

## Parte II

### POSICIONAMIENTO Y SEGUIMIENTO CON INTEGRACIÓN DE SENSORES

Se introducen estrategias de integración de sensores con diferentes fundamentos físicos con el objetivo de mejorar la precisión y la robustez de la estimación de la posición



## INTEGRACIÓN DE SENSORES

---

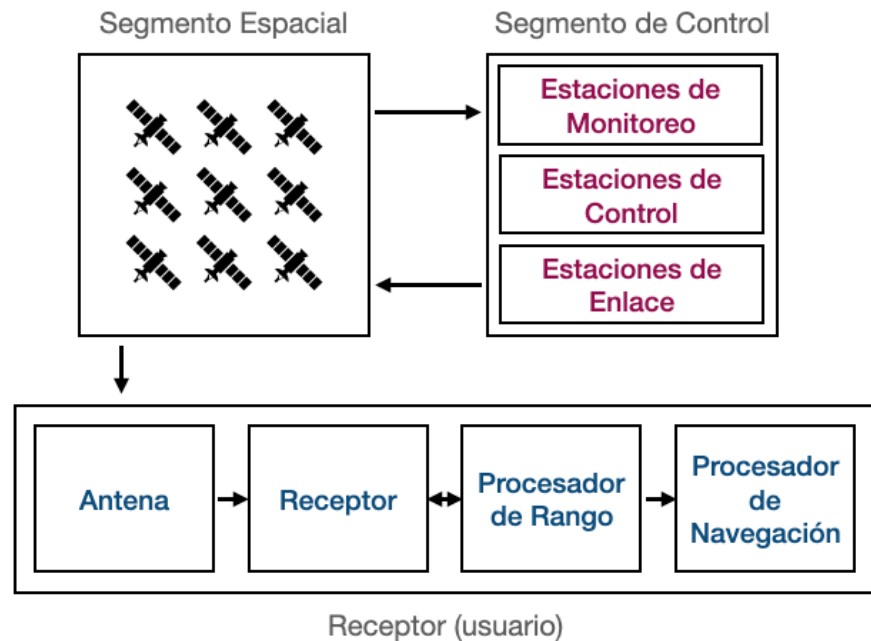
### 2.1 SISTEMAS DE NAVEGACIÓN SATELITAL (GNSS)

GNSS es un término genérico que abarca a sistemas de localización satelital como GPS, GLONASS, Galileo, BeiDou/Compass y sus predecesores y sucesores. En particular, la fenomenal aceptación del GPS para uso empresarial y de los consumidores ha hecho que la antigua tecnología militar pase a utilizarse masivamente.

Los sistemas de navegación por satélite se basan en referencias artificiales (los satélites o segmento espacial) que se distribuyen en el cielo y tienen su movimiento calculado y controlado para estar en órbitas apropiadas (por el segmento de Control). El satélite emite señales de RF que contienen información conocida y códigos que varían con el tiempo. El receptor escucha estas señales, calcula el tiempo de vuelo de la señal y extrae los mensajes relevantes que transporta, e introduce esta información en un algoritmo de localización que devuelve una estimación de la posición del receptor. La ventaja más evidente de esta estrategia es su disponibilidad. La desventaja es el costo de mantenimiento de toda la infraestructura. La figura 2.1 muestra un esquema de esta estructura.

Por mucho, el sistema de posicionamiento satelital más popular es el GPS. Es proporcionado por el Departamento de Defensa de Estados Unidos, de forma gratuita para uso civil con una precisión reducida (unos 10 m) respecto a la utilizada por las agencias gubernamentales estadounidenses y sus aliados. Sin embargo, el GPS adolece de una serie de problemas inherentes. Al menos cuatro satélites deben tener una línea de visión clara hacia el receptor, lo que no siempre es posible en zonas urbanas densas y en el interior de los edificios; puede ser lento en la sincronización, por lo que las lecturas sólo se siguen después de varios minutos; y hay consideraciones políticas.

**LUCES Y SOMBRAS DEL GPS.** Dada la disponibilidad del GPS, es un buen punto de partida a la hora de considerarlo para conocer la posición de las cosas o personas. El GPS fue un hito tan importante en la resolución del problema de la localización, que es difícil imaginar un mundo sin él, pero existen sin embargo algunas preocupaciones y limitaciones. Quizá la mayor preocupación no sea la tecnología en sí, sino la propiedad y el control del servicio. Aparte de los grupos de usuarios para los que se opera, no hay acuerdos de calidad de servicio para nadie más que los militares de un grupo de países aliados. En una economía cada vez más globalizada, con más personas



**Figura 2.1:** Esquema de la estructura del sistema GNSS donde participan un segmento satelital y el receptor al que estamos acostumbrados, pero se suma un segmento de control que posee estaciones de monitoreo y de corrección del sistema.

dependientes del GPS en más países, no es de extrañar que el control político del GPS plantee dudas sobre su valor. Una forma de sortear el problema han sido las propuestas y despliegues de sistemas similares al GPS de otros bloques de poder.

El GPS depende de una red de estaciones terrestres fijas para su seguimiento y control. Y si bien sería poco problemático que uno de los 24 satélites individuales fallara, sería mucho más problemático que se produjera un fallo del sistema en general. Es posible que los aviones civiles no utilicen el GPS como único dispositivo de navegación debido a esta vulnerabilidad. Otra razón para desarrollar otros sistemas como GLONASS y Galileo es proporcionar diversidad al sistema (utilizando receptores multiestándar).

Aunque el GPS funciona bien en muchos entornos al aire libre, sólo tiene una cobertura limitada en interiores. Sería útil que existieran métodos para proporcionar un posicionamiento de precisiones similares, cuando no es posible ver el cielo. Aunque los sistemas de radio convencionales, como la telefonía celular y los servicios de radiodifusión, pueden utilizarse para el posicionamiento en interiores, su precisión es baja, aunque probablemente esto cambie con el estándar 5G. Incluso, dadas las pequeñas escalas de la “geografía” interior, sería útil tener una precisión mejor que la del GPS. Esto está en cambio continuo y aparecen en trabajos científicos y de ingeniería todas las semanas proponiendo soluciones.

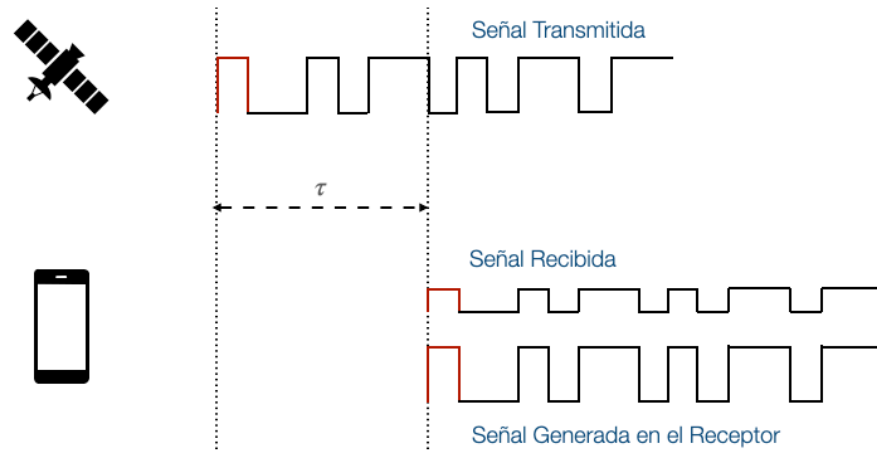
Pero aquí aparece otro tema. Y es el marco de referencia común. Por un lado con medidas de GPS si intercambio las posiciones entre sistemas estas pueden ser manejadas sin problema como Latitud y Longitud. No es así para las personas que raramente compartirían su localización en estos términos. Incluso para los sistemas (como vimos en Fingerprinting), este dato debe ser convertido para poder trabajar en metros. Pero hay otro tema en que este problema es de mayor relevancia. Y es en la integración de sensores. Raramente los esquemas de localización comparten un eje de referencia espacial y pueden estar desplazados y rotados entre sí.

**UN COMENTARIO SOBRE POSICIONAMIENTO MASIVO: SMARTPHONES.** Los dispositivos comerciales que incorporan un receptor de GNSS, como los smartphones, presentan un entorno difícil en términos de recepción de la señal. En estos equipos, la antena suele ser compartida por otro hardware de comunicación, como los receptores Bluetooth y Wi-Fi. La antena está diseñada para una señal polarizada lineal y no circular del GNSS, así como diferentes frecuencias de señal. Los componentes electrónicos cercanos al receptor y a la antena emiten radiación electromagnética ruidosa, lo que deteriora aún más las señales y aumenta la relación señal/ruido. Además, debido a las limitaciones de la batería de los teléfonos inteligentes y a la disponibilidad de canales, el seguimiento de ciertos satélites y constelaciones tiene prioridad sobre otros. Estos factores restringen la capacidad del teléfono inteligente de un buen rendimiento y solo pueden alcanzar de forma realista unas decenas de metros de precisión en condiciones ideales.

### 2.1.1 Principios generales de la navegación satelital

Los sistemas GNSS permiten a los usuarios estimar en tiempo real su posición, velocidad y tiempo con respecto a un marco de referencia en los dominios del espacio y el tiempo. La posición del usuario es determinada por los sistemas de navegación por satélite basándose en la estimación de ToA. El receptor de estos sistemas mide el tiempo de propagación de las señales emitidas por un conjunto de satélites en ubicaciones conocidas que representan las ER.

La medición básica realizada por un receptor GNSS es el tiempo necesario para que la señal se propague desde un satélite al receptor. Esto se puede obtener rastreando la modulación del código pseudoaleatorio de la señal como se muestra en la figura 2.2. Dentro del receptor se genera una copia local de la secuencia pseudoaleatoria, que se compara y alinea continuamente con la señal recibida del satélite. Este bucle de seguimiento proporciona medidas continuas de la fase de código instantánea y, por tanto, del tiempo de transmisión corres-



**Figura 2.2:** Principio de medida del tiempo de propagación comparando los códigos pseudoaleatorios

pondiente a la señal recibida. Comparando este tiempo con el tiempo del receptor local, se obtiene el tiempo de propagación de la señal.

El tiempo de propagación medido, se multiplica por la velocidad de la luz para obtener la distancia entre el receptor y el satélite. Las distancias obtenidas se combinan para estimar la posición del receptor. Si el  $j$ -ésimo satélite transmite un pulso en  $t_0$ , y se recibe en el tiempo  $t_0 + \tau$ , la distancia  $R_j$  entre el transmisor ( $j$ -ésimo satélite) y el receptor se puede estimar como:

$$R_j = c \cdot \tau \quad (2.1)$$

En general, las señales GNSS permiten tres tipos básicos de medidas:

- *Seudorango*: una medida de la diferencia entre el reloj del receptor en la recepción de la señal y el reloj del satélite en la transmisión de la señal (escalado por la velocidad de la luz). Excepto por la asincronicidad de los dos relojes y algunos otros retrasos, el pseudorango mide la distancia satélite-receptor.
- *Fase de la portadora*: la medición de la fase de la portadora es una medida del rango entre un satélite y un receptor expresada en unidades de ciclos de la frecuencia de la portadora. Esta medición se puede realizar con una precisión muy alta (del orden de milímetros), pero no se puede medir el número total de ciclos entre el satélite y el receptor. Una buena analogía con esto es imaginar una cinta métrica que se extiende desde el satélite hasta el receptor que tiene marcadores numerados cada milímetro. Pero, el esquema de numeración vuelve a cero con cada longitud de onda (aproximadamente 20 centímetros para GPS L1). Esto nos permite medir el rango con mucha precisión,

pero con una ambigüedad en el número de ciclos completos de portadora. En caso de que se interrumpa el seguimiento, el recuento de ciclos acumulados se pierde y las mediciones de la fase portadora presentan errores.

- *Doppler*: el cambio en la frecuencia de la portadora recibida causado por el efecto Doppler es una medida de la velocidad mejor que diferenciando dos puntos de posición. También puede ser usada para reducir el ruido en las medidas de pseudorango y aumentar la precisión.

En el espacio tridimensional, cada distancia  $R_j$  define una superficie esférica cuyo centro es la posición del  $j$ -ésimo satélite. A través de la intersección de al menos tres de estas esferas, es posible calcular un punto que representa una posición precisa del usuario. Estrictamente hablando, tres esferas se cruzan en dos puntos, pero un punto de intersección puede rechazarse fácilmente debido al hecho de que está ubicado en una ubicación inverosímil (por ejemplo, en el espacio profundo).

Sin embargo, en una situación real, el reloj del receptor no está sincronizado con el transmisor. Si bien todos los satélites poseen relojes sincronizados, no es posible alinear con bajo costo y complejidad los relojes de los usuarios con los de los satélites. Además, los GNSS están concebidos para ser sistemas unidireccionales (es decir, que no requieren que el usuario interactúe con la constelación de satélites), por lo que no requieren métodos bidireccionales para mantener los satélites y los receptores alineados en la misma escala de tiempo.

Por esta razón, la medida de la distancia sufre un sesgo que es común a cada satélite porque representa el desplazamiento de la escala de tiempo del receptor con respecto a la escala de tiempo del GNSS. La medición realizada por el receptor se denomina pseudorango  $\rho$  y se define como la suma de la distancia real  $R_j$  y un término debido a la falta de alineación de la escala de tiempo. Analíticamente, la pseudodistancia para el  $j$ -ésimo satélite se puede escribir como

$$\rho_j = R_j + \varepsilon = R_j + c \cdot \delta t_r \quad (2.2)$$

donde  $c$  es la velocidad de la luz y  $\delta t_r$  es el sesgo del reloj del receptor. Este término no puede estimarse de las señales disponibles, por lo tanto se agrega la esfera de un cuarto satélite para resolver esa incógnita adicional. El pseudorango  $j$ -ésimo puede ser escrito como

$$\rho_j = \sqrt{(x_{sj} - x_r)^2 + (y_{sj} - y_r)^2 + (z_{sj} - z_r)^2} + b_{rt} \quad (2.3)$$

donde  $[x_r, y_r, z_r]$  son las coordenadas del receptor,  $[x_{sj}, y_{sj}, z_{sj}]$  son las coordenadas del  $j$ -ésimo satélite y  $b_{rt} = c \cdot \delta t_r$  es el sesgo del reloj.

La intersección de cuatro esferas correspondiente a cuatro satélites nos da el siguiente sistema de ecuaciones:

$$\begin{cases} \rho_1 = \sqrt{(x_{s1} - x_r)^2 + (y_{s1} - y_r)^2 + (z_{s1} - z_r)^2} + b_{rt} \\ \rho_2 = \sqrt{(x_{s2} - x_r)^2 + (y_{s2} - y_r)^2 + (z_{s2} - z_r)^2} + b_{rt} \\ \rho_3 = \sqrt{(x_{s3} - x_r)^2 + (y_{s3} - y_r)^2 + (z_{s3} - z_r)^2} + b_{rt} \\ \rho_4 = \sqrt{(x_{s4} - x_r)^2 + (y_{s4} - y_r)^2 + (z_{s4} - z_r)^2} + b_{rt} \end{cases} \quad (2.4)$$

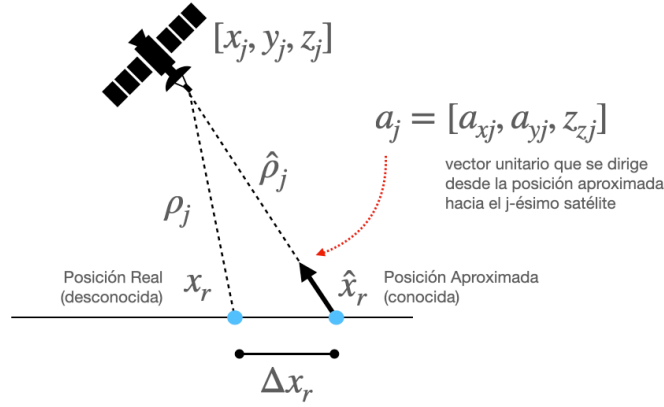
La solución de este sistema en el receptor provee su localización y el valor del sesgo temporal, lo que le permite sincronizar su propio receptor con la escala de tiempo GNSS. De nuevo, para que un receptor estime su posición y el sesgo del reloj, el receptor debe tener al menos cuatro satélites a la vista o en línea de visión. Si hay un mayor número de satélites a la vista, es posible una mejor estimación. En el pasado, debido a limitaciones de cálculo, se elegía la combinación de los cuatro satélites que ofrecían el mejor rendimiento. Los receptores modernos pueden usar incluso más de 12 canales (explotando señales de múltiples GNSS) para realizar la estimación de posición y tiempo.

¿QUÉ PRECISIÓN TIENE EL GPS?. Esta es una pregunta que se hacen casi todos los recién llegados al GPS. ¿Y la respuesta? Depende. Depende de si hablamos de posicionamiento autónomo (un solo receptor) o diferencial, de receptores de frecuencia única o doble, de funcionamiento en tiempo real o pos-procesado, etc. Incluso si nos limitamos al Servicio de Posicionamiento Estándar (SPS en inglés), el servicio oficial e independiente que el gobierno de Estados Unidos proporciona a todos los usuarios del mundo, la respuesta sigue siendo: depende.

La precisión especificada del SPE se da en términos de “niveles mínimos de rendimiento”; es decir, la precisión no será peor que un determinado nivel durante un determinado porcentaje de tiempo. Para cualquier punto del globo, la precisión horizontal es igual o mejor que 15 metros según el doble del error medio cuadrático. Esto significa que, en un periodo de 24 horas, las coordenadas horizontales de una posición determinada por el GPS estarán dentro de los 15 metros de la posición verdadera aproximadamente el 95 % de las veces. La precisión correspondiente para las alturas es de 33 metros y de 30 nanosegundos para la el tiempo.

Estas precisiones previstas se basan en una constelación de 24 satélites, un ángulo de máscara de elevación de los satélites de 5 grados sin obstrucciones y al menos cuatro satélites a la vista con una dilución de precisión de la posición (PDOP, ya veremos como se calcula) de seis o menos. Por lo tanto, incluso la precisión básica del SPS está cuantificada. Esto significa que, dependiendo de dónde nos encontremos y de la hora del día, la precisión real del SPS variará. En los cañones urbanos, es posible que ni siquiera tengamos cuatro satélites a la vista, y si los tenemos, la PDOP puede ser superior a seis.





**Figura 2.3:** Ejemplos de la linealización en una dimensión

La variabilidad de la precisión real del SPS de un lugar a otro y de un momento a otro está dominada, como veremos, por los efectos de la dilución de la precisión, un factor geométrico que cuando se multiplica por los errores de medición y otros errores de entrada, da el error en la posición y el tiempo.

### 2.1.2 Impacto de los Errores de Medición

Una vez que un receptor ha detectado con éxito cuatro (o más) satélites, puede usar los  $n$  pseudorangs medidos para resolver un conjunto de ecuaciones con cuatro incógnitas para obtener las coordenadas de posición  $[x_r, y_r, z_r]$  y el sesgo de reloj  $b_{rt}$ . Conociendo una aproximación de la posición y el sesgo verdaderos  $[\hat{x}_r, \hat{y}_r, \hat{z}_r, \hat{b}_{rt}]$  y expandiendo 2.3 en una serie de Taylor centrada en  $[\hat{x}_r, \hat{y}_r, \hat{z}_r, \hat{b}_{rt}]$  es posible obtener el error de posición  $[\Delta x_r, \Delta y_r, \Delta z_r, \Delta b_{rt}]$  como una función lineal de las coordenadas conocidas y de las medidas de pseudorango.

La figura 2.3 muestra en forma simplificada las diferentes variables involucradas en el proceso de linealización que sería como:

$$\begin{aligned}
 \rho_j &= f(x_r, y_r, z_r, b_{rt}) \\
 &= f(\hat{x}_r + \Delta x_r, \hat{y}_r + \Delta y_r, \hat{z}_r + \Delta z_r, \hat{b}_{rt} + \Delta b_{rt}) \\
 &= f(\hat{x}_r, \hat{y}_r, \hat{z}_r, \hat{b}_{rt}) + \frac{\partial f(\hat{x}_r, \hat{y}_r, \hat{z}_r, \hat{b}_{rt})}{\partial x_r} \Delta x_r + \frac{\partial f(\hat{x}_r, \hat{y}_r, \hat{z}_r, \hat{b}_{rt})}{\partial y_r} \Delta y_r \dots \\
 &\quad \frac{\partial f(\hat{x}_r, \hat{y}_r, \hat{z}_r, \hat{b}_{rt})}{\partial z_r} \Delta z_r + \frac{\partial f(\hat{x}_r, \hat{y}_r, \hat{z}_r, \hat{b}_{rt})}{\partial b_{rt}} \Delta b_{rt} + \dots \\
 &= \hat{\rho}_j - \frac{x_{sj} - \hat{x}_r}{\hat{r}_j} \Delta x_r - \frac{y_{sj} - \hat{y}_r}{\hat{r}_j} \Delta y_r - \frac{z_{sj} - \hat{z}_r}{\hat{r}_j} \Delta z_r + \Delta b_{rt} \\
 &= \hat{\rho}_j - a_{xj} \Delta x_r - a_{yj} \Delta y_r - a_{zj} \Delta z_r + \Delta b_{rt}
 \end{aligned} \tag{2.5}$$

donde  $\hat{r}_j = \sqrt{(x_{sj} - \hat{x}_r)^2 + (y_{sj} - \hat{y}_r)^2 + (z_{sj} - \hat{z}_r)^2}$  y  $\hat{\rho}_j = f(\hat{x}_r, \hat{y}_r, \hat{z}_r, \hat{b}_{rt})$ .

La expansión ha sido truncada en el término de primer orden para eliminar los términos no lineales. La diferencia de pseudorangos  $\Delta\rho_j = \hat{\rho}_j - \rho_j$  puede ser escrita como,

$$\Delta\rho_j = a_{xj}\Delta x_r + a_{yj}\Delta y_r + a_{zj}\Delta z_r - \Delta b_{rt} \quad (2.6)$$

donde  $1 \leq j \leq n$ , con  $n$  el número de satélites disponibles. Estas ecuaciones pueden ser puestas en forma matricial definiendo:

1.  $\Delta\mathbf{p} = [\Delta\rho_1, \Delta\rho_1, \dots, \Delta\rho_n]^T$  como el desplazamiento vectorial de los valores de pseudorangos sin errores (correspondientes a la posición real del usuario) y los valores de pseudorango correspondientes al punto de linealización;
2.  $\Delta\mathbf{x} = [\Delta x_r, \Delta y_r, \Delta z_r, \Delta b_{rt}]^T$  como el desplazamiento del vector desde el punto de linealización de la posición;
3.  $\mathbf{H}$  como la matriz geométrica que contiene, en las 3 primeras columnas, los vectores unitarios que apuntan desde el punto de linealización a cada  $j$ -ésimo satélite

donde,

$$\mathbf{H} = \begin{bmatrix} a_{x1} & a_{y1} & a_{z1} & -1 \\ a_{x1} & a_{y2} & a_{z2} & -1 \\ \vdots & \vdots & \ddots & \vdots \\ a_{xn} & a_{yn} & a_{zn} & -1 \end{bmatrix}. \quad (2.7)$$

Por lo tanto,

$$\Delta\mathbf{p} = \mathbf{H}\Delta\mathbf{x} \quad (2.8)$$

cuya solución es

$$\Delta\mathbf{x} = \mathbf{H}^{-1}\Delta\mathbf{p} \quad (2.9)$$

La ecuación 2.9 relaciona el desplazamiento  $\Delta\mathbf{x}$  en la posición real del usuario y el sesgo de tiempo, con respecto al punto de linealización con el desplazamiento en los valores de pseudorangos libres de error  $\Delta\mathbf{p}$ . Cuando se dispone de más de cuatro mediciones de pseudorango, se puede utilizar el método de los mínimos cuadrados para calcular el desplazamiento  $\Delta\mathbf{x}$ . Este método permite calcular el valor de  $\Delta\mathbf{x}$  para que  $\mathbf{H}\Delta\mathbf{x}$  se lo más cercano posible a  $\Delta\mathbf{p}$ . Para cualquier valor particular de  $\Delta\mathbf{x}$ , la cantidad vectorial  $\mathbf{r} = \mathbf{H}\Delta\mathbf{x} - \Delta\mathbf{p}$  se denomina residuo.

La solución ordinaria de mínimos cuadrados se define como el valor de  $\Delta\mathbf{x}$  que minimiza el cuadrado del residuo, que es

$$\mathbf{R}_{SE}(\Delta\mathbf{x}) = (\mathbf{H}\Delta\mathbf{x} - \Delta\mathbf{p})^2 \quad (2.10)$$

La solución a esta ecuación puede obtenerse derivando 2.10 con respecto a  $\Delta \mathbf{x}$  para obtener el gradiente de  $\mathbf{R}_{SE}$ . Por lo tanto, el gradiente se iguala a cero y se resuelve para  $\Delta \mathbf{x}$  para buscar un valor que minimice  $\mathbf{R}_{SE}$ . Su solución es:

$$\Delta \mathbf{x} = (\mathbf{H}^T \mathbf{H})^{-1} \mathbf{H}^T \Delta \boldsymbol{\rho} \quad (2.11)$$

La condición de que  $\mathbf{H}^T \mathbf{H}$  no sea singular equivale a la condición de que las puntas de los vectores unitarios desde el punto de linealización hasta los satélites no sean linealmente dependientes. Normalmente, cualquier receptor GNSS busca el mínimo de 2.11 utilizando métodos recursivos o un filtro de Kalman. Una vez que las incógnitas son calculadas, se obtienen las coordenadas del usuario  $[x_r, y_r, z_r]$  y el desplazamiento del reloj  $b_{rt}$ .

### 2.1.3 El impacto de la geometría

Para entender el impacto de las relaciones geométricas entre los satélites y el móvil, debe tenerse en cuenta que las mediciones de pseudorango no están libres de errores. En general, las mediciones entre el usuario y el satélite están corrompidas por errores independientes producidas por distintos fenómenos sobre la señal que atraviesa distintas capas de la atmósfera. El conjunto de ecuaciones que hay que resolver debe escribirse como

$$\Delta \boldsymbol{\rho} + \delta \boldsymbol{\rho} = \mathbf{H}(\Delta \mathbf{x} + \delta \mathbf{x}) \quad (2.12)$$

El error en el pseudorango  $\delta \boldsymbol{\rho}$  puede ser relacionado con el error en posición  $\delta \mathbf{x}$  como

$$\delta \mathbf{x} = ((\mathbf{H}^T \mathbf{H})^{-1} \mathbf{H}^T) \delta \boldsymbol{\rho} \quad (2.13)$$

donde  $\delta \boldsymbol{\rho}$  representa el error neto en los valores de pseudorango. Estos errores son considerados variables aleatorias y expresa a  $\delta \mathbf{x}$  como una variable aleatoria funcionalmente relacionada con  $\delta \boldsymbol{\rho}$ . La covarianza de  $\delta \mathbf{x}$  se puede obtener a partir del producto  $\delta \mathbf{x} \delta \mathbf{x}^T$  y calculando la esperanza,

$$\text{cov}(\delta \mathbf{x}) = E \left\{ (\mathbf{H}^T \mathbf{H})^{-1} \mathbf{H}^T \delta \boldsymbol{\rho} \delta \boldsymbol{\rho}^T \mathbf{H} (\mathbf{H}^T \mathbf{H})^{-1} \right\} \quad (2.14)$$

$$\text{cov}(\delta \mathbf{x}) = (\mathbf{H}^T \mathbf{H})^{-1} \mathbf{H}^T \text{cov}(\delta \boldsymbol{\rho}) \mathbf{H} (\mathbf{H}^T \mathbf{H})^{-1} \quad (2.15)$$

Con la hipótesis de que las contribución al error del pseudorango puede ser modelado como una variable aleatoria Gaussiana de media cero, idénticamente distribuida, independiente y con varianza conocida  $\sigma_\rho^2$ ,

$$\text{cov}(\delta \boldsymbol{\rho}) = \mathbf{I}_{n \times n} \sigma_\rho^2 \quad (2.16)$$

donde  $\mathbf{I}_{n \times n}$  es la matriz identidad de  $n \times n$  elementos. Cuando combinamos el ruido del receptor, del reloj satelital, el error de efemérides, el error atmosférico, y el multicamino, todos expresados en unidades de distancia, obtenemos una cantidad conocida como el error de rango equivalente total del usuario (UERE en inglés), que en estas ecuaciones expresamos como  $\sigma_\rho^2$ . Para el SPS el  $\sigma_\rho^2$  total suele ser inferior a 5 metros, y el valor real está dominado por los efectos ionosféricos y multicamino. La matriz de covarianza resultante será:

$$\text{cov}(\delta\mathbf{x}) = (\mathbf{H}^T \mathbf{H})^{-1} \sigma_\rho^2 = \begin{bmatrix} \sigma_{x_r}^2 & \sigma_{x_r, y_r}^2 & \sigma_{x_r, z_r}^2 & \sigma_{x_r, b_{rt}}^2 \\ \sigma_{x_r, y_r}^2 & \sigma_{y_r}^2 & \sigma_{y_r, z_r}^2 & \sigma_{y_r, b_{rt}}^2 \\ \sigma_{x_r, z_r}^2 & \sigma_{x_r, z_r}^2 & \sigma_{z_r}^2 & \sigma_{z_r, b_{rt}}^2 \\ \sigma_{x_r, b_{rt}}^2 & \sigma_{y_r, b_{rt}}^2 & \sigma_{z_r, b_{rt}}^2 & \sigma_{b_{rt}}^2 \end{bmatrix} \quad (2.17)$$

donde los elementos que están fuera de la diagonal principal representan el nivel de correlación cruzada entre las variables.

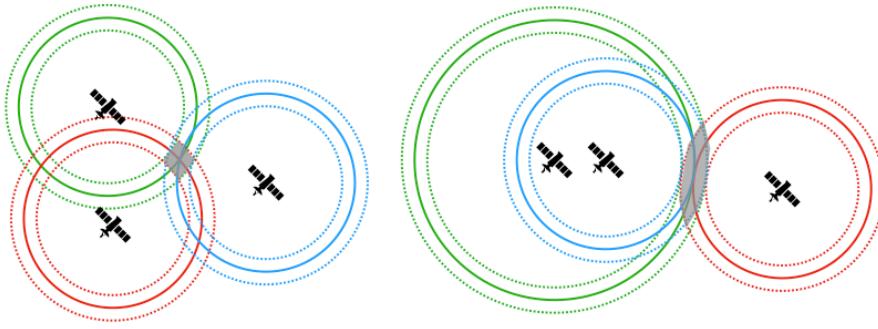
El parámetro mas general utilizado para evaluar al impacto de la geometría en la exactitud final de la posición estimada se denomina *dilución de la precisión geométrica* (GDOP en inglés) que se define como:

$$\text{GDOP} = \frac{\sqrt{\sigma_{x_r}^2 + \sigma_{y_r}^2 + \sigma_{z_r}^2 + \sigma_{x_{rt}}^2}}{\sigma_\rho} \quad (2.18)$$

Este factor geométrico, GDOP, representa un factor de escala de la desviación estándar del error de las mediciones sobre la solución. Como este factor de escala suele ser mayor que uno, amplifica el error del pseudorango, o, diluye la precisión de la determinación de la posición. Podemos simplificar este cálculo introduciendo la varianza del pseudorango en la raíz de la ecuación 2.18,

$$\text{GDOP} = \sqrt{\text{tr}\{(\mathbf{H}^T \mathbf{H})^{-1}\}} \quad (2.19)$$

El GDOP es una medida de cuánto depende el error de posición estimado de los errores de medición del rango y la geometría relativa usuario/satélite. Este efecto geométrico puede entenderse analizando la Figura 2.4, donde se muestran dos escenarios diferentes. Para simplificar, se consideran medidas en 2D. En ambos casos, los anillos libres de errores se cruzan en la ubicación del usuario (líneas continuas). Los anillos adicionales (líneas discontinuas) representan el límite de  $1\sigma$  debido a un error Gaussiano en el rango. La región de incertidumbre es el lugar de las posibles soluciones de navegación en función del proceso aleatorio de error que es el mismo para ambos casos. Sin embargo, el área de la región de incertidumbre es mayor en el segundo caso, en el que se considera un caso peor en términos de geometría (mayor GDOP), y esto sin dudas lleva a una mayor incertidumbre de posición.



**Figura 2.4:** Ejemplos de como opera la dilución geométrica de precisión

Además de esta dilución geométrica, se pueden definir otras diluciones parciales y que se usan comúnmente para analizar la exactitud de varios componentes de la solución. Si definimos la matriz  $\mathbf{G} = (\mathbf{H}^T \mathbf{H})^{-1}$

1. Dilución de precisión en posición (PDOP en inglés)

$$\text{PDOP} = \sqrt{g_{11} + g_{22} + g_{33}} \quad (2.20)$$

2. Dilución de precisión horizontal (HDOP en inglés)

$$\text{PDOP} = \sqrt{g_{11} + g_{22}} \quad (2.21)$$

3. Dilución de precisión vertical (VDOP en inglés)

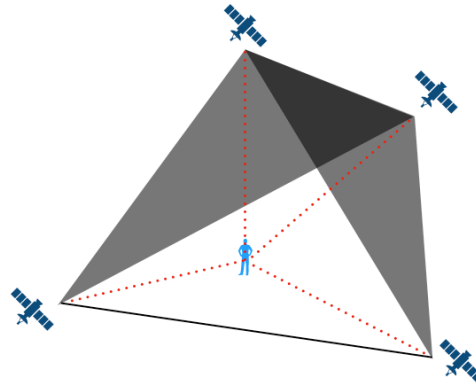
$$\text{PDOP} = \sqrt{g_{33}} \quad (2.22)$$

4. Dilución de precisión temporal (TDOP en inglés)

$$\text{PDOP} = \sqrt{g_{44}} \quad (2.23)$$

El mejor caso para el GDOP es cuando las puntas de los cuatro vectores unitarios receptor-satélite forman un tetraedro. Cuanto mayor sea el volumen del tetraedro, menor será el GDOP. Por supuesto, un receptor GNSS en la superficie de la tierra no puede ver los satélites por debajo del horizonte, por lo que nunca se alcanza este caso ideal. Si se utiliza un número  $n$  de satélites, con  $n > 4$ , en la solución, las puntas de los  $n$  vectores unitarios receptor-satélite forman un sólido más amplio, cuyo volumen es mayor. En consecuencia, los valores de GDOP son menores y, por tanto, el error de la solución es menor (para el mismo error de alcance).

En general, cuantos más satélites se utilicen en la solución, menores serán los valores de DOP y, por tanto, menor será el error de la solución. Para Los valores HDOP suelen estar entre uno y dos. Los valores VDOP son mayores que los valores HDOP (fig. 2.6 de [29]),



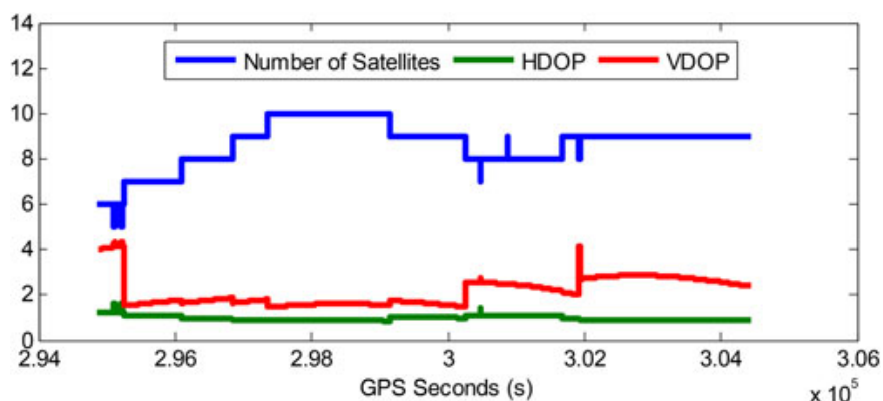
**Figura 2.5:** Si sólo se observan cuatro satélites, las puntas de los vectores unitarios receptor-satélite forman un tetraedro circunscrito por una esfera unitaria. Dos caras del tetraedro -formadas por un satélite en el cenit y tres en un ángulo de elevación de 10 grados, igualmente espaciados en acimut- están sombreadas en esta figura. El volumen del tetraedro está altamente correlacionado con el GDOP. La maximización del volumen tiende a minimizar el GDOP.

lo que indica que los errores de posición vertical son mayores que los horizontales. Sufrimos este efecto porque todos los satélites de los que obtenemos señales están por encima del receptor (la tierra oculta satélites por abajo). Las coordenadas horizontales no sufren ese problema ya que habitualmente recibimos señales de todos los lados.

Interesante es relacionar todos estos valores que estamos viendo. Si el  $\sigma_p^2$  es de 5 metros y el HDOP es de 2, el error medio cuadrático será de 10 metros y, en un nivel de incertidumbre del 95 %, este error será de 20 metros.

El  $\sigma_p$  o  $\sigma_{\text{UERE}}$  se obtiene evaluando los aportes de error en cada enlace. Este ha sido investigado a través de los años en muchos trabajos científicos que analizan los aspectos físicos, de modelado y de mitigación. Las fuentes de este error pueden ser clasificados en

1. *Sistema de Control*: errores de efemérides, relojes, códigos
2. *Ionosfera*: el retardo de propagación depende en la frecuencia y en la densidad de electrones a lo largo del camino de la señal. Los errores mas grandes ocurren en el caso de condiciones anormales de la ionosfera, como el centelleo.
3. *Troposfera*: el retardo de propagación depende de la presión, temperatura y humedad del aire.
4. *Multicamino*: error debido a la posible presencia de replicas de la señal por la reflexión en estructuras y que se suman a la señal directa



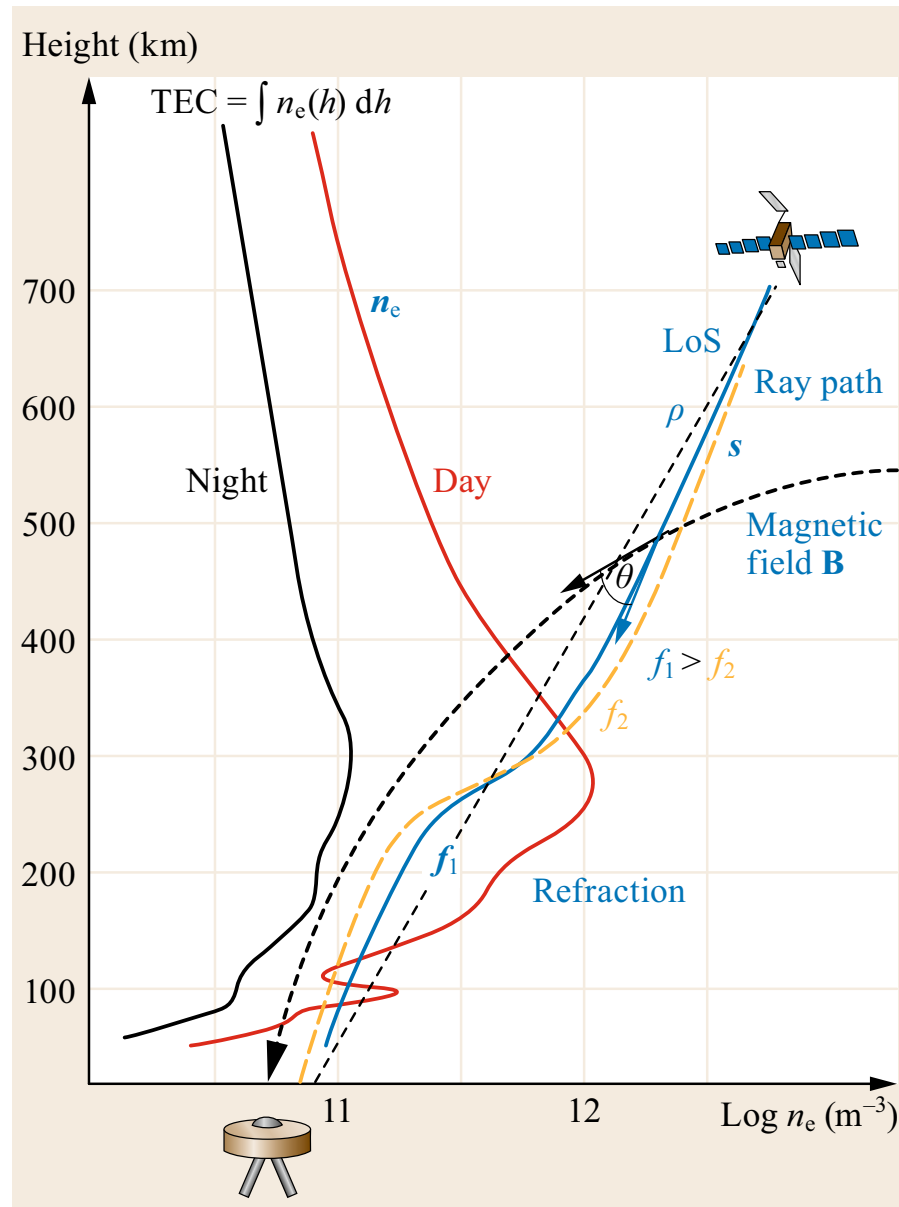
**Figura 2.6:** El número de satélites a la vista y el HDOP y VDOP calculado (de [29]).

5. *Ruido del receptor:* debido al ruido térmico de los componentes de hardware del receptor (recordemos que la sensibilidad típica es de  $-150$  dBm o  $10^{-15}$  mW o 1 femto mili watts o 1 atto watts)
6. *Efectos relativistas sin compensar*

Para mejorar la precisión de la solución, para estas fuentes de error conocidas, cada receptor GNSS compensa en la medida de lo posible los efectos de la contribución “media” de cada una. Este resultado se obtiene utilizando modelos de las fuentes de error o, en receptores más complejos, algoritmos de procesamiento específicos. Sólo el error debido a la ionosfera puede cancelarse casi por completo si se utiliza un receptor de doble frecuencia. La figura 2.7 muestra un ejemplo de propagación de una señal emitida por un satélite y que llega al receptor para dos frecuencias diferentes  $f_1$  y  $f_2$ . El camino directo se muestra en línea punteada. Los cambios mas importantes en el camino de la señal se dan entre los 200 y 300Km de altitud aproximadamente que es la zona ionósfera. Allí la cantidad de electrones (TEC: Total Electron Content) se muestra en línea continua roja y negra y son los que afectan a la refracción de la señal. Al afectar diferente ambas frecuencias, puede utilizarse para reducir ese error.

#### 2.1.4 Modos de posicionamiento

Hay varios modos de posicionamiento (y navegación) usando GNSS, con diferentes grados de complejidad y precisión y exactitud. Estos van desde el método estándar basado en pseudorange de frecuencia única utilizado por la mayoría de los receptores de consumo masivo, incluidos los de teléfonos móviles, pasando por métodos de alta integridad para aplicaciones donde esté en riesgo la vida, hasta técnicas sofisticadas basadas en fase portadora multifrecuencia capaces



**Figura 2.7:** Esquema de propagación de la señal de RF a través de la ionosfera para dos frecuencias en presencia del campo geomagnético y el número de electrones (de [26]).



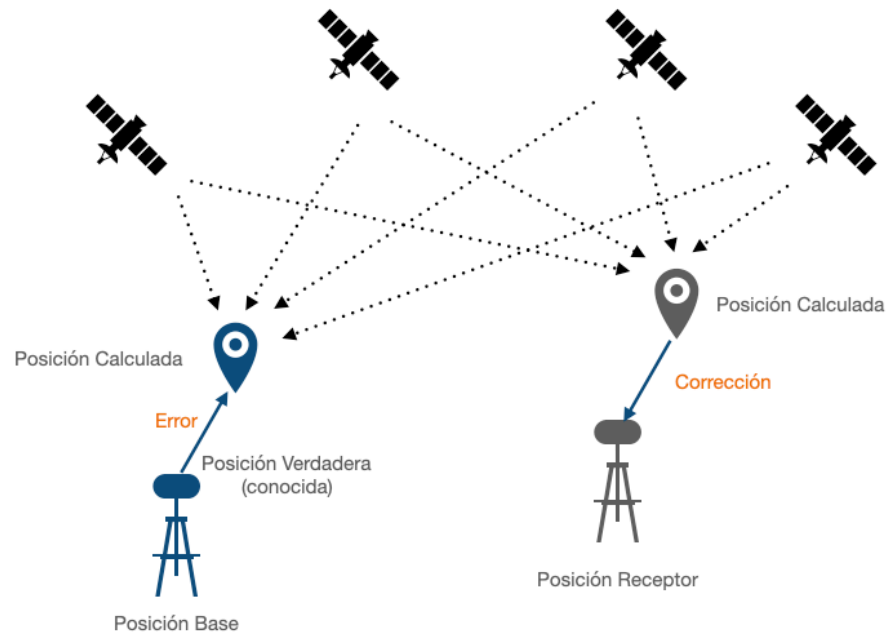
de precisiones del subcentímetro para aplicaciones exigentes como control de equipos autónomos y estudios científicos.

#### 2.1.4.1 PPP: *Precise Point Positioning*

El posicionamiento de punto preciso (PPP) es una versión avanzada de la técnica de posicionamiento común que hemos tratado hasta acá. El PPP utiliza las mediciones de la fase de la portadora como observación principal y las mediciones de pseudodistancia desempeñan un papel secundario. Se combinan las posiciones precisas de los satélites y los relojes, post procesado o en tiempo real, con el uso de dos frecuencias en el receptor para remover el efecto ionosférico de primer orden. Se estiman las ambigüedades de la fase de la portadora (resolviéndolas a valores enteros si es posible), así como el retardo residual de la propagación troposférica tras aplicar un modelo a priori. También se modifican los efectos sutiles, como las mareas terrestres, la carga de las mareas oceánicas, las desviaciones de las antenas de los satélites y de los receptores. A diferencia de otros métodos, que no utilizan estas correcciones, tampoco requiere de estaciones adicionales con posiciones conocidas para generar las correcciones. Esto ultimo elimina a los limites espaciales de operación del método. Por esto mismo es un método que provee mayor consistencia dado que su solución es relativa a los ejes de referencia globales y no son locales relativa a las coordenadas de las estaciones de referencia.

El rendimiento del PPP puede medirse en términos de exactitud, precisión, periodo de convergencia (el tiempo necesario para que una solución de posición converja por debajo de un determinado umbral de exactitud), disponibilidad e integridad. La mejor solución de PPP no suele tener sesgos, por lo que hay poca diferencia entre las estadísticas de exactitud y precisión. La precisión de 1 sigma es de unos pocos centímetros en cada coordenada (norte, este y arriba) para un sitio estático después de la convergencia, mientras que se pueden lograr precisiones de un decímetro para una plataforma móvil. El periodo de convergencia para conseguir una solución a nivel decimétrico suele ser de unos 30 minutos en condiciones normales. Los bloqueos de la señal por árboles y edificios pueden reducir la disponibilidad, pero las observaciones de multiconstelación pueden ser de gran ayuda en este sentido.

Este tiempo de inicialización es una de las limitaciones para su aplicación. Además, la determinación de la fase está corrupta por los sesgos en el satélite y el receptor, lo que limita la precisión. La necesidad de acceder a las órbitas precisas y su disponibilidad, es parte de los retardos en la convergencia. Incluso puede ser una fuente de cargos económicos provistas por privados.



**Figura 2.8:** Esquema de funcionamiento del posicionamiento diferencial por código. La ER conociendo su posición determina el error en los rangos de los satélites y comunica estos errores al receptor para que corrija la posición calculada con menos error.

#### 2.1.4.2 Code Differential Positioning

La ventaja del posicionamiento diferencial sobre el método estándar es que con las técnicas diferenciales se eliminan o reducen en gran medida ciertos efectos (por ejemplo, los errores de órbita y los retrasos atmosféricos que dependen de la correlación espacial). Hay dos tipos básicos de posicionamiento diferencial de código: locales o de grandes áreas (que ampliaremos en la sección 2.1.8).

Las técnicas locales proporcionan correcciones de medición al usuario sin estimar los componentes de error individuales, mientras que las técnicas de grandes áreas proporcionan correcciones de error individuales, como la órbita del satélite y el reloj y el retardo de la propagación ionosférica. Estas correcciones se determinan en las estaciones de referencia y se transmiten a los usuarios mediante enlaces de RF. El usuario necesita un receptor GNSS con un receptor de RF integrado o un receptor de RF separado conectado al receptor GNSS mediante un enlace de comunicaciones en serie como RS-232. La figura 2.8 muestra la esencia de este método, donde la ER que conoce su posición, determina el error en el pseudorango de cada uno de los satélites, y este error es el que se transmite al receptor para que corrija cada pseudorango y obtenga su posición.

La precisión de la posición generalmente se degrada con el aumento de la distancia desde el sitio de transmisión de la baliza. La preci-

sión oficial es de 10 m (horizontal al 95 %) dentro del área de cobertura, pero normalmente el error de una posición DGPS es de 1 a 3 m. El error suele verse como un sesgo en el posicionamiento, que da lugar a un desplazamiento de la posición. Es probable que la dispersión de las coordenadas se mantenga casi constante. En otras palabras no tiene buena exactitud (desplazamiento) pero si mejor precisión (dispersión). Una regla general es un error de 1 m adicional por cada 100 km. Sin embargo, la precisión es peor durante fuertes perturbaciones ionosféricas.

#### 2.1.4.3 Differential Carrier Phase

El posicionamiento diferencial de la fase portadora es una técnica clásica que data de principios de los años ochenta. El procedimiento combina los datos de una (o más) estaciones de referencia con los datos del usuario. Las observaciones del mismo satélite en el mismo momento se diferencian entre los receptores (del móvil y de las ER, diferencia simple) y luego las diferencias simples se diferencian entre pares de satélites (diferencia doble). Este procedimiento elimina los errores residuales del reloj del satélite y del receptor, y reduce los errores de la órbita del satélite y los errores de retardo de la propagación atmosférica. Se pueden obtener precisiones a nivel decimétrico y mejores. En principio, se podría adoptar un enfoque similar utilizando pseudorangos (solos), pero con exactitudes resultantes mucho más bajas.

**DIFERENCIACIÓN SIMPLE.** Consideremos una señal sinusoidal transmitida de un satélite  $s$  en el instante  $t_{tx}$  y recibida en el receptor en el instante  $t_{rx}$ . Esta señal puede ser representada como,

$$y_r^s(t_{rx}) = A_r^s(t_{rx})e^{2j\pi\phi_r^s(t_{rx})} \quad (2.24)$$

donde  $A_r^s(t_{rx})$  es la amplitud recibida y  $\phi_r^s(t_{rx})$  es la fase recibida en ciclos. Para una onda plana que se propaga libre en el espacio, la fase de la señal recibida está dada por:

$$\phi_r^s(t_{rx}) = f_{RF}t_{rx} + \phi^s(0) - \frac{r(t_{rx})}{\lambda} \pmod{1} \quad (2.25)$$

donde  $f_{RF}$  es la frecuencia de transmisión en Hertz (por ejemplo para GPS L1  $f_{RF} = 1575,42\text{MHz}$ ),  $\phi^s(0)$  es la fase inicial del transmisor,  $\lambda = c/f_{RF}$  es la longitud de onda y  $r(t_{rx}) = (t_{rx} - t_{tx})c$  es la distancia del satélite al receptor.

La clave es que la fase recibida de la portadora, tiene información de distancia entre el satélite y el receptor. Puede entonces interesar extraer esta información que idealmente puede ser expresada en unidades de ciclos. Reagrupando la ecuación anterior:

$$\frac{r(t_{rx})}{\lambda} = f_{RF}t_{rx} + \phi^s(0) - [\phi_r^s(t_{rx}) + M_r^s(t_{rx})] \quad (2.26)$$

donde  $M_r^s(t_{rx})$  es el número de longitud de ondas (ciclos) reales entre el satélite y el receptor. Si el receptor es capaz de estimar la fase recibida, quedan tres incógnitas que deben ser estimadas para determinar el rango. Estas incógnitas son:

1. El tiempo del receptor  $t_{rx}$
2. La fase inicial del satélite  $\phi^s(0)$
3. El número de ciclos enteros entre el satélite y el receptor  $M_r^s(t_{rx})$

El sistema de diferenciación simple, se restan las medidas de dos receptores, y se remueve los errores en la fase del satélite y los errores de reloj en el receptor (presente en  $t_{rx}$ ).

En general, el receptor genera un estimado de la fase de la portadora, integrando la frecuencia de Doppler.

**DIFERENCIACIÓN DOBLE.** En este caso, siguiendo con el razonamiento anterior, si diferenciamos la fase entre receptores y luego la restamos entre satélites, tenemos una doble diferenciación que remueve los errores en los relojes del receptor y satélites, la fase inicial y la cantidad de ciclos enteros se reemplaza por otro parámetro entero. Para procesar esta información se modela esta doble diferencia geométrica junto a la doble diferencia del retardo en la troposfera, el efecto de la ionosfera y se utiliza generalmente un estimador como el filtro de Kalman.

#### 2.1.4.4 RTK: Real Time Kinematic.

Este método combina diferenciación de fase con diferenciación local. Una estación base transmite por un enlace de RF adicional ambos datos para corregir en el receptor los errores y lograr exactitudes del orden del centímetro. Estas correcciones tienen validez de 10 o 20 Kms de distancia a la base.

Los datos de la estación de referencia también se transmiten usando protocolos de datos de la Comisión Técnica de Radio para Servicios Marítimos (RTCM) versiones 2.x o 3.x, aunque también existen formatos de datos propietarios. A través de Internet se utiliza el Transporte de red del Protocolo de Internet RTCM (NTRIP) y se puede acceder a ellos mediante un enlace cableado o inalámbrico, como un teléfono móvil. Existen varias técnicas para reducir los efectos de la latencia, los desvíos de ciclo y las ambigüedades.

#### 2.1.4.5 Network RTK

El inconveniente mas importante de la RTK de base única, es que la distancia máxima entre la estación de referencia y el receptor del rover no debe superar los 10-20 km para poder resolver de forma rápida y fiable las ambigüedades de la fase portadora. Esta limitación está

causada por los sesgos dependientes de la distancia, principalmente la refracción de la señal ionosférica, pero también los errores de órbita y la refracción troposférica. Estos errores, sin embargo, pueden ser modelados con precisión utilizando las mediciones de un conjunto de estaciones de referencia GNSS que rodean el sitio del rover. Por lo tanto, la solución a la limitación de distancia de RTK se encuentra en las técnicas multibase que se hicieron populares bajo el nombre de Network RTK (NRTK). De hecho, NRTK también tiene una limitación de distancia. Esta limitación se refiere a las distancias entre las estaciones de referencia. No deberían superar los 100-200 km para poder producir modelos de corrección en tiempo real de gran precisión de los errores dependientes de la distancia.

La técnica NRTK permitió el establecimiento de servicios de posicionamiento que sirven a regiones más grandes o a países enteros mediante la creación y el mantenimiento de redes de estaciones de referencia, la recolección y el preprocesamiento de sus observaciones y la distribución de correcciones de observación en tiempo real a los usuarios de RTK. Por ejemplo un área de 100.000 km<sup>2</sup> requiere una red de unas 20 estaciones de referencia separadas por 75 km.

#### 2.1.4.6 PPP-RTK

Con el método de PPP, se proporcionan órbitas y relojes de satélite precisos para que los usuarios de un solo receptor puedan calcular las posiciones de sus receptores con una gran precisión, decimétrica o centimétrica. PPP-RTK amplía el concepto de PPP proporcionando a los usuarios de un solo receptor, además de las órbitas y los relojes, información sobre los sesgos de fase de los satélites. Esta información, cuando se proporciona adecuadamente, permite recuperar la integridad de las ambigüedades del usuario reduciendo así los tiempos de convergencia en comparación con los de PPP.

Las dos técnicas por separado, podrían resumirse de la siguiente manera muy simplista (donde "localización precisa" significa la localización exacta inferida a partir de las mediciones de código y fase, que es mucho más precisa que la localización inferida a partir de las mediciones de código solamente):

**RTK** Dada la ubicación precisa de la estación de referencia, puedo encontrar la ubicación precisa del receptor

**PPP** Dada la ubicación precisa de los satélites de navegación, puedo encontrar la ubicación precisa del receptor

El concepto de PPP asociado a redes RTK, la combinación de ambos métodos, da la posibilidad de lograr una precisión del centímetro en segundos. Es interesante resaltar que la reducción del tiempo de convergencia del método PPP puede ser logrado con el incremento de la capacidad de cálculo, con el uso de diferentes constelaciones

PPP, o como en este método con la inclusión de bases RTK teniendo en cuenta que mientras el método PPP es un método global, el RTK es local y cercano a las estaciones de referencia.

#### 2.1.5 *Sistemas de Referencia*

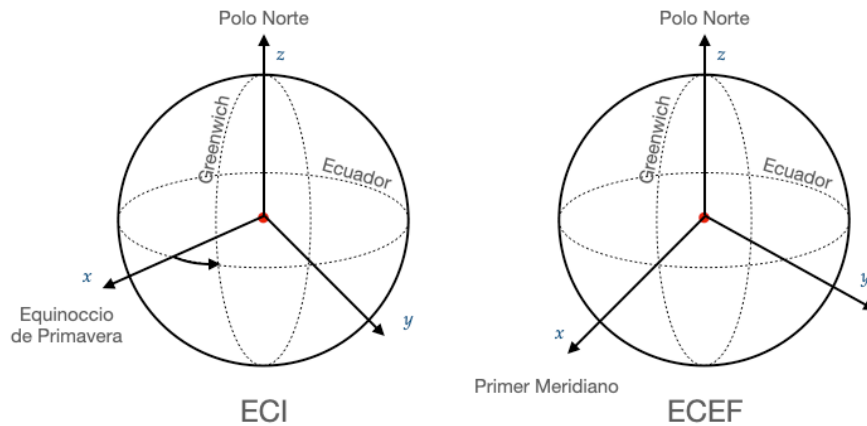
Para formular las matemáticas del problema de la navegación por satélite, es necesario elegir un sistema de coordenadas de referencia en el que se puedan representar los estados tanto del satélite como del receptor. En esta formulación, es típico describir los estados del satélite y del receptor en términos de vectores de posición y velocidad medidos en un sistema de coordenadas cartesianas. Para calcular la distancia entre un usuario y un satélite, es necesario tener la posición de ambos en un sistema de referencia común.

Hay varios sistemas de coordenadas cartesianos de uso común, incluidos los sistemas inerciales y rotativos. Al elegir el “mejor” sistema de referencia, se deben tener en cuenta dos requisitos. La posición del usuario se expresa en forma convencional en un sistema de coordenadas fijo a la tierra y se mueve con ella, de modo que un objeto estacionario permanece fijo. Por otro lado el movimiento de los satélites está regulado por ecuaciones de movimiento, generalmente expresadas en un sistema inercial (es decir, fijo en el espacio o en movimiento uniforme).

**SISTEMA DE REFERENCIA INERCIAL.** A los efectos de medir y determinar las órbitas de los satélites, es conveniente utilizar un sistema de coordenadas inercial centrado en la Tierra (ECI en inglés), en el que el origen es el centro de masa de la Tierra. Un sistema ECI es inercial en el sentido de que las ecuaciones de movimiento de un satélite en órbita terrestre pueden modelarse como si el sistema ECI no estuviera acelerado. En otras palabras, un satélite GNSS obedece a las leyes de movimiento y gravitación de Newton en un sistema de coordenadas ECI. En los sistemas de coordenadas ECI típicos (fig. 2.9), el plano  $xy$  se toma para que coincida con el plano ecuatorial de la Tierra, el eje  $+x$  está permanentemente fijo hacia el equinoccio de primavera (es decir, la dirección de intersección del plano ecuatorial de la Tierra con el plano de la órbita de la tierra alrededor del sol), el eje  $+z$  se toma normal al plano  $xy$  en la dirección del polo norte a lo largo del eje de rotación, y el eje  $+y$  se elige para formar un sistema de coordenadas a la derecha. La órbita de los satélites GNSS se referencia a este sistema de coordenadas ECI.

Un problema en la definición de un sistema de coordenadas ECI surge por las irregularidades en el movimiento de la tierra. La forma de la tierra es achatada, y debido en gran parte a la atracción gravitacional del sol y la luna en la protuberancia ecuatorial de la tierra, el plano ecuatorial se mueve con respecto a la esfera celeste.





**Figura 2.9:** Sistemas de Referencia Inercial (ECI) y Terrestre (ECEF).

Debido a que el eje  $x$  se define en relación con la esfera celeste y el eje  $z$  se define en relación con el plano ecuatorial, las irregularidades en el movimiento de la Tierra harían que el marco ECI definido anteriormente no sea verdaderamente inercial. De hecho, el centro de masa de la tierra se mueve alrededor del sol (segunda ley de Kepler). Además, el eje de rotación de la tierra no está fijo en el espacio en relación a las estrellas distantes debido a la atracción gravitacional del sol y la luna y que no estarían presentes si la tierra fuera esférica y homogénea. Sin embargo estos efectos, las precesiones y nutación del eje de rotación de la Tierra, son fenómenos lentos con períodos de 26.000 años y 18,6 años, respectivamente, y se tienen en cuenta en la definición de los marcos de referencia.

En la definición del sistema ECI para GPS, la solución de este problema ha sido definir la orientación de los ejes en un instante o época particular. El sistema de coordenadas GPS ECI utiliza la orientación del plano ecuatorial a las 12:00 horas UTC del 1 de enero de 2000, como base. El eje  $x$  se toma para apuntar desde el centro de masa de la Tierra a la dirección del equinoccio vernal, y los ejes  $y$  y  $z$  se definen como se describió anteriormente, todo en la época antes mencionada. Dado que la orientación de los ejes permanece fija, el sistema de coordenadas ECI definido de esta manera puede considerarse inercial para fines de GPS.

**SISTEMA DE REFERENCIA TERRESTRE.** Para calcular la posición de un receptor GNSS, es más conveniente utilizar un sistema de coordenadas que gire con la tierra, conocido como sistema centrado y fijo en la tierra (ECEF en inglés). En un sistema de coordenadas de este tipo, es más fácil calcular los parámetros de latitud, longitud y altura que muestra el receptor. Al igual que con el sistema de coordenadas ECI, el sistema de coordenadas ECEF que se utiliza para el GPS tiene

su origen en el centro de masa de la Tierra (fig. 2.9) y su plano  $xy$  coincide con el plano ecuatorial de la Tierra. Sin embargo, en el sistema ECEF, el eje  $x$  se dirige en la dirección de longitud  $0^\circ$  (pasando por la intersección del meridiano de Greenwich con el plano ecuatorial), y el eje  $+y$  apunta en la dirección de  $90^\circ$  de longitud este. Por lo tanto, los ejes  $x$  e  $y$  giran con la Tierra y ya no describen direcciones fijas en el espacio inercial. En este sistema ECEF, el eje  $z$  se elige para que sea normal al plano ecuatorial en la dirección del polo norte geográfico, completando así el sistema de coordenadas de la mano derecha. Dado que el eje de rotación no está fijo en relación con la Tierra sólida (movimiento polar), un polo terrestre convencional (CTP en inglés) se ha definido como el polo medio en los años 1900-1905. En el sistema ECEF, el origen es el centro de masa de la tierra, el eje  $z$  pasa por el CTP, el eje  $x$  pasa por la intersección de un meridiano de referencia (meridiano de Greenwich medio), con el plano ecuatorial del CTP y el eje  $y$  se elige en el plano ecuatorial.

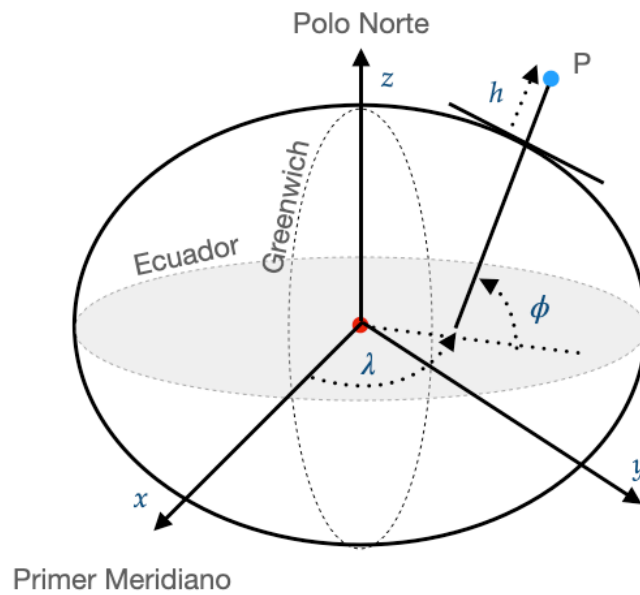
**COORDENADAS ELIPSOIDALES.** La superficie de la Tierra es irregular y cambiante, y el modelo de la Tierra no es una esfera. Es achatada en los polos y que puede aproximarse como un elipsoide de revolución generado al girar un elipsoide sobre su eje menor (elipsoide oblato). El elipsoide se define mediante el semieje mayor ( $a$ ) y el semieje menor ( $b$ ), o de forma equivalente, mediante la excentricidad ( $e$ ) y el aplanamiento ( $f$ )

$$e = \frac{a^2 - b^2}{a^2} \quad f = \frac{a - b}{a} \quad (2.27)$$

Observando la figura 2.10 es posible definir al ángulo  $\phi$  de la latitud geodesica en el plano meridiano que pasa por el punto  $P$  entre el plano ecuatorial ( $xy$ ) de la elipsoide y la línea perpendicular a la superficie del elipsoide que pasa por  $P$  (el norte es positivo); el ángulo  $\lambda$  de la longitud geodésica en el plano ecuatorial entre el meridiano de referencia y el plano meridiano que pasa por  $P$  (el este es positivo); la altura  $h$  geodésica medida a lo largo de la normal al elipsoide que pasa por  $P$ .

**LA GEOIDE.** La Tierra no es esférica y su densidad no es uniforme. Para definir una altura absoluta se lo hace con respecto a un nivel medio del mar idealizado. Se fija entonces una superficie que sirva de referencia global cero para la medición de la altura. El geoide se define como el lugar de todos los puntos con el mismo potencial de gravedad (superficie equipotencial) que mejor se ajusta al nivel medio del mar a nivel global. Esta superficie no es regular, pero tiene una descripción matemática sencilla y tiene en cuenta la formación geológica y el relieve topográfico. Suele representarse numéricamente como una cuadrícula de puntos, y se cartografía en relación con el elipsoide de referencia. En cada punto se define una altura geoi-





**Figura 2.10:** Coordenadas Elipsoidales

dal a lo largo de la línea perpendicular al elipsoide (del elipsoide al geoide).

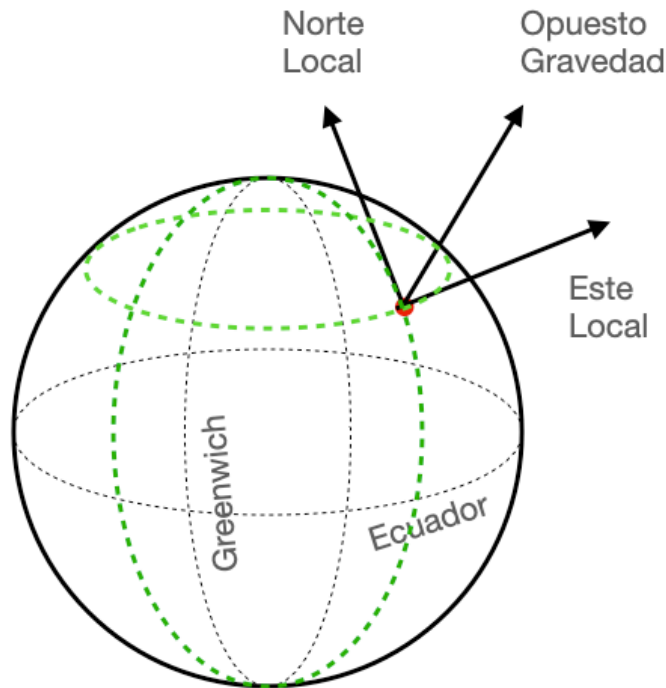
**EL DATUM GLOBAL.** El modelo físico estándar de la Tierra utilizado para la aplicación del GPS es el Sistema Geodésico Mundial 1984 (WGS84), y representa un estándar de facto. Una parte del WGS84 es un modelo detallado de las irregularidades gravitatorias de la Tierra. Esta información es necesaria para derivar la información precisa de las efemérides de los satélites; sin embargo, un receptor GNSS típico sólo pretende estimar su latitud, longitud y altura.

Para ello, WGS84 proporciona un modelo elipsoidal de la forma de la Tierra, como se resume en la Tabla 2.1. En este modelo, las secciones transversales de la Tierra paralelas al plano ecuatorial son circulares. La sección transversal ecuatorial de la Tierra tiene un radio de 6378,137 km, que es el radio ecuatorial medio de la Tierra. En el modelo terrestre WGS84, las secciones transversales de la Tierra normales al plano ecuatorial son elipsoidales. En una sección transversal elipsoidal que contiene el eje  $z$ , el eje mayor coincide con el diámetro ecuatorial de la Tierra. Por lo tanto, el semieje mayor,  $a$ , tiene el mismo valor que el radio ecuatorial medio dado en la tabla 2.1. El eje menor de la sección transversal elipsoidal corresponde al diámetro polar de la Tierra, y el semieje mayor,  $b$ , en WGS84 se toma como 6356,7523142 km.

**MARCO DE REFERENCIA ESTE-NORTE-ARRIBA.** El sistema de coordenadas Este-Norte-Arriba (ENU en inglés) es similar al sistema de coordenadas geocéntrico ECEF. Ambos son cartesianos, lo que signi-

Parámetro	Valor
Semieje Mayor de la Elipsoide ( $a$ )	6378137,0 m
Aplanamiento Elipsoide	298,257223563
Velocidad Angular de la Tierra	$7292115,0 \cdot 10^{-11} \text{ rad/s}$
Constante gravitacional de la Tierra	$3986004,418 \cdot 10^8 \text{ m}^3/\text{s}^2$
Velocidad de la luz en el vacío	$2,99792458 \cdot 10^8 \text{ m/s}$

**Tabla 2.1:** Datos para un gráfico de probabilidad normal



**Figura 2.11:** Coordenadas Este-Norte-Arriba

fica que tienen tres ejes mutuamente perpendiculares y las coordenadas son distancias desde el origen. De hecho, cualquier sistema de coordenadas ENU particular es simplemente la posición xyz representada en un sistema de referencia formado por un plano tangente a la superficie de la Tierra en un origen local, con respecto al cual se definen un Norte local, un Este local y las direcciones Arriba (Fig. 2.11). El sistema de coordenadas ENU puede ser muy útil a la hora de examinar el rendimiento del GNSS desde la perspectiva del usuario o trabajar localmente con una posición. La representación ENU de la posición proporciona información intuitiva para evaluar “lo buena” que es una posición GNSS, dado que la diferencia entre medidas es en metros y no entre minutos y segundos de un ángulo como son la latitud y la longitud.

### 2.1.6 Estructura de la señal de RF

Más allá de que cada sistema GNSS tiene su propia banda de frecuencias y que además con las actualizaciones estas se van ampliando o cambiando, todas están contenidas en la zona del espectro reservada para servicios de navegación por satélite y algunas pocas en el reservado para servicios de radio navegación aeronáutica.

Sin entrar en detalle, las bandas se encuentran entre los 1100MHz y los 1600MHz aproximadamente. Por ejemplo para GPS la frecuencia central de las bandas L1, L2 y L5 están en 1575.42MHz, 1227.60MHz y 1176.45MHz respectivamente. Detalles de las bandas para todos los sistemas GNSS en operación actualmente se pueden consultar en [https://en.wikipedia.org/wiki/Satellite\\_navigation](https://en.wikipedia.org/wiki/Satellite_navigation).

### 2.1.7 Modulación

La señal moduladora, que se denomina en inglés como *Binary Offset Carrier Modulation* (BOC), se representa como BOC(m,n), donde m es un entero que representa múltiplos de la frecuencia de la subportadora de 1023MHz y n otro entero que representa la tasa de código de dispersión en múltiplos de 1023Mchips/s. La forma genérica de la señal puede representarse como,

$$s(t) = \sum_{i=-\infty}^{+\infty} \left\{ c_i d_i r \left( t - i \frac{T_R}{n} \right) \text{sign} \left[ \sin \left( 2\pi \frac{m}{T_R} t \right) \right] \right\} \quad (2.28)$$

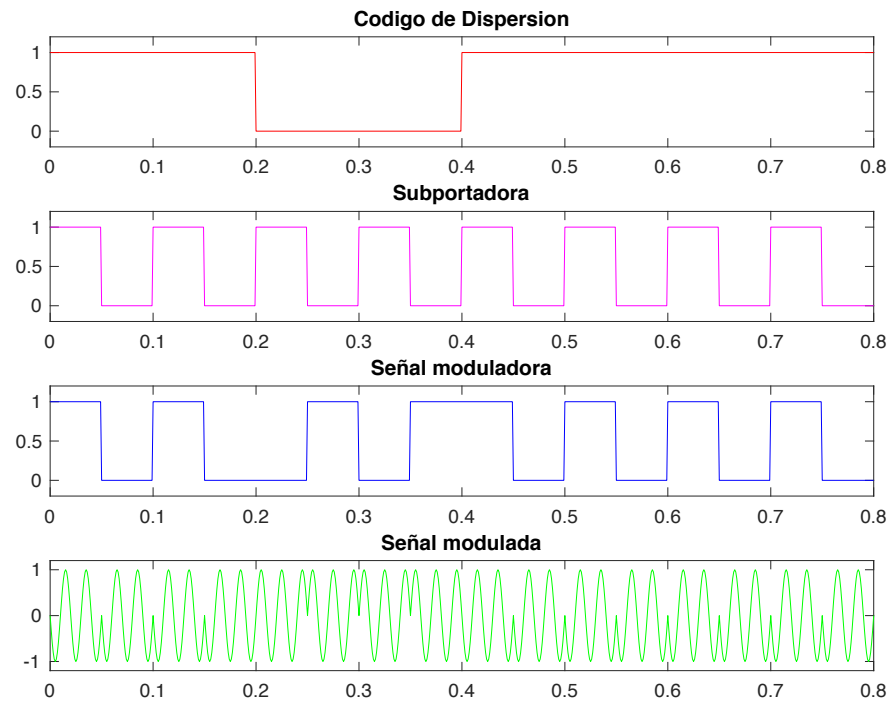
donde  $T_R = \frac{1}{f_R}$ ,  $f_R = 1,023\text{MHz}$ ,  $c_i$  es la secuencia de chips del código pseudo-aleatorio,  $d_i$  son los símbolos dato y  $r(t)$  es un pulso rectangular de amplitud unitaria y duración  $\frac{T_R}{n}$ .

La señal,

$$s_{\sin}(t) = \text{sign} \left[ \sin \left( 2\pi \frac{m}{T_R} t \right) \right] \quad (2.29)$$

representa la subportadora y existen otros tipos de modulaciones que usan coseno. La figura 2.12 muestra un ejemplo de una señal BOC(10,15) para un código con tasa de chip de 5115MChip/s y una subportadora cuadrada de frecuencia fundamental de 10.23MHz. El eje de tiempo está en  $\mu\text{s}$ . Finalmente se muestra una portadora sinusoidal con modulación BPSK (cambio de fase en cada transición de la señal).

**CORRELACIÓN DE LAS SEÑALES.** El rendimiento de los receptores de GNSS depende básicamente de la capacidad de la etapa de procesamiento de señales para alinear un código generado localmente con el código de ruido pseudoaleatorio recibido (PRN). Tal operación se obtiene mediante procedimientos de correlación, y la capacidad

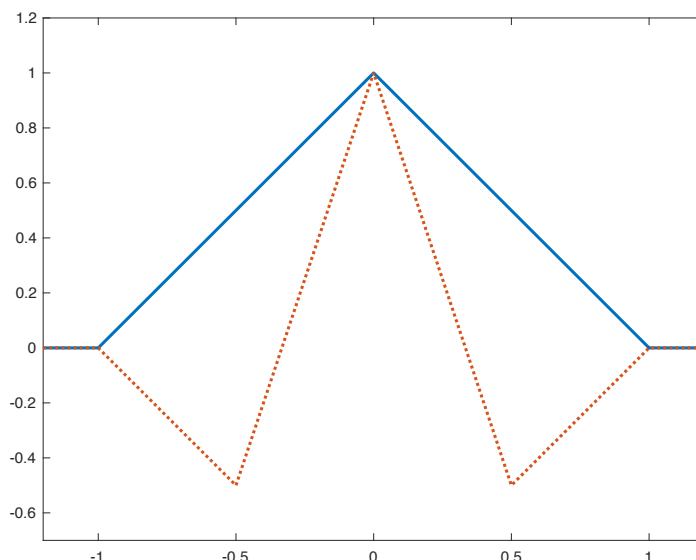


**Figura 2.12:** Ejemplo de señales para BOC(10,5)

del receptor para estimar el tiempo de propagación de la señal en el espacio depende de la forma de la función de correlación de la señal. Debido a las propiedades de ortogonalidad de los códigos de dispersión, la autocorrelación de una señal asume en el intervalo  $(-t_c, +t_c)$  la misma forma de la autocorrelación del pulso de conformación de chip.

Usando un pulso rectangular para la conformación del chip como es el caso de la señal de base de GPS L1 C/A, la función de autocorrelación es triangular entre  $(-t_c, +t_c)$ , mientras que para las señales BOC presenta más de un pico principal y varios picos laterales dependiendo de los valores  $m$  y  $n$ . Las señales moduladas por BOC tienen una función de correlación más estrecha que posiblemente permita una identificación más precisa del tiempo de propagación, como se muestra en la Figura 2.13. Allí se muestra en línea continua la correlación de un pulso de GPS y en línea discontinua la de señales BOC.

Al aumentar la relación  $m/n$  en el BOC( $m, n$ ), aumenta el número de picos laterales, y la separación entre el pico principal y los picos laterales se reduce. Este efecto introduce una ambigüedad potencial en la adquisición y seguimiento del pico de correlación principal, especialmente en el caso de señales que llegan degradadas o incluso son filtradas. En conclusión, los picos principales más estrechos de las señales BOC pueden dar potencialmente una mejor precisión pero, debido a la presencia de los picos laterales, la mejora se compensa con la complejidad del receptor digital.



**Figura 2.13:** Ejemplo de correlación. En línea continua la correlación de un pulso de GPS y en línea discontinua la de señales BOC.

#### 2.1.8 Aumentación

Cuando el rendimiento del posicionamiento GNSS autónomo no es suficiente, se deben utilizar fuentes externas de información que no pertenezcan a la arquitectura central del GNSS para mejorar ese rendimiento. En términos generales, estas fuentes proporcionan correcciones que permiten al receptor de GNSS una mejora. Esta puede deberse principalmente a la exactitud y la precisión, pero también hay otras métricas de rendimiento que pueden beneficiarse, como el tiempo hasta la primera corrección (TTFF, Time To First Fix). En general, los sistemas que apuntan a esto se denominan de *aumentación*.

Los sistemas de aumento se basan en técnicas generalmente denominadas diferenciales (DGNSS en inglés), es decir, un método para mejorar el rendimiento de posicionamiento del GNSS mediante el uso de una o más estaciones de referencia en ubicaciones conocidas, cada una equipada con al menos un receptor GNSS. De esta manera, se pueden proporcionar mejoras de precisión, notificaciones de advertencia de integridad u otros datos a los receptores de los usuarios a través de un enlace de datos.

Existen varias técnicas DGNSS diferentes que varían en sofisticación y complejidad. Desde una sola estación de referencia, que calcula los errores en su posición para informarlos a receptores de GNSS cercanos, hasta redes mundiales que proporcionan datos para estimar errores a partir de modelos de error detallados en cualquier posición cercana a la superficie de la Tierra. Los sistemas DGNSS se pueden clasificar en dos tipos básicos: basados en tierra y basados en satélites.

Hay un tercer tipo centrados en otros sensores en aviones, pero no serán tratados aquí. Un sistema de aumento basado en tierra (GBAS en inglés) a menudo usa radios para transmitir las correcciones diferenciales y es un sistema de área local, ya que las correcciones de transmisión son válidas solo cerca de la estación de referencia. Por otro lado, un sistema de aumentación basado en satélites (SBAS en inglés) transmite correcciones a través de satélites y tiene como objetivo cubrir un área más amplia.

#### 2.1.8.1 *Correcciones diferenciales Locales.*

Como ya se discutió, los sistemas GNSS tienen diferentes fuentes de error que pueden ser mitigadas para aumentar la precisión del cálculo de la posición del usuario. Para mejorar el rendimiento de un sistema de navegación, es posible establecer un sistema GNSS diferencial de área local. Este tipo de sistema explota una estación de referencia (ER) fija que conoce su propia posición georreferenciada, y puede determinar la diferencia entre la distancia medida del satélite y la real. La ER calcula un error para cada satélite a la vista y este error se difunde a los usuarios de la zona de cobertura, transmitiendo las correcciones diferenciales a través de un canal de comunicación utilizando diferentes protocolos como el **RTCM2.3** y el nuevo **RTCM3**. El equipo del usuario puede reducir los errores, aplicando la corrección recibida a cada pseudorange medido. En principio, la ER podría proporcionar una corrección global a la solución calculada, pero su eficacia sólo se lograría si el mismo conjunto de satélites es utilizado para el cálculo de la posición tanto por la ER como por el usuario. Por esta razón, la ER transmite la corrección de cada pseudorange para que el receptor pueda eliminar los errores comunes relacionados con los satélites que están a la vista del usuario y de la ER.

Se sabe que muchas de las fuentes de error del GNSS están altamente correlacionadas en el espacio y el tiempo. Un sistema diferencial eliminará idealmente la mayoría de las contribuciones de error (es decir, los errores del reloj del satélite, los errores de efemérides y los errores de la troposfera y la ionosfera) que son comunes a la ER y al usuario. Los errores debidos al multicamino son locales al receptor y no están correlacionados entre receptores separados aunque estén cercanos (pocas longitudes de onda que es de 20 centímetros aproximadamente) y entonces no pueden ser eliminados por la corrección diferencial.

Hay que tener en cuenta que las correcciones diferenciales son efectivas en una región específica alrededor de la estación de referencia. La eficacia de las correcciones y, por lo tanto, la precisión de la posición final se degradará a medida que aumente la distancia entre el receptor y la estación base (a menudo denominada línea de base). Esto se debe a la descorrelación de los errores experimentados por el usuario y la estación.

En conclusión, la principal limitación de un sistema local es que sólo puede aplicarse fácilmente en zonas limitadas. Las líneas de base típicas para el posicionamiento de la fase portadora (ya presentaremos los detalles) van de 1 a 10 km, mientras que para el posicionamiento por código es de 10 a 100 km. La tendencia de evolución de los sistemas diferenciales locales es crear redes de estaciones capaces de cubrir zonas más amplias.

#### 2.1.8.2 Correcciones diferenciales en grandes áreas.

El GNSS diferencial de grandes áreas es una estrategia de aumento que pretende mejorar las prestaciones del GNSS proporcionando correcciones diferenciales a los usuarios en un área de servicio mayor que los sistemas locales. La arquitectura de un sistema de este tipo incluye una red de estaciones de referencia para monitorear las señales GNSS, un sitio de procesamiento central y un enlace de datos para proporcionar correcciones a los usuarios, normalmente aprovechando uno o más satélites de órbita terrestre geoestacionaria. Por esta razón, estos sistemas también se conocen como sistemas de aumento basados en satélites (SBAS), con un segmento terrestre formado por sitios de monitoreo y procesamiento y un segmento espacial formado por un conjunto de satélites geoestacionarios. Actualmente hay tres SBAS importantes en funcionamiento:

- Wide Area Augmentation System (WAAS), que cubre los Estados Unidos;
- European Geostationary Navigation Overlay System (EGNOS), que cubre Europa;
- Multifunctional Satellite-based Augmentation System (MSAS), que cubre Japón.

India, China y Rusia también han empezado a desarrollar sus propios SBAS, denominados GPS-Aided Geo-Augmented Navigation (GAGAN), Satellite Navigation Augmentation System (SNAS) y System of Differential Correction and Monitoring (SDCM), respectivamente.

El concepto básico de los sistemas diferenciales de grandes áreas es la categorización de las fuentes de error en las variables observables del GNSS, dividiendo el error total de pseudorange experimentado por el receptor en diferentes componentes y estimando la variación de cada componente en una amplia zona en la que se encuentran múltiples ER. Aprovechando un modelo para cada fuente de error, el sistema es capaz de crear la llamada corrección vectorial. A diferencia de la corrección escalar transmitida por el sistema diferencial local, las correcciones vectoriales contienen errores separados del reloj del satélite, errores de efemérides y errores de la ionosfera que tienen que recombinarse para construir la corrección para cada satélite. En



concreto, las correcciones de la ionosfera, que son las que más influyen en el aumento de la precisión de la posición, se evalúan y emiten para los puntos de la grilla ionosférica (IGP en inglés), que son puntos de una grilla virtual de líneas con latitud y longitud constantes a la altura de la ionosfera. El receptor interpola entre estos puntos para desarrollar una corrección de retardo vertical, capaz de mitigar con precisión los errores de la ionosfera, para cada satélite visible utilizando su elevación estimada. La ventaja de la corrección vectorial es su mayor capacidad para captar la descorrelación espacial de las fuentes de error. De esta forma, la precisión alcanzable no depende de la proximidad del usuario a una única ER, como en los sistemas locales, sino que es casi constante en toda la zona de cobertura del sistema SBAS.

En resumen, los principales objetivos de un SBAS son aumentar el rendimiento del sistema básico de navegación para garantizar la precisión, la integridad, la continuidad y la disponibilidad. Proporcionando a los usuarios correcciones diferenciales, para mejorar la precisión del sistema, normalmente de 10 o más metros (por ejemplo, el GPS autónomo) a un par de metros. Proporciona además control de la integridad, para garantizar que los errores están dentro de los límites de confianza (límites tolerables) con una probabilidad muy alta y así garantizar la seguridad.

Hay que señalar que los sistemas SBAS se han diseñado y desarrollado principalmente con el objetivo de cumplir los requisitos de las exigentes aplicaciones de la aviación. Hoy en día se amplía el campo de aplicación a servicios en tierra y mar. Como ejemplo, Europa ha creado recientemente una solución denominada Servicio de Acceso a Datos de EGNOS (EDAS), capaz de transmitir correcciones de EGNOS y datos adicionales a través del Protocolo de Internet (IP). Esto amplía el uso de EGNOS en entornos críticos (por ejemplo, las rutas y sistemas de transporte inteligente) en los que la visibilidad del satélite geostacionario EGNOS es limitada.

**INTEGRIDAD.** La integridad representa el nivel de confianza que un usuario puede tener en la posición calculada y en el error estimado que puede degradarla (nivel de protección). Este nivel de confianza se expresa como un riesgo de integridad, que es la probabilidad de que la posición calculada quede fuera de los niveles de protección requeridos en cualquier punto calculado dentro del área de servicio. Los límites de confianza se presentan en dos valores: el nivel de protección vertical (VPL en inglés) y el nivel de protección horizontal (HPL en inglés). Un sistema de grandes áreas diseñado según la norma asegura la integridad con respecto al GNSS proporcionando al usuario una advertencia adecuada en caso de fallo o superación del nivel de protección. En términos prácticos, la integridad puede considerarse desde el punto de vista del usuario como la probabilidad de que la



información del GNSS (es decir, los pseudorangos y los mensajes de navegación) sea correcta (3 sigma) y que el tiempo de entrega de los avisos de alarma deben ser recibidos por los usuarios en 6 segundos.

#### 2.1.8.3 AGNSS: *Assisted GNSS*

Las técnicas de GNSS asistidas mejoran la sensibilidad de la recepción del GNSS y reducir el TTFF tanto en aplicaciones móviles como en soluciones integradas. Hay que tener en cuenta que un chip GPS autónomo necesita al menos 30 segundos para arrancar en frío, mientras que con las ayudas ese tiempo se reduce a unos pocos segundos. En concreto, la idea central del AGNSS es proporcionar ayuda al terminal a través de una red inalámbrica. Esta ayuda incluye (pero no se limita a): efemérides precisas, posición de referencia, tiempo de referencia, correcciones ionosféricas y parámetros de adquisición (estimación del desplazamiento Doppler). Un servidor de posicionamiento a nivel de red se encarga de generar las asistencias, pero también puede calcular la posición del usuario a partir de los datos observados (enviados por el usuario al servidor), mejorando la precisión gracias a las correcciones diferenciales locales.

La comunicación entre el terminal y el servidor de posicionamiento puede establecerse mediante dos enfoques. El plano de control, en el que las asistencias se envían a través de estructuras de señales de redes celulares predefinidas (GSM, WCDMA). El plano de usuario, en el que las asistencias se envían a través de una conexión de datos TCP/IP general y, por tanto, no requieren ningún mensaje específico del estándar inalámbrico.

#### 2.1.9 *Sistemas GNSS actuales*

En la actualidad hay 6 sistemas GNSS en operación. Cuatro globales: GPS (USA), GLONASS (Rusia), BeiDou (China) y Galileo (UE); y dos regionales: QZSS (Japón) e IRNSS/NavIC (India).

##### 2.1.9.1 *GPS*

El Sistema de Posicionamiento Global (GPS) proporciona servicios gratuitos de posicionamiento y temporizado en todo el mundo. Se desarrolló originalmente para el ejército de EE.UU. y se hizo gratuito para fines civiles. El 22 de febrero de 1978 se lanzó el primer satélite y la capacidad operativa total se completó en junio de 1995 con 24 satélites operativos en órbita. El mantenimiento del GPS corre a cargo del gobierno de Estados Unidos y es de libre acceso para cualquier persona que disponga de un receptor GPS. El GPS ofrece dos servicios de posicionamiento diferentes. El Servicio de Posicionamiento Preciso (PPS) en las frecuencias L1 (1575.42 MHz) y L2 (1227.6MHz), que contienen una señal de encriptación (P) conocida como código Y para

el uso de usuarios autorizados. El Servicio de Posicionamiento Estándar (SPS) en la frecuencia L1, que contiene un código de adquisición grueso (C/A en inglés) con datos de navegación para usuarios civiles. El programa de modernización del GPS comenzó en 2005. Desde ese momento, se han transmitido dos nuevas señales: L2C para los usuarios civiles y una nueva señal militar (código M) en las frecuencias L1 y L2 para ofrecer una mejor resistencia a las interferencias que el código Y. Además, se ha introducido un nuevo enlace de radiofrecuencia L5 (1176:45MHz) para los usuarios civiles. Esta señal, a partir de mayo de 2010, será interoperable con la de Galileo, QZSS e IRNSS/NavIC.

#### 2.1.9.2 GLONASS

La antigua Unión Soviética desarrolló el Global'naya Navigatsionnaya Sputnikovaya Sistema o GLONASS. El primer satélite GLONASS se lanzó el 12 de octubre de 1982. A principios de 1996, una constelación totalmente operativa de 24 satélites estaba en órbita. La constelación completa duró poco debido a las dificultades económicas de Rusia tras el desmantelamiento de la Unión Soviética. En 2002, la constelación se redujo a siete satélites, y sólo seis estaban disponibles durante las operaciones de mantenimiento. GLONASS renació en Diciembre de 2011 se alcanzó de nuevo la plena capacidad operativa.

GLONASS utiliza el acceso múltiple por división de frecuencia (FDMA) para sus señales. Originalmente, el sistema transmitía las señales en dos bandas: L1, 1602-1615:5 MHz, y L2, 1246-1256:5 MHz, a frecuencias espaciadas por 0:5625MHz en L1 y por 0:4375MHz en L2. Los últimos satélites agregados al sistema, transmiten una señal CDMA en una nueva frecuencia L3 (1202:025 MHz).

#### 2.1.9.3 Galileo

El sistema Galileo es una iniciativa conjunta de la Comisión Europea (CE) y la Agencia Espacial Europea (ESA). Los dos primeros satélites se lanzaron en 2011, y otros dos en 2012. En 2018, se han completado 26 satélites, 22 en condiciones de uso. En total, cuando la constelación esté completamente desarrollada, habrá 30 satélites Galileo, con 24 designados como primarios y seis de repuesto.

Los satélites Galileo transmiten tres niveles de servicio en tres bandas de frecuencia utilizando CDMA. El servicio abierto (OS) y el servicio público regulado (PRS) se transmiten en la banda de frecuencias E1 centrada en 1575:46 MHz (la misma que la frecuencia L1 del GPS) y códigos pseudoaleatorios usando modulación binaria (BOC), diferentes para cada satélite al igual que al GPS. La señal del Servicio Comercial (CS) y el PRS se transmiten en la banda de frecuencias E6 centrada en 1278:75MHz utilizando modulación binaria por desplazamiento de fase (BPSK) y BOC, respectivamente. Las distintas señales

contienen mensajes de navegación que proporcionan la información necesaria para adquirir las señales de Galileo y para determinar la posición y la hora del receptor.

#### 2.1.9.4 *BeiDou*

China puso en marcha un sistema de navegación satelital regional conocido como BeiDou (Osa Mayor en chino) tras un programa de investigación y desarrollo que comenzó en 1980. La constelación inicial de tres satélites GEO se completó en 2003. En 2007 se lanzó un cuarto satélite GEO. El sistema regional inicial BeiDou (BeiDou-1) ha sido sustituido por un sistema global conocido como BeiDou-2 (o simplemente BeiDou y anteriormente conocido como Compass). El Sistema de Navegación por Satélite BeiDou (BDS), como se conoce oficialmente ahora, incluirá finalmente cinco satélites GEO, 27 satélites MEO y cinco satélites IGSO (Inclined geosynchronous orbit). BeiDou-2 fue declarado operativo para su uso en China y sus alrededores en 2011. En 2012 se declaró gratis para esta zona. En 2018 comenzó a dar cobertura mundial.

Los satélites transmiten dos niveles de servicio, un servicio abierto y un servicio autorizado principalmente para el gobierno y el ejército chinos, utilizando tres bandas de frecuencia. Las bandas y las frecuencias centrales de los satélites actualmente en uso son B1 a 1561:098 MHz, B2 a 1207:14 MHz y B3 a 1268:52 MHz. Los satélites recientemente lanzados transmiten señales modernizadas en las bandas L1/E1 y L5/E5, así como en la banda BeiDou B3. Por compatibilidad, también está previsto que transmitan la señal de servicio abierto B1 del sistema BeiDou anterior.

#### 2.1.9.5 *QZSS*

El Quasi-Zenith Satellite System (QZSS) utilizará múltiples satélites en órbitas inclinadas, colocados de forma que uno de ellos aparezca siempre cerca del cenit sobre Japón, conocido por sus ciudades de gran altura y donde las señales de los satélites GPS pueden bloquearse fácilmente. El diseño proporciona un servicio de posicionamiento por satélite de alta precisión que cubre casi todo el país, incluidos los cañones urbanos y el terreno montañoso. Los satélites IGSO se complementarán con un satélite GEO. El inicio del servicio completo fue en 2018.

Los satélites generarán y transmitirán sus propias señales, compatibles con las señales modernizadas del GPS. El QZSS también transmite correcciones GPS y datos de disponibilidad, la señal L1-SAIF, por lo que también se considera un satélite de un sistema de aumento. En total, el sistema transmite seis señales con estructuras similares y compatibles con las de GPS y Galileo: L1-C/A (1575:42 MHz), L1C (1575:42 MHz); L2C (1227:6MHz), L5 (1176:45MHz), L1-SAIF 1575:42

MHz), y LEX (L-band Experiment, 1278:75 MHz) y una señal experimental QZSS para un servicio de alta precisión (nivel de 3 cm), que comparte la frecuencia de la señal E6 de Galileo.

#### 2.1.9.6 IRNSS/NavIC

El gobierno indio ha desarrollado el Indian Regional Navigation Satellite System (IRNSS) como un sistema independiente que sirve a la India y sus alrededores. En abril de 2016, el IRNSS pasó a llamarse NavIC, palabra hindi que significa marinero o navegante, así como acrónimo de Navigation with Indian Constellation. El IRNSS proporciona dos tipos de servicio: el SPS, que es un servicio abierto para todos los usuarios, y el Servicio Restringido (RS), que es un servicio codificado disponible sólo para los usuarios autorizados. Se espera que el IRNSS proporcione una precisión de posición en tiempo real basada en la pseudodistancia superior a 20 m en la zona de servicio principal. La constelación del IRNSS consta de tres satélites GEO y dos pares de satélites IGSO. Los satélites transmiten señales de navegación a 1176:45 y 2492:028MHz en las bandas L y S respectivamente. El SPS y el RS se transmiten en ambas frecuencias. El SPS utiliza la modulación BPSK mientras que el RS utiliza la modulación BOC con canales de datos y piloto.

## 2.2 ESTIMACIÓN PROBABILÍSTICA: FILTRO DE KALMAN

Fundamentalmente, un estimador es una regla de decisión que toma como argumento una secuencia de observaciones y cuya acción es calcular un valor para el parámetro o estado de interés. En este proceso de estimación obtenemos una serie de observaciones de un grupo de sensores y, a partir de esta información, deseamos encontrar alguna estimación del verdadero estado del entorno que estamos observando. Los modelos de los sensores son necesarios para entender qué información se proporciona, los modelos del entorno son necesarios para relacionar las observaciones realizadas con los parámetros y los estados que hay que estimar, y se necesita algún concepto de valor de la información para juzgar el rendimiento del estimador.

El filtro de Kalman es un estimador lineal recursivo que calcula sucesivamente una estimación de un estado que evoluciona en el tiempo, sobre la base de observaciones periódicas de este estado. El filtro de Kalman emplea un modelo estadístico explícito de cómo el parámetro de interés  $x(k)$  evoluciona en el tiempo y un modelo estadístico explícito de cómo las observaciones  $z(k)$  que se realizan están relacionadas con este parámetro. Las ganancias empleadas en un filtro de Kalman se eligen para garantizar que, con determinadas hipótesis sobre los modelos de observación y de proceso utilizados, la estimación resultante  $\hat{x}(k)$  minimice el error medio cuadrático.

### 2.2.1 Sistema Lineal

El punto de partida del algoritmo del filtro de Kalman es definir un modelo para los estados a estimar en el espacio de estados

$$\mathbf{x}(k) = \mathbf{F}(k)\mathbf{x}(k-1) + \mathbf{B}(k)\mathbf{u}(k) + \mathbf{G}(k)\mathbf{v}(k), \quad (2.30)$$

donde

- $\mathbf{x}(k) \in \mathfrak{R}^n$  es el estado de interés
- $\mathbf{u}(k) \in \mathfrak{R}^s$  es la entrada de control conocida,
- $\mathbf{v}(k) \in \mathfrak{R}^q$  es un vector aleatorio que describe tanto la dinámica del ruido como las incertidumbres en el modelo de estados en si mismo ( $\mathbf{v}_k$  se supone a menudo aditivo como se muestra en la ecuación).
- $\mathbf{F}(k)$  es la matriz de estados de  $n \times n$
- $\mathbf{B}(k)$  es la matriz de entrada de  $n \times s$
- $\mathbf{G}(k)$  es la matriz de ruido de  $n \times q$

Se supone que las observaciones de los estados de este sistema son hechas de acuerdo a una ecuación lineal de la forma

$$\mathbf{z}(k) = \mathbf{H}(k)\mathbf{x}(k) + \mathbf{D}(k)\mathbf{w}(k) \quad (2.31)$$

donde

- $\mathbf{z}(k) \in \mathfrak{R}^m$  es el vector de observación
- $\mathbf{w}(k) \in \mathfrak{R}^r$  es un vector aleatorio que describe tanto las medidas con ruido como las incertidumbres en el modelo de medida en si mismo ( $\mathbf{w}_k$  se supone a menudo aditivo).
- $\mathbf{H}(k)$  es la matriz de observación que transforma los estados actuales a las observaciones de  $m \times n$ .
- $\mathbf{D}(k)$  es la matriz de ruido de observación de  $m \times r$ .

En el caso del ruido de proceso, se modela como una variable aleatoria con distribución Gaussiana de media cero, temporalmente no correlacionada y covarianza conocida,

$$\begin{aligned} E\{\mathbf{v}_k\} &= \mathbf{0} \quad \forall k \\ E\{\mathbf{v}_i \mathbf{v}_j^T\} &= \delta_{ij} \mathbf{Q}_i \end{aligned}$$

De forma similar se modela el ruido de observación,

$$\begin{aligned} E\{\mathbf{w}_k\} &= \mathbf{0} \quad \forall k \\ E\{\mathbf{w}_i \mathbf{w}_j^T\} &= \delta_{ij} \mathbf{R}_i \end{aligned}$$

Además, se supone que los ruidos de proceso y observación no están correlacionados,

$$E\{\mathbf{v}_i \mathbf{w}_j^T\} = \mathbf{0} \quad \forall i, j$$

Estas condiciones son equivalentes a las propiedades de un proceso de Markov que requiere que observaciones y estados sucesivos sean condicionalmente independientes.

### 2.2.2 Filtro de Kalman Lineal

El filtro de Kalman estima un estado usando los modelos del proceso y de la observación, junto con las suposiciones hechas sobre la calidad del modelo del proceso y el ruido de la observación. El filtro de Kalman produce una estimación que reduce al mínimo el error medio cuadrático

$$\hat{\mathbf{x}}_{i|j} = \arg \min_{\hat{\mathbf{x}}_{i|j} \in \mathcal{R}^n} E\{(\mathbf{x}(i) - \hat{\mathbf{x}})(\mathbf{x}(i) - \hat{\mathbf{x}})^T | \mathbf{Z}^j\} \quad (2.32)$$

El estimado es el valor esperado de estado en el instante  $i$  condicionado a las observaciones hasta el instante  $j$

$$\hat{\mathbf{x}}_{i|j} \triangleq E\{\mathbf{x}(i) | \mathbf{z}_1, \dots, \mathbf{z}_j\} \triangleq E\{\mathbf{x}(i) | \mathbf{Z}^j\} \quad (2.33)$$

La varianza del estimado es la media cuadrada del error de este estimado

$$\mathbf{P}_{i|j} \triangleq E\{(\mathbf{x}(i) - \hat{\mathbf{x}}_{i|j})(\mathbf{x}(i) - \hat{\mathbf{x}}_{i|j})^T | \mathbf{Z}^j\} \quad (2.34)$$

Para presentar el filtro de Kalman usaremos la siguiente notación. El estimado del estado en el instante  $k$  dada toda la información hasta el instante  $k$  será escrito  $\hat{\mathbf{x}}_{k|k}$ . El estimado del estado en el instante  $k$  con la información disponible hasta el instante  $k - 1$  se denomina predicción y se escribe como  $\hat{\mathbf{x}}_{k|k-1}$ . El objetivo es obtener una estimación  $\hat{\mathbf{x}}$  del estado  $\mathbf{x}$  en el instante de tiempo  $k$ , basándose solamente en la estimación en el instante  $k - 1$  y en la observación hecha en el instante  $k$ . La iteración se puede analizar en dos etapas. Una etapa de predicción, que forma una estimación del estado en el paso  $k$  basado solamente en la estimación del paso  $k - 1$ , y una etapa de actualización, que combina la predicción del paso  $k$  con la observación en el instante de tiempo  $k$ , para formar una estimación en el paso  $k$ . La iteración se repite luego en forma recursiva. Este proceso se representa en la figura 2.14.

El algoritmo del filtro de Kalman se presenta sin demostración. Se pueden encontrar derivaciones detalladas en muchos libros sobre el tema (por ejemplo [3]). Se supone que el estado evoluciona en el tiempo según la ecuación 2.30. Las observaciones de este estado se realizan a intervalos de tiempo regulares según la ecuación 2.31. Las





el modelo de observación para generar una actualización. La etapa de actualización del filtro es claramente lineal, con un peso  $\mathbf{W}_k$  asociado a la observación  $\mathbf{z}_k$  y un peso  $1 - \mathbf{W}_k \mathbf{H}_k$  asociado a la predicción  $\hat{\mathbf{x}}_{k|k-1}$ . El filtro de Kalman también proporciona una ecuación de propagación para la covarianza en la predicción y la estimación.

La diferencia entre la observación  $\mathbf{z}_k$  y la predicción de la observación  $\mathbf{H}_k \hat{\mathbf{x}}_{k|k-1}$  se denomina *innovación residual*  $\mathbf{v}_k$ ,

$$\mathbf{v}_k = \mathbf{z}_k - \mathbf{H}_k \hat{\mathbf{x}}_{k|k-1} \quad (2.38)$$

La innovación es una medida importante de la desviación entre la estimación del filtro y la secuencia de observación. De hecho, dado que los *estados verdaderos* no se conocen y no están disponibles para comparar con los estimados, la innovación es la única medida disponible para saber si el filtro está funcionando correctamente. La propiedad más importante de la innovación es que constituye una secuencia blanca, ortogonal y no correlacionada,

$$\mathbb{E}\{\mathbf{v}_k | \mathbf{Z}^{k-1}\} = 0 \quad \mathbb{E}\{\mathbf{v}_i \mathbf{v}_j^T\} = \mathbf{S}_i \delta_{ij} \quad (2.39)$$

donde

$$\mathbf{S}_k = \mathbf{R}_k + \mathbf{H}_k \mathbf{P}_{k|k-1} \mathbf{H}_k^T \quad (2.40)$$

Con estas definiciones, se puede reescribir la etapa de actualización del filtro de Kalman,

$$\begin{aligned} \hat{\mathbf{x}}_{k|k} &= \hat{\mathbf{x}}_{k|k-1} + \mathbf{W}_k \mathbf{v}_k \\ \mathbf{P}_{k|k} &= \mathbf{P}_{k|k-1} - \mathbf{W}_k \mathbf{S}_k \mathbf{W}_k^T \end{aligned} \quad (2.41)$$

donde

$$\mathbf{W}_k = \mathbf{P}_{k|k-1} \mathbf{H}_k^T \mathbf{S}_k^{-1} \quad (2.42)$$

En esta forma, se ve claramente que mientras en la predicción (ecuación 2.35) la varianza crece con la incertidumbre del modelo de propagación de los estados representado por  $\mathbf{Q}_k$ , la varianza en la actualización se reduce proporcionalmente a la varianza de la innovación  $\mathbf{S}_k$ .

**IMPLEMENTACIÓN DEL FILTRO DE KALMAN** Aunque el filtro de Kalman parece algo general y simple su aplicación exitosa depende de los problemas sobre los que se aplica y necesita de criterio ingenieril para ajustar y sintonizar los modelos. Las complicaciones aparecen por dos motivos. Por un lado, lo que tenemos para ver el estado verdadero (las observaciones) son utilizadas por el mismo filtro. Esto implica que no hay nunca una verdad absoluta para juzgar que el filtro funciona. Por el otro lado, no puedo saber si el desempeño del filtro está limitado por las características del sensor o por una mejor sintonía de los parámetros del filtro. Como en todas las cuestiones de ingeniería, seguir un procedimiento ordenado y sistemático es de gran importancia. Un resumen de este procedimiento sería:



- **Entender el Sensor.** Familiarizarse con la física del dispositivo y adquirir la mayor cantidad de datos posibles. Es ilógico pensar en diseñar un estimador sin saber cual es la información que se dispone.
- **Entender el Proceso.** Familiarizarse con su cinemática, su física o geometría. Es evidente que no es factible comenzar a diseñar un filtro sin una comprensión clara del problema que se solucionará.
- **Modelar el Sensor.** Habiendo obtenido toda la información que sea posible del sensor, se debe desarrollar un modelo cinemático y estadístico. Eso significa construir  $\mathbf{H}_k$  y el ruido  $\mathbf{v}_k$ . No hay otra forma de enfrentar al filtro que no sea un modelado detallado y preciso del proceso de medida
- **Modelar el Proceso.** Se debe construir un modelo lo más preciso posible del proceso describiendo todos los aspectos a estimar. Eso significa construir  $\mathbf{F}_k$  y el ruido  $\mathbf{w}_k$ . navaja de Occam (o navaja de Ockham, o principio de economía o de parsimonia)
- **Código del Filtro.** la parte mas fácil!
- **Inicialización.** La formulación recursiva del filtro requiere una inicialización  $\hat{\mathbf{x}}_{0|0}$  y  $\mathbf{P}_{0|0}$ . Esto de de vital importancia en problemas reales
- **Análisis de la secuencia de innovación.** Permite intuir ciertos caminos para sintonizar el filtro.
- **Análisis de desempeño de estado estacionario.** Esto significa analizar el desempeño de las predicciones, estimaciones y sus respectivas covarianzas.
- **Análisis de las condiciones de error.** Esto significa analizar la robustez del filtro ante datos erróneos o espurios.

### 2.2.3 Sistema no Lineal

Los problemas de posicionamiento que ya hemos visto tienen que ver con modelos no lineales. Definimos estos modelos de transición de estados discretos como,

$$\mathbf{x}_k = \mathbf{f}[\mathbf{x}_{k-1}, \mathbf{u}_k, k] + \mathbf{v}_k, \quad (2.43)$$

Definiendo las variables como en el caso lineal y donde  $\mathbf{f}[\cdot, \cdot, \cdot]$  es la función no lineal que mapea el estado y la acción de control en los estados derivados. El modelo de observación también no lineal será,

$$\mathbf{z}_k = \mathbf{h}[\mathbf{x}_k, k] + \mathbf{w}_k \quad (2.44)$$

donde  $\mathbf{h}[\cdot, \cdot]$  es el mapeo de los estados en el espacio de observaciones. Además.

$$\begin{aligned} E\{\mathbf{v}_k\} &= \mathbf{o} \quad \forall k & E\{\mathbf{w}_k\} &= \mathbf{o} \quad \forall k \\ E\{\mathbf{v}_i \mathbf{v}_j^T\} &= \delta_{ij} \mathbf{Q}_i & E\{\mathbf{w}_i \mathbf{w}_j^T\} &= \delta_{ij} \mathbf{R}_i \end{aligned} \quad (2.45)$$

#### 2.2.4 Filtro Extendido de Kalman

Suponemos que está disponible el estimado en el instante  $k-1$  y representa una buena estimación del estado real en ese momento con todas las observaciones hasta entonces,

$$\hat{\mathbf{x}}_{k-1|k-1} \approx E\{\mathbf{x}_{k-1} | \mathbf{Z}^{k-1}\} \quad (2.46)$$

El algoritmo del Filtro Extendido de Kalman (EKF en inglés) se presenta sin demostración. Se pueden encontrar derivaciones detalladas en [3]. Las principales etapas en la derivación del EKF siguen directamente las del filtro de Kalman lineal con el paso adicional de que los modelos de proceso y observación se linealizan como series de Taylor sobre la estimación y la predicción respectivamente. El algoritmo consta de dos etapas:

**Predicción del estado y la varianza**

$$\begin{aligned} \hat{\mathbf{x}}_{k|k-1} &= \mathbf{f}[\hat{\mathbf{x}}_{k-1|k-1}, \mathbf{u}_k] \\ \mathbf{P}_{k|k-1} &= \nabla \mathbf{f}_x \mathbf{P}_{k-1|k-1} \nabla \mathbf{f}_x^T + \mathbf{Q}_k \end{aligned} \quad (2.47)$$

**Actualización del estado y la varianza**

$$\begin{aligned} \hat{\mathbf{x}}_{k|k} &= \hat{\mathbf{x}}_{k|k-1} + \mathbf{W}_k (\mathbf{z}_k - \mathbf{h}[\hat{\mathbf{x}}_{k|k-1}]) \\ \mathbf{P}_{k|k} &= \mathbf{P}_{k|k-1} - \mathbf{W}_k \mathbf{S}_k \mathbf{W}_k^T \end{aligned} \quad (2.48)$$

donde

$$\begin{aligned} \mathbf{W}_k &= \mathbf{P}_{k|k-1} \nabla \mathbf{h}_x^T \mathbf{S}_k^{-1} \\ \mathbf{S}_k &= \nabla \mathbf{h}_x \mathbf{P}_{k|k-1} \nabla \mathbf{h}_x^T + \mathbf{R}_k \end{aligned} \quad (2.49)$$

y donde  $\nabla \mathbf{f}_x$  es el jacobiano de  $\mathbf{f}[\cdot]$  con respecto a  $\mathbf{x}$  evaluado en  $\mathbf{x}_{k-1} = \hat{\mathbf{x}}_{k-1|k-1}$  y  $\nabla \mathbf{h}_x$  es el jacobiano de  $\mathbf{h}[\cdot]$  con respecto a  $\mathbf{x}$  evaluado en  $\hat{\mathbf{x}}_{k+1|k}$ .

El algoritmo del EKF es muy similar al KF lineal con las sustituciones de  $\mathbf{F}_k \rightarrow \nabla \mathbf{f}_{x_k}$  y  $\mathbf{H}_k \rightarrow \nabla \mathbf{h}_{x_k}$ . El EKF es entonces un estimador lineal para un estado error descrito por una ecuación *lineal* y observado por una ecuación *lineal*. El EKF funcionará por tanto como el KF excepto por los jacobianos  $\nabla \mathbf{f}_{x_k}$  y  $\nabla \mathbf{h}_{x_k}$  que en general no son constantes y se los debe calcular en cada iteración. La estimación además, debe permanecer *cercana* al estado verdadero o el filtro será inestable. Por eso se requiere que las no linealidades sean suaves para que las aproximaciones sean válidas. En el mismo sentido, el filtro debe ser inicializado en forma precisa.

De la misma forma que en el filtro lineal, la innovación es la diferencia entre la observación y la predicción de la observación:

$$v_k = z_k - h[\hat{x}_k|_{k-1}] \quad (2.50)$$

y su varianza  $S_k$ .

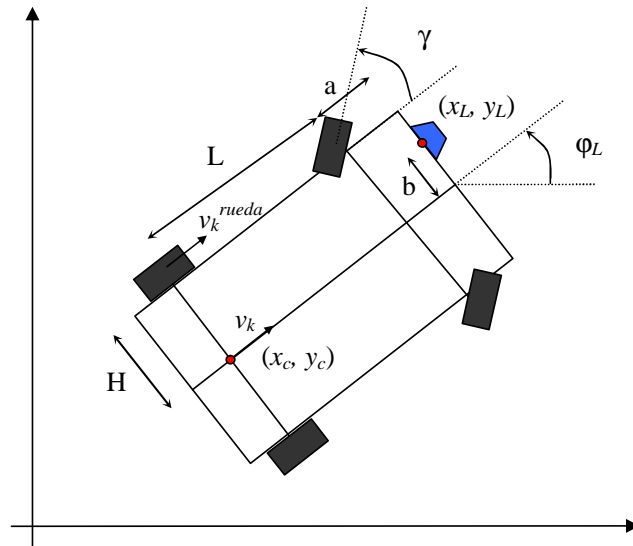
Las aproximaciones lineales para funciones no lineales deben tratarse con cuidado, particularmente en la aplicación del EKF. Todas las recomendaciones descritas para el filtro de Kalman lineal se aplican, con el doble de importancia. En particular la comprensión del sensor y el proceso. Otro tema importante en los problemas no lineales es el de los errores de modelización tanto en el modelo de proceso como en el de observación. Estos son mucho más difíciles de detectar y de tener en cuenta que en el caso lineal. Normalmente, el diseño de un EKF se basa mucho más en el desarrollo de un "modelo de verdad".

### 2.3 NAVEGACIÓN DE UN MÓVIL TERRESTRE

Esta sección presenta un modelo de navegación de un móvil terrestre. Este mismo se utilizará en el resto de las secciones porque es moderadamente complejo y permite introducir varios aspectos del problema de navegación. Recordemos que navegación consiste en determinar la posición, la velocidad y la orientación de un móvil. Por lo tanto extenderemos lo visto hasta el momento, que consiste en determinar una posición estática, a determinarla para un elemento en movimiento.

Las técnicas modernas de navegación se introdujeron hace varios cientos de años para ayudar a los barcos a cruzar el océano. Las primeras técnicas de navegación utilizaban para estimar la posición de un barco las observaciones de la velocidad y el rumbo del barco. Con esta información, se podía predecir su trayectoria, aunque con errores que se acumulaban con el tiempo. El rumbo se determinaba observando la posición del sol o de otras estrellas. La velocidad se hallaba con una corredera (tabla de madera con forma de arco y lastrada con plomo en su borde inferior para que flotase vertical en el agua), compensando los efectos de las corrientes locales. Para compensar los errores acumulados a través de la navegación integrando la velocidad, los puntos de referencia naturales o artificiales bien conocidos en el camino representaban información absoluta. La aparición de nuevos sensores como el compás o la determinación de la latitud midiendo la estrella polar con errores de decenas de millas mejoró las posibilidades de navegar los mares mas allá del Mediterráneo. Para obtener la longitud se necesitarán 300 años con el perfeccionamiento de los relojes para determinar el tiempo con exactitud suficiente.

Esto introduce los elementos esenciales de la navegación, la predicción y la actualización. La predicción puede considerarse como el uso de un modelo de alguna descripción para proporcionar información de navegación a estima que acumula errores de posición con el tiem-



**Figura 2.15:** Descripción del vehículo. El punto  $(x_c, y_c)$  es el centro del eje trasero. El punto  $(x, y)$  es donde se ubican sensores a partir de los cuales se observa el ambiente (GPS, Lasers) y es donde se traslada el modelo del vehículo

po. La corrección es el proceso por el que la observación de puntos de referencia (naturales o artificiales) puede reducir la incertidumbre de localización inherente a la navegación a estima.

Se puede argumentar que, con la llegada de sensores modernos como el Sistema de Posicionamiento Global (GPS), la navegación a estima ya no es necesaria. Sin embargo como ya vimos, no existe un sensor perfecto. Todos tienen alguna medida de error o incertidumbre presente en cada medición del mismo modo que no es posible modelar perfectamente el movimiento del vehículo.

Para cada aplicación suelen ser necesarios diferentes conjuntos de sensores, como la navegación en la minería subterránea, los vehículos submarinos, los barcos y la industria aeroespacial. Cada sistema requiere una combinación particular de sensores de navegación a estima y absolutos. En muchas de estas aplicaciones, la fiabilidad y la integridad son cuestiones centrales y el sistema de navegación debe diseñarse con la suficiente redundancia para poder detectar cualquier posible fallo que pueda producirse.

### 2.3.1 Modelo de un vehículo terrestre

El modelo básico de un vehículo como se muestra en la figura 2.15 en el centro del eje trasero (punto  $P_c$ ) es

$$\begin{bmatrix} \dot{x} \\ \dot{y} \\ \dot{\phi} \end{bmatrix} = \begin{bmatrix} v \cdot \cos(\phi) \\ v \cdot \sin(\phi) \\ \frac{v}{L} \cdot \tan(\gamma) \end{bmatrix} \quad (2.51)$$

Este modelo se puede plantear en la posición donde se colocan los sensores externos. Este punto en la figura 2.15 es el punto  $(x_L, y_L)$  y de la geometría se obtiene que,

$$P_L = (x_L, y_L) = P_C + (L + a) \cdot \vec{T}_\phi + b \cdot \vec{T}_{\phi + \frac{\pi}{2}}$$

donde  $\vec{T}_\phi = (\cos(\phi), \sin(\phi))$  es la orientación del vehículo. Así, el punto  $(x_L, y_L)$  queda explícito como

$$\begin{aligned} x_L &= x_c + (L + a) \cdot \cos(\phi) - b \cdot \sin(\phi) \\ y_L &= y_c + (L + a) \cdot \sin(\phi) + b \cdot \cos(\phi) \end{aligned}$$

Con esto, el modelo de la ecuación 2.51 referido al punto donde están ubicados los sensores externos queda

$$\begin{bmatrix} \dot{x}_L \\ \dot{y}_L \\ \dot{\phi}_L \end{bmatrix} = \begin{bmatrix} v \cdot \cos(\phi) - \frac{v}{L} \cdot ((L + a) \cdot \sin(\phi) + b \cdot \cos(\phi)) \cdot \tan(\gamma) \\ v \cdot \sin(\phi) + \frac{v}{L} \cdot ((L + a) \cdot \cos(\phi) - b \cdot \sin(\phi)) \cdot \tan(\gamma) \\ \frac{v}{L} \cdot \tan(\gamma) \end{bmatrix} \quad (2.52)$$

donde  $v$  es la velocidad en el punto  $(x_c, y_c)$  y  $\gamma$  es el ángulo que se giran las ruedas respecto a la dirección del vehículo. El sensor de velocidad está ubicado en la rueda trasera, si se llama a esta velocidad  $v_r$  su relación con la velocidad deseada es,

$$v = \frac{v_r}{1 - \frac{H}{L} \tan(\gamma)} \quad (2.53)$$

Finalmente, el modelo linealizado aproximado queda de la siguiente forma

$$\begin{aligned} \mathbf{x}_k &= \mathbf{f}[\mathbf{x}_{k-1}, \mathbf{u}_k] \\ \begin{bmatrix} x_{Lk} \\ y_{Lk} \\ \phi_{Lk} \end{bmatrix} &= \begin{bmatrix} x_{Lk-1} + \Delta t \cdot v_k \cdot [\mathbf{G} \cos(\phi_{Lk-1}) - \mathbf{H} \sin(\phi_{Lk-1})] \\ y_{Lk-1} + \Delta t \cdot v_k \cdot [\mathbf{G} \sin(\phi_{Lk-1}) + \mathbf{H} \cos(\phi_{Lk-1})] \\ \phi_{Lk-1} + \Delta t \cdot \frac{v_k}{L} \cdot \tan(\gamma_k) \end{bmatrix} \end{aligned} \quad (2.54)$$

donde  $v_k$  se calcula según la ecuación 2.53

$$\begin{aligned} \mathbf{G} &= \left( 1 - \frac{b \tan(\gamma_k)}{L} \right) \\ \mathbf{H} &= \frac{(L + a) \tan(\gamma_k)}{L} \end{aligned}$$

Como una opción para los ejercicios del final del curso y para algunos ejemplos que mostraremos a continuación, vamos a utilizar los datos siguientes para este modelo:

- $L = 2,38\text{m}$
- $a = 0,95\text{m}$
- $b = 0,5\text{m}$
- $H = 0,75\text{m}$

## 2.4 INTEGRACIÓN DE SENSORES

Existe una gran variedad de estrategias de observación para ayuda a la navegación y guía de vehículos. Esta variedad puede ser clasificada en dos grandes grupos, los sensores internos o de *dead reckoning* y los sensores externos [13].

### 2.4.1 Sensores internos o propioceptivos

La mayoría de los vehículos autónomos terrestres utilizan ruedas para moverse en el ambiente de operación. El método más directo para seguir su posición es medir la velocidad angular de las ruedas y la orientación del mecanismo de dirección. Esta información puede ser utilizada en un modelo de la cinemática del vehículo para predecir la posición. El conjunto de sensores que miden la dinámica de los estados de un vehículo, esto es velocidad, aceleración o variaciones angulares, se denominarán sensores internos o propioceptivos. En este tipo de sensores se incluyen sensores inerciales tales como los acelerómetros y giroscopios, codificadores incrementales o encoders en inglés (ópticos, inductivos, capacitivos, magnéticos), tacómetros, sincros, resolutores, LVDT, compases y potenciómetros.

Los sensores internos proveen información de movimiento en forma incremental a lo largo de una trayectoria y la posición es normalmente obtenida a través de la integración temporal de la secuencia de medidas. Consecuentemente los errores de medición son también integrados lo que implica que estos crezcan sin límite. Se hace necesario por tanto modelar cuidadosamente este error para disminuir sus efectos adversos en la predicción de trayectorias cubriendo distancias largas. En general los errores crecerán con el tiempo y por lo tanto hacen que el sistema de navegación no pueda basarse sólo en este tipo de sensores.

Sin embargo estos sensores tienen otras ventajas que los hacen indispensables en todo lazo de navegación. Son los responsables de proveer la información de alta frecuencia del sistema de navegación lo que resulta indispensable en todas las aplicaciones de guiado y su información es fusionada con la información proveniente de otros sensores como los externos.

### 2.4.2 Sensores externos o exteroceptivos

Los sensores externos o exteroceptivos obtienen medidas que dependen del ambiente que rodea al vehículo. Esto le provee conocimiento de su entorno local, pero la observación depende de la estructura desconocida de ese ambiente y por lo tanto esta sujeta a una gran variedad de errores y ambigüedades.

Los sensores externos pueden medir tanto movimiento incremental como absoluto con respecto a un número de rasgos fijos del ambiente que pueden ser detectados y distinguidos (mojones). Un ejemplo del primer caso son los sensores de velocidad por efecto Doppler que miden la velocidad de una superficie relativa al sensor.

Un caso especial de sensor externo es cuando se colocan mojones artificiales en el ambiente que emiten energía que es detectada por los receptores ubicados en el vehículo. Tal es el caso del Sistema de Posicionamiento Global. Cuando la posición de los emisores es conocida, la ubicación absoluta del vehículo puede ser determinada fácilmente.

Otro tipo interesante de sensores externos son aquellos que miden la ubicación relativa de elementos del ambiente con respecto al vehículo (mojones) cuya posición se conoce en forma absoluta. Por correlaciones sucesivas de observaciones de esos rasgos desde una ubicación del vehículo a la próxima, puede determinarse la posición absoluta de la plataforma y así realizar la navegación.

Los sensores externos pueden ser activos o pasivos. Los sensores activos irradian energía y detectan la energía reflejada por elementos en el ambiente. El tiempo de vuelo, la diferencia de fase, o la información en la amplitud es medida y utilizada para interpretar las propiedades físicas de los objetos. Se aplica luego de la detección un procesamiento de señales para identificar los objetos de interés. Un ejemplo de este tipo de dispositivos es el sensor laser. Los sensores pasivos detectan la energía emitida en el ambiente por los objetos. Ejemplos de estos son las cámaras infrarrojas o de luz visible.

### 2.4.3 Mapas y Rasgos

Los sensores externos permiten observar distintas características del ambiente que rodea al vehículo. La habilidad repetitiva y consistente de detectar y reconocer una entidad específica y bien definida en el ambiente a través de las medidas de un sensor se la refiere como *detección de rasgos*. La entidad detectada, la cual puede ser observada repetidamente tanto en un sensor como en un campo geométrico es denominada *rasgo* o *mojón*. La principal ventaja de detectar y representar rasgos es que pueden ser usados para resumir y describir el ambiente de una forma adecuadamente rica para navegación. Además, la detección y representación de rasgos resultan en una reducción en los requerimientos en almacenamiento y procesamiento.



### 2.4.3.1 *Rasgos Geométricos*

Un rasgo geométrico es un elemento en el ambiente que puede ser parametrizado en un número pequeño de objetos geométricos como puntos, líneas, planos o superficies.

Bajo circunstancias normales los sensores, como cámaras o lasers, retornan observaciones consistentes en un grupo de uno o más puntos; ya sea en coordenadas polares o cartesianas o en el espacio de la imagen. Este conjunto de datos puntuales deben ser agrupados en una primitiva geométrica como un círculo, una línea o una superficie, simplemente ajustándolo a un objeto geométrico parametrizado.

Una ventaja clave de la extracción de rasgos por geométricas es que provee restricciones importantes en la interpretación de las medidas de los sensores y su uso para navegación en ambientes cerrados o interiores es bien conocido. En esos ambientes este tipo de rasgos funciona bien porque el ambiente tiene un número importante de objetos que puedan ser representados así (puertas, ventanas, muebles). Sin embargo en ambientes al aire libre los rasgos naturales pueden ser complicado representarlos de una forma geométrica simple.

### 2.4.3.2 *Rasgos Naturales*

El problema esencial con el uso de rasgos geométricos para modelar ambientes externos naturales es que la mayoría del terreno simplemente no se ajusta un modelo de geometría simple. Los intentos de adaptar los datos externos a tales modelos falla porque es un intento de aplicar una restricción a priori donde no existe.

Una posible solución a este problema es el de modelos basados en grillas. En estos métodos los datos de las medidas simplemente son agrupados en regiones o áreas que corresponden a un mismo tipo de datos. En contraste con el modelado geométrico de rasgos, los métodos basados en grillas esencialmente no proveen restricciones en la interpretación de la información medida. Esto permite obtener una representación muy completa del terreno, sin embargo es posible también una representación errónea y esto es una desventaja.

Los ambientes externos naturales requieren de algún compromiso que permita que los rasgos naturales sean reconocidos. En otras palabras, permitir una descripción general pero a la vez proveer de cierto grado de restricciones para su localización.

### 2.4.3.3 *Rasgos Generalizados*

Existe otra aproximación que es simplemente definir como una entidad aquello que repetidamente, consistentemente e invariablemente es detectado por un sensor en una región del espacio. A esto se los llama rasgos generalizados. Casi cualquier entidad detectable por un sensor puede ser utilizada como un rasgo generalizado. Una región

finita encerrada por una geometría suave y cerrada (burbuja) representa ese rasgo del ambiente. Es un modelo general que se acomoda a un número muy grande de rasgos del ambiente.

Un rasgo generalizado es usualmente una combinación de alguna de tres características. Ellas son las características espaciales (área, perímetro, centro de masa de la burbuja), las características de amplitud (intensidad, color, textura, etc.) y características de procesamiento (media, varianza o transformada de Fourier de la amplitud de los datos). A menudo las características espaciales son suficientes sin embargo es claro que la combinación de todas es una representación mucho mas rica del ambiente que puede ayudar a la robustez y el desempeño de la navegación.

Los rasgos geométricos pueden ser considerados un caso especial y bastante restringido de rasgo generalizado.

#### 2.4.3.4 Mapas Topológicos

Construir el mapa del terreno es una tarea de reconstrucción del espacio mediante un modelo de representación del mundo. En los términos más generales, los tipos de representación pueden dividirse en dos enfoques diferentes: topológico y métrico. Sin embargo, en la práctica, a menudo se aplica un enfoque híbrido que conecta esos dos tipos de mapas principales.

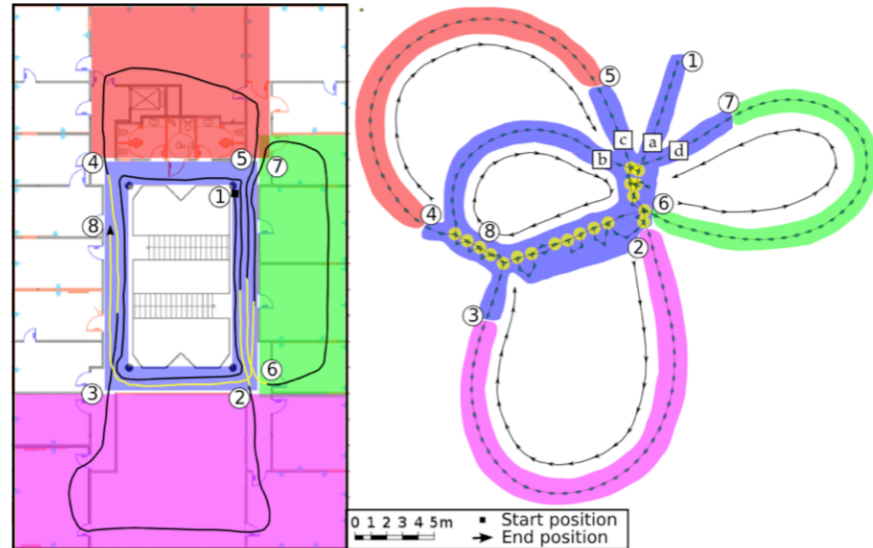
Los mapas topológicos se representan como lugares o puntos conectados entre sí (Fig. 2.16). Probablemente, uno de los ejemplos más frecuentes de representación topológica del mundo es el mapa de las rutas de los subtes. Puede considerarse como un grafo, en el que los vértices o nodos están asociados a lugares exactos (paradas o estaciones), mientras que las aristas de los vértices están relacionadas con las posibles rutas físicas entre estos lugares. Los mapas topológicos están estrechamente relacionados con los métricos, ya que durante la construcción del grafo se utilizan medidas métricas del mundo real para ofrecer posibles transiciones entre los vértices de los nodos.

Los diagramas de Voronoi es otro enfoque asociado a la representación del mundo por la topología. Este método se utiliza para la construcción de mapas con distancias mínimas de seguridad a los obstáculos.

#### 2.4.3.5 Mapas Métricos

Según la forma de representación, todos los mapas métricos pueden dividirse en dos tipos principales: un conjunto de características o de ubicaciones. El mapa basado en la localización se describe como una lista de  $N$  elementos relacionados con cada una de las coordenadas  $x$ ,  $y$  y  $z$ , incluyendo los espacios ocupados y los libres:

$$\mathbf{c} = \{c_{x_1, y_1, z_1}, c_{x_2, y_2, z_2}, \dots, c_{x_N, y_N, z_N}\} \quad (2.55)$$



**Figura 2.16:** Un móvil recorrido un primer bucle en la zona azul, antes de entrar en la zona magenta. Después, la cámara vuelve a entrar en la zona azul y sigue recto hasta la zona roja. A continuación, vuelve de nuevo a la zona azul antes de descubrir la zona verde. El recorrido termina cerca del octavo círculo blanco. En la parte derecha de la figura se muestra el mapa topológico resultante, para el que se utiliza la misma convención de colores con el fin de identificar fácilmente las áreas mapeadas [1].

La familia de mapas métricos basados en la localización más utilizada es la de los mapas de ocupación o de grillas binarias. Los primeros algoritmos de mapeo de cuadrículas binarias fueron presentados a finales de los años 80 utilizando un sonar como sensor. Hoy en día el escaneo láser es una técnica dominante en los enfoques de mapeo de cuadrículas binarias. La idea es dividir el espacio observado en celdas binarias formando una cuadrícula rígida. La precisión en este enfoque depende directamente del tamaño de la cuadrícula y es objeto de compromisos. Las celdas pequeñas proporcionan una representación del mundo más precisa con detalles minúsculos, pero generan mucha carga computacional y a menudo no merecen la pena en espacios vacíos grandes. Por otro lado, para espacios libres enormes, las celdas grandes parecen ser una solución más eficiente. Sin embargo, los obstáculos diminutos pueden ocupar mucho más espacio del que cubren en la realidad. Una de las formas de tratar los compromisos anteriormente mencionados es aplicar la representación en árbol.

Otro tipo de representación del mapa métrico es una lista de puntos de referencia, llamados rasgos o mojonos:

$$\mathbf{m} = \{m_1, m_2, \dots, m_N\} \quad (2.56)$$

A diferencia de las cuadrículas de ocupación, que son densas y almacenan información sobre ubicaciones libres y ocupadas, los mapas

basados en rasgos son malos y almacenan sólo los obstáculos ya descubiertos como puntos de referencia con actualizaciones realizadas durante la exploración añadiendo puntos de referencia recién descubiertos a la lista que representa un mapa. Para los mapas basados en rasgos, la estimación de puntos de referencia mediante filtros de Kalman es natural y robusta.

Los tipos de representación del mundo presentados anteriormente son necesarios para una navegación eficiente de los robots móviles y definen el tipo de enfoques posibles de planificación de trayectorias que podrían aplicarse en los espacios de trabajo objetivo.

#### 2.4.3.6 Mapas Semánticos

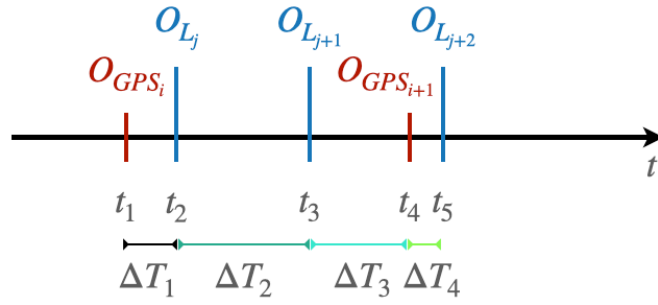
Como vimos, el mapa es una estructura de datos que contiene la información relacionada con el espacio donde se mueve el móvil. Este mapa típico está en armonía con el propósito del móvil, como por ejemplo la navegación. Un mapa semántico, es decir un mapa con información de significado del entorno, aumenta los datos geométricos, por ejemplo, con funcionalidades o eventos que están ubicados en ese espacio.

El propósito principal para plantear la necesidad de un mapa con información semántica es resolver tareas como la planificación, la predicción y la interpretación de los datos de los sensores. Para permitir este razonamiento, se requiere algún conocimiento previo sobre las entidades que puede venir en cualquier formato de representación de ese conocimiento, según sea necesario para el tipo de problemas que se resolverán. Dado que tal conocimiento es típicamente independiente del espacio, no es estrictamente parte del mapa; sin embargo, requerimos que exista para las entidades representadas en el mapa semántico.

En resumen, entonces: Un mapa semántico para un robot móvil es un mapa que contiene, además de información espacial sobre el entorno, asignaciones de características mapeadas a entidades de clases conocidas. Se dispone de más conocimientos sobre estas entidades, independientemente del contenido del mapa, para el razonamiento de tareas asociadas con la navegación. Por ejemplo un plano que es pared y otro que es puerta cerrada. Eventualmente, y de ser necesario, el plano puerta puede abrirse y ser traspasado.

#### 2.4.4 Integración GPS con otros sensores externos

Retomamos el problema de navegación del vehículo terrestre, y suponemos que tenemos para el proceso de actualización al menos dos tipos de sensores externos. Pueden ser más pero, como veremos, la extensión más sensores es muy simple. Uno es un sensor absoluto de posición, el GPS. El otro es un sensor relativo, un lidar.



**Figura 2.17:** Línea de tiempo que representa los momentos donde se reciben las observaciones del GPS ( $O_{GPS_i}$ ) y las del Lidar ( $O_{L_j}$ ).

El lidar (Light Detection and Ranging en inglés) es un tipo de sensor activo de rango que provee directamente información de rango y ángulo y es insensible a condiciones de iluminación externa que perjudica a sensores pasivos como las cámaras de video. El sensor opera con el principio de la medida del tiempo de reflexión de la luz emitiendo una serie corta de pulsos al mismo tiempo que se activa un circuito detector. Si la luz encuentra un objeto en su camino, es reflejada hacia el sensor. Con el tiempo entre la emisión y la recepción el sensor calcula la distancia al objeto. Los láser mas comunes tienen un espejo rotante que desvía los pulsos de luz de manera de barrer un área semicircular. Determinando el ángulo del espejo, se determina en qué dirección está el objeto.

Cuando tenemos sensores que no son sincrónicos y con diferentes tiempos de muestreo, el problema de estimación multisensor se resuelve considerando cada observación del sensor como una actualización independiente y secuencial de la estimación del estado y, para cada observación, calcular una matriz de predicción y ganancia adecuada. De esta forma, se considera cada modelo de sensor individualmente, uno por uno. Esto significa que la dimensión de la innovación y la matriz de covarianza de la innovación en cada etapa de actualización sigue siendo del mismo tamaño que sus equivalentes de un solo sensor, a costa de calcular una nueva matriz de ganancia para cada observación de cada sensor.

La figura 2.17 muestra una línea temporal donde se muestran los momentos  $t_k$  donde se reciben las observaciones del GPS ( $O_{GPS_i}$ ) y del Lidar ( $O_{L_j}$ ). El tiempo de muestreo del GPS es  $t_4 - t_1$  y el del Lidar  $t_3 - t_1 = t_5 - t_3$ .

En la implementación secuencial, en cada instante  $t_k$  que llega una observación, se realiza una predicción del filtro de Kalman y una actualización utilizando el modelo de observación correspondiente. En el caso de  $t_1$  y  $t_4$  será con el modelo del GPS y en  $t_2$ ,  $t_3$  y  $t_5$  el modelo del Lidar. Un tema relevante, es que como estamos trabajando con modelos no lineales y muestreados, los jacobianos y eventualmente

los modelos de propagación de estados, tienen que recalcularse en cada instante de tiempo, utilizando los  $\Delta T_k$  correspondientes.

Entonces, a partir del modelo del vehículo 2.54 y contando con datos de velocidad, orientación, posición absoluta aportada por un GPS y observaciones del ambiente con un láser obtenidos sobre el vehículo en distintos ambientes, se implementa el filtro extendido de Kalman. El objetivo es determinar la pose del vehículo  $(x_L, y_L, \varphi_L)$  a partir de estas medidas respetando los tiempos muestreo de estas señales. Las siguientes variables representan:

- $\mathbf{x}_k$  es el vector de estados que describe la pose del vehículo en el instante  $k$ .
- $\mathbf{u}_k$  es la entrada de control aplicada a los estados del vehículo  $\mathbf{x}_{k-1}$  en el instante  $k-1$  para conducir al vehículo al estado  $\mathbf{x}_k$  en el instante  $k$ .
- $\mathbf{m}_i$  es un vector que describe la ubicación del mojón  $i^{\text{esimo}}$  cuya posición verdadera se supone invariante en el tiempo y en el caso que aquí se trata es un modelo en dos dimensiones con componentes  $m_{xi}, m_{yi}$ .
- $\mathbf{z}_{ik}$  es la observación hecha desde el vehículo de la ubicación del mojón  $i^{\text{esimo}}$  en el instante  $k$ .

Las ecuaciones que implementan un localizador basado en el EKF son las siguientes:

**Inicialización.** Se parte de un estado inicial  $\hat{\mathbf{x}}_{0|0}$  al que se conoce con una incertidumbre  $\mathbf{P}_{0|0}$

**Predicción** del estado y la varianza. El estado y la covarianza del instante anterior (o del inicial) son propagados con las siguientes ecuaciones

$$\begin{aligned}\hat{\mathbf{x}}_{k|k-1} &= \mathbf{f}[\hat{\mathbf{x}}_{k-1|k-1}, \mathbf{u}_k] \\ \mathbf{P}_{k|k-1} &= \nabla \mathbf{f}_x \mathbf{P}_{k-1|k-1} \nabla \mathbf{f}_x^T + \nabla \mathbf{f}_u \mathbf{U}_k \nabla \mathbf{f}_u^T + \mathbf{Q}_k\end{aligned}$$

donde  $\mathbf{f}[\hat{\mathbf{x}}_{k-1|k-1}, \mathbf{u}_k]$  es el modelo de la dinámica del vehículo según las ecuaciones 2.54 y 2.53,  $\mathbf{Q}_k$  y  $\mathbf{U}_k$  son matrices diagonales que representan las covarianzas del ruido de la transición de estados y las entradas respectivamente,  $\nabla \mathbf{f}_x$  es el jacobiano de  $\mathbf{f}[\cdot]$  respecto de los estados y su expresión es

$$\nabla \mathbf{f}_x = \begin{bmatrix} 1 & 0 & -\Delta t \cdot v_k \cdot [\mathbf{G} \sin(\varphi_{Lk-1}) - \mathbf{H} \cos(\varphi_{Lk-1})] \\ 0 & 1 & \Delta t \cdot v_k \cdot [\mathbf{G} \cos(\varphi_{Lk-1}) + \mathbf{H} \sin(\varphi_{Lk-1})] \\ 0 & 0 & 1 \end{bmatrix} \quad (2.57)$$

$\nabla \mathbf{f}_u$  es el jacobiano de  $\mathbf{f}[\cdot]$  respecto de las entradas  $(v_k^{rueda}, \gamma_k)$  y su expresión es

$$\nabla \mathbf{f}_u = \begin{bmatrix} \frac{\Delta t \cdot \mathcal{A}}{1 - \frac{H}{L} \tan(\gamma_k)} & \frac{\Delta t \cdot v_k^{rueda} \cdot \frac{H}{L} \cdot \mathcal{A}}{\left(1 - \frac{H}{L} \tan(\gamma_k)\right)^2 \cdot \cos^2(\gamma_k)} + \frac{\Delta t \cdot v_k \mathcal{C}}{L \cos^2(\gamma_k)} \\ \frac{\Delta t \cdot \mathcal{B}}{1 - \frac{H}{L} \tan(\gamma_k)} & \frac{\Delta t \cdot v_k^{rueda} \cdot \frac{H}{L} \cdot \mathcal{B}}{\left(1 - \frac{H}{L} \tan(\gamma_k)\right)^2 \cdot \cos^2(\gamma_k)} + \frac{\Delta t \cdot v_k \mathcal{D}}{L \cos^2(\gamma_k)} \\ \frac{\Delta t \cdot \tan(\gamma_k)}{L \cdot \left(1 - \frac{H}{L} \tan(\gamma_k)\right)} & \frac{\Delta t \cdot v_k^{rueda} \cdot \frac{H}{L} \cdot \tan(\gamma_k)}{L \cdot \left(1 - \frac{H}{L} \tan(\gamma_k)\right)^2 \cdot \cos^2(\gamma_k)} + \frac{\Delta t \cdot v_k}{L \cos^2(\gamma_k)} \end{bmatrix} \quad (2.58)$$

donde

$$\begin{aligned} \mathcal{A} &= G \cos(\varphi_{Lk-1}) - H \sin(\varphi_{Lk-1}) \\ \mathcal{B} &= G \sin(\varphi_{Lk-1}) + H \cos(\varphi_{Lk-1}) \\ \mathcal{C} &= -b \cos(\varphi_{Lk-1}) - (L + a) \sin(\varphi_{Lk-1}) \\ \mathcal{D} &= -b \sin(\varphi_{Lk-1}) + (L + a) \cos(\varphi_{Lk-1}) \end{aligned}$$

**Actualización del estado y la varianza.** Cada vez que se obtenga una observación y sea asociada con datos del mapa, el estado y covarianza predichos son actualizados por las siguientes ecuaciones

$$\hat{\mathbf{x}}_{k|k} = \hat{\mathbf{x}}_{k|k-1} + \mathbf{W}_k (\mathbf{z}_k - \mathbf{h}[\hat{\mathbf{x}}_{k|k-1}]) \quad (2.59)$$

$$\mathbf{P}_{k|k} = \mathbf{P}_{k|k-1} - \mathbf{W}_k \mathbf{S}_k \mathbf{W}_k^T \quad (2.60)$$

donde

$$\mathbf{W}_k = \mathbf{P}_{k|k-1} \nabla \mathbf{h}_x^T \mathbf{S}_k^{-1} \quad (2.61)$$

y

$$\mathbf{S}_k = \nabla \mathbf{h}_x \mathbf{P}_{k|k-1} \nabla \mathbf{h}_x^T + \mathbf{R}_k \quad (2.62)$$

En el caso que se utilice como sensor externo un sensor relativo (escáner láser por ejemplo), los objetos observados son reportados por el rango  $z_r$  y el ángulo  $z_\theta$  respecto del sensor. La observación en las coordenadas del vehículo será de la forma

$$\mathbf{z}_k = \begin{bmatrix} z_{x,k} \\ z_{y,k} \end{bmatrix} = \begin{bmatrix} z_{r,k} \cos(z_{\theta,k}) \\ z_{r,k} \sin(z_{\theta,k}) \end{bmatrix} \quad (2.63)$$

y la predicción de la observación

$$\mathbf{h}[\hat{\mathbf{x}}_{k|k-1}] = \begin{bmatrix} \cos(\varphi_{Lk-1}) & -\sin(\varphi_{Lk-1}) \\ \sin(\varphi_{Lk-1}) & \cos(\varphi_{Lk-1}) \end{bmatrix} \cdot \begin{bmatrix} m_{xi} - x_L \\ m_{yi} - y_L \end{bmatrix} \quad (2.64)$$

Además  $\nabla \mathbf{h}_x$  es el jacobiano del modelo de observación respecto de los estados,

$$\nabla \mathbf{h}_x = \begin{bmatrix} -\cos(\varphi_{Lk-1}) & -\sin(\varphi_{Lk-1}) & \mathbb{K}_1 \\ \sin(\varphi_{Lk-1}) & -\cos(\varphi_{Lk-1}) & \mathbb{K}_2 \end{bmatrix} \quad (2.65)$$



donde

$$\begin{aligned}\mathbb{K}_1 &= -(m_{xi} - x_L) \sin(\varphi_{Lk-1}) + (m_{yi} - y_L) \cos(\varphi_{Lk-1}) \\ \mathbb{K}_2 &= (m_{xi} - x_L) \cos(\varphi_{Lk-1}) - (m_{yi} - y_L) \sin(\varphi_{Lk-1})\end{aligned}$$

y  $\mathbf{R}_k$  es la incertidumbre en la observación y que resulta en la estadística de la medida trasladada a las coordenadas del vehículo,

$$\mathbf{R}_k = \nabla \mathbf{z}_{z_{r,k}, z_{\theta,k}} \cdot \begin{bmatrix} \sigma_{z_{r,k}}^2 & 0 \\ 0 & \sigma_{z_{\theta,k}}^2 \end{bmatrix} \cdot \nabla \mathbf{z}_{z_{r,k}, z_{\theta,k}}^T \quad (2.66)$$

donde  $\sigma_{z_{\theta,k}}$  y  $\sigma_{z_{r,k}}$  son las desviaciones estándar de la estadística del rango y el ángulo medido y  $\nabla \mathbf{z}_{z_{r,k}, z_{\theta,k}}$

$$\nabla \mathbf{z}_{z_{r,k}, z_{\theta,k}} = \begin{bmatrix} \cos(z_{\theta,k}) & -z_{r,k} \sin(z_{\theta,k}) \\ \sin(z_{\theta,k}) & z_{r,k} \cos(z_{\theta,k}) \end{bmatrix} \quad (2.67)$$

En el caso que la observación sea hecha por un sensor absoluto (GPS), estas ecuaciones son más simples, ya que las medidas estarán dadas en las coordenadas globales y serán directamente la posición del vehículo en dos dimensiones,

$$\mathbf{z}_k = \begin{bmatrix} z_{x,k} \\ z_{y,k} \end{bmatrix} \quad (2.68)$$

y la predicción de la observación simplemente

$$\mathbf{h}[\hat{\mathbf{x}}_{k|k-1}] = \begin{bmatrix} \hat{x}_{Lk|k-1} \\ \hat{y}_{Lk|k-1} \end{bmatrix} \quad (2.69)$$

#### 2.4.4.1 Ejemplo de Navegación

Se presenta una corrida experimental del localizador basado en EKF con el vehículo moviéndose en el techo de una cochera. Los datos son los mismos que utilizaremos en todos los ejemplos. El vehículo usado fue equipado con un codificador óptico incremental en una rueda trasera, un potenciómetro para medir la orientación de las ruedas delanteras, una plataforma inercial, un compás, un receptor de GPS y una escáner láser para dar información en dos dimensiones de la distancia radial y angular de los objetos que están frente al vehículo. En las figuras 2.18 y 2.19 se pueden observar tanto al vehículo con sus sensores como a una vista aérea del ambiente.

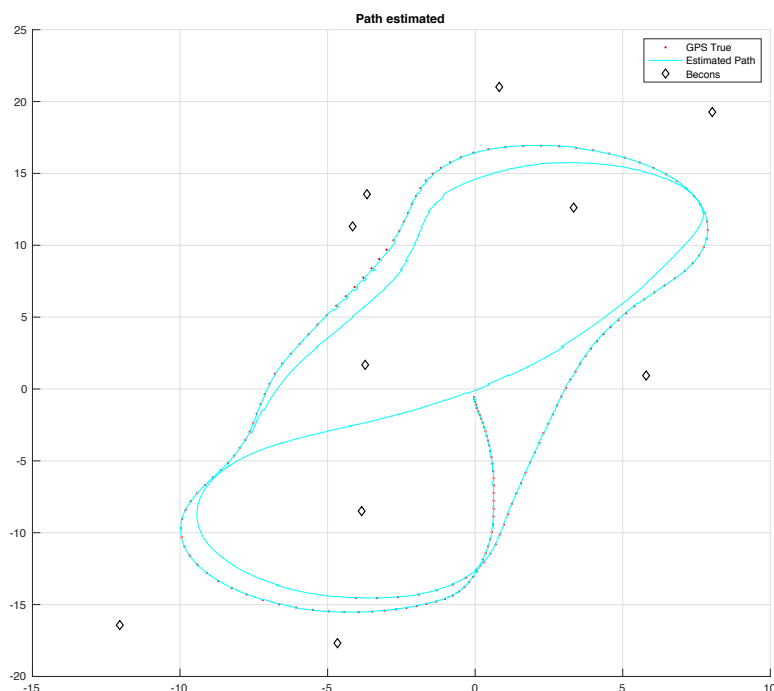
El mapa es conocido y los objetos que son detectados como rasgos son tubos metálicos cubiertos con cinta reflectante para que el sensor láser los detecte fácilmente. Esto hace al análisis independiente del funcionamiento de un detector de rasgos. Dado que la experiencia fue realizada en la terraza de un edificio se pudo obtener información de GPS de muy alta calidad. Esto es posible debido a la gran cantidad



**Figura 2.18:** Vehículo instrumentado utilizado en las experiencias.



**Figura 2.19:** Vista aérea del techo de la cochera donde se realizó la primer experiencia. Sobre impuesto (señalado con una flecha) el trayecto recorrido por el móvil y la ubicación de los mojones.



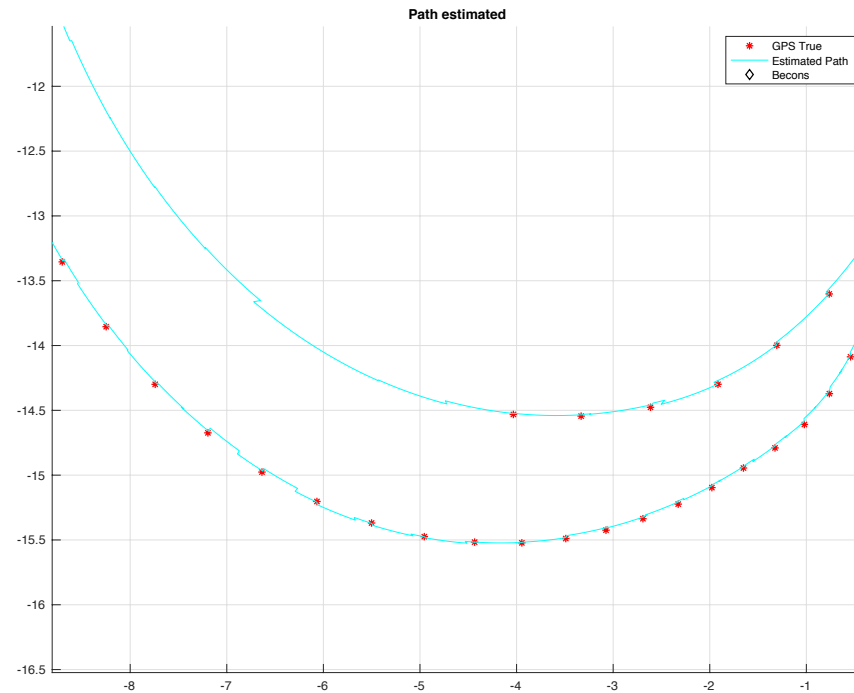
**Figura 2.20:** Resultado de la navegación utilizando EKF, un escáner láser y un GPS. Los ejes están en metros y el punto inicial de navegación corresponde al origen de coordenadas. En puntos se muestra la medida del GPS en modo RTK como referencia de la navegación, en línea continua se muestra el resultado de la estimación. Los diamantes corresponden a la ubicación de los mojones en el mapa.

de satélites que simultáneamente pueden ser seguidos por el receptor cuando tiene visión del cielo sin obstrucciones. La posición de los mojones se tomó con el GPS en el modo más preciso (RTK: Real Time Kinematics) promediando varias medidas, lo que permitió hacer la suposición que el mapa es muy preciso.

En resumen, el vehículo se mueve y a partir de las medidas de velocidad y orientación de las ruedas más el modelo cinemático en 2D, se predice la ubicación del vehículo. Cada vez que el láser observa un objeto del mapa (las cintas reflectantes) y cada vez que se obtiene una medida de GPS, la información es utilizada para estimar y corregir la posición real del vehículo.

El resultado de la navegación se muestra en la figura 2.20. En esta figura se presenta en línea continua el resultado de la navegación y en línea con puntos las medidas aportadas por el GPS en modo RTK y sirve como una referencia de la trayectoria real. Los diamantes representan los mojones del mapa.

En la figura 2.21 se ve un detalle del proceso de navegación en un momento en particular que es cuando se dejan de procesar los datos del GPS. Esto simularía por ejemplo el hecho de que el receptor de GPS pierda la precisión RTK y pase a otro modo, que es usual cuando se ocultan por edificios o árboles, algunos de los satélites.

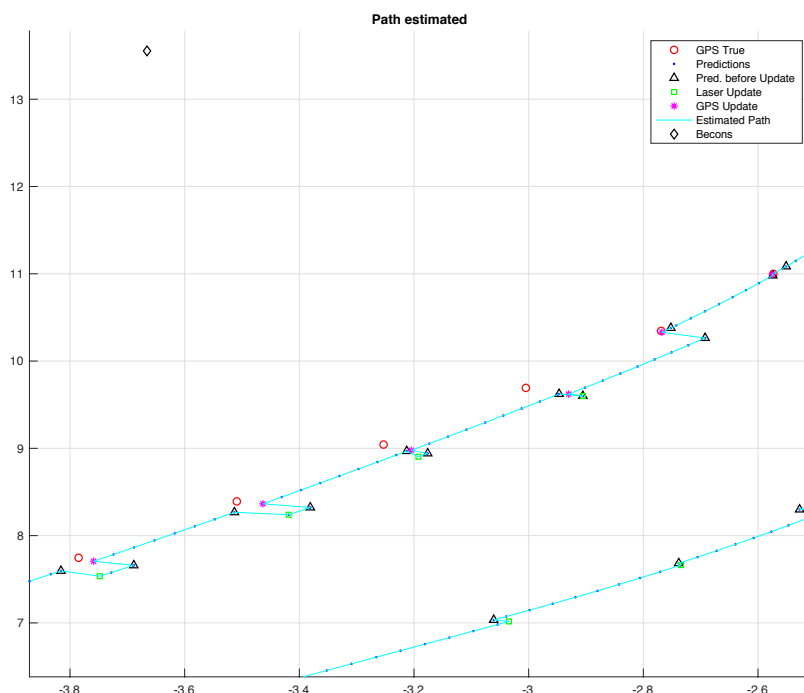


**Figura 2.21:** Ampliación del resultado de la navegación utilizando EKF, un escáner láser y un GPS. En puntos se muestra la medida del GPS en modo RTK como referencia de la navegación, en línea continua se muestra el resultado de la estimación. Los diamantes corresponden a la ubicación de los mojones en el mapa. En un momento se apaga el GPS y se sigue navegando con el láser solamente.

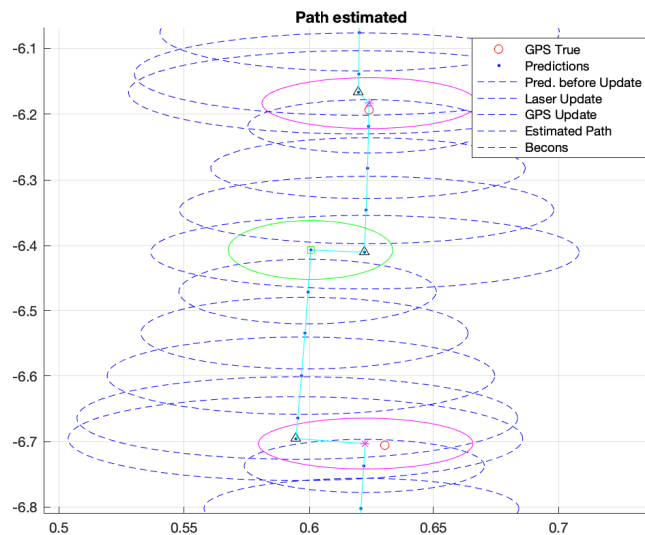
Como se puede observar, la navegación continua en forma suave con correcciones por actualización del sensor láser. Lo que es importante resaltar es uno de los aspectos que aporta la integración de sensores y es la robustez del sistema de navegación ante fallas temporarias de alguno de los sensores.

La figura 2.22 muestra en detalle todas las etapas de predicción y actualización del EKF. Los mojones del mapa se muestran como diamantes y en línea continua se presenta el resultado final del proceso de navegación y con diferentes símbolos los detalles. Los asteriscos magenta representan la actualización hecha con una medida de GPS (círculos rojos que sirven además de referencia). Con cuadrados verdes se muestra la actualización con medidas del láser. Las predicciones se dividen en dos. Como puntos azules se muestran las predicciones que se realizan con las mediciones de velocidad y dirección del volante. Los triángulos negros representan la predicción antes de una actualización con algunos de los sensores externos. En este caso las medidas de los sensores internos se suponen constantes entre muestreos.

Se pueden observar dos cosas interesantes. La primera es que el movimiento producto de las predicciones con un modelo simple, es suave y bastante cercano siempre a la trayectoria real, siendo que en-



**Figura 2.22:** Navegación con las diferentes etapas del filtro mostradas en forma explícita. En línea continua se muestra el resultado final de la estimación. Los puntos negros representan la predicción hecha sin actualización y los triángulos negros la predicción antes de esta. Se puede observar claramente que las predicciones hechas con el modelo tiene movimientos suaves y los saltos representan las correcciones de la actualización.



**Figura 2.23:** Representación aproximada del error de estimación. Las elipses representan el 95 % de confianza de la estimación. En azul la correspondiente a la predicción y en verde y magenta la resultante de la actualización con Láser y GPS respectivamente.

tre actualizaciones hay alrededor de 8 predicciones. Lo segundo son las actualizaciones que muestran saltos (menores a 10 cm) y representan la corrección de la media predicha. Por ejemplo, si observamos actualización con GPS (asterisco magenta en [-3.2,9] por ejemplo), vemos que es un valor intermedio entre la predicción (triángulo) y el valor de GPS (círculo). En la misma figura, la trayectoria inferior muestra solo actualizaciones con Láser dado que el GPS no entrega medidas de calidad.

La figura 2.23 muestra una aproximación del error de estimación. Las elipses representan el 95 % de confianza y en azul se muestra la correspondiente a la predicción y en verde y magenta la correspondiente a la actualización con Láser y GPS respectivamente. Como se observa, la predicción aumenta el error por la integración de los errores en el tiempo y el error de modelo que se acumula. La actualización reduce este error. Se puede observar que el error resultante de la actualización es una elipse contenida (en probabilidad casi en su totalidad) dentro de la elipse de la predicción.

## 2.5 SLAM: NAVEGACIÓN Y CONSTRUCCIÓN SIMULTÁNEA DEL MAPA

La localización y construcción del mapa en forma simultánea (SLAM por su sigla en inglés) es un proceso concurrente en el cual se construye un mapa basado en rasgos del ambiente y este mapa es el que se utiliza para obtener una estimación de la posición del vehículo. Basado en la capacidad para extraer información útil para la navegación de los datos suministrados por sus sensores, el vehículo comienza en



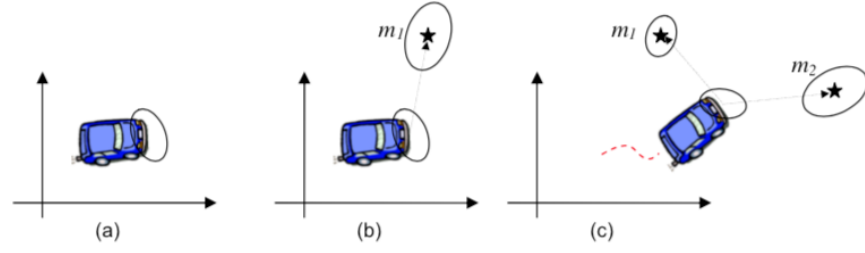
una ubicación desconocida sin el conocimiento a priori de la localización de los rasgos, potencialmente mojones de un mapa métrico. El vehículo logrará observaciones relativas a su posición observando rasgos del ambiente a través de algunos sensores, y observaciones directas o absolutas proveniente de otros (GPS por ejemplo). Si las observaciones relativas de rasgos permanecen consistentes serán incorporados como mojones y a partir de ellos, calcula en forma simultánea una estimación de la localización del vehículo y una estimación de las localizaciones de los mojones. Mientras que continúa en el movimiento, el vehículo va construyendo un mapa completo de mojones y utiliza éstos para proporcionar estimaciones continuas de la localización del vehículo. Siguiendo la posición relativa entre el vehículo y las características reconocibles del ambiente, la posición del vehículo y la posición de los rasgos se pueden estimar simultáneamente.

El SLAM comprende dos tareas, agregar los rasgos del ambiente al mapa usando la pose del vehículo como referencia, mientras que se usan los mojones en el mapa para estimar la pose del vehículo. La incertidumbre en las medidas de los sensores resulta en incertidumbres tanto en la estimación de la pose del vehículo como en la de los mojones y esas incertidumbres son dependientes entre sí (o estarán correlacionadas) como se muestra en la figura 2.24. Un vehículo posee una incertidumbre representada por una elipse y observa un rasgo y lo incorpora al mapa como un mojon  $m_1$ . El estimado de este mojon depende de la incertidumbre en la posición del vehículo con lo cual estará correlacionado con esta. Un momento después, el vehículo vuelve a observar a  $m_1$  e inicializa  $m_2$ . La estimación de  $m_2$  contará con la incertidumbre del vehículo que a su vez depende de  $m_1$  también. Con esto, el estimado de  $m_2$  estará correlacionado con el vehículo y también con  $m_1$ . A medida que los mojones son observados en forma repetida estarán correlacionados entre sí en forma creciente uno con otro lo que hace en el límite un mapa más rígido. La correlación entre las incertidumbres juega un papel esencial en el SLAM de forma que acopla los mojones individuales unos con otros y el vehículo con el mapa. Se ha demostrado que los intentos de estimar la pose del vehículo y los mojones del mapa en forma independiente produce estimados con incertidumbres inconsistentes (optimistas).

Una estimación estocástica consistente requerirá que se mantengan las correlaciones entre los estados. El resultado final para el SLAM basado en el EKF es que el estimado de la pose del vehículo y de los mojones del mapa deben ser almacenados en el mismo vector de estados, y este vector debe ser aumentado a medida que se agregan nuevos de estos.

En el instante  $k$  el estado del vehículo es representado como  $\mathbf{x}_{L_k}$  y el estado del mojon  $m_i$  como  $\mathbf{m}_{i_k}$ . Se define el vector de estados





**Figura 2.24:** Incertidumbres y correlaciones en el SLAM. En (a) un vehículo posee una incertidumbre representada por una elipse. En (b) el vehículo observa un rasgo y lo incorpora al mapa como un mojón  $m_1$ . El estimado de este mojón depende de la incertidumbre en la posición del vehículo con lo cual estará correlacionado con esta. Un momento después, en (c), el vehículo vuelve a observar a  $m_1$  e inicializa  $m_2$ . La estimación de  $m_2$  contará con la incertidumbre del vehículo que a su vez depende de  $m_1$  también. Con esto, el estimado de  $m_2$  estará correlacionado con el vehículo y también con  $m_1$ . A medida que los mojones son observados en forma repetida estarán correlacionados entre sí en forma creciente uno con otro, en el límite se obtiene un mapa más rígido.

aumentado  $\mathbf{x}_k$  que describe los estados del vehículo y de todos los mojones en el ambiente en el instante  $k$  como

$$\mathbf{x}_k = \{\mathbf{x}_{L_k}^T \mathbf{m}_{1_k}^T \mathbf{m}_{2_k}^T \dots \mathbf{m}_{N_k}^T\}^T$$

El estimado de este estado y su correspondiente covarianza serán

$$\begin{aligned} \hat{\mathbf{x}}_{k|k} &= \{\hat{\mathbf{x}}_{L_{k|k}}^T \hat{\mathbf{m}}_{1_{k|k}}^T \hat{\mathbf{m}}_{2_{k|k}}^T \dots \hat{\mathbf{m}}_{N_{k|k}}^T\}^T \\ \mathbf{P}_{k|k} &= E\{\tilde{\mathbf{x}}_{k|k} \tilde{\mathbf{x}}_{k|k}^T\} \\ &= E\left\{ \begin{bmatrix} \tilde{\mathbf{x}}_{L_{k|k}} \\ \tilde{\mathbf{m}}_{1_{k|k}} \\ \vdots \\ \tilde{\mathbf{m}}_{N_{k|k}} \end{bmatrix} \begin{bmatrix} \tilde{\mathbf{x}}_{L_{k|k}}^T \\ \tilde{\mathbf{m}}_{1_{k|k}}^T \\ \vdots \\ \tilde{\mathbf{m}}_{N_{k|k}}^T \end{bmatrix}^T \right\} \\ &= \begin{bmatrix} \mathbf{P}_{LL_{k|k}} & \mathbf{P}_{L1_{k|k}} & \dots & \mathbf{P}_{LN_{k|k}} \\ \mathbf{P}_{1L_{k|k}} & \mathbf{P}_{11_{k|k}} & \dots & \mathbf{P}_{1N_{k|k}} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{P}_{NL_{k|k}} & \mathbf{P}_{N1_{k|k}} & \dots & \mathbf{P}_{NN_{k|k}} \end{bmatrix} \end{aligned}$$

donde

$$\begin{aligned} \mathbf{P}_{LL_{k|k}} &= E\{\tilde{\mathbf{x}}_{L_{k|k}} \tilde{\mathbf{x}}_{L_{k|k}}^T\} \\ \mathbf{P}_{i_{k|k}} &= E\{\tilde{\mathbf{m}}_{i_{k|k}} \tilde{\mathbf{m}}_{i_{k|k}}^T\} \\ \mathbf{P}_{Li_{k|k}} &= E\{\tilde{\mathbf{x}}_{L_{k|k}} \tilde{\mathbf{m}}_{i_{k|k}}^T\} \\ \mathbf{P}_{ij_{k|k}} &= E\{\tilde{\mathbf{m}}_{i_{k|k}} \tilde{\mathbf{m}}_{j_{k|k}}^T\} \end{aligned}$$

Los términos  $\mathbf{P}_{L_{k|k}}$  y  $\mathbf{P}_{i_{k|k}}$  para  $i = 1, \dots, N$  son las matrices de covarianza para el vehículo y los  $N$  mojones estimados respectivamente. El término  $\mathbf{P}_{L_{i_{k|k}}} = \mathbf{P}_{i_{L_{k|k}}}^T$  indica la correlación entre el error en el estimado del vehículo y el error en el estimado de la posición del mojon  $m_i$  y el término  $\mathbf{P}_{ij_{k|k}} = \mathbf{P}_{ji_{k|k}}^T$  señala la correlación entre los errores en las estimaciones de dos mojones  $m_i$  y  $m_j$ . Esto implica que esta estructura de estados, permite al EKF mantener una medida de todas las correlaciones que se presenten entre los errores en sus estimaciones.

El modelo aumentado obtenido a partir del modelo del vehículo de la ecuación 2.54, y con la asunción que los mojones están fijos respecto de un eje o marco de referencia global será entonces

$$\mathbf{x}_k = \begin{bmatrix} \mathbf{x}_{L_k} \\ \mathbf{m}_{i_k} \end{bmatrix} = \mathbf{f}[\mathbf{x}_{k-1}, \mathbf{u}_k] = \begin{bmatrix} \mathbf{f}[\mathbf{x}_{L_{k-1}}, \mathbf{u}_k] + \mathbf{v}_k \\ \mathbf{m}_{i_{k-1}} \end{bmatrix} \quad i = 1, \dots, N$$

donde  $\mathbf{v}_k$  es un ruido aleatorio aditivo que se supone como una variable aleatoria con distribución Gaussiana de media cero, temporalmente no correlacionada y covarianza conocida  $\mathbf{Q}_k$ . Como se puede ver se supone que el modelo de la dinámica de los mojones no posee ruido aditivo, y es natural si se piensa que están fijos respecto del marco de referencia.

Con esto, el filtro extendido de Kalman toma la siguiente forma

**Predicción** del estado y la varianza

$$\hat{\mathbf{x}}_{k|k-1} = \mathbf{f}[\hat{\mathbf{x}}_{k-1|k-1}, \mathbf{u}_k] \quad (2.70)$$

$$\mathbf{P}_{k|k-1} = \nabla \mathbf{f}_x \mathbf{P}_{k-1|k-1} \nabla \mathbf{f}_x^T + \mathbf{Q}_k \quad (2.71)$$

donde

$$\nabla \mathbf{f}_x = \begin{bmatrix} \nabla \mathbf{f}_{x_L} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{I} & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{I} \end{bmatrix}$$

y donde  $\nabla \mathbf{f}_{x_L} = \nabla \mathbf{f}_{x_{L_k}}$  es el Jacobiano como respecto de los estados del vehículo y donde el subíndice  $k$  que representa el instante de tiempo se omite por simpleza de notación.

**Actualización** del estado y la varianza

$$\begin{aligned} \hat{\mathbf{x}}_{k|k} &= \hat{\mathbf{x}}_{k|k-1} + \mathbf{W}_k (\mathbf{z}_k - \mathbf{h}[\hat{\mathbf{x}}_{k|k-1}, i]) \\ \mathbf{P}_{k|k} &= \mathbf{P}_{k|k-1} - \mathbf{W}_k \mathbf{S}_k \mathbf{W}_k^T \end{aligned} \quad (2.72)$$

donde

$$\mathbf{W}_k = \mathbf{P}_{k|k-1} \nabla \mathbf{h}_{x,i}^T \mathbf{S}_k^{-1} \quad (2.73)$$

$$\mathbf{S}_k = \nabla \mathbf{h}_{x,i} \mathbf{P}_{k|k-1} \nabla \mathbf{h}_{x,i}^T + \mathbf{R}_k \quad (2.74)$$

El modelo de observación debe tener en cuenta el vector de estados aumentado y por lo tanto se define un nuevo modelo. Sin embargo, el nuevo modelo de observación representa sólo un cambio en la notación. El nuevo modelo debe describir la observación de uno o varios rasgos que son observados y tendrá la siguiente forma para el caso de uno solo,

$$\mathbf{h}[\mathbf{x}_k, i] = \mathbf{h}_{\mathbf{x}_{L_k}, \mathbf{m}}[\mathbf{x}_{L_k}, \mathbf{m}_i]$$

por lo tanto el jacobiano en la ecuación 2.73 tendrá la siguiente forma

$$\begin{aligned} \nabla \mathbf{h}_{\mathbf{x}, i} &= [\nabla \mathbf{h}_{\mathbf{x}_L} \nabla \mathbf{h}_{\mathbf{m}_1} \dots \nabla \mathbf{h}_{\mathbf{m}_N}] \\ &= [\nabla \mathbf{h}_{\mathbf{x}_L} \mathbf{0} \dots \mathbf{0} \nabla \mathbf{h}_{\mathbf{m}_i} \mathbf{0} \dots] \end{aligned} \quad (2.75)$$

donde  $\nabla \mathbf{h}_{\mathbf{x}_L} = \nabla \mathbf{h}_{\mathbf{x}_{L_k}}$  es el jacobiano respecto de los estados del vehículo y  $\nabla \mathbf{h}_{\mathbf{m}_i} = \nabla \mathbf{h}_{\mathbf{m}_{i_k}}$  lo es respecto de los estados del mojon que está siendo observado y donde como ya se ha dicho, el subíndice  $k$  representa el tiempo y es omitido por simpleza de notación. Este vector al tener la dimensión proporcional a la cantidad de mojon en el mapa, y al ser pocos de ellos los que son efectivamente observados, la mayoría de sus elementos son cero. Esto insinúa la posibilidad de encontrar simplificaciones que permitan una menor complejidad de cálculo.

### 2.5.1 Significado de las correlaciones

El vector de estados aumentado permite al filtro mantener correlaciones explícitas entre los errores del vehículo y los mojon y entre los errores de los mojon. El significado de esas correlaciones pueden ser examinadas considerando una actualización obtenida a partir de la observación del mojon  $\mathbf{m}_i$ .

Considérese primero el problema de la estimación de los estados del vehículo  $\mathbf{x}_{L_k}$  y los de un mojon del mapa  $\mathbf{m}_{i_k}$ , independiente del resto de los mojon. La observación del mojon  $\mathbf{m}_i$  puede ser expresada a partir de una medida relativa al vehículo como

$$\mathbf{z}_k = \begin{bmatrix} z_{r,k} \\ z_{\theta,k} \end{bmatrix} = \begin{bmatrix} \sqrt{(\mathbf{m}_{x,k} - \mathbf{x}_k)^2 (\mathbf{m}_{y,k} - \mathbf{y}_k)^2} \\ \frac{\pi}{2} + \arctan\left(\frac{\mathbf{m}_{y,k} - \mathbf{y}_k}{\mathbf{m}_{x,k} - \mathbf{x}_k}\right) + \varphi_{L_k} \end{bmatrix} = \mathbf{h}[\mathbf{x}_{L_k}, \mathbf{m}_{i_k}]$$

Se puede obtener un modelo linealizado aplicando series de Taylor y la predicción de la observación, alrededor de los estados del predichos del vehículo y los estados predichos del mojon  $\mathbf{m}_i$ , será

$$\hat{\mathbf{z}}_{k|k-1} \approx \mathbf{h}[\hat{\mathbf{x}}_{L_k|k-1}, \hat{\mathbf{m}}_{i_k|k-1}] + \nabla \mathbf{h}_{\mathbf{x}_L} \tilde{\mathbf{x}}_{k|k-1} + \nabla \mathbf{h}_{\mathbf{m}_i} \tilde{\mathbf{m}}_{i_k|k-1}$$

con lo cual la innovación tendrá la forma

$$\begin{aligned} \mathbf{v}_k &= \mathbf{z}_k - \hat{\mathbf{z}}_{k|k-1} \\ &= \nabla \mathbf{h}_{\mathbf{x}_L} \tilde{\mathbf{x}}_{k|k-1} + \nabla \mathbf{h}_{\mathbf{m}_i} \tilde{\mathbf{m}}_{i_k|k-1} \end{aligned}$$

A partir de esta expresión se obtiene la covarianza de la innovación, que será entonces,

$$\begin{aligned}
 \mathbf{S}_k &= E\{\mathbf{v}_k \mathbf{v}_k^T\} \\
 &= \nabla \mathbf{h}_{\mathbf{x}_L} \mathbf{P}_{L|k-1} \nabla \mathbf{h}_{\mathbf{x}_L}^T + \nabla \mathbf{h}_{\mathbf{m}_i} \mathbf{P}_{i|k-1} \nabla \mathbf{h}_{\mathbf{m}_i}^T \\
 &\quad + \nabla \mathbf{h}_{\mathbf{x}_L} \mathbf{P}_{L|i|k-1} \nabla \mathbf{h}_{\mathbf{m}_i}^T + \nabla \mathbf{h}_{\mathbf{m}_i} \mathbf{P}_{i|L|k-1} \nabla \mathbf{h}_{\mathbf{x}_L}^T \\
 &\quad + \mathbf{R}_k
 \end{aligned} \tag{2.76}$$

Ahora, operando con las ecuaciones 2.73 a 2.75 del sistema aumentado para el SLAM, la varianza de la innovación toma la siguiente estructura

$$\begin{aligned}
 \mathbf{S}_k &= \nabla \mathbf{h}_{\mathbf{x}_L} \mathbf{P}_{L|k-1} \nabla \mathbf{h}_{\mathbf{x}_L}^T + \nabla \mathbf{h}_{\mathbf{m}_i} \nabla \mathbf{P}_{i|k-1} \mathbf{h}_{\mathbf{m}_i}^T \\
 &\quad + \nabla \mathbf{h}_{\mathbf{x}_L} \mathbf{P}_{L|i|k-1} \nabla \mathbf{h}_{\mathbf{m}_i}^T + \nabla \mathbf{h}_{\mathbf{m}_i} \mathbf{P}_{i|L|k-1} \nabla \mathbf{h}_{\mathbf{x}_L}^T \\
 &\quad + \mathbf{R}_k
 \end{aligned}$$

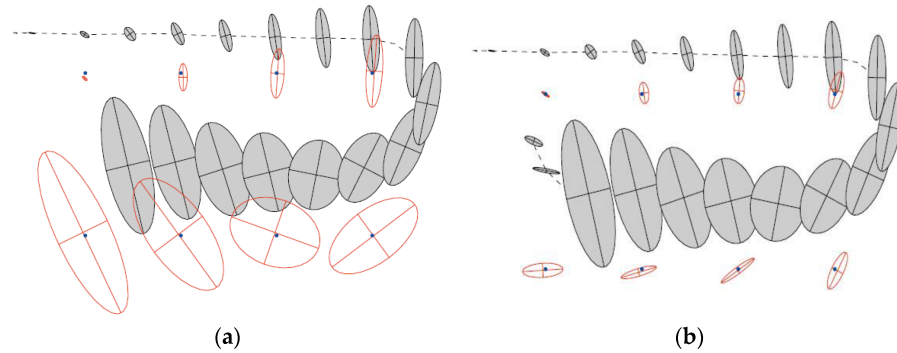
que es de la forma de la expresión 2.76 deducida para el caso de un mojón independiente. Por lo tanto esta estructura de estados aumentada para resolver el SLAM calcula correctamente la covarianza de la innovación. Y dado que cualquiera de los mojones del mapa puede ser observado, todos los términos  $\mathbf{P}_{L|i|k-1}$  para  $i = 1, 2, \dots, N$  deben ser conocidos de manera tal que la covarianza de la innovación puede ser siempre calculada. La matriz de covarianza que mantiene los términos  $\mathbf{P}_{L|i|k-1}$  para  $i = 1, 2, \dots, N$  debe además mantener los términos  $\mathbf{P}_{i|j|k-1}$  para  $i, j = 1, 2, \dots, N$  para completar la matriz de covarianza, si no, la condición  $\mathbf{P}_{k|k-1} = E\{\tilde{\mathbf{x}}_{k|k-1} \tilde{\mathbf{x}}_{k|k-1}^T\}$  presumida en el filtro de Kalman no es satisfecha. Esta condición es necesaria sino las estimaciones del filtro serían *injustificadamente confiables*.

Se examinará ahora la matriz de peso de Kalman  $\mathbf{W}_k$ . En la formulación del vector de estados ampliado, la matriz toma la siguiente forma,

$$\begin{aligned}
 \mathbf{W}_k &= \mathbf{P}_{k|k-1} \nabla \mathbf{h}_{\mathbf{x}_i}^T \mathbf{S}_k^{-1} \\
 &= \begin{bmatrix} \mathbf{P}_{L|L|k-1} \nabla \mathbf{h}_{\mathbf{x}_L}^T + \mathbf{P}_{L|i|k-1} \nabla \mathbf{h}_{\mathbf{m}_i}^T \\ \mathbf{P}_{L|L|k-1} \nabla \mathbf{h}_{\mathbf{x}_L}^T + \mathbf{P}_{L|i|k-1} \nabla \mathbf{h}_{\mathbf{m}_i}^T \\ \vdots \\ \mathbf{P}_{L|i|k-1} \nabla \mathbf{h}_{\mathbf{x}_L}^T + \mathbf{P}_{i|i|k-1} \nabla \mathbf{h}_{\mathbf{m}_i}^T \\ \vdots \\ \mathbf{P}_{j|L|k-1} \nabla \mathbf{h}_{\mathbf{x}_L}^T + \mathbf{P}_{j|i|k-1} \nabla \mathbf{h}_{\mathbf{m}_i}^T \\ \vdots \\ \mathbf{P}_{N|L|k-1} \nabla \mathbf{h}_{\mathbf{x}_L}^T + \mathbf{P}_{N|i|k-1} \nabla \mathbf{h}_{\mathbf{m}_i}^T \end{bmatrix} \mathbf{S}_k^{-1}
 \end{aligned}$$

De aquí se puede observar que las submatrices

$$(\mathbf{P}_{j|L|k-1} \nabla \mathbf{h}_{\mathbf{x}_L}^T + \mathbf{P}_{j|i|k-1} \nabla \mathbf{h}_{\mathbf{m}_i}^T) \mathbf{S}_k^{-1}$$



**Figura 2.25:** Cierre de lazo (de [28]). La figura (a) muestra el momento antes de cerrar el lazo. La figura (b) muestra la reducción de las incertidumbres de los mojones luego de cerrar el lazo y observar los mojones que habían sido incorporados al inicio de la navegación.

ilustran que el filtro, con los estados aumentados, actualiza además de la estimación de los mojones observados, la estimación de los mojones que no están siendo observados debido a la correlación entre la innovación y el error en la estimación del mojón  $m_j$  no observado en el instante  $k$ . Cada vez que se hace una observación todo el mapa se mejora para llegar en el límite a un mapa relativo perfecto.

Este proceso se ve claramente en un fenómeno que se llama cierre de lazo. En lazos cerrados, si un robot detecta una posición en la que ha estado antes y observa los mojones que incorporó al inicio de la navegación, los errores acumulados se reducirán y la precisión del mapa aumentará. La correlación entre la incertidumbre de la pose del robot y la incertidumbre del mapa se muestra gráficamente en la figura 2.25.a (figura extraída de [28]). El robot se mueve a lo largo del camino trazado como la línea discontinua, observando cerca de ocho mojones, dibujados como puntos. Las elipses sombreadas implican las incertidumbres del robot sobre sus propias poses, dibujadas a lo largo del tiempo. Como resultado del error de navegación, la postura del robot se vuelve más incierta cuando se mueve. Las estimaciones de las posiciones de los mojones se especifican mediante elipses blancas. Se puede ver que a medida que la postura del robot se vuelve más incierta, también aumenta la incertidumbre en las ubicaciones estimadas de los mojones recién observados.

En la Figura 2.25.b, el robot completa el bucle y vuelve a visitar un punto de referencia observado previamente. Como la coordenada del primer hito observado tiene una alta precisión, la incertidumbre de la postura del robot predicha se reduce significativamente. Por lo tanto, también disminuyen las incertidumbres de posición de los mojones previamente percibidos. El efecto resultante es que la información se propaga a mojones previamente observados, de modo que la obtención de información sobre la pose del robot es probablemente la característica más importante de la estimación posterior del SLAM.

En la Figura 2.25.b, se puede ver que las elipses sombreadas obtenidas antes del cierre del lazo no se encogen después del cierre, porque representan una serie temporal de incertidumbres de la trayectoria del robot y no se revisan posteriormente.

La capacidad de volver a identificar los mojones detectados previamente es de crucial importancia ya que la revisión del robot mejora la precisión de la localización y el mapeo de mojones. Por eso se suele decir que el algoritmo del SLAM es un algoritmo de localización de visita/revisita. Esto significa que el robot explorará una nueva zona, construirá el mapa y volverá al inicio para mantener las incertidumbres acotadas. Y desde allí continuará la exploración.

### 2.5.2 Asociación de datos

En navegación hay una pregunta que sobreviene cada vez que un rasgo es detectado. Cuando un rasgo es hallado por un sensor que no reporta la identidad del mismo, ¿cómo puedo decir que ese rasgo corresponde a un mojón del mapa y cómo se puede identificar la identidad del mojón?. Esto es importante ya que el algoritmo de localización para realizar una actualización requiere conocer la identidad del mojón que esta viendo y partir de su posición  $m_{xi}$ ,  $m_{yi}$  se puede predecir la observación (ecuación 2.64) que constituye parte del factor de corrección de la ecuación 2.59.

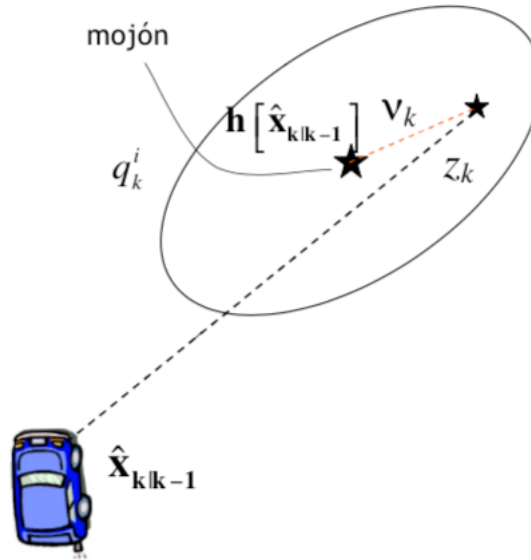
Hay varias formas de solucionar este problema y la primera y más simple es utilizar mojones que no son idénticos; un ejemplo típico es el GPS donde cada satélite envía su nombre codificado con los datos temporales. Pero esto es obviamente poco práctico en el resto de las implementaciones que se intenten cuando se trabaja en un ambiente no estructurado y natural. Un segundo método utiliza herramientas estadísticas para determinar cuán probable es que tal mojón sea observado suponiendo que el vehículo esta en una cierta posición con una determinada confianza. En este caso el problema de la asociación de datos se la refiere como el problema de encontrar la correspondencia entre dos conjuntos de datos o medidas de un evento. En la navegación basada en rasgos, una asociación incorrecta lleva a mapas no consistentes y a una divergencia en la estimación de la posición del vehículo.

La base de la asociación de datos es el uso de la innovación normalizada o la *cota de validación*. Como se ha presentado en capítulos anteriores, la innovación es la diferencia entre la observación y la predicción, expresión que se repite aquí para el Filtro Extendido de Kalman,

$$v_k = z_k - h[\hat{x}_{k|k-1}] \quad (2.77)$$

y cuya varianza es

$$S_k = \nabla h_x P_{k|k-1} \nabla h_x^T + R_k. \quad (2.78)$$



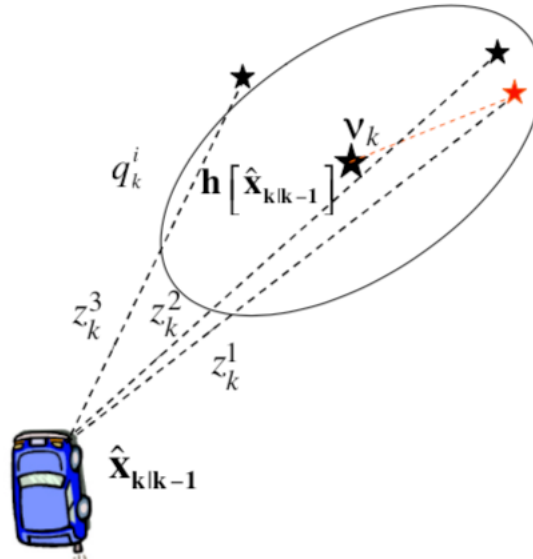
**Figura 2.26:** Cota de validación de la observación. La cota es definida en el espacio de observación y está centrada en la predicción de la observación  $h[\hat{x}_{k|k-1}]$ . La cota es la innovación normalizada y para un valor fijo de  $q_k^i$  es una elipsoide que está definida por los autovalores de la inversa de la covarianza de innovación. El tamaño de la cota es establecido por el requerimiento de que la probabilidad de una correcta asociación esté por encima de un umbral. Las observaciones que caen dentro de ese umbral son consideradas válidas

Finalmente se define a la innovación normalizada como

$$q_k^i = v_k^{(i)T} S_i^{-1} v_k^i, \quad (2.79)$$

que es una cantidad escalar. Se puede demostrar que si la innovación es una variable aleatoria blanca y de media cero entonces la innovación normalizada es una variable aleatoria  $\chi^2$  con  $n_z$  grados de libertad (la dimensión del vector de observación). Es posible por lo tanto establecer un valor para  $q_k^i$  que incluya una probabilidad de que la observación y la observación predicha están siendo correctamente asociadas; esta es la clave para establecer métodos de asociación. Obsérvese la figura 2.26, para un valor fijo de  $q_k^i$  la innovación normalizada describe un volumen elipsoidal (de la dimensión de  $n_z$ ) cuyos ejes son proporcionales a la recíproca de los autovalores de  $S_k$ . Esta elipsoide define un área o volumen en el espacio de observación centrado en la observación predicha. Si una observación cae dentro de este volumen entonces es considerada válida, de ahí el término cota de validación. La innovación normalizada sirve como base para todas las técnicas de asociación de datos, primero porque es prácticamente la única medida de divergencia de la estimación de los estados del filtro y segundo porque permite una medida probabilística precisa de la correcta asociación.





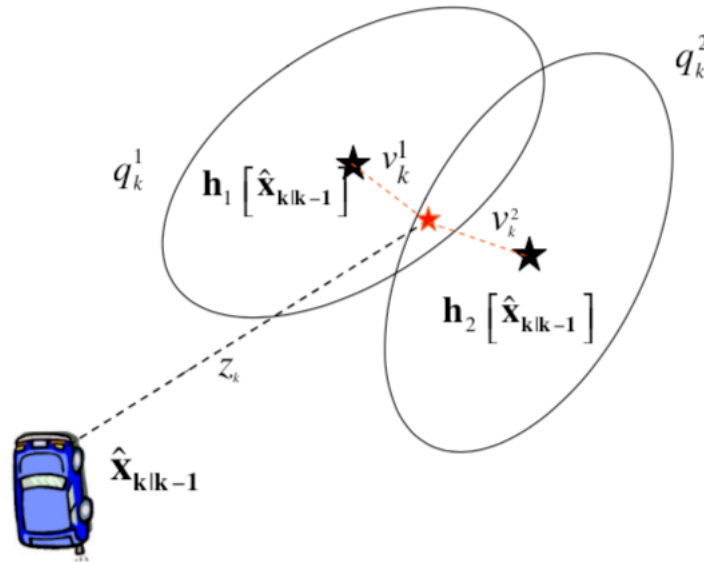
**Figura 2.27:** Para un solo objeto, el algoritmo del vecino más cercano selecciona la observación lo más cerca posible la predicción para la asociación, en este caso  $z_k^1$ . Se rechazan el resto de las observaciones. La medida de la proximidad es la innovación normalizada.

#### 2.5.2.1 El filtro estándar del vecino más cercano

El filtro del vecino más cercano aplica una política obvia de asociación de datos: elige simplemente la medida lo más cerca posible a la medida predicha como la única observación validada, y las observaciones restantes se rechazan de la consideración (figura 2.27). La única medida validada se utiliza para actualizar la estimación del estado. La definición de *lo más cerca posible* es la de aquella observación que logra la innovación normalizada mínima  $q_k^i$ . Si no hay observaciones con valores normalizados de la innovación menos que algún  $q_k^i$  definido, no hay observación asociada. Este algoritmo presume implícitamente una alta probabilidad de una correcta asociación y un índice bajo del perturbación dentro de la cota. En tales situaciones este algoritmo es adecuado y se utiliza mucho en la práctica.

Sin embargo este algoritmo básico tiene algunos problemas, particularmente en situaciones donde hay un alto grado de perturbaciones. La figura 2.28 muestra una situación particular en la cual dos cotas de validación de dos objetos están solapadas. En este, y muchos otros casos, el problema de asociar correctamente una observación a un objeto es complejo. Primero porque la asociación obvia en parejas más cercanas no es necesariamente correcta, y en segundo lugar porque la asociación es dependiente del orden de consideración y por lo tanto todas las asociaciones se deben considerar juntas para alcanzar una asignación correcta.

Existen otros algoritmos que aumentan en complejidad para resolver este problema planteado. **La asociación de datos probabilística**



**Figura 2.28:** Para objetos múltiples, cada objeto selecciona su observación más cercana independientemente de otras posibles asociaciones. Esto puede conducir fácilmente a asignaciones erróneas en altas densidad de objetos. En el caso mostrado, una medida proveniente de un objeto en el ambiente es asociado con dos mojonos distintos en el mapa produciendo dos actualizaciones distintas. Naturalmente una de las dos al menos es incorrecta.

por ejemplo calcula una probabilidad de asociación correcta entre la observación y el mapa que se utiliza para formar una ubicación promedio pesada considerada *correcta*. El problema mas importante con este método es que presume que hay un único mojon sin tener en cuenta resto y donde además se acepta que la detección de este es independiente en cada instante de observación. Sin embargo en el caso de múltiples mojonos en el mapa, la observación de rasgos que se interfieren no puede ser modelada en forma simple como una distribución independiente.

La **asociación de múltiples hipótesis** mantiene una estimación de la posición separada para cada observación asociada posible. En cada paso de tiempo la observación predicha es utilizada para establecer la cota de validación. Para cada medida que se encuentra dentro de la cota de validación, se genera una nueva hipótesis de posición; así una única posición es dividida en  $n$  posiciones, cada una asociada con cada medida válida más una (usualmente denominada 0) para la hipótesis de no asociación. Cada una de esas nuevas posiciones es tratada entonces independientemente y utilizada para generar nuevas predicciones para el próximo instante de tiempo. Dado que el número de ramas en las que se puede separar la posición estimada puede crecer exponencialmente, se calcula una función de probabilidad para cada posición y las poco probables son descartadas. Este método está obviamente afectado por la carga de cálculo y la memoria necesaria

dado que a cada paso cada hipótesis puede generar varias hipótesis y se deben correr varios filtros de estimación en paralelo. Además es fuertemente dependiente del método de poda de las posiciones poco probables. Este algoritmo sin embargo es útil en situaciones donde hay bajo agrupamiento de mojonos pero con alta incertidumbre.

## 2.6 INTEGRACIÓN DE GPS CON UNIDADES INERCIALES

### 2.6.1 *Unidades Inerciales*

Los sensores inerciales realizan mediciones del estado interno del vehículo. Una de las principales ventajas de los sensores inerciales es que no son radiantes ni bloqueables y pueden empaquetarse y sellarse del medio ambiente. Esto los hace potencialmente robustos en condiciones ambientales adversas. Históricamente, los sistemas de navegación inercial (INS) se han utilizado en vehículos aeroespaciales, aplicaciones militares como barcos, submarinos, misiles y, en mucha menor medida, en aplicaciones de vehículos terrestres. Motivados por los requisitos de la industria automotriz y de la electrónica de consumo (teléfonos inteligentes por ejemplo) se dispone en la actualidad una gran variedad de sistemas inerciales de bajo costo en diversas aplicaciones, desde la determinación del rumbo y la actitud a evaluar los ejercicios que una persona hace. El tipo más común de sensores inerciales son los acelerómetros y los giroscopios. Los acelerómetros miden la aceleración con respecto a un marco de referencia inercial. Esto incluye aceleración gravitacional y rotacional, así como aceleración lineal. Los giroscopios miden la velocidad de rotación independientemente del marco de coordenadas. La aplicación más común de los sensores inerciales es el uso de un giróscopo para obtener el rumbo. La integración de la información de la velocidad del giróscopo proporciona la orientación del sensor. Otra aplicación de los sensores inerciales es el uso de acelerómetros como inclinómetros y para realizar análisis de vibraciones. El uso más sofisticado del sensor de inercia es en los sistemas de navegación completos de seis grados de libertad. En este caso, se utilizan al menos tres giroscopios para seguir la orientación de la plataforma y se integran un mínimo de tres acelerómetros para proporcionar velocidad y posición. También pueden proporcionar información de posición en 3-D y, a diferencia de los codificadores, tienen el potencial de observar el deslizamiento de las ruedas.

#### 2.6.1.1 *Acelerómetros*

Los acelerómetros miden la fuerza de inercia generada cuando una masa se ve afectada por un cambio de velocidad. Esta fuerza puede cambiar la tensión de un resorte o provocar una desviación de un haz de luz o incluso puede cambiar la frecuencia de vibración

de una masa. Una especificación importante de los acelerómetros es la aceleración mínima que se puede medir. Esto es de fundamental importancia cuando se trabaja en aplicaciones de vehículos terrestres, donde la aceleración esperada suele estar en el rango de 0,1-0,3 g (g: aceleración gravitacional). Un acelerómetro estándar para este tipo de aplicaciones debe ser capaz de medir una aceleración inferior a 500  $\mu$ g. La dependencia del sesgo con la temperatura y la linealidad del dispositivo son también especificaciones importantes.

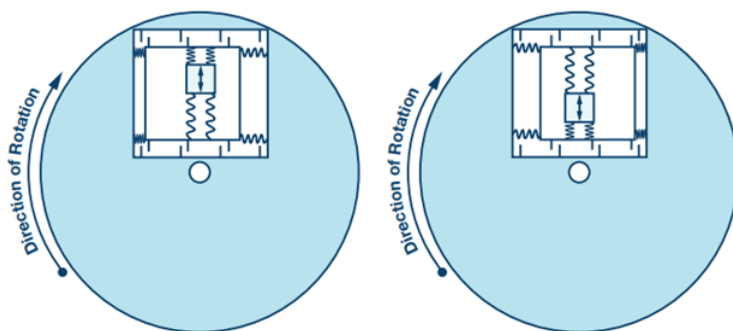
Hay tres familias principales de tecnologías de detección de aceleración lineal: por desplazamiento / reequilibrio de masa traslacional / pendular que incluye estrategias de fabricación por restricción eléctrica, rotacional o elástica; por frecuencia del elemento resonante que incluye fabricaciones por cuerda vibrante, por haz vibrante o por diapasón; y térmicos. Dentro de esta división, se obtienen sensores **MEMS** (silicio resonante / pendular, silicio lateral, acelerómetros de cuarzo resonante) y **electromecánico y o piezoeléctrico**

#### 2.6.1.2 *Giroscopios*

Los giroscopios devuelven una señal proporcional a la velocidad de rotación y su rendimiento se caracteriza principalmente por dos parámetros: el factor de escala y la deriva. Un giroscopio con una baja distorsión en el factor de escala será capaz de detectar con precisión las velocidades angulares en diferentes rangos. La deriva del giroscopio es la salida nominal del giroscopio en estado estacionario.

Hay una gran variedad de giroscopios que se basan en diferentes principios. El precio y la calidad de estos sensores varían considerablemente. Los giroscopios se basan en 3 tecnologías de detección. Los giroscopios mecánicos basados en vibraciones (basados en la fuerza de Coriolis); los giroscopios ópticos (basados en el efecto Sagnac) y los giroscopios resonantes (basados en una masa resonante). Basados en estos principios se pueden identificar siete tipos de giroscopios: giroscopios de masa mecánicos / giratorios; giroscopios de suspensión eléctrica (electrostática) (ESG); giroscopios láser de anillo (RLG); giroscopios de fibra óptica (FOG); giroscopios resonadores hemisféricos (HRG); giroscopios de cuarzo (no MEMS); giroscopios micromecanizados (MEMS); u la tecnología antigua que son giroscopios mecánicos sintonizados dinámicamente, también llamados giroscopios sintonizados dinámicamente (DTG)

Los sensores MEMS los podemos imaginar como un bloque capacitivo que oscila a una velocidad fija en fase opuesta con otro bloque. A medida que el dispositivo gira, las fuerzas de Coriolis de los bloques los mueven ligeramente en direcciones opuestas (ambas perpendiculares al eje de rotación), debido a las diferencias de fase de los bloques. Esta diferencia de fuerza cambia la capacidad de la placa debajo de ella para medir la tasa angular total de un objeto. En realidad, los bloques son más como mallas, pero se aplica el mismo principio bá-

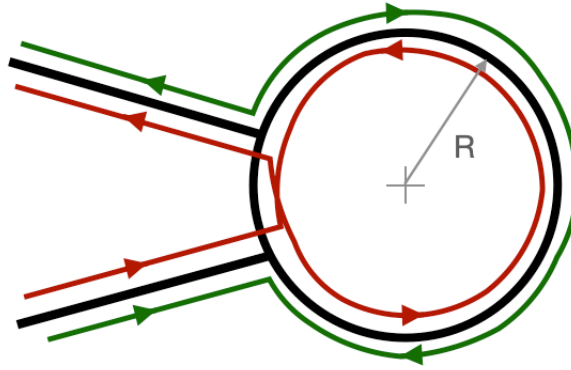


**Figura 2.29:** Esquema de un giroscopio MEMS. El efecto coriolis hace que la diferencia de fuerza que experimenta una masa mas o menos alejada del centro de una superficie que gira. Eso hace que la estructura vibrante se corra hacia los costados, que es medido por los *dedos* arriba y abajo

sico. Los giroscopios MEMS son típicamente de 3 ejes, pero pueden ser de un solo eje o de dos ejes. También se encuentran disponibles diferentes niveles de calidad, según los requisitos de la aplicación. Por ejemplo, los giroscopios de consumo tienden a ser más baratos que los destinados al uso industrial o automotriz, que operan a rangos de temperatura más amplios y están diseñados y probados para tener un rendimiento más consistente con el fin de cumplir con los estrictos requisitos de seguridad. Son más pequeño que otros giroscopios y consumen menos potencia que otros giroscopios. En su contra, poseen menos estabilidad frente a la temperatura, la humedad y el estrés que se pueden compensar con software; poseen mayor error de integración en relación en relación a otras tecnologías y el rendimiento variado entre lotes e incluso unidades.

Los giroscopios de fibra óptica están basados en el efecto Sagnac, descubierto por Georges Sagnac en 1913. Este efecto puede explicarse fácilmente suponiendo dos ondas de luz que circulan en sentido contrario por una trayectoria de radio  $R$ . Si la fuente gira a una velocidad  $\omega$ , la luz que viaja en sentido contrario llegará antes a la fuente que la onda que viaja en el mismo sentido. La mayoría de las implementaciones de bajo costo de estos dispositivos funcionan a lazo abierto. La información de velocidad de rotación es extremadamente precisa, no posee partes móviles y no se basa en la resistencia inercial al movimiento por lo que suelen tener una resolución más alta que los mecánicos. Por el otro lado, tiene mayores exigencias de calibración en la producción y solo se puede usar para un solo eje.

Los giroscopios láser de anillo funcionan también con el efecto Sagnac pero analiza los cambios de frecuencia que se experimenta con la velocidad angular. Cuando se gira, el rayo que se mueve en el sentido del movimiento reduce su frecuencia en el detector. El rayo en sentido antihorario viaja en contra de la rotación, aumentando su frecuencia. Estos giroscopios Poseen una precisión con una incertidumbre de sesgo mejor que  $0,01^\circ/\text{hora}$ ; el tiempo medio entre fallos (MTBF) es



**Figura 2.30:** Esquema de una giróscopo de fibra óptica. El efecto Sagnac explica que ondas de luz que circulan en sentidos contrarios llegarán en momentos diferentes.

superior a 60.000 horas; y no tiene partes mecánicas o móviles que creen fricción por lo que no hay deriva. Por otro lado para rotaciones muy lentas, diferencias entre las ondas pueden volverse tan similares que no se pueden distinguir, que se conoce como bloqueo. Finalmente, solo se puede usar para un solo eje, para obtener información de múltiples ejes, se necesitan múltiples dispositivos.

En navegación los sensores inerciales pueden utilizarse como inclinómetros, para obtener un rumbo y/o como una unidad inercial completa. Para el inclinómetro se usa la detección del vector de gravedad con los acelerómetros. El rumbo o *heading* se obtiene por la integración de los giróscopos pero tienen que ser aquellos con sesgo cero y una muy pequeña varianza de ruido, cosa que no sucede con aquellos disponible para vehículos terrestres. La unidad inercial completa permite una navegación de 6 grados de libertad: *x*, *y*, *z*, guiñada o alabeo, elevación o cabeceo y dirección (*roll*, *pitch*, *yaw*).

#### 2.6.1.3 Modelos de medida

Los errores de medición asociados con los sensores inerciales dependen del principio operativo físico del propio sensor. La velocidad angular puede ser modelada con la siguiente ecuación de observación

$$\tilde{\omega}_{ib}^b = \omega_{ib}^b + \mathbf{b}_g + S_g \omega_{ib}^b + N_g \omega_{ib}^b + \epsilon_g \quad (2.80)$$

donde

$\tilde{\omega}_{ib}^b$	es el vector de medidas del giróscopo [grados/h]
$\omega_{ib}^b$	es el vector de velocidades angulares real [grados/h]
$\mathbf{b}_g$	es el vector del sesgo del giróscopo [grados/h]
$S_g$	es una matriz que representa el factor de escala
$N_g$	es una matriz que representa la no ortogonalidad de la tríada de medidas
$\epsilon_g$	es el vector que representa el ruido del sensor [grados/h]

Las matrices  $N_g$  y  $S_g$  tienen la siguiente forma:

$$N_g = \begin{bmatrix} 1 & \theta_{g,xy} & \theta_{g,xz} \\ \theta_{g,yx} & 1 & \theta_{g,yz} \\ \theta_{g,zx} & \theta_{g,zy} & 1 \end{bmatrix} \quad (2.81)$$

$$N_g = \begin{bmatrix} S_{g,x} & 0 & 0 \\ 0 & S_{g,y} & 0 \\ 0 & 0 & S_{g,z}1 \end{bmatrix} \quad (2.82)$$

donde  $\theta_{(\cdot),(\cdot)}$  son pequeños ángulos que representan los desajustes entre los diferentes ejes del giróscopo y  $S_{(\cdot),(\cdot)}$  son factores de escala para los tres giros.

Los factores que describen el rendimiento y la precisión del acelerómetro son similares a los que caracterizan a los del giroscopio. Las mediciones de la fuerza específica se pueden modelar mediante la ecuación de observación

$$\tilde{\mathbf{f}}^b = \mathbf{f}^b + \mathbf{b}_a + S_1 \mathbf{f}^b + S_2 \mathbf{f}^{b^2} + N_a \mathbf{f}^b + \delta_g + \epsilon_a \quad (2.83)$$

donde

- $\tilde{\mathbf{f}}^b$  es el vector de medidas del acelerómetro [ $\text{m/s}^2$ ]
- $\mathbf{f}^b$  es el vector de aceleraciones real [ $\text{m/s}^2$ ]
- $\mathbf{b}_a$  es el vector del sesgo del acelerómetro [ $\text{m/s}^2$ ]
- $S_1$  es una matriz que representa el factor de escala lineal
- $S_2$  es una matriz que representa el factor de escala no lineal
- $N_a$  es una matriz que representa la no ortogonalidad de la tríada de medidas
- $\delta_g$  es la desviación del valor de gravedad teórico [ $\text{m/s}^2$ ]
- $\epsilon_g$  es el vector que representa el ruido del sensor [ $\text{m/s}^2$ ]

Las matrices  $N_a$ ,  $S_1$  y  $S_2$  tienen la siguiente forma:

$$N_a = \begin{bmatrix} 1 & \theta_{a,xy} & \theta_{a,xz} \\ \theta_{a,yx} & 1 & \theta_{a,yz} \\ \theta_{a,zx} & \theta_{a,zy} & 1 \end{bmatrix} \quad (2.84)$$

$$S_1 = \begin{bmatrix} S_{1,x} & 0 & 0 \\ 0 & S_{1,y} & 0 \\ 0 & 0 & S_{1,z}1 \end{bmatrix} \quad (2.85)$$

$$S_2 = \begin{bmatrix} S_{2,x} & 0 & 0 \\ 0 & S_{2,y} & 0 \\ 0 & 0 & S_{2,z}1 \end{bmatrix} \quad (2.86)$$



donde  $\theta_{(\cdot),(\cdot)}$  son pequeños ángulos que representan los desajustes entre los diferentes ejes del acelerómetro y  $S_{(\cdot),(\cdot)}$  son factores de escala para los tres ejes.

Para ambos sensores, los factores de escala y los sesgos generalmente se consideran constantes (durante un cierto tiempo) pero su valor es desconocido y no están correlacionadas entre los diferentes sensores. En principio, estos errores pueden eliminarse mediante las técnicas de calibración.

La calibración se define como el proceso de comparación de las salidas del instrumento con la información de referencia conocida y así determinar los coeficientes que hagan coincidir la salida con esa información de referencia en el rango deseado de trabajo. La calibración se utiliza para calcular errores deterministas de sensores en el laboratorio. Los parámetros de calibración a determinar pueden cambiar de acuerdo con la tecnología específica de una unidad inercial. Para determinar con precisión todos los parámetros, se necesitan dispositivos de calibración especiales, como plataformas giratorias de tres ejes, para realizar una prueba estática de seis posiciones o una prueba de velocidad angular. Sin embargo, se pueden aproximar esos valores utilizando las siguientes técnicas.

Hay que tener en cuenta que la temperatura y el efecto memoria juegan un papel importante en la estabilidad de la salida de los sensores inerciales de bajo costo. Por esta razón, cuando se compran unidades inerciales de este tipo, no están disponibles todos los valores de los términos de error en sus hojas de datos, por lo que es necesario realizar pruebas en función de la aplicación.

Los acelerómetros normalmente se calibran detectando la gravedad. Cada acelerómetro se coloca en una tabla, calibrada en lo posible, con su eje sensible hacia arriba. Después de tomar alrededor de 10 a 15 minutos de datos, se calcula la  $f_{up}$  media. Luego, se toma una lectura similar,  $f_{down}$ , con el eje sensible apuntando hacia abajo. Las mediciones con el eje sensible del acelerómetro hacia arriba y hacia abajo se pueden expresar como

$$f_{up} = b_a + (1 + S_a)g \quad \text{y} \quad f_{down} = b_a - (1 + S_a)g \quad (2.87)$$

El sesgo  $b_a$  se calcula sumando ambas ecuaciones:

$$b_a = \frac{f_{up} + f_{down}}{2} \quad (2.88)$$

y el factor de escala

$$S_a = \frac{f_{up} - f_{down} - 2g}{2g} \quad (2.89)$$

donde  $g$  es la gravedad. Esto se realiza para cada uno de los ejes y a la temperatura de trabajo. Recordemos que todos estos factores dependerán mucho de la temperatura, por lo que estos sesgos se suelen denominar sesgos de funcionamiento.

Para los giroscopios se emplea un procedimiento similar, pero esta vez se mide la velocidad de rotación de la Tierra ( $\omega_e$ ) en lugar de la gravedad. Para un dispositivo ubicado en una latitud  $\phi_{lat}$ , las proyecciones teóricas de la tasa de rotación de la Tierra sobre los ejes del cuerpo son

$$\omega_x = 0; \omega_y = \omega_e \cos \phi_{lat}; \omega_z = \omega_e \sin \phi_{lat} \quad (2.90)$$

Por lo tanto, un giroscopio vertical (con su eje sensible apuntando hacia arriba) detectará un componente de la gravedad que es  $\omega_e \sin \phi_{lat}$ , es decir el componente vertical de la rotación de la tierra, y esto se usa en la calibración. Las medidas del giroscopio con su eje sensible hacia arriba y hacia abajo son

$$\omega_{up} = b_{go} + (1 + S_g)\omega_e \sin \phi_{lat} \quad y \quad \omega_{down} = b_{go} - (1 + S_g)\omega_e \sin \phi_{lat} \quad (2.91)$$

El sesgo y el factor de escala se calculan de manera similar a la aceleración:

$$b_g = \frac{\omega_{up} + \omega_{down}}{2} \quad (2.92)$$

y el factor de escala

$$S_g = \frac{\omega_{up} - \omega_{down} - 2\omega_e \sin \phi_{lat}}{2\omega_e \sin \phi_{lat}} \quad (2.93)$$

Esto se realiza para cada uno de los ejes y de nuevo, a la temperatura de trabajo. Si los giroscopios son muy baratos puede ser que no detecten la velocidad de rotación de la tierra. Se puede entonces rotar la tabla calibrada a una velocidad que supere la de detección de los giroscopios y se reemplaza el término  $\omega_e \sin \phi_{lat}$  por esta velocidad.

También hay un efecto de histéresis con la temperatura que se encuentra con la mayoría de los sensores inerciales. Por lo tanto, puede ser necesario hacer pruebas recorriendo el rango de temperatura de trabajo de la misma manera cada vez. Esto se conoce como el sesgo de encendido a encendido.

Es posible utilizar máquinas de precisión que aseguren las medidas e incluso integrar las salidas para obtener mas precisión en las correcciones. Sin embargo, los métodos comentados pueden ser suficientes para corregir los errores mas importantes, especialmente en las IMU mas económicas.

**CALIBRACIÓN EN CAMPO.** El método más simple, aunque menos preciso, de obtener las desviaciones de los sensores inerciales instalados en un vehículo es medir las lecturas de cada sensor mientras el vehículo está parado. Estos valores de sesgo se utilizan para calibrar la IMU. Para los giroscopios, el sesgo es simplemente la lectura de estos sensores cuando el vehículo está parado. Sin embargo, se requiere

la alineación de la unidad inercial para determinar los sesgos en los acelerómetros. Esto se logra durante la etapa de alineación, ya que se puede determinar la aceleración *esperada* debida a la gravedad y, por lo tanto, cualquier anomalía en estos valores se atribuye al sesgo. Una estrategia de calibración mas detallada de una unidad inercial montada en un vehículo puede verse en [14].

#### 2.6.1.4 Teorema de Coriolis

La navegación con respecto a un marco giratorio como la Tierra requiere plantear las ecuaciones a partir del teorema de Coriolis. El teorema establece que la velocidad de un vehículo con respecto a un marco inercial fijo  $\mathbf{v}_i$  es igual a la velocidad del móvil con respecto al suelo  $\mathbf{v}_e$  (la velocidad del vehículo con respecto a la tierra), más la velocidad del vehículo debido a la velocidad de rotación de la tierra con respecto al marco inercial  $\boldsymbol{\omega}_{ie}$ , en el punto de la tierra donde se encuentra el vehículo  $\mathbf{r}$ , es decir,

$$\mathbf{v}_i = \mathbf{v}_e + \boldsymbol{\omega}_{ie} \times \mathbf{r} \quad (2.94)$$

donde  $\boldsymbol{\omega}_{ie} = [0 \ 0 \ \Omega]$  y  $\Omega$  es la velocidad de rotación de la tierra. Derivando esta ecuación con respecto a los ejes inerciales obtenemos

$$\dot{\mathbf{v}}_{i/i} = \dot{\mathbf{v}}_{e/i} + \dot{\boldsymbol{\omega}}_{ie/i} \times \mathbf{r} + \boldsymbol{\omega}_{ie} \times \mathbf{v}_{/i} \quad (2.95)$$

Suponiendo que la aceleración angular de la tierra es cero, entonces  $\dot{\boldsymbol{\omega}}_{ie} = 0$  y usando de la ecuación 2.94 el término,

$$\boldsymbol{\omega}_{ie} \times \mathbf{v}_{/i} = \boldsymbol{\omega}_{ie} \times [\mathbf{v}_e + \boldsymbol{\omega}_{ie} \times \mathbf{r}] \quad (2.96)$$

$$= \boldsymbol{\omega}_{ie} \times \mathbf{v}_e + \boldsymbol{\omega}_{ie} \times [\boldsymbol{\omega}_{ie} \times \mathbf{r}] \quad (2.97)$$

la ecuación 2.95 queda

$$\dot{\mathbf{v}}_{i/i} = \dot{\mathbf{v}}_{e/i} + \boldsymbol{\omega}_{ie} \times \mathbf{v}_e + \boldsymbol{\omega}_{ie} \times [\boldsymbol{\omega}_{ie} \times \mathbf{r}] \quad (2.98)$$

Ahora,  $\dot{\mathbf{v}}_{i/i} = \mathbf{f} + \mathbf{g}$  donde  $\mathbf{f}$  es la fuerza específica debida al movimiento del móvil y  $\mathbf{g}$  es la fuerza de la gravedad, por lo que la ecuación 2.98 queda,

$$\mathbf{f} + \mathbf{g} = \dot{\mathbf{v}}_{e/i} + \boldsymbol{\omega}_{ie} \times \mathbf{v}_e + \boldsymbol{\omega}_{ie} \times [\boldsymbol{\omega}_{ie} \times \mathbf{r}] \quad (2.99)$$

con lo cual

$$\dot{\mathbf{v}}_{e/i} = \mathbf{f} - [\boldsymbol{\omega}_{ie} \times \mathbf{v}_e] + [\mathbf{g} - \boldsymbol{\omega}_{ie} \times [\boldsymbol{\omega}_{ie} \times \mathbf{r}]] \quad (2.100)$$

La ecuación 2.100 simplemente establece que la aceleración sobre la superficie de la Tierra es igual a la aceleración medida por los acelerómetros compensados por la aceleración de Coriolis encontrada debido a la velocidad del vehículo sobre una Tierra en rotación,  $\boldsymbol{\omega}_{ie} \times \mathbf{v}_e$  y para la aceleración de la gravedad local, que se compone de la gravedad de la Tierra,  $\mathbf{g}$ , y debido a la rotación de la Tierra, también conocida como aceleración centrípeta  $\boldsymbol{\omega}_{ie} \times [\boldsymbol{\omega}_{ie} \times \mathbf{r}]$ .

### 2.6.1.5 Navegación en áreas locales, el marco de referencia de la tierra

En el marco de referencia de la Tierra, la aceleración del vehículo con respecto a la tierra  $\dot{\mathbf{v}}_{e/e}$ , es igual a la aceleración del vehículo con respecto al marco de inercia  $\dot{\mathbf{v}}_{e/i}$ , menos la aceleración de Coriolis debida a la velocidad del vehículo.  $\mathbf{v}_e$ , sobre una tierra en rotación, es decir,

$$\dot{\mathbf{v}}_{e/e} = \dot{\mathbf{v}}_{e/i} - \boldsymbol{\omega}_{ie} \times \mathbf{v}_e \quad (2.101)$$

El marco de referencia de la Tierra se utilizará a lo largo de este trabajo, ya que su definición es muy adecuada para aplicaciones de vehículos terrestres. Sustituyendo en la ecuación 2.100 se obtiene

$$\dot{\mathbf{v}}_{e/e} = \mathbf{f}_e - 2[\boldsymbol{\omega}_{ie} \times \mathbf{v}_e] + [\mathbf{g} - \boldsymbol{\omega}_{ie} \times [\boldsymbol{\omega}_{ie} \times \mathbf{r}]] \quad (2.102)$$

Nuevamente, dado que la unidad de inercia mide la aceleración en el marco de referencia del móvil, las medidas de aceleración deben transformarse en el marco de la tierra,

$$\mathbf{f}_e = \boldsymbol{\Upsilon} \mathbf{f}_b \quad (2.103)$$

donde  $\boldsymbol{\Upsilon}$  ahora forma parte de las velocidades de rotación  $\boldsymbol{\omega}_{be}$  que relaciona el cuerpo con el marco de la tierra. Sin embargo, los giroscopios miden la rotación inercial total  $\boldsymbol{\omega}_{ie}$  que comprende  $\boldsymbol{\omega}_{be}$  más la rotación de la tierra con respecto al marco inercial transformado en el marco de referencia del móvil, por lo que

$$\boldsymbol{\omega}_{be} = \boldsymbol{\omega}_{ie} - \boldsymbol{\Upsilon}_e^b \boldsymbol{\omega}_{ie} \quad (2.104)$$

La matriz de transformación  $\boldsymbol{\Upsilon}$  tiene que ser precisa ya que la desalineación (errores en la actitud estimada) hace que los componentes del vector de gravedad medidos por los acelerómetros se confundan con la verdadera aceleración del vehículo. Integrados con el tiempo, incluso los pequeños errores de desalineación causarán grandes errores estimados.

La matriz de transformación consta de los ángulos de balanceo, elevación y dirección necesarios para rotar los ejes de referencia del cuerpo los de navegación y, por lo tanto, se actualiza continuamente ya que los ejes del cuerpo siempre están girando con respecto a los de navegación. El proceso de actualización propaga esta matriz en base a los datos obtenidos de los giroscopios, por lo que cualquier error en  $\boldsymbol{\Upsilon}$  es causado tanto por los errores físicos asociados con los giroscopios como por los errores en los algoritmos utilizados para propagar la matriz de transformación.

En principio, existen tres técnicas para propagar la matriz de transformación: de Euler, de cosenos directores (*DCM: Direction Cosine Matrix*) y de cuaternión. El enfoque de Euler no se usa comúnmente debido a la presencia de la singularidad de balanceo y dirección. Esto no

representa un problema para las aplicaciones de vehículos terrestres. Sin embargo, el enfoque de Euler requiere un gran esfuerzo computacional y, por lo tanto, un procesamiento potente si se requieren actualizaciones rápidas.

El cuaternion tiene la ventaja de que solo requiere la actualización de cuatro variables. Se utiliza con mayor frecuencia en aplicaciones militares y espaciales. Para ahorrar en costos computacionales, el cuaternión se convierte en una representación DCM para la transformación de la aceleración desde los ejes del móvil al de navegación. Esto puede ser una carga computacional significativa si se requieren velocidades de muestreo rápidas. Por ejemplo, un sistema Laser INS (LINS) típicamente muestrea los giroscopios de anillo láser a 1200Hz y los acelerómetros a 400Hz. La frecuencia de muestreo más baja es suficiente para las aplicaciones de navegación LINS típicas. Se utilizan tres muestras de los giros para actualizar el cuaternión, formando así un *super* cuaternión que representa una sola rotación. Este cuaternión se convierte luego en la representación DCM y se utiliza para transformar las medidas de aceleración.

El método DCM se usa comúnmente en todas las formas de navegación inercial y encaja en el medio, tanto en términos de comprensión física como en gastos computacionales. Aunque se actualizan nueve variables en cada muestra, es menos costosa computacionalmente que la representación de Euler y, a diferencia del enfoque de Quaternion, no requiere conversión a otra representación para transformar los datos de aceleración del marco del cuerpo al marco de navegación. En el algoritmo se hace una suposición de ángulo pequeño, por lo que las tasas de actualización tienen que ser altas. Por ejemplo, los vehículos terrestres pueden ejecutar velocidades de rotación de 50 grados/seg, por lo que se requiere una velocidad de muestreo de al menos 1 kHz (para mantener los ángulos por debajo de 0,05 grados). Tal frecuencia de muestreo es alta para sistemas inerciales de bajo costo, aunque en sistemas más costosos utilizados en aplicaciones militares y espaciales, 20 kHz no es infrecuente.

### 2.6.2 Integración

La reducción de costos de las unidades inerciales viene acompañada de una reducción sustancial de la calidad. Esto significa que el tiempo en que una IMU puede navegar sin ninguna ayuda es proporcional a su calidad. La solución es ayudar a los sistemas inerciales con información externa a mantener el error dentro de ciertos límites. El sensor de ayuda más común para aplicaciones en exteriores ha sido y es el GPS en todas sus formas (Autónomo / Diferencial / RTK). La tabla 2.2 muestra el desempeño de varias unidades inerciales en función de su costo; la tabla 2.3 presenta los parámetros típicos para

Calidad	Costo	Estabilidad del sesgo del gyro	Tiempo de navegación sin GPS	Aplicaciones
Consumo	<\$10	-	-	Smartphones
Industrial	\$100-\$1.000	< 10°/hora	<1 minuto	UAV/UGV
Táctico	\$5.000-\$50.000	< 1°/hora	<10 minutos	Municiones
Navegación	\$100.000	< 0,1°/hora	Varias horas	Militar

**Tabla 2.2:** Desempeño de sensores inerciales

Calidad	Sesgo Acelerómetro (mg)	Random Walk de velocidad (m/s/ $\sqrt{\text{hr}}$ )	Sesgo Giro (grado/hr)	Random Walk de ángulo (grado/ $\sqrt{\text{hr}}$ )
Consumo	10	1	100	2
Industrial	1	0.1	10	0.2
Táctico	0.1	0.03	1	0.05
Navegación	0.01	0.01	0.01	0.01

**Tabla 2.3:** Errores de sensores inerciales

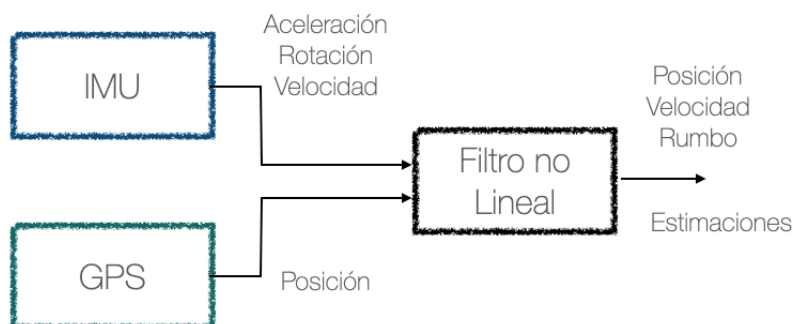
diferentes tipos de calidad de sensores; y la tabla 2.4 el crecimiento del error en posición y como crece en el tiempo.

La implementación más natural de un sistema de navegación inercial asistida es implementar un filtro no lineal con los datos brutos de velocidad de rotación y aceleración proporcionados por la IMU, como se muestra en la Figura 2.31. La implementación se conoce como una estructura de filtro "directa". El modelo de proceso generalmente representa la relación cinemática del vehículo y los estados de interés. El vector de estado se propaga por el modelo y los datos inerciales. La información externa se obtiene de GPS que suministran al sistema observaciones de posición y la velocidad. La estimación estaría en forma de estados del vehículo.

La desventaja de esta implementación es que las ecuaciones de predicción deben evaluarse en cada muestra de los datos inerciales. Esto

Calidad	1s	10s	60s	10min	1hr
Consumo	6cm	6.5m	400m	200km	39.000km
Industrial	6mm	0.7m	40m	20km	3.900km
Táctico	1mm	8cm	5m	2km	400km
Navegación	<1mm	1mm	50cm	100m	10km

**Tabla 2.4:** Crecimiento del error de sensores inerciales

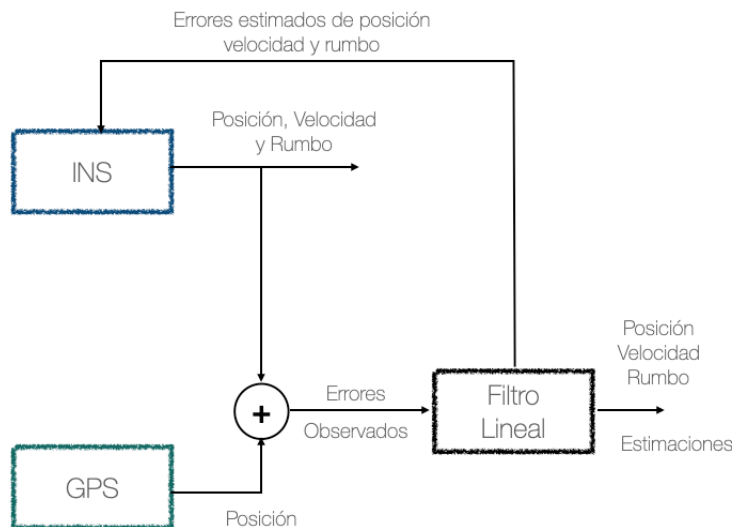


**Figura 2.31:** La estructura "directa" implementa un filtro no lineal para estimar la posición, velocidad y actitud del vehículo. Los datos de inercia los proporciona una IMU y los datos externos un GPS.

requiere un procesamiento sustancial debido a las altas tasas de muestreo de las IMU. Otro problema es que, en general, los movimientos de alta frecuencia se suelen filtrar en el modelo linealizado. La consecuencia de esta omisión es que el filtro atenuará innecesariamente la información de alta frecuencia proporcionada por el INS. Con este enfoque, es posible que el sistema no pueda realizar un seguimiento de las maniobras rápidas. Para superar esto, debe emplearse un sistema de navegación inercial (INS) de modo que un flujo constante de información sobre el estado del vehículo esté disponible fuera del filtro. Para corregir cualquier error, el filtro estima los errores en estos estados. La información inercial aún se puede obtener incluso si no se dispone de información adicional. La Figura 2.32 muestra una implementación que se conoce como el método de *realimentación directa*. La observación que se envía al filtro es el "error observado" de la solución de navegación inercial, es decir, la diferencia entre la solución de navegación inercial y la información del GPS. En esta implementación, la información inercial de alta frecuencia se envía directamente a la salida sin atenuación, mientras que el filtro de Kalman proporciona la corrección de baja frecuencia a la IMU. Dado que la observación es el error observado de la solución de navegación inercial, y dado que el filtro está estimando los errores en la solución de navegación inercial, el modelo de proceso debe tener la forma de un modelo de error de las ecuaciones de navegación inercial estándar. Por tanto, las ecuaciones de navegación inercial se linealizan para formar ecuaciones de error y el filtro adquiere una forma lineal. Por otro lado, los errores estimados se retroalimentan al INS, minimizando así el crecimiento del error observado que se entrega como una observación al filtro.

Como ya se ha dicho, en aplicaciones móviles terrestres no se requiere un muestreo extremadamente alto de sensores de medición inercial, ya que las frecuencias máximas de rotación y aceleración no son altas (con una obvia consideración a la vibración). Además, la potencia de los procesadores modernos es suficiente para manejar el





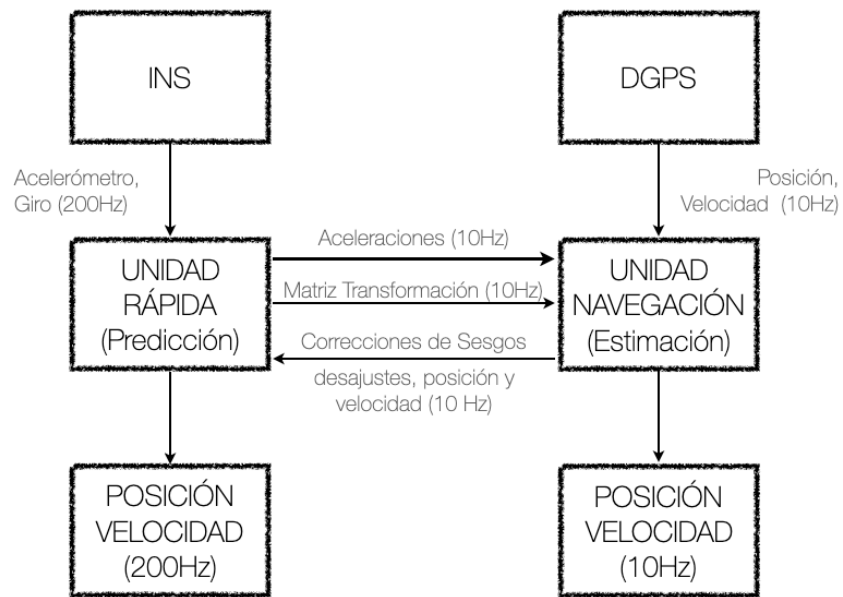
**Figura 2.32:** El método de realimentación directa resuelve los problemas del método directo.

rendimiento de datos requerido en estas aplicaciones. Un esquema de implementación del esquema de navegación por realimentación directa se muestra en la figura 2.33.

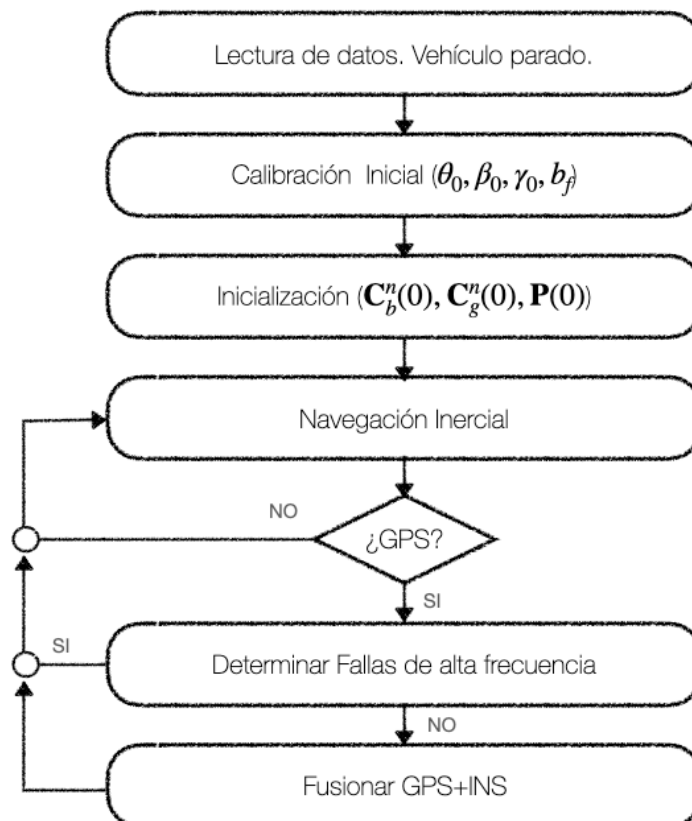
El sistema se puede implementar como dos procesos independientes. El primer proceso se implementa en la unidad rápida de guiado. Este proceso se comunica con la unidad inercial a través de un enlace muy rápido. La unidad de medición inercial transmite información de aceleración y giro triaxial a alta frecuencia. La unidad de guía integra la matriz de coseno de dirección y transforma la aceleración en el marco de navegación. También genera las predicciones de velocidad y posición mediante integración simple y doble, respectivamente. Las posiciones, velocidades, aceleraciones y la matriz de transformación se transmiten a la unidad de navegación a un ritmo mucho menor. Aunque la información del estado del vehículo ya se ha transformado en el marco de coordenadas de navegación local, todavía se requiere la matriz de transformación para implementar el algoritmo de fusión de datos. La unidad de navegación recibe información de posición y velocidad observadas del receptor DGPS a velocidades mucho más bajas de entre 5 y 20 Hz y es responsable de implementar el algoritmo de filtro de Kalman de retroalimentación directa.

El flujo del algoritmo de navegación se puede implementar de la siguiente manera. Varias de las referencias a ecuaciones y secciones las veremos luego. Esta presentación es para entender el proceso de integración y un diagrama de flujo se presenta en la figura 2.34.

1. Dado que el vehículo es generalmente autónomo, todos los comandos de guía se conocen de antemano y, por lo tanto, el filtro de navegación sabe cuándo el vehículo está parado. Mientras el vehículo está parado, el sistema lee todos los datos del sensor



**Figura 2.33:** Arquitectura de Navegación por realimentación directa.



**Figura 2.34:** Flujo del algoritmo de Integración por realimentación directa.

de aceleración, velocidad de rotación e inclinación (si esta información está disponible) de la unidad inercial y proporciona el promedio de estas lecturas del sensor. El tiempo de registro para este promedio depende de la vibración real presente en el sistema.

2. Una vez que se ha completado el proceso de promediado, se logra la calibración de la unidad inercial como se describe en la Sección 2.6.2.3. Esto se usa para determinar los sesgos y obtener la matriz  $C_n^b$  inicial. En esta etapa, la posición del GPS se usa para determinar la matriz  $C_n^g$  inicial, Sección 2.6.2.2.
3. El sistema de navegación pasa entonces al sistema de navegación inercial con la posición inicial determinada por el receptor GPS y la actitud inicial obtenida de la etapa de alineación. Inicialmente, es posible que el sensor GPS no proporcione la posición del vehículo hasta que haya alcanzado su solución más precisa, lo que en el caso de los sistemas RTK puede tardar unos minutos. El filtro se inicializa. A continuación, se informa a la computadora de guía que el sistema de navegación está listo para proporcionar la solución de navegación;
4. A medida que el vehículo se mueve, se leen los valores de aceleración y giroscopio y se eliminan las desviaciones, ecuación 2.125. Se actualiza  $C_n^b$ , ecuación 2.110, y se calcula la aceleración en el marco de navegación, ecuación 2.112. Estos valores se integran luego para proporcionar la posición y rumbo del vehículo;
5. Si no hay una posición de GPS disponible, se reitera el paso 4; de lo contrario, se determina la posición de GPS en el marco de navegación, ecuaciones 2.114. Cuando la latencia es pequeña y la dinámica del vehículo es baja, los datos de velocidad se pueden usar para propagar la información de posición utilizando un modelo de velocidad constante con una precisión razonable;
6. Fusionar los datos inerciales y GPS como se describe en la Sección 2.6.2.4;
7. Utilizar la función de activación, ecuación 2.152, para determinar si hay fallas de alta frecuencia. Si existen fallas de alta frecuencia, se regresa al paso 4. Si la verificación de validación ha pasado, se corrige la unidad inercial como se muestra en la Sección 2.6.2.4 y luego se regresa al paso 4. El procedimiento de estimación ocurre dentro del tiempo de muestreo de la unidad inercial, sin embargo, si este no es el caso y nuevamente la latencia es baja, entonces la velocidad corregida se usa para propagar el valor de posición corregido.

## 2.6.2.1 Matriz de Cosenos Directores (DCM)

La matriz de cosenos directores  $\mathbf{C}_b^n$  es una matriz de  $3 \times 3$  que contiene los cosenos de los ángulos entre los ejes de referencia del móvil y los de navegación. En el tiempo,  $\mathbf{C}_b^n$  se propaga como,

$$\dot{\mathbf{C}}_b^n = \mathbf{C}_b^n \mathbf{\Omega}_{bn} \quad (2.105)$$

donde  $\mathbf{\Omega}_{bn}$  es una matriz simétrica sesgada que representa las tasas de rotación como son medidas por los giros.

$$\mathbf{\Omega}_{bn} = \begin{bmatrix} 0 & -\omega_z & \omega_y \\ \omega_z & 0 & -\omega_x \\ -\omega_y & \omega_x & 0 \end{bmatrix} \quad (2.106)$$

Si  $\theta$ ,  $\beta$  y  $\gamma$  son los ángulos iniciales de balanceo, elevación y dirección respectivamente, la matriz inicial  $\mathbf{C}_b^n$  será,

$$\begin{aligned} \mathbf{C}_b^n &= \begin{bmatrix} \gamma_c & -\gamma_s & 0 \\ \gamma_s & \gamma_c & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \beta_c & 0 & \beta_s \\ 0 & 1 & 0 \\ -\beta_s & 0 & \beta_c \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & \theta_c & -\theta_s \\ 0 & \theta_s & \theta_c \end{bmatrix} \\ &= \begin{bmatrix} \beta_c \gamma_c & -\theta_c \gamma_s + \theta_s \beta_s \gamma_c & \theta_s \gamma_s + \theta_c \beta_s \gamma_c \\ \beta_c \gamma_s & \theta_c \gamma_c + \theta_s \beta_s \gamma_s & -\theta_s \gamma_c + \theta_c \beta_s \gamma_s \\ -\beta_s & \theta_s \beta_c & \theta_c \beta_c \end{bmatrix} \end{aligned} \quad (2.107)$$

donde por simplicidad, los subíndices s y c representa los senos y cosenos de los ángulos.

El procedimiento de actualización en el campo discreto es el siguiente:

- Se obtiene las salidas del giróscopo  $\omega_x$ ,  $\omega_y$  y  $\omega_z$  y se integran para determinar los cambios en ángulo  $\phi_x$ ,  $\phi_y$  y  $\phi_z$
- Con estos ángulos se obtiene la magnitud del vector angular:  

$$\phi = \sqrt{\phi_x^2 + \phi_y^2 + \phi_z^2}$$
- Con este ángulo se determinan los siguientes coeficientes:

$$\begin{aligned} \alpha &= \frac{\sin \phi}{\phi} \\ \beta &= \frac{1 - \cos \phi}{\phi^2} \end{aligned} \quad (2.108)$$

- Además construimos la matriz de sesgo

$$\bar{\Phi} = \begin{bmatrix} 0 & -\phi_z & \phi_y \\ \phi_z & 0 & -\phi_x \\ -\phi_y & \phi_x & 0 \end{bmatrix} \quad (2.109)$$

- Actualizamos la matriz de cosenos directores

$$\mathbf{C}_b^n(i+1) = \mathbf{C}_b^n(i)[\mathbf{I}_{3 \times 3} + \alpha \bar{\Phi} + \beta^2 \bar{\Phi}^2] \quad (2.110)$$

La matriz de transformación puede ser simplificada si se suponen rotaciones angulares pequeñas ( $< 0,05$  grados),

$$\mathbf{C}_b^n(i+1) = \mathbf{C}_b^n(i)[\mathbf{I}_{3 \times 3} + \bar{\Phi}] \quad (2.111)$$

- Con esta matriz, se obtienen las aceleraciones en los ejes de navegación en función de las aceleraciones medidas en el marco de referencia del móvil

$$\mathbf{f}_n = \mathbf{C}_b^n(i+1)\mathbf{f}_b \quad (2.112)$$

- se integra la ecuación 2.112 para obtener la velocidad y nuevamente para obtener la posición.

#### 2.6.2.2 Matriz de Transformación de coordenadas GNSS

Esta sección presenta varias ecuaciones de transformación de coordenadas que convierten la posición y la velocidad dadas por el GPS en el marco de navegación local utilizado por los algoritmos de fusión de datos (Norte, Este, Abajo).

Dadas la Latitud y la Longitud, se puede formar una matriz de transformación  $\mathbf{C}_n^g$  que transforme los datos de posición y velocidad proporcionados por el receptor GNSS al marco de navegación.

$$\mathbf{C}_n^g = \begin{bmatrix} -\sin \phi \cos \lambda & -\sin \phi \sin \lambda & \cos \phi \\ -\sin \lambda & \cos \lambda & 0 \\ -\cos \phi \cos \lambda & -\cos \phi \sin \lambda & -\sin \phi \end{bmatrix} \quad (2.113)$$

donde  $\phi$  es la Latitud en grados y  $\lambda$  es la Longitud también en grados.

Por lo tanto, la posición y la velocidad del vehículo en el marco de navegación local es

$$\begin{aligned} \mathbf{P}_n &= \mathbf{C}_n^g \mathbf{P}_g \\ \mathbf{V}_n &= \mathbf{C}_n^g \mathbf{V}_g \end{aligned} \quad (2.114)$$

Si el área de trabajo del vehículo tiene un cambio insignificante en latitud y longitud, entonces  $\mathbf{C}_n^g$  es fija en los hechos. Las ecuaciones 2.114 forman las observaciones necesarias para el sistema de navegación inercial asistida.

#### 2.6.2.3 Calibración y alineación de una unidad de medida inercial

El objetivo de la calibración es determinar los sesgos en los acelerómetros y giroscopios. Esto se obtiene determinando en primer lugar la alineación inicial de la unidad inercial y, por tanto, evaluando a su vez la matriz de coseno de dirección inicial.

**TÉCNICAS DE ALINEACIÓN** Si las lecturas del acelerómetro se conocen perfectamente, entonces la actitud de la unidad inercial se puede determinar resolviendo el componente de gravedad. Para determinar el componente de gravedad medido por los acelerómetros, la Ecuación 2.112 se reordena de la siguiente forma

$$\mathbf{f}_b = (\mathbf{C}_b^n)^{-1} \mathbf{f}_n \quad (2.115)$$

Dado que  $\mathbf{C}_b^n$  es ortogonal, su inversa es simplemente la traspuesta. La unidad inercial está estacionaria y, por lo tanto, la única aceleración medida es la debida a la gravedad a lo largo del eje vertical. Por lo tanto

$$\mathbf{f}_n = [0 \ 0 \ -g]^T \quad (2.116)$$

la ecuación 2.115 queda

$$\begin{bmatrix} f_{xT} \\ f_{yT} \\ f_{zT} \end{bmatrix} = \begin{bmatrix} \beta_c \gamma_c & \beta_c \gamma_s & -\beta_s \\ -\theta_c \gamma_s + \theta_s \beta_s \gamma_c & \theta_c \gamma_c + \theta_s \beta_s \gamma_s & \theta_s \beta_c \\ \theta_s \gamma_s + \theta_c \beta_s \gamma_c & -\theta_s \gamma_c + \theta_c \beta_s \gamma_s & \theta_c \beta_c \end{bmatrix} \begin{bmatrix} 0 \\ 0 \\ -g \end{bmatrix} \quad (2.117)$$

donde  $\theta$ ,  $\beta$  y  $\gamma$  son los ángulos de balanceo, elevación y dirección respectivamente y el subíndice T representa las componentes de la aceleración verdadera debidas a la gravedad. Por lo tanto

$$f_{xT} = g \sin \beta \quad (2.118)$$

$$f_{yT} = -g \sin \theta \cos \beta \quad (2.119)$$

$$f_{zT} = -g \cos \theta \cos \beta \quad (2.120)$$

Aunque ningún sensor es perfecto, cuanto mayor sea la precisión, menores serán las tolerancias de error y, por lo tanto, la precisión de alineación que se puede obtener. A medida que disminuye la precisión de los sensores, debido a los errores mencionados anteriormente, la precisión de la alineación también disminuye. Reordenando la ecuación 2.118 para determinar la elevación  $\beta$ , y sustituyéndolo en las ecuaciones 2.119 o 2.120 se resolverá el balanceo  $\theta$ . Este procedimiento para determinar la alineación se denomina *alineación aproximada*.

Si la precisión proporcionada por la alineación aproximada no es suficiente para el rendimiento de la navegación, se requerirá información de alineación externa. Esta información puede provenir de sensores de inclinación o información de actitud GNSS, por ejemplo. La alineación aproximada se usa generalmente para una alineación

rápida y/o donde no se posee el tiempo requerido para promediar los datos de los sensores externos.

El término final que debe evaluarse es la dirección del vehículo  $\gamma$ . El girocompás es un tipo de compás no magnético que determina el rumbo o dirección del vehículo basado en la rotación de la tierra. Sin embargo, con giroscopios de bajo costo, el girocompás generalmente no está incluido y se requiere de información externa para determinar la dirección inicial.

**ALINEACIÓN PARA UNIDADES INERCIALES DE BAJO COSTO** Para muchas de las unidades inerciales de costos accesibles, ninguno de los métodos comunes de auto alineación o calibración proporciona resultados lo suficientemente precisos para la navegación. Lo visto en la subsección anterior es para IMUs de alta precisión, de uso táctico o estratégico. En su lugar, se puede incluir mediciones externas de inclinación y elevación para alinear la unidad inercial.

Una inclinación lateral (bank angle: ángulo de inclinación lateral) positivo hará que el acelerómetro mida un componente de gravedad igual a

$$f_{yT} = -g \sin(\text{lateral}) \quad (2.121)$$

De una forma similar, una elevación positiva hará que el acelerómetro mida

$$f_{xT} = g \sin(\text{elevación}) \quad (2.122)$$

Igualando la ecuación 2.118 con 2.122 y la ecuación 2.119 con 2.121, los ángulos de cabeceo/elevación y balanceo serán

$$\beta = \text{elevación} \quad (2.123)$$

$$\theta = \sin^{-1} \left( \frac{\sin(\text{lateral})}{\cos \beta} \right) \quad (2.124)$$

Para resolver el rumbo se utiliza un compás. Las ecuaciones 2.123 y 2.124 se utilizan junto con el ángulo del rumbo inicial para determinar la matriz inicial  $C_b^n$ .

**CALIBRACIÓN** El método más simple de obtener las desviaciones de los sensores inerciales es medir las lecturas de cada sensor mientras el vehículo está parado. Estos valores de sesgo se utilizan para calibrar la IMU. Para los giroscopios, el sesgo es simplemente la lectura de estos sensores cuando el vehículo está parado. Sin embargo, se requiere la alineación de la unidad inercial para determinar los sesgos en los acelerómetros. Esto se logra durante la etapa de alineación,



ya que se puede determinar la aceleración *esperada* debida a la gravedad y, por lo tanto, cualquier anomalía en estos valores se atribuye al sesgo. Por tanto, el sesgo del acelerómetro  $x$  se obtiene mediante;

$$b_{fx} = f_x - f_{xT} \quad (2.125)$$

donde  $f_x$  es la aceleración medida y  $f_{xT}$  es la aceleración esperada obtenida durante la etapa de alineación. El sesgo se obtiene de manera similar para los acelerómetros restantes.

#### 2.6.2.4 Implementación de la realimentación directa lineal

Esta arquitectura de navegación está basada en los errores de observación formados con la posición y / o velocidades predichas y observadas. Para implementar el Filtro de Kalman propuesto, se necesita un modelo de propagación de los errores. Sin entrar en los detalles que se pueden consultar en [25], el lazo de navegación implementa un filtro de Kalman lineal. El vector de estados consiste en los estados error,

$$\mathbf{x} = [\delta p_N, \delta p_E, \delta p_D, \delta v_N, \delta v_E, \delta v_D, \delta \phi_N, \delta \phi_E, \delta \phi_D]^T \quad (2.126)$$

donde  $p(\cdot)$  es la posición,  $v(\cdot)$  es la velocidad y  $\phi(\cdot)$  el rumbo; N representa Norte, E Este y D abajo o *Down*. El Marco terrestre o de la tierra a nivel local es también conocido como marco geográfico o terrestre y se define localmente en relación con el geoide terrestre. El eje  $x$  apunta a la dirección del Norte, el eje  $z$  es perpendicular a la tangente del elipsoide que apunta hacia el interior de la Tierra, no necesariamente al centro de la Tierra. Finalmente, el eje  $y$  apunta hacia el este para completar un sistema ortogonal a la derecha. El marco de navegación usualmente se lo hace coincidir con estos ejes.

El modelo de proceso será

$$\mathbf{F}_g = \begin{bmatrix} 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & -f_D & f_E \\ 0 & 0 & 0 & 0 & 0 & 0 & f_D & 0 & -f_N \\ 0 & 0 & 0 & 0 & 0 & 0 & -f_E & f_N & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \quad (2.127)$$

que lo podemos reescribir en forma compacta  $\dot{\mathbf{x}} = \mathbf{F}_g \mathbf{x} + \mathbf{G} \mathbf{w}$  y teniendo en cuenta las rotaciones de los ejes de referencia del móvil y las mediciones de los giros y los acelerómetros,

$$\begin{bmatrix} \dot{r} \\ \dot{v} \\ \dot{\phi} \end{bmatrix} = \begin{bmatrix} 0 & \mathbf{I} & 0 \\ 0 & 0 & \mathbf{a}_n \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} r \\ v \\ \phi \end{bmatrix} + \begin{bmatrix} 0\mathbf{I}_{3 \times 3} & 0 & 0 \\ 0 & \mathbf{C}_b^n & 0 \\ 0 & 0 & \mathbf{C}_b^n \end{bmatrix} \begin{bmatrix} 0 \\ f_a \\ \omega_{ib} \end{bmatrix} \quad (2.128)$$

Este modelo puede ser extendido para considerar los sesgos y las derivas de los acelerómetros y los giros. Un modelo posible del error de los giros consiste en un proceso de Markov de primer orden con tiempo de correlación  $\tau$  y ruido blanco  $v$

$$\dot{\theta} = -(1/\tau)\theta + v_g \quad E[v_g] = 0 \quad E[v_g v_g^T] = R_g \quad (2.129)$$

La matriz  $\mathbf{T}_g$  tendrá en cuenta las constantes de tiempo para los tres giros, esto es,

$$\mathbf{T}_g = \begin{bmatrix} -1/\tau_x & 0 & 0 \\ 0 & -1/\tau_y & 0 \\ 0 & 0 & -1/\tau_z \end{bmatrix} \quad (2.130)$$

Un modelo de error estándar para los acelerómetros consiste de un componente constante aleatorio más ruido blanco,

$$\dot{a} = v_a \quad E[v_a] = 0 \quad E[v_a v_a^T] = R_a \quad (2.131)$$

La matriz  $\mathbf{T}_a$  modela los errores del acelerómetro, esto es,

$$\mathbf{T}_a = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \quad (2.132)$$

Finalmente, el modelo aumentado que considera los nuevos estados tendrá los siguientes estados

$$\mathbf{x} = [p, v, \phi, f_g, \omega_{ib}]^T \quad (2.133)$$

donde  $p$ ,  $v$  y  $\phi$  son los errores en posición, velocidad y rumbo y  $f_g$  y  $\omega_{ib}$  los errores en los acelerómetros y giros en el marco del móvil; las matrices aumentadas quedan entonces,

$$\mathbf{F} = \begin{bmatrix} & 0 & 0 & & \\ & \mathbf{F}_g & \mathbf{C}_b^n & 0 & \\ & & 0 & \mathbf{C}_b^n & \\ 0 & 0 & 0 & \mathbf{T}_a & 0 \\ 0 & 0 & 0 & 0 & \mathbf{T}_g \end{bmatrix} \quad \mathbf{G} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & \mathbf{C}_b^n & 0 & 0 & 0 \\ 0 & 0 & \mathbf{C}_b^n & 0 & 0 \\ 0 & 0 & 0 & \mathbf{I} & 0 \\ 0 & 0 & 0 & 0 & \mathbf{I} \end{bmatrix} \quad (2.134)$$

Hay que tener en cuenta que en esta implementación, las entradas de aceleración se alimentan directamente al modelo de proceso y, por lo tanto, no hay un vector de control. El modelo de proceso  $F$  comprende términos que varían en el tiempo. Por tanto, para determinar la matriz de transición de estado se requieren métodos numéricos. Si es constante en el intervalo de muestreo, entonces la matriz de transición de estado es simplemente  $F(k) = \exp(F\Delta t)$ , donde  $\Delta t$  es el tiempo de muestreo de la unidad inercial. En el caso de los vehículos terrestres, la dinámica es de una frecuencia mucho menor que la frecuencia de muestreo. Se puede considerar que durante el período de muestreo  $F$  permanece constante, por lo tanto

$$F(k) = \exp(F\Delta t) \approx I + F\Delta t \quad (2.135)$$

truncando los términos de orden mayor a uno que se consideran despreciables. Por lo que las predicciones serán lineales:

$$\hat{x}(k|k-1) = F(k)\hat{x}(k-1|k-1) \quad (2.136)$$

La ventaja de este modelo es que es independiente de la dinámica del vehículo. Inicialmente, los sensores inerciales son calibrados y todos los errores eliminados, luego  $\hat{x}(1|0) = \mathbf{0}$  que representa la inicialización del filtro. Por tanto, a partir de la ecuación 2.136, la predicción de estado en el siguiente ciclo también es cero, y así sucesivamente. Por lo tanto, la predicción de estado es siempre cero y no se produce ninguna corrección de los errores de inercia durante el ciclo de predicción. Es decir, la información de posición, velocidad y actitud obtenida del sistema de navegación son simplemente los datos del INS ya que no hay errores de predicción para corregirlos.

Sin embargo, debido a la deriva en la solución INS, hay un crecimiento correspondiente en la incertidumbre en los estados y esto se evalúa a través de la matriz de covarianza predicha,

$$P_{k|k-1} = F(k)P_{k-1|k-1}F^T(k) + Q(k) \quad (2.137)$$

Esta es una matriz de  $9 \times 9$  que representa la incertidumbre en los errores predichos inerciales.  $Q(k)$  es la matriz de ruido de proceso discreta también de dimensión  $9 \times 9$  y se evalúa utilizando

$$Q(k) = \frac{1}{2}[F(k)G(k)Q_c(k)G^T(k)F^T(k) + G(k)Q_c(k)G^T(k)]\Delta t \quad (2.138)$$

donde  $Q_c(k)$  es la incertidumbre de proceso que se define como

$$Q_c(k) = \begin{bmatrix} \delta\tilde{p} & 0 & 0 \\ 0 & \delta\mathbf{f}_b & 0 \\ 0 & 0 & \delta\boldsymbol{\omega}_{ib} \end{bmatrix} \quad (2.139)$$

$\delta\tilde{p}$  es el ruido inyectado en la evaluación del error de posición y su valor depende puramente de la incertidumbre en la evaluación de la

posición a partir de la integración de la velocidad. Los términos de error restantes en esta matriz son los valores de ruido en las lecturas del sensor, es decir, los errores en el marco del móvil, que se pueden obtener de las especificaciones del fabricante o mediante experimentación.  $\mathbf{G}(k)$  es

$$\mathbf{G}(k) = \begin{bmatrix} \mathbf{I}_{3 \times 3} & 0 & 0 \\ 0 & \mathbf{C}_b^n & 0 \\ 0 & 0 & -\mathbf{C}_b^n \end{bmatrix} \quad (2.140)$$

En una estructura de retroalimentación directa, el modelo implementado en el filtro es un modelo de error lineal que representa los errores en los estados del vehículo, generalmente posición, velocidad y actitud. Cuando una observación está disponible, el filtro estima los errores en estos estados. Dado que el modelo es un modelo de error de las ecuaciones inerciales, la observación  $\mathbf{z}(k)$  es el error observado de la solución de navegación inercial y no la observación proporcionada por el GPS. Por lo tanto, si un sistema de ayuda proporciona datos de posición y velocidad, el vector de observación se convierte en,

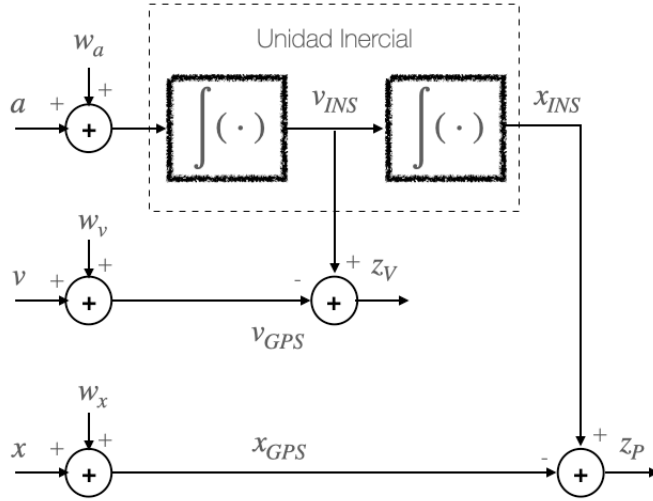
$$\mathbf{z}(k) = \begin{bmatrix} \mathbf{z}_p(k) \\ \mathbf{z}_v(k) \end{bmatrix} = \begin{bmatrix} \mathbf{P}_{inercial}(k) - \mathbf{P}_{GPS}(k) \\ \mathbf{V}_{inercial}(k) - \mathbf{V}_{GPS}(k) \end{bmatrix} \quad (2.141)$$

La Figura 2.35 ilustra la estructura de observación. A la verdadera aceleración, velocidad y posición del vehículo se les agrega ruido para representar las mediciones tomadas por los sensores. La aceleración, medida por el sistema de navegación inercial, se integra dos veces para obtener la velocidad y posición indicadas del vehículo. La información de aceleración se obtiene mediante los acelerómetros y se supone que se ha compensado la aceleración debida a la gravedad.

Definiendo los términos  $\delta\mathbf{P}(k)$  y  $\delta\mathbf{V}(k)$  como los errores de posición y velocidad en los datos inerciales después del proceso de integración,  $\mathbf{P}_T(k)$  y  $\mathbf{V}_T(k)$  la posición y velocidad verdadera y  $v_p(k)$  y  $v_v(k)$  el error en las medidas del GPS, el modelo de observación se convierte en

$$\begin{aligned} \mathbf{z}(k) &= \begin{bmatrix} \mathbf{P}_k^{inercial} - \mathbf{P}_k^{GPS} \\ \mathbf{V}_k^{inercial} - \mathbf{V}_k^{GPS} \end{bmatrix} \\ &= \begin{bmatrix} (\mathbf{P}_T(k) + \delta\mathbf{P}(k)) - (\mathbf{P}_T(k) - v_p(k)) \\ (\mathbf{V}_T(k) + \delta\mathbf{V}(k)) - (\mathbf{V}_T(k) - v_v(k)) \end{bmatrix} \\ &= \begin{bmatrix} \delta\mathbf{P}(k) \\ \delta\mathbf{V}(k) \end{bmatrix} + \begin{bmatrix} v_p(k) \\ v_v(k) \end{bmatrix} \end{aligned} \quad (2.142)$$

La observación es, por tanto, el error entre la posición y velocidad inerciales y la del GPS, y la incertidumbre en esta observación se refleja en el ruido de la observación del GPS. Esto ofrece otro beneficio en



**Figura 2.35:** Ilustración de cómo se obtienen las mediciones de observación  $\mathbf{z}_p(k)$  y  $\mathbf{z}_v(k)$  mediante la información inercial y el GPS.

la implementación de la realimentación directa y tiene que ver con el ajuste del filtro. El ruido de la observación es el ruido del GPS. Por lo tanto, una vez que se fija una unidad inercial y un modelo de proceso, la matriz de ruido de proceso  $\mathbf{Q}(k)$  también se fija, y el ajuste del filtro se basa únicamente en la matriz de ruido de observación  $\mathbf{R}(k)$ . La estimación de los estados de error en el momento  $k$ , dadas todas las observaciones hasta el momento  $k$ , se puede evaluar utilizando las ecuaciones de actualización del filtro de Kalman estándar:

$$\hat{\mathbf{x}}_{k|k} = \hat{\mathbf{x}}_{k|k-1} + \mathbf{W}_k \mathbf{v}_k \quad (2.143)$$

donde  $\mathbf{W}_k = \mathbf{P}_{k|k-1} \mathbf{H}_k^T \mathbf{S}_k^{-1}$  es la matriz de ganancia del filtro de Kalman y  $\mathbf{v}_k = \mathbf{z}_k - \mathbf{H}_k \hat{\mathbf{x}}_{k|k-1}$  es el vector de innovación.  $\mathbf{S}_k$  es la covarianza de la innovación que se obtiene como

$$\mathbf{S}_k = \mathbf{H}_k \mathbf{P}_{k|k-1} \mathbf{H}_k^T + \mathbf{R}_k \quad (2.144)$$

Finalmente, la matriz de covarianza se obtiene a partir de,

$$\mathbf{P}_{k|k} = \mathbf{P}_{k|k-1} - \mathbf{W}_k \mathbf{S}_k \mathbf{W}_k^T \quad (2.145)$$

Sin embargo con esta aproximación, dado que las predicciones de los estados del error  $\hat{\mathbf{x}}_{k|k-1}$  son siempre cero, el estimado es directamente,

$$\hat{\mathbf{x}}_{k|k} = \mathbf{W}_k \mathbf{z}_k \quad (2.146)$$

Esto significa que la actualización es simplemente la suma pesada de las observaciones. El modelo de observación  $\mathbf{H}_k$  es

$$\mathbf{H}_k = \begin{bmatrix} \mathbf{I}_{3 \times 3} & 0 & 0 \\ 0 & \mathbf{I}_{3 \times 3} & 0 \end{bmatrix} \quad (2.147)$$

Los errores actualizados de posición y velocidad ahora se pueden usar para corregir la posición y velocidad del INS,

$$\begin{aligned}\mathbf{p}_{k|k}^{\text{inercial}} &= \mathbf{p}_{k|k-1}^{\text{inercial}} - \delta\mathbf{P}(k) \\ \mathbf{v}_{k|k}^{\text{inercial}} &= \mathbf{v}_{k|k-1}^{\text{inercial}} - \delta\mathbf{V}(k)\end{aligned}\quad (2.148)$$

Una vez que se obtienen los errores de actitud, la matriz de cosenos directores se actualiza de la siguiente manera

$$\mathbf{C}_b^n(k|k) = [\mathbf{I}_{3 \times 3} - [\delta\psi \times]]^{-1} \mathbf{C}_b^n(k|k-1) \quad (2.149)$$

Dado que el término entre corchetes será ortogonal, la inversa será simplemente la traspuesta ( $\mathbf{A}^{-1} = \mathbf{A}^T$ ) y como además también es antisimétrica ( $\mathbf{A}^T = -\mathbf{A}$ ), la ecuación 2.150 queda

$$\mathbf{C}_b^n(k|k) = [\mathbf{I}_{3 \times 3} + [\delta\psi \times]] \mathbf{C}_b^n(k|k-1) \quad (2.150)$$

donde como dijimos,

$$[\delta\psi \times] = \begin{bmatrix} 0 & -\delta\psi_D & \delta\psi_E \\ \delta\psi_D & 0 & -\delta\psi_N \\ -\delta\psi_E & \delta\psi_N & 0 \end{bmatrix} \quad (2.151)$$

Hay que tener en cuenta que  $[\delta\psi \times]$  están en el marco de navegación pero se usa para corregir la matriz  $\mathbf{C}_b^n$  cuyos elementos están definidos en el marco del cuerpo. Se supone que los errores de desalineación son pequeños y, por tanto, los errores asociados sobre el marco de navegación son iguales a los del marco del cuerpo. Cuando se encuentran grandes desalineaciones, los supuestos lineales mantenidos no son válidos. Existe abundante bibliografía que trata esto y otros problemas ([11] y [8] entre muchas) y que quedan fuera del tratamiento en este curso, pero solo deben considerarse en el caso que esto sea necesario o cuando se utilicen unidades inerciales de muy baja calidad.

#### 2.6.2.5 Asuntos relacionados con la implementación en tiempo real

**Consistencia del Filtro de Kalman.** Cuando se implementa un filtro de Kalman no hay forma directa de determinar si está calculando estimaciones correctas. La única información disponible del mundo real es la observación y, por lo tanto, la única forma de medida para determinar el comportamiento del filtro es la diferencia entre la observación y la observación predicha, es decir, la innovación. La innovación tiene la propiedad de que debe ser tanto sin sesgo como blanca, y tener covarianza  $\mathbf{S}_k$  si el filtro está funcionando correctamente. Verificar estas propiedades constituyen las pruebas para analizar el comportamiento del filtro. No las presentaremos en este curso

pero se puede consultar [4]. Cuando se obtiene una observación, se construye la siguiente cantidad escalar

$$\gamma = \mathbf{v}^T \mathbf{S}^{-1} \mathbf{v} \quad (2.152)$$

Si el valor de  $\gamma$  es menor que un umbral predefinido, entonces se acepta la observación. Esto permite un medio de detectar cualquier falla dentro de la observación. El valor umbral se obtiene de las tablas estándar de  $\chi^2$  y se elige en función del nivel de confianza requerido. Por ejemplo, un nivel de confianza del 95 %, y para un vector de estado que incluye tres estados de posición y tres de velocidad, entonces  $\gamma = 12,6$

**Multicamino.** Cuando las señales GNSS sufren errores de multicamino, genera errores de alta frecuencia. El multicamino genera retardo mayores en las señales que afectan tanto a los receptores diferenciales como la fase y por lo tanto a los receptores RTK. Otra falla de alta frecuencia, aunque ocurre con menos frecuencia y con menos efecto, es cuando el receptor utiliza un conjunto diferente de satélites para determinar la posición. La precisión de la corrección depende de la geometría de los satélites observados. Los cambios en la configuración del satélite debido a bloqueos de la vista del satélite, a su vez, alterarán la solución resultante. Ambas formas de fallas de alta frecuencia provocan saltos abruptos en la posición y las correcciones de velocidad obtenidas por el receptor GNSS.

Estas fallas dependen fuertemente del entorno. Un área abierta tendrá menos probabilidades de producir errores de multicamino que otra con construcciones por ejemplo. En consecuencia, el ajuste del filtro que fusiona la unidad inercial y los datos GNSS depende del entorno. El método más común para rechazar los errores de multicamino es el receptor mismo, que internamente pueda distinguir entre la señal verdadera y la señal reflejada<sup>1</sup>.

Otra estrategia es el uso de antenas especiales que minimizan el efecto de la trayectoria múltiple. Las antenas de anillo de choque, por ejemplo, pueden reducir la ganancia de la antena en elevaciones bajas al eliminar las señales de ángulos de elevación bajos. Este diseño contiene una serie de valles circulares concéntricos que tienen un poco más de un cuarto de longitud de onda de profundidad (fig. 2.36). Estas antenas son voluminosas (la de la figura mide 40cm de diámetro, 20cm de alto y pesa 7.6Kg) por lo que no se suelen usar en aplicaciones móviles. Una alternativa es elevar la antena con un plano de tierra que evite las reflexiones en el suelo.

A pesar de estos esfuerzos los errores multicamino no se pueden rechazar por completo, e incluso con la mejora constante de la tecnología GNSS, las fallas de alta frecuencia por multicamino siempre

<sup>1</sup> El receptor *u-blox 8* por ejemplo confía menos en señales de baja potencia para preservar la calidad de la solución de posición en entornos de señal deficiente. Esta característica puede resultar en un rendimiento pobre en situaciones en las que las señales se atenúan por otro motivo, por ejemplo, debido a la ubicación de la antena.





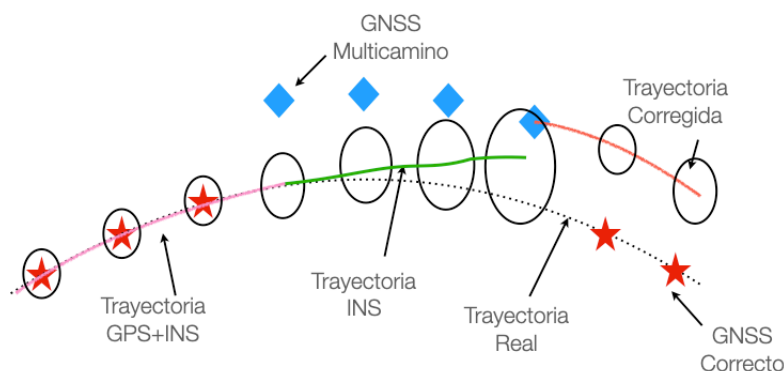
**Figura 2.36:** Antena de anillo de choque comercial.

deben tenerse en cuenta en el desarrollo del sistema de navegación. Se puede utilizar la ecuación 2.152 para mantener la consistencia del filtro teniendo en cuenta que  $\gamma$  debe ser una muestra de una distribución  $\chi^2$ . Esto se puede utilizar para determinar si el proceso o los modelos de observación son válidos o si alguna observación es falsa. Por lo tanto, puede potencialmente determinar si se han producido errores de multicamino.

Debido a la geometría del satélite, la posición GNSS en el plano vertical es significativamente menos precisa que en el plano horizontal. En consecuencia, la corrección en Norte y Este puede estar bien dentro de la región de validación, mientras que la de Abajo puede excederla y forzar el resultado de la función de activación por encima del umbral si se usaran los mismos valores de ruido para todos los términos en la matriz de ruido de observación  $\mathbf{R}_k$ . Sin embargo, si estos términos de ruido se tienen en cuenta tomando los valores del receptor GNSS directamente, la puerta de validación detectará correctamente los errores de trayectos múltiples.

De manera similar, los cambios en la geometría del satélite hacen que varíe la Dilución de Precisión (DOP) como ya hemos visto. Los cambios en la geometría de los satélites ocurren cuando parte del cielo es invisible para la antena del receptor debido a obstáculos que bloquean las señales GNSS. Luego, el receptor debe obtener un nuevo conjunto de satélites y en consecuencia, un cambio en la DOP afectará a la solución GNSS provocando fallas de alta frecuencia. Estas fallas se pueden detectar pero los cambios no son tan grandes y generalmente no se detectan.

**Sintonía del Filtro.** Hay dos etapas en el flujo del filtro. La etapa de predicción donde los errores de inercia predichos son siempre cero y la incertidumbre crece con el tiempo; y la etapa de estimación donde las estimaciones de los errores de inercia se obtienen mediante una suma ponderada de las observaciones y la incertidumbre se reduce.



**Figura 2.37:** Errores de alta frecuencia. Las medidas correctas del GNSS se muestran como estrellas y las debidas al multicamino como rombos. Cuando una medida de GNSS errónea entra dentro de la región de activación representada por  $\gamma$ .

Si no se obtiene ninguna observación durante un período prolongado de tiempo, o de manera equivalente, si se rechazan los puntos GNSS debido a errores, el filtro realizará un ciclo continuo en modo de predicción y no se realizarán correcciones a la solución de navegación inercial. Cuanto mayor sea la duración sin corrección, mayor será la incertidumbre en la solución de navegación. Si esto se mantiene, la observación multicamino puede pasar la prueba de la cota de validación, ya que la incertidumbre de la navegación inercial es grande. Esto se muestra en la Figura 2.37.

Al igual que con cualquier implementación de filtro de Kalman, el ajuste radica en la elección de valores para las matrices de covarianza de proceso  $\mathbf{Q}_k$  y de observación  $\mathbf{R}_k$ . Por ejemplo, una  $\mathbf{Q}_k$  grande implicará un sistema inercial inexacto. Durante la etapa de predicción, la incertidumbre en los datos inerciales aumentará de acuerdo con la magnitud de  $\mathbf{Q}_k$ . Cuando se produce una corrección GNSS, existe una mayor posibilidad de que la unidad inercial se corrija utilizando la primera corrección GNSS disponible, independientemente de la precisión de esta corrección. Asimismo, valores pequeños en  $\mathbf{R}_k$  implicarán correcciones GNSS precisas que pueden plantear un problema cuando la corrección es errónea y se fusiona con sensores inerciales de baja precisión.

Por lo tanto, la sintonización se convierte en un ajuste delicado de las matrices  $\mathbf{Q}_k$  y  $\mathbf{R}_k$  junto con el empleo de la función de activación  $\gamma$  para rechazar las fallas de alta frecuencia del GNSS. Las variaciones a lo largo de la diagonal de  $\mathbf{R}_k$  se determinan simplemente obteniendo los valores de GDOP del receptor y asumiendo que no hay correlación entre los puntos fijos de los ejes de navegación (que es como se genera el GDOP).

La determinación de los valores de  $\mathbf{Q}_k$  depende del nivel de ruido de los sensores, que se puede obtener del fabricante o mediante experimentación. Los principales parámetros de ajuste que deben abor-

darse son las variaciones de velocidad y actitud. Una gran variación en los términos de velocidad implicará una mayor dependencia de los valores fijos de velocidad del GNSS. Además, cuanto mayor sea la precisión de las estimaciones de velocidad, mayor será la amortiguación de los errores de actitud. Si no hay ajustes de actitud, solo la información de velocidad puede contener la deriva en la actitud y la rapidez y precisión con la que esto puede suceder depende de la variación de la actitud.

**Algoritmo.** Hay dos puntos de preocupación en una implementación en tiempo real de un sistema de navegación GNSS asistido por la unidad inercial: potencia de procesamiento y latencia. La latencia de los datos es extremadamente importante, especialmente con el uso de sensores GNSS, donde es común encontrar latencias de posición del orden de decenas de milisegundos y en la velocidad que supere el segundo. En aplicaciones móviles, la potencia de procesamiento internamente en un sensor GNSS se usa para controlar los correladores para que se fijen en los satélites, determinar las efemérides de los satélites y luego calcular la posición y la velocidad del receptor con respecto a algún marco de coordenadas. La complejidad de estos cálculos aumenta con el número de satélites utilizados para determinar la solución.

Las soluciones de pseudodistancia estándar requieren el menor procesamiento computacional. La tecnología RTK requiere la mayor potencia de procesamiento. Además, si se requiere información de actitud, esto también requiere un procesamiento adicional. Para superar este problema de latencia, se debe considerar cuidadosamente la implementación en tiempo real, y cuando la latencia es grande, puede ser necesario un rediseño completo de la estructura del filtro. Cuando la latencia de la solución es baja, la información de posición se puede propagar utilizando datos de velocidad y un modelo de velocidad constante. Esto es suficiente para vehículos de baja velocidad. Los problemas surgen cuando la latencia de la solución GNSS es alta y la dinámica del vehículo es rápida. Una solución de posición GNSS con una latencia de 50ms tendrá un error de 0,8m para un vehículo que viaja a 60km/h. Si el vehículo se mueve con velocidad constante y se mueve en línea recta, entonces la posición puede propagarse simplemente hacia adelante. Sin embargo, cualquier desviación de una línea recta conducirá a estimaciones incorrectas cuando la fijación del GNSS se fusiona con el estado inercial actual. Lo que se requiere es almacenar los datos inerciales y procesar las estimaciones en el momento en que debería haber ocurrido la corrección GNSS, y luego propagar la solución inercial a través de la acumulación de datos. Todo está bien si tanto los datos GNSS de posición como de velocidad tienen latencias iguales.

Sin embargo, si existe el requisito de que el receptor GNSS obtenga la velocidad y la posición con mediciones independientes, entonces

se requieren métodos alternativos. Por ejemplo, en los sistemas RTK se usa el Doppler para obtener la velocidad, mientras que la determinación de la ambigüedad de fase se usa para medir la posición. En tales sistemas, la latencia de los datos de velocidad a veces puede ser superior a 1 s y, por lo tanto, la determinación de la posición y la velocidad se produce en diferentes puntos en el tiempo. Para superar esto, los fabricantes de GNSS propagan la velocidad a través de los datos de posición y, por lo tanto, tanto la salida de posición como la de velocidad ocurren al mismo tiempo con las mismas latencias. Sin embargo, tal enfoque produce una correlación entre los datos de posición y velocidad, lo que no es ideal para los sistemas de navegación.