Hyperspherical Parametrization for Unit-Norm Based Adaptive IIR Filtering

Sundar G. Sankaran, *Student Member, IEEE,* and A. A. Louis Beex, *Senior Member, IEEE*

***Abstract*— The bias problem associated with equation error based adaptive infinite impulse response (IIR) filtering can be surmounted by imposing a unit-norm constraint on the autoregressive (AR) coefficients. We propose a hyperspherical parameterization to convert the unit-norm-constrained optimization into an unconstrained optimization. We show that the hyperspherical parameterization does not introduce any additional minima to the equation error surface.**

***Index Terms*—Adaptive IIR filtering, equation error minimization.**

I. Introduction

TRADITIONALLY, finite impulse response (FIR) structures have been used for adaptive filters, due to their simplicity. However, it could be advantageous to use infinite impulse response (IIR) structures rather than FIR structures for adaptive filters, especially when the desired filter can be modeled with fewer parameters using both poles and zeros. The potential for reduction in computational complexity and improvement in performance has motivated research in adaptive IIR filtering. Equation-error minimization [1] is a preferred form of adaptive IIR filtering, since it leads to a well-behaved convex minimization problem. However, the minimum of the equation-error surface gives a biased estimate for the system parameters.

On the other hand, the minimum of the equation-error surface subject to the constraint that the autoregressive (AR) coefficient vector has unit norm provides an unbiased estimate of the true parameters [3]. The existing adaptive (online) techniques attempt to solve this constrained optimization problem using a Lagrange multiplier method [6] or a generalized Rayleigh quotient [7]. We propose a hyperspherical parameterization that converts this unit-norm-constrained optimization into an unconstrained optimization. Section II presents the proposed hyperspherical parameterization. In Section III, it is shown that this parameterization leads to a nonconvex unconstrained optimization problem that can be solved using gradient based optimization techniques. Section IV provides the conclusion.

Manuscript received July 2, 1999. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. G. Ramponi.

The authors are with the Bradley Department of Electrical and Computer Engineering, Systems Group, DSP Research Laboratory, Virginia Polytechnic Institute and State University, Blacksburg, VA 24061-0111 USA (e-mail: beex@vt.edu).

Publisher Item Identifier S 1070-9908(99)09289-5.

II. Hyper-Spherical Parameterization

All unit-norm based adaptive IIR filtering algorithms in the existing literature [2], [6], [7] use the ARMA coefficient parameterization. These algorithms attempt to estimate the AR and MA coefficients that minimize the equation-error subject to the constraint that the AR coefficient vector has unit norm. It has been proved that this constrained equation-error surface has a unique minimum and that the minimum of the error surface (under the unit-norm constraint) gives the true parameters of the system if the model is of sufficient order and the input sufficiently rich [3].

This constrained optimization problem can be converted into an unconstrained optimization by using a hyperspherical parameterization. That is, rather than directly adapting the MA coefficients $\mathbf{b} = (b_0, b_1, \cdots, b_M)^T$, and the AR coefficients $\mathbf{a} = (a_0, a_1, \cdots, a_M)^T$, where $(\cdot)^T$ is the vector transpose operator, we propose adapting $\overline{\mathbf{b}} = (\overline{b}_0, \overline{b}_1, \cdots, \overline{b}_M)^T$, r, and $\boldsymbol{\theta} = (\theta_1, \theta_2, \cdots, \theta_M)^T$, where

$$\begin{aligned} b_j &= r\overline{b}_j, \qquad \forall j \in \{0, 1, \cdots, M\} \\ a_0 &= r \cos \theta_1, \\ a_k &= r\left(\prod_{i=1}^{k} \sin \theta_i\right) \cos \theta_{k+1}, \quad \forall k \in \{1, 2, \cdots, M-1\} \\ a_M &= r \prod_{i=1}^{M} \sin \theta_i. \end{aligned} \tag{1}$$

It can be shown that $||\mathbf{a}|| = r$ and that $\mathbf{a} = \mathbf{a}(r, \boldsymbol{\theta}) = r\mathbf{a}(1, \boldsymbol{\theta})$. To insure a one-to-one correspondence between $\boldsymbol{\alpha} = [\mathbf{b}^T \quad \mathbf{a}^T]^T$ and $\boldsymbol{\beta} = [\overline{\mathbf{b}}^T \;\, r \;\, \boldsymbol{\theta}^T]^T$, we restrict the values of θ_i such that $\theta_i \in [0, 2\pi) \forall i \in \{1, 2, \cdots, M\}$ and r such that $r > 0$.

III. (Non)-Unimodality Issues

Gradient-based algorithms are the most widely used adaptation algorithms, due to their simplicity and well-understood behavior. For small enough step-sizes, these algorithms are guaranteed to converge to a point where the gradient is zero. Hence, it is desirable to have the gradient being zero at a point as a sufficient condition for that point to be a global optimum of the function being optimized. In this section, we analyze whether the equation-error surface, after hyperspherical transformation, satisfies the global optimality property.

Lemma: If $a_M \neq 0$, the Jacobian $\mathbf{J}_M$ of the hyperspherical transformation is nonsingular.

Proof: In the Appendix, it is shown that the determinant of $\mathbf{J}_M$ is given by

$$\det(\mathbf{J}_M) = r^{2M+1}\sin^{M-1}(\theta_1)\sin^{M-2}(\theta_2)\cdots\sin(\theta_{M-1}). \tag{2}$$

Then $a_M = r\sin(\theta_1)\sin(\theta_2)\cdots\sin(\theta_M) \neq 0$ implies none of $\sin(\theta_i)$ or r in (2) equals zero. Hence $\mathbf{J}_M$ is nonsingular.

Minimizing the equation error (in terms of direct coefficient parameterization) under the unit-norm constraint is equivalent to minimizing the generalized Rayleigh quotient [7] given by

$$E_1(\boldsymbol{\alpha}) \triangleq \frac{[\mathbf{b}^T \quad \mathbf{a}^T]\mathbf{R}\begin{bmatrix}\mathbf{b}\\ \mathbf{a}\end{bmatrix}}{\mathbf{a}^T\mathbf{a}} \tag{3}$$

where $\mathbf{R}$ is a positive-definite autocorrelation matrix. Since there is a one-to-one correspondence between $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$, (3) can be rewritten as follows:

$$\begin{aligned} E_1(\boldsymbol{\alpha}) &= \frac{[r\overline{\mathbf{b}}^T \quad r\mathbf{a}(1,\boldsymbol{\theta})^T]\mathbf{R}\begin{bmatrix} r\overline{\mathbf{b}}\\ r\mathbf{a}(1,\boldsymbol{\theta})\end{bmatrix}}{r^2} \\ &= [\overline{\mathbf{b}}^T \quad \mathbf{a}(1,\boldsymbol{\theta})^T]\mathbf{R}\begin{bmatrix}\overline{\mathbf{b}}\\ \mathbf{a}(1,\boldsymbol{\theta})\end{bmatrix} \\ &\triangleq E_2(\boldsymbol{\beta}). \end{aligned} \tag{4}$$

Thus, $E_2(\boldsymbol{\beta})$ is the unit-norm-constrained equation-error cost function after the hyperspherical transformation.

Theorem: Every stationary point of $E_1(\boldsymbol{\alpha})$ is a stationary point of $E_2(\boldsymbol{\beta})$. Furthermore, any newly formed stationary point of $E_2(\boldsymbol{\beta})$ is a saddle point and the AR polynomial corresponding to these saddle points has degree less than M.

Proof: The gradients of $E_1(\boldsymbol{\alpha})$ and $E_2(\boldsymbol{\beta})$ are related as follows:

$$\nabla_\beta E_2(\boldsymbol{\beta}) = \mathbf{J}_M^T \nabla_\alpha E_1(\boldsymbol{\alpha}) \tag{5}$$

where $\boldsymbol{\alpha} = [\mathbf{b}^T \quad \mathbf{a}^T(r,\boldsymbol{\theta})]^T$. From (5), it follows that if $\nabla_\alpha E_1(\boldsymbol{\alpha}^*) = \mathbf{0}$, then the corresponding point in hyperspherical coordinates $\boldsymbol{\beta}^* = [\overline{\mathbf{b}}^{*T} \; r^* \; \boldsymbol{\theta}^{*T}]^T$, where $\mathbf{b}^* = r^*\overline{\mathbf{b}}^*$ and $\mathbf{a}^* = \mathbf{a}(r^*, \boldsymbol{\theta}^*)$, satisfies $\nabla_\beta E_2(\boldsymbol{\beta}^*) = \mathbf{0}$. This proves the first part of the theorem. Furthermore, since the Hessians of the error surfaces $E_1(\boldsymbol{\alpha})$ and $E_2(\boldsymbol{\beta})$ are related as follows:

$$\nabla_\beta^2 E_2(\boldsymbol{\beta}) = \mathbf{J}_M^T \nabla_\alpha^2 E_1(\boldsymbol{\alpha})\mathbf{J}_M \tag{6}$$

the nature of the stationary points of $E_1(\boldsymbol{\alpha})$ is preserved by the hyperspherical transformation. That is, minima get mapped to minima and saddle points get mapped to saddle points.

If there exists a point $\boldsymbol{\beta}^*$ such that $\nabla_\beta E_2(\boldsymbol{\beta}^*) = \mathbf{0}$ while $\nabla_\alpha E_1(\boldsymbol{\alpha}^*) \neq \mathbf{0}$, where $\boldsymbol{\alpha}^* = [r^*\overline{\mathbf{b}}^{*T} \; \mathbf{a}^T(r^*, \boldsymbol{\theta}^*)]^T$, it follows from (5) that $\mathbf{J}_M$ must be singular. From the Lemma above, $\mathbf{J}_M$ being singular implies $a_M = 0$, which in turn shows that the AR polynomial corresponding to any newly formed stationary point of $E_2(\boldsymbol{\beta})$ has degree less than M. Nayeri and Jenkins have proved that the newly formed stationary points due to any continuous transformation are saddle points [4]. $\square$

From the above theorem, it is clear that the unimodal constrained equation-error surface [7] in ARMA parameterizations, upon hyperspherical transformation, may have stationary points in addition to the global minimum of the error surface. Hence, the gradient being zero at a point is not a sufficient condition for that point to be a global minimum. However, the newly formed stationary points are saddle points. If parameter trajectories go through or near saddle points this can potentially slow down the convergence of the algorithm. However, the gradient algorithm will still converge, albeit perhaps more slowly, to the global minimum provided there is some noise present to perturb the adaptation algorithm away from saddle points and provided the saddle points are not dense. It is worth adding that there is always some noise due to measurement and quantization errors.

We can attempt to minimize the cost function $E_2(\boldsymbol{\beta})$ of (4), in terms of hyperspherical coordinates, by using the LMS algorithm. The LMS algorithm is an approximation of the steepest descent algorithm, where the true gradients are replaced by their (noisy) instantaneous values. Also, since $E_2(\boldsymbol{\beta})$ is independent of r (equivalently $\partial E_2/\partial r = 0$), r can be set to a nonzero positive constant (preferably unity) and need not be adapted. A detailed algorithm based on the instantaneous (stochastic) gradient of (4) is presented elsewhere [5]. Over many simulations, no convergence problems were observed that could be attributed to saddle points. There is merely a temporary slow down of convergence near the saddle point. This suggests that the effects of saddle points (while they exist for some systems) are mitigated by the use of the instantaneous gradient.

IV. Conclusion

A parameterization that converts the unit-norm constrained adaptive filtering problem into an unconstrained adaptive filtering problem is presented. The parameter transformation may introduce additional stationary points in the error surface. However, additional stationary points in the transformed error surface, if any, are saddle points, so that the unique global minimum is preserved.

Appendix
Proof for (2)

The Jacobian of the hyperspherical transformation can be written as shown in (A1)–(A4), shown at the top of the next page, where c_m and s_n denote $\cos\theta_m$ and $\sin\theta_n$, respectively. Using the partitioned-matrix determinant identity [8], $\det(\mathbf{J}_M) = \det(\mathbf{J}_M^{11})\det(\mathbf{J}_M^{22})$. From (A2) follows that $\det(\mathbf{J}_M^{11}) = r^{M+1}$. We claim (proved below) that

$$\det(\mathbf{J}_M^{22}) = r^M \sin^{M-1}(\theta_1)\sin^{M-2}(\theta_2)\cdots\sin(\theta_{M-1}). \tag{A5}$$

Hence

$$\det(\mathbf{J}_M) = r^{2M+1}\sin^{M-1}(\theta_1)\sin^{M-2}(\theta_2)\cdots\sin(\theta_{M-1}). \tag{A6}$$

Inductive Proof for (A5): For $M = 1$, $\det(\mathbf{J}_M^{22}) = r$, which satisfies (A5). Assume that (A5) is true for $\mathbf{J}_{M-1}^{22}$. Using (A4)

$$\begin{aligned} \det(\mathbf{J}_M^{22}) &= rs_1\cdots s_{M-1}c_M^2 \det(\mathbf{J}_{M-1}^{22}) \\ &\quad + rs_1\cdots s_{M-1}s_M^2 \det(\mathbf{J}_{M-1}^{22}) \\ &= rs_1\cdots s_{M-1}\det(\mathbf{J}_{M-1}^{22}). \end{aligned} \tag{A7}$$

$$\mathbf{J}_M = \begin{pmatrix} \mathbf{J}_M^{11} & \mathbf{J}_M^{12} \\ \mathbf{0} & \mathbf{J}_M^{22} \end{pmatrix} \tag{A1}$$

$$\mathbf{J}_M^{11} = r\mathbf{I}_{(M+1)\times(M+1)} \tag{A2}$$

$$\mathbf{J}_M^{12} = \begin{bmatrix} \overline{\mathbf{b}} & \mathbf{0}_{(M+1)\times M} \end{bmatrix} \tag{A3}$$

$$\mathbf{J}_M^{22} = \begin{pmatrix} c_1 & -rs_1 & 0 & \cdots & 0 \\ s_1 c_2 & rc_1 c_2 & -rs_1 s_2 & \ddots & \vdots \\ \vdots & \vdots & \vdots & & \\ & & & \ddots & 0 \\ s_1 \cdots s_{M-1} c_M & rc_1 s_2 \cdots s_{M-1} c_M & rs_1 c_2 s_3 \cdots s_{M-1} c_M & \cdots & -rs_1 \cdots s_M \\ s_1 \cdots s_M & rc_1 s_2 \cdots s_M & rs_1 c_2 s_3 \cdots s_M & \cdots & rs_1 \cdots s_{M-1} c_M \end{pmatrix}_{(M+1)\times(M+1)} \tag{A4}$$

Hence, $\det(\mathbf{J}_M^{22}) = r^M \sin^{M-1}(\theta_1)\sin^{M-2}(\theta_2)\cdots\sin(\theta_{M-1})$ and, by induction, (A5) holds for all positive integer values of M.

REFERENCES

[1] L. Ljung, *System Identification: Theory for the User*. Englewood Cliffs, NJ: Prentice-Hall, 1987.

[2] S. C. Douglas and M. Rupp, "On bias removal and unit norm constraints in equation error adaptive filters," in *Proc. 30th Asilomar Conf. Signals, Systems, and Computers,* Pacific Grove, CA, Nov. 1996, pp. 1093–1097.

[3] P. A. Regalia, "An unbiased equation error identifier and reduced order approximations," *IEEE Trans. Signal Processing,* vol. 42, pp. 1397–1412, June 1994.

[4] M. Nayeri and W. K. Jenkins, "Analysis of alternate realizations of adaptive IIR filters," in *Proc. IEEE Int. Symp. Circuits and Systems,* Espoo, Finland, June 1988, pp. 2157–2160.

[5] S. G. Sankaran and A. A. Beex, "An unbiased equation-error-based adaptive IIR filtering algorithm," in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing,* Seattle, WA, May 1998, pp. 1425–1428.

[6] K. C. Ho and Y. T. Chan, "Bias removal in equation error adaptive IIR filters," *IEEE Trans. Signal Processing,* vol. 43, pp. 51–62, Jan. 1995.

[7] C. E. Davila, "An algorithm for efficient, unbiased, equation-error infinite impulse response adaptive filtering," *IEEE Trans. Signal Processing,* vol. 42, pp. 1221–1226, May 1994.

[8] S. M. Kay, *Modern Spectral Estimation: Theory and Application.* Englewood Cliffs, NJ: Prentice-Hall, 1988.