



UNIVERSIDAD DE CHILE
FACULTAD DE CIENCIAS FÍSICAS Y MATEMÁTICAS
DEPARTAMENTO DE INGENIERÍA MATEMÁTICA

TÍTULO DE LA TESIS

TESIS PARA OPTAR AL GRADO DE MAGÍSTER EN CIENCIAS DE LA INGENIERÍA,
MENCIÓN MATEMÁTICAS APLICADAS

MEMORIA PARA OPTAR AL TÍTULO DE INGENIERO CIVIL MATEMÁTICO

DIEGO ANDRÉS OLGUÍN WENDE

PROFESOR GUÍA:
HÉCTOR ARIEL RAMÍREZ CABRERA
PROFESOR CO-GUÍA:
AXEL ESTEBAN OSSES ALVARADO

MIEMBROS DE LA COMISIÓN:
PROFESOR 2
PROFESOR 3

Este trabajo ha sido parcialmente financiado por:
NOMBRE INSTITUCIÓN

SANTIAGO DE CHILE
2025

RESUMEN DE LA MEMORIA PARA OPTAR
AL TÍTULO DE MAGÍSTER EN CIENCIAS
DE LA INGENIERÍA CON MENCIÓN MATEMÁTICAS APLICADAS
POR: DIEGO ANDRÉS OLGUÍN WENDE
FECHA: 2025
PROF. GUÍA: HÉCTOR ARIEL RAMÍREZ CABRERA
PROF. CO-GUÍA: AXEL ESTEBAN OSSES ALVARADO

TÍTULO DE LA TESIS

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetur id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

Dedicatoria
Saludos

Agradecimientos

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetur id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

Tabla de Contenido

1. Introducción	1
1.1. Problema y algoritmos de filtraje	1
1.1.1. Caso lineal con ruidos aditivos, centrados y Gaussianos	2
1.1.2. Caso no lineal	3
1.2. Reproducing Kernel Hilbert Spaces	6
1.3. Operador de Koopman	11
1.4. Extended Dynamic Mode Decomposition (EDMD)	12
1.5. Aproximación de rango bajo de matrices	14
2. Kernel Extended Dynamic Mode Decomposition	16
2.1. Extended Dynamic Mode Decomposition	16
2.1.1. Resultados numéricos	23
3. Algoritmo de filtraje en tiempo discreto	24
3.1. Filtro de Kalman a tiempo discreto en RKHS	24
4. Algoritmo de filtraje en tiempo continuo	33
5. Algoritmo de suavizado en tiempo continuo y tiempo discreto	34
6. Algoritmo de estimación de parámetros	35
6.1. Problema de optimización estocástica bi-nivel	35
6.2. Observabilidad	35
6.3. Algoritmo propuesto	35
6.4. Resultado de convergencia	35
7. Modelos epidemiológicos	36
Bibliografía	37

Índice de Ilustraciones

1.1.	Diagrama de la evolución temporal tanto en dimensión finita como en dimensión infinita para la dinámica representada por el operador de Koopman, con las respectivas representaciones en dimensión finita de cada uno de los objetos involucrados. Elaboración propia, basado en [11].	14
------	--	----

Capítulo 1

Introducción

1.1. Problema y algoritmos de filtraje

Se considera un sistema controlado-observado estocástico a tiempo discreto

$$\begin{aligned}\mathbf{x}_{k+1} &= \mathbf{f}(t_k, \mathbf{x}_k, \mathbf{u}_k, \mathbf{w}_k) \\ \mathbf{y}_k &= \mathbf{g}(t_k, \mathbf{x}_k, \mathbf{u}_k, \mathbf{v}_k)\end{aligned}$$

Donde:

- $\mathbf{f} : \mathbb{R} \times \mathbb{R}^{n_x} \times \mathbb{R}^{n_u} \times \mathbb{R}^{n_w} \rightarrow \mathbb{R}^{n_x}$ es la función de dinámica.
- $\mathbf{g} : \mathbb{R} \times \mathbb{R}^{n_x} \times \mathbb{R}^{n_u} \times \mathbb{R}^{n_v} \rightarrow \mathbb{R}^{n_y}$ es la función de observación

Siendo n_x la dimensión del estado, n_y la dimensión de la observación, n_u la dimensión del *input*, n_w la dimensión del proceso estocástico en la dinámica y n_v la dimensión del ruido en las observaciones. Se considera una condición inicial \mathbf{x}_0 , por lo general desconocida sobre la que se coloca un *prior* X_0 .

Se supondrá que tanto \mathbf{v}_k como \mathbf{w}_k son vectores aleatorios centrados tal que $\mathbf{w}_k \perp \mathbf{w}_m, \forall k \neq m$, $\mathbf{v}_k \perp \mathbf{v}_m, \forall k \neq m$ y $\mathbf{w}_k \perp \mathbf{v}_k, \forall k$. Diremos que \mathbf{w} y \mathbf{v} son la perturbación en la dinámica y ruido en la medición, respectivamente.

El estudio de sistemas a tiempo continuo, formulados como

$$\begin{aligned}\mathbf{x}'(t) &= \mathbf{f}(t, \mathbf{x}(t), \mathbf{u}(t), \mathbf{w}(t)) \\ \mathbf{y}(t) &= \mathbf{g}(t, \mathbf{x}(t), \mathbf{u}(t), \mathbf{v}(t))\end{aligned}$$

muchas veces se reduce al caso discreto con esquemas de tipo Euler, por lo que durante esta tesis se mantendrá el foco en sistemas a tiempo discreto, a menos de que se necesite lo contrario.

El objetivo en el caso discreto será, en un instante k encontrar un estimador de $\mathbf{x}_{k+\alpha}$, en base a observaciones $\{\tilde{\mathbf{y}}_k\}_k$, que será denotado por $\hat{\mathbf{x}}_{k+\alpha}$. Dependiendo del valor de α el algoritmo recibirá una categoría diferente:

- Predicción: $\alpha > 0$.
- Filtraje: $\alpha = 0$.
- Suavizado: $\alpha < 0$.

Así también se denotará a $\hat{\mathbf{x}}_{k_1|k_2}$ a la estimación del estado en la iteración k_1 dadas mediciones hasta la iteración k_2 , entonces se dirá que $\hat{\mathbf{x}}_{k_1|k_2}$ es un estado

- Suavizado, si $k_2 > k_1$ ($t_2 > t_1$).
- Filtrado, si $k_2 = k_1$ ($t_2 = t_1$).
- Predicho, si $k_1 > k_2$ ($t_1 > t_2$).

El caso del filtraje será el centro durante este trabajo. Un algoritmo de filtraje busca estimar en el tiempo presente el estado en base a observaciones ruidosas, que puede ser entendido como el primer proceso en la estimación del estado de un sistema.

Posterior al filtraje de una medición ruidosa, se siguen los procesos de predicción y/o suavizado, siendo el primero la tarea poder estimar estados futuros en base al pasado, sin información del momento a estimar. Mientras que el suavizado es la capacidad de estimar momentos pasados, lo que se utiliza para mejorar las estimaciones presentes.

Se denota por $\mathbf{y}_{1:k}$, $\mathbf{u}_{0:k}$ a las observaciones e *inputs*, respectivamente, hasta el tiempo $k \in \mathbb{N}$, a $k \in \mathbb{N}$ le llamaremos el horizonte del problema. Notar que se considerará que no se tiene una observación de la condición inicial.

El problema de filtraje se puede formular como un problema de optimización del error cuadrático medio de la estimación, condicional a las observaciones y a los *inputs*

$$(P) \quad \begin{cases} \text{mín} & \sum_{k=0}^N \mathbb{E} [(\hat{\mathbf{x}}_{k|k} - \mathbf{x}_k)^T (\hat{\mathbf{x}}_{k|k} - \mathbf{x}_k) | \mathbf{y}_{0:k}, \mathbf{u}_{0:k}] \\ \text{s.a} & \mathbf{x}_{k+1} = \mathbf{f}(t_k, \mathbf{x}_k, \mathbf{u}_k, \mathbf{w}_k) \\ & \mathbf{y}_k = \mathbf{g}(t_k, \mathbf{x}_k, \mathbf{u}_k, \mathbf{v}_k) \\ & \mathbf{x}_0 \sim X_0 \end{cases}$$

Es entonces que el problema (P) tiene como solución al estimador de mínimo error cuadrático. Es sabido de [1, 2] que este problema tiene un óptimo global dado por la esperanza condicional, que coincide con la noción de dicho estimador.

Proposición 1.1 *El óptimo de (P) viene dado por*

$$\hat{\mathbf{x}}_{k|k} = \mathbb{E} [\mathbf{x}_k | \mathbf{y}_{1:k}, \mathbf{u}_{0:k}], \quad \forall k \in \{0, \dots, N\}$$

Aunque dicha cantidad no sea computable por métodos clásicos, como Monte Carlo, existen algoritmos para poder calcularla tiempo a tiempo, aunque sea de manera aproximada.

1.1.1. Caso lineal con ruidos aditivos, centrados y Gaussianos

Para abordar el caso de sistemas dinámicos controlados-observados generales, es importante estudiar primero el caso lineal y con ruidos aditivos, centrados y Gaussianos, esto es

$$\begin{cases} \mathbf{x}_{k+1} &= \mathbf{A}_k \mathbf{x}_k + \mathbf{B}_k \mathbf{u}_k + \mathbf{v}_k \\ \mathbf{y}_k &= \mathbf{C}_k \mathbf{x}_k + \mathbf{D}_k \mathbf{u}_k + \mathbf{w}_k \end{cases} \quad (1.1)$$

con $\mathbf{A}_k \in \mathbb{R}^{n_x \times n_x}$, $\mathbf{B}_k \in \mathbb{R}^{n_x \times n_u}$, $\mathbf{C}_k \in \mathbb{R}^{n_y \times n_x}$, $\mathbf{D}_k \in \mathbb{R}^{n_y \times n_u}$, $\mathbf{v}_k \sim \mathcal{N}(0, \mathbf{Q}_k)$ y $\mathbf{w}_k \sim \mathcal{N}(0, \mathbf{R}_k)$, $\mathbf{Q}_k \in \mathbb{R}^{n_x \times n_x}$, $\mathbf{R}_k \in \mathbb{R}^{n_y \times n_y}$. matrices de covarianzas, se tiene un algoritmo muy

utilizado en ingeniería y ciencia, con suficiente fundamentos matemáticos e implementaciones eficientes, denotado el filtro de Kalman en honor a Rudolf E. Kalman, quién propuso el algoritmo por primera vez en [1]. Los detalles del algoritmo se presentan en el pseudo-código del Algoritmo 1.

Algorithm 1 Filtro de Kalman

```

1: Entrada: Dinámica discreta como en (1.1),  $\mathbf{x}_0$  prior sobre la condición inicial,  $\mathbf{y}_{1:N}$ 
   observaciones,  $\mathbf{u}_{0:N}$  inputs.
2: Salida:  $(\hat{\mathbf{x}}_{k|k})_{k=0}^N$  estimador de la trayectoria y  $(\hat{\mathbf{P}}_{k|k})_{k=0}^N$  matrices de covarianza.
3:  $\hat{\mathbf{x}}_{0|0} \leftarrow \mathbb{E}[\mathbf{x}_0]$ 
4:  $\mathbf{P}_{0|0} \leftarrow \mathbb{E}[(\mathbf{x}_0 - \hat{\mathbf{x}}_0)(\mathbf{x}_0 - \hat{\mathbf{x}}_0)^T]$ 
5: for  $k = 0, \dots, N - 1$  do
6:    $\hat{\mathbf{x}}_{k+1|k} \leftarrow \mathbf{A}_k \mathbf{x}_{k|k} + \mathbf{B}_k \mathbf{u}_k$  ▷ Estimación a priori
7:    $\mathbf{P}_{k+1|k} \leftarrow \mathbf{A}_k \mathbf{P}_{k|k} \mathbf{A}_k^T + \mathbf{Q}_k$  ▷ Error de covarianza a priori
8:    $\hat{\mathbf{y}}_{k+1|k} \leftarrow \mathbf{C}_{k+1} \hat{\mathbf{x}}_{k+1|k}$  ▷ Estimación de observación a priori
9:    $\mathbf{e}_{\mathbf{y}_{k+1|k}} \leftarrow \mathbf{y}_{k+1} - \hat{\mathbf{y}}_{k+1|k}$  ▷ Error a priori (innovación)
10:   $\mathbf{K}_{k+1} \leftarrow \mathbf{P}_{k+1|k} \mathbf{C}_{k+1}^T (\mathbf{C}_{k+1} \mathbf{P}_{k|k} \mathbf{C}_{k+1}^T + \mathbf{R}_{k+1})^{-1}$  ▷ Ganancia de Kalman
11:   $\hat{\mathbf{x}}_{k+1|k+1} \leftarrow \hat{\mathbf{x}}_{k+1|k} + \mathbf{K}_{k+1} \mathbf{e}_{\mathbf{y}_{k+1|k}}$  ▷ Error a posteriori
12:   $\mathbf{P}_{k+1|k+1} \leftarrow (\mathbf{I} - \mathbf{K}_{k+1} \mathbf{C}_{k+1}) \mathbf{P}_{k+1|k}$  ▷ Error de covarianza a posteriori
13: end for

```

El siguiente resultado le da toda su validez al algoritmo del Filtro de Kalman, que es probado en el artículo original [1].

Proposición 1.2 *El algoritmo del filtro de Kalman devuelve una secuencia $(\hat{\mathbf{x}}_{k|k})_k$ tal que*

$$\hat{\mathbf{x}}_{k|k} = \mathbb{E}[\mathbf{x}_k | \mathbf{y}_{1:k}, \mathbf{u}_{0:k}], \quad \forall k \in \{0, \dots, N\}$$

Como consecuencia del hecho de que en el contexto Gaussiano el estimador de mínimo error cuadrático medio coincide con el estimador de máximo a posteriori, se tiene el siguiente resultado.

Proposición 1.3 *El algoritmo del filtro de Kalman devuelve una secuencia $(\hat{\mathbf{x}}_{k|k})_k$ tal que*

$$\hat{\mathbf{x}}_{k|k} \in \arg \max_{\hat{\mathbf{x}}_{k|k}} p(\hat{\mathbf{x}}_{k|k} | \mathbf{y}_{1:k}, \mathbf{u}_{0:k})$$

1.1.2. Caso no lineal

A diferencia del caso anterior, en el caso no lineal general hay muchos algoritmos propuestos, siendo los más clásicos las variantes del filtro de Kalman, como el *Extended Kalman Filter* (EKF) y el *Unscented Kalman Filter* (UKF), pero se sabe que resultan ser subóptimos [2].

A pesar de que dichos algoritmos reciben dinámicas no lineales, siguen suponiendo que el ruido es aditivo, centrado y Gaussiano. Se verá en secciones posteriores que esto no es un impedimento para poder ejecutarlos, pero una familia de métodos se han construido para el caso más general, cuyos representantes más famosos son los filtros de partículas.

Con el motivo de no extender más este informe solo se expondrá el algoritmo de EKF. Para

esta sección se supondrá que la dinámica es de la forma

$$\begin{aligned}\mathbf{x}_{k+1} &= \mathbf{f}(t_k, \mathbf{x}_k, \mathbf{u}_k) + \mathbf{w}_k \\ \mathbf{y}_k &= \mathbf{g}(t_k, \mathbf{x}_k, \mathbf{u}_k) + \mathbf{v}_k\end{aligned}\tag{1.2}$$

Con $\mathbf{w}_k \sim \mathcal{N}(0, \mathbf{Q}_k)$ y $\mathbf{v}_k \sim \mathcal{N}(0, \mathbf{R}_k)$. El algoritmo de *Extended Kalman Filter*, cuyo pseudo-código puede visualizarse en el algoritmo 2, busca linearizar el sistema a primer orden vía su Jacobiano, tanto para \mathbf{f} como para \mathbf{g} de manera de generar las matrices \mathbf{A}_k y \mathbf{C}_k necesarias para el Filtro de Kalman, respectivamente.

A pesar de lo simple que parece la adaptación de este algoritmo al caso no lineal, ilustra que extender el filtro de Kalman al caso no lineal se basa en una linealización de la dinámica, en este caso vía Jacobiano, pero podrían existir otras, que es en lo que se basará el filtro creado durante este trabajo.

Algorithm 2 *Extended Kalman Filter*

- 1: **Entrada:** Dinámica discreta como en (1.2), \mathbf{x}_0 *prior* sobre la condición inicial, $\mathbf{y}_{1:N}$ observaciones, $\mathbf{u}_{0:N}$ *inputs*.
 - 2: **Salida:** $(\hat{\mathbf{x}}_{k|k})_{k=0}^N$ estimador de la trayectoria y $(\hat{\mathbf{P}}_{k|k})_{k=0}^N$ matrices de covarianza.
 - 3: $\hat{\mathbf{x}}_{0|0} \leftarrow \mathbb{E}[\mathbf{x}_0]$
 - 4: $\mathbf{P}_0 \leftarrow \mathbb{E}[(\mathbf{x}_0 - \hat{\mathbf{x}}_0)(\mathbf{x}_0 - \hat{\mathbf{x}}_0)^T]$
 - 5: **for** $k = 0, \dots, N - 1$ **do**
 - 6: $\mathbf{A}_k \leftarrow \nabla_{\mathbf{x}} \mathbf{f}(t_k, \hat{\mathbf{x}}_{k|k}, \mathbf{u}_k)$ ▷ Linealización de la función de dinámica
 - 7: $\hat{\mathbf{x}}_{k+1|k} \leftarrow \mathbf{f}(t_k, \hat{\mathbf{x}}_{k|k}, \mathbf{u}_k)$ ▷ Estimación a priori
 - 8: $\mathbf{P}_{k+1|k} \leftarrow \mathbf{A}_k \mathbf{P}_{k|k} \mathbf{A}_k^T + \mathbf{Q}_k$ ▷ Error de covarianza a priori
 - 9: $\mathbf{C}_{k+1} \leftarrow \nabla_{\mathbf{x}} \mathbf{g}(t_k, \hat{\mathbf{x}}_{k+1|k}, \mathbf{u}_k)$ ▷ Linealización de la función de observación
 - 10: $\hat{\mathbf{y}}_{k+1|k} \leftarrow \mathbf{g}(t_k, \hat{\mathbf{x}}_{k+1|k}, \mathbf{x}_{k+1})$ ▷ Estimación de observación a priori
 - 11: $\mathbf{e}_{\mathbf{y}_{k+1|k}} \leftarrow \mathbf{y}_{k+1} - \hat{\mathbf{y}}_{k+1|k}$ ▷ Error a priori (innovación)
 - 12: $\mathbf{K}_{k+1} \leftarrow \mathbf{P}_{k+1|k} \mathbf{C}_{k+1}^T (\mathbf{C}_{k+1} \mathbf{P}_{k|k} \mathbf{C}_{k+1}^T + \mathbf{R}_{k+1})^{-1}$ ▷ Ganancia de Kalman
 - 13: $\hat{\mathbf{x}}_{k+1|k+1} \leftarrow \hat{\mathbf{x}}_{k+1|k} + \mathbf{K}_{k+1} \mathbf{e}_{\mathbf{y}_{k+1|k}}$ ▷ Error a posteriori
 - 14: $\mathbf{P}_{k+1|k+1} \leftarrow (\mathbf{I} - \mathbf{K}_{k+1} \mathbf{C}_{k+1}) \mathbf{P}_{k+1|k}$ ▷ Error de covarianza a posteriori
 - 15: **end for**
-

El algoritmo del *Unscented Kalman Filter* (UKF), pseudo-código del algoritmo 3, se basa en generar puntos de muestra (llamados puntos sigma) alrededor de la estimación actual del estado del sistema. Estos puntos permiten representar la distribución estadística de la estimación sin necesidad de calcular derivadas. Al propagar estos puntos sigma a través del modelo no lineal, el UKF logra realizar estimaciones en el siguiente instante de tiempo, capturando de manera precisa las propiedades no lineales del sistema. Esta estrategia hace que el UKF sea una alternativa robusta al *Extended Kalman Filter* (EKF) para el seguimiento y la estimación en sistemas no lineales. Aún así, el algoritmo puede ser costoso en la práctica y es necesaria una gran cantidad de puntos sigma para lograr una estimación fiable.

Los Filtros de Partículas (PF) o algoritmos de *Sequential Monte Carlo* (SMC) [3, 4] son una familia de métodos que estiman el estado de sistemas dinámicos no lineales y/o no gaussianos mediante técnicas de muestreo tipo Monte Carlo. Estos algoritmos representan la distribución de probabilidad del estado mediante un conjunto de partículas, que son muestras aleatorias ponderadas según su probabilidad. La precisión de estos filtros depende críticamente del proceso de resampling, un paso fundamental que evita la degeneración de partículas al

Algorithm 3 *Unscented Kalman Filter*

- 1: **Entrada:** Dinámica discreta como en (1.2), \mathbf{x}_0 *prior* sobre la condición inicial, $\mathbf{y}_{1:N}$ observaciones, $\mathbf{u}_{0:N}$ *inputs*.
- 2: **Salida:** $(\hat{\mathbf{x}}_{k|k})_{k=0}^N$ estimador de la trayectoria y $(\hat{\mathbf{P}}_{k|k})_{k=0}^N$ matrices de covarianza.
- 3: **Inicialización:**
- 4: $\hat{\mathbf{x}}_{0|0} \leftarrow \mathbb{E}[\mathbf{x}_0]$
- 5: $\mathbf{P}_0 \leftarrow \mathbb{E}[(\mathbf{x}_0 - \hat{\mathbf{x}}_0)(\mathbf{x}_0 - \hat{\mathbf{x}}_0)^T]$
- 6: **for** $k = 1, 2, \dots, N$ **do**
- 7: Calcular los puntos sigma χ usando la estimación del estado actual $\hat{\mathbf{x}}_{k-1}$ y la covarianza \mathbf{P}_{k-1} . Asignar pesos a los puntos sigma de acuerdo con pesos predefinidos W_m (para la media) y W_c (para la covarianza).
- 8: $\chi_{k|k-1}^{(i)} \leftarrow \mathbf{f}(t_k, \chi_{k-1}^{(i)}, \mathbf{u}_k)$ \triangleright Propagar cada punto sigma a través de la dinámica.
- 9: $\hat{\mathbf{x}}_{k|k-1} \leftarrow \sum_i W_m^{(i)} \chi_{k|k-1}^{(i)}$ \triangleright Calcular la media del estado predicho.
- 10: Calcular la covarianza predicha:

$$\mathbf{P}_{k|k-1} = \sum_i W_c^{(i)} \left(\chi_{k|k-1}^{(i)} - \hat{\mathbf{x}}_{k|k-1} \right) \left(\chi_{k|k-1}^{(i)} - \hat{\mathbf{x}}_{k|k-1} \right)^\top + \mathbf{Q}_k$$

- 11: $\gamma_k^{(i)} = \mathbf{h}(t_k, \chi_{k|k-1}^{(i)}, \mathbf{u}_k)$ \triangleright Pasar cada punto sigma predicho a través de la observación.
- 12: $\hat{\mathbf{y}}_k = \sum_i W_m^{(i)} \gamma_k^{(i)}$ \triangleright Calcular la media de la medición predicha.
- 13: Calcular la covarianza de la medición:

$$\mathbf{S}_k = \sum_i W_c^{(i)} \left(\gamma_k^{(i)} - \hat{\mathbf{y}}_k \right) \left(\gamma_k^{(i)} - \hat{\mathbf{y}}_k \right)^\top + \mathbf{R}_k$$

- 14: Calcular la covarianza cruzada:

$$\mathbf{C}_k = \sum_i W_c^{(i)} \left(\chi_{k|k-1}^{(i)} - \hat{\mathbf{x}}_{k|k-1} \right) \left(\gamma_k^{(i)} - \hat{\mathbf{y}}_k \right)^\top$$

- 15: $\mathbf{K}_k = \mathbf{C}_k \mathbf{S}_k^{-1}$ \triangleright Calcular la ganancia de Kalman.
 - 16: $\hat{\mathbf{x}}_k = \hat{\mathbf{x}}_{k|k-1} + \mathbf{K}_k (\mathbf{y}_k - \hat{\mathbf{y}}_k)$ \triangleright Actualizar la estimación del estado.
 - 17: $\mathbf{P}_k = \mathbf{P}_{k|k-1} - \mathbf{K}_k \mathbf{S}_k \mathbf{K}_k^\top$ \triangleright Actualizar la covarianza.
 - 18: **end for**
-

eliminar aquellas con pesos bajos y duplicar las más probables. Esto permite que los filtros de partículas mantengan una representación precisa de la distribución posterior en cada paso de tiempo, adaptándose dinámicamente a la complejidad del sistema. Para el algoritmo se considerará que los modelos vienen representados por distribuciones de probabilidad, una de transición para la dinámica $p(\mathbf{x}_k|\mathbf{x}_{k-1})$ y para la observación $p(\mathbf{y}_k|\mathbf{x}_k)$.

Algorithm 4 Filtro de Partículas

```

1: Entrada: Modelo de dinámica  $p(\mathbf{x}_k|\mathbf{x}_{k-1})$ , modelo de medición  $p(\mathbf{y}_k|\mathbf{x}_k)$ , observaciones  $\mathbf{y}_{1:k}$ , número de partículas  $N_p$ , estado inicial  $\{\mathbf{x}_0^{(i)}\}_{i=1}^{N_p}$ 
2: Salida: Estimación del estado basada en la distribución ponderada de partículas  $\{\mathbf{x}_k^{(i)}, w_k^{(i)}\}_{i=1}^{N_p}$ 
3: for  $i = 1, \dots, N_p$  do
4:   Generar la partícula inicial  $\mathbf{x}_0^{(i)}$  de la distribución inicial  $p(\mathbf{x}_0)$ 
5:   Asignar el peso inicial  $w_0^{(i)} = \frac{1}{N_p}$ 
6: end for
7: for  $k = 1, 2, \dots, N$  do
8:   for  $i = 1, \dots, N_p$  do
9:     Muestrear  $\mathbf{x}_k^{(i)} \sim p(\mathbf{x}_k|\mathbf{x}_{k-1}^{(i)})$  ▷ Propagar cada partícula por la dinámica.
10:   end for
11:   for  $i = 1, \dots, N_p$  do
12:      $w_k^{(i)} = w_{k-1}^{(i)} \cdot p(\mathbf{y}_k|\mathbf{x}_k^{(i)})$  ▷ Actualizar peso basado en la observación.
13:   end for
14:    $w_k^{(i)} = \frac{w_k^{(i)}}{\sum_{j=1}^{N_p} w_k^{(j)}}$  ▷ Normalizar los pesos.
15:   if la degeneración de partículas es alta then
16:     Re-sampear las partículas  $\{\mathbf{x}_k^{(i)}, w_k^{(i)}\}_{i=1}^{N_p}$  según sus pesos  $w_k^{(i)}$ 
17:     Reiniciar pesos:  $w_k^{(i)} = \frac{1}{N_p}$  para todas las partículas
18:   end if
19: end for

```

Un aspecto importante de los Filtros de Partículas es que su orden de convergencia ha sido demostrado como $O(N_p^{-1/2})$ [5, 6], es decir, la precisión del método aumenta proporcionalmente a la raíz cuadrada inversa del número de partículas utilizadas. Este número de partículas, que se elige como parámetro del filtro, controla directamente el error de estimación. Esta cota de error se tomará como referencia de comparación en el presente trabajo.

1.2. Reproducing Kernel Hilbert Spaces

Esta sección se basa en lo expuesto en [7] y [8]. Sea E un espacio topológico, se denota por \mathbb{C}^E al espacio de todas las funciones de E en los complejos.

Definición 1.1 (Reproducing kernel Hilbert Space (RKHS)) *Un espacio de Hilbert $H \subset \mathbb{C}^E$ es un RKHS si existe una función $K : E \times E \rightarrow \mathbb{C}$, llamada reproducing kernel, tal que*

1. $K_p \equiv K(\cdot, p) \in \mathcal{H}, \forall p \in E$.
2. $f(p) = \langle f, K_p \rangle_{\mathcal{H}}, \forall p \in E, \forall f \in \mathcal{H}$.

La segunda propiedad es fundamental en la teoría de los RKHS y se denomina la propiedad reproductiva. Se denota por $\mathcal{H}_K(E)$ al espacio de Hilbert de funciones de E en \mathbb{C} cuyo *reproducing kernel* sea K . El siguiente lema da una condición necesaria y suficiente para que un espacio de Hilbert dado posea un *reproducing kernel*.

Lema 1.1 *Un espacio de Hilbert $\mathcal{H} \subset \mathbb{C}^E$ posee un reproducing kernel si y solo si los funcionales de evaluación $e_p : E \rightarrow \mathbb{C}$, $e_p(f) = f(p)$, son continuos en \mathcal{H} .*

Para poder construir un RKHS sobre un espacio de Hilbert, la siguiente definición es necesaria.

Definición 1.2 (Función semi-definida positiva) *Una función $K : E \times E \rightarrow \mathbb{C}$ se dice semi-definida positiva si $\forall n \geq 1$, $\forall (c_1, \dots, c_n) \in \mathbb{C}^n$, $\forall (x_1, \dots, x_n) \in E^n$*

$$\sum_{i=1}^n \sum_{j=1}^n c_i \bar{c}_j K(x_i, x_j) \geq 0$$

Esto es equivalente a que para cada $n \in \mathbb{N}$ y cada $(x_1, \dots, x_n) \in E^n$ la siguiente matriz sea Hermitiana semi-definida positiva

$$G = (K(x_i, x_j))_{1 \leq i, j \leq n}$$

Que se denotará matriz Grammiana del *kernel* asociada a $\{x_i\}_{i=1}^n$, o simplemente matriz Grammiana cuando no haya confusión. Dado un espacio de Hilbert, el siguiente lema entrega una construcción de una función semi-definida positiva.

Lema 1.2 *Sea $\mathcal{H} \subset \mathbb{C}^E$ un Hilbert con producto interno $\langle \cdot, \cdot \rangle_{\mathcal{H}}$ y sea $\varphi : E \rightarrow \mathcal{H}$. Entonces la función $K : E \times E \rightarrow \mathbb{C}$ definida como*

$$K(x, y) = \langle \varphi(x), \varphi(y) \rangle_{\mathcal{H}}$$

Es semi-definida positiva.

Ahora, dado un cierto *kernel* K semi-definido positivo, el siguiente lema y teorema ilustran el comportamiento del espacio de Hilbert $H_K(E)$.

Lema 1.3 *Sea \mathcal{H}_0 cualquier subespacio de \mathbb{C}^E , en el cual está definido un producto interno $\langle \cdot, \cdot \rangle_{\mathcal{H}_0}$, con la norma asociada $\|\cdot\|_{\mathcal{H}_0}$. Luego para que exista un espacio de Hilbert \mathcal{H} tal que*

- (1) $\mathcal{H}_0 \subset \mathcal{H} \subset \mathbb{C}^E$ y la topología definida en \mathcal{H}_0 por el producto interno $\langle \cdot, \cdot \rangle_{\mathcal{H}_0}$ coincida con la topología inducida en \mathcal{H}_0 por \mathcal{H} .
- (2) \mathcal{H} tenga un reproducing kernel K .

Es necesario y suficiente que

- (i) Los funcionales de evaluación $(e_t)_{t \in E}$ sean continuos en \mathcal{H}_0 .
- (ii) Cualquier sucesión de Cauchy (f_n) en \mathcal{H}_0 que converge puntualmente a 0, también converge a 0 en el sentido de la norma.

Teorema 1.1 (Moore–Aronszajn) *Sea K una función semi-definida positiva en $E \times E$. Luego existe un único espacio de Hilbert $\mathcal{H} \subset \mathbb{C}^E$ con K como reproducing kernel. El subespacio \mathcal{H}_0*

de \mathcal{H} generado por las funciones $(K(\cdot, x))_{x \in E}$ es denso en \mathcal{H} y \mathcal{H} es el conjunto de funciones en E que son límites puntuales de sucesiones de Cauchy en \mathcal{H}_0 con el producto interno

$$\langle f, g \rangle_{\mathcal{H}_0} = \sum_{i=1}^n \sum_{j=1}^m \alpha_i \hat{\beta}_j K(y_j, x_i)$$

con

$$f(\cdot) = \sum_{i=1}^n \alpha_i K(\cdot, x_i), \quad g(\cdot) = \sum_{j=1}^m \beta_j K(\cdot, y_j)$$

$$(x_1, \dots, x_n) \in E^n, (y_1, \dots, y_m) \in E^m.$$

El siguiente teorema permite pasar las propiedades de un RKHS a un espacio ℓ^2 , que por lo general es más fácil de trabajar.

Teorema 1.2 *Una función $K : E \times E \rightarrow \mathbb{C}$ es un reproducing kernel si y solo si existe una función $\varphi : E \rightarrow \ell^2(X)$, para algún espacio $\ell^2(X)$ tal que $\forall (x, y) \in E \times E$*

$$K(x, y) = \langle \varphi(x), \varphi(y) \rangle_{\ell^2(X)}$$

Es más, $\ell^2(X)$ es un espacio isométrico a $H_K(E)$ a través de una isometría $T : E \rightarrow \ell^2(X)$ y φ viene dada por $\varphi(x) = T(K(\cdot, x))$. A dicha función φ se le denotará el feature map asociado a K .

Definición 1.3 *Se define el kernel de Matérn de parámetro de suavización $\nu > 0$ y ancho de banda $\gamma > 0$ como*

$$k_\nu(x, y) = \frac{2^{1-\nu}}{\Gamma(\nu)} \left(\sqrt{2\nu} \frac{\|x - y\|}{\gamma} \right)^\nu B_\nu \left(\sqrt{2\nu} \frac{\|x - y\|}{\gamma} \right)$$

donde Γ es la función Gamma y B_ν es la función modificada de Bessel de segundo tipo de parámetro ν .

Definición 1.4 *Un conjunto $\Omega \subset \mathbb{R}^n$ satisface la condición del cono interior si existe un ángulo $\theta \in (0, \pi/2)$ y un radio $r > 0$ tales que lo siguiente se cumple: para cada $x \in \Omega$ existe $\xi(x) \in \mathbb{R}^n$, $\|\xi(x)\|_2 = 1$, tal que el cono*

$$C(x, \xi(x), \theta, r) := \{x + \lambda y : y \in \mathbb{R}^n, \|y\| = 1, y^\top \xi(x) \geq \cos \theta, \lambda \in [0, r]\}$$

está contenido en Ω .

Teorema 1.3 (Tuo & Wu) *Si \mathcal{X} es un compacto que satisface la condición de cono interior, k es el kernel Matern de parámetro $\nu > 0$, entonces para $s = \nu + n/2$ se tiene que $\mathcal{H}_k(\mathcal{X})$ es equivalente en norma a $H^s(\mathcal{X})$.*

Si $\nu = 1/2$, se tiene que

$$B_{1/2}(z) = \sqrt{\frac{\pi}{2}} \frac{e^{-z}}{\sqrt{z}}$$

entonces

$$\begin{aligned} k(x, y) &= \frac{\sqrt{2}}{\Gamma(1/2)} \left(\frac{\|x - y\|}{\gamma} \right)^{1/2} \sqrt{\frac{\pi}{2}} \exp \left(-\frac{\|x - y\|}{\gamma} \right) \left(\frac{\|x - y\|}{\gamma} \right)^{-1/2} \\ &= \exp \left(-\frac{\|x - y\|}{\gamma} \right) \end{aligned}$$

es decir, el kernel de Matern generaliza al kernel Laplace y el RKHS del kernel Laplace es $H^{(n+1)/2}$.

Primero definiremos el elemento medio en un RKHS.

Definición 1.5 (Elemento medio) *Se define el elemento medio en el RKHS $\mathcal{H}_{\mathcal{X}}$ como*

$$\hat{\mu} = \int_{\mathcal{X}} \Phi_{\mathcal{X}}(x) d\mu_{\mathcal{X}}(x) \in \mathcal{H}_{\mathcal{X}}$$

Se define primero el siguiente operador de rango 1 para $x, y \in \mathcal{X}$ fijos

$$\Phi_{\mathcal{X}}(x) \otimes \Phi_{\mathcal{X}}(y) : \mathcal{H}_{\mathcal{X}} \rightarrow \mathcal{H}_{\mathcal{X}}$$

$$[\Phi_{\mathcal{X}}(x) \otimes \Phi_{\mathcal{X}}(y)]\psi = \langle \psi, \Phi_{\mathcal{X}}(y) \rangle \Phi_{\mathcal{X}}(x) = \psi(y) \Phi_{\mathcal{X}}(x)$$

Donde en la última igualdad se utilizó la propiedad reproductiva.

Definición 1.6 *Se define el operador de covarianza asociado al feature map Φ como $C_X : \mathcal{H}_{\mathcal{X}} \rightarrow \mathcal{H}_{\mathcal{X}}$*

$$C_X = \int_{\mathcal{X}} \Phi_{\mathcal{X}}(x) \otimes \Phi_{\mathcal{X}}(x) d\mu_{\mathcal{X}}(x)$$

Definición 1.7 *Se define el operador de covarianza cruzada asociado a la dinámica como $C_{XX^+} : \mathcal{H}_{\mathcal{X}} \rightarrow \mathcal{H}_{\mathcal{X}}$*

$$C_{XX^+} = \mathbb{E}[\Phi_{\mathcal{X}}(X) \otimes \Phi_{\mathcal{X}}(X^+)] = \int_{\mathcal{X}} \int_{\mathcal{X}} \Phi_{\mathcal{X}}(x) \otimes \Phi_{\mathcal{X}}(y) \rho_f(x, dy) d\mu_{\mathcal{X}}(x)$$

y a la observación como $C_{XY} : \mathcal{H}_{\mathcal{Y}} \rightarrow \mathcal{H}_{\mathcal{X}}$

$$C_{XY} = \mathbb{E}[\Phi_{\mathcal{X}}(X) \otimes \Phi_{\mathcal{Y}}(Y)] = \int_{\mathcal{X}} \int_{\mathcal{Y}} \Phi_{\mathcal{X}}(x) \otimes \Phi_{\mathcal{Y}}(y) \rho_g(x, dy) d\mu_{\mathcal{X}}(x)$$

El adjunto de estos operadores es sencillo de calcular

Proposición 1.4 *Sea $x \in \mathcal{X}$ e $y \in \mathcal{Y}$, luego*

$$(\Phi_{\mathcal{X}}(x) \otimes \Phi_{\mathcal{Y}}(y))^* = \Phi_{\mathcal{Y}}(y) \otimes \Phi_{\mathcal{X}}(x)$$

DEMOSTRACIÓN. Sean $h_{\mathcal{X}} \in \mathcal{H}_{\mathcal{X}}$, $h_{\mathcal{Y}} \in \mathcal{H}_{\mathcal{Y}}$, luego

$$\langle (\Phi_{\mathcal{X}}(x) \otimes \Phi_{\mathcal{Y}}(y)) h_{\mathcal{Y}}, h_{\mathcal{X}} \rangle = \langle h_{\mathcal{Y}}(y) \Phi_{\mathcal{X}}(x), h_{\mathcal{X}} \rangle = h_{\mathcal{Y}}(y) \langle \Phi_{\mathcal{X}}(x), h_{\mathcal{X}} \rangle$$

Por propiedad reproductiva

$$h_{\mathcal{Y}}(y) \langle \Phi_{\mathcal{X}}(x), h_{\mathcal{X}} \rangle = h_{\mathcal{Y}}(y) h_{\mathcal{X}}(x)$$

A la vez

$$\langle h_Y, (\Phi_Y(y) \otimes \Phi_X(x)) h_X \rangle = \langle h_Y, h_X(x) \Phi_Y(y) \rangle = h_X(x) \langle h_Y, \Phi_Y(y) \rangle =$$

Nuevamente, por propiedad reproductiva

$$h_X(x) \langle h_Y, \Phi_Y(y) \rangle = h_X(x) h_Y(y)$$

Y se concluye lo pedido. \square

Corolario 1.1 $(C_X)^* = C_X$.

Dado que la integral es lineal

$$(C_X)^* = \int_{\mathcal{X}} (\Phi_X(x) \otimes \Phi_X(x))^* d\mu_X(x) = \int_{\mathcal{X}} \Phi_X(x) \otimes \Phi_X(x) d\mu_X(x) = C_X$$

Corolario 1.2

$$(C_{XX+})^* = C_{X+X}, \quad (C_{XY})^* = C_{YX}$$

Esto es directo ya que el operador adjunto y la esperanza son lineales:

$$(C_{XX+})^* = \mathbb{E}[(\Phi_X(X) \otimes \Phi_X(X^+))^*] = \mathbb{E}[\Phi_X(X^+) \otimes \Phi_X(X)] = C_{X+X}$$

$$(C_{XY})^* = \mathbb{E}[(\Phi_X(X) \otimes \Phi_Y(Y))^*] = \mathbb{E}[\Phi_X(Y) \otimes \Phi_X(X)] = C_{YX}$$

Ahora la primera herramienta importante, que se denomina la Kernel Bayes Rule y que generaliza la regla de Bayes usual cuando hace tiene un embedding de distribuciones en un RKHS

Definición 1.8 (Operador de embedding condicional) *Se define el operador de embedding condicional entre 2 distribuciones X e Y , en \mathcal{X} e \mathcal{Y} , respectivamente, como el operador $C_{X|Y} : \mathcal{H}_X \rightarrow \mathcal{H}_Y$ que satisface*

$$1. \mu_{Y|x} = \mathbb{E}_{Y|X}[\Phi_Y(Y)|X = x] = C_{Y|X} \Phi_X(x)$$

$$2. \mathbb{E}_{Y|X}[h(Y)|X = x] = \langle h, \mu_{Y|x} \rangle$$

El siguiente teorema indica la existencia y forma del operador de embedding condicional [9, 10].

Proposición 1.5 *Si $\mathbb{E}[h(Y)|X = \cdot] \in \mathcal{H}_X$ se cumple para $h \in \mathcal{H}_Y$, entonces*

$$C_{XX} \mathbb{E}[h(Y)|X = \cdot] = C_{XY} g.$$

Proposición 1.6 *Suponiendo que $\mathbb{E}[h(Y)|X = \cdot] \in \mathcal{H}_X$ para cualquier $h \in \mathcal{H}_Y$ entonces se tiene que*

$$C_{Y|X} = C_{YX} C_X^{-1}$$

1.3. Operador de Koopman

Se considera primero un sistema dinámico autónomo determinista.

$$\mathbf{x}_{k+1} = \mathbf{f}(\mathbf{x}_k), \quad k \geq 0 \quad (\text{NL})$$

Con $\mathcal{X} \subset \mathbb{R}^n$ el espacio de estados. En este contexto, a las funciones $\varphi : \mathcal{X} \rightarrow \mathbb{C}$ se les denotará por *observables* en \mathcal{F} un espacio de Banach, con norma $\|\cdot\|_{\mathcal{F}}$. Con esto se puede definir el operador de Koopman asociado a una dinámica en tiempo discreto.

Definición 1.9 (Operador de Koopman, tiempo discreto) *El operador de Koopman asociado a $\mathbf{f} : \mathcal{X} \rightarrow \mathcal{X}$ es el operador $\mathcal{K}_{\mathbf{f}} : \mathcal{F} \rightarrow \mathcal{F}$ que se define mediante composición*

$$\mathcal{K}_{\mathbf{f}}\varphi = \varphi \circ \mathbf{f}, \quad \forall \varphi \in \mathcal{F}$$

El operador de Koopman da un panorama global sobre el sistema estudiado y además resulta ser un operador lineal, aunque infinito-dimensional. En lo que sigue se omitirá el subíndice \mathbf{f} y se denotará simplemente por \mathcal{K} y $\|\mathcal{K}\|$ será su norma de operador definida como

$$\|\mathcal{K}\| = \sup\{\|\mathcal{K}g\|_{\mathcal{F}} : \|g\|_{\mathcal{F}} = 1\}$$

que depende del espacio \mathcal{F} en donde se esté trabajando, aunque en cierto tipos de espacios se cumple lo siguiente.

Proposición 1.7 *Para todo espacio de Banach \mathcal{F} que contenga a las funciones constantes, se cumple que $\|\mathcal{K}\| \geq 1$.*

En lo que sigue se trabajarán con espacios \mathcal{F} en donde \mathcal{K} cumpla $\|\mathcal{K}\| = 1$, ya que tendrán mejores propiedades espectrales. Esto se cumple al menos en espacio de funciones acotadas.

Proposición 1.8 *Si $\mathcal{F} = L^\infty$, entonces el operador de Koopman es un operador con norma unitaria.*

Un caso aún más interesante, pero difícil de cumplir para un sistema dinámico cualquiera y complicado de comprobar en la práctica, es el caso de un sistema que preserve medida.

Definición 1.10 (Sistema dinámico que preserva medida) *Sea \mathcal{X} un conjunto, \mathcal{B} una σ -álgebra sobre \mathcal{X} , $\mu : \mathcal{B} \rightarrow [0, 1]$ una medida de probabilidad y $F : \mathcal{X} \rightarrow \mathcal{X}$ una función medible. Se dice que $(\mathcal{X}, \mathcal{B}, \mu, \mathbf{f})$ preserva medida si $\mu(A) = \mu(\mathbf{f}^{-1}(A))$.*

Proposición 1.9 *Sea \mathcal{B} una σ -álgebra sobre X y $(X, \mathcal{B}, \mu, \mathbf{f})$ un sistema dinámico que preserva medida, entonces si $\mathcal{F} = L^2(\mu)$, se tiene que $\|\mathcal{K}\| = 1$. Es más, \mathcal{K} resulta ser un isomorfismo isométrico, esto es*

$$\langle g, g \rangle = \langle \mathcal{K}g, \mathcal{K}g \rangle, \quad \forall g \in \mathcal{F}$$

Durante el estudio del operador de Koopman y su representación en dimensión finita, será de interés estudiar los valores y funciones propias de Koopman.

Definición 1.11 (Valores y funciones propias) *Una función propia del operador de Koopman*

asociado a la función F es una función $\phi_\mu \in \mathcal{F} \setminus \{0\}$ tal que

$$\mathcal{K}\phi_\mu = \phi_\mu \circ \mathbf{f} = \mu\phi_\mu$$

Con $\mu \in \mathbb{C}$ su valor propio asociado.

Proposición 1.10 Para todo valor propio $\mu \in \mathbb{C}$ de \mathcal{K} , se tiene que $|\mu| \leq \|\mathcal{K}\|$.

Serán de interés las funciones propias y valores propios del operador de Koopman en el caso en que la dinámica del sistema sea lineal, en donde estas se corresponden con vectores y valores propios de la matriz que caracteriza el sistema.

Proposición 1.11 Si $\mathbf{f}(\mathbf{x}) = \mathbf{A}\mathbf{x}$, con $\mathbf{A} \in \mathbb{R}^{n \times n}$, $\mathcal{F} = C(\mathbb{R}^n; \mathbb{C})$ y \mathbf{A} posee valores propios $\{\mu_j\}_{j=1}^n$ y vectores propios por izquierda $\{\mathbf{w}_j\}_{j=1}^n$, es decir $\mathbf{w}_j^T \mathbf{A} = \mu_j \mathbf{w}_j^T$, entonces $\{\mu_j\}_{j=1}^n$ son valores propios de \mathcal{K} con funciones propias asociadas $\phi_j(\mathbf{x}) = \mathbf{w}_j^T \mathbf{x}$.

En el contexto lineal, los valores propios $\{\mu_j\}_{j=1}^n$ y $\{\phi_j\}_{j=1}^n$ se dicen valores y funciones propias principales, respectivamente.

Otra propiedad espectral importante, es que producto de funciones propias es función propia, en el caso en que esto haga sentido.

Proposición 1.12 Si \mathcal{F} es cerrado para la multiplicación y $\phi_{\mu_1}, \phi_{\mu_2}$ son funciones propias de \mathcal{K} con valores propios asociados μ_1, μ_2 , respectivamente, entonces $\phi_{\mu_1}\phi_{\mu_2}$ es función propia de \mathcal{K} con valor propio asociado $\mu_1\mu_2$.

Con las propiedades espectrales del operador de Koopman, se puede entender una descomposición espectral en este contexto.

Definición 1.12 (Desarrollo en modos de Koopman) Suponiendo que $\{\phi_{\mu_j}\}_{j \geq 1}$, las funciones propias de \mathcal{K} , generan un denso en \mathcal{F} , entonces una expansión en modos de Koopman de una función g en $\text{span}(\{\phi_{\mu_j}\}_{j \geq 1})$ es

$$g = \sum_{j \geq 1} \nu_j \phi_{\mu_j}$$

A los coeficientes ν_j se les denota los modos de Koopman de g .

1.4. Extended Dynamic Mode Decomposition (EDMD)

Por lo general, calcular valores y funciones propias del operador de Koopman es muy difícil, tanto analítica como computacionalmente. Es por ello que en la literatura ha nacido la idea de poder obtener aproximación de valores y funciones propias del operador con datos de realizaciones de \mathbf{f} .

Dynamic Mode Decomposition (DMD), descrita por primera vez en [?], es de las primeras con este espíritu, aunque se ha visto mejorada por Extended Dynamic Mode Decomposition (EDMD) presentada en [11] y que es la que se utilizará durante este trabajo.

Sean $(\{\mathbf{x}_j\}_{j=1}^N, \{\mathbf{y}_j\}_{j=1}^N)$ realizaciones de la dinámica, es decir, $\mathbf{y}_j = \mathbf{f}(\mathbf{x}_j)$, usualmente llamados *snapshots* en la literatura, y $\{\Psi_k\}_{k=1}^{N_K}$, con $\Psi_k : \mathcal{X} \rightarrow \mathbb{C}$ funciones generadoras de $\mathcal{F}_{N_K} \subset \mathcal{F}$ de un espacio de funciones finitamente generado. Se define $\Psi : \mathcal{X} \rightarrow \mathbb{C}^{N_K}$ como

$$\Psi(\mathbf{x}) = (\Psi_1(\mathbf{x}), \dots, \Psi_{N_K}(\mathbf{x}))^T$$

que se llamará función de *lifting forward*. Se supone además que existe $\mathbf{B} \in \mathbb{C}^{n \times N_K}$ tal que

$$\mathbf{B}\Psi(\mathbf{x}) = \mathbf{x}, \forall \mathbf{x} \in \mathcal{X}$$

denominada matriz de *lifting back*. Entonces el objetivo es construir una aproximación lineal de la dinámica en dimensión N_K de la forma

$$\mathbf{z}_{k+1} = \mathbf{K}\mathbf{z}_k$$

Notar que al aplicar Ψ a la dinámica se obtiene

$$\Psi(\mathbf{y}_j) = \Psi(\mathbf{f}(\mathbf{x}_j)) = (\mathcal{K}\Psi)(\mathbf{x}_j) = \mathbf{K}\Psi(\mathbf{x}_j) + \mathbf{r}(\mathbf{x}_j)$$

Donde $\mathbf{r} \in \mathcal{F}$ es una función de residuo, que buscamos minimizar en las realizaciones. Con lo que si denotamos

$$\mathbf{P}_X = (\Psi(\mathbf{x}_1) | \dots | \Psi(\mathbf{x}_N)) \in \mathbb{R}^{N_K \times N}, \mathbf{P}_Y = (\Psi(\mathbf{y}_1) | \dots | \Psi(\mathbf{y}_N)) \in \mathbb{R}^{N_K \times N}$$

Se obtiene un problema de regresión lineal

$$\mathbf{P}_Y = \mathbf{K}\mathbf{P}_X + \mathbf{R}(\mathbf{P}_X)$$

Formulado como

$$\min_{\mathbf{K} \in \mathbb{R}^{N_K \times N_K}} J(\mathbf{K}) = \frac{1}{2} \|\mathbf{P}_Y - \mathbf{K}\mathbf{P}_X\|_F^2$$

Donde $\|\cdot\|_F$ es la norma Frobenius de matrices definida como

$$\|A\|_F^2 = \sum_{i=1}^m \sum_{j=1}^n |a_{ij}|^2, \quad A \in \mathbb{R}^{m \times n}$$

Si \mathbf{P}_X tiene rango completo, entonces el problema tiene un único mínimo global dado por

$$\mathbf{K} = \mathbf{P}_Y \mathbf{P}_X^\dagger$$

Donde, para $\mathbf{A} \in \mathbb{R}^{n \times m}$ la matriz $\mathbf{A}^\dagger \in \mathbb{R}^{m \times n}$ es la pseudoinversa de Moore-Penrose, que si \mathbf{A} tiene rango completo tiene forma explícita:

- Si \mathbf{A} inyectiva: entonces $\mathbf{A}^T \mathbf{A}$ es invertible y así

$$\mathbf{A}^\dagger = (\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T$$

- Si \mathbf{A} sobreyectiva: entonces $\mathbf{A} \mathbf{A}^T$ es invertible y así

$$\mathbf{A}^\dagger = \mathbf{A}^T (\mathbf{A} \mathbf{A}^T)^{-1}$$

La matriz \mathbf{K} actúa como una aproximación en dimensión finita del operador de Koopman. Así, llamando $\mathbf{z}_k = \Psi(\mathbf{x}_k)$, se obtiene el siguiente sistema lineal en dimensión finita N_K

$$\begin{aligned} \mathbf{z}_{k+1} &= \mathbf{K}\mathbf{z}_k \\ \mathbf{x}_k &= \mathbf{B}\mathbf{z}_k \\ \mathbf{z}_0 &= \Psi(\mathbf{x}_0) \end{aligned} \tag{L}$$

Uno de los objetivos de esta tesis es entregar resultados que permitan afirmar que trayectorias o soluciones de (L) aproximan, en algún sentido, a las de (NL). Todo lo anterior se puede entender con el diagrama expuesto en la figura 1.1.

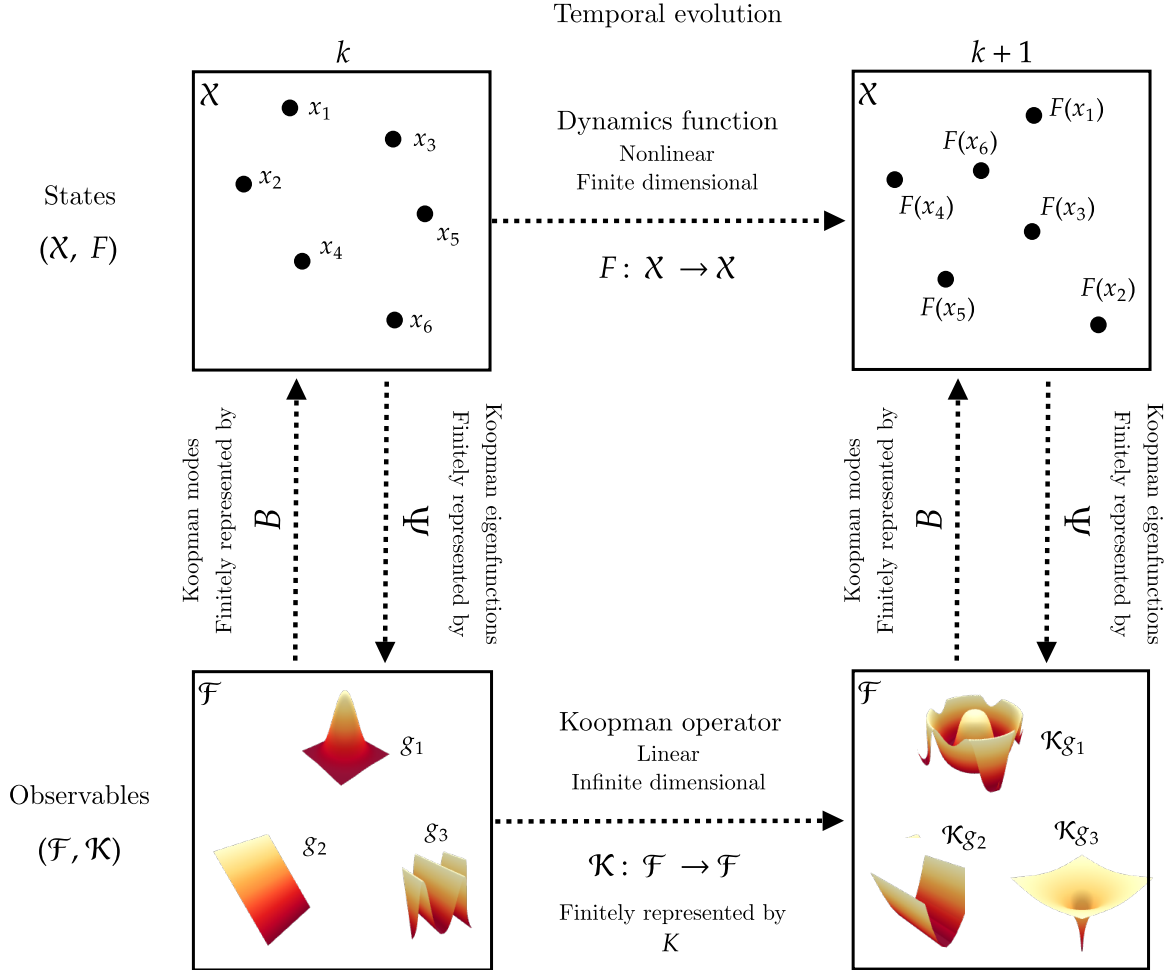


Figura 1.1: Diagrama de la evolución temporal tanto en dimensión finita como en dimensión infinita para la dinámica representada por el operador de Koopman, con las respectivas representaciones en dimensión finita de cada uno de los objetos involucrados. Elaboración propia, basado en [11].

1.5. Aproximación de rango bajo de matrices

La dimensión de aproximación del operador de Koopman puede ser muy alta para ciertas aplicaciones, lo que hace que ciertas operaciones se hagan muy inestables si se mantiene el

rango original. Es por ello que en muchos contextos se propone bajar el rango de las matrices involucradas vía Descomposición en Valores Singulares (SVD). Primero consideremos el problema de obtener la mejor matriz de rango bajo que aproxima otra matriz dada. La solución de este problema fue dada en [?], para ello se denota por $r(\mathbf{A})$ el rango de una matriz $\mathbf{A} \in \mathbb{R}^{m \times n}$.

Lema 1.4 *La mejor aproximación de rango s , en términos de la norma de Frobenius, para una matriz \mathbf{A} de rango t con $t \geq s$, es decir, un minimizador global $\hat{\mathbf{A}}^*$ de*

$$\min_{\hat{\mathbf{A}}} \|\mathbf{A} - \hat{\mathbf{A}}\|_F, \quad \text{s.a.} \quad r(\hat{\mathbf{A}}) \leq s$$

viene dada por

$$\hat{\mathbf{A}}^* = P_s(\mathbf{A}) = \mathbf{U}\Sigma_s\mathbf{V}^T,$$

donde Σ_s es la matriz diagonal con los s valores singulares más grandes de \mathbf{A} y luego solo 0, en donde se denota a la SVD de $\mathbf{A} = \mathbf{U}\Sigma\mathbf{V}^T$.

Es con esto que en [12] se provee una forma cerrada para un problema de regresión lineal en el que se busca que la matriz solución sea de rango bajo también, esto se formula como

$$\min_{\mathbf{M}} \|\mathbf{Y} - \mathbf{MX}\|_F, \quad \text{s.a.} \quad r(\mathbf{M}) \leq s \quad (1.3)$$

Proposición 1.13 *Sea la SVD de $\mathbf{X} = \mathbf{U}\Sigma\mathbf{V}^T$, con $\mathbf{U} \in \mathbb{R}^{m \times m}$, $\mathbf{V} \in \mathbb{R}^{n \times n}$ matrices ortogonales y Σ la matriz diagonal con los valores singulares. Luego, el óptimo debe cumplir*

$$\mathbf{V}^T\mathbf{M}^* = \begin{bmatrix} \Sigma_{r(\mathbf{X})}^{-1} P_s(\mathbf{W}_{r(\mathbf{X})}) \\ \mathbf{a} \end{bmatrix},$$

donde $\Sigma_{r(\mathbf{X})}$ es la matriz diagonal con los valores singulares no nulos de \mathbf{X} , $\mathbf{W}_{r(\mathbf{X})}$ aquella con las primeras $r(\mathbf{X})$ filas de \mathbf{W} , con $\mathbf{W} = \mathbf{U}^T\mathbf{Y}$.

Este resultado será importante para poder trabajar matrices de alta dimensionalidad, como ocurre en el caso de EDMD.

Capítulo 2

Kernel Extended Dynamic Mode Decomposition

2.1. Extended Dynamic Mode Decomposition

Denotemos por $\mathcal{B}_{\mathcal{X}}$ la σ -álgebra Boreliana de \mathcal{X} y $\mathcal{B}_{\mathcal{Y}}$ a la de \mathbb{R}^p . Se definen las medidas de probabilidad

$$\begin{aligned}\rho_f : \mathcal{X} \times \mathcal{B}_{\mathcal{X}} &\rightarrow [0, 1], & \rho_f(\mathbf{x}, A) &= \mathbb{P}(\mathbf{f}(\mathbf{x}, \cdot) \in A) \\ \rho_g : \mathcal{X} \times \mathcal{B}_{\mathcal{Y}} &\rightarrow [0, 1], & \rho_g(\mathbf{x}, A) &= \mathbb{P}(\mathbf{g}(\mathbf{x}, \cdot) \in A)\end{aligned}$$

Es decir, ρ_f es la medida inducida por la dinámica y ρ_g es la medida inducida por la observación.

Supondremos que el espacio de estados \mathcal{X} es compacto y que existe un conjunto compacto $\mathcal{Y} \subseteq \mathbb{R}^p$ tal que

$$\rho_f(\mathbf{x}, \mathcal{X}) = 1, \quad \rho_g(\mathbf{x}, \mathcal{Y}) = 1, \quad \forall \mathbf{x} \in \mathcal{X}$$

Adoptaremos la notación

$$\rho_f(\mathbf{x}, dx) = d\rho_f(\mathbf{x}, \cdot)(x), \quad \rho_g(\mathbf{x}, dy) = d\rho_g(\mathbf{x}, \cdot)(y)$$

Supondremos además que, para $\mu_{\mathcal{X}}$ medida sobre \mathcal{X} y $\mu_{\mathcal{Y}}$ medida sobre \mathcal{Y} , existen

$$p_f : \mathcal{X} \times \mathcal{X} \rightarrow \mathbb{R}_+, \quad p_g : \mathcal{X} \times \mathcal{Y} \rightarrow \mathbb{R}_+$$

tales que

$$\rho_f(\mathbf{x}, A) = \int_A p_f(\mathbf{x}, y) d\mu_{\mathcal{X}}(y), \quad \rho_g(\mathbf{x}, A) = \int_A p_g(\mathbf{x}, y) d\mu_{\mathcal{Y}}(y)$$

Llamaremos:

- X la variable aleatoria asociada a $\mu_{\mathcal{X}}$.
- $X^+|\mathbf{x}$ la variable aleatoria asociada a $\rho_{\mathcal{X}}$, es decir, la variable aleatoria asociada a avanzar un paso, dado \mathbf{x} .
- $Y|\mathbf{x}$ la variable aleatorias asociada a $\rho_{\mathcal{Y}}$, es decir, la variable aleatoria asociada a la observación, dado \mathbf{x} .

Por ejemplo, si la función de dinámica es de la forma

$$\mathbf{f}(\mathbf{x}_k, \mathbf{w}_k) = \tilde{\mathbf{f}}(\mathbf{x}_k) + \mathbf{w}_k$$

con $\mathbf{w}_k \sim \mathcal{N}(0, \mathbf{Q}_k)$, siendo \mathbf{Q}_k definida positiva y $\mu_{\mathcal{X}}$ la medida de Lebesgue sobre \mathcal{X} , se tiene que

$$\rho_f(\mathbf{x}_k, \cdot) \sim \mathcal{N}(\tilde{\mathbf{f}}(\mathbf{x}_k), \mathbf{Q}_k)$$

con lo que p_f es

$$p_f(x, y) = (2\pi \det \mathbf{Q}_k)^{-1} \exp \left(-(\tilde{\mathbf{f}}(x) - y)^\top \mathbf{Q}_k^{-1} (\tilde{\mathbf{f}}(x) - y) \right)$$

que cumple ser acotada, incluso si $\tilde{\mathbf{f}}$ no lo es.

Similar a [13] se asume lo siguiente

- a) $k_{\mathcal{X}} : \mathcal{X} \times \mathcal{X} \rightarrow \mathbb{R}$, $k_{\mathcal{Y}} : \mathcal{Y} \times \mathcal{Y} \rightarrow \mathbb{R}$ dos kernels simétricos, continuos, acotados y semi-definidos positivos. Se denotará por $\mathcal{H}_{\mathcal{X}}$, $\mathcal{H}_{\mathcal{Y}}$ a su RKHS asociados.
- b) Si $\psi_{\mathcal{X}} \in L^2(\mathcal{X})$, $\psi_{\mathcal{Y}} \in L^2(\mathcal{Y})$ son tales que

$$\int_{\mathcal{X} \times \mathcal{X}} k_{\mathcal{X}}(x, y) \psi_{\mathcal{X}}(x) \psi_{\mathcal{X}}(y) d\mu_{\mathcal{X}}(x) d\mu_{\mathcal{X}}(y) = 0$$

$$\int_{\mathcal{Y} \times \mathcal{Y}} k_{\mathcal{Y}}(x, y) \psi_{\mathcal{Y}}(x) \psi_{\mathcal{Y}}(y) d\mu_{\mathcal{Y}}(x) d\mu_{\mathcal{Y}}(y) = 0$$

entonces $\psi_{\mathcal{X}} = 0$, $\psi_{\mathcal{Y}} = 0$ c.s.

- c) Si $\psi_{\mathcal{X}} \in \mathcal{H}_{\mathcal{X}}$, $\psi_{\mathcal{Y}} \in \mathcal{H}_{\mathcal{Y}}$ son tales que $\psi_{\mathcal{X}}(x) = 0$, $\psi_{\mathcal{Y}}(y) = 0$ para todo $x \in \mathcal{X}$ $\mu_{\mathcal{X}}$ -c.s., y para todo $y \in \mathcal{Y}$ $\mu_{\mathcal{Y}}$ -c.s. entonces $\psi_{\mathcal{X}} \equiv 0$, $\psi_{\mathcal{Y}} \equiv 0$.
- d) Se cumplen las siguientes relaciones de $\rho_{\mathcal{X}}$ y $\rho_{\mathcal{Y}}$ con respecto de $\mu_{\mathcal{X}}$ y $\mu_{\mathcal{Y}}$:

$$\int_{\mathcal{X}} \rho_{\mathcal{X}}(x, A_{\mathcal{X}}) d\mu_{\mathcal{X}}(x) \leq L_{\mathcal{X}} \mu_{\mathcal{X}}(A_{\mathcal{X}}), \quad \forall A_{\mathcal{X}} \in \mathcal{B}_{\mathcal{X}}$$

$$\int_{\mathcal{X}} \rho_{\mathcal{Y}}(x, A_{\mathcal{Y}}) d\mu_{\mathcal{X}}(x) \leq L_{\mathcal{Y}} \mu_{\mathcal{Y}}(A_{\mathcal{Y}}), \quad \forall A_{\mathcal{Y}} \in \mathcal{B}_{\mathcal{Y}}$$

Un ejemplo de kernel que cumple a) y b) es Matérn, siendo el punto a) expuesto en la sección anterior y b) debido a la universalidad en L^2 . La suposición c) se cumple si μ tiene densidad con respecto a Lebesgue, mientras que d) se cumple en el caso estocástico cuando existen las funciones p_f y p_g y estas son acotadas, mientras que en el caso determinista se necesita que las funciones asociadas sean difeomorfismos.

Proposición 2.1 Si $p_f \in L^\infty(\mathcal{X} \times \mathcal{X})$, $p_g \in L^\infty(\mathcal{X} \times \mathcal{Y})$ entonces

$$\int_{\mathcal{X}} \rho_{\mathcal{X}}(x, A) d\mu_{\mathcal{X}}(x) \leq L_{\mathcal{X}} \mu_{\mathcal{X}}(A), \quad \int_{\mathcal{X}} \rho_{\mathcal{Y}}(x, A) d\mu_{\mathcal{X}}(x) \leq L_{\mathcal{Y}} \mu_{\mathcal{Y}}(A)$$

con

$$L_{\mathcal{X}} = \mu_{\mathcal{X}}(\mathcal{X}) \|p_f\|_\infty, \quad L_{\mathcal{Y}} = \mu_{\mathcal{X}}(\mathcal{X}) \|p_g\|_\infty$$

DEMOSTRACIÓN. Para $\rho_{\mathcal{X}}$ se tiene que

$$\begin{aligned}
\int_{\mathcal{X}} \rho_{\mathcal{X}}(x, A) d\mu_{\mathcal{X}}(x) &= \int_{\mathcal{X}} \int_A d\rho(x, \cdot)(y) d\mu(x) \\
&= \int_{\mathcal{X}} \int_A p_{\mathbf{f}}(x, y) d\mu_{\mathcal{X}}(y) d\mu_{\mathcal{X}}(x) \\
&\leq \|p_{\mathbf{f}}\|_{\infty} \int_{\mathcal{X}} \int_A d\mu_{\mathcal{X}}(y) d\mu_{\mathcal{X}}(x) \\
&= \|p_{\mathbf{f}}\|_{\infty} \mu_{\mathcal{X}}(\mathcal{X}) \mu_{\mathcal{X}}(A) \\
&= L_{\mathcal{X}} \mu_{\mathcal{X}}(A)
\end{aligned}$$

Mientras que para $\rho_{\mathcal{Y}}$ se tiene que

$$\begin{aligned}
\int_{\mathcal{X}} \rho_{\mathcal{Y}}(x, A) d\mu(x) &= \int_{\mathcal{X}} \int_A d\rho_{\mathcal{Y}}(x, \cdot)(y) d\mu_{\mathcal{X}}(x) \\
&= \int_{\mathcal{X}} \int_A p_{\mathbf{g}}(x, y) d\mu_{\mathcal{Y}}(y) d\mu_{\mathcal{X}}(x) \\
&\leq \|p_{\mathbf{g}}\|_{\infty} \int_{\mathcal{X}} \int_A d\mu_{\mathcal{Y}}(y) d\mu_{\mathcal{X}}(x) \\
&= \|p_{\mathbf{g}}\|_{\infty} \mu_{\mathcal{X}}(\mathcal{X}) \mu_{\mathcal{Y}}(A) \\
&= L_{\mathcal{Y}} \mu_{\mathcal{Y}}(A)
\end{aligned}$$

□

En el caso en que la dinámica o la observación sean deterministas, que es equivalente a que $\mathbf{f}(\mathbf{x}, \mathbf{w}) = \tilde{\mathbf{f}}(\mathbf{x})$ o $\mathbf{g}(\mathbf{x}, \mathbf{v}) = \tilde{\mathbf{g}}(\mathbf{x})$, con lo que $\rho(x, \cdot) = \delta_{\tilde{\mathbf{f}}(x)}(\cdot)$ o $\xi(x, \cdot) = \delta_{\tilde{\mathbf{f}}(x)}(\cdot)$, entonces se necesita mayor regularidad sobre las funciones.

Proposición 2.2 Si $\rho(x, \cdot) = \delta_{\tilde{\mathbf{f}}(x)}(\cdot)$ o $\xi(x, \cdot) = \delta_{\tilde{\mathbf{f}}(x)}(\cdot)$ y $\tilde{\mathbf{f}}$ o $\tilde{\mathbf{g}}$, según corresponda, son difeomorfismos C^1 tal que

$$\inf_{x \in \mathcal{X}} |\det D\tilde{\mathbf{f}}(x)| > 0, \quad \inf_{x \in \mathcal{X}} |\det D\tilde{\mathbf{g}}(x)| > 0$$

entonces

$$\int_{\mathcal{X}} \rho(x, A) d\mu(x) \leq L_{\rho} \mu(A), \quad \int_{\mathcal{X}} \xi(x, A) d\mu(x) \leq L_{\xi} \mu(A)$$

con

$$L_{\rho} = \|\det D\tilde{\mathbf{f}}^{-1}\|_{\infty}, \quad L_{\xi} = \|\det D\tilde{\mathbf{g}}^{-1}\|_{\infty}$$

DEMOSTRACIÓN. La demostración se hace similar a [14]. Notar que para $A \in \mathcal{B}$, se tiene que $\rho(x, A) = \delta_{\tilde{\mathbf{f}}(x)}(A) = \mathbb{1}_A(\tilde{\mathbf{f}}(x))$, con ello

$$\begin{aligned}
\int_{\mathcal{X}} \rho(x, A) d\mu(x) &= \int_{\mathcal{X}} \mathbb{1}_A(\tilde{\mathbf{f}}(x)) d\mu(x) \\
&= \int_{\mathcal{X}} \mathbb{1}_A(x) |\det D\tilde{\mathbf{f}}^{-1}(x)| d\mu(x) \\
&\leq \|\det D\tilde{\mathbf{f}}^{-1}\|_{\infty} \int_{\mathcal{X}} \mathbb{1}_A(x) d\mu(x) \\
&\leq \|\det D\tilde{\mathbf{f}}^{-1}\|_{\infty} \mu(A) \\
&= L_{\rho} \mu(A)
\end{aligned}$$

Y para ξ es análogo. □

Ahora se definen los operadores de Koopman estocásticos para la dinámica y la observación, adaptados en el caso en que se tienen las funciones p_f y p_g como densidades.

Definición 2.1 (Operador de Koopman estocástico para la dinámica) *Se define el operador asociado a \mathbf{f} como $\mathcal{U} : L^2(\mathcal{X}) \rightarrow L^2(\mathcal{X})$*

$$[\mathcal{U}h](x) = \mathbb{E}[h(\mathbf{f}(x, \cdot))] = \int_{\mathcal{X}} h(y) d\rho_{\mathcal{X}}(x, \cdot)(y) = \int_{\mathcal{X}} h(y) p_f(x, y) d\mu_{\mathcal{X}}(y)$$

Definición 2.2 (Operador de Koopman estocástico para la observación) *Se define el operador asociado a \mathbf{g} como $\mathcal{G} : L^2(\mathcal{Y}) \rightarrow L^2(\mathcal{X})$*

$$[\mathcal{G}h](x) = \mathbb{E}[h(\mathbf{g}(x, \cdot))] = \int_{\mathcal{Y}} h(y) d\rho_{\mathcal{Y}}(x, \cdot)(y) = \int_{\mathcal{Y}} h(y) p_g(x, y) d\mu_{\mathcal{Y}}(y)$$

Un objeto que tendrá interés pronto será el operador de Perron-Frobenius

Definición 2.3 (Operador de Perron-Frobenius estocástico para la dinámica) *Se define el operador asociado a \mathbf{f} como $\mathcal{P}_{\mathbf{f}} : L^2(\mathcal{X}) \rightarrow L^2(\mathcal{X})$*

$$[\mathcal{P}_{\mathbf{f}}h](x) = \int_{\mathcal{X}} h(y) p_{\mathbf{f}}(y, x) d\mu_{\mathcal{X}}(y)$$

Definición 2.4 (Operador de Perron-Frobenius estocástico para la observación) *Se define el operador asociado a \mathbf{g} como $\mathcal{P}_{\mathbf{g}} : L^2(\mathcal{X}) \rightarrow L^2(\mathcal{Y})$*

$$[\mathcal{P}_{\mathbf{g}}h](x) = \int_{\mathcal{X}} h(y) p_{\mathbf{g}}(y, x) d\mu_{\mathcal{X}}(y)$$

Entonces notar que, gracias a la representación de los operadores a través de $p_{\mathbf{f}}$ y $p_{\mathbf{g}}$ se tiene que

$$\mathcal{U}^* = \mathcal{P}_{\mathbf{f}}, \quad \mathcal{G}^* = \mathcal{P}_{\mathbf{g}}$$

Definición 2.5 (Feature map) *Se definen $\Phi_{\mathcal{X}} : \mathcal{X} \rightarrow \mathcal{H}_{\mathcal{X}}$, $\Phi_{\mathcal{Y}} : \mathcal{Y} \rightarrow \mathcal{H}_{\mathcal{Y}}$ los feature maps de ambos kernels como*

$$\Phi_{\mathcal{X}}(x) = k_{\mathcal{X}}(x, \cdot), \quad \Phi_{\mathcal{Y}}(y) = k_{\mathcal{Y}}(y, \cdot)$$

Definición 2.6 *Se definen, para $x_1, x_2 \in \mathcal{X}$, $y_1, y_2 \in \mathcal{Y}$ los operadores de rango 1 $C_{x_1, x_2} : \mathcal{H}_{\mathcal{X}} \rightarrow \mathcal{H}_{\mathcal{X}}$, $C_{y_1, y_2} : \mathcal{H}_{\mathcal{Y}} \rightarrow \mathcal{H}_{\mathcal{Y}}$ y $C_{y_1, x_1} : \mathcal{H}_{\mathcal{X}} \rightarrow \mathcal{H}_{\mathcal{Y}}$ respectivamente, como*

$$C_{x_1, x_2} \psi = [\Phi_{\mathcal{X}}(x_1) \otimes \Phi_{\mathcal{X}}(x_2)] \psi = \langle \psi, \Phi_{\mathcal{X}}(x_2) \rangle \Phi_{\mathcal{X}}(x_1) = \psi(x_2) \Phi_{\mathcal{X}}(x_1)$$

$$C_{y_1, y_2} \psi = [\Phi_{\mathcal{Y}}(y_1) \otimes \Phi_{\mathcal{Y}}(y_2)] \psi = \langle \psi, \Phi_{\mathcal{Y}}(y_2) \rangle \Phi_{\mathcal{Y}}(y_1) = \psi(y_2) \Phi_{\mathcal{Y}}(y_1)$$

$$C_{y_1, x_1} \psi = [\Phi_{\mathcal{Y}}(y_1) \otimes \Phi_{\mathcal{X}}(x_1)] \psi = \langle \psi, \Phi_{\mathcal{X}}(x_1) \rangle \Phi_{\mathcal{Y}}(y_1) = \psi(x_1) \Phi_{\mathcal{Y}}(y_1)$$

Definición 2.7 *Se definen los operadores de covarianza $C_{\mathcal{X}} : \mathcal{H}_{\mathcal{X}} \rightarrow \mathcal{H}_{\mathcal{X}}$, $C_{\mathcal{Y}} : \mathcal{H}_{\mathcal{Y}} \rightarrow \mathcal{H}_{\mathcal{Y}}$ como*

$$C_{\mathcal{X}} \psi = \int_{\mathcal{X}} C_{x, x} \psi d\mu_{\mathcal{X}}(x) = \int_{\mathcal{X}} [\Phi_{\mathcal{X}}(x) \otimes \Phi_{\mathcal{X}}(x)] \psi d\mu_{\mathcal{X}}(x)$$

$$C_Y\psi = \int_Y C_{y,y}\psi d\mu_X(y) = \int_Y [\Phi_Y(y) \otimes \Phi_Y(y)]\psi d\mu_Y(y)$$

Definición 2.8 Se define el operador de covarianza cruzada asociada a la dinámica como el operador $C_{\mathcal{X}\mathcal{X}^+} : \mathcal{H}_{\mathcal{X}} \rightarrow \mathcal{H}_{\mathcal{X}}$ dado por

$$C_{\mathcal{X}\mathcal{X}^+}\psi = \int_{\mathcal{X}} \int_{\mathcal{X}} C_{x,y}\psi d\rho_{\mathcal{X}}(x, \cdot)(y) d\mu_{\mathcal{X}}(x) = \int_{\mathcal{X}} \int_{\mathcal{X}} [\Phi_{\mathcal{X}}(x) \otimes \Phi_{\mathcal{X}}(y)] d\rho_{\mathcal{X}}(x, \cdot)(y) d\mu_{\mathcal{X}}(x)$$

Y el operador de covarianza cruzada asociada a la observación como $C_{Y\mathcal{X}} : \mathcal{H}_{\mathcal{X}} \rightarrow \mathcal{H}_Y$ dado por

$$C_{Y\mathcal{X}}\psi = \int_{\mathcal{X}} \int_Y C_{y,x}\psi d\rho_Y(x, \cdot)(y) d\mu_{\mathcal{X}}(x) = \int_{\mathcal{X}} \int_Y [\Phi_Y(y) \otimes \Phi_{\mathcal{X}}(x)] \psi d\rho_Y(x, \cdot)(y) d\mu_{\mathcal{X}}(x)$$

Definición 2.9 (Operadores de embedding condicional [10]) Se definen los operadores de embedding condicional como $C_{\mathcal{X}^+|\mathcal{X}} : \mathcal{H}_{\mathcal{X}} \rightarrow \mathcal{H}_{\mathcal{X}}$, $C_{Y|\mathcal{X}} : \mathcal{H}_{\mathcal{X}} \rightarrow \mathcal{H}_Y$

$$C_{\mathcal{X}^+|\mathcal{X}} = C_{\mathcal{X}^+\mathcal{X}} C_{\mathcal{X}}^{-1}$$

$$C_{Y|\mathcal{X}} = C_{Y\mathcal{X}} C_{\mathcal{X}}^{-1}$$

Asumiremos ahora que $\mathcal{UH}_{\mathcal{X}} \subset \mathcal{H}_{\mathcal{X}}$ y $\mathcal{GH}_Y \subset \mathcal{H}_{\mathcal{X}}$, para ello primero asumiremos que $\mathcal{H}_{\mathcal{X}}$ y \mathcal{H}_Y son espacios de Sobolev (pensar en Matérn) y ahora la siguiente proposición.

Proposición 2.3 (Invarianza Sobolev para Koopman) Si $p_{\mathbf{f}} \in C^{k,k}(\mathcal{X} \times \mathcal{X})$ y $p_{\mathbf{g}} \in C^{k,k}(\mathcal{X} \times \mathcal{Y})$, entonces

$$\mathcal{UH}^k(\mathcal{X}) \subset \mathcal{H}^k(\mathcal{X}), \quad \mathcal{GH}^k(\mathcal{Y}) \subset \mathcal{H}^k(\mathcal{X})$$

DEMOSTRACIÓN. Basta tomar primero $m \in \mathbb{N}$, $m \leq k$ y $|\alpha| = m$ un multiíndice, luego, por teorema de la convergencia dominada, para $h \in \mathcal{H}^k(\mathcal{X})$

$$\partial_x^\alpha(\mathcal{U}h)(x) = \int_{\mathcal{X}} h(y) \partial_x^\alpha p_{\mathbf{f}}(x, y) d\mu_{\mathcal{X}}(y)$$

Así

$$\begin{aligned} \|\partial_x^\alpha(\mathcal{U}h)\|_{L^2}^2 &\leq \int_{\mathcal{X}} \int_{\mathcal{X}} h(y)^2 \partial_x^\alpha p_{\mathbf{f}}(x, y)^2 d\mu_{\mathcal{X}}(y) d\mu_{\mathcal{X}}(x) \\ &\leq \|\partial_x^\alpha p_{\mathbf{f}}\|_{C^{k,k}} \mu(\mathcal{X}) \int_{\mathcal{X}} h(y)^2 d\mu_{\mathcal{X}}(y) \\ &\leq \|\partial_x^\alpha p_{\mathbf{f}}\|_{C^{k,k}} \mu_{\mathcal{X}}(\mathcal{X}) \|h\|_{H^k(\mathcal{X})}^2 \end{aligned}$$

Con lo que

$$\|\mathcal{U}h\|_{H^k} \leq \left(\mu_{\mathcal{X}}(\mathcal{X}) \sum_{|\alpha| \leq k} \|\partial_x^\alpha p_{\mathbf{f}}\|_{C^{k,k}} \right) \|h\|_{H^k}$$

Y se concluye que $\mathcal{U}h \in H^k$, y de hecho

$$\|\mathcal{U}\|_{H^k \rightarrow H^k} \leq \mu_{\mathcal{X}}(\mathcal{X}) \sum_{|\alpha| \leq k} \|\partial_x^\alpha p_{\mathbf{f}}\|_{C^{k,k}}$$

Análogamente, para $h \in H^k(\mathcal{Y})$

$$\|\mathcal{G}h\|_{H^k(\mathcal{X})} \leq \left(\mu_{\mathcal{X}}(\mathcal{X}) \sum_{|\alpha| \leq k} \|\partial_x^\alpha p_{\mathbf{g}}\|_{C^{k,k}} \right) \|h\|_{H^k(\mathcal{Y})}$$

□

Suponiendo la invarianza de los operadores de Koopman a través del RKHS, se tiene lo siguiente

$$\begin{aligned} C_{\mathcal{X}\mathcal{X}}\psi &= \int_{\mathcal{X}} \int_{\mathcal{X}} [\Phi_{\mathcal{X}}(x) \otimes \Phi_{\mathcal{X}}(y)] \psi d\rho_{\mathcal{X}}(x, \cdot)(y) d\mu_{\mathcal{X}}(x) \\ &= \int_{\mathcal{X}} \int_{\mathcal{Y}} \psi(y) \Phi_{\mathcal{X}}(x) d\rho_{\mathcal{X}}(x, \cdot)(y) d\mu_{\mathcal{X}}(x) \\ &= \int_{\mathcal{X}} (\mathcal{U}\psi)(x) \Phi_{\mathcal{X}}(x) d\mu_{\mathcal{X}}(x) \\ &= \int_{\mathcal{X}} [\Phi_{\mathcal{X}}(x) \otimes \Phi_{\mathcal{X}}(x)] (\mathcal{U}\psi) d\mu_{\mathcal{X}}(x) \\ &= C_{\mathcal{X}} \mathcal{U} \psi \end{aligned}$$

Por tanto se tiene que $C_{\mathcal{X}\mathcal{X}} = C_{\mathcal{X}} \mathcal{U}$. Notando que

$$C_{\mathcal{X}} = \mathbb{E}[\Phi_{\mathcal{X}}(\mathcal{X}) \otimes \Phi_{\mathcal{X}}(\mathcal{X})], \quad C_{\mathcal{X}\mathcal{X}^+} = \mathbb{E}[\Phi_{\mathcal{X}}(\mathcal{X}) \otimes \Phi_{\mathcal{X}}(\mathcal{X}^+)]$$

y que

$$(\Phi_{\mathcal{X}}(\mathcal{X}) \otimes \Phi_{\mathcal{X}}(\mathcal{Y}))^* = \Phi_{\mathcal{X}}(\mathcal{Y}) \otimes \Phi_{\mathcal{X}}(\mathcal{X})$$

se tiene $C_{\mathcal{X}}^* = C_{\mathcal{X}}$ y $C_{\mathcal{X}\mathcal{X}^+}^* = C_{\mathcal{X}^+\mathcal{X}}$

$$C_{\mathcal{X}^+\mathcal{X}} C_{\mathcal{X}}^{-1} = \mathcal{U}^*$$

con lo que $C_{\mathcal{X}^+\mathcal{X}} C_{\mathcal{X}}^{-1} = C_{\mathcal{X}^+|\mathcal{X}} = \mathcal{U}^* = \mathcal{P}_{\mathbf{f}}$ y análogamente $C_{\mathcal{Y}|\mathcal{X}} = \mathcal{G}^* = \mathcal{P}_{\mathbf{g}}$. Esto debe entenderse de manera cuidadosa y está bien definido si es que $C_{\mathcal{X}\mathcal{X}}$ es inyectivo [15].

Ahora veremos cómo aproximar los tres operadores (Koopman de dinámica, observación y reconstrucción). Para ello sean N puntos (este será nuestro parámetro de precisión de la aproximación), digamos $\{x_i\}_{i=1}^N \sim \mu_{\mathcal{X}}^N$ y puntos $\{x_i^+\}_{i=1}^N$, $\{y_i\}_{i=1}^N$ sampleados tal que

$$x_i^+ \sim \rho_f(x_i, \cdot), \quad y_i \sim \rho_g(x_i, \cdot), \quad i = 1, \dots, N$$

Llamaremos al espacio

$$\mathcal{H}_{\mathcal{X},N} = \text{span}\{\Phi_{\mathcal{X}}(x_i) : i = 1, \dots, N\}$$

cuya base canónica es $\{\Phi_{\mathcal{X}}(x_i) : i = 1, \dots, N\}$. Llamamos a las matrices

$$X = (x_1 | \dots | x_N), \quad Y = (y_1 | \dots | y_N)$$

$$\Phi_N(X) = (k(x_i, x_j))_{i,j=1}^N$$

$$\Phi_N(X^+) = (k(x_i, x_j^+))_{i,j=1}^N$$

Se definen los operadores

$$C_X^N : \mathcal{H}_{\mathcal{X},N} \rightarrow \mathcal{H}_{\mathcal{X},N}, \quad C_{X^+}^N : \mathcal{H}_{\mathcal{X},N} \rightarrow \mathcal{H}_{\mathcal{X},N}$$

$$\begin{aligned}
C_{XY}^N : \mathbb{R}^p &\rightarrow \mathcal{H}_{\mathcal{X},N}, & C_{XZ}^N : \mathbb{R}^n &\rightarrow \mathcal{H}_{\mathcal{X},N} \\
C_X^N \Phi_{\mathcal{X}}(x_i) &= \frac{1}{N} \sum_{j=1}^N [\Phi_{\mathcal{X}}(x_j) \otimes \Phi_{\mathcal{X}}(x_j)] \Phi_{\mathcal{X}}(x_i) = \frac{1}{N} \sum_{j=1}^N k_{\mathcal{X}}(x_i, x_j) \Phi_{\mathcal{X}}(x_j) \\
C_{XX^+}^N \Phi_{\mathcal{X}}(x_i) &= \frac{1}{N} \sum_{j=1}^N [\Phi_{\mathcal{X}}(x_j) \otimes \Phi_{\mathcal{X}}(x_j)] \Phi_{\mathcal{X}}(x_i^+) = \frac{1}{N} \sum_{j=1}^N k_{\mathcal{X}}(x_i^+, x_j) \Phi_{\mathcal{X}}(x_j) \\
C_{XY}^N \Phi_{\mathcal{Y}}(y_i) &= \frac{1}{N} \sum_{j=1}^N [\Phi_{\mathcal{X}}(x_j) \otimes \Phi_{\mathcal{Y}}(y_j)] \Phi_{\mathcal{Y}}(y_i) = \frac{1}{N} \sum_{j=1}^N k_{\mathcal{Y}}(y_i, y_j) \Phi_{\mathcal{X}}(x_j) \\
C_{XZ}^N \phi(x_i) &= \frac{1}{N} \sum_{j=1}^N [\Phi_{\mathcal{X}}(x_j) \otimes \phi(x_j)] \phi(x_i) = \frac{1}{N} \sum_{j=1}^N \langle x_i, x_j \rangle \Phi_{\mathcal{X}}(x_j)
\end{aligned}$$

Que vienen representados por $\Phi_N(X)$, $\Phi_N(X^+)$, Y , X , respectivamente, con respecto a su base canónica.

Entonces se definen los operadores

$$\mathcal{U}_N : \mathcal{H}_{\mathcal{X},N} \rightarrow \mathcal{H}_{\mathcal{X},N}, \quad \mathcal{G}_N : \mathbb{R}^p \rightarrow \mathcal{H}_{\mathcal{X},N}, \quad \mathcal{B}_N : \mathbb{R}^n \rightarrow \mathcal{H}_{\mathcal{X},N}$$

que vienen representación por las matrices

$$\begin{aligned}
U_N &= (\Phi_N(X))^{-1} \Phi_N(X^+)^{\top} \\
G_N &= (\Phi_N(X))^{-1} Y^{\top} \\
B_N &= (\Phi_N(X))^{-1} X^{\top}
\end{aligned}$$

Definiremos además

$$\|k_{\mathcal{X}}\|_1 = \int_{\mathcal{X}} k_{\mathcal{X}}(x, x) d\mu_{\mathcal{X}}(x), \quad \|k_{\mathcal{X}}\|_{\infty} = \sup_{x \in \mathcal{X}} k_{\mathcal{X}}(x, x)$$

que para el caso de Matérn son ambas finitas.

Una vez definido todo lo anterior, utilizaremos el resultado de Philipp et al.¹.

Teorema 2.1 *Sea un $r \in \mathbb{N}$ arbitrario, se supone que los primeros $r + 1$ valores propios λ_j de C_X son simples, es decir, $\lambda_{j+1} < \lambda_j$ para todo $j = 1, \dots, r$. Se define*

$$\delta_r = \min_{j=1, \dots, r} \frac{\lambda_j - \lambda_{j+1}}{2}, \quad c_r = \frac{1}{\sqrt{\lambda_r}} + \frac{r+1}{\delta_r \lambda_r} (1 + \|k_{\mathcal{X}}\|_1) \|k_{\mathcal{X}}\|_1^{1/2}$$

Además, sea $\varepsilon \in (0, \delta_r)$ y $\delta \in (0, 1)$ arbitrarios, y $N \geq \max\{r, \frac{8\|k\|_{\infty}^2 \ln(4/\delta)}{\varepsilon^2}\}$. Si $\mathcal{U}\mathcal{H}_{\mathcal{X}} \subset \mathcal{H}_{\mathcal{X}}$, entonces, con probabilidad al menos $1 - \delta$, se cumple que

$$\|\mathcal{U} - \mathcal{U}_N\|_{\mathcal{H}_{\mathcal{X},N} \rightarrow L^2(\mathcal{X}; \mu_{\mathcal{X}})} \leq \sqrt{\lambda_{r+1}} \|\mathcal{U}\|_{\mathcal{H}_X \rightarrow \mathcal{H}_{\mathcal{X}}} + c_r \varepsilon$$

La misma cota es válida para \mathcal{G}_N y \mathcal{B}_N . El siguiente teorema culmina esta sección y será la síntesis de la cota necesaria para la sección siguiente.

¹ Error Analysis of Kernel EDMD for Prediction and Control in the Koopman Framework

Teorema 2.2 *Si $\mathcal{H}_{\mathcal{X}}$ es equivalente en norma a $H^{\nu+n/2}$, entonces*

$$\|\mathcal{U} - \mathcal{U}_N\|_{\mathcal{H}_{\mathcal{X},N} \rightarrow L^2(\mathcal{X}; \mu_{\mathcal{X}})} \leq CN^{-1/2}$$

Antes de su demostración, un lema para el orden de los valores propios de la matriz de covarianza.

Lema 2.1 ([16]) *Si $\mathcal{H}_{\mathcal{X}}$ es equivalente en norma a un espacio de Sobolev $H^{\nu+n/2}$ y λ_j son los valores propios de C_X en orden descendente, luego*

$$\sqrt{\lambda_{N+1}} \leq c_1 N^{-(\nu+n)/(2n)}$$

con c_1 alguna constante que no depende de N .

DEMOSTRACIÓN. Gracias a la arbitrariedad de ε en el teorema 2.1 se tiene que

$$\|\mathcal{U} - \mathcal{U}_N\|_{\mathcal{H}_{\mathcal{X},N} \rightarrow L^2(\mathcal{X}; \mu_{\mathcal{X}})} \leq c\sqrt{\lambda_{r+1}}$$

y por el lema 2.1 se tiene que existe una constante C tal que

$$\|\mathcal{U} - \mathcal{U}_N\|_{\mathcal{H}_{\mathcal{X},N} \rightarrow L^2(\mathcal{X}; \mu_{\mathcal{X}})} \leq c_1 N^{-(\nu+n)/(2n)}$$

como $\nu > 0$ se concluye que

$$\|\mathcal{U} - \mathcal{U}_N\|_{\mathcal{H}_{\mathcal{X},N} \rightarrow L^2(\mathcal{X}; \mu_{\mathcal{X}})} \leq c_1 N^{-1/2}$$

□

En principio $N^{-1/2}$ es una cota menos ajustada que la mostrada en [13], pero esta corresponde al orden de error de otras aproximaciones que aparecerán en la sección siguiente.

2.1.1. Resultados numéricos

A pesar de que las cotas de error del kernel Extended Dynamic Mode Decomposition serán útiles para la construcción de la cota de error para el filtro, también son interesantes por sí solas y se puede visualizar al aproximar distintos sistemas dinámicos y operadores de manera empírica.

Capítulo 3

Algoritmo de filtraje en tiempo discreto

3.1. Filtro de Kalman a tiempo discreto en RKHS

El objetivo de esta sección es estudiar el error que se produce entre una regla de Kalman dada y otra, que posteriormente se utilizará para cuantificar el error que comete la regla de Kalman aproximante y la regla de Kalman exacta, ambas por introducirse también en la presente sección.

Primero, se supone que existen dos sistemas dinámicos observados en un espacio de Hilbert, con espacio de estados E_x y de observaciones E_y , de la forma

$$\begin{aligned}\mu_{i,k} &= A_{i,k}\mu_{i,k-1} + \nu_{i,k} \\ y_{i,k} &= C_{i,k}\mu_{i,k} + \xi_{i,k}\end{aligned}$$

Donde $A_{i,k} : E_x \rightarrow E_x$, $C_{i,k} : E_x \rightarrow E_y$ se supondrán lineales y $\nu_{i,k} \in E_x$, $\xi_{i,k} \in E_y$ son variables aleatorias con segundo momento finito y operadores de covarianza $\mathcal{Q}_{i,k}$, $\mathcal{R}_{i,k}$, respectivamente, todo esto para $i \in 1, 2$ y $k \geq 1$.

Entonces, se tiene una regla de Kalman asociada a cada uno de los sistemas, que vienen dadas por

$$\begin{aligned}\mathcal{P}_{i,k}^- &= A_{i,k}^* \mathcal{P}_{i,k-1}^+ A_{i,k} + \mathcal{Q}_{i,k} \\ \mathcal{S}_{i,k} &= C_{i,k} \mathcal{P}_{i,k}^- C_{i,k}^* + \mathcal{R}_{i,k} \\ \mathcal{K}_{i,k} &= \mathcal{P}_{i,k}^- C_{i,k}^* \mathcal{S}_{i,k}^{-1} \\ \mathcal{P}_{i,k}^+ &= (I - \mathcal{K}_{i,k} C_{i,k}) \mathcal{P}_{i,k}^- \\ \hat{\mu}_{i,k} &= A_{i,k} \hat{\mu}_{i,k-1} + \mathcal{K}_{i,k} (y_k - C_{i,k} \hat{\mu}_{i,k-1})\end{aligned}$$

con $i \in \{1, 2\}$, $k \geq 1$, $\mathcal{P}_{i,k}^-, \mathcal{P}_{i,k}^+ : E_x \rightarrow E_x$ los operadores de covarianza de error a priori y a posteriori, respectivamente y $\mathcal{K}_{i,k} : E_y \rightarrow E_x$ el operador de ganancia de Kalman. Estas reglas se inicializan como

$$\hat{\mu}_{i,0} = \mathbb{E}[\mu_{i,0}], \quad \mathcal{P}_{i,0} = \text{Cov}(\mu_{i,0})$$

Con esto se tiene el siguiente resultado que ilustra el hecho de que la discrepancia en norma entre las reglas de Kalman se puede descomponer en la discrepancia en norma de cada uno de los elementos asociados y de la iteración anterior.

Teorema 3.1 (Descomposición de error de Kalman) *Sea $k \geq 1$, si los operadores $\mathcal{S}_{i,k}$ son invertibles entonces existen constantes $c_{k,j}^i$ con $j \in \{1, \dots, 7\}$, $i \in \{1, 2\}$ tales que*

$$\begin{aligned} \|\hat{\mu}_{1,k} - \hat{\mu}_{2,k}\| &\leq c_{1,k}^1 \|A_{1,k} - A_{2,k}\| + c_{2,k}^1 \|C_{1,k} - C_{2,k}\| \\ &\quad + c_{3,k}^1 \|\mathcal{Q}_{1,k} - \mathcal{Q}_{2,k}\| + c_{4,k}^1 \|\mathcal{R}_{1,k} - \mathcal{R}_{2,k}\| \\ &\quad + c_{5,k}^1 \|y_{1,k} - y_{2,k}\| + c_{6,k}^1 \|\hat{\mu}_{1,k-1} - \hat{\mu}_{2,k-1}\| + c_{7,k}^1 \|\mathcal{P}_{1,k-1}^+ - \mathcal{P}_{2,k-1}^+\| \\ \|\mathcal{P}_{1,k}^+ - \mathcal{P}_{2,k}^+\| &\leq c_{1,k}^2 \|A_{1,k} - A_{2,k}\| + c_{2,k}^2 \|C_{1,k} - C_{2,k}\| \\ &\quad + c_{3,k}^2 \|\mathcal{Q}_{1,k} - \mathcal{Q}_{2,k}\| + c_{4,k}^2 \|\mathcal{R}_{1,k} - \mathcal{R}_{2,k}\| \\ &\quad + c_{5,k}^2 \|y_{1,k} - y_{2,k}\| + c_{6,k}^2 \|\hat{\mu}_{1,k-1} - \hat{\mu}_{2,k-1}\| + c_{7,k}^2 \|\mathcal{P}_{1,k-1}^+ - \mathcal{P}_{2,k-1}^+\| \end{aligned}$$

En donde las constante $c_{k,j}^i$ son positivas y dependen de k solo a través de $\|A_{i,k}\|$, $\|C_{i,k}\|$, $\|\mathcal{Q}_{i,k}\|$, $\|\mathcal{R}_{i,k}\|$, $\|\mathcal{S}_{i,k}^{-1}\|$, $\|y_{i,k}\|$, $\|\mu_{i,k-1}\|$ y $\|\mathcal{P}_{i,k-1}^+\|$.

DEMOSTRACIÓN. Notar que

$$\begin{aligned} &\|\hat{\mu}_{1,k} - \hat{\mu}_{2,k}\|_{E_x} \\ &\leq \|A_{1,k}\mu_{1,k-1} - A_{2,k}\mu_{2,k-1}\| \\ &\quad + \|\mathcal{K}_{1,k}(y_{1,k} - C_{1,k}\hat{\mu}_{1,k-1}) - \mathcal{K}_{2,k}(y_{2,k} - C_{2,k}\hat{\mu}_{2,k-1})\| \end{aligned}$$

El primero término, que se denotará error de predicción, entrega

$$\begin{aligned} &\|A_{1,k}\mu_{1,k-1} - A_{2,k}\mu_{2,k-1}\| \\ &\leq \|A_{1,k}\mu_{1,k-1} - A_{1,k}\mu_{2,k-1}\| + \|A_{1,k}\mu_{2,k-1} - A_{2,k}\mu_{2,k-1}\| \\ &\leq \|A_{1,k}\| \|\mu_{1,k-1} - \mu_{2,k-1}\| + \|\mu_{2,k-1}\| \|A_{1,k} - A_{2,k}\| \end{aligned}$$

Mientras que el segundo términos, que se denotará error de actualización, entrega

$$\begin{aligned} &\|\mathcal{K}_{1,k}(y_{1,k} - C_{1,k}\hat{\mu}_{1,k-1}) - \mathcal{K}_{2,k}(y_{2,k} - C_{2,k}\hat{\mu}_{2,k-1})\| \\ &\leq \|\mathcal{K}_{1,k}y_{1,k} - \mathcal{K}_{2,k}y_{2,k}\| + \|\mathcal{K}_{1,k}C_{1,k}\hat{\mu}_{1,k-1} - \mathcal{K}_{2,k}C_{2,k}\hat{\mu}_{2,k-1}\| \\ &\leq \|\mathcal{K}_{1,k}y_{1,k} - \mathcal{K}_{1,k}y_{2,k}\| + \|\mathcal{K}_{1,k}y_{2,k} - \mathcal{K}_{2,k}y_{2,k}\| \\ &\quad + \|\mathcal{K}_{1,k}C_{1,k}\hat{\mu}_{1,k-1} - \mathcal{K}_{1,k}C_{2,k}\hat{\mu}_{2,k-1}\| + \|\mathcal{K}_{1,k}C_{2,k}\hat{\mu}_{2,k-1} - \mathcal{K}_{2,k}C_{2,k}\hat{\mu}_{2,k-1}\| \\ &\leq \|\mathcal{K}_{1,k}\| \|y_{1,k} - y_{2,k}\| + \|y_{2,k}\| \|\mathcal{K}_{1,k} - \mathcal{K}_{2,k}\| \\ &\quad + \|\mathcal{K}_{1,k}\| \|C_{1,k}\hat{\mu}_{1,k-1} - C_{2,k}\hat{\mu}_{2,k-1}\| + \|C_{2,k}\hat{\mu}_{2,k-1}\| \|\mathcal{K}_{1,k} - \mathcal{K}_{2,k}\| \\ &\leq \|\mathcal{K}_{1,k}\| \|y_{1,k} - y_{2,k}\| + \|y_{2,k}\| \|\mathcal{K}_{1,k} - \mathcal{K}_{2,k}\| \\ &\quad + \|\mathcal{K}_{1,k}\| (\|C_{1,k}\hat{\mu}_{1,k-1} - C_{1,k}\hat{\mu}_{2,k-1}\| + \|C_{1,k}\hat{\mu}_{2,k-1} - C_{2,k}\hat{\mu}_{2,k-1}\|) \\ &\quad + \|C_{2,k}\hat{\mu}_{2,k-1}\| \|\mathcal{K}_{1,k} - \mathcal{K}_{2,k}\| \\ &\leq \|\mathcal{K}_{1,k}\| \|y_{1,k} - y_{2,k}\| + \|y_{2,k}\| \|\mathcal{K}_{1,k} - \mathcal{K}_{2,k}\| \\ &\quad + \|\mathcal{K}_{1,k}\| (\|C_{1,k}\| \|\hat{\mu}_{1,k-1} - \hat{\mu}_{2,k-1}\| + \|\hat{\mu}_{2,k-1}\| \|C_{1,k} - C_{2,k}\|) \\ &\quad + \|C_{2,k}\hat{\mu}_{2,k-1}\| \|\mathcal{K}_{1,k} - \mathcal{K}_{2,k}\| \end{aligned}$$

En virtud de lo anterior, se debe ver la diferencia en norma de los operadores de ganancia

$$\begin{aligned}
& \|\mathcal{K}_{1,k} - \mathcal{K}_{2,k}\| \\
& \leq \|\mathcal{P}_{1,k}^- C_{1,k} \mathcal{S}_{1,k}^{-1} - \mathcal{P}_{2,k}^- C_{2,k} \mathcal{S}_{2,k}^{-1}\| \\
& \leq \|\mathcal{P}_{1,k}^- C_{1,k} \mathcal{S}_{1,k}^{-1} - \mathcal{P}_{2,k}^- C_{1,k} \mathcal{S}_{1,k}^{-1}\| + \|\mathcal{P}_{2,k}^- C_{1,k} \mathcal{S}_{1,k}^{-1} - \mathcal{P}_{2,k}^- C_{2,k} \mathcal{S}_{2,k}^{-1}\| \\
& \leq \|C_{1,k} \mathcal{S}_{1,k}^{-1}\| \|\mathcal{P}_{1,k}^- - \mathcal{P}_{2,k}^-\| + \|\mathcal{P}_{2,k}^-\| \|C_{1,k} \mathcal{S}_{1,k}^{-1} - C_{2,k} \mathcal{S}_{2,k}^{-1}\| \\
& \leq \|C_{1,k} \mathcal{S}_{1,k}^{-1}\| \|\mathcal{P}_{1,k}^- - \mathcal{P}_{2,k}^-\| \\
& \quad + \|\mathcal{P}_{2,k}^-\| \|C_{1,k} \mathcal{S}_{1,k}^{-1} - C_{2,k} \mathcal{S}_{2,k}^{-1}\| \\
& \leq \|C_{1,k} \mathcal{S}_{1,k}^{-1}\| \|\mathcal{P}_{1,k}^- - \mathcal{P}_{2,k}^-\| \\
& \quad + \|\mathcal{P}_{2,k}^-\| (\|C_{1,k} \mathcal{S}_{1,k}^{-1} - C_{1,k} \mathcal{S}_{2,k}^{-1}\| + \|C_{1,k} \mathcal{S}_{2,k}^{-1} - C_{2,k} \mathcal{S}_{2,k}^{-1}\|) \\
& \quad + \|\mathcal{P}_{2,k}^-\| (\|C_{1,k}\| \|\mathcal{S}_{1,k}^{-1} - \mathcal{S}_{2,k}^{-1}\| + \|\mathcal{S}_{2,k}^{-1}\| \|C_{1,k} - C_{2,k}\|)
\end{aligned}$$

En donde

$$\|\mathcal{P}_{i,k}^-\| \leq \|A_{i,k}\|^2 \|\mathcal{P}_{i,k-1}^+\| + \|\mathcal{Q}_{i,k}\|, \quad i \in \{1, 2\}$$

Primero para la diferencia en norma de los operadores de covarianza de error a priori se tiene

$$\begin{aligned}
& \|\mathcal{P}_{1,k}^- - \mathcal{P}_{2,k}^-\| \\
& = \|A_{1,k}^* \mathcal{P}_{1,k-1}^+ A_{1,k} + \mathcal{Q}_{1,k} - A_{2,k}^* \mathcal{P}_{2,k-1}^+ A_{2,k} + \mathcal{Q}_{2,k}\| \\
& \leq \|A_{1,k}^* \mathcal{P}_{1,k-1}^+ A_{1,k} - A_{2,k}^* \mathcal{P}_{2,k-1}^+ A_{2,k}\| + \|\mathcal{Q}_{1,k} - \mathcal{Q}_{2,k}\| \\
& \leq \|A_{1,k}^* \mathcal{P}_{1,k-1}^+ A_{1,k} - A_{1,k}^* \mathcal{P}_{2,k-1}^+ A_{2,k}\| + \|A_{1,k}^* \mathcal{P}_{2,k-1}^+ A_{2,k} - A_{2,k}^* \mathcal{P}_{2,k-1}^+ A_{2,k}\| \\
& \quad + \|\mathcal{Q}_{1,k} - \mathcal{Q}_{2,k}\| \\
& \leq \|A_{1,k}^*\| \|\mathcal{P}_{1,k-1}^+ A_{1,k} - \mathcal{P}_{2,k-1}^+ A_{2,k}\| + \|\mathcal{P}_{2,k-1}^+ A_{2,k}\| \|A_{1,k}^* - A_{2,k}^*\| \\
& \quad + \|\mathcal{Q}_{1,k} - \mathcal{Q}_{2,k}\| \\
& \leq \|A_{1,k}\| \|\mathcal{P}_{1,k-1}^+ A_{1,k} - \mathcal{P}_{2,k-1}^+ A_{2,k}\| + \|\mathcal{P}_{2,k-1}^+ A_{2,k}\| \|A_{1,k} - A_{2,k}\| \\
& \quad + \|\mathcal{Q}_{1,k} - \mathcal{Q}_{2,k}\| \\
& \leq \|A_{1,k}\| (\|\mathcal{P}_{1,k-1}^+ A_{1,k} - \mathcal{P}_{1,k-1}^+ A_{2,k}\| + \|\mathcal{P}_{1,k-1}^+ A_{2,k} - \mathcal{P}_{2,k-1}^+ A_{2,k}\|) \\
& \quad + \|\mathcal{P}_{2,k-1}^+ A_{2,k}\| \|A_{1,k} - A_{2,k}\| \\
& \quad + \|\mathcal{Q}_{1,k} - \mathcal{Q}_{2,k}\| \\
& \leq \|A_{1,k}\| (\|\mathcal{P}_{1,k-1}^+\| \|A_{1,k} - A_{2,k}\| + \|A_{2,k}\| \|\mathcal{P}_{1,k-1}^+ - \mathcal{P}_{2,k-1}^+\|) \\
& \quad + \|\mathcal{P}_{2,k-1}^+\| \|A_{2,k}\| \|A_{1,k} - A_{2,k}\| \\
& \quad + \|\mathcal{Q}_{1,k} - \mathcal{Q}_{2,k}\|
\end{aligned}$$

Por último queda ver el término $\|\mathcal{S}_{1,k}^{-1} - \mathcal{S}_{2,k}^{-1}\|$, para ello notar que

$$\mathcal{S}_{1,k}^{-1} - \mathcal{S}_{2,k}^{-1} = \mathcal{S}_{2,k}^{-1} (\mathcal{S}_{2,k} - \mathcal{S}_{1,k}) \mathcal{S}_{1,k}^{-1}$$

Con lo que se tiene

$$\begin{aligned}
& \|\mathcal{S}_{1,k}^{-1} - \mathcal{S}_{2,k}^{-1}\| \\
& \leq \|\mathcal{S}_{2,k}^{-1}(\mathcal{S}_{2,k} - \mathcal{S}_{1,k})\mathcal{S}_{1,k}^{-1}\| \\
& \leq \|\mathcal{S}_{1,k}^{-1}\| \|\mathcal{S}_{2,k}^{-1}\| \|\mathcal{S}_{2,k} - \mathcal{S}_{1,k}\| \\
& \leq \|\mathcal{S}_{1,k}^{-1}\| \|\mathcal{S}_{2,k}^{-1}\| \|C_{1,k}\mathcal{P}_{1,k}^- C_{1,k}^* + \mathcal{R}_{1,k} - C_{2,k}\mathcal{P}_{2,k}^- C_{2,k}^* + \mathcal{R}_{2,k}\|
\end{aligned}$$

En donde, análogo a lo anterior, se obtiene

$$\begin{aligned}
& \|C_{1,k}\mathcal{P}_{1,k}^- C_{1,k}^* + \mathcal{R}_{1,k} - C_{2,k}\mathcal{P}_{2,k}^- C_{2,k}^* + \mathcal{R}_{2,k}\| \\
& \leq \|C_{1,k-1}\| (\|\mathcal{P}_{1,k-1}^+\| \|C_{1,k} - C_{2,k}\| + \|C_{2,k}\| \|\mathcal{P}_{1,k-1}^+ - \mathcal{P}_{2,k-1}^+\|) \\
& \quad + \|\mathcal{P}_{2,k-1}^+\| \|C_{2,k}\| \|C_{1,k} - C_{2,k}\| \\
& \quad + \|\mathcal{R}_{1,k} - \mathcal{R}_{2,k}\|
\end{aligned}$$

Concluyendo que

$$\begin{aligned}
\|\hat{\mu}_{1,k} - \hat{\mu}_{2,k}\| & \leq c_{1,k}^1 \|A_{1,k} - A_{2,k}\| + c_{2,k}^1 \|C_{1,k} - C_{2,k}\| \\
& \quad + c_{3,k}^1 \|\mathcal{Q}_{1,k} - \mathcal{Q}_{2,k}\| + c_{4,k}^1 \|\mathcal{R}_{1,k} - \mathcal{R}_{2,k}\| \\
& \quad + c_{5,k}^1 \|y_{1,k} - y_{2,k}\| + c_{6,k}^1 \|\hat{\mu}_{1,k-1} - \hat{\mu}_{2,k-1}\| + c_{7,k}^1 \|\mathcal{P}_{1,k-1}^+ - \mathcal{P}_{2,k-1}^+\|
\end{aligned}$$

Para la diferencia de los errores de covarianza a posteriori basta notar que

$$\begin{aligned}
& \|\mathcal{P}_{1,k}^+ - \mathcal{P}_{2,k}^+\| \\
& \leq \|(I - \mathcal{K}_{1,k}C_{1,k})\mathcal{P}_{1,k}^- - (I - \mathcal{K}_{2,k}C_{2,k})\mathcal{P}_{2,k}^-\| \\
& = \|\mathcal{K}_{1,k}C_{1,k}\mathcal{P}_{1,k}^- - \mathcal{K}_{2,k}C_{2,k}\mathcal{P}_{2,k}^-\| \\
& \leq \|\mathcal{K}_{1,k}C_{1,k}\mathcal{P}_{1,k}^- - \mathcal{K}_{1,k}C_{2,k}\mathcal{P}_{2,k}^-\| + \|\mathcal{K}_{1,k}C_{2,k}\mathcal{P}_{2,k}^- - \mathcal{K}_{2,k}C_{2,k}\mathcal{P}_{2,k}^-\| \\
& \leq \|\mathcal{K}_{1,k}\| \|C_{1,k}\mathcal{P}_{1,k}^- - C_{2,k}\mathcal{P}_{2,k}^-\| + \|C_{2,k}\mathcal{P}_{2,k}^-\| \|\mathcal{K}_{1,k} - \mathcal{K}_{2,k}\| \\
& \leq \|\mathcal{K}_{1,k}\| (\|C_{1,k}\mathcal{P}_{1,k}^- - C_{1,k}\mathcal{P}_{2,k}^-\| + \|C_{1,k}\mathcal{P}_{2,k}^- - C_{2,k}\mathcal{P}_{2,k}^-\|) + \|C_{2,k}\mathcal{P}_{2,k}^-\| \|\mathcal{K}_{1,k} - \mathcal{K}_{2,k}\| \\
& \leq \|\mathcal{K}_{1,k}\| (\|C_{1,k}\| \|\mathcal{P}_{1,k}^- - \mathcal{P}_{2,k}^-\| + \|\mathcal{P}_{2,k}^-\| \|C_{1,k} - C_{2,k}\|) + \|C_{2,k}\mathcal{P}_{2,k}^-\| \|\mathcal{K}_{1,k} - \mathcal{K}_{2,k}\|
\end{aligned}$$

Observando que todos los términos ya se analizaron antes, se concluye que

$$\begin{aligned}
\|\mathcal{P}_{1,k}^+ - \mathcal{P}_{2,k}^+\| & \leq c_{1,k}^2 \|A_{1,k} - A_{2,k}\| + c_{2,k}^2 \|C_{1,k} - C_{2,k}\| \\
& \quad + c_{3,k}^2 \|\mathcal{Q}_{1,k} - \mathcal{Q}_{2,k}\| + c_{4,k}^2 \|\mathcal{R}_{1,k} - \mathcal{R}_{2,k}\| \\
& \quad + c_{5,k}^2 \|y_{1,k} - y_{2,k}\| + c_{6,k}^2 \|\hat{\mu}_{1,k-1} - \hat{\mu}_{2,k-1}\| + c_{7,k}^2 \|\mathcal{P}_{1,k-1}^+ - \mathcal{P}_{2,k-1}^+\|
\end{aligned}$$

□

Con esto se concluye que para una iteración k , el error depende del error en la condición tanto para la estimación del estado como para el operador de covarianza de error a posteriori. Notemos que

$$\begin{aligned}
\Phi_{\mathcal{X}}(\mathbf{x}_{k+1}) & = \Phi_{\mathcal{X}}(\mathbf{f}(\mathbf{x}_k, \mathbf{w}_k)) \\
& = \mathbb{E}[\Phi_{\mathcal{X}}(\mathbf{f}(\mathbf{x}_k, \cdot))] + \Phi_{\mathcal{X}}(\mathbf{f}(\mathbf{x}_k, \mathbf{w}_k)) - \mathbb{E}[\Phi_{\mathcal{X}}(\mathbf{f}(\mathbf{x}_k, \cdot))] \\
& = (\mathcal{U}\Phi_{\mathcal{X}})(\mathbf{x}_k) + \zeta_k
\end{aligned}$$

donde ζ_k es una variable aleatoria infinito dimensional, centrada y con operador de covarianza acotado, digamos \mathcal{Q}_k . Análogamente

$$\Phi_{\mathcal{Y}}(\mathbf{y}_k) = (\mathcal{G}\Phi_{\mathcal{Y}})(\mathbf{x}_k) + \nu_k$$

donde ν_k es una variable aleatoria infinito dimensional, centrada y con operador de covarianza acotado, digamos \mathcal{R}_k .

Haciendo algo similar a [17], se denota

$$\hat{\mu}_k = \mathbb{E}[\Phi_{\mathcal{X}}(\mathbf{x}_k)|\mathbf{y}_{1:k}], \quad \mathcal{P}_k = \text{Cov}(\Phi_{\mathcal{X}}(\mathbf{x}_k) - \hat{\mu}_k)$$

con

$$\hat{\mu}_0 = \mathbb{E}[\Phi_{\mathcal{X}}(\mathbf{x}_0)], \quad \mathcal{P}_0 = \text{Cov}(\Phi_{\mathcal{X}}(\mathbf{x}_0) - \hat{\mu}_0).$$

Y se define

$$\hat{\mu}_{k+1}^- = \mathbb{E}[\Phi_{\mathcal{X}}(\mathbf{x}_{k+1})|\mathbf{y}_{1:k}], \quad \mathcal{P}_{k+1}^- = \text{Cov}(\Phi_{\mathcal{X}}(\mathbf{x}_{k+1}) - \hat{\mu}_{k+1}^-)$$

Que por la Kernel Bayes Rule [15] cumple

$$\hat{\mu}_{k+1}^- = \mathbb{E}[\Phi_{\mathcal{X}}(\mathbf{x}_{k+1})|\mathbf{y}_{1:k}] = C_{X+|X}\mathbb{E}[\Phi_{\mathcal{X}}(\mathbf{x}_k)|\mathbf{y}_{1:k}] = C_{X+|X}\hat{\mu}_k.$$

Con ello, utilizando la independiencia de ζ_k con $\Phi_{\mathcal{X}}(\mathbf{x}_k)$

$$\begin{aligned} \mathcal{P}_{k+1}^- &= \text{Cov}(\Phi_{\mathcal{X}}(\mathbf{x}_{k+1}) - \hat{\mu}_{k+1}^-) \\ &= \text{Cov}(C_{X+|X}\Phi_{\mathcal{X}}(\mathbf{x}_k) + \zeta_{k+1} - C_{X+|X}\hat{\mu}_k) \\ &= C_{X+|X}\text{Cov}(\Phi_{\mathcal{X}}(\mathbf{x}_k) - \hat{\mu}_k)C_{X+|X}^* + \text{Cov}(\zeta_{k+1}) \\ &= C_{X+|X}\mathcal{P}_k(C_{X+|X})^* + \mathcal{Q}_{k+1} \end{aligned}$$

Ahora, debemos proyectar sobre las observaciones para obtener la estimación a posteriori, es decir

$$\hat{\mu}_{k+1}^- = \mathbb{E}[\Phi_{\mathcal{X}}(\mathbf{x}_{k+1})|\mathbf{y}_{1:k}] \rightarrow \hat{\mu}_{k+1} = \mathbb{E}[\Phi_{\mathcal{X}}(\mathbf{x}_{k+1})|\mathbf{y}_{1:k+1}]$$

Proponiendo que

$$\hat{\mu}_{k+1} = \hat{\mu}_{k+1}^- + \mathcal{K}_{k+1}(\mathbf{y}_{k+1} - C_{Y|X}\hat{\mu}_{k+1}^-)$$

Es decir, actualizar igual que en Kalman lineal, en donde ahora $\mathcal{K}_k : \mathbb{R}^p \rightarrow \mathcal{H}_{\mathcal{X}}$ es el operador de ganancia de Kalman, que cumple

$$\mathcal{K}_{k+1} = \mathcal{P}_{k+1}^-(C_{Y|X})^*(C_{Y|X}\mathcal{P}_{k+1}^-(C_{Y|X})^* + \mathcal{R}_{k+1})^{-1}$$

y entonces el operador de covarianza de error a posteriori es

$$\mathcal{P}_{k+1} = \text{Cov}(\Phi_{\mathcal{X}}(\mathbf{x}_{k+1}) - \hat{\mu}_{k+1})$$

Desarrollando este término, y utilizando independencia

$$\begin{aligned}
\mathcal{P}_{k+1} &= \text{Cov}(\Phi_{\mathcal{X}}(\mathbf{x}_{k+1}) - \hat{\mu}_{k+1}) \\
&= \text{Cov}(\Phi_{\mathcal{X}}(\mathbf{x}_{k+1}) - \hat{\mu}_{k+1}^- - \mathcal{K}_k(\mathbf{y}_{k+1} - C_{Y|X}\hat{\mu}_{k+1}^-)) \\
&= \text{Cov}(\Phi_{\mathcal{X}}(\mathbf{x}_{k+1}) - \hat{\mu}_{k+1}^-) + \text{Cov}(\mathcal{K}_{k+1}(\mathbf{y}_{k+1} - C_{Y|X}\hat{\mu}_{k+1}^-)) \\
&= \mathcal{P}_{k+1}^- + \mathcal{K}_{k+1} \text{Cov}(\mathbf{y}_{k+1} - C_{Y|X}\hat{\mu}_{k+1}^-)(\mathcal{K}_{k+1})^* \\
&= \mathcal{P}_{k+1}^- + \mathcal{K}_{k+1}(C_{Y|X}\Phi_{\mathcal{X}}(\mathbf{x}_{k+1}) + \nu_k - C_{Y|X}\hat{\mu}_{k+1}^-)(\mathcal{K}_{k+1})^* \\
&= \mathcal{P}_{k+1}^- + \mathcal{K}_{k+1}(C_{Y|X}\mathcal{P}_{k+1}^-(C_{Y|X})^* + \text{Cov}(\nu_{k+1}))(\mathcal{K}_{k+1})^* \\
&= \mathcal{P}_{k+1}^- + \mathcal{K}_{k+1}(C_{Y|X}\mathcal{P}_{k+1}^-(C_{Y|X})^* + \mathcal{R}_{k+1})(\mathcal{K}_{k+1})^* \\
&= \mathcal{P}_{k+1}^- + \mathcal{K}_{k+1}C_{Y|X}\mathcal{P}_{k+1}^- \\
&= (I + \mathcal{K}_{k+1}C_{Y|X})\mathcal{P}_{k+1}^-
\end{aligned}$$

Donde se utilizó que

$$(\mathcal{K}_{k+1})^* = (C_{Y|X}\mathcal{P}_{k+1}^-(C_{Y|X})^* + \mathcal{R}_{k+1})^{-1}C_{Y|X}(\mathcal{P}_{k+1}^-)^*$$

y que $C_{Y|X}\mathcal{P}_{k+1}^-(C_{Y|X})^*$, \mathcal{R}_{k+1} son simétricos y \mathcal{P}_{k+1} es autoadjunto. Entonces, las ecuaciones para cada iteración vienen dadas por

$$\begin{aligned}
\hat{\mu}_{k+1}^- &= C_{X+|X}\hat{\mu}_k \\
\mathcal{P}_{k+1}^- &= C_{X+|X}\mathcal{P}_k(C_{X+|X})^* + \mathcal{Q}_{k+1} \\
\mathcal{S}_{k+1} &= C_{Y|X}\mathcal{P}_{k+1}^-(C_{Y|X})^* + \mathcal{R}_{k+1} \\
\mathcal{K}_{k+1} &= \mathcal{P}_{k+1}^-(C_{Y|X})^*\mathcal{S}_{k+1}^{-1} \\
\mathcal{P}_{k+1} &= (I + \mathcal{K}_{k+1}C_{Y|X})\mathcal{P}_{k+1}^- \\
\hat{\mu}_{k+1} &= C_{X+|X}\hat{\mu}_k + \mathcal{K}_{k+1}(\mathbf{y}_{k+1} - C_{Y|X}\hat{\mu}_{k+1}^-)
\end{aligned}$$

Con

$$\hat{\mu}_0 = \mathbb{E}[\Phi_{\mathcal{X}}(\mathbf{x}_0)], \quad \mathcal{P}_0 = \text{Cov}(\Phi_{\mathcal{X}}(\mathbf{x}_0) - \hat{\mu}_0)$$

Ahora dejando todo en función del operador de Koopman, gracias a que

$$C_{X+|X} = \mathcal{U}^*, \quad C_{Y|X} = \mathcal{G}^*$$

Con lo que queda

$$\begin{aligned}
\hat{\mu}_{k+1}^- &= \mathcal{U}^*\hat{\mu}_k \\
\mathcal{P}_{k+1}^- &= \mathcal{U}^*\mathcal{P}_k\mathcal{U} + \mathcal{Q}_{k+1} \\
\mathcal{S}_{k+1} &= \mathcal{G}^*\mathcal{P}_{k+1}^-\mathcal{G} + \mathcal{R}_{k+1} \\
\mathcal{K}_{k+1} &= \mathcal{P}_{k+1}^-\mathcal{G}\mathcal{S}_{k+1}^{-1} \\
\mathcal{P}_{k+1} &= (I + \mathcal{K}_{k+1}\mathcal{G}^*)\mathcal{P}_{k+1}^- \\
\hat{\mu}_{k+1} &= \mathcal{U}^*\hat{\mu}_k + \mathcal{K}_{k+1}(\mathbf{y}_{k+1} - \mathcal{G}^*\hat{\mu}_{k+1}^-)
\end{aligned}$$

Si hacemos las aproximaciones finito dimensionales, queda

$$\begin{aligned}
\hat{\mu}_{N,k+1}^- &= \mathcal{U}_N^* \hat{\mu}_{N,k} \\
\mathcal{P}_{N,k+1}^- &= \mathcal{U}_N^* \mathcal{P}_{N,k} \mathcal{U}_N + \mathcal{Q}_{N,k+1} \\
\mathcal{K}_{N,k+1} &= \mathcal{P}_{N,k+1}^- \mathcal{G}_N (\mathcal{G}_N^* \mathcal{P}_{N,k+1}^- \mathcal{G}_N + \mathcal{R}_{N,k+1})^{-1} \\
\mathcal{P}_{N,k+1} &= (I + \mathcal{K}_{N,k+1} \mathcal{G}_N^*) \mathcal{P}_{N,k+1}^- \\
\hat{\mu}_{N,k+1} &= \mathcal{U}^* \hat{\mu}_{N,k} + \mathcal{K}_{N,k+1} (\mathbf{y}_{k+1} - \mathcal{G}_N^* \hat{\mu}_{N,k+1}^-)
\end{aligned}$$

En donde $\mathcal{Q}_{N,k+1}$, $\mathcal{R}_{N,k+1}$ son los estimadores insesgados de \mathcal{Q}_{k+1} y \mathcal{R}_{k+1} , respectivamente, esto es

$$\mathcal{Q}_{N,k+1} = \frac{1}{N-1} \sum_{j=1}^N (z_{1,j} - \bar{z}_1)^2, \quad \mathcal{R}_{N,k+1} = \frac{1}{N-1} \sum_{j=1}^N (z_{2,j} - \bar{z}_2)^2 \quad (3.1)$$

en donde $\{z_{1,j}\}_{j=1}^N \sim \zeta_k^N$, $\{z_{2,j}\}_{j=1}^N \sim \nu_k^N$ y

$$\bar{z}_i = \frac{1}{N} \sum_{j=1}^N z_{i,j}$$

Si X_0 es la distribución dada para la condición inicial y $\{x_j\}_{j=1}^N \sim X_0$, entonces la inicialización viene dada por

$$\hat{\mu}_{N,0} = \frac{1}{N} \sum_{j=1}^N \Phi_{\mathcal{X}}(x_k), \quad \mathcal{P}_{N,0} = \frac{1}{N-1} \sum_{j=1}^N (\Phi_{\mathcal{X}}(x_k) - \hat{\mu}_{N,0})^2 \quad (3.2)$$

Aplicando el teorema anterior, es directo lo siguiente

Proposición 3.1 *Para $k \geq 1$, existen constantes $c_{k,j}^i$ con $j \in \{1, \dots, 6\}$, $i \in \{1, 2\}$ tales que*

$$\begin{aligned}
\|\hat{\mu}_k - \hat{\mu}_{N,k}\| &\leq c_{1,k}^1 \|\mathcal{U} - \mathcal{U}_N\| + c_{2,k}^1 \|\mathcal{G} - \mathcal{G}_N\| \\
&\quad + c_{3,k}^1 \|\mathcal{Q}_k - \mathcal{Q}_{N,k}\| + c_{4,k}^1 \|\mathcal{R}_k - \mathcal{R}_{N,k}\| \\
&\quad + c_{5,k}^1 \|\hat{\mu}_{k-1} - \hat{\mu}_{N,k-1}\| + c_{6,k}^1 \|\mathcal{P}_{k-1} - \mathcal{P}_{N,k-1}\| \\
\|\mathcal{P}_k - \mathcal{P}_{N,k}\| &\leq c_{1,k}^2 \|\mathcal{U} - \mathcal{U}_N\| + c_{2,k}^2 \|\mathcal{G} - \mathcal{G}_N\| \\
&\quad + c_{3,k}^2 \|\mathcal{Q}_k - \mathcal{Q}_{N,k}\| + c_{4,k}^2 \|\mathcal{R}_k - \mathcal{R}_{N,k}\| \\
&\quad + c_{5,k}^2 \|\hat{\mu}_{k-1} - \hat{\mu}_{N,k-1}\| + c_{6,k}^2 \|\mathcal{P}_{k-1} - \mathcal{P}_{N,k-1}\|
\end{aligned}$$

En donde las constantes $c_{k,j}^i$ son positivas y dependen de k solo a través de $\|\mathcal{U}\|$, $\|\mathcal{G}\|$, $\|\mathcal{Q}_k\|$, $\|\mathcal{R}_k\|$, $\|\mathcal{S}_k^{-1}\|$, $\|\hat{\mu}_{k-1}\|$ y $\|\mathcal{P}_{k-1}\|$.

Proposición 3.2 *Existen constantes C_i^j para $i \in \{1, 2, 3\}$, $j \in \{1, 2\}$ tales que*

$$\begin{aligned}
\|\hat{\mu}_k - \hat{\mu}_{N,k}\| &\leq C_k^1 N^{-1/2} \\
\|\mathcal{P}_k - \mathcal{P}_{N,k}\| &\leq C_k^2 N^{-1/2}
\end{aligned}$$

En donde las constantes C_i^j son positivas y dependen de k solo a través de $\|\mathcal{U}\|$, $\|\mathcal{G}\|$, $\|\mathcal{Q}_j\|$, $\|\mathcal{R}_j\|$, $\|\mathcal{S}_k^{-1}\|$, $\|\hat{\mu}_j\|$ y $\|\mathcal{P}_j\|$, con $j \in \{0, \dots, k-1\}$.

Ahora un lema que permite dar cotas para los elementos y operadores cuya norma se puede acotar por algo de orden $N^{-1/2}$, que son resultados conocidos en la literatura.

Lema 3.1 ([18]) *Sea $N \in \mathbb{N}$ y $\mathcal{Q}_{N,0}, \mathcal{R}_{N,0}, \mu_{N,0}, \mathcal{P}_{N,0}$, definidos en 3.1 y 3.2, respectivamente, luego existe una constante C tal que*

$$\|\hat{\mu}_0 - \hat{\mu}_{N,0}\|_{\mathcal{H}_X}, \|\mathcal{P}_0 - \mathcal{P}_{N,0}\|_{HS}, \|\mathcal{Q}_k - \mathcal{Q}_{N,k}\|_{HS}, \|\mathcal{R}_k - \mathcal{R}_{N,k}\|_{HS} \leq C \cdot N^{-1/2}$$

que, sin pérdida de generalidad, se puede tomar común para todas las cotas.

Ahora se procede con la demotración de la Proposición 3.2.

DEMOSTRACIÓN. Gracias a la proposición 3.1 y el lema 3.1 se obtiene que existen constantes $c_{k,j}^i$ con $j \in \{1, \dots, 6\}$, $i \in \{1, 2\}$ tales que

$$\begin{aligned} \|\hat{\mu}_k - \hat{\mu}_{N,k}\| &\leq c_{1,k}^1 \|\mathcal{U} - \mathcal{U}_N\| + c_{2,k}^1 \|\mathcal{G} - \mathcal{G}_N\| \\ &\quad + c_{3,k}^1 C N^{-1/2} + c_{4,k}^1 C N^{-1/2} \\ &\quad + c_{5,k}^1 \|\hat{\mu}_{k-1} - \hat{\mu}_{N,k-1}\| + c_{6,k}^1 \|\mathcal{P}_{k-1} - \mathcal{P}_{N,k-1}\| \\ \|\mathcal{P}_k - \mathcal{P}_{N,k}\| &\leq c_{1,k}^2 \|\mathcal{U} - \mathcal{U}_N\| + c_{2,k}^2 \|\mathcal{G} - \mathcal{G}_N\| \\ &\quad + c_{3,k}^2 C N^{-1/2} + c_{4,k}^2 C N^{-1/2} \\ &\quad + c_{5,k}^2 \|\hat{\mu}_{k-1} - \hat{\mu}_{N,k-1}\| + c_{6,k}^2 \|\mathcal{P}_{k-1} - \mathcal{P}_{N,k-1}\| \end{aligned}$$

Por teorema 2.2 entonces se concluye que

$$\begin{aligned} \|\hat{\mu}_k - \hat{\mu}_{N,k}\| &\leq C_{1,k}^1 N^{-1/2} + C_{2,k}^1 \|\hat{\mu}_{k-1} - \hat{\mu}_{N,k-1}\| + C_{3,k}^1 \|\mathcal{P}_{k-1} - \mathcal{P}_{N,k-1}\| \\ \|\mathcal{P}_k - \mathcal{P}_{N,k}\| &\leq C_{1,k}^2 N^{-1/2} + C_{2,k}^2 \|\hat{\mu}_{k-1} - \hat{\mu}_{N,k-1}\| + C_{3,k}^2 \|\mathcal{P}_{k-1} - \mathcal{P}_{N,k-1}\| \end{aligned}$$

Para propagar el error hasta la condición inicial se procede por inducción. Para ello primero el caso base $k = 1$ que se tiene directo por el teorema 3.1 aplicado a $k = 1$.

Se supone entonces que para $k \in \mathbb{N}$ se cumple

$$\begin{aligned} \|\hat{\mu}_k - \hat{\mu}_{N,k}\| &\leq C_{1,k}^1 N^{-1/2} + C_{2,k}^1 \|\hat{\mu}_0 - \hat{\mu}_{N,0}\| + C_{3,k}^1 \|\mathcal{P}_0 - \mathcal{P}_{N,0}\| \\ \|\mathcal{P}_k - \mathcal{P}_{N,k}\| &\leq C_{1,k}^2 N^{-1/2} + C_{2,k}^2 \|\hat{\mu}_0 - \hat{\mu}_{N,0}\| + C_{3,k}^2 \|\mathcal{P}_0 - \mathcal{P}_{N,0}\| \end{aligned}$$

Ahora se prueba para $k+1$, que basta hacerlo para $\|\hat{\mu}_{k+1} - \hat{\mu}_{N,k+1}\|$, para la otra cota análoga.

$$\|\hat{\mu}_{k+1} - \hat{\mu}_{N,k+1}\| \leq C_{1,k+1}^1 N^{-1/2} + c_{5,k+1}^1 \|\hat{\mu}_{1,k} - \hat{\mu}_{2,k}\| + c_{6,k+1}^1 \|\mathcal{P}_k - \mathcal{P}_{N,k}\|$$

Ocupando la hipótesis inductiva

$$\begin{aligned} &\leq C_{1,k}^1 N^{-1/2} \\ &\quad + c_{5,k+1}^1 (C_{1,k}^1 N^{-1/2} + C_{2,k}^1 \|\hat{\mu}_0 - \hat{\mu}_{N,0}\| + C_{3,k}^1 \|\mathcal{P}_0 - \mathcal{P}_{N,0}\|) \\ &\quad + c_{6,k+1}^1 (C_{1,k}^2 N^{-1/2} + C_{2,k}^2 \|\hat{\mu}_0 - \hat{\mu}_{N,0}\| + C_{3,k}^2 \|\mathcal{P}_0 - \mathcal{P}_{N,0}\|) \\ &= C_{1,k+1}^1 N^{-1/2} + C_{2,k+1}^1 \|\hat{\mu}_0 - \hat{\mu}_{N,0}\| + C_{3,k+1}^1 \|\mathcal{P}_0 - \mathcal{P}_{N,0}\| \end{aligned}$$

Usando el lema 3.1 se obtiene que existe una constante C_{k+1}^1 tal que

$$\|\hat{\mu}_k - \hat{\mu}_{N,k}\| \leq C_{k+1}^1 N^{-1/2}$$

□

Capítulo 4

Algoritmo de filtraje en tiempo continuo

Capítulo 5

Algoritmo de suavizado en tiempo continuo y tiempo discreto

Capítulo 6

Algoritmo de estimación de parámetros

6.1. Problema de optimización estocástica bi-nivel

Se formula el siguiente problema de optimización estocástica bi-nivel ²

$$(P) \quad \left\{ \begin{array}{ll} \text{mín} & \sum_{k=0}^N \mathbb{E} \left[||\mathbf{y}_k - \mathbf{g}(t_k, \hat{\mathbf{x}}_k, \mathbf{u}_k, \mathbf{w}_k)||^2 \right] \\ \text{s.a} & \hat{\mathbf{x}}_k \in \text{argmin} \left(\sum_{k=0}^N \mathbb{E} \left[||\hat{\mathbf{x}}_{k|k} - \mathbf{x}_k||^2 | \mathbf{y}_{0:k}, \mathbf{u}_{0:k} \right] \right) \\ & \text{s.a} \quad \mathbf{x}_{k+1} = \mathbf{f}(t_k, \mathbf{x}_k, \mathbf{u}_k, \mathbf{w}_k) \\ & \quad \mathbf{y}_k = \mathbf{g}(t_k, \mathbf{x}_k, \mathbf{u}_k, \mathbf{v}_k) \\ & \quad \mathbf{x}_0 \sim X_0, \quad \hat{\mathbf{x}}_0 = \mathbb{E}[\mathbf{x}_0] \end{array} \right.$$

6.2. Observabilidad

6.3. Algoritmo propuesto

6.4. Resultado de convergencia

² An Introduction to Bi-level Optimization: Foundations and Applications in Signal Processing and Machine Learning

Capítulo 7

Modelos epidemiológicos

Bibliografía

- [1] Kalman, R. E., “A new approach to linear filtering and prediction problems”, Transactions of the ASME–Journal of Basic Engineering, vol. 82, pp. 35–45, 1960.
- [2] Setoodeh, P., Habibi, S., y Haykin, S., Nonlinear filters: Theory and applications. Wiley, 2022.
- [3] Kemp, F., “An introduction to sequential monte carlo methods”, Journal of the Royal Statistical Society: Series D (The Statistician), vol. 52, 2003, doi:10.1046/j.1467-9884.2003.t01-6-00383_8.x.
- [4] Wills, A. G. y Schön, T. B., “Sequential monte carlo: A unified review”, 2023, doi:10.1146/annurev-control-042920-015119.
- [5] Crisan, D. y Doucet, A., “A survey of convergence results on particle filtering methods for practitioners”, IEEE Transactions on Signal Processing, vol. 50, 2002, doi:10.1109/78.984773.
- [6] Chopin, N. y Papaspiliopoulos, O., “An introduction to sequential monte carlo”, Springer Series in Statistics, 2020.
- [7] Berlinet, A. y Thomas-Agnan, C., Reproducing Kernel Hilbert Spaces in Probability and Statistics. Springer New York, NY, 1 ed., 2004, doi:10.1007/978-1-4419-9096-9. Springer Science+Business Media Dordrecht 2004.
- [8] Saitoh, S. y Sawano, Y., Theory of Reproducing Kernels and Applications, vol. Reproducing Kernels. Springer Singapore, 1 ed., 2016, doi:10.1007/978-981-10-0530-5. Springer Science+Business Media Singapore 2016.
- [9] Fukumizu, K., Bach, F. R., y Jordan, M. I., “Dimensionality reduction for supervised learning with reproducing kernel hilbert spaces”, Journal of Machine Learning Research, vol. 5, 2004.
- [10] Song, L., Huang, J., Smola, A., y Fukumizu, K., “Hilbert space embeddings of conditional distributions with applications to dynamical systems”, en Proceedings of the 26th International Conference On Machine Learning, ICML 2009, 2009.
- [11] Williams, M. O., Kevrekidis, I. G., y Rowley, C. W., “A data-driven approximation of the koopman operator: Extending dynamic mode decomposition”, Journal of Nonlinear Science, vol. 25, pp. 1307–1346, 2015, doi:10.1007/s00332-015-9258-5.
- [12] Xiang, S., Zhu, Y., Shen, X., y Ye, J., “Optimal exact least squares rank minimization”, 2012, doi:10.1145/2339530.2339609.
- [13] Philipp, F. M., Schaller, M., Worthmann, K., Peitz, S., y Nüske, F., “Error bounds for kernel-based approximations of the koopman operator”, Applied and Computational Harmonic Analysis, vol. 71, 2024, doi:10.1016/j.acha.2024.101657.

- [14] Köhne, F., Philipp, F. M., Schaller, M., Schiela, A., y Worthmann, K., “ l^∞ -error bounds for approximations of the koopman operator by kernel extended dynamic mode decomposition”, 2024, <http://arxiv.org/abs/2403.18809>.
- [15] Fukumizu, K., Song, L., y Gretton, A., “Kernel bayes’ rule: Bayesian inference with positive definite kernels”, *Journal of Machine Learning Research*, vol. 14, 2013.
- [16] Santin, G. y Schaback, R., “Approximation of eigenfunctions in kernel-based spaces”, *Advances in Computational Mathematics*, vol. 42, 2016, doi:10.1007/s10444-015-9449-5.
- [17] Gebhardt, G. H., Kupcsik, A., y Neumann, G., “The kernel kalman rule: Efficient nonparametric inference by recursive least-squares and subspace projections”, *Machine Learning*, vol. 108, pp. 2113–2157, 2019, doi:10.1007/s10994-019-05816-z.
- [18] Zhou, Y., Chen, D. R., y Huang, W., “A class of optimal estimators for the covariance operator in reproducing kernel hilbert spaces”, *Journal of Multivariate Analysis*, vol. 169, 2019, doi:10.1016/j.jmva.2018.09.003.