

RADIOLOCALIZACIÓN Y RADARES

FAVIO MASSON



Curso de Pregrado

2021 – Version 0.1



Co-funded by the
Erasmus+ Programme
of the European Union



Favio Masson: *Radiolocalización y Radares*, Curso de Pregrado, © 2021

ÍNDICE GENERAL

I POSICIONAMIENTO Y SEGUIMIENTO: CONCEPTOS BÁSICOS 1

1	POSICIONAMIENTO: CONCEPTOS BÁSICOS	3
1.1	Definiciones y Repaso Histórico	4
1.1.1	Relación Tiempo y Espacio	6
1.2	Breve historia de los sistemas de localización	9
1.2.1	Telefunken Kompass	9
1.2.2	SONNE/CONSOL	9
1.2.3	GEE	11
1.2.4	LORAN-A	12
1.2.5	GNSS	14
1.3	Taxonomía de los sistemas de posicionamiento	14
1.3.1	Técnicas	14
1.4	Repaso de Radio Frecuencia	17
1.4.1	Métodos de Propagación Básicos	18
1.4.2	Modelos de Propagación	21
1.4.3	Modelo de Pequeña escala	29
1.4.4	Comentarios finales	34
1.5	Multilateración Circular	35
1.5.1	Uso de Tiempo de Arribo	37
1.5.2	Uso del Nivel de Potencia	41
1.6	Precisión	44
1.6.1	DOP: <i>Dilution of Precision</i>	44
1.6.2	CEP: <i>Circular Error Probability</i>	45
1.6.3	CRLB: <i>Cramer Rao Lower Bound</i>	45
1.6.4	Por Ensayo	46
1.6.5	Por Simulación	47
1.7	Multilateración Hiperbólica	50
1.7.1	Filtro de Kalman para TDoA	55
1.7.2	Método de Lavenberg-Marquardt para TDoA	55
1.8	Multilateración Angular	57
1.8.1	Formas de resolución de posición usando AoA	58
1.9	Medición de Tiempos y ángulos	65
1.9.1	Medidas de tiempo entre transmisor y receptor	67
1.9.2	Estimación de RSS	68
1.9.3	Estimación de ángulo de arribo	68
1.10	Outliers: Valor Atípico	70
1.10.1	Análisis de los residuos	71
1.11	Fingerprinting	74
1.11.1	Obtención de la posición (Pattern Matching)	77

1.12	Ejercicio Final	83
II	POSICIONAMIENTO Y SEGUIMIENTO CON INTEGRACIÓN DE SENSORES	87
2	INTEGRACIÓN DE SENSORES	89
2.1	Sistemas de Navegación Satelital (GNSS)	89
2.1.1	Principios generales de la navegación satelital	91
2.1.2	Impacto de los Errores de Medición	95
2.1.3	El impacto de la geometría	97
2.1.4	Modos de posicionamiento	101
2.1.5	Sistemas de Referencia	108
2.1.6	Estructura de la señal de RF	113
2.1.7	Modulación	113
2.1.8	Aumentación	115
2.1.9	Sistemas GNSS actuales	119
2.2	Estimación Probabilística: Filtro de Kalman	122
2.2.1	Sistema Lineal	123
2.2.2	Filtro de Kalman Lineal	124
2.2.3	Sistema no Lineal	127
2.2.4	Filtro Extendido de Kalman	128
2.3	Navegación de un móvil terrestre	129
2.3.1	Modelo de un vehículo terrestre	130
2.4	Integración de sensores	132
2.4.1	Sensores internos o propioceptivos	132
2.4.2	Sensores externos o exteroceptivos	133
2.4.3	Mapas y Rasgos	133
2.4.4	Integración GPS con otros sensores externos	137
2.5	SLAM: Navegación y Construcción Simultánea del Mapa	146
2.5.1	Significado de las correlaciones	150
2.5.2	Asociación de datos	153
2.6	Integración de GPS con Unidades Inerciales	157
2.6.1	Unidades Inerciales	157
2.6.2	Integración	166
III	POSICIONAMIENTO Y SEGUIMIENTO EN REDES INALÁMBRICAS	187
3	SEGUIMIENTO EN REDES INALÁMBRICAS	189
3.1	Redes Celulares [5]	189
3.1.1	Identificación de la Celda (Cell-ID)	190
3.1.2	Localización usando TDOA	191
3.1.3	Localización usando Ángulo de Arribo	192
3.1.4	Patrones de Potencia de señal recibida	192
3.1.5	Posicionamiento en LTE	193
3.2	Redes Inalámbricas de corto alcance y RFID	194
3.2.1	WLAN/Wi-Fi	194
3.2.2	WPAN/Bluetooth	199

3.2.3	RFID	202
IV	RADARES	205
4	RADARES	207
4.1	Introducción [7]	207
4.1.1	Tipos de radares	207
4.1.2	Medidas de Rango	209
4.1.3	Ambigüedad de rango	209
4.1.4	Ventana de Procesamiento	211
4.1.5	Medición de velocidad: Doppler	212
4.1.6	Diagrama en Bloque de un Radar	213
4.2	Ecuación de Alcance de Radar	214
4.2.1	Formulación de la energía de la señal E_s	216
4.2.2	Formulación de la energía del ruido E_N	221
4.2.3	Ejemplo de Alcance de Radar	222
4.2.4	Rango de detección	224
4.2.5	Ecuación del rango de búsqueda de un radar	224
4.2.6	Ejemplo de rango de búsqueda	226
V	APÉNDICE	229
A	APPENDIX TEST	231
	BIBLIOGRAFÍA	233

Parte I

POSICIONAMIENTO Y SEGUIMIENTO: CONCEPTOS BÁSICOS

Este curso tiene como objetivo introducir tanto los conceptos básicos como las técnicas actuales de posicionamiento basados en señales de radiofrecuencia, su precisión, su disponibilidad y sus limitaciones.

POSICIONAMIENTO: CONCEPTOS BÁSICOS

A fines de 1901 Marconi transmitió el primer código Morse a través del Atlántico y un año después aparece la primera patente para detectar la dirección como una ayuda a la navegación. Mucho cambió en el camino desde las comunicaciones hasta la utilidad y uso de los sistemas de localización (rango, disponibilidad, etc.). Pero hay un conjunto de elementos que se mantienen constante como los principios de propagación de las ondas electromagnéticas y los geométricos. El uso de los sistemas de posicionamientos fue avanzando desde la ayuda a la navegación de barcos y aviones a los vehículos terrestres y últimamente la aparición de los teléfonos inteligentes y las aplicaciones que usan intensivamente la posición de las personas. Y más allá de los servicios basados en la localización y orientados al usuario (redes sociales, geo-fencing, etc.), hay algunas aplicaciones en esos teléfonos inteligentes que son inherentes a su funcionamiento, como el cobro basado en la ubicación, llamadas de emergencia, manejo de *hand off*, etc.

En este contexto, la tecnología de posicionamiento se ha convertido en un complemento importante de la vida moderna. Si bien en el pasado, la tecnología estaba restringida a grandes aplicaciones profesionales (navegación de barcos y aeronaves) en la actualidad la disponibilidad de información de posicionamiento se ha convertido en un elemento importante en de las actividades diarias de la gente común. Ya sea navegar en un automóvil, encontrar un restaurante cercano, registrar la ubicación de una foto tomada en un teléfono inteligente o rastrear su recorrido diario en su reloj inteligente. Si bien el sistema de posicionamiento global (GPS) se desarrolló originalmente para fines militares, la tecnología se ha hecho accesible para uso civil en el siglo XXI. Con el desarrollo de receptores GPS de un solo chip, el posicionamiento se ha convertido en importante para una amplia gama de campos como la industria, la agricultura, la defensa nacional, la aeronáutica y la astronáutica, la exploración espacial y los diversos aspectos de la sociedad moderna.

Los algoritmos matemáticos sirven para extraer una posición a partir de datos provistos por sensores. Algunos de ellos en forma individual o en otros casos fusionando varios tipos. En todos ellos hay que lidiar con modelos y medidas imprecisas. En general, tienen que ver con la obtención de un rango, un ángulo o ambos, como se muestra en la figura 1.1. Esas medidas, que pueden provenir de señales de RF o láser o video, tienen que ver con las distancias posiciones relativas de puntos con posición conocida y el punto que se desea localizar.

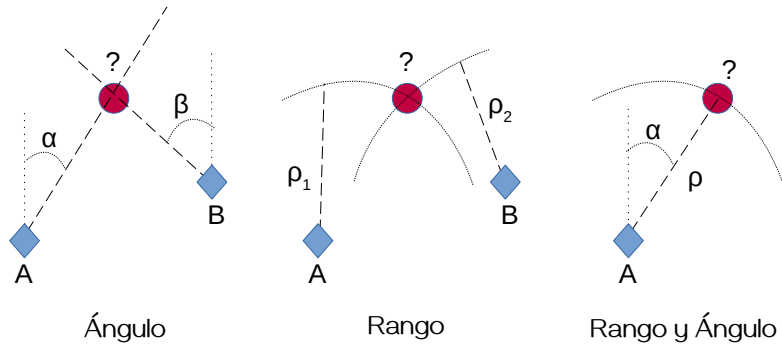


Figura 1.1: Los algoritmos de localización en su forma general consisten en obtener información de ángulo, rango o ambas entre puntos con posición conocida y puntos que se desea localizar

La pregunta que vamos a tratar de responder en este curso es cuáles son los principios fundamentales y los algoritmos básicos para el posicionamiento; antes que en los aspectos científicos del problema de posicionamiento. Aunque, tomaremos estos últimos para entender la profundidad de los problemas.

Este capítulo presenta los algoritmos y conceptos básicos de posicionamiento.

1. **Repaso Histórico** y un conjunto de definiciones para hablar en el mismo idioma
2. **Repaso de RF.** La información provista por los canales de comunicación permiten desarrollar estrategias de posicionamiento pero antes hay que entender algunos aspectos básicos.
3. **Multilateración Circular** o en rango.
4. **Precisión.** Presentación breve de conceptos y formas de conocer la precisión de un método de posicionamiento.
5. **Multilateración Hiperbólica** o de rango diferencial.
6. **Multilateración Angular.**
7. **Medición de tiempos** Complejidades y algoritmos.
8. **Fingerprinting** o posicionamiento basado en firmas.

1.1 DEFINICIONES Y REPASO HISTÓRICO

Definiciones de palabras que usaremos en este curso

LOCALIZACIÓN O POSICIÓN O UBICACIÓN: estos términos para nosotros representan lo mismo; un lugar en el espacio tridimensional o bidimensional de un objeto o grupo de ellos y pueden ser físicas o simbólicas. Las **ubicaciones físicas** se representan en un sistema de referencia fijo; por ejemplo en *Latitud y Longitud* del sistema geodésico mundial o *Northing y Easting* que son distancias hacia el Norte de un parche en la proyección de Mercator. La *ubicación lógica* informa de manera abstracta la posición de interés. Por ejemplo *el segundo piso*. Es decir, usando datos de la infraestructura o arquitectura.

POSICIONAMIENTO: es el proceso en el que se fija una ubicación que puede ser física o simbólica.

NAVEGACIÓN: determinar la pose de uno para lograr llegar a un destino

SEGUIMIENTO: localizar uno o varios objetos durante un periodo de tiempo

LOCAL: posicionamiento en un área específica y limitada (un edificio, una ciudad, etc.)

GLOBAL: una posición en la tierra (por ejemplo LAT-LONG, aunque no es la única)

PARAMÉTRICO: utiliza medidas de distancia o pseudo distancia para determinar la posición

NO PARAMÉTRICO: no implica resolver ecuaciones sino juzgar si un elemento está en la vecindad o en tal habitación

ESTACIÓN MÓVIL (EM): es un dispositivo que se puede mover con instrumentos para poder estimar su posición

ESTACIÓN ANCLA O DE REFERENCIA (ER): es un equipo móvil o fijo que sirve a la EM para su ubicación. Sus coordenadas son conocidas.

DETERMINACIÓN: es el cálculo de posición, velocidad. Si es usando radio se llama radio determinación

LINE-OF-POSITION (LDP): (línea de posición) es el conjunto de puntos en los que se puede ubicar la EM. Las diferentes técnicas de posicionamiento dan diferentes LOP: lineal, circular o hiperbólica. La posición estimada estará dada por la intersección de dos o mas LOP.

EXACTITUD: es la distancia euclídea entre la posición real y la estimada, esto es, si (x_0, y_0) es la posición real en dos dimensiones y (\hat{x}, \hat{y}) la estimación de esa posición

$$\text{error} = \sqrt{(x_0 - \hat{x})^2 + (y_0 - \hat{y})^2} \quad (1.1)$$

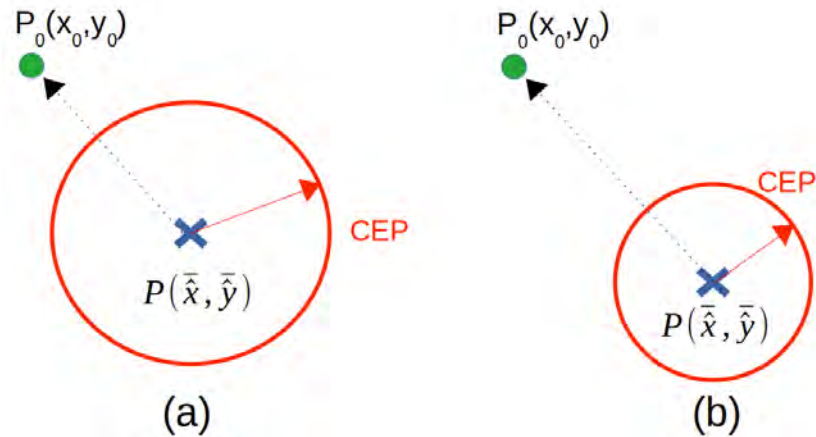


Figura 1.2: Comparación entre exactitud y precisión de dos algoritmos de localización. La estimación (a) es menos precisa y mas exacta que la (b).

PRECISIÓN: es la distribución de probabilidad de la exactitud de un conjunto de estimaciones de posición. Si el objetivo está quieto en un marco de referencia fijo, la precisión puede ser caracterizada por el *error circular probable* (CEP en inglés) que es el radio del círculo dentro del cual están el 50 % de las estimaciones de la posición. El CEP se base en que las estimaciones tienen una distribución normal.

Para comparar el concepto de exactitud y precisión, consideremos M estimaciones $\mathbf{P} = (\hat{x}_i, \hat{y}_i)$ con $i = 1, \dots, M$ y (\bar{x}, \bar{y}) es la media aritmética de esas M estimaciones. Consideremos también a $\mathbf{P}_0 = (x_0, y_0)$ el punto donde se encuentra realmente el móvil. El CEP estará centrado en (\bar{x}, \bar{y}) y el radio del círculo será el que contenga al 50 % de las M estimaciones. A medida que el CEP es de mayor radio, menor será la precisión. La exactitud por su lado, es la distancia que separa a la posición real de las estimaciones. Si se observa la figura 1.2, se puede ver que un método de posicionamiento puede ser a la vez preciso y tener baja exactitud.

1.1.1 Relación Tiempo y Espacio

El tiempo y la posición han evolucionado en forma diferente. En general asociadas, pero diferentes en los resultados. En el inicio de la humanidad, el tiempo se sentía en días, aquellos necesarios para llegar de un punto a otro.

Con la organización del comercio y la agricultura se necesitó de un calendario. En un primer momento, cuando se necesitó de organizar la sociedad en años y meses, aparece el calendario solar. Luego se dio la necesidad de dividir el día, y para esto también se utilizó el sol (sundial). Pero esto era impreciso y variable en el año. De noche había problemas.

Los monjes crearon los relojes para organizar la vida religiosa durante todo el día: los campanarios. La noche seguía siendo otro mundo. Y esos “relojes” se desarrollaron para organizar la noche también. Estos se manejaban en forma centralizada y de a poco los mecanismos se hicieron mas accesibles para que las personas tuvieran su propio horario. Y el perfeccionamiento aparece con la detección de la Longitud para viajar por los mares en un mundo que se expandía a partir de Europa.

En esta expansión del mundo occidental hacia todo el mundo, se plantea la necesidad de sincronizar la diferencia de horario en los distintos lugares por la rotación de la tierra (para las líneas de trenes o la administración de las colonias). Esto motiva la creación de los husos horarios. Luego, la revolución industrial planteó que el trabajo en lugar de ser por objetivos, pasó a ser por cumplimiento de horarios: Tiempo es Dinero. La globalización finalmente implica la necesidad de un tiempo sincronizado con el mundo: para las finanzas, el entretenimiento, los viajes o el trabajo mismo.

1.1.1.1 *Espacio*

La percepción del espacio también fue cambiando a lo largo de los siglos. A medida que el tiempo se sincronizaba en todo el mundo, también se necesitaba una representación más precisa del mundo en términos de mapas, rutas, etc. Y el tiempo está involucrado en esto, por ejemplo es muy importante poder evaluar el tiempo necesario para un viaje determinado, ya sea de personas o de mercancías.

Si intentamos hacer una comparación entre la evolución de la medición del tiempo y la evolución de los sistemas de posicionamiento, es ciertamente posible decir que el posicionamiento está hoy en día en la situación en que se encontraba el tiempo hace más de 150 años con la llegada de los relojes portátiles. Esta fue la hazaña técnica que permitió la apropiación del tiempo por todo el mundo. El equivalente en posicionamiento está ahora disponible con sistemas de posicionamiento basados en satélites (gracias al GPS). Algunas características son similares entre los primeros relojes portátiles y los receptores básicos de GPS de hoy en día.

Hay otro logro técnico fundamental para la difusión de los dispositivos portátiles de posicionamiento y su incorporación a la vida de todos: las telecomunicaciones. Cuando alguien utiliza la hora leída en un reloj de pulsera, se comparte automáticamente con los demás porque basta con la singularidad del referencial común. Este no es



Figura 1.3: El posicionamiento de un equipo (en este caso un montacarga) que se mueve en ambientes exteriores e interiores es un problema complejo de solucionar (figura de maquinac.com).

en absoluto el caso de la posición: incluso considerando un referencial geográfico compartido, la posición es una especificidad de una persona.

En el ámbito de esta evolución, se puede considerar que el posicionamiento podría ser rentable en dominios como la ubicuidad, o en otras palabras, el descubrimiento automático del entorno de cualquier persona, o grupos. Las telecomunicaciones necesarias están disponibles hoy en día, pero no el posicionamiento, especialmente en ambientes interiores.

El impacto de la evolución del posicionamiento que funcione como el tiempo, tendría un impacto tremendo en la vida de las personas y la sociedad. Pero hay varios problemas, además del posicionamiento, que hay que solucionar. Como.

Y porqué estamos tan lejos o tan cerca de esa "solución" tiene que ver con muchas cosas. Pero pensemos en la situación mostrada en la figura 1.3. Este montacarga, lleva material en un ambiente exterior. No digamos que quiero que sea autónomo. Solo que intento hacer un sistema que evite accidentes con otros vehículos. La posición de este vehículo podría razonablemente armarla con GPS, modelos de movimiento y medición de dinámicas del vehículo.

Ahora, ese vehículo ingresa al galpón donde están los materiales. Ya no tengo GPS como medida "directa de posición". Tengo que integrar la información de movimiento. Eso no tiene mucho futuro porque también integro el error. Necesito otro sistema de posicionamiento.

to. Por ejemplo basado en mojones de RF. Puedo lograr posiciones bastante precisas. Pero... no tengo un único reloj que funcione debajo del galpón, o del agua, y en el mar, en el aire o en la tierra. Ahí un ejemplo de la inmadurez.

1.2 BREVE HISTORIA DE LOS SISTEMAS DE LOCALIZACIÓN

Los primeros sistemas de posicionamiento usando señales de RF se basaban en encontrar la dirección desde la que se emitían señales conocidas desde puntos conocidos. Estos sistemas primitivos detectaban el ángulo de arribo de señales emitidas por estaciones con posición conocida en forma omnidireccional. El móvil barría las direcciones de arribo a una velocidad controlada para detectar la señal. Luego se revirtió el sistema y fueron las antenas de las estaciones de referencia giraban y la antena del móvil es omnidireccional.

1.2.1 *Telefunken Kompass*

El sistema Telefunken Kompass fue desarrollado en Alemania en 1908 para guiar los Zeppelin que bombardeaban Gran Bretaña en la 1ra guerra mundial. Cada estación base era un poste de 20m los que se colgaban 32 cables de 60m de largo (figura 1.4). Para transmitir una llave conectaba dos de los cables para formar un dipolo. El sistema primero transmitía en modo omnidireccional su identificación y luego hacía un barrido por cada par de conductores transmitiendo un punto en código morse. En 30 segundos completaba una vuelta. El operador sobre el Zeppelin cuando escuchaba la identificación, esperaba un silencio para detectar la dirección y esta coincidía con un nulo del patrón de radiación del dipolo (figura 1.5). Con dos estaciones base era suficiente para determinar la posición.

1.2.2 *SONNE/CONSOL*

En 1940 los alemanes desarrollaron en la segunda guerra mundial el sistema SONNE que consistía en tres antenas en las bases transmitiendo a 300KHz. Las antenas externas transmitían a 1/4 de la potencia y desfasadas de la central generando un patrón de radiación complejo como se puede apreciar en la figura 1.6.

El sistema operaba en ciclos de 40 segundos. En los primeros 6 segundos la antena central transmitía en forma omnidireccional su identificación en código Morse. Luego de 2 segundos de silencio el transmisor se conecta a las tres antenas formando el patrón de la figura 1.6 y transmite un punto (1/8 segs.) y una raya (3/8 segs.) en forma alternada. Trabajando sobre los desfases, el patrón se gira en sentidos contrarios en cada antena externa. El patrón superpuesto resultante genera zonas de *equiseñal* o donde la señal transmitida es continua por

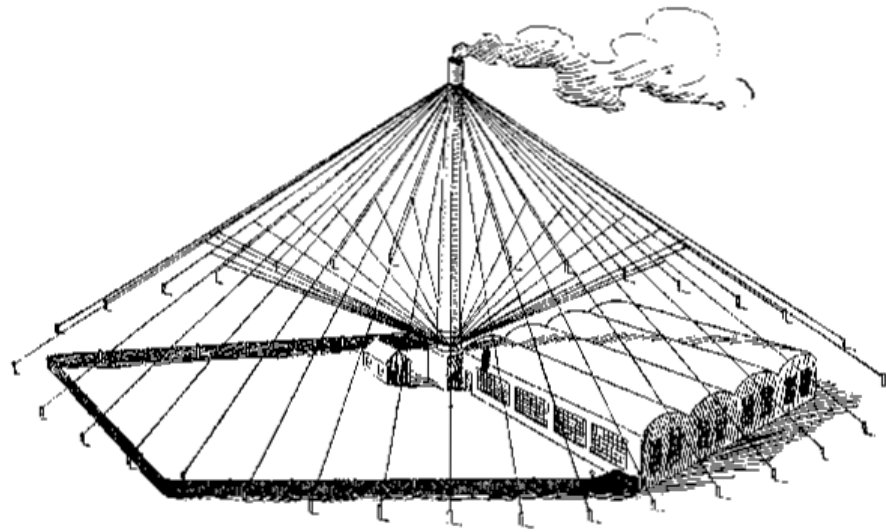


Figura 1.4: Estación de referencia del sistema Telefunken Kompass (figura de earlyradiohistory.us/1918zep.htm).

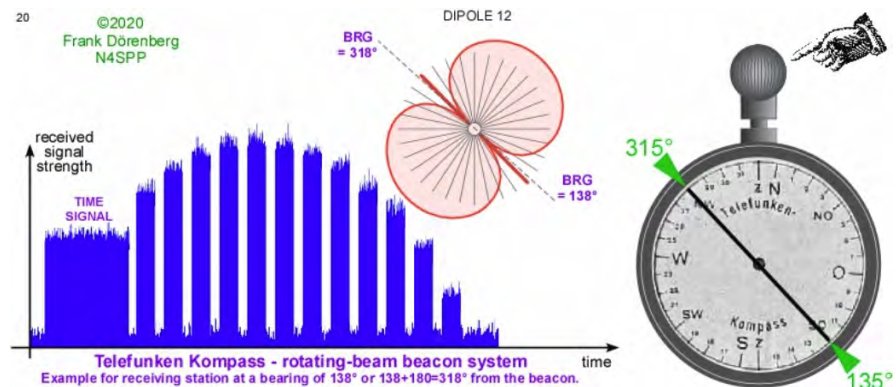


Figura 1.5: Señales del sistema Telefunken Kompass. Una simulación de operación se puede ver en <https://www.nonstopsystems.com/radio/hellschreiber-modes-other-hell-RadNav.htm>.

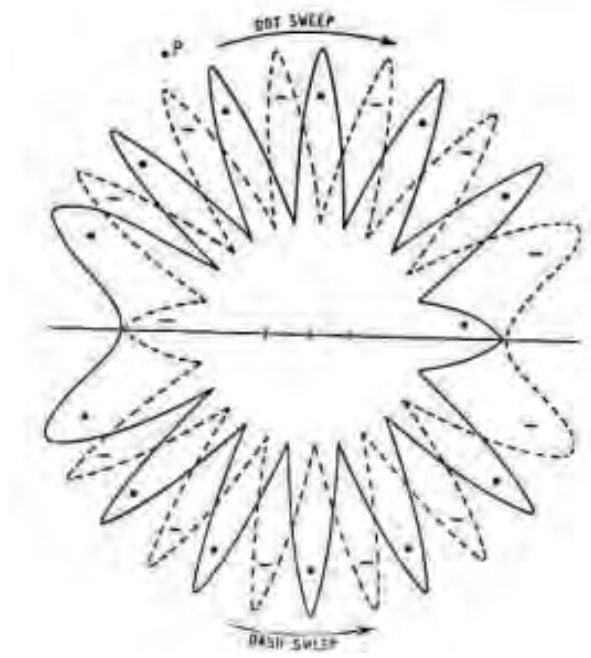


Figura 1.6: Patrón de Radiación.

la superposición del punto y la raya del código Morse. EL operador en el receptor identifica la base en los primeros 6 segundos y luego de los 2 de silencio espera hasta escuchar la señal continua. Cuenta entonces cuantos puntos (o rayas, según lo que haya escuchado primero) hubo hasta la señal continua (fig. 1.7). Traslada esos números a cartas de navegación y determina la posición con un error de $1/4$ de grado .

En 1944 luego de capturar un submarino los ingleses comenzaron a usar el sistema alemán con el nombre de CONSOL y fue usado para fines civiles hasta el año 1991.

1.2.3 GEE

Comenzó a operar en 1942 y fue el primer sistema de posicionamiento hiperbólico. Operaba entre los 20 y los 85 MHz. El sistema consistía en grupos de tres estaciones terrestres denominadas *GEE Chains*. Cada cadena tenía una estación maestra y dos esclavas y una misma base de tiempo controlada por la estación maestra. Cada 4ms usando pulsos de $6\mu s$. En el avión el receptor es conectado a un osciloscopio y se medían las diferencias de tiempo entre la base maestra y cada una de las esclavas (Δt_{AB} y Δt_{AC} , figura 1.8). Cada diferencia de tiempo definía una hipérbola en la superficie de la tierra con cada estación esclava y la maestra definiendo sus focos. Con estos datos se consulta una carta de navegación (figura 1.9) con las hipérbolas dibujadas sobre esta y la intersección de ambas dará dos puntos, uno de los

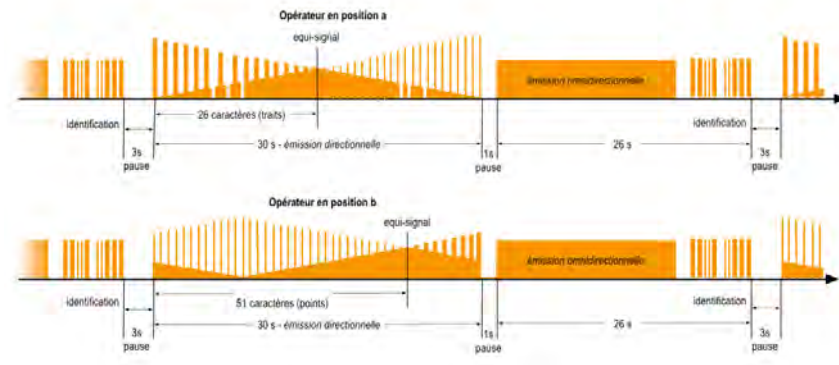


Figura 1.7: Señales del sistema SONNE. Figura tomada de <https://www.rene-reyt.fr/documents/electronique/la-radionavigation-avant-le-g-p-s/le-systeme-elektra-sonne-consol>

cuales se descarta por las posiciones anteriores y el movimiento realizado. La precisión era de 150m a 1600m, suficientes para bombardear.

1.2.4 LORAN-A

el sistema de posicionamiento LORAN (acrónimo de LOng RANGE Navigation) era similar al sistema GEE. Las cadenas de antenas (LORAN Chains) requerían de al menos 4 estaciones de referencia para determinar en forma unívoca un punto, o tres estaciones y el movimiento para descartar una de las posiciones como el sistema GEE. Operaba a 1.95MHz que permitía la propagación de una onda terrestre y otra con rebote en la ionosfera. La primera permitía un alcance de día de 1500Km y la segunda de 2000Km de noche.

1.2.4.1 LORAN-C

Es el mismo principio que el LORAN-A pero operando a 100KHz, que al ser de menor frecuencia tenía mas alcance y mejor precisión (de 19 a 90m)

1.2.4.2 eLORAN

Es una extensión del sistema LORAN, mejorado para servir como respaldo del GPS. Para calcular la posición los receptores eLORAN pueden usar tanto las estaciones de referencia terrestres de una única cadena sino tambien de las 40 estaciones e incluso otra información auxiliar que ayude a determinar la posición (como información de GPS). Opera en la misma frecuencia que el LORAN-C con precisiones de hasta 10m.

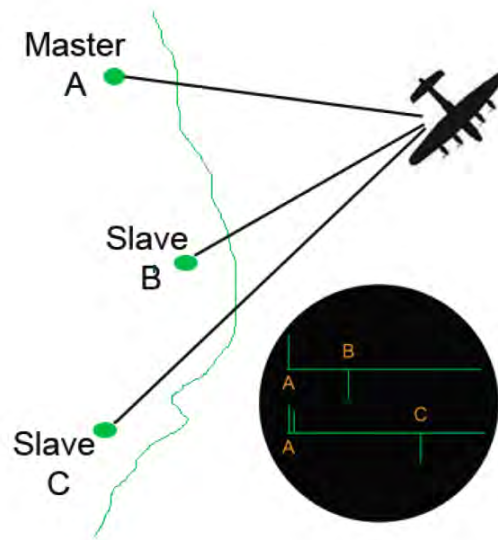


Figura 1.8: Estructura del sistema GEE con una antena maestra y dos esclavas y un ejemplo de lo que se observa en el osciloscopio del receptor.

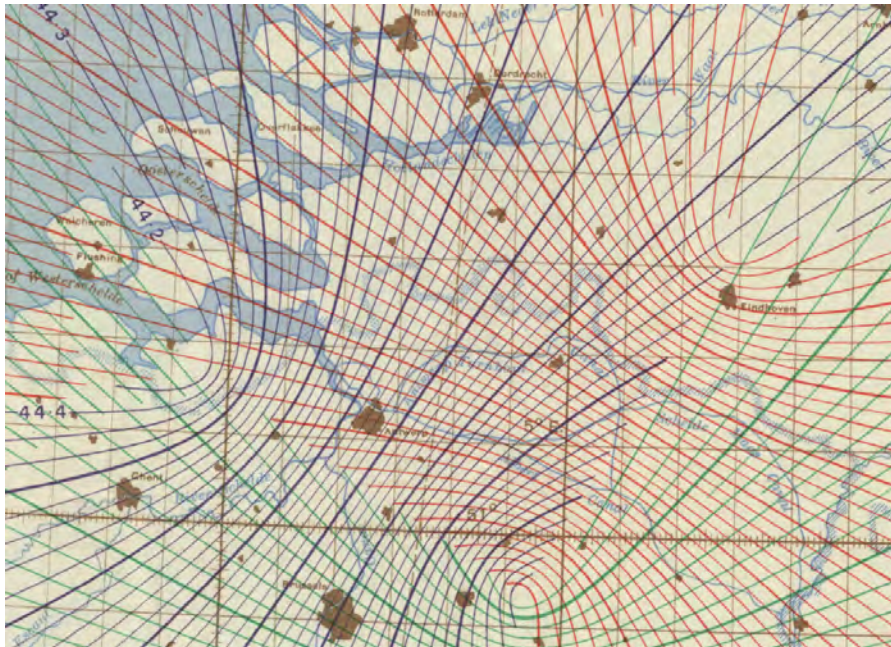


Figura 1.9: Cartas para navegación con el sistema GEE



Figura 1.10: Taxonomía del cálculo de la posición de un móvil

1.2.5 GNSS

El sistema satelital de posicionamiento global (Global Navigation Satellite System) se basa en satélites que se mueven en órbitas medias que emiten señales para que un receptor pueda triangular su posición conociendo la posición actual de los satélites. Hay varias implementaciones de este sistema como el GPS, GLONASS, BEIDU, etc.

1.3 TAXONOMÍA DE LOS SISTEMAS DE POSICIONAMIENTO

Para clasificar las distintas formas de calcular la posición de un móvil se usan tres criterios: la técnica, el involucramiento del móvil y el número mínimo de estaciones de referencia necesarias para fijar la posición. La figura 1.10 muestra en forma simplificada esta clasificación que se amplían a continuación.

1.3.1 Técnicas

1.3.1.1 Proximidad

Es el método mas simple y supone que la posición del EM está cercana a una ER. El problema con estos métodos es que cuando la propagación de la señal no es con línea de vista¹ (LOS en inglés) no se puede asegurar que la ER con mayor potencia medida en el receptor sea la que esté mas cerca. Se puede usar con *access points* de WiFi o en sistemas celulares y la pertenencia a una celda.

¹ La propagación con línea de vista es la que tiene visibilidad directa entre las antenas emisora y receptora, por lo que no debe haber obstáculo entre ambas

Su gran ventaja es la simplicidad pero la exactitud depende mucho de la densidad de las ER. El ejemplo mas claro es en las redes celulares donde el tamaño de las celdas puede ser muy grande y por lo tanto el error. Se usa como un método de respaldo mas que como un esquema de ubicación primario.

1.3.1.2 *Triangulación*

Las técnicas de triangulación usan las medidas de distancia o ángulo (multilateración y multiangulación). Todos estos métodos suponen LOS entre la EM y la ER. Los obstáculos y el multicamino reducen la aplicabilidad o tal vez la precisión. Estos métodos están limitados por el ancho de banda (tiempo) o la sensibilidad de los receptores.

1.3.1.3 *Circular*

Los puntos donde se puede llegar a ubicar la EM son círculos centrados en las ER. Es básicamente una localización basada en rango (o distancia entre la EM y la ER). Técnicamente esto se puede lograr de varias maneras. La mas común es midiendo el tiempo que tarda una señal en llegar entre uno y otro multiplicado por la velocidad de propagación, como por ejemplo hacen los GPSs. También se puede lograr midiendo la perdida de potencia de la señal. Para determinar un único punto se necesitan 3 ER en 2 dimensiones.

1.3.1.4 *Hiperbólica*

Los puntos donde se puede llegar a ubicar la EM son hipérbolas donde las ER son los focos estas curvas. Estos métodos se usan cuando se mide la diferencia de tiempo de llegada de una señal emitida en forma sincrónica por 2 ER. Para determinar un punto en 2 dimensiones se necesitan tres hipérbolas y por lo tanto 4 ER.

1.3.1.5 *Huella Digital (Fingerprinting)*

Estos definen a los métodos que correlacionan datos almacenados en una base de datos con las medidas actuales de esas propiedades. Por ejemplo los niveles de potencia medidos por el receptor y emitidos por varias ER. La idea de esta estrategia es construir una base de datos con propiedades de la señal (por ejemplo niveles de potencia) que están asociados a una posición. Luego, la EM mide ese parámetro y compara sus medidas con las de la base de datos para determinar su posición.

1.3.1.6 *Centroide*

Es un método que se usa en redes usualmente y hay cuatro tipos

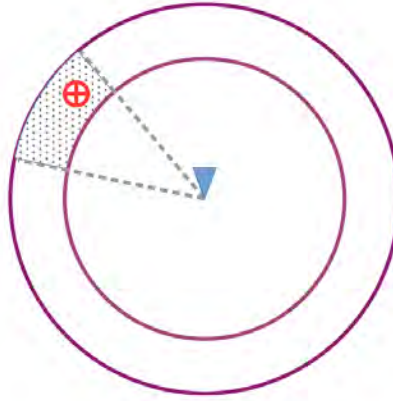


Figura 1.11: Esquema simplificado del esquema de localización de sector de anillo

CENTROIDE SIMPLE. La posición de la EM es el centro de un polígono cuyos vértices son las posiciones de las ER.

$$\begin{bmatrix} \hat{x} \\ \hat{y} \end{bmatrix} = \frac{1}{\#ER} \begin{bmatrix} \sum \hat{x}_i \\ \sum \hat{y}_i \end{bmatrix} \quad (1.2)$$

CENTROIDE PESADO. Es posible reducir el error del método simple, pesando cada vértice en función de una propiedad de la señal recibida. Por ejemplo el nivel de potencia. En situaciones de LOS este método funciona muy bien y es simple.

$$\begin{bmatrix} \hat{x} \\ \hat{y} \end{bmatrix} = \frac{1}{\sum w_k} \begin{bmatrix} \sum w_i \hat{x}_i \\ \sum w_i \hat{y}_i \end{bmatrix}, \quad w_i = \frac{RSS_i}{\sum RSS_k} \quad (1.3)$$

CENTROIDE DE SECTOR DE ANILLO. En redes celulares se suele utilizar sectorización como una forma de mejorar la cobertura y capacidades de la red. La región del sector de anillo se obtiene usando la dirección el radio y la apertura de la antena. La figura 1.11 muestra un esquema simplificado de este principio.

CENTROIDE DEL ÁREA DE SERVICIO PREDICHA. El sector de anillo es una sobre-simplificación de la cobertura de una antena y de cómo la señal llega al receptor. En general la cobertura es irregular y una forma de mejorar las estimaciones es utilizando modelos que predigan la cobertura de la ER y/o una campaña de medición para determinar esa propagación.

1.3.1.7 Híbridos

Es posible combinar métodos para mejorar el posicionamiento. El ejemplo más simple es combinar rango y ángulo (como en radares o lidars). En redes celulares se puede usar el ángulo de servicio y el tiempo de subida o bajada o la potencia. En cualquier caso, la posición se estima a partir de una medida de distancia d , ángulo θ y la posición de la ER (x_0, y_0)

$$\begin{bmatrix} \hat{x} \\ \hat{y} \end{bmatrix} = \begin{bmatrix} x_0 + d \sin \theta \\ y_0 + d \cos \theta \end{bmatrix} \quad (1.4)$$

1.3.1.8 En relación a la participación de la ER

Hay dos tipos

BASADOS EN LA RED. La posición está fijada por la ER en base a la información emitida por el móvil. Se usaba en GSM para hacer el *hand-off*.

BASADOS EN EL MOVIL. La posición la fija el móvil con información que envían las ER. Por ejemplo el GPS.

1.3.1.9 En relación al número mínimo de ER

Esta clasificación es función del número mínimo de estaciones de referencia que son necesarias para establecer una posición en forma unívoca. Esto no tiene que ver con el número de ER disponibles que por supuesto pueden ser muchas más que las mínimas. Tampoco con la distribución geométrica de las ER que pueden requerir mas ER que las mínimas. Por ejemplo, si tenemos ER colineales y estamos usando información de rango, será imposible obtener una posición hasta que no aparezca una ER adicional que esté fuera de la línea donde se encuentran las otras ER. Esta propiedad geométrica la veremos en el curso y se denomina *Dilution of Precision* en inglés.

1.4 REPASO DE RADIO FRECUENCIA

Muchas de las estrategias de localización utilizan señales de RF y debemos entender como esta información se propaga para poder analizar las limitaciones de los algoritmos. Lo que sigue es un breve repaso por conceptos básicos de propagación de señales.

El modelo de espacio libre es el modelo mas simple de propagación de la señal de RF y que considera que la señal viaja del transmisor al receptor en forma directa por un único camino y sin interferencias. Esta propagación es la denominada de línea de vista (*LOS: Line of sight* en inglés) donde el tiempo que tarda la señal en llegar, o tiempo

de vuelo, es único y la potencia recibida será función de la distancia con la siguiente ecuación:

$$P_r = P_t \cdot G_T \cdot G_R \frac{\lambda^2}{(4\pi d)^2} \quad (1.5)$$

donde P_t es la potencia transmitida, P_r la potencia recibida, G_t la ganancia de la antena del transmisor y G_R la del receptor, d la distancia que separa al transmisor del receptor y λ la longitud de onda de la señal que se relaciona con la frecuencia por la velocidad de propagación: $\lambda = c/f$ donde c es la velocidad de la luz.

Es posible expresar la potencia recibida sin conocer datos del transmisor, como la potencia transmitida. En este caso será en función de la distancia que los separa y conociendo la potencia recibida a una distancia conocida, esto es:

$$P_r(d) = P_r(d_0) \left(\frac{d_0}{d} \right)^2 \quad d \geq d_0 \geq d_f \quad (1.6)$$

donde d_0 es la distancia a la que se conoce la potencia recibida $P_r(d_0)$ y d_f es la distancia a partir de la cual es válido este modelo y se denomina campo lejano. Esta distancia es aproximadamente $d_f > 2\lambda$.

Este es, como se dijo, el modelo mas simple de propagación y define el modelo de línea de vista. Esta línea de vista puede estar interferida parcialmente o totalmente. El último es la situación de no línea de vista (*NLOS: No Line of Sight*) y se define por las zonas de Fresnel. Para introducir este y otros conceptos veremos primero los métodos de propagación de una señal de RF.

1.4.1 Métodos de Propagación Básicos

1.4.1.1 Reflexión

Cuando una onda incide sobre una superficie con dos dieléctricos diferentes, parte de la energía se va a reflejar sobre esa superficie y parte se va a transmitir en el segundo medio (figure 1.12). La cantidad de energía reflejada será proporcional al índice de reflexión. Por ejemplo, y solo para referencia, si una onda se propaga con su campo eléctrico paralelo al plano de incidencia, y el medio 1 es el aire, el coeficiente de reflexión tendrá la siguiente expresión:

$$\Gamma_{\parallel} = \frac{-\epsilon_r \sin \theta_i + \sqrt{\epsilon_r - \cos^2 \theta_i}}{\epsilon_r \sin \theta_i + \sqrt{\epsilon_r - \cos^2 \theta_i}} \quad (1.7)$$

La figura 1.12 muestra el coeficiente mostrado en la ecuación 1.7 para el caso paralelo y para el caso de una onda que incide perpendicular a la discontinuidad. En este caso para una onda que se propaga en el aire y llega a la tierra y se refleja. Se muestran situaciones donde la tierra esté seca, húmeda o sea directamente agua y se muestra en función del ángulo de incidencia (figura extraída de [24]).

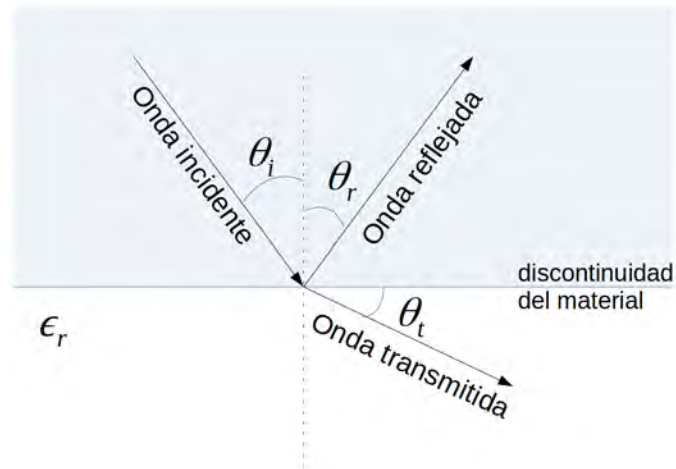


Figura 1.12: Reflexión y transmisión de una onda que se encuentra con una superficie con diferente dieléctrico

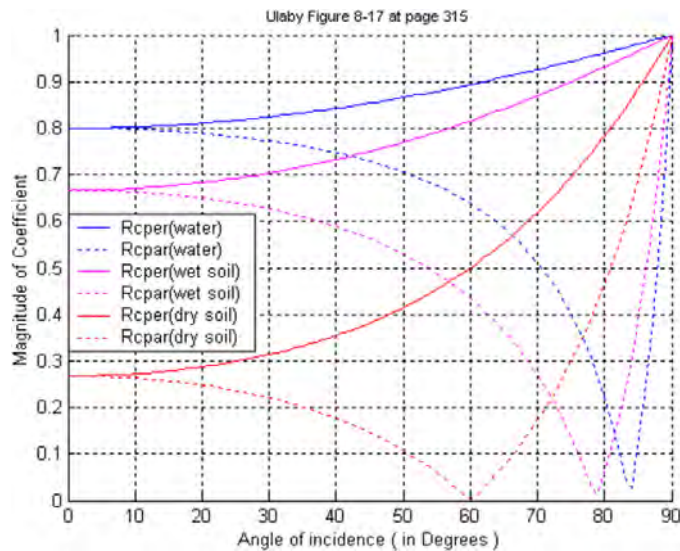


Figura 1.13: Coeficiente de reflexión (de [24]) de una onda paralela y otra perpendicular en la tierra para distintas situaciones de humedad de suelo

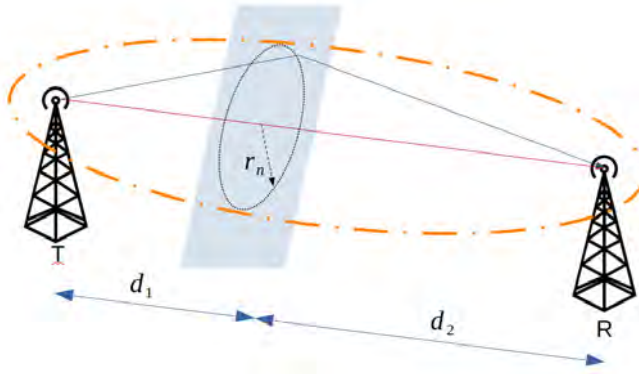


Figura 1.14: Zonas sobre una superficie colocada a una distancia d_1 del transmisor.

1.4.1.2 Difracción

El fenómeno de difracción es que el que permite que las ondas de radio se propague por detrás de las obstrucciones. Antes de introducir este fenómeno definiremos las zonas de Fresnel. Estas, definen zonas donde la fase de la señal de llegada por línea directa tienen la misma fase que otra que recorre un camino mas largo. La figura 1.14 muestra estos caminos en relación a una superficie que se coloca a una distancia d_1 del transmisor. La señal que llega a esta superficie, para que en el receptor tenga la misma fase que la señal que llega directa, tendrá que recorrer un camino mas largo. Todos los caminos posibles, definen circunferencias en la superficie cuyo radio es:

$$r_n = \sqrt{\frac{n\lambda d_1 d_2}{d_1 + d_2}} \quad (1.8)$$

Las zonas de Fresnel me permiten corregir y definir distintos modelos de propagación de la señal de RF. Para esto se define el parámetro de difracción de Fresnel-Kirchhoff:

$$v = h \sqrt{\frac{2(d_1 + d_2)}{\lambda d_1 d_2}} \quad (1.9)$$

donde h es la distancia de la obstrucción a la línea que une al transmisor y el receptor.

El fenómeno de difracción puede ser explicado de la siguiente forma. Todos los puntos de una frente de onda puede ser considerado como una fuente puntual que produce una onda secundaria que se combina con las restantes y produce un nuevo frente de onda en la dirección de propagación. La difracción está causada por la propagación de ese frente secundario de ondas en una región que oculta a la fuente principal, y se suman en el espacio que rodea al obstáculo.

1.4.1.3 Dispersión

Cuando una onda impacta sobre una superficie rugosa cuya magnitud del nivel de su longitud de onda, se hace presente el fenómeno de la dispersión. La señal al llegar a esta superficie es *dispersada* en todas las direcciones y por lo tanto se afecta la cantidad de energía que es reflejada. Se define la altura crítica de la rugosidad h_c para un ángulo de incidencia de la señal θ_i como,

$$h_c = \frac{\lambda}{8 \sin \theta_i} \quad (1.10)$$

Si la distancia entre las protuberancias mínimas y máximas son menores a h_c , la superficie es suave, si no, es rugosa y el coeficiente de reflexión es multiplicado por un factor de dispersión ρ_s para considerar la pérdida de energía. Una definición de ese factor teniendo en cuenta una distribución gaussiana de las rugosidades es,

$$\rho_s = \exp \left[-8 \left(\frac{\pi \sigma_h \sin \theta_i}{\lambda} \right)^2 \right] \quad (1.11)$$

donde σ_h es la desviación estándar de la altura de la rugosidad de la superficie.

El coeficiente de reflexión final será el producto del factor de dispersión por el de reflexión, esto es,

$$\Gamma_{rug} = \rho_s \Gamma \quad (1.12)$$

1.4.2 Modelos de Propagación

Entendiendo que la señal de RF se propaga por alguno o todos estos métodos de propagación, se pueden construir diferentes modelos. Cada uno se aplicará según el objetivo que se persigue. Por ejemplo, una cosa es conocer la cobertura de una antena transmisora y otra predecir el nivel de potencia recibida de una antena particular; o, una cosa es en una ciudad y otra diferente en un ambiente rural. Presentaremos varios modelos que servirán para entender los procesos de localización y sus errores.

1.4.2.1 Reflexión en tierra

En varias situaciones el modelo en el que hay un único camino entre transmisor y receptor, denominado de espacio libre (ec. 1.5), predice una caída de la potencia con el cuadrado de la distancia. Este modelo es a menudo suficiente para considerar escenarios simples, muchas veces con distancias muy cortas. Un modelo más completo es el de dos rayos o de reflexión en tierra (figura 1.15). Este modelo se basa en considerar además del camino directo o de línea de vista, una señal que llega al receptor luego de rebotar en el plano de tierra. Este

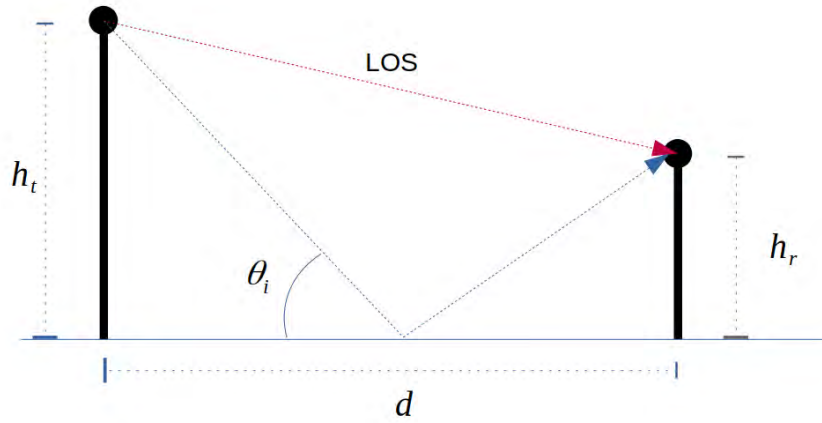


Figura 1.15: Modelo de reflexión en tierra de dos rayos. Uno de los caminos de la señal es por línea de vista y el otro por rebote en tierra.

modelo es bastante bueno para predecir el nivel de potencia recibida cuando transmisor y receptor están separados varios kilómetros y con antenas montadas en torres de transmisión.

Dado entonces la altura del transmisor y el receptor h_t y h_r respectivamente, θ_i el ángulo de incidencia de la señal que rebota en tierra con un coeficiente de reflexión determinado, a medida que el receptor se aleja del transmisor, la fase de este camino va cambiando de forma diferente que la de la línea de vista y la suma es destructiva o constructiva. Si plantea este escenario, se llega a la siguiente relación entre las potencias transmitida y recibida:

$$P_r = \frac{P_T \cdot G_T \cdot G_r \cdot \lambda^2}{(4\pi)^2 d^2} 4 \sin^2(\theta_\Delta/2) \quad (1.13)$$

donde $\theta_\Delta = \frac{2\pi}{\lambda} \frac{2h_r h_t}{d}$. Para ángulos de incidencia chicos, es decir cuando θ_Δ es menor a 0.2 radianes, el seno puede aproximarse por su ángulo, i.e. $\sin(\theta_\Delta/2) \approx \theta_\Delta/2$, entonces la potencia recibida puede simplificarse como:

$$P_r = P_T \cdot G_T \cdot G_r \frac{h_t^2 h_r^2}{d^4} \quad (1.14)$$

Esto quiere decir que cuando la distancia es grande (o $d \gg \sqrt{h_t h_r}$) la señal recibida cae con la potencia cuarta de la distancia. La figura 1.16 simula la ecuación 1.13. Allí puede verse (en rojo ecuación 1.13) que para distancias cortas la potencia cae con el cuadrado de la distancia (curva azul, ecuación 1.5) pero a distancias mayores, esta cae con la distancia a la cuarta (curva gris, ecuación 1.14).

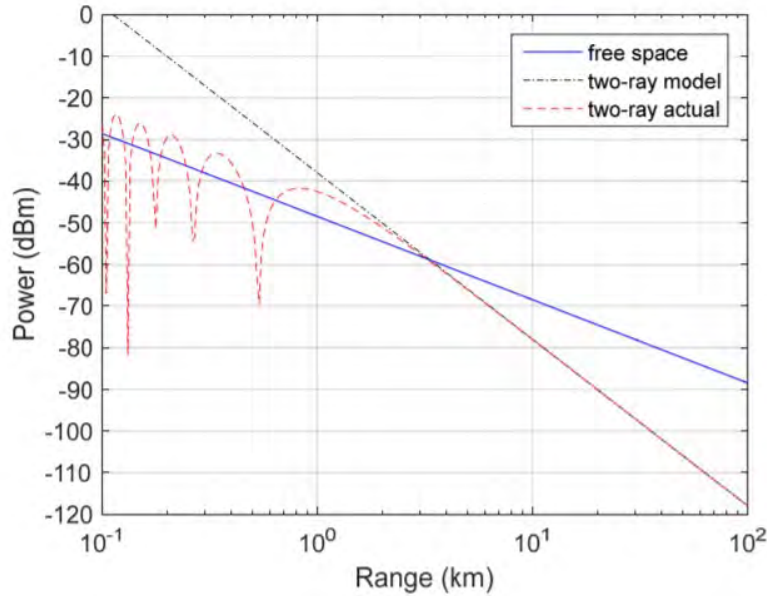


Figura 1.16: Simulación del modelo de 2 rayos y la comparación con una caída de potencia al cuadrado de la distancia la potencia cuarta.

1.4.2.2 Modelo de Filo de Cuchillo

Aunque el cálculo de pérdidas por difracción en terrenos complejos, con obstrucciones e irregulares es un problema matemáticamente difícil, se han derivado expresiones para pérdidas por difracción para muchos casos simples. Como punto de partida, el caso límite de propagación sobre el filo de un cuchillo da una buena idea del orden de magnitud de la pérdida por difracción.

Cuando la sombra es causada por un solo objeto, como una colina, una montaña o un edificio, la atenuación causada por la difracción se puede estimar tratando la obstrucción como un filo de cuchillo difractante. Este es el más simple de los modelos de difracción, y la pérdida por difracción en este caso se puede estimar fácilmente usando la solución clásica de Fresnel para el campo detrás de la obstrucción (figura 1.17).

Considere un receptor en el punto R, ubicado en la región sombreada (también llamada zona de difracción). La intensidad de campo en el punto R de la figura 3.13 es una suma vectorial de los campos debidos a todas las fuentes secundarias de Huygen en el plano por encima del filo de la navaja.

A partir del coeficiente de difracción,

$$v = h \sqrt{\frac{2(d_1 + d_2)}{\lambda d_1 d_2}} \quad (1.15)$$

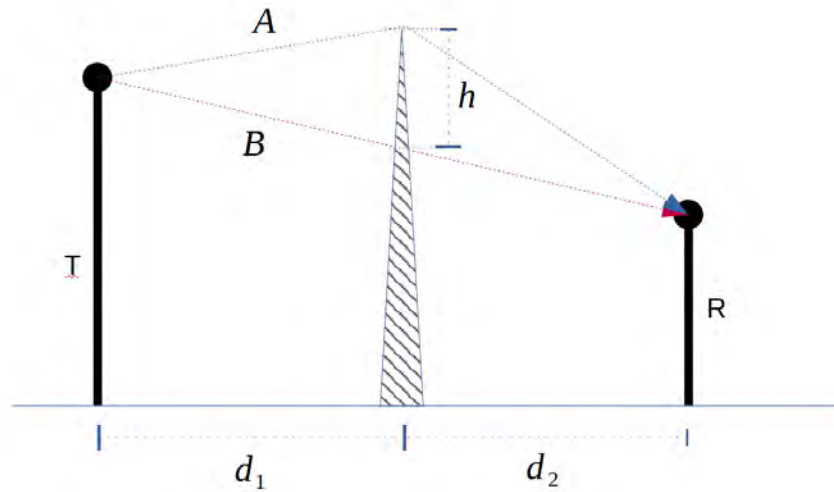


Figura 1.17: Esquema de un modelo de filo de cuchillo con una obstrucción a una distancia d_1 del transmisor.

es posible construir una aproximación a la ganancia (o pérdida) (ec. 1.16) por difracción cuya forma exacta se muestra en la figura 1.18. Esta ganancia se suma (en dB que puede ser positivo o negativo) a la potencia recibida calculada por espacio libre.

$$\begin{aligned}
 G_d(\text{dB}) &= 0 & v &\leq -1 \\
 &= 20 \log(0,5 - 0,62v) & v &\in (-1; 0] \\
 &= 20 \log(0,5 \exp(-0,95v)) & v &\in (0; 1] \quad (1.16) \\
 &= 20 \log(0,4 - \sqrt{0,1184 - (0,38 - 0,1v)^2}) & v &\in (1; 2,4] \\
 &= 20 \log(0,225/v) & v &> 2,4
 \end{aligned}$$

Hay diversas soluciones para calcular la ganancia total cuando son varias las obstrucciones presentes entre el transmisor y el receptor ([18]). El mas simple es reemplazar todas las obstrucciones por una virtual que las substituya mostrando un único camino. La figura 1.19 muestra esta estrategia para dos obstrucciones A y B. Allí, se genera una obstrucción virtual V que llega hasta el punto que une el transmisor y la obstrucción A y la B con el receptor. Las distancias d_1 y d_2 son las de esta obstrucción virtual.

1.4.2.3 Modelos experimentales

La mayoría de los modelos de propagación son derivados utilizando una combinación de métodos analíticos como los presentados y empíricos. La aproximación empírica se basa en ajustar curvas o expresiones analíticas para recrear datos que son medidos en el campo. Esto tiene la ventaja de tener en cuenta implícitamente todos los factores de propagación, tanto conocidos como desconocidos, a través de

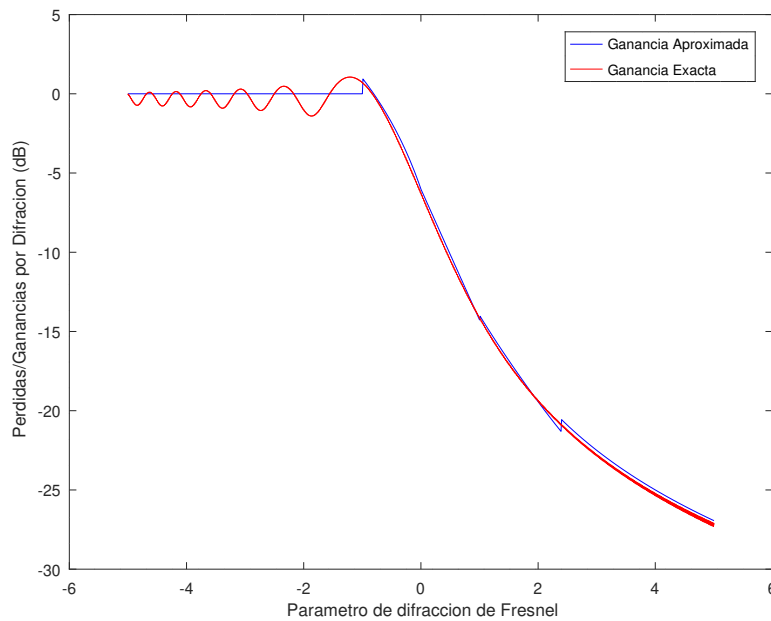


Figura 1.18: Ganancia por difracción en función del coeficiente de Fresnel. En roja la solución numérica exacta y en azul la aproximación dada en la ecuación 1.16

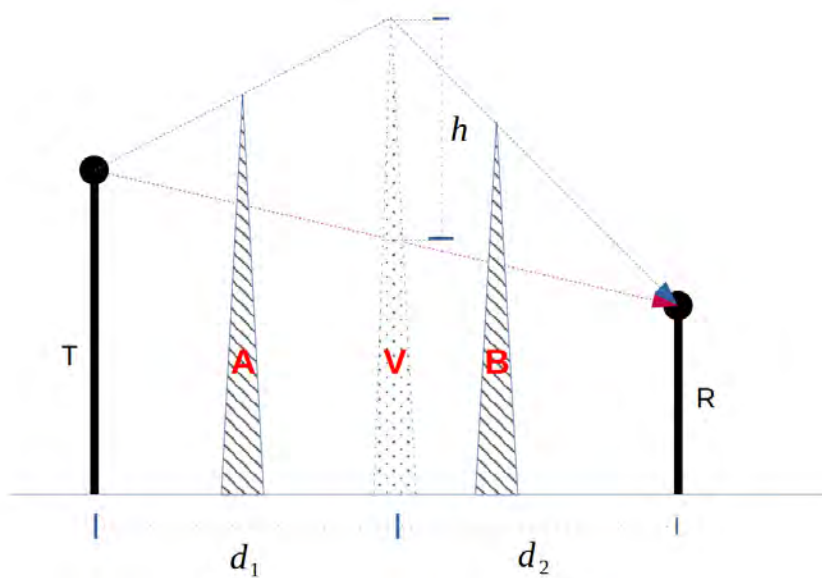


Figura 1.19: Cuando se tienen mas de una obstrucción (A y B) se considera una virtual (V) que representa el camino compuesto por reales.

Ambiente	n	σ
Espacio Libre	2	-
Area urbana 900Khz	2.7 a 3.5	~ 11 dB
Oculto por un edificio	4 a 6	~ 6 dB

Tabla 1.1: Valores característicos de n y σ

mediciones de reales. Sin embargo, la validez de un modelo empírico en frecuencias de transmisión o entornos distintos a los utilizados para derivar el modelo, solo puede establecerse mediante datos medidos adicionales en el nuevo entorno a la frecuencia de transmisión requerida.

Los siguientes son algunos de los tantos modelos empíricos que existen en la literatura y que dan una idea de la metodología de modelado.

LOG-NORMAL. Tanto los modelos de propagación teóricos como los basados en mediciones indican que la potencia promedio de la señal recibida disminuye logarítmicamente con la distancia, ya sea en canales de radio exteriores o interiores. Estos modelos se han utilizado ampliamente en la literatura. La pérdida de trayectoria media a gran escala para una separación T-R (Transmisor-Receptor) arbitraria se expresa como una función de la distancia utilizando un exponente de pérdida de trayectoria, n. La potencia recibida en dB será,

$$P_r[\text{dB}] = P_t[\text{dB}] - PL(d)[\text{dB}] \quad (1.17)$$

donde $PL(d)$ es la pérdida en el camino de la señal que depende de la distancia recorrida d y se expresa como,

$$PL(d)[\text{dB}] = \overline{PL}(d_0) + 10n \log \frac{d}{d_0} + X_\sigma \quad (1.18)$$

donde d_0 es una distancia de referencia donde se conoce la pérdida, n es el exponente de pérdida (que es 2 si el modelo es de espacio libre o 4 si es de reflexión en tierra) y X_σ es una variable aleatoria Gaussiana de media cero en dB y varianza σ .

Estos modelos se pueden ajustar a partir de toma de datos en el campo y ajustando el exponente n y la varianza σ . La tabla 1.1 muestra unos ejemplos de valores típicos de esos parámetros para distintos escenarios.

MODELO DE OKUMURA. El modelo de Okumura es uno de los modelos más utilizados para la predicción de señales de RF en áreas urbanas. Okumura viajó por la ciudad de Tokio e hizo mediciones de la atenuación de la señal desde la estación base hasta la estación móvil y construyó un conjunto de curvas que dan la atenuación media en

relación con la pérdida de trayectoria en el espacio libre. Okumura ideó tres conjuntos de datos para tres escenarios: área abierta, área urbana y área suburbana.

Este modelo es aplicable para frecuencias en el rango de 150 MHz a 1920 MHz (aunque normalmente se extrapola hasta 3000 MHz) y distancias de 1 km a 100 km. Se puede utilizar para alturas de antena de estación base que van desde 30 m a 1000 m. La expresión del modelo completo es:

$$L_{50}[\text{dB}] = L_F + A_{\text{mu}}(f, d) - G(h_{te}) - G(h_{re}) - G_{\text{AREA}} \quad (1.19)$$

donde L_{50} es la mediana del valor de la pérdida por propagación, L_F es la pérdida de espacio libre. $A_{\text{mu}}(f, d)$ es la media de atenuación relativa al espacio libre, depende de la frecuencia y la distancia y se obtiene de cartas como la figura 1.20. G_{AREA} es la ganancia debida a al tipo de ambiente y se muestra en la figura 1.21. $G(h_{te})$ es el factor de ganancia debido a la altura de la antena transmisora. y se aproxima por

$$G(h_{te}) = 20 \log \left(\frac{h_{te}}{200} \right) \quad 1000\text{m} > h_{te} > 30\text{m} \quad (1.20)$$

Finalmente, $G(h_{re})$ es el factor de ganancia debido a la altura de la antena receptora y tiene dos aproximaciones,

$$\begin{aligned} G(h_{re}) &= 10 \log \left(\frac{h_{re}}{3} \right) & h_{re} \leq 3\text{m} \\ G(h_{re}) &= 20 \log \left(\frac{h_{re}}{3} \right) & 10\text{m} > h_{re} > 3\text{m} \end{aligned} \quad (1.21)$$

MODELO DE HATA. Hata, en 1980, ideó expresiones de forma cerrada basadas en el ajuste de curvas de los modelos de Okumura. Además, extendió los modelos de Okumura para incluir los efectos debidos a la difracción, la reflexión y la dispersión de las señales transmitidas por las estructuras circundantes en un entorno urbano. La formula estándar para la pérdida media en áreas urbanas es,

$$\begin{aligned} L_{50}(\text{urbano})[\text{dB}] &= 69,55 + 26,16 \log f_c - 13,82 \log h_{te} - a(h_{re}) + \\ &+ (44,9 - 6,55 \log h_{te}) \log d \end{aligned} \quad (1.22)$$

donde f_c es la frecuencia en MHz entre 150MHz y 1500MHz, h_{te} es la altura efectiva de la antena transmisora que puede ir entre los 30m y los 200m, h_{re} es la altura efectiva de la antena del receptor en metros que puede ir entre 1m y 10m, d es la distancia T-R en kilómetros y $a(h_{re})$ es un factor de corrección que está dado para una ciudad media por

$$a(h_{re}) = (1,1 \log f_c - 0,7)h_{re} - (1,56 \log f_c - 0,8)[\text{dB}] \quad (1.23)$$

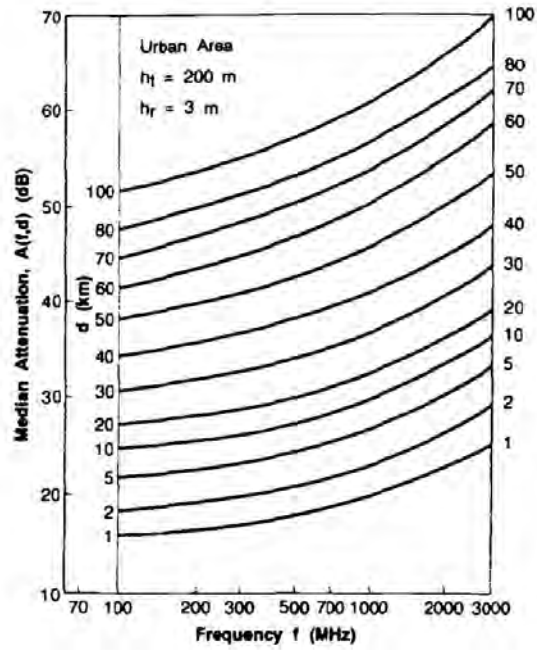


Figura 1.20: Atenuación media relativa al espacio libre $A_{mu}(f, d)$, sobre un terreno suave en función de la frecuencia [17].

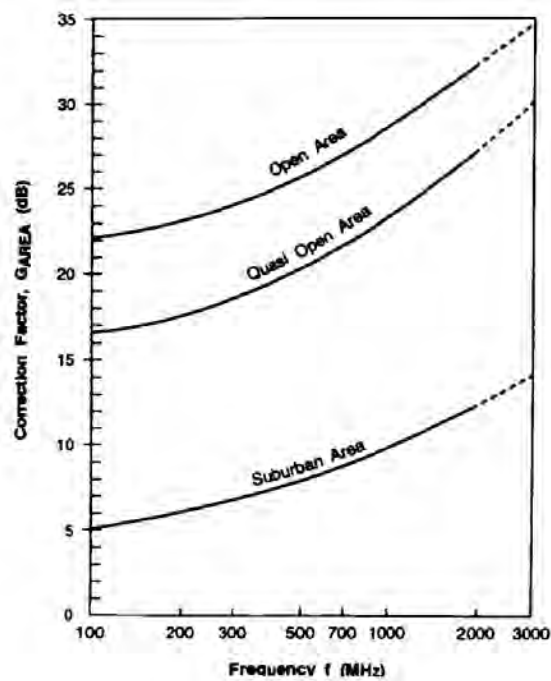


Figura 1.21: Factor de corrección G_{AREA} para diferentes ambientes en función de la frecuencia [17].

Hay ecuaciones similares a 1.23 para grandes ciudades y otras similares a 1.22 para áreas suburbanas y rurales. Y, aunque el modelo de Hata no tiene ninguna de las correcciones específicas que están disponibles en el modelo de Okumura, las expresiones anteriores tienen un valor práctico significativo y aproximan muy bien en mediana los datos reales.

1.4.3 *Modelo de Pequeña escala*

El desvanecimiento a pequeña escala, o simplemente desvanecimiento, se utiliza para describir la fluctuación rápida de la amplitud de una señal de radio durante un período corto de tiempo o movimientos en distancias del orden de la longitud de onda; fluctuaciones que permiten incluso ignorar los efectos de pérdida a gran escala. El desvanecimiento es causado por la interferencia entre dos o más versiones de la señal transmitida que llegan al receptor en momentos ligeramente diferentes. Estas ondas, llamadas ondas multi-camino, se combinan en la antena del receptor para dar una señal resultante que puede variar ampliamente en amplitud y fase, dependiendo de la distribución de la intensidad y el tiempo relativo de propagación de las ondas y el ancho de banda de la señal transmitida.

El multi-camino que crea el desvanecimiento a pequeña escala produce fundamentalmente tres efectos. Por un lado cambios rápidos en la intensidad de la señal en una pequeña distancia de viaje o intervalo de tiempo; también modulación aleatoria de frecuencia debido a cambios Doppler sobre diferentes señales multi-camino; y dispersión temporal provocada por retardos en la propagación de distintos trayectos.

El desvanecimiento ocurre porque la altura de las antenas móviles está muy por debajo de la altura de las estructuras circundantes, por lo que en general no hay una línea de vista única a la estación base. Las ondas de radio entrantes llegan desde diferentes direcciones con diferentes retrasos de propagación. La señal recibida por el móvil en cualquier punto del espacio puede consistir en un gran número de ondas planas con amplitudes, fases y ángulos de llegada distribuidos aleatoriamente. Estos componentes de trayectos múltiples se combinan vectorialmente en la antena del receptor y pueden provocar que la señal recibida por el móvil se distorsione o se desvanezca. Incluso cuando un receptor móvil está estacionario, la señal recibida puede desvanecerse debido al movimiento de los objetos circundantes en el canal de radio.

Existen muchos factores físicos en el canal de radio que influyen en el desvanecimiento a pequeña escala que incluyen:

Propagación por múltiples trayectorias: la presencia de objetos reflectantes y dispersores en el canal crea un entorno en constante cambio que disipa la energía de la señal en amplitud, fase y tiempo. Estos

efectos dan como resultado múltiples versiones de la señal transmitida que llegan a la antena receptora, desplazadas entre sí en tiempo y orientación espacial. La fase aleatoria y las amplitudes de los diferentes componentes de trayectos múltiples provocan fluctuaciones en la intensidad de la señal, lo que induce desvanecimientos a pequeña escala, distorsión de la señal o ambos. Además, puede prolongar el tiempo necesario para que la señal llegue al receptor y provocar la interferencia entre símbolos.

Velocidad del móvil: el movimiento entre el transmisor y el receptor da como resultado una modulación de frecuencia aleatoria debido a los diferentes desplazamientos Doppler en cada uno de los componentes de trayectos múltiples que será positivo o negativo dependiendo de si se alejan o no.

Velocidad de los objetos circundantes: si los objetos en el canal de radio están en movimiento, inducen un cambio Doppler que varía en el tiempo en los componentes multi-camino. Si los objetos circundantes se mueven a una velocidad mayor que el receptor, entonces este efecto domina el desvanecimiento a pequeña escala.

El ancho de banda de transmisión de la señal: la relación entre el ancho de banda de la señal de radio transmitida y el "ancho de banda" del canal genera efectos de desvanecimiento que se pueden cuantificar. Por ejemplo, si la señal transmitida tiene un ancho de banda estrecho en comparación con el canal, la amplitud de la señal cambiará rápidamente, pero la señal no se distorsionará en el tiempo. Si se distorsionará en caso contrario pero no cambiará mucho su amplitud. Consideremos a $x(t)$ que representa la forma de onda transmitida, $y(t)$ la forma de onda recibida, y $h(t, \tau)$ la respuesta de impulso del canal de radio multicamino trayecto variable en el tiempo. La respuesta al impulso $h(t, \tau)$ caracteriza completamente el canal y es una función tanto de t como de τ . La variable t representa las variaciones de tiempo debidas al paso del tiempo, mientras que τ representa señales retardadas por los trayectos múltiples del canal para un valor fijo de t (figura 1.22). La señal recibida $y(t)$ se puede expresar como una convolución de la señal transmitida $x(t)$ con la respuesta de impulso del canal,

$$y(t) = \int_{-\infty}^{+\infty} x(\tau)h(t, \tau)d\tau = x(t) \otimes h(t, \tau) \quad (1.24)$$

La figura 1.22 muestra un ejemplo para ilustrar una respuesta impulsiva de un canal multi-camino. La respuesta posee dos ejes. Sobre el eje del tiempo, la señal recibida sin retardo (en un canal con línea de vista sería la señal recibida por este camino y es la primera que llega al receptor) se representa por un impulso cuya magnitud es la de la señal. Sobre el eje de los retardos, con un impulso se representa a la señal correspondiente a tiempo t_i recibida con un retardo τ_j . Este impulso representa una versión discretizada en $\Delta\tau$ en el tiempo

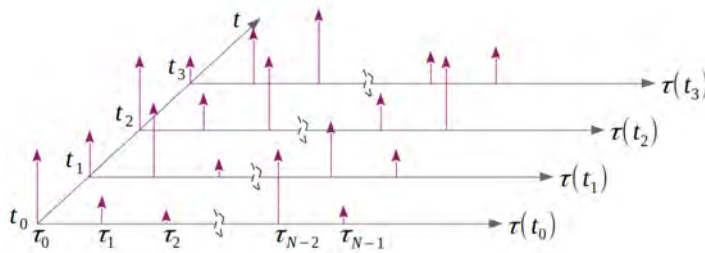


Figura 1.22: Respuesta impulsiva variante en el tiempo para un canal multi-camino.

de los retardos y se denominan retardos por exceso (*excess delay* en inglés). N es el total de retardos posibles en este canal.

1.4.3.1 Parámetros de un Canal Multi-camino

Analizar o modela estos canales puede ser complejo y requiere de equipamiento especial. Por lo que usualmente se recurre a parámetros que describan el tipo de canal sobre el que se está trabajando. Esos parámetros temporales o frecuenciales.

DISPERSIÓN TEMPORAL La figura 1.23 muestra los parámetros estadísticos que se definen sobre una respuesta impulsiva y que caracterizan el canal. Estos se miden a partir de la primera señal detectada sobre el nivel umbral de ruido y son:

- **Retardo Promedio $\bar{\tau}$.** El es promedio (o primer momento) del perfil de potencia
- **Dispersión de Retardo σ_{τ} .** Es la raíz cuadrada de segundo momento del perfil del potencia. Valores típicos para esta dispersión de del orden de los microsegundos en ambientes exteriores y nanosegundos en interiores.
- **Retardo máximo para un nivel X .** Este es el retardo para el cual el perfil de potencia cae por debajo del nivel X dB y no vuelve a subir.

Estos parámetros se definen sobre un único perfil de potencia que se obtiene sobre un promedio temporal y espacial de medidas consecutivas de la respuesta impulsiva.

ANCHO DE BANDA DE COHERENCIA Mientras que la dispersión temporal es un fenómeno natural causado por la propagación de señales reflejadas y dispersas en el canal de radio, el ancho de banda de coherencia es una relación definida derivada de la dispersión del

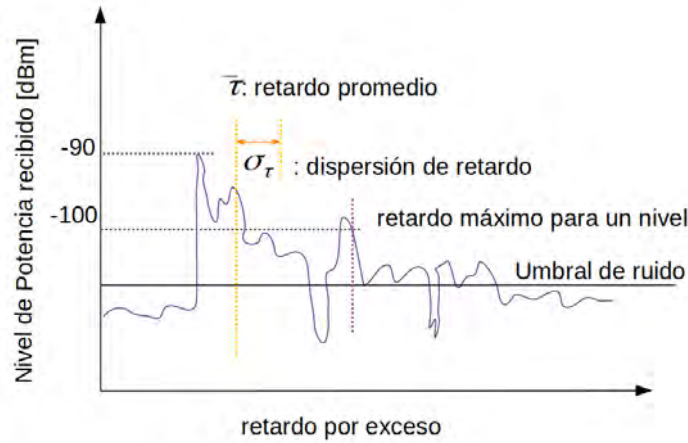


Figura 1.23: Ejemplo de un perfil de retardos de un ambiente interior. Se representan allí los parámetros de dispersión temporal.

retardo. El ancho de banda de coherencia es una medida estadística del rango de frecuencias en el que el canal puede considerarse "plano" (es decir, un canal que pasa todos los componentes espectrales de la señal con una ganancia y una fase lineal aproximadamente iguales). En otras palabras, el ancho de banda de coherencia es el rango de frecuencias en el que dos componentes de frecuencia tienen un fuerte potencial para la correlación de amplitud. Dos sinusoides con una separación de frecuencias mayor que se ven afectadas de manera muy diferente por el canal. Si el ancho de banda de coherencia se define como el ancho de banda sobre el cual la función de correlación de frecuencia está por encima de 0,9, entonces el ancho de banda de coherencia B_c es aproximadamente

$$B_c \approx \frac{1}{50\sigma_\tau} \quad (1.25)$$

DISPERSIÓN DOPPLER Y TIEMPO DE COHERENCIA La dispersión del retardo y el ancho de banda de coherencia son parámetros que describen la naturaleza dispersiva en el tiempo del canal en un área local. Sin embargo, no ofrecen información sobre la naturaleza variable en el tiempo del canal causada por el movimiento relativo entre el móvil y la estación de referencia, o por el movimiento de objetos en el canal. La dispersión Doppler y el tiempo de coherencia son parámetros que describen la naturaleza variable en el tiempo del canal en una región de pequeña escala.

La dispersión Doppler B_D es una medida del ensanchamiento espectral causado por la tasa de cambio de tiempo del canal y se define como el rango de frecuencias sobre el cual el espectro Doppler recibido es esencialmente distinto de cero. Cuando se transmite un tono sinusoidal puro de frecuencia f_c , el espectro de la señal recibida, llamado espectro Doppler, tendrá componentes en el rango de $f_c - f_d$ a

$f_c + f_d$ donde f_d es el desplazamiento Doppler. El ensanchamiento espectral depende de la velocidad relativa de movimiento y del ángulo entre la dirección de movimiento del móvil y la dirección de llegada de las ondas dispersadas. Si el ancho de banda de la señal de banda base es mucho mayor que BD, los efectos de la dispersión Doppler son insignificantes en el receptor. Este es un canal que se desvanece lentamente.

El tiempo de coherencia T_c es el dual de la dispersión Doppler en el dominio temporal y se utiliza para caracterizar la naturaleza variable en el tiempo de la dispersión de frecuencia del canal. La dispersión Doppler y el tiempo de coherencia son inversamente proporcionales entre sí. Es decir,

$$T_c \approx \frac{1}{f_m} \quad (1.26)$$

donde f_m es el máximo corrimiento Doppler. El tiempo de coherencia es la duración temporal durante el cual dos señales recibidas tienen un fuerte correlación de amplitud.

1.4.3.2 Tipos de desvanecimiento de pequeña escala

El tipo de desvanecimiento experimentado por una señal que se propaga a través de un canal de radio móvil depende de la naturaleza de la señal transmitida con respecto a las características del canal. Dependiendo de la relación entre los parámetros de la señal (como el ancho de banda, el período del símbolo, etc.) y los parámetros del canal (como la dispersión del retardo y la dispersión Doppler), las diferentes señales transmitidas experimentarán diferentes tipos de desvanecimiento. Los mecanismos de dispersión de tiempo y dispersión de frecuencia en un canal de radio móvil conducen a cuatro posibles efectos distintos, que se manifiestan dependiendo de la naturaleza de la señal transmitida, el canal y la velocidad. Mientras que la propagación del retardo por trayectos múltiples conduce a la dispersión en el tiempo y al desvanecimiento selectivo en frecuencia, la propagación Doppler conduce a la dispersión en frecuencia y al desvanecimiento selectivo en el tiempo. Los dos mecanismos de propagación son independientes entre sí.

La Figura 1.24 muestra una clasificación de cuatro tipos diferentes de desvanecimientos. Estos pueden deberse a efectos de dispersión temporal y en ese caso se compara la dispersión de los retardos con el tiempo de duración de un símbolo T_c que se transmite por el canal. Si se deben al efecto Doppler, se compara el tiempo de coherencia del canal con la duración del símbolo.

Un canal con desvanecimiento plano es aquel que no distorsiona la señal preservando su forma pero que afecta la amplitud. El canal selectivo en frecuencia por su lado, va a distorsionar la señal que llega al receptor afectando en forma diferente a cada una de las distintas

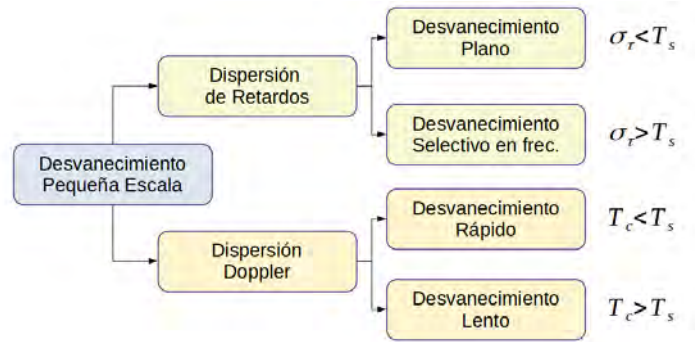


Figura 1.24: Ejemplo de un perfil de retardos de un ambiente interior. Se representan allí los parámetros de dispersión temporal.

reflexiones de la señal original. El canal con desvanecimiento rápido, el canal varía antes que un símbolo sea transmitido completamente y por lo tanto genera una distorsión. El desvanecimiento lento, es el caso donde el canal no cambia durante el tiempo del símbolo y lo afecta en amplitud sin distorsionar.

1.4.4 Comentarios finales

Para entender el problema en su totalidad y las variaciones que podrá tener en el nivel de potencia una señal, además de las caídas de potencia mencionadas anteriormente y que son por efectos de gran escala (o movimientos mayores a una longitud de onda), hay otro efecto que se suma y se denomina de pequeña escala.

La figura 1.25 ilustra las diversas contribuciones que deben tenerse en cuenta al estimar la pérdida de trayectoria para un análisis que requiera predecir la potencia recibida. Estas contribuciones son:

- Pérdida media del trayecto en función de la distancia debida al desvanecimiento a gran escala. Es el tipo de pérdida que vimos en las secciones anteriores.
- Desvanecimiento de pequeña escala o Rayleigh (típicamente 20–30 dB)

En la figura se sugiere en una análisis de peor caso, un área de $\approx 1 - 2\%$ que es la probabilidad debajo de la cual se desprecia el efecto. Esto indica que esa potencia será la recibida en el 98–99 % de los casos.

En función de este modelo, una señal recibida tendrá dos componentes debido a dos variables aleatorias (fig. 1.26). Una correspondiente al desvanecimiento de gran escala o la pérdida con modelos como el de espacio libre y que se modela con una distribución gaussiana (ecuación 1.18). La otra, correspondiente a las variaciones importantes de la señal debido a al fenómeno de pequeña escala con

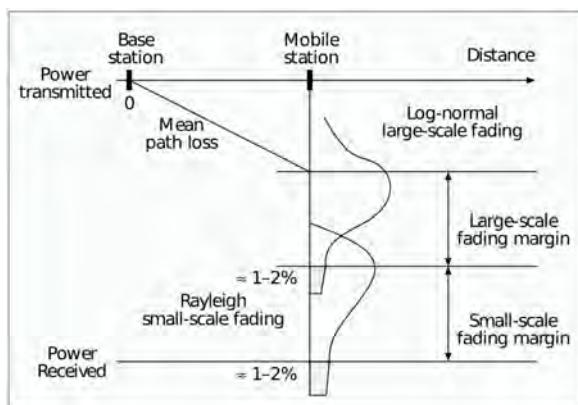


Figura 1.25: Contribuciones a la pérdida de trayectoria en un canal de comunicaciones [23].

una distribución Rayleigh y que se debe a que la señal no llega al receptor desde un único camino sino desde múltiples direcciones y con diferentes fases por el efecto de la difracción y reflexión en las edificaciones e infraestructura. Movimientos de fracciones de la longitud de onda producen por lo tanto variaciones importantes de la señal. A estos efectos se suma el efecto Doppler cuando el receptor está en movimiento.

1.5 MULTILATERACIÓN CIRCULAR

En este caso la línea de posición donde se ubica la EM son círculos centrados en cada ER. La intersección de al menos tres círculos define la posición del móvil de forma unívoca en dos dimensiones (2D) y cuatro en tres (3D). El radio de esos círculos es información de *rango* que se puede obtener de varias maneras. Una es la información de **tiempo de vuelo** de una señal emitida por móvil y recibida por la ER (uplink o camino de subida) o al revés (downlink o camino de bajada).

La figura 1.27 muestra el caso en el que cuatro ER (triángulos negros) observan un móvil (círculo negro) midiendo la distancia entre ellos. Esto define para cada ER una circunferencia. En la figura de la izquierda no hay error en la medición y esto determina un único punto que es donde se encuentra el móvil. En la figura de la derecha, cada medida tiene error y en lugar de un punto existe un *área de confianza* que es la intersección de todos los círculos. Según el algoritmo utilizado, la posición estimada será aproximada. En la figura, la estrella blanca representa como solución al centroide del área; el cuadrado blanco la solución de mínimos cuadrados.

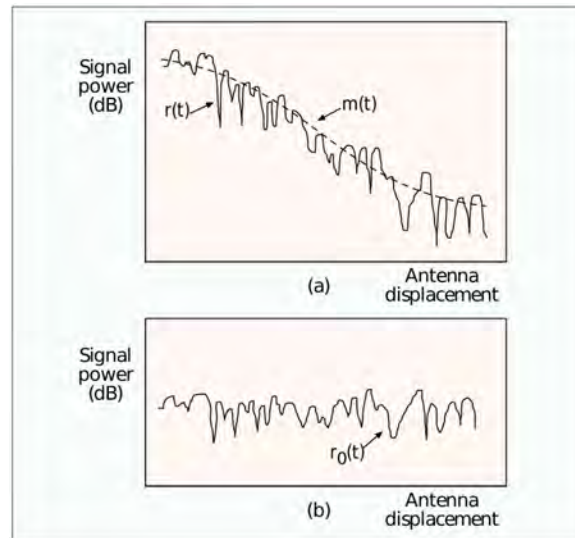


Figura 1.26: Contribuciones a la pérdida de trayectoria en un canal de comunicaciones [23].

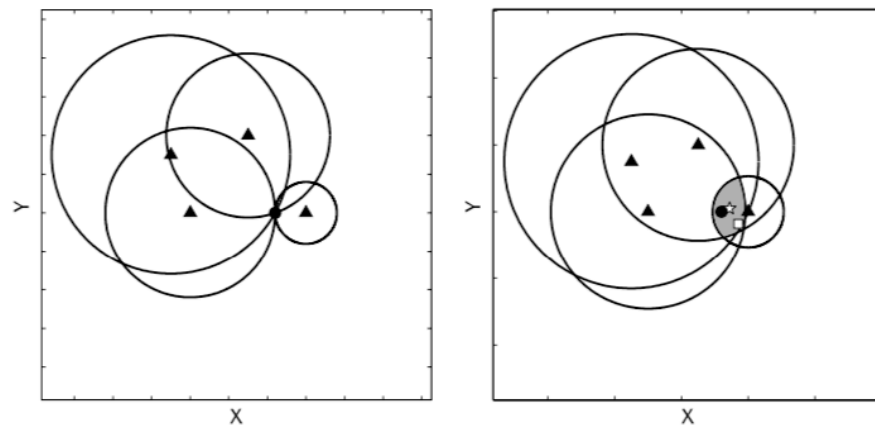


Figura 1.27: Líneas de posición utilizando el rango entre las ER y el móvil.

A la izquierda cuando las medidas son sin incertidumbre y a la derecha cuando hay error en su determinación. Los triángulos representan las ER, la estrella es el centroide de la zona de la región donde se cruzan todos los círculos, el cuadrado es una solución de posición basada en mínimos cuadrados y el círculo la posición real del móvil [21]

1.5.1 Uso de Tiempo de Arribo

El tiempo de arribo (TOA: Time of arrival en inglés) o el tiempo de vuelo (TOF: Time of flight en inglés) es el tiempo que tarda una señal (emitida por el móvil o las ER) en llegar del transmisor al receptor y proporciona información de distancia entre ambos. Esta estrategia requiere de algo fundamental: **sincronización** entre todos, tanto ER como el Móvil.

En todos los casos prácticos las circunferencias que representan las líneas de posición, nunca se intersectan en un punto por el error con que se mide este tiempo. La fuente de estos errores es múltiple. Si la señal es de RF la existencia o no de línea de vista y el multicamino entre otras suman al error. También de qué forma se mide el tiempo afecta el error, aunque puede ser casi despreciable si $c * \tau \ll d$. Es decir que el error en la medición del tiempo (representado por τ) multiplicado por la velocidad de la luz c es mucho menor que la distancia que separan a las ER y el móvil. Podemos definir entonces el rango como:

$$\hat{d}_i = c(t_i + n_i) \quad (1.27)$$

donde \hat{d}_i es la distancia obtenida entre la ER- i ésima y el móvil, t_i es el tiempo sin error del tiempo de vuelo y n_i es el error representado como un retardo aleatorio del que podemos conocer su distribución probabilística por experimentación o simulación. La ecuación de cada circunferencia centrada en las ER será:

$$\begin{aligned} (x - x_1)^2 + (y - y_1)^2 &= \hat{d}_1^2 \\ (x - x_2)^2 + (y - y_2)^2 &= \hat{d}_2^2 \\ \vdots &= \vdots \\ (x - x_N)^2 + (y - y_N)^2 &= \hat{d}_N^2; \end{aligned} \quad (1.28)$$

donde $\mathbf{s} = [x, y]^T$ es la posición del móvil y $\mathbf{b}_i = [x_i, y_i]^T$ la de las ER o "bases". Cuando tenemos error, este sistema de ecuaciones (ec. 1.28) no tienen una solución cerrada.

Una estrategia de solución simple es el *método de las secantes*. Estas secantes son líneas que unen los puntos donde se intersectan dos circunferencias. Con dos líneas, donde estas se cruzan será la solución. Por lo tanto, tomamos una de las circunferencias como referencia y con las otras (dos como mínimo) definimos las rectas (fig 1.28). Matemáticamente, se resta una de las ecuaciones de 1.28 del resto. Por ejemplo, si restamos la primer ecuación del resto obtenemos para los dos primeras:

$$(x - x_1)^2 + (y - y_1)^2 - (x - x_2)^2 - (y - y_2)^2 = \hat{d}_1^2 - \hat{d}_2^2 \quad (1.29)$$

expandiendo esta ecuación,

$$\begin{aligned} x^2 - 2x x_1 + x_1^2 + y^2 - 2y y_1 + y_1^2 \dots \\ - x^2 + 2x x_2 - x_2^2 - y^2 + 2y y_2 - y_2^2 = \hat{d}_1^2 - \hat{d}_2^2 \end{aligned} \quad (1.30)$$

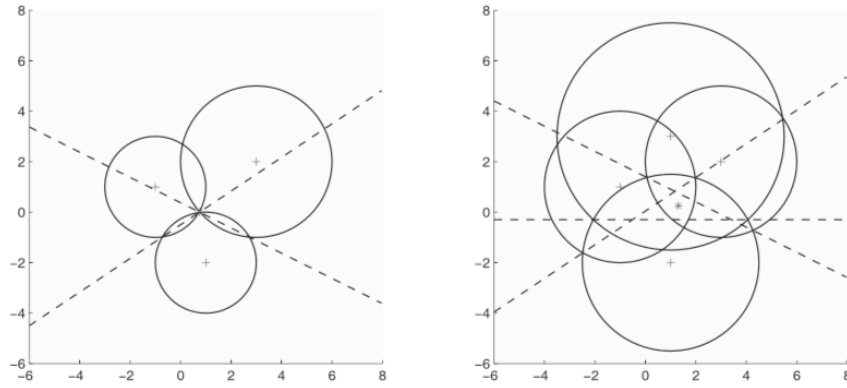


Figura 1.28: Método de las secantes. Cada circunferencia que se cruza con una tomada como referencia define una recta. Donde se cruzan está la solución. A la izquierda cuando no hay error en la medida y a la derecha, medidas con error definen una región de confianza

y separando los términos que contienen la incógnita,

$$2[x(-x_1 + x_2) + y(-y_1 + y_2)] = \hat{d}_1^2 - \hat{d}_2^2 - x_1^2 + x_2^2 - y_1^2 + y_2^2 \quad (1.31)$$

Aplicando las mismas estrategias sobre las ecuaciones restantes, obtenemos un sistema de ecuaciones que representado en forma matricial queda,

$$\begin{bmatrix} (-x_1 + x_2) & (-y_1 + y_2) \\ \vdots & \vdots \\ (-x_1 + x_N) & (-y_1 + y_N) \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} \hat{d}_1^2 - \hat{d}_2^2 - x_1^2 + x_2^2 - y_1^2 + y_2^2 \\ \vdots \\ \hat{d}_1^2 - \hat{d}_N^2 - x_1^2 + x_N^2 - y_1^2 + y_N^2 \end{bmatrix} \quad (1.32)$$

Denominemos **A** a la matriz que multiplica al vector incógnita **s**, de 2 columnas por N filas y **B** al vector columna del miembro de la derecha. Para obtener la solución de la ecuación matricial **A · s = B** se necesita la inversa de **A** o la pseudoinversa en el caso que **A** no sea cuadrada, que es el caso general. La solución **ŝ** será,

$$\hat{\mathbf{s}} = \begin{bmatrix} \hat{x} \\ \hat{y} \end{bmatrix} = \frac{1}{2}(\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T \mathbf{B} \quad (1.33)$$

Esta solución corresponde a la solución de Mínimos Cuadrados Lineal de la ecuación 1.32 o la aproximación de mínimos cuadrados de una función lineal. En el caso de una medida de rango con error, la solución es el centroide de la región de confianza que se ve en la figura 1.28.

1.5.1.1 Pseudo-rango

Como ya se dijo, para implementar este método se necesita sincronización entre todos, el móvil y todas las ER. La sincronización es

relativamente simple de lograr en las ER. Sea porque el reloj es muy preciso (por ejemplo usando información de GPS en exteriores) o porque es posible implementar en ellos algoritmos para lograrlo. Sin embargo, con el objetivo de lograr receptores simples en el Móvil es probable que el reloj de este no pueda estar sincronizado con las ER. En este caso podemos definir el error de sincronización de la siguiente forma:

$$T_i = t_i + \eta \quad T_{\text{Transm}_j} = t_{\text{Transm}_j} + \varepsilon_i \quad (1.34)$$

donde T_i es el tiempo recibido y η el error por falta de sincronización del reloj local y T_{Transm_j} es el tiempo en que se transmitió el mensaje donde ε_i es el error en el reloj de la ER i -ésima. Este último puede ser despreciable si se logra la sincronización entre las ER.

Una posible solución a este problema es definir una nueva medida de distancia que tenga en cuenta el error en el receptor. A esta distancia la denominaremos *seudo-rango*,

$$\rho = \sqrt{(x - x_i)^2 + (y - y_i)^2} - b \quad (1.35)$$

donde $b = c\eta$ es el sesgo en el rango debido al error en el reloj del receptor, $\mathbf{s} = [x, y]^T$ la posición del móvil y $\mathbf{b}_i = [x_i, y_i]^T$ la de la ER i -ésima. Con cada ER podemos construir un sistema de ecuaciones:

$$\begin{aligned} \rho_1 &= \sqrt{(x - x_1)^2 + (y - y_1)^2} - b \\ \rho_2 &= \sqrt{(x - x_2)^2 + (y - y_2)^2} - b \\ &\vdots \\ \rho_N &= \sqrt{(x - x_N)^2 + (y - y_N)^2} - b \end{aligned} \quad (1.36)$$

Una forma de resolver esto es restando las ecuaciones tomando una como referencia de forma que b desaparezca,

$$\rho_1 - \rho_2 = \sqrt{(x - x_1)^2 + (y - y_1)^2} - \sqrt{(x - x_2)^2 + (y - y_2)^2} \quad (1.37)$$

Esta solución significa que la LdP son hipérbolas y la solución la veremos en la próxima sección. Otra forma es construir un estimador que en forma iterativa encuentre la solución. Una alternativa es el filtro de Kalman.

1.5.1.2 Filtro de Kalman para pseudo-rango

Las ecuaciones del filtro de Kalman si tenemos un modelo de transición de estados lineal y un modelo no lineal de observación (como es el caso que nos ocupa) será:

$$\begin{aligned}\hat{\mathbf{x}}_{k|k-1} &= \mathbf{F}_k \hat{\mathbf{x}}_{k-1|k-1} \\ \mathbf{P}_{k|k-1} &= \mathbf{F}_k \mathbf{P}_{k-1|k-1} \mathbf{F}_k^T + \mathbf{Q}_k \\ \hat{\mathbf{x}}_{k|k} &= \hat{\mathbf{x}}_{k|k-1} + \mathbf{W}_k [\mathbf{z}_k - \hat{\mathbf{h}}(\hat{\mathbf{x}}_{k|k-1})] \\ \mathbf{P}_{k|k} &= \mathbf{P}_{k|k-1} - \mathbf{W}_k \mathbf{S}_k \mathbf{W}_k^T\end{aligned}\quad (1.38)$$

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

Para resolver la ecuación 1.38 debemos tener los modelos de movimiento (\mathbf{F}_k y \mathbf{Q}_k) y los de observación ($\mathbf{h}(\mathbf{x}_k)$ y \mathbf{R}_k). El vector de estados lo definimos como $\mathbf{x}_k = [x, y, \dot{x}, \dot{y}, b]^T$, donde $[x, y]$ es la posición en dos dimensiones del móvil, $[\dot{x}, \dot{y}]$ es su velocidad y b el error en el reloj local. Si el móvil está quieto, las matrices de transición de estados serán:

$$\mathbf{F}_k = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} \quad \mathbf{Q}_k = \begin{bmatrix} \sigma_x^2 & 0 & 0 & 0 & 0 \\ 0 & \sigma_y^2 & 0 & 0 & 0 \\ 0 & 0 & \sigma_{\dot{x}}^2 & 0 & 0 \\ 0 & 0 & 0 & \sigma_{\dot{y}}^2 & 0 \\ 0 & 0 & 0 & 0 & \sigma_b^2 \end{bmatrix} \quad (1.39)$$

donde σ_* es la varianza de la variable de estado correspondiente. Si el móvil se mueve, puede plantearse un modelo de **seguimiento** cambiando la matriz \mathbf{F}_k

$$\mathbf{F}_k = \begin{bmatrix} 1 & 0 & \Delta T & 0 & 0 \\ 0 & 1 & 0 & \Delta T & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} \quad (1.40)$$

$$\mathbf{Q}_k = \begin{bmatrix} \frac{\sigma_{\alpha_x}^2}{4} \Delta T^4 & 0 & \frac{\sigma_{\alpha_x}^2}{2} \Delta T^3 & 0 & 0 \\ 0 & \frac{\sigma_{\alpha_y}^2}{4} \Delta T^4 & 0 & \frac{\sigma_{\alpha_y}^2}{2} \Delta T^3 & 0 \\ \frac{\sigma_{\alpha_x}^2}{2} \Delta T^3 & 0 & \sigma_{\alpha_x}^2 \Delta T^2 & 0 & 0 \\ 0 & \frac{\sigma_{\alpha_y}^2}{2} \Delta T^3 & 0 & \sigma_{\alpha_y}^2 \Delta T^2 & 0 \\ 0 & 0 & 0 & 0 & \sigma_b^2 \end{bmatrix} \quad (1.41)$$

donde σ_{α_*} es la varianza que representa el error en aceleración en cada dimensión, es decir, se supone que el error del modelo de movimiento tiene que ver con variaciones alrededor de cero de aceleraciones. En este caso la matriz de incertidumbre del modelo (\mathbf{Q}_k) deberá reflejar las nuevas incertidumbre del modelo. Este modelo representa el movimiento de una partícula libre.

Finalmente, el modelo de observación será no lineal,

$$\mathbf{h}(\mathbf{x}) = \begin{bmatrix} \sqrt{(x-x_1)^2 + (y-y_1)^2} - b \\ \sqrt{(x-x_2)^2 + (y-y_2)^2} - b \\ \vdots \\ \sqrt{(x-x_N)^2 + (y-y_N)^2} - b \end{bmatrix} = \begin{bmatrix} h_1(\mathbf{x}) \\ h_2(\mathbf{x}) \\ \vdots \\ h_N(\mathbf{x}) \end{bmatrix} \quad (1.42)$$

entonces hay que calcular el Jacobiano para aplicar el EKF (que veremos en el capítulo 2 pero presentamos aquí para referencia)

$$\mathbf{H} = \begin{bmatrix} \frac{\partial h_1}{\partial x} & \frac{\partial h_1}{\partial y} & \frac{\partial h_1}{\partial \dot{x}} & \frac{\partial h_1}{\partial \dot{y}} & \frac{\partial h_1}{\partial b} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \frac{\partial h_N}{\partial x} & \frac{\partial h_N}{\partial y} & \frac{\partial h_N}{\partial \dot{x}} & \frac{\partial h_N}{\partial \dot{y}} & \frac{\partial h_N}{\partial b} \end{bmatrix} \quad (1.43)$$

1.5.2 Uso del Nivel de Potencia

Para la localización usando la información de la potencia de RF recibida para obtener la distancia entre la ER y la EM, se utilizan los modelos de gran escala que se vieron en la sección 1.4.2. En general este método tiene la ventaja de que no se necesita sincronización y es usual que los receptores midan este nivel llamado RSS por sus siglas en inglés *Radio Signal Strength*. El problema central es que la potencia recibida tiene mucha variabilidad aún teniendo un buen modelo de propagación. El problema central es la variación de pequeña escala y que la posición de la antena no se puede controlar. Por ejemplo, la figura 1.29 extraída de [10] muestra la variación de potencia recibida en un móvil para una persona en el mismo lugar pero con el teléfono en distintas posiciones (en la oreja, en la mano o en el bolsillo).

Retomamos la ecuación 1.18 pero expresada en función de las medidas de potencia o RSS,

$$\text{RSS}(d)[\text{dB}] = \text{RSS}(d_0) + 10n \log \frac{d}{d_0} + w \quad (1.44)$$

donde d es la distancia entre la ER y la EM y que para una medida se puede estimar invirtiendo este modelo,

$$\hat{d} = d_0 10^{\frac{\text{RSS}(d) - \text{RSS}(d_0)}{10n}} \quad (1.45)$$

Usar esta ecuación es, como ya se comentó, una forma muy imprecisa para determinar la distancia. La figura 1.30 muestra una experimento de medición en un ambiente interior (de [19]) donde se puede determinar que $n = 2,3$ y una varianza $\sigma_w = 6\text{dB}$

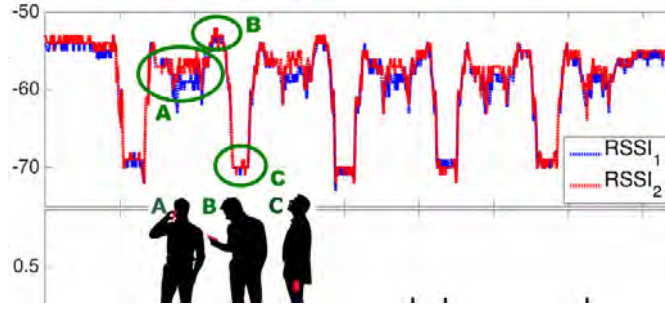


Figura 1.29: Variacion RSS

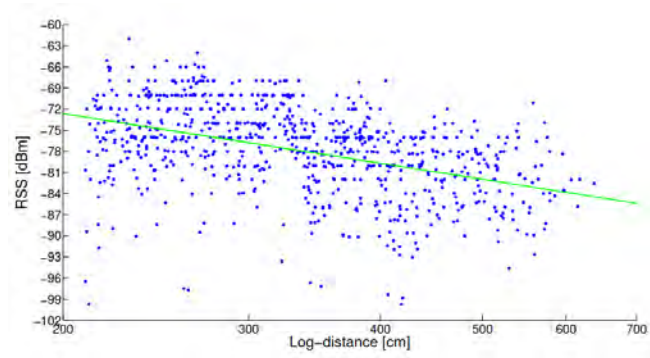


Figura 1.30: Medidas experimentales de RSS (reproducido de [19]).

Otro problema relevante en las mediciones de RSS son los valores atípicos o *outliers*. A partir de las medidas de distancia entre las ER y la EM, la solución de mínimos cuadrados habíamos visto que era (ec. 1.33)

$$\hat{\mathbf{s}} = \begin{bmatrix} \hat{x} \\ \hat{y} \end{bmatrix} = \frac{1}{2}(\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T \mathbf{B} \quad (1.46)$$

Supongamos las siguientes matrices resultantes para A y B,

$$\mathbf{A} = \begin{pmatrix} 2 & 6 \\ 2 & 5,99999 \end{pmatrix} \quad \mathbf{B} = \begin{pmatrix} 8 \\ 8,00001 \end{pmatrix} \quad (1.47)$$

Esto da como solución $\hat{\mathbf{s}} = (7; -1)^T$. Si suponemos que la matriz \mathbf{B} tiene una pequeña desviación y es igual a $\mathbf{B} = (8; 8,001)^T$, el resultado de la estimación será $\hat{\mathbf{s}} = (304, -100)^T$. Esto demuestra que en pequeños errores en las medidas, puede producir resultados muy diferentes. Hay estrategias para aliviar este problema siempre que haya varias ER disponibles y no las mínimas de tres para localizar en dos dimensiones o 4 en tres. Uno simple es el siguiente [19]:

1. La EM recibe la información de las N ER para determinar la distancia a ellas

2. La EM ordena la RSSI de la mas grande a la más pequeña
3. Agrupo de a tres ER y armo la matriz A
4. Calculo la condición de cada matriz A
5. Elijo el grupo con mejor numero de condición y las utilizo para estimar la posición de la EM.

1.5.2.1 Mínimos Cuadrados iterativo

Las estrategias de posicionamiento presentadas son soluciones cerradas de mínimos cuadrados. Sin embargo, es posible plantear soluciones iterativas que permitan trabajar con medidas con niveles de ruido altos. Repetimos la ecuación para la distancia entre la ER y la EM,

$$\hat{d}_i^2 = (x - x_i)^2 + (y - y_i)^2, \quad i = 1, \dots, N \quad (1.48)$$

donde

$$\hat{d}_i = 10^{\frac{RSS_i - RSS_0}{10n}} \quad (1.49)$$

Si diferenciamos 1.48,

$$\delta d_i = \frac{(x_i - x)\delta x + (y_i - y)\delta y}{\sqrt{(x - x_i)^2 + (y - y_i)^2}}, \quad i = 1, \dots, N \quad (1.50)$$

donde δx y δy son los valores desconocidos que deseamos estimar, x e y son conocidas o una estimación actual y $\delta d_i = d_i - \hat{d}_i$. La solución de mínimos cuadrados será,

$$\begin{bmatrix} \delta x \\ \delta y \end{bmatrix} = \begin{bmatrix} \frac{x_1 - x}{d_1} & \frac{y_1 - y}{d_1} \\ \frac{x_2 - x}{d_2} & \frac{y_2 - y}{d_2} \\ \vdots & \vdots \\ \frac{x_N - x}{d_N} & \frac{y_N - y}{d_N} \end{bmatrix}^+ \begin{bmatrix} \delta d_1 \\ \delta d_2 \\ \vdots \\ \delta d_N \end{bmatrix} \quad (1.51)$$

donde $^+$ es la pseudoinversa. El algoritmo comienza desde una posición inicial propuesta y va iterando encontrando una solución en el paso i -ésimo $(x + \delta x, y + \delta y)$ y el método finaliza cuando el error $\sqrt{\delta x^2 + \delta y^2}$ es menor que un umbral predeterminado.

1.5.2.2 Bayesiano

Se puede reformular el problema como una estimación Bayesiana. Reformulamos,

$$P_{r,i} = \theta_i + n_i \quad (1.52)$$

donde $P_{r,i}$ es la RSS recibida y medida y $\theta_i = RSS_0 - 10n \log \frac{d_i}{d_0}$ es una variable aleatoria. Podemos entonces aplicar Bayes para encontrar la distribución a posteriori de la variable aleatoria condicionada a las mediciones,

$$p(\theta_i | P_{r,i}) = \frac{p(P_{r,i} | \theta_i) p(\theta_i)}{p(P_{r,i})} \quad (1.53)$$

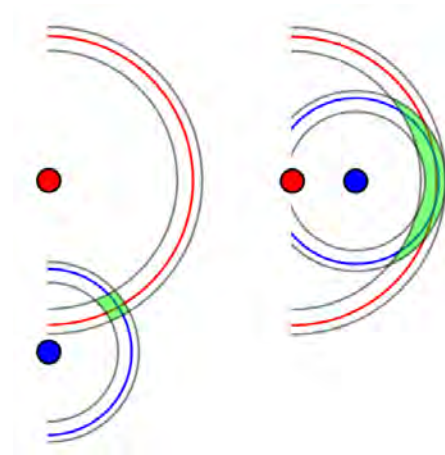


Figura 1.31: Descripción geométrica de la DOP. En la figura de la izquierda las medidas tienen algún error y la verdadera posición se encuentra en la zona coloreada de verde. En la figura de la derecha, el error es el mismo, pero por la posición relativa de las estaciones de referencia la zona verde crece considerablemente

1.6 PRECISIÓN

Sabemos que las medidas no van a ser perfectas. Por lo tanto tendremos errores en la posición estimada. Algunos aspectos que afectan la estimación tendrán que ver con la forma en que se mide el rango y otros con aspectos geométricos.

Recordamos que la precisión se define por las propiedades estadísticas de las medidas y se refiere a la dispersión del conjunto de valores obtenidos en mediciones repetidas de una posición.

1.6.1 DOP: *Dilution of Precision*

La dilución de la precisión o como se conoce mas comúnmente en inglés por *Dilution of precision*, tiene que ver mas que con los errores propios de cada medición sino mas bien con la distribución geométrica de las estaciones de referencia. La figura 1.31 muestra geométricamente la dilución de la precisión. A la izquierda dos estaciones de referencia miden con error la distancia a la que se encuentra la EM y se define una zona (en verde) donde esta se encuentra. A la derecha, dos estaciones con el mismo error de medición pero ubicadas de forma que la zona donde se encuentra la EM creció sin que el error de medición haya cambiado.

Para calcular la DOP, retomemos la ecuación 1.43 que representa los Jacobiandos evaluados en la estimación de la posición. Definimos entonces,

$$\mathbf{A} = (\mathbf{H}^T \mathbf{H})^{-1} = \begin{bmatrix} \sigma_x^2 & \sigma_{x,y}^2 & \sigma_{x,\dot{x}}^2 & \sigma_{x,\dot{y}}^2 & \sigma_{x,b}^2 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \sigma_{b,x}^2 & \sigma_{b,y}^2 & \sigma_{b,\dot{x}}^2 & \sigma_{b,\dot{y}}^2 & \sigma_b^2 \end{bmatrix} \quad (1.54)$$

Esta matriz tiene relación directa con el elipse que define el error de la estimación. Se define entonces a

- DOP Horizontal: $\sigma_{\text{HDOP}} = \sqrt{a_{11} + a_{22}}$
- DOP Temporal: $\sigma_{\text{TDOP}} = \sqrt{a_{55}}$
- DOP Geometría: $\sigma_{\text{GDOP}} = \sqrt{a_{11} + a_{22} + a_{55}}$

Esto corresponde al problema planteado en dos dimensiones y la σ_{HDOP} también es la DOP de la posición (σ_{PDOP}). Si se incorpora una tercera dimensión z , aparece la DOP vertical σ_{VDOP} y la DOP de la posición será $\sigma_{\text{PDOP}} = \sqrt{a_{11} + a_{22} + a_{33}}$ donde a_{33} corresponde al término de z en la ecuación 1.43 expandida a tres dimensiones.

1.6.2 CEP: Circular Error Probability

El CEP es una métrica que fue definida en su momento para aplicaciones militares y es una forma de determinar la precisión. Define el círculo que encierra el 50 % de las medidas de posición estática. Existen también CEP_{90} que corresponde al 90 % de las medidas y R_{95} que define el radio del círculo que cierra al 95 %. Matemáticamente,

$$\text{CEP} : \quad P(\|\mathbf{x}, \hat{\mathbf{x}}\| < d) = 0,5 \quad (1.55)$$

1.6.3 CRLB: Cramer Rao Lower Bound

En teoría de variables aleatorias y probabilidad, el CRLB es un valor que representa un límite inferior en la varianza de la estimación sin sesgo (da una idea de la dispersión en la densidad de probabilidad de la estimación). Esto quiere decir que **cada método** de estimación tendrá en esencia un límite de precisión. Por ejemplo en [12] se muestran los límites de precisión de la estimación de TOA sobre canales multicamino para la estimación de posición en redes celulares usando el CRLB. La figura 1.32 esta extraída de allí y podemos ver para la curva superior que en un ambiente suburbano para una relación señal-ruido de 10dBm, el CRLB es de aproximadamente 30mts; para una relación peor de 5dBm ese límite sube a 60mts.

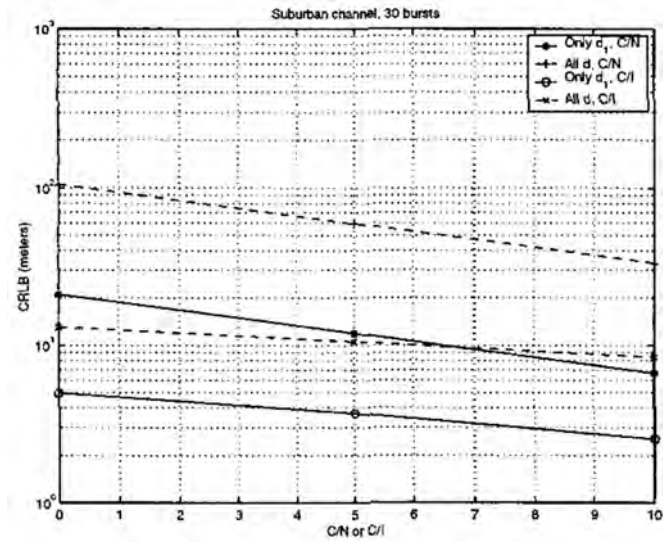


Figura 1.32: Figura extraída de [12]. Se muestra el CRLB para diferentes escenarios de simulación para un ambiente celular suburbano estimando la posición usando TOA.

Parameter	UWB	WASP	PLS
RF Frequency	5 GHz	5.8 GHz	2.4 GHz
RF Channel	500 MHz	125 MHz	60 MHz
Bandwidth (3 dB)	250 MHz	65 MHz	40 MHz
Pulse rise-time	3 ns (0.9 m)	13.5 ns (4 m)	25 ns (7.5 m)
Coverage Area	800 m ²	10000 m ²	1500 m ²

Figura 1.33: Resumen de los parámetros de la campaña de medición.

1.6.4 Por Ensayo

Una forma de evaluar la precisión de un esquema de localización más allá de las hojas de datos de un sensor o los límites teóricos del algoritmo, es a partir de una campaña de medición diseñada para obtenerla estadísticamente. A medida que una EM se aleja del transmisor, siendo que hagamos TOA u otra estrategia, como hemos visto la señal irá variando localmente hasta los 20dB.

Lo que cada campaña va a resaltar es la variabilidad controlando alguna de las variables y variando otras. Por ejemplo en [22] se utilizan varias bandas de transmisión y varias estrategias de medición del TOA (Ultra Wide Band, Wireless Ad Hoc System for Positioning y precision Location System). La tabla mostrada en la figura 1.33 extraída del trabajo muestra las características de la campaña de medición.

Sin ahondar en los detalles de cómo se mide, es interesante observar los resultados que se muestran en la figura 1.34. Allí se pueden ver las medidas de distancia, en línea continua la tendencia lineal y en discontinua representan un pulso temporal en mas y en menos respecto de la tendencia y que sirve como referencia al error correspondiente en la distancia. Los datos se utilizan para determinar el sesgo a distancia y la pendiente del modelo y la desviación estándar respecto de la tendencia. El modelo del error propuesto tendría los siguientes componentes:

$$\varepsilon_{\text{TOA}} = \varepsilon_{\text{TOA}}(\tau_{\text{pulse}}) + \varepsilon_{\text{path}}(R, N, f, BW) \quad (1.56)$$

donde $\varepsilon_{\text{TOA}}(\tau_{\text{pulse}})$ es el error debido a la determinación del pulso y $\varepsilon_{\text{path}}(R, f, BW)$ el error debido al camino y que depende de la distancia R , N el número de paredes que atraviesa la señal en NLOS, f la frecuencia de la portadora y BW el ancho de banda.

1.6.5 Por Simulación

Otra forma de evaluar la precisión es mediante la simulación. Si tomamos también un sistema basado en la medición del tiempo de arribo (TOA), tendremos un canal que va de la EM a la ER (denominado UPLINK) y otro al revés (DOWNLINK). Cada uno tendrá diferentes retardos debido al multicamino. En [6] se presenta un escenario de simulación como el mostrado en la figura 1.35 y los resultados en la figura 1.36. Los autores realizan una campaña de medición y lo comparan con la simulación propuesta y la posición verdadera para una EM que se mueve en un círculo de radio de 2 metros cuyo centro está a 10 metros de la ER.

Otra estrategia de simulación es analizar cómo funcionaría el método considerando un modelo para el retardo medio en exceso (recordar la sección 1.4.3.1). Si se supone que ambos enlaces están distribuidos log-normal (es una variable aleatoria cuyo logaritmo tiene una distribución gaussiana) a una distancia dada, el retardo se modela como:

$$\tau_m = K T_0 D^\epsilon \chi \quad (1.57)$$

donde K es una constante de proporcionalidad entre τ_m y el retardo medio en exceso que caracteriza al canal τ_{rms} , T_0 es el valor medio del τ_{rms} a una distancia D_0 , D la distancia entre la EM y la ER, ϵ una constante y χ una variable aleatoria tal que el $\ln(\chi)$ está distribuida normal con varianza σ . Para dar una idea de los valores de estos parámetros, la tabla 1.2 los muestra en valores típicos para ambientes urbanos y suburbanos.

Si se calcula el τ_m tanto para el enlace de subida como el de bajada, podemos calcular un promedio entre ambos,

$$n_i = \frac{\tau_{m,d} + \tau_{m,u}}{2} \quad (1.58)$$

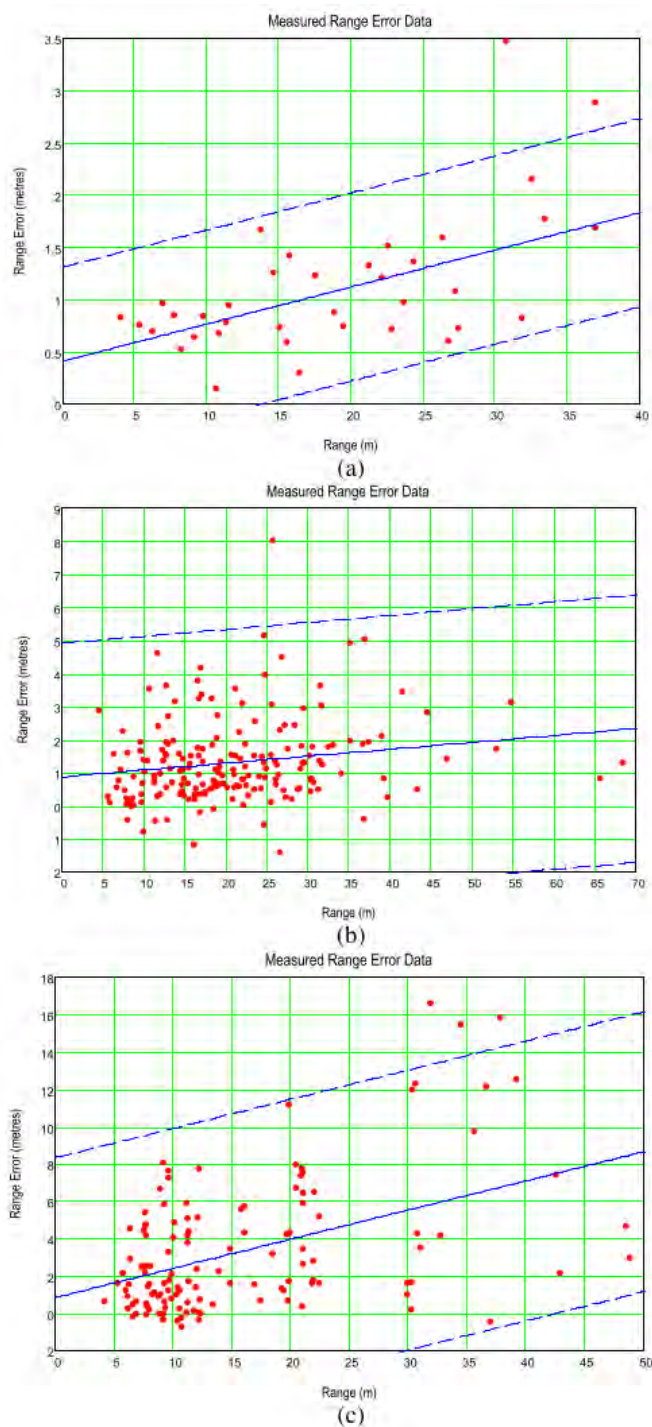


Figura 1.34: Mediciones y tendencias. En (a) para UWB, en (b) para Wasp y en (c) para PLS

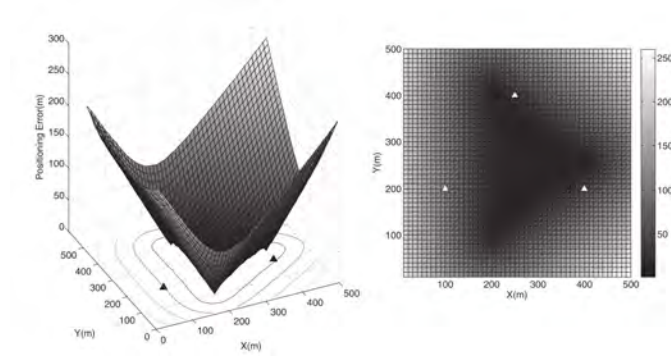


Figura 1.37: El mapa del error para el TOA de un modelo como el planteado en 1.57. Los triángulos representan las ubicaciones de las ER y los parámetros utilizados son $K = 0,5$, $T_0 = 0,28\mu s$, $\sigma = 2dB$ y $\epsilon = 0,5$ ([21]).

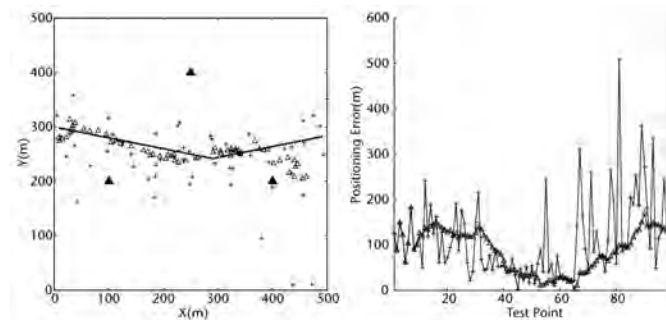


Figura 1.38: Simulación del movimiento de un móvil (línea continua) entre tres estaciones de referencia (triángulos), los parámetros utilizados son $K = 0,5$, $T_0 = 0,28\mu s$, $\sigma = 2dB$ y $\epsilon = 0,5$ ([21]). A la izquierda en cruces las posiciones calculadas y en triángulos el promedio móvil de 10 medidas. A la derecha el error

Es posible simular con esto el mapa del error con una condición de las posibles y listadas en la tabla 1.2. La figura 1.37 muestra un mapa del error donde los triángulos representan las ER y a medida que nos alejamos del centro geométrico de las ubicaciones de las ER donde el error es menor), este va creciendo rápidamente.

Otra simulación posible es analizar el funcionamiento del método en lugar del error. Por ejemplo la figura 1.38 muestra tres ER y usando TOA en para NLOS en un ambiente suburbano, se calcula la posición (a la izquierda) y el error de posición por punto o por promedio móvil para un móvil que se mueve según la línea continua.

1.7 MULTILATERACIÓN HIPERBÓLICA

En forma general la Multilateración Hiperbólica es también denominada ubicación por diferencia de rango. Si se usan los tiempos de llegada de una señal se denomina Diferencia del Tiempo de Arribo

EJERCICIO: El programa de Matlab cho2figo5_o6.m reproduce la figura 1.37 que es la figura 2.5 de [21].

EJERCICIO: El programa de Matlab cho2exo2.m reproduce la figura 1.38 que es el ejemplo 2.2 de [21].

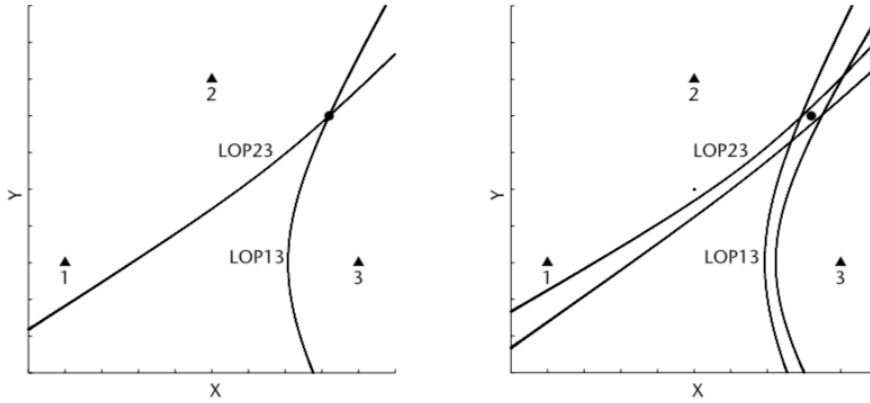


Figura 1.39: Las líneas de puntos posibles donde se encuentra un móvil (círculo negro) en el caso de diferencias de rango es una Hipérbola con focos en las ER.

(TDoA: Time Difference of Arrival en inglés). En este último caso las ER están sincronizadas y emiten un mensaje que es recibido por un móvil que no lo está. El móvil mide la diferencia de tiempo de llegada entre los mensajes de cada ER (tomando una como ancla) y se define una Línea de Posición Hiperbólica que tiene sus focos en las ER (figura 1.39).

Otro aspecto a resaltar es que el TDoA se puede plantear a partir de las medidas hechas por el móvil o en las ER. En lo que sigue consideraremos el primer caso. Si Δt_{ij} es el TDoA medida en el móvil, entonces

$$\Delta \hat{r}_{ij} = (d_i - d_j) = c \cdot (\Delta t_{ij} + n_{ij}), \quad (1.59)$$

donde $(d_i - d_j)$ es la distancia entre las ER i -ésima y j -ésima y el móvil, c es la velocidad de la luz y n_{ij} es un retardo aleatorio que representa el error del tiempo medido. A partir de aquí podemos construir un vector de medidas:

$$\Delta \hat{\mathbf{R}} = \begin{bmatrix} c \cdot (\Delta t_{12} + n_{12}) \\ c \cdot (\Delta t_{13} + n_{13}) \\ \vdots \\ c \cdot (\Delta t_{1M} + n_{1M}) \\ c \cdot (\Delta t_{23} + n_{23}) \\ \vdots \\ c \cdot (\Delta t_{M-1,M} + n_{M-1,M}) \end{bmatrix} = c \cdot (\Delta \mathbf{T} + \mathbf{N}) \quad (1.60)$$

donde $\Delta \mathbf{T}$ es el vector de diferencia de tiempos reales Δt_{ij} y \mathbf{N} el vector que de cada una de diferencias.

Por otro lado, las diferencias de tiempo están relacionadas con las posiciones de las ER y el móvil,

$$c \cdot \Delta t_{ij} = \sqrt{(x - x_i)^2 + (y - y_i)^2} - \sqrt{(x - x_j)^2 + (y - y_j)^2}, \quad (1.61)$$

a partir de esta ecuación y definiendo a las coordenadas del móvil como $\mathbf{s} = [x, y]^T$, se obtiene el vector $\mathbf{F}(\mathbf{s})$

$$\mathbf{F}(\mathbf{s}) = \begin{bmatrix} \sqrt{(x - x_1)^2 + (y - y_1)^2} - \sqrt{(x - x_2)^2 + (y - y_2)^2} \\ \sqrt{(x - x_1)^2 + (y - y_1)^2} - \sqrt{(x - x_3)^2 + (y - y_3)^2} \\ \vdots \\ \sqrt{(x - x_1)^2 + (y - y_1)^2} - \sqrt{(x - x_M)^2 + (y - y_M)^2} \\ \sqrt{(x - x_2)^2 + (y - y_2)^2} - \sqrt{(x - x_3)^2 + (y - y_3)^2} \\ \vdots \\ \sqrt{(x - x_{M-1})^2 + (y - y_{M-1})^2} - \sqrt{(x - x_M)^2 + (y - y_M)^2} \end{bmatrix} \quad (1.62)$$

y combinando las ecs. 1.60 y 1.62,

$$\mathbf{N} = \Delta \hat{\mathbf{R}} - \mathbf{F}(\mathbf{s}) \quad (1.63)$$

La solución a esta ecuación no es cerrada debido a la no linealidad de $\mathbf{F}(\mathbf{s})$, el problema a minimizar puede ser planteado como:

$$\hat{\mathbf{s}} = \underset{\mathbf{s}}{\operatorname{argmin}} \{ (\Delta \hat{\mathbf{R}} - \mathbf{F}(\mathbf{s}))^T (\Delta \hat{\mathbf{R}} - \mathbf{F}(\mathbf{s})) \}; \quad (1.64)$$

esto es minimizar el error al cuadrado $\mathbf{N}^T \mathbf{N}$.

Resolver este problema implica aplicar métodos aproximados e iterativos. Por ejemplo en los sistemas LORAN, medidos los TDoA entre ER, se accede a unos mapas para obtener la solución. Los métodos numéricos pueden ser tan simples como el de búsqueda por gradiente que es intensivo en cómputo y la solución es aproximada porque se basa en la linealización de las ecuaciones. El método de secantes que usamos para ToA no es posible aplicarlo en este caso porque no podemos aislar la posición del móvil en las matrices.

Si realizamos una expansión en series de Taylor alrededor de una estimación inicial $\hat{\mathbf{s}}_0 = [x_0, y_0]^T$ y descartando los términos no lineales, obtenemos la siguiente aproximación para $\mathbf{F}(\mathbf{s})$,

$$\mathbf{F}(\mathbf{s}) \approx \mathbf{F}(\hat{\mathbf{s}}_0) + \frac{\partial \mathbf{F}(\mathbf{s})}{\partial x} \Big|_{\mathbf{s}=\hat{\mathbf{s}}_0} (x - x_0) + \frac{\partial \mathbf{F}(\mathbf{s})}{\partial y} \Big|_{\mathbf{s}=\hat{\mathbf{s}}_0} (y - y_0) \quad (1.65)$$

que se puede expresar en forma matricial como,

$$\mathbf{F}(\mathbf{s}) \approx \mathbf{F}(\hat{\mathbf{s}}_0) + \mathbf{H}(\mathbf{s} - \hat{\mathbf{s}}_0) \quad (1.66)$$

donde

$$\mathbf{H} = \left[\begin{array}{cc} \frac{\partial \mathbf{F}_1(\mathbf{s})}{\partial x} & \frac{\partial \mathbf{F}_1(\mathbf{s})}{\partial y} \\ \frac{\partial \mathbf{F}_2(\mathbf{s})}{\partial x} & \frac{\partial \mathbf{F}_2(\mathbf{s})}{\partial y} \\ \vdots & \vdots \\ \frac{\partial \mathbf{F}_M(\mathbf{s})}{\partial x} & \frac{\partial \mathbf{F}_M(\mathbf{s})}{\partial y} \end{array} \right] \Big|_{\mathbf{s}=\hat{\mathbf{s}}_0} \quad (1.67)$$

Reemplazamos la ecuación 1.66 en la ecuación 1.63,

$$\begin{aligned} \mathbf{N} &= \Delta \hat{\mathbf{R}} - \mathbf{F}(\hat{\mathbf{s}}_0) - \mathbf{H}(\mathbf{s} - \hat{\mathbf{s}}_0) \\ \mathbf{N} &= (\Delta \hat{\mathbf{R}} - \mathbf{F}(\hat{\mathbf{s}}_0) + \underbrace{\mathbf{H} \cdot \hat{\mathbf{s}}_0}_{\Delta \tilde{\mathbf{R}}}) - \mathbf{H} \cdot \mathbf{s} \\ \mathbf{N} &= \Delta \tilde{\mathbf{R}} - \mathbf{H} \cdot \mathbf{s} \end{aligned} \quad (1.68)$$

Reemplazando ahora 1.68 en la ecuación 1.64

$$\hat{\mathbf{s}} = \underset{\mathbf{s}}{\operatorname{argmin}} \{ (\Delta \tilde{\mathbf{R}} - \mathbf{H} \cdot \mathbf{s})^T (\Delta \tilde{\mathbf{R}} - \mathbf{H} \cdot \mathbf{s}) \}, \quad (1.69)$$

La solución de mínimos cuadrados de 1.69 será,

$$\hat{\mathbf{s}} = (\mathbf{H}^T \mathbf{H})^{-1} \mathbf{H}^T \Delta \tilde{\mathbf{R}}, \quad (1.70)$$

donde $\mathbf{H} \rightarrow \mathbf{H}(\mathbf{s}_0)$, expandiendo:

$$\begin{aligned} \hat{\mathbf{s}} &= (\mathbf{H}^T \mathbf{H})^{-1} \mathbf{H}^T (\Delta \hat{\mathbf{R}} - \mathbf{F}(\hat{\mathbf{s}}_0) + \mathbf{H} \cdot \hat{\mathbf{s}}_0) \\ \hat{\mathbf{s}} &= \hat{\mathbf{s}}_0 + (\mathbf{H}^T \mathbf{H})^{-1} \mathbf{H}^T (\Delta \hat{\mathbf{R}} - \mathbf{F}(\hat{\mathbf{s}}_0)) \end{aligned} \quad (1.71)$$

La ecuación 1.71 representa una solución a partir de una posición inicial $\hat{\mathbf{s}}_0$. Con esta expresión se puede construir un método iterativo reescribiéndola como:

$$\hat{\mathbf{s}}_{k+1} = \hat{\mathbf{s}}_k + (\mathbf{H}^T \mathbf{H})^{-1} \mathbf{H}^T (\Delta \hat{\mathbf{R}} - \mathbf{F}(\hat{\mathbf{s}}_k)), \quad (1.72)$$

y donde $\mathbf{H} \rightarrow \mathbf{H}(\mathbf{s}_k)$. Este método no es complejo y a partir de una estimación inicial da una solución como la que podemos observar en la figura 1.40. Allí las ER están representadas con círculos y a partir de una condición inicial (cruz) y una posición real del móvil (círculo negro), los círculos blancos representan la solución de este algoritmo. Las diferentes soluciones tiene que ver con diferentes medidas ruidosas de los tiempos.

Luego se puede construir un mapa de error como se muestra en la figura 1.41. Como se puede observar el error es bajo cuando el objetivo se encuentra dentro del polígono definido por las ER. A medida que nos alejamos las hipérbolas se vuelven líneas *casi* paralelas y se cortan en un ángulo muy pequeño.

EJERCICIO: El programa de Matlab `cho2fig11.m` reproduce la figura 1.40. Se puede jugar con las condiciones iniciales para ver el funcionamiento del algoritmo

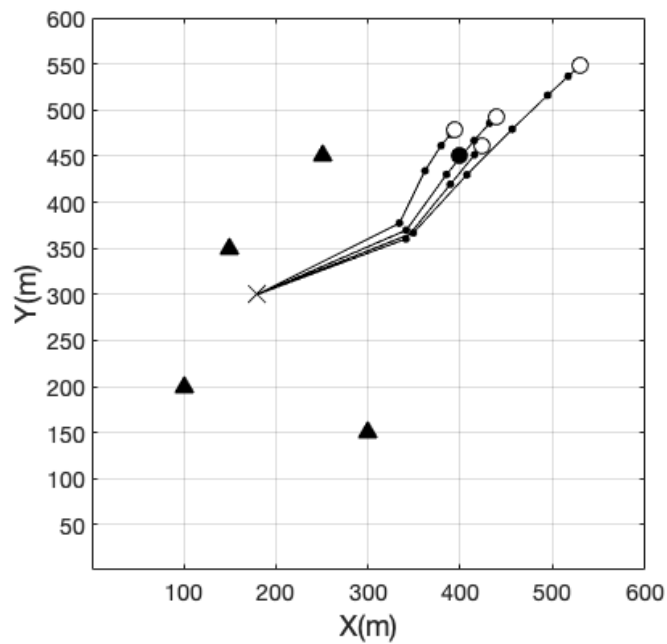


Figura 1.40: Simulación del algoritmo presentado en 1.72. Las ER están representadas con círculos, la condición inicial con una cruz, la posición real del móvil con un círculo negro, y los círculos blancos representan la diferentes soluciones de este algoritmo para medidas ruidosas de tiempos.

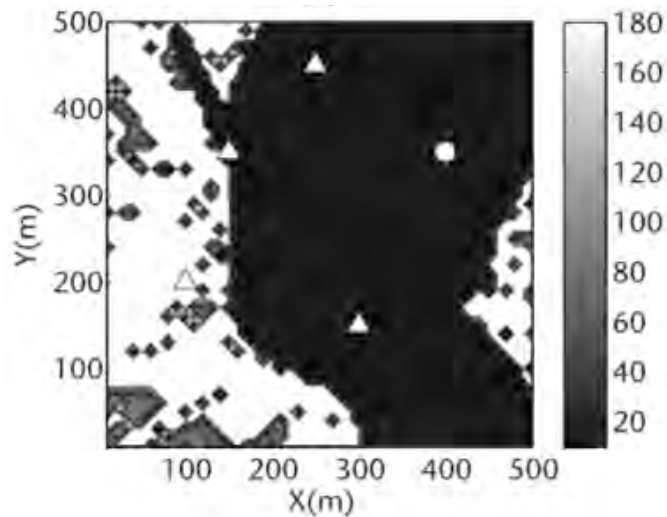


Figura 1.41: Mapa de error para el algoritmo 1.72. El error es mayor a medida que nos alejamos del polígono que definen las ER.

1.7.1 Filtro de Kalman para TDoA

Es posible plantear una solución basada en el filtro de Kalman como se presentó en la sección 1.5.1.2. La matriz de transición de estados pierde la fila y columna correspondiente al estado del sego b. Lo que si cambia en forma fundamental es la matriz de observación dada por la ecuación 1.62 y el Jacobiano que en este caso será:

$$\nabla \mathbf{h}_k = \begin{bmatrix} \frac{x_k - x_1}{\|x_k - x_1\|} & \frac{y_k - y_1}{\|y_k - y_1\|} & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots \\ \frac{x_k - x_4}{\|x_k - x_4\|} & \frac{y_k - y_4}{\|y_k - y_4\|} & 0 & 0 \end{bmatrix} \quad (1.73)$$

1.7.2 Método de Lavenberg-Marquardt para TDoA

Este método es iterativo y su objetivo es minimizar la función de costo

$$\hat{\mathbf{S}} = \underset{\mathbf{s}}{\operatorname{argmin}} \{ (\Delta \hat{\mathbf{R}} - \mathbf{F}(\mathbf{s}))^T (\Delta \hat{\mathbf{R}} - \mathbf{F}(\mathbf{s})) \}, \quad (1.74)$$

mejorando el problema de la convergencia. En este algoritmo la posición del móvil se obtiene en forma iterativa con la siguiente ecuación:

$$\mathbf{s}_{k+1} = \mathbf{s}_k + (\mathbf{A}_k + \lambda^k \mathbf{I})^{-1} \cdot \mathbf{g}_k, \quad (1.75)$$

donde

$$\mathbf{A}_k = [\phi_k^T(\mathbf{x}_k) \Sigma_n^{-1} \phi_k(\mathbf{x}_k)], \quad (1.76)$$

$$\phi_k = \begin{bmatrix} \frac{(x_k - x_2)}{d_{2,k}} - \frac{(x_k - x_1)}{d_{1,k}} & \frac{(y_k - y_2)}{d_{2,k}} - \frac{(y_k - y_1)}{d_{1,k}} \\ \vdots & \vdots \\ \frac{(x_k - x_N)}{d_{N,k}} - \frac{(x_k - x_1)}{d_{1,k}} & \frac{(y_k - y_N)}{d_{N,k}} - \frac{(y_k - y_1)}{d_{1,k}} \end{bmatrix} \quad (1.77)$$

y $d_{i,k} = \sqrt{(x_i - x_k)^2 + (y_i - y_k)^2}$ es la distancia entre la ER i -ésima y la estimación actual del móvil. Además,

$$\Sigma_k = \begin{bmatrix} \sigma_{21} & 0 & \dots & 0 \\ 0 & \sigma_{31} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sigma_{N1} \end{bmatrix}. \quad (1.78)$$

modela con esta varianza la incertidumbre en las medidas de las diferencia de tiempos. Finalmente,

$$\mathbf{g}_k = \phi_k^T \Sigma_n^{-1} (\mathbf{d} - \hat{\mathbf{d}}_k) \quad (1.79)$$

donde \mathbf{d} es el vector de distancias medidas

$$\mathbf{d} = [d_{2,1}, d_{3,1}, \dots, d_{N,1}] \quad (1.80)$$

y $\hat{\mathbf{d}}_k$ es el vector de distancias estimadas

$$\hat{\mathbf{d}}_k = [\hat{d}_{2,1}, \hat{d}_{3,1}, \dots, \hat{d}_{N,1}] \quad (1.81)$$

El factor λ^k asegura que la matriz \mathbf{A}_k sea invertible, lo que hace al algoritmo mas robusto. Una elección posible para ese parámetro es:

$$\lambda^k = \tau \cdot \max\{\mathbf{A}_k / i, i\} \quad (1.82)$$

es decir el máximo valor de la diagonal multiplicado por un factor τ que va a tener influencia en el paso de la iteración. Como ejemplo, para ER separadas por algunos kilometros (como las redes celulares), un posible valor de τ es cinco.

Una estrategia que hace el algoritmo subóptimo pero mas estable, es definiendo un parámetro ρ como sigue,

$$\rho^k = \frac{\varepsilon(\mathbf{s}_k) - \varepsilon(\mathbf{s}_{k-1})}{\mathbf{h}_k^T (\lambda_k \mathbf{h}_k + \mathbf{g}_k)} \quad (1.83)$$

donde

$$\varepsilon(\mathbf{s}_k) = [\mathbf{d} - \hat{\mathbf{d}}_k]^T \Sigma_n^{-1} [\mathbf{d} - \hat{\mathbf{d}}_k] \quad (1.84)$$

y

$$\mathbf{h}_k = (\mathbf{A}_k + \lambda^k \mathbf{I})^{-1} \cdot \mathbf{g}_k \quad (1.85)$$

El algoritmo analiza el signo de ρ . Así, si $\rho > 0$, se utiliza el algoritmo normal ya presentado. Sin embargo, si $\rho \leq 0$ se preservan los valores de las matrices \mathbf{A}_k y \mathbf{g}_k , i.e.

$$\mathbf{A}_{k+1} = \mathbf{A}_k \quad \text{y} \quad \mathbf{g}_{k+1} = \mathbf{g}_k \quad (1.86)$$

Además de estas matrices, el parámetro λ_k también puede ser actualizado de la siguiente manera:

$$\lambda_{k+1} = \begin{cases} \lambda_k \cdot \max\{\frac{1}{3}, 1 - (2\rho_k - 1)\} & \text{si } \rho > 0 \\ \lambda_k \cdot \nu_k & \text{si } \rho \leq 0 \end{cases} \quad (1.87)$$

donde

$$\nu_{k+1} = \begin{cases} 2 & \text{si } \rho > 0 \\ 2 \cdot \nu_k & \text{si } \rho \leq 0 \end{cases} \quad (1.88)$$

considerando a $\nu_0 = 2$. Estos algoritmos no son complejos de implementar solo que requiere el cálculo de la distancia entre las ER y la estimación de la posición del móvil (ec. 1.81).

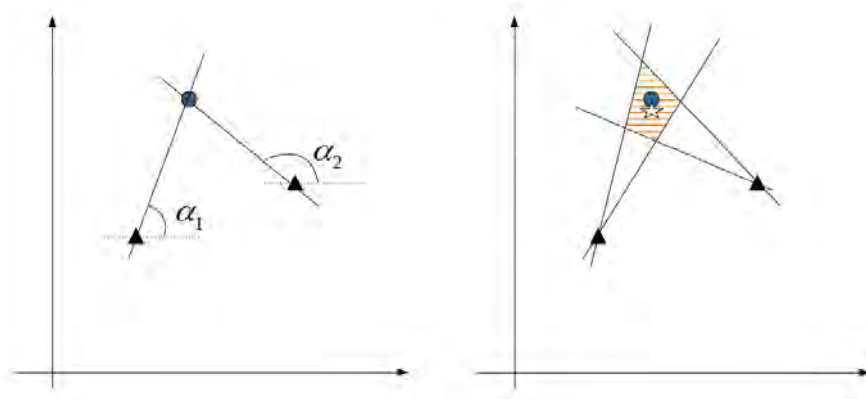


Figura 1.42: Posicionamiento AoA en condiciones de Línea de Vista (izq.) y NLOS (de.). En el primer caso todas las líneas de posición se intersectan en un punto, en el segundo define una región de confianza. El círculo azul indica la verdadera ubicación del móvil, la estrella blanca indica el centroide de la región de confianza. Los triángulos negros indican las verdaderas ubicaciones de las ER.

1.8 MULTILATERACIÓN ANGULAR

También denominado Multiangulación, usa información de ángulo para determinar la posición. En el caso de señales de RF se usa el Angulo de Arribo (AoA: Angle of Arrival en inglés) de la señal. La visión monocular es otro ejemplo de información angular. La intersección de dos ángulos es suficiente para resolver la posición en dos dimensiones. Cuando las medidas son con error, mayor cantidad de ER reduce el área de confianza donde se encuentra el móvil (figura 1.42)

En el caso de que no tengamos error en las mediciones, si la posición de la ER i -ésima en metros es $[x_i, y_i]^T$, α_i el ángulo en radianes medido por estas, la posición estimada será:

$$\begin{aligned}\hat{x} &= \frac{[(x_2 \tan \alpha_2 - x_1 \tan \alpha_1) - (y_2 - y_1)]}{(\tan \alpha_2 - \tan \alpha_1)} \\ \hat{y} &= \frac{[(x_2 - x_1) \tan \alpha_1 \tan \alpha_2 + (y_1 \tan \alpha_2 - y_2 \tan \alpha_1)]}{(\tan \alpha_2 - \tan \alpha_1)}\end{aligned}\quad (1.89)$$

De la ecuación 1.89 se puede apreciar que si $\alpha_1 = \alpha_2$, los denominadores son cero y no se puede determinar la posición. Cuestión lógica ya que las rectas no se cortarían en un único punto. En el caso del uso de esta técnica en RF, el multicamino (figura 1.43) puede generar problemas de indeterminación. Como se observa en la figura 1.44 tenemos cuatro casos posibles. El primero 1.44(a) es la situación deseable en la que todas las medidas se cruzan en una única zona de confianza y es donde se encuentra el móvil. El caso 1.44(b) representa una situación típica de multicamino; las medidas se cruzan en una única zona pero esta no coincide con la posición del móvil. En 1.44(c)

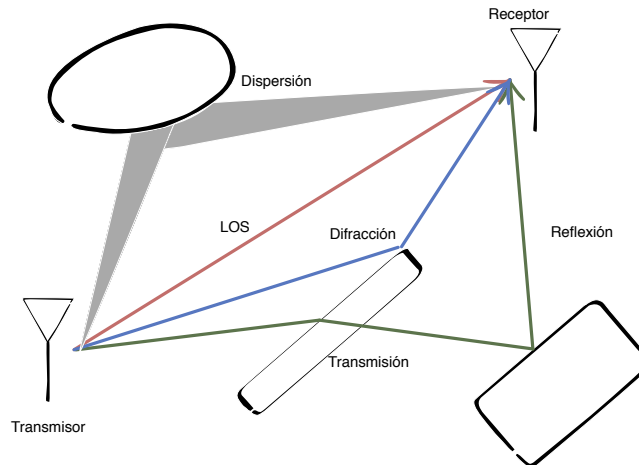


Figura 1.43: La señal de RF que parte de un transmisor llega al receptor por varios caminos y diferentes medios de propagación como muestra la figura. Esto implica que la señal llega al receptor por varios caminos y por lo tanto con diferentes ángulos.

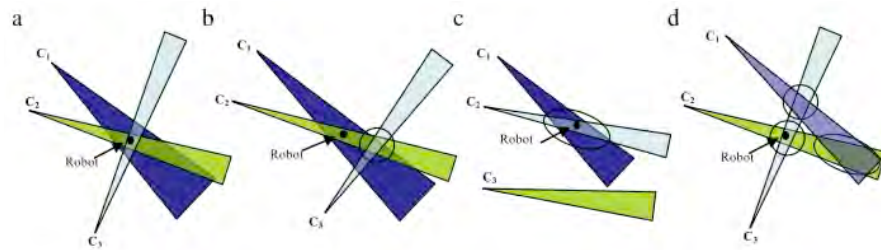


Figura 1.44: El multicamino y la falta de línea de vista entre transmisor y receptor provoca que no siempre se pueda obtener una zona de confianza que incluya la posición real del móvil.

se ve un caso donde por ejemplo tenemos dos transmisores con LOS y uno que no; este caso es simple de detectar en un posible algoritmo de detección de *outliers*. Finalmente, en 1.44(d) se obtienen tres zonas de confianza y alguna de ellas puede contener a la solución.

Existen estrategias para determinar estos *outliers*. Varias de estas utilizan las propiedades estadística de la señal. Estas junto a estrategias de medición del ángulo, los veremos en detalle en la próxima sección 1.9.

1.8.1 Formas de resolución de posición usando AoA

Presentaremos a continuación algunos algoritmos que resuelven la posición a partir de las medidas de ángulo. Para esto definiremos,

$$\mathbf{s}^0 = [x^0, y^0]^T \quad \text{son las coordenadas del móvil} \quad (1.90)$$

$$\mathbf{b}_i^0 = [x_i^0, y_i^0]^T \quad \text{son las coordenadas de la ER } i \quad (1.91)$$

En ausencia de ruido, el ángulo que se mide entre el móvil y la ER i -ésima será

$$\alpha_i^0 = \tan^{-1} \frac{y^0 - y_i^0}{x^0 - x_i^0} \quad (1.92)$$

Con esta mediciones construimos un vector de mediciones que tendrán asociado un ruido de medición:

$$\boldsymbol{\alpha} = [\alpha_1, \alpha_2, \dots, \alpha_M] = \boldsymbol{\alpha}^0 + \mathbf{n} \quad (1.93)$$

donde \mathbf{n} es el error de medida del ángulo modelado con una distribución Gaussiana de media cero y matriz de covarianza $\mathbf{C}_{\alpha\alpha}$. Si la posición de las ER no es perfectamente conocida,

$$\mathbf{b}_i = \mathbf{b}_i^0 + \mathbf{v} \quad \text{donde } \mathbf{b}_i = [b_1, b_2, \dots, b_M] \quad (1.94)$$

donde \mathbf{v} es el error en la posición de las ER que se supone Gaussiano de media cero y cuya matriz de covarianza es

$$\mathbf{C}_{xx} = \text{blockdiag}[\Psi_{b_1}, \Psi_{b_2}, \dots, \Psi_{b_M}] \quad (1.95)$$

Se supone además que \mathbf{n} y \mathbf{v} son independientes (aunque no siempre debe ser así, es el caso donde se esté estimando no solo la posición del móvil sino también las de las ER)

1.8.1.1 AoA por Maximum Likelihood (ML) [27]

Llamamos al parámetro de interés a ser estimado

$$\boldsymbol{\Theta} = [\boldsymbol{\Theta}_1^T, \boldsymbol{\Theta}_2^T]^T = [\mathbf{s}^{0T}, \mathbf{b}_i^{0T}]^T = [x^0, y^0, x_1^0, y_1^0, \dots, x_M^0, y_M^0]^T \quad (1.96)$$

y al vector de observaciones

$$\boldsymbol{\beta} = [\boldsymbol{\alpha}^T, \mathbf{b}^T]^T \quad (1.97)$$

El problema que se debe solucionar planteando máxima verosimilitud (maximum likelihood)

$$\hat{\boldsymbol{\Theta}}_{1,ML} = \underset{\boldsymbol{\Theta}_1}{\text{argmin}} (\boldsymbol{\alpha} - \mathbf{g}_1(\boldsymbol{\Theta}_1))^T \mathbf{C}_{\alpha\alpha}^{-1} (\boldsymbol{\alpha} - \mathbf{g}_1(\boldsymbol{\Theta}_1)) \quad (1.98)$$

donde

$$\mathbf{g}_1(\boldsymbol{\Theta}_1) = [g_{1,1}(\boldsymbol{\Theta}_1), g_{1,2}(\boldsymbol{\Theta}_1), \dots, g_{1,M}(\boldsymbol{\Theta}_1)] \quad (1.99)$$

y

$$g_{1,i}(\boldsymbol{\Theta}_1) = \tan^{-1} \frac{y - y_i}{x - x_i} \quad (1.100)$$

La resolución de este problema puede plantearse por el método de Gauss-Newton,

$$\Theta_1^{k+1} = \Theta_1^k + (\mathbf{H}_{1k}^T \mathbf{C}_{\alpha\alpha}^{-1} \mathbf{H}_{1k})^{-1} \mathbf{H}_{1k}^T \mathbf{C}_{\alpha\alpha}^{-1} (\alpha - \mathbf{g}_1(\Theta_1^k)) \quad (1.101)$$

donde

$$\mathbf{H}_{1k}^T = \frac{\partial \mathbf{g}_1(\Theta_1)}{\partial \Theta_1} \Big|_{\Theta_1 = \Theta_1^k} \quad (1.102)$$

1.8.1.2 AoA por Mínimos Cuadrados (LS) [27]

Como se dijo, la medición del ángulo se relaciona idealmente con el móvil y las ER por,

$$\alpha_i^0 = \tan^{-1} \frac{y^0 - y_i^0}{x^0 - x_i^0} \quad (1.103)$$

Podemos trabajar esta ecuación para encontrar una solución lineal y con forma cerrada. Para esto, trabajamos sobre la ecuación 1.103

$$\tan \alpha_i^0 = \frac{\sin(\alpha_i^0)}{\cos(\alpha_i^0)} = \frac{y^0 - y_i^0}{x^0 - x_i^0} \quad (1.104)$$

Trabajando sobre los dos últimos términos,

$$\underbrace{x^0 \sin(\alpha_i^0) - y^0 \cos(\alpha_i^0)}_{\mathbf{G}\Theta_1} = \underbrace{x_i^0 \sin(\alpha_i^0) - y_i^0 \cos(\alpha_i^0)}_{\mathbf{h}} \quad (1.105)$$

donde

$$\mathbf{G} = \begin{bmatrix} \sin(\alpha_1^0) & -\cos(\alpha_1^0) \\ \vdots & \vdots \\ \sin(\alpha_M^0) & -\cos(\alpha_M^0) \end{bmatrix}. \quad (1.106)$$

y

$$\mathbf{h} = \begin{bmatrix} x_1^0 \sin(\alpha_1^0) - y_1^0 \cos(\alpha_1^0) \\ \vdots \\ x_M^0 \sin(\alpha_M^0) - y_M^0 \cos(\alpha_M^0) \end{bmatrix}. \quad (1.107)$$

La solución LS, en forma cerrada será:

$$\Theta_{1,LS} = (\mathbf{G}^T \mathbf{G})^{-1} \mathbf{G}^T \mathbf{h} \quad (1.108)$$

Aunque esta solución es cerrada puede resultar en una no muy buena estimación. Y esto tiene que ver con el ruido que posee cada medición de ángulo poniéndolo en forma explícita en la ecuación 1.105. Por eso, una solución pesada será:

$$\Theta_{1,WLS} = (\mathbf{G}^T \mathbf{G})^{-1} \mathbf{G}^T \mathbf{W} \mathbf{h} \quad (1.109)$$

donde

$$\mathbf{W} = (\mathbf{B}_1 \mathbf{C}_{\alpha\alpha} \mathbf{B}_1^T + \mathbf{C}_{\alpha\alpha} \mathbf{D}_2 \mathbf{C}_{xx} \mathbf{D}_2^T + \mathbf{D}_1 \mathbf{C}_{xx} \mathbf{D}_1^T)^{-1} \quad (1.110)$$

En esta ecuación definimos a

$$\mathbf{B}_1 = \text{blockdiag}[d_1, d_2, \dots, d_M] \quad (1.111)$$

$$\mathbf{D}_1 = \text{blockdiag}[g_1^T, g_2^T, \dots, g_M^T] \quad (1.112)$$

$$\mathbf{D}_2 = \text{blockdiag}[\dot{g}_1^T, \dot{g}_2^T, \dots, \dot{g}_M^T] \quad (1.113)$$

En estas matrices,

$$d_i = \|\mathbf{s}^0 - \mathbf{b}_i\| \quad (1.114)$$

es la distancia entre la posición real del móvil y las ER i -ésima. Como la posición del móvil no la tenemos, se hace una estimación de \mathbf{s}^0 con $\mathbf{W} = \mathbf{I}_{M \times M}$, y con esta estimación se calcula 1.114 y el peso para refinar la estimación en dos pasos. Además debemos definir,

$$g_i = [\sin(\alpha_1^0), -\cos(\alpha_1^0)]^T \quad (1.115)$$

$$\dot{g}_i = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} g_i \quad (1.116)$$

1.8.1.3 AoA Total Least Square (TLS) [27]

Este algoritmo es una extensión del algoritmo de LS y considera errores en las medidas en la ubicación de las ER. La solución será:

$$\Theta_{1,TLS} = (\mathbf{G}^T \mathbf{G} - \sigma_s^2 \mathbf{I})^{-1} \mathbf{G}^T \mathbf{h} \quad (1.117)$$

donde σ_s es el valor singular mas chico de la matriz construida como $[\mathbf{G}\mathbf{h}]$. Recordemos que la descomposición en valores singulares, generaliza para matrices no cuadradas el concepto de autovalores y autovectores. Si una matriz M tiene tamaño $m \times n$, la descomposición será:

$$M_{m \times n} = U_{m \times m} \Sigma_{m \times n} V_{n \times n}^* \quad (1.118)$$

El valor σ_s de nuestro algoritmo será el menor valor de la diagonal principal de Σ .

La figura 1.45 muestra el desempeño de los algoritmos presentados a partir de simulaciones. El móvil obtiene M medidas de ángulo de estaciones de referencia, $\mathbf{C}_{xx} = 0,1^2 \mathbf{I}_{2M \times M}$, $\mathbf{C}_{\alpha\alpha} = \sigma_\alpha^2 \mathbf{I}_{M \times M}$ siendo σ_α el eje x . El eje y muestra el error medio cuadrático calculado promediando 1000 iteraciones.

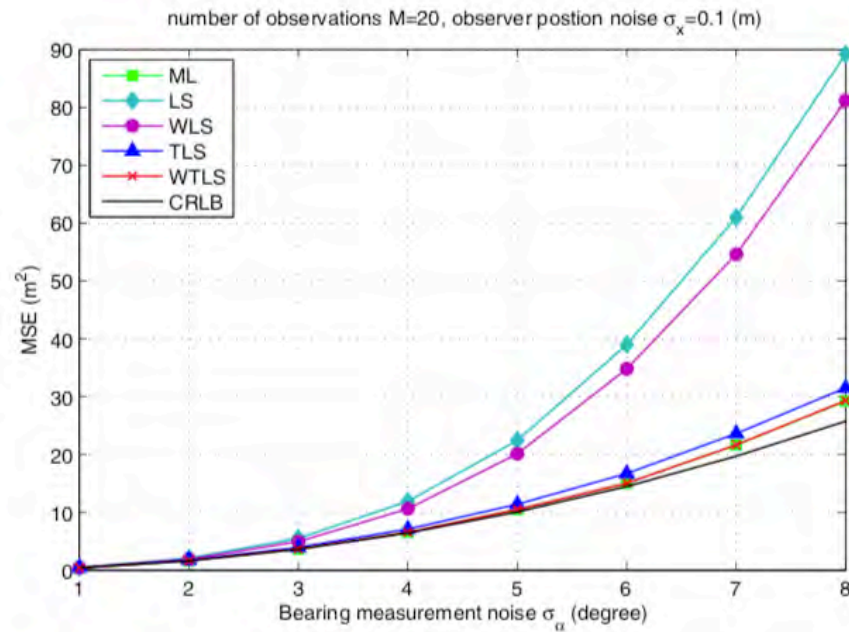


Figura 1.45: Desempeño de los algoritmos presentados hasta acá comparados con el CRLB. El eje x muestra la variación del ruido de ángulo en la medida y en el eje y el error medio cuadrático para cada algoritmo. Se supone que hay 20 ER y cada usa se conoce con un error de 0.1m (extraído de [27]).

1.8.1.4 Estimador Bayesiano pseudo-lineal

Fundamentalmente, la localización por AoA es un problema de estimación no lineal relacionado con las mediciones del ángulo y la posición del móvil. Los algoritmos que ya presentamos no tienen una forma cerrada y la solución es iterativa (estimador ML). Para superar este problema, se han presentado dos estimadores algebraicos de forma cerrada. El principio detrás de estos estimadores algebraicos es reorganizar las ecuaciones de medición de ángulo no lineal en un conjunto de ecuaciones que son lineales en la posición de la fuente desconocida. Esto permite que la posición de la fuente se determine por mínimos cuadrados y la solución se las englobe como estimadores *pseudo-lineales*. El problema más importante de estos estimadores es que sufren de sesgo por correlaciones entre las medidas y el vector pseudo-lineal de ruido.

En muchas aplicaciones prácticas, existe alguna información previa sobre la posición del móvil. Por ejemplo por datos disponibles de otros sensores. Una forma de considerar esta información previa es utilizando herramientas Bayesianas. La posición del móvil se modela como una variable aleatoria desconocida con una función de densidad de probabilidad previa dada, y el objetivo es estimar una

realización particular de esta densidad condicionada a las mediciones hechas desde las ER.

Recordemos que la posición del móvil es \mathbf{s}^0 , la de las estaciones de referencia \mathbf{b}_i^0 con $i \in [1, 2, \dots, M]$ y α_i^0 el ángulo verdadero medido por la ER i -ésima. En este algoritmo, la posición del móvil la modelamos como una variable aleatoria \mathbf{s} con una función de densidad de probabilidad Gaussiana, es decir, $p(\mathbf{s}) \propto \mathcal{N}(\bar{\mathbf{s}}, \mathbf{C}_{ss})$, donde $\bar{\mathbf{s}}$ es la media y \mathbf{C}_{ss} es la matriz de covarianza de \mathbf{s} .

La medida de ángulo obtenida por la ER k será,

$$\alpha_k = \alpha_k^0 + n_k, \quad \alpha_k^0 = \tan^{-1} \frac{y^0 - y_k^0}{x^0 - x_k^0} \quad (1.119)$$

donde $p(n_k) \propto \mathcal{N}(0, \sigma_k^2)$ es un ruido Gaussiano en las medidas de ángulo, de media cero y varianza σ_k^2 , es decir que puede ser diferente para cada ER. Suponemos además que este ruido es independiente entre las ER y entre estas y la posición del móvil. Puede incluso modelarse que σ_k^2 varía con la posición relativa entre el móvil y las ER, con la condición que se conoce a priori. Si usamos información de ángulo de arribo de señales de RF esto puede modelar el patrón variable con el ángulo de la antena.

Recordemos que la solución Bayesiana es la siguiente:

$$p(\mathbf{s}|\boldsymbol{\alpha}) = \frac{p(\boldsymbol{\alpha}|\mathbf{s}) \cdot p(\mathbf{s}_0)}{p(\boldsymbol{\alpha})} \quad (1.120)$$

donde $p(\mathbf{s}|\boldsymbol{\alpha})$ es la estimación de la densidad de probabilidad de \mathbf{s} condicionada a las mediciones $\boldsymbol{\alpha}$ hechas desde cada una de las ER (denominada densidad de probabilidad a posteriori), $p(\boldsymbol{\alpha}|\mathbf{s})$ es la función de verosimilitud que modela los sensores, $p(\mathbf{s}_0)$ es la información a priori que se tiene de la posición del móvil y $p(\boldsymbol{\alpha})$ es un factor de normalización para que la estimación sea una densidad de probabilidad.

La función de verosimilitud puede definirse como una función exponencial como sigue:

$$p(\boldsymbol{\alpha}|\mathbf{s}) = \frac{1}{(2\pi)^{M/2} |\mathbf{C}_{MM}|^{1/2}} \times \exp \left\{ -\frac{1}{2} (\boldsymbol{\alpha} - \hat{\boldsymbol{\alpha}}(\mathbf{s}))^T \mathbf{C}_{MM}^{-1} (\boldsymbol{\alpha} - \hat{\boldsymbol{\alpha}}(\mathbf{s})) \right\} \quad (1.121)$$

donde $\boldsymbol{\alpha} = [\alpha_1, \alpha_2, \dots, \alpha_M]^T$ es el vector de medidas de ángulos de $N \times 1$ y $\hat{\boldsymbol{\alpha}}(\mathbf{s}) = [\hat{\alpha}_1(\mathbf{s}), \hat{\alpha}_2(\mathbf{s}), \dots, \hat{\alpha}_M(\mathbf{s})]$ es el vector de ángulos entre las ER y la posición del móvil. Además, $|\cdot|$ representa el determinante de la matriz, $\mathbf{C}_{MM} = \text{diag}[\sigma_1, \sigma_2, \dots, \sigma_M]$ es la matriz de covarianza de $N \times N$ del ruido de ángulo, es decir del vector $\mathbf{n} = [n_1, n_2, \dots, n_M]$. Al ser una función de verosimilitud, es función de dos variables aleatorias, las medidas de ángulos y la posición del móvil.

Un estimador que maximice la densidad de probabilidad a posteriori puede usarse para resolver este problema, pero como vemos no

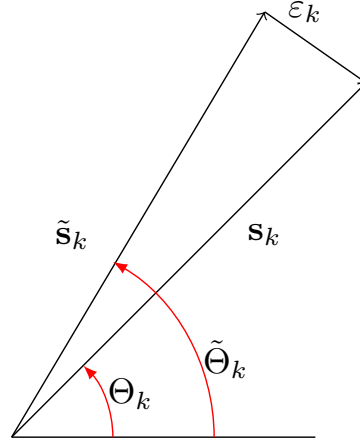


Figura 1.46: Relación entre el vector que une a la ER con el móvil y el vector producto del ruido en el ángulo de medición ([9]).

tiene una solución cerrada que además puede sesgarse por la información a priori. Una solución del tipo lineal es la denominada *estimador Bayesiano pseudo-lineal* [15].

La ecuación 1.119 puede ser linealizada como se hizo en la ecuación 1.105. Obtenemos que,

$$A_k \mathbf{s} = h_k + \eta_k \quad (1.122)$$

donde $A_k = [\sin \alpha_k - \cos \alpha_k]$, $h_k = [\sin \alpha_k - \cos \alpha_k] \cdot \mathbf{b}_k$. Por otro lado, el término

$$\eta_k = \|\mathbf{s} - \mathbf{b}_k\| \sin n_k \quad (1.123)$$

es la transformación no lineal del ruido de ángulo n_k de media cero y varianza $E\{\eta_k^2\} = d_k E\{\sin^2 n_k\}$ y donde d_k es la distancia que separa a la ER *i-esima* y el móvil. Esta transformación se puede apreciar geométricamente como muestra la figura 1.46. Para ruidos de ángulos pequeños, $\sin n_k \approx n_k$ y la varianza de η_k puede aproximarse como $d_k^2 \sigma_k^2$. Como ya hemos dicho, el error entre la posición estimada y la verdadera, dependerá del error de ángulo y la distancia que separa al móvil de la ER (ϵ_k en la figura 1.46).

Tomando las M ecuaciones correspondientes a todas las medidas,

$$\mathbf{G}\mathbf{s} = \mathbf{h} + \boldsymbol{\eta} \quad (1.124)$$

donde \mathbf{G} y \mathbf{h} se obtienen de 1.122 y ya fueron definidas en las ecuaciones 1.106 y 1.107 respectivamente y $\boldsymbol{\eta} = [\eta_1, \eta_2, \dots, \eta_M]^T$. La ecuación 1.124 se la denomina ecuación pseudo-lineal y hay que considerar que la densidad de $\boldsymbol{\eta}$ no es Gaussiana.

Usando la ecuación 1.123 se puede plantear la matriz de covarianza del vector de ruido $\boldsymbol{\eta}$ como,

$$\mathbf{C}_{\eta\eta} = \text{diag} \{ \|\mathbf{s}_{\text{mean}} - \mathbf{b}_1\|^2 \mu_1^2, \|\mathbf{s}_{\text{mean}} - \mathbf{b}_2\|^2 \mu_2^2, \dots, \|\mathbf{s}_{\text{mean}} - \mathbf{b}_M\|^2 \mu_M^2 \}$$

(1.125)

donde $\mu_k^2 = E \{ \sin^2 n_k \} = (1 - \exp(-2\sigma_k^2))/2$ y s_{mean} es la media de la información a priori ($p(s_0)$) en la ecuación 1.120 que se obtiene de otros sensores o conocimiento previo. La varianza de esta densidad de probabilidad también es conocida y la denominamos C_{ss}

Aplicando entonces mínimos cuadrados, obtenemos un estimador Bayesiano lineal de error cuadrático mínimo. La solución será:

$$\hat{s}_{BPLE} = s_{mean} + (C_{ss}^{-1} + G^T C_{\eta\eta}^{-1} G)^{-1} G^T C_{\eta\eta}^{-1} (h - G \cdot s_{mean}) \quad (1.126)$$

Esta solución es muy similar a la que ya presentamos, salvo que pone de manifiesto el conocimiento a priori que tenemos de la posición del móvil en forma aproximada y modelada como una distribución Gaussiana.

1.9 MEDICIÓN DE TIEMPOS Y ÁNGULOS

La forma básica para determinar el TOF es medir el tiempo que tarda un pulso. Este pulso se puede transmitir en banda base como en UWB o como en la mayoría de los casos, un pulso reconstruido en el receptor. En cualquiera de los casos la fuente del error es la debida a la trayectoria de la señal (multicamino) y al ruido en el receptor. Este último podría ser trabajado por el diseñador del sistema, pero el primero solo puede ser mitigado ².

El elemento final para determinar el desempeño es la Relación Señal A Ruido (SNR: Signal to Noise Ratio en inglés) que también determina el rango de operación del sistema. Esto es fundamental en los sistemas exteriores. En ambientes interiores la fuente de error más importante es el multicamino. Si analizamos un poco un ejemplo con un pulso como el de la figura 1.47 transmitido en banda base, podemos ver el efecto de estos errores en la figura 1.48.

Como se puede apreciar en la figura 1.48 si la SNR umbral (marcada con una línea punteada), que representa el piso a partir del cual se considera que lo que se recibe es señal, es muy alta puede ocultar el primer pulso recibido y detectar un tiempo más grande que el real. Por otro lado, detectar el pico es más complejo que detectar el flanco de crecimiento del pulso. Como podemos apreciar, la forma del pulso recibido tiene mucho que ver con el ancho de banda. Si se achica el ancho de banda los pulsos retrasados se van a convolucionar entre sí y el resultado es un pulso grande como suavizado. Lo ideal es un ancho de banda lo más grande posible según la aplicación.

² En este curso no vamos a tratar las formas de mitigación pero para referencia estas son la ecualización, la diversidad y la codificación de canal

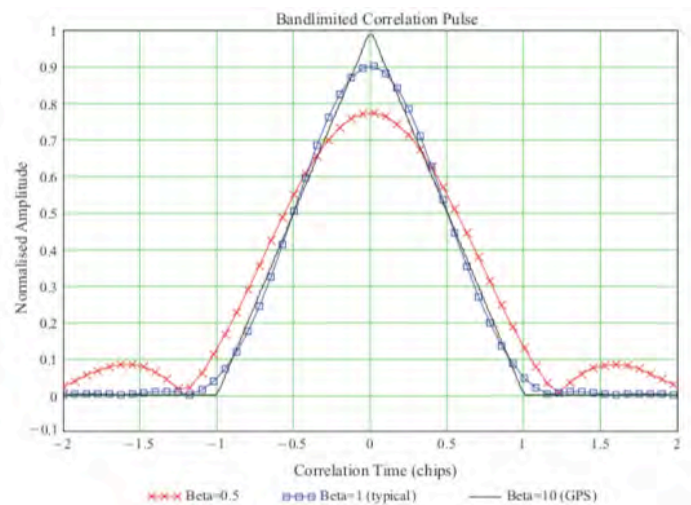


Figura 1.47: Pulso transmitido con diferentes filtros de salida para reducir el ancho de banda.

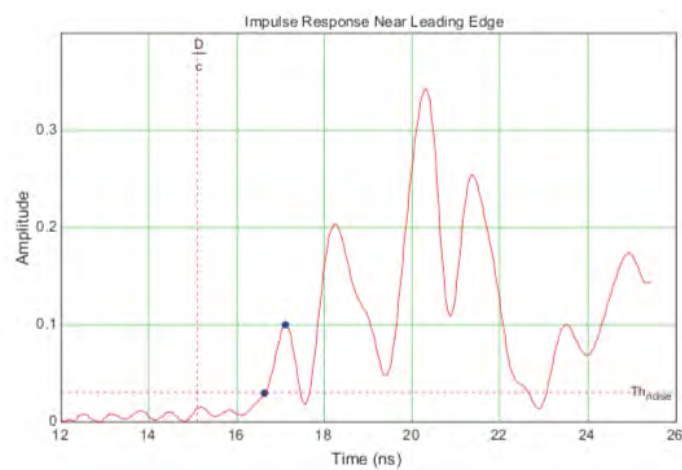


Figura 1.48: Recepción del pulso (1.47) con repeticiones por el efecto del multicamino donde se muestra el piso de ruido en línea punteada.

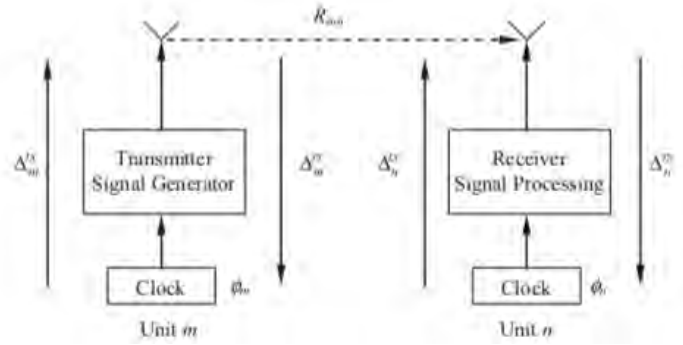


Figura 1.49: Todos los aspectos que influyen en los errores del tiempo medido respecto del tiempo de vuelo ($R_{m,n}$ en la figura) que se desea medir.

1.9.1 Medidas de tiempo entre transmisor y receptor

El retardo total entre el transmisor y el receptor no solo está compuesto por el tiempo de vuelo. En la figura 1.49 se pueden apreciar todos los componentes que están incluidos en un tiempo de vuelo medido. Ese tiempo puede ser expresado como:

$$\text{TOA}_{m,n} = \phi_m + \Delta_m^{\text{Tx}} + D_{m,n} + \Delta_n^{\text{Rx}} - \phi_n \quad (1.127)$$

donde $\phi_{(\cdot)}$ es la diferencia entre los relojes y los tiempos del pulso, $\Delta_{(\cdot)}^{(\cdot)}$ es el tiempo de procesamiento en el transmisor y el receptor y $D_{m,n} = \frac{R_{m,n}}{c}$ es la distancia a partir del tiempo de vuelo donde c es la velocidad de la luz.

Una forma de reducir la cantidad de retardos es usar la medida de tiempo de ida y vuelta a uno de los extremos, denominado *Round Trip Delay (RTT)* en inglés. Una vez que el receptor recibe el pulso, lo retransmite, entonces

$$\text{RTT}_{m,n} = (\text{TOA}_{m,n} + \text{TOA}_{n,m})/2 = D_{m,n} + (\Delta_m + \Delta_n)/2 \quad (1.128)$$

El tiempo final puede entonces ser aproximado por $\text{RTT}_{m,n} \approx D_{m,n} + \Delta_0$ donde Δ_0 es un tiempo promedio del procesamiento.

Los algoritmos para medir el tiempo recibido pueden ser muchas veces complejo. Los siguientes son algunos de los algoritmos utilizados

ALGORITMO DE PICO. Es el más simple y consiste en detectar el primer pico de la señal por arriba del umbral de ruido. Obviamente en situaciones de LOS estas medidas son precisas pero no lo son en NLOS. Es el método utilizado por el sistema GPS por lo que hay mucha literatura al respecto, y por sus características se reduce a transmisiones *Spread Spectrum*³.

³ No se tratará esta forma de comunicación en esta etapa del curso, pero sí más adelante cuando se trate el sistema de posicionamiento satelital 2.1.7

ALGORITMO DE UMBRAL. El algoritmo también es simple y tiene que ver con detectar el borde delantero del pulso cuando este excede un umbral que se define como proporcional a la altura del pulso y se elige para hacer que el algoritmo no dependa de su amplitud. Obviamente en NLOS el umbral puede detectar un pico pero no el correcto, lo que daría un error en el TOA. Además el valor del umbral debe ser mayor al piso de ruido pero lo mas bajo posible para no perder el primer pulso recibido.

ALGORITMO DE PROPORCIÓN. Otra estrategia para que el algoritmo no dependa de la amplitud o el período del pulso, es utilizando la relación o proporción entre dos muestras del pulso y cuando esa relación excede una proporción preestablecida. El punto #1 es el que supera un determinado umbral y punto #2 es el que excede la proporción.

ALGORITMO DE PROYECCIÓN. Este algoritmo, de mejor desempeño que los anteriores porque sirve para SNR bajas, detecta cuando la señal excede el piso de ruido y luego proyecta hacia atrás en el tiempo al inicio del pulso oculto en el piso de ruido. El procedimiento es como sigue: cuando la señal es mayor que el umbral se toma un segundo punto hacia atrás (por ejemplo a un $0.3 \cdot \text{UMBRALE}$) y con esos dos puntos se proyecta a 0 para determinar el inicio del pulso.

1.9.2 Estimación de RSS

En general el nivel de potencia recibida se realiza en base a una señal piloto, es decir una señal cuyo perfil completo de potencia se conoce. En redes celulares hay una canal de control que emiten las bases con este fin y se utiliza para diseñar estrategias de *Hand Off*⁴.

Una de los temas a considerar es que para obtener una buena comunicación, la recepción tiene usualmente un **control automático de ganancia** que amplifica la señal luego de la antena a niveles aceptables para el detector. Esto no significa que no pueda obtenerse el nivel de potencia pero debemos saber como se genera el RSS informado. Si por ejemplo la amplitud detectada se compara con un nivel de referencia y este error corrige la ganancia, el nivel de corrección es proporcional al nivel de potencia recibida.

1.9.3 Estimación de ángulo de arribo

En el caso de RF, se utiliza usualmente un arreglo de antenas para detectar el ángulo de arribo de la señal. La señal recibida por un

⁴ HandOff, o también Handover en inglés, es una estrategia en comunicaciones móviles celulares que tiene como objetivo de transferir el servicio de una estación base a otra cuando la calidad del enlace es insuficiente en una de ellas.

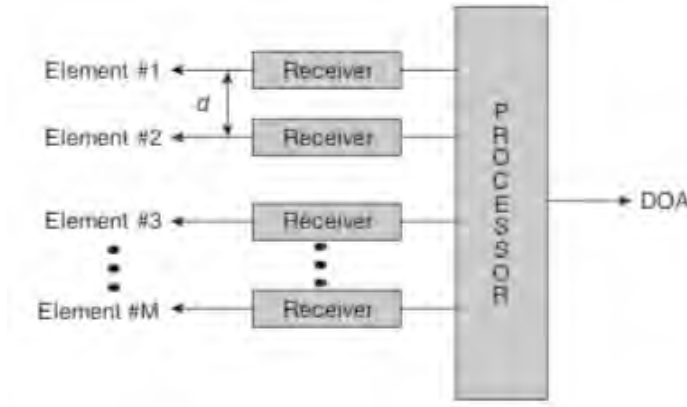


Figura 1.50: Esquema de un receptor construido para detectar el ángulo de arribo de una señal.

conjunto de elementos (antenas) es post procesada para detectar el ángulo de arribo respecto del arreglo.

La figura 1.50 muestra un esquema general de esta idea. Los elementos de las antenas están separadas físicamente una distancia $d = \lambda/2$ y suma la complejidad de que es necesario sincronizar los receptores. La etapa de procesamiento de la información recibida en todos los elementos es tratada por un algoritmo. A continuación se presenta el más común de ellos, el *clasificador múltiple de señales*.

1.9.3.1 MUSIC: Multiple Signal Classification

En un arreglo de M antenas, a cada señal recibida se le aplica un peso complejo w_i cuyo objetivo es cambiar la fase de la señal de manera que cuando se suman todas las señales recibidas la potencia de salida recibida es máxima. La fase relativa de la señal ϕ_i en función del ángulo de arribo θ_{DOA} será,

$$\phi(\theta_{DOA}) = -2\pi \frac{(i-1)d}{\lambda} \sin(\theta_{DOA}) \quad (1.129)$$

donde λ es la longitud de onda de la señal y $d = \lambda/2$ es la distancia entre las antenas. El peso aplicado a cada señal será entonces,

$$w_i(\theta_{DOA}) = \exp(j\phi_i(\theta_{DOA})) \quad (1.130)$$

La señal recibida en cada una de las M antenas será,

$$x_i(t) = w_i(\theta_{DOA}) \cdot y_i(t) + n_i(t) \quad (1.131)$$

donde $w_i(\theta_{DOA})$ es el peso definido en 1.130, y_i y n_i es la señal recibida y el error recibido por la antena i -ésima respectivamente. Definimos $\vec{w}(\theta_{DOA})$, \vec{y} , \vec{n} y \vec{x} arreglos de todas estas señales correspondientes a las M antenas.

La matriz de covarianza que incluye a todas las antenas como sensor, será:

$$\mathbf{R}_{xx} = \mathbf{E}[\mathbf{x}\mathbf{x}^T] = \tilde{\mathbf{w}}(\theta_{DOA})\mathbf{R}_{ss}\tilde{\mathbf{w}}(\theta_{DOA})^T + \sigma_n^2\mathbf{I} \quad (1.132)$$

donde σ_n^2 es la matriz de covarianza del ruido y \mathbf{R}_{ss} es la matriz de covarianza de la señal y por lo tanto no se conoce. Entonces, \mathbf{R}_{xx} puede estimarse a partir del promedio de N muestras del sensor como sigue:

$$\hat{\mathbf{R}}_{xx} = \frac{1}{N} \sum_{k=1}^N \mathbf{x}_k \mathbf{x}_k^T \quad (1.133)$$

Como el número de antenas M es mayor que el número de fuentes D que emiten la señal a medir, el algoritmo MUSIC explota el hecho que $\tilde{\mathbf{w}}(\theta_{DOA})\mathbf{R}_{ss}\tilde{\mathbf{w}}(\theta_{DOA})^T$ tendrá rango $D < M$. Esto significa que la matriz tendrá D autovalores distintos de cero y por lo tanto el subespacio de los autovectores asociados a los autovalores cero, estarán relacionados solamente con el ruido. El subespacio del ruido será ortogonal a $\tilde{\mathbf{w}}(\theta_{DOA})$, esto es $\tilde{\mathbf{w}}^H(\theta_{DOA}) \cdot \mathbf{V}_C = 0$ donde \mathbf{V}_C es el conjunto de autovectores relacionados con el ruido.

Descomponemos la matriz de covarianza estimada en sus autovectores y autovalores como $\hat{\mathbf{R}}_{xx} = \hat{\mathbf{V}}\mathbf{\Lambda}\hat{\mathbf{V}}^H$. Se identifican los $C = M - D$ autovalores mas chicos y con esto se arma la matriz $\hat{\mathbf{V}}_C$ que corresponden a los C autovectores relacionados a esos autovalores. El algoritmo MUSIC estima entonces la DOA de las señales incidentes localizando los picos en el espectro definido como:

$$\hat{\mathbf{H}}_{DOA} = \frac{1}{\tilde{\mathbf{w}}(\theta_{DOA})^H \cdot \hat{\mathbf{V}}_C \cdot \hat{\mathbf{V}}_C^H \cdot \tilde{\mathbf{w}}(\theta_{DOA})} \quad (1.134)$$

donde H significa el operador Hermitiano, es decir la matriz traspuesta compleja conjugada. Al graficar este espectro, se obtendrán picos que corresponderán con la dirección de arribo de la señal. La figura 1.51 muestra un ejemplo de un espectro construido con este método para un arreglo de seis antenas que recibe señales provenientes de las direcciones $[10, 50]$ grados a la izquierda y $[10, 14]$ grados a la derecha.

1.10 OUTLIERS: VALOR ATÍPICO

Una definición simple de valor atípico es *una observación que se desvía tanto de otras observaciones como para despertar sospechas de que fue generado por un mecanismo diferente de esas otras*. En el caso de señales de RF puede ser el caso de varias que llegan al receptor con LOS y una a partir de NLOS. El objetivo es poder desarrollar una estrategia que me permita determinar cuales medidas son esos outliers. Podríamos definir un valor atípico como cualquier medida que tenga un error *grande*. El significado de *grande* obviamente lo tenemos que definir.

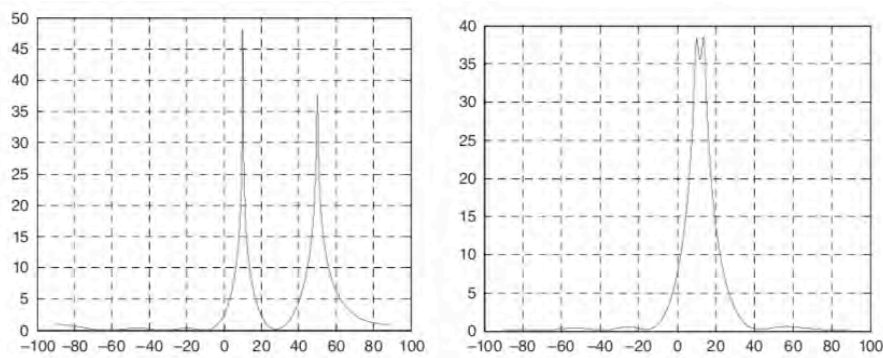


Figura 1.51: Espectro obtenido por el algoritmo MUSIC para señales que arriban desde $[10, 50]$ grados a la izq. y $[10, 14]$. a la der. El arreglo de antenas está compuesto por 6 elementos.

Se han propuesto muchos métodos de detección de valores atípicos, cada uno tiene su propia definición y en general se ajusta a la metodología usada para su detección. Algunos métodos utilizan la prueba de bondad de ajuste de la estadística, como la prueba de chi-cuadrado. La prueba proporciona una medida para determinar qué tan bien concuerdan las medidas entre sí. Si suponemos que n es el número de mediciones y k es el número de mediciones corrompidas, la idea es encontrar las $n - k$ medidas que no incluyan outliers. En cada paso, el algoritmo elige un conjunto de $n - k$ medidas del conjunto total y utiliza la prueba. Por lo general, si la prueba cruza un umbral predeterminado, el conjunto se declara como "libre de valores atípicos". De lo contrario, se prueban otras $n - k$ mediciones. Este procedimiento da buenos resultados. Sin embargo, dado que existen muchos subconjuntos de $n - k$ mediciones, requiere de muchos cálculos. Por ejemplo con $n = 30$ mediciones y $k = 10$ valores atípicos, el número de posibles subconjuntos a verificar son $\binom{n}{n-k} = \binom{30}{20} = 30,045,000$. Además, normalmente se desconoce el número de valores atípicos y, este número puede crecer mucho.

1.10.1 *Análisis de los residuos*

El número de observaciones requeridas (pseudo-rangos y / o ángulos de arribo) depende de la geometría particular y del algoritmo utilizado. En las implementaciones típicas de multi-angulación por ejemplo, se necesitan dos mediciones en 2D. Para poder construir un algoritmo que sea robusto a las medidas erróneas, es útil tener más mediciones que las mínimas y descartar aquellas sospechosas de ser outlier.

Un algoritmo para lograr descartar las medidas atípicas es obtener más medidas que las necesarias para la localización y analizar el residuo. Esto es, la diferencia entre la medida y la estimación de la medida. Se puede aplicar a cualquiera de los métodos que ya vimos.

Tomemos por ejemplo el algoritmo de AoA por mínimos cuadrados. La solución la repetimos acá,

$$\Theta_{1,LS} = [\hat{x}, \hat{y}]^T = (\mathbf{G}^T \mathbf{G})^{-1} \mathbf{G}^T \mathbf{h} \quad (1.135)$$

donde

$$\mathbf{G} = \begin{bmatrix} \sin(\alpha_1^0) & -\cos(\alpha_1^0) \\ \vdots & \vdots \\ \sin(\alpha_M^0) & -\cos(\alpha_M^0) \end{bmatrix}. \quad (1.136)$$

y

$$\mathbf{h} = \begin{bmatrix} x_1^0 \sin(\alpha_1^0) - y_1^0 \cos(\alpha_1^0) \\ \vdots \\ x_M^0 \sin(\alpha_M^0) - y_M^0 \cos(\alpha_M^0) \end{bmatrix}. \quad (1.137)$$

Recordamos que las medidas son $\alpha = [\alpha_1, \alpha_2, \dots, \alpha_M]$. Entonces, el vector de residuos será

$$\varepsilon_i = \alpha_i - \tan^{-1} \frac{\hat{y} - y_i}{\hat{x} - x_i} \quad (1.138)$$

Si se cumple que el ruido que aparece en las medidas α es como dijimos de media cero, Gaussiano y varianza conocida, ε_i tendrá la misma estadística. Esto implica que si dibujamos ese error para cada medida debería mostrar valores razonablemente aleatorios.

A partir de 1.138, se puede calcular el error estándar

$$e = \sqrt{\frac{\sum_{i=1}^M \varepsilon_i^2}{M - B}} \quad (1.139)$$

donde B es el número mínimo de medidas necesarias (2 en 2D) y M el conjunto total de medidas. Si las observaciones dadas no se corroboran entre sí, el resultado tendrá una estimación del error estándar e alto. Una estrategia es establecer un umbral u_e y si $e > u_e$, se elimina la observación con el residuo estándar más alto y vuelve a calcular e. Este proceso continúa hasta que el error estándar de la estimación cae por debajo del u_e . Esta situación es útil y simple de aplicar en casos como el que se muestra en la figura 1.52. Allí se descartaría el dato señalado.

Esta estrategia depende de la definición del umbral u_e y descarta en forma determinística los valores. Una alternativa es utilizar la estadística de los residuos para hacer el descarte. Como dijimos, los residuos serán gaussianos de media cero entonces una estrategia es verificar su distribución. Esto se denomina *Normal Probability plot*. El algoritmo sería como sigue (ejemplo en la tabla 1.3):

1. Tabular los valores de los residuos en orden, del mas negativo al mas positivo (1ra columna de la tabla 1.3).

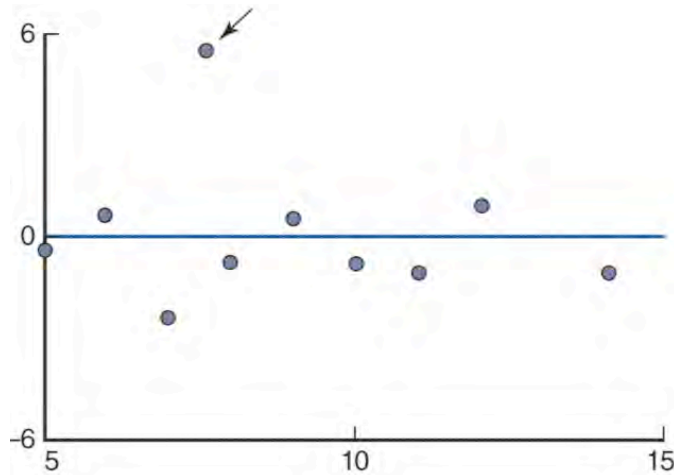


Figura 1.52: Residuos individuales de las M medidas que muestran valores razonablemente aleatorios. Salvo el elemento señalado con la flecha que parece ser un valor atípico

Residuo	Rango	Frec. Acumulada
-1.92	1	0.125
-0.05	2	0.375
0.18	3	0.625
1.24	4	0.875

Tabla 1.3: Datos para un gráfico de probabilidad normal

2. Numerar los residuos de 1 a M , donde 1 corresponde al mas negativo y así hasta el mas positivo (2da columna de la tabla 1.3).
3. Calcular la frecuencia acumulada de los residuos como

$$F_i = \frac{i - 0,5}{M},$$

donde i es el índice del paso 2 (3ra columna de la tabla 1.3).

4. Armar un gráfico donde la abscisa es el valor del residuo (1ra columna de la tabla 1.3) y la ordenada es la frecuencia acumulada (3ra columna de la tabla 1.3).
5. todos los puntos debe estar en una línea recta que indica la normalidad de los valores. Cualquier desviación implica valores atípicos (figura 1.53).

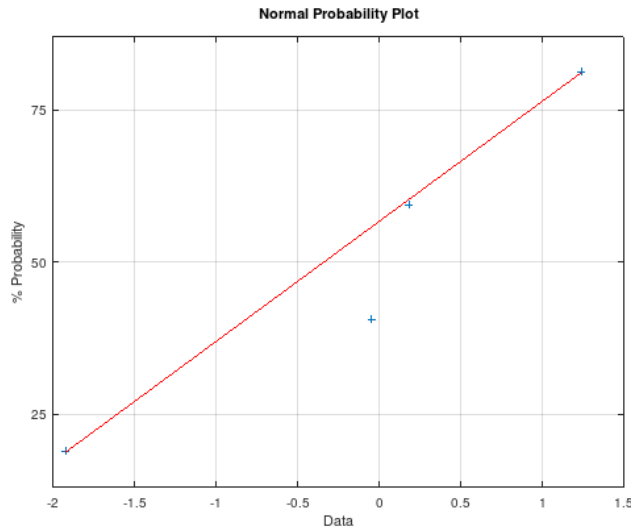


Figura 1.53: Grafico de probabilidad normal para el ejemplode la tabla 1.3. El segundo punto está fuera de la curva y es un probable outlier

Otro método que puedo aplicar, es a partir de la aplicación de la siguiente idea. Si los residuos son variables aleatorias Gaussianas y S_ε representa una aproximación de la varianza de los residuos,

$$S_\varepsilon = \sqrt{\frac{\sum_{i=1}^M \varepsilon_i^2}{M}} \quad (1.140)$$

la variable aleatoria $\varepsilon_i^2/S_\varepsilon$ tendrá una distribución χ^2 con los grados de libertad de la dimensión de ε_i . Podemos entonces definir una cota γ tal que todas las medidas que están por afuera de esta cota tendrán una probabilidad de que sea un valor atípico con un grado de certeza ζ definido por nosotros (por ejemplo 0.95 o 0.99). Es decir el $100 * (1 - \zeta) \%$ de medidas verdaderas serán descartadas, pero aquellas que violan las condiciones estadísticas supuestas quedarán afuera y pueden ser descartadas. Esta estrategia se denomina también la estrategia del *vecino mas cercano*. Por ejemplo, para dos dimensiones, y un grado de certeza $\zeta = 0,95$, la cota tendrá un valor de $\gamma = 5,99$ y para $\zeta = 0,99$ corresponde a $\gamma = 9,21$.

1.11 FINGERPRINTING

Si se utiliza información de radio frecuencia para construir un algoritmo de localización, el multicamino y su consecuencia el fadding o desvanecimiento de la señal de RF que arriba a un receptor, degrada su desempeño. Según el ambiente, este efecto puede ser muy importante y negativo si usamos las técnicas de posicionamiento vistas hasta ahora. Sin embargo, desde el punto de vista estadístico, cuanto peor es el canal, significa que las señales que arriban al receptor

pueden ser independientes estadísticamente para cada punto en el espacio donde yo esté recibéndolas. En otras palabras, para cada punto del espacio, la variedad de señales que se suman y restan al llegar a la antena, construyen una *firma* de ese punto y esta idea es la base para la construcción de un algoritmo de localización basado en firmas o *Fingerprinting*. A medida que el canal es peor, las firmas de cada punto serán *mas diferentes* entre sí.

Las técnicas de localización por firmas se basan en comparar las características de las señales de RF medidas con patrones típicos que corresponden a un conjunto de coordenadas. Cuanto más similares sean los patrones, más probable es que el terminal esté cerca de la posición asociada con la coincidencia. Lo importante es que los algoritmos de posicionamiento basados en firmas no requieren un modelado a priori del canal o el tiempo de vuelo.

Aunque la forma mas simple para construir una firma de un punto es usando la información de RSS de los mensajes transmitidos por las ER, obviamente no es el único dato que puede constituir la firma. La orientación dada por un compás, la altitud dada por un barómetro, etc., se pueden sumar a esta firma. El objetivo es lograr que cada punto sea lo mas *distintivo* posible.

Aunque existe una gran variedad de técnicas de localización por firmas, todas ellas comparten los mismos elementos básicos:

- **Huella digital/Firma:** Se trata de un conjunto de parámetros dependientes de la ubicación. Cuantas más señales se observen o cuantos más parámetros por señal se observen, más exclusiva será la huella digital y, por tanto, mayor será la precisión de la localización. Si en la firma de RF sólo se incluyen parámetros propios de la red (como RSSI, MAC, etc.), la técnica de localización puede estar totalmente basada en la red y no es necesario modificar la estructura existente
- **CDB/Base de Datos de Correlación:** Las firmas se recogen en pruebas de campo o se generan utilizando modelos de simulación, y se almacenan en una base de datos denominada CDB (Correlation DataBase), a la que puede acceder directamente el dispositivo que va a realizar la localización. Cada firma almacenada en la CDB está asociada a una posición específica.
- **Servidor de localización:** Si se piensa al esquema de localización como una *red de localización*, el servidor es el elemento responsable de recibir las solicitudes de localización, consultar el CDB y estimar la localización del móvil.
- **Reducción del espacio de búsqueda dentro de la CDB:** La CDB puede ser bastante grande, y el análisis de todas las firmas de RF almacenadas en ella puede llevar mucho tiempo. Por lo tanto, todas las técnicas de localización por firmas aplican algún

método para reducir el espacio de búsqueda dentro de la CDB. Como consecuencia, también se reduce el tiempo necesario para producir una fijación de posición.

- **Comparación de patrones:** Para estimar la posición del móvil, el servidor de localización debe comparar la firma de RF medida por él con un subconjunto de firmas almacenadas en la CDB. Esta comparación o concordancia de patrones puede realizarse utilizando diferentes técnicas, algunas de las cuales presentaremos acá.

Las dificultades de este método tienen que ver con dos aspectos. Por un lado la construcción y mantenimiento de la CDB. Lleva planificación y tiempo y tiene que ver con que el canal puede ser muy cambiante. En ambientes interiores muebles que se corren o puertas que se cierran. En ambientes exteriores ER que se apagan o mueven o el movimiento de equipos o vehículos. Por otro lado, el tamaño de la CDB crece con el espacio cubierto y el detalle de la grilla en la que se levantan los datos.

La alternativa para mantener la CDB actualizada son varias. El mas popular es el de *Crowdsourcing*, que es el que utilizan las grandes compañías que ofrecen aplicaciones o sistemas operativos gratis para dispositivos inteligentes o aquellas compañías que prestan servicios de red. Otra alternativa es Simultaneous Navigation and Mapping. Esto es, el móvil se va moviendo en un espacio desconocido, construye su mapa a medida que ve elementos diferenciales nuevos, y a la vez usa ese mapa para localizarse. También se puede organizar un mapa a partir de modelado, simulación e interpolación de datos medidos. En el curso nos vamos a centrar en el problema de localización por lo que no vamos a profundizar en ninguna de estas estrategias.

Como ya se dijo, las firmas pueden estar compuestas por distinta información. Así por ejemplo:

- **Firmas visuales.** Básicamente se analiza el contenido de una imagen para construir un conjunto de elementos distintivos que representen una firma del punto donde se tomó la imagen. El tipo de algoritmos que se utilizan para extraer firmas son por ejemplo el SIFT (Scale-Invariant Feature Transform) o el GLOH (Gradient Location and Orientation Histogram). La correlación con la firma puede ser por Nearest Neighbor Search.
- **Firmas de Movimiento.** Se utilizan datos de movimiento como aceleraciones o velocidades angulares. Este método de localización usando este tipo de firmas es incremental. A partir de la posición previa y la huella de movimiento se determina la nueva posición. Es bastante impreciso por lo que raramente se utiliza como único método o información.

- **Firma de Señales.** Es el tipo de firmas de las que venimos hablando, es decir información de señales de RF.

Antes de presentar las formas de localización presentamos las siguientes definiciones. El móvil mide una firma que es aquella que trataremos de correlacionar con las almacenadas en el mapa de la CBD. Esa firma T_{FING} , suponemos que solo contiene información de RSS y la identificación de la ER que envió el mensaje con esa potencia,

$$T_{\text{FING}} = \begin{bmatrix} \text{ID}_1 & \text{RSS}_1 \\ \vdots & \vdots \\ \text{ID}_N & \text{RSS}_N \end{bmatrix} \quad (1.141)$$

La CBD tiene como ya dijimos datos de firmas asociados con posiciones, suponiendo que las firmas las tenemos en el plano,

$$\text{CBD} : \{(x_i, y_i), R_{\text{FING}_i}\} \quad (1.142)$$

donde R_{FING_i} son las firmas de referencia,

$$R_{\text{FING}_i} = \begin{bmatrix} \text{ID}_{i,1} & \text{RSS}_{i,1} \\ \vdots & \vdots \\ \text{ID}_{i,N} & \text{RSS}_{i,N} \end{bmatrix} \quad (1.143)$$

Tanto R_{FING_i} como T_{FING} pueden tener mas columnas, como la información de orientación provista por un compás, el retardo de la señal (puede ser el tiempo de vuelo o mejor el *round trip delay* que es el tiempo de un mensaje que viaja ida y vuelta entre el Tx y el Rx). Es posible organizar el mapa con una grilla uniforme. En ese caso, las coordenadas pueden reemplazarse por la posición en la grilla,

$$\text{CBD} : \{(i, j), R_{\text{FING}_{i,j}}\} \quad (1.144)$$

1.11.1 Obtención de la posición (Pattern Matching)

La forma mas simple de determinar la posición del móvil es minimizar la distancia entre las RSSI_i del T_{FING} con las de la CBD. Definimos a

$$\vec{T} = [T_{\text{FING}}(1, 2), \dots, T_{\text{FING}}(N, 2)] \quad (1.145)$$

y

$$\vec{R}_i = [R_{\text{FING}_i}(1, 2), \dots, R_{\text{FING}_i}(N, 2)] \quad (1.146)$$

los vectores de la RSS de las firmas del móvil y la de cada una de las entradas de la CBD. Para simplificar este primer análisis, suponemos que $R_{\text{FING}_i}(u, 1) = T_{\text{FING}}(u, 1)$, es decir corresponden al mismo ID

y cada elemento de los vectores \vec{T} y \vec{R} se corresponden también. El objetivo es entonces encontrar el índice k que determina el \vec{R} mas parecido a \vec{T} . Algunas estrategias son,

$$k = \underset{i}{\operatorname{argmin}} \frac{\langle \vec{T}, \vec{R}_i \rangle}{\|\vec{T}\| \|\vec{R}_i\|} \quad (1.147)$$

$$k = \underset{i}{\operatorname{argmin}} \|\vec{T} - \vec{R}_i\| \quad (1.148)$$

$$k = \underset{i}{\operatorname{argmin}} [(\vec{T} - \vec{R}_i)(\vec{T} - \vec{R}_i)^T] \quad (1.149)$$

donde

$$\langle \vec{T}, \vec{R}_i \rangle = \sum_i x_i \cdot y_i \quad (1.150)$$

y

$$\|X\| = \sqrt{\sum_i x_i^2} \quad (1.151)$$

La solución a la ecuación 1.147 va a corresponder a aquella en la cual el vector R_{FING_k} estará mas alineado con T_{FING} . Recordemos que el producto interno lo podemos expresar como

$$\langle a, b \rangle = \|a\| \|b\| \cos \theta \quad (1.152)$$

donde θ es el ángulo en radianes entre los vectores. Por otro lado, la ecuación 1.149 es como la 1.148 pero sin la operación de raíz cuadrada. Hay otras posibles medidas de distancia. Por ejemplo,

$$k = \underset{i}{\operatorname{argmin}} \frac{\langle \vec{T}, \vec{R}_i \rangle}{\langle \vec{T}, \vec{T} \rangle + \langle \vec{R}_i, \vec{R}_i \rangle - 2 \langle \vec{T}, \vec{R}_i \rangle} \quad (1.153)$$

o la siguiente que reduce la complejidad de cálculo

$$k = \underset{i}{\operatorname{argmin}} \sum_{n=1}^N [\vec{T} - \vec{R}_n] \quad (1.154)$$

Sin embargo este caso no es una medida de distancia y por lo tanto no la minimiza. En cualquiera de estos casos, el resultado de localización es la posición correspondiente al R_{FING_k} .

Estos métodos son simples de calcular y entender pero no tienen en cuenta las incertidumbres en las medidas de las RSS_i ni tampoco los errores de cuantización al digitalizar las medidas de potencia. La otra cuestión ideal del planteo hasta aquí es que suponemos que todas las

ER están disponibles para todas las firmas. Si las firmas tienen solo información de RSS, una alternativa es agregar en las ER que no están medidas y ponerles un valor muy bajo de potencia. Por ejemplo, si la sensibilidad del receptor es de 100dBm, poner un valor de 120dBm.

Una alternativa es considerar una penalización para las R_{FING_i} que tienen menos ER en T_{FING} que otras,

$$d_{i,j} = \sqrt{\sum_{k=1}^N (T_{mk} - \vec{R}_{nk})^2 + 2\beta(N_a - N)} \quad (1.155)$$

donde mk y nk son los índices de las líneas cuyos ID coinciden en ambas firmas, N_a es el número total de ID en T_{FING} , N el número de ID de T_{FING} que están en R_{FING_i} , β es el rango dinámico de la RSS en dB. Naturalmente si $N_a = N$ la penalización es cero. En el caso de β depende del sistema de comunicación y de la diferencia de potencia en dBm entre aquella a partir de la cual el amplificador empieza a distorsionar (límite superior) y la mínima potencia detectable o sensibilidad del receptor (límite inferior). Usualmente es alrededor de 60dBm.

Otro aspecto que se puede tener en cuenta es que los valores de RSS mas pequeños son mas imprecisos. Puede entonces penalizarse con un peso. Una estrategia es ordenar los RSS de mayor a menor, es decir $RSS_p \geq RSS_q$ si $p < q$, entonces el peso sería,

$$w_k = 10^{0,1\psi(RSS_k - RSS_1)} \quad (1.156)$$

donde RSS_1 es la mayor potencia medida y $\psi \in [1,3]$ y se define empíricamente.

1.11.1.1 Correlación por Nivel

Dada la variabilidad en los dispositivos que miden la RSS, tanto en la construcción de la CDB como en la etapa de localización, es necesario construir una estrategia para mitigar ese efecto. Una forma de hacerlo es analizando el rango o nivel de potencia que ocupa en forma relativa cada RSS en la R_{FING} . La idea es que mientras los niveles de potencia en valor absoluto pueden ser diferentes, cuando se miden varias ER su rango relativo se mantendría en general. Por ejemplo, si dos móviles diferentes miden determinado nivel de potencia diferente para una ER_i , es probable pero en ambos casos se mantengan mayor a otra ER_j .

Una forma de implementar esta idea es con el *Spearman Rank Correlation Coefficient* (SRC). Este coeficiente se puede utilizar como un índice de correspondencia entre T_{FING} y R_{FING} . Sin embargo, estas firmas no tienen necesariamente el mismo número de ER o los mismos ID de ER. En consecuencia, antes de aplicar el factor de corre-

lación, se requiere un paso de ordenación y selección. Se crean dos matrices V_t y V_r de $N_c \times 2$, cuyos valores iniciales serán,

$$V_t(n, 1) = V_r(n, 1) = ID_n \quad (1.157)$$

y

$$V_t(n, 2) = V_r(n, 2) = N_c \quad (1.158)$$

donde N_c es el número de ER en el área e ID_n es el número de identificación de la ER n -ésima. Las filas de T_{FING} están organizadas en orden descendiente de RSS.

$$V_t(n_k, 2) = k \quad (1.159)$$

donde $V_t(n_k, 1) = T_{\text{FING}}(k, 1)$, $n_k \in [1, N_c]$ y $k = 1, 2, \dots, N_a$, además

$$V_r(n_k, 2) = k \quad (1.160)$$

donde $V_r(n_k, 1) = R_{\text{FING}}(k, 1)$, $n_k \in [1, N_c]$ y $k = 1, 2, \dots, N_a$.

El coeficiente SRC se calcula como,

$$\rho = \frac{\sum_{n=1}^{N_c} (V_t(n, 2) - \bar{\rho}_t)(V_r(n, 2) - \bar{\rho}_r)}{\sqrt{\sum_{n=1}^{N_c} (V_t(n, 2) - \bar{\rho}_t)^2 \sum_{n=1}^{N_c} (V_r(n, 2) - \bar{\rho}_r)^2}} \quad (1.161)$$

donde

$$\bar{\rho}_t = \frac{1}{N_c} \sum_{n=1}^{N_c} \{V_t(n, 2)\} \quad (1.162)$$

$$\bar{\rho}_r = \frac{1}{N_c} \sum_{n=1}^{N_c} \{V_r(n, 2)\} \quad (1.163)$$

La distancia entre la firma y la medición será entonces,

$$d = 1 - \rho \quad \text{con } \rho \in [-1, 1] \quad (1.164)$$

EJEMPLO: Calcular la distancia Spearman entre la medida hecha por el móvil

$$T_{\text{FING}} = \begin{bmatrix} 100 & 110 & 2 & 5 & 99 \\ 62 & 60 & 54 & 43 & 40 \end{bmatrix} \quad (1.165)$$

y la siguiente firma en la CDB

$$R_{\text{FING}} = \begin{bmatrix} 100 & 5 & 110 & 111 & 10 & 200 & 201 \\ 54 & 50 & 49 & 45 & 34 & 30 & 29 \end{bmatrix} \quad (1.166)$$

Supongamos que la RRS está cuantizada en 64 valores en pasos de 1dB, donde 0 corresponde a -110dBm y 63 a -48dBm (esto es un receptor típico de GSM por ejemplo). Armamos la matriz V_t , considerando que el número total de ER $N_c = 9$, como

$$V_t = \begin{bmatrix} 2 & 5 & 10 & 99 & 100 & 110 & 111 & 200 & 201 \\ 3 & 4 & 9 & 5 & 1 & 2 & 9 & 9 & 9 \end{bmatrix} \quad (1.167)$$

y

$$V_t = \begin{bmatrix} 2 & 5 & 10 & 99 & 100 & 110 & 111 & 200 & 201 \\ 9 & 2 & 5 & 9 & 1 & 3 & 4 & 6 & 7 \end{bmatrix} \quad (1.168)$$

De estas matrices podemos encontrar $\bar{p}_t = 5,7$ y $\bar{p}_r = 5,1$. Con estos datos,

$$d = 0,69 \quad (1.169)$$

Como se ve, el resultado no tiene que ver con el RSS_i sino su valor relativo. Además el número de IDs puede ser diferente para T_{FING} y R_{FING}

1.11.1.2 Reducción del espacio de Búsqueda

Cada elemento de la CDB contiene una firma de referencia y sus coordenadas geográficas. El espacio de búsqueda es el conjunto de elementos de la CDB cuyas firmas de RF se comparan con la firma medida por el móvil. Las coordenadas geográficas de los elementos del espacio de búsqueda son los candidatos para la localización del móvil.

Inicialmente, el espacio de búsqueda comprende todos los elementos CDB. Sin embargo, a medida que la base de datos es mas grande, no es factible por la carga de cálculo y el tiempo necesario comparar la firma medida con todas las almacenadas en el CDB. Por lo tanto, es necesario una estrategia para reducir el espacio de búsqueda sin que la precisión de la localización se vea afectada de forma significativa. El tamaño de la CBD puede crecer por el área cubierta pero también por la granularidad espacial.

Hay diversas estrategias, pero la más simple es la siguiente:

1. El primer paso en la reducción, es restringir el espacio a aquellas R_{FING} que compartan con T_{FING} la ER con mayor RSS, es decir

$$\{i \mid R_{\text{FING}_i}(1,1) = T_{\text{FING}}(1,1)\} \quad (1.170)$$

donde se supone que las firmas están ordenadas de mayor a menor por su RSS.

2. El segundo paso se basa en la utilización de otros parámetros presentes en las firmas además del RSS. Por ejemplo si se usa el RTD (*round trip delay*), entonces de los elementos seleccionados en el paso anterior ordenados por su RTD,

$$\{i \mid R_{\text{FING}_i}(1,3) = T_{\text{FING}}(1,3)\} \quad (1.171)$$

3. Puede agregarse un tercer paso, que es establecer un número de ER mínimas que tienen que compartir las firmas de la CBD del paso anterior y la firma del móvil.

Este método es simple pero no el único. Otras estrategias, deben evaluarse no solo en la complejidad sino que también mejoren los resultados de localización. Esas estrategias incluyen algoritmos de optimización como el algoritmo genético, o a partir de agrupaciones de ER. Este último es, luego del presentado, muy popular y consiste en un primer paso y fuera de línea en agrupar las ER en base a alguna estrategia (como k-means). Cada grupo tiene una firma representativa y cuando se inicie el proceso de localización, se compara esta forma representativa con la medida por el móvil. La mas cercana determina el grupo donde se va hacer la estimación (ver por ejemplo [20]).

1.11.1.3 Estimación de la posición como promedio

El método mas simple es asignar la posición del móvil con la de la ER con la firma mas cercana. Esta estrategia es directa a partir de lo que ya vimos. Pero es raro que el móvil se encuentre muy cerca de la ER como para que esta estrategia sea buena. Una mejora es construir un promedio de varias ER cercanas.

K-VECINOS MAS CERCANOS (KNN). La idea es simple. Seleccionamos los k R_{FING_i} mas cercanos a la firma del móvil y la estimación será:

$$(\hat{x}, \hat{y}) = \sum_k w_k (\hat{x}_k, \hat{y}_k) \quad (1.172)$$

donde $\sum_k w_k = 1$ y cada w_k puede elegirse de distintas formas. La mas simple es $w_k = 1/k$ y se denomina *k-means*.

MOVING AVERAGE FILTER (MAF). Si el móvil no se mueve o lo hace muy poco en relación a la longitud de onda de RF (por ejemplo en 2.4Ghz la longitud de onda es de 12.5cm), se puede hacer una estimación como

$$(\hat{x}, \hat{y}) = \sum_{u \in [0, L-1]} w_u (\hat{x}_{j-u}, \hat{y}_{j-u}) \quad (1.173)$$

donde $w_u = 1/L$ en el caso más simple.

KNN PROBABILÍSTICO. No vamos a abundar en esta estrategia porque es necesario entender varios aspectos sobre estimación bayesiana. Sin embargo podemos plantear que el tipo de soluciones serán

$$(\hat{x}, \hat{y}) = \frac{1}{\sum_{j=1}^k p(x|z)_j} \sum_{i=1}^k (\hat{x}_i, \hat{y}_i) p(x|z)_i \quad (1.174)$$

donde $p(x|z)_i$ es la probabilidad de que la posición que buscamos (x) condicionada a la medida (z) que en este caso es la posición del R_{FING_i} . Para encontrar esta densidad, se puede aplicar Bayes, esto es,

$$p(x|z)_i = \frac{p(z|x)_i p(x)}{p(z)_i} \quad (1.175)$$

donde $p(x)$ es la información a priori que tenemos sobre la posición del móvil y $p(z|x)_i$ es la likelihood o verosimilitud y se pueden usar distintas estrategias para modelarla (Gaussianas, Gaussian Kernels, etc.).

1.12 EJERCICIO FINAL

La propuesta de ejercicio final es que recolecten datos de RF de Access Points de WiFi en dos o tres pasadas por recorridos que ustedes elijan. Como ejemplo, en la figura 1.54 se muestran tres recorridos diferentes. Por ejemplo lineal, una vuelta a la manzana o una par de manzanas consecutivas. Con una o dos de las pasadas, construir la base de datos y la restante utilizarla para la localización. De esta última pueden tomar algunas de las medidas (como las correspondientes a las de las esquinas en los recorridos no lineales) como el objetivo a localizar.

RECOLECCIÓN DE DATOS. Para juntar los datos les propongo que tomen datos en días, horas y direcciones diferentes de ser posible. Si no, los pueden tomar en forma consecutiva pero al menos en diferentes direcciones. Para esto utilizan su teléfono. Les propongo dos aplicaciones.

- **WiFi Analyzer (open-source) VREM Software Development.** Es una aplicación que les permite recolectar la RSS de todos los AP de WiFi que el teléfono puede ver. Esos datos pueden almacenarse organizados a partir del nivel de RSS, en un txt para ser utilizados luego. Los autores de la aplicación solo solicitan compartir la información de WiFi, la posibilidad de conectar o desconectar el WiFi y la ubicación que no la usa en el algoritmo sino que es un requerimiento de Android para acceder a la info de WiFi.

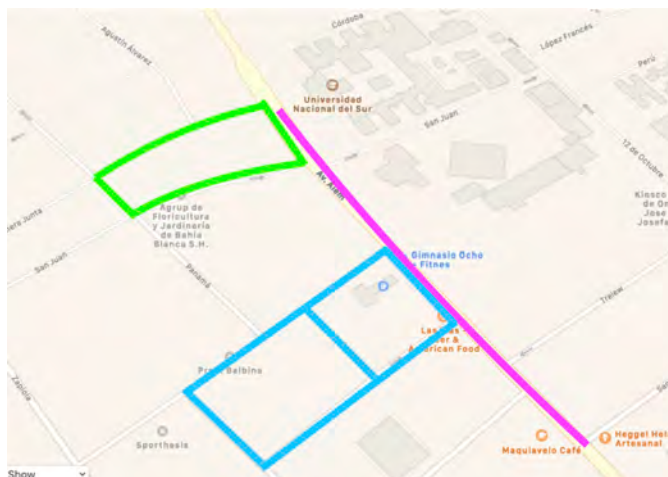


Figura 1.54: Ejemplos de recorridos para la recolección de datos

- **My Location - Where Am I. Financept.** Según lo que estuve viendo, es una de las que menos acceso a su teléfono le requiere. Permite salvar la posición en Latitud y Longitud pero también en la dirección física (por ejemplo Av. Alem 1253)

La cantidad de datos por pasada queda a criterio de ustedes. Les propongo al menos 5 puntos por cuadra (uno cada 20 metros mas o menos). La mínima distancia entre puntos tiene que ser 5 veces la longitud de onda en la frecuencia. En este caso 60cms aproximadamente. Si quieren (no es obligatorio), pueden tomar para cada posición un promedio de tres a cinco medidas. Para esto, para cada medida, se mueven una longitud de onda en diferentes direcciones. Por ejemplo, el primer punto y un paso hacia adelante, atrás, a la derecha y a izquierda de la primera. La coordenada del punto es la del primero medido.

LOCALIZACIÓN. El algoritmo de localización lo pueden implementar en Matlab, Octave, Phytion o lo que ustedes quieran. Deben presentar los resultados de los dos algoritmos de promedios: el KNN y el MAF.

PREPARACIÓN DE DATOS. Los datos de posición de cada firma fueron obtenidos como dirección o LAT/LONG. Si usamos la posición de la firma mas cercana como la del móvil esto es suficiente. Pero como debemos implementar un promedio de posiciones, ninguno de lo dos me sirven para localizar en forma directa al móvil.

Una forma simple, pero no muy precisa, es tomar un punto como el (0,0) y convertirlo las direcciones físicas en una coordenada. Por ejemplo, si Av. Alem 1200 es el punto (0,0), Av. Alem 1253 es (53,0). De esta forma puedo ir al mapa en forma directa como referencia al punto inicial. El problema es en movimientos que no son lineales.

Algo un poco mas preciso es convertir las coordenadas LAT/LONG en unas coordenadas locales. La estrategia mas simple es

1. Multiplicar los grados por 60 (la cantidad de minutos en un grado)
2. Sumar los minutos
3. Si se desea, si la latitud es sur o si la longitud es oeste multiplicar por -1
4. Si se desea, se puede tomar un punto como referencia (una esquina por ejemplo o el primer punto) y restarlo a la cantidad encontrada.

Para volver a LAT/LONG hacemos el procedimiento inverso.

Otra estrategia, que es la mas correcta y precisa, es utilizar una conversión formal de la latitud y longitud al sistema de coordenadas universal transversal de Mercator (en inglés Universal Transverse Mercator, UTM). donde las coordenadas están expresadas en metros. En este sistema el mundo es partido en parches y la posición en metros es un punto del parche, tomando como origen el vértice inferior izquierdo. Hay mapas con los parches, y el que corresponde a Bahía Blanca es la 20H, y la de Neuquén es la 19H. La conversión se puede hacer a partir directamente de los valores de LAT/LONG. Existen páginas que lo hacen en línea⁵ (pueden buscar, hay varias que convierten varios puntos a la vez), funciones de Matlab que probablemente también funcionen en Octave⁶ y planillas Excel. Obviamente, se puede ir de un sistema al otro sin problemas.

Los valores de RSSI por otro lado tienen que cuantizarlos en función de la sensibilidad del receptor y el valor maximo de potencia recibida por el amplificador de entrada. En general los receptores van entre los -85dBm, los -105dBm de la mayoría de los receptores hasta los profesionales con -120dBm. Tomemos -105dBm como referencia inferior. EL límite superior usual es -10dBm (si prueban con la aplicación sobre el modem les dará entre -15 y -20 dBm). Pueden cuantizar de a 1dBm con el cero en -110dBm y el máximo 100 corresponde a -10dBm.

⁵ Ejemplo <https://tool-online.com/es/conversion-coordenadas.php>

⁶ Ejemplo <https://www.mathworks.com/matlabcentral/fileexchange/10915-deg2utm>

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

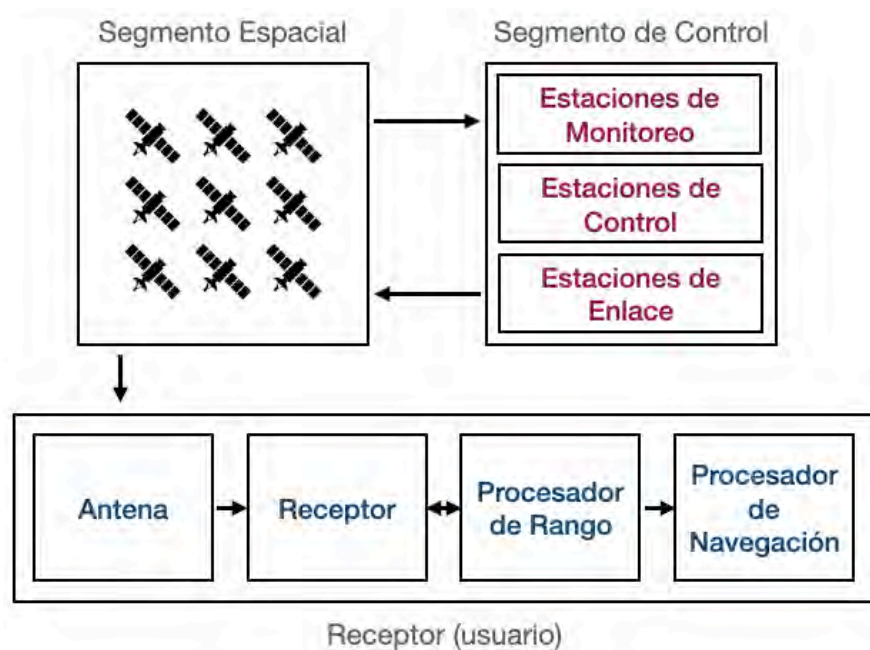


Figura 2.1: Esquema de la estructura del sistema GNSS donde participan un segmento satelital y el receptor al que estamos acostumbrados, pero se suma un segmento de control que posee estaciones de monitoreo y de corrección del sistema.

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-

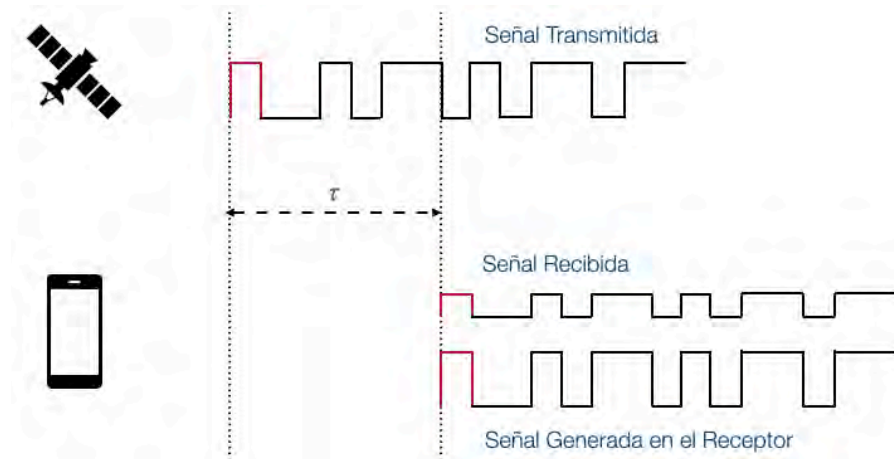


Figura 2.2: Principio de medida del tiempo de propagación comparando los códigos pseudoaleatorios

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 pseudorange 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 pseudorange ρ 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 δt_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 pseudorange 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.

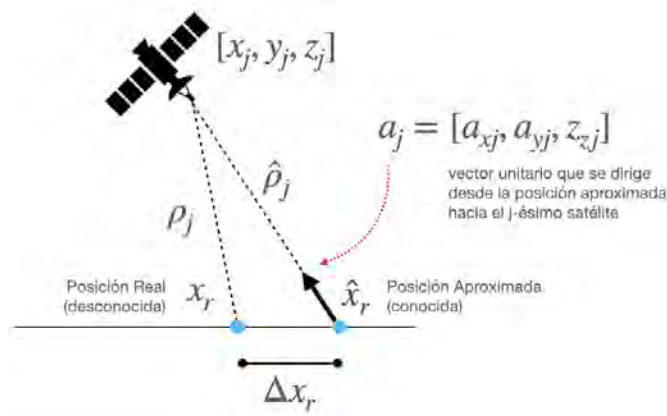


Figura 2.3: Ejemplos de la linealización en una dimensión

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 pseudorangos 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:

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

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

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

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

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

donde $\delta \rho$ 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 $\delta \rho$. La covarianza de $\delta \mathbf{x}$ se puede obtener a partir del producto $\delta \mathbf{x} \delta \mathbf{x}^T$ y calculando la esperanza,

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

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

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 σ_ρ^2 ,

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

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 σ_ρ^2 . Para el SPS el σ_ρ^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:

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

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 1σ 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.

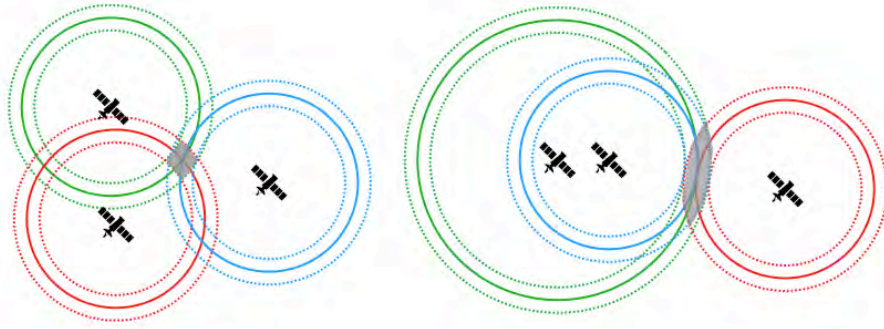


Figura 2.4: Ejemplos de como opera la dilución geométrica de precisión

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 $G = (H^T H)^{-1}$

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

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

2. Dilución de precisión horizontal (HDOP en inglés)

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

3. Dilución de precisión vertical (VDOP en inglés)

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

4. Dilución de precisión temporal (TDOP en inglés)

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

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]),

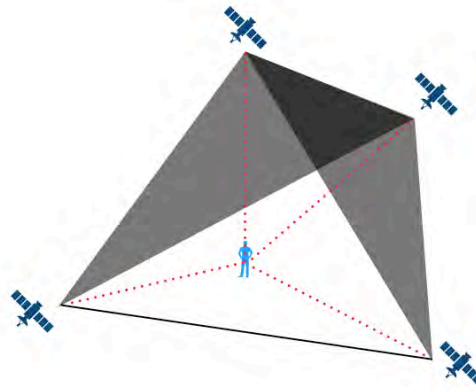


Figura 2.5: Si sólo se observan cuatro satélites, las puntas de los vectores unitarios receptor-satélite forman un tetraedro circunscrito por una esfera unitaria. Dos caras del tetraedro -formadas por un satélite en el cenit y tres en un ángulo de elevación de 10 grados, igualmente espaciados en acimut- 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 σ_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 σ_p o σ_{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

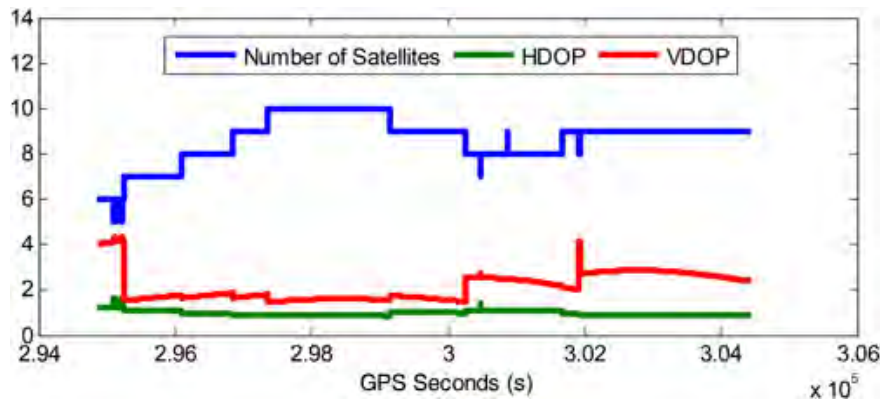


Figura 2.6: El número de satélites a la vista y el HDOP y VDOP calculado (de [29]).

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 1e-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

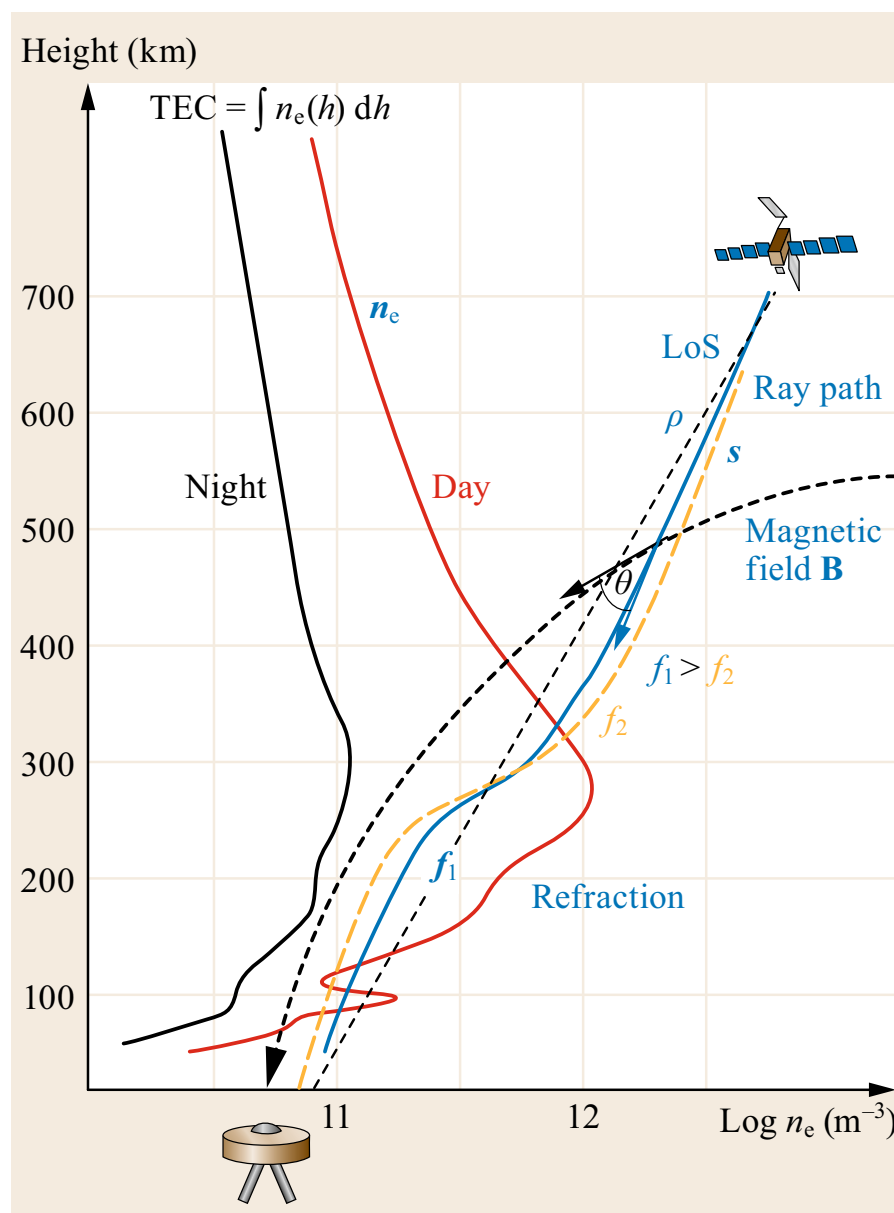


Figura 2.7: Esquema de propagación de la señal de RF a través de la ionosfera para dos frecuencias en presencia del campo geomagnético y el número de electrones (de [26]).

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.

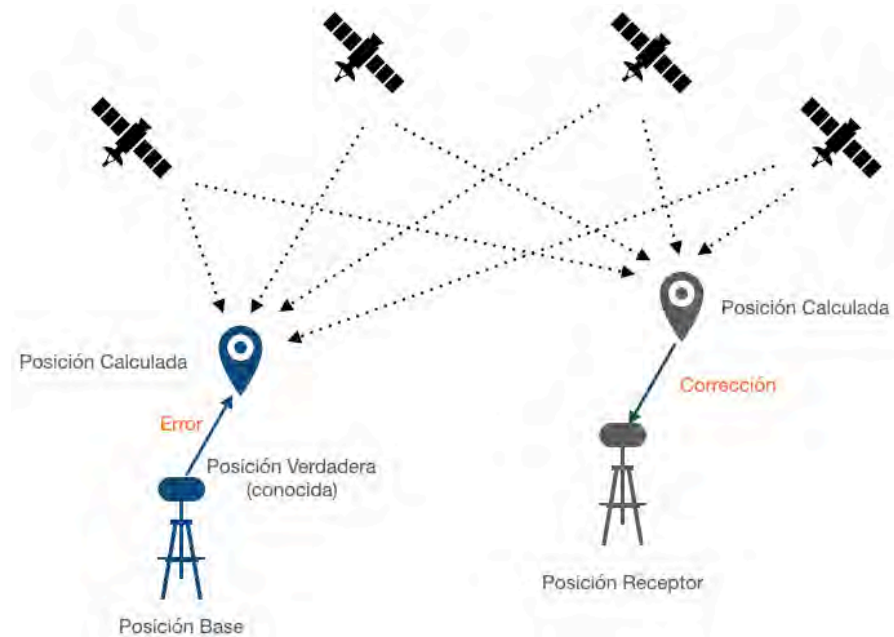


Figura 2.8: Esquema de funcionamiento del posicionamiento diferencial por código. La ER conociendo su posición determina el error en los rangos de los satélites y comunica estos errores al receptor para que corrija la posición calculada con menos error.

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 km² 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.

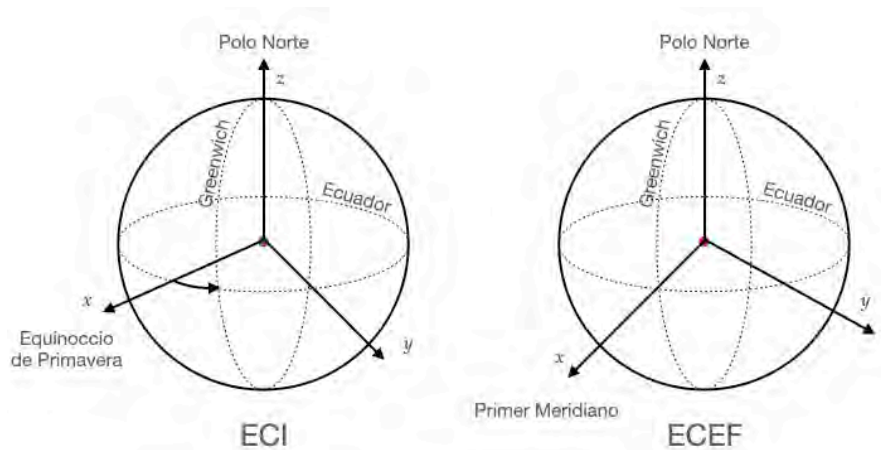


Figura 2.9: Sistemas de Referencia Inercial (ECI) y Terrestre (ECEF).

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° (pasando por la intersección del meridiano de Greenwich con el plano ecuatorial), y el eje $+y$ apunta en la dirección de 90° 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 aplastamiento (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 ϕ 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 λ 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-

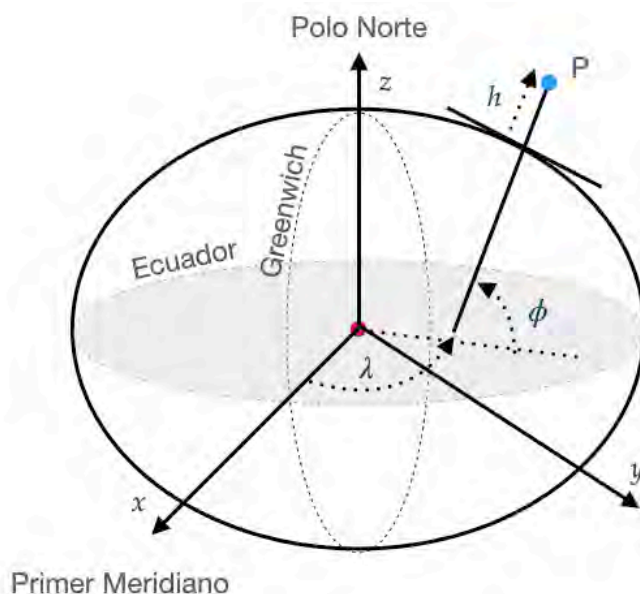


Figura 2.10: Coordenadas Elipsoidales

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

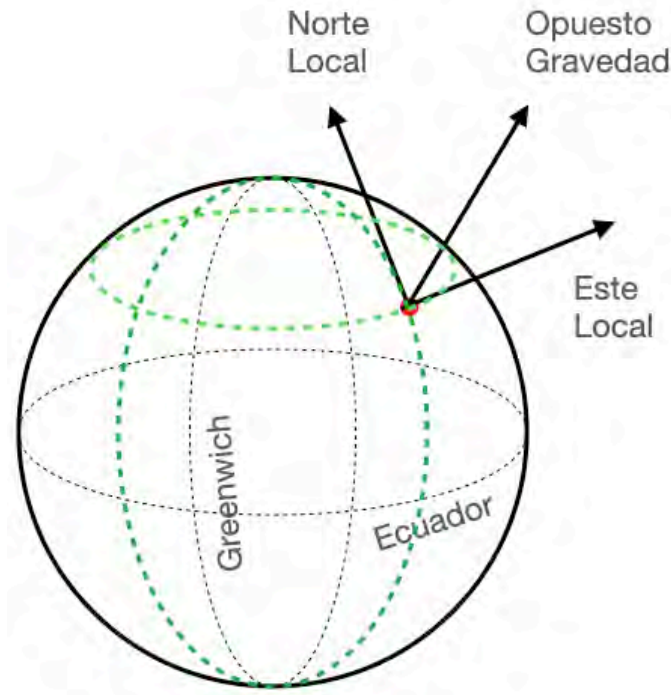


Figura 2.11: Coordenadas Este-Norte-Arriba

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.

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 μs . 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

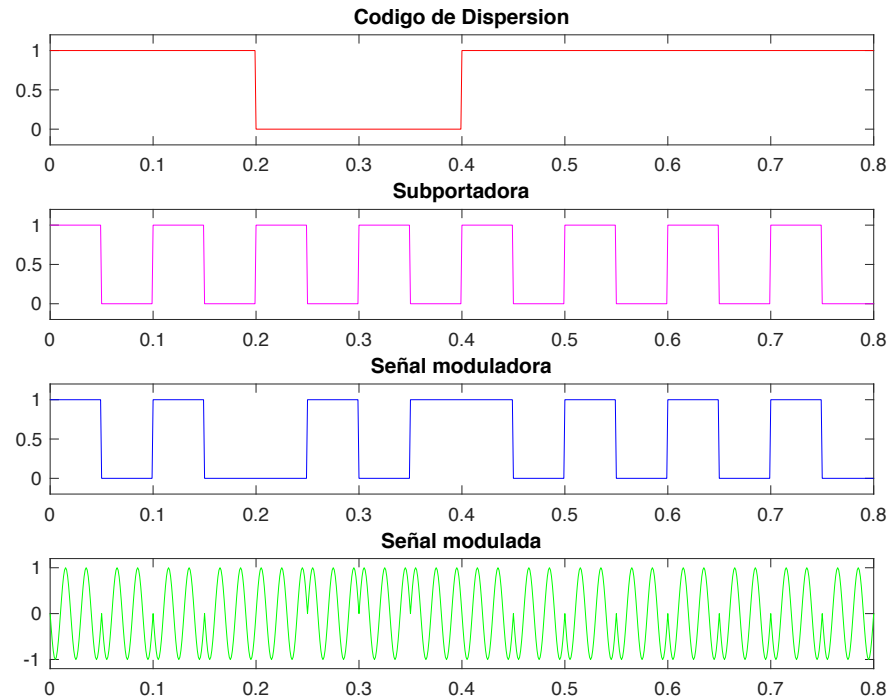


Figura 2.12: Ejemplo de señales para BOC(10,5)

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.

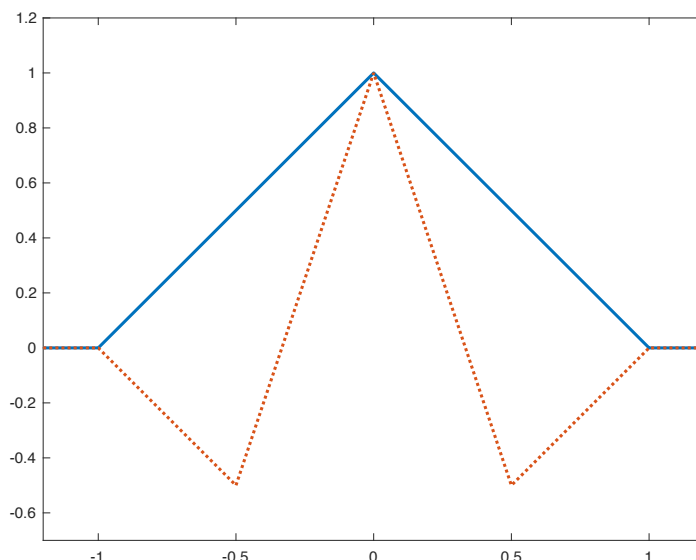


Figura 2.13: Ejemplo de correlación. En línea continua la correlación de un pulso de GPS y en línea discontinua la de señales BOC.

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 2492.028 MHz 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 extendaremos 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-

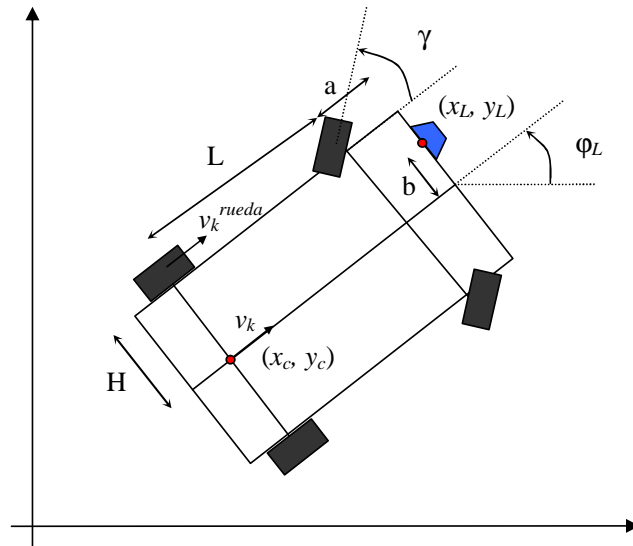


Figura 2.15: Descripción del vehículo. El punto (x_c, y_c) es el centro del eje trasero. El punto (x, y) es donde se ubican sensores a partir de los cuales se observa el ambiente (GPS, Lasers) y es donde se traslada el modelo del vehículo

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 γ 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:

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

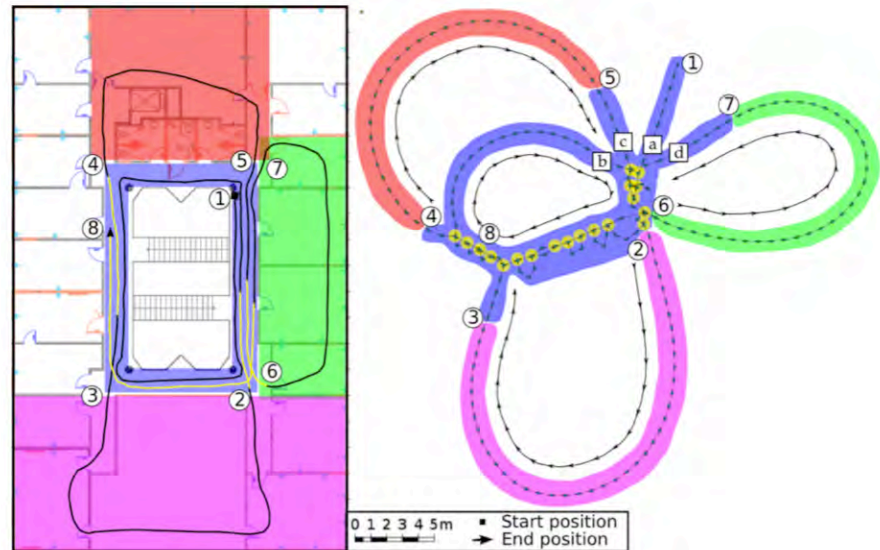


Figura 2.16: Un móvil recorrido un primer bucle en la zona azul, antes de entrar en la zona magenta. Después, la cámara vuelve a entrar en la zona azul y sigue recto hasta la zona roja. A continuación, vuelve de nuevo a la zona azul antes de descubrir la zona verde. El recorrido termina cerca del octavo 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.

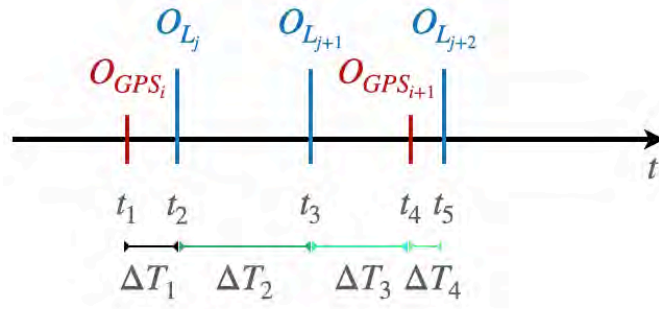


Figura 2.17: Línea de tiempo que representa los momentos donde se reciben las observaciones del GPS (O_{GPS_i}) y las del Lidar (O_{L_j}).

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 ΔT_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, φ_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^{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^{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}, γ_k) y su expresión es

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

donde

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

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_θ 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.

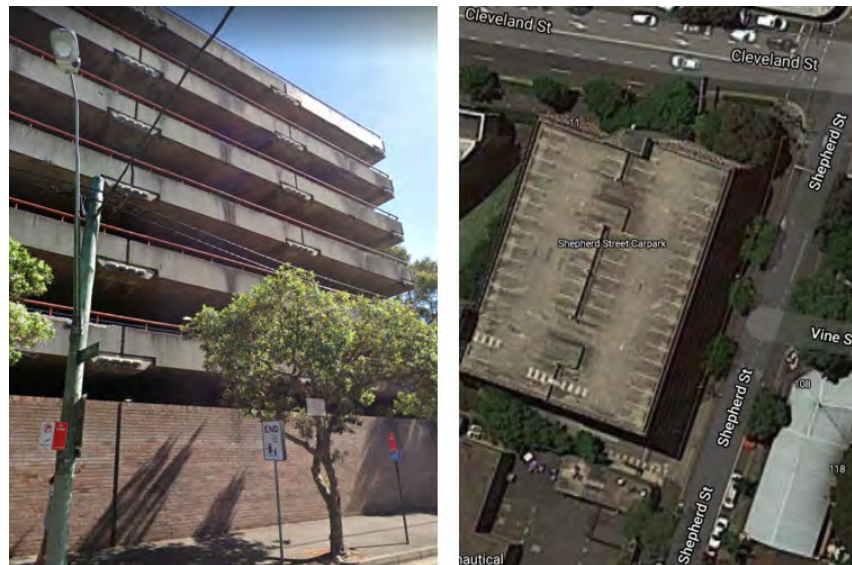


Figura 2.19: Vista aérea del techo de la cochera donde se realizó la primer experiencia. Sobre impuesto (señalado con una flecha) el trayecto recorrido por el móvil y la ubicación de los mojones.

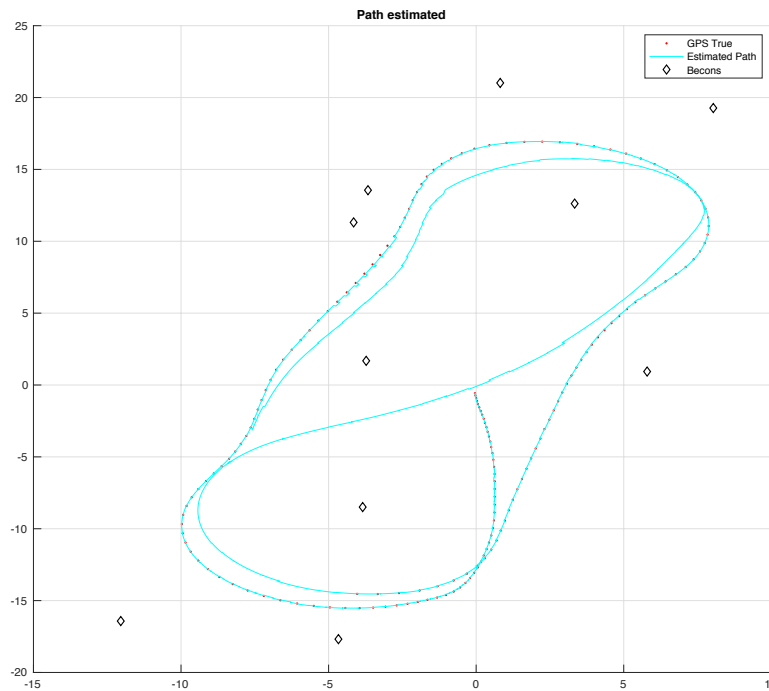


Figura 2.20: Resultado de la navegación utilizando EKF, un escáner láser y un GPS. Los ejes están en metros y el punto inicial de navegación corresponde al origen de coordenadas. En puntos se muestra la medida del GPS en modo RTK como referencia de la navegación, en línea continua se muestra el resultado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

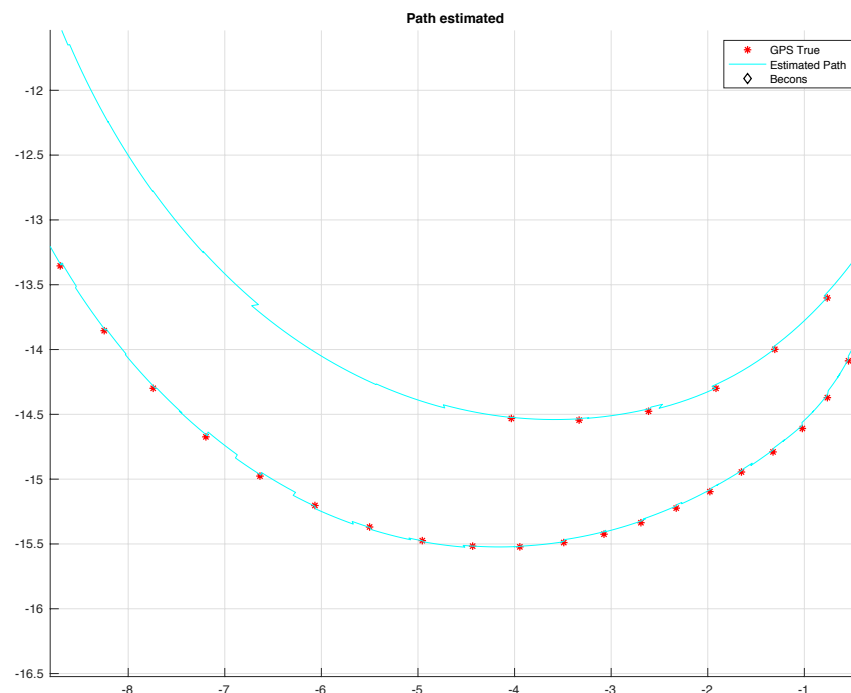


Figura 2.21: Ampliación del resultado de la navegación utilizando EKF, un escáner láser y un GPS. En puntos se muestra la medida del GPS en modo RTK como referencia de la navegación, en línea continua se muestra el resultado de la estimación. Los diamantes corresponden a la ubicación de los mojones en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

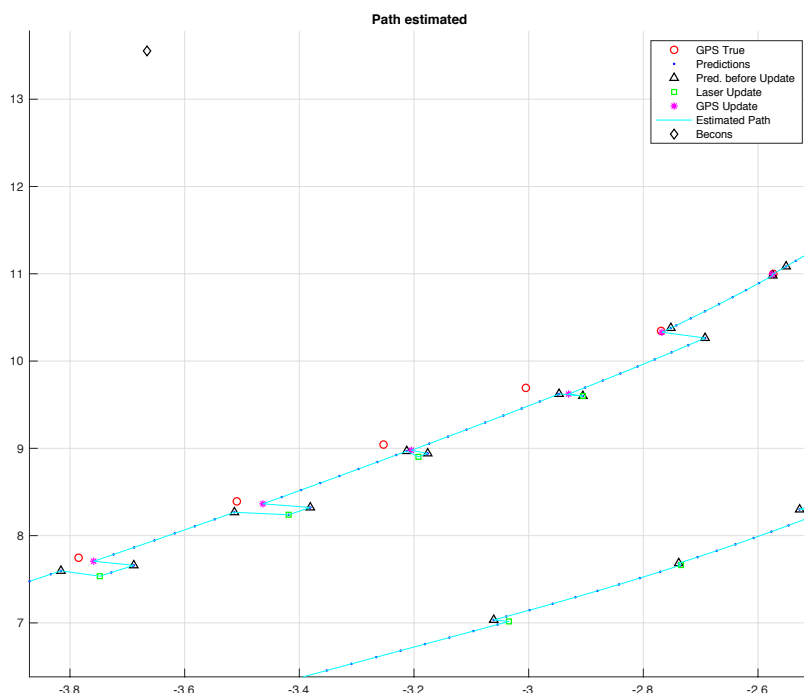


Figura 2.22: Navegación con las diferentes etapas del filtro mostradas en forma explícita. En línea continua se muestra el resultado final de la estimación. Los puntos negros representan la predicción hecha sin actualización y los triángulos negros la predicción antes de esta. Se puede observar claramente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

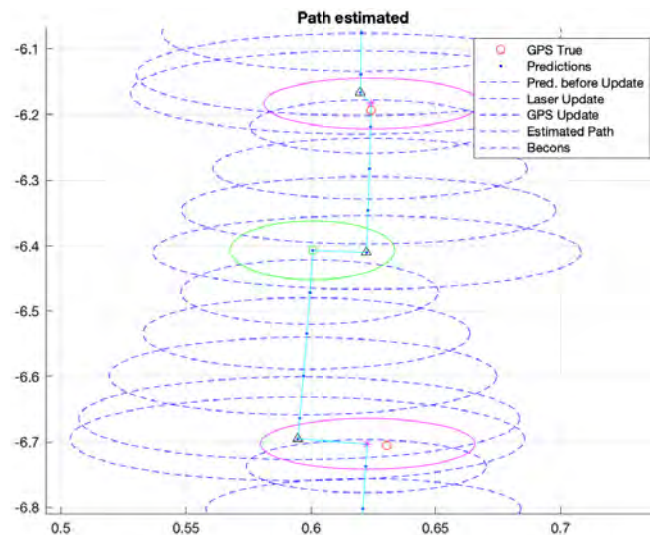


Figura 2.23: Representación aproximada del error de estimación. Las elipses representan el 95 % de confianza de la estimación. En azul la correspondiente a la predicción y en verde y magenta la resultante de la actualización con Láser y GPS respectivamente.

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 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, 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 mojón m_i como \mathbf{m}_{i_k} . Se define el vector de estados

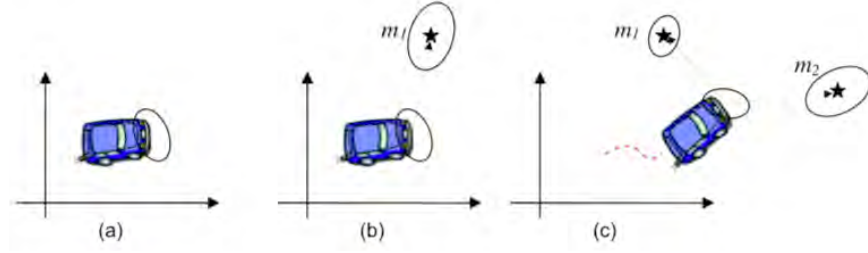


Figura 2.24: Incertidumbres y correlaciones en el SLAM. En (a) un vehículo posee una incertidumbre representada por una elipse. En (b) el vehículo observa un rasgo y lo incorpora al mapa como un mojón m_1 . El estimado de este mojón depende de la incertidumbre en la posición del vehículo con lo cual 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

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

donde

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

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 mojón 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,

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

De aquí se puede observar que las submatrices

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

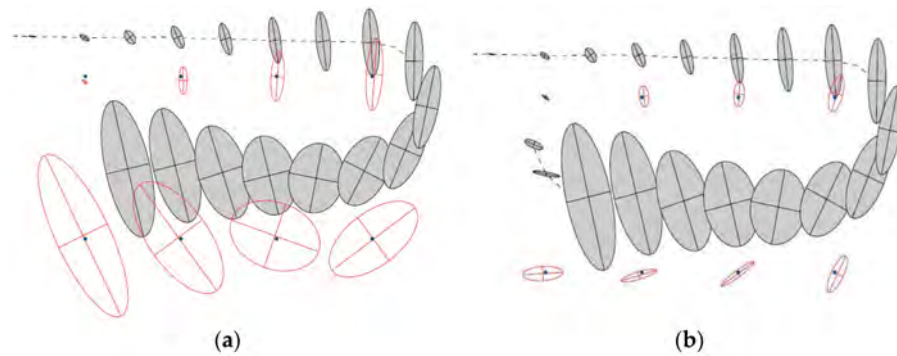


Figura 2.25: Cierre de lazo (de [28]). La figura (a) muestra el momento antes de cerrar el lazo. La figura (b) muestra la reducción de las incertidumbres de los mojones luego de cerrar el lazo y observar los mojones que habían sido incorporados al inicio de la navegación.

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

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

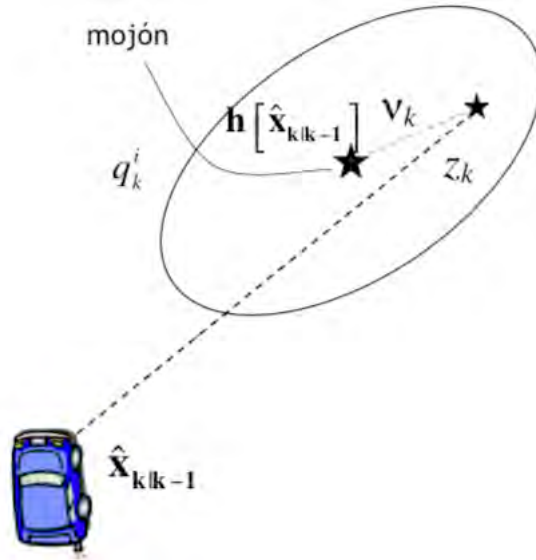


Figura 2.26: Cota de validación de la observación. La cota es definida en el espacio de observación y está centrada en la predicción de la observación $h[\hat{x}_{k|k-1}]$. La cota es la innovación normalizada y para un valor fijo de q_k^i es una elipsoide que está definida por los autovalores de la 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 χ^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.

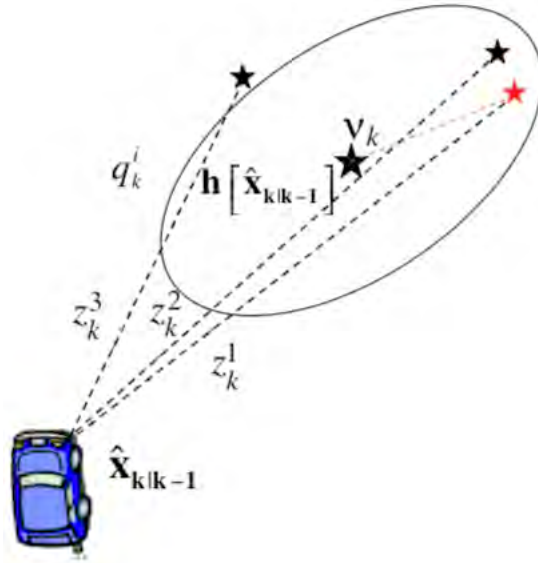


Figura 2.27: Para un solo objeto, el algoritmo del vecino más cercano selecciona la observación lo más cerca posible la predicción para la asociación, en este caso z_k^1 . Se rechazan el resto de las observaciones. La medida de la proximidad es la innovación normalizada.

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**

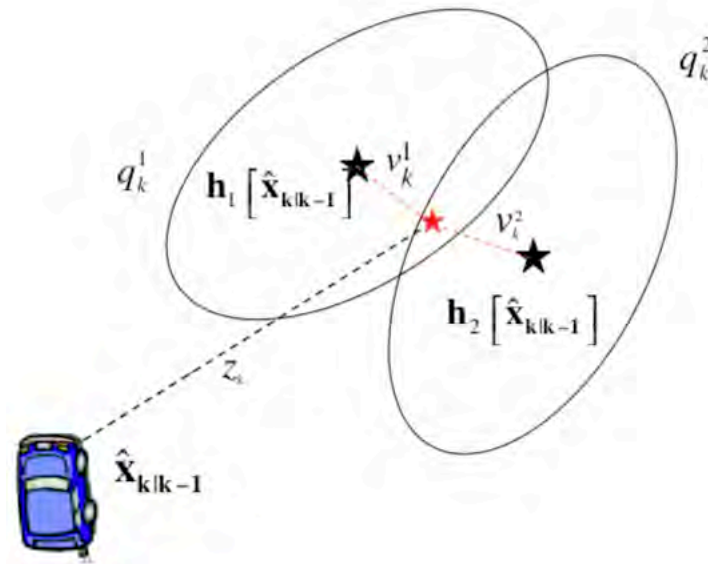


Figura 2.28: Para objetos múltiples, cada objeto selecciona su observación más cercana independientemente de otras posibles asociaciones. Esto puede conducir fácilmente a asignaciones erróneas en altas densidad de objetos. En el caso mostrado, una medida proveniente de un objeto en el ambiente es asociado 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 μ 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-

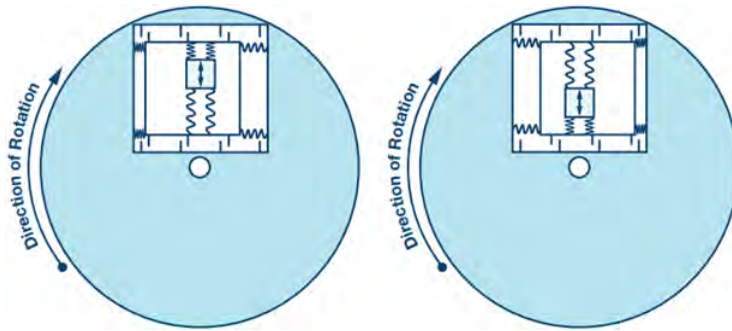


Figura 2.29: Esquema de un giroscopio MEMS. El efecto coriolis hace que la diferencia de fuerza que experimenta una masa mas o menos alejada del centro de una superficie que gira. Eso hace que la estructura vibrante se corra hacia los costados, que es medido por los *dedos* arriba y abajo

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 ω , 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

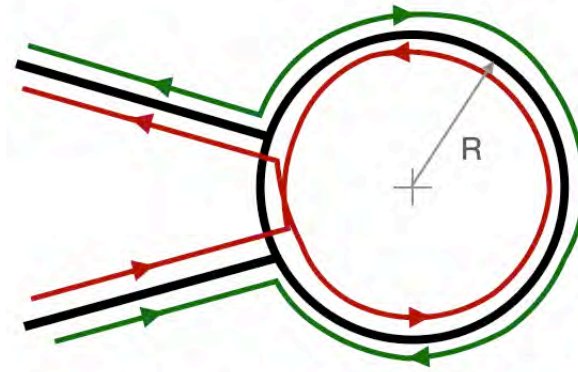


Figura 2.30: Esquema de una giróscopo de fibra óptica. El efecto Sagnac explica que ondas de luz que circulan en sentidos contrarios llegarán en momentos diferentes.

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]
ω_{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
ϵ_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 [m/s^2]
- \mathbf{f}^b es el vector de aceleraciones real [m/s^2]
- \mathbf{b}_a es el vector del sesgo del acelerómetro [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
- δ_g es la desviación del valor de gravedad teórico [m/s^2]
- ϵ_g es el vector que representa el ruido del sensor [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 (ω_e) en lugar de la gravedad. Para un dispositivo ubicado en una latitud ϕ_{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 Ω 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

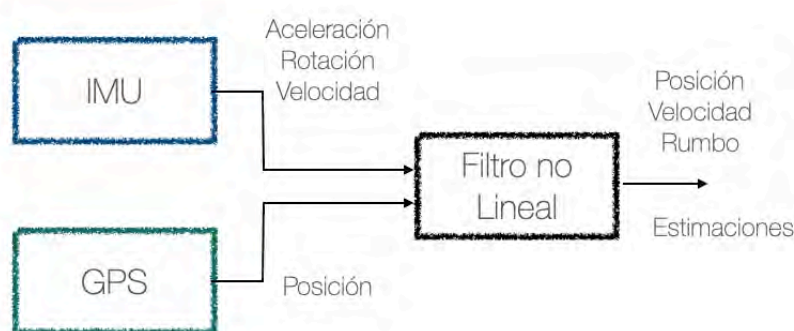


Figura 2.31: La estructura "directa" implementa un filtro no lineal para estimar la posición, velocidad y actitud del vehículo. Los datos de inercia los proporciona una IMU y los datos externos un GPS.

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

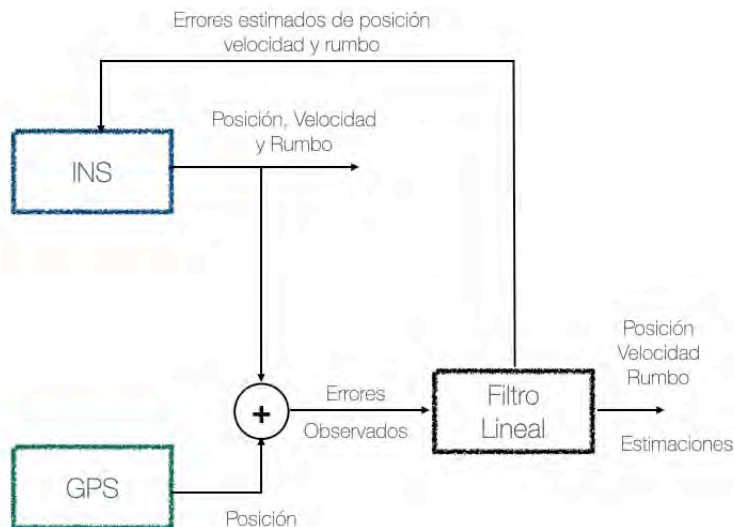


Figura 2.32: El método de realimentación directa resuelve los problemas del método directo.

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

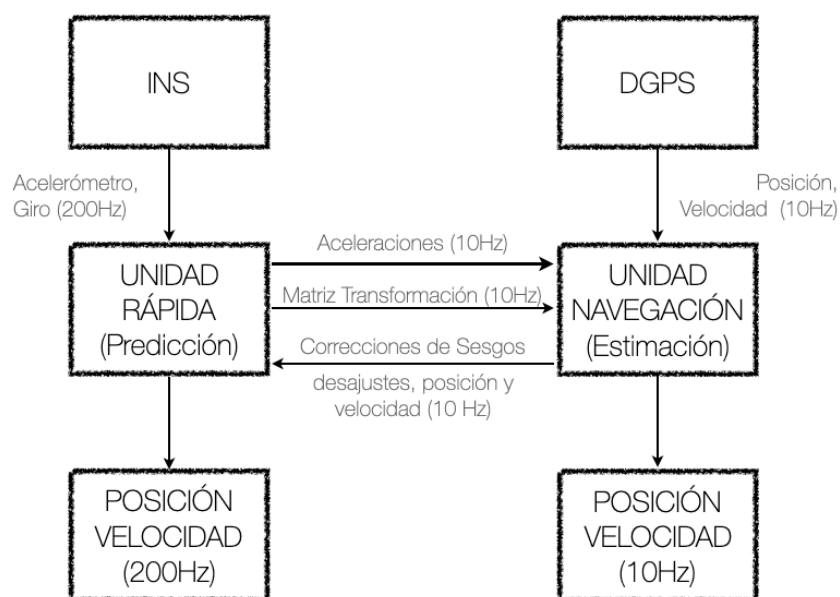


Figura 2.33: Arquitectura de Navegación por realimentación directa.

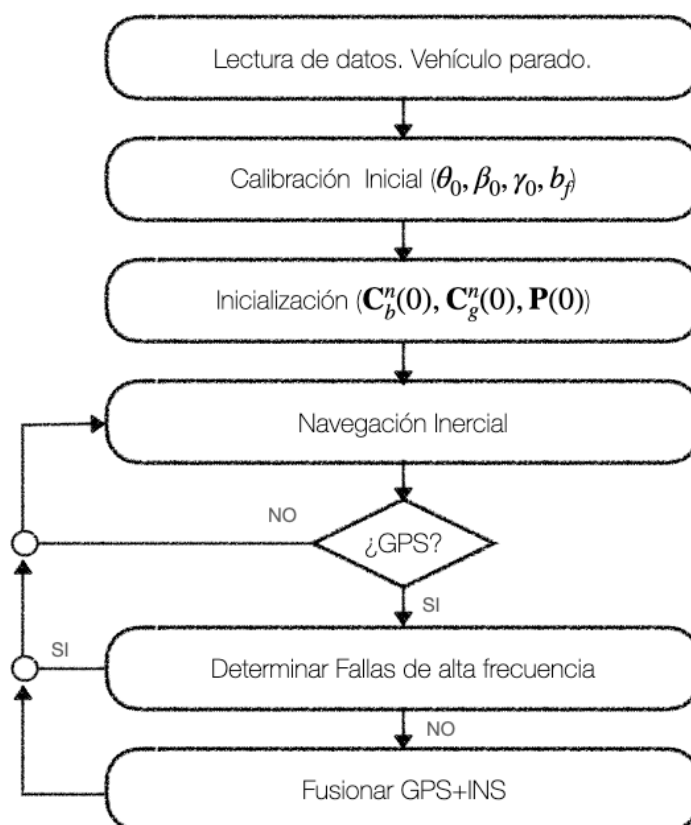


Figura 2.34: Flujo del algoritmo de Integración por realimentación directa.

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 3×3 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,

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

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

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

Si θ , β y γ 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 ω_x , ω_y y ω_z y se integran para determinar los cambios en ángulo ϕ_x , ϕ_y y ϕ_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 ϕ es la Latitud en grados y λ es la Longitud también en grados.

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

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

Si el área de trabajo del vehículo tiene un cambio insignificante en latitud y longitud, entonces \mathbf{C}_{ng} 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 θ , β y γ 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 β , y sustituyéndolo en las ecuaciones 2.119 o 2.120 se resolverá el balanceo θ . 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 γ . 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 τ 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 ϕ son los errores en posición, velocidad y rumbo y f_g y ω_{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 Δt 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 9×9 que representa la incertidumbre en los errores predichos inerciales. $Q(k)$ es la matriz de ruido de proceso discreta también de dimensión 9×9 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

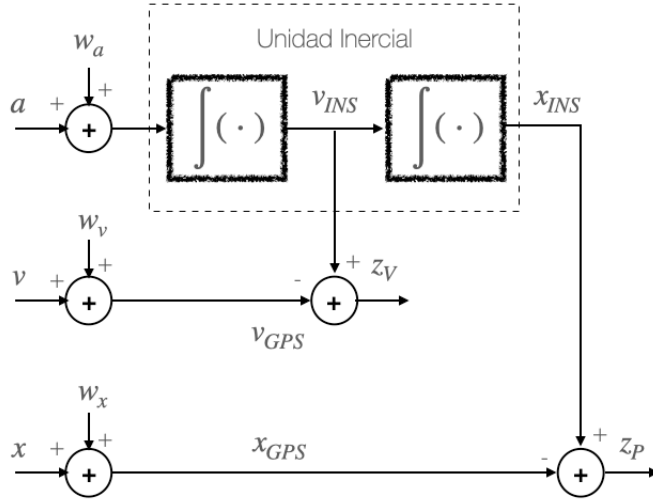


Figura 2.35: Ilustración de cómo se obtienen las mediciones de observación $\mathbf{z}_p(k)$ y $\mathbf{z}_v(k)$ mediante la información inercial y el GPS.

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 γ 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 χ^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 reflejada¹.

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

¹ 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 γ debe ser una muestra de una distribución χ^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.

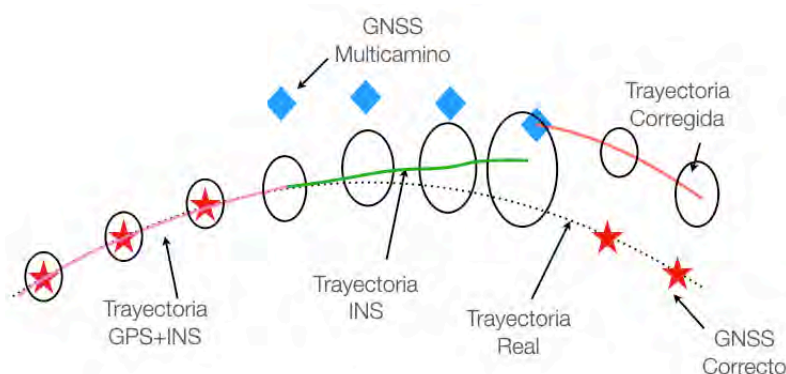


Figura 2.37: Errores de alta frecuencia. Las medidas correctas del GNSS se muestran como estrellas y las debidas al multicamino como rombos. Cuando una medida de GNSS errónea entra dentro de la región de activación representada por γ .

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 γ 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.

Parte III

POSICIONAMIENTO Y SEGUIMIENTO EN REDES INALÁMBRICAS

Se plantean soluciones de posicionamiento utilizando en forma oportunista la información de los protocolos de comunicación sobre redes inalámbricas como Bluetooth, Wi-Fi, LTE, etc.

Como hemos visto, conocer la ubicación es una capacidad muy importante para los teléfonos inteligentes. Estos dispositivos incluyen capacidad de comunicación, acceso a la red y sensores integrados y la información de ubicación se utiliza para proporcionar servicios basados en la ubicación (*LBS: Location Based Services* en inglés). En muchos casos se puede usar la localización basada en GNSS pero no en entornos interiores o de transición con el exterior. Hay estrategias desarrolladas que utilizan frecuencias de radio para la localización absoluta del usuario en función del despliegue de hardware especial en el entorno interior. Pero una alternativa es el uso de señales ya disponibles, las llamadas señales de oportunidad. Wi-Fi es una de las señales de oportunidad más utilizadas para el posicionamiento y el seguimiento de usuarios móviles debido a la disponibilidad de infraestructura ya implementada para comunicaciones y la capacidad de los dispositivos móviles para procesar estas señales. Hoy en día, se puede detectar una gran cantidad de puntos de acceso (AP) de redes públicas y privadas que brindan una alta ubicuidad de la señal.

Al investigar los sistemas de última generación y ya estandarizados, por ejemplo, GSM o LTE, se vuelve obvio que la información de posición no es a priori parte de los sistemas, sino que se ha visto más como un complemento de estas comunicaciones. Estos sistemas han sido diseñados únicamente con fines de comunicación. Sin embargo, alentados por los requisitos de reguladores (como la FCC de EEUU) y la demanda de información de ubicación sin GPS en los móviles, los sistemas se ampliaron para permitir también el posicionamiento independiente. Por esto, las organizaciones de estándares de telecomunicaciones trabajaron para incluir tecnologías de ubicación en sus estándares, por ejemplo, para GSM, UMTS, LTE, cdmaOne, CDMA2000, Wideband CDMA (W-CDMA) o WiMAX.

En esta parte veremos distintas formas de utilizar de forma oportunista para la localización, información disponible en redes de comunicaciones existentes y desarrolladas con el objetivo de transferir datos o comunicar equipos. Por ejemplo las redes celulares, Wi-Fi, Bluetooth, ZigBee, UltraWideband, RFID, NFC, entre otras.

3.1 REDES CELULARES [5]

Las redes celulares son la plataforma principal para los servicios inalámbricos basados en la ubicación (LBS). Esto es natural considerando la amplia distribución de teléfonos celulares y la cobertura de las

redes entre una gran parte de la población mundial. Las demandas de precisión de posición y los métodos utilizados para el posicionamiento dependen en gran medida de la naturaleza de estos servicios. La razón principal para agregar la capacidad de ubicación a la comunicación celular fue la seguridad física para los titulares de teléfonos móviles, al menos en los Estados Unidos, donde los proveedores de telefonía celular están obligados por las regulaciones de telecomunicaciones a brindar posicionamiento como un servicio sin suscripción. Una vez que la infraestructura y/o los modelos de terminales estuvieron disponibles para brindar ubicación, era natural que la gama de servicios basados en la ubicación comenzara a florecer. En Europa y otras regiones del mundo son estos servicios comerciales los que están generando la inclusión de la capacidad de localización en las redes celulares. Algunos de los servicios basados en la ubicación más comunes, además de la seguridad personal, son las instrucciones de navegación, la identificación de comercios cercanos, el rastreo de personas, animales y cosas, la facturación sensible a la ubicación, etc.

La localización no solo puede ser utilizada en los teléfonos sino también en para objetos o equipos a través de modems. Estos no solo pueden comunicar información sino también ubicarse con algunas de las estrategias de localización usando información de redes celulares que analizaremos a continuación.

3.1.1 *Identificación de la Celda (Cell-ID)*

La tecnología de posicionamiento más básica disponible para sistemas celulares es utilizando la identificación de las celdas (Cell-ID). Puede estar basado en un teléfono ¹ o basado en la red. Para llevar a cabo la comunicación, un teléfono está asociado con una base individual ubicado en una celda de red celular. Por supuesto, la estación base conoce la identidad y la ubicación de la celda. La identidad de la celda también la conoce el teléfono, que puede obtener la posición de la celda de la red. Este método de ubicación se conoce como proximidad. Su precisión básica depende únicamente de las dimensiones de la celda, pero se puede mejorar con el apoyo de otros métodos de ubicación. Se logra una reducción del área de ubicación y por lo tanto una mayor resolución en celdas que están divididas en tres o seis sectores por antenas direccionales de estación base. Esto se muestra en la figura 3.1.a donde EM es la estación móvil que se desea ubicar. Con antenas con un ancho de haz de 120° , el área de la celda y, por lo tanto, la inexactitud de la tecnología de identificación de la celda se reduce en aproximadamente un tercio. La precisión de la ubicación se puede mejorar utilizando la intensidad de la señal recibida (fig. 3.1.b). La intensidad de la señal varía considerablemente, debido al desva-

¹ o modem, a continuación nos referiremos a teléfonos pero en general queremos referirnos a un dispositivo que transmite información usando las redes celulares

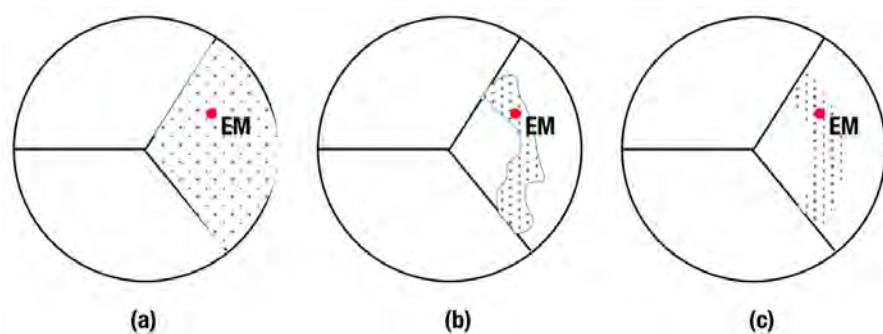


Figura 3.1: Método de identificación por celda y sus mejoras. De izquierda a derecha, por sectorizado de celda, utilizando RSS y con uso de temporizado

necimiento, la topografía, los patrones de antena, la potencia radiada y la frecuencia de funcionamiento.

Otra forma de limitar la incertidumbre de ubicación en la tecnología de ID de celda es usar el avance de tiempo que calcula la estación base en los sistemas GSM. El avance de tiempo es la corrección que la red hace a la temporización de la estación móvil para tener en cuenta el retraso de propagación de modo que los intervalos asignados a las EM no se superpongan (el método de acceso al medio de GSM es TDMA). La resolución del avance de tiempo en GSM es de $3,69\mu\text{s}$, equivalente a una distancia de 554m en un solo sentido. La Figura 3.1.c muestra cómo se reduce la incertidumbre de posición cuando el avance de tiempo se combina con la ID de celda para estimar la ubicación.

3.1.2 Localización usando TDOA

Se han desarrollado varias tecnologías de ubicación que se basan en el método TDOA de medición de distancia. Sus detalles dependen de si están basados en un teléfono o en una red, y cómo se aplican a un protocolo de red específico. En todos los casos dependen de la estimación del tiempo de llegada de la señal celular. Este se encuentra buscando la correlación cruzada máxima de una señal recibida con una réplica retrasada de la señal transmitida que es conocida. Dentro de las tramas de transmisión de señales celulares se encuentran secuencias de sincronización o entrenamiento que son conocidas por el receptor y que tienen baja autocorrelación cuando la señal y su réplica no están alineadas bit a bit.

Los relojes del teléfono y la estación base usualmente no están sincronizados, por lo que solo se pueden estimar los retrasos de tiempo relativos de las diferentes estaciones. En un sistema basado en un teléfono, este estima los retardos de tiempo relativos de las secuencias de entrada de tres o más estaciones base, en comparación con el tiem-

po de la secuencia conocida que se basa en el reloj del teléfono. A menudo no hay sincronización entre los relojes de diferentes estaciones base. Para estimar su propia posición mediante el método TDOA de hipérbolas (en dos dimensiones) el terminal debe recibir de una fuente externa los tiempos de transmisión de las secuencias de cada una de las estaciones base según un reloj común, o las diferencias entre ellas, y también las posiciones geográficas de esas estaciones base. Como alternativa, el teléfono puede enviar sus diferencias de tiempo observadas a un terminal fijo especial que tiene la información de posición de la estación base y el tiempo y puede usar esas diferencias de tiempo observadas para calcular la posición del teléfono.

En GSM se utiliza un método denominado *Enhanced Observed Time Differences*. En este caso, el móvil mide las diferencias de tiempo y las comunica a la *Mobile Location Center* que con la información de las estaciones base, calcula la posición del móvil. En UMTS el método implementado es el del *Observed Time Difference of Arrival*, similar al sistema GSM donde la posición es obtenida por una *Radio Network Controller* que sirve al móvil.

3.1.3 Localización usando Ángulo de Arribo

El método del ángulo de arribo es práctico para su consideración solo en aplicaciones basadas en red, ya que requiere conjuntos de antenas direccionales para su funcionamiento. Incluso entonces, no se puede esperar una adopción a gran escala debido al costo de reemplazar o superponer antenas de estaciones base celulares de segunda y tercera generación instaladas. Sin embargo, la tecnología es particularmente atractiva en circunstancias y entornos particulares. Mientras que para los arreglos TDOA se necesitan tres y preferentemente más estaciones base, para AOA sólo se necesitan dos estaciones base con antenas direccionales. Las regiones rurales, donde las celdas son grandes y, a menudo, no más de dos estaciones base pueden escuchar un terminal móvil, AOA puede ser una solución útil. Además, las rutas de línea de vista desde el móvil a las estaciones base son imprescindibles para este método, y las áreas urbanas pueden ser muy inadecuadas para AOA. AOA podría combinarse con un método de medición de distancia, como TOA, para brindar una solución de posicionamiento unilateral. Se espera que el despliegue de antenas inteligentes en redes celulares de generaciones más modernas aumente el interés en AOA para el posicionamiento celular.

3.1.4 Patrones de Potencia de señal recibida

Los datos de intensidad de la señal recibida, utilizados por sí solos, no son útiles para el posicionamiento celular para la mayoría de las aplicaciones. Sin embargo, cuando se puede armar una base de datos

para una región definida, como se describe en 1.11, se puede obtener una precisión satisfactoria. El sistema puede estar basado en un teléfono o en una red. En el primer caso, las lecturas de RSS obtenidas de varias estaciones base se envían desde el móvil a un servidor de ubicación para la búsqueda y comparación de bases de datos. Particularmente en el caso de un sistema basado en red, la base de datos puede contener, además de la intensidad de la señal recibida, información relacionada con otros parámetros de señal y canal, por ejemplo, respuesta de impulso de canal en una ubicación particular. La tarea principal en un sistema de reconocimiento de patrones es compilar y mantener la base de datos. La compilación se realiza mediante mediciones directas en ubicaciones en toda el área de cobertura, mediante cálculos basados en topografía conocida, o ambos. El método tiene el potencial de brindar una precisión utilizable en lugares donde otros métodos de posicionamiento dan resultados deficientes, donde no hay una ruta de línea de vista, por ejemplo, común en ubicaciones urbanas e interiores.

3.1.5 Posicionamiento en LTE

Originalmente, LTE especificó el soporte para tres métodos de posicionamiento: identificación de celda mejorada (o E-CID y mostrado en la figura 3.1.c), OTDOA y GNSS asistido. El cuarto método, basado en TDOA pero resuelto desde la red y no desde el móvil, se incluyó en versiones posteriores. Los cuatro métodos fueron ya descritos a lo largo del curso.

El posicionamiento del móvil involucra dos pasos: (1) medición de la señal y (2) estimación de la posición (y, opcionalmente, la velocidad) a través de cálculos basados en las mediciones. Las señales se miden en el móvil o la estación base. Los cálculos se realizan en el móvil o en la red (en el *Serving mobile location center*). Usamos los términos **basado en teléfono** y **basado en red** en cada uno de estos casos y para clasificar los sistemas de posicionamiento usados en redes celulares. La palabra *basado* significa que el nodo es responsable de realizar el cálculo de posicionamiento y puede (pero no necesariamente) también realizar las mediciones. Un nodo asistido proporciona las medidas pero no realiza los cálculos.

Hay dos formas en que el cliente puede comunicarse con el servidor de ubicación. Una es usando la conexión de datos estándar que se conecta a la plataforma de ubicación con un protocolo de posicionamiento de propósito general desarrollado por Open Mobile Alliance (OMA). La segunda es en los canales de control de la red. Ambas son entidades lógicas y pueden estar ubicadas en el mismo servidor físico y se considera más confiable y robusta en una situación de emergencia, particularmente cuando hay congestión en los canales de la red.

3.2 REDES INALÁMBRICAS DE CORTO ALCANCE Y RFID

Las redes inalámbricas que se analizan en esta sección tienen varias características notables que son significativas desde el punto de vista del posicionamiento inalámbrico. Todos funcionan con baja potencia en distancias relativamente cortas, generalmente hasta 100 m. En consecuencia, para los métodos de tiempo de llegada, el tiempo de vuelo es corto y para obtener precisiones útiles, la resolución de tiempo no debe ser superior a 10 ns, equivalente a un rango de 3 m. Además, todas las redes se utilizan básicamente en interiores, lo que somete las señales inalámbricas a condiciones severas de rutas múltiples y sin visibilidad directa. Estos sistemas operan en las bandas sin licencia y, por lo tanto, son susceptibles a la interferencia de una amplia gama de tipos de señales. Los terminales móviles son pequeños y deben tener un consumo de corriente muy bajo para alargar la vida de la batería. También están usualmente diseñados con bajo costo. Las tecnologías para incorporar servicios de localización en estas redes deben tener en cuenta todas las características mencionadas anteriormente.

3.2.1 WLAN/Wi-Fi

Probablemente la mayoría de los sistemas de localización que utilizan Wi-Fi se basan en la proximidad. Existen empresas llamadas agregadores de ubicación que compilan bases de datos escaneando las señales de Wi-Fi existentes en la vía pública, anotando las MAC y/o el SSID de los Access Points (AP) y asociándolos a una ubicación estimada obtenida en un vehículo con una antena y un GNSS. Debido a que las direcciones MAC transmitidas no están encriptadas, las ubicaciones de los puntos de acceso comerciales y privados pueden estar disponibles para los servicios de ubicación, generalmente sin el conocimiento y permiso de la fuente. Los suscriptores del servicio prestado por estas empresas utilizan cualquier dispositivo móvil equipado con Wi-Fi y pueden obtener una indicación de su ubicación, dentro del rango de cobertura de Wi-Fi de varias decenas de AP, simplemente usando la RSS y el ID y luego consultando la base de datos, que puede ser de alcance internacional. La amplia distribución de puntos de acceso Wi-Fi, particularmente en áreas urbanas, hace que estos servicios sean útiles con precisiones de hasta alrededor de 10 m.

Un método común de posicionamiento de WLAN es por lo tanto el de Fingerprinting basada en RSS, discutido en la sección 1.11. Sin embargo, para evitar la necesidad de prepararse con anticipación, actualizar las bases de datos y mejorar la precisión, existen propuestas basadas en TOA o TDOA. Las tecnologías utilizadas se dividen además por aquellas que utilizan dispositivos comerciales estándar sin modificación de hardware pero con software o firmware especial,

y aquellas que recurren a modificaciones de hardware o hardware especialmente diseñado para la función de posicionamiento.

3.2.1.1 TOA

La medición de distancia por tiempo de vuelo unidireccional requiere que el receptor que mide y estima el tiempo de llegada de la señal de un transmisor distante, conozca el tiempo de transmisión de esa señal en términos de su propio reloj. Para ello, los relojes del receptor y del transmisor deben estar sincronizados con el grado de precisión requerido de la medición de la distancia. El estándar Wi-Fi IEEE 802.11 especifica la sincronización del reloj mediante un temporizador de función de sincronización de tiempo (TSF) que cuenta en incrementos de microsegundos. El mecanismo descrito en el estándar está diseñado para mantener la sincronización de todos los temporizadores TSF en una red dentro de $4\mu\text{s}$, más el retardo de propagación. La precisión del temporizador, $1\mu\text{s}$, equivale a una distancia de 300 m, que es demasiado grande para el uso directo en la red, donde los terminales normalmente están separados por solo decenas de metros. Sin embargo, cuando se emplea TOA bidireccional (RTT: *Round Trip Time*), la sincronización del temporizador de transmisión y recepción no es necesaria. En este caso, el iniciador/interrogador transmite un mensaje a un segundo terminal, que responde con un mensaje de retorno.

Lo que hace que la medición de distancia TOA bidireccional sea una posibilidad en redes Wi-Fi es que una respuesta de mensaje después de un intervalo de tiempo fijo desde la recepción del mensaje puede generarse automáticamente en el hardware del terminal como parte del protocolo de acceso. El método de acceso del estándar IEEE 802.11 se conoce como acceso múltiple con detección de portadora y evitado de colisiones (CSMA/CA). Una estación que tiene un mensaje para transmitir y desea obtener acceso al canal debe asegurarse de que el canal esté libre antes de transmitir. Lo hace monitoreando el canal y solo si está despejado por un período de tiempo requerido puede intentar acceder. La probabilidad de colisiones entre transmisiones de dos o más terminales se reduce mediante un procedimiento que rige la interrupción aleatoria o períodos de espera adicionales, una vez que un terminal descubre que el canal parece estar libre. Se hace una excepción al requisito de monitoreo del canal antes de la transmisión en el caso de mensajes de acuse de recibo (ACK). El propósito del ACK es informar al remitente que su mensaje fue recibido correctamente. Si no recibe el ACK durante un período de tiempo determinado después de que se envió su mensaje, puede intentar enviarlo nuevamente, mientras observa la rutina prescrita para evitar colisiones. La función de mensaje-ACK reduce en gran medida la posibilidad de perder datos, incluso en redes densas y canales de transmisión de RF difíciles.

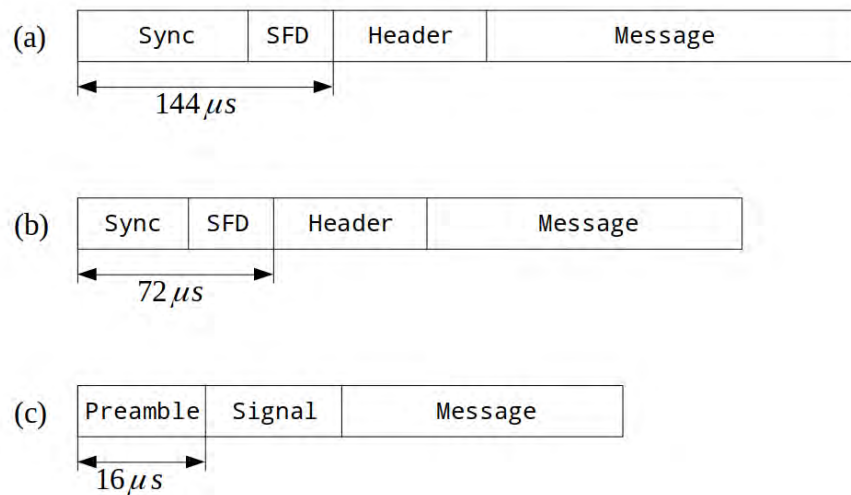


Figura 3.2: Formatos de las tramas de la capa física del estándar IEEE 802.11; (a) formato largo, (b) formato corto y (c) formato IEEE 802.11a

Dado que el ACK se envía sin verificar si el canal está ocupado o no, el protocolo tiene disposiciones para protegerlo de posibles colisiones. Un aspecto de esta protección es asignar al ACK un período de tiempo, llamado espacio entre tramas corto (SIFS), entre el final del mensaje recibido y el comienzo de la transmisión del ACK. SIFS es más corto que el período mínimo antes de que un terminal comience la transmisión, por lo que otro terminal que desee acceder al canal escuchará la trama ACK y pospondrá su transmisión de acuerdo con las reglas. La respuesta automática del mensaje ACK después de un retraso conocido es la base de la medición de distancia TOA bidireccional en IEEE 802.11.

La medición de la distancia TOA depende de la determinación precisa del tiempo en un momento específico de un paquete recibido. En la figura 3.2 se muestran tres tipos de formatos de trama de capa física IEEE 802.11. Todos tienen un preámbulo que facilita la sincronización de bits de la trama recibida e incluye un delimitador de trama de inicio (SFD) o equivalente, que indica el comienzo de la cabecera de trama. El comienzo de la cabecera puede ser el mejor momento para referirse a él como el momento de recepción de la trama, aunque cualquier otro punto podría utilizarse para contar los períodos de chip o símbolo.

El principio TOA bidireccional se muestra en la figura 3.3. Una trama enviada desde el terminal Wi-Fi iniciador a un respondedor llega después de un retraso de propagación de T_p . Si la trama se recibe correctamente, se envía una transmisión ACK al iniciador. La diferencia de tiempo entre el final del mensaje recibido y el comienzo del acuse de recibo, ACK, debe ser menor que SIFS, que en 802.11b y g tienen un valor nominal de $10\mu s$. Este tiempo incluye retrasos de

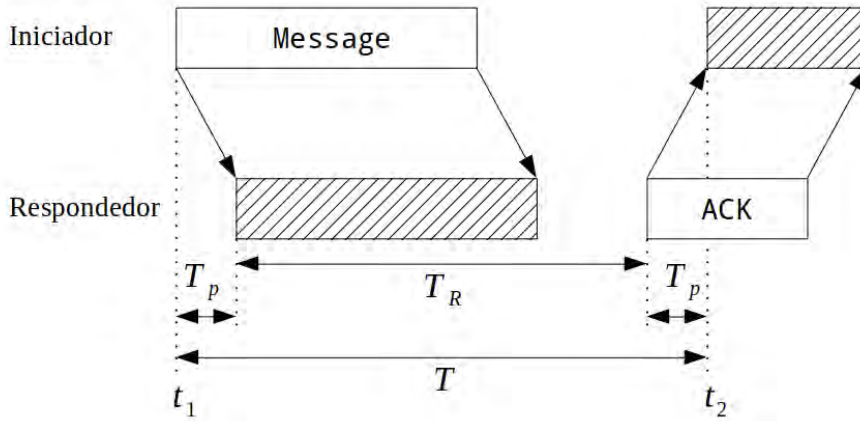


Figura 3.3: Proceso de TOA bidireccional en WLAN

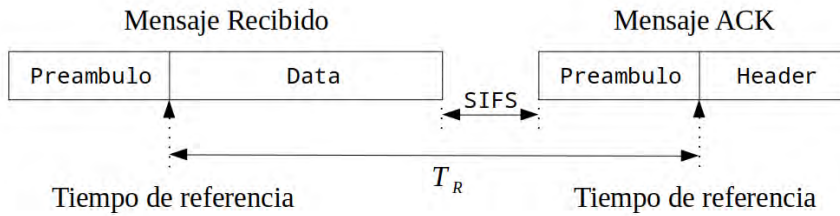


Figura 3.4: Trama de datos mas la de ACK

hardware y procesamiento. El ACK es enviado automáticamente por el hardware del chip y el iniciador lo recibe después de un retraso de propagación de T_p . De la Figura 3.3 se ve que el tiempo total de propagación es:

$$2T_p = (t_2 - t_1) - T_R \quad (3.1)$$

En la Figura 3.3, la marca del comienzo de un mensaje o trama ACK es el final del preámbulo. Como se ve en la figura 3.4, T_R es igual a la duración total de la trama del mensaje de datos menos el preámbulo, más el SIFS, más la longitud del preámbulo de la trama ACK de respuesta. La precisión de la estimación de T_p depende en su mayor parte de la precisión de SIFS.

Hay varias fuentes de incertidumbre que dificultan obtener la precisión deseada cuando se cumple con el estándar IEEE 802.11. Los tiempos t_1 , t_2 y T_R (figura 3.3) cuando se miden en la capa física tienen la precisión del reloj del chip. Para IEEE 802.11b la fuente de reloj típica tiene una frecuencia de 44 MHz, con un periodo de 22,7 ns, equivalente a una distancia de 6,8 m. El SIFS es nominalmente de 10 μ s y puede variar en diferentes implementaciones. Esta es quizás la mayor fuente de incertidumbre. Las bases de tiempo inexactas y el hecho de que las frecuencias de reloj de transmisión y recepción

a menudo no están sincronizadas también agrega imprecisión que es una función de la longitud del paquete. Por ejemplo, supongamos que el reloj del receptor está sincronizado con el reloj de transmisión mediante un PLL con el oscilador local durante el preámbulo del paquete. Si la frecuencia del reloj difiere de la nominal en 20 ppm, y si el valor verdadero de $t_2 - t_1$ es digamos, 300 μ s, entonces el error en la medición debido a la imprecisión del reloj es:

$$\Delta T_{\text{error}} = 300\mu\text{s} \times 20 \times 10^{-6} = 6\text{ns} \quad (3.2)$$

Esto, para una velocidad de propagación de 3×10^8 m/s, significa un error de distancia de 1.8m.

Mientras que el intervalo de tiempo ($t_2 - t_1$) se mide en el lado del iniciador, la estimación más precisa de T_R tendría que determinarse en el lado del respondedor del enlace. Esto se debe a que el respondedor determina, en hardware, el SIFS. Cuando el respondedor mide T_R , podría transmitir el resultado de la estimación al iniciador para el cálculo de $(t_2 - t_1) - T_R = 2T_P$. Sin embargo, en muchos casos, esto es indeseable o imposible. Por ejemplo, si se pretende que la estimación de la distancia se utilice para la seguridad contra el acceso no deseado a la red, el intruso no cooperará. Sin embargo, el iniciador puede aproximar T_R . Conoce la longitud del mensaje y del preámbulo ACK y puede suponer que un tiempo SIFS es el mismo que en el terminal iniciador, en cuyo caso se cancelará exactamente. Si este fuera el caso, la estimación del tiempo de propagación sería, a partir de la ecuación 3.1 y la figura 3.3:

$$T_P = \frac{1}{2} [(t_2 - t_1) - (T_{\text{Data}} + \text{SFIS} + T_{\text{Preambulo}})] \quad (3.3)$$

Las estimaciones de rango se pueden mejorar con relojes de baja precisión aumentando el número de mediciones y por lo tanto el tiempo de medición. La precisión se mejora promediando un conjunto de medidas hechas con varios mensajes de datos y ACK (>100) en forma consecutiva.

Son varias las ventajas de usar este método sobre otros en redes WiFi:

- Incluso con un AP, se puede determinar el rango con una precisión de una fracción de metro en dispositivos que tienen relojes de alta precisión.
- Se puede encontrar una ubicación bidimensional inequívoca a partir de la triangulación utilizando un mínimo de tres puntos de acceso. Las mediciones de tiempo simultáneas de los AP no son necesarias cuando el movimiento del móvil no es significativo y no existe un requisito de sincronización.
- Usando el promedio, se puede lograr una resolución TOA mucho mejor que la del reloj local.

- No se necesita conocimiento de la potencia radiada.

Pero, existen dos limitaciones muy importantes:

- En interiores el multicamino reduce la precisión
- Se requiere un tiempo de medición relativamente largo para promediar lograr una alta precisión cuando no se dispone de un reloj de alta resolución.

3.2.1.2 *Fingerprinting*

La técnica de posicionamiento basada en Fingerprinting es ampliamente utilizada para WLAN. Se basa en un conjunto de medidas de RSS tomadas de un objetivo desde múltiples puntos de acceso y la comparación de los resultados con una base de datos compilada previamente (como vimos en 1.11). El método tiene varias ventajas. No se requiere sincronización de tiempo. La lectura de RSS es inherente al protocolo IEEE 802.11 y no se necesita ningún hardware especial. Se pueden rastrear etiquetas o dispositivos basados en el estándar IEEE 802.11. El método es especialmente aplicable a las redes de interior, ya que los problemas de la propagación por trayectos múltiples se tienen en cuenta automáticamente en la base de datos de referencia. En lo negativo, el método implica la creación de una base de datos para el área a cubrir, y los cambios en la implementación de AP y las características físicas del entorno requieren la actualización de la base de datos.

3.2.2 *WPAN/Bluetooth*

Así como el IEEE 802.11, o Wi-Fi, domina la WLAN, existen varias tecnologías y estándares que conforman la WPAN pero Bluetooth es la mas común de todas. Estas redes no tienen la infraestructura que forma parte de la mayoría de las aplicaciones WLAN y el alcance es generalmente mucho más corto, nominalmente 10 m. Los estándares WPAN difieren significativamente en sus capas físicas y además de Bluetooth (IEEE 802.15.1) se pueden nombrar ZigBee o IEEE 802.15.4 o Wireless-USB.

Un método para realizar estimaciones aproximadas de la ubicación del dispositivo Bluetooth utiliza una aproximación del alcance máximo de un enlace. Los servidores de posición de Bluetooth están ubicados en un área donde se va a estimar la posición de un cliente (en la figura 3.5 las estaciones A, B C y D). Estos servidores están programados para dar sus coordenadas de ubicación a un cliente solicitante. También se pueden utilizar dispositivos Bluetooth que no tengan una programación especial para manejar una solicitud de posición. En este caso, la posición del dispositivo está contenida en una

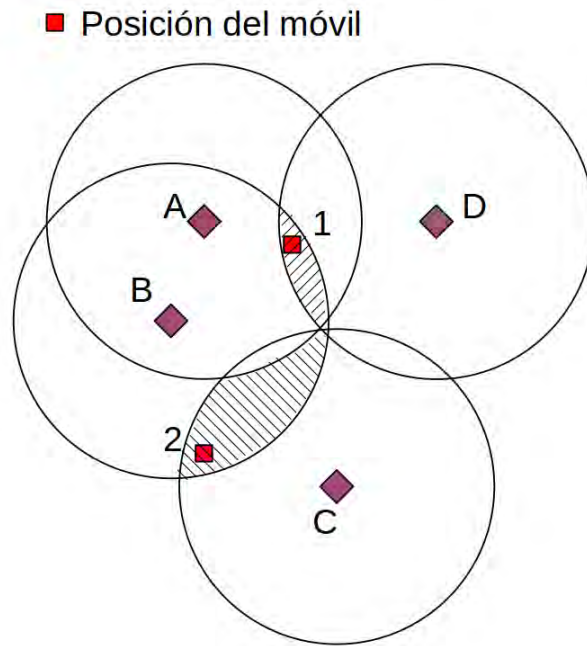


Figura 3.5: Localización Bluetooth usando el área máxima de cobertura. Cada estación base (A,B,C y D) cubre un área máxima (10m) y la ubicación está determinada por la intersección de las áreas que conectan al móvil

tabla referenciada por la ID del dispositivo fijo a la que puede acceder un cliente desde un servidor de ubicación especial que se ha configurado para este propósito. La clase de potencia 3 especificada tiene una salida de potencia máxima de 0 dBm, lo que permite un alcance de alrededor de 10 m. El método de posicionamiento asume que si se realiza una conexión, el dispositivo cliente está dentro de los 10 m del servidor. Se obtiene una mayor precisión como se muestra en la figura 3.5 cuando el cliente realiza una conexión con dos o más servidores de posición o dispositivos cuyas ubicaciones están en la tabla de servidores de ubicación. El cliente en la posición 1 ha establecido conexiones con los dispositivos A, B y C cuyas coordenadas se pueden obtener. El dispositivo D está fuera de rango. El cliente se encuentra en la región sombreada. Las coordenadas de posición estimadas se pueden encontrar calculando el centroide de la región, o una aproximación a partir de las intersecciones de los círculos de radio constante que delimitan la región. Cuando el cliente se mueve a la posición 2, está dentro del alcance solo de los dispositivos C y D, y está ubicado en la región sombreada correspondiente.

Para realizar estimaciones de ubicación, un cliente móvil debe intentar conectarse a varios dispositivos Bluetooth y luego recibir coordenadas de referencia de aquellos con los que la conexión fue exitosa. El tiempo promedio para hacer una estimación de ubicación cuando se consultaron cinco dispositivos base es de entre 20 y 30 segundos.

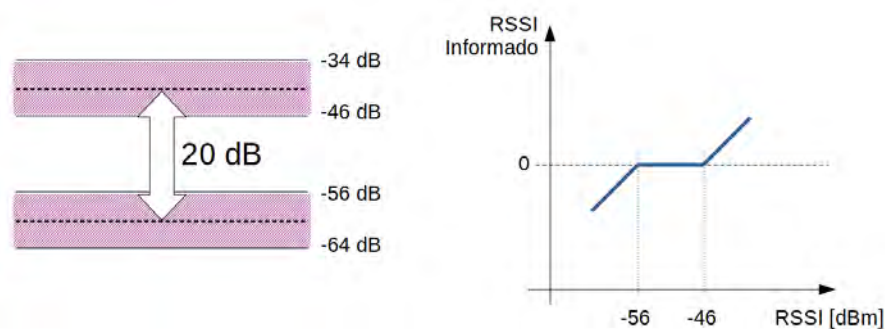


Figura 3.6: El estándar Bluetooth reporta una RSSI que no es lineal. En la zona denominada *Golden Range* entre -46 y -56 dBm informa un valor 0

Estos tiempos son largos en comparación con el tiempo de 16,7 segundos que le toma a un cliente atravesar los 20 m de diámetro del círculo de cobertura a una velocidad de caminata promedio de 1,2 m/s. Por lo tanto, el método no es una buena estrategia cuando el cliente está en movimiento. También afecta la precisión el hecho de que el rango máximo puede variar ampliamente de los 10 m supuestos. La potencia radiada y el patrón de antena de los dispositivos portátiles varían según la forma en que se sostienen y la distancia del cuerpo, y las obstrucciones en la ruta de transmisión en interiores. Todo esto limita la precisión del rango supuesto. A pesar de esto, el método no requiere modificaciones de hardware ni cambios en el protocolo Bluetooth y puede ser adecuado para ciertos servicios de ubicación.

3.2.2.1 RSSI

Otra posibilidad de posicionamiento con Bluetooth es usando las mediciones de RSSI informadas en la capa física. Mas allá de los problemas usuales de utilizar estas medidas en ambientes interiores, las mediciones de RSSI en Bluetooth son imprecisas y no se relacionan de manera lineal con la intensidad de la señal recibida. El propósito principal de la lectura de RSSI en Bluetooth es proporcionar realimentación para el ajuste de potencia del transmisor al que está conectado el receptor. Cuando la señal recibida está dentro de un rango de 20 dB entre un umbral superior (señal fuerte) y un umbral inferior (señal débil), denominado rango dorado de potencia recibida, el RSSI informado es 0 (fig. 3.6). El umbral inferior está entre 6 dB por encima la sensibilidad real del receptor y -56 dBm. El umbral superior es 20 dB mayor, ± 6 dB. El RSSI es negativo cuando la señal está por debajo del umbral inferior, lo que indica que el transmisor debe aumentar la potencia, y positivo cuando la señal está por encima del umbral superior, lo que hace que el transmisor opuesto disminuya su potencia.

La no linealidad de RSSI y los problemas de tiempo de medición prolongado se han resuelto esencialmente en dispositivos Bluetooth de baja energía. La versión 2.1 y posteriores brindan dos características principales que mejoran significativamente la viabilidad de Bluetooth para la medición de distancia y la ubicación. Las lecturas de intensidad de la señal de varios dispositivos Bluetooth dentro del alcance se obtienen rápidamente a través de una función de escaneo, sin la rutina de conexión, y las lecturas de RSSI se informan como valores consecutivos en el rango de sensibilidad del dispositivo. La versión 4, que agregó funciones de Bluetooth Low Energy (LE), hizo que Bluetooth fuera aún más atractivo para el posicionamiento porque brinda a las unidades Bluetooth la capacidad de operar durante períodos prolongados con baterías pequeñas y acorta significativamente el tiempo para realizar conexiones y transferir información. Bluetooth LE tiene una función RSSI completa, por lo que su uso en el posicionamiento es esencialmente el mismo que el de otros dispositivos donde el posicionamiento se basa en la intensidad de la señal, como Wi-Fi.

3.2.3 RFID

RFID (Radio-Frequency IDentification) es una tecnología inalámbrica asimétrica utilizada para la identificación y ubicación de artículos. Un terminal es un *lector* o *interrogador* y el segundo terminal es un *transpondedor* o una etiqueta simple adherida a una estructura, producto, animal o persona que responde con un código de identidad único a las consultas del lector. El RFID opera en bandas de frecuencia sin licencia: baja frecuencia (LF), típicamente 125 kHz y 134,2 kHz; alta frecuencia (HF), 13,56 MHz; y bandas de frecuencia ultra alta (UHF), 433,92 MHz, 860 a 960 MHz, 2,45 GHz y 5,8 GHz. El rango de detección de las etiquetas LF y HF es de hasta 1 m, mientras que las etiquetas UHF pueden cubrir decenas de metros.

Las etiquetas son activas o pasivas. Las etiquetas activas tienen su propia fuente de energía y generan una respuesta de RF modulada en respuesta a la transmisión de un lector. Las etiquetas pasivas funcionan con energía en la señal del lector y responden a través de la tecnología de *backscattering*. La radiación incidente del lector es reflejada por la antena de la etiqueta, modulada a través de cambios en la impedancia de la antena controlada por el IC en la etiqueta. Una tercera categoría de etiqueta es semipasiva, que aumenta el rango de backscattering al alimentar el circuito integrado de la etiqueta para eliminar la carga en las etiquetas pasivas necesarias para crear un voltaje de suministro para operar el chip.

3.2.3.1 Proximidad

El método básico de localización mediante etiquetas RFID es la proximidad. O bien la etiqueta o el lector puede ser el terminal móvil cuya ubicación se va a estimar. Por ejemplo, los lectores de etiquetas se instalan en ubicaciones definidas dentro de una oficina, a menudo en las entradas de las habitaciones ya lo largo de los pasillos. Las personas o equipos cuya ubicación y movimiento están siendo monitoreados están equipados con etiquetas RFID. Cuando una etiqueta está dentro del alcance de un lector, el lector comunica la identificación de la etiqueta a un servidor de ubicación central, ya sea a través de un cableado de infraestructura o una red inalámbrica. El servidor de ubicación recupera de una base de datos la ubicación del lector que luego se considera que es la ubicación del titular de la etiqueta. La operación es análoga en aplicaciones donde las etiquetas se implementan en ubicaciones fijas y conocidas y el lector está conectado al objetivo. En este caso, el lector transmite la identificación de la etiqueta al servidor de ubicación, que ubica al lector cerca de la etiqueta. Es importante resaltar que las etiquetas pasivas que funcionan a 125 kHz suelen tener un alcance de hasta 50 cm. mientras que el alcance de un sistema de 13,56 MHz puede ser de alrededor de 1 m. Esto afecta la densidad de elementos para lograr este tipo de localización.

3.2.3.2 Tags de Referencia

Las etiquetas UHF activas tiene un rango de detección mucho mayor entre el lector y el respondedor que los esquemas de proximidad mencionados. Existen varias propuestas para construir un sistema de localización basados en estas etiquetas. Como ejemplo, tomemos el sistema LANDMARC [16]. Este utiliza etiquetas de referencia y lectores colocadas en un área de hasta decenas de metros cuadrados. El método de ubicación parece similar al *fingerprinting* pero en lugar de crear una base de datos mediante el registro de la intensidad de la señal durante una etapa de inspección, en LANDMARC, las etiquetas de referencia RFID se fijan permanentemente a ubicaciones conocidas. La figura 3.7 muestra un ejemplo del diseño del sistema. Las etiquetas a localizar y las etiquetas de referencia son dispositivos UHF similares. Como el RSSI no está disponible directamente en los lectores de etiquetas, se hacen estimaciones aproximadas de la distancia de la etiqueta ajustando la potencia radiada por el lector en pasos y anotando el número del paso de potencia mínima en el que se detecta la etiqueta. No hay una base de datos preparada de antemano. En la operación de ubicación, las lecturas de intensidad de la señal (paso de potencia) se realizan en todos los lectores de las etiquetas de referencia y la etiqueta de destino. Cada etiqueta de referencia posee un vector de medidas $R_i = (r_{i1}, r_{i2}, \dots, r_{iN})$ donde i es la etiqueta i -ésima, r es el nivel de potencia y N el número de lectores. El vector de

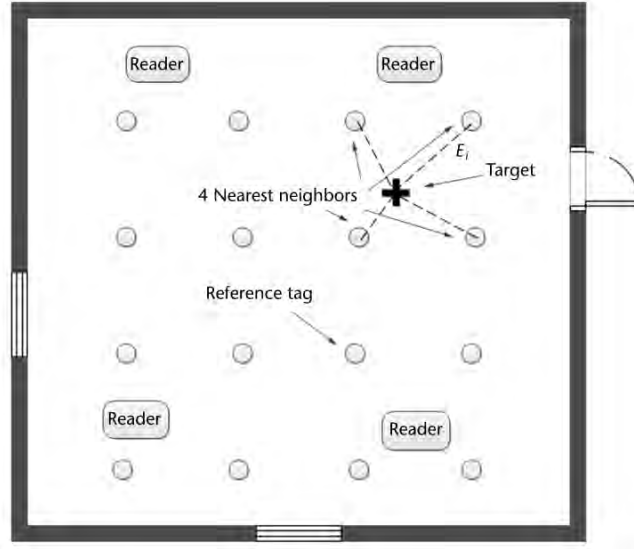


Figura 3.7: Diseño del sistema de ubicación RFID LANDMARC que muestra un ejemplo del despliegue de etiquetas de referencia UHF, lectores y la etiqueta a localizar [5].

medidas del objetivo a localizar es $S = (s_1, s_2, \dots, s_N)$. La distancia euclidiana entre el objetivo y cada una de las etiquetas de referencia será,

$$E_i = \sqrt{\sum_{j=1}^N (r_{ij} - s_j)^2} \quad (3.4)$$

donde j es el índice del lector e i es el índice de la etiqueta de referencia, donde $i \in (1, \dots, M)$ para M etiquetas. Asociado a la etiqueta que se intenta localizar existe un vector de distancias a cada una de las etiquetas de referencia, $\mathbf{E} = (E_1, E_2, \dots, E_M)$. Se supone que el objetivo está cercado a la etiqueta con la menor distancia E_i .

De la misma manera que en el método *Fingerprinting*, la posición del objetivo puede ser estimada por un promedio de las coordenadas de un conjunto de etiquetas cercanas que puede ser mas de una. Los vecinos más cercanos son aquellos que tienen los valores más bajos de E_i . Para L vecinos más cercanos, la estimación de ubicación es

$$(x, y) = \sum_{i=1}^L w_i (x_i, y_i) \quad \text{donde} \quad \sum_{i=1}^L w_i = 1 \quad (3.5)$$

donde (x_i, y_i) son las coordenadas de cada uno de las etiquetas de referencia cercanas, w_i es un factor de peso que enfatiza aquella etiqueta que está mas cercana al objetivo. Un valor posible para ese peso es,

$$w_i = \frac{1/E_i^2}{\sum_{i=1}^L 1/E_i^2} \quad (3.6)$$

Parte IV

RADARES

Se introducen los principios básicos y tecnologías de radares.



RADARES

4.1 INTRODUCCIÓN [7]

Más allá de que las primeras ideas para intentar detectar objetivos usando radiación electromagnética se puede remontar a inicios del siglo XX, y con un desarrollo importante luego de la primer guerra mundial, la palabra **radar** fue introducida en 1940 como un acrónimo de *RA*dio *DE*tectio*N* *AN*d *R*ang*ing*. Al igual que con muchos otros avances tecnológicos, los primeros logros importantes en el radar ocurrieron durante la Segunda Guerra Mundial. Desde entonces, la tecnología de radar ha crecido rápidamente y continúa avanzando a un ritmo acelerado. Vemos ahora una amplia aplicación de radares tanto en aplicaciones comerciales (radares de aeropuerto, radares policiales, radares meteorológicos) como militares (radares de búsqueda, radares de seguimiento)¹.

4.1.1 Tipos de radares

Los radares pueden utilizar dos tipos de señales:

- Pulsado, donde el radar transmite una secuencia de pulsos de radiofrecuencia (RF);
- Onda continua (CW: *Continuous wave* en inglés), donde el radar transmite una señal continua de RF.

Los radares de CW suelen utilizar antenas de transmisión y recepción separadas porque, por lo general, no es posible simultáneamente en una antena recibir una señal mientras se transmite otra a alta potencia. Los radares pulsados evitan este problema utilizando lo que podríamos considerar como multiplexación en el tiempo. Específicamente, la antena se conecta al transmisor mientras el pulso transmite y se conecta al receptor después de la fase de transmisión. Tales radares de señal pulsada constituyen el tipo más común porque requieren solo una antena.

Los dos tipos básicos de radares son radares monoestáticos y bistáticos:

- En un radar monoestático, el transmisor y el receptor (así como sus antenas asociadas) están coubicados. Este es el tipo de radar más común porque es el más compacto. Un radar monoestático

¹ Gran parte de este capítulo es un extracto y resumen simplificado en muchos casos de [7]. En caso que no sea así, se aclara en la sección correspondiente

pulsado suele emplear la misma antena para transmitir y recibir. Un radar monoestático CW generalmente emplea antenas de transmisión y recepción separadas, con un escudo entre ellas.

- En un radar biestático, el transmisor y el receptor están separados, a menudo por distancias muy grandes (>1 km).

Los radares detectan la presencia de un objetivo y determinan su ubicación a partir de la medición del rango y el ángulo. La medición de dichos parámetros permite cálculos razonablemente precisos de la ubicación en 3 dimensiones y la velocidad de un objetivo y, en algunos casos, estimaciones razonables de las derivadas superiores en esas dimensiones.

Los radares operan en la banda de RF del espectro electromagnético entre aproximadamente 5 MHz y 300 GHz. Sin embargo, según la aplicación, la banda de frecuencias utilizadas típicamente son:

- Los radares de búsqueda operan en muy alta frecuencia (VHF) a banda C (4-8GHz);
- Los radares de seguimiento funcionan en las bandas S, C, X y Ku (2-18GHz), ya veces en la banda Ka (27-40GHz);
- Los buscadores de misiles activos operan en las bandas X, Ku, K y Ka (8-40GHz); y
- Los radares de instrumentación y los radares de corto alcance a veces funcionan en la banda Ka y por encima (>27 GHz).

Respecto de la frecuencia de operación de los radares podemos comentar

- Los radares de baja frecuencia requieren antenas grandes y tienen haces más amplios por lo que en general no están asociados con la medición precisa de ángulos.
- Los radares de baja frecuencia tienen limitaciones en cuanto a la precisión de la medición de alcance porque la medición de alcance fino implica un gran ancho de banda instantáneo de la señal de transmisión. Esto causa problemas con el diseño del transmisor y el receptor porque el ancho de banda podría representar un porcentaje significativo de la frecuencia de transmisión.
- La medición de la velocidad no es precisa en los radares de baja frecuencia porque la frecuencia Doppler está directamente relacionada con la frecuencia de transmisión.
- La alta potencia es más fácil de generar a bajas frecuencias porque los dispositivos pueden ser más grandes, lo que permite acomodar voltajes y corrientes más altos.

- La búsqueda requiere alta potencia, pero no necesariamente un rango fino o una medición de ángulo.
- El seguimiento requiere valores precisos de rango y ángulo pero no necesariamente alta potencia.

Por lo tanto, los radares de búsqueda tienden a utilizar frecuencias más bajas y los radares de seguimiento tienden a utilizar frecuencias más altas.

4.1.2 Medidas de Rango

La forma común de medir el rango con un radar es medir el tiempo de retraso desde la transmisión hasta la recepción de un pulso. Dado que la energía de RF viaja a la velocidad de la luz, $c \approx 3 \times 10^8 \text{ m/s}$, el tiempo requerido para que el pulso de transmisión viaje a un objetivo en un rango de R es $\tau_R = R/c$ que es el mismo tiempo que le toma volver al radar. El tiempo medido en el radar entonces será $2\tau_R$ con lo que el rango R puede ser obtenido como

$$R = \frac{c\tau_r}{2} \quad (4.1)$$

Es importante resaltar que el rango que se mide es una línea (a menudo inclinada) desde el radar hasta el objetivo y no es el alcance terrestre, es decir no es la distancia desde el radar hasta la proyección vertical del objetivo sobre el suelo.

Una regla simple para transformar rápidamente la medida de tiempo en un rango es la siguiente. Reemplazamos en la ecuación de rango los valores de velocidad de propagación y el tiempo en microsegundos,

$$R = \frac{c\tau_r}{2} = \frac{3 \times 10^8 \text{ m/s}}{2} (\tau_{\mu\text{s}} \times 10^{-6} \text{ s}/\mu\text{s}) = 150\tau[\text{m}] \quad (4.2)$$

En otras palabras, podemos calcular el rango multiplicando el retardo en microsegundos por $150 \text{ m}/\mu\text{s}$

4.1.3 Ambigüedad de rango

Dado que los radares pulsados transmiten una secuencia de pulsos, la determinación de la distancia al objetivo plantea un problema que es dónde elegimos $t = 0$ para calcular el retraso del rango. El método común es elegir $t = 0$ al comienzo de un pulso de transmisión que se restablece en cada nuevo pulso transmitido. Para definir el problema, considere la Figura 4.1, que muestra pulsos de transmisión separados por $400 \mu\text{s}$. Dado un rango objetivo de 90 km , el retraso del rango al objetivo es

$$\tau_R = \frac{2R}{c} = \frac{2 \times 90 \times 10^3}{3 \times 10^8} = 60 \times 10^{-5} = 600 \mu\text{s} \quad (4.3)$$

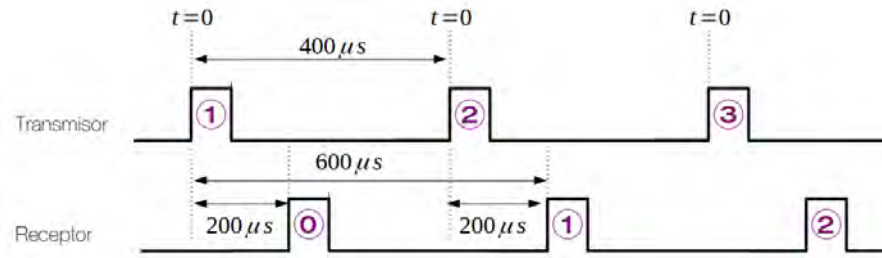


Figura 4.1: Ilustración del problema de la ambigüedad.

Esto significa que el retorno del pulso 1 no se recibe hasta que se transmite el pulso 2; el retorno del pulso 2 no se recibe hasta que se transmite el pulso 3; y así. Dado que todos los pulsos de transmisión son iguales y todos los pulsos recibidos son iguales, no tenemos forma de asociar el pulso recibido 1 con el pulso de transmisión 1. De hecho, dado que el radar restablece $t = 0$ en cada pulso de transmisión, asociará el pulso recibido k con el pulso de transmisión $k + 1$. Además, mediría el retraso del rango como $200 \mu s$ y concluiría erróneamente que el rango objetivo es

$$R_A = 150\tau = 150 \times 200 = 30,000m = 30km \quad (4.4)$$

Este problema es a lo que denominamos ambigüedad o incertidumbre en el rango medido. Si el tiempo entre pulsos es τ_{PRI} (PRI: *pulse repetition interval*, $400 \mu s$ en la figura 4.1), el radar tendrá un rango de detección sin ambigüedades de:

$$R_{amb} = \frac{c\tau_{PRI}}{2} \quad (4.5)$$

Esto significa que si el rango es menor a R_{amb} , el radar puede medirlo sin ambigüedades y con ella si el rango es mayor. En el caso de la figura 4.1 ese rango es de $60km$.

El rango ambiguo a veces presenta un problema en la búsqueda, pero generalmente no en el seguimiento donde los algoritmos de seguimiento del radar proporcionan una estimación del alcance del objetivo, lo que permite que el radar *busque* en el lugar adecuado, incluso con resultados ambiguos.

El uso de formas de onda con múltiples tiempos de repetición de pulsos proporciona otro método para eludir el problema de la ambigüedad de rango. La figura 4.2 muestra un ejemplo de una forma de onda con este principio. En este ejemplo, los espacios entre pulsos son $400 \mu s$, $300 \mu s$ y $350 \mu s$. Como en el ejemplo anterior, se postula un retraso de rango objetivo de $600 \mu s$. Se observará que el retardo de tiempo entre el pulso recibido número 1 y el pulso de transmisión número 2 es de $200 \mu s$, y el retardo de tiempo entre el pulso recibido número 2 y el pulso de transmisión número 3 es de $300 \mu s$. El hecho

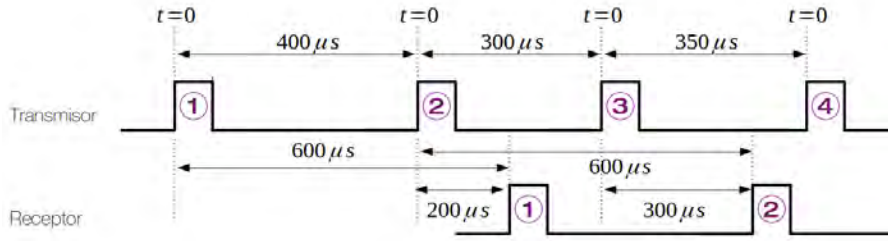


Figura 4.2: Forma de onda para evitar la ambigüedad.

de que el retraso de tiempo entre los pulsos más recientemente transmitidos y recibidos esté cambiando puede usarse para indicar una operación de rango ambiguo. El radar puede usar esta propiedad para ignorar los retornos ambiguos.

Los radares con arreglos de antenas en fase, que dirigen el haz de la antena electrónicamente en lugar de mecánicamente, a menudo transmiten un solo pulso y luego vuelven a dirigir el haz a una posición angular diferente. En esta situación, el concepto de PRI y, por lo tanto, el rango no ambiguo, no está estrictamente definido ya que solo hay un pulso. En este caso, el rango inequívoco se toma como el retraso del rango durante el cual el haz permanece en una posición antes de moverse a otra posición. El tiempo que el haz permanece en una posición se denomina permanencia del haz.

4.1.4 Ventana de Procesamiento

La discusión anterior sobre el rango ambiguo podría llevar a la conclusión de que un radar puede detectar (y rastrear) objetivos en todos los rangos entre 0 y R_{amb} . Sin embargo, en la práctica, este no es el caso. El pulso recibido de un objetivo en un rango de cero llegaría al radar simultáneamente con la transmisión del pulso de sondeo. Dado que el receptor está apagado durante este tiempo, no puede procesar el pulso. El rango mínimo utilizable es por lo tanto igual a

$$R_{min} = \frac{c\tau_p}{2} \quad (4.6)$$

donde τ_p es el ancho de pulso del radar y se denomina rango ciego al que va entre 0 y R_{min} . Entonces, el intervalo de tiempo entre R_{min} y R_{max} se denomina *ventana de procesamiento*. Aunque R_{max} define el rango máximo útil de medida, la mayoría de los radares operan en un rango menor, que se denomina *rango instrumentado*. Este rango se establece según los requisitos del sistema y tiene en cuenta factores como los límites de visualización, los transitorios del circuito, la calibración del radar, los cambios de modo del radar y similares.

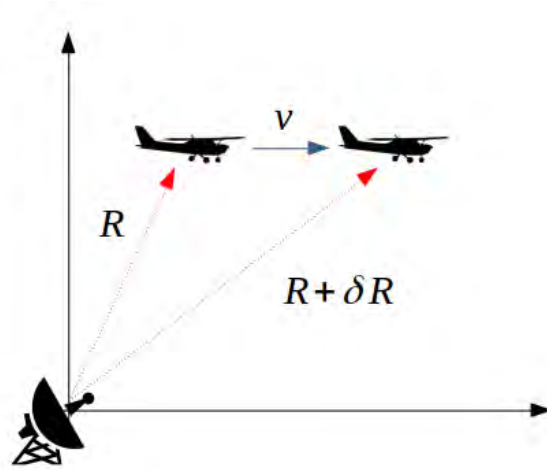


Figura 4.3: Geometría para el cálculo de la velocidad.

4.1.5 Medición de velocidad: Doppler

Además de medir el rango, los radares también pueden medir la tasa de cambio del rango o velocidad. El radar logra esto midiendo la frecuencia Doppler; es decir, la diferencia de frecuencia entre las señales transmitidas y recibidas. Para examinar esto más a fondo, considere la geometría de la figura 4.3. El avión de esta figura se mueve en línea recta a una velocidad de v . Como resultado, la distancia al objetivo cambia continuamente. De hecho, durante un tiempo diferencial de δt , el rango cambia en una cantidad δR , de R a $R + \delta R$. La tasa de rango resultante es

$$\dot{R} = \frac{\delta R}{\delta t} \quad (4.7)$$

En general, $\dot{R} \neq v$ dado que el objetivo no se mueve en la dirección radial, que es el único caso donde la igualdad si existe.

En su forma más simple, el pulso de transmisión constituye un fragmento de una senoide, cuya frecuencia es igual a la operativa del radar (por ejemplo, 10×10^9 Hz o 10 GHz para un radar que funciona en banda X, figura 4.4). Normalmente llamamos a esta la frecuencia portadora del radar y la denotamos como f_c . Para un radar de banda X y una duración de pulso, o ancho de pulso, de $\tau_p = 1 \mu s$, habrá 10 000 ciclos de la portadora durante la duración del pulso.

La frecuencia de la señal recibida será $f_c + f_d$ en lugar de f_c . La frecuencia Doppler o corrimiento Doppler f_d será función de la velocidad y la longitud de onda de la portadora,

$$f_d = \frac{-2\dot{R}}{\lambda_c} \quad (4.8)$$

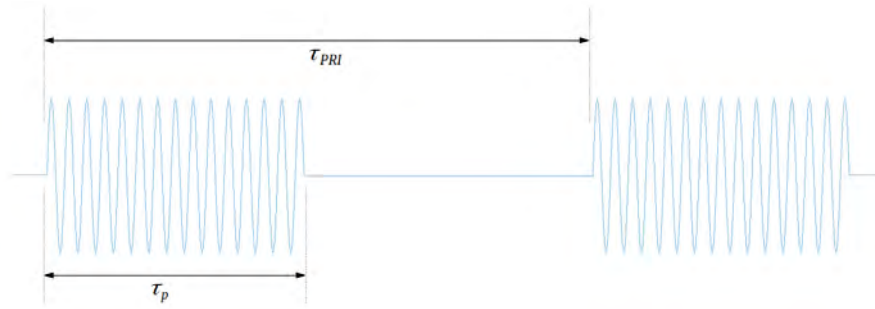


Figura 4.4: Señal de radar.

Por lo tanto, si se compara la frecuencia de la señal transmitida con la de la recibida nos permite determinar f_d y por lo tanto calcular la velocidad.

En la práctica no es simple medir la frecuencia Doppler dada las magnitudes relativas de f_d y f_c . Por ejemplo, si el objetivo se mueve directamente hacia el radar a 540km/h o 150m/s, la velocidad será $\dot{R} = -v = -150\text{m/s}$. Si suponemos que el radar opera en banda X a una frecuencia de portadora $f_c = 10\text{GHz}$ que le corresponde una longitud de onda

$$\lambda_c = \frac{c}{f_c} = \frac{3 \times 10^8}{10 \times 10^9} = 0,03\text{m} \quad (4.9)$$

y la frecuencia de corrimiento será

$$f_d = -\frac{2\dot{R}}{\lambda_c} = -\frac{2(-150)}{0,03} = 10,000\text{Hz} = 10\text{kHz} \quad (4.10)$$

Si comparamos esto con la frecuencia de transmisión, el corrimiento es un millón de veces menor. Para poder medirlo se necesita un pulso de transmisión largo, del orden de los mili-segundos o el procesamiento de varios pulsos.

4.1.6 Diagrama en Bloque de un Radar

La figura 4.5 muestra un diagrama simple de un Radar. En ella se pueden ver las partes básicas que poseen todos los radares. Como se puede observar hay un grupo de tareas que se corresponden con la parte de transmisión (generador de ondas y transmisor de RF), un a la recepción (el receptor de RF, un filtro de acoplamiento, un procesador de las señales recibidas y una etapa final que puede ser de seguimiento o detección o ambas). Entre ambos, está la antena

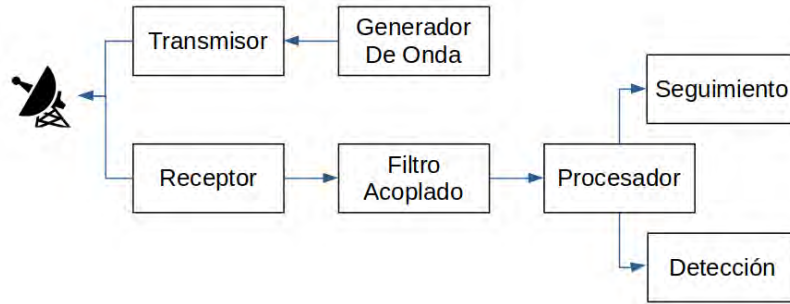


Figura 4.5: Diagrama en bloques genérico de un Radar.

4.2 ECUACIÓN DE ALCANCE DE RADAR

La ecuación básica del alcance de radar que a veces también se denomina ecuación del radar o ecuación de señal a ruido de un solo pulso, es

$$\text{SNR} = \frac{E_S}{E_N} = \frac{P_T G_T G_R \lambda^2 \sigma \tau_p}{(4\pi)^3 R^4 k T_s L} \quad [\text{joule/joule}] \quad (4.11)$$

donde

SNR expresa la relación señal-a-ruido en Joules por Joules o sus unidades equivalentes de watt-segundos por watt-segundos

E_s expresa la energía de la señal, en Joules, en algún punto en el receptor del radar (usualmente a la salida del filtro acoplado o del procesador de señales)

E_N expresa la energía del ruido, en Jules, en el mismo punto en que se especifica la E_s .

P_T denominada potencia máxima de transmisión, expresa la potencia promedio, en watts, durante la transmisión de señales de radar. Aunque P_T se puede especificar en varios puntos, como la salida de la alimentación de la antena, aquí se lo especifica como la potencia a la salida del transmisor.

G_T expresa la directividad, o ganancia directiva, de la antena transmisora en watts por watts.

G_R expresa la directividad, o ganancia directiva, de la antena receptora en watts por watts. En muchos casos, $G_R = G_T$.

λ expresa la longitud de onda del radar en unidades de metros.

σ expresa la sección radar equivalente (RCS: *Radar Cross Section en inglés*) promedio del objetivo en metros cuadrados.

τ_p expresa es el ancho de pulso de transmisión, en segundos.

R es el rango inclinado desde el radar hasta el objetivo en metros.

k es la constante de Boltzmann y es igual a $1,3806503 \times 10^{-23} \text{ W}/(\text{HzK})$ que suele truncarse a $1,38 \times 10^{-23} \text{ W}/(\text{HzK})$.

$T_s = T_a + (F_n - 1)T_0$ Kelvin (K) es la temperatura de ruido del sistema, donde T_a es la temperatura de ruido de la antena utilizada para caracterizar el ruido ambiental.

T_0 expresa una temperatura de referencia. El IEEE define la figura de ruido en términos de una temperatura de ruido $T_0 = 290\text{K}$, lo que da como resultado la aproximación de $kT_0 = 4 \times 10^{-21} \text{ W/Hz}$.

F_n expresa la figura de ruido de radar general y no tiene dimensiones o tiene unidades de watts por watts.

L expresa todas las pérdidas que se tienen en cuenta al utilizar la ecuación de alcance del radar. Este término, que da cuenta de las pérdidas que se aplican a la señal y no al ruido, tiene las unidades de watts por watts. L representa una multitud de factores que degradan el rendimiento del radar, incluidos los relacionados con el propio radar, el entorno en el que opera el radar, el operador del radar y, a menudo, la inexperiencia del analista de radar.

La ecuación de alcance del radar 4.11 se calcula en un solo impulso de transmisión. Si el radar transmite y procesa varios pulsos, la ecuación se puede modificar agregando un término multiplicativo, G_{SP} , que representa la transmisión y el procesamiento de múltiples pulsos. Por otro lado, la formulación de la ecuación como una relación de energía evita la especificación del ancho de banda del ruido y utiliza en su lugar τ_p . Este es el ancho de pulso sin comprimir o el ancho de la envolvente del pulso de transmisión. Cualquier modulación de fase (o frecuencia) en el pulso no es un factor en la ecuación del alcance del radar, excepto posiblemente como una pérdida debido a las diferencias entre el pulso recibido y la respuesta de impulso del filtro adaptado.

La necesidad de calcular la ecuación relación señal a ruido SNR tiene que ver con la probabilidad de detectar una señal que retorna de un objetivo. Esta señal está embebida en ruido (del entorno o de los componentes del receptor) y por lo tanto afecta directamente a esa probabilidad.

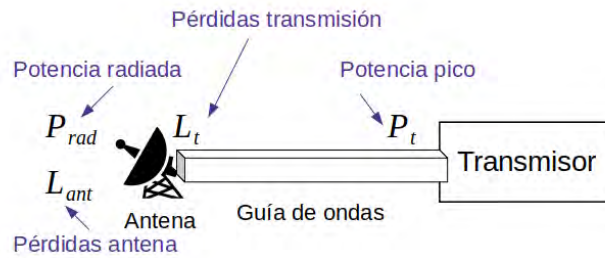


Figura 4.6: Sección del transmisor y las potencias en cada etapa.

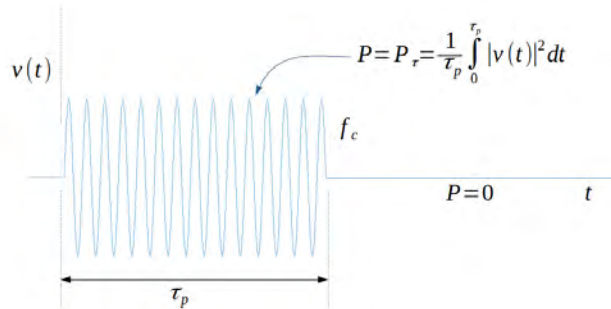


Figura 4.7: Pulso transmitido.

4.2.1 Formulación de la energía de la señal E_s

4.2.1.1 El transmisor

Como se ve en la figura 4.6, desde la salida del transmisor se pasa por la guía de ondas a la antena hacia el espacio. Suponemos que el radar está en el espacio libre dado que se pueden considerar los efectos de la atmósfera en el término de pérdida, L . Suponemos que el transmisor genera un solo pulso con una envolvente rectangular que tiene un ancho de τ_p . La Figura 4.7 contiene una representación simplificada de este pulso que se modula con la frecuencia constante de la portadora f_c . La potencia indicada en la Figura 4.7 es la potencia normalizada, es decir la entregada a una resistencia de 1 ohm.

La potencia de transmisión promedio en la señal durante la duración del pulso se denomina potencia de transmisión máxima P_T y se denomina máxima porque más adelante consideraremos la potencia de transmisión promediada sobre muchos pulsos. La guía de ondas por otro lado lleva la señal desde el transmisor hasta la entrada de alimentación de la antena. La única característica de interés de la guía de ondas en la ecuación de alcance del radar es que es un dispositivo con pérdida que atenúa la señal. Aunque aquí solo nos referimos a la "guía de ondas", hay varios dispositivos incluidos entre el transmisor y la alimentación de la antena de un radar práctico. La caracterizamos

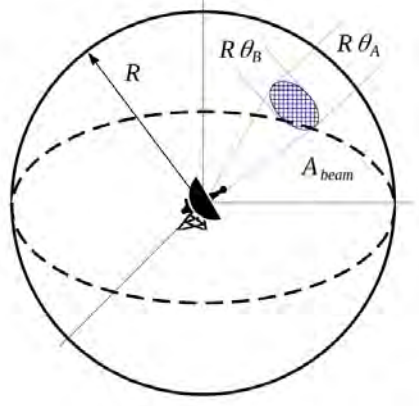


Figura 4.8: Esfera de radiación y el haz de la antena.

entonces por el término L_t o pérdida de transmisión que es mayor a la unidad. La potencia al alimentador de la antena será

$$P_{feed} = \frac{P_T}{L_t} \quad [W] \quad (4.12)$$

En general, el alimentador y otros componentes de la antena introducen pérdidas adicionales que las consolidamos en uno sólo que llamaremos pérdida en la antena L_{ant}

$$P_{rad} = \frac{P_{feed}}{L_{ant}} = \frac{P_T}{L_t L_{ant}} \quad [W] \quad (4.13)$$

Finalmente, dado que la envolvente del pulso es τ_p , la energía radiada por la antena será

$$E_{rad} = P_{rad} \tau_p \quad [Joules] \quad (4.14)$$

4.2.1.2 La antena

El propósito de la antena del radar es concentrar y enfocar la energía radiada en un pequeño sector angular del espacio. Suponemos que toda la energía radiada se concentra en un área A_{beam} como se muestra en la figura 4.8. La densidad de energía sobre esta área será

$$S_R = \frac{E_{rad}}{A_{beam}} = \frac{P_T \tau_p / L_t L_{ant}}{A_{beam}} \quad [Ws/m^2] \quad (4.15)$$

Dados los ejes menores y mayores ($R\theta_A$ y $R\theta_B$), el área de la elipse A_{beam} será

$$A_{ellipse} = \frac{\pi}{4} R^2 \theta_A \theta_B \quad [m^2] \quad (4.16)$$

Dado que toda la energía no se concentrará perfectamente en A_{beam} dado que el enfoque de la antena no es perfecto, se tiene en cuenta esta pérdida, reemplazando $\pi/4$ por un factor K_A ,

$$A_{ellipse} = K_A R^2 \theta_A \theta_B \quad [m^2] \quad (4.17)$$

que lo sustituimos en la ecuación 4.15

$$S_R = \frac{P_T \tau_p / L_t L_{ant}}{K_A R^2 \theta_A \theta_B} \quad [Ws/m^2] \quad (4.18)$$

La directividad de la antena se define como

$$G_T = \frac{4\pi}{K_A \theta_A \theta_B} \quad [W/W] \quad (4.19)$$

que podemos usarla en 4.18

$$S_R = \frac{G_T P_T \tau_p}{r^2 R^2 L_t L_{ant}} \quad [Ws/m^2] \quad (4.20)$$

Tanto la expresión de G_T como la ecuación del alcance del radar asumen tácitamente una antena apuntando directamente al objetivo. Si esto no es así, se debe modificar G_T para tenerlo en cuenta a partir del patrón de radiación de la antena.

4.2.1.3 Directividad de la antena

El factor K_A en la ecuación 4.19 está relacionado con las características de la antena y tiene que ver con que la energía no está uniformemente distribuida en la elipse A_{beam} y que toda la energía no está concentrada en el haz y parte se derrama en los lóbulos laterales.

En la literatura, un valor común para K_A es 1,65, por lo que se puede reescribir la ecuación 4.19

$$G_T = \frac{4\pi}{1,65 \theta_A \theta_B} \quad [W/W] \quad (4.21)$$

con los ángulos expresados en radianes. Si los expresamos en grados, se reescribe esta ecuación como

$$G_T = \frac{25,000}{\theta_A^\circ \theta_B^\circ} \quad [W/W] \quad (4.22)$$

Hay varios autores en la literatura que aproximan de diferentes formas el factor K_A y por lo tanto el numerador de 4.22. Por ejemplo, algunos autores lo aproximan a 32,383 que se aplica a una apertura rectangular con iluminación uniforme y 33,709 para una apertura circular. En general, entre 25,000 y 26,000 es un valor razonable para antenas que aplican algún tipo de estrategia para reducir los lóbulos laterales.

Esto se puede visualizar en la figura 4.9 donde se grafica la ganancia G_T en función del ángulo de observación (p. 765 [2]) y representa la directividad de la antena bocina piramidal que apunta directamente a un objetivo en la dirección de observación 0° . Como se ve, en esa dirección la directividad es máxima y existe energía que se dispersa para diferentes ángulos con lóbulos a los laterales. Si la antena convencional es construida con las paredes corrugadas para disminuir esta energía dispersada.

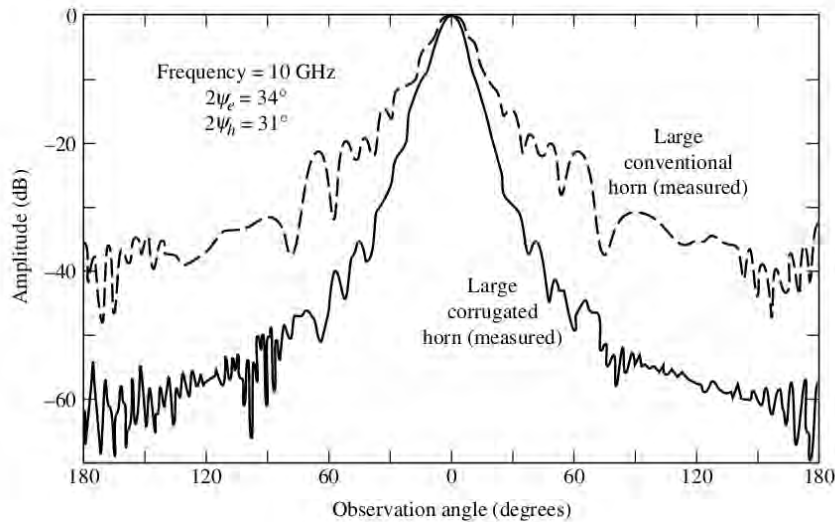


Figura 4.9: Características de radiación de una antena bocina piramidal convencional y otra corrugada (p. 765 [2]).

Si definimos el ancho de lóbulo o haz principal aquel rango de ángulos para el que la potencia de pico cae la mitad, o 3dB menos del pico, en la figura 4.9 esto sucede para ambas antenas casi en el mismo punto y son aproximadamente 5 grados. Si en el plano perpendicular a este, el ángulo es el mismo, la directividad puede ser calculada como,

$$G_T = \frac{25,000}{2 \times 5} = 2500 \quad [W/W] = 34 \quad [dB] \quad (4.23)$$

4.2.1.4 El objetivo y la sección de radar equivalente

A medida que la onda electromagnética pasa por el objetivo, este captura parte de su energía y la vuelve a irradiar hacia el radar. Más exactamente, la onda electromagnética induce corrientes en el objetivo y las corrientes generan otra onda electromagnética que se propaga lejos del objetivo. El proceso de capturar y volver a irradiar energía es muy complicado y es objeto de muchas investigaciones. Por ahora, adoptamos un enfoque simplificado del proceso mediante el uso del concepto de sección de radar equivalente (RCS: *Radar Cross Section* en inglés).

Si multiplicamos la densidad de energía S_R con unidades de $[W \cdot s/m^2]$ por σ que representa la sección de radar equivalente promedio del objetivo, la convertimos en energía. La energía capturada y re-irradiada será

$$E_{tgt} = \sigma S_R \quad [W \cdot s] \quad (4.24)$$

Podemos continuar con la suposición ideal de que el objetivo actúa como un radiador isotrópico, y la energía E_{tgt} se irradia en todas las

direcciones y de hecho actúa como una antena. Dada esto, podemos representar la densidad de energía en el radar como,

$$S_{\text{rec}} = \frac{E_{\text{tgt}}}{4\pi R^2} = \frac{P_T G_T \sigma \tau_p}{(4\pi)^2 R^4 L_t L_{\text{ant}}} \quad [W \cdot s/m^2] \quad (4.25)$$

4.2.1.5 La antena de nuevo

Cuando la onda electromagnética que proviene del objetivo vuelve al radar, la antena captura una parte de la energía de esta onda y la envía al receptor del radar. Si extendemos la lógica que aplicamos al objetivo, podemos formular la energía en la salida de la alimentación de la antena como

$$E_{\text{ant}} = S_{\text{rec}} A_e \quad [W \cdot s] \quad (4.26)$$

donde A_e es el área efectiva de la antena y es una medida que describe la capacidad de la antena para capturar la energía electromagnética que retorna y convertirla en potencia utilizable. También se denomina *apertura efectiva* de la antena. Esta, está relacionada con las características físicas de la antena,

$$A_e = \rho_{\text{ant}} A_{\text{ant}} \quad [m^2] \quad (4.27)$$

donde A_{ant} es el área de la antena proyectada sobre un plano colocado directamente frente a ella. Por ejemplo, si la antena es un paraboloide el área real sería el área de la superficie paraboloide, mientras que A_{ant} es el área del disco definida por el borde frontal de la antena. Por otro lado, ρ_{ant} representa la eficiencia de la antena y puede tomar un valor entre 0 y 1 aunque raramente esté por debajo de 0.4 o arriba de 0.7. Un valor típico es 0.6.

Reorganizando, la energía en la salida de la alimentación de la antena será

$$E_{\text{ant}} = \frac{P_T G_T \sigma \tau_p A_e}{(4\pi)^2 R^4 L_t L_{\text{ant}}} \quad [W \cdot s] \quad (4.28)$$

4.2.1.6 Directividad de la antena de nuevo

El término A_e no es muy fácil de usar por lo que se utiliza la directividad para determinar la energía en la antena. De acuerdo con la teoría de la antena, podemos relacionar la directividad con la apertura efectiva mediante la ecuación

$$G_R = \frac{r\pi A_e}{\lambda^2} \quad [W/W] \quad (4.29)$$

Reemplazamos esto en la ecuación 4.28

$$E_{\text{ant}} = \frac{P_T G_T G_R \lambda^2 \sigma \tau_p}{(4\pi)^3 R^4 L_t L_{\text{ant}}} \quad [W \cdot s] \quad (4.30)$$

Si a esto agregamos una ganancia G que tenga en cuenta los componentes del receptor hasta el punto donde se mide la SNR, obtenemos

$$E_{\text{rec}} = \frac{P_T G_T G_R \lambda^2 \sigma \tau_p}{(4\pi)^3 R^4 L_t L_{\text{ant}}} G \quad [\text{W} \cdot \text{s}] \quad (4.31)$$

Finalmente, hay varias pérdidas que deben ser tenidas en cuenta además de $L_t L_{\text{ant}}$, si consolidamos entonces a estas y todas las pérdidas en L , la energía de la señal será finalmente

$$E_S = \frac{P_T G_T G_R \lambda^2 \sigma \tau_p}{(4\pi)^3 R^4 L} G \quad [\text{W} \cdot \text{s}] \quad (4.32)$$

4.2.2 Formulación de la energía del ruido E_N

Los dos principales contribuyentes al ruido en los radares son el ambiente y los componentes electrónicos del receptor. El ruido ambiental incluye la radiación de la tierra, el ruido galáctico e intergaláctico, el ruido atmosférico y, en algunos casos, el ruido creado por el hombre, como los bloqueadores de ruido. El ruido galáctico e intergaláctico incluye la radiación cósmica de fondo y el ruido solar o de otras estrellas. El ruido ambiental que consideramos es el ruido terrestre, galáctico e intergaláctico.

El ruido de los equipos electrónicos se denomina ruido térmico y surge de la agitación de los electrones causada por el calor. Una de las ecuaciones que conduce a una definición de densidad espectral de potencia de ruido, o energía, para dispositivos resistivos es

$$N_0 = kT \quad [\text{W} \cdot \text{s}] \quad (4.33)$$

donde $k = 1,38 \times 10^{-23} \text{ [W} \cdot \text{s/K]}$ es la constante de Boltzmann y T es la temperatura de ruido de una resistencia en kelvin [K]. Esta ecuación predice que la energía de ruido es independiente de la frecuencia. En general se adopta una versión modificada para componentes electrónicos que está dada por

$$N_0 = kT_0 F \quad [\text{W} \cdot \text{s}] \quad (4.34)$$

donde F se la denomina *figura de ruido* del dispositivo y T_0 es una temperatura de referencia que normalmente se refiere a la temperatura ambiente.

Aunque N_0 se desarrolló para componentes electrónicos y no para ruido ambiental, los analistas de radar la adoptan como una forma de caracterizar la energía en un radar debida al ruido en el entorno, así como en la electrónica. Por lo tanto, definimos la energía del ruido a la entrada del filtro adaptado como

$$E_N = GkT_s \quad [\text{W} \cdot \text{s}] \quad (4.35)$$

donde

$$T_s = T_a + (F_n - 1)T_0 \quad (4.36)$$

En la ecuación 4.35, G es la misma ganancia del receptor que aparece en la ecuación 4.32. Además, T_s es la temperatura de ruido del sistema y T_a es la temperatura de la antena; F_n es el factor de ruido general del radar desde la “cara de la antena” hasta la entrada del filtro acoplado. Incluye las figuras de ruido de todos los dispositivos activos y pasivos en el radar, incluidos los componentes de la antena (p. ej., radomo, desfasadores, guías de ondas, alimentaciones) que presenten una pérdida óhmica o disipativa.

La temperatura de la antena, T_a , proporciona un medio para caracterizar el ruido ambiental en el radar. Para el caso en que el haz del radar apunte hacia el cielo pero no directamente al sol o a una estrella, su ecuación es

$$T_a \approx 0,876T'_a + 36 \quad [\text{K}] \quad (4.37)$$

Esta ecuación también tiene en cuenta que el ruido terrestre entra a través de los lóbulos laterales y traseros de la antena y se supone una antena sin pérdidas óhmicas, que están incluidas en F_n . La temperatura T'_a es la temperatura de una antena idealizada y plasmada en manuales en curvas.

Ahora, si suponemos que $T_a = T_0$, que es razonable cuando $F_n > 7\text{dB}$,

$$T_s = T_a + (F_n - 1)T_0 = f_n T_0 \quad (4.38)$$

Combinando entonces las ecuaciones 4.32 y 4.35, obtenemos finalmente la relación señal a ruido,

$$\text{SNR} = \frac{E_S}{E_N} = \frac{P_T G_T G_R \lambda^2 \sigma \tau_p}{(4\pi)^3 R^4 k T_s L} \quad [\text{W} \cdot \text{s} / \text{W} \cdot \text{s}] \quad (4.39)$$

donde el término G se cancela presente en ambas energías lo que significa que la ganancia general del receptor del radar no afecta a SNR

4.2.3 Ejemplo de Alcance de Radar

Para ilustrar el uso de la ecuación de alcance de radar, consideremos un radar monoestático con los parámetros que se listan en la tabla 4.1. Se trata de calcular la SNR en un objetivo con una RCS de 6dBsm y en un rango de 60km . Para realizar el cálculo, necesitamos encontrar los parámetros de la ecuación del alcance del radar 4.39 y asegurarnos de que estén en unidades consistentes. La mayoría de los parámetros están en la Tabla 4.1 o pueden derivarse de ellos. Los dos parámetros que no están en la tabla son el rango objetivo y el RCS objetivo, que

Parámetros del Radar	Valor
Potencia pico transmitida P_T	1 MW
Perdidas de transmisión $L_t L_{ant}$	2dB
Ancho de pulso τ_p	0.4 μs
Directividad de la antena G_T, G_R	38dB
Frecuencia de operación	8 GHz
Figura de ruido del sistema F_n	8 dB
Otras pérdidas L_o	2dB

Tabla 4.1: Parámetros de un radar ejemplo

Parámetros del Radar	Valor Sist. Métrico	Valor dB
P_T	10^6 [W]	60[dBW]
G_T	6309.6 [W/W]	38 [dB]
G_R	6309.6 [W/W]	38 [dB]
$\lambda = c/f_c$	0.0375 [m]	-14.26 [dB(m)]
Ancho de pulso τ_p	$0,410^{-6}s$ -64 [dB(s)]	
σ	3,98[m ²]	6 [dB(sm)]
R	60×10^3 [m]	47.78 [dB(m)]
k	$1,38 \times 10^{-23}$ [W · s/K]	-228.6 [dB(W · s/K)]
$L = L_o L_t L_{ant}$	2.51 [W/W]	4 [dB]
T_s	1602 [K]	32[dB(K)]

Tabla 4.2: Parámetros finales para el calculo de SNR de un radar ejemplo

son dato. Necesitaremos calcular la longitud de onda, λ , las pérdidas totales y la temperatura de ruido del sistema, T_s . La Tabla 4.2 da los parámetros en las unidades apropiados y los valores de λ y L .

T_s se calculó a partir de

$$T_s = T_a + (F_n - 1)T_0 \quad (4.40)$$

donde para un $T'_a \approx 30K$ obtenida da curvas

$$T_a \approx 0,876T'_a + 36 = 0,876 \times 30 + 36 = 62,3 \quad (4.41)$$

Substituyendo entonces los valores de la tabla 4.2 en la ecuación 4.39,

$$\begin{aligned} \text{SNR} &= \frac{(10^6)(6309,6)(6309,6)(0,0375)^2(3,98)(0,4 \times 10^{-6})}{(4\pi)^3(60 \times 10^3)^4(1,38 \times 10^{-23})(1602)(2,51)} \\ &\approx 62,5 \quad [W \cdot s / W \cdot s] = 17,96\text{dB} \end{aligned} \quad (4.42)$$

4.2.4 Rango de detección

Un uso importante de la ecuación del alcance del radar es la determinación del alcance de detección, o el alcance máximo en el que un objetivo tiene una probabilidad dada de ser detectado por un radar. El criterio para detectar un objetivo es que la SNR esté por encima de algún valor de umbral. Si consideramos la ecuación de alcance del radar anterior, observamos que la SNR varía inversamente con la cuarta potencia del alcance. Esto significa que si la SNR es un cierto valor en un rango dado, aumentará a medida que el rango disminuya. Por lo tanto, definimos el rango de detección como el rango en el que logramos una cierta SNR. Para encontrar el rango de detección resolvemos la ecuación del rango del radar para R ,

$$R = \left[\frac{P_T G_T G_R \lambda^2 \sigma \tau_p}{(4\pi)^3 (\text{SNR}) k T_s L} \right]^{1/4} \quad [\text{m}] \quad (4.43)$$

Para el ejemplo anterior si suponemos que el umbral de detección es para $\text{SNR} = 13\text{dB}$, el rango de detección será de 80km aproximadamente. El nivel de 13dB es una suerte de estándar y se demuestra que esto implica la probabilidad de detección de 0.5 para un objetivo del tipo de un avión.

4.2.5 Ecuación del rango de búsqueda de un radar

Se puede plantear una extensión de la ecuación de alcance del radar que es utilizada para analizar y diseñar radares de búsqueda. Su uso más común es en el dimensionamiento inicial de los radares de búsqueda en términos de potencia y tamaño físico. De hecho, la medida de rendimiento que suele utilizarse para caracterizar este tipo

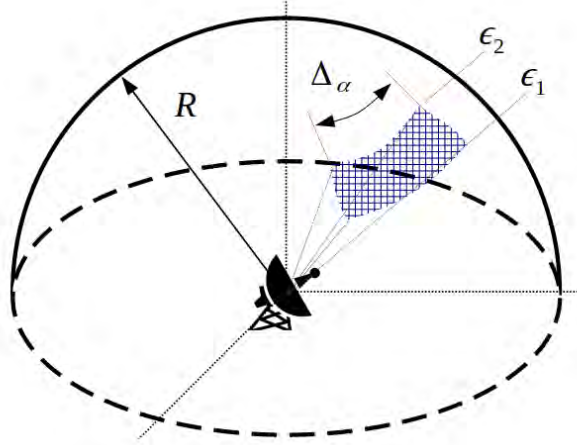


Figura 4.10: Ilustración del sector de búsqueda de un radar.

de radares es el producto medio potencia-apertura, $P_A A_e$, que es el producto de la potencia media por la apertura efectiva del radar.

Comenzamos suponiendo que el radar busca una región angular, o sector, ω que tiene unidades de estereorradianes². Uno de los sectores de búsqueda más comunes es una sección de la superficie de una esfera delimitada por elevación y azimut como se muestra en la figura 4.10. Allí se muestra una extensión de azimut de $\Delta\alpha$ y una extensión de elevación de ϵ_1 a ϵ_2 . El área angular de este sector de búsqueda con todos los ángulos en radianes es

$$\Omega = \Delta\alpha(\sin \epsilon_2 - \sin \epsilon_1) \quad [\text{estereorradianes}] \quad (4.44)$$

El área sobre la superficie de una esfera de radio R será

$$A_{\text{beam}} = K_A R^2 \theta_A \theta_B \quad [\text{m}^2] \quad (4.45)$$

donde como en 4.17, K_A tiene en cuenta la pérdida de energía debido a que parte de la energía se pierde fuera de esta área. Si dividimos esta ecuación por R^2 , se obtiene el área angular del haz del radar,

$$\Omega_{\text{beam}} = K_A \theta_A \theta_B \quad [\text{estereorradianes}] \quad (4.46)$$

El número de haces necesarios para cubrir el sector de búsqueda será

$$n = \Omega / \Omega_{\text{beam}} = \Omega / K_A \theta_A \theta_B \quad (4.47)$$

Este número es ideal y supone un sector de búsqueda y haces rectangulares. En la práctica el número de haces necesarios está afectado por un factor de empaquetado K_{pack} que tiene en cuenta como se

² Es una unidad que mide ángulos sólidos y equivalente tridimensional del radián

organizan y para haces rectangulares o elípticos que se tocan en los 3dB $K_{\text{pack}} = 1$, en general será distinto de 1 y entonces

$$n = K_{\text{pack}} \Omega / \Omega_{\text{beam}} \quad (4.48)$$

Uno de los requerimientos que se le impone a los radares de búsqueda es que debe cubrir el sector en T_{scan} segundos. Esto significa que el radar debe procesar la señal de n haces en T_{scan} segundos. Dado este requerimiento, el tiempo asignada a cada haz será

$$T_{\text{beam}} = T_{\text{scan}} / n \quad [\text{s}] \quad (4.49)$$

que será el intervalo de repetición de pulso (PRI) T y reemplazando a n ,

$$T = T_{\text{scan}} K_A A \theta_A \theta_B / K_{\text{pack}} \Omega \quad [\text{s}] \quad (4.50)$$

Usamos entonces al ciclo de trabajo $d = \tau_p / T$ para obtener

$$d = \frac{\tau_p K_{\text{pack}} \Omega}{T_{\text{scan}} K_A A \theta_A \theta_B} \quad (4.51)$$

Además, dado que el radar tiene un ancho de pulso τ_p y un PRI de T , la potencia promedio será

$$P_A = P_T (\tau_p / T) = P_T d \quad [\text{W}] \quad (4.52)$$

Substituimos 4.51 y 4.52 en la ecuación 4.11

$$\begin{aligned} \text{SNR} &= \frac{P_T G_T G_R \lambda^2 \sigma \tau_p}{(4\pi)^3 R^4 k T_s L} = \frac{P_A}{d} \frac{G_T G_R \lambda^2 \sigma \tau_p}{(4\pi)^3 R^4 k T_s L} \\ &= \frac{P_A}{\tau_p K_{\text{pack}} \frac{\Omega}{T_{\text{scan}} K_A A \theta_A \theta_B}} \frac{G_T G_R \lambda^2 \sigma \tau_p}{(4\pi)^3 R^4 k T_s L} \\ &= \frac{T_{\text{scan}} K_A A \theta_A \theta_B P_A G_T G_R \lambda^2 \sigma}{K_{\text{pack}} \Omega (4\pi)^3 R^4 k T_s L} \quad [\text{W} \cdot \text{s} / \text{W} \cdot \text{s}] \end{aligned} \quad (4.53)$$

Usando 4.19 y 4.29 en la ecuación 4.53

$$\text{SNR} = \frac{P_A A_e \sigma T_{\text{scan}}}{4\pi R^4 k T_s L K_{\text{pack}} \Omega} \quad [\text{W} \cdot \text{s} / \text{W} \cdot \text{s}] \quad (4.54)$$

4.2.6 Ejemplo de rango de búsqueda

Suponemos que la búsqueda cubriría una región del espacio entre 0° y 45° en elevación y 30° en azimut que debemos recorrer en 10s. La potencia está limitada a $3 \times 10^6 [\text{W/m}^2]$. El objetivo de interés tiene un RCS de -10dBsm y se necesita una $\text{SNR} = 13\text{dB}$ para obtener un detección. La tecnología actual para este radar hipotético supone una figura de ruido de 4dB y unas pérdidas totales de 6dB. Suponemos

Parámetros del Radar	Valor
Búsqueda en Azimut	30°
Búsqueda en Elevación	0° – 0°
Potencia de apertura	$3 \times 10^6 [\text{Wm}^2]$
Tiempo de búsqueda T_{scan}	10s
RCS σ	-10 dBsm
SNR para detección	13 dB
Pérdidas totales L	6 dB
Factor de empaquetado K_{pack}	1
Temperatura de ruido de sistema T_s	487 K

Tabla 4.3: Parámetros de un radar de búsqueda ejemplo

una haz (elevación) promedio de 10° lo que da un $T'_a = 15\text{K}$. Esto significa una $T_s = 487\text{K}$. Suponemos $K_{\text{pack}} = 1$. Todos estos parámetros se resumen en la tabla 4.3.

Resolvemos la ecuación 4.54 para obtener R

$$R = \left[\frac{P_A A_e \sigma T_{\text{scan}}}{4\pi(\text{SNR})kT_s L K_{\text{pack}} \Omega} \right]^{-1/4} [\text{m}] \quad (4.55)$$

Calculamos Ω

$$\Omega = \Delta\alpha(\sin \epsilon_2 - \sin \epsilon_1) = \left(\frac{\pi}{6}\right) \left(\sin \frac{\pi}{4} - \sin 0\right) = 0,118\pi \quad (4.56)$$

Reemplazamos todos los valores en 4.55

$$R = \left[\frac{3 \times 10^6 \times 0,1 \times 10}{4\pi(20) \times 1,38 \times 10^{-23} \times 487 \times 10^{0,6} \times 0,118\pi} \right] = 1,047[\text{km}] \quad (4.57)$$



Parte V

APÉNDICE

BIBLIOGRAFÍA

- [1] Adrien Angeli, Stephane Doncieux, Jean-Arcady Meyer y David Filliat. "Incremental vision-based topological SLAM". En: *2008 IEEE/RSJ International Conference on Intelligent Robots and Systems*. 2008, págs. 1031-1036.
- [2] Constantine A Balanis. *Antenna theory: analysis and design*. John Wiley & sons, 2015.
- [3] Yaakov Bar-Shalom, Thomas E Fortmann y Peter G Cable. *Tracking and data association*. Acoustical Society of America, 1990.
- [4] Yaakov Bar-Shalom, X Rong Li y Thiagalingam Kirubarajan. *Estimation with applications to tracking and navigation: theory algorithms and software*. John Wiley & Sons, 2004.
- [5] Alan Bensky. *Wireless positioning technologies and applications*. Artech House, 2016.
- [6] Andreas Bergström, Gustaf Hendeby, Fredrik Gunnarsson y Fredrik Gustafsson. "TOA estimation improvements in multipath environments by measurement error models". En: *2017 IEEE 28th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC)*. IEEE. 2017, págs. 1-7.
- [7] Mervin C Budge y Shawn R German. *Basic RADAR analysis*. Artech House, 2020.
- [8] Xiao Cui, Chunbo Mei, Yongyuan Qin, Gongmin Yan y Qiangwen Fu. "In-motion alignment for low-cost SINS/GPS under random misalignment angles". En: *The Journal of Navigation* 70.6 (2017), págs. 1224-1240.
- [9] Kutluyıl Doğançay. "Passive emitter localization using weighted instrumental variables". En: *Signal Processing* 84.3 (2004), págs. 487-497. ISSN: 0165-1684.
- [10] José Luis Gomez-Tornero, David Cañete-Rebenaque, Jose Antonio López-Pastor y Alejandro Santos Martínez-Sala. "Hybrid Analog-Digital Processing System for Amplitude-Monopulse RSSI-Based MiMo WiFi Direction-of-Arrival Estimation". En: *IEEE Journal of Selected Topics in Signal Processing* 12.3 (2018), págs. 529-540.
- [11] Xiaoying Kong, Eduardo Mario Nebot y Hugh Durrant-Whyte. "Development of a nonlinear psi-angle model for large misalignment errors and its application in INS alignment and calibration". En: *Proceedings 1999 IEEE International Conference on Robotics and Automation (Cat. No. 99CH36288C)*. Vol. 2. IEEE. 1999, págs. 1430-1435.

- [12] Havish Koorapaty. "Cramer-Rao bounds for time of arrival estimation in cellular systems". En: *2004 IEEE 59th Vehicular Technology Conference. VTC 2004-Spring (IEEE Cat. No. 04CH37514)*. Vol. 5. IEEE. 2004, págs. 2729-2733.
- [13] EM Nebot. "Sensors used for autonomous navigation". En: *Advances in Intelligent Autonomous Systems*. Springer, 1999, págs. 135-156.
- [14] Eduardo Nebot y Hugh Durrant-Whyte. "Initial calibration and alignment of low-cost inertial navigation units for land vehicle applications". En: *Journal of Robotic Systems* 16.2 (1999), págs. 81-92.
- [15] Ngoc Hung Nguyen y Kutluyıl Doğançay. "Closed-form algebraic solutions for angle-of-arrival source localization with Bayesian priors". En: *IEEE Transactions on Wireless Communications* 18.8 (2019), págs. 3827-3842.
- [16] Lionel M Ni, Yunhao Liu, Yiu Cho Lau y Abhishek P Patil. "LANDMARC: Indoor location sensing using active RFID". En: *Proceedings of the First IEEE International Conference on Pervasive Computing and Communications, 2003.(PerCom 2003)*. IEEE. 2003, págs. 407-415.
- [17] Yoshihisa Okumura. "Field strength and its variability in VHF and UHF land-mobile radio service". En: *Rev. Electr. Commun. Lab.* 16 (1968), págs. 825-873.
- [18] J. D. Parsons. *The Mobile Radio Propagation Channel*. John Wiley & Sons Ltd, 2001.
- [19] Paolo Pivato, Luigi Fontana, Luigi Palopoli y Dario Petri. "Experimental assessment of a RSS-based localization algorithm in indoor environment". En: *2010 IEEE Instrumentation & Measurement Technology Conference Proceedings*. IEEE. 2010, págs. 416-421.
- [20] Pampa Sadhukhan. "Performance analysis of clustering-based fingerprinting localization systems". En: *Wireless Networks* 25.5 (2019), págs. 2497-2510.
- [21] R Saraiva y L Lovisolo. *RF Positioning: Fundamentals, Applications and Tools*. Artech House Publishers: Boston, MA, USA, 2015.
- [22] Ian Sharp y Kegen Yu. "Indoor TOA error measurement, modeling, and analysis". En: *IEEE Transactions on Instrumentation and Measurement* 63.9 (2014), págs. 2129-2144.
- [23] Bernard Sklar. "Rayleigh fading channels in mobile digital communication systems. I. Characterization". En: *IEEE Communications magazine* 35.7 (1997), págs. 90-100.

- [24] Sakthitharan Subramanian, Thirumaraiselvan Packirisamy, Mariappan Mahalingam, Sudharshan Ramanathan y Lenin Gopal. "Scattering coefficients and pathgain of a multilayer tissue structure using ABCD matrix method". En: *Journal of Ambient Intelligence and Humanized Computing* 12 (jun. de 2021).
- [25] Salah Sukkarieh. "Low Cost, High Integrity, Aided Inertial Navigation Systems for Autonomous Land Vehicles". Tesis doct. University of Sydney, 2000.
- [26] Peter Teunissen y Oliver Montenbruck. *Springer handbook of global navigation satellite systems*. Springer, 2017.
- [27] Reza M Vaghefi, Mohammad Reza Gholami y Erik G Ström. "Bearing-only target localization with uncertainties in observer position". En: *2010 IEEE 21st international symposium on personal, indoor and mobile radio communications workshops*. IEEE. 2010, págs. 238-242.
- [28] Xin Yuan, José-Fernán Martínez-Ortega, José Antonio Sánchez Fernández y Martina Eckert. "AEKF-SLAM: A New Algorithm for Robotic Underwater Navigation". En: *Sensors* 17.5 (2017). ISSN: 1424-8220. URL: <https://www.mdpi.com/1424-8220/17/5/1174>.
- [29] Xiaohong Zhang, Mingkui Wu y Wanke Liu. "Receiver time misalignment correction for GPS-based attitude determination". En: *The Journal of Navigation* 68.4 (2015), págs. 646-664.

DECLARATION

Estos apuntes pueden contener errores de tipeo, en las formulas y hasta en los dibujos! Si encuentran alguno por favor comentenme por email a fmasson@uns.edu.ar. Solo pretende ser un acompañamiento al curso de posgrado por si no se entiende mi letra en las clases.

Bahía Blanca, 2021

Favio Masson