

# GENERALIZED ADAPTIVE NOTCH FILTERS

Maciej Niedźwiecki and Piotr Kaczmarek

Faculty of Electronics, Telecommunications and Computer Science, Department of Automatic Control  
Gdańsk University of Technology, ul. Narutowicza 11/12, Gdańsk, Poland  
maciekn@eti.pg.gda.pl, piokacz@proterians.net.pl

## ABSTRACT

The problem of identification/tracking of quasi-periodically varying systems is considered. This problem is a generalization, to the system case, of a classical signal processing task of either elimination or extraction of nonstationary sinusoidal signals buried in noise. The proposed solution is based on the exponentially weighted basis function (EWBF) approach. First, the global EWBF algorithm is derived and its decomposed, parallel-form and cascade-form variants, are described. Then the frequency-adaptive versions of both schemes are obtained using the recursive prediction error method. In the (special) signal processing case the paper offers new attractive solutions to the problem of adaptive notch filtering.

## 1. PROBLEM STATEMENT

Consider the problem of identification/tracking of coefficients of a complex time varying system governed by

$$y(t) = \sum_{l=1}^n \theta_l(t)u(t-l+1) + v(t) \\ = \boldsymbol{\varphi}^T(t)\boldsymbol{\theta}(t) + v(t) \quad (1)$$

where  $t = 1, 2, \dots$  denotes the normalized discrete time,  $y(t)$  denotes the system output,  $\boldsymbol{\varphi}(t) = [u(t), \dots, u(t-n+1)]^T$  is the regression vector made up of the past input samples,  $v(t)$  is an additive (white) noise, uncorrelated with  $u(t)$ , and  $\boldsymbol{\theta}(t) = [\theta_1(t), \dots, \theta_n(t)]^T$  denotes the vector of time varying impulse response coefficients, modeled as weighted sums of complex exponentials

$$\theta_l(t) = \sum_{i=1}^k a_{li}(t)e^{j\sum_{s=1}^t \omega_i(s)}, \quad l = 1, \dots, n \quad (2)$$

We will assume that for every frequency component  $i$ ,  $i = 1, \dots, k$  the quantities  $a_{li}(t)$ ,  $l = 1, \dots, n$  and  $\omega_i(t)$  are slowly time-varying. The system, governed by (1) - (2),

which obeys the above-mentioned limitation, will be further referred to as *quasi-periodically* time-varying.

One of the challenging potential applications, which under certain conditions admits formulation presented above, is adaptive equalization of rapidly fading communication channels – see e.g. [1] and [2] for more details.

For  $n = 1$  and  $u(t) = 1, \forall t$  the model (1) - (2) becomes a description of a noisy nonstationary multifrequency signal

$$y(t) = \sum_{i=1}^k a_i(t)e^{j\sum_{s=1}^t \omega_i(s)} + v(t). \quad (3)$$

Hence, when restricted to the special case discussed above, the results developed in the paper offer a new solution to the problem of frequency tracking and adaptive notch filtering.

## 2. KNOWN FREQUENCIES

Suppose, for the time being, that both the amplitudes and angular frequencies in (2) are constant, i.e. that the changes in system parameters are governed by

$$\theta_l(t) = \sum_{i=1}^k a_{li}e^{j\omega_i t}, \quad l = 1, \dots, n \quad (4)$$

Let  $\boldsymbol{\alpha}_i = [a_{1i}, \dots, a_{ni}]^T$ ,  $\boldsymbol{\psi}_i(t) = \boldsymbol{\varphi}(t)e^{j\omega_i t}$ ,  $i = 1, \dots, k$ . Using this short-hand notation, (1) can be rewritten in the form

$$y(t) = \sum_{i=1}^k \boldsymbol{\psi}_i^T(t)\boldsymbol{\alpha}_i + v(t) = \boldsymbol{\psi}^T(t)\boldsymbol{\alpha} + v(t) \quad (5)$$

where  $\boldsymbol{\alpha} = [\boldsymbol{\alpha}_1^T, \dots, \boldsymbol{\alpha}_k^T]^T$ .  $\boldsymbol{\psi}(t) = [\boldsymbol{\psi}_1^T(t), \dots, \boldsymbol{\psi}_k^T(t)]^T = \mathbf{f}(t) \otimes \boldsymbol{\varphi}(t)$ ,  $\mathbf{f}(t) = [e^{j\omega_1 t}, \dots, e^{j\omega_k t}]^T$  and  $\otimes$  denotes the Kronecker product. Note that  $\boldsymbol{\alpha}_i$  is the vector of coefficients associated with a particular frequency  $\omega_i$  and *not* with a particular impulse response parameter  $\theta_i(t)$ . Similarly,  $\boldsymbol{\psi}_i(t)$  is the generalized regression vector associated

This work was supported by KBN under Grant 4 T11A 01225.

with the  $i$ th frequency component. This nonstandard parameterization was adopted deliberately. Later on it will allow us to easily derive decoupled versions of the basic estimation algorithm.

Suppose now that the vector  $\alpha$  is slowly varying with time. It is known that, in the case considered, one can track  $\alpha(t)$  using the method of exponentially weighted least squares (EWLS). The EWLS estimate of  $\alpha(t)$  can be obtained from

$$\hat{\alpha}(t) = \arg \min_{\alpha} \sum_{s=1}^t \lambda^{t-s} |y(s) - \psi^T(s)\alpha|^2 \quad (6)$$

where  $\lambda$  ( $0 < \lambda < 1, 1 - \lambda \ll 1$ ) denotes the so-called forgetting constant - the design parameter which controls the memory of the estimator, and hence allows one to trade off between its tracking speed and tracking accuracy. The recursive algorithm for evaluation of  $\hat{\alpha}(t)$  is given by

$$\begin{aligned} \hat{\alpha}(t) &= \hat{\alpha}(t-1) + (\mathbf{R}_{\alpha}^*(t))^{-1} \psi^*(t) \varepsilon(t) \\ \varepsilon(t) &= y(t) - \psi^T(t) \hat{\alpha}(t-1) \\ \mathbf{R}_{\alpha}(t) &= \lambda \mathbf{R}_{\alpha}(t-1) + \psi(t) \psi^H(t) \end{aligned} \quad (7)$$

Based on (7) one can estimate system parameters using

$$\hat{\theta}(t) = \mathbf{D}(t) \hat{\alpha}(t) \quad (8)$$

where  $\mathbf{D}(t) = \mathbf{f}^T(t) \otimes \mathbf{I}_n$  and  $\mathbf{I}_n$  denotes the  $n \times n$  identity matrix.

Since the algorithm (7) - (8) combines the basis function parameterization (it is assumed that system parameters can be expressed as linear combinations of known functions of time, called basis functions) with exponentially weighted least squares estimation, it will be further referred to as the exponentially weighted basis function (EWBF) algorithm [3], [4].

Another, equivalent form of the EWBF estimator, which will be very useful for our purposes, can be obtained by rewriting (7) - (8) in a different system of coordinates. Using the linear time-varying transformation

$$\hat{\beta}(t) = \mathbf{A}_n^{t+1} \hat{\alpha}(t), \quad \mathbf{R}_{\beta}(t) = \mathbf{A}_n^{-(t+1)} \mathbf{R}_{\alpha}(t) \mathbf{A}_n^{t+1}$$

where  $\mathbf{A}_n = \mathbf{A} \otimes \mathbf{I}_n$  and  $\mathbf{A} = \text{diag}\{e^{j\omega_1}, \dots, e^{j\omega_k}\}$ , one can easily convert (7) - (8) into

$$\begin{aligned} \hat{\theta}(t) &= \mathbf{D}_0 \hat{\beta}(t) \\ \hat{\beta}(t) &= \mathbf{A}_n \hat{\beta}(t-1) + (\mathbf{R}_{\beta}^*(t))^{-1} \mathbf{A}_n \varphi_n^*(t) \varepsilon(t) \\ \varepsilon(t) &= y(t) - \varphi_n^T(t) \hat{\beta}(t-1) \\ \mathbf{R}_{\beta}(t) &= \mathbf{A}_n^* [\lambda \mathbf{R}_{\beta}(t-1) + \varphi_n(t) \varphi_n^H(t)] \mathbf{A}_n \end{aligned} \quad (9)$$

where  $\varphi_n(t) = \mathbf{A}_n^{-t} \psi(t) = \mathbf{f}(0) \otimes \varphi(t) = [\varphi^T(t), \dots, \varphi^T(t)]^T$  and  $\mathbf{D}_0 = \mathbf{D}(t) \mathbf{A}_n^{-(t+1)} = (\mathbf{f}^T(t) \otimes \mathbf{I}_n) (\mathbf{A}_n^{-(t+1)} \otimes \mathbf{I}_n) = \mathbf{f}^H(1) \otimes \mathbf{I}_n$ . The last transformation follows from the identity  $(\mathbf{X} \otimes \mathbf{Y})(\mathbf{P} \otimes \mathbf{Q}) = \mathbf{X}\mathbf{P} \otimes \mathbf{Y}\mathbf{Q}$  which holds for Kronecker products.

## 2.1. Parallel decomposition

Denote by

$$y_i(t) = \psi_i^T(t) \alpha_i + v(t)$$

the output of the  $i$ th subsystem of (5), i.e. subsystem associated with the frequency  $\omega_i$ . Even though the signal  $y_i(t)$  is not available, one can easily estimate it using the formula

$$\hat{y}_i(t) = y(t) - \sum_{\substack{l=1 \\ l \neq i}}^k \hat{y}_l(t|t-1) \quad (10)$$

where  $\hat{y}_i(t|t-1) = \psi_i^T(t) \hat{\alpha}_i(t-1) = \varphi^T(t) \hat{\beta}_i(t-1)$  is the predicted value of  $y_i(t)$  yielded by the estimation algorithm designed to track parameters of the  $i$ th subsystem.

Estimation of  $y_i(t)$ , in the way described above, allows one to decompose the tracking algorithm, i.e. to replace one 'global' algorithm (9) with  $k$  mutually coupled 'local' algorithms, each of which takes care of a particular subsystem. The  $i$ th component algorithm can be easily derived from (9) by setting  $\mathbf{A}_n(t) = \rho_i \mathbf{I}_n$ ,  $\rho_i = e^{j\omega_i}$  and  $\varphi_n(t) = \varphi(t)$ . To add some extra design flexibility, we will equip each subalgorithm with an independently assigned forgetting factor  $\lambda_i$ . The resulting decoupled algorithm can be written down in the form

$$\begin{aligned} \hat{\beta}_i(t) &= \rho_i \left[ \hat{\beta}_i(t-1) + (\mathbf{R}_i^*(t))^{-1} \varphi^*(t) \varepsilon_i(t) \right] \\ \varepsilon_i(t) &= \hat{y}_i(t) - \varphi^T(t) \hat{\beta}_i(t-1) \\ \mathbf{R}_i(t) &= \lambda_i \mathbf{R}_i(t-1) + \varphi(t) \varphi^H(t) \\ & \quad i = 1, \dots, k \\ \hat{\theta}(t) &= \sum_{i=1}^k \rho_i^* \hat{\beta}_i(t) \end{aligned} \quad (11)$$

Note that  $\varepsilon_1(t) = \dots = \varepsilon_k(t) = y(t) - \sum_{i=1}^k \varphi^T(t) \hat{\beta}_i(t-1)$  i.e. all subalgorithms are driven by the same global prediction error. Therefore the algorithm (11) can be rewritten in the form

$$\begin{aligned} \hat{\beta}_i(t) &= \rho_i \left[ \hat{\beta}_i(t-1) + \mathbf{k}_i^*(t) \varepsilon(t) \right] \\ \varepsilon(t) &= y(t) - \sum_{i=1}^k \varphi^T(t) \hat{\beta}_i(t-1) \\ \mathbf{k}_i(t) &= \frac{\mathbf{P}_i(t-1) \varphi(t)}{\lambda_i + \varphi^H(t) \mathbf{P}_i(t-1) \varphi(t)} \\ \mathbf{P}_i(t) &= \frac{1}{\lambda_i} [\mathbf{I}_n - \mathbf{k}_i(t) \varphi^H(t)] \mathbf{P}_i(t-1) \\ & \quad i = 1, \dots, k \\ \hat{\theta}(t) &= \sum_{i=1}^k \rho_i^* \hat{\beta}_i(t) \end{aligned} \quad (12)$$

where  $\mathbf{P}_i(t) = \mathbf{R}_i^{-1}(t)$ .

In order to shed more light on the algorithm described above, consider the problem of extraction of a complex sinusoidal signal  $\theta(t) = ae^{j\omega_i t}$ , of known frequency  $\omega_i$ , from noisy measurements  $y(t) = \theta(t) + v(t)$ . In the case considered the EWBF recursions (12) reduce down to ( $k = 1, n = 1, u(t) = 1$ )

$$\begin{aligned}\widehat{\theta}(t) &= \rho_i^* \widehat{\beta}_i(t) \\ \widehat{\beta}_i(t) &= \rho_i \left[ \widehat{\beta}_i(t-1) + \frac{\varepsilon(t)}{r_i(t)} \right] \\ \varepsilon(t) &= y(t) - \widehat{\beta}_i(t-1) \\ r_i(t) &= \lambda_i r_i(t-1) + 1\end{aligned}$$

Note that the gain  $r_i(t)$  stabilizes at the value  $1/(1 - \lambda_i)$  as time goes to infinity, resulting in the following steady-state relationship between the prediction error  $\varepsilon(t)$  and the input signal  $y(t)$

$$\varepsilon(t) = N_i(q^{-1})y(t)$$

where the filter  $N_i(q^{-1}) = (1 - \rho_i q^{-1})/(1 - \lambda_i \rho_i q^{-1})$  can be easily recognized as the notch filter centered at the notch frequency  $\omega_i$ , with bandwidth dependent on the forgetting factor  $\lambda_i$ . Hence, when used for extraction or cancellation of nonstationary multifrequency signals buried in noise, the EWBF algorithm (12) can be regarded as a bank of interconnected notch filters.

## 2.2. Cascade decomposition

The decoupled algorithm presented in the previous subsection is a parallel structure made up of  $k$  identical (from the functional viewpoint) blocks. Each block is designed to track a particular frequency component of the parameter vector  $\theta(t)$ . Connecting the same blocks so that they form a cascade, one obtains an interesting alternative to the parallel decomposition. To obtain the cascade variant of the EWBF algorithm, the first two recursions in (12) should be replaced with

$$\begin{aligned}\widehat{\beta}_i(t) &= \rho_i \left[ \widehat{\beta}_i(t-1) + \mathbf{k}_i^*(t) \varepsilon_i(t) \right] \\ \varepsilon_i(t) &= \varepsilon_{i-1}(t) - \boldsymbol{\varphi}^T(t) \widehat{\beta}_i(t-1)\end{aligned}\quad (13)$$

where  $\varepsilon_0(t) = y(t)$ .

The cascade-form implementations are widespread in signal processing. Consider a multi-frequency notch filter realized as a cascade of single-frequency filters. When the bandwidths of the component filters are sufficiently narrow, so that they have negligible overlap, each section can eliminate one sinusoid. The first section cancels one sinusoid and transmits the rest of the signal, with small distortion, to the second stage, which cancels the second sinusoid (if present) etc. The cascade-form parameter tracking algorithm (13) exploits the same frequency decoupling property of narrow-band filters.

## 3. UNKNOWN FREQUENCIES

Even though the EWBF filter is robust to small local changes in frequencies, it will fail to identify the system correctly in the presence of a frequency drift. For this reason in this section we will derive two frequency-adaptive EWBF algorithms, capable of tracking the time-varying frequencies  $\omega_i(t), i = 1, \dots, k$ .

Consider the following standardized form of the  $i$ th subalgorithm, which fits both realizations discussed in Section II and explicitly shows dependence of various terms on  $\omega_i$

$$\begin{aligned}\widehat{\beta}(t, \omega_i) &= e^{j\omega_i} \left[ \widehat{\beta}(t-1, \omega_i) + \mathbf{k}_i^*(t) \varepsilon(t, \omega_i) \right] \\ \varepsilon(t, \omega_i) &= z_i(t) - \boldsymbol{\varphi}^T(t) \widehat{\beta}(t-1, \omega_i)\end{aligned}$$

where  $z_i(t) = \widehat{y}_i(t)$  in the parallel implementation, and  $z_i(t) = \varepsilon_{i-1}(t)$  in the cascade implementation.

Denote by  $V(t, \omega_i)$  the local exponentially weighted measure of fit

$$V(t, \omega_i) = \frac{1}{2} \sum_{s=1}^t \gamma_i^{t-s} |\varepsilon(s, \omega_i)|^2 \quad (14)$$

where  $\gamma_i, 0 < \gamma_i < 1$ , is the forgetting constant, which will be used to control the speed of the frequency adaptation. To evaluate the estimate  $\widehat{\omega}_i(t) = \arg \min_{\omega_i} V(t, \omega_i)$  we will use the recursive prediction error (RPE) approach – see e.g. Söderström and Stoica [5].

Let  $\varepsilon_i(t) = \varepsilon(t, \widehat{\omega}_i(t))$ ,  $\widehat{\beta}_i(t) = \widehat{\beta}(t, \widehat{\omega}_i(t))$ ,  $\eta_i(t) = \partial \varepsilon(t, \widehat{\omega}_i(t-1)) / \partial \omega_i$ ,  $\boldsymbol{\xi}_i(t) = \partial \widehat{\beta}(t, \widehat{\omega}_i(t)) / \partial \omega_i$ , and  $\nu_i(t) = V''(t, \widehat{\omega}_i(t-1))$ . Using the notation introduced above, the RPE-type frequency-adaptive EWBF algorithm can be written down in the form

$$\begin{aligned}\varepsilon_i(t) &= z_i(t) - \boldsymbol{\varphi}^T(t) \widehat{\beta}_i(t-1) \\ \eta_i(t) &= -\boldsymbol{\varphi}^T(t) \boldsymbol{\xi}_i(t-1) \\ \nu_i(t) &= \gamma_i \nu_i(t-1) + |\eta_i(t)|^2 \\ \widehat{\omega}_i(t) &= \widehat{\omega}_i(t-1) - \nu_i^{-1}(t) \text{Re}[\varepsilon_i(t) \eta_i^*(t)] \\ \widehat{\rho}_i(t) &= e^{j\widehat{\omega}_i(t)} \\ \mathbf{k}_i(t) &= \frac{\mathbf{P}_i(t-1) \boldsymbol{\varphi}(t)}{\lambda_i + \boldsymbol{\varphi}^H(t) \mathbf{P}_i(t-1) \boldsymbol{\varphi}(t)} \\ \mathbf{P}_i(t) &= \frac{1}{\lambda_i} [\mathbf{I}_n - \mathbf{k}_i(t) \boldsymbol{\varphi}^H(t)] \mathbf{P}_i(t-1) \\ \widehat{\beta}_i(t) &= \widehat{\rho}_i(t) [\widehat{\beta}_i(t-1) + \mathbf{k}_i^*(t) \varepsilon_i(t)] \\ \boldsymbol{\xi}_i(t) &= \widehat{\rho}_i(t) [\boldsymbol{\xi}_i(t-1) + \mathbf{k}_i^*(t) \eta_i(t)] + j \widehat{\beta}_i(t)\end{aligned}$$

$$i = 1, \dots, k$$

$$\widehat{\theta}(t) = \sum_{i=1}^k \widehat{\rho}_i^*(t) \widehat{\beta}_i(t) \quad (15)$$

In the parallel realization one should set  $z_i(t) = \hat{y}_i(t)$ , which leads to  $\varepsilon_1(t) = \dots = \varepsilon_k(t) = y(t) - \sum_{i=1}^k \varphi^T(t) \beta_i(t-1) = \varepsilon(t)$ . In the cascade realization one should use  $z_1(t) = y(t)$  and  $z_i(t) = \varepsilon_{i-1}(t), i > 1$ .

It should be noted that the proposed algorithms are novel, both in the system case and in the signal case. The simple gradient search algorithms, which bear some resemblance to (15), were proposed in [1] and [2].

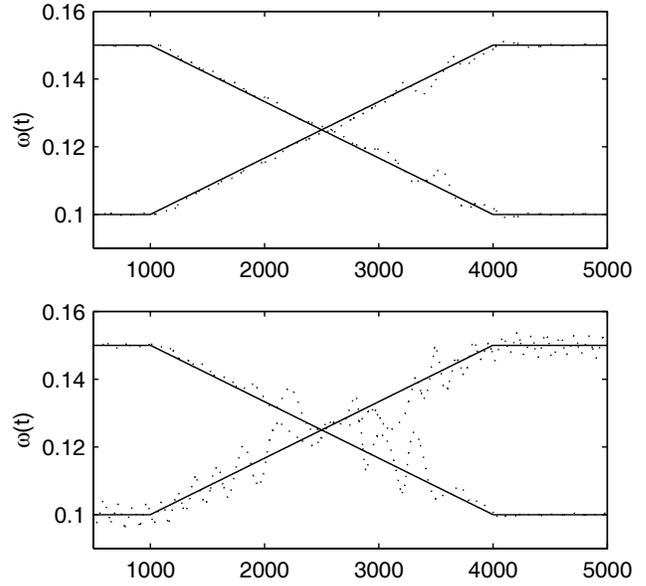
#### 4. COMPUTER SIMULATIONS

The system identification/tracking results, shown in the figures 1 and 2, were obtained for a hypothetical time-varying communication channel with two impulse response coefficients  $\theta_1(t)$  and  $\theta_2(t)$  ( $n=2$ ), each of which was modeled as a linear combination of two complex exponentials ( $k=2$ ). The weighting coefficients in (2) had constant values  $\alpha = [a_{11}, a_{12}, a_{21}, a_{22}]^T = [2, 0.5j, -j, 1.5]^T$ . The input signal was the white 4-QAM sequence ( $u(t) = \pm 1 \pm j, \sigma_u^2 = 2$ ) and the noise was complex Gaussian with variance  $\sigma_v^2 = 0.2$ . The forgetting constants  $\lambda_1 = \lambda_2 = \lambda$  and  $\gamma_1 = \gamma_2 = \gamma$  were set to 0.99 and 0.98, respectively.

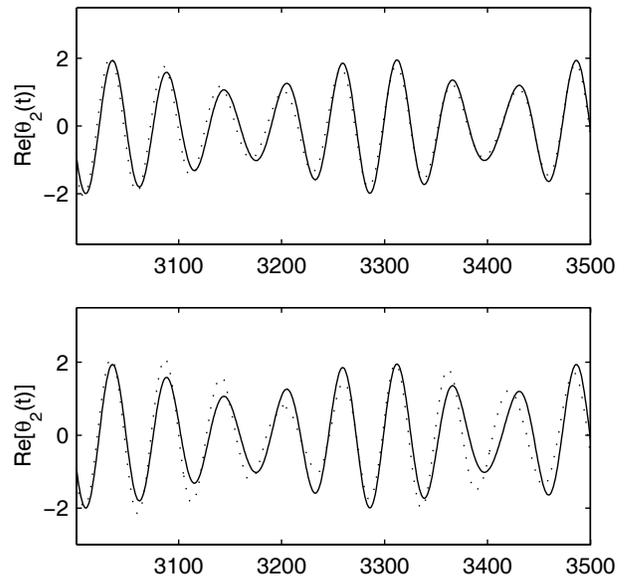
The simulated frequency changes were of the ramp type. To check the 'steady state' tracking capabilities of the compared algorithms, the linear changes in frequencies were enforced only after the initial convergence period was over. Since the trajectories of  $\omega_1(t)$  and  $\omega_2(t)$  intersect in the middle of the analysis interval, one of our main concerns was the behavior of the tracking algorithms in the vicinity of the crossover point. Although both algorithms show satisfactory tracking performance, the parallel-form adaptive notch filter yields consistently better results than the cascade-form filter. This is confirmed by both the frequency estimation plots and parameter tracking plots.

#### 5. REFERENCES

- [1] Tsatsanis, M.K. and G.B. Giannakis (1996). Modeling and equalization of rapidly fading channels. *Int. J. Adaptive Contr. Signal Process.* **vol. 10**, 159–176.
- [2] Bakkoury, J., D. Roviras, M. Ghogho and F. Castanie (2000). Adaptive MLSE receiver over rapidly fading channels. *Signal Processing* **vol. 80**, 1347–1360.
- [3] Niedźwiecki, M. (1990). Recursive functional series modeling approach to identification of time-varying plants - more bad news than good?. *IEEE Trans. Automat. Contr.* **AC-35**, 610–616.
- [4] Niedźwiecki, M. (2000). *Identification of Time-varying Processes*. Wiley. New York.
- [5] Söderström, T. and P. Stoica (1988). *System Identification*. Prentice Hall. Englewood Cliffs NJ.



**Fig. 1.** Instantaneous Doppler frequencies of a simulated mobile radio channel (solid lines) and their estimates (dotted lines) obtained using the parallel-form algorithm (upper plot) and the cascade-form algorithm (lower plot).



**Fig. 2.** Evolution of the real part of the true system parameter  $\theta_2(t)$  (solid line) and its estimates (dotted line) obtained using the parallel-form algorithm (upper plot) and the cascade-form algorithm (lower plot).