



UNIVERSIDAD DE LA RIOJA

TRABAJO FIN DE ESTUDIOS

Título

El filtro de Kalman

Autor/es

Jon Ander Alonso Ortiz De Elguea

Director/es

JOSÉ MARÍA PÉREZ IZQUIERDO

Facultad

Facultad de Ciencia y Tecnología

Titulación

Grado en Matemáticas

Departamento

MATEMÁTICAS Y COMPUTACIÓN

Curso académico

2022-23



El filtro de Kalman, de Jon Ander Alonso Ortiz De Elguea
(publicada por la Universidad de La Rioja) se difunde bajo una Licencia Creative
Commons Reconocimiento-NoComercial-SinObraDerivada 3.0 Unported.
Permisos que vayan más allá de lo cubierto por esta licencia pueden solicitarse a los
titulares del copyright.



UNIVERSIDAD DE LA RIOJA

Facultad de Ciencia y Tecnología

TRABAJO FIN DE GRADO

Grado en Matemáticas

El filtro de Kalman
The Kalman filter

Realizado por:

Jon Ander Alonso Ortiz de Elguea

Tutelado por:

José María Pérez Izquierdo

Logroño, julio de 2023

Resumen

En este trabajo profundizamos en el ámbito del filtro de Kalman, una poderosa herramienta ampliamente usada para la estimación de estados. Comenzamos presentando el ruido blanco gaussiano como derivada del proceso de Wiener dentro del marco de las funciones generalizadas. Después proporcionamos la base teórica de la integral de Itô y las ecuaciones diferenciales estocásticas.

Tras estos conceptos presentamos el filtro de Kalman en su forma discreta como un estimador óptimo del estado de un sistema dinámico inexacto complementado con mediciones ruidosas. Más adelante presentamos el filtro de Kalman discreto generalizado, que amplía la formulación original para manejar sistemas con dinámica no lineal. Exploramos además el filtro Kalman-Bucy, que es su versión continua, y por último su versión híbrida.

Como ejemplo de una aplicación práctica tratamos el problema del seguimiento de la actitud de un sólido en movimiento. Usamos para ello el filtro de Kalman híbrido, empleamos cuaternios para la representación eficaz de las rotaciones tridimensionales y les dotamos de una estructura diferenciable mediante atlas.

Abstract

In this work we delve into the realm of the Kalman filter, a powerful tool widely used for state estimation. We begin by unveiling the concept of white noise as the derivative of the Wiener process, in the framework of generalized functions. Next, we provide a foundation for the Itô integral and stochastic differential equations.

After these concepts, we introduce the Kalman filter in its discrete form as an optimal estimator of the state of an inaccurate dynamic system complemented by noisy measurements. Later we present the generalized discrete Kalman filter, which extends the original formulation to handle systems with non-linear dynamics. We additionally explore the Kalman-Bucy filter, which is its continuous version, and last its hybrid version.

As an example of a practical application, we deal with the problem of tracking the attitude of a moving solid. For that, we use the hybrid Kalman filter, we employ quaternions as an effective representation of three-dimensional rotations and provide them with a differentiable structure through atlases.

Índice general

Resumen	I
Abstract	I
Capítulo 1. Introducción	1
1. Historia	1
2. Presentación sencilla	2
3. Aplicación	3
4. Estructura del trabajo	4
5. Notación	5
Capítulo 2. Probabilidad	7
1. Espacios medibles	7
2. Espacios de probabilidad	10
3. Distribución gaussiana	13
4. Procesos estocásticos	14
5. Ruido blanco gaussiano	17
Capítulo 3. Ecuaciones diferenciales estocásticas	23
1. Motivación	23
2. Integral de Itô	23
3. Ecuaciones diferenciales estocásticas	26
Capítulo 4. Filtro de Kalman	29
1. Estimador de mínima varianza	29
2. Filtro de Kalman discreto	32
3. Filtro de Kalman discreto extendido	36
4. Filtro de Kalman-Bucy	38
5. Filtro de Kalman híbrido	41
Capítulo 5. Seguimiento de la actitud	43
1. Actitud de un sólido	43
2. Cuaternios y rotaciones	43
3. Velocidad angular	47
4. Modelo del sistema	48
5. La variedad S^3	48
6. Estructura del proceso de filtrado	50
7. Linealización de la ecuación para la dinámica	51
8. Linealización de la ecuación para la medición	52
9. Matriz de covarianzas del error	52
10. Etapa de predicción	53
11. Etapa de estimación	54
12. Transformación tras el cambio de carta	54
13. Resumen del algoritmo de filtrado	56

Referencias

57

Conclusiones

59

Introducción

1. Historia

A principios de los sesenta, la URSS llevaba la delantera a los EE.UU. en la carrera espacial. En 1957 envió el primer satélite artificial y al primer ser vivo al espacio, la perra Laika, y en 1961 al primer humano, Yuri Gagarin.

La NASA estaba trabajando en el problema de guiar naves espaciales y para ello había formado en 1958 el Space Task Group en el Centro de Investigación Ames (ARC) de Palo Alto, California. Quería enviar al primer humano a la Luna.

El problema era que en la época la capacidad de computación era escasa. Las simulaciones se realizaban en un IBM 704 en el lenguaje Fortran. El estimador iterativo mínimo-cuadrático que habían planteado inicialmente resultaba pesado, y no encontraban alternativas con precisión.

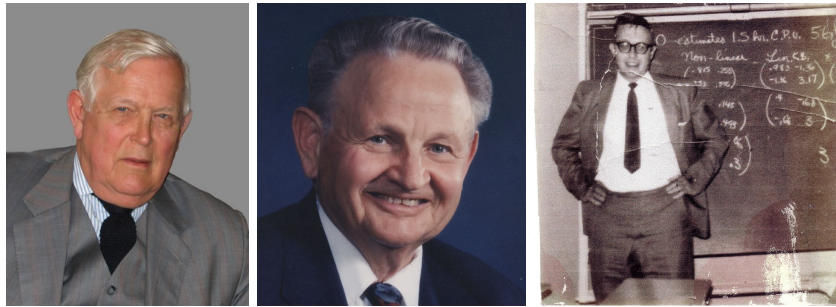


FIGURA 1. Rudolf E. Kálmán, Stanley F. Schmidt y Richard S. Bucy, de izquierda a derecha.

Stanley F. Schmidt (1926-2015), que trabajaba en Ames, conocía al ingeniero húngaroestadounidense Rudolf E. Kálmán (1930-2016) desde hacía años. Por pura casualidad, en otoño de 1960 ambos se reunieron para charlar.

Kalman le presentó [1], un trabajo muy técnico que acababa de publicar, donde presentaba el filtro de Kalman. Schmidt, que había considerado emplear técnicas de filtrado en el ARC, enseguida lo conectó con su trabajo. De vuelta en Ames el filtro resultó encajar. La colaboración culminó con el filtro de Kalman extendido. El ingeniero Richard S. Bucy (1935-2019) también contribuyó a su desarrollo teórico.

Finalmente el filtro se integró en el Computador de Navegación del Apolo (AGC), instalado en la nave Apolo 11. A bordo de esta Neil Armstrong, Michael Collins y Buzz Aldrin aterrizaron en la Luna en 1969.

Aunque en un principio se usara en ingeniería aeroespacial, la simplicidad del filtro ha permitido su aplicación en multitud de ámbitos, incluso en la estimación del comportamiento de la COVID-19 (como muestra [2]). Sigue siendo un algoritmo muy usado para conciliar predicciones teóricas y mediciones, gracias a que es óptimo, requiere pocas operaciones y tiene propiedades recursivas favorables. A continuación exponemos una derivación sencilla del filtro.

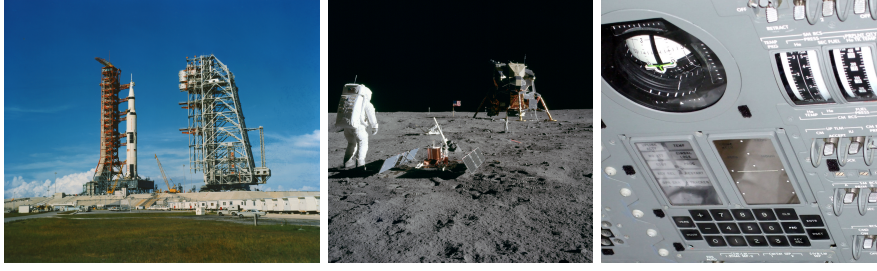


FIGURA 2. El Apolo 11 en su lanzamiento y en la Luna, y el AGC.

2. Presentación sencilla

Consideramos un sistema dinámico lineal en tiempo discreto que en el instante $t_k = t_0 + k\Delta t$ se encuentra en un *estado* \mathbf{x}_k , para cada *etapa* $k \in \mathbb{N}$. La predicción teórica del modelo es:

$$\mathbf{x}_{k+1} = F_k \mathbf{x}_k, \quad k \geq 0.$$

En realidad, el estado estará *oculto* para nosotros. No podremos conocerlo, ni aunque sea con error. Lo *observable* será la *medición* \mathbf{z}_k que tomemos en cada etapa. Esta dependerá del estado, supondremos que de manera lineal:

$$\mathbf{z}_k = H_k \mathbf{x}_k, \quad k \geq 0.$$

El modelo simplifica la realidad. Siempre quedarán pequeñas contribuciones de distinta índole no consideradas que darán lugar a cierto error. Es más fiel reflejar esta inexactitud añadiendo la variable aleatoria \mathbf{w}_k :

$$(1) \quad \mathbf{x}_{k+1} = F_k \mathbf{x}_k + \mathbf{w}_k, \quad k \geq 0.$$

Supondremos que su media es $\bar{\mathbf{w}}_k = 0$.

Las mediciones se verán afectadas por fenómenos aleatorios y no serán exactas sino orientativas. Para reflejar la imprecisión se introduce la variable aleatoria \mathbf{v}_k :

$$(2) \quad \mathbf{z}_k = H_k \mathbf{x}_k + \mathbf{v}_k, \quad k \geq 0.$$

Supondremos que su media es $\bar{\mathbf{v}}_k = 0$.

Por último supondremos que todas las variables \mathbf{w}_k y \mathbf{v}_l son independientes entre ellas y conocemos sus matrices de covarianzas; las cruzadas se anulan y

$$\text{Cov}(\mathbf{w}_k) = Q_k, \quad \text{Cov}(\mathbf{v}_k) = R_k, \quad k \geq 0.$$

En cada etapa nos interesan dos cosas. Por una parte, estimar el estado oculto \mathbf{x}_k con el estimador $\hat{\mathbf{x}}_{k|k}$, que nos dará una idea de cómo es. Por otra, calcular la covarianza del error $\mathbf{P}_{k|k}$, que mide su desviación o incertidumbre.

Una vez tengamos $\hat{\mathbf{x}}_{k|k}$, ¿cómo calculamos $\hat{\mathbf{x}}_{k+1|k+1}$? En una primera fase, si nos atenemos a la dinámica, una *predicción* razonable es

$$\hat{\mathbf{x}}_{k+1|k} := F_k \hat{\mathbf{x}}_{k|k}.$$

Ahora bien, este valor debe ser actualizado para ser compatible con la nueva medición \mathbf{z}_{k+1} . Esta trae consigo una *innovación* que definimos como

$$\begin{aligned} \tilde{\mathbf{z}}_{k+1} &:= \mathbf{z}_{k+1} - H_{k+1} \hat{\mathbf{x}}_{k+1|k} \\ &= H_{k+1} (\mathbf{x}_{k+1} - \hat{\mathbf{x}}_{k+1|k}) + \mathbf{v}_{k+1}. \end{aligned}$$

Usamos esta nueva información para actualizar $\hat{\mathbf{x}}_{k+1|k}$, eligiendo la *estimación*

$$\begin{aligned} \hat{\mathbf{x}}_{k+1|k+1} &:= \hat{\mathbf{x}}_{k+1|k} + \mathbf{K}_{k+1} \tilde{\mathbf{z}}_{k+1} \\ &= (I - \mathbf{K}_{k+1} H_{k+1}) \hat{\mathbf{x}}_{k+1|k} + \mathbf{K}_{k+1} \mathbf{z}_{k+1} \end{aligned}$$

para alguna matriz \mathbf{K}_{k+1} adecuada, a la que llamaremos *ganancia de Kalman*.

Si el estimador de \mathbf{x}_k cumple que $E(\hat{\mathbf{x}}_{k|k}) = E(\mathbf{x}_k)$, entonces la predicción y la estimación de \mathbf{x}_{k+1} cumplen que $E(\hat{\mathbf{x}}_{k+1|k}) = E(\hat{\mathbf{x}}_{k+1|k+1}) = E(\mathbf{x}_{k+1})$. Usaremos las covarianzas de su error y la covarianza de la innovación:

$$\begin{aligned}\mathbf{P}_{k|k} &:= E((\mathbf{x}_k - \hat{\mathbf{x}}_{k|k})(\mathbf{x}_k - \hat{\mathbf{x}}_{k|k})^T), \\ \mathbf{P}_{k+1|k} &:= E((\mathbf{x}_{k+1} - \hat{\mathbf{x}}_{k+1|k})(\mathbf{x}_{k+1} - \hat{\mathbf{x}}_{k+1|k})^T) \\ &= F_k \mathbf{P}_{k|k} F_k^T + Q_k, \\ \mathbf{S}_{k+1} &:= E(\tilde{\mathbf{z}}_{k+1} \tilde{\mathbf{z}}_{k+1}^T) \\ &= H_{k+1} \mathbf{P}_{k+1|k} H_{k+1}^T + R_{k+1}.\end{aligned}$$

Pero, ¿qué matriz \mathbf{K}_{k+1} debemos usar? Como nuestra intención es incorporar toda la información de $\tilde{\mathbf{z}}_{k+1}$ en el estimador $\hat{\mathbf{x}}_{k+1|k+1}$, buscaremos la incorrelación entre $\mathbf{x}_{k+1} - \hat{\mathbf{x}}_{k+1|k+1}$ y $\tilde{\mathbf{z}}_{k+1}$ ¹. Esto obliga a que

$$\begin{aligned}0 &= E((\mathbf{x}_{k+1} - (\hat{\mathbf{x}}_{k+1|k} + \mathbf{K}_{k+1} \tilde{\mathbf{z}}_{k+1})) \tilde{\mathbf{z}}_{k+1}^T) \\ &= E((\mathbf{x}_{k+1} - \hat{\mathbf{x}}_{k+1|k}) \tilde{\mathbf{z}}_{k+1}^T) - \mathbf{K}_{k+1} \mathbf{S}_{k+1} \\ &= E((\mathbf{x}_{k+1} - \hat{\mathbf{x}}_{k+1|k})(\mathbf{x}_{k+1} - \hat{\mathbf{x}}_{k+1|k})^T) H_{k+1}^T - \mathbf{K}_{k+1} \mathbf{S}_{k+1} \\ &= \mathbf{P}_{k+1|k} H_{k+1}^T - \mathbf{K}_{k+1} \mathbf{S}_{k+1}.\end{aligned}$$

Si \mathbf{S}_{k+1} es invertible, entonces podemos calcular \mathbf{K}_{k+1} a partir de $\mathbf{P}_{k+1|k}$, que se calcula a partir de $\mathbf{P}_{k|k}$. Para calcular $\hat{\mathbf{x}}_{k+1|k+1}$ en cada etapa solo necesitaremos llevar cuenta de $\hat{\mathbf{x}}_k$ y $\mathbf{P}_{k|k}$. Ahora, ¿cómo calculamos $\mathbf{P}_{k+1|k+1}$? Notemos que

$$\begin{aligned}\mathbf{x}_{k+1} - \hat{\mathbf{x}}_{k+1|k+1} &= \mathbf{x}_{k+1} - (\hat{\mathbf{x}}_{k+1|k} + \mathbf{K}_{k+1}(H_{k+1}(\mathbf{x}_{k+1} - \hat{\mathbf{x}}_{k+1|k}) + \mathbf{v}_{k+1})) \\ &= (I - \mathbf{K}_{k+1} H_{k+1})(\mathbf{x}_{k+1} - \hat{\mathbf{x}}_{k+1|k}) - \mathbf{K}_{k+1} \mathbf{v}_{k+1}.\end{aligned}$$

Despejando $R_{k+1} = \mathbf{S}_{k+1} - H_{k+1} \mathbf{P}_{k+1|k} H_{k+1}^T$, obtenemos

$$\begin{aligned}\mathbf{P}_{k+1|k+1} &= E((\mathbf{x}_{k+1} - \hat{\mathbf{x}}_{k+1|k+1})(\mathbf{x}_{k+1} - \hat{\mathbf{x}}_{k+1|k+1})^T) \\ &= (I - \mathbf{K}_{k+1} H_{k+1}) \mathbf{P}_{k+1|k} (I - \mathbf{K}_{k+1} H_{k+1})^T + \mathbf{K}_{k+1} R_{k+1} \mathbf{K}_{k+1}^T \\ &= \mathbf{P}_{k+1|k} - \mathbf{K}_{k+1} H_{k+1} \mathbf{P}_{k+1|k} - (\mathbf{P}_{k+1|k} H_{k+1}^T - \mathbf{K}_{k+1} \mathbf{S}_{k+1}) \mathbf{K}_{k+1}^T \\ &= \mathbf{P}_{k+1|k} - \mathbf{K}_{k+1} H_{k+1} \mathbf{P}_{k+1|k} \\ &= (I - \mathbf{K}_{k+1} H_{k+1}) \mathbf{P}_{k+1|k}.\end{aligned}$$

Por tanto, para calcular $\mathbf{P}_{k+1|k+1}$ en cada etapa solo necesitaremos $\mathbf{P}_{k+1|k}$.

3. Aplicación

En resumen, nuestra nueva mejor predicción se basa tan solo en la anterior mejor estimación, con la incertidumbre añadida del ruido:

$$\begin{aligned}\hat{\mathbf{x}}_{k+1|k} &= F_k \mathbf{x}_{k|k}, \\ \mathbf{P}_{k+1|k} &= F_k \mathbf{P}_{k|k} F_k^T + Q_k.\end{aligned}$$

Podemos adelantarnos a la medición entrante preparando los cálculos de la nueva mejor estimación que no dependen de ella:

$$\begin{aligned}\mathbf{S}_{k+1} &= H_{k+1} \mathbf{P}_{k+1|k} H_{k+1}^T + R_{k+1}, \\ \mathbf{K}_{k+1} &= \mathbf{P}_{k+1|k} H_{k+1}^T \mathbf{S}_{k+1}^{-1}.\end{aligned}$$

¹Alternativamente, podemos minimizar $\|\mathbf{x}_{k+1} - \hat{\mathbf{x}}_{k+1|k+1}\|^2$, haciendo que la derivada en $t = 0$ de $\|\mathbf{x}_{k+1} - \hat{\mathbf{x}}_{k+1|k} - (\mathbf{K}_{k+1} + t\Delta)\tilde{\mathbf{z}}_{k+1}\|^2$ se anule para toda matriz Δ . Exigiéndolo a matrices elementales E_{ij} con solo un 1 en la posición (i, j) se llega a la misma condición.

Al llegar la medición los finalizamos, dejando todo listo para la siguiente etapa:

$$\begin{aligned}\tilde{\mathbf{z}}_{k+1} &= \mathbf{z}_{k+1} - H_{k+1}\hat{\mathbf{x}}_{k+1|k}, \\ \hat{\mathbf{x}}_{k+1|k+1} &= \hat{\mathbf{x}}_{k+1|k} + \mathbf{K}_{k+1}\tilde{\mathbf{z}}_{k+1}, \\ \mathbf{P}_{k+1|k+1} &= \mathbf{P}_{k+1|k}(I - \mathbf{K}_{k+1}H_{k+1}).\end{aligned}$$

Esta es la esencia detrás de todas las versiones del *filtro de Kalman* (FK). Además del filtro *discreto* que acabamos de derivar, estudiaremos el filtro:

1. *Extendido*, para sistemas dinámicos no lineales.
2. *Kalman-Bucy*, su versión en tiempo continuo.
3. *Híbrido*, con predicción continua y medición discreta.

Para más variantes del filtro de Kalman, puede consultarse cualquiera de las referencias [3], [4] y [5].

EJEMPLO 1.1. Consideramos un sistema en una dimensión que predecimos constante y del que nos van llegando mediciones. En cada etapa $k \geq 0$,

$$\begin{aligned}\mathbf{x}_{k+1} &= \mathbf{x}_k + \mathbf{w}_k, \\ \mathbf{z}_k &= \mathbf{x}_k + \mathbf{v}_k.\end{aligned}$$

El sistema depende de la estimación inicial $(\hat{\mathbf{x}}_{0|0}, P_{0|0})$, de Q (la varianza del ruido \mathbf{w}_k en la predicción) y de R (la varianza del ruido \mathbf{v}_k en la medición), que tomaremos constantes. Las ecuaciones de predicción son

$$\begin{aligned}\hat{\mathbf{x}}_{k+1|k} &= \mathbf{x}_{k|k}, \\ \mathbf{P}_{k+1|k} &= \mathbf{P}_{k|k} + Q_k, \\ \mathbf{K}_{k+1} &= \frac{\mathbf{P}_{k+1|k}}{\mathbf{P}_{k+1|k} + R_{k+1}} = \frac{1}{1 + \frac{R_{k+1}}{\mathbf{P}_{k|k} + Q_k}}, \\ \hat{\mathbf{x}}_{k+1|k+1} &= \hat{\mathbf{x}}_{k+1|k} + \mathbf{K}_{k+1}(\mathbf{z}_{k+1} - \hat{\mathbf{x}}_{k+1|k}), \\ \mathbf{P}_{k+1|k+1} &= \mathbf{P}_{k+1|k}(1 - \mathbf{K}_{k+1}).\end{aligned}$$

Notemos que $\mathbf{K}_{k+1} \in [0, 1]$, y la ecuación de estimación nos dice que en cada etapa tomamos $\hat{\mathbf{x}}_{k+1|k+1}$ como un punto del segmento que une la predicción teórica $\hat{\mathbf{x}}_{k|k}$ y la medición \mathbf{z}_{k+1} . Depende de \mathbf{K}_{k+1} a cuál se acerca más.

Con la predicción inicial $(\hat{\mathbf{x}}_{0|0}, P_{0|0}) = (1, 1^2)$ como punto de partida, vamos a aplicar las primeras iteraciones del filtro de Kalman en tres casos:

1. $Q = 1^2, R = 3^2$. En este caso $Q < R$, es decir, nos fiamos más de la predicción. Por tanto, la estimación se toma más próxima a la predicción.
2. $Q = 3^2, R = 3^2$. En este caso $Q = R$, es decir, nos fiamos lo mismo de la predicción que de la medición. Por tanto, la estimación se toma en un punto intermedio (no la mitad, por la propagación de $\mathbf{P}_{k|k}$).
3. $Q = 5^2, R = 3^2$. En este caso $Q > R$, es decir, nos fiamos más de la medición. Por tanto, la estimación se toma más próxima a la medición.

En el caso 1, $\hat{\mathbf{x}}_k$ acaba con una varianza de 2.466; en el 2, con una de 5.562; en el 3, con una de 7.025.

Aunque el filtro de Kalman se aplique en modelos n -dimensionales, la filosofía siempre es la misma: tomar una media entre medición y predicción, ponderada por cuánto más nos fiamos de una que de otra.

4. Estructura del trabajo

Tras esta introducción, el trabajo se divide en dos partes.

La primera, que comprende los siguientes dos capítulos, desarrolla la teoría necesaria para las distintas versiones del filtro de Kalman. Estos contenidos incluyen

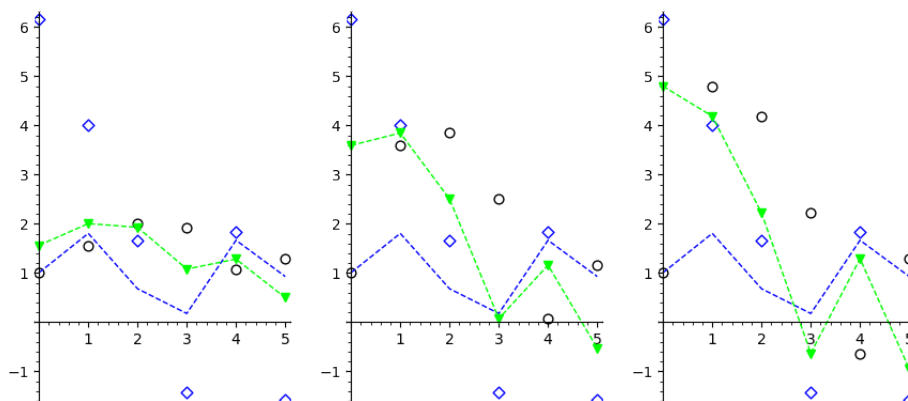


FIGURA 3. Casos 1, 2 y 3. Dinámica real en línea discontinua azul, mediciones en rombo azul, predicciones en círculo negro y estimaciones en triángulo verde unidas por línea discontinua.

el ruido blanco, con el que termina el Capítulo 2, y las ecuaciones diferenciales estocásticas, con las que termina el Capítulo 3, de gran interés matemático.

La segunda, que comprende los últimos dos capítulos, trata íntegramente acerca del filtro de Kalman. En el Capítulo 4 se exponen las versiones mencionadas del filtro de Kalman, junto con resultados teóricos. En el Capítulo 5 se desarrolla completamente un ejemplo que aplica el filtro de Kalman para el control de la actitud, que entre otros contextos precisamente se usa en navegación.

5. Notación

Hacemos uso de los siguientes conjuntos:

- Los números naturales \mathbb{N} , cero incluido.
- Los reales \mathbb{R} , y los reales no negativos \mathbb{R}^+ , cero incluido.
- La recta real extendida $\overline{\mathbb{R}} = \mathbb{R} \cup \{\pm\infty\}$ y $\overline{\mathbb{R}}^+ = \mathbb{R}^+ \cup \{+\infty\}$.
- El conjunto de n -tuplas de reales \mathbb{R}^n , para $n \in \mathbb{N} \setminus \{0\}$. Un vector $\mathbf{X} \in \mathbb{R}^n$ podrá expresarse componente a componente como $\mathbf{X} = (X^1, \dots, X^n)^T$.
- El conjunto de matrices $n \times m$ en los reales $\mathcal{M}_{n \times m}(\mathbb{R})$, para $n, m \in \mathbb{N} \setminus \{0\}$.
- Los complejos \mathbb{C} , para los que las dos últimas definiciones son análogas.
- Dados dos espacios vectoriales V y U , el conjunto $L(V, U)$ de todas las aplicaciones lineales del primero al segundo.

Los vectores son columnas. Las desigualdades entre vectores han de darse componente a componente. Las matrices pueden verse como vectores de longitud nm .

Por comodidad en las expresiones, definimos $0 \cdot \infty = \infty$.

Dados $m, n \in \mathbb{N}$, definimos $\delta_{m,n}$ como 0 salvo si $m = n$; entonces, vale 1.

Dados $s, t \in \mathbb{R}$, definimos $\delta(s, t)$ como 0 salvo si $s = t$; entonces, vale t .

Dados dos conjuntos $A \subseteq X$, definimos la *función indicatriz* del primero como

$$\chi(x) = \begin{cases} 1 & x \in A \\ 0 & x \notin A \end{cases}, \quad x \in X.$$

Probabilidad

1. Espacios medibles

1.1. Conceptos básicos. Primero repasamos los conceptos básicos de la teoría de la medida. Las demostraciones de esta sección se encuentran en [6] y [7]. En esta sección, Ω es un conjunto arbitrario.

DEFINICIÓN 2.1. Una familia $\mathcal{U} \subseteq \mathcal{P}(\Omega)$ es una σ -álgebra de Ω si cumple que:

1. $\emptyset \in \mathcal{U}$.
2. Si $S \in \mathcal{U}$, entonces $S^c \in \mathcal{U}$, con $S^c := \Omega \setminus S$.
3. Si $S_k \in \mathcal{U}$ para cada $k \in \mathbb{N}$, entonces $\bigcup_{k \in \mathbb{N}} S_k \in \mathcal{U}$.

Llamamos a sus elementos *conjuntos medibles* y al par (Ω, \mathcal{U}) , *espacio medible*. Si \mathcal{U} queda clara por el contexto, identificaremos al espacio por Ω . Una subfamilia de \mathcal{U} que sea σ -álgebra de Ω es una *sub- σ -álgebra* de \mathcal{U} .

Como la intersección arbitraria de σ -álgebras de Ω proporciona una σ -álgebra de Ω , queda justificada la siguiente definición.

DEFINICIÓN 2.2. La σ -álgebra generada por una familia $A \subseteq \mathcal{P}(\Omega)$ es

$$\sigma(A) := \bigcap \{ \mathcal{U} \subseteq \mathcal{P}(\Omega) \mid A \subseteq \mathcal{U}, \mathcal{U} \text{ } \sigma\text{-álgebra} \},$$

es decir, la menor σ -álgebra de Ω que contiene a A .

EJEMPLO 2.1. Si Ω es un espacio topológico cuyos abiertos son τ , llamamos σ -álgebra de Borel a $\mathcal{B}_\Omega := \sigma(\tau)$, y a sus elementos, *borelianos*. Sobre subconjuntos de \mathbb{R}^n por defecto consideramos la σ -álgebra de Borel con la topología usual¹. Denotamos $\mathcal{B}_n := \mathcal{B}_{\mathbb{R}^n}$.

DEFINICIÓN 2.3. Sea (Ω, \mathcal{U}) un espacio medible. Una función $\mu: \mathcal{U} \rightarrow \overline{\mathbb{R}}^+$ es una *medida* en (Ω, \mathcal{U}) si cumple que:

1. $\mu(\emptyset) = 0$.
2. Si $(S_k)_{k \in \mathbb{N}}$ es una familia de conjuntos de \mathcal{U} disjuntos dos a dos, entonces

$$\mu\left(\bigcup_{k=1}^{\infty} S_k\right) = \sum_{k=1}^{\infty} \mu(S_k).$$

Llamamos a la terna $(\Omega, \mathcal{U}, \mu)$ *espacio de medida*. Si μ queda clara por el contexto, identificaremos al espacio por Ω . El espacio o la medida μ se denomina:

- *σ -finito* si Ω es la unión contable de conjuntos medibles de medida finita.
- *Completo* si para todo conjunto de medida nula, sus subconjuntos son medibles. En tal caso, tendrán medida nula.

En este capítulo, $(\Omega, \mathcal{U}, \mu)$ es un espacio de medida y (E, \mathcal{E}) un espacio medible. Si quedan claros por el contexto, no insistiremos en ellos.

DEFINICIÓN 2.4. Una propiedad se cumple *μ -en-casi-todo-punto* (μ -c.t.p.) si se cumple en todo Ω salvo en algún conjunto de medida μ nula.

¹En la compactificación $\overline{\mathbb{R}}$ de \mathbb{R} con la topología usual, la topología está generada por los intervalos de la forma $[-\infty, a)$ e $(b, \infty]$ con $a, b \in \mathbb{R}$.

EJEMPLO 2.2. En \mathbb{R}^n , establecemos la medida de un ortoedro como

$$\mu\left(\prod_{k=1}^n [a_k, b_k]\right) = \prod_{k=1}^n (b_k - a_k),$$

y la extendemos a la *medida exterior* $\mu^*: \mathcal{P}(\mathbb{R}^n) \rightarrow \overline{\mathbb{R}}^+$ definida por

$$\mu^*(S) = \inf\left\{\sum_{k=1}^{\infty} \mu(C_k) \mid S \subseteq \bigcup_{k=1}^{\infty} C_k, C_k \text{ es ortoedro}\right\}, \quad S \subseteq \mathbb{R}^n.$$

Definimos la σ -álgebra de Lebesgue como

$$\mathcal{L} = \{S \subseteq \mathbb{R}^n \mid \mu^*(A) = \mu^*(A \cap S) + \mu^*(A \cap S^c) \text{ para todo } A \subseteq \mathbb{R}^n\}.$$

Por el Teorema de Carathéodory, $(\mathbb{R}^n, \mathcal{L}, \mu^*|_{\mathcal{L}})$ es un espacio de medida completo. Además es σ -finito. A la restricción $\mu^*|_{\mathcal{L}}$ se le llama *medida de Lebesgue* para \mathbb{R}^n .

DEFINICIÓN 2.5. Una aplicación $\varphi: \Omega \rightarrow E$ es \mathcal{U} -medible si para todo $R \in \mathcal{E}$, su preimagen $\varphi^{-1}(R) \in \mathcal{U}$. En tal caso, la σ -álgebra inducida por φ en Ω es

$$\mathcal{U}^\varphi := \varphi^{-1}(\mathcal{E}) := \{\varphi^{-1}(R) \mid R \in \mathcal{E}\} \subseteq \mathcal{U},$$

y la *medida inducida* por φ en \mathcal{E} es $\mu_\varphi(R) := \mu(\varphi^{-1}(R))$ para todo $R \in \mathcal{E}$.

\mathcal{U}^φ es la menor σ -álgebra de Ω para la que φ es medible.

1.2. Integral de Lebesgue. La construiremos por pasos.

DEFINICIÓN 2.6. Una función $f: \Omega \rightarrow \mathbb{R}$ es *medible simple* si $f = \sum_{k=1}^m s_k \chi_{S_k}$ con $m \in \mathbb{N}$, $s_k \in \mathbb{R}$ y $S_k \in \mathcal{U}$ disjuntos dos a dos. Su *integral de Lebesgue* es

$$\int_{\Omega} f d\mu := \sum_{k=1}^m s_k \mu(S_k).$$

DEFINICIÓN 2.7. Sea $f: \Omega \rightarrow \overline{\mathbb{R}}^+$ una función medible. Su *integral de Lebesgue* es $\int_{\Omega} f d\mu := \sup\{\int_{\Omega} h d\mu \mid h \text{ medible simple}, 0 \leq h \leq f\}$.

DEFINICIÓN 2.8. Sea $f: \Omega \rightarrow \overline{\mathbb{R}}$ una función medible, para la que usamos la notación $f^+ := \max\{0, f\}$ y $f^- := \max\{0, -f\}$. Si $\int_{\Omega} f^+ d\mu$ y $\int_{\Omega} f^- d\mu$ son ambos finitos, f no posee integral de Lebesgue. Si no, su *integral de Lebesgue* es

$$\int_{\Omega} f d\mu := \int_{\Omega} f^+ d\mu - \int_{\Omega} f^- d\mu.$$

DEFINICIÓN 2.9. Una función $f: \Omega \rightarrow \overline{\mathbb{R}}$ es *integrable Lebesgue* si es medible y $|f|$ tiene integral de Lebesgue finita.

DEFINICIÓN 2.10. Sea una función $f: \Omega \rightarrow \overline{\mathbb{R}}^n$, donde $f = (f^1, \dots, f^n)^T$ con componentes medibles y con integral de Lebesgue. Su *integral de Lebesgue* es

$$\int_{\Omega} f d\mu := \left(\int_{\Omega} f^1 d\mu, \dots, \int_{\Omega} f^n d\mu \right)^T.$$

La definición es válida para matrices.

DEFINICIÓN 2.11. Dado $S \in \mathcal{U}$, una función $f: S \rightarrow \overline{\mathbb{R}}^n$ es *integrable sobre S* si lo es $f\chi_S$, su extensión que vale 0 en S^c . En tal caso, su *integral sobre S* es

$$\int_S f d\mu := \int_{\Omega} f\chi_S d\mu.$$

DEFINICIÓN 2.12. Dados dos espacios de medida $(\Omega_1, \mathcal{U}_1, \mu_1)$ y $(\Omega_2, \mathcal{U}_2, \mu_2)$, su *espacio producto* es $(\Omega_1 \times \Omega_2, \mathcal{U}_1 \otimes \mathcal{U}_2)$, con $\mathcal{U}_1 \otimes \mathcal{U}_2 = \sigma(\mathcal{U}_1 \times \mathcal{U}_2)$. En él, una medida μ es una *medida producto* si cumple que

$$\mu(S_1 \times S_2) = \mu_1(S_1)\mu_2(S_2), \quad S_1 \in \mathcal{U}_1, S_2 \in \mathcal{U}_2.$$

TEOREMA 2.1 (Medida producto). *Dados dos espacios medibles, siempre existe una medida producto. Si ambos son σ -finitos, esta es única.*

TEOREMA 2.2 (Fubini-Tonelli). *Sean $(\Omega_1, \mathcal{U}_1, \mu_1)$ y $(\Omega_2, \mathcal{U}_2, \mu_2)$ dos espacios σ -finitos y μ su medida producto. Si $f: \Omega_1 \times \Omega_2 \rightarrow \mathbb{R}$ es medible, entonces*

$$\int_{\Omega_1 \times \Omega_2} |f(x, y)| d\mu = \int_{\Omega_1} \left(\int_{\Omega_2} |f(x, y)| d\mu_2 \right) d\mu_1 = \int_{\Omega_2} \left(\int_{\Omega_1} |f(x, y)| d\mu_1 \right) d\mu_2.$$

Si además esta integral es finita, entonces

$$\int_{\Omega_1 \times \Omega_2} f(x, y) d\mu = \int_{\Omega_1} \left(\int_{\Omega_2} f(x, y) d\mu_2 \right) d\mu_1 = \int_{\Omega_2} \left(\int_{\Omega_1} f(x, y) d\mu_1 \right) d\mu_2.$$

La integral de Lebesgue generaliza a la de Riemann, pues toda función integrable Riemann en \mathbb{R}^n es integrable Lebesgue en \mathbb{R}^n (con la medida de Lebesgue) y las integrales coinciden. En estos casos, es común usar la notación de Riemann.

TEOREMA 2.3 (Cambio de variables). *Sea $U \subseteq \mathbb{R}^n$ un abierto y $\varphi: U \rightarrow \mathbb{R}^n$ una función inyectiva diferenciable con derivadas parciales continuas cuyo jacobiano $J\varphi$ no se anula. Si $f: V \rightarrow \mathbb{R}$ es continua con dominio compacto $V \subseteq \varphi(U)$, entonces*

$$\int_{\varphi(U)} f(v) dv = \int_U (f \circ \varphi)(u) |\det(J\varphi)(u)| du.$$

1.3. Espacios L^p . Asumimos familiaridad con los *espacios métricos* (dotados de una *métrica* d), los espacios vectoriales *normados* (dotados de una norma $\|\cdot\|$) y *prehilbert* (dotados de un *producto escalar* $\langle \cdot, \cdot \rangle$), y sus versiones completas respecto del espacio métrico que inducen, los espacios *de Banach* y *de Hilbert*.

Sea $p \in [1, +\infty)$. Dada una función $f: \Omega \rightarrow \overline{\mathbb{R}}$, definimos

$$\|f\|_p := \left(\int_{\Omega} |f|^p d\mu \right)^{1/p}$$

y $\mathcal{L}^p(\Omega, \mathcal{U}, \mu) := \{f: \Omega \rightarrow \overline{\mathbb{R}} \mid f \text{ medible}, \|f\|_p < \infty\}$. Este es un espacio vectorial, pero $\|\cdot\|_p$ no es una norma sino una *seminorma*, ya que puede haber varias funciones de norma nula. Concretamente, $\|f\|_p = 0$ si y solo si $f = 0$ μ -c.t.p.

Para convertirlo en espacio normado, se toma la *identificación métrica*

$$f \sim g \quad \text{si y solo si} \quad f = g \quad \mu\text{-c.t.p.},$$

que es una relación de equivalencia. El espacio $L^p(\Omega, \mathcal{U}, \mu) := \mathcal{L}^p / \sim$ junto con la aplicación $\|\cdot\|_p$ bien definida por $\|[f]\|_p = \|f\|_p$ es un espacio normado. Es más, es un espacio de Banach. De todos, L^2 es el único de Hilbert, con el producto

$$\langle f, g \rangle = \int fg d\mu, \quad f, g \in L^2.$$

Por el contexto quedará claro si hablamos de funciones f o de clases de equivalencia de funciones $[f]$. A menudo identificaremos unas con otras.

Llamaremos L_n^p al espacio de funciones $f: \Omega \rightarrow \overline{\mathbb{R}}^n$ con componentes en L^p .

1.4. Derivada de Radon-Nikodym. Este concepto se usa más adelante.

DEFINICIÓN 2.13. Dadas dos medidas μ y ν en (Ω, \mathcal{U}) , ν es *absolutamente continua respecto de μ* ($\nu \ll \mu$) si dado $S \in \mathcal{U}$, $\mu(S) = 0$ implica que $\nu(S) = 0$.

TEOREMA 2.4 (Radon-Nikodym). *Sean μ y ν medidas σ -finitas en (Ω, \mathcal{U}) . Si $\nu \ll \mu$, entonces existe una función medible $f: \Omega \rightarrow \mathbb{R}^+$ única μ -c.t.p. tal que*

$$(3) \quad \nu(S) = \int_S f d\mu, \quad S \in \mathcal{U}.$$

DEFINICIÓN 2.14. Una función f que cumpla (3) se llama *densidad* o *derivada de Radon-Nikodym* de ν con respecto de μ y se escribe

$$f = \frac{d\nu}{d\mu}.$$

Es una notación muy sugerente, pues la derivada representa la tasa de cambio de una medida respecto a otra.

TEOREMA 2.5. Sean μ y ν medidas σ -finitas en (Ω, \mathcal{U}) tales que $d\nu/d\mu$ existe. Para cualquier función medible $g: \Omega \rightarrow \overline{\mathbb{R}}$, cuando la integral esté definida,

$$\int_{\Omega} g d\nu = \int_{\Omega} g \frac{d\nu}{d\mu} d\mu.$$

TEOREMA 2.6 (Regla de la cadena). Sean λ , μ y ν medidas σ -finitas definidas en (Ω, \mathcal{U}) . Si $d\nu/d\mu$ y $d\mu/d\lambda$ existen entonces $d\nu/d\lambda$ existe y

$$\frac{d\nu}{d\lambda} = \frac{d\nu}{d\mu} \frac{d\mu}{d\lambda} \quad \lambda\text{-c.t.p.}$$

2. Espacios de probabilidad

2.1. Conceptos básicos. Ahora repasamos algunos conceptos básicos de probabilidad. Las demostraciones de esta sección se encuentran en [8].

DEFINICIÓN 2.15. Una *medida de probabilidad* en un espacio medible (Ω, \mathcal{U}) es una medida $P: \mathcal{U} \rightarrow [0, 1]$ tal que $P(\Omega) = 1$. Llamamos a la terna (Ω, \mathcal{U}, P) *espacio de probabilidad*, donde:

- Ω es el *espacio muestral*, cuyos elementos son los *resultados*.
- \mathcal{U} es el *espacio de sucesos*, cuyos elementos son los *sucesos* o *eventos*.
- Dado $S \in \mathcal{U}$, $P(S)$ es su *probabilidad*.

En este contexto, una aplicación medible $X: \Omega \rightarrow E$ es una *variable aleatoria* y su probabilidad inducida $P_X: \mathcal{E} \rightarrow [0, 1]$, su *distribución de probabilidad*. La noción de P -c.t.p. es la de P -casi-seguro (P -c.s.).

Es común que el espacio de salida Ω sea oculto y solo lo conozcamos a través de una variable aleatoria X con datos observables en \mathbb{R}^n . Para ignorar Ω , se trabaja con la σ -álgebra inducida en Ω o directamente con la distribución P_X .

2.2. Funciones de distribución conjunta. Con estas funciones nos desvinculamos completamente de Ω para calcular probabilidades.

DEFINICIÓN 2.16. Una función $F: \mathbb{R} \rightarrow \mathbb{R}$ es *de distribución* si es creciente (aunque no lo sea estrictamente), continua a derecha y cumple que

$$\lim_{x \rightarrow -\infty} F(x) = 0, \quad \lim_{x \rightarrow +\infty} F(x) = 1.$$

Una probabilidad Q en $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$ tiene asociada la *función de distribución*

$$F_Q(x) := Q((-\infty, x]), \quad x \in \mathbb{R}.$$

Una variable aleatoria $X: \Omega \rightarrow \mathbb{R}$ tiene asociada la *función de distribución*

$$F_X(x) = P(\{\omega \in \Omega \mid X(\omega) \leq x\}), \quad x \in \mathbb{R},$$

que es la de su distribución de probabilidad.

Recíprocamente, dada una función de distribución F , existe una única medida de probabilidad Q en $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$ tal que $F = F_Q$. También existe una variable aleatoria X tal que $F = F_X$, tomando por ejemplo $\Omega = (0, 1)$ y haciendo

$$X(\omega) = \inf\{x \in \mathbb{R} \mid F(x) \geq \omega\}, \quad \omega \in \Omega,$$

pero esta no es única porque podríamos haber tomado muchos otros Ω . Así, toda función de distribución está asociada a alguna distribución.

DEFINICIÓN 2.17. Dada una función $F: \mathbb{R} \rightarrow \mathbb{R}$ de distribución, $f: \mathbb{R} \rightarrow \mathbb{R}^+$ es su *función de densidad* si es localmente² integrable y

$$F(x) = \int_{-\infty}^x f(t)dt, \quad x \in \mathbb{R}.$$

Si F está asociada a una probabilidad o a una variable aleatoria, también decimos que f es la función de densidad de esa probabilidad o variable aleatoria.

f es la derivada F en el sentido usual y en el de Radon-Nikodym con respecto de la medida de Lebesgue. No tiene por qué existir. Recíprocamente, toda función $f: \mathbb{R} \rightarrow \mathbb{R}^+$ localmente integrable tal que $\int_{-\infty}^{\infty} f(t)dt = 1$ es la función de densidad de alguna distribución.

DEFINICIÓN 2.18. La variable aleatoria $\mathbf{X}: \Omega \rightarrow \mathbb{R}^n$ tiene asociada la *función de distribución* $F_{\mathbf{X}}: \mathbb{R}^n \rightarrow [0, 1]$ dada por

$$F_{\mathbf{X}}(\mathbf{x}) := P(\{\omega \in \Omega \mid \mathbf{X}(\omega) \leq \mathbf{x}\}), \quad \mathbf{x} \in \mathbb{R}^n.$$

DEFINICIÓN 2.19. Dada una variable aleatoria $\mathbf{X}: \Omega \rightarrow \mathbb{R}^n$, $f_{\mathbf{X}}: \mathbb{R}^n \rightarrow \mathbb{R}^+$ es su *función de densidad* si es integrable y

$$F_{\mathbf{X}}(x_1, \dots, x_n) = \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_n} f_{\mathbf{X}}(y_1, \dots, y_n) dy_n \cdots dy_1, \quad x_1, \dots, x_n \in \mathbb{R}.$$

Con esta, podemos calcular la probabilidad de los sucesos de interés mediante

$$P(\mathbf{X} \in B) = \int_B f_{\mathbf{X}}(y_1, \dots, y_n) dy_n \cdots dy_1.$$

DEFINICIÓN 2.20. Las variables aleatorias $\mathbf{X}_1, \dots, \mathbf{X}_m: \Omega \rightarrow \mathbb{R}^n$ tienen asociada la *función de distribución conjunta* $F_{\mathbf{X}_1, \dots, \mathbf{X}_m}: (\mathbb{R}^n)^m \rightarrow [0, 1]$ dada por

$$F_{\mathbf{X}_1, \dots, \mathbf{X}_m}(x_1, \dots, x_m) = P(\mathbf{X}_1 \leq x_1, \dots, \mathbf{X}_m \leq x_m), \quad x_1, \dots, x_m \in \mathbb{R}^n.$$

2.3. Esperanza, varianza, covarianza y correlación. Estos operadores nos ayudan a entender las variables aleatorias.

DEFINICIÓN 2.21. Dada una variable aleatoria $\mathbf{X}: \Omega \rightarrow \mathbb{R}^n$, cuando exista, su:

1. *Esperanza, valor medio o media* es el vector

$$m_{\mathbf{X}} := E(\mathbf{X}) := \int_{\Omega} \mathbf{X} dP.$$

2. *Varianza* es el escalar, donde $\|\cdot\|$ denota la norma euclídea en \mathbb{R}^n ,

$$\sigma_{\mathbf{X}}^2 := \text{Var}(\mathbf{X}) := \int_{\Omega} \|\mathbf{X} - m_{\mathbf{X}}\|^2 dP = E(\|\mathbf{X} - m_{\mathbf{X}}\|^2).$$

3. *Desviación estándar* es $\sigma_{\mathbf{X}} := \sqrt{\sigma_{\mathbf{X}}^2}$.

La media es un valor que razonablemente podemos esperar de la variable, y la varianza da una idea del grado de dispersión con respecto de la media.

Si la variable aleatoria $\mathbf{X}: \Omega \rightarrow \mathbb{R}^n$ tiene función de densidad $f_{\mathbf{X}}: \mathbb{R}^n \rightarrow \mathbb{R}^+$ y $g: \mathbb{R}^n \rightarrow \mathbb{R}$ es una función medible para la que tanto g como $g(\mathbf{X})$ son integrables, se cumple la *ley del estadístico inconsciente*:

$$E(g(\mathbf{X})) = \int_{\mathbb{R}^n} g(\mathbf{x}) f_{\mathbf{X}}(\mathbf{x}) d\mathbf{x}.$$

De este modo, podemos calcular

$$E(\mathbf{X}) = \int_{\mathbb{R}^n} \mathbf{x} f_{\mathbf{X}}(\mathbf{x}) d\mathbf{x}, \quad \text{Var}(\mathbf{X}) = \int_{\mathbb{R}^n} \|\mathbf{x} - m_{\mathbf{X}}\|^2 f_{\mathbf{X}}(\mathbf{x}) d\mathbf{x}.$$

²Localmente significa que lo es sobre compactos.

DEFINICIÓN 2.22. Dadas dos variables aleatorias $\mathbf{X}, \mathbf{Y}: \Omega \rightarrow \mathbb{R}^n$, su *matriz de:*

1. *Covarianzas* es, si existe, $\text{Cov}(\mathbf{X}, \mathbf{Y}) := E((\mathbf{X} - E(\mathbf{X}))(\mathbf{Y} - E(\mathbf{Y}))^T)$.
Notemos que $\text{Cov}(\mathbf{X}, \mathbf{Y}) = (\text{Cov}(X^i, Y^j))_{i,j=1}^n$ y $\text{Cov}(X^k, X^k) = \text{Var}(X^k)$.
Además $\text{Cov}(\mathbf{X}) = \text{Cov}(\mathbf{X}, \mathbf{X})$.
2. *Correlaciones* es, si existe, $\text{Cor}(\mathbf{X}, \mathbf{Y}) := (\frac{\text{Cov}(X_i, Y_j)}{\sigma_{X_i} \sigma_{Y_j}})_{i,j=1}^n$. Las variables se dicen *incorreladas* si esta es nula. Si no, existe correlación entre ellas.

$\text{Cov}(\mathbf{X})$ es semidefinida positiva y $\text{Cov}(A\mathbf{X}) = A \text{Cov}(\mathbf{X}) A^T$ para cualquier matriz $A \in \mathcal{M}_n(\mathbb{R})$. Más tarde denotaremos $P_{\mathbf{X}, \mathbf{Y}} := \text{Cov}(\mathbf{X}, \mathbf{Y})$.

2.4. Probabilidad y esperanza condicionada e independencia. Presentamos definiciones muy generales de estos conceptos.

DEFINICIÓN 2.23. Sea $\mathbf{X}: \Omega \rightarrow \mathbb{R}^n$ una variable aleatoria integrable y \mathcal{V} una sub- σ -álgebra de \mathcal{U} . Una variable aleatoria $\mathbf{Z}: \Omega \rightarrow \mathbb{R}$ es una *esperanza de \mathbf{X} condicionada a \mathcal{V}* si cumple que:

1. Es \mathcal{V} -medible.
2. Si $S \in \mathcal{V}$, entonces $\int_S \mathbf{X} dP = \int_S \mathbf{Z} dP$.

En tal caso, a \mathbf{Z} se la denota por $E(\mathbf{X}|\mathcal{V})$.

TEOREMA 2.7. $E(\mathbf{X}|\mathcal{V})$ existe y es única c.s.

La idea es construir una estimación de \mathbf{X} que sea consistente con la información que \mathcal{V} proporciona a través de las integrales. Ampliamos la definición.

DEFINICIÓN 2.24. Sea $\mathbf{X}: \Omega \rightarrow \mathbb{R}^n$ una variable aleatoria integrable. La *esperanza de \mathbf{X} condicionada a:*

- La variable aleatoria $\mathbf{Y}: \Omega \rightarrow \mathbb{R}^m$ es $E(\mathbf{X}|\mathbf{Y}) := E(\mathbf{X}|\Omega^{\mathbf{Y}})$.
- El evento $R \in \Omega$ es $E(\mathbf{X}|R) := E(\mathbf{X}|\chi_R)$.

PROPOSICIÓN 2.8. La esperanza condicionada cumple que:

1. Si $a, b \in \mathbb{R}$, entonces $E(a\mathbf{X} + b\mathbf{Y}|\mathcal{V}) = aE(\mathbf{X}|\mathcal{V}) + bE(\mathbf{Y}|\mathcal{V})$ c.s.
2. Si $\mathbf{X} \leq \mathbf{Y}$ c.s. entonces $E(\mathbf{X}|\mathcal{V}) \leq E(\mathbf{Y}|\mathcal{V})$ c.s.
3. $E(E(\mathbf{X}|\mathcal{V})) = E(\mathbf{X})$.
4. Si \mathbf{X} es \mathcal{V} -medible, entonces $E(\mathbf{X}|\mathcal{V}) = \mathbf{X}$ c.s.
5. Si Y es \mathcal{V} -medible e \mathbf{X} es integrable, entonces $E(Y\mathbf{X}|\mathcal{V}) = YE(\mathbf{X}|\mathcal{V})$.
6. Si \mathcal{W} es una sub- σ -álgebra de \mathcal{V} , entonces

$$E(\mathbf{X}|\mathcal{W}) = E(E(\mathbf{X}|\mathcal{V})|\mathcal{W}) = E(E(\mathbf{X}|\mathcal{W})|\mathcal{V}) \quad \text{c.s.}$$

DEFINICIÓN 2.25. Sea $S \in \mathcal{U}$ y \mathcal{V} una sub- σ -álgebra de \mathcal{U} . La *probabilidad de S condicionada a \mathcal{V}* es $P(S|\mathcal{V}) := E(\chi_S|\mathcal{V})$. Análogamente se define su probabilidad condicionada a variables aleatorias y a sucesos.

DEFINICIÓN 2.26. Las sub- σ -álgebras $(\mathcal{U}_k)_{k=1}^n$ de \mathcal{U} son *independientes* si para todo $m \in \mathbb{N}$ y $S_{k_1} \in \mathcal{U}_{k_1}, \dots, S_{k_m} \in \mathcal{U}_{k_m}$ se tiene que $P(\bigcap_{i=1}^m S_{k_i}) = \prod_{i=1}^m P(S_{k_i})$. Análogamente se definen variables aleatorias y sucesos independientes.

Si dos variables son independientes, entonces son incorreladas.

TEOREMA 2.9. Si $\mathbf{X}_1, \dots, \mathbf{X}_{k+m}: \Omega \rightarrow \mathbb{R}^n$ son variables independientes, dadas dos funciones $f: (\mathbb{R}^n)^k \rightarrow \mathbb{R}$ y $g: (\mathbb{R}^n)^m \rightarrow \mathbb{R}$ cualesquiera, si

$$\mathbf{Y} = f(\mathbf{X}_1, \dots, \mathbf{X}_k) \quad \text{y} \quad \mathbf{Z} = g(\mathbf{X}_{k+1}, \dots, \mathbf{X}_{k+m})$$

son variables aleatorias, entonces son independientes.

Esto ocurre porque las σ -álgebras generadas por \mathbf{Y} y \mathbf{Z} son menos finas.

TEOREMA 2.10. Sean $\mathbf{X}_1, \dots, \mathbf{X}_m: \Omega \rightarrow \mathbb{R}^n$ variables aleatorias. Si tienen función de distribución $F_{\mathbf{X}_1, \dots, \mathbf{X}_m}$ o de densidad $f_{\mathbf{X}_1, \dots, \mathbf{X}_m}$, respectivamente:

1. Son independientes si y solo si para todo $x_1, \dots, x_m \in \mathbb{R}^n$,

$$F_{\mathbf{X}_1, \dots, \mathbf{X}_m}(x_1, \dots, x_m) = \prod_{k=1}^m F_{\mathbf{X}_k}(x_k).$$

2. Son independientes si y solo si para todo $x_1, \dots, x_m \in \mathbb{R}^n$,

$$f_{\mathbf{X}_1, \dots, \mathbf{X}_m}(x_1, \dots, x_m) = \prod_{k=1}^m f_{\mathbf{X}_k}(x_k).$$

TEOREMA 2.11. Sean $\mathbf{X}_1, \dots, \mathbf{X}_m: \Omega \rightarrow \mathbb{R}^n$ variables independientes. Si las siguientes matrices de covarianzas existen, entonces

$$\text{Cov} \left(\sum_{k=1}^m \mathbf{X}_k, \sum_{k=1}^m \mathbf{X}_k \right) = \sum_{k=1}^m \text{Cov}(\mathbf{X}_k, \mathbf{X}_k).$$

2.5. Función característica. Esta construcción siempre existe y determina completamente la distribución de una variable aleatoria.

DEFINICIÓN 2.27. Dada una variable aleatoria $\mathbf{X}: \Omega \rightarrow \mathbb{R}^n$, definimos su *función característica* como la función $\phi_{\mathbf{X}}: \mathbb{R}^n \rightarrow \mathbb{C}$ dada por

$$\phi_{\mathbf{X}}(\mathbf{u}) := E(e^{i\langle \mathbf{u}, \mathbf{X} \rangle}) := E(\cos(\langle \mathbf{u}, \mathbf{X} \rangle)) + iE(\sin(\langle \mathbf{u}, \mathbf{X} \rangle)), \quad \mathbf{u} \in \mathbb{R}^n,$$

donde $\mathbf{i} = \sqrt{-1}$ y $\langle \cdot, \cdot \rangle$ denota el producto escalar euclídeo en \mathbb{R}^n .

TEOREMA 2.12. Dadas las variables independientes $\mathbf{X}_1, \dots, \mathbf{X}_m: \Omega \rightarrow \mathbb{R}^n$,

$$\phi_{\mathbf{X}_1 + \dots + \mathbf{X}_m}(\mathbf{u}) = \phi_{\mathbf{X}_1}(\mathbf{u}) \cdots \phi_{\mathbf{X}_m}(\mathbf{u}), \quad \mathbf{u} \in \mathbb{R}^n.$$

TEOREMA 2.13. Si las variables aleatorias \mathbf{X} e \mathbf{Y} cumplen que $\phi_{\mathbf{X}} = \phi_{\mathbf{Y}}$, entonces $F_{\mathbf{X}} = F_{\mathbf{Y}}$ y su distribución coincide.

2.6. Ley de los grandes números y teorema del límite central. Abreviaremos la expresión “independientes idénticamente distribuidas” como “i.i.d.”.

TEOREMA 2.14 (Ley de los grandes números). Sea $(\mathbf{X}_k)_{k \in \mathbb{N}}$ una sucesión de variables reales i.i.d. Si su media m existe, entonces definiendo $S_n = \sum_{k=1}^n \mathbf{X}_k$,

$$P \left(\lim_{n \rightarrow \infty} \frac{S_n}{n} = m \right) = 1.$$

TEOREMA 2.15 (Teorema del límite central). Sea $(\mathbf{X}_k)_{k \in \mathbb{N}}$ una sucesión de variables reales i.i.d. Si su media m y su varianza σ^2 existen, definiendo S_n de la misma forma, entonces para cualesquiera $-\infty < a < b < \infty$ se tiene que

$$\lim_{n \rightarrow \infty} P \left(a \leq \frac{S_n - nm}{\sqrt{n}\sigma} \leq b \right) = \frac{1}{\sqrt{2\pi}} \int_a^b e^{-\frac{x^2}{2}} dx.$$

Aquí nos aparece por primera vez la distribución gaussiana.

3. Distribución gaussiana

3.1. Definición. Una variable $X: \Omega \rightarrow \mathbb{R}$ es *normal* o *gaussiana no degenerada* si para ciertos $m, \sigma \in \mathbb{R}$ con $\sigma \neq 0$, su función de densidad es

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp \left(-\frac{(x-m)^2}{2\sigma^2} \right).$$

Su media es m y su varianza, σ^2 . Lo denotaremos por $X \sim \mathcal{N}(m, \sigma^2)$.

Una variable $\mathbf{X}: \Omega \rightarrow \mathbb{R}^n$ es *normal* o *gaussiana no degenerada* si para cierta matriz C simétrica definida positiva y $m \in \mathbb{R}^n$, su función de densidad es

$$f(\mathbf{x}) = \frac{1}{\sqrt{(2\pi)^n \det(C)}} \exp\left(-\frac{1}{2}(\mathbf{x} - \mathbf{m})^T C^{-1}(\mathbf{x} - \mathbf{m})\right).$$

Su media es \mathbf{m} y su matriz de covarianzas, C . Lo denotaremos por $\mathbf{X} \sim \mathcal{N}(\mathbf{m}, C)$. Las no degeneradas son más difíciles de definir y no insistiremos en ellas.

3.2. Propiedades. Mencionaremos algunas de las más relevantes. Las demostraciones de esta sección se encuentran en [9].

TEOREMA 2.16. Si $\mathbf{X} \sim \mathcal{N}(\mathbf{m}, C)$, entonces $\phi_{\mathbf{X}}(\mathbf{u}) = e^{i(\mathbf{u}, \mathbf{m}) - \frac{1}{2}\sqrt{\mathbf{u}^T C \mathbf{u}}}$.

TEOREMA 2.17. Si $\mathbf{X} \sim \mathcal{N}(\mathbf{m}, C)$, $A \in \mathcal{M}_{m \times n}$ y $b \in \mathbb{R}^m$, entonces

$$A\mathbf{X} + b \sim \mathcal{N}(A\mathbf{m} + b, ACA^T).$$

TEOREMA 2.18. Si $\mathbf{X}_k \sim \mathcal{N}(\mathbf{m}_k, C_k)$ y $a_k \in \mathbb{R}$ para $k \in \{1, 2, \dots, m\}$, y además las variables aleatorias son independientes, entonces

$$\sum_{k=1}^m a_k X_k \sim \mathcal{N}(\mathbf{m}, C), \quad \text{con} \quad \mathbf{m} = \sum_{k=1}^m a_k \mathbf{m}_k, \quad C = \sum_{k=1}^m a_k^2 C_k.$$

4. Procesos estocásticos

4.1. Conceptos básicos. En este capítulo, \mathbb{T} es un conjunto de parámetros. Asumiremos que $\mathbb{T} \subseteq \mathbb{R}$ y $0 \in \mathbb{T}$, y que al operar con ellos no nos salimos de \mathbb{T} .

DEFINICIÓN 2.28. Una familia $\mathcal{F} = (\mathcal{F}_t)_{t \in \mathbb{T}}$ de σ -álgebras de Ω es una *filtración* de \mathcal{U} si cuando $s \leq t$, se tiene que $\mathcal{F}_s \subseteq \mathcal{F}_t \subseteq \mathcal{U}$.

DEFINICIÓN 2.29. Un *proceso estocástico* sobre (Ω, \mathcal{U}, P) con valores en (E, \mathcal{E}) es una familia $(\mathbf{X}(t))_{t \in \mathbb{T}}$ de variables aleatorias de Ω a E , junto con una filtración \mathcal{F} de \mathcal{U} tal que $\mathbf{X}(t)$ es \mathcal{F}_t -medible para cada $t \in \mathbb{T}$.

Pueden definirse sin filtración, tomando luego para cada $t \in \mathbb{T}$ como \mathcal{F}_t a

$$\mathcal{U}_{\leq t}^{\mathbf{X}} := \sigma\left(\bigcup\{\mathbf{X}_s^{-1}(\mathcal{E}) \mid s \leq t\}\right).$$

\mathbb{T} puede ser *discreto* (\mathbb{N} , \mathbb{Z} , etc.) o *continuo* (\mathbb{R} , \mathbb{R}^+ , $[0, T]$ etc.), y al proceso también se le aplicarán estos adjetivos. Respecto a la notación:

1. Fijado $t \in \mathbb{T}$, $\mathbf{X}(t, \cdot): \Omega \rightarrow E$ es la variable parametrizada por t . También podemos denotarla por $\mathbf{X}(t)$ o \mathbf{X}_t , aunque puede causar ambigüedad.
2. Fijado $\omega \in \Omega$, $\mathbf{X}(\cdot, \omega): \mathbb{T} \rightarrow E$ es una *realización* o *trayectoria* del proceso.
3. Fijados ambos, denotamos por $\mathbf{X}(t, \omega)$ al valor de $\mathbf{X}(t)$ en ω .
4. Con $\mathbf{X}(\cdot): \mathbb{T} \times \Omega \rightarrow E$ nos referiremos a todo el proceso. Escribiéndolo como $\mathbf{X}(\cdot): \Omega \rightarrow (\mathbb{T} \rightarrow E)$, podemos verlo como una función aleatoria, que induce una distribución en el conjunto de las trayectorias.

Denotaremos las sucesiones estocásticas por $(\mathbf{x}_k)_{k \in \mathbb{T}}$, en minúscula.

4.2. Tipos. Detallamos algunos tipos de procesos estocásticos según las propiedades de sus funciones de distribución conjuntas.

DEFINICIÓN 2.30. El proceso $\mathbf{X}(\cdot)$ es *gaussiano* si para todo $m \in \mathbb{N} \setminus \{0\}$ y $t_1 < \dots < t_m \in \mathbb{T}$, la distribución de $F_{\mathbf{X}(t_1), \dots, \mathbf{X}(t_m)}$ es gaussiana.

DEFINICIÓN 2.31. El proceso $\mathbf{X}(\cdot)$ es:

- *Estrictamente estacionario* si para todo $m \in \mathbb{N} \setminus \{0\}$, $t_1, \dots, t_m \in \mathbb{T}$ y $\tau \geq 0$,

$$F_{\mathbf{X}(t_1+\tau), \dots, \mathbf{X}(t_m+\tau)} = F_{\mathbf{X}(t_1), \dots, \mathbf{X}(t_m)}.$$

- *Débilmente estacionario* si $E(\mathbf{X}(t))$ y $\text{Var}(\mathbf{X}(t))$ son constantes como funciones de t y para todo $s, t \in \mathbb{T}$ y $\tau \geq 0$,

$$\text{Cov}(\mathbf{X}(t + \tau), \mathbf{X}(t)) = \text{Cov}(\mathbf{X}(s + \tau), \mathbf{X}(s)).$$

DEFINICIÓN 2.32. El proceso $\mathbf{X}(\cdot)$ es de *Markov* si para todo $m \in \mathbb{N} \setminus \{0\}$ y $t_1 \leq \dots \leq t_m \in \mathbb{T}$ se cumple la *condición de Markov*:

$$P(\mathbf{X}(t_m) | \mathbf{X}(t_1), \dots, \mathbf{X}(t_{m-1})) = P(\mathbf{X}(t_m) | \mathbf{X}(t_{m-1})).$$

DEFINICIÓN 2.33. El proceso $\mathbf{X}(\cdot)$ con valores en \mathbb{R}^n tiene *incrementos*:

- *Independientes* si para todo $m \in \mathbb{N} \setminus \{0\}$ y $t_1 \leq \dots \leq t_m \in \mathbb{T}$, los *incrementos* $(\mathbf{X}(t_{k+1}) - \mathbf{X}(t_k))_{k=1}^{m-1}$ son variables independientes.
- *Estacionariamente independientes* si además de cumplir la condición anterior, para cualesquiera $s < t \in \mathbb{T}$ y $\tau \geq 0$, $\mathbf{X}(t + \tau) - \mathbf{X}(s + \tau)$ tiene la misma distribución que $\mathbf{X}(t) - \mathbf{X}(s)$.

PROPOSICIÓN 2.19. La función característica de un proceso $\mathbf{X}(\cdot)$ que tiene incrementos independientes es, para $\mathbf{u}_1, \dots, \mathbf{u}_m \in \mathbb{R}^n$,

$$\phi_{\mathbf{X}(t_1), \dots, \mathbf{X}(t_m)}(\mathbf{u}_1, \dots, \mathbf{u}_m) = \phi_{\mathbf{X}(t_1)}\left(\sum_{i=1}^m \mathbf{u}_i\right) \prod_{k=2}^m \phi_{\mathbf{X}(t_k) - \mathbf{X}(t_{k-1})}\left(\sum_{i=k}^m \mathbf{u}_i\right).$$

Por tanto, la distribución queda determinada sin más que especificar las de $\mathbf{X}(t)$ y $\mathbf{X}(t) - \mathbf{X}(s)$ para cualesquiera $s < t \in \mathbb{T}$.

4.3. Proceso de Wiener. Está relacionado con el ruido blanco que queremos filtrar. Las demostraciones de esta sección se encuentran en [10].

DEFINICIÓN 2.34. Un proceso $W(\cdot)$ con valores en \mathbb{R} parametrizado por \mathbb{R}^+ es un *proceso de Wiener* o *movimiento browniano unidimensional* si cumple que:

1. $W(0) = 0$ c.s.
2. Tiene incrementos independientes.
3. $W(t) - W(s) \sim \mathcal{N}(0, t - s)$ para todo $0 \leq s < t$.

Como $W(t) = W(t) - W(0)$ c.s., entonces $W(t) \sim \mathcal{N}(0, t)$.

TEOREMA 2.20. Sea $W(\cdot)$ un proceso de Wiener unidimensional. Para todo $m \in \mathbb{N} \setminus \{0\}$, $0 = t_0 < \dots < t_m$ y $((a_k, b_k))_{k=1}^m$ tales que $a_k \leq b_k$,

$$\begin{aligned} & P(a_1 \leq W(t_1) \leq b_1, \dots, a_m \leq W(t_m) \leq b_m) \\ &= \int_{a_1}^{b_1} \dots \int_{a_m}^{b_m} \left(g(0; x_1, t_1) \prod_{k=2}^m g(x_{k-1}; x_k, t_k - t_{k-1}) \right) dx_m \dots dx_1. \end{aligned}$$

donde $g(y; x, t) = \frac{1}{\sqrt{2\pi t}} e^{-\frac{(x-y)^2}{2t}}$.

TEOREMA 2.21. Para un proceso de Wiener unidimensional $W(\cdot)$ y $s, t \geq 0$,

$$E(W(t)) = 0, \quad E(W(t)^2) = t, \quad E(W(s)W(t)) = \min(s, t).$$

DEFINICIÓN 2.35. Un proceso $\mathbf{W}(\cdot)$ con valores en \mathbb{R}^n parametrizado por \mathbb{R}^+ es un *proceso de Wiener* o *movimiento browniano n -dimensional* si:

1. Sus componentes son procesos de Wiener unidimensionales.
2. $\sigma(\bigcup\{(W^k(t))^{-1}(\mathbb{R}^n) \mid t \geq 0\})$ son independientes para $k \in \{1, 2, \dots, n\}$.

PROPOSICIÓN 2.22. Cuando $\mathbf{W}(\cdot)$ es un proceso de Wiener n -dimensional, para $k, l \in \{1, 2, \dots, n\}$ se tiene que

$$\begin{aligned} E(W^k(s)W^l(t)) &= \min(s, t)\delta_{k,l}, \quad s, t \geq 0, \\ E((W^k(t) - W^k(s))(W^l(t) - W^l(s))) &= (t - s)\delta_{k,l}, \quad t \geq s \geq 0. \end{aligned}$$

Así pues, un proceso de Wiener $\mathbf{W}(\cdot)$ es de Markov. Al estar compuesto por variables gaussianas independientes, es gaussiano y $\mathbf{W}(t) \sim \mathcal{N}(0, tI)$.

4.4. Construcción del proceso de Wiener. Las demostraciones de esta sección se encuentran en [10, Sec. 3.3]. Supongamos que en cada instante una partícula se desplaza por una distancia de Δx hacia arriba o abajo con la misma probabilidad. Su posición tras el paso $n \in \mathbb{N}$ es

$$S(n) = \sum_{k=1}^n X(k),$$

donde las variables $X(k)$ son independientes y valen $\pm\Delta x$ con igual probabilidad. Además tomamos $S(0) = 0$.

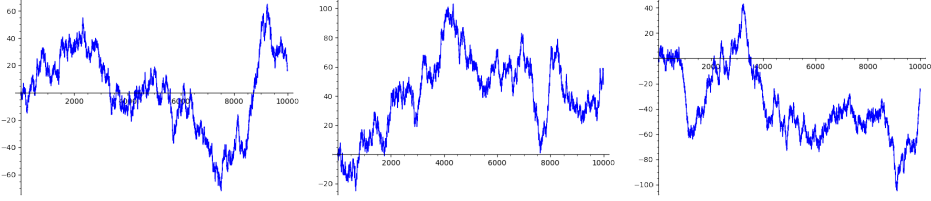


FIGURA 1. Varias realizaciones del paseo aleatorio.

En el espacio adecuado, $S(\cdot)$ es un proceso estocástico discreto. Queremos una variable $S(t)$ continua. Tomaremos el límite de intervalos cada vez menores. Si repartimos cada unidad de tiempo t en n incrementos de Δt segundos,

$$E(X(t)) = 0, \quad \text{Var}(X(t)) = \frac{t}{\Delta t}(\Delta x)^2.$$

Como $t = n\Delta t$, hacer tender Δt a 0 equivale a hacer tender n a infinito. Para que la varianza tenga límite finito, tomaremos $(\Delta x)^2 = \Delta t$. Por el teorema (2.15),

$$\frac{1}{\sqrt{2\pi}} \int_a^b e^{-\frac{x^2}{2}} dx = \lim_{\substack{n \rightarrow \infty \\ t = n\Delta t \\ (\Delta x)^2 = \Delta t}} P\left(a \leq \frac{X(t) - n \cdot 0}{\sqrt{n}\Delta x} \leq b\right) = \lim_{t \rightarrow \infty} P\left(a \leq \frac{X(t)}{\sqrt{t}} \leq b\right),$$

por lo que $S(t) \sim \mathcal{N}(0, t)$. Como $S(n)$ tiene incrementos independientes, $S(t)$ también. La distribución de $S(t) - S(s)$ y $S(t - s) - S(0) \sim \mathcal{N}(0, t - s)$ coincide. Por esto, en el límite el paseo aleatorio es un proceso de Wiener.

Seguiremos otra ruta para la construcción rigurosa del proceso de Wiener; el método Lévy-Ciesielski.

DEFINICIÓN 2.36. Si $k \in \mathbb{N}$, la k -ésima función de Haar es $h_k : [0, 1] \rightarrow \mathbb{R}$, donde $h_0 = 1$, $h_1 = \chi_{[0, \frac{1}{2}]} - \chi_{(\frac{1}{2}, 1]}$ y si $k \in [2^n, 2^{n+1}]$ con $n \in \mathbb{N} \setminus \{0\}$, entonces

$$h_k(t) = \begin{cases} 2^{\frac{n}{2}} & t \in [\frac{k-2^n}{2^n}, \frac{k-2^n+1/2}{2^n}] \\ -2^{\frac{n}{2}} & t \in (\frac{k-2^n+1/2}{2^n}, \frac{k-2^n+1}{2^n}] \\ 0 & \text{en otro caso} \end{cases}, \quad t \in [0, 1].$$

La k -ésima función de Schauder es $s_k : [0, 1] \rightarrow \mathbb{R}$, donde

$$s_k(t) = \int_0^t h_k(s) ds, \quad t \in [0, 1].$$

PROPOSICIÓN 2.23. $(h_k)_{k=0}^\infty$ es una base ortonormada completa de $L^2([0, 1])$.

PROPOSICIÓN 2.24. Si $s, t \in [0, 1]$, entonces $\sum_{k=0}^\infty s_k(s)s_k(t) = \min(s, t)$.

Construiremos procesos “a la carta” con variables adecuadas como coeficientes.

TEOREMA 2.25. Consideramos el proceso estocástico parametrizado por $[0, 1]$

$$W(t, \omega) = \sum_{k=0}^{\infty} A_k(\omega) s_k(t), \quad (t, \omega) \in [0, 1] \times \Omega,$$

donde $(A_k)_{k \in \mathbb{N}}$ son variables independientes de tipo $\mathcal{N}(0, 1)$ definidas en (Ω, \mathcal{U}, P) . Se cumple que:

1. $W(\cdot, \omega)$ converge uniformemente para casi todo $\omega \in \Omega$.
2. $W(\cdot)$ es un proceso de Wiener unidimensional en $[0, 1]$.
3. $W(\cdot, \omega)$ es una trayectoria continua para casi todo $\omega \in \Omega$.

Así, cuando sobre (Ω, \mathcal{U}, P) podemos definir:

- Infinitas variables independientes de tipo $\mathcal{N}(0, 1)$, podemos definir un proceso de Wiener unidimensional parametrizado por \mathbb{R}^+ .
- n procesos de Wiener unidimensionales independientes, podemos definir un proceso de Wiener n -dimensional usándolos como componentes.

4.5. Propiedades de los procesos de Wiener. Las demostraciones de estas propiedades pueden consultarse en [10, Sec. 3.4].

DEFINICIÓN 2.37. Una aplicación $f: [0, T] \rightarrow \mathbb{R}^n$ es Hölder continua en el punto $t \in [0, T]$ si existen $K > 0$ y $\gamma \in (0, 1)$ tales que para todo $s \in [0, T]$,

$$\|f(t) - f(s)\| \leq K \|t - s\|^\gamma.$$

f es Hölder continua en $[0, T]$ si lo es en todo $t \in [0, T]$. Para $\gamma = 1$, esta condición se conoce como *continuidad según Lipschitz*.

Aunque para casi todo ω las trayectorias del movimiento browniano son continuas, distan mucho de ser derivables, como cabía esperar.

DEFINICIÓN 2.38. Dada una aplicación $f: [a, b] \rightarrow \mathbb{R}$, decimos que tiene *variación finita* si el supremo tomado en m y las particiones $(t_k)_{k=0}^{m+1}$ de $[a, b]$ de

$$\sum_{k=0}^m |f(t_{k+1}) - f(t_k)|$$

es finito. En caso contrario, tiene *variación infinita*.

TEOREMA 2.26. Si $T > 0$, para casi todo ω , la trayectoria $\mathbf{W}(\cdot, \omega)$:

1. Es Hölder continua en $[0, T]$, para cualquier exponente $\gamma \in (0, \frac{1}{2})$.
2. No es Hölder continua en $[0, T]$, para ningún exponente $\gamma \in (\frac{1}{2}, 1)$.
3. No es derivable en ningún $t \in [0, T]$.
4. Tiene variación infinita en cada subintervalo de $[0, T]$.

Esto último es un serio problema ya que no vamos a poder integrar en el sentido de Riemann-Stieltjes funciones respecto de estas trayectorias. El cálculo diferencial o integral con procesos que incorporen ruido queda, pues, limitado.

5. Ruido blanco gaussiano

5.1. Conceptos básicos. Esta sección está basada en [11], [12, Cap. 3] y [13, Sec. 2.12]. Para un tratamiento más técnico puede consultarse [14].

Empezamos por el caso discreto.

DEFINICIÓN 2.39. Una *sucesión de ruido blanco* $\mathbf{x}(\cdot)$ es un proceso discreto con valores en \mathbb{R}^n para el que si $s, t \in \mathbb{T}$, entonces:

1. $E(\mathbf{x}(s)) = 0$.
2. $\text{Cov}(\mathbf{x}(s), \mathbf{x}(t)) = \delta_{s,t} Q$ para cierta Q simétrica definida positiva.

Si el proceso es gaussiano, es una *sucesión de ruido blanco gaussiano*.

En el caso gaussiano, además de incorreladas las variables son independientes, y la sucesión queda determinada por Q . Por el teorema 2.15, la suma de muchas contribuciones pequeñas de variables i.i.d. da lugar a este tipo de ruido.

El caso continuo es más complicado. Supongamos que $X^\rho(\cdot)$ es un proceso gaussiano con valores en \mathbb{R} parametrizado por $\rho \in \mathbb{R}^+$, y que además es débilmente estacionario tal que para $t, s \geq 0$,

$$E(X^\rho(t)) = 0, \quad \text{Cov}(X^\rho(t+s), X^\rho(t)) = \sigma^2 \frac{\rho}{2} e^{-\rho|s|},$$

con tal de hacerlo lo más parecido al caso discreto.

Cuando ρ crece, el proceso parece acercarse a lo que deberíamos considerar un proceso de ruido blanco gaussiano $X(t)$, pero la correlación se acerca a

$$\text{Cov}(X(t), X(s)) = \sigma^2 \delta(t-s),$$

donde δ es la *delta de Dirac*, que vale $+\infty$ en 0 y en el resto de puntos se anula. El problema es que δ no es una función en el sentido que le hemos dado hasta ahora.

5.2. Acerca del adjetivo “blanco”. Esta interpretación puede encontrarse en [15, p. 83]. La *densidad espectral de potencia* de $X^\rho(\cdot)$, denotada f^ρ , es la transformada de Fourier de la función de correlación. Así,

$$\begin{aligned} f^\rho(u) &= \int_{-\infty}^{\infty} e^{-isu} \text{Cov}(X^\rho(t+s), X^\rho(t)) ds \\ &= \sigma^2 \frac{\rho}{2} \int_0^{\infty} e^{-usi-\rho s} ds + \sigma^2 \frac{\rho}{2} \int_{-\infty}^0 e^{-usi+\rho s} ds \\ &= \sigma^2 \frac{\rho}{2} \left(\frac{1}{\rho + u\mathbf{i}} + \frac{1}{\rho - u\mathbf{i}} \right) = \frac{\sigma^2}{1 + (u/\rho)^2}. \end{aligned}$$

Notamos que $\lim_{\rho \rightarrow \infty} = \sigma^2$, luego en el límite todas las frecuencias contribuyen lo mismo a la potencia espectral. De igual modo, al descomponer la luz blanca se obtiene la misma contribución de cada frecuencia o “color”.

5.3. Funciones generalizadas. Puede encontrarse una introducción a la teoría de distribuciones en [16, Cap. 6].

DEFINICIÓN 2.40. Una función $\varphi : \mathbb{R}^n \rightarrow \mathbb{R}$ se dice *función test* si es infinitamente diferenciable y su soporte es compacto.

DEFINICIÓN 2.41. $\mathfrak{D}(\mathbb{R}^n, \mathbb{R})$ es el *espacio de funciones test* sobre \mathbb{R}^n .

PROPOSICIÓN 2.27. $\mathfrak{D}(\mathbb{R}^n, \mathbb{R})$ es un \mathbb{R} -espacio vectorial.

DEFINICIÓN 2.42. Una sucesión de funciones test $(\varphi_k)_{k \in \mathbb{N}}$ converge a una función test φ si se cumple que:

1. Existe un conjunto acotado $M \subseteq \mathbb{R}^n$ que contiene a todos sus soportes.
2. Para todo $m_1, \dots, m_n \geq 0$ se tiene convergencia uniforme en

$$\lim_{i \rightarrow \infty} \frac{\partial^{m_1}}{\partial x_1^{m_1}} \cdots \frac{\partial^{m_n}}{\partial x_n^{m_n}} (\varphi_i - \varphi) = 0.$$

DEFINICIÓN 2.43. Un *funcional*³ es una aplicación $F : \mathfrak{D}(\mathbb{R}^n, \mathbb{R}) \rightarrow \mathbb{R}$. Se dice *continuo* si para toda sucesión de funciones test $(\varphi_k)_{k \in \mathbb{N}}$ convergente a φ ,

$$\lim_{k \rightarrow \infty} F(\varphi_k) = F(\varphi).$$

Una *distribución* o *función generalizada* es un funcional lineal y continuo.

³También se puede introducir dotando a $\mathfrak{D}(\mathbb{R}^n, \mathbb{R})$ de topología y considerando su *dual*.

\mathbb{R}^n puede sustituirse por un subconjunto abierto o incluso su clausura sin mayor modificación, algo que haremos sin mención explícita.

EJEMPLO 2.3. La función generalizada δ de Dirac concentrada en 0 es

$$\begin{aligned}\delta: \mathfrak{D}(\mathbb{R}^n, \mathbb{R}) &\rightarrow \mathbb{R} \\ \varphi &\mapsto \delta(\varphi) = \varphi(0).\end{aligned}$$

El siguiente ejemplo justifica el uso extensivo de la palabra “distribución”.

EJEMPLO 2.4. Si $f: \mathbb{R}^n \rightarrow \mathbb{R}$ es una función localmente integrable, entonces

$$F_f: \varphi \mapsto \int_{\mathbb{R}^n} \varphi(x)f(x)dx$$

es una función generalizada. Decimos que f representa a F_f y que F_f es *regular*.

Las distribuciones generalizan a las funciones. Presentan problemas como que el producto no se les puede extender directamente, pero tienen muchas bondades. La que ha motivado este apartado es que si f es una función diferenciable, entonces

$$F_{\partial/\partial x_j(f)}(\varphi) = \int_{\mathbb{R}^n} \frac{\partial}{\partial x_j}(f)\varphi dx = - \int_{\mathbb{R}^n} f \frac{\partial}{\partial x_j}(\varphi) dx = -F_f \left(\frac{\partial}{\partial x_j} \varphi \right).$$

Por tanto, si F_f es una función generalizada, le asociamos su *derivada*

$$\frac{\partial}{\partial x_j} F: \varphi \mapsto -F_f \left(\frac{\partial}{\partial x_j} \varphi \right),$$

sin importarnos si f es derivable o siquiera continua. Igualmente podemos derivar funcionales multilineales $F: \mathfrak{D}(\mathbb{R}^n, \mathbb{R})^m \rightarrow \mathbb{R}$ en este nuevo sentido.

DEFINICIÓN 2.44. Una medida μ en \mathbb{R}^n es *de Radon* si es:

1. *Localmente finita*: todo punto posee un entorno abierto de medida finita.
2. *Internamente regular*: $\mu(S) = \sup\{\mu(K) \mid K \subseteq S, K \text{ compacto}\}$.

EJEMPLO 2.5. Para una medida μ de Radon, la integral de Lebesgue

$$F_\mu: \varphi \rightarrow \int_{\mathbb{R}^n} \varphi d\mu$$

es una distribución. Si μ tiene derivada de Radon-Nikodym f , entonces $F_\mu = F_f$.

5.4. Procesos estocásticos generalizados. El último paso para entender el ruido blanco es parametrizar procesos estocásticos por $\mathfrak{D}(\mathbb{R}, \mathbb{R})$.

DEFINICIÓN 2.45. Un *proceso estocástico generalizado* unidimensional es una aplicación $X: \mathfrak{D}(\mathbb{R}, \mathbb{R}) \times \Omega \rightarrow \mathbb{R}$ que a cada función test φ asigna una variable aleatoria $X(\varphi): \Omega \rightarrow \mathbb{R}$ y cumple:

1. (Linealidad) Dados $a, b \in \mathbb{R}$ y $\varphi, \psi \in \mathfrak{D}(\mathbb{R}, \mathbb{R})$, se cumple que

$$X(a\varphi + b\psi) = aX(\varphi) + bX(\psi) \quad \text{c.t.p.}$$

2. (Continuidad) Si $m \in \mathbb{N}$ y $(\varphi_{l,k})_{k \in \mathbb{N}}^m_{l=1}$ son m sucesiones de funciones test tales que $\lim_{k \rightarrow \infty} \varphi_{l,k} = \varphi_l$ para $l \in \{1, 2, \dots, m\}$ y $g: \mathbb{R}^m \rightarrow \mathbb{R}$ es una función continua y acotada, entonces

$$\lim_{k \rightarrow \infty} E(g(X_{\varphi_{1,k}}, \dots, X_{\varphi_{m,k}})) = E(g(X_{\varphi_1}, \dots, X_{\varphi_m})).$$

EJEMPLO 2.6. Un proceso estocástico $X(\cdot)$ parametrizado por \mathbb{R}^+ con trayectorias localmente integrables *representa* el proceso generalizado

$$X(\cdot, \omega): \varphi \mapsto \int_0^\infty X(t, \omega)\varphi(t)dt, \quad \omega \in \Omega.$$

DEFINICIÓN 2.46. Sea $X(\cdot)$ un proceso estocástico generalizado unidimensional. Si para todo $\varphi \in \mathfrak{D}(\mathbb{R}^+, \mathbb{R})$ existe:

1. $E(X(\varphi))$, entonces está bien definido el *funcional de media*

$$\begin{aligned} \mathbf{m}_X: \mathfrak{D}(\mathbb{R}^+, \mathbb{R}) &\rightarrow \mathbb{R} \\ \varphi &\mapsto E(X(\varphi)). \end{aligned}$$

2. $\text{Var}(X(\varphi))$, entonces está bien definido el *funcional de covarianza*

$$\begin{aligned} \mathfrak{C}_X: \mathfrak{D}(\mathbb{R}^+, \mathbb{R})^2 &\rightarrow \mathbb{R} \\ (\varphi, \psi) &\mapsto E((X(\varphi) - \mathbf{m}(\varphi))(X(\psi) - \mathbf{m}(\psi))). \end{aligned}$$

Se suele exigir que sean continuos. El primero es lineal.

Dado un proceso estocástico $X(\cdot)$, lo vemos como el proceso generalizado que representa. Si cumple las condiciones para la existencia el cambio en el orden de integración, entonces la función $E(X(t))$ representa al funcional \mathbf{m} :

$$\mathbf{m}_X(\varphi) = E(X(\varphi)) = E\left(\int_0^\infty X(t)\varphi(t)dt\right) = \int_0^\infty E(X(t))\varphi(t)dt,$$

Igualmente, la función covarianza representa al funcional \mathfrak{C} :

$$\begin{aligned} \mathfrak{C}_X(\varphi, \psi) &= E((X(\varphi) - \mathbf{m}_X(\varphi))(X(\psi) - \mathbf{m}_X(\psi))) \\ &= \int_0^\infty \int_0^\infty \text{Cov}(X(t), X(s))\varphi(t)\psi(s)dt ds. \end{aligned}$$

La ventaja de los procesos generalizados es que podemos derivarlos.

DEFINICIÓN 2.47. La *derivada* del proceso estocástico generalizado $X(\cdot)$ es el proceso estocástico generalizado $X'(\cdot) := -X\left(\frac{d}{dt}(\cdot)\right)$.

TEOREMA 2.28. Sea $X(\cdot)$ un proceso estocástico generalizado unidimensional.

1. Si \mathbf{m} existe, entonces su derivada es

$$\mathbf{m}'_X(\varphi) = -\mathbf{m}_X(\varphi') = -E(X(\varphi')) = E(X'(\varphi)) = \mathbf{m}_{X'}(\varphi),$$

2. Si \mathfrak{C} existe, entonces su derivada es

$$\begin{aligned} \mathfrak{C}'_X(\varphi, \psi) &= \mathfrak{C}_X(\varphi', \psi') = E((X(\varphi') - \mathbf{m}_X(\varphi'))(X(\psi') - \mathbf{m}_X(\psi'))) \\ &= E((X'(\varphi) - \mathbf{m}_{X'}(\varphi))(X'(\psi) - \mathbf{m}_{X'}(\psi))) = \mathfrak{C}_{X'}(\varphi, \psi) \end{aligned}$$

5.5. Procesos estocásticos generalizados gaussianos. Esta sección puede consultarse en [12, Sec. 2.4].

DEFINICIÓN 2.48. Un proceso estocástico generalizado $X(\cdot)$ con funcionales \mathbf{m}_X y \mathfrak{C}_X continuos es *gaussiano* si para cualesquiera funciones test $\varphi_1, \dots, \varphi_m$ linealmente independientes la variable aleatoria $(X(\varphi_1), \dots, X(\varphi_m))^T$ es gaussiana. Si es no degenerada, el proceso se dice *no degenerado* o *propriadamente gaussiano*.

Un proceso estocástico generalizado propriadamente gaussiano queda determinado por sus funcionales de media y de covarianza. En [12, Sec. 2.3] se demuestra la existencia de este tipo de procesos para funcionales de media y covarianza prefijados.

TEOREMA 2.29. La derivada de un proceso estocástico generalizado gaussiano X es otro con funcionales de media y de covarianza \mathbf{m}'_X y \mathfrak{C}'_X respectivamente.

TEOREMA 2.30. El proceso generalizado de Wiener es gaussiano.

5.6. El ruido blanco gaussiano como derivada del proceso de Wiener. En esta sección seguimos [11, Teorema 2.4].

En el contexto de procesos estocásticos no generalizados, para calcular la derivada del proceso de Wiener deberíamos considerar

$$\frac{W(t+h) - W(t)}{h},$$

cuya distribución es $\mathcal{N}(0, \frac{1}{h})$. Al hacer tender h a 0, el proceso límite $w(\cdot)$, que no existe, debería constar de variables aleatorias $w(t)$ independientes de media nula y varianza infinita.

Si lo consideramos como un proceso estocástico generalizado, entonces su funcional de media es $\mathbf{m}_W = 0$, y el de covarianza

$$\mathfrak{C}_W(\varphi, \psi) = \int_0^\infty \int_0^\infty \min(t, s) \varphi(t) \psi(s) dt ds.$$

Su derivada tendrá funcional de media $\mathbf{m}_{W'} = 0$ y de covarianza

$$\begin{aligned} \mathfrak{C}_{W'}(\varphi, \psi) &= \mathfrak{C}'_W(\varphi, \psi) = \mathfrak{C}_W(\varphi', \psi') = \int_0^\infty \int_0^\infty \min(t, s) \varphi'(t) \psi'(s) dt ds \\ &= \int_0^\infty \left(\int_0^t s \psi'(s) ds \right) \varphi'(t) dt + \int_0^\infty \left(\int_t^\infty \psi'(s) ds \right) t \varphi'(t) dt \\ &= \int_0^\infty \left(\int_0^t s \psi'(s) ds \right) \varphi'(t) dt + \int_0^\infty \psi(t) t \varphi'(t) dt = \int_0^\infty \varphi(t) \psi(t) dt. \end{aligned}$$

En el último paso hemos desarrollado por partes la primera integral, cancelado sumandos y cambiado el orden de integración. Así, observamos que el funcional de covarianza de la derivada del proceso de Wiener es

$$\mathfrak{C}_{W'}(\varphi, \psi) = \int_0^\infty \varphi(t) \psi(t) dt = \int_0^\infty \int_{-\infty}^\infty \varphi(t) \psi(s) \delta(t, s) dt ds.$$

En cuanto a $W'(\varphi)$, observamos que

$$W'(\varphi, u) = -W(\varphi', u) = - \int_0^\infty W(t, u) \varphi'(t) dt.$$

DEFINICIÓN 2.49. Un proceso estocástico generalizado gaussiano $w(\cdot)$ es un *proceso de ruido blanco gaussiano* en $\mathbb{R}_{\geq 0}$ si cumple que:

1. $\mathbf{m}_w = 0$.
2. $\mathfrak{C}_w(\varphi, \psi) = \int_0^\infty \varphi(t) \psi(t) dt$.

TEOREMA 2.31. Si $X(\cdot)$ es un proceso estocástico generalizado gaussiano con $\mathbf{m}_X = 0$ y $\mathfrak{C}_X(\varphi, \psi) = \int_0^\infty \int_0^\infty \min(t, s) \varphi(t) \psi(s) dt ds$, entonces su derivada $X'(\cdot)$ es un proceso generalizado de ruido blanco gaussiano.

DEFINICIÓN 2.50. Un proceso estocástico generalizado $X(\cdot)$ tiene *valores independientes* si cuando los soportes de φ y ψ tienen intersección no trivial, entonces $\mathfrak{C}_X(\varphi, \psi) = 0$.

Este concepto, tratado en [12, Sec. 4.1], generaliza el concepto de sucesión de variables aleatorias independientes. El ruido blanco gaussiano es uno de estos procesos. Existen otros tipos de ruido blanco, como el de Poisson que es el modelo en procesos de emisión de electrones en ciertos semiconductores como diodos, transistores, etc. y se trata en [11, p. 56].

5.7. El ruido blanco como límite de sucesiones de ruido. El ruido blanco gaussiano continuo puede interpretarse como límite de sucesiones del ruido blanco gaussiano visto en el caso discreto. Este ejemplo completa la justificación al inicio de la Sección 5 y puede encontrarse en [15, Ejemplo 3.20].

Sea $\mathbf{x}^\epsilon(\cdot)$ un proceso dependiente de un parámetro $\epsilon > 0$ de modo que

$$\mathbf{x}^\epsilon(t) = \mathbf{x}_n, \quad t \in [n\epsilon, (n+1)\epsilon),$$

donde $(\mathbf{x}_n)_{n \in \mathbb{N}}$ es una sucesión de variables aleatorias independientes gaussianas de media nula y para todo $n, m \in \mathbb{N}$, $\text{Cov}(\mathbf{x}_n, \mathbf{x}_m) = Q\delta_{n,m}$ para alguna matriz simétrica definida positiva Q . Haremos tender ϵ a cero tomando $\epsilon = \frac{1}{m}$ para números naturales m crecientes.

Consideramos la variable aleatoria

$$\mathbf{y}^\epsilon = \int_0^1 \mathbf{x}^\epsilon(t) dt,$$

que tiene media nula y cumple que

$$\begin{aligned} \text{Var}(\|\mathbf{y}^\epsilon\|^2) &= E \left(\int_0^1 \int_0^1 \mathbf{x}^\epsilon(t)^T \mathbf{x}^\epsilon(t) dt \right) = \int_0^1 \int_0^1 E(\mathbf{x}^\epsilon(t)^T \mathbf{x}^\epsilon(t)) dt \\ &= \sum_{i,j=1}^m \int_{(i-1)\epsilon}^{i\epsilon} \int_{(j-1)\epsilon}^{j\epsilon} E(\mathbf{x}_i^T \mathbf{x}_j) ds dt \\ &= \sum_{i,j=1}^m \int_{(i-1)\epsilon}^{i\epsilon} \int_{(j-1)\epsilon}^{j\epsilon} \text{traza}(Q)\delta_{i,j} ds dt = \epsilon \text{traza}(Q). \end{aligned}$$

Representa un movimiento sometido a velocidades aleatorias en el intervalo de tiempo $[0, 1]$. Cada vez usamos más variables $\mathbf{x}(t)$ para describir su aleatoriedad. En el límite no tiene sentido físico que la varianza de \mathbf{y} sea nula. Para evitarlo, se puede sustituir Q por Q/ϵ . En tal caso, cuando $s \neq t$,

$$\text{Cov}(\mathbf{x}^\epsilon(s), \mathbf{x}^\epsilon(t)) = Q/\epsilon$$

para ϵ suficientemente pequeño. Si existe $\mathbf{x}(t) = \lim_{\epsilon \rightarrow 0} \mathbf{x}^\epsilon(t)$, este debería tener matriz de covarianzas $\text{Cov}(\mathbf{x}(t), \mathbf{x}(s)) = Q\delta(t, s)$.

Concluimos que \mathbf{x}^ϵ se aproxima a un ruido blanco gaussiano.

Ecuaciones diferenciales estocásticas

1. Motivación

En este capítulo seguimos [10, Cap. 4], [11, Cap. 3] y [17]. El filtro de Kalman-Bucy trata con sistemas dinámicos continuos con ruido. Es por eso que en los siguientes capítulos aparecerán expresiones del tipo

$$(4) \quad \frac{dX}{dt} = f(t, X) + g(t, X)w(t),$$

donde $w(\cdot)$ es un proceso de ruido blanco gaussiano. En este capítulo nuestro objetivo es dar sentido a expresiones como (4).

Para empezar, podemos dar una interpretación en términos de diferencias. Como, por el teorema 2.31, $w(\cdot)$ es la derivada del proceso de Wiener $W(\cdot)$, la ecuación en diferencias asociada es

$$X_{k+1} - X_k = f(t_k, X_k)\Delta t_k + g(t_k, X_k)\Delta W_k,$$

donde $X_k = X(t_k)$, $W_k = W(t_k)$, $\Delta t_k = t_{k+1} - t_k$ y $\Delta W_k = W_{k+1} - W_k$. Mantendremos esta notación a lo largo del capítulo.

Sumando las ecuaciones anteriores para $k \in \{0, 1, \dots, m+1\}$, obtenemos

$$X_{m+1} = X_0 + \sum_{k=0}^m f(t_k, X_k)\Delta t_k + \sum_{k=0}^m g(t_k, X_k)\Delta W_k.$$

Al hacer tender cada Δt_k a 0, la relación podría interpretarse como

$$(5) \quad X(t) = X(0) + \int_0^t f(s, X(s))ds + \int_0^t g(s, X(s))dW(s),$$

si el resultado tuviese sentido. Aún no disponemos de un significado para el último sumando; precisamente este el objetivo de la integral de Itô.

2. Integral de Itô

2.1. Inconvenientes. En el cálculo convencional, cuando tenemos dos funciones $f, \alpha: [0, T] \rightarrow \mathbb{R}$ adecuadas, podemos considerar la expresión

$$S(P, f, \alpha) := \sum_{k=0}^m f(c_k)\Delta\alpha_k$$

asociada a una partición $P = \{0 = t_0 < t_1 < \dots < t_{m+1} = T\}$ del intervalo $[0, T]$, donde se toman $c_k \in [t_k, t_{k+1}]$ arbitrarios. Para obtener su *integral de Riemann-Stieltjes*

$$\int_0^T f(t)d\alpha(t),$$

se toma el *diámetro* de la partición, $\text{diam}(P) = \max\{t_{k+1} - t_k | k = 0, 1, \dots, m\}$, y se hace tender a 0. En el contexto de los procesos estocásticos, algunas de las reglas más básicas del cálculo cambian y la construcción no es tan sencilla; no basta con cambiar α por W . El siguiente teorema plantea un primer problema; informalmente, que $dW \sim (dt)^{1/2}$.

PROPOSICIÓN 3.1 (Variación cuadrática). Si $P = \{t_0 < t_1 < \dots < t_{m+1}\}$ es una partición de $[0, T]$, entonces, cuando $\text{diam}(P) \rightarrow 0$, converge en $L^2(\Omega)$ el límite

$$\sum_{k=0}^m (W(t_{k+1}) - W(t_k))^2 \rightarrow T.$$

DEMOSTRACIÓN. Notemos que $Y_k := \frac{\Delta W_k}{\sqrt{t_{k+1} - t_k}} \sim \mathcal{N}(0, 1)$ y como W tiene incrementos independientes, por el teorema 2.9 los Y_k son independientes. Así,

$$\begin{aligned} E \left(\sum_{k=1}^m ((\Delta W_k)^2 - T)^2 \right) &= \sum_{k=1}^m E \left(((\Delta W_k)^2 - \Delta t_k)^2 \right) = \sum_{k=1}^m E \left((Y_k^2 - 1)^2 (\Delta t_k)^2 \right) \\ &= E \left((Y^2 - 1)^2 \right) \sum_{k=1}^m (\Delta t_k)^2 \leq E \left((Y^2 - 1)^2 \right) T \text{diam}(P) \rightarrow 0 \end{aligned}$$

cuando $\text{diam}(P) \rightarrow 0$, donde $Y \sim \mathcal{N}(0, 1)$. \square

Un problema más profundo es que dependiendo del punto $c_k \in [t_k, t_{k+1}]$ a elegir, la integral cambia. Esto da pie a varios tipos de integrales estocásticas. Aquí solo se expondrá la de Itô, pero también existe, por ejemplo, la de Stratonovich, que además fue un precursor del filtro de Kalman.

2.2. Integral de Itô de un proceso unidimensional elemental. En esta sección, sea $\mathbf{X}(\cdot)$ un proceso estocástico n -dimensional en un espacio (Ω, \mathcal{U}, P) .

DEFINICIÓN 3.1. La *historia* de $\mathbf{X}(\cdot)$ hasta el momento t es la σ -álgebra

$$\mathcal{U}_{\leq t}^{\mathbf{X}} = \sigma(\{\mathbf{X}_s^{-1}(\mathcal{B}_{\mathbb{R}}) | s \leq t\}).$$

DEFINICIÓN 3.2. El *futuro* de $\mathbf{X}(\cdot)$ posterior al momento t es la σ -álgebra

$$\sigma(\{\mathbf{X}_s - \mathbf{X}_t\}^{-1}(\mathcal{B}_{\mathbb{R}}) | s \geq t\}).$$

DEFINICIÓN 3.3. Una *filtración* de $\mathbf{X}(\cdot)$ es una filtración \mathcal{F} de \mathcal{U} parametrizada por \mathbb{R}^+ tal que para todo $t \geq 0$ cumple que:

1. \mathcal{F}_t contiene la historia de $\mathbf{X}(\cdot)$ hasta el momento t .
2. \mathcal{F}_t es independiente del futuro de $\mathbf{X}(\cdot)$ posterior al momento t .

DEFINICIÓN 3.4. Sea \mathcal{F} una filtración de $\mathbf{X}(\cdot)$. Un proceso estocástico n -dimensional $\mathbf{Y}(\cdot)$ se dice:

1. *Adaptado* a \mathcal{F} si $\mathbf{Y}(t)$ es \mathcal{F}_t -medible para todo $t \geq 0$.
2. *Progresivamente medible* si $\mathbf{Y}(\cdot)$ es $\mathcal{B}_{[0,t]} \otimes \mathcal{F}_t$ -medible para todo $t \geq 0$.

DEFINICIÓN 3.5. Dado $p \in [1, \infty)$, llamamos $\mathbb{L}^p([0, T])$ al espacio de procesos progresivamente medibles tales que

$$E \left(\int_0^T |X(t)|^p dt \right) < \infty.$$

Emplearemos el caso donde $p \in \{1, 2\}$.

DEFINICIÓN 3.6. $Y(\cdot)$ es un *proceso estocástico elemental* o *simple* si cumple que $Y(\cdot) \in \mathbb{L}^2([0, T])$ y para cierto $m \in \mathbb{N}$, partición $P = \{t_0 < t_1 < \dots < t_{m+1}\}$ de $[0, T]$ y variables acotadas Y_1, \dots, Y_m en Ω ,

$$(6) \quad Y(t, \omega) = \sum_{k=0}^m Y_k(\omega) \chi_{[t_k, t_{k+1})}(t), \quad \omega \in \Omega.$$

Su *integral de Itô* es la variable aleatoria

$$\int_0^T Y dW := \sum_{k=0}^{m-1} Y_k (W(t_{k+1}) - W(t_k)).$$

Para $T = 0$, definimos $\int_0^0 Y dW = 0$.

Puede comprobarse que la elección de partición o variables aleatorias en (6) no afecta a la integral y por tanto está bien definida.

2.3. Integral de Itô de procesos unidimensionales no elementales.

Denotamos por $\mathbb{E}([0, T])$ al espacio vectorial de procesos elementales, que es subespacio de $\mathbb{L}^2([0, T])$ dotado de la seminorma $\|\cdot\|$ dada por

$$\|Y\|^2 = E \left(\int_0^T Y(t)^2 dt \right), \quad Y \in \mathbb{L}^2([0, T]).$$

Podemos ver la integral de Itô como una aplicación

$$\begin{aligned} \mathbb{E}([0, T]) &\rightarrow \mathbb{L}^2(\Omega) \\ X &\mapsto \int_0^T X dW. \end{aligned}$$

Usando que $W(\cdot)$ tiene incrementos independientes puede probarse el siguiente resultado, que dice que es una isometría (no necesariamente suprayectiva).

TEOREMA 3.2. *Para cualquier proceso elemental $Y(\cdot)$,*

$$E \left(\int_0^T Y(t)^2 dt \right) = E \left(\left(\int_0^T Y dW \right)^2 \right).$$

Nos gustaría extender la integral a procesos que no sean elementales. Nos es útil el siguiente teorema, que requiere cierta familiaridad con la topología.

TEOREMA 3.3 (Teorema de extensión). *Sea (M, d) un espacio pseudo-métrico y (M', d') un espacio métrico completo. Si $\xi: M_0 \rightarrow M'$ es una aplicación Lipschitz continua desde un subconjunto M_0 denso en M , entonces existe una extensión continua $\tilde{\xi}: M \rightarrow M'$ de ξ . Si ξ es una isometría, entonces $\tilde{\xi}$ también.*

DEMOSTRACIÓN. Si $z \in M_0$, entonces $\xi(z) = \tilde{\xi}(z)$. Si $z \notin M_0$, existe alguna sucesión $(x_n)_{n \in \mathbb{N}}$ en M_0 que converge a z , luego es de Cauchy en M y también lo es $(\xi(x_n))_{n \in \mathbb{N}}$ en M' , pues ξ es Lipschitz continua. Por completitud, converge a un límite que tomamos como $\xi(z)$. Que $\tilde{\xi}$ sea continua o isometría es inmediato. \square

Afortunadamente, $\mathbb{E}([0, T])$ es denso en $\mathbb{L}^2([0, T])$, por el siguiente teorema.

TEOREMA 3.4. $\text{cl}(\mathbb{E}([0, T])) = \mathbb{L}^2([0, T])$.

Es un resultado fundamental cuya demostración se encuentra en [17, Sec. 3.1]. Gracias a él, la *integral de Itô* queda definida para todo $\mathbb{L}^2([0, T])$. Si podemos aproximar un proceso por procesos simples, entonces su integral de Itô existe. En realidad, puede tomarse un espacio mayor (véase [17, Sec. 3.3] y [10, Sec. 4.2.4]).

DEFINICIÓN 3.7. La *integral indefinida de Itô* de $X \in \mathbb{L}^2([0, T])$ es la aplicación

$$\begin{aligned} [0, T] &\rightarrow \mathbb{L}^2(\Omega) \\ t &\mapsto \int_0^t X dW. \end{aligned}$$

3. Ecuaciones diferenciales estocásticas

3.1. El Lema de Itô. Las demostraciones de esta sección se encuentran en [10, Sec. 4.3.4] y [11, Teorema 3.5]. Ahora que (5) tiene sentido, tratamos de hallar soluciones en forma de procesos estocásticos.

DEFINICIÓN 3.8. Sea $X(\cdot)$ un proceso estocástico real para el que existen procesos $F \in \mathbb{L}^1([0, T])$ y $G \in \mathbb{L}^2([0, T])$ de modo que si $0 \leq s \leq r \leq T$,

$$X(r) = X(s) + \int_s^r F dt + \int_s^r G dW.$$

En tal caso diremos que $X(\cdot)$ tiene en $[0, T]$ *diferencial estocástico*

$$dX = F dt + G dW.$$

En el cálculo convencional, si $x, y: [0, T] \rightarrow \mathbb{R}$ son funciones derivables y x es solución de la ecuación diferencial

$$dx = f dt + g dw,$$

entonces, por la regla de la cadena, la función $z = x \circ y$ es solución de

$$dz = y' dx = y' f dt + y' g dw.$$

En el cálculo estocástico, si $X(\cdot)$ es solución de una ecuación diferencial, dada una aplicación $\theta(t, x)$, la ecuación que soluciona $\theta(t, X(t))$ la da el lema de Itô.

TEOREMA 3.5 (Lema de Itô). Sea $X(\cdot)$ un proceso con diferencial estocástico

$$dX = F dt + G dW$$

en $[0, T]$. Dada una aplicación $\theta: [0, T] \times \mathbb{R} \rightarrow \mathbb{R}$ continua en las variables (t, x) y con derivadas parciales θ_t, θ_x y θ_{xx} continuas,

$$d(\theta(t, X)) = \theta_t dt + \theta_x dX + \frac{1}{2} \theta_{xx} G^2 dt = \left(\theta_t + \theta_x F + \frac{1}{2} \theta_{xx} G^2 \right) dt + \theta_x G dW,$$

evaluando las derivadas parciales en $(t, X(t))$.

Cuando \mathbf{X} , F o G son n -dimensionales, se tiene un resultado similar.

TEOREMA 3.6. Sea $\mathbf{X}(\cdot)$ un proceso n -dimensional con diferencial estocástico

$$dX^k = F^k dt + G^k dW, \quad k \in \{1, 2, \dots, n\}$$

en $[0, T]$. Dada una aplicación $\theta: [0, T] \times \mathbb{R}^n \rightarrow \mathbb{R}$ continua en las variables (t, \mathbf{x}) y con derivadas parciales θ_t, θ_{x_k} y θ_{x_k, x_l} continuas para $k, l \in \{1, 2, \dots, n\}$,

$$d(\theta(t, X^1, \dots, X^n)) = \theta_t dt + \sum_{k=1}^n \theta_{x_k} dX^k + \frac{1}{2} \sum_{k, l=1}^n \theta_{x_k, x_l} G^k G^l dt.$$

Cuando \mathbf{W} es n -dimensional la regla es similar, pero necesitamos notación.

DEFINICIÓN 3.9. Para $p \in [1, \infty)$, $\mathbb{L}_{n \times m}^p([0, T])$ es el espacio de procesos estocásticos con valores en $\mathcal{M}_{n \times m}(\mathbb{R})$ cuyas componentes están en $\mathbb{L}^p([0, T])$. Denotamos $\mathbb{L}_n^p([0, T]) := \mathbb{L}_{n \times 1}^p([0, T])$.

DEFINICIÓN 3.10. Si $\mathbf{G} \in \mathbb{L}_{n \times m}^2([0, T])$, su *integral de Itô* es el vector aleatorio cuya k -ésima componente es

$$\left(\int_0^T \mathbf{G} d\mathbf{W} \right)^k = \sum_{l=1}^m \int_0^T G^{k,l} dW^l, \quad k \in \{1, 2, \dots, n\}.$$

DEFINICIÓN 3.11. Sea $\mathbf{X}(\cdot)$ un proceso estocástico n -dimensional para el que existen $\mathbf{F} \in \mathbb{L}_n^1([0, T])$ y $\mathbf{G} \in \mathbb{L}_{n \times m}^2([0, T])$ de modo que si $0 \leq s \leq r \leq T$,

$$\mathbf{X}(r) = \mathbf{X}(s) + \int_s^r \mathbf{F} dt + \int_s^r \mathbf{G} d\mathbf{W}.$$

En tal caso diremos que $\mathbf{X}(\cdot)$ tiene en $[0, T]$ *diferencial estocástico*

$$d\mathbf{X} = \mathbf{F} dt + \mathbf{G} d\mathbf{W}.$$

La ecuación escrita por componentes queda

$$dX^k = F^k dt + \sum_{l=1}^m G^{k,l} dW^l, \quad k \in \{1, 2, \dots, n\}.$$

Ahora ya podemos generalizar la regla de Itô.

TEOREMA 3.7. Sea \mathbf{X} un proceso con diferencial estocástico

$$d\mathbf{X} = \mathbf{F} dt + \mathbf{G} d\mathbf{W}$$

en $[0, T]$. Dada una aplicación $\theta: [0, T] \times \mathbb{R}^n \rightarrow \mathbb{R}^m$ continua en las variables (t, \mathbf{x}) y con derivadas parciales $\theta_t, \theta_{x_k}, \theta_{x_k, x_l}$ continuas para $k, l \in \{1, 2, \dots, n\}$,

$$d\theta(t, \mathbf{X}(t)) = \theta_t dt + \sum_{k=1}^n \theta_k dX^k + \frac{1}{2} \sum_{k,l=1}^n \theta_{x_k, x_l} \sum_{r=1}^m G^{k,l} G^{l,r} dt,$$

evaluando las derivadas parciales en $(t, \mathbf{X}(t))$.

3.2. Existencia y unicidad de soluciones. La demostración de esta sección está en [10, Sec. 5.2.3]. En este capítulo \mathbf{W} es un proceso de Wiener n -dimensional.

TEOREMA 3.8. Sean $\mathbf{b}: [0, T] \times \mathbb{R}^n \rightarrow \mathbb{R}^n$ y $\mathbf{B}: [0, T] \times \mathbb{R}^n \rightarrow \mathcal{M}_{n \times m}(\mathbb{R})$ aplicaciones continuas tales que si $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ y $0 \leq t \leq T$, existe una constante $L \in \mathbb{R}^+$ para la que se cumple que:

- $\|\mathbf{b}(t, \mathbf{x}) - \mathbf{b}(t, \mathbf{y})\| \leq L\|\mathbf{x} - \mathbf{y}\|$ y $\|\mathbf{B}(t, \mathbf{x}) - \mathbf{B}(t, \mathbf{y})\| \leq L\|\mathbf{x} - \mathbf{y}\|$.
- $\|\mathbf{b}(t, \mathbf{x})\| \leq L(1 + \|\mathbf{x}\|)$ y $\|\mathbf{B}(t, \mathbf{x})\| \leq L(1 + \|\mathbf{x}\|)$.

Sea \mathbf{X}_0 una variable aleatoria en \mathbb{R}^n tal que:

- $E(\|\mathbf{X}_0\|^2) < \infty$.
- \mathbf{X}_0 es independiente del futuro de \mathbf{W} posterior al momento $t = 0$.

En tal caso, existe un proceso $\mathbf{X} \in \mathbb{L}_n^2([0, T])$ único c.s. con diferencial estocástico

$$d\mathbf{X} = \mathbf{b}(t, \mathbf{X}) dt + \mathbf{B}(t, \mathbf{X}) d\mathbf{W}$$

en $[0, T]$ tal que $\mathbf{X}(0) = \mathbf{X}_0$.

3.3. Ecuaciones diferenciales estocásticas lineales. En esta sección se siguen [18, Sec. 4.4] y [10, Sec. 5.4].

DEFINICIÓN 3.12. La ecuación diferencial estocástica

$$d\mathbf{X} = \mathbf{b}(t, \mathbf{X}) dt + \mathbf{B}(t, \mathbf{X}) d\mathbf{W}$$

es *lineal* si $\mathbf{b}(t, \mathbf{x}) = \mathbf{c}(t) + \mathbf{D}(t)\mathbf{x}$ y $\mathbf{B}(t, \mathbf{x}) = \mathbf{E}(t) + \mathbf{F}(t)\mathbf{x}$, con

$$\begin{aligned} \mathbf{c}: [0, T] &\rightarrow \mathbb{R}^n, & \mathbf{D}: [0, T] &\rightarrow \mathcal{M}_{n \times n}(\mathbb{R}) \\ \mathbf{E}: [0, T] &\rightarrow \mathcal{M}_{n \times m}(\mathbb{R}), & \mathbf{F}: [0, T] &\rightarrow L(\mathbb{R}^n, \mathcal{M}_{n \times m}(\mathbb{R})). \end{aligned}$$

Desarrollando la ecuación y añadiendo una condición inicial, obtenemos

$$(7) \quad \begin{cases} d\mathbf{X}(t) = (\mathbf{c}(t) + \mathbf{D}(t)\mathbf{X}(t))dt + (\mathbf{E}(t) + \mathbf{F}(t)\mathbf{X}(t))d\mathbf{W}, \\ \mathbf{X}(0) = \mathbf{X}_0. \end{cases}$$

Podemos reescribirla con el convenio de Einstein para sumatorios,

$$d\mathbf{X}(t) = (\mathbf{c}(t) + \mathbf{D}(t)\mathbf{X}(t))dt + (\mathbf{E}^l(t) + \mathbf{F}^l(t)\mathbf{X}(t))dW^l,$$

donde ahora $\mathbf{E}^l(t) \in \mathbb{R}^n$ es la l -ésima columna de $\mathbf{E}(t)$, $\mathbf{F}^l(t)\mathbf{X}(t)$ es la l -ésima columna de $\mathbf{F}(t)\mathbf{X}(t)$, y W^l es la l -ésima componente del proceso de Wiener m -dimensional. Bajo las condiciones del teorema 3.8 se tiene el siguiente resultado.

TEOREMA 3.9. *La solución \mathbf{X} de (7) puede expresarse como*

$$\mathbf{X}(t) = \Phi(t) \left(\mathbf{X}_0 + \int_0^t \Phi(s)^{-1} \left(\mathbf{c}(s) - \mathbf{F}^l(s)\mathbf{E}^l(s) \right) ds + \mathbf{E}^l(s)dW^l \right),$$

donde $\Phi(t) \in \mathcal{M}_{n \times n}(\mathbb{R})$ es solución de

$$\begin{cases} d\Phi = \mathbf{D}(t)\Phi dt + \mathbf{F}^l(t)\Phi dW^l, \\ \Phi(0) = I. \end{cases}$$

Si $\mathbf{F} = 0$ y $\mathbf{D} = D$ son constantes, entonces $\Phi(t) = e^{Dt}$.

COROLARIO 3.10. *La solución \mathbf{X} de (7) para $\mathbf{F} = 0$ y $\mathbf{D} = D$ constante es*

$$\mathbf{X}(t) = e^{Dt}\mathbf{X}_0 + \int_0^t e^{D(t-s)} \left(\mathbf{c}(s)ds + \mathbf{E}^l dW^l \right).$$

TEOREMA 3.11. *La solución \mathbf{X} de (7) para $\mathbf{F} = 0$ satisface que:*

1. *La esperanza dada por $\mathbf{m}(t) := E(\mathbf{X}(t))$ es*

$$\mathbf{m}(t) = \Phi(t) \left(E(\mathbf{X}_0) + \int_0^t \Phi(s)^{-1} \mathbf{c}(s) ds \right),$$

y es solución de

$$\begin{cases} \frac{d\mathbf{m}}{dt}(t) = \mathbf{D}(t)\mathbf{m} + \mathbf{c}(t), \\ \mathbf{m}(0) = E(\mathbf{X}_0). \end{cases}$$

2. *La matriz de covarianzas dada por $\mathbf{L}(s, t) = \text{Cov}(\mathbf{X}(s), \mathbf{X}(t))$ es*

$$\mathbf{L}(s, t) = \Phi(s) \left(\text{Cov}(\mathbf{X}_0, \mathbf{X}_0) + \int_0^{\min(t, s)} \Phi(r)^{-1} \mathbf{E}(r)\mathbf{E}(r)^T (\Phi(r)^{-1})^T dr \right) \Phi(t)^T,$$

y $\mathbf{L}(t) := \mathbf{L}(t, t)$ es solución de

$$\begin{cases} \frac{d\mathbf{L}}{dt}(t) = \mathbf{D}(t)\mathbf{L}(t) + \mathbf{L}(t)\mathbf{D}(t)^T + \mathbf{E}(t)\mathbf{E}(t)^T, \\ \mathbf{L}(0) = \text{Cov}(\mathbf{X}_0, \mathbf{X}_0). \end{cases}$$

Filtro de Kalman

1. Estimador de mínima varianza

1.1. Proyección ortogonal. En esta sección, sea (Ω, \mathcal{U}, P) un espacio de probabilidad, $\mathbf{x} \in L_n^2(\Omega)$ y $\mathbf{z} \in L_m^2(\Omega)$.

TEOREMA 4.1 (Proyección ortogonal). *Sea H un espacio de Hilbert y M un subespacio cerrado. Entonces, existen dos aplicaciones*

$$P_M : H \rightarrow M, \quad P_{M^\perp} : H \rightarrow M^\perp$$

únicas tales que para todo $x \in H$, $x = P_M(x) + P_{M^\perp}(x)$. Estas cumplen que

$$\begin{aligned} \|x - P_M(x)\| &= \min\{\|x - m\| \mid m \in M\}, \\ \|x - P_{M^\perp}(x)\| &= \min\{\|x - m\| \mid m \in M^\perp\} \end{aligned}$$

y son lineales y acotadas.

La demostración de este teorema puede consultarse en [16, Theorem 12.4].

1.2. Estimador de mínima varianza. La notación MVE y BAMVE viene inspirada por el trabajo [19].

DEFINICIÓN 4.1. Una variable aleatoria $\hat{\mathbf{x}} \in L_n^2(\Omega)$ es un *estimador de mínima varianza* (o *MVE*) de \mathbf{x} basado en \mathbf{z} si

$$\|\mathbf{x} - \hat{\mathbf{x}}\| = \min\{\|\mathbf{x} - h(\mathbf{z})\| \mid h : L_m^2(\Omega) \rightarrow L_n^2(\Omega)\}.$$

PROPOSICIÓN 4.2. *Dada una aplicación $h : L_m^2(\Omega) \rightarrow L_n^2(\Omega)$ cualquiera, esta cumple que $\langle \mathbf{x} - E(\mathbf{x}|\mathbf{z}), h(\mathbf{z}) \rangle = 0$.*

DEMOSTRACIÓN. Por la proposición 2.8, se sigue que

$$\langle E(\mathbf{x}|\mathbf{z}), h(\mathbf{z}) \rangle = E(h(\mathbf{z})^T E(\mathbf{x}|\mathbf{z})) = E(E(h(\mathbf{z})^T \mathbf{x}|\mathbf{z})) = E(h(\mathbf{z})^T \mathbf{x}) = \langle \mathbf{x}, h(\mathbf{z}) \rangle. \quad \square$$

El siguiente teorema se puede encontrar en [5, Theorem 3.2].

TEOREMA 4.3. *El MVE de \mathbf{x} basado en \mathbf{z} es $E(\mathbf{x}|\mathbf{z})$.*

DEMOSTRACIÓN. Por el teorema de Pitágoras, dada $g : L_m^2(\Omega) \rightarrow L_n^2(\Omega)$ una aplicación cualquiera, aplicando el teorema 4.2 a $h(\cdot) := E(\mathbf{x}|\cdot) - g(\cdot)$,

$$\|\mathbf{x} - g(\mathbf{z})\|^2 = \|\mathbf{x} - E(\mathbf{x}|\mathbf{z})\|^2 + \|E(\mathbf{x}|\mathbf{z}) - g(\mathbf{z})\|^2 \geq \|\mathbf{x} - E(\mathbf{x}|\mathbf{z})\|^2.$$

Esto muestra que $g(\mathbf{z}) = E(\mathbf{x}|\mathbf{z})$ es un MVE, y cualquier otro es este c.s. \square

1.3. Estimador afín de mínima varianza.

DEFINICIÓN 4.2. Sea \mathcal{S} un subespacio de aplicaciones medibles entre \mathbb{R}^m y \mathbb{R}^n tal que el espacio $\mathcal{S}(\mathbf{z}) := \{f(\mathbf{z}) \mid f \in \mathcal{S}\}$ de \mathcal{S} -estimadores basados en \mathbf{z} es un subespacio cerrado de $L_n^2(\Omega)$. El *mejor \mathcal{S} -MVE basado en \mathbf{z}* es la proyección ortogonal $\hat{\mathbf{x}}$ de \mathbf{x} en $\mathcal{S}(\mathbf{z})$. Si \mathcal{S} es el conjunto de transformaciones afines de \mathbb{R}^m en \mathbb{R}^n entonces tal estimador es el *mejor estimador afín de mínima varianza para \mathbf{x} basado en \mathbf{z}* (BAMVE).

Considerando $\mathbf{z} = (\mathbf{z}_1^T \dots \mathbf{z}_l^T)^T$, las definiciones anteriores se pueden ajustar para definir el concepto de mejor estimador afín de mínima varianza para \mathbf{x} basado en varias variables $\mathbf{z}_1, \dots, \mathbf{z}_l$.

Como no siempre es posible calcular el MVE, se recurre al BAMVE, que es más sencillo. Para ello, necesitamos el concepto de la pseudoinversa, una generalización de la matriz inversa para matrices no invertibles que permite expresar de modo sencillo la solución por mínimos cuadrados de sistemas de ecuaciones lineales.

DEFINICIÓN 4.3. Dada $A \in \mathcal{M}_{m \times n}(\mathbb{C})$, se llama *pseudoinversa (de Moore-Penrose)* de A a cualquier matriz $A^\# \in \mathcal{M}_{n \times m}(\mathbb{C})$ que cumpla:

1. $AA^\#A = A$.
2. $A^\#AA^\# = A^\#$.
3. $(AA^\#)^* = AA^\#$.
4. $(A^\#A)^* = A^\#A$.

Aquí A^* denota la conjugada de la transpuesta de A .

Su existencia y unicidad queda justificada en [20, Theorem 12.11 y Theorem 12.9] (también conviene consultar [21, Sec. 7.7]). Generaliza a la inversa, pues si A es invertible, entonces $A^\# = A^{-1}$. Notemos que $(A^\#)^*$ es la pseudoinversa de A^* .

DEFINICIÓN 4.4. Dada $A \in \mathcal{M}_{m \times n}(\mathbb{C})$ y $b \in \mathbb{C}^n$, una *solución por mínimos cuadrados* del sistema $AX = b$ es un $X_0 \in \mathbb{C}^m$ tal que para todo $X \in \mathbb{C}^m$,

$$\|b - AX_0\| \leq \|b - AX\|.$$

El siguiente teorema puede encontrarse en [21, Theorem 7.7.1]. Por $\text{col}(A)$ nos referimos al subespacio de columnas de A .

TEOREMA 4.4. Sea $A \in \mathcal{M}_{m \times n}(\mathbb{C})$ y $b \in \mathbb{C}^n$. Si $X_0 = A^\#b$, entonces el vector:

1. AX_0 es la proyección ortogonal de b sobre $\text{col}(A)$.
2. X_0 es la solución de $AX = b$ por mínimos cuadrados de norma mínima.

DEMOSTRACIÓN. (1): Claramente $AA^\#b \in \text{col}(A)$, y

$$\langle b - AA^\#b, AX \rangle = (b - AA^\#b)^* AX = b^* AX - b^* AA^\# AX = b^* AX - b^* AX = 0.$$

Esto prueba que $b - AA^\#b \in \text{col}(A)^\perp$. Para (2), por el teorema de Pitágoras,

$$\|b - AX\|^2 = \|b - AA^\#b\|^2 + \|A(A^\#b - X)\|^2 \geq \|b - AA^\#b\|^2,$$

por lo que $AA^\#b$ es solución por mínimos cuadrados. Si Y_0 es otra, entonces

$$\|b - AY_0\|^2 = \|b - AA^\#b\|^2 + \|A(A^\#b - Y_0)\|^2,$$

luego $AA^\#b - AY_0 = 0$. Desarrollamos la expresión

$$\begin{aligned} \langle A^\#b, Y_0 - A^\#b \rangle &= (A^\#b)^*(Y_0 - A^\#b) = b^*(A^\#)^* A^*(A^\#)^*(Y_0 - A^\#b) \\ &= b^*(A^\#)^*(A^\#A)^*(Y_0 - A^\#b) = b^*(A^\#)^* A^\#A(Y_0 - A^\#b) = 0. \end{aligned}$$

Así, por el Teorema de Pitágoras, $\|Y_0\|^2 = \|Y_0 - A^\#b\|^2 + \|A^\#b\|^2 \geq \|A^\#b\|^2$, con igualdad si y solo si $Y_0 = A^\#b$. \square

Los siguientes resultados sobre el BAMVE aparecen en [5, Sec. 5.2].

TEOREMA 4.5. Un BAMVE de \mathbf{x} basado en \mathbf{z} es

$$E^\ell(\mathbf{x} | \mathbf{z}) := E(\mathbf{x}) + P_{\mathbf{x}, \mathbf{z}} P_{\mathbf{z}}^\# (\mathbf{z} - E(\mathbf{z})).$$

Cualquier otro se diferencia de este en $B(\mathbf{z} - E(\mathbf{z}))$ para alguna matriz B tal que $B(\mathbf{z} - E(\mathbf{z})) = 0$ c.s. Además, $\text{Cov}(\mathbf{x} - E^\ell(\mathbf{x} | \mathbf{z})) = P_{\mathbf{x}} - P_{\mathbf{x}, \mathbf{z}} P_{\mathbf{z}}^\# P_{\mathbf{x}, \mathbf{z}}^T$.

DEMOSTRACIÓN. Toda aplicación afín puede darse como $g(\mathbf{z}) = A(\mathbf{z} - \bar{\mathbf{z}}) + b$. Por el teorema de Pitágoras, la expresión que queremos minimizar es

$$\|\mathbf{x} - g(\mathbf{z})\|^2 = \|\mathbf{x} - \bar{\mathbf{x}} - A(\mathbf{z} - \bar{\mathbf{z}})\|^2 + \|\bar{\mathbf{x}} - b\|^2 \geq \|\mathbf{x} - \bar{\mathbf{x}} - A(\mathbf{z} - \bar{\mathbf{z}})\|,$$

ya que $\langle \bar{\mathbf{x}} - b, \mathbf{x} - \bar{\mathbf{x}} \rangle = \langle \bar{\mathbf{x}} - b, \mathbf{z} - \bar{\mathbf{z}} \rangle = 0$. Por tanto, debe cumplirse que $b = \bar{\mathbf{x}}$. Cambiando \mathbf{x} por $\mathbf{x} - \bar{\mathbf{x}}$ y \mathbf{z} por $\mathbf{z} - \bar{\mathbf{z}}$, podemos asumir que $\bar{\mathbf{x}} = 0$ y $\bar{\mathbf{z}} = 0$ y debemos encontrar una A que minimice $\|\mathbf{x} - A\mathbf{z}\|$. Denotando las covarianzas por P ,

$$\begin{aligned} \|\mathbf{x} - A\mathbf{z}\|^2 &= \text{traza}(\overline{(\mathbf{x} - A\mathbf{z})(\mathbf{x} - A\mathbf{z})^T}) \\ &= \text{traza}(P_{\mathbf{x}}) - 2\text{traza}(AP_{\mathbf{z},\mathbf{x}}) + \text{traza}(AP_{\mathbf{z}}A^T). \end{aligned}$$

Para continuar, necesitamos probar que $P_{\mathbf{z}}P_{\mathbf{z}}^{\#}P_{\mathbf{z},\mathbf{x}} = P_{\mathbf{z},\mathbf{x}}$. Esto es obvio si $P_{\mathbf{z}}$ es invertible, pero debemos justificarlo en general. Si \mathbf{u} es un vector de modo que $\mathbf{u}^T P_{\mathbf{z}} = 0$, entonces $\mathbf{u}^T P_{\mathbf{z}} \mathbf{u} = 0$, luego $\overline{\mathbf{u}^T \mathbf{z} \mathbf{z}^T \mathbf{u}} = 0$ y $\mathbf{u}^T \mathbf{z} = 0$ c.s. Por lo tanto, $\mathbf{u}^T P_{\mathbf{z},\mathbf{x}} = 0$ y $\text{col}(P_{\mathbf{z}})^{\perp} \subseteq \text{col}(P_{\mathbf{z},\mathbf{x}})^{\perp}$. En consecuencia, las columnas de $P_{\mathbf{z},\mathbf{x}}$ son combinación lineal de las de $P_{\mathbf{z}}$. Ahora basta aplicar que $P_{\mathbf{z}}P_{\mathbf{z}}^{\#}P_{\mathbf{z}} = P_{\mathbf{z}}$.

Sea ahora $\alpha(A) := -2\text{traza}(AP_{\mathbf{z},\mathbf{x}}) + \text{traza}(AP_{\mathbf{z}}A^T)$. Debido a que $P_{\mathbf{z}}$ es semi-definida positiva y a las propiedades de la pseudoinversa,

$$\begin{aligned} (8) \quad 0 &\leq \text{traza}((A - P_{\mathbf{x},\mathbf{z}}P_{\mathbf{z}}^{\#})P_{\mathbf{z}}(A - P_{\mathbf{x},\mathbf{z}}P_{\mathbf{z}}^{\#})^T) \\ &= \text{traza}(AP_{\mathbf{z}}A^T) - 2\text{traza}(AP_{\mathbf{z}}(P_{\mathbf{z}}^{\#})^T P_{\mathbf{x},\mathbf{z}}) + \text{traza}(P_{\mathbf{x},\mathbf{z}}P_{\mathbf{z}}^{\#}P_{\mathbf{z}}(P_{\mathbf{z}}^{\#})^T P_{\mathbf{x},\mathbf{z}}^T) \\ &= \text{traza}(AP_{\mathbf{z}}A^T) - 2\text{traza}(AP_{\mathbf{z}}P_{\mathbf{z}}^{\#}P_{\mathbf{z},\mathbf{x}}) + \text{traza}(P_{\mathbf{x},\mathbf{z}}P_{\mathbf{z}}^{\#}P_{\mathbf{x},\mathbf{z}}^T) \\ &= \text{traza}(AP_{\mathbf{z}}A^T) - 2\text{traza}(AP_{\mathbf{z},\mathbf{x}}) + \text{traza}(P_{\mathbf{x},\mathbf{z}}P_{\mathbf{z}}^{\#}P_{\mathbf{x},\mathbf{z}}^T) \\ &= \alpha(A) + \text{traza}(P_{\mathbf{x},\mathbf{z}}P_{\mathbf{z}}^{\#}P_{\mathbf{x},\mathbf{z}}^T). \end{aligned}$$

Por tanto, las matrices A que minimizan $\|\mathbf{x} - A\mathbf{z}\|$ son las que minimizan $\alpha(A)$ o, equivalentemente, la expresión $\text{traza}(BP_{\mathbf{z}}B^T)$, con $B = A - P_{\mathbf{x},\mathbf{z}}P_{\mathbf{z}}^{\#}$.

Como $P_{\mathbf{x},\mathbf{z}}P_{\mathbf{z}}^{\#}$ lo hace, $g(\mathbf{z}) = \bar{\mathbf{x}} + P_{\mathbf{x},\mathbf{z}}P_{\mathbf{z}}^{\#}(\mathbf{z} - \bar{\mathbf{z}})$ es una posibilidad para el BAMVE. Ahora bien, entonces $\text{traza}(BP_{\mathbf{z}}B^T) = 0$. En tal caso, $B(\mathbf{z} - \bar{\mathbf{z}}) = 0$ c.s. y $\bar{\mathbf{x}} + A(\mathbf{z} - \bar{\mathbf{z}}) = \bar{\mathbf{x}} + P_{\mathbf{x},\mathbf{z}}P_{\mathbf{z}}^{\#}(\mathbf{z} - \bar{\mathbf{z}}) + B(\mathbf{z} - \bar{\mathbf{z}})$.

El cálculo de $\text{Cov}(\mathbf{x} - E^{\ell}(\mathbf{x}|\mathbf{z}))$ es sencillo y se omite. \square

TEOREMA 4.6. Sean \mathbf{x} y $\mathbf{z}_1, \dots, \mathbf{z}_k$ variables aleatorias en \mathbb{R}^n y $\mathbb{R}^{m_1}, \dots, \mathbb{R}^{m_k}$, respectivamente. Si para todo $i \neq j$, $\text{Cov}(\mathbf{z}_i, \mathbf{z}_j) = 0$, entonces

$$E^{\ell}(\mathbf{x}|\mathbf{z}_1, \dots, \mathbf{z}_k) = \sum_{i=0}^k E^{\ell}(\mathbf{x}|\mathbf{z}_i) - (k-1)\bar{\mathbf{x}}.$$

DEMOSTRACIÓN. Sea $\mathbf{z} := (\mathbf{z}_1^T \dots \mathbf{z}_k^T)^T$. Se cumple que

$$\begin{aligned} E^{\ell}(\mathbf{x}|\mathbf{z}_1, \dots, \mathbf{z}_k) &= E^{\ell}(\mathbf{x}|\mathbf{z}) = \bar{\mathbf{x}} + P_{\mathbf{x},\mathbf{z}}P_{\mathbf{z}}^{\#}(\mathbf{z} - \bar{\mathbf{z}}) \\ &= \bar{\mathbf{x}} + (P_{\mathbf{x},\mathbf{z}_1} \dots P_{\mathbf{x},\mathbf{z}_k}) \begin{pmatrix} P_{\mathbf{z}_1}^{\#} & & \\ & \ddots & \\ & & P_{\mathbf{z}_k}^{\#} \end{pmatrix} \begin{pmatrix} \mathbf{z}_1 - \bar{\mathbf{z}}_1 \\ \vdots \\ \mathbf{z}_k - \bar{\mathbf{z}}_k \end{pmatrix} \\ &= \bar{\mathbf{x}} + \sum_{i=1}^k P_{\mathbf{x},\mathbf{z}_i}P_{\mathbf{z}_i}^{\#}(\mathbf{z}_i - \bar{\mathbf{z}}_i) = \sum_{i=0}^k E^{\ell}(\mathbf{x}|\mathbf{z}_i) - (k-1)\bar{\mathbf{x}}. \quad \square \end{aligned}$$

Recalcamos la propiedad de ortogonalidad que caracteriza a $E^{\ell}(\mathbf{x}|\mathbf{z})$.

TEOREMA 4.7 (Proyección). Sean \mathbf{x} y \mathbf{z} variables aleatorias en \mathbb{R}^n y \mathbb{R}^m , respectivamente. El error $\tilde{\mathbf{x}} = \mathbf{x} - E^{\ell}(\mathbf{x}|\mathbf{z})$ cumple que $\overline{\tilde{\mathbf{x}}\mathbf{z}^T} = 0$. De hecho:

1. Si $\overline{(\mathbf{x} - A\mathbf{z} - b)\mathbf{z}^T} = 0$ y $\overline{\mathbf{x} - A\mathbf{z} - b} = 0$, entonces $A\mathbf{x} + b = E^{\ell}(\mathbf{x}|\mathbf{z})$ c.s.
2. Si además $P_{\mathbf{z}}$ es invertible, entonces $A\mathbf{x} + b = E^{\ell}(\mathbf{x}|\mathbf{z})$.

DEMOSTRACIÓN. Desarrollando la expresión con el valor del BAMVE,

$$\begin{aligned} \overline{(\mathbf{x} - P_{\mathbf{x},\mathbf{z}}P_{\mathbf{z}}^{\#}(\mathbf{z} - \bar{\mathbf{z}}) - \bar{\mathbf{x}})\mathbf{z}^T} &= \overline{(\mathbf{x} - \bar{\mathbf{x}})\mathbf{z}^T} - P_{\mathbf{x},\mathbf{z}}P_{\mathbf{z}}^{\#}\overline{(\mathbf{z} - \bar{\mathbf{z}})\mathbf{z}^T} \\ &= P_{\mathbf{x},\mathbf{z}} - P_{\mathbf{x},\mathbf{z}}P_{\mathbf{z}}^{\#}P_{\mathbf{z}} = P_{\mathbf{x},\mathbf{z}} - P_{\mathbf{x},\mathbf{z}} = 0. \end{aligned}$$

Si $\overline{\mathbf{x} - A\mathbf{z} - b} = 0$, entonces $b = \bar{\mathbf{x}} - A\bar{\mathbf{z}}$, luego $A\mathbf{z} + b = \bar{\mathbf{x}} + A(\mathbf{z} - \bar{\mathbf{z}})$. Si además $\overline{(\mathbf{x} - A\mathbf{z} - b)\mathbf{z}^T} = 0$, entonces $\overline{(\mathbf{x} - \bar{\mathbf{x}} - A(\mathbf{z} - \bar{\mathbf{z}}))\mathbf{z}^T} = 0$, por lo que $P_{\mathbf{x},\mathbf{z}} - AP_{\mathbf{z}} = 0$. Así, $B := A - P_{\mathbf{x},\mathbf{z}}P_{\mathbf{z}}^{\#}$ cumple que $A = P_{\mathbf{x},\mathbf{z}}P_{\mathbf{z}}^{\#} + B$ y por tanto

$$P_{\mathbf{x},\mathbf{z}} = AP_{\mathbf{z}} = P_{\mathbf{x},\mathbf{z}}P_{\mathbf{z}}^{\#}P_{\mathbf{z}} + BP_{\mathbf{z}} = P_{\mathbf{x},\mathbf{z}} + BP_{\mathbf{z}},$$

es decir, $BP_{\mathbf{z}} = 0$. En tal caso $B(\mathbf{z} - \bar{\mathbf{z}}) = 0$ c.s. y, por tanto,

$$(9) \quad A\mathbf{z} + b = \bar{\mathbf{x}} + (B + P_{\mathbf{x},\mathbf{z}}P_{\mathbf{z}}^{\#})(\mathbf{z} - \bar{\mathbf{z}}) = E^{\ell}(\mathbf{x}|\mathbf{z}) \quad \text{c.s.} \quad \square$$

Si $\mathbf{z} = 0$, entonces toda A cumple $\overline{(\mathbf{x} - A\mathbf{z} - \bar{\mathbf{x}})\mathbf{z}^T} = 0$ y $\overline{\mathbf{x} - A\mathbf{z} - \bar{\mathbf{x}}} = 0$, por lo que se necesitan condiciones adicionales en \mathbf{z} para que A quede determinada.

La interpretación geométrica es la natural. Consideramos en $L_n^2(\Omega)$ el subespacio $\mathcal{S}(\mathbf{z}) = \{A\mathbf{z} + b \mid A \in \mathcal{M}_{n \times m}(\mathbb{R}), b \in \mathbb{R}^n\}$ de estimadores afines basados en \mathbf{z} . El estimador afín más cercano a \mathbf{x} (el de menor error cuadrático medio) es la proyección ortogonal de \mathbf{x} en $\mathcal{S}(\mathbf{z})$. Como

$$\mathcal{S}(\mathbf{z}) = \{A(\mathbf{z} - \bar{\mathbf{z}}) \mid A \in \mathcal{M}_{n \times m}(\mathbb{R})\} \perp \{b \mid b \in \mathbb{R}^n\}$$

y la proyección de \mathbf{x} en $\{b \mid b \in \mathbb{R}^n\}$ es $\bar{\mathbf{x}}$, entonces basta encontrar la proyección de $\mathbf{x} - \bar{\mathbf{x}}$ en $\{A(\mathbf{z} - \bar{\mathbf{z}}) \mid A \in \mathcal{M}_{n \times m}(\mathbb{R})\}$. Dado un A_0 válido, este cumple que, para todo A , $\langle \mathbf{x} - \bar{\mathbf{x}} - A_0(\mathbf{z} - \bar{\mathbf{z}}), A(\mathbf{z} - \bar{\mathbf{z}}) \rangle = 0$, o equivalentemente $P_{\mathbf{x},\mathbf{z}} - A_0P_{\mathbf{z}} = 0$. Esto ocurre si y solo si $A_0 = P_{\mathbf{x},\mathbf{z}}P_{\mathbf{z}}^{\#} + B$ con $B(\mathbf{z} - \bar{\mathbf{z}}) = 0$ c.s.

2. Filtro de Kalman discreto

Ampliamos el planteamiento del primer capítulo. Permitimos que el ruido tenga correlación, aparezca multiplicado por una matriz determinista y se le sume a la dinámica una influencia externa $\Gamma_k u_k$ conocida, que refleja el *control* ejercido al sistema. Es el único vector determinista; incluso el estado inicial \mathbf{x}_0 es aleatorio, pues puede haber incertidumbre sobre su valor. Nuestra exposición sigue [5, Sec. 5.4].

Las ecuaciones del sistema son:

1. *Modelo de la dinámica:*

$$(10) \quad \mathbf{x}_{k+1} = F_k \mathbf{x}_k + \Gamma_k u_k + G_k \mathbf{w}_k, \quad k \geq 0.$$

2. *Modelo de la medición:*

$$(11) \quad \mathbf{z}_k = H_k \mathbf{x}_k + \mathbf{v}_k, \quad k \geq 0.$$

Todas las hipótesis que suponemos quedan aquí recogidas:

1. Del estado inicial \mathbf{x}_0 conocemos su media $\bar{\mathbf{x}}_0$ y su matriz de covarianzas $P_{\mathbf{x}_0} = \text{Cov}(\mathbf{x}_0)$.
2. Del ruido \mathbf{w}_k y \mathbf{v}_k conocemos su media $\bar{\mathbf{w}}_k = 0$ y $\bar{\mathbf{v}}_k = 0$, y la matriz de covarianzas de su distribución conjunta

$$\text{Cov} \begin{pmatrix} \mathbf{w}_k \\ \mathbf{v}_l \end{pmatrix} = \begin{pmatrix} P_{\mathbf{w}_k, \mathbf{w}_l} & P_{\mathbf{w}_k, \mathbf{v}_l} \\ P_{\mathbf{v}_k, \mathbf{w}_l} & P_{\mathbf{v}_k, \mathbf{v}_l} \end{pmatrix} = \delta_{k,l} \begin{pmatrix} Q_k & S_k \\ S_k^T & R_k \end{pmatrix}.$$

3. R_k es invertible.
4. \mathbf{x}_0 es independiente de $(\mathbf{w}_k^T, \mathbf{v}_k^T)^T$.
5. $P_{\bar{\mathbf{z}}_k}$ es invertible, donde $\bar{\mathbf{z}}_k := \mathbf{z}_k - E^{\ell}(\mathbf{z}_k | \mathcal{Z}_{k-1})$ y consideramos las observaciones $\mathcal{Z}_k = (\mathbf{z}_0^T, \dots, \mathbf{z}_k^T)^T$ hasta el momento k , con $\mathcal{Z}_{-1} = \emptyset$.

Vector	Longitud	Aleatorio	Matriz	Orden	Aleatoria
\mathbf{x}_k	n	Sí	F_k	$n \times n$	No
\mathbf{u}_k	l	No	Γ_k	$n \times l$	No
\mathbf{w}_k	r	Sí	G_k	$n \times r$	No
\mathbf{z}_k	m	Sí	H_k	$m \times n$	No
\mathbf{v}_k	p	Sí			

CUADRO 1. Vectores y matrices del modelo.

Un posible filtro podría abordar el cálculo del MVE de \mathbf{x}_k basado en \mathcal{Z}_k , pero hemos de rebajar las expectativas y considerar el BAMVE $\hat{\mathbf{x}}_{k|k}$ en su lugar.

Usaremos la siguiente notación para los estimadores, errores y covarianzas involucrados en el siguiente teorema:

$$\begin{aligned}
\hat{\mathbf{x}}_{k|k} &= E^\ell(\mathbf{x}_k | \mathcal{Z}_k), \\
\tilde{\mathbf{x}}_{k|k} &= \mathbf{x}_k - \hat{\mathbf{x}}_{k|k}, \\
\mathbf{P}_{k|k} &= P_{\tilde{\mathbf{x}}_k} = \text{Cov}(\tilde{\mathbf{x}}_{k|k}), \\
\hat{\mathbf{x}}_{k+1|k} &= E^\ell(\mathbf{x}_{k+1} | \mathcal{Z}_k), \\
\tilde{\mathbf{x}}_{k+1|k} &= \tilde{\mathbf{x}}_{k+1} = \mathbf{x}_{k+1} - \hat{\mathbf{x}}_{k+1|k}, \\
\mathbf{P}_{k+1|k} &= P_{\tilde{\mathbf{x}}_{k+1|k}} = \text{Cov}(\tilde{\mathbf{x}}_{k+1|k}), \\
\hat{\mathbf{z}}_{k+1|k} &= E^\ell(\mathbf{z}_{k+1} | \mathcal{Z}_k), \\
\tilde{\mathbf{z}}_{k+1|k} &= \tilde{\mathbf{z}}_{k+1} = \mathbf{z}_{k+1} - \hat{\mathbf{z}}_{k+1|k}.
\end{aligned}$$

TEOREMA 4.8 (Algoritmo de Kalman). *El BAMVE $\hat{\mathbf{x}}_{k|k}$ para el sistema de (10) y (11) puede calcularse recursivamente con el siguiente procedimiento:*

1. Inicialización: $\hat{\mathbf{x}}_{0|0} := \bar{\mathbf{x}}_0$ y $\mathbf{P}_{0|0} := P_{\mathbf{x}_0}$.
2. Etapa de predicción:

$$\begin{aligned}
\hat{\mathbf{x}}_{k+1|k} &= (F_k - G_k S_k R_k^{-1} H_k) \hat{\mathbf{x}}_{k|k} + G_k S_k R_k^{-1} \mathbf{z}_k + \Gamma_k \mathbf{u}_k, \\
\mathbf{P}_{k+1|k} &= (F_k - G_k S_k R_k^{-1} H_k) \mathbf{P}_{k|k} (F_k - G_k S_k R_k^{-1} H_k)^T \\
&\quad + G_k (Q_k - S_k R_k^{-1} S_k^T) G_k^T.
\end{aligned}$$

3. Etapa de estimación (o actualización):

$$\begin{aligned}
\mathbf{K}_{k+1} &= \mathbf{P}_{k+1|k} H_{k+1}^T (H_{k+1} \mathbf{P}_{k+1|k} H_{k+1}^T + R_{k+1})^\#, \\
\hat{\mathbf{x}}_{k+1|k+1} &= \hat{\mathbf{x}}_{k+1|k} + \mathbf{K}_{k+1} (\mathbf{z}_{k+1} - H_{k+1} \hat{\mathbf{x}}_{k+1|k}), \\
\mathbf{P}_{k+1|k+1} &= (I - \mathbf{K}_{k+1} H_{k+1}) \mathbf{P}_{k+1|k}.
\end{aligned}$$

La matriz \mathbf{K}_{k+1} se llama ganancia de Kalman.

DEMOSTRACIÓN. En primer lugar observamos que de las hipótesis del modelo podemos deducir que para todo $k \in \mathbb{N}$,

$$(12) \quad P_{\mathbf{x}_l, \mathbf{w}_k} = 0, \quad P_{\mathbf{x}_l, \mathbf{v}_k} = 0, \quad l \in \{0, 1, \dots, k\}.$$

De igual modo,

$$(13) \quad P_{\mathbf{z}_l, \mathbf{w}_k} = 0, \quad P_{\mathbf{z}_l, \mathbf{v}_k} = 0, \quad l \in \{0, 1, \dots, k-1\}.$$

No descartamos, sin embargo, correlación entre \mathbf{z}_k y \mathbf{v}_k , ni entre \mathbf{z}_k y \mathbf{w}_k .

Vamos a calcular el estimador $\hat{\mathbf{x}}_{k+1|k}$ para \mathbf{x}_{k+1} basado en \mathcal{Z}_k . El principal problema en este cálculo es la correlación que pueden presentar \mathbf{w}_k y \mathbf{z}_k . Por eso vamos a cambiar \mathbf{w}_k por otra variable $\mathbf{w}_k^\circ := \mathbf{w}_k - B\mathbf{v}_k$ eligiendo B adecuadamente para que su covarianza se anule, es decir,

$$0 = \text{Cov}(\mathbf{w}_k - B\mathbf{v}_k, \mathbf{z}_k) = \text{Cov}(\mathbf{w}_k, \mathbf{z}_k) - B \text{Cov}(\mathbf{v}_k, \mathbf{z}_k)$$

$$= \text{Cov}(\mathbf{w}_k, \mathbf{v}_k) - B \text{Cov}(\mathbf{v}_k, \mathbf{v}_k) = S_k - BR_k.$$

Por tanto, basta tomar $B = S_k R_k^{-1}$. Despejando \mathbf{w}_k en esta expresión y \mathbf{v}_k en el modelo de la medición, obtenemos la expresión para la dinámica

$$(14) \quad \mathbf{x}_{k+1} = (F_k - G_k S_k R_k^{-1} H_k) \mathbf{x}_k + G_k \mathbf{w}_k^\circ + \Gamma_k \mathbf{u}_k + G_k S_k R_k^{-1} \mathbf{z}_k.$$

Usando la linealidad en la primera componente de E^ℓ obtenemos

$$\begin{aligned} \hat{\mathbf{x}}_{k+1|k} &= E^\ell(\mathbf{x}_{k+1} | \mathcal{Z}_k) \\ &= (F_k - G_k S_k R_k^{-1} H_k) \hat{\mathbf{x}}_{k|k} + G_k E^\ell(\mathbf{w}_k^\circ | \mathcal{Z}_k) + \Gamma_k \mathbf{u}_k + G_k S_k R_k^{-1} \mathbf{z}_k \\ &= (F_k - G_k S_k R_k^{-1} H_k) \hat{\mathbf{x}}_{k|k} + G_k S_k R_k^{-1} \mathbf{z}_k + \Gamma_k \mathbf{u}_k. \end{aligned}$$

ya que \mathbf{w}_k° y \mathcal{Z}_k son incorreladas, \mathbf{z}_k es constante cuando se condiciona respecto de \mathcal{Z}_k y la media de \mathbf{w}_k° es nula.

Ahora calculamos la estimación $\hat{\mathbf{x}}_{k+1|k+1}$ a partir de la predicción $\hat{\mathbf{x}}_{k+1|k}$. Como $\tilde{\mathcal{Z}}_{k+1} := \{\tilde{\mathbf{z}}_0, \dots, \tilde{\mathbf{z}}_{k+1}\}$ y \mathcal{Z}_{k+1} generan el mismo subespacio, $E^\ell(\mathbf{x}_{k+1} | \mathcal{Z}_{k+1}) = E^\ell(\mathbf{x}_{k+1} | \tilde{\mathcal{Z}}_{k+1})$. La ventaja de $\tilde{\mathcal{Z}}_{k+1}$ es que sus variables son incorreladas y con media nula. Así, gracias al teorema 4.6,

$$(15) \quad \begin{aligned} \hat{\mathbf{x}}_{k+1|k+1} &= E^\ell(\mathbf{x}_{k+1} | \mathcal{Z}_{k+1}) = E^\ell(\mathbf{x}_{k+1} | \tilde{\mathcal{Z}}_{k+1}) \\ &= E^\ell(\mathbf{x}_{k+1} | \tilde{\mathcal{Z}}_k) + E^\ell(\mathbf{x}_{k+1} | \tilde{\mathbf{z}}_{k+1}) - \bar{\mathbf{x}}_{k+1} \\ &= \hat{\mathbf{x}}_{k+1|k} + P_{\mathbf{x}_{k+1}, \tilde{\mathbf{z}}_{k+1}} P_{\tilde{\mathbf{z}}_{k+1}}^\# \tilde{\mathbf{z}}_{k+1}. \end{aligned}$$

Esto nos muestra que para deducir una fórmula para $\hat{\mathbf{x}}_{k+1|k+1}$ debemos desviarnos un momento y llevar cuenta de ciertas matrices de covarianzas. Partimos de

$$\tilde{\mathbf{z}}_{k+1} = \mathbf{z}_{k+1} - \hat{\mathbf{z}}_{k+1|k} = (H_{k+1} \mathbf{x}_{k+1} + \mathbf{v}_{k+1}) - H_{k+1} \hat{\mathbf{x}}_{k+1|k} = H_{k+1} \tilde{\mathbf{x}}_{k+1} + \mathbf{v}_{k+1}.$$

Como no hay correlación entre \mathbf{x}_{k+1} y \mathbf{v}_{k+1} ni entre \mathcal{Z}_k y \mathbf{v}_{k+1} , tampoco la hay entre $\tilde{\mathbf{x}}_{k+1}$ y \mathbf{v}_{k+1} , lo que a la vista de (2) nos dice que

$$(16) \quad P_{\tilde{\mathbf{z}}_{k+1}} = H_{k+1} P_{\tilde{\mathbf{x}}_{k+1}} H_{k+1}^T + R_{k+1}$$

y que, como $\text{Cov}(\tilde{\mathbf{x}}_{k+1}, \mathcal{Z}_k) = 0$,

$$(17) \quad P_{\mathbf{x}_{k+1}, \tilde{\mathbf{z}}_{k+1}} = P_{\mathbf{x}_{k+1}, H_{k+1} \tilde{\mathbf{x}}_{k+1} + \mathbf{v}_{k+1}} = P_{\mathbf{x}_{k+1}, \tilde{\mathbf{x}}_{k+1}} H_{k+1}^T = P_{\tilde{\mathbf{x}}_{k+1}} H_{k+1}^T.$$

Sustituyendo (16) y (17) en (15) llegamos a

$$(18) \quad \hat{\mathbf{x}}_{k+1|k+1} = \hat{\mathbf{x}}_{k+1|k} + P_{\tilde{\mathbf{x}}_{k+1}} H_{k+1}^T (H_{k+1} P_{\tilde{\mathbf{x}}_{k+1}} H_{k+1}^T + R_{k+1})^\# \tilde{\mathbf{z}}_{k+1}.$$

La ganancia de Kalman es la matriz

$$\mathbf{K}_{k+1} = P_{k+1|k} H_{k+1}^T (H_{k+1} P_{k+1|k} H_{k+1}^T + R_{k+1})^\#.$$

Como $\tilde{\mathbf{z}}_{k+1} = \mathbf{z}_{k+1} - H_{k+1} \hat{\mathbf{x}}_{k+1|k}$, podemos reescribir (18) como

$$\begin{aligned} \hat{\mathbf{x}}_{k+1|k+1} &= \hat{\mathbf{x}}_{k+1|k} + \mathbf{K}_{k+1} (\mathbf{z}_{k+1} - H_{k+1} \hat{\mathbf{x}}_{k+1|k}) \\ &= (I - \mathbf{K}_{k+1} H_{k+1|k}) \hat{\mathbf{x}}_{k+1|k} + \mathbf{K}_{k+1} \mathbf{z}_{k+1}. \end{aligned}$$

Para actualizar las covarianzas primero debemos obtener $\mathbf{P}_{k+1|k+1}$ a partir de $\mathbf{P}_{k+1|k}$. Puesto que $\mathbf{x}_{k+1} - \hat{\mathbf{x}}_{k+1|k} = (\mathbf{x}_{k+1} - \hat{\mathbf{x}}_{k+1|k+1}) + \mathbf{K}_{k+1} \tilde{\mathbf{z}}_{k+1}$ y ambos sumandos son incorrelados por la proposición 4.2, entonces, por (16),

$$\begin{aligned} \mathbf{P}_{k+1|k} &= \mathbf{P}_{k+1|k+1} + \mathbf{K}_{k+1} (H_{k+1} \mathbf{P}_{k+1|k} H_{k+1}^T + R_{k+1}) \mathbf{K}_{k+1}^T \\ &= \mathbf{P}_{k+1|k+1} + \mathbf{K}_{k+1} H_{k+1} P_{\tilde{\mathbf{x}}_{k+1}} = \mathbf{P}_{k+1|k+1} + \mathbf{K}_{k+1} H_{k+1} \mathbf{P}_{k+1|k}, \end{aligned}$$

donde en la penúltima igualdad hemos usado la propiedad (2) de la pseudoinversa y que la matriz $H_{k+1} \mathbf{P}_{k+1|k} H_{k+1}^T + R_{k+1}$ es simétrica. Así,

$$\mathbf{P}_{k+1|k+1} = (I - \mathbf{K}_{k+1} H_{k+1}) \mathbf{P}_{k+1|k}.$$

Solo queda encontrar la fórmula que expresa $\mathbf{P}_{k+1|k}$ en términos de $\mathbf{P}_{k|k}$. Volviendo a usar las variables \mathbf{w}_k° , a partir de (14) se tiene que

$$(19) \quad \mathbf{x}_{k+1} - \hat{\mathbf{x}}_{k+1|k} = (F_k - G_k S_k R_k^{-1} H_k)(\mathbf{x}_k - \hat{\mathbf{x}}_{k|k}) + G_k \mathbf{w}_k^\circ.$$

Tomando matrices de covarianzas,

$$\mathbf{P}_{k+1|k} = (F_k - G_k S_k R_k^{-1} H_k) \mathbf{P}_{k|k} (F_k - G_k S_k R_k^{-1} H_k)^T + G_k (Q_k - S_k R_k^{-1} S_k^T) G_k^T.$$

ya que $\text{Cov}(\mathbf{w}_k^\circ) = Q_k - S_k R_k^{-1} S_k^T$ y \mathbf{w}_k° no está correlada con \mathbf{x}_k ni con $\hat{\mathbf{x}}_{k|k}$. \square

Observamos que si $S_k = 0$, las fórmulas se simplifican notablemente.

EJEMPLO 4.1. Supongamos que \mathbf{s} es la posición de un tren en movimiento a lo largo de un carril. En nuestro modelo, su velocidad es constante y podemos controlar su aceleración usando unos frenos. En cada momento tan solo nos llegan mediciones de su posición. Podemos escribir el sistema como, para $k \geq 0$,

$$\begin{aligned} \begin{pmatrix} \mathbf{s}_{k+1} \\ \dot{\mathbf{s}}_{k+1} \end{pmatrix} &= \begin{pmatrix} 1 & \Delta t \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \mathbf{s}_k \\ \dot{\mathbf{s}}_k \end{pmatrix} + \begin{pmatrix} \frac{1}{2}(\Delta t)^2 \\ \Delta t \end{pmatrix} (\ddot{\mathbf{s}}_k) + \mathbf{w}_k, \\ \mathbf{z}_k &= \begin{pmatrix} 1 & 0 \end{pmatrix} \begin{pmatrix} \mathbf{s}_k \\ \dot{\mathbf{s}}_k \end{pmatrix} + \mathbf{v}_k. \end{aligned}$$

En este caso, la variable de estado es $\mathbf{x} = (\mathbf{s}, \dot{\mathbf{s}})^T$, compuesta por la posición del tren y su velocidad. El vector de control es $\mathbf{u} = (\ddot{\mathbf{s}}_k)$, compuesto por la aceleración, que es determinista. La medición es unidimensional. Tomamos $\Delta t = 1$ y no ejercemos control sobre el sistema.

Suponemos que en cada etapa se mantienen las covarianzas

$$Q = \begin{pmatrix} 50^2 & 0 \\ 0 & 20^2 \end{pmatrix}, \quad R = (200^2), \quad S = \begin{pmatrix} 0 \\ 0 \end{pmatrix}.$$

En la dinámica, la desviación estándar del ruido en la posición es de 50 unidades, y en la velocidad, de 20. Las componentes del ruido están incorreladas. En la medición, la desviación estándar del ruido en la posición es de 200 unidades. Esto es mucho ruido. En este caso, nos fiamos más de la dinámica que de las mediciones.

Para completar la información necesaria, tomamos la predicción inicial

$$\hat{\mathbf{x}}_0 = \begin{pmatrix} 0 \\ 100 \end{pmatrix}, \quad P_{\mathbf{x}_0} = \begin{pmatrix} 5^2 & 0 \\ 0 & 1^2 \end{pmatrix}.$$

Basta esta información para poner en marcha el filtro de Kalman. Podemos ver su ejecución en la figura 1. Se acerca más a la trayectoria real que las mediciones, pero no tira ningún dato a la basura; por muy ruidosas que sean, las tiene en cuenta en cada etapa e influyen su estimación.

COROLARIO 4.9. *Si las variables aleatorias \mathbf{w}_k y \mathbf{v}_k del teorema 4.8 son además gaussianas independientes y \mathbf{x}_0 es o bien determinista o bien gaussiano, entonces el MVE es gaussiano, satisface las relaciones del teorema y*

$$\begin{aligned} \mathbf{x}_{k+1} | \mathcal{Z}_k &\sim \mathcal{N}(\hat{\mathbf{x}}_{k+1|k}, \mathbf{P}_{k+1|k}), \\ \mathbf{x}_{k+1} | \mathcal{Z}_{k+1} &\sim \mathcal{N}(\hat{\mathbf{x}}_{k+1|k+1}, \mathbf{P}_{k+1|k+1}). \end{aligned}$$

En este sentido, puede verse una explicación intuitiva en [22].

Por limitaciones de espacio no podemos entrar en el análisis de la estabilidad del filtro de Kalman, pero durante la preparación de este trabajo hemos consultado la tesis [19] donde este aspecto se trata de un modo muy riguroso.

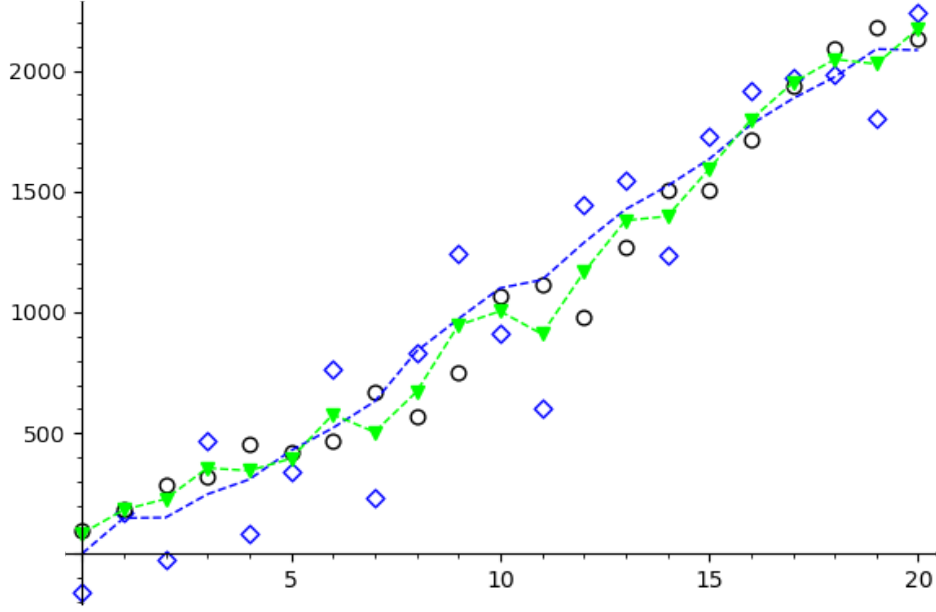


FIGURA 1. Posición s del ejemplo 4.1 actualizada para las primeras 20 etapas. Dinámica real en línea discontinua azul, mediciones en rombo azul, predicciones en círculo negro y estimaciones en triángulo verde unidas por línea discontinua.

3. Filtro de Kalman discreto extendido

En este caso el sistema es discreto, pero no necesariamente lineal:

$$\begin{aligned}\mathbf{x}_{k+1} &= f_k(\mathbf{x}_k, \mathbf{w}_k), \\ \mathbf{z}_k &= h_k(\mathbf{x}_k, \mathbf{v}_k).\end{aligned}$$

Esta vez las hipótesis son:

1. \mathbf{w}_k es ruido con estadística hasta segundo orden conocida (media nula y covarianza Q_k) y $E(\mathbf{w}_k \mathbf{x}_j^T) = 0$ si $j \in \{0, 1, \dots, k\}$.
2. \mathbf{v}_k es ruido con estadística hasta segundo orden conocida (media nula y covarianza R_k) y $E(\mathbf{v}_k \mathbf{w}_j^T) = 0$, $E(\mathbf{v}_k \mathbf{x}_j^T) = 0$ si $j \in \{0, 1, \dots, k\}$.
3. \mathbf{x}_0 es una variable aleatoria con estadística hasta segundo orden conocida (media $\bar{\mathbf{x}}_0$ y covarianza $P_{\mathbf{x}_0}$).

Por sencillez, nos restringimos al caso en que no hay correlación entre los ruidos. Esto corresponde a $S_k = 0$ en las fórmulas del filtro de Kalman discreto. Nuevamente el problema es encontrar el BAMVE para \mathbf{x}_k basado en \mathcal{Z}_k junto con la matriz de covarianzas del error de la estimación.

Tomamos una *trayectoria de referencia* o *trayectoria nominal*

$$\mathbf{x}_0^* \in \mathbb{R}^n, \quad \mathbf{x}_{k+1}^* = f_k(\mathbf{x}_k^*, 0), \quad k \geq 0.$$

Consideramos dos procesos asociados

$$\begin{aligned}\boldsymbol{\delta}_k &= \mathbf{x}_k - \mathbf{x}_k^*, \\ \boldsymbol{\nu}_k &= \mathbf{z}_k - h_k(\mathbf{x}_k^*, 0).\end{aligned}$$

$\boldsymbol{\delta}_k$ mide la desviación de la trayectoria real de la de referencia. Cuando la frecuencia de muestreo sea alta, quedará limitada a un pequeño entorno de 0. $\boldsymbol{\nu}_k$ nos da el desajuste entre la medición real y la de la trayectoria de referencia.

Aproximamos las funciones por la parte afín de su desarrollo de Taylor

$$\begin{aligned} f_k(\mathbf{x}_k, \mathbf{w}_k) &\approx f_k(\mathbf{x}_k^*, 0) + F_k(\mathbf{x}_k - \mathbf{x}_k^*) + G_k \mathbf{w}_k, \\ h_k(\mathbf{x}_k, \mathbf{v}_k) &\approx h_k(\mathbf{x}_k^*, 0) + H_k(\mathbf{x}_k - \mathbf{x}_k^*) + L_k \mathbf{v}_k, \end{aligned}$$

en el punto de la trayectoria de referencia, donde

$$\begin{aligned} F_k &= D_x|_{(\mathbf{x}_k^*, 0)} f_k, & G_k &= D_w|_{(\mathbf{x}_k^*, 0)} f_k, \\ H_k &= D_x|_{(\mathbf{x}_k^*, 0)} h_k, & L_k &= D_v|_{(\mathbf{x}_k^*, 0)} h_k. \end{aligned}$$

Esto nos proporciona una aproximación lineal para

$$(20) \quad \boldsymbol{\delta}_{k+1} \approx F_k \boldsymbol{\delta}_k + G_k \mathbf{w}_k,$$

$$(21) \quad \boldsymbol{\nu}_k \approx H_k \boldsymbol{\delta}_k + L_k \mathbf{v}_k.$$

Podemos aplicar el filtro de Kalman a este modelo aproximado para lograr una estimación del error $\hat{\boldsymbol{\delta}}_{k|k}$ junto con la matriz de covarianza $\mathbf{P}_{k|k}$ y considerar

$$\begin{aligned} \hat{\mathbf{x}}_{k+1|k} &= \mathbf{x}_{k+1}^* + \hat{\boldsymbol{\delta}}_{k+1|k}, \\ \hat{\mathbf{x}}_{k+1|k+1} &= \mathbf{x}_{k+1}^* + \hat{\boldsymbol{\delta}}_{k+1|k+1}. \end{aligned}$$

Así $\mathbf{x}_{k+1} - \hat{\mathbf{x}}_{k+1|k+1} = \boldsymbol{\delta}_{k+1} - \hat{\boldsymbol{\delta}}_{k+1|k+1}$ y $\mathbf{x}_{k+1} - \hat{\mathbf{x}}_{k+1|k} = \boldsymbol{\delta}_{k+1} - \hat{\boldsymbol{\delta}}_{k+1|k}$ tienen como matrices de covarianzas $\mathbf{P}_{k+1|k+1}$ y $\mathbf{P}_{k+1|k}$ respectivamente.

Aunque la trayectoria de referencia es útil para el planteamiento, queda determinada desde el inicio. Esto no tiene sentido porque el ruido la separará mucho de la real. Es preferible ir incorporando las predicciones y estimaciones según vayan llegando, tal y como hace el filtro de Kalman discreto extendido, que resumimos:

1. *Inicialización:* $\hat{\mathbf{x}}_{0|0} := \bar{\mathbf{x}}_0$ y $\mathbf{P}_{0|0} := \mathbf{P}_{\mathbf{x}_0}$.
2. *Etapa de predicción:*

$$\begin{aligned} F_k &= D_x|_{(\hat{\mathbf{x}}_{k|k}, 0)} f_k, \\ G_k &= D_w|_{(\hat{\mathbf{x}}_{k|k}, 0)} f_k, \\ \hat{\mathbf{x}}_{k+1|k} &= f_k(\hat{\mathbf{x}}_{k|k}, 0), \\ \mathbf{P}_{k+1|k} &= F_k \mathbf{P}_{k|k} F_k^T + G_k Q_k G_k^T. \end{aligned}$$

Notemos que como predicción se toma simplemente el punto que nos sugiere la nueva trayectoria de referencia recalculada a partir de t_k con las funciones linealizadas en la última estimación $(\hat{\mathbf{x}}_{k|k}, 0)$.

3. *Etapa de estimación (o actualización):*

$$\begin{aligned} H_{k+1} &= D_x|_{(\hat{\mathbf{x}}_{k+1|k}, 0)} h_{k+1}, \\ L_{k+1} &= D_v|_{(\hat{\mathbf{x}}_{k+1|k}, 0)} h_{k+1}, \\ \mathbf{K}_{k+1} &= \mathbf{P}_{k+1|k} H_{k+1}^T (H_{k+1} \mathbf{P}_{k+1|k} H_{k+1}^T + L_{k+1} R_{k+1} L_{k+1}^T)^{\#}, \\ \hat{\mathbf{x}}_{k+1|k+1} &= \hat{\mathbf{x}}_{k+1|k} + \mathbf{K}_{k+1} (\mathbf{z}_{k+1} - h_{k+1}(\hat{\mathbf{x}}_{k+1|k}, 0)), \\ \mathbf{P}_{k+1|k+1} &= (I - \mathbf{K}_{k+1} H_{k+1}) \mathbf{P}_{k+1|k}. \end{aligned}$$

Notemos que, puesto que se dispone de la predicción para el nuevo estado, la aplicación h_{k+1} que describe la medición esperada en el momento $k+1$ se linealiza en $(\hat{\mathbf{x}}_{k+1|k}, 0)$.

El desempeño del filtro de Kalman extendido depende de la covarianza del ruido, la inicialización y especialmente lo buena que sea la aproximación de Taylor.

4. Filtro de Kalman-Bucy

El filtro de Kalman-Bucy es la versión continua del filtro de Kalman. Como el tratamiento del ruido es mucho más complejo en este caso, conviene usar el lenguaje de los procesos de Wiener y las ecuaciones diferenciales estocásticas. El interés para el problema del siguiente capítulo se concentra únicamente en la ecuación diferencial de Riccati para la covarianza, recogida en el teorema 4.19.

4.1. Procesos de Wiener generales. Este apartado sigue las referencias [23, Cap. 4] y [18].

DEFINICIÓN 4.5. Un proceso estocástico gaussiano $\mathbf{W}(\cdot)$ es un *proceso de Wiener n -dimensional general* parametrizado por $[0, T]$ si

$$(22) \quad E(\mathbf{W}(t)) = 0, \quad \text{Cov}(\mathbf{W}(s), \mathbf{W}(t)) = \int_0^s \mathbf{B}(u) du, \quad 0 \leq s \leq t \leq T,$$

donde $\mathbf{B}: [0, T] \rightarrow \mathcal{M}_{n \times n}(\mathbb{R})$ es una aplicación continua tal que $\mathbf{B}(t)$ es simétrica y semidefinida positiva para todo $t \in T$.

Un proceso de Wiener estándar 1-dimensional corresponde a $\mathbf{B}(u) = 1$, para así recuperar $\text{Cov}(\mathbf{W}(s), \mathbf{W}(t)) = \min(s, t)$. En general, \mathbf{B} permite incorporar ruidos gaussianos no necesariamente blancos.

El siguiente resultado deja más claro el papel de los procesos generales.

TEOREMA 4.10. Sea $\mathbf{W}^0(\cdot)$ un proceso de Wiener estándar n -dimensional. Se tiene que $\mathbf{W}(\cdot)$ es un proceso de Wiener general n -dimensional si y solo si existe una aplicación $\mathbf{Q}: [0, T] \rightarrow \mathcal{M}_{n \times n}(\mathbb{R})$ continua tal que

$$\mathbf{W}(t) = \int_0^t \mathbf{Q}(u) d\mathbf{W}^0(u), \quad t \in [0, T].$$

Por tanto, es un proceso con diferencial estocástico $d\mathbf{W}(t) = \mathbf{Q}(t)d\mathbf{W}^0(t)$ y lo podemos usar en ecuaciones diferenciales estocásticas.

4.2. Sistemas dinámicos continuos. La versión continua de un sistema dinámico discreto es la solución de una ecuación diferencial estocástica

$$(23) \quad \begin{cases} d\mathbf{X} = (\mathbf{c}(t) + \mathbf{D}(t)\mathbf{X})dt + d\mathbf{W}, \\ \mathbf{X}(0) = \mathbf{X}_0. \end{cases}$$

Las observaciones también vienen dadas por una ecuación diferencial estocástica. En su versión integral, se considera una *observación parcial acumulativa*

$$(24) \quad \mathbf{Z}(t) = \int_0^t \mathbf{H}(u)\mathbf{X}(u)du + d\mathbf{W}^*, \quad 0 \leq t \leq T,$$

con $\mathbf{H}: [0, T] \rightarrow \mathcal{M}_{m \times n}(\mathbb{R})$ continua y $\mathbf{W}^*(\cdot)$ un proceso de Wiener m -dimensional independiente de $\mathbf{W}(\cdot)$.

DEFINICIÓN 4.6. La *parte estocástica* de un proceso estocástico $\mathbf{X}(\cdot)$ es

$$\mathbf{X}_e(t) := \mathbf{X}(t) - E(\mathbf{X}(t)),$$

mientras que $E(\mathbf{X}(t))$ es su *parte determinista*.

A partir de (23) se puede obtener una ecuación integral para \mathbf{X}_e .

TEOREMA 4.11. La *parte estocástica* \mathbf{X}_e de \mathbf{X} satisface que

$$\mathbf{X}_e(t) = \mathbf{X}_{e,0} + \left(\int_0^t \mathbf{D}(s)\mathbf{X}_e(s)ds \right) + \mathbf{W}(t), \quad t \in [0, T],$$

con $\mathbf{X}_{e,0} = \mathbf{X}_0 - E(\mathbf{X}_0)$. En este caso la solución es única c.s. y viene dada por

$$(25) \quad \begin{cases} \mathbf{X}_e(t) = \Phi(t) \left(\mathbf{X}_{e,0} + \int_0^t \Phi(s)^{-1} d\mathbf{W}(s) \right), \\ d\Phi(t) = \mathbf{D}(t)\Phi(t)dt, \\ \Phi(0) = I. \end{cases}$$

Además, la matriz $\text{Cov}(\mathbf{X}(s), \mathbf{X}(t)) = \text{Cov}(\mathbf{X}_e(s), \mathbf{X}_e(t))$ se puede expresar como

$$\text{Cov}(\mathbf{X}_e(s), \mathbf{X}_e(t)) = \Phi(s) \left(E(\mathbf{X}_{e,0}\mathbf{X}_{e,0}^T) + \int_0^s \Phi^{-1}(u)\mathbf{B}(u)(\Phi(u)^{-1})^T du \right) \Phi(s)^T$$

cuando $0 \leq s \leq t \leq T$. Por tanto, $\mathbf{L}(t) = \text{Cov}(\mathbf{X}_e(t), \mathbf{X}_e(t))$ es solución de

$$(26) \quad \begin{cases} \frac{d}{dt}\mathbf{L}(t) = \mathbf{D}(t)\mathbf{L}(t) + \mathbf{L}(t)\mathbf{D}(t)^T + \mathbf{B}(t), \\ \mathbf{L}(0) = E(\mathbf{X}_{e,0}\mathbf{X}_{e,0}^T). \end{cases}$$

La demostración se encuentra en [23, Teorema 3.5]. (25) es consecuencia inmediata del teorema 3.9 y (26) es sencilla de obtener, aunque se puede apelar al teorema 3.11; lo mismo para la covarianza si se usa la siguiente versión de la isometría de Itô.

TEOREMA 4.12. Sean $\mathbf{F}, \mathbf{G}: [0, t] \rightarrow \mathcal{M}_{n \times n}(\mathbb{R})$ aplicaciones continuas. Cuando $0 \leq s \leq t \leq T$, se tiene que

$$E \left(\int_0^s \mathbf{F}(u) d\mathbf{W}(u) \left(\int_0^t \mathbf{G}(v) d\mathbf{W}(v) \right)^T \right) = \int_0^s \mathbf{F}(u) \mathbf{B}(u) \mathbf{G}(u)^T du.$$

4.3. El estimador de Kalman-Bucy. En este capítulo, sea $H(\mathbf{Z}_e, t)$ la clausura¹ en $L^2(\Omega)$ del espacio vectorial generado por las componentes $Z_e^k(s)$ para $k \in \{1, 2, \dots, m\}$ y $s \in [0, t]$. De modo análogo al caso discreto, el *estimador de Kalman-Bucy* es el proceso $\hat{\mathbf{X}}_e = (\hat{X}_e^1, \dots, \hat{X}_e^n)^T$ que para todo $t \in [0, T]$ cumple

$$E((X_e^k(t) - \hat{X}_e^k(t))^2) = \min_{\Upsilon \in H(\mathbf{Z}_e, t)} E((X_e^k(t) - \Upsilon)^2), \quad k \in \{1, 2, \dots, n\}.$$

Añadiendo $E(\mathbf{X}(t))$ a $\hat{\mathbf{X}}_e(t)$ obtenemos la mejor aproximación por mínimos cuadrados de \mathbf{X} condicionada a las observaciones parciales acumuladas.

A partir de ahora prescindiremos del subíndice e , suponiendo medias nulas. Asumiremos que la matriz $\mathbf{B}^*(\cdot)$ asociada a \mathbf{W}^* es simétrica definida positiva para todo $t \in [0, T]$. Nuestro objetivo es describir $\hat{\mathbf{X}}$.

4.4. El estimador de Kalman-Bucy como proyección ortogonal. Primero debemos caracterizar los elementos de $H(\mathbf{Z}, t)$.

TEOREMA 4.13. $\Upsilon \in H(\mathbf{Z}, t)$ si y solo si existe una función $f: [0, t] \rightarrow \mathbb{R}^n$ en $L^2([0, t])$ tal que

$$\Upsilon = \int_0^t f(s)^T d\mathbf{Z}(s).$$

Además, dado $\Upsilon \in H(\mathbf{Z}, t)$, la función f es única c.t.p.

La demostración se encuentra en [23, Teorema 4.11].

Dado que la observación satisface (24), \mathbf{Z} es solución de la ecuación

$$d\mathbf{Z}(s) = \mathbf{H}(s)\mathbf{X}(s)ds + d\mathbf{W}^*(s),$$

por lo que, en términos de procesos de Wiener, la función $\Upsilon(t)$ del teorema 4.13 es

$$\int_0^t f(s)^T d\mathbf{Z}(s) = \int_0^t f(s)^T \mathbf{H}(s)\mathbf{X}(s)ds + \int_0^t f^T d\mathbf{W}^*(s).$$

¹Esta clausura no solo es lineal sino también topológica.

$H(\mathbf{Z}, t)$ es un subespacio cerrado de $L^2(\Omega)$. Como $\hat{X}^k(t)$ es la k -ésima componente de $\hat{\mathbf{X}}(t)$ si y solo si es la proyección ortogonal de $X^k(t)$ sobre $H(\mathbf{Z}, t)$, por el teorema 4.1, \hat{X}^k existe y es única. Lo resumimos en el siguiente resultado.

PROPOSICIÓN 4.14. *Tomamos el complemento ortogonal $H(\mathbf{Z}, t)^\perp$ en $L^2(\Omega)$.*

1. *El estimador de Kalman-Bucy $\hat{\mathbf{X}}$ para \mathbf{X} es único c.t.p.*
2. *El proceso $\hat{\mathbf{X}}$ es el estimador de Kalman-Bucy para \mathbf{X} si y solo si para todo $t \in [0, T]$ y $k \in \{1, 2, \dots, n\}$,*

$$\hat{X}_k(t) \in H(\mathbf{Z}, t), \quad X^k(t) - \hat{X}^k(t) \in H(\mathbf{Z}, t)^\perp.$$

Equivalentemente, $\Upsilon = \hat{X}^k(t)$ minimiza $\|X^k(t) - \Upsilon\|$ entre los $\Upsilon \in H(\mathbf{Z}, t)$. Por el teorema 4.13 y la ortogonalidad concluimos el siguiente resultado.

TEOREMA 4.15. *Un proceso estocástico $\hat{\mathbf{X}}$ es el estimador de Kalman-Bucy para \mathbf{X} si y solo si cumple que, para la observación parcial acumulativa \mathbf{Z} :*

1. *$\hat{\mathbf{X}}$ se puede expresar como*

$$\hat{\mathbf{X}}(t) = \int_0^t \mathbf{K}(t, s) d\mathbf{Z}(s), \quad t \in [0, T],$$

donde las entradas de \mathbf{K} están en $L^2([0, t])$.

2. *Se satisfacen las ecuaciones de Wiener-Hopf:*

$$E(\mathbf{X}(t)\mathbf{Z}(s)^T) = E(\hat{\mathbf{X}}(t)\mathbf{Z}(s)^T), \quad 0 \leq s \leq t \leq T.$$

TEOREMA 4.16. *Un proceso estocástico $\hat{\mathbf{X}}$ es el estimador de Kalman-Bucy si y solo si \mathbf{X} se puede escribir como*

$$\hat{\mathbf{X}}(t) = \int_0^t \mathbf{K}(t, u) d\mathbf{Z}(u),$$

donde las entradas de \mathbf{K} están en $L^2([0, t])$ y, tomando $\mathbf{R}(u) = \mathbf{H}(u)\mathbf{X}(u)$,

$$\mathbf{K}(t, s)\mathbf{B}^*(s) + \int_0^t \mathbf{K}(t, u) E(\mathbf{R}(u)\mathbf{R}(s)^T) du = E(\mathbf{X}(t)\mathbf{R}(s)^T), \quad 0 \leq s \leq t \leq T.$$

Hemos caracterizado la matriz \mathbf{K} que proporciona el estimador de Kalman-Bucy. El problema es que calcularla a partir de esta ecuación no suele ser viable. Por eso se adopta otro método similar al visto en el caso discreto, pero donde en vez de ser independientes, las fases de predicción y estimación están acopladas.

4.5. Cálculo del estimador de Kalman-Bucy. Definimos el *error* de la estimación como $\tilde{\mathbf{X}}(t) := \mathbf{X}(t) - \hat{\mathbf{X}}(t)$, y la *matriz de covarianzas del error* como

$$\mathbf{P}(t) = E(\tilde{\mathbf{X}}(t)\tilde{\mathbf{X}}(t)^T).$$

Para el cálculo de esta matriz puede usarse el proceso \mathbf{X} y su estimador $\hat{\mathbf{X}}$, con medias incluidas, en lugar de sus partes estocásticas. Como las componentes de $\tilde{\mathbf{X}}(t)$ son ortogonales a $H(\mathbf{Z}, t)$ y $\hat{\mathbf{X}}(t) \in H(\mathbf{Z}, t)$, entonces

$$\mathbf{P}(t) = E(\tilde{\mathbf{X}}(t)\mathbf{X}(t)^T).$$

Aparte de darnos información de la dispersión del error, determina a $\mathbf{K}(t, t)$.

TEOREMA 4.17. *Se tiene que $\mathbf{K}(t, t) = \mathbf{P}(t)\mathbf{H}(t)^T\mathbf{B}^*(t)^{-1}$.*

Con la matriz $\mathbf{K}(t, t)$ ya podemos calcular el estimador.

PROPOSICIÓN 4.18. *El estimador de Kalman-Bucy $\hat{\mathbf{X}}(t)$ cumple que*

$$\hat{\mathbf{X}}(t) = \int_0^t \mathbf{K}(s, s) d\mathbf{Z}(s) + \int_0^t (\mathbf{D}(u) - \mathbf{K}(u, u)\mathbf{H}(u)) \hat{\mathbf{X}}(u) du.$$

La expresión es similar a la que aparece en la fase de estimación del caso discreto. Todavía falta encontrar la matriz de covarianzas del error. Lo haremos resolviendo la siguiente ecuación diferencial.

TEOREMA 4.19. *La matriz de covarianzas $\mathbf{P}(t)$ es la solución de la ecuación diferencial de Riccati matricial*

$$\begin{cases} \frac{d}{dt}\mathbf{P}(t) = \mathbf{B}(t) + \mathbf{D}(t)\mathbf{P}(t) + \mathbf{P}(t)\mathbf{D}(t)^T - \mathbf{P}(t)\mathbf{H}(t)^T\mathbf{B}^*(t)^{-1}\mathbf{H}(t)\mathbf{P}(t), \\ \mathbf{P}(0) = E(\mathbf{X}_0\mathbf{X}_0^T). \end{cases}$$

5. Filtro de Kalman híbrido

En este apartado seguimos [3, Sec. 13.2.2]. En muchas aplicaciones prácticas se conoce una descripción continua de la dinámica, pero solo se dispone de mediciones en tiempo discreto. En tales casos se plantea una aproximación híbrida: la estimación se hace como en el filtro de Kalman discreto y la predicción, como en el filtro de Kalman-Bucy. Usaremos este método en el capítulo siguiente.

Se parte de un sistema dinámico con la formulación

$$d\mathbf{X} = (\mathbf{c}(t) + \mathbf{D}(t)\mathbf{X})dt + d\mathbf{W}.$$

Para calcular la matriz de covarianzas del error en el momento t_{k+1} , por el teorema 4.19, debemos resolver la ecuación diferencial

$$(27) \quad \frac{d}{dt}\mathbf{P} = \mathbf{B} + \mathbf{D}\mathbf{P} + \mathbf{P}\mathbf{D}^T, \quad t \in [t_k, t_{k+1})$$

con condición inicial $\mathbf{P}(t_k)$.

El término $\mathbf{P}(t)\mathbf{H}(t)^T\mathbf{B}^*(t)^{-1}\mathbf{H}(t)\mathbf{P}(t)$ no se ha tenido en cuenta ya que entre t_k y t_{k+1} no hay más que una medición, lo que puede interpretarse como que en ese intervalo $H(t) = 0$. La matriz \mathbf{B} proviene del movimiento de Wiener general que corresponde al ruido de la ecuación de la dinámica. La predicción del estado se realiza usando la ecuación de la dinámica sin ruido $d\mathbf{X} = (\mathbf{c}(t) + \mathbf{D}(t)\mathbf{X})dt$.

La medición \mathbf{Z}_k en tiempo t_k cumple que

$$\mathbf{Z}_k = \mathbf{H}_k\mathbf{X}_k + \mathbf{v}_k,$$

donde $\mathbf{X}_k = \mathbf{X}(t_k)$, \mathbf{v}_k es una sucesión de ruido gaussiano independiente de $\mathbf{W}(\cdot)$ y $\text{Cov}(\mathbf{v}_k, \mathbf{v}_l) = \delta_{k,l}R_k$. La estimación en tiempo t_{k+1} se realizará a partir de la predicción y de la estimación en tiempo t_k mediante las fórmulas del filtro discreto.

Seguimiento de la actitud

1. Actitud de un sólido

La *actitud* de un sólido en movimiento es su orientación con respecto a un sistema de referencia *inercial* (es decir, que no está acelerado). Su seguimiento es relevante para la navegación de vehículos de tierra, mar, aire o espacio.

Este capítulo expone uno de los métodos descritos en [24] para el seguimiento de la actitud de un sólido. Las observaciones las proporciona una IMU (unidad de medición inercial) que combinando acelerómetros (o magnetómetros) y giroscopios registra su orientación relativa al vector de gravedad y su velocidad angular. Asumimos un movimiento suave, de forma que entre observaciones la velocidad angular sea constante, tan solo afectada por perturbaciones accidentales que tratamos como ruido aleatorio. Estimamos la actitud con un filtro de Kalman híbrido.

Describimos la actitud mediante el álgebra de cuaternios, en particular la superficie esférica S^3 . Esta parametrización de las rotaciones tridimensionales es continua y carece de singularidades, pero plantea dos problemas:

1. Podemos definir variables aleatorias con valores en S^3 , que es un espacio topológico y por tanto medible. Sin embargo, no queda claro cómo extender el concepto de esperanza dentro de S^3 .
2. Las operaciones del filtro de Kalman no están disponibles en S^3 , que no está cerrado ni por suma ni por producto por escalares.

La solución adoptada en [24] es tratar S^3 como una variedad diferenciable y trabajar con cartas. Como S^3 es un grupo de Lie con el producto heredado de los cuaternios, basta considerar una carta con centro en el elemento neutro y trasladarla para obtener una carta con centro en cualquier otro elemento.

2. Cuaternios y rotaciones

El álgebra real de *cuaternios de Hamilton* consta del espacio vectorial

$$\mathbb{H} = \mathbb{R}1 \oplus \mathbb{R}\mathbf{i} \oplus \mathbb{R}\mathbf{j} \oplus \mathbb{R}\mathbf{k}$$

dotado de un producto \mathbb{R} -bilineal que queda determinado por la tabla de multiplicación del cuadro 1. Como álgebra asociativa unitaria, los productos entre elementos de la base se deducen de $\mathbf{i}^2 = \mathbf{j}^2 = \mathbf{k}^2 = \mathbf{ijk} = -1$.

Dado un cuaternio $\mathbf{x} = x_01 + x_1\mathbf{i} + x_2\mathbf{j} + x_3\mathbf{k}$ cualquiera, su *componente real* es $\mathbf{x}_s := \text{Re}(\mathbf{x}) := x_01$ y su *componente imaginaria* es $\mathbf{x}_v := \text{Im}(\mathbf{x}) := x_1\mathbf{i} + x_2\mathbf{j} + x_3\mathbf{k}$.

	1	\mathbf{i}	\mathbf{j}	\mathbf{k}
1	1	\mathbf{i}	\mathbf{j}	\mathbf{k}
\mathbf{i}	\mathbf{i}	-1	\mathbf{k}	$-\mathbf{j}$
\mathbf{j}	\mathbf{j}	$-\mathbf{k}$	-1	\mathbf{i}
\mathbf{k}	\mathbf{k}	\mathbf{j}	$-\mathbf{i}$	-1

CUADRO 1. Producto entre los elementos de la base de \mathbb{H} .

El subespacio

$$\mathbb{H}_0 := \mathbb{R}\mathbf{i} \oplus \mathbb{R}\mathbf{j} \oplus \mathbb{R}\mathbf{k}$$

es el de los *cuaternios imaginarios puros*.

En \mathbb{H} siempre consideraremos la base $(1, \mathbf{i}, \mathbf{j}, \mathbf{k})$, y en \mathbb{H}_0 , la base $(\mathbf{i}, \mathbf{j}, \mathbf{k})$. En ambos espacios tomamos el producto escalar usual $\langle \cdot, \cdot \rangle$ y el producto vectorial usual \times con respecto a ellas. Las dos bases serán ortonormadas, por lo que $x_0 = \langle \mathbf{x}, 1 \rangle$, $x_1 = \langle \mathbf{x}, \mathbf{i} \rangle$, $x_2 = \langle \mathbf{x}, \mathbf{j} \rangle$ y $x_3 = \langle \mathbf{x}, \mathbf{k} \rangle$.

Reducimos la notación identificando un vector $\mathbf{x} \in \mathbb{H}$ con sus coordenadas $(x_0, x_1, x_2, x_3)^T$ en $\mathbb{R}^4 \cong \mathbb{H}$, y un vector $\mathbf{x}_v \in \mathbb{H}_0$ con sus coordenadas $(x_1, x_2, x_3)^T$ en $\mathbb{R}^3 \cong \mathbb{H}_0$. Así, multiplicaremos libremente una matriz por un cuaternio.

Si tomamos otro cuaternio $\mathbf{y} = y_0 1 + y_1 \mathbf{i} + y_2 \mathbf{j} + y_3 \mathbf{k}$, el producto de ambos es

$$\begin{aligned} \mathbf{xy} &= (x_0 y_0 - x_1 y_1 - x_2 y_2 - x_3 y_3) 1 \\ &\quad + (x_0 y_1 + x_1 y_0 + x_2 y_3 - x_3 y_2) \mathbf{i} \\ &\quad + (x_0 y_2 - x_1 y_3 + x_2 y_0 + x_3 y_1) \mathbf{j} \\ &\quad + (x_0 y_3 + x_1 y_2 - x_2 y_1 + x_3 y_0) \mathbf{k}. \end{aligned}$$

Queda claro que el producto a izquierda por \mathbf{x} y el producto a derecha por \mathbf{y} pueden expresarse como un producto matricial respecto de la base $(\mathbf{i}, \mathbf{j}, \mathbf{k})$. Alternativamente, podemos descomponer el producto como

$$(28) \quad \mathbf{xy} = (\mathbf{x}_s \mathbf{y}_s - \langle \mathbf{x}_v, \mathbf{y}_v \rangle) 1 + (\mathbf{x}_s \mathbf{y}_v + \mathbf{y}_s \mathbf{x}_v + \mathbf{x}_v \times \mathbf{y}_v).$$

El producto vectorial es *anticonmutativo*, es decir, para todo $\mathbf{x}_v, \mathbf{y}_v \in \mathbb{H}_0$ se cumple que $\mathbf{x}_v \times \mathbf{y}_v = -\mathbf{y}_v \times \mathbf{x}_v$. Dado $\mathbf{x}_v \in \mathbb{H}_0$, la matriz coordenada del endomorfismo lineal $\mathbf{y}_v \mapsto \mathbf{x}_v \times \mathbf{y}_v$ en \mathbb{H}_0 respecto de la base $(\mathbf{i}, \mathbf{j}, \mathbf{k})$ es

$$[\mathbf{x}_v]_{\times} := \begin{pmatrix} 0 & -x_3 & x_2 \\ x_3 & 0 & -x_1 \\ -x_2 & x_1 & 0 \end{pmatrix},$$

y el de $\mathbf{x}_v \mapsto \mathbf{x}_v \times \mathbf{y}_v$ es $-[\mathbf{y}_v]$. El *conmutador* de dos cuaternios es $[\mathbf{x}, \mathbf{y}] := \mathbf{xy} - \mathbf{yx}$. Como en (28) tan solo el producto vectorial no conmuta, se tiene que

$$\frac{1}{2}[\mathbf{x}, \mathbf{y}] = \mathbf{x}_v \times \mathbf{y}_v.$$

Esta fórmula permite extender el producto vectorial a \mathbb{H} haciendo que

$$1 \times \mathbf{x} = \mathbf{x} \times 1 = 0, \quad \mathbf{x} \in \mathbb{H}.$$

El *centro* del álgebra de cuaternios es

$$Z(\mathbb{H}) := \{\mathbf{x} \in \mathbb{H} \mid \forall \mathbf{y} \in \mathbb{H}, \mathbf{xy} = \mathbf{yx}\}.$$

PROPOSICIÓN 5.1. $Z(\mathbb{H}) = \mathbb{R}1$.

DEMOSTRACIÓN. Sea $\mathbf{x} \in Z(\mathbb{H})$. Usamos que $(1, \mathbf{i}, \mathbf{j}, \mathbf{k})$ es una base de \mathbb{H} . Como $\mathbf{i}\mathbf{x} = \mathbf{x}\mathbf{i}$, necesariamente $x_2 = x_3 = 0$. Como $\mathbf{j}\mathbf{x} = \mathbf{x}\mathbf{j}$, necesariamente $x_1 = 0$. Por otro lado, todos los reales están en el centro. \square

El *conjugado* de un cuaternio \mathbf{x} es

$$\bar{\mathbf{x}} := x_0 1 - x_1 \mathbf{i} - x_2 \mathbf{j} - x_3 \mathbf{k}.$$

PROPOSICIÓN 5.2. La conjugación es una involución, esto es, una aplicación lineal que para todo $\mathbf{x}, \mathbf{y} \in \mathbb{H}$ cumple que $\bar{\bar{\mathbf{x}}} = \mathbf{x}$ y que $\overline{\mathbf{xy}} = \bar{\mathbf{y}}\bar{\mathbf{x}}$.

DEMOSTRACIÓN. La linealidad y la primera propiedad se tienen directamente. Basta probar la segunda para los elementos de la base. Si los dos son iguales o alguno es 1, es trivial. Si no, se sigue por la tabla del producto. \square

Asociadas a cualquier cuaternio \mathbf{x} podemos definir su *traza* $t(\mathbf{x})$ y una forma cuadrática $n(\mathbf{x})$ (a menudo llamada *norma*, lo que puede causar ambigüedad):

$$t(\mathbf{x}) = 2\langle 1, \mathbf{x} \rangle = 2x_0, \quad n(\mathbf{x}) := \|\mathbf{x}\|^2 = \langle \mathbf{x}, \mathbf{x} \rangle = x_0^2 + x_1^2 + x_2^2 + x_3^2.$$

PROPOSICIÓN 5.3. *Dados $\mathbf{x}, \mathbf{y}, \mathbf{z} \in \mathbb{H}$, se tiene que:*

1. $t(\mathbf{x}) = \mathbf{x} + \bar{\mathbf{x}}$.
2. $n(\mathbf{x}) = \mathbf{x}\bar{\mathbf{x}} = \bar{\mathbf{x}}\mathbf{x}$.
3. $n(\mathbf{xy}) = n(\mathbf{x})n(\mathbf{y})$.
4. Si $\mathbf{x} \neq 0$, entonces es invertible $\mathbf{y} \mathbf{x}^{-1} = \frac{1}{n(\mathbf{x})}\bar{\mathbf{x}}$. Además, $n(\mathbf{x}^{-1}) = \frac{1}{n(\mathbf{x})}$.
5. $\mathbf{x}^2 - t(\mathbf{x})\mathbf{x} + n(\mathbf{x}) = 0$.
6. $\langle \mathbf{x}, \mathbf{y} \rangle = \frac{1}{2}(\bar{\mathbf{x}}\mathbf{y} + \bar{\mathbf{y}}\mathbf{x}) = \frac{1}{2}(\mathbf{x}\bar{\mathbf{y}} + \mathbf{y}\bar{\mathbf{x}})$.
7. $\langle \bar{\mathbf{x}}, \bar{\mathbf{y}} \rangle = \langle \mathbf{x}, \mathbf{y} \rangle$.
8. $\langle \mathbf{xz}, \mathbf{y} \rangle = \langle \mathbf{x}, \mathbf{y}\bar{\mathbf{z}} \rangle$, $\langle \mathbf{zx}, \mathbf{y} \rangle = \langle \mathbf{x}, \bar{\mathbf{z}}\mathbf{y} \rangle$.
9. $\langle \mathbf{xy}, \mathbf{xz} \rangle = n(\mathbf{x})\langle \mathbf{y}, \mathbf{z} \rangle = \langle \mathbf{yx}, \mathbf{zx} \rangle$.

DEMOSTRACIÓN. (1) es directa. Para (2), descomponemos $\mathbf{x} = \mathbf{x}_s + \mathbf{x}_v$, por lo que $\bar{\mathbf{x}} = \mathbf{x}_s - \mathbf{x}_v$. Como $\mathbf{x}_s \in Z(\mathbb{H})$, \mathbf{x} y $\bar{\mathbf{x}}$ conmutan. Además,

$$\begin{aligned} \mathbf{x}\bar{\mathbf{x}} &= \mathbf{x}_s^2 - \mathbf{x}_v^2 \\ &= x_0^2 + x_1^2 + x_2^2 + x_3^2 + x_1x_2(\mathbf{ij} + \mathbf{ji}) + x_2x_3(\mathbf{jk} + \mathbf{kj}) + x_1x_3(\mathbf{ki} + \mathbf{ik}) \end{aligned}$$

y los últimos tres sumandos se anulan. Con esto, es fácil ver (3) y (5):

$$n(\mathbf{xy}) = \mathbf{xy}\bar{\mathbf{xy}} = \mathbf{xy}\bar{\mathbf{y}}\bar{\mathbf{x}} = n(\mathbf{y})\mathbf{x}\bar{\mathbf{x}} = n(\mathbf{x})n(\mathbf{y}),$$

$$\mathbf{x}^2 - t(\mathbf{x})\mathbf{x} + n(\mathbf{x}) = (\mathbf{x} - t(\mathbf{x}))\mathbf{x} + n(\mathbf{x}) = -\bar{\mathbf{x}}\mathbf{x} + \bar{\mathbf{x}}\mathbf{x} = 0.$$

Si $\mathbf{x} \neq 0$, entonces $n(\mathbf{x}) \neq 0$ y dividiendo en la ecuación tenemos (4). Además, $1 = n(1) = n(\mathbf{xx}^{-1}) = n(\mathbf{x})n(\mathbf{x}^{-1})$. Como el producto escalar usual es la forma bilineal asociada a la forma cuadrática n , a partir de la relación $\langle \mathbf{x}, \mathbf{y} \rangle = \frac{1}{2}(n(\mathbf{x} + \mathbf{y}) - n(\mathbf{x}) - n(\mathbf{y}))$ podemos obtener que

$$\begin{aligned} \langle \mathbf{x}, \mathbf{y} \rangle &= \frac{1}{2} \left((\mathbf{x} + \mathbf{y})\overline{(\mathbf{x} + \mathbf{y})} - \mathbf{x}\bar{\mathbf{x}} - \mathbf{y}\bar{\mathbf{y}} \right) = \frac{1}{2}(\bar{\mathbf{x}}\mathbf{y} + \bar{\mathbf{y}}\mathbf{x}), \\ \langle \mathbf{x}, \mathbf{y} \rangle &= \frac{1}{2} \left(\overline{(\mathbf{x} + \mathbf{y})}(\mathbf{x} + \mathbf{y}) - \bar{\mathbf{x}}\mathbf{x} - \bar{\mathbf{y}}\mathbf{y} \right) = \frac{1}{2}(\mathbf{x}\bar{\mathbf{y}} + \mathbf{y}\bar{\mathbf{x}}), \end{aligned}$$

y se tiene (6). Con esta expresión, es fácil ver (7) y (8):

$$\langle \bar{\mathbf{x}}, \bar{\mathbf{y}} \rangle = \frac{1}{2}(\bar{\bar{\mathbf{x}}}\bar{\bar{\mathbf{y}}} + \bar{\bar{\mathbf{y}}}\bar{\bar{\mathbf{x}}}) = \frac{1}{2}(\mathbf{x}\bar{\mathbf{y}} + \bar{\mathbf{y}}\mathbf{x}) = \langle \mathbf{x}, \mathbf{y} \rangle,$$

$$\langle \mathbf{xz}, \mathbf{y} \rangle = \frac{1}{2}(\mathbf{xz}\bar{\mathbf{y}} + \mathbf{y}\bar{\mathbf{xz}}) = \frac{1}{2}(\mathbf{x}\bar{\mathbf{y}}\bar{\mathbf{z}} + \mathbf{y}\bar{\mathbf{z}}\bar{\mathbf{x}}) = \langle \mathbf{x}, \mathbf{y}\bar{\mathbf{z}} \rangle,$$

donde la segunda igualdad de (8) es análoga, y (9) se deduce de ambas. \square

Notemos que si $n(\mathbf{x}) = 1$, entonces $\mathbf{x}^{-1} = \bar{\mathbf{x}}$.

La *adjunción* (o *aplicación adjunta*) por $\mathbf{x} \neq 0$ es la aplicación

$$\begin{aligned} \text{Ad}_{\mathbf{x}}: \mathbb{H} &\rightarrow \mathbb{H} \\ \mathbf{y} &\mapsto \mathbf{xyx}^{-1}. \end{aligned}$$

PROPOSICIÓN 5.4. *Para cualquier $0 \neq \mathbf{x} \in \mathbb{H}$, la aplicación $\text{Ad}_{\mathbf{x}}: \mathbb{H}_0 \rightarrow \mathbb{H}_0$ induce, por restricción, una isometría.*

DEMOSTRACIÓN. Claramente $\text{Ad}_{\mathbf{x}}$ es inyectiva. Veamos que $\text{Ad}_{\mathbf{x}}(\mathbb{H}_0) = \mathbb{H}_0$, por lo que la aplicación está bien definida y es suprayectiva. Por la relación

$$\langle \text{Ad}_{\mathbf{x}}(\mathbf{y}), 1 \rangle = \langle \mathbf{xyx}^{-1}, \mathbf{xx}^{-1} \rangle = n(\mathbf{x})n(\mathbf{x}^{-1})\langle \mathbf{y}, 1 \rangle = \langle \mathbf{y}, 1 \rangle,$$

tenemos que $\mathbf{y} \in \mathbb{H}_0$ si y solo si $\text{Ad}_{\mathbf{x}}(\mathbf{y}) \in \mathbb{H}_0$. Si $\mathbf{y} \in \mathbb{H}_0$, tenemos que

$$n(\text{Ad}_{\mathbf{x}}(\mathbf{y})) = n(\mathbf{xyx}^{-1}) = n(\mathbf{x})n(\mathbf{y})n(\mathbf{x}^{-1}) = n(\mathbf{y}),$$

por lo que $\text{Ad}_{\mathbf{x}}$ es una isometría. \square

De aquí en adelante siempre trabajaremos con la restricción de $\text{Ad}_{\mathbf{x}}$ a \mathbb{H}_0 .

En cualquier caso, $\text{Ad}_{\mathbf{q}_1} = \text{Ad}_{\mathbf{q}_2}$ si y solo si $\mathbf{q}_1 = r\mathbf{q}_2$ para cierto $r \in \mathbb{R}$. Cambiando \mathbf{q}_1 por $(\sqrt{n(\mathbf{q}_1)})^{-1}\mathbf{q}_1$, que tiene norma 1, para operar con estas isometrías podemos restringirnos a los cuaternios de norma 1, que forman la 3-esfera.

TEOREMA 5.5. Sean $S^3 := \{\mathbf{q} \in \mathbb{H} \mid n(\mathbf{q}) = 1\}$ la 3-esfera en \mathbb{H} y $\mathbf{q}_1, \mathbf{q}_2 \in S^3$:

1. $\text{Ad}_{\mathbf{q}_1} = \text{Ad}_{\mathbf{q}_2}$ si y solo si $\mathbf{q}_1 = \pm\mathbf{q}_2$.
2. $\text{Ad}_{\mathbf{q}_1} \circ \text{Ad}_{\mathbf{q}_2} = \text{Ad}_{\mathbf{q}_1\mathbf{q}_2}$.
3. Si $\mathbf{q} \in S^3 \setminus \{\pm 1\}$, este se puede escribir de forma única como

$$\mathbf{q} = \cos\left(\frac{\alpha}{2}\right) + \sin\left(\frac{\alpha}{2}\right)\mathbf{u}$$

con $\alpha \in (0, 2\pi)$ y $\mathbf{u} \in \mathbb{H}_0$ un cuaternio imaginario de norma 1. Además $\text{Ad}_{\mathbf{q}}$ es una rotación de eje $\mathbb{R}\mathbf{u}$ con ángulo α en orientación positiva¹.

4. La aplicación

$$\begin{aligned} \text{Ad}: S^3 &\rightarrow \text{SO}(\mathbb{H}_0) \\ \mathbf{q} &\mapsto \text{Ad}_{\mathbf{q}} \end{aligned}$$

es un epimorfismo de grupos con núcleo $\{\pm 1\}$.

DEMOSTRACIÓN. Para (1), sabemos que $\text{Ad}_{\mathbf{q}_1} = \text{Ad}_{\mathbf{q}_2}$ si y solo si $\mathbf{q}_1\mathbf{q}_2^{-1} = r$ para algún $r \in \mathbb{R}$. Como $n(r) = 1$, entonces $r = \pm 1$, y ambos sirven.

Para (2), sea un $\mathbf{x} \in \mathbb{H}_0$ arbitrario. Entonces,

$$\text{Ad}_{\mathbf{q}_1} \circ \text{Ad}_{\mathbf{q}_2}(\mathbf{x}) = \mathbf{q}_1\mathbf{q}_2\mathbf{x}\mathbf{q}_2^{-1}\mathbf{q}_1^{-1} = (\mathbf{q}_1\mathbf{q}_2)\mathbf{x}(\mathbf{q}_1\mathbf{q}_2)^{-1} = \text{Ad}_{\mathbf{q}_1\mathbf{q}_2}(\mathbf{x}).$$

Para (3), como $\mathbf{q}_v \neq 0$, podemos escribir $\mathbf{q} = \lambda_0 + \lambda_1\mathbf{u}$, con $\lambda_0 = \mathbf{q}_s \in \mathbb{R}1$, $\lambda_1 = \sqrt{n(\mathbf{q}_v)} > 0$ y $\mathbf{u} = (\sqrt{n(\mathbf{q}_v)})^{-1}\mathbf{q}_v$ imaginario de norma 1. Como

$$1 = n(\mathbf{q}) = n(\mathbf{q}_s) + n(\mathbf{q}_v) = n(\lambda_0) + n(\lambda_1\mathbf{u}) = \lambda_0^2 + \lambda_1^2,$$

existe cierto $\alpha \in (0, 2\pi)$ tal que $\cos(\frac{\alpha}{2}) = \lambda_0$ y $\sin(\frac{\alpha}{2}) = \lambda_1$. Este está unívocamente determinado y con él la parte imaginaria y \mathbf{u} . Como \mathbf{q} y \mathbf{u} conmutan, dado $\lambda\mathbf{u}$ con $\lambda \in \mathbb{R}$, $\text{Ad}_{\mathbf{q}}(\lambda\mathbf{u}) = \lambda\mathbf{u}$. Así, $\text{Ad}_{\mathbf{q}}$ fija a todo vector de $\mathbb{R}\mathbf{u}$.

Completemos \mathbf{u} hasta una base ortonormada $(\mathbf{u}, \mathbf{v}, \mathbf{w})$ de \mathbb{H}_0 , donde $\mathbf{w} = \mathbf{u} \times \mathbf{v} = \mathbf{u}\mathbf{v}$. Teniendo en cuenta que $0 = 2\langle \mathbf{u}, \mathbf{v} \rangle = \bar{\mathbf{u}}\mathbf{v} + \bar{\mathbf{v}}\mathbf{u} = -(\mathbf{u}\mathbf{v} + \mathbf{v}\mathbf{u})$, y que al ser imaginarios de norma uno, $0 = (\mathbf{u}\mathbf{v})^2 + t(\mathbf{u}\mathbf{v})\mathbf{u}\mathbf{v} + n(\mathbf{u}\mathbf{v}) = (\mathbf{u}\mathbf{v})^2 + 0 + 1$ y entonces se cumple que $\mathbf{v} = \mathbf{u}\mathbf{v}\mathbf{u}$. Así,

$$\begin{aligned} \text{Ad}_{\mathbf{q}}(\mathbf{v}) &= \left(\cos\left(\frac{\alpha}{2}\right) + \sin\left(\frac{\alpha}{2}\right)\mathbf{u}\right)\mathbf{v}\left(\cos\left(\frac{\alpha}{2}\right) - \sin\left(\frac{\alpha}{2}\right)\mathbf{u}\right) \\ &= \cos^2\left(\frac{\alpha}{2}\right)\mathbf{v} + \cos\left(\frac{\alpha}{2}\right)\sin\left(\frac{\alpha}{2}\right)(\mathbf{u}\mathbf{v} - \mathbf{v}\mathbf{u}) - \sin^2\left(\frac{\alpha}{2}\right)\mathbf{u}\mathbf{v}\mathbf{u} \\ &= \left(\cos^2\left(\frac{\alpha}{2}\right) - \sin^2\left(\frac{\alpha}{2}\right)\right)\mathbf{v} + \left(2\cos\left(\frac{\alpha}{2}\right)\sin\left(\frac{\alpha}{2}\right)\right)\mathbf{u}\mathbf{v} = \cos(\alpha)\mathbf{v} + \sin(\alpha)\mathbf{w}. \end{aligned}$$

Del mismo modo, como $-\mathbf{v} = \mathbf{u} \times \mathbf{w} = \mathbf{u}\mathbf{w}$, $\text{Ad}_{\mathbf{q}}(\mathbf{v}) = -\sin(\alpha)\mathbf{v} + \cos(\alpha)\mathbf{w}$. Por tanto, el movimiento es el descrito. Los resultados anteriores se resumen en (4). \square

EJEMPLO 5.1. Para representar el giro en \mathbb{R}^3 de 120° según la recta que pasa por el origen y el punto $(1, 1, 1) \equiv \mathbf{i} + \mathbf{j} + \mathbf{k}$ usamos el cuaternio

$$\mathbf{q} = \cos\left(\frac{\pi}{3}\right) + \sin\left(\frac{\pi}{3}\right)\frac{\mathbf{i} + \mathbf{j} + \mathbf{k}}{\sqrt{3}} = \frac{1}{2}(1 + \mathbf{i} + \mathbf{j} + \mathbf{k})$$

Comprobamos, por ejemplo, que $\text{Ad}_{\mathbf{q}}(\mathbf{i}) = \mathbf{j}$, $\text{Ad}_{\mathbf{q}}(\mathbf{j}) = \mathbf{k}$ y $\text{Ad}_{\mathbf{q}}(\mathbf{k}) = \mathbf{i}$.

Hemos visto que para componer rotaciones basta con multiplicar los cuaternios que las representan. Esto agiliza las operaciones con muchas rotaciones.

¹Siguiendo la mano derecha, respecto de la base $(\mathbf{i}, \mathbf{j}, \mathbf{k})$. Si $\mathbf{q} = \pm 1$, $\alpha = 0$ y \mathbf{u} no es único.

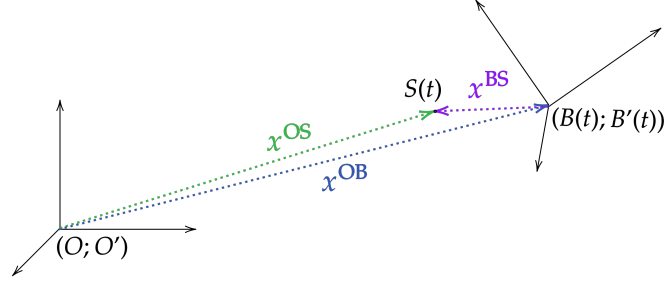


FIGURA 1. Situación del modelo.

3. Velocidad angular

En esta sección trabajamos con la geometría afín en $\mathbb{H}_0 \cong \mathbb{R}^3$. Dado un sistema de referencia $(C; C')$ y un vector v , denotamos por ${}^C v$ sus coordenadas respecto de la base C' . Dados dos puntos A y B , denotamos por x^{AB} el vector que los une. En particular, sus coordenadas respecto de la base C' son ${}^C x^{AB}$.

Consideramos un sistema de referencia euclídeo $(O; O')$ que permanece inmóvil; es el inercial. Si además consideramos un sólido rígido en movimiento con centro de masas $B(t)$ y en él anclamos una base ortonormada $B'(t)$, tenemos otro sistema de referencia euclídeo $(B(t); B'(t))$ en cada instante $t \in \mathbb{R}$.

Si en $t = 0$ ambas bases coinciden, entonces están relacionadas en cada instante mediante una rotación. Así, existe un cuaternio $\mathbf{q}(t) \in S^3$ de modo que

$$B'(t) = \text{Ad}_{\mathbf{q}(t)}(O').$$

Este describe la orientación de $B'(t)$ respecto a O' . La matriz coordenada $R(\mathbf{q}(t))$ de la rotación $\text{Ad}_{\mathbf{q}}$ con la base O' en el espacio de llegada y salida coincide con la de cambio de coordenadas de la base $B'(t)$ a O' . Para todo vector v ,

$${}^O v = R(\mathbf{q})^B v,$$

donde a partir de ahora denotaremos $B = B(t)$. Si $S = S(t)$ es un punto del sólido, su movimiento respecto del origen se descompone como

$$x^{OS} = x^{OB} + x^{BS}.$$

Nos es irrelevante x^{OB} , relativo al movimiento del centro de masas. Nuestro interés reside en x^{BS} , relativo al movimiento del punto. Como los puntos del sólido se mueven junto con su base, las coordenadas $v := {}^B x^{BS} \in \mathbb{H}_0$ son constantes. Las que cambian son ${}^O x^{BS} = R(\mathbf{q})^B x^{BS}$, las relativas al origen, ya que \mathbf{q} cambia.

Supongamos que el cuaternio $\mathbf{q} : \mathbb{R} \rightarrow S^3$ que relaciona las bases describe una curva derivable, y además al empezar las bases coinciden, es decir, $\mathbf{q}(0) = 1$. Si en $t = 0$ trasladamos el vector x^{BS} al punto O , entonces

$$\begin{aligned} f : \mathbb{H}_0 &\rightarrow \mathbb{H}_0 \\ t &\mapsto f(t) := \text{Ad}_{\mathbf{q}(t)}(v) \end{aligned}$$

refleja la rotación de S respecto de B vista desde O , empezando en $f(0) = v$. Notemos que en todo momento ${}^O f = {}^O x^{BS}$.

Como $\bar{\mathbf{q}}\mathbf{q} = 1$, derivando, $\bar{\dot{\mathbf{q}}}\mathbf{q} + \bar{\mathbf{q}}\dot{\mathbf{q}} = 0$, por la bilinearidad del producto. Definimos la *velocidad angular* o *de rotación* como $\boldsymbol{\omega} := 2\bar{\mathbf{q}}\dot{\mathbf{q}}$, pues

$$\dot{f} = \dot{\mathbf{q}}v\bar{\mathbf{q}} + \mathbf{q}v\bar{\dot{\mathbf{q}}} = \mathbf{q}(\bar{\mathbf{q}}\dot{\mathbf{q}}v + v\bar{\dot{\mathbf{q}}}\bar{\mathbf{q}}) = \text{Ad}_{\mathbf{q}}(\bar{\mathbf{q}}\dot{\mathbf{q}}v - v\bar{\dot{\mathbf{q}}}\bar{\mathbf{q}}) = \text{Ad}_{\mathbf{q}}(\boldsymbol{\omega} \times v).$$

Debe cumplirse la ecuación diferencial $2\dot{\mathbf{q}} = \boldsymbol{\omega}\mathbf{q}$. Cuando $\boldsymbol{\omega}$ es una constante no nula, $\mathbf{q}(t) = e^{\frac{1}{2}t\boldsymbol{\omega}}$ es su solución para la condición inicial $\mathbf{q}(0) = 1$. En este caso, f describe la rotación de v respecto del eje $\frac{\boldsymbol{\omega}}{\|\boldsymbol{\omega}\|}$ con ángulo $t\|\boldsymbol{\omega}\|$; de ahí el nombre.

4. Modelo del sistema

El estado del modelo es $\mathbf{x}(t) = (\mathbf{q}(t), \boldsymbol{\omega}(t))^T$, donde:

1. El cuaternio $\mathbf{q}(t) \in S^3$ describe la rotación.
2. El cuaternio $\boldsymbol{\omega}(t) \in \mathbb{H}_0$ describe la velocidad angular.

Con todo lo visto hasta ahora, las ecuaciones del modelo de la dinámica son

$$(29) \quad \begin{cases} \frac{d}{dt}\mathbf{q}(t) = \frac{1}{2}\mathbf{q}(t)\boldsymbol{\omega}(t), \\ \frac{d}{dt}\boldsymbol{\omega}(t) = W(t), \end{cases}$$

donde $\mathbf{w}(t) = (0, W(t))^T$ es el ruido que afecta al sistema. $W(t) \sim \mathcal{N}(0, Q^\omega)$ es ruido blanco gaussiano con valores en \mathbb{H}_0 . Su matriz de covarianzas es

$$(30) \quad Q(t) = \begin{pmatrix} 0 & 0 \\ 0 & Q^\omega \end{pmatrix},$$

la misma en cada etapa.

Como el modelo no es lineal, necesitaremos el filtro de Kalman extendido. Usamos como trayectoria de referencia a $\mathbf{q}^*(t)$, la solución del sistema

$$(31) \quad \begin{cases} \frac{d}{dt}\mathbf{q}^*(t) = \frac{1}{2}\mathbf{q}^*(t)\boldsymbol{\omega}^*, \\ \frac{d}{dt}\boldsymbol{\omega}^* = 0, \end{cases}$$

donde $\boldsymbol{\omega}^*$ es constante. El modelo se basa en elegir intervalos de tiempo tan pequeños como para asumir que en ellos $\boldsymbol{\omega}$ se mantiene constante salvo ruido.

La observación del modelo es $\mathbf{z}(t) = (\mathbf{a}(t), \mathbf{b}(t))^T$, donde:

1. El cuaternio $\mathbf{a}(t) \in \mathbb{H}_0$ recoge la información $\mathbf{g}(t)$ sobre la rotación en forma de gravedad o campo magnético. Vendrá dada respecto de (B, B') , por lo que la convertiremos con $R(\mathbf{q})^T$ para tenerla respecto de (O, O') .
2. El cuaternio $\mathbf{b}(t) \in \mathbb{H}_0$ recoge la velocidad angular $\boldsymbol{\omega}(t)$.

Las ecuaciones del modelo de la medición son

$$(32) \quad \begin{cases} \mathbf{a}(t) = R(\mathbf{q}(t))^T(\mathbf{g}(t) + \mathbf{r}_0(t)) + \mathbf{r}_1(t), \\ \mathbf{b}(t) = \boldsymbol{\omega}(t) + \mathbf{r}_2(t), \end{cases},$$

donde $\mathbf{r}_1(t)$ y $\mathbf{r}_2(t)$ son ruido y $\mathbf{r}_0(t)$ es la corrección que $\mathbf{g}(t)$ necesita al estar midiéndose en puntos distintos a O . Esta última puede tener media no nula. Definimos $\mathbf{r} = (\mathbf{r}_0(t), \mathbf{r}_1(t), \mathbf{r}_2(t))^T$, $\mathbf{r}^* = (E(\mathbf{r}_0), 0, 0)^T$ y sus matrices de covarianza $R_0(t)$, $R_1(t)$ y $R_2(t)$, respectivamente. En realidad, el ruido que afecta a la medición es $\mathbf{v}(t) = (R(\mathbf{q}(t))^T\mathbf{r}_0(t) + \mathbf{r}_1(t), \mathbf{r}_2(t))^T$. y su matriz de covarianzas es

$$\begin{pmatrix} (R(\mathbf{q}^*)^T R_0 R(\mathbf{q}^*) + R_1 & 0 \\ 0 & R_2 \end{pmatrix}.$$

5. La variedad S^3

En esta sección, M es un espacio topológico.

DEFINICIÓN 5.1. Una *carta* en M es un par (U, ϕ_U) donde $\phi_U : U \rightarrow A$ es un homeomorfismo entre el abierto U de M y el abierto A de \mathbb{R}^n para cierto $n \in \mathbb{N}$. Si $\phi_U(x) = 0$, entonces la carta está *centrada* en x .

DEFINICIÓN 5.2. Una colección de cartas que cubren M es un *atlas*.

Dadas dos cartas (U, ϕ_U) y (V, ϕ_V) , si $U \cap V \neq \emptyset$, entonces queda bien definida la *aplicación de transición* entre cartas $\phi_U \circ \phi_V^{-1} : \phi_V(U \cap V) \rightarrow \mathbb{R}^n$ e induce un homeomorfismo entre abiertos de \mathbb{R}^n . Estos deben tener la misma dimensión debido al Teorema de invarianza de dominios (véase [25, Theorem 2.55] para el contexto y [26, Theorem 6.10.7] para una demostración).

Dado un atlas de M , una función $f: M \rightarrow \mathbb{R}$ se dice *diferenciable* en un punto $x \in U$ si al tomar una carta (U, ϕ_U) tal que $x \in U$, $f \circ \phi_U^{-1}$ es diferenciable en el punto $\phi_U(x)$. Como un punto puede pertenecer a varias cartas, es preciso imponer que si $U \cap V \neq \emptyset$, entonces $\phi_U \circ \phi_V^{-1}$ sea diferenciable.

DEFINICIÓN 5.3. Una *variedad diferenciable* es un espacio topológico Hausdorff y segundo numerable con un atlas cuyas funciones de transición son diferenciables.

Puede exigirse un atlas maximal. También pueden considerarse variedades con distintos tipos de diferenciability, según el de las aplicaciones de transición.

DEFINICIÓN 5.4. Un *grupo de Lie* es un grupo que a la vez es variedad diferenciable, de modo que la multiplicación e inversión son operaciones diferenciables².

En un grupo de Lie, al trasladar una carta (U, ϕ_U) que contiene al elemento identidad e hasta otro elemento g mediante la multiplicación a izquierda L_g , se obtiene una carta $(gU, \phi_U L_g^{-1}|_{gU})$ que contiene a g . Por esto, para especificar el atlas basta dar una carta que contenga a e .

Como hemos visto, la esfera S^3 es un grupo con el producto heredado de los cuaternios y además es de Lie. Las cartas usuales que contienen a 1 son:

1. Carta ortográfica (O):

$$\begin{aligned} \phi: \{\mathbf{q} \in S^3 \mid \operatorname{Re}(\mathbf{q}) > 0\} &\rightarrow \{\mathbf{e} \in \mathbb{H}_0 \mid \|\mathbf{e}\| < 2\} \\ \mathbf{q} &\mapsto 2 \operatorname{Im}(\mathbf{q}) \\ \sqrt{1 - \frac{\|\mathbf{e}\|^2}{4}} + \frac{1}{2}\mathbf{e} &\leftarrow \mathbf{e}. \end{aligned}$$

2. Carta de Rodrigues (RP):

$$\begin{aligned} \phi: \{\mathbf{q} \in S^3 \mid \operatorname{Re}(\mathbf{q}) > 0\} &\rightarrow \mathbb{H}_0 \\ \mathbf{q} &\mapsto \frac{2}{\operatorname{Re}(\mathbf{q})} \operatorname{Im}(\mathbf{q}) \\ \frac{1}{\sqrt{4 + \|\mathbf{e}\|^2}}(2 + \mathbf{e}) &\leftarrow \mathbf{e}. \end{aligned}$$

3. Carta de Rodrigues modificada (MRP):

$$\begin{aligned} \phi: \{\mathbf{q} \in S^3 \mid \operatorname{Re}(\mathbf{q}) > 0\} &\rightarrow \{\mathbf{e} \in \mathbb{H}_0 \mid \|\mathbf{e}\| < 4\} \\ \mathbf{q} &\mapsto \frac{4}{1 + \operatorname{Re}(\mathbf{q})} \operatorname{Im}(\mathbf{q}) \\ \frac{1}{16 + \|\mathbf{e}\|^2}(16 - \|\mathbf{e}\|^2 + 8\mathbf{e}) &\leftarrow \mathbf{e}. \end{aligned}$$

4. Carta exponencial (RV):

$$\begin{aligned} \phi: \{\mathbf{q} \in S^3 \mid \operatorname{Re}(\mathbf{q}) > 0\} &\rightarrow \{\mathbf{e} \in \mathbb{H}_0 \mid \|\mathbf{e}\| < \pi\} \\ e^{\frac{1}{2}\mathbf{e}} &\leftarrow \mathbf{e}. \end{aligned}$$

Todas estas cartas están centradas en 1 y para cada una de ellas se tiene que

$$\phi^{-1}(\mathbf{e}) = 1 - \frac{\|\mathbf{e}\|^2}{8} + \frac{1}{2}\mathbf{e} + O(\|\mathbf{e}\|^3).$$

Fijada una carta (U, ϕ) centrada en 1 y un elemento $\mathbf{q}^* \in S^3$, obtenemos una carta $(\mathbf{q}^*U, \mathbf{e}_{\mathbf{q}^*})$ centrada en \mathbf{q}^* (y el homeomorfismo inverso) mediante

$$\begin{array}{ccccc} \mathbf{q}^*U & \longrightarrow & U & \longrightarrow & \phi(U) \\ \mathbf{q} & \longmapsto & \overline{\mathbf{q}^*}\mathbf{q} & \longmapsto & \mathbf{e}_{\mathbf{q}^*}(\mathbf{q}) := \phi(\overline{\mathbf{q}^*}\mathbf{q}) \\ \mathbf{q}\mathbf{q}^*(\mathbf{e}) := \mathbf{q}^*\phi^{-1}(\mathbf{e}) & \longleftarrow & \phi^{-1}(\mathbf{e}) & \longleftarrow & \mathbf{e}. \end{array}$$

²El producto cartesiano de variedades diferenciables es una variedad diferenciable.

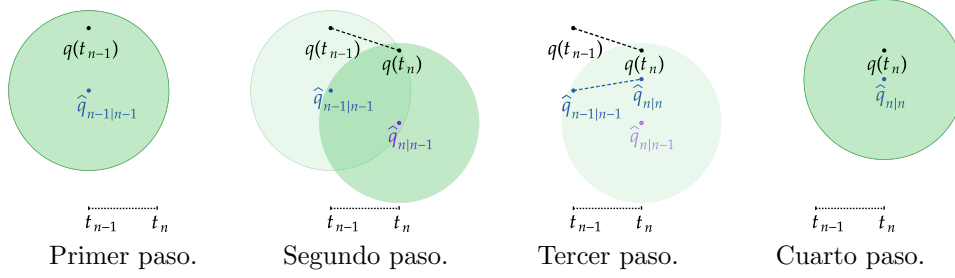


FIGURA 2. Pasos del proceso de filtrado. El estado es negro; la predicción, azul; la estimación, morada; el abierto, verde.

6. Estructura del proceso de filtrado

Describimos las operaciones que se producen en cada iteración. Con los superíndices señalamos el centro de la carta que estamos usando para calcular un dato. Recordamos que el estado \mathbf{x} y en particular la actitud \mathbf{q} son desconocidos. En el cambio de la etapa $n-1$ a la n ocurren los siguientes pasos:

1. En tiempo t_{n-1} tenemos una estimación $\hat{\mathbf{q}}_{n-1|n-1}$ de la actitud $\mathbf{q}(t_{n-1})$. Esta es el centro de una carta que hemos usado como referencia para calcular la estimación $\hat{\mathbf{x}}_{n-1|n-1}^{\hat{\mathbf{q}}_{n-1|n-1}}$ y la matriz $\mathbf{P}_{n-1|n-1}^{\hat{\mathbf{q}}_{n-1|n-1}}$.
2. En tiempo t_n el estado cambia. Hacemos una predicción $\hat{\mathbf{q}}_{n|n-1}$ de la nueva actitud $\mathbf{q}(t_n)$ a partir de la dinámica. Esta nos sirve como centro de una carta provisional que usamos para calcular por linealización la predicción $\hat{\mathbf{x}}_{n|n-1}^{\hat{\mathbf{q}}_{n|n-1}}$ y la matriz $\mathbf{P}_{n|n-1}^{\hat{\mathbf{q}}_{n|n-1}}$ con el filtro Kalman-Bucy.
3. Al llegar la nueva medición, sobre la misma carta calculamos por linealización la estimación $\hat{\mathbf{x}}_{n|n}^{\hat{\mathbf{q}}_{n|n-1}}$ y la matriz $\mathbf{P}_{n|n}^{\hat{\mathbf{q}}_{n|n-1}}$ con el filtro discreto.
4. Finalmente cambiamos de carta a una centrada en $\hat{\mathbf{q}}_{n|n}$ y obtenemos la estimación $\hat{\mathbf{x}}_{n|n}^{\hat{\mathbf{q}}_{n|n}}$ y la matriz $\mathbf{P}_{n|n}^{\hat{\mathbf{q}}_{n|n}}$ respecto de ella. Para aliviar la carga computacional se usa una aproximación de primer orden de la aplicación de cambio de coordenadas entre cartas.

Aunque en el transcurso de t_{n-1} a t_n la velocidad angular $\boldsymbol{\omega}$ se asuma constante, la medición en t_n incorporará el cambio en $\boldsymbol{\omega}$ y nos permitirá seguir la actitud del sólido; el filtro corregirá los cambios debidos al ruido.

Para las siguientes secciones, nos queda pendiente:

1. Linealizar el modelo para la dinámica.
2. Linealizar el modelo para la medición.
3. Calcular la covarianza del error, resolviendo la ecuación (27).
4. Calcular la predicción, resolviendo una ecuación diferencial.
5. Calcular la estimación, de manera natural.
6. Calcular los cambios de carta del estado y la matriz de covarianzas.

Desarrollamos los cálculos de un modo distinto a [24]. Para reducir la notación, a menudo obviaremos la variable t , conscientes de que solo $\boldsymbol{\omega}^*$ es constante. Denotaremos $\Delta t = t_n - t_{n-1}$, $\Delta \boldsymbol{\omega} = \boldsymbol{\omega} - \boldsymbol{\omega}^*$ y $\Delta \mathbf{e}_{\mathbf{q}^*}(\mathbf{q}) = \mathbf{e}_{\mathbf{q}^*}(\mathbf{q}) - 0$. Con el subíndice n haremos referencia a valores en el intervalo $[t_{n-1}, t_n)$.

Tomamos una carta $(\mathbf{q}^*U, \mathbf{e}_{\mathbf{q}^*})$ centrada en \mathbf{q}^* pasando a través de alguna de las cuatro cartas (U, ϕ) centradas en 1 recién descritas. En este sistema,

$$(33) \quad \mathbf{e}_{\mathbf{q}^*}(\mathbf{q}) = \phi(\boldsymbol{\gamma})$$

son las “coordenadas” de \mathbf{q} , con $\gamma = \overline{\mathbf{q}^*} \mathbf{q}$. Solo desarrollaremos en profundidad las fórmulas para la carta ortográfica, pues el resto son similares.

7. Linealización de la ecuación para la dinámica

Buscamos linealizar la dinámica; si $\dot{\mathbf{x}} = f(\mathbf{x}, \mathbf{w})$, debemos encontrar los términos de primer orden del desarrollo de Taylor de f en el punto $(\mathbf{x}^*, 0)$. Aquí,

$$(34) \quad \mathbf{x} = (\mathbf{e}_{\mathbf{q}^*}(\mathbf{q}), \mathbf{w})^T$$

y $\mathbf{x}^* = (0, \mathbf{w}^*)^T$, pero desconocemos f . Para obtener los términos de primer grado, trabajamos *módulo* ellos, ignorando los términos de orden mayor que uno. Indicamos esta relación de equivalencia con el símbolo “ \equiv ”.

Para comenzar observamos que como

$$(35) \quad \dot{\mathbf{x}} = \begin{pmatrix} \frac{d}{dt} \mathbf{e}_{\mathbf{q}^*}(\mathbf{q}) \\ W \end{pmatrix} = \begin{pmatrix} \frac{d}{dt} \phi(\gamma) \\ W \end{pmatrix},$$

podemos asumir que $f(\mathbf{x}, \mathbf{w}) = (f_1(\mathbf{x}), \mathbf{w})^T$ para una aplicación $f_1(\mathbf{x})$ tal que

$$f_1(\mathbf{x}) = \frac{d}{dt} \phi(\gamma)$$

y reducir el problema a calcular el término de primer orden del desarrollo de Taylor de $f_1(\mathbf{x})$ en el punto \mathbf{x}^* .

Calculemos primero $\frac{d}{dt} \gamma$. Como $\mathbf{q} = \mathbf{q}^* \gamma$ entonces $\dot{\mathbf{q}} = \dot{\mathbf{q}}^* \gamma + \mathbf{q}^* \dot{\gamma}$, y por las ecuaciones diferenciales, $\frac{1}{2} \dot{\mathbf{q}} \mathbf{w} = \frac{1}{2} \dot{\mathbf{q}}^* \mathbf{w}^* \gamma + \mathbf{q}^* \dot{\gamma}$. Esto nos permite despejar

$$(36) \quad \begin{aligned} \dot{\gamma} &= \frac{1}{2} (\overline{\dot{\mathbf{q}}^*} \mathbf{q} \mathbf{w} - \mathbf{w}^* \dot{\gamma}) = \frac{1}{2} (\gamma \dot{\omega} - \dot{\gamma} \omega^* + \gamma \dot{\omega}^* - \dot{\gamma} \omega^*) \\ &= \frac{1}{2} \gamma \Delta \omega - \mathbf{w}^* \times \dot{\gamma} = \frac{1}{2} (\text{Re}(\gamma) + \text{Im}(\gamma)) \Delta \omega - \mathbf{w}^* \times \dot{\gamma} \\ &= \frac{1}{2} \text{Re}(\gamma) \Delta \omega + \frac{1}{2} (\text{Im}(\gamma) \times \Delta \omega - \langle \gamma, \Delta \omega \rangle 1) - \mathbf{w}^* \times \dot{\gamma} \\ &= \frac{1}{2} (\text{Re}(\gamma) \Delta \omega - (\Delta \omega + 2\mathbf{w}^*) \times \dot{\gamma}) - \frac{1}{2} \langle \gamma, \Delta \omega \rangle 1. \end{aligned}$$

Continuemos con el desarrollo de $g(\mathbf{x})$. Por la regla de la cadena,

$$f_1(\mathbf{x}) = D\phi|_{\gamma} \frac{d}{dt} \gamma,$$

donde D denota la diferencial. Cambiando γ por $\phi^{-1}(\mathbf{e}_{\mathbf{q}^*}(\mathbf{q}))$ en la derecha podemos llegar a la expresión de f_1 en términos de \mathbf{x} .

En realidad no es necesario, pues el término de grado cero en el desarrollo de $\dot{\gamma}$ corresponde a hacer $(\gamma, \omega) = (1, 0)$, y por (36) se anula. Por tanto, solo consideramos el término independiente en $D\phi|_{\gamma}$ y el de primer orden de $\dot{\gamma}$.

El primero se calcula fácilmente para todos los casos de ϕ como

$$D\phi|_1 : \mathbf{x} \mapsto 2 \text{Im}(\mathbf{x}).$$

En cuanto al segundo, como $\gamma \equiv 1 + \frac{1}{2} \mathbf{e}_{\mathbf{q}^*}(\mathbf{q})$, por (36),

$$\frac{d}{dt} \gamma \equiv \frac{1}{2} \left(\Delta \omega - \mathbf{w}^* \times \frac{1}{2} \mathbf{e}_{\mathbf{q}^*}(\mathbf{q}) \right).$$

Así pues, el término de orden 1 es

$$2 \text{Im} \left(\frac{1}{2} \left(\Delta \omega - \mathbf{w}^* \times \frac{1}{2} \mathbf{e}_{\mathbf{q}^*}(\mathbf{q}) \right) \right) = \Delta \omega - \mathbf{w}^* \times \frac{1}{2} \mathbf{e}_{\mathbf{q}^*}(\mathbf{q}).$$

La aproximación hasta el primer orden queda

$$(37) \quad \frac{d}{dt} \begin{pmatrix} \Delta \mathbf{e}_{\mathbf{q}^*}(\mathbf{q}) \\ \Delta \mathbf{w} \end{pmatrix} \equiv \begin{pmatrix} \Delta \mathbf{w} - \mathbf{w}^* \times \mathbf{e}_{\mathbf{q}^*}(\mathbf{q}) \\ 0 \end{pmatrix} + \begin{pmatrix} 0 \\ W \end{pmatrix},$$

e identificando cuaternios con sus coordenadas tenemos el siguiente resultado.

PROPOSICIÓN 5.6. *Las matrices F y G que aparecen en (20) para (29) son*

$$F = \begin{pmatrix} -[\mathbf{w}^*]_{\times} & I \\ 0 & 0 \end{pmatrix}, \quad G = I.$$

La matriz F solo depende de $\boldsymbol{\omega}^*$, que es constante en cada intervalo. Esto simplifica notablemente las fórmulas.

8. Linealización de la ecuación para la medición

Buscamos linealizar la medición; si $\mathbf{z} = h(\mathbf{x}, \mathbf{r})$, debemos encontrar los términos de primer orden del desarrollo de Taylor de h en el punto $(\mathbf{x}^*, 0)$. Los aislaremos reordenando \mathbf{z} en términos de $\Delta \mathbf{e}_{\mathbf{q}^*}(\mathbf{q})$, $\Delta \mathbf{w}$ y \mathbf{r} . Ahora desconocemos h .

Llamando $\mathbf{a}^* := \overline{\mathbf{q}^*}(\mathbf{g} + \mathbf{r}_0)\mathbf{q}^*$, como $\mathbf{q} = \mathbf{q}^* \phi^{-1}(\mathbf{e}_{\mathbf{q}^*}(\mathbf{q})) \equiv \mathbf{q}^* + \frac{1}{2}\mathbf{q}^* \mathbf{e}_{\mathbf{q}^*}(\mathbf{q})$,

$$\begin{aligned} R(\mathbf{q})^T(\mathbf{g} + \mathbf{r}_0) - R(\mathbf{q}^*)^T(\mathbf{g} + \mathbf{r}_0) \\ &\equiv \left(1 - \frac{1}{2}\mathbf{e}_{\mathbf{q}^*}(\mathbf{q})\right) \overline{\mathbf{q}^*}(\mathbf{g} + \mathbf{r}_0)\mathbf{q}^* \left(1 + \frac{1}{2}\mathbf{e}_{\mathbf{q}^*}(\mathbf{q})\right) - \overline{\mathbf{q}^*}(\mathbf{g} + \mathbf{r}_0)\mathbf{q}^* \\ &\equiv -\frac{1}{2}\mathbf{e}_{\mathbf{q}^*}(\mathbf{q})\overline{\mathbf{q}^*}(\mathbf{g} + \mathbf{r}_0)\mathbf{q}^* + \frac{1}{2}\overline{\mathbf{q}^*}(\mathbf{g} + \mathbf{r}_0)\mathbf{q}^* \mathbf{e}_{\mathbf{q}^*}(\mathbf{q}) = \frac{1}{2}[\mathbf{a}^*, \mathbf{e}_{\mathbf{q}^*}(\mathbf{q})]. \end{aligned}$$

Por tanto,

$$(38) \quad \mathbf{z} - \mathbf{z}^* \approx \begin{pmatrix} [\mathbf{a}^*]_{\times} & 0 \\ 0 & I \end{pmatrix} (\mathbf{x} - \mathbf{x}^*) + \begin{pmatrix} R(\mathbf{q}^*)^T & I & 0 \\ 0 & 0 & I \end{pmatrix} (\mathbf{r} - \mathbf{r}^*),$$

e identificando cuaternios con sus coordenadas tenemos el siguiente resultado.

PROPOSICIÓN 5.7. *Las matrices H y L que aparecen en (21) para (32) son*

$$H = \begin{pmatrix} [\mathbf{a}^*(t)]_{\times} & 0 \\ 0 & I \end{pmatrix}, \quad L = I.$$

9. Matriz de covarianzas del error

Buscamos resolver (27). Que la matriz F no dependa del tiempo nos facilitará mucho las cosas. Siguiendo el teorema 4.19, se tiene el siguiente resultado.

TEOREMA 5.8. *La covarianza del error $\mathbf{P}(t) = E((\mathbf{x} - \hat{\mathbf{x}})(\mathbf{x} - \hat{\mathbf{x}})^T)$ cumple*

$$(39) \quad \frac{d}{dt}\mathbf{P}(t) = F\mathbf{P}(t) + \mathbf{P}(t)F^T + GQG^T.$$

Procedamos, pues, a resolver (39). Tomamos la versión homogénea

$$\frac{d}{dt}\mathbf{P}_h(t) = F\mathbf{P}_h(t) + \mathbf{P}_h(t)F^T,$$

cuya solución es $\mathbf{P}_h(t) = e^{tF}C_0e^{tF^T}$ para alguna matriz C_0 . Si ahora dejamos variar C en función de t , obtenemos $\mathbf{P}(t) = e^{tF}C(t)e^{tF^T}$, y debe cumplir que

$$F\mathbf{P}(t) + e^{tF}\frac{d}{dt}C(t)e^{tF^T} + \mathbf{P}(t)F^T = \frac{d}{dt}\mathbf{P}(t) = F\mathbf{P}(t) + \mathbf{P}(t)F^T + GQG^T.$$

Como aquí $G = I$, la ecuación que debe cumplir $C(t)$ es³

$$\begin{cases} \frac{d}{dt}C(t) = e^{-tF}Qe^{-tF^T}, \\ C(0) = \mathbf{P}(0). \end{cases}$$

³En la práctica, la condición inicial vendrá dada en tiempo t_{n-1} .

Basándonos en la aproximación

$$e^{tF} = \exp\left(\begin{pmatrix} -t[\boldsymbol{\omega}^*]_{\times} & tI \\ 0 & 0 \end{pmatrix}\right) \approx \begin{pmatrix} e^{-t[\boldsymbol{\omega}^*]_{\times}} & tI \\ 0 & I \end{pmatrix} = \begin{pmatrix} R(e^{\frac{1}{2}t\boldsymbol{\omega}^*})^T & tI \\ 0 & I \end{pmatrix},$$

podemos desarrollar $C(t)$ en serie de potencias de t y aproximarla por

$$C(t) \approx \mathbf{P}(0) + \begin{pmatrix} Q^w \frac{t^3}{3} & -Q^w \frac{t^2}{2} \\ -Q^w \frac{t^2}{2} & Q^w t \end{pmatrix}.$$

Incorporando la matriz $C(t)$ a $\mathbf{P}(t)$, tenemos el siguiente resultado.

PROPOSICIÓN 5.9. *En $[t_{n-1}, t_n)$, basándonos en la aproximación anterior,*

$$\mathbf{P}(t) \approx \begin{pmatrix} R(e^{\frac{1}{2}t\boldsymbol{\omega}^*})^T & tI \\ 0 & I \end{pmatrix} \left(\mathbf{P}(0) + \begin{pmatrix} Q^w \frac{t^3}{3} & -Q^w \frac{t^2}{2} \\ -Q^w \frac{t^2}{2} & Q^w t \end{pmatrix} \right) \begin{pmatrix} R(e^{\frac{1}{2}t\boldsymbol{\omega}^*}) & 0 \\ tI & I \end{pmatrix}.$$

Con esto podemos predecir la matriz de covarianzas.

COROLARIO 5.10. *Basándonos en la aproximación anterior,*

$$(40) \quad \mathbf{P}_{n|n-1} := \mathbf{P}(t_n) \approx E_n (\mathbf{P}_{n-1|n-1} + Q_n) E_n^T,$$

donde

$$(41) \quad E_n := \begin{pmatrix} R(e^{\frac{1}{2}\Delta t\boldsymbol{\omega}^*})^T & \Delta t I \\ 0 & I \end{pmatrix}, \quad Q_n = \begin{pmatrix} Q^w \frac{(\Delta t)^3}{3} & -Q^w \frac{(\Delta t)^2}{2} \\ -Q^w \frac{(\Delta t)^2}{2} & Q^w \Delta t \end{pmatrix}.$$

10. Etapa de predicción

La predicción se hace siguiendo la ecuación de la dinámica sin ruido

$$\begin{cases} \frac{d}{dt} \mathbf{q}^*(t) = \frac{1}{2} \mathbf{q}^*(t) \boldsymbol{\omega}^*, \\ \frac{d}{dt} \boldsymbol{\omega}^* = 0, \end{cases}$$

con las condiciones iniciales $\hat{\mathbf{q}}_{n-1|n-1}$ y $\hat{\boldsymbol{\omega}}_{n-1|n-1}$ en tiempo t_{n-1} . Como la velocidad angular permanece constante, su predicción es

$$\hat{\boldsymbol{\omega}}_{n|n-1} = \hat{\boldsymbol{\omega}}_{n-1|n-1}.$$

Con esta, la solución del sistema es

$$\mathbf{q}^*(t) = \hat{\mathbf{q}}_{n-1|n-1} e^{\frac{t-t_{n-1}}{2} \hat{\boldsymbol{\omega}}_{n|n-1}}.$$

Marca la trayectoria de referencia en $[t_{n-1}, t_n)$. En t_n predecimos

$$\hat{\mathbf{q}}_{n|n-1} = \hat{\mathbf{q}}_{n-1|n-1} e^{\frac{\Delta t}{2} \hat{\boldsymbol{\omega}}_{n|n-1}}.$$

De acuerdo al corolario 5.10 y con su misma notación, predecimos

$$\mathbf{P}_{n|n-1}^{\hat{\mathbf{q}}} = E_n \left(\mathbf{P}_{n-1|n-1}^{\hat{\mathbf{q}}} + Q_n \right) E_n^T.$$

La predicción $\hat{\mathbf{z}}_{n|n-1} = (\hat{\mathbf{a}}_{n|n-1}, \hat{\mathbf{b}}_{n|n-1})^T$ de la medición para tiempo t_n es⁴

$$\begin{aligned} \hat{\mathbf{a}}_{n|n-1} &= R^T(\hat{\mathbf{q}}_{n|n-1})(E(\mathbf{r}_0)_n + \mathbf{g}_n), \\ \hat{\mathbf{b}}_{n|n-1} &= \hat{\boldsymbol{\omega}}_{n|n-1}, \end{aligned}$$

donde $\mathbf{g}_n = \mathbf{g}(t_n)$ y hemos calculado $\hat{\mathbf{a}}_{n|n-1}$ según el modelo de la medición. Si el sólido estuviese muy alejado del sistema inercial o en lugar de la gravedad estamos midiendo el campo magnético, $E(\mathbf{r}_0)_n$ podría no ser nula.

Por (38), la covarianza de $\hat{\mathbf{z}}_{n|n-1}$ es

$$\mathbf{S}_{n|n-1} = H_n \mathbf{P}_{n|n-1} H_n^T + R_n.$$

⁴Recordamos que las aplicaciones de la medición se linealizan con base en la nueva predicción.

11. Etapa de estimación

La estimación se realiza de modo natural, según el filtro de Kalman extendido:

$$\begin{aligned}\mathbf{K}_n &= \mathbf{P}_{n|n-1}^{\hat{\mathbf{q}}_{n|n-1}} H_n^T \mathbf{S}_{n|n-1}^{-1}, \\ \hat{\mathbf{x}}_{n|n}^{\hat{\mathbf{q}}_{n|n-1}} &= \hat{\mathbf{x}}_{n|n-1}^{\hat{\mathbf{q}}_{n|n-1}} + \mathbf{K}_n (\mathbf{z}_n - \hat{\mathbf{z}}_{n|n-1}), \\ \mathbf{P}_{n|n}^{\hat{\mathbf{q}}_{n|n-1}} &= (I - \mathbf{K}_n H_n) \mathbf{P}_{n|n-1}^{\hat{\mathbf{q}}_{n|n-1}} (I - \mathbf{K}_n H_n)^T + \mathbf{K}_n R_n \mathbf{K}_n^T,\end{aligned}$$

donde $\hat{\mathbf{x}}_{n|n}^{\hat{\mathbf{q}}_{n|n-1}} = (\hat{\mathbf{e}}_{n|n-1}^{\hat{\mathbf{q}}_{n|n-1}} = 0, \hat{\boldsymbol{\omega}}_{n|n-1})^T$.

12. Transformación tras el cambio de carta

Debemos transformar la estimación según la carta con centro en $\hat{\mathbf{q}}_{n|n-1}$ a una según la carta con centro en $\hat{\mathbf{q}}_{n|n}$. Esta transformación tiene notables efectos en la matriz de covarianzas. Dados $\mathbf{q} \in \hat{\mathbf{q}}_{n|n} U \cap \hat{\mathbf{q}}_{n|n-1} U$ y $\delta := \hat{\mathbf{q}}_{n|n-1}^{-1} \hat{\mathbf{q}}_{n|n}$,

$$\begin{aligned}\mathbf{e}_{\hat{\mathbf{q}}_{n|n}}(\mathbf{q}) &= \phi(\hat{\mathbf{q}}_{n|n}^{-1} \mathbf{q}) = \phi(\hat{\mathbf{q}}_{n|n}^{-1} \hat{\mathbf{q}}_{n|n-1} \phi^{-1}(\phi(\hat{\mathbf{q}}_{n|n-1}^{-1}(\mathbf{q})))) \\ &= \phi(\hat{\mathbf{q}}_{n|n-1}^{-1} \hat{\mathbf{q}}_{n|n-1} \phi^{-1}(\mathbf{e}_{\hat{\mathbf{q}}_{n|n-1}}(\mathbf{q}))) = \phi L_{\delta}^{-1} \phi^{-1}(\mathbf{e}_{\hat{\mathbf{q}}_{n|n-1}}(\mathbf{q})),\end{aligned}$$

donde L_{δ}^{-1} es el inverso del operador de multiplicación a izquierda por δ . Como δ tiene norma 1, este coincide con $L_{\bar{\delta}}$. Usando la aplicación ϕ de la carta ortográfica tenemos que para $\mathbf{e} \in \phi(U)$,

$$\begin{aligned}\phi L_{\bar{\delta}} \phi^{-1}(\mathbf{e}) &= \phi L_{\bar{\delta}} \left(\sqrt{1 - \frac{\|\mathbf{e}\|^2}{4}} + \frac{1}{2} \mathbf{e} \right) = 2 \operatorname{Im} \left(\bar{\delta} \left(\sqrt{1 - \frac{\|\mathbf{e}\|^2}{4}} + \frac{1}{2} \mathbf{e} \right) \right) \\ &= \bar{\delta} \left(\sqrt{1 - \frac{\|\mathbf{e}\|^2}{4}} + \frac{1}{2} \mathbf{e} \right) - \overline{\bar{\delta} \left(\sqrt{1 - \frac{\|\mathbf{e}\|^2}{4}} + \frac{1}{2} \mathbf{e} \right)} \\ &= \bar{\delta} \left(\sqrt{1 - \frac{\|\mathbf{e}\|^2}{4}} + \frac{1}{2} \mathbf{e} \right) - \left(\sqrt{1 - \frac{\|\mathbf{e}\|^2}{4}} - \frac{1}{2} \mathbf{e} \right) \delta \\ &= 2 \operatorname{Re}(\delta) \mathbf{e} - \sqrt{4 - \|\mathbf{e}\|^2} \operatorname{Im}(\delta) - \operatorname{Im}(\delta) \times \mathbf{e}\end{aligned}$$

En realidad se usa la aproximación de primer orden en el punto $\phi(\delta)$. Para calcular la diferencial de esta aplicación, partimos de una curva derivable γ con $\gamma(0) = \phi(\delta) = 2 \operatorname{Im}(\delta)$ y $\dot{\gamma}(0) = \mathbf{v} \in \mathbb{H}_0$ y derivamos la curva $\phi L_{\bar{\delta}} \phi^{-1}(\gamma)$ en $t = 0$:

$$\begin{aligned}D\phi L_{\bar{\delta}} \phi^{-1}|_{\phi(\delta)}(\mathbf{v}) &= \left. \frac{d}{dt} \right|_{t=0} \phi L_{\bar{\delta}} \phi^{-1}(\gamma(t)) \\ &= \left. \frac{d}{dt} \right|_{t=0} \left(2 \operatorname{Re}(\delta) \gamma(t) - \sqrt{4 - \|\gamma(t)\|^2} \operatorname{Im}(\delta) - \operatorname{Im}(\delta) \times \gamma(t) \right) \\ &= 2 \operatorname{Re}(\delta) \mathbf{v} - \left. \frac{d}{dt} \right|_{t=0} \left(\sqrt{4 - \|\gamma(t)\|^2} \right) \operatorname{Im}(\delta) - \operatorname{Im}(\delta) \times \mathbf{v} \\ &= 2 \operatorname{Re}(\delta) \mathbf{v} - \frac{-2 \langle \dot{\gamma}(0), \gamma(0) \rangle}{2 \sqrt{4 - \|\gamma(0)\|^2}} \operatorname{Im}(\delta) - \operatorname{Im}(\delta) \times \mathbf{v} \\ &= 2 \operatorname{Re}(\delta) \mathbf{v} + \frac{\langle \mathbf{v}, \gamma(0) \rangle}{2 \sqrt{1 - \|\operatorname{Im}(\delta)\|^2}} \operatorname{Im}(\delta) - \operatorname{Im}(\delta) \times \mathbf{v} \\ &= 2 \operatorname{Re}(\delta) \mathbf{v} + \frac{\langle \mathbf{v}, \operatorname{Im}(\delta) \rangle}{\operatorname{Re}(\delta)} \operatorname{Im}(\delta) - \operatorname{Im}(\delta) \times \mathbf{v} \\ &= 2 \operatorname{Re}(\delta) \mathbf{v} + \frac{\operatorname{Im}(\delta) \operatorname{Im}(\delta)^T}{\operatorname{Re}(\delta)} \mathbf{v} - \operatorname{Im}(\delta) \times \mathbf{v}\end{aligned}$$

Carta	Aplicación $\phi L_{\bar{\delta}} \phi^{-1}(\mathbf{e})$
O	$2 \operatorname{Re}(\delta) \mathbf{e} - \sqrt{4 - \ \mathbf{e}\ ^2} \operatorname{Im}(\delta) - \operatorname{Im}(\delta) \times \mathbf{e}$
RP	$2 \frac{\operatorname{Re}(\delta) \mathbf{e} - 2 \operatorname{Im}(\delta) - \operatorname{Im}(\delta) \times \mathbf{e}}{2 \operatorname{Re}(\delta) + \operatorname{Im}(\delta) \cdot \mathbf{e}}$
MRP	$4 \frac{8 \operatorname{Re}(\delta) \mathbf{e} - (16 - \ \mathbf{e}\ ^2) \operatorname{Im}(\delta) - 8 \operatorname{Im}(\delta) \times \mathbf{e}}{16 + \ \mathbf{e}\ ^2 + \operatorname{Re}(\delta)(16 - \ \mathbf{e}\ ^2) + 8 \operatorname{Im}(\delta) \cdot \mathbf{e}}$
RV	Véase [24, Table 3.3]

CUADRO 2. Aplicaciones de transición $\phi L_{\bar{\delta}} \phi^{-1}(\mathbf{e})$.

Carta	Matriz T	Dominio
O	$2 \operatorname{Re}(\delta) I - [\operatorname{Im}(\delta)]_{\times} + \frac{\operatorname{Im}(\delta) \operatorname{Im}(\delta)^T}{\operatorname{Re}(\delta)}$	$\{\delta \in S^3 \mid \operatorname{Re}(\delta) > 0\}$
RP	$\operatorname{Re}(\delta) (\operatorname{Re}(\delta) I - [\operatorname{Im}(\delta)]_{\times})$	$\{\delta \in S^3 \mid \operatorname{Re}(\delta) \neq 0\}$
MRP	$\frac{1}{2} [(1 + \operatorname{Re}(\delta)) (\operatorname{Re}(\delta) I - [\operatorname{Im}(\delta)]_{\times}) + \operatorname{Im}(\delta) \operatorname{Im}(\delta)^T]$	$\{\delta \in S^3 \mid \operatorname{Re}(\delta) \geq 0\}$
RV	Véase [24, Table 5.1]	Véase [24, Table 5.1]

CUADRO 3. Matrices T .

donde en la última igualdad identificamos $\operatorname{Im}(\delta)$ con sus coordenadas para escribir su traspuesta. Así, alrededor de $\phi(\delta)$,

$$\phi L_{\bar{\delta}} \phi^{-1}(\mathbf{e}) \approx \phi(\delta) + 2 \operatorname{Re}(\delta) \mathbf{e} + \frac{\operatorname{Im}(\delta) \operatorname{Im}(\delta)^T}{\operatorname{Re}(\delta)} \mathbf{e} - \operatorname{Im}(\delta) \times \mathbf{e}.$$

Llamamos T a la matriz que se aplica a \mathbf{e} :

$$T = \operatorname{Re}(\delta) I - [\operatorname{Im}(\delta)]_{\times} + \frac{\operatorname{Im}(\delta) \operatorname{Im}(\delta)^T}{\operatorname{Re}(\delta)}.$$

Los cálculos en los distintos tipos de cartas se recogen en el cuadro 2 y en el cuadro 3. La aplicación de transición para (RV) es $2 \ln(\bar{\delta} e^{\frac{1}{2} \mathbf{e}})$, pero la expresión exacta, al igual que la de su matriz T asociada, puede encontrarse en [24].

Si denotamos por $\mathbf{x}^{\hat{\mathbf{q}}_{n|n-1}}$ y $\mathbf{x}^{\hat{\mathbf{q}}_{n|n}}$ al estado del sólido referido a la carta con centro en $\hat{\mathbf{q}}_{n|n-1}$ y $\hat{\mathbf{q}}_{n|n}$ respectivamente, el cambio de carta produce la relación

$$\begin{aligned} \mathbf{x}^{\hat{\mathbf{q}}_{n|n}} - \hat{\mathbf{x}}_{n|n}^{\hat{\mathbf{q}}_{n|n}} &= \begin{pmatrix} \mathbf{e}_{\hat{\mathbf{q}}_{n|n}}(\mathbf{q}(t_n)) - \mathbf{e}_{\hat{\mathbf{q}}_{n|n}}(\hat{\mathbf{q}}_{n|n}) \\ \boldsymbol{\omega}(t_n) - \hat{\boldsymbol{\omega}}_{n|n} \end{pmatrix} \\ &\approx \begin{pmatrix} T(\delta_n)(\mathbf{e}_{\hat{\mathbf{q}}_{n|n-1}}(\mathbf{q}(t_n)) - \mathbf{e}_{\hat{\mathbf{q}}_{n|n-1}}(\hat{\mathbf{q}}_{n|n})) \\ \boldsymbol{\omega}(t_n) - \hat{\boldsymbol{\omega}}_{n|n} \end{pmatrix} \\ &= \begin{pmatrix} T(\delta_n) & 0 \\ 0 & I \end{pmatrix} (\mathbf{x}^{\hat{\mathbf{q}}_{n|n-1}} - \hat{\mathbf{x}}_{n|n}^{\hat{\mathbf{q}}_{n|n-1}}), \end{aligned}$$

donde $\delta_n = \hat{\mathbf{q}}_{n|n-1}^{-1} \mathbf{q}_{n|n}$. Así, la matriz de covarianza del error se transforma por el cambio de carta de acuerdo a la regla

$$\mathbf{P}_{n|n}^{\hat{\mathbf{q}}_{n|n}} \approx \begin{pmatrix} T(\delta_n) & 0 \\ 0 & I \end{pmatrix} \mathbf{P}_{n|n}^{\hat{\mathbf{q}}_{n|n-1}} \begin{pmatrix} T(\delta_n) & 0 \\ 0 & I \end{pmatrix}^T.$$

13. Resumen del algoritmo de filtrado

Recopilamos el filtrado para el seguimiento de la actitud desarrollado en este capítulo. En tiempo t_{n-1} disponemos del centro $\hat{\mathbf{q}}_{n-1|n-1}$ de la carta, de la estimación $\hat{\mathbf{x}}_{n-1|n-1}^{\hat{\mathbf{q}}_{n-1|n-1}}$ y de la matriz $\mathbf{P}_{n-1|n-1}^{\hat{\mathbf{q}}_{n-1|n-1}}$ de covarianzas del error en la estimación. El incremento de tiempo es $\Delta t = t_n - t_{n-1}$.

Etapa de predicción. Se calculan

$$\begin{aligned}\hat{\boldsymbol{\omega}}_{n|n-1} &= \hat{\boldsymbol{\omega}}_{n-1|n-1}, \\ \hat{\mathbf{q}}_{n|n-1} &= \hat{\mathbf{q}}_{n-1|n-1} e^{\frac{\Delta t}{2} \hat{\boldsymbol{\omega}}_{n|n-1}}, \\ \mathbf{P}_{n|n-1}^{\hat{\mathbf{q}}_{n|n-1}} &= E_n \left(\mathbf{P}_{n-1|n-1}^{\hat{\mathbf{q}}_{n-1|n-1}} + Q_n \right) E_n^T,\end{aligned}$$

usando las matrices auxiliares

$$Q_n = \begin{pmatrix} Q^\omega \frac{(\Delta t)^3}{3} & -Q^\omega \frac{(\Delta t)^2}{2} \\ -Q^\omega \frac{(\Delta t)^2}{2} & Q^\omega \Delta t \end{pmatrix}, \quad E_n = \begin{pmatrix} R(e^{\frac{1}{2}\Delta t \boldsymbol{\omega}^*})^T & \Delta t I \\ 0 & I \end{pmatrix}.$$

Las predicciones para la nueva medición vienen dadas por

$$\begin{aligned}\hat{\mathbf{z}}_{n|n-1} &= \begin{pmatrix} R^T(\hat{\mathbf{q}}_{n|n-1})(E(\mathbf{r}_0)_n + \mathbf{g}_n) \\ \hat{\boldsymbol{\omega}}_{n|n-1} \end{pmatrix}, \\ \mathbf{S}_{n|n-1} &= H_n \mathbf{P}_{n|n-1}^{\hat{\mathbf{q}}_{n|n-1}} H_n^T + R_n,\end{aligned}$$

usando las matrices auxiliares

$$H_n = \begin{pmatrix} [\hat{\mathbf{a}}_{n|n-1}]^\times & 0 \\ 0 & I \end{pmatrix}, \quad R_n = \begin{pmatrix} R(\hat{\mathbf{q}}_{n|n-1})^T (R_0)_n R(\hat{\mathbf{q}}_{n|n-1}) + (R_1)_n & 0 \\ 0 & (R_2)_n \end{pmatrix},$$

donde $\hat{\mathbf{a}}_{n|n-1} = R^T(\hat{\mathbf{q}}_{n|n-1})(E(\mathbf{r}_0)_n + \mathbf{g}_n)$ y $(R_k)_n$ representa la matriz de covarianzas para $\mathbf{r}_k(t)$ durante el intervalo $[t_{n-1}, t_n]$ para $k \in \{1, 2, 3\}$.

Etapa de estimación. Se calculan

$$\begin{aligned}\mathbf{K}_n &= \mathbf{P}_{n|n-1}^{\hat{\mathbf{q}}_{n|n-1}} H_n^T \mathbf{S}_{n|n-1}^{-1}, \\ \hat{\mathbf{x}}_{n|n}^{\hat{\mathbf{q}}_{n|n-1}} &= \hat{\mathbf{x}}_{n|n-1}^{\hat{\mathbf{q}}_{n|n-1}} + \mathbf{K}_n (\mathbf{z}_n - \hat{\mathbf{z}}_{n|n-1}), \\ \mathbf{P}_{n|n}^{\hat{\mathbf{q}}_{n|n-1}} &= (I - \mathbf{K}_n H_n) \mathbf{P}_{n|n-1}^{\hat{\mathbf{q}}_{n|n-1}} (I - \mathbf{K}_n H_n)^T + \mathbf{K}_n R_n \mathbf{K}_n^T,\end{aligned}$$

donde $\hat{\mathbf{x}}_{n|n}^{\hat{\mathbf{q}}_{n|n-1}} = (\hat{\mathbf{e}}_{n|n-1}^{\hat{\mathbf{q}}_{n|n-1}} = 0, \hat{\boldsymbol{\omega}}_{n|n-1})^T$.

Cambio de carta. Se calculan

$$\begin{aligned}\delta_n &= \phi^{-1}(\hat{\mathbf{e}}_{n|n}^{\hat{\mathbf{q}}_{n|n-1}}), \\ \mathbf{q}_{n|n} &= \hat{\mathbf{q}}_{n|n-1} \delta_n, \\ \mathbf{P}_{n|n}^{\hat{\mathbf{q}}_{n|n}} &= \begin{pmatrix} T(\delta_n) & 0 \\ 0 & I \end{pmatrix} \mathbf{P}_{n|n}^{\hat{\mathbf{q}}_{n|n-1}} \begin{pmatrix} T(\delta_n) & 0 \\ 0 & I \end{pmatrix}^T,\end{aligned}$$

donde $\hat{\mathbf{e}}_{n|n}^{\hat{\mathbf{q}}_{n|n-1}}$ es la primera componente de $\hat{\mathbf{x}}_{n|n}^{\hat{\mathbf{q}}_{n|n-1}}$ y $T(\delta_n)$ viene dado por el cuadro 3 según el tipo de carta que se esté usando.

Referencias

- [1] R. E. Kalman. «A New Approach to Linear Filtering and Prediction Problems». En: *Journal of Basic Engineering* (1960).
- [2] A. G. Expósito y J. A. R. Macías y M. Á. G. Cagigal. *Modelado y análisis de la evolución de una epidemia vírica mediante filtros de Kalman: el caso del COVID-19 en España*. 2020.
- [3] D. Simon. *Optimal State Estimation. Kalman, H_∞ and nonlinear approaches*. Wiley, 2006.
- [4] B. Hassibi T. Kailath y A. H. Sayed. *Linear estimation*. Prentice Hall, 2000.
- [5] B. D. O. Anderson y J. B. Moore. *Optimal Filtering*. Prentice-Hall, 1979.
- [6] R. G. Bartle. *The Elements of Integration and Lebesgue Measure*. Wiley-Interscience, 1995.
- [7] W. Rudin. *Principles of Mathematical Analysis*. McGraw-Hill, Inc., 1953.
- [8] B. Fristedt y L. Gray. *A Modern Approach to Probability Theory*. Birkhäuser, 1997.
- [9] Y. L. Tong. *The Multivariate Normal Distribution*. Springer, 1990.
- [10] L. C. Evans. *An introduction to Stochastic Differential Equations*. AMS, 2014.
- [11] S. Schäffler. *Generalized Stochastic Processes: Applications of Harmonic Analysis*. Birkhäuser, 2018.
- [12] I. M. Gelf'and y N. Ya. Vilenkin. *Generalized Functions*. Vol. 4. Academic Press, 1964.
- [13] V. Capasso y D. Bakstein. *An introduction to Continuous-Time Stochastic Processes. Theory, Models, and Applications to Finance, Biology, and Medicine*. 4^a. Birkhäuser, 2021.
- [14] H-H. Kuo. *White Noise Distribution Theory*. CRC Press, 1996.
- [15] A. H. Jazwinski. *Stochastic Processes and Filtering Theory*. Academic Press, 1970.
- [16] W. Rudin. *Functional Analysis*. McGraw-Hill, Inc., 1991.
- [17] B. Øksendal. *Stochastic Differential Equations*. Universitext. Springer, 2013.
- [18] J. Aguilera. *Control de ecuaciones diferenciales estocásticas*. Trabajo fin de grado. Dpto. Análisis Matemático y Matemática Aplicada, UCM, 2020.
- [19] K. Riadh. «Different Aspects of Nonlinear Stochastic Filtering Theory». Tesis doct. Mannheim, Alemania: Universität Mannheim, 2006.
- [20] B. N. Cooperstein. *Advanced Linear Algebra*. 2^a. CRC Press, 2015.
- [21] S. J. Leon. *Linear Algebra with Applications*. 9^a. Pearson, 2015.
- [22] R. Faragher. «Understanding the Basis of the Kalman Filter Via a Simple and Intuitive Derivation». En: *IEEE Signal Processing Magazine* (2012).
- [23] P. A. Ruymgaart y T. T. Soong. *Mathematics of the Kalman-Bucy Filtering*. Springer-Verlag, 1988.
- [24] P. Bernal. «Kalman Filtering for Orientation Estimation using IMUs». Tesis de mtría. Universidad de Murcia, 2020.
- [25] J. M. Lee. *Introduction to Topological Manifolds*. 2^a. Springer, 2011.
- [26] S. Deo. *Algebraic Topology. A Primer*. 2^a. Springer, 2018.

La fotografía de Kalman está tomada de su cuenta de Facebook. La de Schmidt tiene una licencia CC BY-SA 3.0 a nombre del usuario Schmidt62 de Wikimedia Commons. La de Bucy está tomada de su obituario. Las imágenes de la NASA son de dominio público. Los diagramas han sido generados por el autor con software como SageMath. El trabajo se ha desarrollado en \LaTeX .

Conclusiones

Mientras preparaba este trabajo me encontré con la siguiente cita del matemático Alfred N. Whitehead: «No es ninguna paradoja decir que en nuestras disposiciones más teóricas podemos estar lo más próximos a nuestras aplicaciones más prácticas». Me parece una muy adecuada conclusión para este trabajo.

Para entender el filtro he tenido que mezclar técnicas de análisis real, cálculo matricial, álgebra abstracta, análisis funcional, topología, probabilidad, programación e incluso un poco de física. Situar-me en la intersección de tantos campos ha hecho que me dé cuenta de lo importante que es el intercambio de ideas y el interés por los diferentes ámbitos de las matemáticas.

El filtro de Kalman se mantiene vigente a día de hoy en toda clase de aplicaciones dentro de la ciencia y la ingeniería por su eficacia y simpleza, pero si lo hace es gracias a que se fundamenta sobre robustos pilares teóricos.

Doy gracias a mi familia y amigos por su apoyo, y a mi tutor Chema por su disposición y amabilidad.