Estimación y Predicción en Series Temporales

Filtro de Kalman

Departamento de Procesamiento de Señales

Instituto de Ingeniería Eléctrica
Facultad de Ingeniería

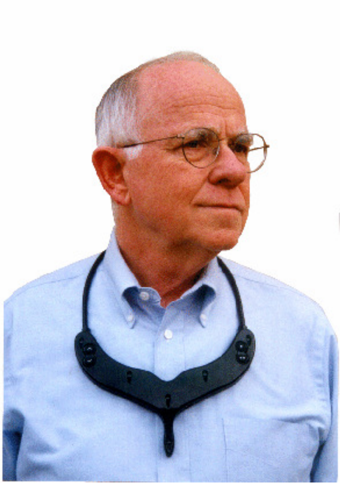
2023

Biblio:

- Brown and Hwang, *Introduction to Random Signals and Applied Kalman Filtering*, 3ra edición, 1996.
- Anderson and Moore, *Optimal Filtering*, 1979 (re-edición Dover 2005).

Motivación

Óptimo y elegante, pero no adaptivo.



Adaptivo y rápido, pero lento y ruidoso.

Recapitulación
Filtro RLS (vox populi)

Adaptivo, rápido y preciso, pero más costoso.

Adaptivo, rápido, preciso, más general. Más complicado.

Filtros adaptivos clásicos (LMS, RLS)
Modelo: relación lineal entre entrada y salida:

$$\hat{\mathbf{x}}_n = \hat{\mathbf{w}}^\intercal \hat{\mathbf{u}}_n$$

Filtro de Kalman
El modelo es un sistema dinámico lineal

$$\begin{aligned}\mathbf{x}_{n+1} &= \mathbf{F}_n \mathbf{x}_n + \mathbf{w}_n \,,\; \mathbf{x}_n \in \mathbb{R}^M, \mathbf{F}_n \in \mathbb{R}^{M\times M},\; \mathbf{w}_n \sim \mathcal{N}(0, \mathbf{Q}_n)\\ \mathbf{y}_n &= \mathbf{C}_n \mathbf{x}_n + \mathbf{v}_n \,,\; \mathbf{y}_n \in \mathbb{R}^N, \mathbf{C}_n \in \mathbb{R}^{N\times M},\; \mathbf{v}_n \sim \mathcal{N}(0, \mathbf{R}_n)\end{aligned}$$

- $\mathbf{x}_n$ es el estado del sistema (no observable)
- $\mathbf{y}_n$ son sus salidas observables
- $\mathbf{w}_n$, $\mathbf{v}_n$ ruido en estado y observación

Dado el sistema dinámico lineal:

$$\mathbf{x}_{n+1} = \mathbf{F}_n\mathbf{x}_n + \mathbf{w}_n\,,\ \mathbf{x}_n \in \mathbb{R}^M, \mathbf{F}_n \in \mathbb{R}^{M\times M},\ \mathbf{w}_n \sim \mathcal{N}(0, \mathbf{Q}_n)$$
$$\mathbf{y}_n = \mathbf{C}_n\mathbf{x}_n + \mathbf{v}_n\,,\ \mathbf{y}_n \in \mathbb{R}^N, \mathbf{C}_n \in \mathbb{R}^{N\times M},\ \mathbf{v}_n \sim \mathcal{N}(0, \mathbf{R}_n)$$

- Deseamos estimar $\mathbf{x}$
- Estimación a priori de $\mathbf{x}$: dinámica del sistema ($\mathbf{F}$)
- Refinamiento: incorpora información provista por $\mathbf{y}_n$.

Filtro de Kalman

Qué agrega el filtro de Kalman?

- Adaptivo y rápido, como RLS
- Muchísimo más general: no se limita a modelos AR/MA

Impacto Enorme

- No habría ni satélites, ni exploración espacial
- Tampoco misiles balísticos, pero bueh.
- Tracking de todo tipo (desde touchpads a GPS)

Planteo

$$\mathbf{x}_{n+1} = \mathbf{F}_n \mathbf{x}_n + \mathbf{w}_n \, , \; \mathbf{x}_n \in \mathbb{R}^M, \mathbf{F}_n \in \mathbb{R}^{M \times M}, \; \mathbf{w}_n \sim \mathcal{N}(0, \mathbf{Q}_n)$$
$$\mathbf{y}_n = \mathbf{C}_n \mathbf{x}_n + \mathbf{v}_n \, , \; \mathbf{y}_n \in \mathbb{R}^N, \mathbf{C}_n \in \mathbb{R}^{N \times M}, \; \mathbf{v}_n \sim \mathcal{N}(0, \mathbf{R}_n)$$

- $\mathbf{x}_n$ es un vector de estados del sistema
- $\mathbf{y}_n$ son observaciones del sistema
- $\mathbf{w}_n \in \mathcal{N}(0, \mathbf{Q}_n)$,$\mathbf{v}_n \in \mathcal{N}(0, \mathbf{R}_n)$, independientes
- $\mathbf{w}_n$ y $\mathbf{v}_n$ mutualmente independientes (ortogonales)

Todo proceso ARMA puede ser modelado de esta manera

En el tiempo n se tiene

- Estimación a priori de $\mathbf{x}_n$, $\bar{\mathbf{x}}_n \approx \mathbb{E}\left[\, \mathbf{x}_n \,\right]$
- Covarianza de su error: $\bar{\mathbf{P}}_n \approx \mathbb{E}\left[\, \bar{\mathbf{e}}_n^{\intercal} \bar{\mathbf{e}}_n \,\right]\,, \quad \bar{\mathbf{e}}_n = \mathbf{x}_n - \bar{\mathbf{x}}_n$

Idea: refinar usando observación $\mathbf{y}_n$

$$\hat{\mathbf{x}}_n = \bar{\mathbf{x}}_n + \mathbf{K}_n(\mathbf{y}_n - \mathbf{C}_n \bar{\mathbf{x}}_n)$$

Objetivo: hallar $\mathbf{K}_n$ que minimice

$$\mathbb{E}\left[\, \|\mathbf{e}_n\|^2 \,\right], \mathbf{e}_n = \mathbf{x}_n - \hat{\mathbf{x}}_n$$

La matrix $\mathbf{K}_n$ se llama ganancia de Kalman

Desarrollo

Objetivo:

- Encontrar $\mathbf{K}_n$ que minimice $\mathbb{E}\left[\,\|\mathbf{e}_n\|^2\,\right]$
- Podemos reescribirlo de manera equivalente como:

$$\mathbf{K}_n = \arg\min_{\mathbf{K}}\ \mathrm{Tr}(\mathbf{P}_n)$$

en donde $\mathbf{P}_n$ es la matriz de covarianza del error efectivo

$$\mathbf{P}_n = \mathbb{E}[\mathbf{e}_n\mathbf{e}_n^\intercal] = \mathbb{E}[(\mathbf{x}_n - \hat{\mathbf{x}}_n)(\mathbf{x}_n - \hat{\mathbf{x}}_n)^\intercal].$$

Notar que $(\mathbf{x}_n - \hat{\mathbf{x}}_n)^\intercal(\mathbf{x}_n - \hat{\mathbf{x}}_n) = \mathrm{Tr}((\mathbf{x}_n - \hat{\mathbf{x}}_n)(\mathbf{x}_n - \hat{\mathbf{x}}_n)^\intercal)$

Desarrollo (ii)

Objetivo: hallar $\mathbf{K}_n$ que minimice $\mathbf{Tr}(\mathbf{P}_n)$:

$$\mathbf{P}_n = \mathbb{E}[\mathbf{e}_n \mathbf{e}_n^\intercal] = \mathbb{E}[(\mathbf{x}_n - \hat{\mathbf{x}}_n)(\mathbf{x}_n - \hat{\mathbf{x}}_n)^\intercal].$$

Tenemos:

$$\begin{aligned} \text{i)}\ \hat{\mathbf{x}}_n &= \bar{\mathbf{x}}_n + \mathbf{K}_n(\mathbf{y}_n - \mathbf{C}_n \bar{\mathbf{x}}_n), \\ \text{ii)}\ \mathbf{y}_n &= \mathbf{C}_n \mathbf{x}_n + \mathbf{v}_n. \end{aligned}$$

Combinando i e ii:

$$\hat{\mathbf{x}}_n = \bar{\mathbf{x}}_n + \mathbf{K}_n(\mathbf{C}_n \mathbf{x}_n + \mathbf{v}_n - \mathbf{C}_n \bar{\mathbf{x}}_n)$$

Sustituyendo en el error efectivo:

$$\begin{aligned} \mathbf{e}_n &= \mathbf{x}_n - \bar{\mathbf{x}}_n - \mathbf{K}_n(\mathbf{C}_n \mathbf{x}_n + \mathbf{v}_n - \mathbf{C}_n \bar{\mathbf{x}}_n) \\ &= (\mathbf{I} - \mathbf{K}_n \mathbf{C}_n)(\mathbf{x}_n - \bar{\mathbf{x}}_n) - \mathbf{K}_n \mathbf{v}_n \\ &= (\mathbf{I} - \mathbf{K}_n \mathbf{C}_n)\bar{\mathbf{e}}_n - \mathbf{K}_n \mathbf{v}_n \end{aligned}$$

Desarrollo (iii)

Objetivo: hallar $\mathbf{K}_n$ que minimice $\mathbf{Tr}(\mathbf{P}_n)$:

$$\mathbf{P}_n = \mathbb{E}[\mathbf{e}_n \mathbf{e}_n^\intercal]$$

donde

$$\mathbf{e}_n = (\mathbf{I} - \mathbf{K}_n \mathbf{C}_n)\bar{\mathbf{e}}_n - \mathbf{K}_n \mathbf{v}_n$$

Ahora: Si sustituimos la expresión de arriba y asumimos

$$\text{(i) } \mathbb{E}[\mathbf{e}^\intercal \mathbf{v}_n^\intercal] = 0, \quad \text{(ii) } \bar{\mathbf{P}}_n = \mathbb{E}[\bar{\mathbf{e}}_n \bar{\mathbf{e}}_n^\intercal]$$

llegamos a:

$$\begin{aligned}
\mathbb{E}[\mathbf{e}_n \mathbf{e}_n^\intercal] &= (\mathbf{I} - \mathbf{K}_n \mathbf{C}_n)\mathbb{E}[\bar{\mathbf{e}}_n \bar{\mathbf{e}}_n^\intercal](\mathbf{I} - \mathbf{K}_n \mathbf{C}_n)^\intercal + \mathbf{K}_n \mathbb{E}[\mathbf{v}_n \mathbf{v}_n^\intercal]\mathbf{K}_n^\intercal \\
&= (\mathbf{I} - \mathbf{K}_n \mathbf{C}_n)\bar{\mathbf{P}}_n(\mathbf{I} - \mathbf{K}_n \mathbf{C}_n)^\intercal + \mathbf{K}_n \mathbf{R}_n \mathbf{K}_n^\intercal \\
&= \bar{\mathbf{P}}_n - \mathbf{K}_n \mathbf{C}_n \bar{\mathbf{P}}_n - \bar{\mathbf{P}}_n^\intercal \mathbf{C}_n^\intercal \mathbf{K}_n^\intercal + \mathbf{K}_n(\mathbf{C}_n \bar{\mathbf{P}}_n \mathbf{C}_n^\intercal + \mathbf{R}_n)\mathbf{K}_n^\intercal.
\end{aligned}$$

Derivación (iv)

Queremos minimizar:

$$\mathbf{Tr}(\mathbf{P}_n) = \mathbf{Tr}\left[\bar{\mathbf{P}}_n - \mathbf{K}_n\mathbf{C}_n\bar{\mathbf{P}}_n - \bar{\mathbf{P}}_n^\intercal\mathbf{C}_n^\intercal\mathbf{K}_n^\intercal + \mathbf{K}_n(\mathbf{C}_n\bar{\mathbf{P}}_n\mathbf{C}_n^\intercal + \mathbf{R}_n)\mathbf{K}_n^\intercal.\right]$$

Como la traza es *lineal*:

$$\mathbf{Tr}(\mathbf{P}_n) = \mathbf{Tr}\left[\bar{\mathbf{P}}_n\right] - \mathbf{Tr}\left[\mathbf{K}_n\mathbf{C}_n\bar{\mathbf{P}}_n\right] - \mathbf{Tr}\left[\bar{\mathbf{P}}_n^\intercal\mathbf{C}_n^\intercal\mathbf{K}_n^\intercal\right] + \mathbf{Tr}\left[\mathbf{K}_n(\mathbf{C}_n\bar{\mathbf{P}}_n\mathbf{C}_n^\intercal + \mathbf{R}_n)\mathbf{K}_n^\intercal.\right]$$

Ahora usando $\mathbf{Tr}(A^\intercal) = \mathbf{Tr}(A)$,

$$\mathrm{Tr}[\bar{\mathbf{P}}_n^\intercal\mathbf{C}_n^\intercal\mathbf{K}_n^\intercal] = \mathrm{Tr}[\mathbf{K}_n\mathbf{C}_n\bar{\mathbf{P}}_n] :$$

$$\mathbf{Tr}(\mathbf{P}_n) = \mathbf{Tr}\left[\bar{\mathbf{P}}_n\right] - 2\mathbf{Tr}\left[\mathbf{K}_n\mathbf{C}_n\bar{\mathbf{P}}_n\right] + \mathbf{Tr}\left[\mathbf{K}_n(\mathbf{C}_n\bar{\mathbf{P}}_n\mathbf{C}_n^\intercal + \mathbf{R}_n)\mathbf{K}_n^\intercal.\right]$$

Para hallar el $\mathbf{K}_n$ óptimo, basta derivar e igualar a cero:

$$\frac{\partial\mathbf{Tr}(\mathbf{P}_n)}{\partial\mathbf{K}_n} = 0$$

Tenemos que hallar $\mathbf{P}_n$ tal que:

$$\frac{\partial \mathbf{Tr}(\mathbf{P}_n)}{\partial \mathbf{K}_n} = \frac{\partial \mathbf{Tr}\left[\bar{\mathbf{P}}_n\right]}{\partial \mathbf{K}_n} - 2\frac{\partial \mathbf{Tr}\left[\mathbf{K}_n\mathbf{C}_n\bar{\mathbf{P}}_n\right]}{\partial \mathbf{K}_n} + \frac{\partial \mathbf{Tr}\left[\mathbf{K}_n(\mathbf{C}_n\bar{\mathbf{P}}_n\mathbf{C}_n^\intercal + \mathbf{R}_n)\mathbf{K}_n^\intercal.\right]}{\partial \mathbf{K}_n} = 0$$

Ahora usamos las siguientes propiedades:

$$\frac{\partial}{\partial \mathbf{A}}\mathrm{Tr}[\mathbf{AB}] = \mathbf{B}^\intercal \quad \frac{\partial}{\partial \mathbf{A}}\mathrm{Tr}[\mathbf{ACA}^\intercal] = 2\mathbf{AC}$$

con $\mathbf{A}, \mathbf{B}$ cuadradas y $\mathbf{C}$ simétrica. Llegamos a:

$$\frac{\partial \mathbf{Tr}(\mathbf{P}_n)}{\partial \mathbf{K}_n} = -2\bar{\mathbf{P}}_n^\intercal\mathbf{C}_n^\intercal + 2\mathbf{K}_n(\mathbf{C}_n\bar{\mathbf{P}}_n\mathbf{C}_n^\intercal + \mathbf{R}_n) = 0$$

Finalmente despejamos (usando $\mathbf{P}_n^\intercal = \mathbf{P}_n$):

$$\mathbf{K}_n = \bar{\mathbf{P}}_n\mathbf{C}_n^\intercal\left(\mathbf{C}_n\bar{\mathbf{P}}_n\mathbf{C}_n^\intercal + \mathbf{R}_n\right)^{-1}$$

Matriz de convarianza óptima

Una vez hallado:

$$\mathbf{K}_n = \bar{\mathbf{P}}_n \mathbf{C}_n^\intercal \left(\mathbf{C}_n \bar{\mathbf{P}}_n \mathbf{C}_n^\intercal + \mathbf{R}_n \right)^{-1}$$

Podemos sustituirlo en la expresión de $\mathbf{P}_n$:

$$\mathbf{P}_n = (\mathbf{I} - \mathbf{K}_n \mathbf{C}_n) \bar{\mathbf{P}}_n (\mathbf{I} - \mathbf{K}_n \mathbf{C}_n)^\intercal + \mathbf{K}_n \mathbf{R}_n \mathbf{K}_n^\intercal$$

Luego de unas cuentas sencillas llegamos a:

$$\mathbf{P}_n = (\mathbf{I} - \mathbf{K}_n \mathbf{C}_n) \bar{\mathbf{P}}_n$$

Paso predictivo

- Asumimos conocidas las predicciones $\bar{\mathbf{x}}_n$, $\bar{\mathbf{P}}_n$
- Resta actualizarlas para usar en el paso $n+1$, $\bar{\mathbf{x}}_{n+1}$
- Predecimos $\bar{\mathbf{x}}_n$ como su valor esperado dado $\hat{\mathbf{x}}_n$:

$$\bar{\mathbf{x}}_{n+1} = \mathbb{E}\left[\, \mathbf{x}_{n+1} \,\middle|\, \hat{\mathbf{x}}_n \,\right] = \mathbf{F}_n \hat{\mathbf{x}}_n$$

  porque $\mathbf{w}_n$ es blanco y de media nula.
- Ahora la covarianza del error de predicción:

$$\begin{aligned} \bar{\mathbf{e}}_{n+1} &= \mathbf{x}_{n+1} - \bar{\mathbf{x}}_{n+1} = (\mathbf{F}_n \mathbf{x}_n + \mathbf{w}_n) - \mathbf{F}_n \hat{\mathbf{x}}_n \\ &= \mathbf{F}_n \mathbf{e}_n + \mathbf{w}_n \end{aligned}$$

  Luego,

$$\begin{aligned} \bar{\mathbf{P}}_{n+1} &= \mathbb{E}[\bar{\mathbf{e}}_{n+1} \bar{\mathbf{e}}_{n+1}^{\intercal}] \\ &= \mathbb{E}[(\mathbf{F}_n \mathbf{e}_n + \mathbf{w}_n)(\mathbf{F}_n \mathbf{e}_n + \mathbf{w}_n^{\intercal}] \\ &= \mathbf{F}_n \mathbf{P}_n \mathbf{F}_n^{\intercal} + \mathbf{Q}_n \end{aligned}$$

Modelo

$$\begin{aligned}\mathbf{x}_{n+1} &= \mathbf{F}_n \mathbf{x}_n + \mathbf{R}_n \mathbf{w}_n \\ \mathbf{y}_n &= \mathbf{C}_n \mathbf{x}_n + \mathbf{Q}_n \mathbf{v}_n\end{aligned}$$

Inicialización

$$\bar{\mathbf{x}}_0, \quad \bar{\mathbf{P}}_0$$

Recursión

1. Gananacia de Kalman: $\mathbf{K}_n = \bar{\mathbf{P}}_n \mathbf{C}_n^\intercal (\mathbf{C}_n \bar{\mathbf{P}}_n \mathbf{C}_n^\intercal + \mathbf{R}_n)^{-1}$
2. Corrección del estimador: $\hat{\mathbf{x}}_n = \bar{\mathbf{x}}_n + \mathbf{K}_n(\mathbf{y}_n - \mathbf{C}_n \bar{\mathbf{x}}_n)$
3. Corrección de covarianza: $\mathbf{P}_n = (\mathbf{I} - \mathbf{K}_n \mathbf{C}_n)\bar{\mathbf{P}}_n$
4. Proyección a $n+1$:

$$\bar{\mathbf{x}}_{n+1} = \mathbf{F}_n \hat{\mathbf{x}}_n, \quad \bar{\mathbf{P}}_{n+1} = \mathbf{F}_n \mathbf{P}_n \mathbf{F}_n^\intercal + \mathbf{Q}_n$$

Filtro de Kalman

Esquema recursivo

$$
\begin{aligned}
\bar{\mathbf{x}}_n &= \mathbf{F}_{n-1}\hat{\mathbf{x}}_{n-1} \\
\bar{\mathbf{P}}_n &= \mathbf{F}_{n-1}\mathbf{P}_{n-1}\mathbf{F}_{n-1}^\top + \mathbf{Q}_{n-1}
\end{aligned}
$$

$$
\mathbf{K}_n = \bar{\mathbf{P}}_n\mathbf{C}_n^\top(\mathbf{C}_n^\top\mathbf{P}_n\mathbf{C}_n^\top + \mathbf{R}_n)^{-1}
$$

$$
\begin{aligned}
\hat{\mathbf{x}}_n &= \bar{\mathbf{x}}_n + \mathbf{K}_n(\mathbf{y}_n - \mathbf{C}_n\bar{\mathbf{x}}_n) \\
\mathbf{P}_n &= (\mathbf{I} - \mathbf{K}_n\mathbf{C}_n)\bar{\mathbf{P}}_n
\end{aligned}
$$

$$
n \leftarrow n + 1
$$