

El sistema GPS

José A. López-Salcedo
José López Vicario
Gonzalo Seco Granados

PID_00185449



Los textos e imágenes publicados en esta obra están sujetos –excepto que se indique lo contrario– a una licencia de Reconocimiento-NoComercial-SinObraDerivada (BY-NC-ND) v.3.0 España de Creative Commons. Podéis copiarlos, distribuirlos y transmitirlos públicamente siempre que citéis el autor y la fuente (FUOC. Fundació para la Universitat Oberta de Catalunya), no hagáis de ellos un uso comercial y ni obra derivada. La licencia completa se puede consultar en <http://creativecommons.org/licenses/by-nc-nd/3.0/es/legalcode.es>

Índice

| | |
|--|----|
| Introducción | 5 |
| Objetivos | 7 |
| 1. Segmentos del sistema GPS | 9 |
| 1.1. Segmento espacial | 9 |
| 1.2. Segmento de control | 12 |
| 1.3. Segmento de usuario | 14 |
| 2. Bandas frecuenciales de operación | 16 |
| 3. Estructura de señal | 17 |
| 3.1. Formato de señalización | 17 |
| 3.1.1. Código C/A | 20 |
| 3.1.2. Código P | 25 |
| 3.2. Mensaje de navegación | 25 |
| 3.3. Señales modernizadas | 27 |
| 4. Arquitectura del receptor GPS | 29 |
| 4.1. Etapa de adquisición de código | 30 |
| 4.2. Etapa de <i>tracking</i> | 34 |
| 4.2.1. Discriminadores de DLL | 36 |
| 4.2.2. Discriminadores de PLL | 40 |
| 4.2.3. Filtros de lazo | 42 |
| 4.2.4. Caracterización en lazo cerrado | 44 |
| 5. Cálculo de la posición | 51 |
| 5.1. Cálculo de los pseudorrangos | 51 |
| 5.2. Cálculo de las fases de portadora | 55 |
| 5.3. Solución de navegación | 58 |
| 5.3.1. Métodos basados en la linealización | 62 |
| 5.3.2. Métodos con solución cerrada | 64 |
| 5.3.3. Algoritmos completos para la solución de navegación | 66 |
| 5.4. Fuentes de error | 69 |
| 5.4.1. Factores DOP. Factores geométricos que afectan al cálculo de la solución de navegación | 71 |
| 5.4.2. Errores de medida en los observables que afectan al cálculo de la solución de navegación | 74 |
| 5.4.3. Errores del segmento de control: reloj del satélite y efemérides | 75 |

| | | |
|---|---|-----|
| 5.4.4. | Efectos de la propagación: la ionosfera | 77 |
| 5.4.5. | Efectos de la propagación: la troposfera | 85 |
| 5.4.6. | Efectos locales en el receptor: el ruido | 90 |
| 5.4.7. | Efectos locales en el receptor: las interferencias | 92 |
| 5.4.8. | Efectos locales en el receptor: el multicamino | 97 |
| 5.5. | Posicionamiento diferencial | 109 |
| 5.5.1. | Clasificación y reducción de las fuentes de error | 110 |
| 5.5.2. | Posicionamiento absoluto basado en el código | 116 |
| 5.5.3. | Posicionamiento relativo basado en el código | 117 |
| 5.5.4. | Posicionamiento relativo con fases de portadora | 118 |
| 5.5.5. | Resolución de las ambigüedades de fase | 124 |
| 5.5.6. | Métodos de resolución de la ambigüedad <i>geometry-free</i> .. | 127 |
| 5.5.7. | Métodos de resolución de la ambigüedad <i>geometry-</i> <i>based</i> | 136 |
| 5.5.8. | Posicionamiento absoluto con fases de portadora | 140 |
| 6. | Algoritmos de alta sensibilidad | 141 |
| 6.1. | Algoritmos de HS-GNSS basados en la DFT | 145 |
| Resumen | | 152 |
| Ejercicios de autoevaluación | | 159 |
| Solucionario | | 163 |
| Abreviaturas | | 164 |
| Bibliografía | | 166 |

Introducción

El sistema de radionavegación por satélite GPS es ciertamente uno de los sistemas más extendidos y utilizados actualmente. Su origen proviene del interés que se suscitó entre varias organizaciones gubernamentales de Estados Unidos en la década de los sesenta para desarrollar un sistema de posicionamiento de alta precisión y con cobertura global.

Tal como se ha comentado en el módulo "Sistemas de radionavegación", fruto de ese interés nació el primer sistema de radionavegación por satélite, conocido como Transit. No obstante, dicho sistema sufría varias limitaciones, como la obtención de medidas de posición con muy baja frecuencia (dependiendo de la latitud del blanco, el tiempo entre medidas disponibles podía variar entre 10 y 30 minutos) y el gran tiempo de cálculo requerido por el receptor para determinar la posición en sí (entre 10 y 15 minutos). Por este motivo, varias instituciones y organizaciones llevaron a cabo el estudio de diferentes alternativas. Entre estas instituciones se encontraban la propia desarrolladora de Transit, es decir, el Laboratorio de Física Aplicada de la Universidad de John Hopkins, las Fuerzas Aéreas de Estados Unidos y la Armada de Estados Unidos. Dichas alternativas proponían el cálculo de la posición mediante diferentes métodos basados en la obtención de distancias a partir de la estimación del tiempo de llegada de señales basadas en secuencias pseudoaleatorias, la determinación de ángulos o medidas basadas en el efecto Doppler, etc.

Finalmente, el Departamento de Defensa de Estados Unidos estableció en el año 1969 el programa de DNSS (Defense Navigation Satellite System) con el objetivo de unir en un mismo sistema todos los esfuerzos llevados a cabo hasta el momento por las diferentes organizaciones gubernamentales. Dicho sistema fue denominado como NAVSTAR (NAVigation Satellite Time and Ranging) GPS, basado en la transmisión, por parte de los satélites, de señales formadas por secuencias pseudoaleatorias, a partir de las cuales se obtiene la distancia entre el blanco y cada satélite mediante la obtención del tiempo de propagación de la propia señal. Otra de las características de este sistema es que los satélites se encuentran equipados con relojes atómicos para poder asegurar la sincronización del sistema.

El primer satélite de la constelación GPS fue lanzado en 1978, y se declaró en el año 1993 como sistema con capacidad operacional inicial. A principios del año 1995, una vez se habían lanzado ya 24 satélites y se había testado intensivamente el segmento de control terrestre, el sistema pasó a declararse con capacidad operacional total.

Tiempo de propagación

En inglés, *time of arrival* (TOA).

En este módulo, se presentan las características principales del sistema GPS. En primer lugar, se describe la arquitectura del sistema, basada en tres segmentos: el segmento espacial, el segmento de control y el segmento de usuarios. Hecho esto, se presentan las bandas frecuenciales de operación del sistema y la estructura de las señales de navegación.

En la segunda parte del módulo se presenta el subsistema de procesado de señal del receptor. Este subsistema es el encargado de generar las medidas de retardo y fase a partir de la señal recibida. Estas medidas dan lugar a los observables, que constituyen el punto de partida del cálculo de la posición, que se estudia en la última parte del módulo. Se analizan las dos fases fundamentales del proceso de sincronización, que son la adquisición y el seguimiento, y las arquitecturas más habituales para llevarlas a cabo.

En la última parte del módulo se explica en detalle el proceso que permite calcular la posición a partir de los diferentes observables generados por el subsistema de procesado de señal del receptor. Se presentan diferentes métodos de posicionamiento en función del tipo de observable que se utiliza y de la precisión requerida. También se analizan las técnicas de procesado de señal y de datos que se han de adoptar si se quieren utilizar los sistemas GNSS en lugares donde la señal se recibe fuertemente atenuada, como en zonas urbanas o entornos interiores.

Objetivos

Los objetivos de este módulo son:

1. Introducir la arquitectura del sistema GPS basada en tres segmentos: segmento espacial, segmento de control y segmento de usuarios.
2. Mostrar las bandas frecuenciales utilizadas en el sistema.
3. Describir la estructura de la señal GPS, tanto el formato de señalización como el mensaje de navegación.
4. Describir la arquitectura de un receptor GPS y el funcionamiento de los bloques que la integran.
5. Describir los fundamentos matemáticos que permiten calcular la posición del usuario a partir de la solución de navegación.
6. Conocer las fuentes de error que afectan al posicionamiento.
7. Comprender los métodos de procesado de datos que permiten combatir dichas fuentes de error.
8. Formular diferentes tipos de ecuaciones de navegación y deducir algoritmos para solucionarlas.
9. Entender los fundamentos del posicionamiento diferencial y con fases de portadora.
10. Presentar los fundamentos de los receptores GPS de alta-sensibilidad, aquellos que operan en condiciones de trabajo para las que el sistema GPS no fue inicialmente diseñado.

1. Segmentos del sistema GPS

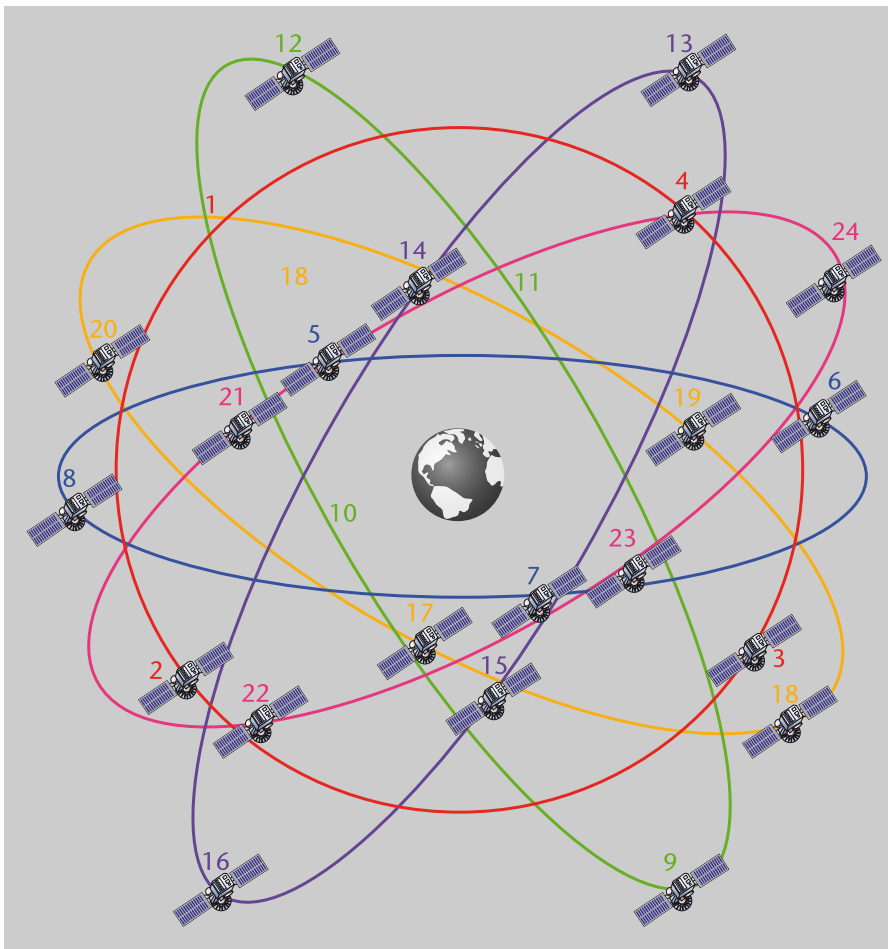
La arquitectura del sistema GPS¹ está dividida en tres segmentos: segmento espacial, segmento de control y segmento de usuario. A continuación se describe cada segmento con más detalle.

⁽¹⁾GPS son las siglas de *Global Positioning System*.

1.1. Segmento espacial

El segmento espacial de GPS está formado por la constelación de satélites del sistema. Estos satélites son los encargados de transmitir las señales de radionavegación. La configuración base consta de 24 satélites distribuidos en 6 planos orbitales con 4 satélites en cada plano (figura 1).

Figura 1. Constelación base de GPS



Las órbitas de los satélites son prácticamente circulares (excentricidad de 0,02), con unos 26.600 km de radio, y presentan una inclinación de 55° y un período orbital de 11 horas y 58 minutos (medio día sideral). Esta órbita permite que

como mínimo 4 satélites sean visibles las 24 horas del día en cualquier punto de la Tierra. Los parámetros orbitales (o parámetros keplerianos) de los satélites son los siguientes (sujetos a tolerancias):

- Semieje mayor de la órbita (a): 26.600 km.
- Excentricidad de la órbita (e): 0,02.
- Inclinação del plano de órbita (i): 55° .
- Longitud del nodo ascendente (Ω): $\pm 2^\circ$.
- Argumento del perigeo u orientación (ω): $\pm 180^\circ$.

La identificación de cada satélite se puede llevar a cabo de diferentes maneras, las más comunes son:

- **Posición en plano orbital.** En este caso se asigna una letra a cada plano orbital (es decir, A, B, C, D, E y F) y a cada uno de los satélites en dicha órbita se le asigna un número del 1 al 4. Por ejemplo, el identificador A2 correspondería al segundo satélite del plano orbital A.
- **SVN².** Es el número asignado por las Fuerzas Aéreas de Estados Unidos. Un ejemplo sería que el identificador de un satélite fuera el SVN 28.
- **Código PRN.** Cada satélite tiene un código pseudoaleatorio³ asignado. Por tanto, otra manera de identificar los satélites es mediante el número asignado a dicho código.

⁽²⁾SVN son las siglas de *space vehicle number*.

⁽³⁾En inglés, *pseudorandom noise* (PRN).

Ved también

Los código PRN se estudian en el apartado 3 de este módulo.

Tal como se ha mencionado anteriormente, el primer satélite del sistema GPS fue lanzado en 1978. Este satélite correspondía al primer bloque de satélites lanzados al espacio para llevar a cabo tareas de experimentación. En concreto, el desarrollo de GPS se ha llevado a cabo de manera continua desde mediados de los años setenta, dividiéndose este desarrollo en distintas fases. Cada una de estas fases tiene asociado un conjunto o bloque de satélites:

- **Bloque I.** Es el bloque asociado a la fase de experimentación del sistema. Se lanzaron 11 satélites entre los años 1978 y 1985; quedó una constelación de 10 satélites al fallar uno de ellos. De los 10 satélites, 4 disponían de relojes de cuarzo, mientras que en el resto se incorporaron relojes atómicos (3 de rubidio y 3 de cesio). Estos satélites estaban diseñados para tener una vida de 5 años, pero algunos de ellos llegaron a los 10 años.
- **Bloque II.** Es el bloque compuesto por los primeros satélites de la fase de producción. Gracias a la experiencia acumulada en el despliegue de los satélites del bloque I, se pudieron introducir mejoras significativas para alargar la vida de los satélites y su fiabilidad. Algunos ejemplos son el incremento de la robustez contra la radiación para evitar el malfuncionamiento de la memoria del equipo de a bordo debido a rayos cósmicos o modificaciones en la señalización utilizada para detectar errores en el sistema.

Ved también

En el apartado 3 se presentan con detalle los principales tipos de nuevas señales.

En este caso se lanzaron 9 satélites entre los años 1989 y 1990. Todos los satélites venían provistos de relojes atómicos de rubidio o cesio y fueron diseñados para tener una vida de 7,5 años, que se vio alargada a una media de 11,8 años (con algún caso que llegó a los 15 años).

- **Bloque IIA.** Los satélites de este bloque eran muy similares a los del bloque II pero se introdujeron algunas modificaciones para permitir un período de operación más largo (180 días frente a los 14 de los satélites del bloque II). Por otro lado, estos satélites tenían capacidad para operar continuamente durante 6 meses sin contacto con el control de tierra. No obstante, la precisión del sistema se degrada a medida que pasa el tiempo si no se aplican correcciones en las efemérides (discutidas en el apartado 3). Por tanto, se requiere una actualización diaria para conseguir buenas prestaciones. En este bloque se lanzaron 19 satélites entre los años 1990 y 1997. Aunque la vida esperada era igual a la del bloque II, aún quedan 10 satélites operativos en órbita. Cabe comentar que fue en el año 1995, una vez que se disponía ya de 24 satélites operativos y se había validado el segmento de control terrestre, cuando se declaró GPS como sistema con capacidad operacional total.
- **Bloque IIR.** Este bloque se conoce como bloque de satélites de reposición, ya que uno de sus principales objetivos era el ir reponiendo los satélites que habían sido lanzados en los bloques previos. Estos satélites se lanzaron entre los años 1997 y 2004. En total se lanzaron 13 satélites, de los cuales 1 falló y los 12 restantes siguen estando operativos.
- **Bloque IIR-M.** El bloque IIR-M (modernizado) está compuesto por un conjunto de satélites provistos de un diseño mejorado y funcionalidades adicionales. Algunos ejemplos son el uso de nuevas señales en la banda L, diseñadas especialmente para mejorar la precisión del sistema, y un rediseño de las cadenas de radio-frecuencia y de las antenas del satélite. En cuanto a las nuevas señales (tal como se describe en los apartados posteriores), se introducen un nuevo código militar en las bandas L1 y L2, conocido como código M, y una nueva señal civil en la banda L2, conocida como L2C. La banda L1 de los satélites de los bloques previos se utilizaba únicamente para transmitir una señal civil, conocida con el nombre de código C/A (*coarse acquisition*), y un código militar, conocido como código P (*precision*), mientras que la banda L2 se centraba en transmitir únicamente el código P. Finalmente, se ha de comentar que entre los años 2005 y 2009 se lanzaron 8 satélites, de los cuales quedan 7 operativos.
- **Bloque IIF.** Tal como se comenta en el siguiente punto, se prevé lanzar en el futuro un bloque de satélites orientado a ofrecer una arquitectura lo suficientemente flexible como para satisfacer los requisitos del sistema durante las próximas décadas. No obstante, hasta que este bloque no se encuentre totalmente operativo, se dispone de un programa de mantenimiento del sistema basado en los satélites del bloque IIF. El primero de

estos satélites fue lanzado en el año 2010 y el segundo fue lanzado en el 2011. En total, el bloque constará de 11 satélites, los cuales incluyen algunas mejoras, como la inclusión de una nueva señal civil en L5 y procesadores embarcados a bordo reprogramables que permiten actualizaciones del software del satélite.

- **Bloques III-A, III-B y III-C.** Tal como se ha comentado en el punto anterior, el programa III de GPS tiene como objetivo satisfacer los requisitos militares y civiles que puedan aparecer en las próximas décadas. Teóricamente, los satélites del bloque III-A empezarán a lanzarse en el 2014 y se tiene previsto lanzar un total de 12 satélites. Aún no existen previsiones para las fechas de lanzamiento de los bloques III-B y III-C, que dispondrán de 8 y 16 satélites respectivamente.

Finalmente, cabe comentar que la constelación de satélites actual (en el momento de redacción de este material) está formada por 31 satélites: 10 del bloque IIA, 12 del bloque IIR, 7 del bloque IIR-M y 2 del bloque IIF.

Plena capacidad del sistema GPS

El primer satélite del sistema GPS fue lanzado en el año 1978 y, desde entonces, se han lanzado un total de 60 satélites, de los cuales 31 siguen en órbita. El sistema fue declarado como sistema con capacidad operacional total en el año 1995, una vez se disponía de 24 satélites operativos y se había validado el segmento de control terrestre. No obstante, uno de los primeros casos donde el sistema fue ampliamente utilizado fue durante la Guerra del Golfo en los años 1990 y 1991.

1.2. Segmento de control

El segmento de control terrestre es el encargado de monitorizar y controlar la constelación de satélites de GPS. Concretamente, las principales funcionalidades de este segmento son:

- Monitorizar las señales de navegación de los satélites.
- Actualizar los mensajes de navegación.
- Resolver anomalías de los satélites.
- Monitorizar el estado de cada satélite.

El segmento de control está formado por los siguientes elementos:

1) **Estación de control maestra**⁴. Es la encargada de centralizar el control de la constelación GPS. Dicha estación se encuentra bajo el control de las Fuerzas Aéreas de Estados Unidos y está localizada en la Base Aérea Schriever, en Colorado Springs (Estados Unidos). A su vez, existe una estación redundante en Maryland. Las principales funciones de la MCS son:

- Monitorizar el estado de los satélites y llevar a cabo acciones de mantenimiento.

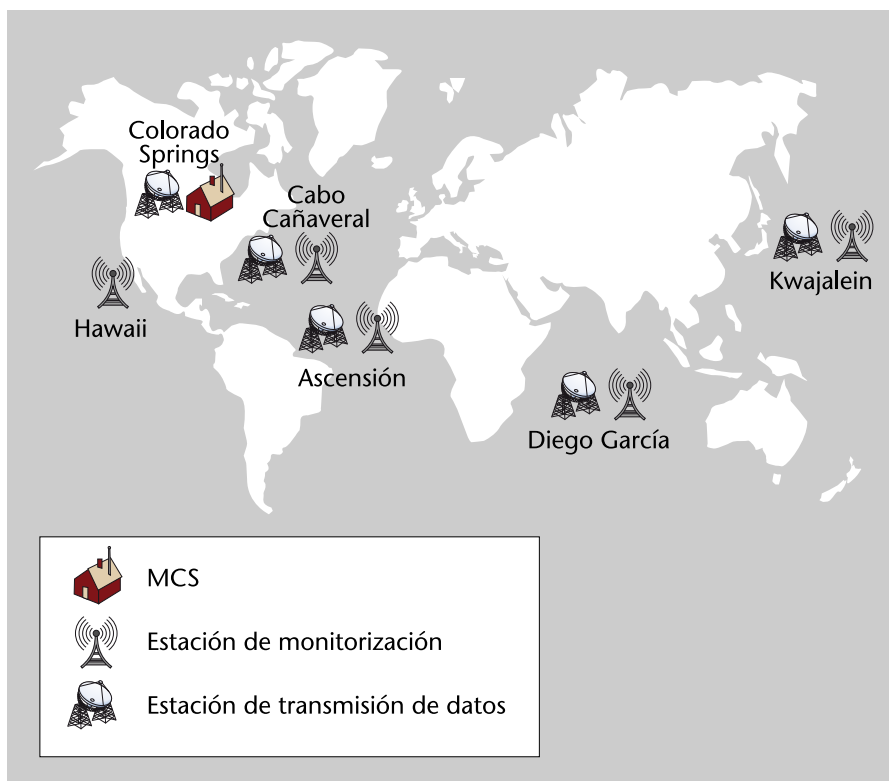
⁽⁴⁾En inglés, *master control station* (MCS).

- Monitorizar las órbitas de los satélites.
- Estimar y predecir los relojes de los satélites y sus efemérides.
- Generar los mensajes de navegación.
- Mantener la sincronización temporal del sistema GPS.
- Verificar los datos de navegación que se destinan a los usuarios del sistema.
- Maniobrar los satélites para mantener las órbitas y para llevar a cabo reposiciones de los satélites defectuosos.

2) **Estaciones de monitorización**⁵. Están encargadas de llevar a cabo las funciones de seguimiento de las señales de navegación de los diferentes satélites. Tal como se puede observar en la figura 2, las estaciones se encuentran en emplazamientos próximos al ecuador para maximizar su cobertura.

⁽⁵⁾En inglés, *monitor stations*.

Figura 2. Distribución geográfica del segmento de control de GPS



Cada una de estas estaciones está controlada por la estación MCS y se encargan de recoger datos referentes a las distancias ente la estación y los diferentes satélites, el estado de estos y el estado meteorológico del lugar (para modelar el retardo que introduce la troposfera en la señal GPS). Estos datos son enviados a la MCS para que los procese. Concretamente, una estación de monitorización está formada por:

- Un receptor en la banda L.
- Dos relojes atómicos de cesio.
- Sensores meteorológicos.
- Un conjunto de servidores locales.
- Un equipo de comunicaciones (para establecer conexión con la MCS).

Seis estaciones de monitorización

Existen seis estaciones de monitorización repartidas por todo el mundo: Isla de Ascensión (Atlántico), Diego García (Índico), Kwajalein (Pacífico), Hawaii (Pacífico), Colorado Springs (Estados Unidos) y Cabo Cañaveral (Estados Unidos).

⁽⁶⁾En inglés, *ground antennas*.

3) Estaciones de transmisión de datos⁶. Están encargadas de transmitir las órdenes de control y los mensajes de navegación generados por la MCS a los satélites. Existen 4 estaciones localizadas en el mismo emplazamiento que el de algunas de las estaciones de monitorización: Isla de Ascensión, Diego García, Kwajalein y Cabo Cañaveral (podéis ver la figura 2). Estas estaciones operan en la banda S, ya que es la banda utilizada para el enlace de control entre el segmento de la Tierra y los satélites. Por otra parte, estas estaciones también reciben en esa banda datos de telemetría de los diferentes satélites, que son, a su vez, reenviados a la MCS.

1.3. Segmento de usuario

El segmento de usuario está formado por el conjunto de receptores GPS.

Existen diferentes formas de clasificación, tal como se muestra a continuación:

1) Tipos de usuario:

a) Civil. Son los receptores utilizados por todos aquellos usuarios "no autorizados" por el Gobierno de Estados Unidos. Tal como se muestra en el apartado 3, estos usuarios se encuentran limitados en cuanto al tipo de señales GPS que pueden utilizar y, por tanto, las prestaciones que obtienen del servicio son menores.

b) Militar. Es el grupo de usuarios formado por las fuerzas armadas de Estados Unidos y otras organizaciones con autorización por parte del Departamento de Defensa de Estados Unidos, como son, por ejemplo, las fuerzas armadas de la OTAN. Disponen de mayor precisión al no sufrir de ninguna limitación del servicio.

2) Tipos de arquitectura de recepción:

a) Canales independientes. Son los receptores que disponen de un número determinado de canales que se utilizan de manera independiente para tratar las señales de los satélites bajo seguimiento. Normalmente, el número de canales suele ser de entre cuatro y doce.

b) Secuencial. Son los receptores con un único canal que se utiliza para recibir la señal de un solo satélite. El tiempo que el receptor opera con el satélite es de al menos un segundo, ya que es el tiempo mínimo para poder obtener la señal y la medida de pseudorange. Pasado el tiempo asignado a un satélite, el receptor pasa a recibir la señal de otro y, de este modo, se van obteniendo de manera secuencial los pseudorange o estimaciones de pseudodistancia de diferentes satélites para poder calcular la medida de posición.

c) **Multiplex:** como en el caso anterior, el seguimiento de los satélites se lleva a cabo de manera secuencial, pero la frecuencia de salto entre los diferentes satélites es más rápida para poder llevar cabo, mediante multiplexación temporal, el seguimiento de más de un satélite de modo simultáneo. Cada fracción de tiempo utilizada para un satélite debe ser el múltiplo del período de un bit de señal GPS (20 ms), pudiéndose llevar a cabo el seguimiento de hasta 5 satélites de manera simultánea en el intervalo de un segundo.

3) Capacidad frecuencial:

a) **Monofrecuencia.** Son los receptores que operan con una única banda de frecuencia (banda L1).

b) **Multifrecuencia** Son los receptores que pueden tratar diferentes bandas frecuenciales (bandas L1, L2 y, para el caso más reciente de satélites del bloque IIR-M, L5). Esta capacidad multifrecuencia permite que se puedan corregir efectos adversos introducidos por la ionosfera.

2. Bandas frecuenciales de operación

El sistema GPS utiliza tres bandas frecuenciales para transmitir las señales de los satélites. Estas bandas se encuentran en la banda L, localizada en la región UHF (*ultra high frequency*) del espectro electromagnético. Concretamente, se utilizan las siguientes bandas:

- L1: banda centrada en $f_{L1} = 1.575,42$ MHz (longitud de onda $\lambda = 19$ cm), es utilizada para transmitir los códigos C/A y P. En el caso de los satélites correspondientes al bloque IIR-M en adelante, se utiliza también dicha banda para transmitir el nuevo código militar M.
- L2: centrada en $f_{L2} = 1.227,60$ MHz ($\lambda = 24$ cm), es utilizada para transmitir el código P. Al igual que en el caso anterior, el número de señales en esta banda se vio incrementado en el momento de lanzamiento de los satélites del bloque IIR-M. En concreto, se pasó a transmitir también el código M y la nueva señal civil L2C.
- L5: centrada en $f_{L5} = 1.176,45$ MHz ($\lambda = 25$ cm). Esta banda se empezó a utilizar muy recientemente por el primer satélite del bloque IIF. En concreto, se emplea para transmitir una nueva señal de uso civil conocida con el nombre de *Safety-of-Life Signal*.

Ved también

Los códigos C/A y P se describen en el apartado 3.

Cabe destacar que dichas frecuencias se obtienen a partir de una frecuencia fundamental de $f_0 = 10,23$ MHz, que es la frecuencia utilizada por el oscilador atómico del satélite ($f_{L1} = 154f_0$, $f_{L2} = 120f_0$ y $f_{L5} = 115f_0$). Tal como se ha comentado en el apartado anterior, existen receptores que disponen de la capacidad de operar con diferentes bandas simultáneamente. Este hecho ofrece la posibilidad de calcular el retardo introducido por la ionosfera mediante la comparación de dos señales en bandas diferentes, ya que dicho retardo es proporcional a la longitud de onda de la señal.

En cuanto a las bandas utilizadas para la comunicación entre las estaciones de transmisión de datos (*ground antennas*) y los satélites, se utiliza la banda S. Concretamente se utiliza una banda frecuencial centrada en 1.783,74 MHz para el enlace de subida y una banda frecuencia centrada en 2.227,5 MHz para el enlace de bajada.

3. Estructura de señal

En este apartado se describe la estructura de las señales emitidas por los satélites de la constelación GPS para ofrecer el servicio de posicionamiento a los usuarios. Tal como se ha comentado anteriormente, los satélites de los primeros bloques transmitían únicamente señales basadas en los códigos C/A y P, y hasta que se lanzaron los primeros satélites modernizados, del bloque IIR-M, no se empezaron a introducir nuevas señales, como las basadas en el código M y la señal *Safety-of-Life*. Con el objetivo de presentar los conceptos fundamentales de la estructura de la señal GPS, este apartado se centrará principalmente en la descripción de las señales basadas en los códigos C/A y P, presentando al final de esta algunos detalles de las señales modernizadas.

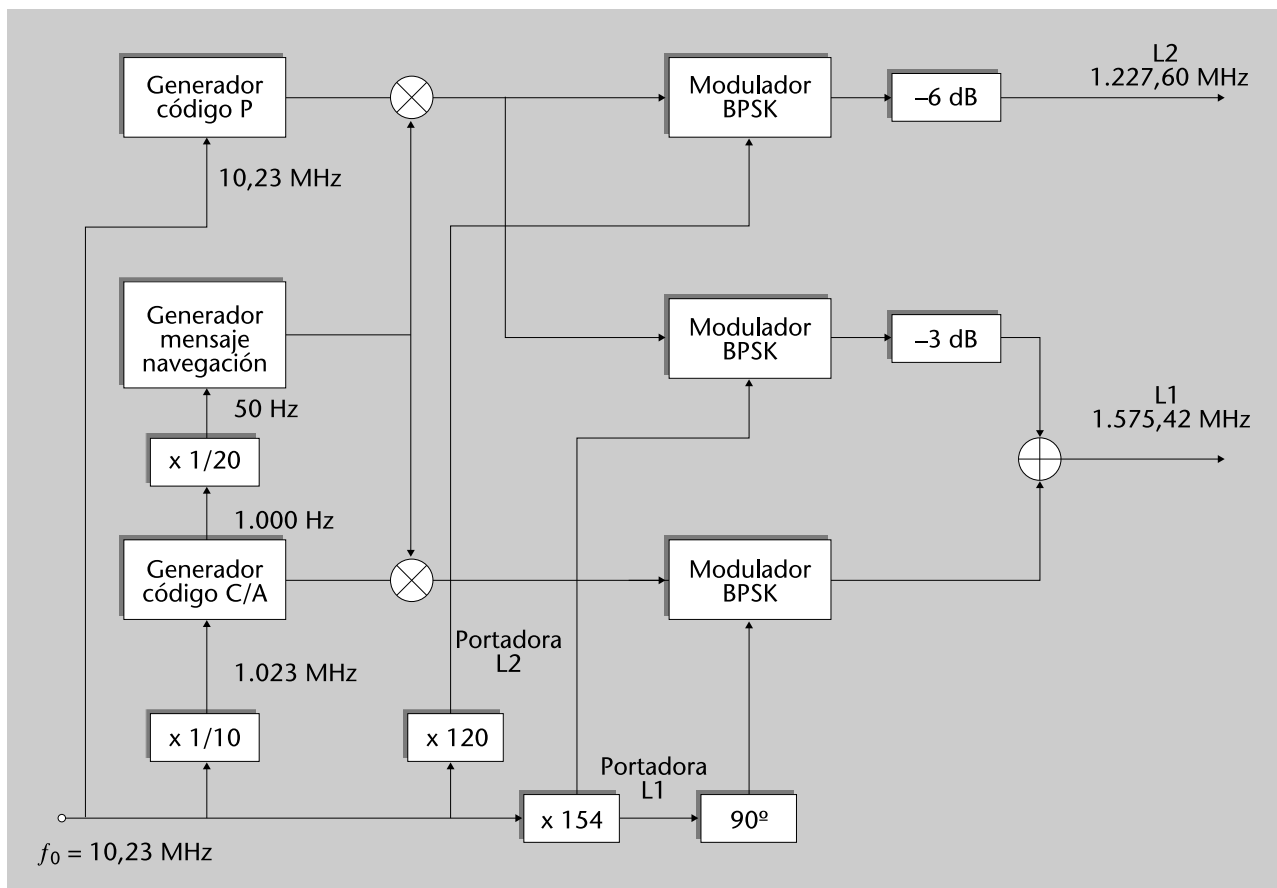
3.1. Formato de señalización

En la figura 3 se presenta el esquema de generación de las señales transmitidas por un satélite GPS, centrándonos en el caso tradicional donde la banda L1 se utilizaba únicamente para transmitir los códigos C/A y P y la banda L2 para transmitir la señal P.

L2

Aunque L2 también puede utilizarse para transmitir el código C/A, no es el caso normal y, por tanto, no se tiene en cuenta.

Figura 3. Esquema de transmisión de señales GPS



Caso tradicional basado en transmisión de C/A en L1 y P en L1 y L2

Tal como se puede observar, la señal transmitida por un satélite, denotado con el índice k , viene dada por la siguiente expresión:

$$s_k(t) = \sqrt{2P_C} C_k(t) D_k(t) \cos(2\pi f_{L1} t) + \sqrt{2P_{PL1}} P_k(t) D_k(t) \sin(2\pi f_{L1} t) + \sqrt{2P_{PL2}} P_k(t) D_k(t) \sin(2\pi f_{L2} t) \quad (1)$$

donde P_C , P_{PL1} y P_{PL2} son las potencias de las señales asociadas a los códigos C/A, P en banda L1 y P en banda L2, respectivamente. $C_k(t)$ y $P_k(t)$ son las secuencias temporales correspondientes a los códigos C/A y P asignados al satélite k -ésimo, mientras que $D_k(t)$ es el mensaje de navegación transmitido por dicho satélite.

Tal como se puede observar en la expresión, los datos del mensaje de navegación son combinados con los diferentes códigos mediante el operador producto. Debido a que tanto $D_k(t)$ como $C_k(t)$ y $P_k(t)$ son señales de naturaleza binaria, representándose el "0" lógico con una amplitud igual a -1 y un "1" lógico con una amplitud $+1$ (o viceversa), el tipo de modulación utilizada en estas señales es de tipo BPSK. Es decir, se modula una serie de portadoras con amplitudes binarias $+1$ y -1 (equivalente a meter desfases de 0° o 180° en la portadora según el símbolo que se quiera transmitir). Por otro lado, cabe comentar que en el caso de la banda L1, la señal BPSK asociada al código C/A se envía en la componente en fase de la señal (mediante el uso del $\cos(2\pi f_{L1} t)$) y la señal BPSK⁷ asociada al código P en la componente en cuadratura (mediante el uso del $\sin(2\pi f_{L1} t)$). Por tanto, el receptor puede separar ambas contribuciones perfectamente.

Los códigos C/A y P, por su parte, son conocidos también con el nombre de códigos de ensanchamiento (*spreading*). Esto se debe a que las señales asociadas a estos códigos están formadas por unas secuencias de bits – denominados en la jerga GPS como *chips* para denotar que no transportan ninguna información– con un período de chip mucho menor que el período de bit del mensaje de navegación.

En el caso de C/A, por ejemplo, el período de chip es de $1/1.023$ ms, mientras que, en el caso de los bits de navegación, el período de bit es de 20 ms. Por tanto, al hacer el producto del mensaje de navegación $D_k(t)$ con la señal $C_k(t)$ se obtiene como resultado una señal de navegación con un tiempo de bit menor que el de la señal original. Esto en espectro se traduce en que se tiene una señal de navegación con el espectro más ancho, ya que el ancho de banda de la señal, B , es proporcional al inverso del período de bit, T_b , es decir $B \propto 1/T_b$. En

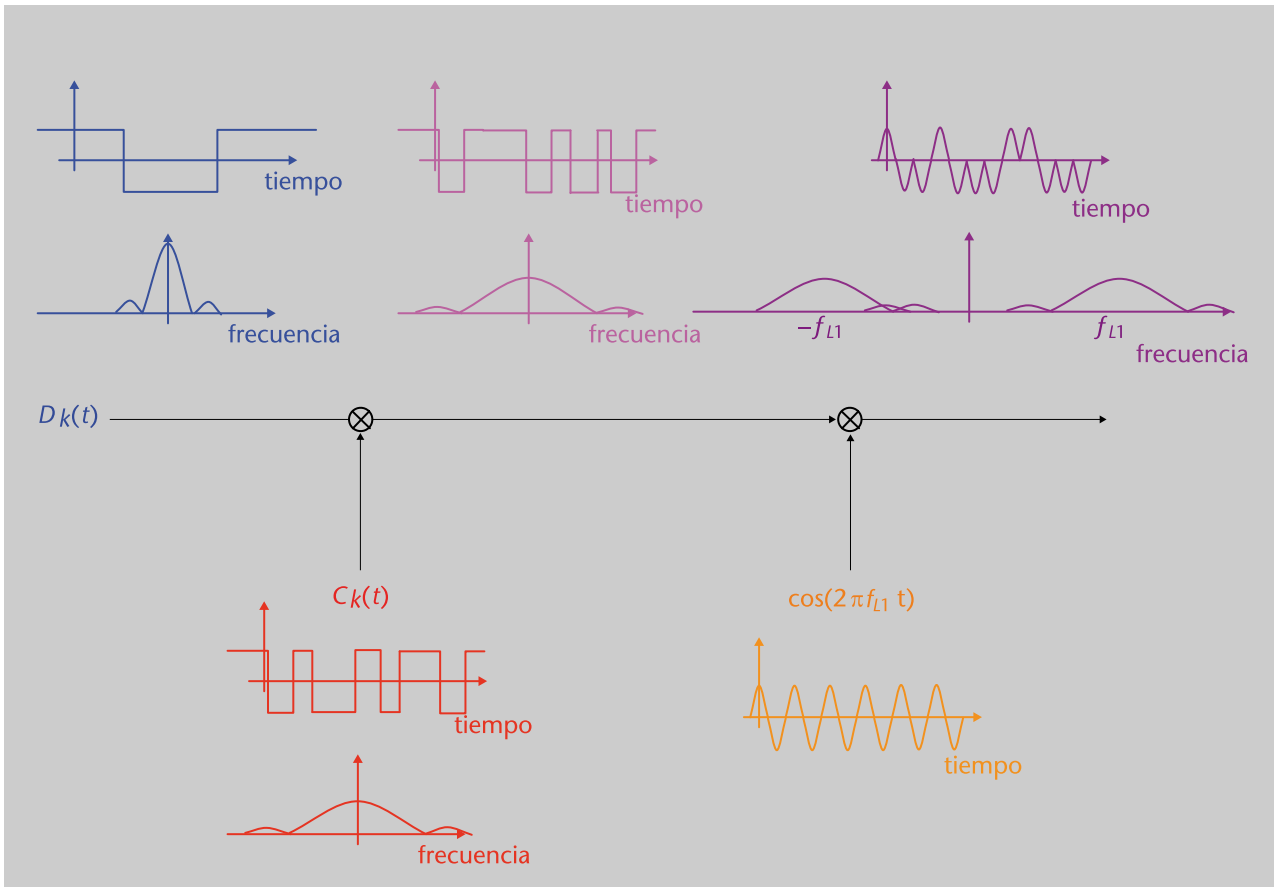
Ved también

Para una descripción más detallada del mensaje de navegación podéis ver el subpartado 3.2 de este módulo.

⁽⁷⁾BPSK son las siglas de *binary phase-shift keying*.

la figura 4 se presenta el proceso de modulación comentado arriba para el caso de la señal asociada al código C/A, observándose el efecto de ensanchamiento del espectro.

Figura 4. Proceso de modulación de la señal asociada al código C/A



Cada satélite dispone de sus propios códigos C/A y P, que son diferentes al resto de los de otros satélites. El objetivo de usar estos códigos viene motivado por las siguientes razones:

- Los códigos están basados en secuencias pseudoaleatorias que permiten obtener buenas precisiones en las estimaciones de las distancias entre el usuario y los satélites.
- El uso de diferentes códigos en cada uno de los satélites permite que el receptor pueda distinguir señales recibidas de distintos satélites simultáneamente en la misma frecuencia. Esto se conoce como técnica de acceso al medio CDMA⁸.
- El distribuir la energía de la señal en un espectro más amplio permite mejorar la robustez del sistema ante interferencias de banda estrecha. Es decir, una interferente de banda estrecha distorsionará parte de la señal GPS pero, dado el mayor ancho de banda de esta, el resto del espectro de la señal no se verá afectado.

⁽⁸⁾CDMA son las siglas *code division multiple access*.

A continuación se describen las características de los códigos C/A y P con más detalle, mostrándose además cómo su estructura cumple las propiedades mencionadas en los puntos de arriba.

3.1.1. Código C/A

El código C/A está formado a partir de una secuencia $p_k(t)$ binaria (+1 y -1) con una longitud de 1.023 chips. El código se transmite periódicamente mediante la señal $C_k(t)$, de modo que una secuencia completa se transmite en el período temporal $T_{per} = 1$ ms.

Es decir, el aspecto que presenta la señal $C_k(t)$ es:

$$C_k(t) = \sum_{i=-\infty}^{\infty} p_k(t - iT_{per}) \quad (2)$$

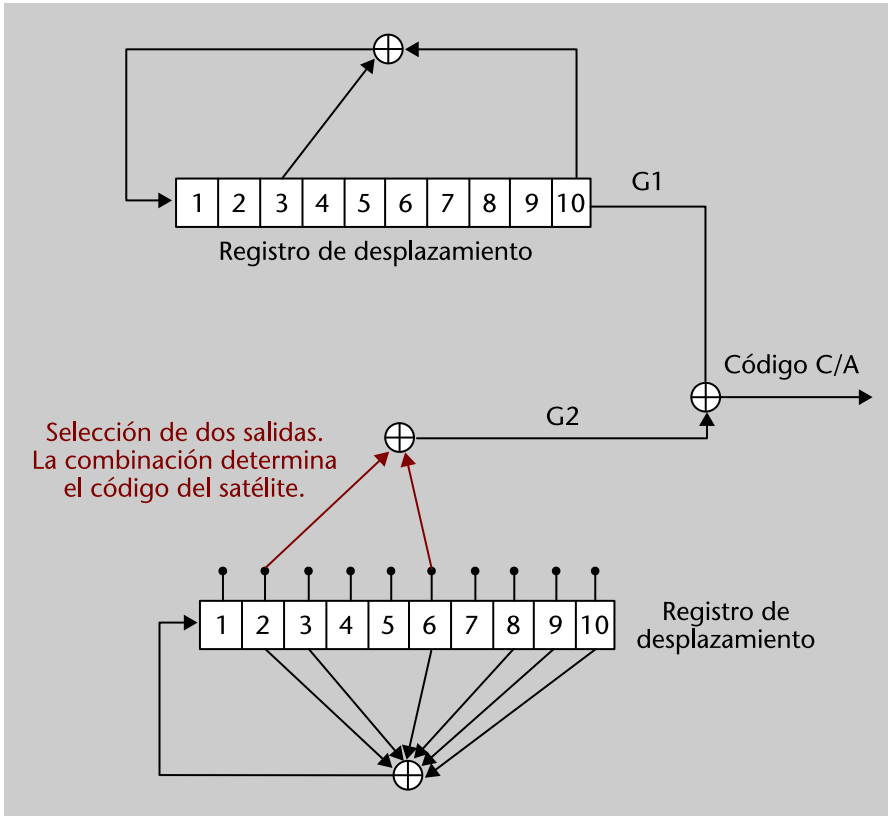
El tiempo de chip, por tanto, es de $1/1.023$ ms ≈ 1 μ s, lo que da por tanto una tasa de chip igual a 1,023 Mchips/s. Tal como se ha comentado previamente, este código se multiplica al mensaje de navegación $D_k(t)$. Teniendo en cuenta que los datos transmitidos en el mensaje de navegación tienen una velocidad de bit igual a 50 bps, es decir, cada bit ocupa 20 ms, se obtiene como resultado que un bit de navegación contiene 20 períodos del código C/A concatenados en el tiempo.

El código C/A en sí está basado en un tipo de secuencias conocidas como secuencias pseudoaleatorias, denominadas en inglés como *pseudorandom noise (PRN) codes*. Estas secuencias son deterministas, pero tienen propiedades similares a las que tiene el ruido blanco (de ahí su denominación). Estas características son que la autocorrelación de una señal C/A presenta un máximo en 0, mientras que, para el resto de valores, los niveles de autocorrelación son muy bajos. Esto se asemeja al comportamiento del ruido blanco, que tiene una función de autocorrelación igual a una delta de Dirac, ya que el ruido en un instante de tiempo determinado está totalmente incorrelado con el existente el resto del tiempo. Por otro lado, si se calcula la correlación cruzada con otros códigos C/A (de otros satélites), se obtienen unos valores de correlación muy bajos.

Los códigos PRN utilizados, concretamente, se basan en los códigos Gold. Estos códigos se generan usando dos registros de desplazamiento de 10 bits, tal como se representa en la figura 5. Cada uno de estos registros de desplazamiento genera una secuencia de longitud máxima, de longitud $2^{10} - 1$ bits, la cual ya presenta por sí misma características de secuencia PRN. No obstante, el código Gold se basa en combinar dichas secuencias haciendo la suma en módulo 2

de estas. Cabe mencionar que consideramos que los registros trabajan con niveles lógicos "0" y "1" y, una vez generado el código Gold, se lleva a cabo la transformación "0" \rightarrow -1 y "1" \rightarrow +1 para generar la secuencia $p_k(t)$.

Figura 5. Esquema de generación de un código C/A



En la misma figura se observa también que una de las secuencias, la secuencia G1, se genera siempre del mismo modo, mientras que la segunda secuencia, G2, se genera de manera diferente dependiendo de las combinaciones seleccionadas a la salida del segundo registro de desplazamiento. Esto es debido a que cada satélite genera una secuencia distinta variando las combinaciones utilizadas. En la tabla 1 se muestran las combinaciones de salidas del segundo registro utilizadas para generar los códigos C/A reservados para su empleo en satélites GPS.

Tabla 1. Combinaciones de salidas del registro G2 utilizadas en la generación de códigos C/A reservados para satélites GPS

| Código PRN | Combinación salidas G2 | Código PRN | Combinación salidas G2 |
|------------|------------------------|------------|------------------------|
| 1 | 2 \oplus 6 | 17 | 1 \oplus 4 |
| 2 | 3 \oplus 7 | 18 | 2 \oplus 5 |
| 3 | 4 \oplus 8 | 19 | 3 \oplus 6 |
| 4 | 5 \oplus 9 | 20 | 4 \oplus 7 |
| 5 | 1 \oplus 9 | 21 | 5 \oplus 8 |
| 6 | 2 \oplus 10 | 22 | 6 \oplus 9 |

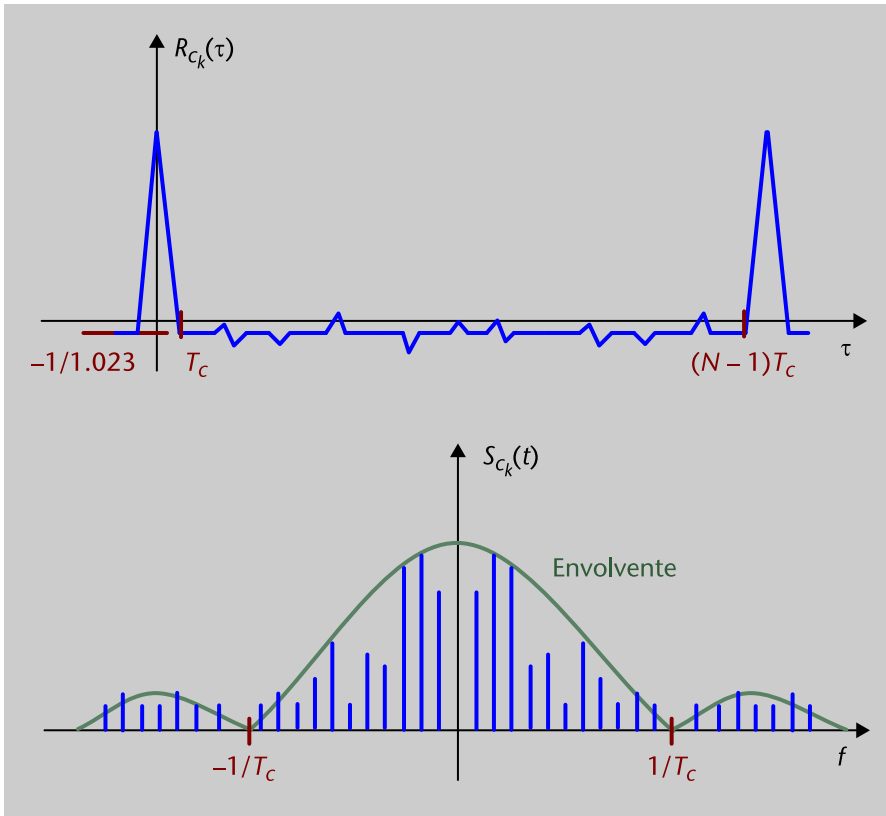
| Código PRN | Combinación salidas G2 | Código PRN | Combinación salidas G2 |
|------------|------------------------|------------|------------------------|
| 7 | 1 ⊕ 8 | 23 | 1 ⊕ 3 |
| 8 | 2 ⊕ 9 | 24 | 4 ⊕ 6 |
| 9 | 3 ⊕ 10 | 25 | 5 ⊕ 7 |
| 10 | 2 ⊕ 3 | 26 | 6 ⊕ 8 |
| 11 | 3 ⊕ 4 | 27 | 7 ⊕ 9 |
| 12 | 5 ⊕ 6 | 28 | 8 ⊕ 10 |
| 13 | 6 ⊕ 7 | 29 | 1 ⊕ 6 |
| 14 | 7 ⊕ 8 | 30 | 2 ⊕ 7 |
| 15 | 8 ⊕ 9 | 31 | 3 ⊕ 8 |
| 16 | 9 ⊕ 10 | 32 | 4 ⊕ 9 |

Las secuencias que aparecen en la tabla son el conjunto de secuencias (de todas las secuencias posibles) que ofrecen las mejores en términos de correlación cruzada, lo cual es interesante en la aplicación GPS al ofrecer la posibilidad de separar de modo más eficiente las señales de los distintos satélites. En términos de autocorrelación, estas secuencias ofrecen un comportamiento similar al que ofrece una secuencia de longitud máxima convencional. Como resultado, se obtiene que el valor de autocorrelación de la señal resultante,

$C_k(t) = \sum_{i=-\infty}^{\infty} p_k(t - iT_{per})$, expresado como:

$$R_{C_k}(\tau) = \lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^T C_k(t) C_k(t + \tau) dt \quad (3)$$

presenta un aspecto como el mostrado en la figura 6.

Figura 6. Autocorrelación y densidad espectral de potencia de $C_k(t)$ 

Es decir, el máximo está centrado en $\tau = 0$, obteniéndose el valor:

$$R_{C_k}(0) = 1 \quad (4)$$

mientras que en el rango $-(N-1)T_c \leq \tau \leq (N-1)T_c$, donde $N = 1.023$ es la longitud de la secuencia Gold, se obtiene un valor de autocorrelación igual a $-1/1.023$ salvo algunas fluctuaciones. Notad que la autocorrelación presentaría un valor constante en esa zona (igual a $-1/1.023$) en el caso de que se tratara de una secuencia de longitud máxima. Estas fluctuaciones, no obstante, cumplen lo siguiente:

$$|R_{C_k}(\tau)| \leq \frac{2^{(10+2)/2} + 1}{1.023} = \frac{65}{1.023} \approx \frac{1}{16} \quad \text{si } -(N-1)T_c \leq \tau \leq (N-1)T_c \quad (5)$$

Es decir, la autocorrelación presenta un comportamiento muy similar al del caso de la secuencia de longitud máxima. En cuanto a la correlación cruzada existente con los códigos de otros satélites, se cumple lo siguiente:

$$|R_{C_k C_l}(\tau)| = \left| \lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^T C_k(t) C_l(t + \tau) dt \right| \leq \frac{65}{1.023} \approx \frac{1}{16} \quad \forall \tau \quad (6)$$

que constituyen unos niveles mucho mejores que los obtenidos con las secuencias de longitud máxima convencionales. También cabe comentar que gran parte de los valores obtenidos al realizar la correlación cruzada son del orden de $-1/1.023$.

Por tanto, estas características ofrecen las siguientes propiedades:

- Dada la forma de la autocorrelación, se puede determinar con gran precisión cuál es el retardo de la señal a la entrada del receptor. Es decir, el receptor de un sistema GPS utiliza una réplica del código del satélite para realizar operaciones de correlación entre esta réplica y la señal entrante. Para ello, se utilizan diferentes retardos de código a la hora de generar la réplica. Aquel retardo que ofrezca un máximo a la salida del correlador es el retardo que se considera como retardo de la señal de entrada. Por tanto, este código facilita la tarea de estimación de la distancia a partir de la estimación del retardo de propagación (o TOA) de la señal de entrada.
- Por otra parte, dadas las propiedades de la autocorrelación, el sistema tiene una cierta robustez ante señales multicamino, ya que la correlación cruzada existente entre la señal de visión directa y los rebotes de la señal que lleguen con retardos diferentes a la del retardo presentará niveles bajos, es decir, el correlador del receptor podrá cancelar las señales multicamino. No obstante, tal como se presenta más adelante en el módulo, algunos efectos multicamino no pueden ser totalmente cancelados y presentan algunos problemas en el caso de que se requieran grandes precisiones.
- Teniendo en cuenta el aspecto de la correlación cruzada, las señales provenientes de otros satélites pueden ser también canceladas por el correlador del receptor. Por tanto, se pueden recibir señales de diferentes satélites de manera simultánea y en la misma frecuencia. Para seleccionar cada una de ellas, el correlador del receptor debe configurarse para que genere la réplica del código del satélite deseado.

En cuanto al espectro de dichas secuencias, $S_{C_k}(f) = \int_{-\infty}^{\infty} R_{C_k}(\tau) e^{-j2\pi f \tau} d\tau$, se observa el aspecto representado en la figura 6. Tal como se observa, gran parte de la potencia se concentra en la banda existente entre $-1/T_c$ y $1/T_c$. Por ese motivo, cuando se aplica dicho código al mensaje de navegación y se modula después la portadora $\cos(2\pi f_{L1} t)$, se tiene como resultado una señal con un espectro más ensanchado que el de la señal de navegación original a 50 Hz, ya que se concentra el 90% de la potencia de la señal en aproximadamente 2 MHz ($= 2/T_c$) de ancho de banda.

3.1.2. Código P

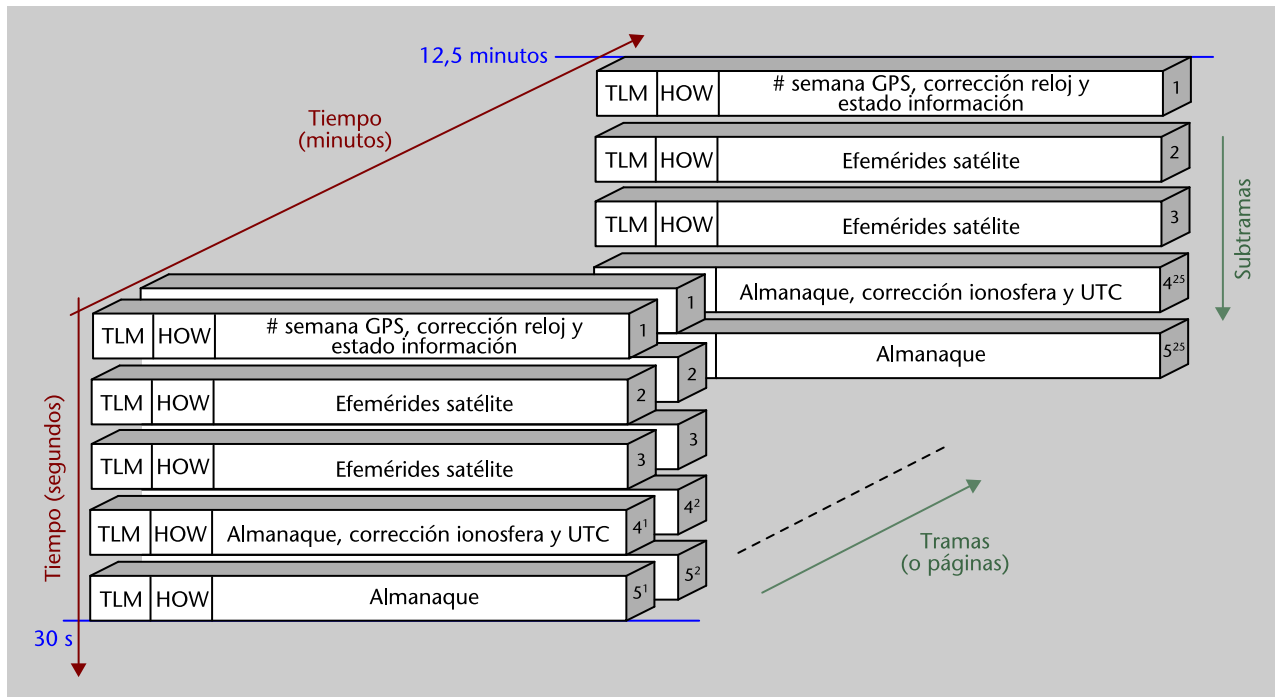
Este código, de uso exclusivamente militar, se construye de modo algo similar a los códigos Gold usados en C/A, pero, en este caso, se utilizan cuatro registros de desplazamiento de 12 bits. Concretamente, el código P tiene una longitud de $6,1871 \times 10^{12}$ chips ocupando un período temporal de 7 días, una velocidad de chip igual a 10,23 Mchips/s y, por tanto, un período de chip T_c de aproximadamente $0,1 \mu\text{s}$. Dada la alta velocidad de chip y la gran longitud de la secuencia, el código P presenta unas propiedades de autocorrelación y correlación cruzada excelentes. Concretamente, se obtienen niveles de autocorrelación del orden de $-1/6,1871 \times 10^{12}$ en la zona $-(N-1)T_c \leq \tau \leq (N-1)T_c$, presentándose los mismos niveles en el caso de correlación cruzada con los códigos de otros satélites. En cuanto al espectro, la mayor parte de la potencia de la señal modulada con este código se concentra en un ancho de banda de aproximadamente 20 MHz. Por tanto, el ensanchamiento del espectro es aún mayor, lo que ofrece al sistema una mayor robustez ante interferencias.

El código P es normalmente encriptado, activando un modo en el satélite conocido como modo AS (*anti-spoofing*), lo cual imposibilita que los usuarios civiles puedan utilizarlo. Cuando se activa el modo AS, el código P se transforma en lo que se conoce como código Y. No obstante, la velocidad de chip es la misma en ambos casos y, por ese motivo, la versión encriptada del código se suele denominar como código P(Y). Finalmente, cabe comentar que el cálculo de posición no se puede obtener con el código P(Y) por sí solo. Concretamente, se requiere una adquisición previa del código C/A.

3.2. Mensaje de navegación

El mensaje de navegación $D_k(t)$ contiene la información necesaria para que el usuario pueda calcular la posición del satélite k -ésimo y el tiempo de transmisión de la señal recibida. Aparte, también proporciona información útil, como, por ejemplo, el estado de los datos y la información de otros satélites para facilitar la adquisición de las señales de estos. Tal como se ha comentado anteriormente, dicha información se transmite a 50 bps y se organiza en tramas (o páginas) de la manera representada en la figura 7.

Figura 7. Estructura del mensaje de navegación GPS



Cada trama, de 30 s de duración, se divide a su vez en 5 subtramas de 6 s y 300 bits. Organizándose los 300 bits en 10 palabras de 30 bits, se tiene que las dos primeras palabras de cada subtrama son las palabras TLM (*TeLeMetry*) y HOW (*HandOver Word*), que se detallan a continuación:

- **TLM:** formada por 8 bits de preámbulo (secuencia 10001011) usados para que el receptor pueda reconocer el inicio de una subtrama, seguidos de 14 bits reservados para los usuarios autorizados, 2 bits que se dejan vacíos y 6 bits de paridad.
- **HOW:** llamado palabra *handover* porque permite al receptor realizar el paso de un código C/A a un código P(Y). Está formado por una versión trunca de 17 bits del TOW (*time of week*), que indica en qué instante de la semana se ha iniciado la transmisión de la subtrama (con una resolución de 6 segundos por el truncamiento), seguido de 2 bits de *flag* (el primero para indicar si el modo AS está activo y el segundo para indicar que la precisión de la señal puede ser muy mala), 3 bits para identificar la subtrama en sí, 6 bits de paridad y 2 bits vacíos.

En cuanto al resto de las palabras que contiene cada subtrama, se utilizan para transportar la siguiente información:

- **Subtrama 1:** la información contiene el número de semana GPS en la que se está transmitiendo la señal, además de términos de corrección del reloj del satélite e información sobre el estado de la información. La información de esta subtrama se actualiza en cada nueva trama (o página).

⁽⁹⁾UTC son las siglas de *universal coordinated time*.

- **Subtramas 2 y 3:** contienen las efemérides del satélite, compuestas por los parámetros orbitales y las anomalías, para poder determinar su posición. La información de estas subtramas también se actualiza en cada trama.
- **Subtramas 4 y 5:** a diferencia de las subtramas 1 a 3, la información de estas subtramas se distribuye en 25 tramas (o páginas). Por ese motivo, en la figura 7, se observa que dichas subtramas tienen 25 versiones (indicadas con el superíndice añadido). En concreto, estas subtramas contienen los almanagues (efemérides con menor resolución) de todos los satélites de la constelación para permitir que el receptor pueda adquirir la señal de otros satélites con más facilidad. Contienen además parámetros de corrección ionosféricos y parámetros para relacionar el reloj del sistema GPS con el sistema de tiempos UTC⁹.

Teniendo en cuenta la estructura de las diferentes subtramas, se observa que para enviar un mensaje de navegación completo se requieren 12,5 minutos (transmisión de 25 tramas de 30 s).

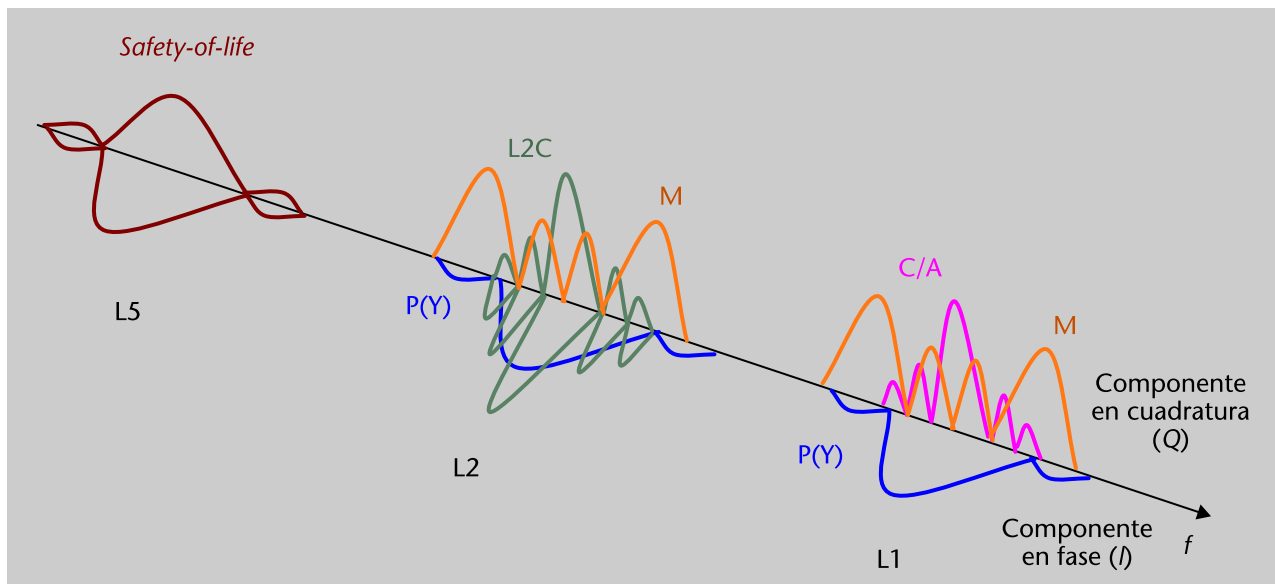
3.3. Señales modernizadas

Aparte de las señales C/A y P, existen también un conjunto de señales conocidas como señales modernizadas: L2C, M y *Safety-of-Life* (podéis ver la figura 8). Concretamente estas señales empezaron a utilizarse a partir de los lanzamientos de los satélites de los bloques IIR-M (L2C y M) e IIF (*Safety-of-Life*).

Ved también

Las señales C/A y P se estudian en el subapartado 3.1 de este módulo.

Figura 8. Representación espectral del conjunto de señales GPS



A continuación se presentan las principales características de estas señales:

- **L2C.** Esta nueva señal de uso civil se transmite en la portadora L2. Su objetivo es proporcionar mejores precisiones que las ofrecidas por la señal C/A. Está formada por dos secuencias PRN, transmitiéndose una en la compo-

nente en fase y la otra en la componente en cuadratura, con las siguientes propiedades:

- CM (*civilian moderate length code*): código con longitud de 10.230 bits y período de repetición igual a 20 milisegundos. Conjuntamente, se transmite un nuevo mensaje de navegación a 50 bps utilizando sistema de corrección de errores FEC (*forward error correction*) con tasa de código de 1/2. Por tanto, la tasa de bits de información útil (información de navegación exclusivamente) es de 25 bps.
- CL (*civilian long length code*): código con longitud igual a 767.250 bits y con 1.500 ms de período de repetición. No utiliza mensaje de navegación.
- M: nueva señal militar transmitida en las portadoras L1 y L2 orientada a ofrecer mejor robustez anti-*jamming* y un mayor nivel de seguridad. Sus principales características son las siguientes:
 - La señal se ha diseñado de modo que su espectro no interfiera con el de las señales ya existentes en esas bandas (podéis ver la figura 8). Esto se consigue utilizando un tipo de modulación conocida como BOC (*binary offset carrier*). En el caso de la señal M, concretamente, se utiliza el tipo de modulación BOC (10,5).
 - A diferencia del código P, el código M se puede utilizar para calcular la posición del usuario sin necesidad de adquirir el código C/A previamente.
 - En el futuro, en los satélites del bloque III, cada satélite transmitirá dos códigos M en dos antenas diferentes. Uno se transmitirá en una antena diseñada para ofrecer cobertura global y otro en una antena más directiva (con una ganancia 20 dB mayor) para ofrecer mejor cobertura en una zona determinada. Actualmente, solo se utiliza la antena con cobertura global.
- **Safety-of-Life**: señal transmitida en la banda L5 orientada a ofrecer servicios de búsqueda y salvamento. Para ello se utilizan dos códigos, uno en fase y otro en cuadratura, de 10,23 Mchip/s y longitud igual a 10.230 bits transmitidos en 1 ms. Por tanto, se utiliza un ancho de banda 10 veces superior al de la señal C/A, lo que ofrece una mayor robustez de la señal, y un código 10 veces más largo, dando mejores prestaciones en cuanto a autocorrelación y correlación cruzada. Por otro lado, el satélite emplea el doble de potencia de transmisión a la usada en las bandas L1 y L2 para transmitir esta señal.

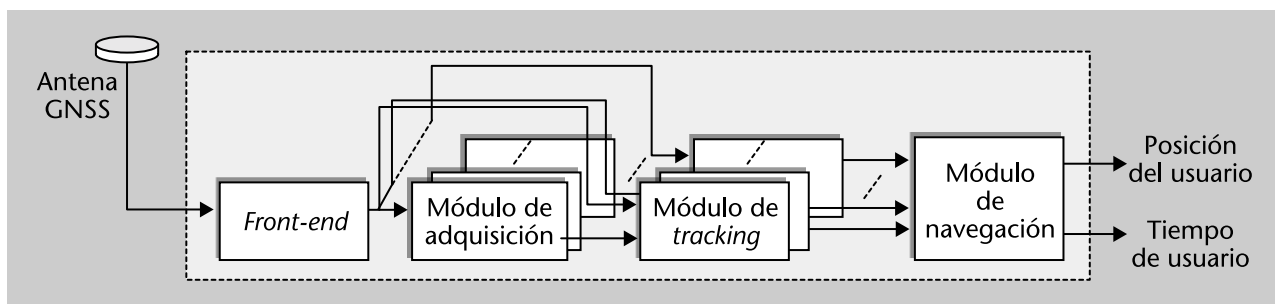
Ved también

La modulación BOC es descrita en el módulo "Evolución de la navegación por satélite: Galileo y sistemas híbridos" de esta asignatura.

4. Arquitectura del receptor GPS

El objetivo de todo receptor de GPS (o de GNSS, en general) es poder determinar la posición del usuario a partir de analizar una serie de señales que son transmitidas por los satélites que forman el segmento de espacio. El procedimiento que ha de seguir el receptor hasta llegar a determinar la posición del usuario no es directo, sino que implica realizar toda una serie de tareas que han de llevarse a cabo de manera secuencial. A grandes rasgos, estas tareas son las que se ilustran mediante los bloques que aparecen en la figura 9. En realidad, la arquitectura mostrada en esta figura corresponde al caso general de un receptor GNSS cualquiera (ya sea de GPS, Galileo o Glonass) pues las tareas que han de llevarse a cabo, dejando a un lado sus detalles correspondientes, pueden considerarse comunes a todos estos sistemas.

Figura 9. Arquitectura general de un receptor de GNSS, como por ejemplo GPS



Tal y como se puede apreciar en la figura 9, el primer paso, como en todo receptor de radiofrecuencia, es realizar el proceso de conversión a banda base de la señal paso banda captada por la antena. Este proceso lo lleva a cabo el bloque indicado como *front-end*, el cual habitualmente consta de un amplificador de bajo ruido (que de manera simultánea filtra paso banda y amplifica la señal resultante), una o varias etapas de conversión a frecuencia intermedia, un filtrado anti-aliasing ya en banda base y una conversión analógico/digital (ADC). A la salida del front-end, pues, el receptor dispone de un flujo constante de muestras discretas cuantificadas, típicamente a dos bits por muestra.

Cuantificación de 2 bits

La adopción de una cuantificación de tan solo 2 bits es habitual en sistemas de espectro ensanchado (como GPS) en donde la densidad espectral de señal útil resulta estar habitualmente por debajo de la densidad espectral de ruido. En estas circunstancias, utilizar más de 2 bits por muestra no aporta ninguna ventaja sustancial puesto que la señal recibida está dominada por ruido. Una excepción sería el caso de receptores GPS operando en escenarios donde es posible la presencia de señales interferentes, en cuyo caso sí que interesa aumentar el número de bits de cuantificación para evitar que el ADC se sature en presencia de una señal interferente de banda estrecha y alta potencia.

En el otro extremo del receptor, el módulo de navegación es el encargado propiamente de resolver la posición del usuario. A medio camino entre el front-end y el módulo de navegación es donde aparecen los módulos de adquisición y de *tracking*, que son realmente los encargados de procesar la señal recibida y de extraer de ella la información necesaria para el módulo de navegación sea capaz, posteriormente, de calcular la posición del usuario. La adquisición y el *tracking* son los dos elementos clave de la arquitectura de un receptor de GNSS, por lo que respecta al procesado de señal.

Ved también

El módulo de adquisición se estudia en el subapartado 4.1 de este módulo y el de *tracking* en el subapartado 4.2.

4.1. Etapa de adquisición de código

Una vez que la señal recibida ha sido correctamente acondicionada, la primera tarea propia de un receptor de GPS es buscar los satélites que actualmente tiene visibles y que, en caso de estarlo, habrán dejado su impronta en la señal captada por la antena. Estos satélites en vista sirven como puntos de anclaje sobre los que el receptor se apoyará para calcular la posición del usuario, una misión que será llevada a cabo a partir de medir el tiempo que tarda la señal transmitida por los satélites en llegar al receptor.

La obtención de este retardo, o al menos de una estimación tentativa, es el objetivo de la etapa de adquisición de código, la cual recibe este nombre puesto que el tiempo de retardo se mide en relación a un inicio del código PRN que emite cada satélite en cuestión.

En el proceso de buscar los satélites visibles y estimar el retardo en la señal recibida de cada uno de ellos, existen varios factores que hacen que esta tarea sea más complicada de lo previsto a priori. El factor más importante es la presencia de un desplazamiento de frecuencia Doppler en la frecuencia portadora que se recibe de cada satélite. Este desplazamiento Doppler varía de satélite a satélite, puesto que depende de la velocidad aparente que percibe el receptor para cada uno de ellos (es decir, la proyección del vector velocidad del satélite sobre el vector de posición que une al receptor con el satélite). La existencia de este error de frecuencia, desconocido, juntamente con el retardo que experimenta la señal recibida, también desconocido, hace que el proceso de adquisición de código implique una búsqueda bidimensional en tiempo y en frecuencia. Esta búsqueda se implementa correlando la señal recibida con una réplica local del código del satélite bajo análisis, la cual se genera con un cierto retardo y con una frecuencia, ambas tentativas. El proceso se repite probando para todos los posibles valores de retardo de código y de frecuencia Doppler, hasta que en una de las pruebas, el valor a la salida de la correlación sea suficientemente grande como para superar un umbral de decisión. En ese caso, el resultado confirma la presencia del satélite que se está analizando y, además, indica el valor aproximado de retardo de código y de error Doppler que presenta respecto la réplica local que está generando el receptor.

La figura 10 muestra un ejemplo de la representación que se obtendría en la búsqueda tiempo-frecuencia anteriormente comentada, en el caso de que el satélite bajo análisis estuviera presente. Como puede verse, el resultado del proceso de adquisición es una matriz en cuyas celdas se almacenan los valores a la salida de la correlación. En aquellas celdas cuya hipótesis de retardo de código y error de frecuencia se acerque más al valor exacto que presenta la señal, se observa un pico que puede ser detectable fácilmente fijando un umbral previamente diseñado para garantizar una cierta probabilidad de falsa alarma constante (esto es, según el criterio de detección de Neyman-Pearson).

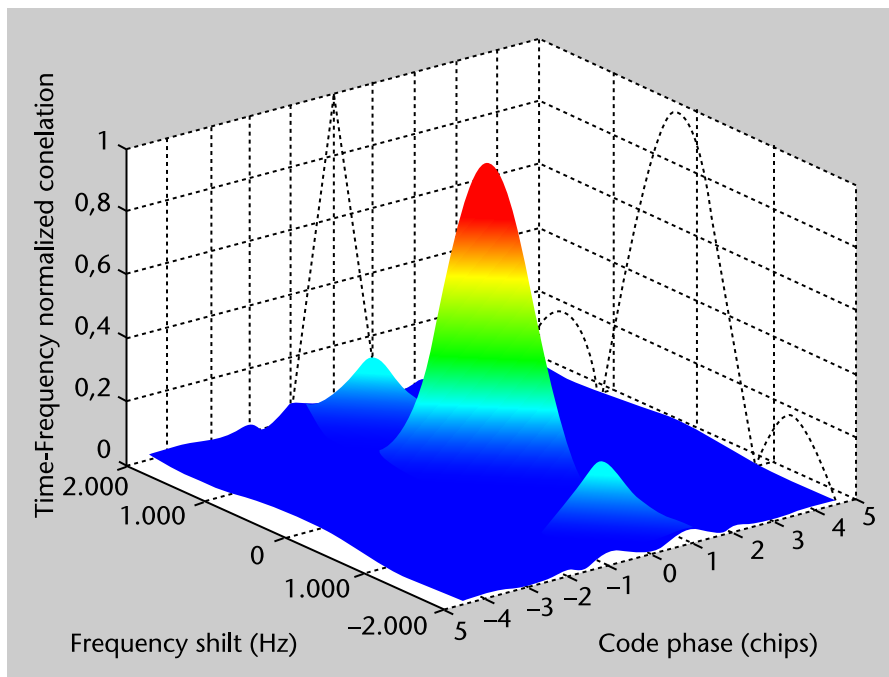
Ved también

El criterio de detección de Neyman-Pearson se estudia en el módulo "Sistemas radar" de esta asignatura.

Señal detectable

Para las celdas cuya hipótesis de retardo de código y error de frecuencia se acerque al valor exacto que presenta la señal, se observa un pico detectable fácilmente siempre y cuando el escenario de operación del receptor sea el previsto en el diseño del sistema GPS, es decir, principalmente en entornos exteriores. En interiores o en escenarios con dificultades de visión directa, el pico de correlación puede sufrir una gran atenuación que lo haga prácticamente imperceptible. En ese caso, es necesario aplicar técnicas de alta sensibilidad como las que se presentarán en el apartado 6.

Figura 10. Ejemplo de la respuesta en tiempo-frecuencia



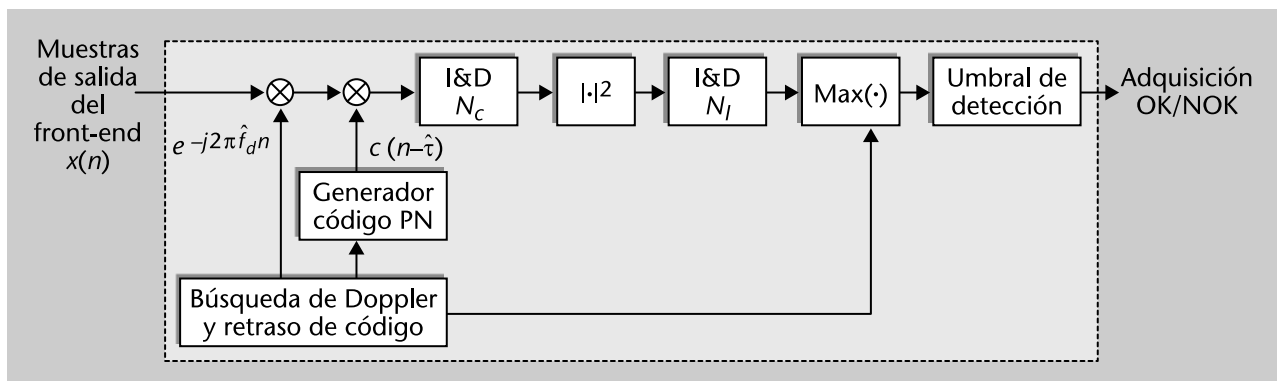
Por lo que respecta a la arquitectura interna de la etapa de adquisición de código, un ejemplo de esta se ilustra en la figura 11. En esta figura puede observarse cómo las muestras discretas procedentes del front-end son corregidas en frecuencia según el valor tentativo \hat{f}_d y posteriormente se lleva a cabo la correlación con el código PRN del satélite bajo análisis. Esta correlación puede entenderse también como un proceso de *despreading* o desensanchado, puesto que lo que se hace es eliminar la modulación de espectro ensanchado con la que el satélite ha encapsulado el mensaje de información que envía a la Tierra. Esta correlación se muestra en la figura 11 como una multiplicación por el código PRN seguida de una integración coherente de N_{code} muestras, que es precisamente la longitud del código PRN en muestras. La integración se lleva

a cabo en el bloque indicado como "I&D", nombre que responde a las siglas en inglés de *integrate and dump*, puesto que se realiza una integración y una vez finalizada la misma, es cuando de vuelca su valor a la salida y resetea la acumulación.

En definitiva, una de las operaciones básicas en el proceso para detectar la señal del satélite es la acumulación o integración de muestras, y el mínimo número de muestras que puede integrarse de manera coherente (es decir, sin utilizar no-linealidades como el módulo) es de N_{scode} muestras, que se corresponde con el tiempo de un código PRN, que para el caso de GPS-L1 es $T_{code}=1$ ms.

El diagrama de bloques de la figura 11, sin embargo, contempla la posibilidad de que puedan llevarse a cabo integraciones más largas, acumulando de manera coherente N_c salidas de correlación. En ese caso el tiempo total de integración sería de $N_c T_{code}$ segundos (o ms, según las unidades de T_{code}). Es posible extender aún más el tiempo de integración en caso de que la señal recibida llegue muy atenuada y sea necesario aumentar la integración para conseguir reducir el efecto del ruido y hacer así visible el pico de la correlación. En ese caso, la correlación se extiende de manera no-coherente (utilizando el módulo al cuadrado, habitualmente) para evitar que el error Doppler residual se vaya acumulando y acabe plegando la fase. Si ello sucediera, la integración total, lejos de ir aumentando progresivamente, iría disminuyendo, lo cual imposibilitaría una detección fiable de la señal del satélite. Los receptores que incorporan este tipo de correlación *extendida* mediante el uso de integraciones no-coherentes se conocen como receptores GPS de *alta sensibilidad* y su diagrama de bloques se muestra en la figura 11.

Figura 11. Arquitectura interna de la etapa de adquisición de código de un receptor GPS de alta sensibilidad, el cual incorpora integraciones no-coherentes para aumentar el tiempo total de correlación



El encargado de realizar la integración no-coherente es el bloque de I&D de N_l muestras que aparece tras el bloque de elevar al cuadrado en la figura 11. Nótese que para el caso de receptores GPS tradicionales, en donde se asumen condiciones de trabajo ideales (entornos exteriores, habitualmente), no es ne-

cesario realizar integraciones no-coherentes y simplemente el mismo diagrama de bloques de la figura 11 sigue siendo válido, con la excepción de que se fija $N_f=1$.

Es importante mencionar que el proceso de correlación y acumulación (coherente o no-coherente) se realiza para todos y cada uno de los valores tentativos de retraso de código y frecuencia Doppler que forman parte de la búsqueda bidimensional anteriormente mencionada.

Los resultados de las acumulaciones suelen ser guardados en una matriz tiempo-frecuencia y cuando todos los valores ya han sido almacenados, se procede a seleccionar el valor máximo de todos ellos. A continuación, ese valor acumulado máximo se compara con un umbral para determinar si el satélite bajo análisis está realmente presente o no. Si este proceso de detección resulta ser negativo, se considera que el satélite no está presente y se vuelve a iniciar todo el proceso de adquisición para buscar otro satélite diferente. Si, por el contrario, el proceso de detección es positivo y el valor acumulado supera el umbral, se determina que el satélite está presente. La probabilidad de que esta afirmación sea falsa (es decir, que el satélite realmente no esté presente aunque el valor acumulado supere el umbral) viene dada por la probabilidad de falsa alarma bajo la que ha sido fijado el umbral de detección, tal y como ya se mencionó en los fundamentos de teoría de detección.

En cualquier caso, una detección de señal positiva nos permite asumir que el satélite está presente y por tanto, dedicar mayor atención a afinar con precisión los parámetros de sincronización de la señal recibida. Ello es necesario puesto que durante la fase de adquisición, los valores de retraso de código τ y frecuencia Doppler f_d que se han evaluado son simplemente valores tentativos, resultantes de discretizar el eje de tiempo y el eje de frecuencia con una cierta granularidad.

Por ejemplo, la búsqueda en frecuencia suele hacerse en pasos de 500 Hz para el caso de utilizar correlaciones coherentes de 1 ms^{10} y la búsqueda en retraso de código en pasos de un chip, un tiempo que en distancia equivaldría a unos 300 metros. Nótese que esta granularidad es suficientemente grande como para que el espacio de búsqueda se divida en un número razonable de hipótesis a testear, y por lo tanto, que el tiempo total de búsqueda no sea demasiado elevado. Sin embargo, esta discretización no aporta toda la precisión que un receptor de GPS necesita. Hay que tener en cuenta que el objetivo último de un receptor GPS es determinar la posición del usuario con la mayor precisión posible, partiendo de medidas de tiempo de retardo que obtiene de los satélites visibles. Estas medidas de tiempo de retardo han de ser convertidas posteriormente en distancias, para lo cual son multiplicadas por $\beta \text{ m/s}$, la velocidad de la luz, el cual es un número ciertamente grande. Por lo tanto, cualquier error o imprecisión en las medidas de tiempo de retardo queda enormemente am-

Ved también

Los fundamentos de la teoría de detección se exponen en el módulo "Sistemas radar" de esta asignatura.

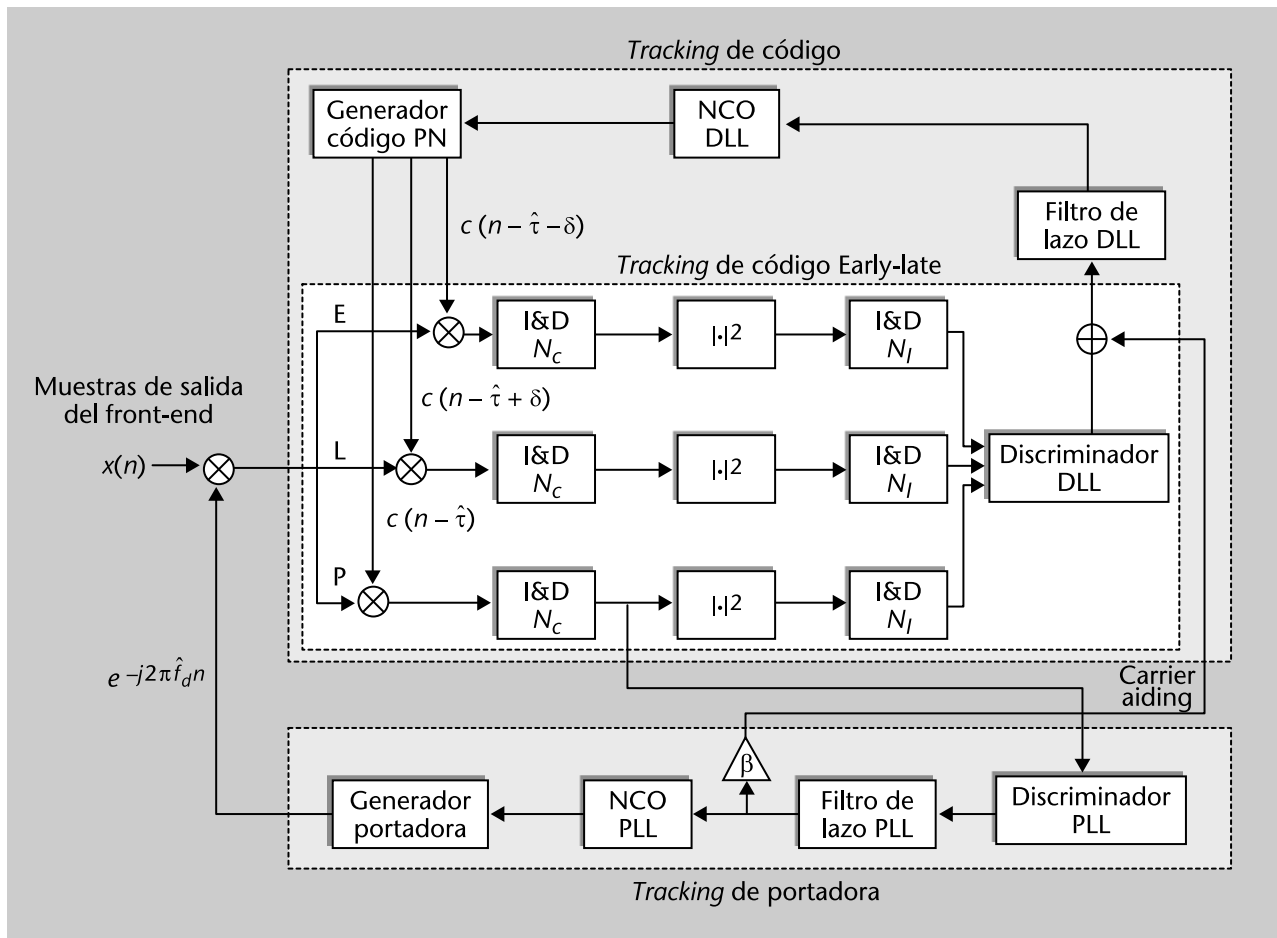
⁽¹⁰⁾Véase la sección 7.4 de la obra de J. B.-Y. Tsui (2004).

plificado cuando es convertido a distancia. Es por ello por lo que los valores tentativos de retraso de código (así como de frecuencia Doppler) obtenidos en el proceso de adquisición han de ser refinados y monitorizados para poder seguir cualquier pequeña variación que puedan sufrir a lo largo del tiempo. Este proceso de *seguimiento* de precisión es el que se conoce como *tracking* y será objeto de estudio en el subapartado siguiente.

4.2. Etapa de *tracking*

Tal y como se acaba de mencionar en el subapartado anterior, los valores tentativos de retraso de código y frecuencia Doppler obtenidos durante la fase de adquisición han de ser refinados y seguidos a lo largo del tiempo para monitorizar cualquier pequeña variación que puedan sufrir. Dicha tarea es la misión de la etapa de *tracking*, la cual consta principalmente de dos arquitecturas en paralelo: una dedicada al seguimiento del retraso de código, conocida como *delay lock loop* (DLL), y otra dedicada al seguimiento de la frecuencia, conocida como *phase lock loop* (PLL). Ambas arquitecturas se muestran en la figura 12.

Figura 12. Arquitectura interna de la etapa de *tracking*



Tanto el DLL como el PLL son arquitecturas en lazo cerrado las cuales obtienen una estimación precisa del parámetro de interés a base de comparar la señal de entrada con una réplica local. La señal de error resultante se utiliza para corregir las estimaciones de los parámetros que se han usado para generar la réplica, hasta que se consigue una alineación con la entrada lo más precisa posible (esto es, haciendo que la señal de error resultante sea lo más cercana a cero).

Para el caso del DLL, la réplica local es simplemente la salida de un generador de código PRN que está controlado por un NCO¹¹, el cual no deja de ser más que la versión discreta de un tradicional oscilador por tensión o VCO¹².

⁽¹¹⁾NCO son las siglas de *numerically controlled oscillator*.

⁽¹²⁾VCO son las siglas de *voltage controlled oscillator*.

Para el caso del PLL, la réplica local es la salida de un generador de portadora, el cual ofrece a su salida la componente en fase y la componente en cuadratura de una exponencial compleja según la fase indicada por un NCO. En ambos casos, la réplica local se correla con la señal de entrada y el resultado se introduce a un discriminador, a cuya salida se obtiene una señal proporcional (en media) al error del parámetro que se está siguiendo. Tal y como se muestra en la figura 12, el *tracking* del código necesita tres correladores: uno conocido como *prompt* y centrado en el retraso de interés, y dos correladores más situados simétricamente antes y después del *prompt*, y conocidos por tanto como correladores *early* y *late*. Esta configuración forma parte de la familia de métodos basados en la filosofía de *early-late* (Kaplan; Hegarty, 1996, cap. 5). La salida del discriminador es posteriormente filtrada con el objetivo de ser suavizada, para lo cual se utiliza un filtro de lazo a cuya salida se alimenta al NCO, cerrando así el lazo.

En la figura 12 puede observarse que hay una conexión entre el módulo de *tracking* de portadora y el módulo de *tracking* de código. Esta conexión es lo que se conoce como *carrier aiding* y, mediante ella, el PLL ayuda al DLL proporcionándole una señal de referencia. Dicha señal de referencia incorpora las variaciones temporales que experimenta la señal recibida como consecuencia de la dinámica del usuario y se trata, además, de una señal de gran precisión, puesto que el PLL trabaja con un ruido de fase órdenes de magnitud inferior al que sufre el DLL. Es por ello que, desde el punto de vista del DLL, es interesante poder aprovechar esta información a priori que aporta el *carrier aiding* para, así, reducir el ancho de banda del lazo, conseguir un mayor rechazo al ruido y centrarse en seguir las variaciones de código que surgen debido a la dinámica de la ionosfera (Kaplan; Hegarty, 1996, sección 5.2.4).

Antes de poder utilizar la señal de *carrier aiding*, es importante aplicarle un factor de escala para tener en cuenta el diferente efecto que tiene la frecuencia Doppler sobre la portadora y sobre el código. Es decir, el factor β que aparece en la figura 12 viene dado por $\beta = R_c / F_L$, en donde R_c es la frecuencia de chip y F_L es la frecuencia de portadora. Como puede deducirse, la relación entre uno

y otro término es de varios órdenes de magnitud, y de ahí la importancia de aplicar correctamente este factor de escala antes de introducir la información de portadora en el DLL.

Tras esta breve introducción a los elementos que forman la etapa de *tracking*, se está ya en posición de adentrarse en los fundamentos y principios de operación de cada uno de ellos, lo cual os aportará una visión más detallada acerca del funcionamiento global de un receptor GNSS. En primer lugar, comenzaremos por describir los diferentes tipos de discriminadores que pueden utilizarse tanto en el DLL (subapartado 4.2.1) como en el PLL (subapartado 4.2.2). A continuación, en el subapartado 4.2.3, se presentará la estructura de los filtros de lazo y la motivación de este tipo de elementos dentro de una arquitectura de *tracking*. Finalmente, en el subapartado 4.2.4, se realizará un breve repaso de las propiedades de estabilidad de los sistemas en lazo cerrado y se proporcionarán unas sencillas reglas para la caracterización de este tipo de sistemas.

4.2.1. Discriminadores de DLL

El discriminador es una pieza clave de todo esquema de sincronización en lazo cerrado, puesto que es el encargado de generar la señal de error que sirve como referencia para aplicar las correcciones necesarias y alinear así la réplica local con la señal de entrada.

En el caso del DLL, el discriminador se encarga de proporcionar a su salida un valor proporcional al error de retardo de código τ que presenta la señal de entrada respecto al código PN local. Para el caso del código C/A de GPS-L1, y suponiendo por simplificar, que la portadora está perfectamente adquirida, la salida del correlador *prompt* puede expresarse según:

$$y_p(\tau) = \begin{cases} 1 - \frac{|\tau|}{T_c}, & |\tau| < T_c \\ 0, & |\tau| \geq T_c \end{cases} \quad (7)$$

con T_c el tiempo de chip. Para el caso de los otros dos correladores que se muestran en la figura 12, los correladores *early* y *late*, sus salidas vienen dadas respectivamente por

$$y_E(\tau) \doteq y_p\left(\tau + \frac{\Delta}{2}\right), \quad (8)$$

$$y_L(\tau) \doteq y_p\left(\tau - \frac{\Delta}{2}\right) \quad (9)$$

en donde la separación Δ entre *early-late* suele ser habitualmente igual a T_c . A partir de las salidas de estos tres correladores se definen diferentes tipos de discriminadores, pero principalmente pueden dividirse en dos categorías: discriminadores coherentes y discriminadores no-coherentes.

El **discriminador coherente** más habitual es el conocido como *early-minus-late* (E-L), el cual como su nombre indica, consiste en tomar como señal de error la resta entre el correlador *early* y el correlador *late*. Si denotamos la salida del discriminador de código como $\Delta\tau(\tau)$, para el caso E-L coherente tenemos:

$$\Delta\tau_{E-L,c}(\tau) = \frac{1}{2} [y_E(\tau) - y_L(\tau)]. \quad (10)$$

Los **discriminadores no-coherentes** son utilizados cuando la hipótesis que se ha supuesto anteriormente, acerca de que la portadora está perfectamente adquirida, no es cierta. En ese caso, la salida del correlador se ve afectada por varios efectos: una atenuación α debida a la integración de la correlación en presencia de errores residuales de frecuencia, y una rotación compleja θ_ε resultado de la integración de muestras complejas. Ambos efectos se indican a continuación en un modelo de señal más realista que el utilizado en la expresión 7 para la salida del correlador *prompt*:

$$y_p(\tau) = \begin{cases} \alpha \left(1 - \frac{|\tau|}{T_c}\right) e^{j\theta_\varepsilon}, & |\tau| < T_c \\ 0, & |\tau| \geq T_c. \end{cases} \quad (11)$$

La atenuación que sufre la señal a la salida del correlador es ya irreparable. Es por ello que su impacto debe ser evaluado previamente, ajustando de forma adecuada el tiempo de integración del correlador, a fin de evitar que incluso la señal a la salida pueda llegar a cancelarse para algunos valores de error residual de frecuencia. Por lo que respecta al término de fase, este puede neutralizarse trabajando con la envolvente de la señal en vez de con su amplitud compleja. Este es el caso, por ejemplo, del discriminador *no-coherente early-minus-late*, cuya salida proporciona la señal de error:

$$\Delta\tau_{E-L,nc}(\tau) = \frac{1}{2} [|y_E(\tau)|^2 - |y_L(\tau)|^2]. \quad (12)$$

Una variante de este discriminador conocida como *early-minus-late de envolvente normalizada*, incluye un término de normalización para reducir la sensibilidad del discriminador a cambios en amplitud de la señal de entrada,

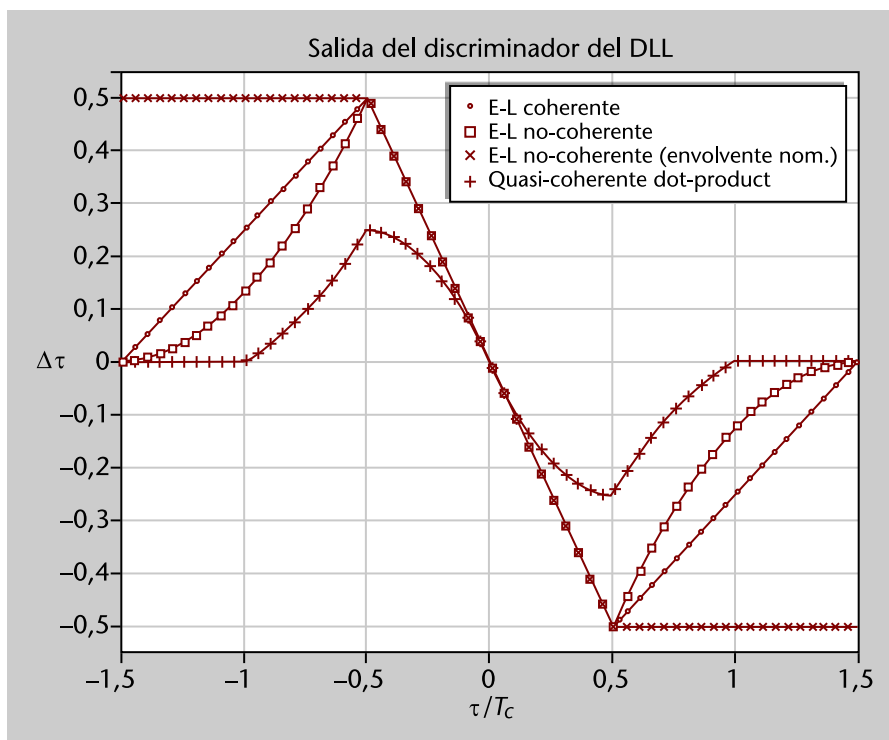
$$\Delta\tau_{E-L,env}(\tau) = \frac{1}{2} \frac{|y_E(\tau)|^2 - |y_L(\tau)|^2}{|y_E(\tau)|^2 + |y_L(\tau)|^2}. \quad (13)$$

Finalmente, otra de las opciones utilizadas es el conocido como discriminador *cuasi-coherente*, denominado así debido a que evita utilizar la no-linealidad de elevar al cuadrado la salida del correlador. En lugar de eso, utiliza productos cruzados entre los tres correladores,

$$\Delta\tau_{E-L,qcoh}(\tau) = \frac{1}{2} [\text{Re}[y_E(\tau) - y_L(\tau)]\text{Re}[y_P(\tau)] + \text{Im}[y_E(\tau) - y_L(\tau)]\text{Im}[y_P(\tau)]] \quad (14)$$

La respuesta de los cuatro discriminadores de código que se acaban de presentar se muestra en la figura 13.

Figura 13. Salida de diferentes discriminadores de DLL en función del error de retraso de código τ entre la señal de entrada y la réplica local. Representación que es conocida como *curva S*.



Según se observa en la figura 13, la principal característica de los discriminadores que se han presentado es que proporcionan un comportamiento lineal para valores pequeños del error de retraso de código τ (típicamente, para valores de $\tau/T_c \in [-0,5, 0,5]$). Aunque discriminadores como el de tipo E-L de envolvente normalizada poseen un margen de trabajo más pequeño, comprendido apenas entre $\tau/T_c \in [-0,25, 0,25]$.

Nótese que cuanto mayor sea el margen de visibilidad, más robusto es el discriminador en lo que a seguir posibles cambios de τ se refiere. Si el margen es pequeño, un cambio significativo en τ puede hacer que el valor de salida quede fuera de la zona de trabajo lineal y que el discriminador entre en una zona de falso equilibrio, en donde la salida del discriminador puede llegar a ser cero aunque el error de entrada no lo sea. Este es el caso de la región $|\tau|/T_c > 0,5$ o $|\tau|/T_c > 0,25$ para el caso del discriminador cuasi-coherente. Ante esta posible ambigüedad, se hacen necesarios mecanismos que se encarguen de confirmar

Lectura recomendada

Sobre los detectores de *lock* podéis ver la sección 5.11 de la obra de Kaplan y Hegarty (1996).

si la salida del discriminador realmente vale cero porque el error que se persigue se ha llegado a compensar por completo, o si por el contrario, se ha entrado en una zona de falso equilibrio. Dichos mecanismos se conocen como *detectores de lock*.

De todas formas, es importante remarcar que reducir la zona lineal del discriminador haciendo que Δ sea pequeño tiene como resultado que el discriminador se vuelva más sensible a variaciones en el retardo de código τ a la entrada. Este aumento de la sensibilidad hace que el discriminador se vuelva más preciso y que, por tanto, reduzca la varianza en la señal de corrección que ofrece a su salida.

Aunque la reducción de Δ lleva consigo también una reducción del margen de visibilidad del discriminador, esto habitualmente no es demasiado problema, ya que las variaciones del error de retardo de código τ debidas a la dinámica del usuario (es decir, aquellas variaciones que realmente pueden ocasionar que el discriminador se salga fuera de su zona lineal) son compensadas a través de la señal de carrier aiding que asiste al módulo de *tracking* de código. Esta práctica es habitual en la mayoría de receptores actuales y permite operar al discriminador de código con una separación más pequeña, la cual no solo aporta una menor varianza a la salida del discriminador sino que también lo protege frente al multipath.

Para el caso del discriminador *early-late no-coherente*, que es el más utilizado, la varianza que ofrece a su salida puede expresarse según Gernot (2009),

$$\sigma_{\tau,E-L,nc}^2 \approx \frac{1}{\frac{c}{N_0} T} \left(\Delta + \frac{\Delta}{1-\Delta} \frac{1}{2 \frac{c}{N_0} T} \right) \quad (15)$$

en donde T es el tiempo de integración del correlador. A partir de la expresión 15 se puede apreciar cómo una disminución en Δ tiene como consecuencia directa una disminución en la varianza del discriminador. Sin embargo, la mejora que aporta reducir progresivamente la separación Δ tiene un límite, el cual viene dado por el ancho de banda del front-end. En ese caso, la reducción de la separación Δ , especialmente para valores $\Delta < \frac{1}{2}$, ha de ir acompañada de un aumento del ancho de banda del front-end. De lo contrario, llegaría un momento en que los correladores *early-late* acabarían trabajando en puntos que han quedado suavizados debido al truncado que realiza el front-end sobre el ancho de banda de la señal recibida. En estos puntos se ha perdido la precisión que posee la función de autocorrelación ideal (la cual presenta un pico abrupto) y por tanto, ya no es posible mejorar la precisión del discriminador por mucho que se acerquen los correladores *early-late*.

front-ends y anchos de banda

Recuérdese que, por ejemplo, el ancho de banda de la señal GPS emitida por los satélites es del orden de 20 MHz, mientras que los receptores mass-market suelen tener *front-ends* con anchos de banda del orden de tan solo 2 MHz.

Lecturas recomendadas

Por lo que respecta al resto de discriminadores y sus prestaciones, encontraréis un análisis detallado, por ejemplo, en la sección 5.6 de la obra de Kaplan y Hegarty (1996), en la sección IV.2 de la obra de Simon y otros (2002), o en los capítulos 7 y 8 en las obras de Parkinson y Spilker (1996) o de Betz y Kolodziejewski (2009), así como en otras referencias de la literatura en sistemas de espectro ensanchado.

4.2.2. Discriminadores de PLL

El seguimiento de la portadora se realiza, según muestra la figura 12, tomando la salida del correlador *prompt*. El primer paso es introducir esta señal al discriminador del PLL para obtener una señal de error que proporcione información acerca de la corrección que es necesaria aplicar sobre la fase del generador de portadora local. Los dos discriminadores de fase más habituales son el basado en la operación de arcotangente y el basado en la componente en cuadratura del correlador *prompt*. La salida del primero puede expresarse según,

$$\Delta\theta_{\text{atan}}(\theta_\epsilon) = \text{atan2}\left(\frac{\text{Im}[y_p(\hat{\tau})]}{\text{Re}[y_p(\hat{\tau})]}\right) \quad (16)$$

en donde atan2 representa la operación de arcotangente sobre los cuatro cuadrantes y $\hat{\tau}$ representa el valor correcto de retraso de código, por lo que se supone que el bloque de DLL ha alcanzado el equilibrio. En la práctica, ambos bloques (el DLL y PLL) suelen operar en paralelo, pero, para simplificar el problema, supondremos primero que el DLL ya ha convergido mientras ilustremos el funcionamiento del PLL. Por lo que respecta al discriminador del PLL basado en la componente en cuadratura, su salida puede expresarse según:

$$\Delta\theta_Q(\theta_\epsilon) = \frac{\text{Im}[y_p(\hat{\tau})]}{E_T[|y_p(\tau)|]} \quad (17)$$

en donde $E_T[|y_p(\tau)|]$ representa el promedio temporal durante un tiempo de observación T , de la envolvente del correlador *prompt*, y sirve de normalización para hacer al discriminador insensible frente a variaciones en la amplitud de la señal recibida.

Aunque hasta ahora se ha hecho mención a los discriminadores de PLL, esta nomenclatura en realidad, corresponde solo a uno de los tres tipos de discriminadores que existen. Los otros dos corresponden al **discriminador de Costas** y al **discriminador de frecuencia** o FLL. El primero es muy utilizado en la práctica puesto que posee la propiedad de ser robusto a los cambios de fase que presenta la señal debido a la modulación de los bits del mensaje de navegación. Estos saltos de fase son inevitables en la mayoría de casos, a no ser que

se tenga acceso al mensaje de navegación (Kaplan; Hegarty, 1996, p. 166), lo cual no suele ser habitual en la práctica. Existen diversos discriminadores de fase de tipo Costas, siendo los más habituales los siguientes:

$$\Delta\theta_{\text{Costas,I}\times\text{Q}}(\theta_e)=\text{Re}[y_p(\hat{t})]\text{Im}[y_p(\hat{t})] \propto \sin 2\theta_e \quad (18)$$

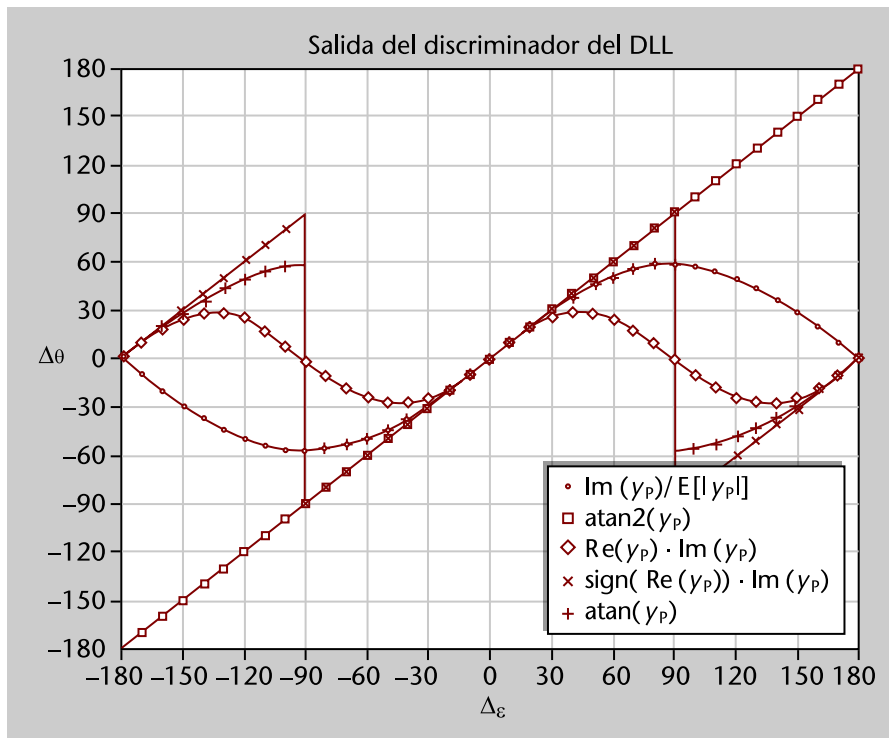
$$\Delta\theta_{\text{Costas,sign(I)}\times\text{Q}}(\theta_e)=\text{sign}(\text{Re}[y_p(\hat{t})])\text{Im}[y_p(\hat{t})] \propto \sin\theta_e \quad (19)$$

$$\Delta\theta_{\text{Costas,atan}}(\theta_e)=\text{atan}\left(\frac{\text{Im}[y_p(\hat{t})]}{\text{Re}[y_p(\hat{t})]}\right) \propto \theta_e. \quad (20)$$

A priori, los dos primeros discriminadores parecen presentar una relación no lineal entre su salida y el error de fase θ_e . Sin embargo, y suponiendo régimen estacionario en donde el error de fase es ya pequeño, puede utilizarse la relación $\sin(x) \stackrel{x \ll 1}{\approx} x$, dejando patente la relación lineal entre la salida de esos discriminadores y el valor de interés, θ_e , al menos para valores pequeños de este.

La respuesta de los discriminadores aquí mencionados puede observarse en la figura 14 para el margen de valores de error de fase comprendido entre $[-180,180)$ grados. Como puede apreciarse, el discriminador basado en la función atan2 es el único que proporciona un comportamiento lineal para todos los valores de fase de entrada. El segundo mayor comportamiento lineal lo proporciona el discriminador basado en la función atan , el cual se limita al margen entre $[-90,90)$ ya que esta función queda definida únicamente en los dos cuadrantes del semiplano positivo. Cabe recordar, sin embargo, que pese a ofrecer un buen comportamiento lineal, estas dos funciones son las que implican un mayor nivel de complejidad. Es por ello por lo que el resto de discriminadores presentados, los basados en operaciones con la parte real y la parte imaginaria del correlador *prompt*, ofrecen una solución de baja complejidad bajo un compromiso de reducción en el margen lineal a su salida.

Figura 14. Salida de diferentes discriminadores de PLL en función del error de fase entre la señal de entrada y la réplica local



En cuanto a las prestaciones del discriminador de PLL, la varianza en lazo abierto que ofrece cuando a su entrada hay únicamente ruido, puede expresarse para el caso de un discriminador basado en atan según Kaplan y Hegarty (1996, cap. 5).

$$\sigma_{\theta, \text{atan}}^2 = \frac{1}{\frac{C}{N_0}} \left(1 + \frac{1}{2T \frac{C}{N_0}} \right). \quad (21)$$

Por último, el discriminador FLL ofrece a su salida una señal proporcional al error de frecuencia entre la señal de entrada y la réplica local. Nótese que el discriminador de FLL monitoriza la frecuencia, pero no la fase, y aunque el valor de fase puede obtenerse artificialmente integrando las correcciones de frecuencia, existe siempre una ambigüedad que será necesario compensar en posteriores etapas antes de proceder a la demodulación del mensaje de navegación.

Lecturas recomendadas

Aunque la adopción de discriminadores FLL no es tan habitual como el caso del PLL, si estás interesado encontrarás información detallada en la sección 5.6 de la obra de Kaplan y Hegarty (1996) y en la obra de Juang y Chen (2009).

4.2.3. Filtros de lazo

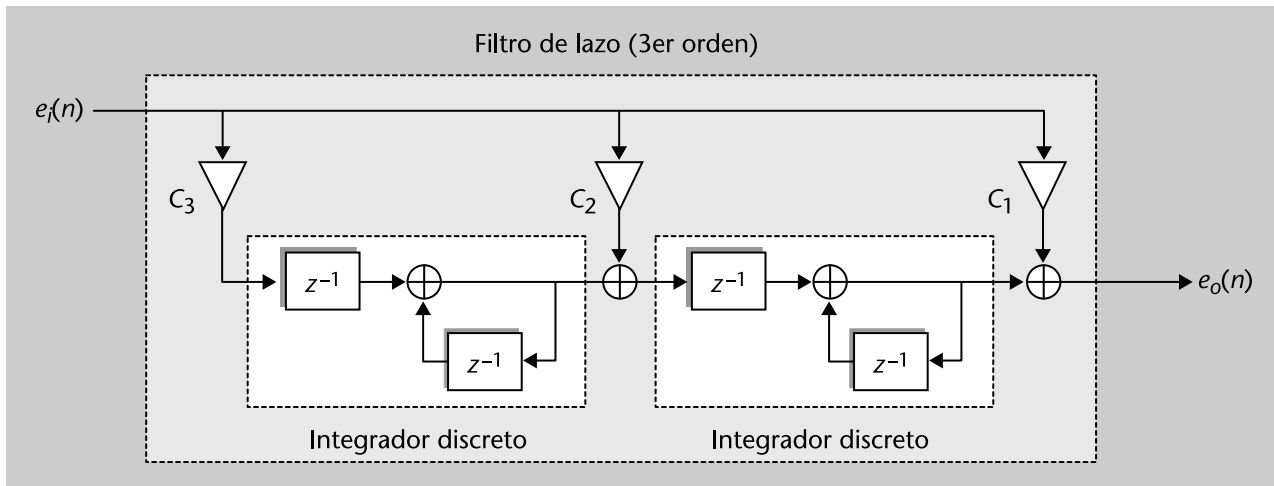
En los subpartados anteriores se ha comentado que la salida del discriminador ofrece una señal de referencia que resulta ser proporcional al error cometido entre la señal recibida y la réplica local que está siendo generada en el

receptor. Esta señal de referencia es de gran importancia, puesto que permite al sistema implementar las correcciones necesarias para generar la réplica local con el mayor parecido posible (en términos de retraso de código y fase de portadora) a la señal que está siendo recibida y, así, lograr que ambas señales estén sincronizadas.

En la práctica, sin embargo, hay que tener en cuenta que la señal recibida estará afectada por ruido térmico y, por tanto, también lo estará la salida del discriminador. Es por ello por lo que en la práctica no es recomendable conectar directamente la salida del discriminador al NCO, ya que, en ese caso, la generación de señal absorbería toda la inestabilidad y variabilidad del ruido de entrada, enmascarando la corrección que realmente hay que aplicar en cada caso. Para evitar este problema, la mayoría de sistemas de *tracking* o sincronización en lazo cerrado implementan un filtrado de la salida del discriminador con el objetivo de reducir la variabilidad del ruido y proporcionar una señal suave con que atacar al NCO. Este bloque de filtrado es precisamente el que se conoce como *filtro de lazo* y aparece en la figura 12, tanto en el módulo de *tracking* de código como en el de *tracking* de portadora.

En el fondo, un filtro de lazo no es más que un conjunto de factores de escala e integradores que se encargan de suavizar la señal de entrada. La estructura de este filtro se muestra en la figura 15 para el caso de orden 3, a partir de la cual se puede obtener la estructura para el caso de orden 2 (el más habitual) y orden 1, simplemente fijando las constantes $C_3=0$ y $\{C_3, C_2\}=0$, respectivamente. A partir de la figura 15, puede observarse cómo el caso de un filtro de lazo de orden 1 equivale a aplicar únicamente un factor de escala a la señal de entrada, sin llevarse a cabo propiamente un filtrado de la misma. En caso de que la salida del discriminador fuera idealmente cero (es decir, en régimen estacionario, una vez ya se ha sincronizado la señal recibida con la réplica local) la salida del filtro de lazo también sería cero y el NCO se mantendría fijo, generando una señal constante. Por tanto, y en régimen estacionario, un esquema de *tracking* que incorpora un filtro de lazo de orden 1 únicamente es capaz de seguir señales constantes.

Figura 15. Diagrama de bloques de un filtro de lazo de tercer orden



Si la señal de entrada presenta una cierta dinámica, entonces es necesario aumentar el orden del lazo. Para el caso de orden 2, el filtro de lazo incorpora un integrador que va procesando la salida del discriminador escalada en un factor C_2 , tal y como se observa en la figura 15. Una vez que la salida del discriminador se hace cero al alcanzar el régimen estacionario, el filtro de lazo es capaz de ofrecer una salida constante gracias al valor acumulado en el registro retroalimentado del integrador. Ese valor constante a la salida del filtro de lazo hará mover de manera constante el NCO que viene a continuación en el esquema de *tracking*, con lo cual la réplica de señal generada experimentará una deriva constante. Por lo tanto, y en régimen estacionario, un esquema de *tracking* que incorpora un filtro de lazo de orden 2 es capaz de seguir señales cuya dinámica presenta una derivada constante. Por ejemplo, el caso de seguir la fase de una señal afectada por error de frecuencia constante. Este comportamiento puede generalizarse para el caso de orden 3, en donde el filtro de lazo en régimen estacionario es capaz de proporcionar a su salida una señal con deriva constante, y por lo tanto, el esquema de *tracking*, incluyendo el efecto de integración adicional del NCO, será capaz de seguir señales cuya dinámica presente una segunda derivada constante y diferente de cero.

Filtros de lazo de orden superior a tres

Los filtros de lazo de orden superior a tres no son habituales en la práctica y además, suelen presentar problemas de estabilidad. Es por ello por lo que no serán considerados en esta asignatura.

4.2.4. Caracterización en lazo cerrado

Una vez se ha presentado de manera intuitiva el funcionamiento y comportamiento de un filtro de lazo, el siguiente paso es implementarlo en la práctica. Para ello, sin embargo, es necesario tener en cuenta todos los bloques que forman el sistema en lazo cerrado. A partir de este análisis es posible hallar el valor de las constantes correspondientes $\{C_1, C_2, C_3\}$ que hacen que todo el lazo global de *tracking* sea estable y proporcione las prestaciones deseadas.

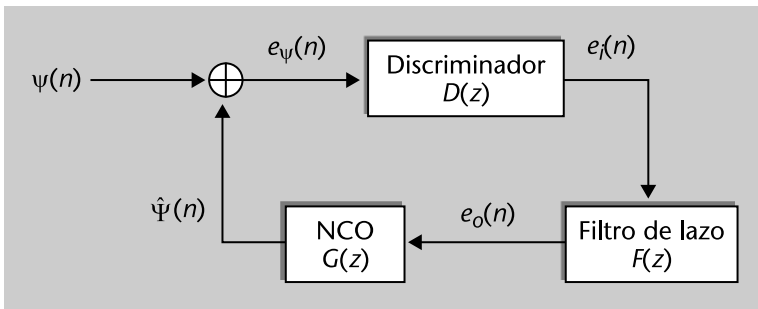
Lecturas recomendadas

Estudiar las prestaciones del sistema de manera detallada es un tema que requiere de una complejidad y de una extensión que lo hacen quedar fuera del alcance de la asignatura. Es por ello por lo que aquí nos centraremos en abordar el problema desde una visión pragmática, ofreciendo una serie de herramientas básicas y de propiedades para determinar si el sistema es estable o no y qué prestaciones ofrece. Los que estéis interesados en

un análisis más profundo encontraréis información detallada al respecto en la obra de Gardner (2005), en la de Mao y Chen (2008) para el caso de sistemas en tiempo discreto, o en las de Perdikaris (1996) y de Jury (1964) para un análisis más amplio en el caso de sistemas en tiempo continuo.

La metodología tradicional consiste en analizar el sistema mediante su función de transferencia en el dominio de la transformada Z. Ello implica combinar la función de transferencia del filtro de lazo junto con la del resto de bloques que forman el lazo global, el cual se representa de manera esquemática en la figura 16.

Figura 16. Representación general de un esquema de seguimiento en lazo cerrado



A partir del esquema de esta figura, el cual aplica tanto al caso de módulo de *tracking* de código como de portadora, hay varias relaciones de entrada-salida que es interesante definir.

En primer lugar, se define la *función de transferencia del sistema* como,

$$H(z) \doteq \frac{\hat{\Psi}(z)}{\Psi(z)} \quad (22)$$

en donde $\Psi(z) \doteq \text{TZ}\{\psi(n)\}$ es la transformada Z de la señal de entrada y $\hat{\Psi}(z) \doteq \text{TZ}\{\hat{\psi}(n)\}$ la transformada Z de la réplica generada por el lazo cerrado.

De manera similar, la *función de transferencia en lazo abierto* se define como,

$$U(z) \doteq \frac{\hat{\Psi}(z)}{E_\psi(z)} = D(z)F(z)G(z) \quad (23)$$

en donde $E_\psi(z) \doteq \text{TZ}\{e_\psi(n)\}$ es la transformada Z de la señal de error entre la entrada y la réplica, y las transformadas Z representadas por las funciones $D(z)$, $F(z)$ y $G(z)$ corresponden al discriminador, filtro de lazo y NCO, tal y como se indica en la figura 16.

Por lo que respecta al discriminador, es habitual suponer en primera instancia que este opera en su zona lineal, y por tanto, su función de transferencia viene dada únicamente por un factor de escala, $D(z) = \kappa_d$. En cuanto al filtro de lazo,

su función de transferencia puede obtenerse fácilmente a partir de analizar el esquema de la figura 15. Teniendo en cuenta que la función de transferencia de un integrador discreto viene dada por

$$H_{\text{int}}(z) = \frac{z^{-1}}{1 - z^{-1}}, \quad (24)$$

la función de transferencia del filtro de lazo resulta ser

$$F(z) = \frac{C_1(1 - z^{-1})^2 + C_2z^{-1}(1 - z^{-1}) + C_3z^{-2}}{(1 - z^{-1})^2}. \quad (25)$$

Finalmente, el NCO es un acumulador con aritmética modular, por lo que llegado a un cierto valor, se resetea y vuelve a comenzar a acumular desde 0. Este comportamiento es ciertamente no lineal, pero a efectos de calcular la función de transferencia, este se suele ignorar y se supone que el NCO opera linealmente. En ese caso, puede modelarse como un integrador discreto como los utilizados en el filtro de lazo, y por tanto su respuesta impulsional puede modelarse como $G(z) = \frac{z^{-1}}{1 - z^{-1}}$. Con estos tres ingredientes, la función de transferencia en lazo abierto resulta ser

$$U(z) = \kappa_d z^{-1} \frac{C_1(1 - z^{-1})^2 + C_2z^{-1}(1 - z^{-1}) + C_3z^{-2}}{(1 - z^{-1})^3}. \quad (26)$$

Mediante la respuesta en lazo abierto, la función de transferencia global del sistema se puede expresar según

$$H(z) = \frac{U(z)}{1 + U(z)} = \frac{\kappa_d z^{-1} [C_1(1 - z^{-1})^2 + C_2z^{-1}(1 - z^{-1}) + C_3z^{-2}]}{(1 - z^{-1})^3 + \kappa_d z^{-1} [C_1(1 - z^{-1})^2 + C_2z^{-1}(1 - z^{-1}) + C_3z^{-2}]}. \quad (27)$$

Para que el sistema sea estable, los coeficientes $\{C_1, C_2, C_3\}$ han de fijarse de manera que los polos de la función de transferencia $H(z)$, o lo que es lo mismo, los ceros del polinomio característico de $H(z)$, estén todos dentro del círculo unidad. Es por ello por lo que, aparte de que uno pueda dar ciertos valores iniciales¹³ a estos coeficientes en función del ancho de banda o el tiempo de respuesta del lazo que se desea, en última instancia, hay que verificar siempre que las siguientes condiciones de estabilidad se cumplen (Gardner, 2005, sección 4.3):

- **Orden 1** ($C_2 = C_3 = 0$). El sistema posee un único polo situado en $z = 1 - \kappa_d C_1$ por lo que para que sea estable se ha de cumplir que $C_1 < 2/\kappa_d$. Este resultado contrasta con el que se obtiene para el caso de sistemas en lazo cerrado en tiempo continuo, donde para orden 1 el sistema es siempre estable mientras $\kappa_d C_1 > 0$. Este hecho pone de relevancia las diferencias que, en casos

⁽¹³⁾Por ejemplo, la práctica habitual recomienda que C_2 sea alrededor de un orden de magnitud inferior a C_1 y a su vez, que C_3 también sea alrededor de un orden de magnitud inferior a C_2 .

particulares, pueden ocurrir entre las versiones en tiempo discreto y en tiempo continuo de un mismo sistema en lazo cerrado.

- **Orden 2** ($C_3=0$). En este caso el sistema posee un par de polos que están situados en $z=1-\frac{\kappa_d C_1}{2}\left(1\pm\sqrt{1-\frac{4C_2}{\kappa_d C_1^2}}\right)$. Teniendo en cuenta que habitualmente $\frac{C_2}{C_1} < 1$, el sistema es estable mientras se cumpla que

$$2C_1 - C_2 < \frac{4}{\kappa_d}. \quad (28)$$

Nótese que en caso de que $C_2=0$ y el sistema pase a ser de orden 1, la condición anterior se convierte en $C_1 < \frac{2}{\kappa_d}$, que es precisamente la misma expresión que la ya presentada en el punto anterior.

- **Orden 3**. A diferencia de los casos anteriores, los sistemas de orden 3 resultan ser condicionalmente inestables, lo cual indica que para valores muy pequeños de $\kappa_d C_1$, el sistema puede dejar de ser estable. La condición de estabilidad en este caso viene dada por el siguiente intervalo de valores para la ganancia del lazo:

$$\frac{4C_1 C_2}{(4C_1 - C_2)(2C_1 - C_2)^2} < \kappa_d < \frac{32C_1}{(4C_1 - \frac{C_2}{2})^2} \quad (29)$$

Estas condiciones de estabilidad han de hacerse compatibles, tal y como se comentaba anteriormente, con las prestaciones que se desea obtener del sistema. Para analizar qué prestaciones son estas y cómo se modelan, nos centraremos a continuación en el caso de un sistema de orden 2, puesto que son los sistemas más habituales en la práctica. Haciendo la analogía con el caso de sistemas en lazo cerrado en tiempo continuo, los dos parámetros que caracterizan las prestaciones de un sistema de segundo orden son la frecuencia natural ω_n y el factor de amortiguamiento ζ , los cuales pueden expresarse en tiempo discreto como

$$\omega_n = \sqrt{\kappa_d C_1 C_2} \quad (30)$$

$$\zeta = \frac{1}{2} \sqrt{\frac{\kappa_d}{C_1 C_2}} \quad (31)$$

A nivel intuitivo, ω_n está relacionada con el ancho de banda o equivalentemente, el tiempo de respuesta del sistema, mientras que ζ está relacionado con el tipo de respuesta que presenta el sistema ante una excitación en forma de escalón.

En particular, para valores de $\zeta > 1$ se dice que el sistema está **sobre-amortiguado**, presentando una respuesta muy lenta al escalón. Para $\zeta=1$ se dice que el sistema está en **amortiguamiento crítico**, ofreciendo la respuesta más rápida posible. Finalmente, para $0 < \zeta < 1$ el sistema está **subamortiguado** y presenta una respuesta al escalón muy rápida, excediendo su amplitud y presentando oscilaciones pronunciadas.

Finalmente, y a partir de la función de transferencia en la expresión 27, es posible calcular también el ancho de banda equivalente de ruido del sistema, B_n ,

$$B_n \doteq \frac{1}{2\pi} \int_{-\pi}^{\pi} S_n(e^{j\omega}) |H(e^{j\omega})|^2 d\omega \quad (32)$$

en donde $S_n(e^{j\omega})$ es la densidad espectral de ruido a la entrada del sistema, y $H(e^{j\omega}) = H(z)_{|z=e^{j\omega}}$. Para el caso de un sistema de orden 2, el ancho de banda discreto resulta ser (Gardner, 2005, p. 130):

$$B_n = \frac{\kappa_d}{4} \frac{1 + C_1 C_2 \left(\frac{1}{\kappa_d} - \frac{3}{2} + \frac{1}{2} C_1 C_2 \right)}{1 - \frac{\kappa_d}{2} + C_1 C_2 \left(-1 + \frac{\kappa_d}{4} (1 - C_1 C_2) \right)} \quad (33)$$

en donde puede observarse que la ganancia del discriminador en régimen lineal, κ_d juega un papel muy importante en el valor que toma finalmente el ancho de banda del sistema. Teniendo en cuenta que el producto $C_1 C_2$ es habitualmente pequeño y suponiendo $\kappa_d \ll 1$ para asegurar que el discriminador trabaja en la zona lineal (en principio es suficiente $\kappa_d < 0,2$), entonces el ancho de banda de ruido se puede aproximar por

$$B_n \approx \frac{\kappa_d + C_1 C_2}{4} \quad (34)$$

con lo cual se puede observar que fijar constantes $\{C_1, C_2\}$ con valores grandes tiende a aumentar el ancho de banda del sistema, y por tanto a hacerlo más sensible a cambios en la entrada, mientras que fijar constantes de valores pequeños tiende a reducir el ancho de banda. En caso de que $C_2=0$, el ancho de banda resulta $B_n \approx \kappa_d/4$, el cual coincide con el ancho de banda de un sistema de primer orden en tiempo continuo.

El ancho de banda del lazo es un parámetro clave en el diseño de sistemas de *tracking* y juega un papel muy importante en la definición de la ganancia que puede obtenerse en lazo cerrado, medida a través de la relación señal-a-ruido a la salida del lazo, SNR_L . Si bien en general, la definición de la relación señal-a-ruido no implica mayor dificultad que implementar un cociente entre la potencia de señal y la potencia de ruido, en el caso de un filtro de lazo la definición no es directa ya que propiamente, no hay ninguna señal dentro del

lazo. Esto es debido a que la entrada al filtro de lazo es en realidad el error entre la señal recibida y la réplica local, con lo que en régimen estacionario y dejando de lado el efecto del ruido térmico, es una entrada que tiende a ser cero.

Como consecuencia, la definición de la relación señal-a-ruido a la salida del lazo no es directa, y en la literatura se pueden encontrar diversas definiciones o interpretaciones. La más utilizada es aquella que interpreta el filtro de lazo de la misma manera como si fuera un filtro tradicional, con un cierto ancho de banda B_n y que ante una entrada de ruido espectralmente blanco con densidad espectral $N_0/2$, proporciona una potencia de ruido $P_n=N_0B_n$ con lo cual, la relación señal a ruido a la salida del lazo resulta ser

$$\text{SNR}_L \doteq \frac{C}{N_0B_n}. \quad (35)$$

A partir del valor de $\frac{C}{N_0}$ de entrada y del ancho de banda del lazo, es posible determinar el impacto del ruido sobre las prestaciones del módulo de *tracking*. En particular, el ruido de fase a la salida del lazo viene dado por $\sigma_\theta^2 = \frac{1}{\text{SNR}_L}$ el cual, traducido a varianza en distancia (medida en m^2), resulta

$$\sigma_{\theta \rightarrow \tau}^2 = \frac{1}{(2\pi F_L)^2 \text{SNR}_L}. \quad (36)$$

La expresión 36 representa el impacto del ruido en las medidas de distancia que son obtenidas a partir de la portadora (módulo de *tracking* de portadora). Esa expresión puede convertirse a medidas en distancia pero obtenidas a partir del código (módulo de *tracking* de código) teniendo en cuenta que, tal y como se ha mencionado anteriormente, los errores en código son F_L/R_c veces mayores que los que aparecen en portadora. Como consecuencia, el impacto del ruido en las medidas de distancia obtenidas a partir del código resulta

$$\sigma_\tau^2 = T_c \frac{1}{\text{SNR}_L}. \quad (37)$$

Tanto la ecuación 36 como la 37 son expresiones que, como primera aproximación, permiten conocer de antemano las prestaciones de los módulos de *tracking*. Como puede observarse, ambas dependen del ancho de banda del lazo, lo cual tiene un impacto directo en el ajuste de las constantes $\{C_1, C_2, C_3\}$ que se han descrito a lo largo de este subapartado. En la práctica, y a partir de unos requerimientos concretos en términos de errores en distancia (ya sea a nivel de portadora o a nivel de código), el proceso de diseño del módulo de *tracking* consiste en ajustar los valores de las constantes del lazo de manera que el sis-

tema sea estable, se satisfagan los requerimientos establecidos, y el sistema sea capaz de seguir la dinámica de usuario prevista, haciendo que el lazo tenga capacidad de reacción suficiente como para no perder el seguimiento del error de entrada. Es decir, un compromiso entre disponer de un ancho de banda suficientemente estrecho como para filtrar el máximo posible de ruido, pero suficientemente ancho como para permitir seguir la dinámica de la entrada. Todo ello es un proceso complejo que habitualmente se resuelve mediante procedimientos de prueba y error, realizando extensos estudios numéricos y simulaciones por ordenador para confirmar que el diseño de las constantes del lazo se ajusta al escenario de trabajo y a los requerimientos especificados.

5. Cálculo de la posición

En el apartado anterior hemos estudiado la obtención de las medidas de retardo y fase. Este apartado se dedica a la obtención de la estimación de la posición, que es el objetivo último en la mayoría de las aplicaciones GNSS. El punto de partida para el cálculo de la posición son los observables de código (o pseudorrangos) y de fase. La definición y caracterización de estos se trata en los subapartados 5.1 y 5.2. Los métodos que permiten pasar de los pseudorrangos a la estimación de las coordenadas se estudian en el subapartado 5.3, mientras que en el subapartado 5.4 se analizan las diferentes fuentes de error que intervienen en todo el proceso y que hacen que el posicionamiento no pueda ser perfecto. Las técnicas diferenciales, explicadas en detalle en el subapartado 5.5, son la base para obtener actualmente las mejores prestaciones de los sistemas GNSS y alcanzar precisiones centrimétricas.

5.1. Cálculo de los pseudorrangos

La medida básica realizada por un receptor GPS es el tiempo aparente de propagación de la señal desde el satélite hasta el receptor, definido como la diferencia entre el instante de recepción de la señal y el instante de transmisión de esta. El instante de recepción se obtiene simplemente leyendo el reloj que tiene el receptor. El instante de transmisión es más complicado de obtener. Más adelante explicaremos concretamente cómo se realiza esto, pero a los efectos prácticos de comprender el cálculo de la posición en GPS, lo más conveniente es pensar que el tiempo de transmisión va marcado en la señal. Evidentemente, tanto el instante de recepción como el de transmisión se refieren a un mismo punto en la señal, como, por ejemplo, el inicio de un determinado chip. En definitiva, se trata de calcular cuándo se transmitió y cuándo se recibe un determinado punto de la señal.

El producto de este tiempo de tránsito por la velocidad de la luz en el vacío proporciona el observable básico de GPS, que se denomina pseudorango o pseudodistancia.

El prefijo *pseudo* se añade porque este observable no representa una verdadera distancia, sino que está sesgado debido a que los instantes de transmisión y de recepción hacen referencia a escalas de tiempo diferentes.

Reflexión

Pensad qué pasaría si en una carrera el tiempo de salida de un corredor se lee de un reloj y el tiempo de llegada se lee de otro reloj. ¿Sería la diferencia entre ambas lecturas el tiempo que realmente ha tardado el corredor? ¿Qué relación habría entre la diferencia

de las lecturas, el tiempo verdadero que ha tardado el corredor y el retraso o adelanto que tienen los relojes con los que se han realizado las medidas?

Concretamente, hay tres escalas de tiempo involucradas en el cálculo del pseudorrango. Dos de ellas están determinadas por el reloj del receptor y el del satélite. La tercera es una escala de tiempos de referencia, llamada Tiempo GPS (GPST), que se genera en el segmento terreno como un promedio de los relojes de todos los satélites y de las estaciones terrenas. Vamos a denotar las tres escalas de tiempo con las variables t_u (reloj del receptor), t_s (reloj del satélite) y t (tiempo GPS), respectivamente.

Consideremos que una determinada transición de la señal llega al receptor en el instante t según la escala GPST y que el tiempo de tránsito o propagación verdadero de esta transición ha sido T . El instante de recepción medido según el reloj del receptor es $t_u(t)$, y el instante de transmisión (medido mediante el reloj de satélite) correspondiente es $t_s(t - \tau)$. Entonces el pseudorrango se calcula como

$$P(t) = c(t_u(t) - t_s(t - \tau)) \quad (38)$$

donde c es la velocidad de la luz en el vacío. Tanto t como T son desconocidos y deben estimarse.

Observación

Notad que el modelo que aquí estamos presentando es genérico y es válido para cualquier señal de GPS (de la banda L1, L2 o L5) y de hecho también para otro sistema GNSS.

La relación de las escalas de tiempos del receptor y del satélite con el tiempo GPST son las siguientes:

$$t_u(t) = t + \delta t_u(t), \quad (39)$$

$$t_s(t - \tau) = (t - \tau) + \delta t_s(t - \tau), \quad (40)$$

donde $\delta t_u(t)$ y $\delta t_s(t)$ son los sesgos de los relojes del receptor y del satélite, respectivamente, ambos medidos respecto a la escala de tiempos GPST, y que en general pueden variar con el tiempo.

Según nuestra notación, estos sesgos representan el adelanto de los dos relojes respecto a la escala de referencia del sistema. De hecho, el segmento terreno estima el sesgo del reloj del satélite (δt_s) y aproxima su evolución temporal mediante un polinomio. Los coeficientes de dicho polinomio se transmiten en el mensaje de navegación. Por lo tanto, se puede considerar que δt_s es una función conocida, pero lo que desconocemos por el momento es el instante $t - \tau$ en el que se tiene que evaluar.

Si introducimos las expresiones 39 y 40 en la 38, el pseudorrango se puede escribir como

$$\begin{aligned}
 P(t) &= c(t + \delta t_u(t) - (t - \tau + \delta t^s(t - \tau))) + \varepsilon_\rho(t) \\
 &= c\tau + c(\delta t_u(t) - \delta t^s(t - \tau)) + \varepsilon_\rho(t)
 \end{aligned}
 \quad (41)$$

donde $\varepsilon_\rho(t)$ es un término que engloba todo el resto de los fenómenos no incluidos explícitamente en la ecuación anterior. Estos fenómenos son tanto errores adicionales o como ruido de medida, o efectos no modelados, o incluso errores en el modelo.

El tiempo de tránsito multiplicado por la velocidad de la luz en el vacío admite la siguiente expresión:

$$c\tau = r(t, t - \tau) + I_\rho(t) + T_\rho(t), \quad (42)$$

donde $r(t, t - \tau)$ es la distancia geométrica (o distancia verdadera) entre la posición del usuario en el instante t y el satélite en el instante $t - \tau$; I_ρ y T_ρ representan los retardos debidos a la propagación de la señal a través de la ionosfera y de la troposfera. Ambos términos son positivos. Realmente estos retardos aparecen en la expresión 42 porque el tiempo de tránsito se multiplica por la velocidad de la luz en el vacío y, debido a la presencia de la atmósfera, esta no es exactamente la velocidad de propagación de las señales.

Finalmente, si introducimos la expresión 42 en la 41 y, por sencillez, obviamos la referencia explícita al instante de medida t , obtenemos la **ecuación fundamental para el posicionamiento con pseudorrangos**:

$$P = r + c(\delta t_u - \delta t^s) + I_\rho + T_\rho + \varepsilon_\rho \quad (43)$$

Idealmente nos habría gustado tener unos observables que fueran iguales a la distancia geométrica r . Pero debido a los principios intrínsecos de funcionamiento del sistema obtenemos la pseudodistancia P , que es una versión sesgada y ruidosa de la distancia verdadera. En los siguientes subapartados estudiaremos cómo se puede obtener realmente una estimación muy precisa de la posición y del tiempo a partir de estas medidas que pueden diferir notablemente de la distancia real.

Los receptores GPS calculan los pseudorrangos de todos los satélites visibles simultánea y típicamente a una frecuencia alrededor de 1 Hz (es decir, miden un conjunto de pseudorrangos una vez por segundo). Los receptores orientados a aplicaciones donde hay una alta dinámica toman medidas más frecuentemente (hasta llegar a los 50 Hz) para poder seguir más fielmente el movimiento de la plataforma a la que van unidos. En aplicaciones donde se monitoriza un fenómeno que cambia muy lentamente o si se quiere reducir la cantidad de datos a procesar, los receptores generan medidas a tasas inferiores a 1 Hz.

Ved también

Volveremos sobre el tema de los errores cuando analicemos las fuentes de error en el posicionamiento en el subapartado 5.4 de este módulo.

En la práctica, el cálculo de un pseudorrango se realiza utilizando la propia definición dada por la ecuación 38. La obtención de $t_u(t)$ es inmediata puesto que el receptor solo ha de leer un registro interno que mantiene su propia escala de tiempo. Por otro lado, el cómputo de $t^u(t - T)$ es algo más intrincado y se realiza de la siguiente manera.

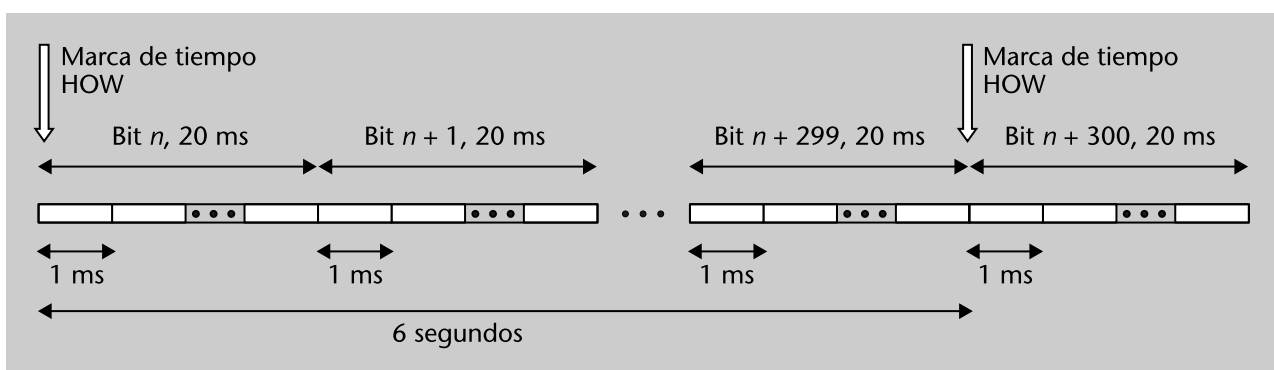
En la fase de seguimiento o *tracking*, el receptor mantiene una réplica de la señal GPS sincronizada con la señal que está realmente recibiendo. Por lo tanto, el receptor conoce en todo momento qué chip o qué muestra correspondiente a un chip está siendo recibida en un instante determinado, ya que coincide con la muestra que el receptor está generando para la réplica. El instante en el que nos tenemos que fijar es justamente el instante en el que el receptor lee su registro interno para calcular $t_u(t)$. Entonces el receptor ha de calcular el instante de tiempo en el que el satélite transmitió la muestra que ahora está recibiendo (el término "ahora" lo entendemos como el momento en el que el receptor decide medir el valor de t_u).

Para calcular el instante de transmisión correspondiente a una muestra cualquiera de la señal, se utilizan las marcas temporales que existen en la señal GPS. El satélite marca determinados eventos de la señal; concretamente marca el inicio de cada subtrama (*subframe*) mediante el TOW (*time of week*), tal como podéis ver en la figura 17. El TOW es un contador de 19 bits que contiene el instante de transmisión del inicio de la subtrama. El valor del bit menos significativo es de 1,5 segundos. Realmente, en la señal únicamente se incluyen los 17 bits más significativos del TOW, que forman la denominada HOW (*Hand-Over Word*). Al haber eliminado los 2 bits menos significativos, la HOW cuenta en unidades de 6 segundos, que es justamente la duración de una subtrama.

Ved también

El *tracking* se estudia en el subapartado 4.2 de este módulo.

Figura 17. Marcas de tiempo en la señal GPS

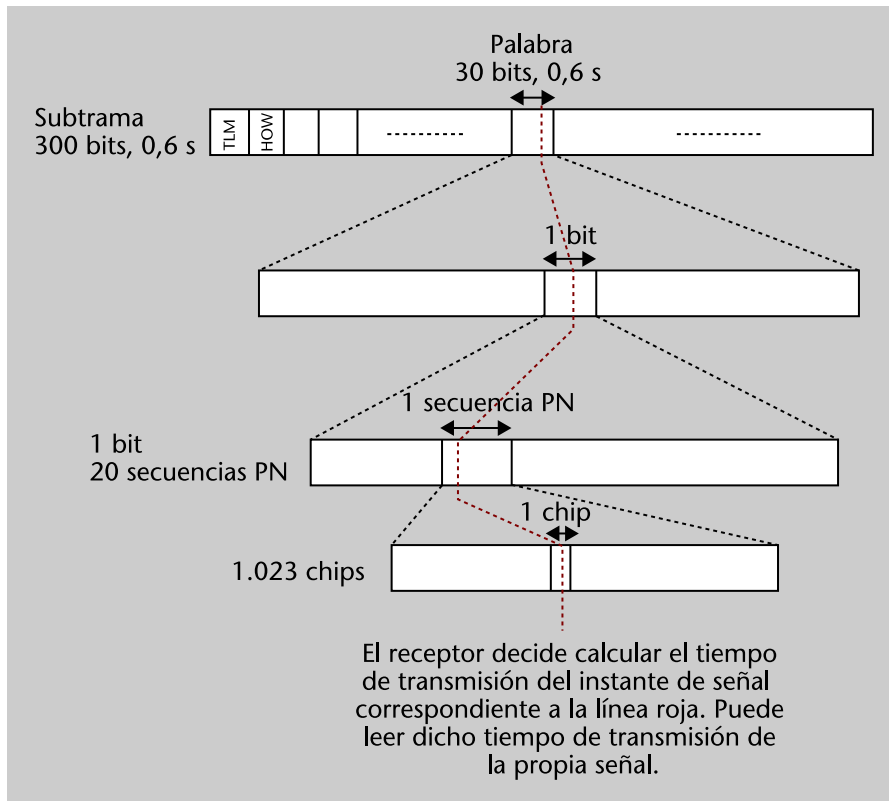


Dado que el receptor está demodulando y detectando los bits que contiene la señal, obtiene de manera inmediata, a través del HOW, el valor del tiempo de transmisión correspondiente al inicio de la subtrama en la que se encuentra la muestra de la cual desea calcular su tiempo de transmisión. Para pasar del tiempo de inicio de la subtrama al tiempo de transmisión de la muestra en cuestión, el receptor simplemente ha de contar la parte entera y fraccional del número de chips que han pasado desde el inicio de la subtrama hasta dicha

⁽¹⁴⁾Esto es factible para el receptor porque él mismo está generando una réplica de la señal recibida, y por tanto conoce los chips que está generando en cualquier momento.

muestra¹⁴. Como la duración de los chips en transmisión es conocida (por ejemplo, $1/1,023 \mu\text{s}$ para la señal GPS C/A), simplemente se debe multiplicar el número de chips por su duración y así se obtiene el tiempo que se le tiene que añadir al TOW para calcular el instante correspondiente a la muestra en la que se realiza la medida. Este proceso se explica de manera gráfica y simplificada en la figura 18.

Figura 18. Ilustración del cálculo del instante de transmisión



5.2. Cálculo de las fases de portadora

Una estimación mucho más precisa del retardo de propagación se puede obtener a partir de la medida de la fase de la portadora de la señal recibida del satélite. Este nuevo observable se denomina simplemente fase de portadora u observable de fase. Por semejanza, el pseudorrango se denomina alternatively como fase de código, ya que se obtiene calculando la posición o fase de una determinada muestra dentro de la secuencia pseudoaleatoria que forma la señal. De modo similar, se puede interpretar que la fase de portadora se calcula a partir de la posición de la muestra en cuestión dentro de un ciclo de la portadora.

Concretamente el observable de fase se define como la diferencia en el instante de medida entre la fase de la señal recibida y la fase de una portadora generada en el receptor.

No obstante, esta diferencia de fases no se mide módulo 1 ciclo (o 2π radianes), sino que se acumula de forma continua desde un determinado instante de inicio, por lo que a veces se denomina también *integrated Doppler* o *accumulated Doppler range* (pero no se debe confundir con una medida directa de la frecuencia de la señal, aunque se utilice el término *Doppler*). Como veremos a continuación, este observable es una medida indirecta y ambigua del tiempo de tránsito. Se suele decir que la fase de portadora es un medida precisa (*precise* en inglés) pero no exacta (*accurate* en inglés), mientras que el pseudorrango es lo contrario, esto es, es exacto pero no preciso. Esto es así porque el pseudorrango es una medida no ambigua pero con un nivel de ruido más elevado, mientras que la fase de portadora proporciona una estimación de la distancia con un nivel de ruido muy bajo pero con una ambigüedad dada por un número desconocido de longitudes de onda.

La fase de la portadora, en unidades de ciclos, generada por el receptor se puede expresar como

$$\phi_u(t) = f_0 t_u(t) + N_u = f_0 t + f_0 \delta t_u(t) + N_u \quad (44)$$

donde hemos utilizado la expresión 39, N_u simboliza número entero arbitrario de ciclos (dado que cualquier fase instantánea está definida salvo un número entero de ciclos y f_0 es la frecuencia portadora nominal de la señal). El término $f_0 \delta t_u(t)$ representa la desviación de fase causada por la diferencia entre la frecuencia nominal y la frecuencia real del receptor porque una variación del error en tiempo $\delta t_u(t)$ equivale a un error en frecuencia, dado que el reloj del receptor no es más que un oscilador y un contador de los ciclos de dicho oscilador. Por otro lado, la fase de la señal transmitida por el satélite se puede escribir como

$$\phi^s(t) = f_0 t^s(t) + N^s = f_0 t + f_0 \delta t^s(t) + N^s \quad (45)$$

donde el significado de cada uno de los términos es el mismo que en la expresión 44, pero haciendo referencia al satélite. El observable de fase se define como

$$\phi(t) = \phi_u(t) - \phi^s(t - \tau) \quad (46)$$

o sea, la diferencia entre la fase de la portadora local generada por el receptor en el instante t y la fase de la señal recibida en ese mismo instante, donde esta última coincide con la fase de la señal transmitida T segundos antes, dado que T es el tiempo de tránsito o propagación. Sustituyendo las expresiones 44 y 45 en la expresión 46, obtenemos

$$\phi(t) = \frac{f_0}{c} r(t, t - \tau) + f_0 (\delta t_u(t) - \delta t^s(t)) + \frac{f_0}{c} (I_\phi(t) + T \phi(t)) + N + \varepsilon_\phi(t), \quad (47)$$

donde $I_\phi(t)$ y $T_\phi(t)$ son los desplazamientos de fase, expresados en metros, causados por la ionosfera y troposfera, respectivamente; $\varepsilon_\phi(t)$ representa el ruido de media en $\phi^s(t)$, así como el resto de los efectos no modelados o errores en el modelo; y $N = N_u - N^s$ es la ambigüedad y sigue siendo un número entero. Por las propiedades físicas de la troposfera y de la ionosfera, el efecto de la primera es el mismo en el código (o sea, en la envolvente de la señal) que en la fase, mientras que la segunda se comporta como un medio dispersivo y provoca un adelanto de fase de igual valor al retraso que sufre en el código.

Por lo tanto, se cumple que $T_\phi = T_\rho$ y $I_\phi = -I_\rho$. Finalmente, teniendo en cuenta que la longitud de onda es $\lambda = \frac{c}{f_0}$ y omitiendo por sencillez la dependencia con t , la expresión final del observable de fase (en unidades de ciclos) es

$$\phi = \frac{r}{\lambda} + \frac{c}{\lambda}(\delta t_u - \delta t^s) + \frac{1}{\lambda}(T_\rho - I_\rho) + N + \varepsilon_\phi \quad (48)$$

La ambigüedad N se puede interpretar como el número entero de longitudes de onda entre el receptor y el satélite, cambiado de signo, en el instante inicial de medida. En este justo instante, la primera medida del fase del receptor es necesariamente un número entre 0 y 1 (expresado en ciclos, ya que fases que difieren entre sí en un número entero de ciclos son indistinguibles), mientras que los términos debidos a la distancia, errores de relojes y efectos atmosféricos en la parte derecha de la expresión 48 tendrán en principio un valor muy alejado del intervalo $[0,1)$.

Fijémonos por ejemplo en que r es del orden de decenas de miles de kilómetros, mientras que λ son decenas de centímetros, por lo tanto, r/λ es del orden de 10^8 . Para que la parte izquierda y derecha de la expresión 48 puedan ser iguales en el instante de la primera medida, es necesario añadir un valor entero N . Para las medidas sucesivas, el receptor puede seguir la evolución de las fases $\phi_u(t)$ y $\phi^s(t)$, así que la fase $\phi(t)$ puede seguir las evoluciones de los términos de la derecha de la ecuación 48 sin estar limitada a tomar valores en el intervalo $[0,1)$.

Ejemplo

Para ilustrar esto, consideremos a título de ejemplo que el satélite se aleja del receptor en cinco longitudes de onda entre el instante t_1 y el instante t_2 . En ausencia del resto de los efectos, el observable de fase en t_2 satisfará $\phi(t_2) = \phi(t_1) + 5$. Esto es cierto porque la ambigüedad entera presente en $\phi(t_1)$ y $\phi(t_2)$ es la misma, lo que equivale a decir que ambas fases están referidas al mismo instante inicial y que no se ha perdido el seguimiento o *tracking* de la fase entre t_1 y t_2 .

No obstante, a veces la fase estimada por el lazo de seguimiento de fase (por ejemplo, el PLL) experimenta saltos en un número entero de ciclos debido al ruido, interferencias u otras perturbaciones de la señal. Este fenómeno es conocido como *cycle slips*. Un *cycle slip* actúa como una reinicialización de la ambigüedad, de manera que se produce un cambio en el valor de N . Es decir,

Reflexión

Podéis hacer memoria de los contenidos de otras asignaturas y recordar que las guías de onda también se comportaban como medios dispersivos y las implicaciones que esto tenía para la transmisión de señales.

si se produce un *cycle slip* entre t_1 y t_2 , las fases $\phi(t_1)$ y $\phi(t_2)$ no son directamente comparables (por ejemplo, su diferencia no se debe solo a la variación de la distancia, efectos de los relojes y efectos de la propagación), sino que contienen ambigüedades diferentes. Por este motivo el posicionamiento con fases de portadora es un proceso delicado, ya que debe tener en cuenta que el valor de N en la expresión 48 es solo constante mientras no se produzcan *cycle slips*, pero cambia cuando estos ocurren, y esto implica que es necesario realizar un procesamiento previo de detección y corrección de *cycle slips*. Este procesamiento es un paso esencial en la mayoría de los métodos de posicionamiento que trabajan con el observable de fase.

La mayor precisión de los observables de fase se puede combinar con la ausencia de ambigüedad en los observables de código para obtener estimaciones que son al mismo tiempo precisas y no ambiguas. Esta forma de combinar las observaciones de código y fase se denomina *code-carrier smoothing* o simplemente *carrier smoothing* (que se traduciría como suavizado del código utilizando la fase).

5.3. Solución de navegación

El término *solución de navegación* es la forma habitual de referirse al cálculo de la posición y el tiempo a partir de algunos de los observables GPS. Esto también se suele denominar de forma genérica como estimación del PVT (posición, velocidad, tiempo), aunque en muchas ocasiones no se calcula explícitamente la velocidad. Existen multitud de algoritmos de posicionamiento dependiendo del tipo de observables que se utilizan, de si se realiza en tiempo real o en posprocesado, del número de señales en diferentes bandas que se emplean, de si se incluye un modelo dinámico, de si existe algún tipo de corrección diferencial o sistema de aumentación, etc. No obstante, para centrar la discusión, ahora nos vamos a ocupar del caso de posicionamiento instantáneo a partir de los pseudorrangos medidos a una única frecuencia. Se trata de un caso básico, pero resulta útil porque las bases en las que se fundamenta son comunes y extrapolables a los casos más complicados, y además permite analizar de una manera más clara los efectos de la geometría de los satélites y de los errores de los observables sobre la precisión del posicionamiento.

El punto de partida es la ecuación 43, donde de aquí en adelante indicaremos las magnitudes que hacen referencia al satélite k -ésimo mediante el superíndice (k) . Vamos a denotar como $P_c^{(k)}$ al pseudorango corregido, esto es, el pseudorango medido por el receptor ($P^{(k)}$) del que se han abstraído las estimaciones que disponga el receptor de los efectos del lado derecho de la expresión 43. Este siempre dispone al menos de una estimación del *offset* del reloj de satélite ($\delta t^{(k)}$) porque lo transmite el propio satélite en el mensaje de navegación. También es posible obtener una medida aproximada del retardo de la ionosfera a través de los coeficientes del modelo de Klobuchar (también incluidos en el mensaje de navegación) o se puede obtener una medida mucho más precisa

si dispone de observaciones a dos frecuencias o correcciones provenientes de un sistema de aumentación. Suponiendo que el receptor solo tuviera valores del *offset* del satélite y de la ionosfera, el pseudorange corregido se obtendría como $P_c^{(k)} = P^{(k)} + c\delta t^{(k)} - \hat{I}_\rho^{(k)}$. Esto se comenta únicamente a título de ejemplo. En el subapartado 5.5 explicaremos con más detalle cómo se pueden corregir u obtener estimaciones de cada uno de los efectos. Es evidente cómo se corregiría el pseudorange en otros casos. Observad que utilizamos el símbolo $\hat{}$ para referirnos a los valores estimados o aproximaciones, y así distinguirlos de los valores reales.

La relación entre el pseudorange corregido y las posiciones viene dada por

$$P_c^{(k)} = r^{(k)} + c\delta t_u + \varepsilon_{\rho,c} \quad (49)$$

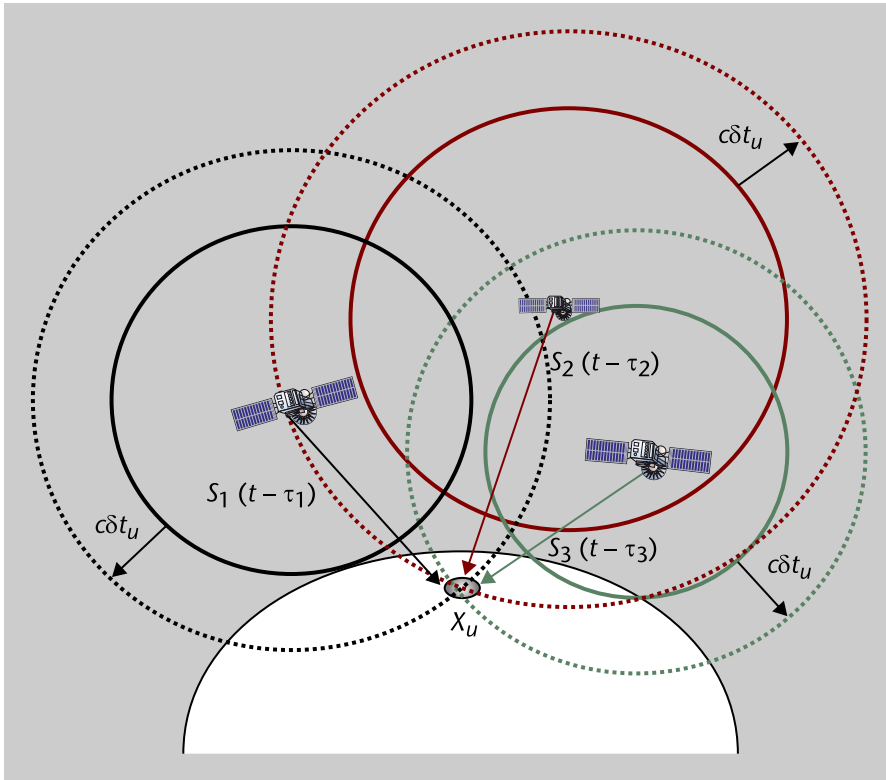
donde $\varepsilon_{\rho,c}$ incluye el término ε_ρ de la expresión 43 más por un lado todos los efectos que no se han podido corregir en $P_c^{(k)}$ (en el ejemplo anterior, sería la troposfera) y, por otro lado, los residuos entre las estimaciones de los efectos que se han corregido y los efectos reales (en el ejemplo anterior, sería el caso de $\hat{I}_\rho^{(k)} - \hat{I}_\rho^{(k)}$). A grandes rasgos, podemos decir que la desviación estándar de $\varepsilon_{\rho,c}$ es inferior a 1 metro en el caso de disponer de correcciones diferenciales, y está entre 5 y 10 metros para un receptor funcionando de modo autónomo. La distancia geométrica es el módulo del vector que va desde la posición del receptor $\mathbf{x}_u = [x_u, y_u, z_u]^T$ en el instante de medida y la posición del satélite en el instante de transmisión $\mathbf{x}^{(k)} = [x^{(k)}, y^{(k)}, z^{(k)}]^T$:

$$r^{(k)} = \|\mathbf{x}^{(k)} - \mathbf{x}_u\| = \sqrt{(x^{(k)} - x_u)^2 + (y^{(k)} - y_u)^2 + (z^{(k)} - z_u)^2} \quad (50)$$

Se puede plantear una ecuación del tipo de la expresión 49 para cada uno de los K satélites que esté recibiendo el receptor. Cada ecuación incluye 4 incógnitas: el *offset* del reloj y las tres coordenadas de la posición. Por tanto, se necesita un número de satélites $K \geq 4$ para poder calcular la posición en 3 dimensiones. En caso de posicionamiento en 2 dimensiones, se necesitarían al menos 3 satélites. La causa de necesitar al menos un satélite más que el número de coordenadas se ilustra en la figura 19. El receptor se encuentra en el punto de intersección de varias circunferencias en el caso de posicionamiento 2D o de varias superficies esféricas en el caso de posicionamiento 3D, cuyo radio viene dado por el pseudorange correspondiente menos el *offset* del reloj ($c\delta t_u$). Si nos centramos en el caso de 2D por sencillez, vemos que los puntos de intersección de cualquier par de circunferencias cambia con el valor $c\delta t_u$, y como este es desconocido, no se puede calcular la posición en 2D a partir de dos satélites únicamente. La clave del posicionamiento basado en GNSS

es el hecho que 3 circunferencias se intersectan en un único punto para un valor adecuado de $c\delta t_u$, y esto justifica que se necesiten al menos 3 satélites en posicionamiento 2D.

Figura 19. Ilustración de la solución de navegación y efecto del *offset* del reloj del receptor



Si se dispone de más satélites, no será posible hacer que todas las circunferencias se intersecten en un único punto debido al ruido de medida, pero el principio de funcionamiento seguirá siendo el mismo: el valor de $c\delta t_u$ será aquel que haga que un determinado punto (que será la estimación de las coordenadas del receptor) esté lo más próximo posible a todas las circunferencias. En función de cómo se definamos matemáticamente "lo más próximo posible", se obtendrán diferentes variaciones del método de posicionamiento.

La argumentación anterior ha dejado patente que en el proceso de cálculo de la posición en GNSS se obtiene también necesariamente valor del *offset* $c\delta t_u$, aunque para una aplicación de posicionamiento este parámetro podría no ser necesario. No obstante, esto indica que sin ninguna complejidad adicional un receptor GNSS puede corregir su base de tiempos con el *offset* calculado y obtener una estimación del tiempo del sistema (GPST en el caso de GPS). Esta es la base de las aplicaciones de *timing* de GNSS, ya que los receptores se pueden utilizar como generadores de frecuencia o de una base de tiempos muy precisa y teóricamente idéntica para todos los receptores aunque se encuentren muy alejados entre ellos.

Antes de retomar los métodos para solucionar el sistema de ecuaciones de posicionamiento, debemos hacer una puntualización sutil pero importante sobre la ecuación 50. Para que la expresión sea correcta, las coordenadas del receptor y de los satélites deben estar expresadas en el mismo sistema de coordenadas, que puede ser un sistema inercial (ECI) o un sistema rotatorio con la tierra (ECEF). Resulta mucho más fácil e intuitivo expresar la posición del usuario en el sistema ECEF correspondiente al instante de medida t (en este sistema las coordenadas de un usuario estático no cambian con el tiempo). Por lo tanto, tenemos que expresar la posición del satélite ($x^{(k)}$) en el instante de emisión $t-\tau$ en el sistema de coordenadas ECEF del instante t , pero las efemérides del mensaje de navegación nos permiten calcular la posición del satélite en el sistema ECEF correspondiente al instante $t-\tau$ (la denominaremos $\bar{x}^{(k)}$). Durante el tiempo que necesita la señal para viajar desde el satélite hasta la antena del usuario, la tierra y el ECEF rotan alrededor del eje z en un ángulo $\omega_E\tau$, donde ω_E es la velocidad angular de rotación de la tierra ($\sim 7.29\text{e-}5$ rad/s). Por lo tanto, la relación entre $\bar{x}^{(k)}$ y $x^{(k)}$ viene dada por

$$x^{(k)} = \begin{bmatrix} \cos\omega_E\tau & \sin\omega_E\tau & 0 \\ -\sin\omega_E\tau & \cos\omega_E\tau & 0 \\ 0 & 0 & 1 \end{bmatrix} \bar{x}^{(k)} \quad (51)$$

Para receptores situados en tierra, el tiempo de tránsito T está entre 67 y 86 ms, lo que implica una rotación de entre $4,9 \times 10^{-6}$ y $6,3 \times 10^{-6}$ radianes. Aunque un valor de esta magnitud podría parecer despreciable, realmente no lo es, porque si lo despreciamos se introduce un error de 10-20 m en el cálculo de la distancia según la expresión 50, lo que se traduce en un error de la misma magnitud en la posición en la dirección este-oeste. El problema que se plantea es que T realmente no es conocido porque depende de la distancia entre el satélite y el receptor, y la posición de este último es justamente una de las incógnitas del problema. Volveremos a la cuestión del cálculo de las posiciones de los satélites más adelante, después de explicar los métodos básicos de posicionamiento.

Regresando a la solución de navegación, tenemos que resolver el siguiente sistema, formado por K ecuaciones de la forma dada por la expresión 49:

$$\begin{bmatrix} P_c^{(1)} \\ P_c^{(2)} \\ \vdots \\ P_c^{(k)} \end{bmatrix} = \begin{bmatrix} r^{(1)}(x_u) \\ r^{(2)}(x_u) \\ \vdots \\ r^{(k)}(x_u) \end{bmatrix} + \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix} c\delta t_u \quad (52)$$

donde, para facilitar la comprensión, hemos explicitado la dependencia de la distancia verdadera con la posición del receptor. La principal dificultad reside en que se trata de un problema no lineal debido a la presencia del operador norma, y por lo tanto la manera de resolverlo no es inmediata.

Existen en general dos grandes familias de métodos para calcular la solución de navegación:

- Una familia se basa en linealizar la dependencia de la distancia con la posición.
- Los métodos de la segunda familia calculan la posición de forma cerrada y sin necesitar un proceso iterativo.

Los **métodos basados en la alineación** proporcionan mayor precisión pero requieren el conocimiento de un valor aproximado de la posición del receptor y son iterativos (por ejemplo, el resultado de aplicar el algoritmo se toma como nuevo valor a priori de la posición para la siguiente iteración). Ninguno de

estos aspectos es un problema serio porque muy pocas iteraciones suelen ser necesarias y porque en la práctica suele resultar sencillo conocer una posición aproximada inicial del receptor, bien utilizando la última posición que calculó en receptor cuando se empleó por última vez o bien aplicando un método de la segunda familia.

Los **métodos con solución cerrada** proporcionan una precisión peor que los métodos basados en la linealización, pero no necesitan ningún conocimiento previo de la posición.

5.3.1. Métodos basados en la linealización

Vamos a presentar a continuación el método tradicional de posicionamiento basado en la linealización, que es representativo de cualquier otra variante dentro de esta primera familia. Dado que el sistema de ecuaciones de la expresión 52 no se puede cumplir con igualdad debido la existencia del término ε_{pc} en 49, es habitual plantear un ajuste de mínimo cuadrados. Por lo tanto, la posición y *offset* temporal del receptor se calculan como los valores que proporcionan un mejor ajuste entre las medidas y el modelo. Esto se puede formular como el valor que minimiza la siguiente función de coste:

$$f(r_u, \delta t_u) = (p_c - r(x_u) - 1\delta t_u)^T W(\mathbf{p}_c - \mathbf{r}(x_u) - \mathbf{1}\delta t_u) \quad (53)$$

donde hemos definido los siguientes vectores $\mathbf{p}_c = [p_c^{(1)}, \dots, p_c^{(k)}]^T$, $\mathbf{r}(x_u) = [r^{(1)}(x_u), \dots, r^{(k)}(x_u)]^T$; $\mathbf{1}$ es un vector columna de todos unos, y la W es una matriz de pesos¹⁵.

⁽¹⁵⁾Comentaremos más adelante posibles alternativas para construir esta matriz, siendo no obstante la matriz identidad una opción válida y sencilla.

Observamos que la función $f(\cdot)$ no es más que una suma ponderada de los residuos. La minimización de expresiones como la 53 es un problema muy estudiado en matemáticas y denominado regresión no lineal (Seber; Wild, 1989). En nuestro caso, además de ser no lineal, la función $f(r_u, \delta t_u)$ no es convexa, por lo que puede tener varios mínimos locales, lo que siempre es una dificultad añadida. Una técnica habitual para resolver problemas no lineales con restricciones es el método de Newton-Raphson, que sería también aplicable en nuestro caso. No obstante, un método similar es el de Gauss-Newton, que es el que habitualmente se utiliza en el ámbito de la solución de navegación de GNSS. Este método se basa en linealizar $r^{(k)}(x_u)$ alrededor de un valor tentativo de la posición del usuario, que denominaremos $\mathbf{x}_{u,0} = [x_{u,0}, y_{u,0}, z_{u,0}]^T$, utilizando el polinomio de Taylor de primer orden. Se puede comprobar fácilmente que

$$\begin{aligned}
r^{(k)}(\mathbf{x}_u) &\approx r^{(k)}(\mathbf{x}_{u,0}) - \frac{(x^{(k)}-x_{u,0})\Delta x_u + (y^{(k)}-y_{u,0})\Delta y_u + (z^{(k)}-z_{u,0})\Delta z_u}{\sqrt{(x^{(k)}-x_{u,0})^2 + (y^{(k)}-y_{u,0})^2 + (z^{(k)}-z_{u,0})^2}} \\
&= r^{(k)}(\mathbf{x}_{u,0}) - \frac{(\mathbf{x}^{(k)}-\mathbf{x}_{u,0})^T}{\|\mathbf{x}^{(k)}-\mathbf{x}_{u,0}\|} \Delta \mathbf{x}_u \\
&= r^{(k)}(\mathbf{x}_{u,0}) - (\mathbf{v}_u^{(k)}) \Delta \mathbf{x}_u
\end{aligned} \tag{54}$$

donde $\Delta \mathbf{x}_u = [\Delta x_u, \Delta y_u, \Delta z_u]^T = \mathbf{x}_u - \mathbf{x}_{u,0}$ es la posición del usuario relativa al valor tentativo a priori, y $\mathbf{v}_u^{(k)}$ es el vector unitario que va en la dirección del receptor al satélite k -ésimo.

Vector unitario

El vector unitario también se suele denominar como el vector de cosenos directores.

Introduciendo este desarrollo en la expresión 52, obtenemos

$$\frac{\mathbf{p}_c - \mathbf{r}(\mathbf{x}_{u,0})}{\Delta \mathbf{p}_c} = \underbrace{\begin{bmatrix} -(\mathbf{v}_u^{(1)})^T & 1 \\ -(\mathbf{v}_u^{(2)})^T & 1 \\ \vdots & \vdots \\ -(\mathbf{v}_u^{(k)})^T & 1 \end{bmatrix}}_{\mathbf{H}} \begin{bmatrix} \Delta \mathbf{x}_u \\ c\delta t_u \\ \Delta c \end{bmatrix} \tag{55}$$

Es inmediato identificar cada uno de los términos que aparecen en la formulación habitual de sistemas de ecuaciones lineales: los datos $\Delta \mathbf{p}_c$, las incógnitas $\Delta \mathbf{c}$ y la matriz de coeficientes \mathbf{H} . Si $K = 4$, el sistema es determinado y tiene una única solución (excepto en casos degenerados). Si $K > 4$, el sistema es, en general, sobredeterminado (presenta más de una solución válida), y lo que se hace en este caso es buscar una solución de mínimos cuadrados. Este valor es el que minimiza la suma de residuos, que ahora se puede expresar como

$$f(\Delta \mathbf{c}) = (\Delta \mathbf{p}_c - \mathbf{H}\Delta \mathbf{c})^T \mathbf{W}(\Delta \mathbf{p}_c - \mathbf{H}\Delta \mathbf{c}) \tag{56}$$

La solución es

$$\Delta \hat{\mathbf{c}} = \begin{bmatrix} \Delta \hat{\mathbf{x}}_u \\ c\delta \hat{t}_u \end{bmatrix} = (\mathbf{H}^T \mathbf{W} \mathbf{H})^{-1} \mathbf{H}^T \mathbf{W} \Delta \mathbf{p}_c \tag{57}$$

y la nueva estimación de la posición es $\hat{\mathbf{x}}_u = \mathbf{x}_{u,0} + \Delta \mathbf{x}_u$. Este procedimiento se puede iterar de nuevo, linealizando las ecuaciones alrededor de esta nueva posición. La convergencia suele ser rápida, incluso si el primer valor tentativo de la posición tiene un error de kilómetros. Se suelen realizar entre dos y cuatro iteraciones, siendo dos el valor más habitual.

La matriz \mathbf{W} se debería elegir de tal manera que se minimizaran los errores en la posición calculada. Si los errores en los pseudorrangos fueran de media cero, la matriz \mathbf{W} que conseguiría esto sería la inversa de la matriz de correlación de los

pseudorrangos. No obstante, esta matriz de correlación suele ser desconocida, así que se utilizan algunas reglas heurísticas para construir la matriz \mathbf{W} . Esta matriz debe introducir idealmente dos efectos:

- Reducir el peso de aquellas ecuaciones en las que las medidas sean más ruidosas o, lo que es lo mismo, dar más importancia a aquellas ecuaciones con medidas de más calidad.
- Acoplar las ecuaciones cuyas medidas estén correladas.

Para conseguir el primer efecto, es habitual hacer que las componentes de la diagonal de \mathbf{W} sean directamente proporcionales a alguna magnitud relacionada cualitativamente con la calidad de los pseudorrangos, como puede ser la relación C/No o el ángulo de elevación. El segundo efecto se consigue con las componentes que están fuera de la diagonal de \mathbf{W} . No obstante, obtener una medida cualitativa de la correlación suele ser complicado y es habitual que la matriz \mathbf{W} sea diagonal.

5.3.2. Métodos con solución cerrada

Vamos a presentar un método perteneciente a lo que hemos denominado anteriormente como la segunda familia de técnicas de posicionamiento, es decir, aquellas que calculan la posición de forma cerrada. Concretamente, vamos a describir el **método de Bancroft**, que es posiblemente el representante mejor conocido de esta familia. Cada una de las ecuaciones de la expresión 52 se puede expresar como

$$P_c^{(k)} - x\delta t_u = \sqrt{(x^{(k)} - x_u)^2 + (y^{(k)} - y_u)^2 + (z^{(k)} - z_u)^2} \quad (58)$$

Si la elevamos al cuadrado y desarrollamos todos los términos cuadráticos, obtenemos

$$\begin{aligned} (P_c^{(k)})^2 + (c\delta t_u)^2 - 2P_c^{(k)}c\delta t_u &= (x^{(k)})^2 + (y^{(k)})^2 + (z^{(k)})^2 \\ &+ (x_u)^2 + (y_u)^2 + (z_u)^2 - 2x^{(k)}x_u - 2y^{(k)}y_u - 2z^{(k)}z_u \end{aligned} \quad (59)$$

Ahora, definiendo $B = (x_u)^2 + (y_u)^2 + (z_u)^2 - (c\delta t_u)^2$ y reagrupando términos, resulta

$$(P_c^{(k)})^2 - (x^{(k)})^2 - (y^{(k)})^2 - (z^{(k)})^2 = -2x^{(k)}x_u - 2y^{(k)}y_u - 2z^{(k)}z_u + 2P_c^{(k)}c\delta t_u + B \quad (60)$$

En este punto se puede proceder de dos maneras. La primera de ellas consiste en suponer que B es una variable adicional, omitiendo su relación con las coordenadas y con el *offset*. En este caso podemos plantear el siguiente sistema de ecuaciones lineales

$$\begin{bmatrix} (P_c^{(1)})^2 - (x(1))^2 - (y(1))^2 - (z(1))^2 \\ (P_c^{(2)})^2 - (x(2))^2 - (y(2))^2 - (z(2))^2 \\ \vdots \\ (P_c^{(k)})^2 - (x(k))^2 - (y(k))^2 - (z(k))^2 \end{bmatrix} = \begin{bmatrix} -2x(1) & -2y(1) & -2z(1) & 2P_c^{(1)} & 1 \\ -2x(2) & -2y(2) & -2z(2) & 2P_c^{(2)} & 1 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ -2x(k) & -2y(k) & -2z(k) & 2P_c^{(k)} & 1 \end{bmatrix} \begin{bmatrix} x_u \\ y_u \\ z_u \\ c\delta t_u \\ B \end{bmatrix} \quad (61)$$

que es determinado si $K = 5$. Para $K > 5$, es un sistema sobredeterminado y se puede resolver por mínimos cuadrados, como se ha discutido anteriormente para el método basado en la linealización. No obstante, parte del peaje que pagamos en este caso por obtener la posición sin requerir un valor tentativo previo es que necesitamos medidas de al menos 5 satélites; el resto del peaje es una peor precisión con respecto a los métodos basados en la linealización.

La segunda manera de proceder a partir de la expresión 60 nos conduce a lo que se conoce propiamente como método de Bancroft, y permite evitar que se necesiten 5 satélites como mínimo. Para un valor dado de B , el sistema de ecuaciones resultante es

$$\begin{aligned} & \underbrace{\begin{bmatrix} (P_c^{(1)})^2 - (x(1))^2 - (y(1))^2 - (z(1))^2 \\ (P_c^{(2)})^2 - (x(2))^2 - (y(2))^2 - (z(2))^2 \\ \vdots \\ (P_c^{(k)})^2 - (x(k))^2 - (y(k))^2 - (z(k))^2 \end{bmatrix}}_{\mathbf{p}'} + \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix} B = \\ & = \underbrace{\begin{bmatrix} -2x(1) & -2y(1) & -2z(1) & 2P_c^{(1)} & 1 \\ -2x(2) & -2y(2) & -2z(2) & 2P_c^{(2)} & 1 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ -2x(k) & -2y(k) & -2z(k) & 2P_c^{(k)} & 1 \end{bmatrix}}_H \begin{bmatrix} x_u \\ y_u \\ z_u \\ c\delta t_u \\ B \end{bmatrix} \quad (62) \end{aligned}$$

cuya solución de mínimos cuadrados es

$$\hat{\mathbf{c}} = \begin{bmatrix} \hat{x}_u \\ \hat{y}_u \\ \hat{z}_u \\ c\delta t_u \end{bmatrix} = (\mathbf{H}'\mathbf{W}\mathbf{H})^{-1} \mathbf{H}'\mathbf{W}(\mathbf{p}' + \mathbf{1}B) \quad (63)$$

donde \mathbf{W} representa también aquí una determinada matriz simétrica de ponderación. El punto clave consiste en imponer que B no es un valor arbitrario, sino que debe cumplir una determinada relación con el resto de las variables,

que se puede expresar como $B = \hat{\mathbf{c}}^T \mathbf{D} \hat{\mathbf{c}}$, siendo \mathbf{D} una matriz diagonal con unos en todas las posiciones excepto en la última, que vale -1 . Aplicando esto a la expresión 63, obtenemos

$$B = (\mathbf{p}' + 1B)^T \mathbf{W} \mathbf{H} (\mathbf{H}^T \mathbf{W} \mathbf{H})^{-1} \mathbf{D} (\mathbf{H} \mathbf{W} \mathbf{H}')^{-1} \mathbf{H}^T \mathbf{W} (\mathbf{p}' + 1B) \quad (64)$$

que es una ecuación cuadrática en B :

$$1^T \mathbf{T} 1 B^2 + (1^T \mathbf{T} \mathbf{p}' - 1)^{-1} B + \mathbf{p}'^T \mathbf{T} \mathbf{p}' = 0 \quad (65)$$

donde por simplificar la notación hemos definido $\mathbf{T} = \mathbf{W} \mathbf{H} (\mathbf{H}^T \mathbf{W} \mathbf{H})^{-1} \mathbf{D} (\mathbf{H} \mathbf{W} \mathbf{H}')^{-1} \mathbf{H}^T \mathbf{W}$. Los siguientes pasos del algoritmo consisten en resolver la ecuación 65 e introducir el valor obtenido de B en la 63, que ya proporciona directamente entonces las estimaciones de las coordenadas y del reloj.

En general, los métodos que calculan la posición de forma cerrada implican elevar al cuadrado las ecuaciones del modelo para eliminar la raíz cuadrada de la expresión de la distancia. Al hacer esto también se eleva al cuadrado la contribución del ruido que hay en los pseudorrangos, y cualitativamente se puede entender que esto provoca que los errores aumenten con estos métodos. Además, se modifica la distribución estadística del ruido. Por ejemplo, si el ruido es Gaussiano, al elevar al cuadrado el ruido deja de ser Gaussiano y entonces el ajuste de mínimos cuadrados no tiene por qué ser el método de estimación óptimo.

5.3.3. Algoritmos completos para la solución de navegación

Una dificultad añadida en el cálculo de la posición del receptor y que hemos pasado por alto en las explicaciones anteriores es el cálculo de la posición de los satélites. Esto sorprende altamente a primera vista porque pensamos que las incógnitas son las coordenadas del receptor y que las coordenadas de los satélites son parte de los datos. Esto no es cierto completamente, y en general no se puede suponer que las coordenadas de los satélites sean conocidas directamente. Ya se intuye aquí la dificultad del problema, porque puede parecer imposible poder determinar la posición del receptor si simultáneamente también se debe determinar la posición de los puntos de referencia, que son los satélites. No obstante, el propio funcionamiento del sistema GPS demuestra que sí que es factible.

La dificultad en determinar la posición de cada satélite se debe a que hay dos movimientos simultáneos implicados: el propio movimiento del satélite y el giro de la Tierra, que implica que también gire cualquier sistema de coordenadas ligado a ella. Por lo tanto, para calcular la posición del satélite hay que

saber el instante de transmisión de la señal (en el tiempo GPS) para calcular la posición del satélite en el sistema ECI y también el tiempo de tránsito de la señal desde el propio satélite hasta el receptor para aplicar la rotación necesaria al sistema de coordenadas y poder expresar todas las coordenadas en un sistema ECEF referido a un único instante de tiempo (típicamente el del instante de recepción). Ya se ve que el cálculo del ángulo de rotación necesario es complicado porque el tiempo de tránsito depende de la posición del satélite y del receptor, y al menos esta última es una incógnita. Parece que estamos ante un "problema del gato y el ratón" porque para calcular la posición del receptor necesitamos las coordenadas de los satélites en un sistema de coordenadas adecuado, pero para obtener esto último necesitamos las coordenadas del receptor... Por lo tanto, el cálculo de la solución de navegación no es tan sencillo como resolver algunos de los sistemas de ecuaciones presentados anteriormente, sino que se debe utilizar algún algoritmo para solventar esta interdependencia entre "datos" e incógnitas.

Existen diferentes variaciones de estos algoritmos en función de cuáles sean los datos de partida. Vamos a presentar los dos algoritmos que corresponden a las situaciones más habituales:

1) **Algoritmo 1.** Corresponde al caso en el que la determinación del instante de transmisión de la señal es directa o casi directa. Esto ocurre cuando se trabaja "dentro del receptor" (es decir, teniendo acceso a la parte de procesado de señal del receptor) o cuando se dispone de los pseudorrangos. En la primera situación se calcula el tiempo de emisión de la señal en el tiempo del satélite simplemente observando la posición de la muestra recibida dentro de la señal recibida. Como se conoce en qué instante comienzan las subtramas según el reloj del satélite y la duración de los chips, se puede calcular el tiempo de emisión que corresponde a una muestra dada (esto se discutió juntamente con la figura 18). En la segunda situación, la propia definición de pseudorango (ecuación 38) nos proporciona el instante de emisión medido según el reloj del satélite:

$$t^s = t_u - \frac{P}{c} \quad (66)$$

En ambas situaciones, podemos calcular el instante de emisión en la escala de tiempo GPS (t_{GPS}^s) corrigiendo el *offset* del reloj del satélite, que se transmite a través del mensaje de navegación, resultando en

$$t_{GPS}^s = t^s - \delta t^s \quad (67)$$

Una vez que se dispone de t_{GPS}^s se puede aplicar el siguiente algoritmo:

- a) Calcular las coordenadas del satélite en el instante t_{GPS}^s y en el sistema de coordenadas ECI.
- b) Calcular una aproximación del tiempo de tránsito. Se puede suponer que es simplemente 70 ms o calcularlo a partir de la distancia geométrica entre la posición del satélite calculada en el paso *a* y una estimación inicial de la posición del receptor (aunque en este punto los sistemas de coordenadas en los que se expresan cada una de las posiciones sean diferentes no hay ningún problema).
- c) Una vez calculado el tiempo de tránsito, se rota la posición del satélite en el ángulo correspondiente utilizando la expresión 51.
- d) Se pueden realizar varias iteraciones de los pasos *b* y *c*. Lo habitual es realizar una sola iteración más o ninguna. En cualquier caso, las iteraciones consisten en: calcular la distancia geométrica entre la posición del satélite obtenida en el paso *c* y la posición a priori del receptor; a partir de aquí, se obtiene un nuevo tiempo de tránsito, se aplica la rotación correspondiente a la posición obtenida en el paso *a*, y se puede continuar iterando a partir de la posición resultante de la rotación.
- e) Se resuelven las ecuaciones de navegación y se obtiene la posición del receptor.
- f) Con la nueva posición del receptor, se puede volver al paso *b*. Es habitual realizar esta iteración una vez.

2) **Algoritmo 2.** Corresponde al caso en el que la determinación del instante de transmisión no es inmediato y se tiene que obtener a partir del instante de recepción. La clave radica en que las posiciones de los satélites no son arbitrarias (como lo es en principio la del receptor), sino que cada satélite solo se puede encontrar en una curva de \mathfrak{R}^3 que es función del tiempo. Dado que el instante de transmisión no es observable directamente como en el caso anterior, sino que se ha de estimar, las coordenadas del satélite en el instante de transmisión y en el sistema ECI tampoco se pueden obtener directamente (a diferencia de como se hacía en el punto *a* del algoritmo anterior). O sea, a diferencia del caso 1, en el que había que averiguar la rotación que se debía aplicar a las coordenadas del satélite, ahora, además de esto, también hay que averiguar de forma implícita qué instante de transmisión es compatible con la órbita que siguen los satélites y con el instante de recepción.

Un posible algoritmo adecuado para esta situación es el siguiente:

- a) Calcular una estimación del instante de transmisión restando al instante de recepción (que es el dato de partida) una estimación del tiempo de tránsito. En la primera iteración, tiene sentido utilizar como tiempo de tránsito un valor de 70 ms o algo cercano. En el resto de las iteraciones, se puede obtener a partir de la distancia geométrica entre la última estimación de las coordenadas del satélite y del receptor.
- b) Calcular las coordenadas del satélite en el sistema ECI y en el instante obtenido en el punto anterior.
- c) Calcular el tiempo de tránsito a partir de las coordenadas obtenidas en el paso *b* y una posición a priori del receptor.
- d) Rotar las coordenadas del satélite.
- e) Se puede volver al volver al punto *a*. En este caso es habitual realizar al menos una iteración más o varias, y entonces parar las iteraciones cuando las posiciones del satélite obtenidas en iteraciones sucesivas difieran menos que un cierto umbral.
- f) Se resuelven las ecuaciones de navegación, y se obtiene la posición del receptor y el *offset* del reloj del receptor.
- g) Con la nueva posición del receptor y el tiempo de recepción corregido por el *offset* calculado en el paso *f*, se puede volver al punto *a*. Es habitual realizar esta iteración una vez.

En el caso 2, el instante de transmisión estimado estará afectado por los *offsets* presentes en el instante de recepción y en la posición del receptor, y estos *offsets* se trasladan también a las coordenadas de los satélites. Esto provoca que las derivadas parciales que forman el sistema de ecuaciones linealizado (ecuación 55) se tengan que modificar. No obstante, este es un tema notablemente avanzado que excede el nivel de este material.

5.4. Fuentes de error

Para el análisis de los errores vamos a retomar la solución de navegación obtenida en la expresión 57 y vamos a suponer, como es habitual en la literatura, que los errores en $\Delta \mathbf{p}_c$ tienen media cero y covarianza \mathbf{c}_p . Entonces, los errores en las estimaciones de las coordenadas y del *offset* también tiene media cero y su matriz de covarianza es

$$\mathbf{c}_{\text{pos}} = \text{cov} \left\{ \begin{bmatrix} \Delta \hat{\mathbf{x}}_u \\ c \delta \hat{t}_u \end{bmatrix} \right\} = (\mathbf{H}^T \mathbf{W} \mathbf{H})^{-1} \mathbf{H}^T \mathbf{W} \mathbf{C}_p \mathbf{W} \mathbf{H} (\mathbf{H}^T \mathbf{W} \mathbf{H})^{-1} \quad (68)$$

A continuación vamos a analizar esta expresión para unos determinados valores de las matrices \mathbf{W} y \mathbf{C}_ρ . Aunque estos sean unos valores particulares, este es el tipo de análisis más habitual en GNSS. El objetivo no es tanto realizar un modelado o predicción precisos de los valores de los errores, sino entender cualitativamente la dependencia de los errores en la posición con los diferentes efectos físicos, y para esto nos van a resultar útiles las suposiciones que vamos a realizar a continuación. Si elegimos la matriz de ponderación como $\mathbf{W} = \mathbf{C}_\rho^{-1}$, entonces

$$\mathbf{C}_{\text{pos}} = (\mathbf{H}^T \mathbf{C}_\rho^{-1} \mathbf{H}) \quad (69)$$

Si suponemos además que los errores en los pseudorrangos están incorrelados y tienen la misma varianza, su matriz de correlación adopta la expresión $\mathbf{C}_\rho = \sigma_{\text{URE}}^2 \mathbf{I}$, y obtenemos

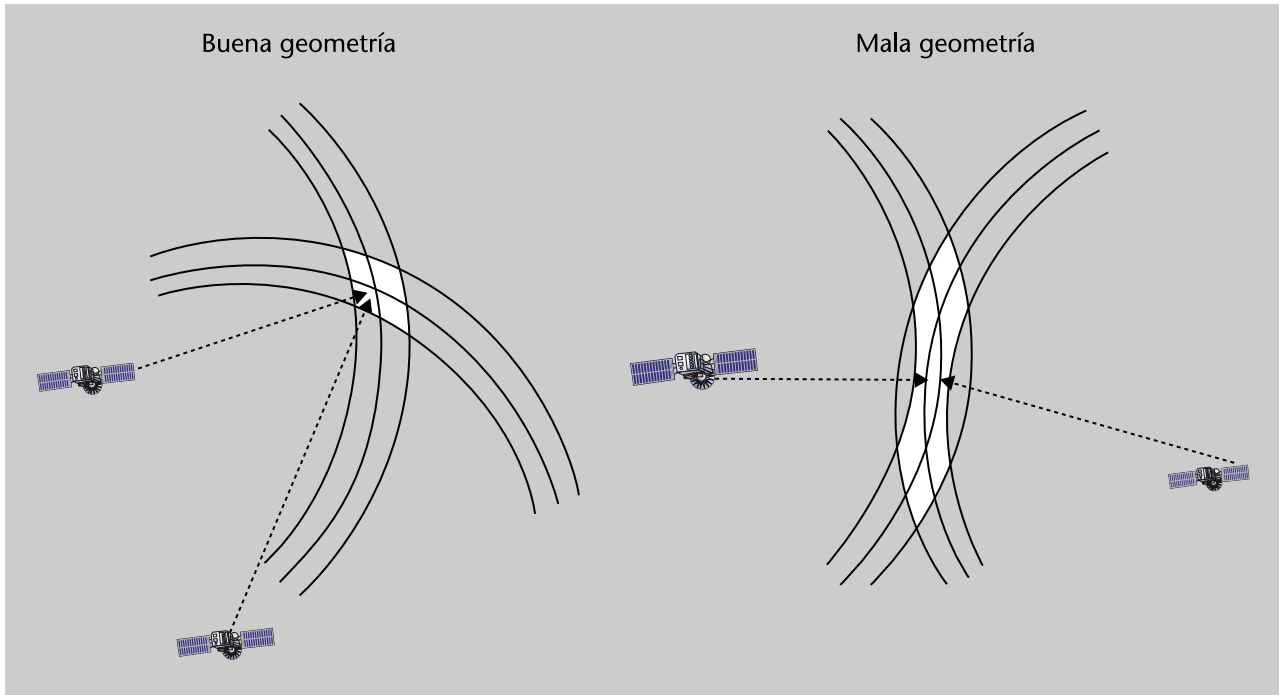
$$\mathbf{C}_{\text{pos}} = \sigma_{\text{URE}}^2 (\mathbf{H}^T \mathbf{H})^{-1} \quad (70)$$

donde UERE son las siglas correspondientes a *user equivalent range error*. Al final obtenemos que la magnitud de los errores depende de dos factores por separado: de los errores en las medidas, σ_{URE}^2 , y de un término que es función únicamente de la geometría, $\mathbf{G} = (\mathbf{H}^T \mathbf{H})^{-1}$. Este resultado que se ha obtenido a partir de fórmulas es en realidad intuitivo.

Ejemplo

Para ilustrar esto, consideremos un ejemplo muy sencillo en el que un usuario se quiere posicionar realizando medidas de distancia (no de pseudodistancia) a dos satélites. La posición se obtendrá como el punto de intersección de dos circunferencias centradas en las estaciones. Esto se representa en los esquemas de la figura 20. Los posibles errores en las medidas se muestran como una variación del radio de las circunferencias, y es la misma en los dos casos. Lo que varía es el ángulo con el que se ven los satélites desde el usuario. La zona sombreada representa la región de posibles errores en la posición causada por los errores en las medidas. Para unos mismos errores en las medidas, los errores en la posición cambian en función de la posición relativa de los satélites. Los errores aumentan cuando los satélites se ven desde el usuario en direcciones cercanas u opuestas, lo que corresponde a la idea intuitiva de que las medidas tienen menor diversidad espacial. Por otro lado, es evidente que para cualquiera de las dos geometrías los errores aumentarían si creciera la zona de incertidumbre de las medidas (es decir, la distancia entre las circunferencias concéntricas).

Figura 20. Ejemplo simple para ilustrar la dependencia de la precisión en la posición en los errores de medida y en la geometría



5.4.1. Factores DOP. Factores geométricos que afectan al cálculo de la solución de navegación

Vamos a analizar a continuación en detalle cada uno de los factores que afectan a los errores. Empezamos por los términos ligados a la geometría, para los que se introducen los llamados factores DOP (*dilution of precision*).

Los factores DOP son los factores de escala entre los errores en las medidas y los errores en las variables de interés (por ejemplo, posición y *offset* de reloj).

Por ejemplo, el error RMS en la posición (definido como la raíz cuadrada de la suma de varianzas en las tres coordenadas) es:

$$\begin{aligned} \text{Error RMS en posición} &= \sqrt{\sigma_x^2 + \sigma_y^2 + \sigma_z^2} \\ &= \sqrt{[C_{\text{pos}}]_{1,1} + [C_{\text{pos}}]_{2,2} + [C_{\text{pos}}]_{3,3}} = \sigma_{\text{URE}} \sqrt{[G]_{1,1} + [G]_{2,2} + [G]_{3,3}} \end{aligned} \quad (71)$$

Notación
El elemento (m,n) de una matriz se representa como $[\cdot]_{m,n}$.

Por lo tanto, podemos definir el PDOP (*position dilution of precision*) como $\sqrt{[G]_{1,1} + [G]_{2,2} + [G]_{3,3}}$. De la misma manera se pueden definir otros factores, cuyo significado es obvio a partir del nombre. A continuación se muestra la lista completa.

$$\begin{aligned}
\text{Position dilution of precision (PDOP)} &= \sqrt{[\mathbf{G}]_{1,1} + [\mathbf{G}]_{2,2} + [\mathbf{G}]_{3,3}} \\
\text{Time dilution of precision (TDOP)} &= \sqrt{[\mathbf{G}]_{4,4}} \\
\text{Geometric dilution of precision (GDOP)} &= \sqrt{[\mathbf{G}]_{1,1} + [\mathbf{G}]_{2,2} + [\mathbf{G}]_{3,3} + [\mathbf{G}]_{4,4}}
\end{aligned} \tag{72}$$

Las coordenadas con las que hemos trabajado hasta ahora corresponden al sistema ECEF, pero desde el punto de vista del usuario, a la hora de evaluar los errores, resulta mucho más práctico e intuitivo pensar en las componentes horizontal y vertical del error, que están definidas según el sistema de coordenadas local *east-north-up* (ENU). Las coordenadas ECEF se pueden transformar en coordenadas ENU mediante una matriz unitaria \mathbf{R}_L . Por lo tanto, la solución de navegación en el sistema ENU es

$$\begin{bmatrix} \Delta \hat{x}_{u,L} \\ \Delta \hat{y}_{u,L} \\ \Delta \hat{z}_{u,L} \\ c\delta \hat{t}_u \end{bmatrix} = \begin{bmatrix} \mathbf{R}_L & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \Delta \hat{x}_u \\ \Delta \hat{y}_u \\ \Delta \hat{z}_u \\ c\delta \hat{t}_u \end{bmatrix} \tag{73}$$

siendo su covarianza

$$C_{\text{pos},L} = (\mathbf{R}_L C_{\text{pos}} \mathbf{R}_L^T) = \sigma_{\text{URE}}^2 \mathbf{R}_L (\mathbf{H}^T \mathbf{H})^{-1} \mathbf{R}_L^T = \sigma_{\text{URE}}^2 (\mathbf{R}_L \mathbf{H}^T \mathbf{H} \mathbf{R}_L^T)^{-1} \tag{74}$$

donde en la última igualdad hemos utilizado que \mathbf{R}_L es una matriz unitaria.

De modo similar a como hemos hecho con la matriz \mathbf{G} , podemos definir nuevos factores DOP a partir de la matriz $\tilde{\mathbf{G}} = (\tilde{\mathbf{H}}^T \tilde{\mathbf{H}})^{-1} = (\mathbf{R}_L \mathbf{H}^T \mathbf{H} \mathbf{R}_L^T)^{-1}$. Son los siguientes.

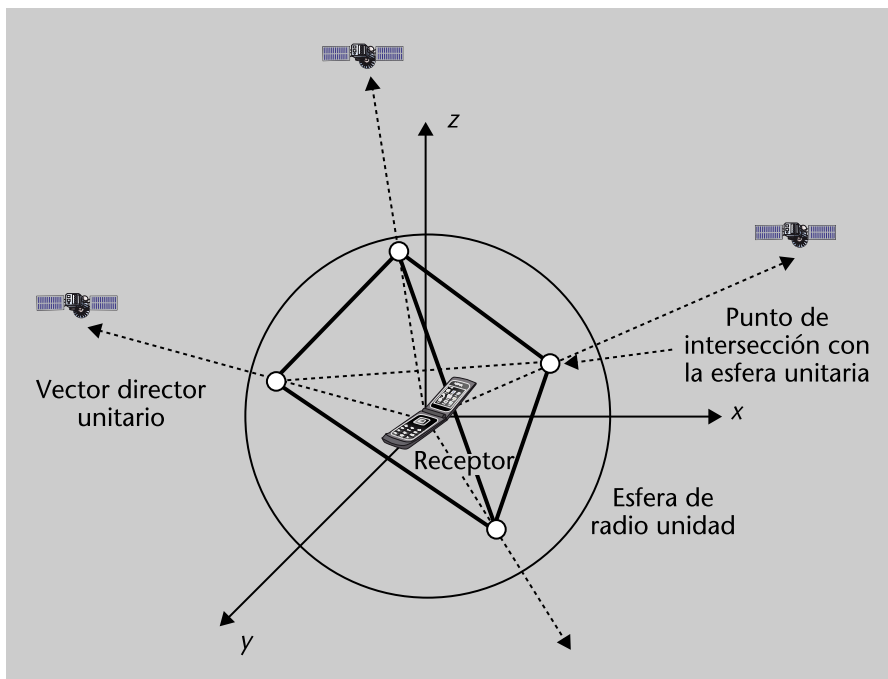
$$\begin{aligned}
\text{Horizontal dilution of precision (HDOP)} &= \sqrt{[\tilde{\mathbf{G}}]_{1,1} + [\tilde{\mathbf{G}}]_{2,2}} \\
\text{Vertical dilution of precision (VDOP)} &= \sqrt{[\tilde{\mathbf{G}}]_{3,3}}
\end{aligned} \tag{75}$$

Los valores de DOP se hacen menores al aumentar el número de satélites visibles. HDOP y TDOP son típicamente próximos a uno, y generalmente no sobrepasan el valor de dos, mientras que VDOP tiende a ser superior a HDOP. Una explicación cualitativa de esto último es que la precisión horizontal suele ser mejor que la vertical porque los satélites están en acimuts distribuidos de manera más o menos uniforme alrededor del usuario, pero solo con elevaciones positivas. De hecho, si la tierra fuera un globo transparente, entonces, debido a la simetría del problema, las precisiones en las direcciones norte, este y vertical serían muy similares. A grandes rasgos, un buen valor (o sea, un valor bajo)

de HDOP requiere una distribución de satélites razonablemente uniforme en sentido acimutal y situados no muy por encima del horizonte. Para reducir el VDOP se necesita disponer de algún satélite próximo al cénit. El VDOP se podría reducir más utilizando satélites situados por debajo del horizonte, pero estos satélites solo son visibles en receptores embarcados en otros satélites.

En definitiva, un buen PDOP se obtiene cuando existe algún satélite con una elevación grande y el resto de los satélites cubren todos los acimuts. Se puede demostrar que con 4 satélites el GDOP es mínimo cuando los vectores directores unitarios de los satélites forman un tetraedro regular de volumen máximo (podéis ver la Figura 21), es decir, cuando hay tres satélites en un plano perpendicular a la dirección de un cuarto satélite (el que están en el cénit o cerca, por ejemplo) y separados 120° entre ellos.

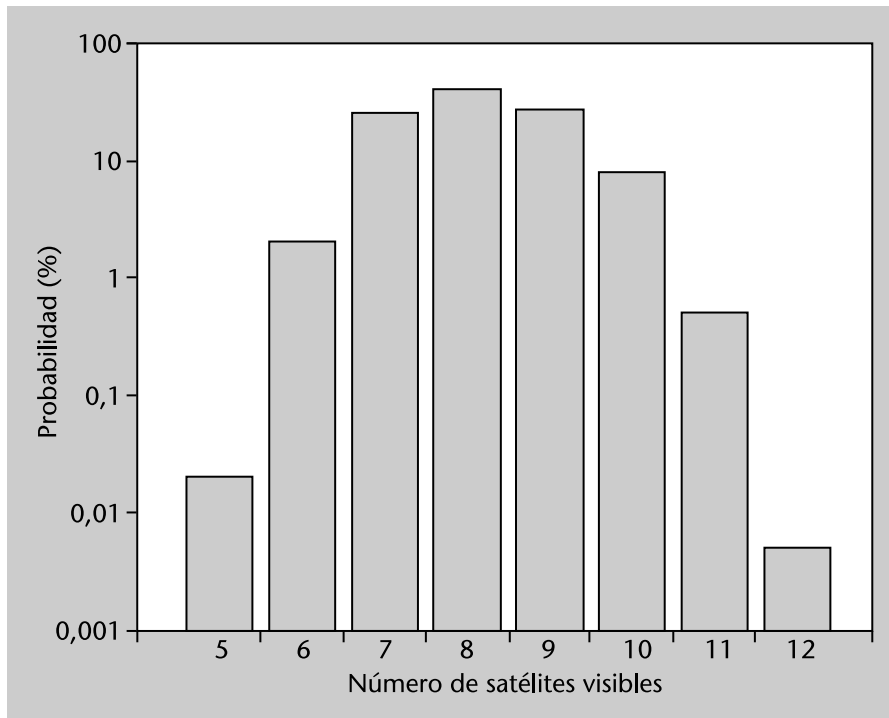
Figura 21. Geometría de una constelación de 4 satélites para posicionamiento en 3D



Con la constelación GPS completa, el valor medio y el percentil del 99% del HDOP son 1 y 1,6, respectivamente; los valores correspondientes para el VDOP son 1,6 y 3,7, respectivamente. Podemos observar la estadística en la figura 22. No obstante, hay que tener en cuenta que en la mayor parte de las aplicaciones los usuarios no gozan de una visibilidad sin obstáculos de todo el cielo, con lo que se obtienen valores de HDOP y VDOP mayores. Los receptores actuales suelen tener un número elevado de canales y entonces siguen todos los satélites visibles. En algunos casos, no obstante, puede que este no sea el caso, como por ejemplo para reducir el consumo de energía del receptor, se puede forzar a que este siga solo un número reducido de satélites. En esta situación, se debe realizar una selección de satélites que seguir. Esto mismo ocurría en receptores más antiguos que tenían un número menor de canales debido a las limitaciones tecnológicas. Por estos motivos, el problema de selección de satélites ha recibido tradicionalmente mucha atención. Aunque existen muchos

critérios en los que basar la selección de un subconjunto de satélites que seguir entre todos los visibles, la selección basada en la minimización del PDOP o del GDOP es muy habitual.

Figura 22. Histograma del número de satélites visibles por encima de una elevación de 5° para un usuario en cualquier punto de la tierra con una visibilidad sin obstrucciones y con una constelación de 24 satélites



Una vez concluido el estudio del efecto de la geometría sobre los errores, vamos a analizar el otro factor que contribuye a ellos. Como hemos explicado antes, se trata de los errores en las propias medidas, que en el modelo sencillo que acabamos de presentar se engloban en el término σ_{URE} . Veremos que este modelo presenta sin embargo algunas limitaciones.

5.4.2. Errores de medida en los observables que afectan al cálculo de la solución de navegación

Los errores en los observables (denominados a veces errores de medida o simplemente errores) se suelen clasificar como *ruido* o *sesgo*.

El **ruido** se refiere generalmente a un error que varía rápidamente y cuyo promedio durante un "corto" intervalo de tiempo es cero, donde "correlación" se define en relación o bien al tiempo de integración o bien al tiempo de suavizado de los observables del receptor.

Lo que aquí denominamos *ruido* es un concepto más general que el efecto del ruido térmico, que comentaremos más adelante. En lugar de *ruido* también se podría haber utilizado el término *variabilidad aleatoria*, pero *ruido* es una terminología más habitual.

El **sesgo** se refiere a un error que persiste más allá de un período corto de tiempo.

Esta distinción entre ruido y sesgo es útil para entender los efectos de los errores de medida.

Dependiendo de su origen, los errores se pueden clasificar en:

- Errores en los valores de los parámetros difundidos por el satélite en su mensaje de navegación, de los cuales el segmento de control es responsable.
- Incertidumbre asociada con el medio de propagación, que afecta al tiempo de propagación de la señal entre el satélite y el receptor.
- Interferencias y ruido captados o generados por el receptor, que afectan a la precisión de las medidas, y efectos locales de la propagación en la vecindad del receptor, refiriéndose fundamentalmente la propagación multicamino debida a las señales reflejadas en la proximidad de la antena receptora.

Vamos a examinar estas fuentes de error a continuación. Nuestro objetivo es entender la naturaleza y magnitud del error introducido por cada fuente, y analizar la eficacia de diferentes métodos de mitigación.

5.4.3. Errores del segmento de control: reloj del satélite y efemérides

Las efemérides y los parámetros del reloj difundidos por el satélite se calculan en el segmento de control a partir de las medidas en las estaciones monitoras. Los valores se obtienen mediante un complicado filtro de Kalman que incluye entre sus estados variables que describen el movimiento del satélite (posición y velocidad) y su reloj (sesgo de fase, frecuencia y de tasa de variación de la frecuencia). A partir de las estimaciones realizadas por el filtro de Kalman, otro modelo predice las efemérides y los parámetros del reloj, que se deben cargar

en los satélites y que estos difundirán mediante el mensaje de navegación a 50 bps (en el caso de GPS). Por lo tanto, hay errores asociados tanto a la estimación de los valores actuales de los parámetros como a la predicción de los valores futuros. El error de predicción crece con la antigüedad de los datos (AoD, *age of data*), definido como el tiempo que ha transcurrido desde la última carga de los parámetros. Obviamente, cuanto más precisos sean los modelos utilizados para estimar y predecir las efemérides y los parámetros del reloj, y más frecuentes sean las recargas de la información en los satélites, menores serán los errores asociados al segmento de control. Una mayor densidad de estaciones monitoras incrementaría la información disponible en el filtro de Kalman y redundaría también en unos errores menores.

El error rms en la distancia atribuido a las efemérides y al reloj estaba limitado a 6 m en la especificación del PPS de GPS. Realmente, las prestaciones de los relojes han sido mejores que sus especificaciones y los errores de predicción se han mantenido bajos gracias a transmisiones frecuentes, limitando los errores originados en el segmento terreno a 3 metros para los satélites de los bloques II/IIA y IIR. Siendo esto así, finalmente se han reducido los errores a 3 m en la especificación del servicio a partir del bloque IIF.

El error en las efemérides se suele descomponer en sus componentes según tres direcciones ortogonales definidas a partir de la órbita del satélite: radial, *along-track* y *cross-track*. En la estimación de una órbita a partir de medidas de distancia, el error en la dirección radial suele ser varias veces menor que en las otras dos direcciones. Esto es un aspecto favorable porque el error en los pseudorrangos es la proyección del error de posición del satélite en la dirección receptor-satélite, y depende fundamentalmente de la componente radial del vector, siendo la contribución de las componentes *along-track* y *cross-track* pequeña.

El error en distancia causado por los errores en el reloj y en las efemérides se define como la raíz de la suma de los errores de reloj y del error de la componente radial de las efemérides al cuadrado. El segmento de control monitoriza en tiempo real el valor aproximado de este error con una precisión de 1 m rms. Con una recarga diaria del mensaje de navegación (que es la periodicidad habitual), el error rms en las predicciones es aproximadamente de 1,5 metros debido al reloj y de la misma magnitud debido a las efemérides. Si el segmento de control considera que el error supera un determinado umbral, programa una recarga de contingencia.

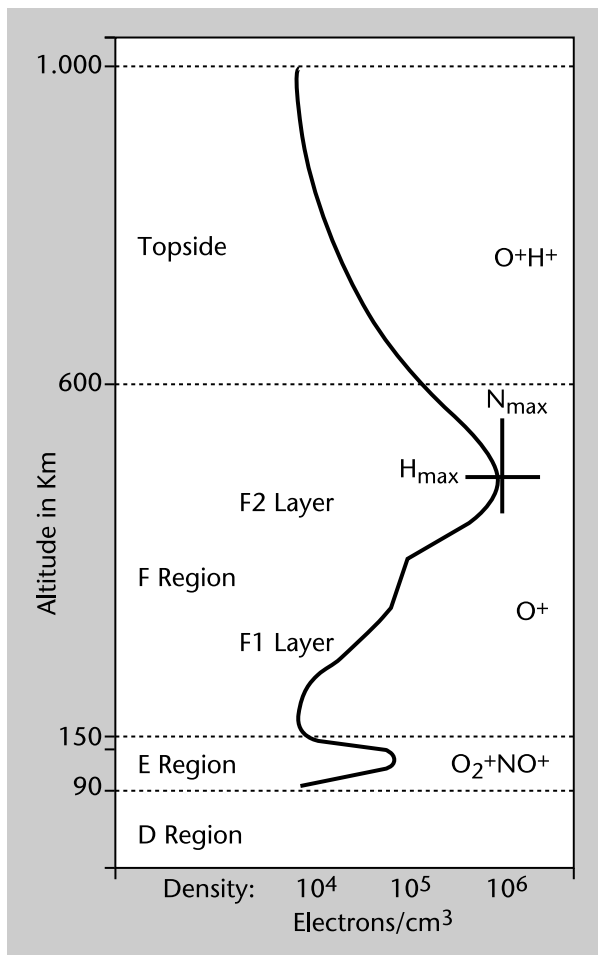
Los satélites de los bloques IIR y IIF disponen de un modo de funcionamiento denominado *Autonav* que les permite mantener bajos los errores de reloj y efemérides aunque se haya perdido el contacto con el segmento de control (en el bloque IIF, el error se mantiene por debajo de 3 m –sin contar el movimiento polar y cambios en la escala de tiempos UT1– aun estando 60 días sin

contacto con el control). En el modo *Autonav* se utilizan enlaces entre enlaces para medir la distancia entre ellos. No obstante, parece que dicho modo no está operativo habitualmente.

5.4.4. Efectos de la propagación: la ionosfera

La ionosfera es una capa de la atmósfera formada por gases ionizados (por ejemplo, electrones e iones libres) situada a una altitud entre 50 km y 1.000 km. La radiación solar es la causa de la ionización y, por lo tanto, el estado de la ionosfera depende principalmente de la intensidad de la actividad solar. La ionosfera se compone de varias capas (denominadas D, E, F1 y F2) a diferentes alturas, cada una con diferentes tasas de producción y pérdida de electrones libres (podéis ver la figura 23). La máxima densidad de electrones están en una franja entre 250 y 400 km de altura (se trata de la capa F2).

Figura 23. Capas de la ionosfera y valores típicos del contenido de electrones



Fuente: extraído de Space Environment TOPICS SE-14, Space Environment Center 325 Broadway, Boulder, CO 80303-3326

Las características físicas de la ionosfera o, lo que es lo mismo, la densidad de electrones libres varían entre uno y dos órdenes de magnitud entre el día y la noche, siendo evidentemente más alta durante el día. Puede haber también variaciones importantes entre diferentes días, estaciones o años, en función por ejemplo de la fase del ciclo solar (de periodicidad igual a 11 años) en la

que nos encontremos. Existe también una acusada dependencia con la latitud: la ionosfera presenta una actividad mucho más intensa en las zonas ecuatoriales y en los polos magnéticos, y es más moderada en las zonas templadas. La región con el mayor retardo ionosférico es la franja de $\pm 20^\circ$ alrededor del ecuador magnético.

El efecto de la ionosfera en la propagación de las señales electromagnéticas se caracteriza cuantitativamente a partir del TEC (*total electron content*), que es número de electrones en un cilindro ficticio de sección recta igual a 1 m^2 y que une el receptor con el satélite:

$$\text{TEC} = \int_{S \rightarrow R} n_e(l) dl \quad (76)$$

donde $n_e(l)$ es la densidad de electrones a lo largo del camino de la señal y donde la integración se realiza para la línea que une al satélite y el receptor. Está claro que el camino de propagación por el interior de la ionosfera es más corto cuando el satélite se encuentra en el cenit y, por tanto, el valor del TEC en la dirección vertical (TECV) es más bajo que el valor para cualquier otra elevación. Realmente, la ionosfera se suele caracterizar con el valor TECV, ya que es independiente de la elevación, y es fácil calcular, al menos de manera aproximada, el valor del TEC para una determinada elevación a partir del valor del TECV. Tanto es así, que es habitual referirse al valor de TECV como simplemente TEC. La unidad de medida del TEC (o del TECV) es el TECU (TEC units), definido como 10^{16} electrones/ m^2 . TECV típicamente varía entre 1 y 150 TECU. Para un lugar e instante dado, TECV puede cambiar un 20-25% con respecto a su media mensual, lo que supone una variación muy importante. La mayoría de los modelos actuales de la ionosfera no proporcionan una representación suficientemente realista de la variación diaria del TEC como para servir de base para la navegación de precisión.

Refracción y dispersión

La presencia de un medio diferente al vacío cambia la velocidad (magnitud y dirección) de propagación de las ondas electromagnéticas. Este fenómeno se denomina refracción. El cambio en la velocidad de propagación modifica el tiempo de tránsito de la señal por el medio, que es el principal observable en GPS. El índice de refracción (n) de un medio se define como el cociente entre la velocidad de propagación de la onda en el vacío (c) y en el medio (v):

$$n = \frac{c}{v} \quad (77)$$

Simplemente cabe mencionar aquí que en la expresión anterior la velocidad se refiere a lo que se suele denominar velocidad de fase. Más adelante se entenderá por qué hacemos esta puntualización. Muchos medios, y en particular la ionosfera y la troposfera, no son uniformes en cuanto a su composición y el índice de refracción cambia a lo largo del camino que sigue la onda. Esto provoca que la velocidad de propagación también cambie y, entonces, el tiempo de tránsito viene dado por:

$$\tau_\phi = \frac{1}{c} \int_{S \rightarrow R} n(l) dl \quad (78)$$

donde la integral se realiza a lo largo del camino que va del satélite al receptor y $n(l)$ es el índice de refracción en cada punto de este camino. La variación del índice de refracción también causa una curvatura de los rayos (ley de Snell). No obstante, el impacto de esto en GNSS en el tiempo de propagación es despreciable.

Si el índice de refracción también depende de la frecuencia, se dice que el medio es dispersivo. Por ejemplo, es bien conocido que un prisma de cristal o gotas de agua en suspensión en el aire son medios dispersivos, y por ello crean un arcoíris cuando la luz los cruza. Consideremos ahora que transmitimos una señal modulada en amplitud en un medio dispersivo:

$$s(t) = \cos(2\pi f_m t) \cos(2\pi f_c t) \quad (79)$$

donde f_m es la frecuencia moduladora, f_c es la frecuencia portadora y se cumple que $f_m \ll f_c$. En el contexto de GPS, f_m podría corresponder a la velocidad de chip (por ejemplo, ~ 1 MHz) y f_c a la portadora de la banda L1 (por ejemplo, $\sim 1,5$ GHz). Recordemos que el número de onda se define como

$$k = \frac{2\pi}{\lambda} = \frac{2\pi}{v|f} = \frac{2\pi}{c|(nf)} = nf \frac{2\pi}{c} \quad (80)$$

donde hemos utilizado la expresión 77, y que la longitud de onda es $\lambda = \frac{v}{f}$.

La señal (se puede ver alternativamente como la suma de dos portadoras ligeramente desplazadas:

$$s(t) = \frac{1}{2} [\cos(2\pi(f_c + f_m)t) + \cos(2\pi(f_c - f_m)t)] \quad (81)$$

que se propagan también a velocidades ligeramente diferentes, y que vienen dadas por sus respectivos números de onda (los subíndices + y - hacen referencia a las frecuencias suma y diferencia respectivamente):

$$\begin{aligned} k_+ &\approx k_c + \frac{dk}{df} f_m \\ k_- &\approx k_c - \frac{dk}{df} f_m \end{aligned} \quad (82)$$

En la expresión anterior se ha linealizado la variación del número de onda con la frecuencia. El valor del número de onda a la frecuencia portadora se ha denominado k_c . Si el medio no fuera dispersivo, la variación sería exactamente lineal con la frecuencia, es decir: $dk/df = n2\pi/c$.

Si la distancia geométrica entre el transmisor y el receptor es r , la señal recibida es (ignorando por sencillez cambios en la amplitud):

$$\begin{aligned} r(t) &= \frac{1}{2} [\cos(2\pi(f_c + f_m)t - k_+ r) + \cos(2\pi(f_c - f_m)t - k_- r)] \\ &= \frac{1}{2} [\cos(2\pi(f_c + f_m)t - k_c r - \frac{dk}{df} f_m r) + \cos(2\pi(f_c - f_m)t - k_c r + \frac{dk}{df} f_m r)] \\ &= \cos(2\pi f_m t - \frac{dk}{df} f_m r) \cos(2\pi f_c t - k_c r) \\ &= \cos(2\pi f_m (t - \frac{1}{2\pi} \frac{dk}{df} r)) \cos(2\pi f_c (t - \frac{1}{2\pi} \frac{k_c}{f_c} r)) \end{aligned} \quad (83)$$

Observamos que los retardos que experimentan la envolvente (o moduladora) y la portadora son diferentes, lo que indica que viajan a dos velocidades diferentes (podéis ver la figura 24). Identificando los términos adecuados en la expresión anterior, se deduce que la velocidad de la portadora es

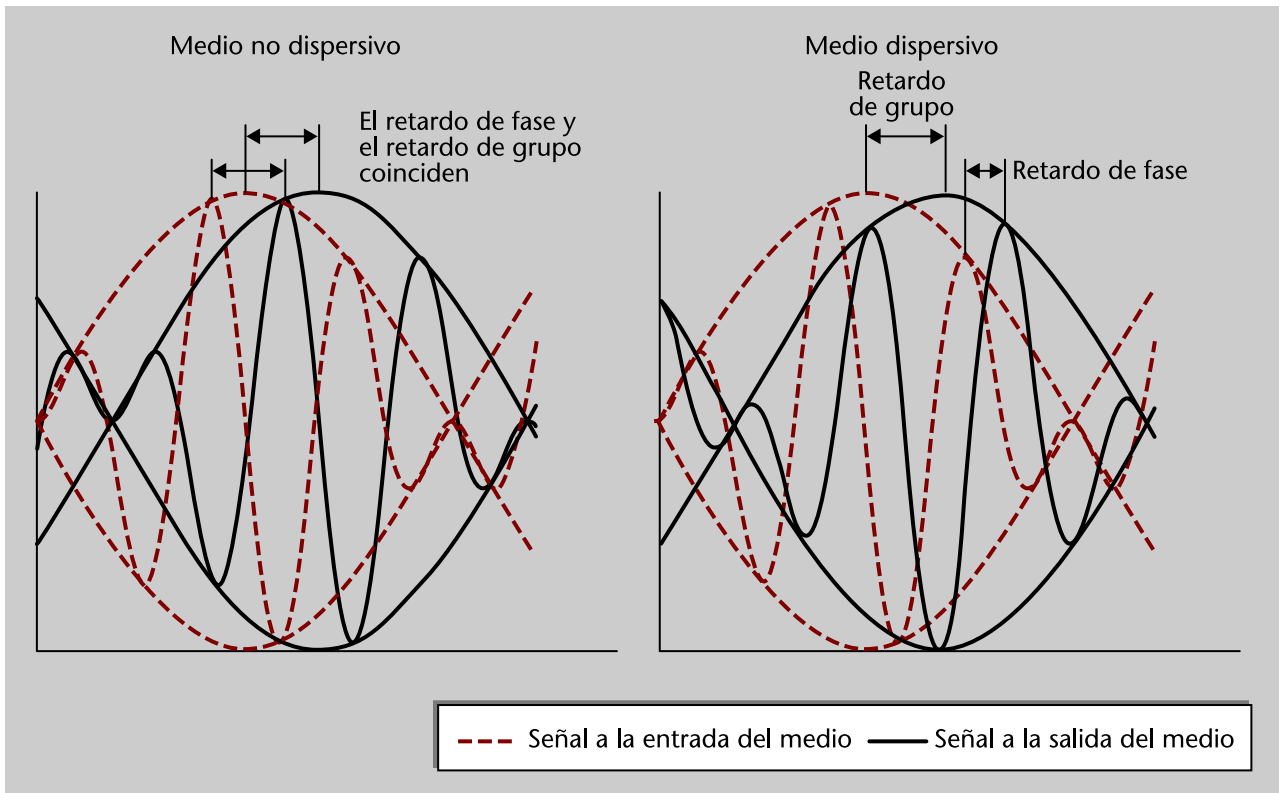
$$v_p = 2\pi \frac{f_c}{k_c} \quad (84)$$

que se suele denominar velocidad de fase y es la velocidad que ya habíamos presentado anteriormente al introducir el índice de reflexión en la expresión 77, mientras que la velocidad de propagación de la envolvente es

$$v_g = 2\pi \frac{df_c}{dk_c} \quad (85)$$

que recibe el nombre de velocidad de grupo. Los respectivos retardos se denominan retardo de fase y retardo de grupo. La velocidad de grupo es realmente la velocidad a la que viaja la energía de la señal y, por lo tanto, siempre es menor o igual a la velocidad de la luz en el vacío. Por otro lado, la velocidad de fase es una velocidad aparente y puede ser mayor que c . Es inmediato comprobar que si el índice de refracción no depende de la frecuencia, entonces la velocidad de fase y de grupo son iguales.

Figura 24. Propagación de una señal modulada en un medio no dispersivo y en uno dispersivo



Por coherencia en la notación, el índice de refracción que hemos definido en la ecuación 77 se puede llamar también índice de refracción de fase (n_p) y se define de modo similar un nuevo parámetro, denominado índice de refracción de grupo, como

$$n_g = \frac{c}{v_g} \quad (86)$$

La ionosfera, como cualquier gas ionizado, es un medio dispersivo para las ondas electromagnéticas. El valor de su índice de refracción es (despreciando términos de orden superior en f):

$$n_p \approx 1 - \frac{40,3n_e}{f^2} \quad (87)$$

El retardo de fase en exceso (por ejemplo, en exceso con respecto a la propagación en el espacio vacío) es

$$\Delta\tau_\phi = \frac{1}{c} \int_{S \rightarrow R} (n_p(l) - 1) dl = -\frac{1}{c} \int_{S \rightarrow R} \frac{40,3 n_e}{f^2} dl = -\frac{40,3 \text{ TEC}}{c f^2} \quad (88)$$

y expresado en metros

$$I_\phi = c\Delta\tau_\phi = -\frac{40,3 \text{ TEC}}{f^2} \quad (89)$$

que es la magnitud que habíamos utilizado en la expresión 47. Estos retardos son negativos (o sea, son adelantos), que es coherente con el hecho de que el índice de refracción es menor que 1 y, por lo tanto, la velocidad de fase es mayor que en el vacío.

Podemos comprobar que el índice de refracción de grupo y el retardo de grupo en metros son

$$\begin{aligned} n_g &= 1 + \frac{40,3n_e}{f^2} \\ I_\rho &= -I_\phi = \frac{40,3\text{TEC}}{f^2} \end{aligned} \quad (90)$$

Un cambio del TEC en 1 TECU corresponde aproximadamente a un cambio de 16 cm en el retardo ionosférico a la frecuencia L1.

Para calcular el retardo a una determinada elevación a partir del retardo cenital se puede utilizar el denominado modelo de capa fina (*thin shell model*), que consiste en suponer que la ionosfera se concentra en una capa muy delgada a una determinada altura h_I de la tierra (entre 300 y 400 km). De esta manera, el cálculo de la longitud del camino de propagación dentro de la ionosfera se reduce únicamente a una cuestión geométrica, representada cuantitativamente por el factor de oblicuidad (OF), evitando tener que realizar la integral de la densidad de electrones a lo largo del camino en cuestión. El factor de oblicuidad es la constante de proporcionalidad entre el TEC para una determinada elevación (el) y el TECV, o, lo que es lo mismo, entre el retardo ionosférico a una elevación y en dirección cenital:

$$\begin{aligned} \text{TEC}(el) &= \text{OF}(el) \cdot \text{TECV} \\ I_\rho(el) &= \text{OF}(el) \cdot I_\rho(90^\circ) \end{aligned} \quad (91)$$

Su expresión más habitual es

$$\text{OF}(el) = \left(1 - \left(\frac{R_E \cos(el)}{R_E + h_I} \right)^2 \right)^{-1/2} \quad (92)$$

El retardo ionosférico cenital típicamente varía a latitudes medias de 1-3 m por la noche a 5-15 m al principio de la tarde. Valores tan altos como 36 m han sido observados cerca del ecuador en el período más activo del ciclo so-

lar. La magnitud de estos valores indica que la contribución de la ionosfera es significativa en los pseudorrangos y no se puede obviar si se desea obtener un error de posicionamiento por debajo de una decena de metros. Una opción para eliminar los errores ionosféricos es utilizar algún modelo que proporciona una estimación del retardo ionosférico a partir de ciertos parámetros. El modelo más habitual es el de Klobuchar, que es el que utiliza el propio sistema GPS. Este modelo calcula el retardo cenital a la frecuencia L1 como una constante durante la noche y como un semiarco de coseno en función del tiempo durante el día:

$$\hat{I}_{\rho}(t, 90^{\circ}) = \begin{cases} cA_1 + cA_2 \cos\left(\frac{2\pi(t-A_3)}{A_4}\right), & \text{si } |t - A_3| < \frac{A_4}{4} \\ cA_1, & \text{en cas contrari} \end{cases} \quad (93)$$

Los parámetros de A_1 y A_3 toman valores predeterminados, mientras que los valores de A_2 y A_4 se transmiten de manera indirecta en el mensaje de navegación de GPS (concretamente en la subtrama 4) a través de los 8 coeficientes de dos polinomios (4 coeficientes por polinomio) que se utilizan para aproximar estos valores. El factor de oblicuidad que se utiliza en el modelo de Klobuchar es

$$\text{OF}_{kl}(el) = 1 + 16\left(0,53 - \frac{el}{\pi}\right)^3 \quad (94)$$

El modelo de Klobuchar es capaz de corregir aproximadamente el 50% (rms) de los errores ionosféricos. Esto implica que a latitudes medias, el error ionosférico residual en la dirección del cenit puede llegar a 10 m durante el día y ser mucho mayor durante períodos de alta actividad solar. Por lo tanto, no se puede decir que el modelo de Klobuchar sea la solución adecuada cuando se desea un posicionamiento preciso. Existen otros modelos mejorados, como el modelo de NeQuick, que corrige aproximadamente un 75% de los errores. Este es el modelo propuesto para el sistema Galileo, que en contrapartida incrementa los requisitos de memoria del receptor porque es necesario guardar un mapa de información ionosférica.

La solución más efectiva para eliminar los errores ionosféricos se basa en utilizar observables a dos frecuencias (por ejemplo, L1 y L2, pero se puede utilizar cualquier otro par de frecuencias). Vamos a reescribir la ecuación 43 haciendo explícita la dependencia con la frecuencia

$$\begin{aligned} P_{L1} &= r + c(\delta t_u - \delta t^s) + \frac{40,3\text{TEC}}{f_{L1}^2} + T_{\rho} + \varepsilon_{\rho,L1} \\ P_{L2} &= r + c(\delta t_u - \delta t^s) + \frac{40,3\text{TEC}}{f_{L2}^2} + T_{\rho} + \varepsilon_{\rho,L2} \end{aligned} \quad (95)$$

La contribución de la ionosfera no se puede eliminar directamente porque el valor del TEC es desconocido, pero el punto clave aquí es aprovechar la dependencia del retardo ionosférico con $1/f^2$. Podemos realizar la siguiente combinación lineal de los pseudorrangos

$$P_{IF} = \frac{f_{L1}^2 P_{L1} - f_{L2}^2 P_{L2}}{f_{L1}^2 - f_{L2}^2} = r + c(\delta t_u - \delta t^s) + T_\rho + \frac{f_{L1}^2 \varepsilon_{\rho,L1} - f_{L2}^2 \varepsilon_{\rho,L2}}{f_{L1}^2 - f_{L2}^2} \quad (96)$$

que se denomina combinación libre de ionosfera. Hemos construido un nuevo pseudorango donde el error ionosférico ha desaparecido completamente. Esto es cierto al menos para la parte de la contribución de la ionosfera que depende de $1/f^2$, que es aproximadamente un 99,9% de la contribución total. Existen términos de orden superior que no desaparecen al hacer la combinación en la expresión 96, pero estos términos suelen tomar valores muy pequeños y solo es necesario considerarlos en aplicaciones muy concretas de alta precisión. El precio que se paga por eliminar el sesgo causado por la ionosfera es un incremento del ruido. Para determinar únicamente el orden de magnitud de este incremento, supongamos que $\varepsilon_{\rho,L1}$ y $\varepsilon_{\rho,L2}$ tuvieran la misma varianza y fueran incorrelados, entonces la desviación típica del último término de la expresión 96 sería un factor

$$\frac{\sqrt{f_{L1}^4 + f_{L2}^4}}{f_{L1}^2 - f_{L2}^2} \approx 2,98 \quad (97)$$

mayor que la de los términos correspondientes en la ecuación 95.

Existen aplicaciones científicas donde lo interesante es estimar el retardo ionosférico (y, a partir de aquí, el TEC), en lugar de cancelarlo. Esto también es posible cuando se dispone de observables a dos frecuencias. Para ello se utiliza la denominada combinación ionosférica

$$P_I = P_{L2} - P_{L1} = 40,3 \cdot \text{TEC} \cdot \left(\frac{1}{f_{L2}^2} - \frac{1}{f_{L1}^2} \right) + \varepsilon_{\rho,L2} - \varepsilon_{\rho,L1} \quad (98)$$

La combinación libre de ionosfera y la combinación ionosférica que hemos definido para los pseudorrangos también se pueden aplicar a las fases de portadora. Para ello, definimos el observable de fase en metros (L) a partir del observable de fase en ciclos (Φ) que presentamos en la expresión 48: $L = \lambda \times \Phi$. Las combinaciones para las fases son:

$$L_{IF} = \frac{f_{L1}^2 L_{L1} - f_{L2}^2 L_{L2}}{f_{L1}^2 - f_{L2}^2} = r + c(\delta t_u - \delta t^s) + cT_\rho + c \frac{f_{L1} N_{L1} - f_{L2} N_{L2}}{f_{L1}^2 - f_{L2}^2} + c \frac{f_{L1} \varepsilon_{\phi,L1} - f_{L2} \varepsilon_{\phi,L2}}{f_{L1}^2 - f_{L2}^2} \quad (99)$$

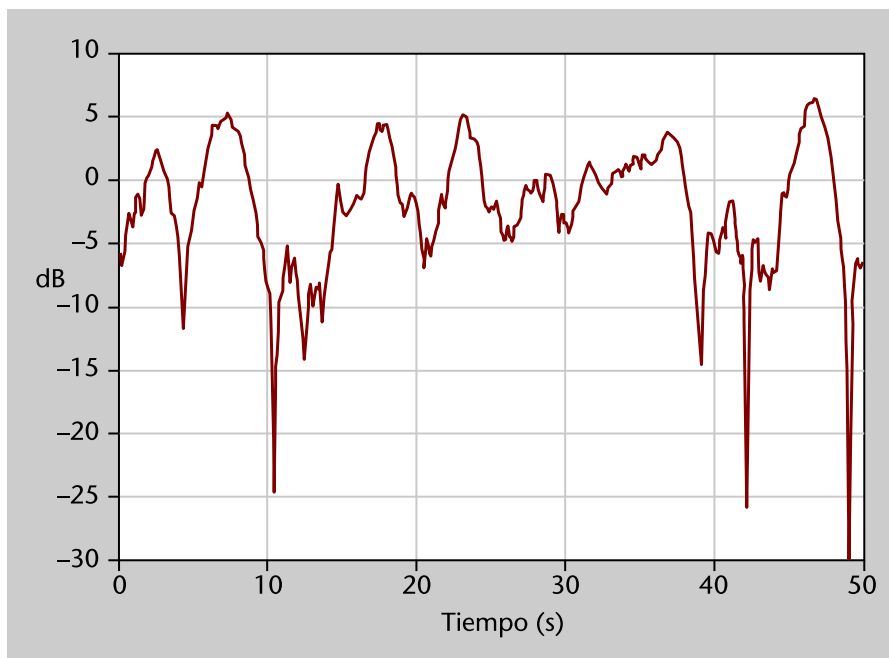
$$L_I = L_{L2} - L_{L1} = 40,3 \cdot \text{TEC} \cdot \left(\frac{1}{f_{L2}^2} - \frac{1}{f_{L1}^2} \right) + \lambda_{L2} N_{L2} - \lambda_{L1} N_{L1} + \lambda_{L2} \epsilon_{\phi L2} - \lambda_{L1} \epsilon_{\phi L1} \quad (100)$$

En L_I se ha invertido el orden de los términos en la resta para que la contribución de la ionosfera aparezca con el mismo signo que en P_I . El comportamiento de estas combinaciones con respecto a la ionosfera es el mismo que el comentado para los pseudorrangos (por ejemplo, en una de ellas se elimina la ionosfera y la otra sirve para estimarla), pero son ambiguas. Un hecho destacable es que las ambigüedades en estos nuevos observables dejan de ser números enteros de ciclos, como ocurría en los observables de fase originales.

Al realizar combinaciones de observables medidos a varias frecuencias hay que considerar en los modelos un efecto que hasta ahora no habíamos incluido, se trata del *interfrequency bias* (IFB). El receptor introduce retardos/desfases diferentes a las señales que llegan a diferentes frecuencias. La diferencia de dichos retardos/desfases es el IFB.

Además del retardo de grupo y de fase, que son constantes en intervalos cortos de tiempo, la ionosfera también causa unas fluctuaciones rápidas en la fase y en la amplitud de la señal, que se denomina centelleo ionosférico (*scintillation*, en inglés). Podemos observar unos ejemplos de esto en la figura 25. Se debe a las irregularidades en la distribución de electrones a lo largo del camino de propagación, que a su vez vienen causadas por llamaradas y tormentas magnéticas originadas en el Sol.

Figura 25. Desvanecimientos en la amplitud de la señal



Fuente: extraído de IEEA - GISM Global Ionospheric Scintillation Model RP_257_08/01/2008

5.4.5. Efectos de la propagación: la troposfera

Las señales GPS sufren también el efecto de la refracción en las capas bajas de la atmósfera debido a la presencia de gases secos (principalmente N_2 y O_2) y vapor de agua. Este último se encuentra a alturas inferiores a 12 km respecto al nivel del mar, y la mayor concentración está en los 4 km más bajos. Los gases secos se encuentran en concentraciones decrecientes gradualmente hasta alturas de cientos de kilómetros. Las tres cuartas partes de los gases secos están en la troposfera, que es la parte más baja de la ionosfera y que tiene una altura aproximada de 16 km en el ecuador y 9 km en los polos. El efecto combinado de los gases secos y el vapor de agua recibe la denominación de efecto troposférico.

A diferencia de la ionosfera, la troposfera no es dispersiva a las frecuencias de GNSS (es decir, el índice de refracción no depende de la frecuencia). El índice de refracción de los gases es ligeramente superior a la unidad: $n \approx 1,003$ al nivel del mar, y está todavía mucho más próximo a uno en la zona superior de la troposfera. La velocidad de propagación (tanto la de fase como la de grupo, que en este medio son iguales) es ligeramente inferior a la del espacio libre y, en consecuencia, la distancia aparente al satélite es superior a la geométrica en 2,5-25 m dependiendo de la elevación del satélite. El efecto de la troposfera es el mismo en los pseudorrangos y en las fases (lo que significa, en la nomenclatura utilizada en los apartados anteriores, que $T_\rho = T_\phi$), y también el mismo en las diferentes frecuencias. Este retardo común a todas las medidas que se realizan a un satélite no puede ser estimado a partir de las propias medidas, y es necesario recurrir a modelos para corregirlo.

La magnitud del retardo troposférico experimentado por una señal depende del índice de refracción del aire a lo largo del camino de propagación. El índice de refracción de una masa de aire es función de su densidad, que se puede expresar como la suma de las densidades del aire seco y del vapor de agua. Estas densidades son a su vez función de la presión de los gases y del vapor de agua, y de la temperatura. Los gases secos y el vapor de agua afectan de modo diferente a la propagación de una señal de RF y se modelan por separado. La composición de los primeros varía con la latitud, estación y altitud, pero es moderadamente constante a lo largo del tiempo. Sin embargo, el contenido de vapor de agua de la troposfera es mucho más difícil de modelar, ya que cambia con la situación meteorológica local y puede cambiar rápidamente. Afortunadamente, la mayor parte del retardo troposférico (aproximadamente el 90%) se debe a las componentes secas, que son más predecibles.

Resulta apropiado definir la refractividad como $N = (n - 1) \times 10^6$, que se puede expresar como la suma de las componentes debidas a los gases secos y al vapor de agua en la atmósfera

$$N_t = N_{t,d} + N_{t,w} \quad (101)$$

donde $N_{t,d}$ y $N_{t,w}$ se denominan como refractividad seca y húmeda, respectivamente.

Siguiendo los mismos pasos que en el caso de la ionosfera, el exceso de retardo (en metros) introducido por la troposfera es

$$T_\rho = 10^{-6} \int_{S \rightarrow R} N_t(l) dl = 10^{-6} \int_{S \rightarrow R} (N_{t,d}(l) + N_{t,w}(l)) dl = T_{\rho,d} + T_{\rho,w} \quad (102)$$

donde $T_{\rho,d}$ y $T_{\rho,w}$ se denominan retardo seco y húmedo, respectivamente. También al igual que en la ionosfera, despreciamos el efecto de la curvatura de los rayos debido a la variación del índice de refracción y nos centramos en el retardo ocasionado por la reducción en la velocidad de propagación.

Los modelos de la troposfera intentan estimar las refractividades seca y húmeda a partir de varios parámetros físicos, como la presión total atmosférica (P), la presión parcial del vapor de agua (e) y la temperatura (T_a). Expresiones aproximadas con unos coeficientes determinados empíricamente son:

$$N_{t,d} = 77,64 \frac{P}{T_a}, \quad N_{t,w} = 3,73 \cdot 10^5 \frac{e}{T_a^2} \quad (103)$$

donde las presiones están en unidades de milibares y la temperatura en kelvin. El conocimiento de estas magnitudes físicas a lo largo del camino de propagación permite obtener los retardos troposféricos de forma precisa a partir de la expresión 102. Estas magnitudes se suelen determinar para la dirección vertical a partir de medidas realizadas con globos sonda. El problema es que los usuarios GPS rara vez tiene acceso a estos datos meteorológicos detallados (o no están dispuestos a pagar por ellos). Un procedimiento más sencillo y barato consiste en medir P , e y T_a en la posición del receptor y extrapolar a partir de modelos su evolución con la altura. No obstante, esto no deja de ser también complejo y solo es utilizado por usuarios que necesitan una gran precisión (un ejemplo de ello son las estaciones dedicadas a medir el movimientos de las placas tectónicas). Para un usuario convencional, realizar medidas meteorológicas es poco práctico, así que en la mayoría de las situaciones se recurre a estimar el retardo troposférico a partir de condiciones meteorológicas promedio proporcionadas por un modelo de *atmósfera estándar* para un día dado y unos valores de latitud y altitud.

Un modelo de *atmósfera estándar* es una especificación del perfil promedio de presión, humedad y temperatura en función de la altitud, así como las variaciones con las estaciones y la latitud.

Supongamos que disponemos de valores de la temperatura y las dos presiones en la posición de la antena obtenidas bien a través de medidas o bien de modelos de la atmósfera estándar. Seguidamente hay que encontrar una manera de extrapolar estos valores o los correspondientes índices de refracción a lo largo de la atmósfera y realizar la integración en la expresión 102. Existen dos alternativas principales, que se conocen como el modelo de Saastamoinen y el modelo de Hopfield.

El **modelo de Saastamoinen** se basa en unos determinados gradientes de presión y temperatura según la altitud y concluye que los retardos cenitales son

$$\begin{aligned} T_{\rho,d}(90^\circ) &= 0,002277 \left(1 + 0,0026 \cos(2\phi_{\text{lat}}) + 0,00028H \right) P \\ T_{\rho,w}(90^\circ) &= 0,002277 \left(\frac{1,255}{T_a} + 0,05 \right) e \end{aligned} \quad (104)$$

donde ϕ_{lat} es la latitud y H la altura ortométrica en km.

El **modelo de Hopfield** propone las siguientes evoluciones de los índices de refracción

$$\begin{aligned} N_{t,d} &= N_{t,d0} \left(1 - \frac{h}{h_d} \right)^4, \quad \text{para } h \leq h_d \\ N_{t,w} &= N_{t,w0} \left(1 - \frac{h}{h_w} \right)^4, \quad \text{para } h \leq h_w \end{aligned} \quad (105)$$

donde h es la altura sobre la antena, $h_d = 43$ km y $h_w = 12$ km. Se supone que los índices de refracción son cero a partir de estas alturas. Los valores de $N_{t,d0}$ y $N_{t,w0}$ se obtienen aplicando la expresión 103 con los valores de presión y temperatura disponibles. Finalmente, los retardos cenitales son

$$\begin{aligned} T_{\rho,d}(90^\circ) &= 77,64 \cdot 10^{-6} \frac{P}{T_a} \frac{h_d}{5} \\ T_{\rho,w}(90^\circ) &= 0,373 \frac{e}{T_a^2} \frac{h_w}{5} \end{aligned} \quad (106)$$

En cualquier caso, una vez encontrados los valores se deben encontrar los que corresponden a una determinada elevación del satélite. Para ello se utilizan, al igual que en la ionosfera, los factores de oblicuidad, aunque en el contexto de la troposfera suelen denominarse funciones de mapeo (*mapping functions*). Los retardos troposféricos son entonces

$$\begin{aligned} T_{\rho,d}(el) &= T_{\rho,d}(90^\circ) m_d(el) \\ T_{\rho,w}(el) &= T_{\rho,w}(90^\circ) m_w(el) \end{aligned} \quad (107)$$

donde se utilizan en general funciones de mapeo diferentes para las componentes secas y húmedas, debido a que la distribución de los gases y del vapor de agua es diferente con la altura (independientemente del perfil concreto, el vapor de agua ocupa una región de menor altura que los gases).

Existe una gran variedad de funciones de mapeo. La más sencilla es $1/\sin(el)$, que ignora la curvatura de la tierra y, por lo tanto, no funciona bien para elevaciones bajas ($el < 15^\circ$). Otra opción algo más precisa es utilizar una función del mismo tipo que el factor de oblicuidad empleado en la ionosfera (ecuación 92):

$$m(el) = \left(1 - \left(\frac{\cos(el)}{1.001} \right)^2 \right)^{-1/2} \quad (108)$$

que se conoce como la función de Black&Eisner. A día de hoy la complejidad de cálculo no es una gran limitación y es habitual utilizar funciones de mapeo basadas en un desarrollo truncado de funciones continuas. Si se trunca a partir del tercer término, el resultado es:

$$m_i(el) = \frac{1 + \frac{a_i}{b_i}}{\sin(el) + \frac{a_i}{\sin(el) + \frac{b_i}{\sin(el) + c_i}}} \quad (109)$$

donde los coeficientes a_i , b_i y c_i se calculan por separado para la componente seca y para el vapor de agua, bien de forma empírica o bien a partir de las variables físicas: latitud, altura, temperatura y presión en la antena, y día del año.

La función de mapeo de la troposfera aumenta rápidamente al bajar la elevación del satélite, y son mucho más grandes de los factores de oblicuidad de la ionosfera a bajas elevaciones¹⁶. Valores típicos de la función de mapeo son dos a 30° , cuatro a 15° , seis a 10° y diez a 5° . Un pequeño error de 5-10 cm en

⁽¹⁶⁾ Esto es fácil de justificar cualitativamente observando la geometría del problema y los caminos que sigue la señal por dentro de la troposfera y de la ionosfera

la estimación del retardo troposférico vertical crece hasta 0,5-1 m si el satélite está a 5° de elevación, que no es vital en navegación a escala de metros pero sí en navegación centimétrica.

El valor de $T_{\rho,d}(90^\circ)$ es de 2,3-2,6 m a nivel del mar, y se reduce con la altitud, siendo de 2 m aproximadamente a 2 km de altitud y de 1 m en las cumbres del Himalaya. Su valor se puede predecir con gran precisión (errores de pocos milímetros) si se dispone de medidas precisas de la presión en la superficie. El valor de $T_{\rho,w}(90^\circ)$ varía entre casi cero y 80 cm. Este valor es más difícil de calcular que el retardo seco porque la distribución del vapor de agua a lo largo del camino de la señal es altamente variable. Si se dispone de medidas meteorológicas de superficie, el error es de 1-2 cm. Si se utilizan valores meteorológicos medios en lugar de las medidas reales, el error en la estimación del error troposférico vertical total puede ser de 5-10 cm.

El enfoque que aquí hemos adoptado es el de un usuario que desea calcular la posición con GPS y, por tanto, debe utilizar modelos y/o medidas para eliminar el error ionosférico de los observables GPS. Pero también sería posible adoptar el enfoque de la observación de la atmósfera, donde la magnitud de interés es justamente el retardo troposférico (especialmente el debido al vapor de agua) para extraer a partir de él las variables meteorológicas. Se podría utilizar una red de estaciones GPS en posiciones conocidas, que proporcionarían observables a partir de los cuales se estimaría el efecto de la troposfera. Esto sustituiría el costoso lanzamiento de sondas que se realiza hoy en día para monitorizar la atmósfera (solo en Estados Unidos el U. S. Weather Service lanza más de 160 sondas cada día). No obstante, eliminar de los observables el resto de los efectos (efemérides, ionosfera, multicamino, etc.) con la precisión adecuada para estimar en tiempo real el retardo número sigue siendo un reto técnico y científico.

La tabla 2 resume las principales características de la troposfera y de la ionosfera que hemos explicado en estos subapartados.

Tabla 2. Resumen de las características de la ionosfera y de la troposfera para la propagación de señales GNSS

| | Ionosfera | Troposfera |
|--|---|---|
| Variabilidad | Alta (depende de la hora del día, de la estación, de los ciclos solares y de la actividad solar, de la latitud) | Baja (los cambios significativos solo se producen en la componente del vapor de agua, que representa menos del 10% del total) |
| Retardo cenital | Metros a decenas de metros | 2,3-3,4 m a nivel del mar |
| Factor de oblicuidad (= 1 en la dirección del cénit) | 1,8 a 30° de elevación, 2,5 a 15°, 3 a 5° | 2 a 30° de elevación, 4 a 15°, 10 a 5° |
| Error de modelado para el retardo cenital | 1-10 m, o más en situaciones de alta actividad | 5-10 cm (sin utilizar datos meteorológicos) |

$T_{\rho,w}(90^\circ)$

El valor de $T_{\rho,w}(90^\circ)$ varía bastante. Es de milímetros en los polos, de pocos centímetros en los desiertos, y de decenas de centímetros en las áreas tropicales.

| | Ionosfera | Troposfera |
|--------------------------------------|--|-------------------|
| Dispersivo a las frecuencias de GNSS | Sí. El 99,9% del efecto se puede eliminar con medidas a dos frecuencias. | No |

5.4.6. Efectos locales en el receptor: el ruido

Los observables de código y de fase de portadora están afectados por las fluctuaciones aleatorias que contiene la señal, que se suelen englobar en el término genérico de ruido del receptor. Este es un término que incluye la radiación de RF captada por la antena diferente a la de la señal de interés, el ruido generado en los componentes (antena, amplificadores, cables, filtros, mezcladores, etc.) que forman el receptor, así como el ruido de cuantificación en la conversión analógico-digital. El ruido debido a los componentes tiene fundamentalmente origen térmico, pero también existen otros tipos de ruido en los componentes activos, como por ejemplo el ruido *flicker* en los osciladores. La radiación de RF contiene la radiación de fondo captada por la antena (que normalmente se caracteriza mediante la temperatura de ruido de la antena) y la contribución de todas las otras fuentes que están transmitiendo en la misma banda. Esta contribución se suelen englobar bajo el término *interferencia*, en cuyo caso se entiende que el ruido incluye el resto de los efectos exceptuando las perturbaciones causadas por las señales generadas por otros transmisores.

Como en todos los sistemas y, por tanto, también en GNSS, el ruido enmascara la señal útil. Las consecuencias primordiales en el caso de receptores GNSS son que:

- El DLL no puede identificar con total precisión las transiciones de los chips o, lo que es lo mismo, la medida del retardo de la señal no es perfecta, sino que contiene unas fluctuaciones alrededor del valor ideal. Estas fluctuaciones se suelen denominar *jitter* y, en ocasiones, el propio término *jitter* se toma como sinónimo de la desviación típica de dichas fluctuaciones.
- Las medidas proporcionadas por el PLL también contienen un *jitter* de fase.

Hemos de recordar que estos dos efectos se producen realmente en cualquier receptor, pero lo que ocurre es que no son los efectos más relevantes en el caso habitual de los receptores de comunicaciones. En estos últimos, lo primordial son los errores que causa el ruido en la detección de los datos. En cambio, en los receptores GNSS, la probabilidad de error en la detección del mensaje de navegación no suele ser un factor de preocupación bien porque en escenarios benignos (por ejemplo, exteriores, sin bloqueos de la señal, etc.) es sumamente baja o incluso la señal es una señal piloto que no contiene datos, o bien porque cuando el receptor no va a operar en escenarios benignos se puede complementar con un sistema de asistencia (este es el concepto de *Assisted-GNSS*).

Bibliografía

Las interferencias se tratan en el subapartado 5.4.7 de este módulo.

Ved también

El concepto de *Assisted-GNSS* se explica en el módulo "Introducción a los sistemas de radionavegación y radiodeterminación" y se trata también en el apartado 6 de este módulo.

En el ámbito de GNSS, se suele utilizar una regla aproximada según la cual un receptor puede medir el retardo o desfase de una señal con una precisión inferior a 0,5% (aquí estamos teniendo en cuenta solo el efecto del ruido térmico y al nivel nominal en GNSS) de la "longitud de onda" asociada a la señal, donde "longitud de onda" se entiende en este caso como la separación que existe entre las transiciones de la señal. En el caso del pseudorrango, esta separación es la duración de un chip, que para la señal GPS L1 C/A es aproximadamente $1 \mu\text{s}$ o 300 m. Por lo tanto, la regla diría que la precisión en los pseudorrangos es al menos de 1,5 m (considerando solo el ruido térmico como fuente de degradación). Para el observable de fase, la longitud de onda es la correspondiente a la frecuencia portadora. De nuevo para GPS L1, esto es 19 cm, y la precisión de las medidas de fase de portadora es de 1 mm.

Reflexión

Podemos observar que esta regla aproximada que se utiliza en GNSS sobre la precisión de los pseudorrangos tiene un paralelismo con la que se utiliza en radar, según la cual la precisión en la medida de distancias es una fracción del inverso del ancho de banda de la señal utilizada por el radar. De hecho, podemos comprobar que realmente vienen a decir lo mismo porque el inverso del ancho de banda es una medida cualitativa del tiempo entre transiciones en la señal, y este tiempo expresado en metros es la "longitud de onda" que hemos mencionado.

De hecho, la regla anterior solo sirve para realizar cálculos aproximados del orden de magnitud de la precisión porque esta no depende únicamente de la estructura de la señal (por ejemplo, de la longitud de onda asociada), sino que depende también de la relación señal a ruido (SNR). En los sistemas de posicionamiento, la calidad de la señal no se suele caracterizar por la SNR, como en los sistemas de comunicaciones, sino por la C/N_0 , que es el cociente entre la potencia de la señal deseada (C) y la densidad espectral de ruido (N_0). El motivo es que en los sistemas de comunicaciones, lo normal en la gran mayoría de los casos es que el receptor tenga un ancho de banda que deje pasar toda la señal deseada, porque en caso contrario se filtraría la señal y se produciría interferencia intersimbólica. Por lo tanto, el receptor capta toda la potencia de la señal y la potencia de ruido que hay en el ancho de banda que la ocupa. En cambio, en los sistemas GNSS actuales se transmiten señales formadas por diferentes combinaciones de pulsos rectangulares, que en el espectro tienen numerosos lóbulos secundarios. Aunque no se transmiten todos los lóbulos secundarios porque esto implicaría un ancho de banda infinito, es habitual transmitir un número elevado. El receptor tiene un ancho de banda que deja pasar más o menos lóbulos, pero no necesariamente todo el ancho de banda de la señal porque, a diferencia de en comunicaciones, se puede filtrar la señal y seguir midiendo el retardo y fase de la señal. La selección del ancho de banda del receptor es una cuestión de diseño que responde al compromiso entre precisión, complejidad y susceptibilidad a interferencias. Mientras que al aumentar el ancho de banda la precisión mejora, la complejidad, el consumo de energía del receptor y su susceptibilidad a posibles interferencias empeoran. La SNR realmente depende del ancho de banda del receptor porque según su

valor se deja pasar más o menos potencia de señal y de ruido y, por tanto, la SNR no es un parámetro que dependa únicamente de la señal como tal. En GNSS se utiliza entonces habitualmente la C/N_0 (que también se denota indistintamente como CNo) para tener un parámetro que caracterice la calidad de la señal recibida independientemente del receptor.

La relación exacta entre el *jitter* de código y de fase con la CNo depende del tipo concreto de DLL y PLL que se utiliza. No obstante, en condiciones normales, no existen grandes variaciones para los diferentes tipos de lazos, y se pueden dar unas expresiones válidas en general que proporcionen valores aproximados de estos dos *jitters*:

$$\begin{aligned}\sigma_P &= cT_c \sqrt{\frac{B_{DLL} d_{E-L}}{2K C/N_0}} \text{ [m]} \\ \sigma_\phi &= \frac{\lambda}{2\pi} \sqrt{\frac{B_{PLL}}{C/N_0}} \text{ [m]}\end{aligned}\quad (110)$$

donde es la duración de un chip, B_{DLL} es el ancho de banda del DLL, B_{PLL} es el ancho de banda del PLL, d_{E-L} es el espaciado *early-late* del DLL en unidades de chips, y K es un factor que depende del tipo de señal utilizado. Este factor mide cómo de agudo es el pico de la curva de correlación en relación con el caso de una señal BPSK (como por ejemplo la señal C/A de GPS), en cuyo caso $K = 1$. Para otras señales, tenemos los siguientes valores aproximados: $K = 3$ para BOC(1,1), $K = 23$ para BOC(15, 2,5), y $K = 6$ para AltBOC(15,10).

5.4.7. Efectos locales en el receptor: las interferencias

Los sistemas GNSS son vulnerables como cualquier otro sistema a la presencia de señales no deseadas en la misma banda.

En general, se engloba bajo el término *interferencia* cualquier señal que no sea la del satélite que se pretende recibir ni el ruido de fondo que capta la antena.

Existen varios criterios para clasificar las interferencias. Uno de ellos es según el ancho de banda de la interferencia comparado con el ancho de banda de la señal deseada, que da lugar a los casos de interferencia de banda ancha (*wideband*) y de banda estrecha (*narrowband*). Una misma interferencia puede ser de banda ancha para la señal GPS L1 C/A y de banda estrecha para la señal GPS L5. El caso límite de interferencia de banda estrecha es un único tono, que se denomina interferencia de onda continua (CW, *continuous wave*).

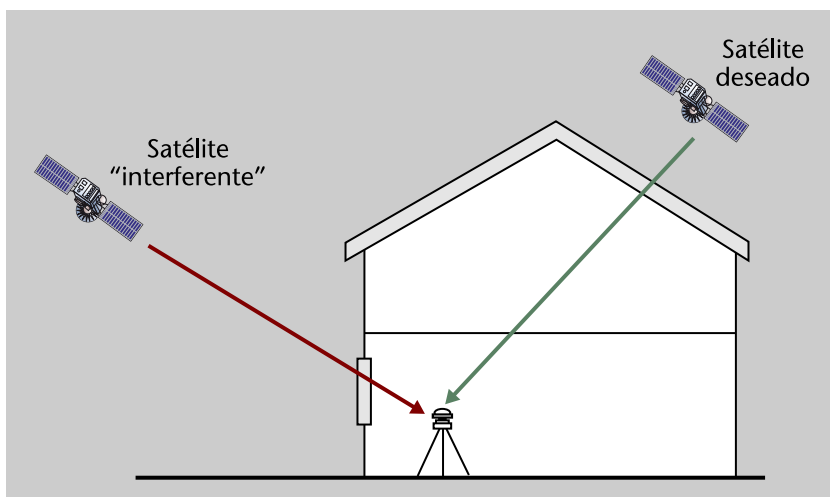
Las señales recibidas del resto de los satélites del mismo sistema GNSS también interfieren a la componente que se pretende recibir. Esto se denomina autointerferencia o interferencia intrasistema (*self-interference* o *intra-system interference*).

La interferencia intersistemas (*inter-system interference*) hace referencia a la interferencia que varios sistemas GNSS, por ejemplo GPS y GALILEO, se generan uno al otro. En general, los sistemas están diseñados de manera que las interferencias intrasistema e intersistemas tengan un efecto casi despreciable. Existen, no obstante, excepciones en dos situaciones. Una se da normalmente en entornos urbanos o interiores, en los que la señal deseada se recibe atenuada debido a los obstáculos (podéis ver la figura 26), mientras que alguna de las otras señales del propio sistema se reciben con una potencia notablemente más elevada. Por ejemplo, esto ocurre cuando la señal deseada cruza paredes y las señales de otros satélites se reciben a través de una ventana. Cuando la diferencia de potencias es suficientemente grande como para que no pueda ser compensada por la ganancia de procesamiento de las señales DS-CDMA, el sistema se aparta de las condiciones nominales de diseño y las componentes intrasistema dejan de ser despreciables; entonces, se deben aplicar técnicas específicas para combatir sus efectos.

Este fenómeno de desbalance de potencias debido a las condiciones de propagación se suele denominar *near-far*, por similitud con el efecto que se produce en sistemas de comunicaciones celulares debido a las diferencias de distancias de los móviles a las estaciones base.

En el caso que acabamos de describir sobre GNSS, la diferencia de potencias no se debe a la diferencia de distancias, pero se sigue manteniendo el mismo nombre de *near-far*.

Figura 26. Ilustración del efecto *near-far*



En la segunda situación, el efecto *near-far* sí que está ocasionado por la diferencia de distancias a dos transmisores. Esto ocurre cuando se utilizan sistemas terrestres de aumentación (GBAS) o pseudolitos. El receptor puede estar muy próximo al transmisor terrestre, recibiendo una señal muy potente de este, que enmascararía las señales que llegan de los satélites. Una solución habitual en estos casos es que la señal terrestre sea pulsada, para que solo interfiera a los satélites durante una fracción del tiempo.

La interferencia puede ser intencionada o no. El primer caso se denomina *jamming*, y además, en algunas situaciones, su objetivo no solo es interrumpir el funcionamiento del receptor, sino también hacer que este calcule una posición incorrecta mediante la transmisión de señales aparentemente idénticas a las señales del sistema GNSS pero falsas. Esto último se ha denominado *spoofing*.

Existen numerosas técnicas de *spoofing*. Aquí solo vamos a mencionar un ejemplo ilustrativo, que recibe el nombre de *repeat-back spoofing*. Consiste simplemente en utilizar equipos que retransmiten las señales reales que captan de los satélites. Si un receptor próximo se engancha a las señales retransmitidas en lugar de a las directas, calculará la posición de retransmisor en lugar de la suya propia.

Los avances que se han producido recientemente, tanto en el ámbito de los sistemas como en los receptores, han reducido significativamente el efecto del resto de las fuentes de error y han provocado que el *jamming* y, especialmente, el *spoofing* sean unos de los elementos que más preocupan hoy en día en calidad de ser la fuente de vulnerabilidad más grave de los sistemas GNSS. Anteriormente, este problema se circunscribía casi únicamente a aplicaciones militares, es decir, en escenarios donde es normal esperar la existencia de interferencias intencionadas. Pero a día de hoy también es un problema en el ámbito civil porque los sistemas GNSS ya no se utilizan solo en aplicaciones puramente informativas o relacionadas con el ocio, sino que cada vez son más habituales en aplicaciones donde la seguridad de personas está en juego (que se suelen denominar *safety-of-life applications*) o la seguridad de un sistema crítico puede verse afectada (por ejemplo, la datación de las transacciones financieras). Por todo esto, el estudio y desarrollo de técnicas de anti-*jamming* y anti-*spoofing* está recibiendo gran atención actualmente.

Ejemplo de interferencias

En el año 2010, se instaló en el aeropuerto de Newark (Nueva York) un sofisticado sistema GBAS para utilizar GPS en las operaciones de aproximación y de control del tráfico aéreo. Se observaron problemas en su funcionamiento, y después de meses de investigaciones se llegó a la conclusión de que los problemas se debían a unos pequeños transmisores que llevaban los camiones que pasaban por una autopista cercana. Estos transmisores, extremadamente sencillos y que se alimentan con la toma de 12 V del vehículo, son utilizados para interferir el receptor GPS del camión y evitar que su posición vaya quedando registrada. Este es un ejemplo de cómo un sistema GBAS de varios millones de dólares puede dejar de funcionar debido a unos dispositivos que cuestan 20 dólares.

En la región del Mediterráneo, se ha reportado en numerosas ocasiones problemas en la recepción de GPS causados por productos de intermodulación generados en las estaciones de TV digital.

Aunque todas las bandas de GNSS están expuestas a interferencias, existen unas más problemáticas que otras. La banda L1 está asignada a los sistemas GNSS de diferentes países y, por tanto, está bastante protegida y libre de interferencias externas. La situación es diferente en L2 y L5. La banda L2 está compartida con algunos sistemas radar. La banda L5 se solapa parcialmente con la utilizada por otros sistemas de navegación aérea, como son DME y TACAN.

La degradación que ocasiona una interferencia depende de muchos factores, como el tipo de interferencia, el tipo de señal deseada, la arquitectura y las técnicas de mitigación del receptor, etc., por lo que no es posible dar un resultado único sobre el aumento del *jitter* de código y fase para una determinada potencia de interferencia. Podemos cuantificar de manera genérica el efecto de la interferencia, de potencia P_I , a través de la CNo efectiva, que se define como:

$$CN_{0,\text{eff}} = \frac{1}{CN_0^{-1} + \frac{P_I}{CT_c^{-1}Q}} \quad (111)$$

donde Q es el factor que mide la resistencia de la señal deseada frente a la interferencia, y recoge toda la casuística mencionada antes. Esta CNo efectiva se puede utilizar en las expresiones 110 en lugar de la CNo original para obtener la desviación típica en el código y en la fase¹⁷.

⁽¹⁷⁾Esto es válido obviamente si el receptor continúa en seguimiento, y la interferencia no hace que el receptor se desenganche o sature las etapas analógicas del receptor impidiendo completamente su funcionamiento.

En la tabla 3, podemos ver ejemplos del valor de Q en algunos escenarios. Estos valores corresponden a la situación en la que el receptor no implementa ninguna contramedida especial frente a interferencias (excepto en el último caso). Además, son valores aproximados porque solo tienen en cuenta la forma del pulso conformador de chip de la señal deseada y no la secuencia pseudoaleatoria que la forma. El primer caso de la tabla, para el cual $Q = 1$, da el mismo resultado que el análisis básico que se suele realizar al estudiar las señales DS-SS: Al realizar el proceso de *despreading* en el receptor, se ensancha el espectro de todas las señales que no comparten la secuencia aleatoria de la señal deseada. Por lo tanto, la interferencia de banda estrecha pasa a ocupar un ancho de banda igual $1/T_c$ después del *despreading*, y se puede asimilar aproximadamente a ruido blanco de densidad espectral $P_I T_c$.

Tabla 3. Factor que mide la protección de la señal deseada frente a la interferencia (Kaplan, 2006)

| Tipo de interferencia y señal deseada | Q |
|---|---------------------|
| Interferencia CW y señal deseada BPSK | 1 (peor caso) |
| Interferencia CW y señal deseada BOC | 1,9-2,5 (peor caso) |
| Interferencia con el mismo espectro que la señal deseada, que es BPSK | 1,5 |

| Tipo de interferencia y señal deseada | Q |
|---|--|
| Interferencia con el mismo espectro que la señal deseada, que es BOC | 3-4,5 |
| Interferencia es ruido blanco con un ancho de banda B_I que incluye a la señal deseada | $B_I T_C^{-1}$ |
| Interferencia es ruido blanco en un ancho de banda $B_I = 2/T_C$, centrado en el máximo espectral de la señal deseada BPSK | 2,22 |
| Interferencia es ruido blanco en un ancho de banda $B_I = 2/T_C$, centrado en uno de los dos lóbulos principales de la deseada BOC | 4,6-5,5 |
| Interferencia pulsada (suponiendo que el receptor utiliza <i>pulse blanking</i>) | Inverso del ciclo de trabajo (<i>duty cycle</i>) |

Si el receptor incluye alguna técnica de mitigación de interferencias, el valor de Q se puede hacer más elevado. Existe también una larga variedad de técnicas de mitigación, cuya descripción en detalle podría ser objeto de un curso completo. Vamos a comentar brevemente aquí algunas de las más representativas. En el caso de interferencias CW, se pueden utilizar filtros *notch* o filtros de escisión, que sitúan un nulo muy estrecho en el espectro, cancelando la interferencia pero casi sin afectar a la señal. La posición del nulo puede ser conocida a priori o estimarse de forma adaptativa. Si el filtro se sitúa después del ADC, este debe tener un margen dinámico suficiente (lo que se traduce en una profundidad de varios bits) para dejar pasar la señal con poca distorsión; si no, la interferencia bloquearía completamente el ADC y el filtro no podría hacer nada para eliminarla.

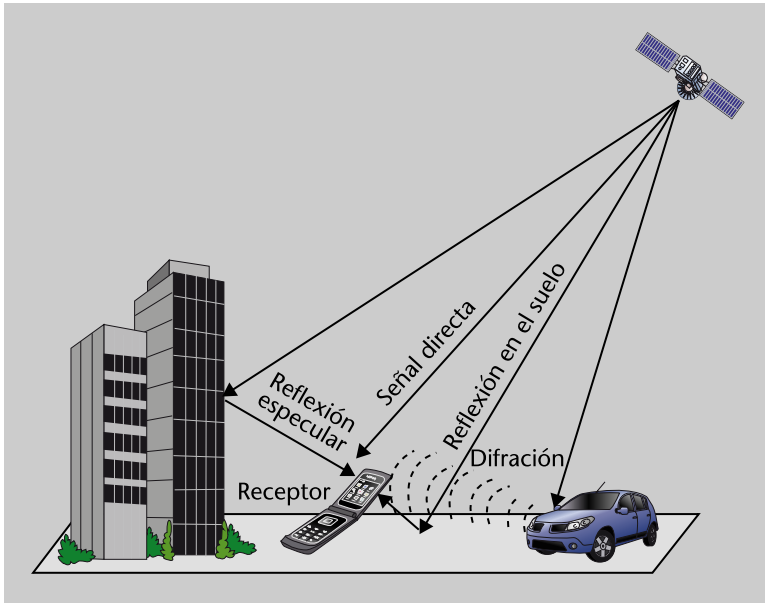
Si las interferencias son pulsadas, es efectivo utilizar *pulse blanking*. Consiste en anular la señal que continúa hacia etapas posteriores del receptor cuando se detecta una amplitud muy grande a la entrada. Se suele implementar justo antes del ADC. La alternativa que es útil ante cualquier tipo de interferencia, incluidas las intencionadas y el *spoofing*, son los arrays adaptativos de antenas, que permiten situar nullos de recepción en las direcciones de llegada de las interferencias. Los arrays también permiten averiguar las direcciones de las señales recibidas, lo que es muy interesante, por ejemplo, frente al *repeat-back spoofing* porque se puede observar que todas las señales aparentemente deseadas llegan de una única dirección en lugar de llegar de satélites distribuidos de manera aleatoria en el cielo. El uso de arrays de antenas para receptores GNSS está bastante extendido en el ámbito militar y, recientemente, también se está considerando seriamente para algunas aplicaciones civiles.

5.4.8. Efectos locales en el receptor: el multicamino

La propagación multicamino¹⁸ consiste en que la señal transmitida por el satélite llega al receptor a través de varios caminos, de manera que este recibe la señal directa y varias versiones retardadas y atenuadas de ella (podéis ver la figura 27).

⁽¹⁸⁾En inglés, *multipath*.

Figura 27. Ilustración de la propagación multicamino



Estas réplicas siempre llegan con un retardo mayor que el de la señal directa por el motivo obvio de que la señal directa recorre el camino más corto entre el satélite y el receptor. También suelen llegar con una amplitud más pequeña porque el fenómeno que las genera (normalmente reflexión o difracción en alguna superficie u objetos próximos al receptor) introduce una cierta atenuación. No obstante, esto no tiene por qué ser siempre así, ya que a veces también se da el efecto de *shadowing*, por el cual algunos objetos obstruyen el camino directo y la señal directa llega atenuada. En este caso, algunas componentes multicamino (que denominaremos en ocasiones y por comodidad reflexiones, aunque como hemos visto la reflexión no es el único proceso que las produce) pueden ser más potentes que la propia señal directa.

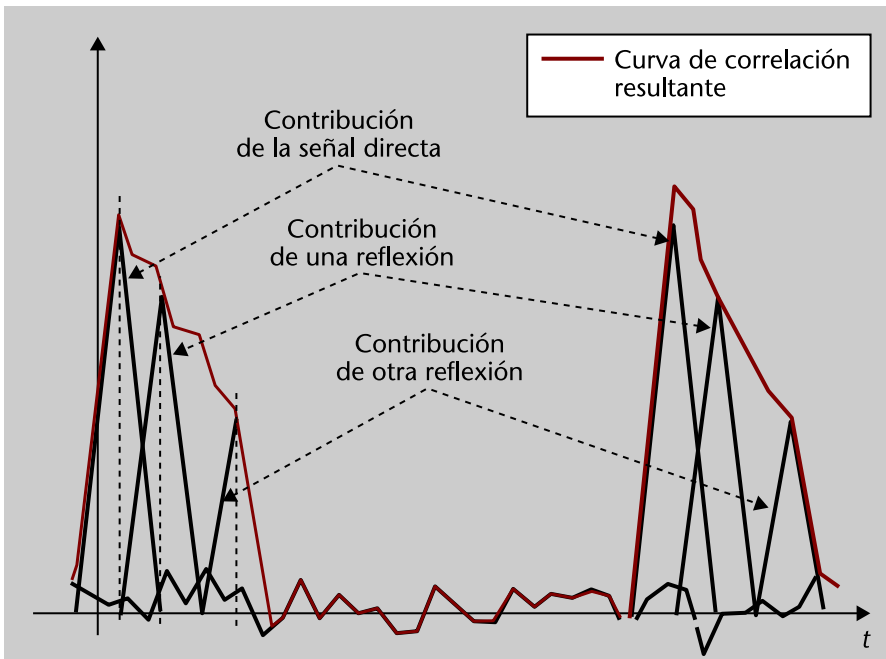
Por efecto multicamino se suele entender el error que causa la propagación multicamino en el seguimiento de código y de fase.

Aunque las componentes multicamino también afectan a la adquisición y a la detección del mensaje, los efectos no son tan importantes como en el *tracking* y, además, no tienen por qué ser negativos, como ocurre en sistemas de comunicaciones.

El error en el *tracking* de código se produce porque al estar la señal deseada solapada con réplicas retardadas de ella misma, se distorsiona también la curva de correlación a partir de la cual el DLL determina el retardo. Esta distorsión se puede observar cualitativamente en la figura 28, que muestra la salida de un filtro adaptado al código PN cuando a la entrada se recibe la señal directa y dos réplicas retardadas con amplitudes ligeramente más pequeñas. El DLL acaba midiendo un retardo que no depende únicamente del instante de llegada de la señal directa, sino también de los instantes de llegada de las réplicas. Hemos de darnos cuenta de que una reflexión no aporta ninguna información útil para el posicionamiento mediante GNSS (existen algunas excepciones a esta afirmación en aplicaciones científicas o de posicionamiento en interiores) porque su retardo de propagación no es una medida de la distancia geométrica entre el receptor y el satélite (que es la base del posicionamiento GNSS), sino que corresponde al camino que sigue la señal desde el satélite hasta la zona en la que se produce la componente multicamino más la distancia desde esa zona hasta la antena del receptor. Esta es una situación manifiestamente diferente a lo que ocurre en comunicaciones *spread spectrum* de secuencia directa (DS-SS). Si los retardos relativos de las diferentes componentes son comparables o mayores al tiempo de chip, el multicamino es una fuente de diversidad que permite mejorar las prestaciones del sistema utilizando esquemas RAKE que capturan y combinan la energía que llega a través de los diferentes rayos. Si los retardos relativos son menores al tiempo de chip, el multicamino provoca básicamente *fading*, es decir, fluctuaciones en la amplitud de la señal.

Podemos observar en la figura 28 que un punto cualquiera de la curva gris en la zona de los picos de correlación contiene la contribución de la señal directa y también de las reflexiones. Por lo tanto, la fase medida por el PLL tampoco será la fase de la señal directa, sino que estará afectada por los valores de la correlación de las reflexiones.

Figura 28. Distorsión de la curva de correlación causada por el multicamino



Se muestra la contribución de las cuatro componentes por separado en negro (señal directa, dos reflexiones y ruido) y el resultado global en gris.

Los errores causados por el multicamino dependen de:

- la amplitud,
- el retardo,
- la fase y
- la desviación de frecuencia o Doppler

de cada de una las reflexiones respecto a la de la señal directa.

Todos estos parámetros son diferentes en general para cada uno de los satélites. Por esto, el multicamino causa un error diferente en los observables de cada satélite que se traslada a la posición porque no se puede incluir en el término del error del reloj del receptor.

Antes de pasar a analizar concretamente dichos errores, vamos a comentar posibles clasificaciones del multicamino. Este puede estar compuesto por componentes especulares y/o componentes difusas, dependiendo del entorno físico del receptor. Las primeras están producidas por superficies planas, como las alas de un avión, la superficie del mar y de la tierra, y ciertos edificios. Las componentes difusas se pueden interpretar como un conjunto de pequeñas contribuciones indistinguibles entre ellas y normalmente producidas por difracción o reflexión en superficies rugosas. Hay que tener en cuenta que la potencia de una determinada reflexión es normalmente inferior (los casos excepcionales ya se han comentado anteriormente) a la de la señal directa, pero la potencia de todo el conjunto de reflexiones puede superar a la de esta última.

El efecto que causa una determinada reflexión en los lazos de código o de fase depende de la velocidad de cambio de su fase con respecto a la fase de la señal directa, que es justamente la diferencia entre las frecuencias Doppler de la reflexión y de la señal directa. Si esta diferencia es menor que el ancho de banda del lazo, se dice estar en la caso de desvanecimiento lento (*slow fading*), y, en caso contrario, desvanecimiento rápido (*fast fading*). El primero de ellos suele ser más perjudicial para el receptor. El rango de frecuencias que cubren todas las frecuencias Doppler de todas las reflexiones, incluyendo la señal directa, se denomina ancho de banda de desvanecimientos o *fading bandwidth*.

La diferencia entre las frecuencias Doppler de una reflexión y de la señal directa depende de la geometría del conjunto satélite-receptor-reflector y de la velocidad del receptor. A continuación vamos a presentar unos valores aproximados para esta diferencia. Para receptores estáticos, la frecuencia Doppler de la reflexión en el receptor es la misma que en el punto de reflexión, ya que el proceso de reflexión no cambia la frecuencia de la señal si el reflector es estático. Por lo tanto, si las posiciones del receptor y del reflector son conocidas, la diferencia Doppler se puede calcular para una determinada órbita del satélite. La diferencia es máxima cuando el satélite está en el cénit y el reflector está en el plano de la órbita. Si la distancia del reflector al receptor es d , entonces la máxima diferencia Doppler para la constelación GPS (expresada en milihertzios) vale:

$$\Delta f_{d,\max} = 1,38 d \text{ [mHz]} \quad (112)$$

donde d está expresada en metros.

Veremos que solo nos preocupan las reflexiones con un retardo inferior a $1,5 T_c$, ya que las reflexiones con un retardo mayor no afectan a las medidas. Un tiempo de $1,5 T_c$ para el código C/A de GPS representa 450 m y, como se puede aproximar la distancia entre el receptor y el reflector por la diferencia de caminos entre el rayo directo y el reflejado, se obtiene que la máxima diferencia Doppler es:

$$1,38 \times 450 = 621 \text{ mHz} = 0,621 \text{ Hz} \quad (113)$$

Para otro tipo de geometrías la diferencia siempre es menor. Si el satélite, receptor y reflector están alineados, la diferencia es cero. Pero es difícil predecir durante cuánto tiempo estará la diferencia por debajo de un cierto nivel porque la geometría cambia continuamente, con lo que resulta en una frecuencia Doppler también cambiante.

Para receptores móviles, la diferencia de frecuencias Doppler es mucho mayor. Un receptor con velocidad v da lugar a:

Reflectores móviles

El caso de reflectores móviles se puede analizar como el caso de receptor móvil, que lo presentaremos más adelante.

$$\Delta f_d = v \frac{f}{c} (\cos(\alpha_r) \cos(\beta_r) - \cos(\alpha_s) \cos(\beta_s)) \quad (114)$$

donde α_r , β_r y α_s , β_s son las diferencias en elevación y acimut entre el vector velocidad del receptor y los vectores que van del receptor al reflector y al satélite, respectivamente. Por ejemplo, si $v = 15$ m/s, la diferencia puede valer 160 Hz. Para geometrías especiales la diferencia de frecuencias puede ser cero, pero es muy improbable que estas situaciones duren más que algunos segundos, ya que la geometría receptor-reflector cambia muy rápidamente debido a la velocidad del receptor (o del reflector).

Hay que tener en cuenta que las reflexiones que provienen de partes del propio móvil, como pueden ser las reflexiones en las alas de un avión, se comportan como si el receptor fuera estático.

El multicamino es especialmente perjudicial porque causa un sesgo en las medidas del PLL y del DLL. Se dice que son sesgos porque son errores constantes, o al menos aproximadamente constantes, que no se pueden reducir promediando los observables durante intervalos de tiempo relativamente cortos, como ocurre con el ruido. En la práctica, los errores causados por una reflexión se decorrelan a partir de tiempos superiores al inverso de la diferencia entre la frecuencia Doppler de la reflexión y de la señal directa. En escenarios estáticos esta diferencia puede ser muy pequeña, lo que da lugar a duraciones de correlación muy largas de los errores causados por el multicamino, lo que provoca que estos errores sean a efectos prácticos casi constantes.

Para calcular los errores en el observable de fase, consideramos que el valor esperado de la medida que proporciona el PLL es la fase de muestra de correlación *prompt* en ausencia de ruido:

$$R_p = \sum_{i=0}^M a_i e^{j\varphi_i} R_c(\hat{\tau}_0 - \tau_i) \quad (115)$$

donde M es el número de reflexiones, cada una con amplitud, fase y retardo igual a a_i , φ_i y τ_i , respectivamente. La autocorrelación del código es $R_c(\tau)$, y el subíndice 0 hace referencia a la señal directa, cuya estimación del retardo es $\hat{\tau}_0$. En el caso de *slow fading* podemos suponer que la amplitud y fase de las reflexiones es constante a bloques (por ejemplo, durante el intervalo entre medidas sucesivas del PLL). Para obtener una expresión cerrada de la fase, vamos a suponer que solo existe una reflexión (por ejemplo, $M = 1$). Los máximos errores de fase se producen cuando el desfase entre la señal directa y la reflexión cumple:

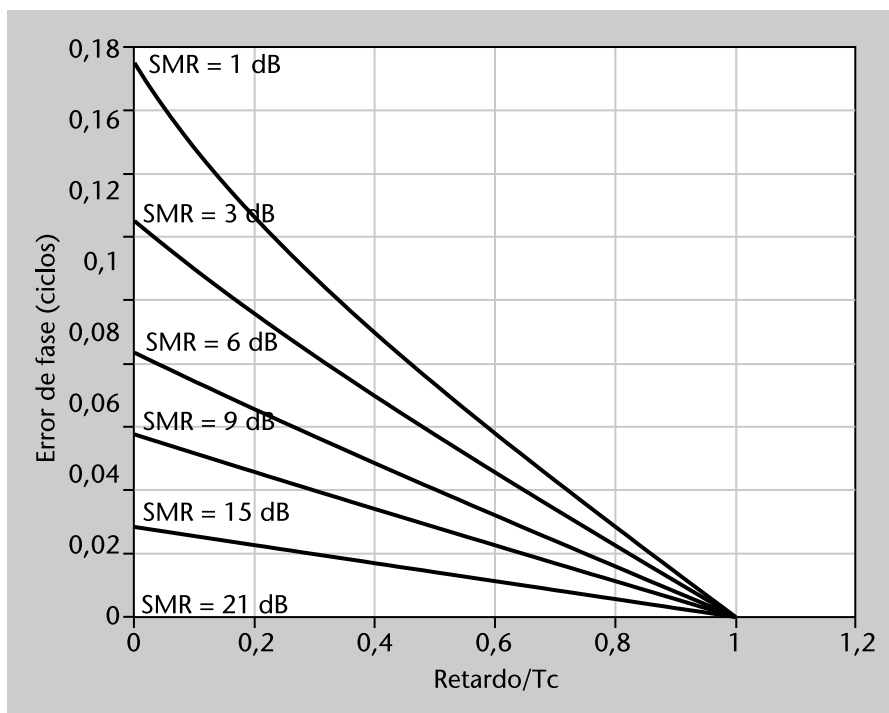
$$|\varphi_1 - \varphi_0| = \pi - \arccos \left(\frac{R_c(\hat{\tau}_0 - \tau_1)}{R_c(0) \sqrt{\text{SMR}}} \right) \quad (116)$$

donde es la relación señal a multicamino. Entonces, el error de fase (en metros) vale

$$\varphi_e = \pm \frac{\lambda}{2\pi} \arcsen\left(\frac{R_c(\hat{\tau}_0 - \tau_1)}{R_c(0)\sqrt{\text{SMR}}}\right) \quad (117)$$

En la figura 29 se muestran los valores positivos que toma esta expresión. Podemos observar que los errores son relativamente pequeños, ya que en ningún caso superan un cuarto de la longitud de onda. En el caso de *fast fading*, los errores de fase son más difíciles de cuantificar, pero menores que en el caso aquí estudiado.

Figura 29. Errores de fase en función de la diferencia de retardos entre la reflexión y la señal directa

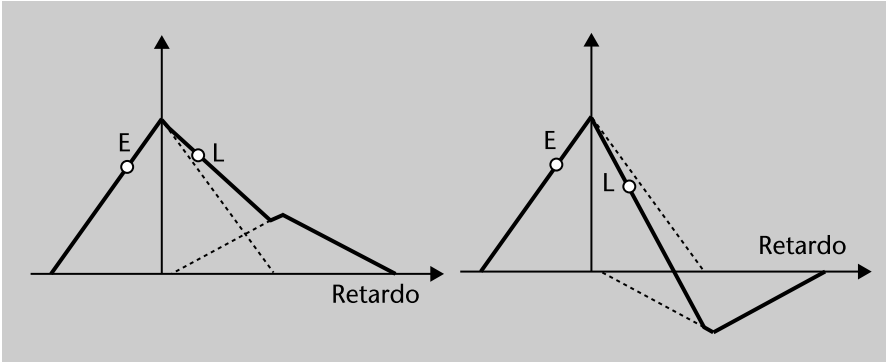


Aunque el modelo basado en un único rayo que hemos utilizado puede no ser muy representativo de escenarios reales, es muy habitual en el estudio del efecto del multicamino en GNSS porque permite ganar conocimiento sobre el funcionamiento de diferentes técnicas y sirve para establecer comparaciones entre ellas. Por estos motivos, este modelo de dos rayos es el que también vamos a utilizar para analizar los errores de código.

La reflexión distorsiona la curva de correlación y por tanto también la curva S, cuyo cruce por cero se desplaza y deja de coincidir con el retardo de la señal directa. Podemos observar cualitativamente la distorsión de la curva de correlación en la figura 30, que representa lo que ocurriría en el caso de GPS puesto que se han utilizado correlaciones individuales de forma triangular, lo que equivale a suponer que los pulsos de chip son rectangulares. Cuando las muestras *early* y *late* están centradas en el retardo de la señal directa, no tienen la misma amplitud, por lo que no se anula la curva S. Para que tengan la mis-

ma amplitud y la curva S se anule, hay que desplazar las muestras ligeramente hacia la derecha en el caso de que la reflexión esté en fase o hacia la izquierda en el caso de que la reflexión esté en contrafase con la señal directa. Esto explica el resultado aparentemente sorprendente de que el retardo medido es menor que el retardo real de la señal directa cuando se recibe una reflexión que se combina destructivamente, dando la impresión de que la señal total se adelanta, cuando en realidad la reflexión llega retrasada en todos los casos.

Figura 30. Distorsión de las curvas de correlación causada por una reflexión



En la primera gráfica, la reflexión y la señal directa están en fase, mientras que en la segunda están en contrafase. Se utiliza un pulso conformador de chip rectangular.

Los valores *early* y *late* de la correlación se pueden expresar como:

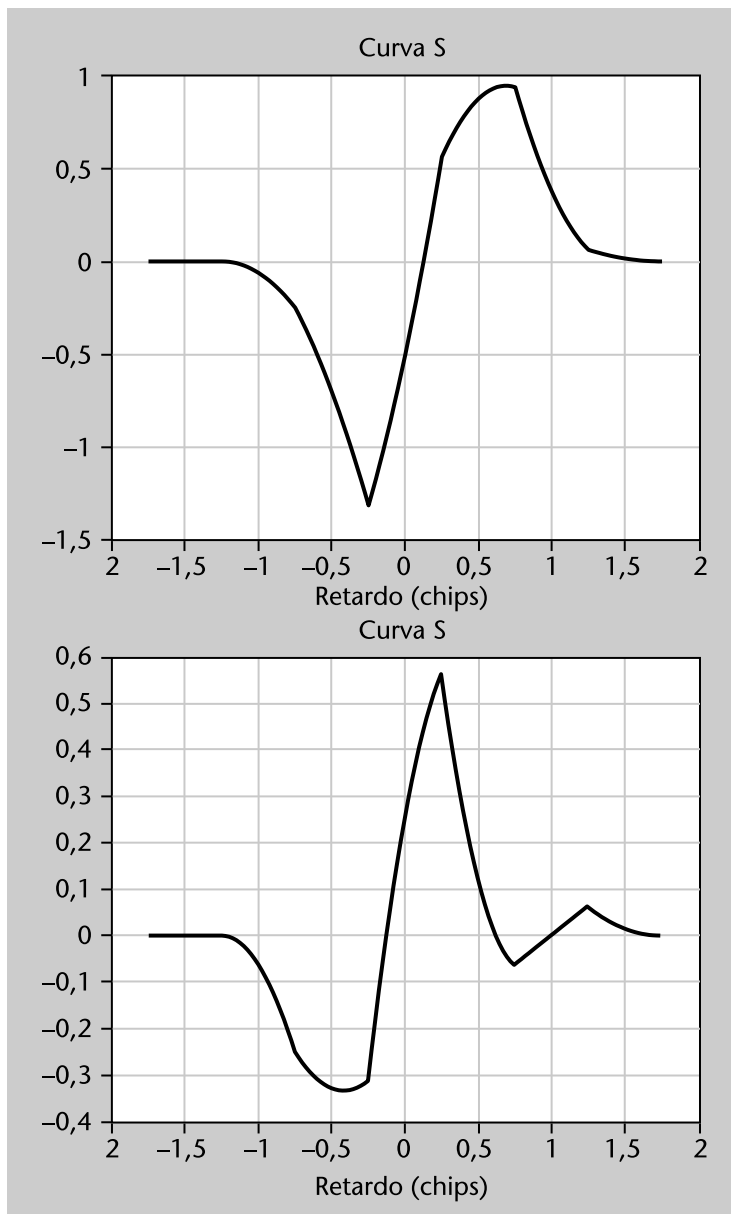
$$\begin{aligned} R_E &= a_0 e^{j\varphi_0} c\left(\hat{\tau}_0 - \frac{d_{E-L}}{2} T_c - \tau_0\right) + a_1 e^{j\varphi_1} c\left(\hat{\tau}_0 - \frac{d_{E-L}}{2} T_c - \tau_1\right) \\ R_L &= a_0 e^{j\varphi_0} c\left(\hat{\tau}_0 + \frac{d_{E-L}}{2} T_c - \tau_0\right) + a_1 e^{j\varphi_1} c\left(\hat{\tau}_0 + \frac{d_{E-L}}{2} T_c - \tau_1\right) \end{aligned} \quad (118)$$

Suponiendo que utilizamos un discriminador no-coherente *early-minus-late*, su salida queda como:

$$\begin{aligned} S(\hat{\tau}_0) &= |R_E|^2 - |R_L|^2 = a_0^2 \left(c^2\left(\hat{\tau}_0 - \frac{d_{E-L}}{2} T_c - \tau_0\right) - c^2\left(\hat{\tau}_0 + \frac{d_{E-L}}{2} T_c - \tau_0\right) \right) \\ &+ a_1^2 \left(c^2\left(\hat{\tau}_0 - \frac{d_{E-L}}{2} T_c - \tau_1\right) - c^2\left(\hat{\tau}_0 + \frac{d_{E-L}}{2} T_c - \tau_1\right) \right) \\ &+ 2a_0 a_1 \cos(\varphi_1 - \varphi_0) c\left(\hat{\tau}_0 - \frac{d_{E-L}}{2} T_c - \tau_0\right) c\left(\hat{\tau}_0 - \frac{d_{E-L}}{2} T_c - \tau_1\right) \\ &+ 2a_0 a_1 \cos(\varphi_1 - \varphi_0) c\left(\hat{\tau}_0 + \frac{d_{E-L}}{2} T_c - \tau_0\right) c\left(\hat{\tau}_0 + \frac{d_{E-L}}{2} T_c - \tau_1\right) \end{aligned} \quad (119)$$

Podemos observar en la figura 31 la forma que adopta esta expresión en dos ejemplos concretos. A causa de la reflexión, la curva S pierde su simetría y deja de pasar por cero en el origen (que se supone que es el retardo que corresponde a la señal directa). El punto en el que se produce el cruce por cero indica justamente el sesgo que produce la reflexión en la medida del pseudorango. Efectivamente se puede comprobar que cuando la señal directa y la reflexión están en fase (gráfica de la izquierda), se mide un retardo mayor que el de la señal directa; mientras que cuando las señales tienen signo diferente (gráfica de la derecha), se mide un retardo que indica que aparentemente la señal llega incluso antes que la señal directa.

Figura 31. Curva S en presencia de una reflexión retardada $0,5 T_c$ con respecto a la señal directa

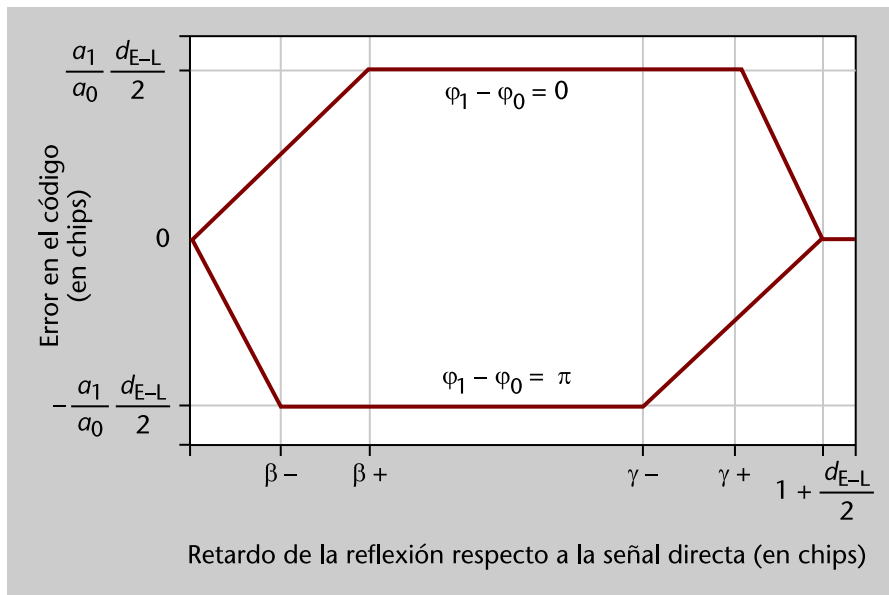


SMR = 6 dB. Espaciamiento *early-late* de medio chip. En la gráfica de la izquierda, la reflexión y la señal directa están en fase, mientras que en la gráfica de la derecha están en contrafase. Se utiliza un pulso conformador de chip rectangular.

Este ejemplo que acabamos de describir lo podríamos repetir para cualquier retardo de la reflexión, y podríamos representar en cada caso el error que obtendríamos en la medida cuando la reflexión está en fase y en contrafase con la señal directa. Entonces obtendríamos el resultado de la figura 32, que se denomina habitualmente envolvente de los errores multicamino (*multipath envelope*). Los parámetros que aparecen en los ejes de esta figura valen:

$$\beta_{\pm} = \frac{a_0 \pm a_1}{a_0} \frac{d_{E-L}}{2}, \quad \gamma_{\pm} = 1 - d_{E-L} \left(1 - \frac{a_0 \pm a_1}{2a_0} \right) \quad (120)$$

Figura 32. Envolvente de los errores causados por una reflexión en el DLL. Se utiliza un pulso conformador de chip rectangular



Esta gráfica representa los errores máximos en valor absoluto que causa una reflexión en el DLL. Esto es así porque los casos más desfavorables ocurren cuando la reflexión está en fase o en contrafase con la señal directa; para otros desfases, el error es menor en valor absoluto que los que se producen en esos dos casos. Por ello, el error causado por una reflexión con desfase arbitrario siempre se encuentra dentro de la región delimitada por las líneas de la figura 32, y de ahí viene el nombre de envolvente de los errores. Dado que esta figura determina los errores máximos causados por una reflexión, es una herramienta muy habitual para comparar los errores debidos al multicamino en diferentes señales.

Ejemplo

Vamos a calcular un ejemplo de error causado por una reflexión en la señal C/A de GPS. Dado que esta señal contiene chips rectangulares, podemos utilizar la figura 32. Supongamos que la reflexión recorre 60 m más que la señal directa y se recibe con una SMR de 6 dB. Asumimos también que se utiliza un DLL con espaciamiento *early-late* de 0,3 chips.

Como la longitud equivalente de los chips del código C/A son 300 m aproximadamente, el retardo de la reflexión es 0,2 chips. Su amplitud es $a_1 = a_0/2$. Los valores que marcan el inicio de las diferentes regiones de la envolvente son:

$$\beta^- = 0,075, \beta^+ = 0,225, \gamma^- = 0,0775, \gamma^+ = 1,075 \quad (121)$$

Por lo tanto, para un retardo de 0,2 chips, nos encontramos en la región lineal de la parte superior de la envolvente y en la región plana de la parte inferior. Los errores resultantes son 0,067 chips o -0,075 chips cuando la reflexión está en fase o en contrafase, respectivamente. Para otros valores del desfase, el error se encuentra entre estos dos valores, que, aunque pueden parecer que son una fracción muy pequeña de un chip, realmente representan 20 y -22,5 metros. Este cálculo revela la gran importancia que tienen los errores de multicamino en GNSS, ya que una sola reflexión no excesivamente potente y con un retardo de 0,2 μ s (lo que hace que sea una situación que se puede dar en la práctica con relativa facilidad) puede causar un error cuya magnitud supera al de cualquiera de las otras fuentes. Esto no deja de ser un ejemplo; para otro conjunto de parámetros más desfavorables, una única reflexión puede llegar a causar errores que superan los 100 metros.

Siguiendo con los valores numéricos planteados, la envolvente se anula a partir de un retardo de 1,15 chips, lo que significa que reflexiones que llegan con un retardo de hasta

345 m pueden causar errores. Pero si llegan con un retardo mayor, ya no afectan a la medida de distancia.

A partir de estos cálculos se hace evidente el interés en utilizar señales con frecuencias de chip mayores, como en la componente P de GPS o en la señal E5a/b de Galileo. Como ambos ejes de la envolvente son proporcionales al tiempo de chip, al utilizar frecuencias de chip más altas se reduce tanto la magnitud de los errores como el rango de reflexiones que los causan. Esto último es especialmente relevante porque la zona de alrededor de la antena que puede contener objetos susceptibles de producir las reflexiones también se reduce. Al ser una zona más pequeña, es más fácil controlarla o buscar emplazamientos con buenas condiciones. Por ejemplo, se puede afirmar que aproximadamente los errores debidos al multicamino son 10 veces menores con la señal GPS-P que con la señal GPS-C/A, ya que la primera tiene una frecuencia de chip 10 veces superior.

En el contexto de GNSS es muy habitual dibujar la envolvente de errores para diferentes tipos de señales y se considera que aquellas para las cuales la envolvente delimita un área (que se suele denominar *multipath envelope running average*) menor son más robustas frente al multicamino. Aunque, como ya se ha comentado, el modelo de una reflexión puede no ser representativo de un escenario real, su interés no solo radica en su sencillez, sino también en que se considera válido en general para establecer comparaciones entre señales porque si una señal se comporta mejor que otra en presencia de una reflexión, probablemente también se comportará mejor cuando se reciban varias reflexiones. Observad que hemos dicho "probablemente" porque no se puede realizar la afirmación en general, ya que el error causado por una suma de reflexiones no es la suma de errores individuales; y esto es así porque no aplica el principio de linealidad o superposición a los errores a causa de las operaciones cuadráticas que realiza el discriminador (podéis ver la ecuación 119). De hecho, que el efecto de las reflexiones no es lineal se puede observar en la propia figura 32, donde queda patente que un cambio del signo de la reflexión no causa únicamente un cambio del signo del error en todos los casos.

Podemos observar en la figura 32 que existen tres zonas en función del retardo de la reflexión. En la primera, el error aumenta (en valor absoluto) con el retardo de la reflexión, a continuación se llega a la zona de errores máximos, y luego los errores disminuyen y se anulan a partir de un retardo de $(1 + d_{E-L}/2)$ chips. Los puntos de inicio y final de cada zona vienen determinados por las expresiones 120. Se puede observar que en todos los casos los errores son proporcionales a la amplitud de la reflexión, a_1 . Además, los errores máximos son también proporcionales al espaciado *early-late* d_{E-L} . Esta observación conduce a la técnica más habitual de reducción de errores multicamino, que se denomina *narrow-correlator spacing*. Simplemente consiste en utilizar valores pequeños de d_{E-L} en el DLL. Por valores pequeños se entienden habitualmente valores de alrededor de 0,1 chips, en contraposición al valor estándar de 0,5 chips. Nos podríamos preguntar si haciendo d_{E-L} arbitrariamente pequeño, podríamos reducir los errores tanto como quisiéramos. Esto no es así porque la envolvente que hemos dibujado en la figura 32 presupone que el ancho de banda de la señal es infinito. Cuando el ancho de banda es finito, la envolvente se suaviza y la altura de la parte plana de la curva es proporcional a d_{E-L} solo a partir de un cierto valor mínimo de d_{E-L} (y este valor mínimo

es inversamente proporcional al ancho de banda), de manera que para valores de d_{E-L} inferiores a dicho valor mínimo, una reducción del espaciamiento no se traduce en una reducción de los errores en el pseudorrango.

La técnica del *narrow-correlator spacing* hace disminuir la altura de la envolvente, es decir, la magnitud de los errores, pero no reduce significativamente el rango de retardos de las reflexiones que causan errores, ya que, incluso para valores de d_{E-L} arbitrariamente pequeños, todas las reflexiones con retardos relativos de hasta un chip provocan errores. Para conseguir que la envolvente tenga menor amplitud (por ejemplo, menores errores para un retardo de la reflexión dado) y también menor longitud (por ejemplo, un intervalo más corto de reflexiones que causan errores) se deben utilizar otras técnicas. A grandes rasgos, se pueden clasificar en tres grupos, según se basen en:

- 1) Utilizar señales de referencia locales modificadas en lugar de una señal idéntica al código que se quiere recibir.
- 2) Métodos de estimación paramétricos, que calculan los parámetros (amplitud, retardo, etc.) de la señal directa y también de las reflexiones. El primer método de este tipo que se propuso fue el MEDLL (*multipath estimating DLL*).
- 3) Arrays de antenas inteligentes, que permiten discriminar las señales espacialmente y no solo en función de su retardo.

Técnicas de mitigación del multicamino

Una descripción detallada de las técnicas de mitigación del multicamino queda fuera del alcance de esta asignatura. Aunque se ha investigado ampliamente sobre este tema en los últimos 20 años, el problema de mitigar el multicamino que llega con retardos pequeños sigue siendo una línea activa de investigación.

En el análisis anterior, hemos supuesto que la fase de la reflexión podía tomar un valor cualquiera, pero que no variaba con el tiempo. Es decir, estábamos trabajando bajo la hipótesis de *slow fading*. Vamos a comentar brevemente qué ocurriría en el caso de *fast fading*. La curva S se puede obtener a partir de la expresión 119 eliminando los dos últimos términos. Los cosenos que hay en esta expresión dependen de una diferencia de fases que ahora sí que varía durante un intervalo de tiempo igual al inverso del ancho de banda del lazo. El lazo realiza por tanto un filtrado de la salida del discriminador, y esto se puede ver como el cálculo del valor medio de dicha salida. El valor medio de los cosenos es cero si la diferencia de fases tiene distribución uniforme (que es una hipótesis razonable) y, finalmente, la expresión de la curva S es:

$$S(\hat{\tau}_0) = |R_E|^2 - |R_L|^2 = a_0^2 \left(c^2 \left(\hat{\tau}_0 - \frac{d_{E-L}}{2} T_c - \tau_0 \right) - c^2 \left(\hat{\tau}_0 + \frac{d_{E-L}}{2} T_c - \tau_0 \right) \right) + a_1^2 \left(c^2 \left(\hat{\tau}_0 - \frac{d_{E-L}}{2} T_c - \tau_1 \right) - c^2 \left(\hat{\tau}_0 + \frac{d_{E-L}}{2} T_c - \tau_1 \right) \right) \quad (122)$$

Lectura recomendada

Una buena revisión de estas técnicas se puede encontrar en la obra de G. A. McGraw y M. S. Braasch (1999).

Lectura recomendada

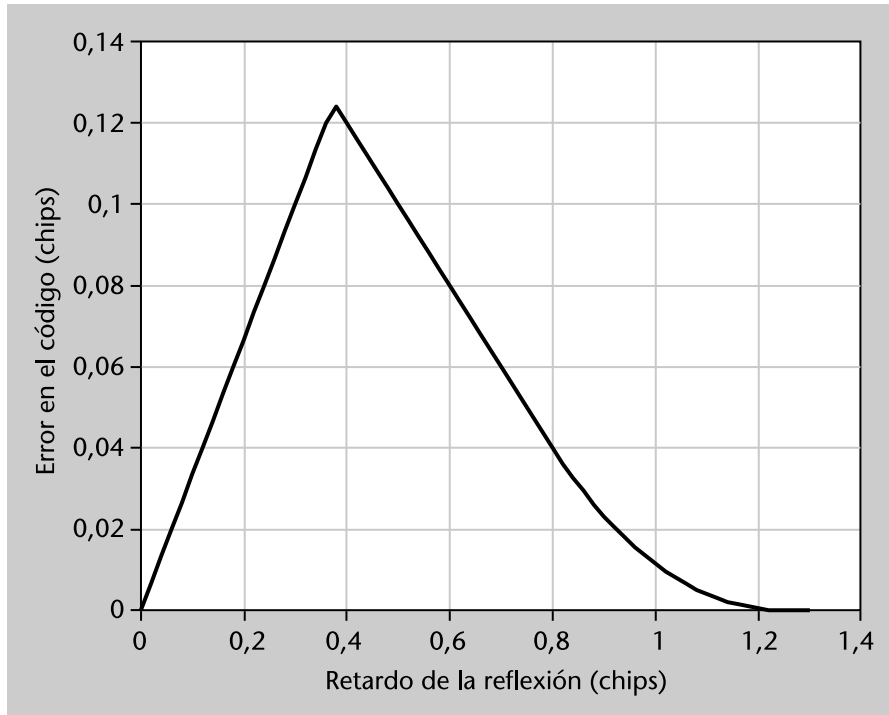
Una buena revisión de estas técnicas se puede encontrar en la obra de van Nee (1993).

Lectura recomendada

Una buena revisión de estas técnicas se puede encontrar en la obra de Seco-Granados, Swindlehurst, Astely (2000) y en la obra de Cebrián y otros (2010).

A diferencia del caso de *slow fading*, ahora la curva S sí que es la suma de las curvas S de cada una de las componentes de la señal, cada una afectada por el retardo y la amplitud correspondientes. Como no hay cambio de signo en los términos que se suman, en este caso el retardo que se mide (que viene dado por el cruce por cero de la expresión 122) siempre será mayor que el de la señal directa. Un ejemplo de los errores que provoca una reflexión en el caso de *fast fading* se muestra en la figura 33. Los errores son menores que los dados por la envolvente del caso *slow fading*.

Figura 33. Error causado por una reflexión con SMR = 6 dB en el caso de *fast fading*



Efectos beneficiosos del multicamino

A pesar de los serios efectos que tiene el multicamino en la medida de los pseudorrangos, existen algunos casos en los que sorprendentemente las reflexiones y el multicamino pueden llegar a ser beneficiosos. No se trata obviamente de aplicaciones de posicionamiento convencionales, ya que los errores en los pseudorrangos se traducirían directamente a errores en la posición, sino de aplicaciones científicas de GNSS. Un ejemplo es el estudio del estado del mar (altura, rugosidad, salinidad, etc.). Las reflexiones de las señales GNSS sobre la superficie del mar pueden aportar información sobre dicho estado. A partir del retardo y la frecuencia Doppler de dichas reflexiones se puede llegar a inferir diferentes parámetros físicos. En este caso, las reflexiones son señales que se desea recibir.

Tradicionalmente, siempre se ha considerado que la ionosfera y el multicamino son los principales elementos que limitan la precisión de los sistemas GNSS. Hoy en día con la existencia de señales a varias frecuencias o con la disponibilidad de información precisa sobre la ionosfera, los errores ionosféricos se han cancelado o reducido en gran medida incluso en aplicaciones no diferenciales. Además, todos los errores originados en el segmento del control o de espacio del sistema también se han reducido por la mejora de los algoritmos y equipos en dichos segmentos. Esto deja al multicamino como la principal fuente de errores y a las interferencias como la principal fuente de vulnerabilidad de los sistemas GNSS. Tanto el multicamino como las interferencias son difícilmente controlables porque se pueden producir muy fácilmente, y tienen la capacidad de ocasionar errores muy grandes o incluso impedir el funcionamiento del receptor.

La tabla 4 muestra un resumen de los valores de los errores causados por diversas fuentes en "condiciones favorables". Por "condiciones favorables" nos referimos al caso en el que las diferentes fuentes de error están presentes (o sea, no son condiciones ideales), pero no de la peor manera posible. La tabla representa el nivel de errores que uno puede esperar cuando no se dan condiciones extremas. Los errores pueden variar notablemente (en general, aumentar) según el escenario. Por ejemplo, en el caso del multicamino ya hemos comentado que los errores pueden superar fácilmente la decena de metros.

Ved también

Podéis consultar también la tabla 5 en la que se resume el efecto de cada fuente de error en el caso convencional y cómo se reduce al utilizar alguna técnica de mitigación.

Tabla 4. Resumen de las diferentes fuentes de errores

| Fuente de error | Error (desviación típica en metros) |
|--|-------------------------------------|
| Reloj de satélite | 1,5 |
| Efemérides | 1,5 |
| Ionosfera (modelos para una frecuencia) | 2,50-25 |
| Troposfera | 0,20-0,40 |
| Ruido en el pseudorrango | 0,25-0,50 |
| Ruido en la fase | 0,001-0,002 |
| Multicamino en el pseudorrango (entorno benigno) | 0,25-0,50 |
| Multicamino en la fase (entorno benigno) | 0,005-0,01 |

5.5. Posicionamiento diferencial

En este subapartado vamos a presentar las bases del posicionamiento diferencial y las diferentes familias de técnicas que existen. Empezaremos analizando cómo y en qué medida se reducen las diferentes fuentes de errores en el subapartado 5.5.1. Los tres siguientes subapartados, desde el 5.5.2 al 5.5.4,

presentan diferentes formas de utilizar el concepto de posicionamiento diferencial, según si se persigue obtener un posicionamiento absoluto o relativo, o de si se utilizan únicamente los pseudorrangos o también las fases de portadora, para tener una precisión mayor. Un paso clave para poder utilizar el observable de fase es resolver la ambigüedad intrínseca que presenta (subapartado 5.5.5). Se introduce al estudiante en esta problemática utilizando las técnicas más sencillas, que se denominan técnicas libres de la geometría (subapartado 5.5.6), y seguidamente se pasa a las técnicas más avanzadas basadas en la geometría (subapartado 5.5.7). Finalmente, en el subapartado 5.5.8 se presenta brevemente una forma alternativa de posicionamiento preciso absoluto basado también en las fases. Aunque no es estrictamente una técnica diferencial, se presenta dentro de este subapartado porque tiene muchos elementos comunes con el resto de las técnicas tratadas.

5.5.1. Clasificación y reducción de las fuentes de error

La precisión que se obtiene con el servicio SPS de GPS (que implica el uso del pseudorrango del código C/A a una única frecuencia) es del orden de 10 m en posición y 20 ns en tiempo (ambos valores referidos a un percentil del 95%). Sin embargo, existen multitud de aplicaciones que requieren precisiones mucho mejores, que no son ni tan solo alcanzables mediante el uso de medidas a dos frecuencias. Un modo de reducir los errores en la posición y/o en la medida del tiempo consiste en reducir los errores presentes en los observables. Una alternativa consiste en utilizar las medidas de fase en lugar de pseudorrango, pero esto conlleva complicaciones adicionales debido a la presencia de las ambigüedades. Otra alternativa, que sirve tanto para las medidas de código como para las de fase, implica cambiar el modo como se utiliza el sistema GNSS, pasando del posicionamiento autónomo basado en un único receptor al posicionamiento diferencial.

Las alternativas no son excluyentes, sino todo lo contrario, puesto que las medidas de fase normalmente se utilizan en modo diferencial. En el caso particular de GPS, el modo de utilización se suele conocer como DGPS, que será el que asumiremos por defecto. No obstante, los principios del posicionamiento diferencial son idénticos en todos los sistemas GNSS, como veremos a continuación.

La idea básica que subyace en el DGPS es aprovechar el hecho de que los errores asociados al reloj del satélite, las efemérides y la atmósfera son similares para usuarios separados decenas, o incluso centenas de kilómetros, y estos errores varían lentamente con el tiempo. Es decir, estos errores exhiben una alta correlación espacial y temporal. Cuanto más cercanos están dos usuarios y más próximos se encuentran los instantes de medida, más similares son los errores mencionados más arriba. Se dice que los errores se decorrelan cuando aumenta la distancia entre los usuarios o la separación entre los instantes de cálculo de los observables.

Si la posición de un receptor GPS es conocida, en lugar de calcular realmente su posición (como se hace en la operación normal de GPS), se puede realizar el proceso inverso, consistente en estimar el efecto global de los errores antes mencionados en observables. Si estas estimaciones están disponibles para otros receptores, estos pueden utilizarlas para corregir sus observables, y de esta manera mitigar dichos errores y mejorar la calidad de la posición y del tiempo calculados. Esta idea es aplicable tanto a las medidas de fase como a los pseudorrangos. Veremos además que no es imprescindible conocer la posición del receptor GPS a partir del cual se calculan las correcciones. Este receptor se suele denominar estación de referencia. El concepto de DGPS se puede extender a varias estaciones de referencia. Dichas estaciones pueden transmitir diferentes tipos de información:

- Estimaciones de los errores, para que sirvan como correcciones a los observables (este caso corresponde al principio básico de funcionamiento que hemos comentado en los párrafos anteriores, pero no es la única opción).
- Datos que reemplacen al mensaje de navegación de los satélites en el cálculo del error de reloj del satélite y las efemérides.
- Los propios observables de código y fase.
- Información de integridad acerca de cada satélite (por ejemplo, una indicación sobre si se puede utilizar o no el satélite, o una medida estadística del nivel de precisión de las correcciones).
- Datos auxiliares, como la localización, la salud o las variables meteorológicas de las estaciones de referencia.

Tradicionalmente, esta información se ha transmitido mediante comunicaciones inalámbricas (por ejemplo, enlaces radio) en las bandas LF, MF, HF o UHF. El formato de la información ha sido habitualmente el definido por el estándar

dar RTCM SC-104, que define la estructura de los paquetes y el interfaz entre el receptor de los mensajes y el receptor GPS. La velocidad de transmisión necesaria suele ser muy modesta, alrededor de 100-200 bps.

Sin embargo, el avance de Internet y el uso masivo de terminales móviles con conexión de datos ha motivado el desarrollo de mecanismos para poder distribuir las correcciones GPS a través de redes de datos y, en particular, a través de Internet. Esta es la idea detrás del proyecto NTRIP.

NTRIP¹⁹ es un protocolo de comunicación que permite la distribución de flujos de datos GNSS desde estaciones de referencia hacia usuarios fijos o móviles, a través de Internet. La distribución de información se realiza en modo difusión o *broadcasting*, con lo que diversos dispositivos pueden acceder de manera simultánea a la información. NTRIP se ha desarrollado como un protocolo abierto, no propietario, basado en el estándar HTTP y que permite el acceso a dispositivos móviles mediante TCP/IP. Existen cientos de servicios a nivel mundial que utilizan los datos de corrección recibidos por NTRIP para implementar aplicaciones de navegación de alta precisión. Otros dos ejemplos de transmisión de correcciones a través de Internet son el sistema GDGPS (NASA's Global Differential GPS) y el sistema SISNET de la Agencia Espacial Europea, que permite acceder a los mensajes del sistema EGNOS. En el otro extremo de la sofisticación encontramos otro mecanismo para implementar DGPS, que aún se sigue utilizando hoy en día y que consiste en utilizar dos receptores GPS que graban por separado sus observables GPS. Más adelante, estos observables se vuelcan en un ordenador y allí se aplican los algoritmos DGPS.

⁽¹⁹⁾NTRIP son las siglas de *networked transport of RTCM via Internet protocol*.

proyecto NTRIP

NTRIP es un proyecto iniciado por la Agencia Federal Alemana de Cartografía y Geodesia (BKG) junto con la Universidad de Dortmund y la empresa Trimble Terrasat GmbH.

Sea cual sea el mecanismo de transmisión que se utilice, el aspecto crítico en todos los casos es la latencia, es decir, el período tiempo comprendido entre el instante al que corresponde la información diferencial y el instante en el que se aplica. Cuanto más cerca está el usuario de la estación de referencia y menor es la latencia, mejor funcionan las técnicas DGPS.

Observemos que el último mecanismo explicado en anteriormente es un mecanismo de posprocesado y por lo tanto puede tener un retraso importante en la generación de la posición, pero tiene una latencia en principio nula.

Las técnicas DGPS se pueden clasificar de diferentes maneras:

- **Absolutas o relativas.** En el posicionamiento absoluto diferencial se calcula la posición del usuario respecto a un sistema de coordenadas ECEF, mientras que en el posicionamiento relativo, la posición del usuario se calcula respecto a un punto de referencia cuyas coordenadas absolutas no tie-

nen por qué ser conocidas (si se conocen, entonces a partir de la posición relativa del usuario es evidente calcular también su posición absoluta).

- **De área local, regional o amplia**²⁰. Este criterio hace referencia al área a la que da servicio el sistema DGPS. Los más sencillos permiten que el usuario esté separado como máximo entre 10 y 100 km de la estación de referencia, y se denominan sistemas de área local. Cuando se pretende cubrir regiones mayores, es necesario utilizar varias estaciones de referencia. El término *regional* se suele reservar para distancias de hasta 1.000 km (aunque esto son valores meramente cualitativos), y por encima de este valor se habla de sistemas de área amplia, que pueden llegar a tener cobertura global.
- **Basadas en el código o en la fase**, en función del tipo de observables que se utilicen para el cálculo de la posición. Las técnicas diferenciales basadas en pseudorrangos pueden proporcionar precisiones en el rango de metros, mientras que las técnicas basadas en las fases alcanzan precisiones del orden de centímetros, pero se necesitan algoritmos notablemente más complejos.

⁽²⁰⁾En inglés, *local, regional or wide area*.

Vamos a revisar a continuación cada una de las fuentes de error que hemos visto en subpartados anteriores y cómo se mitigan en modo diferencial:

- **Reloj del satélite.** El error en el modelado del reloj de satélite ya es pequeño en condiciones normales (2 m rms, y se va reduciendo a medida que el sistema se moderniza), y cambia lentamente en un marco de tiempo de horas. Este error no se decorrela con la distancia por motivos evidentes, y se puede corregir prácticamente por completo si la latencia de los mensajes es inferior a una hora.
- **Efemérides.** También contienen un error pequeño (2 m rms, y va mejorando con la modernización de los sistemas) que cambia lentamente a nivel de minutos. Recordemos que prácticamente solo la componente radial del error de posición de los satélites causa un error en los pseudorrangos; y la componente radial es la proyección del error de posición del satélite sobre el vector unitario receptor-satélite. La dirección de este vector cambia con la posición del receptor. Por lo tanto, la componente radial del error en las efemérides no es exactamente la misma vista desde el usuario o desde la estación de referencia. No obstante, se ve a simple vista que el cambio de dirección es muy pequeño incluso cuando la separación entre usuario y estación de referencia es de centenares de kilómetros, dado que la distancia a los satélites es de 20.000 km. Se puede comprobar que la separación angular entre los vectores que apuntan a un satélite desde dos puntos en la Tierra separados 100 km es como máximo 0,3°. Una cota conservadora de la diferencia de errores en la componente radial de las efemérides es

$$\varepsilon_R \leq \frac{d \cdot \delta r}{r} \quad (123)$$

donde r es la distancia al satélite, δr es el error en las efemérides, y d es la distancia entre el usuario y la referencia.

Por ejemplo, para $d = 100$ km y $\delta r = 10$ m (que realmente es un valor muy alto porque como hemos dicho un valor razonable es 2 m rms), la contribución del error de las efemérides en posicionamiento diferencial es inferior a 5 cm.

- **Ionosfera.** Recordemos que el retardo ionosférico depende del contenido total de electrones (TEC) a lo largo del camino de propagación. Si consideramos dos receptores sobre la superficie terrestre separados una distancia d , la separación entre los caminos que siguen las señales que llegan a cada uno de los receptores en su tránsito por la ionosfera también es aproximadamente d (gráficamente se puede entender fácilmente que esto es así porque la ionosfera se concentra a unas alturas bastante inferiores a los 20.000 km de altura de los satélites). Por lo tanto, los errores diferenciales causados por la ionosfera dependerán de su correlación espacial. Un valor típico del error en la dirección cenital sería de 0,1-0,2 m (que se tiene que escalar adecuadamente para otras elevaciones) para $d = 100$ km, aunque existe una importante variabilidad dependiendo la actividad solar y de las tormentas electromagnéticas, y se pueden alcanzar valores superiores a 1 m. Todo esto implica que en técnicas diferenciales que utilizan los observables de fase y pretenden resolver las ambigüedades (esto lo estudiaremos más adelante), la distancia entre usuario y estación de referencia está limitada a 15-20 km para latitudes medias, que es donde se producen altos gradientes en la ionosfera.
- **Troposfera.** El retardo troposférico depende del perfil de densidad del aire a lo largo del camino de propagación. Dos receptores separados unos pocos kilómetros pueden estar sujetos a condiciones meteorológicas completamente diferentes. El contenido en vapor de agua muestra una alta variabilidad tanto temporal como espacial. Evidentemente, los errores diferenciales son mayores para satélites observados a una baja elevación, y un valor habitual es 0,1-0,2 m para distancias horizontales de 10 km. No obstante, estos errores pueden aumentar en un orden de magnitud si la separación entre los receptores viene dada por la altitud. A bajas elevaciones, el error residual por metro de diferencia en altitud es de 2-7 mm. Para separaciones horizontales superiores a 10 km o diferencias importantes en altitud, es preferible corregir por separado las medidas tanto en el receptor del usuario como en la estación de referencia utilizando un modelo troposférico que confiar en la cancelación de forma diferencial.
- **Multicamino, ruido e interferencias.** Estos errores son completamente locales y, por lo tanto, incorrelados entre el usuario y la estación de referencia. Hemos de observar que, a diferencia del multicamino y del ruido (que son completamente específicos para un receptor), una misma inter-

ferencia sí que puede afectar a los dos receptores. No obstante, aunque la fuente de la degradación sea común, los errores instantáneos que producirá en cada uno de los receptores serán en general diferentes. Esto significa que los errores del multicamino, ruido e interferencias no se pueden combatir de forma diferencial. Más bien viene a ser lo contrario, ya que la posición del usuario calculada está afectada por los errores tanto en las medidas la estación de referencia como en las medidas del propio usuario, y existen el doble de fuentes de errores debido al multicamino, ruido e interferencias que en el posicionamiento estándar. Por este motivo, en posicionamiento diferencial es fundamental realizar una buena selección de los emplazamientos del usuario y de la estación de referencia (al menos de esta última, puesto que puede ser que no se tenga control sobre el usuario, y además los errores en la estación de referencia afectarían a todos los usuarios que la utilizaran).

La tabla 5 resume la discusión previa, mostrando el efecto de cada fuente de error en el caso convencional y cómo se reduce al utilizar alguna técnica de mitigación.

Tabla 5. Comparación de las fuentes de error en posicionamiento estándar y diferencial

| Fuente de error | Error en los observables | Técnica de mitigación y tamaño del error |
|-------------------|--|--|
| Reloj de satélite | Error del modelo: 2 m (rms) | DGPS: cancelación perfecta |
| Efemérides | Error del modelo: 2 m (rms) | DGPS: 0,1 m (rms) |
| Ionosfera | En la dirección cenital: 2-10 m. Factor de oblicuidad: 1,8, 3 a 30° y 5° de elevación, respectivamente. Los errores en el pseudorrango y en la fase tienen la misma magnitud pero signos diferentes. | DGPS: 0,2 m (rms) Utilizando un receptor de una única frecuencia y el modelo de Klobuchar: 1-5 m. Receptor bifrecuencia: cancela el error ionosférico en un 99,9%, pero aumenta el efecto del ruido a 1 m (rms) |
| Troposfera | Retardo en la dirección cenital y a la altura de mar: 2,3-2,5 m. Disminuye con la altura. Factor de oblicuidad: 2, 4, 10 a 30°, 15° y 5° de elevación, respectivamente. Afecta por igual a pseudorrango y fase. | DGPS: 0,2 m (rms) más 2-7 mm por metro de diferencia en altura. Modelos basados en condiciones meteorológicas medias: 0,1-1 m. |
| Multicamino | En un entorno "limpio": Pseudorrango: 0,5-1 m Fase: 0,5-1 cm | Incorrelado entre receptores, no me puede mitigar con DGPS. Las técnicas de mitigación habituales son: diseño de la antena y emplazamiento adecuado, técnicas de procesado de señal en el receptor, <i>carrier-smoothing</i> , arrays de antenas. |

La comparación se realiza bajo la hipótesis de que la distancia entre receptores es de decenas de km y la latencia de decenas de segundos.

| Fuente de error | Error en los observables | Técnica de mitigación y tamaño del error |
|-----------------|--|---|
| Ruido | Pseudorrango: 0,25-0,5 m (rms) Fase: 1-2 mm (rms) | Incorrelado entre receptores, no se puede mitigar con DGPS. Su influencia se puede reducir mediante un diseño adecuado del receptor. |

La comparación se realiza bajo la hipótesis de que la distancia entre receptores es de decenas de km y la latencia de decenas de segundos.

Vamos a presentar a continuación dos formas básicas de posicionamiento DGPS utilizando medidas de pseudorrango en el usuario y en la estación de referencia, que los identificaremos en el resto del módulo con los subíndices u y r . Después pasaremos a tratar el posicionamiento utilizando las fases de la portadora.

5.5.2. Posicionamiento absoluto basado en el código

En el posicionamiento absoluto basado en el código es necesario conocer las coordenadas exactas de la estación de referencia: x_r . Partiendo de la notación del subapartado 5.3 y de la expresión 43, podemos expresar los pseudorrangos de los dos receptores para un determinado satélite como:

$$\begin{aligned} P_u &= r_u + c(\delta t_u - \delta t^s) + I_{\rho,u} + T_{\rho,u} + \varepsilon_{\rho,u} \\ P_r &= r_r + c(\delta t_r - \delta t^s) + I_{\rho,r} + T_{\rho,r} + \varepsilon_{\rho,r} \end{aligned} \quad (124)$$

Dado que la estación de referencia conoce su propia posición \mathbf{x}_r , puede calcular $r_r = \|\mathbf{x}^s - \mathbf{x}_r\|$ y obtener el error en los pseudorrangos como $e_r = r_r - P_r$, que resultará ser igual a

$$e_r = -c(\delta t_r - \delta t^s) - I_{\rho,r} - T_{\rho,r} - \varepsilon_{\rho,r} \quad (125)$$

El valor de e_r se retransmite a los usuarios, que lo utilizan para corregir sus pseudorrangos de la siguiente manera

$$\begin{aligned} P_{u,\text{corr}} &= P_u + e_r = r_u + c(\delta t_r - \delta t_u) + (I_{\rho,u} - I_{\rho,r}) + (T_{\rho,u} - T_{\rho,r}) + \varepsilon_{\rho,u} - \varepsilon_{\rho,r} \\ &\approx r_u + c(\delta t_r - \delta t_u) + \varepsilon_{\rho,u} - \varepsilon_{\rho,r} \end{aligned} \quad (126)$$

Así se obtiene un pseudorrango corregido donde ha desaparecido por completo el término del error del reloj de satélite, y en buena medida también los términos de la ionosfera y la troposfera si las medidas P_u y P_r se han tomado suficientemente próximas en tiempo y espacio. Vemos que efectivamente la expresión 126 incluye dos contribuciones de ruido. Aunque no se observa explícitamente en las ecuaciones, los errores debidos a las efemérides también se cancelan aproximadamente porque el efecto de un error en \mathbf{x}^s es aproximadamente el mismo en r_r y r_u (esto se entenderá mejor en el siguiente subapartado). La expresión 126 tiene el mismo aspecto que la de un pseudorrango

convencional y permite aplicar las soluciones de navegación que hemos visto para posicionamiento no diferencial, simplemente considerando un nuevo término de sesgo $\delta t'_u = \delta t_u - \delta t_r$, que deja de tener el significado del sesgo del reloj del receptor, pero que permite mantener la formulación y el número de incógnitas igual que en el posicionamiento no diferencial (es decir, el número de incógnitas no se ve incrementado en uno).

Existen numerosos sistemas DGPS en todo el mundo que fundamentalmente funcionan del modo que acabamos de describir. Muchos de ellos son comerciales (por ejemplo, el usuario puede suscribirse para recibir las correcciones) y otros se ofrecen por instituciones públicas. Por ejemplo, el US Coast Guard ofrece correcciones DGPS en todas las zonas costeras. Este sistema se ha extendido al resto del territorio de Estados Unidos en lo que se ha denominado el Nationwide DGPS (NDGPS).

5.5.3. Posicionamiento relativo basado en el código

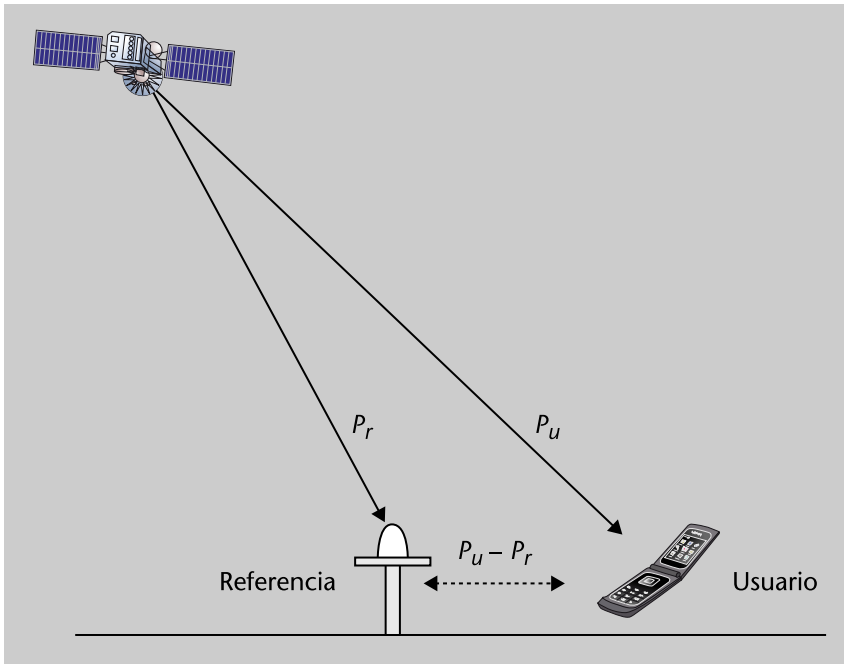
La técnica diferencial descrita en el subapartado anterior se puede también aplicar con ciertas modificaciones a la situación en la que se desconozca la posición de la estación de referencia. En este caso, se obtendrá la posición relativa del usuario con respecto a la referencia. Pero no es una limitación grave; además, en muchas aplicaciones realmente lo que interesa es obtener un posicionamiento relativo. Dos ejemplos de esto son los vuelos en formación o el aterrizaje de aviones en portaviones. En este último caso, difícilmente se puede conocer la posición del portaviones con precisión centimétrica, pero lo realmente importante es tener este nivel de precisión en la posición del avión relativa al portaviones cuando se vaya a producir el aterrizaje.

Si calculamos la diferencia de los dos observables en la expresión 124, tenemos

$$\begin{aligned} P_{ur} &\equiv P_u - P_r = (r_u - r_r) + c(\delta t_u - \delta t_r) + (I_{\rho,u} - I_{\rho,r}) + (T_{\rho,u} - T_{\rho,r}) + (\varepsilon_{\rho,u} - \varepsilon_{\rho,r}) \\ &\approx (r_u - r_r) + c\delta t'_u + \varepsilon'_{\rho,u} \end{aligned} \quad (127)$$

donde se han cancelado (al menos de manera aproximada) de nuevo los términos de la ionosfera y de la troposfera, y se han definido unos nuevos sesgos y ruidos equivalentes, indicados con el símbolo prima. Realizar la diferencia entre los observables medidos en dos receptores diferentes es una operación muy habitual en el procesado de datos GNSS, y el resultado (P_{ur} , en este caso) tiene nombre propio: diferencia simple (SD, *single difference*).

Figura 34. Formación de la diferencia simple



El aspecto clave aquí es que la diferencia $r_u - r_r$ depende en primera aproximación del vector diferencia $\mathbf{x}_{ur} \equiv \mathbf{x}_u - \mathbf{x}_r$, y no de los vectores \mathbf{x}_u y \mathbf{x}_r por separado, aunque esto no sea evidente a partir de la expresión:

$$r_u - r_r = \sqrt{(x_u - x^s)^2 + (y_u - y^s)^2 + (z_u - z^s)^2} - \sqrt{(x_r - x^s)^2 + (y_r - y^s)^2 + (z_r - z^s)^2} \quad (128)$$

Si aproximamos la primera raíz por su desarrollo de Taylor de primer orden alrededor de \mathbf{x}_r , obtenemos

$$r_u - r_r \approx -\frac{x^s - x_r}{r_r} x_{ur} - \frac{y^s - y_r}{r_r} y_{ur} - \frac{z^s - z_r}{r_r} z_{ur} \quad (129)$$

En la práctica, esta expresión depende de las coordenadas relativas del usuario con respecto a la referencia, (x_{ur}, y_{ur}, z_{ur}) , y no de las coordenadas precisas de la referencia porque los cosenos directores $(x^s - x_r)/r_r$, $(y^s - y_r)/r_r$, $(z^s - z_r)/r_r$, se pueden calcular para un valor aproximado de \mathbf{x}_r sin cometer un error perceptible. A partir de las diferencias simples en la expresión 127 y de la linealización en la 129, es inmediato plantear la solución de navegación para x_{ur} siguiendo los mismos pasos que en el posicionamiento estándar.

5.5.4. Posicionamiento relativo con fases de portadora

Los métodos anteriores, así como cualquier otro método de posicionamiento diferencial o relativo basado en el código, proporcionan precisiones alrededor de un metro. Para dar el salto a presiones centimétricas, es necesario procesar

las fases de portadora. Vamos a partir de las ecuaciones equivalentes de la expresión 124 pero planteadas para las fases de portadora para describir cómo utilizar estos observables para posicionamiento:

$$\begin{aligned} L_u^s &= r_u^s + c(\delta t_u - \delta t^s) - I_{\rho,u}^s + T_{\rho,u}^s + R_u^s + \varepsilon_{L,u}^s \\ L_r^s &= r_u^s + c(\delta t_r - \delta t^s) - I_{\rho,r}^s + T_{\rho,r}^s + R_u^s + \varepsilon_{L,r}^s \end{aligned} \quad (130)$$

donde hemos introducido dos modificaciones con respecto a la notación básica presentada en el subapartado 5.2. En primer lugar, expresamos las ecuaciones en metros, y por ello denotamos la fase como L , en lugar de Φ , que la reservamos para la fase en ciclos (no obstante, la diferencia entre una y otra es solo el escalado por la longitud de onda). En segundo lugar, hemos incluido el término R en lugar de simplemente un número entero de longitudes de onda porque para estudiar correctamente el posicionamiento con fases de portadora es necesario utilizar un modelo más preciso que el planteado en los subapartados introductorios. El término R incluye de hecho la ambigüedad en el número entero de longitudes de onda y además los desfases/retardos instrumentales que experimenta la señal en el satélite (k^s) y en el receptor (k_u):

$$\begin{aligned} R_u^s &= \lambda N_u^s + k^s + k_u \\ R_r^s &= \lambda N_r^s + k^s + k_r \end{aligned} \quad (131)$$

Para eliminar los errores comunes entre el usuario y la referencia, es habitual proceder al cálculo de las diferencias simples:

$$L_{ur}^s \equiv L_u^s - L_r^s = r_{ur}^s + c(\delta t_u - \delta t_r) + R_{ur}^s + \varepsilon_{L,ur}^s \quad (132)$$

En el término correspondiente a la ambigüedad y a los desfases instrumentales

$$R_{ur}^s = \lambda(N_u^s - N_r^s) + (k_u - k_r) = \lambda N_{ur}^s + k_{ur} \quad (133)$$

solo se ha podido eliminar el desfase del satélite. Para eliminar las fuentes de error o de indeterminación propias de cada uno de los receptores, cuando se trabaja con fases de portadora, es muy habitual plantear las dobles diferencias (DD, *double difference*), que se definen como la diferencia entre las SD correspondientes a dos satélites. Si los superíndices s y v denotan los dos satélites involucrados, la doble diferencia es

$$L_{ur}^{sv} \equiv L_{ur}^s - L_{ur}^v = r_{ur}^{sv} + R_{ur}^{sv} + \varepsilon_{L,ur}^{sv} \quad (134)$$

Nota

La misma discusión que la realizada para los pseudorrangos es válida aquí y las limitaciones en la cancelación de errores siguen siendo las de la tabla 5.

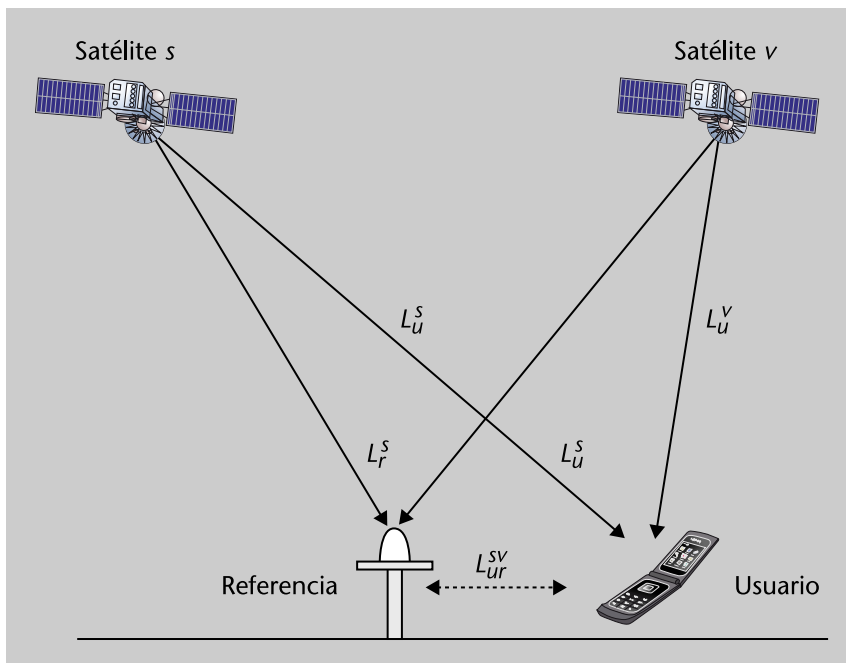
Gracias a esta segunda diferenciación se han conseguido eliminar los *offsets* de los relojes de los receptores y sus retardos instrumentales. Este último es el aspecto clave por el que es tan habitual trabajar con las diferencias dobles; es decir, al haber eliminado los retardos instrumentales, el término residual en las dobles diferencias de fase es un número entero de longitudes de onda:

$$R_{ur}^{sv} = \lambda(N_{ur}^s - N_{ur}^v) = \lambda(N_u^s - N_u^v - N_r^s + N_r^v) = \lambda N_{ur}^{sv} \quad (135)$$

Explotar este hecho es lo que finalmente permitirá obtener la máxima precisión en el posicionamiento. Si se conociera el valor del número de entero de longitudes de onda presente en los observables de fase, estos se comportarían como pseudorrangos (es decir, como medidas no ambiguas) con un nivel de ruido muy bajo²¹.

⁽²¹⁾Del orden de pocos milímetros, en contrapartida al nivel de ruido de un metro o algo inferior en los pseudorrangos propiamente dichos.

Figura 35. Formación de las diferencias dobles



Hemos presentado las dobles diferencias como el resultado de restar dos diferencias simples para dos satélites diferentes, mientras que las diferencias simples corresponden a la diferencia de los observables entre dos receptores y un satélite dado. No obstante, las dos dobles diferencias también se pueden obtener calculando la diferencia entre los observables para dos satélites y un receptor, realizando a continuación la diferencia para dos receptores. Quizá la expresión mediante fórmulas es más sencilla en este caso que el texto. Para una variable genérica *A* (que puede ser la fase, la ambigüedad, la distancia, etc.) se cumple:

$$\begin{aligned} A_{ur}^{sv} &= A_{ur}^s - A_{ur}^v = (A_u^s - A_r^s) - (A_u^v - A_r^v) \\ &= A_u^s - A_r^s - A_u^v + A_r^v = (A_u^s - A_u^v) - (A_r^s - A_r^v) \end{aligned} \quad (136)$$

Hemos de observar que al realizar cualquier operación de diferenciación el ruido "se acumula"; así, las SD contienen dos términos de ruido y las DD contienen cuatro términos de ruido. Bajo la hipótesis razonable de que los términos de ruido están incorrelados y tienen la misma potencia, podemos afirmar que al realizar un proceso de diferenciación la desviación típica del ruido aumenta en un factor $\sqrt{2}$. Por lo tanto, el nivel de ruido en las DD es el doble (o cuatro veces en potencia) que en los observables originales que lo forman. Este aumento del ruido no es crítico, especialmente en los observables de fase, porque es un precio relativamente pequeño que pagar por la cancelación de las otras fuentes de error que tienen valores mayores y que desaparecen gracias a la diferenciación.

A título de ejemplo, si suponemos que la desviación típica de los observables de código y fase es 0,5 metros y 5 milímetros respectivamente, al hacer las dobles diferencias, las desviaciones se convierten en 1 metro y 1 centímetro.

Las ecuaciones de navegación cuando se utilizan dobles diferencias siguen los mismos pasos que en el caso estándar de posicionamiento con pseudorrangos. El punto clave es la linealización de la distancia r_{ur}^{sv} en función del vector de posición relativa \mathbf{x}_{ur} . Como se ha visto en el subapartado 5.5.2, la diferencia simple de la distancia se puede expresar como

$$r_{ur}^s \approx -(\mathbf{v}_u^s)^T \mathbf{x}_{ur} \quad (137)$$

donde \mathbf{v}_u^s es el vector unitario que apunta de la estación de referencia al satélite (para ser más precisos, habría que decir que el punto de partida debería ser el punto medio de la línea de base, aunque para líneas de base más cortas que 10 km, no se produce ninguna variación notable al tomar cualquier punto de la línea de base). Por tanto, la doble diferencia de la distancia satisface

$$r_{ur}^{sv} \approx -(\mathbf{v}_u^s - \mathbf{v}_u^v)^T \mathbf{x}_{ur} \quad (138)$$

Ahora supongamos que disponemos de medidas a K satélites, entonces podemos plantear el siguiente sistema de ecuaciones lineales a partir de las dobles diferencias de fase:

$$\begin{bmatrix} L_{ur}^{21} \\ L_{ur}^{31} \\ \vdots \\ L_{ur}^{K1} \end{bmatrix} = \begin{bmatrix} -(\mathbf{v}_u^2 - \mathbf{v}_u^1)^T & \lambda & 0 & \cdots & 0 \\ -(\mathbf{v}_u^3 - \mathbf{v}_u^1)^T & 0 & \lambda & \cdots & \\ \vdots & \vdots & \vdots & \ddots & \ddots \\ -(\mathbf{v}_u^K - \mathbf{v}_u^1)^T & 0 & 0 & \cdots & \lambda \end{bmatrix} \begin{bmatrix} \mathbf{x}_{ur} \\ N_{ur}^{21} \\ N_{ur}^{31} \\ \vdots \\ N_{ur}^{K1} \end{bmatrix} \quad (139)$$

El vector de incógnitas contiene $K + 2$ variables (las tres coordenadas de la posición y las $K-1$ ambigüedades). Al utilizar diferencias dobles, el *offset* del reloj ya no aparece como incógnita, a diferencia del caso de posicionamiento no-diferencial (también conocido como *point positioning*).

Línea de base

La línea de base o *baseline* es el segmento que va del usuario a la estación de referencia.

Podemos observar que hemos identificado con el superíndice 1 el satélite elegido como referencia. En teoría, a partir de las medidas a K satélites, se pondrían plantear $K(K-1)$ diferencias dobles, pero en realidad no tiene sentido trabajar con más de $K-1$, ya que un conjunto más amplio de DD sería linealmente dependiente. Existen muchas maneras de generar $K-1$ dobles diferencias que sean linealmente independientes, pero es habitual utilizar un mismo satélite "de referencia" en todas ellas, y que este sea el de mayor elevación. Esto es así porque en general una mayor elevación conlleva un menor nivel de ruido, y es importante que el satélite "de referencia" sea el que proporciona un observable de más calidad ya que su ruido aparece en todas las DD y en todas las ecuaciones.

El sistema de ecuaciones de la expresión 139 está formado por $K-1$ ecuaciones y $K + 2$ incógnitas, por lo tanto, no es resoluble (es decir, no tiene solución única). Para contrarrestar este problema, lo que hay que hacer es añadir nuevas ecuaciones que no añadan el mismo número de incógnitas. Existen muchísimas maneras de hacer esto, de hecho se diría que infinitas, ya que hay tantas como usuarios o fabricantes, y cada uno puede proponer diferentes formas para acabar calculando la posición relativa precisa a partir de las fases. Vamos a describir diferentes alternativas, teniendo en cuenta que las soluciones prácticas, más allá de los detalles, utilizan una o varias de estas alternativas.

El sistema de ecuaciones de navegación anterior se puede modificar o ampliar de varias maneras:

- Incluyendo ecuaciones correspondientes a los pseudorrangos, que dependen de las incógnitas de posición, pero no añaden ambigüedades.
- Introduciendo medidas en diferentes instantes de tiempo (épocas, en la nomenclatura de GNSS). Si el usuario y la referencia son estáticos, esto no añade incógnitas de posición. Aunque este no sea el caso y en cada época el vector de coordenadas sea diferente, aún hay un efecto neto favorable porque el número de medidas por época es, en condiciones normales, mayor que el número de coordenadas. Estamos presuponiendo que las ambigüedades son constantes, o sea, que no cambian entre épocas, al menos durante períodos de tiempo relativamente largos. Esto es así siempre que no se produzcan *cycle-slips*. Cada vez que se produce un *cycle-slip*, se debe determinar un nuevo valor de la ambigüedad. Esto no es problemático si son poco frecuentes, porque en caso contrario la ventaja de procesar varias épocas se diluye. Cuando se utilizan medidas de diferentes épocas no es habitual incluirlas todas ellas en un sistema de ecuaciones, sino que lo normal es aplicar un método de estimación secuencial, como el método de mínimos cuadrados secuencial o su versión bayesiana, que es el filtro de Kalman. Esta última es la opción más extendida.
- Utilizando medidas a varias frecuencias. En teoría, esto se puede hacer de dos maneras. Una consiste simplemente en calcular las DD para cada fre-

cuencia e incluirlas como ecuaciones en el sistema. La otra consiste en calcular una determinada combinación (por ejemplo, la combinación *wide-lane*, que veremos más adelante, la combinación libre de ionosfera, etc.) a partir de los observables de una pareja receptor-satélite a dos frecuencias, y entonces construir la DD del observable combinación, que es la ecuación que finalmente se incluye en el sistema.

- Variando el carácter de las ambigüedades entre número reales o enteros. Si se resuelve el sistema de navegación bajo la suposición de que las ambigüedades son números reales, se obtiene lo que se denomina la solución flotante (*floating solution* y las *floating ambiguities*). Está claro que en este caso las ambigüedades contribuyen a la dimensión del espacio vectorial en el que se encuentran las incógnitas (y es bien conocido que se necesitan al menos tantas ecuaciones como la dimensión de dicho espacio para poder calcular la solución). Si se impone que las ambigüedades sean realmente números enteros, entonces se obtienen la solución fija (*fixed solution* o las *fixed ambiguities*). El proceso de determinar los valores enteros de las ambigüedades se denomina fijar la ambigüedades (*to fix the ambiguities*). Como se ha mencionado más arriba, al fijar las ambigüedades, el observable de fase se comporta como un pseudorange ultrapreciso (por ejemplo, con ruido del orden de pocos milímetros).

Cuando se impone en la resolución del sistema que las ambigüedades sean enteras, estas solo contribuyen al aumento de la dimensión del espacio de la incógnitas en una unidad (por ejemplo, una forma coloquial de expresarlo sería decir que solo consumen un grado de libertad). Esto implica que el sistema se podría resolver con tal de que el número de ecuaciones fuera igual o superior al número de incógnitas reales (en el caso de las DD que hemos planteado, las 3 coordenadas de la posición relativa) más uno (es decir, con cuatro ecuaciones se podría resolver el sistema de la expresión 139, al menos en teoría). Esto es así porque en ausencia de ruido, solo una combinación (o quizá varias, pero será fácilmente identificar la correcta) de número enteros haría que el sistema se cumpliera con igualdad. En presencia de ruido, en general el sistema no se cumplirá exactamente para ninguna combinación, sino que se cumplirá de manera aproximada para un conjunto de combinaciones, y el número de elementos de dicho conjunto aumentará a medida que el ruido aumente. Entonces en función del nivel de ruido en las ecuaciones, el número de ecuaciones necesarias para identificar las ambigüedades enteras se alejará más o menos del límite inferior que hemos mencionado. Y, en definitiva, la forma de contrarrestar el aumento del nivel de ruido es añadir más ecuaciones. Se puede interpretar de forma cualitativa que el hecho de tener más ecuaciones permite realizar una especie de "promediado" que reduce el ruido.

5.5.5. Resolución de las ambigüedades de fase

Cualquier tipo de posicionamiento diferencial preciso con fase de portadora y resolviendo las ambigüedades se engloba habitualmente bajo el término de posicionamiento RTK (*real-time kinematic position*). Se trata de una nomenclatura muy extendida, aunque supone en realidad cierto abuso del lenguaje porque originalmente el calificativo de *real-time* se refería a las técnicas en utilizaban un sistema de comunicaciones en tiempo real para transmitir las medidas de la estación de referencia y/o del usuario y calcular la posición también en tiempo real. Es lo opuesto a técnicas de posprocesado. El calificativo *kinematic* se asociaba a las técnicas en las que receptor del usuario no era necesariamente estático. Además, se acuñó el término *on-the-fly* (OTF), que indicaba que era posible resolver las ambigüedades mientras el usuario estaba en movimiento (esto es, que la fase de inicialización no tenía por qué ser estática).

Antes de pasar a un ejemplo concreto sobre cómo quedaría el sistema de ecuaciones de navegación al utilizar alguna de las alternativas que acabamos de plantear, hemos de comentar dos aspectos. En primer lugar, para obtener la máxima precisión (por ejemplo, centimétrica) del posicionamiento hay dos condiciones necesarias:

1) Se debe fijar el valor de las ambigüedades. Si se deja que las ambigüedades floten, se puede afirmar en términos generales que la precisión estará en el orden del decímetro.

2) Es necesario que el modelo que relaciona los observables de fase con las ambigüedades y el resto de las variables sea preciso al nivel de una fracción de centímetro. Es decir, no se puede pretender fijar el valor de las variables enteras y obtener una precisión centimétrica cuando el propio modelo no alcanza ese nivel de exactitud. Este aspecto hace referencia fundamentalmente a los errores ionosféricos y troposféricos residuales presentes en las medidas y las suposiciones correspondientes realizadas en el modelo. Por ejemplo, si estamos trabajando con dobles diferencias, podemos suponer que la contribución ionosférica se ha cancelado. Si esto es realmente cierto, no hay ningún problema. Pero realmente hemos visto anteriormente que el error ionosférico residual al realizar la diferencia de observables depende de la actividad de la ionosfera y de la distancia entre receptores, acumulándose un error (como máximo) de 2 cm cada 10 km de separación. Esto implica que en términos generales solo podemos decir que las diferencias simples o dobles cancelan el error ionosférico al nivel requerido para fijar las ambigüedades cuando la línea de base no supera los 10-20 km.

Se podría pensar que la ionosfera no debería ser un problema de cara a determinar las ambigüedades cuando se dispone de medidas a dos frecuencias porque en este caso se podría trabajar con las dobles diferencias del observable de fase libre de ionosfera:

$$L_{IF,ur}^{sv} = \frac{f_1^2}{f_1^2 - f_2^2} L_{L1,ur}^{sv} - \frac{f_2^2}{f_1^2 - f_2^2} L_{L2,ur}^{sv} = -(\mathbf{v}_u^s - \mathbf{v}_u^v)^T \mathbf{x}_{ur} + c \frac{f_1 N_{L1,ur}^{sv} - f_2 N_{L2,ur}^{sv}}{f_1^2 - f_2^2} \quad (140)$$

que se puede obtener como la combinación libre de ionosfera de las dobles diferencias (que corresponde a como se expresa en el segundo término de la igualdad anterior) o como la doble diferencia de las combinaciones libres de ionosfera. Para aligerar la notación, en la expresión anterior y en algunas de las siguientes obviamos el "ruido" que siempre aparece al final del modelo.

Efectivamente es cierto que la contribución ionosférica se cancela completamente (entendido como al menos el 99,9%, que es el porcentaje del retardo ionosférico que es inversamente proporcional a la frecuencia al cuadrado) en $L_{IF,ur}^{sv}$ y esto ocurre incluso si $L_{L1,ur}^{sv}$ y $L_{L2,ur}^{sv}$ incluyen una contribución de la ionosfera debida a la decorrelación espacial causada por la separación entre usuario y referencia. El problema es entonces que en $L_{IF,ur}^{sv}$ la ambigüedad no se puede expresar como un cierto múltiplo de un número entero y, por lo tanto, no es posible aplicar directamente los algoritmos de resolución de ambigüedades a las combinaciones libres de ionosfera.

En definitiva, existe una limitación intrínseca en la determinación de la ambigüedad y la cancelación simultáneamente de la ionosfera. La cancelación de la ionosfera destruye el carácter entero de las ambigüedades, mientras que conservar la naturaleza entera de las ambigüedades conlleva mantener un error ionosférico residual en los observables. Esta limitación se hace manifiesta especialmente cuando las líneas de base son largas (por ejemplo, superan los 10-20 km). El posicionamiento diferencial con fases de portadora en áreas amplias y con el usuario en movimiento, que se ha denominado WARTK (*wide area real-time kinematic*), es un área activa de investigación. Se han propuesto técnicas en esta dirección que se basan en utilizar fuentes adicionales de información ionosférica precisa (Hernández-Pajares y otros; 2004). La descripción de estas técnicas está fuera del alcance de esta asignatura.

Ejemplo

Una vez realizada esta discusión, vamos a pasar a plantear un ejemplo de cómo quedaría el sistema de ecuaciones en unas condiciones concretas. Supondremos que disponemos de medidas de código y fase, a dos frecuencias, y en una única época. En este caso, las ecuaciones de posicionamiento resultan ser:

$$\begin{pmatrix} P_{L1,ur}^{21} \\ \vdots \\ P_{L1,ur}^{K1} \\ L_{L1,ur}^{21} \\ \vdots \\ L_{L1,ur}^{K1} \\ P_{L2,ur}^{21} \\ \vdots \\ P_{L2,ur}^{K1} \\ L_{L2,ur}^{21} \\ \vdots \\ L_{L2,ur}^{K1} \end{pmatrix} = \begin{pmatrix} -(\sqrt{v_u^2 - v_u^1}) & f_1^{-2} & \dots & 0 & 0 & \dots & 0 & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ -(\sqrt{v_u^k - v_u^1}) & 0 & \dots & f_1^{-2} & 0 & \dots & 0 & 0 & \dots & 0 \\ -(\sqrt{v_u^2 - v_u^1}) & -f_1^{-2} & \dots & 0 & \lambda_1 & \dots & \dots & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots \\ -(\sqrt{v_u^k - v_u^1}) & 0 & \dots & -f_1^{-2} & \dots & \dots & \lambda_1 & 0 & \dots & 0 \\ -(\sqrt{v_u^2 - v_u^1}) & -f_2^{-2} & \dots & 0 & 0 & \dots & 0 & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ -(\sqrt{v_u^k - v_u^1}) & 0 & \dots & -f_2^{-2} & 0 & \dots & 0 & 0 & \dots & 0 \\ -(\sqrt{v_u^2 - v_u^1}) & -f_2^{-2} & \dots & 0 & 0 & \dots & 0 & \lambda_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \\ -(\sqrt{v_u^k - v_u^1}) & 0 & \dots & -f_2^{-2} & 0 & \dots & 0 & 0 & \dots & \lambda_2 \end{pmatrix} \begin{pmatrix} x_{ur} \\ b_{ur}^{21} \\ \vdots \\ b_{ur}^{K1} \\ N_{L1,ur}^{21} \\ \vdots \\ N_{L1,ur}^{K1} \\ N_{L2,ur}^{21} \\ \vdots \\ N_{L2,ur}^{K1} \end{pmatrix} \tag{141}$$

Por completitud hemos incluido explícitamente el retardo ionosférico, modelado por los parámetros b_{ur}^{sv} dividido por las frecuencias al cuadrado. Si se considera que el término ionosférico está suficientemente atenuado por las dobles diferencias, entonces simplemente hay que eliminar la variable correspondiente en las ecuaciones. A partir de aquí es inmediato deducir cómo se podrían obtener los sistemas correspondientes a otros casos, consistentes por ejemplo en utilizar una sola frecuencia o más de dos frecuencias, varias épocas, etc. El sistema anterior consta de $4(K - 1)$ ecuaciones con $3 + (K - 1)$ incógnitas reales y $2(K - 1)$ incógnitas que pueden ser enteras o reales, con lo que si $K \geq 4$ es en teoría resoluble considerando que las ambigüedades son número reales. No obstante, que el sistema sea resoluble no garantiza que la determinación de las ambigüedades enteras vaya a tener éxito porque esto último depende de forma genérica del nivel de ruido presente en las medidas con respecto a la magnitud que representa un cambio de una unidad en las ambigüedades ("tener éxito" significa en este contexto que la probabilidad de que el número entero calculado coincida con el real es alta). Analizaremos esto con un poco más de detalle a continuación.

La determinación de las ambigüedades ha recibido mucha atención casi desde los inicios de GPS, siendo el primer trabajo el de Counselman y Gourevitch (1981), y desde entonces se han publicado cientos de artículos. Una posible clasificación de todos los métodos propuestos consiste en agruparlos en las categorías de *geometry-free* o *geometry-based*.

Los métodos libres de la geometría (*geometry-free*) no explotan la dependencia de la distancia geométrica con las coordenadas del receptor, sino que consideran que cada distancia es una variable independiente. Se determina la ambigüedad (o ambigüedades si hay medidas a dos frecuencias) para cada satélite de forma independiente a las del resto de los satélites (por ello en ocasiones estos métodos también se denominan métodos de búsqueda en el dominio de las medidas o métodos uno a uno). Los métodos basados en la geometría (*geometry-based*) pretenden determinar todas las ambigüedades en conjunto explotando el hecho de que existe una ligadura entre ellas dado por la dependencia de todas las distancias con las coordenadas del receptor. La mayor parte de la investigación en los últimos años se ha centrado en este tipo de técnicas, ya que al aprovechar la dependencia entre las medidas son las que proporcio-

nan mejores prestaciones, aunque también conllevan en general una mayor complejidad. No obstante, vamos a empezar explicando algunos casos representativos de técnicas *geometry-free* no solo porque ayudan a comprender las otras técnicas, sino porque en ocasiones también se pueden aplicar como un paso previo a la utilización de una técnica *geometry-based*.

5.5.6. Métodos de resolución de la ambigüedad *geometry-free*

La idea básica de las técnicas *geometry-free* queda recogida en este ejemplo sencillo. Partimos de una medida de pseudorrango y de fase a una determinada frecuencia, y consideramos que utilizamos las dobles diferencias para eliminar o mitigar las fuentes de errores comunes (por ejemplo, consideraremos aquí de momento que los errores ionosféricos residuales son despreciables). Podemos expresar las medidas como:

$$\begin{aligned} P_{ur}^{sv} &= r_{ur}^{sv} + \varepsilon_p \\ L_{ur}^{sv} &= r_{ur}^{sv} + \lambda N_{ur}^{sv} + \varepsilon_L \end{aligned} \quad (142)$$

donde los términos ε representan los ruidos de medida (fundamentalmente el ruido térmico y posiblemente interferencias y *multipath*). A partir de estas ecuaciones, se puede estimar la ambigüedad entera como

$$\hat{N} = \text{round}\left\{\frac{L_{ur}^{sv} - P_{ur}^{sv}}{\lambda}\right\} \quad (143)$$

donde $\text{round}\{\}$ es simplemente la operación de redondeo. Se trata de un estimador tremendamente sencillo, pero sus prestaciones no son satisfactorias porque el ruido presente en P_{ur}^{sv} tiene una desviación típica del orden de 1 metro, y dado que el ruido en L_{ur}^{sv} es mucho menor (desviación típica de 1 centímetro) y que la longitud de onda es $\lambda = 20$ cm (o un valor similar), la desviación típica de $(L_{ur}^{sv} - P_{ur}^{sv})/\lambda$ es del orden de 5 unidades. No se puede esperar que al cuantificar en niveles separados una unidad una magnitud cuyo nivel de ruido es 5 se obtenga el valor correcto. Para conseguir una probabilidad de error suficientemente baja, la desviación típica de la cantidad que se va redondear debería ser al menos inferior a 0,2 ciclos. Una forma de reducir la varianza sería promediar las medidas en diferentes épocas. Con los números anteriores deberíamos promediar 625 valores observables, que implica un tiempo de observación de al menos 625 segundos, dado que el tiempo de correlación entre observables en un modo de funcionamiento normal del receptor es de al menos de un segundo. En cualquier caso, este ejemplo revela que la resolución de las ambigüedades no es un problema cuando se dispone de tiempo y datos en abundancia; el reto es realizarlo en unas pocas épocas (idealmente en una sola).

Nomenclatura

A lo largo de este subapartado utilizaremos el símbolo "acento circunflejo" $\hat{}$ para denotar los "flotantes", y el símbolo "sombreado" $\hat{}$ para los valores "fijos".

Podemos extender el ejemplo anterior al caso de dos frecuencias. Las medidas se pueden escribir entonces como

$$\begin{bmatrix} P_{L1,ur}^{sv} \\ L_{L1,ur}^{sv} \\ P_{L2,ur}^{sv} \\ L_{L2,ur}^{sv} \end{bmatrix} = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & -1 & 1 & 0 \\ 1 & \gamma & 0 & 0 \\ 1 & -\gamma & 0 & 1 \end{bmatrix} \begin{bmatrix} r_{ur}^{sv} \\ I_{L1,ur}^{sv} \\ \lambda_{L1} N_{L1,ur}^{sv} \\ \lambda_{L2} N_{L2,ur}^{sv} \end{bmatrix} \quad (144)$$

donde $\gamma = f_1^2 / f_2^2$. Hemos incluido explícitamente el efecto ionosférico en las ecuaciones, aunque también se podría haber omitido si se considera que está suficientemente atenuado por el cálculo de las dobles diferencias. Se puede calcular la inversa de la matriz del sistema anterior, que vale:

$$\begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & -1 & 1 & 0 \\ 1 & \gamma & 0 & 0 \\ 1 & -\gamma & 0 & 1 \end{bmatrix}^{-1} = \frac{1}{\gamma - 1} \begin{bmatrix} \gamma & 0 & -1 & 0 \\ -1 & 0 & 1 & 0 \\ -\gamma - 1 & \gamma - 1 & 2 & 0 \\ -2\gamma & 0 & \gamma + 1 & \gamma - 1 \end{bmatrix} \quad (145)$$

Por lo tanto, las estimaciones de los parámetros son²²:

⁽²²⁾Donde omitimos los subíndices y superíndices correspondientes a los receptores y a los satélites para aligerar la notación, pero está claro por el contexto que los observables se referirán a las dobles diferencias; ecuaciones similares se podrían plantear para los observables originales aunque no sea tan habitual.

$$\begin{aligned} \hat{r} &= \frac{\gamma}{\gamma - 1} P_{L1} - \frac{1}{\gamma - 1} P_{L2} = \frac{f_1^2 P_{L1} - f_2^2 P_{L2}}{f_1^2 - f_2^2} \\ \hat{I}_{L1} &= \frac{f_2^2}{f_1^2 - f_2^2} (P_{L2} - P_{L1}) \\ \hat{N}_{L1} &= \frac{1}{\lambda_1} (L_{L1} + \hat{I}_{L1} - \hat{r}) \\ \hat{N}_{L2} &= \frac{1}{\lambda_2} \left(L_{L2} + \frac{f_1^2}{f_2^2} \hat{I}_{L1} - \hat{r} \right) \end{aligned} \quad (146)$$

Observamos que la primera ecuación es simplemente la combinación de pseudorrangos libre de ionosfera, y la segunda ecuación es la combinación ionosférica. Podríamos pensar que es posible fijar las ambigüedades redondeando los resultados de la tercera y cuarta ecuaciones. Sin embargo, hacerlo así tiene los mismos inconvenientes que el método de una frecuencia planteado anteriormente. Como el nivel de ruido en \hat{r} y \hat{I}_{L1} es mucho mayor que las longitudes de onda, la probabilidad de cometer un error en el redondeo es muy alta. De hecho, conviene mencionar que el nivel de ruido en la combinación libre de ionosfera y en la combinación ionosférica son mayores que en los propios pseudorrangos, porque en dichas combinaciones los pseudorrangos aparecen

Frecuencias y longitudes de onda

Recordamos la relación entre las frecuencias y las longitudes de onda, aunque ya es bien conocido a estas alturas:

$$\lambda_{L1} = c / f_1, \quad \lambda_{L2} = c / f_2.$$

multiplicados por los factores $f_1^2/(f_1^2 - f_2^2)$ y $f_2^2/(f_1^2 - f_2^2)$, que son más grandes que 1. Para contrarrestar este problema, vamos a calcular la diferencia entre la ecuación tercera y cuarta en la expresión 146, que resulta en

$$\begin{aligned}\hat{N}_{L1L2} &\equiv \hat{N}_{L1} - \hat{N}_{L2} = \frac{1}{c}(f_1 L_{L1} - f_2 L_{L2}) + \frac{1}{c}\left(f_1 \hat{I}_{L1} - f_1 \hat{I} - \frac{f_1^2}{f_2} \hat{I}_{L1} + f_2 \hat{I}\right) \\ &= \frac{1}{c}(f_1 L_{L1} - f_2 L_{L2}) + \frac{1}{c} \frac{f_1 - f_2}{f_1 + f_2} (f_1 P_{L1} + f_2 P_{L2}) \\ &= \frac{1}{\lambda_w} \left(\frac{f_1 L_{L1} - f_2 L_{L2}}{f_1 - f_2} - \frac{f_1 P_{L1} + f_2 P_{L2}}{f_1 + f_2} \right)\end{aligned}\quad (147)$$

donde hemos utilizado la definición $\lambda_w = c/(f_1 - f_2)$, que se suele denominar como longitud onda *wide-lane* (porque es mucho mayor que λ_{L1} y λ_{L2}). Podemos definir la combinación *wide-lane* del observable de fase y la combinación *narrow-lane* del observable de código como:

$$\begin{aligned}LW &\equiv \frac{f_1 L_{L1} - f_2 L_{L2}}{f_1 - f_2} \\ PW &\equiv \frac{f_1 P_{L1} + f_2 P_{L2}}{f_1 + f_2}\end{aligned}\quad (148)$$

Entonces podemos reescribir la expresión 147 de la siguiente manera:

$$\hat{N}_{L1L2} = \frac{1}{\lambda_w} (LW - PW) \quad (149)$$

que es una expresión bien conocida en el ámbito de GNSS. Esta expresión se suele deducir habitualmente de forma heurística, es decir, simplemente sustituyendo en el término de la derecha de la expresión 149 las expresiones de los observables y comprobando que proporciona una estimación de la diferencia de ambigüedades. La diferencia $LW - PW$ suele denominarse como combinación Melbourne-Wübbena. No obstante, aquí hemos seguido un procedimiento que permite mostrar que efectivamente la expresión 149 se obtiene de la resolución del sistema de ecuaciones en 144. La combinación Melbourne-Wübb es muy interesante porque en ella desaparece completamente el efecto de la ionosfera. Es fácil comprobar que el efecto de la ionosfera es el mismo en LW y PW , y, por lo tanto, al hacer la diferencia desaparece.

La ventaja de este método es que ahora sí que podemos fijar la ambigüedad N_{L1L2} simplemente redondeando y con cierta probabilidad de éxito (o al menos, habrá que promediar los datos durante menos épocas para que obtener una cierta probabilidad de éxito):

$$\widehat{N}_{L1L2} = \text{round}\left\{\frac{1}{\lambda_w}(LW - PW)\right\} \quad (150)$$

porque el ruido en PW en relación con la longitud de onda no es tan grande como en los casos anteriores. Por un lado, esto ocurre porque la longitud de onda *wide-lane* es bastante grande, ya que corresponde a la frecuencia diferencia entre $L1$ y $L2$, que en el caso de GPS es 347,82 MHz y, por lo tanto, $\lambda_w = 86,2$ cm. Por otro lado, también sucede que el nivel de ruido en PW es mejor que en los pseudorrangos originales porque estos intervienen afectados por los factores $f_1/(f_1 + f_2)$ y $f_2/(f_1 + f_2)$, que son inferiores a 1.

Una vez obtenido el valor de \widehat{N}_{L1L2} se puede plantear dos opciones. Una consiste en utilizar LW como el observable de fase en el que basar el posicionamiento, ya que una vez determinado \widehat{N}_{L1L2} , está libre de ambigüedad. El problema de esta opción es que, a diferencia de lo que ocurre en PW , el nivel de ruido en LW es mayor que en las fases originales, ya que estas están afectadas por los factores $f_1/(f_1 - f_2)$ y $f_2/(f_1 - f_2)$, que son mayores que la unidad. La segunda opción consiste en fijar las ambigüedades en los observables de fase originales, que tienen un nivel de ruido inferior y que, por lo tanto, trasladarán un menor nivel de ruido a la posición.

Para calcular el valor de las ambigüedades originales podemos retomar el sistema de ecuaciones original (expresión 144), y sustituir una de las ecuaciones por la nueva ecuación $\widehat{N}_{L1L2} = N_{L1} - N_{L2}$. Si sustituimos por ejemplo la tercera ecuación, nos queda

$$\begin{bmatrix} P_{L1} \\ L_{L1} \\ \widehat{N}_{L1L2} \\ L_{L2} \end{bmatrix} = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & -1 & 1 & 0 \\ 0 & 0 & \lambda_{L1}^{-1} & -\lambda_{L2}^{-1} \\ 1 & -\gamma & 0 & 1 \end{bmatrix} \begin{bmatrix} r \\ I_{L1} \\ \lambda_{L1} N_{L1} \\ \lambda_{L2} N_{L2} \end{bmatrix} \quad (151)$$

La inversa de la matriz del sistema de ecuaciones es:

$$\begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & -1 & 1 & 0 \\ 0 & 0 & \lambda_{L1}^{-1} & -\lambda_{L2}^{-1} \\ 1 & -\gamma & 0 & 1 \end{bmatrix}^{-1} = \frac{1}{(1+\gamma)\lambda_{L1} - 2\lambda_{L2}} \begin{bmatrix} \gamma\lambda_{L1} - \lambda_{L2} & -\lambda_{L2} & \lambda_{L1}\lambda_{L2} & \lambda_{L1} \\ \lambda_{L1} - \lambda_{L2} & \lambda_{L2} & -\lambda_{L1}\lambda_{L2} & -\lambda_{L1} \\ (1-\gamma)\lambda_{L1} & (1+\gamma)\lambda_{L1} & -2\lambda_{L1}\lambda_{L2} & -2\lambda_{L1} \\ (1-\gamma)\lambda_{L2} & (1+\gamma)\lambda_{L2} & -(1+\gamma)\lambda_{L1}\lambda_{L2} & -2\lambda_{L2} \end{bmatrix} \quad (152)$$

Por lo tanto, las estimaciones de las ambigüedades originales se puede obtener como:

$$\begin{aligned} \widehat{N}_{L1} &= \frac{1}{\lambda_{L1}((1+\gamma)\lambda_{L1} - 2\lambda_{L2})} ((1-\gamma)\lambda_{L1}P_{L1} + (1+\gamma)\lambda_{L1}L_{L1} - 2\lambda_{L1}\lambda_{L2}\widehat{N}_{L1L2} - 2\lambda_{L1}L_{L2}) \\ \widehat{N}_{L2} &= \frac{1}{\lambda_{L2}((1+\gamma)\lambda_{L1} - 2\lambda_{L2})} ((1-\gamma)\lambda_{L2}P_{L1} + (1+\gamma)\lambda_{L2}L_{L1} - (1+\gamma)\lambda_{L1}\lambda_{L2}\widehat{N}_{L1L2} - 2\lambda_{L2}L_{L2}) \end{aligned} \quad (153)$$

Podemos reescribir estas expresiones de la siguiente manera:

$$\begin{aligned}\widehat{N}_{L1} &= \frac{1}{\lambda_{L1}} \left(\frac{\lambda_{L1}^2 - \lambda_{L2}^2}{(\lambda_{L1} - \lambda_{L2})^2} P_{L1} + \frac{\lambda_{L1}^2 - \lambda_{L2}^2}{(\lambda_{L1} - \lambda_{L2})^2} L_{L1} - \frac{2\lambda_{L1}^2}{(\lambda_{L1} - \lambda_{L2})^2} (\lambda_{L2} \widehat{N}_{L1L2} + L_{L2}) \right) \\ \widehat{N}_{L2} &= \frac{1}{\lambda_{L2}} \left(\frac{\lambda_{L1}^2 - \lambda_{L2}^2}{(\lambda_{L1} - \lambda_{L2})^2} \lambda_{L1} P_{L1} + \frac{\lambda_{L1}^2 + \lambda_{L2}^2}{(\lambda_{L1} - \lambda_{L2})^2} \lambda_{L2} L_{L1} - \frac{2(\lambda_{L1}^2 + \lambda_{L2}^2)\lambda_{L2}}{(\lambda_{L1} - \lambda_{L2})^2} \widehat{N}_{L1L2} - \frac{2\lambda_{L1}\lambda_{L2}}{(\lambda_{L1} - \lambda_{L2})^2} L_{L2} \right)\end{aligned}\quad (154)$$

donde se observa que el factor que multiplica a los psedorrangos es:

$$\left| \frac{\lambda_{L1}^2 - \lambda_{L2}^2}{(\lambda_{L1} - \lambda_{L2})^2} \right| = \frac{f_{L1}^2 - f_{L2}^2}{(f_{L1} - f_{L2})^2} = \frac{f_{L1} + f_{L2}}{(f_{L1} - f_{L2})^2} = \frac{154 + 120}{154 - 120} = 8,06 \quad (155)$$

Se trata de un valor muy superior a la unidad, por lo que hace que el nivel de ruido en los pseudorrangos aún aumente mucho más con respecto a las longitudes de onda de las portadoras, y esto provoca que no sea fiable en absoluto obtener las ambigüedades enteras redondeando las ecuaciones en 154. Una alternativa consiste en despreciar la contribución de la ionosfera (suponiendo, por ejemplo, que ya ha sido suficientemente atenuada por el cálculo de las dobles diferencias). De esta manera, podemos eliminar una incógnita y una ecuación en la expresión 151, resultando en

$$\begin{bmatrix} L_{L1} \\ \widehat{N}_{L1L2} \\ L_{L2} \end{bmatrix} = \begin{bmatrix} 1 & 1 & 0 \\ 0 & \lambda_{L1}^{-1} & -\lambda_{L2}^{-1} \\ 1 & 0 & 1 \end{bmatrix} \begin{bmatrix} r \\ \lambda_{L1} N_{L1} \\ \lambda_{L2} N_{L2} \end{bmatrix} \quad (156)$$

cuya solución es

$$\begin{aligned}\widehat{N}_{L1} &= \frac{1}{\lambda_{L2} - \lambda_{L1}} (L_{L2} - L_{L1} + \lambda_{L2} \widehat{N}_{L1L2}) \\ \widehat{N}_{L2} &= \frac{1}{\lambda_{L2} - \lambda_{L1}} (L_{L2} - L_{L1} + \lambda_{L1} \widehat{N}_{L1L2})\end{aligned}\quad (157)$$

En estas ecuaciones no aparecen los pseudorrangos, por lo que el nivel de ruido es mucho menor y, por lo tanto, es factible obtener las ambigüedades enteras bien redondeando directamente dichas expresiones, o bien promediando un número no muy elevado de veces dichas expresiones y redondeado a continuación.

Otro algoritmo propuesto por Clyde C. Goad para obtener las ambigüedades consiste en los siguientes pasos (Strang; Borre; 1997):

$$\begin{aligned}K_2 &= \text{round}\{60\widehat{N}_{L1} - 77\widehat{N}_{L2}\} \\ \widehat{N}_{L2} &= \text{round}\left\{\frac{60\widehat{N}_{L1L2} - K_2}{77}\right\} \\ \widehat{N}_{L1} &= \widehat{N}_{L2} + \widehat{N}_{L1L2}\end{aligned}\quad (158)$$

donde \widehat{N}_{L1} y \widehat{N}_{L2} se obtienen con las expresiones 154. Los factores 60 y 70 surgen del hecho de que $\lambda_{L1}/\lambda_{L2} = 60/77$.

Las alternativas no se agotan aquí ni mucho menos porque el tema de la resolución de las ambigüedades ha recibido mucha atención en los últimos años. Así que podemos mencionar una variación más. Se trata de una alternativa a la expresión 150 mencionada en el artículo "Wide-Area, Carrier-Phase Ambiguity Resolution Using a Tomographic Model of the Ionosphere" que permite calcular \widehat{N}_{L1L2} como

$$\widehat{N}_{L1L2} = \text{round}\left\{\frac{1}{\lambda_w}(LW - L_{IF} + \hat{R}_{IF})\right\} \quad (159)$$

Podemos observar que hemos substituido PW por $L_{IF} - \hat{R}_{IF}$, donde L_{IF} es la combinación de fases libre de ionosfera y \hat{R}_{IF} es una estimación del sesgo debido a las ambigüedades en dicha combinación (no vamos a entrar en cómo se obtiene esta estimación). Está claro que la estimación del sesgo debe tener un error pequeño comparado con λ_w . La ventaja de la expresión 159 con respecto a la 150 es que en ella solo intervienen observables de fase (o sea, con menor ruido), pero tiene la desventaja de que los errores ionosféricos no se cancelan completamente en general. Esto se debe a que si bien la ionosfera está cancelada en los términos libres de ionosfera (L_{IF} y \hat{R}_{IF}), no tiene por qué estarlo completamente en L_w . En cambio, en la expresión 150, tanto LW como PW incluyen exactamente la misma contribución ionosférica, y al hacer la diferencia se cancela completamente.

Por último, vamos a presentar otra variación más porque el objetivo no es realizar una descripción exhaustiva de todas las técnicas que existen (lo que podría ser objeto de un curso completo), sino que el estudiante se familiarice con el tipo de razonamiento que hay detrás de las diferentes técnicas de resolución de ambigüedades.

A partir de la expresión 140 podemos observar que el sesgo en la combinación libre de ionosfera admite la expresión

$$R_{IF} = c \frac{f_1 N_{L1} - f_2 N_{L2}}{f_1^2 - f_2^2} \quad (160)$$

Si al mismo tiempo definimos $N_s = N_{L1} + N_{L2}$, entonces podemos expresar $N_{L1} = 0,5(N_s + N_{L1L2})$ y $N_{L2} = 0,5(N_s - N_{L1L2})$, que al sustituirlos en 160 resulta en

$$R_{IF} = \frac{1}{2} \lambda_n N_s + \frac{1}{2} \lambda_w N_{L1L2} \quad (161)$$

donde se ha definido $\lambda_n = c / (f_1 + f_2)$, que recibe el nombre de longitud de onda *narrow-lane* (en el caso de GPS L1-L2 toma el valor de 10,7 cm). Si se dispone de \hat{R}_{IF} y \hat{N}_{L1L2} en lugar de calcular las ambigüedades a las frecuencias individuales mediante la expresión 157, se puede calcular

$$\begin{aligned}\hat{N}_s &= \text{round}\left\{\frac{1}{\lambda_n}(2\hat{R}_{IF} - \lambda_w\hat{N}_{L1L2})\right\} \\ \hat{N}_{L1} &= \text{round}\left\{\frac{1}{2}(\hat{N}_s + \hat{N}_{L1L2})\right\} \\ \hat{N}_{L2} &= \hat{N}_{L1} - \hat{N}_{L1L2}\end{aligned}\quad (162)$$

La ventaja de este procedimiento con respecto a la ecuación 157 es que, gracias a disponer de \hat{R}_{IF} , no es necesario presuponer que los errores ionosféricos son despreciables (que es una suposición implícita en la ecuación 157), pero por otro lado se necesita una gran exactitud en \hat{R}_{IF} , puesto que λ_n es una longitud de onda muy pequeña.

Hemos obtenido la expresión 159 y las posteriores a ella de forma heurística, mientras que obtuvimos 150 de forma rigurosa partiendo del sistema de ecuaciones 144. Aunque ya no entraremos en los detalles, es fácil comprobar que dichas expresiones se pueden obtener también a partir de otros sistemas de ecuaciones. La ventaja de derivar las expresiones a partir de sistemas de ecuaciones en lugar de hacerlo de forma heurística es que las suposiciones realizadas para su obtención quedan claramente manifiestas. Por ejemplo, la expresión 159 se puede obtener resolviendo

$$\begin{bmatrix} L_{L1} \\ L_{L2} \\ \hat{R}_{IF} \end{bmatrix} = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & \lambda_n\lambda_w/\lambda_{L1}^2 & -\lambda_n\lambda_w/\lambda_{L2}^2 \end{bmatrix} \begin{bmatrix} r \\ \lambda_{L1}N_{L1} \\ \lambda_{L2}N_{L2} \end{bmatrix}\quad (163)$$

donde queda patente que la contribución de la ionosfera se ha despreciado. Esto nos permite recalcar una vez más que el problema de todas estas técnicas es que para acabar obteniendo las ambigüedades enteras en las portadoras originales, se termina en alguna de estas dos situaciones:

- Se supone despreciable la contribución de la ionosfera (que al trabajar con diferencias dobles deja de ser una buena aproximación para líneas de base de más de 10-20 km) y esto permite resolver las ambigüedades utilizando expresiones en las que no intervienen (al menos, directamente) los pseudorrangos (podéis ver por ejemplo la expresión 157 o la 162).
- Se considera explícitamente la presencia de la ionosfera, pero entonces los pseudorrangos aparecen en la operación de redondeo que conduce a las ambigüedades enteras (podéis ver por ejemplo 154). El alto nivel de ruido en los pseudorrangos comparado con la longitud de onda provoca que la operación de redondeo sea muy poco fiable. El primer paso consistente en el cálculo de \hat{N}_{L1L2} mediante la expresión 150 es bastante fiable porque λ_w

es grande, pero el problema surge en el cálculo de N_{L1} o N_{L2} porque N_{L2} y λ_{L1} son mucho más pequeñas que λ_{L2} .

Con el objetivo de superar las limitaciones reflejadas en los dos puntos anteriores, se propuso el uso de señales GNSS transmitidas en tres frecuencias portadoras (en lugar de en solo dos), lo que dio lugar a las técnicas llamadas TCAR (*three-carrier ambiguity resolution*). La idea fundamental que subyace es que el uso de tres frecuencias permite realizar el paso de las longitudes de onda *wide-lane* a las longitudes de onda de las portadoras de forma gradual, de manera que en cada paso la relación entre el ruido en los observables y la longitud de onda es menor que en el caso de utilizar solo dos frecuencias.

Vamos a describir a continuación de una manera cualitativa la lógica que hay detrás de los métodos TCAR. Conviene insistir en que se trata únicamente de una descripción cualitativa y no de un algoritmo que se aplique exactamente de esta manera en la práctica. Supongamos que tenemos medidas a tres frecuencias: L1, L2 y L5. Con estas tres frecuencias podemos formar otras tres frecuencias *wide-lane*, que, suponiendo que utilizamos el caso de GPS, toman los valores mostrados en la tabla 6.

Tabla 6. Frecuencias portadoras en GPS y sus combinaciones *wide-lane*

| Frecuencia (MHz) | Longitud de onda (cm) |
|----------------------------|-------------------------|
| $f_{L1} = 1.575,42$ | $\lambda_{L1} = 19,0$ |
| $f_{L2} = 1.227,60$ | $\lambda_{L2} = 24,4$ |
| $f_{L5} = 1.176,45$ | $\lambda_{L5} = 25,5$ |
| $f_{L1} - f_{L5} = 398,97$ | $\lambda_{w15} = 75,1$ |
| $f_{L1} - f_{L2} = 347,82$ | $\lambda_{w12} = 86,2$ |
| $f_{L2} - f_{L5} = 51,15$ | $\lambda_{w25} = 586,1$ |

En el primer paso se aprovecha que la combinación L2-L5 tiene una longitud de onda muy grande para calcular la diferencia entre las ambigüedades en L2 y L5 mediante la adaptación adecuada de la expresión 150:

$$\widehat{N}_{L2L5} = \text{round}\left\{\frac{1}{\lambda_{w25}}(LW_{25} - PW_{25})\right\} \quad (164)$$

donde la definición de cada uno de los términos es evidente a estas alturas:

$$\begin{aligned}
\lambda_{w25} &= c / (f_{L2} - f_{L5}) \\
LW_{25} &= (f_{L2}L_{L2} - f_{L5}L_{L5}) / (f_{L2} - f_{L5}) \\
PW_{25} &= (f_{L2}P_{L2} + f_{L5}P_{L5}) / (f_{L2} + f_{L5}) \\
N_{L2L5} &= N_{L2} - N_{L5}
\end{aligned} \tag{165}$$

En la descripción original del método (Forsell; Martín-Neira; Harris; 1997), en 164 se utiliza P_{L5} (o el pseudorrango a la frecuencia que corresponda) en lugar de PW_{25} ; la ventaja de utilizar esta última es que se cancelan los errores ionosféricos en este primer paso.

La probabilidad de elegir el entero adecuado (ecuación 164) es elevada porque λ_{w25} es varias veces superior al ruido en los pseudorrangos. A partir de la ambigüedad que acabamos de obtener, podemos deducir una medida de distancia mejorada (por ejemplo, una medida de fase con la ambigüedad fijada) como:

$$\widehat{L}_{L2L5} = LW_{25} - \lambda_{w25} \widehat{N}_{L2L5} \tag{166}$$

Ahora podemos repetir el proceso con la longitud de onda *wide-lane* siguiente en orden decreciente de valor. Obtenemos una nueva ambigüedad y una nueva distancia mejorada como:

$$\begin{aligned}
\widehat{N}_{L1L2} &= \text{round}\left\{\frac{1}{\lambda_{w12}}(LW_{12} - \widehat{L}_{25})\right\} \\
\widehat{L}_{L1L2} &= LW_{12} - \lambda_{w12} \widehat{N}_{L1L2}
\end{aligned} \tag{167}$$

Por último, pasamos a resolver la ambigüedad en una de las portadoras originales y obtenemos la medida de fase con la ambigüedad fijada:

$$\begin{aligned}
\widehat{N}_{L1} &= \text{round}\left\{\frac{1}{\lambda_{L1}}(L_{L1} - \widehat{L}_{L1L2})\right\} \\
\widehat{L}_{L1} &= L_{L1} - \lambda_{L1} \widehat{N}_{L1}
\end{aligned} \tag{168}$$

Variaciones de TCAR

La idea básica del TCAR ha dado lugar a múltiples variaciones: *integrated three carrier ambiguity resolution* (ITCAR), *factorized multi-carrier ambiguity resolution* (FAMCAR), *multiple carrier ambiguity resolution* (MCAR). De todos modos, todas las técnicas descritas anteriormente se enmarcan dentro de la categoría de técnicas *geometry-free*. La mejora notable en las técnicas de resolución de la ambigüedad se obtiene con las técnicas *geometry-based*, para las cuales también se han propuesto multitud de métodos.

5.5.7. Métodos de resolución de la ambigüedad *geometry-based*

Vamos a presentar el problema como un problema de mínimos cuadrados con enteros²³ con el objetivo de dar una descripción bastante genérica. Muchas de las técnicas propuestas se pueden interpretar como una manera particular de solucionar un problema ILS. Otra ventaja de la formulación ILS es que también es válida para las técnicas *geometry-free*.

⁽²³⁾En inglés, *integer least squares* (ILS).

Resolver el problema de la ambigüedad como un problema ILS es la manera óptima de hacerlo, en el sentido de que maximiza la probabilidad de encontrar la combinación correcta de ambigüedades. La dificultad de los problemas ILS es que son problemas combinatorios, lo que quiere decir que, para encontrar la solución exacta del problema en general, habría que evaluar todas las posibles combinaciones de las variables enteras, lo que computacionalmente puede resultar muy costoso. No obstante, como veremos, existen alternativas para realizar la búsqueda de las combinaciones de manera mucho más rápida. Algunas de estas alternativas son subóptimas y otras son óptimas (quiere decir que garantizan que se encuentra el mejor candidato en la región de búsqueda).

Podemos representar mediante un modelo muy general cualquier problema que nos aparece relacionado con la estimación de la posición y/o las ambigüedades a partir de los observables. Se trata del siguiente modelo:

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{A}\mathbf{z} + \mathbf{n} \quad (169)$$

donde $\mathbf{y} \in \mathcal{R}^n$ son las medidas, $\mathbf{x} \in \mathcal{R}^p$ son las incógnitas reales y $\mathbf{z} \in \mathcal{Z}^q$ son las incógnitas enteras. El ruido está representado por \mathbf{n} , y simplemente refleja que en general no se puede encontrar un ajuste perfecto entre las medidas y las variables del modelo. Justamente la dependencia de las medidas con estas variables viene dado por las matrices \mathbf{H} y \mathbf{A} , que son conocidas y dependen del problema concreto que se trate. Las medidas pueden ser los observables de código o de fase, cualquier tiempo de combinación de estos (diferencias simples, dobles, u otras), etc. Las incógnitas reales hacen normalmente referencia a las coordenadas, pero también pueden incluir parámetros relacionados con la propagación, el reloj, etc. Para que la dependencia con las coordenadas sea lineal, es necesario haber linealizado previamente la dependencia de la distancia con la posición. Las incógnitas enteras corresponden a las ambigüedades. Es fácil comprobar que todos los sistemas de ecuaciones presentados en el módulo se pueden obtener particularizando este modelo.

El objetivo es encontrar los valores de las variables reales y enteras que minimizan el error cuadrático entre las medidas y el modelo. Es decir, queremos encontrar los valores $\hat{\mathbf{x}}$ y $\hat{\mathbf{z}}$ que resuelven

$$\min_{\mathbf{x} \in \mathcal{R}^p, \mathbf{z} \in \mathcal{Z}^q} \|\mathbf{y} - \mathbf{H}\mathbf{x} - \mathbf{A}\mathbf{z}\|_{\mathbf{Q}_y}^2 \quad (170)$$

donde \mathbf{Q}_y es la matriz de covarianza de las medidas (si no se conoce, se supone que es la identidad). Este problema no admite la solución cerrada habitual de los problemas de mínimos cuadrados debido a la restricción de que \mathbf{z} debe ser entero. Para explicar cómo se pueden resolver los problemas ILS, vamos a encontrar algunas expresiones equivalentes del problema de la expresión 170.

El valor de \mathbf{x} que minimiza el error cuadrático para una \mathbf{z} fija es

$$\hat{\mathbf{x}}(\mathbf{z}) = (\mathbf{H}^T \mathbf{Q}_y^{-1} \mathbf{H})^{-1} \mathbf{H}^T \mathbf{Q}_y^{-1} (\mathbf{y} - \mathbf{A}\mathbf{z}) \quad (171)$$

Sumando y restando un mismo término en el error cuadrático, podemos expresarlo como

$$\begin{aligned} \|\mathbf{y} - \mathbf{H}\mathbf{x} - \mathbf{A}\mathbf{z}\|_{\mathbf{Q}_y}^2 &= \|(\mathbf{I} - \mathbf{H}(\mathbf{H}^T \mathbf{Q}_y^{-1} \mathbf{H})^{-1} \mathbf{H}^T \mathbf{Q}_y^{-1})(\mathbf{y} - \mathbf{A}\mathbf{z}) - \mathbf{H}(\mathbf{x} - \hat{\mathbf{x}}(\mathbf{z}))\|_{\mathbf{Q}_y}^2 \\ &= \|(\mathbf{I} - \mathbf{H}(\mathbf{H}^T \mathbf{Q}_y^{-1} \mathbf{H})^{-1} \mathbf{H}^T \mathbf{Q}_y^{-1})(\mathbf{y} - \mathbf{A}\mathbf{z})\|_{\mathbf{Q}_y}^2 + \|\mathbf{H}(\mathbf{x} - \hat{\mathbf{x}}(\mathbf{z}))\|_{\mathbf{Q}_y}^2 \\ &= \|\mathbf{P}_{\mathbf{Q}_y^{-1/2} \mathbf{H}}^\perp \mathbf{Q}_y^{-1/2} (\mathbf{y} - \mathbf{A}\mathbf{z})\|^2 + \|\mathbf{x} - \hat{\mathbf{x}}(\mathbf{z})\|_{(\mathbf{H}^T \mathbf{Q}_y^{-1} \mathbf{H})^{-1}}^2 \end{aligned} \quad (172)$$

donde hemos definido implícitamente la matriz de proyección ortogonal

$\mathbf{P}_{\mathbf{Q}_y^{-1/2} \mathbf{H}}^\perp$ como

$$\mathbf{P}_{\mathbf{Q}_y^{-1/2} \mathbf{H}}^\perp = \mathbf{I} - \mathbf{Q}_y^{-1/2} \mathbf{H} (\mathbf{H}^T \mathbf{Q}_y^{-1} \mathbf{H})^{-1} \mathbf{H}^T \mathbf{Q}_y^{-1/2} \quad (173)$$

y hemos utilizado que $\mathbf{P}_{\mathbf{Q}_y^{-1/2} \mathbf{H}}^\perp \mathbf{Q}_y^{-1/2} \mathbf{H} = 0$ para separar la norma inicial en dos normas. Suponiendo que \mathbf{z} pueda ser un vector real, el valor que minimiza el error cuadrático se obtiene a partir de la primera norma que aparece en la última expresión en la ecuación 172:

$$\hat{\mathbf{z}} = \left(\mathbf{A}^T \mathbf{Q}_y^{-1/2} \mathbf{P}_{\mathbf{Q}_y^{-1/2} \mathbf{H}}^\perp \mathbf{Q}_y^{-1/2} \mathbf{A} \right)^{-1} \mathbf{A}^T \mathbf{Q}_y^{-1/2} \mathbf{P}_{\mathbf{Q}_y^{-1/2} \mathbf{H}}^\perp \mathbf{Q}_y^{-1/2} \mathbf{y} \quad (174)$$

que se denomina como solución (*float*, en inglés). Podemos sumar y restar el término $\mathbf{A}\hat{\mathbf{z}}$ en la primera norma, resultando que esta se puede expresar como

$$\begin{aligned}
& \| \mathbf{P}_{\mathbf{Q}_y}^\perp \mathbf{H}^{-1/2} \mathbf{Q}_y^{-1/2} (\mathbf{Y} - \mathbf{A}\mathbf{z}) \|^2 \\
&= \| \mathbf{P}_{\mathbf{Q}_y}^\perp \mathbf{H}^{-1/2} \mathbf{Q}_y^{-1/2} (\mathbf{y} - \mathbf{A}\hat{\mathbf{z}}) + \mathbf{Q}_y^{1/2} \mathbf{P}_{\mathbf{Q}_y}^\perp \mathbf{H}^{-1/2} \mathbf{Q}_y^{-1/2} \mathbf{A} (\hat{\mathbf{z}} - \mathbf{z}) \|^2 \\
&= \| \mathbf{P}_{\mathbf{Q}_y}^\perp \mathbf{H}^{-1/2} \mathbf{Q}_y^{-1/2} \left(\mathbf{I} - \mathbf{A} \left(\mathbf{A}^T \mathbf{Q}_y^{-1/2} \mathbf{P}_{\mathbf{Q}_y}^\perp \mathbf{H}^{-1/2} \mathbf{Q}_y^{-1/2} \mathbf{A} \right)^{-1} \mathbf{A}^T \mathbf{Q}_y^{-1/2} \mathbf{P}_{\mathbf{Q}_y}^\perp \mathbf{H}^{-1/2} \mathbf{Q}_y^{-1/2} \right) \mathbf{y} \|^2 + \\
&+ \| \mathbf{Q}_y^{1/2} \mathbf{P}_{\mathbf{Q}_y}^\perp \mathbf{H}^{-1/2} \mathbf{Q}_y^{-1/2} \mathbf{A} (\hat{\mathbf{z}} - \mathbf{z}) \|^2_{\mathbf{Q}_y} \tag{175} \\
&= \| \left(\mathbf{I} - \mathbf{P}_{\mathbf{Q}_y}^\perp \mathbf{H}^{-1/2} \mathbf{Q}_y^{-1/2} \mathbf{A} \left(\mathbf{A}^T \mathbf{Q}_y^{-1/2} \mathbf{P}_{\mathbf{Q}_y}^\perp \mathbf{H}^{-1/2} \mathbf{Q}_y^{-1/2} \mathbf{A} \right)^{-1} \mathbf{A}^T \mathbf{Q}_y^{-1/2} \mathbf{P}_{\mathbf{Q}_y}^\perp \mathbf{H}^{-1/2} \mathbf{Q}_y^{-1/2} \right) \mathbf{P}_{\mathbf{Q}_y}^\perp \mathbf{H}^{-1/2} \mathbf{Q}_y^{-1/2} \mathbf{y} \|^2 + \\
&+ \| \hat{\mathbf{z}} - \mathbf{z} \|^2_{\mathbf{Q}_y}
\end{aligned}$$

donde el primer término del resultado es una constante (independiente de \mathbf{z} y \mathbf{x}) y donde se ha utilizado la definición de $\mathbf{Q}_z = \left(\mathbf{A}^T \mathbf{Q}_y^{-1/2} \mathbf{P}_{\mathbf{Q}_y}^\perp \mathbf{H}^{-1/2} \mathbf{Q}_y^{-1/2} \mathbf{A} \right)^{-1}$. Sustituyendo 175 en 172 obtenemos la expresión final del error cuadrático:

$$\| \mathbf{y} - \mathbf{H}\mathbf{x} - \mathbf{A}\mathbf{z} \|^2_{\mathbf{Q}_y} = cte + \| \hat{\mathbf{z}} - \mathbf{z} \|^2_{\mathbf{Q}_z} + \| \mathbf{x} - \bar{\mathbf{x}}(\mathbf{z}) \|^2_{(\mathbf{H}^T \mathbf{Q}_y^{-1} \mathbf{H})^{-1}} \tag{176}$$

La utilidad de esta expresión es que nos permite descomponer la solución del problema ILS en tres pasos:

Paso 1: Cálculo de la solución flotante de las variables enteras. Consiste simplemente en calcular la solución del problema sin la restricción de que una parte de las variables sean enteras, de manera que se puede resolver como un problema de mínimos cuadrados estándar. La solución flotante viene dada por la expresión 174. Alternativamente, también se puede encontrar $\hat{\mathbf{z}}$ resolviendo el problema original como un problema de mínimos cuadrados estándar, cuya solución satisface:

$$\begin{bmatrix} \mathbf{H}^T \mathbf{Q}_y^{-1} \mathbf{H} & \mathbf{H}^T \mathbf{Q}_y^{-1} \mathbf{A} \\ \mathbf{A}^T \mathbf{Q}_y^{-1} \mathbf{H} & \mathbf{A}^T \mathbf{Q}_y^{-1} \mathbf{A} \end{bmatrix} \begin{bmatrix} \hat{\mathbf{x}} \\ \hat{\mathbf{z}} \end{bmatrix} = \begin{bmatrix} \mathbf{H}^T \\ \mathbf{A}^T \end{bmatrix} \mathbf{Q}_y^{-1} \mathbf{y} \tag{177}$$

Podemos observar que el primer bloque de filas de la ecuación anterior implica que

$$(\mathbf{H}^T \mathbf{Q}_y^{-1} \mathbf{H}) \hat{\mathbf{x}} + (\mathbf{H}^T \mathbf{Q}_y^{-1} \mathbf{A}) \hat{\mathbf{z}} = \mathbf{H}^T \mathbf{Q}_y^{-1} \mathbf{y} \tag{178}$$

que es una relación útil para derivar la expresión 179.

Paso 2. Se minimiza el término $\| \hat{\mathbf{z}} - \mathbf{z} \|^2_{\mathbf{Q}_z}$ bajo la restricción de que \mathbf{z} sea entero. Este es el paso clave en la resolución del problema ILS porque esta minimización es en general compleja y la forma de realizarla es lo que diferencia

a unos métodos de otros. Más abajo comentaremos brevemente algunos de estos métodos. El resultado de este paso es el vector de variables enteras fijadas (*fixed solution*), denominado $\widehat{\mathbf{z}}$.

Paso 3. La minimización del último término en la expresión 176 fijando el valor de \mathbf{z} al encontrado en el paso anterior es directa, ya que no hay restricciones sobre \mathbf{x} . La solución final para las incógnitas reales teniendo en cuenta el valor fijado para las incógnitas enteras es:

$$\begin{aligned}\widehat{\mathbf{x}} &\equiv \widehat{\mathbf{x}}(\widehat{\mathbf{z}}) = (\mathbf{H}^T \mathbf{Q}_y^{-1} \mathbf{H})^{-1} \mathbf{H}^T \mathbf{Q}_y^{-1} (\mathbf{y} - \mathbf{A} \widehat{\mathbf{z}}) \\ &= \widehat{\mathbf{x}} - (\mathbf{H}^T \mathbf{Q}_y^{-1} \mathbf{H})^{-1} \mathbf{H}^T \mathbf{Q}_y^{-1} \mathbf{A} (\widehat{\mathbf{z}} - \widehat{\mathbf{z}})\end{aligned}\quad (179)$$

Podemos observar que en general $\widehat{\mathbf{x}}$ no coincide con $\widehat{\mathbf{x}}$ (que sería la solución del problema si todas las incógnitas fueran reales), sino que se le aplica una corrección que depende de la diferencia entre la solución *float* y *fixed* para \mathbf{z} . En el caso de posicionamiento, $\widehat{\mathbf{x}}$ podrían ser las coordenadas que se obtienen una vez que se han fijado las ambigüedades al valor $\widehat{\mathbf{z}}$.

La solución del paso 2 se puede ver como un mapeo entre \mathcal{R}^q y Z^q , de manera que a un cada vector real se le asigna un vector entero. Según el criterio óptimo de minimizar la distancia en la norma de \mathbf{Q}_z , las zonas de atracción o regiones en la que se divide \mathcal{R}^q tienen una forma que depende de \mathbf{Q}_z y que en general no facilitan determinar a qué región pertenece un vector real dado. Existen varias alternativas subóptimas para evitar la complejidad del mapeo óptimo, como el redondeo de todas las componentes del vector $\widehat{\mathbf{z}}$ independientemente (que equivale a utilizar como zonas de atracción hiper-cubos) o el método de *bootstrapping*, que consiste en resolver el problema de mínimos cuadrados $\min_{\mathbf{z} \in Z^q} \|\widehat{\mathbf{z}} - \mathbf{z}\|_{\mathbf{Q}_z}^2$ de forma secuencial, redondeando una componente en cada iteración.

El redondeo de todas las componentes independientemente es sin duda el método más sencillo pero solo es óptimo cuando la matriz \mathbf{Q}_z es diagonal. Justamente en esta observación se basa el método LAMBDA (*least-squares ambiguity decorrelation adjustment*) (Teunissen, 1995), que es posiblemente la técnica más rigurosa y con mejores prestaciones para abordar el problema de resolución ambigüedades en GNSS. El método LAMBDA consiste en resolver el problema ILS aplicando en el paso 2 una transformación (denominada transformación Z) para decorrelar las ambigüedades. De esta manera, la matriz de correlación resultante es aproximadamente diagonal y entonces se puede aplicar una técnica similar (pero no igual exactamente) al *bootstrapping* obteniendo muy buenas prestaciones. De hecho, está garantizado que el método LAMBDA siempre encuentra la solución óptima si esta se halla dentro de la región de búsqueda.

Si se tiene conocimientos de comunicaciones digitales, uno puede observar que el problema de resolver las ambigüedades en GNSS es idéntico al problema de detección de símbolos que pertenecen a una constelación en comunicaciones digitales. Este problema ha sido ampliamente estudiado, alcanzando un nivel de sofisticación muy importante. Las técnicas avanzadas consisten en dos pasos: una etapa de reducción de la rejilla de búsqueda (los algoritmos más conocidos son el LLL (Lenstra; Lenstra; Lovász; 1982) y el BKZ (Schnorr; Euchner; 1994), seguida de una búsqueda en árbol (utilizando por ejemplo el algoritmo Schnorr-Euchner (Schnorr; Euchner; 1994)). Todas estas técnicas de comunicaciones digitales se pueden trasladar al problema de resolución de ambigüedades (Hassibi; Boyd; 1998).

5.5.8. Posicionamiento absoluto con fases de portadora

Las técnicas RTK han sido el estándar de la industria del posicionamiento para obtener precisiones mejores que 10 cm. En estas técnicas es necesario disponer de un receptor de referencia, y la capacidad para resolver las ambigüedades viene determinada fundamentalmente por la distancia entre el usuario y la estación de referencia, y los efectos atmosféricos. En los últimos años ha aparecido una alternativa a RTK que aún está en una fase experimental pero que está atrayendo muchísimo interés. Se trata de las técnicas PPP (*precise point positioning*), que tienen la gran ventaja de no necesitar una estación de referencia y que tienen el objetivo de proporcionar una precisión similar a RTK (Bisnath; Gao; 2009). La base del funcionamiento de las técnicas PPP se basa en utilizar información de las órbitas y de los relojes de los satélites mucho más precisas que las transmitidas por los propios satélites. Esta información se puede obtener de sistemas que utilizan redes densas de receptores y modelos más precisos para estimar las órbitas y los relojes, y normalmente está disponible a través de Internet. La fuente de información más utilizada es el International GNSS Service (IGS). Los errores ionosféricos se eliminan utilizando medidas a dos frecuencias (aunque actualmente se está investigando sobre el PPP de una frecuencia).

6. Algoritmos de alta sensibilidad

En este apartado vamos a describir de manera sucinta las estrategias para aumentar la sensibilidad de los receptores GNSS y también un modo de implementar dichas estrategias.

Los sistemas GNSS han sido diseñados para funcionar correctamente en condiciones de visibilidad directa de los satélites. No obstante, existen muchos intereses comerciales y también cuestiones relacionadas con el cumplimiento de regulaciones (como la normativa E911 en Estados Unidos o la recomendación E112 en Europa) que provocan que sea interesante extender el funcionamiento de los receptores (aunque sea con una precisión en la posición mucho más pobre) a entornos más complicados. Estos entornos son básicamente entornos interiores y escenarios urbanos donde las señales pueden estar en mayor o menor medida bloqueadas por los edificios. El principal problema en estos entornos es que la potencia recibida es mucho menor que la esperada en condiciones nominales (por ejemplo, entornos exteriores) y, por lo tanto, para que los receptores funcionen se deben utilizar técnicas que permitan aumentar su sensibilidad. De aquí viene el término HS-GNSS (*high-sensitivity* GNSS) utilizado habitualmente, que fundamentalmente se refiere a todo el conjunto de algoritmos que se pueden utilizar para que los receptores GNSS funcionen con niveles de C/No varias decenas de dB (desde 10 dB hasta valores de 40 dB o incluso superiores) por debajo de los valores nominales.

Aparte de la integración con otros sistemas de posicionamiento, la clave para conseguir que el receptor funcione con niveles de potencia de señal más bajos consiste en que el receptor capte más energía de la señal recibida, y esto solo se puede conseguir correlando la señal durante intervalos de tiempo más largos. La primera opción para alargar el intervalo de correlación consiste en aumentar el número de acumulaciones o integraciones coherentes N_c . No obstante, existen varios motivos que impiden que N_c se pueda aumentar arbitrariamente:

1) Si la señal contiene datos²⁴, existe el riesgo de cancelación de la señal si hay transiciones de bit que ocurren durante la integración coherente de la señal. El valor medio de la degradación de la SNR debido a este efecto es

$$L_c = -10 \log_{10} \left(1 - \frac{T_{coh}}{3T_{bit}} \right) \quad (180)$$

Lectura recomendada

Este apartado solo es una introducción superficial al tema. Si estáis interesados podéis profundizar en él en la obra de Van Diggelen (2009) y en la de Seco-Granados; López-Salcedo; Jiménez-Baños; López-Risueño (2012).

Ved también

El concepto de integración coherente se explica en el subapartado 5.1 de este módulo.

⁽²⁴⁾ Hecho que no ocurre en las componentes piloto de los nuevos sistemas GNSS; podéis ver, por ejemplo, la descripción de Galileo en el módulo "Evolución de la navegación por satélite: Galileo y sistemas híbridos".

donde T_{coh} es la duración total de la integración coherente y T_{bit} es la duración de un bit. Existen alternativas para mitigar la pérdida de SNR causada por la falta de alineación entre los bits y los intervalos de correlación coherente. Una posibilidad es utilizar el *half-bit method* (Psiaki, 2001), que consiste en realizar integraciones coherentes durante intervalos de medio tiempo de chip. De cada par de intervalos consecutivos, solo se acaba eligiendo uno (el de mayor valor) y el otro se descarta, de manera que existe una gran probabilidad de elegir los intervalos que no sufren transiciones de chip y eliminar los que sí que las sufren. Otra alternativa consiste efectivamente en alinear las integraciones con los bits, lo que es lo mismo que calcular la sincronización de bit. Esta es la alternativa que proporciona mejores prestaciones a costa de una mayor complejidad.

Para extender la integración coherente más allá de la duración de un bit, es necesario conocer su valor. Esta información se puede incluir en información de asistencia que se puede enviar al receptor GPS a través de un sistema de comunicaciones móviles (típicamente un sistema WLAN o celular, como WiFi o UMTS/LTE). Este es el concepto denominado A-GNSS (Assisted GNSS). No obstante, no es muy habitual utilizar los bits que puede haber en la información de asistencia, porque para utilizar estos datos el receptor GPS debe utilizar continuamente el sistema de comunicaciones y se necesita una sincronización precisa entre los datos enviados por el sistema de asistencia y los datos que realmente están en la señal GPS. Pero sí que es muy habitual utilizar el concepto A-GNSS para enviar al receptor otro tipo de información, como valores aproximados del retardo y del Doppler de los satélites (lo que permite acelerar la adquisición y aumentar la sensibilidad), y los datos del mensaje de navegación (así puede calcular la posición incluso en el caso de que no haya recibido el mensaje directamente de la señal GPS debido a la atenuación) (Van Diggelen, 2009).

2) La señal recibida está afectada por errores residuales de frecuencia, y existe el riesgo de cancelación de la señal cuando T_{coh} excede un cierto valor. Este efecto se puede caracterizar analizando cuánto aumenta la SNR en función de T_{coh} y del error de frecuencia F_ϵ en relación con el caso de integrar coherente solo durante la duración de un código PN T_{code} :

$$\Delta SNR = 10 \log_{10} \left(\frac{T_{coh}}{T_{code}} \right) + 20 \log_{10} \left(\left| \frac{T_{code} \sin(\pi F_\epsilon T_{coh})}{T_{coh} \sin(\pi F_\epsilon T_{code})} \right| \right) \quad (181)$$

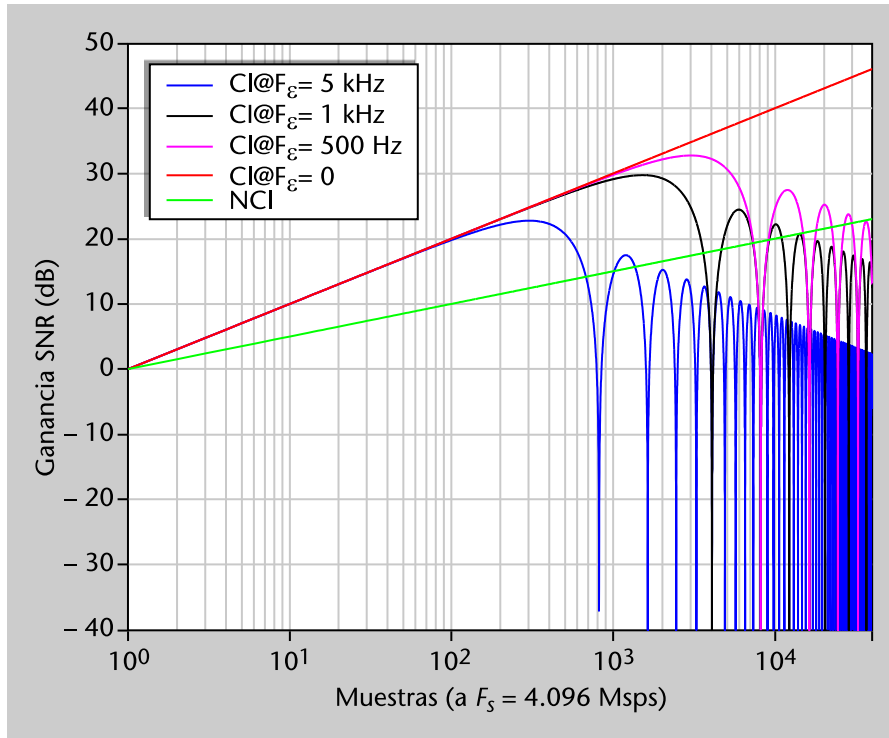
Esta expresión se dibuja en la figura 36, donde se puede observar que cuanto más pequeño es el error residual en frecuencia, más largo puede ser el intervalo de integración coherente antes de que la ganancia en SNR se reduzca debido a los plegados de fase durante la integración. Este efecto indica que existe un compromiso entre ganancia en SNR mediante integración coherente y

Ved también

En la segunda parte de este apartado explicaremos el método *doble-DFT*, que permite alinear las integraciones con los bits.

la complejidad computacional, porque para poder aumentar la ganancia hay que reducir la magnitud del error residual de frecuencia, y esto implica una búsqueda de frecuencia más fina, que a su vez implica más operaciones.

Figura 36. Ganancia en SNR para GPS de la integración coherente para diferentes errores de frecuencia y también de la integración no coherente



3) La presencia de ruido de fase debido fundamentalmente al oscilador del receptor también provoca, al igual que el error residual de frecuencia, una autocancelación parcial de la señal a medida que se van acumulando muestras coherentemente.

Incluso en el caso favorable de que la señal no tenga datos y de que no haya limitación en la carga computacional, la duración de la integración coherente estaría limitada por el ruido de fase del oscilador. Además, este ruido es un efecto importante en los receptores destinados a posicionamiento *indoor* porque estos receptores suelen ser teléfonos móviles (o similares) que por cuestiones de coste y tamaño disponen de osciladores con relativamente bastante ruido. Por ello, es inevitable recurrir a la combinación de integraciones coherentes y no coherentes para poder posicionar el receptor a niveles de C/N_0 por debajo de 25 dB-Hz.

Al utilizar ambos tipos de integraciones, la curva de correlación resultante se puede escribir como

$$R_{NC}(\tau, f) = \frac{1}{N_I} \sum_{k=0}^{N_I-1} |R_C(\tau, f; k)|^2 \quad (182)$$

Ved también

El concepto de integración no coherente se explica en el subapartado 5.1 de este módulo.

donde es el número de intervalos de integración coherente que se acumulan de forma no coherente, y la correlación coherente se define como

$$R_C(\tau, f; k) = \sum_{n=[KN_C N_{scode}]^{[(k+1)N_C N_{scode}-1]}} x(n)c(((1+\beta f)n-\tau)_{\text{mod}N_{scode}})e^{-j2\pi fn} \quad (183)$$

donde $x(n)$ es la señal recibida, $c(n)$ es la secuencia PN, N_{scode} es el número de muestras en un período de dicha secuencia, y β es el factor de proporcionalidad entre la frecuencia portadora y la frecuencia de chip de la señal. Por lo tanto, el tiempo de integración total (también conocido como *total dwell time*) es $T_{tot} = N_I N_C N_{scode} T_s$, siendo T_s el período de muestreo. El significado de $R_{NC}(\tau, f)$ es conceptualmente el mismo que el de la curva de correlación estándar mostrada en la figura 10, y una vez calculada sobre una rejilla de valores (τ, f) se procede de manera similar a como allí se explicó. Es decir, se determinan las estimaciones gruesas (en inglés, se suelen denominar *coarse estimates*) del retardo y de la frecuencia como el par de puntos de la rejilla en el que $R_{NC}(\tau, f)$ es máximo, siempre y cuando este máximo supere un umbral γ , que se fija de acuerdo con una probabilidad de falsa alarma. Si el máximo no supera el umbral, se considera que el satélite que se está buscando no está presente.

Hemos visto que hay limitaciones para aumentar el valor de N_C arbitrariamente, pero podríamos pensar que no hay problema en mejorar la sensibilidad tanto como queramos aumentando el valor de N_I . Como ya podemos intuir a estas alturas, esto tampoco va a ser así. En primer lugar, aumentar N_I también hace crecer la latencia que existe en el cálculo de la posición, y puede llegar un momento en el que sea excesiva para el tipo de uso que se quiere hacer. Pero lo que es más interesante, el error Doppler residual también tiene un efecto en las integraciones no coherentes, porque dicha frecuencia residual se traduce así mismo en un lento desplazamiento de la señal con respecto a la referencia. El hecho de que la frecuencia también afecta a la señal banda base se observa en el término de $c(((1+\beta f)n-\tau)_{\text{mod}N_{scode}})$ de la expresión 183, porque al igual que el efecto Doppler cambia la duración de los ciclos de portadora (y por tanto la frecuencia, como es bien conocido) también cambia la duración de los chips. La consecuencia de dicho desplazamiento es que curvas de correlación coherentes sucesivas se van desplazando con respecto a las anteriores, y al realizar la suma no coherente se "emborrona" la posición de pico. Para poder asegurar que el error que esto causa en la posición del pico no excede un valor $\Delta\tau$, es necesario que se cumpla $T_{tot} \leq 2 \Delta\tau / \varepsilon_f$, donde ε_f es el error relativo en la estimación de la frecuencia (es decir, el error de frecuencia, F_e , dividido por la frecuencia portadora nominal de la señal). El valor de ε_f se puede reducir haciendo la rejilla en la que se busca el máximo de $R_{NC}(\tau, f)$ más fina en la dimensión de la frecuencia. Por lo tanto, se observa un comportamiento si-

milar a lo que ocurría con la integración coherente, ahora también podemos aumentar las acumulaciones no coherentes, N_I , si se acepta una mayor complejidad computacional.

De todos modos, aunque consiguiéramos que el efecto del "emborronamiento" del pico de correlación debido a los errores residuales de frecuencia fueran despreciables, irían surgiendo nuevos efectos que limitarían el tiempo total de integración. De manera que, aunque en teoría se podría llegar a detectar la señal a niveles arbitrariamente bajos de C/N_0 , en la práctica surgen multitud de efectos que hacen muy complicado llegar a valores de C/N_0 inferiores a 10 dB-Hz. Incluso en el caso de que la frecuencia Doppler estuviera perfectamente estimada, la dinámica del reloj de receptor (por "dinámica" se entiende las derivas respecto a un reloj ideal) también distorsionaría el pico de correlación. Esto se puede mitigar mediante técnicas de estimación de la dinámica del reloj, pero esto no es factible en receptores de mano. La tecnología actual de relojes limita el tiempo de integración a 10 segundos como máximo. Está reconocido que el desarrollo de relojes con mejores prestaciones y de bajo coste permitirá realizar progresos en el posicionamiento GNSS de alta sensibilidad.

Otro ejemplo de estos efectos es el propio movimiento de los satélites y del receptor. Dado que el resultado de observar la señal durante T_{tot} es una única medida de distancia para cada satélite, pero el satélite y/o el receptor se han movido durante T_{tot} , no está nada claro a qué posición del satélite y del receptor corresponde la distancia calculada. Esta incertidumbre se traduce finalmente en incertidumbre en la posición calculada.

Una vez halladas las estimaciones gruesas de T y f , se deben encontrar las estimaciones finas. En los receptores convencionales esto se realiza mediante los algoritmos de seguimiento (o *tracking*), que también se pueden modificar para utilizarlos en HS-GNSS. No obstante, en este último caso también es muy habitual utilizar las llamadas implementaciones del tipo *snapshot* o *push-to-fix* o de adquisición continua, donde la idea fundamental es evitar la etapa de *tracking* y obtener las estimaciones precisas de T y f realizando una evaluación en una rejilla más fina. Este modo de funcionamiento encaja bien con las aplicaciones en entornos interiores, donde en muchas ocasiones puede no ser necesario disponer de la posición continuamente, sino que tiene sentido que el usuario solicite el cálculo de la posición solo en determinados momentos, cuando quiere realizar una determinada acción. En respuesta a la pulsación de un botón por parte del usuario, el terminal captura un conjunto de muestras de la señal y, a partir de ellas, calcula la posición y desconecta la funcionalidad de posicionamiento hasta que el usuario lo vuelve a solicitar.

6.1. Algoritmos de HS-GNSS basados en la DFT

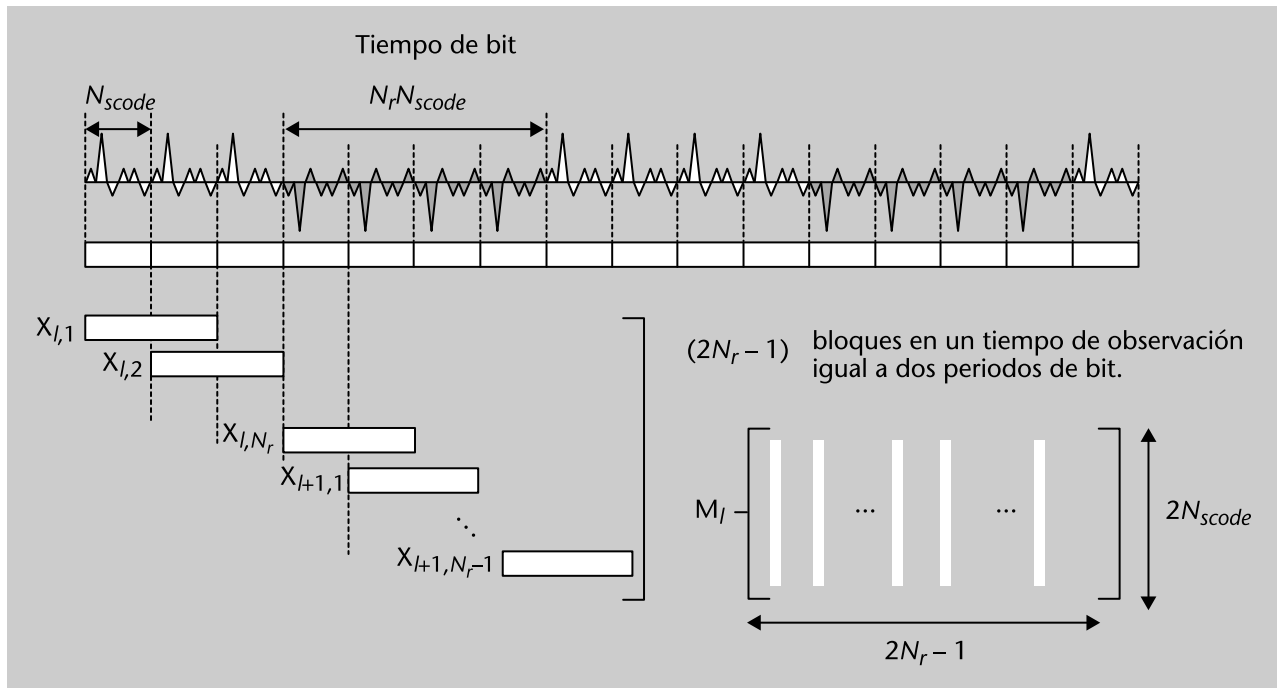
Vamos a describir a continuación una posible manera de realizar las integraciones coherentes y no coherentes de forma computacionalmente más sencilla que siguiendo literalmente la expresión 182. Esta implementación se ha

utilizado en la práctica en proyectos experimentales de la Agencia Espacial Europea, como se explica en la obra de López-Salcedo y otros (2008) y en la de Seco-Granados; López-Salcedo; Jiménez-Baños; López-Risueño (2012). La evaluación de la correlación entre la señal recibida y la réplica local es sin duda el proceso más costoso desde el punto de vista del número de operaciones en un receptor GNSS, y más aún cuando se trata de un receptor de alta sensibilidad debido a los largos intervalos de correlación. El cálculo de la correlación implica procesar miles de muestras cada milisegundo para satélite y para cada celda de retardo y frecuencia. Una opción habitual es realizar las correlaciones en el dominio de la frecuencia, es decir, multiplicando muestra a muestra las representaciones frecuenciales de dos señales y haciendo seguidamente la transformación inversa. Suponiendo que las dos señales tienen una longitud de N muestras, así se consigue que el orden del número de operaciones sea $O(N \log N)$, en lugar de $O(N^2)$, que resulta cuando la correlación se realiza en el dominio temporal mediante el desplazamiento, la multiplicación y la suma de las señales.

El algoritmo que vamos a esbozar se denomina *doble-DFT*. Permite realizar integraciones coherentes (que a su vez se pueden acumular no coherentemente) de duración igual a la de los bits, sin ninguna información a priori sobre las transiciones de estos. Es una forma eficiente de implementar el filtro adaptado para diferentes valores del desplazamiento de frecuencia utilizando solo transformaciones DFT y algunas operaciones básicas. El hecho de utilizar solo transformaciones DFT es muy útil para la inclusión de receptores GNSS en dispositivos móviles de comunicaciones. Todos los estándares de comunicaciones móviles utilizan ya o evolucionan hacia señales OFDM (*orthogonal frequency division multiplex*), en las que el elemento clave en su demodulación es la transformación DFT, y por lo tanto, los terminales suelen incluir un módulo especializado en realizar la DFT de manera muy eficiente. Un algoritmo de procesado HS-GNSS que utilice casi únicamente DFT se puede incluir en un terminal móvil de comunicaciones aprovechando el módulo de DFT que ya existe en el propio terminal, reduciendo al máximo la inclusión de nuevos componentes de procesado de señal GNSS en el receptor.

El primer paso del algoritmo *doble-DFT* aplicado a la señal GPS-C/A consiste en tomar un bloque de muestras de duración igual a dos bits. Así se asegura que dicho intervalo incluye necesariamente el trozo de señal que corresponde a un bit, sea cual sea el retardo de la señal. Siguiendo el mismo razonamiento, un intervalo de duración igual a dos veces la duración del período de la secuencia PN incluye necesariamente un período entero de dicha secuencia. Fragmentos sucesivos de $2N_{scode}$ muestras tomados con un desplazamiento ente ellos de un período del código se apilan para construir la matriz $\mathbf{M}_l \in \mathbb{C}^{2N_{scode} \times (2N_r - 1)}$, donde $N_r = 20$ es el número de períodos de la secuencia PN que forman un bit y el subíndice l indica el intervalo de bit que se está procesando. La forma de construir la matriz se muestra en la figura 37.

Figura 37. Procedimiento para rellenar demuestras la matriz \mathbf{M}_I utilizada en el algoritmo *doble-DFT*



Hemos supuesto que el valor de N_{scodes} es entero, pero en la práctica, debido a errores residuales de frecuencia o al uso prácticamente obligatorio de frecuencias de muestreo inconmensurables con la tasa de chip, dicho valor no es entero. Esto causa ciertos problemas a la hora de rellenar \mathbf{M}_I porque no está claro en qué muestra acaba un período del código y empieza el siguiente. La solución trivial es redondear N_{scodes} y empezar cada columna de \mathbf{M}_I con la muestra que está una posición múltiplo de ese valor, pero no funciona satisfactoriamente porque se introduce un pequeño retardo entre columnas que se va acumulando. Este retardo hará que los picos de correlación asociados a cada columna no estén alineados entre ellos y, al sumarlos posteriormente, se produce un efecto similar al "emborronamiento" discutido antes. Aunque puede parecer un error marginal, en realidad no lo es tanto. Cualquier pequeño error en retardo queda "amplificado" al multiplicar por la velocidad de la luz y pasarlo a metros, y se convierte en relevante dada la precisión requerida a los sistemas GNSS. Para poner un ejemplo imaginemos una frecuencia de muestreo igual a 3,5 veces la velocidad de chip, que equivale a una distancia de 86 metros entre muestras. Si se incurre en un error de media muestra entre columnas, esto equivale a un desplazamiento de 43 metros entre curvas de correlación. Para evitar este efecto, se interpolan las muestras de la señal al introducirlas en la matriz. Esto se puede hacer de modo muy eficiente en el dominio de la frecuencia multiplicando por una matriz \mathbf{T}_I de exponenciales complejas.

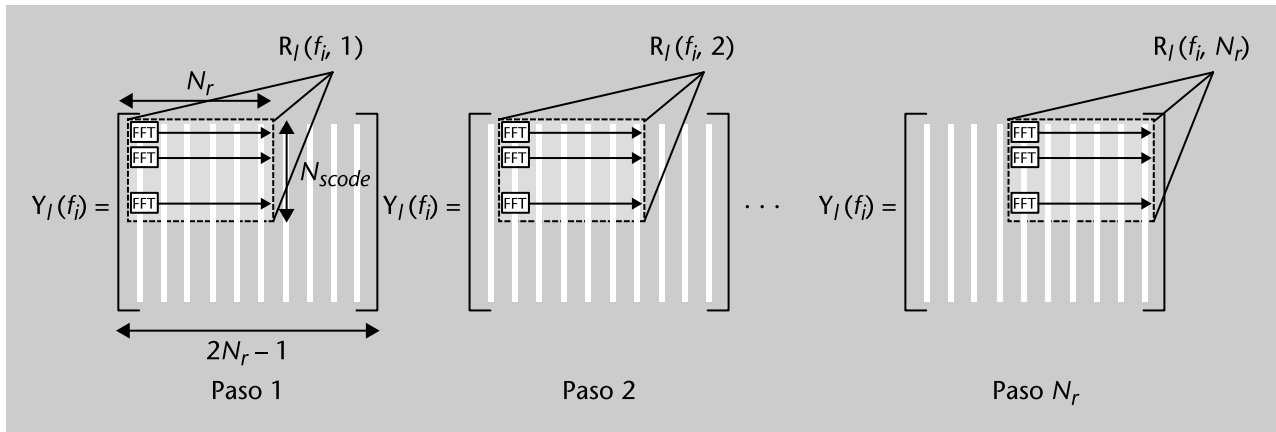
La correlación de cada una de las columnas de \mathbf{M}_I con un período de la secuencia PN desplazada a una frecuencia f_i se puede realizar mediante la DFT, y se puede expresar como

$$\mathbf{Y}(f_i) = \mathbf{F}^H((\mathbf{F}\mathbf{M}_l) \bullet \mathbf{T}_l) \bullet ((\mathbf{J}_i \mathbf{F} \mathbf{c}) \otimes \mathbf{1}_{2N_r-1}^T) \quad (184)$$

donde F es la matriz de Fourier (esto es, multiplicar por F equivale a realizar la DFT; y multiplicar por \mathbf{F}^H equivale a la DFT inversa), \bullet representa el producto de Hadamard o componente a componente, \otimes es el producto de Kronecker, c contiene las muestras de un período de la secuencia PN. El vector c contiene ceros en su segunda mitad, mientras que las columnas de \mathbf{M}_l están completamente llenas de muestras de la señal, lo que implica que estamos utilizando el método *overlap-and-save* para llevar a cabo la correlación (Oppenheim; Schafer; Buck; 1999), y por lo tanto solo la mitad superior de la matriz \mathbf{Y}_l es relevante (la parte inferior se puede descartar). \mathbf{J}_i es un matriz que realiza un rotación circular del vector que tiene a su derecha. Hay que recordar que una rotación en el dominio de la frecuencia discreta equivale a un desplazamiento de frecuencia en tiempo, que es justamente lo que se desea. La granularidad de la búsqueda *coarse* en frecuencia viene dada por el mínimo desplazamiento de frecuencia, que es el que se obtiene al rotar una muestra, y que a su vez es igual al inverso de la duración de las columnas. Por lo tanto, en este caso es 500 Hz. En dispositivos móviles dotados de un reloj con estabilidad de 1 ppm, habría que buscar en una ventana de $\pm 1,5$ kHz (esto es 1 ppm sobre la frecuencia portadora L1), que se traduce en explorar 7 posiciones de frecuencia en saltos de 500 Hz.

Una vez que se dispone de las matrices \mathbf{Y}_l , el siguiente paso es buscar las transiciones de bit para poder integrar coherentemente durante un bit. El aspecto fundamental es que esta búsqueda también se puede realizar mediante la DFT, y de aquí el nombre de *doble-DFT* del algoritmo. El proceso se ilustra en la figura 38. En cada matriz \mathbf{Y}_l , se toman sucesivos conjuntos de N_r columnas para comprobar si esas son las columnas que forman un bit. Sin embargo, los valores de las filas de estas submatrices no se pueden sumar directamente porque la existencia de un error de frecuencia haría que la señal se cancelase. Por lo tanto, se debe sumar pero probando al mismo tiempo diferentes *offsets* de frecuencia, y esto se puede hacer directamente aplicando la DFT a cada una de las filas de las submatrices. El resultado son las matrices $\mathbf{R}(f_i, m)$, donde cada uno de los parámetros corresponde a lo siguiente: l al período de bit que se está procesando, f_i a la celda gruesa de frecuencia y m a la posición de la transición de bit.

Figura 38. Procedimiento para la sincronización de bit en el algoritmo *doble-DFT*



Si se añaden N_r ceros a cada una de las filas antes de hacer la DFT, el tamaño de cada matriz $\mathbf{R}_I(f_i, m)$ es $N_{scodes} \times 2N_r$, y la granularidad de la búsqueda fina en frecuencia es $1/(2N_r N_{scodes} T_s)$, que en el caso de GPS-C/A es 25 Hz. Recapitulando, se ha obtenido un conjunto de matrices $\mathbf{R}_I(f_i, m)$ a partir de las muestras originales aplicando poco más que la DFT en dos direcciones (primero por columnas y luego por filas) y alguna reorganización y selección de los elementos de las matrices involucradas.

Por comodidad para completar la detección del satélite, se pueden definir unos hipercubos que $\mathbf{C}_I(f_i)$, cuyas dimensiones son $N_{scodes} \times 2N_r \times N_r$, que contienen los elementos del conjunto de matrices $\mathbf{R}_I(f_i, m)$ para todos los valores de m . Los hipercubos correspondientes a N_I intervalos de bit se pueden acumular de forma no coherente:

$$\mathbf{R}_{NC}(f_i) = \frac{1}{N_I} \sum_{l=0}^{N_I-1} |\mathbf{R}_I(f_i, l)|^2 \quad (185)$$

que es la estructura final de datos resultante del algoritmo *doble-DFT*. El último paso es la determinación del máximo de esta estructura, de manera que si dicho máximo supera un umbral (como ya se comentó previamente), indicará la presencia de la señal de un satélite. El máximo se tiene que buscar en 4 dimensiones, esto es, las tres de cada hipercubo $\mathbf{R}_{NC}(f_i)$ y además el parámetro f_i . La descripción que habíamos presentado al principio del apartado decía que la búsqueda se hacía en una rejilla de dos dimensiones en (τ, f) , pero se trataba de una descripción a nivel conceptual. Cuando esto se lleva a la práctica, es necesario hacer adaptaciones, etc., y el algoritmo *doble-DFT* es un ejemplo de esto. Vemos que en la práctica el resultado es una búsqueda en 4D aunque conceptualmente no deja de ser una búsqueda en las dos dimensiones de tiempo-frecuencia. La búsqueda en 4D no tiene por qué ser ni más ni menos complicada que la búsqueda en 2D, simplemente es una forma diferente de representar los datos. La posición de máximo indica el desplazamiento en

tiempo y en frecuencia de la señal recibida. Concretamente, la posición en la primera dimensión (es decir, en la dirección de las columnas) hace referencia al desplazamiento dentro de un período del código, mientras que la tercera dimensión hace referencia a la posición de la transición de bit dentro los N_r períodos que forman un bit. De esta manera, entre la primera y tercera dimensión determinan el retardo de la señal. Por otra parte, el valor de f_i para el cual se encuentra el máximo proporciona el valor del desplazamiento grueso de frecuencia, mientras que la posición del máximo en la segunda dimensión (es decir, en la dirección de las filas) corresponde a la corrección fina de frecuencia a añadir al desplazamiento grueso. Entre ambos valores se obtiene la estimación de la frecuencia de la señal.

Por último, cabe mencionar que los algoritmos para resolver la ecuación de navegación también se deben modificar en el caso de HS-GNSS. En los escenarios con baja C/No puede ocurrir que no se puedan detectar los bits del mensaje de navegación. Hemos de mencionar que es posible medir el retardo de la señal a niveles de C/No para los cuales no se pueden detectar los bits, y este es el motivo último de interés en los sistemas HS-GNSS/A-GNSS. El hecho de no disponer del mensaje de navegación tiene dos implicaciones. La primera es evidente y es que no se dispone de la información contenida en el mensaje de navegación. Esto se soluciona fácilmente utilizando el mecanismo A-GNSS.

La segunda implicación es que no se pueden observar las marcas temporales que contiene la señal GNSS (en el caso de GPS es el TOW, como se explicó en el subapartado 5.1) y el aspecto esencial es que esto imposibilita obtener el instante de transmisión a partir de la observación de la señal (recordad la figura 18). En el subapartado 5.3 se presentaron dos algoritmos para resolver las ecuaciones de navegación. El primero de ellos tomaba como dato de partida el tiempo de transmisión de la señal, por lo que claramente no es aplicable en el caso de HS-GNSS. El segundo algoritmo partía del tiempo de recepción de la señal, pero este se debería conocer con una precisión considerable (del orden del milisegundo) para no introducir errores excesivos en la posición de los satélites. Dado que conocer el tiempo GPS con una precisión menor al milisegundo tampoco es habitual en las aplicaciones habituales de HS-GNSS, el segundo algoritmo tampoco es utilizable directamente. Aunque no vamos a entrar en los detalles, la solución es modificar las ecuaciones de navegación para incluir una 5.^a variable (además de las habituales, que son las coordenadas del receptor y el *offset* del reloj del receptor), que es el instante de transmisión. Este tipo de navegación se suele conocer en la literatura avanzada como *coarse-time navigation*.

El hecho de no observar las marcas temporales también provoca que los retardos y, por tanto también los pseudorrangos, medidos en HS-GNSS sean ambiguos. No obstante, el período de la ambigüedad suele ser bastante grande (por ejemplo, 1 ms en GPS, que equivale a 300 km, o también puede ser 20 ms) y si se dispone de unos valores aproximados de la posición del receptor (incluso con errores de kilómetros) y del tiempo en el receptor (incluso del

Lectura recomendada

Si queréis saber más sobre la *coarse-time navigation*, podéis encontrar más detalles en la sección VII de la obra de Secco-Granados, López-Salcedo, Jiménez-Baños y López-Risueño (2012) o en el excelente capítulo 4 de Van Diggelen (2009).

orden de segundos) se puede saber la distancia aproximada entre el receptor y los satélites con un error mucho menor al período de la ambigüedad, y por lo tanto resolverla.

Resumen

En este módulo se ha presentado el sistema de radionavegación por satélite más utilizado actualmente: el sistema GPS. Este sistema nació de un programa que estableció el Departamento de Defensa de Estados Unidos a finales de los años sesenta con el objetivo de unir los esfuerzos llevados hasta el momento en materia de radionavegación por satélite. No obstante, hasta el año 1995 el sistema no pasó a declararse sistema con capacidad operacional total.

En primer lugar, se han descrito los segmentos del sistema GPS: segmento espacial, segmento de control y segmento de usuario. El segmento espacial está formado por una constelación de satélites base que consta de 24 satélites distribuidos en 6 planos orbitales de 4 satélites cada uno. La órbita de dichos satélites es prácticamente circular, con un período orbital de medio día sideral. Estos satélites se encuentran organizados en bloques, según la fase de desarrollo a la que pertenezcan. Los bloques son el bloque I, asociado a la fase de experimentación del sistema y formado por 11 satélites lanzados entre los años 1978 y 1985; el bloque II, compuesto por los primeros 9 satélites de producción lanzados entre 1989 y 1990; el bloque IIA, 19 satélites similares a los del bloque II pero con algunas mejoras, siendo en este caso lanzados entre 1990 y 1997; el bloque IIR, bloque de reposición de satélites formado por 13 satélites lanzados en el período comprendido entre 1997 y 2004; el bloque IIR-M, bloque consistente de la versión modernizada de los satélites que incluían nuevas señales y formado por 8 satélites lanzados entre 2005 y 2009; el bloque IIF, formado por 11 satélites de mantenimiento, que incluyen la nueva señal *Safety-of-Life*, el primero de los cuales se lanzó en el año 2010, y, finalmente, los bloques previstos para el futuro, orientados a satisfacer los requisitos de las próximas décadas (bloques III-A, III-B y III-C). La constelación de satélites en el momento de redacción del módulo está formada por 30 satélites: 10 del bloque IIA, 12 del bloque IIR, 7 del bloque IIR-M y 1 del bloque IIF. El segmento de control se encarga de monitorizar y controlar la constelación de satélites de GPS. Para ello, el sistema está compuesto de una estación de control maestra, encargada de centralizar el control de la constelación GPS; seis estaciones de monitorización, repartidas por el globo terrestre y encargadas de llevar a cabo las funciones de seguimiento de las señales de navegación de los diferentes satélites, y cuatro estaciones de transmisión de datos, encargadas de transmitir las órdenes de control y los mensajes de navegación generados por la estación de control maestra a los satélites. En cuanto al segmento de usuario, está formado por el conjunto de receptores GPS, existiendo diferentes formas de clasificación según su naturaleza: tipo de usuario (civil o militar), tipos de arquitectura de recepción (canales independientes, secuencia o multiplex) y capacidad frecuencial (monofrecuencia o multifrecuencia).

A continuación, se han presentado las bandas frecuenciales de operación y la estructura de las señales de navegación. Las bandas frecuenciales utilizadas para enviar las señales de GPS son las bandas L1, centrada en $f_{L1} = 1.575,42$ MHz y utilizada para enviar los códigos C/A, P y M; la banda L2, centrada en $f_{L2} = 1.227,60$ MHz y utilizada para transmitir los códigos P, M y L2C, y la banda L5, centrada en $f_{L5} = 1.176,45$ MHz y utilizada para enviar la señal *Safety-of-Life*. En cuanto a la señal GPS tradicional (previa a las versiones modernizadas), se compone principalmente del mensaje de navegación $D_k(t)$ y de las secuencias de *spreading* $C_k(t)$ y $P_k(t)$ asociadas a los códigos C/A y P, respectivamente. El mensaje de navegación $D_k(t)$ contiene la información necesaria para que el usuario pueda calcular la posición del satélite k -ésimo, como es el caso de las efemérides del satélite, y el tiempo de transmisión de la señal recibida. Es decir, la información necesaria para que el usuario calcule su posición. Aparte proporciona información relacionada con el estado de la señal, los almanaques, parámetros de corrección ionosféricos y parámetros para relacionar el reloj del sistema GPS con el sistema de tiempos UTC. Dicha información se transmite a 50 bps, por lo que se requieren 12,5 minutos para enviar un mensaje de navegación completo. Las secuencias de *spreading*, por su parte, no contienen ninguna información y su objetivo es:

- 1) dotar a la señal del satélite de un tipo de identificador (basado en el código en sí que contienen) para distinguirla de las señales de los otros satélites;
- 2) modificar el aspecto de la señal para que ofrezca mejores prestaciones a la hora de estimar el tiempo de propagación de la señal y, con esto, la distancia entre el usuario y el satélite, y
- 3) mejorar la robustez de la señal ante interferencias y posibles ataques intencionados (*jamming*).

Los códigos C/A y P se forman a partir de secuencias PRN con diferentes características según el caso. El código C/A, por su parte, se forma con códigos Gold de 1.023 chips de longitud y velocidad de chip igual a 1,023 Mchips/s. Presenta autocorrelaciones y correlaciones cruzadas con buenos niveles de rechazo ante versiones retardadas del propio código y códigos de otros satélites, concretamente del orden de $-1/1.023$ salvo algunas perturbaciones. En el caso del código P, este tiene una longitud de $6,1871 \times 10^{12}$ chips, un período temporal de 7 días y una velocidad de chip igual a 10,23 Mchips/s. En cuanto a los niveles de autocorrelación (en las zonas de rechazo) y las correlaciones cruzadas, el código P ofrece un comportamiento excelente al presentar niveles del orden de $-1/6,1871 \times 10^{12}$. Finalmente, se han presentado las características principales de las señales modernizadas que empezaron a utilizarse con los lanzamientos de los satélites de los bloques IIR-M (señales L2C y M) e IIF (señal *Safety-of-Life*).

A continuación se ha estudiado cómo se utilizan estas señales para acabar calculando finalmente la posición. En una primera fase, se ha analizado el proceso que va desde la recepción de la señal a la generación de las medidas de retardo y fase. Este proceso se lleva a cabo en dos etapas, denominadas etapa de adquisición (subapartado 4.1) y etapa de *tracking* (subapartado 4.2). La etapa de adquisición tiene como objetivo proporcionar una estimación aproximada de los valores de retardo de código y frecuencia Doppler, para los satélites que se encuentran visibles en el momento en el que fue capturada la señal por el receptor GPS. Estas estimaciones son el resultado de una búsqueda bidimensional en tiempo y frecuencia en la que la métrica bajo análisis es la integración coherente y/o no-coherente de la salida del correlador. Si bien la integración coherente presenta una mayor ganancia, la presencia de los bits del mensaje de navegación limita esta integración a un máximo de 20 ms, además de la propia limitación que pueda venir impuesta por el error residual de frecuencia existente. Por el contrario, la integración no-coherente no presenta estos problemas y puede ser extendida más allá del tiempo de bit, llegando a valores de integración de varios segundos en condiciones de alta sensibilidad. En ambos casos, el objetivo de estas integraciones es el mismo: reducir el efecto del ruido y permitir que la señal del satélite bajo análisis, en caso de estar presente, pueda ser fácilmente detectada. Este proceso, de hecho, puede modelarse como un problema de detección de señal en ruido, en donde pueden aplicarse los conceptos y las herramientas sobre teoría de la detección ya presentadas en el módulo "Sistemas radar". Una vez que la señal del satélite de interés ha sido detectada, automáticamente se dispone de estimaciones aproximadas acerca de su retardo de código y frecuencia Doppler. Ambos valores corresponden a la celda de la búsqueda bidimensional en donde se ha detectado la presencia de señal.

Estas estimaciones aproximadas de retardo de código y frecuencia Doppler podrían ser ya suficientes para el caso de un receptor de comunicaciones. Sin embargo, un receptor GPS requiere una precisión mucho mayor que la de un receptor de comunicaciones, ya que las medidas de retardo de código (en unidades de tiempo) han de ser convertidas posteriormente a distancia, y ello implica un factor de conversión que viene dado por la velocidad de la luz $c = 3 \cdot 10^8$ m/s. Por tanto, pequeñas variaciones en términos de retardo de código ocasionan grandes variaciones en términos de distancia. Por ello, las estimaciones aproximadas de retardo de código y frecuencia Doppler que proporciona la etapa de adquisición han de ser refinadas y monitorizadas a lo largo del tiempo (para seguir cualquier posible variación temporal), lo cual se lleva a cabo durante la etapa de *tracking*. En ella, se realiza un seguimiento del retardo de código y de la fase de la portadora mediante dos esquemas en paralelo, el Delay Lock Loop (DLL) y el Phase Lock Loop (PLL), respectivamente. En el subapartado 4.2 se han presentado los principios de funcionamiento de ambos esquemas, los cuales están basados en arquitecturas en lazo cerrado. Los principales elementos de este tipo de arquitecturas son:

- el discriminador, que facilita a su salida una señal proporcional al error entre la señal de entrada y la réplica local;
- el filtro de lazo, que suaviza la variabilidad que pueda presentar la salida del discriminador debido al efecto del ruido;
- el NCO (Numerically Controlled Oscillator), que acumula la salida del filtro de lazo y proporciona una señal que es utilizada directamente para generar la fase del código o la fase de la portadora de la réplica local.

Para el análisis de este tipo de sistemas en lazo cerrado, el subapartado 4.2 ha introducido las herramientas básicas y ha presentado las condiciones de estabilidad y prestaciones que rigen dichos sistemas.

Una vez explicado cómo se obtienen las medidas de retardo y de fase para cada satélite, se ha pasado a estudiar cómo se utilizan dichas medidas para calcular la posición del receptor. El primer paso de este proceso consiste en generar los observables de pseudodistancia y de fase de portadora. Estas magnitudes son esencialmente las medidas proporcionadas por el subsistema de procesamiento de señal del receptor, pero es necesario hacer algunas modificaciones para que se pueda plantear el modelo que relaciona los observables con las distancias entre el receptor y el satélite. Se ha presentado en detalle dicho modelo, explicando el porqué de la dependencia de los observables no solo de la distancia geométrica, sino también de parámetros atmosféricos y de los relojes del receptor y de los satélites. El segundo paso es lo que se denomina habitualmente como "resolución de las ecuaciones de navegación" y consiste en despejar de dicho modelo las incógnitas de posición. Resolver las ecuaciones de navegación no es trivial porque la dependencia de la distancia con las coordenadas es altamente no lineal. Se han presentado dos familias de soluciones. La primera de ellas se basa en linealizar dicha dependencia mediante un desarrollo de Taylor alrededor de una posición a priori aproximada del receptor, y da lugar a un procedimiento iterativo porque la posición obtenida se puede utilizar como nuevo punto de partida para la linealización en una siguiente iteración. La segunda familia de soluciones, cuyo representante más conocido es el método de Bancroft, permite obtener la estimación de la posición de manera cerrada. Para ello se elevan las ecuaciones de navegación al cuadrado para eliminar la raíz cuadrada que aparece en la expresión de la distancia, pero esto también ocasiona un aumento del ruido. Por lo tanto, este tipo de soluciones son menos exactas, pero resultan útiles en general para obtener un punto de partida para las técnicas basadas en la linealización. Un problema añadido en el cálculo de la solución de navegación es el del cálculo de las posiciones de los satélites, que no es un dato directamente conocido a priori puesto que sus coordenadas varían con el tiempo debido a su propio movimiento y al movimiento del sistema de coordenadas ligado a la Tierra. Se han explicado dos algoritmos que permiten resolver esta incertidumbre inherente a las posiciones de los satélites al mismo tiempo que se calcula la posición del receptor.

Evidentemente, la posición que obtiene en el receptor no coincide perfectamente con la real porque existen múltiples fuentes de errores que afectan a la precisión de la posición. Se han analizado el origen y los efectos de las diferentes fuentes de error, que se han clasificado en tres categorías según si se deben a errores en las efemérides o el reloj de los satélites, a la atmósfera o a los efectos locales en el receptor (ruido, interferencias y propagación multicamino). La atmósfera introduce errores en la medida de distancias porque la velocidad de propagación de la señal en la ionosfera y en la troposfera no es exactamente la del vacío. En términos absolutos, los errores ionosféricos son los más importantes de todos, pero se pueden corregir realizando medidas a dos o más frecuencias o mediante técnicas diferenciales.

Estas técnicas diferenciales también permiten prácticamente eliminar los errores provenientes de los satélites (efemérides y relojes), de manera que finalmente los errores debidos al multicamino y a las interferencias se convierten en la principal limitación en muchas aplicaciones GNSS. En particular, se ha estudiado en detalle el efecto de la propagación multicamino. La recepción de varias réplicas de la señal GNSS no tiene un efecto potencialmente beneficioso, como ocurre a veces en comunicaciones al permitir captar más energía de la señal, sino que distorsiona tanto las medidas de retardo como las de fase. Las señales reflejadas se propagan a través de caminos más largos que el camino directo entre satélite y receptor y, por tanto, no aportan información sobre la distancia entre ambos, que es la base del posicionamiento. En la práctica, esto se traduce en un desplazamiento de la curva de discriminación del DLL y PLL. Además, dado que los errores debidos al multicamino suelen variar lentamente con el tiempo, provocan un sesgo en la posición, que a penas se puede reducir promediando a lo largo del tiempo. Se ha mostrado que los errores debidos al multicamino pueden alcanzar varias decenas de metros en los pseudorrangos con la señal GPS C/A, y que en la fase no exceden un cuarto de ciclo. Se ha recordado la diferencia entre la velocidad de grupo y la de fase en los medios dispersivos para justificar las distintas maneras como la ionosfera afecta a las medidas de retardo y de fase.

El modo habitual de conseguir una precisión en el posicionamiento del orden de un metro o mejor es recurrir a técnicas diferenciales. A grandes rasgos, consisten en utilizar una estación de referencia próxima al receptor del usuario que se vaya a posicionar, de manera que al realizar combinaciones de los observables medidos por ambos receptores los errores comunes (que son básicamente todos menos los errores locales a los receptores) se cancelen. Los dos principales criterios para clasificar las técnicas diferenciales hacen referencia a si permiten realizar un posicionamiento absoluto o relativo del usuario, o a si se utilizan medidas de código y/o de fase. Se han analizado todos los casos, poniendo especial énfasis en el posicionamiento relativo mediante medidas de fase porque es el modo de conseguir precisiones centrimétricas, y que con cierto abuso de lenguaje se suele denominar posicionamiento RTK (*real time kinematic*). El paso clave para poder aprovechar las medidas de fase es lo que se denomina la resolución de la ambigüedad de fase, que consiste en calcular el

número entero de longitudes de onda que se ha de añadir a las medidas de fase para que representen una medida absoluta de distancia. Se han presentado las técnicas llamadas "libres de geometría" para entender y asimilar el problema, aunque la manera de obtener mejores prestaciones pasa por utilizar técnicas "basadas en la geometría". La base de este último grupo de técnicas es plantear un conjunto de ecuaciones donde las incógnitas son las coordenadas de la posición y los valores enteros de las ambigüedades. En este caso, la obtención de la solución de navegación implica resolver un problema de optimización mixto, es decir, con variables enteras y reales.

Finalmente, se ha estudiado la extensión de los sistemas GNSS a entornos para los cuales no fueron diseñados inicialmente. Las condiciones nominales de funcionamiento de un sistema GNSS corresponden a la recepción de la señal en condiciones de cielo abierto, pero existe un gran interés en poder utilizar dichos sistemas en cañones urbanos y en entornos interiores. Esto tiene implicaciones a dos niveles: es necesario aumentar la sensibilidad de los receptores y se deben modificar las ecuaciones de navegación. El modo de aumentar la sensibilidad pasa por aumentar el tiempo de observación de la señal para captar más energía de esta antes de proporcionar una nueva estimación de la posición. Se han analizado las limitaciones que existen en alargar dicho tiempo de observación debido a la presencia del mensaje de navegación, desviaciones de frecuencia, ruido de fase, etc., para poder entender los compromisos fundamentales que existen en el diseño de un receptor de alta sensibilidad. La acumulación de la energía de la señal durante largos períodos de tiempo también supone una mayor carga computacional en la determinación de la frecuencia y del retardo de la señal. Se ha esbozado una manera eficiente de realizar la integración de la señal y de las búsquedas correspondientes utilizando repetidamente la transformada discreta de Fourier. En cuanto a la segunda implicación de la recepción en entornos degradados, esta se debe a que el receptor no es capaz de observar las marcas de tiempo que se hallan en la señal GNSS y que le permiten calcular el tiempo de transmisión de la señal, que se convierte en una incógnita más en las ecuaciones de navegación.

Ejercicios de autoevaluación

1. Indicad qué afirmación es correcta:

- a) La configuración base del sistema GPS consta de 30 satélites distribuidos en 6 planos orbitales con 5 satélites en cada plano.
- b) Las órbitas de los satélites GPS presentan una gran excentricidad.
- c) El período orbital de los satélites GPS es de medio día sideral.
- d) El radio de la órbita de los satélites GPS es de aproximadamente 11.000 km.

2. Los satélites del bloque IIR-M (modernizado)...

- a) transmiten la nueva señal *Safety-of-Life* en L5.
- b) presentan un diseño mejorado y una arquitectura flexible para satisfacer los requisitos de las próximas décadas.
- c) fueron puestos en órbita en el período 2005-2009, y se lanzaron un total de 11 satélites.
- d) fueron los primeros en transmitir la señal civil L2C.

3. Indicad qué afirmación es correcta:

- a) El sistema GPS fue declarado como sistema de capacidad operacional inicial durante la Guerra del Golfo (años 1990-1991).
- b) Los satélites del bloque IIA ofrecen buenas prestaciones contactando con la Tierra una vez cada seis meses.
- c) Los satélites del bloque IIF no incluían ninguna nueva señal de navegación, ya que formaban parte del programa de mantenimiento del sistema.
- d) Algunos de los satélites del bloque I llegaron a doblar la vida útil de diseño.

4. Indicad cuál de las siguientes aseveraciones es falsa:

- a) Las estaciones de monitorización del segmento terrestre disponen de un receptor en la banda L y uno en la banda S.
- b) El número de estaciones de monitorización del segmento terrestre es igual a 6, mientras que únicamente existen 4 estaciones de transmisión de datos.
- c) Las estaciones de transmisión de datos operan en la banda S para enviar y recibir datos de los satélites.
- d) Las estaciones de monitorización disponen de sensores meteorológicos.

5. Las señales de GPS se transmiten...

- a) usando la banda L1 en el caso de que se quiera transmitir la señal *Safety-of-Life*.
- b) usando la banda L5 si se quiere transmitir la nueva señal militar basada en el código M.
- c) usando la banda L2 si se quiere transmitir las dos nuevas señales L2C y *Safety-of-Life*.
- d) usando la banda L1 si se quiere transmitir las señales basadas en el código C/A y en el nuevo código militar M.

6. Indicad cuál de las siguientes aseveraciones es correcta:

- a) La señal basada en el código C/A es más robusta ante *jamming* que la señal basada en el código P debido a su mayor tiempo de chip.
- b) La señal basada en el código P se transmite únicamente en la banda L2.
- c) Un bit de navegación contiene 20 secuencias del código C/A.
- d) El código P se forma a partir de un código Gold resultante de la combinación de dos registros de desplazamiento.

7. Indicad cuál de las siguientes aseveraciones es falsa:

- a) Uno de los objetivos de los códigos de *spreading* es incrementar la velocidad con la que se envían los bits de navegación.
- b) Uno de los objetivos de los códigos de *spreading* es separar las contribuciones de diferentes satélites.
- c) Uno de los objetivos de los códigos de *spreading* es ofrecer una mayor robustez a señales interferentes de banda estrecha.
- d) Uno de los objetivos de los códigos de *spreading* es mejorar la estimación de las medidas de TOA.

8. El mensaje de navegación...

- a) requiere una semana para enviar un mensaje completo.
- b) está organizado en transmisiones de 25 páginas por mensaje completo.
- c) está organizado en transmisiones de 5 subtramas por mensaje completo.
- d) se centra en enviar únicamente las efemérides del satélite que transmite el mensaje para que el usuario pueda localizar su posición.

9. Considerad un módulo de adquisición como el mostrado en la figura 11. Indicad qué afirmación es correcta:

- a) El tiempo máximo de integración coherente es 20 ms.
- b) El tiempo máximo de integración coherente es 10 ms.
- c) El tiempo máximo de integración coherente es 1 ms.
- d) El tiempo de integración coherente debe ser lo más pequeño posible para no verse afectado por la presencia de errores residuales de frecuencia.

10. Considerad un módulo de adquisición como el mostrado en la figura 11. Indicad qué afirmación es correcta:

- a) En presencia de errores de frecuencia, es preferible mantener la integración coherente lo más corta posible y extender el tiempo de integración de manera no-coherente.
- b) En presencia de errores de frecuencia, es preferible implementar una integración coherente de 20 ms para extraer toda la energía del bit, y posteriormente extender el tiempo de integración de manera no-coherente.
- c) El tiempo de integración coherente depende del error máximo de frecuencia permisible.
- d) La integración coherente no se puede implementar si hay errores de frecuencia.

11. Indicad qué afirmación es correcta:

- a) El módulo de adquisición implica un mayor número de operaciones que el módulo de *tracking*.
- b) El módulo de *tracking* implica un mayor número de operaciones que el módulo de adquisición.
- c) El módulo de *tracking* es obligatorio implementarlo para poder obtener estimaciones muy precisas de la posición del usuario.
- d) El módulo de adquisición nunca puede llegar a proporcionar la precisión que ofrece el módulo de *tracking*.

12. Considerad el diseño del bloque encargado de la búsqueda en tiempo y frecuencia que tiene lugar en el módulo de adquisición de un receptor GPS. Indicad qué afirmación es falsa:

- a) El diseño de ese bloque tiene un impacto en el tiempo total que se tardará en completar la búsqueda del satélite de interés.
- b) El diseño de ese bloque tiene un impacto en la precisión de las estimaciones de retardo de código y frecuencia Doppler que posteriormente serán refinadas por el módulo de *tracking*.
- c) El diseño de ese bloque es independiente del resto de bloques que forman el módulo de adquisición.
- d) El diseño de ese bloque es independiente del módulo de *tracking*.

13. Indicad qué afirmación es correcta:

- a) La curva S del discriminador *early-late cuasi-coherente* presenta una zona lineal más estrecha, y por lo tanto, al ser más sensible a cambios en la entrada, las correcciones que ofrece a su salida son más precisas.
- b) Aumentar la precisión del discriminador es únicamente una cuestión de reducir la separación Δ entre correladores *early-late*.
- c) Los discriminadores de DLL presentan todos la misma pendiente en la zona lineal y se diferencian únicamente en la anchura de esta zona lineal y en el comportamiento fuera de ella.
- d) En presencia de multipath, es recomendable utilizar discriminadores *early-late* con una separación Δ pequeña.

14. Indicad qué afirmación es correcta:

- a) Un sistema en lazo cerrado de primer orden es estable siempre y cuando la ganancia del discriminador κ_d sea inferior a $2/C_1$.
- b) Un filtro de lazo de primer orden es capaz de seguir la evolución de la señal de entrada en caso de que las derivadas de esta sean todas iguales a cero excepto la de primer orden.

- c) Un sistema en lazo cerrado es estable si los ceros de la función de transferencia del sistema están todos dentro del círculo unidad.
- d) Un sistema en lazo cerrado está sobreamortiguado si $\zeta < 1$.

15. Indicad qué afirmación es correcta:

- a) En un filtro de lazo de segundo orden, el ancho de banda del lazo depende exclusivamente de los valores de las constantes $\{C_1, C_2\}$.
- b) En cuanto al impacto del ruido térmico, la ganancia que proporciona un filtro de lazo viene dada por su ancho de banda B_n .
- c) Las medidas de retardo de la señal recibida que pueden obtenerse a partir de la fase de la portadora son F_L^2 veces más precisas que las medidas obtenidas a partir del propio retardo del código.
- d) Cuando la C/N_0 es muy baja, las medidas de tiempo de retardo que se obtienen a partir del código son muy ruidosas. En estas circunstancias, es preferible utilizar las medidas de tiempo de retardo obtenidas a partir de la portadora, para compensar así el incremento de ruido con la mayor precisión que ofrece la portadora.

16. ¿Cuál es la definición del observable de pseudodistancia?

- a) La diferencia entre el instante de transmisión de la señal medido según el reloj del satélite y el instante de recepción de la señal medido según el reloj del receptor, todo ello multiplicado por la velocidad de la luz en la atmósfera.
- b) La distancia entre el satélite y el receptor, incluyendo los retardos troposféricos e ionosféricos.
- c) La diferencia entre el instante de transmisión de la señal medido según el reloj del satélite y el instante de recepción de la señal medido según el reloj del receptor, todo ello multiplicado por la velocidad de la luz en el vacío.
- d) La diferencia entre el instante de transmisión de la señal y el instante de recepción medidos según la escala de referencia de tiempos del sistema.

17. El instante de transmisión de la señal se obtiene a partir...

- a) del contenido del mensaje de navegación.
- b) de la triangulación de las medidas de varios satélites.
- c) del reloj del receptor, corregido por el *offset* de este.
- d) de una marcas temporales que contiene la señal, que en la práctica suelen ser unos determinados conjuntos de bits en el mensaje de navegación.

18. ¿Cuál de las siguientes afirmaciones es cierta?

- a) El observable de fase siempre toma un valor entre menos medio y más medio ciclo de portadora.
- b) El observable de fase puede tomar cualquier valor entero.
- c) El observable de fase tiene una ambigüedad intrínseca igual a un valor entero de ciclos de portadora.
- d) El observable de fase es una versión escalada del pseudorange.

19. ¿Cuál de las siguientes afirmaciones sobre el método de Bancroft es falsa?

- a) Proporciona en general una solución con menor precisión que los métodos de solución de la ecuación de navegación basados en la linealización.
- b) Es necesario disponer de medidas a 5 satélites como mínimo para poderlo aplicar.
- c) Requiere resolver una ecuación de segundo orden.
- d) No requiere conocer un valor aproximado de la posición.

20. Los factores DOP...

- a) están directamente relacionados con la calidad de los observables.
- b) son una forma de cuantificar el efecto de la geometría del conjunto de satélites sobre la precisión del posicionamiento.
- c) son la covarianza del residuo de la solución de navegación.
- d) dependen de la potencia de transmisión del satélite.

21. El uso de técnicas diferenciales es útil para reducir los errores debidos a...

- a) ionosfera, troposfera y efemérides.
- b) ionosfera y multicamino, si se utilizan medidas a dos frecuencias.

- c) únicamente los efectos locales (ruido, multicamino e interferencias).
- d) la rotación de la Tierra entre el instante de transmisión y de recepción.

22. ¿Cuál de las siguientes afirmaciones es cierta?

- a) El multicamino provoca un error mayor cuando mayor es el retardo de las reflexiones.
- b) Como las reflexiones siempre llegan después que la señal directa, el efecto del multicamino siempre es un sesgo positivo en el pseudorange.
- c) Las medidas de fase no están afectadas por el multicamino.
- d) Los errores del multicamino no se pueden mitigar con técnicas diferenciales ni con medidas a dos frecuencias.

23. La diferencia entre las técnicas de resolución de ambigüedades de fase *geometry-free* (GF) y *geometry-based* (GB) es:

- a) Las GF se pueden aplicar a receptores en movimiento y las GB no.
- b) Las GF requieren la transmisión de las correcciones diferenciales en tiempo real.
- c) Las GB utilizan las coordenadas del receptor como parte de las incógnitas, en lugar de las distancias entre el receptor y cada satélite.
- d) Las GB solo se pueden utilizar con las dobles diferencias.

24. ¿Cuál de las siguientes opciones no es útil para aumentar la sensibilidad de los receptores?

- a) Dotar al receptor de mayor capacidad de cálculo y un reloj con menor ruido de fase.
- b) Aumentar el período de observación de la señal.
- c) Ignorar el mensaje de navegación para poder integrar durante más tiempo la señal.
- d) Recibir información de asistencia (A-GNSS).

25. ¿Cuál de las siguientes opciones es más adecuada para obtener un posicionamiento relativo con precisión centimétrica?

- a) Utilizar diferencias dobles de medidas de pseudorange a dos frecuencias.
- b) Utilizar diferencias dobles de medidas de fase a dos frecuencias si la distancia entre usuario y referencia no excede de 20 km
- c) Utilizar diferencias dobles de medidas de fase a dos frecuencias si la distancia entre usuario y referencia excede de 20 km, y a una frecuencia cuando no excede dicho valor.
- d) Utilizar el concepto de PPP.

26. ¿Qué afirmación referente a la troposfera y la ionosfera es cierta?

- a) La troposfera produce un adelanto de fase debido al contenido de vapor de agua de la atmósfera.
- b) Conociendo el perfil de temperatura y presión de la atmósfera se puede estimar de manera precisa el retardo de la ionosfera.
- c) La actividad de la troposfera aumenta por el día debido actividad solar.
- d) La velocidad de grupo y de fase coinciden en la troposfera, pero no en la ionosfera.

27. La afirmación: "No tiene sentido realizar medidas a más de cuatro satélites porque la ecuación de navegación solo tiene 4 incógnitas (las tres coordenadas de la posición y el *offset* del reloj del satélite)"...

- a) es cierta.
- b) puede ser cierta o falsa dependiendo de si se dispone de medidas a dos frecuencias o no.
- c) es falsa porque disponer de más de 4 medidas permite tener un sistema de ecuaciones de navegación con más ecuaciones que incógnitas y esto añade redundancia que puede ayudar en la determinación de la posición.
- d) solo es cierta cuando el receptor es estático.

Solucionario

Ejercicios de autoevaluación

1. c

2. d

3. d

4. a

5. d

6. c

7. a

8. b

9. b

10. c

11. a

12. c

13. d

14. a

15. b

16. c

17. d

18. c

19. b

20. b

21. a

22. d

23. c

24. c

25. b

26. d

27. c

Abreviaturas

- ADC** Analog-to-Digital Converter
- A-GNSS** *Assisted GNSS*
- AS** Anti-Spoofing
- BPSK** Binary Phase-Shift Keying
- BOC** Binary Offset Carrier
- C/A** Coarse Acquisition code
- C/No** *Carrier to NOise spectral densisty ratio*
- CDMA** Code Division Multiple Access
- CL** Civilian Long length code
- CM** Civilian Moderate length code
- DFT** *Discrete Fourier Transform*
- DLL** Delay Lock Loop
- DNSS** Defense Navigation Satellite System
- DOP** *Dilution Of Precision*
- DS-SS** *Direct Sequence Spread Spectrum*
- ECEF** *Earth Centered Earth Fixed*
- ECI** *Earth Center Inertial*
- E-L** Early-Late
- ENU** *East North Up*
- FEC** Forward Error Correction
- FLL** Frequency Lock Loop
- GBAS** *Ground Based Augmentation System*
- GDOP** *Geometric Dilution Of Precision*
- GNSS** Global Navigation Satellite System
- GPS** Global Positioning System
- GPST** *GPS Time*
- HDOP** *Horizontal Dilution Of Precision*
- HOW** HandOver Word
- HS-GNSS** *High Sensitivity GNSS*
- I&D** Integer and Dump
- IFB** *InterFrequency Bias*
- ILS** *Integer Least Squares*
- LAMBDA** *Least-squares AMBiguity Decorrelation Adjustment*
- MCS** Master Control Station
- NAVSTAR** NAVigation Satellite Time and Ranging

NCO Numerically Controlled Oscillator

NTRIP *Networked Rransport of RTCM via Internet Protocol*

OF Obliquity Factor

OTK *On-The-Fly*

P Precision code

PDOP *Position Dilution Of Precision*

PLL Phase Lock Loop

PPP *Precise Point Positioning*

PPS *Precise Positioning Service / Pulse Per Second*

PRN Pseudo Random Noise

RMS *Root Mean Square*

RTK *Real Time Kinematic*

SBAS *Satellite Based Augmentation System*

SMR *Signal to Multipath power Ratio*

SPS *Standard Positioning Service*

SVN Space Vehicle Number

TCAR *Three-Carrier Ambiguity Resolution*

TDOP *Time Dilution Of Precision*

TEC *Total Electron Content*

TECU/TECV *Total Electron Content Units / Total Electron Content Vertical*

TLM TeLeMetry

TOA Time Of Arrival

TOW Time Of Week

TZ Transformada Z

UERE *User Equivalent Range Error*

UHF Ultra High Frequency

UT1 *Universal Time 1*

VCO Voltage Controlled Oscillator

VDOP *Vertical Dilution Of Precision*

Bibliografía

"Wide-Area, Carrier-Phase Ambiguity Resolution Using a Tomographic Model of the Ionosphere". *Navigation* (Spring, 2002).

Betz, J. W.; Kolodziejcki, K. R. (2009) *Generalized theory of code tracking with an Early-Late discriminator*. Part I: lower bound and coherent processing. *IEEE J AES*, 45(4):1538-1550.

Bisnath, S. B.; Gao, Y. (2009). "Innovation: Precise Point Positioning". *GPS World* (abril).

Borre, K.; Akos, D. M.; Bertelsen, N.; Rinder, P.; Jensen, S. H. (2007). *A Software-Defined GPS and Galileo Receiver*. Birkhäuser.

Cebrián, J. M.; Picanyol, J.; González, L.; Lavín, C.; Gago, O.; Amarillo, F.; Seco-Granados, G.; Vicario, J. L.; Barceló, M.; Mañosas-Caballú, M.; Antreich, F.; Basta, N.; Cuntz, M.; Heckler, M. V. T.; Sgammini, M. (2010). "ADIBEAM: Design and Experimental Validation of a Robust Beamforming Platform for Galileo Reference Ground Stations". *Proc. 5th ESA Workshop on Satellite Navigation User Equipment Technologies (NAVITEC)*.

Corbasí, A. (1998). *Sistemas de Navegación. Desde el Compás Magnético a la Navegación por Satélite*. McGraw Hill.

Counselman, C. C.; Gourevich, S. (1981). "Miniature Interferometer Terminals for Earth Surveying: Ambiguity and Multipath with Global Positioning System". *IEEE Transactions on Geosciences and Remote Sensing* (vol. 91, núm. 4).

Van Diggelen, F. (2009). *A-GPS: Assisted GPS, GNSS, and SBAS*. Artech House.

Forssell, B.; Martín-Neira, M.; Harris, R. A. (1997). "Carrier Phase Ambiguity Resolution in GNSS-2". *ION-GPS*.

Forssell, B. (2008). *Radionavigation Systems*. Artech House.

Gardner, F. M. (2005). *Phaselock techniques*. John Wiley and Sons, 3a edición.

Gernot, C. (2009) *Development of combined GPS L1/L2C acquisition and tracking methods for weak signals environments*. PhD thesis, Dept. of Geomatics Engineering, University of Calgary.

Hassibi, A.; Boyd, S. (1998). "Integer parameter estimation in linear models with applications to GPS". *IEEE Transactions on Signal Processing*.

Hernández-Pajares, M.; Juan, J. M.; Sanz, J.; García-Rodríguez, A.; Colombo, O. L. (2004). "Wide Area Real Time Kinematics with Galileo and GPS Signals". *Proc. ION GNSS'2004*. Long Beach, USA.

Jaime, R. (1995). *Radionavegació*. Edicions UPC.

Juang, J. C.; Chen, Y. H. (2009). *Phase/frequency tracking in a GNSS software receiver*. *IEEE J JSTSP*, 3(4):651-660.

Jury, E. I. (1964). *Theory and applications of the z-transform method*. R. E. Krieger Publishing.

Kaplan, E. D.; Hegarty, C. J. (2006). *Understanding GPS Principles and Applications* (2ª. ed.). Artech House.

Lenstra, A. K.; Lenstra, H. W.; Lovász, L. (1982). "Factoring polynomials with rational coefficients". *Mathematische Annalen* (261 (4), pág. 515-534).

López-Salcedo, J. A.; Capelle, Y.; Toledo, M.; Seco-Granados, G.; Vicario, J.; Ku-brak, D.; Monnerat, M.; Mark, A. (2008). "DINGPOS: A Hybrid Indoor Navigation Platform for GPS and GALILEO". *Proc. ION GNSS*.

Mao, W. L.; Chen, A. B. (2008). Mobile GPS carrier phase tracking using a novel intelligent dual-loop receiver. *International Journal of Satellite Communications and Networking*, 26:119-139.

McGraw, G. A.; Braasch, M. S. (1999). "GNSS Multipath Mitigation Using Gated and High Resolution Correlator Concepts". *Proc. ION-NTM*.

Misra, P.; Enge, P. (2006). *Global Positioning System* (2ª. ed.). Ganga-jamuna Press.

- van Nee, R. D. J.** (1993). "Optimum DGPS Receiver Structures". *Proc of Second International Symposium on Differential Satellite Navigation Systems (DSNS)*.
- Oppenheim, A. V.; Schafer, R. W.; Buck, J. R.** (1999). *Discrete-Time Signal Processing*. Prentice Hall.
- Parkinson, B. W; Spilker, J. J. (eds.)** (1996). *Global Positioning System: theory and applications* (volumen I). American Institute of Aeronautics and Astronautics.
- Perdikaris, G. A.** (1996). *Computer controlled systems. Theory and applications*. Kluwer Academic.
- Psiaki, M. I** (2001). "Block Acquisition of Weak GPS Signals in a Software Receiver". *Proc. of ION-GNSS*.
- Schnorr, C.; Euchner, M.** (1994). "Lattice basis reduction: Improved practical algorithms and solving subset sum problems". *Mathematical Programming* (vol. 66, pág. 181-199).
- Seber, G. A. F; Wild, C. J.** (1989). *Nonlinear Regression*. John Wiley & Sons.
- Seco-Granados, G.; Swindlehurst, A. L.; Astely, D.** (2000). "Exploiting Antenna Arrays for Synchronization". En G. B. Giannakis, Y. Hua, P. Stoica, L. Tong (eds.). *Signal Processing Advances in Wireless and Mobile Communications* (vol. 2: Trends in Single- and Multi-User Systems, pág. 403-426).
- Seco-Granados, G.; López-Salcedo, J. A.; Jiménez-Baños, D.; López-Risueño, G.** (2012). "Signal Processing Challenges in Indoor GNSS". *IEEE Signal Processing Magazine*.
- Simon, M. K.; Omura, J. K.; Scholtz, R. A.; Levitt, B. K.** (2002). *Spread spectrum communications handbook*. McGraw-Hill.
- Strang, G.; Borre, K.** (1997). "Linear Algebra, Geodesy, and GPS". *Wellesley College*.
- Teunissen, P. J. G.** (1995). "The Least-Squares Ambiguity Decorrelation Adjustment: a Method for Fast GPS Integer Ambiguity Estimation". *Journal of Geodesy* (vol. 70, pág. 65-82).
- Tsui, J. B. Y.** (2004). *Fundamentals of Global Positioning System Receivers*. Wiley & Sons.

