Kalman filter in speech and music

Orchisama Das

Center for Computer Research in Music and Acoustics, Stanford University

odas@stanford.edu

July 26, 2017

Overview

1. Kalman filter introduction
2. KF in speech - used for speech denoising [Das 2016]
3. KF in music - used for monophonic pitch tracking [Das 2017]

Kalman filter

- Recursive MMSE estimator proposed by R.Kalman in [Kalman 1960]
- Basic idea - Update current state $\hat{x}(n)$ based on present noisy observation $y(n)$ and all past predictions $\hat{x}(0), \hat{x}(1), \ldots, \hat{x}(n-1)$.
- Generalized Model -

$$\begin{aligned} x(n) &= Ax(n-1) + u(n); \mathbf{u} \sim N(0, Q) \\ y(n) &= Hx(n) + w(n); \mathbf{w} \sim N(0, R) \end{aligned} \tag{1}$$

  - $x(n)$ - Current state, $y(n)$ - Current noisy observation
  - $u(n)$ - Process noise, $w(n)$ - Observation noise
  - Q - Process noise covariance matrix, R - Measurement noise covariance matrix
  - A - State transition matrix, H - Observation vector

Kalman filter equations

- Time update equations

$$
\begin{aligned}
\hat{x}_n^- &= A\hat{x}_{n-1} \\
P_n^- &= AP_{n-1}A^T + Q
\end{aligned} \tag{2}
$$

- Measurement update equations

$$
\begin{aligned}
K_n &= P_n^- H^T (HP_n^- H^T + R)^{-1} \\
\hat{x}_n &= \hat{x}_n^- + K_n(y_n - H\hat{x}_n^-) \\
P_n &= (I - K_n H)P_n^-
\end{aligned} \tag{3}
$$

  P is the error covariance matrix and K is the Kalman gain that acts as a weighting factor between observed and predicted values.

- Kalman filter best suited to track model dynamics in the presence of noise, for e.g position of robotic arm with noisy sensors.

Use in speech denoising

- Original paper proposed by Paliwal and Basu in [Paliwal 1987] that made use of autoregressive model of speech.
- Model -

$$
\begin{aligned}
x(n) &= -\sum_{i=1}^{p} \alpha_i x(n-i) + u(n) \\
y(n) &= x(n) + w(n)
\end{aligned}
\tag{4}
$$

- α_i's also known as linear prediction coefficients (LPC) usually calculated by solving the Yule-Walker equations.
- α_i's are all we need to know to form transition matrix, the observation matrix is $H = [0, 0, \dots, 1]$. Remaining KF equations are same.

Kalman filter tuning [Das 2016]

- Tune other filter parameters - Q and R
- R - measurement noise variance can be estimated relatively easily by getting noisy samples from silent regions and calculating their variance.
- Q - process noise variance, part of model. Use robustness and sensitivity metrics to find best compromise value of Q [Saha 2014].

$$\begin{aligned} A_n &= H(AP_n^- A^T)H^T \\ J_1 &= \frac{\sigma_w^2}{\sigma_w^2 + A_n + \sigma_u^2} \\ J_2 &= \frac{\sigma_u^2}{\sigma_u^2 + A_n} \end{aligned} \tag{5}$$

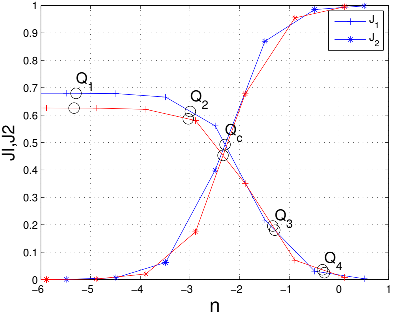


Figure: J_1 - *sensitivity metric,* J_2 - *robustness metric for blue - voiced frames, red- silent frames.*

- Two values of Q used - Q_c for voiced frames for balance between sensitivity and robustness and $Q_2 < Q_c$ for more sensitive performance in silent frames.

Contd. II

- Kalman gain adjustment according to value of Q and frame type - smaller K for silent frames, higher K for voiced frames.

$$\hat{x}_n = (I - HK_n)\hat{x}_n^- + K_n y(n) \tag{6}$$

- Smaller K denotes less weight on noisy observation and more weight on past prediction to predict current state - ideal for silent frames which only have noise, no speech information that needs to be preserved.

Contd. III

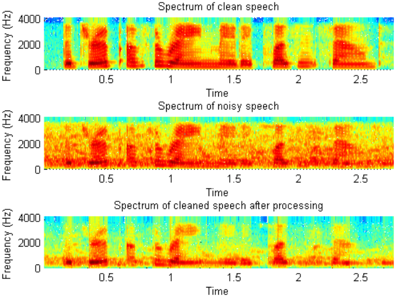


Figure: Example of clean, noisy and enhanced speech at 5dB SNR.

Use in pitch tracking

- Extended complex Kalman filter proposed in [Dash 2000] to track distorted power system signal frequency in presence of noise and harmonics.
- Same model (sines+noise) can be applied to music. ECKF Pitch tracker proposed in [Das 2017].
- Original Kalman filter proposed for linear systems. Extended Kalman filter works on non-linear systems, where the current state estimate is a non-linear function of the previous estimate. EKF linearizes the function by using a first order Taylor series expansion.

ECKF Pitch Tracker I

- Model-

$$y_k = a_1 \cos\left(\omega_1 k T_s + \phi_1\right) + w_k \tag{7}$$

where a_1, w_1 and ϕ_1 are the fundamental amplitude, frequency and phase respectively and T_s is the sampling interval. The state vector is constructed as:

$$x_k = \begin{bmatrix} \alpha \\ u_k \\ u_k^* \end{bmatrix} \tag{8}$$

where

$$\begin{aligned} \alpha &= \exp(j\omega_1 T_s) \\ u_k &= a_1 \exp(j\omega_1 k T_s + j\phi_1) \\ u_k^* &= a_1 \exp(-j\omega_1 k T_s - j\phi_1) \end{aligned} \tag{9}$$

- The state vector estimate update rule x_{k+1} relates to x_k as

$$\begin{aligned}\begin{bmatrix}\alpha \\ u_{k+1} \\ u^*_{k+1}\end{bmatrix} &= \begin{bmatrix}1 & 0 & 0 \\ 0 & \alpha & 0 \\ 0 & 0 & \frac{1}{\alpha}\end{bmatrix}\begin{bmatrix}\alpha \\ u_k \\ u^*_k\end{bmatrix} \\ x_{k+1} &= f(x_k) \\ f(x_k) &= \begin{bmatrix}\alpha & \alpha u_k & \frac{u^*_k}{\alpha}\end{bmatrix}^T\end{aligned} \tag{10}$$

- y_k relates to x_k as

$$\begin{aligned}y_k &= Hx_k + w_k \\ H &= \begin{bmatrix}0 & 0.5 & 0.5\end{bmatrix}\end{aligned} \tag{11}$$

ECKF Pitch Tracker III

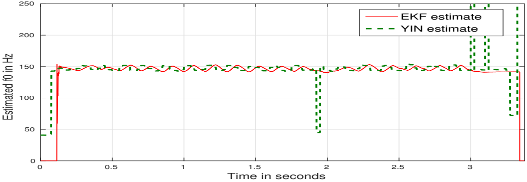
- The transition matrix gets replaced by the Jacobian of $f(x_k)$ w.r.t x_k.

$$F_k = \frac{\partial f(x_k)}{\partial x_k}\bigg|_{x_k = \hat{x}_{k|k}} = \begin{bmatrix} 1 & 0 & 0 \\ \hat{x}_{k|k}(2) & \hat{x}_{k|k}(1) & 0 \\ -\frac{\hat{x}_{k|k}(3)}{\hat{x}^2_{k|k}(1)} & 0 & \frac{1}{\hat{x}_{k|k}(1)} \end{bmatrix} \tag{12}$$

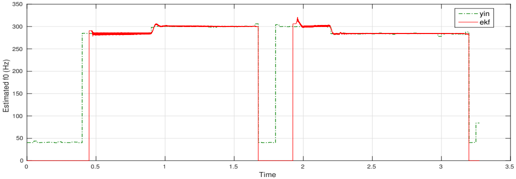
- Detect silent regions (unpitched moments) by dividing signal into frames and looking at spectral flatness. Silent frames will have spectral flatness close to 1.
- Reset covariance matrix and recalculate initial estimates with FFT every time there is a transition from non-silent to silent frame to enable ECKF to track rapidly changing pitch dynamics.
- Adaptive process noise:

$$\log_{10}(\sigma_u^2) = -c + |y_k - H\hat{x}_{k|k}| \tag{13}$$

ECKF Pitch Tracker IV



(a) ECKF tracking vibrato played on the double bass



(b) ECKF tracking two notes hammered on the guitar

References I

R. Kalman (1960)
A new approach to linear filtering and prediction problems.
Journal of basic Engineering 82.1 (1960): 35-45.

K.Paliwal and A.Basu
A speech enhancement method based on Kalman filtering.
International Conference on Acoustics, Speech, and Signal Processing, IEEE ICASSP'87 Vol. 12. IEEE, 1987.

M. Saha, R. Ghosh, and B. Goswami
Robustness and sensitivity metrics for tuning the extended Kalman filter
IEEE Transactions in Instrumentation and Measurement, vol. 63, no. 4, pp. 964971, 2014.

O.Das, B.Goswami, R.Ghosh (2016)
Application of the tuned Kalman filter in speech enhancement
IEEE International Conference on Control, Measurement and Instrumentation (CMI 2016)

References II

P. K. Dash et al.
An extended complex Kalman filter for frequency measurement of distorted signals.
Power Engineering Society Winter Meeting, 2000. IEEE. Vol. 3. IEEE, 2000.

O.Das, J. Smith, C.Chafe (2017)
Real-time pitch tracking in audio signals with the extended complex Kalman filter
International Conference on Digital Audio Effects (DAFx 2017)

The End