Statistical Signal Processing

Lecture 11

Statistical Signal Modeling, Learning and Processing

- chapter 4: Adaptive Filtering
  Tracking stationary Time-Varying Parameters
  - model, LMS & RLS performance
  - optimal approach: Kalman filtering
- chapter 3: Optimal Filtering
  Kalman Filtering
  - state-space models
  - basic Kalman filter derivation
  - extensions
- chapter 5: Sinusoids in Noise prototype problem

Outline

State-Space Model

- The signal model can be written as

$$
\begin{array}{rcl}
& \text{state update equation:} & \\
\boldsymbol{x}_{k+1} & = & \boldsymbol{F}_k\, \boldsymbol{x}_k + \boldsymbol{G}_k\, \boldsymbol{w}_k \\
& \text{measurement equation:} & \\
\boldsymbol{y}_k & = & \boldsymbol{H}_k\, \boldsymbol{x}_k + \boldsymbol{v}_k
\end{array}
\tag{1}
$$

  where

$$
\mathrm{E}\begin{bmatrix} \boldsymbol{x}_0 \\ \boldsymbol{w}_k \\ \boldsymbol{v}_k \end{bmatrix} = \begin{bmatrix} \widehat{\boldsymbol{x}}_0 \\ \mathbf{0} \\ \mathbf{0} \end{bmatrix}, \quad \mathrm{E}\begin{bmatrix} \boldsymbol{x}_0 \\ \boldsymbol{w}_k \\ \boldsymbol{v}_k \end{bmatrix}\begin{bmatrix} \boldsymbol{x}_0 \\ \boldsymbol{w}_l \\ \boldsymbol{v}_l \end{bmatrix}^T = \begin{bmatrix} \boldsymbol{P}_0 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{Q}\,\delta_{kl} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \boldsymbol{R}\,\delta_{kl} \end{bmatrix}
\tag{2}
$$

- State vector $\boldsymbol{x}_k$: the state summarizes the past to produce the current output $\boldsymbol{y}_k$.
- State update: vector AR(1) process.
- Kalman filter: recursive LMMSE estimation of $\boldsymbol{x}_k$ on basis of $\boldsymbol{y}_{1:k}$.

State-Space Model example: AR(n) in noise

- Consider an AR(n) signal in white noise:

$$\begin{aligned} s_k &= -\sum_{i=1}^{n} a_i \, s_{k-i} + w_{k-1} \\ y_k &= s_k + v_k \end{aligned} \tag{3}$$

- corresponding n-dimensional state-space model:

$$\begin{aligned} \boldsymbol{x}_k = \begin{bmatrix} s_k \\ s_{k-1} \\ \vdots \\ s_{k-n+1} \end{bmatrix}, \ \boldsymbol{F} = \begin{bmatrix} -a_1 & -a_2 & \cdots & -a_{n-1} & -a_n \\ 1 & 0 & \cdots & & 0 \\ & \ddots & & & \vdots \\ 0 & \cdots & & 1 & 0 \end{bmatrix}, \ \boldsymbol{G} = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \\ \boldsymbol{H} = \begin{bmatrix} 1 & 0 & \cdots & 0 \end{bmatrix} \end{aligned} \tag{4}$$

- State-space model not unique: if $\boldsymbol{T}$ invertible,
  $\boldsymbol{x}_k^{'} = \boldsymbol{T}\boldsymbol{x}_k$, $\boldsymbol{F}^{'} = \boldsymbol{T}\boldsymbol{F}\boldsymbol{T}^{-1}$, $\boldsymbol{G}^{'} = \boldsymbol{T}\boldsymbol{G}$, $\boldsymbol{H}^{'} = \boldsymbol{H}\boldsymbol{T}^{-1}$
  also valid state-space model for y_k.
  Depends whether $\boldsymbol{x}_k$ (components) have meaning.

State-Space Model example: Position Tracking

White Noise Acceleration

Let $\boldsymbol{p}_k$ be the position at sampling instant k, $\boldsymbol{v}_k$ the velocity (not to be confused with the measurement noise) and $\boldsymbol{a}_k$ the acceleration. In the case of e.g. 3D positioning, $\boldsymbol{p}_k$ is of the form $\boldsymbol{p} = [x\ y\ z]^T$. By simple discretization of the differential equations of motion, we get

$$\begin{array}{rcl} \boldsymbol{p}_{k+1} & = & \boldsymbol{p}_k + \boldsymbol{v}_k \\ \boldsymbol{v}_{k+1} & = & \boldsymbol{v}_k + \boldsymbol{a}_k \\ \boldsymbol{a}_k & = & \boldsymbol{w}_k \ . \end{array} \tag{5}$$

In the case of modeling the acceleration as (temporally) white noise, the acceleration is the process noise. To simplify the equations, we assume here that the unit of time for velocity and acceleration is the sampling period. The physical speed and acceleration are then $t_s\,\boldsymbol{v}_k$ and $t_s^2\,\boldsymbol{a}_k$ where t_s is the sampling period expressed in seconds, assuming $\boldsymbol{p}_k$ is expressed in meters. We get for the state-space model

$$\boldsymbol{x}_k = \begin{bmatrix} \boldsymbol{p}_k \\ \boldsymbol{v}_k \end{bmatrix},\ \boldsymbol{F} = \begin{bmatrix} \boldsymbol{I}_n & \boldsymbol{I}_n \\ \boldsymbol{0}_{n,n} & \boldsymbol{I}_n \end{bmatrix},\ \boldsymbol{G} = \begin{bmatrix} \boldsymbol{0}_{n,n} \\ \boldsymbol{I}_n \end{bmatrix},\ \boldsymbol{y}_k = \widehat{\boldsymbol{p}}_k, \boldsymbol{H} = [\boldsymbol{I}_n\ \ \boldsymbol{0}_{n,n}] \tag{6}$$

where $\boldsymbol{0}_{n,m}$ is a $n \times m$ matrix of zeros.
The only unknown system parameter in this case is the acceleration covariance matrix $\boldsymbol{Q}$.

State-Space Model example: Position Tracking

AR(1) (Markov) Acceleration

In this case we assume a first-order autoregressive model for the acceleration $\boldsymbol{a}_{k+1} = \boldsymbol{A}\,\boldsymbol{a}_k + \boldsymbol{w}_k$ where now $\boldsymbol{A}$ and $\boldsymbol{Q}$ are unknown (need to be estimated also).

We have (pseudo-inverse) $\boldsymbol{G}^+ = [\mathbf{0}_{n,2n}\ \ \boldsymbol{I}_n]$ and $\boldsymbol{a}_k = \boldsymbol{G}^+\boldsymbol{x}_k$. Note that $\boldsymbol{G}^+\boldsymbol{F} = \boldsymbol{A}\boldsymbol{G}^+$.

We get for the state-space model

$$\boldsymbol{x}_k = \begin{bmatrix} \boldsymbol{p}_k \\ \boldsymbol{v}_k \\ \boldsymbol{a}_k \end{bmatrix},\ \boldsymbol{F} = \begin{bmatrix} \boldsymbol{I}_n & \boldsymbol{I}_n & \mathbf{0}_{n,n} \\ \mathbf{0}_{n,n} & \boldsymbol{I}_n & \boldsymbol{I}_n \\ \mathbf{0}_{n,n} & \mathbf{0}_{n,n} & \boldsymbol{A} \end{bmatrix},\ \boldsymbol{G} = \begin{bmatrix} \mathbf{0}_{n,n} \\ \mathbf{0}_{n,n} \\ \boldsymbol{I}_n \end{bmatrix},\ \boldsymbol{H} = [\boldsymbol{I}_n\ \ \mathbf{0}_{n,2n}]\,. \tag{7}$$

Outline

KF derivation: Innovations approach

- notation $\boldsymbol{y}_{1:k} = \{\boldsymbol{y}_1, \ldots, \boldsymbol{y}_k\}$.
- The KF performs Gram-Schmidt orthogonalization (decorrelation) of the measurements $\boldsymbol{y}_k$.
  $\widehat{\boldsymbol{y}}_{k|k-1}$ = LMMSE predictor of $\boldsymbol{y}_k$ on the basis of $\boldsymbol{y}_{1:k-1}$,
  leading to the orthogonalized prediction error (or innovation) $\widetilde{\boldsymbol{y}}_k = \widetilde{\boldsymbol{y}}_{k|k-1} = \boldsymbol{y}_k - \widehat{\boldsymbol{y}}_{k|k-1}$.
- Correlation matrix notation $R_{\boldsymbol{x}\boldsymbol{y}} = \mathrm{E}\,\boldsymbol{x}\,\boldsymbol{y}^T$. Denote $R_{\widetilde{\boldsymbol{y}}_k\widetilde{\boldsymbol{y}}_k} = \boldsymbol{S}_k$.
- Innovations approach idea: linear estimation in terms of $\boldsymbol{y}_{1:k}$ is equivalent to estimation in terms of $\widetilde{\boldsymbol{y}}_{1:k}$ since one set is obtained from the other by an invertible linear transformation. Now, since the $\widetilde{\boldsymbol{y}}_k$ are decorrelated, estimation in terms of $\widetilde{\boldsymbol{y}}_{1:k}$ simplifies:

$$\widehat{\boldsymbol{x}}_{|k} = \sum_{i=1}^{k} R_{\boldsymbol{x}\widetilde{\boldsymbol{y}}_i} R^{-1}_{\widetilde{\boldsymbol{y}}_i\widetilde{\boldsymbol{y}}_i} \widetilde{\boldsymbol{y}}_i = \widehat{\boldsymbol{x}}_{|k-1} + R_{\boldsymbol{x}\widetilde{\boldsymbol{y}}_k} \boldsymbol{S}_k^{-1} \widetilde{\boldsymbol{y}}_k \ . \tag{8}$$

  Used to obtain predicted estimates $\widehat{\boldsymbol{x}}_{k|k-1}$ with estimation error $\widetilde{\boldsymbol{x}}_{k|k-1} = \boldsymbol{x}_k - \widehat{\boldsymbol{x}}_{k|k-1}$ and error covariance matrix $\boldsymbol{P}_{k|k-1} = R_{\widetilde{\boldsymbol{x}}_{k|k-1}\widetilde{\boldsymbol{x}}_{k|k-1}}$ and also filtered estimates $\widehat{\boldsymbol{x}}_{k|k}$ with estimation error $\widetilde{\boldsymbol{x}}_{k|k} = \boldsymbol{x}_k - \widehat{\boldsymbol{x}}_{k|k}$ and error covariance matrix $\boldsymbol{P}_{k|k} = R_{\widetilde{\boldsymbol{x}}_{k|k}\widetilde{\boldsymbol{x}}_{k|k}}$.

Kalman Filter: Time Update

- Notation: $\mathcal{L}\{\boldsymbol{y}_{1:k}\}$ = linear span of $\boldsymbol{y}_1, \ldots, \boldsymbol{y}_k$ = linear vector space spanned by $\boldsymbol{y}_1, \ldots, \boldsymbol{y}_k$ (using matrices of appropriate dimension as combination coefficients). Then e.g. $\mathcal{L}\{\boldsymbol{y}_{1:k}\} = \mathcal{L}\{\widetilde{\boldsymbol{y}}_{1:k}\}$.
- $\boldsymbol{v}_k \perp \mathcal{L}\{\boldsymbol{v}_{1:k-1}\}$ means decorrelation of $\boldsymbol{v}_k$ from the indicated space due to the whiteness of the measurement noise process.
- We have $\boldsymbol{x}_k \in \mathcal{L}\{\boldsymbol{x}_0, \boldsymbol{w}_{1:k-1}\}$, $\boldsymbol{y}_k \in \mathcal{L}\{\boldsymbol{x}_0, \boldsymbol{w}_{1:k-1}, \boldsymbol{v}_k\}$, hence $\mathcal{L}\{\boldsymbol{y}_{1:k-1}\} \subset \mathcal{L}\{\boldsymbol{x}_0, \boldsymbol{w}_{1:k-2}, \boldsymbol{v}_{1:k-1}\} \perp \boldsymbol{v}_k$.
- Hence

$$\begin{aligned}
\widehat{\boldsymbol{y}}_{k|k-1} &= \boldsymbol{H}_k\, \widehat{\boldsymbol{x}}_{k|k-1} + \widehat{\boldsymbol{v}}_{k|k-1} = \boldsymbol{H}_k\, \widehat{\boldsymbol{x}}_{k|k-1} \\
\widetilde{\boldsymbol{y}}_{k|k-1} &= \boldsymbol{y}_k - \widehat{\boldsymbol{y}}_{k|k-1} = \boldsymbol{H}_k\, \widetilde{\boldsymbol{x}}_{k|k-1} + \boldsymbol{v}_k \\
\boldsymbol{S}_k &= R_{\widetilde{\boldsymbol{y}}_{k|k-1}\widetilde{\boldsymbol{y}}_{k|k-1}} = \boldsymbol{H}_k\, \boldsymbol{P}_{k|k-1}\, \boldsymbol{H}_k^T + \boldsymbol{R}_k
\end{aligned} \tag{9}$$

- $\widehat{\boldsymbol{w}}_{k|k} = 0$ since $\mathcal{L}\{\boldsymbol{y}_{1:k}\} \subset \mathcal{L}\{\boldsymbol{x}_0, \boldsymbol{w}_{1:k-1}, \boldsymbol{v}_{1:k}\} \perp \boldsymbol{w}_k$. Hence

$$\begin{aligned}
\widehat{\boldsymbol{x}}_{k+1|k} &= \boldsymbol{F}_k\, \widehat{\boldsymbol{x}}_{k|k} + \boldsymbol{G}_k\, \widehat{\boldsymbol{w}}_{k|k} = \boldsymbol{F}_k\, \widehat{\boldsymbol{x}}_{k|k} \\
\widetilde{\boldsymbol{x}}_{k+1|k} &= \boldsymbol{x}_{k+1} - \widehat{\boldsymbol{x}}_{k+1|k} = \boldsymbol{F}_k\, \widetilde{\boldsymbol{x}}_{k|k} + \boldsymbol{G}_k\, \boldsymbol{w}_k \\
\boldsymbol{P}_{k+1|k} &= \boldsymbol{F}_k\, \boldsymbol{P}_{k|k}\, \boldsymbol{F}_k^T + \boldsymbol{G}_k\, \boldsymbol{Q}_k\, \boldsymbol{G}_k^T
\end{aligned} \tag{10}$$

Kalman Filter: Measurement Update

- $$\begin{aligned}
\widehat{\boldsymbol{x}}_{k|k} &= \widehat{\boldsymbol{x}}_{k|k-1} + R_{\boldsymbol{x}_k\widetilde{\boldsymbol{y}}_k}\,\boldsymbol{S}_k^{-1}\,\widetilde{\boldsymbol{y}}_k \\
R_{\boldsymbol{x}_k\widetilde{\boldsymbol{y}}_k} &= \mathrm{E}\,\boldsymbol{x}_k(\boldsymbol{H}_k\,\widetilde{\boldsymbol{x}}_{k|k-1} + \boldsymbol{v}_k)^T = \mathrm{E}\,\boldsymbol{x}_k\widetilde{\boldsymbol{x}}_{k|k-1}^T\,\boldsymbol{H}_k^T + \mathrm{E}\,\boldsymbol{x}_k\boldsymbol{v}_k^T = \boldsymbol{P}_{k|k-1}\,\boldsymbol{H}_k^T \\
\text{Let } \boldsymbol{K}_k &= \boldsymbol{P}_{k|k-1}\,\boldsymbol{H}_k^T\,\boldsymbol{S}_k^{-1} \quad \text{Kalman gain} \\
\widehat{\boldsymbol{x}}_{k|k} &= \widehat{\boldsymbol{x}}_{k|k-1} + \boldsymbol{K}_k\,\widetilde{\boldsymbol{y}}_k \\
R_{\widehat{\boldsymbol{x}}_{k|k}\widehat{\boldsymbol{x}}_{k|k}} &= R_{\widehat{\boldsymbol{x}}_{k|k-1}\widehat{\boldsymbol{x}}_{k|k-1}} + \boldsymbol{K}_k\,\boldsymbol{S}_k\,\boldsymbol{K}_k^T \quad \text{because } \widetilde{\boldsymbol{y}}_k \perp \mathcal{L}\{\widetilde{\boldsymbol{y}}_{1:k}\} \\
\boldsymbol{P}_{k|k} &= R_{\boldsymbol{x}_k\boldsymbol{x}_k} - R_{\widehat{\boldsymbol{x}}_{k|k}\widehat{\boldsymbol{x}}_{k|k}} = R_{\boldsymbol{x}_k\boldsymbol{x}_k} - R_{\widehat{\boldsymbol{x}}_{k|k-1}\widehat{\boldsymbol{x}}_{k|k-1}} - \boldsymbol{K}_k\,\boldsymbol{S}_k\,\boldsymbol{K}_k^T \\
&= \boldsymbol{P}_{k|k-1} - \boldsymbol{K}_k\,\boldsymbol{S}_k\,\boldsymbol{K}_k^T
\end{aligned} \tag{11}$$
- $R_{\boldsymbol{x}_k\boldsymbol{x}_k} = R_{\widehat{\boldsymbol{x}}_{k|k}\widehat{\boldsymbol{x}}_{k|k}} + \boldsymbol{P}_{k|k}$: orthogonality property LMMSE : $R_{\widehat{\boldsymbol{x}}\widetilde{\boldsymbol{x}}} = \boldsymbol{0}$
- Note: cannot do $\widetilde{\boldsymbol{x}}_{k|k} = \widetilde{\boldsymbol{x}}_{k|k-1} - \boldsymbol{K}_k\,\widetilde{\boldsymbol{y}}_k \;\Rightarrow\; \boldsymbol{P}_{k|k} \neq \boldsymbol{P}_{k|k-1} + \boldsymbol{K}_k\,\boldsymbol{S}_k\,\boldsymbol{K}_k^T$ because $\boldsymbol{x}_k \not\perp \mathcal{L}\{\widetilde{\boldsymbol{y}}_k\}$!

Outline

Kalman Filter summary

- two-step recursive procedure to go from $|k-1$ to $|k$:
- Measurement Update

$$\begin{array}{rcl} \widetilde{\boldsymbol{y}}_{k|k-1} & = & \boldsymbol{y}_k - \boldsymbol{H}_k\, \widehat{\boldsymbol{x}}_{k|k-1} \\ \boldsymbol{S}_k & = & \boldsymbol{H}_k\, \boldsymbol{P}_{k|k-1}\, \boldsymbol{H}_k^T + \boldsymbol{R}_k \\ \boldsymbol{K}_k & = & \boldsymbol{P}_{k|k-1}\, \boldsymbol{H}_k^T\, \boldsymbol{S}_k^{-1} \\ \widehat{\boldsymbol{x}}_{k|k} & = & \widehat{\boldsymbol{x}}_{k|k-1} + \boldsymbol{K}_k\, \widetilde{\boldsymbol{y}}_{k|k-1} \\ \boldsymbol{P}_{k|k} & = & \boldsymbol{P}_{k|k-1} - \boldsymbol{K}_k\, \boldsymbol{H}_k\, \boldsymbol{P}_{k|k-1} \end{array} \tag{12}$$

- Time Update (prediction)

$$\begin{array}{rcl} \widehat{\boldsymbol{x}}_{k+1|k} & = & \boldsymbol{F}_k\, \widehat{\boldsymbol{x}}_{k|k} \\ \boldsymbol{P}_{k+1|k} & = & \boldsymbol{F}_k\, \boldsymbol{P}_{k|k}\, \boldsymbol{F}_k^T + \boldsymbol{G}_k\, \boldsymbol{Q}_k\, \boldsymbol{G}_k^T \end{array} \tag{13}$$

- There are various other ways to formulate these update equations, including performing both steps in one step.

$$\begin{array}{rcl} \widehat{\boldsymbol{x}}_{k+1|k} & = & \boldsymbol{F}_k(\boldsymbol{I} - \boldsymbol{K}_k \boldsymbol{H}_k)\, \widehat{\boldsymbol{x}}_{k|k-1} + \boldsymbol{F}_k \boldsymbol{K}_k\, \boldsymbol{y}_k \\ \boldsymbol{P}_{k+1|k} & = & \boldsymbol{F}_k\, \boldsymbol{P}_{k|k-1}\, \boldsymbol{F}_k^T + \boldsymbol{G}_k\, \boldsymbol{Q}_k\, \boldsymbol{G}_k^T - \boldsymbol{F}_k \boldsymbol{P}_{k|k-1} \boldsymbol{H}_k^T \boldsymbol{S}_k^{-1} \boldsymbol{H}_k \boldsymbol{P}_{k|k-1} \boldsymbol{F}_k^T \;\text{Riccati equation} \end{array} \tag{14}$$

- The choice of the initial conditions crucially affects the initial convergence (transient behavior). In the usual case of total absence of prior information on the initial state, one can choose $\widehat{\boldsymbol{x}}_{0|0} = \widehat{\boldsymbol{x}}_0 = 0$, $\boldsymbol{P}_{0|0} = \boldsymbol{P}_0 = p_0\, I$ with p_0 a (very) large number.

Outline

Kalman Filtering for causal Wiener Filtering

- time-invariant state-space model: ARMA signal in white noise
- steady-state Kalman filter solution leads to causal Wiener filter
- but the Kalman filter also applies to time-varying state-space models

Outline

Kalman Filter variations

- error feedback, stabilizing closed-loop, controllability, observability
- numerical stability: symmetry and positive (semi-)definiteness of $\boldsymbol{P}_k$, square-root Kalman filters, exponential forgetting
- smoothing: fixed-lag, fixed-interval
- nonlinear filtering: Extended Kalman Filter, ...
  special case: adapting hyperparameters (when added to the state)