Self-tuning Information Fusion Kalman Predictor Weighted by Diagonal Matrices and Its Convergence Analysis

DENG Zi-Li[1] LI Chun-Bo[1]

Abstract For the multisensor systems with unknown noise statistics, using the modern time series analysis method, based on on-line identification of the moving average (MA) innovation models, and based on the solution of the matrix equations for correlation function, estimators of the noise variances are obtained, and under the linear minimum variance optimal information fusion criterion weighted by diagonal matrices, a self-tuning information fusion Kalman predictor is presented, which realizes the self-tuning decoupled fusion Kalman predictors for the state components. Based on the dynamic error system, a new convergence analysis method is presented for self-tuning fuser. A new concept of convergence in a realization is presented, which is weaker than the convergence with probability one. It is strictly proved that if the parameter estimation of the MA innovation models is consistent, then the self-tuning fusion Kalman predictor will converge to the optimal fusion Kalman predictor in a realization, or with probability one, so that it has asymptotic optimality. It can reduce the computational burden, and is suitable for real time applications. A simulation example for a target tracking system shows its effectiveness.

Key words Multisensor information fusion, decoupled fusion, identification, self-tuning Kalman predictor, convergence analysis

1 Introduction

The multisensor information fusion Kalman filtering has widely applied to many fields including guidance, defense, robotics, integrated navigation, target tracking, GPS positioning, communication, and signal processing, and has received great attention in recent years. It is only effective when the model parameters and noise statistics are exactly known. This restricts its practical applications. In many realistic applications, the model parameters and noise statistics are completely or partially unknown[1~3]. For the state or signals with unknown model parameters and/or noise statistics, several self-tuning filters have been presented[2~5]. Their basic principle is that the optimal filter with a recursive identifier of the autoregressive moving average (ARMA) innovation model will yield a self-tuning filter[2~5]. But, so far, the convergence of self-tuning filter has not been proved strictly, and the strict convergence analysis approach has not been presented. Recently, for the distributed fusion (weighted fusion) Kalman filters, the optimal fusion criteria weighted by matrices, diagonal matrices, and scalars have been presented in the linear minimum variance sense[6,7], where the optimal fusion criterion weighted by diagonal matrices is equivalent to the optimal fusion criterion weighted by scalars for the state components, which realizes a decoupled fused estimation for the state components in the sense that only the component estimators with the same physical sense are weighted by scalar to obtain the fused component estimator which is independent of other component estimators. The modern time series analysis method proposed by Deng, *et al*[5,6] provides an important methodology for solving optimal and self-tuning filtering problems. Its basic tool is the AMRA innovation model. Compared with the classical Kalman filtering method[8], the Riccati equation is avoided, and based on on-line identification of the ARMA innovation, the self-tuning filter can be designed[5]. In this paper, for the multisensor systems with unknown noise statistics, using the modern time series analysis method, based on on-line identification of the moving average (MA) innovation models, applying the optimal fusion criterion weighted by diagonal matrices, a self-tuning information fusion Kalman predictor is presented, which realizes the self-tuning decoupled fusion Kalman predictors for the state components. If the interest is to find the fuser for a component of the state, then the proposed decoupled fusers can avoid the computation of local and fused predictors for other components, so that the computational burden can be reduced. In addition, compared with the Kalman fuser with the matrix weights[7], the Kalman fuser with the diagonal matrix weights avoids the on-line computation of a high-dimension inverse matrix, and only requires a less computational burden. The convergence analysis of a self-tuning fuser is a very difficult open problem. A new convergence analysis method is presented in this paper. First, a new concept of convergence in a realization is presented, which is weaker than convergence with the probability one. In the second place, the general mathematical method and tool for solving the convergence problems are presented based on a dynamic error system. Thirdly, it is strictly proved for the first time that if the parameter estimation of the MA innovation models is consistent, the self-tuning fusion Kalman predictor will converge to the optimal fusion Kalman predictor in a realization, or with probability one.

Received November 9, 2005; in revised form March 22, 2006
Supported by National Natural Science Foundation of P. R. China (60374026)
1. Department of Automation, Heilongjiang Uninvesity, Harbin 150080, P. R. China
DOI: 10.1360/aas-007-0156

2 Optimal fusion steady-state Kalman predictor weighted by diagonal matrices

Consider the multisensor linear discrete-time stochastic system

$$x(t+1) = \Phi x(t) + \Gamma w(t) \tag{1}$$

$$y_i(t) = H_i x(t) + v_i(t), i = 1, \cdots, L \tag{2}$$

where $x(t) \in R^n$, $y_i(t) \in R^{m_i}$, $w(t) \in R^r$, $v_i(t) \in R^{m_i}$ are the state, measurement, process noise, and measurement noise of the ith sensor subsystem, respectively, and Φ, Γ, and H_i are constant matrices with compatible dimensions.

Assumption 1. $w(t)$ and $v_i(t)$ are uncorrelated white noises with zero mean and

$$\begin{aligned} &\mathrm{E}[w(t)w^{\mathrm{T}}(k)] = Q_w\delta_{tk}, \mathrm{E}[w(t)v_j^{\mathrm{T}}(k)] = 0 \\ &\mathrm{E}[v_i(t)v_j^{\mathrm{T}}(k)] = Q_{v_j}\delta_{ij}\delta_{tk}, \forall t, k, i, j = 1, \cdots, L \end{aligned} \tag{3}$$

where E denotes the mathematical expectation, the superscript T denotes the transpose, $\delta_{\alpha\beta}$ is the Kronecker delta function, *i.e.* $\delta_{\alpha\alpha}=1, \delta_{\alpha\beta}=0(\alpha\neq\beta)$.

Assumption 2. (Φ, H_i) is a completely observable pair with the observability index β_i, and (Φ,Γ) is a complete controllable pair, or Φ is a stable matrix (*i.e* all eigenvalues of Φ lie inside the unit circle).

Assumption 3. Φ, Γ, and H_i are known, the noise variance matrices Q_w and Q_{vi} are completely or partially unknown.

Assumption 4. The measurement data $y_i(t)$ are bounded, *i.e.* a realization of the stochastic process $y_i(t)$ is bounded, so that $\| y_i(t) \| \leq c$, $\forall t, i=1,\cdots,L$, with constant c, and $\|\cdot\|$ denotes the norm of the vetor.

From (1) and (2) we have $y_i(t) = H_i(I_n - q^{-1}\Phi)^{-1}\Gamma q^{-1}w(t)+v_i(t)$, where q^{-1} is the backward shift operator, $q^{-1}w(t)=w(t-1)$, and I_n is the $n\times n$ unit matrix. Introducing the left-coprime factorization

$$H_i(I_n-q^{-1}\Phi)^{-1}\Gamma q^{-1} = A_i^{-1}(q^{-1})B_i(q^{-1}) \tag{4}$$

where $A_i(q^{-1})$ and $B_i(q^{-1})$ are polynomial matrices having the form $X_i(q^{-1})=X_{i0}+X_{i1}q^{-1}+\cdots+X_{in_{xi}}q^{-n_{xi}}$, with $X_{in_{xi}}\neq 0, X_{ij}=0(j>n_{xi}), A_{i0}=I_{m_i}$, and $B_{i0}=0$, we obtain the local ARMA innovation models

$$A_i(q^{-1})y_i(t)=D_i(q^{-1})\varepsilon_i(t), i=1,2,\cdots,L \tag{5}$$

where $D_i(q^{-1})=D_{i0}+D_{i1}q^{-1}+\cdots+D_{in_{di}}q^{-n_{di}}$ is stable (*i.e.* all zeros of $\det D_i(t)$ lie outside the unit circle), $D_{i0}=I_{m_i}$, and the innovation process $\varepsilon_i(t)\in R^{m_i}$ is white noise with zero mean and variance matrix $Q_{\varepsilon i}$, and

$$D_i(q^{-1})\varepsilon_i(t)=B_i(q^{-1})w(t)+A_i(q^{-1})v_i(t) \tag{6}$$

where $D_i(q^{-1})$ and $Q_{\varepsilon i}$ can be obtained by the Gevers-Wouters algorithm[6].

Lemma 1. For the multisensor system (1) and (2) with known model parameters and noise statistics, the ith sensor subsystem has the local N-step-ahead steady-state Kalman predictor of $x(t+N)(N\geq 1)$ as

$$\hat{x}_i(t+N|t)=\Phi^{N-1}\hat{x}_i(t+1|t) \tag{7}$$

$$\begin{gathered}\hat{x}_i(t+1|t)=\Psi_{pi}\hat{x}_i(t|t-1)+K_{pi}y_i(t)\\ \Psi_{pi}=\Phi-K_{pi}H_i\end{gathered} \tag{8}$$

$$K_{pi}=\begin{bmatrix}H_i\\ H_i\Phi\\ \vdots\\ H_i\Phi^{\beta_i-1}\end{bmatrix}^{+}\begin{bmatrix}M_{i1}\\ M_{i2}\\ \vdots\\ M_{i,\beta_i}\end{bmatrix} \tag{9}$$

where the pseudo-inverse is defined as $X^+=(X^{\mathrm{T}}X)^{-1}X^{\mathrm{T}}$. The matrices M_{ij} can recursively be computed as

$$M_{ij}=-A_{i1}M_{i,j-1}-\cdots-A_{in_{ai}}M_{i,j-n_{ai}}+D_{ij} \tag{10}$$

where we define that $M_{ij}=0(j<0)$. Defining the steady-state local prediction error cross-covariances as $P_{ij}(N)=\mathrm{E}[\tilde{x}_i(t+N|t)\tilde{x}_j^{\mathrm{T}}(t+N|t)], i,j=1,\cdots,L$, with $\tilde{x}_i(t+N|t)=x(t+N)-\hat{x}_i(t+N|t)$, we have the relation

$$\begin{gathered}P_{ij}(N)=\\ \Phi^{N-1}\Sigma_{ij}\Phi^{(N-1)\mathrm{T}}+\sum_{k=2}^{N}\Phi^{N-k}\Gamma Q_w\Gamma^{\mathrm{T}}\Phi^{(N-k)\mathrm{T}}\\ N\geq 2\end{gathered} \tag{11}$$

with the definition $\Sigma_{ij}=P_{ij}(1)$. Σ_{ij} satisfy the Lyapunov equation

$$\begin{gathered}\Sigma_{ij}=\Psi_{pi}\Sigma_{ij}\Psi_{pj}^{\mathrm{T}}+\Gamma Q_w\Gamma^{\mathrm{T}}+K_{pi}Q_{vi}\delta_{ij}K_{pj}^{\mathrm{T}}\\ i,j=1,\cdots,L\end{gathered} \tag{12}$$

which can be solved by iteration[6]. The proof of Lemma 1 was given in [6].

Lemma 2. For the multisensor system (1) and (2) with known model parameters and noise statistics, the optimal information fusion steady-state Kalman predictor weighted by diagonal matrices is given as

$$\begin{gathered}\hat{x}_0(t+N|t)=\sum_{j=1}^{L}\Omega_j(N)\hat{x}_j(t+N|t)\\ \Omega_j(N)=\mathrm{diag}(\omega_{j1}(N),\cdots,\omega_{jn}(N))\end{gathered} \tag{13}$$

Denoting the local and optimal fused Kalman predictors in the component form as

$$\begin{gathered}\hat{x}_j(t+N|t)=[\hat{x}_{j1}(t+N|t),\cdots,\hat{x}_{jn}(t+N|t]^{\mathrm{T}}\\ \hat{x}_0(t+N|t)=[\hat{x}_{01}(t+N|t),\cdots,\hat{x}_{0n}(t+N|t)]^{\mathrm{T}}\end{gathered} \tag{14}$$

the decoupled optimal fused Kalman predictors for the state components are given by

$$\hat{x}_{0i}(t+N|t)=\sum_{j=1}^{L}\omega_{ji}(N)\hat{x}_{ji}(t+N|t), i=1,\cdots,n \tag{15}$$

where the optimal weighting coefficient vectors $\omega_i(N)=[\omega_{1i}(N),\cdots,\omega_{Li}(N)], i=1,\cdots,n$, are given by

$$\omega_i(N)=[e^{\mathrm{T}}(P^{ii}(N))^{-1}e]^{-1}e^{\mathrm{T}}(P^{ii}(N))^{-1}, i=1,\cdots,n \tag{16}$$

where $e^{\mathrm{T}}=[1,\cdots,1]$, and $P^{ii}(N)=(P_{kj}^{(ii)}(N)), k,j=1,\cdots,L$, is the $L\times L$ matrix whose (k,j) element $P_{kj}^{(ii)}(N)$ is the (i,i) diagonal element of $P_{kj}(N)$. Defining $P_{0i}(N)=\mathrm{E}[\tilde{x}_{0i}^2(t+N|t)]$ with $\tilde{x}_{0i}(t+N|t)=x_i(t+N)-\hat{x}_{0i}(t+N|t)$, and $x(t+N)=[x_1(t+N),\cdots,x_n(t+N)]^{\mathrm{T}}$, we have

$$P_{0i}(N)=[e^{\mathrm{T}}(P^{ii}(N))^{-1}e]^{-1}, i=1,\cdots,n \tag{17}$$

and we have the accuracy relation that $P_{0i}(N)\leq P_{jj}^{ii}(N), i=1,\cdots,n; j=1,\cdots,L$. The proof of Lemma 2 was given in [6].

3 Self-tuning decoupled fusion Kalman predictor weighted by diagonal matrices

Theorem 1. For multisensor system (1) and (2) with Assumptions $1\sim 3$, the unknown noise variance matrices Q_w and Q_{vi} can be obtained by solving the following matrix equations

$$\begin{gathered}\sum_{j=\tau}^{n_{di}}D_{ij}Q_{\varepsilon i}D_{i,j-\tau}^{\mathrm{T}}=\\ \sum_{j=\tau}^{n_{bi}}B_{ij}Q_wB_{i,j-\tau}^{\mathrm{T}}+\sum_{j=\tau}^{n_{ai}}A_{ij}Q_{vi}A_{i,j-\tau}^{\mathrm{T}}\\ \tau=0,1,\cdots,n_{di}, i=1,\cdots,L\end{gathered} \tag{18}$$

where A_{ij} and B_{ij} are known, and D_{ij} and $Q_{\varepsilon i}$ are assumed to be known.

Proof. Computing the correlation functions of the two sides of MA process (6) yields (18). Let θ_i denote an $n_i \times 1$ column vector which consists of all unknown elements in Q_w and Q_{vi}. For the fixed i, expanding (18)($\tau = 0, \cdots, n_{di}$) for each element of matrices, equations (18) can be rewritten as an equivalent set of linear equations

$$\Delta_i \theta_i = \delta_i, i = 1, \cdots, L \tag{19}$$

where the matrix Δ_i is known, the vetor δ_i is obtained from the elementary operations of elements of $D_{ij}(j = 1, \cdots, n_{di}$ and $Q_{\varepsilon i}$, *i.e.*, δ_i is a continuous function of elements of $D_{ij}(j = 1, \cdots, n_{di})$ and $Q_{\varepsilon i}$, which is denoted by $\delta_i = f_i(D_{i1}, \cdots, D_{in_{di}}, Q_{\varepsilon i})$. Since θ_i satisfies (19), the linear equaitons (19) have consistency. If Δ_i has full column rank, *i.e.* rank$\Delta_i = n_i$, then Δ_i has the same row rank, so that for fixed i, from (19) we can select n_i linear independent equations as

$$\Delta_{i0} \theta_i = \delta_{i0}, \delta_{i0} = f_{i0}(D_{i1}, \cdots, D_{in_{di}}, Q_{\varepsilon i}) \tag{20}$$

where Δ_{i0} is known $n_i \times n_i$ non-singular matrix, δ_{i0} is the $n_i \times 1$ vector and f_{i0} is a continuous function. From (20), θ_i can be solved as

$$\theta_i = \Delta_{i0}^{-1} \delta_{i0} \tag{21}$$

□

When only Q_w and $Q_{vi}(i = 1, \cdots, L)$ are unknown, $A_i(q^{-1})$ and $B_i(q^{-1})$ are known. Introducing the new measurement processes $z_i(t)$ as $z_i(t) = A_i(q^{-1})y_i(t)$, then (5) becomes the MA innovation models

$$z_i(t) = D_i(q^{-1})\varepsilon_i(t), i = 1, \cdots, L \tag{22}$$

The unknown MA parameter matrices D_{ij} can be estimated by a recurive identifier [9], and in the convergence analysis we assume that the MA parameter estimation is consistent, *i.e.* $\hat{D}_{ij} \to D_{ij}$, as $t \to \infty$, where $\hat{D}_{ij}$ denotes the estimate of D_{ij} at time t, and time t is omitted, *i.e.* $\hat{D}_{ij}$ means $\hat{D}_{ij}(t)$. From (22), the estimate $\hat{\varepsilon}_i(t)$ of innovation process $\varepsilon_i(t)$ at time t is defined as

$$\hat{\varepsilon}_i(t) = z_i(t) - \hat{D}_{i1}\hat{\varepsilon}_i(t-1) - \cdots - \hat{D}_{in_{di}}\hat{\varepsilon}_i(t - n_{di}) \tag{23}$$

which yields the relation

$$z_i(t) = \hat{D}_i(q^{-1})\hat{\varepsilon}_i(t) \tag{24}$$

where we define the estimate $\hat{D}_i(q^{-1})$ of $D_i(q^{-1})$ at time t as $\hat{D}_i(q^{-1}) = I_{m_i} + \hat{D}_{i1}q^{-1} + \cdots + \hat{D}_i n_{di} q^{-n_{di}}$, and define the sampled covariance estimate $\hat{Q}_{\varepsilon i}$ of $Q_{\varepsilon i}$ at time t as

$$\hat{Q}_{\varepsilon i} = \frac{1}{t}\sum_{j=1}^{t} \hat{\varepsilon}_i(j)\hat{\varepsilon}_i^{\mathrm{T}}(j) \tag{25}$$

Substituting the estimates $\hat{D}_{ij}$ and $\hat{Q}_{\varepsilon i}$ into (20) yields the estimates $\hat{\theta}_i$ and $\hat{\delta}_{i0}$ at time t as

$$\hat{\theta}_i = \Delta_{i0}^{-1}\hat{\delta}_{i0}, \hat{\delta}_{i0} = f_{i0}(\hat{D}_{i1}, \cdots, \hat{D}_{in_{di}}, \hat{Q}_{\varepsilon i}) \tag{26}$$

Hence, based on the ith subsystem, from (26) we obtain the estimates $\hat{Q}_{wi}$ and $\hat{Q}_{vi}$ of Q_w and Q_{vi} at time t. Based on all subsystems, the estimate $\hat{Q}_w$ of Q_w at time t is defined as

$$\hat{Q}_w = \frac{1}{L}\sum_{i=1}^{L} \hat{Q}_{wi} \tag{27}$$

The self-tuning decoupled fusion Kalman predictor consists of the following three steps:

Step 1. Applying a recursive identifier[9] of the MA innovation models (22), the estimates $\hat{D}_{ij}$ of D_{ij} can be obtained, and substituting $\hat{D}_{ij}$ into (23)~(27), (9)~(13) and (16) yields estimates $\hat{Q}_{\varepsilon i}$, $\hat{Q}_w$, $\hat{Q}_{vi}$, $\hat{M}_{ij}$, $\hat{K}_{pi}$, $\hat{\Psi}_{pi}$, $\hat{P}_{ij}(N)$, $\hat{\Sigma}_{ij}$, $\hat{\omega}_{ji}(N)$ and $\hat{\Omega}_j(N)$.

Step 2. From (7) and (8), the local self-tuning Kalman predictors are given as

$$\begin{aligned} \hat{x}_i^s(t+1|t) &= \hat{\Psi}_{pi}\hat{x}_i^s(t|t-1) + \hat{K}_{pi}y_i(t) \\ \hat{x}_i^s(t+N|t) &= \Phi^{N-1}\hat{x}_i^s(t+1|t), N > 1 \end{aligned} \tag{28}$$

Step 3. From (13), the self-tuning fused Kalman predictor weighted by diagonal matrices is given as

$$\hat{x}_0^s(t+N|t) = \sum_{j=1}^{L} \hat{\Omega}_j(N)\hat{x}_j^s(t+N|t) \tag{29}$$

The above three steps are repeated at each time t.

Remark 1. In order to reduce the on-line computational burden of solving the Lyapunov equations (12) with estimates $\hat{Q}_w$ and $\hat{Q}_{vi}$ by iteration, we can select a computing period (dead band) T_d of (12). In a dead band T_d, the estimates $\hat{\Sigma}_{ij}$ are not changed, so that the estimates $\hat{\Omega}_i(N)$ are also not changed in T_d.

4 The convergence analysis of self-tuning fused Kalman predictor

The known measurement data $y_i(t)$ can be viewed as a realization of the measurement stochastic process $y_i(t)$.

Definition 1. If based on known measurement data, the estimate $\hat{D}_{ij}$ of the MA parameter D_{ij} converges to the true value D_{ij}, *i.e.* $\hat{D}_{ij} \to D_{ij}$, as $t \to \infty$, then we call that the estimate $\hat{D}_{ij}$ converges to D_{ij} in a realization.

Definition 2. If the self-tuning Kalman perdictors $\hat{x}_i^s(t+N|t)$ and steady-state Kalman predictor $\hat{x}_i(t+N|t)$ obtained based on known measurement data have the relation that $[\hat{x}_i^s(t+N|t) - \hat{x}_i(t+N|t)] \to 0$, as $t \to \infty$, $i = 1, \cdots, L$, then we call that $\hat{x}_i^s(t+N|t)$ converges to $\hat{x}_i(t+N|t)$ in a realization.

Definition 3. If the self-tuning and optimal fused Kalman predictors $\hat{x}_0^s(t+N|t)$ and $\hat{x}_0(t+N|t)$ obtained based on known measurement data $y_i(t)(i = 1, \cdots, L)$ have the relation $[\hat{x}_0^s(t+N|t) - \hat{x}_0(t+N|t)] \to 0$, as $t \to \infty$, then we call that $\hat{x}_0^s(t+N|t)$ converges to $\hat{x}_0(t+N|t)$ in a realization.

Remark 2. The convergence in a realization or a realization-based convergence is weaker than the convergence with probability one. If the convergence with probability one holds, according to the statistical inference principle, for known measurement data as a realization of the measurement process, the convergence in a realization holds. Inversely, if for each realization, except the realizations with probability zero, the convergence in a realization holds, then the convergence with probability one holds. But, if the convergence in a realization holds, generally, we do not conclude whether the convergence with probability one holds.

Remark 3. The concept of convergence in a realization has an important application value. Because in many application problems, we only know a realization of a stochastic process, for example, meteorological data, hydrological data.

Lemma 3. Consider a time-varying dynamic error system

$$\delta(t) = F(t)\delta(t-1) + u(t) \tag{30}$$

where $t \geq 0$, the output (dynamic error) $\delta(t) \in R^n$, the input $u(t) \in R^n$. Assume that $F(t) \to F$, as $t \to \infty$, where F is a stable matrix, and $u(t)$ is bounded, *i.e.* $\| u(t) \| \leq c_1, t \geq 0$, with constant c_1. Then $\delta(t)$ is bounded.

Proof. See Appendix A.

Corollary 1. Assume that $\delta(t) \in R^m$ satisfies the time-varying non-homogeneous difference equation

$$\hat{\Lambda}(q^{-1})\delta(t) = u(t) \tag{31}$$

where the input $u(t) \in R^m$ is bounded, and we define that

$$\begin{aligned}\hat{\Lambda}(q^{-1}) &= I_m + \hat{\Lambda}_1(t)q^{-1} + \cdots + \hat{\Lambda}_{n_\lambda}(t)q^{-n_\lambda} \\ \Lambda(q^{-1}) &= I_m + \Lambda_1 q^{-1} + \cdots + \Lambda_{n_\lambda} q^{-n_\lambda}\end{aligned} \tag{32}$$

Assume that $\hat{\Lambda}_i(t) \to \Lambda_i$, as $t \to \infty$, $i = 1, \cdots, n_\lambda$, and $\Lambda(q^{-1})$ is stable. Then the output $\delta(t)$ also is bounded.

Proof. See Appendix A.

Lemma 4. Consider a stable dynamic error system

$$\delta(t) = F\delta(t-1) + u(t) \tag{33}$$

where $t \geq 0, \delta(t) \in R^n$, $u(t) \in R^n$. Assume that F is a stable matrix, and $u(t) \to 0$ as $t \to \infty$. Then $\delta(t) \to 0$, as $t \to \infty$.

Proof. See Appendix A.

Corollary 2. Consider a dynamic error system described by a stable non-homogenous different equation

$$\Lambda(q^{-1})\delta(t) = u(t) \tag{34}$$

where $u(t) \in R^m$, $\delta(t) \in R^m$, $\Lambda(q^{-1})$ as defined in (32) is stable, and $u(t) \to 0$, as $t \to \infty$. Then $\delta(t) \to 0$, as $t \to \infty$.

Proof. See Appendix A.

Theorem 2. For the multisensor system (1) and (2) with Assumptions $1 \sim 4$, if

$$\begin{aligned}&\hat{D}_{ij} \to D_{ij}, \text{ as } t \to \infty, \\ &\text{in a realization}, i = 1, \cdots, L, j = 1, \cdots, n_{di}\end{aligned} \tag{35}$$

then the innovation estimator $\hat{\varepsilon}_i(t)$ defined in (23) converges to $\varepsilon_i(t)$ in the sense that

$$\begin{aligned}&[\hat{\varepsilon}_i(t) - \varepsilon_i(t)] \to 0, \text{as } t \to \infty \\ &i = 1, \cdots, L, \text{in a realization}\end{aligned} \tag{36}$$

Proof. Setting $\hat{D}_{ij} = D_{ij} + \Delta\hat{D}_{ij}$, $\hat{D}_i(q^{-1}) = D_i(q^{-1}) + \Delta\hat{D}_i(q^{-1})$, $\Delta\hat{D}_i(q^{-1}) = \Delta\hat{D}_{i1}q^{-1} + \cdots + \Delta\hat{D}_{in_{di}}q^{-n_{di}}$, from (35) we have that $\Delta\hat{D}_{ij} \to 0$, $\hat{D}_i(q^{-1}) \to D_i(q^{-1})$, and $\Delta\hat{D}_i(q^{-1}) \to 0$, as $t \to \infty$. Defining $\delta_i(t) = \hat{\varepsilon}_i(t) - \varepsilon_i(t)$, and subtracting (22) from (24) yields the dynamic error system

$$D_i(q^{-1})\delta_i(t) = u_i(t), u_i(t) = -\Delta\hat{D}_i(q^{-1})\hat{\varepsilon}_i(t) \tag{37}$$

From Assumption 4 we have that $z_i(t)$ is bounded. Applying (24), the stability of $D_i(q^{-1})$ and Corollary 1 yields that $\hat{\varepsilon}_i(t)$ is bounded. Hence from (37) we have that $u_i(t) \to 0$, as $t \to \infty$. From (37), applying Corollary 2 and the stability of $D_i(q^{-1})$ yields that $\delta_i(t) \to 0$, as $t \to \infty$, *i.e.*(36) holds. □

Theorem 3. For the multisensor system (1) and (2) with Assumptions $1 \sim 4$, if (35) holds, we have

$$\hat{Q}_{\varepsilon i} \to Q_{\varepsilon i}, \text{as } t \to \infty, \text{in a realization} \tag{38}$$

Proof. Since white noise $\varepsilon_i(t)$ is a stationary stochastic process, according to the ergodicity[9], we have

$$\frac{1}{t}\sum_{j=1}^{t}\varepsilon_i(j)\varepsilon_j^{\mathrm{T}}(j) \to Q_{\varepsilon i}, \text{as } t \to \infty, \text{with probability 1} \tag{39}$$

According to the statistical inference principle, we conclude that (39) holds in a realization.

Applying (25) and $\delta_i(t) = \hat{\varepsilon}_i(t) - \varepsilon_i(t)$, we have

$$\begin{aligned}&\hat{Q}_{\varepsilon i} - Q_{\varepsilon i} = \\ &\frac{1}{t}\sum_{j=1}^{t}[\varepsilon_i(j)\delta_i^{\mathrm{T}}(j) + \delta_i(j)\varepsilon_i^{\mathrm{T}}(j) + \delta_i(j)\delta_i^{\mathrm{T}}(j)] + \\ &\frac{1}{t}\sum_{j=1}^{t}\varepsilon_i(j)\varepsilon_i^{\mathrm{T}}(j) - Q_{\varepsilon i}\end{aligned} \tag{40}$$

Notice that $\| \varepsilon_i(t) \| \leq \| \delta_i(t) \| + \| \hat{\varepsilon}_i(t) \|$. Since $\delta_i(t) \to 0$, as $t \to \infty$, $\delta_i(t)$ is bounded. Hence applying the boundedness of $\hat{\varepsilon}_i(t)$ yields that $\varepsilon_i(t)$ is bounded. Therefore, we have that $[\varepsilon_i(j)\delta_i^{\mathrm{T}}(j) + \delta_i(j)\varepsilon_i^{\mathrm{T}}(j) + \delta_i(j)\delta_i^{\mathrm{T}}(j)] \to 0$, as $j \to \infty$, and from (39) and (40) we have that (38) holds in a realization. □

Theorem 4. For the multisensor system (1) and (2) with Assumptions $1 \sim 4$, if (35) holds, then

$$\begin{aligned}&\hat{Q}_w \to Q_w, \hat{Q}_{vi} \to Q_{vi}, \hat{K}_{pi} \to K_{pi}, \hat{\Psi}_{pi} \to \Psi_{pi}, \\ &\hat{\Omega}_i(N) \to \Omega_i(N), \text{as } t \to \infty, \text{in a realization}\end{aligned} \tag{41}$$

Proof. Since f_{i0} is a continuous function of the elements of $D_{ij} (i = 1, \cdots, n_{di})$ and $Q_{\varepsilon i}$, from (20) and (26), applying (35) and (38) yields that $\hat{\theta}_i \to \theta_i$, *i.e.* $\hat{Q}_{wi} \to Q_w$, $\hat{Q}_{vi} \to Q_{vi}$, and applying (27) yields that $\hat{Q}_w \to Q_w$ in a realization. From (10), each element of M_{ij} is a continuous function of elements of $D_{ij} (j = 1, \cdots, n_{di})$, applying (35) yields that $\hat{M}_{ij} \to M_{ij}$ in a realization. From (9), each element of K_{pi} is a continuous function of elements of $M_{ij} (j = 1, \cdots, \beta_i)$, so that $\hat{K}_{pi} \to K_{pi}$, and $\hat{\Psi}_{pi} \to \Psi_{pi}$ in a realization. For the Lyapunov equation (12), applying the existence theorem for implicit function[10], in a sufficiently small neighborhood, each element of Σ_{ij} is a continuous function of elements of $\Psi_{pi}, \Psi_{pj}, K_{pi}, K_{pj}, Q_w$ and Q_{vi}, so that $\hat{\Sigma}_{ij} \to \Sigma_{ij}$, and from (11), $\hat{P}_{ij}(N) \to P_{ij}(N)$ in a realization. From (13) and (16), each element of $\Omega_i(N)$ is a continuous function of elements of $P_{kj}(N)$, so that $\hat{\Omega}_i(N) \to \Omega_i(N)$ in a realization. □

Theorem 5. For the multisensor system (1) and (2) with Assumptions $1 \sim 4$, if (35) holds, then the local self-tuning Kalman predictor $\hat{x}_i^s(t+N|t)$ converges to the local steady-state Kalman predictor $\hat{x}_i(t+N|t)$ in the sense that

$$\begin{aligned}&[\hat{x}_i^s(t+N|t) - \hat{x}_i(t+N|t)] \to 0 \\ &\text{as } t \to \infty, \text{in a realization}\end{aligned} \tag{42}$$

Proof. Setting $\hat{K}_{pi} = K_{pi} + \Delta\hat{K}_{pi}$, $\hat{\Psi}_{pi} = \Psi_{pi} + \Delta\hat{\Psi}_{pi}$, from (41) we have that $\Delta\hat{K}_{pi} \to 0, \Delta\hat{\Psi}_{pi} \to 0$ in a realization. Denoting $\delta_i(t) = \hat{x}_i^s(t+1|t) - \hat{x}_i(t+1|t)$, subtracting

(8) from (28) yields a dynamic error system

$$\begin{aligned}\delta_i(t) &= \Psi_{pi}\delta_i(t-1) + u_i(t)\\ u_i(t) &= \Delta\hat{\Psi}_{pi}\hat{x}_i^s(t|t-1) + \Delta\hat{K}_{pi}y_i(t)\end{aligned} \tag{43}$$

In (28), applying Assumption 4 and $\hat{K}_{pi} \rightarrow K_{pi}$ yields that $\hat{K}_{pi}y_i(t)$ is bounded, and noting that $\hat{\Psi}_{pi} \rightarrow \Psi_{pi}$, and Ψ_{pi} is a stable matrix[8], according to Lemma 3, $\hat{x}_i^s(t+N|t)$ is bounded in a realization. Hence $u_i(t) \rightarrow 0$ in (43), and applying Lemma 4 yields that $\delta_i(t) \rightarrow 0$, and from (8) and (28) we have that (42) holds. □

Theorem 6. For the multisensor system (1) and (2) with Assumptions 1 $\sim$ 4, if (35) holds, then the self-tuning fused Kalman predictor weighted by diagonal matrices converges to the optimal fused Kalman predictor weighted by diagonal matrices in the sense that

$$\begin{aligned}&[\hat{x}_0^s(t+N|t) - \hat{x}_0(t+N|t)] \rightarrow 0\\ &\text{as } t \rightarrow \infty, \text{in a realization}\end{aligned} \tag{44}$$

Proof. Setting $\hat{\Omega}_i(N) = \Omega_i(N) + \Delta\hat{\Omega}_i(N)$, from (41) we have that $\Delta\hat{\Omega}_i(N) \rightarrow 0$. Subtracting (13) from (29) yields

$$\begin{aligned}&\hat{x}_0^s(t+N|t) - \hat{x}_0(t+N|t) =\\ &\sum_{i=1}^{L}\Omega_i(N)[\hat{x}_i^s(t+N|t) - \hat{x}_i(t+N|t)] +\\ &\sum_{i=1}^{L}\Delta\hat{\Omega}_i(N)\hat{x}_i^s(t+N|t)\end{aligned} \tag{45}$$

Applying (42), $\Delta\hat{\Omega}_i(N) \rightarrow 0$, and boundedness of $\hat{x}_i^s(t+N|t)$ yields that (44) holds. □

Theorem 7. For the multisensor system (1) and (2) with Assumptions 1 $\sim$ 3, if $\hat{D}_{ij}$ converges to D_{ij} with probability one, and the measurement processes $y_i(t)(i=1,\cdots,L)$ are bounded with probability one, then the self-tuning fused Kalman predictor converges to the optimal fused Kalman predictor with probability one, *i.e.*,

$$[\hat{x}_0^s(t+N|t) - \hat{x}_0(t+N|t)] \rightarrow 0, \text{as } t \rightarrow \infty, \text{w.p.1} \tag{46}$$

Proof. From Theorem 6, except the realizations with probability zero, for each realization of the measurement processes $y_i(t)(i=1,\cdots,L)$, (44) holds which yields that (46) holds. □

Remark 4. Theorem 7 shows that the problem of the stochastic convergence with probability one can be converted into the problem of the non-stochastic or determinate convergence in a realization.

Remark 5. From the point of view of methodology, the problem of the convergence in a realization is converted into the stability problems of dynamic error system: the stability of bounded input to bounded output, and the stability of infinite small input to infinite small output. The convergence problem in a realization is essentially a determinate (non-stochastic) limit problem which can easily be solved by a strict mathematical tool as shown in Lemmas 3 $\sim$ 4, and Corollaries 1 $\sim$ 2.

5 Simulation example

Consider the target tracking system (1) and (2) with 3 sensors, and

$$\begin{aligned}&\Phi = \begin{bmatrix}1 & T_0 & 0.5T_0^2\\ 0 & 1 & T_0\\ 0 & 0 & 1\end{bmatrix}, \Gamma = \begin{bmatrix}0\\0\\1\end{bmatrix}, H_i = \begin{bmatrix}1 & 0 & 0\\ 0 & 1 & 0\end{bmatrix}\\ &i = 1, 2, 3\end{aligned} \tag{47}$$

where T_0 is the sampling period, $x(t)=[x_1(t), x_2(t), x_3(t)]^{\mathrm{T}}$ is the state, the components $x_1(t)$, $x_2(t)$ and $x_3(t)$ are the position, velocity, and acceleration of target at the sample time tT_0, respectively, $n=3$, $L=3$, $m_i=2$, $w(t)$ and $v_i(t)$ are independent Gaussian white noises with zero mean and unknown variances σ_w^2 and Q_{vi}, respectively. The problem is to find the self-tuning fused Kalman predictor weighted by diagonal matrices, $\hat{x}_0^s(t+2|t)$. In simulation we take $T_0 = 1$, $\sigma_w^2 = 0.36$, $Q_{v1} = \sigma_{v1}^2 I_2, Q_{v2} = \sigma_{v2}^2 I_2$, $Q_{v3} = \sigma_{v3}^2 I_2, \sigma_{v1}^2 = 0.01$, $\sigma_{v2}^2 = 0.02$, $\sigma_{v3}^2 = 0.04$. Introducing the left-coprime factorization (4) yields the MA innovation models of subsystems as[11]

$$\begin{aligned}&z_i(t) = (I_2 + D_{i1}q^{-1} + D_{i2}q^{-1})\varepsilon_i(t)\\ &z_i(t) =\\ &\left\{\begin{bmatrix}1 & 0\\ 0 & 1\end{bmatrix} + \begin{bmatrix}-1 & -1.5T_0\\ 0 & -2\end{bmatrix}q^{-1} + \begin{bmatrix}0 & 0.5T_0\\ 0 & 1\end{bmatrix}q^{-2}\right\} y_i(t)\end{aligned} \tag{48}$$

The simulation results are shown in Fig.1 $\sim$ Fig.13. The convergence of the estimates D_{ij} obtained by a modified recursive extended least squares (MRELS) method[11] is shown in Fig.1 $\sim$ Fig.6, where the curved lines denote the estimates, the straight lines denote the true values. The convergence of estimates $\hat{\sigma}_w^2$ and $\hat{Q}_{vi} = \hat{\sigma}_{vi}^2 I_2$ is shown in Fig.7 $\sim$ Fig.9. The error curves between the self-tuning and optimal fused Kalman predictors are shown in Fig.10 $\sim$ Fig.12, where we see that the self-tuning fused Kalman predictor converges to the optimal fused Kalman predictor, so that it has the asymptotic optimality. The curves of accumulated error squares for the self-tuning local and fused Kalman predictors are shown in Fig.13, where we see that the accuracy of the self-tuning fused Kalman predictor is higher than that of each local self-tuning Kalman predictor.

6 Conclusion

For the multisensor system with unknown noise statistics, by the modern time series analysis method, a self-tuning information fusion Kalman predictor weighted by diagonal matrices has been presented based on on-line identification of the MA innovation models. It has been realized the self-tuning decoupled Kalman predictors for state components. The estimators of the noise variances are obtained by solving the matrix equations for correlation function. A

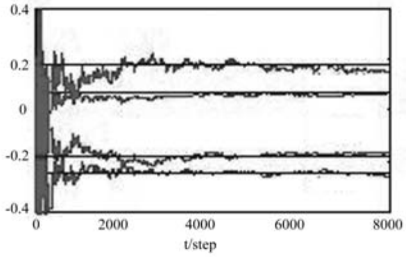


Fig. 1 The convergence of MA parameter estimate $\hat{D}_{11}$ for sensor 1

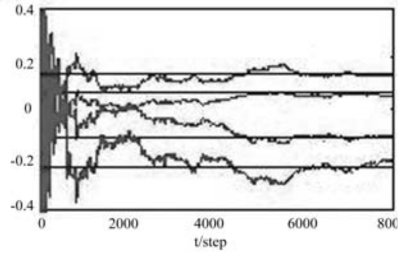


Fig. 2 The convergence of MA parameter estimate $\hat{D}_{12}$ for sensor 1

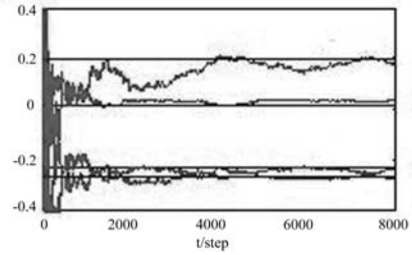


Fig. 3 The convergence of MA parameter estimate $\hat{D}_{21}$ for sensor 2

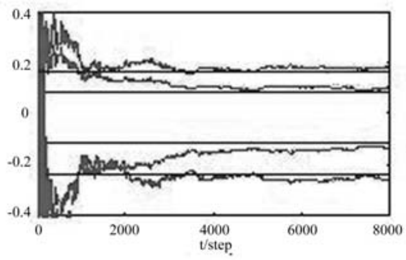


Fig. 4 The convergence of MA parameter estimate $\hat{D}_{22}$ for sensor 2

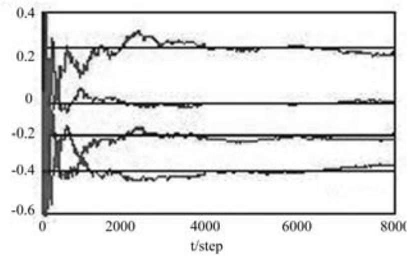


Fig. 5 The convergence of MA parameter estimate $\hat{D}_{31}$ for sensor 3

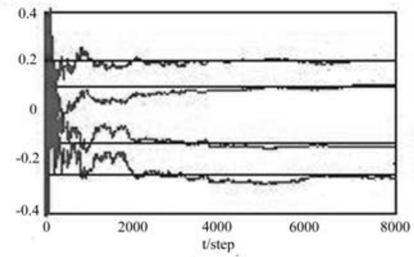


Fig. 6 The convergence of MA parameter estimate $\hat{D}_{32}$ for sensor 3

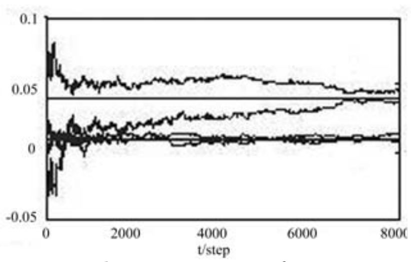


Fig. 7 The convergence of estimate $\hat{\sigma}^2_{vi}$,$i = 1, 3$ for sensor i

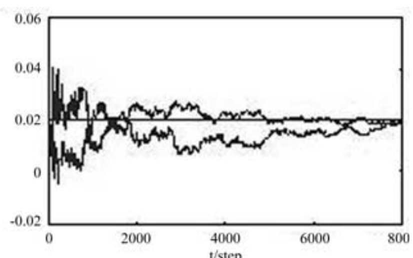


Fig. 8 The convergence of estimate $\hat{\sigma}^2_{v2}$ for sensor 2

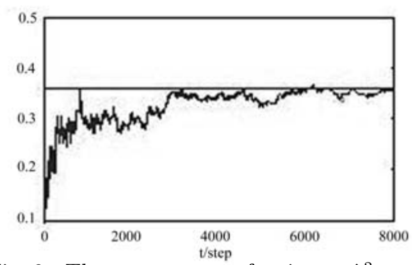


Fig. 9 The convergence of estimate $\hat{\sigma}^2_{w}$

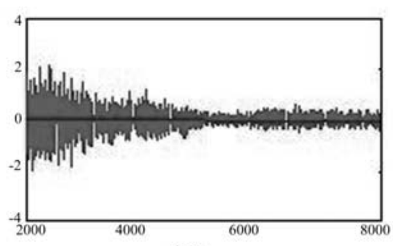


Fig. 10 The error curve between self-tuning and optimal fused Kalman position predictors

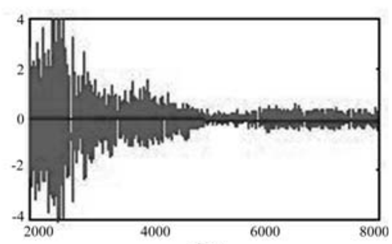


Fig. 11 The error curve between self-tuning and optimal fused Kalman velocity predictors

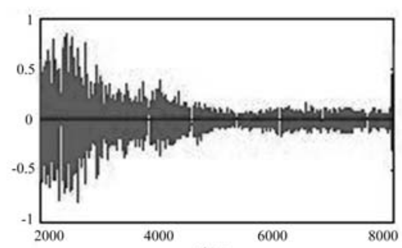
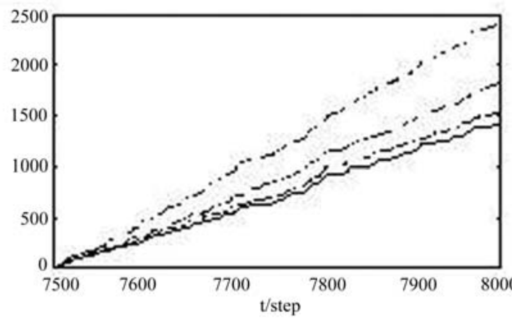
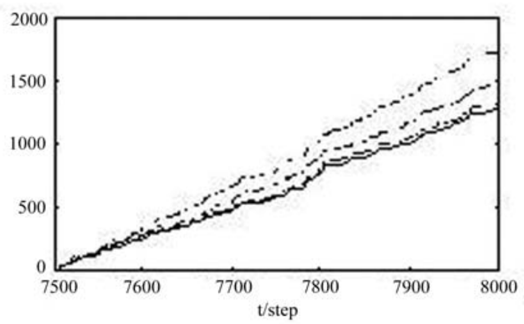


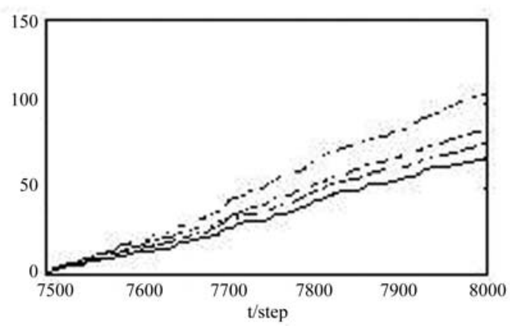
Fig. 12 The error curve between self-tuning and optimal fused Kalman acceleration predictors



(a) The curves of accumulated error squares of local and fused self-tuning Kalman position predictors



(b) The curves of accumulated error squares of local and fused self-tuning Kalman velocity predictors



(c) The curves of accumulated error squares of local and fused self-tuning Kalman acceleration predictors

– – – – – : The curves of accumulated error squares of self-tuning Kalman predictor for sensor 1
– · – · – · : The curves of accumulated error squares of self-tuning Kalman predictor for sensor 2
· · – · · – : The curves of accumulated error squares of self-tuning Kalman predictor for sensor 3
———— : The curves of accumulated error squares of self-tuning Kalman fused predictor weighted by diagonal matrices

Fig. 13 The curves of accumulated error squares of position, velocity, and acceleration for local and fused self-tuning Kalman predictors

new concept of convergence in a realization is presented, which is weaker than the convergence with probability one. The new convergence analysis method and tool based on the dynamic error systems have been presented. It has been proved strictly that the self-tuning fused Kalman predictor converges to the optimal fused Kalman predictor in a realization or with probability one, so that it has the asymptotic optimality. The proposed self-tuning fuser and its convergence theory, method, and tool open up a new field − self-tuning information fusion filtering theory and applications for systems with unknown model parameters and noise statistics.

Acknowledgement

The authors wish to thank the reviewers for their constructive comments.

References

1 Mehra R K. On the identification of variances and adaptive Kalman filtering. *IEEE Transactions on Automatic Control*, 1970, **15**(2): 175 ∼ 184

2 Hagander P, Wittenmark B. A self-tuning filter for fixed-lag smoothing. *IEEE Transactions on Informaiton Theory*, 1977, **23**(3): 377 ∼ 384

3 Moir T, Grimble M. Optimal self-tuning filtering, prediction, and smoothing for discrete multivariable processes. *IEEE Transactions on Automatic Control*, 1984, **29**(2): 128 ∼ 137

4 Deng Zi-Li, Li Bei-Xin. A self-tuning α-β tracking filter. *Acta Automatica Sinica*, 1992, **18**(6): 720 ∼ 723

5 Deng Zi-Li, Zhang Huan-Shui, Liu Shu-Jun, Zhou L. Optimal and self-tuning white noise estimators with applications to deconvolution and filtering problems. *Automatica*, 1996, **32**(2): 199 ∼ 216

6 Deng Zi-Li, Gao Yuan, Mao Lin, Li Yun, Hao Gang. New approach to information fusion steady-state Kalman filtering. *Automatica*, 2005, **41**(10):1695 ∼ 1707

7 Sun Shu-Li, Deng Zi-Li. Multi-sensor optimal information fusion Kalman filter. *Automatica*, 2004, **40**(6): 1017 ∼ 1023

8 Kailath T, Sayed A H, Hassibi B. *Linear Estimation*. Upper Saddle River, New Jersey: Prentice-Hall, 2000. 311∼405

9 Ljung L. *System Identification−Theory for the User*, 2nd Edition. Prentice Hall PTR, Beijing: Tsinghua University Press, 2002. 361∼398

10 Chen Chuan-Zhang. *Mathematical Analysis*. Shanghai: Shanghai Science Technology Press, 1962. 276∼296

11 Deng Zi-Li. *Optimal Estimation Theory with Applications–Modeling, Filtering, and Information Fusion Estimation*. Harbin: Harbin Institute of Technology Press, 2005

12 Cheng Yun-Peng. *Matrix Theory*, 2nd Edition. Xi′an: Northwestern Polytechnical University Press, 2001. 132∼138

DENG Zi-Li Professor of Heilongjiang University. His research interests include state estimation, signal processing, time series analysis, multisensor information fusion, identification, and self-tuning filtering. Corresponding author of this paper. E-mail: dzl@hlju.edu.cn

LI Chun-Bo Received her master degree from Heilongjiang University in 2006. Her research interests include multisensor information fusion, optimal and self-tuning Kalman filtering, and identification.

Appendix A

Proof of Lemma 3. Let $\lambda_1, \cdots, \lambda_n$ be the eigenvalues of F. Since F is a stable matrix, then $|\lambda_i| < 1, i = 1, \cdots, n$, so that its spectral radius $\rho = \max(|\lambda_1|, \cdots, |\lambda_n|) < 1$. Applying the matrix theory[12], there exists a matrix norm $\|\cdot\|$ such that $\|F\| = \rho + \mu = \rho_0 < 1, \mu > 0$. Setting $F(t) = F + \Delta F(t)$, from $F(t) \rightarrow F$ we have that $\Delta F(t) \rightarrow 0$, as $t \rightarrow \infty$. Hence taking $\alpha > 0$ such that $\alpha < 1 - \rho_0$, there exists $t_0 > 0$ such that $\|\Delta F(t)\| < \alpha$, as $t > t_0$. Defining $\rho_m = \rho_0 + \alpha$, we have that $0 < \rho_m < 1$, and $\|F(t)\| \leq \|F\| + \|\Delta F(t)\| < \rho_0 + \alpha = \rho_m < 1$, as $t > t_0$. By the iteration for (30), we obtain the relation

$$\delta(t) = F(t, t_0)\delta(t_0) + \sum_{i=t_0+1}^{t} F(t, i)u(i) \tag{A1}$$

with the definitions that $F(t, t) = I_n$, $F(t, i) = F(t)F(t-1)\cdots F(i+1), t > i$. Thus $\|F(t, i)\| \leq \|F(t)\| \cdots \|F(i+1)\| < \rho_m^{t-i}$, and

$$\begin{aligned} \|\delta(t)\| \leq \|F(t, t_0)\| \|\delta(t_0)\| + \sum_{t=t_0+1}^{t} \|F(t, i)\| \|u(i)\| \leq \\ \rho_m^{t-t_0} \|\delta(t_0)\| + \sum_{i=t_0+1}^{t} \rho_m^{t-i} c_1 \end{aligned} \tag{A2}$$

Noting that $0 < \rho_m < 1$, we have that $0 < \rho_m^{t-t_0} < 1$, and

$$\sum_{i=t_0+1}^{t} \rho_m^{t-i} = \sum_{j=0}^{t-t_0-1} \rho_m^j = \frac{1 - \rho_m^{t-t_0}}{1 - \rho_m} < \frac{1}{1 - \rho_m} \tag{A3}$$

From (A2) and (A3) we easily yield the boundedness of $\delta(t)$. □

Proof of Corollary 1. The different equation (31) has the state space model

$$x(t) = \hat{\Lambda}(t)x(t-1) + b(t) \tag{A4}$$

with the definitions

$$x(t) = \begin{bmatrix} \delta(t) \\ \delta(t-1) \\ \vdots \\ \delta(t - n_\lambda + 1) \end{bmatrix}, \hat{\Lambda}(t) = \begin{bmatrix} -\hat{\Lambda}_1(t) & \cdots & \cdots & -\hat{\Lambda}_{n_\lambda}(t) \\ I_m & & & 0 \\ \vdots & & & \vdots \\ 0 & \cdots & I_m & 0 \end{bmatrix}$$

$$b(t) = \begin{bmatrix} u(t) \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \Lambda = \begin{bmatrix} -\Lambda_1 & \cdots & \cdots & -\Lambda_{n_\lambda} \\ I_m & & & 0 \\ \vdots & & & \vdots \\ 0 & \cdots & I_m & 0 \end{bmatrix} \tag{A5}$$

Since $u(t)$ is bounded, then $b(t)$ is bounded. Noting[11] that $\det(I - q^{-1}\Lambda) = \det\Lambda(q^{-1})$, where $\Lambda(q^{-1})$ is defined in (32), the stability of $\Lambda(q^{-1})$ yields that Λ is a stable matrix. From $\Lambda_i(t) \rightarrow \Lambda_i$, we have that $\hat{\Lambda}(t) \rightarrow \Lambda$, as $t \rightarrow \infty$. Applying Lemma 3 for (A4) yields that $x(t)$ is bounded, and from (A5), $\delta(t)$ is bounded. □

Proof of Lemma 4. Iterating (33) yields the relation

$$\delta(t) = F^t\delta(0) + \sum_{j=0}^{t-1} F^j u(t-j) \tag{A6}$$

Because F is a stable matrix, it has the spectral radius ρ, $0 \leq \rho < 1$, and there exists a matrix norm $\|\cdot\|$ such that $\|F\| = \rho + \mu = \rho_0 < 1, \mu > 0$. From (A6) it follows that

$$\|\delta(t)\| \leq \rho_0^t \|\delta(0)\| + \sum_{j=0}^{t-1} \rho_0^j \|u(t-j)\| \tag{A7}$$

Noting that $0 < \rho_0 < 1$, it follows that $\rho_0^t \rightarrow 0$, as $t \rightarrow \infty$. Applying the assumption that $u(t) \rightarrow 0$, as $t \rightarrow \infty$, we have that $\|u(t)\| \rightarrow 0$, as $t \rightarrow \infty$. Hence for arbitrarily small $\varepsilon > 0$, there exist t_ρ and t_μ such that $\rho_0^t < \varepsilon$, as $t > t_\rho$, $\|u(t)\| < \varepsilon$, as $t > t_u$. Here ρ_0 is a fixed number, so that t_ρ only depends on ε. Consider the decomposition

$$\begin{aligned} \sum_{j=0}^{t-1} \rho_0^j \|u(t-j)\| = \\ \sum_{j=0}^{t_\rho} \rho_0^j \|u(t-j)\| + \sum_{j=t_\rho+1}^{t-1} \rho_0^j \|u(t-j)\| \end{aligned} \tag{A8}$$

It is obvious that $u(t)$ is bounded, *i.e.* $\|u(t)\| \leq c_2$ with constant c_2, $t \geq 0$. When $t > t_\delta = t_u + t\rho$, we have that

$$\begin{aligned} \sum_{j=0}^{t_\rho} \rho_0^j \|u(t-j)\| < \varepsilon \sum_{j=0}^{t_\rho} \rho_0^j = \frac{\varepsilon(1 - \rho_0^{t_\rho+1})}{1 - \rho_0} < \frac{\varepsilon}{1 - \rho_0} \\ \sum_{j=t_\rho+1}^{t-1} \rho_0^j \|u(t-j)\| < c_2 \sum_{j=t_\rho+1}^{t-1} \rho_0^j = \\ \frac{c_2\rho_0^{t_\rho+1}(1 - \rho_0^{t-t_\rho-1})}{1 - \rho_0} < \frac{c_2\varepsilon}{1 - \rho_0} \end{aligned} \tag{A9}$$

Hence taking $t > t_\delta$, from $(A7) \sim (A9)$ we have that $\|\delta(t)\| \leq c_3\varepsilon$, with constant $c_3 = \|\delta(0)\| + [(1 + c_2)/(1 - \rho_0)]$ which is independent of ε. Since $\varepsilon > 0$ can be taken as an arbitarily small number, then $\|\delta(t)\|$ can be arbitarily small for sufficiently large t, $t > t_\delta$, *i.e.* $\delta(t) \rightarrow 0$, as $t \rightarrow \infty$. □

Proof of Corollary 2. The difference equation (34) has the state space model

$$x(t) = \Lambda x(t-1) + b(t) \tag{A10}$$

where $x(t)$, $b(t)$ and Λ are defined in (A5). The stability of $\Lambda(q^{-1})$ and the relation[11] $\det(I - q^{-1}\Lambda) = \det\Lambda(q^{-1})$ yield that Λ is a stable matrix. The assumption that $u(t) \rightarrow 0$, as $t \rightarrow \infty$, and (A5) yield that $b(t) \rightarrow 0$, as $t \rightarrow \infty$. Applying Lemma 4 for (A10) yields that $x(t) \rightarrow 0$, as $t \rightarrow \infty$, and from (A5) we have that $\delta(t) \rightarrow 0$, as $t \rightarrow \infty$. □