Estimación y Predicción en Series Temporales

Kalman optimalidad, extensiones

Departamento de Procesamiento de Señales

Instituto de Ingeniería Eléctrica
Facultad de Ingeniería

2023

Optimalidad MMSE del filtro de Kalman

Introducción

- Por definición, Kalman es óptimo en sentido MSE
- Veremos que es óptimo cuando el ruido es Gaussiano

Esperanza condicional

- Sean $\mathcal{Y}_n = \{\mathbf{y}_0, \mathbf{y}_1, \ldots, \mathbf{y}_{n-1}\}$ las obs. hasta el tiempo n
- Consideremos el MSE condicionado a $\mathbf{y}_n$,

$$\mathbb{E}[\mathbf{e}_n^H \mathbf{e}_n | \mathcal{Y}_n], \quad \mathbf{e}_n = \mathbf{x}_n - \hat{\mathbf{x}}_n$$

- La idea es encontrar el estimador $\hat{\mathbf{x}}_n$ que la minimice

Minimización de la esperanza condicional

$$\begin{aligned}\mathbb{E}[\mathbf{e}_n^H \mathbf{e}_n|\mathcal{Y}_n] &= \mathbb{E}[(\mathbf{x}_n - \hat{\mathbf{x}}_n)^H(\mathbf{x}_n - \hat{\mathbf{x}}_n)|\mathcal{Y}_n] \\ &= \mathbb{E}[\mathbf{x}_n^H \mathbf{x}_n - \mathbf{x}_n^H \hat{\mathbf{x}}_n - \hat{\mathbf{x}}_n^H \mathbf{x}_n + \hat{\mathbf{x}}_n^H \hat{\mathbf{x}}_n|\mathcal{Y}_n].\end{aligned}$$

Ahora aplicamos la linealidad de $\mathbb{E}[\cdot]$.
Como $\hat{\mathbf{x}}_n$ queda determinado por $\mathcal{Y}_n$, $\mathbb{E}[\hat{\mathbf{x}}_n|\mathcal{Y}_n] = \hat{\mathbf{x}}_n$ y tenemos:

$$\mathbb{E}[\mathbf{e}_n^H \mathbf{e}_n|\mathcal{Y}_n] = \mathbb{E}[\mathbf{x}_n^H \mathbf{x}_n|\mathcal{Y}_n] - \mathbb{E}[\mathbf{x}_n^H|\mathcal{Y}_n]\hat{\mathbf{x}}_n - \hat{\mathbf{x}}_n^H \mathbb{E}[\mathbf{x}_n|\mathcal{Y}_n] + \hat{\mathbf{x}}_n^H \hat{\mathbf{x}}_n$$

Para minimizarla, derivamos en $\hat{\mathbf{x}}_n$ e igualamos a 0:

$$0 = -2\mathbb{E}[\mathbf{x}_n^H|\mathcal{Y}_n] + 2\hat{\mathbf{x}}_n \Rightarrow \hat{\mathbf{x}}_n = \mathbb{E}[\mathbf{x}_n|\mathcal{Y}_n]$$

Esta fórmula es general para cualquier estimador MMSE,
indep. del modelo y las densidades de los ruidos.

El caso Gaussiano

Supongamos

- $\bar{\mathbf{x}}_n$ es cond. óptimo en sentido MSE dado $\mathcal{Y}_n$
- Su matriz de covarianza es $\bar{\mathbf{P}}_n$

Entonces, por def. del MMSE:

$$f_{\mathbf{x}_n|\mathcal{Y}_n}(\mathbf{x}_n) = \mathcal{N}(\hat{\mathbf{x}}_n, \bar{\mathbf{P}}_n)$$

Al medir tenemos:

$$\mathbf{y}_n = \mathbf{C}_n\mathbf{x}_n + \mathbf{v}_n \Rightarrow f_{\mathbf{y}_n|\mathbf{x}_n}(\mathbf{y}_n) = \mathcal{N}(\mathbf{C}_n\mathbf{x}_n, \mathbf{R}_n)$$

Si, en lugar de $\mathbf{x}_n$, condicionamos respecto a $\bar{\mathbf{x}}_n$:

$$f_{\mathbf{y}_n|\bar{\mathbf{x}}_n}(\mathbf{y}_n) = \mathcal{N}(\mathbf{C}_n\bar{\mathbf{x}}_n, \mathbf{C}_n\bar{\mathbf{P}}_n\mathbf{C}_n^H + \mathbf{R}_n)$$

que refleja el ruido adicional debido a usar $\bar{\mathbf{x}}_n$ en lugar de $\mathbf{x}_n$.

Fórmula de de Bayes

Nos interesa el estimador óptimo luego (a posteriori) de observar $\mathbf{y}_n$:

$$f_{\mathbf{x}_k|\mathcal{Y}_n,\mathbf{y}_n} = \frac{f_{\mathbf{y}_n|\mathbf{x}_n,\cancel{\mathcal{Y}_n}} \cdot f_{\mathbf{x}_n|\mathcal{Y}_n}}{f_{\mathbf{y}_n|\mathcal{Y}_n}} = \frac{f_{\mathbf{y}_n|\mathbf{x}_n} \cdot f_{\mathbf{x}_n|\mathcal{Y}_n}}{f_{\mathbf{y}_n|\bar{\mathbf{x}}_n}}$$

Lo anterior surge de:

- $\mathbf{y}_n$ es cond. indep. de $\mathcal{Y}_n$ dado $\mathbf{x}_n$
- $\mathbf{y}_n$ es cond. indep. de $\mathcal{Y}_n$ dado $\bar{\mathbf{x}}_n$

Tenemos

$$
\begin{aligned}
f_{\mathbf{x}_n|\mathcal{Y}_n,\mathbf{y}_n} &= \frac{f_{\mathbf{y}_n|\mathbf{x}_n} \cdot f_{\mathbf{x}_n|\mathcal{Y}_n}}{f_{\mathbf{y}_n|\bar{\mathbf{x}}_n}} \\
f_{\mathbf{x}_n}(\mathbf{x}_n|\mathcal{Y}_n) &= \mathcal{N}(\hat{\mathbf{x}}_n, \bar{\mathbf{P}}_n) \\
f_{\mathbf{y}_n|\mathbf{x}_n}(\mathbf{y}_n) &= \mathcal{N}(\mathbf{C}_n\mathbf{x}_n, \mathbf{R}_n) \\
f_{\mathbf{y}_n|\bar{\mathbf{x}}_n}(\mathbf{y}_n) &= \mathcal{N}(\mathbf{C}_n\bar{\mathbf{x}}_n, \mathbf{C}_n\bar{\mathbf{P}}_n\mathbf{C}_n^H + \mathbf{R}_n)
\end{aligned}
$$

Sustituyendo y haciendo cuentas, se llega a:

$$
\begin{aligned}
\hat{\mathbf{x}}_n &= \mathbb{E}[\mathbf{x}_n|\mathcal{Y}_n] = \bar{\mathbf{x}}_n + \bar{\mathbf{P}}_n\mathbf{C}_n^H(\mathbf{C}_n\bar{\mathbf{P}}_n\mathbf{C}_n^H + \mathbf{R}_n)^{-1}(\mathbf{y}_n - \mathbf{C}_n\bar{\mathbf{x}}_n) \\
\mathbf{P}_n &= \mathrm{Cov}[\mathbf{x}_n|\mathcal{Y}_n] = (\bar{\mathbf{P}}_n^{-1} + \mathbf{C}_n^H\mathbf{R}^{-1}\mathbf{C}_n)^{-1}
\end{aligned}
$$

El caso Gaussiano
Comentarios

- La expresión de $\hat{\mathrm{x}}_n$ es idéntica a la de Kalman.
- Se puede demostrar que la covarianza del error también es equivalente a la de Kalman.
- **No se impuso que el filtro fuese lineal!**
- Eso surgió naturalmente de suponer:
  - ruidos Gaussianos
  - que el estimado sea la esperanza condicional

Conclusión:

En el caso Gaussiano, los filtros de Kalman/Wiener son óptimos aún incluyendo los filtros no-lineales

Kalman para sistemas dinámicos no lineales

Modelo lineal

$$\mathbf{x}_{n+1} = \mathbf{F}_n \mathbf{x}_n + \mathbf{v}_n$$
$$\mathbf{y}_n = \mathbf{C}_n \mathbf{x}_n + \mathbf{w}_n$$

En este caso:

- Ruido arbitrario: Kalman es el filtro lineal óptimo en sentido MMSE
- Ruido Gaussiano: Kalman es universalmente óptimo en sentido MMSE
- El modelo lineal incluye a los ARMA

Consideremos el modelo dinámico no lineal:

$$\mathbf{x}_{n+1} = \mathbf{f}_n(\mathbf{x}_n, \mathbf{v}_n)$$
$$\mathbf{y}_n = \mathbf{h}_n(\mathbf{x}_n, \mathbf{w}_n)$$

Qué se puede hacer en este caso?

- Extended Kalman Filter (EKF):
  1. linearizar modelo en tiempo n
  2. aplicar Kalman al modelo resultante
- Unscented Kalman Filter (UKF):
  1. mapeo no lineal de medias y matrices de covarianza
  2. Estimaciones: ponderación de muchos puntos
- El EKF es trivial; el UKF es más complicado

Extended Kalman Filter

Modelo no lineal:

$$\begin{aligned} \mathbf{x}_{n+1} &= \mathbf{f}_n(\mathbf{x}_n, \mathbf{v}_n) \\ \mathbf{y}_n &= \mathbf{h}_n(\mathbf{x}_n, \mathbf{w}_n) \end{aligned}$$

Linearización

- Ecuación de estado: entorno a $(\mathbf{x}, \mathbf{v}) = (\hat{\mathbf{x}}_n, 0)$.
- Ecuación de medida: entorno a $(\mathbf{x}, \mathbf{w}) = (\bar{\mathbf{x}}_n, 0)$.

$$\begin{aligned} \mathbf{x}_{n+1} &= \mathbf{f}_n(\hat{\mathbf{x}}_n, 0) + \mathbf{F}_n(\mathbf{x}_n - \hat{\mathbf{x}}_n) + \mathbf{G}_n\mathbf{v}_n \\ \mathbf{y}_n &= \mathbf{h}_n(\bar{\mathbf{x}}_n, 0) + \mathbf{H}_n(\mathbf{x}_n - \bar{\mathbf{x}}_n) + \mathbf{U}_n\mathbf{w}_n, \end{aligned}$$

donde

$$\begin{aligned} \mathbf{F}_n &= \frac{\partial \mathbf{f}_n}{\partial \mathbf{x}}(\hat{\mathbf{x}}_n, 0), & \mathbf{G}_n &= \frac{\partial \mathbf{f}_n}{\partial \mathbf{v}}(\hat{\mathbf{x}}_n, 0), \\ \mathbf{H}_n &= \frac{\partial \mathbf{h}_n}{\partial \mathbf{x}}(\bar{\mathbf{x}}_n, 0), & \mathbf{U}_n &= \frac{\partial \mathbf{h}_n}{\partial \mathbf{w}}(\bar{\mathbf{x}}_n, 0). \end{aligned}$$

- Inicialización: $\mathbf{x}_0$, $\mathbf{P}_0$.
- Para $n = 0, 1, 2, \ldots,$ calcular:

$$
\begin{aligned}
\bar{\mathbf{x}}_n &= \mathbf{f}_n(\hat{\mathbf{x}}_{n-1}, 0) \\
\bar{\mathbf{P}}_n &= \mathbf{F}_n \hat{\mathbf{P}}_{n-1} \mathbf{F}_n^T + \mathbf{G}_n \mathbf{Q}_n \mathbf{G}_n^T \\
\mathbf{K}_n &= \bar{\mathbf{P}}_n \mathbf{H}_n^T (\mathbf{H}_n \bar{\mathbf{P}}_n \mathbf{H}_n^T + \mathbf{U}_n \mathbf{R}_n \mathbf{U}_n^T)^{-1} \\
\hat{\mathbf{x}}_n &= \bar{\mathbf{x}}_n + \mathbf{K}_n (\mathbf{y}_n - \mathbf{h}_n(\bar{\mathbf{x}}_n, 0)) \\
\hat{\mathbf{P}}_n &= (\mathbf{I} - \mathbf{K}_n \mathbf{H}_n) \bar{\mathbf{P}}_n
\end{aligned}
$$

Observaciones

- Problema (i): linealización introduce un sesgo en $\hat{\mathbf{x}}_{n|k}$
- Problema (ii): linearización produce covarianza errónea
- Cuanto menor sean $\|\mathbf{x} - \hat{\mathbf{x}}_n\|^2$ y $\|\mathbf{x} - \bar{\mathbf{x}}_n\|^2$, mejor
- Es de esperar que el EKF sea mejor si SNR es alto
- Ver [Anderson & Moore] para más detalles

Unscented Kalman Filter

Julier & Uhlmann 1996

- Evitar linearización (y sus sesgos)
- Aproximar $\hat{\mathbf{y}}$ y $\mathbf{P}_{\mathbf{y}}$ de manera insesgada
- Capturar geometría de $\mathbf{h}(\cdot), \mathbf{f}(\cdot)$ mediante muestreo
- El muestreo es determinístico
- Se elige de manera de preservar estadísticas relevantes

- Se asume $\mathbf{x} \sim \mathcal{N}(\bar{\mathbf{x}}, \mathbf{P_x})$
- Se eligen $2n+1$ puntos $\mathcal{X}_0, \mathcal{X}_1, \ldots, \mathcal{X}_{2n+1}$ tales que su covarianza empírica igual a la covarianza de $\mathbf{x}$
  - estos son llamados σ-puntos
  - Cada punto tiene un peso W_i
- Se obtienen $2n+1$ puntos $\mathcal{Y}_i = \mathbf{h}(\mathcal{X}_i)$
- Se calculan media y covarianza empíricas:

$$\bar{\mathbf{y}} = \sum_i W_i \mathcal{Y}_i$$

$$\mathbf{P_y} = \sum_i W_i (\mathcal{Y}_i - \bar{\mathbf{y}})(\mathcal{Y}_i - \bar{\mathbf{y}})^\intercal$$

- Éstas se insertan en las ecuaciones de Kalman

- A los efectos del muestreo, se considera la variable conjunta $\mathbf{x}^a = [\mathbf{x},\ \mathbf{v},\ \mathbf{w}]$, cuya varianza es:

$$\mathbf{P}^a = \begin{bmatrix} \hat{\mathbf{P}}_{n-1} & 0 & 0 \\ 0 & \mathbf{Q}_n & 0 \\ 0 & 0 & \mathbf{R}_n \end{bmatrix}$$

- Los puntos $\mathcal{X}_i$ se toman en base a $\mathbf{P}^a_n$
- Luego se desarma $\mathcal{X}_i = [\mathcal{X}_i^x,\ \mathcal{X}_i^v, \mathcal{X}_i^w]$

Unscented Kalman Filter (UKF)

Algoritmo

1. Inicialización: $\hat{\mathbf{x}}_{n-1}$, $\hat{\mathbf{P}}_{n-1}$
2. Para $n = 1, 2, \ldots$

$$
\begin{aligned}
\mathcal{X}_i &= [\mathcal{X}_i^x, \mathcal{X}_i^v, \mathcal{X}_i^w] \;\; i = 0, \ldots, 2n \;(\sigma - \text{puntos}) \\
\bar{\mathcal{X}}_i^x &= \mathbf{f}_n(\mathcal{X}_i^x, \mathcal{X}_i^v)\,; \quad \bar{\mathbf{x}}_n = \sum_i W_i \bar{\mathcal{X}}_i^x \\
\bar{\mathbf{P}}_n^{\mathbf{x}} &= \sum_i W_i (\bar{\mathcal{X}}_i^x - \bar{\mathbf{x}}_n)(\bar{\mathcal{X}}_i^x - \bar{\mathbf{x}}_n)^T \\
\bar{\mathcal{Y}}_i &= \mathbf{h}(\bar{\mathcal{X}}_i^x, \bar{\mathcal{X}}_i^w)\,; \quad \bar{\mathbf{y}}_n = \sum_i W_i \bar{\mathcal{Y}}_i \\
\bar{\mathbf{P}}_n^{\mathbf{y}} &= \sum_i W_i (\bar{\mathcal{Y}}_i - \bar{\mathbf{y}}_n)(\bar{\mathcal{Y}}_i - \bar{\mathbf{y}}_n)^\intercal \\
\bar{\mathbf{P}}_n^{\mathbf{xy}} &= \sum_i W_i (\bar{\mathcal{X}}_i^x - \bar{\mathbf{x}}_n)(\bar{\mathcal{Y}}_i - \bar{\mathbf{y}}_n)^\intercal \\
\mathbf{K}_n &= \bar{\mathbf{P}}_n^{\mathbf{xy}} (\bar{\mathbf{P}}_n^{\mathbf{y}})^{-1} \\
\hat{\mathbf{x}}_n &= \bar{\mathbf{x}}_n + \mathbf{K}_n(\mathbf{y}_n - \bar{\mathbf{y}}_n) \\
\hat{\mathbf{P}}_n &= \bar{\mathbf{P}}_n - \mathbf{K}_n \bar{\mathbf{P}}_n^{\mathbf{y}} \mathbf{K}_n^\intercal.
\end{aligned}
$$

UKF: elección de los σ-puntos

- Sup. $\mathbf{x} \in \mathbb{R}^n$, $\mathbf{x} \sim \mathcal{N}(\bar{\mathbf{x}}, \mathbf{P}_{\mathbf{x}})$.
- Sea $\mathbf{L}$ una factorización de $\mathbf{P}_{\mathbf{x}}$: $\mathbf{P}_{\mathbf{x}} = \mathbf{L}\mathbf{L}^\intercal$ (ej. Cholesky)
- Sea $\kappa > 0$ un escalar a elegir

Los σ-puntos y sus pesos se definen como:

$$
\begin{aligned}
\mathcal{X}_0 &= \bar{\mathbf{x}} & W_0 &= \frac{\kappa}{\kappa + n} \\
\mathcal{X}_{2i-1} &= \bar{\mathbf{x}} - \sqrt{\kappa + n} \cdot \mathbf{L}_i & W_{2i-1} = W_{2i} &= \frac{1}{2(\kappa + n)} \\
\mathcal{X}_{2i} &= \bar{\mathbf{x}} + \sqrt{\kappa + n} \cdot \mathbf{L}_i & i &= 1, \dots, n.
\end{aligned}
$$

donde $\mathbf{L}_i$ es la i-ésima columna de $\mathbf{L}$.

Justificación:
Se puede ver que con esta elección la media y covarianza empíricas de $\bar{\mathcal{Y}}$ coinciden con las de $\bar{\mathbf{y}}$ hasta tercer orden.