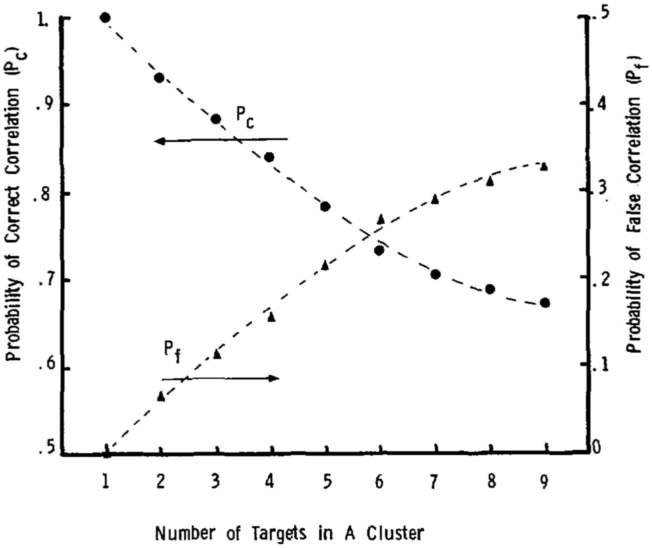

Fig. 1. Correlation performance versus target density; no redundant correlations allowed.

4) Using one such subset of sensors, compute the sufficient statistic, $l_{i_1,i_2,\cdots,i_M}$, using their measurements with (2.1), (2.2) or (2.3), (2.4) for the Gaussian measurement noise case.

5) For a desired leakage probability, a threshold for $l_{i_1,\cdots,i_M}$ can be chosen using x^2 statistics. Apply the threshold to all $l_{i_1,\cdots,i_M}$'s. For those cases which are above the threshold, the corresponding measurements are declared uncorrelated. This may again result in some measurements not being correlated at all.

6) Apply the Hungarian/Munkre-type algorithm to process the submatrices resulting from thresholding. This gives the correlation solution.

7) Let $l_{i_1,i_2,\cdots,i_M}$ denote an entry representing a set of correlated measurements. Fix $i_2,\cdots,i_M$; search those $i_1,i_2,\cdots,i_M$ for all i's that are within a threshold value of $l_{i_1,i_2,\cdots,i_M}$. Those entries which are within the threshold define a set of multiple (or redundant) correlations. This step is repeated for all i_j's and all correlations accepted in step 3).

8) Repeat steps 4)–7) to process all sensor subsets identified by steps 1)–3).

III. A Numerical Example

In this section, we consider an example which illustrates the target handover problem between two sensors. Consider the case of two optical sensors tracking a complex of ballistic targets. The first sensor tracks a subset of targets for 100 s, then predicts their state vectors 1000 s later. The second sensor tracks another subset of targets which partially overlaps that of the first sensor also for 100 s. The end of track time of the second sensor is the same as the end of the prediction time of the first sensor. The correlation of estimates from two sensors constitutes a typical handover problem.

The probability of correct correlation (P_c) and the probability of false correlation (P_f) are defined as follows:

$$P_c = \frac{\text{number of correct correlations}}{\text{number of targets}}$$

$$P_f = \frac{\text{number of incorrect correlations}}{\text{number of correlations declared}}.$$

A covariance analysis technique (the Cramer–Rao bound) was applied in [6] to obtain the track accuracy. The simulated state estimate used in this note is obtained by corrupting true states with random vectors with covariances being equal to the Cramer–Rao bound. Correlation performance in terms of P_c and P_f is obtained with one hundred repetitions on each target.

The P_c and P_f as functions of the number of targets in a cluster is shown in Fig. 1. We note that a chi-square threshold of 30 was used for these results. Redundant matches were not allowed; P_c and P_f therefore sum to one. If one allows for redundant correlations, the P_c can be improved substantially together with the increase of P_f. Fig. 2 shows such a tradeoff. We note that the redundant correlation is possible by accepting those targets whose chi-squares are within a certain range of that of the optimum choice (see step 7) of the algorithm). The dotted line denotes the case of not allowing redundant correlations. Notice that the increase in P_f for maintaining $P_c > 0.9$ is rather small.

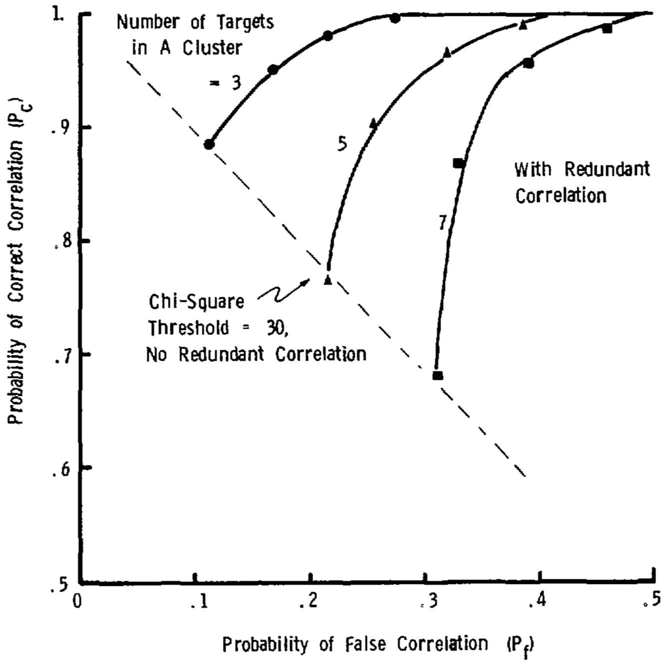

Fig. 2. Operating characteristics for sensor-to-sensor correlation.

References

[1] Y. Bar-Shalom, "Tracking methods in a multi-target environment," *IEEE Trans. Automat. Contr.*, vol. AC-23, pp. 618–626, Aug. 1978.
[2] D. Reid, "An algorithm for tracking multiple targets," *IEEE Trans. Automat. Contr.*, vol. AC-24, pp. 843–854, Dec. 1979.
[3] D. Willner, C. B. Chang, and K. P. Dunn, "Kalman filter algorithms for a multi-sensor system," in *Proc. 1976 IEEE Conf. Decision Contr.*, Dec. 1976.
[4] J. Munkres, "Algorithms for the assignment and transportation problems," *SIAM J.*, vol. 5, pp. 32–38, Mar. 1957.
[5] C. B. Chang and L. C. Youens, "Measurement correlation for multiple sensor tracking in a dense target environment," Lincoln Lab., M.I.T., Cambridge, MA, Tech. Rep. 549, Jan. 1981.
[6] C. B. Chang, "Ballistic trajectory estimation with angle-only measurements," *IEEE Trans. Automat. Contr.*, vol. AC-25, pp. 474–480, June 1980.

Iterative Least-Squares Parameter Estimation for ARMA Pulse Response and Output Disturbance

P. LÖHNBERG AND G. H. J. WISSELINK

Abstract—**The parameters of an ARMA system and of additive ARMA output disturbance with mutually different poles were estimated from a record of the disturbed output on a single pulse input. Alternating the steps of iterative inverse filtering with iterations according to generalized least squares appeared very economical for this system. This procedure converged for second-order simulations. Statistics of the estimates were evaluated from these simulation results for two signal-to-noise ratios.**

Manuscript received September 24, 1981; revised March 29, 1982 and May 27, 1982.
P. Löhnberg is with the Department of Electrical Engineering, Twente University of Technology, 7500 AE Enschede, The Netherlands.
G. H. J. Wisselink is with Vitatron Medical B.V., 6950 AB Dieren, The Netherlands.

0018-9286/82/1200-1252$00.75 ©1982 IEEE

INTRODUCTION

In several (e.g., biological) systems, control may only be impulse-like (e.g., administration of drugs, injections, or electrical stimulation of nerves and muscles). Output measurements are generally performed by taking samples (e.g., blood samples, number of nerve cell depolarizations per interval, or analog-to-digital conversions). The measurements are often corrupted by offset and disturbance, which may be assumed uncorrelated and additive. Insight into the system environment, unbiased estimation, and optimal control requires estimation of disturbance characteristics apart from system characteristics.

The response could be separated from the disturbance by averaging over a considerable number of responses to equal impulses. As an effect, nonparametric models could be obtained. In practice, the costs of each impulse (e.g., discomfort for a patient) may be very high. Further, time variance may occur. Therefore, it may often be preferred to identify the system by a single impulse. Then separation of response and disturbance requires a parametric model for each.

A generally applicable response model with a few identifiable parameters is an autoregressive moving average (ARMA) model. The disturbance can also be modeled most economically by an ARMA model with an uncorrelated sequence as input [1]. This yields the system approximation in the top of Fig. 1, characterized by output samples

$$y = v + w + g \tag{1}$$

with z transforms (omitting z^{-1} and (z) for simplicity)

$$v = \frac{A}{B} u, \tag{2}$$

$$w = \frac{C}{D} e \tag{3}$$

with shift operators

$$A = \sum_{i=0}^{p} a_i z^{i-p}, \qquad B = 1 + \sum_{i=1}^{q} b_i z^{i-q},$$
$$C = 1 + \sum_{i=1}^{r} c_i z^{i-r}, \qquad D = 1 + \sum_{i=1}^{s} d_i z^{i-s} \tag{4}$$

and $u = S$ for the Mth sample and 0 for the other ones. e is assumed to be normally distributed with average 0 and standard deviation σ_e.

Identification from a single record of N samples of y, containing as much of the response as possible, should make optimal use of all information contained. That can be carried out most economically by a nonrecursive identification procedure. Of known nonrecursive methods for this model, gradient procedures like prediction error methods [2] require high computational effort, quasi-linearization [3] requires $D = B$, and the instrumental variable procedure is not efficient [1] and requires a lot of art. Most simple and economic procedures are least-squares (LS) methods.

For the generalized least-squares (GLS) method [4], C must be 1. For the extended matrix method [5], B and D may not contain common factors, as this procedure is insensitive to these. For the general case that such requirements cannot be guaranteed or accurately met by approximation, an efficient procedure is described here. It consists of alternating the iteration steps of the GLS procedure for the response parameters with those of the iterative inverse filtering (ITIF) procedure [6] for the disturbance parameters.

Section II describes this generalized least-squares iterative inverse filtering (GLSITIF) procedure, and Section III describes estimator statistics from simulations. The procedure appeared applicable to a real world example from neuropathophysiology [7].

PROCEDURE

The procedure, which is motivated and described below, is indicated schematically in the lower part of Fig. 1. First, for $\bar{u} = 0$, g is estimated by the average $\hat{g}$ of y. This yields

$$y' = y - \hat{g} \approx v + w. \tag{5}$$

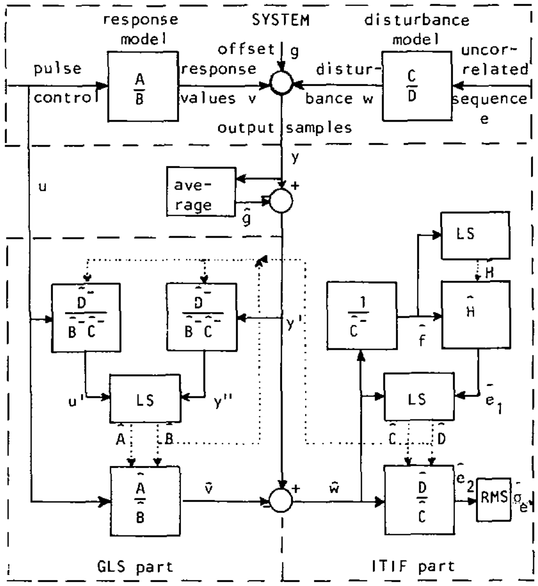

Fig. 1. Schematic diagram of the GLSITIF procedure.

It follows from (5) with (2), (3) that the equation error

$$By' - Au \approx \frac{BC}{D} e \tag{6}$$

is not white (= uncorrelated). As a result, direct LS estimation of A and B from y' and u would yield biased estimates. Therefore the signals are filtered as in the GLS procedure [4]:

$$u' = \frac{\hat{D}^-}{\hat{B}^- \hat{C}^-} u,$$
$$y'' = \frac{\hat{D}^-}{\hat{B}^- \hat{C}^-} y' \tag{7}$$

with $\hat{B}^-$, $\hat{C}^-$, and $\hat{D}^-$ estimates from the former iteration step. Considerations for the first step are given in the simulation section. Substitution of (7) into (6) yields

$$By'' - Au' \approx e \tag{8}$$

which is white. Hence the subsequent estimation from y'' and u' according to (8) yields estimates $\hat{A}$ and $\hat{B}$ which may converge without bias.

Response and disturbance are approximated by

$$\hat{v} = \frac{\hat{A}}{\hat{B}} u, \tag{9}$$

$$\hat{w} = y' - \hat{v}. \tag{10}$$

C and D have to be estimated using the result from (3)

$$Ce = Dw \tag{11}$$

which requires an estimate $\hat{e}$ of e. As such, the inverse of (3)

$$\hat{e}^- = \frac{\hat{D}^-}{\hat{C}^-} \hat{w} \tag{12}$$

would depend heavily on $\hat{C}^-$ and $\hat{D}^-$ and therefore cause slow convergence. One-step procedures, which do not make use of $\hat{C}^-$ or $\hat{D}^-$, are quite laborious. Hence it was decided to make use of the economic iteration steps of the ITIF procedure [6], which only use $\hat{C}^-$, by forming an estimate

$$\hat{f} = \frac{1}{\hat{C}^-} \hat{w} \tag{13}$$

TABLE I
SIMULATED PARAMETER VALUES WITH ESTIMATE STATISTICS

Parameter	Simulated value	36 realizations with S = 30		43 realizations with S = 90	
		Relative bias	Relative standard deviation	Relative bias	Relative standard deviation
a_0	0.038	0.014	0.109	0.004	0.032
a_1	-0.014	-0.060	0.339	0.001	0.113
a_2	-0.029	0.021	0.158	-0.002	0.052
b_1	-1.791	-0.003	0.009	0.000	0.002
b_2	0.954	-0.004	0.017	0.000	0.004
c_1	-0.320	0.065	0.198	0.079	0.165
c_2	-0.005	1.929	10.767	-1.779	14.870
d_1	-1.680	-0.005	0.020	-0.002	0.018
d_2	0.920	-0.004	0.033	-0.005	0.033
σ_e	0.140	-0.031	0.044	-0.015	0.042

of intermediate signal $f = e/D$. Substitution into (13) of (3) yields

$$\hat{f} \approx \frac{1}{\hat{C}^-}\frac{C}{D}e \approx \frac{1}{H}e \Rightarrow e \approx H\hat{f} \tag{14}$$

with H of order r (of C)$+s$ (of D). Estimate $\hat{H}$ of H is obtained from $\hat{f}$ by LS estimation according to (14), after which a first estimate of e is obtained according to (14) as

$$\hat{e}_1 = \hat{H}\hat{f}. \tag{15}$$

Then $\hat{C}$ and $\hat{D}$ are LS estimated from $\hat{w}$ and $\hat{e}_1$ according to (11).

As $\hat{e}_1$ depends strongly on $\hat{C}^-$, a second estimate of e is obtained by the inverse of (3)

$$\hat{e}_2 = \frac{\hat{D}}{\hat{C}}\hat{w} \tag{16}$$

in order to estimate σ_e by the root mean square (rms) $\hat{\sigma}_e$ of $\hat{e}_2$.

Now that all parameters have been estimated, some stopping criterion may be checked. A useful criterion is whether or not the mean-square value of all relative changes in parameter estimates per iteration is lower than a small factor (e.g., 10^{-10}), or if an allowed number of iterations (e.g., 30) has been reached. If not, the next iteration step starts at formula (7).

After the stopping criterion has been met, it may be checked whether or not the loss function $\hat{\sigma}_e$ is minimal within the direct neighborhood of the other parameter estimates [8]. If the model structure and procedure are correct, $\hat{e}_2$ should approximate samples of an uncorrelated process. This may be checked by a whiteness test as described in [9]. In that procedure, it is checked whether or not the normalized integrated sample power spectral estimate of $\hat{e}_2$ remains within the acceptance interval at a given level of significance.

APPROXIMATE ESTIMATOR STATISTICS FROM SIMULATIONS

The model was simulated for pulse instant $M = 100$, number of samples $N = 300$, orders $p = q = r = s = 2$, and parameter values as shown in Table I. This represents near-oscillatory response and disturbance models. Noise e was started 500 samples before the first simulation in order to reach a reasonable initial state. For next simulations, each realization of e was obtained as the continuation of the former realization. This was repeated for the number of realizations of e and the values of pulse height S indicated in Table I.

The described stopping criterion was met in 2–8 iterations of 1 s each on a DEC 10 computing system, if no filtering (7) was applied for the first iteration step. Substitution of approximate parameter values for $\hat{B}^-$, $\hat{C}^-$, and $\hat{D}^-$ in (7) in the first iteration step did not improve the final results, but caused convergence in considerably fewer iteration steps.

The estimates appeared approximately normally distributed. Table I shows their relative bias and standard deviation. It can be seen from Table I that all parameters except the relatively small c_2 were estimated reasonably well and that the response parameter estimates improve for higher signal-to-noise ratio S/σ_e. This improvement of estimate quality should be weighed against adverse effects such as increasing discomfort for a patient, reaching a nonlinear range and increase of quantization errors by the finite ADC range.

CONCLUSION

The GLSITIF procedure appears to converge quickly for this most general model type and with reasonable estimate statistics for these simulations. Further use, comparison and/or development of analysis methods are needed to thoroughly test merits, convergence conditions, and estimate statistics for arbitrary models.

ACKNOWLEDGMENT

The authors would like to thank Prof. H. Kwakernaak and Prof. H. B. K. Boom for their discussions about this procedure, and G. G. M. Steijlen for drawings and typework.

REFERENCES

[1] D. Q. Mayne and F. Firoozan, "Linear estimation of ARMA systems," in *Proc. 7th IFAC World Congr.*, Helsinki, Finland, June 12–16, 1978, vol. 3, pp. 1907–1912.
[2] L. Ljung, "On the consistency of prediction error identification methods," in *System Identification: Advances and Case Studies*, R. K. Mehra and D. G. Lainiotis, Eds. New York: Academic, 1976, pp. 121–164.
[3] B. P. Furth, "Maximum likelihood identification of Åström model by quasilinearization," in *Proc. 3rd IFAC Symp. Ident. Syst. Parameter Est.*, Delft, The Netherlands, P. Eykhoff, Ed. Amsterdam, The Netherlands: North-Holland, part 2, pp. 737–740.
[4] R. Hastings-James and M. W. Sage, "Recursive generalised-least-squares procedure for online identification of process parameters," *Proc. IEE*, vol. 116, pp. 2057–2062, Dec. 1969.
[5] J. L. Talmon and A. J. W. Van den Boom, "On the estimation of the transfer function parameters of process- and noise dynamics using a single-stage estimator," in *Proc. 3rd IFAC Symp. Ident. Syst. Parameter Est.*, Delft, The Netherlands, 1973, P. Eykhoff, Ed. Amsterdam, The Netherlands: North-Holland, part 2, pp. 711–720.
[6] I. S. Konvalinka and M. R. Matausek, "Simultaneous estimation of poles and zeros in speech analysis and ITIF—Iterative inverse filtering algorithm," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-27, pp. 485–492, Oct. 1979.

[7] P. Löhnberg, H. B. K. Boom, and H. M. Greebe, "Undamped natural frequency estimation of tremor and reflex response for Parkinson patients," in *Proc. 6th IFAC Symp. Ident. Syst. Parameter Est.*, Washington, DC, 1982, G. A. Bekey and G. N. Saridis, Eds. Oxford, England: Pergamon.
[8] K. J. Aström, "Maximum likelihood and prediction error methods," in *Tutorials on System Identification*, R. Iserman, Ed. Darmstadt, Germany: Institut für Regelungstechnik, 1979, pp. 53–67.
[9] G. M. Jenkins and D. G. Watts, *Spectral Analysis and Its Applications*. San Francisco, CA: Holden-Day, 1968.

Asymptotic Behavior of an Adaptive Estimation Algorithm with Application to M-Dependent Data

S. ABU EL ATA

***Abstract*—Theoretical results are presented concerning the asymptotic behavior of the parameter estimates generated by an adaptive algorithm in stationary dependent random situations. Proofs are exhibited in the general case and derived for the case of statistical dependence in the input for a finite number of lags. It is found that the estimate mean-error norm converges to an asymptotic bound, giving a finite bias, generally nonzero, and that the asymptotic mean-norm square error is bounded and can be arbitrarily reduced by decreasing the adaptation factor.**

I. Introduction

Let $\{s_n\}$ and $\{e_n\}$ denote the sequences of scalar outputs and N-dimensional vector inputs of the system, respectively. We suppose that at any instant n the input–output relation is given by

$$s_n = a^T e_n \tag{1}$$

where a represents the N-dimensional parameter vector to be estimated and T denotes transposition.

Assuming e_n never tends to the null vector, we use the following gradient-type estimation algorithm:

$$a_n = a_{n-1} + \mu \frac{(s_n - a_{n-1}^T e_n) e_n}{\|e_n\|^2} \tag{2}$$

initiating it at a chosen vector a_0. μ is a positive adaptation constant and the symbol $\|\cdot\|$ denotes the Euclidian norm and will also be used to denote the induced matrix norm.

Putting $v_n = a_n - a$ and $A_n = e_n e_n^T / \|e_n\|^2$ and using (1), (2) gives

$$v_n = (I - \mu A_n) v_{n-1}, \tag{3}$$

I being the $N \times N$ identity matrix. We note that the eigenvalues of A_n are 0, with multiplicity $N-1$, and 1.

In many practical situations, the measured output contains a residual term $r_n = s_n - a^T e_n$ not taken into account by model (1). This term may be due to noise, truncation errors, nonlinear terms, etc. The homogeneous difference equation (3) must thus be replaced by the nonhomogeneous difference equation

$$v_n = (I - \mu A_n) v_{n-1} + \mu \frac{r_n e_n}{\|e_n\|^2}. \tag{4}$$

Introducing the transition matrices $\Phi(i,j)$ $i \leqslant j$ defined by

$$\Phi(i,j) = \begin{cases} \prod_{\substack{k=i+1 \\ i<j}}^{j} (I - \mu A_k) = (I - \mu A_j)(I - \mu A_{j-1}) \cdots (I - \mu A_{i+1}) \\ I \qquad i = j, \end{cases}$$

Manuscript received April 13, 1981; revised February 22, 1982 and April 21, 1982. This work was supported by Direction des Recherches, Etudes et Techniques of the Ministry of Defense under Grant 80/111.
The author is with Adersa-Gerbios, 91120 Palaiseau, France.

the error equations (3) and (4) give, respectively,

$$v_n = \Phi(0,n) v_0 \tag{5}$$

and

$$v_n = \Phi(0,n) v_0 + \mu \sum_{j=1}^{n} \Phi(j,n) \frac{r_j e_j}{\|e_j\|^2}. \tag{6}$$

In the following section, we consider the asymptotic behavior of the algorithm for stationary dependent data, extending the analysis of [4]. It is shown that under conditions easily satisfied in practice, the algorithm has "good" asymptotic behavior. Section III is devoted to a specified example of the dependence of the inputs: the M-dependence case.

II. Asymptotic Behavior of the Algorithm

Sufficient conditions for the exponential convergence of the homogeneous algorithm (5) in stationary dependent situations are expressed in the following theorem given in [3].

Theorem 1: With ergodic inputs $\{e_n\}$, the series $\{v_n\}$ converges exponentially to 0 almost surely and in mean square for $\mu \in (0,2)$ if there exists a finite integer l such that

$$E\left[\lambda_{\min} \sum_{i=1}^{l} A_i\right] > 0, \tag{7}$$

E being the mathematical expectation and $\lambda_{\min}$ being the minimum eigenvalue. Furthermore (7) holds, provided $\{e_n\}$ has full-rank covariance and its fourth moment exists.

A calculation of the convergence rate of the homogeneous algorithm is given in [2] which indicates that under the conditions of Theorem 1, there exists ζ, $0 < \zeta < 1$ such that

$$\|v_n\|^2 \leqslant (1-\zeta)^n \|v_0\|^2 \tag{8}$$

with ζ depending on μ and the statistical properties of the inputs. Also ζ is linear in μ to the first order.

Now assuming that the homogeneous algorithm is exponentially convergent, we analyze the asymptotic behavior of the nonhomogeneous algorithm (6) by studying $\lim_{n\to\infty} \|E(v_n)\|$ which gives the bias of the estimates and $\lim_{n\to\infty} E\|v_n\|^2$ which is the asymptotic value of the mean-square error. It is found that these quantities are bounded, their bounds being reached exponentially fast. The obtained results are presented in the following two theorems.

Theorem 2: If

$$E\|e_n\|^2 \leqslant C, \qquad E(r_n^2) \leqslant C,^1 \tag{9}$$

then

$$\lim_{n\to\infty} \|E(v_n)\| \leqslant C. \tag{10}$$

Proof: From (6) and the triangle inequality, we can write

$$\|E(v_n)\| \leqslant E\|v_n\| \leqslant E\|\Phi(0,n) v_0\| + \mu \sum_{j=1}^{n} E\left\|\Phi(j,n) \frac{r_j e_j}{\|e_j\|^2}\right\|$$

and from Jensen and Hölder inequalities, we obtain

$$\|E(v_n)\| \leqslant E^{1/2}\|\Phi(0,n)\|^2 E^{1/2}\|v_0\|^2 + \mu \sum_{j=1}^{n} E^{1/2}\|\Phi(j,n)\|^2 E^{1/2}\left(\frac{r_j^2}{\|e_j\|^2}\right). \tag{11}$$

[1] C, C_1, and C_2 will denote bounded quantities; the repeated use of any of these symbols does not imply equality.