but with $m = l = 12$, only 8% of the number of flops were required for STAN_1 and about 7% for EIV-PAST. From Fig. 1, we conclude that all considered algorithms offer similar tracking accuracy.

Example 2: As previously stated, one of the main advantages with STAN, compared with, for example EIV-PAST, is that estimates of the signal singular values are available. In Fig. 2, the estimated signal singular values and the singular values of $\hat{\boldsymbol{R}}_{x\xi}(t)$ are shown in a scenario where the number of signals is time varying. The STAN_3 algorithm is run with the hypothesis that $n = 2$. We note that the estimates obtained from STAN_3 allow us to detect changes in n.

Example 3: This example considers the stationary accuracy of the proposed algorithms. In this scenario, two planar wavefronts arrive from DOA's $[0^\circ 20^\circ]$. Further, the estimates of the first 300 samples are discarded so that the effects of the initial conditions are negligible. The results in Fig. 3 are based on 200 independent realizations, and $\mu = 0.97$. The first conclusion is that STAN_1 performs as well as EIV-PAST and the SVD approach. The second conclusion is that STAN_3 performs as wel as EIV-PAST and the SVD approach, only for "medium to high" values of the SNR.

V. Conclusions

In this correspondenec, we have proposed an $O((m + l)n^2)$ class of perturbation-based low-rank tracking algorithms, which are referred to as STAN. The low complexity of STAN is achieved by applying a novel approximation of a residual matrix. For the AC case with exponential forgetting, none of the STAN algorithms are new. However, the analysis that led to the different variants is useful. It was, for example, shown that Karasalo's algorithm and the algorithm F2 in [1] are closely related. Only the approximation of the residual matrix differs. Furthermore, the introduced perturbation-based framework allowed us to propose a novel sliding window algorithm for the AC case, and attractive algorithms for the CC case were found.

References

[1] P. Comon and G. H. Golub, "Tracking a few extreme singular values and vectors in signal processing," *Proc. IEEE*, vol. 78, pp. 1327–1343, Aug. 1990.

[2] G. H. Golub and C. F. VanLoan, *Matrix Computations*, 3rd ed. Baltimore, MD: Johns Hopkins Univ. Press, 1996.

[3] T. Gustafsson, "Instrumental variable subspace tracking using projection approximation," *IEEE Trans. Signal Processing*, vol. 46, pp. 669–681, Mar. 1998.

[4] ——, "Subspace methods for system identification and signal processing," Ph.D. dissertation, Chalmers Univ. Technol., Goteburg, Sweden, 1999.

[5] I. Karasalo, "Estimating the covariance matrix by signal subspace averaging," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-34, pp. 8–12, Feb. 1986.

[6] H. Krim and M. Viberg, "Two decades of array signal processing research: The parametric approach," *IEEE Signal Processing Mag.*, vol. 13, pp. 67–94, July 1996.

[7] M. Viberg, P. Stoica, and B. Ottersten, "Array processing in correlated noise fields based on instrumental variables and subspace fitting," *IEEE Trans. Signal Processing*, vol. 43, pp. 1187–1199, May 1995.

[8] B. Yang, "Projection approximation subspace tracking," *IEEE Trans. Signal Processing*, vol. 43, pp. 95–107, Jan. 1995.

Convergence Analysis of the Binormalized Data-Reusing LMS Algorithm

José Apolinário, Jr., Marcello L. R. Campos, and Paulo S. R. Diniz

***Abstract*—Normalized least mean squares algorithms for FIR adaptive filtering with or without the reuse of past information are known to converge often faster than the conventional least mean squares (LMS) algorithm. This correspondence analyzes an LMS-like algorithm: the binormalized data-reusing least mean squares (BNDR-LMS) algorithm. This algorithm, which corresponds to the affine projection algorithm for the case of two projections, compares favorably with other normalized LMS-like algorithms when the input signal is correlated. Convergence analyses in the mean and in the mean-squared are presented, and a closed-form formula for the mean squared error is provided for white input signals as well as its extension to the case of a colored input signal. A simple model for the input-signal vector that imparts simplicity and tractability to the analysis of second-order statistics is fully described. The methodology is readily applicable to other adaptation algorithms of difficult analysis. Simulation results validate the analysis and ensuing assumptions.**

***Index Terms*—Adaptive filters, adaptive signal processing, least mean squared methods.**

I. Introduction

The least mean squares (LMS) algorithm is very popular and has been widely used due to its simplicity. Its convergence speed, however, is highly dependent on the eigenvalue spread (conditioning number) of the input-signal autocorrelation matrix [1], [2]. Alternative schemes that try to improve performance at the cost of additional computational complexity have been proposed and extensively discussed in the past [1]–[5].

The data-reusing LMS (DR-LMS) algorithm [3] uses current desired and input signals repeatedly within each iteration in order to improve its convergence speed. With the algorithms proposed in [6], which are called normalized and unnormalized new data-reusing LMS (NNDR-LMS and UNDR-LMS), performance can be further improved when past data are also used within each iteration. The binormalized data-reusing LMS (BNDR-LMS) algorithm [7], [8] analyzed in this correspondence employs normalization on two orthogonal directions obtained from current and past data. It can be shown that the BNDR-LMS algorithm belongs to the family of normalized algorithms discussed in [9] when sample-by-sample updating is used. Although not mentioned, the affine-projections (AP) algorithm [10]–[12] can also be classified in the same set of algorithms studied in [9]. Actually, the BNDR-LMS algorithm is an alternative implementation of the special case of the AP algorithm when the number of projections is equal to two.

A thorough analysis of convergence in the mean and mean-squared for the BNDR-LMS algorithm coefficient vector is provided, together with a graphical description of the coefficient-vector updating for several LMS-like algorithms. Stability limits for the convergence factor, as well as closed-form formulas for mean squared error (MSE) after convergence for white and colored input signals, are obtained from

Manuscript received January 9, 1998; revised February 22, 2000. The associate editor coordinating the review of this paper and approving it for publication was Dr. Segios Theodoridis.

J. A. Apolinário, Jr. is with the Departamento de Engenharia Elétrica, Instituto Militar de Engenharia, Rio de Janeiro, Brazil.

M. L. R. de Campos and P. S. R. Diniz are with the Programa de Engenharia Elétrica, COPPE/EE-Universidade Federal do Rio de Janeiro, Rio de Janeiro, Brazil.

Publisher Item Identifier S 1053-587X(00)09312-0.

the analysis. The inadequacy of the independence assumption [13] for analyses of data-reusing algorithms [6], [9] is overcome by adopting a simplified model for the input-signal vector that is consistent with the first two moments and renders a tractable analysis [4], [14]. This analysis can be readily extended to other data-reusing algorithms.

II. BNDR-LMS and the Affine-Projections Algorithms

The BNDR-LMS algorithm combines data reusing, orthogonal projections of two consecutive gradient directions, and normalization in order to achieve faster convergence when compared with other LMS-like algorithms. At iteration $(k+1)$, the coefficient vector is calculated such that it belongs to hyperplanes $\mathcal{S}(k)$ (which contains all vectors $\boldsymbol{w}$ such that $\boldsymbol{x}^T(k)\boldsymbol{w} = d(k)$) and $\mathcal{S}(k-1)$, and it is at a minimum distance from $\boldsymbol{w}(k)$, i.e., $\boldsymbol{w}(k+1)$ is the solution of

$$\min_{\boldsymbol{w}(k+1)} \|\boldsymbol{w}(k+1) - \boldsymbol{w}(k)\|^2 \tag{1}$$

subjected to

$$\boldsymbol{x}^T(k)\boldsymbol{w}(k+1) = d(k) \tag{2}$$

and

$$\boldsymbol{x}^T(k-1)\boldsymbol{w}(k+1) = d(k-1) \tag{3}$$

where $d(k)$ is the desired signal, and $\boldsymbol{x}(k)$ is the input-signal vector containing the $N+1$ most recent input-signal samples, i.e.,

$$\boldsymbol{x}(k) = [x(k) \quad x(k-1) \quad \cdots \quad x(k-N)]. \tag{4}$$

The function to be minimized is therefore

$$\begin{aligned} f[\boldsymbol{w}(k+1)] = {} & \|\boldsymbol{w}(k+1) - \boldsymbol{w}(k)\|^2 \\ & + \lambda_1[d(k) - \boldsymbol{x}^T(k)\boldsymbol{w}(k+1)] \\ & + \lambda_2[d(k-1) - \boldsymbol{x}^T(k-1)\boldsymbol{w}(k+1)] \end{aligned} \tag{5}$$

which, for linearly independent input-signal vectors $\boldsymbol{x}(k)$ and $\boldsymbol{x}(k-1)$, has the unique solution

$$\boldsymbol{w}(k+1) = \boldsymbol{w}(k) + \frac{\lambda_1}{2}\boldsymbol{x}(k) + \frac{\lambda_2}{2}\boldsymbol{x}(k-1) \tag{6}$$

where λ_1 and λ_2 are the Lagrange multipliers given by

$$\begin{aligned} \frac{\lambda_1}{2} = {} & \frac{[d(k) - \boldsymbol{x}^T(k)\boldsymbol{w}(k)]\|\boldsymbol{x}(k-1)\|^2}{\|\boldsymbol{x}(k)\|^2\|\boldsymbol{x}(k-1)\|^2 - [\boldsymbol{x}^T(k)\boldsymbol{x}(k-1)]^2} \\ & - \frac{[d(k-1) - \boldsymbol{x}^T(k-1)\boldsymbol{w}(k)]\boldsymbol{x}^T(k-1)\boldsymbol{x}(k)}{\|\boldsymbol{x}(k)\|^2\|\boldsymbol{x}(k-1)\|^2 - [\boldsymbol{x}^T(k)\boldsymbol{x}(k-1)]^2} \end{aligned} \tag{7}$$

and

$$\begin{aligned} \frac{\lambda_2}{2} = {} & \frac{[d(k-1) - \boldsymbol{x}^T(k-1)\boldsymbol{w}(k)]\|\boldsymbol{x}(k)\|^2}{\|\boldsymbol{x}(k)\|^2\|\boldsymbol{x}(k-1)\|^2 - [\boldsymbol{x}^T(k)\boldsymbol{x}(k-1)]^2} \\ & - \frac{[d(k) - \boldsymbol{x}^T(k)\boldsymbol{w}(k)]\boldsymbol{x}^T(k-1)\boldsymbol{x}(k)}{\|\boldsymbol{x}(k)\|^2\|\boldsymbol{x}(k-1)\|^2 - [\boldsymbol{x}^T(k)\boldsymbol{x}(k-1)]^2}. \end{aligned} \tag{8}$$

The derivation presented above is valid for any $\boldsymbol{w}(k)$, which may or may not belong to $\mathcal{S}(k-1)$. However, if successive optimized steps are taken for $\boldsymbol{w}(k)$ for all k, then

$$\boldsymbol{x}^T(k-1)\boldsymbol{w}(k) = d(k-1) \tag{9}$$

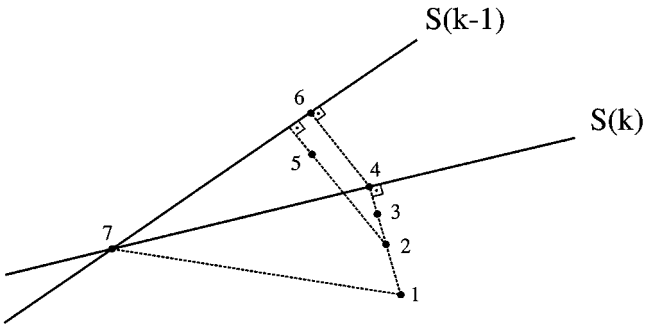

Fig. 1. Coefficient vector update for $L = 1$: Position 1 $\boldsymbol{w}(k)$. Position 2 $\boldsymbol{w}_{LMS}(k+1)$, first step toward $\boldsymbol{w}_{\text{DR-LMS}}(k+1)$ and $\boldsymbol{w}_{\text{UNDR-LMS}}(k+1)$. Position 3 $\boldsymbol{w}_{\text{DR-LMS}}(k+1)$. Position 4 $\boldsymbol{w}_{\text{NLMS}}(k+1)$ and first step toward $\boldsymbol{w}_{\text{NNDR-LMS}}(k+1)$. Position 5 $\boldsymbol{w}_{\text{UNDR-LMS}}(k+1)$. Position 6 $\boldsymbol{w}_{\text{NNDR-LMS}}(k+1)$. Position 7 $\boldsymbol{w}_{\text{BNDR-LMS}}(k+1)$ and $\boldsymbol{w}_{\text{AP}}(k+1)$.

and a simplified set of updating equations for the algorithm results:

$$\boldsymbol{w}(k+1) = \boldsymbol{w}(k) + \frac{\lambda_1'}{2}\boldsymbol{x}(k) + \frac{\lambda_2'}{2}\boldsymbol{x}(k-1) \tag{10}$$

where

$$\frac{\lambda_1'}{2} = \frac{[d(k) - \boldsymbol{x}^T(k)\boldsymbol{w}(k)]\|\boldsymbol{x}(k-1)\|^2}{\|\boldsymbol{x}(k)\|^2\|\boldsymbol{x}(k-1)\|^2 - [\boldsymbol{x}^T(k)\boldsymbol{x}(k-1)]^2} \tag{11}$$

and

$$\frac{\lambda_2'}{2} = \frac{-[d(k) - \boldsymbol{x}^T(k)\boldsymbol{w}(k)]\boldsymbol{x}^T(k-1)\boldsymbol{x}(k)}{\|\boldsymbol{x}(k)\|^2\|\boldsymbol{x}(k-1)\|^2 - [\boldsymbol{x}^T(k)\boldsymbol{x}(k-1)]^2}. \tag{12}$$

The AP algorithm relates to the NLMS and BNDR-LMS algorithms directly for it is a generalization for L data reuses of an algorithm that yields at every iteration k a solution that belongs to the intersection of hyperplanes defined by the present and all L previous data pairs. The coefficients are updated as follows:

$$\boldsymbol{w}(k+1) = \boldsymbol{w}(k) + \mu\boldsymbol{X}(k)\boldsymbol{t}(k) \tag{13}$$

where

$$\boldsymbol{t}(k) = [\boldsymbol{X}^T(k)\boldsymbol{X}(k) + \delta\boldsymbol{I}]^{-1}\boldsymbol{e}(k) \tag{14}$$

$$\boldsymbol{e}(k) = \boldsymbol{d}(k) - \boldsymbol{X}^T(k)\boldsymbol{w}(k) \tag{15}$$

and, for $L+1$ projections, the desired-signal vector and input-signal matrix are

$$\boldsymbol{d}(k) = [d(k) \quad d(k-1) \quad \cdots \quad d(k-L)]^T \tag{16}$$

and

$$\boldsymbol{X}(k) = [\boldsymbol{x}(k) \quad \boldsymbol{x}(k-1) \quad \cdots \quad \boldsymbol{x}(k-L)] \tag{17}$$

respectively, with $\boldsymbol{x}(k)$ denoting the input-signal vector as in (4).

Fig. 1 illustrates the updating of the coefficient vector for a two-dimensional problem for the algorithms mentioned above, starting with an arbitrary $\boldsymbol{w}(k)$. As we are interested in comparing algorithms of similar complexity, it was considered the case of one reuse, i.e., $L = 1$.

In order to control the excess of MSE of the BNDR-LMS algorithm, a step-size μ may be used. Although maximum convergence rate is usually obtained with $\mu = 1$, the use of a smaller value for the step-size may be required in applications where measurement error is too high. Alternatively, a variable step-size μ has been derived, yielding fast convergence and low mean squared error [15]. In these cases, we must emphasize that the solution $\boldsymbol{w}(k+1)$ obtained at each iteration is not at

TABLE I
BINORMALIZED DATA-REUSING LMS ALGORITHM

ϵ = small positive value

for each k

{ $\alpha = \boldsymbol{x}^T(k)\boldsymbol{x}(k-1)$

$\rho(k) = \boldsymbol{x}^T(k)\boldsymbol{x}(k)$

$y_1 = \boldsymbol{x}^T(k)\boldsymbol{w}(k)$

$e_1 = d(k) - y_1$

$den = \rho(k)\rho(k-1) - \alpha^2$

if $den < \epsilon$

{ $\boldsymbol{w}(k+1) = \boldsymbol{w}(k) + \mu e_1 \boldsymbol{x}(k)/\rho(k)$

}

else

{ $y_2 = \boldsymbol{x}^T(k-1)\boldsymbol{w}(k)$

$e_2 = d(k-1) - y_2$

$\frac{\lambda_1}{2} = (e_1\rho(k-1) - e_2\alpha)/den$

$\frac{\lambda_2}{2} = (e_2\rho(k) - e_1\alpha)/den$

$\boldsymbol{w}(k+1) = \boldsymbol{w}(k) + \mu[\frac{\lambda_1}{2}\boldsymbol{x}(k) + \frac{\lambda_2}{2}\boldsymbol{x}(k-1)]$

}

}

the intersection of hyperplanes $\mathcal{S}(k-1)$ and $\mathcal{S}(k)$, and therefore, the simplified version of this algorithm should not be used.

It is worth mentioning that if $\boldsymbol{x}(k)$ and $\boldsymbol{x}(k-1)$ are linearly dependent, then $\mathcal{S}(k)$ is parallel to $\mathcal{S}(k-1)$, and the BNDR-LMS algorithm should correspond to the NLMS algorithm. In order to avoid the case of almost parallel hyperplanes, a small test is carried out. The BNDR-LMS algorithm is summarized in Table I.

III. CONVERGENCE ANALYSIS OF THE COEFFICIENT VECTOR

In this section, we assume that an unknown FIR filter with coefficient vector given by $\boldsymbol{w}_o$ is to be identified by an adaptive filter of same order employing the BNDR-LMS algorithm, i.e., $d(k)$ can be modeled as

$$d(k) = \boldsymbol{x}^T(k)\boldsymbol{w}_o + n(k) \tag{18}$$

where $n(k)$ is measurement noise. It is also assumed that input signal and measurement noise are taken from independent and identically distributed zero-mean white noise processes with variances σ_x^2 and σ_n^2, respectively.

We are interested in analyzing the convergence behavior of the coefficient vector in terms of a step-size μ. Let

$$\Delta\boldsymbol{w}(k) = \boldsymbol{w}(k) - \boldsymbol{w}_o \tag{19}$$

be the error in the adaptive filter coefficients as related to the ideal coefficient vector. For the BNDR-LMS algorithm as described in Table I and using (18), $\Delta\boldsymbol{w}(k+1)$ is given by

$$\Delta\boldsymbol{w}(k+1) = [\boldsymbol{I} + \mu\boldsymbol{A}]\Delta\boldsymbol{w}(k) + \mu\boldsymbol{b} \tag{20}$$

where

$$\begin{aligned}\boldsymbol{A} = &\frac{\boldsymbol{x}(k)\boldsymbol{x}^T(k)\boldsymbol{x}(k-1)\boldsymbol{x}^T(k-1)}{\|\boldsymbol{x}(k)\|^2\|\boldsymbol{x}(k-1)\|^2 - [\boldsymbol{x}^T(k)\boldsymbol{x}(k-1)]^2} \\ &+ \frac{\boldsymbol{x}(k-1)\boldsymbol{x}^T(k-1)\boldsymbol{x}(k)\boldsymbol{x}^T(k)}{\|\boldsymbol{x}(k)\|^2\|\boldsymbol{x}(k-1)\|^2 - [\boldsymbol{x}^T(k)\boldsymbol{x}(k-1)]^2} \\ &- \frac{\|\boldsymbol{x}(k-1)\|^2\boldsymbol{x}(k)\boldsymbol{x}^T(k)}{\|\boldsymbol{x}(k)\|^2\|\boldsymbol{x}(k-1)\|^2 - [\boldsymbol{x}^T(k)\boldsymbol{x}(k-1)]^2} \\ &- \frac{\|\boldsymbol{x}(k)\|^2\boldsymbol{x}(k-1)\boldsymbol{x}^T(k-1)}{\|\boldsymbol{x}(k)\|^2\|\boldsymbol{x}(k-1)\|^2 - [\boldsymbol{x}^T(k)\boldsymbol{x}(k-1)]^2}\end{aligned} \tag{21}$$

and

$$\begin{aligned}\boldsymbol{b} = &\frac{n(k)\|\boldsymbol{x}(k-1)\|^2 - n(k-1)\boldsymbol{x}^T(k)\boldsymbol{x}(k-1)}{\|\boldsymbol{x}(k)\|^2\|\boldsymbol{x}(k-1)\|^2 - [\boldsymbol{x}^T(k)\boldsymbol{x}(k-1)]^2}\,\boldsymbol{x}(k) \\ &+ \frac{n(k-1)\|\boldsymbol{x}(k)\|^2 - n(k)\boldsymbol{x}^T(k-1)\boldsymbol{x}(k)}{\|\boldsymbol{x}(k)\|^2\|\boldsymbol{x}(k-1)\|^2 - [\boldsymbol{x}^T(k)\boldsymbol{x}(k-1)]^2}\,\boldsymbol{x}(k-1).\end{aligned} \tag{22}$$

By taking the expected value on both sides of (20), for $n(k)$ and $x(k)$ samples from independent zero-mean random processes, we have $E[\boldsymbol{b}] = 0$ and

$$\begin{aligned}&E[\Delta\boldsymbol{w}(k+1)] \\ &\quad = E[(\boldsymbol{I} + \mu\boldsymbol{A})\Delta\boldsymbol{w}(k)] \\ &\quad = E\left(\left\{\boldsymbol{I} + \mu\left[\frac{\boldsymbol{x}(k)\boldsymbol{x}^T(k)\boldsymbol{x}(k-1)\boldsymbol{x}^T(k-1)}{\|\boldsymbol{x}(k)\|^2\|\boldsymbol{x}(k-1)\|^2 - [\boldsymbol{x}^T(k)\boldsymbol{x}(k-1)]^2}\right.\right.\right. \\ &\qquad + \frac{\boldsymbol{x}(k-1)\boldsymbol{x}^T(k-1)\boldsymbol{x}(k)\boldsymbol{x}^T(k) - \|\boldsymbol{x}(k-1)\|^2\boldsymbol{x}(k)\boldsymbol{x}^T(k)}{\|\boldsymbol{x}(k)\|^2\|\boldsymbol{x}(k-1)\|^2 - [\boldsymbol{x}^T(k)\boldsymbol{x}(k-1)]^2} \\ &\qquad \left.\left.\left. - \frac{\|\boldsymbol{x}(k)\|^2\boldsymbol{x}(k-1)\boldsymbol{x}^T(k-1)}{\|\boldsymbol{x}(k)\|^2\|\boldsymbol{x}(k-1)\|^2 - [\boldsymbol{x}^T(k)\boldsymbol{x}(k-1)]^2}\right]\right\}\Delta\boldsymbol{w}(k)\right).\end{aligned} \tag{23}$$

Expression (23) can be further simplified if the following assumptions are made:

1) $\Delta\boldsymbol{w}(k)$ is statistically independent of $\boldsymbol{x}(k)\boldsymbol{x}^T(k)$ (independence assumption [13]).
2) $E[num/den] \approx E[num]/E[den]$, where num and den are the elements in the numerator and denominator of (23), respectively, which implies independence between num and den as well as a first-order approximation[1] in the evaluation of $E[1/den]$.

Moreover, the following relations can be easily verified when the elements of $\boldsymbol{x}(k)$ are samples of a white Gaussian process (see Appendix A):

$$E\{[\boldsymbol{x}^T(k)\boldsymbol{x}(k-1)]^2\} = (N+1)(\sigma_x^2)^2 \tag{24}$$

$$\begin{aligned}&E\{\|\boldsymbol{x}(k)\|^2\|\boldsymbol{x}(k-1)\|^2 - [\boldsymbol{x}^T(k)\boldsymbol{x}(k-1)]^2\} \\ &\qquad = N(N+3)(\sigma_x^2)^2\end{aligned} \tag{25}$$

$$\begin{aligned}&\{E[\boldsymbol{x}(k-1)\boldsymbol{x}^T(k-1)\boldsymbol{x}(k)\boldsymbol{x}^T(k)]\}_{ij} \\ &\qquad = \begin{cases}(\sigma_x^2)^2, & i = j \text{ or } i = j-2 \\ 0, & \text{otherwise}\end{cases}\end{aligned} \tag{26}$$

[1]For a more in-depth discussion on this approximation, see [4] and [16].

for $[\cdot]_{ij}$, which is the (i,j) element of matrix $[\cdot]$.

$$E[\|\boldsymbol{x}(k-1)\|^2\boldsymbol{x}(k)\boldsymbol{x}^T(k)] = (N+3)(\sigma_x^2)^2\boldsymbol{I} \tag{27}$$

$$\boldsymbol{x}^T(k-1)\Delta\boldsymbol{w}(k) = (1-\mu)\boldsymbol{x}^T(k-1)\Delta\boldsymbol{w}(k-1) + \mu n(k-1). \tag{28}$$

Based on these assumptions and relations, (23) can be rewritten as

$$\begin{aligned} E[\Delta\boldsymbol{w}(k+1)] &\approx E\left(\left\{\boldsymbol{I}+\mu\left[\frac{\boldsymbol{x}(k-1)\boldsymbol{x}^T(k-1)\boldsymbol{x}(k)\boldsymbol{x}^T(k)}{N(N+3)(\sigma_x^2)^2}\right.\right.\right. \\ &\quad \left.\left.\left. - \frac{\|\boldsymbol{x}(k-1)\|^2\boldsymbol{x}(k)\boldsymbol{x}^T(k)}{N(N+3)(\sigma_x^2)^2}\right]\right\}\Delta\boldsymbol{w}(k) \\ &\quad + \mu(1-\mu)\left[\frac{\boldsymbol{x}(k)\boldsymbol{x}^T(k)\boldsymbol{x}(k-1)\boldsymbol{x}^T(k-1)}{N(N+3)(\sigma_x^2)^2}\right. \\ &\quad \left. - \frac{\|\boldsymbol{x}(k)\|^2\boldsymbol{x}(k-1)\boldsymbol{x}^T(k-1)}{N(N+3)(\sigma_x^2)^2}\right]\Delta\boldsymbol{w}(k-1)\Big) \\ &\approx \left(1-\frac{\mu}{N}\right)E[\Delta\boldsymbol{w}(k)] - \frac{\mu(1-\mu)}{N}E[\Delta\boldsymbol{w}(k-1)]. \end{aligned} \tag{29}$$

The last relation of (29) was obtained by considering $\|\boldsymbol{x}(k-1)\|^2$ statistically independent of $\Delta\boldsymbol{w}(k)$ and by making a first-order approximation in the calculation of the numerators with the help of (26) to (28). From (29), it is clear that convergence in the mean of the BNDR-LMS algorithm to an unbiased solution is guaranteed for values of step size μ such that all elements of $E[\Delta\boldsymbol{w}(k+1)]$ in (29) go to zero as $k \to \infty$. This is achieved if the poles of the second-order difference equation are strictly inside the unit circle, i.e.,

$$|z_{1,2}| = \left|\frac{1-\frac{\mu}{N} \pm \sqrt{\left(1-\frac{\mu}{N}\right)^2 - \frac{4\mu(1-\mu)}{N}}}{2}\right| < 1 \tag{30}$$

which is always true for $N \geq 1$ and μ satisfying

$$0 < \mu < 2. \tag{31}$$

Note that for the case $N = 0$, the BNDR-LMS algorithm degenerates to the NLMS algorithm, and (31) is still valid [4].

IV. Second-Order Statistics Analysis

A. White Input Signal

Although $\Delta\boldsymbol{w}(k)$ converges in average to zero as k goes to infinity, which characterizes unbiasedness of the estimate, consistency of coefficient estimates can seldom be achieved for nonvanishing values of μ. In general, an excess of MSE, which depends on the second-order statistics of vector $\Delta\boldsymbol{w}(k)$, will be present. The excess of MSE is defined as [1], [2]

$$\xi_{\text{exc}} = \lim_{k\to\infty}\xi(k) - \xi_{\min} \tag{32}$$

where $\xi(k) = E[e^2(k)]$, and $\xi_{\min}$ is the minimum mean-squared error due to nonexact-modeling or presence of additive noise, or both [1].

The difference $\Delta\xi(k) = \xi(k) - \xi_{\min}$ is known as excess in the MSE [1] and can be expressed as

$$\begin{aligned} \Delta\xi(k) &= E\{[n(k) - \Delta\boldsymbol{w}^T(k)\boldsymbol{x}(k)]^2\} - \xi_{\min} \\ &= E[\Delta\boldsymbol{w}^T(k)\boldsymbol{R}\Delta\boldsymbol{w}(k)] \\ &= \text{tr}\{\boldsymbol{R}\,\text{cov}[\Delta\boldsymbol{w}(k)]\}. \end{aligned} \tag{33}$$

It is necessary, therefore, to derive an expression for the coefficient-error-vector covariance matrix $\text{cov}[\Delta\boldsymbol{w}(k+1)]$. From (20)

$$\begin{aligned} \text{cov}[\Delta\boldsymbol{w}(k+1)] &= E[\Delta\boldsymbol{w}(k+1)\Delta\boldsymbol{w}^T(k+1)] \\ &= E\{[\boldsymbol{I}+\mu\boldsymbol{A}]\Delta\boldsymbol{w}(k)\Delta\boldsymbol{w}^T(k)[\boldsymbol{I}+\mu\boldsymbol{A}]\} \\ &\quad + E\{\mu[\boldsymbol{I}+\mu\boldsymbol{A}]\Delta\boldsymbol{w}(k)\boldsymbol{b}^T\} \\ &\quad + E\{\mu\boldsymbol{b}\Delta\boldsymbol{w}^T(k)[\boldsymbol{I}+\mu\boldsymbol{A}]\} + E[\mu^2\boldsymbol{b}\boldsymbol{b}^T]. \end{aligned} \tag{34}$$

Recalling (21) and (22), we can foresee the enormous complexity to evaluate (34), even with a number of assumptions [6], [9], which thus far had prevented the complete analysis of this family of algorithms. An interesting alternative is the use of a simplified model for the input-signal vector $\boldsymbol{x}(k)$, which can be consistent with the first- and second-order statistics of a general input signal but has a reduced and countable number of possible directions of excitation. This model was successfully employed in [4] and [14]. The input-signal vector for the model is

$$\boldsymbol{x}(k) = s_k r_k \boldsymbol{V}_k \tag{35}$$

where we have the following.

- s_k is ± 1 with probability of occurrence 1/2.
- r_k^2 has the same probability distribution function of $\|\boldsymbol{x}(k)\|^2$ or, for the case of interest, is a sample of an independent process with χ-square distribution of $(N+1)$ degrees of freedom $E[r_k^2] = (N+1)\sigma_x^2$.
- $\boldsymbol{V}_k$ is equal to one of the $N+1$ orthonormal eigenvectors of $\boldsymbol{R}$, which are denoted $\mathcal{V}_i, i = 1, \cdots, N+1$. We will also assume that for a white Gaussian input signal, $\boldsymbol{V}_k$ is uniformly distributed, and consequently, if $P(\cdot)$ denotes the probability of occurrence of event $(\cdot)$

$$P(\boldsymbol{V}_k = \mathcal{V}_i) = \frac{1}{N+1}. \tag{36}$$

For the given input-signal model, we may express $\Delta\xi(k+1)$ as

$$\begin{aligned} \Delta\xi(k+1) &= \Delta\xi(k+1)|_{\boldsymbol{x}(k)\|\boldsymbol{x}(k-1)} \times P[\boldsymbol{x}(k)\|\boldsymbol{x}(k-1)] \\ &\quad + \Delta\xi(k+1)|_{\boldsymbol{x}(k)\perp\boldsymbol{x}(k-1)} \times P[\boldsymbol{x}(k) \perp \boldsymbol{x}(k-1)] \end{aligned} \tag{37}$$

Conditions $\boldsymbol{x}(k)\|\boldsymbol{x}(k-1)$ and $\boldsymbol{x}(k) \perp \boldsymbol{x}(k-1)$ in the adopted model are equivalent to $\boldsymbol{V}_k = \boldsymbol{V}_{k-1}$ and $\boldsymbol{V}_k \neq \boldsymbol{V}_{k-1}$, respectively, such that $\boldsymbol{V}_k$ and $\boldsymbol{V}_{k-1}$ can only be parallel or orthogonal to each other.

As remarked before, the BNDR-LMS algorithm behaves exactly like the NLMS algorithm when the input signal vector at instants k and $k-1$ are parallel. In this case, the excess of MSE is given by [4]

$$\Delta\xi(k+1)_{\|} = \left[1 + \frac{\mu(\mu-2)}{N+1}\right]\Delta\xi(k) + \frac{\mu^2}{(N+2-\nu_x)}\sigma_n^2 \tag{38}$$

where $\nu_x = E[x^4(k)/\sigma_x^4]$ is the *kurtosis* of the input signal, which varies from 1 for a binary distribution to 3 for a Gaussian distribution to ∞ for a Cauchy distribution [4], [17]. It must be stressed, however, that (38) holds only for $\nu_x \ll N+1$ [4].

For the case where $\boldsymbol{x}(k)$ and $\boldsymbol{x}(k-1)$ are always orthogonal, from (33) and (34), we have, for $\boldsymbol{R} = \sigma_x^2 \boldsymbol{I}$, i.e., white-noise input signals (see Appendix B)

$$\begin{aligned}\Delta\xi(k+1)_{\perp} = &\left[1 + \frac{\mu(\mu-2)}{N+1}\right]\Delta\xi(k) \\ &+ \frac{\mu(1-\mu)^2(\mu-2)}{N+1}\,\Delta\xi(k-1) \\ &+ \frac{\mu^2(\mu-2)^2}{N+2-\nu_x}\,\sigma_n^2. \end{aligned} \tag{39}$$

A final expression for the excess in the MSE may now be obtained from (38) and (39) combined and weighted accordingly, as suggested in (37). For a white input signal, the probabilities of $\boldsymbol{V}_k = \boldsymbol{V}_{k-1}$ and $\boldsymbol{V}_k \neq \boldsymbol{V}_{k-1}$ are equal to $1/N+1$ and $N/N+1$, respectively. The excess in the MSE is, therefore, given by

$$\begin{aligned}\Delta\xi(k+1) = &\left[1 + \frac{\mu(\mu-2)}{N+1}\right]\Delta\xi(k) \\ &+ \frac{N\mu(1-\mu)^2(\mu-2)}{(N+1)^2}\,\Delta\xi(k-1) \\ &+ \frac{\mu^2[1+N(\mu-2)^2]}{(N+1)(N+2-\nu_x)}\,\sigma_n^2. \end{aligned} \tag{40}$$

B. Colored Input Signal

Using the input-signal-vector model given in (35), we may now extend the analysis to colored input signals. The angular distribution of $\boldsymbol{x}(k)$ needs be changed in order to incorporate different probabilities for the directions given by the $(N+1)$ eigenvectors of $\boldsymbol{R}$. In other words, (37)–(39) are maintained, and only probabilities $P[\boldsymbol{x}(k)\|\boldsymbol{x}(k-1)]$ and $P[\boldsymbol{x}(k) \perp \boldsymbol{x}(k-1)]$ need to be recalculated. Each eigenvector of $\boldsymbol{R}$, which is denoted as $\mathcal{V}_i$, $i = 1, \cdots, N+1$ will now have the following probability of occurrence [4]:

$$P(\boldsymbol{V}_k = \mathcal{V}_i) = \frac{\lambda_i}{\mathrm{tr}(\boldsymbol{R})} \tag{41}$$

where λ_i is the eigenvalue associated to the eigenvector $\mathcal{V}_i$. For an easy association between $P[\boldsymbol{x}(k)\|\boldsymbol{x}(k-1)]$ and input-signal correlation, let us suppose that the input signal $x(k)$ is correlated by an allpole filter as in

$$x(k) = \gamma x(k-1) + (1-\gamma)\eta(k), \qquad 0 \leq \gamma \leq 1 \tag{42}$$

where $\eta(k)$ is a sample from an independent zero-mean process with variance given by σ_η^2. The autocorrelation matrix for this input signal can be easily derived and is expressed as

$$\boldsymbol{R} = \frac{1-\gamma}{1+\gamma}\,\sigma_\eta^2 \begin{bmatrix} 1 & \gamma & \gamma^2 & \cdots & \gamma^N \\ \gamma & 1 & \gamma & \cdots & \gamma^{N-1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \gamma^N & \gamma^{N-1} & \gamma^{N-2} & \cdots & 1 \end{bmatrix}. \tag{43}$$

From (43), we have all the necessary eigenvalues and eigenvectors such that we can compute

$$\begin{aligned} &P[\boldsymbol{x}(k)\|\boldsymbol{x}(k-1)] \\ &\quad = P[\boldsymbol{V}_k\|\boldsymbol{V}_{k-1}] \\ &\quad = P[\boldsymbol{V}_k\|\boldsymbol{V}_{k-1}|_{\boldsymbol{V}_{k-1}=\boldsymbol{V}_1}] \times P[\boldsymbol{V}_{k-1} = \boldsymbol{V}_1] + \cdots \\ &\qquad + P[\boldsymbol{V}_k\|\boldsymbol{V}_{k-1}|_{\boldsymbol{V}_{k-1}=\boldsymbol{V}_{N+1}}] \times P[\boldsymbol{V}_{k-1} = \boldsymbol{V}_{N+1}] \\ &\quad = \sum_{i=1}^{N+1} \left(\frac{\lambda_i}{\mathrm{tr}(\boldsymbol{R})}\right)^2 \end{aligned} \tag{44}$$

and

$$P[\boldsymbol{x}(k) \perp \boldsymbol{x}(k-1)] = 1 - P[\boldsymbol{x}(k)\|\boldsymbol{x}(k-1)]. \tag{45}$$

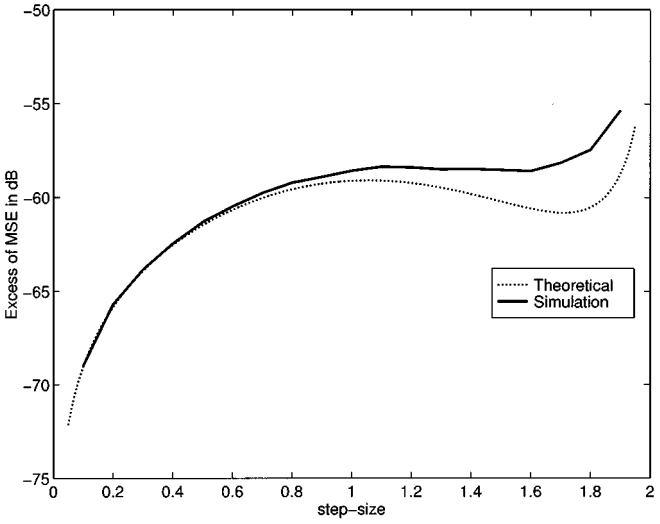

Fig. 2. Excess of MSE for $N = 10$ as a function of μ.

Equations (44) and (45) are in accordance with the white-input situation because this case corresponds to $\gamma = 0$, and all eigenvalues will be equal to σ_x^2 such that $P(\boldsymbol{V}_k = \mathcal{V}_i) = (1/N+1)$, as already described. When the input signal is correlated through a first-order allpole filter and modeled with (35) and (41), the excess of MSE is given by (37)–(39) with probabilities given by (44) and (45). Although (38) and (39) have been obtained based on a white Gaussian model for the input signal, simulations have shown that our reasoning is valid when the input signal is generated according to (35) with probabilities given by (41) and λ_i obtained from (43). Moreover, for $\mu = 1$ and a modeled input signal where only parallel or perpendicular vectors may occur, the BNDR-LMS algorithm degrades to the NLMS algorithm, and the steady-state MSE becomes independent of the radial distribution of $\boldsymbol{x}(k)$ [4]. This is perfectly described by (38) and (39), supporting the validity of our reasoning.

V. Simulation Results

In order to test the theoretical results obtained from the convergence analysis, the following experiment was carried out in a system identification problem: The input signal was white noise, and the excess of MSE was measured for different values of the step-size (μ varied from 0.1 to 1.9). As it was shown that the filter order N has a great influence on the theoretical results, the experiment was repeated for $N = 10$ and $N = 63$. The results are depicted in Figs. 2 and 3, respectively, where we can see that the theoretical curve is closer to the experimental curve as N is increased. Furthermore, as N is increased, the probability of $\boldsymbol{V}_k = \boldsymbol{V}_{k-1}$ occurring becomes less likely, and the curves tend to the situation where consecutive input-signal vectors are always orthogonal.

A second experiment was designed to test the influence of colored signals on the excess of MSE and the accuracy of the expressions derived in the analysis. Four situations were contemplated, corresponding to input signals having different characteristics. In the first two situations, signals were obtained from zero-mean white-Gaussian sequences filtered by first-order allpole IIR filters with poles at 0.8 and 0.9, yielding autocorrelation matrices with eigenvalue rates of 50.85 and 145.44, respectively. In the other two situations, input-signal vectors were generated with discrete radial probability distributions, and autocorrelation matrices with eigenvalue rates also equal to 50.85 and 145.44, respectively. The excess of MSE in decibels for these simulations are depicted in Fig. 4, where simulation results and theoretical curves are confronted. Theoretical values were calculated using (37)–(39) with probabilities given by (44) and (45). The analysis

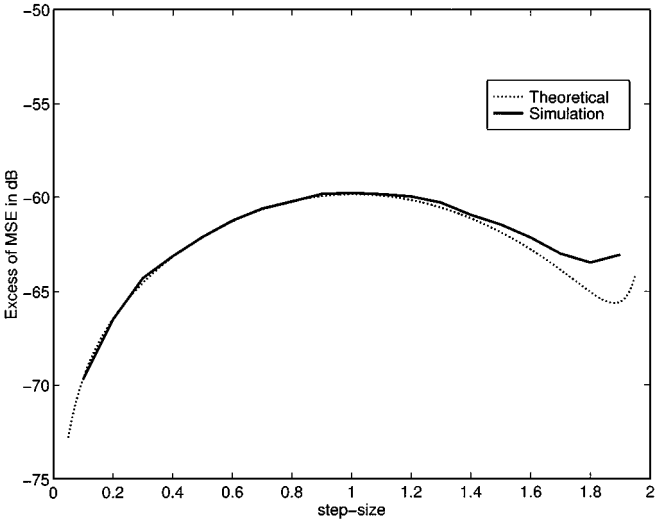

Fig. 3. Excess of MSE for $N = 63$ as a function of μ.

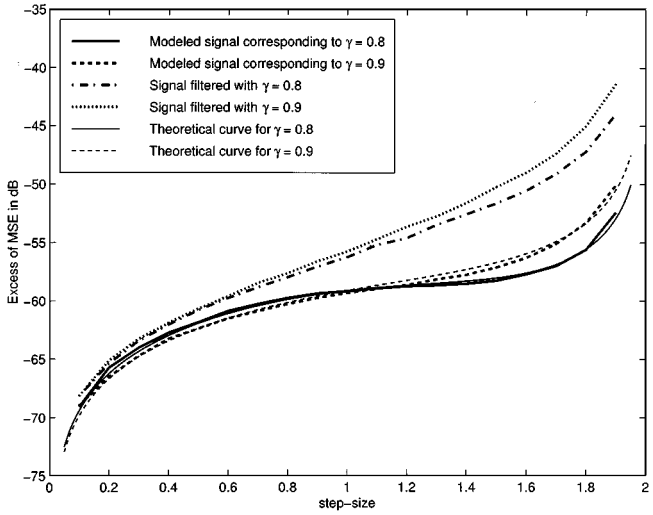

Fig. 4. Excess of MSE for colored input signals.

for colored input-signals presented very good agreement with the simulations carried out for input-signal vectors presenting discrete angular probability distributions. For the signals obtained by filtering white-Gaussian sequences with first-order allpole IIR filters, only a qualitative description of the evolution of the excess of MSE with respect to the step-size could be observed. This can be explained by the fact that the model characterizes correlated signals by increasing the probability of occurrence of parallel subsequent data vectors. It is assumed in the analysis that for this situation the two hyperplanes $\mathcal{S}(k-1)$ and $\mathcal{S}(k)$ are the same and that the algorithm adapts the coefficients in a single step, performing like the NLMS algorithm. In a more practical implementation, correlated subsequent data vectors are not likely to be parallel. In fact, the algorithm attempts to reach the intersection of nonparallel hyperplanes $\mathcal{S}(k-1)$ and $\mathcal{S}(k)$, reaching a solution that may be very distant from that of the NLMS algorithm. However, in the range of interest $(0 < \mu \leq 1)$, the difference between the simulated and the theoretical curves is less than 3 dB.

VI. Conclusions

This correspondence presented the analyses of convergence and mean-squared error for the BNDR-LMS algorithm.

These analyses in the mean and covariance were based on a simplified model for the input signal that rendered tractable expressions for the complex problem of analyzing data-reusing algorithms. Consistency with the first two moments of the input signal is maintained by the model. For white input signals, analysis of mean-squared error, which is in excellent agreement with simulations, was carried out. Limits for convergence in the mean and the covariance of the coefficient vector were also established. Moreover, a closed-form expression for the excess of MSE as a function of the step-size was derived for the case of white input signals. The applicability of this expression for the case of colored input signals was also addressed. The model and the analyses can be readily extended to other data-reusing algorithms that have not been considered in the past due to exceeding complexity.

Appendix A

1) Equation (24):

$$E[\boldsymbol{x}^T(k)\boldsymbol{x}(k-1)\boldsymbol{x}^T(k-1)\boldsymbol{x}(k)]$$
$$= E\left[\sum_{i=0}^{N}\sum_{j=0}^{N} x(k-i)x(k-1-i)x(k-1-j)x(k-j)\right]$$
$$= \sum_{i=0}^{N} E[x^2(k-i)x^2(k-1-i)]$$
$$= (N+1)\,(\sigma_x^2)^2. \tag{46}$$

2) Equation (25):

$$E\{\|\boldsymbol{x}(k)\|^2\|\boldsymbol{x}(k-1)\|^2 - [\boldsymbol{x}^T(k)\boldsymbol{x}(k-1)]^2\}$$
$$= E\left[\sum_{i=0}^{N}\sum_{j=0}^{N} x^2(k-i)x^2(k-1-j)\right] - E\{[\boldsymbol{x}^T(k)\boldsymbol{x}(k-1)]^2\}$$
$$= \sum_{i=0}^{N} E[x^2(k-i)] \sum_{j=0, j\neq i-1}^{N} E[x^2(k-1-j)] + \sum_{i=1}^{N} E[x^4(k-i)] - E\{[\boldsymbol{x}^T(k)\boldsymbol{x}(k-1)]^2\}$$
$$= (N^2+N+1)\,(\sigma_x^2)^2 + NE[x^4(k)] - (N+1)\,(\sigma_x^2)^2. \tag{47}$$

For stationary Gaussian-distributed signals, using the fourth-moment factoring theorem, we have $E[x^4(k)] = 3(\sigma_x^2)^2$ [17], and therefore

$$E\{\|\boldsymbol{x}(k)\|^2\|\boldsymbol{x}(k-1)\|^2 - [\boldsymbol{x}^T(k)\boldsymbol{x}(k-1)]^2\} = N(N+3)\,(\sigma_x^2)^2. \tag{48}$$

3) Equation (26): Let $[\cdot]_{ij}$ be the (i,j) element of matrix $[\cdot]$; then

$$E\{[\boldsymbol{x}(k-1)\boldsymbol{x}^T(k-1)\boldsymbol{x}(k)\boldsymbol{x}^T(k)]_{ij}\} = E\left[x(k-1-i)x(k-j)\sum_{l=0}^{N} x(k-1-l)x(k-l)\right]. \tag{49}$$

For Gaussian-distributed signals, we may use the fourth-moment factoring theorem to obtain

$$\begin{aligned}E\{[\boldsymbol{x}(k-1)\boldsymbol{x}^T(k-1)\boldsymbol{x}(k)\boldsymbol{x}^T(k)]_{ij}\}&=\sum_{l=0}^{N}\{E[x(k-1-i)x(k-j)]E[x(k-1-l)x(k-l)]\\&\quad+E[x(k-1-i)x(k-1-l)]E[x(k-j)x(k-l)]\\&\quad+E[x(k-1-i)x(k-l)]E[x(k-j)x(k-1-l)]\}\\&=\begin{cases}(\sigma_x^2)^2, & i=j \text{ or } i=j-2\\ 0, & \text{otherwise.}\end{cases}\end{aligned}\tag{50}$$

4) Equation (27): Once again, using the fact that for stationary Gaussian-distributed signals $E[x^4(k)]=3(\sigma_x^2)^2$ [17], we have

$$\begin{aligned}E\{[\|\boldsymbol{x}(k-1)\|^2\boldsymbol{x}(k)\boldsymbol{x}^T(k)]_{ij}\}&=E\left[\sum_{l=0}^{N}x^2(k-1-l)x(k-i)x(k-j)\right]\\&=\begin{cases}N(\sigma_x^2)^2+E[x^4(k)], & i=j\\ 0, & \text{otherwise}\end{cases}\\&=\begin{cases}(N+3)(\sigma_x^2)^2, & i=j\\ 0, & \text{otherwise.}\end{cases}\end{aligned}\tag{51}$$

APPENDIX B

1) Equation (39): In the derivation of (39), $\boldsymbol{x}(k)$ and $\boldsymbol{x}(k-1)$ were replaced by $s_k r_k V_k$ and $s_{k-1}r_{k-1}V_{k-1}$, respectively, with $V_k \perp V_{k-1}$. Therefore, $\boldsymbol{x}^T(k)\boldsymbol{x}(k-1)=0$. Furthermore, a second-order approximation for $E[1/r_k^2]$ was used [4], i.e.,

$$\begin{aligned}E\left[\frac{1}{\|\boldsymbol{x}(k)\|^2}\right]=E\left[\frac{1}{\|\boldsymbol{x}(k-1)\|^2}\right]=E\left[\frac{1}{r_k^2}\right]\\ \approx\frac{1}{(N+2-\nu_x)\sigma_x^2}\end{aligned}\tag{52}$$

where ν_x is the kurtosis of the input signal.

For $\boldsymbol{R}=\sigma_x^2\boldsymbol{I}$, using (33) and (34), the expression for $\Delta\xi(k)$ may be rewritten as

$$\begin{aligned}\Delta\xi(k+1)&=\sigma_x^2\,\mathrm{tr}\{\mathrm{cov}[\Delta\boldsymbol{w}(k+1)]\}\\&=\sigma_x^2\,\mathrm{tr}\{E[\Delta\boldsymbol{w}(k+1)\Delta\boldsymbol{w}^T(k+1)]\}\\&=\sigma_x^2\,\mathrm{tr}(E\{[\boldsymbol{I}+\mu\boldsymbol{A}]\Delta\boldsymbol{w}(k)\Delta\boldsymbol{w}^T(k)\,[\boldsymbol{I}+\mu\boldsymbol{A}]\})\\&\quad+\sigma_x^2\,\mathrm{tr}(E\{\mu[\boldsymbol{I}+\mu\boldsymbol{A}]\Delta\boldsymbol{w}(k)\boldsymbol{b}^T\})\\&\quad+\sigma_x^2\,\mathrm{tr}(E\{\mu\boldsymbol{b}\Delta\boldsymbol{w}^T(k)[\boldsymbol{I}+\mu\boldsymbol{A}]\})\\&\quad+\sigma_x^2\,\mathrm{tr}(E[\mu^2\boldsymbol{b}\boldsymbol{b}^T])\\&=\rho_1+\rho_2+\rho_3+\rho_4.\end{aligned}\tag{53}$$

Evaluating each of these terms separately, we obtain

$$\begin{aligned}\rho_1&=\sigma_x^2\,\mathrm{tr}(E\{[\boldsymbol{I}+\mu\boldsymbol{A}]\Delta\boldsymbol{w}(k)\Delta\boldsymbol{w}^T(k)[\boldsymbol{I}+\mu\boldsymbol{A}]\})\\&=\sigma_x^2\,\mathrm{tr}\{\,\mathrm{cov}[\Delta\boldsymbol{w}(k)]\}\\&\quad-\mu\sigma_x^2\,\mathrm{tr}\left\{E\left[\frac{\boldsymbol{x}(k-1)\boldsymbol{x}^T(k-1)\Delta\boldsymbol{w}(k)\Delta\boldsymbol{w}^T(k)}{\|\boldsymbol{x}(k-1)\|^2}\right]\right\}\\&\quad-\mu\sigma_x^2\,\mathrm{tr}\left\{E\left[\frac{\boldsymbol{x}(k)\boldsymbol{x}^T(k)\Delta\boldsymbol{w}(k)\Delta\boldsymbol{w}^T(k)}{\|\boldsymbol{x}(k)\|^2}\right]\right\}\\&\quad-\mu\sigma_x^2\,\mathrm{tr}\left\{E\left[\frac{\Delta\boldsymbol{w}(k)\Delta\boldsymbol{w}^T(k)\boldsymbol{x}(k-1)\boldsymbol{x}^T(k-1)}{\|\boldsymbol{x}(k-1)\|^2}\right]\right\}\\&\quad-\mu\sigma_x^2\,\mathrm{tr}\left\{E\left[\frac{\Delta\boldsymbol{w}(k)\Delta\boldsymbol{w}^T(k)\boldsymbol{x}(k)\boldsymbol{x}^T(k)}{\|\boldsymbol{x}(k)\|^2}\right]\right\}\\&\quad+\mu^2\sigma_x^2\,\mathrm{tr}\left(E\left\{\frac{\boldsymbol{x}(k-1)\boldsymbol{x}^T(k-1)\Delta\boldsymbol{w}(k)}{\|\boldsymbol{x}(k-1)\|^2}\right.\right.\\&\quad\left.\left.\times\frac{\Delta\boldsymbol{w}^T(k)\boldsymbol{x}(k-1)\boldsymbol{x}^T(k-1)}{\|\boldsymbol{x}(k-1)\|^2}\right\}\right)\\&\quad+\mu^2\sigma_x^2\,\mathrm{tr}\left\{E\left[\frac{\boldsymbol{x}(k)\boldsymbol{x}^T(k)\Delta\boldsymbol{w}(k)\Delta\boldsymbol{w}^T(k)\boldsymbol{x}(k-1)\boldsymbol{x}^T(k-1)}{\|\boldsymbol{x}(k)\|^2\|\boldsymbol{x}(k-1)\|^2}\right]\right\}\\&\quad+\mu^2\sigma_x^2\,\mathrm{tr}\left\{E\left[\frac{\boldsymbol{x}(k-1)\boldsymbol{x}^T(k-1)\Delta\boldsymbol{w}(k)\Delta\boldsymbol{w}^T(k)\boldsymbol{x}(k)\boldsymbol{x}^T(k)}{\|\boldsymbol{x}(k)\|^2\|\boldsymbol{x}(k-1)\|^2}\right]\right\}\\&\quad+\mu^2\sigma_x^2\,\mathrm{tr}\left(E\left\{\frac{\boldsymbol{x}(k)\boldsymbol{x}^T(k)\Delta\boldsymbol{w}(k)\Delta\boldsymbol{w}^T(k)\boldsymbol{x}(k)\boldsymbol{x}^T(k)}{[\|\boldsymbol{x}(k)\|^2]^2}\right\}\right)\\&=\psi_1+\psi_2+\cdots+\psi_9\end{aligned}\tag{54}$$

where

$$\begin{aligned}\psi_1&=\Delta\xi(k)\\\psi_2&=-\frac{\mu(1-\mu)^2}{(N+1)}\Delta\xi(k-1)-\frac{\mu^3\sigma_n^2}{(N+2-\nu_x)}\\\psi_3&=-\frac{\mu}{(N+1)}\Delta\xi(k).\end{aligned}\tag{55}$$

Recalling that $\mathrm{tr}[\boldsymbol{AB}]=\mathrm{tr}[\boldsymbol{BA}]$ for any square matrices $\boldsymbol{A}$ and $\boldsymbol{B}$, we find that

$$\begin{aligned}\psi_4&=\psi_2\\\psi_5&=\psi_3\\\psi_6&=-\mu\psi_2\\\psi_7&=\psi_8=0\\\psi_9&=\frac{\mu^2}{N+1}\Delta\xi(k).\end{aligned}\tag{56}$$

Therefore

$$\begin{aligned}\rho_1=&\left(1+\frac{\mu(\mu-2)}{N+1}\right)\Delta\xi(k)+\frac{(1-\mu)^2\mu(\mu-2)}{N+1}\Delta\xi(k-1)\\&+\frac{\mu^3(\mu-2)}{N+2-\nu_x}\sigma_n^2.\end{aligned}\tag{57}$$

Similarly

$$\rho_2=\rho_3=\frac{\mu^2(1-\mu)}{(N+2-\nu_x)}\sigma_n^2\tag{58}$$

$$\rho_4=\frac{2\mu^2\sigma_n^2}{(N+2-\nu_x)}.\tag{59}$$

From (57)–(59), the difference equation for $\Delta\xi(k)$ is finally obtained as in (39)

$$\begin{aligned}\Delta\xi(k+1)=&\left[1+\frac{\mu(\mu-2)}{N+1}\right]\Delta\xi(k)\\&+\frac{(1-\mu)^2\mu(\mu-2)}{N+1}\Delta\xi(k-1)\\&+\frac{\mu^2(\mu-2)^2}{N+2-\nu_x}\sigma_n^2.\end{aligned}\tag{60}$$

REFERENCES

[1] P. S. R. Diniz, *Adaptive Filtering—Algorithms and Practical Implementation.* Norwell, MA: Kluwer, 1997.

[2] S. Haykin, *Adaptive Filter Theory*, 2nd ed. Englewood Cliffs, NJ: Prentice-Hall, 1991.

[3] S. Roy and J. J. Shynk, "Analysis of the data-reusing LMS algorithm," in *Proc. 32nd Midwest Symp. Circuits Syst.*, Urbana, IL, 1990, pp. 1127–1130.

[4] D. T. Slock, "On the convergence behavior of the LMS and the normalized LMS algorithms," *IEEE Trans. Signal Processing*, vol. 41, pp. 2811–2825, Sept. 1993.
[5] W. K. Jenkins, A. W. Hull, J. C. Strait, B. A. Schnaufer, and X. Li, *Advanced Concepts in Adaptive Signal Processing*. Norwell, MA: Kluwer, 1996.
[6] B. A. Schnaufer, "Practical Techniques for Rapid and Reliable Real-Time Adaptive Filtering," Ph.D. dissertation, Univ. Illinois Urbana-Champaign, 1995.
[7] J. A. Apolinário, Jr., M. L. R. de Campos, and P. S. R. Diniz, "The binormalized data-reusing LMS algorithm," in *Proc. XV Simpósio Brasileiro de Telecomunicações*, Recife, Brazil, 1997, pp. 77–80.
[8] ——, "Convergence analysis of the binormalized data-reusing LMS algorithm," in *Proc. Euro. Conf. Circuit Theory Des.*, Budapest, Hungary, 1997, pp. 972–977.
[9] M. Montazeri and P. Duhamel, "A set of algorithms linking NLMS and block RLS algorithms," *IEEE Trans. Signal Processing*, vol. 43, pp. 444–453, Feb. 1995.
[10] K. Ozeki and T. Umeda, "An adaptive filtering algorithm using an orthogonal projection to an affine subspace and its properties," *Electron. Commun. Japan*, vol. 67-A, no. 5, pp. 19–27, 1984.
[11] S. L. Gay, "Fast Projection Algorithms with Application to Voice Echo Cancellation," Ph.D. dissertation, Rutgers Univ., New Brunswick, NJ, Oct. 1994.
[12] S. L. Gay and S. Tavathia, "The fast affine projection algorithm," in *Proc. Int. Conf. Acoust., Speech, Signal Process.*, Detroit, MI, 1995, pp. 3023–3026.
[13] J. E. Mazo, "On the independence theory of equalizer convergence," *Bell Syst. Tech. J.*, vol. 58, pp. 962–993, May–June 1979.
[14] M. L. R. de Campos and A. Antoniou, "A new quasi-Newton adaptive filtering algorithm," *IEEE Trans. Circuits Syst. II*, vol. 44, pp. 924–934, Nov. 1997.
[15] J. A. Apolinário, Jr., P. S. R. Diniz, T. Laakso, and M. L. R. de Campos, "Step-size optimization of the BNDR-LMS algorithm," in *Proc. Euro. Signal Process. Conf.*, Rhodes, Greece, Sept. 1998.
[16] P. S. R. Diniz, M. L. R. de Campos, and A. Antoniou, "Analysis of LMS-Newton adaptive filtering algorithms with variable convergence factor," *IEEE Trans. Signal Processing*, vol. 43, pp. 617–627, Mar. 1995.
[17] A. Papoulis, *Probability, Random Variables, and Stochastic Processes*, 3rd ed. New York: McGraw-Hill, 1991.

Multiresolution Circular Harmonic Decomposition

Giovanni Jacovitti and Alessandro Neri

***Abstract*—A dictionary of complex waveforms suited for multiresolution analysis and individually steerable by multiplication by a complex factor is presented. It is based on circular harmonic wavelets (CHW) and is useful for pattern analysis under rotations. The main theoretical aspects of CHWs are illustrated, and an example of application to motion estimation is provided.**

***Index Terms*—Circular harmonic wavelets, complex multiscale representation, self-steerable pyramid.**

I. Introduction

In many applications of computer vision, it is desired that pattern rotations are characterized by few parameters in the representation domain. In order to get a basis for multiresolution image representation possessing such a property, a wavelet expansion based on "steerable" functions has been introduced in [1]. Steerable filters are obtained as linear combinations of a finite set of basic point-spread functions (PSFs) so that the spatial orientation is determined by properly tuning the values of the weighting coefficients (steering coefficients) [2]. By extension, steerable pyramids are obtained by combination of basic multiresolution analyzes.

In this work, we present another approach to steerability based on complex wavelets called circular harmonic wavelets (CHWs). The CHWs are polar-separable wavelets, with harmonic angular shape. This implies that they are steered in any wanted direction by simple multiplication with a complex steering factor (we will refer this property to as *self-steerability*). The CHWs have been introduced by the authors in [3] and [4] starting from the circular harmonic functions (CHFs), yet employed in optical correlators for rotation invariant pattern recognition (see, for instance, [5]). The same functions also appear in harmonic tomographic decomposition [6] and have been considered for local image symmetry analysis [7]. In addition, recently, CHFs have been employed for the definition of rotation-invariant pattern signatures [8]. In this contribution, a family of orthogonal CHWs forming a multiresolution pyramid called the circular harmonic pyramid (CHP) is presented. In essence, each CHW pertaining to the pyramid represents the image by translated, dilated, and rotated versions of a CHF. At the same time, for a fixed resolution, the CHP orthogonal system provides a local representation of the image around any point in terms of CHFs. The self-steerability of each component of the CHP can be exploited for pattern analysis in presence of rotation (other than translation and dilation) and, in particular, for pattern recognition, irrespective of orientation.

This paper is organized as follows. In Section II, basic definitions and theorems are provided. In Section III, the Laguerre–Gauss expansion of the image around a generic point of the complex plane is presented along with the related Laguerre–Gauss pyramid (LGP). In Section IV, an application of the CHPs to local motion estimation in presence of non-negligible rotations is briefly illustrated.

Manuscript received March 5, 1997; revised July 20, 2000. The associate editor coordinating the review of this paper and approving it for publication was Dr. Stephane Mallat.

G. Jacovitti is with the INFOCOM Department, University of Rome "La Sapienza," Rome, Italy.

A. Neri is with the Department of Electronic Engineering, University of ROMA TRE, Rome, Italy.

Publisher Item Identifier S 1053-587X(00)09316-8.