint fourth-order moments, whereas the performance of the NLMS algorithm (with the exception of steady-state error) is independent of the probability density function. The method presented in this paper provides a means for algorithm comparison. For some special cases analytical expressions for optimal step sizes could be determined.

In contrast to describing behavior by empirical obser-

vation, spherically invariant processes enable one to calculate the behavior of some derived LMS algorithms (see also [10]), for instance, the signed LMS algorithm and an LMS algorithm with delayed coefficient update:

$$\underline{w}(n + 1) = \underline{w}(n) + \mu_s \, \text{sign} \, (\underline{u}(n)) e(n)$$

$$\underline{w}(n + 1) = \underline{w}(n) + \mu_D \underline{u}(n - D) e(n - D).$$

The latter appears useful in many applications (see also [9]).

APPENDIX A

Although the derivations here are similar to those in [1], they are briefly listed again to point out the extensions for spherically invariant processes, since Bershad used exclusively the properties of the Gaussian pdf. It will be shown under what conditions the function $F_{ij}(0)$ is di-

agonizable by the same unitary matrix Q as the ACF matrix. The calculation is started with (4.5) from Section IV,

$$F_{ij}(\beta) = \frac{1}{\pi^{M/2}\sqrt{\det{(R_{uu})}}} \underbrace{\int \cdots \int}_{M\,\text{times}} \frac{(\underline{u}\underline{u}^T)_{ij}}{\underline{u}^T\underline{u}}$$

$$\cdot f(\underline{u}^T L^{-1}\underline{u}, M)\, d\underline{u} \qquad (A.1)$$

where $L = (R_{uu}^{-1} + 2\beta I)^{-1} = A(\beta) A^T(\beta)$. In the following calculations, $A(\beta)$ is simply written as A, which should not be confused with $A(0)$. Differentiating $F_{ij}(\beta)$ with respect to β leads to

$$\frac{dF_{ij}(\beta)}{d\beta} = \frac{2}{\pi^{M/2}\sqrt{\det{(R_{uu})}}} \underbrace{\int \cdots \int}_{M\,\text{times}} (\underline{u}\underline{u}^T)_{ij}$$

$$\cdot \frac{\partial f(r^2, M)}{\partial r^2}\, d\underline{u}. \qquad (A.2)$$

Here, the fact is used that $\underline{u}^T L^{-1}\underline{u}$ can be written as the squared radius r^2. Substituting the correlated process $\underline{u}(n)$ by the matrix A and the uncorrelated process $\underline{x}(n)$ changes the equation above to

$$\frac{dF_{ij}(\beta)}{d\beta} = \frac{2\det{(A)}}{\pi^{M/2}\sqrt{\det{(R_{uu})}}} \left(A \underbrace{\int \cdots \int}_{M\,\text{times}} \underline{x}\underline{x}^T \right.$$

$$\left. \cdot \frac{\partial f(r^2, M)}{\partial r^2}\, d\underline{x} A^T \right)_{ij}. \qquad (A.3)$$

The introduction of spherically invariant processes leads to

$$\frac{dF_{ij}(\beta)}{d\beta} = \frac{2\det{(A)}}{\pi^{M/2}\sqrt{\det{(R_{uu})}}} \left(A \int_0^\infty r^{M+1} \frac{\partial f(r^2, M)}{\partial r^2}\, dr \right.$$

$$\left. \cdot \underbrace{\int \cdots \int}_{M-1\,\text{times}} \underline{\varphi}(\underline{\phi})\underline{\varphi}^T(\underline{\phi})\, \Delta(\underline{\phi})\, d\underline{\phi}\, A^T \right)_{ij}. \qquad (A.4)$$

where the vector $\underline{\varphi}(\underline{\phi}) = \underline{x}/r$ is filled with the sine and cosine terms from (2.3c) to (2.3d) without the factor r, and the scalar $\Delta(\underline{\phi})$ is the product term as defined in (2.3e) without r^{M-1}. Using the substitution $s = r^2$, the radial term is

$$\int_0^\infty r^{M+1} \frac{\partial f(r^2, M)}{\partial r^2}\, dr = \frac{1}{2} \int_0^\infty s^{M/2} \frac{\partial f(s, M)}{\partial s}\, ds$$

$$= \frac{1}{2} s^{M/2} f(s, M)\Big|_0^\infty - \frac{M}{4}$$

$$\cdot \int_0^\infty s^{(M/2)-1} f(s, M)\, ds. \qquad (A.6)$$

To eliminate the first term of (A.6), it is necessary that the density function $f(s, M)$ show a certain property at $s = 0$ and at $s \to \infty$. For the typical density functions, Gaussian, Laplace, gamma, and K_0 this condition is fulfilled:

$$\frac{dF_{ij}(\beta)}{d\beta} = \frac{-M\det{(A)}}{\pi^{M/2}\sqrt{\det{(R_{uu})}}} \left(A \underbrace{\int \cdots \int}_{M\,\text{times}} \frac{\underline{x}\underline{x}^T}{\underline{x}^T\underline{x}} \right.$$

$$\left. \cdot f(r^2, M)\, d\underline{x}\, A^T \right)_{ij}. \qquad (A.7)$$

Because the determinant of the matrix A is the square root of the determinant of the matrix L, (A.7) can be expressed as

$$\frac{dF_{ij}(\beta)}{d\beta} = \frac{-\sqrt{\det{(L)}}}{\sqrt{\det{(R_{uu})}}} \left(A \frac{M}{\pi^{M/2}} \right.$$

$$\left. \cdot \underbrace{\int \cdots \int}_{M\,\text{times}} \frac{\underline{x}\underline{x}^T}{\underline{x}^T\underline{x}} f(r^2, M)\, d\underline{x}\, A^T \right)_{ij} \qquad (A.8)$$

$$= \frac{-\sqrt{\det{(L)}}}{\sqrt{\det{(R_{uu})}}} (AA^T)_{ij}. \qquad (A.9)$$

Equation (A.9) holds because the term between the two matrices A and A^T is just the expectation of the normalized uncorrelated ACF matrix, which is $1/M$ times the identity matrix. Integrating with respect to β leads to the desired result for the matrix:

$$F(0) = Q \int_0^\infty \frac{(\Lambda^{-1} + 2\beta I)^{-1}}{\sqrt{\prod_{k=1}^M (1 + 2\beta\lambda_k)}}\, d\beta\, Q^T. \qquad (A.10)$$

APPENDIX B

The inequality in (4.10) is to be proven. Beginning with (4.9):

$$\lambda_i^N = \int_0^\infty \frac{\lambda_i\, d\beta}{(1 + 2\beta\lambda_i) \sqrt{\prod_{j=1}^M (1 + 2\beta\lambda_j)}}. \qquad (B.1)$$

Applying Schwarz' inequality to the square root term and the rest:

$$(\lambda_i^N)^2 \le \int_0^\infty \frac{\lambda_i^2\, d\beta}{(1 + 2\beta\lambda_i)^2} \int_0^\infty \frac{d\beta}{\prod_{j=1}^M (1 + 2\beta\lambda_j)}. \qquad (B.2)$$

The first integral is simply $\lambda_i/2$. The second integral is written as a sum:

$$\int_0^\infty \frac{d\beta}{\prod_{j=1}^M (1 + 2\beta\lambda_j)} = \int_0^\infty \sum_{j=1}^M \frac{A_j\, d\beta}{1 + 2\beta\lambda_j} \qquad (B.3)$$

with

$$A_j = \frac{\lambda_j^{M-1}}{\prod_{j \neq i} (\lambda_j - \lambda_i)}. \tag{B.4}$$

First, (B.3) is solved for a particular fixed i and a finite limit R,

$$\int_0^R \frac{A_j \, d\beta}{1 + 2\beta\lambda_j} = \frac{A_j}{2\lambda_j} \ln (1 + 2R\lambda_j). \tag{B.5}$$

Summing all terms for $j = 1 \cdots M$ and increasing R to infinity,

$$\int_0^\infty \sum_{j=1}^M \frac{A_j \, d\beta}{1 + 2\beta\lambda_j} = \lim_{R \to \infty} \frac{1}{2} \sum_{j=1}^{M-1} \frac{A_j}{\lambda_j} \ln \frac{1 + 2R\lambda_j}{1 + 2R\lambda_M} \tag{B.6}$$

$$= \frac{1}{2} \sum_{j=1}^M \frac{\lambda_j^{M-2}}{\prod_{j \neq i} (\lambda_j - \lambda_i)} \ln \lambda_j. \tag{B.7}$$

Finally, the desired result is:

$$(\lambda_i^N)^2 \leq \lambda_i / 4 \sum_{j=1}^M \frac{\lambda_j^{M-2}}{\prod_{k=1, k \neq j}^M (\lambda_j - \lambda_k)} \ln \lambda_j. \tag{B.8}$$

APPENDIX C

The proof for $\alpha \leq 2$ is still open. For this we introduce some abbreviations:

$$\lambda_i^N = \int_0^\infty \frac{\lambda_i \, d\beta}{(1 + 2\beta\lambda_i) \sqrt{\prod_{j=1}^M (1 + 2\beta\lambda_j)}} \tag{C.1}$$

$$\lambda_{ij}^N = \int_0^\infty \int_{\beta_2}^\infty \frac{\lambda_i \lambda_j \, d\beta_1 \, d\beta_2}{(1 + 2\beta_1\lambda_i)(1 + 2\beta_1\lambda_j) \sqrt{\prod_{k=1}^M (1 + 2\beta_1\lambda_k)}}. \tag{C.2}$$

First it will be shown that the following is true:

$$\lambda_i^N = \sum_{j=1}^M \lambda_{ij}^N + 2\lambda_{ii}^N. \tag{C.3}$$

Because the integrands in (C.1) and (C.2) are positive, the normalized eigenvalues are also positive values. The integrand of (C.3) equals:

$$\frac{\lambda_i}{1 + 2\beta_2\lambda_i} \frac{1}{\sqrt{\prod_{k=1}^M (1 + 2\beta_2\lambda_k)}}$$

$$= \int_{\beta_2}^\infty \frac{\lambda_i}{1 + 2\lambda_i\beta_1} \left(\sum_{j=1}^M \frac{\lambda_j}{1 + 2\lambda_j\beta_1} + \frac{2\lambda_i}{1 + 2\lambda_i\beta_1} \right)$$

$$\cdot \frac{d\beta_1}{\sqrt{\prod_{k=1}^M (1 + 2\beta_1\lambda_k)}}. \tag{C.4}$$

Differentiating the equation by β_2 leads to the desired result. After these calculations we can prove the stability bound for the step-size α. Therefore, the L_1-norm $v(n)$ of the homogeneous solution $\underline{c}_h(n)$ of (5.9) is desired: $v(n) = \sum_{i=1}^M (\underline{c}_h(n))_i$, where $(\underline{c}_h(n))_i$ denotes the ith element of the vector $\underline{c}_h(n)$. The following is obtained:

$$v(n + 1) = v(n) - 2\alpha \sum_{i=1}^M \lambda_i^N (\underline{c}_h(n))_i$$

$$+ \alpha^2 \left(\sum_{i=1}^M \left(\sum_{j=1}^M \lambda_{ij}^N + 2\lambda_{ii}^N \right) (\underline{c}_h(n))_i \right) \tag{C.5}$$

$$= v(n) - \alpha(2 - \alpha) \sum_{i=1}^M \lambda_i^N (\underline{c}_h(n))_i \tag{C.6}$$

$$< (1 - \alpha(2 - \alpha)\lambda_{\min}^N) v(n). \tag{C.7}$$

The parabola in α lies between zero and one for $0 \leq \alpha \leq 2$. The minimum of the parabola is $1 - \lambda_{\min}^N$ for $\alpha = 1$, which is an indication of the optimal step size.

REFERENCES

[1] N. J. Bershad, "Analysis of the normalized LMS algorithm with Gaussian inputs," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-34, no. 4, pp. 793-806, Aug. 1986.
[2] H. Brehm, "Sphärisch invariante stochastische Prozesse," *Ausgewählte Arbeiten über Nachrichtensysteme*, nr. 47, Habilitationsschrift, Erlangen, 1981.
[3] H. Brehm and W. Stammler, "Description and generation of spherically invariant speech-model signals," *Signal Processing*, vol. 9, pp. 119-141, 1987.
[4] A. Feuer and E. Weinstein, "Convergence analysis of LMS filters with uncorrelated Gaussian data," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-33, no. 1, pp. 222-230, Feb. 1985.
[5] W. A. Gardner, "Learning characteristics of stochastic- gradient-descent algorithms: A general study, analysis, and critique," *Signal Processing*, vol. 6, pp. 113-133, 1984.
[6] A. Gersho, "Adaptive equalization of highly dispersive channels for data transmission," *Bell Syst. Tech. J.*, pp. 55-70, Jan. 1969.
[7] I. S. Gradsteyn and I. M. Ryzhik, *Table of Integrals, Series and Products*. New York: Academic, 1980.
[8] Simon Haykin, *Adaptive Filter Theory*. Englewood Cliffs, NJ: Prentice-Hall, Information and System Sciences Series, 1986.
[9] G. Long, F. Ling, and J. G. Proakis, "The LMS algorithm with de-

layed coefficient adaptation," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 37, no. 9, pp. 1397–1405, Sept. 1989.

[10] O. Macchi, J. C. Pescquet, and G. Tziritas, "Second-order statistical analysis of two constrained LMS algorithms," in *Signal Processing V., Proc. Eusipco-90*, 1990, pp. 193–196.

[11] J. E. Mazo, "On the independence theory of equalizer convergence," *Bell Syst. Tech. J.*, vol. 58, no. 5, pp. 963–993, May–June 1979.

[12] R. Nitzberg, "Application of the normalized LMS algorithm to MSLC," *IEEE Trans. Aerosp. Electron. Syst.*, vol. AES-21, no. 1, pp. 79–91, Jan. 1985.

[13] R. Nitzberg, "Normalized LMS algorithm degradation due to estimation noise," *IEEE Trans. Aerosp. Electron. Syst.*, vol. AES-22, no. 6, pp. 740–750, Nov. 1986.

[14] W. Stammler, "Beiträge zur Analyse und Synthese sphärisch invarianter Sprachmodellprozesse," *Ausgewählte Arbeiten über Nachrichtensysteme*, nr. 60, Erlangen, 1985.

[15] M. Tarrab and A. Feuer, "Convergence and performance analysis of the normalized LMS algorithms with uncorrelated Gaussian data," *IEEE Trans. Inform. Theory*, vol. IT-34, no. 4, pp. 680–691, July 1988.

Markus Rupp was born in Völklingen, Germany, on May 14, 1963. He received the Diploma degree in electrical engineering from the Fachhochschule des Saarlandes in 1984 and from the Universität des Saarlandes, Saarbrücken, Germany, in 1988. He is currently working toward the Ph.D. degree in the area of the analysis methods of gradient-based algorithms.

Since 1988 he has been a Research Assistant in the Institut für Netzwerk- und Signaltheorie, TH Darmstadt.