Kalman filter

Yuxin Chen

Princeton University, Fall 2018

Outline

- Innovation sequence
- Kalman filter

Orthogonal sequences

- X: signal; $\boldsymbol{Y}$: observation vector (all zero mean)
- Suppose $\{Y_i\}$ are orthogonal, i.e. $\mathbb{E}[Y_i Y_j] = 0$ $(i \neq j)$, and let $\hat{X}(\boldsymbol{Y})$ (resp. $\hat{X}(Y_i)$) be the linear MMSE estimate of X given $\boldsymbol{Y}$ (resp. Y_i). Then one can verify that

$$\hat{X}(\boldsymbol{Y}) = \sum_{i=1}^{n} \hat{X}(Y_i),$$

$$\mathsf{MSE} = \mathsf{Var}(X) - \sum_{i=1}^{n} \frac{\mathsf{Cov}^2(X, Y_i)}{\mathsf{Var}(Y_i)}$$

Hence the computation of the linear MMSE estimate and its MSE are very simple

Orthogonal sequences

This means that the estimates and the MSE can be computed recursively

$$\hat{X}(Y^{i+1}) = \hat{X}(Y^i) + \hat{X}(Y_{i+1})$$

$$\mathsf{MSE}_{i+1} = \mathsf{MSE}_i - \frac{\mathsf{Cov}^2(X, Y_{i+1})}{\mathsf{Var}(Y_{i+1})}$$

where Y^i denotes $\{Y_1, \cdots, Y_i\}$, and $\hat{X}(Y^i)$ is linear MMSE estimate of X given $\{Y_1, \cdots, Y_i\}$

Nonorthogonal sequences

Now suppose the Y_is are not orthogonal. We can still express the estimate and its MSE as sums

- We first whiten $\boldsymbol{Y}$ to obtain $\boldsymbol{Z}$. The linear MMSE estimate of X given $\boldsymbol{Y}$ is exactly the same as that given $\boldsymbol{Z}$ (why?)
- The estimate and its MSE can then be computed as

$$\hat{X}(\boldsymbol{Y}) = \sum_{i=1}^{n} \hat{X}(Z_i)$$

$$\mathsf{MSE} = \mathsf{Var}(X) - \sum_{i=1}^{n} \frac{\mathsf{Cov}^2(X, Z_i)}{\mathsf{Var}(Z_i)}$$

Whitening

We can compute an orthogonal observation sequence $\tilde{\boldsymbol{Y}}$ from $\boldsymbol{Y}$ recursively

- Given Y^i, we compute the error of the best linear MSE estimate of Y_{i+1},
$$\tilde{Y}_{i+1}(Y^i) = Y_{i+1} - \hat{Y}_{i+1}(Y^i)$$
- From the orthogonality principle, $\tilde{Y}_{i+1} \perp \mathrm{span}(\tilde{Y}_1, \tilde{Y}_2, \ldots, \tilde{Y}_i)$, hence we can write

$$\hat{Y}_{i+1}(Y^i) = \sum_{j=1}^{i} \hat{Y}_{i+1}(\tilde{Y}_j)$$

Interpretation

- $\hat{Y}_{i+1}$ is the part of Y_{i+1} predictable by Y^i, hence carries no useful new information for estimating X beyond Y^i
- $\tilde{Y}_{i+1}$ by comparison is the unpredictable part, hence carries new information
- As such, $\tilde{\boldsymbol{Y}}$ is called the innovation sequence of $\boldsymbol{Y}$
- Remark: If we normalize $\tilde{\boldsymbol{Y}}$ (by dividing each $\tilde{Y}_i$ by its standard deviation), we obtain the same sequence as using the Cholesky decomposition

Example

Let the observation sequence be $Y_i = X + Z_i$ for $i = 1, 2, \ldots, n$, where X, Z_1, …, Z_n are zero mean, uncorrelated r.v.s with $\mathbb{E}[X^2] = P$ and $\mathbb{E}[Z_i^2] = N$ for $i = 1, 2, \ldots, n$. Find the innovation sequence of $\boldsymbol{Y}$

$$\tilde{Y}_1 = Y_1,$$

$$\tilde{Y}_2 = Y_2 - \hat{Y}_2(Y_1) = Y_2 - \frac{\mathsf{Cov}(Y_1, Y_2)}{\mathsf{Var}(Y_1)} Y_1 = Y_2 - \frac{P}{P+N} Y_1,$$

$$\tilde{Y}_3 = Y_3 - \hat{Y}_3(Y^2) = Y_3 - c_3 \sum_{j=1}^{2} Y_j$$

Now, $(Y_3 - c_3 \sum_{j=1}^{2} Y_j) \perp \tilde{Y}_1$, which gives $c_3 = \frac{P}{2P+N}$. Hence

$$\tilde{Y}_3 = Y_3 - \frac{P}{2P+N} \sum_{j=1}^{2} Y_j$$

In general, $\tilde{Y}_{i+1} = Y_{i+1} - \frac{P}{iP+N} \sum_{j=1}^{i} Y_j$

Innovation sequence

- Using the innovation sequence, the MMSE linear estimate of X given $\tilde{Y}^{i+1}$ and its MSE can be computed recursively

$$\hat{X}(\tilde{Y}^{i+1}) = \hat{X}(\tilde{Y}^{i}) + \hat{X}(\tilde{Y}_{i+1}),$$

$$\mathsf{MSE}_{i+1} = \mathsf{MSE}_i - \frac{\mathsf{Cov}^2(X, \tilde{Y}_{i+1})}{\mathsf{Var}(\tilde{Y}_{i+1})}$$

- The innovation sequence will prove useful in deriving the Kalman filter

Kalman filter

The Kalman filter is an efficient, recursive algorithm for computing the MMSE linear estimate and its MSE when the signal $\boldsymbol{X}$ and observations $\boldsymbol{Y}$ evolve according to a state-space model

State-space model

- Consider a linear dynamical system described by the state-space model:

$$\boldsymbol{X}_{i+1} = \boldsymbol{A}_i \boldsymbol{X}_i + \boldsymbol{U}_i, \quad i = 0, 1, \ldots, n$$

  with noisy observations (output)

$$\boldsymbol{Y}_i = \boldsymbol{X}_i + \boldsymbol{V}_i, \quad i = 0, 1, \ldots, n,$$

  where $\boldsymbol{X}_0$, $\boldsymbol{U}_0, \boldsymbol{U}_1, \ldots, \boldsymbol{U}_n$, $\boldsymbol{V}_0, \boldsymbol{V}_1, \ldots, \boldsymbol{V}_n$ are zero mean, uncorrelated RVs with $\boldsymbol{\Sigma}_{\boldsymbol{X}_0} = \boldsymbol{P}_0$, $\boldsymbol{\Sigma}_{\boldsymbol{U}_i} = \boldsymbol{Q}_i$, $\boldsymbol{\Sigma}_{\boldsymbol{V}_i} = \boldsymbol{N}_i$; $\boldsymbol{A}_i$ is a known sequence of matrices

This state space model is used in many applications, e.g. navigation, computer vision (e.g. face tracking), economics, etc.

Kalman filter

The goal is to compute the MMSE linear estimate of the state from causal observations:

- Prediction: Find the estimate $\hat{\boldsymbol{X}}_{i+1|i}$ of $\boldsymbol{X}_{i+1}$ from $\boldsymbol{Y}^i$ and its MSE $\boldsymbol{\Sigma}_{i+1|i}$
- Filtering: Find the estimate $\hat{\boldsymbol{X}}_{i|i}$ of $\boldsymbol{X}_i$ from $\boldsymbol{Y}^i$ and its MSE $\boldsymbol{\Sigma}_{i|i}$

Kalman filter provides clever recursive equations for computing these estimates and their error covariance matrices

Scalar Kalman filter

Consider the scalar state space system:

$$X_{i+1} = a_i X_i + U_i, \quad i = 0, 1, \ldots, n$$

with noisy observations

$$Y_i = X_i + V_i, \quad i = 0, 1, \ldots, n,$$

where X_0, $U_0, U_1, \ldots, U_n$, $V_0, V_1, \ldots, V_n$ are zero mean, uncorrelated r.v.s with Var$(X_0) = P_0$, Var$(U_i) = Q_i$, Var$(V_i) = N_i$, and a_i is a known sequence

Scalar Kalman filter (prediction)

Initialization: $\hat{X}_{0|-1} = 0$, $\sigma^2_{0|-1} = P_0$

Update equations: For $i = 0, 1, 2, \ldots, n$, the estimate is

$$\hat{X}_{i+1|i} = a_i \hat{X}_{i|i-1} + k_i (Y_i - \hat{X}_{i|i-1}),$$

where the filter gain is

$$k_i = \frac{a_i \sigma^2_{i|i-1}}{\sigma^2_{i|i-1} + N_i}$$

The MSE of $\hat{X}_{i+1|i}$ is

$$\sigma^2_{i+1|i} = a_i (a_i - k_i) \sigma^2_{i|i-1} + Q_i$$

Example

Let $a_i = 1$, $Q_i = 0$, $N_i = N$, and $P_0 = P$ (so $X_0 = X_1 = \cdots = X$), and $Y_i = X + V_i$. The Kalman filter is given as follows:

Initialization: $\hat{X}_{0|-1} = 0$ and $\sigma^2_{0|-1} = P$
Update equation:

$$\hat{X}_{i+1|i} = (1 - k_i)\hat{X}_{i|i-1} + k_i Y_i$$

with

$$k_i = \frac{\sigma^2_{i|i-1}}{\sigma^2_{i|i-1} + N}$$

The MSE is

$$\sigma^2_{i+1|i} = (1 - k_i)\sigma^2_{i|i-1}$$

Example

We can solve for $\sigma^2_{i+1|i}$ explicitly

$$\sigma^2_{i+1|i} = \left(1 - \frac{\sigma^2_{i|i-1}}{\sigma^2_{i|i-1} + N}\right)\sigma^2_{i|i-1} = \frac{N\sigma^2_{i|i-1}}{\sigma^2_{i|i-1} + N}$$

$$\frac{1}{\sigma^2_{i+1|i}} = \frac{1}{N} + \frac{1}{\sigma^2_{i|i-1}}$$

$$\sigma^2_{i+1|i} = \frac{1}{i/N + 1/P} = \frac{NP}{iP + N}$$

The gain is

$$k_i = \frac{P}{iP + N}$$

The recursive estimate is

$$\hat{X}_{i+1|i} = \frac{(i-1)P + N}{iP + N}\hat{X}_{i|i-1} + \frac{P}{iP + N}Y_i$$

Derivation of Kalman filter

We use innovations. Let $\tilde{Y}_i$ be the innovation r.v. for Y_i, then we can write

$$\begin{aligned}\hat{X}_{i+1|i} &= \hat{X}_{i+1|i-1} + k_i\tilde{Y}_i,\\ \sigma^2_{i+1|i} &= \sigma^2_{i+1|i-1} + k_i\mathsf{Cov}(X_{i+1}, \tilde{Y}_i)\end{aligned}$$

where $\hat{X}_{i+1|i-1}$ and $\sigma^2_{i+1|i-1}$ are the MMSE linear estimate of X given Y^{i-1} and its MSE, and

$$k_i = \frac{\mathsf{Cov}(X_{i+1}, \tilde{Y}_i)}{\mathsf{Var}(\tilde{Y}_i)}$$

Derivation of Kalman filter

Now, since $X_{i+1} = a_i X_i + U_i$, by linearity of MMSE linear estimate, we have

$$\hat{X}_{i+1|i-1} = a_i \hat{X}_{i|i-1}$$

and

$$\sigma^2_{i+1|i-1} = a_i^2 \sigma^2_{i|i-1} + Q_i$$

Derivation of Kalman filter

Now, the innovation r.v. for Y_i is $\tilde{Y}_i = Y_i - \hat{Y}_i(Y^{i-1})$
Since $Y_i = X_i + V_i$ and V_i is uncorrelated with Y_j, $j = 1, 2, \ldots, i-1$,

$$\hat{Y}_i(Y^{i-1}) = \hat{X}_{i|i-1}$$

Hence,

$$\tilde{Y}_i = Y_i - \hat{X}_{i|i-1}$$

This yields

$$\begin{aligned}
\hat{X}_{i+1|i} &= a_i \hat{X}_{i|i-1} + k_i \tilde{Y}_i = a_i \hat{X}_{i|i-1} + k_i (Y_i - \hat{X}_{i|i-1}) \\
\sigma^2_{i+1|i} &= \sigma^2_{i+1|i-1} - k_i \mathsf{Cov}(X_{i+1}, \tilde{Y}_i),
\end{aligned}$$

Derivation of Kalman filter

Now, consider

$$
\begin{aligned}
k_i &= \frac{\mathsf{Cov}(X_{i+1}, \tilde{Y}_i)}{\mathsf{Var}(\tilde{Y}_i)} = \frac{\mathsf{Cov}(a_i X_i + U_i, X_i - \hat{X}_{i|i-1} + V_i)}{\mathsf{Var}(X_i - \hat{X}_{i|i-1} + V_i)} \\
&= \frac{\mathsf{Cov}(a_i X_i, X_i - \hat{X}_{i|i-1})}{\mathsf{Var}(X_i - \hat{X}_{i|i-1} + V_i)} = \frac{a_i \mathsf{Cov}(X_i, X_i - \hat{X}_{i|i-1})}{\mathsf{Var}(X_i - \hat{X}_{i|i-1} + V_i)} \\
&= \frac{a_i \mathsf{Cov}(X_i - \hat{X}_{i|i-1}, X_i - \hat{X}_{i|i-1})}{\mathsf{Var}(X_i - \hat{X}_{i|i-1} + V_i)} \quad \text{since } (X_i - \hat{X}_{i|i-1}) \perp \hat{X}_{i|i-1} \\
&= \frac{a_i \mathsf{Var}(X_i - \hat{X}_{i|i-1})}{\mathsf{Var}(X_i - \hat{X}_{i|i-1}) + N_i} = \frac{a_i \sigma^2_{i|i-1}}{\sigma^2_{i|i-1} + N_i}
\end{aligned}
$$

Derivation of Kalman filter

The MSE is

$$\begin{aligned}\sigma^2_{i+1|i} &= \sigma^2_{i+1|i-1} - k_i \mathsf{Cov}(a_i X_i + U_i, X_i - \hat{X}_{i|i-1} + V_i) \\ &= \sigma^2_{i+1|i-1} - k_i a_i \sigma^2_{i|i-1} \\ &= a_i(a_i - k_i)\sigma^2_{i|i-1} + Q_i\end{aligned}$$

This completes the derivation of the scalar Kalman filter

Vector Kalman Filter

The above scalar Kalman filter can be extended to the vector state space model:

- **Initialization:** $\hat{\boldsymbol{X}}_{0|-1} = \mathbf{0}$, $\boldsymbol{\Sigma}_{0|-1} = \boldsymbol{P}_0$

  Update equations: For $i = 0, 1, 2, \ldots, n$,

  $$\hat{\boldsymbol{X}}_{i+1|i} = \boldsymbol{A}_i \hat{\boldsymbol{X}}_{i|i-1} + \boldsymbol{K}_i (\boldsymbol{Y}_i - \hat{\boldsymbol{X}}_{i|i-1}),$$

  where the filter gain matrix

  $$\boldsymbol{K}_i = \boldsymbol{A}_i \boldsymbol{\Sigma}_{i|i-1} (\boldsymbol{\Sigma}_{i|i-1} + \boldsymbol{N}_i)^{-1}$$

  The covariance of the error is

  $$\boldsymbol{\Sigma}_{i+1|i} = \boldsymbol{A}_i \boldsymbol{\Sigma}_{i|i-1} \boldsymbol{A}_i^{\top} - \boldsymbol{K}_i \boldsymbol{\Sigma}_{i|i-1} \boldsymbol{A}_i^{\top} + \boldsymbol{Q}_i$$

Smoothing

Now assume the goal is to compute the MMSE linear estimate of X_i given Y^i, i.e. instead of predicting the next state, we are interested in estimating the current state

We denote this estimate by $\hat{X}_{i|i}$ and its MSE by $\sigma^2_{i|i}$

Smoothing

The Kalman filter can be adapted to this case as follows:
Initialization:

$$\hat{X}_{0|0} = \frac{P_0}{P_0 + N_0} Y_0$$
$$\sigma^2_{0|0} = \frac{P_0 N_0}{P_0 + N_0}$$

Update equations: For $i = 1, 2, \ldots, n$, the estimate is

$$\hat{X}_{i|i} = a_{i-1}(1 - k_i)\hat{X}_{i-1|i-1} + k_i Y_i$$

with filter gain

$$k_i = \frac{a^2_{i-1}\sigma^2_{i-1|i-1} + Q_{i-1}}{a^2_{i-1}\sigma^2_{i-1|i-1} + Q_{i-1} + N_i}$$

and MSE recursion

$$\sigma^2_{i|i} = (1 - k_i)\left(a^2_{i-1}\sigma^2_{i-1|i-1} + Q_{i-1}\right)$$

Smoothing: vector case

Initialization:

$$\hat{\boldsymbol{X}}_{0|0} = \boldsymbol{P}_0(\boldsymbol{P}_0 + \boldsymbol{N}_0)^{-1}\boldsymbol{Y}_0$$
$$\boldsymbol{\Sigma}_{0|0} = \boldsymbol{P}_0(\boldsymbol{I} - (\boldsymbol{P}_0 + \boldsymbol{N}_0)^{-1}\boldsymbol{P}_0)$$

Update equations: For $i = 1, 2, \ldots, n$, the estimate is

$$\hat{\boldsymbol{X}}_{i|i} = (\boldsymbol{I} - \boldsymbol{K}_i)\boldsymbol{A}_{i-1}\hat{\boldsymbol{X}}_{i-1|i-1} + \boldsymbol{K}_i\boldsymbol{Y}_i$$

with filter gain

$$\boldsymbol{K}_i = (\boldsymbol{A}_{i-1}\boldsymbol{\Sigma}_{i-1|i-1}\boldsymbol{A}_{i-1}^\top + \boldsymbol{Q}_{i-1})\left(\boldsymbol{A}_{i-1}\boldsymbol{\Sigma}_{i-1|i-1}\boldsymbol{A}_{i-1}^\top + \boldsymbol{Q}_{i-1} + \boldsymbol{N}_i\right)^{-1}$$

and MSE recursion

$$\boldsymbol{\Sigma}_{i|i} = (\boldsymbol{A}_{i-1}\boldsymbol{\Sigma}_{i-1|i-1}\boldsymbol{A}_{i-1}^\top + \boldsymbol{Q}_{i-1})(\boldsymbol{I} - \boldsymbol{K}_i^\top)$$

Reference

[1] "*Lecture notes for Statistical Signal Processing*," A. El Gamal.