Identification of Bilinear Forms with the Kalman Filter

Laura Dogariu\textsuperscript{1,3}, Constantin Paleologu\textsuperscript{1}, Silviu Ciochină\textsuperscript{1}, Jacob Benesty\textsuperscript{2}, and Pablo Piantanida\textsuperscript{3}

\textsuperscript{1} University Politehnica of Bucharest, Romania
\textsuperscript{2} INRS-EMT, University of Quebec, Montreal, Canada
\textsuperscript{3} L2S, CentraleSupélec, F-91192 Gif-sur-Yvette, France

e-mail: ldogariu@comm.pub.ro

April 20, 2018
Outline

1. Introduction and Motivation
2. System Model
3. Kalman Filter for Bilinear Forms (KF - BF)
4. Simplified Kalman Filter for Bilinear Forms (SKF - BF)
5. Practical Considerations
6. Simulation Results
7. Conclusions
Kalman filter - used before in system identification
→ reference (desired) signal:

\[ d(t) = h^T(t)x(t) + v(t) \]

\( h \) - unknown system of length \( L \)
\( x(t) = [x(t)x(t-1)\ldots x(n-L+1)]^T \) - input signal
\( v(t) \) - system noise
Kalman filter - used before in system identification
→ reference (desired) signal:
\[ d(t) = h^T(t)x(t) + v(t) \]
\( h \) - unknown system of length \( L \)
\( x(t) = [x(t)x(t-1)\ldots x(n-L+1)]^T \) - input signal
\( v(t) \) - system noise

Our approach - identification of **bilinear forms**
→ reference (desired) signal:
\[ d(t) = h^T(t)X(t)g(t) + v(t) \]
\( h, g \): unknown systems of lengths \( L, M \)
\( X(t) = [x_1(t) x_2(t) \ldots x_M(t)] \) - input signal matrix
\( x_m(t) = [x_m(t) x_m(t-1) \ldots x_m(t-L+1)]^T \quad m = 1, 2, \ldots, M \)
Motivation

- **Target:** a Kalman algorithm for the identification of bilinear forms
- Timely topic
- Numerous applications:
  - nonlinear acoustic echo cancellation
  - identification of Hammerstein systems
  - tensor algebra - Big Data
Signal model:
\[ d(t) = h^T(t)X(t)g(t) + v(t) \]
→ bilinear form with respect to the impulse responses
System Model

- **Signal model:**
  \[ d(t) = h^T(t)X(t)g(t) + v(t) \]
  → bilinear form with respect to the impulse responses

- **System impulse responses:**
  \[
  h(t) = h(t-1) + w_h(t) \\
  g(t) = g(t-1) + w_g(t)
  \]
  \[ w_h(t), w_g(t): \text{zero-mean WGN} \]
  \[
  R_{w_h}(t) = \sigma_{w_h}^2 I_L \\
  R_{w_g}(t) = \sigma_{w_g}^2 I_M
  \]
System Model

- Signal model:
  \[ d(t) = h^T(t)X(t)g(t) + v(t) \]
  \( \rightarrow \) bilinear form with respect to the impulse responses

- System impulse responses:
  \[ h(t) = h(t - 1) + w_h(t) \quad g(t) = g(t - 1) + w_g(t) \]
  \( w_h(t), w_g(t) \): zero-mean WGN
  \[ R_{w_h}(t) = \sigma_{w_h}^2 \bm{I}_L \quad R_{w_g}(t) = \sigma_{w_g}^2 \bm{I}_M \]

- Equivalent model:
  \[ d(t) = f^T(t)\tilde{x}(t) + v(t) \]
  \[ f(t) = g(t) \otimes h(t) \] - Kronecker product of length \( ML \)
  \[ \tilde{x}(t) = \text{vec}[X(t)] = [x_1^T(t) \quad x_2^T(t) \quad \ldots \quad x_M^T(t)]^T \]
Scaling Ambiguity

\[ f(t) = g(t) \otimes h(t) = [\eta g(t)] \otimes \left[ \frac{1}{\eta} h(t) \right] \quad \eta \in \mathcal{R}^* \text{ - scaling factor} \]

\[
\left[ \frac{1}{\eta} h(t) \right]^T X(t) [\eta g(t)] = h^T(t)X(t)g(t)
\]
Scaling Ambiguity

\[ f(t) = g(t) \otimes h(t) = [\eta g(t)] \otimes \left[ \frac{1}{\eta} h(t) \right] \quad \eta \in \mathbb{R}^* \text{ - scaling factor} \]

\[
\left[ \frac{1}{\eta} h(t) \right]^T X(t) [\eta g(t)] = h^T(t) X(t) g(t) \quad \Rightarrow \quad \hat{h}(t) \rightarrow \frac{1}{\eta} h(t) \]
\[
\hat{g}(t) \rightarrow \eta g(t) \]
\[
\hat{f}(t) \rightarrow f(t) \]
Scaling Ambiguity

- \( f(t) = g(t) \otimes h(t) = [\eta g(t)] \otimes \left[ \frac{1}{\eta} h(t) \right] \quad \eta \in \mathcal{R}^* \) - scaling factor

- Normalized projection misalignment (NPM) \(^1\)

\[
\text{NPM}[h(t), \hat{h}(t)] = 1 - \left[ \frac{h^T(t)\hat{h}(t)}{||h(t)|| ||\hat{h}(t)||} \right]^2
\]

\[
\text{NPM}[g(t), \hat{g}(t)] = 1 - \left[ \frac{g^T(t)\hat{g}(t)}{||g(t)|| ||\hat{g}(t)||} \right]^2
\]

- Normalized misalignment (NM)

\[
\text{NM}[f(t), \hat{f}(t)] = \frac{||f(t) - \hat{f}(t)||^2}{||f(t)||^2}
\]

\(^1\) [Morgan et al., IEEE Signal Processing Letters, July 1998]
Kalman Filter for Bilinear Forms (KF - BF)

- estimated output signal: $\hat{y}(t) = \hat{h}^T(t - 1)X(t)\hat{g}(t - 1)$
Kalman Filter for Bilinear Forms (KF - BF)

- estimated output signal: \( \hat{y}(t) = \hat{h}^T(t-1)X(t)\hat{g}(t-1) \)
- error signal:

\[
\begin{align*}
e(t) &= d(t) - \hat{f}^T(t-1)\tilde{x}(t) \\
&= d(t) - \hat{h}^T(t-1)\tilde{x}_g(t) \leftarrow e_g(t) \\
&= d(t) - \hat{g}^T(t-1)\tilde{x}_h(t) \leftarrow e_h(t)
\end{align*}
\]

\[
\begin{align*}
\tilde{x}_g(t) &= [\hat{g}(t-1) \otimes I_L]^T\tilde{x}(t) \\
\tilde{x}_h(t) &= [I_M \otimes \hat{h}(t-1)]^T\tilde{x}(t)
\end{align*}
\]
Kalman Filter for Bilinear Forms (KF - BF)

- estimated output signal: \( \hat{y}(t) = \hat{h}^T(t - 1)\textbf{X}(t)\hat{g}(t - 1) \)
- error signal:

\[
e(t) = d(t) - \hat{f}^T(t - 1)\tilde{\textbf{x}}(t)
= d(t) - \hat{h}^T(t - 1)\tilde{\textbf{x}}_g(t) \leftarrow e_g(t)
= d(t) - \hat{g}^T(t - 1)\tilde{\textbf{x}}_h(t) \leftarrow e_h(t)
\]

\[
\tilde{\textbf{x}}_g(t) = [\hat{\textbf{g}}(t - 1) \otimes \textbf{I}_L]^T\tilde{\textbf{x}}(t)
\tilde{\textbf{x}}_h(t) = [\textbf{I}_M \otimes \hat{\textbf{h}}(t - 1)]^T\tilde{\textbf{x}}(t)
\]

- optimal estimates of the state vectors:

\[
\hat{\textbf{h}}(t) = \hat{\textbf{h}}(t - 1) + k_h(t)e(t)
\hat{\textbf{g}}(t) = \hat{\textbf{g}}(t - 1) + k_g(t)e(t)
\]

\( k_h(t), k_g(t): \) Kalman gain vectors
a posteriori misalignments:
\[ \mu_h(t) = \frac{h(t)}{\eta} - \hat{h}(t) \]
\[ \mu_g(t) = \eta g(t) - \hat{g}(t) \]

a priori misalignments:
\[ m_h(t) = \frac{h(t)}{\eta} - \hat{h}(t - 1) \]
\[ = \mu_h(t - 1) + \frac{w_h(t)}{\eta} \]
\[ m_g(t) = \eta g(t) - \hat{g}(t - 1) \]
\[ = \mu_g(t - 1) + \eta w_g(t) \]
a posteriori misalignments:
\[ \mu_h(t) = h(t)/\eta - \hat{h}(t) \]
\[ \mu_g(t) = \eta g(t) - \hat{g}(t) \]

a priori misalignments:
\[ m_h(t) = h(t)/\eta - \hat{h}(t-1) \]
\[ m_h(t) = \mu_h(t-1) + w_h(t)/\eta \]
\[ m_g(t) = \eta g(t) - \hat{g}(t-1) \]
\[ m_g(t) = \mu_g(t-1) + \eta w_g(t) \]

simplifying notations:
\[ \overline{w}_h(t) = w_h(t)/\eta \]
\[ R_{m_h}(t) = R_{\mu_h}(t-1) + \sigma^2_{w_h} I_L \]
\[ \overline{w}_g(t) = \eta w_g(t) \]
\[ R_{m_g}(t) = R_{\mu_g}(t-1) + \sigma^2_{w_g} I_M \]
a posteriori misalignments:

\[ \mu_h(t) = \frac{h(t)}{\eta} - \hat{h}(t) \]
\[ \mu_g(t) = \eta g(t) - \hat{g}(t) \]

a priori misalignments:

\[ m_h(t) = \frac{h(t)}{\eta} - \hat{h}(t-1) \]
\[ = \mu_h(t-1) + \frac{w_h(t)}{\eta} \]
\[ m_g(t) = \eta g(t) - \hat{g}(t-1) \]
\[ = \mu_g(t-1) + \eta \frac{w_g(t)}{\eta} \]

simplifying notations:

\[ \bar{w}_h(t) = \frac{w_h(t)}{\eta} \]
\[ \bar{w}_g(t) = \eta \frac{w_g(t)}{\eta} \]
\[ R_{m_h}(t) = R_{\mu_h}(t-1) + \sigma^2_{w_h} I_L \]
\[ R_{m_g}(t) = R_{\mu_g}(t-1) + \sigma^2_{w_g} I_M \]

minimizing \( \frac{1}{L} \text{tr} \left[ R_{\mu_h}(t) \right] \), \( \frac{1}{M} \text{tr} \left[ R_{\mu_g}(t) \right] \) yields:

\[ k_h(t) = R_{m_h}(t) \tilde{x}_g(t) [\tilde{x}_g^T(t) R_{m_h}(t) \tilde{x}_g(t) + \sigma^2_v]^{-1} \]
\[ k_g(t) = R_{m_g}(t) \tilde{x}_h(t) [\tilde{x}_h^T(t) R_{m_g}(t) \tilde{x}_h(t) + \sigma^2_v]^{-1} \]
Simplifying assumptions:
Simplifying assumptions:

- after convergence was reached:

\[ R_{mh}(t) \approx \sigma_{mh}^2(t)I_L \quad R_{mg}(t) \approx \sigma_{mg}^2(t)I_M \]
Simplifying assumptions:

1. 
   - after convergence was reached:
     \[ R_{mh}(t) \approx \sigma^2_{mh}(t)I_L \]
     \[ R_{mg}(t) \approx \sigma^2_{mg}(t)I_M \]

2. 
   - misalignments of the individual coefficients: uncorrelated
   \[ \rightarrow \approx \]

\[ I_L - k_h(t)\hat{x}_g^T(t) \approx \left[ 1 - \frac{1}{L}k_h^T(t)\hat{x}_g(t) \right]I_L \]
\[ I_M - k_g(t)\hat{x}_h^T(t) \approx \left[ 1 - \frac{1}{M}k_g^T(t)\hat{x}_h(t) \right]I_M \]
Simplifying assumptions:

- after convergence was reached:
  \[ R_{mh}(t) \approx \sigma^2_{mh}(t)I_L \]
  \[ R_{mg}(t) \approx \sigma^2_{mg}(t)I_M \]

- misalignments of the individual coefficients: uncorrelated → approximate:
  \[ I_L - k_h(t)\tilde{x}_g^T(t) \approx \left[ 1 - \frac{1}{L} k_h^T(t)\tilde{x}_g(t) \right] I_L \]
  \[ I_M - k_g(t)\tilde{x}_h^T(t) \approx \left[ 1 - \frac{1}{M} k_g^T(t)\tilde{x}_h(t) \right] I_M \]

⇒ Simplified Kalman Filter for bilinear forms (SKF - BF)
Practical Considerations

The parameters related to uncertainties in $h, g$: $\sigma^2_w, \sigma^2_w$:

- small $\Rightarrow$ good misalignment, poor tracking
- large (i.e., high uncertainty in the systems) $\Rightarrow$
  $\Rightarrow$ good tracking, high misalignment
The parameters related to uncertainties in $h$, $g$: $\sigma_{\tilde{w}_h}^2$, $\sigma_{\tilde{w}_g}^2$:

- small $\Rightarrow$ good misalignment, poor tracking
- large (i.e., high uncertainty in the systems) $\Rightarrow$ good tracking, high misalignment

A good compromise is needed!
The parameters related to uncertainties in $h$, $g$: $\sigma^2_{wh}$, $\sigma^2_{wg}$:
- small $\Rightarrow$ good misalignment, poor tracking
- large (i.e., high uncertainty in the systems) $\Rightarrow$ good tracking, high misalignment

A good compromise is needed!

In practice $\rightarrow$ some a priori information
(e.g., if $g$ - time-invariant $\Rightarrow \sigma^2_{wg} = 0$)
The parameters related to uncertainties in $h$, $g$: $\sigma^2_{wh}$, $\sigma^2_{wg}$:

- small $\Rightarrow$ good misalignment, poor tracking
- large (i.e., high uncertainty in the systems) $\Rightarrow$ good tracking, high misalignment

A good compromise is needed!

In practice $\rightarrow$ some a priori information
(e.g., if $g$ - time-invariant $\Rightarrow \sigma^2_{wg} = 0$)

By applying the $\ell_2$ norm on the state equation:

$$\hat{\sigma}^2_{wh}(t) = \frac{1}{L} \left\| \hat{h}(t) - \hat{h}(t-1) \right\|^2_2$$
Simulation Setup

Conditions

- Input signals $x_m(t), m = 1, 2, \ldots, M$ - independent WGN, respectively AR(1) generated by filtering a white Gaussian noise through a first-order system $1 / (1 - 0.8z^{-1})$
- $h, g$ - Gaussian, randomly generated, of lengths $L = 64, M = 8$
- $v(t)$ - independent white Gaussian noise signal
Simulation Setup

**Conditions**
- Input signals $x_m(t)$, $m = 1, 2, \ldots, M$ - independent WGN, respectively AR(1) generated by filtering a white Gaussian noise through a first-order system $1 / (1 - 0.8z^{-1})$
- $h, g$ - Gaussian, randomly generated, of lengths $L = 64$, $M = 8$
- $v(t)$ - independent white Gaussian noise signal

**Compared algorithms**
- KF-BF and KF
- SKF-BF and SKF when $\sigma_{wh}^2 = \sigma_{wg}^2 = \sigma_w^2 = 10^{-9}$
- SKF-BF and SKF when $\sigma_{wg}^2 = 0$ and $\hat{\sigma}_{wh}^2(t) = \frac{\|\hat{h}(t) - \hat{h}(t-1)\|_2^2}{L}$
Figure 1: Normalized misalignment of the KF-BF and regular KF for different types of input signals. $ML = 512$, $\sigma_v^2 = 0.01$, $\sigma_{wh}^2 = \sigma_{wg}^2 = \sigma_w^2 = 10^{-9}$, and $\epsilon = 10^{-5}$. 
Figure 2: Normalized misalignment of the SKF-BF and regular SKF for different types of input signals. Other conditions are the same as in Fig. 1.
Figure 3: Normalized misalignment of the SKF-BF and regular SKF for different types of input signals, using the recursive estimates $\hat{\sigma}^2_{w_h}(t)$ and $\hat{\sigma}^2_w(t)$, respectively.

$ML = 512$, $\sigma^2_v = 0.01$, $\sigma^2_w = 0$, and $\epsilon = 10^{-5}$. 

$\sigma^2_v = 0.01$, $\sigma^2_w = 0$, and $\epsilon = 10^{-5}$. 

IEEE ICASSP, 15-20 April 2018
Calgary, Alberta, Canada
April 20, 2018
Conclusions

- **KF-BF, SKF-BF**: improvement in convergence rate and tracking with respect to regular KF, SKF
Conclusions

- **KF-BF, SKF-BF**: improvement in convergence rate and tracking with respect to regular KF, SKF

- The proposed solution uses 2 filters of lengths $L$, $M$, compared to one filter of length $ML$ for the regular one $\Rightarrow$ computationally simpler

- SKF-BF provides: reduced computational complexity, but also slower convergence rate, especially for correlated inputs with respect to KF-BF

IEEE ICASSP, 15-20 April 2018
Calgary, Alberta, Canada
Conclusions

- **KF-BF, SKF-BF**: improvement in convergence rate and tracking with respect to regular KF, SKF

- The proposed solution uses 2 filters of lengths $L, M$, compared to one filter of length $ML$ for the regular one $\Rightarrow$ computationally simpler

- SKF-BF provides:
  - reduced computational complexity, but also
  - slower convergence rate, especially for correlated inputs

with respect to KF-BF
Conclusions

- **KF-BF, SKF-BF**: improvement in convergence rate and tracking with respect to regular KF, SKF

- The proposed solution uses 2 filters of lengths $L, M$, compared to one filter of length $ML$ for the regular one $\Rightarrow$ computationally simpler

- SKF-BF provides:
  - reduced computational complexity , but also
  - slower convergence rate, especially for correlated inputs

  with respect to KF-BF

- The experimental results indicate the good performance of the proposed algorithms
Thank you!