A MULTIVARIABLE STEIGLITZ-McBRIDE METHOD

Mehdi Ashari[1], *Mamadou Mboup*[2] *and Phillip A. Regalia*[3]

[1]Laboratoire des Signaux et Systèmes, CNRS-ESE-UPS, 91192 Gif-sur-Yvette, France
[2]Univ. René Descartes-Paris V, UFR Math-Info, 45 rue des Saints Pères, 75270 Paris cedex 06.
[3]Inst. National des Télécomm., Départ. Signal & Image, 9 rue Charles Fourier, 91011 Evry cedex, France.

ABSTRACT

In this paper, we present an off-line multi-input/multi-output version of the Steiglitz-McBride method, as well as an analytic description of the set of its stationary points. As in the scalar case [13], the description is given in terms of first- and second-order interpolation constraints, respectively, on the model impulse response and covariance sequences. The constraints are related to the theory of q-Markov covariance equivalent realizations and generalize the work of Inouye [7] and King *et al.* [9].

1 INTRODUCTION

The Steiglitz-McBride Method [17] was introduced in 1965, for linear system identification purposes. The first formal study of the method appeared in 1981 when Stoica and Söderström [18] showed that the method can claim an exact matching property in sufficient order cases, like many identification schemes, provided the measurement noise is white. An on-line version of the method followed five years later with Fan and Jenkins [3], as an adaptive infinite impulse response (IIR) filtering algorithm. A formal analysis of this on-line variant followed in Fan [4], who showed that, as with the off-line counterpart, the on-line version is capable of correct identification in sufficient order cases if the disturbance term is white. Since then, many simulation studies [3], [5] have shown that the algorithm is a well behaved candidate among the numerous methods for adaptive IIR filtering.

An important feature of the method is that it does not rely on a gradient-based search of the mean square error (MSE) surface. As such, it is free from local minima traps. Paradoxally, this advantage also raises a serious difficulty in the sense that no meaningful criteria can be, so far, attached to the method in reduced order cases. More seriously, the question of whether a convergent point even exists in all cases has never been satisfactorily answered. Indeed, for such non-gradient algorithms, inferring existence of a stationary point, much less convergence towards one, is in general a difficult problem, owing to the nonlinear character of the involved equations.

Several important results have nonetheless been established recently for the Steiglitz-McBride algorithm in the reduced order scenario, assuming a white noise input:

- An analytic description of the set of stationary points of the method has been obtained [12], [13];
- This description has allowed us to prove the following [14]: *If M is the order of the model, then the L_2 norm of the error function at any stationary point can be no larger than the $M+1^{st}$ Hankel singular value of the unknown system.* This *a priori* error bound fully explains Fan and Doroslovački's conjecture [5] on the closeness of the Steiglitz-McBride model to the global minimum of the error surface, as has been observed in many simulation studies [13], [5];
- The method can be viewed [17] as a sequence of weighted equation error problems which, when convergent, reduce to a classical output error problem. Now, the equation error method may be interpreted, to some extent, as a first- and second-order interpolation problem in which the interpolation points cluster at the origin. This formulation was introduced by Mullis and Roberts in the scalar case [11], and has been extended to the multivariable case by Inouye [7] and King *et al.* [9]. Following the same formulation, the analytic description of the set of possible stationary points of the Steiglitz-McBride method (SMM) has been shown [15] to fall into the framework of Nevanlinna-Pick interpolation theory. The well-developed machinery of this theory afforded [15] a complete parametrization of the transfer functions covering this set;
- Based on this Nevanlinna-Pick problem, a sufficient condition for the existence of a stationary point was derived [15], and recently strengthened [16] to show that a stationary point exists whenever the

Mehdi Ashari is supported by a grant from the CNET/GRETSI of France, contract no. 936B005.

(larger order) unknown system satisfies a mild stability constraint.

In this contribution, we present an off-line version of the Steiglitz-McBride method in the multivariable case, and show that a given rational matrix-valued function corresponds to a stationary point for the multivariable SMM if and only if it is a solution of a first- and second-order interpolation problem. Our motivation originates from the interesting properties of the Steiglitz-McBride method in the scalar case, as listed above, combined with the interest in multivariable identification from such signal processing applications as stereo echo cancellation.
The paper is organized as follow: in Section 2 we review some basic definitions related to $L_2^{p\times p}$ functions. In Section 3 we present the philosophy of multivariable Steiglitz-McBride method. Section 4 then presents our main result on the stationary points. We show how a stationary point can be related to the theory of q-Markov covariance equivalent realizations in Section 5. Finally, the last Section concerns some concluding remarks.

2 PRELIMINARIES

Consider the set $L_2^{p\times p}$ of all $\mathbb{C}^{p\times p}$-matrix valued functions $F(z)$, square integrable on the unit circle $|z|=1$. They are representable in the form of a convergent power series expansion:

$$F(z)=\sum_{i=-\infty}^{\infty} F_i\, z^i, \quad \text{with} \quad \sum_{i=-\infty}^{\infty} \|F_i\|^2<\infty \qquad (1)$$

where $\|F_i\|^2=\mathrm{tr}(F_iF_i^\dagger)$ is the Frobenius norm of $F_i\in\mathbb{C}^{p\times p}$ and where the superscript $\dagger$ and the symbol tr denote, respectively, the transpose-conjugation and the trace of a matrix.
We interpret z as the unit delay operator. The set of all stable and causal $\mathbb{C}^{p\times p}$-matrix valued transfer functions is denoted by $\mathcal{H}_2^{p\times p}$. For any $F(z)$ in $L_2^{p\times p}$ define its *anti-causal projection* as

$$[F(z)]_-=\sum_{i=-\infty}^{0} F_i\, z^i=\cdots+F_{-2}z^{-2}+F_{-1}z^{-1}+F_0, \quad (2)$$

which converges to an analytic matrix function in $|z|>1$, and its *strictly causal projection* as

$$[F(z)]_+=\sum_{i=1}^{\infty} F_i\, z^i=F_1z+F_2z^2+F_3z^3+\cdots \qquad (3)$$

which converges to an analytic matrix function in $|z|<1$, vanishing at the origin. These projections are related to the decomposition of $L_2^{p\times p}$ into orthogonal complement Hardy subspaces: $L_2^{p\times p}=\mathcal{H}_2^{p\times p}\oplus[\mathcal{H}_2^{p\times p}]^\perp$, of $L_2^{p\times p}$ functions analytic in the open unit disc $\mathbb{D}$ and outside the closed unit disc, respectively [6].

To each pair of functions $F(z)$ and $G(z)$ in $L_2^{p\times p}$ we associate two (matrix-valued) inner products

$$\langle F(z),G(z)\rangle_\Omega \triangleq \int_0^{2\pi} F(e^{j\theta})\, d\Omega\, G^\dagger(e^{-j\theta}), \qquad (4a)$$

$$\langle F(z),G(z)\rangle_\Omega^\sim \triangleq \int_0^{2\pi} F^\dagger(e^{-j\theta})\, d\Omega\, G(e^{j\theta}), \qquad (4b)$$

where Ω is a positive (Hermitian) matrix-valued measure.
The norm of a function $F(z)$ in $L_2^{p\times p}$, in the metric Ω, can be computed by either inner product (4a) or (4b), i.e.,

$$\begin{aligned}\|F(z)\|_\Omega^2 &= \mathrm{tr}\,\langle F(z),F(z)\rangle_\Omega\,, &(5a)\\ &= \mathrm{tr}\,\langle F(z),F(z)\rangle_\Omega^\sim &(5b)\\ &= \|F(z)\|_\Omega^{\sim 2}. &(5c)\end{aligned}$$

In the interest of clarity, we shall always retain the "$\sim$" sign when the norm is computed via (5b).

3 ALGORITHM DESCRIPTION

Let $H(z)\in\mathcal{H}_2^{p\times p}$ denote the $\mathbb{C}^{p\times p}$ matrix-valued transfer function of the p-input/p-output system to be identified:

$$\boldsymbol{y}(n)=H(z)\cdot\boldsymbol{x}(n)+\boldsymbol{\zeta}(n). \qquad (6)$$

Here $\boldsymbol{x}(\cdot),\boldsymbol{y}(\cdot)\in\mathbb{C}^p$ are, respectively, the input and the output of the system, and $\boldsymbol{\zeta}(\cdot)$ is the measurement noise. To simplify some notations to follow, we henceforth assume that the input $\boldsymbol{x}(\cdot)$ is a normalized white noise process. We also suppose that the measurement noise $\boldsymbol{\zeta}(\cdot)$ is a white noise process. Denote by $G(z)\in\mathcal{H}_2^{p\times p}$ the transfer matrix of an adjustable model, given in a right Matrix Fraction Description (MFD) form [8]:

$$G(z) \;=\; B(z)A^{-1}(z) \qquad (7)$$

where both $A(z)$ and $B(z)$ have formal degree M, and $A(z)$ is a minimum phase polynomial, i.e., $\det A(z)\neq 0$ for all $|z|\leq 1$.
As in the scalar case, the Multivariable Steiglitz-McBride (MSM) algorithm can be described (see figure 1) as follows:

1. The input $\boldsymbol{x}(\cdot)$ and output $\boldsymbol{y}(\cdot)$ sequences are prefiltered using an initial estimate, $A_k^{-1}(z)$, of the *denominator* of the right MFD of the model;

2. Find the polynomials $A_{k+1}(z)$ and $B_{k+1}(z)$ to minimize the variance of the vector equation error signal

$$\boldsymbol{e}(n)=A_{k+1}(z)\boldsymbol{y}_f(n)-B_{k+1}(z)\boldsymbol{x}_f(n), \qquad (8)$$

where the prefiltered signals are given by

$$\left\{\begin{aligned}\boldsymbol{y}_f(n) &= A_k^{-1}(z)\,\boldsymbol{y}(n)\\ &= A_k^{-1}(z)\,H(z)\boldsymbol{x}(n)+A_k^{-1}(z)\,\boldsymbol{\zeta}(n),\\ \boldsymbol{x}_f(n) &= A_k^{-1}(z)\,\boldsymbol{x}(n);\end{aligned}\right. \qquad (9)$$

3. Update the prefilter by $A_{k+1}^{-1}(z)$ for the next experiment and iterate.

The MSM method may thus be viewed as the sequence of minimization problems:

$$\min_{A_{k+1},B_{k+1}} \{\|A_{k+1}(z)A_k^{-1}(z)H(z) - B_{k+1}(z)A_k^{-1}(z)\|^2 \\ +\sigma^2 \|A_{k+1}(z)A_k^{-1}(z)\|^2\}, \quad (10)$$

where σ^2 is the energy of the measurement noise. Of course, to avoid the trivial solution $A_{k+1}(z) = B_{k+1}(z) = \bigcirc$, we must put some constraint on the minimizing arguments, which will be described in the next section.

One can show [10] that this algorithm generates a sequence of rational function $B_k(z)A_k^{-1}(z)$ in $\mathcal{H}_2^{p\times p}$, and convergence of this algorithm—if it occurs—reduces the previous cost function (10) to

$$\|H(z) - B_{k+1}(z)A_{k+1}^{-1}(z)\|^2 + p\,\sigma^2, \quad (11)$$

which is a classical multivariable output error problem in presence of an additive white noise term. We should emphasize that this particular $G(z)$ is not claimed to be optimal in $L_2^{p\times p}$ norm.

The minimization problem (10) can be rearranged as

$$\min_{A_{k+1},B_{k+1}} \{\|A_{k+1}(z)A_k^{-1}(z)H(z)A_k(z) - B_{k+1}(z)\|^2_{\Omega_k} \\ +\sigma^2 \|A_{k+1}(z)\|^2_{\Omega_k}\}, \quad (12)$$

with $d\Omega_k(e^{j\theta}) = A_k^{-1}(e^{j\theta})\,A_k^{-\dagger}(e^{-j\theta})\,d\theta$.

Here, $d\Omega_k/d\theta$ is an spectral density matrix that can be described in a dual form [2]:

$$d\Omega_k(e^{j\theta}) = C_k^{-\dagger}(e^{-j\theta})\,C_k^{-1}(e^{j\theta})\,d\theta, \quad (13)$$

where $C_k(z)$, like $A_k(z)$, is a minimum phase polynomial of formal degree M. This decription of $d\Omega_k$ leads us to a left MFD of $G(z)$ [8]:

$$G(z) = C^{-1}(z)\,D(z). \quad (14)$$

By using this left MFD for $G(z)$ we can set up a dual form of the MSM method:

1. Find the polynomials $C_{k+1}(z)$ and $D_{k+1}(z)$ to minimize the variance of the vector equation error signal

   $$\boldsymbol{e}(n) = C_k^{-1}(z)C_{k+1}\boldsymbol{y}(n) - C_k^{-1}(z)D_{k+1}(z)\boldsymbol{x}(n), \quad (15)$$

   where $C_k^{-1}(z)$, an initial estimate of the *denominator* of the left MFD of the model, is used as a postfilter.

2. Update the postfilter by $C_{k+1}^{-1}(z)$ for the next experiment and iterate.

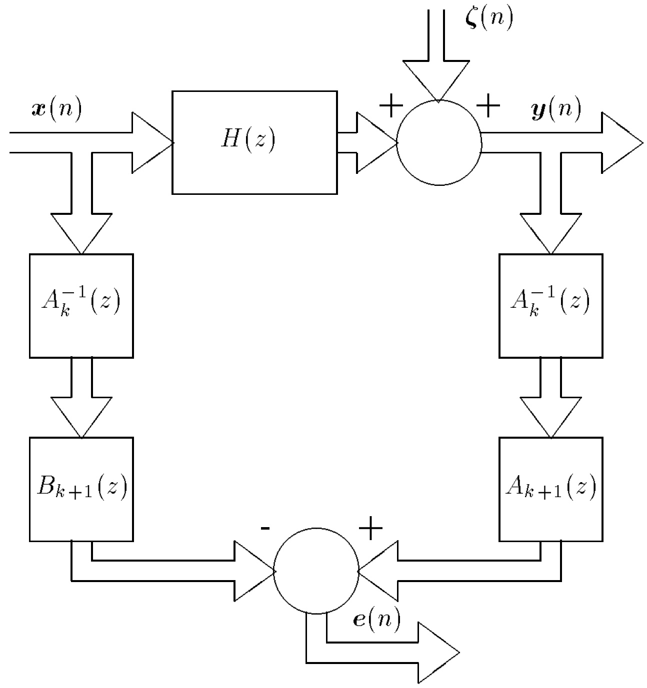

Figure 1: Steiglitz-McBride Algorithm using a right MFD for the model

Note that $C_k(z)$ obtained in the k^{th} iteration will be used as a postfilter, rather than a prefilter as it was in the previous algorithm.

When a left MFD is used for $G(z)$, the problem in equation (12) is replaced by the dual form:

$$\min_{C_{k+1},D_{k+1}} \left\{\|C_{k+1}(z)H(z) - D_{k+1}(z)\|^2_{\tilde{\Omega}_k} \\ +\sigma^2 \|C_{k+1}(z)\|^2_{\tilde{\Omega}_k}\right\}. \quad (16)$$

4 STATIONARY POINTS

In this section we give an analytic description of stationary point(s) of the off-line MSM method in term of first- and a second-order constraints.

To obtain this description, we suppose that $A_k(z)$ is monic, i.e., $A_k(0) = I$, and as a consequence, the minimization (10) is subject to $A_{k+1}(0) = I$. Our main result is:

Theorem 4.1. *Suppose the input $\{\boldsymbol{x}(\cdot)\}$ is white noise, and let $G(z) = B(z)A^{-1}(z) = C^{-1}(z)D(z)$ be an adjustable model, where all the involved polynomials are of formal degree M. Then $G(z)$ is a stationary point of the multivariable Steiglitz-McBride method if and only if*

$$H(z) - G(z) = P(z)V(z) \quad (17a)$$

$$\left[H(z)H^{\dagger}(z^{-1})\right]_+ - \left[G(z)G^{\dagger}(z^{-1})\right]_+ = Q(z)V(z), \quad (17b)$$

where $V(z)$, defined by $V(z) = z^M C^{\dagger}(z^{-1})A^{-1}(z)$, is paraunitary, i.e., $V(z)V^{\dagger}(z^{-1}) = I$ for all z, and $P(z)$ and $Q(z)$ are two strictly causal $\mathcal{H}_2^{p\times p}$ functions.

Proof. See [10] for a proof. □

5 INTERPOLATION CONSTRAINTS

We show in this section that the contraints (17) are related to a generalized q-Markov covariance equivalent realization problem [9].

If in theorem 4.1 we set $V(z) = z^M I$ (equation error case [7]), which would mean that all the zeros of $V(z)$ cluster at 0, then any solution of (17) would fulfill:

$$G_i = H_i, \quad \text{for } i = 0, \cdots, M \tag{18a}$$

$$\widehat{R}_i = R_i, \quad \text{for } i = 1, \cdots, M \tag{18b}$$

with

$$\begin{cases} H(z) = \sum_{i=0}^{\infty} H_i z^i, \\ G(z) = \sum_{i=0}^{\infty} G_i z^i, \\ R(z) = H(z)H^\dagger(z^{-1}) = \displaystyle\sum_{i=-\infty}^{\infty} R_i z^i, \\ \widehat{R}(z) = G(z)G^\dagger(z^{-1}) = \displaystyle\sum_{i=-\infty}^{\infty} \widehat{R}_i z^i. \end{cases} \tag{19}$$

Thus in this special case, any solution $G(z)$ in theorem 4.1, appears as a solution of a q-Markov covariance equivalent realization problem, which can also be interpreted as a matricial Nevanlinna-Pick type interpolation problem [1, 15].

This interpretation still holds for the stationary points of the MSM algorithm. In this case, the constraints in theorem 4.1 can be put (see [10]) in the context of tangential Nevanlinna-Pick interpolation theory [1].

6 CONCLUDING REMARKS

In this paper, we have given two *off-line* version of the Multivariable Steiglitz-McBride method based respectively on the right and left Matrix Fraction Description of the model. We hope that this could serve as a starting point for the development of an *on-line* version of this algorithm.

References

[1] D. Alpay, V. Bolotnikov and P. Loubaton, "On tangential $\mathcal{H}_2$ interpolation with second order norm constraints," *Integral Equations and Operator Theory*, (to appear).

[2] P. Dewilde, A. C. Vieira and T. Kailath, "On a Generalized Szegö-Levinson Realisation Algorithm for Optimal Linear Predictors Based on a Network Synthesis Approach," *IEEE Trans. Circuits and Systems*, vol. CAS-25, No. 9, pp. 663-675, September 1978.

[3] H. Fan, W.K. Jenkins, " A new adaptive IIR filter," *IEEE Trans. Circ. and Syst.*, vol.33, pp. 939-947, 1986.

[4] H. Fan, "Application of Benveniste's convergence results in the study of adaptive IIR filtering algorithms," *IEEE Trans. Information Theory*, vol. 34, pp. 679-709, July 1988.

[5] H. Fan, M. Doroslovački, "On 'Global Convergence' of Steiglitz-McBride Adaptive Algorithm", *IEEE Trans. Circ. and Syst.*, II, vol.40, No. 2, pp. 73-87, 1993.

[6] K. Hoffman, *Banach Spaces of Analytic Functions*, Prentice-Hall, Englewood Cliffs, NJ, 1962.

[7] Y. Inouye, "Approximation of multivariable linear systems with impulse response and autocorrelation sequences," *Automatica*, vol. 19, no. 3, pp. 265-277, 1983.

[8] T. Kailath, *Linear Systems*, Prentice-Hall, Englewood Cliffs, NJ, 1980.

[9] A. M. King, U. B. Desai and R. E. Skelton, "A generalized approach to q-Markov covariance equivalent realizations for discrete systems," *Automatica*, vol. 24, no. 4, pp. 507-515, 1988.

[10] M. Mboup, M. Ashari, P. A. Regalia, "Analytical Description of the Stationary Points for the Multivariable Steiglitz-McBride Method," *in preparation.*

[11] C. T. Mullis and R. A. Roberts, "The use of second-order information in the approximation of discrete-time linear systems," *IEEE Trans. Acoustics, Speech, and Signal Processing*, vol. 24, pp. 226-238, 1976.

[12] P.A. Regalia, M. Mboup, "On the approximation character of some recursive identification schemes," *I.M.A. conf. on Math. in Sign. Proc.*, Warwick, U.K., Dec. 1992.

[13] P. A. Regalia, *Adaptive IIR Filtering in Signal Processing and Control*, Marcel Dekker, New York, 1995.

[14] P.A. Regalia and M. Mboup, "Undermodelled adaptive filtering: An a priori error bound for Steiglitz-McBride algorithm," t *IEEE Trans. Circuits and Systems–II*,vol. 43, no. 2, pp. 105-116, February 1996.

[15] P. A. Regalia, M. Mboup, and M. Ashari, "A class of first- and second-order interpolation problems in model reduction," *Archiv für Elektronik und Übertragungstechnik*, vol. 43, pp. 343-344, Sept. 1995.

[16] P. A. Regalia, M. Mboup, and M. Ashari, "On the existence of stationary points for the Steiglitz-McBride algorithm," submitted to *IEEE Trans. Aut. Contr.*, June 1995.

[17] K. Steiglitz, L.E. McBride, "A technique for the identification of linear systems," *IEEE Trans. Aut. Contr.*, vol. 10, pp. 461-464, 1965.

[18] P. Stoica, T. Söderström, "The Steiglitz–McBride algorithm revisited–convergence analysis and accuracy aspects," *IEEE Trans. Aut. Contr.*, vol. 26, pp. 712-717, 1981.