

Estimación y Predicción en Series Temporales

Optimalidad basada en la densidad condicional y filtro de Kalman

Departamento de Procesamiento de Señales

Instituto de Ingeniería Eléctrica
Facultad de Ingeniería

2022

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).

Introducción - Motivación

En la derivación del filtro de Kalman presentada en la clase anterior nos restringimos a:

- Filtros lineales
- Error cuadrático medio como índice de performance.

Hoy veremos que en el caso que los ruidos de estado y de medida son Gaussianos, los resultados son más generales.

Esperanza condicional

Sean $\mathcal{Y}_k = \{\mathbf{y}_0, \mathbf{y}_1, \dots, \mathbf{y}_k\}$ las medidas tomadas hasta el tiempo k .

Calculemos la norma del error de estimación $\mathbf{e}_k = \mathbf{x}_k - \hat{\mathbf{x}}_k$, conocido \mathcal{Y}_k , i.e. $\mathbb{E}[\mathbf{e}_k^H \mathbf{e}_k | \mathcal{Y}_k]$. La idea es encontrar el estimador $\hat{\mathbf{x}}_k$ que la minimice.

Minimización de la esperanza condicional

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

Ejercicio: observando que todo estimador es función de los datos, i.e. $\hat{\mathbf{x}}_k = \text{fct}(\mathcal{Y}_k)$, demostrar que el estimador óptimo es $\hat{\mathbf{x}}_k = \mathbb{E}[\mathbf{x}_k | \mathcal{Y}_k]$. Como el estimador es función de los datos,

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

El término en azul es el único que depende de $\hat{\mathbf{x}}_k$, y es inmediato ver que el MSE se minimiza si y sólo si

$$\hat{\mathbf{x}}_k = \mathbb{E}[\mathbf{x}_k | \mathcal{Y}_k]$$

Esta fórmula es general para cualquier estimador que minimiza el MSE, independientemente del modelo y las densidades de los ruidos.

El caso Gaussiano

Suponemos ahora que tenemos un estimador óptimo a priori $\hat{\mathbf{x}}_k^-$ con matriz de covarianza de error \mathbf{P}_k^- asociada.

Consideramos la variable \mathbf{x}_k , que por definición ya está implícitamente condicionada por las medidas \mathcal{Y}_{k-1} .

- Densidad de probabilidad de \mathbf{x}_k : $f_{\mathbf{x}_k}(\mathbf{x}_k) = \mathcal{N}(\mathbf{x}_k; \hat{\mathbf{x}}_k^-, \mathbf{P}_k^-)$
- $\mathbf{y}_k = \mathbf{C}_k \mathbf{x}_k + \mathbf{v}_k \Rightarrow$
$$\begin{cases} f_{\mathbf{y}_k}(\mathbf{y}_k) = \mathcal{N}(\mathbf{y}_k; \mathbf{C}_k \hat{\mathbf{x}}_k^-, \mathbf{C}_k \mathbf{P}_k^- \mathbf{C}_k^H + \mathbf{R}_k) \\ f_{\mathbf{y}_k|\mathbf{x}_k}(\mathbf{y}_k) = \mathcal{N}(\mathbf{y}_k; \mathbf{C}_k \mathbf{x}_k, \mathbf{R}_k) \end{cases}$$

Usando la fórmula de Bayes tenemos

$$f_{\mathbf{x}_k|\mathcal{Y}_k} = f_{\mathbf{x}_k|\mathbf{y}_k} = \frac{f_{\mathbf{y}_k|\mathbf{x}_k} f_{\mathbf{x}_k}}{f_{\mathbf{y}_k}} = \frac{\mathcal{N}(\mathbf{y}_k; \mathbf{C}_k \mathbf{x}_k, \mathbf{R}_k) \mathcal{N}(\mathbf{x}_k; \hat{\mathbf{x}}_k^-, \mathbf{P}_k^-)}{\mathcal{N}(\mathbf{y}_k; \mathbf{C}_k \hat{\mathbf{x}}_k^-, \mathbf{C}_k \mathbf{P}_k^- \mathbf{C}_k^H + \mathbf{R}_k)}$$

Haciendo cuentas (**ejercicio**) se obtiene para la variable $\mathbf{x}_k|\mathcal{Y}_k$:

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

Tenemos entonces:

$$\hat{\mathbf{x}}_k = \hat{\mathbf{x}}_k^- + \mathbf{P}_k^- \mathbf{C}_k^H (\mathbf{C}_k \mathbf{P}_k^- \mathbf{C}_k^H + \mathbf{R}_k)^{-1} (\mathbf{y}_k - \mathbf{C}_k \hat{\mathbf{x}}_k^-)$$
$$\mathbf{P}_k = ((\mathbf{P}_k^-)^{-1} + \mathbf{C}_k^H \mathbf{R}^{-1} \mathbf{C}_k)^{-1}$$

Observaciones:

- La expresión para la actualización del estimador de \mathbf{x}_k es idéntica a la del filtro de Kalman.
- Se puede demostrar que la expresión de covarianza del error de estimación en k es equivalente a la obtenida para el filtro de Kalman.
- Como resultado de la optimización vemos que la elección de actualizar el estimador $\hat{\mathbf{x}}_k$ como el promedio de la predicción $\hat{\mathbf{x}}_k^-$ y del error de predicción de la observación \mathbf{y}_k ponderados por la ganancia de Kalman, $\hat{\mathbf{x}}_k = \hat{\mathbf{x}}_k^- + \mathbf{K}_k (\mathbf{y}_k - \mathbf{C}_k \hat{\mathbf{x}}_k^-)$, fue acertada, al menos para el caso Gaussiano.
- En ningún momento impusimos que el filtro fuese lineal. Surgió naturalmente de imponer: (i) Ruidos Gaussianos:

Modelo de estado a partir de un modelo ARMA

$$y(k+n) + \alpha_{n-1}y(k+n-1) + \cdots + \alpha_0y(k) = \beta_mw(k+m) + \beta_{m-1}w(k+m-1) + \cdots + \beta_0w(k).$$

- Imponemos que esto vale para $m = n - 1, k = 0, 1, 2, \dots$
- Definimos una variable temporal $r(k)$ que cumple

$$r(k+n) + \alpha_{n-1}r(k+n-1) + \cdots + \alpha_0r(k) = w(k). \quad (*)$$

- Definimos los estados como:

$$\begin{aligned}x_1(k) &= r(k) \\x_2(k) &= r(k+1) \\&\vdots \\x_n(k) &= r(k+n-1).\end{aligned}$$

De esta forma, reemplazando en (*) podemos escribir:

$$\begin{bmatrix} x_1(k+1) \\ x_2(k+1) \\ \vdots \\ x_n(k+1) \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 & 0 & \dots \\ 0 & 0 & 1 & 0 & \dots \\ \vdots & & & & \\ -\alpha_0 & -\alpha_1 & -\alpha_2 & \dots & -\alpha_{n-1} \end{bmatrix} \begin{bmatrix} x_1(k) \\ x_2(k) \\ \vdots \\ x_n(k) \end{bmatrix} + \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix} w$$

Luego, para obtener $y(k)$:

$$y(k) = [\beta_0 \ \beta_1 \ \dots \ \beta_m] \begin{bmatrix} x_1(k) \\ x_2(k) \\ \vdots \\ x_n(k) \end{bmatrix}.$$

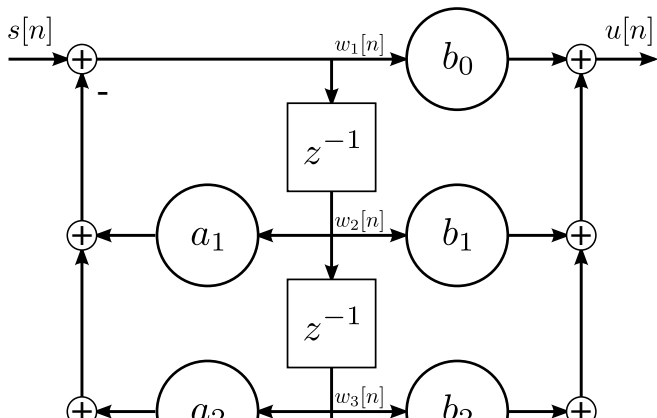
Modelo de estado a partir de un modelo ARMA (cont.)

En forma de diagrama de bloque, calculando primero la transferencia correspondiente a $(*)$,

$$R(z)z^n + \alpha_{n-1}R(z)z^{n-1} + \dots + \alpha_0R(z) = W(z)$$

$$\Rightarrow R(z) = \frac{W(z)}{z^n + \alpha_{n-1}z^{n-1} + \dots + \alpha_0},$$

tenemos:



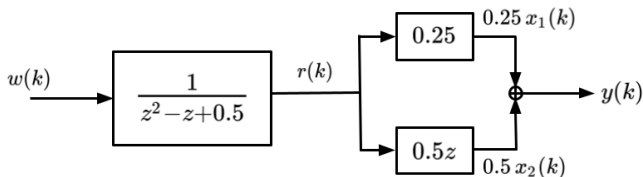
Ejemplo

Tenemos el siguiente proceso ARMA:

$$y(k+2) - y(k+1) + \frac{1}{2}y(k) = \frac{1}{2}w(k+1) + \frac{1}{4}w(k), \quad k = 0, 1, 2, \dots$$

Representar el proceso como diagrama de bloques, y proponer una representación en variables de estado.

Solución:



$$\begin{bmatrix} x_1(k+1) \\ x_2(k+1) \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ -0.5 & 1 \end{bmatrix} \begin{bmatrix} x_1(k) \\ x_2(k) \end{bmatrix} + \begin{bmatrix} 0 \\ 1 \end{bmatrix} w(k).$$

$$y(k) = \begin{bmatrix} 0.25 & 0.5 \end{bmatrix} \begin{bmatrix} x_1(k) \\ x_2(k) \end{bmatrix}.$$

Forma alternativa del filtro de Kalman discreto

En la formulación original teníamos

$$\mathbf{P}_k = (\mathbf{I} - \mathbf{K}_k \mathbf{C}_k) \mathbf{P}_k^-$$
$$\mathbf{K}_k = \mathbf{P}_k^- \mathbf{C}_k^T (\mathbf{C}_k \mathbf{P}_k^- \mathbf{C}_k^T + \mathbf{R}_k)^{-1}$$

Sustituyendo, tenemos

$$\mathbf{P}_k = \mathbf{P}_k^- - \mathbf{P}_k^- \mathbf{C}_k^T (\mathbf{C}_k \mathbf{P}_k^- \mathbf{C}_k^T + \mathbf{R}_k)^{-1} \mathbf{C}_k \mathbf{P}_k^-.$$

- Se puede demostrar que si \mathbf{P}_k , \mathbf{P}_k^- y \mathbf{R}_k son invertibles, entonces

$$\mathbf{P}_k^{-1} = (\mathbf{P}_k^-)^{-1} + \mathbf{C}_k^T \mathbf{R}_k^{-1} \mathbf{C}_k.$$

- $\mathbf{K}_k = \mathbf{P}_k^- \mathbf{C}_k^T (\mathbf{C}_k \mathbf{P}_k^- \mathbf{C}_k^T + \mathbf{R}_k)^{-1}$. Insertando $\mathbf{P}_k \mathbf{P}_k^{-1}$ y $\mathbf{R}_k^{-1} \mathbf{R}_k$ obtenemos

$$\begin{aligned} \mathbf{K}_k &= \mathbf{P}_k \mathbf{P}_k^{-1} \mathbf{P}_k^- \mathbf{C}_k^T \mathbf{R}_k^{-1} \mathbf{R}_k (\mathbf{C}_k \mathbf{P}_k^- \mathbf{C}_k^T + \mathbf{R}_k)^{-1} \\ &= \mathbf{P}_k \mathbf{P}_k^{-1} \mathbf{P}_k^- \mathbf{C}_k^T \mathbf{R}_k^{-1} (\mathbf{C}_k \mathbf{P}_k^- \mathbf{C}_k^T \mathbf{R}_k^{-1} + \mathbf{I})^{-1} \end{aligned}$$

Sustituyendo \mathbf{P}_k^{-1} en la expresión previa,

$$\mathbf{K}_k = \mathbf{P}_k \mathbf{C}_k^T \mathbf{R}_k^{-1}$$

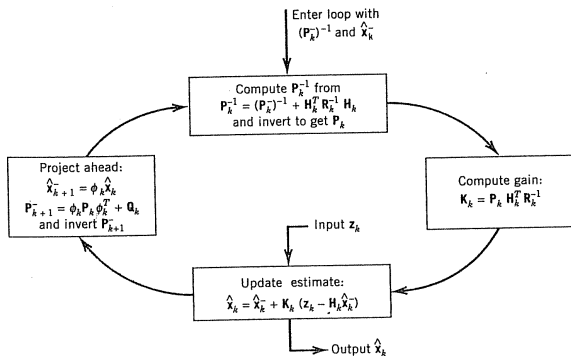
Forma alternativa del filtro de Kalman discreto (cont.)

Tenemos así una formulación alternativa del filtro de Kalman, con

$$\mathbf{P}_k^{-1} = (\mathbf{P}_k^-)^{-1} + \mathbf{C}_k^T \mathbf{R}_k^{-1} \mathbf{C}_k$$

$$\mathbf{K}_k = \mathbf{P}_k \mathbf{C}_k^T \mathbf{R}_k^{-1}$$

Obs. En esta formulación se calcula primero \mathbf{P}_k^{-1} .



En la Figura: $\mathbf{H}_k = \mathbf{C}_k$, $\mathbf{z}_k = \mathbf{y}_k$.

Ejemplo (ejercicio): estimación de una constante x escalar

x constante $\Rightarrow \Phi_k = 1, \mathbf{Q}_k = 0$.

- Tenemos N medidas independientes entre si y de x en $t = 0$, con varianza σ_v^2 :

$$\mathbf{y}_k = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix} x + \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_N \end{bmatrix}, \quad \mathbf{R}_k = \sigma_v^2 \mathbf{I}_N.$$

- Condiciones iniciales: $\hat{x}_0^- = 0, \mathbf{P}_0^- = +\infty$ (esto último no es posible en la formulación anterior porque la ganancia queda ∞/∞).

Ejercicio: plantear el filtro y obtener la primera estimación \hat{x}_0 .

Estimación de una constante: solución

$$\begin{aligned} \textcircled{1} \quad \mathbf{P}_0^{-1} &= (\mathbf{P}_0^-)^{-1} + \mathbf{C}_0^T \mathbf{R}_0^{-1} \mathbf{C}_0 = \\ &= (+\infty)^{-1} + \begin{bmatrix} 1 & \dots & 1 \end{bmatrix} \frac{\mathbf{I}_N}{\sigma_v^2} \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix} = \frac{N}{\sigma_v^2} \end{aligned}$$

$$\begin{aligned} \textcircled{2} \quad \mathbf{K}_0 &= \mathbf{P}_0 \mathbf{C}_0^T \mathbf{R}_0^{-1} \\ &= \frac{\sigma_v^2}{N} \begin{bmatrix} 1 & \dots & 1 \end{bmatrix} \text{diag}(1/\sigma_v^2, \dots, 1/\sigma_v^2) = \frac{1}{N} \begin{bmatrix} 1 & \dots & 1 \end{bmatrix} \end{aligned}$$

$$\textcircled{3} \quad \hat{x}_0 = \hat{x}_0^- + \mathbf{K}_0(\mathbf{y}_0 - \mathbf{C}_0 \hat{x}_0^-) = \mathbf{K}_0 \mathbf{y}_0 = \frac{1}{N} \sum_{n=1}^N \mathbf{y}_0(n).$$

El objetivo es proyectar N pasos hacia adelante (hasta ahora fue $N = 1$).

$$\hat{\mathbf{x}}(k + N|k) = \Phi(k + N, k)\hat{\mathbf{x}}(k|k)$$

$$\mathbf{P}(k + N|k) = \Phi(k + N, k)\mathbf{P}(k|k)\Phi^T(k + N, k) + \mathbf{Q}(k + N, k).$$

Dos tipos de predicción: 1. N fijo 2. k fijo, $N = 1, 2, 3, \dots$

