On the Composite Squared Error Algorithm for Adaptive IIR Filters

Sergio L. Netto and Panajotis Agathoklis

Depto. de Eletrônica – EE
Universidade Federal do Rio de Janeiro
C.P. 68564 – Rio de Janeiro – RJ
21945 – 970 – BRAZIL
sergioln@coe.ufrj.br

Dept. Elec. Comp. Eng.
University of Victoria
P.O. Box 3055 – Victoria – BC
V8W3P6 – CANADA
pan@sirius.uvic.ca

Abstract

The composite squared error (CSE) adaptation algorithm for IIR filters is derived as a combination of the equation error (EE) and the output error (OE) algorithms. In this paper, the convergence process of the CSE adaptation algorithm is investigated. Steady-state analysis is given relating the CSE stationary points to the EE and OE stationary points. Transient analyses of the CSE algorithm are obtained with the ordinary-difference-equation approach and the local linearization method.

1 Introduction

The block diagram of a general adaptive system is seen in Figure 1, where $x(n)$ is the input signal, $\hat{y}(n)$ is the adaptive-filter output signal, $y(n)$ is the desired output signal, and $e_{OE}(n)$ is an error signal.

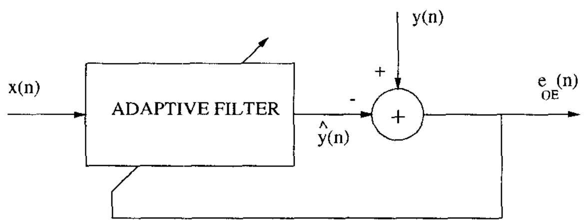

Figure 1: Block diagram of a basic adaptive system.

The adaptive filter is commonly described by

$$\hat{y}(n) = \left[\frac{\hat{B}(q,n)}{\hat{A}(q,n)}\right]\{x(n)\} = \frac{\sum_{j=0}^{n_{\hat{b}}} \hat{b}_j(n)q^{-j}}{1+\sum_{i=1}^{n_{\hat{a}}} \hat{a}_i(n)q^{-i}}$$

A general adaptation algorithm has the form

$$\hat{\boldsymbol{\theta}}(n+1) = \hat{\boldsymbol{\theta}}(n) + \mu(n)e(n)\hat{\boldsymbol{\phi}}(n)$$

where $\hat{\boldsymbol{\theta}}(n) = [\hat{a}_1(n)\ldots\hat{a}_{n_{\hat{a}}}(n)\ \hat{b}_0(n)\ldots\hat{b}_{n_{\hat{b}}}(n)]^T$ is the set of parameters being adapted, $\mu(n)$ is a gain factor, $e(n)$ is an estimation error, and $\hat{\boldsymbol{\phi}}(n)$ is the regression vector associated to the algorithm. In this framework, the EE and OE algorithms are respectively characterized by

$$\begin{aligned} e(n) &\equiv e_{EE}(n) \\ &= \hat{A}(q,n)\{y(n)\} - \hat{B}(q,n)\{x(n)\} \\ \hat{\boldsymbol{\phi}}(n) &\equiv \hat{\boldsymbol{\phi}}_{EE}(n) \\ &= \left[-y(n-1)\ldots-y(n-n_{\hat{a}})\ x(n)\ldots x(n-n_{\hat{b}})\right]^T \\ e(n) &\equiv e_{OE}(n) \\ &= y(n) - \hat{y}(n) \\ \hat{\boldsymbol{\phi}}(n) &\equiv \hat{\boldsymbol{\phi}}_{OE}(n) \\ &= \left[-\hat{y}^f(n-1)\ldots-\hat{y}^f(n-n_{\hat{a}})\ x^f(n)\ldots x^f(n-n_{\hat{b}})\right]^T \end{aligned}$$

where the superscript ^{f} indicates that the corresponding signal is processed by the all-pole filter $\frac{1}{\hat{A}(q,n)}$. The gradient-type CSE adaptation algorithm is then described by [3]

$$\begin{aligned} \hat{\boldsymbol{\theta}}(n+1) &= \hat{\boldsymbol{\theta}}(n) \\ &+\mu\left[\gamma e_{EE}(n)\hat{\boldsymbol{\phi}}_{EE}(n)+(1-\gamma)e_{OE}(n)\hat{\boldsymbol{\phi}}_{OE}(n)\right] \end{aligned}$$

It can be shown that the CSE algorithm is the routine that attempts to minimize the composite squared error (CSE) signal, defined as

$$e^2_{CSE}(n) = \gamma e^2_{EE}(n) + (1-\gamma)e^2_{OE}(n) \quad (1)$$

where γ is the composite parameter commonly restrained to the interval $0 \leq \gamma \leq 1$. From (1), the mean composite square error (MCSE) performance surface is given by

$$E\left[e^2_{CSE}(n)\right] = \gamma E\left[e^2_{EE}(n)\right] + (1-\gamma)E\left[e^2_{OE}(n)\right]$$

2 Steady-State Analysis

The stationary points of the MCSE performance surface are the solutions of

$$E\left[\gamma e_{EE}(n)\hat{\boldsymbol{\phi}}_{EE}(n)+(1-\gamma)e_{OE}(n)\hat{\boldsymbol{\phi}}_{OE}(n)\right] = 0$$

*This work was supported by CAPES and CNPq - Ministry of Education (Brazil) and Micronet (Canada).

1058-6393/98 $10.00 © 1998 IEEE

For that surface the following result applies.

Property 1: The stationary points of the CSE algorithm are given by[1]

$$[\gamma \mathbf{R}^*_{EE}\mathbf{P}^*_{EE} + (1-\gamma)\mathbf{R}^*_{OE}\mathbf{P}^*_{OE}]\mathbf{h}^* - \gamma E\left[\mathbf{v}(n)\hat{A}^*(q)\{v(n)\}\right] = \mathbf{0}$$

where the asterisk symbol indicates the respective variable to be a solution of that equation, the vector $\mathbf{h} = [h_0 \ldots h_{n_h}]^T$ is composed by the coefficients of $H(q) = h_0 + \ldots + h_{n_h}q^{-n_h} = \bar{A}(q)B(q) - A(q)\bar{B}(q)$ and $\mathbf{v}(n) = [v(n-1)\ldots v(n-n_{\hat{a}})\ 0\ldots 0]^T$.

The implication of this property is that the stationary points of the CSE algorithm are anywhere from unique, as in the case of the EE algorithm ($\gamma = 1$), to unbiased, as in the case of the OE algorithm ($\gamma = 0$).

3 Transient Analysis

3.1 Ordinary-Difference-Equation Method

Property 2: Let $x(n)$ and $v(n)$ be stationary processes with finite first, second, and fourth moments [4]. If $\hat{A}(q,n)$ is stable for all n and $x(n)$ and $v(n)$ are ϕ-mixing as defined in [1], then the behavior of the CSE algorithm converges to the solution of the ordinary difference equation (ODE)

$$\begin{aligned}\frac{\mathrm{d}\hat{\mathbf{\Theta}}(t)}{\mathrm{d}t} &= V\left[\hat{\mathbf{\Theta}}(t)\right] \\ &= E\left[\gamma e_{EE}(n)\hat{\phi}_{EE}(n) + (1-\gamma)e_{OE}(n)\hat{\phi}_{OE}(n)\right]\end{aligned}$$

with $\hat{\mathbf{\Theta}}(0) = \hat{\mathbf{\Theta}}_0$,in probability, such that

$$P\left\{\sup_{0\le n\tau \le S} \|\hat{\theta}(n) - \hat{\mathbf{\Theta}}^*(n\tau)\| > C\varepsilon(\tau)\right\} < C'\varepsilon(\tau) \quad (2)$$

where S, C, and C' are positive constants and $\varepsilon(\tau)$ is a positive function going to zero as τ decreases.

Example 1: Let the plant and the adaptive filter be respectively given by

$$H(q) = \frac{0.05 - 0.4q^{-1}}{1 - 0.0003q^{-1} - 0.68915q^{-2}}$$

$$\hat{H}(q,n) = \frac{\hat{b}_0(n)}{1 - \hat{a}_1(n)q^{-1}}$$

With $\hat{\mathbf{\Theta}}(t) = [\hat{a}_1(t)\ \hat{b}_0(t)]^T$, the CSE algorithm can be associated to the ODE given by

$$\frac{\mathrm{d}V\left[\hat{\mathbf{\Theta}}(t)\right]}{\mathrm{d}t} = \begin{bmatrix}\gamma\,[f_{1A}+f_{1B}] + (1-\gamma)\,[f_{2A}+f_{2B}] \\ g_{1A}+g_{1B}\end{bmatrix} \quad (3)$$

[1] A complete definition of the matrices $\mathbf{R}^*_{EE}$, $\mathbf{P}^*_{EE}$, $\mathbf{R}^*_{OE}$, and $\mathbf{P}^*_{OE}$ is given in the Appendix A.

where

$$f_{1A} = \frac{(b_0^2 + b_1^2)(a_1 - \hat{a}_1(t) - a_2\hat{a}_1(t))}{(1-a_2)(1+2a_2 - a_1^2 - a_2^2)} \quad (4a)$$

$$f_{1B} = \frac{b_0 b_1(2a_1\hat{a}_1(t) - 1 - a_1^2 + a_2^2)}{(1-a_2)(1+2a_2 - a_1^2 - a_2^2)} \quad (4b)$$

$$f_{2A} = \frac{\hat{b}_0(t)(a_1 b_0 - b_1 - 2a_2 b_0 \hat{a}_1(t) + a_2 b_1 \hat{a}_1^2(t))}{(1 - a_1\hat{a}_1(t) + a_2\hat{a}_1^2(t))^2} \quad (4c)$$

$$f_{2B} = -\frac{\hat{a}_1(t)\hat{b}_0^2(t)}{(1-\hat{a}_1^2(t))^2} \quad (4d)$$

$$g_{1A} = \frac{\gamma\hat{a}_1(t)(b_1 - a_1 b_0 + a_2 b_0\hat{a}_1(t)) + (b_0 - b_1\hat{a}_1(t))}{(1 - a_1\hat{a}_1(t) + a_2\hat{a}_1^2(t))} \quad (4e)$$

$$g_{1B} = -\frac{\hat{b}_0(t)(1-\gamma\hat{a}_1^2(t))}{(1-\hat{a}_1^2(t))} \quad (4f)$$

Figure 2 shows the predicted trajectories (left-hand side), using (3)-(4), and the actual adaptive coefficient trajectories (right-hand side) for distinct values of γ. In this figure, it is easy to verify how well the proposed analysis complies with the actual results.

3.2 Local Linearization Method

The ensemble mean of the parameter trajectories $E[\hat{\theta}(n)]$ generated by an adaptation algorithm is a deterministic locus [2]. For the CSE algorithm, one has

$$\begin{aligned}&E\left[\hat{\theta}(n+1) - \hat{\theta}(n)\right] = \\ &\mu E\left[\gamma e_{EE}(n)\hat{\phi}_{EE}(n) + (1-\gamma)e_{OE}(n)\hat{\phi}_{OE}(n)\right]\end{aligned}$$

This function can be approximated around a stationary point $\hat{\theta}^*$, using a first-order Taylor series, by

$$E\left[\hat{\theta}(n+1) - \hat{\theta}(n)\right] \simeq \mathbf{Q}E\left[\hat{\theta}(n) - \hat{\theta}^*\right] \quad (5)$$

where $\mathbf{Q}$ is a sensitivity matrix given by

$$\mathbf{Q} = -\frac{\mu}{2}\frac{\mathrm{d}^2 E\left[e^2_{CSE}(n)\right]}{\mathrm{d}\hat{\theta}^2}\Big|_{\hat{\theta}=\hat{\theta}^*} \quad (6)$$

Defining the coefficient error vector as $\tilde{\theta}(n) = \hat{\theta}(n) - \hat{\theta}^*$, equation (5) yields

$$E\left[\tilde{\theta}(n)\right] \simeq (\mathbf{I}+\mathbf{Q})^n\,\tilde{\theta}(0) \quad (7)$$

This equation indicates that the trajectories followed by the CSE algorithm present a dumped exponential property controlled by the absolute values of the eigenvalues of the matrix $(\mathbf{I}+\mathbf{Q})$.

Property 3: Let $\hat{\theta}^*$ be the minimum of a concave region $\mathcal{R}$ of the MCSE performance surface and assume that the adaptive filter remains stable throughout the adaptation process. Convergence of the CSE algorithm to $\hat{\boldsymbol{\theta}}^*$ is then guaranteed, for a sufficiently small value of μ, if the adaptive filter is initialized in $\mathcal{R}$.

Example 2: Consider an identification case where $n_a = n_{\hat{a}} = 1$, $n_b = n_{\hat{b}} = 0$, and $v(n) \equiv 0$. Thus

$$y(n) = b_0 x(n) - a_1 y(n-1)$$
$$\hat{y}(n) = \hat{b}_0(n)x(n) - \hat{a}_1(n)\hat{y}(n-1)$$

Assume the input signal $x(n)$ is a Gaussian noise with zero mean and unitary variance. Then, in a neighborhood of the stationary point $\hat{\boldsymbol{\theta}}^* = [a_1 \; b_0]^T$, with the CSE algorithm, $E[\tilde{\boldsymbol{\theta}}(n)]$ evolves as in (7) with

$$\mathbf{Q} = \mu \begin{bmatrix} \frac{(3a_1^2\gamma - a_1^4\gamma - a_1^2 - 1)b_0^2}{(1-a_1^2)^3} & -(1-\gamma)\frac{a_1 b_0}{(1-a_1^2)^2} \\ -(1-\gamma)\frac{a_1 b_0}{(1-a_1^2)^2} & \frac{(a_1^2\gamma - 1)}{(1-a_1^2)} \end{bmatrix}$$

For the case with $a_1 = 0.7$, $b_0 = 0.5$, and $\mu = 0.002$, the predicted (left-hand side), using the local linearization method, and the actual (right-hand side) trajectories followed by the CSE algorithm are shown in Figure 3.

When $a_1 = 0.85$, $b_0 = 1.7$, $\mu = 0.001$, and $\hat{\boldsymbol{\theta}}(0) = [0.85 \; 1.84]^T$, the convergence speed of the CSE algorithm was measured, for several values of γ, by the number of iterations N for which $\|\tilde{\boldsymbol{\theta}}(n)\|_2^2 \leq 10^{-5}, \forall n \geq N$. The averaged results over an ensemble of 12 experiments are listed in Table 1 along with the eigenvalues of the matrix $(\mathbf{Q} + \mathbf{I})$. Notice that $|\lambda_2|$ becomes the major factor to determine the convergence speed of the algorithm, as the value of $|\lambda_1|$, of the same magnitude order of $|\lambda_2|$, remains practically unchanged.

Table 1: Example 2 - Eigenvalues of the $(\mathbf{Q}+\mathbf{I})$ Matrix and Number of Iterations N for Convergence of the CSE Algorithm as Functions of Composite Factor γ.

γ	$\lvert\lambda_1\rvert$	$\lvert\lambda_2\rvert$	N
0.0	0.9979	0.7655	1591
0.1	0.9980	0.7879	1719
0.2	0.9981	0.8103	1808
0.3	0.9982	0.8328	1908
0.4	0.9983	0.8552	2094
0.5	0.9984	0.8776	2204
0.6	0.9985	0.9000	2474
0.7	0.9986	0.9224	2690
0.8	0.9987	0.9448	2949
0.9	0.9989	0.9672	3285
1.0	0.9990	0.9896	3762

4 Conclusion

In this paper, a thorough study of the composite squared error (CSE) algorithm was performed. In that manner, the transient and the steady-state parts of the CSE convergence process were analyzed. The results showed that the CSE algorithm can take advantage of the quadratic nature of the equation error scheme and the unbiased global minimum of the output error method.

References

[1] H. Fan, "Application of Benveniste's convergence results in a study of adaptive IIR filtering algorithms," *IEEE Trans. Inf. Theory*, vol. 34, no. 4, pp. 692–709, July 1988.

[2] M. G. Larimore, J. R. Treichler, and C. R. Johnson, Jr., "SHARF: an algorithm for adapting IIR digital filters," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. ASSP-28, no. 4, pp. 428–440, Aug. 1980.

[3] S. L. Netto and P. Agathoklis, "A new composite adaptive IIR algorithm," *Proc. 28th Asilomar Conf. Signals, Systems & Computers*, Pacific Grove, USA, pp. 1506–1510, Oct./Nov. 1994.

[4] A. Papoulis, *Probability, Random Variables, and Stochastic Processes*, McGraw-Hill, New York, USA, 3rd Edition, 1991.

[5] B. Widrow and S. Stearns, *Adaptive Signal Processing*, Prentice-Hall, Englewood-Cliffs, USA, 1985.

Appendix A

Definition 1: The plant coefficient matrix of order $(n_{\hat{a}}+n_{\hat{b}}+1)\times(n_h+n_l+1)$, the adaptive filter coefficient matrix of order $(n_{\hat{a}}+n_{\hat{b}}+1)\times(n_{\hat{a}}+n_{\hat{b}}-n_l+1)$, and the matrices of orders $(n_h+n_l+1)\times(n_h+1)$ and $(n_{\hat{a}}+n_{\hat{b}}-n_l+1)\times(n_h+1)$ are respectively defined by

$$\mathbf{R}_{EE} = \begin{bmatrix} 0 & -b_0 & \cdots & -b_{n_b} & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & \cdots & -b_0 & \cdots & -b_{n_b} \\ 1 & a_1 & \cdots & a_{n_a} & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & 1 & a_1 & \cdots & a_{n_a} \end{bmatrix}$$

$$\mathbf{R}_{OE} = \begin{bmatrix} 0 & -\bar{b}_0 & \cdots & -\bar{b}_{n_{\bar{b}}} & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & -\bar{b}_0 & \cdots & \cdots & -\bar{b}_{n_{\bar{b}}} \\ 1 & \bar{a}_1 & \cdots & \bar{a}_{n_{\bar{a}}} & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & 1 & \bar{a}_1 & \cdots & \bar{a}_{n_{\bar{a}}} \end{bmatrix}$$

$$\mathbf{P}_{EE} = E\left[\left(\frac{1}{A(q)}\right)\left\{\begin{bmatrix} x(n) \\ \vdots \\ x(n-n_h-n_l) \end{bmatrix}\right\}\right] \cdot \left[\left(\frac{L(q)}{A(q)}\right)\{[x(n)\ldots x(n-n_h)]\}\right]$$

$$\mathbf{P}_{OE} = E\left[\left(\frac{1}{\hat{A}(q)\bar{A}(q)}\right)\left\{\begin{bmatrix} x(n) \\ \vdots \\ x(n-n_{\hat{a}}-n_{\hat{b}}+n_l) \end{bmatrix}\right\}\right] \cdot \left[\left(\frac{1}{A(q)\bar{A}(q)}\right)\{[x(n)\ldots x(n-n_h)]\}\right]$$

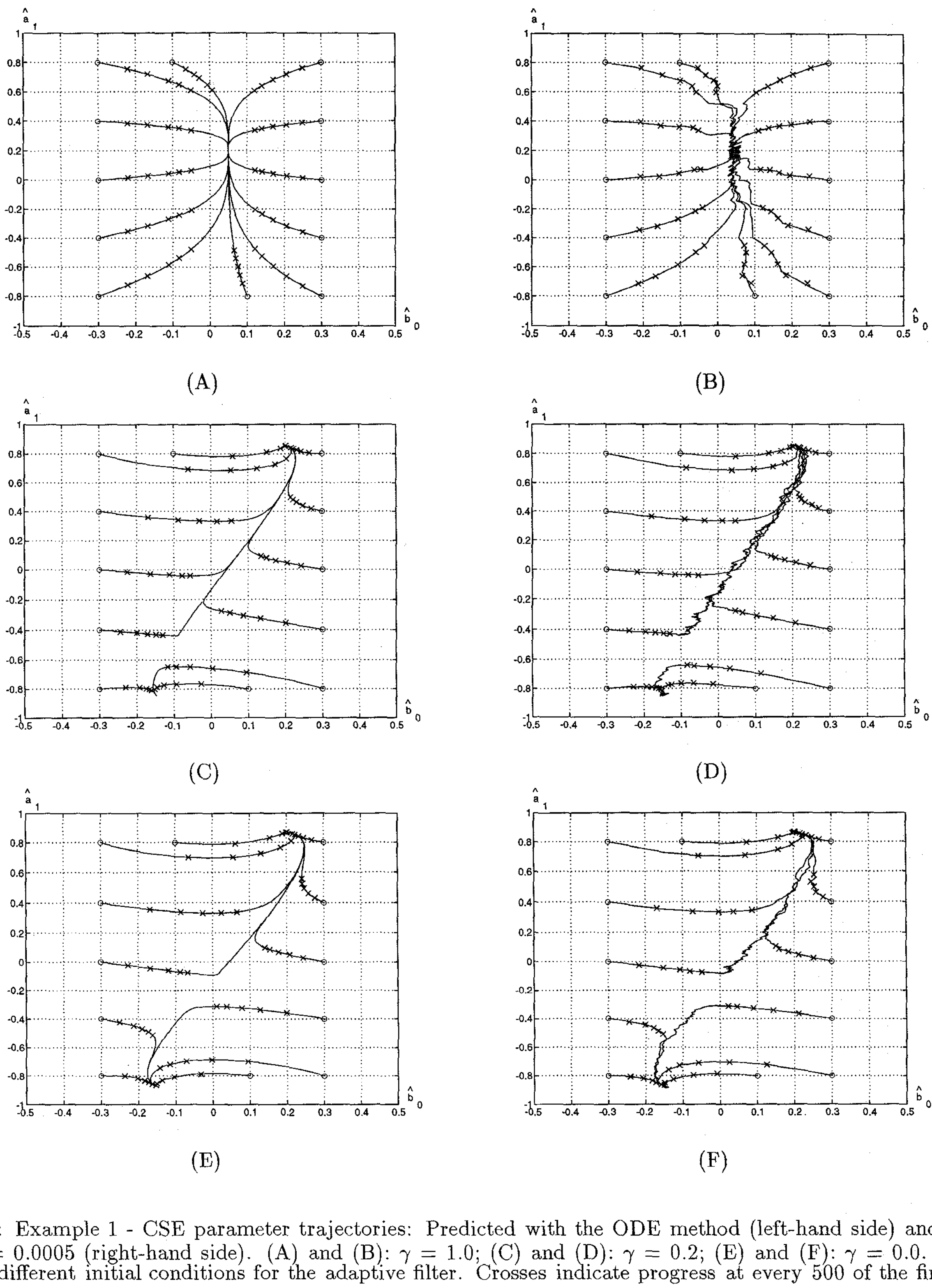

Figure 2: Example 1 - CSE parameter trajectories: Predicted with the ODE method (left-hand side) and actual with $\mu = 0.0005$ (right-hand side). (A) and (B): $\gamma = 1.0$; (C) and (D): $\gamma = 0.2$; (E) and (F): $\gamma = 0.0$. Circles indicate different initial conditions for the adaptive filter. Crosses indicate progress at every 500 of the first 2500 iterations.

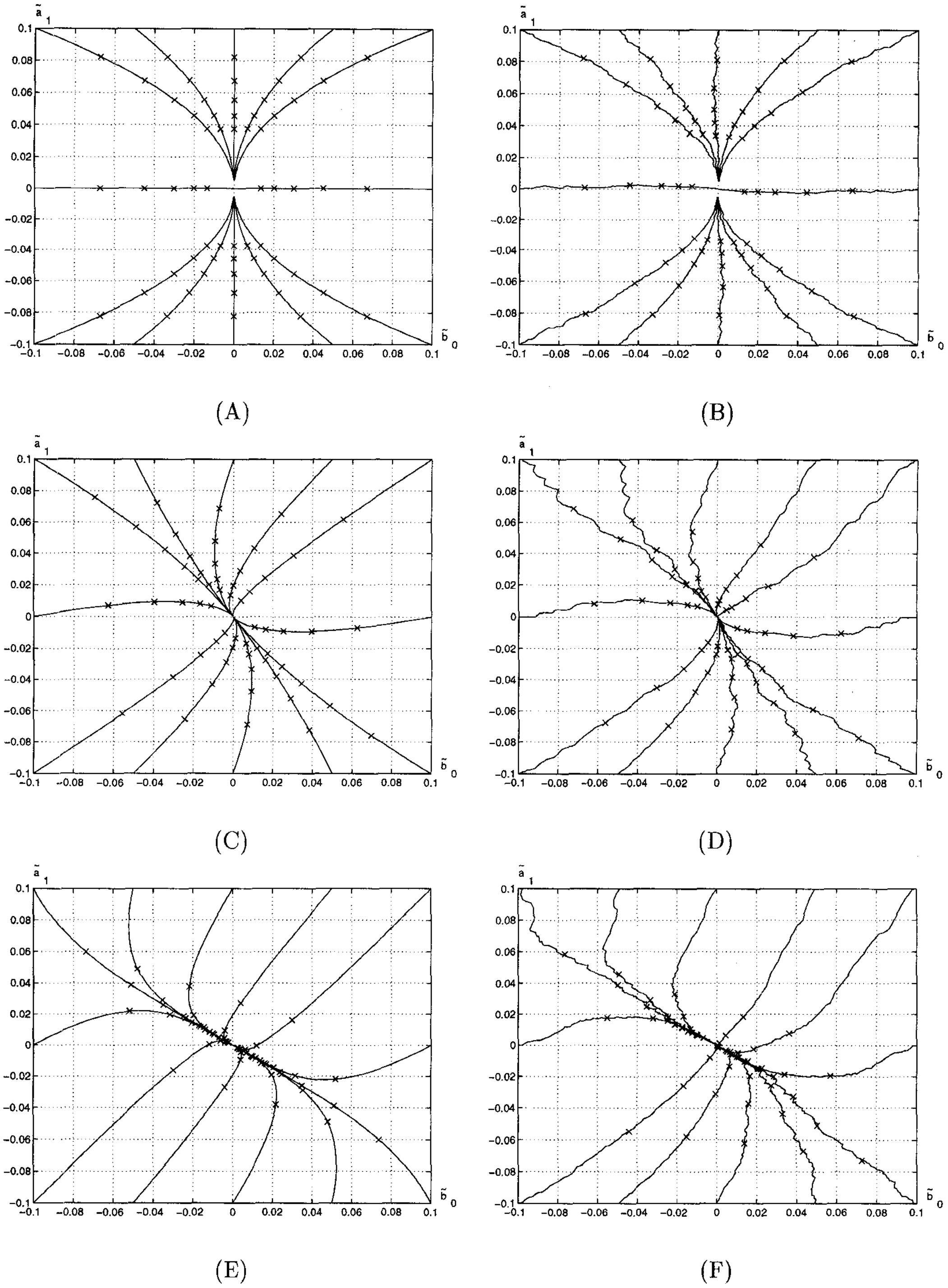

Figure 3: Example 2 - CSE parameter trajectories: Predicted with the local linearization method (left-hand side) and actual (right-hand side). (A) and (B): $\gamma = 1.0$; (C) and (D): $\gamma = 0.8$; (E) and (F): $\gamma = 0.0$. Crosses indicate progress at every 200 of the first 1000 iterations.